Universidade Federal do Piauí
Centro de Educação Aberta e a Distância

PROBABILIDADE E ESTATÍSTICA II

Juarez Rodrigues Martins

Ministério da Educação - MEC
Universidade Aberta do Brasil - UAB
Universidade Federal do Piauí - UFPI
Centro de Educação Aberta e a Distância - CEAD

Probilidade e Estatística II

Juarez Rodrigues Martins

UFPI
2011

M386p Martins, Juarez Rodrigues.
Probabilidade e estatística II / Juarez Rodrigues Martins.
– Teresina : CEAD/UFPI, 2011.
136 p.

ISBN: 978-85-7463-493-7

1. Estatística. 2. Probabilidade. 3. Estatística Matemática.
I. Título.

C.D.D. - 310

Apresentação

O presente material é destinado aos alunos aprendizes que participam do programa de Educação à Distância da Universidade Aberta do Piauí (UAPI), vinculada ao consórcio formado pela Universidade Federal do Piauí (UFPI), Universidade Estadual do Piauí (UESPI), Cento Federal de Ensino Tecnológico do Piauí (CEFET –PI), com apoio do Governo do Estado do Piauí, através da Secretaria de Educação.

Este material está organizado de sete unidades, contendo subunidades, estruturadas de modo sequencial, as quais discorrem sobre Amostragem, Distribuições Amostrais, Estimação de Parâmetros, Estatística Paramétrica – Teste de Hipóteses, Estatística Não–Paramétrica, Correlação e Regressão Linear e Análise de Variância ou Comparação de Várias Médias.

Na **unidade 1**, apresentamos o conceito de amostragem, dimensionamento da amostra, amostragem probabilística e seus tipos, amostragem não-probabilística e seus principais tipos, além de uma lista de exercícios no final da unidade.

Na **unidade 2**, apresentamos as distribuições amostrais, com uma introdução, distribuição normal padrão e o uso da tabela de distribuição normal padrão, distribuição amostral das médias, assim como seus principais teoremas, distribuição amostral das frequências relativas, distribuição amostral de variâncias, distribuição t de Student e a distribuição F de Snedecor, além de uma lista de exercícios no final da unidade.

Na **unidade 3**, apresentamos a estimação de parâmetros, que são intervalos de confiança, os tipos: intervalo de confiança para a média quando a variância é conhecida e quando a variância for

desconhecida, intervalo de confiança para a variância, intervalo de confiança para o desvio padrão da população e intervalo de confiança para a proporção populacional, além de uma lista de exercícios no final da unidade.

Na **unidade 4**, apresentamos a estatística paramétrica ou teste de hipóteses, como também, os principais conceitos: hipótese estatística, teste de hipóteses, tipos de hipóteses, tipos de erros; passos para a realização dos testes de hipóteses, teste de hipótese para a média populacional, teste de hipóteses para proporções, além de uma lista de exercícios no final da unidade.

Na **unidade 5**, apresentamos a estatística não-paramétrica, com uma introdução, teste qui-quadrado, teste qui-quadrado para independência ou associações, teste dos sinais, teste de Mann-Whitney e teste de Kruskal-Wallis, além de uma lista de exercícios no final da unidade.

Na **unidade 6**, apresentamos o estudo da correlação e regressão linear, com uma introdução, correlação linear simples: medida de correlação e os tipos de correlação, regressão linear simples e o poder explicativo do modelo, além de uma lista de exercícios no final da unidade.

Na **unidade 7**, apresentamos a análise de variância ou comparação de várias médias: com uma introdução, hipótese do modelo, classificação única ou experimento de um fator e estimadores da variância comum , fundamentos da análise da variância (ANOVA), quadro de análise da variância; classificação de dois critérios ou experimentos de dois fatores e estimadores de variância comum σ^2, além de uma lista de exercícios no final da unidade.

BONS ESTUDOS!!!

Sumário

UNIDADE 01

Amostragem

Resumindo

Nesta unidade, abordamos o estudo dos elementos que compõem uma amostragem extraída de uma população. O conceito de população é intuitivo. O estudo de todos os elementos da população possibilita o conhecimento preciso das variáveis que estão sendo pesquisados. É importante ressaltar que a representatividade da amostra depende do seu tamanho e de outras considerações de ordem metodológica. Na teoria da amostragem, são consideradas duas dimensões: dimensionamento da amostra e a composição da amostra.

1 AMOSTRAGEM

INTRODUÇÃO

A amostragem é o processo de retirada de amostras de uma **população***. É uma das etapas importantes na tomada de decisões nos diversos níveis gerenciais, pois o pesquisador procurará acercar-se de cuidados, visando à obtenção de uma **amostra*** significativa, ou seja, que de fato represente "o melhor possível" toda à população.

O objetivo principal desta unidade é apresentar alguns conceitos e definições necessárias para conduzir convenientemente uma operação de amostragem, visando principalmente à coleta de dados de uma forma mais econômica.

Se considerarmos uma população de clientes, podemos determinar o tempo médio em que o cliente fica, por exemplo, utilizando no dia o aparelho de telefone fixo (média populacional μ), que corresponde geralmente a um valor desconhecido, chamado de **parâmetro***. Como nós não vamos medir toda a população, podemos obter uma amostra que represente esta população e, estudando esta amostra, nós teremos condições de calcular a média amostral, que corresponde ao **estimador***. O resultado obtido (valor numérico) corresponderá à estimativa.

*População** – é o conjunto de elementos que apresentam uma ou mais características em comum.
*Amostra**– é um subconjunto da população.
*Parâmetro** – é um valor desconhecido associado a uma característica da população.
*Estimador** – é uma estatística usada para estimar um parâmetro. É a fórmula utilizada para o cálculo (média, proporção e outros).

Os problemas de amostragem podem ser mais ou menos complexos e sutis, dependendo das populações e das variáveis que se deseja estudar. Na indústria, onde amostras são frequentemente retiradas para efeito de controle de qualidade dos produtos e materiais, em geral, os problemas de amostragem são mais simples de resolver. Por outro lado, em pesquisas sociais, econômicas ou de opinião, a complexidade dos problemas de amostragem é normalmente bastante grande.

Em tais casos, deve ser tomado extremo cuidado quanto à caracterização da população e ao processo usado para selecionar a amostra, a fim de evitar que os elementos desta constituam um conjunto com características diferentes das da população.

Em resumo, a obtenção de soluções adequadas para o problema de amostragem exige, em geral, muito bom senso e experiência. Além disso, é muitas vezes conveniente que o trabalho do estatístico seja complementado pelo de um especialista no assunto em questão.

Na teoria da amostragem, são consideradas duas dimensões:

a) Dimensionamento da amostra

b) Composição da amostra.

DIMENSIONAMENTO DA AMOSTRA

Como proceder:

1º) Analise o questionário, ou roteiro da entrevista e escolha uma variável que julgue mais importante para o estudo. Se possível, escolha mais do que uma.
2º) Verifique o nível de mensuração da variável: se nominal, ordinal ou intervalar.
3º) Considere o tamanho da população: infinita ou finita:
4º) Se a variável escolhida for intervalar e a população considerada infinita, você poderá determinar o tamanho da amostra pela fórmula:

$$\left(\frac{Z \cdot \sigma}{d}\right)^2 \qquad 1.1$$

Onde: Z = abscissa da curva normal padrão, fixado um nível de confiança.
Se o nível for 95,5%, Z = 2
Se o nível for 95%, Z = 1,96

Se o nível for 99%, Z = 2,57
Geralmente, utiliza-se Z = 2.

σ = desvio padrão da população, expresso na unidade variável. Você poderá determiná-lo de pelo menos três maneiras:

- Especificações técnicas;
- Resgatar o valor de estudos semelhantes;
- Fazer conjeturas sobre possíveis valores.

d = erro amostral, expresso na unidade da variável. O erro amostral é a máxima diferença que o investigador admite suportar entre μ e $\bar{x}$, isto é: $|\mu - \bar{x}| < d$, onde μ é a verdadeira média populacional, que ele não conhece, e $\bar{x}$ será a média amostral a ser calculada a partir da amostra.

5º) Se a variável escolhida for intervalar e a população finita, têm-se:

$$n = \frac{Z^2 \cdot \sigma^2 \cdot N}{d^2 (N-1) + Z^2 \cdot \sigma^2} \qquad \textbf{1.2}$$

Onde Z = abscissa da normal padrão
σ = desvio padrão da população
N = tamanho da população
d = erro amostral.

6º) Se a variável escolhida for nominal ou ordinal, e a população considerada infinita, você poderá determinar o tamanho da amostra pela fórmula:

$$n = \frac{Z^2 \cdot \hat{p} \cdot \hat{q}}{d^2} \qquad \textbf{1.3}$$

Onde: Z = abscissa da normal padrão; $\hat{p}$ .estimativa da verdadeira proporção de um dos níveis da variável escolhida. Por exemplo, se a variável for porte da empresa, $\hat{p}$ poderá ser a estimativa da verdadeira proporção de grandes empresas do setor que está sendo estudado. Será expresso em decimais. Assim, se $\hat{p}$ = 30%, teremos:
$\hat{p}$ = 0,30.
$\hat{q}$ = 1 - $\hat{p}$
d = erro amostral, expresso em decimais. O erro amostral neste caso será a máxima diferença que o investigador admite suportar entre p e $\hat{q}$, isto é:

$|p - \hat{q}| \leq d$, em que *p é a verdadeira proporção, que ele não conhece, e* $\hat{p}$ será a proporção (frequência relativa) do evento a ser calculado a partir da amostra.

7º) Se a variável escolhida for nominal ou ordinal e a população finita, tem-se:

$$n = \frac{Z^2 \cdot \hat{p} \cdot \hat{q} \cdot N}{d^2 (N-1) + Z^2 \cdot \hat{p} \cdot \hat{q}}$$

Onde Z = abscissa da normal padrão;

n = tamanho da população;

$\hat{p}$ = estimativa da proporção;

$\hat{q} = 1 - p$

d = erro amostral.

Todas essas fórmulas são básicas para qualquer tipo de composição da amostra; todavia, existem fórmulas específicas segundo o critério de composição da amostra. Se o investigador escolher mais de uma variável, deve optar pelo maior n obtido.

Aplicações:

1) Suponha que a variável escolhida em um estudo seja o peso de certa peça e que a população seja infinita. Pelas especificações do produto, o desvio padrão (dispersão em torno da média) é de 10 kg. Logo se admitindo um nível de confiança de 95,5% e um erro amostral de 1,5 kg, determine o tamanho da amostra n.

Solução: a variável é intervalar e a população infinita, logo usaremos a fórmula (1.1) desta unidade.

$Z = 2$, $\sigma = 10$ *e* $d = 1,5$

$$N = \left(\frac{Z \cdot \sigma}{d}\right)^2 = \left(\frac{2 \cdot 10}{1,5}\right)^2 = 177,77 \cong 178.$$

Logo, o valor de n será de 178 elementos.

2) Admita os mesmos dados do exemplo anterior e que a população seja

finita de 600 peças, Qual é o tamanho da amostra n?

Solução: Aqui a variável é intervalar e a população finita, logo usaremos a fórmula (1.2) desta unidade.

Dados: $Z = 2,\ \sigma = 10\ ,\ N = 600\ e\ d = 1,5$

$$n = \frac{Z^2 \cdot \sigma^2 \cdot N}{d^2 (N-1) + Z^2 \cdot \sigma^2} = \frac{2^2 \cdot 10^2 \cdot 600}{1,5^2 (600-1) + 2^2 \cdot 10^2} = 137,31 \cong 138$$

Logo, o tamanho da amostra n será de 138 elementos.

3) Suponha que a variável escolhida em um estudo seja a proporção de eleitores favoráveis ao candidato X e que o investigador tenha elementos para suspeitar que essa porcentagem seja de 30%. Admita a população infinita e que se deseja um nível de confiança de 99% e um erro amostral de 2% (ou seja, que a diferença entre a verdadeira proporção de eleitores do candidato X e a estimativa a ser calculada na amostra seja no máximo de 2%). Determine o tamanho da amostra n.

Solução: A variável aqui é ordinal e a população é infinita, logo usaremos a fórmula (1.3).

Dados: Z = 2,57, $\hat{p}$ = 30% = 0,30, $\hat{q}$ =1 - 0.30 = 0,70 e d = 2% = 0,02, então n será:

$$n = \frac{(2,57)^2 \cdot (0,30) \cdot (0,70)}{(0,02)^2} = 3.467,57 \cong 3468.$$

Logo o tamanho da amostra será de 3468 eleitores.

4) Admita os mesmos dados do exemplo anterior e que a população de eleitores seja finita de 20.000 eleitores. Encontre o valor de n.

Solução: A variável escolhida é ordinal e a população finita, logo usaremos a fórmula (1.4).

Dados: Z = 2,57, $\hat{p}$ = 0,30, $\hat{q}$ = 0,70 e N = 20.000.

$$n = \frac{(2{,}57)^2.0{,}30).(0{,}70).(20.000)}{(0{,}0{,}2)^2.(20.000 - 1)+(2{,}57)^2.(0{,}30).(0{,}70)} = 2955{,}33$$

$$= 2956.$$

Logo, o tamanho da amostra será de 2956 eleitores.

AMOSTRAGEM PROBABILÍSTICA

Distinguimos dois tipos de amostragem: a probabilística e a não-probabilística. A amostragem será probabilística se todos os elementos da população tiverem probabilidade conhecida, e diferente de zero, de pertencer a amostra. Caso contrário, a amostra será não-probabilística.

Segundo essa definição, a amostragem probabilística implica um sorteio com regras bem determinadas, cuja realização só será possível se a população for finita e totalmente acessível.

Note que a decisão de abandonar os grupos maiores que 800 ou repetidos deve ser tomada antes de iniciado o processo, prevendo-se já tais ocorrências para evitar eventuais interferências do julgamento pessoal durante a retirada da amostra.

A utilização de uma amostragem probabilística é a melhor recomendação que se deve fazer no sentido de se garantir a representatividade da amostra, pois o acaso será o único responsável por eventuais discrepâncias entre população e amostra, o que é levado em consideração pelos métodos de análise da estatística indutiva.

Veremos a seguir algumas das principais técnicas de amostragem probabilística. Outras poderão também ser usadas, como combinação ou não das descritas.

Amostragem casual simples

Este tipo de amostragem, também chamada de simples ao acaso, aleatória, casual, simples, etc., é equivalente a um sorteio lotérico. Nela, todos os elementos da população têm igual probabilidade de pertencer à amostra, e todas as possíveis amostras têm também igual probabilidade de ocorrer.

Sendo N o número de elementos da população e n o número de elementos da amostra, cada elemento da população tem probabilidade n/N de

pertencer à amostra. A essa relação n/N denomina-se fração de amostragem. Por outro lado, sendo a amostragem feita sem reposição, o que suporemos em geral, existem $\binom{N}{m}$ possíveis amostras, todas igualmente prováveis.

Na prática, a amostragem simples ao acaso pode ser realizada numerando-se a população de 1 a N, sorteando-se, a seguir, por meio de um dispositivo aleatório qualquer, n elementos sorteados para a amostra.

Proporção do estrato h será igual ao número de elementos presente neste estrato (N_h) dividido pelo tamanho da população

$$(N) \rightarrow {}^{N_h}/_{N}.$$

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

$${}^{N_h}/_{N}.$$

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

Ex: Seja uma população de 800 elementos, da qual desejamos tirar uma amostra casual simples de 50 elementos. Consideramos a população numerada de 001 a 800, sendo os números tomados sempre com três algarismos. A seguir, sorteamos um dígito qualquer na nossa tabela em anexo (Tabela A1.1), a partir do qual iremos considerar os grupos de três algarismos subsequente formados, os quais irão indicar os elementos da amostra. Assim, se, a partir do ponto sorteado para início do processo, os dígitos observados forem 5 3 7 4 1 8 0 2 3 8 5 6 7 0 6 ..., os elementos sorteados para a amostra serão os de ordem 537, 418, 023, 706, etc. Evidentemente, o grupo 856 foi desprezado, pois não consta da população, como seria também abandonado um grupo que já tivesse aparecido (a não ser, é claro, que se desejasse amostragem com reposição). Prosseguindo o processo, obtêm-se os 50 elementos desejados.

Amostragem sistemática

Quando os elementos da população se apresentam ordenados e a retirada dos elementos da amostra é feita periodicamente, temos uma

amostragem sistemática. Assim, por exemplo, em uma linha de produção, podemos, a cada dez itens produzidos, retirar um para pertencer a uma amostra da produção diária.

Voltando ao exemplo anterior com N = 800, n = 50 e a população já ordenada, poderíamos adotar o seguinte procedimento; sortear um número de 1 a 16 (note-se que $^{800}/_{50}$ =16), o qual indicaria o primeiro elemento sorteado para a amostra e os demais elementos seriam periodicamente retirados de 16 em 16. Equivalentemente, poderíamos considerar os números de 1 a 800 dispostos sequencialmente em uma matriz com 50 linhas e 16 colunas, sorteando-se a seguir uma coluna, cujos números indicariam os elementos da amostra. Observamos que, nesse caso, cada elemento da população ainda teria probabilidade $^{50}/_{800}$ de pertencer à amostra.

A principal vantagem da amostragem sistemática está na grande facilidade na determinação dos elementos da amostra. O perigo em adotá-la está na possibilidade de existirem ciclos de variação da variável de interesse, especialmente se o período desses ciclos coincidir com o período de retirada dos elementos da amostra. Por outro lado, se a ordem dos elementos na população não tiver qualquer relacionamento com a variável de interesse, então a amostragem sistemática terá efeitos equivalentes à casual simples, podendo ser utilizada sem restrições.

Amostragem por conglomerados

Quando a população apresenta uma subdivisão em pequenos grupos, chamados conglomerados, é possível – e muitas vezes conveniente – fazer-se a amostragem por conglomerados, a qual consiste em sortear um número suficiente de conglomerados, cujos elementos constituirão a amostra. Ou seja, as unidades de amostragem, sobre as quais é feito o sorteio, passam a ser conglomerados e não mais os elementos individuais da população, Este tipo de amostragem é às vezes adotado por motivos de ordem prática e econômica, ou mesmo por razões de viabilidade.

Assim, por exemplo, num levantamento da população de uma cidade, podemos dispor do mapa indicando cada quarteirão e não dispor de uma relação atualizada dos seus moradores. Podemos, então, colher uma amostra dos quarteirões e fazer a contagem completa de todos os que residem naqueles quarteirões sorteados.

Amostragem estratificada

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

Podemos verificar que pesquisas eleitorais apresentam uma grande heterogeneidade em relação a intenção de votos, quando consideramos, por exemplo, a faixa salarial ou o nível de escolaridade. Então, se fizéssemos uma amostragem aleatória simples, poderíamos incluir na amostra uma quantidade de elementos de um grupo e, proporcionalmente, este grupo seria pequeno em relação à população. Desta forma, não teríamos uma amostra representativa da população a ser estudada. Então, podemos dividir a população em grupos (estratos) que são homogêneos para a característica que estamos avaliando, neste caso, a intenção de votos. Como estamos dividindo a população em estratos (grupos) que são homogêneos dentro de si, podemos, então, caracterizar a amostragem estratificada. Para efetuarmos a amostragem estratificada de forma proporcional, precisamos primeiramente definir a proporção do estrato em relação à população.

Exemplos em que uma amostragem estratificada parece ser recomendável é a estratificação de uma cidade em bairros, quando se deseja investigar alguma variável relacionada à renda familiar; a estratificação de uma população humana em homens e mulheres, ou por faixas etárias; a estratificação de uma população de estudantes conforme suas especializações, etc.

Amostragem múltipla

Em uma amostragem múltipla, a amostra é retirada em diversas etapas sucessivas. Dependendo dos resultados observados, etapas suplementares podem ser dispensadas. Esse tipo de amostragem é, muitas vezes, empregado na inspeção por amostragem, sendo particularmente importante a amostragem dupla. Sua finalidade é diminuir o número médio de itens inspecionados em longo prazo, baixando assim o custo da inspeção.

Um caso extremo de amostragem múltipla é a amostragem sequencial. A amostra vai sendo acrescida item por item, até se chegar a

uma conclusão no sentido de se aceitar ou rejeitar uma dada hipótese. Com a amostragem sequencial, pretende-se tornar mínimo o número médio de itens inspecionados em longo prazo.

AMOSTRAGEM NÃO-PROBABILÍSTICA

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

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

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

Amostragem a esmo

Imagine uma caixa de 1000 parafusos. A enumeração destes parafusos ficaria muito difícil, e a amostragem aleatória simples se torna inviável. Então, em situações deste tipo, supondo que a população de parafusos seja homogênea, escolhemos a esmo a quantidade relativa ao tamanho da amostra.Quanto mais homogênea for a população, mais podemos supor a equivalência com uma AAS.

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

Amostragem intencional

A amostragem intencional corresponde àquela em que o amostrador deliberadamente escolhe certos elementos para pertencer à amostra, por julgar tais elementos bem representativos da população. Um exemplo deste tipo de amostragem corresponde à situação em que se deseja saber a aceitação em relação a uma nova marca de uísque a ser inserida no mercado

de uma cidade. Somente entrarão para a amostra pessoas que façam uso da bebida e que tenham condições financeiras de comprar essa nova marca (classe social de maior poder aquisitivo).

Amostragem por cotas

Um dos métodos de amostragem mais comumente usados em levantamentos de mercado e em prévias eleitorais é o método de amostragem por quotas. Ele abrange três fases:

1) Classificação da população em termos de propriedades que se sabe, ou presume, serem relevantes para a característica a ser estudada;
2) Determinação da proporção da população para cada característica, com base na constituição conhecida, presumida ou estimada da população;
3) Fixação de quotas para cada observador ou entrevistador que terá a responsabilidade de selecionar interlocutores ou entrevistados, de modo que a amostra total observada ou entrevistada contenha a proporção de cada classe tal como determinada em (2).

Exemplificando: Admite-se que se deseja pesquisar o “trabalho das mulheres”. Provavelmente se terá interesse em considerar: a divisão cidade/campo, a habitação, o número de filhos, a idade dos filhos, a renda média, as faixas etárias...

A primeira tarefa é descobrir as proporções (porcentagens) dessas características na população. Imagine-se que haja 47% de homens e 53% de mulheres na população. Logo, uma amostra de 50 pessoas deverá ter 23 homens e 27 mulheres. Então o pesquisador receberá uma “quota” para entrevistar 27 mulheres. A consideração de várias categorias exigirá uma composição amostral que atenda ao n determinado e às proporções populacionais estipuladas.

EXERCÍCIO

1) Dada a seguinte população (rendas em R$ 1000)

29	6	34	12	15	31	34	20	8	30
8	15	24	22	35	31	25	26	20	10
30	4	16	21	14	21	16	18	20	12
31	20	12	18	12	25	26	13	10	5
13	19	30	17	25	29	25	28	32	15
10	21	18	7	16	14	11	22	21	36
32	17	15	13	8	12	23	25	13	21
5	12	32	21	10	30	30	10	14	17
34	22	30	48	19	12	8	7	15	20
2625	22	30	33	14	17	13	10	9	

Fonte: Fonseca, Jairo Simon da. 2006: pág. 184.

a) Calcule o tamanho da amostra para se estimar a média, sendo d = R$ 2000, σR$7000 e 1 - α = 95,5%
b) Retire uma amostra aleatória simples, considerando o tamanho amostral obtido em (a);
c) Agrupe os elementos da amostra em classes;
d) Calcule sua média;
e) Calcule o desvio padrão amostral;
f) Calcule a média da população e verifique se $|\mu - x| \leq d$.

2) Sendo p = q = 0,5 população infinita, d = 0,05 e 1 - α = 95,5%, determine o tamanho amostral.

3) Sendo p = q = 0,5, população de 200.000, d = 0,05 e 1 - α = 95,5%, determine o tamanho amostral. Compare com o resultado obtido no exercício 2.

4) Uma população se encontra dividida em três estratos, com tamanhos, respectivamente, de N_1 = 80, N_2 = 120 e N_3 = 60. Ao se realizar uma amostragem estratificada proporcional, 12 elementos da amostra foram retirados do primeiro estrato. Qual é o número total de elementos da amostra?

UNIDADE 02

Distribuições Amostrais

Resumindo

As distribuições amostrais, que são objeto de estudo desta unidade, são a base para aplicação das técnicas de inferências estatísticas apresentadas nas unidades seguintes. Nesta unidade, juntam-se os principais modelos de distribuições contínuas de probabilidade e as medidas que caracterizam uma amostra (que foram objetos de estudo anteriores), obtendo-se, assim as distribuições amostrais dos principais estimadores.

2 DISTRIBUIÇÕES AMOSTRAIS

INTRODUÇÃO

O capítulo que abordaremos agora é, de certa forma, uma ponte entre a Estatística Descritiva e a Estatística Indutiva. Sua apresentação é fundamental para a boa compreensão de como se constroem os métodos estatísticos de análise e interpretação dos dados, ou seja, os métodos da Estatística Indutiva. É aqui que o cálculo de probabilidades vai se apresentar como a ferramenta básica de que se vale a Estatística Indutiva para a elaboração de sua metodologia.

