UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
CADERNOS DE MATEMÁTICA E ESTATÍSTICA
SÉRIE B: TRABALHO DE APOIO DIDÁTICO

ELEMENTOS BÁSICOS DE ESTATÍSTICA INFERENCIAL PARA OS ALUNOS DE MAT02215

GLÁUCIA MICHEL DE OLIVA

SÉRIE B, N° 71
PORTO ALEGRE, ABRIL DE 2007

LISTA DE FIGURAS

LISTA DE TABELAS

CAPÍTULO 1. INTRODUÇÃO

A crescente abordagem quantitativa utilizada no estudo e desenvolvimento das mais diversas áreas de conhecimento vem revestindo cada vez mais a estatística de suma importância no mundo científico. As técnicas estatísticas aliadas a programas adequados de computação constituem valiosos instrumentos de análise e interpretação da realidade manifesta.

Os métodos estatísticos são especialmente apropriados para estudos que dispõem de dados sujeitos a variabilidades. É fácil perceber que se um fenômeno particular fosse completamente uniforme (constante), uma única observação seria suficiente para investigar a sua natureza. No entanto, esse não é o caso da maioria dos fenômenos que não podem ser controlados experimentalmente. Assim, é necessário utilizar e/ou desenvolver técnicas pelas quais se possa testar hipóteses sobre fenômenos das mais diversas áreas.

Resumidamente, pode-se dizer que a estatística é uma particular maneira de ler o mundo, de tentar organizá-lo com o fim de compreendê-lo. É particular porque é essencialmente numérica. A Estatística não é uma ciência; é, antes, um conjunto de técnicas. É um método. De uma maneira geral, estuda-se estatística para: (i) tirar informações significativas a partir de dados brutos; (ii) fazer inferências sobre a natureza de uma população com base em observações de uma amostra; (iii) realizar previsões sobre as taxas de ocorrências de eventos aleatórios e (iii) aplicar seus conceitos como auxílio às tomadas de decisão diante de incertezas, justificando cientificamente as decisões.

As técnicas estatísticas, aliadas a programas adequados de computador, constituem valiosos instrumentos de análise e interpretação da realidade manifesta.

O objetivo deste Caderno de Estatística é apresentar os elementos básicos de estatística para os alunos de Estatística Geral II, MAT 2215, de modo a facilitar o acompanhamento dos trabalhos desenvolvidos em sala de aula.

1.1 POPULAÇÃO E AMOSTRA

O estudo da estatística é o estudo da chamada população, que consiste na totalidade de unidades de observação. Essas unidades de observações podem ser pessoas, objetos ou eventos, a partir dos quais se pretende tomar uma decisão. É importante diferenciar a unidade de observação das características desta população.

Uma unidade de observação é um objeto do qual se coletam dados, e que pode ter muitas características, embora o interesse geralmente recaia sobre apenas uma ou poucas dessas características. Em termos estatísticos população é o conjunto dos valores atribuídos à determinada característica que é comum a todos os elementos sob estudo. Ou seja, é o conjunto de todos os possíveis valores de uma variável relativamente a um universo de referência. O levantamento exaustivo dos valores de uma população chama-se censo.

A experiência diária mostra a necessidade de se fazer generalizações sobre assuntos e circunstâncias. Esses assuntos e circunstâncias freqüentemente ultrapassam as fronteiras do quotidiano, indicando que, dentro de certos limites, o conhecimento de uma parte do todo é uma informação prática e útil que pode ser aplicável à totalidade. Essa parte representativa do todo é denominada amostra. Uma amostra é um subconjunto de medições selecionado da população de interesse.

O procedimento de obtenção de uma amostra chama-se levantamento por amostragem. As características numéricas de uma população chamam-se parâmetros, enquanto que estatísticas são características de uma amostra. Normalmente, as estatísticas são utilizadas como base para se estimar os parâmetros populacionais. Costuma-se, por convenção, indicar as estatísticas com letras do alfabeto latino, utilizando-se letras gregas para os respectivos parâmetros. A Figura 1 exemplifica alguns símbolos utilizados para diferenciar estatísticas de parâmetros.

Característica	**Na população**	**Na amostra**
Média	μ	$\overline{x}$
Desvio padrão	σ	**S**
Proporção	π	**p**
Coef. Correlação	ρ	**r**

Figura 1 Símbolos usados para as características numéricas

1.2 ESTATÍSTICA DESCRITIVA E INFERENCIAL

Conforme estejamos interessados em descrever os dados observados ou em realizar inferências a partir de uma amostra, devemos utilizar técnicas da estatística descritiva ou da indutiva.

A Estatística descritiva tem por objetivo descrever a realidade observada (população ou amostra), usando métodos numéricos e gráficos, realizando comentários simples de maneira mais informativa possível. A Estatística Descritiva se ocupa da descrição de um conjunto de dados, sejam eles amostrais ou populacionais.

A indutiva se ocupa em fazer inferências (isto é, estimativas, previsões, tomadas de decisões) sobre um conjunto populacional com base nas informações contidas na amostra. Ou seja, através de técnicas inferenciais conclui-se para o todo, a partir da observação de uma parte; as conclusões transcendem a parte manifesta. Assim, os métodos de estatística inferencial só podem ser utilizados, com algum sentido, sobre dados amostrais.

1.3 VARIÁVEIS FREQÜÊNCIAS E GRÁFICOS

Variável em estatística, é a atribuição de um número a cada característica da unidade de observação, ou seja, é uma função matemática definida na população. É importante que o pesquisador saiba distinguir entre variáveis quantitativas e qualitativas.

Uma outra distinção importante é entre variáveis discretas e variáveis contínuas. As variáveis discretas são aquelas associadas a processos de contagem (por exemplo, número de empregados de uma empresa), enquanto as contínuas são as derivadas de procedimentos de mensuração (por exemplo, as alturas de um grupo de pessoas). A diferença fundamental entre elas é que, ao contrário das discretas, as continuas podem assumir um número infinito de valores para qualquer intervalo dado, por menor do que ele seja.

Convém salientar que o fato de uma variável ser expressa por números não significa que ela seja necessariamente quantitativa porque a classificação da variável depende de como foi medida (escala nominal, ordinal, Intervalar e razões), e não da maneira como se manifesta.

A freqüência de um determinado valor que uma variável assume é o número de vezes que tal valor ocorre em um dado conjunto de observações. A freqüência relativa é a proporção com que o valor ocorre com respeito ao número total de observações. Em condições bem gerais, a freqüência relativa pode ser também interpretada como a probabilidade de tal valor ocorrer. A freqüência acumulada de um dado valor é o número de observações menores ou iguais ao valor considerado. A distribuição de freqüências de uma variável é o conjunto das freqüências de todos os diferentes valores observados da variável. O perfil da distribuição de freqüências identifica a forma (o padrão, o tipo) de distribuição da própria variável.

A Figura 2 apresenta um exemplo de distribuição de freqüência discreta e um de contínua

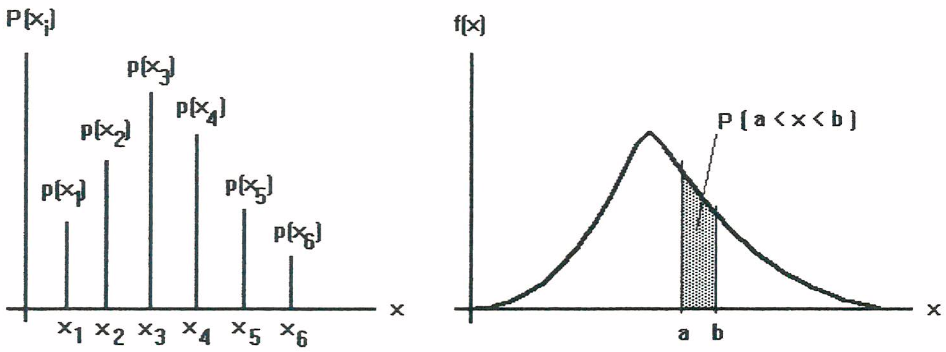


Figura 2 Distribuição de Freqüências: variáveis discretas e contínuas

Quando a variável for contínua ou quando houver um grande número de valores observados, devem-se agrupar as observações em intervalos de classe. Nesse caso, as freqüências estarão associadas ao intervalo como um todo e não mais a valores individuais. Para efeitos de cálculos, supõe-se que todos os valores observados dentro de um mesmo intervalo sejam iguais ao respectivo ponto médio.

Intervalos de classe	Freqüência absoluta	Freqüência relativa	Freqüência percentual %	Freqüência acumulada
19,5 – 22,5	12	0,30	30	12
22,5 – 25,5	3	0,07	7	15
25,5 – 28,5	9	0,23	23	24
28,5 – 31,5	5	0,13	13	29
31,5 – 34,5	6	0,15	15	35
34,5 – 37,5	2	0,05	5	37
37,5 – 40,5	3	0.07	7	40
Total	**40**	**1.00**	**100**	

Figura 3 Distribuição de Freqüências de dados agrupados

A representação gráfica mais adequada de uma distribuição de freqüências é feita quer pelo histograma quer pelo polígono de freqüências. O Histograma é um diagrama de colunas tal que a área de cada retângulo (coluna) seja proporcional à freqüência (absoluta ou relativa) da respectiva classe. No caso, de todos os intervalos de classe apresentarem amplitudes iguais, a altura de cada coluna fica diretamente proporcional à freqüência. O polígono de freqüências é o polígono que se constrói ligando os pontos obtidos elevando-se os pontos médios de cada classe à altura da coluna correspondente do histograma. Ogiva representa uma distribuição de freqüências acumuladas (absolutas ou relativas).

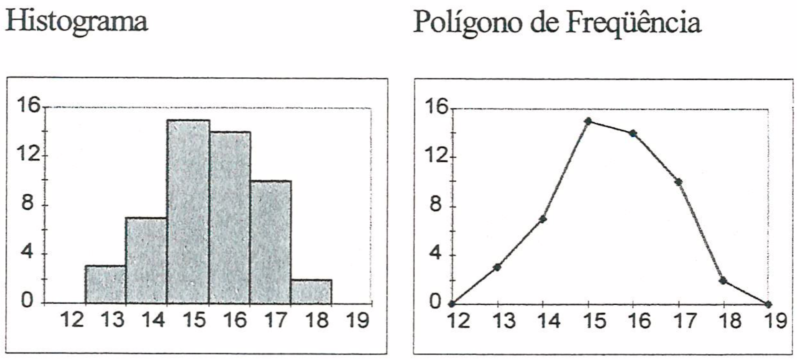


Figura 4 Histograma e Polígono de Freqüências

1.4 MEDIDAS DE TENDÊNCIA CENTRAL

As principais medidas de tendência central são a moda, a mediana e a média aritmética (ou simplesmente a média). A moda é o valor que ocorre com maior freqüência. A mediana é o valor que divide o conjunto de observações exatamente ao meio, isto é, de tal maneira que o número de observações maiores do que a mediana seja igual ao número de observações menor do que a mediana. Já a média aritmética é igual à soma de todos os valores observados dividida pelo número de observações. Ela é o centro de gravidade da distribuição da variável.

A moda é a medida de tendência central mais simples; é aplicável a qualquer variável quantitativa ou qualitativa. Trata-se, porém, de uma medida extremamente sensível a pequenas alterações nos valores observados. Além disso, há casos em que ou ela não existe (como no conjunto 8, 8, 9 , 9 , 15 ,15).

A aplicação da mediana requer que a variável possa ser ordenável. Uma variável como sexo (masculino/feminino), por exemplo, comporta apenas a moda, mas não a mediana. Uma desvantagem da mediana é ser muito afetada por ligeiras variações nos valores centrais da variável. Em compensação, ela permanece inalterada para mudanças ainda que grandes nos valores extremos da distribuição.

Já a aplicação da média requer que a variável seja quantitativa. Ao contrário da mediana, a média é pouco sensível a variações nos valores centrais, enquanto que - e esta é sua maior desvantagem - é grandemente afetada pelos valores extremos da variável.

A comparação entre média, mediana e moda permite identificar a condição de simetria da distribuição. De fato, em distribuições:

a) simétricas: média = mediana = moda
b) com assimetria positiva ou para a direita:moda < mediana < média
c) com assimetria negativa ou para a esquerda: média < mediana < moda.

1.4.1 Cálculo das Medidas de tendência central

Moda (Mo): é o valor que ocorre com maior freqüência em uma distribuição.

Moda para dados não-agrupados:

$$Mo=\frac{d_1}{d_1+d_2}h_i+l_i \qquad (1)$$

onde "i" indica a classe modal e d1 = fi - fi-1 e d2 = fi - fi+1.

Mediana, (Md): é o valor que se encontra no meio ou no centro quando se ordenam os valores em ordem crescente, i.e., o valor que divide o conjunto de observações de tal maneira que o número de observações maiores do que a mediana seja igual ao número de observações menor do que a mediana.

Mediana para dados em séries: Md = x tal que F'x = 0,5

$$Md = l_i + \frac{\left(\frac{n}{2} - F_{i-1}\right)}{f_i} h_i \quad (2)$$

Média aritmética

A média aritmética da amostra é dada pela fórmula (3) e a média da amostra é dada pela fórmula a seguir:

Média dados em série:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \quad (3)$$

Média dados agrupados:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i f_i \quad (4)$$

População: dados em série;

$$\mu_X = \frac{1}{N} \sum_{i=1}^{N} X_i \quad (5)$$

População: dados agrupados;

$$\mu_X = \frac{1}{N}\sum_{i=1}^{N} X_i f_i \qquad (6)$$

A média aritmética é o centro de gravidade da distribuição, i.e.,

$$\sum (x-\mu)^2 = min \qquad (7)$$

se y=a+bx, então:

$$\mu_y = a + b\mu_x \qquad (8)$$

1.5 MEDIDAS DE DISPERSÃO (OU DE VARIABILIDADE)

As principais medidas de variabilidade são a amplitude, a variância e o desvio padrão. A amplitude R é a diferença entre os valores máximo e mínimo da distribuição. A variância (σ^2 *ou* S^2) é a média dos quadrados dos desvios em relação à média, e o desvio padrão (σ *ou* S) é a raiz quadrada positiva da variância.

1.5.1 Cálculo das Medidas de Variabilidade

Variância para dados em série:

População:

$$\sigma_X^2 = \frac{1}{N}\sum_{i=1}^{N} (X_i - \mu_X)^2 \qquad (9)$$

Amostra:

$$S_X^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^2 \qquad (10)$$

Variância para dados agrupados:

População:

$$\sigma_X^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu_X)^2 f_i \qquad (11)$$

Amostra:

$$S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 f_i \qquad (12)$$

Desvio padrão:

$$\sigma = +\sqrt{\sigma^2} \quad \text{ou} \quad S = +\sqrt{S^2} \qquad (13)$$

Variância Relativa:

$$\gamma^2 = \frac{\sigma^2}{\mu^2} \quad \text{ou} \quad g^2 = \frac{S^2}{\overline{X}^2} \qquad (14)$$

