

ESTATÍSTICA APLICADA

FILIFE ANTÓNIO MAHALUÇA






2016

APRESENTAÇÃO

A Estatística é uma ferramenta imprescindível a qualquer pesquisador ou pessoa que necessite tomar decisões. O seu estudo não representa uma tarefa muito fácil, principalmente no início, quando são apresentados muitos conceitos novos que exigem um tipo especial de raciocínio.

Uma boa base teórica é importante e necessária para que o estudo da Estatística seja prazeroso e não muito sofrido.

Com o intuito de facilitar a apresentação e aprendizado da Estatística desenvolveu-se este material para servir de apoio didático às disciplinas de Estatística, apresentada aos estudantes das áreas de:

-  Ciências Contabilísticas e económicas
-  Ciências Sociais e Humanas
-  Ciências Naturais
-  Ciências de Saúde
-  Engenharias

Filipe António Mahaluça

(Statistic and MSc in Health Sciences)

Contact: +258-848407234

*Às vezes construímos sonhos em cima de
grandes pessoas....O tempo passa e
descobrimos que grandes mesmo eram os
sonhos e as pessoas pequenas demais para
torna-los reais!*

- Bob Marley-

Índice

1. INTRODUÇÃO.....	1
1.1. Evolução Histórica de Estatística	1
2. NOÇÕES SOBRE AMOSTRAGEM	4
2.1. Estatística	4
2.2. População e amostra	4
2.3. Técnicas de Amostragem.....	5
2.3.1. Amostra aleatória, casual, ou probabilística.....	6
2.3.2. Amostragem Semi-probabilística	7
2.3.3. Amostra não probabilística	9
2.4. EXERCÍCIOS PROPOSTOS	11
3. APRESENTAÇÃO DE DADOS.....	14
3.1. Dados e Variável.....	14
3.1.1. Classificação de variáveis	14
3.2. Fases do método estatístico	16
3.3. Apresentação de dados em tabelas.....	17
3.3.1. Tabelas de distribuição de frequências	17
3.4. Apresentação de dados em gráficos.....	21
3.4.1. Gráfico de sectores	21
3.4.2. Gráfico de Barras.....	22
3.4.3. Histograma.....	23
3.4.4. Polígono de frequências	24
3.4.5. Polígono de frequências acumuladas (OGIVA).....	25
3.5. EXERCÍCIOS PROPOSTOS	26
4. MEDIDAS DESCRITIVAS.....	33
4.1. Medidas de Localização.....	33
4.1.1. Medidas de Tendência Central.....	33
4.1.2. Medidas de Tendência não Central ou Separatizes.....	42

4.2.	Medidas de Dispersão ou Variabilidade	46
4.2.1.	Amplitude total	47
4.2.2.	Momentos Centrais.....	47
4.2.3.	Desvio Médio	48
4.2.4.	Desvio Padrão	48
4.2.5.	Coeficiente de Variação.....	50
4.2.6.	Intervalo interquartil.....	52
4.3.	Medidas de Assimetria	52
4.4.	Medidas de Achatamento ou Curtose	53
4.5.	EXERCICIOS PROPOSTOS	55
5.	MÉTODO DE CONTAGEM	68
5.1.	Factorial.....	68
5.2.	Princípios fundamentais.....	68
5.3.	Permutações.....	70
5.3.1.	Permutações de objectos distintos.....	70
5.3.2.	Permutações de objectos nem todos distintos.....	71
5.3.3.	Permutações de objectos distintos com repetições	72
5.4.	Combinações.....	73
5.4.1.	Combinações de objectos distintos.....	73
5.4.2.	Triângulo de Pascal.....	74
5.5.	EXERCÍCIOS PROPOSTOS	74
6.	PROBABILIDADES	78
6.1.	Experimento aleatório.....	78
6.2.	Espaço amostral	78
6.3.	Eventos	79
6.4.	Definições de probabilidade.....	80
6.4.1.	Probabilidade clássica.....	80
6.4.2.	Probabilidade frequentista.....	81

6.4.3.	Probabilidade Geométrica	81
6.4.4.	Probabilidade subjectiva	81
6.4.5.	Teoria dos conjuntos: revisão de conceitos	82
6.4.6.	Axiomas	82
6.5.	Principais Teoremas	82
6.5.1.	Probabilidade do vazio.....	82
6.5.2.	Probabilidade da união finita de eventos disjuntos 2 a 2	83
6.5.3.	Probabilidade do evento complementar	83
6.5.4.	Probabilidade de eventos aninhados	83
6.5.5.	Probabilidade entre 0 e 1	84
6.5.6.	Probabilidade da subtração	84
6.5.7.	Desigualdade de Boole	84
6.5.8.	Probabilidade da união de 2 eventos	85
6.5.9.	Probabilidade da união de 3 eventos	85
6.5.10.	Probabilidade da união finita	86
6.6.	Probabilidade condicional e principais teoremas	86
6.6.1.	Probabilidade Condicional	86
6.6.2.	Teorema de multiplicação.....	88
6.6.3.	Partição de um conjunto	88
6.6.4.	Teorema de probabilidade total.....	89
6.6.5.	Teorema de Bayes	90
6.6.6.	Sensibilidade, Especificidade, Valor Preditivo Positivo e Negativo.....	91
6.7.	Independência	92
6.7.1.	Independência entre dois eventos.....	92
6.7.2.	Independência de eventos dois a dois	93
6.7.3.	Independência mútua	93
6.8.	EXERCÍCIOS PROPOSTOS	93
7.	VARIÁVEL ALEATÓRIA DISCRETA	112

7.1.	Variável aleatória discreta.....	112
7.1.1.	Função de probabilidade (f.d.p)	113
7.1.2.	Função Distribuição Acumulada.....	114
7.1.3.	Esperança Matemática ou Valor Esperado.....	115
7.1.4.	Variância	116
7.2.	EXERCÍCIOS PROPOSTOS	117
8.	ALGUMAS DISTRIBUIÇÕES DISCRETAS.....	122
8.1.	Distribuição Uniforme Discreta	123
8.1.1.	Esperança e Variância.....	123
8.2.	Distribuição de Bernoulli	124
8.2.1.	Esperança e Variância.....	124
8.3.	Distribuição de Binomial.....	125
8.3.1.	Esperança e Variância.....	125
8.4.	Distribuição Geométrica	126
8.4.1.	Esperança e Variância.....	127
8.5.	Distribuição Binomial Negativa	127
8.5.1.	Esperança e Variância.....	128
8.6.	Distribuição Hipergeométrica.....	128
8.6.1.	Esperança e Variância.....	129
8.7.	Distribuição de Poisson.....	129
8.7.1.	Esperança e Variância.....	130
8.8.	EXERCÍCIOS PROPOSTOS	131
9.	VARIÁVEL ALEATÓRIA CONTÍNUA	141
9.1.	Função densidade de probabilidades (f.d.a)	141
9.2.	Função distribuição acumulada (f.d.a)	141
9.3.	Esperança de variáveis aleatórias contínuas	142
9.4.	Variância de variáveis aleatórias contínuas	142
9.5.	Propriedades da média e variância.....	142

9.6.	EXERCÍCIOS PROPOSTOS	144
10.	ALGUMAS DISTRIBUIÇÕES CONTÍNUAS	147
10.1.	Distribuição Uniforme	147
10.1.1.	Função de distribuição	147
10.1.2.	Esperança e Variância	147
10.2.	Distribuição Exponencial	148
10.2.1.	Função de distribuição	148
10.2.2.	Esperança e Variância	149
10.3.	Distribuição Normal.....	149
10.3.1.	Densidade normal padrão.....	150
10.3.2.	Esperança e Variância.....	151
10.3.3.	Tabela Normal Padrão.....	151
10.3.4.	Cálculos com a distribuição normal	153
10.3.5.	Encontrando a abcissa da normal para uma probabilidade específica	154
10.3.6.	Exemplos de Aplicação da Distribuição Normal	156
10.4.	EXERCÍCIOS PROPOSTOS	157
11.	DISTRIBUIÇÃO AMOSTRAL E ESTIMAÇÃO PONTUAL DE PARÂMETROS	164
11.1.	Inferência Estatística.....	164
11.2.	Técnicas de Amostragem (Revisão)	164
11.3.	Planeando um experimento	164
11.4.	Erro Amostral	164
11.5.	Erro Não-Amostral	165
11.6.	Estimação de Parâmetros.....	165
11.7.	Conceitos Fundamentais	166
11.7.1.	Amostra aleatória	166
11.7.2.	Parâmetro	166
11.7.3.	Estatísticas	166
11.7.4.	Distribuição amostral.....	166

11.7.5.	Espaço paramétrico	167
11.7.6.	Estimador	167
11.8.	Teorema do Limite Central.....	168
11.9.	Propriedades de um estimador.....	169
11.9.1.	Estimador não viciado ou não tendencioso	169
11.9.2.	Estimador Consistente.....	171
11.9.3.	Estimador Eficiente	172
11.10.	EXERCÍCIOS PROPOSTOS	172
12.	INTERVALO DE CONFIANÇA	177
12.1.	Estimação por intervalo.....	177
12.2.	Nível de confiança	177
12.3.	Valor crítico.....	178
12.4.	Margem do erro	178
12.5.	Intervalo de confiança para uma média.....	178
12.5.1.	Intervalo de confiança para uma média quando a variância é conhecida	178
12.5.2.	Intervalo de confiança para uma média quando a variância é desconhecida (n grande) 182	
12.5.3.	Intervalo de confiança para uma média quando a variância é desconhecida (n pequeno) 183	
12.5.4.	Intervalo de confiança para a diferença entre duas médias quando as variâncias são conhecidas.....	185
12.5.5.	Intervalo de confiança para a diferença entre duas médias quando as variâncias são desconhecidas e iguais.....	186
12.5.6.	Intervalo de confiança para a diferença entre observações.....	187
12.5.7.	Intervalo de confiança para a variância	189
12.5.8.	Intervalo de confiança para quociente entre duas variâncias	191
12.5.9.	Intervalo de confiança para uma proporção.....	192
12.5.10.	Intervalo de confiança para a diferença entre as proporções.....	195
12.6.	EXERCÍCIOS PROPOSTOS	196

13.	TESTE DE HIPÓTESES.....	217
13.1.	Hipótese estatística	217
13.2.	Estatística de teste.....	218
13.3.	Regra de decisão estatística	218
13.4.	Erros de inferência	218
13.4.1.	Relação entre os erros tipo I e tipo II	219
13.5.	Valor de p ou $p - value$	220
13.6.	Região de Rejeição ou região crítica (RC).....	220
13.7.	Passos de um teste de hipóteses através da RC	222
13.8.	Teste de hipótese para média, quando a variância é conhecida.....	222
13.9.	Teste de hipótese para média, quando a variância é desconhecida.....	224
13.10.	Teste de hipótese da diferença entre duas médias populacionais	226
13.10.1.	Teste de hipótese da diferença entre duas médias populacionais com variâncias conhecidas	227
13.10.2.	Teste de hipótese da diferença entre duas médias populacionais com variâncias desconhecidas e iguais.....	228
13.10.3.	Teste de hipótese da diferença entre duas médias populacionais com variâncias desconhecidas e diferentes.....	229
13.11.	Teste de hipótese de duas amostras emparelhadas.....	231
13.12.	Teste de hipótese para variância numa população normal	232
13.13.	Teste de hipóteses para variâncias de duas populações.....	234
13.14.	Teste de hipóteses para proporção.....	236
13.15.	Teste de hipóteses para duas proporções	238
13.16.	Testes não-paramétricos	241
13.16.1.	Teste de independência estatística	241
13.17.	EXERCÍCIOS PROPOSTOS	245
14.	CORRELAÇÃO E REGRESSÃO LINEAR	266
14.1.	Introdução	266
14.2.	Coeficiente de Correlação de Pearson.....	266

14.2.1. Teste de significância do coeficiente de correlação de Pearson.....	268
14.3. Análise de Regressão.....	270
14.3.1. Regressão Linear Simples	270
14.3.2. Método dos Mínimos Quadrados.....	272
14.3.3. Estimativa da variância do termo erro.....	273
14.3.4. A Adequação do modelo de regressão linear ajustado	273
14.3.5. Interpretação dos parâmetros do modelo	275
Referências Bibliográficas	276
Apêndice.....	277

1. INTRODUÇÃO

1.1. Evolução Histórica de Estatística

As primeiras aplicações da estatística estavam voltadas para as necessidades de Estado, na formulação de políticas públicas, fornecendo dados demográficos e económicos à administração pública. A abrangência da estatística aumentou no começo do século XIX para incluir a acumulação e análise de dados de maneira geral. Hoje, a estatística é largamente aplicada nas ciências naturais, e sociais, inclusive na administração pública e privada.

Seus fundamentos matemáticos foram postos no século XVII com o desenvolvimento da teoria das probabilidades por Pascal e Fermat, que surgiu com o estudo dos jogos de azar. O método dos mínimos quadrados foi descrito pela primeira vez por Carl Friedrich Gauss cerca de 1794. O uso de computadores modernos tem permitido a computação de dados estatísticos em larga escala e também tornaram possível novos métodos antes impraticáveis a nossa sociedade.

O termo estatística deriva do *neolatim statisticum collegium* ("conselho de Estado") e do Italiano *statista* ("estadista" ou "político"). O alemão *Statistik*, introduzido pela primeira vez por Gottfried Achenwall (1749), designava originalmente a análise de dados sobre o Estado, significando a "ciência do Estado" (então chamada aritmética política (*political arithmetic*) em inglês). A palavra adquiriu o significado de colecta e classificação de dados em geral através de Sir John Sinclair.

Assim, o propósito original da *Statistik* era fornecer os dados a serem usados pelo governo e outras organizações. A colecta de dados sobre estados e localidades continua, em grande parte através de órgãos estatísticos nacionais e internacionais. Em particular, os censos fornecem informação regular sobre as populações.

Os métodos matemáticos da estatística emergiram da teoria das probabilidades, que remonta à correspondência entre Pierre de Fermat e Blaise Pascal (1654). Christiaan Huygens (1657) deu o tratamento científico mais antigo que se conhece sobre o assunto. A obra póstuma *Ars Conjectandi* (1713) de Jakob Bernoulli e Abraham de Moivre, *The Doctrine of Chances* (1718) tratou o assunto como um ramo da matemática. Na era moderna, a obra de Kolmogorov tem sido útil na formulação dos modelos fundamentais da teoria das probabilidades, imprescindíveis à estatística.

A teoria dos erros remonta à obra póstuma *Opera Miscellanea* (1722) de Roger Cotes, mas uma edição de memórias preparada por Thomas Simpson em 1755 (impressa em 1756) aplicou pela primeira vez a teoria à discussão dos erros na observação. A reimpressão (de 1757) dessas memórias estabelece o axioma de que erros positivos e negativos são igualmente prováveis, e que existem certos limites dentro dos quais todos os erros irão ocorrer; erros contínuos são discutidos e é fornecida uma curva de probabilidades.

Pierre-Simon Laplace (1774) fez a primeira tentativa de deduzir a regra para a combinação de observações dos princípios da teoria das probabilidades. Ele representou a lei das probabilidades dos erros através de uma curva. Ele deduziu uma fórmula para a média de três observações. Ele também deu (em 1781) uma fórmula para a lei de 'facilidade de erro' (um termo devido a Joseph Louis Lagrange, 1774), mas que levou a equações não tratáveis. Daniel Bernoulli (1778) introduziu o princípio do produto máximo de probabilidade de um sistema de erros concorrentes.

O método dos mínimos quadrados, que foi usado para minimizar erros na medição de dados, foi publicado independentemente por Adrien-Marie Legendre (1805), Robert Adrain (1808) e Carl Friedrich Gauss (1809). Gauss usou o método na sua famosa previsão de onde se localizava o planeta anão Ceres. Outras provas foram dadas por Laplace (1810, 1812), Gauss (1823), James Ivory (1825, 1826), Hagen (1837), Friedrich Bessel (1838), W. F. Donkin (1844, 1856), John Herschel (1850) e Morgan Crofton (1870).

Outros que contribuíram foram Ellis (1844), De Morgan (1864), Glaisher (1872) e Giovanni Schiaparelli (1875). A fórmula de Peters (1856) para r , o erro provável de uma única observação, é bastante conhecida.

No século XIX, autores que trataram da teoria geral incluem Laplace, Sylvestre Lacroix (1816), Littrow (1833), Richard Dedekind (1860), Helmert (1872), Hermann Laurent (1873), Liagre, Didion e Karl Pearson. Augustus De Morgan e George Boole fizeram melhorias na apresentação da teoria.

Adolphe Quetelet (1796-1874), outro importante fundador da estatística, introduziu a noção de "homem médio" (*l'homme moyen*) como um meio de compreender fenômenos sociais complexos como taxas de criminalidade, de casamento e de suicídio.

Durante o século XX, a criação de instrumentos precisos para a agronomia, saúde pública (epidemiologia, bioestatística, etc.), controle de qualidade industrial e propósitos económicos e sociais (taxa de desemprego, econometria, etc.) necessitavam avanços substanciais nas práticas estatísticas.

Hoje, a utilização da estatística se expandiu para muito além das suas origens. Indivíduos e organizações usam a estatística para compreender dados e tomar decisões bem-informadas nas ciências naturais e sociais, na medicina, nos negócios e em outras áreas.

A estatística é geralmente tida não como um ramo da matemática, mas como uma área distinta, ainda que intimamente relacionada. Muitas universidades mantêm departamentos separados de matemática e estatística.

2. NOÇÕES SOBRE AMOSTRAGEM

Grande parte das pessoas já ouviu falar de prévias eleitorais, de censo, de pesquisa de opinião. A maioria das pessoas já respondeu perguntas sobre a qualidade dos serviços de uma telefonia móvel, já assistiu na televisão programas em que pedem para o ouvinte ou telespectador votar em um cantor ou em uma música, ou dar opinião sobre determinado assunto por telefone ou por correio electrónico.

O uso tão difundido de levantamento de dados faz pensar que esse é um trabalho fácil. Por conta disso, ao ler um relatório de pesquisa no jornal da cidade, muita gente se acha capaz de fazê-lo, e até melhor, pois entende que, para levantar dados, basta fazer perguntas e depois contar as respostas. Mas não é bem assim. Um bom levantamento de dados exige conhecimentos de Estatística.

2.1. Estatística

Estatística é uma ciência que fornece os princípios e técnicas para colecta, organização, descrição, análise e interpretação de dados experimentais.

A estatística subdivide-se em:

- **Estatística descritiva:** Colecta, descrição e organização de dados experimentais
- **Estatística Inferencial:** Análise e interpretação de dados experimentais

2.2. População e amostra

População é o conjunto de indivíduos que possuem em comum mesmas características e com interesse para o estudo. Na estatística, a população é classificada como finita e infinita

- **População finita:** nesses casos o número de elementos de um grupo não é muito grande, a entrevista e a análise das informações devem abordar a todos do grupo. Por exemplo:
 - ✓ As condições das escolas particulares na cidade de Maputo. Se observarmos o grupo chegaremos à conclusão de que o número de escolas particulares em Maputo é considerado finito.

- **População infinita:** o número de elementos nesse caso é muito elevado, sendo considerado infinito. Por exemplo:
 - ✓ A população da cidade de Maputo.

Muitas vezes se torna difícil, ou até mesmo impossível, observar todo um grupo, especialmente se esse for muito grande. Nesses casos, podemos utilizar apenas uma parte desse, denominado **amostra**.

Amostra diz respeito a um subconjunto da população, fracção ou uma parte do grupo. Em alguns casos seria impossível entrevistar todos os elementos de uma população, pois levaria muito tempo para concluir o trabalho ou até mesmo seria financeiramente inviável, dessa forma, o número de entrevistados corresponde a uma quantidade determinada de elementos do conjunto, uma amostra.

A amostra deve ser representativa da população, isto é, deve conter todas as características, já que por meio dessa amostra serão tiradas as conclusões para toda a população.

As razões que levam os pesquisadores a trabalhar com amostras e não com toda a população são poucas, mas absolutamente relevantes.

- Custo e demora dos censos
- Populações muito grandes
- Impossibilidade física de examinar toda a população
- Comprovado valor científico das informações colectadas por meio de amostras

2.3. Técnicas de Amostragem

Antes de obter uma amostra, é preciso definir os critérios que serão usados para seleccionar as unidades que compõem essa amostra. De acordo com a técnica usada, tem-se um tipo de amostra. Serão definidas aqui:

- Amostragem aleatória, casual, ou probabilística;
- Amostragem semi-probabilística;
- Amostragem não-probabilística ou de conveniência.

2.3.1. Amostra aleatória, casual, ou probabilística

A amostra aleatória ou probabilística é constituída por n unidades retiradas ao acaso da população. Em outras palavras, a amostra aleatória é obtida por sorteio. Logo, toda unidade da população tem probabilidade conhecida de pertencer à amostra.

Para obter uma amostra aleatória, é preciso que a população seja conhecida e cada unidade esteja identificada por nome ou por número. Os elementos que constituirão a amostra são escolhidos por sorteio. Algumas pessoas acreditam que o sorteio: por computador é mais "sério", ou mais "exacto". Hoje em dia, é mais fácil. No entanto, o sorteio feito com papezinhos em uma caixa ou bolas em uma urna (usados em programas de televisão) ajuda entender as regras do procedimento aleatório.

Uma amostra aleatória pode ser:

- Simples
- Estratificada.

2.3.1.1. Amostragem aleatória Simples

A amostragem aleatória simples é obtida por sorteio de uma população constituída por unidades homogêneas para a variável que você quer estudar.

Exemplo 2.1: Uma amostragem aleatória simples

Imagine que você precisa obter uma amostra de 2% dos 500 pacientes de uma clínica para entrevistá-los sobre a qualidade de atendimento da secretária. Qual seria o procedimento para obter uma amostra aleatória simples?

Solução

Para obter uma amostra aleatória de 2% dos 500 clientes, você precisa sortear 10. Você pode fazer isso da maneira mais antiga e conhecida (e também a mais trabalhosa). Comece escrevendo o nome de todos os clientes em pedaços de papel. Coloque todos os pedaços de papel em uma urna, misture bem e retire um nome. Repita o procedimento até ter os nomes dos 10 clientes que comporão sua amostra.

2.3.1.2. Amostragem aleatória Estratificada

A amostra aleatória estratificada é usada quando a população é constituída por unidades heterogêneas para a variável que se quer estudar. Nesse caso, as unidades da população devem ser identificadas; depois, as unidades similares devem ser reunidas em subgrupos chamados estratos. O sorteio é feito dentro de cada estrato.

Exemplo 2.2: Uma amostragem aleatória Estratificada

Imagine que você precisa obter uma amostra de 2% dos 500 clientes de um banco de micro finanças para entrevistá-los sobre a qualidade de atendimento da secretária. Você suspeita que homens sejam mais bem atendidos do que mulheres. Aproximadamente metade dos clientes é do sexo masculino. Você quer obter dados dos dois sexos. Qual seria o procedimento?

Solução

Comece separando homens de mulheres. Você tem, então, dois estratos, um de homens, outro de mulheres. Depois você obtém uma amostra aleatória de cada sexo (ou cada estrato) e reúne os dados dos dois estratos numa só amostra aleatória estratificada.

2.3.2. Amostragem Semi-probabilística

A amostra semi-probabilística é constituída por n unidades retiradas da população por procedimento parcialmente aleatório. Dentre as amostras semiprobabilísticas, temos:

- *Amostra sistemática*
- *Amostra por conglomerados;*
- *Amostra por quotas.*

2.3.2.1. Amostragem sistemática

A amostra sistemática é constituída por n unidades retiradas da população segundo um sistema preestabelecido. Por exemplo, se você quiser uma amostra constituída por 1/8 da população, você sorteia um número que caia entre 1 e 8. Se for sorteado o número 3, por exemplo, a terceira unidade (número 3) será seleccionada para a amostra. A partir daí, tome, sistematicamente, a terceira unidade de cada oito, em sequência. No

caso do exemplo, a primeira unidade é 3. Seguem, de oito em oito, as unidades de números: 11, 19, 27 etc.

Exemplo 2.3: Amostragem sistemática

Imagine que você precisa obter uma amostra de 2% dos 500 clientes de um banco de micro finanças para entrevistá-los sobre a qualidade de atendimento da secretária. Como você obteria uma amostra sistemática?

Solução

Uma amostra de 2% dos 500 clientes significa amostra de tamanho 10. Para obter a amostra, você pode dividir 500 por 10, e obter 50. Sorteie então um número entre 1 e 50, inclusive. Se sair o número 27, por exemplo, esse será o número do primeiro cliente que será incluído na amostra. Depois, a partir do número 27, conte 50 e chame esse cliente. Proceda dessa forma até completar a amostra de 10 clientes.

2.3.2.2. Amostragem por conglomerados

A amostra por conglomerados é constituída por n unidades tomadas de alguns conglomerados. O conglomerado é um conjunto de unidades que estão agrupadas, qualquer que seja a razão. Um asilo é um conglomerado de idosos, uma universidade pública é um conglomerado de pessoas com bom nível socioeconómico, um serviço militar é um conglomerado de adultos jovens saudáveis. Como exemplo, imagine que um dentista quer levantar dados sobre a necessidade de aparelho ortodôntico em crianças de 12 anos. Ele pode sortear três escolas de primeiro grau (conglomerados) e examinar todas as crianças com 12 anos dessas escolas.

Exemplo 2.4: Amostragem por conglomerados

Um professor de Educação Física quer estudar o efeito da terapia de reposição hormonal (uso de hormônios por mulheres depois da menopausa) sobre o desempenho nos exercícios. Como obteria uma amostra por conglomerados?

Solução

O professor de Educação Física pode sortear duas academias de ginástica da cidade e avaliar o desempenho das mulheres que frequentam a academia e já tiveram a menopausa

(tanto as que fazem como as que não fazem uso da terapia de reposição hormonal) para posterior comparação.

2.3.2.3. Amostragem por quotas

A amostra por quotas é constituída por n unidades retiradas da população segundo quotas estabelecidas de acordo com a distribuição desses elementos na população. A ideia de quota é semelhante à de estrato, com uma diferença básica: você selecciona a amostra por julgamento e depois confirma as características das unidades amostradas.

A amostragem por quotas não é aleatória, embora muitos pensem que é. A grande vantagem é ser relativamente barata. Por esta razão, é muito usada em levantamentos de opinião e pesquisas de mercado.

Exemplo 2.5: Amostragem por quotas

Considere uma pesquisa sobre a preferência de modelo de carro. Como se faz uma amostragem por quotas?

Solução

Você possivelmente irá entrevistar homens e mulheres com mais de 18 anos que vivem em uma cidade (por exemplo, Matola), na proporção apresentada pelo censo demográfico em termos de sexo, idade e renda. Você então sai às ruas para trabalhar com a incumbência de entrevistar determinada quota de pessoas com determinadas características.

Por exemplo, você pode ser incumbido de entrevistar 30 homens com "mais de 50 anos que recebam mais de seis e menos de 10 salários mínimos". Então você deverá julgar pela aparência da pessoa, se ela se enquadra nas características descritas homem de mais de 50 anos que ganha entre seis e 10 salários mínimos. Se achar que viu a pessoa certa, deve fazer a abordagem e depois confirmar as características com perguntas. O número de pessoas em determinada quota depende do número delas na população.

2.3.3. Amostra não probabilística

As amostragens não probabilísticas utilizam-se em três tipos de situações:

- Estudos em grupos cujos elementos são difíceis de identificar e contactar (por exemplo, membros de gangs juvenis);
- Estudos com grupos específicos em que razões éticas impedem que se identifiquem todos os elementos desses grupos, pelo que se entrevistam apenas voluntários (por exemplo, sujeitos portadores de determinada doença);
- Investigações em situações piloto – (por exemplo, sujeitos que participam numa acção de formação)

As principais técnicas de amostragem não probabilística podem ser intencionais e não intencionais (ou de conveniência):

2.3.3.1. Amostragens intencionais

1. **Amostragem Bola-de-neve:** O entrevistador estabelece contacto inicial com alguns sujeitos previamente identificados como membros do grupo que se pretende estudar e estes sujeitos põem o investigador em contacto com outros membros desse grupo e assim sucessivamente (por exemplo, consumidores de drogas que indicam ao investigador outros consumidores que aceitam participar no estudo).
2. **Amostragem por quotas**
3. **Focus Grupo:** Consiste na entrevista a um pequeno grupo de sujeitos- 10 a 20 elementos – escolhidos pelo investigador por serem representativos de uma população particular (jovens, potenciais consumidores de um produto, membros de uma profissão, etc.
4. **Casos críticos**

2.3.3.2. Amostragens não intencionais ou de conveniência

Entrevistam-se sujeitos a que se tem acesso imediato e directo (por exemplo, estudantes de uma turma de que o investigador é professor, doentes de um determinado serviço hospitalar, ...)

Exemplo 2.6: Amostragem por quotas

Imagine que um nutricionista quer entrevistar 50 mães de crianças com idades de 3 e 4 anos para conhecer os hábitos alimentares dessas crianças. Como obteria essa amostra?

Solução

Se o nutricionista trabalha em uma escola, para obter a amostra de 50 mães de crianças de 3 e 4 anos, provavelmente procurará as mães de crianças matriculadas na escola em que trabalha.

2.4. EXERCÍCIOS PROPOSTOS

- 2.4.1.** Os prontuários dos pacientes de um hospital estão organizados em um arquivo, por ordem alfabética. Qual é a maneira mais rápida de amostrar $1/3$ do total de prontuários?
- 2.4.2.** Um pesquisador tem 10 gaiolas, cada uma com seis ratos. Como o pesquisador pode seleccionar 10 ratos para uma amostra?
- 2.4.3.** Para levantar dados sobre o número de filhos por mulher, em uma comunidade, um pesquisador organizou um questionário que enviou, pelo correio, a todas as residências. A resposta ao questionário era facultativa, pois o pesquisador não tinha condições de exigir a resposta. Nesse questionário perguntava-se o número de filhos por mulher moradora na residência. Você acha que os dados assim obtidos seriam tendenciosos?
- 2.4.4.** Um pesquisador pretende levantar dados sobre o número de moradores por domicílio, usando a técnica de amostragem sistemática. Para isso, o pesquisador visitará cada domicílio seleccionado. Se nenhuma pessoa estiver presente na ocasião da visita, o pesquisador excluirá o domicílio da amostra. Esta última determinação torna a amostra tendenciosa. Por quê?
- 2.4.5.** Muitas pessoas acreditam que as famílias se tornaram menores. Suponha que, para estudar essa questão, um pesquisador seleccionou uma amostra de 2.000 casais e perguntou quantos filhos eles tinham, quantos filhos tinham seus pais e quantos filhos tinham seus avós. O procedimento produz dados tendenciosos. Por quê?
- 2.4.6.** Para estudar atitudes religiosas, um sociólogo sorteia 10 membros de uma grande igreja para compor uma amostra casual simples. Nota, então, que a

amostra ficou composta por nove mulheres e um homem. O sociólogo se espanta: "A amostra não é aleatória! Quase só tem mulher." O que você diria?

- 2.4.7.** Para avaliar a expectativa de pais de adolescentes em relação às possibilidades de estudo de seus filhos; foram distribuídos 5.000 questionários pelos distritos da província de Zambézia. Retornaram 1.032. Cerca de 60% dos respondentes diziam que a maior preocupação deles era com o preço que se paga para um jovem frequentar a universidade. Você considera esse resultado uma boa estimativa para o número de pais preocupados com essa questão?
- 2.4.8.** Um dentista quer levantar o tipo de documentação que seus colegas arquivam, quando fazem um tratamento ortodôntico. A documentação depende do caso, mas também envolve questões legais e de bom senso do ortodontista. Para essa pesquisa, o dentista elabora um questionário que envia, por correio, a todos os profissionais inscritos no conselho de odontologia. O dentista provavelmente não receberá respostas de todos. Você saberia dizer algumas das razões de isso acontecer?
- 2.4.9.** Para estudar o uso de serviços de saúde por mulheres em idade reprodutiva moradoras de uma grande capital, um pesquisador buscou no Instituto Nacional de Estatística (INE) as subdivisões da idade utilizadas em censos, conhecidas como sectores censitários. Como você procederia para tomar uma amostra de mulheres moradoras nesses sectores e em idade reprodutiva?
- 2.4.10.** Um editor de livros técnicos quer saber se os leitores preferem capas de cores claras com desenhos, ou capas simples de cores mais escuras. Se o editor pedir a você para estudar a questão, como você definiria a população do estudo?
- 2.4.11.** Para cada um dos seguintes casos, caracterize a população quanto ao tamanho.
- a) Conjunto de pessoas que chegam numa repartição bancária num determinado dia.
 - b) Conjunto de parafusos defeituosos produzidos por uma máquina.
 - c) Conjunto de raparigas matriculadas na ISCAM.
- 2.4.12.** Uma empresa está interessada em testar a eficácia da propaganda de um novo comercial de televisão. Como parte do teste, o comercial é mostrado em um programa de notícias locais, às 18h 30min. Dois dias mais tarde, uma firma de

pesquisa de mercados realizou um levantamento telefónico para obter informações sobre os índices de respostas (percentagem de telespectadores que responderam ter visto o comercial) e impressões sobre o comercial.

- a) Qual é a população desse estudo?
- b) Qual é a amostra para esse estudo?
- c) Por que se usaria uma amostra nessa situação? Explique.

2.4.13. A maneira de fazer a pergunta pode influenciar a resposta da pessoa que responde. Basicamente, existem dois tipos de questões: a “questão fechada” e a “questão aberta “. Na “questão fechada” o pesquisador fornece uma série de respostas possíveis e a pessoa que responde deve apenas assinalar a alternativa, ou as alternativas, que lhe convém. A “questão aberta” deve ser respondida livremente. Imagine que um dentista quer levantar dados sobre hábitos de higiene oral das pessoas de uma comunidade. Escreva então uma “questão fechada” e uma “questão aberta”.

2.4.14. Em uma pesquisa de mercado para serviços de telefonia móvel tomou-se a lista telefónica, onde os nomes dos assinantes estão organizados em ordem alfabética do último sobrenome, e se amostrou o décimo de cada 10 assinantes. Critique esse procedimento.

3. APRESENTAÇÃO DE DADOS

Você já aprendeu que os estatísticos colectam informações. Essas informações podem ser sobre peso de pessoas, eficiência dum serviço, incidência de doenças, causas de acidentes, quantidade de carros acidentados, etc. Neste Capítulo vamos aprender como essas informações são organizadas para facilitar a leitura. Mas antes vamos aprender o que são dados e o que são variáveis.

3.1. Dados e Variável

Variável é uma condição ou característica das unidades da população; a variável pode assumir valores diferentes em diferentes unidades. Por exemplo, a idade das pessoas residentes em Moçambique é uma variável. **Dados** são os valores da variável em estudo, obtidos por meio de uma amostra.

Exemplo 3.1: Dados e Variáveis

O dono de um supermercado quer saber a opinião de seus clientes sobre a qualidade dos serviços que presta. O que é variável e o que são dados nesse problema?

Solução

A variável de interesse é a opinião dos clientes. Os dados serão obtidos somente quando o dono do supermercado começar a pedir aos clientes que dêem uma nota a cada serviço. Então, se for pedido que o cliente dê uma nota de zero e 5 a cada serviço que utiliza os dados colectados poderão ser, por exemplo, 4, 3, 2, 4, 1 etc., por serviço.

3.1.1. Classificação de variáveis

As variáveis são classificadas em dois tipos:

- Quantitativas ou numéricas;
- Qualitativas ou categorizadas.

Uma variável é qualitativa ou categorizada quando os dados são distribuídos em categorias mutuamente exclusivas. São exemplos de variáveis qualitativas: marca de automóvel (Toyota, Mazda, etc.); sexo (Masculino ou Feminino);

Uma variável é quantitativa ou numérica quando é expressa por números. São exemplos de variáveis quantitativas: idade, altura, número de crianças numa escola, número de lápis numa caixa.

As variáveis qualitativas ou categorizadas são classificadas em dois tipos:

- Nominal
- Ordinal

A variável é nominal quando os dados são distribuídos em categorias mutuamente exclusivas, mas são indicadas em qualquer ordem. São variáveis nominais: cor de cabelos (loiro, castanho, preto, ruivo), tipo de sangue (O, A, B, AB), género (masculino, feminino), etc.

A variável é ordinal quando os dados são distribuídos em categorias mutuamente exclusivas que têm ordenação natural. São variáveis ordinais: escolaridade (primeiro grau, segundo grau, terceiro grau), classe social (A, B, C, D, E), gravidade de uma doença (leve, moderada, severa) etc.

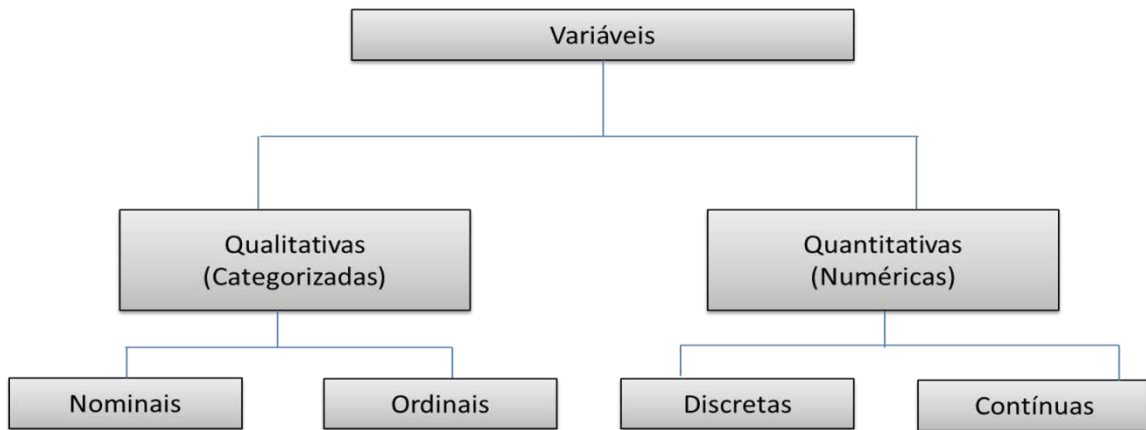
As variáveis quantitativas ou numéricas são classificadas em dois tipos:

- Discreta
- Contínua.

A variável discreta só pode assumir alguns valores em um dado intervalo. São variáveis discretas: número de filhos (0, 1, 2, 3, 4 etc.), quantidade de moedas num bolso (zero, 1, 2, 3 etc.), número de pessoas numa sala.

A variável contínua assume qualquer valor num dado intervalo. São variáveis contínuas: peso, tempo de espera, quantidade de chuva etc.

Os dados são do mesmo tipo que o das variáveis. Por exemplo, uma variável discreta produz dados discretos. Veja o organograma:



3.2. Fases do método estatístico

Num estudo estatístico, normalmente, segue-se um conjunto de passos que se designam por fases do método estatístico, a saber:

- Definição de problema: A primeira fase consiste na definição e formulação correcta do problema a ser estudado;
- Planificação: Definido o problema, é preciso determinar um processo para o resolver e, em especial, como obter informações sobre a variável em estudo. é nesta fase que se decide pela observação de toda a população ou de uma amostra
- Recolha de dados: Os dados podem ser recolhidos através de :
 - ✓ Questionários
 - ✓ Observação
 - ✓ Experimentação
 - ✓ Pesquisa Bibliográfica
- Organização de dados: Há duas formas de apresentação que não excluem mutuamente:
 - ✓ Apresentação por tabelas
 - ✓ Apresentação por gráficos
- Análise e interpretação de dados: Nesta fase calculam-se novos números com base nos dados estatísticos. Estes novos números permitem fazer uma descrição do fenómeno evidenciando algumas das suas características

3.3. Apresentação de dados em tabelas

Tabela: é um quadro que resume um conjunto de observações. Uma tabela compõe-se de:

- Corpo: conjunto de linhas e colunas que contém informações sobre a variável em estudo;
- Cabeçalho: parte superior da tabela que especifica o conteúdo das colunas;
- Coluna indicadora: parte da tabela que especifica o conteúdo das linhas;
- Linhas: rectas imaginárias que facilitam a leitura, no sentido horizontal, de dados que se inscrevem nos seus cruzamentos com as colunas;
- Casa ou célula: espaço destinado a um só número.
- Título: conjunto de informações, as mais completas possíveis, respondendo a perguntas como: O quê? Quando? Onde? Localizado no topo da tabela.

Exemplo 3.2: Tabelas de entrada simples

Produção de castanha Moçambique -1991-1995	
Anos	Produção (1000 ton.)
1991	2535
1992	2666
1993	2122
1994	3750
1995	2007

Exemplo 3.3: Tabelas de dupla entrada

Produção de castanha Moçambique -1991-1995		
Anos	Produção (1000 ton.)	Exportação (1000 ton.)
1991	2535	1478
1992	2666	1987
1993	2122	1241
1994	3750	2136
1995	2007	1729

3.3.1. Tabelas de distribuição de frequências

Dá-se o nome de distribuição de frequências ao conjunto de todos os valores de uma variável estatística com as correspondentes frequências:

- **Frequências absolutas:** é o número de vezes que esse valor foi observado;
- **Frequências relativas:** é o quociente entre a frequência absoluta da variável e o número total de observações

As distribuições de frequências podem-se classificar:

- Ordinárias: a cada valor ou classe de valores da variável corresponde a sua frequência;
- Acumulada: a cada valor ou classe de valores da variável corresponde a sua frequência mais a de todos os valores, ou classes de valores anteriores (ou posteriores).

A sua disposição prática é designada por quadro de frequências (dados agrupados e dados agrupados em intervalo de classes)

Exemplo 3.4: Distribuição de dados agrupados

Foram examinados 100 lotes de 50 peças produzidas por uma máquina, para verificação do número de peças defeituosas por lote. Os resultados apresentam-se no seguinte quadro:

Nº de peças defeituosas por lote	Nº de lotes
0	3
1	11
2	21
3	30
4	23
5	7
6	5
Total	100

Represente os dados em frequências absolutas e relativa

Solução

Valor da variável	Frequências Absolutas		Frequências Relativas	
	Ordinárias (f_i)	Acumuladas (F_i)	Ordinárias (f_r)	Acumuladas (F_r)
0	3	3	0.03	0.03
1	11	14	0.11	0.14
2	21	35	0.21	0.35
3	30	65	0.30	0.65
4	23	88	0.23	0.88
5	7	95	0.07	0.95
6	5	100	0.05	1.00
Total	100		1.00	

Exemplo 3.5: Distribuição de dados agrupados em intervalo de classe

Utilizando os dados apresentados no exemplo 3.3, podemos agrupar os valores da variável nas classes [0---2], [3---4], [5---6], obtendo a distribuição de valores da variável agrupados em intervalos de classe:

Classes de valor	Frequências Absolutas		Frequências Relativas	
	Ordinárias (f_i)	Acumuladas (F_i)	Ordinárias (f_r)	Acumuladas (F_r)
[0---2]	35	35	0.35	0.35
[3---4]	53	88	0.53	0.88
[5---6]	12	100	0.12	1.00
Total	100		1.00	

Neste caso de distribuição de frequências devemos considerar outros elementos e conceitos além dos mencionados anteriormente

- a) **Intervalo de variação da variável x:** é o intervalo que contém todos os valores da variável x, isto é:

$$I = [\min\{x_i\}; \max\{x_i\}]$$

- b) **Classes:** são intervalos cuja reunião contém o intervalo de variação da variável observada.

$$I_1 = [l_1; l_2[, [l_2; l_3[, \dots, [l_k; l_{k+1}[,$$

onde $l_1 < l_2 < l_3 < \dots < l_k < l_{k+1}$

- c) **Amplitude da classe:**

$$a_i = l_{i+1} - l_i$$

- d) **Centro da classe:** é o ponto médio do intervalo, isto é,

$$x_i = \frac{l_{i+1} + l_i}{2}$$

- e) **O número k de classes:**

- ✓ Deve depender do número N de observações efectuadas;
- ✓ Não deve ser tão elevado que sobressaia irregularidades acidentais devido ao pequeno número de indivíduos por classe
- ✓ Não deve ser tão pequeno que conduza a uma perda de informação
- ✓ Situa-se em geral entre 5 e 15.

Depois de determinado k e se as classe tiverem amplitude constante temos que calcular a amplitude da classe, usando a seguinte fórmula:

$$a_i = \frac{x_{max} - x_{min}}{k}$$

Sempre que possível, é vantajoso que os intervalos de classes possuam a mesma amplitude, a fim de que seja mais sugestiva a comparação das frequências de cada classe.

No que se refere á determinação do número k de classes a tomar, não há regras fixas. Iremos usar a seguinte regra para amostras de pequenas dimensões:

$$\begin{cases} k = 5 & \text{se } n < 30 \\ k = \sqrt{n} & \text{se } n \geq 30 \end{cases}$$

Para amostras grandes desse usar-se a fórmula de Sturges:

$$k = \text{int}[1 + 3.22 * \log_{10} n]$$

Onde $\text{int}[x]$ é a parte inteira de x

Exemplo 3.6: Distribuição de dados agrupados em intervalo de classe

Os dados seguintes referem-se á percentagem de algodão, no material usado para confeccionar camisas de homem:

34.2	33.6	33.8	34.7	37.8	32.6
33.1	34.7	34.2	33.6	36.6	33.1
34.5	35.0	33.4	32.5	35.4	34.6
35.6	35.4	34.7	34.1	34.6	35.9
36.3	36.2	34.6	35.1	33.8	34.7

Organize os dados em tabela de frequências de dados agrupados em intervalos de classe.

Solução

1º Determinar o nº de classes

$$k = \text{int}[1 + 3.22 * \log_{10} 30] = 6$$

2º Derminar a amplitude da classe

$$a_i = \frac{x_{max} - x_{min}}{k} = \frac{37.8 - 32.5}{6} = 0.9$$

3º Tabela de frequência

Classe	x_i	f_i	F_i	f_r	F_r
[32.5---33.4[32.95	3	3	0.1	0.1
[33.4---34.3[33.85	8	11	0.27	0.37
[34.3---35.2[34.75	9	20	0.3	0.67
[35.2---36.1[35.65	6	26	0.2	0.87
[36.1---37.0[36.55	3	29	0.1	0.97
[37.0---37.9[37.45	1	30	0.03	1.00
Total		30		1.00	

3.4. Apresentação de dados em gráficos

3.4.1. Gráfico de sectores

Gráfico de sectores ou gráfico circular, como é tradicionalmente chamado gráfico de pizza é um diagrama circular em que os valores de cada categoria estatística representada são proporcionais às respectivas medidas dos ângulos (1% no gráfico de sector equivale a $3,6^\circ$).

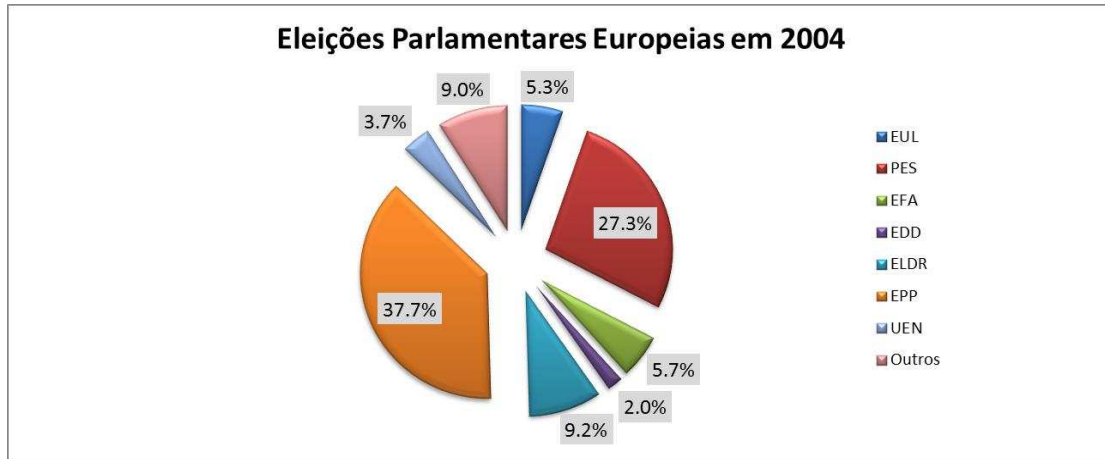
Exemplo 3.7: Gráfico de sectores

O exemplo a seguir é baseado no resultado preliminar das Eleições Parlamentares Europeias em 2004. A tabela consiste no número de assentos alocados para cada partido, além de uma percentagem do grupo total que eles compõem. Os valores da última coluna, que são o ângulo central de cada um dos sectores podem ser encontrados multiplicando as percentagens por 360°

Grupo	Assentos	Percentagem (%)	Angulo Central ($^\circ$)
EUL	39	5.3	19.2
PES	200	27.3	98.4
EFA	42	5.7	20.7
EDD	15	2.0	7.4
ELDR	67	9.2	33.0
EPP	276	37.7	135.7
UEN	27	3.7	13.3
Outros	66	9.0	32.5
Total	732	99.9*	360.2*

*Devido ao arredondamento, o valor total não termina como 100 ou 360.

Solução



Nota que este gráfico pode ser substituído pelo gráfico de barras.

3.4.2. Gráfico de Barras

No gráfico de barras a altura de cada barra traduz o valor da frequência (absoluta ou relativa) respeitante a cada valor da variável. No eixo horizontal assinalam-se os valores possíveis da variável. No eixo vertical as frequências absolutas ou relativa.

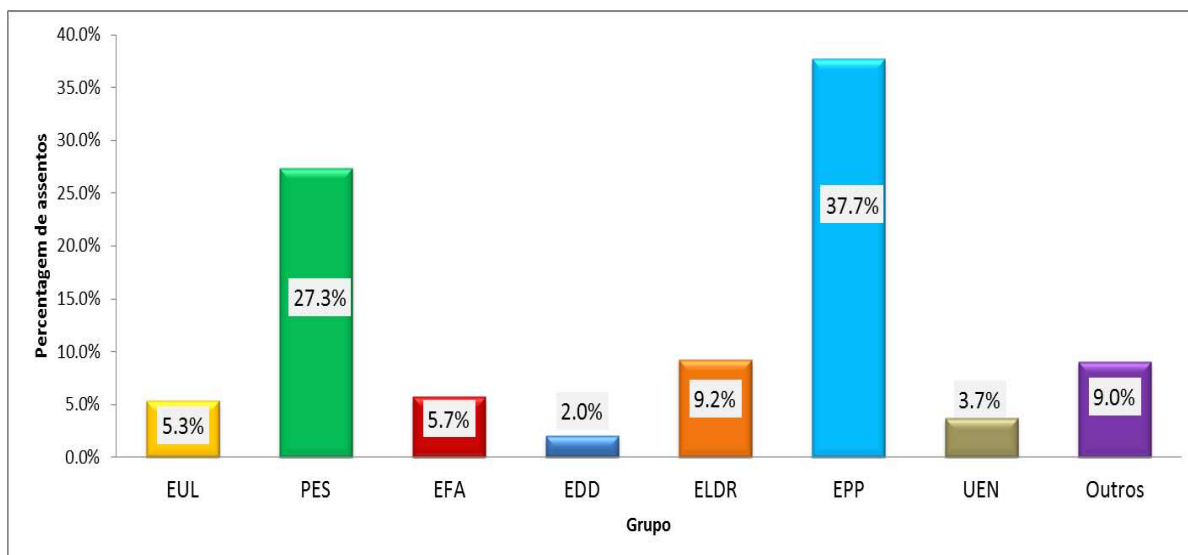
Exemplo 3.8: Gráfico de barras

O exemplo a seguir é baseado no resultado preliminar das Eleições Parlamentares Europeias em 2004. A tabela consiste no número de assentos alocados para cada partido, além de uma percentagem do grupo total que eles compõem. Os valores da última coluna, que são o ângulo central de cada um dos sectores podem ser encontrados multiplicando as percentagens por 360°

Grupo	Assentos	Percentagem (%)	Angulo Central (°)
EUL	39	5.3	19.2
PES	200	27.3	98.4
EFA	42	5.7	20.7
EDD	15	2.0	7.4
ELDR	67	9.2	33.0
EPP	276	37.7	135.7
UEN	27	3.7	13.3
Outros	66	9.0	32.5
Total	732	99.9*	360.2*

*Devido ao arredondamento, o valor total não termina como 100 ou 360.

Solução



3.4.3. Histograma

No caso dos valores agrupados em intervalos de classe é muito frequente representar a distribuição através de um histograma. É um gráfico formado por rectângulos adjacentes em que a área dos rectângulos é proporcional às frequências ordinárias (absolutas ou relativas). Se todos os intervalos tiverem a mesma amplitude, as alturas dos rectângulos serão proporcionais às frequências das classes e então, tomam-se as alturas numericamente iguais a essas frequências. Se os intervalos de classe não tiverem a mesma amplitude, essas alturas deverão ser ajustadas.

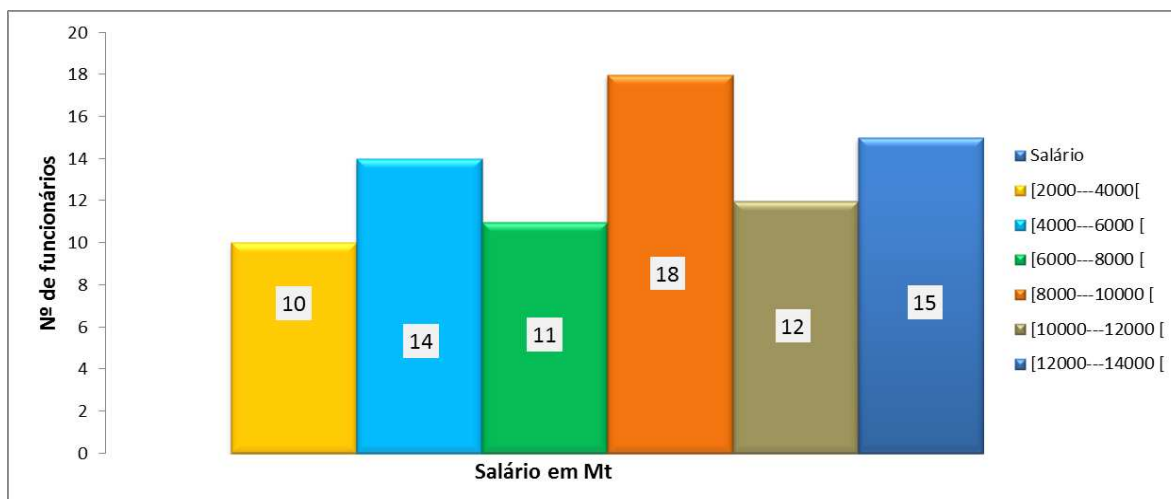
Exemplo 3.9: Histograma

A tabela a seguir representa o salário mensal em meticais de funcionários de uma fábrica de tijolos:

Salário	f_i
[2000---4000[10
[4000---6000 [14
[6000---8000 [11
[8000---10000 [18
[10000---12000 [12
[12000---14000 [15
Total	80

Esboce o histograma correspondente.

Solução



3.4.4. Polígono de frequências

Um polígono de frequência é um gráfico que se realiza através da união dos pontos mais altos das colunas num histograma de frequência (que utiliza colunas verticais para mostrar as frequências).

Os polígonos de frequência para dados agrupados, por sua vez, constroem-se a partir da marca de classe que coincide com o ponto médio de cada coluna do histograma.

Geralmente, os polígonos de frequência são usados quando se pretende mostrar mais de uma distribuição ou a classificação cruzada de uma variável quantitativa contínua com uma qualitativa ou quantitativa discreta num mesmo gráfico. O ponto que tiver mais altura num polígono de frequência representa a maior frequência, ao passo que a área abaixo da curva inclui a totalidade dos dados existentes

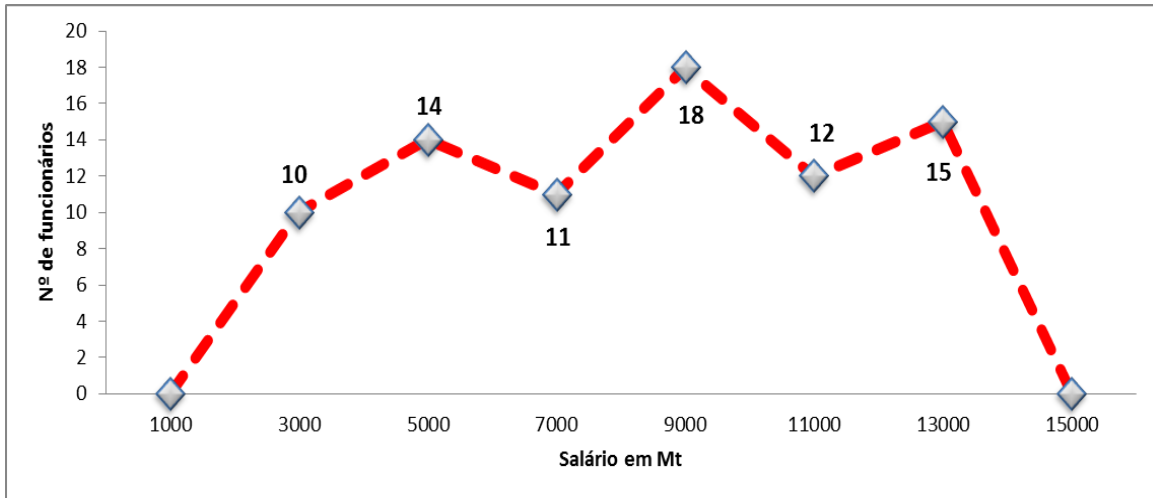
Exemplo 3.10: Polígono de frequências

A tabela a seguir representa o salário mensal em meticais de funcionários numa fábrica de tijolos:

Salário	x_i	f_i
[2000---4000[3000	10
[4000---6000 [5000	14
[6000---8000 [7000	11
[8000---10000 [9000	18
[10000---12000 [11000	12
[12000---14000 [13000	15
Total		80

Esboce o polígono de frequências correspondente.

Solução



3.4.5. Polígono de frequências acumuladas (OGIVA)

Unindo os limites superiores das classes, obtém-se, analogamente o polígono de frequências acumulada ou Ogiva.

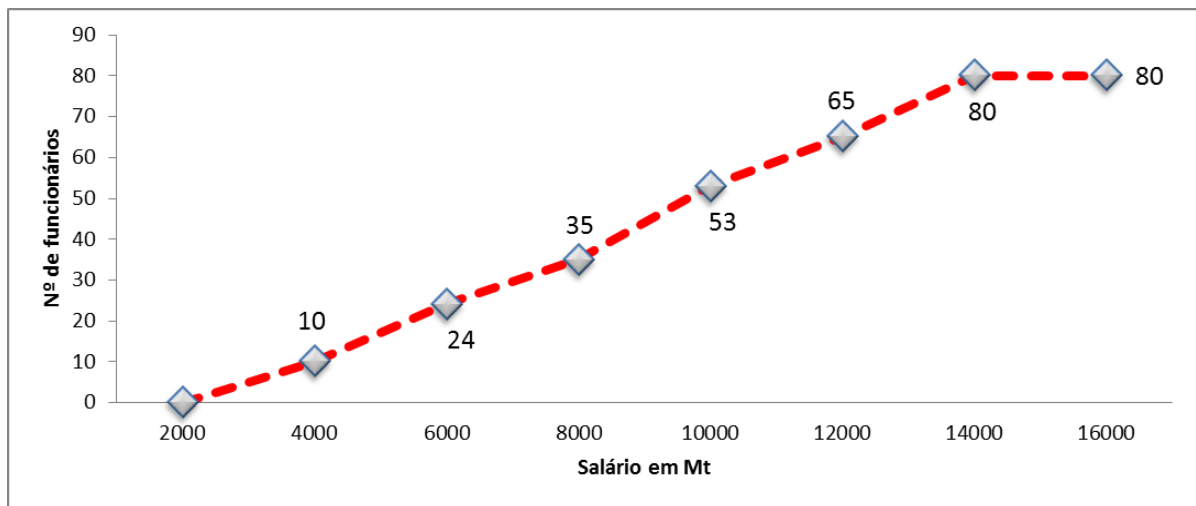
Exemplo 3.11: Polígono de frequências acumuladas

A tabela a seguir representa o salário mensal em metcais de funcionários duma fábrica de tijolos:

Salário	x_i	f_i	F_i
[2000---4000[3000	10	10
[4000---6000 [5000	14	24
[6000---8000 [7000	11	35
[8000---10000 [9000	18	53
[10000---12000 [11000	12	65
[12000---14000 [13000	15	80
Total		80	

Esboce o polígono de frequências acumuladas correspondente.

Solução



3.5. EXERCÍCIOS PROPOSTOS

3.5.1. Classifique cada uma das variáveis abaixo em qualitativo (nominal e ordinal) e quantitativo (discreta e contínua):

- a) Idade;
- b) Altura de um indivíduo;
- c) Sexo;
- d) Classe social;
- e) Marca de automóvel;
- f) Peso de um indivíduo;
- g) Salários de empregados de uma indústria;
- h) Número de ações vendidas diariamente na Bolsa de Valores;
- i) Temperaturas registadas cada meia hora em um posto de Meteorologia;
- j) Comprimentos de 1000 parafusos produzidos numa fábrica;
- k) Número G de litros de água numa máquina de lavar roupa;
- l) Número B de livros em uma estante de biblioteca;
- m) Diâmetro D de uma esfera.
- n) Nº único de identificação tributária
- o) Cor dos olhos;
- p) Número de pessoas favoráveis à pena de morte;
- q) Vendas anuais;
- r) Situação socio económica de um indivíduo;
- s) Programa televisivo com maior audiência e o número de vezes que vai ao ar;

- t) Zona de origem;
- u) Valor de um imóvel
- v) Conceitos em certa disciplina
- w) Classificação em um concurso.
- x) Número de famílias de residentes dum prédio;
- y) Cor do cabelo de um individuo;
- z) Produto interno bruto de 15 países da região;
- aa) Ganhos por acção;
- bb) Método de pagamento (à vista, com cheque, com cartão de crédito).
- cc) Gastos com alimentação.
- dd) Tempo para fazer um teste.

3.5.2. Converta as seguintes proporções em percentagens: 0,09; 0,955; 0,33; 0,017.

3.5.3. Converta as seguintes percentagens em proporções: 35,5%; 53,1%; 50%; 46,57%.

3.5.4. Assinale a afirmativa verdadeira:

- a) Um gráfico de barras ou colunas é aquele em que os rectângulos que o compõem estão dispostos horizontalmente.
- b) Um gráfico de barras ou colunas é aquele em que os rectângulos que o compõem estão dispostos verticalmente.
- c) Um gráfico de barras é aquele em que os rectângulos que o compõem estão dispostos verticalmente e um gráfico de colunas, horizontalmente.
- d) Um gráfico de barras é aquele em que os rectângulos que o compõem estão dispostos horizontalmente e um gráfico de colunas, verticalmente.
- e) Todas as alternativas anteriores são falsas.

3.5.5. Um dado foi lançado 50 vezes e foram registados os seguintes resultados

5	4	6	1	2	5	3	1	3	3
4	4	1	5	5	6	1	2	5	1
3	4	5	1	1	6	6	2	1	1
4	4	4	3	4	3	2	2	2	3
6	6	3	2	4	2	6	6	2	1

Construa uma distribuição de frequência sem intervalo de classe e determine:

- a) O número de classe;
- b) A amplitude total;
- c) A freqüência total;
- d) A freqüência simples absoluta do primeiro elemento;
- e) A freqüência simples relativa do primeiro elemento;
- f) A freqüência acumulada do primeiro elemento;
- g) A freqüência acumulada relativa do primeiro elemento;
- h) A freqüência simples absoluta do segundo elemento;
- i) A freqüência simples relativa do quinto elemento;
- j) A freqüência acumulada relativa do sexto elemento;

3.5.6. Dado o rol de medidas das alturas (dadas em cm) de uma amostra de 100 indivíduos de uma faculdade:

151	152	154	155	158	159	159	160	161	161
161	162	163	163	163	164	165	165	165	166
166	166	166	167	167	167	167	167	168	168
168	168	168	168	168	168	168	168	169	169
169	169	169	169	169	170	170	170	170	170
170	170	171	171	171	171	172	172	172	173
173	173	174	174	174	175	175	175	175	176
176	176	176	177	177	177	177	178	178	178
179	179	180	180	180	180	181	181	181	182
182	182	183	184	185	186	187	188	190	190

calcule:

- a) A amplitude amostral;
- b) O número de classes;
- c) A amplitude de classes;
- d) Os limites de classes;
- e) As frequências absolutas das classes;
- f) As frequências relativas;
- g) Os pontos médios das classes;
- h) As frequências acumuladas;
- i) O histograma e o polígono de frequência;
- j) O polígono de frequência acumulada;
- k) Faça um breve comentário sobre os valores das alturas desta amostra através da distribuição de frequência.

3.5.7. Considere a seguinte distribuição de frequência correspondente aos diferentes preços em u.m de um determinado produto em vinte lojas pesquisadas.

Preços	Nº de lojas
50	2
51	5
52	6
53	6
54	1
Total	20

- Quantas lojas apresentaram um preço de 52,00 u.m?
- Construa uma tabela de frequências simples relativas.
- Construa uma tabela de frequências absolutas acumuladas.
- Quantas lojas apresentaram um preço de até 52,00 u.m (inclusive)?
- Qual o percentual de lojas com preço maior de que 51,00 u.m e menor de que 54,00 u.m?

3.5.8. O quadro seguinte representa as alturas (em cm) de 40 alunos de uma classe.

162 163 148 166 169 154 170 166
164 165 159 175 155 163 171 172
170 157 176 157 157 165 158 158
160 158 163 165 164 178 150 168
166 169 152 170 172 165 162 164

- Calcular a amplitude total.
- Admitindo-se 6 classes, qual a amplitude do intervalo de classe?
- Construir uma tabela de frequência das alturas dos alunos.

3.5.9. Construa uma tabela para mostrar que, em determinado curso, o número de alunos matriculados na 1ª, 2ª e 3ª classe era, respectivamente, 40, 35 e 29 em 1997 e 42, 36 e 32 em 1998.

3.5.10. Construa uma tabela para mostrar que, de acordo com a Pesquisa Nacional por Amostra de Domicílios, PNAD, em 1992 havia no Brasil 73,1 milhões de pessoas com renda familiar mensal até 330 reais (pobres e miseráveis), 45 milhões de pessoas com renda familiar mensal de 330 reais até 1300 reais (emergentes) e 13,6 milhões de pessoas com renda familiar mensal acima de 1300 reais (classe média e ricos). Apresente, também, percentuais.

3.5.11. Um restaurante usa um questionário para solicitar aos seus clientes uma avaliação do garçom, da qualidade da comida, dos serviços, dos preços e do ambiente no restaurante. Cada característica é avaliada numa escala de

excelente (E), óptimo (O), bom (B), médio (M), e fraco (F). Use a estatística descritiva para sintetizar os seguintes dados colectados sobre a qualidade da comida.

O O M F O B B M F E E O O B B E E M
 O O F O O B E M M O O O O O M F E F
 E O M O O E O O O E O O M E B M O E
 F F E M F M E O E E M E E E F O E O

- Represente os dados em gráfico de sectores e barras;
- Qual é a sua impressão sobre a qualidade da comida apresentada no restaurante?

3.5.12. Os dados seguintes representam 20 observações relativas ao índice pluviómetro em determinados municípios do país:

144 152 159 160 141 160 151 157 146 150
 154 145 141 150 143 142 146 142 141 158

- Determinar o número de classes por regra de Sturges.
- Construir a tabela de frequências absolutas Simples.
- Determinar as frequências absolutas acumuladas.
- Determinar as frequências relativas Simples.
- Determinar as frequências relativas acumuladas.

3.5.13. Os dados a seguir referem-se ao número de livros adquiridos, no ano passado, pelos 40 alunos da Turma A:

4	2	1	0	3	1	2	0	2	1
0	2	1	1	0	4	3	2	3	5
8	0	1	6	5	3	2	1	6	4
3	4	3	2	1	0	2	1	0	3

- Classifique a variável.
- Organize os dados em uma tabela adequada.
- Qual o percentual de alunos que adquiriram menos do que 3 livros?
- Qual o percentual de alunos que adquiriram pelo menos 4 livros?
- A partir do item (b), quantos livros foram adquiridos pelos 40 alunos?

3.5.14. Considere os dados abaixo referentes ao consumo de energia Kw, de 75 contas da EDM:

32	40	22	11	34	40	16	26	23	31	49
10	38	17	13	45	25	50	18	23	35	56
22	30	14	18	20	13	24	35	29	33	19
48	20	12	31	39	17	58	19	16	12	11
21	15	12	20	51	12	19	15	41	29	15
25	13	23	32	14	27	43	37	21	28	
37	26	44	11	53	38	46	17	36	28	

- Organize os dados numa distribuição de frequências com 9 classes de amplitudes iguais;
- A partir da distribuição de frequências construída no item anterior, determine e interprete: f_3 ; fr_4 ; $Fr_4 - Fr_2$;
- Construa o correspondente histograma de frequências relativas;
- Determine as frequências simples e acumuladas (absolutas e relativas).

3.5.15. A altura de 60 alunos da FACE-PUC foi registada abaixo, em cm:

174	170	156	168	176	178	162	182	172	168
168	156	169	168	162	160	163	168	162	172
168	167	170	153	171	166	168	156	160	172
163	170	175	176	182	158	176	161	175	161
173	163	172	167	170	179	179	170	151	175
151	172	173	170	174	167	158	174	164	173

- Construa uma distribuição de frequência com 8 classes de amplitudes iguais, adoptando como limite inferior da distribuição 150 cm;
- Qual o percentual de alunos com altura mínima de 166 cm?
- Quantos alunos tem menos de 162 cm?
- Qual o percentual de alunos com altura média de 164 cm? Qual a soma total aproximada das alturas dos 60 alunos?

3.5.16. Abaixo são mostrados os saldos médios de 48 contas de clientes do BB Novo S.A. (dados brutos em US\$ 1,00).

450	500	150	1000	250	275	550	500
225	475	150	450	950	300	800	275
600	750	375	650	150	500	1000	700
475	900	800	275	600	750	375	650
150	500	225	250	150	120	250	360
230	500	350	375	470	600	1030	270

- Agrupe os dados numa distribuição de frequências.
- Determine as frequências relativas: simples e acumulada.
- Apresente o histograma de frequências relativas.
- Interprete fr_2 , f_3 e $(Fr_4 - Fr_2)$.
- Apresente dados em polígonos de frequências simples e acumulada.

3.5.17. As informações abaixo indicam o número de acidentes ocorridos com 70 motoristas de uma empresa de ônibus nos últimos 5 anos:

Nº de acidentes	0	1	2	3	4	5	6	7
Nº de motoristas	15	11	20	9	6	5	3	1

- a) Determine o número de motoristas com menos de 1 acidente.
- b) Determine o percentual de motoristas com pelo menos 3 acidentes.
- c) Determine o percentual de motoristas com no máximo 2 acidentes.
- d) Qual o número total de acidentes ocorrido no período?

3.5.18. Na administração de um sistema escolar de certo município 70% da despesa vai para o ensino, 12% para a administração e manutenção e 18% para órgãos auxiliares, encargos fixos e despesas ocasionais. Qual o gráfico que melhor representaria essa situação? Se o valor total da despesa é de 10,000.00 unidades monetárias, qual é o valor gasto com administração e manutenção?

4. MEDIDAS DESCRITIVAS

As medidas descritivas classificam-se em medidas de localização (de tendência central ou de posição não central), dispersão (ou de variabilidade), assimetria e achatamento (ou curtose).

4.1. Medidas de Localização

4.1.1. Medidas de Tendência Central

As medidas de tendência central indicam os pontos em torno dos quais se encontram os valores da variável estatística, ou seja, localizam a distribuição. As principais medidas de localização são:

- Média
- Mediana
- Moda

Média aritmética

É o tipo de média mais utilizada pelas pessoas no dia-a-dia e subdivide-se em dois tipos: simples e ponderada. A diferença entre elas é que na média aritmética a importância (peso) de cada ocorrência é igual e na ponderada cada termo possui uma importância relativa, ou seja, possuem pesos diferentes.