Portanto, torna-se necessário um estudo detalhado das distribuições amostrais, que são base para intervalos de confiança e testes de hipóteses. Para que você tenha condições de fazer afirmações sobre um determinado parâmetro populacional (ex: μ), baseadas na estimativa x, obtido a partir dos dados amostrais, é necessário conhecer a relação existente entre x e μ, isto é, o comportamento de x , quando se extraem todas as amostras possíveis da população, ou seja, sua distribuição amostral.

Distribuição amostral – Considere todas as possíveis amostras de tamanho n que podem ser extraídas de determinada população. Se para cada uma delas se calcular um valor do estimador, tem-se uma distribuição amostral desse estimador.

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

Veremos a seguir algumas distribuições amostrais que terão grande utilização nos capítulos seguintes. Outras serão mencionadas e comentadas em outros pontos do texto, sempre que necessário.

DISTRIBUIÇÃO NORMAL

É a mais importante distribuição de probabilidade, sendo aplicada em inúmeros fenômenos e utilizada para o desenvolvimento teórico da estatística. É também conhecida como distribuição de Gauss, Laplace ou Laplace-Gauss.

Vejamos uma aplicação desta distribuição:

Seja X uma variável aleatória contínua. X terá distribuição normal se:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)}, \quad -\infty < x < \infty \qquad \textbf{2.1}$$

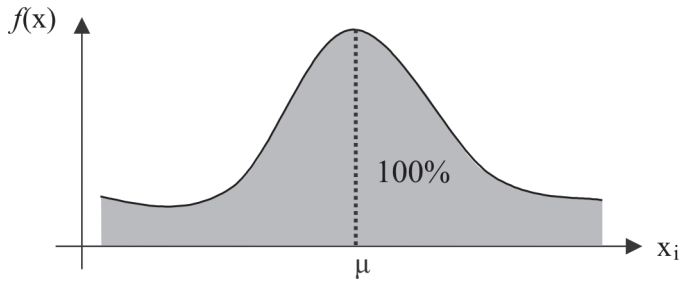
onde: μ = média da distribuição

σ = desvião padrão da distribuição

$\pi = 3{,}1416...$

e = 2,7...

Sendo seu gráfico:



Para o cálculo das probabilidades, surgem dois grandes problemas: primeiro, para a integração de $f(x)$, pois para o cálculo é necessário o desenvolvimento em séries; segundo, seria a elaboração de uma tabela de probabilidades, pois $f(x)$ depende de dois parâmetros, o que acarretaria um grande trabalho para tabelar essas probabilidades considerando-se as várias combinações de $\mu e \sigma^2$.

Esses problemas podem ser solucionados por meio de uma mudança de variável, obtendo-se, assim, a distribuição normal padronizada (Distribuição Normal Padrão) ou reduzida.

Distribuição normal padrão

Seja Z uma variável aleatória tal que:

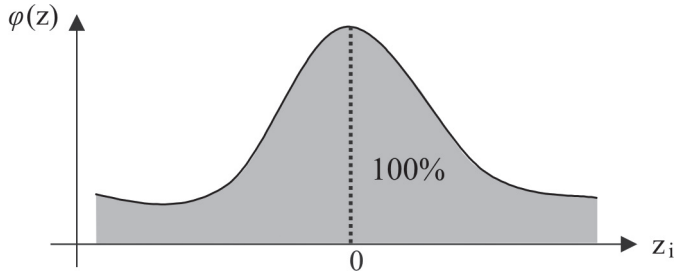
$$Z_i = \frac{x_i - \mu}{\sigma} \qquad \textbf{2.2}$$

Em X é a variável normal de média μ e variância σ^2.

Então a média de z será: $E[z] = 0$ e sua variância: $var[z] = 1$. Logo a função densidade será:

$$\varphi(z) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2} \qquad \infty < z < \infty$$

Sendo o gráfico de φ(z) igual a



Como a média de z é 0 e a variância 1, as probabilidades (áreas) são calculadas e tabeladas. Nos exemplos seguintes será explicado o uso da tabela da distribuição normal padrão.

Para se registrarem distribuições normais usa-se a seguinte notação:

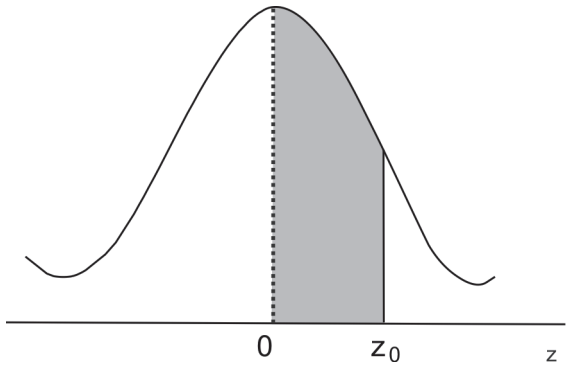
$X = N(\mu,\sigma^2)$ (lê-se "a variável X tem distribuição normal com média μ e variância σ^2. ")

$Z = N(0, 1)$ (lê-se "a variável aleatória Z tem distribuição normal com média 0 e variância 1." Ou, simplesmente distribuição normal padrão.)

Uso da tabela de distribuição normal padrão

Há vários tipos de tabelas que oferecem as áreas (probabilidades) sob a curva normal padrão. O tipo mais frequente é a tabela de faixa central.

A tabela de faixa central dá a área sob a curva normal padrão entre z = 0 e qualquer valor positivo de z. A simetria em torno de z = 0 permite obter a área entre quaisquer valores de z (positivos ou negativos).



A tabela oferece a área entre 0 e z_0 ou $P(0 \leq z \leq z_0)$.

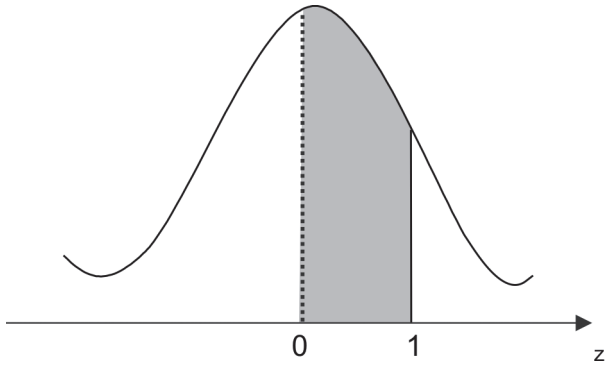
Exemplo: Desejam-se as probabilidades:

a) $P(0 \leq z \leq 1)$
b) $P(-2{,}55 < z < 1{,}2)$
c) $P(z \geq 1{,}93)$

Solução:

Tem-se:

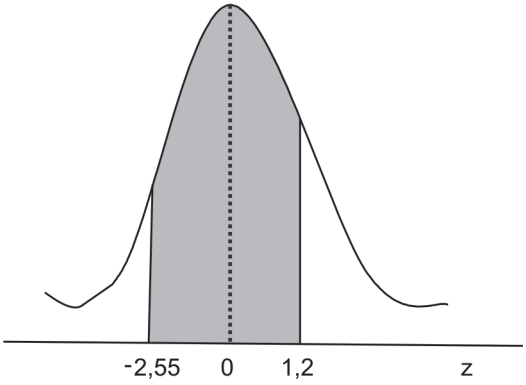
a)



Para se obter probabilidade, basta entrar com a abscissa 1,0 (na primeira coluna) e 0,00 (na primeira linha) da tabela. Assim:

$P(0 \leq z \leq 1) =$ **0,3413**

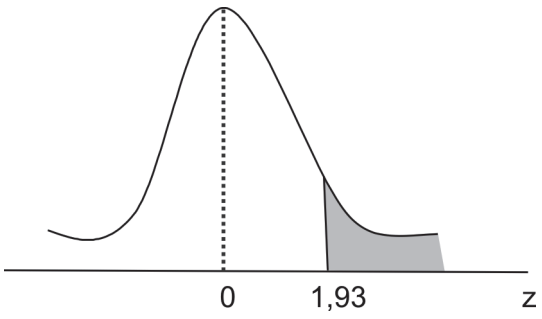
b)



Entra-se na tabela com o valor 1,2 na primeira coluna e 0 na primeira linha, obtendo 0,3849. A propriedade da simetria em relação a $z = 0$. Entra-se com 2,5 na primeira coluna e 0,05 na primeira linha, obtendo-se 0,4946. Portanto,

$$P(-2,55 < z < 1,2) = 0,3849 + 0,4946 = \mathbf{0,8795}$$

c)



Entra-se na tabela com 1,9 na primeira coluna e 0,03 na primeira linha obtendo 0,4732. Porém, essa é a área compreendida entre 0 e 1,93. Lembrando que a área embaixo da curva vale 1 e que a função é simétrica em relação à origem $z = 0$, tem-se:

$$P(z > 1,93) = 0,5000 - 0,4732 = 0,0268$$

Ex: As alturas dos alunos de determinada faculdade são normalmente

distribuídas com média 1,60 m e desvio padrão 0,30 m. Encontre a probabilidade de um aluno medir:

a) Entre 1,50 e 1,80 m;
b) Mais de 1,75 m;
c) Menos de 1,48 m;
d) Qual deve ser a média mínima para escolhermos 10% dos mais altos?

Solução: Sabe-se que = 1,60 e =0,30.

Faça X a variável altura dos alunos. Então:

a) $P(1,50 \leq x \leq 1,80) = P(z_1 \leq x \leq z_2) =$

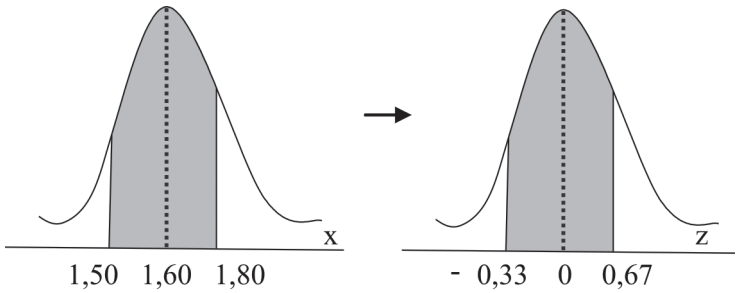
$= P(-0,33 \leq z \leq 0,67) =$

$= 0,1293 + 0,2486 =$

$= 0,3779 = 37,79\%$

Em que $z_1 = \frac{X - \mu}{\sigma} = \frac{1,50 - 1,60}{0,30} = \mathbf{-0,33}$ e

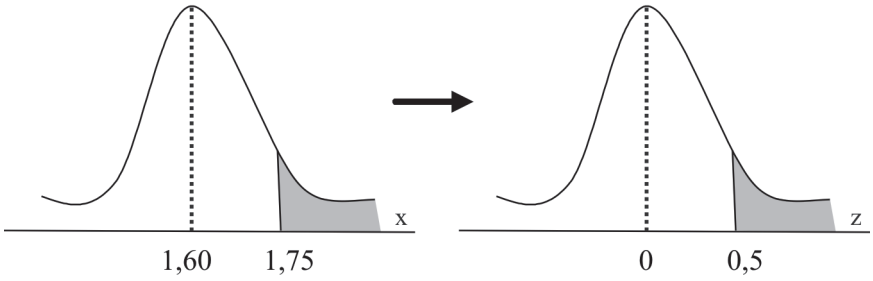
$$z_2 = \frac{X - \mu}{\sigma} = \frac{1,80 - 1,60}{0,30} = \mathbf{0,67}$$



b) $P(X > 1,75) = P(z > z_1) = P(z > 0,5) =$

$= 0,5000 - 0,1915 = \mathbf{0,3085}$

Em que $z_1 = \frac{1,75 - 1,60}{0,30} = \mathbf{0,5}$

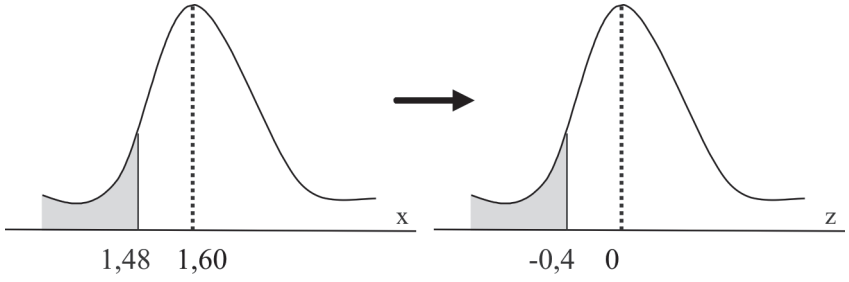


c) $P(X < 1,48) = P(z < z_1) = P(z < -0,4) =$

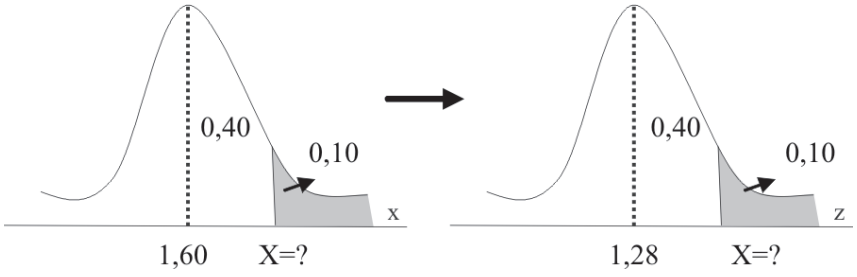
$= 0,5000 - 0,1554 =$ **0,3446**

Em que $z_1 = \frac{1,48 - 1,50}{0,30} \cong - 0,4$

Média aritmética ($\overline{x}$), ou média de um conjunto de n observações, x_1, x_2, x_3,..., x_n, é definida como sendo: $\overline{x} = \sum_{i=1}^{n} x_i/n$, onde: x_i = valor genérico da observação e *n* = nº de observações.



d) É o problema inverso dos itens anteriores, pois neste caso, tem-se a probabilidade e deseja-se a medida:



Para se encontrar o valor de z que deixa 0,10 à direita, deve-se entrar

na tabela com 0,40. Assim, descobrimos que z = 1,28. Logo,

$$Z = \frac{X - \mu}{\sigma} \rightarrow 1{,}28 = \frac{X - 1{,}60}{0{,}30} \rightarrow X = 1{,}98 \text{ m.}$$

Portanto X = 1,98 m deve ser a medida para se encontrar 10% dos mais altos.

DISTRIBUIÇÃO AMOSTRAL DAS MÉDIAS ($\overline{X}$)

Lembrando o conceito de distribuição amostral, visto anteriormente, busca-se descobrir qual é a distribuição da média aritmética $\overline{x}$.

Sabe-se que $\overline{x} = \Sigma\, x_i/n$ = (média aritmética) é um estimador da média populacional μ. O estimador x é uma variável aleatória, portanto, busca-se conhecer sua distribuição de probabilidade.

Teorema 1

A média da distribuição amostral das médias, denotada por μ(x), é igual à média populacional μ. Isto é:

$$E[\overline{x}] = \mu(\overline{x}) = \mu \qquad \textbf{2.3}$$

Assim, é provado que a média das médias amostrais é igual à média populacional.

Teorema 2

Se a população é infinita, ou se a amostragem é com reposição, então a variância da distribuição amostral das médias, denotada por é dada por:

$$E[(\overline{x} - \mu)^2] = \sigma^2(\overline{x}) = \frac{\sigma^2}{n} \qquad \textbf{2.4}$$

onde σ^2 é a variância da população. Isto é, pode-se afirmar que, para populações infinitas, ou amostragens com reposição, a variância da distribuição das médias é igual à variância da população dividida pelo tamanho da amostra.

Teorema 3

Se a população é finita, ou se a amostragem é sem reposição, então a variância da distribuição amostral das médias é dada por:

$$\sigma^2(\bar{x}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) \quad \textbf{2.5}$$

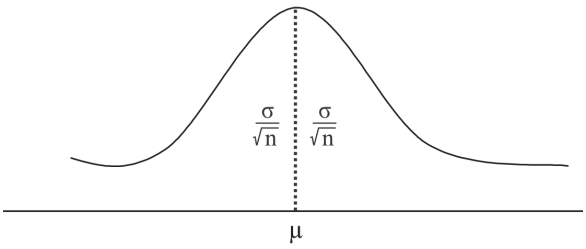
Sendo que: $\mu\ (\bar{x}) = \mu$.

Teorema 4

Se a população tem ou não distribuição normal com média μ e variância σ^2, então a distribuição das médias amostrais será normalmente distribuída com média μ e variância σ^2/n.

Esses quatro Teoremas provam que a média amostral (x) tem distribuição normal com média igual à média da população μ e variância dada por $\sigma^2/2$ para populações infinitas, e $\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$ para populações finitas.

Graficamente:



(fig 2.1)

ou ainda:

$$\bar{x} = N(\mu;\frac{\sigma^2}{n}) \quad \text{ou} \quad \bar{x} = N(\mu;\frac{\sigma^2}{n}(\frac{N-n}{N-1}))$$

Com distribuições padronizadas dadas por:

$$Z_i = \frac{x_i - \mu}{\frac{\sigma}{\sqrt{n}}} \quad ou \quad Z_i = \frac{x_i - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\left(\frac{N-n}{N-1}\right)}} \quad \textbf{2.6}$$

Distribuição amostral das frequências relativas

Seja X uma população infinita, e $\boldsymbol{p}$ a probabilidade (ou proporção) e certo evento de X. Logo $1 - p = q$ será a probabilidade de o evento não ocorrer.

Seja (x_1, x_2, ... , x_n) uma amostra aleatória de n elementos dessa população e x o número de sucessos na amostra. É fácil identificar como uma variável aleatória com distribuição Binomial (nº de sucessos na amostra), de média $\boldsymbol{np}$ e variância $\boldsymbol{npq}$.

Então, a distribuição amostral da frequência relativa

$\hat{p} = f = {}^{x}/_{n}$ será dada por:

$$E[f] = E[\frac{x}{n}] = \frac{np}{n} = p \qquad \textbf{2.7}$$

$$Var[f] = Var[\frac{x}{n}] = \frac{npq}{n^2} = \frac{pq}{n} \qquad \textbf{2.8}$$

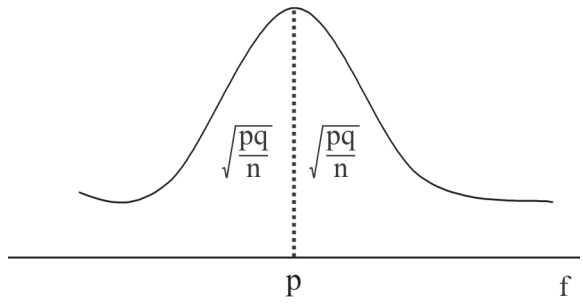
Para $n \geq 30$ a distribuição amostral de f será normal:

$$F = N(p, \frac{pq}{n})$$

Assim a sua distribuição padronizada será:

$$Z_i = \frac{f_i - p}{\sqrt{\frac{pq}{n}}} \qquad \textbf{2.9}$$

Ou graficamente:



(fig 2.2)

Distribuição amostral de variâncias

Sabe-se que a variância da população é designada por σ^2. Seja S^2 (variância amostral) o estimador de σ^2.

Variância (S^2) – é definida como sendo o quociente entre a soma dos quadrados dos desvios e o número de elementos.

$$Z_i = \frac{\Sigma_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}$$

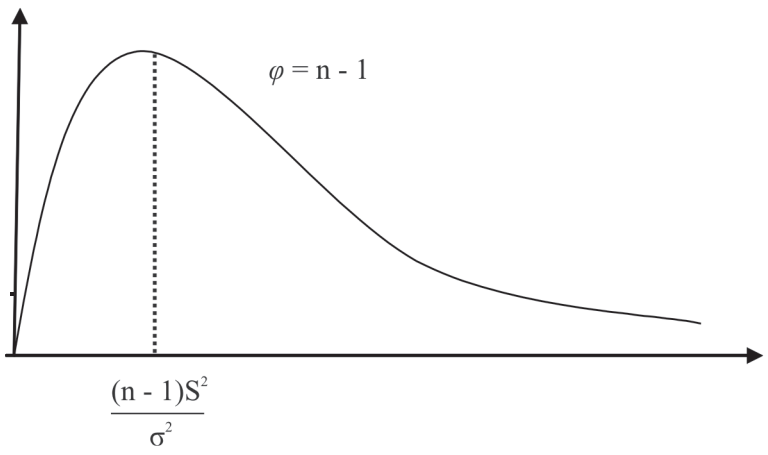
Se desejar saber qual é a distribuição de S^2, pode-se demonstrar que:

$$E[S^2] = \sigma^2 \text{ e } var[S^2] = \frac{2\sigma^4}{n - 1} \quad \textbf{2.10}$$

E que S^2 tem distribuição qui-quadrado com (n – 1) graus de liberdade. Ou seja:

$$\frac{(n - 1)S^2}{\sigma^2} \; x^2_{n-1} \quad \textbf{2.11}$$

Lembre-se que (n – 1) e σ^2 são constantes. Graficamente, a relação entre S^2 e σ^2 é dada por uma distribuição qui-quadrado.



(fig 2.3)

DISTRIBUIÇÃO *t* DE STUDENT

Suponhamos que, a partir de uma amostra de n valores retirados de uma população normal de média μ e desvio padrão $\sigma/\sqrt{n}$, fosse definida a estatística:

$$Z_i = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad \textbf{2.12}$$

Desvio Padrão (S) – é igual à raiz quadrada da variância: $S=\sqrt{S^2}$

Como a distribuição amostral de x seria precisamente normal, com média μ e desvio padrão $\sigma/\sqrt{n}$, segue-se que essa estatística teria simplesmente distribuição normal padrão, o que justifica o uso do símbolo z em (2.12).

Entretanto, se usarmos em (2.12) o desvio padrão da amostra, obteremos uma estatística cuja distribuição não mais é normal. De fato, conforme mostrou Student, a estatística:

2.13

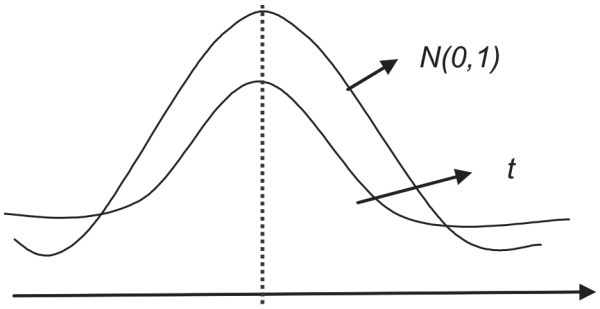
$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

distribui-se simetricamente, com média 0, porém não normalmente. É claro que, para amostras grandes, S deve ser próximo de σ, e as correspondentes distribuições t devem estar próximas da normal padrão. Vemos, pois, que existe uma família de distribuições t cuja forma tende à distribuição normal padrão quando n cresce. Note-se que a estatística definida em (2.13) tem n – 1 graus de liberdade, o que justifica sua denotação por t_{n-1}.

A fig. 2.4 procura ilustrar comparativamente uma distribuição t e a distribuição normal padrão z. Vemos que uma distribuição t genérica é mais alongada que a normal padrão.

Por outro lado, a tabela t de Student fornece valores de t em função de diversos valores do número de graus de liberdade G.L. e de probabilidades notáveis, correspondentes à cauda à direita na respectiva distribuição. Assim, por exemplo, entrando-se na tabela com a probabilidade p = 0,025 e G.L = 50, lemos o valor t_{50} = 2,009. Isso significa, dada a simetria das distribuições t, que $P(t_{50} > 2{,}009) = P(t_{50} < -2{,}009) = 0{,}025$. Note-se que esse valor de t_{50} é

já muito próximo do correspondente valor t_{∞} = z = 1,960.



(fig 2.4)

É importante notar que a expressão (2.13) pode ser escrita

$$t_{n-1} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \cdot \frac{\sigma}{S} = z\,\frac{\sigma}{S} \qquad \textbf{2.14}$$

Ou ainda:

$$t_{n-1} = z\sqrt{\frac{n-1}{x^2_{n-1}}} \qquad \textbf{2.15}$$

Ou ainda:

$$t_{n-1} = z\sqrt{\frac{G.L}{x^2}} \qquad \textbf{2.16}$$

x = Média amostral;
μ = Média populacional;
S = Desvio padrão amostral;
σ = Desvio padrão populacional;
n = Tamanho da amostra;
N = Tamanho da população.

Essa expressão nos mostra o relacionamento existente entre as distribuições t de Student e X^2.

DISTRIBUIÇÃO *F* DE SNEDECOR

Trata-se de um modelo de distribuição contínua também útil para inferências estatísticas.

A distribuição F é a razão entre duas variáveis aleatórias independentes com distribuição qui-quadrado. Assim, a distribuição f com "p" graus de liberdade no numerador e "q" graus de liberdade no denominador é expressa por:

2.17

$$F(p,q) = \frac{\frac{x^2 p}{p}}{\frac{x^2 q}{q}} \quad \frac{x^2 p}{x^2 q} \quad \frac{q}{p}$$

A distribuição F possui dois parâmetros: grau de liberdade do numerador e grau de liberdade do denominador, que são denominados, comumente, por e respectivamente.