Coeficiente de variação:

$$\gamma = +\sqrt{\gamma^2} \qquad \text{ou} \quad g = +\sqrt{g^2} \qquad (15)$$

1.6 DESIGUALDADE DE TCHEBYSHEV

Dado um número k maior do que 1, pelo menos $(1 - 1/k^2)$ das observações de uma variável qualquer se encontram necessariamente dentro do intervalo definido pela média da variável menos/mais k desvios padrão:

a) 68% das observações, para k = 1
b) 95% das observações, para k = 2

c) quase a totalidade das observações, para k = 3

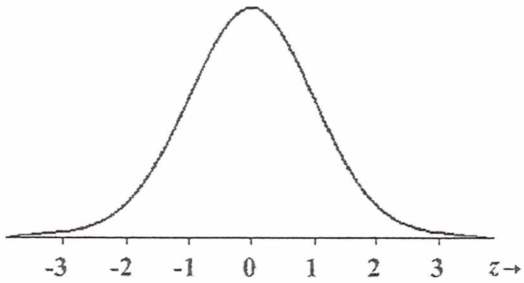


Figura 5 Distribuição Simétrica

1.7 ESTATÍSTICA DESCRITIVA NO EXCEL

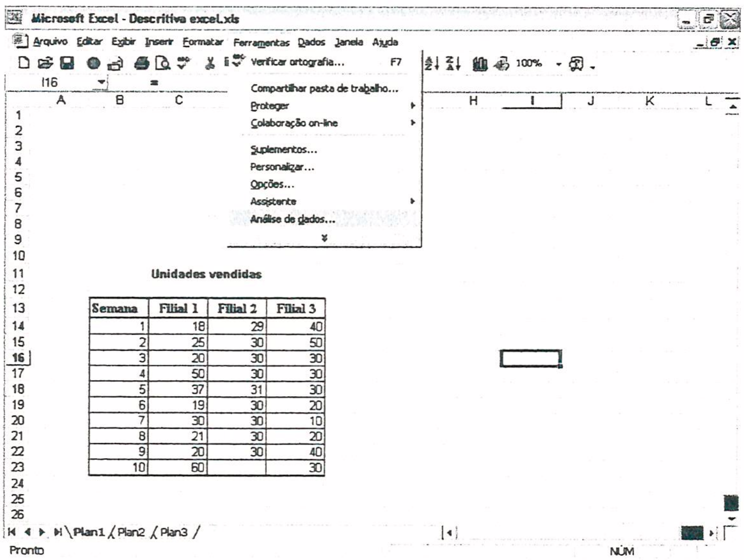
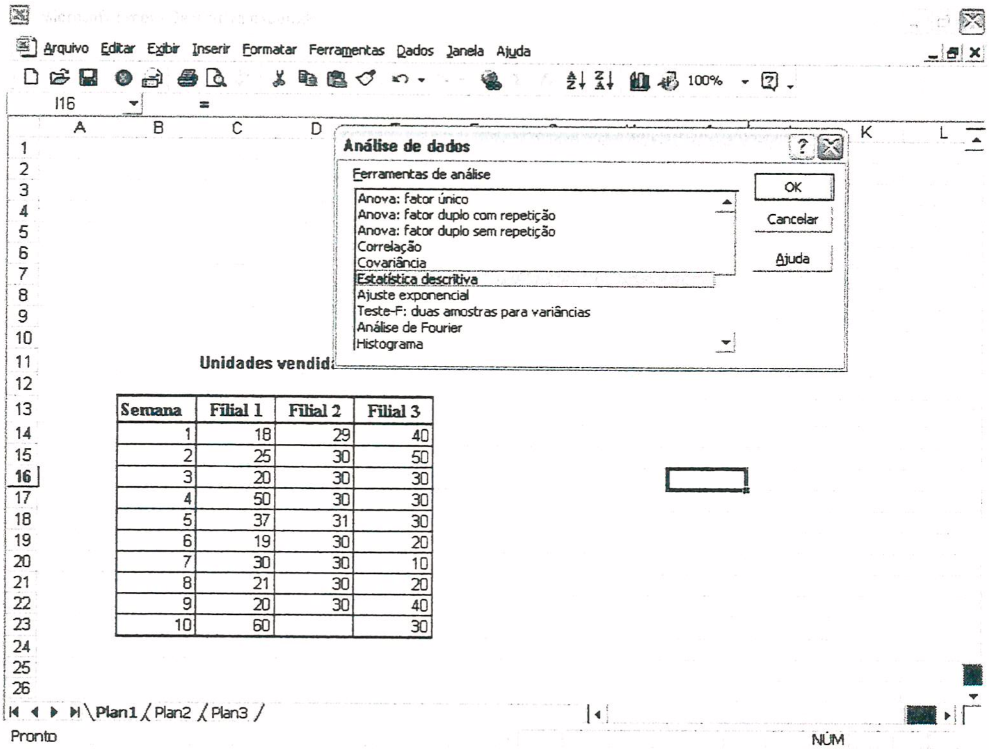
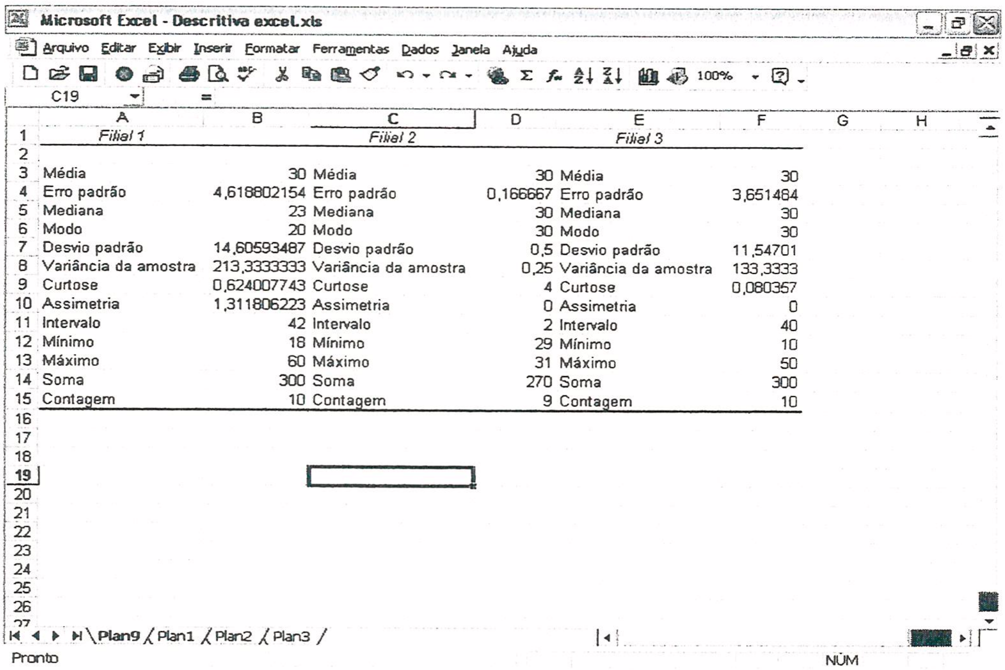


Figura 6 Estatística descritiva no Excel – passo 1



Semana	Filial 1	Filial 2	Filial 3
1	18	29	40
2	25	30	50
3	20	30	30
4	50	30	30
5	37	31	30
6	19	30	20
7	30	30	10
8	21	30	20
9	20	30	40
10	60		30

Figura 7 Estatística Descritiva no Excel – passo 2



Filial 1		*Filial 2*		*Filial 3*	
Média	30	Média	30	Média	30
Erro padrão	4,618802154	Erro padrão	0,166667	Erro padrão	3,651484
Mediana	23	Mediana	30	Mediana	30
Modo	20	Modo	30	Modo	30
Desvio padrão	14,60593487	Desvio padrão	0,5	Desvio padrão	11,54701
Variância da amostra	213,3333333	Variância da amostra	0,25	Variância da amostra	133,3333
Curtose	0,624007743	Curtose	4	Curtose	0,080357
Assimetria	1,311806223	Assimetria	0	Assimetria	0
Intervalo	42	Intervalo	2	Intervalo	40
Mínimo	18	Mínimo	29	Mínimo	10
Máximo	60	Máximo	31	Máximo	50
Soma	300	Soma	270	Soma	300
Contagem	10	Contagem	9	Contagem	10

Figura 8 Resultados obtidos no Excel

1.8 QUESTÕES PARA ESTUDO

1) Foi realizado um estudo para investigar o efeito de ter música ambiental em diversas variáveis que medem o funcionamento de supermercados. Foram entrevistados 200 clientes de um supermercado para determinar se preferem música ambiental enquanto realizam suas compras.

    a) descreva a população e a amostra associadas a esta pesquisa;
    b) é a percentagem de clientes da amostra que preferem música ambiental igual a percentagem da população de clientes com a mesma preferência?

2) Foi feito um estudo para determinar se as donas de casa de Porto Alegre preferem uma determinada marca de detergente. De 50 donas de casa entrevistadas, 30 disseram que preferem esta marca em relação às demais.

    d) descrever a população e a amostra;
    e) qual a estatística e qual o parâmetro?
    f) dizer qual o método utilizado para fazer as seguintes afirmações:

    A1: "a proporção amostral é 0,6";

    ***A2; "a proporção da população está entre 0,50 e 0,70 com uma probabilidade determinada de que se encontra fora destes limites"*.**

3) As predições econômicas baseiam-se às vezes em uma amostra realizada entre economistas e executivos nacionalmente conhecidos. Uma delas amostrou a opinião de 38 economistas e executivos. Como parte da pesquisa, pediu-se a cada um sua previsão pessoal da taxa média mensal de inflação para os próximos cinco meses. A média destas 38 opiniões foi uma taxa de 0,4 % ao mês.

    g) descreva a população da qual se selecionou a amostra;

h) se conhecêssemos todas as previsões da população amostrada, poderíamos determinar a taxa média mensal de inflação para os próximos cinco meses?

4) Dê exemplos de variáveis discretas e de variáveis contínuas. O faturamento mensal de uma empresa é uma variável de que tipo?

5) Considere que você vai fazer uma amostra de 500 eleitores de Porto Alegre, para saber a intenção de voto de cada um para as eleições do próximo ano. De que tipo é a variável "candidato preferido"? Em que sentido o número de eleitores que responderão que irão votar no candidato X pode ser considerado uma variável? Será uma variável discreta ou contínua? E a proporção dos que declaram, na amostra, que irão votar em X pode ser considerada uma variável? De que tipo?

6) Em que condições não teria importância utilizar a média, a mediana ou a moda como medida de tendência central? Por quê? dar uma situação em que seja mais apropriado utilizar como medida de tendência central: a) a mediana; b) a moda.

7) Em certas circunstâncias, a chamada média geométrica é uma medida de tendência central mais apropriada do que a média, a mediana ou a moda. Costuma-se representá-la por G, e é definida como a raiz n-ésima do produto de um conjunto de n valores. Uma aplicação importante de G é prometida proporções de variação. Suponhamos, por exemplo, que as vendas de uma companhia aumentaram de \$2 milhões em 1994 para \$4 milhões em 1995 e para \$6 milhões em 1996. O aumento foi de 200% de 1994 a 1996, com média aritmética de 100% ao ano. Mas esse valor é evidentemente errôneo, já

que a proporção média de aumento anual foi muito menor. Nesta situação devemos utilizar a média geométrica das proporções de variação, tomadas como fatores de variação e não em termos de taxas percentuais. Calcular a proporção média anual do aumento de vendas.

8) Outra medida de tendência central importante é a média harmônica. É geralmente indicada por H, e é igual ao inverso da média aritmética dos inversos dos valores das observações. É utilizada quando as observações estão expressas na dimensão inversa da média buscada, ou seja, quando se quer a média da variável de tal forma que seja considerada constante a dimensão do numerados (e não a do denominador, como faz a média aritmética). Suponha que uma pessoa gastou $10,00 em cada uma de três lojas. na primeira comprou meias a 42,00 o par, na segunda a $2,50 o par e na terceira a $5,00 o par. Qual o preço médio que pagou pelo par de meias?

9) O número de horas que se assiste televisão, por família, e o horário com maior número de telespectadores, são dois fatores que influem nos preços de publicidade. Uma amostra aleatória de 25 famílias em uma região particular produziu os seguintes resultados do tempo que cada família dedica assistir televisão:

3	8,5	5,5	8	10
6	4	6	3,5	6,5
7,5	5	3	9	1
12	12	3,5	2	7,5
15	5	5	6,5	1

Calcule a percentagem de horas de assistência à televisão por família que se encontram no intervalo $\bar{X} \pm 2$ S. Comparar a resposta com a percentagem correspondente dada pela regra empírica.

10) Uma cadeia de supermercados está considerando abrir uma loja em um bairro ainda não atendido por este tipo de comércio. faça uma lista de fatores que os tomadores de decisão deveriam considerar ao decidir se a loja deve ou não ser aberta. Para cada fator, identifique as estatísticas pertinentes que os tomadores de decisão gostariam de considerar, analisando, em cada caso, as dificuldades de obtê-las.

CAPÍTULO 2. PROBABILIDADES

A idéia de probabilidade pode ser entendida a partir da noção de experimento aleatório. Um experimento é aleatório quando seu resultado se dá ao acaso, ao azar. Sua característica básica é a de que não é possível antecipar com exatidão o seu resultado. Um experimento aleatório, por conseguinte, é um experimento que se realiza em condições de incerteza. A probabilidade procura ser uma medida da incerteza que temos sobre a ocorrência de um determinado evento (por exemplo, a ocorrência de duas caras) em um dado experimento aleatório (por exemplo, o lançamento de duas moedas). Existem divergências importantes sobre o que seria uma boa definição de probabilidades. Existem, a respeito, pelo menos quatro conceitos distintos, a saber:

a) conceito clássico: razão de casos favoráveis sobre casos possíveis.
b) conceito empírico: limite da freqüência relativa da ocorrência do evento.
c) conceito subjetivista: grau de crença na verdade de uma afirmação.
d) conceito axiomático: uma função que respeita certos axiomas.

2.1 REGRA DA ADIÇÃO

A probabilidade de ocorrer um evento A ou um evento b é dada por:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{16}$$

observações:

a) o "ou" é inclusivo
b) se A e B são mutuamente excludentes, então

$$P(A \cup B) = P(A) + P(B) \tag{17}$$

2.2 PROBABILIDADE CONDICIONADA:

A probabilidade condicionada de ocorrer o evento A, tendo ocorrido B, é:

$$P(A \cap B) = P(A)\, P(B / A) \tag{18}$$

2.3 INDEPENDÊNCIA

Dois eventos A e B são independentes se e somente se:

$$P(A / B) = P(A) \tag{19}$$

2.4 REGRA DA MULTIPLICAÇÃO

Da regra da probabilidade condicionada deriva-se a fórmula de cálculo para a probabilidade da ocorrência conjunta de dois eventos A e B, conhecida como a Regra da Multiplicação:

$$P(A \cap B) = P(A)\,P(B/A) \quad (20)$$

caso particular em que A e B forem independentes, tem-se

$$P(A \cap B) = P(A)\,P(B) \quad (21)$$

2.4.1 QUESTÕES PARA ESTUDO

1) Uma pesquisa de mercado classificou os clientes de uma grande loja segundo o sexo e segundo a residência, se na cidade ou nos subúrbios. As proporções de clientes em cada das quatro categorias encontram-se na tabela abaixo:

Residência	Sexo Masculino	Sexo Feminino
Subúrbios	0,17	0,67
Cidade	0,04	0,12

Seleciona-se ao acaso um cliente deste grupo. Encontrar a probabilidade de que ele:

a) resida nos subúrbios;
b) seja mulher e viva na cidade;
c) seja homem.

2) Uma loja que vende roupas pelo correio comercializa duas linhas de artigos, uma relativamente cara (Linha 1) e outra mais barata (Linha 2). Uma amostra de 1.000 pedidos produziu os seguintes números:

Sexo	Linha 1	Linha 2	Total
Masculino	132	147	279
Feminino	516	205	721
Total	648	352	1.000

Seleciona-se ao acaso um destes 1.000 pedidos:

a) Calcular a probabilidade do evento A: o consumidor é mulher;
b) Calcular a probabilidade do evento B: o pedido é para a Linha 1;
c) Calcular a probabilidade do evento C: a ocorrência conjunta de A e B;
d) Calcular a probabilidade de que o pedido seja para a Linha 1, dado que o consumidor é mulher;
e) Verificar se A e B são ou não independentes.

3) (Árvore de Eventos) Para se estudar o comportamento do mercado de automóveis, as marcas foram divididas em 3 categorias, a saber, marca F, marca V, e marca X, esta última reunindo todas as marcas que não são F ou V. Um estudo sobre o hábito de mudança de marca mostrou o seguinte quadro de probabilidades:

Possuidor de carro da marca	Probabilidade de mudança para		
	V	F	X
V	0,50	0,25	0,25
F	0,15	0,70	0,15
X	0,30	0,30	0,40

O primeiro carro que um indivíduo compra, o faz segundo as seguintes probabilidades: marca V com 50%, marca F com 30%, e marca X com 20%. Considere um indivíduo qualquer:

a) qual a probabilidade de seu segundo carro ser da marca V?
b) se o segundo carro é V, qual a probabilidade de o primeiro também ter sido V?

CAPÍTULO 3. VARIÁVEIS ALEATÓRIAS

Uma variável aleatória e uma variável que tem associada a si as probabilidades de assumir seus diferentes valores (no caso de a variável ser discreta) ou se encontrar dentro de intervalos determinados (no caso de a variável ser contínua). Uma variável aleatória também possui média e variância. A sua média, também chamado valor esperado ou esperança matemática, e que se representa por E(X), é a média aritmética dos valores assumidos pela variável ponderados por suas respectivas probabilidades. Quando a variável é discreta, tais probabilidades são dadas pela chamada função de probabilidade.

3.1 VARIÁVEIS ALEATÓRIAS DISCRETAS (V.A.D.)

3.1.1 Distribuição Binomial

a) experimento consta de n provas idênticas;
b) cada prova pode ter um de dois resultados, convencionalmente chamados de sucesso e fracasso.
c) a probabilidade de um sucesso em uma prova é igual a p e permanece constante (as provas são independentes). A probabilidade de um fracasso é (1-p) = q

a) interessa conhecer X, o número de sucessos observados em n provas.

Assim, se X é uma v.a.d. com distribuição binomial:

$$P(X=x)=C_n^x\, p^x q^{n-x} \tag{22}$$

com parâmetros:

$$E(x) = np \qquad e \qquad Var(x) = npq \tag{23}$$

3.1.2 Distribuição Multinomial

É uma extensão direta da binomial para o caso em que estamos interessados no número de ocorrência, em n provas independentes, de cada um dos k eventos mutuamente excludentes que definem uma partição do espaço amostral. Seja A1, A2, ...,Ak uma partição do espaço amostral e seja pi a probabilidade do evento Ai, tal que p1 + p2+ ... + pk = 1. Então a probabilidade de que A1 ocorra X1 vezes, que A2 ocorra X2 vezes, ... , que Ak ocorra Xk vezes, em n provas independentes, é:

$$P(x_1, x_2, ..., x_k) = \frac{n!}{x_1! x_2! ... x_k!} p_1^{x_1} p_2^{x_2} ... p_k^{x_k} \quad (24)$$

Com parâmetros dados por:

$$E(x) = np \qquad e \qquad Var(x) = npq \quad (25)$$

3.1.3 Distribuição Geométrica

Quando, em um experimento binomial, interessa o número X de tentativas (provas) até a observação do primeiro sucesso (observe que o número de tentativas poderia seguir indefinidamente até a observação do primeiro sucesso e que, portanto, X é um exemplo de v.a.d. que pode assumir um número infinito, mas enumerável, de valores).

