

ESTATÍSTICA E PROBABILIDADE

LICENCIATURA EM MATEMÁTICA

Ministério da Educação - MEC
Coordenação de Aperfeiçoamento
de Pessoal de Nível Superior
Universidade Aberta do Brasil
Instituto Federal de Educação,
Ciência e Tecnologia do Ceará



MINISTÉRIO DA EDUCAÇÃO
Universidade Aberta do Brasil
Instituto Federal de Educação, Ciência e Tecnologia do Ceará
Diretoria de Educação a Distância

Licenciatura em Matemática
Estatística e Probabilidade

Paulo Maia Ferreira

Fortaleza, CE
2012

CRÉDITOS

Presidente

Dilma Vana Rousseff

Ministro da Educação

Fernando Haddad

Secretário da SEED

Luís Fernando Massonetto

Diretor de Educação a Distância

Celso Costa

Reitor do IFCE

Cláudio Ricardo Gomes de Lima

Pró-Reitor de Ensino

Gilmar Lopes Ribeiro

Diretora de EAD/IFCE e Coordenadora UAB/IFCE

Cassandra Ribeiro Joye

Vice-Coordenadora UAB

Régia Talina Silva Araújo

Coordenador do Curso de Tecnologia em Hotelaria

José Solon Sales e Silva

Coordenador do Curso de Licenciatura em Matemática

Priscila Rodrigues de Alcântara

Elaboração do conteúdo

Paulo Maia Ferreira

Colaboradora

Marília Maia Moreira

Equipe Pedagógica e Design Instrucional

Ana Cláudia Uchôa Araújo

Andréa Maria Rocha Rodrigues

Carla Anaíle Moreira de Oliveira

Cristiane Borges Braga

Eliana Moreira de Oliveira

Gina Maria Porto de Aguiar

Glória Monteiro Macedo

Iraci de Oliveira Moraes Schmidlin

Irene Moura Silva

Isabel Cristina Pereira da Costa

Jane Fontes Guedes

Karine Nascimento Portela

Lívia Maria de Lima Santiago

Lourdes Losane Rocha de Sousa

Luciana Andrade Rodrigues

Maria Irene Silva de Moura

Maria Vanda Silvino da Silva

Marília Maia Moreira

Maria Luiza Maia

Saskia Natália Brígido Batista

Equipe Arte, Criação e Produção Visual

Ábner Di Cavalcanti Medeiros

Benghson da Silveira Dantas

Cícero Felipe da Silva Figueiredo

Elson Felipe Gonçalves Mascarenha

Germano José Barros Pinheiro

Gilvandenys Leite Sales Júnior

José Albério Beserra

José Stelio Sampaio Bastos Neto

Lucas de Brito Arruda

Marco Augusto M. Oliveira Júnior

Equipe Web

Benghson da Silveira Dantas

Fabrice Marc Joye

Hanna França Menezes

Herculano Gonçalves Santos

Luiz Bezerra de Andrade Filho

Lucas do Amaral Saboya

Ricardo Werlang

Samantha Onofre Lóssio

Tibério Bezerra Soares

Revisão Textual

Aurea Suely Zavam

Nukácia Meyre Araújo de Almeida

Revisão

Antônio Carlos Marques Júnior

Débora Liberato Arruda Hissa

Saulo Garcia

Logística

Francisco Roberto Dias de Aguiar

Secretários

Breno Giovanni Silva Araújo

Francisca Venâncio da Silva

Auxiliar

Ana Paula Gomes Correia

Bernardo Matias de Carvalho

Charlene Oliveira da Silveira

Isabella de Castro Britto

Nathália Rodrigues Moreira

Virgínia Ferreira Moreira

Vivianny de Lima Santiago

Wagner Souto Fernandes

Catálogo na Fonte: Islânia Fernandes Araújo (CRB 3 - Nº 917)

F383e Ferreira, Paulo Maia.

Estatística e Probabilidade / Paulo Maia Ferreira; Coordenação Cassandra Ribeiro Joye. - Fortaleza: UAB/IFCE, 2012.
208p. : il. ; 27cm.

ISBN 978-85-63953-99-5

1. CÁLCULO PROBABILÍSTICO. 2. ESTATÍSTICA DESCRITIVA.
3. INFERÊNCIA ESTATÍSTICA. I. Joye, Cassandra Ribeiro. (Coord.). II.
Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE III.
Universidade Aberta do Brasil – UAB. IV. Título.

CDD – 519

Apresentação 8
Referências 208
Currículo 209

SUMÁRIO

AULA 1 Introdução aos cálculos de probabilidade 9

Tópico 1 Teoria dos cálculos das probabilidades 10

Tópico 2 Variáveis aleatórias 18

AULA 2 Distribuições discretas e contínuas de probabilidade 24

Tópico 1 Distribuições discretas de probabilidades 25

Tópico 2 Distribuições contínuas de probabilidades 31

AULA 3 Séries estatísticas 43

Tópico 1 Principais tipos de séries 44

Tópico 2 Representação das séries estatísticas através de tabelas 48

Tópico 3 Representação das séries através de gráficos 52

AULA 4 Estatística descritiva - medidas de posição e medidas de dispersão 57

Tópico 1 Distribuições de frequência 58

Tópico 2 Medidas de posição ou de tendência central 63

Tópico 3 Medidas de dispersão 75

AULA 5 **Miscelânea de exercícios resolvidos 82**

- Tópico 1 Exercícios sobre cálculos de probabilidade 83
- Tópico 2 Exercícios: Modelos de distribuições probabilísticas 90
- Tópico 3 Exercícios: Séries estatísticas 95
- Tópico 4 Exercícios: Estatística Descrita 99

AULA 6 **Correlação e regressão 106**

- Tópico 1 Correlação 107
- Tópico 2 Regressão 114

AULA 7 **Estimações 119**

- Tópico 1 Estimações de médias populacionais 120
- Tópico 2 Estimações de proporções populacionais 130

AULA 8 **Testes estatísticos não-paramétricos 134**

- Tópico 1 Testes não-paramétricos para uma variável 135
- Tópico 2 Testes não-paramétricos para duas variáveis 145

AULA 9

Tópico 1

Tópico 2

Testes estatísticos não-paramétricos 153

Testes estatísticos envolvendo um único parâmetro populacional 154

Testes estatísticos envolvendo dois parâmetros populacionais 166

AULA 10

Tópico 1

Tópico 2

Tópico 3

Tópico 4

Miscelânea de exercícios resolvidos 175

Exercícios sobre correlação e regressão 176

Exercícios sobre estimações 180

Exercícios sobre testes não-paramétricos 186

Exercícios sobre testes paramétricos 199

APRESENTAÇÃO

Olá aluno(a)!

A disciplina de Estatística e Probabilidade terá três focos temáticos principais: cálculos das probabilidades, estatística descritiva e inferência estatística. Nos cálculos das probabilidades, realizaremos um estudo introdutório sobre o assunto. Você conhecerá os principais conceitos probabilísticos, os principais teoremas da probabilidade e as variáveis aleatórias. Iremos solucionar problemas com o auxílio dos principais modelos de distribuições discretas e contínuas de probabilidade.

No tocante à estatística descritiva, primeiramente, conheceremos os principais tipos de séries estatísticas e suas formas de representações tabulares e gráficas. Em seguida, você verá que iremos organizar dados em tabelas de distribuições de frequências, diferenciando estes tipos de frequências. Encontraremos também os principais tipos de medidas estatísticas de posição, de dispersão e de assimetria, para posteriormente realizarmos leituras descritivas dos resultados obtidos.

Por fim, acerca da inferência estatística, com base nas noções de probabilidade e do conhecimento da estatística descritiva, você aprenderá como explorar resultados populacionais. Neste foco, teremos contato com os conceitos de correlação e regressão, realizaremos estimações de parâmetros populacionais desconhecidos e aplicaremos testes estatísticos paramétricos e não-paramétricos para posteriores tomadas de decisões.

Esperamos que você compreenda os conceitos que permeiam a ideia de Estatística e Probabilidade e que esses estudos tenham utilidade em suas vidas, assim como a Matemática, como um todo, tem a sua importância na vida de todos nós. Desejo a você um bom aprendizado e um ótimo desempenho no curso.

Paulo Maia Ferreira

AULA 1

Introdução aos cálculos de probabilidade

Olá caro(a) aluno(a),

Percebendo as divergências das concepções de ensino de Probabilidade e de Estatística que existem entre a real proposta destas disciplinas e o que encontramos na maioria dos livros didáticos, elaboramos este material, para que você, aluno, desenvolva o raciocínio estatístico e probabilístico de maneira integrada, sendo capaz de organizar e analisar informações, formulando argumentos para conseguir realizar inferências convincentes.

Desta forma, iniciaremos com um bom embasamento de raciocínio probabilístico, prosseguiremos com as principais medidas estatísticas a serem usadas como ferramentas essenciais e finalizaremos com a prática de algumas inferências estatísticas.

Então vamos à aula?

Objetivos

- Compreender os principais conceitos de espaço amostral e eventos, bem como os principais tipos de eventos
- Estudar os principais teoremas conceituais da probabilidade

TÓPICO 1

Teoria dos cálculos das probabilidades

OBJETIVOS

- Usar dos conhecimentos de espaço amostral e eventos para uma melhor compreensão da definição de probabilidade
- Resolver situações probabilísticas com base nos principais teoremas de probabilidade

Para estudarmos os fenômenos ocorridos em nosso cotidiano, utilizaremos a ideia de Experimentos Aleatórios (casuais), ou seja, aqueles que não sabemos o resultado *a priori*. Neste tópico, iremos conhecer alguns resultados da teoria de probabilidade, os quais irão contribuir para uma melhor compreensão de todas as aplicações de situações probabilísticas.



VOCÊ SABIA?

Você sabia que, segundo Morettin (1999, p. 1), na natureza, encontramos dois tipos de experimentos: os determinísticos e aleatórios. No entanto, você utilizará somente os Experimentos Aleatórios para o seu estudo.

1.1 CONCEITOS IMPORTANTES

Os conceitos abordados em probabilidade que devem ser estudados por você, aluno e futuro professor de Matemática, serão incorporados ao seu conteúdo intelectual e servirão de apoio a estudos posteriores. Sendo assim, como já definimos o conceito de experimento aleatório, outros conceitos, como espaço amostral e evento, devem ser também conhecidos.

1.1.1 ESPAÇO AMOSTRAL

Definição₁: Chamaremos de espaço amostral (S) o conjunto de todos os possíveis resultados de um experimento aleatório.

Exemplo:

Seja o seguinte experimento aleatório:

E: "lançar duas moedas e observar o resultado".

Assim, teremos o seguinte **espaço amostral**:

$S = \{(k,k) (k,c) (c,k) (c,c)\}$; em que $k = \text{cara}$ e $c = \text{coroa}$.

1.1.2 EVENTO

Definição₂: É um conjunto de resultados de um experimento. Em termos de conjunto, é um subconjunto do espaço amostral S .

ATENÇÃO!

Você se atente que, segundo Hazzan (1993, p.92), um evento, em geral, é indicado por letras maiúsculas de nosso alfabeto: A, B, C, D, E,..., X, Y, Z.

Exemplo:

Seja o seguinte experimento aleatório:

E: "lançar um dado e observar o resultado".

Podemos então definir o seguinte evento, o qual nomearemos com a letra A.

A: "ocorrer múltiplo de 3". Assim, teremos $A = \{3, 6\}$.

Percebendo então que, em termo de conjunto, evento é um subconjunto de um espaço amostral, então, neste exemplo, poderíamos formar um total de $2^6 = 64$ eventos.

1.1.2.1 EVENTOS MUTUAMENTE EXCLUSIVOS

Definição₃: Dois eventos são mutuamente exclusivos quando não podem ocorrer simultaneamente. Assim, se A e B são mutuamente exclusivos, então $A \cap B = \emptyset$.

Exemplo:

Veja o seguinte experimento:

E: "Jogar um dado e observar o resultado".

Assim, teremos o seguinte espaço amostral: $S = \{1, 2, 3, 4, 5, 6\}$.

Observem os seguintes eventos:

A: "Ocorrer múltiplo de dois", ou seja, $A = \{0, 2, 4, 6\}$.

e

B: "Ocorrer nº ímpar", em outras palavras, $B = \{1, 3, 5\}$.

Então $A \cap B = \emptyset$. Assim são dois eventos mutuamente exclusivos.

1.1.2.2 EVENTOS INDEPENDENTES

Definição₄: Dizemos que dois eventos A e B são independentes quando a probabilidade da ocorrência simultânea dos dois eventos for igual ao produto entre as probabilidades. Ou seja, se dois eventos A e B são independentes, então temos: $P(A \cap B) = P(A) \cdot P(B)$.

Exemplo:

Veja o seguinte experimento aleatório.

E: Retirar uma carta de um baralho. Assim teríamos o seguinte espaço amostral:

$$S = \{A_o, \dots, R_o; A_c, \dots, R_c; A_p, \dots, R_p; A_e, \dots, R_e\}$$

Podemos então formar dois eventos a partir deste experimento:

A: "Retirar um dez do baralho"

$$A = \{10_o, 10_c, 10_p, 10_e\}$$

B: "Retirar uma carta de copa"

$$B = \{A_c, \dots, R_c\}$$

Ainda, $A \cap B$: "Retirar um 10 de copa",

ou seja, $A \cap B = \{10_c\}$. Assim, teríamos

$$P(A) = \frac{4}{52} \text{ e } P(B) = \frac{13}{52}, \quad \text{ou seja,}$$

$$P(A \cap B) = \frac{4}{52} \cdot \frac{13}{52} = \frac{1}{52}.$$

Então concluímos que A e B são eventos independentes.



VOCÊ SABIA?

Para sua informação, os índices o, c, p, e que estão na sequência: $A_o, \dots, R_o; A_c, \dots, R_c; A_p, \dots, R_p; A_e, \dots, R_e$, são respectivamente o de ouros, c de copas, p de paus, e de espada, ou seja, todos os naipes de um baralho. Cabe ainda ressaltar que o baralho possui 52 cartas, distribuídas em 4 naipes (ou grupos): espadas, paus, copas e ouros. Cada naipe possui 13 cartas, sendo elas um ás (representado pela letra A), todos os números de 2 a 10, e três figuras: o valete, representado pela letra J (do inglês jack), a dama (também chamada de rainha), representada pela letra Q (do inglês queen), e o rei, com a letra K (do inglês king). Nos exercícios de probabilidade, aparecem muito problemas envolvendo os naipes de baralho.



ATENÇÃO!

O conceito de Eventos Independentes será mais bem compreendido quando você (aluno) tiver tido acesso ao conceito de probabilidade, que será definido logo a seguir.

1.2 DEFINIÇÃO DE PROBABILIDADE

Definição₃: Seja E um experimento aleatório. A probabilidade de um evento A, associada a este experimento e denotada por $P(A)$ será uma função definida em seu espaço amostral S, que irá associar ao evento um n° real, que irá satisfazer aos axiomas:

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

ii) $P(S)=1$.

em que $P(A) = \frac{\text{Nº de casos favoráveis ao evento A}}{\text{Nº total de casos do espaço amostral S}}$, ou ainda $P(A) = \frac{N(A)}{N(S)}$.

