Estatística Aplicada à Administração

Prof. Marcelo Tavares

ORGANIZAÇÃO DO CONTEÚDO
Prof. Marcelo Tavares

PROJETO GRÁFICO
Annye Cristiny Tessaro
Mariana Lorenzetti

DIAGRAMAÇÃO
Annye Cristiny Tessaro
Victor Emmanuel Carlson

REVISÃO DE PORTUGUÊS
Renato Tapado
Patrícia Regina da Costa

Sumário

Introdução

O cidadão comum pensa que a estatística se resume apenas a apresentar tabelas de números em colunas esportivas e/ou econômicas de jornais e revistas, ilustradas com gráficos, pilhas de moedas, etc., ou quando muito associam a estatística à previsão de resultados eleitorais. A estatística não se limita somente a compilar tabelas de dados e os ilustrar graficamente. Sir Ronald Fisher (1890-1962), em seus trabalhos, iniciou a estatística como método científico. Desta forma, o trabalho do estatístico passou a ser o de ajudar a planejar a obtenção de dados, interpretar e analisar os dados obtidos e apresentar os resultados de maneira a facilitar a tomada de decisões razoáveis.

Didaticamente, podemos dividir a estatística em duas partes: a **estatística descritiva** e a **inferência estatística**.

A estatística descritiva preocupa-se com a forma pela qual podemos apresentar um conjunto de dados em tabelas e gráficos, e também resumir as informações contidas nestes dados mediante a utilização de medidas estatísticas.

Para saber mais vá ao site www.ibge.gov.br

Já a inferência estatística baseia-se na teoria das probabilidades para estabelecer conclusões sobre todo um grupo (chamado população), quando se observou apenas uma parte (amostra) representativa desta população.

Uma grande quantidade de informações importantes que auxiliam na tomada de decisões está no site do Instituto Brasileiro de Geografia e Estatística (IBGE). Todo o cálculo das estatísticas pode ser feito por meio de calculadoras científicas e também softwares estatísticos. As planilhas eletrônicas permitem o cálculo de diversas estatísticas e confecção de gráficos de forma mais rápida e eficiente.

Uma sugestão de referência é NEUFELD, John L. *Estatística Aplicada à Administração usando Excel*. v.1. São Paulo: Prantice Hall, 2003

É necessário ter em mente que a estatística é uma ferramenta para o gestor ou executivo, nas respostas dos "porquês" de seus problemas que podem ser explicados por uma análise de dados. Para ela ser bem usada, é necessário conhecer os seus fundamentos e princípios, e acima de tudo que o gestor ou executivo desenvolva um espírito crítico e jamais deixe de pensar. Pois é fácil mentir usando a estatística, o difícil é falar a verdade sem usar a estatística.

Atualmente, as empresas têm procurado profissionais como executivos que tenham um nível de conhecimento de estatística alto, pois este conhecimento tem feito uma diferença grande nos processos decisórios em empresas.

Este livro será dividido em quatro Unidades:

> Unidade 1 – Estatística Descritiva (Descrição de amostras por meio de distribuições de freqüências, e medidas de posição e dispersão);
>
> Unidade 2 – Probabilidades (Conceitos básicos de probabilidades, variáveis aleatórias uni e bidimensionais) e Distribuições de Probabilidades (discretas e contínuas);
>
> Unidade 3 – Amostragem (probabilística e não probabilística), Distribuições Amostrais (Distribuições t de Student, qui-quadrado e F) e Intervalos de Confiança (média, proporção); e
>
> Unidade 4 – Processos Decisórios (Testes de Hipóteses).

UNIDADE

1

Estatística descritiva

Objetivo

Esta Unidade tem por objetivo fazer com que você tenha condições de descrever e apresentar os resultados de um conjunto de observações de forma clara, objetiva e passando o máximo de informações possíveis. Para tal objetivo, serão abordadas as distribuições de freqüências, análises gráficas, medidas de posição e dispersão.

Estatística descritiva

Qualquer conjunto de dados, tais como o tempo de uma ligação telefônica, a velocidade de processamento de um computador, a proporção de participação no mercado das empresas de um determinado setor, suscetibilidade de empresas a uma determinada mudança no mercado, opinião dos alunos quanto à didática de um professor, etc., contém informação sobre algum grupo de indivíduos. As possíveis diferenças entre indivíduos determinam a variação que está sempre presente na análise de dados.

Uma característica que pode assumir diferentes valores de indivíduo para indivíduo é denominada **variável**, pois de outra forma seria denominada **constante**.

A classificação das variáveis em qualitativas e quantitativas foi apresentada na disciplina de Metodologia de Pesquisa. **Caso não se lembre, reveja o material de Metodologia de Pesquisa.**

Desta forma, apenas para relembrar, como você faria a classificação das seguintes variáveis?

a) **Número de páginas desta unidade;**

b) **peso dos funcionários do setor de marketing de uma empresa;**

c) **tipos de empresas em relação a adoção de determinada técnica; e**

d) **tamanho de empresas (pequena, média e grande).**

Respostas: a) quantitativa discreta; b) quantitativa contínua; c) qualitativa nominal; d) qualitativa ordinal.

Os dados qualitativas são divididos em nominais e ordinais; enquanto os dados quantitativas são divididos em discretas e contínuas.

GLOSSÁRIO

*Dados Brutos:– dados na forma em que foram coletados, sem nenhum tratamento. Fonte: Lacombe (2004)

Quando você coleta os dados para uma pesquisa, estas observações são chamadas de dados brutos. Um exemplo de dados brutos corresponde ao tempo em minutos que consumidores de uma determinada operadora de telefonia celular utilizariam em um mês (dados simulados pelo autor a partir de um caso real). Os dados foram obtidos em uma pesquisa de mercado e apresentados na forma em que foram coletados (Tabela 1), por este motivo são denominados **dados brutos***.

Geralmente, este tipo de dado traz pouca ou nenhuma informação ao leitor, sendo necessário organizar os dados, com o intuito de aumentar sua capacidade de informação.

C	T	C	T	C	T	C	T	C	T
1	**104**	9	**122**	17	**129**	25	**144**	33	**183**
2	**108**	10	**142**	18	**138**	26	**151**	34	**138**
3	**138**	11	**106**	19	**122**	27	**146**	35	**115**
4	**101**	12	**201**	20	**161**	28	**82**	36	**179**
5	**163**	13	**169**	21	**167**	29	**137**	37	**142**
6	**141**	14	**120**	22	**189**	30	**132**	38	**111**
7	**90**	15	**210**	23	**132**	31	**172**	39	**140**
8	**154**	16	**98**	24	**127**	32	**87**	40	**136**

Tabela 1: Tempo (T) em minutos de uso de telefone celular por consumidores (C) de uma determinada operadora

GLOSSÁRIO

*Rol – é a mais simples organização numérica. É a ordenação dos dados em ordem crescente ou decrescente.

*Amplitude Total – corresponde à diferença entre o maior e o menor valor observado em um conjunto de dados. Notaremos por **A.**

Como você pode observar na Tabela 1, a simples organização dos dados em um **rol*** aumenta muito a capacidade de informação destes. Na Tabela 2, você pode verificar que o menor tempo observado foi 82 minutos, e o maior, 210 minutos, o que nos fornece uma **amplitude total*** de variação da ordem de 128 minutos.

Outra informação que podemos obter nos dados por meio da Tabela 2 (organizada em rol crescente) é que alguns tempos, como 122 min, 132 min, 138 min e 142 min, foram os mais freqüentes, ou seja, os mais citados na pesquisa.

82	111	132	142	167
87	115	136	142	169
90	120	137	144	172
98	122	138	146	179
101	122	138	151	183
104	127	138	154	189
106	129	140	161	201
108	132	141	163	210

Tabela 2: Tempo em minutos de uso de telefone celular por consumidores de uma determinada operadora (dados em rol crescente).

Então surge uma pergunta:

Como você pode organizar os dados de uma forma mais eficiente, na qual se possa apresentar uma quantidade maior de informações?

Uma maneira de organizar um conjunto de dados para você melhor representá-lo é por meio de uma **tabela de distribuição de freqüências** (tabela onde são apresentadas as freqüências de cada uma das classes).

Distribuindo-se os dados observados em **classes*** e contando-se o número de observações contidas em cada classe, obtém-se a **freqüência de classe**. Veja que a disposição tabular dos dados agrupados em classes, juntamente com as freqüências correspondentes, se denomina distribuição de freqüências.

Por exemplo, para o caso do tempo em minutos do uso de celulares, pode-se desejar incluir em uma única classe todos os indivíduos que possuam tempo entre 128 e 138 minutos assim, a classe irá variar de 128 a 138 minutos.

GLOSSÁRIO

*Classes: – Intervalos nos quais os valores da variável analisada são agrupados.

Para identificar uma classe, deve-se conhecer os valores dos **limites inferior e superior da classe**, que delimitam o **intervalo de classe**.

Neste ponto, surge uma dúvida.

Indivíduos que apresentem tempo exatamente iguais a 128 ou a 138 minutos pertencem ou não a esta classe? (128 a 138)

Deste modo, surge a necessidade de definir a natureza do intervalo de classe, se é **aberto** ou **fechado**. Portanto, podemos ter exemplo de notação dos diferentes tipos de intervalos: **Intervalos abertos*** ➔ 128 min – 138 min; **Intervalos fechados*** ➔ 128 min |–| 138 min. Pode-se ter ainda **intervalos mistos***, como por exemplo: 128 min |– 138 min.

Este tipo de intervalo é o mais utilizado.

GLOSSÁRIO

*Intervalos abertos– os limites da classe (inferior e superior) não pertencem a ela.

*Intervalos fechados – os limites de classe (superior e inferior) pertencem à classe em questão.

*Intervalos mistos – um dos limites pertence à classe, e o outro, não.

Vamos, então, a partir dos dados do exemplo relativo ao tempo de utilização dos celulares, construir uma distribuição de freqüência e ao longo deste exercício identificar conceitos presentes em uma distribuição de freqüências.

Então, vamos exercitar.

Para elaborar uma distribuição de freqüências é necessário que primeiramente, se determine o **número de classes** (k) em que os dados serão agrupados. Por questões de ordem prática e estética, sugere-se utilizar de 5 a 20 classes. O número de classes (k) a ser utilizado, pode ser calculado em função do número de observações (n).

Na pesquisa, como temos n = 40 consumidores, teremos, então, o número de classes definido por $k = \sqrt{n} = \sqrt{40} = 6{,}32$, e como o número de classes é inteiro, usaremos 6 classes. O arredondamento utilizado neste material é o padrão de algarismos significativos (como foi aprendido no Ensino Médio). O número de classes pode também ser definido de uma forma arbitrária sem o uso desta regra.

$k = \sqrt{n}$, para $n \leq 100$; $k = 5\log n$, para $n > 100$

Após você determinar o número de classes (k) em que os dados serão agrupados, deve-se, então, determinar a **amplitude do intervalo de classe (c)**.

$c = \dfrac{A}{k-1}$ onde: c = amplitude de classe; A = amplitude total; e k = número de classes.

Para calcularmos a amplitude do intervalo de classe, vamos primeiramente calcular a **amplitude total dos dados (A), que corresponde à diferença entre o maior valor observado e o menor valor observado.** No nosso caso, teremos A = 210 – 82 =128 mm.

Com base neste valor da amplitude total (A) calculado, vamos obter a amplitude do intervalo de classe (c), como é mostrado a seguir:

$$c = \frac{128}{6-1} = 25{,}6 \ \text{min}$$

Deve ficar claro para você, que existem outros procedimentos para determinação da amplitude do intervalo de classe que podem ser encontrados na literatura.

Conhecida a amplitude de classes, você deve determinar os intervalos de classe. O limite inferior e o superior das classes devem ser escolhidos de modo que o menor valor observado esteja localizado no **ponto médio (PM)** da primeira classe.

$PM = \dfrac{LI+LS}{2}$ onde: LI: Limite inferior; e LS: Limite superior

Partindo deste raciocínio, então, o limite inferior da primeira classe será:

$$\text{Limite inf. } 1^{a} = \text{menor valor} - \frac{c}{2}.$$

No nosso caso, temos: $\text{Limite inf. } 1^{a} = 82 - \dfrac{25{,}6}{2} = 69{,}2 \ \text{min}$

Definindo, então, o limite inferior da primeira classe, para obtermos as classes da nossa distribuição, basta que somemos a amplitude do intervalo de classe a cada limite inferior.

Assim, teremos:

69,2 |– 94,8 ➔ primeira classe

94,8 |– 120,4 ➔ segunda classe

20,4 |– 146,0 ➔ terceira classe

146,0 |– 171,6 ➔ quarta classe

171,6 |– 197,2 ➔ quinta classe

197,2 |– 222,8 ➔ sexta classe

Então, você pode obter uma tabela como a apresentada a seguir.

*A **freqüência absoluta** (fa) corresponde ao numero de observações que temos em uma determinada classe ou em um determinado atributo de uma variável qualitativa e a **freqüência relativa** (fr) corresponde à proporção do número de observações em uma determinada classe em relação ao total de observações que temos.*

Esta freqüência pode ser expressa em termos porcentuais. Para isto, basta multiplicar a freqüência relativa obtida por 100.

Classes (mm)	Freqüência
69,2 |– 94,8	?
94,8 |– 120,4	?
120,4 |– 146,0	?
146,0 |– 171,6	?
171,6 |– 197,2	?
197,2 |– 222,8	?
Total	

Tabela 3: Distribuição de freqüências do tempo em minutos de uso de telefone celular por consumidores de uma determinada operadora

Na tabela, aparece uma nova denominação chamada "freqüência". Podem ter freqüências chamadas **de freqüência absoluta** (*fa*), **freqüência relativa** (*fr*) e freqüência acumulada (discutida posteriormente).

O cálculo da freqüência relativa é obtido por meio da seguinte expressão:

$$fr_i = \frac{fa_i}{\sum_i fa_i}, \text{ com } fa_i = \text{freqüência absoluta da classe i.}$$

Apresentando os dados na forma de distribuição de freqüência, você consegue sintetizar as informações contidas neles, além de facilitar sua visualização. Na Tabela apresentada a seguir, temos as freqüências (*fa* e *fr*) relacionadas ao tempo de utilização do aparelho celular.

Classes (mm)	*fa* (consumidores)	*fr* (proporção de consumidores)
69,2 |– 94,8	3	0,075
94,8 |– 120,4	8	0,200
120,4 |– 146,0	16	0,400
146,0 |– 171,6	7	0,175
171,6 |– 197,2	4	0,100
197,2 |– 222,8	2	0,050
Total	40	1,000

Tabela 4: Distribuição de freqüências do tempo em minutos de uso de telefone celular por consumidores de uma determinada operadora

Como ficaria, então, a interpretação da distribuição de freqüências?

Pode-se verificar claramente na Tabela 4 que os tempos de utilização do celular das 40 pessoas avaliadas em questão estão concentrados nas classes segunda, terceira e quarta, decrescendo em direção às classes do início e do fim da tabela. A apresentação dos dados em forma de distribuição de freqüência facilita ainda o cálculo manual de várias medidas estatísticas de interesse e sua apresentação gráfica.

Além das freqüências absolutas e relativas, muitas vezes pode-se estar interessado na quantidade de observações que existe acima ou abaixo de um determinado ponto na distribuição.

Desta forma, podemos trabalhar com a **freqüência acumulada***.

A título de ilustração, você pode visualizar nas Tabelas 4 e 5, respectivamente, as freqüências acumuladas para cima e para baixo dos tempos de utilização das 40 pessoas avaliadas na pesquisa. A freqüência acumulada apresentada na Tabela 4 pode ser obtida da seguinte forma: abaixo do limite superior da primeira classe, temos três pessoas que estão presentes nesta classe, como pode ser visto na Tabela 3 da distribuição de freqüências absoluta. Quando consideramos a segunda classe, a freqüência acumulada corresponde ao número de

GLOSSÁRIO

*Freqüência Acumulada – Freqüência Acumulada corresponde à soma da freqüência daquela classe às freqüências de todas as classes abaixo dela.

pessoas que temos abaixo do limite superior desta classe, ou seja, as oito pessoas da segunda classe mais as três pessoas da primeira classe, totalizando 11 pessoas abaixo de 120,4 minutos. Para as outras classes, o raciocínio é semelhante.

Tempo (mim)	Freq. acumulada	Freq. acumulada (relativa)
69,2 |– 94,8	0	0,000
94,8 |– 120,4	3	0,075
120,4 |– 146,0	11	0,275
146,0 |– 171,6	27	0,675
171,6 |– 197,2	34	0,850
197,2 |– 222,8	38	0,950
Total	40	1,000

Tabela 5: Distribuição de freqüência acumulada do tempo em minutos de uso de telefone celular por consumidores de uma determinada operadora

Um bom exemplo de aplicação das distribuições de freqüências acumuladas corresponde à identificação de uma determinada freqüência abaixo ou acima de um determinado valor que não corresponde ao limite superior ou inferior de uma classe qualquer.

Podemos, então, querer verificar qual a porcentagem de pessoas que utilizam o celular por um tempo inferior a 146 minutos, e para isto, basta consultar diretamente a Tabela 4 e verificar a freqüência acumulada abaixo deste valor (6,75%), pois o valor 146 minutos corresponde a um dos limites de classe apresentados nesta tabela.

E se você quiser saber a proporção de pessoas que utilizam o celular por menos de 150 minutos?

Para podermos obter esta freqüência, é necessário que venhamos a pressupor que os tempos de utilização estejam uniformemente distribuídos dentro das classes. O cálculo é baseado nos dados da Tabela 4 e são apresentados a seguir.

Freq. acumulada relativa abaixo da classe imediatamente inferior a 150 (abaixo de 146) = 0,675; e

Freq. acumulada relativa abaixo da classe imediatamente superior a 150 (abaixo de 171,6) = 0,850;

Proporção de consumidores com tempo de uso abaixo de 146,0 min = 0,675

Proporção de consumidores com tempo de uso abaixo de 171,6 min = 0,850

Freq. entre 146,0 e 171,6 min = 0,175 **0,850 – 0,675**

de 146,0 a 171,6 min são 25,6 min

de 146,0 a 150,0 min são 4,0 min

assim,

(diferença) no tempo variação (diferença) na proporção

25,6 min --- 0,175

4,0 min -- x

Portanto, fazendo o cálculo da regra de três apresentada anteriormente, teremos o valor de x.

$$x = \frac{4 \cdot 0{,}175}{25{,}6} = 0{,}0273$$

Portanto, como abaixo de 140,0 min existe uma proporção de 0,675 e entre 140,0 e 150 min existe uma proporção de 0,0273, conclui-se, então, que abaixo de 150 min existe uma proporção de 0,7023 (0,675 + 0,0273). Em termos porcentuais, isto corresponde a 70,23% dos consumidores.

É importante ressaltar que este resultado é aproximado, devido à perda de informação pelo fato de a tabela ser intervalar, ou seja, as classes estão em intervalos.

Quando você trabalha com variáveis qualitativas, os atributos são as variações nominativas da variável. A construção da tabela consiste em contar as ocorrências de cada atributo. O resultado da contagem define a freqüência absoluta do atributo. Para podermos entender isto, tomemos como exemplo uma pesquisa na qual se procurou avaliar o número de pessoas de diferentes sexos em uma determinada empresa. Estes resultados são apresentados na Tabela 6.

Sexo	*fa*	*fr*
Masculino	20	0,40
Feminino	30	0,60
Total	50	1,00

Tabela 6: Distribuição de freqüências do número de funcionários em relação ao seu sexo em 2006

Tomando, por exemplo, o caso de uma **variável aleatória discreta (conceito visto em Metodologia de Pesquisa)**, realizou-se no SAC (Serviço de Atendimento ao Consumidor) de uma empresa um estudo referente ao número de reclamações (N.R.) atendidas diariamente, durante um certo mês, obtendo os seguintes resultados:

	N.R.	Dia	N.R.	Dia	N.R.	Dia	N.R.	Dia	N.R.
	0	7	1	13	0	19	1	25	0
	2	8	2	14	0	20	0	26	3
	1	9	2	15	1	21	0	27	4
	5	10	3	16	2	22	2	28	0
	3	11	0	17	3	23	0	29	2
	2	12	3	18	5	24	4	30	1

Dispondo estes dados em um rol (crescente), tem-se:

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

Podemos, então, apresentar a seguir estes dados em uma distribuição de freqüências. Neste caso, não é necessário definir intervalos de classes, porque a variação dos valores é pequena.

Numero de reclamações por dia	Número de dias (fa)	Freq. Relativa
0	9	0.3
1	5	0.17
2	7	0.23
3	5	0.17
4	2	0.07
5	2	0.07
Total	30	1

Tabela 7: Número de reclamações atendidas diariamente, durante certo mês

Verificamos que os valores da variável discreta correspondem a cada uma das classes.

Surge, então, uma pergunta:

As tabelas de distribuição de freqüências são a única forma que você tem de apresentar um conjunto de dados?

Para responder a esta pergunta, vamos falar um pouco sobre algumas formas de representação gráfica de tabelas de freqüência. Logicamente, dependendo do tipo de variável, temos um gráfico mais adequado. Os diferentes tipos de gráfico, (histogramas, polígonos de freqüência, ogivas, gráficos de setores, pictogramas e outros) permitem uma melhor visualização de resultados. Estes gráficos podem ser obtidos utilizando planilhas eletrônicas, como por exemplo, o Excel.

GLOSSÁRIO

*histograma – Histogramas: são constituídos por um conjunto de retângulos, com as bases assentadas sobre um eixo horizontal, tendo o centro da mesma no ponto médio da classe que representa, e cuja altura é proporcional à freqüência da classe.

*Polígono de freqüências – é um gráfico de análise no qual as freqüências das classes são localizadas sobre perpendiculares levantadas nos pontos médios das classes.

Os **histogramas*** são gráficos utilizados para representar tabelas intervalares.

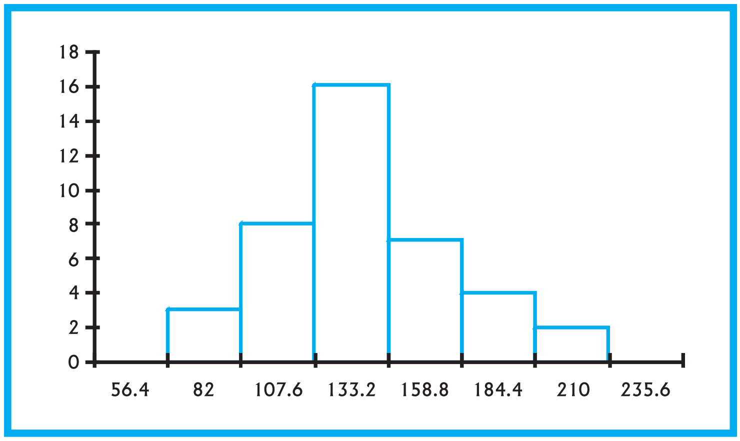


Figura 1: Histograma representativo da distribuição de freqüências do tempo em minutos de uso de telefone celular por consumidores de uma determinada operadora

Já o **polígono de freqüência***, você pode obter pela simples união dos pontos médios dos topos dos retângulos de um histograma. Completa-se o polígono unindo as extremidades da linha que ligam os pontos representativos das freqüências de classe aos pontos médios das classes, imediatamente, anterior e posterior às classes extremas, que têm freqüência nula.

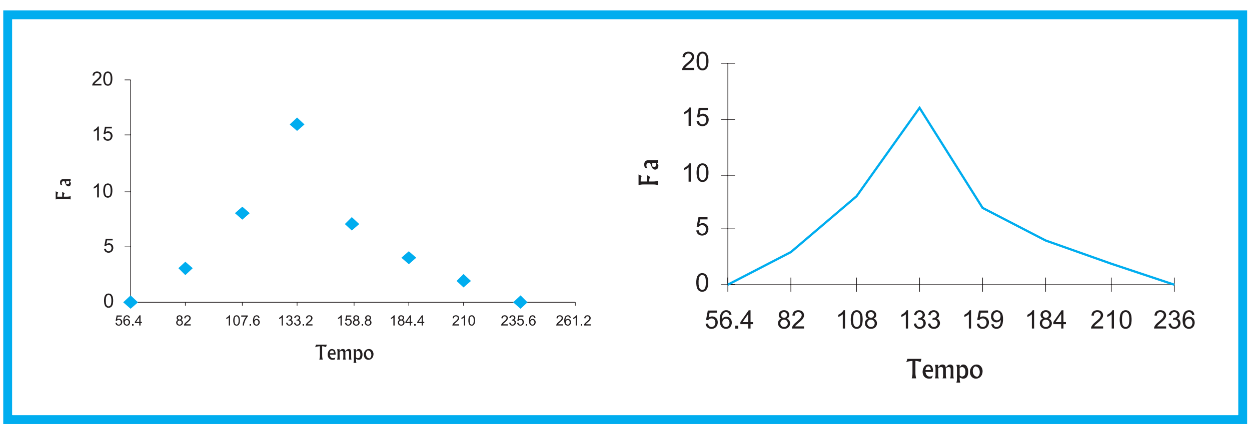


Figura 2: Polígono de Freqüências do tempo em minutos de uso de telefone celular por consumidores de uma determinada operadora

Quando temos uma tabela de variável qualitativa, um tipo de gráfico adequado para apresentar os resultados corresponde ao gráfico de setores, também popularmente conhecido como gráfico tipo pizza. Sua construção é simples: sabe-se que o ângulo de 360º equivale a 100% da área da circunferência; assim, para obter-se o ângulo do setor cuja área representa uma determinada freqüência, basta resolver uma regra de três simples, como a apresentada a seguir:

360º ---------------- 100%
xº ------------------- Freq. Relativa (Porcentual)

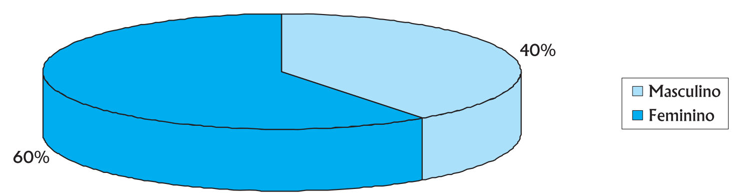


Figura 3: Gráfico do sexo de pessoas que trabalham em uma determinada empresa

Os gráficos chamados de **ogivas** correspondem a um polígono de freqüências acumuladas, nas quais estas freqüências são localizadas sobre perpendiculares levantadas nos limites inferiores ou superiores das classes, dependendo se a ogiva representar as freqüências acumuladas.

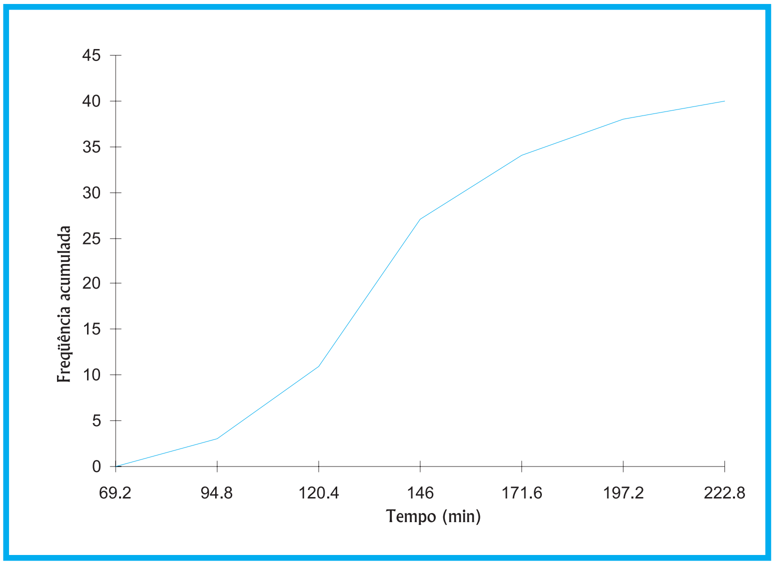


Figura 4: Ogiva "abaixo de" do tempo em minutos que consumidores de uma determinada operadora de telefonia celular utilizariam em um mês se houvesse uma redução na tarifa de 20%

Após o estudo da construção de distribuições de freqüências e gráficos que as representam, você deve ser capaz de organizar um conjunto de dados, por meio de uma distribuição de freqüências (absoluta, relativa, e acumuladas) e representá-lo graficamente.

Saiba mais...

- Visite o site de como usar a planilha Calç, do pacote OpenOffice, nas estatísticas descritivas, em: http://www2.ufpa.br/dicas/open/oo-ind.htm

Vamos, então, fazer um exercício para fixar os conhecimentos adquiridos. (As respostas estão no final do livro.)

Exercício 1: tem-se a seguir o tempo em minutos de reuniões em um setor de uma empresa.