-se X é v.a.d com distribuição geométrica:

$$P(X) = p\, q^{x-1} \quad (26)$$

3.1.4 Distribuição de Poisson

Aplica-se a distribuição de Poisson:

a) a processos em que os eventos são observados (contados) por unidade de tempo ou de espaço (i.e, no continuum).

b) como distribuição limite da binomial, em situações nas quais a probabilidade p é muito pequena e n é grande (para os chamados eventos raros; a lei de Poisson é também conhecida como Lei dos Pequenos Números).

Assim, se X é v.a.d. com distribuição de Poisson,

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{27}$$

Com parâmetros dados por:

$$E(x) = \lambda \ e \ \sigma^2(x) = \lambda \tag{28}$$

3.1.5 Distribuição Hipergeométrica

Em uma binomial, quando a amostragem é feita sem reposição e o tamanho da amostra n é grande relativamente ao tamanho da população N, então o número X de sucessos nas n provas (n observações) terá distribuição hipergeométrica.

3.2 VARIÁVEIS ALEATÓRIAS CONTÍNUAS (V.A.C.)

3.3 FUNÇÃO DENSIDADE E FUNÇÃO DISTRIBUIÇÃO DE PROBABILIDADE.

A função densidade de probabilidade de uma v.a.c. X é uma função f que, para todo x no intervalo $-\infty \leq x \leq +\infty$, satisfaz

$$f(x) \geq 0 \quad e \quad \int_a^b f(x)\,dx \tag{29}$$

e é tal que a probabilidade de x assumir um valor no intervalo a ≤ x ≤ b é igual a:

$$P(a \leq x \leq b) = \int_a^b f(x)\,dx \tag{30}$$

A função de distribuição um certo número x, é definida como de probabilidade de X, que dá a probabilidade de a variável assumir um valor menor ou igual a:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx \tag{31}$$

É claro que:

$$P(a \leq X \leq b) = F(b) - F(a) \quad \text{e que} \quad f(x) = \frac{dF(x)}{dx} \tag{32}$$

3.4 VALOR ESPERADO E VARIÂNCIA DE V.A.C.

Se X é v.a.c., então:

$$\begin{aligned} &E(X) = \int_{-\infty}^{\infty} xf(x)dx \\ &e \\ &Var(x) = \int (x-\mu)^2 f(x)dx \end{aligned} \tag{33}$$

3.5 PRINCIPAIS DISTRIBUIÇÕES CONTÍNUAS

Entre as distribuições teóricas de variável aleatória contínua, as principais são:

a) Distribuição Normal

b) Distribuição Qui-quadrado

c) Distribuição T de Student

d) Distribuição F de Snedecor

3.5.1 Distribuição Normal

Se X é v.a.c. com distribuição normal, então:

a) a variável aleatória X pode assumir todo e qualquer valor real.

b) A representação gráfica da distribuição normal é uma curva em forma de sino, simétrica em torno da média, que recebe o nome de curva normal ou de Gauss.

c) a área total limitada pela curva e pelo eixo das abscissas é igual a 1, já que essa área corresponde à probabilidade de a variável aleatória X assumir qualquer valor real.

d) a curva normal é assintótica em relação ao eixo das abscissas, isto é, aproxima-se indefinidamente do eixo das abscissas sem, contudo, alcançá-lo.

e) como a curva é simétrica em torno da média, a probabilidade de ocorrer valor maior que a média é igual à probabilidade de ocorrer valor menor do que a média, isto é, ambas as probabilidades são iguais a 0,5 ou 50%. Cada metade da curva representa 50% de probabilidade.

Para as variáveis contínuas (ou que podem ser assim consideradas), o modelo de probabilidade mais importante é o da distribuição normal, também conhecida como distribuição de Gauss.

Se X é uma variável aleatória com distribuição normal, então:

$$f(x)=\frac{1}{2\sigma}e^{\frac{-1}{2}(\frac{x-\mu}{\sigma})^2} \qquad (34)$$

onde μ e σ são, respectivamente, o valor esperado (a média) e o desvio padrão da variável. Como o cálculo da probabilidade de uma variável normal estar dentro de um certo intervalo é extremamente difícil, construiu-se uma tabela que fornece tais probabilidades para o caso de uma variável normal que tem média 0 e desvio padrão igual 1. Uma variável que tem média 0 e desvio padrão igual a 1 é chamada de variável reduzida ou padronizada. Para uma variável normal qualquer, é sempre possível transformá-la em uma variável reduzida através da fórmula:

$$z = \frac{X - \mu}{\sigma} \qquad (35)$$

Na prática, isso permite o uso da tabela para quaisquer que sejam a média e o desvio padrão da variável sob estudo. Mais adiante, no estudo da Estatística Inferencial serão apresentadas as demais variáveis aleatórias contínuas.

3.6 QUESTÕES PARA ESTUDO

1) Utilize a tabela normal para calcular as seguintes probabilidades:

a) $P(0 < Z < 1)$
b) $P(0 < Z < 0,34)$
c) $P(-1,5 < Z < 0,25)$

2) O faturamento diário em um pequeno restaurante, excluindo sábados, tem distribuição normal com média de $530,00 e desvio padrão de $120,00.

   a) Qual a probabilidade de que o faturamento exceda, em um dado dia, a $700,00?
   b) O restaurante necessita faturar pelo menos $300,00 diariamente para cobrir seus custos. Qual a probabilidade de que, em um dado dia, o estabelecimento não consiga cobrir seus custos?

3) Um investimento de risco tem uma rentabilidade mensal real que se distribui normalmente com média 1% e com desvio padrão 0,4%. Qual a probabilidade de em um determinado mês o investimento apresentar uma rentabilidade real negativa?

4) Os salários dos executivos financeiros empresas são distribuídos normalmente, em torno da média R$ 10.000,00, com desvio padrão de R$ 800,00. Calcule a probabilidade de um executivo financeiro ter o salário situado entre R$ 9.800,00 e R$ 10.400,00.

5) Um vendedor de carros tem a oportunidade de trabalhar para a revenda RA ou para a revenda RB. Ele estima as possibilidades de vendas em cada revenda como segue, sendo A e B, respectivamente, o número de carros por semana que ele poderia vender em RA e RB

A	P(A)	B	P(B)
0	0,4	0	0,2
1	0,3	1	0,6
2	0,2	2	0,2
3	0,1	3	0

a) Calcule E(A) e VAR(A)
b) Calcule E(B) e Var(B)
c) Com que firma, no longo prazo, o vendedor espera vender mais carros?

6) A revenda RA oferece ao vendedor uma comissão de R$ 100,00 por carro vendido. Qual é o valor esperado dos ganhos semanais do vendedor, de acordo com esta proposta?

7) A revenda RB oferece uma comissão de R$ 200,00 por carro vendido. Relacione os ganhos esperados semanais e as variâncias destes ganhos segundo as propostas de RA e RB.

8) A revenda RA apresenta a seguinte contra proposta: o vendedor ganharia um fixo de R$ 100,00 por semana, além da comissão de R$ 100,00 por carro vendido. Quais, agora, o valor esperado e a variância dos ganhos semanais do

vendedor? Como são agora comparáveis as propostas de RA e de RB.

9) Uma pesquisa mostra que 80 % das famílias de uma certa comunidade possuem no mínimo dois aparelhos de televisão. Supondo verdadeira a afirmação, e tomando uma amostra aleatória de 10 famílias desta comunidade, qual a probabilidade de exatamente 4 famílias da amostra terem no mínimo dois aparelhos de televisão?

CAPÍTULO 4. VARIÁVEL ALEATÓRIA BIDIMENSIONAL E COEFICIENTE DE CORRELAÇÃO LINEAR DE PEARSON

4.1 ANÁLISE BIDIMENSIONAL

Na maioria das vezes, ao descrever os resultados de um experimento, atribui-se a um mesmo ponto da amostra os valores de duas ou mais variáveis aleatórias. Neste caso, o objetivo é o estudo de um par de variáveis aleatórias, indicando que os conceitos apresentados estendem-se facilmente ao conjunto formado por um número finito de variáveis aleatórias. O desenvolvimento deste tema será feito para variáveis aleatórias discretas.

4.1.1 Tabela de Contingência

Uma tabela de contingência apresenta as freqüências observadas da ocorrência de um evento,segundo dois diferentes critérios de classificação. Cada critério é a expressão de uma variável qualitativa, de modo que a tabela pode ser vista como a distribuição de freqüência da ocorrência conjunta de duas variáveis X e Y. A estrutura geral de uma tabela de contingência é a seguinte:

classes	Y(1)	Y(2)	....	Y(j)	....	Y(c)	totais
X(1)	f11	f12	....	f1j	....	f1c	R1
X(2)	f21	f22	....	f2j	....	f2c	R2
......	...	...	....	....	....	....	....
X(i)	fi1	fi2	....	fij	....	fic	Rj
...	...	...	....	....	....	....	....
X(r)	fr1	fr2	....	frj	....	frc	Rr
totais			....	Cj	....	Cc	N

Figura 9 Tabela de Contingência

onde fij é a freqüência observada correspondente à casela da linha "i" e da coluna "j". A ordem de uma tabela de contingência é o número total de classificações conjuntas possíveis, ou seja, é o produto do número de linha pelo número de colunas (isto é, a ordem é igual a r x c). Por exemplo, vamos supor que observássemos os defeitos que ocorrem em uma linha de produção e que os classificássemos segundo o turno em que aconteceram (1 e 2) e segundo a gravidade dos mesmos (A, B e C). Teríamos, no caso, uma tabela r = 2 e c = 3, cuja ordem portanto seria 2 x 3 (lê-se "dois por três")

4.1.2 Distribuição de Probabilidade Conjunta

Um exemplo de distribuição de probabilidade conjunta é dado pela tabela a seguir:

X	Y			
	1	2	3	*P(Y=y)*
1	0,1	0,1	0	0,2
2	0,1	0,2	0,2	0,5
3	0,1	0,1	0,1	0,3
P(X=x)	0,3	0,4	0,3	1

Figura 10 Distribuição de probabilidade conjunta

4.1.3 Distribuições Marginais e Condicionadas

A partir da Figura 10 pode-se obter facilmente as distribuições de X e Y A primeira e a última coluna da tabela indicam a distribuição de Y, enquanto que a primeira e a última linha da tabela indicam a distribuição de X. Estas distribuições são as chamadas distribuições marginais. Quando se estudam os aspectos descritivos das distribuições com mais de uma variável, observa-se que, às vezes, é conveniente calcular proporções em relação a uma linha ou coluna, e não em relação ao total. Isto é equivalente a ao conceito de distribuição condicional, ou distribuição condicionada. Por exemplo, a partir da Figura 10 pode-se obter a distribuição de x, sabendo-se que y =2.

4.1.4 Covariância e Coeficiente de Correlação Linear de Pearson

Sejam X e Y são duas variáveis aleatórias. A covariância de X e Y é uma medida de associação linear existente entre X e Y. É definida por:

na amostra:

$$Cov(X,Y)=\sum_{i=1}^{n}\frac{(X_i-\overline{X})(Y_i-\overline{Y})}{n} \tag{36}$$

$$Cov(X,Y)=E\left[(X-E(X))(Y-E(Y))\right] \tag{37}$$

Demonstração:

$$E(XY-X\overline{Y}-\overline{X}Y+\overline{X}\overline{Y})=E(XY)-\overline{Y}\,E(X)-\overline{X}\,E(Y)+\overline{X}\,\overline{Y}=$$

$$E(XY)-E(Y)\,E(X)-E(X)\,E(Y)+E(X)\,E(Y)=E(XY)-E(X)\,E(Y)$$

$$\log o: \tag{38}$$

$$Cov(XY)=E(XY)-E(X)\,E(Y)$$

O valor da Cov (XY) depende da unidade de medida de X e Y. para se obter uma medida adimensional do relacionamento entre duas variáveis aleatórias, X e Y, divide-se a covariância pelo produto dos respectivos desvios padrão. Obtém-se assim, o chamado coeficiente de correlação de Pearson. Este coeficiente meda a força de dependência linear conjunta entre X e Y. É dado por:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \tag{39}$$

O coeficiente de correlação varia apenas entre -1 e +1. Seu sinal indica se a relação entre X e Y é direta ou inversa. Se a correlação for positiva, então, á medida em que X cresce, Y, em média, também cresce; e vice-versa. Se a correlação for negativa, então Y, cresce à medida que X decresce. E vice-versa. Um coeficiente de correlação igual a zero indica que X e Y são linearmente independentes, isto é, que X e Y não estão linearmente correlacionadas correlacionados. O coeficiente de correlação não se altera se for acrescida uma constante às variáveis, ou se as variáveis forem multiplicadas por constantes de mesmo sinal.

CAPÍTULO 5. AMOSTRAGEM E ESTIMAÇÃO

Há, pelo menos, quatro razões para utilizarmos levantamentos por amostragem em lugar de censos:

a) custo total menor;
b) tempo;
c) destruição do item investigado (testes destrutivos);
d) populações infinitas;

Mas também existem aquelas situações nas quais não se devem utilizar amostras; por exemplo:

a) população relativamente pequena;

b) facilidade de realizar o censo;

c) necessidade de grande precisão nos resultados;

Há dois tipos de levantamentos por amostragem, a saber, o de amostragem probabilística e o de amostragem não probabilística.

5.1 AMOSTRAGEM PROBABILÍSTICA

5.1.1 Amostra aleatória Simples.

A probabilidade de obter uma amostra de tamanho n é a mesma para qualquer combinação de n objetos da mesma população.

5.1.2 Amostragem Estratificada

Divide-se a população em grupos (ou classes, ou estratos), de modo que os elementos pertencentes ao mesmo estrato sejam o mais homogêneo possível, com respeito à característica em estudo. Para cada grupo toma-se uma subamostra pelo procedimento e amostragem aleatória simples (a.a.s)., e a amostra global é o resultado da combinação das subamostras de todos os estratos.

5.1.3 Amostragem por Conglomerados

Seleciona-se primeiro, ao acaso, grupos (conglomerados) de elementos individuais da população e, a seguir, toma-se ou todos os elementos ou uma subamostra de cada conglomerado. Nos conglomerados, as diferenças entre eles devem ser tão pequenas quanto possível, enquanto as diferenças dentro deles devem ser tão grandes quanto possível.

5.2 AMOSTRAGEM NÃO PROBABILÍSTICA

5.2.1 Amostragem Acidental

Os elementos chegam acidentalmente ao pesquisador; não há sorteio de espécie alguma.

5.2.2 Amostragem por Quotas.

Divide-se a população em subgrupos e determina-se a quota (proporcional) a cada subgrupo. A seleção dos objetos individuais não é por sorteio.

5.2.3 Amostragem por Julgamento

Os elementos que compõem a amostra são intencionalmente escolhidos pelo pesquisador.

5.3 PROCESSO DE RETIRADA DE AMOSTRAS

O processo de retirada de uma amostra da população pode seguir o esquema de:

a) com reposição
b) sem reposição.

É importante saber que se a população for infinita então as retiradas com e sem reposição serão equivalentes. Ou seja, quando a população for infinita, ou tiver um grande número de elementos, o fato de se recolocar o elemento retirado de volta na população, antes de retirar o próximo elemento, em nada afetará a probabilidade de ocorrência do próximo elemento.

No entanto, se a população for finita, ou tiver um pequeno número de elementos, será necessário deve-se levar em conta os dois tipos de procedimentos. Isso porque, na retirada com reposição as diversas retiradas serão independentes. Contudo, no processo sem reposição haverá dependência entre as retiradas. Ou seja, a probabilidade de retirada do próximo elemento é alterada pelo fato de um primeiro já ter sido retirado. Convém ressaltar que a amostragem sem reposição é mais eficiente que a amostragem com reposição, pois apresenta menor variabilidade, uma vez que não é possível retirar elementos extremos mais de uma vez.

5.4 NÚMERO DE AMOSTRAS

Seja N o tamanho da população e n o tamanho da amostra. O número “m” de amostras possíveis de serem retiradas de uma população, de acordo com os critérios com e sem reposição será:

com reposição:

$$m = N^n \qquad (40)$$

sem reposição:

$$m = C_N^{\,n} = \frac{N!}{n!(N-n)!} \qquad (41)$$

CAPÍTULO 6. DISTRIBUIÇÃO AMOSTRAL E ESTIMADORES

Estimador é qualquer função dos elementos da amostra. Assim, n a população tem-se o parâmetro θ e na amostra o estimador $\hat{\theta}$ é uma função dos elementos da amostra de tamanho n.

6.1.1 Propriedades dos Estimadores

a) Não tendencioso (viés bias)

$$E(\hat{\theta}) = \theta \begin{cases} justo \\ centrado \end{cases}$$

Figura 11 Estimador Não-tendencioso e tendencioso

b) Eficiente: considerando dos estimadores, $\hat{\theta}_1, \hat{\theta}_2$, se $\operatorname{var}\hat{\theta}_1 \geq \operatorname{var}\hat{\theta}_2$ diz-se que $\hat{\theta}_2$ é mais eficiente do que $\hat{\theta}_1$.

c)Consistência

Tanto a propriedade da consistência como da eficiência se analisa em amostras de dado tamanho, ou seja, para um determinado tamanho de amostra. Por isso, tendensiosidade e eficiência são propriedades de pequenas amostras. Mas, consistência ou divergência é definida para grandes amostras. Tem-se uma seqüência de $\hat{\theta}$. Um $\hat{\theta}$ para cada amostra, cada vez um pouco maior. É essa seqüência $\hat{\theta}_n$ que será consistente ou convergente, a medida que o tamanho da amostra tende ao tamanho da população. Talvez, a maneira mais fácil de entender consistência, seja quando n $\rightarrow$ N, E($\hat{\theta}_n$)$\rightarrow \theta$; o viés tende a zero.