Média aritmética simples

A média aritmética da variável estatística X define-se por:

$$\bar{x} = \frac{\sum x_i}{n} \rightarrow \text{para dados não agrupados}$$

Ou

$$\bar{x} = \frac{\sum x_i * f_i}{n} \rightarrow \text{para dados agrupados, onde } n = \sum f_i$$

Exemplo 4.1: Média aritmética (dados não agrupados)

A tabela abaixo mostra as notas de matemática de um aluno em um determinado ano:

1° Bimestre	3,5
2° Bimestre	7,5
3° Bimestre	9,0
4° Bimestre	6,0

Determine a nota média bimestral

Solução

$$\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + x_3 + x_4}{n} = \frac{3.5 + 7.5 + 9 + 6}{4} = 6.5$$

Interpretação: Ter média 6,5 significa dizer que, apesar de ele ter obtido notas mais altas ou mais baixas em outros bimestres, a soma das notas (26) é a mesma que ele alcançaria se tivesse obtido nota 6,5 em todos os bimestres.

Exemplo 4.2: Média aritmética (dados agrupados)

Os salários semanais dos funcionários de uma empresa estão distribuídos na tabela abaixo:

Salário (em Mt)	Nº de funcionários
400,00	5
600,00	2
1.000,00	2
5.000,00	1

Determine o salário médio semanal dos funcionários

Solução

$$\bar{x} = \frac{\sum x_i * f_i}{n} = \frac{400 * 5 + 600 * 2 + 1000 * 2 + 5000 * 1}{5 + 2 + 2 + 1} = \frac{10200}{10} = 1020$$

Ou (Recorrendo a tabela)

x_i	f_i	$x_i * f_i$
400	5	2000
600	2	1200
1000	2	2000
5000	1	5000
Somatório	10	10200

$$\bar{x} = \frac{\sum x_i * f_i}{n} = \frac{400 * 5 + 600 * 2 + 1000 * 2 + 5000 * 1}{5 + 2 + 2 + 1} = \frac{10200}{10} = 1020$$

Interpretação: Em média cada funcionário recebe semanalmente 1020 Mt.

Exemplo 4.3: Média aritmética (dados agrupados em intervalo de classe)

A tabela a seguir representa o salário mensal em meticais de funcionários duma fábrica de tijolos:

Salário	f_i
[2000---4000[10
[4000---6000 [14
[6000---8000 [11
[8000---10000 [18
[10000---12000 [12
[12000---14000 [15
Total	80

Determine o salário médio mensal dos funcionários desta fábrica.

Solução

Recorrendo a tabela temos:

Classe	x_i	f_i	$x_i * f_i$
[2000---4000[3000	10	30000
[4000---6000 [5000	14	70000
[6000---8000 [7000	11	77000
[8000---10000 [9000	18	162000
[10000---12000 [11000	12	132000
[12000---14000 [13000	15	195000
Total		80	666000

$$\bar{x} = \frac{\sum x_i * f_i}{n} = \frac{666000}{80} = 8325$$

Interpretação: O salário médio mensal dos funcionários desta fábrica é de 8325

Meticais.

Média aritmética ponderada

A média ponderada considera “pesos” para cada item, ou seja, em um conjunto de dados, cada item recebe uma importância. Vamos supor que tenhamos um conjunto com n dados $(x_1; x_2; x_3; x_4; \dots; x_n)$, onde cada dado receberá um peso, respectivamente $(p_1; p_2; p_3; p_4; \dots; p_n)$. Cada item será multiplicado pelo seu peso. A média será dada pela divisão entre esta soma e a soma dos pesos considerados. A média entre esses dados será dada por:

$$\bar{x}_p = \frac{x_1 * p_1 + x_2 * p_2 + x_3 * p_3 + \dots + x_n * p_n}{p_1 + p_2 + p_3 + p_4 + \dots + p_n}$$

Exemplo 4.4: Média aritmética ponderada

Uma aluna fez uma prova e obteve nota 19,1 e um trabalho, com nota 8,7. A média considera que a prova tenha peso 6 e o trabalho peso 4. Determine a média dessa aluna.

Solução

$$\bar{x}_p = \frac{19.1 * 6 + 8.7 * 4}{6 + 4} = 14,94$$

Interpretação: A média dessa aluna é de 14,94 valores.

Média geométrica

Esse tipo de média tem várias aplicações. É muito utilizada na área de finanças e de engenharia. Lembrando que a média geométrica de um conjunto é sempre menor ou igual a média aritmética, vários problemas de desigualdades como na geometria são resolvidos através dela. A **média geométrica** de um conjunto de números positivos é definida como o produto de todos os membros do conjunto elevado ao inverso do número de membros

A média geométrica entre um conjunto de n dados é a raiz n -ésima da multiplicação desses dados. Considere um conjunto de n dados $(x_1; x_2; x_3; x_4; \dots; x_n)$. A média geométrica entre estes dados será dada por:

$$\bar{x}_g = \sqrt[n]{x_1 * x_2 * x_3 * \dots * x_n}$$

Exemplo 4.5: Média geométrica

Qual a média geométrica entre 2, 8 e 32?

Solução

Temos três dados, então a média geométrica será a raiz cúbica de 2, 8, 32:

$$\bar{x}_g = \sqrt[n]{x_1 * x_2 * x_3}$$

$$\bar{x}_g = \sqrt[3]{2 * 8 * 32} = \sqrt[3]{512} = 8$$

Interpretação: A média geométrica entre 2, 8 e 32 é 8.

Exemplo 4.6: Média geométrica

Uma outra utilização para este tipo de média, é com variações percentuais em sequência. Como por exemplo:

Digamos que uma categoria de operários tenha um aumento salarial de 20% após um mês, 12% após dois meses e 7% após três meses. Qual o percentual médio mensal de aumento desta categoria?

Solução

$$\bar{x}_g = \sqrt[3]{20 * 12 * 7} = \sqrt[3]{1.43808} = 1.1287$$

Média harmónica

A média harmónica de um conjunto de n dados é obtida dividindo a quantidade de dados pela soma dos inversos dos dados. Esse tipo de média nunca é maior que a média aritmética ou a geométrica. Ela é utilizada quando se trata de grandezas inversamente proporcionais. Considerando um conjunto de n dados $(x_1; x_2; x_3; x_4; \dots; x_n)$, a média harmónica entre esses dados, será:

$$\bar{x}_h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \frac{1}{x_4} + \dots + \frac{1}{x_n}}$$

Exemplo 4.7: Média harmónica

Qual a média harmónica entre 2, 8 e 32?

Solução

$$\bar{x}_h = \frac{3}{\frac{1}{2} + \frac{1}{8} + \frac{1}{32}} = 4.57$$

Interpretação: A média harmónica entre 2, 8 e 32 é 4.57

Mediana

Mediana é o valor que separa a metade maior e a metade menor de uma amostra, uma população ou uma distribuição de probabilidade. Em termos mais simples, mediana pode ser o valor do meio de um conjunto de dados.

A mediana é uma medida comum das propriedades de conjuntos de dados em estatística e em teoria das probabilidades, com importância central na estatística

robusta. A estatística robusta é mais resistente, com ponto de ruptura de 50%. A mediana não fornece resultados arbitrariamente grandes desde que mais da metade dos dados não esteja contaminada.

A vantagem da mediana em relação à média é que a mediana pode dar uma ideia melhor de um valor típico porque não é tão distorcida por valores extremamente altos ou baixos. Em estudos estatísticos sobre renda familiar ou outros activos voláteis, a média pode ser distorcida por um pequeno número de valores extremamente altos ou baixos.

Em relação a esta medida convém distinguir claramente os dois casos de dados brutos ou agrupados, e agrupados em intervalo de classe.

Para dados brutos ou agrupados, supondo as observações dos valores da variável estatística ordenados sob forma crescente $x_1 \leq x_2 \leq x_3 \leq x_4 \leq \dots \leq x_n$, há a considerar duas hipóteses:

- $n = 2k + 1$ (Ímpar), neste caso $\tilde{x} = x_{k+1}$ (observação central)
- $n = 2k$ (Par), neste caso $\tilde{x} = \frac{x_k + x_{k+1}}{2}$ (observação central)

Exemplo 4.8: Mediana (dados brutos ou agrupados)

Considere os dados em Rol: 3, 4, 4, 5, 6, 8, 8, 8, 10. Determine a mediana.

Solução

$$n = 2k + 1, \text{ isto é, } n = 9 \text{ e } k = 4, \text{ então } \tilde{x} = x_{k+1} = x_5 = 6$$

Ou

(pelo agrupamento dos dados)

x_i	f_i	F_i
3	1	1
4	2	3
5	1	4
6	1	5
8	3	8
10	1	9
Total	9	

$$\text{Sendo } n = 9 \text{ e } \frac{9+1}{2} = 5; \text{ então } \tilde{x} = x_5 = 6$$

Exemplo 4.9: Mediana (dados brutos ou agrupados)

Considere os dados em Rol: 2, 3, 4, 4, 5, 6, 8, 8, 8, 10. Determine a mediana.

Solução

$n = 2k$, isto é, $n = 10$ e $k = 5$, então:

$$\tilde{x} = \frac{x_k + x_{k+1}}{2} = \frac{x_5 + x_6}{2} = \frac{5 + 6}{2} = 5.5$$

Ou (pelo agrupamento dos dados)

x_i	f_i	F_i
2	1	1
3	1	2
4	2	4
5	1	5
6	1	6
8	3	9
10	1	10
Total	10	

Sendo $n = 10$ e $\frac{10}{2} = 5$; então $\tilde{x} = \frac{x_5 + x_6}{2} = \frac{5 + 6}{2} = 5.5$

Para dados agrupados em intervalo de classe, a mediana é o valor tal que a ordenada levantada no ponto do eixo das abscissas divide a área do histograma em duas partes iguais, isto é, a mediana será o valor da abscissa a que corresponde a frequência absoluta (relativa) acumulada $\frac{n}{2}$, e a fórmula será dada por:

$$\tilde{x} = \lim_{inf} + \frac{\frac{1}{2} * n - F_{i-1}}{f_i} * a_i, \text{ Onde:}$$

- \lim_{inf} : limite inferior da classe mediana;
- n : tamanho da amostra
- F_{i-1} : frequência absoluta acumulada anterior a classe mediana
- f_i : frequência absoluta simples da classe mediana
- a_i : amplitude da classe mediana

Exemplo 4.10: Mediana (dados agrupados em intervalo de classes)

A tabela a seguir representa o salário mensal em metcais de funcionários duma fábrica de tijolos:

Salário	f_i
[2000---4000[10
[4000---6000 [14
[6000---8000 [11
[8000---10000 [18
[10000---12000 [12
[12000---14000 [15
Total	80

Determine o salário mediano dos funcionários desta fábrica.

Solução

Salário	f_i	F_i
[2000---4000[10	10
[4000---6000 [14	24
[6000---8000 [11	35
[8000---10000 [18	53
[10000---12000 [12	65
[12000---14000 [15	80
Total	80	

Vamos determinar a mediana desta distribuição, procurando o salário do $\frac{1}{2} * 80 = 40^{\text{º}}$ funcionário. Como se sabe pode observar, até á classe 6000-8000, inclusive, o total das frequências acumuladas é de 35 e adicionando a frequência da classe 8000-10000, obtém se a frequência acumulada 53, já superior a 40. Pode pois garantir-se que a mediana é um dos valores da classe 8000-10000 (classe mediana), então:

$$\tilde{x} = 8000 + \frac{\frac{1}{2} * 80 - 35}{18} * 2000 = 8556$$

Interpretação: 50% dos funcionários recebem um salário máximo de até 8556 Mt

Moda

A moda amostral de um conjunto de dados trata do valor que ocorre com maior frequência ou o valor mais comum em um conjunto de dados. A moda é especialmente útil quando os valores ou as observações não são numéricos, casos em que a média e a mediana não podem ser definidas.

Uma amostra pode ser unimodal (uma moda), bimodal (duas modas), multimodal (várias modas) e amodal (nenhuma moda).

Em relação a esta medida convém distinguir claramente os dois casos de dados brutos ou agrupados, e agrupados em intervalo de classe.

Para dados brutos ou agrupados, a moda será aquele que apresentar maior frequência.

Exemplo 4.11: Moda (dados brutos ou agrupados)

Considere os dados em Rol: 3, 4, 4, 5, 6, 8, 8, 8, 10. Determine a moda.

Solução

$$\hat{x} = 8$$

Exemplo 4.12: Moda (dados brutos ou agrupados)

Considere os dados apresentados na tabela abaixo. Determine a moda.

x_i	f_i
3	1
4	2
5	1
6	1
8	3
10	1
Total	9

$\hat{x} = 8$, Porque apresenta a maior frequência

Para dados agrupados em intervalo de classe, a moda será dada por:

$$\hat{x} = \lim_{inf} + \frac{f_{i\hat{x}} - f_{\hat{x}i-1}}{2 * f_{i\hat{x}} - (f_{\hat{x}i-1} + f_{\hat{x}i+1})} * a_i \quad (\text{Moda do Czuber})$$

$$\hat{x} = \lim_{inf} + \frac{f_{\hat{x}i+1}}{f_{\hat{x}i-1} + f_{\hat{x}i+1}} * a_i \quad (\text{Moda do King})$$

$$\hat{x} = 3 * \tilde{x} - 2 * \bar{x} \quad (\text{Moda do Pearson})$$

Onde

- $f_{i\hat{x}}$: frequência modal
- $f_{\hat{x}i-1}$: frequência anterior a modal
- $f_{\hat{x}i+1}$: frequência posterior a modal
- a_i : amplitude da classe modal

Moda de King considera as classes adjacentes à classe modal, enquanto que, Moda de Czuber considera as classes adjacentes à classe modal e à própria classe modal.

Exemplo 4.13: Moda (dados agrupados em intervalo de classes)

A tabela a seguir representa o salário mensal em meticais de funcionários duma fábrica de tijolos:

Salário	f_i
[2000---4000[10
[4000---6000 [14
[6000---8000 [11
[8000---10000 [18
[10000---12000 [12
[12000---14000 [15
Total	80

Determine o salário modal dos funcionários desta fábrica.

Solução

Fórmula de Czuber

Classe modal [8---10[

$$\hat{x} = 8000 + \frac{18 - 11}{2 * 18 - (11 + 12)} * 2000 = 9077$$

Resposta: O salário modal (Czuber) dos funcionários é de 9077.00 Mt.

Fórmula de King

Classe modal [8---10[

$$\hat{x} = 8000 + \frac{12}{11+12} * 2000 = 9043.5$$

Resposta: O salário modal (King) dos funcionários é de 9043.50 Mt.

4.1.2. Medidas de Tendência não Central ou Separatizes

Estas medidas descritivas permitem localizar a posição de um valor dentro de um conjunto de dados, é calculada para as variáveis qualitativas ordinais e quantitativa (discreta e contínua). Pode ser calculando quando os valores observados são expressos nas mesmas unidades de dados estudo.

As principais medidas são:

- Quartis
- Decis
- Percentis

Quartil

É uma medida que divide um conjunto de dados em 4 partes iguais. Existem três quartis nomeadamente (Q_1 ; Q_2 e Q_3), no qual determina-se da seguinte maneira:

- Q_1 : Obtém-se determinando a mediana do conjunto de valores observados que fica á esquerda da mediana (o 1º quartil será um valor da variável tal que o número de observações para valores inferiores será 25%);
- Q_2 : Corresponde a mediana (o 2º quartil terá a metade das observações, á sua direita e outra metade á sua esquerda);

- Q_3 : Obtém-se determinando a mediana do conjunto de valores observados que fica á direita da mediana (o 3º quartil será um valor tal que á sua esquerda concentrar-se-ão 75% das observações);

Exemplo 4.14: Quartil (dados brutos ou agrupados)

Considere o conjunto de valores observados: 3, 4, 4, 4, 5, 6, 8, 8, 8, 8, 10. Determine os valores de Quartis.

Solução

Começamos por determinar a mediana (Q_2) que é igual ao 2º quartil. O conjunto de valores tem um número ímpar de termos, logo $\tilde{x} = Q_2 = 6$. A partir da mediana o conjunto fica dividido em dois subconjuntos:

3, 4, 4, 5, 5 e 8, 8, 8, 8, 10

As medianas desses subconjuntos são respectivamente iguais a 4 e 8, pelo que $Q_1 = 4$ e $Q_3 = 8$

OU

Localizando as posições da medida Quartílica pela fórmula

$$p = \frac{i}{4} * (n + 1) \rightarrow \text{para } n \text{ ímpar e } p = \frac{1}{4} * (in + 2) \rightarrow \text{para } n \text{ par}$$

$$Q_1: \frac{1}{4} * (11 + 1) = 3 \text{ logo } Q_1 = X_3 = 4$$

$$Q_3: \frac{3}{4} * (11 + 1) = 9 \text{ logo } Q_3 = X_9 = 8$$

Exemplo 4.15: Quartil (dados agrupados em intervalos de classes)

A tabela a seguir representa o salário mensal em meticais de funcionários duma fábrica de tijolos:

Salário	f_i
[2000---4000[10
[4000---6000 [14
[6000---8000 [11
[8000---10000 [18
[10000---12000 [12
[12000---14000 [15
Total	80

Determine os quartis 1, 2 e 3.

Solução

Tratando-se de dados agrupados em intervalos de classe, a fórmula para obter os quartis é idêntica á da mediana:

$$Q_i = \lim_{inf} + \frac{\frac{in}{4} - F_{i-1}}{f_i} * a_i$$

Salário	f_i	F_i
[2000---4000[10	10
[4000---6000 [14	24
[6000---8000 [11	35
[8000---10000 [18	53
[10000---12000 [12	65
[12000---14000 [15	80
Total	80	

Identificação da classe quartílica $\left(\frac{in}{4}, \text{onde } i = 1,2,3\right)$

Quartil 1 (0.25*80=20):

$$P_{25} = 4000 + \frac{20 - 10}{14} * 2000 = 5428.57$$

Quartil 2 (0.50*80=40):

$$P_{50} = 8000 + \frac{40-35}{18} * 2000 = 8555.56$$

Quartil 3: (0.75*80=60)

$$P_{75} = 10000 + \frac{60-53}{12} * 2000 = 11166.67$$

Interpretação: 25%, 50%, 75% dos funcionários auferem um salário máximo de até 5428,57Mt, 8555.56Mt e 11166.67Mt respectivamente.

Decil

Decil é qualquer um dos nove valores que dividem os dados ordenados de uma variável em dez partes iguais, de modo que cada parte representa 1/10 da amostra ou população. Assim:

- O 1° decil é o ponto de corte para 10% dos dados mais baixos, isto é, o percentil 10;
- O 5° decil é o ponto de corte para 50% dos dados, isto é, o percentil 50, 2° quartil, ou mediana;
- O 9° decil é o limite para 90% dos dados mais baixos, isto é, o percentil 90.

Se tratar de dados não agrupados ou agrupados por frequências procede-se de forma idêntica aos quartis.

Para dados agrupados em intervalos de classe a expressão que dá os decis é análoga á dos quartis:

$$D_i = \lim_{inf} + \frac{\frac{i}{10} - F_{i-1}}{f_i} * a_i$$

Com $i = 1, 2, 3, \dots, 9$

Percentil

Percentil é uma medida que divide a amostra ordenada (por ordem crescente dos dados) em 100 partes, cada uma com uma percentagem de dados aproximadamente igual. Portanto:

- O 1° percentil determina o 1% menor dos dados;
- O 98° percentil determina os 98% menores dos dados.
- O 25° percentil é o primeiro quartil; o 50° percentil é a mediana;
- O 10° percentil é o primeiro decil;
- O 80° percentil é o oitavo decil.

A definição de Mendenhall e Sincich para o i -ésimo percentil de N valores ordenados é correspondente ao valor que ocupa a posição $p = \frac{i*(n+1)}{100}$, arredondada para o inteiro mais próximo.

A fórmula generalizada para cálculo de percentil para dados não agrupados ou agrupados em frequências será dada por:

$$P_i = X_m + (p - m) * (X_{m+1} - X_m)$$

onde

- P_i : é a medida percentil a ser utilizada
- X_{m+1} e X_m : são as posições dos dados no rol
- p : é a posição da medida percentil adoptada
- m : é a parte inteira de p

Exemplo 4.16: Percentis (dados não agrupados ou agrupados em frequências)

Considere o conjunto de valores observados: 3, 4, 4, 4, 5, 6, 8, 8, 8, 8, 10. Determine

$$P_{60} \text{ e } P_{90}$$

$$P_i = X_m + (p - m) * (X_{m+1} - X_m)$$

$$P_{60}: p = \frac{i*(n+1)}{100} = \frac{60}{100} * (11 + 1) = 7.2$$

$$P_{60} = X_7 + (7.2 - 7) * (X_8 - X_7)$$

$$P_{60} = 8 + (7.2 - 7) * (8 - 8) = 8$$

$$P_{90}: p = \frac{i*(n+1)}{100} = \frac{90}{100} * (11 + 1) = 10.8$$

$$P_{90} = X_{10} + (10.8 - 10) * (X_{11} - X_{10})$$

$$P_{90} = 8 + (10.8 - 10) * (10 - 8) = 9.6$$

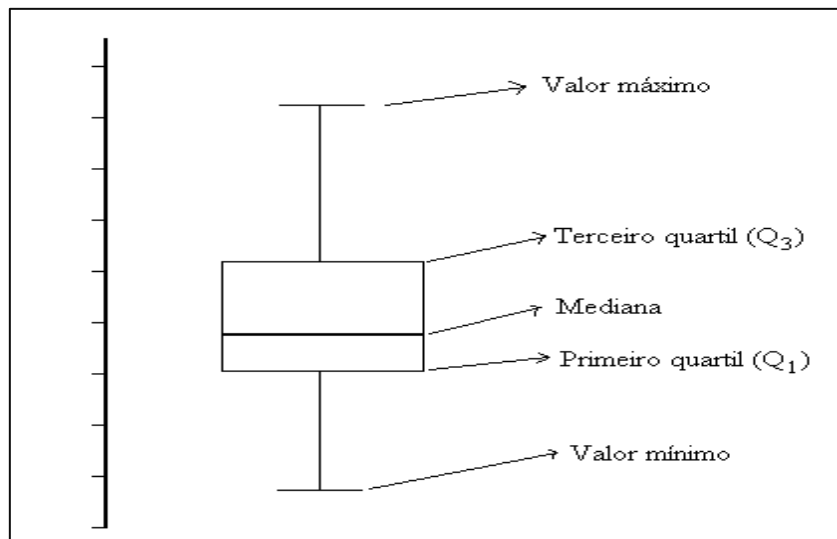
Para dados agrupados em intervalos de classe, a expressão que dá os percentis é análoga á dos quartis e decis:

$$P_i = \lim_{inf} + \frac{\frac{i}{100} - F_{i-1}}{f_i} * a_i$$

Com $i = 1, 2, 3, \dots, 99$

Diagramas em Caixa (Box-plots)

Quando se conhece os quartis e decis de um conjunto de dados, uma maneira bastante comum de se representar a distribuição é através dos diagramas em caixa (Box-plots em inglês). A figura abaixo ilustra o que é um diagrama em caixa.



4.2. Medidas de Dispersão ou Variabilidade

Embora as medidas de localização forneçam indicações sobre os valores mais representativos de uma distribuição, não indicam a sua estrutura interna, isto é, a forma como os diferentes valores se distribuem ao longo do intervalo de variação. As medidas

de dispersão ou variação, permitem conhecer a forma como os valores da variável estatística se distribuem (dispersam) em torno dos valores centrais.

As medidas de dispersão mais importantes são:

- Amplitude total;
- Momentos centrais
- Desvio médio
- Desvio padrão
- Coeficiente de dispersão ou variação
- Intervalo interquartil

4.2.1. Amplitude total

A amplitude total é a medida de dispersão mais simples. É a diferença entre os valores extremos assumidos pela variável estatística.

$$At = \begin{cases} x_{max} - x_{min}, & \text{para dados não agrupados ou agrupados em frequências} \\ & \text{ou} \\ l_{k+1} - l_1, & \text{para dados agrupados em intervalos de classe} \end{cases}$$

O emprego desta medida de dispersão apresenta alguns inconvenientes. A principal desvantagem resulta dela depender apenas dos valores extremos assumidos pela variável e não dos valores intermédios. Duas distribuições podem ter a mesma amplitude total mas dispersão muito diferentes.

4.2.2. Momentos Centrais

São as médias aritméticas da 1ª, 2ª, 3ª, 4ª,... potências dos desvios em relação à média aritmética. A fórmula para o cálculo dos momentos é:

$$M_i = \frac{\sum |(x_i - \bar{x})^i| * f_i}{\sum f_i}$$

Onde $i=1,2,3,4,\dots$

Os momentos são muito importantes em Estatística para caracterizar as distribuições de uma variável ou probabilidade. Por exemplo, a distribuição normal é caracterizada apenas pelo primeiro e pelo segundo momento. O primeiro, segundo, terceiro e quarto

momento caracterizam a tendência central, dispersão, assimetria e curtose, respectivamente, de uma distribuição.

Os momentos mais importantes são os quatro primeiros, que são muito utilizados para caracterizar as distribuições de uma variável ou probabilidade. Entretanto, é quase sempre possível calcular momentos de alta ordem.

4.2.3. Desvio Médio

Da definição da média, é imediato reconhecer que a soma dos desvios das observações em relação à média é nula. Este facto leva à definição de desvio absoluto médio ou simplesmente desvio médio dos valores x_i em relação à média. O desvio médio em relação a média aritmética ou, simplesmente, desvio médio é definido por:

$$D_m = \frac{\sum |x_i - \bar{x}| * f_i}{\sum f_i}, \text{ para dados agrupados}$$

Ou

$$D_m = \frac{\sum |x_i - \bar{x}|}{n}, \text{ para dados não agrupados}$$

O desvio médio é a média aritmética dos desvios absolutos em relação à média. Quanto menos dispersos se encontram os valores x_i relativamente à \bar{x} menor será o desvio médio, e reciprocamente

4.2.4. Desvio Padrão

Dá-se o nome de variância dos valores de uma X de uma amostra a:

$$s^2 = \frac{\sum (x_i - \bar{x})^2 * f_i}{\sum f_i - 1}, \text{ para dados agrupados}$$

Ou

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}, \text{ para dados não agrupados}$$

Nota: A variância corresponde ao momento central da 2ª ordem.

O desvio padrão (ou desvio quadrado médio) é a raiz quadrada da variância:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2 * f_i}{\sum f_i - 1}}, \text{ para dados agrupados}$$

Ou

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}, \text{ para dados não agrupados}$$

O desvio padrão indica a proximidade com que os valores estão agrupados á volta da média. Um valor pequeno do desvio padrão significa que as observações estão pouco dispersos á volta da média.

O desvio padrão pode servir como medida de incerteza. Em ciências, a precisão de medições repetidas é dada pelo desvio padrão. O desvio padrão é crucial para analisar se as medições batem com a previsão teórica. Se a média das medições estiver muito longe da previsão teórica (distância medida pelo desvio padrão), então a teoria testada provavelmente precisa ser revisada. Enquanto o desvio padrão mede a distância dos valores típicos da média, outras medidas estão disponíveis. É o exemplo do desvio médio absoluto, que pode ser considerado uma medida mais directa da distância da média em comparação à distância da raiz quadrada média inerente ao desvio padrão.

Exemplo 4.17: Desvio padrão (dados não agrupados)

A tabela abaixo mostra as notas de matemática de um aluno em um determinado ano:

1º Bimestre	3,5
2º Bimestre	7,5
3º Bimestre	9,0
4º Bimestre	6,0

Determine o desvio padrão

Solução

$$s = \sqrt{\frac{\sum(3.5 - 6.5)^2 + (7.5 - 6.5)^2 + (9 - 6.5)^2 + (6 - 6.5)^2}{4 - 1}} = 2.35$$

Interpretação: A dispersão das notas de matemática deste aluno é de 2.35 medido pelo desvio padrão.

Exemplo 4.18: Desvio padrão (dados agrupados)

A tabela a seguir representa o salário mensal em meticais de funcionários numa fábrica de tijolos:

Salário	f_i
[2000---4000[10
[4000---6000 [14
[6000---8000 [11
[8000---10000 [18
[10000---12000 [12
[12000---14000 [15
Total	80

Determine o desvio padrão do salário dos funcionários desta fábrica.

Solução

Recorrendo a tabela temos:

Classe	x_i	f_i	$(x_i - \bar{x})^2 * f_i$
[2000---4000[3000	10	283556250
[4000---6000 [5000	14	154778750
[6000---8000 [7000	11	19311875
[8000---10000 [9000	18	8201250
[10000---12000 [11000	12	85867500
[12000---14000 [13000	15	327834375
Total		80	879550000

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2 * f_i}{\sum f_i - 1}} = \sqrt{\frac{879550000}{77}} = 3336,7$$

4.2.5. Coeficiente de Variação

O coeficiente de variação de Pearson é uma medida de dispersão relativa, empregada para estimar a precisão de experimentos e representa o desvio-padrão expresso como percentagem da média. Sua principal qualidade é a capacidade de comparação de distribuições diferentes.

O coeficiente de variação em uma carteira de activos serve como medida de risco para cada unidade de activo. O uso do coeficiente de variação é usualmente recomendado para variáveis quantitativas do tipo razão (na qual exista um zero absoluto), tais como altura, peso e velocidade.

Se a variável não é do tipo razão (ex: temperatura em graus Celsius), o coeficiente de variação poderá assumir valores negativos (ex: caso a média seja negativa) e sua interpretação dependerá do ponto de referência (ponto considerado como "0" na escala), levando a interpretações equivocadas e relativas.

A fórmula para o cálculo do coeficiente de variação é:

$$CV = \frac{s}{\bar{x}} * 100\%$$

Deve ser interpretado como a variabilidade dos dados em relação à média. Quanto menor for o coeficiente mais homogêneo será o conjunto de dados;

O coeficiente de variação é adimensional, isto porque: será positivo se a média for positiva e será zero quando não houver variabilidade, isto é, $s = 0$.

Quanto à representatividade em relação à média, podemos dizer que quando o coeficiente de variação (CV) é ou está:

- Menor que 10%: significa que é um ótimo representante da média, pois existe uma pequena dispersão (desvio padrão) dos dados em torno da média;
- Entre 10% e 20%: é um bom representante da média, pois existe uma boa dispersão dos dados em torno da média;
- Entre 20% e 35%: é um razoável representante da média, pois existe uma razoável dispersão dos dados em torno da média;
- Entre 35% e 50%: representa fracamente a média, pois existe uma grande dispersão dos dados em torno da média;
- Acima de 50%: não representa a média, pois existe uma grandíssima dispersão dos dados em torno da média.

Exemplo 4.19: Coeficiente de variação

Um empresário pode investir em dois possíveis produtos A e B, cujos preços variam. Os dados observados permitiram calcular as seguintes medidas:

$$\bar{x}_A = 70; \bar{x}_B = 150$$

$$s_A = 30; s_B = 40$$

Qual dos produtos apresenta maior homogeneidade nos seus preços?

Solução

- *Tomando os desvios padrões, o produto A parece preferível a B, porque possui menor variação de preços.*
 - *Recorrendo ao coeficiente de variação temos:*

Produto	\bar{x}	S	CV
A	70	30	0.429
B	150	40	0.267

Solução: O produto que apresenta os preços mais homogêneos é o A, porque possui menor coeficiente variação.

4.2.6. Intervalo interquartil

Uma medida de dispersão alternativa que pode ser empregada é o chamado intervalo interquartil ou amplitude interquartil. Essa medida só aproveita 50% dos dados e é pouco influenciada pelos valores extremos.

É a diferença entre o terceiro e o primeiro quartil, isto é:

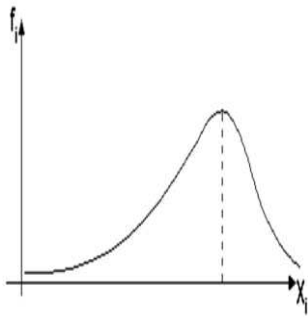
$$IQ = Q_3 - Q_1 = P_{75} - P_{25}$$

4.3. Medidas de Assimetria

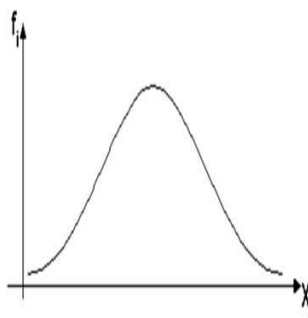
A assimetria é a falta de simetria do histograma, ou da curva de frequências, em relação à vertical que passa pela abcissa correspondente à média assimétrica.

As medidas de assimetria sintetizam até que ponto uma distribuição de frequência é enviesada, deformada ou assimétrica. Estas medidas utilizam-se para classificar distribuições unimodais e elucidam sobre a forma geral da distribuição, isto é, se é simétrica ou, em caso contrário, se afasta muito ou pouco da simetria. Quando a distribuição é simétrica, o coeficiente de assimetria é nulo. Quando não é nulo, a distribuição é assimétrica, sendo o grau de assimetria tanto maior quanto maior for o valor absoluto do coeficiente.

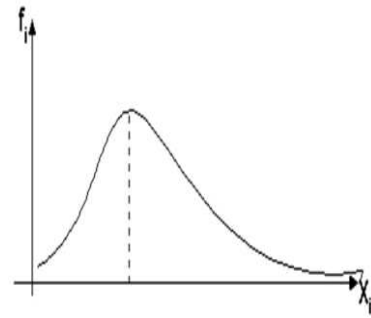
As medidas de assimetria permitem distinguir as distribuições simétricas ($\bar{x} = \hat{x} = \tilde{x}$) das assimétricas. No caso das distribuições assimétricas estas podem ter assimetria positiva ($\hat{x} \leq \tilde{x} \leq \bar{x}$) ou assimetria negativa ($\bar{x} \leq \tilde{x} \leq \hat{x}$).



Curva Assimétrica Negativa



Curva Simétrica



Curva Assimétrica Positiva

Quatro indicadores de assimetria podem calcular-se na ausência de uma imagem esclarecedora e de acordo com os indicadores disponíveis:

$$C_a = \frac{\bar{x} - \hat{x}}{S}$$

- Coeficiente de Assimetria de Pearson

$$C_a = \frac{3 * (\bar{x} - \tilde{x})}{S}$$

- Coeficiente de Assimetria de Pearson

$$C_a = \frac{Q_1 + Q_3 - 2 * Q_2}{Q_3 - Q_1}$$

- Coeficiente Quartílico de Assimetria

$$C_a = \frac{m_3}{S^3}$$

- Coeficiente Momento de Assimetria

Para qualquer dos indicadores:

- Uma distribuição **simétrica** resultará num valor igual a 0 (zero);
- Se a distribuição for **assimétrica positiva** resultará num valor superior a 0 (zero);
- Se a distribuição for **assimétrica negativa** resultará num valor inferior a 0 (zero).

4.4. Medidas de Achatamento ou Curtose

Para definirmos e visualizarmos o achatamento de uma distribuição de frequências, necessitamos da denominada curva normal.

O achatamento de uma distribuição refere-se á intensidade das frequências nos valores vizinhos dos valores centrais. As mediadas de achatamento medem o grau de afunilamento ou achatamento de uma curva simétrica ou aproximadamente simétrica em relação a curva normal.

A classificação da distribuição de frequência, relativamente ao seu achatamento, pode ser feita através do cálculo de dois indicadores:

Coefficiente Percentílico de Curtose

$$K = \frac{Q_3 - Q_1}{2 * (P_{90} - P_{10})}$$

Por comparação com a distribuição normal cujo grau de curtose é 0.263.

- Se $K = 0.263$, a distribuição é **mesocúrtica**;
- Se $K < 0.263$, a distribuição é **leptocúrtica**;
- Se $K > 0.263$, a distribuição é **platicúrtica**.

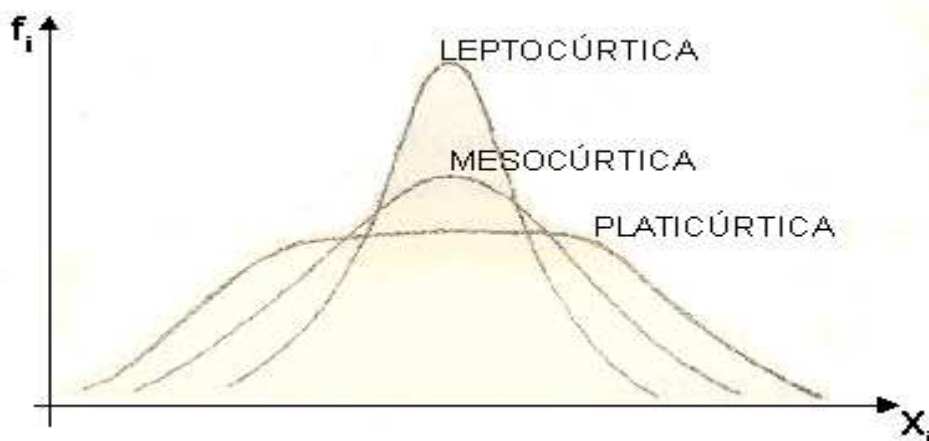
Coefficiente Momento de Curtose

$$K = \frac{m_4}{s^4}$$

Por comparação com a distribuição normal cujo grau de curtose é 3.

- Se $K = 3$, a distribuição é **mesocúrtica**;
- Se $K < 3$, a distribuição é **leptocúrtica**;
- Se $K > 3$, a distribuição é **platicúrtica**.

Numa distribuição unimodal, quanto maior for a concentração de valores em torno do centro da mesma, maior será o valor da sua curtose. Graficamente isto será associado a uma curva com a parte central mais afilada, mostrando um pico de frequência simples mais destacado, mais pontiagudo, caracterizando a moda da distribuição de forma mais nítida, conforme o gráfico abaixo:



4.5. EXERCÍCIOS PROPOSTOS

4.5.1. A poluição causada por óleo em mares e oceanos estimula o crescimento de certos tipos de bactérias. Uma contagem de microorganismos presentes no petróleo (número de bactérias por 100 mililitros), em 10 porções de água do mar, indicou as seguintes medidas:

49 70 54 67 59 40 71 67 67 52

- Determine e interprete a média, mediana e moda.
- Calcule o desvio padrão.

4.5.2. Para os valores (escores) 205, 6, 5, 5, 5, 2 e 1 calcule a moda, a mediana e a média aritmética. Além disso, responda que medida de tendência central não deveria ser usada para descrever esse conjunto de escores? Por quê?

4.5.3. Em certa empresa trabalham 4 analistas de mercado, 2 supervisores, 1 chefe de seção e 1 gerente que ganham, respectivamente: 1.300,00Mt; 1.600,00Mt; 2.750,00Mt; 5.000,00Mt. Qual o valor do salário médio desses funcionários?

4.5.4. A comunidade "A" tem 100 motoristas profissionais, cujo salário médio é de 950,00 Dólares. A comunidade "B", com 300 desses profissionais, remunera-os com uma média de 800,00 Dólares.

- É correto afirmar que "A" remunera melhor seus motoristas que "B"?
- Diante das informações disponíveis há garantia que os 100 salários individuais de "A" são maiores que os 300 de "B"? Por que?

4.5.5. O revisor de um jornal fez durante um mês o levantamento dos erros ortográficos encontrados no editorial do jornal. Os resultados encontrados foram:

0 1 0 1 0 0 0 0 2 3 0 3 5 2 0 1 5 3
0 1 2 3 4 0 0 0 1 4 1 5 2 1 1 2 1 0
0 1 2 1 2 5 4 4 3 4 1 2 4 3 5 1 0 0

- Faça uma distribuição de frequência dos dados.
- Calcule as medidas de tendência central.
- Calcule as medidas de dispersão.

4.5.6. Durante certo período de tempo as taxas de juros para dez ações foram as abaixo registadas:

Acção	1	2	3	4	5	6	7	8	9	10
Taxa (%)	2.59	2.64	2.60	2.62	2.55	2.61	2.50	2.63	2.64	2.69

Calcule:

- A taxa média;
- A taxa mediana;
- A taxa modal;
- O desvio padrão das taxas;
- O coeficiente de variação das taxas;
- O coeficiente de assimetria.

4.5.7. Abaixo são mostrados os saldos médios de 48 contas de clientes do BB Novo S.A. (dados brutos em 1000 Mt).

450	500	150	1000	250	275	550	500
225	475	150	450	950	300	800	275
600	750	375	650	150	500	1000	700
475	900	800	275	600	750	375	650
150	500	225	250	150	120	250	360
230	500	350	375	470	600	1030	270

- Agrupe os dados numa distribuição de frequências.
- A partir de (a), calcule média, mediana e moda. Interprete.

4.5.8. Considere os dados referentes ao consumo de energia em Kw de 75 contas da EDM.

32	40	22	11	34	40	16	26	23	31	49
10	38	17	13	45	25	50	18	23	35	56
22	30	14	18	20	13	24	35	29	33	19
48	20	12	31	39	17	58	19	16	12	11
21	15	12	20	51	12	19	15	41	29	15
25	13	23	32	14	27	43	37	21	28	
37	26	44	11	53	38	46	17	36	28	

- Agrupar os dados em uma distribuição de frequência, em intervalos fechados à direita e com amplitude 10. Utilize o limite inferior da distribuição igual a zero.
- Calcule as medidas de tendência central: moda, média e mediana.
- Encontre as medidas de variabilidade: amplitude, variância absoluta, desvio padrão e coeficiente de variação.

4.5.9. Os 20 alunos de uma turma especial de Estatística obtiveram as notas abaixo.

84 88 78 80 89 94 95 77 81 90
83 87 91 83 92 90 92 77 86 99

Determine:

- A amplitude total das notas;
- O desvio padrão das notas;
- A variância absoluta das notas;
- O coeficiente de variação;
- A proporção de alunos com notas maiores que 90;
- A média, sabendo que o professor acrescentou 5 pontos para cada aluno;
- O desvio padrão, quando foi adicionado 5 pontos.

4.5.10. Um fabricante de embalagens recebeu uma encomenda de um cliente que fabrica margarina. Para isso, apresentou ao cliente três tipos diferentes de embalagens, A, B e C, segundo a pressão média de rompimento para cada uma delas. O cliente optou pelo tipo de embalagem que possuísse a menor variação absoluta na pressão de ruptura.

Tipos de embalagens	A	B	C
Pressão média de ruptura (Bária)	300	150	200
Desvio-padrão das pressões (bária)	60	40	50

- Qual das três embalagens o cliente optou? Por quê?
- Se a sua escolha fosse apoiada na maior variação relativa qual das três embalagens ele teria escolhido? Por quê?

4.5.11. As informações abaixo indicam o número de acidentes ocorridos com 70 motoristas de uma empresa de ônibus nos últimos 5 anos:

Nº DE ACIDENTES	0	1	2	3	4	5	6	7
Nº DE MOTORISTAS	15	11	20	9	6	5	3	1

- Qual a média de acidentes?
- E a moda de acidentes?
- E a mediana?

4.5.12. Uma distribuidora de refrigerantes fez um levantamento sobre o consumo semanal (em litros) por pessoa, em jan/2002, em uma cidade do litoral, obtendo:

Consumo	[0---0.5[[0.5---1.0[[1.0---1.5[[1.5---2.0[[2.0---2.5[
Nº de Pessoas	10	25	9	7	6

- Determine e interprete o consumo médio.
- Qual o percentual de pessoas que consomem menos de 1 litro por semana?
- Determine e interprete o consumo modal e o consumo mediano.
- Se a empresa tem um lucro de 0,50Mt por litro, qual o lucro médio por pessoa?
- Calcule o desvio padrão.
- Represente graficamente o consumo.

4.5.13. Uma companhia distribuidora tem por hipótese que uma chamada telefónica é mais eficiente que uma carta para acelerar a cobrança de contas atrasadas. Esta companhia fez uma experiência usando duas amostras e obteve os seguintes resultados:

Método utilizado	Nº de dias até o pagamento						
Carta	10	8	9	11	11	14	10
Chamada telefónica	7	4	5	4	8	6	9

- Qual dos métodos apresentou resultados mais homogêneos? Justifique através do coeficiente de variação.
- Se houver mais um dia de atraso em TODAS as contas, o que acontecerá neste caso com o coeficiente de variação anterior.
- Se dobrar o tempo até o pagamento de TODAS as contas observadas, o que acontecerá com a variância do grupo que recebeu cobrança através de carta.

4.5.14. Os salários (em 1000 Mt) iniciais de funcionários de uma empresa estão descritos na tabela abaixo:

Classe de salários (1000 Mt)	[8--10[[10--12[[12--14[[14--16[[16--18[[18--20[
Nº de funcionários	7	12	10	8	4	6

- a) Determine o salário médio, mediana e modal dos funcionários;
- b) Determine o intervalo Inter-Quartil.
- c) Classifique a distribuição quanto a assimetria e achatamento

4.5.15. Os operários de um sector industrial têm, em Julho, um salário médio de 5 salários mínimos (s.m.) e desvio padrão de 2 s.m. Um acordo colectivo prevê, para agosto, um aumento de 60%, mais uma parte fixa correspondente a 0,7 s.m. Qual a média do número de salários mínimos e qual o desvio padrão em agosto?

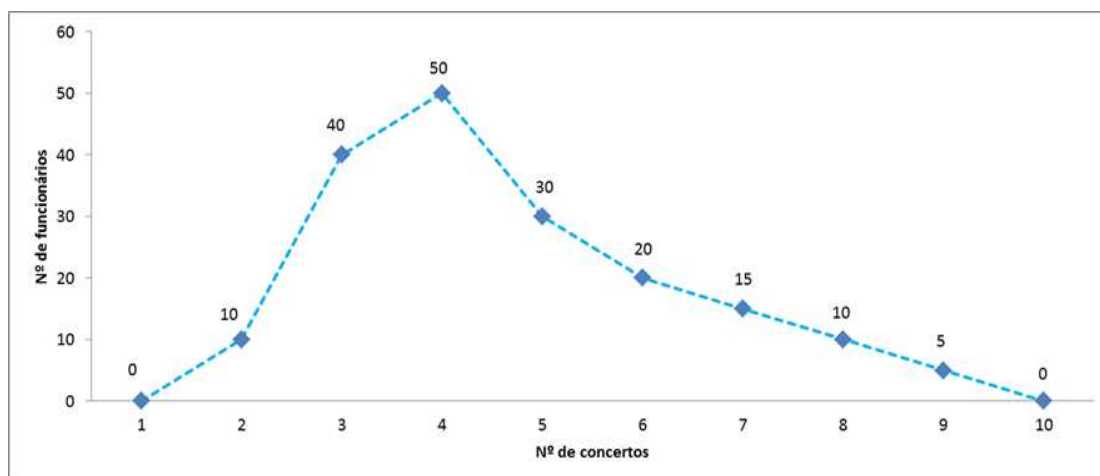
4.5.16. Uma Distribuição A possui média aritmética igual a 50 e desvio padrão igual a 5. Uma outra distribuição B possui média aritmética igual a 40 e variância igual a 36. Qual delas tem a média aritmética mais representativa?

4.5.17. A tabela seguinte apresenta a produção de castanha, em milhões de toneladas, na região norte do país.

Ano	1992	1993	1994	1995	1996	1997	1998	1999	200	2001
Toneladas	12	15	18	22	17	14	18	23	29	32

- a) Calcule o valor da produção média;
- b) Calcule o valor da moda da produção;
- c) Calcule o valor da mediana da produção;
- d) Calcule o valor do desvio padrão da produção;
- e) Calcule o valor do coeficiente de variação;

4.5.18. Considere gráfico seguinte:



- Que tipo de gráfico se trata?
- Indique a variável em estudo.
- Determine a variação relativa de concertos diários realizados pelos funcionários da Delta Limitada.
- Indique o número máximo de concertos diários realizados pelos 10%, 25%, 50%, 68%, 75% e 90% dos funcionários da Delta Limitada.
- Determine o número modal de concertos diários realizados pelos funcionários da Delta Limitada.
- Esboce o Polígono de frequências acumuladas e o Histograma correspondente.

4.5.19. A empresa do ramo calçadista Pé nas Nuvens Calçados foi fundada em 1990. Desde então, sempre buscou trabalhar com uma abordagem de gestão inovadora, e tendo em vista o aumento de produtividade de seus vendedores, resolveu premiar com um aumento de 5% no salário, 36% de seus vendedores mais eficientes. Para isto, fez um levantamento de vendas semanais por vendedor. Os dados obtidos estão descritos na tabela abaixo.

Vendas (USD 1000)	[0--10[[10--20[[20--30[[30--40[[40--50[
Nº de vendedores	10	25	9	7	6

- A partir de qual volume de vendas o vendedor da Pé nas Nuvens Calçados será premiado?
- Determine o número modal do volume de vendas da empresa.

4.5.20. O departamento de pessoal de uma certa firma fez um levantamento dos salários dos funcionários do sector administrativo, tendo obtido os seguintes resultados.

Faixa salarial (1000 Mt)	[0--2[[2--4[[4--6[[6--8[[8--10[
Frequência relativa	0.05	0.20	0.35	0.25	0.15

- Calcule aproximadamente a média, a variância e o desvio padrão dos salários.
- Se for concedido um aumento de 100% a todos os funcionários, haverá alteração na média dos salários? E na variância dos salários? Justifique.
- Classifique a distribuição salarial quanto a assimetria.

4.5.21. Em um grupo de 600 hóspedes de determinado hotel, tem-se os seguintes valores com relação ao tempo de permanência no hotel:

- Média = 9 dias;
- 1º Quartil = 5 dias;
- 3º Quartil = 15 dias;
- Coeficiente de variação = 20%.

Pede-se:

- a) Quantos hóspedes permaneceram mais de 15 dias;
- b) Quantos hóspedes permaneceram entre 5 e 15 dias;
- c) O desvio padrão para o tempo de permanência;

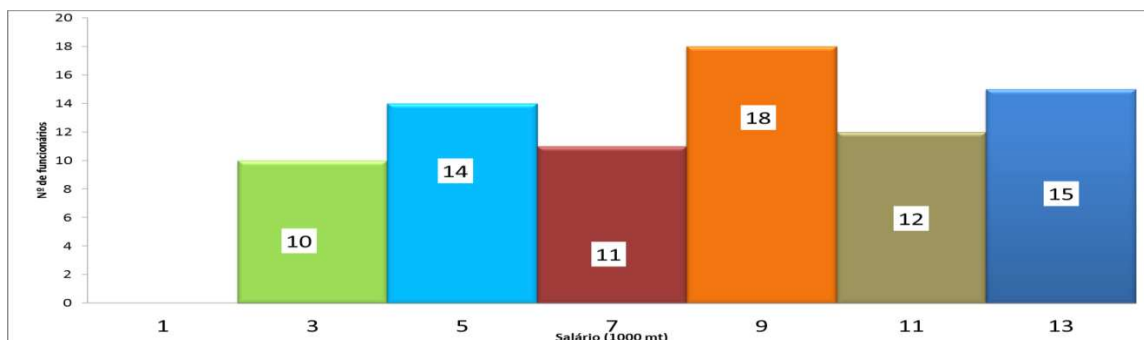
4.5.22. Uma cerâmica fabrica tijolos de acordo com a norma de um grande cliente. A norma estabelece que os tijolos devem suportar no mínimo uma força de compressão média de 10 kg/cm² e que o desvio padrão não deve ser superior a 5% da média. Num ensaio realizado em um lote de tijolos pelo Engenheiro da Qualidade do cliente, foram registrados os seguintes dados de uma amostra de 6 tijolos, para sua resistência à compressão em kg/cm²: 12; 11; 10; 9; 8,5 e 11,5. Nestas condições, o Engenheiro da Qualidade aprovará ou reprovará o lote de tijolos?

4.5.23. Um levantamento dos preços à vista de gasolina e de álcool, em alguns postos da cidade, está mostrado na tabela abaixo (em USD)

Gasolina	2.61	2.64	2.56	2.62	2.60	2.58
Álcool	1.90	1.79	1.88	1.81	1.88	1.84

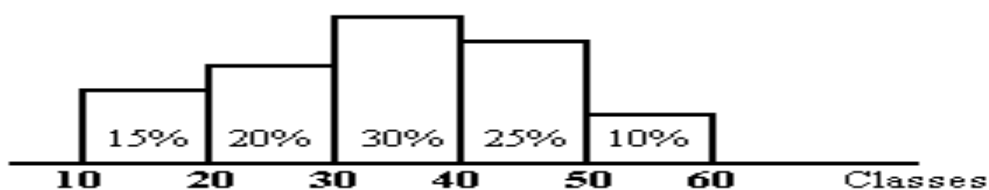
- a) Qual é a média, o desvio padrão e o coeficiente de variação dos preços de cada combustível?
- b) Qual é o combustível que tem seus preços mais homogêneos?

4.5.24. O histograma a seguir representa os salários de 80 funcionários duma certa firma:



- Qual é a percentagem de funcionários com salário inferior a 8000.00 Mt?
- Determine o salário médio;
- Determine o salário mediano;
- Determine o salário modal;
- Classifique a distribuição salarial quanto a simetria.

4.5.25. Dado o histograma a seguir, determinar a média, mediana, moda, o desvio padrão e o 1º quartil da distribuição



4.5.26. Um caminhão cujo peso vazio é de 3.200 kg será carregado com 470 caixas de 11 kg cada, 360 caixas de 9 kg cada, 500 caixas de 4 kg cada e 750 caixas de 6 kg cada. O motorista do caminhão pesa 75 kg e a lona de cobertura da carga pesa 48 kg.

- Sabendo-se que este caminhão tem que passar por uma balança que só permite a passagem de veículos com peso máximo de 16 toneladas, pergunta-se: Ele passará pela balança?
- Qual o peso médio das caixas carregadas no caminhão?

4.5.27. Um pesquisador dispõe das seguintes informações a respeito de uma amostra:

- Média = 50,34
- Soma do quadrado dos valores = 150.000
- Número de elementos da amostra = 54

Calcular as medidas de dispersão possíveis a partir das informações fornecidas.

4.5.28. A série de dados abaixo se refere às medidas tomadas de uma amostra de cães.

Cão	1	2	3	4	5	6	7	8	9	10
Peso (kg)	23,0	22,7	21,2	21,5	17,0	28,4	19,0	14,5	19,0	19,5
Comprimento (cm)	104	105	103	105	100	104	100	91	102	99

Pedem-se, para cada característica avaliada (peso e comprimento), as estatísticas:

- Média;
- Mediana;
- Moda;
- Variância;
- Desvio Padrão;
- Coeficiente de Variação;
- Qual das duas características é mais homogênea?
- Coeficiente de Correlação entre as duas características.

4.5.29. Considere a seguinte tabela de distribuição de frequências com os tempos (em dias) que um corrector demora a concluir um negócio, observado em 40 operações:

Tempo (dias)	[0---2.5[[2.5---5.0[[5.0---7.5[[7.5---10.0[
Nº de operações	10	25	9	7

Calcule:

- A média aritmética, a moda e a mediana.
- A variância, o desvio padrão.
- O coeficiente de variação.
- O 3º quartil e o 4º percentil.

4.5.30. Numa sala estão 5 pessoas. A média das idades é de 30 anos. Entra na sala uma pessoa de 36 anos de idade. Qual será agora a média das idades das pessoas naquela sala?

4.5.31. O Departamento de Comércio Exterior do Banco Central possui 30 funcionários com a seguinte distribuição salarial em dólares.

Nº de funcionários	10	12	5	3
Salário em USD	2000	3600	4000	6000

- Determine o salário máximo auferido pelos 10%, 25%, 50%, 75% e 90% dos funcionários deste departamento.
- Classique a distribuição salarial quanto ao achatamento e assimetria.
- Quantos funcionários que recebem 3600 USD devem ser demitidos para que a mediana desta distribuição de salários seja de 2800 USD?
- Suponha que sejam contratados dois novos funcionários com salários de 3000 USD cada. A média e a variância da nova distribuição de salários ficará menor, igual ou maior que a anterior? Justifique.

4.5.32. No quadro seguinte apresentam-se o número de transacções efectuadas em cada uma das lojas dos Supermercados XXX em u.m, classificadas por níveis de despesa, e o número de empregados existentes em cada uma delas.

Escalão de despesas	[0---10[[10---20[[20---30[[30---40[Nº de Empregados
Loja A	29	44	26	9	20
Loja B	74	78	30	18	30

- Será possível afirmar que “em ambas as lojas, mais de 70% das transacções têm um valor inferior a 20 u.m.”?
- Represente graficamente o polígono de frequências de cada uma das distribuições.
- Determine o valor médio por transacção e o valor médio das transacções por empregado, em cada uma das lojas.
- Calcule o desvio padrão da distribuição das transacções na loja 2 sabendo que o valor correspondente para a outra loja é de 9,1 u.m. Em qual das duas distribuições é mais elevada a dispersão? Justifique.

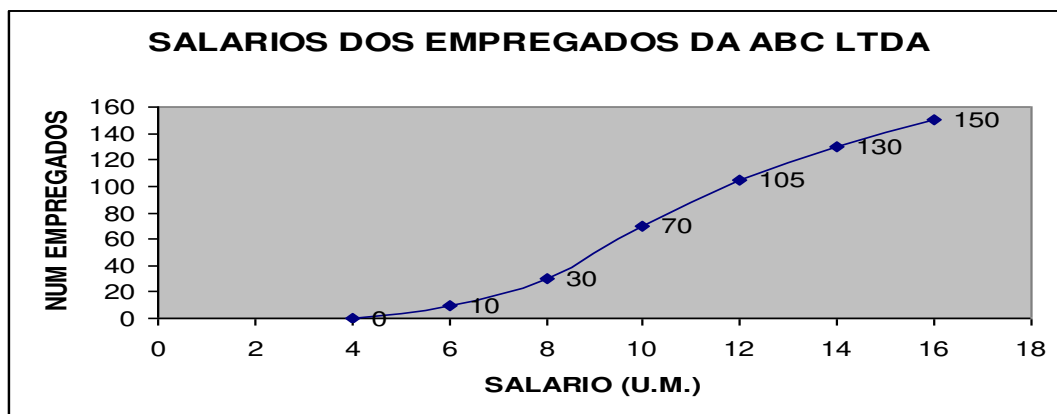
4.5.33. Uma turma obteve as seguintes notas

Notas	[0---2[[2---4[[4---6[[6---8[[8---10[
Frequência	4	16	24	30	6

- O professor da turma ofereceu bolsas para os 5% melhores e um programa de reforço para os 8% piores, Qual a menor nota dos bolsistas? Qual a maior nota dos 8% piores?
- Determine a nota média da turma e o desvio padrão.
- O professor acrescentou 0,5 pontos a nota da prova de todos os alunos por um exercício extra resolvido por estes alunos, determina nota média da turma e o desvio padrão.

4.5.34. Os dados a seguir representam o número de apólices de seguro que um corretor conseguiu vender em cada um de seus 20 primeiros dias em um emprego novo: 2, 4, 6, 3, 2, 1, 4, 3, 5, 2, 1, 1, 4, 0, 2, 2, 5, 2, 2, 1
 Calcule a média, a mediana e a moda desses dados, interpretando os resultados obtidos.

4.5.35. Considere o gráfico seguinte:



- Indique e classifique a variável em estudo quanto ao tipo.
- Determine a variação relativa dos salários auferidos pelos empregados da ABC Limitada.
- Indique o número máximo dos salários auferidos pelos 10%, 25%, 68% e 90% dos empregados da ABC Limitada.
- Determine o intervalo interquartil.
- Determine o salário modal dos empregados da ABC Limitada.
- Esboce o Histograma correspondente.

4.5.36. Em um aviário foi observada a distribuição dos frangos com relação ao peso apresentado conforme a tabela a seguir:

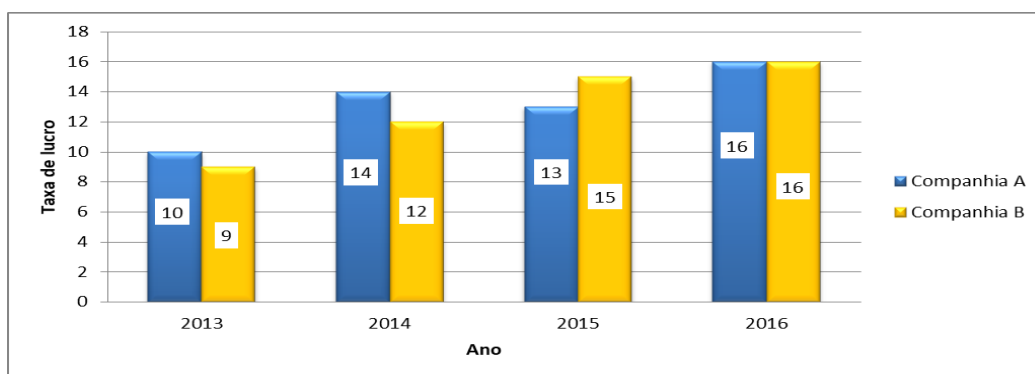
Peso (kg)	[0.96-0.98[[0.98-1.00[[1.00-1.02[[1.02-1.04[[1.04-1.06[[1.06-1.08[
Nº de frangos	60	160	280	260	160	80

- Determine o peso médio?
- Queremos dividir os frangos em 4 categorias, com relação ao peso, de modo que:
 - os 20% mais leves sejam da categoria D
 - os 30% seguintes sejam da categoria C
 - os 30% seguintes sejam da categoria B
 - os 20% restantes sejam da categoria A

Apresente a tabela dos limites de peso entre as categorias A, B, C, D?

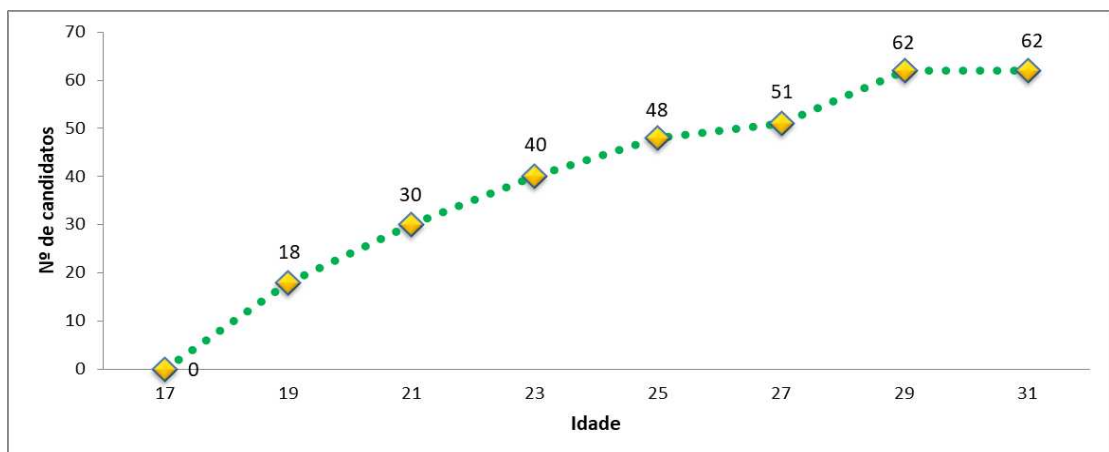
- c) Determine o peso modal.
- d) O proprietário decide separar deste lote os animais com peso inferior a um desvio padrão abaixo da média para receberem ração reforçada e também separar os animais com peso superior a $3/4$ desvios padrões acima da média para usá-los como reprodutores. Qual a percentagem de animais que serão separados em cada caso?

4.5.37. Para se estudar o desempenho de 2 companhias correctoras de acções, seleccionou-se amostras das acções negociadas por cada uma delas. Para cada acção seleccionada, computou-se a taxa de lucro apresentada durante um período de 2013 a 2016, obtendo-se o gráfico abaixo.



- a) Apresente os dados em tabela.
- b) Qual companhia teve melhor desempenho?
- c) Caracterize a distribuição do lucro para cada companhia quanto a achatamento (use a fórmula dos momentos).

4.5.38. A idade média dos candidatos a um determinado curso de aperfeiçoamento oferecido por uma empresa foi sempre baixa, da ordem de 24 anos. Como esse curso foi preparado para todas as idades, decidiu-se fazer uma campanha de divulgação. Para verificar se a campanha foi ou não eficiente, fez-se um levantamento da idade dos candidatos à última promoção, obtendo-se os resultados do gráfico a seguir:



- Que tipo de gráfico se trata?
- Apresente os dados em tabela (limites da classe, ponto médio da classe, frequência absoluta simples e acumulada).
- Baseando-se nesses resultados, você diria que a campanha surtiu o efeito desejado?
- Um outro pesquisador decidiu usar o seguinte critério: se a diferença $24 - \bar{x}$ fosse maior que o valor $\alpha = \frac{2 \cdot S}{\sqrt{n}}$, então a campanha teria sido efectiva. Qual a conclusão dele?
- Determine a idade mediana e modal dos candidatos.
- Se a efectividade do programa fosse baseada em distribuição assimétrica positiva, qual seria a sua conclusão?
- Classifique a distribuição das idades dos candidatos quanto a achatamento.

4.5.39. Considere a tabela abaixo, sobre avaliações de nove lugares para se hospedar na Europa, responda:

- Qual é a percentagem de hotéis com preços de quartos iguais à \$\$\$?
- Calcule a pontuação global média dos quartos.
- Determine a mediana da pontuação dos quartos.

Nome do Estabelecimento	País	Preço do Quarto	Número de Quartos	Pontuação por quarto
Graveteye Manor	Inglaterra	\$\$	18	83.6
Villa d'Este	Itália	\$\$\$\$	166	86.3
Hotel Prem	Alemanha	\$	54	77.8
Hotel d'Europa	França	\$\$	47	76.8
Palace Luzern	Suíça	\$\$	326	80.9
Royal Crescent Hotel	Inglaterra	\$\$\$	45	73.7
Hotel Sacher	Áustria	\$\$\$	120	85.5
Duc de Bourgogne	Bélgica	\$	10	76.9
Villa Gallici	França	\$\$	22	90.6

5. MÉTODO DE CONTAGEM

A princípio, pode parecer desnecessária a existência de métodos para realizar uma contagem. Isto de fato é verdade se o número de elementos que queremos contar for pequeno. Entretanto, se o número de elementos for grande, a contagem pode se tornar uma tarefa árdua.

5.1. Factorial

Seja n um número natural (inteiro não negativo). O factorial de n , indicado por $n!$, é definido por:

$$n! = n * (n - 1) * (n - 2) \dots 3 * 2 * 1, \text{ para } n \geq 2$$

Exemplo 5.1 (Factorial)

- $4! = 4 * 3 * 2 * 1$
- $5! = 5 * 4 * 3 * 2 * 1$
- $0! = 1$
- $1! = 1$

5.2. Princípios fundamentais

Usaremos as palavras selecção e arranjo no seu sentido usual. Assim não deve haver ambiguidade em sentenças como:

- Seleccionemos 2 candidatos a representantes discentes entre 6 alunos.
- Existem 15 maneiras de seleccionarmos 2 candidatos a representantes discentes entre 6 alunos.
- Arranje os 4 CD's na prateleira.
- Existem 24 modos de se arranjar 4 CD's na prateleira.

Usaremos a palavra combinação como sinónimo de selecção e a palavra permutação como sinónimo de arranjo. As selecções ou combinações levam em conta somente a natureza dos objectos enquanto que os arranjos ou permutações consideram tanto a natureza quanto a ordem dos objectos.

Sejam n e r números inteiros positivos. Em todo este capítulo e nos seguintes, vamos denotar por $C(n, r)$ uma r -combinação de n objectos definida como sendo uma selecção

não ordenada de r dos n objectos. Tais objectos podem ser distintos ou indistintos e podemos considerar repetições ou não. Analogamente, vamos denotar por $P(n, r)$ uma r -permutação de n objectos definida como sendo um arranjo ordenado de r destes n objectos. Novamente, tais objectos podem ser distintos ou indistintos e podemos considerar repetições ou não.

Agora, consideremos um exemplo.

Exemplo 5.2

Consideremos as letras a, b, c e os algarismos $1, 2, 3, 4, 5$

- *Existem $3 * 5 = 15$ modos de seleccionarmos uma letra e um algarismo;*
- *Existem $3 + 5 = 8$ modos de seleccionarmos uma letra ou um algarismo*

As regras que acabamos de usar no exemplo acima estão descritas formalmente a seguir.

Regra do produto: Se um evento pode ocorrer m vezes e um outro evento pode ocorrer n vezes, então existem $m * n$ modos de ocorrerem ambos os eventos.

Regra da soma: Se um evento pode ocorrer m vezes e um outro evento pode ocorrer n vezes, então existem $m + n$ modos de ocorrerem ambos os eventos.

Nota: Se um evento pode ocorrer m vezes e um outro evento pode ocorrer n vezes, então existem $m + n$ modos de ocorrer um destes eventos.

Exemplo 5.3 (Soma e Produto)

Consideremos 5 livros escritos em inglês e distintos, 7 livros escritos em japonês e distintos e 10 livros escritos em francês e distintos. Então existem $5 * 7$ modos de escolhermos um livro em inglês e um em japonês, $5 * 10$ modos de escolhermos um livro em inglês e um em francês, e $7 * 10$ modos de escolhermos um livro em japonês e um em francês. Portanto existem:

$$5 * 7 + 5 * 10 + 7 * 10 = 155$$

Modos de escolhermos 2 livros de línguas diferentes. Por outro lado, existem

$$22 * 21 = 462$$

Maneiras de escolhermos 2 livros quaisquer (um depois o outro) entre os 22 livros.

Pela regra do produto, uma r -permutação de n objectos distintos pode ser considerada como uma selecção de r objectos entre n objectos seguida de um arranjo dos r objectos seleccionados, ou seja:

$$P(n, r) = C(n, r) * P(r, r)$$

5.3. Permutações

A seguir, vamos deduzir uma fórmula para $P(n, r)$ e, na secção seguinte, consideraremos combinações $C(n, r)$.

5.3.1. Permutações de objectos distintos

Consideremos inteiros positivos n e r tais que $r < n$. Como a ordem dos objectos num arranjo deve ser levada em conta, segue que, quando arranjamos r entre n objectos distintos, devemos considerar que cada objecto “ocupa” uma posição determinada. Deste modo, existem n modos de ocuparmos a primeira posição. Depois de ocupada a primeira posição, podemos ocupar a segunda posição de $n - 1$ modos diferentes. Similarmente, depois de ocupada a segunda posição (e também a primeira), podemos ocupar a terceira posição de $n - 2$ modos diferentes e assim por diante até chegarmos à r -ésima posição. Finalmente a r -ésima posição pode ser ocupada de $n - (r - 1) = n - r + 1$ modos diferentes. Daí, pela regra do produto, concluímos que:

Teorema 1.5. Se r e n são inteiros positivos com $r < n$, então o numero de r permutações de n objectos distintos é:

$$P(n, r) = n * (n - 1) \dots (n - r + 1)$$

Ou seja:

$$P(n, r) = \frac{n!}{(n - r)!}$$

Definimos $P(n, 0) = 0$ para qualquer inteiro n não negativo.

Exemplo 5.4 (Permutações)

O número de maneiras pelas quais podemos arranjar 5 entre 7 objectos distintos é:

$$P(7, 5) = \frac{7!}{(7 - 5)!} = \frac{7!}{2!} = 2520$$

Teorema 2.5. A n -permutação de n objectos distintos é dada por:

$$P(n, n) = n!$$

Demonstração: Mostremos a tese por indução. É claro que $P(1, 1) = 1 = 1!$ Suponhamos que $P(n - 1, n - 1) = (n - 1)!$ e provemos que $P(n, n) = n!$ Para arranjarmos n objectos em ordem, basta tomarmos um determinado objecto e arranjarmos os demais $n - 1$. Para cada arranjo dos $n - 1$ objectos, existem n posições possíveis para o objecto tomado. Logo, pela regra do produto,

$$P(n, n) = n * P(n - 1, n - 1) = n * (n - 1)! = n!$$

e a prova está completa.

Exemplo 5.5 (Permutações)

O número de maneiras pelas quais n pessoas podem se posicionar em fila é:

$$P(n, n) = n!$$

Exemplo 5.6 (Permutações)

De quantos modos n pessoas podem se posicionar em pé formando um círculo?

Pelo exemplo anterior, $P(n, n) = n!$ é o número de arranjos que podem ser feitos para enfileirar n pessoas. Porém, para arranjarmos as pessoas em círculo, somente as posições relativas das pessoas são importantes, pois consideramos iguais quaisquer 2 arranjos que podem ser obtidos um através do outro por rotação. Desta forma, o número de arranjos

circulares é:

$$P(n, r) = \frac{n * (n - 1)!}{n!} = (n - 1)!$$

5.3.2. Permutações de objectos nem todos distintos

Até o momento, consideramos que os objectos que queremos arranjar podem ser distinguidos um do outro de alguma forma. Agora, vamos considerar objectos nem todos distintos entre si.