Para entender melhor esse conceito, vamos resolver um exercício. Neste exercício, iremos usar conceitos de Análise Combinatória, mais especificamente com o conceito de Combinação visto na disciplina de Matemática Básica 2 (semestre II). Qualquer dúvida leia o material referente a essa disciplina.

EXERCÍCIO RESOLVIDO I

Em uma sala de aula com 20 alunos, há apenas 4 alunos que nunca foram reprovados. Selecionando aleatoriamente dois estudantes desta sala, qual a probabilidade de que:

a) Ambos nunca tenham sido reprovados.

b) Ambos já tenham sofrido reprovação.

c) Ao menos um já tenha sido reprovado.

Solução do item (a):

Definiremos o seguinte evento A: "ambos nunca foram reprovados". Assim temos o seguinte cálculo:

$$P(A) = \frac{N(A)}{N(S)} = \frac{C_{4,2}}{C_{20,2}} = \frac{6}{190}$$

Solução do item (b):

Já no item (b), temos o evento B: "Ambos já foram reprovados" e, usando a definição, obteremos

$$P(B) = \frac{N(B)}{N(S)} = \frac{C_{16,2}}{C_{20,2}} = \frac{120}{190}$$

Solução do item (c):

No entanto ainda temos o item (c), o qual temos o evento C : "Ao menos um já sofreu reprovação". Assim sendo, podemos desmembrar este evento em dois outros eventos:

C_1 : Um aluno nunca sofreu reprovação.

C_2 : Dois alunos nunca sofreram reprovação.

Sabendo que C_1 e C_2 são eventos mutuamente exclusivos, então $(C_1 \cap C_2) = \emptyset$. Portanto, iremos perceber que $P(C) = P(C_1 \cup C_2) = P(C_1) + P(C_2)$. Assim sendo, devemos calcular $P(C_1)$ e $P(C_2)$, ou seja,

$$P(C_1) = \frac{N(C_1)}{N(S)} = \frac{C_{4,1} \cdot C_{16,1}}{C_{20,2}} = \frac{4 \cdot 16}{190} = \frac{64}{190} \quad \text{e} \quad P(C_2) = \frac{6}{190}.$$

$$\text{Logo, o resultado para } P(C) \text{ é } P(C) = \frac{64}{190} + \frac{6}{190} = \frac{70}{190}.$$

Outra forma de construirmos a solução do item (c) é que, se percebermos que a probabilidade pedida neste evento é o complemento da probabilidade pedida no evento, então diríamos que $P(C) = P(B^c) = 1 - P(B) = 1 - \frac{120}{190} = \frac{70}{190}$, chegando assim a solução desejada.

A seguir, veremos os teoremas relacionados à probabilidade, alguns dos quais iremos demonstrar.

1.3 PRINCIPAIS TEOREMAS

Teorema₁: Dado um experimento aleatório, temos o seguinte:

Se \emptyset é o conjunto vazio, então $P(\emptyset) = 0$.

Se A^c é o complementar do evento A , então $P(A^c) = 1 - P(A)$.

Demonstração

O item (i) será deixado como exercício para você fazer. Mas, iremos demonstrar o item (ii). Sendo S o espaço amostral, temos que

$S = A \cup A^c$, no qual esta união é disjunta, pois $A \cap A^c = \emptyset$. Assim,

$$P(S) = P(A) + P(A^c) \Rightarrow P(A^c) = P(S) - P(A) = 1 - P(A).$$

Teorema₂: Se A e B são dois eventos quaisquer, então:
 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Demonstração

Como $A \cup B = A \cup (B - A)$ e $A \cap (B - A) = \emptyset$, então:

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

Também sabemos que $B = (B - A) \cup (A \cap B)$ e $(B - A) \cap (A \cap B) = \emptyset$, portanto

$$P(B) = P(B - A) + P(A \cap B). \text{ Assim, com base nesses dois resultados, temos que}$$

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

Teorema₃: Seja A e B eventos, e A^c e B^c seus complementos, logo temos

$$\text{i) } P(A^c \cup B^c) = P(A \cap B)^c$$

$$\text{ii) } P(A^c \cap B^c) = P(A \cup B)^c$$

Demonstração

Deixaremos para você, caro aluno, essa demonstração como exercício.

Sugestão: Utilize os Diagramas de Venn.

1.4 PROBABILIDADE CONDICIONAL,

TEOREMA DO PRODUTO E O TEOREMA DE BAYES

1.4.1 PROBABILIDADE CONDICIONAL

Definição₆: A probabilidade de ocorrer um evento A, dado que certo evento B ocorreu, será dada por $P(A/B) = \frac{P(A \cap B)}{P(B)}$.

EXERCÍCIO RESOLVIDO 2

Seja o experimento E: “lançar um dado e observar o resultado” e o evento A: “sair o nº 2”; encontre a probabilidade do evento A ocorrer, condicionada a probabilidade de um outro evento B: “sair um nº par”, também ocorrer.

Solução:

Assim teríamos, $P(A) = 1/6$; ou seja, ocorrer o nº 2 ao lançarmos um dado. Assim a probabilidade do evento A ocorrer, condicionada ao evento B também ocorrer, teríamos o espaço amostral reduzido somente aos números pares, sendo agora a probabilidade representada da seguinte forma:

$$P(A/B) = 1/3$$

$$P(A/B) = 1/3.$$

1.4.2 TEOREMA DO PRODUTO

Teorema₄: Podemos tomar um caso particular de dois eventos A e B, assim a probabilidade da ocorrência simultânea dos dois eventos será igual à probabilidade da ocorrência do evento A (ou B) vezes a probabilidade da ocorrência do outro dado que ocorreu o primeiro, ou seja:

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

ou

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

EXERCÍCIO RESOLVIDO 3

Uma caixa contém 15 peças, das quais 4 possuem algum defeito. Selecionam-se duas peças, uma após a outra, sem repor a primeira. Qual a probabilidade de que nenhuma seja defeituosa?

Solução:

Sejam os eventos:

B_1 : " a 1ª peça retirada é boa ".

B_2 : " a 2ª peça retirada é boa ".

$$\text{Assim, a } P(B_1 \cap B_2) = P(B_1) \cdot P(B_2 / B_1) = \frac{11}{15} \cdot \frac{10}{14} = \frac{11}{21}.$$

1.4.3 TEOREMA DE BAYES

Teorema₅: sejam A_1, A_2, \dots, A_n eventos que formam uma partição do espaço amostral. Para qualquer evento B, tal que $P(B) > 0$, então:

$$P(A_i / B) = \frac{P(A_i) \cdot P(B / A_i)}{P(A_1) \cdot P(B / A_1) + \dots + P(A_i) \cdot P(B / A_i) + \dots + P(A_n) \cdot P(B / A_n)}.$$

EXERCÍCIO RESOLVIDO 4

Sejam as seguintes caixas, com as determinadas composições de bolinhas.

CAIXA-I: 3 bolinhas brancas e 5 vermelhas.

CAIXA-II: 4 bolinhas brancas e 2 vermelhas.

CAIXA-III: 2 bolinhas brancas e 7 vermelhas.

Seleciona-se uma caixa e dela retira-se uma bolinha. Verificando que a bolinha é vermelha, qual a probabilidade de ela ter saído da caixa II?

Solução:

Sejam os eventos

C1: "A caixa selecionada é a caixa I"

C2: "A caixa selecionada é a caixa II"

C3: "A caixa selecionada é a caixa III"

B: "A bolinha retirada é branca"

V: "A bolinha retirada é vermelha"

Assim, a probabilidade de ela ter saído da caixa II será calculada através da conta abaixo:

$$\begin{aligned} P(C_2/V) &= \frac{P(C_2) \cdot P(V/C_2)}{P(C_1) \cdot P(V/C_1) + P(C_2) \cdot P(V/C_2) + P(C_3) \cdot P(V/C_3)} = \\ &= \frac{\frac{1}{3} \cdot \frac{2}{6}}{\frac{1}{3} \cdot \frac{5}{8} + \frac{1}{3} \cdot \frac{2}{6} + \frac{1}{3} \cdot \frac{7}{9}} = \frac{72}{125}. \end{aligned}$$

Neste tópico, introduzimos alguns conceitos de probabilidades. Abordamos conceitos como espaço amostral, eventos, definição formal de probabilidade. Apresentamos também alguns teoremas sobre esse assunto. No próximo tópico, iremos falar sobre o conceito de variáveis aleatórias.

TÓPICO 2

Variáveis aleatórias

OBJETIVOS

- Aprender a diferenciar os tipos de variáveis aleatórias
- Conhecer as funções de probabilidades para as variáveis aleatórias discretas e as funções densidades de probabilidade para as variáveis aleatórias contínuas

Daremos continuidade a esta aula, trazendo para este tópico o estudo sobre a variável aleatória, diferenciando as variáveis aleatórias discretas das variáveis aleatórias contínuas. Bom estudo!

2.1 VARIÁVEL ALEATÓRIA

Definição₇ : Trata-se de toda função que irá associar um n° real a cada elemento do espaço amostral.

Os conjuntos dos pontos $(s, f(s))$ é chamado **função de probabilidade**. Podemos ver a Figura 1 ilustrando esta associação.

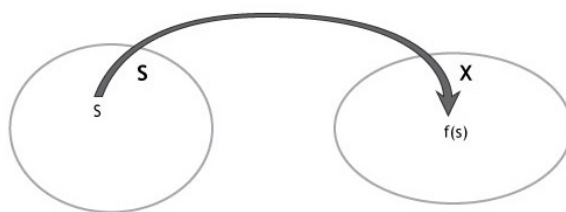


Figura 1: Diagrama de Venn de s em $f(s)$

Exemplo:

Seja E o experimento aleatório: “jogar três vezes uma moeda”. Assim teríamos o seguinte espaço amostral:

$$S = \{(C, C, C); (K, C, C); (C, K, C); (C, C, K); (K, K, C); (K, C, K); (C, K, K); (K, K, K)\}.$$

No qual temos: C = Cara e K = Coroa.

Poderíamos definir como variável aleatória X o n° de caras obtidas ao realizarmos o experimento. Neste caso, teríamos

$$X = 0 \rightarrow \{(K, K, K)\}$$

$$X = 1 \rightarrow \{(K, K, C); (K, C, K); (C, K, K)\}$$

$$X = 2 \rightarrow \{(K, C, C); (C, K, C); (C, C, K)\}$$

$$X = 3 \rightarrow \{(C, C, C)\}$$

Então os valores assumidos pela variável aleatória $X = \{0, 1, 2, 3\}$, e $\forall x \in X$, podem ocorrer com as respectivas probabilidades: $P(X = x) = \left\{ \frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8} \right\}$. Lembre-se de que utilizamos conceitos advindos do tópico 1.

2.2 TIPOS DE VARIÁVEIS ALEATÓRIAS

As variáveis aleatórias podem ser discretas e contínuas. Esses dois conceitos serão definidos logo a seguir.

2.2.1 VARIÁVEIS DISCRETAS

Definição₈: São variáveis que só podem assumir valores inteiros e que o contra domínio da função que envolve a variável é finito ou infinito enumerável.

Exemplo:

N° de aprovações em uma turma ou n° de disciplinas cursadas por um aluno.

Assim, os conjuntos dos pontos $(x, f(x))$ é chamado função de probabilidade, ou distribuição de probabilidade se tivermos diante de uma variável aleatória discreta X, tal que $\sum f(x) = 1$, em que $f(x) = P(X = x)$.

Temos também a **Função Distribuição de Probabilidade (F.D.P)**, ou função repartição (também chamada de função acumulativa), dada por:

$$F_x(X) = P(X \leq x) = \sum P(X = x).$$

Satisfazendo as seguintes propriedades:

1. $F(-\infty)=0$; ou seja, a função acumulativa da variável discreta para valores negativos é zero, pois a variável não assume valores negativos.
2. $F(+\infty)=1$, ou ainda $F(X \leq +\infty)=1$; neste caso, a função acumulativa, passa por todos os valores possíveis que a variável possa assumir. Assim a soma das probabilidades para todos esses valores resultará 1.
3. $P(a < X \leq b) = F(b) - F(a)$; ou seja, se estivermos interessados em calcular a probabilidade intervalar entre dois pontos onde somente o limite superior do intervalo faz parte da probabilidade desejada, devemos encontrar o acúmulo da função até o limite superior e subtrairmos o acúmulo da função até o limite inferior.
4. $P(a \leq X \leq b) = F(b) - F(a) + P(X = a)$: Como os dois limites do intervalo fazem parte da probabilidade desejada, a regra manda que além da diferença entre os resultados das funções acumulativas dos dois limites, devemos acrescentar o resultado da probabilidade do limite inferior do intervalo.
5. $P(a < X < b) = F(b) - F(a) - P(X = b)$: Como nenhum dos limites do intervalo faz parte da probabilidade desejada, a regra manda que além da diferença entre os resultados das funções acumulativas dos dois limites, devemos subtrair o resultado da probabilidade do limite superior do intervalo.

EXERCÍCIO RESOLVIDO 1

Ao lançarmos uma moeda duas vezes, encontre o espaço amostral, a função probabilidade e a FDP.

Solução:

Ora, como você pode observar, o espaço amostral seria o conjunto formado de todas as possibilidades do experimento, ou seja: $S = \{(C C), (C K), (K C), (K K)\}$.

Mas, a Função de probabilidade será dada da seguinte maneira:

x	\rightarrow	$P(X=x)$
0	\rightarrow	$\frac{1}{4}$
1	\rightarrow	$\frac{2}{4}$
2	\rightarrow	$\frac{1}{4}$

Já a F.D.P ou função repartição será dado por:

- $F(0) = P(X \leq 0) = P(X = 0) = \frac{1}{4}$.
- $F(1) = P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{1}{4} + \frac{2}{4} = \frac{3}{4}$.
- $F(2) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{1}{4} + \frac{2}{4} + \frac{1}{4} = 1$.

2.2.2 VARIÁVEIS CONTÍNUAS

Definição: São variáveis que podem assumir qualquer valor em certo intervalo da reta real e que o contra domínio da função é um intervalo ou conjunto de intervalos.

Exemplo:

Notas de uma turma ou idade dos alunos de uma turma.

ATENÇÃO!

Note que, no caso da variável aleatória X ser contínua, os limites do intervalo pertencer ou não a probabilidade desejada para variável não irá alterar o resultado final desta probabilidade, pois, no caso da variável contínua, a probabilidade em um determinado ponto é zero.

Assim, os conjuntos dos pontos $(x, f(x))$, é chamado função densidade de probabilidade se tivermos diante de uma **variável aleatória contínua** X , tal que

$$\int f(x) = 1.$$

Onde: $f(x) \geq 0$ para todo $x \in R(x)$. Onde $R(x)$ é o contradomínio de X .

$$\text{Além disso, } P(a < X < b) = \int_a^b f(x) dx$$

Obs₁: Se X é uma variável aleatória contínua então:

$$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b).$$

Obs₂: A função distribuição densidade acumulativa é obtida: $F(x) = \int_{-\infty}^x f(x) dx$.