População $\longrightarrow$ Amostra

Distribuição de Probabilidade (ou FDP) — Distribuição Amostral

Parâmetros (valor fixo) $\xleftarrow{\text{estimar}}$ Estatísticas (variável aleatória)

Estimação $\begin{cases} \text{pontual (estatísticas)} \\ \text{por intervalo (intervalos de confiança)} \end{cases}$

OBS: estatística: é a v.a. que estima (pontualmente) um parâmetro (populacional)
as vezes é chamada simplesmente de estimador
estimativa: é o valor do estimador obtido para uma amostra específica

Figura 12 Quadro resumo: estimação

6.1.2 Questões para estudos

1) A seguir apresentam-se três diferentes estimadores da média populacional para amostras de tamanho dois. Verifique se estes estimadores são não tendenciosos.

a) $\bar{x} = \frac{1}{2}x_1 + \frac{1}{2}x_2$

b) $\overset{\leftrightarrow}{x} = \frac{1}{2}x_1 + \frac{2}{3}x_2$

c) $\check{x} = \frac{1}{3}x_1 + \frac{1}{3}x_2$

Solução:

a) $E(\bar{x}) = \frac{1}{2}E(x_1) + \frac{1}{2}E(x_2) \Rightarrow \frac{1}{2}\mu + \frac{1}{2}\mu = \mu$

$E(\bar{x}) = \mu$

$$E(\overset{\leftrightarrow}{x}) = \frac{1}{2}x_1 + \frac{2}{3}x_2 \Rightarrow \frac{1}{2}E(x_1) + \frac{2}{3}E(x_2)$$

$E(\overset{\leftrightarrow}{x}) = \mu$, portanto é estimador não tendencioso da média da população.

b) $E(\overset{\mathrm{V}}{x}) = \frac{1}{3}E(x_1) + \frac{1}{3}E(x_2) = \frac{1}{3}\mu + \frac{1}{3}\mu = 2/3\,\mu$, logo é estimador tendencioso da média da populaçãoé tendencioso.

Entre dos estimadores não tendenciosos usa-se o critério da eficiência.

$$Var(\bar{x}) = \frac{1}{4}\sigma_x^2 + \frac{1}{4}\sigma_x^2 = \frac{2}{4}\sigma_x^2 = \frac{1}{2}\sigma_x^2 \qquad Var(\bar{x}) = \frac{\sigma_x^2}{2}$$

$$Var(\overset{\leftrightarrow}{x}) = \frac{1}{9}\sigma_x^2 + \frac{4}{9}\sigma_x^2 = \frac{5}{9}\sigma_x^2 \qquad Var(\overset{\leftrightarrow}{x}) = \frac{5}{9}\sigma_x^2$$

$$Var(\overset{\mathrm{V}}{x}) = \frac{1}{9}\sigma_x^2 + \frac{1}{9}\sigma_x^2 = \frac{2}{9}\sigma_x^2 \qquad Var(\overset{\mathrm{V}}{x}) = \frac{2}{9}\sigma_x^2$$

$Var(\overset{\mathrm{V}}{x}) = \frac{2}{9}\sigma_{\bar{x}}^2$ é tendencioso mas de menor variância.

$\sigma_{\bar{x}}^2 > \sigma_{\bar{x}}^2$ donde se segue que $\bar{x}$ é mais eficiente.

$$\frac{1/2\sigma_x^2}{5/9\sigma_x^2} = 1/2\sigma_x^2 \cdot 9/5 \cdot \sigma_x^2 = 9/10\sigma^2 = 0{,}9\sigma^2$$

2) Sejam x_1, x_2, e x_3 variáveis aleatórias independentes, identicamente distribuídas com média μ e variância σ^2. Seja β = 3/4 x_1 + 1/8 x_2+ 1/8 x_3. Calcule E(β) e σ_β^2 (β). β é um estimador não-tendencioso de μ?

3) Sendo $\hat{\mu} = x_1 + 1/8\ x_2 + 1/2\ x_3$ um estimador de μ, verifique se $\hat{\mu}$ é estimador não tendencioso μ e calcule a $\sigma^2_{\hat{\mu}}$.

4) Sabendo-se que $\hat{y}$ é um estimador não tendencioso do parâmetro populacional y, assinale a alternativa correta.

a) $E(\hat{y}) = y$

b) $\hat{y} = y$

   a) a variância de $\hat{y}$ é igual a variância de y

   b) $\hat{y} = y$, apenas quando a amostra tiver mais do que 30 observações

   c) $\hat{y}$ terá sempre distribuição normal

5) Sabe-se que x_1, x_2 e x_3 representam, respectivamente, o primeiro, segundo e terceiro elemento de uma amostra de tamanho três, retirada de uma população com média μ e variância σ^2. Sabe-se, também, que A e B são estimadores da média populacional, onde $A = 3/4\ x_1 + 1/8\ x_2 + 1/8\ x_3$ e $B = 3/4\ x_1 + 1/8\ x_2$ Com estas informações assinale a alternativa correta.

   a) A é estimador não tendencioso e B é estimador tendencioso

   b) A é estimador não tendencioso e variância de A é igual a 0,5 σ^2

   c) B é estimador não tendencioso e variância de B é igual a 0,5 σ^2

   d) B é estimador tendencioso e variância de B é igual à variância de A

   e) A e B são estimadores tendenciosos e possuem a mesma variância.

6.2 DISTRIBUIÇÃO AMOSTRAL DO ESTIMADOR MÉDIA

6.2.1 Amostragem com reposição

Considere a população P = {1, 2, 3, 4} e todas as possíveis amostras de tamanho, extraídas com reposição e apresentadas na tabela a seguir:

	1	**2**	**3**	**4**
1	1,1,	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3,	3,4
4	4,1	4,2	4,3	4,4

As médias amostrais encontram-se na tabela a seguir:

	1	**2**	**3**	**4**
1	1	1,5	2	2,5
2	1,5	2	2,5	3
3	2	2,5	3	3,5
4	2,5	3	3,5	4

A distribuição amostral da média é dada por:

$\bar{X}$	$P(\bar{X})$
1	1/16
1,5	2/16
2	3/16
2,5	4/16
3	3/16
3,5	2/16
4	1/16

6.2.2 Cálculo da Expectância

E(x) = 1*1/16 +1,5* 2/16+2* 3/16 + 2,5*4/16+3* 3/16 + 3,5 * 2/16 + 4 * 1/16 = 2,5

6.2.3 Cálculo da variância

$\sigma^2_{\bar{X}} = E(\bar{X}^2) - E(\bar{X})^2$, logo:

$$\sigma^2_{\bar{X}} = \left[1^2 * 1/16 + 1,5^2 * 2/16 + 2,5^2 * 3/16 + 2,5^2 * 4/16 + 3^2 * 3/16 + 3,5^2 * 2/16 + 4^2 * 1/16\right] - 2,5^2$$

$$\sigma^2_{\bar{X}} = 0,625$$

Observa-se que a E($\bar{X}$) é igual a μ (média da população)

<u>Demonstração de</u> $E(\bar{X}) = \mu$

$$E(\bar{X}) = \left(\frac{\sum_{1}^{n} xi}{n}\right) = \frac{1}{n} E\left(\sum_{1}^{n} xi\right) = \frac{1}{n}\left[E(x_1 + x_2 + x_3 + \ldots\ldots + x_n)\right]$$

$$E(\bar{X}) = \frac{E(x_1) + E(x_2) + E(x_3) + \ldots\ldots + E(x_n)}{n}$$

$$E(\bar{X}) = \frac{\mu + \mu + \mu + \mu + \ldots.}{n} = \frac{n\mu}{n} = \mu$$

$$E(\bar{X}) = \mu$$

6.2.4 Cálculo da Variância de $\bar{x}$

$\sigma^2_{\bar{X}} = E(\bar{X}^2) - E(\bar{X})^2$, logo:

$$\sigma^2_{\bar{X}} = \left[1^2 * 1/16 + 1,5^2 * 2/16 + 2,5^2 * 3/16 + 2,5^2 * 4/16 + 3^2 * 3/16 + 3,5^2 * 2/16 + 4^2 * 1/16\right] - 2,5^2$$

$$\sigma^2_{\bar{X}} = 0,625$$

Observa-se que a variância da distribuição amostral da média é igual à variância da população dividida pelo tamanho da amostra. Ou seja, $\sigma^2_{\bar{X}}$ é igual a $\frac{\sigma^2}{n}$, onde σ^2 é a variância da população e n o tamanho da amostra.

Demonstração de $\sigma_{\bar{X}}^2 = \dfrac{\sigma^2}{n}$

Sejam x_1, x_2, x_3, ... x_n variáveis aleatórias independentes (amostragem com reposição).

$$Var(\bar{X}) = Var\left(\frac{\sum_{1}^{n} x1}{n}\right) = \frac{1}{n^2}\left[Var\left(x_1 + x_2 + x_3 + ... + x_n\right)\right]$$

$$Var(\bar{X}) = \frac{1}{n^2}\left[Var(x_1) + Var(x_2) + Var(x_3) + ... + Var(x_n)\right]$$

$$Var(\bar{X}) = \frac{1}{n^2}\left[\sigma^2 + \sigma^2 + \sigma^2 + ... + \sigma^2\right]$$

$$Var(\bar{X}) = \frac{1}{n^2} * n\sigma^2 = \frac{\sigma^2}{n}$$

6.2.5 Amostragem sem reposição

Considere a mesma população anterior, P= (1,2,3,4) e todas as amostras possíveis de tamanho (n=2), retiradas, sem reposição,e apresentadas na tabela a seguir:

	1	**2**	**3**	**4**
1	1	1,5	2	2,5
2	1,5	2	2,5	3
3	2	2,5	3	3,5
4	2,5	3	3,5	4

As médias amostrais encontram-se na tabela a seguir:

	1	**2**	**3**	**4**
1		1,2	1,3	1,4
2			2,3	2,4
3				3,4
4				

A distribuição amostral da média é dada por:

$\bar{X}$	$P(\bar{X})$
1,5	1/6
2	1/6
2,5	2/6
3	1/6
3,5	1/6

Calculando-se a expectância da média amostral tem-se:

$$E(\bar{X}) = 1,5 * 1/6 + 2*1/6 + 2,5 *2/6 + 3 *1/6 + 3,5 * 1,6 = 2,5$$

Observa-se que a expectância de todas as médias amostrais obtidas com as possíveis amostras de tamanho 2, extraídas sem reposição, também é igual a média da população, ou seja:

$$E(\bar{X}) = \mu$$

Calculando-se a variância de $\bar{X}$ tem-se que:

$$\sigma_{\bar{X}}^2 = E(\bar{X}^2) - E(\bar{X})^2$$

$$E(\bar{X}^2) = 1,5^2 *1/6 + 2^2 *1/6 + 2,5^2 *2/6 + 3^2 *1/6 + 3,5^2 *1/6 \cong 6,66667$$

$$\sigma_{\bar{X}}^2 = 6,66667 - 2,5^2 = 0,41667$$

Assim, utiliza-se o fator de correção para população finita, dado por $\frac{N-n}{N-1}$. Observa-se que quanto maior for a diferença entre o tamanho da população e o tamanho da amostra mais próximo de um estará este fator. Então como regra prática, utiliza-se o fator de correção quando o tamanho da amostra for superior a 5% da população. Caso contrário não será necessário fazer a distinção entre amostragem com e sem reposição.

Estas considerações apresentadas valem para populações pequenas. Se a população for bastante grande ou infinita, não será possível pensar em construir tabelas para representar a distribuição amostral. Assim, será necessário procurar por modelos probabilísticos que descrevam as distribuições amostrais.

A partir de dois teoremas demonstra-se que a distribuição amostral da média, $\overline{X}$ é uma distribuição normal.

Teorema da combinação linear: a combinação linear de variáveis com distribuição normal é uma normal. Assim, se a população tiver distribuição normal a distribuição amostral da média é normal.

Teorema central do limite: se (x_1, x_2, x_3,....x_n) for uma amostra aleatória extraída de uma população com qualquer distribuição, então aumentando-se o tamanho da amostra a distribuição amostral da média rapidamente converge para uma distribuição normal. Para amostras de tamanho 30 a convergência será suficiente boa, podendo-se utilizar este resultado. Assim, se X tem qualquer distribuição então $\overline{X}$ terá uma distribuição aproximadamente $N(\mu, \frac{\sigma^2}{n})$.

6.3 QUESTÕES PARA ESTUDO

1) Uma população se constitui dos números 2, 6, 8 e 12. Considere todas as amostras possíveis de tamanho n= 2, que podem ser retiradas, com reposição, dessa população. Determine:

   a) a distribuição amostral da média
   b) o desvio padrão da distribuição amostral da média
   c) a probabilidade de se retirar uma amostra com média superior a 8
   d) a probabilidade de se retirar uma amostra com média inferior a 8.

2) Uma variável X tem distribuição normal, com média 20 e desvio padrão 10. Qual a probabilidade de esta população gerar uma amostra, de tamanho 25, cuja média amostral seja maior ou igual a 19?

3) Considere uma população bastante grande cuja média é igual a 50 e desvio padrão igual a 12. Determine a distribuição de amostragem da média para amostras de tamanho n = 36, em termos de valor esperado e desvio padrão.

4) Um auditor retira uma amostra aleatória de tamanho n = 16 de um conjunto de N = 100 contas a receber. O desvio padrão das contas a receber é σ = R$ 57,00. Qual a distribuição amostral da média?

5) Uma população se constitui dos números (1, 3, 5, 6). Considere todas as amostras possíveis de tamanho n=2, que podem ser retiradas dessa população. Considerando que o processo de amostragem é com reposição, determine:

e) a média da população
f) o desvio padrão da população
g) a distribuição amostral da média
h) $E(\bar{X})$ e $\sigma_{\bar{X}}$

6) Com a população do exercício anterior, considerando que o processo de amostragem é sem reposição, determine:

i) a média da população
j) o desvio padrão da população
k) a distribuição amostral da média
l) $E(\bar{X})$ e $\sigma_{\bar{X}}$

7) Um sociólogo extrai uma amostra aleatória de 45 pessoas de uma população cuja renda média é de US$ 900,00 e o desvio padrão US$ 200,00. Qual será a

probabilidade de que a renda média da amostra seja inferior a US$ 850,00?

8) Certas vacinas produzidas por um laboratório têm validade média de 800 horas e desvio padrão de 60 horas. Determine a probabilidade de uma amostra aleatória de 50 vacinas ter a validade média:

   a) entre 790 e 810 horas;
   b) inferior a 785 horas;
   c) superior a 820 horas.

9) Suponhamos que o nível educacional de adultos de certo país tenha uma média de 11,1 anos e um desvio padrão de 3 anos. Qual a probabilidade de que, em uma amostra aleatória de 40 adultos, se encontre um nível médio de escolaridade entre 10 e 12 anos?

10) A média de uma distribuição amostral de médias é 12 e o desvio-padrão é 3. Sabendo que a distribuição amostral das médias é normal calcule:

   a) que percentagem de médias amostrais estará entre 6 e 18?
   b) que percentagem de médias amostrais estará entre 9 e 15?
   c) que percentagem de médias amostrais será menor do que a média populacional?

11) De uma população com 50 elementos, é retirada uma amostra aleatória simples, sem reposição, de n =10.

   a) Qual o número de amostras possíveis?
   b) Qual a probabilidade de cada uma destas amostras ser selecionada?
   c) Uma população é composta dos elementos: x, y, z, e p.
   d) Apresente todas as possíveis amostras aleatórias simples, sem reposição, com n = 2.
   e) Apresente todas as aas, sem reposição, de tamanho n = 3.
   f) Determine a probabilidade de ser sorteada a amostra z e p e x.

g) Determine a probabilidade de ser sorteada a amostra x, y e z.

6.4 DISTRIBUIÇÃO AMOSTRAL DA PROPORÇÃO

Pode-se mostrar que a proporção p com que uma determinada característica ocorre na amostra, $\hat{\pi}$, é um bom estimador, do ponto de vista estatístico, da proporção π com que esta mesma característica ocorre na população amostrada.

Mostra-se ainda que para amostras grandes (isto é, amostras com mais de 30 observações), p tem distribuição aproximadamente normal com:

$$E(\hat{\pi}) = \pi \quad e \quad \sigma_{\hat{\pi}} = \sqrt{\frac{\pi(1-\pi)}{n}} \tag{42}$$

6.5 ESTIMATIVA POR PONTO E POR INTERVALO DE CONFIANÇA

6.5.1 Intervalo de Confiança para média com variância populacional conhecida

De uma população normal com média desconhecida e σ^2=200, retirou-se uma amostra de 50 elementos. Assim, tem-se que:

$X \sim N(\mu,\ \sigma^2=200)$ $\qquad$ n = 50 $\qquad$ $\bar{X} = 80$

$\bar{x} = 80$ $\quad\Rightarrow$ Estimativa por ponto da média de população, ou seja, a estimativa pontual de μ é 80.