O teorema seguinte generaliza a ideia do Exemplo 5.6 acima.

Teorema 3.5. Sejam n objectos não todos distintos entre si. Destes n objectos, sejam q_1 do primeiro tipo, q_2 do segundo tipo, ... e q_t do t -ésimo tipo. Então o número de n -permutações destes n objectos é:

$$P(n, n) = \frac{n!}{q_1! q_2! \dots q_t!}$$

Demonstração. Imaginemos que os n objectos são “marcados” de forma que possamos distinguir cada um dos objectos do mesmo tipo. Então, pelo Teorema 2.5, existem $n!$ modos de permutarmos estes objectos “distintos”. Mas duas permutações serão iguais sempre que, ao “tirarmos” as marcas dos objectos do mesmo tipo, os arranjos feitos forem iguais. Portanto uma permutação de objectos não marcados corresponde a $q_1! * q_2 \dots q_t!$ permutações de objectos marcados e temos a fórmula de produto. (Reveja o Exemplo 5.6)

Exemplo 5.7 (Permutações)

O número de arranjar 3 segmentos e 4 pontos é:

$$\frac{7!}{3! 4!} = \frac{7 * 6 * 5}{3 * 2} = 7 * 5 = 35$$

5.3.3. Permutações de objectos distintos com repetições

A seguir, vamos considerar repetições de objectos distintos.

Teorema 4.5. Sejam r e n inteiros positivos. O número de maneiras pelas quais podemos arranjar r objectos entre n objectos distintos de forma que possam haver repetições é:

$$P(n, r) = n^r$$

Demonstração: Segue directamente aplicando-se a regra do produto, uma vez que há n modos de escolhermos um objecto para ocupar a primeira posição, n modos de escolhermos um objecto para ocupar a segunda posição e assim por diante até a r -ésima posição.

Observação: Em geral, temos $r > n$ no Teorema 4.5. Veja o exemplo a seguir.

Exemplo 5.8 (Permutações)

Dos 10 bilhões de números entre 1 e 10.000.000.000, quantos números contém o algarismo 1 e quantos não contém?

Consideremos os 9 algarismos 0, 2, 3, 4, 5, 6, 7, 8, e 9. A posição das unidades pode ser ocupada por um destes algarismos de 9 modos diferentes; a posição das dezenas

também pode ser ocupada de 9 modos diferentes e assim por diante. Então, dos 10 bilhões de números entre 0 e 9.999.999.999, existem 9^{10} números que não contém o algarismo 1. Portanto, entre os números 1 e 10.000.000.000, existem $9^{10} - 1$ números que não contém o algarismo 1 (já que o número 10.000.000.000 contém o algarismo 1 e não estava sendo contado entre 0 e 9.999.999.999). Logo

$$10^{10} - (9^{10} - 1)$$

números entre 1 e 10.000.000.000 contém o algarismo .

5.4. Combinações

Uma combinação sem repetição, em análise combinatória, é um subconjunto com r elementos em um conjunto U com n elementos. Como é um conjunto, não há repetição de membros dentro do conjunto.

O número de subconjuntos de s elementos diferentes de um conjunto de n elementos diferentes pode ser representado por: $C(n, r) = \binom{n}{r}$.

É importante notar a diferença entre combinação e arranjo sem repetição. Em uma combinação a ordem dos elementos não importa, ou seja, elementos que diferem apenas pela ordem são contados como um único elemento. Já em um arranjo, a ordem importa, ou seja, sequencias com os mesmos elementos, mas em ordem diferente são contadas separadamente.

5.4.1. Combinações de objectos distintos

Em vista aos teoremas 1.5 e 2.5, podemos considerar o resultado seguinte. Se n e r são inteiros positivos, então escrevemos:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Teorema 5.5. Sejam r e n inteiros positivos. O número de r -combinações de n objectos distintos é dado por:

$$\binom{n}{r} = \frac{P(n, r)}{P(r, r)} = \frac{P(n, r)}{r!} = \frac{n!}{r!(n-r)!}, r < n \quad e \quad \binom{n}{n} = 1$$

Definimos $\binom{n}{0} = 1$ para todo o inteiro n não negativo.

Nota: É imediato que:

$$\binom{n}{r} = C(n, n - r)$$

o que é de se esperar uma vez que seleccionar r objectos entre n objectos distintos é equivalente a “tirar” os $n - r$ objetos que não serão seleccionados.

Exemplo 5.9 (Combinações)

Dentre 10 homens e 8 mulheres, quantas comissões de 5 pessoas podem ser formadas, sendo que em cada uma deve haver 3 homens e 2 mulheres?

Solução

Seja A o conjunto dos subconjuntos de 3 homens e B o conjunto dos subconjuntos de 2 mulheres, então:

$$\binom{10}{3} * \binom{8}{2} = 120 * 28 = 3360$$

5.4.2. Triângulo de Pascal

O triângulo de pascal é uma forma de organizar os resultados de $\binom{n}{r}$ para diferentes valores de n e r , conforme indica figura abaixo.

$$\begin{array}{cccccc}
 \binom{0}{0} & \rightarrow & \text{linha zero} & & & & \\
 \binom{1}{0} & & \binom{1}{1} & \rightarrow & 1^{\text{a}} \text{ linha} & & \\
 \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & \rightarrow & 2^{\text{a}} \text{ linha} \\
 \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & \rightarrow & 3^{\text{a}} \text{ linha} \\
 \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} & \rightarrow & 4^{\text{a}} \text{ linha} \\
 \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \\
 \binom{n}{0} & & \binom{n}{1} & & \binom{n}{2} & & \binom{n}{3} & & \binom{n}{4} & & \binom{n}{n} & \rightarrow & n - \text{ésima linha}
 \end{array}$$

5.5. EXERCÍCIOS PROPOSTOS

5.5.1. Três classes diferentes contem 20, 18 e 25 estudantes e nenhum estudante é membro de mais de uma das classes. Se uma equipe deve ser composta por um

estudante de cada classe, de quantos modos diferentes os membros desta equipe podem ser escolhidos?

- 5.5.2.** Em um campeonato de futebol participam 20 times. Quantos resultados são possíveis para os 3 primeiros lugares?
- 5.5.3.** Um cofre possui um disco marcado com os dígitos 0,1,2,3,4,5,6,7,8 e 9. O segredo do cofre é formado por uma sequência de 3 dígitos distintos. Se uma pessoa tentar abrir o cofre, quantas tentativas devera fazer (no máximo) para conseguir abri-lo? Suponha que a pessoa sabe a quantidade de dígitos do segredo e que este é formado por dígitos distintos.
- 5.5.4.** De quantas formas 6 pessoas podem sentar-se numa fileira de 6 cadeiras se duas delas, Joaquim e Rafael, se recusam a sentar um ao lado do outro?
- 5.5.5.** Considere 10 cadeiras numeradas de 1 a 10. De quantas maneiras 2 pessoas podem sentar-se, devendo haver ao menos uma cadeira entre eles?
- 5.5.6.** Quantos anagramas da palavra “estudo” começam e terminam com vogal?
- 5.5.7.** Considere 2 urnas. A primeira com 4 cartas numeradas de 1 a 4 e a segunda com 3 cartas numeradas de 7 a 9. Duas cartas são extraídas da primeira urna, sucessivamente e sem reposição, e em seguida duas cartas são extraídas da segunda urna, sucessivamente e sem reposição. Quantos números de 4 algarismos podem ser formados com os números das cartas na sequência em que foram obtidas?
- 5.5.8.** Com os algarismos 1, 2, 3, 4, 5, 6, 7, 8 e 9, quantos números de 4 algarismos com pelo menos dois algarismos iguais existem?
- 5.5.9.** De quantas formas 5 meninos e 5 meninas podem ficar em fila, de modo que meninos e meninas devem ficar em posições alternadas?
- 5.5.10.** Dez pessoas, dentre elas António e Beatriz, devem ficar em fila. De quantas formas isto pode ser feito de modo que António e Beatriz fiquem sempre juntos?
- 5.5.11.** De quantas formas 4 homens e 5 mulheres podem ficar em fila se

- a) Os homens devem ficar juntos?
- b) E se os homens devem ficar juntos e as mulheres também?

5.5.12. Considere 15 livros em uma estante, dos quais 4 são de probabilidade. De quantas formas podemos coloca-lo em uma prateleira da estante de modo que os livros de probabilidade fiquem sempre juntos?

5.5.13. Quantos anagramas existem da palavra "AMARILIS"?

5.5.14. Uma urna contém 3 bolas vermelhas e 2 amarelas, que se distinguem apenas pela cor. Quantas sequências de cores são possíveis de observar extraindo uma a uma sem reposição?

5.5.15. Quantos números de 7 algarismos existem nos quais aparecem uma só vez os algarismos 3, 4 e 5 e quatro vezes o algarismo 9?

5.5.16. Uma moeda é lançada 20 vezes. Quantas sequências de caras e coroas existem com 10 caras e 10 coroas?

5.5.17. Quantos produtos podemos obter se tomarmos 3 factores distintos escolhidos entre 2,3,5,7 e 11?

5.5.18. Um time de futebol de salão deve ser escalado a partir de um conjunto de 10 jogadores, entre eles Joaquim e Caio. Quantos times de 5 jogadores podem ser formados se Ari e Arnaldo devem ser escalados necessariamente?

5.5.19. Considere 10 homens e 10 mulheres. Quantas comissões de 5 pessoas podemos formar se em cada uma deve haver 3 homens e 2 mulheres?

5.5.20. Uma urna contém 10 bolas brancas e 6 pretas, todas marcadas com símbolos distintos. Quantos conjuntos de 7 bolas (retiradas desta urna) podemos formar de modo que pelo menos 4 bolas do conjunto sejam pretas?

5.5.21. Em uma reunião, cada pessoa cumprimentou todas as outras, havendo ao todo 45 apertos de mão. Quantas pessoas haviam na reunião?

- 5.5.22.** Um químico possui 10 tipos diferentes de substâncias. De quantos modos possíveis poderá associar 6 diferentes tipos destas substâncias, sendo que dois tipos (somente) não podem ser juntados pois produzem mistura explosiva?
- 5.5.23.** Quantas diagonais têm um polígono regular de n lados?
- 5.5.24.** Obter o número de maneiras que nove algarismos iguais a 0 e seis algarismos iguais a 1 podem ser colocados em sequência de modo que dois uns não compareçam juntos.
- 5.5.25.** Quantos subconjuntos de 5 cartas contendo exactamente 3 ases podem ser formados de um baralho de 52 cartas?
- 5.5.26.** A directoria de uma firma é composta por 7 directores brasileiros e 4 japoneses. Quantas comissões podem ser formadas com 3 directores brasileiros e 3 japoneses?
- 5.5.27.** Um homem possui 8 pares de meias distintos. De quantas formas ele pode seleccionar escolher dois pés de meia (um direito e um esquerdo) de modo que eles sejam de pares diferentes?
- 5.5.28.** Com os algarismos 1, 2, 3, 4, 5, 6, 7, 8 e 9, quantos números de algarismos distintos existem entre 500 e 1000?
- 5.5.29.** Com os algarismos 1, 2, 3, 4, 5 e 6, quantos números pares de 3 algarismos distintos podemos formar?
- 5.5.30.** Quantos números pares de 3 algarismos podemos formar com os algarismos 1, 3, 6, 7, 8 e 9?
- 5.5.31.** Suponhamos que todos os números obtidos a partir da permutação dos algarismos 1, 2, 4, 6 e 8 foram dispostos em ordem crescente. Qual posição ocupa o número 68412?

6. PROBABILIDADES

As origens históricas da teoria das probabilidades estão vinculadas á teoria dos jogos e aos nomes de Fermat e Pascal, que na metade do século XVII formalizaram pela primeira vez o conceito de probabilidade. Falamos aqui de história escrita (mesmo que isto seja uma redundância), já que existem indícios de que o trabalho de Fermat e Pascal consolidou ideias que foram desenvolvidas a partir do século XII.

No decorrer do tempo a teoria das probabilidades foi superando o marco original da teoria dos jogos para constituir na actualidade um ramo da matemática pura com aplicações nas ciências de um modo geral. Faremos aqui uma breve revisão dos aspectos fundamentais da evolução do conceito de probabilidade.

6.1. Experimento aleatório

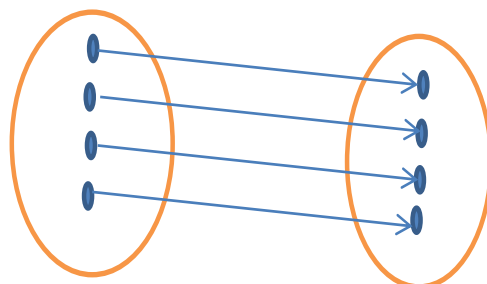
É um experimento que pode fornecer diferentes resultados, se repetido essencialmente sob as mesmas condições.

6.2. Espaço amostral

Suponhamos um experimento realizado sob certas condições fixas. O espaço amostral Ω do experimento é um conjunto que contém representações de todos os resultados possíveis, onde por “resultado possível”, entende-se resultado elementar e indivisível do experimento. Ω deve satisfazer as seguintes condições:

- A todo resultado possível corresponde um, e somente um, elemento $\omega \in \Omega$.
- Resultados distintos correspondem a elementos distintos em Ω , ou seja, $\omega \in \Omega$ não pode representar mais de um resultado.

Resultados possíveis Espaço amostral



Exemplo 6.1 (Espaço amostral)

Considere um experimento que consiste em arremessar uma moeda duas vezes e observar as faces obtidas voltadas para cima. Defina um espaço amostral par este experimento

Solução

Não é difícil encontrar quem defina $\Omega = \{cara; coroa\}$ como espaço amostral deste experimento. No entanto, esta definição esta incorrecta, pois no experimento é arremessada a moeda duas vezes e não uma. Lembre-se que o espaço amostral deve conter representações de todos os resultados possíveis do experimento. Um espaço amostral para este experimento é:

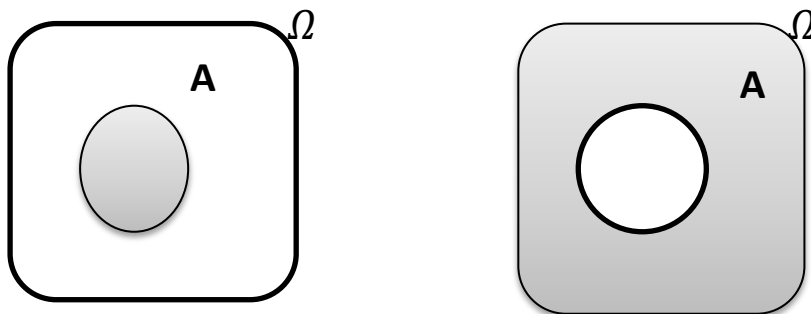
$$\Omega = \{(cara, cara); (cara, coroa); (coroa, cara); (coroa, coroa)\}$$

6.3. Eventos

Seja Ω o espaço amostral do experimento. Todo subconjunto $A \subset \Omega$ será chamado evento.

- Ω é o evento certo.
- \emptyset é o evento impossível.
- Para $\omega \in \Omega$, o evento ω é dito elementar (ou simples).
- Eventos com uma distribuição de probabilidade são chamados de eventos aleatórios.

O complementar de um evento A , denotado por A^c , é o conjunto formado pelos elementos de Ω que não pertencem á A . Assim, $A^c = \{\omega \in \Omega : \omega \notin A\}$



Exemplo 6.2 (Evento)

Joga-se uma moeda duas vezes. Indique o evento aparecer cara apenas uma vez.

Solução

Seja o evento $A = \text{aparecimento de cara uma vez}$, então:

$$A = \{(cara; coroa); (coroa; cara)\}$$

6.4. Definições de probabilidade

Há várias interpretações da probabilidade. A seguir, veremos as quatro mais importantes.

6.4.1. Probabilidade clássica

Se Ω é finito, a definição clássica da probabilidade $P(A)$ de um evento $A \subset \Omega$ é dada por:

$$P(A) = \frac{\#A}{\#\Omega} = \frac{\text{número de elementos de } A}{\text{número de elementos de } \Omega}$$

Observação: Esta definição basea-se no conceito de resultados equiparáveis, ou melhor, no princípio da indiferença. Por exemplo, em um experimento que consiste em lançar um dado e observar o resultado, podemos usar $\Omega = \{1, 2, \dots, 6\}$ e, diante da indiferença entre os resultados, temos $P(i) = \frac{1}{6}, \forall i \in \Omega$

Exemplo 6.3 (Definição clássica)

Considere um experimento que consiste em arremessar uma moeda duas vezes e observar as faces obtidas voltadas para cima. Qual é a probabilidade de se obter cara apenas uma vez?

Solução

Seja

$$\Omega = \{(ca, ca); (ca, co); (co, ca); (co, co)\} \text{ e}$$

$$A = \{(ca, co); (co, ca)\}$$

$$P(A) = \frac{\#A}{\#\Omega} = \frac{2}{4}$$

6.4.2. Probabilidade frequentista

A definição frequentista baseia-se na frequência relativa de um número grande de realizações do experimento. Mais especificamente, definimos a probabilidade $P(A)$ de um evento A usando o limite da frequência relativa da ocorrência de A em n repetições independentes do experimento, com n tendendo ao infinito, ou seja,

$$P(A) = \lim_{n \rightarrow \infty} \frac{1}{n} * (n^{\circ} \text{ de ocorrências de } A \text{ em } n \text{ repetições independentes})$$

A grande dificuldade da definição frequentista é que os experimentos nunca são realizados infinitas vezes, logo não há como avaliar a probabilidade de forma estrita.

6.4.3. Probabilidade Geométrica

Consideremos um experimento que consiste em escolher um ponto ao acaso em uma região $\Omega \subset R^p$. A definição geométrica da probabilidade $P(A)$ de um evento $A \subset \Omega$ é dada por:

$$P(A) = \frac{\text{volume de } A}{\text{volume de } \Omega}$$

Nota: Naturalmente, em espaços unidimensionais ($p = 1$) o volume é substituído por comprimento e em espaços bidimensionais ($p = 2$), por área.

6.4.4. Probabilidade subjectiva

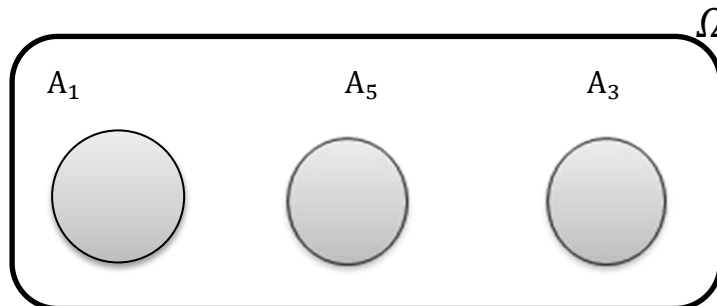
A definição subjectiva de probabilidade baseia-se em crenças e/ou informações do observador a respeito do fenómeno em estudo.

Exemplo 6.4 (Definição subjectiva)

Consideremos o evento $A =$ "chove em Nampula". Para alguém em Maputo podemos ter a seguinte avaliação: $P(A) = 0,5$. Para alguém de Inhambane, podemos ter $P(A) = 0.8$ se chove em Inhambane e $P(A) = 0.3$ se não chove em Inhambane. Para alguém de Nampula, $P(A) = 1$ se está chovendo em Nampula e $P(A) = 0$ se não está chovendo em Nampula.

6.4.5. Teoria dos conjuntos: revisão de conceitos

Os conjuntos da sequência (finita ou enumerável) A_1, A_2, \dots são disjuntos 2 a 2, se $A_i \cap A_j = \emptyset, \forall i \neq j$



6.4.6. Axiomas

Não vamos nos preocupar, doravante, com o problema de como definir probabilidade para cada experimento. Simplesmente, vamos admitir que as probabilidades estão definidas em um certo conjunto γ de eventos, chamados eventos aleatórios. Vamos supor que a todo $A \in \gamma$ seja associado um número real $P(A)$, chamado de probabilidade de A , de modo que os **axiomas** a seguir sejam satisfeitos.

- Axioma 1: $P(A) \geq 0, \forall A \in \gamma$;
- Axioma 2: $P(\Omega) = 1$
- Se $A_1, A_2, \dots \in \gamma$ são disjuntos 2 a 2, então:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

Um espaço de probabilidade é um trio Ω, γ, P , onde

- Ω é um conjunto não vazio,
- γ é um conjunto de eventos aleatórios e
- P é uma probabilidade em γ

6.5. Principais Teoremas

6.5.1. Probabilidade do vazio

Teorema 6.1: $P(\emptyset) = 0$

Demostração: Temos que:

$$\begin{aligned} P(\Omega) &= P(\Omega \cup \emptyset \cup \emptyset \cup \emptyset \cup \dots) \rightarrow \\ P(\Omega) &+ P(\emptyset) + P(\emptyset) + \dots \rightarrow \\ 0 &= P(\emptyset) + P(\emptyset) + \dots \rightarrow \\ P(\emptyset) &= 0 \end{aligned}$$

6.5.2. Probabilidade da união finita de eventos disjuntos 2 a 2

Teorema 6.2: Se $A_1, A_2, \dots, A_n \in \gamma$ são eventos aleatórios disjuntos 2 a 2 então:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

Demostração: Fazendo $A_i = \emptyset, \forall i \in \{n+1, n+2, \dots\}$, temos que:

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= \sum_{i=1}^{\infty} P(A_i) = (\text{pelo axioma 3}) = \\ &= \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^n P(A_i) + \sum_{i=n+1}^{\infty} P(A_i) = \\ &= \sum_{i=1}^n P(A_i) + \sum_{i=n+1}^{\infty} P(\emptyset) = (\text{pelo teorema 1}) = \\ &= \sum_{i=1}^n P(A_i) + \sum_{i=n+1}^{\infty} 0 = \sum_{i=1}^n P(A_i) \end{aligned}$$

6.5.3. Probabilidade do evento complementar

Teorema 6.3: $P(A^c) = 1 - P(A), \forall A \in \gamma$

Demostração: Temos que:

$$\begin{aligned} \Omega &= A \cup A^c \rightarrow \\ P(\Omega) &= P(A \cup A^c) \rightarrow (\text{aplicando os axiomas 2 e 3}) \rightarrow \\ 1 &= P(A) + P(A^c) \rightarrow \\ P(A^c) &= 1 - P(A) \end{aligned}$$

6.5.4. Probabilidade de eventos aninhados

Teorema 6.4: $\forall A, B \in \gamma; A \subset B \rightarrow P(A) \leq P(B)$

Demostração: Pelo axioma 1, temos que $P(B \cap A^c) \geq 0$. Assim:

$$\begin{aligned}
P(B \cap A^c) &\geq 0 \rightarrow \\
P(B \cap A^c) + P(A) &\geq P(A) \rightarrow (\text{pelo axioma 3}) \rightarrow \\
P((B \cap A^c) \cup P(A)) &\geq P(A) \rightarrow \\
P(A) &\leq P(B)
\end{aligned}$$

6.5.5. Probabilidade entre 0 e 1

Teorema 6.5: $0 \leq P(A) \leq 1, \forall A \in \gamma$

Demonstração: Como $A \subset \Omega$, aplicando o teorema 6.4 temos que:

$$P(A) \leq P(\Omega) \rightarrow (\text{pelo axioma 2}) \rightarrow P(A) \leq 1$$

Além disso, pelo axioma 1, $P(A) \geq 0$. Logo $0 \leq P(A) \leq 1$

6.5.6. Probabilidade da subtração

Teorema 6.6: $\forall A, B \in \gamma, P(A \cap B^c) = P(A) - P(A \cap B)$

Demonstração: Temos que:

$$\begin{aligned}
(A \cap B^c) \cup (A \cap B) &= A \rightarrow \\
P((A \cap B^c) \cup (A \cap B)) &= P(A) \rightarrow \\
P(A \cap B^c) + P(A \cap B) &= P(A) \rightarrow \\
P(A \cap B^c) &= P(A) - P(A \cap B)
\end{aligned}$$

Nota: Em alguns livros o evento $A \cap B^c$ é definido como $A - B$. Aqui não adoptamos esta notação para evitar confusões entre a subtração de probabilidades e a subtração de conjuntos.

6.5.7. Desigualdade de Boole

Teorema 6.7: Supondo que A_1, A_2, A_3, \dots são eventos aleatórios,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$$

Demonstração: Consideremos a seguinte sequência de eventos:

$$B_1 = A_1$$

$$\begin{aligned}
B_2 &= A_2 \cap A_1^c \\
B_3 &= A_3 \cap (A_1 \cap A_2)^c \\
&\dots \\
B_3 &= A_3 \cap (A_1 \cap A_2)^c \\
B_i &= A_i \cap (A_1 \cup A_2 \cup \dots \cup A_{i-1})^c
\end{aligned}$$

Note que esta sequência é de eventos disjuntos 2 a 2. Além disso, temos que $B_i \subset A_i$, o que implica $P(B_i) \subset P(A_i)$. Deste modo, temos que:

$$\begin{aligned}
P\left(\bigcup_{i=1}^{\infty} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} B_i\right) \rightarrow \text{pelo axioma 3} \\
\sum_{i=1}^{\infty} P(B_i) &\leq \sum_{i=1}^{\infty} P(A_i)
\end{aligned}$$

Teorema 6.8: Supondo que $A_1, A_2, A_3, \dots, A_n$ são eventos aleatórios, temos que:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$$

Demonstração: Análoga á demonstração do resultado anterior

6.5.8. Probabilidade da união de 2 eventos

Teorema 6.9: Se A e B forem eventos quaisquer, então:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Demonstração:

$$P(A \cup B) = P((A \cap B^c) \cup B)$$

repare que $A \cap B^c$ e B são disjuntos \rightarrow

$$P(A \cap B^c) + P(B) = P(A) + P(B) - P(A \cap B)$$

6.5.9. Probabilidade da união de 3 eventos

Teorema 6.10: Se A, B e C forem eventos quaisquer, então:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Demonstração:

$P(A \cup B \cup C) =$ (pelo teorema 6.9) temos:

$P(A \cup B) + P(C) - P((A \cup B) \cap C)$ (pelo teorema 6.9) temos:

$$P(A) + P(B) + P(C) - P(A \cap B) - P((A \cup B) \cap C)$$

$$P(A) + P(B) + P(C) - P(A \cap B) - P((A \cap C) \cup (B \cap C)) \rightarrow$$

$$= \text{(pelo teorema 6.9)} =$$

$$P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

6.5.10. Probabilidade da união finita

Teorema 6.11: Supondo uma sequência $A_1, A_2, A_3, \dots, A_n$ de eventos aleatórios, temos que:

$$\begin{aligned} P(A_1 \cup A_2 \cup \dots \cup A_n) &= \sum_{i=1}^n P(A_i) - \sum_{i < j=2}^n P(A_i \cap A_j) + \sum_{i < j < r=2}^n P(A_i \cap A_j \cap A_r) \\ &\quad + (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n) \end{aligned}$$

Demonstração: Por indução finita

Nota: Os dois últimos teoremas são casos particulares deste teorema

6.6. Probabilidade condicional e principais teoremas

6.6.1. Probabilidade Condicional

Seja (Ω, γ, P) um espaço de probabilidade. Se $B \in \gamma$ e $P(B) > 0$, a probabilidade condicional de $A \in \gamma$ dado B é definida por:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Nota:

- Se $P(B) = 0$, $P(A|B)$ pode ser arbitrariamente definida. Mas por independência, é conveniente fazer $P(A|B) = P(A)$ como veremos adiante.

- Decorre da definição que $P(A \cap B) = P(B) * P(A|B)$ e esta igualdade é válida quando $P(B) = 0$

Exemplo 6.5 (Probabilidade Condicional)

Suponhamos que uma fábrica possui 310 máquinas de soldar. Algumas destas máquinas são eléctricas (E), enquanto outras são manuais (M). Por outro lado, temos também que algumas são novas (N) e outras são usadas (U). A tabela abaixo informa o número de máquinas de cada categoria.

	Eléctricas (E)	Manuais (M)
Novas (N)	10	60
Usadas (U)	200	40

- a) Sabendo que uma determinada peça foi soldada usando uma máquina nova, qual é a probabilidade de ter sido soldada por uma máquina eléctrica?

Solução

$$P(E|N) = \frac{P(E \cap N)}{P(N)} = \frac{10}{70} = 0.1429$$

- b) Sabendo que uma determinada peça foi soldada usando uma máquina eléctrica, qual é a probabilidade de ter sido soldada por uma máquina nova?

Solução

$$P(N|E) = \frac{P(N \cap E)}{P(E)} = \frac{10}{210} = 0.0476$$

Teorema 6.12: Uma probabilidade condicional dado um evento B qualquer é uma probabilidade.

Demonstração:

Para mostrar que a probabilidade condicional é uma probabilidade, devemos verificar que:

- $P(A|B) \geq 0, \forall A \in \gamma$
- $P(\Omega|B) = 1$ e que
- Se $A_1, A_2, A_3, \dots \in \gamma$ são disjuntos 2 a 2, então:

$$P\left(\bigcup_{i=1}^{\infty} A_i \setminus B\right) = \sum_{i=1}^{\infty} P(A_i \setminus B)$$

Vamos verificar então as condições acima.

- $P(A \setminus B) = \frac{P(A \cap B)}{P(B)}$, com $P(A \cap B) \geq 0$ e $P(B) > 0$ temos que $P(A \setminus B) \geq 0$ e a 1ª condição foi satisfeita.
- Temos também que $P(\Omega \setminus B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$ e a 2ª condição foi satisfeita.
- Por fim temos:

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i \setminus B\right) &= \frac{P\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right)}{P(B)} = \frac{P\left(\bigcup_{i=1}^{\infty} A_i \cap B\right)}{P(B)} = (\text{pelo axioma 3}) \\ &= \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)} = \sum_{i=1}^{\infty} \frac{P(A_i \cap B)}{P(B)} = \sum_{i=1}^{\infty} P(A_i \setminus B) \end{aligned}$$

6.6.2. Teorema de multiplicação

Teorema 6.13: Seja (Ω, γ, P) um espaço de probabilidade com $A_1, A_2, A_3, \dots \in \gamma$ então:

$$\begin{aligned} P(A_1 \cap A_2 \cap \dots \cap A_n) &= \\ &= P(A_n \setminus A_1 \cap \dots \cap A_{n-1}) * P(A_{n-1} \setminus A_1 \cap \dots \cap A_{n-2}) * \dots * P(A_2 \setminus A_1) * P(A_1) \end{aligned}$$

Demonstração: Por indução finita

Nota: Especificamente, para $n=2$, temos:

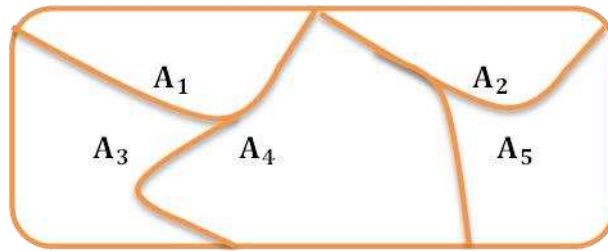
$$P(A_1 \cap A_2) = P(A_2 \setminus A_1) * P(A_1) = P(A_1 \setminus A_2) * P(A_2)$$

6.6.3. Partição de um conjunto

Uma seqüência A_1, A_2, A_3, \dots finita ou enumerável de conjuntos é uma partição de um conjunto A quando:

- For uma seqüência de conjuntos disjuntos 2 a 2 e

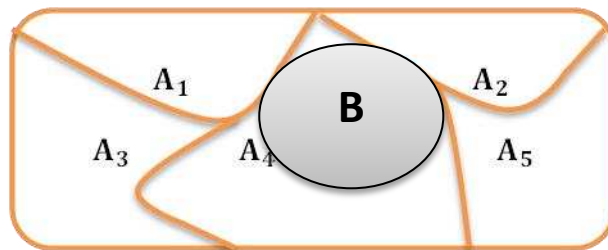
$$\bigcup_i A_i = A$$



6.6.4. Teorema de probabilidade total

Teorema 6.14: Seja (Ω, γ, P) um espaço de probabilidade. Se a sequência (finita ou enumerável) $A_1, A_2, A_3, \dots \in \gamma$ formar uma partição de Ω , então:

$$P(B) = \sum_i P(B \setminus A_i) * P(A_i)$$



Demonstração:

$$\begin{aligned} P(B) &= \bigcup_i (B \cap A_i) = (\text{pelo axioma 3}) \\ &= \sum_i P(B \cap A_i) = \sum_i P(B \setminus A_i) * P(A_i) \end{aligned}$$

Exemplo 6.6 (Teorema de Probabilidade Total)

Uma empresa produz circuitos em três fábricas, denotadas por I, II e III. A fábrica I produz 40% dos circuitos, enquanto a II e a III produzem 30% cada uma. As probabilidades de que um circuito produzido por essas fábricas não funcione são 0.01, 0.04 e 0.03 respectivamente. Escolhido ao acaso um circuito da produção conjunta das três fábricas, qual é a probabilidade do circuito não funcionar?

Solução

Considere os eventos:

- I = o circuito foi produzido pela fábrica I

- $II = \text{o circuito foi produzido pela fábrica II}$
- $III = \text{o circuito foi produzido pela fábrica III}$
 - $B = \text{o circuito não funciona}$

Primeiro repare que os conjuntos I , II e III formam uma partição do espaço amostral.

Assim, aplicando o teorema da probabilidade total, temos que:

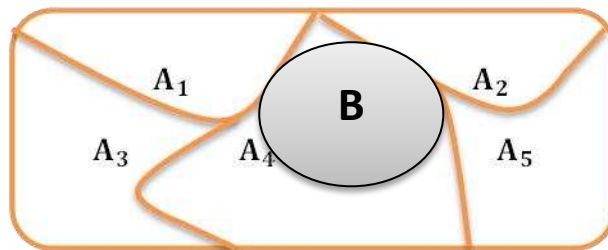
$$P(B) = P(B \setminus I) * P(I) + P(B \setminus II) * P(II) + P(B \setminus III) * P(III)$$

$$P(B) = 0.01 * 0.4 + 0.04 * 0.3 + 0.03 * 0.3 = 0.025$$

6.6.5. Teorema de Bayes

Teorema 6.14: Seja (Ω, γ, P) um espaço de probabilidade. Se a sequência (finita ou enumerável) $A_1, A_2, A_3, \dots \in \gamma$ formar uma partição de Ω , então:

$$P(A_i \setminus B) = \frac{P(B \setminus A_i) * P(A_i)}{\sum_j P(B \setminus A_j) * P(A_j)}$$



Demonstração:

$$P(A_i \setminus B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B \setminus A_i) * P(A_i)}{P(B)} = (\text{pelo teorema de probabilidade total})$$

Exemplo 6.7 (Teorema de Bayes)

Uma empresa produz circuitos em três fábricas, denotadas por I , II e III . A fábrica I produz 40% dos circuitos, enquanto a II e a III produzem 30% cada uma. As probabilidades de que um circuito produzido por essas fábricas não funcione são 0.01, 0.04 e 0.03 respectivamente. Escolhido ao acaso um circuito da produção conjunta das três fábricas, qual é a probabilidade do circuito ter sido produzido pela fábrica I ?

Solução

Considere os eventos:

- $I = \text{o circuito foi produzido pela fábrica I}$
- $II = \text{o circuito foi produzido pela fábrica II}$
- $III = \text{o circuito foi produzido pela fábrica III}$

- $B = \text{o circuito não funciona}$

Primeiro repare que os conjuntos I, II e III formam uma partição do espaço amostral.

Assim, aplicando o teorema da probabilidade total, temos que:

$$P(A_1|B) = \frac{P(B|I) * P(I)}{P(B|I) * P(I) + P(B|II) * P(II) + P(B|III) * P(III)}$$

$$P(A_1|B) = \frac{0.01 * 0.4}{0.01 * 0.4 + 0.04 * 0.3 + 0.03 * 0.3} = 0.16$$

6.6.6. Sensibilidade, Especificidade, Valor Preditivo Positivo e Negativo

Em qualquer dado teste administrado a uma dada população é importante calcular a sensibilidade, a especificidade, o valor preditivo positivo e o valor preditivo negativo, para determinar o quão útil o teste é para detectar doenças ou características de uma certa população. Se deseja usar um teste para sobre características específicas numa amostra da população, é importante saber:

- O quão provável é que o teste detecte a presença de uma característica em alguém com a característica (**sensibilidade**)?
- O quão provável é que o teste detecte a ausência de uma característica em alguém sem a característica (**especificidade**)?
- O quão provável é que alguém com um resultado positivo no teste verdadeiramente tenha a característica (**valor preditivo positivo**)?
- O quão provável é que alguém com um resultado negativo no teste verdadeiramente não tenha a característica (**valor preditivo negativo**)?

Suponhamos um exame médico qualquer com dois resultados possíveis. Vamos assumir que $X = 1$ quando o exame acusa a doença e que $X = 0$ caso contrário. Por outro lado, vamos usar θ para indicar o verdadeiro estado do indivíduo submetido ao exame, onde $\theta = 1$ indica um indivíduo doente e $\theta = 0$ indica um indivíduo saudável, então:

$$\text{Sensibilidade} = P(X = 1 | \theta = 1)$$

$$\text{Especificidade} = P(X = 0 | \theta = 0)$$

$$\text{Valor preditivo positivo} = P(\theta = 1 | X = 1)$$

$$\text{Valor preditivo negativo} = P(\theta = 0 | X = 0)$$

Exemplo 6.8 (Sensibilidade, Especificidade, Valor Preditivo Positivo e Negativo)

Recomenda-se que, a partir dos 40 anos, as mulheres façam mamografias anuais. Nesta idade, aproximadamente uma em cada 100 mulheres são portadoras de um tumor assintomático de mama.

Seja θ uma quantidade desconhecida que indica se uma paciente desta faixa etária tem a doença ou não. Se ela possui a doença então $\theta = 1$, caso contrário $\theta = 0$. Assim, podemos assumir que $P(\theta = 1) = 0.01$ e $P(\theta = 0) = 0.99$

Sabe-se que a mamografia indica a doença em 80% das mulheres com câncer de mama, mas esse mesmo resultado ocorre também com 9,6% das mulheres sem o câncer. Assim, seja X uma variável aleatória associada ao resultado da mamografia, de modo que se $X = 1$ o exame acusou a doença e $X = 0$ caso contrário. Temos então que:

$P(X = 0 \mid \theta = 0)$ (especificidade: exame não acusar sem a doença)

$P(X = 1 \mid \theta = 1)$ (sensibilidade: exame acusar com a doença)

Imagine agora que você encontra uma amiga de 40 e poucos anos aos prantos, desesperada, porque fez uma mamografia de rotina e o exame acusou a doença. Qual a probabilidade de ela ter um câncer de mama?

Solução

$$P(\theta = 1 \mid X = 1) = \frac{P(X = 1 \mid \theta = 1) * P(\theta = 1)}{P(X = 1 \mid \theta = 1) * P(\theta = 1) + P(X = 1 \mid \theta = 0) * P(\theta = 0)}$$

$$P(\theta = 1 \mid X = 1) = \frac{0.80 * 0.01}{0.80 * 0.01 + 0.96 * 0.99} = 0.0776$$

Logo, a probabilidade dela ter a doença é de aproximadamente 7,8%.

6.7. Independência

6.7.1. Independência entre dois eventos

Dois eventos aleatórios A e B são independentes quando:

- $P(A \cap B) = P(A) * P(B)$
- $P(A \setminus B) = P(A)$
- $P(B \setminus A) = P(B)$

6.7.2. Independência de eventos dois a dois

Seja $\{A_i: i \in I\}$ uma colecção de eventos aleatórios indexada por um conjunto (de índices) I . Os eventos desta colecção são ditos independentes 2 a 2 se:

$$P(A_i \cap A_j) = P(A_i) * P(A_j) \forall i, j \in I \text{ tais que } i \neq j$$

6.7.3. Independência mútua

Seja $B = \{A_i: i \in I\}$ uma colecção de eventos aleatória indexada por um conjunto (de índices) I . Os eventos desta colecção são (mutuamente) independentes se, para toda subfamília finita $\{A_{i_1}, A_{i_2}, \dots, A_{i_n}\}$ de eventos em B , tivermos:

$$P(A_{i_1}, A_{i_2}, \dots, A_{i_n}) = P(A_{i_1}) * P(A_{i_2}) \dots P(A_{i_n})$$

Em particular, para todas as subfamílias (A_i, A_j) com $i \neq j$, temos que:

$$P(A_i \cap A_j) = P(A_i) * P(A_j)$$

Logo, B é uma colecção de eventos independentes 2 a 2.

6.8. EXERCÍCIOS PROPOSTOS

6.8.1. Sejam A e B dois eventos. Se $P(A) = 0.3$, $P(B) = 0.2$ e $P(A \cap B) = 0.1$, então calcule:

- a) $P(\bar{A})$
- b) $P(A \cup B)$
- c) $P(\bar{A} \cap B)$
- d) $P(A \cap \bar{B})$
- e) $P(\overline{A \cup B})$
- f) $P(\bar{A} \cup B)$
- g) $P(\bar{A} \cap \bar{B})$
- h) $P(\bar{A} \cup \bar{B})$

6.8.2. Uma empresa efectuou um estudo sobre fraudes em cartões de crédito, com resultados resumidos no quadro abaixo.

Tipo de fraude	Quantidade
Cartão roubado	243
Cartão falsificado	85
Pedido por telefone	52
Outros	46

Seleccionando aleatoriamente um caso de fraude nos casos resumidos acima, qual é a probabilidade (clássica) da fraude resultar de um cartão falsificado?

6.8.3. Num restaurante registaram-se, durante bastante tempo, os pedidos dos clientes, tendo-se chegado à conclusão que, para terminar a refeição, 20% dos clientes pedem só sobremesa, 40% pedem só café e 30% pedem sobremesa e café.

- Construa um diagrama de Venn para ilustrar a situação anterior.
- Determine a probabilidade do acontecimento “pedir café”.
- Determine a probabilidade do acontecimento “não pedir sobremesa”.
- Determine a probabilidade do acontecimento “nem pede café nem sobremesa”.
- Determine a probabilidade do acontecimento “pedir café ou sobremesa”.
- Os acontecimentos “pedir café” e “pedir sobremesa” são independentes?

6.8.4. Suponhamos que uma fábrica possui 310 máquinas de soldar. Algumas destas máquinas são eléctricas (E), enquanto outras são manuais (M). Por outro lado, temos também que algumas são novas (N) e outras são usadas (U). A tabela abaixo informa o número de máquinas de cada categoria.

	Eléctricas	Manuais
Novas	10	60
Usadas	200	40

- Sabendo que uma determinada peça foi soldada usando uma máquina nova, qual é a probabilidade de ter sido soldada por uma máquina eléctrica?
- Sabendo que uma determinada peça foi soldada usando uma máquina eléctrica, qual é a probabilidade de ter sido soldada por uma máquina nova?

6.8.5. Uma empresa produz circuitos em três fábricas, denotadas por I, II e III. A fábrica I produz 40% dos circuitos, enquanto a II e a III produzem 30% cada uma. As probabilidades de que um circuito produzido por essas fábricas não funcione são 0.01, 0.04 e 0.03 respectivamente. Escolhido ao acaso um circuito da produção conjunta das três fábricas,

- Qual é a probabilidade do circuito não funcionar?

- b) Dado que o circuito escolhido não funciona, qual é a probabilidade do circuito ter sido produzido pela fábrica I?

6.8.6. Um empreiteiro apresentou orçamentos separados para a execução da parte eléctrica e da parte de encanamento de um edifício. Ele acha que a probabilidade de ganhar a concorrência da parte eléctrica é de 50%. Caso ele ganhe a parte eléctrica, a chance de ganhar a parte de encanamento é de $3/4$, caso contrário esta probabilidade é de $1/3$.

- a) Qual a probabilidade do empreiteiro ganhar os dois contractos?
 b) Qual a probabilidade do empreiteiro ganhar apenas um?
 c) Qual a probabilidade do empreiteiro perder a parte eléctrica e perder a parte de encanamento?

6.8.7. Óleos de cozinha são produzidos em duas principais variedades: monoinsaturados e polinsaturados. Duas matérias-primas para óleos de cozinha são: milho e canola. A tabela a seguir mostra o número de garrafas destes óleos em um supermercado.

		Tipo de óleo	
		Canola	Milho
Tipo de insaturação	Mono	7	13
	Poly	93	77

- a) Se uma garrafa de óleo é seleccionada aleatoriamente, qual a probabilidade de ser um óleo polinsaturado?
 b) Se uma garrafa de óleo é seleccionada aleatoriamente, qual a probabilidade de ser monoinsaturado de canola?

6.8.8. Um certo tipo de motor eléctrico falha se ocorrer uma das seguintes situações: emperramento dos mancais, queima dos rolamentos ou desgaste das escovas. Suponha que o emperramento seja duas vezes mais provável do que a queima, esta sendo quatro vezes mais provável do que o desgaste das escovas. Qual é a probabilidade de que a falha seja devida a cada uma dessas circunstâncias?

6.8.9. Suponha que A , B , C sejam eventos tais que $P(A) = P(B) = P(C) = \frac{1}{4}$; $P(A \cap B) = P(B \cap C) = 0$ e $P(A \cap C) = 1/8$. Calcule a probabilidade de que ao menos um dos eventos (A , B ou C) ocorra.

6.8.10. Um equipamento electrónico é formado por 2 componentes, I e II. Suponha que:

- A chance do componente I falhar é 0,20;
- A chance de apenas o componente II falhar é 0,15 e
- A chance de I e II falharem simultaneamente é 0,15.

a) Calcule a probabilidade de apenas o componente I falhar.

b) Calcule a probabilidade do componente I falhar dado que o componente II falhou.

6.8.11. Um operador de rádio envia pontos e traços com igual probabilidade, mas devido a perturbações atmosféricas, os pontos são muitas vezes entendidos pelo receptor como traços e vice-versa. Seja $1/5$ a probabilidade de um ponto ser recebido como traço e $1/4$ a probabilidade de um traço ser recebido como ponto. Supondo que o receptor interpreta todos os pontos aparentes como pontos verdadeiros (o mesmo valendo para os traços), qual é a probabilidade de haver um erro na transmissão?

6.8.12. Uma caixa contém 4 válvulas defeituosas e 6 perfeitas. Duas válvulas são extraídas juntas. Sabendo que uma delas é perfeita, qual é a probabilidade da outra válvula também ser perfeita?

6.8.13. Suponha que temos duas urnas I e II, cada uma com duas gavetas. A urna I contém uma moeda de ouro em uma gaveta e uma moeda de prata na outra gaveta, enquanto a urna II contém uma moeda de ouro em cada gaveta. Uma urna é escolhida ao acaso e, em seguida, uma de suas gavetas é aberta ao acaso. Sabendo que a moeda encontrada nesta gaveta é de ouro, qual é a probabilidade de que a moeda provenha da urna II?

6.8.14. Uma montagem electrónica é formada de dois subsistemas. Supondo que a probabilidade do primeiro sistema falhar é igual a 0.20, que a probabilidade ambos falharem é 0.15 e que a probabilidade do segundo sistema falhar sozinho é 0.15, calcule:

a) A probabilidade do primeiro sistema ter falhado dado que o segundo sistema falhou.

b) A probabilidade de ocorrer falha apenas no primeiro sistema.

6.8.15. Em um lote de 100 chips semicondutores 20 são defeituosos. Dois deles são seleccionados ao acaso e sem reposição.

- Qual é a probabilidade do primeiro chip seleccionado ser defeituoso?
- Qual é a probabilidade do segundo chip seleccionado ser defeituoso, dado que o primeiro deles é defeituoso?
- Como a resposta do item (b) mudaria se os chips seleccionados fossem repostos antes da próxima selecção?

6.8.16. Amostras de uma peça de alumínio fundido são classificadas em duas categorias de acabamento: “excelente” e “bom”. Uma outra classificação divide as peças em duas categorias de comprimento: “excelente” e “bom”. A tabela abaixo exhibe o número de peças por categoria de um determinado lote:

		Comprimento	
		Excelente	Bom
Acabamento da superfície	Excelente	75	7
	Bom	10	8

Suponhamos que uma peça é seleccionada aleatoriamente deste lote.

- Qual é a probabilidade da peça ter um excelente acabamento na superfície;
- Qual é a probabilidade da peça ter um excelente comprimento;
- Se a peça seleccionada tiver excelente acabamento na superfície, qual é a probabilidade do comprimento ser excelente?
- Se a peça seleccionada tiver bom comprimento, qual é a probabilidade do acabamento na superfície ser excelente?

6.8.17. Duas válvulas defeituosas se misturam com duas válvulas perfeitas. As válvulas são seleccionadas, uma a uma e sem reposição, até que ambas as defeituosas sejam encontradas.

- Qual é a probabilidade de encontrar a última válvula defeituosa no segundo ensaio?
- Qual é a probabilidade de encontrar a última válvula defeituosa no terceiro ensaio?
- Qual é a probabilidade de encontrar a última válvula defeituosa no quarto ensaio?
- Some os números obtidos em (a), (b) e (c) acima. O resultado surpreende?

6.8.18. Suponha que A e B são eventos independentes associados a um experimento. Se a probabilidade de A ou B ocorrerem for igual a 0.6 e a probabilidade da ocorrência de A for igual a 0.4, determine a probabilidade da ocorrência de B.

6.8.19. Vinte peças, 12 das quais são defeituosas e 8 perfeitas, são inspeccionadas uma após a outra. Se estas peças forem extraídas ao acaso, qual é a probabilidade de que:

- a) Qual é a probabilidade das duas primeiras peças serem defeituosas?
- b) Qual é a probabilidade das duas primeiras peças serem perfeitas?
- c) Dentre as duas primeiras peças inspeccionadas, qual é a probabilidade de uma ser perfeita e a outra defeituosa?

6.8.20. No design preliminar de produtos são utilizadas avaliações de clientes. No passado, 95% dos produtos de alto sucesso receberam boas avaliações, 60% dos produtos de sucesso moderado receberam boas avaliações, e 10% dos produtos de pobre desempenho receberam boas avaliações. Além disso, 40% dos produtos tiveram alto sucesso, 35% tiveram sucesso moderado e 25% tiveram desempenho pobre.

- a) Qual é a probabilidade de que o produto consiga uma boa avaliação?
- b) Se um novo design obtém uma boa avaliação, qual a probabilidade de que ele tenha alto sucesso?
- c) Se um produto não recebe uma boa avaliação, qual é a probabilidade de que ele tenha sucesso moderado?

6.8.21. Um software que detecta fraudes em cartões telefónicos detecta o número de áreas metropolitanas onde as chamadas são originadas a cada dia. São obtidos os seguintes dados:

- 1% dos usuários legítimos chamam de duas ou mais áreas metropolitanas em um mesmo dia.
- 30% dos usuários fraudulentos chamam de duas ou mais áreas metropolitanas em um mesmo dia.
- A proporção de usuários fraudulentos é de 0.01%.

Se um mesmo usuário faz chamadas de duas ou mais áreas metropolitanas em um mesmo dia, qual é a probabilidade de que o usuário seja fraudulento?

6.8.22. Em uma fábrica de parafusos, as máquinas A, B e C produzem 25%, 35% e 40% do total, respectivamente. Da produção de cada máquina, 5%, 4% e 2%, respectivamente, são parafusos defeituosos. Escolhe-se ao acaso um parafuso e verifica-se que é defeituoso.

- a) Qual a probabilidade de que o parafuso tenha sido produzido na máquina A?
- b) Qual a probabilidade de que o parafuso tenha sido produzido na máquina B?
- c) Qual a probabilidade de que o parafuso tenha sido produzido na máquina C?

6.8.23. Um inspetor trabalhando para uma companhia de manufatura tem uma probabilidade de 99% de identificar correctamente um item com defeito e 0.5% de probabilidade de classificar incorrectamente um produto bom como defeituoso. A companhia tem evidências de que sua linha produz 0.9% de itens defeituosos.

- a) Qual é a probabilidade de um item seleccionado para inspecção ser classificado como defeituoso?
- b) Se um item seleccionado aleatoriamente é classificado como não-defeituoso, qual é a probabilidade dele ser realmente bom?

6.8.24. Sejam A; B; C três eventos de um espaço amostral. Exprima os eventos abaixo usando as operações união, intersecção e complementação:

- a) Somente A ocorre
- b) A, B e C ocorrem
- c) Pelo menos um ocorre
- d) Exactamente dois ocorrem

6.8.25. Em um arquivo há 4 balancetes de orçamento e 3 balancetes de custos. Em uma auditoria, o auditor selecciona aleatoriamente um destes balancetes.

- a) Qual é a probabilidade de que seja um balancete de custos? E de orçamento?
- b) Supondo que o auditor retira sequencialmente 2 balancetes sem reposição. Qual é a probabilidade de serem sorteados:
 - I. 2 balancetes de custos?
 - II. 2 balancetes de orçamento?

III. 2 balancetes de tipos diferentes?

6.8.26. De um total de 500 empregados, 200 possuem plano pessoal de aposentadoria complementar, 400 contam com o plano de aposentadoria complementar oferecido pela empresa e 200 empregados possuem ambos os planos. Sorteie-se aleatoriamente um empregado dessa empresa.

- a) Qual é a probabilidade de que ele tenha algum plano de aposentadoria complementar?
- b) Qual é a probabilidade de que ele não possua qualquer plano de aposentadoria complementar?
- c) Se o empregado conta com o plano de aposentadoria complementar oferecido pela empresa, qual é a probabilidade de que ele tenha plano pessoal de aposentadoria complementar?
- d) Se o empregado tem plano pessoal de aposentadoria complementar, qual é a probabilidade de que ele conte com o plano de aposentadoria complementar da empresa?
- e) Verifique se os eventos “empregado tem o plano de aposentadoria complementar da empresa” e o “empregado possui um plano pessoal de aposentadoria complementar” são ou não independentes.

6.8.27. Se um avião está presente em determinada área, um radar detecta sua presença com probabilidade 0,99. No entanto, se o avião não está presente, o radar detecta erradamente a presença de um avião com probabilidade 0,02. A probabilidade de um avião estar presente nesta área é de 0,05.

- a) Qual é a probabilidade de um falso alarme?
- b) Qual é a probabilidade de o radar deixar de detectar um avião?
- c) Dado que o radar acaba de detectar o avião, qual é a probabilidade de não estar presente?

6.8.28. A probabilidade de que uma nova campanha publicitária fique pronta antes do prazo estipulado pela directoria foi estimada em 0,60. A probabilidade de que a directoria aprove essa campanha publicitária é de 0,50. A probabilidade de que ambos os objectivos sejam atingidos é 0,30.

- a) Qual é a probabilidade de que pelo menos um dos objectivos seja atingido?
- b) Qual é a probabilidade de que nenhum objectivo seja atingido?

c) Se a campanha ficou pronta antes do prazo estipulado, qual é a probabilidade de que a directoria a aprove?

6.8.29. O Ministério da Economia da Espanha acredita que a probabilidade da inflação ficar abaixo de 3% este ano é de 0,20; entre 3% e 4% é de 0,45 e acima de 4% é de 0,35. O Ministério acredita que, com inflação abaixo de 3%, a probabilidade de se criar mais 200.000 empregos é de 0,6, diminuindo essa probabilidade para 0,3 caso a inflação fique entre 3% e 4%; no entanto, com inflação acima de 4%, isso é totalmente impossível.

a) Qual é a probabilidade de se criarem 200.000 empregos nesse ano?

b) No ano seguinte, um economista estrangeiro constata que foram criados 200.000 empregos novos. Qual é a probabilidade de a inflação ter ficado abaixo de 3%.

6.8.30. O chefe do Sector de Compras de uma empresa trabalha com 3 grandes distribuidores de material de escritório. O distribuidor 1 é responsável por 70% dos pedidos, enquanto cada um dos outros 2 distribuidores responde por 15% dos pedidos. Dos registos gerais de compra, sabe-se que 6% dos pedidos chegam com atraso. A proporção de pedidos com atraso do distribuidor 1 é a metade da proporção do distribuidor 2 que, por sua vez, é o dobro da proporção do distribuidor 3. Calcule a percentagem de pedidos com atraso de cada um dos distribuidores.

6.8.31. O gerente de Recursos Humanos de uma empresa escolhe estagiários oriundos de dois cursos de Administração. No curso 1, a proporção de alunos com boa formação em informática é de 60%, enquanto no outro curso, essa proporção cai para 40%. Um estagiário acaba de ser contratado. A probabilidade de que tenha boa formação em informática é 0,44. Qual é na preferência (probabilidade) do gerente pelo curso 1?

6.8.32. O chefe dum escritório de contabilidade tem três auxiliares, um que trabalha em tempo integral e os outros dois que trabalham em tempo parcial. O funcionário de tempo integral é responsável por 50% dos balancetes, enquanto cada um dos funcionários de tempo parcial responde pela metade dos balancetes restantes. Nos últimos 2 meses, a proporção de balancetes com erros oriundos do funcionário de tempo integral foi de 5%, enquanto para os funcionários de

tempo parcial essas proporções foram de 6% e 8%. O chefe resolve, então, fazer um novo treinamento, discutindo os principais erros encontrados. No mês seguinte ao treinamento, a proporção de balancetes com erro cai pela metade, com cada funcionário de tempo parcial produzindo a mesma proporção de balancetes com erro, igual à metade da proporção de erros do funcionário de tempo integral. Quais são as novas proporções de balancetes com erro de cada funcionário?

6.8.33. Suponha que a entrevista de candidatura a um emprego, é uma experiência aleatória e que a nossa atenção se centra em dois acontecimentos “o candidato tem boa aparência” (acontecimento A) e “o candidato conseguiu o emprego” (acontecimento B). Sabendo que $P(A) = 0.4$, $P(B) = 0.2$ e $P(A \cup B) = 0.4$

- a) Calcule a probabilidade de o candidato ter boa aparência e conseguir o emprego;
- b) Os acontecimentos A e B são independentes?
- c) Calcule a probabilidade de o candidato ter boa aparência sabendo que conseguiu o emprego.

6.8.34. Uma pesquisa realizada entre 1000 consumidores, registou que 650 deles trabalham com cartões de crédito da bandeira MasterCard, que 550 trabalham com cartões de crédito da bandeira VISA e que 200 trabalham com cartões de crédito de ambas as bandeiras. Qual a probabilidade de ao escolhermos deste grupo uma pessoa que utiliza a bandeira VISA, ser também um dos consumidores que utilizam cartões de crédito da bandeira MasterCard?

6.8.35. Um gerente de banco tem que decidir se concede ou não empréstimo aos clientes que o solicitam. Ele analisa diversos dados para estudar a possibilidade de o cliente vir a inadimplente. Com base em dados passados, ele estima em 15% a taxa de inadimplência. Dentre os inadimplentes, ele tem 80% de chance de tomar a decisão certa, enquanto que essa chance aumenta para 95% entre os clientes adimplentes. Esse gerente acaba de recusar um empréstimo. Qual é a probabilidade de ele ter tomado a decisão correta?

6.8.36. Duas máquinas A e B produzem 3000 peças em um dia. A máquina A produz 1000 peças, das quais 3 % são defeituosas. A máquina B produz as restantes 2000, das quais 1 % são defeituosas. Da produção total em um dia uma peça é

escolhida ao acaso e, examinando-a, constata-se que é defeituosa. Qual é a probabilidade de que a peça tenha sido produzida pela máquina A?

6.8.37. Um exame de laboratório têm eficiência de 95 % para detectar uma doença quando essa doença existe de fato. Entretanto o teste aponta um resultado “falso positivo” para 1 % das pessoas sadias testadas. Se 0,5 % da população tem a doença, qual é a probabilidade de uma pessoa ter a doença dado que seu exame foi positivo?

6.8.38. Um sindicato de trabalhadores locais consiste de associados encanadores e electricistas, classificado de acordo com grau:

	Aprendiz	Jornalista	Oficial	Total
Encanadores	25	20	30	75
Electricistas	15	40	20	75
Total	40	60	50	150

Um associado do sindicato é seleccionado ao acaso. Dado que a pessoa seleccionada é um encanador, a probabilidade de que ele é um jornalista é:

6.8.39. Os depositantes do Banco X são categorizados por idade. Seleccionaremos aleatoriamente um indivíduo desse grupo de 2.000 depositantes:

Sexo/Idade	Homem	Mulher
30 ou menos	800	600
31 ou mais	400	200

Qual é a probabilidade de:

- Ser mulher de 30 anos ou menos?
- Ser homem ou (31 ou mais)?
- Ser mulher?
- Ter 30 anos ou menos, dado que ele é homem?
- São as idades e sexos dos depositantes independentes para o Banco X

6.8.40. Um empresário é conhecido por ser muito cauteloso com relação a suas informações. Ele tem um registro minucioso da composição de cada área de sua empresa e sabe que:

Área/Departamento	Executivo Sénior	Executivo Pleno	Executivo júnior
Financeiro	2	3	4
Advocacia	3	2	2
Contabilidade	4	1	1

Ele realiza visita surpresa dum dos departamentos, escolhendo aleatoriamente um deles e consegue identificar de modo imediato dois executivos. Um deles é sénior e o outro, júnior. Assuma que os três departamentos são igualmente prováveis de serem visitados (as portas das salas são idênticas e equiprováveis). Com base nesses dados, calcule:

- a) Qual é a probabilidade de visitar a área financeira?
- b) Qual é a probabilidade de a visita não ser na área de advocacia?
- c) Qual é a probabilidade de ser o departamento de contabilidade o visitado?
- d) Qual é a probabilidade de avistar um executivo júnior e um sénior no espaço amostral considerado?

6.8.41. O problema de Monty Hall é um problema matemático que surgiu a partir de um concurso televisivo dos Estados Unidos da América chamado Let's Make a Deal, exibido na década de 1970. O jogo consiste no seguinte: Monty Hall (o apresentador) apresentava 3 portas aos concorrentes, sabendo que atrás de uma delas está um carro (prémio bom) e que as outras têm prémios de pouco valor.

1. Na 1ª etapa o concorrente escolhe uma porta (que ainda não é aberta).
2. Em seguida, Monty abre uma das outras duas portas que o concorrente não escolheu, sabendo que o carro não se encontra nela.
3. Agora, com duas portas apenas para escolher e sabendo que o carro está atrás de uma delas, o concorrente tem que se decidir se permanece com a porta que escolheu no início do jogo ou se muda para a outra porta que ainda está fechada.

Qual é a estratégia mais lógica? Ficar com a porta escolhida inicialmente ou mudar de porta? Com qual das duas portas ainda fechadas o concorrente tem mais probabilidades de ganhar? Porquê?

6.8.42. Suponha que 80% de todos os estatísticos sejam tímidos, enquanto apenas 15% de todos os economistas sejam tímidos. Suponha também que 90% das pessoas em um congresso sejam economistas e os outros 10% sejam estatísticos. Se você estiver no congresso e conhecer, ao acaso, uma pessoa tímida, qual é a probabilidade de que ela seja um estatístico?

6.8.43. Os estudos epidemiológicos indicam que 20% dos idosos sofrem de uma deterioração neuropsicológica. Sabe-se que a tomografia axial computadorizada (TAC) é capaz de detectar esse transtorno em 80% dos que sofrem disso, mas que também resulta 3% de falso positivo entre pessoas com boa saúde. Se for escolhido um idoso ao acaso, sendo o resultado do seu TAC positivo, qual é a probabilidade de que ele realmente esteja enfermo?

6.8.44. Com o propósito de aferir a capacidade diagnóstica da ausência de fosfatidilglicerol (FG) no último apósito vulvar para pré-determinar doença pulmonar da membrana hialina (DPMH) foi conduzida por ESTOL, em 1987, uma pesquisa que obteve os resultados seguintes:

"Foram diagnosticados em 253 recém-nascidos 27 casos, dos quais 23 foram correctamente identificados como imaturos na gravidez por apresentar ausência de FG no apósito vulvar. Em 191 dos 226 neonatos que não apresentaram DPMH foi possível fazer o diagnóstico correto ante-natal de "maduro" pela identificação de FG no último apósito vulvar dentro de 72 horas anteriores ao nascimento. Dos 58 casos com diagnóstico de "imaturado" (FG ausente), 23 neonatos apresentaram DPMH".

Determine a sensibilidade, especificidade, valor preditivo negativo e positivo.

6.8.45. Uma família viaja ao litoral para passar um final de semana. A probabilidade de congestionamento na estrada é de 0.6. Havendo congestionamento, a probabilidade de seus filhos brigarem no carro é de 0.8 e, sem congestionamento, a briga pode aparecer com probabilidade 0.4. Quando há briga, com ou sem congestionamento, a probabilidade do pai perder a paciência com os filhos é de 0.7. É claro que havendo congestionamento o pai pode perder a paciência com os filhos mesmo sem brigas o que aconteceria com probabilidade 0.5. Quando não há congestionamentos e nem brigas, o pai dirige tranquilo e não perde a paciência. Determine a probabilidade de:

- a) Não ter havido congestionamento se o pai não perdeu a paciência com seus filhos.
- b) Ter havido briga, dado que o pai perdeu a paciência.

6.8.46. Na verificação rotineira de máquinas, observam-se as partes eléctricas, mecânica e estrutural. A probabilidade de aparecer uma falha em qualquer

uma das partes é 0.01, independente das demais. O tempo de conserto é de 10, 20 ou 50 minutos para falha eléctrica, mecânica ou estrutural, respectivamente. Se a falha eléctrica aparece junto com a falha mecânica, teremos um acréscimo de 20 minutos devido à complicações no conserto. Para uma máquina escolhida ao caso, qual a probabilidade do tempo de conserto:

- a) Durar menos de 25 minutos?
- b) Ultrapassar 40 minutos?

6.8.47. Um médico desconfia que um paciente tem tumor no abdómen, pois isto ocorreu em 70% dos casos similares que tratou. Se o paciente de fato tiver o tumor, o exame ultra-som o detectará com probabilidade 0.9. Entretanto, se ele não tiver o tumor, o exame pode, erroneamente, indicar que tem com probabilidade de 0.1. Se o exame detectou um tumor, qual é a probabilidade do paciente tê-lo de fato?

6.8.48. Três candidatos disputam as eleições para o Governo do Estado. O candidato do partido de direita tem 30% da preferência eleitoral, o de centro tem 30% e o da esquerda 40%. Sendo eleitos, a probabilidade de dar efetivamente prioridade para Educação e Saúde é de 0.4, 0.6 e 0.9 para os candidatos de direita, centro e esquerda respectivamente.

- a) Qual é a probabilidade de não ser dada prioridade a essas áreas no próximo governo?
- b) Se a área teve prioridade, qual a probabilidade do candidato de direita ter ganho a eleição?

6.8.49. Uma pesquisa realizada sobre a preferência dos consumidores por três categorias de veículos A, B e C de uma indústria automobilística revelou que dos 100 entrevistados, 65 preferiam o veículo A; 68 preferiam o veículo B; 50 preferiam o veículo C; 40 preferiam os veículos A e B; 35 preferiam os veículos A e C; 30 preferiam os veículos B e C e 4 não preferem nenhuma das três categorias de veículos. Um consumidor é selecionado ao acaso entre os entrevistados. Calcule a probabilidade de que:

- a) Ele prefira simultaneamente as três categorias;
- b) Ele prefira somente uma das categorias;
- c) Ele prefira somente a categoria B ou C;
- d) Ele prefira somente a categoria B e C;

- e) Ele prefira pelo menos uma das categorias

6.8.50.Três candidatos concorrem, independentemente, para vagas de contabilistas numa empresa. As probabilidades de apuramento do primeiro, segundo e terceiro são, respectivamente, 80%, 60% e 30%. Determine a probabilidade de selecção do primeiro e segundo, admitindo que no concurso foram apurados apenas dois candidatos.

- a) Determine a probabilidade de ser admitido apenas um candidato;
- b) Determine a probabilidade de ser admitido apenas o primeiro e terceiro candidato;
- c) Determine a probabilidade de ser admitido apenas o primeiro e segundo candidato;
- d) Determine a probabilidade de ser admitido apenas o segundo e terceiro candidato;
- e) Determine a probabilidade de serem admitidos os três candidatos candidato;
- f) Qual é a probabilidade de serem admitidos dois candidatos?
- g) Determine a probabilidade de selecção do primeiro e segundo, admitindo que no concurso foram apurados apenas dois candidatos.

6.8.51.A probabilidade de que uma bateria de automóvel, sujeita a alta temperatura no compartimento do motor, sofra baixa corrente de carga é de 0.7. A probabilidade da bateria estar sujeita a alta temperatura no compartimento do motor é de 0.05. Determine a probabilidade da bateria estar sujeita a baixa corrente de carga e a alta temperatura no compartimento do motor.

6.8.52.Suponha que um dia de produção de 850 peças fabricadas contenha 50 peças que não satisfaçam as exigências dos consumidores. Suponha que duas peças sejam seleccionadas da batelada, porém a primeira peça é repostada antes da segunda peça ser seleccionada. Qual é a probabilidade de que a segunda peça seja defeituosa, dado que a primeira peça é defeituosa?

6.8.53.Um estudo realizado por uma empresa de recursos humanos mostrou que 45% dos funcionários de uma multinacional saíram da empresa porque estavam insatisfeitos com seus salários, 28% porque consideraram que a empresa não possibilitava o crescimento profissional e 8% indicaram insatisfação tanto com o

salário como com sua impossibilidade de crescimento profissional. Qual é a probabilidade de um funcionário sair desta empresa devido a insatisfação, com o salário ou insatisfação com sua impossibilidade de crescimento profissional?

6.8.54. Uma empresária sabe por experiência, que 65% das mulheres que compram em sua loja preferem sandálias plataformas. Qual é a probabilidade de as duas próximas clientes comprarem cada uma delas, uma sandália plataforma?

6.8.55. O Sr Rungo, ao dirigir-se ao trabalho usa TPM (Transporte Público) ou Chapa (Transporte Privado) com probabilidade de 0.2 e 0.8 respectivamente. Quando toma TPM, chega atrasado 30% das vezes e quando toma o chapa, atrasa-se 20% dos dias. Se o Sr Rungo chegar atrasado ao trabalho em determinado dia, qual é a probabilidade dele haver tomado um TPM?

6.8.56. Para se estudar o comportamento do mercado automobilístico, as marcas foram divididas em 3 categorias: marca **F**, marca **W** e as demais reunidas como marca **X**. Um estudo sobre os hábitos de mudança de marca mostrou o seguinte quadro de probabilidades:

		Probabilidade de mudança para		
		W	F	X
Pesquisador de carro da marca	W	0.50	0.25	0.25
	F	0.15	0.70	0.15
	X	0.30	0.30	0.40

O primeiro carro que o indivíduo compra é de marca **W** com probabilidade de 0.50.

- a) Qual é a probabilidade de o indivíduo comprar o terceiro carro da marca **W**?
- b) Se o terceiro carro é da marca **W**, qual é a probabilidade de o primeiro também ter sido **W**?

6.8.57. Um restaurante popular apresenta dois tipos de refeição: salada completa ou um prato a base de carne. Vinte por cento dos fregueses do sexo masculino preferem salada; trinta por cento das mulheres escolhem carne; setenta e cinco por cento dos fregueses são homens. Num certo dia o primeiro freguês a sair do restaurante escolheu a salada completa. Qual é a probabilidade do freguês ser do sexo feminino?

6.8.58. Para selecionar seus funcionários, uma empresa oferece aos candidatos um curso de treinamento durante uma semana. No final do curso, eles são submetidos a uma prova e 25% são classificados como bons (B), 50% como médios (M) e os restantes 25% como fracos (F). Para facilitar a seleção, a empresa pretende substituir o treinamento por um teste contendo questões referentes a conhecimentos gerais e específicos. Para isso, gostaria de conhecer qual a probabilidade de um indivíduo aprovado no teste ser considerado fraco, caso fizesse o curso. Assim, neste ano, antes do início do curso, os candidatos foram submetidos ao teste e receberam o conceito aprovado (A) ou reprovado (R). No final do curso 80% dos funcionários bons foram aprovados, metade dos médios reprovados e 20% dos fracos aprovados. Um funcionário acaba de ser aprovado, qual é a probabilidade de ser fraco?

6.8.59. A probabilidade de um homem sobreviver mais 10 anos, a partir de uma certa data, é 0,4, e de que sua esposa sobreviva mais 10 anos a partir da mesma data é 0,5. Qual a probabilidade de:

- a) Ambos sobreviverem mais 10 anos a partir daquela data?
- b) Ao menos um deles sobreviver mais 10 anos a partir daquela data?