EXERCÍCIO RESOLVIDO 2

Se X é uma variável aleatória contínua com a seguinte função:

$$f(x) = \begin{cases} 3x^2, & \text{se } 0 < x < 1 \\ 0, & \forall x \in \mathbb{R} \end{cases}$$

a) $f(x)$ é uma f.d.p (função densidade de probabilidade) ?

b) Qual a Função densidade acumulativa $F(x)$?

c) A probabilidade de $f(x)$ no intervalo de $\frac{1}{3} \leq x < \frac{4}{5}$?

Solução do item (a):

Sim, pois $f(x) \geq 0$ e para qualquer $x \in R(x)$.

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^0 0dx + \int_0^1 x^2dx + \int_1^{+\infty} 0dx = 1$$

Solução do item (b):

Veja que teremos que analisar a solução para os intervalos, $x < 0$, $0 \leq x < 1$ e $x \geq 1$.

- Para $x < 0$, $F(x) = \int_{-\infty}^x 0dx = 0$.
- Para $0 \leq x < 1$, $F(x) = \int_{-\infty}^0 0dx + \int_0^x 3x^2dx = x^3$.
- E finalmente, para $x \geq 1$,

$$F(x) = \int_{-\infty}^0 0dx + \int_0^1 3x^2dx + \int_1^{+\infty} 0dx = 0 \cdot x + 3 \frac{x^3}{3} + 0 \cdot x = 3 \cdot \frac{1^3}{3} = \frac{3}{3} = 1$$

Solução do item (c):

No intervalo $\frac{1}{3} \leq x < \frac{4}{5}$, teremos a seguinte probabilidade,

$$P\left(\frac{1}{3} \leq x < \frac{4}{5}\right) = \int_{\frac{1}{3}}^{\frac{4}{5}} 3x^2dx = 3 \frac{x^3}{3} \Bigg|_{\frac{1}{3}}^{\frac{4}{5}} = 3 \left(\frac{\left(\frac{4}{5}\right)^3}{3} - \frac{\left(\frac{1}{3}\right)^3}{3} \right) = 0,475.$$

Caro aluno, até agora conseguimos entender os conceitos teóricos associados ao cálculo das probabilidades. Formulamos uma definição para probabilidade, com base nesses conceitos, e realizamos um estudo básico sobre variável aleatória. Agora, iremos realizar cálculos de probabilidades através de modelos pré-estabelecidos, tanto para experimentos aleatórios discretos, como para experimentos aleatórios contínuos.



ATIVIDADES DE APROFUNDAMENTO

1) Seja o experimento aleatório E “lançar uma moeda e um dado simultaneamente e observar o resultado”.

A partir dessa situação, faça o que se pede:

- Qual o espaço amostral?
 - Formule dois eventos associados ao experimento que sejam independentes
 - Calcule a probabilidade do complementar da união desses dois eventos
- 2) Num sorteio de m Bingo, as peças chamadas estão numeradas de 1 a 70. Qual a probabilidade de ser chamado:
- Um múltiplo de 5 ou de 4
 - Uma peça não múltipla de 6
- 3) Sejam as seguintes caixas com as suas composições de bolas:

CAIXA-I: 2 bolas verdes, 6 bolas vermelhas e 1 bola branca

CAIXA-II: 5 bolas verdes, 1 bola vermelha e 3 bolas brancas

Selecione-se uma das caixas e em seguida uma bola. Se esta bola é verde, qual a probabilidade dela ter sido extraída da CAIXA-I ?

4) Se X é uma variável aleatória contínua com a seguinte função:
$$f(x) = \begin{cases} 4x^3 - x, & \text{se } 0 < x < 3 \\ 0, & \forall x \in \mathbb{R} \end{cases}$$

Então faça o que se pede a seguir.

- $f(x)$ é uma função densidade de probabilidade ?
- Calcule a $P(-1 < x < 4)$.

AULA 2

Distribuições discretas e contínuas de probabilidade

Olá aluno(a),

Na aula passada, estudamos alguns conceitos de probabilidade e variáveis aleatórias, você lembra? Agora, nesta segunda aula, será justamente o assunto de variáveis que nos dará uma direção e nos ajudará a conhecermos um pouco mais sobre distribuição e probabilidade. Nesta aula, iremos conhecer modelos de distribuições probabilísticas que irão servir para solucionar problemas de probabilidades associados tanto a variáveis discretas como a variáveis contínuas.

Então, vamos à aula?

Objetivos

- Conhecer e aplicar os principais modelos de distribuições discreta e contínua da uniforme
- Reconhecer a importância da distribuição normal como o mais importante modelo contínuo de probabilidade

TÓPICO 1

Distribuições discretas de probabilidades

OBJETIVO

- Conhecer as características de experimentos aleatórios que possam ser solucionados através da distribuição binomial, da distribuição hipergeométrica e da distribuição de Poisson

Inicaremos este primeiro tópico tratando acerca das aplicações dos principais modelos de distribuições discretas de probabilidade: **binomial; hipergeométrica e de Poisson**. Então, vamos aos estudos?

1.1 DISTRIBUIÇÃO BINOMIAL

VOCÊ SABIA?

A distribuição binomial é também chamada de distribuição de Bernoulli, em homenagem a James Bernoulli (SPIEGEL, 1993).



Segundo Spiegel (1993), se p é a probabilidade de um evento acontecer em uma tentativa única (chamada de probabilidade de sucesso) e $q=1-p$ é a de que o evento não ocorra em qualquer tentativa única (chamada de insucesso), então a probabilidade do evento ocorrer exatamente n repetições, em k tentativas é chamado de **distribuição binomial**.

Vamos considerar um experimento aleatório com as seguintes características:

Existem dois resultados possíveis (sucesso e fracasso)

- i) A probabilidade de sucesso p é constante ao longo das repetições.
- ii) As repetições são independentes.
- iii) Existe um número pré-fixado n de repetições.

Depois vamos associar a uma variável aleatória X “nº de sucesso”. Veremos, portanto, que a distribuição binomial será calculada através da seguinte expressão matemática.

$$P(X=k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$$

na qual temos as seguintes legendas:

n = nº de repetições.

k = valor desejado para a variável aleatória X .

p = probabilidade de sucesso em uma amostra.

q = probabilidade de fracasso em uma amostra.

Agora, veja um exercício resolvido para melhor compreendermos esse conceito.

EXERCÍCIO RESOLVIDO 1

Se a probabilidade de certo estudante acertar um problema de probabilidade é de 93%, qual probabilidade de que em cinco problemas solucionáveis ele:

- Acerte quatro.
- Erre no máximo um.

SOLUÇÃO DO ITEM (A):

Para você resolver essa questão, lembre-se de que se deve ter as seguintes informações bem claras: $n = 5; k = 4; p = 93\% = 0,93$ e $q = 7\% = 0,07$.

Usando a expressão matemática estudada anteriormente:

$$P(X=k) = \binom{n}{k} \cdot p^k \cdot q^{n-k} \text{ . Obteremos o seguinte cálculo:}$$

$$P(X=4) = \binom{5}{4} \cdot 0,93^4 \cdot 0,07^{5-4} = 0,304 = 30,4\%$$

SOLUÇÃO DO ITEM (B):

Aqui, você deve se atentar para o tipo de variável aleatória. Trata-se de números de erros e não mais acertos. Então teremos agora as seguintes informações:

$$n = 5; k = [0,1]; p = 7\% = 0,07 \text{ e } q = 93\% = 0,93.$$

Assim sendo, obteremos o seguinte cálculo:

$$\begin{aligned} P(X \leq 1) &= P(X=0) + P(X=1) = \binom{5}{0} \cdot 0,07^0 \cdot 0,93^{5-0} + \binom{5}{1} \cdot 0,07^1 \cdot 0,93^{5-1} = \\ &= 0,957 = 95,7\% \end{aligned}$$

Conseguiu compreender? A seguir iremos estudar outro conceito de distribuição discreta, chamada distribuição hipergeométrica.

1.2 DISTRIBUIÇÃO HIPERGEOMÉTRICA

Apresentaremos o conceito de **distribuição hipergeométrica**, considerando um experimento aleatório com as seguintes características:

- i) Existem dois resultados possíveis (sucesso e fracasso).
- ii) A probabilidade de sucesso p , varia ao longo das repetições.
- iii) As repetições são dependentes.
- iv) Existe um n° pré-fixado n de repetições.

Assim, se associarmos a uma variável aleatória X “ n° de sucesso”, obteremos

$$P(X=k) = \frac{\binom{n_x}{k} \cdot \binom{N-n_x}{n-k}}{\binom{N}{n}}, \text{ em que}$$

n_x = n° de casos da variável aleatória.

k = valor desejado para a variável aleatória X .

N = n° total de casos.

n = n° de repetições.

Agora observe e analise o próximo exercício resolvido para que você aluno entenda melhor, porque neste caso o modelo hipergeométrico é o indicado para encontrarmos a solução do problema.

EXERCÍCIO RESOLVIDO 2

Uma caixa contém 7 bolas azuis e 5 bolas brancas. Retiram-se, sem reposição, quatro bolas. Qual a probabilidade de pelo menos uma das bolas retiradas ser da cor branca?

Solução:

Alunos vejam que, neste problema, temos as seguintes informações: $n_x = 5$, $k = [1, 2, 3, 4]$, $N = 12$ e $n = 4$. Se sabemos que a variável aleatória X está inserida em “ N° de bolas brancas”, então, aluno, acompanhe o raciocínio:

O problema pede a probabilidade de serem retiradas pelo menos uma bola branca. Assim, se a variável aleatória é definida como sendo:

X : “ n° de bolas brancas”.

Então devemos encontrar $P(X \geq 1)$ cujas possibilidades são $(X = 1)$ ou $(X = 2)$ ou $(X = 3)$ ou $(X = 4)$. Desta forma podemos encontrar a probabilidade pedida pela seguinte relação:

$P(X \geq 1) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$. Em vez de se calcular as quatro probabilidades para somarmos, devemos lembrar que, em algumas situações, podemos aplicar a regra do complementar da probabilidade pedida, para diminuir o nosso trabalho e chegarmos a mesma resposta final. Ou seja, podemos ainda dizer que a $P(X \geq 1)$ poderá ser encontrada, subtraindo de 1 a probabilidade do complementar que é dada por $P(X < 1)$, ou seja, $P(X \geq 1) = 1 - P(X < 1)$. Como $(X < 1)$ é igual a $(X = 0)$, pois a variável por ser discreta só pode assumir valores inteiros e positivos, podemos ainda dizer que $P(X \geq 1) = 1 - P(X = 0)$. Assim iremos encontrar primeiro a $P(X = 0)$. Lembre aluno, a variável aleatória X : “nº de bolas brancas”. Percebemos que o problema segue as quatro características do modelo hipergeométrico. Se não vejamos:

i) Existem dois resultados possíveis, a bola retirada é ou não branca.

ii) A Probabilidade de sucesso varia ao longo das repetições, pois a probabilidade da 1ª bola ser branca é $\frac{5}{12}$, e a probabilidade da 2ª bola ser branca já não é a mesma por que as retiradas das bolas ocorrem sem reposição.

iii) As repetições são dependentes, pois a 2ª bola ser branca, dependerá se a 1ª foi ou não branca.

iv) É pedida uma probabilidade a partir de um nº pré-fixado de repetições, ou seja, 4 retiradas.

Então, teremos através da distribuição hipergeométrica:

$$P(X=k) = \frac{\binom{n_x}{k} \cdot \binom{N-n_x}{n-k}}{\binom{N}{n}}.$$

Assim sendo, podemos usar essa fórmula a seguir:

$$P(X=0) = \frac{\binom{5}{0} \binom{12-5}{4-0}}{\binom{12}{4}} = \frac{\binom{5}{0} \binom{7}{4}}{\binom{12}{4}} = \frac{5!}{0!5!} \times \frac{7!}{4!3!} = \frac{12!}{4!8!} = 0.0707. \text{ Assim, teremos:}$$

$$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.0707 = 0.9293$$

A seguir, iremos ver algo sobre distribuição de Poisson.

1.3 DISTRIBUIÇÃO DE POISSON

Neste item, iremos resolver problemas de probabilidade referentes a experimentos aleatórios que possuam uma determinada particularidade. Considere um experimento aleatório em que o número médio de sucesso que ocorre em certa região específica, ou em certo intervalo que pode ser de tempo, distância, área, volume (ou outra unidade análoga) seja conhecido e proporcional ao tamanho total da região ou comprimento total desse intervalo.

Assim, associamos a uma variável aleatória X “nº de sucesso”. Então:

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

em que:

$$e \cong 2,71828.$$

k = valor desejado para a variável aleatória X .

λ = nº médio de sucesso.

Esse tipo de distribuição é chamado de distribuição de **Poisson**. A seguir observe um exercício resolvido.



SAIBA MAIS!

Siméon Denis Poisson nasceu em 1781 na cidade francesa de Pithivierse foi um matemático e físico francês. Foi considerado por muitos como o sucessor de Laplace. Desenvolveu pesquisas nas áreas de mecânica, eletricidade, elasticidade, calor e som. Teve grandes contribuições matemáticas aplicadas na medicina, astronomia e estatística.

Ver mais informações no site:
www.brasilecola.com/biografia/simeon-denis.htm

EXERCÍCIO RESOLVIDO 3

Em uma produção de tecidos, ocorrem em média 6 falhas a cada 150 m produzidos. Qual a probabilidade de que:

- a) Ocorram 10 falhas em 200m observados de uma produção.
- b) Ocorram no máximo 2 falhas em 15 metros.

SOLUÇÃO DO ITEM (A):

Ora, o número de falhas, equivalentemente para 200m, poderá ser calculado através de uma regra de três simples, isto é:

$$\begin{array}{l} 6 \text{ falhas} \rightarrow 150\text{m} \\ \lambda \text{ falhas} \rightarrow 200\text{m} \end{array} \left\langle \frac{6}{\lambda} = \frac{150}{200} \right\rangle \lambda = 8 \text{ falhas}$$

Assim sendo, para a variável aleatória X “Nº de falhas” pedida na questão, tem então

$$P(X = 10) = \frac{e^{-8} \cdot 8^{10}}{10!} = 0,0993 = 9,93\%$$

SOLUÇÃO DO ITEM (B):

Novamente, você deve atentar para saber qual o número de falhas que realmente é equivalente para 200m, o qual será calculado através da regra de três simples, assim:

$$\begin{array}{l} 6 \text{ falhas} \rightarrow 150\text{m} \\ \lambda \text{ falhas} \rightarrow 15\text{m} \end{array} \left\langle \frac{6}{\lambda} = \frac{150}{15} \right\rangle \lambda = 0,6 \text{ falhas}$$

Logo, teremos a probabilidade pedida no item (b). Lembre que X irá ter um limite até 2. Então:

$$\begin{aligned} P(X \leq 2) &= P(X=0) + P(X=1) + P(X=2) = \\ &= 0,5488 + 0,3293 + 0,0988 = 0,9769 = 97,69\% \end{aligned}$$

Neste tópico, conseguimos resolver problemas que envolviam variáveis aleatórias discretas, fazendo uso de modelos de distribuições discretas de probabilidade. No próximo tópico, você irá conhecer os principais modelos de distribuições contínuas.