A distribuição de amostragem de $\bar{x}$ é:

$\bar{x}$~N(μ, σ²/n) $\qquad \sigma^2_{\bar{x}} = \sigma^2 / n = \dfrac{200}{50} = 4 \qquad \sigma^2_{\bar{x}} = 4$ e $\sigma_{\bar{x}} = 2$

ξ⇒ Define-se como erro absoluto de estimação.

$\xi = \lambda \sigma_{\bar{x}}$

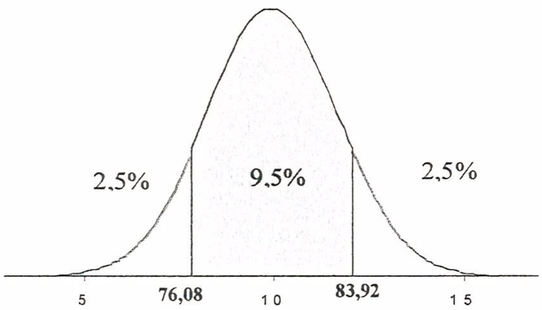
λ é um afastamento na distribuição amostral do estimador (um certo número de desvios padrão)

$\xi = z\sigma_{\bar{x}}$, z = 1,96

Tamanho da amostra é dado por: $n = \left(\dfrac{z_{\alpha/2}\, \sigma_{\bar{X}}}{\varepsilon} \right)^2$

ξ= 1,96*2 = 3,92 é o erro absoluto de estimação; dá uma idéia da precisão da estimativa

Assim: $P(\bar{X} - z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}) = 1 - \alpha$ =P(76,08 ≤ μ ≤ 83,92) = 95% .



A probabilidade de que este intervalo contenha a verdadeira média da população é igual a 95 %.

Se o processo de amostragem for sem reposição e o tamanho da amostra, n, for maior do que 5 % do tamanho da população, N, utiliza-se o fator de correção:

$$\boxed{P\left(\bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}<\mu<\bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right)=1-\alpha}$$

Observe que não é a probabilidade de que a média caia dentro do intervalo, pois quem varia de amostra para amostra é o intervalo. A média da população μ não tem probabilidade associada. Não é variável aleatória. É um parâmetro.

6.5.2 Nível de Confiança

Em 95% das amostras de tamanho 50, o intervalo que se construir vai conter a média populacional. A Figura 13 apresenta um resumo para o cálculo do intervalo de confiança para média da população.

$X \sim N(\mu,\sigma^2)$ μ desconhecido, mas σ^2 conhecido

$$\bar{X} \sim N(\mu,\frac{\sigma^2}{n})$$

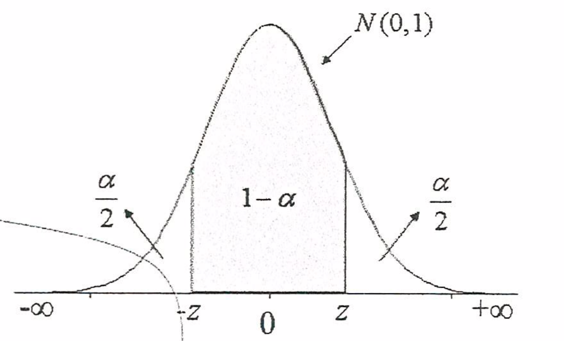
$$Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \to N(0,1) \text{ (Normal Padrão)}$$

$$P(-z<\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}<z)=1-\alpha$$

$$P(-z\frac{\sigma}{\sqrt{n}}<\bar{X}-\mu<z\frac{\sigma}{\sqrt{n}})=1-\alpha$$

$$\boxed{P(\bar{X}-z\frac{\sigma}{\sqrt{n}}<\mu<\bar{X}+z\frac{\sigma}{\sqrt{n}})=1-\alpha}$$

IC para μ



$P(|Z|>z)=\alpha$ nível de significância

$P(-z<Z<z)=1-\alpha$ nível de confiança

Figura 13 Quadro resumo: I. C. variância populacional conhecida

6.5.3 Intervalo de Confiança para a média com variância populacional desconhecida

De uma população normal com média e variância desconhecida foi retirada a seguinte amostra:

3, 7 e 5.

$\overline{x} = \dfrac{3+7+5}{3} = 5$ $\qquad\qquad$ $\overline{x} = 5$ Estimativa por ponto.

$\overline{x} = N(\mu, \sigma^2 / n)$, mas não se conhece σ^2, logo é necessário estimá-la.

Conforme visto anteriormente, tem-se que:

$\boxed{\hat{\sigma}^2 = \dfrac{\sum(x-\overline{x})^2}{n-1}}$ é um estimador não tendencioso da variância populacional.

Assim:

$$\hat{\sigma}^2 = \frac{\sum(x-\overline{x})^2}{n-1} = \frac{(3-5)^2+(7-5)^2+(5-5)^2}{2} = \frac{8}{2} = 4$$

$\hat{\sigma}_x^2 = 4$ estimativa não tendenciosa da variância da população.

$\hat{\sigma}_{\overline{x}}^2 = \dfrac{\hat{\sigma}^2}{n} \Rightarrow \hat{\sigma}_{\overline{x}}^2 = 4/3$ estimativa da variância de $\overline{x}$

$\hat{\sigma}_{\overline{x}} = 1{,}15$ estimativa da variância de $\overline{x}$

Agora abre-se o intervalo, sabendo-se que $\overline{x} \sim N$

$$\frac{\overline{x}-\mu}{\sigma_x} \sim N(0,1) = z$$

Mas, como não se conhece $\sigma_{\overline{x}}$ e sim $\hat{\sigma}_{\overline{x}}$, tem-se que:

$$\boxed{\frac{\overline{x}-\mu}{\hat{\sigma}_{\overline{x}}} \sim t_{n-1}}.$$

Assim:

$$\boxed{P(\bar{X}-t_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}<\mu<\bar{X}+t_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}})=1-\alpha}$$

Se o processo de amostragem for sem reposição e o tamanho da amostra, n, for maior do que 5 % do tamanho da população, N, utiliza-se o fator de correção:

$$\boxed{P(\bar{X}-t_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}<\mu<\bar{X}+z_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}})=1-\alpha}$$

A Figura 14 apresenta o quadro resumo para o intervalo de confiança para a média populacional, quando a variância populacional for desconhecida.

$X \sim N(\mu,\sigma^2)$ $\quad$ μ e σ^2 desconhecidos

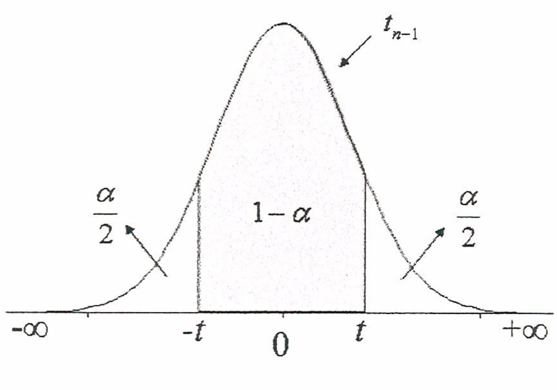
$$T = \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$

$$P(-t<\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}<t)=1-\alpha$$

$$P(-t\frac{s}{\sqrt{n}}<\bar{X}-\mu<t\frac{s}{\sqrt{n}})=1-\alpha$$

$$\boxed{P(\bar{X}-t\frac{s}{\sqrt{n}}<\mu<\bar{X}+t\frac{s}{\sqrt{n}})=1-\alpha}$$

IC para μ



$P(-t<T<t)=1-\alpha$

Figura 14 Quadro resumo: I. C. variância populacional desconhecida

6.5.4 Questões para estudo

1. Uma amostra aleatória de 15 peças produzidas por uma máquina forneceu um comprimento médio de 20 mm, com desvio-padrão estimado de 0,1 mm. Supondo que o comprimento das peças tem distribuição normal de probabilidades, determine um intervalo de confiança de 95% para o comprimento médio das peças produzidas por esta máquina.

2. Uma restaurante do tipo *self-service* cobra refeições por peso. Uma amostra aleatória de 12 refeições selecionadas em um período em que foram servidas 80 refeições apresentou um peso médio de 460 g com desvio-padrão estimado de 80g. Supondo que o peso das refeições se distribua normalmente, determine um intervalo de confiança de 98% para o peso médio das refeições servidas neste período.

3. Uma amostra de 5 elementos selecionados de uma população normal de 200 elementos apresentou os valores: 170, 98, 106, 145 e 92. Calcule a média e o desvio-padrão amostral. Construa um intervalo de 95 % de confiança para a média da população.

4. Uma amostra de 60 elementos obtidos de uma população muito grande apresentou média de 46 e desvio-padrão estimado de 8. Calcular: a) o erro-padrão de estimação da média populacional, ao nível de 90%; b) o intervalo de confiança para a média populacional ao nível de 98%.

5. Uma amostra de 60 elementos obtidos de uma população muito grande apresentou média de 46 e desvio-padrão estimado de 8. Calcular: a) o erro-padrão de estimação da média populacional, ao nível de 90%; b) o intervalo

de confiança para a média populacional ao nível de 98%.

6. Uma amostra de 60 elementos obtidos de uma população muito grande apresentou média de 46 e desvio-padrão estimado de 8. Calcular: a) o erro-padrão de estimação da média populacional, ao nível de 90%; b) o intervalo de confiança para a média populacional ao nível de 98%.

6.6 INTERVALO DE CONFIANÇA PARA A PROPORÇÃO

Determinado atributo na população tem determinada proporção π.

$$\boxed{\hat{\pi} \sim N(\pi; \frac{\pi(1-\pi)}{n})} \qquad \sigma_{\pi} = \sqrt{\frac{\pi(1-\pi)}{n}}$$, independente da normalidade.

$$\boxed{P(\hat{\pi} - z_{\alpha/2}\sigma_{\hat{\pi}} < \pi < \hat{\pi} + z_{\alpha/2}\sigma_{\hat{\pi}}) = 1-\alpha}$$

Para pequenas amostras, n < 30, Distribuição Binomial.

Para grandes amostras n >3 0, Distribuição Normal.

Exemplo:

Suponha uma amostra de 200 pessoas, n =200, e dessas 200 pessoas 30 são fumantes.

$\hat{\pi}$ = 30/200 = 0,15

$\hat{\pi} = 0,15$ **é** uma estimativa não tendenciosa de π.

O desvio padrão de $\hat{\pi}$ é dado por: $\sigma_{\hat{\pi}} = \sqrt{\dfrac{\pi(1-\pi)}{n}}$

Contudo, não se conhece π, logo usa-se $\hat{\pi}$.

$$\hat{\sigma}_{\hat{\pi}} = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} = \sqrt{\frac{0,15*0,85}{200}} = 0,0252$$

$\varepsilon = z_{\alpha/2}\sigma_{\hat{\pi}} = 1,96*0,0252 = 0,05$ erro absoluto de estimação de 5%. Assim, tem-se que:

$$P(0,10 \leq \pi \leq 0,20) = 95\%$$

Probabilidade de 95% de que o intervalo contenha a verdadeira proporção de fumantes, π.

6.7 TAMANHO DA AMOSTRA

$$\varepsilon = z_{\frac{\alpha}{2}}\, \sigma_{\hat{\pi}}$$

$$\varepsilon = z_{\frac{\alpha}{2}} \sqrt{\frac{\pi(1-\pi)}{n}}$$

Agora n é uma incógnita. Então é preciso saber a margem de erro que se suporta. Suponha que a margem de erro seja 2 %.

$\xi = 2\%$ e $1 - \alpha = 0,95$

$0,02 = 1,96\sqrt{\dfrac{\pi(1-\pi)}{n}}$, não se conhece π, é o que se quer estimar.

π	$(1 - \pi)$	$\pi (1 - \pi)$
0,1	0,9	0,09
0,2	0,8	0,16
0,3	0,7	0,21
0,4	0,6	0,24
0,5	0,5	0,25

Observa-se que $\pi (1 - \pi)$ tem um valor máximo, pois $0 \leq \pi \leq 1$.

No lugar de fazer uma amostra piloto, pode-se supor variância máxima, ou seja:

$\pi (1 - \pi) = 0,25$

$$0,2 = 1,96\sqrt{\frac{0,25}{n}} = \left(\frac{0,02}{1,96}\right)^2 = \frac{0,25}{n} \Rightarrow n = 2400$$

Tamanho mínimo da amostra que garante erro máximo de estimação de 2%, pois está se supondo variância máxima.

$$n = \frac{z^2 * 0,25}{\xi^2}$$

Se quisesse 3%

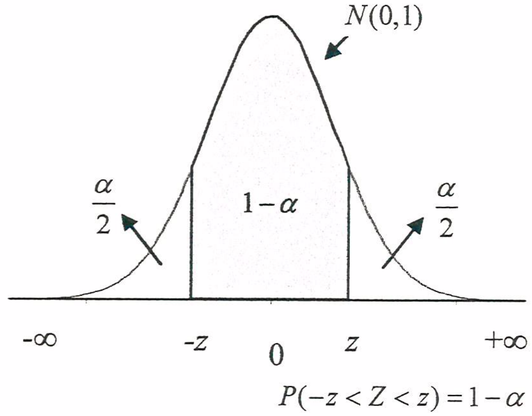
$n = \dfrac{1,96^2 * 0,25}{0,03^2} = 1067$ é o tamanho mínimo que garante erro máximo de 3 %.

A seguir, apresenta-se um quadro resumo do intervalo de confiança para a proporção populacional.

$$\hat{p} \sim N(p, \frac{pq}{n})$$

$$\frac{\hat{\pi} - \pi}{\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}} \sim N(0,1)$$

$$P(-z < \frac{\hat{\pi} - \pi}{\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}} < z) = 1 - \alpha$$



$$P(\hat{\pi} - z\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} < \pi < \hat{\pi} + z\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}) = 1 - \alpha$$

$$P(\hat{\pi} - z\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} < \pi < \hat{\pi} + z\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}) = 1 - \alpha$$

Figura 15 Quadro resumo: I. C. a proporção

Se o processo de amostragem for sem reposição e o tamanho da amostra, n, for maior do que 5 % do tamanho da população, N, utiliza-se o fator de correção

$$\boxed{P(\hat{\pi} - z_{\alpha/2}\sigma_{\hat{\pi}}\sqrt{\frac{N-n}{N-1}} < \pi < \hat{\pi} + z_{\alpha/2}\sigma_{\hat{\pi}}\sqrt{\frac{N-n}{N-1}}) = 1 - \alpha}$$

6.8 EXERCÍCIOS RESOLVIDOS

1. Em certa cidade deseja-se estimar a proporção de pessoas que são favoráveis a construção de um viaduto no centro da cidade. Supondo que numa amostra de 100 pessoas desta cidade, 75 não são favoráveis à construção de tal viaduto, qual será o intervalo com 99% de confiança para a proporção de

pessoas não favoráveis a construção do viaduto? Qual o tamanho da amostra para se ter uma confiança de 90, mantendo-se o mesmo erro de estimação?

2. Uma amostra de 500 eleitores selecionados ao acaso mostra que 52% são favoráveis ao Partido Democrático. Poderia esta amostra ter sido retirada de uma população que tivesse 50% de eleitores democratas? Admita $1-\alpha = 0,05$.

3. Em uma cidade coletou-se uma amostra de 500 famílias que possuem aparelho de televisão. Nesta amostra, 340 famílias que possuem televisor de tela plana. Estime o número de famílias que possuem televisor tela plana, qual o tamanho da amostra necessário para que se tenha 95% de confiança em que o erro de nossa estimativa não seja superior a 0,02? Para que essa confiança seja, no mínimo, 95%?

4. Antes de uma eleição, deseja-se fazer uma pesquisa para verificar a proporção de eleitores que pretende votar no candidato A. Para isto, consultou-se uma amostra de 1.600 eleitores, onde se obteve 35% dos eleitores favoráveis ao candidato A. Neste caso, qual será o intervalo com 95% de confiança para a proporção de eleitores que são favoráveis ao candidato A?

5. A produção anual de determinado produto é de 1000 unidades. Desta produção, retirou-se uma amostra de 100 unidades, observando-se uma proporção amostral de 30% de embalagens defeituosas. Estime um intervalo de 95 % confiança para a proporção de embalagens defeituosas.