45	51	50	58
50	44	46	57
42	41	60	58
41	50	54	60
52	46	52	51

a) Construa a distribuição de freqüências absoluta, relativa e acumulada; e

b) Determine o número de reuniões em que o tempo foi menor do que 50, a partir da distribuição de freqüências.

Medidas de posição

As medidas de posição ou de tendência central constituem uma forma mais sintética de apresentar os resultados contidos nos dados observados, pois representam um valor central, em torno do qual os dados se concentram. As medidas de tendência central mais empregadas são a média, a mediana e a moda.

A **média aritmética*** é a mais usada das três medidas de posição mencionadas, por ser a mais comum e compreensível delas, bem como pela relativa simplicidade do seu cálculo, além de prestar-se bem ao tratamento algébrico.

Considerando o caso do número de reclamações em um SAC (ver em distribuições de freqüência), se você somar todos os valores do número de reclamações e dividir pelo número de dias, você terá então a **média aritmética** ($\bar{x}$) do número de reclamações.

Então, o valor obtido será: $\bar{x} = 1{,}73$ *reclamações por dia*.

GLOSSÁRIO

*A **média aritmética**, ou simplesmente média de um conjunto de n observações, x_1, x_2,...,x_n é definida como:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

O somatório (Σ) corresponde à soma de todos os valores obtidos.

Você pode interpretar este resultado da média como sendo o número de reclamações médio por dia, que é de 1,73, podendo ocorrer dias em que o número de reclamações pode ser maior ou menor que o valor médio encontrado.

Portanto, de uma forma mais geral, podemos interpretar a média como sendo um valor típico do conjunto de dados. Pode ser um valor que não pertence ao conjunto de dados.

GLOSSÁRIO

*Hipótese Tabular Básica– todas as observações contidas numa classe são consideradas iguais ao ponto médio da classe.

Se os dados estiverem agrupados na forma de uma distribuição de freqüência em classes, lança-se mão da **Hipótese Tabular Básica*** para o cálculo da média.

Então, você vai calcular a média por meio da seguinte expressão:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i fa_i}{\sum_{i=1}^{n} fa_i} = \sum_{i=1}^{n} x_i fr_i$$

sendo x_i: ponto médio da classe i; fa_i: freqüência absoluta da classe i; fr_i : freqüência relativa da classe i.

Considerando o caso do tempo de uso em minutos do celular (ver no item distribuições de freqüências), a média será dada por:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i fa_i}{\sum_{i=1}^{n} fa_i} = \frac{(82.3) + (107{,}6\ .\ 8) + ... + (210\ .\ 2)}{14 + 46 + ... + 6} = 137{,}68 \text{ minutos}$$

ou

$$\bar{x} = \sum_i x_i fr_i = (82\ .\ 0{,}075) + (107{,}6\ .\ 0{,}2) + ... + (210\ .\ 0{,}05) = 137{,}68 \text{ minutos}$$

O valor de 82 apresentado na expressão corresponde ao ponto médio da primeira classe, o qual foi obtido pela soma dos limites superior e inferior, dividido por dois, ou seja, a média aritmética. Os pontos médios das outras classes são obtidos de forma similar.

Existem outros tipos de média que podem ser utilizados, como por exemplo, média ponderada (utilizada quando existe algum fator de ponderação), média geométrica (quando os dados apresentam uma distribuição que não é simétrica) e outras.

Às vezes, associam-se às observações $x_1, x_2, ..., x_n$ determinadas ponderações ou pesos $w_1, w_2, ..., w_n$, que dependem da importância atribuída a cada uma das observações; neste caso, a média ponderada é dada por:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i \; w_i}{\sum_{i=1}^{n} w_i}$$

Como exemplo, você pode considerar um processo de avaliação de um funcionário em três etapas. Um funcionário apresentou as seguintes notas durante a avaliação: 1ª etapa = 90; 2ª etapa = 70; 3ª etapa = 85, e os pesos de cada etapa são 1, 1 e 3, respectivamente. Qual o score médio final do funcionário?

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i \; w_i}{\sum_{i=1}^{n} w_i} = \frac{(1 \times 70) + (1 \times 90) + (3 \times 85)}{1 + 1 + 3} = \frac{415}{5} = 83$$

Outro tipo de média corresponde à geométrica (Mg). Ela é calculada pela raiz n-ésima do produto de um conjunto de n observações, $x_1, x_2, ..., x_n$, associadas às freqüências absolutas $f_1, f_2, ..., f_n$, e, respectivamente, é dada por:

$$Mg = \sqrt[n]{x_1^{f_1} \times x_2^{f_2} \times \ldots \times x_n^{f_n}}$$

Este tipo de média, você vai trabalhar na disciplina de Matemática Financeira.

Em algumas situações, você verá que é necessária a informação do número de observações que mais ocorre em um conjunto de dados.

No caso do número de reclamações no SAC, verifica-se que o que mais ocorre é zero, ou seja, em vários dias não ocorre nenhuma reclamação. Assim, podemos, então, definir a **moda (Mo) como sendo o valor em um conjunto de dados que ocorre com maior freqüência.** Um conjunto de dados pode ser unimodal (uma moda) ou amodal (não possuir moda, pois não existe nenhum valor que ocorre com maior freqüência) ou multimodal (possui mais de uma moda).

Quando os dados não estão em intervalos de classes, basta olhar o valor que ocorre com maior freqüência.

Para dados agrupados em intervalos de classes, você pode calcular a moda por meio do método de Czuber, que se baseia na influência das classes adjacente na moda, deslocando-se no sentido da classe de maior freqüência. A expressão que você utilizará é:

$$Mo = L_i + \frac{d_1}{d_1 + d_2} \times c$$

onde:

L_i: limite inferior da classe modal;

d_1: diferença entre a freqüência da classe modal e a imediatamente anterior;

d_2: diferença entre a freqüência da classe modal e a imediatamente posterior; e

c: amplitude da classe modal.

No caso do tempo de uso de aparelhos celulares (ver a tabela no item distribuição de freqüências), teremos que a classe modal é a terceira, pois apresenta maior freqüência. Utilizando a expressão mostrada anteriormente, teremos:

$$Mo = L_i + \frac{d_1}{d_1 + d_2} \times c = 120{,}4 + \frac{8}{8+9} \times 25{,}6 = 132{,}44 \text{ min}$$

Uma característica importante da moda é que ela não é afetada pelos valores extremos da distribuição, desde que estes valores não constituam a classe modal.

Desta forma, a moda deve ser utilizada quando desejamos obter uma medida rápida e aproximada de posição ou quando a medida deva ser o valor mais freqüente da distribuição.

Outra medida de posição que você pode utilizar é a **mediana (Md).**

Em um conjunto de valores dispostos segundo uma ordem (crescente ou decrescente), é o valor situado de tal forma no conjunto que o separa em dois subconjuntos de mesmo número de elementos, ou seja, 50% dos dados são superiores à mediana, e 50% são inferiores.

O símbolo da mediana é dado por Md ou $\tilde{x}$. A posição da mediana é dada por meio da expressão: E (elemento central) = (n+1) / 2.

Considerando um conjunto de dados com número ímpar de elementos como (1, 2, 5, 9, 10, 12, 13), a posição da mediana será dada por (7 + 1)/2 = 4ª posição. Portanto, a partir dos dados ordenados, o número que se encontra na 4ª posição é o 9, e assim a mediana será igual a 9. (Temos três valores abaixo e três valores acima ou 50% acima da mediana, e 50% abaixo)

Caso o número de elementos do conjunto de dados for par, como por exemplo, (1, 2, 6, 8, 9, 12, 11, 13), encontra-se a posição da mediana ((**8 + 1)/2 = 4,5ª posição**). Como a posição 4,5 está entre a 4ª e a 5ª posição, calcula-se a média entre os valores que ocupam estas posições. O valor encontrado de 8,5 corresponde à mediana.

Quando os dados estão agrupados na mediana devemos encontrar a classe mediana.

Se os dados estão agrupados em intervalos de classe, como no caso do tempo de utilização do telefone, utilizaremos a seguinte expressão:

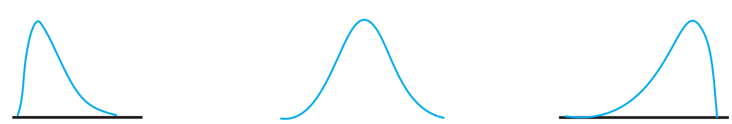
$$Md = li + \left(\frac{(n / 2) - f_{antac}}{f_{med}} \right) \times c$$

onde: l_i: limite inferior da classe mediana; n: número total de elementos; f_{antac}: freqüência acumulada anterior à classe mediana; f_{med}: freqüência absoluta da classe mediana e a amplitude da classe mediana.

Portanto, resolvendo o caso do tempo de utilização dos celulares, teremos que a posição da mediana será dada por E = 40/2 = 20ª elemento, o qual está na terceira classe (120,4 |– 146), que corresponde à classe mediana.

$$Md = li + \left(\frac{(n/2) - f_{antac}}{f_{med}} \right) \times c = 120{,}4 + \left(\frac{20 - 11}{16} \right) \times 25{,}6 = 134{,}8 \text{ min}$$

Em um conjunto de dados, a mediana, a moda e a média não necessariamente devem apresentar o mesmo valor. Uma informação importante é de que a mediana não é influenciada pelos valores extremos. Comparando os resultados encontrados para uma amostra em relação às medidas de posição estudadas e verificando a inter-relação entre elas, você pode concluir que seus valores podem nos dar um indicativo da natureza da distribuição dos dados, em função das regras definidas a seguir:



$\bar{x} > Md > Mo$, então a distribuição é assimétrica à direita (positiva)

$\bar{x} = Md = Mo$, então a distribuição é simétrica

$\bar{x} < Md < Mo$, então, a distribuição é assimétrica à esquerda (negativa)

Outras medidas de posição denominadas separatrizes serão definidas a seguir.

A principal característica das medidas separatrizes consiste na separação da série em partes iguais que apresentam o mesmo número de valores.

As principais são os quartis, decis e percentis.

Os quartis são valores de um conjunto de dados ordenados, que os dividem em quatro partes iguais. É necessário, portanto, três quartis

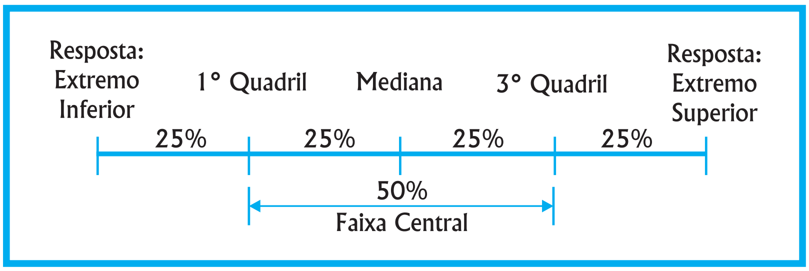
(Q_1, Q_2 e Q_3) para dividir um conjunto de dados ordenados em quatro partes iguais.

Q_1 : deixa 25% dos elementos abaixo dele.

Q_2 : deixa 50% dos elementos abaixo dele e coincide com a mediana.

Q_3 : deixa 75% dos elementos abaixo dele.

A figura abaixo mostra bem o quartis:



Se considerarmos o exemplo do número de reclamações por dia em um SAC, teremos de forma semelhante a figura anterior:

$$\frac{\textit{Mínimo}}{0} - - - - - \frac{Q_1}{0} - - - - - \frac{Q_2}{2} - - - - - \frac{Q_3}{3} - - - - - \frac{\textit{Máximo}}{5}$$

Para valores não tabelados, pode ser dito que o primeiro quartil pode ser obtido como a mediana da primeira metade dos dados, e para o terceiro quartil, como a mediana da segunda metade. Para dados tabelados, a fórmula da mediana pode ser adaptada para os demais quartis.

Medidas de dispersão

Como foi visto anteriormente, podemos sintetizar um conjunto de observações em alguns valores representativos como média, mediana, moda e quartis. Em várias situações, torna-se necessário visualizar como os dados estão dispersos. Tomando como exemplo várias empresas que apresentem salários médios iguais, podemos concluir, então, que a contribuição social (% do salário) será a mesma? Somente com base no salário médio, sim, mas estaríamos chegando a uma conclusão errada. A variação em termos de faixas salariais pode ser diferente, apesar de apresentarem a mesma média. Pensando no que foi dito anteriormente, considere o valor (em reais) ganho por dia de três grupos de empregados (A: 70, 70, 70, 70, 70; B: 50, 60, 70, 80, 90; C: 5, 15, 50, 120, 160).

Podemos verificar que, apesar de apresentarem a mesma média (70), os três grupos apresentam comportamento diferenciado, pois o grupo A é o mais homogêneo, e o grupo C é o que apresenta maior variação de ganho por dia. Portanto, devemos sempre inserir junto a uma medida de posição uma medida que avalie esta distribuição, ou seja, a variabilidade de um conjunto de dados. Portanto, quanto maior a variabilidade, maior será a dispersão das observações.

Uma primeira medida de dispersão que vamos comentar é a **amplitude total**. No caso dos ganhos diários, podemos obter os seguintes resultados:

O termo amplitude total foi visto anteriormente na construção de uma distribuição de freqüência em classes. Relembrando, é a diferença entre o maior e o menor valor observados.

$A_A = 70 - 70 = 0$ $\qquad$ $A_B = 90 - 50 = 40$ $\qquad$ $A_C = 160 - 5 = 155$

Verificamos, então, que o grupo C é o que apresenta maior variabilidade, e que o grupo A corresponde ao de menor variabilidade.

Deste modo, o grupo C corresponde àquele que teve maior variabilidade em torno da média.

No caso de dados agrupados, a amplitude total é calculada por meio da diferença entre o ponto médio da última classe e o ponto médio da primeira classe.

A amplitude total tem a desvantagem de só levar em conta os dois valores extremos, por isso é apenas uma indicação aproximada da dispersão. Outra desvantagem é que a amplitude total apresenta muita variação de uma amostra para outra, mesmo que ambas sejam extraídas da mesma população.

Portanto, você deve trabalhar com uma medida que leve em consideração todas as observações. Desta forma, podemos querer verificar o quanto um conjunto de observações está mais próximo ou mais distante de uma medida, que no caso será a média. Então, você pode calcular o desvio de cada valor em relação à média ($d_i = (x_i - \bar{x})$), e se fizermos o somatório destes desvios, o resultado será igual a zero. Se você elevar este desvio ao quadrado e somar, teremos o que chamamos de soma de quadrado dos desvios. Dividindo este somatório pelo total de observações, teremos uma idéia da dispersão das observações em relação à média. Esta medida que acabamos de visualizar de forma intuitiva corresponde à variância. Portanto, você pode concluir que a variância sempre assumirá valores positivos.

Quando o nosso interesse é o de tirar inferências válidas para toda a população a partir de uma amostra (porção representativa da população), deve-se trocar na fórmula da variância N por n – 1, onde:

- ***N** corresponde ao tamanho da população; e*
- ***n** corresponde ao tamanho da amostra utilizada.*

As expressões para cálculo das variâncias populacional e amostral são apresentadas a seguir.

	Populacional (σ^2)	Amostral (s^2)
Dados não agrupados	$\sigma^2 = \frac{SQD}{N} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$	$s^2 = \frac{SQD}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
Dados agrupados em classes	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 . fi$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 . fi$

Quando temos os dados agrupados em intervalos de classes, o x_i corresponde ao ponto médio da classe, e f_i à freqüência da classe.

Como a variância é calculada a partir dos quadrados dos desvios, ela é um número que apresenta a unidade elevada ao quadrado em relação à variável que não está elevada ao quadrado; isto se torna um inconveniente em termos de interpretação do resultado. Por isso, definiu-se uma nova medida, o **desvio-padrão**, que é a raiz quadrada da variância, com mais utilidade e interpretação práticas, representada por s ou σ. A variância é uma medida que tem pouca utilidade na Estatística Descritiva, mas será extremamente importante na Inferência Estatística e em combinações de amostras. Também é importante frisar que, na grande maioria das situações, trabalhamos com amostras, então devemos utilizar o desvio-padrão amostral.

No caso dos ganhos diários, calculando a variância de cada um dos grupos que correspondem a uma amostra, encontramos os seguintes resultados:

s_A = 0 reais; s_B = 15,81 reais; s_C=67,54 reais.

O desvio-padrão, quando analisado isoladamente, não dá margem a muitas conclusões. Por exemplo, para uma distribuição cuja média é 300, um desvio-padrão de 2 unidades é pequeno, mas para uma distribuição cuja média é 20, ele já não é tão pequeno.

Importante!

Condições para se usar o desvio-padrão ou variância para comparar a variabilidade entre grupos:

- mesmo número de observações;
- mesma unidade; e
- mesma média.

Além disso, se quisermos comparar duas ou mais amostras de valores expressas em unidades diferentes, não poderá ser possível fazer a comparação por meio do desvio-padrão, pois ele é expresso na mesma unidade dos dados. Também é necessário que os conjuntos de observações tenham o mesmo tamanho. Podemos, então, considerar a situação na qual se avaliou o custo indireto de fabricação (CIF) de um produto em reais e o tempo gasto em uma máquina para fabricação deste produto em segundos.

	$\bar{x}$	S
CIF	175 reais	5 reais
Tempo	68 segundos	2 segundos

A princípio, você poderia concluir que o CIF apresenta maior variabilidade. Entretanto, as condições citadas anteriormente deveriam ser satisfeitas para que se pudesse utilizar o desvio-padrão para comparar a variabilidade. Como as condições não são satisfeitas, devemos tentar expressar a dispersão dos dados em torno da média, em termos porcentuais. Então, utilizaremos uma medida estatística chamada de coeficiente de variação (CV). O coeficiente será dado por meio da expressão:

$CV = \frac{s}{\bar{x}} \times 100$, onde s e $\bar{x}$ foram definidos anteriormente

Para a situação do CIF e tempo, teremos:

$$CV_E = \frac{s}{\bar{x}} \times 100 = \frac{5}{175} \times 100 = 2{,}85\%$$

$$CV_P = \frac{s}{\bar{x}} \times 100 = \frac{2}{68} \times 100 = 2{,}94\%$$

Portanto, nesse grupo de indivíduos, o tempo de horas, máquina apresenta maior dispersão do que o custo indireto de fabricação (CIF), mudando, assim, a conclusão anterior.

Ao final desta parte de medidas de posição e dispersão, você deve ser capaz de calcular as medidas de posição e dispersão, e interpretá-las.

Caso não consiga, você deve voltar ao texto e fixar melhor os conceitos.

Seguem abaixo exercícios para fixação dos conhecimentos adquiridos nos assuntos de medida de posição e de dispersão. Você deve resolver todos eles.

Exercício 2: a tabela abaixo apresenta uma distribuição de freqüências das áreas de 400 lotes:

A partir da tabela acima, calcule:

Áreas (m^2)	Nº de lotes
300 |– 400	14
400 |– 500	46
500 |– 600	58
600 |– 700	76
700 |– 800	68
800 |– 900	62
900 |– 1000	48
1000 |– 1100	22
1100 |– 1200	6

a) média, mediana e moda;

b) desvio padrão e coeficiente de variação;

c) o ponto médio da sétima classe;

d) a amplitude do intervalo da segunda classe;

e) a freqüência relativa da sexta classe;

f) a freqüência acumulada da quinta classe;

g) o nº de lotes cuja área não atinge 700 m^2;

h) o nº de lotes cuja área atinge e ultrapassa 800 m^2; e

i) a classe do 72º lote.

Exercício 3: os dez funcionários de uma pequena empresa receberam os seguintes salários, em reais:

230, 210, 100, 140, 160, 120, 390, 450, 100 e 200

Calcule as medidas de posição e dispersão em relação aos salários

Exercício 4: uma loja vende cinco produtos básicos A, B, C, D e E. O lucro por unidade comercializada destes produtos vale, respectivamente $ 200,00; $ 300,00; $ 500,00; $ 1.000,00; $ 5.000,00. A loja vendeu em determinado mês 20; 30; 20; 10; 5 unidades, respectivamente. Qual foi o lucro médio por unidade comercializada por esta loja?

Exercício 5: uma empresa tem duas filiais praticamente idênticas quanto às suas características funcionais. Um levantamento sobre os salários dos empregados dessas filiais resultou nos seguintes valores:

Filial A: $\overline{x}_A = 400$ e $S_A = 20$

Filial B: $\overline{x}_B = 500$ e $S_B = 25$

Podemos afirmar que as duas filiais apresentam a mesma dispersão?

Saiba mais...

- Sobre cálculo de médias e funções em planilhas, visite o site: http://www.juliobattisti.com.br/tutoriais/celsonunes/openoffice007.asp
- Mais exercícios referentes ao assunto estão no site: http://www.famat.ufu.br/prof/marcelo/exercicios.htm

UNIDADE

2

Introdução a probabilidades

Introdução a probabilidades

Quando estamos falando de probabilidade, queremos identificar a chance de ocorrência de um determinado resultado de interesse, em situações nas quais não é possível calcular com exatidão o valor real do evento. Desta forma, trabalhamos com chances ou probabilidades.

Uma situação, para exemplificarmos este fato, está associada à seguinte pergunta: meu vendedor poderá cumprir sua meta de venda na semana que vem? O **espaço amostral*** simbolizado por S ou Ω nesta situação será **atinge a meta** e **não atinge a meta**. Para calcular a probabilidade de cumprir a meta, você pode usar a intuição (subjetivo) ou usar a freqüência relativa das últimas dez semanas em que o vendedor esteve trabalhando (objetivo).

Portanto, para calcularmos uma probabilidade, é necessário que tenhamos um **experimento aleatório***, que apresenta as seguintes características: a) cada experimento pode ser repetido indefinidamente sob as mesmas condições (n); b) não se conhece a priori o resultado do experimento, mas podem-se descrever todos os possíveis resultados; e c) quando o experimento for repetido um grande número de vezes, surgirá uma regularidade do resultado, isto é, haverá uma estabilidade da fração $f = \frac{r}{n}$ (freqüência relativa) da ocorrência de um particular resultado, onde r corresponde ao número de vezes que um determinado resultado aconteceu.

Nos experimentos ou situações mencionadas, você pode notar que a incerteza sempre está presente, o que quer dizer que, se estes experimentos forem repetidos em idênticas condições, não se pode determinar qual o resultado ocorrerá.

A incerteza está associada à chance de ocorrência que atribuímos ao resultado de interesse.

Consideremos, como exemplo, os funcionários que trabalham no setor de marketing de uma determinada empresa. Sabe-se que nes-

GLOSSÁRIO

*Espaço amostral– conjunto de possibilidades, ou seja, os possíveis resultados associados a um experimento aleatório.

*Experimento aleatório – qualquer processo que venha a gerar um resultado incerto ou casual.

te setor trabalham seis funcionários. Um experimento ao acaso seria a escolha aleatória de um dos funcionários. Podemos considerar como evento de interesse o sexo do funcionário escolhido. Você, então, vai aplicar os conceitos vistos acima e novos conceitos associados a probabilidades.

GLOSSÁRIO

*Evento – qualquer subconjunto de um espaço amostral.

- Conjunto de possibilidades (Espaço amostral): S = {Carlos, Jackeline, Giulyana, Girlene, Cláudio, Larissa}.
- Conjunto de possibilidades favoráveis (Funcionários do sexo masculino que correspondem a um **evento***): {Carlos, Cláudio}.
- Qual a probabilidade de escolher um funcionário ao acaso e ele ser do sexo masculino? (Sugestão: verificar o número de funcionários do sexo masculino).

$$\frac{2}{6} = \frac{\text{número de funcionários do sexo masculino}}{\text{número total de funcionários}}$$

Dados três eventos, A (funcionário ser do sexo feminino) e B (seu nome começa com a letra G) e C (seu nome começa com a letra C) dos funcionários do setor de marketing apresentado anteriormente:

- $A \cap B$ é o evento em que A e B ocorrem simultaneamente: **{Giulyana, Girlene}**.
- $A \cup C$ é o evento em que A ocorre ou C ocorre (ou ambos): **{Carlos, Jackeline, Giulyana, Girlene, Cláudio, Larissa}**.
- $\overline{A}$ é o evento em que A não ocorre (complementar de A): **{Carlos, Claudio}**.

Dois eventos são considerados mutuamente exclusivos, se a ocorrência de um exclui a ocorrência do outro.

Você pode, então, definir a probabilidade como uma função que atribui um número real aos eventos de Ω (se A é um evento de Ω, P(A) é a probabilidade de A), que satisfaz:

1. $P(\varnothing) = 0$, $P(\Omega) = 1$

2. $0 \leq P(A) \leq 1$

3. Regra da soma: dados dois **eventos mutuamente exclusivos** A e C de Ω,

$P(A \cup C) = P(A) + P(C)$

O símbolo $\varnothing$ corresponde à ocorrência de um evento impossível, ou seja, que não pode ocorrer no espaço amostral considerado.

OBS: Caso os eventos não sejam mutuamente exclusivos, na regra da soma, devemos considerar que a intersecção será contada duas vezes. Então, devemos retirar na regra da soma a intersecção. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Considerando os eventos A, B e C definidos anteriormente, calcule as probabilidades mencionadas na página anterior para fixação dos conceitos.

Para inserirmos outros conceitos de probabilidade, você deve considerar os dados a seguir referentes ao acesso e cadastro em dois sites, por pessoas em uma determinada região. O site 1 segue o padrão normal, enquanto o site 2 corresponde a uma nova proposta de apresentação de informações.

	Acessa e cadastra no site	Acessa e não cadastra no site	Total
Site 1	39.577	8.672	48.249
Site 2	46.304	7.297	53.601
Total	85.881	15.969	101.850

Um acesso a um dos sites é escolhido ao acaso. Podemos considerar, então, que o nosso espaço amostral (Ω) corresponderá ao conjunto de 101.850 acessos.

Há os seguintes eventos de interesse:

- S1 = número de acessos feitos no site 1.

- S2 = número de acessos feitos no site 2.
- AC = o site é acessado, e o cadastro é feito pelo internauta.
- S1 ∩ AC = o internauta acessa o site 1 e faz o cadastro no site.
- S1 ∪ AC = o internauta acessa o site 1 ou faz o cadastro no site.

Você pode obter, então, algumas probabilidades como:

$$P(AC) = \frac{\text{nº de acessos que cadastram}}{\text{nº total de acessos}} = \frac{85.881}{101.850} = 0{,}843$$

$$P(S1) = \frac{\text{nº de acessos ao site 1}}{\text{nº total de acessos}} = \frac{48.249}{101.850} = 0{,}473$$

$$S2 = \overline{S1} \Rightarrow P(S2) = P(\overline{S1}) = 1 - P(S1) = 1 - 0{,}473 = 0{,}527$$

$$P(S1 \cup AC) = P(S1) + P(AC) - P(S1 \cap AC)$$

$$= 0{,}473 + 0{,}843 - 0{,}388$$

$$= 0{,}928$$

$$\frac{39577}{101850} = 0{,}388$$

Se você relembrar a interpretação da probabilidade, considerando A um evento de um espaço amostral associado a um experimento aleatório, você pode ter duas formas de atribuir probabilidades aos eventos de um espaço amostral:

- P{A} é uma intuição (subjetiva) que se deposita na ocorrência de A.
- Interpretação freqüêntista (objetiva).

$$f_n(A) = \frac{\text{nº de repetições que A ocorre}}{n}$$

Quando n cresce: $f_n(A)$ se aproxima da P(A), por isso foram realizadas n repetições independentes do experimento.

No exemplo anterior, se você souber que um acesso sorteado é do site 1, qual é a probabilidade de que ocorra a efetuação do cadastro?

Temos uma informação parcial: o acesso é do site 1.

Vamos designar a probabilidade de AC, quando se sabe que o acesso ocorreu no site 1, que chamaremos de P(AC /S1) e denominá-la **probabilidade (condicional) de AC dado S1 (lembre-se que o símbolo / não corresponde a uma divisão).**

É natural atribuirmos:

$$\text{P (AC/S1)} = \frac{\text{n}^{\text{o}} \text{ de acessos em que ocorreu cadastro}}{\text{n}^{\text{o}} \text{ total acessos no site 1}}$$

$$= \frac{39.577}{48.249} = 0{,}820$$

Note que:

$$\text{P (AC/S1)} = \frac{\dfrac{\text{n}^{\text{o}} \text{ de acessos em que ocorreu cadastro}}{\text{n}^{\text{o}} \text{ total acessos}}}{\dfrac{\text{n}^{\text{o}} \text{ de acessos ao site 1}}{\text{n}^{\text{o}} \text{ total de acessos}}}$$

$$\text{P (AC/S1)} = \frac{\text{P}(AC \cap S)}{\text{P}(S1)}$$

Portanto, você pode generalizar para dois eventos A e B quaisquer de um experimento aleatório. Desta forma, podemos dizer que a probabilidade condicional de A dado B (nota-se por P (A / B)) é definida como:

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

Podemos, então, definir a **regra do produto**, ou seja, a partir da probabilidade condicionada definida anteriormente, obteremos a chamada regra do produto para a probabilidade da interseção de dois eventos A e B de um espaço amostral:

Passe a probabilidade de ocorrência de B na probabilidade condicionada e multiplique pela probabilidade de ocorrência de A sabendo que B já aconteceu.

$$P\ (A \cap B) = P\ (A\ /\ B)\ .\ P\ (B)$$

Se dois eventos A e B são independentes, então P{A / B} = P{A} ou P{B / A} = P(B).