TÓPICO 2

Distribuições contínuas de probabilidades

OBJETIVO

- Conhecer as características de experimentos aleatórios que possam ser solucionados através da distribuição uniforme, normal e normal padrão

Daremos continuidade a esta aula percorrendo, neste tópico, sobre as aplicações dos principais modelos de distribuições contínuas de probabilidade, que no caso são 3: **uniforme, normal e normal padrão**.

2.1 DISTRIBUIÇÃO UNIFORME

Uma variável aleatória contínua X possui uma **distribuição uniforme** no intervalo $[a;b]$ se a sua função densidade de probabilidade for dada pela fórmula matemática

$$\begin{cases} f(x) = \frac{1}{b-a}, & \text{se } a \leq x \leq b; \\ 0, & \text{caso contrário.} \end{cases}$$

EXERCÍCIO RESOLVIDO 1

A ocorrência de quedas em qualquer localidade de uma rede de *internet* em 10km segue uma distribuição uniforme no intervalo $[0,10]$.

- a) Qual a probabilidade de que uma queda venha ocorrer nos primeiros 600 metros?
- b) Qual a probabilidade de que ocorra uma queda nos 2 km centrais da rede?

Obs.: Veja bem, alguns conceitos da aula passada serão usados aqui. Qualquer dúvida retorne à aula 1 para estudar esses conceitos.

SOLUÇÃO DO ITEM (A):

A função de probabilidade (f.d.p) da distribuição uniforme será dada por $f(x) = \frac{1}{10}$, se $0 \leq x \leq 10$. Assim, a probabilidade de ocorrer queda nos primeiros 600m será dada por

$$P(X \leq 0,6) = \int_0^{0,6} f(x).dx = \frac{0,6 - 0}{10} = 0,06 = 6\%$$

SOLUÇÃO DO ITEM (B):

Já a probabilidade de ocorrer uma queda nos 2 km centrais, será dada por

$$P(5 \leq X \leq 6) = \int_5^6 f(x).dx = P(X \leq 6) - P(X \leq 5) = \frac{6}{10} - \frac{5}{10} = \frac{1}{10} = 0,1 = 10\%$$

A seguir, você verá o conceito de distribuição normal seguido de um exercício resolvido.

2.2 DISTRIBUIÇÃO NORMAL

A **distribuição normal** é a mais importante distribuição estatística, considerando a questão prática e teórica. A probabilidade de uma observação assumir um valor entre dois pontos quaisquer é igual à área compreendida entre esses dois pontos. Notemos conforme o gráfico abaixo que a curva da distribuição normal, nos transmite algumas informações iniciais já pré-estabelecidas. Se uma variável está normalmente distribuída, ou seja, se a curva da distribuição segue as características abaixo informadas, poderemos por exemplo dizer que cerca de 68,3% das observações desta variável, estará no intervalo de um desvio-padrão em torno de sua média. E ainda, quase a totalidade das observações desta variável, 99,7% delas, estarão no intervalo de três desvios em torno de sua média (veremos com mais detalhes nas próximas aulas sobre medidas estatísticas).

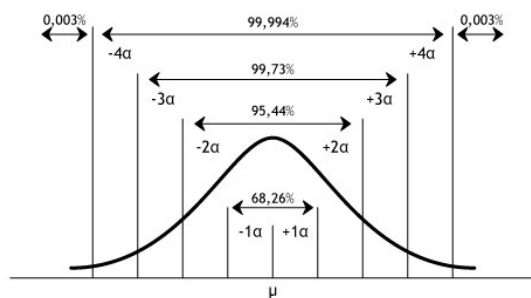


Figura 1: Curva simétrica da distribuição normal

CARACTERÍSTICAS:

- I. A curva da distribuição é simétrica em relação a sua média μ . Ou seja, note na curva que o comportamento a direita da sua média μ é simétrico ao comportamento a esquerda desta média. Assim, se estivermos calculando uma área a uma certa distância à direita da média da distribuição, esta área será igual a uma área que se situar a uma mesma distância a esquerda desta média.
- II. A curva prolonga-se de $-\infty$ a $+\infty$. Ou seja, podemos encontrar probabilidades intervalares, através de áreas sob esta curva da normal, desde um menor valor possível que esta variável possa assumir, até um maior valor possível.
- III. A distribuição é especificada por uma média μ e um desvio padrão σ . Ou seja, para calcularmos qualquer probabilidade de uma variável que siga uma normal, devemos conhecer a média e o desvio-padrão desta variável.
- IV. A área total sob a curva é de 100%, que seria o mesmo de estarmos calculando toda a área sob a curva da normal.

Assim, teremos que a probabilidade de uma observação assumir um valor entre dois pontos quaisquer é igual à área compreendida entre esses dois pontos.

Se X é uma V.A, que segue uma normal com média μ e desvio padrão σ , em que, por notação, podemos escrever uma $N(\mu; \sigma^2)$. Então, $P(a \leq X \leq b) = \int_a^b f(X).dx$. Onde:

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$

Principais notações:

μ = média populacional.

σ = desvio padrão populacional.

V.A = Variável Aleatória.

EXERCÍCIO RESOLVIDO 2

Se estivermos diante de uma variável aleatória X , que esteja normalmente distribuída com uma média de 8,8 e um desvio padrão de 1,1, qual seria a probabilidade de que:

- a) um valor de X seja menor que 9.

b) um valor de X seja maior que 7,5.

Bem, sabendo-se que a variável segue uma distribuição normal com média 8,8 e desvio padrão de 1,1, ou seja:

$$\mu = 8,8$$

$$\sigma = 1,1.$$

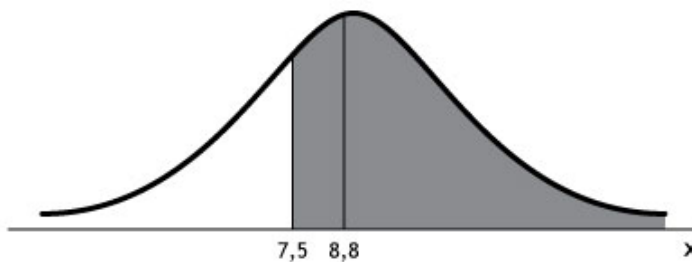
Então, no item a), segundo a curva da normal e a probabilidade pedida, iríamos encontrar a área em destaque abaixo:



Para tanto, teríamos que usar da integralização para encontrarmos tal probabilidade, ou seja: $P(X < 9) = \int_{-\infty}^9 f(X)dx$, em que $f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$.

Ou seja, substituindo os valores de $\mu = 8,8$ e $\sigma = 1,1$. Para diminuirmos nosso trabalho, poderemos padronizar a curva e encontrarmos a mesma probabilidade pedida, fazendo uso das tabelas da normal padronizada (Tal aplicação, caro aluno, veremos no próximo item 2.3).

b) Neste item foi pedida a $P(X > 7,5)$. Analogamente ao item anterior podemos observar a área pedida para a variável X , e usar da integral para buscar a solução. Vejamos:



$$\text{Ou seja, } P(X > 7,5) = \int_{7,5}^{+\infty} f(X)dx = \int_{7,5}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}; \text{ substituindo os}$$

valores de $\mu = 8,8$ e $\sigma = 1,1$. Também iremos encontrar esta probabilidade pedida, através da padronização desta curva e da utilização da tabela da normal padrão (Tal aplicação, aluno, também veremos no próximo item 2.3).

Então, no próximo item, iremos conhecer como trabalhar com a curva da normal padronizada, para calcularmos as áreas de probabilidades pedidas para variáveis contínuas que sigam uma $N(\mu; \sigma^2)$, sem ser necessário usarmos da integralização.

Vamos ver como funciona?

2.3 DISTRIBUIÇÃO NORMAL PADRÃO

Se tivermos uma variável aleatória X que siga uma distribuição $N(\mu; \sigma^2)$, poderemos transformar a variável X em uma variável padronizada Z que siga uma normal cuja média seja sempre igual a zero e o desvio padrão igual a 1, ou seja, em uma $N(0,1)$.

A padronização faz-se chamando de $Z_i = \frac{X_i - \mu}{\sigma}$. Assim a função densidade da normal padronizada seria dada por: $\varphi(Z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1(z)^2}{2}}$, $-\infty < Z < +\infty$, tal que a média da normal padrão, também conhecida como a esperança de Z , seria igual a zero, ou seja, $E(Z)=0$ e a variância da normal padrão seria igual a 1, ou seja, $\text{Var}(Z)=1$.

DEMONSTRAÇÃO - 1:

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

$$E(Z) = E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma} [E(X - \mu)] = \frac{1}{\sigma} [E(X) - E(\mu)] = \frac{1}{\sigma} [\mu - \mu] = 0$$

DEMONSTRAÇÃO - 2

$$\text{Var}(Z) = \text{Var}\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2} [\text{Var}(X - \mu)] = \frac{1}{\sigma^2} [\text{Var}(X)] = \frac{\sigma^2}{\sigma^2} = 1$$

Obs.: A variância é igual ao quadrado do desvio padrão. Veremos em medidas estatísticas na aula 4.

Assim, aluno, se tivermos de calcular uma probabilidade intervalar de uma variável X , que siga uma distribuição $N(\mu; \sigma^2)$, teremos apenas que transformar numa normal padronizada $N(0;1)$, através do escore de transformação $Z_i = \frac{X_i - \mu}{\sigma}$

e com o uso da tabela da normal padronizada abaixo, buscar o resultado da probabilidade solicitada.

TABELA DA DISTRIBUIÇÃO NORMAL PADRÃO

P(Z< z)	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999

3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Fonte: www.pucrs.br/format/rossana/psicologia/tabela_normal.pdf

P(Z<z)	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
-2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
-2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
-2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
-2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
-2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
-3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
-3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
-3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
-3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
-3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,6	0,0002	0,0002	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,7	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001

-3,8	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Fonte: www.pucrs.br/format/rossana/psicologia/tabela_normal.pdf

Vamos ver como funciona terminando o Exercício Resolvido 2.

Ali nós tínhamos uma variável X que seguia uma normal com média $\mu = 8,8$

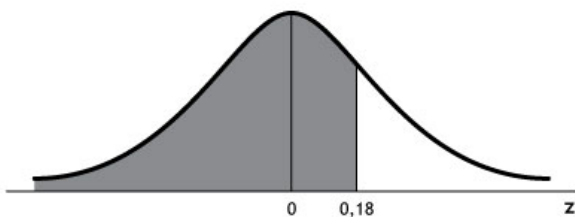
e $\sigma = 1,1$ e queríamos:

a) $P(X < 9)$, ou seja, tínhamos que encontrar a seguinte área sob a curva normal:



Iremos usar do escore de padronização $Z = \frac{X - \mu}{\sigma}$, para encontrarmos a curva transformada Z . Assim, como a média da normal padrão é sempre zero, teríamos apenas que saber qual é o valor de Z , quando X for 9. Então vejamos:

$Z = \frac{X - \mu}{\sigma} = \frac{9 - 8,8}{1,1} = 0,18$. Então, teríamos a seguinte curva similar e padronizada Z :



Assim, podemos ver com base nas tabelas da normal padrão que:

$$P(X < 9) = P(Z < 0,18) = 0,5714.$$

Conseguiu compreender ? Então, vamos resolver o item b.

Nele era pedido $P(X > 7,5)$. De forma análoga, iremos transformar a curva X , em uma curva em Z , ou seja:



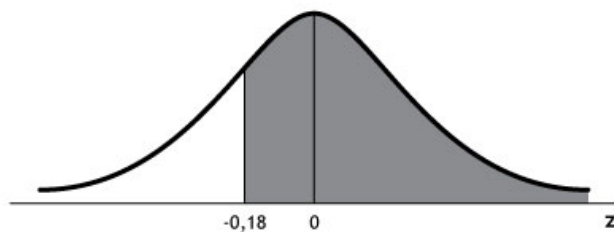
ATENÇÃO!

Os valores fornecidos na tabela da normal padrão acima são referentes às áreas de Z menores que certo valor z , ou seja, $P(Z < z)$. Os valores de Z são vistos na tabela cruzando o valor da linha e o valor da coluna cuja soma é o valor desejado. No Ex. $Z < 0,18$, temos que olhar o valor de probabilidade correspondente a 0,10 da coluna e 0,08 da linha.



Através do escore

$Z = \frac{X - \mu}{\sigma} = \frac{7,5 - 8,8}{1,1} = -1,18$. Assim, teríamos a seguinte curva padronizada:



Consequentemente, podemos dizer que

$P(X > 7,5) = P(Z > -1,18)$. Vimos que uma das características da curva normal é que ela é simétrica em relação a sua média. Assim podemos escrever da seguinte forma:

$$P(Z > -1,18) = P(Z < 1,18).$$

Agora, podemos encontrar a probabilidade pedida nas tabelas da normal padronizadas. Então teremos $P(Z < 1,18) = 0,8810$.

Para reforçar, vamos a mais um exercício resolvido.

EXERCÍCIO RESOLVIDO 3

As notas dos alunos de uma turma estão normalmente distribuídas com média $7,9$ e desvio padrão $0,5$. Se selecionarmos aleatoriamente um aluno desta turma, qual a probabilidade de que sua nota

- Seja menor que $8,5$.
- Seja menor que 6 .
- Esteja entre 7 e 9 .

SOLUÇÃO DO ITEM (A):

Como a variável aleatória X : “notas” segue uma normal com média $\mu = 7,9$ e desvio padrão $\sigma = 0,5$, e teremos que encontrar no item “a” a seguinte probabilidade:

$$P(X < 8,5) = ?$$

Vimos que a melhor maneira é transformarmos a probabilidade pedida em X , numa padronizada Z , assim usaremos o escore de transformação:

$$Z = \frac{X - \mu}{\sigma} = \frac{8,5 - 7,9}{0,5} = 1,2$$

Assim a $P(X < 8,5) = P(Z < 1,2)$. Agora faremos uso da tabela (p. 34 e 35) e encontraremos que

$$P(Z < 1,2) = 0,8849$$

SOLUÇÃO DO ITEM (B)

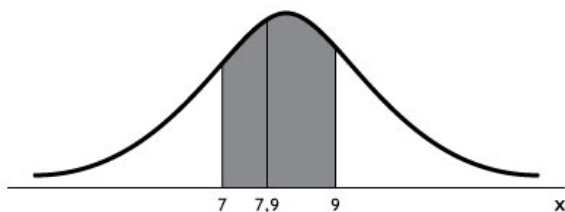
Neste item é pedida a probabilidade da nota do aluno selecionado ser menor que seis, ou seja, $P(X < 6)$. Então faremos da mesma maneira anterior:

$$Z = \frac{X - \mu}{\sigma} = \frac{6 - 7,9}{0,5} = -3,8 \Rightarrow P(X < 6) = P(Z < -3,8) = 0,00001$$

SOLUÇÃO DO ITEM (C)

Agora temos que encontrar a probabilidade da nota do aluno estar entre 7 e 9. Então vamos colocar a área correspondente na curva, para uma melhor compreensão.