1)

$n = 100$

$\hat{\sigma} = 75\,\%\, não\ favoraveis$

$\hat{\sigma} = 25\,\%\ favoraveis$

$1 - \alpha = 95\%$

$P(\hat{\pi} - z_{\alpha/2}\,\hat{\sigma}_{\pi} < \pi < \hat{\pi} - z_{\alpha/2}\hat{\sigma}_{\pi}) = 1 - \alpha$

$$P(0,75 - 1,96\sqrt{\frac{0,75 * 0,25}{100}} < \pi < 0,75 + 1,96\sqrt{\frac{0,75 * 0,25}{100}}) = 95\%$$

$P(0,75 - 0,085 < \pi < 0,75 + 0,085) = 95$

$P(0,67 < \pi < 0,83) = 95$

2)

$n = 500$

$\hat{\sigma} = 52\,\%\ favoraveis$

$\sigma = 50\,\%$

$1 - \alpha = 95\%$

$P(\hat{\pi} - z_{\alpha/2}\,\hat{\sigma}_{\pi} < \pi < \hat{\pi} - z_{\alpha/2}\hat{\sigma}_{\pi}) = 1 - \alpha$

$$P(0,5 - 1,96\sqrt{\frac{0,5 * 0,5}{500}} < \pi < 0,5 + 1,96\sqrt{\frac{0,5 * 0,5}{500}}) = 95\%$$

$P(0,5 - 0,1 < \pi < 0,5 + 0,1) = 95\%$

$P(0,4 < \pi < 0,6) = 95\%$

3)

$n = 500$

$\widehat{\sigma} = 0,68$

$1 - \alpha = 95\%$

$$P(\hat{\pi} - z_{\alpha/2}\,\widehat{\sigma}_{\pi} < \pi < \hat{\pi} - z_{\alpha/2}\widehat{\sigma}_{\pi}) = 1 - \alpha$$

$$P(0,68 - 1,96\sqrt{\frac{0,68*0,32}{500}} < \pi < 0,5 + 1,96\sqrt{\frac{0,68*0,32}{500}}) = 95\%$$

$$P(0,68 - 0,02 < \pi < 0,68 + 0,02) = 95\%$$

$$P(0,66 < \pi < 0,7) = 95\%$$

$$\varepsilon = \frac{z^2{}_{\alpha/2}\,\widehat{\sigma}_{\pi}{}^2}{n^2}$$

$$n = \frac{z^2{}_{\alpha/2}\,\widehat{\sigma}_{\pi}}{\varepsilon^2}$$

$\log o:$

$$n = \frac{1,96^2\, 0,68*0,32}{0,02^2} = 2090 \ \ observações$$

4)

$n = 1600$

$\hat{\sigma} = 35\%$

$1 - \alpha = 95\%$

$$P(\hat{\pi} - z_{\alpha/2}\,\hat{\sigma}_{\pi} < \pi < \hat{\pi} - z_{\alpha/2}\hat{\sigma}_{\pi}) = 1 - \alpha$$

$$P(0,35 - 1,96\sqrt{\frac{0,35*0,65}{1600}} < \pi < 0,35 + 1,96\sqrt{\frac{0,35*0,65}{1600}}) = 95\%$$

$$P(0,35 - 0,02 < \pi < 0,35 + 0,02) = 95\%$$

$$P(0,33 < \pi < 0,37) = 95\%$$

5)

$N = 1000$

$n = 100$

$\hat{\sigma} = 30\%$

$1 - \alpha = 95\%$

$$P(\hat{\pi} - z_{\alpha/2}\,\hat{\sigma}_{\pi}\sqrt{\frac{N-n}{N-1}} < \pi < \hat{\pi} + z_{\alpha/2}\hat{\sigma}_{\pi}\sqrt{\frac{N-n}{N-1}}) = 1 - \alpha$$

$$P(0,30 - 1,96\sqrt{\frac{0,30*0,7}{100}}\sqrt{\frac{900}{999}} < \pi < 0,30 + 1,96\sqrt{\frac{0,30*0,7}{100}}\sqrt{\frac{900}{999}}) = 95\%$$

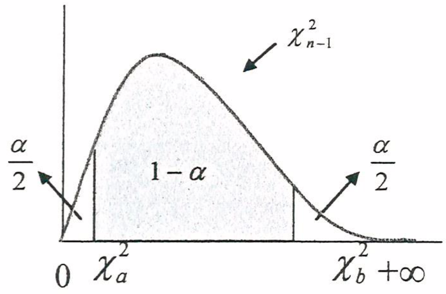
$$P(0,30 - 0,085 < \pi < 0,30 + 0,085) = 95\%$$

$$P(0,215 < \pi < 0,385) = 95\%$$

6.9 INTERVALO DE CONFIANÇA PARA A VARIÂNCIA POPULACIONAL

$$\frac{(n-1)\sigma^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$P(\chi^2_a < \frac{(n-1)\hat{\sigma}^2}{\sigma^2} < \chi^2_b) = 1-\alpha$$



$$P\left(\frac{1}{\chi^2_b} < \frac{\sigma^2}{(n-1)\hat{\sigma}^2} < \frac{1}{\chi^2_a}\right) = 1-\alpha$$

$$P(\chi^2_a < \chi^2_{n-1} < \chi^2_b) = 1-\alpha$$

$$\boxed{P\left(\frac{(n-1)\hat{\sigma}^2}{\chi^2_b} < \sigma^2 < \frac{(n-1)\hat{\sigma}^2}{\chi^2_a}\right) = 1-\alpha}$$

Figura 16 Quadro Resumo: I. C. para a variância

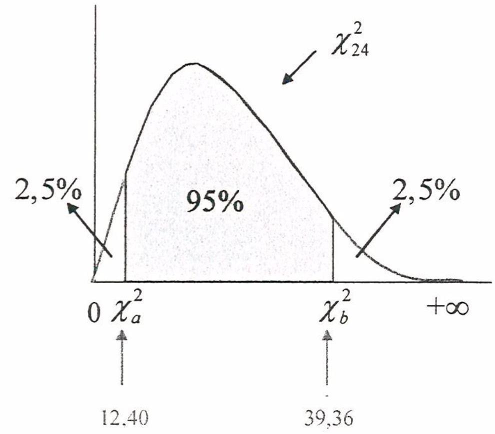
Exemplo:

Uma v.a. qualquer tem uma distribuição desconhecida com média μ e variância σ^2 desconhecidas. Retira-se uma amostra de 25 valores e calcula-se a variância amostral. Construa um I.C de 95% para σ^2 supondo que $\hat{\sigma}^2 = 2{,}34$.

$$P\left(\frac{(n-1)s^2}{\chi_b^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_a^2}\right) = 0,95$$

$$P\left(\frac{24*2,34}{39,36} < \sigma^2 < \frac{24*2,34}{12,40}\right) = 0,95$$

$$\boxed{P\left(1,43 < \sigma^2 < 4,53\right) = 0,95}$$



6.10 INTERVALO DE CONFIANÇA PARA A SOMA OU DIFERENÇA DE DUAS MÉDIAS

Duas populações normais, X_1 e X_2 apresentam $\sigma_{X1} = 5\, e\ \sigma_{X2} = 2$. Uma amostra aleatória de 12 elementos da primeira população apresentou $\overline{X} = 34$. Uma amostra de 8 elementos da segunda população apresentou $\overline{X} = 9,4$. Calcule o intervalo de confiança de 98 % para a diferença entre as médias, $\mu_{X1} - \mu_{X2}$.

Solução:

$$P\left[(\overline{x}_1 \pm \overline{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}} < \mu_1 \pm \mu_2 < (\overline{x}_1 \pm \overline{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}}\right] = 1 - \alpha$$

$$P\left[(34-9,4) - 2,33\sqrt{\frac{5^2}{12} + \frac{2^2}{8}} < \mu_1 - \mu_2 < (34-9,4) + 2,33\sqrt{\frac{5^2}{12} + \frac{2^2}{8}}\right] = 0,98$$

$$P(20,86 < \mu_1 - \mu_2 < 28,36) = 0,98$$

6.10.1 Com variância populacional desconhecida

$$P\left[(\bar{x}_1 \pm \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{\hat{\sigma}_1}{n_1} + \frac{\hat{\sigma}_2}{n_2}} < \mu_1 \pm \mu_2 < (\bar{x}_1 \pm \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{\hat{\sigma}_1}{n_1} + \frac{\hat{\sigma}_2}{n_2}}\right] = 1 - \alpha$$

6.10.2 Exercício resolvido:

Uma amostra de 10 elementos selecionados ao acaso de uma população normal apresentou média igual a 18 e desvio padrão estimado igual a 3. Uma amostra de 15 elementos selecionada, ao acaso, de uma população normal apresentou média igual a 25 e desvio padrão estimado igual a. Supondo as populações independentes, calcule o intervalo de confiança de 95% para $\mu_{X1} + \mu_{X2}$.

Solução:

$$P\left[(\bar{x}_1 \pm \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{\hat{\sigma}_1}{n_1} + \frac{\hat{\sigma}_2}{n_2}} < \mu_1 \pm \mu_2 < (\bar{x}_1 \pm \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{\hat{\sigma}_1}{n_1} + \frac{\hat{\sigma}_2}{n_2}}\right] = 1 - \alpha$$

$$P\left[(18 + 25) - 2,06\sqrt{\frac{3^2}{10} + \frac{4^2}{15}} < \mu_1 + \mu_2 < (18 + 25) + 2,33\sqrt{\frac{3^2}{10} + \frac{4^2}{15}}\right] = 0,95$$

$$P(40,11 < \mu_1 + \mu_2 < 45,89) = 0,95$$

6.11 INTERVALO DE CONFIANÇA PARA A SOMA OU DIFERENÇA DE DUAS PROPORÇÕES

$$P\left[(\hat{\pi}_1 \pm \hat{\pi}_2) - z_{\alpha/2}\sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}} < \pi_1 \pm \pi_2 < (\hat{\pi}_1 \pm \hat{\pi}_2) + z_{\alpha/2}\sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}\right] = 1-\alpha$$

6.11.1 Exemplo Resolvido

A assessoria de um candidato à Presidência da República efetuou um levantamento amostral em dois estados. No estado A foram selecionados 120 eleitores ao acaso e verificou-se que 36 votariam neste candidato. No Estado B foram selecionados 80 eleitores, também ao acaso, e verificou-se que 22 votariam neste candidato. Construa um intervalo de confiança de 95 % para a diferença $\pi_1 - \pi_2$, entre as proporções de eleitores deste candidato nos estados A e B.

Solução

$$P\left[(\hat{\pi}_1 \pm \hat{\pi}_2) - z_{\alpha/2}\sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}} < \pi_1 \pm \pi_2 < (\hat{\pi}_1 \pm \hat{\pi}_2) + z_{\alpha/2}\sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}\right] = 1-\alpha$$

$$P\left[(0,3_1 + 0,275) - 1,96\sqrt{\frac{0,3*0,7}{120} + \frac{0,275*0,725}{80}} < \pi_1 \pm \pi_2 < (0,3_1 - 0,275) + -1,96\sqrt{\frac{0,3*0,7}{120} + \frac{0,275*0,725}{80}}\right] = 95\%$$

$$P[-0,025 < \pi_1 \pm \pi_2 < 0,230\ ] = 95\%$$

CAPÍTULO 7. TESTE DE HIPÓTESES

7.1 CONSIDERAÇÕES INICIAIS

A principal consideração ao iniciar um teste de hipóteses refere-se a importância de se saber perfeitamente o objetivo com que o teste será feito. Isso porque, uma vez que você está trabalhando com amostras, está se está trabalhando com informação incompleta ou informação parcial. Ou seja, o que a amostra dá é uma idéia; é uma quantidade suficiente ou não sobre evidência de alguma coisa que pode estar ocorrendo ou não na população. Em resumo, o objetivo de um teste de cotizes é obter a evidência de alguma coisa que está ocorrendo na população.

Uma hipótese é uma afirmação sobre a população; afirmação sobre os parâmetros da população. É necessário, portanto, ver se a amostra fornece evidências suficientes para que se aceite que na população sua informação é verdadeira.

É importante ressaltar que o teste de hipótese jamais provará alguma coisa; ele não é um processo de demonstração da verdade. É um mecanismo que auxilia na tomada de decisão em condições de incerteza. Ele dá o peso que a evidência que a amostra fornece em uma ou outra direção.

7.2 INTERVALO DE CONFIANÇA PARA A PROPORÇÃO

Para exemplificar os fundamentos e o objetivo de um teste de hipóteses, suponha que se queira analisar se uma moeda é viciada ou não.

A primeira pergunta que se faz aqui é: a moeda é viciada? Para responder esta pergunta, pode-se fazer um experimento, ou seja, lançar a moeda 100 vezes ($n = 100$) e contar, por exemplo, a freqüência com que a face cara aparece nos 100 lançamentos. Assim, tem-se uma certa proporção na amostra.

A segunda pergunta pertinente a este processo é: Como decidir se esta moeda é honesta já que, por exemplo, obteve-se 56 % de caras? Não se sabe e nunca se saberá. Assim, fica evidente que para decidir se a moeda é honesta ou não deve-se construir uma regra de decisão. Sabe-se que uma moeda é perfeita estatística mente quando □ = 0,5.

Como não se conhece π, só se conhece p (proporção amostral) pode-se estabelecer a seguinte regra de decisão:

0,402　　　0,51　　　0,598

Essa é a regra de decisão onde estão ocorrendo todos os possíveis valores para p na amostra. Em síntese, esta regra está considerando que se p estiver entre 0,402 e 0,598 tem-se evidências suficientes para aceitar que a moeda é honesta.; caso contrário tem-se evidências suficientes para acreditar que a moeda é viciada. Ou seja, para $0,402 < p < 0,598$ a moeda é boa caso contrário, é viciada.

A terceira pergunta refere-se às hipóteses que estão sendo confrontadas. As hipóteses que estão sendo confrontadas são; neste caso, as de a moeda ser ou não viciada.

Tem-se, portanto, duas hipóteses: a chamada hipótese nula (Ho) (hipótese contra o pesquisador). Esta hipótese indica a ausência do fenômeno, ou seja, a ausência de vício. A outra hipótese é a chamada hipótese alternativa (Ha).

Assim,

Ho: $\pi = 0,5$

Ho: $\pi \neq 0,5$

Em um teste de hipótese, saber especificar as hipóteses é muito importante, pois as decisões são sempre sobre a hipótese nula.

Testar uma hipótese consiste tipicamente em decidir por uma de duas alternativas. A primeira, indicada por Ho, é chamada hipótese nula e expressa a ausência, na população, da propriedade ou da relação de interesse do pesquisador. A segunda, indicada por Ha, é chamada hipótese alternativa e é complementar a Ho, indicando a presença, na população, da característica estudada. A hipótese nula é a hipótese de referência do teste. É sobre ela que decidimos: ou aceitamos Ho como verdadeira (e conseqüentemente rejeitamos Ha), ou não aceitamos Ho como verdadeira (e conseqüentemente aceitamos Ha). Neste processo, há quatro combinações possíveis entre nossa decisão e os estados do mundo, a saber:

Decisão que tomamos (resultado do teste)	Estados do Mundo	
	Ho é Verdadeira	Ho é Falsa
Aceita Ho	Decisão Correta	ERRO TIPO II
Rejeita Ho	ERRO TIPO I	Decisão Correta

O nível de significância de um teste de hipótese é a probabilidade de cometermos o Erro Tipo I, isto é, a probabilidade de rejeitarmos Ho quando Ho é verdadeira. O nível de significância é indicado por α e é igual a 1–nível de confiança. Em geral utilizamos um nível de significância em torno de 5%.

$$\alpha + (1 - \alpha) = 1$$
$$\beta + (1 - \beta) = 1$$

O erro tipo I acontece quando a proporção p encontrada na amostra cai fora do intervalo e a moeda é não viciada.

0,402 0,51 0,598

Assim, α = P(erro I) = P(não aceitar $H_0/\pi = 0{,}5$). Sob esta regra de decisão, 5% das vezes que não se aceitar Ho, comete-se erro Tipo I.

O Erro tipo II só ocorre quando se aceita Ho e Ho é falsa.

Suponha agora que $\hat{\pi} = 0{,}6$. Qual a probabilidade de se cometer erro Tipo II? Agora está se supondo Ho falso pois $\pi = 0{,}5$ e a distribuição

$$p : N(E(\hat{\pi}) = 0{,}5; \sigma_{\hat{\pi}}^{2} = \sqrt{\frac{\pi(1-\pi)}{100}})$$

onde

$$\sqrt{\frac{\pi(1-\pi)}{100}} = \sqrt{\frac{0{,}5*0{,}5}{100}} = 0{,}05$$

Para calcular a probabilidade tem-se:

$$z_{0,598} = \frac{0{,}598 - 0{,}6}{0{,}049} = -0{,}04$$

$$z_{0,402} = \frac{0{,}402 - 0{,}6}{0{,}049} = -4{,}05$$

P(erro II) = 0,484 se $\pi = 0{,}6$

Mas se $\pi = 0{,}7$?

$$E(\hat{\pi}) = \pi \qquad e \qquad \sigma_{\pi} = \sqrt{\frac{0,7 * 0,3}{100}}$$

Exemplo:

Em 100 lançamentos de uma moeda foram observadas 62 caras. Teste ao nível de significância de 5% a hipótese de que a moeda é honesta.

$H_0 : \pi = 0,5$

$H_0 : \pi \neq 0,5$

Na realidade precisa-se saber se a distância, "diferença" entre o modelo padrão e o efetivamente observado é grande ou não; se é significante ou não.