Deste modo, se A e B forem independentes, você pode verificar que:

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

Veja esta outra situação, utilizando os conceitos de probabilidade condicionada e independência de eventos. Considere a tabela a seguir, representativa da distribuição da renda anual de produtores rurais e duas cooperativas em uma determinada região.

Faixa de renda anual (em R$1.000)	Cooperativas		Total
	A	B	
15 a 20 (R1)	70	40	110
20 a 25 (R2)	15	15	30
25 a 30 (R3)	10	20	30
30 a 35 (R4)	20	10	30
Total	115	85	200

Observando-se os dados acima, se verifica que a probabilidade de um cooperado aleatoriamente escolhido ser:

a) da cooperativa A: P(A) = 115/200 = 0,575

b) da cooperativa B: P(B) = 85/200 = 0,425

c) de ter renda entre R$ 15.000,00 e R$ 20.000,00: P(R1) = 110/200 = 0,550

d) da cooperativa B e ter renda entre R$ 15.000,00 e R$ 20.000,00: P(B ∩ R1) = 40/200 = 0,20

e) ter renda entre R$ 15.000,00 e R$ 20.000,00 dado que é da cooperativa B: P(R1/B) = 40/85 = 0,4706 ou

$$P(R1/B) = \frac{P(R1 \cap B)}{P(B)} = \frac{0{,}20}{0{,}425} = 0{,}4706$$

Como P(R1) $\neq$ P(R1/B), conclui-se que os eventos cooperativa e renda são dependentes.

Para que sejam considerados independentes, a relação de independência deve ser válida para todas as intersecções presentes na Tabela 11.

Um exemplo de aplicação dos conceitos de independência de eventos pode ser visualizado por meio do lançamento de uma moeda não viciada (não existe preferência para cara ou coroa) três vezes. Considere os seguintes eventos:

A = no primeiro lançamento da moeda, sai cara, e

B = no segundo lançamento da moeda, sai cara.

Obs: considere C = cara e R = coroa

Verifique se é verdadeira a hipótese de que os eventos A e B são independentes. O espaço amostral e os eventos são apresentados a seguir:

Ω = {CCC, CCR, CRC, CRR, RCC, RCR, RRC, RRR}

(A) = {**CCC, CCR**, CRC, CRR}

(B) = {**CCC, CCR**, RCC, RCR}

Os resultados que estão em negrito ocorrem no espaço amostral (8) somente duas vezes.

$P(A \cap B) = 2/8 = 1/4$

$P(A) = 4/8 = 1/2$

$P(B) = 4/8 = 1/2$

Portanto,

$P(A \cap B) = P(A) \,.\, P(B) \Rightarrow 1/4 = 1/2 \,.\, 1/2$ ou

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{1/4}{1/2} = \frac{2}{4} = \frac{1}{2}$$

$P(A/B) = P(A) \Rightarrow 1/2 = 1/2$

Então, provamos que os eventos são independentes.

Vamos resolver alguns exercícios relacionados aos conceitos de probabilidade vistos anteriormente. (Os resultados estão no final do livro.)

Exercício 1: as probabilidades de três vendedores, A, B e C, que trabalham independentemente, efetivarem uma venda quando abordam um cliente são 2/3, 4/5 e 7/10, respectivamente. Se cada um abordar um cliente, qual a probabilidade de que pelo menos um efetive a venda?

Exercício 2: A e B são dois mestres que já estão suficientemente treinados em partidas de xadrez e jogam 120 partidas, das quais A ganha 60, B ganha 40, e 20 terminam empatadas. A e B concordam em jogar três partidas. Determinar a probabilidade de:

a) A ganhar todas as três partidas;

b) duas partidas terminarem empatadas; e

c) A e B ganharem alternadamente.

Exercício 3: num período de um mês, cem funcionários de uma empresa que trabalha com resíduos nucleares, sofrendo de determinada doença, foram tratados. Informações sobre o método de tratamento aplicado a cada funcionário e o resultado final obtido estão na tabela abaixo:

		Tratamento	
		A	**B**
	Cura total	24	16
Resultado	Cura parcial	24	16
	Morte	12	8

a) Sorteando-se aleatoriamente um desses funcionários, determine a probabilidade de que o funcionário escolhido:

a1) tenha sido submetido ao tratamento A;

a2) tenha sido totalmente curado;

a3) tenha sido submetido ao tratamento A e tenha sido parcialmente curado; e

a4) tenha sido submetido ao tratamento A ou tenha sido parcialmente curado.

Exercício 4: para selecionar seus funcionários, uma empresa oferece aos candidatos um curso de treinamento durante uma semana. Ao final, eles são submetidos a uma prova, e 25% são classificados como bons (B), 50%, como médios (M), e os demais 25%, como fracos (F). Como medida de economia, o departamento de seleção pretende substituir o treinamento por um teste contendo perguntas de conhecimentos gerais e específicos. Mas, para isso, gostaria de conhecer qual a probabilidade de um indivíduo aprovado no teste ser considerado fraco caso fizesse o teste. De acordo com os resultados, receberam os conceitos: aprovado (A) ou reprovado (R). Sabendo que P(A B) = 0,20; P(A M) = 0,25 e P(A F) = 0,05; encontrar P(A/F).

Variáveis aleatórias

Você pode definir uma variável aleatória como sendo uma função que associa valores reais aos eventos de um espaço amostral, e que pode ser discreta ou contínua.

Um exemplo de uma variável aleatória discreta (v.a) consiste em verificar o número de ações que tiveram queda em um determinado dia, em uma carteira composta por cinco ações diferentes. A função será dada por:

X= “número de ações que tiveram queda em um determinado dia”. Define uma variável aleatória discreta, que pode assumir os valores 0, 1, 2, 3, 4, 5.

Vamos considerar agora uma situação na qual se verificou o tempo gasto por um vendedor para convencer um cliente a adquirir um determinado produto. A função será:

Y= "tempo gasto por um vendedor para convencer um cliente a adquirir um determinado produto". Define uma variável aleatória contínua, que pode assumir infinitos valores.

Se uma variável aleatória X pode assumir os valores $x_1, x_2, ..., x_n$ com probabilidades respectivamente iguais a $p_1, p_2, ..., p_n$, e $\sum_{i=1}^{n} p_i = 1$, tem-se definida uma **distribuição de probabilidade**.

É importante ressaltar que a variável aleatória tem notação de letra maiúscula, e seus possíveis valores, minúsculos, como apresentado no parágrafo anterior.

Se a variável X em questão for discreta, sua distribuição é caracterizada por uma **função de probabilidade (P(X=x))**, que associa probabilidades não nulas aos possíveis valores da variável aleatória.

Para o exemplo do número ações da carteira, as probabilidades obtidas são mostradas na função de probabilidade que corresponde à tabela abaixo.

X	0	1	2	3	4	5	
P(X=x)	1/10	1/10	2/10	3/10	4/10	5/10	Σ =1,00

Se a variável X for contínua, somente haverá interesse na probabilidade de que a variável assuma valores dentro de determinados intervalos, sendo sua distribuição de probabilidades caracterizada por uma função densidade de probabilidade (f.d.p.), f(x), a qual deverá possuir as seguintes propriedades:

- $f(x) \geq 0$;
- $\int_{-\infty}^{\infty} f(x)dx = 1$

- $\int_a^b f(x)dx = P(a \le x \le b)$

A área compreendida entre os pontos a e b, da função f(x) e o eixo das abscissas, corresponde à probabilidade da variável X assumir valores entre a e b.

Para o caso do tempo gasto para convencer um cliente a adquirir um produto, podemos, por exemplo, ter a função abaixo, que corresponde a uma distribuição normal **que será vista posteriormente**:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$, que é a distribuição normal.

A função repartição ou distribuição acumulada, representada por F(x), corresponde à probabilidade de a variável aleatória ser menor ou igual a um determinado valor de x.

Se a variável for discreta, a distribuição acumulada será dada por $F(x) = P(X \le x)$, ou seja, você deve somar todas as probabilidades que se tem abaixo de um determinado valor, inclusive este.

Já no caso de uma variável contínua, o F(x) será dado pela área que vai de $-\infty$ até o ponto x a ser considerado. Portanto, teremos:

Integral de $-\infty$ até o ponto x (visto no módulo de Matemática).

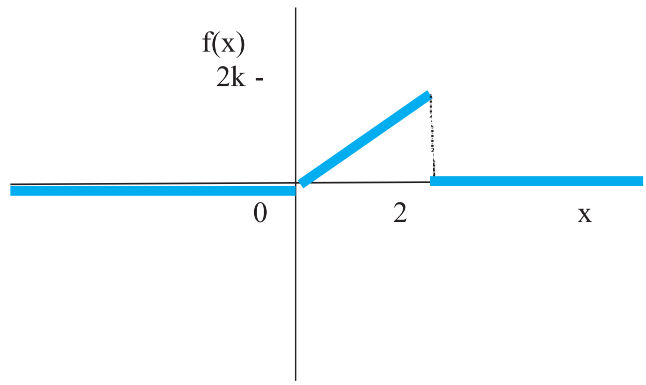
$$F(x) = \int_{-\infty}^{x} f(x)dx$$

Agora, você vai ver um exemplo de utilização destes conceitos.

Seja a seguinte variável aleatória contínua, definida pela função densidade de probabilidade (f.d.p):

$$\begin{cases} f(x) = 0 \text{ para } x < 0 \\ f(x) = kx \text{ para } 0 \le x \le 2 \\ f(x) = 0 \text{ para } x > 2 \end{cases}$$

graficamente, tem-se:



a) Obtenha o valor de k.

Como f(x) é uma fdp:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$\int_{-\infty}^{0} 0\,dx + \int_{0}^{2} kxdx + \int_{2}^{\infty} 0dx = 1$$

$$0 + k\int_{0}^{2} xdx + 0 = 1 \;; \quad k\left.\frac{x^2}{2}\right|_{0}^{2} = 1 \;;\; 4\frac{k}{2} = 1 \;;$$

$$k = 1/2$$

O resultado encontrado corresponde à inclinação da reta, ou seja, o quanto que a função aumenta, quando a variável x é acrescida de uma unidade.

---

b) calcular F(1).

F(1) = P(X 1) = $\displaystyle\int_{0}^{1} \frac{x}{2}dx = \frac{1}{4}$

Para o estudo de variáveis aleatórias, até este ponto, considerou-se que o resultado do experimento em questão seria registrado como um único valor x. Todavia, existem casos em que há interesse por dois resultados simultâneos, como por exemplo, observar o peso e altura de uma pessoa, o sexo e desempenho no trabalho, etc. Para tanto, faz-se necessária a seguinte definição:

Sejam E um experimento aleatório, e S o espaço amostral associado a E.

*Sejam X e Y duas variáveis aleatórias. Então, (X,Y) define uma variável **aleatória bidimensional**, que pode ser discreta, contínua ou mista.*

O principal objetivo da análise de variáveis aleatórias bidimensionais é avaliar simultaneamente dois resultados de uma situação associando as probabilidades individuais e conjuntas.

Vamos, então, definir probabilidades ou distribuições conjuntas e marginais.

A distribuição conjunta é a distribuição simultânea das duas variáveis, ou seja, a intersecção das variáveis e as distribuições marginais são as distribuições isoladas de cada variável. Estas distribuições são assim chamadas por ocuparem, em uma tabela, a parte central e as margens das tabelas, respectivamente. Você pode visualizar este fato na tabela apresentada na página seguinte.

Se (X,Y) é uma variável aleatória bidimensional discreta, sua função de probabilidade, representada por $P(X = x_i ; Y = y_i)$ que associa um valor $p(x_i, y_i)$ a cada valor do par (X,Y), deve satisfazer as seguintes condições:

- $P(x_i, y_i) = 0$
- $\sum_i \sum_j p(x_i, y_i) = 1$

Veja a seguinte situação: uma pesquisa foi realizada para verificar a existência de relação entre a utilização de um produto (baixa, média ou alta) e o grau de instrução das pessoas (Fundamental, Médio e Superior). Como os resultados associam duas variáveis, então temos uma distribuição de probabilidade de variáveis aleatórias bidimensionais. No resultado encontrado (mostrado a seguir), temos um quadro chamada de contingência.

Utilização	Instrução			
	Fundamental	Médio	Superior	Total
Alta	100	120	65	285
Média	350	100	20	470
Baixa	150	80	15	245
Total	600	300	100	1000

Vamos, então, calcular o quadro das probabilidades (dividindo cada valor por 1.000 que é o tamanho da amostra utilizada.

Utilização	Instrução			
	Fundamental	Médio	Superior	Total
Alta	0.100	0.120	0.065	0.285
Média	0.350	0.100	0.020	0.470
Baixa	0.150	0.080	0.015	0.245
Total	0.600	0.300	0.100	1.000

Distribuição Marginal da Utilização.

Distribuição Conjunta do Grau de Instrução e Utilização.

Distribuição Marginal do Grau de Instrução.

Resolva agora estes dois exercícios e, caso tenha dúvida, releia o texto relativo a variáveis aleatórias (os resultados estão no final do livro).

Exercício 5: uma empresa tem quatro caminhões de aluguel. Sabendo-se que o aluguel é feito por dia e que a distribuição diária do número de caminhões alugado é a seguinte, determine:

Nº de caminhões alugados / dia	0	1	2	3	4
Probabilidade de alugar	0,1	0,2	0,3	0,3	0,1

a) Qual é a probabilidade de alugar num dia mais de dois caminhões?

b) Qual é a probabilidade de alugar no mínimo um caminhão?

c) Qual a probabilidade de alugar no máximo dois caminhões?

d) Determine a função de distribuição acumulada.

e) Qual o valor de F(3)? O que significa este resultado?

Exercício 6: a proporção de álcool em um certo composto pode ser considerada uma variável aleatória com a seguinte função de densidade:

$$f(x) = \begin{cases} 20x^3(1-x), & para\ 0 \le x \le 1 \\ 0 & , para\ outros\ valores\ de\ x \end{cases}$$

Calcule a probabilidade da proporção de álcool neste composto entre 0,20 e 0,25.

Distribuições de variáveis aleatórias discretas

Distribuição Uniforme Discreta

Enquadram-se aqui as distribuições em que os possíveis valores da variável aleatória tenham todos a mesma probabilidade de ocorrência. Logo, se existem n valores possíveis, cada um terá probabilidade igual a 1/n.

Ex. Seja o lançamento de um dado e a variável aleatória X = "face superior do dado",

tem-se que:

X	1	2	3	4	5	6	
P(X=x)	1/6	1/6	1/6	1/6	1/6	1/6	$\sum = 1$

ou P(X=x) = 1/6

Desta forma, você pode verificar que esta variável segue uma distribuição uniforme discreta, pois a variável é discreta, e todos os possíveis resultados da variável aleatória têm a mesma probabilidade (1/6).

Distribuição de Bernoulli

Imagine uma situação na qual só podem ocorrer dois possíveis resultados, "sucesso" e "fracasso". Veja alguns exemplos:

- uma venda é efetuada ou não em uma ligação de call center;
- um cliente pode ser adimplente ou inadimplente;
- uma peça produzida por uma cia. pode ser perfeita ou defeituosa; e
- um consumidor que entra numa loja pode comprar ou não comprar um produto.

Associando-se uma variável aleatória X aos possíveis resultados do experimento, de forma que:

X= 1 se o resultado for "sucesso",

X= 0 se o resultado for "fracasso".

Está relacionado com o objetivo do trabalho a ser realizado.

Então, a variável aleatória X, assim definida, tem distribuição Bernoulli, com p sendo a probabilidade de ocorrer **"sucesso"**, e q = (1-p) a probabilidade de ocorrer **"fracasso"**.

A função de probabilidade da Distribuição de Bernoulli é dada por:

$$P(X = x) = \begin{cases} p \text{ para } x = 1, \\ q = 1 - p \text{ para } x = 0 \\ 0 \text{ para } x \text{ diferente de } 0 \text{ ou } 1 \end{cases}$$

A média e a variância serão obtidas por:

Média = p

Variância = pq

Contextualizando a distribuição de Bernoulli, temos a seguinte situação: a experiência tem mostrado que, durante as vendas de Natal, um cliente que entra em uma determinada loja tem 60% de chance de comprar um produto qualquer. Temos, portanto, uma probabilidade de sucesso (**o cliente adquirir um produto qualquer**) de 0,6 e uma probabilidade de não adquirir um produto de 0,4 (**vem da diferença q = 1-0,6**).

Distribuição binomial

Para que uma situação possa se enquadrar em uma distribuição binomial, deve atender às seguintes condições:

Pense, como exemplo de forma didática no lançamento de uma moeda 50 vezes, e veja se esta situação se enquadra nas condições da distribuição binomial.

- são realizadas n repetições (tentativas) independentes;
- cada tentativa é uma prova de Bernoulli (só podem ocorrer dois possíveis resultados); e
- a probabilidade p de sucesso em cada prova é constante.

Se uma situação atende a todas as condições acima, então a variável aleatória X = número de sucessos obtidos nas n tentativas terá uma distribuição binomial, com n tentativas e p (probabilidade de sucesso).

Simbolicamente, temos: X ~ B(n,p) com a interpretação dada a seguir:

A variável aleatória x tem distribuição binomial com n ensaios e uma probabilidade p de sucesso. (em cada ensaio).

A função de probabilidade utilizada para cálculo de probabilidades, quando a situação se enquadra na distribuição binomial, será dada por meio da seguinte expressão:

$C_n^x = \frac{n!}{x!(n-x)!}$, *onde n! corresponde ao fatorial de n.*

p = probabilidade de "sucesso" em cada ensaio

q = 1-p = probabilidade de "fracasso" em cada ensaio

Para exemplificar a utilização da distribuição binomial, você deve considerar que pessoas entram em uma loja no período próximo ao Dia das Mães. Sabe-se que a probabilidade de uma pessoa do sexo masculino comprar um presente é de 1/3. Se entrarem quatro pessoas do sexo masculino nesta loja, qual a probabilidade de que duas venham a comprar presentes?

Se as quatro pessoas entram na loja e duas delas compram, podemos colocar as possibilidades da seguinte forma (C ➔ compra e não-C ➔ não compra). O espaço amostral associado ao experimento é:

> C, C, não-C, não-C ou C, não-C, não-C, C ou C, não-C, C, não-C ou
>
> não-C, não-C, C, C ou não-C, C, não-C, C ou não-C, C, C, não-C

Logo, calculando as probabilidades usando as regras do "**e**" (multiplicação, pois são independentes) e do "**ou**" (soma), a probabilidade de dois clientes do sexo masculino comprarem presentes é:

$$p = \frac{1}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{2}{3} + \frac{1}{3}\cdot\frac{2}{3}\cdot\frac{2}{3}\cdot\frac{1}{3} + \frac{1}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}\cdot\frac{2}{3} + \frac{2}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}\cdot\frac{1}{3} + \frac{2}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{1}{3} + \frac{2}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}$$

$$p = 6\cdot\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{2}{3}$$

$$p = 6\cdot\left(\frac{1}{3}\right)^2\cdot\left(\frac{2}{3}\right)^2$$

$$p = \frac{24}{81} \cong 29{,}63\%$$

Como na binomial são n ensaios de Bernoulli e a distribuição tem média p, a média da binomial será np. Raciocínio semelhante é feito para a variância.

Agora, você deve calcular utilizando a função de probabilidade apresentada anteriormente e verificar que o resultado será o mesmo.

Os valores da média e da variância da distribuição binomial são:

> Média = np
>
> Variância = npq

Um outro exemplo de utilização da distribuição binomial é o seguinte. Em um determinado processo de fabricação, 10% das peças produzidas são consideradas defeituosas. As peças são acondicionadas em caixas com cinco unidades cada uma. Considere que cada peça tem a mesma probabilidade de ser defeituosa (como se houvesse repetição no experimento de retirar uma peça).

a) Qual a probabilidade de haver exatamente três peças defeituosas numa caixa?

P = 0,1 n = 5

b) Qual a probabilidade de haver duas ou mais peças defeituosas em uma caixa?

c) Qual a probabilidade de uma caixa não apresentar nenhuma peça defeituosa?

d) Supondo que a empresa pague uma multa de R$ 10,00 por caixa que apresente peças defeituosas, qual o valor esperado desta multa em um lote de 1.000 caixas?

P(uma caixa ter peça defeituosa) = 1- P(X=0) = 0,4095

Temos, então, uma nova variável aleatória (número de caixas com peças defeituosas), a qual chamaremos de Y em um lote de 1.000 caixas, que segue uma distribuição binomial com n=1.000 e p=0,4095.

E(Y) = np = 1000.0,4095 = 409,5 caixas.

Multa Esperada = 409,5 . R$ 10,00 = R$ 4.095,00

Distribuição de Poisson

Você pode empregar a distribuição de Poisson em situações nas quais não está interessado no número de sucessos obtidos em n tentativas, como ocorre no caso da distribuição binomial, entretanto este número de sucessos deve estar dentro de um intervalo contínuo, ou seja, **o número de sucessos ocorridos durante um intervalo contínuo**, que pode ser um intervalo de tempo, espaço, etc. Imagine que você queira estudar o número de suicídios ocorridos em uma cidade durante um ano ou o número de acidentes automobilísticos ocorridos numa rodovia em um mês, ou o número de defeitos encontrados em um rolo de arame ovalado de 500 m. Estas situações são exemplos que se enquadram na distribuição de Poisson.

Note que, nos exemplos acima, não há como você determinar a probabilidade de ocorrência de um sucesso, mas sim a freqüência média de sua ocorrência, como por exemplo, dois suicídios por ano, que denominaremos λ.

Em uma situação com estas características, a variável aleatória X = número de sucessos em um intervalo contínuo terá uma distribuição Poisson, com (freqüência média de sucesso). Simbolicamente, podemos utilizar a notação X ~ P(λ).

A variável aleatória x tem uma distribuição de Poisson com uma freqüência média de sucesso λ.

A função de probabilidade da distribuição de Poisson será dada por meio da seguinte expressão:

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

Onde e =2,7182 (base dos logaritmos neperianos) e λ corresponde à freqüência média de sucesso no intervalo contínuo em que se deseja calcular a probabilidade.

Vamos considerar que o Corpo de Bombeiros de uma determinada cidade recebe, em média, três chamadas por dia. Queremos saber, então, qual a probabilidade de do Corpo de Bombeiros receber:

a) quatro chamadas num dia:

Como o intervalo em que se deseja calcular a probabilidade é um dia, então, o λ será igual a 3.

X~P(3)

$$P(X = 4) = e^{-3} \frac{3^4}{4!} = 0{,}1680$$

b) nenhuma chamada em um dia:

$$P(X = 0) = e^{-3} \frac{3^0}{0!} = 0{,}0498$$

c) 20 chamadas em uma semana:

λ = 21 chamadas por semana

Como o intervalo desejado é uma semana, ou seja, sete dias, então, em uma semana a freqüência média de chamadas será de sete dias vezes 3 chamadas/dia.

$$P(X = 20) = e^{-21}\frac{21^{20}}{20!} = 0{,}0867$$

Uma característica da distribuição de Poisson é que as estatísticas da distribuição (média e variância) apresentam o mesmo valor, ou seja, são iguais a λ. Então, teremos:

$$\text{Média} = \text{Variância} = \lambda$$

Vamos fazer alguns exercícios relativos à distribuição binomial e de Poisson.

Exercício 7: no Brasil, a proporção de microempresas que fecham em até um ano é de 10%. Em uma amostra aleatória de 20 microempresas, qual a probabilidade de cinco terem fechado em até um ano de criação?

R: $P(X = 5) = C_{20}^{5}\, 0{,}1^{5} 0{,}9^{15} = 0{,}03192$

Exercício 8: entre 2.000 famílias de baixa renda, com quatro crianças e considerando que a chance de nascer uma criança do sexo masculino é igual à do sexo feminino, em quantas famílias se esperaria que tivessem:

$n = 4$ e $p = ½$

a) dois meninos? R: P(x=2) . 2.000 = 0,3750 . 2.000 = 750 famílias.

b) Um ou dois meninos? R: [P(1) + P(2)] . 2.000 = (0,25 + 0,375) . 2.000 = 1.250 famílias.

c) Nenhum menino? R: P(0) . 2.000 = 0,0625 . 2.000 = 125 famílias.

Exercício 9: a probabilidade de compra de um aparelho de celular é igual a 30%. Observando oito compradores, qual a probabilidade de quatro deles comprarem este aparelho?

R: P(X = 4) = $C_8^4 0{,}3^4 0{,}7^4$ = 0,13614

Exercício 10: chegam caminhões a um depósito à razão de 2,8 caminhões/hora, segundo uma distribuição de Poisson. Determine a probabilidade de chegarem dois ou mais caminhões:

a) num período de 30 minutos;

b) num período de 1 hora; e

c) num período de 2 horas.

R: 1- [P(0) + P(1)]

a) $\lambda = 1{,}4$ R= 0,40817

b) $\lambda = 2{,}8$ R=0,76892

c) $\lambda = 5{,}6$ R=0,97559

Distribuições de Probabilidade Contínuas

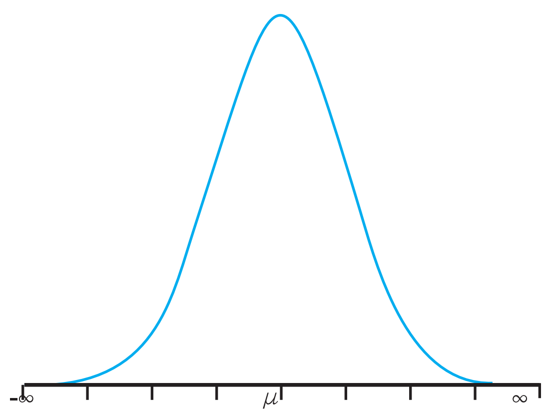
Dentre as várias distribuições de probabilidade contínuas, será abordada aqui apenas a distribuição normal, pois apresenta grande aplicação em pesquisas científicas e tecnológicas. Grande parte das variáveis contínuas de interesse prático segue esta distribuição, aliada ao Teorema do Limite Central (TLC), que é a base das estimativas e dos testes de hipóteses, realizados sobre a média de uma população qualquer, e garante que a distribuição amostral das médias segue uma distribuição normal, independentemente da distribuição da variável em estudo, como será visto mais adiante.

A função densidade de probabilidade da distribuição normal é dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, x \in R.$$

onde μ e σ são a média e desvio-padrão, respectivamente, da distribuição de probabilidade e π corresponde a 3,1415 e exp a uma função exponencial.

O gráfico da distribuição normal, utilizando a função mostrada anteriormente, e os conceitos vistos no módulo de Matemática, são dados por:



Você encontrará a seguir as principais propriedades da distribuição normal.

1) É simétrica em relação ao ponto $x = \mu$ (50% abaixo e 50% acima da média).

2) Tem forma campanular (sino).

3) As três medidas de posição, média, mediana e moda se confundem no ponto de máximo da curva ($x = \mu$).

4) Fica perfeitamente definida conhecendo-se a média e o desvio-padrão, pois outros termos da função são constantes.

5) Tem dois pontos de inflexão em $x = \mu \pm \sigma$.

6) É assintótica em relação ao eixo das abscissas $\lim_{x \to \pm\infty} f(x) = 0$.

7) A área compreendida entre a curva e eixo x é igual a $1 \int_{-\infty}^{\infty} f(x)dx = 1$.

Condição para ser uma função densidade de probabilidade.

Portanto, a área sob a curva entre os pontos a e b, em que a < b, representa a probabilidade da variável X assumir um valor entre a e b.

$$P(a < x < b) = \int_a^b f(x)dx$$

Conceitos vistos em Matemática e em variáveis aleatórias.

Deste modo, a probabilidade de um ponto qualquer é nula:

$$P(x = a) = \int_a^a f(x)dx = 0$$

A notação utilizada para a distribuição normal será a apresentada a seguir:

A variável X tem distribuição normal com média m e desvio s.

$$X \sim N(\mu, \sigma)$$

Verifica-se que a probabilidade em um ponto é zero, pois de um ponto a ele mesmo não existe área, e como nas distribuições contínuas a área entre a função e o eixo das abscissas corresponde à probabilidade.

Como você pode notar, o cálculo de probabilidades via distribuição normal envolve a solução de integrais que não são nada triviais. Em virtude da grande aplicação da distribuição normal, procurou-se tabelar os valores de probabilidade, que seriam obtidos por meio da integração da função densidade de probabilidade normal num determinado intervalo.