Quanto à média, é dada por:

$$\mu = \frac{\varphi_2 - \varphi_1}{\varphi_2} \quad \text{com} \quad \varphi_2 > \varphi_1 \qquad \textbf{2.18}$$

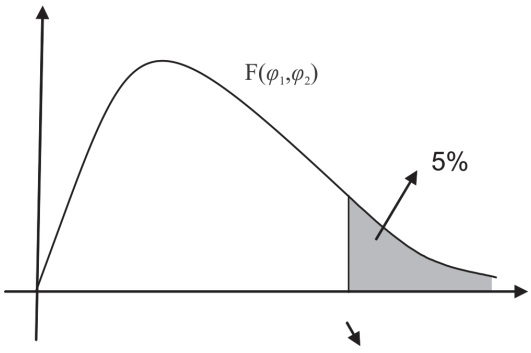
A variância é expressa por:

$$\sigma^2 = \frac{2\varphi^2_2(\varphi_1 + \varphi_2 - 2)}{\varphi_1(\varphi_1 - 4)(\varphi_2 - 2)} \quad \text{com } \varphi_2 > 4 \qquad \textbf{2.19}$$

E a média:

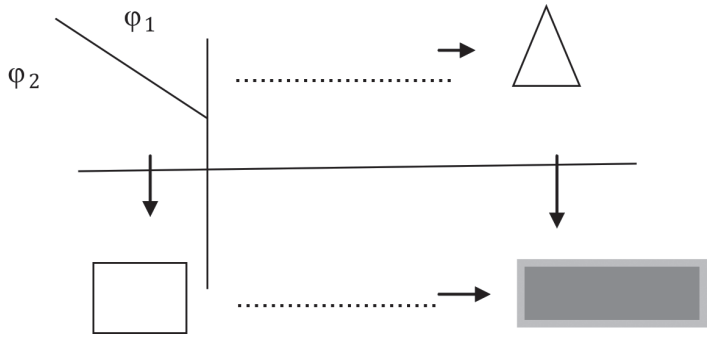
$$M_0 = \left(\frac{\varphi_1 - 2}{\varphi_1}\right)\left(\frac{\varphi_1}{\varphi_1 + 2}\right) \text{ com } \varphi_1 > 2 \qquad \textbf{2.20}$$

A distribuição F está tabelada (A2.3). Esta tabela nos dá as abscissas que deixam 5% na cauda à direita, dados os parâmetros φ_1 e φ_2 . Assim:



encontra-se na Tabela 2 (pág. 154).

Na tabela, procede-se assim:

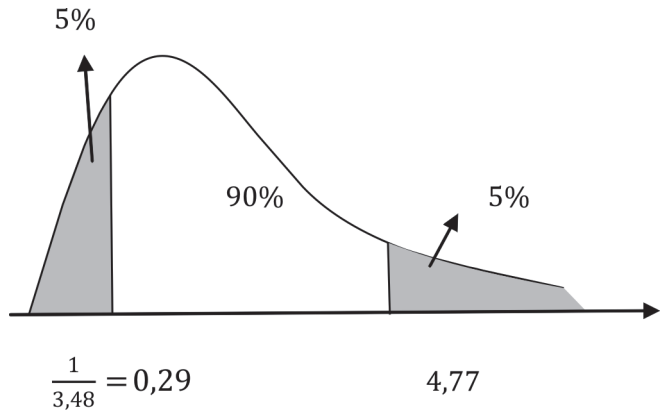


Valor da abscissa

Para se encontrar o valor da abscissa $F_{1-\alpha}(\varphi_1, \varphi_2)$, utiliza-se a seguinte fórmula:

$$F_{1-\alpha}(\varphi_1,\varphi_2) = \frac{1}{F_{\alpha}(\varphi_1, \varphi_2)} \qquad \textbf{2.21}$$

Exemplo: Sendo φ_1 = 9, φ_2 = 5 e α = 5%, determine as abscissas superior e inferior



EXERCÍCIO

1) Faça Z uma variável com distribuição normal padronizada e encontre (use a tabela):

a) P(0 ≤ z ≤ 1,44)

b) P(-0,85 < z < 0)

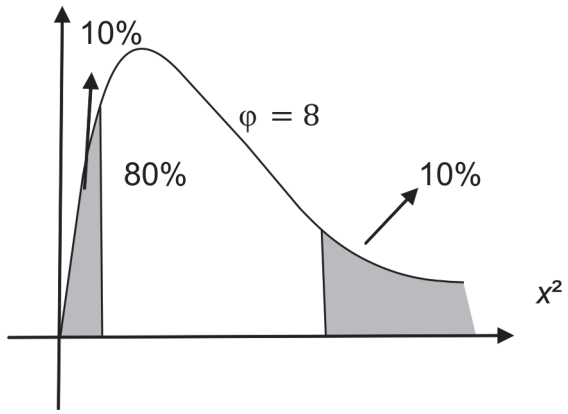
c) P(-1,48 < z < 2,05)

2. Os pesos de 600 estudantes são normalmente distribuídos com média 65,3 kg e desvio padrão 5,5 kg. Encontre o número de alunos que pesam:

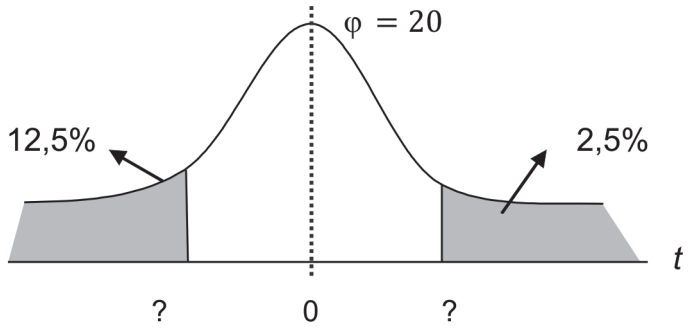
a) Entre 60 e 70 kg;

b) Mais que 63,2 kg.

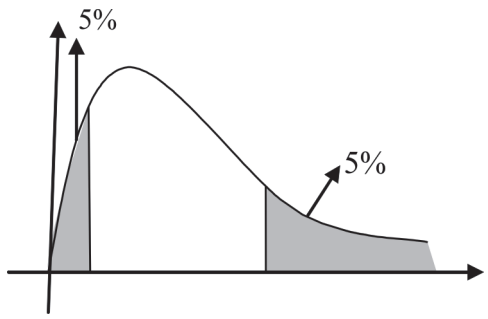
3) Determine os valores de x^2_{sup} e x^2_{inf}



4) Consulte a tabela para descobrir os valores das abscissas.



5) Admita uma distribuição F com $\varphi_1 = 8$ e $\varphi_2 = 10$. Determine a média, a variância, o desvio padrão e as abscissas para:



UNIDADE 03

Estimação de Parâmetros

Resumindo

Nesta unidade abordamos uma técnica para se fazer inferência estatística. A partir de um intervalo de confiança, construído com elementos amostrais, pode-se inferir sobre um parâmetro populacional. A construção de intervalos de confiança se fundamenta nas distribuições amostrais vistas na unidade anterior. Esta técnica se diferencia da estimação por ponto, onde se calcula um único valor (estimativa) para o parâmetro populacional. No caso do intervalo de confiança, busca-se um 'segmento', ou intervalo que contenha o parâmetro desconhecido.

3 ESTIMAÇÃO DE PARÂMETROS

INTRODUÇÃO

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

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

A estimativa pode ser por ponto ou intervalar. A estimativa pontual infere sobre a população, considerando apenas o valor da estimativa. Essas estimativas por ponto não nos dão uma ideia sobre confiança e as margens de erro que deveriam ser aplicadas ao resultado de uma pesquisa, por exemplo. Já a estimativa por intervalos nos fornece uma informação mais precisa em relação ao parâmetro, esta é a melhor forma de estimar o parâmetro populacional. Então, para você estimar parâmetros populacionais por meios de dados, é necessário o conhecimento da distribuição amostral da

estatística que está sento usada como estimador (visto anteriormente). Por isso, estudaremos a seguir a estimação intervalar.

INTERVALO DE CONFIANÇA

Trata-se de uma técnica para se fazer inferência estatística. Ou seja, a partir de um intervalo de confiança, construído com os elementos amostrais, podemos inferir sobre um parâmetro populacional.

Ao intervalo que, com probabilidade conhecida, deverá conter o valor real do parâmetro chamaremos de intervalo de confiança para esse parâmetro. À probabilidade, que designaremos por 1 - α, de que um intervalo de confiança contenha o valor do parâmetro chamaremos de nível ou grau de confiança do respectivo intervalo. Vemos que será a probabilidade de erro na estimação por intervalo, isto é, a probabilidade de errarmos ao afirmar que o valor do parâmetro está contido no intervalo de confiança.

Veremos a seguir como construir intervalos de confiança para os parâmetros usuais.

Intervalo de Confiança para a Média (μ) quando a Variância (σ^2) é Conhecida.

Como se sabe, o estimador de μ é $\bar{x}$. Também é conhecida a distribuição de probabilidade de $\bar{x}$.

Devemos construir um intervalo em torno de de forma tal que esse intervalo contenha o valor do parâmetro com confiança 1 – α ou (1 – α) . 100 = %, esse intervalo, sendo simétrico em probabilidade, será também geometricamente simétrico em relação a $\bar{x}$, devido à simetria da distribuição amostral. Observa-se na tabela da distribuição normal padrão o valor das abscissas que deixam $\alpha/2$ em cada uma das caudas. Com os valores de $\bar{x}$ (média amostral), σ = desvio padrão populacional, que neste caso é conhecido e n (tamanho da amostra), temos:

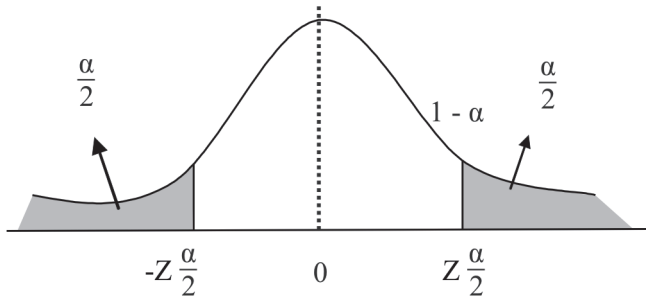
$\bar{x} = N(\mu ; \frac{\sigma^2}{n})$ para populações infinitas.

$\bar{x} = N[\mu ; \frac{\sigma^2}{n} (\frac{N - n}{N - 1})]$ para populações infinitas.

Para o caso de populações infinitas, a variável padronizada de $\bar{x}$ é:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad 3.1$$

Fixando-se um nível de confiança: 1 - α, temos:



Ou seja: P(-z ≤ z ≤ z) = 1 - α.

Substituindo o valor de z , temos:

$$P\left(-z\,\frac{\alpha}{2} \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z\,\frac{\alpha}{2}\right) = 1 - \alpha$$

Resolvendo-se as duas inequações para μ, temos o intervalo de confiança para a média populacional (μ) quando a variância (σ^2) é conhecida:

$$P\left(\bar{x} - z\,\frac{\alpha}{2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x}\, z\frac{\alpha}{2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \qquad 3.2$$

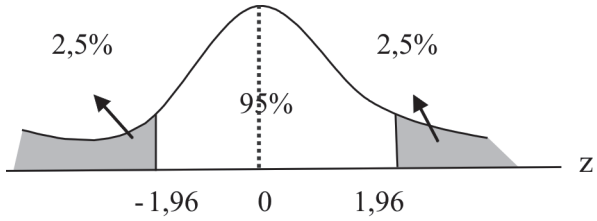
Aplicação:

A duração da vida média de uma peça de equipamento é tal que σ = 5 horas. Foram amostradas 100 dessas peças obtendo-se a média de 500 horas. Deseja-se construir um intervalo de confiança para a verdadeira duração média da peça com um nível de 95%.

Solução: Temos

σ = 5; n = 100; x = 500 e (1 – α).100 = 95%

Observe o gráfico da distribuição normal padrão:



Lembrando que para descobrir a abscissa 1,96, entrou-se na tabela de distribuição normal com 0,475 = (0,5 – 0,025) = 47,5%, sabendo que a tabela é de faixa central.

Substituindo os dados na fórmula, temos:

$$P(500 - 1{,}96 \cdot \frac{5}{\sqrt{100}} \leq \mu \leq 500 + 1{,}96 \cdot \frac{5}{\sqrt{100}}) = 95\%.$$

Efetuando-se os cálculos:

P(499,02 ≤ μ ≤ 500,98) = 95%

Interpretação:

O intervalo [499,02 ; 500,98] contém a verdadeira duração média da peça com 95% de confiança.

Isto significa que se forem construídos intervalos dessa mesma maneira, para um grande número de amostras, em 955 dos casos tais intervalos incluiriam a média populacional μ.

No caso de populações finitas, usa-se a seguinte fórmula:

$$P(\bar{x} - z\frac{\alpha}{2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \leqslant \mu \leqslant \bar{x}\, z\frac{\alpha}{2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}) = 1 - \alpha \qquad \textbf{3.3}$$

Intervalo de Confiança para a Média (μ) quando a Variância (σ^2) é Desconhecida

Vejamos agora como proceder para construir o intervalo de confiança para a média da população quando o desvio padrão populacional é também

desconhecido, o que, em geral, ocorre nos problemas práticos.

Ora, se desconhecemos σ, devemos estimar seu valor com base na amostra disponível. Devemos adotar como estimativa o desvio padrão da amostra, já visto anteriormente.

O processo para se construir o intervalo de confiança é semelhante àquele estudado no item anterior. Como não se conhece σ, porém, é preciso substituí-lo por S (desvio padrão amostral) que, contrariamente a σ é uma variável aleatória. Portanto, o quociente entre duas variáveis aleatórias, $\bar{x}$ e S, pois:

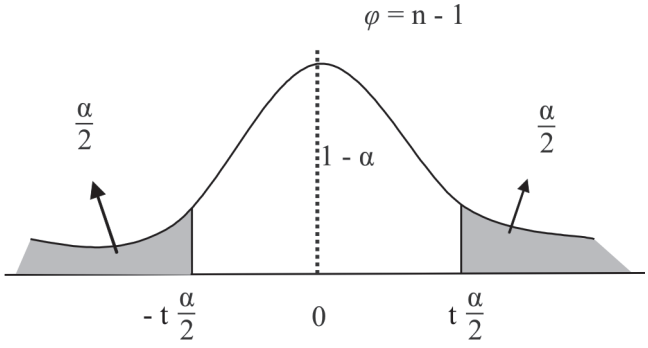
$$\frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \qquad \textbf{3.4}$$

Pode-se demonstrar que:

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \qquad \textbf{3.5}$$

tem distribuição "t" de Student com (n - 1) graus de liberdade.

Fixando-se um nível de confiança: 1 - α tem–se:



Ou seja: $P(-t\frac{\alpha}{2} \leq t \leq t\frac{\alpha}{2}) = 1 - \alpha$

Substituindo-se o valor de "t" e resolvendo-se as inequações para μ, obtém-se o intervalo para a média quando a variância (σ^2) é desconhecida.

$$P(\bar{x} - t\frac{\alpha}{2}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + t\frac{\alpha}{2}\frac{S}{\sqrt{n}}) = 1 - \alpha \qquad \textbf{3.6}$$

Onde a variável "t" possui (n – 1) graus de liberdade.

Aplicação:

A amostra: 9, 8, 12, 7, 9, 6, 11, 6, 10, 9 foi extraída de uma população normal. Construir um intervalo de confiança para a média ao nível de 95%.

Solução:

Cálculo da média e do desvio padrão:

$$\bar{x} = \frac{9+8+12+7+9+6+11+6+10+9}{10} \quad \frac{87}{10} = 8,7$$

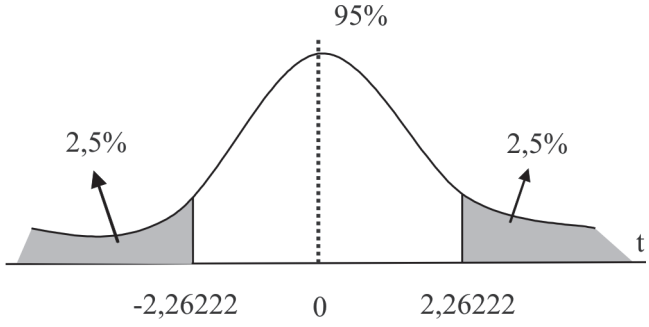
$$\boxed{\bar{x} = 8,7}$$

$$S = \sqrt{\frac{1}{9}\left(793 - \frac{(87)^2}{10}\right)} = \sqrt{\frac{1}{9}\left(793 - \frac{7569}{10}\right)}$$

$$S = \sqrt{\frac{1}{9}(793 - 7596,9)} = \sqrt{\frac{1}{9}(36,1)} = \sqrt{4,011111} \cong 2,0$$

$$\boxed{S = 2}$$

Como: $1 - \alpha = 95\%$ e g.l. = $\varphi = n - 1 = 10 - 1 = 9$, temos:



As abscissas na tabela t de Student ($t_{0,025}$), logo:

$$P(8,7 - 2,2622.\frac{2}{\sqrt{10}} \leq \mu \leq 8,7 + 2,262\ .\quad) = 95\%$$

$$P(7,27 \leq \mu \leq 10,13) = 95\%$$

Interpretação do resultado:

O intervalo [7,27; 10,13] contém a verdadeira média com 95% de confiança.

Para o caso de populações finitas, usa-se a seguinte fórmula:

$$P(\bar{x} - t\,\frac{\alpha}{2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \leqslant \mu \leqslant \bar{x}\,t\,\frac{\alpha}{2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}) = 1 - \alpha \qquad \textbf{3.7}$$

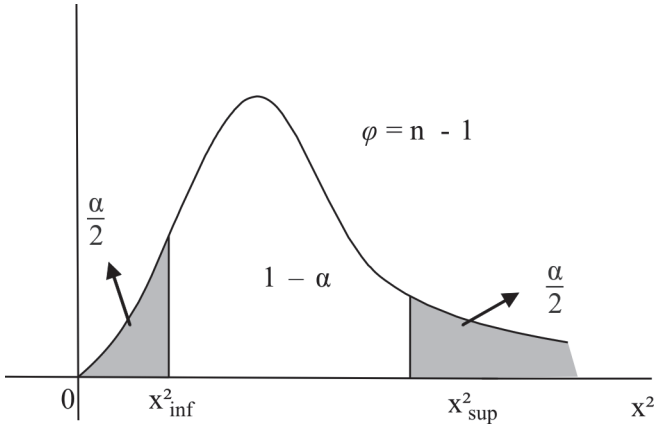
Intervalo de Confiança para a Variância

Consideremos agora o problema da construção do intervalo de confiança ao nível 1 - α para a variância σ^2 da população. O conhecimento das distribuições x², visto anteriormente, será fundamental para esse propósito.

O estimador de σ^2 é S². Demonstra-se que $(n-1).S^2/\sigma^2$ tem distribuição qui-quadrado com (n – 1) graus de liberdade. Ou seja:

$$X_{n-1} = \frac{(n-1)S^2}{\sigma^2} \qquad \textbf{3.8}$$

Então, o intervalo será:



Substituindo-se o valor de x², e isolando-se σ^2, obtém-se:

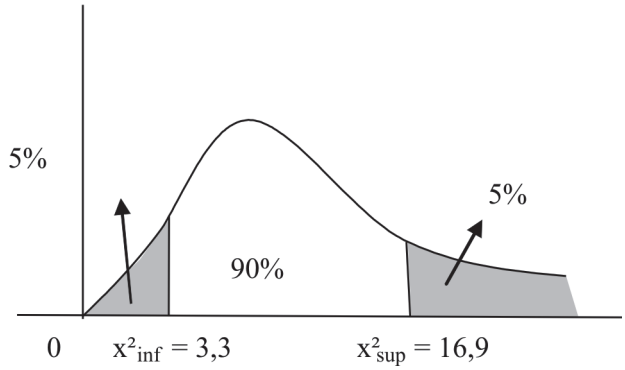
$$P\left(\frac{(n-1)S^2}{x^2sup} \leqslant \sigma^2 \leqslant \frac{(n-1)S^2}{x^2inf}\right) = 1 - \alpha \qquad \textbf{3.9}$$

Com a distribuição qui-quadrado de parâmetro: φ = (n - 1)

Aplicação:

Admitindo-se n = 10, S^2 = 4, $1 - \alpha$ = 90% e φ = 10 - 1 = 9

Consultando a tabela da distribuição qui-quadrado, temos:



Logo:

$$P\left(\frac{1,69}{9,4} \leqslant \sigma^2 \leqslant \frac{9,4}{3,33}\right) = 90\%$$

$$P(2,13 \leq \sigma^2 \leq 10,81) = 90\%$$

Interpretação:

O intervalo [2,13; 10,81] contém a verdadeira variância com 90% de confiança.

Intervalo de Confiança para o Desvio Padrão da População

Vimos anteriormente que o desvio padrão da amostra, S, não é um estimador justo do desvio padrão da população, σ e que, por essa razão, deveríamos introduzir uma correção, especialmente no caso de amostras pequenas. Entretanto, se desejarmos um intervalo de confiança ao nível de 1 - α, para o parâmetro σ não será necessário investigar a distribuição por amostragem do correto estimador de σ pois decorre imediatamente do resultado obtido no item anterior que, com probabilidade 1 - α, temos:

$$\sqrt{\frac{(n-1)S^2}{x^2 sup}} \leqslant \sqrt{\sigma^2} \leqslant \sqrt{\frac{(n-1)S^2}{x^2 inf}}$$

Logo:

$$P\left(S.\sqrt{\frac{(n-1)S^2}{x^2 sup}} \leqslant \sigma^2 \leqslant S.\sqrt{\frac{(n-1)S^2}{x^2 inf}} \right) = 1 - \alpha \quad \textbf{3.10}$$

Com distribuição qui-quadrado de parâmetro: $\varphi = (n - 1)$.

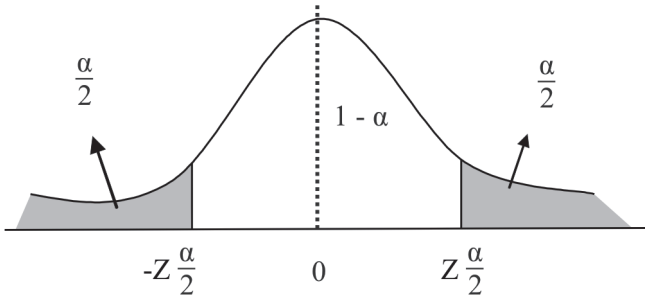
Intervalo de Confiança para Proporção Populacional ou Probabilidade (P)

Você deve considerar que, geralmente, a proporção de sucessos em uma população é desconhecida. Então, o que fazemos? Calculamos uma estimativa da proporção de sucessos na população a partir de uma amostra retirada desta.

Para construirmos o intervalo de confiança para p desconhecido, determinamos f na amostra e consideramos $\sigma_f \cong \sqrt{\frac{p.q}{n}}$. Assim, para o caso de populações infinitas, a variável padronizada de f é dada por:

$$Z = \frac{f - p}{\sqrt{\frac{p.q}{n}}} \quad \textbf{3.11}$$

Fixando-se um nível de confiança 1- α temos:



Ou seja: $P(-z\frac{\alpha}{2} \leq z \leq z\frac{\alpha}{2}) = 1 - \alpha$

Substituindo-se o valor de z:

$$P(-z\frac{\alpha}{2} \leq \frac{f - p}{\sqrt{\frac{p.q}{n}}} \leq z\frac{\alpha}{2}) = 1 - \alpha$$

Isolando-se p do denominador, encontraremos:

$$P(f - z\frac{\alpha}{2}.\sqrt{\frac{p.q}{n}} \leq p \leq f + z\frac{\alpha}{2}.\sqrt{\frac{p.q}{n}}) = 1 - \alpha$$

Para amostras grandes (n > 30), pode-se substituir p e $q = 1 - p$ do radicando por f e $(1 - f)$. Assim, o IC para a proporção será:

$$P(f - z\frac{\alpha}{2}.\sqrt{\frac{f.(1-f)}{n}} \leq p \leq f + z\frac{\alpha}{2}.\sqrt{\frac{f.(1-f)}{n}}) = 1 - \alpha \qquad \textbf{3.12}$$

Para o caso de populações finitas, o IC será:

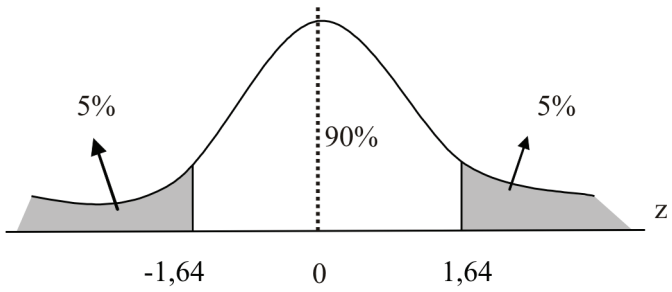
$$P(f - z\frac{\alpha}{2}.\sqrt{\frac{f.(1-f)}{n}.(\frac{N-n}{N-1})} \; p \leq f + z\frac{\alpha}{2}.\sqrt{\frac{f.(1-f)}{n}.(\frac{N-n}{N-1})}) \; 1 - \alpha \qquad \textbf{3.13}$$

Aplicação:

Ao serem examinadas 500 peças de uma grande produção, foram encontradas 260 defeituosas. No nível de 90%, construa um IC para a verdadeira proporção de peças defeituosas.