6.8.60. Determinado veículo pode ter problemas mecânicos ou elétricos. Se ele tiver problemas mecânicos, não para, mas se tiver problema elétrico tem de parar imediatamente. A chance de esse veículo ter problemas mecânicos é de 0,2. Já a chance do mesmo veículo ter problemas elétricos é de 0,15 se não houve problema mecânico precedente, e de 0,25 se houve problema mecânico precedente. Agora, calcule:

- a) Qual é a probabilidade de o veículo parar em determinado dia?
- b) Se o veículo parou em certo dia, qual a chance de que tenha havido defeito mecânico?
- c) Qual é a probabilidade de que tenha havido defeito mecânico em determinado dia se o veículo não parou nesse dia?

6.8.61. Uma companhia de seguros analisou a frequência com que 2000 segurados (1000 homens e 1000 mulheres) usaram hospital. Os resultados são apresentados na tabela a seguir:

	Homens	Mulheres
Usaram	100	150
Não usaram	900	850

- Qual a probabilidade de que uma pessoa segurada use o hospital?
- O uso do hospital independe do sexo do segurado?

6.8.62. Sejam A e B eventos tais que $P(A) = 0,4$; $P(A \cup B) = 0,7$ e $P(B) = p$.

- Para qual valor de p A e B são disjuntos?
- Para qual valor de p A e B são independentes?

6.8.63. Uma empresa de segurança na internet divulgou um estudo com usuários de redes sociais no qual foi constatado que:

- 54% dos usuários não divulgam seus verdadeiros primeiros nomes nas redes;
- Dentre estes usuários, 72% divulgam seus verdadeiros sobrenomes;
- Dentre os que não mentem sobre o seu primeiro nome, 25% divulgam seu verdadeiro sobrenome.

Sabendo que um usuário divulgou seu sobrenome verdadeiro, qual é a probabilidade de que o primeiro nome que ele divulgou também seja verdadeiro?

6.8.64. Dois equipamentos, A e B, para processamento de dosagens bioquímicas são colocados para teste de controlo de qualidade por 120 horas. A probabilidade de que um erro de cálculo aconteça em um equipamento do tipo A é de $1/30$, no tipo B, $1/80$ e em ambos, $1/1000$. Qual a probabilidade de que:

- Pelo menos um dos equipamentos tenha apresentado erro?
- Nenhum equipamento tenha apresentado erro?
- Apenas o equipamento A tenha apresentado erro?

6.8.65. Sabe-se que um soro da verdade, quando ministrado a um suspeito, é 90% eficaz quando a pessoa é culpada e 99% eficaz quando é inocente. Em outras palavras, 10% dos culpados são julgados inocentes, e 1% dos inocentes é julgado culpado. Se o suspeito foi retirado de um grupo em que 95% jamais cometeram qualquer crime, e o soro indica culpado qual a probabilidade de o suspeito ser inocente?

6.8.66. Em um levantamento em um bairro de 1.000 moradores, verifica-se que:

- 220 têm curso superior;
- 160 são casados;
- 100 estão desempregados;
- 50 têm curso superior, são casados e estão empregados;
- 60 têm curso superior e estão desempregados;
- 20 têm curso superior, são casados e estão desempregados.

Escolhe-se ao acaso um morador desse bairro. Qual é a probabilidade de que ele:

- a) Tenha curso superior e seja casado?
- b) Ou tenha curso superior e seja casado ou esteja empregado
- c) Ou tenha curso superior ou esteja desempregado?

6.8.67. Uma comissão de dois estudantes deve ser sorteada de um grupo de 5 alunas e 3 alunos. Sejam os eventos:

- M_1 = “primeiro estudante sorteado é mulher”
 - M_2 = “segundo estudante sorteado é mulher”
- a) Calcule a probabilidade de M_1 e M_2 .
 - b) Verifique se M_1 e M_2 são independentes.

6.8.68. Num super mercado há 2000 lâmpadas provenientes de três fábricas distintas X, Y e Z. X produziu 500, das quais 400 são boas. Y produziu 700, das quais 600 são boas e Z as restantes, das quais 500 são boas. Se sortearmos ao acaso uma das lâmpadas nesse super mercado, qual a probabilidade de que seja boa?

6.8.69. Num grupo de 100 pessoas da zona rural, 25 estão afetadas por uma parasitose intestinal A e 11 por uma parasitose intestinal B, não se verificando nenhum caso de incidência conjunta de A e B. Duas pessoas desse grupo são escolhidas, aleatoriamente, após a outra. Determine a probabilidade de que, dessa dupla, a primeira pessoa esteja afectada por A e a segunda por B.

7. VARIÁVEL ALEATÓRIA DISCRETA

Variável Aleatória (v.a.) é toda e qualquer variável associada a uma probabilidade, isto é, os seus valores estão relacionados a um experimento aleatório.

Exemplo 7.1 (Variáveis aleatórias)

Ao jogar uma moeda duas vezes, o espaço amostral associado a este experimento aleatório será: $S = \{CaCa ; CaCo ; CoCa ; CoCo\}$.

Considere que “X” represente o número de caras na face superior do lançamento da moeda. Temos então uma função definida no espaço amostral:

Ponto Amostral	X
Ca Ca	2
Ca Co	1
Co Ca	1
Co Co	0

Uma função definida em um espaço amostral é denominada variável aleatória, sendo designada, em geral, por uma letra maiúscula (X, Y, Z, ...).

Uma variável aleatória pode ser classificada como **Variável Aleatória Discreta** (v.a.d.) ou **Variável Aleatória Contínua** (v.a.c.).

7.1. Variável aleatória discreta

Considere X uma Variável Aleatória. Se o conjunto de valores de X for finito ou infinito enumerável, então X será uma Variável Aleatória Discreta (v.a.d.), sendo obtida mediante a alguma forma de contagem.

Exemplo 7.2 (v.a.d)

- Número de acidentes ocorridos em uma semana;
- Número de peças defeituosas produzidas por uma máquina;
- Número de filhos do sexo masculino de um casal.

Exemplo 7.3 (v.a.d)

Um homem possui 4 chaves em seu bolso. Como está escuro, ele não consegue ver qual a chave correta para abrir a porta de sua casa. Ele testa cada uma das chaves até encontrar a correta. Defina a v.a. X = número de chaves experimentadas até conseguir abrir a porta (inclusive a chave correta). Quais são os valores de X ?

Solução

$$X = 1,2,3,4$$

7.1.1. Função de probabilidade (f.d.p)

A Função de Probabilidade de uma Variável Aleatória Discreta X é uma função que define a probabilidade de ocorrência de cada resultado x_i desta variável, isto é, se X assume os valores $\{x_1; x_2; \dots; x_n\}$, então:

$$P(X = x_i) = P(x_i) = p_i,$$

em que a cada valor x_i associa-se a sua probabilidade de ocorrência.

A Função de Probabilidade satisfaz as seguintes condições:

- $P(x_i) \geq 0$, para todo x_i
- $\sum_{i=1}^n P(x_i) = 1$

A coleção dos pares $[x_i; P(x_i)]$, com $i = 1, 2, \dots, n$, denominaremos de **Distribuição de Probabilidade da Variável Aleatória Discreta X** , podendo ser representada por meio de tabelas e/ou gráficos (gráfico de barras).

Exemplo 7.4 (Função de probabilidades)

Um homem possui 4 chaves em seu bolso. Como está escuro, ele não consegue ver qual a chave correta para abrir a porta de sua casa. Ele testa cada uma das chaves até encontrar a correta. Defina a v.a. X = número de chaves experimentadas até conseguir abrir a porta (inclusive a chave correta). Encontre a função de probabilidade de X ?

Solução

x_i	1	2	3	4	$\sum P(x_i)$
$P(x_i)$	1/4	3/12	6/24	6/24	1

7.1.2. Função Distribuição Acumulada

A Função Distribuição Acumulada ou Função Distribuição de Probabilidade (F.D.A) também chamada função de distribuição, é a probabilidade da V.A X assumir valores menores ou iguais a t , onde t é um número real. É representada por $F(t)$, de modo que:

$$F(t) = P(X \leq t)$$

A $F(t)$ tem as seguintes propriedades:

- $F(t)$ é uma função não decrescente
- $\lim_{t \rightarrow +\infty} F(t) = 1$
- $\lim_{t \rightarrow -\infty} F(t) = 0$

Exemplo 7.5 (F.D.A)

Um homem possui 4 chaves em seu bolso. Como está escuro, ele não consegue ver qual a chave correta para abrir a porta de sua casa. Ele testa cada uma das chaves até encontrar a correta. Defina a v.a. X = número de chaves experimentadas até conseguir abrir a porta (inclusive a chave correta). Encontre a Função Distribuição Acumulada de X ?

Solução

$$F(x) = \begin{cases} 0 & \text{se } x < 1 \\ \frac{1}{4} & \text{se } 1 \leq x < 2 \\ \frac{2}{4} & \text{se } 2 \leq x < 3 \\ \frac{3}{4} & \text{se } 3 \leq x < 4 \\ 1 & \text{se } x \geq 4 \end{cases}$$

7.1.3. Esperança Matemática ou Valor Esperado

A Esperança Matemática ou Valor Esperado ($E(X)$ ou μ_x) quantifica a média de uma Variável Aleatória Discreta (v.a.d).

Seja X uma v.a.d. com a seguinte Distribuição de Probabilidade:

X_i	X_1	X_2	...	X_n	Total
$P(X_i)$	$P(X_1)$	$P(X_2)$...	$P(X_n)$	1,0

Define-se Esperança Matemática de X por:

$$E(X) = x_1 * P(x_1) + x_2 * P(x_2) + \dots + x_n * P(x_n)$$

$$E(x) = \sum_{i=1}^n x_i * P(x_i)$$

Propriedades da Esperança Matemática:

- A Esperança Matemática de uma constante é a própria constante

$$E(K) = K$$

- A Esperança Matemática do produto de uma constante por uma variável é igual ao produto da constante pela Esperança Matemática da variável

$$E(KX) = K.E(X)$$

- Se X e Y são duas variáveis aleatórias independentes

$$E(XY) = E(X).E(Y)$$

- Esperança Matemática da soma ou da subtração de duas variáveis quaisquer é igual à soma ou subtração das Esperanças Matemáticas das duas variáveis aleatórias

$$E(X \pm Y) = E(X) \pm E(Y)$$

- A Esperança Matemática da soma ou subtração de uma variável aleatória com uma constante é igual à soma ou subtração da Esperança Matemática da variável com a constante

$$E(X \pm K) = E(X) \pm K$$

Exemplo 7.6 (Esperança matemática)

Considere a v.a. X com função de probabilidade dada por:

x	-3	-1	0	2	5	8	9	Total
$P(x)$	0.25	0.30	0.20	0.10	0.07	0.05	0.03	1

Determine a Esperança da v.a. X

Solução

$$E(x) = \sum_{i=1}^n x_i * P(x_i) = -3 * 0.25 + (-1) * 0.30 + 0 * 0.20 + \dots + 9 * 0.03 = 0.17$$

7.1.4. Variância

A Variância (**V(X) ou Var(X) ou σ_x^2**) é uma medida que quantifica a dispersão dos valores em torno da média. A Variância de uma Variável Aleatória Discreta X é definida por:

$$V(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2$$

em que:

$$E(X^2) = \sum_{i=1}^n x_i^2 * P(x_i)$$

Propriedades da Variância:

- A variância de uma constante é igual à zero

$$V(K) = 0$$

- Somando ou subtraindo uma constante a uma variável aleatória sua variância não se altera

$$V(X \pm K) = V(X)$$

- Multiplicando uma variável aleatória por uma constante sua variância fica multiplicada pelo quadrado da constante

$$V(K.X) = K^2.V(X)$$

- A variância da soma ou subtração de duas Variáveis Aleatórias Independentes (X e Y) é igual à soma de suas variâncias

$$V(X \pm Y) = V(X) \pm V(Y)$$

- A variância do produto de duas variáveis independentes é igual á:

$$Var(X * Y) = E(X)^2 * Var(X) + E(Y)^2 * Var(Y)$$

Exemplo 7.7 (Variança)

Considere a v.a. X com função de probabilidade dada por:

x	-3	-1	0	2	5	8	9	Total
$P(x)$	0.25	0.30	0.20	0.10	0.07	0.05	0.03	1

Determine a Variança da v.a. X

$$V(X) = E(x^2) - [E(x)]^2$$

$$V(X) = (-3)^2 \cdot 0.25 + (-1)^2 \cdot 0.30 + \dots + (9)^2 \cdot 0.03 - 0.17^2 = 10.33$$

7.2. EXERCÍCIOS PROPOSTOS

7.2.1. Um homem possui 4 chaves em seu bolso. Como está escuro, ele não consegue ver qual a chave correta para abrir a porta de sua casa. Ele testa cada uma das chaves até encontrar a correta.

- Defina um espaço amostral para esse experimento.
- Defina a v.a. X = número de chaves experimentadas até conseguir abrir a porta (inclusive a chave correta). Quais são os valores de X ?
- Encontre a f.d.p da v.a. X .

7.2.2. Numa urna há 7 bolas brancas e 4 bolas verdes. Cinco bolas são extraídas dessa urna. Defina a v.a. X = número de bolas verdes.

- Quais são os possíveis valores de X se as extracções são feitas (1) sem reposição; (2) com reposição.
- Encontre a fdp da v.a. X .

7.2.3. Seja uma v.a. X com fdp dada na tabela a seguir:

x	0	1	2	3	4	5
$P(x)$	0	p^2	p^2	p	p	p^2

- Encontre o valor de p ;
- Calcule $P(x \geq 4)$ e $P(x < 3)$
- Calcule $P(|x - 3| \geq 2)$
- Calcule $E(x)$ e $Var(x)$

7.2.4. Em determinado sector de uma loja de departamentos, o número de produtos vendidos em um dia pelos funcionários é uma variável aleatória P com a seguinte distribuição de probabilidades (esses números foram obtidos dos resultados de vários anos de estudo):

<i>Nº de produtos</i>	0	1	2	3	4	5	6
<i>Probabilidade de venda</i>	0.1	0.4	0.2	0.1	0.1	0.05	0.05

Cada vendedor recebe comissões de venda, distribuídas da seguinte forma: se ele vende até 2 produtos em um dia, ele ganha uma comissão de 10,00 u.m por produto vendido. A partir da terceira venda, a comissão passa para 50,00 u.m.

- Qual é o número médio de produtos vendidos por cada vendedor;
- Qual a comissão média de cada um deles?

7.2.5. Um lojista mantém extensos registos das vendas diárias de um certo aparelho. O quadro a seguir dá a distribuição de probabilidades do número de aparelhos vendidos em uma semana.

<i>$x = n^\circ$ de aparelhos</i>	0	1	2	3	4	5
<i>$P(x)$</i>	0.1	0.1	0.2	0.3	0.2	0.1

- Se é de 500,00 USD o lucro por unidade vendida, qual o lucro esperado em uma semana?
- Qual é o desvio padrão do lucro?

7.2.6. O tempo t , em minutos, necessário para um operário processar certa peça é uma v.a. com fdp dada na tabela abaixo.

<i>t</i>	2	3	4	5	6	7
<i>$P(t)$</i>	0.1	0.1	0.3	0.2	0.2	0.1

Para cada peça processada, o operário ganha um fixo de 2 u.m. (unidade monetária) mas, se ele processa a peça em menos de 6 minutos, ganha 0,50 u.m. por cada minuto poupado.

- Encontre a função de distribuição da v.a. $G =$ quantia (em u.m.) ganha por peça.
- Determine a quantia média e o desvio padrão ganha por peça.

7.2.7. Um vendedor de equipamentos pesados pode visitar, num dia, um ou dois clientes, com probabilidade de $1/3$ e $2/3$ respectivamente. De cada contacto, pode resultar a venda de um equipamento por 50.000 u.m com probabilidade

1/10 ou nenhuma venda, com probabilidade 9/10. Indicando por Y o valor total de vendas diárias deste vendedor, escreva a função de probabilidade de Y e calcule o valor total esperado de vendas diárias.

7.2.8. Na produção de uma peça são empregadas 2 máquinas. A primeira é utilizada para efectivamente produzir as peças e o custo de produção é de 50,00 u.m por peça. Das peças produzidas nessa máquina, 90% são perfeitas. As peças defeituosas são colocadas na segunda máquina para a tentativa de recuperação. Nessa segunda máquina o custo de produção é de 25,00 u.m mas apenas 60% das peças são de fato recuperadas. Cada peça perfeita é vendida por 90,00 u.m e cada peça defeituosa é vendida por 20,00 u.m. Seja L o lucro por peça. Obtenha:

- A função de distribuição de probabilidades de L ;
- A função de distribuição acumulada de L ;
- O lucro esperado por peça;
- A variância do lucro.

7.2.9. Um revendedor de produtos veterinários recebe de vários laboratórios certo tipo de antibiótico, que tem custo diferenciado. Levando-se em conta a proporção fornecida e o preço apresentado por cada laboratório, pode-se considerar que o custo de uma dose de antibiótico em reais, escolhida ao acaso, é uma variável aleatória C . Admitindo a seguinte distribuição de probabilidade para C :

ci	1,00	1,10	1,20	1,30	1,40
$P(ci)$	0,2	0,3	0,2	0,2	0,1

- Determinar a média e a variância da variável aleatória C ;
- Supondo que o revendedor venda cada um desses antibióticos acrescentando 50% sobre o custo, além de um adicional de 0,10 u.m pelo frete, calcular a média e a variância da nova variável aleatória preço de revenda R .

7.2.10. Uma loteria distribui prêmios entre seus clientes da seguinte forma:

- 400 prêmios de 100,00 u.m;
- 50 prêmios de 200,00 u.m;
- 10 prêmios de 400,00 u.m.

Admitindo-se que em um concurso sejam emitidos e vendidos 10.000 bilhetes, qual o preço justo a se pagar por um bilhete?

7.2.11. Um jogador A paga 5,00 u.m a B e lança um dado. Se sair face 3, ganha 20,00 u.m. Se sair face 4, 5, ou 6, perde. Se sair face 1 ou 2, tem o direito de jogar novamente. Desta vez, lança dois dados. Se saírem duas faces 6, ganha 50,00 u.m. Se sair uma face 6, recebe o dinheiro de volta. Nos demais casos, perde. Seja L o lucro líquido do jogador A nesse jogo. Calcule a função de probabilidade de L e o lucro esperado do jogador A.

7.2.12. A demanda por um certo produto pode ser vista como uma variável aleatória X cuja função de probabilidade $f_x(x)$ é estimada por:

<i>nº de unidades demandadas x</i>	1	2	3	4
$f_x(x) = P(X = x)$	0.25	0.45	0.15	0.15

- Verifique que $f_x(x)$ realmente define uma função de probabilidade;
- Obtenha a função de distribuição acumulada de X;
- Usando a função de distribuição calculada no item anterior, calcule $P(X \leq 3.5)$.

7.2.13. Dentre os cinco alunos de um curso com coeficiente de rendimento (CR) superior a 8.5, dois serão sorteados para receber uma bolsa de estudos. Os CRs desses alunos são: 8.8; 9.2; 8.9; 9.5; 9.0.

- Designando por A; B; C; D; E os alunos, defina um espaço amostral para esse experimento;
- Seja $X = CR$ médio dos alunos sorteados. Liste os possíveis valores de X;
- Liste o evento $X \geq 9.0$;
- Encontre a função de probabilidade de X e calcule $P(X \geq 9)$.

7.2.14. Um bandido é preso em uma cela que contém 3 portas. A primeira porta o leva a um túnel que o conduz à própria cela depois de 2 dias de viagem. A segunda porta leva-o a um túnel que o conduz à própria cela depois de 4 dias de viagem. A terceira porta o conduz à liberdade depois de um dia de viagem. Se assumirmos que o bandido selecciona as portas 1, 2 e 3 com probabilidades 0.5, 0.3 e 0.2 respectivamente, qual o número esperado de dias para que alcance a liberdade?

7.2.15. Um investimento pode resultar em uma das possibilidades possíveis: lucro de US\$ 4.000, lucro de US\$ 8.000 ou prejuízo de US\$ 10.000 com probabilidades iguais a 45%, 55% e 26%, respectivamente. Determine o valor esperado para um investimento potencial.

7.2.16. A organização financeira Betha Ltda. verificou que o lucro unitário L , obtido numa operação financeira é dado pela seguinte expressão:

$$L = 1,9V - 0,9C - 4,5$$

Sabendo-se que o preço de venda unitário V tem uma distribuição de média US\$ 50,00 e desvio padrão de US\$ 2,00, e que o preço de custo unitário C tem uma distribuição de média US\$ 45,00 e desvio padrão US\$ 1,50, qual é a média e o desvio padrão do lucro unitário?

7.2.17. Um produto tem custo médio de US\$ 10,00 e desvio padrão de US\$ 0,80. Calcular o preço de venda médio, bem como seu desvio padrão, de forma que o lucro médio seja de US\$ 4,00 e seu desvio padrão de US\$ 1,00.

7.2.18. Uma empresa comercializa um produto que possui um determinado prazo de validade. Para cada unidade vendida do produto, a empresa recebe 800,00Mt e tem um custo de 200,00Mt. Nestas condições a probabilidade de um produto ser vendido antes de vencer seu prazo de validade é de 80%.

- Quanto a empresa espera lucrar em uma unidade do produto?
- Qual é a variabilidade do lucro da empresa em uma unidade do produto?

7.2.19. Seja X uma variável aleatória com função de probabilidade apresentada na tabela abaixo:

x	-2	-1	0	1	2
$P(x)$	0.1	0.3	0.1	0.2	0.3

- Calcule $E(X)$.
- Calcule $Var(X)$.
- Determine a acumulada de X .
- Calcule $P(X \geq 0 | X < 2)$.
- Seja $Y = x^2$. Determine a função de probabilidade de Y .

8. ALGUMAS DISTRIBUIÇÕES DISCRETAS

Considere as seguintes situações:

- (1) Lança-se uma moeda viciada e observa-se o resultado obtido e (b) pergunta-se a um eleitor se ele vai votar no candidato A ou B.
- (2) Lança-se uma moeda n vezes e observa-se o número de caras obtidas e (b) de uma grande população, extrai-se uma amostra de n eleitores e pergunta-se a cada um deles em qual dos candidatos A ou B eles votarão e conta-se o número de votos do candidato A.
- (3) De uma urna com P bolas vermelhas e Q bolas brancas, extraem-se n bolas sem reposição e conta-se o número de bolas brancas e (b) de uma população com P pessoas a favor do candidato A e Q pessoas a favor do candidato B, extrai-se uma amostra de tamanho n sem reposição e conta-se o número de pessoas a favor do candidato A na amostra.

Em cada uma das situações anteriores, os experimentos citados têm algo em comum: em certo sentido, temos a “mesma situação”, mas em contextos diferentes. Por exemplo, na **situação 1**, cada um dos experimentos tem dois resultados possíveis e observamos o resultado obtido. Na **situação 3**, temos uma população dividida em duas categorias e dela extraímos uma amostra sem reposição; o interesse está no número de elementos de uma determinada categoria.

Na prática, existem muitas outras situações que podem se “encaixar” nos modelos acima e mesmo em outros modelos. O que veremos nesse capítulo são alguns modelos de variáveis aleatórias discretas que podem descrever situações como as listadas anteriormente. Nesse contexto, um modelo será definido por uma variável aleatória e sua função de probabilidade, explicitando-se claramente as hipóteses de validade. De posse desses elementos, poderemos analisar diferentes situações práticas para tentar “encaixá-las” em algum dos modelos dados.

Neste capítulo, serão descritas as distribuições de probabilidade discretas mais usuais. A introdução de cada uma delas será feita através de um exemplo clássico (moeda, urna, baralho etc.) e, em seguida, serão explicitadas as características do experimento. Tais características são a ferramenta necessária para sabermos qual modelo se aplica a uma determinada situação prática. Definida a distribuição, calculam-se a média e a variância.

8.1. Distribuição Uniforme Discreta

A variável aleatória discreta X , que assume os valores x_1, x_2, \dots, x_n , têm **distribuição uniforme** se

$$f_x(x_i) = P(X = x_i) = \frac{1}{n} \quad i = 1, 2, \dots, n$$

Note que, em uma distribuição discreta uniforme, todos os valores são igualmente prováveis. Além disso, para que uma v.a. X tenha distribuição uniforme discreta, é necessário que X assuma um número finito de valores, já que $\sum_x f_x(x) = 1$

8.1.1. Esperança e Variância

Seja X uma v.a. discreta uniforme que assume valores x_1, x_2, \dots, x_n . Por definição, a esperança de X é

$$E(x) = \frac{1}{n}x_1 + \frac{1}{n}x_2 + \dots + \frac{1}{n}x_n = \bar{x}$$

ou seja, $E(X)$ é a média aritmética dos valores possíveis de X .

Com relação à variância, temos, por definição, que:

$$V(X) = E[X - E(X)]^2 = \frac{1}{n}(x_1 - \bar{x})^2 + \frac{1}{n}(x_2 - \bar{x})^2 + \dots + \frac{1}{n}(x_n - \bar{x})^2 = \sigma_x^2$$

Exemplo 7.1 (Distribuição uniforme discreta)

Considere o lançamento de uma moeda. Vamos definir a seguinte variável aleatória X associada a esse experimento:

$$X = 0; \text{ se ocorre cara}$$

$$X = 1; \text{ se ocorre coroa}$$

Para que essa v.a. tenha distribuição uniforme, é necessário supor que a moeda seja honesta e, nesse caso:

$$f_x(0) = f_x(1) = \frac{1}{2}$$

$$E(x) = \frac{0 + 1}{2} = \frac{1}{2}$$

$$Var(x) = \frac{1}{2} * \left(0 - \frac{1}{2}\right)^2 + \frac{1}{2} * \left(1 - \frac{1}{2}\right)^2 = \frac{1}{4}$$

8.2. Distribuição de Bernoulli

Considere o lançamento de uma moeda. A característica de tal experimento aleatório é que ele possui apenas dois resultados possíveis. Uma situação análoga surge quando da extração da carta de um baralho, em que o interesse está apenas na cor (preta ou vermelha) da carta sorteada.

Um experimento de Bernoulli é um experimento aleatório com apenas dois resultados possíveis; por convenção, um deles é chamado “sucesso” e o outro, “fracasso”.

A v.a. de Bernoulli é a v.a. X associada a um experimento de Bernoulli, em que se define:

$$X = \begin{cases} 1 & \text{se ocorre sucesso} \\ 0 & \text{se ocorre o fracasso} \end{cases}$$

Seja p a probabilidade de sucesso ($0 < p < 1$), a distribuição de Bernoulli é mostrada na tabela a seguir:

x	0	1
$f_x(x)$	$1 - p$	p

Obviamente, as condições definidoras de uma função de probabilidade são satisfeitas, uma vez que $p > 0$; $1 - p > 0$ e $p + (1 - p) = 1$.

O modelo probabilístico que representa este acontecimento, é:

$$P(X = x) = f(x) = p^x * (1 - p)^{1-x}$$

O valor de p é o único valor que precisamos conhecer para determinar completamente a distribuição; ele é, então, chamado parâmetro da distribuição de Bernoulli. Vamos denotar a distribuição de Bernoulli com parâmetro p por $Bern(p)$.

A função de distribuição acumulada é dada por:

$$f_x(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - p & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1 \end{cases}$$

8.2.1. Esperança e Variância

$$X \sim Bern(p) \rightarrow \begin{cases} E(X) = p \\ Var(x) = p(1 - p) \end{cases}$$

É comum denotar a probabilidade de fracasso por q , isto é, $q = 1 - p$.

Exemplo 7.2 (Distribuição de Bernoulli)

Um auditor de Receita governamental examina declarações de Imposto de Renda de pessoas físicas, cuja variação patrimonial ficou acima do limite considerado aceitável. De dados históricos, sabe-se que 10% dessas declarações são fraudulentas.

Vamos considerar o experimento correspondente ao sorteio aleatório de uma dessas declarações. Esse é um experimento de Bernoulli, em que o sucesso equivale à ocorrência de declaração fraudulenta e o parâmetro da distribuição de Bernoulli é $p = 0,1$.

Esse exemplo ilustra o fato de que “sucesso”, nesse contexto, nem sempre significa uma situação feliz na vida real. Aqui, sucesso é definido de acordo com o interesse estatístico no problema. Em uma situação mais dramática, “sucesso” pode indicar a morte de um paciente, por exemplo.

8.3. Distribuição de Binomial

Um experimento binomial consiste em repetições independentes de um experimento de Bernoulli com probabilidade p de sucesso, probabilidade essa que permanece constante em todas as repetições.

Para um experimento binomial consistindo em n repetições independentes de um experimento de Bernoulli com parâmetro p , defina a variável aleatória

$$X = \text{“número de sucessos”}$$

Então, X tem distribuição binomial com parâmetros n e p ; cuja função de probabilidade é dada por:

$$P(x = k) = \binom{n}{k} * p^k * q^{n-k} \text{ onde } k = 0, 1, \dots, n$$

8.3.1. Esperança e Variância

$$X \sim \text{Bin}(n, p) \rightarrow \begin{cases} E(X) = n * p \\ \text{Var}(x) = n * p * (1 - p) \end{cases}$$

Exemplo 7.3 (Distribuição Binomial)

Um atirador acerta, na mosca do alvo, 20% dos tiros. Se ele dá 10 tiros, qual a probabilidade de ele acertar na mosca no máximo uma vez?

Solução

Podemos pensar os tiros como experimentos de Bernoulli independentes, em que o sucesso é acertar no alvo e a probabilidade de sucesso é 0,20. Então, o problema pede $P(X \leq 1)$, em que $X =$ número de acertos em 10 tiros. Logo, $X \sim \text{bin}(10; 0,20)$

$$P(X \leq 1) = P(X = 0) + P(X = 1)$$

$$P(X \leq 1) = \binom{10}{0} * 0.20^0 * 0.80^{10} + \binom{10}{1} * 0.20^1 * 0.80^9 = 0.37581$$

8.4. Distribuição Geométrica

Consideremos, então, repetições independentes de um experimento de Bernoulli com probabilidade p de sucesso. Vamos definir a variável aleatória que representa o número de repetições até a ocorrência do primeiro sucesso. Os possíveis valores dessa variável são 1 (primeiro sucesso na primeira repetição), 2 (primeiro sucesso na segunda repetição e, portanto, fracasso na primeira), 3 (primeiro sucesso na terceira repetição e, portanto, fracasso nas duas primeiras), etc. Esse é um exemplo de v.a. discreta em que o espaço amostral, enumerável, é infinito.

Para calcular a probabilidade do evento $X = k, k = 1, 2, 3, \dots$ devemos notar que tal evento corresponde à ocorrência de fracassos nas $k - 1$ primeiras repetições e sucesso na $k - \text{ésima}$ repetição.

Para repetições independentes de um experimento de Bernoulli com parâmetro p , defina a variável aleatória

$$X = \text{“número de repetições até a ocorrência do primeiro sucesso”}$$

Então, X tem distribuição geométrica com parâmetro p cuja função de probabilidade é dada por:

$$P(X = k) = (1 - p)^{k-1} * p \text{ onde } k = 1, 2, 3, \dots$$

Dizemos que X tem distribuição geométrica com parâmetro p (o único valor necessário para especificar completamente a fdp) e vamos representar tal fato por:

$$X \sim \text{Geom}(p)$$

8.4.1. Esperança e Variância

$$X \sim \text{Geom}(p) \rightarrow \begin{cases} E(X) = \frac{1}{p} \\ \text{Var}(x) = \frac{1-p}{p^2} \end{cases}$$

Exemplo 7.4 (Distribuição Geométrica)

Determinado tipo de parafuso é vendido em caixas com 50 peças. É uma característica da fabricação produzir 90% de boas. Qual é a probabilidade de serem necessárias 8 peças para encontrar a primeira defeituosa?

Solução

Seja

v. a: $x =$

nº de parafusos necessários até se encontrar o primeiro defeituoso

Então, $x \sim \text{Geom}(p)$

Pela fórmula: $P(x = k) = p * q^{k-1}$

onde $p = 0.10$ e $n = 8$

$$P(x = 8) = 0.10 * 0.90^{8-1} = 0.048$$

8.5. Distribuição Binomial Negativa

Consideremos novamente repetições independentes de um experimento de Bernoulli com probabilidade p de sucesso. Vamos considerar agora uma generalização da v.a. geométrica, no seguinte sentido:

$X =$ número de repetições necessárias até a obtenção do r – ésimo sucesso, $r \geq 1$

Note que $r = 1$ corresponde à distribuição geométrica.

Para definir os possíveis valores de X , devemos notar que para ter r sucessos, são necessários no mínimo r repetições. Logo, os possíveis valores de X são:

$$r, r + 1, r + 2, \dots$$

O evento $X = k$ indica que foram necessárias k repetições para obter r sucessos e, portanto, $k - r$ fracassos. Pela definição da variável, a última repetição resultou em sucesso e os outros $r - 1$ sucessos podem estar em quaisquer das $k - 1$ posições restantes.

Para repetições independentes de um experimento de Bernoulli com parâmetro p , defina a variável aleatória

$$X = \text{“número de repetições até a ocorrência do } r - \text{ésimo sucesso”}$$

Então, X tem distribuição binomial negativa com parâmetros r e p cuja função de probabilidade é dada por:

$$P(x = k) = \binom{k-1}{r-1} * p^r * q^{k-r} \text{ onde } k \geq r$$

Essa distribuição é também conhecida como *distribuição de Pascal*. Se X tem tal distribuição, vamos representar esse fato por $X \sim BinNeg(r, p)$.

8.5.1. Esperança e Variância

$$X \sim BinNeg(r, p) \rightarrow \begin{cases} E(X) = \frac{r}{p} \\ Var(x) = \frac{r * (1 - p)}{p^2} \end{cases}$$

Exemplo 7.5 (Distribuição Binomial Negativa)

Deseja-se produzir 5 peças boas, em uma máquina que dá 20% de peças defeituosas. Qual é a probabilidade de ser necessário fabricar 8 peças para se conseguir as 5 peças boas?

Solução

Seja

v. a: $x = n^{\circ}$ de peças fabricadas necessários para conseguir 5 boas

Então, $X \sim BinNeg(r, p)$

$$\text{Pela fórmula: } P(x = k) = \binom{k-1}{r-1} * p^r * q^{k-r}$$

onde $p = 0.80$ e $r = 5$

$$P(x = 8) = \binom{8-1}{5-1} * 0.80^5 * 0.20^{8-5} = 0.0917504$$

8.6. Distribuição Hipergeométrica

Seja uma população de tamanho N dividida em 2 classes, uma composta de r “sucessos” e a outra composta de $N - r$ “fracassos”. Dessa população, vamos extrair uma amostra de tamanho n sem reposição. Se a variável aleatória de interesse é:

$X =$ número de sucessos na amostra, então:

$$P(X = k) = \frac{\binom{r}{k} * \binom{N-r}{n-k}}{\binom{N}{n}} \text{ onde } k = 0, 1, \dots, n$$

Essa é a distribuição hipergeométrica com parâmetros N ; r e n : Notação:

$$X \sim \text{hiper}(N; r; n)$$

8.6.1. Esperança e Variância

$$X \sim \text{hiper}(N; r; n) \rightarrow \begin{cases} E(X) = \frac{n * r}{N} \\ \text{Var}(x) = n * \frac{r}{N} * \frac{N-r}{N} * \frac{N-n}{N-1} \end{cases}$$

Exemplo 7.6 (Distribuição Hipergeométrica)

Entre os 16 programadores de uma empresa, 12 são do sexo masculino. A empresa decide sortear 5 programadores para fazer um curso avançado de programação. Qual é a probabilidade dos 5 sorteados serem do sexo masculino?

Solução

Seja $v. a: x = n^{\circ}$ de sorteados do sexo masculino

Então, $X \sim \text{hiper}(N = 16; r = 12; n = 5)$

$$\text{Pela fórmula: } P(X = k) = \frac{\binom{r}{k} * \binom{N-r}{n-k}}{\binom{N}{n}}$$

$$P(X = 5) = \frac{\binom{12}{5} * \binom{16-12}{5-5}}{\binom{16}{5}} = 0.181319$$

8.7. Distribuição de Poisson

Suponhamos que estejamos observando um determinado fenômeno de interesse por um certo período de tempo de comprimento t com o interesse de contar o número de vezes X que determinado evento ocorre.

Vamos fazer as seguintes hipóteses sobre a forma como esse evento ocorre:

- H_1 : Em um intervalo de tempo suficientemente curto, apenas 0 ou 1 evento ocorre, ou seja, 2 ou mais ocorrências não podem acontecer simultaneamente. Então, em cada um desses intervalos temos um experimento de Bernoulli.
- H_2 : A probabilidade de exatamente 1 ocorrência nesse pequeno intervalo de tempo, de comprimento Δt , é proporcional a esse comprimento, ou seja, é $\lambda \Delta t$. Logo, a probabilidade de nenhuma ocorrência é $1 - \lambda \Delta t$.

- H₃: As ocorrências em intervalos pequenos e disjuntos são experimentos de Bernoulli independentes.

Sejam eventos gerados de acordo com as hipóteses H₁ a H₃ acima. Se X é o número de eventos em um intervalo de tempo de comprimento t, então a função de probabilidade de X é:

$$P(X = k) = \frac{(\lambda t)^k * e^{-\lambda t}}{k!} \text{ Onde } k = 0, 1, \dots$$

Assumindo $\mu = \lambda t$, então:

$$P(X = k) = \frac{(\mu)^k * e^{-\mu}}{k!}$$

8.7.1. Esperança e Variância

$$X \sim \text{Poisson}(\lambda t) \rightarrow \begin{cases} E(X) = \mu = \lambda t \\ \text{Var}(x) = \mu = \lambda t \end{cases}$$

Exemplo 7.7 (Distribuição de Poisson)

Uma fábrica de pneus verificou que ao testar seus pneus nas pistas, havia em média um estouro de pneu a cada 5.000 km. Qual a probabilidade que num teste de 3.000 km haja no máximo um pneu estourado?

Solução

Seja *v. a:* $x = \text{estouro do pneu num determinado espaço em Km}$

Pelos dados temos:

Sabe – se que em cada 5000 km regista se em média um estouro

Pede – se $P(x \leq 1)$ em cada 3000 km

$$x \sim \text{Poisson}(\mu)$$

$$\text{onde } \mu = \frac{1}{5000} * 3000 = 0.6$$

Pela fórmula $P(x = k) = \frac{e^{-\mu} * \mu^k}{k!}$ temos:

$$P(x \leq 1) = P(x = 0) + P(x = 1)$$

$$P(x \leq 1) = e^{-0.6} * \left(\frac{0.6^0}{0!} + \frac{0.6^1}{1!} \right) = 0.878$$

8.8. EXERCÍCIOS PROPOSTOS

- 8.8.1.** Os defeitos em determinada máquina ocorrem aproximadamente na mesma frequência. Dependendo do tipo de defeito, o técnico leva 1, 2, 3, 4 ou 5 horas para consertar a máquina.
- Descreva o modelo probabilístico apropriado para representar a duração do tempo de reparo da máquina.
 - Qual é o tempo médio de reparo desta máquina? E o desvio-padrão deste tempo de reparo?
 - São 15 horas e acaba de ser entregue uma máquina para reparo. A jornada normal de trabalho do técnico termina às 17 horas. Qual é a probabilidade de que o técnico não precise fazer hora extra para terminar o conserto desta máquina?
- 8.8.2.** A probabilidade de se encontrar um determinado semáforo aberto é igual a 20%. Qual é a probabilidade de passar pelo semáforo sucessivas vezes e encontrá-lo aberto pela primeira vez na quinta passagem?
- 8.8.3.** Em uma linha de produção, a probabilidade de um determinado componente ser defeituoso é de 10%. Qual é a probabilidade de se produzir 4 componentes defeituosos antes de 20 perfeitos.
- 8.8.4.** Deseja-se produzir 5 peças boas, em uma máquina que dá 20% de peças defeituosas. Qual é a probabilidade de ser necessário fabricar 8 peças para se conseguir as 5 peças boas?
- 8.8.5.** Um atirador acerta na mosca do alvo, 20% dos tiros. Se ele dá 10 tiros, qual a probabilidade de ele acertar na mosca no máximo 1 vez?
- 8.8.6.** Dois adversários A e B disputam uma série de 8 partidas de um determinado jogo. A probabilidade de A ganhar uma partida é 0,6 e não há empate. Qual é a probabilidade de A ganhar a série?
- 8.8.7.** Um caçador, após um dia de caça, verificou que matou 5 andorinhas e 2 aves de uma espécie rara, proibida de ser caçada. Como todos os espécimes tinham o mesmo tamanho, ele os colocou na mesma bolsa, pensando em dificultar o trabalho dos fiscais. No posto de fiscalização há dois fiscais, Manoel e Pedro,

que adoptam diferentes métodos de inspeção. Manuel retira três espécimes de cada bolsa dos caçadores. Pedro retira um espécime, classifica-o e o repõe na bolsa, retirando em seguida um segundo espécime. Em qualquer caso, o caçador é multado se é encontrado pelo menos um espécime proibido. Qual dos dois fiscais é mais favorável para o caçador em questão?

8.8.8. Entre os 16 programadores de uma empresa, 12 são do sexo masculino. A empresa decide sortear 5 programadores para fazer um curso avançado de programação. Qual é a probabilidade dos 5 sorteados serem do sexo masculino?

8.8.9. Uma central telefónica recebe uma média de 5 chamadas por minuto. Supondo que as chamadas que chegam constituam uma distribuição de Poisson.

- a) Qual é a probabilidade de a central não receber nenhuma chamada em um minuto?
- b) Qual é a probabilidade de receber no máximo 2 chamadas em 2 minutos?

8.8.10. Seja $X \sim bin(n; p)$. Se $E(X) = 12$ e $Var(X) = 4$, determine os valores de n e p .

8.8.11. A probabilidade de uma máquina produzir uma peça defeituosa em um dia é 0,1.

- a) Qual a probabilidade de que, em 20 peças produzidas em um dia, exactamente 5 sejam defeituosas?
- b) Qual a probabilidade de que a 10ª peça produzida em um dia seja a primeira defeituosa?

8.8.12. Um atirador acerta na mosca do alvo, 20% dos tiros.

- a) Qual a probabilidade de ele ter de dar 10 tiros para acertar 6 vezes na mosca?
- b) Se ele dá 10 tiros, qual a probabilidade de ele acertar na mosca 6 vezes?

8.8.13. Suponha que 100 peixes especiais são pescados, marcados e colocados no lago de um sitiante, que passa a ter, então, um total de 2000 peixes. Num certo dia, o sitiante autoriza a realização de uma pescaria. Assuma que a quantidade de

peixes no lago até esse dia não se alterou. Se 60 peixes são pescados, calcule a probabilidade de 5 serem especiais

- a) Supondo que os peixes pescados são colocados de volta no lago;
- b) Supondo que os peixes pescados não são colocados de volta no lago.

8.8.14. Um supermercado faz a seguinte promoção: o cliente, ao passar pelo caixa, lança um dado.

Se sair face 6 tem um desconto de 30% sobre o total de sua conta. Se sair face 5 o desconto é de 20%. Se sair face 4 o desconto é de 10% e se ocorrerem faces 1, 2 ou 3, o desconto é de 5%. Seja $X = \text{desconto concedido}$:

- a) Encontre a função de distribuição de probabilidade de X ;
- b) Calcule o desconto médio concedido;
- c) Calcule a probabilidade de que, num grupo de 5 clientes, pelo menos um consiga um desconto maior que 10%;
- d) Calcule a probabilidade de que o quarto cliente seja o primeiro a receber 30% de desconto.

8.8.15. As probabilidades de que haja 1, 2, 3, 4 ou 5 pessoas nos carros que passam por um pedágio são, respectivamente, 0,05; 0,20; 0,40; 0,25 e 0,10. Seja

$X = \text{número de passageiros por veículo}$:

- a) Explícite a função de distribuição de probabilidade de X ;
- b) Calcule o número médio de passageiros por veículo;
- c) Calcule a probabilidade de que, num grupo de 5 carros, pelo menos um tenha mais que 3 pessoas;
- d) Calcule a probabilidade de que o quarto carro seja o primeiro a ter 5 passageiros.

8.8.16. Numa estrada há 2 acidentes para cada 100 km. Qual a probabilidade de que:

- a) Ocorram pelo menos 3 acidentes em 250 km?
- b) Ocorram 5 acidentes em 300 km?

8.8.17. Na manufatura de certo artigo, é sabido que 1 entre 10 artigos é defeituoso. Uma amostra de tamanho 4 é retirada com reposição, de um lote da produção. Qual a probabilidade de que a amostra contenha

- a) Nenhum defeituoso?
- b) Pelo menos 2 defeituosos?
- c) Exactamente 1 defeituoso?

8.8.18. Um fabricante de peças de automóveis garante que uma caixa de suas peças conterà, no máximo, 2 defeituosas. Se a caixa contém 18 peças e a experiência mostra que esse processo de fabricação produz 5% de peças defeituosas, qual a probabilidade de que uma caixa satisfaça a garantia?

8.8.19. Certo curso de treinamento aumenta a produtividade de uma certa população de funcionários em 80% dos casos. Se 10 funcionários quaisquer participam deste curso, encontre a probabilidade de:

- a) Exactamente 7 funcionários aumentarem a produtividade;
- b) Pelo menos 3 funcionários não aumentarem a produtividade;
- c) Não mais que 8 funcionários aumentarem a produtividade.

8.8.20. Determinado tipo de parafuso é vendido em caixas com 1000 peças. É uma característica da fabricação produzir 10% de defeituosos. Normalmente, cada caixa é vendida por 13,50 u.m. Um comprador faz a seguinte proposta para o produtor: de cada caixa, ele escolhe uma amostra de 20 peças; se ele encontrar 0 defeituoso, ele paga 20,00 u.m pela caixa; 1 ou 2 defeituosos, ele paga 10,00 u.m.; 3 ou mais defeituosos, ele paga 8,00 u.m. Qual é a alternativa mais vantajosa para o fabricante.

8.8.21. Numa central telefónica, o número de chamadas chega segundo uma distribuição de Poisson, com a média de 8 chamadas por minuto. Determinar qual a probabilidade de que em um minuto se tenha:

- a) 10 ou mais chamadas;
- b) Menos de 5 chamadas.

8.8.22. As chegadas de petroleiros a uma refinaria em cada dia ocorrem segundo uma distribuição de Poisson, com parâmetro $\lambda = 2$. As atuais instalações podem atender, no máximo, a 3 petroleiros por dia. Se mais de 3 petroleiros chegarem num dia, o excesso é enviado a outro porto.

- a) Em um dia, qual a probabilidade de se enviar petroleiros para outro porto?
- b) De quanto deverão ser aumentadas as instalações para permitir atender a todos os navios que chegam em pelo menos 95% dos dias?

8.8.23. Um industrial fabrica peças, das quais 20% são defeituosas. Dois compradores, A e B, classificam as partidas adquiridas em categorias I e II, pagando 1,20 u.m. e 0,80 u.m respectivamente, do seguinte modo:

- Comprador A: retira uma amostra de 5 peças; se encontrar mais que uma defeituosa, classifica como II;
- Comprador B: retira uma amostra de 10 peças; se encontrar mais que 2 defeituosas, classifica como II.

Em média, qual comprador oferece maior lucro para o fabricante?

8.8.24. Sabendo que 30% das vendas de uma loja são feitas a comerciantes e que foram atendidos 10 clientes, construa a distribuição de probabilidades do número de comerciantes atendidos e desenhe o gráfico da mesma. Calcule a esperança (média), a variância e o desvio padrão da distribuição.

8.8.25. Uma moeda é lançada 20 vezes. Qual a probabilidade de saírem 8 caras?

8.8.26. Numa criação de coelhos, 40% são machos. Qual a probabilidade de que nasçam pelo menos 2 coelhos machos num dia em que nasceram 20 coelhos?

8.8.27. Uma prova tipo teste tem 50 questões independentes. Cada questão tem 5 alternativas. Apenas uma das alternativas é correta. Se um aluno resolve a prova respondendo a todas as questões, qual a probabilidade de tirar nota 10?

8.8.28. Achar a esperança (média) e a variância da variável aleatória

$Y = 3X + 2$, sabendo que X é um variável aleatória com distribuição binomial com $n = 20$ e $p = 0,3$.

8.8.29. Uma caixa contém 12 lâmpadas das quais 5 estão queimadas. São escolhidas 6 lâmpadas ao acaso. Qual a probabilidade de que:

- a) Exactamente duas estejam queimadas?

- b) Pelo menos uma esteja boa?
- c) Pelo menos duas estejam queimadas?
- d) O número esperado de lâmpadas queimadas?
- e) A variância do número de lâmpadas queimadas?

8.8.30. Pequenos motores são guardados em caixas de 50 unidades. Um inspetor de qualidade examina cada caixa, antes da posterior remessa, testando 5 motores. Se nenhum motor for defeituoso, a caixa é aceita. Se pelo menos um motor for defeituoso, todos os 50 motores são testados. Há 6 motores defeituosos numa caixa. Qual a probabilidade de que seja necessário examinar todos os motores dessa caixa?

8.8.31. Uma firma compra lâmpadas por centenas. Examina sempre uma amostra de 15 lâmpadas para verificar se estão boas. Se uma centena inclui 12 lâmpadas queimadas, qual a probabilidade de se escolher uma amostra com pelo menos uma lâmpada queimada?

8.8.32. Num livro de 800 páginas há 800 erros de impressão. Qual a probabilidade de que uma página contenha pelo menos 3 erros?

8.8.33. Numa central telefónica chegam 300 telefonemas por hora. Qual a probabilidade de que:

- a) Num minuto não haja nenhum chamado;
- b) Em 2 minutos haja 2 chamados;
- c) Em t minutos não haja chamados;

8.8.34. Em um certo tipo de fabricação de fita magnética, ocorrem defeitos a uma taxa de 1 a cada 2000 metros. Qual a probabilidade de que um rolo com 2000 metros de fita magnética:

- a) Não tenha defeitos?
- b) Tenha no máximo dois defeitos?
- c) Tenha pelo menos dois defeitos?

8.8.35. Um dado é formado por chapas de plástico de 10x10 cm. Em média aparecem 50 defeitos por metro quadrado de plástico, segundo uma distribuição de Poisson.

- a) Qual a probabilidade de uma determinada face apresentar exactamente 2 defeitos?
- b) Qual a probabilidade de o dado apresentar no mínimo dois defeitos?
- c) Qual a probabilidade de que pelo menos 5 faces sejam perfeitas?

8.8.36. A Air America adopta a política de vender 15 passagens para um avião que dispõe de apenas 14 assentos. (A experiência passada mostra que apenas 85% dos que reservam lugar comparecem efectivamente ao embarque.) Determine a probabilidade de não haver assentos suficientes no caso de a Air América vender 15 passagens.

8.8.37. A Telektronic Company compra grandes lotes de lâmpadas fluorescentes e adopta o seguinte método: seleccionar aleatoriamente e testar 24 lâmpadas e aceitar todo o lote se no máximo uma não funcionar. Se determinado lote de lâmpadas tem efectivamente 4 % de unidades defeituosas, qual é a probabilidade de todo o lote ser aceito?

8.8.38. A Loja de Departamentos Newtower constatou uma taxa de 3,2% de queixas de clientes e decidiu reduzir essa taxa mediante um programa de treinamento de seus empregados. Ao fim do programa, observaram-se 850 clientes.

- a) Admitindo que o programa de treinamento não tenha produzido efeito, determine a média e o desvio-padrão do número de queixas nesses grupos de 850 clientes.
- b) No grupo de 850 clientes observados, 7 tiveram alguma queixa. Esse resultado é excepcional? O programa de treinamento parece ter sido eficaz?

8.8.39. As pacientes diagnosticadas com câncer de mama precocemente têm 80% de probabilidade de serem completamente curadas. Para um grupo de 12 pacientes nessas condições, calcule a probabilidade de:

- a) Oito ficarem completamente curadas.
- b) Entre 3 e 5 (inclusive) não ficarem curadas.
- c) Não mais de 2 permanecerem com a doença.

8.8.40. A probabilidade de um sapato apresentar defeito de fabricação é de 2 %. Para um par de sapatos ser rejeitado pelo controle de qualidade basta que um dos pés, direito ou esquerdo, apresente defeito. Numa partida de 10.000 pares,

qual o valor esperado e o desvio-padrão do número de pares totalmente defeituosos?

8.8.41. Certa empresa fabricante de artigos para desenho resolveu inserir em seus produtos determinados tipos de lápis, cujos grafites são importados. Esses grafites vêm acondicionados em embalagens contendo seis unidades cada. Após a primeira remessa recebida, verificou-se que 3% deles são recebidos com quebra. Calcular a probabilidade de:

- a) Menos da metade dos grafites de certa caixa apresentarem defeitos;
- b) No mínimo três caixas, de um grupo de oito, apresentarem um grafite quebrado.

8.8.42. Determinada empresa tem quatro eventuais compradores de seu produto, que pagam preços em função da qualidade:

- O comprador A paga 1.300 dólares por peça, se em uma amostra de cinco peças não encontrar nenhuma defeituosa e 650 dólares pelo restante;
- O comprador B paga 900 dólares por peça, desde que encontre no máximo uma peça defeituosa em cinco peças, pagando pelo restante 700 dólares;
- O comprador C paga 620 dólares por peça, aceitando até três defeituosas em uma amostra de cinco, e paga pelo restante 430 dólares;
- O comprador D não exige nenhuma inspeção, mas paga apenas 540 dólares por peça.

Qual dos compradores não deveria ser escolhido pelo empresário, se ele sabe que na produção 8% são totalmente defeituosas?

8.8.43. A Empresa Spelunke S.A. adota o seguinte critério no sector de controlo de qualidade: para cada lote de 90 unidades de seu produto, testa, por amostragem, apenas oito. O critério de avaliação final é feito da seguinte maneira: se forem encontradas no máximo duas peças defeituosas, o lote é aceito normalmente, caso contrário, deve-se passar por outra inspeção. Admitindo-se a existência de três peças defeituosas por lote, calcular:

- a) A probabilidade de não haver inspeção total em certo lote;

- b) A probabilidade de somente dois lotes, de um grupo de cinco lotes iguais, apresentarem no máximo uma peça defeituosa por lote;
- c) Se o custo operacional para cada lote for de 600 dólares, estimar o custo médio de inspeção para 60 lotes recebidos.

8.8.44. O fluxo de carros que passam em determinado pedágio é 1,7 carro por minuto. Qual a probabilidade de passarem exactamente dois carros em dois minutos?

8.8.45. Sabe-se por experiência que 1,5% das pastilhas de freio fabricadas por determinada empresa apresentam defeito. O controlo de qualidade da empresa, para tal, escolheu, ao acaso, 100 peças de pastilhas. Determinar a probabilidade de que:

- a) No máximo duas sejam defeituosas;
- b) Pelo menos duas apresentem defeitos.

8.8.46. Um distribuidor de gasolina tem capacidade de receber, nas condições atuais, no máximo, três camiões por dia. Se chegarem mais que três camiões, o excesso deve ser enviado a outro distribuidor, e, nesse caso, há uma perda média de 800 dólares, por dia, em que não se pode aceitar todos camiões. Sabendo-se que o número de camiões que chegam diariamente obedece à distribuição de Poisson de média 2, calcular:

- a) A probabilidade de chegarem de três a cinco camiões no total de dois dias;
- b) A probabilidade de, em certo dia, ter-se que mandar camiões para outro distribuidor;
- c) A perda média mensal (30 dias) por causa de camiões que não puderam ser aceites.

8.8.47. Uma loja vende, em média, 2,5 fogões por dia. Certo dia, ao encerrar o expediente, verifica-se existirem três fogões em estoque, e sabe-se que a nova remessa só chegará depois de dois dias. Qual a probabilidade de, no fim desses dois dias, a loja não ter deixado de atender, por falta de estoque, às pessoas que vierem comprar?

8.8.48. O Departamento de Atendimento da Empresa Mondubim Ltda. está dimensionado a poder atender, no período diário normal, até cinco pedidos de clientes; se chegarem mais que cinco pedidos, o pessoal deve recorrer a horas

extras para cumprir o atendimento. Sabendo-se que o número de pedidos que chegam diariamente é distribuído segundo Poisson de média 4,2 pedidos, calcular:

- a) A probabilidade de se ter que fazer horas extras em certo dia;
- b) Sendo o custo diário de horas extras de US\$ 4.500, qual será o custo médio semanal em virtude das mesmas? Considerar semana de seis dias.

8.8.49. Ao observar-se a duração das baterias de vídeo-games do tipo Gameboy, conclui-se que esta vida nada mais é do que o intervalo entre falhas sucessivas das baterias; para essas falhas, pode-se aplicar o processo de Poisson. Desse modo, o tempo médio entre falhas vem a ser a vida média da bateria. Considere que inúmeras baterias foram usadas e anotou-se (algo raro de ocorrer no dia-a-dia, somente as fábricas o fazem) que a cada sete dias havia necessidade de trocá-las (ou seja, a vida média da bateria é de uma semana). As falhas das baterias são aleatórias e independentes e atendem às condições da distribuição de Poisson:

- a) Determine a probabilidade de a bateria durar pelo menos 2 semanas;
- b) Determine a probabilidade de uma bateria falhar dentro de 3 dias;
- c) Determine a probabilidade de uma bateria durar de 3 a 4 semanas;

9. VARIÁVEL ALEATÓRIA CONTÍNUA

Uma variável aleatória é uma função real (isto é, que assume valores em \mathbb{R}), definida no espaço amostral Ω de um experimento aleatório. Dito de outra forma, uma variável aleatória é uma função que associa um número real a cada evento de Ω .

Uma variável aleatória é **discreta** se sua imagem (ou conjunto de valores que ela assume) for um conjunto finito ou enumerável. Se a imagem for um conjunto não enumerável, dizemos que a variável aleatória é **contínua**.

9.1. Função densidade de probabilidades (f.d.a)

Uma função densidade de probabilidade é uma função $f(x)$ que satisfaz as seguintes propriedades:

- $f(x) \geq 0$
- $\int_{-\infty}^{+\infty} f(x)dx = 1$
- Dada uma função $f(x)$ satisfazendo as propriedades acima, então $f(x)$ representa alguma variável aleatória contínua X , de modo que:

$$P(a \leq X \leq b) = \int_a^b f(x)dx = 1$$

9.2. Função distribuição acumulada (f.d.a)

A função distribuição acumulada $F(t)$ é:

$$F(t) = P(X \leq t)$$

$$F(t) = \int_{-\infty}^t f(t)dx = 1$$

A $F(t)$ tem as seguintes propriedades:

- $F(t)$ é uma função não decrescente
- $\lim_{t \rightarrow +\infty} F(t) = 1$
- $\lim_{t \rightarrow -\infty} F(t) = 0$

9.3. Esperança de variáveis aleatórias contínuas

Seja X uma variável aleatória contínua com função densidade de probabilidade f_x . A esperança (ou média ou valor esperado) de X é definida como:

$$E(x) = \int_{-\infty}^{+\infty} x * f_x(x) dx$$

9.4. Variância de variáveis aleatórias contínuas

Seja X uma variável aleatória contínua com função densidade de probabilidade f_x . A variância de X é definida como:

$$Var(x) = E(x^2) - [E(X)]^2 = \int_{-\infty}^{+\infty} x^2 * f_x(x) dx - \left[\int_{-\infty}^{+\infty} x * f_x(x) dx \right]^2$$

9.5. Propriedades da média e variância

As mesmas propriedades vistas para variáveis aleatórias discretas continuam valendo no caso contínuo:

Esperança	Variância
$E(a) = a$	$Var(a) = 0$
$E(X + a) = E(X) + a$	$Var(X + a) = Var(X)$
$E(bx) = b * E(X)$	$Var(bx) = b^2 * Var(X)$
$x_{min} \leq E(X) \leq x_{max}$	$Var(X) \geq 0$

Esses resultados podem ser facilmente demonstrados a partir das propriedades da integral definida e das definições vistas.

Exemplo 9.1 (Variáveis aleatórias contínuas)

Considere a função:

$$f(x) = \begin{cases} 0.06 + 0.04x & \text{se } 1 \leq x \leq 6 \\ 0 & \text{Caso contrário} \end{cases}$$

a) Calcule $P(2 \leq x \leq 3)$

Solução

$$P(2 \leq X \leq 3)$$

$$= \int_2^3 (0.06 + 0.04x)dx = (3 - 2) * 0.06 + \frac{9 - 4}{2} * 0.04 = 0.16$$

b) Encontre a função de distribuição de X.

$$F(x) = \begin{cases} 0 & \text{se } t < 1 \\ \int_1^x (0.06 + 0.04t)dt & \text{se } 1 \leq t < 6 \\ 1 & \text{se } t \geq 6 \end{cases}$$

$$\rightarrow F(x) = \begin{cases} 0 & \text{se } x < 1 \\ 0.02x^2 + 0.06x - 0.08 & \text{se } 1 \leq x < 6 \\ 1 & \text{se } x \geq 6 \end{cases}$$

c) Determine o valor de k tal que $P(x \leq k) = 0.6$

$$\int_1^k (0.06 + 0.04x)dx = 0.6$$

$$0.02 * (k^2 - 1) + 0.06 * (k - 1) = 0.6$$

$$k = 5.8$$

Nota: O valor de k aqui determinado corresponde ao percentil 60

d) Calcule a esperança e a variância de X.

$$E(x) = \int_1^6 x * (0.06 + 0.04x)dx = \frac{0.06}{2} * (6^2 - 1^2) + \frac{0.04}{3} * (6^3 - 1^3)$$

$$= 3.92$$

e

$$Var(x) = \int_{-\infty}^{+\infty} x^2 * f_x(x)dx - \left[\int_{-\infty}^{+\infty} x * f_x(x)dx \right]^2$$

$$\frac{0.06}{3} * (6^3 - 1^3) + \frac{0.04}{4} * (6^4 - 1^4) - 3.92^2 = 1.88$$

9.6. EXERCÍCIOS PROPOSTOS

9.6.1. Considere a função:

$$f(x) = \begin{cases} 0.06 + 0.04x & \text{se } 1 \leq x \leq 6 \\ 0 & \text{Caso contrário} \end{cases}$$

- Calcule $P(2 \leq x \leq 3)$
- Encontre a função de distribuição de X .
- Determine o valor de k tal que $P(x \leq k) = 0.6$
- Calcule os quartis da distribuição.
- Calcule a esperança e a variância de X .

9.6.2. Considere a função:

$$f(x) = \begin{cases} \frac{x}{4} & \text{se } 0 \leq x < 2 \\ 1 - \frac{x}{4} & \text{Se } 2 \leq x \leq 4 \\ 0 & \text{Caso contrário} \end{cases}$$

- Calcule $P(1 \leq x \leq 3)$
- Calcule $E(X)$ e $Var(X)$:
- Encontre a função de distribuição de X .
- Determine o valor de k tal que $P(x \leq k) = 0.6$
- Calcule $P\left(x \leq \frac{5}{4} / \frac{3}{4} \leq x \leq \frac{9}{4}\right)$

9.6.3. A demanda diária de arroz num supermercado, em centenas de quilos, é uma variável aleatória com função densidade de probabilidade

$$f(x) = \begin{cases} \frac{2x}{3} & \text{se } 0 \leq x < 1 \\ 1 - \frac{x}{3} & \text{Se } 1 \leq x \leq 3 \\ 0 & \text{Caso contrário} \end{cases}$$

- Qual é a probabilidade de se vender mais de 150 kg num dia escolhido ao acaso?
- Qual a quantidade de arroz que deve ser deixada à disposição dos clientes diariamente para que não falte arroz em 95% dos dias?

9.6.4. A função de distribuição de uma variável aleatória X é dada por

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{3x^2 - x^3}{2} & \text{Se } 0 \leq x < 1 \\ 1 & \text{Se } x \geq 1 \end{cases}$$

a) Calcule $E(x)$ e $Var(x)$

b) $P(0 \leq x \leq \frac{1}{4} / x \leq \frac{1}{2})$

9.6.5. Considere a seguinte função:

$$f(x) = \begin{cases} k(2x - x^2) & \text{se } 0 \leq x \leq 1 \\ 0 & \text{Caso contrário} \end{cases}$$

Calcule $P(x \leq \frac{1}{2} / \frac{1}{3} \leq x \leq \frac{2}{3})$

9.6.6. Considere a seguinte função:

$$g(x) = \begin{cases} k(2 - x) & \text{se } 0 \leq x \leq 1 \\ 0 & \text{Caso contrário} \end{cases}$$

a) Encontre o valor de K para que $g(x)$ seja função densidade de probabilidade de uma variável aleatória X.

b) Encontre a função de distribuição acumulada.

c) Calcule os quartis da distribuição.

d) Calcule a esperança e a variância de X.

9.6.7. Considere a seguinte função:

$$f(x) = \begin{cases} -0.125x + 0.5 & \text{se } 0 \leq x \leq 4 \\ 0 & \text{Caso contrário} \end{cases}$$

a) Calcule $P(x > 2)$

b) Determine m tal que $P(x > m) = \frac{1}{8}$

c) Calcule a esperança e a variância de X

d) Calcule a função de distribuição acumulada

9.6.8. Uma variável aleatória X tem função densidade de probabilidade dada por:

$$f(x) = \begin{cases} 6x(1 - x) & \text{se } 0 \leq x \leq 1 \\ 0 & \text{Caso contrário} \end{cases}$$

Se $\mu = E(x)$ e $\sigma^2 = Var(x)$, Calcule $P(\mu - 2 * \sigma < x < \mu + 2 * \sigma)$

9.6.9. Uma variável aleatória X tem função de distribuição acumulada F dada por:

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ x^5 & \text{Se } 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Calcule a esperança e a variância de X

9.6.10. Seja X: tempo durante o qual um equipamento eléctrico é usado em carga máxima, num certo período de tempo, em minutos. A função densidade de probabilidade de X é dada por:

$$f(x) = \begin{cases} \frac{1}{1500^2}x & \text{se } 0 \leq x \leq 1500 \\ \frac{1}{1500^2}(-x + 3000) & \text{se } 1500 < x \leq 3000 \\ 0 & \text{c aso contrário} \end{cases}$$

- Encontre a função distribuição de probabilidade;
- Determine o tempo médio em que o equipamento será utilizado em carga máxima;

10. ALGUMAS DISTRIBUIÇÕES CONTÍNUAS

10.1. Distribuição Uniforme

Uma v.a. contínua X tem distribuição uniforme no intervalo $[a; b]$ se sua função densidade é dada por:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{se } x \in [a, b] \\ 0 & \text{caso contrário} \end{cases}$$

Os valores a e b são chamados parâmetros da distribuição uniforme; note que ambos têm de ser finitos para que a integral seja igual a 1. Quando $a = 0$ e $b = 1$ temos a uniforme padrão, denotada por $U(0; 1)$.

10.1.1. Função de distribuição

Por definição, a função de distribuição acumulada é

$$F_X(x) = P(X \leq x)$$

e essa probabilidade é dada pela área sob a curva densidade à esquerda de x , isto é,

$$F(x) = \begin{cases} 0 & \text{se } x < a \\ \frac{x-a}{b-a} & \text{Se } a \leq x \leq b \\ 1 & x > b \end{cases}$$

10.1.2. Esperança e Variância

$$X \sim \text{Unif}(a, b) \rightarrow \begin{cases} E(x) = \frac{a+b}{2} \\ \text{Var}(x) = \frac{(b-a)^2}{12} \end{cases}$$

Exemplo 10.1 (Distribuição Uniforme)

Latas de coca-cola são enchidas num processo automático segundo uma distribuição uniforme no intervalo (em ml) $[345; 355]$.

- a) Qual é a probabilidade de uma lata conter mais de 353 ml?

Solução

Seja v. a $X =$ "conteúdo da lata de coca – cola". Então, $X \sim U[345; 355]$

$$P(x > 353) = \frac{355 - 353}{355 - 345} = 0.2$$

b) Qual é a probabilidade de uma lata conter menos de 346 ml?

Solução

$$P(x < 353) = \frac{346 - 345}{355 - 345} = 0.1$$

c) Qualquer lata com volume 4 ml abaixo da média pode gerar reclamação do consumidor e com volume 4 ml acima da média pode transbordar no momento de abertura, devido à pressão interna. Qual é a proporção de latas problemáticas?

Solução

$$P(x < 350 - 4) + P(x > 350 + 4) = \frac{346 - 345}{355 - 345} + \frac{355 - 354}{355 - 345} = 0.2$$

10.2. Distribuição Exponencial

Diz-se que uma variável aleatória contínua X tem distribuição exponencial com parâmetro λ se sua função densidade de probabilidade é dada por:

$$f(x) = \begin{cases} \lambda * e^{-\lambda x} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases}$$

Usaremos a seguinte notação para indicar que uma variável aleatória tem distribuição exponencial com parâmetro λ : $x \sim \exp(\lambda)$

10.2.1. Função de distribuição

Por definição, temos que:

$$F(x) = P(X \leq x) = \int_0^x f(t) dt = \int_0^x \lambda * e^{-\lambda t} dt = -(e^{-\lambda x} - 1)$$

Ou seja

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases}$$

10.2.2. Esperança e Variância

$$X \sim \text{exp}(\lambda) \rightarrow \begin{cases} E(x) = \frac{1}{\lambda} \\ \text{Var}(x) = \frac{1}{\lambda^2} \end{cases}$$

Exemplo 10.2 (Distribuição Exponencial)

Seja X uma variável aleatória exponencial com média 4. Calcule

a) $P(x > 1)$

Solução

$$P(x > 1) = \int_1^{\infty} \frac{1}{4} * e^{-\frac{x}{4}} dx = 0 - (e^{-0.25}) = 0.7788$$

ou pelo uso directo da função de distribuição

$$P(x > 1) = 1 - P(x \leq 1) = 1 - F(1) = 1 - \left(1 - e^{-\frac{1}{4}}\right) = 0.7788$$

b) $P(1 \leq x \leq 2)$

Solução

$$\begin{aligned} P(1 \leq x \leq 2) &= P(x \leq 2) - P(x < 1) = P(x \leq 2) - P(x \leq 1) \\ &= F(2) - F(1) = \left(1 - e^{-\frac{2}{4}}\right) - \left(1 - e^{-\frac{1}{4}}\right) = 0.17227 \end{aligned}$$

10.3. Distribuição Normal

A distribuição Normal também é conhecida como de Gauss, em referência ao emprego pioneiro dessa distribuição no tratamento dos erros aleatórios de medidas experimentais, atribuído ao matemático alemão Karl Friedrich Gauss (1777-1855).

A distribuição Normal é utilizada para descrever o comportamento de uma variável aleatória que flutua de forma simétrica em torno de um valor central. Algumas de suas propriedades matemáticas, a serem discutidas no presente item, fazem do modelo Normal a distribuição apropriada à modelação de variáveis que resultam da soma de um grande número de outras variáveis independentes. Além disso, a distribuição Normal está na origem de toda a formulação teórica acerca da construção de intervalos de

confiança, testes estatísticos de hipóteses, bem como da teoria de regressão e correlação.

A distribuição Normal é um modelo a dois parâmetros μ e σ , cujas funções densidade e de probabilidades acumuladas são expressas, respectivamente, por:

$$f_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} * e^{\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]}$$
$$F(x) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} * e^{\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]} dx$$

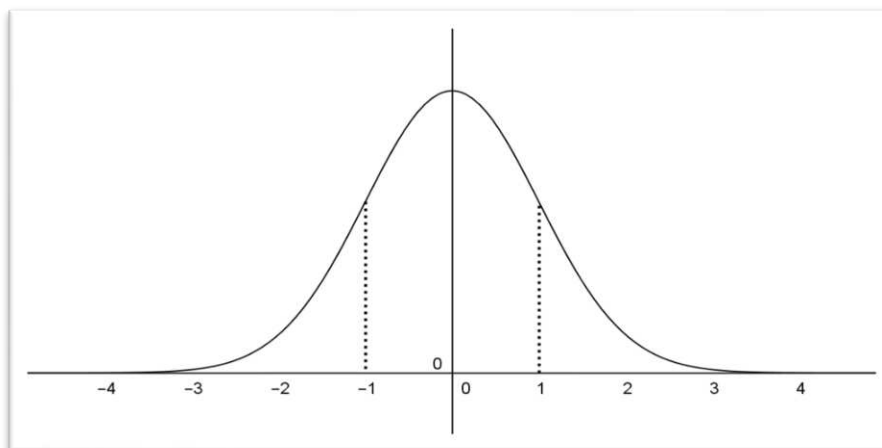
Diz-se que X é normalmente distribuída com média μ e desvio padrão σ , ou, sinteticamente, que $X \sim N(\mu, \sigma^2)$. Portanto, a média de uma variável Normal X é igual ao parâmetro de posição, em torno do qual os valores de X se dispersam simetricamente. O grau com que a variável X se dispersa em torno de μ , é dado pelo parâmetro de escala, o qual é igual ao desvio padrão σ .