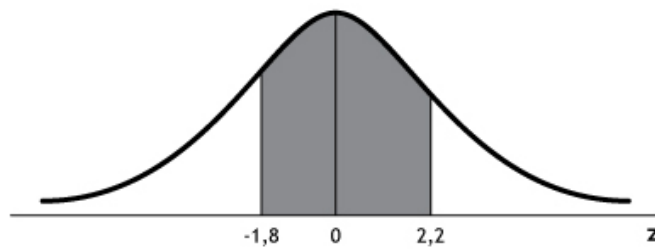
$$P(7 \leq X \leq 9) = ?$$



Assim, faremos a transformação através dos seguintes escores de $Z = \frac{X - \mu}{\sigma}$

$$Z_1 = \frac{7 - 7,9}{0,5} = -1,8 \quad Z_2 = \frac{9 - 7,9}{0,5} = 2,2$$

Dessa maneira, estaríamos diante da seguinte área desejada para Z:



Ou seja, podemos dizer que a $P(7 \leq X < 9) = P(-1,8 \leq Z < 2,2)$, e esta área de Z conforme a curva acima poderá ser encontrada subtraindo a área de Z até -1,8 da área de Z menor que 2,2. Consultando a tabela da normal padronizada teremos:

$$P(-1,8 \leq Z < 2,2) = P(Z < 2,2) - P(Z \leq -1,8) = 0,9861 - 0,0359 = 0,9502 .$$

Com isso, finalizamos nosso tópico conhecendo os mais importantes modelos de distribuições contínuas de probabilidade (**distribuição uniforme, distribuição normal e distribuição normal padrão**).

Nesta aula, você aprendeu a fazer uso de modelos de distribuições probabilísticas, pré-definidos tanto para experimentos que envolvem variáveis aleatórias discretas, como para experimentos que envolvem variáveis aleatórias contínuas. Na próxima aula, iremos dar início ao estudo das séries estatísticas para aprendermos a organizar as informações da melhor forma possível. Até lá!



ATIVIDADE DE APROFUNDAMENTO!

- 1) Em uma determinada delegacia, têm-se 88% de chances de deter um preso, quando este foge. Sendo assim, caso 3 presos fujam:
 - a) qual a chance de detê-los?
 - b) qual a chance de deter no máximo um deles?
- 2) Se nesta delegacia do exercício anterior são registradas 15 denúncias por hora, qual a probabilidade de que
 - a) em 10 minutos ocorram pelo menos três denúncias
 - b) em cinco minutos não ocorra denúncia
- 3) Em uma determinada sala desta delegacia, estão detidos 8 suspeitos, dos quais 5 são suspeitos de homicídios. Qual a probabilidade de que em três verificações
 - a) todos sejam suspeitos de homicídio
 - b) no máximo um não seja suspeito de homicídio
- 4) As idades dos presos estão distribuídas, segundo uma distribuição normal com média de 21,3 anos e desvio padrão de 1,9 anos. Ao selecionarmos um entre todos os presos, qual a probabilidade de que
 - a) ele seja menor de 18 anos
 - b) sua idade esteja entre 20 e 25 anos

AULA 3

Séries estatísticas

Olá caro aluno(a),

Na aula passada, você estudou conceitos acerca da noção de modelos de distribuições probabilísticas. Eles serviram para solucionar situações-problema de probabilidades que envolviam tanto a variáveis discretas como a variáveis contínuas. Nesta aula, iremos iniciar a organização de informações que serão representadas por variáveis com as mais diversas características. Estas variadas formas de representações corresponderão aos principais tipos de séries estatísticas, as quais você irá aprender como representá-las através de tabelas ou gráficos da melhor forma possível.

Objetivos

- Diferenciar os principais tipos de séries estatísticas
- Conhecer os elementos mais importantes e alguns elementos complementares que deverão ser utilizados na representação das séries através de tabelas e gráficos

TÓPICO 1

Principais tipos de séries

OBJETIVO

- Conhecer as séries cronológica ou temporal; geográfica ou de localização; específica ou categórica; bem como as séries mistas

No início dessa aula, você aprenderá a diferenciar os principais tipos de agrupamentos das informações e conhecerá as principais séries estatísticas. Irá perceber que as séries estatísticas podem ser representadas através de tabelas ou gráficos. Vamos a estas séries.

1.1 SÉRIE TEMPORAL OU CRONOLÓGICA

Aluno, uma série Temporal é definida como sendo aquela que, em sua representação, só ocorre variação da época da ocorrência.

Veja bem que, na série **temporal ou cronológica**, as informações são agrupadas de acordo com a época da ocorrência do fenômeno. A localidade da ocorrência e o fenômeno ocorrido permanecerão fixos. Vejamos um exemplo.

Produção Brasileira de Trigo
2005-2010

Anos	Quantidade (1000 t)
2005(1)	6455
2006	6512
2007	6960
2008	7047
2009	7256
2010	7500

Fonte: Dados Fictícios

Nota: Produção voltada para o consumo interno.

(1) Parte da produção exportada.

Note, aluno, que, na série acima, a localidade da produção é apenas no Brasil e que todas as quantidades da produção se referem apenas ao trigo. A única variação na informação refere-se à época da produção que varia de 2005 a 2010.

A seguir, você irá conhecer um segundo modelo de série estatística. Vamos a ela.

1.2 SÉRIE GEOGRÁFICA OU DE LOCALIZAÇÃO

Chamaremos de série Geográfica ou de localização aquela em que, na sua representação, só ocorre variação na localidade da ocorrência. Ou seja, neste tipo de série, os dados serão agrupados, segundo localidades distintas da ocorrência. Notaremos também que a época da ocorrência e o fenômeno ocorrido ficarão fixos. Veja um exemplo.

Produção Brasileira de Trigo, por Unidade da Federação - 2010

Unidades da Federação	Quantidade (1000 ton)
São Paulo	980
Santa Catarina	487
Paraná	778
Goiás	454
Rio de Janeiro	399
Rio Grande do Sul	870

Fonte: Fictícia

Você pôde notar pelo exemplo acima que a variação da informação só ocorre na localidade da produção. A produção trata-se apenas do trigo, assim todos os dados desta produção também estão se referindo apenas ao ano de 2010.

Veja a seguir um terceiro modelo de agrupamento de informações.

1.3 SÉRIE ESPECÍFICA OU CATEGÓRICA

Agora, aluno, chamaremos de série Específica ou Categórica aquela em que a variação só ocorre no próprio fenômeno ocorrido. Ou seja, nesta série estatística, a época e o local da ocorrência permanecerão fixos, enquanto os dados que se referem ao fenômeno ocorrido serão agrupados de acordo com a modalidade da ocorrência do fenômeno. Veja a situação a seguir:

Rebanhos Brasileiros - 2009

Espécie	Quantidade (1000 cabeças)
Bovinos	21000
Suínos	2 171
Caprinos	5 491
Equinos	6 200

Fonte: IBGE

Você notou que, no exemplo acima, a produção se refere a um único local (Brasil) e a um único período (2009), mas são apresentados resultados para mais de uma categoria de produção.

1.4 SÉRIES MISTAS

Depois de apresentar estas séries: temporal, geográfica, e específica vamos apresentar a você a combinação destas séries.

Como exemplo, teremos a situação abaixo:



SAIBA MAIS!

Ainda podemos combinar dois ou mais modelos de séries estatísticas, formando as séries das quais chamamos de conjugadas ou mistas.

Exportação Brasileira de alguns produtos agrícolas - 2007 - 2009

Produto	Quantidade (1000 ton)		
	2007	2008	2009
Feijão	7700	8300	9400
Arroz	10700	12000	12300
Soja	6100	7100	8100

Fonte: Dados imaginários

Nota: Produtos mais exportados no período.

Assim estaríamos diante de uma série estatística que em sua informação teríamos a variação da época da informação (2007 a 2009), caracterizando assim uma série cronológica, e também teríamos a variação das categorias do produto (feijão, arroz e soja), caracterizando também uma série categórica. Ou seja, teríamos uma variação simultânea de época e categoria, identificando-se desta maneira como sendo uma série mista do tipo cronológica-categórica.

Poderíamos até mesmo transformar a série do exemplo acima numa série que se identificasse ao mesmo tempo como sendo uma Cronológica, Geográfica e Categórica. Vejamos como seria:

**Exportação Brasileira de alguns produtos agrícolas pelas regiões
Norte e Sul (2007 – 2009)**

Produto	Quantidade (1000 ton)					
	2007		2008		2009	
	N	S	N	S	N	S
Feijão	5200	8100	7300	10000	5400	4600
Arroz	8700	11040	9480	8000	8300	6750
Soja	6900	9700	11000	6990	8100	4700

Fonte: Dados imaginários

Nota: Produtos mais exportados no período.

Note que, no exemplo acima, além das variações da época da exportação e das categorias dos produtos exportados, também ocorre a variação das localidades destas exportações por regiões, caracterizando assim uma série Cronológica, Categórica e também Geográfica.

Neste tópico, abordamos os principais tipos de séries estatísticas, agora vamos aprender quais são as formas corretas de representá-las. No próximo tópico, você verá algumas considerações importantes para representação das séries estatísticas por meio de tabelas.

TÓPICO 2

Representação das séries estatísticas através de tabelas

OBJETIVOS

- Identificar quais elementos são indispensáveis a composição de uma tabela
- Conhecer alguns elementos complementares

Iniciaremos este tópico conhecendo elementos que são indispensáveis às tabelas. Você identificará os elementos que compõe uma tabela: título; corpo e fonte, e conhecerá também os elementos que são complementares (não obrigatórios) e poderão ser usados em uma tabela, tais como notas e sinais convencionais. Então, vamos estudá-los?

2.1 ELEMENTOS FUNDAMENTAIS À COMPOSIÇÃO DE UMA TABELA

Quando estamos representando uma série estatística através de uma tabela, devemos atentar para o fato de que existem elementos que são obrigatórios na sua composição, como o título e o corpo da tabela. Assim, podemos enunciá-los a seguir:

2.1.1 TÍTULO

Toda tabela deverá conter um título, o qual deverá informar o que ocorreu, quando ocorreu e onde ocorreu.

2.1.2 CORPO

O corpo das tabelas é dividido em zonas da seguinte maneira:

ZONA DESIGNATIVA	ZONA DESIGNATIVA	
	ZONA DESIGNATIVA	ZONA DESIGNATIVA
ZONA INDICATIVA	ZONA ENUMERATIVA	

Figura 1: Detalhamento das características de uma tabela

- **Zona designativa:** situa-se logo abaixo do título, compreendendo o que chamamos de cabeçalho. Especifica os conteúdos das colunas.
- **Zona indicativa:** indica o tipo de série simples que estamos informando. Especifica os conteúdos das linhas.
- **Zona enumerativa:** também é conhecida como **zona de resultados**. Nela especificamos as grandezas ocorridas na série.

2.1.3 FONTE

Situa-se logo abaixo das tabelas, informando o órgão responsável pela divulgação dos conteúdos da tabela.

O exemplo da figura 2, mostra com detalhes quais são os elementos essenciais na composição de uma tabela.

O diagrama mostra uma tabela com o título 'Transcrições de série em 1982, taxas oficiais'. A tabela é dividida em quatro zonas: Zona Designativa (cabeçalho), Zona Indicativa (linhas 1-3), Zona Enumerativa (linhas 4-8) e Zona Fonte (rodapé). As setas indicam a localização de cada zona em relação à tabela.

Transcrições de série em 1982, taxas oficiais			
Série	Repetência	Promoção	Evação
1	0,296	0,449	0,255
2	0,207	0,703	0,090
3	0,169	0,738	0,093
1	0,134	0,818	0,048
5	0,227	0,634	0,138
6	0,199	0,700	0,102
7	0,170	0,730	0,100
8	0,123	0,764	0,114

Fonte: Serviço de Estatística da Educação e Cultura, Síntese Estatística da Educação Básica 1981/1982/1983 (Brasília: SEEC, 1984), Tabelas 3.2 e 3.5, pp. 35, 38, 85 e 88.

Figura 2: Exemplo de tabela

A seguir, você verá que nem todos os elementos são obrigatórios nas tabelas. Alguns são utilizados em algumas situações específicas. Vamos a eles?

2.2 ELEMENTOS COMPLEMENTARES À COMPOSIÇÃO DE UMA TABELA

Existem alguns elementos que só deveremos usar em algumas situações específicas. Dentre esses elementos, podemos destacar:

2.2.1 NOTAS

Só serão utilizadas quando for necessário esclarecer alguma composição que apareça na tabela.

2.2.2 SINAIS CONVENCIONAIS

Existem algumas simbologias que seguem certa convenção, das quais servem para esclarecer algumas informações. As principais são traço, três pontos e zero.

- **Traço (-):** é utilizado quando a informação inexistir.
- **Três pontos (...):** são utilizados quando não dispomos da informação.
- **Zero (Ø):** usado quando o valor numérico for muito pequeno para ser expresso na unidade de medida adotada.

Observação: o fechamento das tabelas só deverá ocorrer na parte superior e na parte inferior, nunca nas laterais, pois tecnicamente estaríamos diante de um quadro e não de uma tabela.

→ QUADRO

→ TABELA

Vejamos então a seguinte situação tabular:

Nº DE ANALFABETOS P/ SEXO DOS MUNICÍPIOS/ESTADO-X (2010)

MUNICÍPIOS	SEXO	
	MASCULINO	FEMININO
A	15.635	18.324
B	10.090	7.034
C	7*	-
D	5.489	7.543
E	...	8.423

FONTE: Fictícia

NOTAS: - (Nenhuma pessoa do sexo feminino do Município-C é analfabeta)

... (Não foi informada a quantidade de pessoas do Município-E do sexo masculino que é analfabeta)

*Poderíamos substituir este valor pela simbologia Ø, ressaltando em nota que praticamente não existe índice de analfabetismo de pessoas do sexo masculino do Município-X.

EXERCÍCIO RESOLVIDO 1

Crie uma tabela referente a uma série estatística mista, que se identifique como sendo Categórico-Cronológica e que contenha dois erros em sua composição. Em seguida, comente os erros.

Solução:

Nº DE MATRÍCULAS EM DUAS DISCIPLINAS NA UNIVERSIDADE-X (CEARÁ)

SEMESTRES	DISCIPLINA	
	A	B
2008.2	112	87
2009.1	132	91
2009.2	126	88
2010.1	145	100
2010.2	164	95

ERRO-1: O título está incompleto, pois não responde a uma das três perguntas básicas, que é obrigatória em todo título (quando ocorreu a informação?)

ERRO-2: A tabela não informa a fonte, que é obrigatória em toda composição de tabela.

Desta forma, aluno, concluímos neste tópico a maneira correta de representarmos as séries estatísticas através de tabelas. No próximo tópico, veremos as principais considerações a respeito destas representações de formas gráficas.

TÓPICO 3

Representação das séries através de gráficos

OBJETIVOS

- Conhecer todos os elementos obrigatórios para a composição dos gráficos, quando estivermos representando séries estatísticas
- Verificar que outros elementos também poderão ser usados nas representações gráficas, porém sem obrigatoriedade de utilização

Vamos primeiramente conhecer os elementos de uso obrigatórios aos gráficos.