Sabe-se que a distribuição de p sob Ho é dada por:

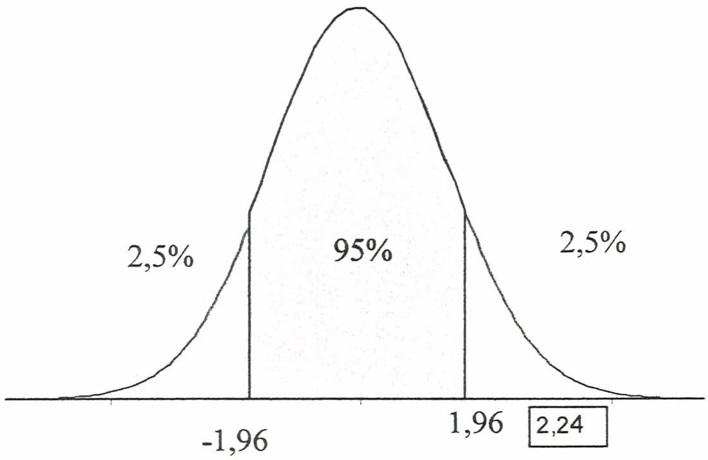
$$z = \frac{\hat{\pi} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \sim N(0,1) \qquad \sigma_{\hat{\pi}} = \sqrt{\frac{\pi(1-\pi)}{n}}$$

z: estatística de teste: é o valor na amostra da estimativa do parâmetro que se está estimando.

$$z = \frac{\hat{\pi} - E(\hat{\pi})}{\sigma_{\hat{\pi}}} \sim N(0,1)$$

$$z = \frac{0,62 - 0,5}{\sqrt{\frac{0,5(1-0,5)}{100}}} = 0,12$$ $\boxed{\mathbf{z = 0,12}}$ afastamento

$$z = \frac{0,12}{0,05} = 2,4$$ $\boxed{\mathbf{z = 2,4}}$ estatística de teste para a amostra.



Isso quer dizer que: com esta amostra e com esse nível de significância têm-se evidências significantes para rejeitar a hipótese de que a moeda é honesta.

7.3 INTERVALO DE CONFIANÇA PARA MÉDIA

$X \sim N(\mu, \sigma^2)$ $\quad \mu$ desconhecido, mas σ^2 conhecido

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

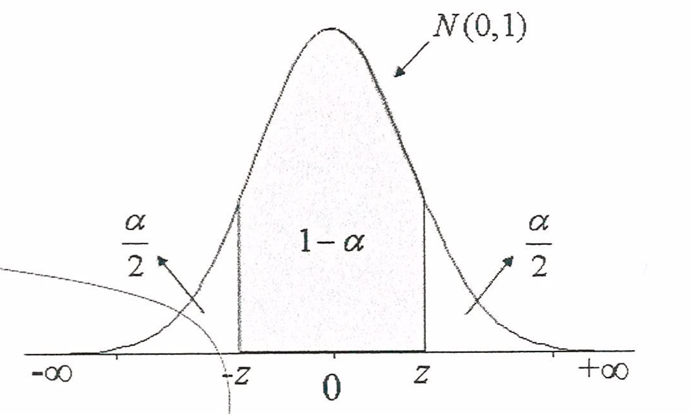
$$Z \quad \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \quad \text{(Normal Padrão)}$$

$$P(-z < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z) = 1 - \alpha$$

$$P(-z\frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

$$\boxed{P(\bar{X} - z\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z\frac{\sigma}{\sqrt{n}}) = 1 - \alpha}$$

IC para μ



$P(|Z| > z) = \alpha$ nível de significância

$P(-z < Z < z) = 1 - \alpha$ nível de confiança

7.3.1 Exercício Resolvido

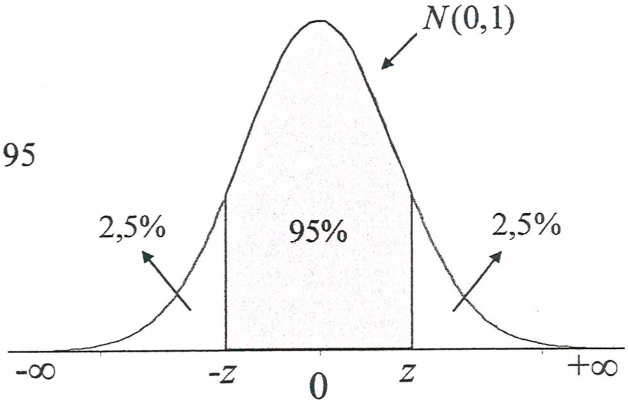
Uma v.a. qualquer tem uma distribuição desconhecida com média μ, também desconhecida e variância s^2 = 16. Retira-se uma amostra de 25 valores e calcula-se a média amostral. Construa um IC de 95% para μ supondo que $\bar{x} = 12,7$.

$$P(\bar{X} - z\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z\frac{\sigma}{\sqrt{n}}) = 0,95$$

$$P(12,7 - 1,96\frac{4}{\sqrt{25}} < \mu < 12,7 + 1,96\frac{4}{\sqrt{25}}) = 0,95$$

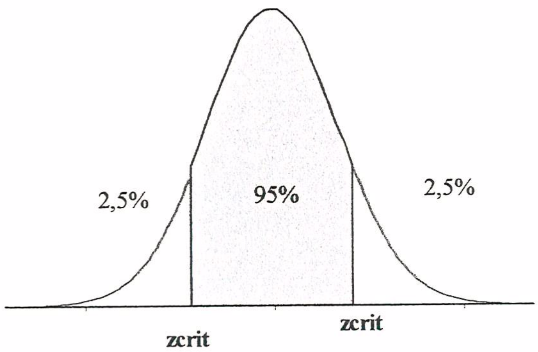
$$P(12,7 - 1,568 < \mu < 12,7 + 1,568) = 0,95$$

$$\boxed{P(11,132 < \mu < 14,268) = 0,95}$$



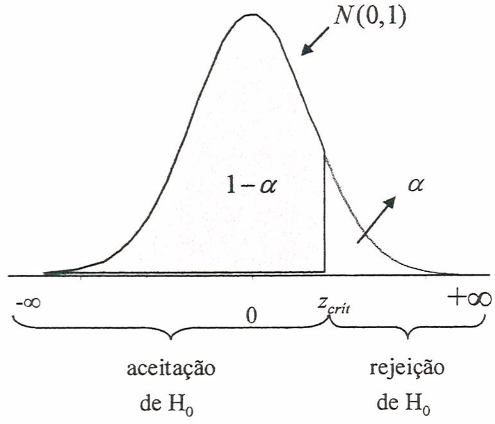
7.3.2 Teste bilateral

$$H_0 : \mu = \mu_0$$
$$H_a : \mu \neq \mu_0$$



$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

7.3.3 Teste unilateral



Se H_0 é verdadeira, então

$$z = \frac{\bar{X} - 100}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Região Crítica:

•aceito H_0 se $z < z_{crit}$ $\rightarrow P(z < z_{crit}) = 1 - \alpha$
•rejeito H_0 caso contrário $\rightarrow P(z > z_{crit}) = \alpha$

Conclusão (sempre associada a um nível de significância)

$$z_{crit} = \frac{\bar{X}_{crit} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \Rightarrow \bar{X}_{crit} = \mu_0 + z_{crit}\frac{\sigma}{\sqrt{n}}$$

Existe a possibilidade de se selecionar uma amostra de uma população com média μ_0 e obter $\bar{x}$ alto de forma que leve a conclusão errada de que H_0 é falsa. Este erro é chamado de erro do tipo I , cuja probabilidade de ocorrência é o nível de significância α.

P(rejeitar H_0 / H_0 é verdadeira) = α

P(aceitar H_0 / H_0 é verdadeira) = 1 - α

7.3.4 Erro Tipo II

Existe a possibilidade de se selecionar uma amostra de uma população com média $\mu > \mu_0$ e obter $\bar{x}$ de forma que leve a conclusão errada de que H_0 é verdadeira. Este tipo de erro é denominado Erro Tipo II.

P(aceitar H0 / H1 é verdadeira) = β

P(rejeitar H0 / H1 é verdadeira) = 1 - β (poder do teste)

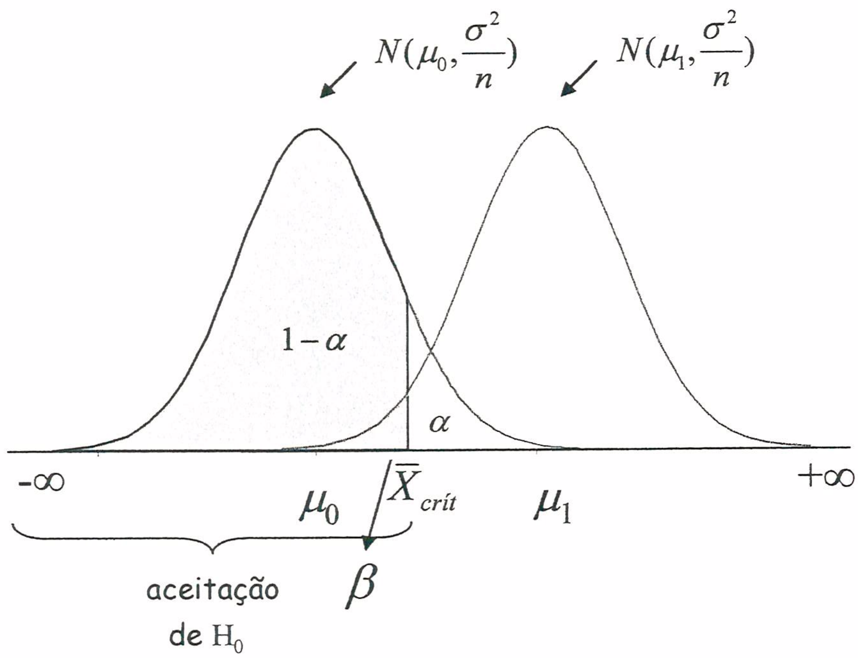


Figura 17 Erro Tipo II

7.3.5 Exercício Resolvido

Exemplo: Usando o mesmo exemplo anterior, calcule a probabilidade de se aceitar H0 (μ = 1800), para o caso da verdadeira média ser 1850 kg.

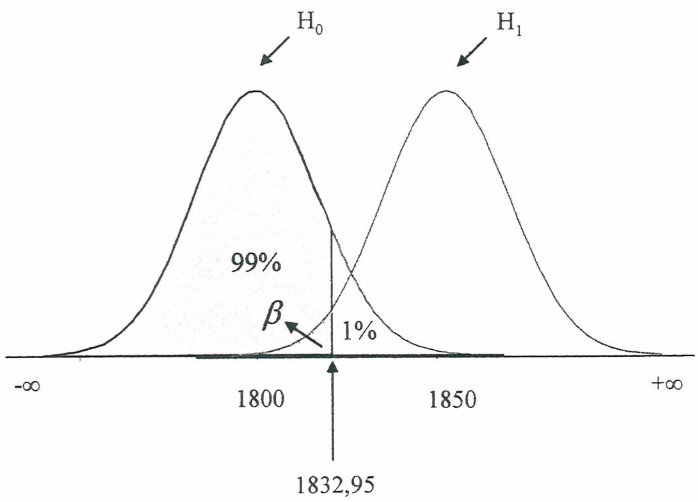


Figura 18 Erro Tipo II

$H_0 : \mu = 1800$
$H_1 : \mu = 1850$

$\beta = P(\text{aceitar } H_0 \,/\, H_1 \text{ é verdadeiro})$

$$\beta = P(Z < 2,33 / \mu = 1850)$$

$$\beta = P(\frac{\overline{X} - 1800}{\frac{100}{\sqrt{50}}} < 2,33 / \mu = 1850)$$

$$\beta = P(\overline{X} < 1832,95 / \mu = 1850)$$

$$\beta = P(\frac{\overline{X} - 1850}{\frac{100}{\sqrt{50}}} < \frac{1832,95 - 1850}{\frac{100}{\sqrt{50}}})$$

$$\beta = P(Z < -1,21) = P(Z > 1,21) = 0,5 - 0,3869 = 0,1131$$

7.3.6 Interpretação p-value

Foram coletadas amostras (50 pontos) em mapas a fim de avaliar sua exatidão. Procedeu-se o teste z para verificar quais deles possuíam exatidão (p) de 0,90. A tabela abaixo apresenta a exatidão estimada, o resultado do teste (estatística z) e o valor-P de cada mapa.

$$z = \frac{\hat{\pi} - \pi}{\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}}$$

		z	valor-P
Mapa 1	0,87	-0,707	0,2397
Mapa 2	0,62	-6,600	2,07e-11
Mapa 3	0,82	-1,886	0,0297
Mapa 4	0,84	-1,414	0,0786

7.4 TESTE DE HIPÓTESES PARA A DIFERENÇA DE MÉDIAS

$$\boxed{\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2} \quad \text{e} \quad \boxed{\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\left(\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}\right)}} \quad \boxed{z = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\left(\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}\right)}}}$$

Hipótese alternativa	**Rejeitar a hipótese nula se**	***Aceitar a hipótese nula ou reservar julgamento se***
$\mu_1 - \mu_2 < \delta$	$z \leq -z_\alpha$	$z > -z_\alpha$
$\mu_1 - \mu_2 > \delta$	$z \geq z_\alpha$	$z < z_\alpha$
$\mu_1 - \mu_2 \neq \delta$	$z \leq -z_{\alpha/2}$ ou $z \geq z_{\alpha/2}$	$-z_{\alpha/2} < z < z_{\alpha/2}$

7.4.1 Exercícios resolvidos

Em um estudo para testar se há diferenças de alturas médias entre as mulheres adultas nascidas em dois países diferentes, obtiveram-se os seguintes resultados de amostras aleatórias:

$n_1 = 120$	$\bar{x}_1 = 62{,}7$	$s_1 = 2{,}50$
$n_1 = 150$	$\bar{x}_2 = 61{,}8$	$s_2 = 2{,}62$

(medidas em polegadas). Ao nível de 0,05 de significância, teste a hipótese nula de que as médias populacionais correspondentes são iguais, contra a hipótese alternativa de que não são iguais. Considere que as variâncias populacionais sejam iguais.

$H_0 : \delta = 0$

$H_A : \delta \neq 0$

$\alpha = 0{,}05$

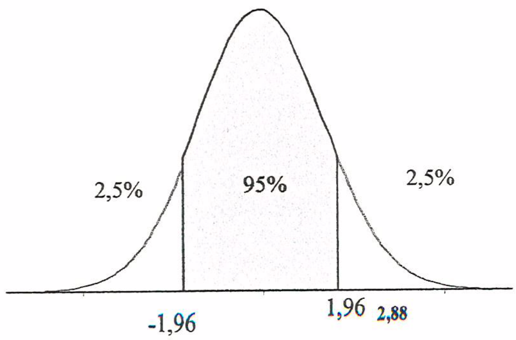
Rejeitar a hipótese nula se $z \leq -1{,}96$ ou $z \geq 1{,}96$, onde

$$\boxed{z = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\left(\frac{\sigma_1^{\ 2}}{n_1} + \frac{\sigma_2^{\ 2}}{n_2}\right)}}}$$

com σ_1 e σ_2 substituídos por s_1 e s_2 e fazendo $\delta = 0$;

E substituindo s_1 e s_2 por $\sigma_1 = 2{,}50$ e s_2 por $\sigma_2 = 2{,}62$ na fórmula de z, obtemos

$$z = \frac{(62{,}7 - 61{,}8)}{\sqrt{\left(\frac{2{,}50^2}{120} + \frac{2{,}62^2}{150}\right)}} \approx 2{,}88$$



Como z = 2,88 excede 1,96, a hipótese nula deve ser rejeitada; em outras palavras, a diferença entre $\bar{x}_1 = 62{,}7$ e $\bar{x}_2 = 61{,}8$ é significativa.

Refaçamos agora este exemplo, utilizando o valor-p, o que exige modificação dos passos 3, 4 e 5.

$$\boxed{z = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}}} \qquad z = \frac{(62{,}7 - 61{,}8)}{\sqrt{\left(\frac{2{,}50^2}{120} + \frac{2{,}62^2}{150}\right)}} \approx 2{,}88$$

A área sobre a curva normal a direita de z = 2,88 é 0,0020. Esta é também a área sob a curva à esquerda de z = -2,88; assim, o valor de p é

$$0{,}0020 + 0{,}0020 = 0{,}0040$$

Como o valor p é inferior a $\alpha = 0{,}05$, a hipótese nula deve ser rejeitada.

Note-se que é um tanto estranho o fato de compararmos médias quando σ_1 e σ_2 têm valores diferentes (ou quando os valores amostrais s_1 e s_2 não estão próximos um do outro). Consideremos, por exemplo, duas populações normais com médias $\mu_1 = 50$ e $\mu_2 = 52$ e desvios-padrão $\sigma_1 = 5$ e $\sigma_2 = 15$. Embora a Segunda população tenha maior média, é muito mais provável que ela apresente um valor abaixo de 40. Defrontando situações como esta, o pesquisador deve decidir se a comparação de μ_1 e μ_2 realmente focaliza o problema de interesse.

7.4.2 Diferenças entre médias: Pequenas amostras

Estatística para teste em pequenas amostras da diferença entre duas médias

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\left(\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)}}$$

Desvio-Padrão combinado

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

7.4.3 Estatística para o teste de diferença entre duas médias (Pequenas Amostras)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

7.4.3.1 Exercício Resolvido

As amostras aleatórias seguintes são medidas dos pesos de sacos de carvão retirados de duas minas diferentes.

Mina 1	8.400	8.230	8.380	7.860	7.930
Mina 2	7.510	7.690	7.720	8.070	7.660

Ao nível de 0,05 de significância, teste se a diferença entre as médias destas duas amostras é significativa.

Solução:

Inicialmente para a suposição de que as duas amostras provêm de populações normais com mesmo desvio-padrão.