A dificuldade para processar esse tabelamento se prendeu na infinidade de valores que μ (média) e μ (desvio padrão) poderiam assumir. Nestas condições, teria que se dispor de uma tabela para cada uma das infinitas combinações de μ e σ, ou seja, em cada situação que se quisesse calcular uma probabilidade.

Para resolver este problema, podemos obter uma nova forma para a distribuição normal, que não seja influenciada por μ e σ. O problema foi solucionado mediante o emprego de uma nova variável, definida por $z = \frac{x - \mu}{\sigma}$, que transforma todas as distribuições normais em uma distribuição normal reduzida, ou padronizada, de média zero e desvio-padrão um, z ~ N(0,1).

Assim, utilizamos apenas uma tabela para o cálculo de probabilidades, para qualquer que seja a curva correspondente a uma distribuição normal.

Portanto, para um valor de x =   numa distribuição normal qualquer, corresponde o valor:

$$z = \frac{x - \mu}{\sigma} = \frac{\mu - \mu}{\sigma} = 0, \text{ na distribuição normal reduzida.}$$

Para $x = \mu + \sigma$, tem-se

$$z = \frac{x - \mu}{\sigma} = \frac{\mu + \sigma - \mu}{\sigma} = \frac{\sigma}{\sigma} = 1, \text{ e assim por diante.}$$

Então, podemos definir a distribuição normal reduzida ou padronizada como sendo uma distribuição da variável Z que apresenta distribuição normal com média zero e variância um (Z ~ N (0;1)).

A Figura da distribuição normal padronizada é apresentada a seguir:

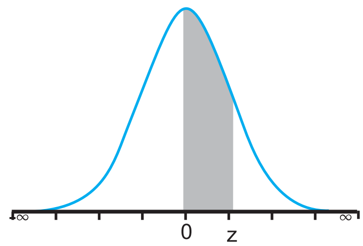


Figura 5: Área sob a curva normal padronizada compreendida entre os valores 0 e Z
Fonte: elaborado pelos autores

Z	0	1	2	3	4	5	6	7	8	9
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

Veja que na tabela da distribuição normal, os valores apresentados na primeira coluna correspondem à parte inteira e decimal do valor de Z, enquanto os valores da primeira linha correspondem à parte centesimal. Já os valores encontrados no meio da tabela correspondem às probabilidades dos respectivos valores compreendidos entre zero e Z.

Para que você possa entender a utilização da distribuição normal, vamos considerar a situação em que se estudou a durabilidade de um certo tipo de pneu. Verificou-se que esta durabilidade seguia uma distribuição normal com duração média 60.000 km e desvio-padrão 10.000 km. Procurou-se, então, responder os seguintes questionamentos:

a) Qual a probabilidade de um pneu aleatoriamente escolhido durar mais de 75.000 km?

X~ N(60000;100002) e procura-se calcular a P(X > 75000) = ?

$$z = \frac{x-\mu}{\sigma} = \frac{75000-60000}{10000} = 1{,}50$$

P(X > 75000) = P(z > **1,50**)

Olhando este valor na tabela de z, encontraremos no meio da tabela o valor de 0,4332, que corresponde à probabilidade de z estar entre zero e 1,5.

= 0,5 - P(0 < z < 1,50) = 0,4332 =

= 0,5 **-0,4332**

Retirou-se a probabilidade encontrada de 0,5, pois este valor corresponde à probabilidade de zero até o infinito.

= 0,0668

b) Qual a probabilidade de um pneu aleatoriamente escolhido durar entre 50.000 e 70.000 km?

P(50000 < X < **70000**) = ?

Como sugestão, faça o desenho da curva com os valores de x e depois transforme para os valores de z.

$$z_1 = \frac{x-\mu}{\sigma} = \frac{50000-60000}{10000} = -1{,}00$$

$$z_2 = \frac{x-\mu}{\sigma} = \frac{70000-60000}{10000} = 1{,}00$$

P(50000 < X < 70000) = P(-1,00 < z < 1,00) = 0,3413 + 0,3413 = 0,6826

c) Qual a probabilidade de um pneu aleatoriamente escolhido durar entre 63.000 e 70.000 km?

P(63.000 < X < 70.000) = ?

$$z_1 = \frac{x-\mu}{\sigma} = \frac{63.000-60.000}{10.000} = 0{,}30$$

$$z_2 = \frac{x-\mu}{\sigma} = \frac{70.000 - 60.000}{10.000} = 1{,}00$$

P(63000 < X < 70000) = P(0,30 < z < 1,00) = 0,3413 + 0,1179 = 0,2234

d) Qual a probabilidade de um pneu aleatoriamente escolhido durar exatamente 70.000 km?

P (X = 70000) = 0

e) O fabricante deseja fixar prazo de garantia, em quilômetros, de tal modo que, se a duração do pneu for inferior à garantia, o pneu seja trocado. De quantos quilômetros deve ser este prazo, para que somente 1% dos pneus sejam trocados?

x // P(X < x) = 0,01 ==> z // P(Z < z) = 0,01

z = -2,33

$$z = \frac{x-\mu}{\sigma}$$

$$-2{,}33 = \frac{x - 60000}{10000}$$

$$x = 36700\,\text{Km}$$

Resolva a seguir os exercícios relativos à distribuição normal e confira os resultados no final do livro.

Exercício 11: as rendas mensais dos graduados em um curso de especialização em uma grande empresa são normalmente distribuídas com uma média de R$ 2.000 e um desvio-padrão de R$ 200. Qual é o valor de Z para uma renda X de R$ 2.200? R$ 1.700?

Exercício 12: o uso diário de água por pessoa em uma determinada cidade é normalmente distribuído com média μ igual a 20 litros e desvio-padrão σ igual a 5 litros.

a) Que percentagem da população usa entre 20 e 24 litros por dia?

b) Que percentagem usa entre 16 e 20 litros?

c) Qual é a probabilidade de que uma pessoa selecionada ao acaso use mais do que 28 litros?

Exercício 13: um consultor verificou que as médias obtidas em uma avaliação após um treinamento tem distribuição normal com uma média igual a 72 e desvio-padrão 5. Ele decide atribuir conceitos para o seu treinamento tal que os melhores 15 % recebem conceito A. Qual é a média mínima que o funcionário submetido ao treinamento precisa receber para obter um conceito A?

Saiba mais...

- Mais exercícios referentes ao assunto estão no site: http://www.famat.ufu.br/prof/marcelo/exercicios.htm

UNIDADE

3

Amostragem

Amostragem

A amostragem (processo de retirada de amostras de uma população) é uma das etapas fundamentais na tomada de decisões nos diversos níveis gerenciais, pois uma amostragem mal executada, com certeza, resultará em estatísticas pouco confiáveis e em uma tomada de decisão possivelmente imprecisa.

Esta Unidade tem como objetivo apresentar a você alguns conceitos e definições essenciais para conduzir convenientemente uma operação de amostragem, visando principalmente à coleta de dados socioeconômicos.

Não são contemplados todos os aspectos referentes da amostragem, como por exemplo, o aspecto teórico das várias técnicas disponíveis. Você verá os aspectos mais importantes da amostragem e de maior aplicabilidade dentro das várias áreas da Administração.

Primeiramente, torna-se necessário definirmos população e amostra. Se considerarmos todos os atuais clientes de uma empresa da área de telefonia, podemos considerar estas pessoas como sendo a **população*** que caracteriza os clientes da empresa de telefonia, pois a população apresenta características em comum, sendo, neste caso, o fato de utilizarem esta empresa de telefonia.

Se quisermos utilizar uma parte desta população, que apresente as mesmas características daquela teremos uma **amostra**, ou seja, uma porção ou fração da população que preserva todas as características importantes dos elementos que a integram.

Se considerarmos esta população de clientes, você pode determinar o tempo médio em que o cliente fica utilizando no dia o aparelho de telefone fixo (média populacional (μ), que corresponde geralmente a um valor desconhecido, chamado de **parâmetro***. Como você não vai medir toda a população, podemos obter uma amostra que represente esta população, e estudando a amostra, você terá condições de calcular a média amostral ($\overline{x}$), que corresponde ao **estimador***, e o resultado obtido (valor numérico) corresponderá à **estimativa**.

GLOSSÁRIO

*População– é o conjunto de elementos que apresentam uma ou mais características em comum.

*Parâmetro – é um valor desconhecido associado a uma característica da população.

*Estimador – é uma função (fórmula) que permite estimar o valor de um parâmetro (estimativa), baseando-se nas observações de uma amostra.

A amostragem é o estudo das relações existentes entre a amostra, a população de onde ela foi extraída e a forma como ocorre esta extração. É útil na avaliação de grandezas desconhecidas da população, freqüentemente denominadas **parâmetros**, com base no conhecimento de grandezas correspondentes das amostras, geralmente chamadas **estimativas** ou **estatísticas** (Teoria da Estimação). Também auxilia na verificação de diferenças observadas entre duas ou mais amostras (tratamentos), para você saber se estas diferenças são devidas a uma variação casual ou se são verdadeiramente relacionadas aos efeitos de tratamentos (Teoria da Decisão).

Portanto, a amostragem tem por objetivo principal determinar meios e métodos para estudar as populações através de amostras. Observe que, quando obtemos informações a partir das amostras e tentamos atingir as populações, estamos realizando uma inferência.

Em resumo, podemos dizer que **amostra** é um subconjunto da população, necessariamente finito, pois todos os seus elementos serão examinados para efeito da realização do estudo estatístico desejado. Se considerarmos, nesta população de clientes, o tempo de utilização diária do telefone fixo, teremos uma variável aleatória, cuja média populacional μ corresponde ao parâmetro média do tempo de utilização diária do telefone fixo de todos os clientes atuais da empresa telefônica. Em geral, a população é muito grande, e a média populacional μ é estimada por meio de uma amostra retirada desta população, pelo cálculo da média amostral $\overline{x}$, que corresponde ao **estimador**, e o resultado obtido (valor numérico) corresponderá à **estimativa**. O problema de estimação de parâmetros é um dos importantes tópicos da estatística inferencial e será estudado posteriormente.

A utilização da amostragem ocorre, geralmente, quando queremos avaliar populações muito grandes ou infinitas.

As principais vantagens da utilização do estudo **por amostras representativas** (aquelas que mantêm as características da população de onde a amostra foi retirada) em relação ao **censo (avaliação de toda a população)** são:

- ocorre uma redução no custo, pois sendo os dados obtidos apenas de uma fração da população, as despesas são menores do que as oriundas de um censo. Tratando-se de grandes populações, podem-se obter resultados suficientemente precisos, para serem úteis, de amostras que representam apenas uma pequena fração da população;
- na prática ou no dia-a-dia das organizações, é necessário que os resultados sejam obtidos com a maior rapidez possível. Portanto, com a amostragem, você pode apurar os dados e sintetizá-los mais rapidamente do que em uma contagem completa. Este é um fator primordial, quando se necessita urgentemente das informações. Se o resultado de uma pesquisa for conhecido muito tempo depois, é bem possível que a situação que você pretendia resolver, seja, nesse momento, completamente diferente da que existia no momento da coleta dos dados;
- outra vantagem corresponde a uma maior amplitude e flexibilidade. Em certos tipos de investigação, como pesquisas de mercado, tem-se que utilizar pessoal bem treinado e equipamento altamente especializado, cuja disponibilidade é limitada para a obtenção de dados. O censo completo torna-se impraticável, e resta a escolha em obter as informações por meio de uma amostra. Portanto, com um número reduzido de entrevistadores, por exemplo, o treinamento a ser aplicado neles é de qualidade muito maior do que em um grupo maior de entrevistadores; e
- a última vantagem a ser citada aqui é a maior exatidão dos resultados. Em virtude de se poder empregar pessoal de melhor qualidade e intensivamente treinado, e por se tornar exeqüível a supervisão mais cuidadosa do campo de trabalho e do processamento de dados, favorecendo a uma redução no volume de trabalho, portanto, uma amostragem "pode", na realidade, proporcionar resultados mais exatos do que o censo.

Desta forma, podemos dizer que as amostras a serem trabalhadas devem apresentar uma característica importante, que corresponde à **representatividade***. Para que as conclusões da teoria de amostragem sejam válidas, as amostras devem ser escolhidas de modo

GLOSSÁRIO

*Representatividade – corresponde à possibilidade de manter as mesmas características presentes na população.

a serem representativas da população. Isso significa que a amostra deve possuir as mesmas características básicas da população, no que diz respeito à(s) variável(eis) que desejamos estudar. Desta forma, o plano de amostragem deve ser formulado para garantir esta representatividade.

Com um plano amostral apropriado, você considera que seja possível garantir a representatividade da amostra devido a um erro amostral?
Reflita em uma situação.

Uma vez tendo decidido realizar a pesquisa selecionando uma amostra da população, é preciso elaborar o plano de amostragem. O plano de amostragem consiste em definir as **unidades amostrais***, maneira pela qual a amostra será retirada (o tipo de amostragem), e o próprio tamanho da amostra.

GLOSSÁRIO
*Unidades amostrais – correspondem às unidades selecionadas na amostragem para calcular as estatísticas.

Estas unidades amostrais podem corresponder aos próprios elementos da população, quando há acesso direto a eles, ou qualquer outra unidade que possibilite chegar até eles. Você pode considerar como população os domicílios de uma cidade e que se deseje avaliar o perfil socioeconômico. A unidade amostral será cada um dos domicílios, que corresponderá aos elementos da população. Caso a unidade amostral for definida como os quarteirões, a unidade amostral não corresponderá aos elementos populacionais.

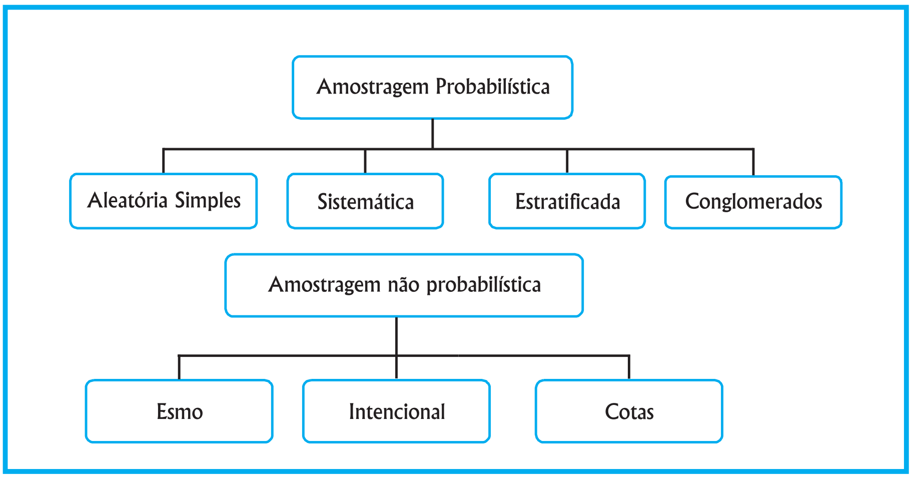
Podemos ter dois tipos de amostragem, as probabilísticas e as não probabilísticas, as quais serão definidas a seguir.

- **Amostragem probabilística**: quando todos os elementos da população tiveram uma probabilidade conhecida e diferente de zero de pertencer à amostra (ex: 50 funcionários em uma atividade de treinamento, e você deve selecionar dez funcionários). A realização deste tipo de amostragem só é possível se a população for finita e totalmente acessível.
- **Amostragem não probabilística**: quando não se conhece a probabilidade de um elemento da população pertencer à amos-

tra. Por exemplo, quando somos obrigados a colher a amostra na parte da população a que temos acesso.

Você pode notar que a utilização de uma amostra probabilística é melhor para garantir a representatividade da amostra, pois o acaso será o único responsável por eventuais discrepâncias entre população e amostra. Estas discrepâncias são levadas em consideração nas inferências estatísticas.

Os principais esquemas amostrais são apresentados a seguir.



Amostragem aleatória (casual) simples

Você deve utilizar a amostragem aleatória simples somente quando a população for homogênea em relação à variável que se deseja estudar. Geralmente, atribuímos uma numeração a cada indivíduo da população, e através de um **sorteio aleatório** os elementos que vão compor a amostra são selecionados. Todos os elementos da população têm a mesma probabilidade de pertencer à amostra.

Pode ser feito por meio da geração de um número aleatório.

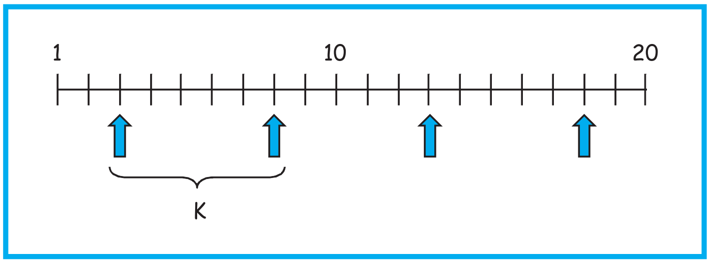
Imagine que você queira amostrar um número de pessoas que estão fazendo um determinado concurso com n inscritos. Como a população é finita, devemos enumerar cada um dos n candidatos e sortear n deles.

Amostragem sistemática

Em algumas situações, é conveniente retirar da população os elementos que vão compor a amostra de forma cíclica (em períodos), por exemplo, quando os elementos da população se apresentam ordenados. Porém, é de fundamental importância que a variável de interesse não apresente ciclos de variação coincidente com os ciclos de retirada, pois este fato tornará a amostragem não aleatória. Esta técnica de amostragem se chama amostragem sistemática. Para podermos entender melhor, vamos imaginar que você queira retirar uma amostra de currículos apresentados para um processo seletivo, e a variável de interesse corresponde à idade dos candidatos. Pode ocorrer que pessoas de uma determinada faixa etária deixem para entregar o currículo no último dia. Então, se pegássemos os currículos de forma aleatória, poderíamos estar subestimando ou superestimando a idade média. Nesta situação foram recebidos 500 currículos ordenados por ordem alfabética. Deseja-se amostrar 50 currículos para estimar a idade média dos candidatos. Será utilizada a técnica de amostragem sistemática, supondo que as idades estejam aleatoriamente distribuídas na população, ou seja, sem qualquer ciclo de repetição.

Primeiramente, deve-se enumerar a população de 1 a 500 e calcular uma constante (K) que servirá como fator de ciclo para retirada dos currículos amostrados. Então, podemos dividir os 500 currículos pelo tamanho da amostra (50) que se deseja trabalhar. Teremos uma constante igual a 10, e os elementos serão amostrados a cada dez elementos. Generalizando, então, teremos que a constante (K) será dado por $K = N/n$, onde N é o tamanho da população, e n, o tamanho da amostra.

Após a definição do valor de K, sorteia-se o ponto inicial da amostragem, ou seja, um dos elementos do primeiro intervalo constituído pelos elementos populacionais numerados de 1 até 10. Escolhe-se o seguinte, que será o elemento de ordem $(i + K)$; e assim por diante, sempre somando-se K à ordem do elemento anterior, até completar a escolha dos n elementos que vão compor a amostra. Um esquema é apresentado a seguir.



Para fixar os conceitos de amostragem sistemática, faça um esquema de amostragem para saber a opinião dos usuários de um banco em relação ao tempo de atendimento.

O banco possui uma listagem de 33.400 clientes em uma determinada cidade. A pesquisa será feita por telefone, utilizando uma estrutura de call center. Deseja-se trabalhar com uma amostra de 300 clientes. Como seria organizada a amostragem sistemática?

Amostragem Estratificada

Quando a variável de interesse apresenta uma heterogeneidade na população e esta heterogeneidade permite a identificação de grupos homogêneos, você pode dividir a população em grupos (estratos) e fazer uma amostragem dentro de cada estrato, garantindo, assim, a representatividade de cada estrato na amostra.

Podemos verificar que pesquisas eleitorais apresentam uma grande heterogeneidade em relação à intenção de votos, quando consideramos, por exemplo, a faixa salarial ou o nível de escolaridade. Então, se fizéssemos uma amostragem aleatória simples, poderíamos incluir na amostra uma maior quantidade de elementos de um grupo, e, proporcionalmente, este grupo é pequeno em relação à população. Desta forma, não teríamos uma amostra representativa da população a ser estudada. Então, podemos dividir a população em grupos (estratos) que são homogêneos para a característica que estamos avaliando, ou seja, neste caso, a intenção de votos.

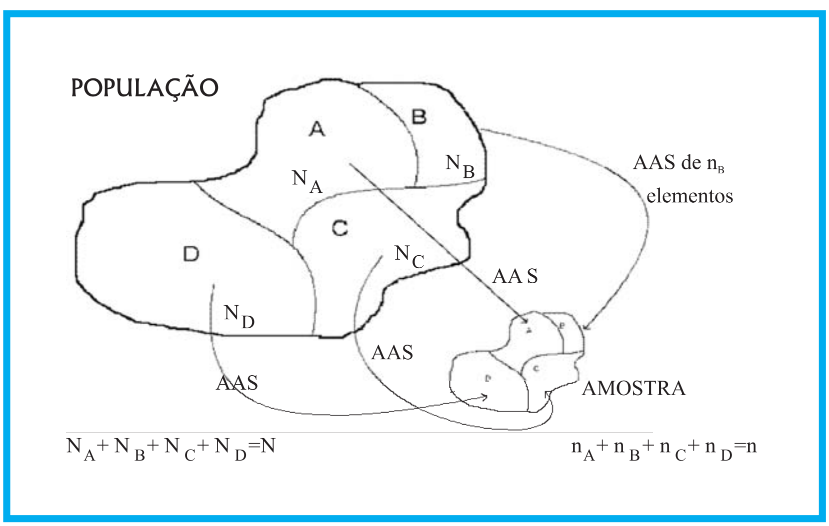
Como estamos dividindo a população em estratos (grupos) que são homogêneos dentro de si, podemos, então, caracterizar a amostragem estratificada. Para efetuarmos a amostragem estratificada de forma proporcional, precisamos primeiramente definir a **proporção do estrato em relação à população**.

Proporção do estrato h será igual ao número de elementos presentes neste estrato (N_h) dividido pelo tamanho da população (N) ➔ (N_h/N).

Após você obter esta proporção do estrato em relação à população, deve-se multiplicar o tamanho total da amostra (n) pela proporção de cada estrato na população (N_h/N).

Assim, teremos um tamanho de amostra em cada estrato, proporcional ao tamanho do estrato em relação à população.

A Figura 11 mostra como é feita a escolha dos elementos de cada estrato (A, B, C, D) que você pode fazer usando amostragem aleatória simples devido ao fato de os estratos serem homogêneos individualmente, considerando a variável de interesse.



Para que você possa fixar os conceitos de amostragem estratificada, resolva a seguinte situação.

Exercício 1: uma franquia de *fast food* com foco em sanduíches, apresenta lojas em todo o mundo. Para fazer uma pesquisa de satisfação dos clientes, dividiu-se a população de lojas em três estratos (países desenvolvidos, países em desenvolvimento e países do grupo asiático). Pretende-se trabalhar com uma amostra de tamanho n = 200. Com as informações a seguir, faça o esquema de uma amostragem estratificada.

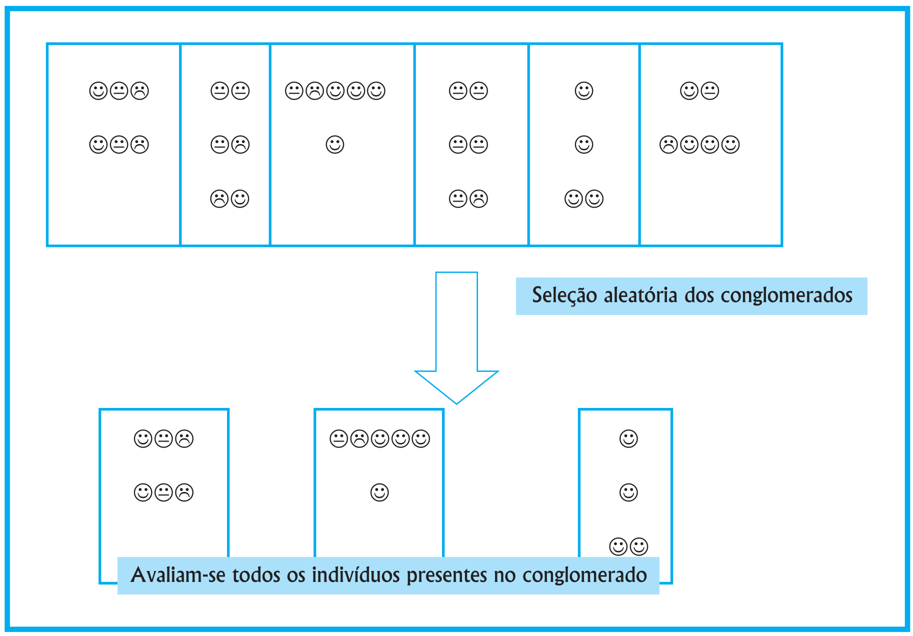
Estratos	Tamanho do estrato (no de lojas)
Países desenvolvidos	$N_1 = 700$
Países em desenvolvimento	$N_2 = 420$
Países do grupo asiático	$N_3 = 270$

Amostragem por conglomerados

Apesar de a amostragem estratificada apresentar resultados satisfatórios, a sua implementação é dificultada pela falta de informações sobre a população para fazer a estratificação. Para poder contornar este problema, você pode trabalhar com o esquema de amostragem chamado amostragem por conglomerados.

Os conglomerados são definidos em função da experiência do gestor ou pesquisador. Geralmente, podemos definir os conglomerados por fatores geográficos, como por exemplo, bairros e quarteirões. A utilização da amostragem por conglomerados possibilita uma redução significativa do custo do processo de amostragem. Portanto, um conglomerado é um subgrupo da população, que individualmente reproduz a população, ou seja, individualmente os elementos que o compõem são muito heterogêneos entre si. Este tipo de amostragem é muito útil quando a população é grande, por exemplo, no caso de uma pesquisa em nível nacional.

Para efetuarmos a amostragem por conglomerados, primeiramente definimos o conglomerado e assim dividimos a população nos conglomerados. Sorteamos os conglomerados por meio de um processo aleatório e avaliamos todos os indivíduos presentes no conglomerado, que é chamado de amostragem por conglomerados em um estágio. Caso façamos um sorteio de elementos dentro de cada conglomerado, teremos uma amostragem por conglomerados em dois estágios. Segue abaixo o esquema de uma amostragem por conglomerados em um único estágio. Cada quadrado corresponde a uma residência.



A Pesquisa Nacional por Amostra de Domicílios (PNAD) do IBGE é feita por conglomerados em três estágios.

Saiba mais...

- Sobre a Pesquisa Nacional por Amostra de Domicílios (PNAD), consulte o site www.ibge.com.br.

O cálculo do tamanho amostral será visto em conjunto com a parte de intervalos de confiança.

Amostragem Não Probabilística

Quando trabalhamos com a amostragem não probabilística, não conhecemos a priori a probabilidade que um elemento da população tem de pertencer à amostra. Neste caso, não é possível calcular o erro decorrente da generalização dos resultados das análises estatísticas da amostra para a população de onde a amostra foi retirada.

Utilizamos, geralmente, a amostragem não probabilística por simplicidade ou por impossibilidade de se obter uma amostra probabilística, como seria desejável.

Os principais tipos de amostragem não probabilística que temos são amostragem sem norma ou a esmo, intencional e por cotas.

Amostragem a esmo

Imagine uma caixa com 1.000 parafusos. A enumeração destes parafusos ficaria muito difícil, e a amostragem aleatória simples se torna inviável. Então, em situações deste tipo, supondo que a população de parafusos seja homogênea, escolhemos a esmo a quantidade relativa ao tamanho da amostra. Quanto mais homogênea for a população, mais podemos supor a equivalência com uma AAS.

Desta forma, os parafusos serão escolhidos para compor a amostra de um determinado tamanho sem nenhuma norma ou a esmo. Daí vem o nome deste tipo de amostragem.

Amostragem intencional

A amostragem intencional corresponde àquela em que o amostrador deliberadamente escolhe certos elementos para pertencer à amostra, por julgar tais elementos bem representativos da população. Um exemplo deste tipo de amostragem corresponde à situação em que se deseja saber a aceitação em relação a uma nova marca de whisky a ser inserida no mercado de uma cidade. Somente entrarão para compor a amostra pessoas que façam uso da bebida e que tenham condições financeiras de comprar esta nova marca (classe social de maior poder aquisitivo).