3.1 ELEMENTOS FUNDAMENTAIS A COMPOSIÇÃO DE UM GRÁFICO

Igualmente às tabelas, os gráficos possuem elementos que são indispensáveis à sua composição. São eles o título, o corpo e a fonte.

3.1.1 – TÍTULO

Igualmente às tabelas, os gráficos necessitam de um título, o qual, para estar completo, também deverá responder àquelas três perguntas: o quê? Quando? Onde?

3.1.2 – CORPO

Este se definirá de acordo com o tipo de série que ele esteja representando. Veremos nos principais tipos de gráficos, logo a seguir no item 3.3.

3.1.3 – FONTE

A fonte indicará o órgão ou entidade responsável pelo fornecimento ou elaboração dos conteúdos apresentados. Assim, como o gráfico é gerado a partir de um levantamento de dados, então devemos responsabilizar alguém pelas informações.

Você verá a seguir que, em algumas situações, existem elementos que irão colaborar para esclarecer tipos específicos de situações gráficas, como as notas e as legendas.

3.2 ELEMENTOS COMPLEMENTARES À COMPOSIÇÃO DE UM GRÁFICO

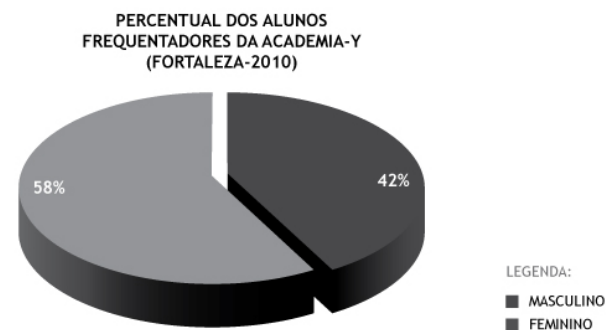
Existem alguns elementos que deverão ser utilizados somente em algumas circunstâncias. Vejamos os mais importantes:

3.2.1 – NOTAS

Alguns gráficos aparecem com composições, que necessitam ser esclarecidas. Nestas situações, podemos usar do recurso das notas. Preferencialmente, devem estar localizadas logo abaixo da fonte.

3.2.2 – LEGENDAS

São utilizadas para diferenciar as notações gráficas. Na maioria dos casos, diferenciamo-las por cores.



FONTE: ARBITRÁRIA

Vamos conhecer agora os principais tipos de gráficos.

ATENÇÃO!

Quando os dados são criados por conta própria, podemos dizer que a fonte pode ser arbitrária, hipotética ou fictícia.

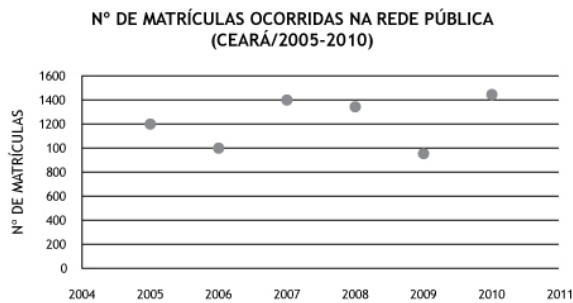
3.3 PRINCIPAIS TIPOS DE GRÁFICOS

3.3.1 – DIAGRAMAS

São considerados gráficos de análises, devido ao rigor e exatidão que eles apresentam. Através destes gráficos, podemos realizar análises descritivas das informações estatísticas com uma maior precisão. Os mais importantes

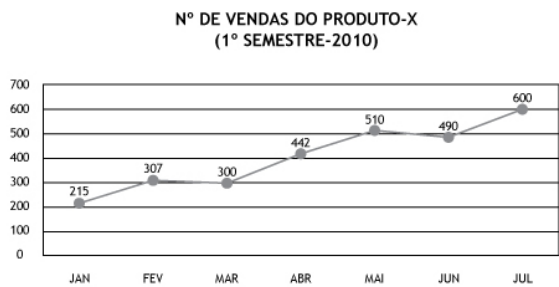
são por pontos, por linhas, por superfície, setorial, polar, histograma e polígonos de frequência.

1º) **Por pontos:** deverão ser utilizados quando forem poucas as classes a serem representadas.



Fonte: FICTÍCIA

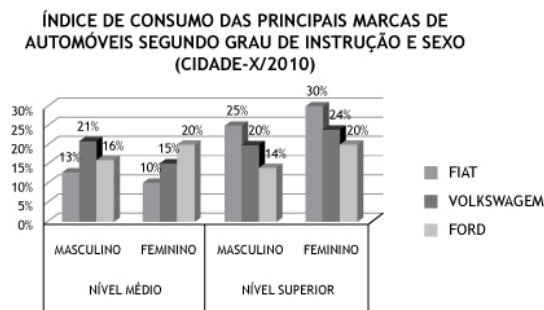
2º) **Por linhas:** mostram claramente a evolução do fenômeno ocorrido, por isso são indicados geralmente para realizar a representação das séries cronológicas.



Fonte: HIPOTÉTICA

3º) **Por Superfície:** são representados por áreas. Dentre todos os gráficos tipo superfície, iremos destacar os mais relevantes:

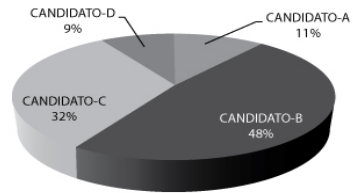
- **Colunas ou barras:** são os mais indicados para representar as séries específicas. Eles também são bastante eficientes na representação das séries mistas.



FONTE: ARBITRÁRIA

- **Setorial:** são representados por setores circulares. Usamos quando queremos ressaltar a participação de cada classe no total da informação.

ÍNDICE DE ACEITAÇÃO DOS CANDIDATOS A PREFEITURA
(CIDADE-X/2010)



FONTE: HIPOTÉTICA

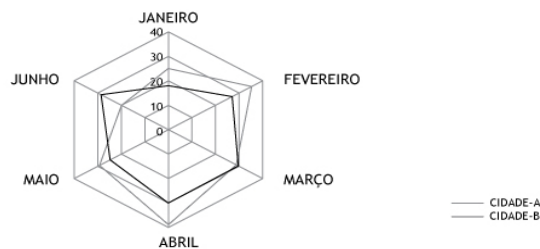
ATENÇÃO!

Na maioria dos softwares gráficos, este gráfico polar é conhecido pelo nome popular como gráfico radar.

- **Polar:** quando queremos comparar ocorrências de dois ou mais ciclos periódicos, a utilização do gráfico polar torna-se a melhor solução.

Vamos conferir no exemplo a seguir:

TEMPERATURA MÉDIA (1º SEMESTRE DE 2010)
CIDADE-A x CIDADE-B



FONTE: ARBITRÁRIA

- **Histograma e Polígonos de frequência:** representam as distribuições de frequências.

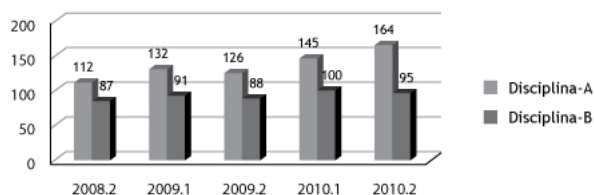
EXERCÍCIO RESOLVIDO 1

Qual seria o gráfico ideal para representar a série estatística do Exercício Resolvido 1 do tópico 1?

Solução:

Como estamos diante de uma série estatística mista, vimos que o gráfico ideal seria um gráfico em colunas composta. Assim, a representação ficaria da seguinte maneira:

Nº DE MATRÍCULAS EM DUAS DISCIPLINAS
NA UNIVERSIDADE-X (CEARÁ/2008-2010)



FONTE: ARBITRÁRIA

Estudamos, nesta aula, a organização de dados através das chamadas séries estatísticas. Vimos que, de acordo com o agrupamento das informações, estaremos diante de diferentes modelos de séries. Aprendemos também a representar corretamente as séries, de forma tabular e gráfica. Iremos, na próxima aula, dar início ao estudo da estatística descritiva. Primeiramente aprenderemos a organizar informações em tabelas de distribuições de frequências, para em seguida encontrar medidas estatísticas e realizar as análises descritivas cabíveis.



SAIBA MAIS!

Para a estatística descritiva, os gráficos tipos Diagramas, que vimos neste tópico, são os mais importantes. Vale salientar, porém, que existem outros tipos de gráficos conhecidos, tais como ORGANOGRAMAS, FLUXOGRAMAS, CARTOGRAMAS, ESTEREOGRAMAS e PICTOGRAMAS.



ATIVIDADES DE APROFUNDAMENTO!

- 1) Crie um levantamento de dados que possa ser representado através de uma série Cronológica.
- 2) Acrescente, ao levantamento anterior, dados suficientes para que o novo conjunto de dados passe a caracterizar uma série mista.
- 3) Crie um novo levantamento de dados para representar uma série estatística Categórica.
- 4) Acrescente, ao levantamento anterior, dados suficientes para que o novo conjunto represente uma série estatística mista do tipo Cronológica-Categórica-Geográfica.
- 5) Diga qual seria o modelo de gráfico ideal para representar as séries dos itens 1) 2) e 3).

AULA 4

Estatística descritiva - medidas de posição e medidas de dispersão

Saudações aluno(a),

Antes de iniciarmos esta aula, julgamos relevante para a uma melhor compreensão de nossa matéria definir o conceito de estatística descritiva. A estatística descritiva é aquela que envolve a coleta, a organização, a apuração e análises descritivas de resultados provenientes de dados que irão descrever os mais diversos tipos de ocorrências.

Nesta aula, iremos tratar primeiramente da organização dessas ocorrências, para depois podermos apurar e analisar essas ocorrências, através das principais medidas estatísticas de posição e de dispersão.

Vamos iniciar esta quarta aula fazendo um estudo da organização de dados em tabelas de distribuições de frequências.

Então, vamos iniciar a aula?

Objetivos

- Conhecer a organização de dados em tabelas de distribuições de frequências discretas e contínuas
- Calcular e interpretar as principais medidas estatísticas referentes a variáveis discretas e contínuas

TÓPICO 1

Distribuições de frequência

OBJETIVOS

- Organizar dados em tabelas de distribuições de frequências
- Diferenciar frequência absoluta, frequência acumulada e frequência relativa

Neste primeiro tópico, você vai conhecer o que são as distribuições de frequências. Antes, porém, é importante aprender a diferenciar os principais tipos de variáveis. Vamos lá?

1.1 PRINCIPAIS TIPOS DE VARIÁVEIS

1.1.1 - VARIÁVEL QUALITATIVA

São variáveis cujos valores são expressos por certos atributos ou qualidades. Podem ser do tipo qualitativa nominal e ordinal.

- **Qualitativa nominal:** são variáveis qualitativas que não são ordenáveis.
Exemplo: sexo, raça, religião.
- **Qualitativa ordinal:** são variáveis qualitativas que são ordenáveis.
Exemplo: classe social, grau de instrução.

1.1.2 - VARIÁVEL QUANTITATIVA

São variáveis cujos valores são expressos por números, obtidos através de um processo de medição ou de contagem. Podem ser do tipo discreta ou contínua.

- **Quantitativa discreta:** são variáveis que só podem assumir valores inteiros em pontos da reta real.

Exemplo: nº de habitantes, nº de filhos, nº de alunos matriculados na EaD.

- **Quantitativa contínua:** são variáveis que podem assumir qualquer valor entre dois pontos da reta real.

Exemplo: peso dos alunos de uma turma de EaD, notas dos alunos de uma turma de EaD.

Podemos tanto organizar todo tipo de levantamento de dados referentes aos mais diversos tipos de variáveis, através de distribuições de frequências, como condensar um conjunto de dados, tanto discretos como contínuos, conforme as frequências ou repetições de seus valores. Desta forma, estaremos construindo tabelas de distribuições de frequências.

Na distribuição de frequência referente à variável discreta, os dados não são agrupados em classes.

Nº de faltas	Nº de alunos
0	13
1	7
2	8
3	5
4	4
5	3
Σ	40

Note que na tabela o nº de faltas representa a variável discreta e o nº de alunos, representa a frequência com que cada resultado da variável ocorre. Assim podemos notar que 13 alunos não tiveram faltas, 7 alunos tiveram apenas 1 falta, 8 alunos tiveram duas faltas, e assim por diante. Ok?

Já na distribuição de frequência referente à variável contínua, os dados deverão ser agrupados por intervalos de classe, por exemplo:



ATENÇÃO!

Note que a notação $|$ diz que o limite inferior pertence ao intervalo e o limite superior não pertence.

Notas	Nº de alunos
0 2	3
2 4	5
4 6	11
6 8	15
8 10	63
Σ	40

Ou seja, neste exemplo, teríamos 3 alunos que tiraram nota de 0 a 1,9; 5 alunos que tiraram nota de 2 a 3,9; 11 alunos com nota de 4 a 5,9; e assim por diante.

Segundo Simon (1995, p.111), existe mais de uma metodologia para determinarmos o número de classes e a amplitude ou tamanho de cada classe.

Poderemos aplicar a seguinte regra para se determinar o nº de classes (k):

- Se $n \leq 25 \Rightarrow k = 5$
- Se $n > 25 \Rightarrow k \cong \sqrt{n}$.

$n \Rightarrow$ nº de observações. Após determinarmos o nº de classes, poderíamos encontrar o $h \Rightarrow$ Amplitude ou tamanho de cada classe.

$$h \cong A_t \div k.$$

$A_t \Rightarrow$ Amplitude total (diferença entre a maior e a menor observação).

Também podemos aplicar a seguinte fórmula de Sturges para determinarmos o nº de classes k .

$k \cong 1 + 3,22 \log n$. Ex. se $n=40$ então teríamos:

1º método:

$$n > 40 \Rightarrow k \cong \sqrt{40} \cong 7$$

2º método:

$k \cong 1 + 3,22 \log 40 \cong 7$. Nos dois métodos o arredondamento deverá ser sempre para o maior inteiro, fugindo da regra de arredondamento convencional.

Vamos diferenciar agora os principais tipos de frequências existentes, para posteriormente podermos aplicá-las nos cálculos de algumas medidas estatísticas.

1.2 FREQUÊNCIA ABSOLUTA, ACUMULADA E RELATIVA

Aluno, você irá perceber que há diferença entre os três tipos de frequências. Vamos descrever cada uma a seguir.

1.2.1 - FREQUÊNCIA ABSOLUTA (F_i)

É o nº de vezes que cada observação ocorre em um levantamento de dados discretos e o nº de observações ocorridas dentro de cada intervalo de observações no caso contínuo. Sendo assim no ex. da tabela abaixo, a coluna dos alunos



ATENÇÃO!

Se somarmos todas as frequências absolutas em uma tabela de frequência, teremos o total de observações que estaremos trabalhando, ou seja,

$$\sum_{i=1}^n F_i = n$$

representam as frequências com que os alunos tiram notas dentro de cada intervalo de notas.