Solução:

Temos: $n = 500,\ x = 260,\ 1 - \alpha = 90\%$ e $f = \frac{x}{n} = \frac{260}{500} = \mathbf{0,52}$.



Então, o intervalo de confiança será:

$$P(0,52 - 1,64.\sqrt{\frac{0,52.(1-0,52)}{500}} \leq p \leq 0,52 + 1,64.\sqrt{\frac{0,52.(1-0,52)}{500}}) = 90\%$$

ou $P(0,488 \leq p \leq 0,552) = 90\%$; ou ainda:

$$P(48,8\% \leq p \leq 55,2\%) = 90\%.$$

Interpretação:

O intervalo $[48,8\%; 55,2\%]$ contém a verdadeira porcentagem (ou proporção) de peças defeituosas.

EXERCÍCIO

1) Foram retiradas 25 peças da produção diária de uma máquina, encontrou-se para uma medida uma média de 5,2 mm. Sabendo-se que as medidas têm distribuição normal com desvio padrão populacional 1,2 mm, construir intervalo de confiança para a média aos níveis de 90% e 95%.

2) De uma distribuição normal com σ^2 = 1,96, obteve-se a seguinte amostra: 25,2; 26,0; 26,4; 27,1; 28,2; 28,4. Determinar o intervalo de confiança para a média da população, sendo α = 0,05 e α = 0,10.

3) Supondo que uma amostra de n = 10 fornecesse s^2 = 2,25. Quais os limites de confiança a 80% para a verdadeira variância?

4) Qual é o intervalo de confiança que conterá com 90% a verdadeira variância de uma população normal que resultou Σx_i = 700,8 e Σx^2_i = 23.436,80 de uma amostra de 30 elementos?

5) Uma centena de componentes foi ensaiada e 93 deles funcionaram mais de 500 horas. Determinar um intervalo de confiança da proporção de 95% para a proporção.

6) Uma amostra aleatória de 400 domicílios mostra-nos que 25% deles são casas de aluguel. Qual é o intervalo de confiança da proporção de casas de aluguel? α = 2%.

UNIDADE 04

Estatística Paramétrica - Teste de Hipóteses

Resumindo

Nesta unidade abordamos uma técnica muito importante para se fazer inferência estatística. Ou seja, a partir de um teste de hipóteses, realizado com os dados amostrais, pode-se tirar conclusões sobre a população. Na unidade anterior estudamos os ICs, com os quais busca-se 'cercar' o parâmetro populacional e pelos elementos amostrais faz-se um teste que indicará a aceitação ou rejeição da hipótese formulada.

4 ESTATÍSTICA PARAMÉTRICA - TESTE DE HIPÓTESES

INTRODUÇÃO

Agora abordaremos o segundo tipo de problema de estatística indutiva: o dos testes de hipóteses referentes à população. Nesta unidade, trataremos dos testes chamados paramétricos, pois se referem a hipóteses sobre parâmetros populacionais.

No caso dos Intervalos de Confiança, busca-se "cercar" o parâmetro populacional desconhecido. Já no teste de hipótese, formula-se uma hipótese quanto ao valor do parâmetro populacional, e pelos elementos amostrais faz-se um teste que indicará a aceitação ou rejeição da hipótese formulada.

PRINCIPAIS CONCEITOS

Veremos em seguida, os principais conceitos que usaremos no estudo sobre teste de hipóteses: Hipótese Estatística, Testes de Hipóteses, Tipos de Hipóteses e Tipos de Erro.

HIPÓTESE ESTATÍSTICA

É uma suposição quanto ao valor de um parâmetro populacional, ou quanto à natureza da distribuição de probabilidade de uma variável populacional.

Aqui serão apresentados os testes referentes aos parâmetros da população.

Ex: a) A altura média da população brasileira é 1,65 m, ou seja: H: μ = 1,65m .

a) A proporção de piauienses com a doença y é 40%, ou seja: H:p = 0,40.

TESTE DE HIPÓTESE

É uma regra de decisão para aceitar ou rejeitar uma hipótese estatística com base nos elementos amostrais.

TIPOS DE HIPÓTESE

Hipótese nula é a hipótese estatística a ser testada e será designada por H_0 e por H_1 a hipótese alternativa. A hipótese nula é expressa por uma igualdade, enquanto a hipótese alternativa por uma desigualdade.

Exemplos:

a) H_0: μ = 1,65m. Originará um teste
H_1: $\mu \neq$ 1,65m. Bicaudal

b) H_0: μ = 1,65m. Originará um teste
H_1: $\mu >$ 1,65m. Unicaudal à direita

c) H_0: μ = 1,65m. Originará um teste
H_1: $\mu <$ 1,65m. Unicaudal à esquerda

TIPOS DE ERROS

Ao testar uma hipótese estatística, podemos cometer dois tipos de erro. Pode-se rejeitar uma hipótese, quando ela é, de fato, verdadeira, ou aceitar uma hipótese quando ela é, de fato, falsa. A rejeição de uma hipótese verdadeira é chamada “erro tipo I”. A aceitação de uma hipótese falsa constitui um “erro tipo II”.

As probabilidades desses dois tipos de erros são designadas, respectivamente, pela probabilidade do erro tipo I é denominada “nível de significância” do teste. Resumindo, temos:

Erro tipo I: rejeitar H_0, sendo H_0 verdadeira;

Erro tipo II: aceitar H_0, sendo H_0 falsa.

A faixa de valores da variável de teste que leva à rejeição de H_0 é

denominada de **região crítica** (R.C.) do teste. A faixa restante constitui a **região de aceitação** (R.A.)

Passos para realização dos testes de hipóteses (significância)

O procedimento para realização dos testes de significância é resumido nos seguintes passos:

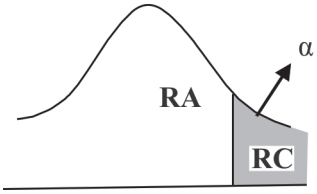
1. Enunciar as hipóteses H_0 e H_1. Primeiramente, vamos estabelecer as hipóteses nula e alternativa.
2. Definir o nível de significância (α) e identificar a variável do teste; o nível de significância de um teste é dado pela probabilidade de se cometer erro tipo I. Com o valor desta probabilidade fixada, você pode determinar o chamado valor crítico, que separa a região de rejeição da hipótese H_0 da região de aceitação da hipótese H_0.
3. Usando as tabelas estatísticas e considerando α e a variável do teste, determinar as RC (região crítica) e RA (região de aceitação) para H_0;

Na figura abaixo, as áreas hachuradas correspondem à significância do teste, ou seja, à probabilidade de se cometer o erro tipo I (rejeitar H_0 quando ela é verdadeira). Esta probabilidade é representada por α e o complementer dela, que é chamado de nível de confiança,por 1 - α.

Unilateral à direita

$H_0 : \mu = \mu_0$

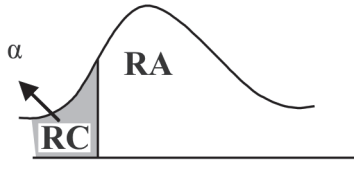
$H_1: \mu > \mu_0$



Unilateral à esquerda

$H_0: \mu = \mu_0$

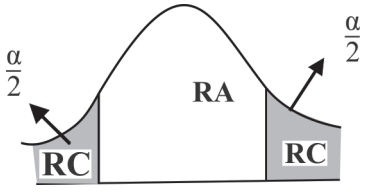
$H_1: \mu < \mu_0$



Bilateral

$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$



4. Com os elementos amostrais, calcular o valor da variável do teste;

Dependendo da variável, temos:

$$Z_{cal} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad \textbf{4.1} \qquad \text{ou} \qquad t_{cal} = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \quad \textbf{4.2}$$

Onde: $\bar{x}$ é a média amostral

μ é a média populacional

S é o desvio padrão amostral

σ é o desvio padrão populacional

n é o tamanho da amostra.

5. A conclusão pela aceitação ou rejeição de H_0 pela comparação do valor obtido no passo anterior com RA e RC.

Para tomar a decisão, você deve observar a estimativa do teste estatístico já calculado no item anterior, para rejeitar ou não a hipótese H_0.

Se o valor da estatística calculado (z_{cal} ou t_{cal}) estiver na região crítica (de rejeição), rejeita-se H_0, caso contrário, aceita-se H_0, ou seja, se estiver na região de aceitação, aceita-se H_0.

Quando trabalhamos com amostras grandes, ou seja, n > 30, a distribuição de z e t de student apresentam comportamentos próximos e valores da estatística próximos também.

Teste de hipótese para a média populacional

Quando você retira uma amostra de uma população e calcula a média desta amostra, é possível verificar se a afirmação sobre a média populacional é verdadeira. Para isso, basta verificar se a estatística do teste estará na região de aceitação ou de rejeição da hipótese H_0.

Aqui você verá duas situações diferentes:

1ª.) Se o desvio padrão da população é conhecido ou a amostra é considerada grande (n > 30), a distribuição amostral a ser utilizada será a normal ou z e a estatística – teste que você utilizará – será:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad \textbf{4.3}$$

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

2ª.) Agora, se você não conhece o desvio padrão populacional e a amostra for pequena (n < 30), então, a distribuição amostral a ser utilizada será a t de student, e a estatística – teste será:

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \qquad \textbf{4.4}$$

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

Aplicações:

a) O desvio padrão de uma população é conhecido e igual a 22 unidades. Se uma amostra de 100 elementos, retirada dessa população, forneceu $\bar{x}$ = 115,8, podemos afirmar que a média dessa população é inferior a 120 unidades, ao nível de 5% de significância? Qual a significância do resultado obtido, face as hipótese testada?

Solução:

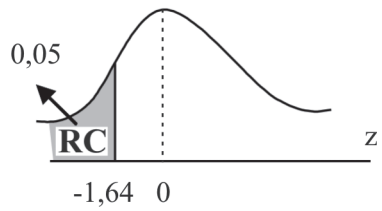
1º) Testando as hipóteses

H_0: $\mu = 120$

H_1: $\mu < 120$

2º) A variável do teste será z e α = 0,05

3º) z tabelado



4º) Cálculo de z_{cal}

$$Z_{cal} = \frac{115,8 - 120}{\frac{22}{\sqrt{100}}} = \frac{-4,2}{2,2} = \textbf{-1,91}$$

5º) Conclusão

Como z_{cal} z_{tab}, ou seja, -1,91 ϵ RC, então se rejeita H_0. Portanto, podemos afirmar que, nesse nível de significância, que a média da população é inferior a 120 unidades.

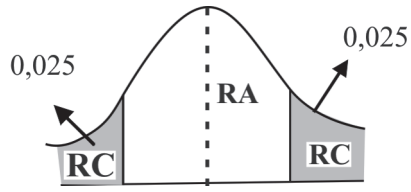
b) Os dois registros dos últimos anos da UAPI atestam para os calouros admitidos uma nota média 115 (teste vocacional). Para testar a hipótese de que a média de uma prova é a mesma, tirou-se, ao acaso, uma amostra de 20 notas, obtendo-se média 118 e desvio padrão 20. Admitir que α = 5% = 0,05, para efetuar o teste.

Solução:

1º) H_0: $\mu = 115$
H_1: $\mu \neq 115$

2º) α = 0,05 e a variável do teste é "t" com
φ = 20 - 1 = **19** g.l.

3º)



4º) $t_{cal} = \frac{118 - 115}{\frac{20}{\sqrt{20}}} = 0,67$

5º) Como t_{cal} ϵ RA, ou seja, -2,093 < 0,67 < 2,093 , não se pode rejeitar H_0: μ = 115cm com esse nível de significância, ou seja, se aceita que a verdadeira média populacional é igual a 115.

Teste de hipótese para variâncias

As mesmas ideias apresentadas no caso do teste resultaram em uma média que pode ser utilizada para se realizar testes envolvendo a variância da população. Assim, iremos testar as hipóteses

H_0: $\sigma^2 = \sigma_0^2$
H_1: $\sigma_2 \neq \sigma_0^2$
$\sigma^2 < \sigma_0^2$
$\sigma^2 > \sigma_0^2$

A variável de teste deverá ser a variância da amostra, definida anteriormente, pois é o estimador justo da variância populacional, conforme já visto. Sendo normal a distribuição da população, a quantidade $(n - 1).S^2 / \sigma^2$ tem distribuição x^2 com (n – 1) graus de liberdade. Logo, supondo verdadeira a hipótese H_0, ou seja, admitindo que a variância da população seja igual ao valor testado S_0^2, podemos escrever:

$$x^2_{cal} = \frac{(n - 1).S^2}{\sigma^2_0} \qquad \textbf{4.5}$$

Onde: n: tamanho da amostra; S^2: variância amostral e σ^2: valor da hipótese nula.

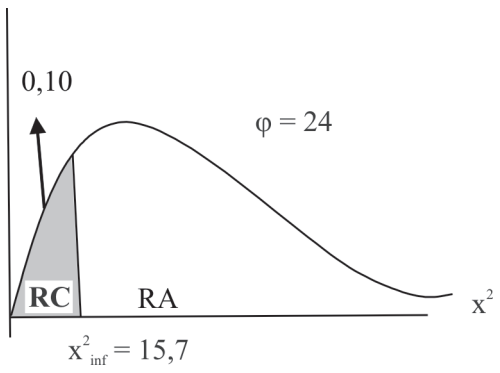
Aplicação:

Para testar a hipótese de que a variância de uma população é 25, tirou-se uma amostra aleatória de 25 elementos, obtendo-se $S^2 = 18,3$. Admitindo-se $\alpha = 0,10$, efetuar o teste de significância unicaudal à esquerda.

1º) H_0: $\sigma^2 = 25$
H_1: $\sigma^2 < 25$

2º) $\alpha = 0,10$; variável x^2 com $\varphi = 25 - 1 = 24$ g.l.

3º)



4º) $x^2_{cal} = \frac{(25 - 1).18,3}{25} = \mathbf{17,56}$

5º) Como x^2_{cal} 15,7, ou seja, x^2_{cal} ϵ RA, não se pode rejeitar H_0: $\sigma^2 = 25$ ao nível de significância de 10%, ou seja, se aceita a hipótese de que a variância da população é igual a 25.

Teste de hipótese para proporções

Já sabemos que, ao realizarmos induções sobre uma proporção populacional p, devemos nos basear na proporção observada na amostra f. Sabemos, também, que podemos aproximar a distribuição amostral de f pela distribuição normal de média p e desvio padrão $\sqrt{P(1 - P)}/p$. Isso nos permite realizar facilmente testes envolvendo proporções populacionais, de forma análoga ao que foi visto para os testes de uma média. Para a realização desse teste, temos:

1º) H_0: $p = p_0$
H_1: $p \neq p_0$
$p > p_0$
$p < p_0$

2º) Fixar α. Escolher a variável normal padrão z.

3º) Com o auxílio da tabela de distribuição normal padrão, determina-se RA e RC.

4º) Calcular o valor da variável:

$$Z_{cal} = \frac{f - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} \qquad \mathbf{4.6}$$

onde: f = frequência relativa do evento na amostra,
p_0 = valor da hipótese nula,
n = tamanho da amostra.

5º) Conclusões:
Se z_{cal} ϵ RA, então aceita-se H_0, ou seja, não se pode rejeitar H_0.

Se $z_{cal} \in RC$, então rejeita-se H_0, ou seja, não se pode aceitar H_0.

Aplicação:

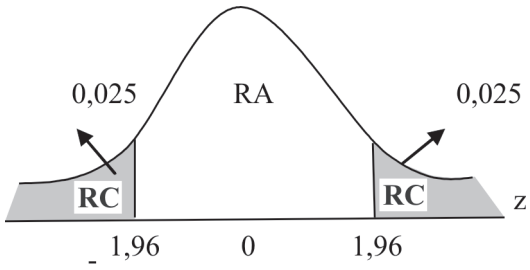
As condições de mortalidade de uma região são tais que a proporção de nascidos que sobrevivem até 60 anos é de 0,6. Testando essa hipótese ao nível de 5%, em 1000 nascimentos amostrados aleatoriamente, verificou-se 530 sobreviventes até 60 anos.

Solução:

1º) $H_0 : p = 0,6$
$H_1 : p \neq 0,6$

2º) $\alpha = 0,05$ e a variável escolhida, é a normal z.

3º) Determinação da RA e RC



4º) z_{cal}

$$Z_{cal} = \frac{f - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{0,53 - 0,6}{\sqrt{\frac{0,6(1 - 0,6)}{1000}}} = -4,42$$

5º) Como $z_{cal} \in RC$, rejeita-se H_0, concluindo-se ao nível de 5%, que a verdadeira proporção de sobreviventes é diferente de 0,6, ou seja, $p \neq 0,6$.

EXERCICIO

1) Uma amostra de 25 elementos resultou em média de 13,5 com desvio padrão 4,4. Efetue o teste ao nível de 0,05 para a hipótese que $\mu = 16$ contra $\mu \neq 16$.

2) Retirando-se uma amostra de 15 parafusos, obtiveram-se as seguintes medidas para seus diâmetros:

10 10 10 11 11 12 12 12 12
13 13 14 14 14 15

Teste H0 : $\mu = 12,5$ contra $\mu \neq 12,5$; $\mu > 12,5$. Adotando $\alpha = 0,05$.

3) Um laboratório fez 8 determinações da quantidade de impurezas em porções de certo composto. Os valores eram: 12,4; 12,6; 12,0; 12,0 12,1; 12,3 12,5 e 12,7 mg.
a) Estime a variância de impurezas entre porções.
b) Teste a hipótese de que a variância é 1, ao nível de $\alpha = 0,05$ Contra H_1: $\sigma^2 < 1$.

4) Suponha $X = N(\mu,\sigma^2)$ em que μ e σ^2 são desconhecidos. Uma amostra de tamanho 15 forneceu $\Sigma x_i = 8,7$ e $\Sigma x^2_i = 27,3$. Teste a hipótese de que a variância da população é 4. Adote $\alpha = 1\%$. (Teste uni e bicaudal).

5) Uma amostra de 500 eleitores selecionados ao acaso dá 52% ao Partido Democrático. Poderia esta amostra ter sido retirada de uma população que tivesse 50% de eleitores democratas? Admita $\alpha = 0,05$.

6) Uma pesquisa revelou que das 500 donas de casa consultadas, 300 preferiram o detergente A. Teste a hipótese ao nível de 0,04 para H_0: $p = 0,5$, contra H_1: $p \neq 0,5$.

UNIDADE 05

Estatística não Paramétrica

Resumindo

Nesta unidade, vemos as técnicas da Estatística Não Paramétrica que são, particularmente, adaptáveis aos dados das ciências do comportamento. A aplicação dessas técnicas não exige suposições quanto à distribuição da população da qual se tenha retirado amostras para análise. Contrariamente ao que ocorre na Estatística Paramétrica, onde as variáveis são, na maioria das vezes, intervalares, como foi visto nas unidades 3 e 4, os testes não paramétricos são extremamente interessantes para análises de dados qualitativos.

5 ESTATÍSTICA NÃO PARAMÉTRICA

INTRODUÇÃO

As técnicas da estatística não-paramétrica são, particularmente, adaptáveis aos dados das ciências do comportamento. A aplicação dessas técnicas não exige suposições quanto à distribuição da população da qual se tenha retirado amostras para análises. Podem ser aplicadas a dados que se disponham simplesmente em ordem, ou mesmo para estudo de variáveis nominais. Contrariamente ao que acontece na estatística paramétrica, onde as variáveis são, na maioria, intervalares, como visto nas unidades 3 e 4. Os testes não - paramétricos são extremamente interessantes para análises de dados qualitativos.

Os testes da estatística não–paramétrica exigem poucos cálculos e são aplicáveis para análise de pequenas amostras (n < 30).

Como o próprio nome indica, a estatística não - paramétrica independe dos parâmetros populacionais $(\mu; \sigma^2; \sigma; p \ldots)$ e de suas respectivas estimativas $(\bar{x}\ ; S^2; S, f \ldots)$

TESTE QUI-QUADRADO

O mais popular teste não-paramétrico é o teste qui-quadrado, ou teste de adequação do ajustamento.

Seja ε um experimento aleatório. Sejam E_1, E_2, ..., E_k, "K" eventos associados a ε. Admita que o experimento seja realizado "n" vezes.

Sejam F_{01}, F_{02}, F_{0k} as frequências observadas dos "k" eventos.

Sejam Fe_1, Fe_2, Fe_k as frequências esperadas, ou frequências teóricas dos "k" eventos.

Deseja-se realizar um teste estatístico para verificar se há adequação de ajustamento entre as frequências observadas e as frequências esperadas. Isto é, se as discrepâncias ($Fo_i - Fe_i$), i = 1, 2, ... k, são devidas ao acaso, ou se de fato existe diferença significativa entre as frequências.

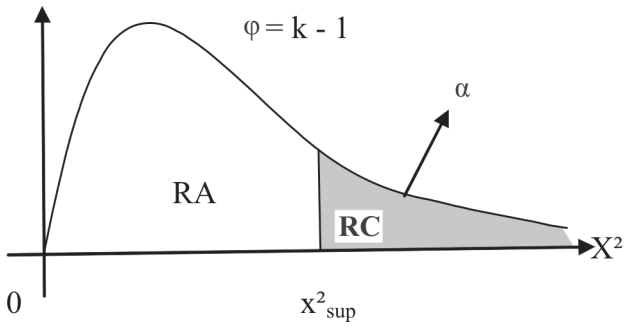
Passos para efetuar o teste:

1. Enunciar as hipóteses H_0 e H_1.

H_0 afirmará não haver discrepância entre as frequências observadas e esperadas, enquanto H_1 afirmará que as frequências observadas e esperadas são discrepantes.

2. Fixar α. Escolher a variável qui-quadrado com φ = k - 1. Lembrando que k é igual ao número de eventos.

3. Com o auxílio da tabela x², determinam-se RA e RC.



4. Cálculo do valor da variável:

$$X^2_{cal} = \sum_{i=1}^{k} \frac{(Fe_i - Fe_i)^2}{Fe_i} = \frac{(Fe_1 - Fe_1)^2}{Fe_1} + \ldots + \frac{(Fe_k - Fe_k)^2}{Fe_k} \qquad \textbf{5.1}$$

5. Conclusão

Se $x^2_{cal} < x^2_{sup}$, não se pode rejeitar H_0, ou seja, as frequências observadas e esperadas não são discrepantes.

Se $x^2_{cal} > x^2_{sup}$, rejeita-se H_0, concluindo-se com o risco que a discrepância entre as frequências observadas é esperada. Ou seja, não há adequação do ajustamento.

Aplicações:

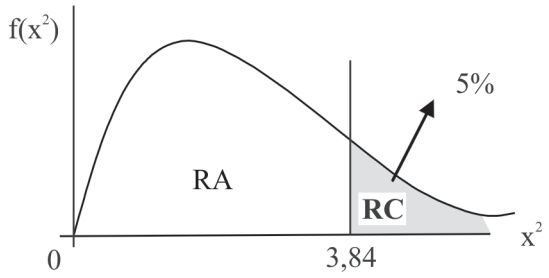
1. Em 100 lances de uma moeda, observaram-se 65 coroas e 35 caras. Testar a hipótese de a moeda ser honesta, adotando-se α = 5%.

Solução:

1. H_0: A moeda é honesta
   H_1: A moeda não é honesta

2. α = 5%. Escolhe-se uma x^2_1, pois
   k = 2 – 1 = 1.

3. Determinação da RC e RA.



4. Cálculo do valor da variável

Eventos	Cara	Coroa
F. observadas	35	65
F. esperadas	50	50

$$X^2_{cal} = \sum_{i=1}^{2} \frac{(F0_i - Fe_i)^2}{Fe_i} = \frac{(35 - 50)^2}{50} + \frac{(65 - 50)^2}{50} = 9 \qquad \textbf{5.2}$$

5. Conclusão:

Como $x^2_{cal} \geq 3{,}84$, ou seja, $x^2_{cal} \in$ RC, rejeita-se H_0, concluindo-se, com risco de 5%, que a moeda não é honesta.

TESTE QUI-QUADRADO PARA INDEPENDÊNCIA OU ASSOCIAÇÃO

O teste qui-quadrado de associação é aconselhável quando o tamanho da amostra é razoavelmente grande e deve ser aplicado com maior cuidado se existem frequências esperadas (Fesp) menores do que 5. Nestes casos, a solução é juntar classes adjacentes, evitando-se que $F_{esp} < 5$.