10.3.1. Densidade normal padrão

A variável aleatória Z tem distribuição normal padrão se sua função densidade é dada por:

$$f(Z) = \frac{1}{\sqrt{2\pi}} * e^{-\frac{z^2}{2}}$$

para todo $z \in R$. Vamos denotar por $N(0; 1)$ a densidade normal padrão e, se uma variável aleatória Z é distribuída segundo uma normal padrão, representaremos esse fato como $Z \sim N(0; 1)$.



10.3.2. Esperança e Variância

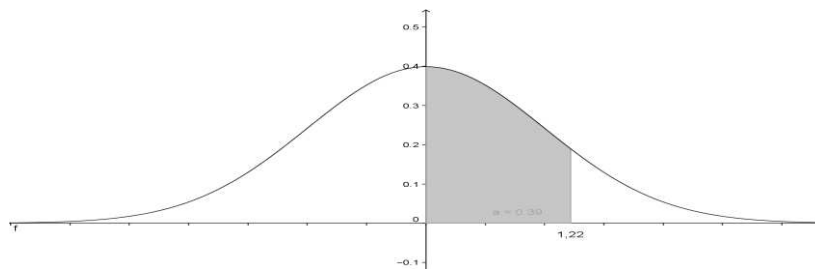
$$Z \sim N(0; 1) \rightarrow \begin{cases} E(Z) = 0 \\ \text{Var}(Z) = 1 \end{cases}$$

10.3.3. Tabela Normal Padrão

Vimos anteriormente que o cálculo de probabilidades associadas a variáveis aleatórias contínuas envolve cálculo de áreas sob a curva densidade (mais precisamente, cálculo de integral da fdp). Isso, obviamente, continua valendo para a densidade normal. A diferença está no fato de que o cálculo de áreas sob a curva normal envolve métodos numéricos mais complexos e, para facilitar esses cálculos, podemos usar uma tabela em que alguns valores já se encontram calculados.

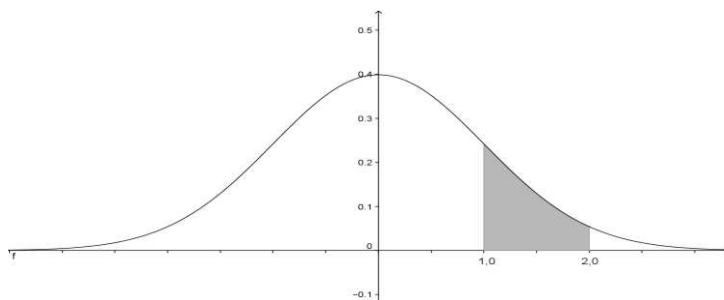
Exemplo 10.3 (Distribuição Normal)

$$P(0 \leq Z \leq 1.22) = \text{tab}(1.22) = 0.3888$$



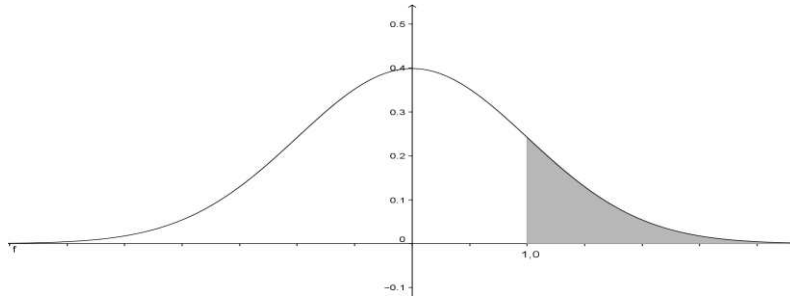
Exemplo 10.4 (Distribuição Normal)

$$\begin{aligned} P(1 \leq Z \leq 2) &= P(0 \leq Z \leq 2) - P(0 \leq Z \leq 1) = \text{tab}(2.0) - \text{tab}(1.0) \\ &= 0.4772 - 0.3413 = 0.1359 \end{aligned}$$



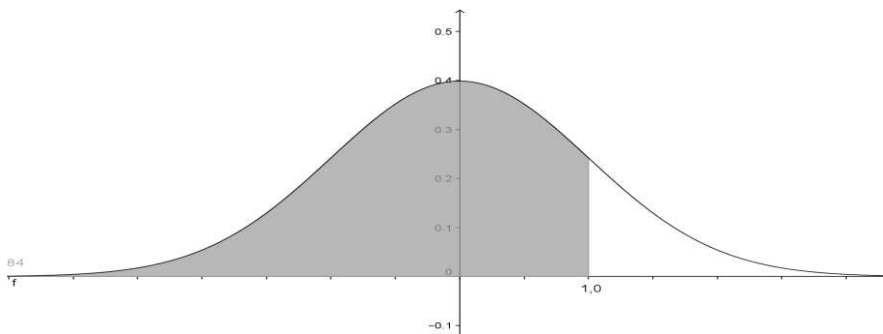
Exemplo 10.5 (Distribuição Normal)

$$P(Z \geq 1) = 0.5 - P(0 \leq Z \leq 1) = 0.5 - \text{tab}(1.0) = 0.5 - 0.3413 = 0.1587$$



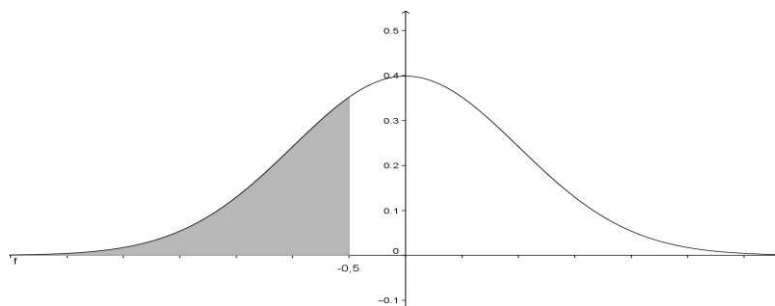
Exemplo 10.6 (Distribuição Normal)

$$P(Z \leq 1) = P(Z \leq 0) + P(0 \leq Z \leq 1) = 0.5 + tab(1.0) = 0.5 + 0.3413 = 0.8413$$



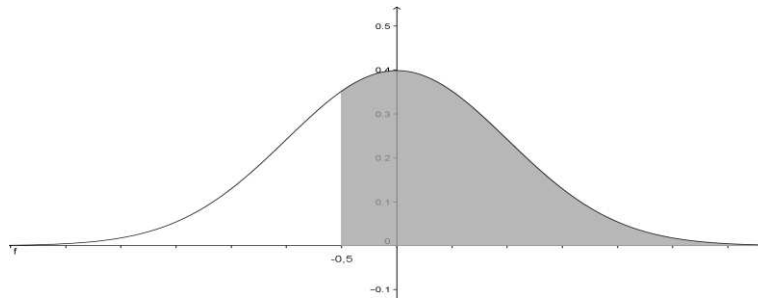
Exemplo 10.7 (Distribuição Normal)

$$\begin{aligned} P(Z \leq -0.5) &= P(Z \geq 0.5) = 0.5 - P(0 \leq Z \leq 0.5) = 0.5 - tab(0.5) \\ &= 0.5 - 0.1915 = 0.3085 \end{aligned}$$



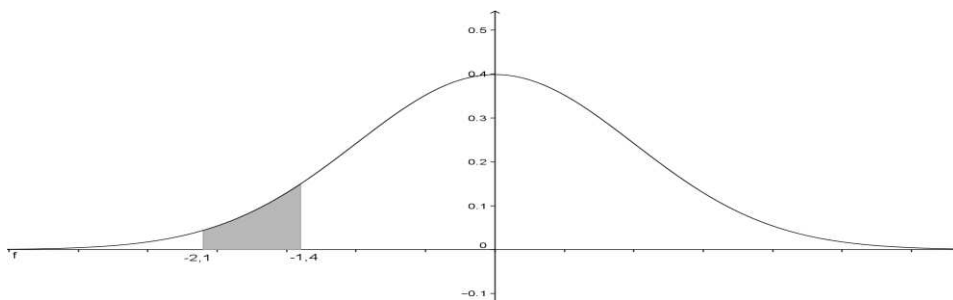
Exemplo 10.8 (Distribuição Normal)

$$\begin{aligned} P(Z \geq -0.5) &= P(-0.5 \leq Z \leq 0) + P(Z \geq 0) = P(0 \leq Z \leq 0.5) + 0.5 \\ &= tab(0.5) + 0.5 = 0.1915 + 0.5 = 0.6915 \end{aligned}$$



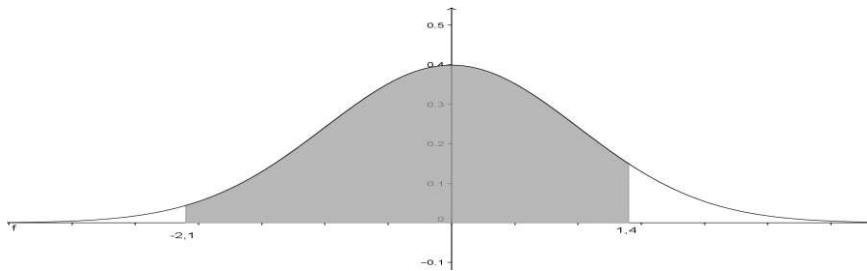
Exemplo 10.9 (Distribuição Normal)

$$P(-2.1 \leq Z \leq -1.4) = P(1.4 \leq Z \leq 2.1) = P(0 \leq Z \leq 2.1) - P(0 \leq Z \leq 1.4) \\ = tab(2.1) - tab(1.4) = 0.4821 - 0.4192 = 0.0629$$



Exemplo 10.10 (Distribuição Normal)

$$P(-2.1 \leq Z \leq 1.4) = P(0 \leq Z \leq 1.4) + P(-2.1 \leq Z \leq 0) \\ = P(0 \leq Z \leq 1.4) + P(0 \leq Z \leq 2.1) = tab(1.4) + tab(2.1) \\ = 0.4821 + 0.4192 = 0.9013$$



10.3.4. Cálculos com a distribuição normal

Nesta seção serão apresentados resultados básicos sobre a distribuição normal, que permitirão que você calcule probabilidades associadas a qualquer variável aleatória normal, e isso ampliará o escopo de aplicações práticas.

Na seção anterior, você viu como usar tabelas da distribuição normal padrão para calcular probabilidades associadas à variável $Z \sim N(0; 1)$. Essas tabelas, ou softwares especializados, são necessários para fazer os cálculos, pois não existem métodos

directos para calcular áreas sob a curva da densidade normal padrão. Mas as tabelas vistas referiam-se à distribuição $N(0; 1)$.

Padronização da distribuição normal

Se $X \sim N(\mu; \sigma^2)$, então:

$$Z = \frac{X - \mu}{\sigma}$$

Tem a distribuição $N(0; 1)$

Exemplo 10.11 (Distribuição Normal)

Suponhamos que $X \sim N(3; 9)$, determine $P(-1 \leq X \leq 4)$

Solução

$$\begin{aligned} P(-1 \leq X \leq 4) &= P\left(\frac{-1 - 3}{\sqrt{9}} \leq \frac{X - 3}{\sqrt{9}} \leq \frac{4 - 3}{\sqrt{9}}\right) = P(1.33 \leq Z \leq 0.33) \\ &= \text{tab}(0.33) + \text{tab}(1.33) = 0.12930 + 0.40824 = 0.53754 \end{aligned}$$

10.3.5. Encontrando a abcissa da normal para uma probabilidade específica

Nos exemplos vistos até o momento, consideramos situações em que tínhamos uma abcissa de uma distribuição normal e queríamos a probabilidade associada a essa abcissa.

Agora, vamos lidar com a situação inversa: dada uma probabilidade, qual é a abcissa correspondente? Eis algumas situações que envolvem esse tipo de problema:

- Em uma turma de Estatística, os 10% melhores alunos receberão um livro de presente;
- Em uma comunidade, as famílias com as 15% piores rendas irão receber um auxílio da prefeitura.

Como antes, vamos apresentar vários exemplos que ilustram essa situação.

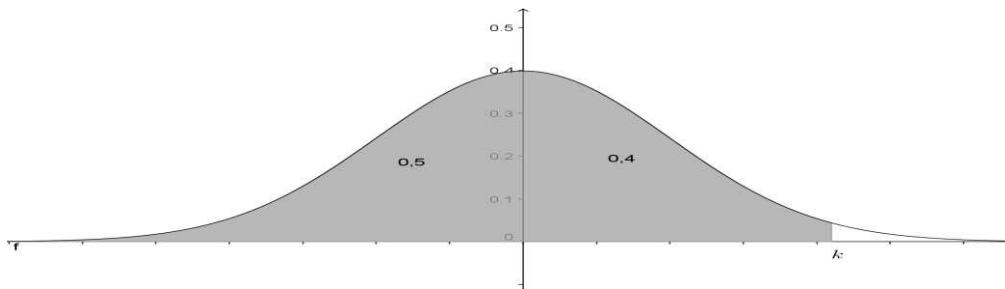
Exemplo 10.12 (Distribuição Normal)

Suponhamos que $Z \sim N(0; 1)$

Determine o valor de k tal que $P(Z \leq k) = 0.90$

Solução

Vamos “traduzir” esse problema: queremos encontrar a abscissa k da normal padrão com 0,90 de área (probabilidade) à esquerda dela. Como 0,9 é a área à esquerda de k , resulta que k tem que ser maior que zero, isto é, temos que ter $k > 0$. Veja a Figura abaixo: à esquerda de k temos área 0,90 e à esquerda de 0 temos área 0,5. Logo, entre 0 e k temos que ter área 0,40.



Escrevendo essas observações em termos de probabilidade, temos:

$$P(Z \leq k) = 0.90 \leftrightarrow P(Z \leq 0) + P(0 < Z \leq k) = 0.90 \leftrightarrow$$

$$0.5 + P(0 < Z \leq k) = 0.90 \leftrightarrow$$

$$P(0 < Z \leq k) = 0.40$$

$$tab(k) = 0.40$$

Esta última igualdade nos diz que k é a abscissa correspondente ao valor 0,40 na tabela. Para identificar k ; temos que buscar no corpo dessa tabela, o valor mais próximo de 0,40. Na linha correspondente ao valor 1,2 encontramos as entradas 0,39973 e 0,40147. Como a primeira está mais próxima de 0,40, olhamos qual é a abscissa correspondente: a linha é 1,2 e a coluna é 8, o que nos dá a abscissa de 1,28, ou seja, $k = 1,28$ e $P(Z \leq 1,28) = 0,90$, completando a solução.

Exemplo 10.13 (Distribuição Normal)

Suponhamos que $X \sim N(3; 4)$

Determine o valor de k tal que $P(X \leq k) = 0.90$

Solução

Com a probabilidade à esquerda de k é maior que 0,5, resulta que k tem de ser maior que a média. O primeiro passo na solução é escrever a probabilidade dada em termos da normal padrão.

$$P(X \leq k) = 0.90 \leftrightarrow P\left(\frac{X - 3}{2} \leq \frac{k - 3}{2}\right) = 0.90 \leftrightarrow$$

$$\begin{aligned}
P\left(Z \leq \frac{k-3}{2}\right) &= 0.90 \\
P(Z \leq 0) + P\left(0 \leq Z \leq \frac{k-3}{2}\right) &= 0.90 \\
P\left(0 \leq Z \leq \frac{k-3}{2}\right) &= 0.40 \\
tab\left(\frac{k-3}{2}\right) &= 0.40 \\
\frac{k-3}{2} = 1.28 &\leftrightarrow k = 5.56
\end{aligned}$$

10.3.6. Exemplos de Aplicação da Distribuição Normal

A distribuição normal é um modelo probabilístico que se aplica a diversas situações práticas. Vamos finalizar este capítulo com alguns exemplos práticos, mas, na última parte do curso, você verá mais aplicações no contexto da inferência estatística, em que decisões têm de ser tomadas com base nos resultados obtidos a partir de uma amostra.

Exemplo 10.14 (Distribuição Normal)

Uma fábrica de carros sabe que os motores de sua fabricação têm duração normal com média de 150000 km e desvio-padrão de 5000 km. Qual a probabilidade de que um carro, escolhido ao acaso, dos fabricados por essa firma, tenha um motor que dure:

- a) Menos de 170000 km?

Solução

Do problemas temos:

$$\mu = 150 \text{ e } \sigma = 5$$

$$x \sim N(\mu = 150000; \sigma^2 = 5000^2) \rightarrow Z \sim N(\mu = 0; \sigma = 1), \text{ onde } Z = \frac{x - \mu}{\sigma}$$

$$\begin{aligned}
P(X < 170000) &= P\left(Z < \frac{170000 - 150000}{5000}\right) = P(Z < 4) \\
&= 0.5 + P(0 \leq Z \leq 4) = 0.5 + tab(4.0) = 0.5 + 0.49998 = 0.99998
\end{aligned}$$

- b) Entre 140000 km e 165000 km?

Solução

$$\begin{aligned}
&P(140000 < X < 165000) \\
&= P\left(\frac{140000 - 150000}{5000} < Z < \frac{165000 - 150000}{5000}\right)
\end{aligned}$$

$$\begin{aligned}
&= P(-2 < Z < 3) = P(-2 < Z < 0) + P(0 < Z < 3) \\
&= \text{tab}(-2.0) + \text{tab}(3.0) \\
&= 0.4772 + 0.4987 \\
&= 0.9759
\end{aligned}$$

- c) Se a fábrica substitui o motor que apresenta duração inferior à garantia, qual deve ser esta garantia para que a percentagem de motores substituídos seja inferior a 0,2%?

Solução

$$P(X < k) = 0.002$$

$$Z_{0.002} = -2.28$$

$$\frac{k - 150000}{5000} = -2.88$$

$$k = 150000 - 2.88 * 5000 = 135600$$

10.4. EXERCÍCIOS PROPOSTOS

10.4.1. Uma variável aleatória X tem distribuição uniforme no intervalo $[a; b]$ com média 7,5 e variância 6,75. Determine os valores de a e b , sabendo que $b > a > 0$.

10.4.2. Você está interessado em dar uma proposta em um leilão de um lote de terra. Você sabe que existe um outro licitante. Pelas regras estabelecidas para este leilão, a proposta mais alta acima de 100.000,00 u.m será aceite. Suponha que a proposta do seu competidor seja uma variável aleatória uniformemente distribuída entre 100.000,00 u.m e 150.000,00 u.m. Supondo que tens dois conselheiros identificados por A e B. O conselheiro A propõe lhe uma proposta de 120.000,00 u.m, enquanto que o B uma proposta de 140.000,00 u.m. Você optaria por proposta do conselheiro A ou B?

10.4.3. A LAM apresenta um tempo de voo de 2 horas e 5 minutos para seus voos de Maputo a Pemba. Suponha que acreditamos que os tempos de voo reais sejam uniformemente distribuídos entre 2 horas e 2 horas e 20 minutos.

- a) Qual é a probabilidade de que o voo terá não mais que cinco minutos de atraso?
- b) Qual é a probabilidade de que o voo terá não mais que 10 minutos de atraso?
- c) Qual é a expectativa do tempo de voo?

10.4.4. A distância de impulso para as 60 principais golfistas no torneio LPGA está entre 238,9 e 261,2 jardas (*Golfweek*, 6 de Dezembro de 1997). Considere que a distância de impulso para essas mulheres está uniformemente distribuída nesse intervalo.

- a) Qual é a probabilidade de que a distância de impulso para uma dessas mulheres seja de menos de 250 jardas?
- b) Qual é a probabilidade de que a distância de impulso para uma dessas mulheres seja de pelo menos 255 jardas?
- c) Qual é a probabilidade de que a distância de impulso para uma dessas mulheres esteja entre 245 e 260 jardas?
- d) Quantas dessas mulheres impulsionam a bola pelo menos 250 jardas?

10.4.5. Latas de coca-cola são enchidas num processo automático segundo uma distribuição uniforme no intervalo (em ml) [345;355].

- a) Qual é a probabilidade de uma lata conter mais de 353 ml?
- b) Qual é a probabilidade de uma lata conter menos de 346 ml?
- c) Qualquer lata com volume 4 ml abaixo da média pode gerar reclamação do consumidor e com volume 4 ml acima da média pode transbordar no momento de abertura, devido à pressão interna. Qual é a proporção de latas problemáticas?

10.4.6. A dureza X de uma peça de aço pode ser pensada como sendo uma variável aleatória uniforme no intervalo (50,70) da escala Rockwel. Qual é a probabilidade de que uma peça tenha dureza entre 55 e 60?

10.4.7. Seja $X \sim N(3; 4)$. Determine o valor de k tal que:

- a) $P(x \leq k) = 0.90$
- b) $P(x \leq k) = 0.05$
- c) $P(|x - 3| \leq k) = 0.95$

10.4.8. Seja $X \sim N(100; 25)$. Determine:

- a) $P(100 \leq X \leq 106)$
- b) $P(89 \leq X \leq 107)$
- c) $P(112 \leq X \leq 116)$
- d) $P(X \geq 108)$

10.4.9. O consumo mensal em minutos por conta de celular em uma região é uma variável aleatória normal com média 36 e desvio padrão 12.

- a) Qual é a probabilidade de uma pessoa desta região usar o telefone celular por menos de 48 minutos?
- b) Qual é a probabilidade de uma pessoa desta região usar o telefone celular por mais de 30 minutos?
- c) Qual o tempo mínimo que alguém deve gastar ao telefone no mês para estar entre os 5% que MAIS usam o celular?
- d) Qual o tempo máximo que alguém deve gastar ao telefone no mês para estar entre os 12% que MENOS usam o celular?

10.4.10. Uma máquina de empacotar determinado produto oferece variações de peso que se distribuem segundo uma distribuição normal com desvio padrão de 20 gramas. Em quanto deve ser regulado o peso médio desses pacotes para que apenas 10% deles tenham menos que 500 gramas?

10.4.11. O saldo médio dos clientes de um banco é uma v.a. normal com média 2.000,00 u.m e desvio-padrão 250,00 u.m. Os clientes com os 10% maiores saldos médios recebem tratamento VIP, enquanto aqueles com os 5% menores saldos médios receberão propaganda extra para estimular maior movimentação da conta.

- a) Quanto você precisa de saldo médio para se tornar um cliente VIP?
- b) Abaixo de qual saldo médio o cliente receberá a propaganda extra?

10.4.12. Uma máquina de empacotar determinado produto oferece variações de peso que se distribuem segundo uma distribuição normal com desvio padrão de 20 gramas. Em quanto deve ser regulado o peso médio desses pacotes para que apenas 10% deles tenham menos que 500 gramas?

10.4.13. Uma máquina fabrica tubos metálicos cujos diâmetros podem ser considerados uma variável aleatória normal com média 200mm e desvio-padrão 2mm. Verifica-se que 15% dos tubos estão sendo rejeitados como grandes e 10% como pequenos.

- a) Quais são as tolerâncias de especificação para esse diâmetro?
- b) Mantidas essas especificações, qual deverá ser a regulagem média da máquina para que a rejeição por diâmetro grande seja nula? Nesse caso, qual será a percentagem de rejeição por diâmetro pequeno?

10.4.14. Em um grande complexo industrial, o departamento de manutenção tem instruções para substituir as lâmpadas antes que se queimem. Os registros indicam que a duração das lâmpadas, em horas, tem distribuição normal, com média de 900 horas e desvio-padrão de 75 horas. Quando devem ser trocadas as lâmpadas, de modo que no máximo 5% delas queimem antes de serem trocadas?

10.4.15. Suponha que os tempos de vida de 2 marcas de aparelhos elétricos sejam variáveis aleatórias D_1 e D_2 , onde $D_1 \sim N(42; 36)$ e $D_2 \sim N(45; 9)$. Se o aparelho deve ser usado por um período de 45 horas, qual marca deve ser preferida? E se for por um período de 49 horas?

10.4.16. As vendas de um determinado produto têm distribuição aproximadamente normal com média de 500 unidades e desvio padrão de 50 unidades. Se a empresa decide fabricar 600 unidades no mês em estudo, qual a probabilidade de que não possa atender a todos os pedidos desse mês, por estar com a produção esgotada?

10.4.17. Um produto alimentício é ensacado automaticamente, sendo o peso médio de 50 kg por saco, com desvio padrão de 1,6 kg. Os clientes exigem que, para cada saco fornecido com menos de 48 kg, o fornecedor pague uma indenização de 5 u.m.

- a) Para 200 sacos fornecidos, qual o custo médio com indenização?
- b) Para que o custo calculado no item anterior caia para 50 u.m., qual deveria ser a nova regulagem média da máquina?
- c) Como o fornecedor acha que, no custo global, é desvantajoso aumentar a regulagem da máquina, ele quer comprar uma nova máquina. Qual

deveria ser o desvio padrão dessa máquina para que, trabalhando com peso médio de 50 kg, em apenas 3% dos sacos se pague indenização?

10.4.18. Um teste de aptidão para o exercício de uma certa profissão exige uma sequência de operações a serem executadas rapidamente uma após a outra. Para passar no teste, o candidato deve completá-lo em, no máximo, 80 minutos. Admita que o tempo, em minutos, para completar a prova seja uma variável aleatória normal com média 90 minutos e desvio padrão 20 minutos.

- a) Que percentagem dos candidatos tem chance de ser aprovada?
- b) Os 5% melhores receberão um certificado especial. Qual o tempo máximo para fazer jus a tal certificado?

10.4.19. Uma empresa produz televisores e garante a restituição da quantia paga se qualquer televisor apresentar algum defeito grave no prazo de 6 meses. Ela produz televisores do tipo A comum e do tipo B de luxo, com um lucro respectivo de 1000 u.m. e 2000 u.m. caso não haja restituição, e com prejuízo de 3000 u.m. e 8000 u.m., se houver restituição. Suponha que o tempo para ocorrência de algum defeito grave seja, em ambos os casos, uma variável aleatória com distribuição normal com médias 9 meses e 12 meses e desvios padrões 2 meses e 3 meses. Se tivesse que planejar uma estratégia de marketing para a empresa, você incentivaria as vendas dos aparelhos tipo A ou tipo B?

10.4.20. Um fabricante de baterias sabe, por experiência passada, que as baterias de sua fabricação têm vida média de 600 dias e desvio-padrão de 100 dias, sendo que a duração tem aproximadamente distribuição normal. Oferece uma garantia de 312 dias, isto é, troca as baterias que apresentarem falhas nesse período. Fabrica 10000 baterias mensalmente. Quantas deverá trocar pelo uso da garantia, mensalmente?

10.4.21. Uma fábrica de carros sabe que os motores de sua fabricação têm duração normal com média de 150000 km e desvio-padrão de 5000 km. Qual a probabilidade de que um carro, escolhido ao acaso, dos fabricados por essa firma, tenha um motor que dure:

- a) Menos de 170.000 km?

- b) Entre 140000 km e 165000 km?
- c) Se a fábrica substitui o motor que apresenta duração inferior à garantia, qual deve ser esta garantia para que a percentagem de motores substituídos seja inferior a 0,2%?

10.4.22. Impostos pagos por uma grande amostra de contribuintes distribuem-se normalmente de tal forma que 30% são inferiores a 1.200,00 u.m e 10% são superiores a 3.000,00 u.m. Pede-se determinar o imposto médio.

10.4.23. O tempo de vida de transistores produzidos pela Indústria Zeppelin Ltda. tem distribuição aproximadamente normal, com valor esperado e desvio-padrão igual a 500 horas e 50 horas, respectivamente. Se o consumidor exige que pelo menos 95% dos transistores fornecidos tenham vida superior a 400 horas, pergunta-se se tal especificação é atendida. Justifique!

10.4.24. Há dois procedimentos para possibilitar que um determinado tipo de avião esteja pronto para a decolagem. O procedimento A requer um tempo médio de 27 minutos com desvio padrão de 5 minutos. Para o procedimento B, tempo médio de 30 minutos com desvio padrão de 2 minutos, respectivamente. Qual procedimento deve ser utilizado se o tempo disponível é de 30 minutos? 34 minutos?

10.4.25. Os prazos de substituição para CD players têm distribuição normal com média de 7,1 anos e desvio-padrão de 1,4 ano (com base em dados do "Getting Things Fixed", *Consumer Reports*). Determine a probabilidade de um CD player escolhido aleatoriamente ter um prazo de substituição inferior a 8,0 anos.

10.4.26. Uma aplicação clássica da distribuição normal é inspirada em uma carta a Dear Abby, em que uma esposa alegava ter dado a luz 308 dias após uma rápida visita de seu marido que estava servindo na Marinha. Os prazos da gravidez têm distribuição normal com média de 268 dias e desvio-padrão de 15 dias. Com base nessa informação, determine a probabilidade de uma gravidez durar 308 dias ou mais. Que é que o resultado sugere?

- 10.4.27.** De acordo com a *Opinion Research Corporation*, os homens gastam em média 11,4 minutos no chuveiro. Suponha que esses tempos tenham distribuição normal com desvio-padrão de 1,8 min. Escolhido um homem aleatoriamente, determine a probabilidade de ele gastar ao menos 10,0 min no chuveiro.
- 10.4.28.** Um subfornecedor da IBM foi contratado para fabricar substratos de cerâmica, utilizados para transmitir sinais entre chips de silício para computador. As especificações exigem uma resistência entre 1,500 ohm e 2,500 ohms, mas a população tem resistências distribuídas normalmente com média de 1,978 ohm e desvio-padrão de 0,172 ohm. Que percentagem dos substratos de cerâmica foge às especificações do fabricante? Esse processo de fabricação parece estar funcionando bem?
- 10.4.29.** Você pode escolher entre 2 empregos. Em uma indústria seus ganhos mensais terão distribuição normal com média de 4000 Mt e desvio padrão de 500 Mt. Como vendedor de uma firma seus ganhos mensais terão distribuição normal com média de 3200 Mt e desvio padrão de 2600 Mt.
- Você ganha actualmente (salário fixo) 3500 Mt. Qual é a probabilidade de ganhar mais nos dois possíveis empregos?
 - Com base no resultado do item a, qual dos dois empregos você escolheria?

11. DISTRIBUIÇÃO AMOSTRAL E ESTIMAÇÃO PONTUAL DE PARÂMETROS

11.1. Inferência Estatística

O campo da inferência estatística consiste naqueles métodos usados para tomar decisões ou tirar conclusões acerca de uma **população**. Esses métodos utilizam a informação contida em uma **amostra** da população para extrair conclusões.

A inferência estatística pode ser dividida em duas grandes áreas: **estimação de parâmetros e testes de hipóteses**.

11.2. Técnicas de Amostragem (Revisão)

São procedimentos a ser adoptado na selecção dos elementos da amostra e tem como objectivo central obter uma amostra representativa.

Uma amostra é representativa se representa toda a população da melhor maneira possível. Essa representatividade depende de:

- Metodologia adoptada para a selecção da amostra
- Tamanho da amostra

11.3. Planeando um experimento

No planeamento de um experimento é necessário que sejam:

- Identificados os objectivos;
- Colectados dados amostrais
- Usados procedimentos aleatórios para evitar vícios
- Analisados dados e tirar conclusões

11.4. Erro Amostral

É a diferença entre um resultado amostral e o verdadeiro resultado populacional.

Decorre da própria noção de amostra. Quando se recolhe uma amostra, alguma coisa se perde da população de onde foi retirada, pelo que, embora cuidadosamente recolhida, uma amostra pode não ser representativa da população.

Do mesmo modo, não se pode esperar que duas amostras, independentemente retiradas da mesma população, forneçam resultados iguais. Porque existe esta variabilidade nas estimativas e porque a amostra não é uma perfeita representação da população, os resultados que ela fornece são de alguma forma errados.

O erro amostral pode ser controlável com acções do tipo:

- Técnica de amostragem - optando por aquela que, no caso concreto, se revela mais eficiente; mediante a escolha de um processo de amostragem aleatório e do aumento do tamanho da amostra, pode-se assegurar a representatividade e associar os resultados com grau de confiança elevado.
- Estimadores - optando por aquele que seja mais eficiente, isto é, com menor variabilidade.

O erro amostral é um erro aleatório, pois as estimativas comportam-se aleatoriamente em torno do verdadeiro valor do parâmetro. Ou seja, não coincidem com o parâmetro, estando umas estimativas acima e outras abaixo deste, mas concentram-se em torno de um valor central que coincide com o verdadeiro valor do parâmetro.

11.5. Erro Não-Amostral

É a incorrecção na colecta, registo ou análise de dados amostrais, isto é:

- Colecta tendeciosa de amostra
- Utilização de instrumento descalibrado
- Registo incorrecto de dados amostrais

11.6. Estimação de Parâmetros

Estimação é o processo que consiste em utilizar dados amostrais para estimar valores para a média e o desvio padrão de uma população e a proporção populacional de parâmetros populacionais desconhecidos. Essencialmente, qualquer característica de uma população pode ser estimada a partir de uma amostra aleatória. Entre os mais comuns as estatísticas amostrais são utilizadas como estimativas de parâmetros populacionais que podem ser classificadas em pontual ou intervalar.

11.7. Conceitos Fundamentais

11.7.1. Amostra aleatória

As variáveis aleatórias $X_1, X_2, X_3, \dots, X_n$ são uma amostra aleatória de tamanho n se:

- Forem independentes
- Cada X_i tiver mesma distribuição de probabilidades

11.7.2. Parâmetro

É a quantidade de interesse da população, onde em geral são desconhecidas a média (μ) e desvio padrão (σ) dessa população.

$$\mu = \frac{\sum_{i=1}^n X_i}{N} \text{ e } \sigma^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{N}$$

11.7.3. Estatísticas

É qualquer função da amostra que não depende de parâmetros desconhecidos

11.1. Exemplo (Estatísticas de amostra aleatória)

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$
$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

11.7.4. Distribuição amostral

É a distribuição de probabilidades de uma medida estatística baseada em uma amostra aleatória. Distribuições amostrais são importantes porque fornecem uma grande simplificação, usada para inferência estatística. Mais especificamente, elas permitem considerações analíticas serem baseadas na distribuição amostral de uma estatística, em vez de na distribuição conjunta.

O conceito de distribuição de probabilidade de uma variável aleatória será agora utilizado para caracterizar a distribuição dos diversos valores de uma variável em uma população.

Ao retirar uma amostra aleatória de uma população estaremos considerando cada valor da amostra como um valor de uma variável aleatória cuja distribuição de

probabilidade é a mesma da população no instante da retirada desse elemento para a amostra.

Em consequência do fato de os valores de amostra serem aleatórios, decorre que qualquer quantidade calculada em função dos elementos da amostra também será uma variável aleatória.

A distribuição amostral de uma estatística depende da distribuição da população, do tamanho da amostra e do método de selecção da amostra.

11.2 Exemplo (Distribuição amostral da média)

$$E(\bar{X}) = \mu_{\bar{x}}$$

11.3 Exemplo (Parâmetros da distribuição amostral da média)

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

11.7.5. Espaço paramétrico

Um parâmetro (θ) estatístico é uma função definida sobre os valores numéricos de uma população. Trata-se, portanto, de um valor representativo que permite modelizar a realidade.

A utilidade dos parâmetros estatísticos prende-se com a dificuldade de trabalhar com uma grande quantidade de dados individuais de uma mesma população. Este tipo de parâmetros permite obter um panorama geral da população e realizar comparações e previsões.

Exemplo 11.4 (Parâmetros)

Seja a amostra aleatórias $X_1, X_2, X_3, \dots, X_n$ da variável $X \sim N(\mu; \sigma^2)$

- Se $\sigma^2 = 0$ então $\theta = \mu$ é o parâmetro desconhecido e $\theta = \{\mu, -\infty < \mu < \infty\}$
- Se $\mu = 0$, então $\theta = \sigma^2$ é o parâmetro desconhecido e $\theta = \{\sigma^2, \sigma^2 > 0\}$

11.7.6. Estimador

Um estimador ($\hat{\theta}$) é uma regra para calcular uma estimativa de uma determinada quantidade baseada em dados observados: assim a regra e seu resultado (a estimativa) são distinguidos.

Um "estimador " ou "ponto estimado" é uma estatística (isto é, uma função dos dados) que é utilizada para inferir o valor de um parâmetro desconhecido em um modelo estatístico. O parâmetro a ser estimado por vezes é chamado estimando. Ele pode ser de dimensão finita (no paramétrico e modelo semi-paramétrico), ou de dimensão infinita (não semi-paramétrico e modelo não paramétrico).

Sendo uma função dos dados, o estimador é em si uma variável aleatória, uma realização particular desta variável aleatória é chamada "estimativa". Às vezes, as palavras "estimador" e "estimativa" são usados alternadamente. A definição coloca, praticamente sem restrições, sobre quais funções dos dados podem ser chamadas de "estimadores ". A atratividade de diferentes estimadores pode ser julgada ao olhar para as suas propriedades, tais como viés, erro quadrático médio, consistência, distribuição assintótica, etc.. A construção e comparação de estimadores são os temas da teoria da estimação. No contexto da teoria da decisão, um estimador é um tipo de regra de decisão, e seu desempenho pode ser avaliado através do uso de funções de perda.

Os estimadores podem ser:

- **Estimativa de parâmetro populacional:** é um valor específico, ou um intervalo de valores, usado para estimar parâmetro populacional.
- **Estimativa pontual:** é um valor numérico de uma estatística $\hat{\theta}$.

11.8. Teorema do Limite Central

O Teorema central do limite é um importante resultado da estatística e a demonstração de muitos outros teoremas estatísticos dependem dele. Em teoria das probabilidades, esse teorema afirma que quando o tamanho da amostra aumenta, a distribuição amostral da sua média aproxima-se cada vez mais de uma distribuição normal. Este resultado é fundamental na teoria da inferência estatística.

Na inferência estatística a utilidade do teorema central do limite vai desde estimar os parâmetros como a média populacional ou o desvio padrão da média populacional, a partir de uma amostra aleatória dessa população, ou seja, da média amostral e do desvio padrão da média amostral até calcular a probabilidade de um parâmetro ocorrer dado um intervalo, sua média amostral e o desvio padrão da média amostral.

Seja $X_1, X_2, X_3, \dots, X_n$ uma amostra aleatória de tamanho n de uma população (finita ou infinita), com média μ e variância finita σ^2 . Então:

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ quando } n \rightarrow \infty$$

Comentários Gerais

- A aproximação normal para a média amostral depende do tamanho da amostra;
- Com população contínua, unimodal e simétrica, na maioria dos casos, o teorema de limite central trabalha bem para pequenas amostras ($n = 4$ ou $n = 5$);
- Em muitos casos de interesse prático, a aproximação normal será satisfatória para $n \geq 30$;
- Se $n < 30$, o teorema de limite central funcionará se a distribuição da população não for muito diferente da normal

11.9. Propriedades de um estimador

Os estimadores devem ser escolhidos de forma adequada, isto é, devem apresentar as seguintes características:

- Não Viciado
- Consistente
- Eficiente

11.9.1. Estimador não viciado ou não tendencioso

Um estimador $\hat{\theta}$ é não viciado (não viesado ou não tendencioso) para um parâmetro θ se

$$E(\hat{\theta}) = \theta$$

A esperança de um estimador está relacionada com sua **exactidão**.

Exemplo 11.5 (Estimador não viciado)

- A média amostral é não viesada para estimar a média verdadeira (populacional):

$$E(\bar{X}_n) = \mu_x$$

- X_1 (primeiro item colectado da amostra) é não viciado para estimar a verdadeira média

$$E(X_1) = \mu_x$$

- A variância amostral é não viciada para estimar a variância populacional (σ^2)?

$$\begin{aligned} E(S^2) &= \frac{1}{n-1} * \sum_{i=1}^n E[(x_i - \bar{x})^2] = \frac{1}{n-1} * E \left[\sum_{i=1}^n (x_i - \mu + \mu - \bar{x})^2 \right] \\ &= \frac{1}{n-1} * E \left[\sum_{i=1}^n (x_i - \bar{x})^2 \right] - \frac{1}{n-1} * E[n(\bar{x} - \mu)^2] \\ &= \frac{1}{n-1} * \sum_{i=1}^n E[(x_i - \mu)^2] - \frac{n}{n-1} * E[(\bar{x} - \mu)^2] \\ &= \frac{n}{n-1} * \sigma^2 - \frac{1}{n-1} * \sigma^2 = \sigma^2 \end{aligned}$$

11.9.1.1. Variância de um estimador

$\hat{\theta}_1$ e $\hat{\theta}_2$ estimadores não viciados de θ de variâncias diferentes.

Se $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$ é mais provável que $\hat{\theta}_1$ produza uma estimativa mais próxima do valor verdadeiro de θ .

11.9.1.2. Estimador de variância mínima

Se considerarmos todos os estimadores não tendenciosos de θ , aquele com menor variância será chamado de estimador não tendencioso de variância mínima.

Este estimador é o mais provável, dentre todos os não viciados, para produzir uma estimativa que seja próxima do valor verdadeiro.

11.9.1.3. Erro padrão

O erro padrão de um estimador $\hat{\theta}$ é o seu desvio padrão:

$$\sigma_{\hat{\theta}} = \sqrt{Var(\hat{\theta})}$$

- O erro padrão (ou de variância) do estimador está relacionado com sua precisão.

- Se o erro padrão envolver parâmetros desconhecidos que possam ser estimados, então a substituição daqueles valores produz um erro padrão estimado.

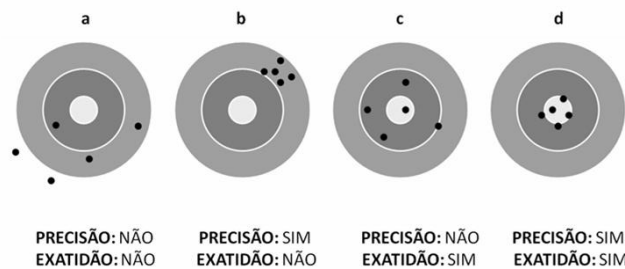
✓ O erro padrão da média amostral é:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- ✓ Se não conhecemos σ , mas substituimos pelo desvio padrão amostral, então o erro padrão amostral estimado da média média amostral é:

$$\sigma_{\bar{x}} = \frac{S}{\sqrt{n}}$$

- Quando o estimador seguir uma distribuição normal, podemos estar confiantes que o valor verdadeiro do parâmetro estará entre dois erros-padrão da estimativa (para grandes valores de n este é um resultado útil)
- Nos casos em que o estimador é não viciado e não normalmente distribuído, a estimativa do parâmetro, em no máximo 6% das vezes, se desviará do valor verdadeiro tanto quanto 4 erros-padrão.



11.9.2. Estimador Consistente

Um estimador $\hat{\theta}$ é consistente se á medida em que o tamanho amostral aumenta, o seu valor esperado converge para o parâmetro de interesse e sua variância converge para zero.

$$\lim_{n \rightarrow \infty} E[\hat{\theta}] = \theta$$

$$\lim_{n \rightarrow \infty} Var[\hat{\theta}] = 0$$

- A média amostral é consistente é cosintente para estimar a média verdadeira;
- O primeiro item colectado da amostra não é consistente para estimar a média populacional.

11.9.3. Estimador Eficiente

Dados dois estimadores $\hat{\theta}_1$ e $\hat{\theta}_2$, não viciados para um parâmetro θ , dizemos que $\hat{\theta}_1$ é mais eficiente que $\hat{\theta}_2$ se $Var[\hat{\theta}_1] < Var[\hat{\theta}_2]$

Exemplo 11.6 (Estimador eficiente)

- No caso de amostra proveniente de distribuição Normal:
 - ✓ Média amostral e mediana amostral são não viciadas para estimar a média populacional:

$$E[\bar{x}] = \mu \text{ e } E[\tilde{x}] = \mu$$

- ✓ A média amostral e mediana amostral são consistentes para estimar a média verdadeira:

$$Var[\bar{x}] = \frac{\sigma^2}{n} \text{ e } Var[\tilde{x}] = \frac{\pi\sigma^2}{2n}$$

- ✓ A média amostral é mais eficiente que a mediana amostral para estimar a média populacional:

$$\frac{Var[\bar{x}]}{Var[\tilde{x}]} = \frac{\frac{\sigma^2}{n}}{\frac{\pi\sigma^2}{2n}} = \frac{2}{\pi} = 0.63 < 1$$

11.10. EXERCÍCIOS PROPOSTOS

11.10.1. Seja X_1, X_2, X_3, X_4, X_5 , variáveis aleatórias independentes, identicamente distribuídas de forma normal com a média μ , e variância σ^2 . Defina os seguintes estimadores:

$$1. \hat{\theta}_1 = \frac{X_1 + X_2 + X_3 + 3X_5}{5}$$

$$2. \hat{\theta}_2 = \frac{X_1 + X_2 + X_3 + X_4}{4}$$

$$3. \hat{\theta}_3 = \frac{2X_1 + X_2 + X_3 + X_4}{4}$$

$$4. \hat{\theta}_4 = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{6}$$

$$5. \hat{\theta}_5 = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$$

Diga qual desses estimadores é não viciado e qual é o mais eficiente.

11.10.2. De uma população com $N = 12$ elementos é retirada uma amostra aleatória simples, sem reposição, de $n = 5$.

- Quantas são as possíveis amostras?
- Qual a probabilidade de cada uma destas amostras ser seleccionada?

11.10.3. Uma população é composta dos elementos: A, B, C, D e F.

- Liste todas as possíveis amostras aleatórias simples, sem reposição, com $n = 2$.
- Liste todas as aas, sem reposição, de tamanho $n = 3$.
- Determine a probabilidade de ser sorteada a amostra BC.
- Determine a probabilidade de ser sorteada a amostra ACD.

11.10.4. A tabela abaixo é a distribuição de frequências de uma amostra proveniente de determinada população.

X	1	2	3	4
f	40	45	8	9

- Determine o tamanho da amostra.
- Determine uma estimativa da média da população.
- Determine uma estimativa da variância da população.
- Determine uma estimativa da proporção de valores pares na população.

11.10.5. A tabela abaixo apresenta valores amostrais.

<i>Elementos</i>	A	B	C	D	E
X	5	7	12	15	10

- Qual o tamanho da amostra?
- Determine uma estimativa para a média da população.
- Determine uma estimativa do desvio padrão populacional.
- Determine uma estimativa dos valores ímpares de X .

11.10.6. Utilize os valores da amostra tabelada abaixo, extraída aleatoriamente e sem reposição, de uma população com $N = 2000$ elementos, para estimar:

<i>Elementos</i>	[0---2[[2---4[[4---6[[6---8[[8---10[
<i>X</i>	27	51	49	48	25

- A média da população.
- A variância da população.
- O percentual de elementos menores que 6.
- O erro amostral da média.

11.10.7. De uma população com $N = 4000$ pessoas de uma região foi obtida uma amostra aleatória, sem reposição, de 400 pessoas que revelou 60 analfabetos.

Estime:

- A proporção de analfabetos da região.
- O erro amostral do estimador da proporção.

11.10.8. Uma aas de tamanho 900 extraída de uma população bastante grande apresentou 40% de pessoas do sexo masculino. Estime o erro amostral do estimador da proporção de pessoas do sexo masculino.

11.10.9. Uma população tem média 500 e desvio padrão 30.

- Determinar a probabilidade que uma aas de 100 elementos apresentar um valor médio superior a 504,50.
- Calcule a probabilidade de que uma aas com $n = 64$ valores apresentar média entre 492,5 e 507,5.
- Se uma aas de $n = 144$ for extraída desta população, qual o percentual de médias amostrais que estarão entre 495,5 e 504,5?

11.10.10. Uma população é normalmente distribuída com média 800 e desvio padrão 60.

- Determine a probabilidade de que uma aas de tamanho 9 apresentar média menor que 780.
- Calcule a probabilidade de que uma aas de tamanho $n = 16$ tenha média entre os valores 781,4 e 818,6.
- Que percentual de médias amostrais de uma amostra de tamanho $n = 25$ estará no intervalo [776; 824]?

11.10.11. A proporção de eleitores de um candidato é 20%.

- d) Qual a probabilidade de uma amostra aleatória simples de 100 eleitores apresentar uma proporção amostral superior a 26%?
- a) Qual a probabilidade de uma amostra aleatória simples de 400 eleitores apresentar uma proporção de eleitores do candidato entre 17% e 23%?
- b) Se a amostra aleatória for de 625 eleitores, qual a percentual de valores do estimador proporção amostral que estarão no intervalo $[0,16864; 0,23136]$?

11.10.12. Admitindo que a probabilidade nascer um menino ou uma menina seja iguais, determine a probabilidade de que das próximas 400 crianças a nascerem:

- a) Menos de 45% sejam meninas.
- b) Mais de 54% sejam meninos.

11.10.13. Numa certa cidade, a duração de conversas telefônicas em minutos, segue um modelo Exponencial com parâmetro 3. Observando se uma amostra aleatória de 50 dessas chamadas, qual será a probabilidade de em média, a duração de conversas telefônicas não ultrapassarem 4 minutos.

11.10.14. A proporção de peças fora de especificação num lote é de 0,4. Numa amostra de tamanho 30, calcule a probabilidade de que a proporção de peças defeituosas seja menor do que 0,5.

11.10.15. Uma população consta de 4 números: 3, 7, 11 e 15. Considerar todas as amostras possíveis que podem ser retiradas com reposição. Determinar:

- a) A média populacional;
- b) O desvio padrão da população;
- c) A média da distribuição amostral das médias;
- d) O desvio padrão da distribuição amostral das médias.

11.10.16. Certas válvulas fabricadas por uma companhia têm uma vida média de 800 horas e desvio padrão de 60 horas. Determinar a probabilidade de uma amostra aleatória de 16 válvulas, retiradas do grupo, ter a vida média:

- a) Entre 790 e 810 horas;
- b) Inferior a 785 horas.

11.10.17. Os pesos de fardos recebidos por um depósito têm média de 150 kg e um desvio padrão de 25 kg. Qual é a probabilidade de 25 fardos, recebidos ao acaso e carregados em um elevador, não exceder o limite específico desse último, que é de 4100 kg?

11.10.18. Uma companhia fabrica cilindros que tem uma média de 2 polegadas de diâmetro. O desvio padrão dos diâmetros dos cilindros é de 10 polegadas. Os diâmetros de uma amostra de 4 cilindros são medidos todas as horas. A média amostral é usada para decidir se o processo de fabricação está operando satisfatoriamente ou não. A seguinte regra de decisão é aplicada: se diâmetro médio da amostra de 4 cilindros é maior ou igual a 2,15 polegadas, ou menor ou igual a 1,85 polegadas, interrompe-se o processo.

- a) Qual é a probabilidade de parar o processo se a média do processo permanece constante no valor de 2,00 polegadas?
- b) Qual é a probabilidade de parar o processo se a média do processo muda para $\mu = 2,10$ polegadas?
- c) Qual é a probabilidade do processo continuar operando se a média do processo mudar para $\mu = 2,15$ polegadas?

12. INTERVALO DE CONFIANÇA

Um intervalo de confiança para um parâmetro populacional é um intervalo com uma proporção associada p gerada por uma amostra aleatória de uma população subjacente, de tal forma que se o experimento for repetido várias vezes e o intervalo de confiança for recalculado para cada experimento com mesmo procedimento, uma proporção p dos intervalos de confiança conteria o parâmetro estatístico em questão.

Os intervalos de confiança são usados para indicar a confiabilidade de uma estimativa.

12.1. Estimação por intervalo

Normalmente, no processo de investigação de um parâmetro θ , necessitamos ir além da sua estimativa pontual $\hat{\theta}$. O fato de não se conhecer o valor de θ pode causar uma “insegurança” e levar a um questionamento:

- Quão próximo estamos do valor real de θ quando obtemos sua estimativa?

A resposta depende da precisão (ou variância) do estimador e , também, do valor real do parâmetro. Uma maneira de contornar esse problema consiste em se encontrar um intervalo em torno de $\hat{\theta}$ que tenha alta probabilidade de englobar θ .

Um intervalo de confiança ou estimativa intervalar é uma amplitude ou um intervalo de valores que tem probabilidade de conter o verdadeiro valor da população.

$$\hat{\theta} - erro \leq \theta \leq \hat{\theta} + erro$$

Onde

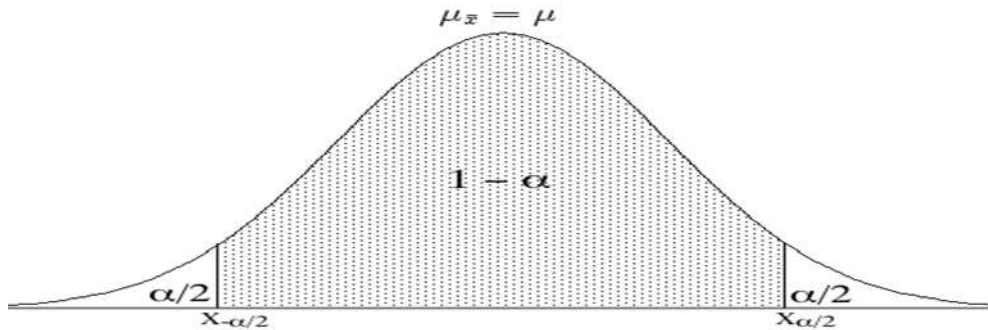
$$\theta \in [\hat{\theta} - erro; \hat{\theta} + erro]$$

12.2. Nível de confiança

O grau de confiança ou nível de confiança é a frequência ou probabilidade $1 - \alpha$ com a qual o intervalo observado contém o parâmetro real de interesse quando o experimento é repetido várias vezes. Em outras palavras, o nível de confiança seria a proporção de intervalos de confiança construídos em experimentos separados da mesma população e com o mesmo procedimento que contém o parâmetro de interesse real.

São escolhas comuns para grau de confiança: 90% (com $\alpha = 0.10$); 95% (com $\alpha = 0.05$) e 99% (com $\alpha = 0.01$).

Dentre essas, a mais utilizada é 95%.



12.3. Valor crítico

É um número na fronteira que separa os valores das estatísticas amostrais prováveis de ocorrerem, dos valores que têm pouca chance de ocorrer. O número $Z_{\alpha/2}$ é um valor crítico que é um escore Z com a propriedade de separar uma área de $\frac{\alpha}{2}$ na cauda direita da distribuição normal padrão. Há uma área de $1 - \alpha$ entre as fronteiras verticais $-Z_{\alpha/2}$ e $Z_{\alpha/2}$.

12.4. Margem do erro

É a diferença entre a média amostral observada \bar{x} e a verdadeira média populacional μ . A margem do erro ε é chamada também do erro máximo da estimativa e pode ser obtida multiplicando-se o valor crítico pelo desvio padrão das médias amostrais.

12.5. Intervalo de confiança para uma média

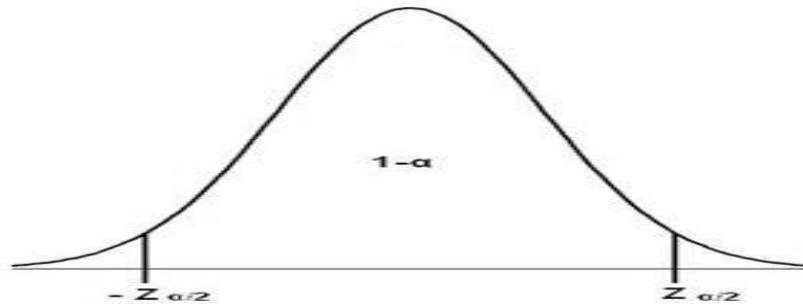
12.5.1. Intervalo de confiança para uma média quando a variância é conhecida

Seja $X_1, X_2, X_3, \dots, X_n$ uma amostra aleatória de tamanho n , de uma população com média μ (desconhecida) e variância σ^2 (conhecida). A média amostral \bar{x} , tem distribuição normal com média μ e variância $\frac{\sigma^2}{n}$. Isto é:

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0; 1)$$

Logo, fixando um nível de confiança $(1 - \alpha)$, pode-se determinar $Z_{\alpha/2}$ de tal forma:

$$P\left(-Z_{\frac{\alpha}{2}} \leq Z \leq Z_{\frac{\alpha}{2}}\right) = 1 - \alpha \rightarrow P\left(-Z_{\frac{\alpha}{2}} \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$



Logo, intervalo de 100 $(1 - \alpha)$ % de confiança para μ é dado por:

$$\bar{x} - Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$$

Para populações finitas o intervalo será dado por:

$$IC(\mu, 1 - \alpha) = \bar{x} - Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N - n}{N - 1}} \leq \mu \leq \bar{x} + Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N - n}{N - 1}}$$

Onde

$$\frac{N - n}{N - 1} \text{ vé o factor de correcção.}$$

- **Intervalo de confiança unilateral esquerdo**

$$\bar{x} - Z_{\alpha} * \frac{\sigma}{\sqrt{n}} \leq \mu$$

ou

$$\bar{x} - Z_{\alpha} * \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N - n}{N - 1}} \leq \mu$$

- **Intervalo de confiança unilateral direito**

$$\bar{x} + Z_{\alpha} * \frac{\sigma}{\sqrt{n}} \geq \mu$$

ou

$$\bar{x} + Z_{\alpha} * \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N - n}{N - 1}} \geq \mu$$

Exemplo 12.1 (Intervalo de confiança para uma média)

Em uma indústria de cerveja, a quantidade de cerveja inserida em latas tem-se comportado como uma distribuição normal com desvio padrão 3 ml. Após alguns problemas na linha de produção, suspeita-se que houve alteração na média. Uma amostra de 20 latas acusou uma média 346 ml. Obtenha um intervalo de 95% para a quantidade média de cerveja inserida em latas, supondo que não tenha ocorrido alteração na variabilidade.

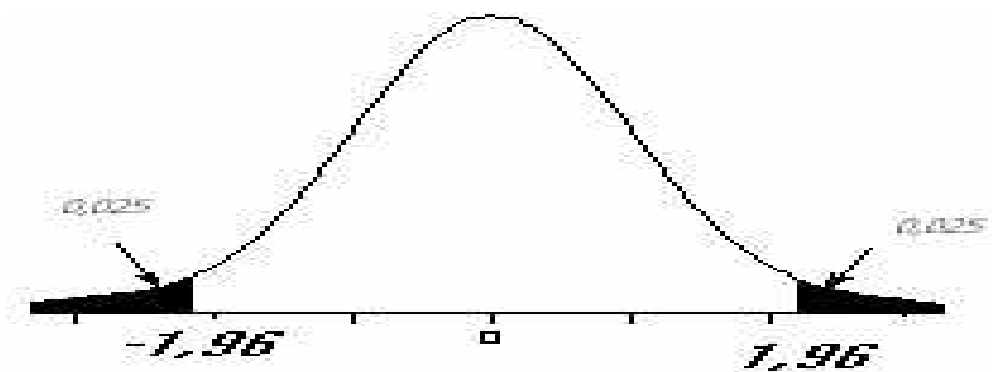
Solução

Pelos dados temos:

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0; 1)$$

Já que $1 - \alpha = 0.95$, temos da tabela normal padrão $Z_{0.975} = 1.96$

$$\sigma = 3; n = 20; \bar{x} = 346$$



$$IC(\mu, 1 - \alpha) = \bar{x} - 1.96 * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96 * \frac{\sigma}{\sqrt{n}} \rightarrow$$

$$IC(\mu, 0.95) = 346 - 1.96 * \frac{3}{\sqrt{20}} \leq \mu \leq 346 + 1.96 * \frac{3}{\sqrt{20}} = (346.69; 347.31)$$

Interpretação

$\mu \in [346.69; 347.31]$: Isto significa que se colectássemos um número infinito de amostras, se para cada uma dessas amostras calculássemos o intervalo dessa forma 95% deles iriam conter o verdadeiro valor de μ .

Nota:

- O comprimento do intervalo de confiança está associado à precisão, quanto menor for o comprimento mais precisa é a média.

- Se diminuirmos α , isto é, aumentarmos $1 - \alpha$ (grau de confiança), mantendo n fixo, a vai aumentar e conseqüentemente o comprimento do intervalo. Não é possível fazer $\alpha = 0$ pois nesse caso $amplitude = +\infty$.

12.5.1.1. Tamanho de amostra e Erro Máximo de Estimação

Quando estimamos μ usando \bar{x} cometemos um erro. Fixando um nível de confiança $(1 - \alpha)$ podemos escolher um tamanho de amostra n dependendo do erro máximo (margem de erro) que queremos cometer.

Escolhemos n tal que o erro máximo cometido é:

$$\varepsilon = Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$$

Ou

$$\varepsilon = Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N - n}{N - 1}}$$

Isolando o n temos que:

$$n = \left(\frac{Z_{\frac{\alpha}{2}} * \sigma}{\varepsilon} \right)^2$$

Ou

$$n = \frac{Z_{\frac{\alpha}{2}}^2 * \sigma^2 * N}{\varepsilon^2 * (N - 1) + Z_{\frac{\alpha}{2}}^2 * \sigma^2}$$

Exemplo 12.2 (tamanho de amostra)

Uma firma construtora deseja estimar a resistência média das barras de aço utilizadas na construção de casas. Qual o tamanho amostral necessário para garantir que haja um risco de 0,001 de ultrapassar um erro de 5 kg ou mais na estimação? O desvio padrão da resistência para este tipo de barra é de 25 kg.

Solução

Pelos dados temos:

$$\sigma = 25; \alpha = 0.001; \varepsilon = 5; Z_{0.9995} = 3.29$$

$$\varepsilon = Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} \rightarrow n = \left(\frac{3.29 * 25}{5} \right)^2 \approx 271$$

12.5.2. Intervalo de confiança para uma média quando a variância é desconhecida (n grande)

Seja $X_1, X_2, X_3, \dots, X_n$ uma amostra aleatória de tamanho n , de uma população não normal com média μ (desconhecida) e variância σ^2 (conhecida). Se assumirmos que ($n \geq 30$), pelo teorema de limite central diz-se que:

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0; 1)$$

á medida que $n \rightarrow \infty$. Pode-se mostrar que esse resultado continua valendo se substituirmos σ por S .

A conclusão dessas duas observações é a seguinte:

Dada uma amostra aleatória simples $X_1, X_2, X_3, \dots, X_n$ de uma população X com média μ e variância σ^2 , então:

$$Z = \frac{\bar{x} - \mu}{S/\sqrt{n}} \approx N(0; 1)$$

Para n suficientemente grande. Nesse caso, o intervalo de confiança aproximado de nível de confiança $1 - \alpha$ para μ é:

$$IC(\mu, 1 - \alpha) = \left(\bar{x} - Z_{\frac{\alpha}{2}} * \frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\frac{\alpha}{2}} * \frac{S}{\sqrt{n}} \right)$$

- **Intervalo de confiança unilateral esquerdo**

$$\bar{x} - Z_{\alpha} * \frac{S}{\sqrt{n}} \leq \mu$$

- **Intervalo de confiança unilateral direito**

$$\bar{x} + Z_{\alpha} * \frac{S}{\sqrt{n}} \geq \mu$$

12.5.2.1. Margem do erro

Note, mais uma vez a fórmula do intervalo de confiança, fornece a seguinte fórmula da margem do erro:

$$\varepsilon = Z_{\frac{\alpha}{2}} * \frac{S}{\sqrt{n}}$$

Exemplo 12.3 (Intervalo de confiança para uma média)

A partir de uma amostra aleatória simples de tamanho $n = 100$, os seguintes valores foram obtidos: $\bar{x} = 12,36$ e $S^2 = 132,56$. Obtenha um intervalo de confiança de nível de confiança 90% para a média populacional μ .

Solução

Como o tamanho amostral é grande, podemos usar a aproximação normal. Como $1 - \alpha = 0,90$, em cada cauda temos que ter 5% e, assim, devemos procurar no corpo da tabela da distribuição normal o valor mais próximo de 0,45. Resulta que

$Z_{0,05} = 1,64$; o que nos dá o seguinte intervalo de confiança:

$$\begin{aligned} \bar{x} - Z_{\frac{\alpha}{2}} * \frac{S}{\sqrt{n}} &\leq \mu \leq \bar{x} + Z_{\frac{\alpha}{2}} * \frac{S}{\sqrt{n}} \rightarrow \\ 12.36 - 1.64 * \frac{11.51}{\sqrt{100}} &\leq \mu \leq 12.36 + 1.64 * \frac{11.51}{\sqrt{100}} \\ 10.472 &\leq \mu \leq 14.248 \end{aligned}$$

Interpretação: A um nível de confiança de 90% pode-se concluir que o intervalo [10.472; 14.248] contém o verdadeiro médio populacional.

12.5.3. Intervalo de confiança para uma média quando a variância é desconhecida (n pequeno)

O intervalo de confiança para a média de uma população normal com variância desconhecida é obtido com base no seguinte resultado:

Se $X_1, X_2, X_3, \dots, X_n$ é uma amostra aleatória simples de uma população $X \sim N(\mu; \sigma^2)$, então:

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

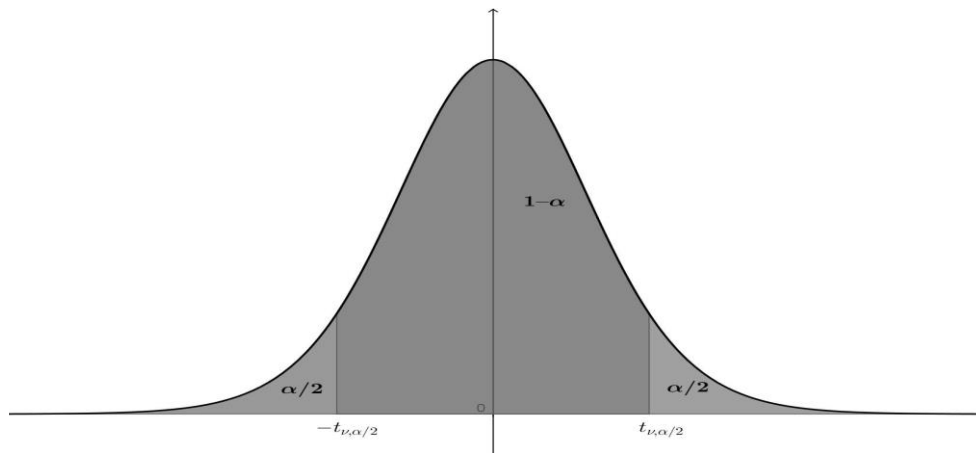
Onde t é a distribuição t de Student, $n - 1$ graus de liberdades e:

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

O intervalo de confiança para μ de nível de confiança $1 - \alpha$ é:

$$\bar{x} - t_{\frac{\alpha}{2}; n-1} * \frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2}; n-1} * \frac{S}{\sqrt{n}}$$

Onde $t_{\frac{\alpha}{2};n-1}$ é o valor crítico da distribuição t-student com $n - 1$ graus de liberdade que deixa a área $\frac{\alpha}{2}$ acima dele.



- **Intervalo de confiança unilateral esquerdo**

$$\bar{x} - t_{\alpha;n-1} * \frac{S}{\sqrt{n}} \leq \mu$$

- **Intervalo de confiança unilateral direito**

$$\bar{x} + t_{\alpha;n-1} * \frac{S}{\sqrt{n}} \geq \mu$$

A distribuição **T de Student** é uma distribuição de probabilidade estatística, publicada por um autor que se chamou de Student, pseudônimo de William Sealy Gosset, que não podia usar seu nome verdadeiro para publicar trabalhos enquanto trabalhasse para a cervejaria Guinness.

12.5.3.1. Margem do erro

Note, mais uma vez a fórmula do intervalo de confiança, fornece a seguinte fórmula da margem do erro:

$$\varepsilon = t_{\frac{\alpha}{2};n-1} * \frac{S}{\sqrt{n}}$$

Exemplo 12.4 (Intervalo de confiança para uma média)

De uma população normal com média e variância desconhecidas, extrai-se uma amostra de tamanho 15 obtendo-se: $\bar{x} = 12$ e $S^2 = 49$. Obtenha um intervalo de confiança de para a verdadeira média populacional, utilizando um nível de confiança 95%.

Solução

Os seguintes requisitos para o IC para μ são satisfeitos: a população é normal e a amostra é pequena. Dessa forma, temos que usar a distribuição t com $n - 1 = 14$ graus de liberdade. Como o nível de confiança é de 95%, em cada cauda da distribuição temos que ter 2,5%. Assim, devemos procurar a abscissa t procurando na linha correspondente a 14 graus de liberdade e na coluna correspondente à área de 0,025. Encontramos

$$t_{0,025;14} = 2.145$$

Então o intervalo de confiança é dada por:

$$\begin{aligned} \bar{x} - t_{\frac{\alpha}{2};n-1} * \frac{S}{\sqrt{n}} &\leq \mu \leq \bar{x} + t_{\frac{\alpha}{2};n-1} * \frac{S}{\sqrt{n}} \\ 12 - 2.145 * \frac{7}{\sqrt{15}} &\leq \mu \leq 12 + 2.145 * \frac{7}{\sqrt{15}} \\ 8.1231 &\leq \mu \leq 15.8769 \end{aligned}$$

Interpretação: Temos 95% de certeza de que o intervalo [8.1231; 15.8769] contém o verdadeiro médio populacional.

12.5.4. Intervalo de confiança para a diferença entre duas médias quando as variâncias são conhecidas

Sejam X_1 e X_2 duas variáveis aleatórias normais com médias μ_1 e μ_2 desconhecidas e variâncias σ_1^2 e σ_2^2 conhecidas.

Um intervalo de confiança $1 - \alpha$ para a diferença entre as médias pode ser construído a partir dos resultados de amostras aleatórias de cada uma dessas populações.

Pode ser demonstrada que a variância das diferenças entre as médias vem dada por:

$$\sigma^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Assim o intervalo de confiança bilateral de $1 - \alpha$ será:

$$(\bar{x}_1 - \bar{x}_2) - Z_{\frac{\alpha}{2}} * \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + Z_{\frac{\alpha}{2}} * \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

E os correspondentes intervalos unilaterais serão:

$$(\bar{x}_1 - \bar{x}_2) - Z_\alpha * \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2$$

$$(\bar{x}_1 - \bar{x}_2) + Z_\alpha * \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \geq \mu_1 - \mu_2$$

Se pelos menos uma das amostras é maior ou igual que 30, pode se substituir σ_1^2 por S_1^2 e σ_2^2 por S_2^2 obtendo-se:

$$(\bar{x}_1 - \bar{x}_2) - Z_{\frac{\alpha}{2}} * \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + Z_{\frac{\alpha}{2}} * \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

12.5.5. Intervalo de confiança para a diferença entre duas médias quando as variâncias são desconhecidas e iguais

Sejam X_1 e X_2 duas variáveis aleatórias normais com médias μ_1 e μ_2 desconhecidas e variâncias σ_1^2 e σ_2^2 também desconhecidas.

Se for possível assumir que as variâncias sejam iguais, ou seja, $\sigma_1^2 = \sigma_2^2$ uma estimativa da variância pode ser obtida como:

$$S_p^2 = \frac{(n_1 - 1) * S_1^2 + (n_2 - 1) * S_2^2}{n_1 + n_2 - 2}$$

Uma vez encontrada a estimativa da variância dos valores individuais, pode ser demonstrado que a estimativa da variância da diferença entre as médias será:

$$S^2 = \frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}$$

Com graus de liberdade $v = n_1 + n_2 - 2$.

De modo que o intervalo de confiança bilateral $1 - \alpha$ será:

$$(\bar{x}_1 - \bar{x}_2) - t_{\frac{\alpha}{2};v} * \sqrt{S_p^2 * \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + t_{\frac{\alpha}{2};v} * \sqrt{S_p^2 * \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Os correspondentes intervalos de confiança unilaterais serão:

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha;v} * \sqrt{S_p^2 * \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \leq \mu_1 - \mu_2$$

$$(\bar{x}_1 - \bar{x}_2) + t_{\alpha;v} * \sqrt{S_p^2 * \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \geq \mu_1 - \mu_2$$

Exemplo 12.5 (Intervalo de confiança para a diferença entre duas médias)

Um eixo deve ser montado no interior de um rolamento. Uma amostra de doze unidades indicou para o diâmetro inteiro do rolamento $\bar{x}_1 = 2.538cm$ e $S_1 = 0.008cm$; e para o diâmetro do eixo $\bar{x}_2 = 2.520cm$ e $S_2 = 0.006cm$. Calcule o intervalo de confiança de 99% para a folga de montagem.