1.2.2 - FREQUÊNCIA ACUMULADA (FAC)

É a soma das frequências anteriores ou iguais a uma determinada frequência absoluta observada. Por exemplo, na tabela abaixo o valor 8 da coluna das frequências acumuladas (Fac), representa a soma das frequências absolutas (F_i) 3 e 5.

Você deve observar que a última frequência acumulada deverá ser igual ao nº de observações. No Ex. das notas dos alunos, teríamos

Notas	Nº de alunos (F_i)	Fac
0 2	3	3
2 4	5	8
4 6	11	19
6 8	15	34
8 10	63	40
Σ	40	

Agora vejamos o que são frequências relativas.

1.2.3 - FREQUÊNCIA RELATIVA (f_i)

É a porcentagem representativa de cada observação ou intervalo de observações. Será obtida por $f_i = \frac{F_i}{n} \cdot 100$. Assim, no nosso exemplo, a primeira frequência relativa ao intervalo 0 | 2, seria dada por $f_i = \frac{3}{40} \times 100 = 7,5$. Desta forma, você já poderia construir a tabela de distribuição de frequência do exemplo das notas, da seguinte forma:

Notas	Nº de alunos (F_i)	Fac	$f_i(\%)$
0 2	3	3	7,5
2 4	5	8	12,5
4 6	11	19	27,5
6 8	15	34	37,5
8 10	6	40	15
Σ	40		100

Após ter visto e aprendido a organizar os dados em tabelas de frequências, você irá para o próximo tópico. Nesta segunda parte de nossa aula, aprenderá a calcular e interpretar as principais medidas de posição, que são as medidas as quais nos forneceram as primeiras informações sobre o comportamento das variáveis.

TÓPICO 2

Medidas de posição ou de tendência central

OBJETIVO

- Encontrar e analisar o resultado da média aritmética, mediana e moda, bem como os resultados das separatrizes: quartis, decis e percentis

As medidas de posição podem apresentar-se de várias formas, dependendo do que se pretende conhecer a respeito dos dados. Geralmente os dados resultantes das medidas de posição se concentram em torno do centro da distribuição, por isso são também chamadas de medidas de tendência central. Daremos início com a mais conhecida das medidas de posição, que é a média aritmética. Vamos lá então?

2.1 MÉDIA ARITMÉTICA (\bar{X})

Definição: é a representação de um conjunto de dados, o qual poderá estar de forma não-agrupada ou agrupada, através de um único valor.

A seguir, você irá ver as definições para o cálculo de dados não-agrupados e agrupados. Haverá exemplos para melhor compreensão do assunto. Iniciaremos com o cálculo para dados não-agrupados.

2.1.1 - MÉDIA ARITMÉTICA PARA DADOS NÃO-AGRUPADOS

Chamaremos de dados não-agrupados, ao conjunto de dados em que nenhum elemento se repete, não havendo assim a necessidade de agrupar estes dados por frequências, visto que todos teriam frequência igual a 1, conforme veremos no exemplo abaixo.

Assim, se X uma variável quantitativa que assume os valores X_1, X_2, \dots, X_n não agrupados, poderemos encontrar a média aritmética de X , através da seguinte relação.

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n}$$

Exemplo

o nº de jornais vendidos em uma determinada semana em uma banca ocorre na seguinte grandeza: 39, 18, 24, 20, 27, 19, 25. Assim, teríamos o seguinte nº médio de jornais vendidos:

$$\bar{X} = \frac{39 + 18 + 24 + 20 + 27 + 19 + 25}{7} = \frac{172}{7} = 24,6$$

Vejamos agora o cálculo para dados agrupados. Você irá perceber que, ao contrário dos dados não-agrupados, esses dados poderão ser agrupados por frequências, visto que aparecem elementos repetidos dentro do conjunto.

2.1.2 - MÉDIA ARITMÉTICA PARA DADOS AGRUPADOS

Se os valores da variável forem agrupados em uma distribuição de frequência, usaremos

$$\bar{X} = \frac{\sum_{i=1}^n X_i F_i}{n}$$

Em que termos as seguintes legendas:

X_i : valores observados da variável discreta, ou ponto médio das classes no caso contínuo.

F_i : frequências absolutas simples

n (nº de observações) = $\sum_{i=1}^n F_i$.

Nos exemplos anteriores das variáveis nº de faltas e notas, teríamos

Nº de faltas (Xi)	Nº de alunos (Fi)	XiFi
0	13	0
1	7	7
2	8	16
3	5	15
4	4	16
5	3	15
Σ	40	69

Observação: O preenchimento da terceira coluna foi resultante do produto entre cada resultado da primeira e cada resultado da segunda. Assim poderemos calcular o nº médio de faltas, da seguinte forma:

$$\bar{X} = \frac{\sum_{i=1}^n X_i F_i}{n} = \frac{232}{40} = 5,8 \text{ .}$$

Então, ocorreu em média 1,7 faltas por aluno

Já no exemplo das notas dos alunos, teríamos:

Notas	Nº de alunos (Fi)	Xi	XiFi
0 2	3	1	3
2 4	5	3	15
4 6	11	5	55
6 8	15	7	105
8 10	6	9	54
Σ	40		232

Observação: Neste caso, a coluna dos Xis é preenchida calculando o ponto médio de cada intervalo. Por ex. o 1º valor Xi=1 é resultante do ponto médio do intervalo 0 | 2. Já os resultados da coluna XiFi foram obtidos através do produto da coluna Xi pela coluna Fi.

Assim podemos calcular a nota média dos alunos, da seguinte maneira:

$$\bar{X} = \frac{\sum_{i=1}^n X_i F_i}{n} = \frac{232}{40} = 5,8 \text{ .}$$

Verificamos, assim, que a nota média para os 40 alunos foi de 5,8.

Veja como exemplo a seguinte série:

$1, 4, 5, 7, 8 \Rightarrow$ sua média será igual a 5. Se somarmos a constante $k=2$ aos termos da série teríamos uma nova série:

$3, 6, 7, 9, 10 \Rightarrow$ cuja média seria igual a $7 = 5 + 2$. Se multiplicarmos a constante $k=2$ aos termos da série inicial, teríamos a seguinte série resultante:

$2, 8, 10, 14, 16 \Rightarrow$ cuja média seria igual a $10 = 5 \times 2$

Veja agora mais uma medida de posição: a **mediana**.



SAIBA MAIS!

Se somarmos, subtrairmos, multiplicarmos ou dividirmos uma série de dados por uma constante e calcularmos a média da nova série obtida, esta também ficará acrescida, subtraída, multiplicada ou dividida por essa constante.

2.2 MEDIANA (MD)

Definição₂: A trata-se do elemento central de um conjunto de dados, ou seja, é um ponto de equilíbrio do conjunto. Veja a Figura 1.

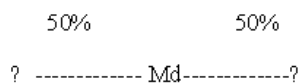


Figura 1: Mediana: ponto de equilíbrio

Existem dois procedimentos, um para distribuição de frequência discreta e outro para distribuição contínua. Vejamos a seguir:

2.2.1 – MEDIANA PARA VARIÁVEIS DISCRETAS

Teremos ainda duas situações:

1º caso: Quando n (n° de observações) for ímpar a mediana será o elemento de ordem: $\frac{n+1}{2}$.

Exemplo

Seja a seguinte série de dados, referentes a uma determinada variável discreta: $3, 5, 3, 2, 5, 6, 3, 4, 2, 7, 2$. Então teremos $n = 11$ (ímpar), assim a mediana será o elemento de ordem: $\frac{11+1}{2} = 6^\circ$ elemento de ordem, ou seja: $2, 2, 2, 3, 3, 3, 4, 5, 5, 6, 7 \Rightarrow Md = 3$

2º caso: Quando n (nº de observações) for par a mediana será a média aritmética entre os elementos de ordem: $\frac{n}{2}$ e $\frac{n}{2} + 1$.

Vejamos então o exemplo a seguir.

Exemplo

Seja a seguinte série de valores: 4, 2, 6, 3, 7, 5, 5, 4, 2, 3. Então $n = 10$ (par). Assim, teríamos $\frac{10}{2} = 5^\circ$ e $\frac{10}{2} + 1 = 6^\circ$. Ou seja: 2, 2, 3, 3, 4, 4, 5, 5, 6, 7. Desta forma a mediana seria dada por $\frac{4 + 4}{2} = 4$.

2.2.2. - MEDIANA PARA VARIÁVEL CONTÍNUA

Usaremos o seguinte processo de identificação da mediana

$$Md = \ell + \frac{\left(\frac{n}{2} - \sum f\right)_{xh}}{Fmd} \text{ em que teremos a seguinte legenda.}$$

$\ell \rightarrow$ limite inferior da classe mediana.

$\frac{n}{2} \rightarrow$ elemento identificador da classe mediana.

$\sum f \rightarrow$ soma das frequências anteriores a classe mediana.

$h \rightarrow$ amplitude da classe mediana.

$Fmd \rightarrow$ frequência da classe mediana.

Lembre-se que a amplitude é a diferença entre os limites da classe.

Aluno, voltaremos ao exemplo das notas para vermos como funciona:

GUARDE BEM ISSO!

Lembre-se que a amplitude é a diferença entre os limites da classe.

Notas	Nº de alunos	Fac
0 2	3	03
2 4	5	8
4 6	11	19
6 8	15	34
8 10	6	40
Σ	40	

$$\text{Assim, teríamos } Md = \ell + \frac{\left(\frac{n}{2} - \sum f\right)_{xh}}{Fmd}.$$

1º) Vamos identificar a classe mediana, pois todos os elementos se referem a ela.

$\frac{n}{2} = \frac{40}{2} = 20^\circ$, o qual pela Fac podemos verificar que se encontra na classe $6 \mid 8$.

2º) A partir da classe mediana, iremos identificar todos os elementos que compõem a fórmula de identificação, da seguinte maneira:

$$Md = 6 + \frac{(20 - 19) \times 2}{15} = 6,13.$$
 Ou seja, 50% dos alunos tiveram nota no máximo igual a 6,13 e outros 50% tiveram nota no mínimo igual a 6,13.

Agora você vai conhecer mais um conceito de medida de posição: a moda.

2.3 MODA (MO)

Definição₃: Trata-se do elemento mais comum em qualquer conjunto de dados, ou seja, aquele elemento que mais se repete neste conjunto.

Analogamente à mediana, teremos duas situações: uma quando a variável for discreta e outra quando a variável for contínua.

2.3.1 - MODA PARA VARIÁVEL DISCRETA

Neste caso, iremos apenas identificar o elemento modal.

Exemplo

Seja a seguinte série de dados, referentes a uma determinada variável discreta: 1, 4, 2, 5, 4, 6, 2, 4, 7, 5, 5, 4, 1, 4. Teríamos como resultado $Mo = 4$, pois é o valor mais frequente na série.

Usaremos o seguinte processo de identificação da moda:

$$Mo = \ell + \frac{\Delta_1}{\Delta_1 + \Delta_2} xh$$
, na qual teremos as seguintes legendas:

$\ell \rightarrow$ limite inferior da classe modal.



ATENÇÃO!

Como já vimos anteriormente, Fac é a frequência acumulada, obtida através do acúmulo das frequências absolutas simples.



VOCÊ SABIA?

Uma série de dados pode ser do tipo unimodal (uma única moda), bimodal (duas modas), trimodal (três modas) e multimodal (a partir de quatro modas).

$\Delta_1 \rightarrow$ diferença entre a frequência simples da classe modal e a frequência simples da classe anterior.

$\Delta_2 \rightarrow$ diferença entre a frequência simples da classe modal e a frequência simples da classe posterior.

$h \rightarrow$ amplitude da classe modal.

Obs: Classe modal trata-se da classe de maior frequência absoluta

Então, vamos agora encontrar a nota modal no nosso Ex. das notas, ok?

Notas	Nº de alunos
0 - 2	3
2 - 4	5
4 - 6	11
6 - 8	15
8 - 10	6
Σ	40

ATENÇÃO!

A classe modal é a classe de maior frequência simples. Então, no exemplo, a classe modal será de 6 - 8.

1º) Vamos identificar a classe modal, pois todos os elementos se referem a ela.

2º) A partir da classe modal, iremos identificar todos os elementos que compõem o processo de Czuber, da seguinte maneira:
 $Mo = 6 + \frac{4}{4 + 9} \times 2 = 6,6$, ou seja, a nota mais comum entre os 40 alunos foi 6,6.

Vamos finalizar as medidas de posição, conhecendo as **separatrizes**.

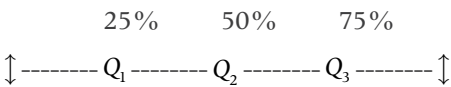
2.3 SEPARATRIZES

Dependendo do elemento limite do conjunto de dados que estivermos querendo encontrar, podemos necessitar dos quartis, decis ou dos percentis

2.3.1 - QUARTIS

Definição_q: são medidas estatísticas que dividem um conjunto de dados em quatro partes iguais.

Assim, temos o 1°, 2° e 3° quartil que poderão ser encontrados através da seguinte relação:



$$Q_i = \ell + \frac{\left(\frac{in}{4} - \sum f\right) xh}{FQ_i}, \text{ em que teremos as legendas a seguir:}$$

$\ell \rightarrow$ limite inferior da classe quartílica Q_i

$\frac{in}{4}$ elemento identificador da classe Q_i

$\sum f \rightarrow$ soma das frequências anteriores a classe Q_i

$h \rightarrow$ amplitude da classe Q_i

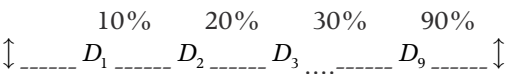
$FQ_i \rightarrow$ frequência da classe Q_i .

Veja aplicação na qual foi encontrado o resultado do terceiro quartil:

$$Q_3 = \ell + \frac{\left(\frac{3n}{4} - \sum f\right) xh}{FQ_3} = 6 + \frac{(30 - 19) x 2}{15} = 7,47. \text{ Ou seja,}$$

No exemplo das notas, podemos afirmar que 75% dos alunos tiraram no máximo nota igual a 7,5 e que 25% da turma tirou nota no mínimo igual a 7,5.

Definição₅: são medidas estatísticas que dividem um conjunto de dados em dez partes iguais. Assim, teremos o 1°, 2°, ..., 9° decil, ou seja:



Todos os decis poderão ser encontrados através da seguinte relação:

$$D_i = \ell + \frac{\left(\frac{in}{10} - \sum f\right) xh}{FD_i}, \text{ onde:}$$

$\ell \rightarrow$ limite inferior da classe decílica D_i .

$\frac{in}{10}$ elemento identificador da classe D_i .

$\sum f \rightarrow$ soma das frequências anteriores a classe D_i .

$h \rightarrow$ amplitude da classe D_i .

$FD_i \rightarrow$ frequência da classe D_i .

Como exemplo, vamos encontrar o oitavo decil das notas.

Exemplo: No Ex. das notas o oitavo decil, seria dado por:

$$D_8 = \ell + \frac{\left(\frac{8n}{10} - \sum f\right) x h}{FD_8} = 6 + \frac{(32 - 19) x 2}{11} = 7,7. \text{ Ou seja,}$$

No exemplo das notas, podemos afirmar que 80% dos alunos tiraram no máximo nota igual a 7,7 e que 20% da turma tirou nota no mínimo igual a 7,7. Finalizaremos as separatrizes, mostrando a você o último conceito estudado aqui: os percentis. Veja a seguir.