O teste qui-quadrado tem uma aplicação importante quando se quer estudar a associação ou independência, entre duas variáveis. A representação das frequências observadas é dada por uma tabela de dupla entrada ou tabela de contingência.

O cálculo das frequências esperadas fundamenta-se na definição de variáveis aleatórias independentes, conforme visto em variáveis aleatórias. Isto é: diz-se que X e Y são independentes se a distribuição conjunta de (X,Y) é igual ao produto das distribuições marginais de X e de Y. Isto é:

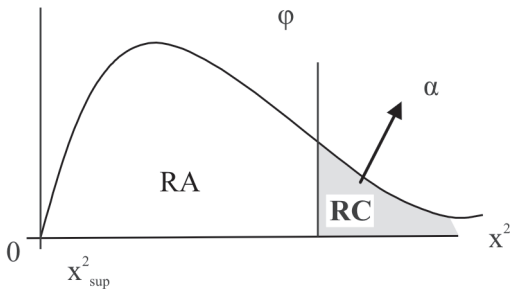
$$P(x_i, y_j) = p(x_i) \cdot p(y_j) \text{ para todo i e j.}$$

Passos para efetuar o teste:

1. H_0: as variáveis são independentes, ou as variáveis não estão associadas.
   H_1: as variáveis são dependentes, ou as variáveis estão associadas.

2. Fixar α. Escolher a variável qui-quadrado com $\varphi = (L - 1)(C - 1)$ onde L = nº de linhas da tabela de contingência, e C = nº de colunas.

3. Com o auxílio da tabela x^2, determinam-se RA e RC.



4. Cálculo do valor da variável.

$$X^2_{cal} = \sum_{i=1}^{L} \sum_{j=1}^{C} \frac{(Fo_{ij} - Fe_{ij})^2}{Fe_{ij}}$$

onde cada Fe_{ij} é determinado por:

$$Fe_{ij} = \frac{(\text{soma da linha } i)\ (\text{soma da coluna } j)}{\text{total de observações}}$$

5. Conclusão

Se $x^2_{cal} \in RA$, não se pode rejeitar H_0, isto é, não se pode dizer que as variáveis sejam dependentes.

Se $x^2_{cal} \in RC$, rejeita-se H_0, ou seja, concluindo-se com risco α que as variáveis são dependentes, ou estão associadas.

Aplicação:

Testar ao nível de 5% se há dependência entre as preferências por sabor da pasta de dentes e o bairro.

Sabor da pasta	**Bairros**			$\sum$
	A	B	C	
Limão	70	44	86	200
Chocolate	50	30	45	125
Hortelã	10	6	34	50
Outros	20	20	85	125
$\sum$	150	100	250	500

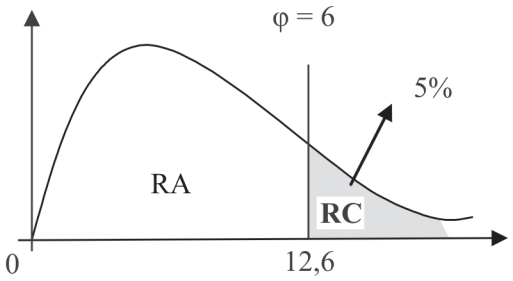
Fonte: Fonseca, Jairo Simon da. 2006: pág.231.

Solução:

1. H_0: A preferência pelo sabor independente do bairro.
   H_1: A preferência pelo sabor depende do bairro.

2. α = 5%. Escolher um x^2 com:
   φ = (4-1).(3-1) = 6 gl.

3. *RA* e *RC*



4. Cálculo do valor da variável.

A tabela das frequências esperadas é dada por:

Sabor	Bairros		
	A(1)	B(2)	C(3)
(1) Limão	60	40	100
(2) Chocolate	37,5	25	62,5
(3) Hortelã	15	10	25
(4) Outros	37,5	25	62,5

Onde, por exemplo,

$$Fe_{11} = \frac{(\text{soma da linha 1})\,(\text{soma da coluna 1})}{\text{total de observações}}$$

$$Fe_{11} = \frac{(150)\,(200)}{500} = 60$$

$$Fe_{43} = \frac{(\text{soma da linha 4})\,(\text{soma da coluna 3})}{\text{total de observações}}$$

$$Fe_{43} = \frac{(125)\,(150)}{500} = 62,5$$

Assim:

$$X^2_{cal} = \frac{(70-60)^2}{60} + \frac{(50-35,5)^2}{37,5} + \frac{(10-15)^2}{15} + \frac{(20-37,5)^2}{37,5} + \frac{(44-40)^2}{40} + \frac{(30-25)^2}{25}$$

$$\frac{(6-10)^2}{10} + \frac{(20-25)^2}{25} + \frac{(86-100)^2}{100} + \frac{(45-62,5)^2}{62,5} + \frac{(34-25)^2}{25} + \frac{(85-62,5)^2}{62,5} = \mathbf{37,88}$$

5. Conclusão

Como $x^2_{cal} \in RC$, rejeita-se H_0, concluindo-se, com risco de 5%, que há dependência entre sabor da pasta de dentes e o bairro.

TESTE DOS SINAIS

É utilizado para análise de dados emparelhados (o mesmo individuo é submetido a duas medidas). É aplicado em situações em que o pesquisador deseja determinar se duas condições são diferentes.

A variável em estudo poderá ser intervalar ou ordinal. O nome “teste dos sinais” se deve ao fato de serem utilizados os sinais “mais” e “menos”, em lugar dos dados numéricos. Assim, se houve alteração para maior, usa-se (+), se para menor, (-). Não havendo alteração, atribui-se (0). Para o teste, desconsideram-se os casos de empates, ou seja, os pares em que foram atribuídos zeros.

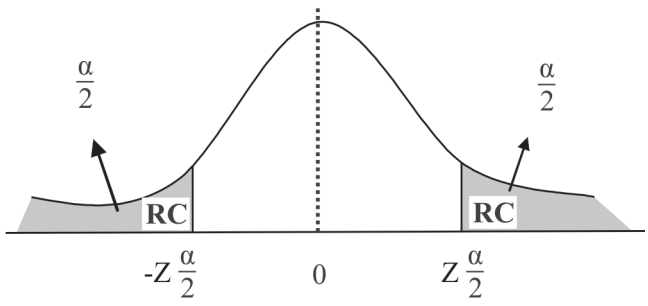
A “lógica” do teste é que as condições podem ser consideradas iguais quando as quantidades de “+” e “-“ forem aproximadamente iguais. Isto é, a proporção de sinais “+” equivale a 50%, ou seja: p = 0,5.

Procedimento para realização do teste:

1. H_0: não há diferença entre os grupos, ou seja: P = 0,5.
   H_1: há diferença, ou seja: uma das alternativas
   p ≠ (a)
   p > (b)
   p < (c)

2. Fixar α. Escolher a distribuição N(0,1) se n > 25, ou binomial se n ≤ 25.

3. Com o auxílio da tabela, determina-se RA e RC (para n > 25), caso n ≤ 25 utiliza-se a distribuição binomial.



4. Cálculo do valor da variável (n > 25)

$$Z_{cal} = \frac{y - n.p}{\sqrt{n.p.q}} \qquad 5.3$$

onde: y = número de sinais "+".

n = tamanho da amostra descontados os empates.

$p = 0,5$ *e* $q = 1 - p = 0,5$

5. Conclusões:

Se $z_{cal} \in RA$, não se pode rejeitar H_0. Se $z_{cal} \in RC$, rejeita-se H_0, concluindo-se, com risco α, que há diferença entre os dois grupos, ou duas condições.

Aplicações:

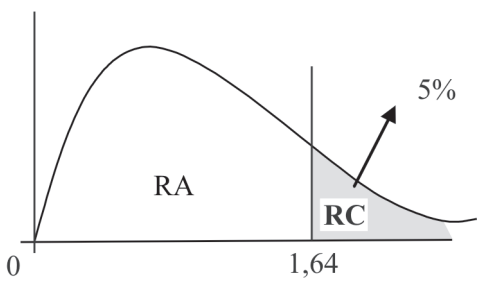
Sessenta alunos se matriculam num curso de inglês. Na primeira aula, aplica-se um teste que avalia o conhecimento da língua. Após seis meses, aplica-se um segundo teste. Os resultados mostram que 35 alunos apresentaram melhora (35 "+"), 20 se conduziram melhor no primeiro teste (20 "-") e 5 não apresentaram modificações (5 "0"). Testar, no nível de 5%, se o curso alterou o conhecimento de inglês do grupo de 60 alunos.

Solução:

1. H_0: O curso não alterou (p = 0,5).
   H_1: O curso melhorou o conhecimento de inglês.

2. α = 5%, Variável N(0,1)

3. *RA* e *RC*



Observe, devido ao enunciado de H_1, que se optou pelo teste unicaudal à direita. Caso H_1 fosse "piorou", o teste seria unicaudal à esquerda.

4. Cálculo do valor da variável:

$$Z_{cal} = \frac{35 - 55.(0,5)}{\sqrt{55.(0,5).(0,5)}} = \mathbf{2,02}$$

Onde $y = 35$

$n = 60\text{-}5 = 55 e\ p = q = 0,5$

5. Conclusão:

Como $z_{cal} \in RC$, rejeita-se H_0, concluindo-se com risco de 5%, que o curso melhorou o conhecimento de inglês.

TESTE DE MANN–WHITNEY

É usado para testar se duas amostras independentes foram retiradas de populações com médias iguais. Trata-se de uma interessante alternativa ao teste paramétrico para igualdade de médias, pois este teste não exige nenhuma consideração sobre as distribuições populacionais e suas variâncias. Como já vimos, o teste paramétrico para igualdade de médias exige populações com distribuições normais de mesma variância. Este teste poderá ser aplicado para variáveis intervalares ou ordinais.

Procedimento:

a) Considerar n_1 = nº de casos do grupo com menor quantidade de observações e n_2 = nº de casos do maior grupo.

b) Considere todos os dados dos dois grupos e coloque-os em ordem crescente. Atribua primeiro ao escore que algebricamente for menor e prossiga até $N = n_1 + n_2$.

Às observações empatadas, atribuir a média dos pontos correspondentes:

c) Calcular R_1 = soma dos postos do grupo n_1.

R_2 = soma dos postos do grupo n_2.

d) Escolher a melhor soma entre R_1 e R_2.

e) Calcular a estatística:

$$\mu_1 = n_1.n_2 + \frac{n_1(n_1+1)}{2} - R_1 \quad \textbf{5.4}$$

ou

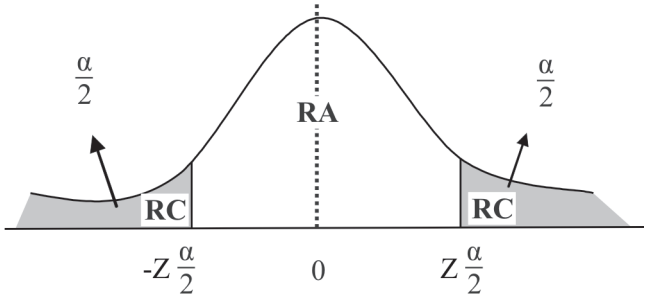
$$\mu_2 = n_1.n_2 + \frac{n_2(n_2+1)}{2} - R_2 \quad \textbf{5.5}$$

Teste:

1. H_0: não há diferença entre os grupos.
   H_1: há diferença.
Para n_1, n_2 < 10 há tabela própria.

2. Fixar α. Escolher a variável N(0,1).

3. Com o auxílio da tabela N(0,1) determinam-se RA e RC.



4. Cálculo do valor da variável.

$$Z_{cal} = \frac{\mu - \mu(u)}{\sigma(u)} \qquad \textbf{5.6}$$

onde:

$$\mu(u) = \frac{n_1 \cdot n_2}{2} \qquad \textbf{5.7}$$

$$\sigma(u) = \sqrt{\frac{n_1 . n_2 (n_1 + n_2 + 1)}{12}} \qquad \textbf{5.8}$$

5. Conclusão:

Se $z_{cal} \in RA$, não se pode rejeitar H_0.

Se $z_{cal} \in RC$, rejeita-se H_0, concluindo, com risco α, que há diferença entre os grupos.

Aplicação:

Determine no nível de 10%, se as vendas médias de dois “shopping centers” são diferentes.

Shopping A (em 10^6 R$)	Shopping B (em 10^6 R$)
10	22
18	17
9	15
8	10
2	7
11	7
4	8
3	14
9	15
12	-
10	-

Solução:

a) $n_1 = 9$ (shopping B) e $n_2 = 11$

b) Postos de todas as vendas.

A	B
11º	20º
19º	18º
8,5º	16,5º
6,5º	11º
1º	4,5º
13º	4,5º
3º	6,5º
2º	15º
8,5º	16,5º
14º	
11º	
Soma = 97,5	112,5

c) $R_2 = 97,5$ $R_1 = 112,5$

d) Escolher $R_2 = 97,5$

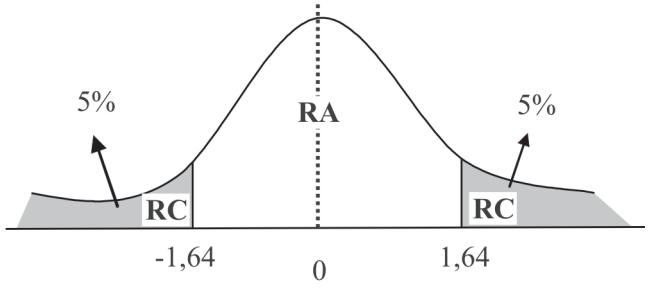
e) $u_2 = 9.(11) + \frac{11.(11+1)}{2} - 97,5 = 67,5$

Teste:

1. H_0: as vendas são iguais.
   H_1: as vendas são diferentes.

2. α = 10%. Escolher N(0,1)

3. Com o auxílio da tabela:



4. Cálculo do valor da variável

$$\mu(u) = \frac{(9).(11)}{2} = 49,5$$

$$\sigma(u) = \sqrt{\frac{9.(11)(9+11+1)}{12}} = 13,16$$

$$Z_{cal} = \frac{67,5 - 49,5}{13,16} = \mathbf{1,37}$$

5. Conclusão:

Como $z_{cal} \in$ RA, não se pode rejeitar a hipótese de que as vendas são iguais.

TESTE KRUSKAL–WALLIS

Este teste é extremamente útil para decidir se k amostras (k > 2) independentes provêm de populações com médias iguais. Poderá ser aplicado para variáveis intervalares ou ordinais.

Procedimento:

a) Dispor, em ordem crescente, as observações de todos os k grupos, atribuindo-lhes postos de 1 a n. Caso haja empates, atribuir o posto médio.

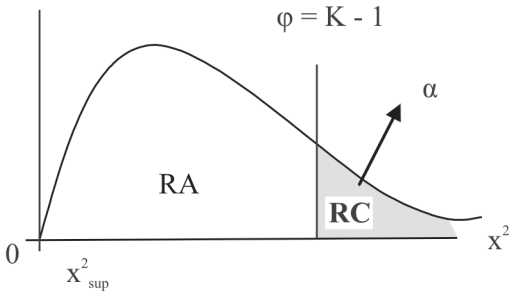
b) Determinar o valor da soma dos postos para cada um dos k grupos: R_i i = 1, 2, 3, ..., k.

c) Realizar o teste:

1) H_0: As médias são iguais.
H_1: Há pelo menos um par diferente.

2) Fixar α. Escolher uma variável qui-quadrado com $\varphi = k - 1$.

3) Com auxílio da tabela qui-quadrado, determinam-se RA e RC.



4) Calcula-se a estatística:

$$H = \frac{12}{n(n+1)} \cdot \sum_{i=1}^{k} \frac{(R_i)^2}{n_i} - 3.(n+1) \qquad \textbf{5.9}$$

5) Conclusão:

Se H ϵ RA, não se pode rejeitar H_0.

Se H_1 ϵ RC, rejeita-se H_0, concluindo-se com risco α que há diferença entre as médias dos k grupos.

Aplicação:

Testar, no nível de 5%, a hipótese da igualdade das médias para os três grupos de alunos que foram submetidos a esquemas diferenciados de aulas. Foram registradas as notas obtidas para uma mesma prova.

Aulas expositivas	Aulas com recursos audiovisuais	Aulas através de ensino programado
65	60	61
62	71	69
68	66	67
70	63	72
60	64	74
-	59	-

Solução:

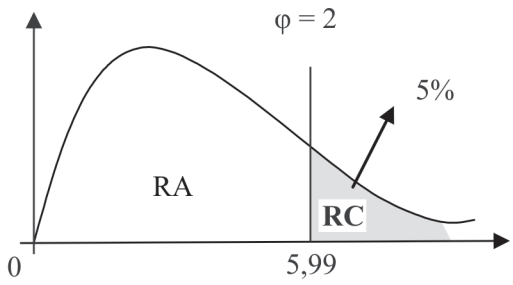
Atribuem-se postos às notas:

Postos		
Aulas Expositivas	**Aulas com recursos audiovisuais**	**Aulas através de ensino programado**
8º	2,5º	4º
5º	14º	12º
11º	9º	10º
13º	6º	15º
2,5º	7º	16º
-	1º	-
$\sum$ = 39,5	$\sum$ = 39,5	$\sum$ = 57

1) H_0: as notas médias são iguais.
H_1:as notas médias são diferentes.

2) α = 5%. Escolhe-se uma distribuição qui-quadrado com 2 gl, pois $\varphi = k - 1 = 3 - 1 = 2$.

3) Com auxílio da tabela da distribuição determinam-se RA e RC.



4) Cálculo da estatística H .

$$H = \frac{12}{16(16+1)} . [\frac{(39,5)^2}{5} + \frac{(39,5)^2}{6} + \frac{(57)^2}{5}\ ] - 3.(16+1)$$

$$H = 2,90$$

5) Conclusão:

Como H ϵ RA , não se pode rejeitar H_0. Assim, as notas médias podem ser consideradas iguais, ao nível de 5%.

EXERCÍCIO

1) Uma moeda é lançada 200 vezes e verifica-se 110 caras e 90 coroas. Teste a honestidade da moeda, sendo α = 0,10.

2) O número de livros emprestados por uma biblioteca, durante uma determinada semana, está indicado a seguir. Teste a hipótese de o número de livros emprestados não depender do dia da semana, sendo α = 0,01.

Dias da semana	Seg	Ter	Qua	Qui	Sex
Nº de livros emprestados	110	135	120	146	114

3) Teste (α = 5%) se há alguma relação entre as notas escolares e o salário.

Notas Escolares				
SALÁRIO	Alto	18	17	5
	Médio	26	38	16
	Baixo	6	15	9

4) Para a situação abaixo, aplique o teste dos sinais. Adote α = 2,5%.

Indivíduos submetidos a um programa de dieta

Peso (kg) Pré-dieta	Peso (kg) Pós-dieta	Continuação
55	50	48 50
63	65	49 51
78	78	90 91
81	79	93 85
68	70	90 90
58	57	56 58
60	58	66 64
60	62	67 68

5) Use o teste de Mann-Whitney para determinar se a média do grupo X é maior do que a média do grupo Y. Adote α = 1%

X :63 65 70 48 50 81 88 99 35 47 75 85 61
Y: 90 50 60 70 40 38 89 47 51 65 87.

UNIDADE 06

Correlação e Regressão Linear

Resumindo

Nesta unidade, abordamos um conteúdo muito importante para verificar se existe relação entre duas ou mais variáveis. A verificação da existência e do grau de relação entre variáveis é do objeto de estudo da correlação. Depois de caracterizada, procura-se descrever uma relação de forma matemática, através de uma função. A estimação dos parâmetros dessa função é objeto de estudo da regressão.

6 CORRELAÇÃO E REGRESSÃO LINEAR

INTRODUÇÃO

Frequentemente, procura-se verificar se existe relação entre duas ou mais variáveis. O peso, por exemplo, pode estar relacionado com a idade das pessoas; o consumo das famílias pode estar relacionado com sua renda; as vendas de uma empresa e os gastos promocionais podem relacionar-se, bem como a demanda de um determinado produto e seu preço. A verificação da existência e do grau de relação entre variáveis **é objeto de estudo da correlação.**

Uma vez caracterizada, procura-se descrever uma relação sob forma matemática, através de uma função. A estimação dos parâmetros dessa função matemática **é o objeto da regressão.**

Nota: *Para que uma relação possa ser descrita por meio de , é imprescindível que ela se aproxime de uma função linear. Uma maneira prática de verificarmos a linearidade da relação é a inspeção do diagrama de dispersão: se a elipse apresenta saliências ou reentrâncias muito acentuadas, provavelmente, trata-se de correlação curvilínea.*

CORRELAÇÃO LINEAR SIMPLES

O objetivo principal do estudo da correlação é medir e avaliar o grau de relação existente entre duas variáveis aleatórias. Assim, por exemplo, podemos medir se a relação entre o número de filhos de uma família e sua renda é forte, fraca ou nula.

A correlação linear procura medir a relação entre as variáveis X e Y através da disposição dos pontos (X, Y) em torno de uma reta.

Medida de Correlação

O instrumento de medida da correlação linear é dado pelo coeficiente de correlação de Pearson:

$$r_{XY} = \frac{\Sigma_{XY} - \frac{\Sigma_X . \Sigma_Y}{n}}{\sqrt{\left[\Sigma_{X^2} - \frac{(\Sigma_X)^2}{n}\right]\left[\Sigma_{Y^2} - \frac{(\Sigma_Y)^2}{n}\right]}} \qquad \textbf{6.1}$$

Onde: n = número de observações.

Símbolos comumente utilizados:

$$S_{XY} = \Sigma_{XY} - \frac{\Sigma_X . \Sigma_Y}{n} = \Sigma(Y - \bar{Y}).(X - \bar{X})$$

$$S_{XY} = \Sigma(X - \bar{X})^2 = \Sigma_{X^2} - \frac{(\Sigma_X)^2}{n}$$

$$S_{XY} = \Sigma(Y - \bar{Y})^2 = \Sigma_{Y^2} - \frac{(\Sigma_Y)^2}{n}$$

O campo de variação do coeficiente r situa-se entre -1 e +1.

$$-1 \leq r_{XY} \leq 1 \qquad \textbf{6.2}$$

Sua interpretação dependerá do valor numérico e do sinal.

Nota:
Para podermos tirar algumas conclusões significativas sobre o comportamento simultâneo das variáveis analisadas, é necessário que:

$$0{,}6 \leq |r_{XY}| \leq 1$$

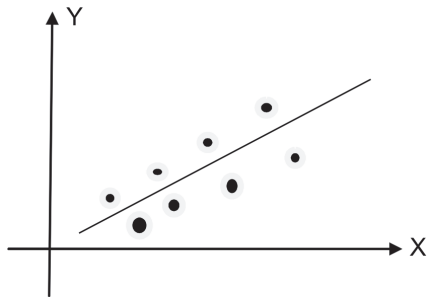
Se $0{,}3 \leq |r_{XY}| < 0{,}6$, há uma correlação relativamente fraca entre as variáveis.

Se $0 \leq |r_{XY}| < 0{,}3$, a correlação é muito fraca e, praticamente, nada podemos concluir sobre a relação entre as variáveis em estudo.

Correlação Linear Positiva

A correlação será considerada positiva se valores crescentes de X

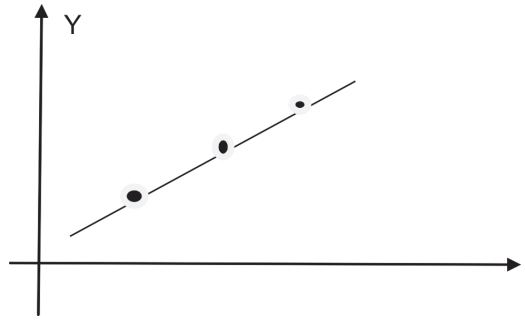
estiverem associados a valores crescentes de Y, ou valores decrescentes de X estiverem associados a valores decrescentes da variável Y.



$$0 \leq r_{XY} \leq 1$$

Correlação Linear Perfeita Positiva

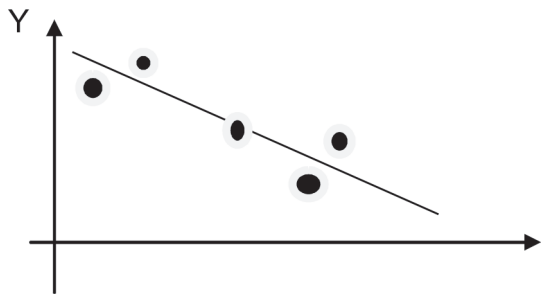
Essa correlação corresponde ao caso anterior, só que os pontos (X, Y) estão perfeitamente alinhados.



$$R_{XY} = 1$$

Correlação Linear Negativa

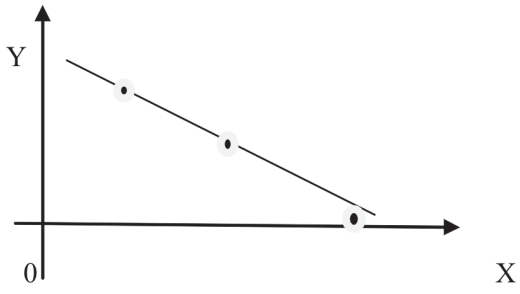
Essa correlação é considerada negativa quando valores crescentes da variável X estiverem associados a valores decrescentes de Y, ou valores decrescentes de X estiverem associados a valores crescentes da variável Y.