Solução

Supondo que as variâncias são iguais têm-se:

$$S_p^2 = \frac{(n_1 - 1) * S_1^2 + (n_2 - 1) * S_2^2}{n_1 + n_2 - 2}$$

$$S_p^2 = \frac{(12 - 1) * 0.008^2 + (12 - 1) * 0.006^2}{12 + 12 - 2} = 0.000050$$

$$\text{Graus de liberdade} = v = n_1 + n_2 - 2 = 22$$

$$t_{0.005;22} = 2.82$$

$$(\bar{x}_1 - \bar{x}_2) - t_{\frac{\alpha}{2};v} * \sqrt{S_p^2 * \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + t_{\frac{\alpha}{2};v} * \sqrt{S_p^2 * \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$(2.538 - 2.52) \pm 2.82 * \sqrt{0.000050 * \left(\frac{1}{12} + \frac{1}{12}\right)} \in \mu_1 - \mu_2$$

$$0.00986 \leq \text{Folga} \leq 0.026$$

12.5.6. Intervalo de confiança para a diferença entre observações

No caso em que se deseja comparar dois sistemas é possível, e as vezes necessário, trabalhar com a diferença entre as observações.

Por exemplo, para comparar dois métodos de tratamento contra corrosão, pode-se escolher diversos blocos de terreno, colocar dois tubos (de marcas diferentes) em cada bloco e observar as diferenças.

Seja:

- X_1 os resultados do sistema 1;
- X_2 os resultados do sistema 2;
- $d = X_1 - X_2$ as diferenças medidas bloco a bloco.

A partir dos resultados de n blocos, calcula-se e usa-se a distribuição t para construir o intervalo de confiança para a média da diferença μ_d :

$$\bar{d} - t_{\frac{\alpha}{2}; n-1} * \frac{S_d}{\sqrt{n}} \leq \mu_d \leq \bar{d} + t_{\frac{\alpha}{2}; n-1} * \frac{S_d}{\sqrt{n}}$$

Se o valor zero estiver contido neste intervalo, então não pode ser descartada a hipótese que o desempenho dos dois sistemas seja o mesmo.

Exemplo 12.6 (Intervalo de confiança para a diferença entre as observações)

Uma empresa quer verificar se o conhecimento de seus alunos a respeito de um determinado assunto melhorou após 30 horas de treinamento. Para isso foi realizado com os quinze alunos do treinamento um teste antes e após o treinamento. Os dados a seguir representam as notas obtidas pelos alunos. Conclua a respeito da eficiência do treinamento com 95% de confiança.

Alunos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Antes	6.5	6.7	7.0	7.0	6.5	7.3	7.8	6.9	6.7	7.2	7.5	7.5	7.2	7.0	6.8
Depois	7.5	7.7	7.9	8.0	7.4	8.3	8.8	8.9	7.7	8.2	8.5	8.5	8.2	8.0	8.8
Diferença	1.0	1.0	0.9	1.0	0.9	1.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0

Solução

Pelos dados temos:

$$\bar{d} = 1.12; S_d = 0.36; t_{0.025; 14} = 2.145$$

$$\bar{d} - t_{\frac{\alpha}{2}; n-1} * \frac{S_d}{\sqrt{n}} \leq \mu_d \leq \bar{d} + t_{\frac{\alpha}{2}; n-1} * \frac{S_d}{\sqrt{n}}$$

$$1.12 - 2.145 * \frac{0.36}{\sqrt{15}} \leq \mu_d \leq 1.12 + 2.145 * \frac{0.36}{\sqrt{15}}$$

$$0.92 \leq \mu_d \leq 1.32$$

Conclusão: Como o valor zero não está incluído no intervalo, rejeita-se a hipótese de que as notas antes e depois sejam as mesmas, logo conclui-se que o treinamento foi eficiente.

12.5.7. Intervalo de confiança para a variância

Outra distribuição importante, definida a partir da distribuição Normal é a distribuição do Chi-quadrado χ^2 .

Seja $X \rightarrow N(0; 1)$ e seja $X_1, X_2, X_3, \dots, X_n$ é uma amostra aleatória deste processo.

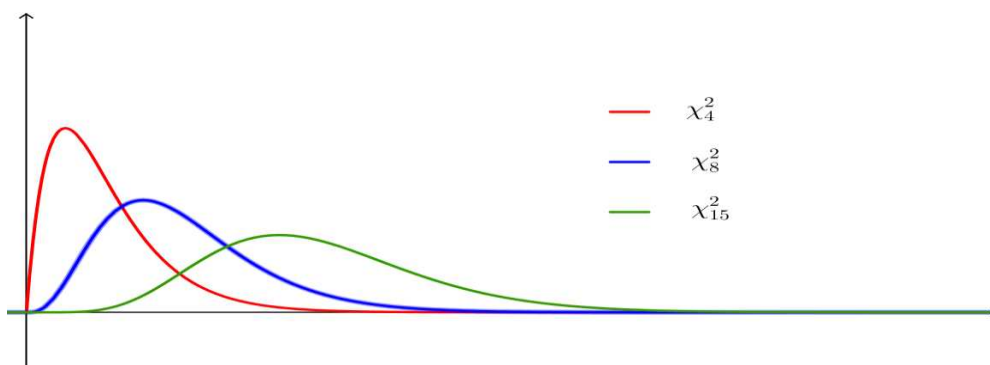
Então, a variável aleatória $\chi_n^2 = x_1^2 + x_2^2 + \dots + x_n^2$ distribui-se de acordo com a distribuição Chi-quadrado cuja a distribuição amostral é:

$$\frac{(n - 1) * S^2}{\sigma^2} \approx \chi_{n-1}^2$$

Ou seja, a distribuição χ^2 é a base para inferências a respeito da variância σ^2 .

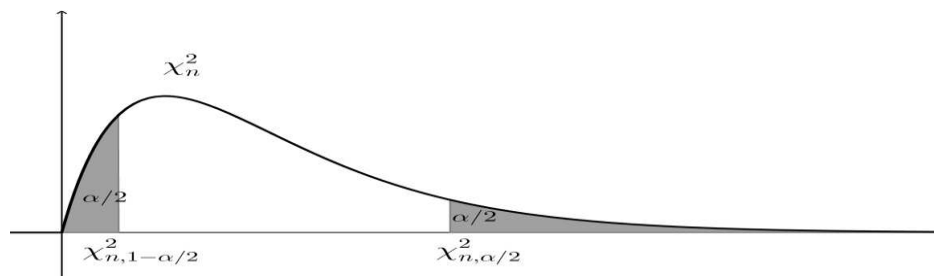
É uma distribuição assimétrica á direita, com média e variância dadas por:

$$\mu = n \text{ e } \sigma^2 = 2n$$



Suponha que X é uma variável aleatória Normal com média e variância desconhecida. Seja que a variância amostral S^2 é calculada para uma amostra de n observações. Então, um intervalo bilateral de confiança $1 - \alpha$ é obtido usando-se a distribuição do Chi-quadrado:

$$\frac{(n-1) * S^2}{\chi^2_{\frac{\alpha}{2}; n-1}} \leq \sigma^2 \leq \frac{(n-1) * S^2}{\chi^2_{1-\frac{\alpha}{2}; n-1}}$$



No caso do interesse residir em intervalos unilaterais de $1 - \alpha$ teremos:

$$\text{Limite inferior: } \frac{(n-1) * S^2}{\chi^2_{\alpha; n-1}} \leq \sigma^2$$

$$\text{Limite superior: } \frac{(n-1) * S^2}{\chi^2_{1-\alpha; n-1}} \geq \sigma^2$$

Exemplo 12.7 (Intervalo de confiança para a variância)

De uma população normal com média e variância desconhecidas, extrai-se uma amostra de tamanho 15 obtendo-se $\bar{x} = 12$ e $S^2 = 49$. Obtenha um intervalo de confiança para a variância populacional, utilizando o nível de confiança de 95%.

Solução

O requisito para o IC para σ^2 é satisfeito, uma vez que a população é normal. Temos que usar a distribuição χ^2 com $n - 1 = 14$ graus de liberdade. Como o nível de confiança é de 95%, em cada cauda da distribuição temos que ter 2,5%. Assim, para a cauda superior, devemos usar o valor crítico $\chi^2_{0.025; 14}$, procurando na linha correspondente a 14 graus de liberdade e na coluna correspondente à probabilidade de 0,025. Encontramos

$$\text{que } \chi^2_{0.025; 14} = 26.119$$

Para a cauda inferior, devemos usar o valor crítico $\chi^2_{0.975; 14}$, procurando na linha correspondente a 14 graus de liberdade e na coluna correspondente à probabilidade de 0,975.

Encontramos que χ

$\chi^2_{0.025; 14} = 26.119$. Logo, o intervalo de confiança é:

$$\frac{14 * 49}{26.119} \leq \sigma^2 \leq \frac{14 * 49}{5.629}$$

12.5.8. Intervalo de confiança para quociente entre duas variâncias

Seja $X_1 \rightarrow N(\mu_1; \sigma_1)$ e $X_2 \rightarrow N(\mu_2; \sigma_2)$. Se S_1 e S_2 são variâncias amostrais, medidas em amostras de tamanho n_1 e n_2 , teremos:

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \approx F_{n_1-1; n_2-1}$$

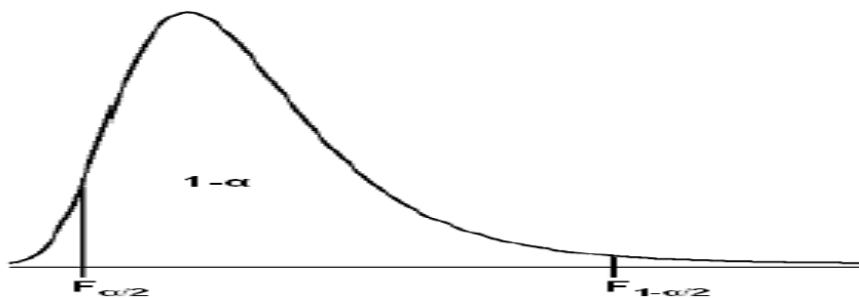
Assim, a distribuição F pode ser usada para fazer inferências sobre a variância de duas distribuições Normais.

Para comparar duas variâncias, σ_1^2 e σ_2^2 , oriundas de populações com distribuição Normal, é vantajoso trabalhar com o quociente σ_1^2/σ_2^2 uma vez que este se distribui conforme a distribuição F.

O intervalo de confiança para este quociente virá dado por:

$$\frac{S_1^2}{S_2^2} * F_{1-\frac{\alpha}{2}; n_1-1; n_2-1} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2} * F_{\frac{\alpha}{2}; n_1-1; n_2-1}$$

Onde $F_{\alpha; u; v}$ são os pontos percentuais da distribuição F com u e v graus de liberdade, tais que $P\{F \geq F_{\alpha; u; v}\} = \alpha$



Se o valor um estiver contido neste intervalo, então não pode ser descartada a hipótese de que a variância das duas populações seja a mesma.

Os respectivos intervalos unilaterais serão dados por:

$$\text{Limite inferior: } \frac{S_1^2}{S_2^2} * F_{1-\alpha; n_1-1; n_2-1} \leq \frac{\sigma_1^2}{\sigma_2^2}$$

$$\text{Limite superior: } \frac{S_1^2}{S_2^2} * F_{\alpha; n_1-1; n_2-1} \geq \frac{\sigma_1^2}{\sigma_2^2}$$

Os valores da distribuição F costumam fornecer apenas os valores de F_{α} , mas $F_{1-\alpha}$ pode ser obtido a partir da seguinte relação:

$$F_{1-\alpha;u;v} = \frac{1}{F_{\alpha;v;u}}$$

Exemplo 12.8 (Intervalo de confiança para a razão entre duas variâncias)

Os valores a seguir representam os tempos de produção de duas máquinas. Analise os dados e conclua a respeito da variabilidade das máquinas A e B a um nível de confiança de 95%.

A	91.0	90.3	90.2	92.1	91.8	91.3	89.3	91.0	91.2	89.6
B	91.8	91.2	89.4	89.2	90.7	92.6	91.3	91.2		

Solução:

$$S^2_A = 0.8307 \text{ e } S^2_B = 1.316$$

$$F_{0.025;9;7} = 4.82 \text{ e } F_{0.925;9;7} = \frac{1}{F_{0.025;7;9}} = \frac{1}{4.20} = 0.238$$

$$\frac{0.8307}{1.316} * 0.238 \leq \frac{\sigma^2_1}{\sigma^2_2} \leq \frac{0.8307}{1.316} * 4.82$$

Conclusão: O intervalo inclui o valor 1, assim não pode ser descartada a hipótese de que a variabilidade das duas máquinas seja a mesma.

Nota: Além de servir para a comparação de duas variâncias, a distribuição F é a chave para a comparação de vários grupos, o que é feito usando o procedimento conhecido como **Análise de Variância**.

12.5.9. Intervalo de confiança para uma proporção

Seja que uma amostra de n observações, $X_1, X_2, X_3, \dots, X_n$, é extraída de um processo de Bernoulli, com probabilidade de sucesso constante igual a p. Então, a soma das observações seguirá o modelo Binomial com parâmetros n e p. Além disso, como cada x_i pode ser 0 ou 1, a média:

$$\bar{X} = \frac{1}{n} * \sum_{i=1}^n x_i$$

Será uma variável discreta contida no espaço $\{0, \frac{1}{n}, \frac{2}{n}, \dots, 1\}$.

A distribuição pode ser obtida a partir da Binomial, uma vez que:

$$P(\bar{X} \leq a) = P(X \leq an) = \sum_{k=0}^{an} \binom{n}{k} * p^k * (1-p)^{n-k}$$

Onde $[an]$ é o maior inteiro menor que an . A média e a variância de \bar{X} são:

$$\mu_{\bar{X}} = p \text{ e } \sigma_{\bar{X}}^2 = \frac{p * (1-p)}{n}$$

Logo se n é grande ($n \geq 30$) e $p \geq 0,1$, então a aproximação Normal para a Binomial pode ser usada, resultando no seguinte intervalo de confiança de $1 - \alpha$:

$$Z = \frac{p - \pi}{\sqrt{\frac{\pi * (1 - \pi)}{n}}} \sim N(0; 1)$$

$$p - Z_{\frac{\alpha}{2}} * \sqrt{\frac{p * (1 - p)}{n}} \leq \pi \leq p + Z_{\frac{\alpha}{2}} * \sqrt{\frac{p * (1 - p)}{n}}$$

Se n é pequeno o problema deve ser resolvido usando tabelas de distribuição binomial.

Se p é pequeno, é possível usar a distribuição de Poisson.

Para as populações N finitas e conhecidas, o intervalo será dado por:

$$p - Z_{\frac{\alpha}{2}} * \sqrt{\frac{p * (1 - p)}{n}} * \sqrt{\frac{N - n}{N - 1}} \leq \pi \leq p + Z_{\frac{\alpha}{2}} * \sqrt{\frac{p * (1 - p)}{n}} * \sqrt{\frac{N - n}{N - 1}}$$

12.5.9.1. Erro amostral e Tamanho de amostra

A partir do intervalo de confiança pode se obter o erro amostral e tamanho de amostra definida por:

$$\varepsilon = Z_{\frac{\alpha}{2}} * \sqrt{\frac{p * (1 - p)}{n}}$$

$$n = \frac{Z_{\frac{\alpha}{2}}^2 * p * (1 - p)}{\varepsilon^2}$$

E para as populações finitas:

$$\varepsilon = Z_{\frac{\alpha}{2}} * \sqrt{\frac{p * (1 - p)}{n}} * \sqrt{\frac{N - n}{N - 1}}$$

$$n = \frac{Z_{\frac{\alpha}{2}}^2 * p * (1 - p) * N}{\varepsilon^2 * (N - 1) + Z_{\frac{\alpha}{2}}^2 * p * (1 - p)}$$

Exemplo 12.9 (Intervalo de confiança para uma proporção)

Um empresário deseja conhecer a satisfação de seus clientes em relação aos serviços prestados por sua empresa. Em uma amostra aleatória de 100 clientes entrevistados, 4 pessoas demonstraram insatisfação com os serviços prestados. Construa um intervalo de 95% de confiança para a proporção de clientes insatisfeitos.

Solução

Pelos dados temos:

$$Z = \frac{p - \pi}{\sqrt{\frac{\pi * (1 - \pi)}{n}}} \sim N(0; 1)$$

Já que $1 - \alpha = 0.95$, temos da tabela normal padrão $Z_{0.975} = 1.96$

$$p = \frac{4}{100} = 0.04 ; n = 100$$

Substituindo na fórmula temos:

$$p - Z_{\frac{\alpha}{2}} * \sqrt{\frac{p * (1 - p)}{n}} \leq \pi \leq p + Z_{\frac{\alpha}{2}} * \sqrt{\frac{p * (1 - p)}{n}}$$

$$0.04 - 1.96 * \sqrt{\frac{0.04 * (1 - 0.04)}{100}} \leq \pi \leq 0.04 + 1.96 * \sqrt{\frac{0.04 * (1 - 0.04)}{100}}$$

$$0.03 \leq \pi \leq 0.05$$

Exemplo 12.10 (Intervalo de confiança para uma proporção – tamanho de amostra)

Um fornecedor alega que entrega 10% de produtos defeituosos. Qual o tamanho de amostra suficiente para estimar a proporção de produtos defeituosos entregues por este fornecedor com precisão de 0.03 e 95% de confiança?

Solução

Já que $1 - \alpha = 0.95$, temos da tabela normal padrão $Z_{0.975} = 1.96$

$$p = 0.10 ; \varepsilon = 0.03$$

$$n = \frac{Z^2_{\frac{\alpha}{2}} * p * (1 - p)}{\varepsilon^2} = \frac{1.96^2 * 0.10 * (1 - 0.10)}{0.03^2} = 384.16$$

Logo, é necessário uma amostra de 385 produtos.

12.5.10. Intervalo de confiança para a diferença entre as proporções

Seja X_1 e X_2 variáveis aleatórias simples que representam o número de sucessos contidos nas amostras independentes retiradas, respectivamente, da População 1 e da População 2; e

- n_1 : Tamanho da amostra da População 1;
- n_2 : Tamanho da amostra da População 2;

$$\text{Se } p_1 = \frac{X_1}{n_1} \text{ e } p_2 = \frac{X_2}{n_2}$$

Para estimar $\pi_1 - \pi_2$ pontualmente, usamos o valor do estimador pontual $p_1 - p_2$.

$$\text{Tem-se: } X_1 \sim B(n_1; \pi_1) \text{ e } X_2 \sim B(n_2; \pi_2)$$

Se $n_1 > 30$ e $n_2 > 30$ então:

$$X_1 \sim N(n_1\pi_1; n_1\pi_1(1 - \pi_1)) \text{ e } X_2 \sim N(n_2\pi_2; n_2\pi_2(1 - \pi_2))$$

$$p_1 \sim N\left(\pi_1; \frac{\pi_1(1 - \pi_1)}{n_1}\right) \text{ e } p_2 \sim N\left(\pi_2; \frac{\pi_2(1 - \pi_2)}{n_2}\right)$$

$$(p_1 - p_2) \sim N\left(\pi_1 - \pi_2; \frac{p_1 * q_1}{n_1} + \frac{p_2 * q_2}{n_2}\right)$$

O o intervalo de confiança a $1 - \alpha$, para $\pi_1 - \pi_2$ será dado por:

$$(p_1 - p_2) - Z^2_{\frac{\alpha}{2}} * \sqrt{\frac{p_1 * q_1}{n_1} + \frac{p_2 * q_2}{n_2}} \leq \pi_1 - \pi_2 \leq (p_1 - p_2) + Z^2_{\frac{\alpha}{2}} * \sqrt{\frac{p_1 * q_1}{n_1} + \frac{p_2 * q_2}{n_2}}$$

12.6. EXERCÍCIOS PROPOSTOS

12.6.1. Defina estimação pontual e estimação por intervalos de confiança. Diga qual das duas será melhor, justificando.

12.6.2. Seja X uma população com distribuição normal de média μ e desvio padrão igual a 2. Uma amostra aleatória de dimensão $n = 25$ foi extraída desta população e revelou uma média de 78,3.

- Calcule o intervalo de confiança para média a 99%.
- Qual a amplitude do intervalo de confiança (a 99% de confiança) ao estimar μ por $\bar{x} = 78.3$?
- Qual deverá ser a dimensão da amostra para que a amplitude (a 99% de confiança), ao estimar μ por \bar{x} , não exceda os 0.1?
- Calcule o intervalo de confiança a 95% para μ .
- Qual o efeito de variar o grau de confiança?
- Qual deverá ser a dimensão da amostra para a amplitude, a 95% de confiança, ao estimar μ por \bar{x} não exceda os 0.1? E a 99, 9% de confiança?

12.6.3. Considere uma v.a. normal de variância igual a 4. Recolheu-se a seguinte amostra:

3, 7, 9, 10, 11, 12, 12, 14

- Determine um intervalo de confiança a 90% para a média.
- Qual deveria ser o grau de confiança a utilizar para que a amplitude do intervalo fosse 1.813?
- Indique a dimensão da amostra que consideraria para que a amplitude do intervalo seja inferior a um, nas condições da alínea 1.
- Explique sucintamente o que aconteceria se aumentasse para 99% o grau de confiança, mantendo a amostra.

12.6.4. Uma fábrica que produz papel quer estimar o tempo médio requerido para uma nova máquina produzir uma resma de papel. Sabe-se que uma amostra de 36 resmas produzidas por essa máquina requereu em média cerca de 1.5 minutos/resma. Assumindo que $\sigma = 0.30$ minutos, construa um intervalo de confiança a 95%.

12.6.5. O dono de um café quer calcular o lucro médio diário por cliente. Numa amostra de 100 clientes verificou que o gasto médio por cliente era de 350 unidades monetárias (u.m.), sendo o desvio padrão dessa amostra de 75 u.m. Estime um intervalo de confiança para o verdadeiro gasto médio com 90% de confiança.

12.6.6. Admita que a densidade de construção num projecto de urbanização tem um comportamento Normal. Uma amostra aleatória de 51 lotes desse projecto forneceu os seguintes dados:

$$\sum_{i=1}^{51} x_i = 227.2 \qquad \sum_{i=1}^{51} x_i^2 = 2242.6$$

- Indique uma estimativa pontual para a densidade média de construção e respectiva variância.
- Deduzza, calcule e interprete um intervalo de confiança a 95% para a densidade média de construção.
- Que dimensão deveria ter a amostra para que a amplitude do intervalo anterior fosse reduzida a metade?

12.6.7. Uma fábrica de relógios de alta precisão pretende estudar a fiabilidade da sua produção. É escolhida uma amostra aleatória de 10 relógios. Ao fim de um mês estes relógios são confrontados com um relógio padrão e o seu desvio é registado; resulta que a média da amostra é de 0,7 segundos e o seu desvio padrão (modificado) 0,4 segundos. Admitindo que a distribuição dos erros dos relógios (relativamente ao relógio padrão) é normal, o que pode afirmar com 90% de confiança, quanto à fiabilidade média dos relógios da fábrica?

12.6.8. Um minimercado pretende estimar o número médio de litros de água que vende diariamente (fenómeno com comportamento normal), para efeitos de controlo de encomendas a fornecedores. Ao fim de 20 dias de negócio, verificou que em média vendia 32 litros de água/dia, sendo o desvio padrão desta amostra igual a 12 litros. Admitindo a normalidade, calcule os limites de confiança para um grau de confiança de 95%.

12.6.9. Com a finalidade de estimar o peso médio (em quilos) das crianças de 15 anos de idade em determinada região geográfica, seleccionaram-se aleatoriamente 10 crianças que forneceram uma média de 38.4 quilos e um desvio padrão de 5.5 quilos. Admitindo a normalidade.

- a) Determine um intervalo de confiança a 95% para o peso médio de todas as crianças.
- b) Considerando que a estimativa para o valor médio não é suficientemente precisa (dado que o intervalo de confiança é demasiado grande), pergunta-se: qual deve ser o tamanho da amostra para que o intervalo de confiança a 95%, tenha uma amplitude de 3 quilos?

12.6.10. Pretende-se estudar o comportamento de um rio, por isso retiraram-se 19 medições do caudal do rio em diferentes alturas do ano. Concluiu-se dessa amostra que o caudal médio é 6.94 e o desvio padrão é 1,1. Admitindo a normalidade da população.

- a) Deduza um intervalo de confiança a $(1 - \alpha)100\%$ para o caudal médio do rio.
- b) Obteve-se o seguinte intervalo de confiança para o valor médio do caudal [6.503; 7.376]. Indique a confiança que deve ser atribuída a esse intervalo

12.6.11. Pesaram-se 16 sacos de café e com os pesos observados, em gramas, construiu-se o seguinte intervalo de confiança a 95%, para o valor médio do peso de um saco: [1000.74; 1009.26].

- a) Deduza o valor médio e o desvio padrão (corrigido) do peso dos sacos que constituem a amostra, admitindo a normalidade da população.
- b) Para construir um intervalo de confiança com uma amplitude de 3 gramas, qual deverá ser a dimensão da amostra, mantendo-se o grau de confiança do intervalo?

12.6.12. A concentração activa de um ingrediente num detergente líquido é supostamente afectada pelo catalisador usado no processo. O desvio padrão da concentração activa é 3 gramas/litro independentemente do catalisador utilizado, sendo o comportamento do processo normal. Foram recolhidas 10 observações cada uma com o seu catalisador:

Cat. 1	57.9	66.2	65.4	65.2	62.6	67.6	63.7	67.2	71.0	65.4
Cat. 2	66.4	71.7	70.3	69.3	64.8	69.6	68.6	69.4	65.3	68.8

Determine um intervalo de confiança a 95% para a diferença de médias dos dados obtidos nos dois catalisadores.

12.6.13. Um engenheiro civil tenciona medir a força compressiva de dois tipos de betão. De duas amostras aleatórias independentes de 10 elementos dos dois tipos resultaram:

Tipo I	3250	3268	4302	3184	3266	3297	3332	3502	3064	3116
Tipo II	3094	3268	4302	3184	3266	3124	3316	3212	3380	3018

Considerando que as amostras provêm de populações Normais com desvio padrão igual a 353 e 133, respectivamente, determine um intervalo de confiança a 95% para a diferença entre os valores esperados das duas populações.

12.6.14. Pretende-se investigar o nível de remuneração salarial dos homens e mulheres de certa categoria profissional. De duas amostras obtidas entre dois grupos, destacam-se os seguintes resultados (em unidades monetárias):

Amostra de 250 homens	$\bar{x} = 33.8$	$S^2_1 = 5.7$
Amostra de 150 mulheres	$\bar{x} = 31$	$S^2_2 = 10.3$

Construa um intervalo de confiança a 99% para as diferenças salariais médias entre os dois sexos e conclua sobre a possível existência de discriminação sexual na atribuição de remunerações.

12.6.15. Em duas populações de cobaias de laboratório (com comportamentos normais e variâncias iguais), uma de animais do sexo masculino e outra de animais do sexo feminino, foram recolhidas duas amostras com dimensões 11 e 31 respectivamente. Os dados amostrais relativos aos pesos, em gramas, destas cobaias foram os seguintes:

n_1	$\bar{x}_1 = 818$	$S_1 = 40$
n_2	$\bar{x}_2 = 715$	$S_2 = 50$

Determine um intervalo de confiança a 98% para a diferença dos pesos médios e verifique se uma das populações é, em média, mais pesada do que a outra.

12.6.16. Para comparar a eficiência de dois métodos de ensino, uma turma de 24 alunos foi dividida aleatoriamente em dois grupos. Cada grupo é ensinado de acordo com um método diferente. Os resultados no fim do semestre são os seguintes (numa escala de 0 a 100):

1º Grupo	$n_1 = 13$	$\bar{x}_1 = 74.5$	$S^2_1 = 82.6$
2º Grupo	$n_2 = 11$	$\bar{x}_2 = 71.8$	$S^2_2 = 112.6$

Assumindo que as populações são normais (com variâncias iguais), obteve-se o seguinte intervalo de confiança para a diferença entre os valores esperados das duas populações:

$$[-5.635; 11.035]$$

Indique qual o grau de confiança utilizado no cálculo deste intervalo.

12.6.17. Para avaliar a dureza de um material plástico recolheu-se a seguinte amostra de 8 elementos:

$$5.0; 4.9; 4.6; 5.1; 4.7; 4.8; 4.9; 5.1$$

Supondo a normalidade da população:

- Indique estimadores pontuais de μ e de σ^2 e com base na amostra obtenha estimativas pontuais para cada um dos parâmetros.
- Deduza um intervalo de confiança a 95% para a variância da população.
- Indique, justificando, o valor lógico da seguinte afirmação: "O efeito conjugado de um aumento simultâneo da dimensão da amostra e do grau de confiança, conduz obrigatoriamente à redução da amplitude de um intervalo de confiança."

12.6.18. Suponha-se em presença de uma população normal, com parâmetros desconhecidos. Com base numa amostra casual, com 16 observações, foi construído o seguinte intervalo de confiança para a média da população:

$$[7.398, 12.602]$$

- Sabendo que, com a informação da amostra, obteve-se $s = 4$, qual o grau de confiança que pode atribuir ao intervalo atrás referido?
- Com base na mesma amostra construa um intervalo de confiança a 95% para a variância da população.
- Suponha que a verdadeira variância da população é 44. Se pretender construir um intervalo de confiança, a 95%, para a média da população cuja amplitude não exceda 2.5, qual deverá ser a dimensão da amostra a considerar?

12.6.19. Recolheram-se 9 observações de uma v.a $N(8; \sigma)$ obtendo-se os seguintes valores:

7.2; 7.8; 7.5; 8.6; 7.9; 8.3; 6.4; 8.4; 9.8

Construa um intervalo de confiança para σ^2 a 95%.

12.6.20. O gerente de uma rede de hipermercados está a analisar os desvios observados no volume de vendas mensais. Este gerente sabe que o volume de vendas mensais segue uma lei Normal com média 160. Após ter recolhido 20 meses de observações, resumiu-os no seguinte resultado:

$$\sum_{i=1}^{20} (x_i - 160)^2 = 8000$$

Deduz e calcule um intervalo de confiança, a 95%, para o desvio padrão do volume mensal de vendas e comente-o.

12.6.21. Considere-se uma população com distribuição Normal de parâmetros desconhecidos. Dessa população foi retirada uma amostra casual de dimensão 25. Suponha-se que a amostra forneceu os seguintes resultados:

$$\sum_{i=1}^{25} x_i = 75 \qquad \sum_{i=1}^{25} x_i^2 = 321$$

Construa um intervalo de confiança a 95% para o desvio padrão.

12.6.22. A altura (em mm) da espuma de sabão numa bacia é importante para os fabricantes de detergentes e supõe-se que o seu comportamento é Normal. Foi efectuada uma experiência, colocando a mesma quantidade de detergente em 10 bacias de tamanho standard e, depois de uma certa agitação da água, mediu-se a altura da espuma. Obtiveram-se os seguintes resultados:

$$\sum_{i=1}^{10} x_i = 229 \qquad \sum_{i=1}^{10} (x_i - \bar{x})^2 = 1553$$

- Determine uma estimativa pontual para a média e para o desvio padrão.
- Determine um intervalo a 99% de confiança para o desvio padrão.
- Comente os dois tipos de estimativas obtidas (nas alíneas anteriores) para o desvio padrão.

12.6.23. Num estudo de mercado foi encontrado o seguinte intervalo de confiança a 95% para a proporção de pessoas receptivas a um novo tipo de espuma de banho a lançar em breve no mercado: $[0.52; 0.61]$. Comente as seguintes afirmações, indicando se estas lhe parecem correctas ou incorrectas:

- a) 95% das pessoas vão passar a usar a nova espuma de banho.
- b) A probabilidade da nova espuma de banho alcançar uma quota de mercado de 50%, é de 0.95.
- c) A quota de mercado poderá ser, com 95% de confiança, de 56.5% (valor intermédio do intervalo);
- d) O resultado obtido indica apenas que é oportuno proceder ao lançamento da nova espuma de banho.

12.6.24. Numa região afectada por um surto epidémico, observou-se uma amostra de 2500 indivíduos, tendo-se encontrado 850 contaminados. Determine intervalos de confiança a 95% e 98% de confiança para a proporção de contaminados na população.

12.6.25. Em certo distrito, 840 dos 2000 eleitores inquiridos numa sondagem, declararam ir votar no plano A.

- a) Deduza um intervalo a $(1 - \alpha)$ 100% de confiança para a proporção de eleitores do plano A.
- b) Calcule o intervalo a 95% de confiança para p .
- c) Se tivessem sido inquiridos 4000 eleitores e 1680 tivessem declarado preferir o plano A, qual seria agora o intervalo a 95% de confiança.

12.6.26. Dois inquéritos realizados (em 1990 e 1999), relativamente ao consumo de bebidas alcoólicas, em idades entre os 15 e os 35 anos, forneceram os seguintes dados:

Ano	Nº de inquiridores	Consumidores	Não consumidores
1990	4000	1750	2250
1999	5000	2250	2750

Através de um intervalo de confiança, a 98%, indique a veracidade da afirmação: “A percentagem de consumidores de bebidas alcoólicas, em indivíduos com idades compreendidas entre os 15 e os 35 anos, registou um grande aumento na década de 90.”

12.6.27. Com o objectivo de identificar factores de risco de doença coronária analisaram-se duas amostras de 215 homens e de 1140 mulheres, tendo-se registado que 58 dos homens e 217 das mulheres tinham diabetes. Estime um intervalo de confiança a 90% para a diferença das proporções de diabéticos nas duas populações e interprete o resultado obtido.

12.6.28. Duas amostras extraídas de duas populações normais consistindo em 21 e 9 observações têm variâncias dadas por $S^2_1 = 24$ e $S^2_2 = 9$ respectivamente. Elabore um intervalo de confiança para o quociente das variâncias a 95% de confiança.

12.6.29. Para elaborar um estudo sobre o aproveitamento na disciplina de Estatística em dois cursos, analisaram-se as notas obtidas pelos alunos em cada um deles

Curso A	31 <i>alunos</i>	$\bar{x}_1 = 13$	$S^2_1 = 10.3$
Curso B	61 <i>alunos</i>	$\bar{x}_2 = 10.8$	$S^2_2 = 5.7$

- Construa um intervalo de confiança a 95%, para a razão entre variâncias e retire conclusões sobre as dispersões de notas (suponha que as populações têm um comportamento normal).
- Utilizando um intervalo de confiança a 98%, averigúe se as médias nos dois cursos diferem de forma expressiva.

12.6.30. Para averiguar o grau de preferência dos consumidores de duas cidades em relação a uma marca de detergente, foi efectuada uma sondagem onde os inquiridos classificavam o produto numa escala de 0 a 20. Assim foram recolhidas aleatoriamente 21 opiniões de consumidores da cidade A e 11 da cidade B, sendo os resultados obtidos os seguintes:

$\bar{x}_1 = 12.9$	$S^2_1 = 2.1$
$\bar{x}_2 = 14.7$	$S^2_2 = 1.8$

Suponha um comportamento normal na distribuição de opiniões de ambas as cidades.

- Deduza e calcule um intervalo de confiança a 95% para o quociente das variâncias das opiniões e comente o resultado obtido.
- Indique, justificando, de que forma poderia reduzir a amplitude do intervalo anterior.

- c) Verifique, justificando, se há diferença significativa (para um grau de confiança de 99%) entre as classificações médias que os consumidores das duas cidades atribuem ao referido detergente (considere $\sigma^2_1 = \sigma^2_2$)

12.6.31. O tempo de atendimento em um restaurante apresenta variância $\sigma^2 = 0.0015$. Uma amostra aleatória de 12 meses indicou tempo médio de atendimento de $\bar{x} = 12.258$ min.

- a) Construa um intervalo de 95% de confiança para o tempo médio de atendimento no restaurante.
b) Recalcule o intervalo de confiança supondo que a variância não é conhecida e o valor $S^2 = 0.0015$

12.6.32. Uma amostra aleatória de 36 elementos retirados de uma população aproximadamente normal forneceu media de 15,50 e desvio padrão de 1.50. Construir um intervalo de 90% confiança para a média dessa população.

12.6.33. O peso de frangos apresenta variância conhecida igual a 900g. Uma amostra aleatória de 20 unidades indica média igual 508g. Construa um intervalo com 90% de confiança para o peso médio desses frangos.

12.6.34. A característica X em certo artigo produzido em série segue distribuição normal com desvio padrão igual a 3. Com base numa amostra aleatória de 25 unidades que forneceu um valor médio de 48, construa um intervalo de confiança a 95% para o valor médio do universo.

12.6.35. Foi obtido um intervalo de confiança a 95% para a despesa média que cada cliente realiza em determinado restaurante. O resultado obtido foi o seguinte [390; 440]. Critique as interpretações que se lhe afigurem incorrectas:

- a) O próximo cliente vai gastar entre 390 u.m. e 440 u.m.
b) A probabilidade de a despesa média estar entre 390 u.m. e 440 u.m. é 0.95.
c) A probabilidade de o intervalo conter a despesa média é 0.95.
d) Apenas 5% dos clientes do restaurante gasta mais de 440 u.m. com a refeição

12.6.36. Em um processo químico, as características dimensionais do produto resultante segue o modelo normal.

35.2	36.7	37.5	38.2	38.7	39.5
36.3	37.3	37.8	38.3	39.3	40.1

- A partir da amostra apresentada a seguir, defina o limite inferior de um intervalo unilateral de 95% de confiança para a característica dimensional média;
- Construa um intervalo de 90% para a variância da característica dimensional. Depois converta esse intervalo apresentando-o em termos de desvio padrão.

12.6.37. Uma máquina é usada para encher pacotes de leite. O volume segue aproximadamente o modelo normal. Uma amostra de 16 pacotes indicou:

1021	1016	1012	1011	1014	1018	1022	1027
1008	1015	1013	1013	1017	1019	1007	1003

- Construa um intervalo unilateral de 99% com limite inferior para a média;
- Construa um intervalo de 95% para a média;
- Construa um intervalo de 90% para o desvio padrão do volume dos pacotes do leite.
- Imagine que há uma segunda máquina de enchimento para a qual uma amostra de 16 pacotes indicou:

1011	1015	1017	1015	1021	1021	1010	1007
1022	1018	1016	1015	1020	1022	1025	1030

- Construa um intervalo de 95% para a diferença entre as duas médias das máquinas.
- Baseado nos resultados desses cálculos você concluiria que as duas máquinas fornecem o mesmo volume médio?

12.6.38. Em uma indústria química, os engenheiros desejam saber se o alongamento de um composto de borracha permanece inalterado ao passar por uma máquina estrussora. Como o alongamento do composto depende do lote de matéria-prima usado na sua confecção, os dados foram colectados aos pares:

Lote	1	2	3	4	5	6	7	8	9	10
Antes	360	370	380	345	365	380	390	395	385	410
Depois	360	365	355	340	350	370	390	375	375	395

- a) Construa um intervalo de confiança para a diferença entre os pares observados;
- b) Calcula o quociente entre as variâncias dos alongamentos medidos antes e depois do composto passar pela estrussora. Depois construa um intervalo de confiança para esse quociente.

12.6.39. Determine um intervalo de confiança a 80% para a média de uma variável aleatória normal com base na amostra:

9; 14; 10; 12; 7; 3; 11; 12

Qual deveria ser o nível de confiança a utilizar para que a amplitude do intervalo fosse de 2,77?

12.6.40. Para estudar a viabilidade de lançamento de um novo produto no mercado, o gerente de uma grande empresa contrata uma firma de consultoria estatística para estudar a aceitação do produto entre os clientes potenciais. O gerente deseja obter uma estimativa com um erro máximo de 1% com probabilidade 80% e pede ao consultor estatístico que forneça o tamanho de amostra necessário.

- a) De posse das informações dadas, o consultor calcula o tamanho da amostra necessário no pior cenário. O que significa “pior cenário” nesse caso? Qual o tamanho de amostra obtido pelo consultor?
- b) O gerente acha que o custo de tal amostra seria muito alto e autoriza o consultor a realizar um estudo piloto com uma amostra de 100 pessoas para obter uma estimativa da verdadeira proporção. O resultado desse estudo piloto é uma estimativa $P = 0.76$ de aceitação do novo produto. Com base nessa estimativa, o consultor recalcula o tamanho da amostra necessário. Qual é esse tamanho?
- c) Selecionada a amostra com o tamanho obtido no item anterior, obteve-se uma proporção de 72% de clientes favoráveis ao produto. Construa um intervalo de confiança para a verdadeira proporção com nível de confiança de 90%.

12.6.41. Uma amostra aleatória de 250 dispositivos electrónicos apresentou 27 unidades defeituosas. Estime a fracção de não conformes e construa um intervalo de 95% de confiança para o verdadeiro valor da fracção de não conformes.

- 12.6.42.** Em uma pesquisa eleitoral, 60 das 180 pessoas entrevistadas responderam que votariam no candidato da oposição. Essa amostra é suficiente para estimar a verdadeira proporção de eleitores desse candidato, com uma precisão de 0.04 e confiança de 95%?
- 12.6.43.** Numa pesquisa de opinião pública, entre 600 pessoas, 240 responderam que sim a certa pergunta. Determinar o intervalo de confiança para a percentagem populacional que deve responder sim, ao nível de confiança de 95%.
- 12.6.44.** Uma amostra de 300 habitantes de uma cidade mostrou que 180 desejavam a fluoretação da água. Encontrar limites de confiança para a proporção real de habitantes não favoráveis à fluoretação, para:
- a) Nível de confiança de 90%;
 - b) Nível de confiança de 95%.
- 12.6.45.** Numa cidade, entre 1000 residências, 288 assinam TV cabo. Determinar o intervalo de confiança para a proporção de assinantes de TV cabo, nesta cidade, ao nível de confiança de 98%.
- 12.6.46.** Determinar o número mínimo de elementos de uma amostra, se desejamos estimar a média populacional, com 95% de confiança, e erro amostral de 0,4, sendo que de uma amostra piloto com 70 elementos obtivemos variância de 36.
- 12.6.47.** Suponhamos que se pretenda estimar a renda média por família numa grande cidade. Com base em informações passadas, admite-se que o desvio padrão das rendas das famílias é de 200,00mt. Qual deve ser o tamanho da amostra, a fim de que o erro de estimativa da renda média seja no máximo de 10,00mt, com probabilidade igual a 96%?
- 12.6.48.** Para se estimar a proporção de pessoas interessadas em água fluorada, qual o tamanho da amostra, para se estar confiante em 95% de que o erro seja de no máximo 1%?
- 12.6.49.** Um fabricante de peças acredita que aproximadamente 5% de seus produtos são defeituosos. Se ele deseja estimar a verdadeira percentagem,

dentro de 0,5% de erro, com uma probabilidade de estar certo de 90%, qual o tamanho da amostra a ser tomada?

12.6.50. Um analista de um departamento de pessoal selecciona aleatoriamente os registos de 16 empregados a tempo parcial e acha que a taxa de salários por hora é 75,00 mt. Supõe-se que os salários da firma sejam distribuídos normalmente. Se o desvio padrão dos salários conhecido é igual a 10,00mt, estimar a taxa média de salários na firma usando um intervalo de confiança de 95%.

12.6.51. Para uma mostra de 50 firmas tomada de uma determinada indústria, o número médio de empregados por firma é 420,4 com um desvio padrão na amostra de 55,7. Nesta indústria, há um total de 380 firmas.

- a) Determinar o intervalo de confiança de 90%, para estimar o número médio de trabalhadores por firma da indústria.
- b) Qual o erro absoluto de estimação obtido?

12.6.52. Um departamento de manutenção recebeu um carregamento de 100 máquinas defeituosas. Sabe-se por experiência anterior que o desvio padrão em relação ao tempo necessário para conserto é de 15 minutos. Estimar o tempo médio, por máquina, necessário para consertar as máquinas do carregamento, usando um nível de confiança de 90%. Para tal finalidade tomou se uma amostra aleatória de 10 máquinas e o tempo médio necessário para conserto de 85,0 minutos.

12.6.53. Uma centena de componentes foi ensaiada e 93 deles funcionaram mais de 500 horas. Determinar um intervalo de confiança de 95% para a proporção.

12.6.54. Um mini - mercado pretende estimar o número médio de litros de água que vende diariamente (fenómeno com comportamento normal), para efeitos de controlo de encomendas a fornecedores. Ao fim de 20 dias de negócio, verificou que em média vendia 32 litros de água/dia, sendo o desvio padrão desta amostra igual a 12 litros. Admitindo a normalidade, calcule os limites de confiança para um grau de confiança de 95%.

12.6.55. Para avaliar a precisão de uma balança de laboratório, pesa-se repetidas vezes um objecto padrão de peso conhecido igual a 10 gramas. As leituras da balança têm distribuição normal com média desconhecida (essa média é 10 gramas, se a balança é equilibrada). Sabe-se que o desvio padrão das leituras é 0,0002 grama. Pesa-se o objecto 5 vezes e o resultado médio é 10,0023 gramas.

- a) Estabeleça um intervalo de 95% de confiança para a média de repetidas pesagens do objecto.
- b) Quantas observações ou medidas devem entrar no cálculo da média, a fim de que se obtenha uma margem de $\pm 0,0001$ de erro com 95% de confiança?

12.6.56. Suponha que estejamos interessados em estimar a percentagem de consumidores de certo produto. Se a amostra de tamanho 300 forneceu 100 indivíduos que consomem o dado produto, determine:

- a) O intervalo de confiança de p , a proporção de pessoas que consomem o produto, com coeficiente de 95% (interprete o resultado).
- b) O tamanho da amostra para que o erro da estimativa não exceda a 2% com probabilidade de 95% (interprete o resultado).

12.6.57. Numa pesquisa sobre a opinião dos moradores de duas cidades, A e B, com relação a um determinado projecto, obteve-se a tabela abaixo. Utilize o intervalo confiança para avaliar a diferença entre os percentuais de favoráveis nas duas cidades.

Cidade	A	B
Nº de entrevistados	400	600
Nº favoráveis	180	350

12.6.58. Um estudo de saúde envolve 1000 mortes seleccionados aleatoriamente, dentre as quais 131 causadas por intoxicação alimentícia.

- a) Com os dados amostrais, construa um intervalo de confiança de 99% para a proporção de mortes causadas por intoxicação.
- b) Utilizando os dados amostrais como estudo piloto, determine o tamanho da amostra necessário para estimar a proporção de mortes por

intoxicação em uma cidade. Admita um nível de confiança de 95%, em que o erro da estimativa não supere 0,01.

- c) Sabe-se que a cidade tem cerca de 250.000 habitantes. Você acha que esse dado poderia ser utilizado para melhorar a estimativa do tamanho da amostra? Como?

12.6.59. Foi realizada uma pesquisa envolvendo uma amostra de 600 pacientes de um certo hospital. Cada um desses pacientes foi submetido a uma série de exames clínicos e, entre outras coisas, mediu-se o Índice Cardíaco (em litros/min/m²) de todos eles. Os 600 pacientes foram então classificados, de forma aleatória, em 40 grupos de 15 pacientes cada. Para um desses grupos os valores medidos do Índice Cardíaco foram:

405, 348, 365, 291, 135, 260, 300, 155, 34, 294, 758, 472, 559, 143, 172

- a) Com base nos valores acima, construa um Intervalo de Confiança para o valor médio do Índice Cardíaco ao nível de 95%.
- b) Se para cada um desses 40 grupos de 15 pacientes fosse construído um Intervalo de confiança para μ ao nível de 95%, quantos desses intervalos se espera que não conteriam a verdadeira média populacional no seu interior? Porquê?
- c) Determine o intervalo de confiança para variancia.

12.6.60. Um tratamento a base de cloridrato de Metformina é aplicado a doze pacientes diabéticos. Os teores de glicose antes e depois de duas semanas de tratamento estão apresentados na tabela seguir:

Antes	129	132	139	132	148	126	128	137	131	118	136	116
Depois	122	127	134	126	144	128	122	138	125	110	130	133

Determine Intervalos de confiança de 99% para:

- a) Os níveis de Glicose antes do tratamento;
- b) Os níveis de Glicose depois do tratamento;
- c) A diferença entre os teores de glicose antes e depois do tratamento

12.6.61. Um grupo de doentes é alimentado com uma determinada dieta e um segundo grupo de 50 doentes é alimentado com uma dieta diferente. Depois de um certo período de tempo, o aumento médio no peso para o primeiro grupo foi

de 124.7 g, com um desvio padrão de 9g. Já para o segundo grupo, aumento médio no peso foi de 130.8g, com um desvio padrão de 12g.

- a) Obtenha um Intervalo de confiança de 90% para a diferença entre as médias populacionais dos aumentos de peso nos dois grupos.
- b) Determine o intervalo de confiança de 90% para o quociente das varianças e conclua a respeito.

12.6.62. Um estudo de saúde envolve 1000 mortes seleccionadas aleatoriamente, dentre as quais 131 causadas por intoxicação alimentícia.

- a) Com os dados amostrais, construa um intervalo de confiança de 99% para a proporção de mortes causadas por intoxicação.
- b) Utilizando os dados amostrais como estudo piloto, determine o tamanho da amostra necessário para estimar a proporção de mortes por intoxicação em uma cidade. Admita um nível de confiança de 95%, em que o erro da estimativa não supere 0,01.
- c) Sabe-se que a cidade tem cerca de 250.000 habitantes. Você acha que esse dado poderia ser utilizado para melhorar a estimativa do tamanho da amostra? Como?

12.6.63. A Leishmaniose Visceral é uma doença perigosa e que, se não for tratada correctamente, pode levar a óbito. Todo caso diagnosticado de Leishmaniose Visceral deve ser notificado às autoridades de saúde. Em estudo sobre o número de dias entre o início dos sintomas da Leishmaniose Visceral e a notificação do caso às autoridades, uma pesquisadora deseja estimar o número médio de dias entre os sintomas e a notificação usando um intervalo de 95% de confiança. Sabendo que ela gostaria que o erro de estimação fosse a metade do desvio-padrão do número de dias e supondo que o número de dias entre o início dos sintomas e a notificação tenha distribuição Gaussiana. Quantos casos de Leishmaniose Visceral, no mínimo, ela deve estudar?

12.6.64. Suponhamos dois grupos A e B, formados respectivamente por 50 e 100 doentes semelhantes. Ao grupo A foi administrado um novo tipo de medicamento para dormir e, ao grupo B, um medicamento convencional. Os doentes do grupo A estiveram uma média de 7,83 horas a dormir, com um desvio padrão de 0,24 horas; os doentes do grupo B estiveram $6,75 \pm 0,30$ horas a dormir.

- a) Encontre os limites de confiança a 95% e 99% para a diferença do número médio de horas a dormir, induzidas por ambos os medicamentos.
- b) Verifique se o efeito do medicamento foi igual para os dois grupos ($1 - \alpha = 0.95$).

12.6.65. Um banco possui 800 terminais de auto-atendimento instalados no estado de SC. Avaliando 48 terminais, 6 apresentaram defeitos. Estime a proporção de terminais com defeitos a um nível de confiança de 95%.

12.6.66. Determine o tamanho de amostra necessário para estimar o volume médio de vendas de carros novos nacionais entre as concessionárias, fixando um nível de confiança de 99% para um erro de estimação igual a 1 automóvel. É conhecido que existem 200 concessionárias na região em estudo. Em uma pesquisa similar feita 5 anos antes, o desvio padrão amostral foi igual a 2,8. Supor que foi feita uma amostragem aleatória simples.

12.6.67. Líderes estudantis de uma faculdade querem conduzir uma pesquisa para determinar a proporção de estudantes a favor de uma mudança no horário de aulas. Como é impossível entrevistar todos os 2000 estudantes em um tempo razoável, decide-se fazer uma amostragem aleatória simples dos estudantes:

- a) Determinar o tamanho de amostra (número de estudantes a serem entrevistados) necessário para estimar a proporção com um erro máximo de 0,05 e nível de confiança de 95%. Assumir que não há nenhuma informação a priori disponível para estimar a proporção.
- b) Os líderes estudantis também querem estimar a proporção de estudantes que sentem que a representação estudantil atende adequadamente as suas necessidades. Com um erro máximo de 7% e nível de confiança de 95%, determinar o tamanho de amostra para estimar a proporção. Utilizar a informação de uma pesquisa similar conduzida a alguns anos, quando 60% dos estudantes acreditavam que estavam bem representados.
- c) Qual o tamanho de amostra adequado para atingir ambos os objectivos da pesquisa?

12.6.68. Uma rede de lanchonetes deseja estimar a quantia média que cada cliente gasta por lanche. Foram seleccionados dados de uma amostra de 22

clientes que revelou uma quantia média de 15 u.m com um desvio-padrão de 5. Construir um intervalo de confiança de 95% para a média populacional.

12.6.69. Um centro de ortodontia deseja conhecer a estimativa do tempo médio que um membro da equipe gasta para atender a cada paciente. Suponha que uma amostra de 38 especialistas revelou que a média foi de 45 minutos com um desvio padrão de 6 minutos. Determine um intervalo de 99% de confiança para o parâmetro.

12.6.70. Um grupo de pesquisadores pretende estimar o real salário médio de trabalhadores bancários. Para tal seleccionou-se aleatoriamente 100 trabalhadores de diferentes bancos, tendo-se obtido um salário médio de 1500 u.m e desvio padrão de 1200 u.m. Os responsáveis pela pesquisa consideram que uma boa estimativa para o salário médio mensal deve ter no máximo um erro de 200 u.m. Determine a probabilidade do erro máximo ser de 200 u.m.

12.6.71. Uma empresa que presta serviços de acessória a outras empresas está interessada em comparar a taxa de reclamações sobre os seus serviços em dois dos seus escritórios em duas cidades diferentes. Suponha que a empresa tenha seleccionado aleatoriamente 100 serviços realizados pelo escritório da cidade A e foi constatado que em 12 deles houve algum tipo de reclamação. Já do escritório da cidade B foram seleccionados 120 serviços e 18 receberam algum tipo de reclamação. Construa um intervalo de confiança a 98% para a diferença entre as taxas de reclamações das duas cidades.

12.6.72. Suponha que as produções (em gramas) de castanhas, em intervalos de tempo fixos, aleatoriamente seleccionados de duas máquinas M1 e M2 de uma fábrica se podem considerar normais. Os pesos obtidos em duas amostras permitiram determinar as quantidades seguintes:

M1	M2
$\sum_{i=1}^8 x_i = 80.8$	$\sum_{i=1}^9 y_i = 96.3$
$\sum_{i=1}^8 x_i^2 = 816.664$	$\sum_{i=1}^9 (y_i - \bar{y})^2 = 0.549$

- a) Indique uma estimativa pontual para a produção média de cada máquina e respectivas variâncias.
- b) Verifique se é plausível considerar que a variabilidade em gramas da produção das duas máquinas é idêntica (use 95% de confiança).
- c) Construa um intervalo de confiança a 95% para a diferença das produções médias entre as duas máquinas.

12.6.73. Um teste de auditoria, para estabelecer com que frequências ocorrem falhas no processamento de determinado procedimento de controlo interno, está para ser feito. O auditor decide que a precisão deve ser igual a 2% do desvio padrão máximo do erro tolerável permitido. Qual deve ser o tamanho de amostra a um nível de confiança de 97%, sabendo que as falhas no processamento seguem uma distribuição normal?

12.6.74. Um fabricante de insecticida descobriu, em uma pesquisa, que muitos consumidores da categoria achavam que o cheiro do seu produto era suave demais, e enfraquecia o seu apelo publicitário de produto fortíssimo contra os insectos. Após alguns *sniff tests*, foi seleccionado um novo aroma para o produto. Faltava testar o novo aroma no produto em uso pela consumidora. O fabricante solicitou à empresa de pesquisa que fizesse um teste de produto *in home* em que uma amostra de 15 consumidores usaria durante uma semana o produto com o cheiro actual, enquanto uma outra amostra, também de 12 consumidores com o mesmo perfil, usaria o produto com o novo cheiro. A avaliação em escala de 7 pontos resultou nos seguintes parâmetros.

	Aroma actual	Aroma novo
Média	5.94	6.27
Desvio Padrão	5.22	2.39

- a) Assumindo que as populações são normais e as variâncias são diferentes, qual é a probabilidade de o intervalo $[-3.3401; 2.6801]$ conter a diferença entre os valores esperados das duas populações?
- b) Assumindo que as populações são normais e as variâncias são iguais, qual é a probabilidade de o intervalo $[-3.868; 3.208]$ conter a diferença entre os valores esperados das duas populações?

12.6.75. Uma empresa que presta serviços de acessória a outras empresas está interessada em comparar a taxa de reclamações sobre os seus serviços em dois dos seus escritórios em duas cidades diferentes. Suponha que a empresa tenha seleccionado aleatoriamente serviços realizados pelo escritório da cidade A e da cidade B, tendo obtido a 95% de confiança as seguintes taxas de reclamações:

Cidade	Cidade A	Cidade B
<i>Taxas de reclamações</i>	0.54 ± 0.0977	0.50 ± 0.0800

Construa um intervalo de confiança a 98% para a diferença entre as taxas de reclamações das duas cidades.

12.6.76. Uma votação será feita entre os residentes de uma cidade e a região rural ao redor desta cidade para determinar se um projecto químico deverá ser construído. A construção é dentro dos limites da cidade e por esta razão muitos eleitores do campo sentem que o projecto passará por causa da grande proporção dos eleitores da cidade, os quais são favoráveis. Para determinar se existe diferença na proporção de eleitores da cidade e do campo a favor do projecto, uma amostragem foi feita. Supondo que 80 de 200 eleitores da cidade não são a favor do projecto e 240 de 500 eleitores do campo são a favor, construa um intervalo a 97% de confiança para a diferença das proporções de eleitores favoráveis ao projecto.

12.6.77. Uma das maneiras de medir o grau de satisfação dos empregados de uma mesma categoria quanto a política salarial é por meio da variabilidade dos seus salários. A Fábrica A diz ser mais coerente na política salarial do que a Fabrica B. Para verificar essa afirmação, sorteou-se uma amostra de 17 funcionários não especializados da fábrica A e 15 da fábrica B, obtendo-se varianças de 1000 u.m e 1200 u.m respectivamente. Usando um nível de confiança de 95%, qual seria a sua conclusão?

12.6.78. Um teste de auditoria, para estabelecer com que frequências ocorrem falhas no processamento de determinado procedimento de controlo interno, está para ser feito. O auditor decide que a taxa máxima de erro tolerável permitida é de 5%.

- c) Que tamanho de amostra é necessário para atingir uma precisão de amostra de 2%, com 99% de confiança?
- d) Qual seria sua resposta em (a) se a taxa máxima tolerável de erro fosse 10%?

13. TESTE DE HIPÓTESES

Um importante tipo de problema em Inferência Estatística é determinar se uma amostra pode ter vindo de uma população tendo uma distribuição parcial ou completamente especificada. Por exemplo, se sabemos que uma amostra veio de uma distribuição normal, é razoável dizer que ela veio de uma distribuição com média μ_0 ?

Ou, se duas amostras vieram de distribuições normais, é razoável dizer que estas vieram de distribuições que têm médias iguais? Como estatísticas como estas são variáveis aleatórias que têm suas próprias distribuições de probabilidade, afirmações sobre seus parâmetros devem ser feitas em termos de probabilidade.

No caso das inferências através dos intervalos de confiança, busca-se “cercar” o parâmetro populacional desconhecido. Aqui formula-se uma hipótese quanto ao valor do parâmetro populacional, e pelos elementos amostrais faz-se um teste que indicará a aceitação ou rejeição da hipótese formulada

13.1. Hipótese estatística

Uma hipótese estatística é uma afirmação sobre a distribuição (ou parâmetros) de uma ou mais variáveis aleatórias. Uma hipótese estatística pode ser verdadeira ou não.

Por exemplo, suponha que \bar{X} seja a média de uma amostra de tamanho n retirada de uma distribuição $N(\mu, \sigma^2)$, onde σ^2 é conhecida e μ é desconhecida. Suponha que se deseje verificar se é razoável que esta amostra tenha vindo de uma população $N(\mu_0, \sigma^2)$ considerando a possibilidade de que esta poderia ter vindo de alguma distribuição normal $N(\mu_1, \sigma^2)$, onde $\mu_1 \neq \mu_0$. Podemos abreviar esta questão dizendo que desejamos testar a hipótese estatística:

$$H_0: \mu = \mu_0$$

Contra a alternativa

$$H_1: \mu \neq \mu_0$$

Usando a amostra de tamanho n e a média \bar{X} .

H_0 é chamada de **Hipótese Nula** e H_1 é chamada de **Hipótese Alternativa**.

- **Hipótese Nula:** é uma afirmação sobre o valor de um parâmetro populacional, deve conter a condição de igualdade, ou seja:

$$H_0: \theta = \theta_0 \text{ ou } H_0: \theta \geq \theta_0 \text{ ou } H_0: \theta \leq \theta_0$$

Ao fazermos efectivamente o teste, trabalhamos com a hipótese de que o parâmetro é igual a um valor específico. Testamos a hipótese nula directamente no sentido de que, supondo-a verdadeira, procuramos chegar a uma conclusão que nos leve a rejeitar H_0 ou não rejeitar H_0 .

- **Hipótese alternativa:** é uma afirmação que deve ser verdadeira se a hipótese nula é falsa. A hipótese alternativa comporta apenas uma das três formas:

$$H_1: \theta \neq \theta_0 \text{ ou } H_1: \theta > \theta_0 \text{ ou } H_1: \theta < \theta_0$$

Normalmente, H_1 é a negação de H_0 , embora nem sempre seja necessariamente assim.

13.2. Estatística de teste

Uma estatística de teste é uma função das observações amostrais cujo valor vai determinar a conclusão a retirar do teste estatístico. No caso de se testar um parâmetro, a estatística de teste é, habitualmente, um estimador desse parâmetro.

13.3. Regra de decisão estatística

A regra de decisão estatística é o princípio que determina a conclusão a retirar (rejeitar ou não H_0) a partir da comparação do valor da estatística de teste com um ou mais valores críticos. Os valores críticos determinam o conjunto de valores da estatística de teste que conduz à rejeição da hipótese nula. Este conjunto de valores denomina-se região crítica ou região de rejeição da hipótese nula (ou de teste). Os valores que não pertencem à região de rejeição pertencem à região de aceitação.

13.4. Erros de inferência

Ao testarmos uma hipótese nula, chegamos a uma conclusão: rejeitá-la ou não rejeitá-la. Como na inferência estatística parte-se da amostra para a população (em consequência, do particular para o geral), pode incorrer-se em erros na tomada de decisão (quando forem tomadas decisões incorrectas).

Há dois tipos diferentes de erro que podemos cometer. A tabela seguinte resume as diferentes possibilidades e mostra que tomamos uma decisão correcta quando, ou rejeitamos uma hipótese nula que é falsa, ou deixamos de rejeitar uma hipótese nula que é verdadeira. Todavia, cometemos um erro quando rejeitamos uma hipótese nula verdadeira, ou deixamos de rejeitar uma hipótese nula falsa.

		Natureza da situação	
		H_0 é verdadeira	H_0 é falsa
Decisão	Rejeita-se a H_0	Erro Tipo I	Decisão correcta
	Não se rejeita-se a H_0	Decisão correcta	Erro Tipo II

O erro tipo I consiste em rejeitar a hipótese nula quando esta é verdadeira. O erro tipo I não é um cálculo mal feito ou uma fase do processo mal desempenhada, é um erro que pode ocorrer como consequência casual de um acontecimento raro. A probabilidade de rejeitar a hipótese nula quando esta é verdadeira é inferior ou igual ao nível de significância α :

$$P(\text{erro tipo I}) = P(\text{rejeitar } H_0 \mid H_0 \text{ verdadeiro}) \leq \alpha$$

O valor de α é tipicamente pré-determinado sendo comuns as escolhas de 0.05 e 0.01.

O erro tipo II consiste em não rejeitar a hipótese nula quando ela é falsa. Usa-se o símbolo β para representar a probabilidade de um erro tipo II:

$$\beta = P(\text{erro tipo II}) = P(\text{não rejeitar } H_0 \mid H_0 \text{ falsa})$$

13.4.1. Relação entre os erros tipo I e tipo II

Vimos que α é a probabilidade de um erro tipo I (rejeitar uma hipótese nula verdadeira) e β é a probabilidade de um erro tipo II (não rejeitar uma hipótese nula falsa). Uma das etapas do processo de teste de hipóteses envolve a escolha do nível de significância α , no entanto, não seleccionamos β . Seria óptimo se pudéssemos ter sempre $\alpha = 0$ ou $\beta = 0$, mas isto não é possível; devemos, pois, procurar controlar as probabilidades de erro α e β .

O tamanho da amostra n , α e β estão todos interrelacionados, de forma que, escolhidos quaisquer dois deles, o terceiro está automaticamente determinado.

Poderíamos, pois, escolher α e β (e o tamanho n da amostra estaria determinado), mas a prática comum na pesquisa consiste em determinar previamente os valores de α e n , de modo que o valor de β fica determinado. Escolhe-se então um tamanho n de amostra tão grande quanto razoável em face do tempo, custo e outros factores relevantes. Valem as seguintes considerações de ordem prática:

1. Para α fixo, um aumento do tamanho n da amostra ocasiona uma redução de β , isto é, uma amostra maior reduz a hipótese de cometermos o erro de não rejeitar a hipótese nula quando ela é falsa;
2. Para um tamanho n , fixo, de amostra, uma diminuição de α acarreta um aumento de β ; reciprocamente, um aumento de α acarreta uma diminuição de β ;
3. Para reduzir α e β , devemos aumentar o tamanho n da amostra.

13.5. Valor de p ou $p - value$

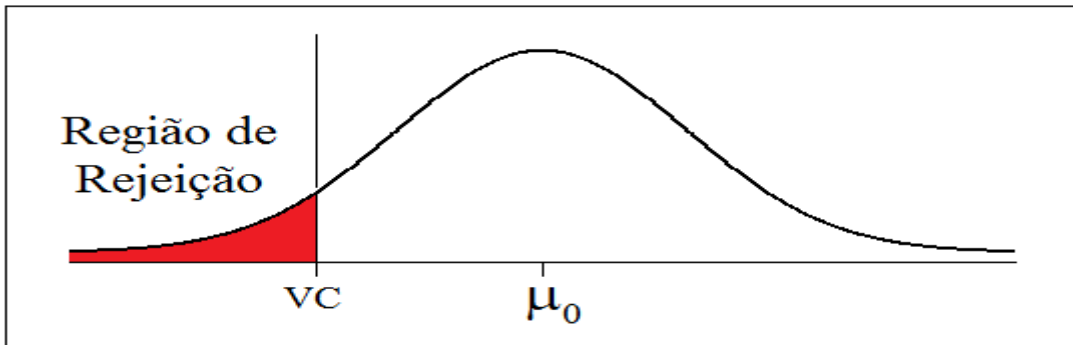
O $p - value$ é o menor nível de significância α , a partir do qual se começa a rejeitar a hipótese nula, isto é, se $\alpha \geq p - value$ então deve-se rejeitar a H_0 .

Tipicamente, em muitas Ciências, resultados que produzem $p \leq 0,05$ são considerados estatisticamente significantes, mas lembre que este nível de significância ainda envolve uma probabilidade de erro razoavelmente grande (5%). Resultados que são significantes ao nível de $p \leq 0,01$ são estatisticamente significantes e níveis de $p \leq 0,005$ ou $p \leq 0,001$ considerados “altamente” significantes. Estas classificações são meramente arbitrárias e são convenções informalmente baseadas em experiência de pesquisa de modo geral.

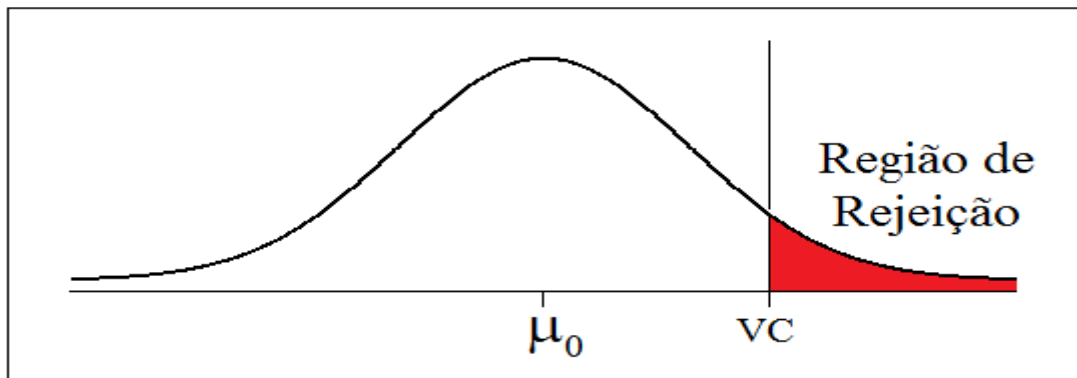
13.6. Região de Rejeição ou região crítica (RC)

Região de Rejeição ou região crítica, é formada pelo conjunto de valores que levam à rejeição de H_0 . É subconjunto do espaço paramétrico θ . A região que não leva à rejeição de H_0 será chamada de região de não rejeição. O valor que delimita a região de rejeição e a região de não rejeição será chamado de valor crítico.

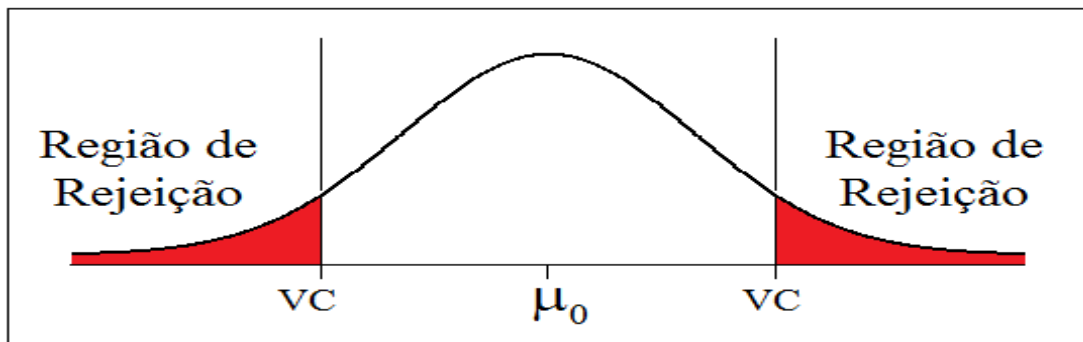
- Região de rejeição para o teste unicaudal para a média (cauda inferior)



- Região de rejeição para o teste unicaudal para a média (cauda superior)



- Região de rejeição para o teste bicaudal para a média



Nota: existem 3 procedimentos para realizar um teste de hipótese ao nível de significância α

1. **Com base na região crítica RC:** rejeitar H_0 se o valor da estatística do teste encontrar –se na região crítica;
2. **Através do p-value:** Rejeitar H_0 se o valor de $p - value \leq \alpha$;
3. **Através do intervalo de confiança:** rejeitar H_0 se o valor do parâmetro específico em H_0 não pertencer ao intervalo de confiança

13.7. Passos de um teste de hipóteses através da RC

1. Identificar a afirmação ou hipótese específica a ser testada e colocá-la em forma simbólica. Dar a forma simbólica que deve ser verdadeira quando a afirmação original é falsa. Das duas expressões simbólicas obtidas, a hipótese nula H_0 é a que contém a condição de igualdade e a hipótese alternativa H_1 é a outra afirmação;
2. Escolher o nível de significância α . São comuns os valores 0,05 e 0,01;
3. Identificar a estatística relevante para este teste e determinar a sua distribuição amostral;
4. Determinar a estatística de teste, os valores críticos e a região crítica. Esboçar um gráfico e incluir a estatística de teste, o(s) valor(es) crítico(s) e a região ou regiões críticas;
5. Rejeitar H_0 se a estatística de teste está na região crítica. Não rejeitar H_0 se a estatística de teste não está na região crítica;
6. Um teste estatístico é conduzido para se determinar se a amostra traz evidência suficiente para se rejeitar H_0 e, assim, concluir que H_1 é verdadeira. Ou seja, o teste estatístico é usado para se concluir a favor de H_1 ao se concluir que H_0 pode ser rejeitada. Neste sentido, testar uma hipótese pode ser visto como “testar a hipótese nula”.