$H_0 : \delta = 0$

$H_A : \delta \neq 0$

$\alpha = 0,05$

Rejeitar a hipótese nula se $t \leq -2,306$ ou $t \geq 2,306$, onde

$$\boxed{t=\frac{(\bar{x}_1-\bar{x}_2)-\delta}{\sqrt{\left(\frac{(n_1-1)\hat{\pi}_1^{\,2}+(n_2-1)\hat{\pi}_2^{\,2}}{n_1+n_2-2}\times\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right)}}}$$

com $\delta = 0$ e 2,306 é o valor de $t_{0,025}$ para

$$5+5-2=8$$

graus de liberdade; em caso contrário, manter a afirmação de que a diferença entre as médias das duas amostras não é significativa.

As médias e as variâncias das duas amostras são $\bar{x}_1 = 8.160$, $\bar{x}_2 = 7.730$, s_1^2 = 63.450, s_2^2 = 42.650. Levando estes valores, juntamente com $n_1 = 5$, $n_2 = 5$ e $\delta = 0$, na fórmula de t, obtemos:

$$t=\frac{(8.160-7.730)}{\sqrt{\left(\frac{(63.450)4+4(42.650)}{5+5-2}\cdot\left(\frac{1}{5}+\frac{1}{5}\right)\right)}}\approx 2{,}95$$

Como t = 2,95 excede 2,306, a hipótese nula deve ser rejeitada; em outras palavras, concluímos que a capacidade média de geração de calor não é a mesma para o carvão de uma ou de outra mina.

7.4.4 Diferenças entre duas médias: (dados emparelhados)

7.4.5 Exercício Resolvido

Os dados abaixo referem-se, respectivamente, às notas de 10 operários obtidas antes e depois de um programa de treinamento para operar um novo tipo de máquina.

45 e 36	73 e 60	46 e 44	124 e 119	33 e 35
57 e 51	83 e 77	34 e 29	26 e 24	17 e 11

Teste a eficácia do programa de treinamento, ao nível de 0,05.

Como o tamanho n = 10 da amostra é pequeno, partimos da suposição de que as diferenças constituem uma amostra de uma população normal. As diferenças são:

9	13	2	5	-2	6	6	5	2	6

e para estes dados fazemos o seguinte teste:

$H_0 : \mu = 0$

$H_A : \mu > 0$

(A alternativa é que, em média, haja mais acidentes "antes" do que "depois".)

α =0,05

Rejeitar a hipótese nula se $t \geq 1{,}833$, onde

$$\boxed{t = \frac{(\bar{x} - \mu_0)}{s / \sqrt{n}}}$$

e 1,833 é o valor de $t_{0{,}05}$ para 10 – 1 = 9 graus de liberdade; em caso contrário, aceitar a hipótese nula ou reservar julgamento (conforme a situação recomenda).

Calculando inicialmente a média e o desvio-padrão das dez diferenças, obtemos $\bar{x} = 5{,}2$ e $s = 4{,}08$. Levando, então, na fórmula de t esses valores, juntamente com $n = 10$ e $\mu_0 = 0$, vem $t = \frac{(5{,}2 - 0)}{4{,}08 / \sqrt{10}} \approx 4{,}03$. Como $t = 4{,}03$ excede 1,833, a hipótese nula deve ser rejeitada; em outras palavras, mostramos que o programa de segurança industrial é eficaz.

CAPÍTULO 8. EXERCÍCIOS PROPOSTOS

1) Você está participando de um estudo sobre a velocidade de condução e a distância percorrida em quilômetros para os automóveis de tamanho médio. Você solicita para seu auxiliar, o cálculo da covariância e do coeficiente de correlação entre essas duas variáveis (velocidade de condução e distância percorrida). Após alguns minutos, seu auxiliar lhe informa que com os cálculos feitos na planilha Excel ele obteve um coeficiente de correlação amostral igual a 0,89 e uma covariância igual a 0. Com base nesses resultados, que tipo de comentário você faria com seu auxiliar?

2) Numa urna tem-se 5 tiras de papel, numeradas 1,3, 3 8, e 7. Uma tira de papel é sorteada, recolocada na urna, e uma segunda tira é sorteada. Sejam X1 e X2 , respectivamente, o primeiro e o Segundo número sorteado:

a) Determine a distribuição conjunta de X1 e X2;
b) Ache as distribuições marginais de X1 e X2. Elas são independentes?

Suponha que X e Y tenham a seguinte tabela de distribuição conjunta:

Y \ X	1	2	3
1	0,1	0,1	0
2	0,1	0,2	0,2
3	0,1	0,1	0,1

a) Determinar a função de probabilidade de X + Y e, a partir daí, calcular E(X+Y)
b) Determinar a função de probabilidade de X Y e, em seguida, calcular E(XY)
c) Calcule a covariância de X+Y e XY

3) Uma urna contém 3 bolas numeradas 1, 2, 3. Duas bolas são retiradas ao acaso e sucessivamente. X = número da primeira bola retirada e Y o número da segunda bola retirada. Calcule:

a) E(X Y);
b) Cov (X,Y)
c) Var (X + Y)

4) O que você entende por distribuição amostral da média? E por distribuição amostral da variância?

5) Uma população se constitui dos números 2, 4, 6, 8. Considere todas as amostras possíveis de tamanho 2, que podem ser retiradas, sem reposição, desta população. Determine:

   a) média da população
   b) desvio padrão da população
   c) a distribuição amostral da média.
   d) a média da distribuição amostral da média
   e) o desvio padrão da distribuição amostral da média

6) Uma amostra de 25 observações de uma variável normal apresentou uma média igual a 110 e uma variância estimada igual a 960. Construir os intervalos de 90 % de confiança para os parâmetros média e variância.

7) As companhias de seguro estão ficando preocupadas com o fato de que o número crescente de telefones celulares resulte em maior número de acidentes de carros; estão, por isso, pensando em cobrar prêmios mais elevados para os motoristas que utilizam celulares. Deseja-se estimar, com uma margem de erro de 3 pontos percentuais, a percentagem de motoristas que falam ao celular enquanto estão dirigindo. Considerando que se pretende um nível de confiança de 95% nos resultados, quantos motoristas devem ser pesquisados:

   a) supondo que, com base em estudos realizados, tem-se que 18% dos motoristas falam ao celular enquanto dirigem, isto é: $\hat{\pi}=18\%$;
   b) supondo que não se tenha qualquer informação que possa sugerir um valor de p, ou seja, um valor para a estimativa da verdadeira proporção, π.

8) O que você pode concluir a partir dos resultados obtidos nos itens (a) e (b)?

9) Considere a seguinte amostra de uma variável de média e variância desconhecidas: 3, 1, 5, 2, 3, 4, 5, 2, 2, 4, 6. Pede-se:

a) estimar a média e a variância populacionais
b) construir o intervalo para média populacional com uma probabilidade de confiança de 95%.

10) Os valores da média e do desvio padrão de 25 vendas a crédito realizadas por uma grane loja, em um determinado mês, foram iguais, respectivamente, a R$ 240,00 e R$ 36,00. Construir o intervalo de confiança de 90% para o valor da média de to

11) Uma amostra de 30 torcedores porto-alegrenses revelou as seguintes preferências clubísticas, onde G indica "Grêmio", C indica "!Internacional", e N indica "outro clube":

C G G N G N C C N G G G N C G C N G G C G N N N C G C C G C

Com base nesta amostra, estimar a proporção dos torcedores porto-alegrenses que preferem o Grêmio.

12) Quer-se estimar a proporção de porto-alegrenses maiores de 16 anos que são favoráveis à flexibilização das leis trabalhistas. Qual o tamanho mínimo da amostra necessário para um erro absoluto de estimação máximo de 0,02, com um nível de confiança de 90%? E com um nível de 95%?

13) Uma moeda "chumbada" foi lançada 100 vezes, obtendo-se 38 caras. A partir desta amostra, construir o intervalo de confiança de 95% para a verdadeira probabilidade do resultado "cara" dessa moeda. Suponha, agora, que uma moeda "perfeita" vai ser lançada 100 vezes. Qual a probabilidade de obtermos 38 caras ou menos?

14) Em certa cidade deseja-se estimar a proporção de pessoas que são favoráveis a construção de um viaduto no centro da cidade.. Supondo que numa amostra de 100 pessoas desta cidade, 75 não são favoráveis a construção de tal viaduto, qual será o intervalo com 99% de confiança para a proporção p? Qual o tamanho da amostra para se Ter uma confiança de 90?

15) No ano passado, uma empresa amostrou aleatoriamente 36 funcionários da área de produção e constatou que 8 deles usavam o próprio carro para ir ao trabalho. A partir deste ano, a empresa começou a fornecer o vale-transporte aos seus funcionários. Agora, a empresa amostrou aleatoriamente 40 funcionários da mesma área de produção e constatou que 8 deles usam o próprio carro para ir ao trabalho. Calcule um intervalo de confiança de 90 % para a diferença, antes e depois do fornecimento do vale-transporte, entre as proporções de funcionários que vão ao trabalho com carro próprio.

16) Qual o tamanho mínimo que deve ter uma amostra para se estimar a idade média, em anos, dos estudantes de segundo grau em escolas estaduais de Porto Alegre, para que o erro absoluto de estimação seja no máximo 2,5 anos, a um nível de confiança de 98 %?

17) Uma pesquisa efetuada com 135 funcionários selecionados ao acaso, e sem reposição, entre os 600 funcionários de uma empresa, revelou que 54 deles não mantinham convênio com qualquer empresa particular de assistência médica. Construa um intervalo de confiança de 98 % para a proporção de funcionários desta empresa que mantém convênio com alguma empresa particular de assistência médica.

18) Para estudar a situação salarial em uma empresa, um consultor coletou uma amostra aleatória de 50 salários recebidos na empresa. O salário médio amostral é igual a em 120 u.m e o desvio padrão estimado é igual a 20 u.m.. Determine um intervalo de confiança de 95 % para o salário médio pago por esta empresa.

19) Qual o tamanho mínimo que deve ter uma amostra para se estimar a idade média, em anos, dos estudantes de segundo grau em escolas estaduais de Porto Alegre, para que o erro absoluto de estimação seja no máximo 2,4 anos, a um nível de confiança de 98 %?

20) Uma empresa recebe de dois fornecedores, fornecedor 1 e fornecedor 2, caixas de papelão que são utilizadas como embalagens. O gerente de produção desta empresa alega que o peso médio das caixas de papelão recebidas do fornecedor 1 são diferentes que o peso médio das caixas de papelão recebidas do fornecedor 2. Para testar essa alegação, o gerente de produção retirou aleatoriamente uma amostra de 10 caixas de cada fornecedor. O peso médio das caixas entregues pelo fornecedor 1 foi igual a 39 gramas com um desvio padrão estimado de 7 gramas. O peso médio das caixas entregues pelo fornecedor 2 apresentaram uma média de 43 gramas e um desvio padrão estimado de 9 gramas. Sabendo-se que as variâncias populacionais são iguais e utilizando um nível de significância de 5%, o que você pode concluir sobre a alegação do gerente de produção:

21) Para testar se uma moeda é equilibrada, adota-se a seguinte regra de decisão. Rejeitar a hipótese de equilíbrio se X, o número de caras obtidas em 100 jogadas, for ou menor do que 40 ou maior do que 60. Determinar a

probabilidade de um erro tipo I.

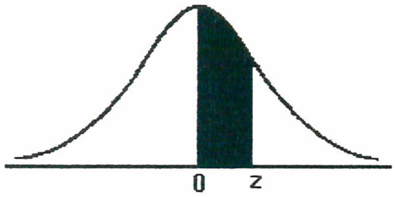
22) O gerente de um hotel estabeleceu que a quantia média gasta pelos hóspedes em um fim de semana é de R$ 400,00 ou menos. Um funcionário do departamento de contabilidade do hotel notou que as despesas totais para os hóspedes de fim de semana têm aumentado nos últimos meses. O contador usará uma amostra de contas de hóspedes de fim de semanas para testar a afirmação do gerente. Especifique as hipóteses adequadas para testar a afirmação do gerente.

23) Uma nova série de televisão precisa ter evidências que tem mais do que 25% da audiência de telespectadores depois das 13 primeiras semanas de exibição para ser julgada bem-sucedida. Em uma amostra de 400 famílias, 112 estavam vendo a nova série. Com um nível de significância de 5%, a série pode ser julgada bem-sucedida?

24) Ana afirma que o salário médio de uma classe de trabalhadores é duas populações X1 e X2, normalmente distribuídas apresentam desvio 4 e desvio 3,4, respectivamente. Uma amostra aleatória de 20 elementos, selecionados da população X1, apresentam média 45, enquanto que outra amostra de 30 elementos, selecionados da população X2, apresentam média 43. Teste ao nível de significância de 5% a hipótese de que as médias populacionais são iguais.

25) Uma grande loja de departamentos planeja oferecer para seus clientes o serviço de compras via Internet. O serviço ficará disponível se mais de 40% dos usuários de Internet tiverem o hábito de fazer compras via Internet. A partir de uma amostra de 30 usuários de Internet, 17 deles declararam que a

APÊNDICE A DISTRIBUIÇÃO NORMAL PADRÃO



z	,00	,01	,02	,03	,04	,05	,06	,07	,08	,09
0,0	,0000	,0040	,0080	,0120	,0160	,0199	,0239	,0279	,0319	,0359
0,1	,0398	,0438	,0478	,0517	,0557	,0596	,0636	,0675	,0714	,0753
0,2	,0793	,0832	,0871	,0910	,0948	,0987	,1026	,1064	,1103	,1141
0,3	,1179	,1217	,1255	,1293	,1331	,1368	,1406	,1443	,1480	,1517
0,4	,1554	,1591	,1628	,1664	,1700	,1736	,1772	,1808	,1844	,1879
0,5	,1915	,1950	,1985	,2019	,2054	,2088	,2123	,2157	,2190	,2224
0,6	,2257	,2291	,2324	,2357	,2389	,2422	,2454	,2486	,2518	,2549
0,7	,2580	,2612	,2642	,2673	,2704	,2734	,2764	,2794	,2823	,2852
0,8	,2881	,2910	,2939	,2967	,2995	,3023	,3051	,3078	,3106	,3133
0,9	,3159	,3186	,3212	,3238	,3264	,3289	,3315	,3340	,3365	,3389
1,0	,3413	,3438	,3461	,3485	,3508	,3531	,3554	,3577	,3599	,3621
1,1	,3643	,3665	,3686	,3708	,3729	,3749	,3770	,3790	,3810	,3830
1,2	,3849	,3869	,3888	,3907	,3925	,3944	,3962	,3980	,3997	,4015
1,3	,4032	,4049	,4066	,4082	,4099	,4115	,4131	,4147	,4162	,4177
1,4	,4192	,4207	,4222	,4236	,4251	,4265	,4279	,4292	,4306	,4319
1,5	,4332	,4345	,4357	,4370	,4382	,4394	,4406	,4418	,4429	,4441
1,6	,4452	,4463	,4474	,4484	,4495	,4505	,4515	,4525	,4535	,4545
1,7	,4554	,4564	,4573	,4582	,4591	,4599	,4608	,4616	,4625	,4633
1,8	,4641	,4649	,4656	,4664	,4671	,4678	,4686	,4693	,4699	,4706
1,9	,4713	,4719	,4726	,4732	,4738	,4744	,4750	,4756	,4761	,4767
2,0	,4772	,4778	,4783	,4788	,4793	,4798	,4803	,4808	,4812	,4817
2,1	,4821	,4826	,4830	,4834	,4838	,4842	,4846	,4850	,4854	,4857
2,2	,4861	,4864	,4868	,4871	,4875	,4878	,4881	,4884	,4887	,4890
2,3	,4893	,4896	,4898	,4901	,4904	,4906	,4909	,4911	,4913	,4916
2,4	,4918	,4920	,4922	,4925	,4927	,4929	,4931	,4932	,4934	,4936
2,5	,4938	,4940	,4941	,4943	,4945	,4946	,4948	,4949	,4951	,4952
2,6	,4953	,4955	,4956	,4957	,4959	,4960	,4961	,4962	,4963	,4964
2,7	,4965	,4966	,4967	,4968	,4969	,4970	,4971	,4972	,4973	,4974
2,8	,4974	,4975	,4976	,4977	,4977	,4978	,4979	,4979	,4980	,4981
2,9	,4981	,4982	,4982	,4983	,4984	,4984	,4985	,4985	,4986	,4986
3,0	,49865	,4987	,4987	,4988	,4988	,4989	,4989	,4989	,4990	,4990
4,0	,49997									

APÊNDICE B DISTRIBUIÇÃO "T" DE STUDENT

g. l.	0,10	0,05	0,025	0,01	0,005
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
60	1,296	1,671	2,000	2,390	2,660
120	1,289	1,658	1,980	2,358	2,617
∞	1282	1,645	1,960	2,326	2,576

APÊNDICE C DISTRIBUIÇÃO χ^2

$P(t \geq t_t)$

g. l.	$P(x \geq \chi^2) = 0{,}05$
1	3,84
2	5,99
3	7,81
4	9,49
5	11,07
6	12,59
7	14,07
8	15,51
9	16,92
10	18,31
11	19,68
12	21,03
13	22,36
14	23,68
15	25,00
16	26,30
17	27,59
18	28,87
19	30,14
20	31,41
21	32,67
22	33,92
23	35,17
24	36,42
25	37,65
26	38,89
27	40,11
28	41,34
29	42,56
30	43,77
40	55,76
50	67,50
60	79,08
70	90,53
80	101,88
90	113,14
100	124,34