$-1 \leq r_{XY} \leq 0$

Correlação Linear Perfeita Negativa

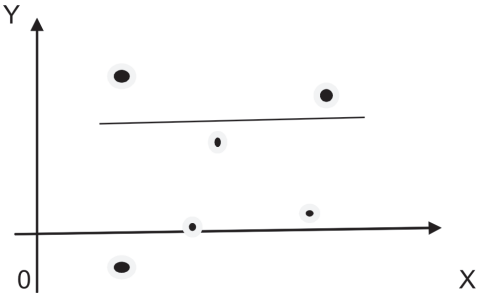
Quando os pontos estiverem perfeitamente alinhados, mas em sentido contrário, a correlação é denominada perfeita positiva.



$r_{XY} = -1$

Correlação Nula

Quando não houver relação entre X e Y, ou seja, quando as variações de X e Y ocorrerem independentemente, não existe correlação entre elas.



$R_{XY} = 0$

Cálculo Prático do Coeficiente de Correlação Linear

Para o cálculo do coeficiente de correlação, é conveniente a construção de uma tabela, onde, a partir dos valores X e Y, são determinadas todas as somas necessárias.

Y	X	X^2	Y^2	XY
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-
$\sum Y$	$\sum X$	$\sum X^2$	$\sum Y^2$	$\sum xY$

Exemplo 1:

Calcular o coeficiente de correlação linear entre as variáveis X e Y, usando os dados da tabela abaixo:

Tabela 6.1

Y	10	8	6	10	12
X	2	4	6	8	10

Fonte: Estatística Básica: Toledo, Geraldo Luciano, pag. 416.

Solução:

Tabela 6.2

Y	X	X^2	Y^2	XY
10	2	4	100	20
8	4	16	64	32
6	6	36	36	36
10	8	64	100	80
12	10	100	144	120
46	30	220	444	288

$n = 5$

$$r_{XY} = \frac{288 - \frac{(30)(46)}{5}}{\sqrt{[200 - \frac{(30)^2}{5}][444 - \frac{(46)^2}{5}]}} = \frac{12}{\sqrt{(40)(20,8)}} = \mathbf{0,416}$$

O resultado mostra que a correlação linear entre as variáveis X e Y é positiva (quando X cresce linearmente, Y também cresce linearmente), porém, é baixa.

Exemplo 2:

A tabela seguinte mostra os resultados de uma pesquisa com 10 famílias de determinada região:

Tabela 6.3

Famílias	Renda (R$ 100)	Poupança (R$ 1000)	Número de filhos	Média de anos de estudo da família
A	10	4	8	3
B	15	7	6	4
C	12	5	5	5
D	70	20	1	12
E	80	20	2	16
F	100	30	2	18
G	20	8	3	8
H	30	8	2	8
I	10	3	6	4
J	60	15	1	8

Fonte: Estatística Básica: Toledo, Geraldo Luciano, pág. 417

a) Calcular o coeficiente de correlação linear entre renda familiar e a poupança das dez famílias.

Solução:

Tabela 6. 4

Renda (Y)	Poupança (X)	X²	Y²	XY
10	4	16	100	40
15	7	49	225	105
12	5	25	144	60
70	20	400	4900	1400
80	20	400	6400	1600
100	30	900	10000	3000
20	8	64	400	160
30	8	64	900	240
10	3	9	100	30
60	15	225	3600	900
407	120	2152	26769	7535

$$r_{XY} = \frac{7535 - \frac{(407)(120)}{10}}{\sqrt{[2152 - \frac{(120)^2}{10}][26769 - \frac{(407)^2}{10}]}} = \mathbf{0,9835}$$

Este resultado r_{XY} = 0,9835 revela uma forte correlação linear entre renda e poupança familiar. O sinal do coeficiente mostra que as duas variáveis variam no mesmo sentido.

b) Calcular o coeficiente de correlação linear entre renda e número de filhos para as dez famílias.

Solução:

Tabela 6.5

Renda (Y)	**Nº de filhos (X)**	**X²**	**Y²**	**XY**
10	8	64	100	80
15	6	36	225	90
12	5	25	144	60
70	1	1	4900	70
80	2	4	6400	160
100	2	4	10000	200
20	3	9	400	60
30	2	4	900	60
10	6	36	100	60

60	1	1	3600	60
407	36	184	26769	900

$$r_{XY} = \frac{900 - \frac{(407)(36)}{10}}{\sqrt{[184 - \frac{(36)^2}{10}][26769 - \frac{(407)^2}{10}]}} = \mathbf{0,7586}$$

Este resultado r_{XY} = -0,758 revela uma correlação forte e inversa (negativa), ou seja, as famílias com maiores rendas têm menor número de filhos.

REGRESSÃO LINEAR SIMPLES

A análise de regressão tem como objetivo descrever, através de um modelo matemático, a relação existente entre duas variáveis, a partir de n observações dessas variáveis. Supondo X a variável explicativa e Y a variável explicada, dizemos que Y = f(x), ou seja, a variável Y é uma função da variável X. Em regressão, considera-se apenas a variável Y como aleatória e a variável X como supostamente sem erro. Então, a relação entre X e Y não é regida apenas por uma lei matemática, ou seja, para um dado valor de x, não observaremos necessariamente o mesmo Y. Portanto, a relação entre X e Y deverá ser escrita como segue: Y = f(x) + e, onde a variável e captará todas as influências sobre Y não devidas a X.

Dado um conjunto de valores observados de X e Y, construir um modelo de regressão linear de Y sobre X tendo como objetivo obter, a partir desses valores, uma reta que melhor represente a relação verdadeira entre essas variáveis. A determinação dos parâmetros dessa reta é denominada **ajustamento**.

O processo de ajustamento deve partir da escolha da função através da qual os valores de X explicarão os de Y. Para tanto, recorre-se a um gráfico conhecido como diagrama de dispersão. O mesmo é construído anotando, em um sistema de coordenadas retangulares, os pontos correspondentes aos pares de observações de X e de Y.

A reta ajustada é representada por $\hat{Y} = a + bX$, onde a e b são os parâmetros do modelo; **a** é o ponto onde a reta ajustada corta o eixo da variável Y, e **b** é a tangente do ângulo que a reta forma com uma paralela

ao eixo da variável X. A reta ajustada é denominada, também, reta de mínimos quadrados, pois os valores de a e b são obtidos de tal forma que é mínima a soma dos quadrados das diferenças entre os valores observados de Y e os obtidos a partir da reta ajustada para os mesmos valores de X. Simbolicamente, temos:

$\sum \hat{e} = \sum(\hat{Y} - Y)^2 = \sum(Y - a - bX)^2$ mínima, onde: ê = Y - $\hat{Y}$.

Para obter os parâmetros a e b, aplica-se a condição necessária de mínimo à função $\sum(Y - \hat{Y})^2$. Para isto, basta derivá-la com relação a esses parâmetros e igualamos as derivadas a zeros. As demonstrações das fórmulas você verá nas páginas 426 e 427 do livro: Estatística básica, de Geraldo Luciano Toledo. As fórmulas para o calculo dos parâmetros a e b são:

$$a = \overline{Y} - b\overline{X} \qquad \textbf{6.3}$$

Onde: $\overline{Y} = \frac{\Sigma Y}{n}$ e $\overline{X} = \frac{\Sigma X}{n}$

$$b = \frac{\Sigma XY - \frac{\Sigma X.\Sigma Y}{n}}{\Sigma X^2 - \frac{(\Sigma Y)^2}{n}} \qquad \textbf{6.4}$$

Costuma-se usar os seguintes símbolos para diminuir o numerador e denominador da expressão que definirá o valor de b:

$$S_{YX} = \Sigma XY - \frac{\Sigma X.\Sigma Y}{n}$$

$$b = \frac{S_{YX}}{S_{XX}}$$

$$S_{XX} = \Sigma X^2 - \frac{(\Sigma X)^2}{n}$$

Podemos escrever, então,

$$\hat{Y} = a + bX \qquad \textbf{6.5}$$

Aplicações:

1) Os dados abaixo se referem ao volume de precipitação pluviométrica (mm)

e ao volume de produção de leite tipo C (milhões de litros), em determinada região do país.

Anos	Produção de Leite C(1000.000 l)	Índice Pluviométrico (mm)
1970	26	23
1971	25	21
1972	31	28
1973	29	27
1974	27	23
1975	31	28
1976	32	27
1977	28	22
1978	30	26
1979	30	25

Fonte: pág. 427, Estatística Básica, Geraldo Luciano Toledo.

a) Ajustar os dados através de um modelo linear.

b) Admitindo-se, em 1980, um índice pluviométrico de 24 mm, qual deverá ser o volume esperado de produção de leite tipo C?

Solução:

a) Para efetuarmos os cálculos necessários ao ajustamento, recorremos a uma tabela com 4 colunas, contendo os valores de Y, X, X^2 e XY em cada uma das colunas, como segue:

Tabela 6.7

Y	X	X^2	XY
26	23	529	598
25	21	441	525
31	28	784	868
29	27	729	783
27	23	529	621
31	28	784	868
32	27	729	864
28	22	484	616
30	26	676	780

30	25	625	750
$\Sigma Y = 289$	$\Sigma X = 250$	$\Sigma X^2 = 6310$	$\Sigma XY = 7273$

I – determinação do valor do parâmetro b

$$b = \frac{S_{YX}}{S_{XX}} = \frac{\Sigma_{XY} - \frac{\Sigma_X . \Sigma_Y}{n}}{\Sigma_{X^2} - \frac{(\Sigma_X)^2}{n}} = \frac{7273 - \frac{(250)(289)}{10}}{6310 - \frac{(250)^2}{10}} = \frac{48}{60} = \mathbf{0,8}$$

$S_{XY} = 48$, $S_{XX} = 60$ e $b = 0,8$

II – determinação do parâmetro a

$$a = \overline{Y} - b\overline{X} = \frac{\Sigma_Y}{n} - b\,\frac{\Sigma_X}{n} = \frac{289}{10} - 0,8.\frac{250}{10} = \mathbf{8,9}$$

$a = 8,9$

III – equação da reta ajustada

$\hat{Y} = a + bX = 8,9 + 0,8$, logo,

$\hat{Y} = \mathbf{8,9 + 0,8X}$

b) Fazendo x = 24 mm, temos:

$\overline{Y} = 8,9 + 0,8.(24) = \mathbf{28,1}$

Logo, de acordo com o modelo, podemos esperar 28,1 milhões de litros produzidos para um índice pluviométrico de 24 mm.

2) Uma empresa está estudando a variação da demanda de certo produto em função do seu preço de venda. Para isso, levantou as seguintes informações:

Tabela 6.8

Meses	Unidades vendidas (Y)	Preço de venda (X) por unidade
J	248	162,00
F	242	167,00
M	234	165,00
A	216	173,00
M	230	170,00
J	220	176,00
J	213	178,00
A	205	180,00
S	198	182,00
O	195	187,00

Com base nestes dados, mostrar que a demanda do produto decresce linearmente com o acréscimo de preço.

Solução:

$\sum Y = 2201$ $\quad\quad$ $\sum XY = 381703$

$\sum Y = 1740$ $\quad\quad$ $\sum X^2 = 303340$

$\overline{X} = 174$ $\quad\quad$ $\overline{Y} = 220,1$

$$b = \frac{381703 - \frac{(1740).(2201)}{10}}{303340 - \frac{(1740)^2}{10}} = \mathbf{-2,19}$$

$$a = 220,1 - (-2,19).(174) = \mathbf{601,4}$$

$$\mathbf{\overline{Y} = 601,4 - 2,19X}$$

O resultado b = -2,19 significa que para cada unidade de variação positiva de preço (X), a quantidade procurada (Y) decresce 2,19 unidades.

O PODER EXPLICATIVO DO MODELO

Símbolo: R^2

Também denominado coeficiente de determinação, o poder explicativo da regressão tem como objetivo avaliar a "qualidade" do ajuste. Seu valor

fornece a proporção da variação total da variável Y explicada pela variável X através da função ajustada. Assim, podemos expressar R^2 por:

$$R^2 = \frac{b^2 S_{XX}}{S_{YY}} \qquad \textbf{6.6}$$

Com $0 \leq R^2 \leq 1$

$$R^2 = \frac{b.S_{YX}}{S_{YY}} \qquad \textbf{6.7}$$

Com $0 \leq R^2 \leq 100\%$

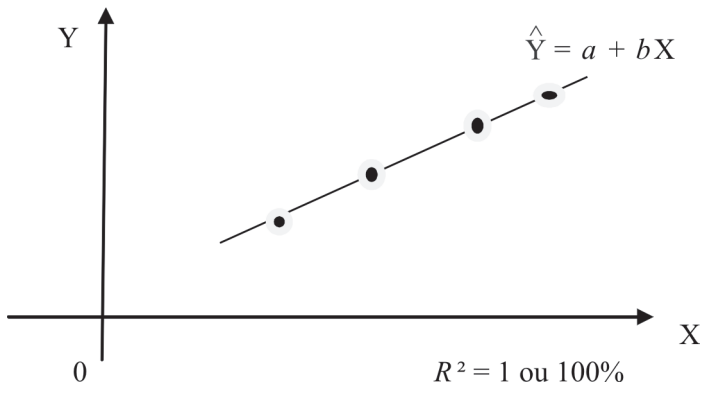
Onde: $S_{XX} = \Sigma X^2 - \frac{(\Sigma X)^2}{n}$

$S_{YY} = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n}$

$S_{YX} = \Sigma XY - \frac{\Sigma X.\Sigma Y}{n}$

Quando $R^2 = 0$, a variação explicada de Y é zero, ou seja, a reta ajustada é paralela ao eixo da variável X. Se R^2 for igual a 1, a reta ajustada explicará toda a variação de Y. Assim sendo, quanto mais próximo da unidade estiver o valor de R^2, melhor a "qualidade"do ajuste da função aos pontos do diagrama de dispersão e quanto mais próximo de zero pior será a "qualidade" do ajuste.

Se o poder explicativo for, por exemplo, 98%, isto significa que 98% das variações de Y são explicadas por X através da função escolhida para relacionar as duas variações e 2% são atribuídas a causas aleatórias. Observe o gráfico abaixo.



$bS_{YX} = S_{YY}$

Aplicações:

01) Calcular o poder explicativo da regressão para os dados da tabela 6.8 do exemplo 2 da página 67.

Solução:

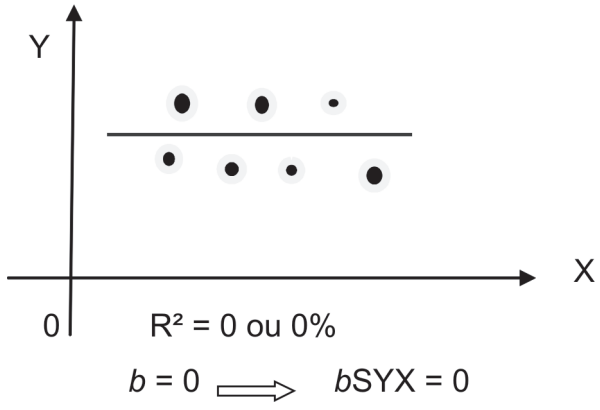
$\sum X$ = 3.915,5 $\sum Y$ = 3.273,4

$\sum X^2$ = 1.150.349,73 $\sum Y^2$ = 800.330,16

S_{XX} = 55.268,28 S_{YY} = 34.962,48

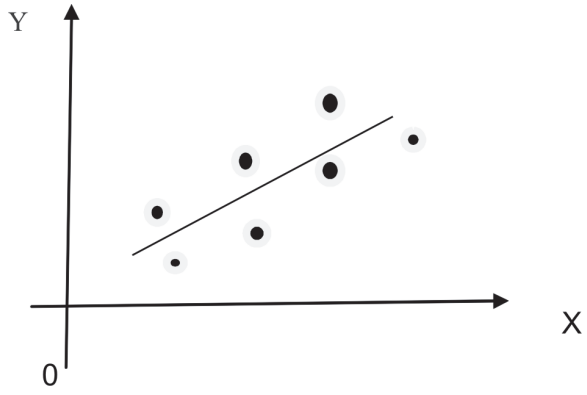
b = 0,971

Usando a fórmula (6.6), obtemos:



$R^2 = 0$ ou 0%

$b = 0 \Longrightarrow bSYX = 0$



$0 \leq R^2 \leq 1$

O resultado mostra um excelente grau de ajuste da reta aos pontos. A relação linear obtida explica 98,9% das variações totais da variável Y. Somente 1,1% das variações de Y são consideradas aleatórias caso seja adotado o modelo linear.

02) Seja Y uma variável que representa o valor do frete rodoviário de determinada mercadoria e X a variável distância (em km) ao destino da mercadoria. Uma amostra de 10 observações das variáveis apresentou os seguintes resultados:

n = 10 $\quad\quad$ $\sum XY$ = 842.060

$\sum X$ = 1.200 $\quad\quad$ $\sum Y^2$ = 4.713.304,03

$\sum Y$ = 6.480,50 $\quad\quad$ $\sum X^2$ = 186.400

a) Determine a regressão: $\hat{Y}$ = a + bX.

b) Interprete os valores encontrados para a e b.

c) Calcule e interprete o poder explicativo da regressão. Agora tente resolvê-los e depois confira as respostas abaixo:

Respostas:

a) $\hat{Y}$ = 624,05 + 0,50X.

b) a = 624,05 = parte do frete que não depende da distância; b = 0,50 = acréscimo no frete por quilômetro rodado.

c) R^2 = 0,311 ou 31,1%. A distância explica muito pouco das variações do frete.

03) Calcule o poder explicativo da regressão usando os dados da tabela 5.8 do exemplo 2.

Solução:

$\sum Y$ = 2.201 $\quad\quad$ $\sum Y^2$ = 487.403

$\sum X$ = 1.740 $\quad\quad$ $\sum X^2$ = 303.340

n = 10 $\quad\quad$ b = -2,19

$$R^2 = \frac{(-2,19)^2 .[303.340 - \frac{(1740)^2}{10}]}{487.403 - \frac{(2201)^2}{10}} = \frac{2781,74}{2962,90} = \mathbf{0,939}$$

ou R^2 = 93,9%. Ou seja, 93,9% das variações da demanda são

explicadas por variações de preço.

EXERCÍCIO

1) Com os dados da tabela 6.5, calcule o coeficiente de correlação entre
a) Poupança e Nº de filhos.
b) Média dos anos de estudo e Nº de filhos
c) Renda familiar e Média de anos de estudo.

2) A tabela abaixo apresenta informações sobre o custo de determinada mercadoria, em reais, e a distância em km do destino para onde deve ser enviada.

Custo	Distância
125,8	229
134,8	287
127,0	209
122,8	174
123,4	190
122,2	196
122,5	186
122,8	202
122,4	178
122,0	168
124,1	192
124,7	210
122,3	168

a) Estime, por regressão linear, o custo fixo e o custo por km rodado. Escreva a equação ajustada.
b) Calcule o poder explicativo do modelo.

UNIDADE 07

Análise de Variância - Comparação de Várias Médias

Resumindo

Nas unidades anteriores foram apresentados testes paramétricos e não paramétricos para verificar a igualdade entre duas Médias: teste T e de Man-Whitney e ainda no capítulo 6 e o teste Kruskal - Wallis foi aplicado para testar a igualdade de k médias, k>2. Uma alternativa ao teste de kruskal - Wallis é a análise de variância, que é um método estatístico, desenvolvido por Fisher, o qual através de testes de igualdade de médias é possível verificar se fatores produzem mudanças sistemáticas em alguma variável de interesse. Os fatores propostos podem ser variáveis quantitativas ou qualitativas, enquanto a variável depende deve ser quantitativa (intervalar) e é observada dentro das classes dos fatores – os tratamentos. A finalidade desta unidade é apresentar os fundamentos desse método.

7 ANÁLISE DE VARIÂNCIA - COMPARAÇÃO DE VÁRIAS MÉDIAS

INTRODUÇÃO

Devido à importância da questão, dedicaremos toda esta unidade ao estudo dos problemas envolvendo a comparação de várias médias.

Importante:

A análise de variância é um teste de hipótese usado para comparação de mais de duas populações. Imagine que você queira comparar o grande endividamento de empresas de três setores (indústria, comércio e prestação de serviços). Para a comparação, é necessário que você tenha repetições, pois são elas que medirão a variação do acaso. Então, você deve selecionar uma amostra de dez empresas de cada setor (repetições).

A principal e mais importante técnica que utilizamos para a solução do problema é a análise de variância, que foi inicialmente desenvolvida pelo grande estatístico britânico R. A. Fisher, como instrumento para a análise de experimentos agrícolas. Concomitantemente, foram sendo desenvolvidos diversos modelos de planejamento de experimentos, os quais, entretanto, serão apenas parcialmente examinados nesta unidade.

A análise de variância é um método suficientemente poderoso para identificar diferenças entre as médias populacionais devidas a várias causas atuando simultaneamente sobre os elementos da população.

Nosso objetivo é apresentar a ideia fundamental do método de forma simplificada, sem grande aprofundamento teórico, já que isso demandaria um vasto espaço e fugiria à nossa meta.

HIPÓTESES DO MODELO

Há três suposições básicas que devem ser satisfeitas para que se possa aplicar a análise da variância.

1. As amostras devem ser aleatórias e independentes.
2. As amostras devem ser extraídas de populações normais.
3. As populações devem ter variâncias iguais.

Classificação única ou experimento de um fator

Admite-se um único fator (variável independente) que é subdividido em tratamentos (níveis do fator). A variável de estudo (variável dependente) é medida através de amostras de cada tratamento. Eis a configuração desse tipo de experimento:

Tratamentos / Elemento da amostra	1	2	3	. . .	k
1	x_{11}	x_{21}	x_{31}	. . .	x_{k1}
2	x_{12}	x_{22}	x_{32}	. . .	x_{k2}
3	x_{13}	x_{23}	x_{33}	. . .	x_{k3}
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
n_i	$x_1 n_1$	$x_2 n_2$	$x_3 n_3$	. . .	$x_k n_k$
Somas					Total
Médias	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	. . .	$\bar{x}_k$ $\quad \bar{X}$

$i = 1, 2, 3, \ldots, k$

$j = 1, 2, 3, \ldots, n_i$

Assim x_{ij}, denota o valor da j-ésima observação sujeito ao i-ésimo tratamento.

A média dos valores observados no i-ésimo grupo será:

$$\bar{x}_i = \frac{\sum_{i=1}^{n_i} x_{ij}}{n_i} \qquad \textbf{7.1}$$

i = 1, 2, 3, . . . ,k

A média geral é dada por:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij} \quad \text{ou} \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i \qquad \textbf{7.2}$$

Em que $n = \sum_{i=1}^{k} n_i$ é o número total de observações.

A hipótese nula é de que todos os tratamentos tenham médias iguais, isto é:

H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$

E que todas as "k" populações dos tratamentos tenham a mesma variância: σ^2.

A hipótese alternativa é de que pelo menos um par de médias seja diferente:

H_1: $\mu_p \neq \mu_q$ para $p \neq q$

A aceitação de H_0 revelará que o fator considerado não acarreta mudanças significativas na variável de estudo. Por outro lado, a rejeição de H_0 indicará, com risco α, que o fator considerado exerce influência sobre a variável de estudo.

A base da análise da variância está nas comparações que podem ser feitas com os estimadores da variância comum de todos os tratamentos (σ^2).

Estimadores da variância comum σ^2

Admitindo-se H_0 como verdadeira, pode-se estimar a variância comum de três maneiras diferentes:

1º) No primeiro caso, consideram-se os k tratamentos como uma única amostra de tamanho n e a média geral $\overline{X}$. Se H_0 : $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k = \mu$ é verdadeira, tem-se que:

$$S^2_t = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2}{n - 1} \qquad \textbf{7.3}$$

Será um estimador justo de σ^2, isto é, $E[S^2_t] = \sigma^2$. Por outro lado, se H_0 não for verdadeira, S^2t irá superestimar σ^2.

Ao numerador $\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2$, denomina-se variação total (Qt).

Pelo teorema de Fisher: $\frac{\Sigma\Sigma_{j=1}^{n_i}(x_{ij} - \bar{x})^2}{\sigma^2}$, tem distribuição qui-quadrado com (n - 1) graus de liberdade.