Amostragem por cotas

Neste tipo de amostragem, a população é dividida em grupos, e seleciona-se uma cota proporcional ao tamanho de cada grupo. Entretanto, dentro de cada grupo não é feito sorteio, e sim os elementos são procurados até que a cota de cada grupo seja cumprida. Em pesquisas eleitorais, a divisão de uma população em grupos (considerando, por exemplo, o sexo, o nível de escolaridade, a faixa etária e a renda) pode servir de base para a definição dos grupos, partindo da suposição de que estas variáveis definem grupos com comportamentos diferenciados no processo eleitoral. Para se ter uma idéia do tamanho destes grupos, pode-se recorrer a pesquisas feitas anteriormente pelo IBGE (Instituto Brasileiro de Geografia e Estatística).

Distribuições amostrais

Com as distribuições amostrais, você pode inferir propriedades de um agregado maior (a população) a partir de um conjunto menor (a amostra), ou seja, inferir sobre parâmetros populacionais, dispondo apenas de estatísticas amostrais.

Portanto, torna-se necessário um estudo detalhado das distribuições amostrais, que são base para intervalos de confiança e testes de hipóteses.

Portanto, para que você tenha condições de fazer afirmações sobre um determinado parâmetro populacional (ex: μ), baseadas na estimativa $\overline{x}$, obtido a partir dos dados amostrais, é necessário conhecer a relação existente entre $\overline{x}$ e μ, isto é, o comportamento de $\overline{x}$, quando se extraem todas as amostras possíveis da população, ou seja, sua distribuição amostral.

Para obtermos a distribuição amostral de um estimador, é necessário conhecer o processo pelo qual as amostras foram retiradas, isto é, se amostras foram retiradas **com reposição** ou **sem reposição**.

Portanto, a partir do comportamento da estatística amostral, pode-se aplicar um teorema muito conhecido na estatística como Teorema do Limite Central. Este teorema propõe que, se retirarmos todas as possíveis amostras de tamanho n de uma população independente de sua distribuição, e verificarmos como as estatísticas amostrais obtidas se distribuem, teremos uma distribuição aproximadamente normal, com $\mu_{\bar{x}} = \mu$ **(média das médias amostrais igual à média populacional)** e variância das médias $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$ **(variância das médias mostrais igual à variância da população dividida pelo tamanho da amostra)**, se a amostragem for realizada com reposição, ou $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$, se a amostragem for realizada sem reposição em uma população finita ($n/N > 0{,}05$), independentemente da distribuição da variável em questão.

Portanto, considerando a distribuição amostral de médias, quando se conhece a variância ou a amostra é grande (n > 30), utilizamos a estatística z da distribuição normal vista anteriormente, independente da distribuição da população. Então, por meio do teorema do limite central, a estatística será dada por: $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$.

Para saber mais

Considere uma população formada pelos números {1 , 2, 3}. Sabemos que esta população apresenta $\mu = 2$ e variância $\sigma^2 = 2/3$. Retire todas as amostras possíveis com n = 2, fazendo com e sem reposição e calcule a média das médias amostrais (μ_2) e a variância das médias amostrais ($\sigma^2_{\bar{x}}$). Compare com os resultados da população e veja se o teorema é verdadeiro. Pesquise este problema em sites da internet ou outros livros de Estatística.

Porém, ocorre que, na prática, muitas das vezes não se conhece σ^2 e trabalha-se com amostras pequenas, ou seja, menores ou iguais a 30. Assim, você conhece apenas sua estimativa s (desvio-padrão amostral). Substituindo σ por seu estimador s, na expressão da variável padronizada, obtém-se a variável:

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$ **(expressão semelhante a Z)**

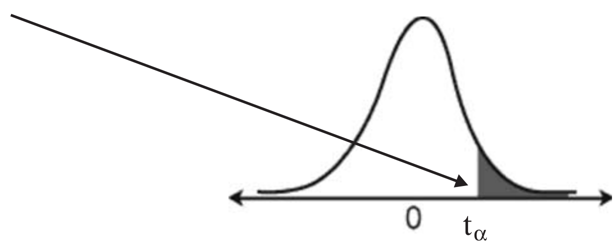
Corresponde ao divisor do cálculo da variância amostral, ou seja, n - 1. Número de variáveis na amostra que variam livremente, na definição da Estatística.

a qual segue uma distribuição t de Student com (n-1) **graus de liberdade.**

A distribuição t apresenta as seguintes características:

- é simétrica em relação à média, que é zero;
- tem forma campanular (semelhante à normal);
- quando n tende para infinito, a distribuição t tende para a distribuição normal, na prática, a aproximação é considerada boa quando n >30; e
- possui n-1 graus de liberdade.

Vamos aprender a utilizar a Tabela da distribuição de t de Student. Na Tabela t de Student, na primeira linha temos o valor de α, **que corresponde à probabilidade (área) acima de um determinado valor da tabela**. Na figura a seguir, temos o conceito de α (área mais escura).



Observe que na Tabela de t (a seguir), temos na primeira coluna os graus de liberdade (GL) e no centro da tabela, teremos os valores da **estatística t de Student**. Na primeira linha temos os valores de α.

					α				
GL	0.250	0.200	0.150	0.100	0.050	0.025	0.010	0.005	0.001
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.656	318.289
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.328
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.894
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.261
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.160
240	0.676	0.843	1.039	1.285	1.651	1.970	2.342	2.596	3.125
480	0.675	0.842	1.038	1.283	1.648	1.965	2.334	2.586	3.107
700	0.675	0.842	1.037	1.283	1.647	1.963	2.332	2.583	3.102
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098

Tabela 8: Limites unilaterais da distribuição t de Student ao nível α de probabilidade

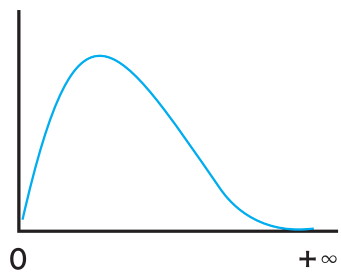
Fonte: www.dpi.inpe.br/~camilo/estatistica/06estimação.ppt

Para exemplificar o uso da tabela, consideremos que desejamos encontrar a probabilidade de ser maior do que um valor de t igual a 2,764, trabalhando com uma amostra de tamanho n = 11. Portanto, teremos 10 graus de liberdade e nesta linha procuramos o valor que desejamos encontrar, 2,764. Subindo na Tabela em direção ao α encontraremos um valor de 0,01 na primeira linha, ou seja, esta é a probabilidade de ser maior do que 2,764, com 10 graus de liberdade.

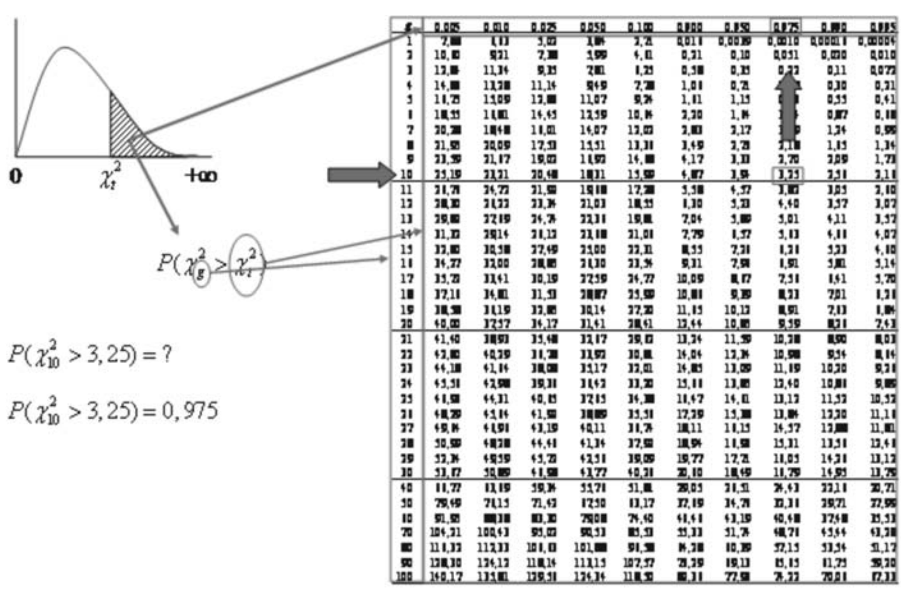
Retirando-se uma amostra de n elementos de uma população normal com média μ e variância α^2, então, pode-se demonstrar que a distribuição amostral da variância amostral segue uma **distribuição de χ2 (qui-quadrado)** com n-1 graus de liberdade. A variável da estatística de qui-quadrado será dada por:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \text{ tem distribuição } \boldsymbol{\chi}^2 \text{ com n-1 graus de liberdade.}$$

Esta distribuição é sempre positiva, o que pode ser comprovado pela própria definição da variável. Esta distribuição é assimétrica, como pode ser visto no gráfico da distribuição mostrado a seguir.



No esquema a seguir, temos como é feita a utilização da distribuição de qui-quadrado com g graus de liberdade.



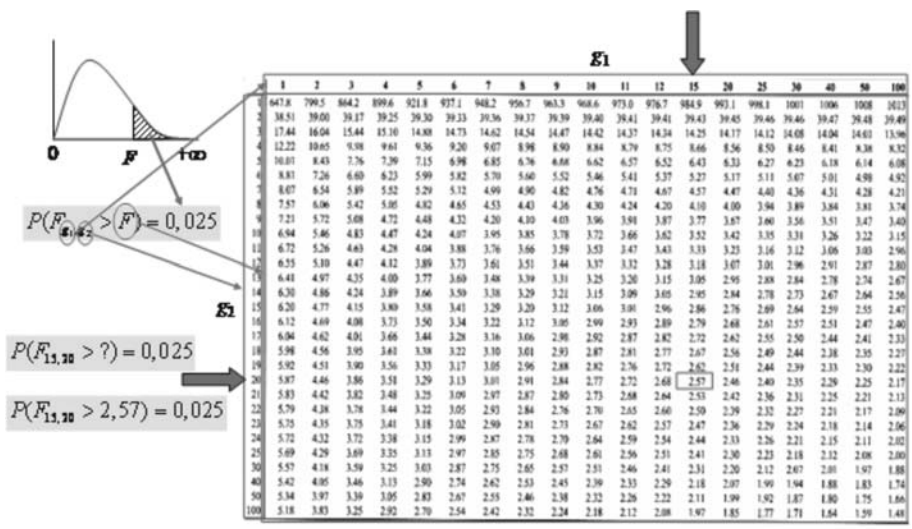
Fonte: www.dpi.inpe.br/~camilo/estatistica/06estimação.ppt

A **distribuição de F de Snedecor** corresponde à distribuição da razão de duas variâncias. Temos, então, duas populações que apresentam variâncias populacionais e delas são retiradas amostras, nas quais são calculadas variâncias amostrais. A relação entre essas variâncias é que nos dá a distribuição de F. A estatística da distribuição é apresentada a seguir:

$$F = \frac{s_A^2 / \sigma_A^2}{s_B^2 / \sigma_B^2}$$ segue uma distribuição F com $g_1 = n_1 - 1$ e $g_2 = n_2 - 1$

graus de liberdade para o numerador e denominador respectivamente.

A utilização da Tabela é apresentada a seguir:



Fonte: www.dpi.inpe.br/~camilo/estatistica/06estimação.ppt

Nota-se que, no caso da tabela de F, o valor de α que corresponde à área extrema à direita da curva, é apresentado no título da tabela, pois para cada valor de α temos uma tabela diferente.

Uma aplicação prática da distribuição de F está na verificação da homogeneidade das variâncias provenientes de duas populações normais e independentes.

A seguir, são apresentados exercícios, que possibilitam o treino de utilização das tabelas de t, χ^2 e F. Como sugestão, sempre faça o desenho da curva para melhor entender como utilizar a tabela.

Exercício 2: obter os seguintes valores da distribuição t de Student:

a) P (-2,160 < t < a) = 0,95 com 13 g.l;

b) P (a < t < 1,708) = 0,90 com 25 g.l;

c) P (t > a) = 0,05 com 20 g.l.

Exercício 3: obter os seguintes valores da distribuição de $\chi 2$:

a) P ($\chi 2$ > a) = 0,025 com 21 g.l;

b) P ($\chi 2$ < a) = 0,025 com 21 g.l;

c) P($\chi 2$ > a) = 0,95 com 15 g. l.

Exercício 4: obter os seguintes valores da distribuição F de Snedecor:

a) P(F > a) = 0,10 com g1 = 5 e g_2 = 25 g.l;

b) P(F < a) = 0,90 com n1 = 6 e n_2 = 26 g.l;

c) P(F > a) = 0,05 com g1 = 13 e g_2 = 29 g.l.

Estimação

Um dos principais objetivos da estatística inferencial consiste em estimar os valores de parâmetros populacionais desconhecidos (estimação de parâmetros) utilizando dados amostrais. Então, qualquer característica de uma população pode ser estimada a partir de uma amostra aleatória, desde que esta amostra represente bem a população. Os parâmetros populacionais mais comuns a serem estimados são a média, o desvio-padrão e a proporção. A estatística inferencial apresenta uma relevância alta, já que na maioria das decisões que um gestor ou pesquisador deve tomar, estão associadas à utilização de dados amostrais. Consiste em tirar conclusões de uma população a partir de amostra representativa dela, tendo uma grande importância em muitas áreas do conhecimento.

A partir de uma amostra de 800 clientes (escolhidos aleatoriamente entre todos os clientes que abasteceram na primeira quinzena de um determinado mês) de um posto de gasolina que possuem carros populares, verificou-se que o consumo médio de gasolina foi de R$ 200,00 por quinzena.

Reflita sobre a afirmação abaixo:
Então, podemos inferir que o consumo médio da população de clientes da primeira quinzena do mês em estudo, proprietários de carros populares que abastecem neste posto de gasolina é de R$ 200,00.

Esta é uma estimativa que chamamos de **pontual**, ou seja, inferimos sobre a população, considerando apenas o valor da estimativa. Essas estimativas por ponto não nos dão uma idéia sobre confiança e as margens de erro que deveriam ser aplicadas ao resultado. Tudo que nós sabemos, por exemplo, é que o consumo médio de gasolina foi estimado como R$ 200,00 por quinzena, independente do tamanho da amostra e da variabilidade inerente dos dados. Se fosse usado um tamanho grande de amostra e houvesse pouca variabilidade, teríamos grandes razões para acreditar no resultado. Mas não sabemos nada, se tivermos apenas uma estimativa por ponto. No entanto, podemos estimar ou fazer inferências sobre os valores da população usando uma segunda abordagem, chamada **estimativas por intervalos ou intervalos de confiança**, que dão o intervalo dentro do qual se espera que esteja o valor da população, com uma dada probabilidade ou um nível de confiança. Neste caso, poderíamos inferir, por exemplo, que o consumo de carros populares que abastecem no posto de gasolina está no intervalo de R$ 180,00 a R$ 220,00, e ainda afirmamos isto com, por exemplo, uma certeza de 95%.

Em resumo, podemos dizer que a estimativa pontual fornece uma estimativa única de um parâmetro e que a estimativa intervalar nos dá um intervalo de valores possíveis, no qual se admite que esteja o parâmetro populacional com uma probabilidade conhecida.

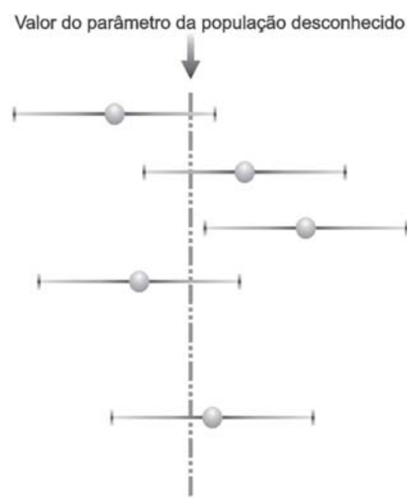
Como a estimativa por intervalos nos fornece uma informação mais precisa em relação ao parâmetro, esta é a melhor forma de estimar o parâmetro populacional. Então, para você estimar parâmetros populacionais por meio de dados amostrais, é necessário o conhecimento da distribuição amostral da estatística que está sendo usada como estimador (visto anteriormente).

Considere que, em uma loja em um shopping, você coletou uma amostra de clientes do mês anterior e verificou que eles apresentavam uma idade média de 24,2 anos. Surge, então, uma pergunta: este valor encontrado está próximo da média da população de clientes do mês anterior? A distribuição da média amostral, segundo o teorema central do limite apresenta uma distribuição aproximadamente normal (considerando um tamanho amostral suficientemente grande). Pelo que foi visto na Unidade 2 (distribuição normal), se considerarmos uma probabilidade de 99%, teremos que 99% médias amostrais estarão dentro do intervalo correspondente ao limite inferir de 2,57 desvios-padrão abaixo da média da variável, média amostral e ao limite superior de 2,57 desvios-padrão acima da média. Em função da dificuldade de encontrarmos um valor exato, então trabalhamos com intervalos de confiança. Podemos concluir que existe uma chance de 1% de que a média não esteja dentro deste afastamento da média de 2,57 desvios-padrão, e convertendo este afastamento na unidade dos dados, teremos o chamado intervalo de confiança.

*Então, um **intervalo de confiança** dá um intervalo de valores, centrado na estatística amostral, no qual julgamos, com um risco conhecido de erro, estar o parâmetro da população.*

Este 1% do exemplo anterior pode ser chamado de nível de significância ou α, que nos dá a medida da incerteza desta **inferência**. O α geralmente assume valores entre 1 e 10%.

Então, a partir de informações de amostras, devemos calcular os limites de um intervalo, valores críticos, que em (1-α)% dos casos inclua o valor do parâmetro a estimar e em α% dos casos não inclua o valor do parâmetro, como pode ser visto na Figura.



O nível de confiança 1 - α é a probabilidade de o intervalo de confiança conter o parâmetro estimado. Em termos de variável normal padrão Z, isto representa a área central sob a curva normal entre os pontos -Z e Z.

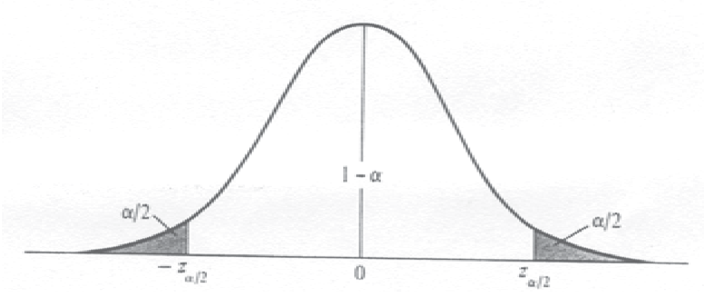


Figure 8.4 Location of $z_{\alpha/2}$ and $-z_{\alpha/2}$

Você pode observar que a área total sob a curva normal é unitária. Se a área central é 1 -α, o ponto -z representa o valor de Z, que deixa à sua esquerda a área $\alpha/2$, e o ponto z representa o valor de Z, que deixa à sua direita a área $\alpha/2$.

Vamos então aprender como construir alguns intervalos de confiança.

Intervalo de confiança para a média populacional quando o desvio-padrão populacional é conhecido.

Vamos imaginar a seguinte situação: o Departamento de Recursos Humanos de uma grande empresa informa que o tempo de execução de tarefas que envolvem participação manual varia de tarefa para tarefa, mas que o desvio-padrão permanece aproximadamente constante, em 3 minutos. Uma nova tarefa está sendo implantada na empresa. Uma amostra aleatória do tempo de execução de 50 destas novas tarefas forneceu o valor médio de 15 minutos. Determine um intervalo de confiança de 95% para o tempo médio de execução desta nova tarefa.

Primeiramente, você precisa identificar que o desvio-padrão populacional é conhecido, e também a amostra é considerada grande ($n > 30$). Então, a construção do intervalo de confiança será feita utilizando a média amostral, que é aproximadamente normal. Utilizaremos para a obtenção dos limites de confiança a curva normal padrão Z.

Como os limites são dados por meio da estatística calculada a partir dos dados amostrais e da margem de erro (fornecida pela estatística da distribuição multiplicada pelo desvio-padrão da distribuição amostral), teremos, nesta situação, os limites calculados por meio da seguinte expressão:

$$\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}}\text{, onde } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

Margem de erro
$(e = z_{\alpha/2}\sigma_{\bar{x}})$

Erro-padrão da média, obtido a partir do cálculo do desvio-padrão da média de uma amostra aleatória simples, retirada de uma população de tamanho muito maior do que o tamanho da amostra e de variância conhecida σ^2.

Logo, o intervalo de confiança tem centro na média amostral: calculando na nossa situação, teremos:

$1 - \alpha = 0{,}95 \rightarrow \alpha = 0{,}05 \rightarrow \alpha/2 = 0{,}025$

olhando na tabela de Z, você encontrará Z /2 = 1,96

$$e = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1{,}96.\frac{3}{\sqrt{50}} = 0{,}8315$$

$$P(\bar{x} - e < \mu < \bar{x} + e) = (1 - \alpha)$$

$$P(15 - 0{,}8315 < \mu < 15 + 0{,}8315) = 0{,}95$$

$$P(14{,}168 < \mu < 15{,}831) = 0{,}95$$

Interpretação do resultado: em cada grupo de 100 amostras retiradas de 50 clientes, espera-se que em 95 delas a média esteja dentro do intervalo de 14,168 a 15,831.

Fator de correção para população finita

Quando você estiver trabalhando com um tamanho de amostra que não seja tão pequeno em relação ao tamanho da população, e se a amostragem realizada for sem reposição, é necessário que se faça uma correção na estimativa do erro-padrão da distribuição amostral. O fator de correção para população finita, amostragem sem reposição, é dado por: $\frac{N-n}{N-1}$ onde N: tamanho da população e n: tamanho da amostra.

Uma regra prática para uso da correção é dada por meio da relação entre o tamanho da amostra e o tamanho da população. Então, se $\frac{n}{N} > 0{,}05$, devemos fazer a correção para população finita, e teremos o desvio-padrão igual a $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$.

Esta correção, quando necessária, deve ser utilizada para qualquer distribuição amostral.

Resolva agora o exercício proposto a seguir.

Exercício 5: considere, por exemplo, que as despesas mensais com alimentação das 1.000 cabeças de gado de uma fazenda são normalmente distribuídas com desvio-padrão de US$ 3,00. Uma amostra de cem bois revelou uma despesa média mensal de US$ 27,00. Determine o intervalo de confiança de 90% para a despesa média com alimentação dos bois desta fazenda.

Desenvolvendo a expressão do erro mostrada anteriormente, teremos o tamanho de amostra para estimação da média populacional, quando o desvio-padrão populacional é conhecido, como é mostrado a seguir:

$$e = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad \sqrt{n} = z_{\alpha/2} \cdot \frac{\sigma}{e} \quad \Rightarrow \quad n = \left(\frac{z_{\alpha/2}.\sigma}{e}\right)^2 \quad \Rightarrow \quad n = \frac{(z_{\alpha/2})^2.\sigma^2}{e^2}$$

Se você verificar que a amostragem foi feita sem reposição em uma população finita, conforme dito anteriormente, devemos utilizar o fator de correção.

Imagine a seguinte situação: que tamanho de amostra será necessário para produzir um intervalo de 95% de confiança para a verdadeira média populacional, com erro de 1,0, se o desvio-padrão da população é 10,0?

$$n_o = \frac{Z_{\alpha/2}^2.\sigma^2}{e^2} = \frac{1{,}96^2.10^2}{1^2} = 384{,}16 \cong 385$$

Intervalo de confiança para a média populacional, quando o desvio-padrão populacional não é conhecido e a amostra é pequena (n < 30)

Nesta situação, não temos uma boa estimativa do desvio-padrão populacional, devido ao tamanho reduzido da amostra utilizado (n < 30). A única diferença em relação ao intervalo de confiança, quando o desvio-padrão é conhecido ou a amostra é grande (apresentado anteriormente), é que, no lugar da distribuição normal (Z) vamos trabalhar com a distribuição t de Student. Portanto, a expressão para o intervalo de confiança é apresentada a seguir:

$$\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Todas as demais considerações apresentadas anteriormente permanecem, bem como a interpretação do intervalo de confiança.

Considere que, para uma dada semana, foi tomada uma amostra aleatória de 28 empregados horistas selecionados de um grande número de funcionários de uma fábrica, a qual apresentou um salário médio de R$ 180,00 com um desvio-padrão de R$ 14,00. Estimar o salário médio para todos os empregados horistas da fábrica, de tal maneira que se tenha uma confiança de 95% de que o intervalo estimado inclua a média da população.

Calculando na nossa situação, teremos:

$$1- \alpha = 0{,}95 \Rightarrow \alpha = 0{,}05 \Rightarrow \alpha/2 = 0{,}025$$

Olhando na tabela de t, você encontrará $t_{\alpha/2}$ = 2,052 (com 27 graus de liberdade)

$$e = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 2{,}052 . \frac{14}{\sqrt{28}} = 5{,}42$$

$$P(\bar{x} - e < \mu < \bar{x} + e) = (1 - \alpha)$$

$$P(180 - 5{,}42 < \mu < 180 + 5{,}42) = 0{,}95$$

$$P(174{,}57 < \mu < 185{,}42) = 0{,}95$$

Intervalo de confiança para a estimação da proporção populacional

Você deve considerar que, geralmente, a proporção de sucessos em uma população (p), na maioria das vezes é desconhecida. Então, o que fazemos? Calculamos uma estimativa da proporção de sucessos na população a partir de uma amostra retirada desta, a qual denominamos $\hat{p}$. A distribuição amostral de uma proporção apresenta uma média p e desvio-padrão $\sqrt{\frac{pq}{n}}$.

Para construirmos o intervalo de confiança para p desconhecida, determinamos $\hat{p}$ na amostra e consideramos $\sigma_{\hat{p}} \cong \sqrt{\frac{\hat{p}.\hat{q}}{n}}$.

Portanto, considerando um nível de significância α, teremos:

a) intervalo de confiança para p: $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$;

b) margem de erro da estimativa: $e = z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$; e

c) tamanho da amostra: $n = \frac{(z_{\alpha/2})^2\,\hat{p}\,\hat{q}}{e^2}$.

Todas as demais considerações apresentadas anteriormente permanecem, bem como a interpretação do intervalo de confiança.

Um despachante que cuida da documentação de automóveis está interessado em estimar a proporção de clientes que trocaram de carro no último ano para oferecer seus serviços. Para isto, amostrou 80 clientes do seu cadastro e consultou-os por telefone, verificando que 30 deles teriam trocado de carro no último ano. Determine o tamanho da amostra necessário para estimar com 95% de confiança esta proporção com erro máximo de 4%.

$$\hat{p} = \frac{30}{80} = 0{,}375 \quad e \quad \hat{q} = 1 - \hat{p} = 1 - 0{,}375 = 0{,}625$$

$$n = \frac{Z^2_{\alpha/2} . \hat{p}\hat{q}}{e^2} = \frac{1{,}96^2 .0{,}375.0{,}625}{0{,}04^2} = 562{,}73 \cong 563$$

Veja esta outra situação agora: considere que uma empresa de pesquisa de mercado faz contato com uma amostra de cem homens em uma grande comunidade e verifica que uma proporção de 0,40 na amostra prefere lâminas de barbear fabricadas por seu cliente, em vez de qualquer outra marca. Determine o intervalo de confiança de 95% para a proporção de todos os homens na comunidade que preferem essa marca.

$$1 - \alpha = 0{,}95 \rightarrow \alpha = 0{,}05 \rightarrow \alpha/2 = 0{,}025$$

Olhando na tabela de Z, você encontrará Z /2 = 1,96

$$e = z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} = 1{,}96 \sqrt{\frac{0{,}40.0{,}60}{100}} = 0{,}096$$

$$P(\hat{p} - e < p < \hat{p} + e) = (1 - \alpha)$$

$$P(0{,}40 - 0{,}096 < p < 0{,}40 + 0{,}096) = 0{,}95$$

$$P(0{,}304 < p < 0{,}496) = 0{,}95$$

Saiba mais...

- Mais exercícios referentes ao assunto estão no site: http://www.famat.ufu.br/prof/marcelo/exercicios.htm

UNIDADE

4

Testes de Hipóteses

Testes de Hipóteses

Na teoria de decisão estatística, os testes de hipóteses assumem uma importância fundamental, já que estes permitem nos dizer, por exemplo, se duas populações são de fato iguais ou diferentes, utilizando para isso amostras destas populações. Desta forma, a tomada de decisão de um gestor, deve estar baseada na análise de dados a partir de um teste de hipótese.

Então, você pode definir as hipóteses a serem testadas, retirar as amostras das populações a serem estudadas, calcular as estatísticas delas e, por fim, determinar o grau de aceitação de hipóteses baseadas na teoria de decisão, ou seja, se uma determinada hipótese será validada ou não.

Para você decidir se uma hipótese é verdadeira ou falsa, ou seja, se ela deve ser aceita ou rejeitada, considerando uma determinada amostra, precisamos seguir uma série de passos. Os passos são mostrados a seguir.