13.8. Teste de hipótese para média, quando a variância é conhecida

Suponha que X é uma variável aleatória com média μ desconhecida e variância σ^2 conhecida. E queremos testar a hipótese de que a média é igual a um certo valor especificado μ_0 .

O teste de hipótese pode ser formulado como segue:

$$\text{Teste bilateral } \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$$

$$\text{Teste unilateral á direita } \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases} \text{ ou } \begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$$

$$\text{Teste unilateral á esquerda } \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases} \text{ ou } \begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$$

Para testar a hipótese, toma-se uma amostra aleatória de n observações e se calcula a estatística:

$$Z_0 = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

A hipótese H_0 é rejeitada se $|Z_0| > |Z_{crítico}|$.

Exemplo 13.1 (Teste de hipótese para média)

Uma indústria produz lâmpadas cuja duração segue uma distribuição $N(800; 1600)$. Testar a hipótese de que $\mu = 800$ contra a alternativa de $\mu \neq 800$ se uma amostra aleatória de 30 lâmpadas tem um tempo médio de vida de 788 horas. Um

Solução:

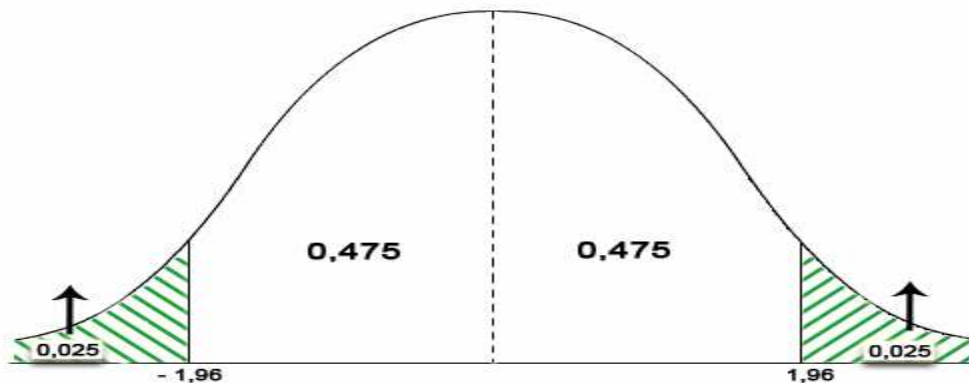
1. $\begin{cases} H_0: \mu = 800 \\ H_1: \mu \neq 800 \end{cases}$
2. $\alpha = 0.05$
3. $Z \sim N(0,1)$

$$Z_0 = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{788 - 800}{40/\sqrt{30}} = -1.64$$

4. $\alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025$ pois o teste é bilateral. Para uma área de 0.475,

$$Z_{crítico} = Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$$

Desenhando a curva, temos:



5. Como $-Z_{\text{crítico}} = -1.96 < Z_{\text{calculado}} = -1.64 < Z_{\text{crítico}} = 1.96$, aceita-se a hipótese nula H_0 .
6. Conclusão: A um nível de significância de 5%, pode-se concluir que a diferença não é significativa.

13.9. Teste de hipótese para média, quando a variância é desconhecida

Suponha que X é uma variável aleatória com média μ e variância σ^2 desconhecida. Para testar a hipótese de que a média é igual a um certo valor especificado μ_0 formulamos as seguintes hipóteses:

$$\text{Teste bilateral } \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$$

$$\text{Teste unilateral á direita } \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases} \text{ ou } \begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$$

$$\text{Teste unilateral á esquerda } \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases} \text{ ou } \begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$$

Como a variância é desconhecida, é necessário fazer a suposição adicional de que a variável tenha distribuição Normal.

Essa suposição é necessária para poder desenvolver a estatística do teste; contudo, os resultados ainda serão válidos se o afastamento da normalidade não for forte.

- Quando $n < 30$ usa-se a distribuição de t-student para construir a estatística do teste:

$$t_0 = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

A hipótese $H_0: \mu = \mu_0$ é rejeitada se $|t_0| > \left| t_{\frac{\alpha}{2}; n-1} \right|$.

- Quando $n \geq 30$ usa-se a distribuição aproximada de Z para construir a estatística do teste:

$$Z_0 = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

Exemplo 13.2 (Teste de hipótese para média)

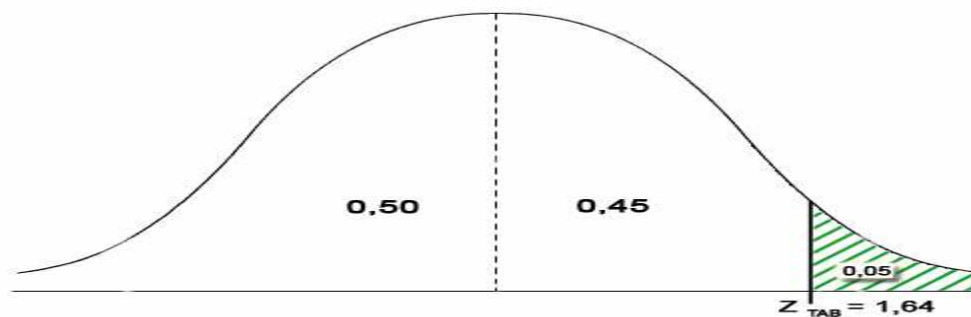
Uma amostra de 36 elementos de uma variável X normalmente distribuída forneceu: $\bar{X} = 42,3$ e $S = 5,2$. Testar, no nível de significância 0,05, a hipótese de que $\mu > 40$.

Solução:

1. $\begin{cases} H_0: \mu = 40 \\ H_1: \mu > 40 \end{cases}$
2. $\alpha = 0.05$
3. $Z \sim N(0,1)$

$$Z_0 = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{42.3 - 40}{5.2/\sqrt{36}} = 2.65$$

4. $\alpha = 0.05$ pois o teste é unilateral direito. Para uma área de 0.45, $Z_{crítico} = Z_\alpha = Z_{0.05} = 1.64$. Desenhando a curva, temos:



5. Como $Z_{crítico} = 1.64 < Z_{calculado} = 2.65$, rejeita-se a hipótese nula H_0 .
6. Conclusão: A um nível de significância de 5%, pode-se concluir que a diferença é significativa.

Exemplo 13.3 (Teste de hipótese para média)

Uma amostra de 20 elementos de uma variável X normalmente distribuída forneceu: $\bar{X} = 53,4$ e $S = 7,5$. Testar, no nível de significância 0,05, a hipótese de que $\mu = 50$.

Solução:

1. $\begin{cases} H_0: \mu = 50 \\ H_1: \mu \neq 50 \end{cases}$
2. $\alpha = 0.05$

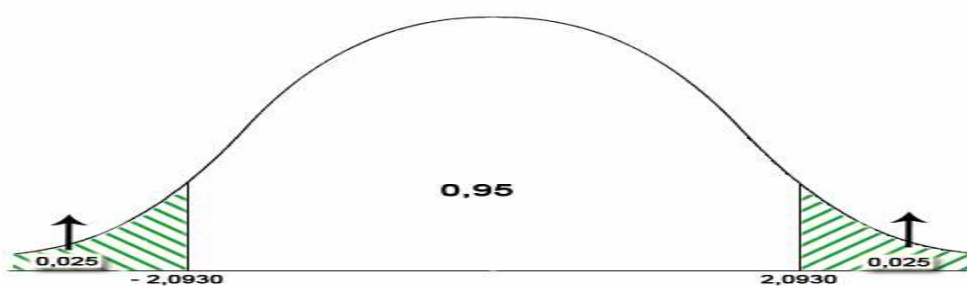
3. $t \sim t_{n-1}$

$$t_0 = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{53.4 - 50}{7.5/\sqrt{20}} = 2.027$$

4. $\alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025$ pois o teste é bilateral.

$$t_{\frac{\alpha}{2}; n-1} = t_{0.025; 19} = 2.0930$$

Desenhando a curva, temos:



5. Como $-t_{\text{crítico}} = -2.0930 < t_{\text{calculado}} = 2.027 < t_{\text{crítico}} = 2.0930$, aceita-se a hipótese nula H_0 .
6. Conclusão: A um nível de significância de 5%, pode-se concluir que a diferença não é significativa.

13.10. Teste de hipótese da diferença entre duas médias populacionais

A hipótese nula (H_0) usualmente testada é a de que as duas amostras tenham sido obtidas de populações com médias iguais, ou seja $(\mu_1 - \mu_2)_0 = 0$.

O uso da distribuição normal para duas amostras independentes é utilizada sempre que as variâncias populacionais sejam conhecidas, caso contrário utiliza-se a distribuição *t-student*.

Estudaremos os seguintes casos:

13.10.1. Teste de hipótese da diferença entre duas médias populacionais com variâncias conhecidas

Consideremos duas populações normais com médias μ_1 e μ_2 e variâncias σ^2_1 e σ^2_2 , sendo n_1 e n_2 duas amostras independentes obtidas respectivamente, dessas populações, e \bar{x}_1 e \bar{x}_2 suas médias.

A estatística do teste a ser usada para as hipóteses $\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$

é:

$$Z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}}$$

Exemplo 13.4 (Teste de hipótese para diferença entre duas médias)

Duas marcas de veículos pretendem comparar o desempenho de seus modelos populares. Para isso, a marca **A** selecionou $n = 12$ veículos de sua produção e fez um teste de consumo. A marca **B** também retirou uma amostra com $n = 12$ veículos e realizou o mesmo teste. As empresas afirmam que ambas têm a mesma variabilidade, $\sigma^2 = 0.81$ (os dados estão em km/litro). Conclua a um nível de confiança de 5%.

A	13.5	12.8	11.4	10.9	11.9	12.3	10.7	11.9	10.9	11.5	11.8	12.1
B	12.8	12.8	13.6	13.8	10.1	11.1	11.9	11.4	10.8	12.2	12.4	12.5

Solução:

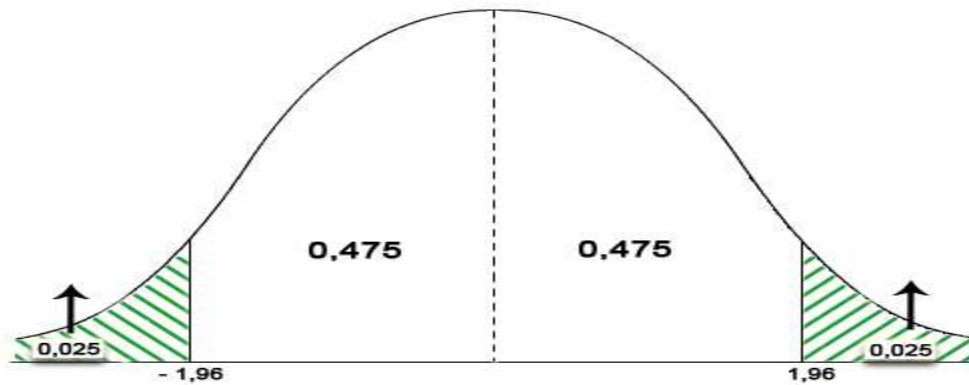
- $\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$
- $\alpha = 0.05$
- $Z \sim N(0,1)$

$$Z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}} = \frac{12.12 - 11.81}{\sqrt{\frac{0.81}{12} + \frac{0.81}{12}}} = 0.844$$

- $\alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025$ pois o teste é bilateral. Para uma área de 0.475,

$$Z_{crítico} = Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$$

Desenhando a curva, temos:



5. Como $-Z_{\text{crítico}} = -1,96 < Z_{\text{calculado}} = 0,844 < Z_{\text{crítico}} = 1,96$, aceita-se a hipótese nula H_0 .
6. Conclusão: A um nível de significância de 5%, pode-se concluir que as duas marcas de veículos tem o mesmo desempenho.

13.10.2. Teste de hipótese da diferença entre duas médias populacionais com variâncias desconhecidas e iguais

Quando há duas populações normais com médias μ_1 e μ_2 e variâncias σ^2_1 e σ^2_2 desconhecidas, as hipóteses para testar se as médias são iguais, são as seguintes:

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

O procedimento do teste irá depender de que $\sigma^2_1 = \sigma^2_2$. Se essa suposição for razoável, então calcula-se a variância combinada:

$$S_p^2 = \frac{(n_1 - 1) * S_1^2 + (n_2 - 1) * S_2^2}{n_1 + n_2 - 2}$$

E a seguir calcula-se a estatística:

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_p * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

A H_0 será rejeitada se $|t_0| > t_{\frac{\alpha}{2}, n_1 + n_2 - 2}$

Exemplo 13.5 (Teste de hipótese para diferença entre duas médias)

O Estatístico contratado para fazer os cálculos (**exemplo 13.4**) não confiando no valor da variância $\sigma^2 = 0.81$, decidiu refazer as contas considerando as variâncias desconhecidas, porém iguais. Supondo que usou o mesmo nível de significância que conclusões obteve?

Solução:

$$1. \begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

$$2. \alpha = 0.05$$

$$3. t \sim t_{n_1+n_2-2}$$

$$S_p^2 = \frac{(n_1 - 1) * S_1^2 + (n_2 - 1) * S_2^2}{n_1 + n_2 - 2} = \frac{11 * 1.1118^2 + 11 * 0.8163^2}{12 + 12 - 2} = 0.9512$$

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}} = \frac{12.12 - 11.81}{\sqrt{\frac{0.9512}{12} + \frac{0.9512}{12}}} = 0.7786$$

$$4. \alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025 \text{ pois o teste é bilateral.}$$

$$t_{\text{crítico}} = t_{\frac{\alpha}{2}} = t_{0.025,22} = 2.0739$$

5. Como $-t_{\text{crítico}} = -2.0739 < t_{\text{calculado}} = 0.7786 < t_{\text{crítico}} = 2.0739$, aceita-se a hipótese nula H_0 .

6. Conclusão: A um nível de significância de 5%, pode-se concluir que as duas marcas de veículos tem o mesmo desempenho.

13.10.3. Teste de hipótese da diferença entre duas médias populacionais com variâncias desconhecidas e diferentes

Quando as variâncias σ_1^2 e σ_2^2 forem diferentes e desconhecidos devemos utilizar as suas estimativas amostrais S_1^2 e S_2^2 .

A estatística de teste, neste caso, é dada por:

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Que tem distribuição t – *student* com v graus de liberdade, onde:

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$

A H_0 será rejeitada se $|t_0| > t_{\frac{\alpha}{2}, v}$

Exemplo 13.6 (Teste de hipótese para diferença entre duas médias)

Alguns inspectores, ainda mais desconfiados (**exemplo 13.4**), foram investigar as informações e descobriram que a empresa A havia omitido 4 valores de sua amostra. São eles: 10.0; 10.5; 10.6 e 11.5. Com isso, refeitos os cálculos, obteve-se:

$$\bar{x}_A = 11.75 \text{ e } S_A = 1.1894$$

Considerando, agora, variâncias diferentes e desconhecidas, testar para a igualdade de médias com um nível significância de 0.05.

Solução:

1. $\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$
2. $\alpha = 0.05$
3. $t \sim t_v$

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}} = \frac{\left[\frac{1.1894^2}{16} + \frac{0.8163^2}{12}\right]^2}{\frac{[(1.1894)^2/16]^2}{15} + \frac{[(0.8163)^2/12]^2}{11}} \approx 26$$

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{11.75 - 11.81}{\sqrt{\frac{1.1894^2}{16} + \frac{0.8163^2}{12}}} = -0.1590$$

4. $\alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025$ pois o teste é bilateral.

$$t_{\text{critico}} = t_{\frac{\alpha}{2}} = t_{0.025, 26} = 2.056$$

5. Como $-t_{\text{critico}} = -2.056 < t_{\text{calculado}} = 0.7786 < t_{\text{critico}} = 2.056$, aceita-se a hipótese nula H_0 .

6. Conclusão: A um nível de significância de 5%, pode-se concluir que as duas marcas de veículos tem o mesmo desempenho.

13.11. Teste de hipótese de duas amostras emparelhadas

Duas amostras são dependentes se membros de uma amostra podem ser usados para determinarem os membros da outra amostra. Ou seja, com dados emparelhados, há alguma relação, de modo que cada valor em uma amostra está emparelhado com um valor correspondente na outra amostra.

Agora, necessitamos analisar duas populações relacionadas, isto é, duas populações dependentes.

Neste caso, a variável de interesse será a diferença entre os pares das duas amostras, no lugar das próprias amostras, que devem ter o mesmo tamanho.

Como premissa, a população das diferenças tem distribuição aproximadamente normal e a amostra das diferenças é extraída aleatoriamente da população das diferenças.

A hipótese testada é se existe diferenças significativas entre pares de observações:

$$\begin{cases} H_0: \mu_d = 0 \\ H_1: \mu_d \neq 0 \end{cases}$$

O teste baseia-se na estatística .

$$t_0 = \frac{\bar{x}_d - \mu_d}{\frac{S_d}{\sqrt{n}}}$$

A hipótese H_0 será rejeitada se $|t_0| > \left| t_{\frac{\alpha}{2}; n-1} \right|$

Exemplo 13.7 (Teste de hipótese amostras emparelhadas)

Dois espécies de um certo tipo de cereal estão sendo testadas quanto ao seu crescimento. O experimento foi feito escolhendo 10 blocos de terreno e plantando em cada bloco mudas de ambas as espécies. Os resultados a seguir são as alturas medidas ao final do primeiro mês. Utilizar $\alpha = 0,05$.

Terreno	1	2	3	4	5	6	7	8	9	10
Espécie 1	22	27	18	33	25	21	15	33	21	24
Espécie 2	21	31	24	32	29	23	19	37	22	27

Os dados deste experimento foram colectados aos pares para impedir que as diferenças de fertilidade entre os blocos de terreno (que podem ser grandes) mascarem os resultados.

Solução:

1. $\begin{cases} H_0: \mu_d = 0 \\ H_1: \mu_d \neq 0 \end{cases}$
2. $\alpha = 0.05$
3. $t \sim t_{n-1}$

$$t_0 = \frac{\bar{x}_d - \mu_d}{\frac{S_d}{\sqrt{n}}} = \frac{-2.6}{2.32/\sqrt{10}} = -3.54$$

4. $\alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025$ pois o teste é bilateral.

$$t_{\frac{\alpha}{2}; n-1} = t_{0.025; 9} = 2.262$$

5. Como $t_{\text{crítico}} = 2.262 < |t_{\text{calculado}} = 3.54|$, rejeita-se a hipótese nula H_0 .
6. Conclusão: A um nível de significância de 5%, pode-se concluir que existe uma diferença significativa dos blocos em relação ao crescimento dos cereais.

13.12. Teste de hipótese para variância duma população normal

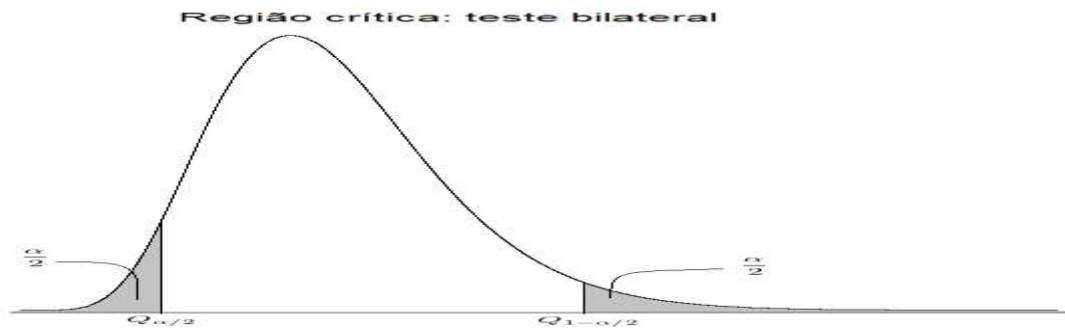
Seja $X_1, X_2, X_3, \dots, X_n$ uma amostra aleatória de tamanho n retirada de uma população normal $N(\mu, \sigma^2)$. Suponha que desejamos testar uma hipótese sobre a variância σ^2 desta população.

Sabemos que a estatística $Q = \frac{(n-1) \cdot S^2}{\sigma^2}$ tem distribuição qui-quadrado com $n - 1$ graus de liberdade.

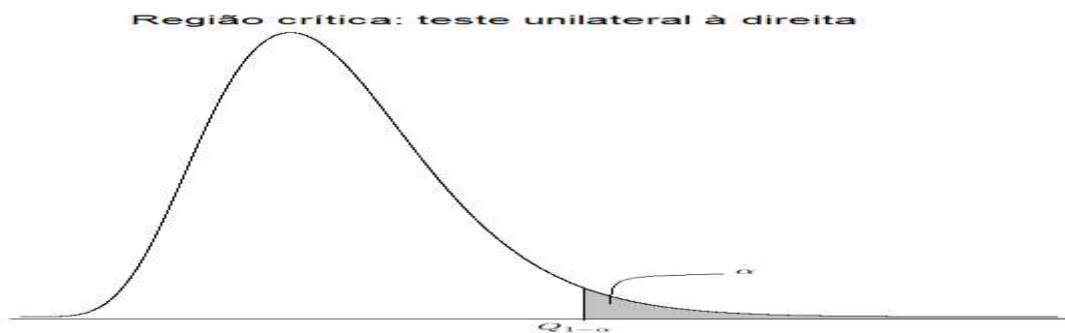
Denotando $\chi^2 \sim \chi^2_{n-1}$. Para executar este tipo de teste podemos formular os seguintes hipóteses:

$$\begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 \neq \sigma_0^2 \\ H_1: \sigma^2 > \sigma_0^2 \\ H_1: \sigma^2 < \sigma_0^2 \end{cases}$$

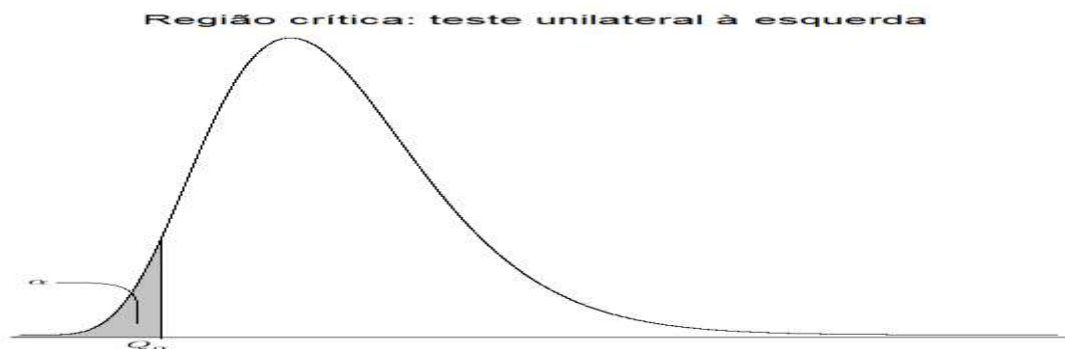
Nota: as hipóteses H_0 podem ser substituídas por $H_0: \sigma^2 \geq \sigma_0^2$ ou $H_0: \sigma^2 \leq \sigma_0^2$.



Se $\chi^2_{calculado} > \chi^2_{\frac{\alpha}{2}}$ ou $\chi^2_{calculado} < \chi^2_{1-\frac{\alpha}{2}}$ Rejeitamos a hipótese nula.



Se $\chi^2_{calculado} > \chi^2_{1-\alpha}$ rejeitamos a hipótese nula



Se $\chi^2_{calculado} < \chi^2_{\alpha}$ rejeitamos a hipótese nula

Exemplo 13.7 (Teste de hipótese para variância)

Seleccionou-se aleatoriamente uma amostra da cotação diária, em euros, de uma empresa, em relação aos dois últimos meses. Os dados obtidos, após tratamento resultaram na seguinte informação:

10,1; 10,3; 9,9; 9,8; 10,0; 10,2; 10,4; 10,6; 10,1.

A cotação diária da empresa é normalmente distribuída. Afirma-se que a variância das cotações da empresa é inferior a 0,04. Considerando um nível de significância de 5%, verifique a validade desta afirmação.

Solução:

$$1. \begin{cases} H_0: \sigma^2 \geq 0.04 \\ H_1: \sigma^2 < 0.04 \end{cases}$$

$$2. \alpha = 0.05$$

$$3. \chi^2 \sim \chi^2_{n-1}$$

$$\chi^2 = \frac{(n-1) * S^2}{\sigma^2_0} = \frac{(9-1) * 0.063}{0.04} = 12.6$$

4. $\alpha = 0.05 \rightarrow$ teste unilateral esquerdo.

$$\chi^2_{0.05,8} = 2.733$$

5. Como $\chi^2_{crítico} = 2.733 < \chi^2_{calculado} = 12.6$, não se rejeita-se a hipótese nula H_0 .

6. Conclusão: A um nível de significância de 5%, pode-se concluir que a variância das cotações da empresa é inferior a 0,04

13.13. Teste de hipóteses para variâncias de duas populações

A distribuição F (de Snedcor): Considere que temos duas amostras independentes de tamanhos n_1 e n_2 , retiradas de duas populações normais com mesma variância σ^2 . Então:

$$U = \frac{(n_1 - 1) * S^2}{\sigma^2} \sim \chi^2_{n_1-1} \text{ e } V = \frac{(n_2 - 1) * S^2}{\sigma^2} \sim \chi^2_{n_2-1}$$

A estatística definida por:

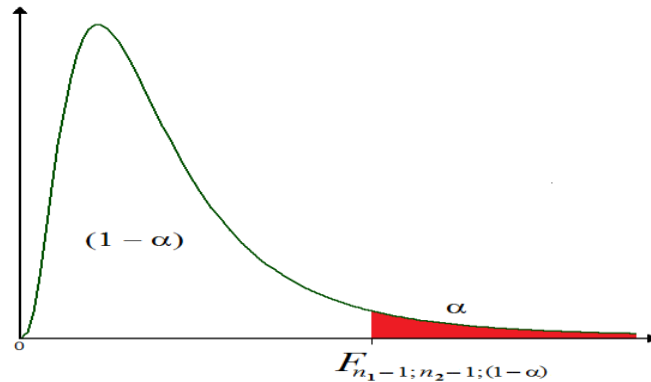
$$F = \frac{S^2_1}{S^2_2} = \frac{U/(n_1 - 1)}{V/(n_2 - 1)}$$

Tem distribuição F com $(n_1 - 1; n_2 - 1)$ graus de liberdade, ou seja:

$$F = \frac{S^2_1}{S^2_2} \sim F_{n_1-1; n_2-1}$$

- **Teste unicaudal para comparação de duas variância**

$$\text{Hipóteses: } \begin{cases} H_0: \sigma^2_1 = \sigma^2_2 \\ H_1: \sigma^2_1 > \sigma^2_2 \end{cases}$$



Como a distribuição de F é exacta, comparamos o seu valor observado,

$$F_{\text{calculado}} = \frac{S^2_1}{S^2_2}$$

directamente com o percentil da distribuição F.

Portanto:

$$\text{Se } \begin{cases} F_{\text{calculado}} > F_{\alpha; n_1-1; n_2-1} \rightarrow \text{Rejeita - se } H_0 \\ F_{\text{calculado}} \leq F_{\alpha; n_1-1; n_2-1} \rightarrow \text{N\~{a}o se Rejeita } H_0 \end{cases}$$

▪ **Teste bicaudal para comparação de duas variância**

$$\text{Hipóteses: } \begin{cases} H_0: \sigma^2_1 = \sigma^2_2 \\ H_1: \sigma^2_1 \neq \sigma^2_2 \end{cases}$$

Como a distribuição de F é exacta, então, comparamos o seu valor observado,

$$F_{\text{calculado}} = \frac{S^2_1}{S^2_2}$$

directamente com os percentis $\alpha/2$ da distribuição F com $n_1 - 1$ e $n_2 - 1$ graus de liberdade

Portanto:

$$\text{Se } \begin{cases} F_{\text{calculado}} > F_{\alpha/2; n_1-1; n_2-1} \text{ ou } F_{\text{calculado}} < F_{1-\alpha/2; n_1-1; n_2-1} \rightarrow \text{Rejeita - se } H_0 \\ F_{1-\alpha/2; n_1-1; n_2-1} \leq F_{\text{calculado}} \leq F_{\alpha/2; n_1-1; n_2-1} \rightarrow \text{N\~{a}o se Rejeita } H_0 \end{cases}$$

Exemplo 13.8 (Teste de hipótese para variâncias de duas populações)

Um docente universitário está indeciso entre dois modelos de exame. A sua escolha recairá naquele que demorar menos tempo a concluir. Para efectuar a sua escolha, o docente aplicou o teste A, a uma amostra de 8 alunos e o teste B, a uma amostra de 10 alunos, tendo obtido os seguintes tempos, em minutos, para a conclusão do exame:

- Teste A: 120; 90; 110; 100; 80; 85; 95; 80;
- Teste B: 110; 95; 100; 85; 90; 95; 110; 80; 90; 110.

Admitindo que os tempos de resolução se comportam segundo uma lei normal, teste a um nível de significância de 5%, se existe diferença na variabilidade do tempo de resolução dos dois tipos de exame.

Solução

$$1. \begin{cases} H_0: \sigma^2_1 = \sigma^2_2 \\ H_1: \sigma^2_1 \neq \sigma^2_2 \end{cases}$$

$$2. F \sim F_{n_1-1; n_2-1}$$

$$F_{\text{calculado}} = \frac{S^2_1}{S^2_2} = \frac{207.14}{116.94} = 1.77$$

$$3. \alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025 \text{ teste bicaudal.}$$

$$4. F_{\frac{\alpha}{2}; n_1-1; n_2-1} = F_{0.025; 7; 9} = 4.197$$

$$F_{1-\alpha/2; n_1-1; n_2-1} = F_{0.975; 7; 9} = \frac{1}{F_{0.025; 9; 7}} = \frac{1}{4.823} = 0.207$$

5. Como $F_{0.975; 7; 9} = 0.207 < F_{\text{calculado}} = 1.77 < F_{0.025; 7; 9} = 4.197$, não se rejeita-se a hipótese nula H_0 .

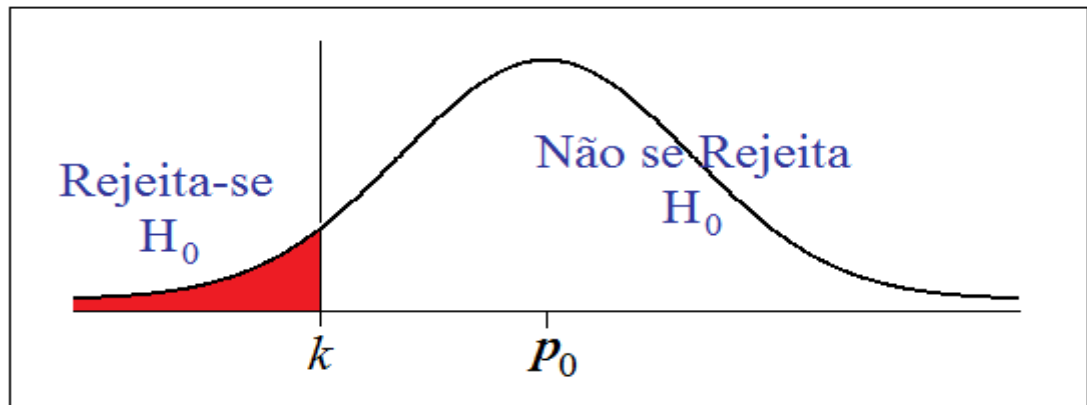
6. Conclusão: Não existe diferença na variabilidade do tempo de resolução dos dois tipos de exame, ao nível de significância de 5%.

13.14. Teste de hipóteses para proporção

O teste de hipótese para a proporção é similar ao teste para a média, uma vez que a proporção π é a média de uma variável dicotômica.

- **Teste unicaudal na cauda inferior:**

$$\text{Hipóteses: } \begin{cases} H_0: \pi = \pi_0 \\ H_1: \pi < \pi_0 \end{cases}$$



O valor observado da estatística do teste é:

$$Z_0 = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 * (1 - \pi_0)}{n}}}$$

$$\text{Se } \begin{cases} Z_0 < Z_\alpha \rightarrow \text{Rejeita-se } H_0 \\ Z_0 \geq Z_\alpha \rightarrow \text{Não se Rejeita } H_0 \end{cases}$$

- **Teste unicaudal na cauda superior:**

$$\text{Hipóteses: } \begin{cases} H_0: \pi = \pi_0 \\ H_1: \pi > \pi_0 \end{cases}$$

O valor observado da estatística do teste é:

$$Z_0 = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 * (1 - \pi_0)}{n}}}$$

$$\text{Se } \begin{cases} Z_0 > Z_\alpha \rightarrow \text{Rejeita-se } H_0 \\ Z_0 \leq Z_\alpha \rightarrow \text{Não se Rejeita } H_0 \end{cases}$$

- **Teste unicaudal na cauda inferior:**

$$\text{Hipóteses: } \begin{cases} H_0: \pi = \pi_0 \\ H_1: \pi > \pi_0 \end{cases}$$

O valor observado da estatística do teste é:

$$Z_0 = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 * (1 - \pi_0)}{n}}}$$

$$Se \begin{cases} Z_0 > Z_{\frac{\alpha}{2}} \text{ ou } Z_0 < -Z_{\frac{\alpha}{2}} \rightarrow \text{Rejeita - se } H_0 \\ -Z_{\frac{\alpha}{2}} \leq Z_0 \leq Z_{\alpha/2} \rightarrow \text{N\~{a}o se Rejeita } H_0 \end{cases}$$

Exemplo 13.9 (Teste de hipótese para variâncias de duas populações)

Uma empresa pretende lançar um novo produto numa cidade de um milhão de habitantes. No estudo de mercado realizado foram inquiridas 1000 pessoas, tendo 800 delas afirmado que muito dificilmente iriam utilizar aquele novo produto. Teste, ao nível de significância de 1%, se a verdadeira proporção de habitantes que utilizarão o novo produto pode ser considerada no máximo igual a 0,18.

Solução:

1. $\begin{cases} H_0: \pi \leq 0.18 \\ H_1: \mu > 0.18 \end{cases}$
2. $\alpha = 0.05$
3. $Z \sim N(0,1)$

$$Z_0 = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 * (1 - \pi_0)}{n}}} = \frac{0.2 - 0.18}{\sqrt{\frac{0.18 * 0.82}{1000}}} = 1.6462$$

4. $\alpha = 0.01$ pois o teste é unilateral direito. Para uma área de 0.49,
 $Z_{\text{crítico}} = Z_{\alpha} = Z_{0.01} = 2.33$
5. Como $Z_{\text{calculado}} = 1.6462 < Z_{\text{crítico}} = 2.33$, aceita-se a hipótese nula H_0 .
6. Conclusão: conclui-se que a verdadeira proporção de habitantes que utilizarão o novo produto pode ser considerada no máximo igual a 0,18, ao nível de significância de 1%.

13.15. Teste de hipóteses para duas proporções

Muitas vezes há o interesse em se comparar duas proporções π_1 e π_2 como por exemplo:

- Num determinado grupo a proporção de fumantes do sexo masculino é igual a do sexo feminino?
- A intenção de votos de um candidato é a mesma em duas capitais?
- No tratamento de uma enfermidade, a proporção de cura de um novo tratamento é a mesma que a do convencional?

Nessas circunstâncias pode-se aplicar tanto o teste unicaudal como o bicaudal. A seguir apresentaremos apenas a construção do teste bicaudal, que é o mais aplicado.

$$\text{Hipóteses: } \begin{cases} H_0: \pi_1 = \pi_2 \\ H_1: \pi_1 \neq \pi_2 \end{cases}$$

A estatística do teste é:

$$Z_{\text{calculado}} = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{p_1 * (1 - p_1)}{n_1} + \frac{p_2 * (1 - p_2)}{n_2}}}$$

Senho H_0 verdadeira, temos que $\pi_1 = \pi_2$. Logo, a estimativa da variância $Var(p_1 - p_2)$ pode ser calculada por:

$$Var(p_1 - p_2) = p * (1 - p) * \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

em que:

$$p = \frac{n_1 * p_1 + n_2 * p_2}{n_1 + n_2} = \frac{X_1 + X_2}{n_1 + n_2}$$

Com:

- $X_1 =$ número de sucessos da população 1
- $X_2 =$ número de sucessos da população 2

Desta forma, a estatística de teste é escrita por:

$$Z = \frac{p_1 - p_2}{\sqrt{p * (1 - p) * \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$\text{Se } \begin{cases} Z_0 > Z_{\frac{\alpha}{2}} \text{ ou } Z_0 < -Z_{\frac{\alpha}{2}} \rightarrow \text{Rejeita - se } H_0 \\ -Z_{\frac{\alpha}{2}} \leq Z_0 \leq Z_{\alpha/2} \rightarrow \text{Não se Rejeita } H_0 \end{cases}$$

Exemplo 13.10 (Teste de hipótese para proporções de duas populações)

Um candidato em campanha eleitoral deseja saber se a sua intenção de votos é a mesma em duas cidades importantes no cenário político (Atlântida e Flórida), e definir as estratégias de campanha. Para isso ele contratou um instituto de pesquisa que realizou pesquisas nas duas cidades. Em Atlântida foram entrevistados 500 eleitores, dos quais 116 afirmaram votar no candidato, enquanto que, em Flórida, dos 600 entrevistados, 105 foram favoráveis a ele. Qual a conclusão que se pode tirar com essas informações? (*assumir* $\alpha = 0.05$).

Solução:

1. $\begin{cases} H_0: \pi_1 = \pi_2 \\ H_1: \pi_1 \neq \pi_2 \end{cases}$
2. $\alpha = 0.05$
3. $Z \sim N(0,1)$

$$Z = \frac{p_1 - p_2}{\sqrt{p * (1 - p) * \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.232 - 0.175}{\sqrt{0.201 * 0.799 * \left(\frac{1}{500} + \frac{1}{600}\right)}} = 2.35$$

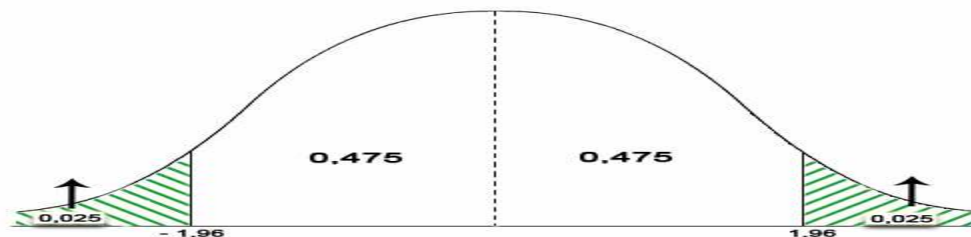
Sendo

$$p = \frac{n_1 * p_1 + n_2 * p_2}{n_1 + n_2} = \frac{500 * 0.232 + 600 * 0.175}{500 + 600} = 0.201$$

4. $\alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025$ pois o teste é bilateral. Para uma área de 0.475,

$$Z_{\text{crítico}} = Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$$

Desenhando a curva, temos:



5. Como $Z_{\text{calculado}} = 2.35 > Z_{\text{crítico}} = 1.96$, Rejeita-se a hipótese nula H_0 .
6. Conclusão: A um nível de significância de 5%, pode-se concluir que a intenção de votos nas duas cidades não é a mesma (a intenção de votos na Flórida é menor).

13.16. Testes não-paramétricos

São procedimentos matemáticos para testes de hipóteses que, diferentemente da estatística paramétrica, não fazem suposições sobre distribuição de probabilidade das variáveis a serem consideradas.

As técnicas da Estatística Não-Paramétrica são, particularmente, adaptáveis aos dados das ciências do comportamento. A aplicação dessas técnicas não exige suposições quanto à distribuição da variável populacional. Os testes não-paramétricos são extremamente interessantes para análises de dados qualitativos. Na Estatística Paramétrica, para aplicação de teste como o *t – Student*, a variável em análise precisa ser numérica. Como o próprio nome sugere, a Estatística Não-Paramétrica independe dos parâmetros populacionais e de suas respectivas estimativas.

Assim, se a variável populacional analisada não segue uma distribuição normal e/ou as amostras forem pequenas, pode-se aplicar um teste Não-Paramétrico.

- **Vantagens dos Métodos Não-Paramétricos**

- ✓ Os métodos Não-Paramétricos podem ser aplicados a uma ampla diversidade de situações, porque não exigem populações distribuídas normalmente.
- ✓ Ao contrário dos métodos Paramétricos, os métodos Não-Paramétricos podem frequentemente ser aplicados a dados não-numéricos.
- ✓ Os métodos Não-Paramétricos em geral envolvem cálculos mais simples do que seus correspondentes Paramétricos, sendo, assim, mais fáceis de entender.

- **Desvantagens dos Métodos Não-Paramétricos**

- ✓ Os métodos Não-Paramétricos tendem a perder informação, porque os dados numéricos são frequentemente reduzidos a uma forma qualitativa.
- ✓ Os testes Não-Paramétricos não são tão eficientes quanto os testes Paramétricos; assim, com um teste Não-Paramétrico, em geral necessitamos de uma amostra maior ou maiores diferenças para então rejeitarmos uma hipótese nula.

13.16.1. Teste de independência estatística

Os testes de independência baseiam-se no uso das chamadas tabelas de contingência. Uma tabela de contingência (ou tabela de frequência de dupla entrada) é uma tabela

em que as frequências correspondem a duas variáveis (uma variável categoriza as linhas e a outra categoriza as colunas).

Os testes de independência são usados para determinar se uma variável linha de uma tabela de contingência é independente da sua variável coluna, isto é, estes testes consistem em decidir entre duas alternativas do tipo:

- H_0 : As variáveis são (estatisticamente) independentes
- H_1 : As variáveis não são (estatisticamente) independentes

É de suma importância reconhecer que, neste contexto, a palavra contingência se refere a dependência estatística e não pode ser usada para estabelecer uma ligação directa de causa e efeito entre as duas variáveis em questão.

Ao testarmos a hipótese nula de independência entre as variáveis linha e coluna numa tabela de contingência, aplicam-se os seguintes pressupostos:

1. Os dados amostrais são seleccionados aleatoriamente;
2. Cada realização de cada variável pode ser classificada numa de várias categorias exaustivas e mutuamente exclusivas;
3. A hipótese nula H_0 é a afirmação de que as variáveis linha e coluna são independentes. A hipótese alternativa H_1 afirma que as variáveis linha e coluna são dependentes;
4. Para cada célula na tabela de contingência, a frequência esperada é no mínimo 5 (não há tal exigência para as frequências observadas).

Depois de realizar e classificar as observações, comparam-se as frequências observadas o_{ij} (na amostra) com as frequências esperadas e_{ij} (no caso de H_0 ser verdadeira), calculando depois a estatística de teste. A estatística de teste permite-nos medir o grau de discordância entre as frequências observadas e as frequências que deveríamos esperar teoricamente no caso de as variáveis serem independentes (no caso de H_0 ser verdadeira). Pequenos valores da estatística de teste indicam acentuada concordância entre as frequências observadas e as frequências esperadas, pelo que a decisão deve ser no sentido de aceitar H_0 . Grandes valores da estatística de teste reflectem diferenças significativas entre as frequências observadas e as esperadas pelo que a decisão deve ser no sentido de rejeitar H_0 . Para realizar o teste construímos a seguinte tabela, cujas colunas ($j = 1, 2, \dots, c$) e as linhas ($i = 1, 2, \dots, l$) se referem a

classes. Os símbolos o_{ij} representam as frequências observadas e os e_{ij} as frequências esperadas das células ij . As frequências marginais:

$$o_{.j} = \sum_{i=1}^l o_{ij} \text{ e } o_{i.} = \sum_{j=1}^c o_{ij}$$

São os totais da coluna j e da linha i , respectivamente. As frequências podem ser absolutas (número total de observações ou relativas n^{-1})

Classe j de uma variável Classe i de uma variável	Frequência observada (frequência esperada)				Total
	1	2	...	c	
1	$o_{11}(e_{11})$	$o_{12}(e_{12})$		$o_{1c}(e_{1c})$	$o_{1.} = e_{1.}$
2	$o_{21}(e_{21})$	$o_{22}(e_{22})$		$o_{2c}(e_{2c})$	$o_{2.} = e_{2.}$
...					
l	$o_{l1}(e_{l1})$	$o_{l2}(e_{l2})$		$o_{lc}(e_{lc})$	$o_{l.} = e_{l.}$
Total	$o_{.1}(e_{.1})$	$o_{.2}(e_{.2})$		$o_{.c}(e_{.c})$	n

As frequências esperadas marginais $e_{.j}$ e $e_{i.}$ são estimadas a partir das frequências observadas na amostra, isto é, faz-se $e_{.j} = o_{.j}$ e $e_{i.} = o_{i.}$. Finalmente, as frequências esperadas e_{ij} são calculadas no pressuposto de que H_0 é verdadeira, pelo que:

$$e_{ij} = \frac{o_{i.} * o_{.j}}{n}$$

Então:

- A estatística de teste é:

$$\chi^2_0 = \sum_{i=1}^l \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(l-1)*(c-1)}$$

- Regras de decisão: se $\chi^2_0 \geq \chi^2_{\text{tabelado}}$ rejeita-se H_0 ,

$$\text{sendo } \chi^2_{\text{crítico}} = \chi^2_{(l-1)*(c-1)}$$

Os testes de independência com tabelas de contingência envolvem apenas regiões críticas unilaterais à direita.

Exemplo 13.11 (Teste de independência)

Num estudo de mercado sobre a audiência de 2 jornais semanais e de 1 revista semanal foram inquiridos 1000 leitores de ambos os sexos sobre o semanário que compram preferencialmente, tendo-se encontrado os seguintes resultados:

Sexo \ Semanário	Savana	Domingo	Zambézia
Feminino	150	50	150
Masculino	350	200	100

Será de admitir que a preferência pelos vários semanários é influenciada pelo sexo dos leitores? (Admita um nível de significância de 5%)

Solução:

Pretende-se saber se a preferência pelos vários semanários é independente do sexo dos leitores. Para isso vai aplicar-se o teste do qui-quadrado da independência entre duas variáveis. Então:

- H_0 : A preferência pelos vários semanários não depende do sexo do eleitor
- H_1 : A preferência pelos vários semanários depende do sexo do eleitor

Cálculos auxiliares: Os o_{ij} correspondem ao número de indivíduos observados das células ij . O cálculo do número esperado de indivíduos esperados e_{ij} , pressupõe que a hipótese nula, H_0 é verdadeira, isto é, as variáveis são independentes:

$$e_{ij} = n * p_{ij} = n * p_{i.} * p_{.j} = n * \frac{o_{i.}}{n} * \frac{o_{.j}}{n} = \frac{o_{i.}}{n} * o_{.j} * n$$

Neste caso ter-se-á:

- $o_{1.} = 150 + 50 + 150 = 350$
- $o_{2.} = 350 + 200 + 100 = 650$
- $o_{.1} = 150 + 350 = 500$
- $o_{.2} = 50 + 200 = 250$
- $o_{.3} = 150 + 100 = 250$
- $e_{11} = \frac{350*500}{1000} = 175$
- $e_{12} = \frac{350*250}{1000} = 87.5$
- $e_{13} = \frac{350*250}{1000} = 87.5$
- $e_{21} = \frac{650*500}{1000} = 325$
- $e_{22} = \frac{650*250}{1000} = 162.5$
- $e_{23} = \frac{650*250}{1000} = 162.5$

Na tabela seguinte estão incluídos para além dos o_{ij} os e_{ij} obtidos:

Sexo \ Semanário	Savana	Domingo	Zambézia	Total
Feminino	150 (175)	50 (87.5)	150 (87.5)	350
Masculino	350 (325)	200 (162.5)	100 (162.5)	650
Total	500	250	250	1000

$$\chi^2_0 = \sum_{i=1}^l \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$= \frac{(150 - 175)^2}{175} + \frac{(50 - 87.5)^2}{87.5} + \frac{(150 - 87.5)^2}{87.5} + \frac{(350 - 325)^2}{325}$$

$$+ \frac{(200 - 162.5)^2}{62.5} + \frac{(100 - 162.5)^2}{162.5} = 98.89$$

Valor crítico:

$$\chi^2_{(l-1)*(c-1); 1-\alpha} = \chi^2_{(2-1)*(3-1); 1-0.05} = \chi^2_{2; 0.95} = 5.991$$

Decisão: Como o valor da estatística de teste $\chi^2_0 = 98.89 > \chi^2_{2; 0.95} = 5.991$, rejeita-se a hipótese nula, isto é, devemos concluir que a preferência pelos semanários depende sexo do eleitor.

13.17. EXERCÍCIOS PROPOSTOS

13.17.1. Um fornecedor apresenta uma caixa, e afirma que o peso médio desta caixa é de 368 gramas. De experiências anteriores sabe-se que o desvio padrão da população vale 15g e que os valores se comportam segundo a distribuição Normal. Para verificar se a afirmação é verdadeira, verifica-se uma amostra de 25 caixas, pesa-se e calcula-se o peso médio da amostra, achando 372,5g. Qual a conclusão a respeito da afirmação do fornecedor, ao nível de significância 0,01?

13.17.2. Uma agência de empregos alega que os candidatos à directoria por ela colocados nos últimos seis meses têm salários de 9000MT, em média. Uma agência governamental extraiu uma amostra aleatória daquele grupo, encontrando salários médios de 8000MT, com desvio padrão de 1000MT, com base em 50 empregados. Teste a afirmação da agência, contra a alternativa, de que o salário médio é inferior a 9000MT, ao nível de significância 0,05.

13.17.3. O gerente de marketing de uma fábrica de automóveis está interessado em determinar a proporção de novos proprietários de carros compactos que teriam adquirido um air-bag inflável para o lado do passageiro se o mesmo estivesse disponível a um custo adicional de 300,00 u.m. Por informações anteriores, o gerente acredita que a proporção é 30%. Suponha que é feito um levantamento com 200 novos proprietários de carros compactos e 79 indiquem que teriam comprado os air-bags infláveis. No nível de significância de 0,05, há evidências de que a proporção da população é diferente de 0,3?

13.17.4. Suponha que o director de produção de uma fábrica de tecidos precise determinar se uma nova máquina está produzindo um tipo de tecido de acordo com as especificações do fabricante. As especificações indicam que o tecido devia ter uma resistência de rompimento superior a 70 libras (1 libra = 433,59 gramas) e um desvio padrão de 3,5 libras. Uma amostra de 36 peças revela uma média aritmética da amostra igual a 69,7 libras. Há evidências de que a máquina não está atendendo às especificações, em termos da média da resistência de rompimento? (utilize um nível de significância de 0,05).

13.17.5. Uma rede de postos de gasolina afirma que, em seus estabelecimentos não se vende gasolina adulterada. Sabe-se que, de acordo com os padrões de qualidade, a gasolina não pode conter mais de 240 ml de álcool por litro. O órgão de fiscalização colheu 25 medições do produto nos postos dessa rede, obtendo a partir delas uma média de 240,75 ml de álcool/litro. Admitindo-se que a quantidade de álcool presente na gasolina tem uma distribuição normal com desvio-padrão de 2,5 ml/litro. Ao nível de significância 5%, pode-se afirmar que a gasolina é adulterada?

13.17.6. Um psicólogo de indústrias deseja estudar os efeitos da motivação nas vendas, em determinada empresa. Foi seleccionada uma amostra aleatória de 24 indivíduos, 12 de cada grupo. Os dados a seguir representam o volume de vendas (em milhares de reais) alcançado durante o primeiro mês de emprego. Há evidências de que o volume médio de vendas seja diferente entre os grupos? (utilize nível de significância 0,05)

Por hora	256	245	212	219	222	236	317	329	234	231	236	227
Comissão	224	261	254	228	273	234	285	225	237	232	277	245

13.17.7. No caso judicial EUA versus Cidade de Chicago, foram postas em dúvida as práticas honestas de emprego. Um grupo minoritário (A) e um grupo maioritário (B) fizeram o exame para capitão do corpo de bombeiros, com os seguintes resultados:

	Aprovados	Reprovados
Grupo A	10	14
Grupo B	417	145

Com os resultados acima, e com nível de significância de 5%, teste a afirmação de que o sucesso no teste é independente do grupo.

13.17.8. Solicitou-se a quatro amostras de 30 funcionários de uma grande empresa que opinassem sobre a nova direção da empresa. Ao nível de significância 0,01, o que se pode concluir?

	Estagiários	Formadores	Técnicos	Gerentes
Aprovam	5	4	20	27
Desaprovam	25	26	10	3

13.17.9. Um estudo de usuários e não usuários do cinto de segurança resultou nos dados amostrais aleatórios resumidos na tabela a seguir. Teste a afirmação de que a quantidade de fumo é independente do uso do cinto de segurança. Uma teoria plausível é que as pessoas que fumam mais estão menos preocupadas com a sua saúde e segurança, sendo assim, menos propensas a usar cintos. Com nível de significância 0,01, os dados amostrais apoiam esta teoria?

	Número de cigarros por dia			
	0	1-14	15-34	35 ou mais
Aprovam	175	20	42	6
Desaprovam	149	17	41	9

13.17.10. Para elaborar um estudo sobre o aproveitamento na disciplina de Estatística em dois cursos, analisaram-se as notas obtidas pelos alunos em cada um deles:

Curso	Nº de alunos	\bar{x}	s^2
Economia	31	13	10.3
Contabilidade	61	10.8	5.7

Utilizando um nível de significância de 2%, averigúe se as médias nos dois cursos diferem de forma expressiva.

- 13.17.11.** Numa linha de engarrafamento de azeite a quantidade deitada em cada garrafa é uma variável aleatória que se admite ter distribuição normal. O processo de enchimento considera-se regulado se $\mu = 1$ litro, não sendo de admitir grandes desvios. Para controlar o processo de enchimento escolheram-se ao acaso 20 garrafas da produção diária. Suponha que se obteve uma média de 0.965 litros com um desvio padrão de 0.08 litros. Pode-se afirmar que o processo não está regulado?
- 13.17.12.** As especificações de uma dada droga veterinária exigem 23,2g de álcool etílico. Uma amostra de 10 análises do produto apresentou um teor médio de álcool de 23,5g com desvio padrão de 0,24g. Pode-se concluir ao nível de significância de 1% que o produto satisfaz as condições exigidas?
- 13.17.13.** Um laboratório de vacinas contra febre aftosa reivindicou que ela imuniza 90% dos animais. Em uma amostra de 200 animais, nos quais foram aplicadas a vacina, 160 foram imunizados. Verificar se a declaração do fabricante é verdadeira ao nível de 5%.
- 13.17.14.** A associação dos proprietários de indústrias metalúrgicas está muito preocupada com o tempo perdido com acidentes de trabalho, cuja média, nos últimos tempos, tem sido da ordem de 60 horas/homem por ano e desvio padrão de 20 horas/homem. Tentou-se um programa de prevenção de acidentes, após o qual foi tomada uma amostra de nove indústrias e medido o número de horas/homens perdidos por acidentes, que foi de 50 horas. Você diria, no nível de 5%, que há evidência de melhoria?
- 13.17.15.** População de eleitores portugueses. Sondagem (aleatória) a 1200 eleitores revelou que 683 tencionam votar no partido ABC. Entretanto o presidente do partido tinha afirmado "estou convencido que vamos obter mais de 50% dos votos". Concordamos com esta afirmação?
- 13.17.16.** Uma amostra aleatória de 1785 adultos foi inquirida relativamente à seguinte questão: Assistiu a algum serviço religioso nos últimos 7 dias? 980 pessoas responderam "Sim". Ao nível de significância de 0.01, podemos

considerar que os dados fornecem evidência suficiente de que mais de metade da população adulta assistiu a um serviço religioso?

13.17.17. O ministro da saúde afirmou, numa conferência de imprensa, que o tempo médio de espera para uma consulta de oftamologia no SNS é de 25 dias. Os jornalistas desconfiam que o tempo de espera é maior. Numa amostra de 36 utentes encontrou-se um tempo médio de espera de 28 dias, com desvio padrão igual a 5.

13.17.18. Em Junho de 1975, foram obtidas 12 amostras de água de diferentes zonas de um lago de um dado distrito. Com o objectivo de estudar o nível médio de concentração de cloro foram efectuadas análises químicas a essas amostras. Dois anos depois, foram analisadas 10 amostras, tendo-se obtido os seguintes valores:

	1975	1977
Desvio padrão	1.2	1.3
Média	18.3	17.8

Os dados evidenciam a diminuição do nível médio de concentração de cloro nas águas do lago em 1977, em relação ao nível médio obtido em 1975? (Use $\alpha = 0.01$)

13.17.19. A *DeBug Company* vende um repelente de insectos que alega ser eficiente pelo prazo de 400 horas no mínimo. Uma análise de nove itens escolhidos aleatoriamente acusou uma média de eficiência de 380 horas. Teste a afirmação da empresa, contra a alternativa que a duração é inferior a 400 horas, ao nível de significância de 1%, se o desvio-padrão amostral é de 60 horas (considere distribuição normal).

13.17.20. Uma companhia de seguros iniciará uma campanha extensa de propaganda para vender apólices de seguro de vida, se verificar que a quantia média segurada por família é inferior a 10.000,00 Mt. Tomou-se uma amostra aleatória de 50 famílias que acusou um seguro médio de 9.600,00 Mt com desvio padrão de 1.000,00 Mt. Com base na evidência amostral, a campanha deve ser iniciada ou não (nível de 1% de significância)?

- 13.17.21.** Um ensaio de tensões de ruptura de 6 cabos produzidos por uma companhia mostrou a tensão média de ruptura de 7.750kg e o desvio padrão de 145kg, ao passo que o fabricante declara que aquela tensão média é de 8.000kg. Será verdadeira a declaração do fabricante, ao nível de significância de 5%?
- 13.17.22.** Suponhamos que em indivíduos normais quanto à visão, a pressão intra-ocular seja uma variável aleatória normalmente distribuída com média 20 e variância 4 (em unidade de mm de mercúrio). Um cientista, querendo por à prova a sua hipótese de que o glaucoma causa um aumento tencional, mediu as pressões de 16 pacientes portadores de glaucoma, obtendo uma média igual a 24. O cientista deve ou não manter sua hipótese, ao nível de significância de 5%?
- 13.17.23.** O presidente do Clube A, afirma que 58% da população de sua cidade torce para seu time. O presidente do clube rival com o intuito de desmentir a afirmação, contrata uma pesquisa que entrevistou 200 pessoas na qual 107 afirmaram realmente torcer para o clube A. Formule a hipótese e realize o teste ao nível de significância de 10%.
- 13.17.24.** A experiência tem comprovado que mais de 40% dos estudantes são reprovados em uma prova de estatística. Se 45 de 90 estudantes amostrados fossem reprovados, o que se pode concluir a respeito desta afirmação. Teste esta hipótese ao nível de significância de 4%.
- 13.17.25.** Estudos anteriores levam a supor que, quando uma criança de 2 meses começa a tomar exclusivamente leite do tipo A, o seu peso sofre um aumento que no primeiro mês é suposto seguir uma distribuição Normal de variância 9000 gramas². Foram escolhidas, ao acaso 20 crianças de 2 meses a quem se deu o leite referido e o seu peso aumentou em média 475 gramas. Averigúe, ao nível de significância 5% se o aumento de peso foi significativamente inferior a 500 gramas.
- 13.17.26.** O consumo mensal de calorias (kcal/g) dum certa espécie de esquilos é bem modelado por uma distribuição Normal com desvio padrão 0.16 (parâmetro populacional). Recolheu-se uma amostra aleatória de dimensão 18

cuja média amostral foi de 0.41. Teste, ao nível de significância 5%, se o consumo médio é significativamente superior a 0.4.

13.17.27. Mediu-se o comprimento (em mm) da cauda de 25 ratos do campo escolhidos aleatoriamente. A média da amostra foi 56.22 e o respectivo desvio padrão 1.33. Admita que a população dos comprimentos das caudas é bem modelada por uma distribuição Normal.

- a) Teste, ao nível de significância 5%, se o comprimento médio das caudas dos ratos é significativamente superior a 55mm.
- b) Teste, ao nível de significância 5%, se a variância do comprimento das caudas é significativamente diferente de 1mm.

13.17.28. Uma amostra de 29 alturas de plantas de membros de uma determinada espécie forneceu uma variância amostral de 14.62 cm² e uma amostra de 25 alturas de plantas de membros de uma outra espécie forneceu uma variância amostral de 8.45 cm². Teste, ao nível de significância 5%, se a variância da segunda amostra é significativamente inferior à da primeira.

13.17.29. Em um estudo publicado no “*Canadian Medical Association Journal*” em novembro de 1972, procurou-se investigar o efeito do uso da vitamina C na prevenção de resfriados. Para isso, realizou-se o seguinte experimento: por um determinado período de tempo, 407 indivíduos tomaram fortes doses de vitamina C e 411 receberam placebo. No grupo da vitamina, 105 participantes ficaram livres de doenças do trato respiratório, enquanto, no grupo placebo, esse número foi de apenas 76 participantes. O que os pesquisadores puderam concluir? (Use $\alpha = 0.05$)

13.17.30. O tempo médio por operário, para executar uma tarefa, tem sido de 100 min com desvio padrão de 15 min. Foi introduzida uma modificação para reduzir este tempo e após alguns meses foi seleccionada uma amostra de 16 operários medindo se o tempo de execução de cada um. Obteve se um tempo médio amostral de 90 min e um desvio padrão amostral de 16 min.

- a) Verifique se existem evidências, ao nível de significância 0.05, de que a modificação sortiu efeito;
- b) Verifique se há evidências ao nível de significância 0.05, de que a modificação alterou a variância populacional.

13.17.31. Uma empresa que presta serviços de acessória a outras empresas esta interessada em comparar a taxa de reclamações sobre os seus serviços em dois dos seus escritórios em duas cidades diferentes. Suponha que a empresa tenha seleccionado aleatoriamente 100 serviços realizados pelo escritório da cidade A e foi constatado que em 12 deles houve algum tipo de reclamação. Já do escritório da cidade B foram seleccionados 120 serviços e 18 receberam algum tipo de reclamação. A empresa deseja saber se estes resultados são suficientes para se concluir que os dois escritórios apresentam diferenças significativas entre suas taxas de reclamações. (Use $\alpha = 0.05$)

13.17.32. Uma máquina está calibrada pra encher garrafas de refrigerantes com uma média de 25 cl. Uma amostra aleatória de 16 garrafas revelou os seguintes valores (em cl):

25.3 25.9 24.3 26.1 27.0 23.8 24.6 23.7
26.3 25.2 24.8 23.7 25.5 26.2 26.2 25.4

Estes dados apresentam evidência de a máquina não estar devidamente calibrada? (Use $\alpha = 0.05$)

13.17.33. Pretende-se introduzir um novo processo na produção de um certo tipo de esferas para uso industrial. O novo processo mantém a pressão média mas espera-se que consiga reduzir a sua variabilidade, que até agora era de 14,5. Como a introdução completa do novo processo acarreta custos, resolveu-se proceder a um teste, para o qual foram obtidas 16 esferas produzidas de acordo com o novo método.

Ao nível de significância de 0,05 e sabendo que a variância da amostra é 6.8, qual a decisão a tomar?

13.17.34. Um investigador pretende averiguar se uma determinada pílula tem como efeito secundário o aumento da pressão arterial. Inicialmente, são efectuadas medições da pressão arterial a 14 mulheres. Após 6 meses de utilização regular da pílula, são efectuadas novas medições da pressão arterial. As medições obtidas foram:

Antes	70	80	72	76	76	76	76	72	82	64	74	92	74	68
Depois	68	72	62	70	58	66	68	52	64	72	74	60	74	72

Levam-nos os dados a crer, com uma significância de 5%, de que o uso desta pílula reduz a pressão arterial?

- 13.17.35.** Uma população tem desvio padrão conhecido, sendo igual a 5mm. Se uma amostra de 50 elementos, obtida dessa população, tem média igual a 46mm, podemos afirmar que a média dessa população é superior a 43mm, ao nível de significância de 1%?
- 13.17.36.** Um fabricante afirma que a tensão média de ruptura dos cabos produzidos por sua companhia não é inferior a 500kgf. Uma amostra de 7 cabos foi ensaiada, obtendo-se os resultados (em kgf):
490, 495, 480, 493, 475, 478 e 485
Testar a afirmação do fabricante, utilizando o nível de significância de 5%.
- 13.17.37.** Uma marca B de bebidas divulga ser detentora de pelo menos 35,5% das vendas naquela região. O concorrente regional C resolveu contestar e, para tanto, pesquisou em 20 pontos de vendas, chegando no percentual de 35,8% de vendas sendo da marca B. Com esta pesquisa, podemos concluir, à significância de 10%, que a marca B tinha razão?
- 13.17.38.** Uma fábrica anuncia que o índice de nicotina dos cigarros da marca X apresenta-se abaixo de 26mg por cigarro. Um laboratório realiza 10 análises do índice obtendo 26, 24, 23, 22, 28, 25, 27, 26, 28 e 24. Sabe-se que o índice de nicotina dos cigarros da marca X se distribui normalmente com variância de 5,36mg². Pode-se aceitar a afirmação do fabricante ao nível de 5%?
- 13.17.39.** Uma peça ao ser fabricada, foi planejada de tal maneira que uma de suas dimensões é 10,0cm. A variância do processo produtivo é de 0,0095cm². Se uma amostra de 40 peças fornece essa dimensão média igual a 10,05cm, devemos rejeitar a hipótese nula de que a média $\mu = 10,0\text{cm}$, em favor da alternativa de que a média μ seja diferente de 10,0 cm? Usar o nível de significância de 5%.
- 13.17.40.** Uma fábrica produz certo tipo de reguladores de pressão. Estes reguladores são produzidos para suportar uma pressão de 20atm. Um ensaio é realizado com uma amostra de 7 reguladores de pressão e verificou-se que as pressões suportadas são (em atm):

19,5, 18,9, 19,0, 19,1, 18,9, 19,3 e 19,0

Com base no ensaio realizado, podemos concluir que a pressão suportada é na realidade menor que 20atm? Usar o nível de significância de 1%.

13.17.41. Uma marca B de bebidas alega ser detentora de mais que 35,5% das vendas naquela região. O concorrente regional C resolveu contestar e, para tanto, pesquisou em 20 pontos de vendas, chegando no percentual de 35,8% de vendas sendo da marca B. Com esta pesquisa, podemos concluir, à significância de 10%, que a marca B tinha razão?

13.17.42. Uma fábrica de lajotas de cerâmica introduz um novo material em sua fabricação e acredita que aumentará a resistência média, que é de 206kg. A resistência das lajotas tem distribuição normal com desvio padrão de 12kg. Retira-se uma amostra de 30 lajotas, obtendo média amostral igual a 210kg. Ao nível de 10%, pode o fabricante aceitar que a resistência média de suas lajotas tenha aumentado?

13.17.43. Uma máquina automática que empacota o alimento A é programada para colocar 100g de peso. Para verificar a precisão da máquina, uma amostra de 60 pacotes do citado alimento fornece peso médio de 98g e desvio padrão de 6g. O que se pode concluir ao nível de 1%?

13.17.44. Um auditor deseja testar a hipótese de que o valor médio de todas as contas a receber de uma dada firma é, no mínimo, 260,00 u.m, tomando para tanto uma amostra de $n = 36$ e supondo que a média amostral é igual a 240,00 u.m. Testar a hipótese ao nível de significância de 5%, dado que se conhece o desvio padrão dos valores das contas a receber, isto é, $\sigma = 43,00 u. m.$

13.17.45. Um experimento foi realizado para comparar o desgaste abrasivo de dois diferentes materiais laminados. A variância da medida do desgaste (codificado) é conhecida como sendo 16 para o material 1 e 25 para o material 2. No experimento, 20 peças do material 1 foram testadas, expondo cada peça a uma máquina e medindo o desgaste e 30 peças do material 2 foram testadas da mesma forma. Em cada caso, a profundidade do desgaste foi observada. A amostra do material 1 deu uma média (codificada) de 85 unidades, enquanto que a amostra do material 2 deu uma média de 81. Teste, ao nível de

significância de 0,10, a hipótese de que os dois tipos de materiais apresentam a mesma média de desgaste abrasivo.

13.17.46. Um caçador de faisão afirma que ele acerta 80% dos pássaros em que ele atira. Você concorda com esta afirmação se num dia qualquer ele acerta 9 dos 15 pássaros em que ele atira? Use 0,05 como nível de significância.

13.17.47. Uma votação será feita entre os residentes de uma cidade e a região rural ao redor desta cidade para determinar se um projecto químico deverá ser construído. A construção é dentro dos limites da cidade e por esta razão muitos eleitores do campo sentem que o projecto passará por causa da grande proporção dos eleitores da cidade, os quais são favoráveis. Para determinar se existe diferença significativa na proporção de eleitores da cidade e do campo a favor do projecto, uma amostragem foi feita. Se 120 de 200 eleitores da cidade são a favor do projecto e 240 de 500 eleitores do campo são a favor, você concordaria que a proporção de eleitores da cidade favoráveis ao projecto é maior do que a proporção de eleitores do campo favoráveis ao projecto? Use $\alpha = 0,025$.

13.17.48. O consumidor de um certo produto acusou o fabricante, dizendo que mais de 20% das unidades fabricadas apresentam defeito. Para confirmar sua acusação, ele usou uma amostra de tamanho 50, onde 27% das peças eram defeituosas. Mostre como o fabricante poderia refutar a acusação. Utilize um nível de significância de 10%.

13.17.49. Suponha que um medicamento existente no mercado produza o efeito desejado em 60% dos casos nos quais o mesmo é aplicado. Um laboratório produz um novo medicamento e afirma que ele é melhor do que o existente. Aplicou-se o medicamento em 10 pacientes. Verificar estatisticamente se a afirmação do laboratório é verdadeira. Use um nível de significância de 5%.

13.17.50. A produção do remédio M em uma máquina no laboratório Y tem uma média de 72/h e um desvio-padrão de 2/h com distribuição normal. Recentemente a máquina foi ajustada. A fim de determinar o efeito do ajuste, 10 amostras foram testadas. Os testes apresentaram uma produção média de 75/h. Considere que o desvio-padrão não mudou. Com base nesses dados, com

um nível de significância de 5% é possível afirmar que o valor médio não mudou?

13.17.51. Um farmacêutico, ao receber de um fornecedor um grande lote de remédios, decidiu inspeccionar 200 delas. Decidiu, também, que o lote será rejeitado se ficar convencido, ao nível de 5% de significância. Qual será sua decisão (aceitar ou rejeitar o lote) se na amostra foram encontradas nove remédios com problemas. Sabendo que a média de problemas na população é 4 a cada 100 e desvio padrão de 1.

13.17.52. Um fabricante de insecticida descobriu, em uma pesquisa, que muitos consumidores da categoria achavam que o cheiro do seu produto era suave demais, o enfraquecia o seu apelo publicitário de produto fortíssimo contra os insectos. Após alguns *sniff tests*, foi seleccionado um novo aroma para o produto. Faltava testar o novo aroma no produto em uso pela consumidora. O fabricante solicitou à empresa de pesquisa que fizesse um teste de produto *in home* em que uma amostra de 15 consumidores usaria durante uma semana o produto com o cheiro actual, enquanto uma outra amostra, também de 12 consumidores com o mesmo perfil, usaria o produto com o novo cheiro. Definiu como padrão de acção: mudar o aroma de sua marca se, e somente se, o novo produto for superior ao actual. Suponha que o fabricante tenha escolhido o nível de significância 0,05. A avaliação em escala de 7 pontos resultou nos seguintes parâmetros.

	Aroma actual	Aroma novo
Média	5.94	6.27
Desvio Padrão	5.22	2.39

Qual é a decisão do fabricante?

13.17.53. Em um hospital, o tempo médio, por médico, para executar uma cirurgia, tem sido 100 minutos, com um desvio padrão de 15 minutos. Introduziu-se uma modificação para diminuir esse tempo, e, após certo período, sorteou-se uma amostra de 16 médicos, medindo-se o tempo de cirurgia de cada um. O tempo médio da amostra foi de 85 minutos, e o desvio padrão foi 15 minutos. Estes resultados trazem evidências estatísticas da melhora desejada (nível de significância 0,05)?