Desenvolvendo-se o quadrado, obtém-se uma fórmula prática para o cálculo da variação total. Assim:

$$Variação\ total = Qt = \Sigma_{i=1}^{k}\Sigma_{j=1}^{n_i} x_{ij}^2 - C \qquad \textbf{7.4}$$

Onde:

$$C = \frac{(\Sigma_{i=1}^{k}\Sigma_{j=1}^{n_i} x_{ij})^2}{n} \qquad \textbf{7.5}$$

2º) A segunda forma de se estimar a variância comum σ^2 é pela consideração das médias dos grupos e a média geral $\overline{X}$. Se H_0 for verdadeira, teremos para cada amostra:

$E[\bar{x}] = \mu$ e $\sigma^2(\bar{x}_i) = \frac{\sigma^2}{n}$, ou $\bar{x}_i = N(\mu ; \frac{\sigma^2}{n})$. Então

$$S^2 = \frac{\Sigma_{i=1}^{k}(\bar{x}_i - \bar{x})^2}{k - 1} \qquad \textbf{7.6}$$

Será um estimador justo de $\frac{\sigma^2}{n}$ e

$$S_e^2 = n_i S^2 = \frac{\Sigma_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2}{k - 1} \qquad \textbf{7.7}$$

Será um estimador justo de σ^2, isto é: $E[S_e^2] = \sigma^2$.

Porém, se H_0 não for verdadeira, S_e^2 irá superestimar σ^2.

Ao numerador $\Sigma_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2$, denomina-se variação entre tratamentos. Pelo Teorema de Fisher: $\frac{\Sigma_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2}{\sigma^2}$, tem distribuição qui-quadrado com $(k - 1)$ graus de liberdade.

Fórmula prática para o cálculo de Q_e .

$$Q_e = \Sigma_{i=1}^{k}\left[\frac{(\Sigma_j x_{ij})^2}{n_i}\right] - C \qquad \textbf{7.8}$$

3º) A terceira maneira de se estimar a variância σ^2 comum será por meio de cada uma das k amostras. Assim, para o i-ésimo tratamento, tem-se:

$$S_i^2 = \frac{\Sigma_{i=1}^{n_i}(x_{ij} - \bar{x})}{k - 1} \qquad \textbf{7.9}$$

Como i = 1, 2, 3, . . . , k, tem-se "k" estimadores do tipo S_i^2 . Então o estimador da variância comum será dado pela média aritmética dos S_i^2

ponderadas pelos respectivos graus de liberdade (ϕ_i = n – 1), assim:

$$S_r^2 = \frac{\phi_1 S_i^2 + \phi_2 S_2^2 + ... + \phi_k S_k^2}{\phi_1 + \phi_2 + ... + \phi_k}$$, ou seja:

$$S_r^2 = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{n - k} \qquad \textbf{7.10}$$

Sob a condição de H_0 ser verdadeira ou não, tem-se $E[S_r^2] = \sigma^2$, isto é, S_n^2 é justo. (Veja configuração à frente).

Ao numerador $\sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$ denomina variação dentro dos tratamentos ou variação residual (Q_r). Pelo Teorema de Fisher:

$$\frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{\sigma^2}$$

tem distribuição qui-quadrado com $(n - k)$ graus de liberdade.

Fórmula prática para o cálculo de Q_r.

$$Q_r = \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}^2 \sum_i \left[\frac{(\sum_j x_{ij})^2}{n_i}\right] \qquad \textbf{7.11}$$

Pode-se demonstrar que:

$$Q_t = Q_e + Q_r \qquad \textbf{7.12}$$

Isto é, a variação total é igual à soma da variação entre tratamentos e a variação residual. Ou seja, $Q_t - Q_e - Q_r = 0$.

Resumindo, as variações Q_t , Q_e e Q_r têm distribuição x^2 respectivamente com $(n - 1)$, $(k - 1)$ e $(n - k)$ graus de liberdade.

Isto é: = $x^2_{n-1} = x^2_{k-1} + x^2_{n-k}$.

Nota-se que: $(n - 1) = (k - 1) + (n - k)$.

Sendo esta a condição necessária e suficiente para que x^2_{k-1} e x^2_{n-k} e sejam independentes. Assim:

$$F = \frac{\frac{x^2_{k-1}}{k-1}}{\frac{x^2_{n-k}}{n-k}} = \frac{\frac{Q_e}{k-1}}{n-k} = \frac{S_e^2}{S_r^2} \qquad \textbf{7.13}$$

Terá distribuição F com $(k - 1)$ g.l. no numerador e $(n - k)$ g.l. no denominador

O quociente F será utilizado para testar a hipótese H_0.

Quanto mais próximo de 1 for o quociente, H_0 deverá ser aceita; ao contrário, quanto maior o valor de F, o teste irá indicar a rejeição de H_0, e

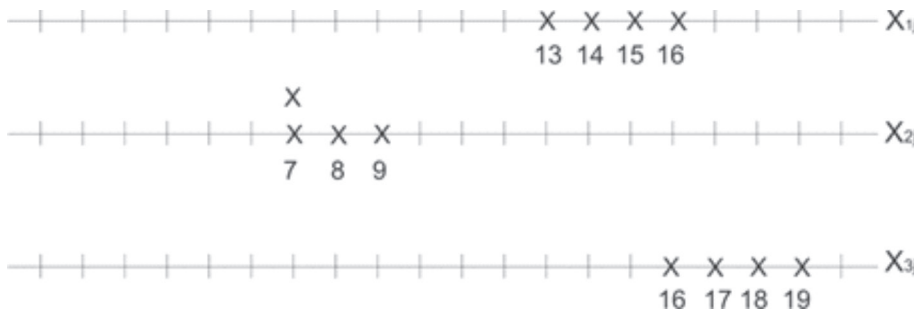
nesse caso conclui-se com risco α que o fator considerado tem influência sobre a variável dependente.

Fundamentos da análise da variância (ANOVA)

A configuração seguinte ilustra a base do método e o consequente uso do quociente f para se testar a hipótese da igualdade das médias. Para tanto, vamos supor três amostras de quatro elementos cada uma, cujos valores são:

Amostra 1: X_{1j} :	14	16	15	13
Amostra 2: X_{2j} :	7	7	8	9
Amostra 3: X_{2j} :	16	17	18	19

E que estão mostradas no gráfico:



Vê-se claramente um caso em que a hipótese H_0 será rejeitada pela análise de variância. As três amostras parecem confirmar a hipótese de homocedasticidade (variâncias iguais); todavia, as médias diferem claramente de amostra para amostra: ($\bar{x}_1$ = 14,4; $\bar{x}_2$ = 7,75; $\bar{x}_3$ = 17,5). Calculando-se, dessa forma, os valores das estimativas da variância encontraremos:

S^2_t = 19,30; S^2_e = 99,75; S^2_e = 1,42.

Nota-se pela análise do gráfico e dos resultados que, sendo H_0 falsa, haverá uma tendência de S^2_t e S^2_e superestimarem (19,30 e 99,75, respectivamente) σ^2. O que não ocorre com S^2_r, já que S^2_r = 1,42 é uma boa estimativa de σ^2.

Contrariamente, se H_0 for verdadeira, S^2_t, S^2_e e S^2_r fornecerão boas estimativas para σ^2. Imagine, olhando para o gráfico da página anterior com os três grupos alinhados em torno de um eixo vertical ($\bar{x}$).

Tem-se aí, o fundamento lógico da análise da variância. Na verdade,

o teste de igualdade de médias é substituído por um teste de igualdade de variâncias: $\sigma^2_e = \sigma^2_r$.

Assim, se a hipótese de igualdade das variâncias for aceita, pode-se concluir que as médias são iguais, pois neste caso o estimador S^2_e terá a mesma dimensão que S^2_r, ou seja, o quociente entre ambos estará próximo da unidade. Porém, se a hipótese da igualdade das médias não é verificada, se terá S^2_e bem maior do que S^2_r e, consequentemente, o valor do quociente será bem maior do que a unidade.

É fácil compreender que o teste da análise da variância será unicaudal à direita, com risco concentrado na cauda à direita.

Quadro de análise da variância

Os resultados obtidos poderão ser reunidos no quadro de Análise da Variância, assim:

Quadro de Análise da Variância				
Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Quadrados Médios	Teste F
Entre tratamentos	Q_e	$K - 1$	$S^2_e = \frac{Q_e}{k - 1}$	$F_{cal} = \frac{S^2_e}{S^2_r}$
Dentro das Amostras (Residual)	$Q_r = Q_t - Q_e$	$n - k$	$S^2_r = \frac{Q_t - Q_e}{n - k}$	
Total	Q_t	$n - 1$	-	-

Fonte: Fonseca, Jairo Simon da. 2006: pág.260

Para testar a hipótese H_0: $\mu_1 = \mu_2 = \mu_3 = ... = \mu_k = \mu$ contra H_1: $\mu_1 \neq \mu_2$, para $p \neq q$, compara-se o valor F_{cal} com o valor F tabelado com $(n - 1)$ g.l. no numerador e $(n - k)$ no denominador, fixando certo nível α de significância.

Se $F_{cal} < F_{tab}$, então aceita-se H_0 e conclui-se com risco que o fator considerado não causa efeito sobre a variável em estudo. Por outro lado, se $F_{cal} > F_{tab}$, rejeita-se H_0, concluindo-se pela diferença das médias a consequente influência do fator sobre a variável analisada.

Segue procedimento para a realização do teste:

1º) Dispor os elementos, segundo a tabela a seguir, obtendo as somas das colunas e suas respectivas médias.

Tratamentos / Elemento da Amostra	1 2 3 . . . k	
1 2 . . n_i	$x_{11}\ x_{21}\ x_{31} \ldots x_{k1}$ $x_{12}\ x_{22}\ x_{32} \ldots x_{k2}$ . . . . $x_1n_1\ x_2n_2\ x_3n_3 \ldots x_kn_k$	
$\sum$		
$\bar{x}_i$		$\bar{X}$

2º) Calcula-se a constante

$$C = \frac{(\Sigma_{i=1}^{k} \Sigma_{j=1}^{n_i} x_{ij})^2}{n} \qquad \textbf{7.14}$$

3º) Avalia-se a soma: $\Sigma_{i=1}^{k} \Sigma_{j=1}^{n_i} x_{ij}^2$, obtendo a variação total

$$Q_t = \Sigma_{i=1}^{k} \Sigma_{j=1}^{n_i} x_{ij}^2 - C \qquad \textbf{7.15}$$

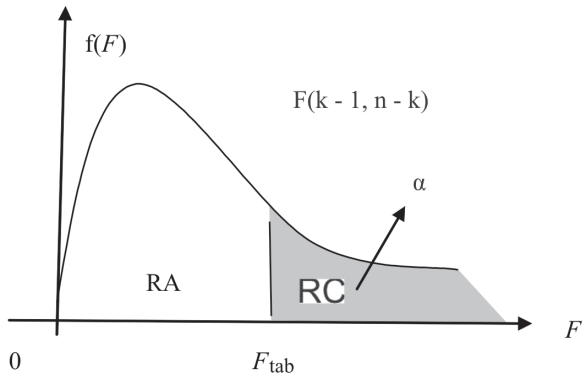
4º) Calcula-se a variação entre tratamentos

$$Q_e = \Sigma \left[\frac{(\Sigma_j x_{ij})^2}{n_i}\right] - C \qquad \textbf{7.16}$$

5º) Obtém-se a variação residual por diferença: $Q_r = Q_t - Q_e$.

6º) Constrói-se o Quadro de Análise da Variância, avaliando o F_{cal}.

7º) Determina-se a região crítica e de aceitação da hipótese H_0 por meio da tabela F.



8º) Compara-se F_{cal} com F_{tab}, obtendo-se a conclusão.

Aplicação:

O resultado das vendas efetuadas por 3 vendedores de uma indústria durante certo período é dado a seguir. Deseja-se saber, ao nível de 5%, se há diferença de eficiência entre os vendedores.

Vendedores		
A	B	C
29	27	30
27	27	30
31	30	31
29	28	27
32	-	29
30	-	-

Fonte: Fonseca, Jairo Simon da. 2006: pág.262

Solução:

Sem efetuar os resultados, pode-se subtrair uma constante, digamos 28, a todos os valores, simplificando dessa forma os cálculos.

Assim:

1º) Dispor os elementos segundo a tabela abaixo:

Vendedores			
A	B	C	
1	-1	2	Total
-1	-1	2	
3	2	3	
1	0	-1	
4	-	1	
2	-	-	
$\sum = 10$	$\sum = 0$	$\sum = 7$	17

2º) $C = \frac{(17)^2}{15} = \mathbf{19,27}$

3º) $\Sigma_{i=1}^{3}\Sigma_{j=1}^{n_i} x_{ij}^2 = 1^2 + (-1)^2 + 3^2 + 1^2 + \ldots + (-1)^2 + 1^2 = 57$

$Q_t = 57 - 19,27 = \mathbf{37,73}$

4º) $Q_e = \Sigma_{i=1}^{3}\left[\frac{(\Sigma_j x_{ij})^2}{n_i}\right] - C$

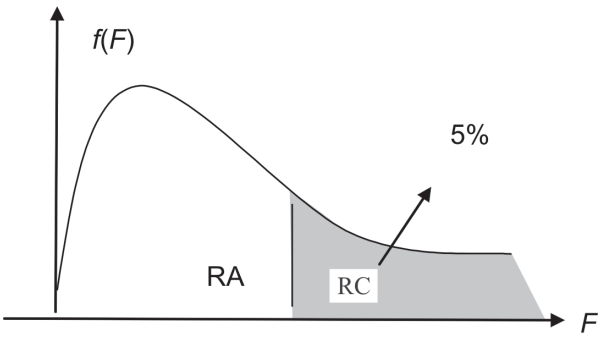
$= \frac{(10)^2}{6} + \frac{0^2}{4} + \frac{7^2}{5} - 19,27 = \mathbf{7,20}$

5º) $Q_r = Q_t - Q_e = 37,73 - 7,20 = \mathbf{30,53}$

6º) QAV

Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Quadrados Médios	Teste F
Entre os Tratamentos	$Q_e = 7,20$	$K - 1$ $3 - 1 = 2$	$S_e^2 = \frac{7,2}{2} = 3,6$	$F_{cal} = \frac{3,6}{2,54} = 1,44$
Residual	$Q_r = 30,53$	$n - k$ $15 - 3 = 12$	$S_r^2 = \frac{30,53}{12} = 2,54$	
Total	$Q_t = 37,73$	$n - 1$ $15 - 1 = 14$	-	

7º)



8º) Como F_{cal} = 1,42 < F_{tab} = 3,89, aceita-se H_0, concluindo-se com nível de 5% que não há diferença entre os vendedores, isto é, aceita-se H_0: $\mu_A = \mu_B = \mu_C$.

CLASSIFICAÇÃO DE DOIS CRITÉRIOS OU EXPERIMENTO DE DOIS FATORES

Admitem-se dois fatores (variáveis independentes). Variável de estudo (variável dependente) é observada em cada cela, combinação dos tratamentos do fator 1, e dos blocos do fator 2.

Tem-se uma tabela de "k" colunas e "L" linhas. Ou seja, $K.L = n$ observações.

	Primeiro critério (colunas) = Tratamento
Segundo critério (linhas) = Blocos	x_{11} x_{21} . . . x_{k1} x_{12} x_{22} . . . x_{k2} . . . .
	x_{1L} x_{2L} x_{kL}

Considere um experimento de natureza agrícola consistindo no exame das safras por are de 3 variedades de soja, em que cada variedade é plantada em 4 lotes diferentes de terra. Há um total de 3.4 =12 lotes. Em tal caso é conveniente combinar os lotes em blocos, digamos 3 lotes constituindo um bloco, com uma variedade distinta de soja plantada em cada lote do bloco. São, então, necessários 4 blocos. Neste caso, há duas classificações ou fatores, pois pode haver diferença na produção por are devida: ao tipo de soja; ou ao particular bloco considerado.

Denota-se por x_i a média de uma coluna i qualquer, por x_j a média de uma linha j qualquer por X a média geral:

$$\overline{x} = \frac{1}{L}\Sigma_{j=1}^{L} x_{ij} \qquad \textbf{7.17}$$

$$\overline{x}_j = \frac{1}{k}\Sigma_{i=1}^{k} x_{ij} \qquad \textbf{7.18}$$

$$\overline{X} = \frac{1}{n}\Sigma_{i=1}^{k}\,\Sigma_{j=1}^{L} x_{ij} \qquad \textbf{7.19}$$

Assim como no caso da classificação única, admite-se que todas as amostras provenham de populações normais com a mesma variância:

Para a comparação das médias entre colunas (tratamentos), a hipótese nula será: H_0^c: $\mu_i = \mu$ para qualquer i = 1, 2, .., k será testada contra a hipótese alternativa H_1^c : $\mu_i \neq \mu$.

Analogamente, para a comparação das médias entre linhas (blocos), a hipótese que será colocada à prova será: $H_0^L : \mu_j \neq \mu$ para qualquer j = 1, 2, ... , L, enquanto $H_1^L: \mu_j \neq \mu$.

ESTIMADORES DA VARIÂNCIA COMUM σ^2

Pode-se estimar a variância de 4 formas diferentes. A estimativa total S^2_t, a estimativa entre linhas S^2_L, a estimativa entre colunas S^2_C, e a estimativa residualS^2r. Assim:

1º) Estimativa total: S^2_t

$$S^2_t = \frac{\Sigma_{i=1}^{k} \Sigma_{j=1}^{L} (x_{ij} - \bar{x})^2}{n - 1} = \frac{Q_t}{n - 1} \qquad \textbf{7.20}$$

O numerador (Q_t) representa a variação total. Por outro lado, sabe-se que $\frac{Q_t}{\sigma^2}$ tem distribuição x^2 com (k - 1) g.l. A fórmula prática da variação total é dada por:

$$S^2_t = \Sigma_{i=1}^{k} \Sigma_{j=1}^{L} x^2_{ij} - C \qquad \textbf{7.21}$$

$$C = \frac{(\Sigma_i \Sigma_j^L x_{ij})^2}{n} \qquad \textbf{7.22}$$

2º) Estimativa entre Colunas: S^2_C

$$S^2_C = L \frac{\Sigma_{i=1}^{k} (\bar{x}_i - \bar{x})^2}{k - 1} = \frac{Q_{EC}}{k - 1} \qquad \textbf{7.23}$$

O numerador (Q_{EC}) representa a variação entre colunas. Demonstra-se que $\frac{Q_{EC}}{\sigma^2}$ tem distribuição x^2 com (k - 1)g.l. A fórmula prática para o cálculo de Q_{EC} é dada por:

$$Q_{EC} = \Sigma_{i=1}^{k} \left[\frac{(\Sigma_i x_{ij})^2}{L} \right] - C \qquad \textbf{7.24}$$

3º) Estimativa entre Linhas: S^2_L

$$S^2_L = k \frac{\Sigma_{j=1}^{L} (\bar{x}_j - \bar{x})^2}{L - 1} = \frac{Q_{EL}}{L - 1} \qquad \textbf{7.25}$$

A variação entre linhas é dada pelo numerador da expressão. Da mesma maneira, $\frac{Q_{EL}}{\sigma^2}$ tem distribuição x^2 com (L – 1) g.l. A fórmula prática para seu cálculo é dada por:

$$Q_{EL} = \sum_{i=1}^{L} \left[\frac{(\sum_i x_{ij})^2}{k} \right] - C \qquad \textbf{7.26}$$

4º) Estimativa Residual: S^2_r

$$S^2_t = \frac{\sum_{i=1}^{k} \sum_{j=1}^{L} [(x_{ij} - \bar{x}) - (\bar{x}_i - \bar{x}) - (\bar{x}_j - \bar{x})]^2}{(k-1)(L-1)} \qquad \textbf{7.27}$$

ou

$$S^2_r = \frac{\sum_{i=1}^{k} \sum_{j=1}^{L} (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2}{(k-1)(L-1)} = \frac{Q_r}{(k-1)(L-1)} \qquad \textbf{7.28}$$

A variação residual é dada pelo numerador e também neste caso $\frac{Q_r}{\sigma^2}$ em distribuição x^2 com $(k-1)(L-1)$ graus de liberdade. A avaliação de Q_r é obtida por diferença, já que também neste caso é válida a igualdade:

$$Q_t = Q_{EC} + Q_{EL} + Q_r \qquad \textbf{7.29}$$

Assim:

$$Q_r = Q_t + Q_{EC} + Q_{EL} \qquad \textbf{7.30}$$

Convém observar também neste caso que S^2_C; S^2_L serão estimadores justos se tanto H^C_0 como H^L_0 forem verdadeiras, ao passo que S^2_r será um estimador justo sob quaisquer hipóteses sobre o comportamento das médias.

Por outro lado, nota-se que, se

$$Q_t = Q_{EC} + Q_{EL} + Q_r$$

Então

$$\sigma^2 x^2_{n-1} = \sigma^2 x^2_{k-1} + \sigma^2 x^2_{L-1} + \sigma^2 x^2_{(k-1)(L-1)}$$

Nota-se que a soma dos graus de liberdade dos qui-quadrados do segundo membro é igual ao número de g.l. do qui-quadrado particionado, isto é:

$$n - 1 = (k - 1) + (L - 1) + (k - 1)(L - 1)$$
$$n - 1 = k - 1 + L - 1 + kL - k - L + 1$$
$$n - 1 = kL - 1$$

Lembre-se que $k.L = n$, logo: x^2_{k-1}; x^2_{L-1} e $x^2_{(k-1)(L-1)}$ são qui-quadrados independentes e, dessa maneira, pode-se testar a igualdade das médias segundo as colunas e/ou linhas mediante o cálculo de:

para colunas:

$$F^c_{cal} = \frac{S^2_C}{S^2_r} \qquad \textbf{7.31}$$

para linhas:

$$F^L_{cal} = \frac{S^2_L}{S^2_r} \qquad \textbf{7.32}$$

Deve-se, também, salientar que o fato de H^C_0 não ser verdadeira não impede que se teste H^L_0 e vice-versa. O quadro da análise da variância a seguir resume ambos os testes:

QAV				
Fonte de variação	Soma dos Quadrados	G.L	Quadrados Médios	Teste F
Entre colunas	$Q_{EC} = \Sigma_{i=1}^{k}\left[\frac{(\Sigma_i x_{ij})^2}{L}\right] - C$	$k - 1$	$S^2_C = \frac{Q_{EC}}{k-1}$	$F^c_{cal} = \frac{S^2_C}{S^2_r}$ $F^L_{cal} = \frac{S^2_L}{S^2_r}$
Entre linhas	$Q_{EL} = \Sigma_{i=1}^{L}\left[\frac{(\Sigma_i x_{ij})^2}{k}\right] - C$	$L - 1$	$S^2_r = \frac{Q_r}{(k-1)(L-1)}$	
Residual	$Q_r = Q_t - Q_{EC} - Q_{EL}$	$(K - 1)$ $(L - 1)$	$S^2_r = \frac{Q_{EL}}{}$	
Total	$Q_t = \Sigma_{i=1}^{k}\Sigma_{j=1}^{L} x^2_{ij} - C$	$n - 1$		

Regra de Decisão: Fixando certo nível de significância α, tem-se:

1. Se $F^L_{cal} \epsilon RA$, então, aceita-se H_{0_C}; $\mu_i = \mu$ para qualquer i = 1, 2, ..., k, e conclui-se com risco que o fator 1(tratamentos) não causa efeito na variável dependente. Por outro lado, se $F^C_{cal} \epsilon RC$, rejeita-se H_0, concluindo-se pela diferença das médias das colunas e consequente influência do fator sobre a variável analisada.

2. Se $F^L_{cal} \epsilon RA$, então, aceita-se H^L_0: $\mu_j = \mu$ para qualquer j = 1, 2, ..., L, e conclui-se com risco α que o fator 2 (blocos) não causa efeito na variável dependente. Por outro lado, se $F^L_{cal} \epsilon RC$, rejeita-se H_0 , concluindo-se pela

diferença das médias das linhas e consequente influência do fator sobre a variável em estudo.