1) Definir a hipótese de igualdade (H_0) e a hipótese alternativa (H_1) para tentar rejeitar H_0 (possíveis erros associados à tomada de decisão).

2) Definir o nível de significância (α).

3) Definir a distribuição amostral a ser utilizada.

4) Definir os limites da região de rejeição e aceitação.

5) Calcular a estatística da distribuição escolhida a partir dos valores amostrais obtidos e tomar a decisão.

Você deve tomar a decisão baseada na seguinte regra: se o valor da estatística da distribuição calculado estiver na região de rejeição, rejeitar, então, a hipótese nula, senão a decisão será que a hipótese nula não poderá ser rejeitada ao nível de significância determinada.

Diversos conceitos serão apresentados ao longo do detalhamento dos passos a serem seguidos na formulação de um teste de hipótese.

Detalhamento dos passos na formulação de um teste de **hipótese**:

1) **Formular as hipóteses (H_o e H_1).**

Primeiramente, vamos estabelecer as **hipóteses nula e alternativa**. Para exemplificar, você deve considerar um teste de hipótese para uma média. Então, a hipótese de igualdade é chamada de **hipótese de nulidade ou H_o**. Suponha que você queira testar a hipótese de que o tempo médio de ligações é igual a 50 segundos. Então, esta hipótese será simbolizada da maneira apresentada a seguir:

H^o: $\mu = 50$ (hipótese de nulidade)

Esta hipótese, na maioria dos casos, será de igualdade.

Se você rejeitar esta hipótese, vai aceitar, neste caso, outra hipótese, que chamamos de **hipótese alternativa**. Este tipo de hipótese é simbolizada por **H_1 ou H_a**.

As hipóteses alternativas mais comuns são as apresentadas a seguir a partir do nosso exemplo:

H1: $\mu > 50$ (teste unilateral ou unicaudal à direita)

O tempo médio de ligação é superior a 50 segundos

H1: $\mu < 50$ (teste unilateral ou unicaudal à esquerda)

O tempo médio de ligação é inferior a 50 segundos

H1: $\mu \neq 50$ (teste bilateral ou bicaudal)

O tempo médio de ligação pode ser superior ou inferior a 50 segundos.

Surge uma dúvida. Qual hipótese alternativa você utilizará? A resposta é bem simples.

A hipótese alternativa será definida por você, em função do tipo de decisão que deseje tomar.

Veja o seguinte exemplo: você inspeciona uma amostra de uma grande remessa, encontrando-se 8% defeituosa. O fornecedor garante que não haverá mais de 6% de peças defeituosas em cada remessa. O que devemos responder, com auxílio dos testes de significância, é se a afirmação do fornecedor é verdadeira.

As hipóteses que você vai formular são:

H_0: p=0,06; H_1: p>0,06.

A hipótese alternativa só pode ser maior, pois o fornecedor garante que não haverá mais de 6%.

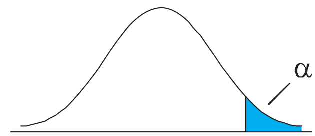
É importante ressaltar que o sinal de igual para a hipótese H_o corresponde a um sinal de menor ou igual (neste exemplo), pois o teste é unilateral à direita (p1 > 0,06). Portanto, sempre que o teste for unilateral, deve ser feita esta consideração.

2) Definir o nível de significância.

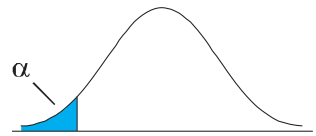
O nível de significância de um teste é dado pela probabilidade de se cometer erro do tipo I (**ocorre quando você rejeita a hipótese Ho e esta hipótese é verdadeira**). Com o valor desta probabilidade fixada, você pode determinar o chamado **valor crítico**, que separa a chamada **região de rejeição** da hipótese H_o da região de aceitação da hipótese H_o.

Na Figura abaixo, as áreas escuras correspondem à significância do teste, ou seja, à probabilidade de se cometer o chamado erro tipo I (rejeitar H_o quando ela é verdadeira). Esta probabilidade é chamada de α, e geralmente os valores mais utilizados são 0,01 e 0,05. O complementar do nível de significância é chamado de nível de confiança e é dado por $1 - \alpha$.

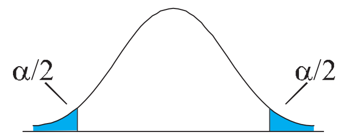
Unilateral à direita:
Ho: $\mu = 50$
H1:: $\mu > 50$



Unilateral à esquerda:
Ho: : $\mu = 50$
H1: : $\mu < 50$



Bilateral:
Ho: : $\mu = 50$
H1:: $\mu \neq 50$

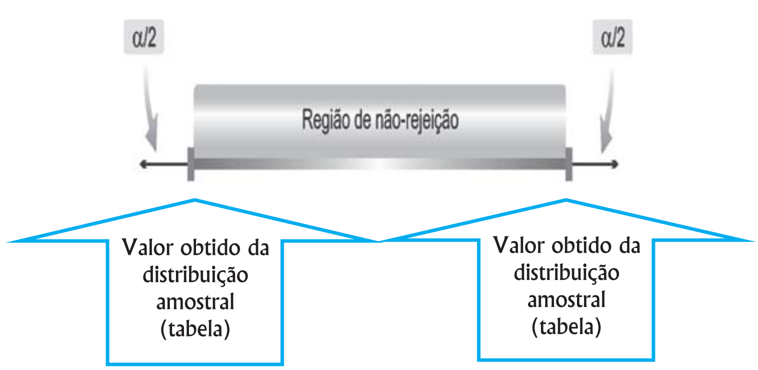


3) Definir a distribuição amostral a ser utilizada.

A estatística a ser utilizada no teste, você definira em função da distribuição amostral a qual os dados seguem. Se você fizer um teste de hipótese para uma média ou diferença entre médias, utilize a distribuição de Z ou t de Student. Outro exemplo é se você quiser comparar a variância de duas populações, então deverá trabalhar com a distribuição F, ou seja, da razão de duas variâncias. Note que o conhecimento das distribuições amostrais vistas na Unidade 3 é muito importante.

4) Definir os limites da região de rejeição.

Os limites entre as regiões de rejeição e aceitação da hipótese H_o, você definirá em função do tipo de hipótese H_1, do valor de (nível de significância) e da distribuição amostral utilizada. Considerando um teste bilateral, você terá a região de aceitação (não-rejeição) com uma probabilidade de 1- α e uma região de rejeição com probabilidade α ($\alpha/2$ + $\alpha/2$).



Através da amostra obtida, você deve calcular a estimativa que servirá para aceitar ou rejeitar a hipótese nula.

5) Tomar a decisão.

Para tomar a decisão, você deve calcular a estimativa do teste estatístico que será utilizado para rejeitar ou não a hipótese H_o. A estrutura deste cálculo para a média de forma generalista é dada por:

$$Estatistica \ \ da \ \ distribuição = \frac{(estimativa \ - \ parâmetro)}{erro \ \ padrão \ da \ \ estimativa}$$

Podemos exemplificar pela distribuição de Z, que será:

$$Z_{cal} = \frac{(\bar{X} - \mu)}{(\sigma / \sqrt{n})}$$

Variabilidade das médias

Estatística do teste

Se o valor da estatística estiver na região crítica (de rejeição), rejeitar H_o; caso contrário, aceitar H_0. O esquema abaixo mostra bem a situação de decisão.

Teste de hipótese para média populacional

Quando você retira uma amostra de uma população e calcula a média desta amostra, é possível verificar se a afirmação sobre a média populacional é verdadeira. Para tanto, basta verificar se a estatística do teste estará na região de aceitação ou de rejeição da hipótese H_o.

Aqui você tem três situações distintas:

1ª) se o desvio-padrão da população é conhecido ou a amostra é considerada grande (n >30), a distribuição amostral a ser utilizada será da Normal ou Z e a estatística-teste que você utilizará será: $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$.

Onde $\bar{x}$: média amostral; μ: média populacional; σ: desvio-padrão populacional e n: tamanho da amostra.

2ª) agora, se você não conhecer o desvio-padrão populacional e a amostra for pequena (   ), então, a distribuição amostral a ser utilizada será a t de Student, e a estatística-teste será: $t = \dfrac{\bar{x} - \mu}{s / \sqrt{n}}$. Onde $\bar{x}$: média amostral; μ: média populacional; s : desvio-padrão amostral e n: tamanho da amostra.

***Uma observação importante**: quando trabalhamos com amostras grandes, ou seja, n > 30, a distribuição de Z e t de Student apresentam comportamentos próximos e valores da estatística próximos também.*

Veja uma situação utilizando o teste de hipótese para uma média usando Z. Registros dos últimos anos de funcionários de uma determinada empresa atestam que sua média num teste de QI foi 115, com um desvio-padrão de 20. Para saber se uma nova equipe de funcionários é típica desta empresa, retirou-se uma amostra aleatória de 50 funcionários desta nova equipe, encontrando-se média de 118. Com uma significância de 5%, teste a hipótese de que esta nova equipe apresente a mesma característica dos funcionários da empresa, com relação ao QI.

Resolução

$H_0 : \mu = 115$

$H_1 : \mu \neq 115$

Estatística a ser utilizada → Z

$\alpha = 0{,}05$

$$Z = \frac{\overline{X} - \mu_O}{\sigma / \sqrt{n}} = \frac{118 - 115}{20 / \sqrt{50}} = 1{,}06$$

Conclusão: Como o valor da estatística calculada está na região de aceitação, então você deve aceitar Ho como verdadeiro.

Agora você deve resolver o seguinte exercício:

Exercício 1: um fabricante afirma que seus pneus radiais suportam em média uma quilometragem com mais de 40.000 km. Para testar essa afirmação, um comprador selecionou uma amostra de 49 pneus. Os testes nessa amostra forneceram uma média de 43.000 km. Sabe-se que a quilometragem de todos os pneus tem desvio-padrão de 6.500 km. Se o comprador testar essa afirmação ao nível de significância de 5%, qual será sua conclusão?

Resposta no final do livro.

Veja agora uma situação aplicando o teste t de Student.

O tempo médio gasto para profissionais da área de Ciências Contábeis realizarem um determinado procedimento tem sido de 50 minutos. Um novo procedimento está sendo implementado. Neste novo procedimento, retirou-se uma amostra de 12 pessoas, com um tempo médio de 42 minutos e um desvio-padrão de 11,9 minutos. Teste a hipótese de que a média populacional no novo procedimento é menor do que 50.

Resolução

$H_0 : \mu = 50$

$H_1 : \mu < 50$

$\alpha = 0{,}05$

Estatística a ser utilizada → t

$$t = \frac{\overline{X} - \mu_O}{\sigma / \sqrt{n}} = \frac{42 - 50}{11{,}9 / \sqrt{12}} = -2{,}53$$

$$t_\alpha = t_{0,05} = -1{,}796$$

Conclusão: Como o valor da estatística calculada está na região de rejeição, então você deve rejeitar Ho como verdadeiro.

Teste de hipótese para a razão de duas variâncias.

Este teste de hipótese é utilizado para saber se duas variâncias populacionais são estatisticamente iguais ou se uma é maior do que a outra. Então, utilizando a distribuição F, poderemos formular o teste de hipótese da razão entre duas variâncias e chegar à conclusão baseados apenas nas estimativas calculadas a partir das amostras.

As hipóteses H_o e H_1 serão:

$$H_0 : \sigma^2{}_1 = \sigma^2{}_2$$
$$H_1 : \sigma^2{}_1 > \sigma^2{}_2$$

Como estamos utilizando um teste unilateral à direita (questões didáticas), então, no cálculo da estatística de F, teremos a maior variância dividida pela menor variância.

A maior variância amostral encontrada será chamada de S_1^2 (proveniente de uma amostra de tamanho n_1), e a menor variância amostral será chamada S_2^2 (proveniente de amostra de tamanho n_2).

Vamos supor que tivéssemos duas amostras provenientes de duas populações. Desejamos saber se as variâncias das populações são estatisticamente iguais ou uma é maior do que a outra. Considere uma significância de 2,5%. Os resultados amostrais são apresentados a seguir:

$S_1^2 = 0{,}5184$ com $n_1 = 14$

$S_2^2 = 0{,}2025$ com $n_2 = 21$

A estatística será dada por:

Então, a variável de teste do teste F será:

$$F = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2}$$

Como em H_o, estou considerando que as variâncias populacionais são iguais, então, na expressão acima as duas variâncias populacionais vão se cancelar. No nosso exemplo, teremos:

$$F = \frac{s_1^{\ 2}}{s_2^{\ 2}} = \frac{0{,}5184}{0{,}2025} = 2{,}56$$

O valor tabelado (crítico) da distribuição de F será obtido na tabela da distribuição com uma significância de 2,5%. Considerando como graus de liberdade iguais a 13 ($n_1 - 1$) para o numerador (v_1) e 20 ($n_2 - 1$) para o denominador (v_2), chegaremos ao seguinte resultado: valor tabelado igual a 2,637.

O valor calculado da estatística foi menor do que o tabelado, então, ele caiu na região de aceitação de H_o. Assim, aceitamos H_o e consideramos que a variância da população 1 é estatisticamente igual à variância da população 2, ou seja, não ocorre uma diferença entre elas.

Este teste servirá de base na escolha do próximo teste (diferença entre médias para amostras independentes), ou seja, escolher o tipo de teste a ser utilizado.

Teste de hipótese para a diferença entre médias populacionais

Quando queremos comparar a média de duas populações, retiramos amostras das duas, e estas amostras podem apresentar tamanhos diferentes. Vamos considerar as situações de amostras independentes (as populações não apresentam nenhuma relação entre si) e amostras dependentes (uma população sofre uma intervenção e avalia-se antes e depois da intervenção para saber se esta resultou em algum efeito).

1ª situação: amostras independentes e grandes (n>30).

2ª situação: amostras independentes e pequenas, mas que apresentam variâncias populacionais estatisticamente iguais.

3ª situação: amostras independentes e pequenas, mas que apresentam variâncias populacionais estatisticamente desiguais.

4ª situação: amostras dependentes.

Agora você vai estudar cada uma destas situações. Lembre-se que as considerações anteriores em relação aos passos para formulação dos testes de hipóteses permanecem as mesmas.

A grande diferença que você vai ver ocorre só na determinação das hipóteses a serem testadas. A hipótese H_o será:

$$H_o: \mu_1 - \mu_1 = d_0$$

Onde: μ_1: média da população 1 e μ_2: média da população 2.

Já $\mathbf{d_o}$ corresponde a uma diferença qualquer que você deseje testar. Geralmente, quando queremos saber se as médias das duas populações são estatisticamente iguais, utilizamos o valor de $\mathbf{d_o}$ igual a zero.

As hipóteses alternativas seguem a mesma linha de raciocínio. Abaixo temos um quadro que nos auxiliará a visualizar estas considerações.

H_o	H_1
$\mu_1 - \mu_2 = d_0$	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$

É importante ressaltar que, se as hipóteses alternativas forem unilaterais, o sinal da hipótese H_o será **menor ou igual, maior ou igual**, dependendo da hipótese alternativa, apesar de utilizarmos a notação de igual (conforme comentado anteriormente).

Todas as outras considerações em relação aos testes de hipótese permanecem as mesmas. Vamos, então, procurar entender cada situação para os testes de hipóteses para diferença entre médias.

1ª situação: amostras independentes e grandes (n>30).

Como estamos trabalhando aqui com amostras grandes, ou quando se conhecem os desvios-padrão populacionais, devemos trabalhar com a distribuição amostral de Z (raciocínio semelhante ao utilizado no teste de hipótese para uma média). Portanto, a estatística do teste será dada por:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}}$$

Onde: $\overline{x}_1$: média da amostra 1; $\overline{x}_2$: média da amostra 2 μ_1: média da população 1; μ_2: média da população 2; σ_1^2: variância da população 1; σ_2^2: variância da população 2; n_1: tamanho da amostra 1 e n_2 tamanho da amostra 2.

OBS: se trabalharmos com amostras grandes poderemos substituir as variâncias populacionais pelas variâncias amostrais.

Vamos, então, ver como podemos aplicar o teste de hipótese para a diferença entre médias nesta situação.

Foram retiradas amostras de aparelhos usados de duas marcas, e os resultados são apresentados na tabela a seguir. Verifique se as duas marcas têm uma mesma durabilidade ou se são diferentes, com uma significância de 0,05.

Marcas	A	B
Média	1.160	1.140
Desvio-padrão	90	80
tamanho amostra	100	100

Resolução

$H_0 : \mu_A - \mu_B = 0$

$H_1 : \mu_A - \mu_B \neq 0$

$\alpha = 0{,}05$ **Estatística a ser utilizada ➔ Z**

$$Z = \frac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} = \frac{(1160 - 1140) - (0)}{\sqrt{\frac{90^2}{100} + \frac{80^2}{100}}} = 1{,}67$$

$$Z_{\alpha/2} = Z_{0,025} = 1{,}96$$

Conclusão: Como o valor da estatística calculada está na região de aceitação, então você deve aceitar Ho como verdadeiro.

2ª situação: amostras independentes e pequenas, mas que apresentam variâncias populacionais estatisticamente iguais.

Como as amostras com que estamos trabalhando são pequenas, e as variâncias populacionais, desconhecidas, então, você deve trabalhar com a distribuição t de Student.

Aqui consideraremos que as variâncias populacionais são estatisticamente iguais, pois esta situação influenciará nos cálculos e, conseqüentemente, no processo decisório. Para saber se as variâncias podem ser consideradas iguais, deve-se fazer um teste da razão de duas variâncias (teste F) mostrado anteriormente.

A estatística do teste será dada por:

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{1/n_1 + 1/n_2}}$$

Aqui aparece um termo novo (Sp). Ele corresponde ao desvio-padrão ponderado pelos graus de liberdade, ou seja, calculamos um novo desvio-padrão, no qual o fator de ponderação corresponde ao grau de liberdade de cada amostra. Veja a seguir:

$$Sp = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Para você encontrar o valor tabelado que limita as regiões de aceitação e rejeição na tabela t de Student (revise na Unidade 3), o número de graus de liberdade (v) será dado por:

$$v = n_1 + n_2 - 2$$

Vamos agora resolver um exemplo.

Um treinamento na área contábil de um grupo empresarial é ministrado a 12 profissionais pelo método convencional. Um segundo grupo de dez profissionais recebeu o mesmo treinamento por um método programado. Os resultados de notas dos dois métodos são apresentados na tabela a seguir. Determine se há diferença entre os dois métodos considerando uma significância de 0,01.

Método	Convencional	Programado
Média	85	81
Desvio-padrão	4	5

OBS: no teste F, não foram encontradas diferenças entre as variâncias populacionais.

Resolução

$H_0: \mu_C - \mu_P = 0$

$H_1: \mu_C - \mu_P \neq 0$

$\alpha = 0{,}01$

Estatística a ser utilizada→ t

$$t = \frac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{s_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} = \frac{(85-81)-(0)}{4{,}478\sqrt{\frac{1}{12}+\frac{1}{10}}} = 2{,}07$$

v=22-2=20 gl

$t_{\alpha/2} = t_{0{,}005} = 2{,}845$

Conclusão: Como o valor da estatística calculada está na região de aceitação, então você deve aceitar Ho como verdadeiro.

Agora você deve resolver o seguinte exercício:

Exercício 2: duas técnicas de venda são aplicadas em dois grupos de vendedores. A técnica A foi aplicada em um grupo de 12 vendedores, resultando em um número de vendas efetivadas em média de 76 e uma variância de 50. Já a técnica B foi aplicada em um grupo de 15 vendedores, resultando em um número de vendas efetivadas em média de 68 e uma variância de 75. Considerando as variâncias estatisticamente iguais, e com uma significância de 0,05, verifique se as médias são estatisticamente iguais.

3ª situação: amostras independentes e pequenas, mas que apresentam variâncias populacionais estatisticamente desiguais. A diferença desta situação para a anterior é que você considera que as populações apresentam variâncias estatisticamente desiguais. Também utilizaremos a estatística do teste a partir da distribuição t de Student. A estatística-teste será dada por:

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_1^{\,2} / n_1 + s_2^{\,2} / n_2}}$$

Outra diferença esta no cálculo do número de graus de liberdade, pois nesta situação utilizaremos uma aproximação que é dada pela expressão a seguir:

$$v = gl = \frac{(s_1^{\,2} / n1 + s_2^{\,2} / n_2)^2}{\dfrac{(s_1^{\,2} / n_1)^2}{n_1 - 1} + \dfrac{(s_2^{\,2} / n_2)^2}{n_2 - 1}}$$

Se este valor calculado apresentar valores decimais, deve ser feito o arredondamento para um número inteiro. Vamos a um exemplo.

Para estudar o efeito da certificação ambiental no valor de empresas, consideraram-se amostras de empresas da mesma área, com e sem certificação ambiental. Obtiveram-se os seguintes resultados. Após ter sido testado, verificou-se que as populações apresentam variâncias desiguais. Teste a hipótese de que os dois padrões de empresas apresentam médias de valor diferentes.

Método	Com certificação ambiental	Semcertificação ambiental
Média	24,0	13,3
Desvio-padrão	1,7	2,7
N	8	21

Resolução

$H_0 : \mu_C - \mu_P = 0$

$H_1 : \mu_C - \mu_P \neq 0$

$\alpha = 0{,}01$ $\boxed{\textbf{Estatística a ser utilizada} \rightarrow \textbf{t}}$

$$t = \frac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{s_p \sqrt{\dfrac{1}{n_A} + \dfrac{1}{n_B}}} = \frac{(85 - 81) - (0)}{4{,}478 \sqrt{\dfrac{1}{12} + \dfrac{1}{10}}} = 2{,}07$$

v=22-2=20 gl

$$t_{\alpha/2} = t_{0,005} = 2{,}845$$

Conclusão: Como o valor da estatística calculada está na região de aceitação, então você deve aceitar Ho como verdadeiro.

Resolva agora este exercício:

Exercício 3: um empresário deseja saber se há futuros profissionais mais promissores em escolas de regiões pobres e de regiões ricas. Uma amostra de 16 estudantes de uma zona pobre resultou em um teste específico, uma média de 107 pontos e um desvio-padrão de 10 pontos. Já 14 estudantes de região rica apresentaram uma média de 112 pontos e um desvio-padrão de 8 pontos. Você deve verificar se a média dos pontos dos dois grupos é diferente ou igual, para que o empresário possa saber se ele pode investir em qualquer uma das áreas ou uma das áreas é mais promissora (primeiro, verifique se as variâncias são estatisticamente iguais ou diferentes).

4ª situação: amostras dependentes.

Relembrando, amostras dependentes ocorrem quando se faz uma intervenção e se deseja saber se os resultados antes da intervenção são iguais aos resultados depois da intervenção.

Um ponto importante nesta situação é que são calculadas primeiramente as diferenças de antes e depois. Esta diferença é chamada de d_i. Então, você pode ver que:

d_i = valor antes - valor depois

Com base nestas diferenças (d_i) você vai calcular a média ($\overline{D}$) e o desvio- padrão destas diferenças (SD)

$$\overline{D} = \frac{\sum_{i=1}^{n} d_i}{n} \quad \text{e} \quad S_D = \frac{\sum_{i=1}^{n} d_i^2 - \frac{\left(\sum_{i=1}^{n} d_i^2\right)}{n}}{n-1}$$

Veja que estas fórmulas são iguais às de cálculo da média e desvio-padrão apresentados na Unidade 1. Neste caso, no lugar da variável x, são utilizados os valores de d_i (diferenças).

Com estes valores, a estatística teste será dada por:

$$t = \frac{\overline{D} - d_O}{S_D / \sqrt{n}}$$

O valor de n corresponde ao número de diferenças calculadas, e o grau de liberdade para ser olhado na tabela t de Student será dado por n - 1.

Em um estudo, procurou-se investigar a não-eficácia de uma propaganda na percepção de clientes. O Quadro a seguir dá os resultados de pessoas selecionadas anteriormente. No nível de 5% de significância, teste a afirmação de que as percepções sensoriais são inferiores após a propaganda, ou seja, a propaganda não é eficaz. (Os valores se referem a antes e depois da propaganda; medidas em uma escala de zero a doze.)

Pessoa	A	B	C	D	E	F	G	H
Antes	6,6	6,5	9,0	10,3	11,3	8,1	6,3	11,6
Depois	6,8	2,4	7,4	8,5	8,1	6,1	3,4	2,0

Resolução

$H_0 : \mu_D = 0$

$H_1 : \mu_D < 0$

$\alpha = 0,05$

Estatística a ser utilizada ➔ t

$$t = \frac{\overline{D} - \mu_D}{s_D / \sqrt{n}} = \frac{2,9114 - 0}{3,036 / \sqrt{8}} = 2,71$$

$$t_\alpha = t_{0,05} = 1,895$$

Conclusão: Como o valor da estatística calculada está na região de rejeição, então você deve rejeitar Ho como verdadeiro.

Teste de hipótese para diferença entre proporções

Em diversas situações, o que nos interessa é saber se a proporção de sucessos (evento de interesse) em duas populações apresenta a mesma proporção ou não. Neste caso, os dados seguem uma distribuição de Bernoulli (vista na Unidade 2) com média p e variância pq. Portanto, a expressão da estatística-teste (no caso utilizaremos a distribuição de Z) será dada por:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$$

Nesta expressão, você tem:

> $\hat{p}_1$ e $\hat{p}_2$ correspondem à proporção de sucesso nas amostras 1 e 2, respectivamente.
>
> p_1 e p_2 correspondem à proporção de sucesso nas populações 1 e 2, respectivamente.

Você deve se lembrar que a proporção de fracasso (q) é dada por um, menos a proporção de sucesso.

Vejamos, então, como aplicar o teste da diferença de proporções.

Uma questão de teste é considerada boa, se permitir discriminar entre estudantes preparados e estudantes não preparados. A primeira questão de um teste foi respondida corretamente por 62, dentre 80 alunos preparados, e por 23, dentre 50 alunos não preparados. Com um nível de 5% de significância, teste a afirmação de que esta questão foi respondida corretamente por uma proporção maior de estudantes preparados.

Resolução

$H_0: p_1 - p_2 = 0$

$H_1: p_1 - p_2 > 0$

$\alpha = 0{,}05$

Estatística a ser utilizada ➔ Z

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{(\hat{p}_1\hat{q}_1 / n_1) + (\hat{p}_2\hat{q}_2 / n_2)}} = \frac{(0{,}775 - 0{,}46) - (0)}{\sqrt{(0{,}775 . 0{,}225 / 80) + (0{,}46 . 0{,}54 / 50)}} = 3{,}73$$

$$Z_\alpha = Z_{0,05} = 1{,}65$$

Conclusão: Como o valor da estatística calculada está na região de rejeição, então você deve rejeitarHo como verdadeiro.

Teste do qui-quadrado de independência

O teste do qui-quadrado de independência está associado a duas variáveis qualitativas, ou seja, uma análise bidimensional (visto na Unidade 2). Você se lembra que as tabelas de contingência permitem verificar a relação de dependência entre as duas variáveis analisadas.

Neste caso, procura-se calcular a freqüência de ocorrência das características dos eventos a serem estudados. Por exemplo, podemos estudar a relação entre o sexo de pessoas (masculino e feminino) e o grau de aceitação do governo (ruim, médio e bom). Então, você vai obter, por exemplo, o número de pessoas (freqüência) que são do sexo feminino e que acham o governo bom. Todos os cruzamentos das duas variáveis são calculados.

Vamos apresentar a você, como exemplo, os possíveis resultados da situação apresentada anteriormente (dados simulados).

Sexo	Função			
	Ruim	Médio	Bom	Total
Masculino	157	27	74	258
Feminino	206	0	10	216
Total	363	27	84	474

Podemos, então, querer determinar o grau de associação entre essas duas variáveis, ou seja, se o grau de aceitação do governo depende do sexo ou existe uma relação de dependência.

As hipóteses a serem testadas são:

H_o: variável linha independe da variável coluna
H_1: variável linha está associada com a variável coluna

A estatística de qui-quadrado será dada por meio da seguinte expressão:

$$\chi_c^2 = \sum_{i=1}^{k} \frac{(fo_i - fe_i)^2}{fe_i}$$

Onde o valor k corresponde ao número de classes (freqüências encontradas). Você pode verificar que **fo** corresponde à freqüência observada, ou seja, o valor encontrado na tabela de contingência.

Já **fe** corresponde à freqüência esperada caso as variáveis não tenham nenhuma relação de dependência, ou seja, as duas variáveis sejam independentes. Em função desta definição, a freqüência esperada (**fe**) será obtida por:

$$fe = \frac{(total\ linha).(total\ coluna)}{total\ geral}$$

Neste caso, os graus de liberdade (v), para que possamos olhar a tabela de qui-quadrado, são dados por:

v = (h-1) (k-1) nas tabelas com h linhas e k colunas

Então, para cada célula da tabela de contingências, você vai calcular a diferença entre **fe** e. **fo**. Esta diferença é elevada ao quadrado para evitar que as diferenças positivas e negativas se anulem. A divisão pela freqüência esperada é feita para obter diferenças em termos relativos.