13.17.54. Estudos anteriores levam a supor que, quando uma criança de 2 meses começa a tomar exclusivamente leite do tipo A, o seu peso sofre um aumento que no primeiro mês é suposto seguir uma distribuição Normal de variância 9000 gramas². Foram escolhidas, ao acaso 20 crianças de 2 meses a quem se deu o leite referido e o seu peso aumentou em média 475 gramas. Averigúe, ao nível de significância 5% se o aumento de peso foi significativamente inferior a 500 gramas.

13.17.55. Uma dose padrão de um parasita foi transmitida para 20 ratos. Depois de 4 semanas os ratos foram mortos e o número de parasitas adultos foi contado. O número total de parasitas encontrados nos 20 ratos foi 36 dos quais 27 foram machos. Sabe-se que em coelhos a proporção é de 70%. Verifique a um nível de 5% se o mesmo ocorre com os ratos.

13.17.56. Uma nova espécie de trigo desenvolvida em laboratórios será testada quanto a sua produtividade, em comparação com a espécie tradicional. Dados do governo revelam que a produtividade média de lavouras que se utilizam da espécie tradicional é de 25 ton/ha. A produtividade de uma fazenda é uma variável aleatória normalmente distribuída. Dezasseis fazendas foram preparadas para a avaliação da nova espécie. Qual seria o seu parecer sobre a nova espécie se, em seu experimento você observasse na amostra média de 28 ton/ha e variância de 12 (ton/ha)².

13.17.57. As condições de mortalidade de uma região são tais que a proporção de nascidos que sobrevivem até 60 anos é de 0,60. Testar esta hipótese ao nível de 5% de significância se em 1000 nascimentos amostrados aleatoriamente, verificou-se 530 sobreviventes até os 60 anos.

13.17.58. Dez cobaias foram submetidas ao tratamento de engorda com certa ração. Os pesos em gramas, antes e após o teste são dados a seguir (supõe-se que provenham de distribuições normais). A 1% de significância, podemos concluir que o uso da ração contribuiu para o aumento do peso médio dos animais?

Cobaia	1	2	3	4	5	6	7	8	9	10
Antes	635	704	662	560	603	745	698	575	633	669
Depois	640	712	681	558	610	740	707	585	635	682

13.17.59. Uma máquina para colocar vacinas nos frascos enche-os com volume X , $X \sim N(500\text{ml}, \sigma^2)$. Cada dia toma-se uma amostra de 16 frascos para avaliar se a máquina esta regulada. Em um dia obteve-se a amostra:

509,42 476,79 507,27 458,10 473,30 501,41

515,07 518,07 463,76 493,95 521,62 503,41

480,71 495,32 481,55 472,32

- Fazer um teste de hipóteses para testar se a média difere ou não de 500ml ($\alpha = 0,01$).
- Fazer um teste de hipóteses para testar se a variância difere ou não de 400ml² ($\alpha = 0,10$).

13.17.60. FARBER (2009, p.296-297) o limite USDA para contaminação por salmonela por frango é de 20%. Um inspector de carnes reporta que o frango produzido por uma empresa excede o limite USDA. Realizando o teste de hipóteses para determinar se afirmação do inspector de carne é verdadeira. Quando ira ocorrer um erro do tipo I ou tipo II? Qual é mais sério?

13.17.61. O anúncio de uma certa dieta de emagrecimento dizia: “Perca 18 quilos em 4 meses!”. Recolheu-se uma amostra de 30 seguidores desta dieta tendo-se obtido uma média de 14.4 e um desvio padrão de 4.4 (em quilos perdidos). Teste, a um nível de significância de 5%, o anúncio.

13.17.62. Uma clínica pretende comparar dois tipos de dietas. Com esse objectivo, escolheu aleatoriamente e independentemente, uma amostra de 100 pacientes com excesso de peso e durante 10 semanas metade desses pacientes foram sujeitos à dieta 1 e os restantes à dieta 2. Após as 10 semanas, anotou-se o total de peso perdido (em kg) por cada paciente e obteve-se

$$\bar{x}_1 = 9.3, \quad s_1 = 2.4 \text{ para dieta 1}$$

$$\bar{x}_2 = 8.2, \quad s_2 = 2.6 \text{ para dieta 2}$$

Deve-se admitir que as duas dietas têm, em média, o mesmo efeito na perda de peso? Justifique a resposta considerando $\alpha = 0.05$.

13.17.63. Um estudo nutricional detectou numa amostra de 55 hipertensos, 24 com dietas pobres em sódio. Paralelamente, numa amostra de 149 não hipertensos detectaram-se 36 com dietas pobres em sódio. Poderá concluir se, para um nível

de significância de 0.05, que a proporção de indivíduos sujeitos a dietas pobres em sódio é maior entre hipertensos?

13.17.64. As variâncias modificadas de duas amostras aleatórias independentes extraídas de duas populações normais foram $s^2_1 = 6.32$ e $s^2_2 = 2.34$, sendo as amostras de dimensões $n_1 = 11$ e $n_2 = 16$ respectivamente. Comente à luz destes dados a hipótese das duas populações terem a mesma variância ($\alpha = 0.05$).

13.17.65. Suponha que as produções (em gramas) de comprimidos, em intervalos de tempo fixos, aleatoriamente seleccionados de duas máquinas M1 e M2 de um laboratório se podem considerar normais (com variâncias iguais). Os pesos obtidos em duas amostras permitiram determinar as quantidades seguintes:

M1	M2
$\sum_{i=1}^8 x_i = 80.8$	$\sum_{i=1}^9 y_i = 96.3$
$\sum_{i=1}^8 x^2_i = 816.664$	$\sum_{i=1}^9 (y_i - \bar{y})^2 = 0.549$

- Verifique se é plausível considerar que a variabilidade, em gramas, da produção das duas máquinas é idêntica ($\alpha = 0.01$).
- O técnico responsável pelas duas máquinas garante que a máquina 2 produz, em média, mais quantidade que a máquina 1. Ao nível de 5%, esta afirmação é consistente com os dados?

13.17.66. Uma pesquisa conclui que 90% dos médicos recomendam aspirina a pacientes que têm filhos. Teste a afirmação, ao nível de significância de 0,05, contra a alternativa de que a percentagem é inferior a 90%, se numa amostra aleatória de 100 médicos, 80% recomendam aspirina.

13.17.67. O Ministério da Saúde afirma que, com os meios agora postos à disposição dos hospitais civis, o tempo médio de internamento é no máximo 15 dias. Estas declarações foram postas em causa por alguns gestores hospitalares. Estes decidiram proceder à recolha de uma amostra aleatória de 225 doentes, onde se observou que o tempo médio de internamento foi de 18 dias. Com base nestes dados, e supondo que a variável em estudo segue uma distribuição

normal com desvio padrão de 15 dias. Terão os gestores hospitalares razão? Justifique a sua resposta, utilizando um teste adequado a 1% de significância.

13.17.68. A fim de acelerar o tempo que um analgésico leva para penetrar na corrente sanguínea, um químico analista acrescentou certo componente à fórmula original, que acusava um tempo de médio de 43 minutos. Em 36 observações com a nova fórmula, obteve-se um tempo médio de 42 minutos, com desvio padrão de 6 minutos.

- a) O que podemos concluir, ao nível de 5% de significância, sobre a eficiência do novo componente?
- b) Qual seria a resposta ao nível de 1%?
- c) Que tipo de erro pode ser cometido?

13.17.69. Uma amostra aleatória de 100 mortes naturais, numa certa região, forneceu uma média de 78 anos, com desvio padrão de 8,9 anos. Ao nível de 5% de significância, isto indica que o tempo médio de vida nessa região, actualmente, é maior que 70 anos?

13.17.70. Entre um número considerável de casos de pneumonia não tratados com sulfas, a percentagem que desenvolveu complicações foi de 16%. Com o intuito de saber se o emprego de sulfas diminuiria essa percentagem, 250 casos de pneumonia foram tratados com sulfapiridina e destes 26 apresentaram complicações. Admitindo que os pacientes são semelhantes em tudo, excepto quanto ao tratamento, teste a hipótese de que a proporção de casos com complicações entre os pacientes tratados com sulfas é significativamente menor do que os não tratados (considerar $\alpha = 0,05$).

13.17.71. Um médico receitou um medicamento vasodilatador (Nifedipina) para Hipertensão Arterial, mas ele suspeita que o medicamento está aumentando a frequência cardíaca dos pacientes. Sabedor que a população apresenta os seguintes valores: média igual a 69,8 e desvio padrão igual a 1,86, colectou uma amostra aleatória de 50 pacientes e mediu as suas frequências cardíacas, obtendo a média de 70,5. Ele estava correto? Use significância de 5%.

13.17.72. Um laboratório de vacinas contra febre aftosa reivindicou que ela imuniza 90% dos animais. Em uma amostra de 200 animais, nos quais foram

aplicados a vacina, 160 foram imunizados. Verificar se a declaração do fabricante é verdadeira ao nível de 5%.

13.17.73. O ministro da saúde afirmou, numa conferência de imprensa, que o tempo médio de espera para uma consulta de oftamologia no SNS é de 25 dias. Os jornalistas desconfiam que o tempo de espera é maior. Numa amostra de 36 utentes encontrou-se um tempo médio de espera de 28 dias, com desvio padrão 5.

13.17.74. Suponhamos que em indivíduos normais quanto à visão, a pressão intra-ocular seja uma variável aleatória normalmente distribuída com média 20 e variância 4 (em unidade de mm de mercúrio). Um cientista, querendo por à prova a sua hipótese de que o glaucoma causa um aumento tencional, mediu as pressões de 16 pacientes portadores de glaucoma, obtendo uma média igual a 24. O cientista deve ou não manter sua hipótese, ao nível de significância $\alpha = 0,005$?

13.17.75. As alturas de 20 recém nascidos foram medidas no sector de Pediatria dum Hospital. Os resultados obtidos são apresentados abaixo, em cm.

40 52 49 54 50 49 47 52 49 50 52 50 47 49 51 46 50 49 50

Existem evidências estatísticas a um nível de significância de 1% de que a altura média das crianças nascidas nesse hospital seja menor do que 50 cm?

13.17.76. Uma equipe de pesquisadora está disposta a supor que a pressão arterial sistólica em uma população de homens segue uma distribuição aproximadamente gaussiana. Uma amostra aleatória de 64 homens dessa população possui pressão sistólica média de 133 mmHg e desvio padrão de 16 mmHg. A um nível de 5% de significância:

- a) Existem evidências estatísticas de que a pressão sistólica média dessa população de homens seja maior do que 130 mmHg?
- b) Existem evidências estatísticas de que a pressão sistólica média dessa população de homens seja diferente de 130 mmHg?

13.17.77. Um estudo é realizado para determinar a relação entre uma droga e certa anomalia em embriões de frango. Injectou-se 50 ovos fertilizados com a droga no quarto dia de incubação. No vigésimo dia de incubação, os embriões foram examinados e 7 apresentaram a anomalia. Os resultados desse

experimento contêm evidências estatísticas a um nível de significância de que a proporção de embriões com anomalias seja inferior a 25%?

13.17.78. Diana M. Bailey (*The American Journal of Occupational Therapy*, 44,23-29) conduziu um estudo para analisar as razões pelas quais terapeutas ocupacionais deixam a profissão. A amostra foi constituída por mulheres que trabalham como terapeutas ocupacionais e que deixaram a profissão de forma temporária ou permanente. Dos 696 indivíduos que responderam a uma pesquisa para a colecta de dados, 63% decidiram deixar seus empregos para ter e cuidar de seus próprios filhos. Com base nestes dados, existem evidências estatísticas suficientes a um nível de significância de 1% para se concluir que:

- a) A população amostrada possui uma proporção de indivíduos que decidiram deixar seus empregos para dedicar-se a ter e cuidar de seus próprios filhos diferente de 60%?
- b) A população amostrada possui mais de 60% de indivíduos que decidiram deixar seus empregos para dedicar-se a ter e cuidar de seus próprios filhos?

13.17.79. Henning e colegas (*American Journal of Public Health*, 82, 885-888) encontraram que 66% das crianças de uma amostra de 670 crianças completaram a série completa de vacinas contra a Hepatite B. Com base nesses dados, pode-se concluir que a população amostrada possui uma proporção de crianças com a série completa de vacinas contra a Hepatite B estatisticamente maior do que 60%? (use $\alpha = 5\%$)

13.17.80. O estresse afecta a capacidade de memorização das testemunhas oculares? Este problema foi estudado em um experimento que testou a memória visual de 40 testemunhas uma semana após o interrogatório normal de um suspeito que cooperava (situação sem estresse), e de outro grupo de 40 testemunhas de um interrogatório exaustivo de um suspeito que não cooperava (situação com estresse). Os números de detalhes lembrados uma semana após o incidente estão resumidos aqui.

	n	\bar{x}	s^2
Sem estresse	40	53.3	45.3
Com estresse	40	11	13.2

Ao nível de 1% de significância, teste a afirmação do artigo de que “o cansaço concorre para diminuir a quantidade de detalhes lembrados”.

- 13.17.81.** (Adaptado de Triola, 1999) Em um experimento destinado a testar os efeitos do álcool, registaram-se os erros em um teste de habilidade visual e motora para um grupo de tratamento, que bebeu etanol, e também para outro grupo, ao qual foi dado um placebo. Os resultados são mostrados a seguir.

	n	\bar{x}	s
Grupo de tratamento	8	23.1	7.3
Grupo placebo	11	20	9.8

Ao nível de 5% de significância, teste a afirmação de que os dois grupos provêm de populações com a mesma média. Esses resultados apoiam a crença geral de que bebida é prejudicial para motoristas?

- 13.17.82.** (Adaptado de Triola, 1999) O Captopril é um remédio destinado a baixar a pressão sistólica. Feito o teste com este remédio em pacientes, mediram-se suas pressões sistólicas (em mm de mercúrio) antes e depois de tomarem o remédio, com os resultados constantes na tabela a seguir.

Pessoa	A	B	C	D	E	F	G	H	I	J	K	L
Antes	200	174	198	170	179	182	193	209	185	155	169	210
Depois	191	170	177	167	159	151	176	183	159	145	146	177

Ao nível de 1% de significância, pode-se afirmar que existe diferença entre as pressões sistólicas médias antes e depois da ingestão do remédio?

- 13.17.83.** Experimentou-se uma nova máquina de enchimento estéril de frascos de antibióticos, obtendo-se para os 33 frascos, o peso médio de 1 093 mg e um desvio padrão de 36 mg. Pelo processo de enchimento manual, uma amostra de 30 frascos deu o peso médio de 1 122 mg e um desvio padrão de 23 mg. Acha que existe uma diferença significativa entre as médias dos pesos obtidos pelos dois processos?

- 13.17.84.** O peso (kg) ao nascer de 160 bebês do sexo masculino, cujas mães tomavam uma determinada vitamina, foi de 3,675 Kg com desvio padrão de 0,850 Kg. A média da população de bebês do sexo masculino é de 3,39 Kg. Com base nesses dados, é possível afirmar que a vitamina tenha algum efeito sobre o peso dos bebês ao nascer? Adote um nível de significância de 3%.

13.17.85. Um grupo de 20 doentes é alimentado com uma determinada dieta e um segundo grupo de 10 doentes é alimentado com uma dieta diferente. Depois de um certo período de tempo, o aumento médio no peso para o primeiro grupo foi de 124.7 g, com um desvio padrão de 9g. Já para o segundo grupo, aumento médio no peso foi de 130.8g, com um desvio padrão de 12g. Usando um nível de significância de 5%:

- a) Verifique a homogeneidade dos grupos.
- b) A diferença do aumento dos pesos médios dos grupos

13.17.86. Determinaram-se os níveis de fosfatase sérica de 24 crianças com infecção malárica, obtendo se um nível médio de 2.81 mg/dl com desvio padrão de 0.52. Sabendo que o nível de fosfatase sérica padrão é de 2.45 mg/dl, pode se concluir com um nível de significância de 0.01 que a fosfatase sérica é alterada durante a infecção malárica?

13.17.87. Supondo que, de 10000 indivíduos com idade entre 50 e 60 anos e com história familiar de câncer gástrico, 350 apresentaram esse tipo de carcinoma. Os estudos demonstraram que a prevalência dessa doença na população é de 0.03. Verifique se a proporção desses indivíduos com história familiar de neoplasia gástrica é diferente da observada na população (use nível de significância de 5%).

13.17.88. Dois produtos comerciais D1 e D2, recomendados para alívio de dores musculares, foram testados em duas amostras independentes de pacientes da mesma idade, do sexo feminino e com condições clínicas semelhantes. Cada grupo estava constituído de 32 indivíduos e o tempo decorrido entre a ingestão das drogas e o alívio do sintoma para os dois grupos foi de 5.1 e 5.3 minutos respectivamente. As variâncias paramétricas dos produtos D1 e D2 são respectivamente, 1.68 e 1.62. Teste a um nível de significância de 1% se a droga 1 alivia a dor muscular em um tempo inferior ao da droga 2.

13.17.89. Uma pesquisa foi conduzida para estudar o efeito da aspirina no enfarte do miocárdio. Um grupo de 10000 pessoas do sexo masculino, de 40 anos de idade, cada uma recebeu 100 mg de droga por dia e foram observados durante 10 anos. Ocorreram manifestações de enfarte coronariano em 14. Outro grupo de 9500 indivíduos do mesmo sexo e com a mesma idade foram

seguidos por igual período, mas não receberam aspirina, tendo ocorrido 30 casos de enfarte do miocárdio. A um nível de significância de 5% pode se concluir que a aspirina diminui a incidência da doença coronária?

14. CORRELAÇÃO E REGRESSÃO LINEAR

14.1. Introdução

Considere a existência de uma variável quantitativa X a qual acreditamos apresentar alguma relação com uma outra variável quantitativa Y . Por exemplo: consumo de eletricidade e valor da conta de energia eléctrica; idade e tempo de reacção um estímulo; temperatura e tempo de uma reacção química, dentre outros.

Em situações como as citadas, a construção de um gráfico de dispersão dos valores de X versus os valores de Y , se constitui numa ferramenta estatística simples, porém muito útil, para investigar a existência de uma possível relação entre essas duas variáveis. Adicionalmente, podemos também fazer uso dos coeficientes de correlação, como por exemplo, o de Pearson, apresentado a seguir.

14.2. Coeficiente de Correlação de Pearson

O coeficiente de correlação de Pearson é utilizado quando desejamos verificar a existência de associação linear entre duas variáveis quantitativas, X e Y , e é obtido dividindo-se a covariância de X e Y pelo produto dos respectivos desvios-padrão de ambas as variáveis, isto é:

$$\rho = \frac{Cov(X, Y)}{\sigma_x * \sigma_y} = \frac{E(XY) - E(X) * E(Y)}{\sigma_x * \sigma_y}$$

Esse coeficiente resulta sempre em um valor entre -1 e 1 e sua interpretação depende do seu valor numérico e do seu sinal. Quanto mais próximo de -1 e 1 , mais forte é o grau de relação linear existente entre X e Y e, quanto mais próximo de 0 , mais fraco é o grau desta relação. Uma correlação linear negativa indica que quando o valor de uma variável aumenta, o valor da outra diminui e, uma correlação linear positiva, indica que quando o valor de uma variável aumenta, o valor da outra também aumenta.

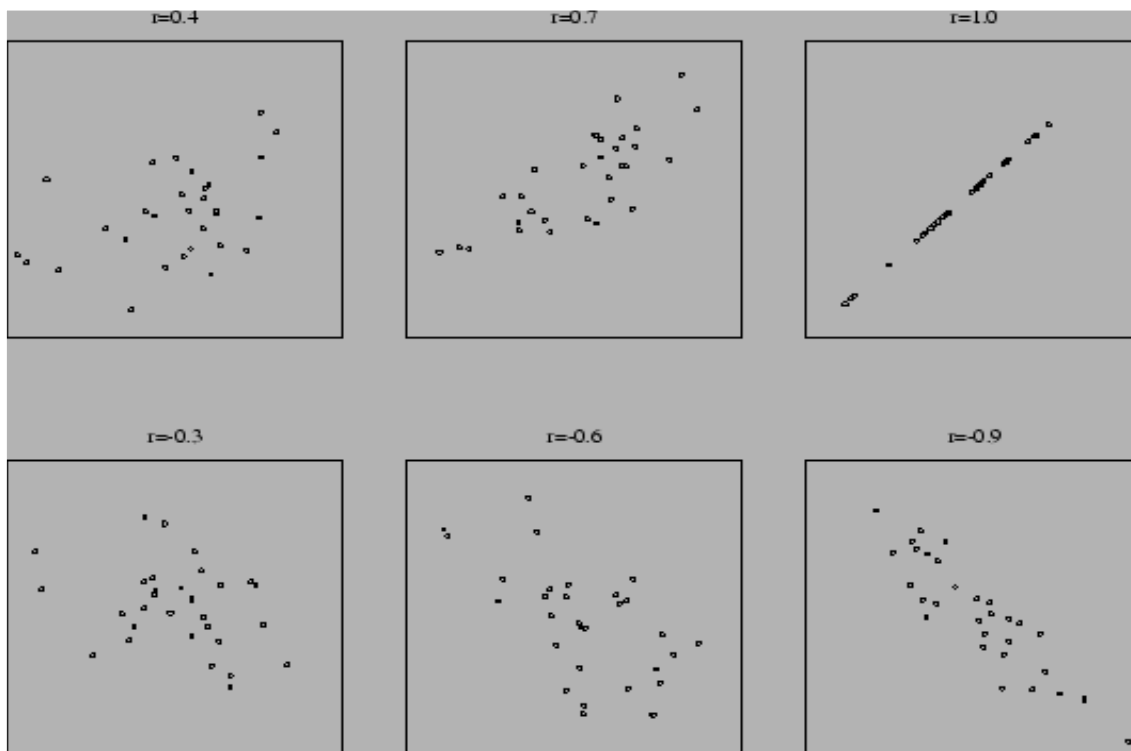
Para uma amostra de tamanho n , em que para cada indivíduo i ($i = 1, \dots, n$) observamos os pares de valores (x_i, y_i) , o coeficiente de correlação linear entre X e Y é:

$$\begin{aligned}
 r &= \frac{s_{xy}}{\sqrt{s_{xx} * s_{yy}}} = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{[\sum_{i=1}^n (x_i - \bar{x})^2] * [\sum_{i=1}^n (y_i - \bar{y})^2]}} \\
 &= \frac{\sum_{i=1}^n x_i * y_i - n * \bar{x} * \bar{y}}{\sqrt{[\sum_{i=1}^n x_i^2 - n\bar{x}^2] * [\sum_{i=1}^n y_i^2 - n\bar{y}^2]}} \\
 &= \frac{n \sum_{i=1}^n x_i * y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n * \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} * \sqrt{n * \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}}
 \end{aligned}$$

A seguinte tabela fornece um guia de como podemos descrever uma correlação em palavras dado o valor numérico. É claro que as interpretações dependem de cada contexto em particular.

Valor de r (+ ou -)	Interpretação
[0.00 – 0.19]	Correlação muito fraca
[0.20 – 0.39]	Correlação fraca
[0.40 – 0.69]	Correlação moderada
[0.70 – 0.89]	Correlação forte
[0.90 – 1.00]	Correlação muito forte

Abaixo estão exemplos de diagramas de dispersão de Pearson com seus coeficientes de correlação correspondentes.



A ausência de relação linear, quando indicada por este coeficiente, não implica a ausência de relação entre elas. Outro tipo de relação pode estar presente, como, por exemplo, a não-linear.

Exemplo 14.1 (Coeficiente de Correlação)

Os dados a seguir correspondem à variável renda familiar (X) e gasto com alimentação (Y) (em unidades monetárias) para uma amostra de 25 famílias.

X	Y	X	Y	X	Y
3	1,5	40	10,0	100	40,0
5	2,0	50	20,0	120	30,0
10	6,0	60	20,0	120	40,0
10	7,0	70	25,0	140	40,0
20	10,0	70	30,0	150	50,0
20	12,0	80	25,0	180	40,0
20	15,0	100	40,0	180	50,0
30	8,0	100	35,0	200	60,0
200	50,0				

Calcular o coeficiente de correlação entre essas variáveis.

Resolução

$$r = \frac{s_{xy}}{\sqrt{s_{xx} * s_{yy}}} = \frac{\sum_{i=1}^n x_i * y_i - n * \bar{x} * \bar{y}}{\sqrt{[\sum_{i=1}^n x_i^2 - n\bar{x}^2] * [\sum_{i=1}^n y_i^2 - n\bar{y}^2]}}$$

$$= \frac{80774.5 - 25 * 83.12 * 27.2}{(271934 - 25 * 83.12) * (24899.25 - 25 * 27.2)}$$

$$r = 0.948$$

Interpretação: Existe uma relação positiva muito forte entre renda familiar (X) e gasto com alimentação (Y) das 25 famílias

14.2.1. Teste de significância do coeficiente de correlação de Pearson

Quando se coleta uma amostra de n pares de valores das variáveis (X, Y) e se calcula o seu coeficiente de correlação r , o que se quer saber é se esse valor de r é significativo. Para se fazer isso, vai-se assumir como hipótese nula H_0 que não existe correlação para a população das variáveis X e Y (o que implicaria que o valor obtido

para r ocorreu por mero acaso). Costuma-se denotar o coeficiente de correlação para a população das variáveis X e Y por ρ (não confundir com o coeficiente de correlação de Spearman).

Portanto, temos dois coeficientes de correlação: um para a população de valores de X e Y , denotado por ρ , que é puramente teórico; e o outro para uma amostra de n pares retirada da população, denotado por r . A figura abaixo ilustra a situação.

O valor de r é usado para estimar o coeficiente de correlação ρ e o teste de significância desse valor consiste em assumir como H_0 que $\rho = 0$ (ausência de correlação) para verificar se sob tal hipótese o valor obtido para r é muito ou pouco provável. Se a probabilidade de se obter o valor de r for menor que um certo valor crítico (por exemplo, 0,05), rejeita-se H_0 e assume-se como mais provável a hipótese alternativa, segundo a qual $\rho \neq 0$.

Para que o teste de significância do coeficiente de correlação de Pearson entre X e Y seja realizado é necessário que a distribuição de probabilidade conjunta para a população das variáveis X e Y seja normal bidimensional.

Em geral, quando se trabalha com amostras de n pares de valores (x, y) onde $n \geq 30$, a condição de normalidade das duas variáveis é satisfeita. Assumindo que a condição acima é válida, temos as seguintes hipóteses:

$$\begin{cases} H_0: \rho = 0 \\ H_1: \rho \neq 0 \end{cases}$$

Exemplo 14.2 (Teste de significância de Coeficiente de Correlação)

Teste a significância do exemplo 14.1

Resolução

1. $\alpha = 0.05$;
2. Como a amostra contém n pares de dados, então:

$$t_{\frac{\alpha}{2}; n-2} = t_{0.025; 25-1} = 2.064$$

3. Estatística do teste:

$$t_0 = r * \sqrt{\frac{n-2}{1-r^2}} = 0.948 * \sqrt{\frac{25-2}{1-0.948^2}} = 14.28$$

4. Decisão:

$$t_0 = 14.28 > t_{0.025; 25-1} = 2.064$$

Rejeita-se H_0

5. Conclusão:

Conclui-se que o valor de $r = 0.948$ obtido para a amostra é significativo e que existe correlação r entre as variáveis X e Y com nível de significância igual a 5%.

Uma vez constatada a existência da relação linear entre duas variáveis, é de usual interesse descrever essa relação por meio de uma equação linear.

14.3. Análise de Regressão

Análise de regressão é uma técnica de modelagem utilizada para analisar a relação entre uma variável dependente (Y) e uma ou mais variáveis independentes $X_1, X_2, X_3, \dots, X_n$. O objetivo dessa técnica é identificar (estimar) uma função que descreve, o mais próximo possível, a relação entre essas variáveis e assim podermos prever o valor que a variável dependente (Y) irá assumir para um determinado valor da variável independente X . O modelo de regressão poderá ser escrito genericamente como:

$$Y = f(X_1, X_2, X_3, \dots, X_n) + \varepsilon$$

onde o termo ε representa uma perturbação aleatória na função, ou o erro da aproximação. O número de variáveis independentes varia de uma aplicação para outra, quando se tem apenas uma variável independente chama-se Modelo de Regressão Simples, quando se tem mais de uma variável independente chama-se de Modelo de Regressão Múltipla. A forma da função também varia, podendo ser representada por um modelo linear, polinomial ou até mesmo uma função não linear.

14.3.1. Regressão Linear Simples

Se uma relação linear é válida para sumarizar a dependência observada entre duas variáveis quantitativas, então a equação que descreve esta relação é dada por:

$$Y = \beta_0 + \beta_1 X$$

Os valores observados não se encontram, contudo, exactamente sobre esta linha recta, ou seja, existe uma diferença entre o valor observado e o valor fornecido pela equação.

Esta diferença é denominada erro e é representada por ε . Este erro é assumido ser um erro estatístico, isto é, uma variável aleatória que quantifica a falha do modelo em ajustar-se aos dados exactamente. Tal erro pode ser devido ao efeito, dentre outros, de variáveis não consideradas e de erros de medição. Incorporando esse erro à equação anterior temos:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

que é denominado modelo de regressão linear simples. Para cada indivíduo i ($i = 1, 2, 3, \dots, n$) na amostra, o modelo fica representado por:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

A variável X denominada variável regressora ou independente, é considerada uma variável controlada pelo analista dos dados e medida com erro desprezível.

Já Y , denominada variável resposta ou dependente, é considerada uma variável aleatória, isto é, existe uma distribuição de probabilidade para Y em cada valor possível de X .

É muito frequente, na prática, encontrarmos situações em que Y tenha distribuição Normal. Nesses casos, os erros ε_i (em que alguns são positivos e outros negativos) são assumidos serem normalmente distribuídos com média zero e variância constante desconhecida σ^2 , bem como independentes, isto é, o valor de um erro independe do valor de qualquer outro erro. Sendo assim, a média e a variância da variável Y serão, respectivamente:

$$E(Y/X = x) = E(\beta_0 + \beta_1 x + \varepsilon) = \beta_0 + \beta_1 x$$

$$V(Y/X = x) = V(\beta_0 + \beta_1 x + \varepsilon) = \sigma^2$$

14.3.2. Método dos Mínimos Quadrados

Com base nos n pares de observações $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$, o método de estimação por MQO consiste em escolher β_0 e β_1 de modo que a soma dos quadrados dos erros, ε_i ($i = 1, \dots, n$), seja mínima. Note que $\varepsilon_i = Y_i - \hat{\beta}_0 - \widehat{B}_1 x_i$

Para minimizar esta soma, que é expressa por:

$$SQ = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \widehat{B}_1 x_i)^2$$

devemos, inicialmente, diferenciar a expressão anterior com respeito a β_0 e β_1 , em seguida, igualar a zero as expressões resultantes. Feito isso, e após algumas operações algébricas, os estimadores resultantes são:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i x_i - n \bar{y} \bar{x}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}$$

Onde, \bar{y} é a média amostral dos y_i e \bar{x} a média amostral dos x_i . Logo:

$$E(\widehat{Y/X}) = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

É o modelo de regressão linear simples ajustado, em que $E(\widehat{Y/X})$, denotado também \hat{Y} por simplicidade, e é o valor médio de Y para qualquer valor $X = x$ que esteja na variação observada de X .

Os desvios $e_i = Y_i - \hat{Y}$ são denominados resíduos e são considerados uma amostra aleatória dos erros. Por este facto, uma análise gráfica dos resíduos é, em geral, realizada para verificar as suposições assumidas para os erros ε_i .

Exemplo 14.3 (Modelo de regressão)

Considere as variáveis gasto com alimentação e renda familiar, obtenha a equação de regressão.

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} = \frac{\sum_{i=1}^n y_i x_i - n \bar{y} \bar{x}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{80774.5 - 25 * 83.12 * 27.2}{271934 - 25 * 83.12^2} = 0.245$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 27.2 - 0.245 * 83.12 = 6.847$$

Logo, a reta de regressão estimada da variável Gasto de alimentação (Y) em função da Renda familiar (X) é

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

$$\hat{Y} = 6.847 + 0.245X$$

Esta reta é o “melhor” ajustamento para estes dados e seria diferente para cada amostra das variáveis X e Y, retiradas desta mesma população. Esta reta pode ser considerada uma estimativa da verdadeira linha de regressão onde 0.245 seria uma estimativa do valor β_1 (parâmetro angular) e 6.847 uma estimativa do valor β_0 (parâmetro linear), que são os verdadeiros coeficientes de regressão.

14.3.3. Estimativa da variância do termo erro

O termo erro, U, é uma variável aleatória, supostamente com média zero e variância constante. Então, intuitivamente parece plausível usar os resíduos da reta de regressão pelos método dos mínimos quadrados para se estimar a variância σ^2 dos termos “erro”. A variância amostral desses resíduos é igual a:

$$s_e^2 = \frac{s_{yy} - \hat{\beta}_1 * s_{xy}}{n - 2}$$

$$= \frac{\sum n * \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2 - \hat{\beta}_1 * [n \sum_{i=1}^n x_i * y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i]}{n - 2}$$

$$s_e = \sqrt{\frac{s_{yy} - \hat{\beta}_1 * s_{xy}}{n - 2}}$$

Exemplo: Determine o erro padrão da regressão

$$s_e = 5.4268$$

14.3.4. A Adequação do modelo de regressão linear ajustado

Após ajustar o modelo de regressão linear simples devemos, antes de adoptá-lo definitivamente para fazer predições (interpolações), verificar:

1. Se o modelo se ajusta bem aos dados e,

2. Se as suposições básicas se encontram satisfeitas.

Quanto a qualidade de ajuste do modelo, podemos fazer uso do coeficiente de determinação, R^2 que nos fornece a percentagem da variação total de Y explicada pelo modelo, ou seja, o percentual da variabilidade da variável dependente Y explicada pela variável independente X . Em regressão linear simples esse coeficiente pode ser obtido por $R^2 = r^2$, em que r é o coeficiente de correlação de Pearson amostral. O coeficiente de determinação varia de 0 a 1 (ou 0 a 100%), sendo que quanto mais próximo de 1 (100%), melhor o ajuste do modelo considerado.

Podemos, também, obter o coeficiente de determinação R^2 a partir da análise de variância da regressão, em que a variação o total de Y é decomposta como mostrado na tabela ANOVA a seguir. Fazendo-se uso da decomposição apresentada, temos que:

$$R^2 = \frac{SQ_{reg}}{SQ_{total}}$$

Tabela ANOVA

Fonte de variação	GL	Soma dos Quadrados	Quadrado Médio
Regressão	$p - 1$	$SQ_{Reg} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$QM_{Reg} = \frac{SQ_{Reg}}{p - 1}$
Resíduos	$n - p$	$SQ_{Res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$	$QM_{Res} = \frac{SQ_{Res}}{n - p}$
Total	$n - 1$	$SQ_{total} = \sum_{i=1}^n (y_i - \bar{y})^2$	

p = número de parâmetros do modelo e n = tamanho amostral.

Para testarmos a significância do parâmetro β_1 , o que, na prática, significa verificar se a covariável X influencia a resposta Y , testamos as hipóteses:

$$\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$$

A estatística de teste utilizada para esta finalidade é dada por:

$$F = \frac{QM_{Reg}}{QM_{Res}}$$

Onde QM_{Reg} e QM_{Res} são, respectivamente, os quadrados médios da regressão e dos resíduos apresentados na tabela ANOVA. Sob H_0 , tal estatística tem distribuição F de

Snedecor-Fisher com $p - 1$ e $n - p$ graus de liberdade. Assim, rejeitamos-se H_0 se o valor calculado de F for maior que o valor de F tabelado a um nível α de significância pré-estabelecido.

14.3.5. Interpretação dos parâmetros do modelo

Se o modelo de regressão linear simples (MRLS) for considerado adequado para descrever a relação linear entre Y e X , os coeficientes β_0 e β_1 são interpretados do seguinte modo:

1. Se a variação dos dados em X incluir $x = 0$, então o intercepto β_0 é a resposta esperada (média) em $x = 0$. Caso contrário, β_0 não apresenta interpretação prática;
2. O parâmetro β_1 é interpretado como a mudança no valor esperado de Y produzido por uma unidade de mudança em X .

Referências Bibliográficas

BUSSAB, W.O. e Morettin, P.A. *Estatística Básica*. São Paulo: Atual, 2007.

FONSECA, J.S. e Martins, G.A. *Curso de Estatística*. São Paulo: Atlas, 1993.

LAPPONI, J.C. *Estatística usando Excel 5 e 6*. São Paulo: Lapponi Treinamento e Editora, 1997.

MORETTIN, L.G. *Estatística Básica – Vol. 2 – Inferência*. São Paulo: Makron Books, 1999.

MORETTIN, L.G. *Estatística Básica – Vol.1 – Probabilidade*. São Paulo: Makron Books, 1999.

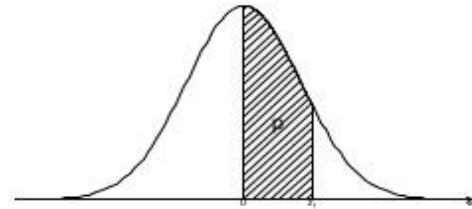
STEVENSON, W.J. *Estatística Aplicada à Administração*. São Paulo: Harbra, 2004.

TIBONI, C.G. R. *Estatística Básica para o curso de Turismo*. São Paulo: Atlas, 2002.

TOLEDO, G. L. e Ovalle, I.I. *Estatística Básica*. São Paulo: Atlas, 1985.

Apêndice

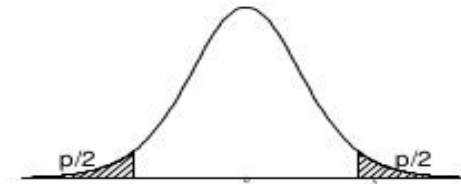
Distribuição Normal



	0	1	2	3	4	5	6	7	8	9
0,0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586
0,1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535
0,2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409
0,3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173
0,4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793
0,5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240
0,6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490
0,7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524
0,8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327
0,9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891
1,0	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214
1,1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298
1,2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147
1,3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41309	0.41466	0.41621	0.41774
1,4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189
1,5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408
1,6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449
1,7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327
1,8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062
1,9	0.47128	0.47193	0.47257	0.47320	0.47381	0.47441	0.47500	0.47558	0.47615	0.47670
2,0	0.47725	0.47778	0.47831	0.47882	0.47932	0.47982	0.48030	0.48077	0.48124	0.48169
2,1	0.48214	0.48257	0.48300	0.48341	0.48382	0.48422	0.48461	0.48500	0.48537	0.48574
2,2	0.48610	0.48645	0.48679	0.48713	0.48745	0.48778	0.48809	0.48840	0.48870	0.48899
2,3	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158
2,4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361
2,5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520
2,6	0.49534	0.49547	0.49560	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643
2,7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.49720	0.49728	0.49736
2,8	0.49744	0.49752	0.49760	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807
2,9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861
3,0	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49896	0.49900
3,1	0.49903	0.49906	0.49910	0.49913	0.49916	0.49918	0.49921	0.49924	0.49926	0.49929
3,2	0.49931	0.49934	0.49936	0.49938	0.49940	0.49942	0.49944	0.49946	0.49948	0.49950
3,3	0.49952	0.49953	0.49955	0.49957	0.49958	0.49960	0.49961	0.49962	0.49964	0.49965
3,4	0.49966	0.49968	0.49969	0.49970	0.49971	0.49972	0.49973	0.49974	0.49975	0.49976
3,5	0.49977	0.49978	0.49978	0.49979	0.49980	0.49981	0.49981	0.49982	0.49983	0.49983
3,6	0.49984	0.49985	0.49985	0.49986	0.49986	0.49987	0.49987	0.49988	0.49988	0.49989
3,7	0.49989	0.49990	0.49990	0.49990	0.49991	0.49991	0.49992	0.49992	0.49992	0.49992
3,8	0.49993	0.49993	0.49993	0.49994	0.49994	0.49994	0.49994	0.49995	0.49995	0.49995
3,9	0.49995	0.49995	0.49996	0.49996	0.49996	0.49996	0.49996	0.49996	0.49997	0.49997

Tabela 1: Probabilidades $p = P[0 \leq Z \leq Z_i]$ da Distribuição Normal padrão com valores de Z_i dados nas margens da tabela.

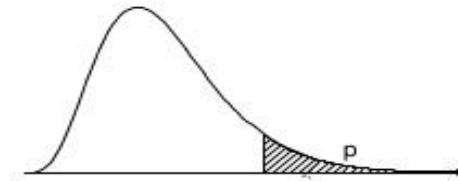
Distribuição t de Student



	90%	80%	70%	60%	50%	40%	30%	20%	10%	9%	8%	7%	6%	5%	4%	3%	2%	1%	0.5%	0.2%	0.1%
2	0.142	0.289	0.445	0.617	0.816	1.061	1.386	1.886	2.920	3.104	3.320	3.578	3.896	4.303	4.849	5.643	6.965	9.925	14.089	22.327	31.599
3	0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.353	2.471	2.605	2.763	2.951	3.182	3.482	3.896	4.541	5.841	7.453	10.215	12.924
4	0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.226	2.333	2.456	2.601	2.776	2.999	3.298	3.747	4.604	5.598	7.173	8.610
5	0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.098	2.191	2.297	2.422	2.571	2.757	3.003	3.365	4.032	4.773	5.893	6.869
6	0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.019	2.104	2.201	2.313	2.447	2.612	2.829	3.143	3.707	4.317	5.208	5.959
7	0.130	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.895	1.966	2.046	2.136	2.241	2.365	2.517	2.715	2.998	3.499	4.029	4.785	5.408
8	0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	1.928	2.004	2.090	2.189	2.306	2.449	2.634	2.896	3.355	3.833	4.501	5.041
9	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	1.899	1.973	2.055	2.150	2.262	2.398	2.574	2.821	3.250	3.690	4.297	4.781
10	0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.812	1.877	1.948	2.028	2.120	2.228	2.359	2.527	2.764	3.169	3.581	4.144	4.587
11	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	1.859	1.928	2.007	2.096	2.201	2.328	2.491	2.718	3.106	3.497	4.025	4.437
12	0.128	0.259	0.395	0.539	0.695	0.873	1.083	1.356	1.782	1.844	1.912	1.989	2.076	2.179	2.303	2.461	2.681	3.055	3.428	3.930	4.318
13	0.128	0.259	0.394	0.538	0.694	0.870	1.079	1.350	1.771	1.832	1.899	1.974	2.060	2.160	2.282	2.436	2.650	3.012	3.372	3.852	4.221
14	0.128	0.258	0.393	0.537	0.692	0.868	1.076	1.345	1.761	1.821	1.887	1.962	2.046	2.145	2.264	2.415	2.624	2.977	3.326	3.787	4.140
15	0.128	0.258	0.393	0.536	0.691	0.866	1.074	1.341	1.753	1.812	1.878	1.951	2.034	2.131	2.249	2.397	2.602	2.947	3.286	3.733	4.073
16	0.128	0.258	0.392	0.535	0.690	0.865	1.071	1.337	1.746	1.805	1.869	1.942	2.024	2.120	2.235	2.382	2.583	2.921	3.252	3.686	4.015
17	0.128	0.257	0.392	0.534	0.689	0.863	1.069	1.333	1.740	1.798	1.862	1.934	2.015	2.110	2.224	2.368	2.567	2.898	3.222	3.646	3.965
18	0.127	0.257	0.392	0.534	0.688	0.862	1.067	1.330	1.734	1.792	1.855	1.926	2.007	2.101	2.214	2.356	2.552	2.878	3.197	3.610	3.922
19	0.127	0.257	0.391	0.533	0.688	0.861	1.066	1.328	1.729	1.786	1.850	1.920	2.000	2.093	2.205	2.346	2.539	2.861	3.174	3.579	3.883
20	0.127	0.257	0.391	0.533	0.687	0.860	1.064	1.325	1.725	1.782	1.844	1.914	1.994	2.086	2.197	2.336	2.528	2.845	3.153	3.552	3.850
21	0.127	0.257	0.391	0.532	0.686	0.859	1.063	1.323	1.721	1.777	1.840	1.909	1.988	2.080	2.189	2.328	2.518	2.831	3.135	3.527	3.819
22	0.127	0.256	0.390	0.532	0.686	0.858	1.061	1.321	1.717	1.773	1.835	1.905	1.983	2.074	2.183	2.320	2.508	2.819	3.119	3.505	3.792
23	0.127	0.256	0.390	0.532	0.685	0.858	1.060	1.319	1.714	1.770	1.832	1.900	1.978	2.069	2.177	2.313	2.500	2.807	3.104	3.485	3.768
24	0.127	0.256	0.390	0.531	0.685	0.857	1.059	1.318	1.711	1.767	1.828	1.896	1.974	2.064	2.172	2.307	2.492	2.797	3.091	3.467	3.745
25	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.316	1.708	1.764	1.825	1.893	1.970	2.060	2.167	2.301	2.485	2.787	3.078	3.450	3.725
26	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.315	1.706	1.761	1.822	1.890	1.967	2.056	2.162	2.296	2.479	2.779	3.067	3.435	3.707
27	0.127	0.256	0.389	0.531	0.684	0.855	1.057	1.314	1.703	1.758	1.819	1.887	1.963	2.052	2.158	2.291	2.473	2.771	3.057	3.421	3.690
28	0.127	0.256	0.389	0.530	0.683	0.855	1.056	1.313	1.701	1.756	1.817	1.884	1.960	2.048	2.154	2.286	2.467	2.763	3.047	3.408	3.674
29	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.311	1.699	1.754	1.814	1.881	1.957	2.045	2.150	2.282	2.462	2.756	3.038	3.396	3.659
30	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.310	1.697	1.752	1.812	1.879	1.955	2.042	2.147	2.278	2.457	2.750	3.030	3.385	3.646
35	0.127	0.255	0.388	0.529	0.682	0.852	1.052	1.306	1.690	1.744	1.803	1.869	1.944	2.030	2.133	2.262	2.438	2.724	2.996	3.340	3.591
40	0.126	0.255	0.388	0.529	0.681	0.851	1.050	1.303	1.684	1.737	1.796	1.862	1.936	2.021	2.123	2.250	2.423	2.704	2.971	3.307	3.551
50	0.126	0.255	0.388	0.528	0.679	0.849	1.047	1.299	1.676	1.729	1.787	1.852	1.924	2.009	2.109	2.234	2.403	2.678	2.937	3.261	3.496
60	0.126	0.254	0.387	0.527	0.679	0.848	1.045	1.296	1.671	1.723	1.781	1.845	1.917	2.000	2.099	2.223	2.390	2.660	2.915	3.232	3.460
120	0.126	0.254	0.386	0.526	0.677	0.845	1.041	1.289	1.658	1.709	1.766	1.828	1.899	1.980	2.076	2.196	2.358	2.617	2.860	3.160	3.373

Tabela 2: Quantis da Distribuição t . Graus de liberdade na margem esquerda da tabela e probabilidades p dadas no topo da tabela tal que $\frac{p}{2} = P[t \geq t_c]$.

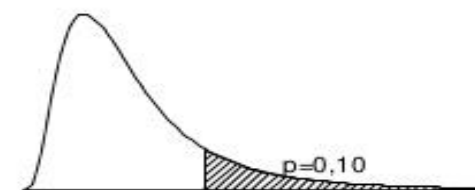
Distribuição χ^2



	99%	98%	97.5%	95%	90%	80%	70%	60%	50%	40%	30%	20%	10%	5%	4%	2.5%	2%	1%	0.2%	0.1%
1	0.000	0.001	0.001	0.004	0.016	0.064	0.148	0.275	0.455	0.708	1.074	1.642	2.706	3.841	4.218	5.024	5.412	6.635	9.550	10.828
2	0.020	0.040	0.051	0.103	0.211	0.446	0.713	1.022	1.386	1.833	2.408	3.219	4.605	5.991	6.438	7.378	7.824	9.210	12.429	13.816
3	0.115	0.185	0.216	0.352	0.584	1.005	1.424	1.869	2.366	2.946	3.665	4.642	6.251	7.815	8.311	9.348	9.837	11.345	14.796	16.266
4	0.297	0.429	0.484	0.711	1.064	1.649	2.195	2.753	3.357	4.045	4.878	5.989	7.779	9.488	10.026	11.143	11.668	13.277	16.924	18.467
5	0.554	0.752	0.831	1.145	1.610	2.343	3.000	3.655	4.351	5.132	6.064	7.289	9.236	11.070	11.644	12.833	13.388	15.086	18.907	20.515
6	0.872	1.134	1.237	1.635	2.204	3.070	3.828	4.570	5.348	6.211	7.231	8.558	10.645	12.592	13.198	14.449	15.033	16.812	20.791	22.458
7	1.239	1.564	1.690	2.167	2.833	3.822	4.671	5.493	6.346	7.283	8.383	9.803	12.017	14.067	14.703	16.013	16.622	18.475	22.601	24.322
8	1.646	2.032	2.180	2.733	3.490	4.594	5.527	6.423	7.344	8.351	9.524	11.030	13.362	15.507	16.171	17.535	18.168	20.090	24.352	26.124
9	2.088	2.532	2.700	3.325	4.168	5.380	6.393	7.357	8.343	9.414	10.656	12.242	14.684	16.919	17.608	19.023	19.679	21.666	26.056	27.877
10	2.558	3.059	3.247	3.940	4.865	6.179	7.267	8.295	9.342	10.473	11.781	13.442	15.987	18.307	19.021	20.483	21.161	23.209	27.722	29.588
11	3.053	3.609	3.816	4.575	5.578	6.989	8.148	9.237	10.341	11.530	12.899	14.631	17.275	19.675	20.412	21.920	22.618	24.725	29.354	31.264
12	3.571	4.178	4.404	5.226	6.304	7.807	9.034	10.182	11.340	12.584	14.011	15.812	18.549	21.026	21.785	23.337	24.054	26.217	30.957	32.909
13	4.107	4.765	5.009	5.892	7.042	8.634	9.926	11.129	12.340	13.636	15.119	16.985	19.812	22.362	23.142	24.736	25.472	27.688	32.535	34.528
14	4.660	5.368	5.629	6.571	7.790	9.467	10.821	12.078	13.339	14.685	16.222	18.151	21.064	23.685	24.485	26.119	26.873	29.141	34.091	36.123
15	5.229	5.985	6.262	7.261	8.547	10.307	11.721	13.030	14.339	15.733	17.322	19.311	22.307	24.996	25.816	27.488	28.259	30.578	35.628	37.697
16	5.812	6.614	6.908	7.962	9.312	11.152	12.624	13.983	15.338	16.780	18.418	20.465	23.542	26.296	27.136	28.845	29.633	32.000	37.146	39.252
17	6.408	7.255	7.564	8.672	10.085	12.002	13.531	14.937	16.338	17.824	19.511	21.615	24.769	27.587	28.445	30.191	30.995	33.409	38.648	40.790
18	7.015	7.906	8.231	9.390	10.865	12.857	14.440	15.893	17.338	18.868	20.601	22.760	25.989	28.869	29.745	31.526	32.346	34.805	40.136	42.312
19	7.633	8.567	8.907	10.117	11.651	13.716	15.352	16.850	18.338	19.910	21.689	23.900	27.204	30.144	31.037	32.852	33.687	36.191	41.610	43.820
20	8.260	9.237	9.591	10.851	12.443	14.578	16.266	17.809	19.337	20.951	22.775	25.038	28.412	31.410	32.321	34.170	35.020	37.566	43.072	45.315
21	8.897	9.915	10.283	11.591	13.240	15.445	17.182	18.768	20.337	21.991	23.858	26.171	29.615	32.671	33.597	35.479	36.343	38.932	44.522	46.797
22	9.542	10.600	10.982	12.338	14.041	16.314	18.101	19.729	21.337	23.031	24.939	27.301	30.813	33.924	34.867	36.781	37.659	40.289	45.962	48.268
23	10.196	11.293	11.689	13.091	14.848	17.187	19.021	20.690	22.337	24.069	26.018	28.429	32.007	35.172	36.131	38.076	38.968	41.638	47.391	49.728
24	10.856	11.992	12.401	13.848	15.659	18.062	19.943	21.652	23.337	25.106	27.096	29.553	33.196	36.415	37.389	39.364	40.270	42.980	48.812	51.179
25	11.524	12.697	13.120	14.611	16.473	18.940	20.867	22.616	24.337	26.143	28.172	30.675	34.382	37.652	38.642	40.646	41.566	44.314	50.223	52.620
26	12.198	13.409	13.844	15.379	17.292	19.820	21.792	23.579	25.336	27.179	29.246	31.795	35.563	38.885	39.889	41.923	42.856	45.642	51.627	54.052
27	12.879	14.125	14.573	16.151	18.114	20.703	22.719	24.544	26.336	28.214	30.319	32.912	36.741	40.113	41.132	43.195	44.140	46.963	53.023	55.476
28	13.565	14.847	15.308	16.928	18.939	21.588	23.647	25.509	27.336	29.249	31.391	34.027	37.916	41.337	42.370	44.461	45.419	48.278	54.411	56.892
29	14.256	15.574	16.047	17.708	19.768	22.475	24.577	26.475	28.336	30.283	32.461	35.139	39.087	42.557	43.604	45.722	46.693	49.588	55.792	58.301
30	14.953	16.306	16.791	18.493	20.599	23.364	25.508	27.442	29.336	31.316	33.530	36.250	40.256	43.773	44.834	46.979	47.962	50.892	57.167	59.703
35	18.509	20.027	20.569	22.465	24.797	27.836	30.178	32.282	34.336	36.475	38.859	41.778	46.059	49.802	50.928	53.203	54.244	57.342	63.955	66.619
40	22.164	23.838	24.433	26.509	29.051	32.345	34.872	37.134	39.335	41.622	44.165	47.269	51.805	55.758	56.946	59.342	60.436	63.691	70.618	73.402
45	25.901	27.720	28.366	30.612	33.350	36.884	39.585	41.995	44.335	46.761	49.452	52.729	57.505	61.656	62.901	65.410	66.555	69.957	77.179	80.077
50	29.707	31.664	32.357	34.764	37.689	41.449	44.313	46.864	49.335	51.892	54.723	58.164	63.167	67.505	68.804	71.420	72.613	76.154	83.657	86.661

Tabela 3: Quantis da Distribuição χ^2 . Graus de liberdade na margem esquerda da tabela e probabilidades p dadas no topo da tabela tal que $p = P[\chi^2 \geq \chi^2_p]$.

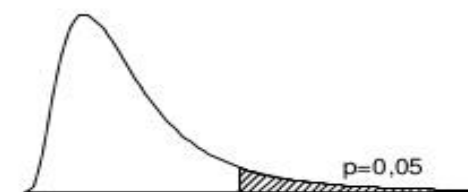
Distribuição F de Snedecor a 10% ($p=0.10$)



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	18	20	30	40	60	120
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.40	9.41	9.41	9.42	9.42	9.43	9.44	9.44	9.46	9.47	9.47	9.48
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.22	5.21	5.20	5.20	5.20	5.19	5.18	5.17	5.16	5.15	5.14
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.91	3.90	3.89	3.88	3.87	3.86	3.85	3.84	3.82	3.80	3.79	3.78
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.28	3.27	3.26	3.25	3.24	3.23	3.22	3.21	3.17	3.16	3.14	3.12
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.92	2.90	2.89	2.88	2.87	2.86	2.85	2.84	2.80	2.78	2.76	2.74
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.68	2.67	2.65	2.64	2.63	2.62	2.61	2.59	2.56	2.54	2.51	2.49
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.52	2.50	2.49	2.48	2.46	2.45	2.44	2.42	2.38	2.36	2.34	2.32
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.40	2.38	2.36	2.35	2.34	2.33	2.31	2.30	2.25	2.23	2.21	2.18
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.30	2.28	2.27	2.26	2.24	2.23	2.22	2.20	2.16	2.13	2.11	2.08
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.23	2.21	2.19	2.18	2.17	2.16	2.14	2.12	2.08	2.05	2.03	2.00
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.17	2.15	2.13	2.12	2.10	2.09	2.08	2.06	2.01	1.99	1.96	1.93
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.12	2.10	2.08	2.07	2.05	2.04	2.02	2.01	1.96	1.93	1.90	1.88
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.07	2.05	2.04	2.02	2.01	2.00	1.98	1.96	1.91	1.89	1.86	1.83
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.04	2.02	2.00	1.99	1.97	1.96	1.94	1.92	1.87	1.85	1.82	1.79
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	2.01	1.99	1.97	1.95	1.94	1.93	1.91	1.89	1.84	1.81	1.78	1.75
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.98	1.96	1.94	1.93	1.91	1.90	1.88	1.86	1.81	1.78	1.75	1.72
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.95	1.93	1.92	1.90	1.89	1.87	1.85	1.84	1.78	1.75	1.72	1.69
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.93	1.91	1.89	1.88	1.86	1.85	1.83	1.81	1.76	1.73	1.70	1.67
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.91	1.89	1.87	1.86	1.84	1.83	1.81	1.79	1.74	1.71	1.68	1.64
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.90	1.87	1.86	1.84	1.83	1.81	1.79	1.78	1.72	1.69	1.66	1.62
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.88	1.86	1.84	1.83	1.81	1.80	1.78	1.76	1.70	1.67	1.64	1.60
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.87	1.84	1.83	1.81	1.80	1.78	1.76	1.74	1.69	1.66	1.62	1.59
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.85	1.83	1.81	1.80	1.78	1.77	1.75	1.73	1.67	1.64	1.61	1.57
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.84	1.82	1.80	1.79	1.77	1.76	1.74	1.72	1.66	1.63	1.59	1.56
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.83	1.81	1.79	1.77	1.76	1.75	1.72	1.71	1.65	1.61	1.58	1.54
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.82	1.80	1.78	1.76	1.75	1.74	1.71	1.70	1.64	1.60	1.57	1.53
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.81	1.79	1.77	1.75	1.74	1.73	1.70	1.69	1.63	1.59	1.56	1.52
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.80	1.78	1.76	1.75	1.73	1.72	1.69	1.68	1.62	1.58	1.55	1.51
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.79	1.77	1.75	1.74	1.72	1.71	1.69	1.67	1.61	1.57	1.54	1.50
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.74	1.71	1.70	1.68	1.66	1.65	1.62	1.61	1.54	1.51	1.47	1.42
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.68	1.66	1.64	1.62	1.60	1.59	1.56	1.54	1.48	1.44	1.40	1.35
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.63	1.60	1.58	1.56	1.55	1.53	1.50	1.48	1.41	1.37	1.32	1.26

Tabela 4: Quantis da Distribuição F para probabilidade $p = P[F \geq F_i] = 0,10$. Graus de liberdade do numerador no topo e do denominador na margem esquerda.

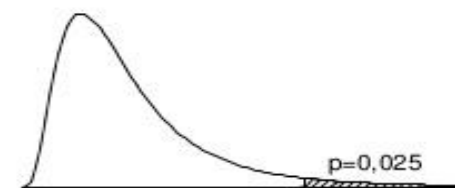
Distribuição F de Snedecor a 5% ($p=0.05$)



	1	2	3	4	5	6	7	8	9	10	12	14	15	16	18	20	30	40	60	120
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.42	19.43	19.43	19.44	19.45	19.46	19.47	19.48	19.49
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.71	8.70	8.69	8.67	8.66	8.62	8.59	8.57	8.55
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.87	5.86	5.84	5.82	5.80	5.75	5.72	5.69	5.66
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.64	4.62	4.60	4.58	4.56	4.50	4.46	4.43	4.40
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.96	3.94	3.92	3.90	3.87	3.81	3.77	3.74	3.70
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.53	3.51	3.49	3.47	3.44	3.38	3.34	3.30	3.27
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.24	3.22	3.20	3.17	3.15	3.08	3.04	3.01	2.97
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.03	3.01	2.99	2.96	2.94	2.86	2.83	2.79	2.75
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.86	2.85	2.83	2.80	2.77	2.70	2.66	2.62	2.58
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.74	2.72	2.70	2.67	2.65	2.57	2.53	2.49	2.45
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.64	2.62	2.60	2.57	2.54	2.47	2.43	2.38	2.34
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.55	2.53	2.51	2.48	2.46	2.38	2.34	2.30	2.25
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.48	2.46	2.44	2.41	2.39	2.31	2.27	2.22	2.18
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.42	2.40	2.38	2.35	2.33	2.25	2.20	2.16	2.11
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.37	2.35	2.33	2.30	2.28	2.19	2.15	2.11	2.06
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.33	2.31	2.29	2.26	2.23	2.15	2.10	2.06	2.01
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.29	2.27	2.25	2.22	2.19	2.11	2.06	2.02	1.97
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.26	2.23	2.21	2.18	2.16	2.07	2.03	1.98	1.93
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.22	2.20	2.18	2.15	2.12	2.04	1.99	1.95	1.90
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.20	2.18	2.16	2.12	2.10	2.01	1.96	1.92	1.87
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.17	2.15	2.13	2.10	2.07	1.98	1.94	1.89	1.84
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.15	2.13	2.11	2.08	2.05	1.96	1.91	1.86	1.81
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.13	2.11	2.09	2.05	2.03	1.94	1.89	1.84	1.79
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.11	2.09	2.07	2.04	2.01	1.92	1.87	1.82	1.77
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.09	2.07	2.05	2.02	1.99	1.90	1.85	1.80	1.75
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.08	2.06	2.04	2.00	1.97	1.88	1.84	1.79	1.73
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.06	2.04	2.02	1.99	1.96	1.87	1.82	1.77	1.71
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.05	2.03	2.01	1.97	1.94	1.85	1.81	1.75	1.70
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.04	2.01	1.99	1.96	1.93	1.84	1.79	1.74	1.68
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.95	1.92	1.90	1.87	1.84	1.74	1.69	1.64	1.58
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.86	1.84	1.82	1.78	1.75	1.65	1.59	1.53	1.47
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.78	1.75	1.73	1.69	1.66	1.55	1.50	1.43	1.35

Tabela 5: Quantis da Distribuição F para probabilidade $p = P[F \geq F_{\alpha}] = 0,05$. Graus de liberdade do numerador dado no topo e do denominador na margem esquerda.

Distribuição F de Snedecor a 2,5% ($p=0.025$)



	1	2	3	4	5	6	7	8	9	10	12	14	15	16	18	20	30	40	60	120
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.43	39.44	39.44	39.45	39.46	39.47	39.48	39.49
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.28	14.25	14.23	14.20	14.17	14.08	14.04	13.99	13.95
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.68	8.66	8.63	8.59	8.56	8.46	8.41	8.36	8.31
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.46	6.43	6.40	6.36	6.33	6.23	6.18	6.12	6.07
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.30	5.27	5.24	5.20	5.17	5.07	5.01	4.96	4.90
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.60	4.57	4.54	4.50	4.47	4.36	4.31	4.25	4.20
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.13	4.10	4.08	4.03	4.00	3.89	3.84	3.78	3.73
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.80	3.77	3.74	3.70	3.67	3.56	3.51	3.45	3.39
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.55	3.52	3.50	3.45	3.42	3.31	3.26	3.20	3.14
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.36	3.33	3.30	3.26	3.23	3.12	3.06	3.00	2.94
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.21	3.18	3.15	3.11	3.07	2.96	2.91	2.85	2.79
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.08	3.05	3.03	2.98	2.95	2.84	2.78	2.72	2.66
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.98	2.95	2.92	2.88	2.84	2.73	2.67	2.61	2.55
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.89	2.86	2.84	2.79	2.76	2.64	2.59	2.52	2.46
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.82	2.79	2.76	2.72	2.68	2.57	2.51	2.45	2.38
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.75	2.72	2.70	2.65	2.62	2.50	2.44	2.38	2.32
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.70	2.67	2.64	2.60	2.56	2.44	2.38	2.32	2.26
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.65	2.62	2.59	2.55	2.51	2.39	2.33	2.27	2.20
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.60	2.57	2.55	2.50	2.46	2.35	2.29	2.22	2.16
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.56	2.53	2.51	2.46	2.42	2.31	2.25	2.18	2.11
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.53	2.50	2.47	2.43	2.39	2.27	2.21	2.14	2.08
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.50	2.47	2.44	2.39	2.36	2.24	2.18	2.11	2.04
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.47	2.44	2.41	2.36	2.33	2.21	2.15	2.08	2.01
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.44	2.41	2.38	2.34	2.30	2.18	2.12	2.05	1.98
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.42	2.39	2.36	2.31	2.28	2.16	2.09	2.03	1.95
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.39	2.36	2.34	2.29	2.25	2.13	2.07	2.00	1.93
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.37	2.34	2.32	2.27	2.23	2.11	2.05	1.98	1.91
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.36	2.32	2.30	2.25	2.21	2.09	2.03	1.96	1.89
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.34	2.31	2.28	2.23	2.20	2.07	2.01	1.94	1.87
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.21	2.18	2.15	2.11	2.07	1.94	1.88	1.80	1.72
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.09	2.06	2.03	1.98	1.94	1.82	1.74	1.67	1.58
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.98	1.94	1.92	1.87	1.82	1.69	1.61	1.53	1.43

Tabela 6: Quantis da Distribuição F para probabilidade $p = P[F \geq F_1] = 0,025$. Graus de liberdade do numerador dado no topo e do denominador na margem esquerda.

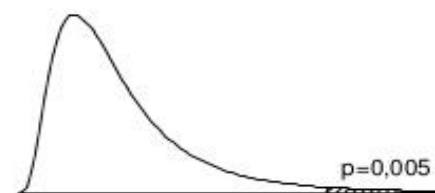
Distribuição F de Snedecor a 1% ($p=0.01$)



	1	2	3	4	5	6	7	8	9	10	12	14	15	16	18	20	30	40	60	120
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.43	99.44	99.44	99.45	99.47	99.47	99.48	99.49
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.92	26.87	26.83	26.75	26.69	26.50	26.41	26.32	26.22
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.25	14.20	14.15	14.08	14.02	13.84	13.75	13.65	13.56
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.77	9.72	9.68	9.61	9.55	9.38	9.29	9.20	9.11
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.60	7.56	7.52	7.45	7.40	7.23	7.14	7.06	6.97
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.36	6.31	6.28	6.21	6.16	5.99	5.91	5.82	5.74
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.56	5.52	5.48	5.41	5.36	5.20	5.12	5.03	4.95
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	5.01	4.96	4.92	4.86	4.81	4.65	4.57	4.48	4.40
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.60	4.56	4.52	4.46	4.41	4.25	4.17	4.08	4.00
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.29	4.25	4.21	4.15	4.10	3.94	3.86	3.78	3.69
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.05	4.01	3.97	3.91	3.86	3.70	3.62	3.54	3.45
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.86	3.82	3.78	3.72	3.66	3.51	3.43	3.34	3.25
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.70	3.66	3.62	3.56	3.51	3.35	3.27	3.18	3.09
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.56	3.52	3.49	3.42	3.37	3.21	3.13	3.05	2.96
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.45	3.41	3.37	3.31	3.26	3.10	3.02	2.93	2.84
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.35	3.31	3.27	3.21	3.16	3.00	2.92	2.83	2.75
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.27	3.23	3.19	3.13	3.08	2.92	2.84	2.75	2.66
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.19	3.15	3.12	3.05	3.00	2.84	2.76	2.67	2.58
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.13	3.09	3.05	2.99	2.94	2.78	2.69	2.61	2.52
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.07	3.03	2.99	2.93	2.88	2.72	2.64	2.55	2.46
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	3.02	2.98	2.94	2.88	2.83	2.67	2.58	2.50	2.40
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.97	2.93	2.89	2.83	2.78	2.62	2.54	2.45	2.35
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.93	2.89	2.85	2.79	2.74	2.58	2.49	2.40	2.31
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.89	2.85	2.81	2.75	2.70	2.54	2.45	2.36	2.27
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.86	2.81	2.78	2.72	2.66	2.50	2.42	2.33	2.23
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.82	2.78	2.75	2.68	2.63	2.47	2.38	2.29	2.20
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.79	2.75	2.72	2.65	2.60	2.44	2.35	2.26	2.17
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.77	2.73	2.69	2.63	2.57	2.41	2.33	2.23	2.14
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.74	2.70	2.66	2.60	2.55	2.39	2.30	2.21	2.11
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.56	2.52	2.48	2.42	2.37	2.20	2.11	2.02	1.92
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.39	2.35	2.31	2.25	2.20	2.03	1.94	1.84	1.73
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.23	2.19	2.15	2.09	2.03	1.86	1.76	1.66	1.53

Tabela 7: Quantis da Distribuição F para probabilidade $p = P[F \geq F_i] = 0,01$. Graus de liberdade do numerador dado no topo e do denominador na margem esquerda.

Distribuição F de Snedecor a 0,5% ($p=0.005$)



	1	2	3	4	5	6	7	8	9	10	12	14	15	16	18	20	30	40	60	120
2	198.50	199.00	199.17	199.25	199.30	199.33	199.36	199.37	199.39	199.40	199.42	199.43	199.43	199.44	199.44	199.45	199.47	199.47	199.48	199.49
3	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88	43.69	43.39	43.17	43.08	43.01	42.88	42.78	42.47	42.31	42.15	41.99
4	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14	20.97	20.70	20.51	20.44	20.37	20.26	20.17	19.89	19.75	19.61	19.47
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.38	13.21	13.15	13.09	12.98	12.90	12.66	12.53	12.40	12.27
6	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	10.03	9.88	9.81	9.76	9.66	9.59	9.36	9.24	9.12	9.00
7	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51	8.38	8.18	8.03	7.97	7.91	7.83	7.75	7.53	7.42	7.31	7.19
8	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.01	6.87	6.81	6.76	6.68	6.61	6.40	6.29	6.18	6.06
9	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.23	6.09	6.03	5.98	5.90	5.83	5.62	5.52	5.41	5.30
10	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.66	5.53	5.47	5.42	5.34	5.27	5.07	4.97	4.86	4.75
11	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54	5.42	5.24	5.10	5.05	5.00	4.92	4.86	4.65	4.55	4.45	4.34
12	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.91	4.77	4.72	4.67	4.59	4.53	4.33	4.23	4.12	4.01
13	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94	4.82	4.64	4.51	4.46	4.41	4.33	4.27	4.07	3.97	3.87	3.76
14	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72	4.60	4.43	4.30	4.25	4.20	4.12	4.06	3.86	3.76	3.66	3.55
15	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.25	4.12	4.07	4.02	3.95	3.88	3.69	3.58	3.48	3.37
16	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38	4.27	4.10	3.97	3.92	3.87	3.80	3.73	3.54	3.44	3.33	3.22
17	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25	4.14	3.97	3.84	3.79	3.75	3.67	3.61	3.41	3.31	3.21	3.10
18	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14	4.03	3.86	3.73	3.68	3.64	3.56	3.50	3.30	3.20	3.10	2.99
19	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04	3.93	3.76	3.64	3.59	3.54	3.46	3.40	3.21	3.11	3.00	2.89
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.68	3.55	3.50	3.46	3.38	3.32	3.12	3.02	2.92	2.81
21	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88	3.77	3.60	3.48	3.43	3.38	3.31	3.24	3.05	2.95	2.84	2.73
22	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81	3.70	3.54	3.41	3.36	3.31	3.24	3.18	2.98	2.88	2.77	2.66
23	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75	3.64	3.47	3.35	3.30	3.25	3.18	3.12	2.92	2.82	2.71	2.60
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69	3.59	3.42	3.30	3.25	3.20	3.12	3.06	2.87	2.77	2.66	2.55
25	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64	3.54	3.37	3.25	3.20	3.15	3.08	3.01	2.82	2.72	2.61	2.50
26	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60	3.49	3.33	3.20	3.15	3.11	3.03	2.97	2.77	2.67	2.56	2.45
27	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56	3.45	3.28	3.16	3.11	3.07	2.99	2.93	2.73	2.63	2.52	2.41
28	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52	3.41	3.25	3.12	3.07	3.03	2.95	2.89	2.69	2.59	2.48	2.37
29	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48	3.38	3.21	3.09	3.04	2.99	2.92	2.86	2.66	2.56	2.45	2.33
30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.18	3.06	3.01	2.96	2.89	2.82	2.63	2.52	2.42	2.30
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	2.95	2.83	2.78	2.74	2.66	2.60	2.40	2.30	2.18	2.06
60	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.74	2.62	2.57	2.53	2.45	2.39	2.19	2.08	1.96	1.83
120	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.54	2.42	2.37	2.33	2.25	2.19	1.98	1.87	1.75	1.61

Tabela 8: Quantis da Distribuição F para probabilidade $p = P[F \geq F_1] = 0,005$. Graus de liberdade do numerador dado no topo e do denominador na margem esquerda.