Encontram-se, a seguir, os principais passos para a efetivação do teste:

1. Dispor os elementos segundo a tabela que segue. Obtendo as somas das colunas e linhas, bem como suas respectivas médias:

Fator 1 (i) / Fator 2 (j)	1	2	...	k	$\sum$	x_j
1	x_{11}	x_{21}	...	x_{k1}		
2	x_{12}	x_{22}	...	x_{k2}		
.	.	.		.		
.	.	.		.		
.	.	.		.		
.	.	.		.		
L	x_{1L}	x_{2L}	...	x_{kL}		
$\sum$						
x_i						

2. Avalia-se a constante

$$C = \frac{(\Sigma_i \Sigma_j^L x_{ij})^2}{n} \qquad \textbf{7.33}$$

Lembre-se de que $n = kL$

3. Calcula-se a variação total

$$Q_t = \Sigma_{i=1}^{k} \Sigma_{j=1}^{L} x_{ij}^2 - C \qquad \textbf{7.34}$$

4. Determina-se a variação entre colunas

$$Q_{EC} = \Sigma_{i=1}^{k} \left[\frac{(\Sigma_i x_{ij})^2}{L} \right] - C \qquad \textbf{7.35}$$

5. Avalia-se a variação entre linhas

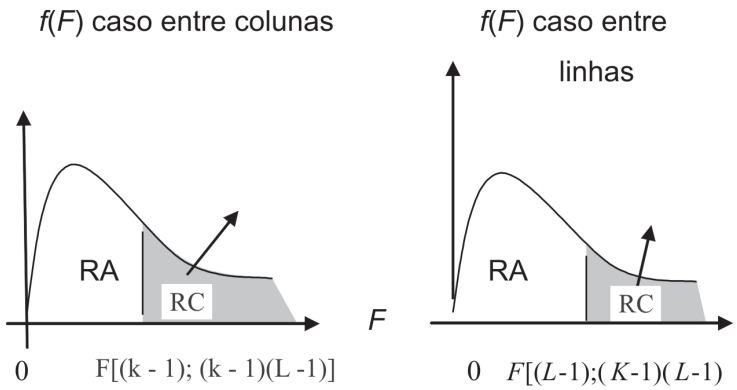
$$Q_{EL} = \Sigma_{i=1}^{L} \left[\frac{(\Sigma_i x_{ij})^2}{k} \right] - C \qquad \textbf{7.36}$$

6. Obtém-se a variação residual por diferença

$$Q_r = Q_t - Q_{EC} - Q_{EL} \qquad \textbf{7.37}$$

7. Constrói-se o quadro de análise de variância, avaliando-se F^{C}_{cal} e F^{L}_{cal} .

8. Determina-se RA e RC por meio da tabela F.



9. Efetuam-se as conclusões pela comparação dos respectivos valores dos F calculados e tabelados.

Aplicação:

Em uma experiência agrícola, foram utilizados 5 diferentes fertilizantes em duas variedades de trigo. A produção está indicada a seguir. Verificar ao nível de 5% se: a) há diferença na produção devido ao fertilizante; b) há diferença na safra devido à variedade do trigo.

Fertilizante	A	B	C	D	E
Variedade 1	54	38	46	50	44
Variedade 2	57	42	45	53	50

Fonte: Fonseca, Jairo Simon da. 2006: pág.271

Solução: Considerando-se o fator 1 como o tipo de fertilizante e o fator 2 como variedade de trigo, constrói-se a tabela, subtraindo 45 a todos os valores observados. Assim:

1.

(i) Fator 1 / (j) Fator 2	A	B	C	D	E	∑
1	9	-7	1	5	-1	7
2	12	-3	0	8	5	22
∑	21	-10	1	13	4	29

2. $C = \frac{(\Sigma_{i=1}^{5}\Sigma_{L=1}^{2}x_{ij})^2}{10} = \frac{(29)^2}{10} = 84{,}1$

3. $Q_t = \Sigma_{i=1}^{5}\Sigma_{j=1}^{2}x_{ij}^2 - C$

$= 9^2 + 12^2 + (-7)^2 + \ldots + 5^2 - 84{,}1 = 314{,}9$

4. $Q_{EC} = \Sigma_{i=1}^{5}\left[\frac{(\Sigma_{i=1}^{2}x_{ij})^2}{2}\right] - C$

$= \frac{(21)^2}{2} + \frac{(-10)^2}{2} + \frac{(1)^2}{2} + \frac{(13)^2}{2} + \frac{(4)^2}{2} - 84{,}1 = 279{,}4$

5. $Q_{EL} = \Sigma_{j=1}^{2}\left[\frac{(\Sigma_{i=1}^{5}x_{ij})^2}{5}\right] - C$

$= \frac{(7)^2}{5} + \frac{(22)^2}{5} - 84{,}1 = 22{,}5$

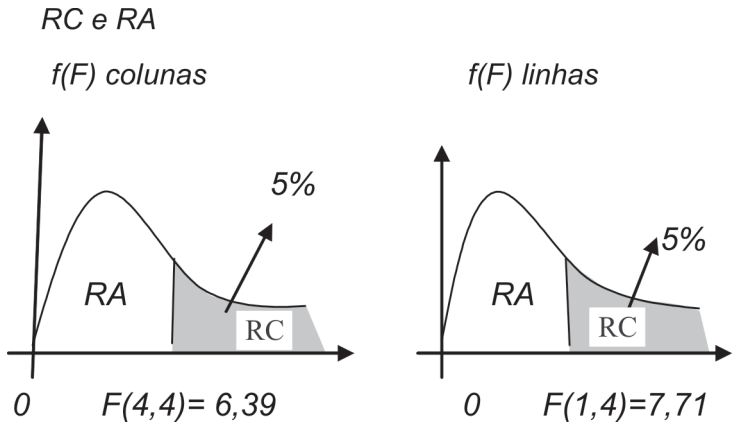
6. $Q_r = Q_t - Q_{EC} - Q_{EL}$

$= 314{,}9 - 279{,}4 - 22{,}5 = 13$

7. QAV

Fonte de Variação	Soma dos Quadrados	G.L.	Quadrados Médios	Teste F
Entre Colunas	279,4	4	69,85	$F^{c}_{cal} = \frac{69{,}85}{3{,}25}$
Entre Linhas	22,5	1	22,5	$F^{L}_{cal} = \frac{22{,}5}{3{,}25}$
Residual	13	4	3,25	
Total	314,9	9		

8. RC e RA



9. Conclusão: Para o primeiro fato, fertilizante, tem-se que $F^{C}_{cal} \in RC$; portanto, rejeita-se $H_0 : \mu_i = \mu$; i = 1, 2, ..., 5, concluindo-se que o tipo de fertilizante tem influência na produção de trigo.

Para o segundo fator, variedade de trigo, tem-se que $F^{L}_{cal} \in RA;$ portanto, aceita-se $H_0 : \mu_j = \mu$; j = 1, 2, concluindo-se que a variedade de trigo não altera a produção.

EXERCÍCIO

1.Quatro analistas determinaram o rendimento de dado processo, obtendo:

Analistas			
1	2	3	4
26	17	36	20
27	20	33	18
24	22	31	17
25	21	29	22
29			21
28			23

Determine:

As médias para os diferentes analistas;

a) A média total;

b) A variação total;

c) A variação entre tratamentos;

d) A variação dentro dos tratamentos (residual);

e) Se há diferença entre as médias, adotando α = 5%.

2. Os dados a seguir representam, em segundos, o tempo gasto por cinco operários para realizar certa tarefa, usando três máquinas diferentes. Considerando α = 5% , verifique se há diferenças entre as máquinas e entre os operários.

Operários	**Máquinas**		
	A	B	C
1	40	59	42
2	39	55	51
3	47	55	45
4	45	50	40
5	52	52	41

Fonte: Fonseca, Jairo Simon da. 2006:, pág. 285

RESPOSTAS DOS EXERCÍCIOS

1º Exercício

1) a) a = 33,da b) até a f) Resposta pessoal
2) n = 400
3) n = 399. Comparando-se os resultado de 2) e 3) verifica-se que uma população de 200.000 dá aproximadamente o resultado de uma população infinita.
4) n = 39

2º Exercício

1) a) 0,4251 b) 0,3023 c) 0,9104
2) a) 380 b) 389
3) x^2_{sup} = 13,4 e x^2_{inf} = 3,49 e
4) -1,1848 e 2,0860
5) $\mu = 1,25$; $\sigma^2 = 1,042$; $\sigma = 1,021$; e abscissas 0,2985 e 3,07

3º Exercício

1) * O intervalo [4,81; 5,59] contém a média populacional com 90% de confiança.
* O intervalo [4,73; 5,67] contém a média populacional com 95% de confiança.
2) * O intervalo [25,76; 28,00] contém a média populacional com 95% de confiança.
* O intervalo [25,94; 27,82] contém a média populacional com 90% de confiança.
3) Os limites de confiança a 80% para a variância são [1,38; 4,86].
4) O intervalo com nível de 90%, será: [165,86; 399,20].
5) O intervalo [0,88; 0,98] contém a proporção com 95% de confiança.
6) O intervalo [16%; 34%] contém a proporção de casas de aluguel com 98% de confiança.

4º Exercício

1) Como t_{cal} = -2,84, rejeita-se H_0, concluindo-se, com risco de 5%, que a média é diferente de 16.
2) Como t_{cal} = -0,72, não se pode rejeitar H_0, ao nível de 5% nos dois testes.
3) a) S^2 = 0,07 mg^2
b) Como x^2_{cal} = 0,49, rejeita-se H_0, concluindo-se com risco de 5% que a variância é menor que 1.
4) Como x^2_{cal} = 5,57, não se pode rejeitar a hipótese de que a variância populacional é 4, ao nível de 1%.
5) Como z_{cal} = 0,89, não se pode rejeitar a hipótese de que a proporção de eleitores democratas é 50%, ao nível de 5%.
6) Como z_{cal} = 4,47 ,rejeita-se H_0 , concluindo-se, com risco de 4%, que a proporção é diferente de 0,5.

5º Exercício

1) Como x^2_{cal} = 2, Não se pode rejeitar a honestidade moeda, ao nível de 10%.
2) Como x^2_{cal} = 7,296 , não se pode rejeitar a hipótese de que o número de livros emprestados independe do dia da semana, ao nível de 1%.
3) Como x^2_{cal} = 6,11, não se pode rejeitar a hipótese de que as variáveis sejam independentes, ao nível de 5%.
4) Como z_{cal} = -0,75, não de pode rejeitar a hipótese de que não houve diferença dos pesos, ao nível de 2,5%.
5) Como z_{cal} = 0,43, não se pode rejeitar a hipótese de igualdade das médias, ao nível de 1%.

6º Exercício

1) a) r_{XY} = -0,711 b) r_{XY} = -0,736 c) r_{XY} = -0,947
2) a) a = 103,57 b = 0,1048 e Y = 103,57 + 0,1048x
b) R^2 = 89,4%

7ª Exercício

1)
a) x_1 = 26,50, x_2 = 20, x_3 = 32,25 e x_4 = 20,17

b) x = 24,45

c) 542,95

d) 457,37

e) 85,08

f) Há diferença. $F_{cal} = 28,99$

2) $F^{C}_{cal} = 6,43$ e $F^{L}_{cal} = 0,29$; só há entre as máquinas.

Referências

COSTA NETO, Pedro de Oliveira. **Estatística**. 2. ed. São Paulo: Blucher, 2002.

CRESPO, Antonio Aenout. **Estatística fácil**. 17. São Paulo: Saraiva, 2002.

FONSECA, Jairo Simon da. **Curso de estatística**. 10. ed. Reimp. Martins, Gilberto de Andrade. São Paulo: Atlas, 2006.

LAPPONI, Juan Carlos. **Estatística usando Excel**. 4. ed. Reimpressão. Rio de Janeiro: Elsevier, 2005.

MEYER, Paul L. **Probabilidade:** aplicações à Estatística. 2. ed. Rio de Janeiro: LTC, 2000.

MORETIN, Pedro Alberto. **Estatística básica**. 5. ed. São Paulo: Saraiva, 2006.

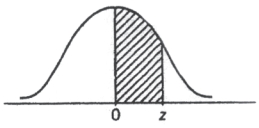
SPIEGEL, M. R. **Probabilidade e estatística**. 1. ed. São Paulo: McGraw-Hill, 2001.

TOLEDO, Geraldo Luciano. **Estatística básica.** 2. ed. São Paulo: Atlas, 1985.

ANEXOS
TABELAS E ESTATÍSTICAS

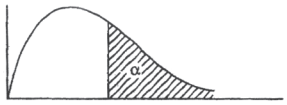
Tabela 1. Áreas de uma distribuição normal padrão

Cada casa na tabela dá a proporção sob a curva inteira entre $z = 0$ e um valor positivo de z. As áreas para os valores de z negativos são obtidas por simetria.



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2703	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

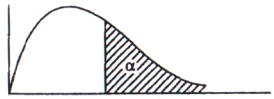
Tabela 2. Distribuição de χ^2



φ \ α	0,995	0,990	0,975	0,950	0,900	0,750	0,500	0,250	0,100	0,050	0,025	0,010	0,005
1	0,0000	0,0002	0,0010	0,0039	0,0158	0,102	0,455	1,32	2,71	3,84	5,02	6,63	7,88
2	0,0100	0,0001	0,0506	0,103	0,211	0,575	1,39	2,77	4,61	5,99	7,38	9,21	10,6
3	0,0717	0,115	0,216	0,352	0,584	1,021	2,37	4,11	6,25	7,81	9,25	11,3	12,8
4	0,207	0,297	0,484	0,711	1,06	1,92	3,36	5,39	7,78	9,49	11,1	13,3	14,9
5	0,412	0,554	0,831	1,15	1,61	2,67	4,35	6,63	9,24	11,1	12,8	15,1	16,7
6	0,676	0,872	1,24	1,64	2,20	3,45	5,35	7,84	10,6	12,6	14,4	16,8	18,5
7	0,989	1,24	1,69	2,17	2,83	4,25	6,35	9,04	12,0	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	3,49	5,07	7,34	10,2	13,4	15,5	17,5	20,1	22,0
9	1,73	2,09	2,70	3,33	4,17	5,90	8,34	11,4	14,7	16,9	19,0	21,7	23,6
10	2,16	2,56	3,25	3,94	4,87	6,74	9,34	12,5	16,0	18,3	20,5	23,2	25,2
11	2,60	3,05	3,82	4,57	5,58	7,58	10,3	13,7	17,3	19,7	21,9	24,7	26,8
12	3,07	3,57	4,40	5,23	6,30	8,44	11,3	14,8	18,5	21,0	23,3	26,2	28,3
13	3,57	4,11	5,01	5,89	7,04	9,30	12,3	16,0	19,8	22,4	24,7	27,7	29,8
14	4,07	4,66	5,63	6,57	7,79	10,2	13,3	17,1	21,1	23,7	26,1	29,1	31,3
15	4,60	5,23	6,23	7,26	8,55	11,0	14,3	18,2	22,3	25,0	27,5	30,6	32,8
16	5,14	5,80	6,91	7,96	8,31	11,9	15,3	19,4	23,5	26,3	28,4	32,0	34,3
17	5,70	6,41	7,56	8,67	10,1	12,8	16,3	20,5	24,8	27,6	30,2	53,4	35,7
18	6,26	7,01	8,23	9,39	10,9	13,7	17,3	21,6	26,0	28,9	31,5	34,8	37,2
19	6,84	7,63	8,91	10,1	11,7	14,6	18,3	22,7	27,2	30,1	32,9	36,2	38,6
20	7,43	8,26	9,59	10,9	12,4	15,5	19,3	23,8	28,4	31,4	34,2	37,6	40,0
21	8,03	8,90	10,3	11,6	13,2	16,3	20,3	24,9	29,6	32,7	35,5	38,9	41,4
22	8,64	9,54	11,0	12,3	14,0	17,2	21,3	26,0	30,8	33,9	36,8	40,5	42,8
23	9,26	10,2	11,7	13,1	14,8	18,1	22,3	27,1	32,0	35,2	38,1	41,6	44,2
24	9,89	10,9	12,4	13,8	15,7	19,0	23,3	28,2	33,1	36,4	39,4	43,0	45,6
25	10,5	11,5	13,1	14,6	16,5	19,9	24,3	29,3	34,4	37,7	40,6	44,3	46,9
26	11,2	12,2	13,8	15,4	17,3	20,8	25,3	30,4	35,6	38,9	41,9	45,6	48,3
27	11,8	12,9	14,6	16,2	18,1	21,7	26,3	31,5	36,7	40,1	43,2	47,0	49,6
28	12,5	13,6	15,3	16,9	18,9	22,7	27,3	32,6	37,9	41,3	44,5	48,3	51,0
29	13,1	14,3	16,0	17,7	19,8	23,6	28,3	33,7	39,1	42,6	45,7	49,6	52,5
30	13,8	15,0	16,8	18,5	20,6	24,5	29,3	34,8	40,3	43,8	47,0	50,9	53,7

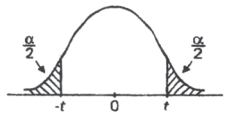
Para $\varphi > 30$ usar a aproximação: $\chi^2_x = \frac{1}{2}\left[\pm Z_\alpha + \sqrt{2\varphi - 1}\right]^2$

Tabela 3. Distribuição de F de Snedecor $\alpha = 5\%$



φ_1 / φ_2	1	2	3	4	5	6	7	8	9	10	20	30	120	∞
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	248,0	250,1	253,3	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,45	19,46	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,66	8,62	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,80	5,75	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,56	4,50	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,87	3,81	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,44	3,38	3,27	2,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,15	3,08	2,97	2,92
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	2,94	2,86	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	4,07	3,02	2,98	2,77	2,70	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,65	2,57	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,90	2,85	2,80	2,75	2,54	2,47	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,46	2,38	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,39	3,31	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,33	2,25	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,28	2,19	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,23	2,15	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,19	2,11	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,16	2,07	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,12	2,04	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,10	2,01	1,87	1,81
22	4,30	3,44	4,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,07	1,98	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,05	1,96	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,03	1,94	1,79	1,73
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	1,93	1,84	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,21	2,18	2,12	2,08	1,84	1,74	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,75	1,65	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,66	1,55	1,35	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,57	1,46	1,22	1,00

Tabela 4. Distribuição *t* de Student



φ \ α	0,50	0,25	0,10	0,05	0, 025	0,01	0,005
1	1,00000	2,4142	6,3138	12.706	25.542	63.657	127,32
2	0,81650	1,6036	2,9200	4,3127	6,2053	9,9248	14,089
3	0,76489	1,4226	2,3534	3,1825	4,1765	5,8409	7,4533
4	0,74070	1,3444	2,1318	2,7764	3,4954	4,6041	5,5976
5	0,72669	1,3009	2,0150	2,5706	3,1634	4,0321	4,7733
6	0,71756	1,2733	1,9432	2,4469	2,9687	3,7074	4,3168
7	0,71114	1,2543	1,8946	2,3646	2,8412	3,4995	4,0293
8	0,70639	1,2403	1,8595	2,3060	2,7515	3,3554	3,8325
9	0,70272	1,2297	1,8331	2,2622	2,6850	3,2498	3,6897
10	0,69981	1,2213	1,8125	2,2281	2,6338	3,1693	3,5814
11	0,69745	1,2145	1,7959	2,2010	2,5931	3,1058	3,4966
12	0,69548	1,2089	1,7823	2,1788	2,5600	3,9545	3,4284
13	0,69384	1,2041	1,7709	2,1604	2,5326	3,0123	3,3725
14	0,692	1,2001	1,7613	2,1448	2,5096	2,9768	3,3257
15	0,69120	1,1967	1,7530	2,1315	2,4899	2,9467	3,2860
16	0,69013	1,1937	1,7459	2,1199	2,4729	2,9208	3,2520
17	0,68919	1,1910	1,7396	2,1098	2,4581	2,8982	3,2225
18	0,68837	1,1887	1,7341	2,1009	2,4450	2,8784	3,1966
19	0,68763	1,1866	1,7291	2,0930	2,4334	2,8609	3,1737
20	0,68696	1,1848	1,7247	2,0860	2,4231	2,8453	3,1534
21	0,68635	1,1831	1,7207	2,0796	2,4138	2,8314	3,1352
22	0,68580	1,1816	1,7171	2,0739	2,4055	2,8188	3,1188
23	0,68531	1,1802	1,7139	2,0687	2,3979	2,8073	3,1040
24	0,68485	1,1789	1,7109	2,0639	2,3910	2,7969	3,0905
25	0,68443	1,1777	1,7081	2,0595	2,3846	2,7874	3,0782
26	0,68405	1,1766	1,7056	2,0555	2,3788	2,7787	3,0669
27	0,68370	1,1757	1,7033	2,0518	2,3734	2,7707	3,0565
28	0,68335	1,1748	1,7011	2,0484	2,3685	2,7633	3,0469
29	0,68304	1,1739	1,6991	2,0452	2,3638	2,7564	3,0380
30	0,68276	1,1731	1,6973	2,0423	2,3596	2,7500	3,0298
40	0,68066	1,1673	1,6839	2,0211	2,3289	2,7045	2,9712
60	0,67862	1,1616	1,6707	2,0003	2,2991	2,6603	2,9146
120	0,67656	1,1559	1,6577	1,9799	2,2699	2,6174	2,8599
∞	0,67449	1,1503	1,6449	1,9600	2,2414	2,5758	2,8070

Tabela 5. Dígitos aleatórios

03991	10461	93716	16894	98953	73231	39528	72484	82474	25593
38555	95554	32886	59780	09958	18065	81616	18711	53342	44276
17546	73704	92052	46215	15917	06253	07586	16120	82641	22820
32643	52861	95819	06831	19640	99413	90767	04235	13574	17200
69572	68777	39510	35905	85244	35159	40188	28193	29593	88627
24122	66591	27699	06494	03152	19121	34414	82157	86887	55087
61196	30231	92692	61773	22109	78508	63439	75363	44989	16822
30532	21704	10274	12202	94205	20380	67049	09070	93399	45547
03788	97599	75867	20717	82037	10268	79495	04146	52162	90286
48228	63379	85783	47619	87481	37220	91704	30552	04737	21031
88618	19161	41290	67312	74857	15957	48545	35247	18619	13674
71299	23853	05870	01119	92784	26340	75122	11724	74627	73707
27954	58909	82444	99005	04921	73701	92904	13141	32392	19763
80863	00514	20247	81759	45197	25332	69902	63742	78464	22501
33564	60780	48460	85558	15191	18782	94972	11598	62095	36787
90899	75754	60833	25983	01291	41349	19152	00023	12302	80783
78038	70267	43529	06318	38384	74761	36024	00867	76378	41605
55986	66485	88722	56736	66164	49431	94458	74284	05041	49807
87539	08823	94813	31900	54155	83436	54158	34243	46978	35482
16818	60311	74457	90561	72848	11834	75051	93029	47665	64382
34677	58300	74910	64345	19325	81540	60365	94653	35075	33949
45305	07521	61318	31855	14413	70951	83799	42402	56623	34442
59747	67277	76503	34513	39663	77544	32960	07405	36409	83232
16520	69676	11654	99893	02181	68161	19322	53845	57620	52606
68652	27376	92852	55866	88448	03584	11220	94747	07399	37408

(*continuação*)

(*continuação*)

79375	95220	01159	63267	10622	48391	31751	57260	68980	05339
33521	26665	55823	47641	86225	31704	88492	99382	14454	04504
59589	49067	66821	41575	49767	04037	30934	47744	07481	83828
20554	91409	96277	48257	50816	97616	22888	48893	27499	98748
59404	72059	43947	51680	43852	59693	78212	16993	35902	91386
42614	29297	01918	28316	25163	01889	70014	15021	68971	11403
34994	41374	70071	14736	65251	07629	37329	33295	18477	65622
99385	41600	11133	07586	36815	43625	18637	37509	14707	93997
66497	68646	78138	66559	64397	11692	05327	82162	83745	22567
48509	23929	27482	45476	94515	25624	95096	67946	16930	33361
15470	48355	88651	22596	83761	60873	43253	84145	20368	07126
20094	98977	74843	93413	14387	06345	80854	09279	41196	37480
73788	06533	28597	20405	51321	92246	80088	77074	66919	31678
60530	45128	74022	84617	72472	00008	80890	18002	35352	54131
44372	15486	65741	14014	05466	55306	93128	18464	79982	68416
18611	19241	66083	24653	84609	58232	41849	84547	46850	52323
58319	15997	08355	60860	29735	47762	46352	33049	69248	93460
61199	67940	55121	29281	59076	07936	11087	96294	14013	31792
18627	90872	00911	98936	76355	93779	52701	08337	56303	87315
00441	58997	14060	40619	29549	69616	57275	36898	81304	48585
32624	68691	14845	46672	61958	77100	20857	73156	70284	24326
65961	73488	41839	55382	17267	70943	15633	84924	90415	93614
20288	34060	39685	23309	10061	68829	92694	48297	39904	02115
59362	95938	74416	53166	35208	33374	77613	19019	88152	00080
99782	93478	53152	67433	35663	52972	38688	32486	45134	63545

Minicurrículo

Juarez Rodrigues Martins

Especialista em Matemática (2001) e em Estatística (2008) pela Universidade Federal do Piauí. Graduou-se em Biologia pela Universidade Estadual do Piauí (1992) e em Matemática pela Universidade Federal do Piauí (1995). Foi professor substituto na Universidade Federal do Piauí, durante quatro anos, nos períodos de março de 2003 a março de 2005, e abril de 2007 a abril de 2009. Atuou como professor do Ensino Médio da rede pública estadual do Piauí e foi professor da rede particular de ensino de Teresina. Atualmente é professor efetivo da Universidade Federal do Piauí.

CONTATO

Email: martins-juarez@bol.com.br.
Web site: http://www.famat.ufu.br/prof/marcelo/exercicios.htm

GOVERNO FEDERAL BRASIL PAÍS RICO É PAÍS SEM POBREZA

Ministério da Educação

CAPES

UAB UNIVERSIDADE ABERTA DO BRASIL

UFPI

www.uapi.ufpi.br