Para entendermos melhor o teste de qui-quadrado do tipo independência, vamos trabalhar com a seguinte situação: para testar se determinada droga era capaz de inibir a absorção de álcool pelo organismo humano, realizou-se um experimento com a participação de 60 voluntários (homens saudáveis, idade entre 25 e 28 anos). Metade dos voluntários tomou a droga, e a outra metade não tomou. Todos os voluntários tomaram duas doses de uísque. Uma hora mais tarde, selecionou-se uma amostra do sangue de cada sujeito, observando-se os resultados a seguir. Usando 5% de significância, pode-se concluir que o resultado do teste está associado à ingestão da droga?

Teste Droga	Presença de álcool	Ausência de álcool
Tomaram	8	32
Não tomaram	16	40

H_o: Presença ou ausência de álcool independe de tomar droga
H_1: Presença ou ausência de álcool está associado a tomar droga

$$\frac{56 \,.\, 24}{96} = 14$$

Teste Droga	Presença de álcool	Ausência de álcool	
Tomaram	8 (10)	32 (30)	40
Não tomaram	16 (14)	40 (42)	56
	24	72	96

Valores entre parênteses (**fe**)

$$\chi_c^2 = \sum_{i=1}^{k} \frac{(fo_i - fe_i)^2}{fe_i} = \frac{(8-10)^2}{10} + \ldots.. + \frac{(40-42)^2}{42} = 0{,}914$$

$$v = (2\text{-}1)\,.\,(2\text{-}1) = 1 \text{ gl}$$

$\alpha = 0{,}05$ ➔ Qui-quadrado tabelado = 3,8415

Como o valor calculado (0,914) foi menor do que o tabelado, então o calculado caiu na região de aceitação de H_o. Portanto, não temos indícios para rejeitar a hipótese H_o, ou seja, o uso da droga não levou a uma inibição da absorção de álcool.

Análise de variância

A análise de variância é um teste de hipótese utilizado para a comparação de mais de duas populações. Imagine que você queira comparar o grau de endividamento de empresas de três setores (indústria, comércio e prestação de serviços). Para a comparação, é necessário que você tenha repetições, pois elas é que medirão a variação do acaso. Então, você deve selecionar uma amostra de dez empresas de cada setor (repetições).

Para realizar uma análise de variância, dividimos a variação total de um conjunto de tratamentos a serem comparados com as suas respectivas repetições. No nosso exemplo, os setores da indústria correspondem aos tratamentos. Você tem, então, dois componentes: variação ENTRE e variação DENTRO.

A variação ENTRE corresponde à variação encontrada entre as médias dos tratamentos, em relação a uma média geral. Esta variação mede a diferença que ocorre entre os tratamentos.

Já a variação DENTRO do tratamento, como o próprio nome diz, é a variação que ocorre entre as repetições de cada tratamento. Você pode ver que as avaliações das repetições dentro de cada tratamento correspondem à variação do acaso. Então, você tem o quadro a seguir, que sintetiza tudo o que discutimos.

VARIAÇÃO TOTAL = VARIAÇÃO ENTRE + VARIAÇÃO DENTRO

OU

VARIÇAÇÃO TOTAL = VARIAÇÃO NÃO ALEATÓRIA + VARIAÇÃO ALEATÓRIA

Vamos, então, aprender a calcular a variação ENTRE, DENTRO E TOTAL.

Aqui serão apresentadas expressões simplificadas para o cálculo das variações. Estas variações correspondem a cálculos de somas de quadrados, semelhantes às aprendidas na Unidade 1 para cálculo da variância.

A variação ENTRE tratamento é aquela atribuída estritamente à variabilidade das médias dos tratamentos em relação à média geral.

A variação DENTRO de tratamentos é aquela devida à variação de cada observação em relação à média do tratamento. É a variação devida a todas as fontes que causam variações nos experimentos (acaso), excetuando os tratamentos.

A variação total é a variação de cada observação em relação à média geral. Então, temos a seguir as expressões para cálculo das somas de quadrados:

$$Variação\ \ Total\ = SQTotal\ = \sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{ij} - \overline{\overline{Y}})^2 = \sum Y^2 - \frac{(\sum Y)^2}{rt}$$

Onde: $\overline{\overline{y}}$: média geral de todos os tratamentos; r: número de repetições e t: número de tratamentos.

$$Variação\ \ Entre\ = SQTratamento = \sum_{i=1}^{t}\frac{T_i^2}{r_j} - C$$

$$em\ que:$$
$$T_i = total\ \ do\ \ trat\ \ i$$
$$C = fator\ \ de\ \ correção\ = \frac{(\sum\sum Y_{ij})^2}{t.r}$$

Como você sabe, a variação TOTAL é igual à variação ENTRE mais variação DENTRO. Portanto, o cálculo da variação DENTRO (efeitos do acaso) ou a soma de quadrados DENTRO é obtida por meio da seguinte diferença:

$$SQDENTRO\ = \ SQTOTAL\ - \ SQENTRE$$

O valor da soma de quadrado DENTRO é obtido por diferença, devido à maior dificuldade de sua obtenção, principalmente em esquemas de análise de variância mais complexos.

Esse tipo de análise de variância é chamado de análise com um fator ou de um critério. É o mais simples de todos os esquemas de análise de variância, sendo recomendado quando todas as condições experimentais são homogêneas (não há uma variação em uma determinada direção). É próprio para situações (experimentos) nas quais se possa garantir homogeneidade.

Este processo foi desenvolvido por Fisher com o objetivo de repartir a variância de uma variável aleatória em partes ortogonais (independentes) correspondentes a tratamentos (fator) e erro experimental (variações do acaso).

Você pode vislumbrar agora que os objetivos da análise de variância são obter estimativas precisas das médias dos tratamentos, diferenças entre médias e testar hipóteses sobre igualdade de médias de tratamentos.

As hipóteses na análise de variância são:

H_o: $t_1 = t_2 = ... = t_t$ (não existe diferença entre as médias dos tratamentos)

H_1: no mínimo, um dos tratamentos difere dos demais.

Na análise de variância, são obtidas os quadrados médios (QM), que são estimativas não tendenciosas das variâncias envolvidas na análise. Daí vem o nome análise de variância. Estes quadrados médios são obtidos pela divisão da soma de quadrado pelo respectivo grau de liberdade. Então, você tem:

$$QMTrat = \frac{SQTrat}{G.L.\,Trat} \qquad QM\,\mathrm{Re}\,síduo = \frac{SQ\,\mathrm{Re}\,síduo}{G.L\,\mathrm{Re}\,síduo}$$

A forma pela qual você obtém os graus de liberdade é apresentada a seguir:

Fonte de Variação	**G. L. (grau de liberdade)**
Tratamento (Entre)	t – 1 = número de tratamentos menos um
Resíduo (Dentro)	t (r – 1) = número de tratamento vezes numero de repetições menos um
Total	tr – 1 = número de tratamento vezes numero de repetições menos um

Então, você precisa testar se a variância do fator (ENTRE) difere da variância do acaso (DENTRO). A distribuição amostral que você estudou na Unidade 3, que compara duas variâncias, é a distribuição F ou da razão de duas variâncias.

Portanto, você pode utilizar o teste de F para verificar a validade da hipótese H_o descrita anteriormente. O teste é apresentado a seguir:

$$F_{calc} = \frac{QMTrat}{QM\,\mathrm{Re}\,síduo}$$

Comparamos o valor calculado na análise de variância com o valor de tabela F α(v1, v2), em que v1 e v2 são, respectivamente, os graus de liberdade de tratamentos e de resíduos. Se Fcalc > Ftab, temos que o experimento foi significativo, ou seja, indica que existe uma probabilidade superior a $1 - \alpha$ de que pelo menos um dos tratamentos difere dos demais.

O quadro da análise de variância pode ser resumido da seguinte forma:

Fonte de Variação	G. L.	S. Q.	Q. M.	Fcalc	F $\alpha(v1,v2)$
Tratamento	t-1	SQTrat	QMTrat	F calculado	F tabela
Resíduo	t(r-1)	SQResíduo	QMResíduo		
Total	tr-1	SQTotal			

G. L. = graus de liberdade

Vamos, então, fazer um exemplo para entender melhor esta análise: com o objetivo de comparar um determinado índice inflacionário em três regiões metropolitanas em um período de cinco meses, você obteve os resultados apresentados a seguir. Verifique, por meio de uma análise de variância, se as médias são estatisticamente iguais ou não.

Regiões Metropolitanas			
Meses	R1	R2	R3
1	1,60	1,20	2,00
2	2,00	1,10	1,80
3	2,20	1,20	1,40
4	1,70	1,30	1,60
5	1,80	1,00	1,90
Total	5,30	5,80	8,70

As hipóteses desta análise de variância são:

H_o: $R_1 = R_2 = R_3$ (não existe diferença entre as médias das regiões)

H_1: pelo menos uma das regiões difere das demais em média.

As repetições, ou seja, os meses são independentes, pois são considerados apenas repetições.

OBS: o teste F para análise de variância será sempre um teste unilateral à direita, em função do tipo de hipótese alternativa.

Cálculos das somas de quadrados:

$$SQENTRE = \sum_{i=1}^{t} \frac{T_i^2}{r_j} - C$$

$$SQENTRE = \frac{1}{5}\left(9{,}3^2 + 5{,}8^2 + 8{,}7^2\right) - \left(\frac{23{,}8^2}{15}\right) = 39{,}164 - 37{,}7626 = 1{,}401$$

$$SQTotal = \sum Y^2 - \frac{(\sum Y)^2}{rt}$$

$$SQTOTAL = \left(1{,}6^2 + 2{,}0^2 + \ldots\ldots\ldots + 1{,}9^2\right) - \left(\frac{23{,}8^2}{15}\right) = 1{,}9173$$

$$SQDENTRO = SQTOTAL - SQENTRE = 0{,}516$$

FV	GL	SQ	QM	F_{cal}	Significância
Entre	2	1,401	0,701	1.,2943	0,0006
Dentro	12	0,516	0,043		
Total	14	1,9173			

Tabela 8: Análise de variância

Conclusão: existe diferença significativa entre as regiões, pois o F tabelado (tabela de 5% e v_1=2 gl e v_2 = 12 gl) foi menor do que o calculado (16,29), fazendo assim com que o F calculado tenha caído na região de rejeição da hipótese H_o.

Saiba mais...

- Mais exercícios referentes ao assunto estão no site: http://www.famat.ufu.br/prof/marcelo/exercicios.htm

Respostas dos exercícios propostos

Unidade 1

Exercício 1:

a)

Classes	Freqüências absolutas
38,5 |– 43,5	3
43,5 |– 48,5	4
48,5 |– 53,5	7
53,5 |– 58,5	4
58,5 |– 63,5	2
Total	20

Classes	Freqüência acumulada para baixo
abaixo de 43,5	3
abaixo de 48,5	7
abaixo de 53,5	14
abaixo de 58,5	18
abaixo de 63,5	20

Classes	Freqüência acumulada para cima
acima de 38,5	20
acima de 43,5	17
acima de 48,5	13
acima de 53,5	6
acima de 58,5	2

b)

variação entre número das classes	variação de freqüências
5	7
1,5	x

$x = \frac{1{,}5 \cdot 7}{5} = 2{,}1$; como abaixo de 48,5 temos 7, e entre 48,5 e 53,5 temos 5, então, abaixo de 50, teremos: 7 +2,1 = 9,1.

Exercício 2:

j) Média = 715,5 reais; Mediana = 708,82; Moda = 669,23

k) Desvio-padrão = 13,79 e coeficiente de variação = 1,92%

l) R: 950

m) R: 100

n) R: 0,155

o) R: 262

p) R: 194

q) R: 138

r) 3ª classe

Exercício 3:

$\overline{x}$ = 210; s = 10,96; CV = 5,22%; Md = 180; Mo = 100

Exercício 4:

Média = 682,35

Exercício 5:

Sim. Apresentam o mesmo CV.

Unidade 2

Exercício 1:

R: 1 – (1/3 * 1/5 * 3/10) = 0,98

Exercício 2:

a) R: 0,125; b) R: 0,0694; c) R: 0,1388

Exercício 3:

a1) R: 60/100; a2) R: 40/100; a3) R: 24/100; a4) R: 76/100

Exercício 4:

R: 0,05/0,25 = 0,2

Exercício 5:

a) R: 0,4; b) R:0,9; c) R:0,6;

d) R:

0	1	2	3	4
0,1	0,3	0,6	0,9	1

e) R: 0,9. Probabilidade de alugar no máximo três caminhões.

Exercício 6:

R: 0,0089

Exercício 7:

R: P (X = 5) = $C_{20}^{5} 0{,}1^{5} 0{,}9^{15}$ = 0,03192

Exercício 8:

Distribuição binomial com n = 4 e p = ½

a) R: P(x=2) . 2.000 = 0,3750 . 2.000 = 750 famílias

b) R: [P(1) + P(2)] . 2.000 = (0,25 + 0,375) . 2.000 = 1.250 famílias

c) R: P(0) . 2.000 = 0,0625 . 2.000 = 125 famílias

Exercício 9:

R: P(X = 4) = $C_{8}^{4} 0{,}3^{4} 0{,}7^{4}$ = 0,13614

Exercício 10:

R: 1 – [P(0) + P(1)], onde a distribuição de probabilidade é uma Poisson com parâmetro lambda.

a) λ = 1,4 R = 0,40817

b) λ = 2,8 R = 0,76892

c) λ = 5,6 R = 0,97559

Exercício 11:

Para X = 2200 ➔ $Z = \dfrac{X - \mu}{\sigma} = \dfrac{2200 - 2000}{200} = 1{,}00$

Para X = 1700 ➔ $Z = \dfrac{X - \mu}{\sigma} = \dfrac{1700 - 2000}{200} = -1{,}50$

Exercício 12:

a) X = 20 ➔ Z = 0

X = 24 ➔ $Z = \dfrac{24 - 20}{5} = 0{,}8$

P(20 < X < 24) = P(0 < Z < 0,8) = 0,2881 (28,81 %).

b) X = 16 ➔ $Z = \dfrac{16 - 20}{5} = -0{,}8$

X = 20 ➔ Z = 0

P(16 < X < 20) = P (-0,8 < Z < 0) = P(0 < Z < 0,8) = 0,2881 = 28,81

c) X = 28 ➔ Z = (28 - 20)/ 5 = 1,6

P(X > 28) = P (Z > 1,6) = 0,5 – 0,4452 = 0,0548

Exercício 13:

seja X' a mínima média. $P(X \geq X') = 0{,}15$

O Z correspondente é 1,04 (aproximadamente)

$\dfrac{X' - 72}{5} = 1{,}04 \quad X' = 77{,}2$

Unidade 3

Exercício 1:

$n_1 = 101$; $n_2 = 60$; $n_3 = 39$

Exercício 2:

a) a = 2,160; b) a = –1,708; c) a = 1,725

Exercício 3:

a) a = 34,4789; b) a = 10,2829; c) a = 7,2609

Exercício 4:

a) a = 2,092; b) a = 2,092; c) a = 2,075

Unidade 4

Exercício 1:

(**Sugestão**: siga os passos para realizar um teste de hipótese.)

$$Z = \frac{\overline{X} - \mu_O}{\sigma \ / \sqrt{n}} = \frac{43000 - 40000}{6500 / \sqrt{49}} = 3{,}23$$

$$Z_\alpha = Z_{0,05} = 1{,}64$$

Como o valor calculado foi maior que o tabelado (1,64), ele caiu na região de rejeição de H_o.

Exercício 2:

$$H_0 : \mu_A - \mu_B = 0$$

$$H_0 : \mu_A - \mu_B \neq 0$$

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{1/n_1 + 1/n_2}} = \frac{(76 - 68) - 0}{8\sqrt{1/12 + 1/15}} = 2{,}56$$

$$t_{0,025} = 2{,}060$$

Como o valor calculado foi maior que o tabelado (2,060), ele caiu na região de rejeição de H_o.

Exercício 3:

$$H_0 : \mu_1 - \mu_2 = 0 \qquad H_0 : \mu_1 - \mu_2 \neq 0$$

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_1^2 / n_1 + s_2^2 / n_2}} = \frac{(112 - 107) - 0}{\sqrt{8^2 / 14 + 10^2 / 16}} = -1{,}52$$

v=29,7425 =30 (graus de liberdade obtido pela aproximação)

$$t_{0,025} = 2{,}042 \quad (com\ 30\ gl)$$

Conclusão: como o valor calculado caiu na região de aceitação, então as médias são estatisticamente iguais, o que indica que as duas regiões apresentam o mesmo potencial.

REFERÊNCIAS

ARANGO, H. G. **Bioestatística Teórica e Computacional**. Rio de Janeiro: Guanabara Koogan, 2001

BARBETTA, P. A. **Estatística Aplicada às Ciências Sociais**. 4 ed. Florianópolis: Ed. da UFSC, 2002.

BEIGUELMAN, B. **Curso Prático de Bioestatística**. Ribeirão Preto: Revista Brasileira de Genética, 1996.

BRAULE, Ricardo. **Estatística Aplicada com Excel**: para cursos de Administração e Economia. Rio de Janeiro: Campus, 2001.

BUSSAB, W. O.; MORETTIN, P. **Estatística Básica**. São Paulo: Atual, 2002.

COSTA NETO, P. L. de O. **Estatística**. São Paulo: Edgard Blücher, 2002.

DOWNING, D.; CLARK, J. **Estatística Aplicada**. São Paulo: Saraiva, 2000.

FONSECA, Jairo Simon da; MARTINS, Gilberto de Andrade. **Curso de Estatística**. Rio de Janeiro: LTC, 1982.

FREUD, J. E.; SIMON, G. A. **Estatística aplicada**. Bookman, 2000, 403 p.

LEVINE, D. M.; BERENSON, M. L.; STEPHAN, D. **Estatística:** teoria e aplicações (usando o Microsoft Excel em português). LTC, 2000, 812 p.

MORETTIN, L. G. **Estatística Básica**: Probabilidade. V. 1. São Paulo: Makron Books, 1999.

________. **Estatística Básica**: Inferência. V. 2. São Paulo: Makron Books, 1999.

SOARES, J. F.; FARIAS, A. A.; CESAR, C. C. **Introdução à Estatística**. Rio de Janeiro: LTC, 1991.

SPIEGEL, M. **Probabilidade e Estatística**. Mc Graw Hill. 1993.

STEVENSON, William J. **Estatística Aplicada à Administração**. São Paulo: Harper, 1981.

TRIOLA, M. F. **Introdução à Estatística**. Rio de Janeiro: LTC, 1999.

WONNACOTT, T. H., WONNACOTT, R. J. **Estatística Aplicada à Economia e à Administração**. Rio de Janeiro: LTC,1981.

Anexos

Z	0	1	2	3	4	5	6	7	8	9
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

Tabela 1: Área sob a curva normal padronizada compreendida entre os valores 0 e Z

	α								
GL	0.250	0.200	0.150	0.100	0.050	0.025	0.010	0.005	0.001
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.656	318.289
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.328
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.894
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.261
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.160
240	0.676	0.843	1.039	1.285	1.651	1.970	2.342	2.596	3.125
480	0.675	0.842	1.038	1.283	1.648	1.965	2.334	2.586	3.107
700	0.675	0.842	1.037	1.283	1.647	1.963	2.332	2.583	3.102
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098

Tabela 2: Limites unilaterais da distribuição t de Student ao nível de probabilidade

	α												
GL	0.995	0.99	0.975	0.95	0.9	0.75	0.5	0.25	0.1	0.05	0.025	0.01	0.005
1	0.0000	0.0002	0.0010	0.0039	0.0158	0.1015	0.4549	1.3233	2.7055	3.8415	5.0239	6.6349	7.8794
2	0.0100	0.0201	0.0506	0.1026	0.2107	0.5754	1.3863	2.7726	4.6052	5.9915	7.3778	9.2104	10.5965
3	0.0717	0.1148	0.2158	0.3518	0.5844	1.2125	2.3660	4.1083	6.2514	7.8147	9.3484	11.3449	12.8381
4	0.2070	0.2971	0.4844	0.7107	1.0636	1.9226	3.3567	5.3853	7.7794	9.4877	11.1433	13.2767	14.8602
5	0.4118	0.5543	0.8312	1.1455	1.6103	2.6746	4.3515	6.6257	9.2363	11.0705	12.8325	15.0863	16.7496
6	0.6757	0.8721	1.2373	1.6354	2.2041	3.4546	5.3481	7.8408	10.6446	12.5916	14.4494	16.8119	18.5475
7	0.9893	1.2390	1.6899	2.1673	2.8331	4.2549	6.3458	9.0371	12.0170	14.0671	16.0128	18.4753	20.2777
8	1.3444	1.6465	2.1797	2.7326	3.4895	5.0706	7.3441	10.2189	13.3616	15.5073	17.5345	20.0902	21.9549
9	1.7349	2.0879	2.7004	3.3251	4.1682	5.8988	8.3428	11.3887	14.6837	16.9190	19.0228	21.6660	23.5893
10	2.1558	2.5582	3.2470	3.9403	4.8652	6.7372	9.3418	12.5489	15.9872	18.3070	20.4832	23.2093	25.1881
11	2.6032	3.0535	3.8157	4.5748	5.5778	7.5841	10.3410	13.7007	17.2750	19.6752	21.9200	24.7250	26.7569
12	3.0738	3.5706	4.4038	5.2260	6.3038	8.4384	11.3403	14.8454	18.5493	21.0261	23.3367	26.2170	28.2997
13	3.5650	4.1069	5.0087	5.8919	7.0415	9.2991	12.3398	15.9839	19.8119	22.3620	24.7356	27.6882	29.8193
14	4.0747	4.6604	5.6287	6.5706	7.7895	10.1653	13.3393	17.1169	21.0641	23.6848	26.1189	29.1412	31.3194
15	4.6009	5.2294	6.2621	7.2609	8.5468	11.0365	14.3389	18.2451	22.3071	24.9958	27.4884	30.5780	32.8015
16	5.1422	5.8122	6.9077	7.9616	9.3122	11.9122	15.3385	19.3689	23.5418	26.2962	28.8453	31.9999	34.2671
17	5.6973	6.4077	7.5642	8.6718	10.0852	12.7919	16.3382	20.4887	24.7690	27.5871	30.1910	33.4087	35.7184
18	6.2648	7.0149	8.2307	9.3904	10.8649	13.6753	17.3379	21.6049	25.9894	28.8693	31.5264	34.8052	37.1564
19	6.8439	7.6327	8.9065	10.1170	11.6509	14.5620	18.3376	22.7178	27.2036	30.1435	32.8523	36.1908	38.5821
20	7.4338	8.2604	9.5908	10.8508	12.4426	15.4518	19.3374	23.8277	28.4120	31.4104	34.1696	37.5663	39.9969
21	8.0336	8.8972	10.2829	11.5913	13.2396	16.3444	20.3372	24.9348	29.6151	32.6706	35.4789	38.9322	41.4009
22	8.6427	9.5425	10.9823	12.3380	14.0415	17.2396	21.3370	26.0393	30.8133	33.9245	36.7807	40.2894	42.7957
23	9.2604	10.1957	11.6885	13.0905	14.8480	18.1373	22.3369	27.1413	32.0069	35.1725	38.0756	41.6383	44.1814
24	9.8862	10.8563	12.4011	13.8484	15.6587	19.0373	23.3367	28.2412	33.1962	36.4150	39.3641	42.9798	45.5584
25	10.5196	11.5240	13.1197	14.6114	16.4734	19.9393	24.3366	29.3388	34.3816	37.6525	40.6465	44.3140	46.9280
26	11.1602	12.1982	13.8439	15.3792	17.2919	20.8434	25.3365	30.4346	35.5632	38.8851	41.9231	45.6416	48.2898
27	11.8077	12.8785	14.5734	16.1514	18.1139	21.7494	26.3363	31.5284	36.7412	40.1133	43.1945	46.9628	49.6450
28	12.4613	13.5647	15.3079	16.9279	18.9392	22.6572	27.3362	32.6205	37.9159	41.3372	44.4608	48.2782	50.9936
29	13.1211	14.2564	16.0471	17.7084	19.7677	23.5666	28.3361	33.7109	39.0875	42.5569	45.7223	49.5878	52.3355
30	13.7867	14.9535	16.7908	18.4927	20.5992	24.4776	29.3360	34.7997	40.2560	43.7730	46.9792	50.8922	53.6719
40	20.7066	22.1642	24.4331	26.5093	29.0505	33.6603	39.3353	45.6160	51.8050	55.7585	59.3417	63.6908	66.7660
50	27.9908	29.7067	32.3574	34.7642	37.6886	42.9421	49.3349	56.3336	63.1671	67.5048	71.4202	76.1538	79.4898
60	35.5344	37.4848	40.4817	43.1880	46.4589	52.2938	59.3347	66.9815	74.3970	79.0820	83.2977	88.3794	91.9518
100	67.3275	70.0650	74.2219	77.9294	82.3581	90.1332	99.3341	109.1412	118.4980	124.3421	129.5613	135.8069	140.1697
120	83.8517	86.9233	91.5726	95.7046	100.6236	109.2197	119.3340	130.0546	140.2326	146.5673	152.2113	158.9500	163.6485
240	187.3241	191.9897	198.9838	205.1354	212.3856	224.8820	239.3337	254.3918	268.4707	277.1377	284.8025	293.8881	300.1826
480	403.9488	410.8739	421.1886	430.1981	440.7454	458.7543	479.3335	500.5192	520.1110	532.0753	542.5989	555.0066	563.5606

Tabela 3: Limites unilaterais da distribuição de 2 ao nível de probabilidade

GL	V1																			
V2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	40	60	120	240
1	39.864	49.500	53.593	55.833	57.240	58.204	58.906	59.439	59.857	60.195	60.473	60.705	60.902	61.073	61.220	61.740	62.529	62.794	63.061	63.194
2	8.526	9.000	9.162	9.243	9.293	9.326	9.349	9.367	9.381	9.392	9.401	9.408	9.415	9.420	9.425	9.441	9.466	9.475	9.483	9.487
3	5.538	5.462	5.391	5.343	5.309	5.285	5.266	5.252	5.240	5.230	5.222	5.216	5.210	5.205	5.200	5.184	5.160	5.151	5.143	5.138
4	4.545	4.325	4.191	4.107	4.051	4.010	3.979	3.955	3.936	3.920	3.907	3.896	3.886	3.878	3.870	3.844	3.804	3.790	3.775	3.768
5	4.060	3.780	3.619	3.520	3.453	3.405	3.368	3.339	3.316	3.297	3.282	3.268	3.257	3.247	3.238	3.207	3.157	3.140	3.123	3.114
6	3.776	3.463	3.289	3.181	3.108	3.055	3.014	2.983	2.958	2.937	2.920	2.905	2.892	2.881	2.871	2.836	2.781	2.762	2.742	2.732
7	3.589	3.257	3.074	2.961	2.883	2.827	2.785	2.752	2.725	2.703	2.684	2.668	2.654	2.643	2.632	2.595	2.535	2.514	2.493	2.482
8	3.458	3.113	2.924	2.806	2.726	2.668	2.624	2.589	2.561	2.538	2.519	2.502	2.488	2.475	2.464	2.425	2.361	2.339	2.316	2.304
9	3.360	3.006	2.813	2.693	2.611	2.551	2.505	2.469	2.440	2.416	2.396	2.379	2.364	2.351	2.340	2.298	2.232	2.208	2.184	2.172
10	3.285	2.924	2.728	2.605	2.522	2.461	2.414	2.377	2.347	2.323	2.302	2.284	2.269	2.255	2.244	2.201	2.132	2.107	2.082	2.069
11	3.225	2.860	2.660	2.536	2.451	2.389	2.342	2.304	2.274	2.248	2.227	2.209	2.193	2.179	2.167	2.123	2.052	2.026	2.000	1.986
12	3.177	2.807	2.606	2.480	2.394	2.331	2.283	2.245	2.214	2.188	2.166	2.147	2.131	2.117	2.105	2.060	1.986	1.960	1.932	1.918
13	3.136	2.763	2.560	2.434	2.347	2.283	2.234	2.195	2.164	2.138	2.116	2.097	2.080	2.066	2.053	2.007	1.931	1.904	1.876	1.861
14	3.102	2.726	2.522	2.395	2.307	2.243	2.193	2.154	2.122	2.095	2.073	2.054	2.037	2.022	2.010	1.962	1.885	1.857	1.828	1.813
15	3.073	2.695	2.490	2.361	2.273	2.208	2.158	2.119	2.086	2.059	2.037	2.017	2.000	1.985	1.972	1.924	1.845	1.817	1.787	1.771
16	3.048	2.668	2.462	2.333	2.244	2.178	2.128	2.088	2.055	2.028	2.005	1.985	1.968	1.953	1.940	1.891	1.811	1.782	1.751	1.735
17	3.026	2.645	2.437	2.308	2.218	2.152	2.102	2.061	2.028	2.001	1.978	1.958	1.940	1.925	1.912	1.862	1.781	1.751	1.719	1.703
18	3.007	2.624	2.416	2.286	2.196	2.130	2.079	2.038	2.005	1.977	1.954	1.933	1.916	1.900	1.887	1.837	1.754	1.723	1.691	1.674
19	2.990	2.606	2.397	2.266	2.176	2.109	2.058	2.017	1.984	1.956	1.932	1.912	1.894	1.878	1.865	1.814	1.730	1.699	1.666	1.649
20	2.975	2.589	2.380	2.249	2.158	2.091	2.040	1.999	1.965	1.937	1.913	1.892	1.875	1.859	1.845	1.794	1.708	1.677	1.643	1.626
21	2.961	2.575	2.365	2.233	2.142	2.075	2.023	1.982	1.948	1.920	1.896	1.875	1.857	1.841	1.827	1.776	1.689	1.657	1.623	1.605
22	2.949	2.561	2.351	2.219	2.128	2.060	2.008	1.967	1.933	1.904	1.880	1.859	1.841	1.825	1.811	1.759	1.671	1.639	1.604	1.586
23	2.937	2.549	2.339	2.207	2.115	2.047	1.995	1.953	1.919	1.890	1.866	1.845	1.827	1.811	1.796	1.744	1.655	1.622	1.587	1.568
24	2.927	2.538	2.327	2.195	2.103	2.035	1.983	1.941	1.906	1.877	1.853	1.832	1.814	1.797	1.783	1.730	1.641	1.607	1.571	1.552
25	2.918	2.528	2.317	2.184	2.092	2.024	1.971	1.929	1.895	1.866	1.841	1.820	1.802	1.785	1.771	1.718	1.627	1.593	1.557	1.538
26	2.909	2.519	2.307	2.174	2.082	2.014	1.961	1.919	1.884	1.855	1.830	1.809	1.790	1.774	1.760	1.706	1.615	1.581	1.544	1.524
27	2.901	2.511	2.299	2.165	2.073	2.005	1.952	1.909	1.874	1.845	1.820	1.799	1.780	1.764	1.749	1.695	1.603	1.569	1.531	1.511
28	2.894	2.503	2.291	2.157	2.064	1.996	1.943	1.900	1.865	1.836	1.811	1.790	1.771	1.754	1.740	1.685	1.592	1.558	1.520	1.500
29	2.887	2.495	2.283	2.149	2.057	1.988	1.935	1.892	1.857	1.827	1.802	1.781	1.762	1.745	1.731	1.676	1.583	1.547	1.509	1.489
30	2.881	2.489	2.276	2.142	2.049	1.980	1.927	1.884	1.849	1.819	1.794	1.773	1.754	1.737	1.722	1.667	1.573	1.538	1.499	1.478
40	2.835	2.440	2.226	2.091	1.997	1.927	1.873	1.829	1.793	1.763	1.737	1.715	1.695	1.678	1.662	1.605	1.506	1.467	1.425	1.402
50	2.809	2.412	2.197	2.061	1.966	1.895	1.840	1.796	1.760	1.729	1.703	1.680	1.660	1.643	1.627	1.568	1.465	1.424	1.379	1.354
60	2.791	2.393	2.177	2.041	1.946	1.875	1.819	1.775	1.738	1.707	1.680	1.657	1.637	1.619	1.603	1.543	1.437	1.395	1.348	1.321
80	2.769	2.370	2.154	2.016	1.921	1.849	1.793	1.748	1.711	1.680	1.653	1.629	1.609	1.590	1.574	1.513	1.403	1.358	1.307	1.278
100	2.756	2.356	2.139	2.002	1.906	1.834	1.778	1.732	1.695	1.663	1.636	1.612	1.592	1.573	1.557	1.494	1.382	1.336	1.282	1.250
120	2.748	2.347	2.130	1.992	1.896	1.824	1.767	1.722	1.684	1.652	1.625	1.601	1.580	1.562	1.545	1.482	1.368	1.320	1.265	1.232
240	2.727	2.325	2.107	1.968	1.871	1.799	1.742	1.696	1.658	1.625	1.598	1.573	1.552	1.533	1.516	1.451	1.332	1.281	1.219	1.180

Tabela 4: Limites unilaterais da distribuição F de Fisher-Snedecor ao nível de 10% de probabilidade

GL	V1																			
V2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	40	60	120	240
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.0	243.9	244.7	245.4	245.9	248.0	251.1	252.2	253.3	253.8
2	18.513	19.000	19.164	19.247	19.296	19.329	19.353	19.371	19.385	19.396	19.405	19.412	19.419	19.424	19.429	19.446	19.471	19.479	19.487	19.492
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.785	8.763	8.745	8.729	8.715	8.703	8.660	8.594	8.572	8.549	8.538
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.936	5.912	5.891	5.873	5.858	5.803	5.717	5.688	5.658	5.643
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.704	4.678	4.655	4.636	4.619	4.558	4.464	4.431	4.398	4.382
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.027	4.000	3.976	3.956	3.938	3.874	3.774	3.740	3.705	3.687
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.603	3.575	3.550	3.529	3.511	3.445	3.340	3.304	3.267	3.249
8	5.318	4.459	4.066	3.838	3.688	3.581	3.500	3.438	3.388	3.347	3.313	3.284	3.259	3.237	3.218	3.150	3.043	3.005	2.967	2.947
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.102	3.073	3.048	3.025	3.006	2.936	2.826	2.787	2.748	2.727
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.943	2.913	2.887	2.865	2.845	2.774	2.661	2.621	2.580	2.559
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.818	2.788	2.761	2.739	2.719	2.646	2.531	2.490	2.448	2.426
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.717	2.687	2.660	2.637	2.617	2.544	2.426	2.384	2.341	2.319
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.635	2.604	2.577	2.554	2.533	2.459	2.339	2.297	2.252	2.230
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.565	2.534	2.507	2.484	2.463	2.388	2.266	2.223	2.178	2.155
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.507	2.475	2.448	2.424	2.403	2.328	2.204	2.160	2.114	2.090
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.456	2.425	2.397	2.373	2.352	2.276	2.151	2.106	2.059	2.035
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.413	2.381	2.353	2.329	2.308	2.230	2.104	2.058	2.011	1.986
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.374	2.342	2.314	2.290	2.269	2.191	2.063	2.017	1.968	1.943
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.340	2.308	2.280	2.256	2.234	2.155	2.026	1.980	1.930	1.905
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.310	2.278	2.250	2.225	2.203	2.124	1.994	1.946	1.896	1.870
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321	2.283	2.250	2.222	2.197	2.176	2.096	1.965	1.916	1.866	1.839
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297	2.259	2.226	2.198	2.173	2.151	2.071	1.938	1.889	1.838	1.811
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275	2.236	2.204	2.175	2.150	2.128	2.048	1.914	1.865	1.813	1.785
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255	2.216	2.183	2.155	2.130	2.108	2.027	1.892	1.842	1.790	1.762
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236	2.198	2.165	2.136	2.111	2.089	2.007	1.872	1.822	1.768	1.740
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.220	2.181	2.148	2.119	2.094	2.072	1.990	1.853	1.803	1.749	1.720
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204	2.166	2.132	2.103	2.078	2.056	1.974	1.836	1.785	1.731	1.702
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190	2.151	2.118	2.089	2.064	2.041	1.959	1.820	1.769	1.714	1.685
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177	2.138	2.104	2.075	2.050	2.027	1.945	1.806	1.754	1.698	1.669
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165	2.126	2.092	2.063	2.037	2.015	1.932	1.792	1.740	1.683	1.654
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077	2.038	2.003	1.974	1.948	1.924	1.839	1.693	1.637	1.577	1.544
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026	1.986	1.952	1.921	1.895	1.871	1.784	1.634	1.576	1.511	1.476
60	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993	1.952	1.917	1.887	1.860	1.836	1.748	1.594	1.534	1.467	1.430
80	3.960	3.111	2.719	2.486	2.329	2.214	2.126	2.056	1.999	1.951	1.910	1.875	1.845	1.817	1.793	1.703	1.545	1.482	1.411	1.370
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032	1.975	1.927	1.886	1.850	1.819	1.792	1.768	1.676	1.515	1.450	1.376	1.333
120	3.920	3.072	2.680	2.447	2.290	2.175	2.087	2.016	1.959	1.910	1.869	1.834	1.803	1.775	1.750	1.659	1.495	1.429	1.352	1.307
240	3.881	3.033	2.642	2.409	2.252	2.136	2.048	1.977	1.919	1.870	1.829	1.793	1.761	1.733	1.708	1.614	1.445	1.375	1.290	1.237

Tabela 5: Limites unilaterais da distribuição F de Fisher-Snedecor ao nível de 5% de probabilidade

GL	V1																			
V2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	40	60	120	240
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.6	963.3	968.6	973.0	976.7	979.8	982.5	984.9	993.1	1005.6	1009.8	1014.0	1016.1
2	38.506	39.000	39.166	39.248	39.298	39.331	39.356	39.373	39.387	39.398	39.407	39.415	39.421	39.427	39.431	39.448	39.473	39.481	39.489	39.494
3	17.443	16.044	15.439	15.101	14.885	14.735	14.624	14.540	14.473	14.419	14.374	14.337	14.305	14.277	14.253	14.167	14.036	13.992	13.947	13.925
4	12.218	10.649	9.979	9.604	9.364	9.197	9.074	8.980	8.905	8.844	8.794	8.751	8.715	8.684	8.657	8.560	8.411	8.360	8.309	8.283
5	10.007	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619	6.568	6.525	6.488	6.456	6.428	6.329	6.175	6.123	6.069	6.042
6	8.813	7.260	6.599	6.227	5.988	5.820	5.695	5.600	5.523	5.461	5.410	5.366	5.329	5.297	5.269	5.168	5.012	4.959	4.904	4.877
7	8.073	6.542	5.890	5.523	5.285	5.119	4.995	4.899	4.823	4.761	4.709	4.666	4.628	4.596	4.568	4.467	4.309	4.254	4.199	4.171
8	7.571	6.059	5.416	5.053	4.817	4.652	4.529	4.433	4.357	4.295	4.243	4.200	4.162	4.130	4.101	3.999	3.840	3.784	3.728	3.699
9	7.209	5.715	5.078	4.718	4.484	4.320	4.197	4.102	4.026	3.964	3.912	3.868	3.831	3.798	3.769	3.667	3.505	3.449	3.392	3.363
10	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717	3.665	3.621	3.583	3.550	3.522	3.419	3.255	3.198	3.140	3.110
11	6.724	5.256	4.630	4.275	4.044	3.881	3.759	3.664	3.588	3.526	3.474	3.430	3.392	3.359	3.330	3.226	3.061	3.004	2.944	2.914
12	6.554	5.096	4.474	4.121	3.891	3.728	3.607	3.512	3.436	3.374	3.321	3.277	3.239	3.206	3.177	3.073	2.906	2.848	2.787	2.756
13	6.414	4.965	4.347	3.996	3.767	3.604	3.483	3.388	3.312	3.250	3.197	3.153	3.115	3.082	3.053	2.948	2.780	2.720	2.659	2.628
14	6.298	4.857	4.242	3.892	3.663	3.501	3.380	3.285	3.209	3.147	3.095	3.050	3.012	2.979	2.949	2.844	2.674	2.614	2.552	2.520
15	6.200	4.765	4.153	3.804	3.576	3.415	3.293	3.199	3.123	3.060	3.008	2.963	2.925	2.891	2.862	2.756	2.585	2.524	2.461	2.429
16	6.115	4.687	4.077	3.729	3.502	3.341	3.219	3.125	3.049	2.986	2.934	2.889	2.851	2.817	2.788	2.681	2.509	2.447	2.383	2.350
17	6.042	4.619	4.011	3.665	3.438	3.277	3.156	3.061	2.985	2.922	2.870	2.825	2.786	2.753	2.723	2.616	2.442	2.380	2.315	2.282
18	5.978	4.560	3.954	3.608	3.382	3.221	3.100	3.005	2.929	2.866	2.814	2.769	2.730	2.696	2.667	2.559	2.384	2.321	2.256	2.222
19	5.922	4.508	3.903	3.559	3.333	3.172	3.051	2.956	2.880	2.817	2.765	2.720	2.681	2.647	2.617	2.509	2.333	2.270	2.203	2.169
20	5.871	4.461	3.859	3.515	3.289	3.128	3.007	2.913	2.837	2.774	2.721	2.676	2.637	2.603	2.573	2.464	2.287	2.223	2.156	2.121
21	5.827	4.420	3.819	3.475	3.250	3.090	2.969	2.874	2.798	2.735	2.682	2.637	2.598	2.564	2.534	2.425	2.246	2.182	2.114	2.079
22	5.786	4.383	3.783	3.440	3.215	3.055	2.934	2.839	2.763	2.700	2.647	2.602	2.563	2.528	2.498	2.389	2.210	2.145	2.076	2.040
23	5.750	4.349	3.750	3.408	3.183	3.023	2.902	2.808	2.731	2.668	2.615	2.570	2.531	2.497	2.466	2.357	2.176	2.111	2.041	2.005
24	5.717	4.319	3.721	3.379	3.155	2.995	2.874	2.779	2.703	2.640	2.586	2.541	2.502	2.468	2.437	2.327	2.146	2.080	2.010	1.973
25	5.686	4.291	3.694	3.353	3.129	2.969	2.848	2.753	2.677	2.613	2.560	2.515	2.476	2.441	2.411	2.300	2.118	2.052	1.981	1.944
26	5.659	4.265	3.670	3.329	3.105	2.945	2.824	2.729	2.653	2.590	2.536	2.491	2.452	2.417	2.387	2.276	2.093	2.026	1.954	1.917
27	5.633	4.242	3.647	3.307	3.083	2.923	2.802	2.707	2.631	2.568	2.514	2.469	2.429	2.395	2.364	2.253	2.069	2.002	1.930	1.892
28	5.610	4.221	3.626	3.286	3.063	2.903	2.782	2.687	2.611	2.547	2.494	2.448	2.409	2.374	2.344	2.232	2.048	1.980	1.907	1.869
29	5.588	4.201	3.607	3.267	3.044	2.884	2.763	2.669	2.592	2.529	2.475	2.430	2.390	2.355	2.325	2.213	2.028	1.959	1.886	1.847
30	5.568	4.182	3.589	3.250	3.026	2.867	2.746	2.651	2.575	2.511	2.458	2.412	2.372	2.338	2.307	2.195	2.009	1.940	1.866	1.827
40	5.424	4.051	3.463	3.126	2.904	2.744	2.624	2.529	2.452	2.388	2.334	2.288	2.248	2.213	2.182	2.068	1.875	1.803	1.724	1.682
50	5.340	3.975	3.390	3.054	2.833	2.674	2.553	2.458	2.381	2.317	2.263	2.216	2.176	2.140	2.109	1.993	1.796	1.721	1.639	1.594
60	5.286	3.925	3.343	3.008	2.786	2.627	2.507	2.412	2.334	2.270	2.216	2.169	2.129	2.093	2.061	1.944	1.744	1.667	1.581	1.534
80	5.218	3.864	3.284	2.950	2.730	2.571	2.450	2.355	2.277	2.213	2.158	2.111	2.071	2.035	2.003	1.884	1.679	1.599	1.508	1.457
100	5.179	3.828	3.250	2.917	2.696	2.537	2.417	2.321	2.244	2.179	2.124	2.077	2.036	2.000	1.968	1.849	1.640	1.558	1.463	1.409
120	5.152	3.805	3.227	2.894	2.674	2.515	2.395	2.299	2.222	2.157	2.102	2.055	2.014	1.977	1.945	1.825	1.614	1.530	1.433	1.376
240	5.088	3.746	3.171	2.839	2.620	2.461	2.341	2.245	2.167	2.102	2.047	1.999	1.958	1.921	1.888	1.766	1.549	1.460	1.354	1.289

GL	V1																			
V2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	40	60	120	240
1	4052.2	4999.3	5403.5	5624.3	5764.0	5859.0	5928.3	5981.0	6022.4	6055.9	6083.4	6106.7	6125.8	6143.0	6157.0	6208.7	6286.4	6313.0	6339.5	6352.6
2	98.502	99.000	99.164	99.251	99.302	99.331	99.357	99.375	99.390	99.397	99.408	99.419	99.422	99.426	99.433	99.448	99.477	99.484	99.491	99.495
3	34.116	30.816	29.457	28.710	28.237	27.911	27.671	27.489	27.345	27.228	27.132	27.052	26.983	26.924	26.872	26.690	26.411	26.316	26.221	26.173
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546	14.452	14.374	14.306	14.249	14.198	14.019	13.745	13.652	13.558	13.511
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051	9.963	9.888	9.825	9.770	9.722	9.553	9.291	9.202	9.112	9.066
6	13.745	10.925	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.790	7.718	7.657	7.605	7.559	7.396	7.143	7.057	6.969	6.925
7	12.246	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.538	6.469	6.410	6.359	6.314	6.155	5.908	5.824	5.737	5.694
8	11.259	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.734	5.667	5.609	5.559	5.515	5.359	5.116	5.032	4.946	4.903
9	10.562	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.178	5.111	5.055	5.005	4.962	4.808	4.567	4.483	4.398	4.354
10	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.772	4.706	4.650	4.601	4.558	4.405	4.165	4.082	3.996	3.953
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539	4.462	4.397	4.342	4.293	4.251	4.099	3.860	3.776	3.690	3.647
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296	4.220	4.155	4.100	4.052	4.010	3.858	3.619	3.535	3.449	3.405
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100	4.025	3.960	3.905	3.857	3.815	3.665	3.425	3.341	3.255	3.210
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939	3.864	3.800	3.745	3.698	3.656	3.505	3.266	3.181	3.094	3.050
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805	3.730	3.666	3.612	3.564	3.522	3.372	3.132	3.047	2.959	2.914
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691	3.616	3.553	3.498	3.451	3.409	3.259	3.018	2.933	2.845	2.799
17	8.400	6.112	5.185	4.669	4.336	4.101	3.927	3.791	3.682	3.593	3.518	3.455	3.401	3.353	3.312	3.162	2.920	2.835	2.746	2.700
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508	3.434	3.371	3.316	3.269	3.227	3.077	2.835	2.749	2.660	2.613
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434	3.360	3.297	3.242	3.195	3.153	3.003	2.761	2.674	2.584	2.537
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.294	3.231	3.177	3.130	3.088	2.938	2.695	2.608	2.517	2.470
21	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310	3.236	3.173	3.119	3.072	3.030	2.880	2.636	2.548	2.457	2.409
22	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258	3.184	3.121	3.067	3.019	2.978	2.827	2.583	2.495	2.403	2.355
23	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3.406	3.299	3.211	3.137	3.074	3.020	2.973	2.931	2.780	2.536	2.447	2.354	2.306
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168	3.094	3.032	2.977	2.930	2.889	2.738	2.492	2.403	2.310	2.261
25	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129	3.056	2.993	2.939	2.892	2.850	2.699	2.453	2.364	2.270	2.220
26	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3.288	3.182	3.094	3.021	2.958	2.904	2.857	2.815	2.664	2.417	2.327	2.233	2.183
27	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3.256	3.149	3.062	2.988	2.926	2.872	2.824	2.783	2.632	2.384	2.294	2.198	2.148
28	7.636	5.453	4.568	4.074	3.754	3.528	3.358	3.226	3.120	3.032	2.959	2.896	2.842	2.795	2.753	2.602	2.354	2.263	2.167	2.117
29	7.598	5.420	4.538	4.045	3.725	3.499	3.330	3.198	3.092	3.005	2.931	2.868	2.814	2.767	2.726	2.574	2.325	2.234	2.138	2.087
30	7.562	5.390	4.510	4.018	3.699	3.473	3.305	3.173	3.067	2.979	2.906	2.843	2.789	2.742	2.700	2.549	2.299	2.208	2.111	2.060
40	7.314	5.178	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801	2.727	2.665	2.611	2.563	2.522	2.369	2.114	2.019	1.917	1.862
50	7.171	5.057	4.199	3.720	3.408	3.186	3.020	2.890	2.785	2.698	2.625	2.563	2.508	2.461	2.419	2.265	2.007	1.909	1.803	1.745
60	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632	2.559	2.496	2.442	2.394	2.352	2.198	1.936	1.836	1.726	1.666
80	6.963	4.881	4.036	3.563	3.255	3.036	2.871	2.742	2.637	2.551	2.478	2.415	2.361	2.313	2.271	2.115	1.849	1.746	1.630	1.566
100	6.895	4.824	3.984	3.513	3.206	2.988	2.823	2.694	2.590	2.503	2.430	2.368	2.313	2.265	2.223	2.067	1.797	1.692	1.572	1.504
120	6.851	4.787	3.949	3.480	3.174	2.956	2.792	2.663	2.559	2.472	2.399	2.336	2.282	2.234	2.191	2.035	1.763	1.656	1.533	1.462
240	6.742	4.695	3.864	3.398	3.094	2.878	2.714	2.586	2.482	2.395	2.322	2.260	2.205	2.157	2.114	1.956	1.677	1.565	1.432	1.351

Tabela 6: Limites unilaterais da distribuição F de Fisher-Snedecor ao nível de 2,5% de probabilidade

GL	V1																			
V2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	40	60	120	240
1	16212	19997	21614	22501	23056	23440	23715	23924	24091	24222	24334	24427	24505	24572	24632	24837	25146	25254	25358	25414
2	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.5	199.5	199.5	199.5
3	55.552	49.800	47.468	46.195	45.391	44.838	44.434	44.125	43.881	43.685	43.525	43.387	43.270	43.172	43.085	42.779	42.310	42.150	41.990	41.910
4	31.332	26.284	24.260	23.154	22.456	21.975	21.622	21.352	21.138	20.967	20.824	20.705	20.603	20.515	20.438	20.167	19.751	19.611	19.469	19.397
5	22.785	18.314	16.530	15.556	14.939	14.513	14.200	13.961	13.772	13.618	13.491	13.385	13.293	13.215	13.146	12.903	12.530	12.402	12.274	12.209
6	18.635	14.544	12.917	12.028	11.464	11.073	10.786	10.566	10.391	10.250	10.133	10.034	9.950	9.878	9.814	9.589	9.241	9.122	9.001	8.941
7	16.235	12.404	10.883	10.050	9.522	9.155	8.885	8.678	8.514	8.380	8.270	8.176	8.097	8.028	7.968	7.754	7.422	7.309	7.193	7.135
8	14.688	11.043	9.597	8.805	8.302	7.952	7.694	7.496	7.339	7.211	7.105	7.015	6.938	6.872	6.814	6.608	6.288	6.177	6.065	6.008
9	13.614	10.107	8.717	7.956	7.471	7.134	6.885	6.693	6.541	6.417	6.314	6.227	6.153	6.089	6.032	5.832	5.519	5.410	5.300	5.244
10	12.827	9.427	8.081	7.343	6.872	6.545	6.303	6.116	5.968	5.847	5.746	5.661	5.589	5.526	5.471	5.274	4.966	4.859	4.750	4.695
11	12.226	8.912	7.600	6.881	6.422	6.102	5.865	5.682	5.537	5.418	5.320	5.236	5.165	5.103	5.049	4.855	4.551	4.445	4.337	4.281
12	11.754	8.510	7.226	6.521	6.071	5.757	5.524	5.345	5.202	5.085	4.988	4.906	4.836	4.775	4.721	4.530	4.228	4.123	4.015	3.960
13	11.374	8.186	6.926	6.233	5.791	5.482	5.253	5.076	4.935	4.820	4.724	4.643	4.573	4.513	4.460	4.270	3.970	3.866	3.758	3.703
14	11.060	7.922	6.680	5.998	5.562	5.257	5.031	4.857	4.717	4.603	4.508	4.428	4.359	4.299	4.247	4.059	3.760	3.655	3.547	3.492
15	10.798	7.701	6.476	5.803	5.372	5.071	4.847	4.674	4.536	4.424	4.329	4.250	4.181	4.122	4.070	3.883	3.585	3.480	3.372	3.317
16	10.576	7.514	6.303	5.638	5.212	4.913	4.692	4.521	4.384	4.272	4.179	4.099	4.031	3.972	3.920	3.734	3.437	3.332	3.224	3.168
17	10.384	7.354	6.156	5.497	5.075	4.779	4.559	4.389	4.254	4.142	4.050	3.971	3.903	3.844	3.793	3.607	3.311	3.206	3.097	3.041
18	10.218	7.215	6.028	5.375	4.956	4.663	4.445	4.276	4.141	4.030	3.938	3.860	3.793	3.734	3.683	3.498	3.201	3.096	2.987	2.931
19	10.073	7.093	5.916	5.268	4.853	4.561	4.345	4.177	4.043	3.933	3.841	3.763	3.696	3.638	3.587	3.402	3.106	3.000	2.891	2.834
20	9.944	6.987	5.818	5.174	4.762	4.472	4.257	4.090	3.956	3.847	3.756	3.678	3.611	3.553	3.502	3.318	3.022	2.916	2.806	2.749
21	9.829	6.891	5.730	5.091	4.681	4.393	4.179	4.013	3.880	3.771	3.680	3.602	3.536	3.478	3.427	3.243	2.947	2.841	2.730	2.673
22	9.727	6.806	5.652	5.017	4.609	4.322	4.109	3.944	3.812	3.703	3.612	3.535	3.469	3.411	3.360	3.176	2.880	2.774	2.663	2.605
23	9.635	6.730	5.582	4.950	4.544	4.259	4.047	3.882	3.750	3.642	3.551	3.474	3.408	3.351	3.300	3.116	2.820	2.713	2.602	2.543
24	9.551	6.661	5.519	4.890	4.486	4.202	3.991	3.826	3.695	3.587	3.497	3.420	3.354	3.296	3.246	3.062	2.765	2.658	2.546	2.488
25	9.475	6.598	5.462	4.835	4.433	4.150	3.939	3.776	3.645	3.537	3.447	3.370	3.304	3.247	3.196	3.013	2.716	2.609	2.496	2.437
26	9.406	6.541	5.409	4.785	4.384	4.103	3.893	3.730	3.599	3.492	3.402	3.325	3.259	3.202	3.151	2.968	2.671	2.563	2.450	2.391
27	9.342	6.489	5.361	4.740	4.340	4.059	3.850	3.687	3.557	3.450	3.360	3.284	3.218	3.161	3.110	2.927	2.630	2.522	2.408	2.348
28	9.284	6.440	5.317	4.698	4.300	4.020	3.811	3.649	3.519	3.412	3.322	3.246	3.180	3.123	3.073	2.890	2.592	2.483	2.369	2.309
29	9.230	6.396	5.276	4.659	4.262	3.983	3.775	3.613	3.483	3.376	3.287	3.211	3.145	3.088	3.038	2.855	2.557	2.448	2.333	2.273
30	9.180	6.355	5.239	4.623	4.228	3.949	3.742	3.580	3.451	3.344	3.255	3.179	3.113	3.056	3.006	2.823	2.524	2.415	2.300	2.239
40	8.828	6.066	4.976	4.374	3.986	3.713	3.509	3.350	3.222	3.117	3.028	2.953	2.888	2.831	2.781	2.598	2.296	2.184	2.064	1.999
50	8.626	5.902	4.826	4.232	3.849	3.579	3.376	3.219	3.092	2.988	2.900	2.825	2.760	2.703	2.653	2.470	2.164	2.050	1.925	1.858
60	8.495	5.795	4.729	4.140	3.760	3.492	3.291	3.134	3.008	2.904	2.817	2.742	2.677	2.620	2.570	2.387	2.079	1.962	1.834	1.764
80	8.335	5.665	4.611	4.028	3.652	3.387	3.188	3.032	2.907	2.803	2.716	2.641	2.577	2.520	2.470	2.286	1.974	1.854	1.720	1.646
100	8.241	5.589	4.542	3.963	3.589	3.325	3.127	2.972	2.847	2.744	2.657	2.583	2.518	2.461	2.411	2.227	1.912	1.790	1.652	1.573
120	8.179	5.539	4.497	3.921	3.548	3.285	3.087	2.933	2.808	2.705	2.618	2.544	2.479	2.423	2.373	2.188	1.871	1.747	1.606	1.524
240	8.027	5.417	4.387	3.816	3.447	3.187	2.991	2.837	2.713	2.610	2.524	2.450	2.385	2.329	2.278	2.093	1.770	1.640	1.488	1.396

Tabela 7: Limites unilaterais da distribuição F de Fisher-Snedecor ao nível de 1,0% de probabilidade