Definição: São medidas estatísticas que dividem um conjunto de dados em cem partes iguais.

Assim, teremos o 1°, 2°, ..., 99° percentil, ou seja:

$$\begin{array}{ccccccc} 1\% & 2\% & 3\% & & 99\% \\ \uparrow & \text{---} & P_1 & \text{---} & P_2 & \text{---} & P_3 \text{ --- } P_{99} & \text{---} & \downarrow \end{array}$$

Assim, poderemos encontrar qualquer destes percentis através da seguinte

relação: $P_i = \ell + \frac{\left(\frac{in}{100} - \sum f\right) x h}{FP_i}$.

A seguir, encontraremos no exemplo das notas o quadragésimo percentil.

$$P_{40} = \ell + \frac{\left(\frac{40n}{100} - \sum f\right) x h}{FP_{40}} = 4 + \frac{(16 - 8) x 2}{11} = 6,4. \text{ Ou seja, no exemplo}$$

das notas, podemos afirmar que 80% dos alunos tiraram no máximo nota igual a 7,7 e que 20% da turma tirou nota no mínimo igual a 7,7. Quando a variável for discreta, poderemos encontrar qualquer percentil através da ordem do elemento, da seguinte maneira:

$$X = (n - 1) x \frac{p}{100} + 1. \text{ Para tanto, devemos saber que}$$

$X \rightarrow$ É a ordem do elemento

$n \rightarrow$ É o nº de elementos

$p \rightarrow$ É o percentil desejado. Veremos como funciona, aplicando o ex. a seguir.

Exemplo: no ex. do nº de faltas dos alunos, o sexagésimo terceiro percentil seria dado por

Nº de faltas (Xi)	Nº de alunos (Fi)	Fac
0	13	13
1	7	20
2	8	28
3	5	33
4	4	37
5	3	40
Σ	40	

$$X = (n - 1) \times \frac{p}{100} + 1 = (40 - 1) \times \frac{63}{100} + 1 = 25,6 \cong 26^\circ$$
 elemento, o qual poderemos identificá-lo através da Fac e verificar que este elemento estará contido na frequência acumulada 28 que corresponde ao nº de falta 2. Ou seja, 63% dos alunos tiveram no máximo duas faltas e ainda 37% da turma tiveram no mínimo duas faltas.

Caro aluno(a), para fixarmos melhor todas as medidas estatísticas de posição que conhecemos neste capítulo, vamos observar o exercício resolvidos a seguir.

EXERCÍCIO RESOLVIDO 1

Seja a seguinte distribuição de frequência referente aos preços cobrados por uma mercadoria em 50 estabelecimentos comerciais:

Preços	Nº de estabelecimentos
2 – 5	8
5 – 8	10
8 – 11	9
11 – 14	12
14 – 17	11
Σ	50

- Qual o preço médio cobrado pela mercadoria?
- Interprete o resultado do preço mediano.
- Qual o preço cobrado pela maioria dos estabelecimentos?

- d. Noventa por cento dos estabelecimentos cobram no máximo quanto pela mercadoria?

Solução:

Neste caso, teremos que encontrar a média aritmética da distribuição:

Preços(R\$)	Nº de estabelecimentos-Fi	Xi	XiFi	Fac
2 5	8	3,5	28	8
5 8	10	6,5	65	18
8 11	9	9,5	85,5	27
11 14	12	12,5	150	39
14 17	11	15,5	170,5	50
Σ	50		499	

Usaremos a seguinte relação para dados agrupados:

$$\bar{X} = \frac{\sum X_i F_i}{n} = \frac{499}{50} = 9,98$$

Como a variável é contínua, podemos encontrar o resultado da mediana da maneira seguinte:

$$Md = \ell + \frac{\left(\frac{n}{2} - \sum f\right) xh}{FMd} = 8 + \frac{\left(\frac{50}{2} - 18\right) x3}{9} = 10,3. \text{ Ou seja, } 50\% \text{ dos}$$

estabelecimentos comerciais cobram, no máximo, R\$ 10,3 pela mercadoria. Também os 50% dos estabelecimentos que cobram mais pela mercadoria, no mínimo, cobram os mesmos R\$ 10,3.

Vimos que a moda por definição representa o elemento mais frequente, assim o preço cobrado pela maioria dos estabelecimentos será dado da seguinte forma:

$$Mo = \ell + \frac{\Delta_1}{\Delta_1 + \Delta_2} xh = 11 + \frac{3}{3+1} x3 = R\$15,6 = R\$13,2$$

Este preço será dado pela separatriz P90, ou seja:

$$P_{90} = \ell + \frac{\left(\frac{90n}{100} - \sum f\right) xh}{FP_{90}} = 14 + \frac{(45 - 39) x3}{11} = R\$15,6$$

Nesse tópico, você conheceu as principais medidas de posição. Agora continuará estudando as medidas estatísticas, conhecendo as principais medidas de dispersão. Vamos lá?

TÓPICO 3

Medidas de dispersão

OBJETIVOS

- Encontrar e analisar o resultado do desvio-médio, da variância e do desvio-padrão
- Aprender a verificar se um conjunto de dados possui uma baixa, uma média, ou uma alta dispersão

Então, aluno, você irá primeiramente reconhecer o desvio médio como uma importante medida de dispersão.

3.1 DESVIO MÉDIO (DM)

Definição₇: Podemos dizer que o trata-se da média das distâncias que existem entre cada observação de um conjunto e a média aritmética deste conjunto.

Poderemos encontrar o desvio-médio, através da seguinte relação

$$DM = \frac{\sum_{i=1}^N |X_i - \bar{X}| \times F_i}{N} .$$

Vejamos uma utilidade do desvio-médio através do seguinte exercício:

EXERCÍCIO RESOLVIDO 1

Qual das duas séries seguinte está sendo representada da melhor forma por sua média aritmética?

SÉRIE-A: 3,5,9,3,6,6,5,3 ($\overline{X_A} = 5$).

SÉRIE-B: 7,8,8,6,9,7,7,4 ($\overline{X_B} = 8$).

Solução:

Através dos desvios-médios, iremos verificar qual das duas séries melhor representa a sua média. Assim teremos

Na Série-A, a seguinte distribuição:

(X_i)	(F_i)	$X_i - \overline{X_A} \times F_i$
3	3	6
5	2	0
6	2	2
9	1	4
Σ	8	12

Então, turma, iremos encontrar o DM da série-A da seguinte forma:

$$DM = \frac{\sum_{i=1}^N |X_i - \overline{X}| \times F_i}{N} = \frac{12}{8} = 1,5$$

Na Série-B, haverá a seguinte distribuição:

(X_i)	(F_i)	$X_i - \overline{X_B} \times F_i$
4	1	4
6	1	2
7	3	3
8	2	0
9	1	1
Σ	8	10

$$DM = \frac{\sum_{i=1}^N |X_i - \overline{X}| \times F_i}{N} = \frac{10}{8} = 1,25. \text{ Assim, como } DM(B) < DM(A),$$

então a série B melhor representa a sua média aritmética.

A seguir, conheceremos a variância como uma das mais importantes medidas de dispersão.

3.2 VARIÂNCIA

ATENÇÃO!

Iremos ver o conceito de Inferências Estatísticas com mais profundidade a partir da aula 7.

Definição: É uma média dos quadrados dos desvios da média \bar{X} . O seu valor, além de analisar a dispersão de um conjunto, é utilizado para realizar algumas inferências estatísticas.

Teoricamente, podemos encontrar a variância tanto para uma população, como para uma amostra. Assim, podemos usar dos seguintes procedimentos:

3.2.1 – VARIÂNCIA POPULACIONAL (σ^2)

$$\text{Usaremos: } \sigma^2 = \frac{1}{N} \left[\sum_{i=1}^N Xi^2 Fi - \frac{(\sum_{i=1}^N Xi Fi)^2}{N} \right]$$

SAIBA MAIS!

Podemos calcular a variância através de outros processos de fórmulas teóricas. O modelo escolhido acima facilita os cálculos se lembrarmos de que os dados, tanto para variáveis discretas como para variáveis contínuas, estarão organizadas em tabelas de distribuições de frequências, as quais nos possibilitam encontrar os somatórios necessários com uma maior facilidade.

3.2.2 - VARIÂNCIA AMOSTRAL (S^2)

Usaremos

$$S^2 = \frac{1}{n-1} \left[\sum_{i=1}^N Xi^2 Fi - \frac{(\sum_{i=1}^N Xi Fi)^2}{n} \right]$$

Obs: Os somatórios que aparecem nas duas fórmulas acima serão encontrados através das tabelas de distribuições de frequências, conforme veremos no ex.abaixo.

Exemplo: considerando a série de dados a seguir como sendo uma amostra de 14 valores referentes a uma variável discreta qualquer: 4, 2, 5, 3, 4, 3, 3, 2, 5, 6, 5, 3, 2, 4; poderemos encontrar a variância desta série da seguinte maneira:

(Xi)	(Fi)	XiFi	Xi²Fi
2	3	6	12
3	4	12	36
4	3	12	48
5	3	15	75
6	1	6	36
Σ	14	10	207

Perceba você que

$X_i \rightarrow$ São os valores que apareceram na série de dados

$F_i \rightarrow$ São as frequências com que cada observação aparece na série

$X_i F_i \rightarrow$ São resultantes do produto entre a coluna do X_i pela coluna do F_i

$X_i^2 F_i \rightarrow$ São resultantes do produto entre a coluna do X_i^2 pela coluna do

F_i

Assim, poderemos substituir na fórmula os resultados dos somatórios encontrados na tabela de distribuição de frequência acima, da seguinte maneira:

$$S^2 = \frac{1}{n-1} \left[\sum_{i=1}^N X_i^2 F_i - \frac{(\sum_{i=1}^N X_i F_i)^2}{n} \right] = \frac{1}{14-1} \left[207 - \frac{(51)^2}{14} \right] = 1,63$$

3.3 DESVIO-PADRÃO

O desvio-padrão é um resultado consequente da variância. É uma das medidas de dispersão mais utilizadas. Através do desvio-padrão, podemos ter algumas informações iniciais a partir de um conjunto de dados. Também poderemos ter o desvio padrão tanto para população, como para amostra. Iniciaremos pelo desvio padrão populacional.

3.3.1 - DESVIO PADRÃO POPULACIONAL (σ)

Pela própria representação do desvio-padrão, percebemos que o desvio-padrão é resultante da raiz quadrada da variância, ou seja: $\sigma = \sqrt{\sigma^2}$.



VOCÊ SABIA?

60% a 80% das observações de um conjunto de dados encontram-se no intervalo de um desvio padrão em torno da média do conjunto. Ainda 100% das observações do conjunto se encontram no intervalo de três desvios em torno da média aritmética do conjunto, ou seja, $\bar{X} \pm 3S$.

3.3.2 - DESVIO PADRÃO AMOSTRAL (S)

De maneira semelhante iremos encontrar o desvio padrão amostral, através da raiz quadrada da variância amostral, ou seja: $S = \sqrt{S^2}$.

Finalizaremos as aplicações das medidas de dispersão, conhecendo o coeficiente de variação, para que possamos atribuir uma classificação ao conjunto de dados.

3.4 COEFICIENTE DE VARIAÇÃO (C.V)

Definição: O trata-se da relação entre uma medida de dispersão (o desvio padrão) e uma medida de posição (a média aritmética).

Através do resultado do coeficiente de variação, poderemos ainda atribuir uma classificação a um conjunto de dados da seguinte forma:

Quando	$C.V \leq 10\%$	→	BAIXA DISPERSÃO
	$10\% < C.V < 20\%$	→	MÉDIA DISPERSÃO
	$C.V \geq 20\%$	→	ALTA DISPERSÃO

em que usaremos $CV = \frac{S}{\bar{X}} \cdot 100$. Aplicaremos a seguir esta medida.

EXERCÍCIO RESOLVIDO 3

Voltando ao Ex. dos preços da mercadoria cobradas em 50 estabelecimentos comerciais:

Preços	Fi	Xi	XiFi	Xi²Fi
2 5	8	3,5	28	98
5 8	10	6,5	65	422,5
8 11	9	9,5	85,5	812,25
11 14	12	12,5	150	1875
14 17	11	15,5	170,5	2642,75
Σ	50		496	5850,5

- Classifique a dispersão dos preços.
- Encontre o intervalo que garante os preços cobrados por 60% a 80% dos estabelecimentos.

Solução:

a) Conforme vimos anteriormente, podemos classificar a dispersão dos preços, através do resultado do Coeficiente de Variação Assim teríamos

$$C.V = \frac{S}{\bar{X}} \cdot 100. \text{ Sabemos que } \bar{X} = \frac{\sum_{i=1}^n X_i F_i}{n} = \frac{499}{50} = 9,98. \text{ E ainda:}$$

$S = \sqrt{S^2}$. A variância amostral será dada por

$$S^2 = \frac{1}{n-1} \left[\sum_{i=1}^N X_i^2 F_i - \frac{(\sum_{i=1}^N X_i F_i)^2}{n} \right] =$$
$$= \frac{1}{49} \left[5850,5 - \frac{(496)^2}{50} \right] = 18,98 \rightarrow S = \sqrt{18,98} = 4,4. \text{ Assim, teríamos}$$

$CV = \frac{4,4}{9,98} \cdot 100 = 44,1\%$ (os preços são cobrados com uma alta dispersão entre os estabelecimentos comerciais).

b) Como vimos anteriormente, o intervalo que garante 60% a 80% será dado por

$$\bar{X} \pm S = 9,98 \pm 4,4 = [R\$5,58 \leftrightarrow R\$14,38].$$

Nesta aula, aprendemos a organizar dados quantitativos em tabelas de distribuições de frequências, para posteriormente usarmos diversos métodos de estatística descritiva, para sintetizar a posição e a variabilidade da distribuição e conseguirmos fazer uma leitura descritiva com uma maior qualidade.

Na aula 5, aplicaremos exercícios que servirão para recordarmos o que foi visto nas aulas anteriores, relembrando com detalhes os tópicos explorados nessas aulas. Esta revisão com exercícios nos qualificará para uma melhor compreensão e facilitará o início do estudo da Inferência Estatística, assunto da aula 7.



ATIVIDADES DE APROFUNDAMENTO!

- 1) Crie uma variável discreta, simule 40 resultados dessa variável e, em seguida:
 - a) Organize os dados em uma tabela de frequência absoluta, acumulada e relativa
 - b) Encontre e interprete o resultado de duas medidas de posição e duas de dispersão
- 2) Considere os mesmos valores da questão anterior como sendo relativos a uma variável contínua e, em seguida:
 - a) Organize os dados em uma distribuição de frequência com amplitude de cada classe com tamanho 3
 - b) Interprete o resultado do septuagésimo percentil
- 3) Qual das duas variáveis criadas anteriormente melhor representam a sua média aritmética?
- 4) Qual das duas distribuições de frequências anteriores possuem uma maior dispersão?

AULA 5

Miscelânea de exercícios resolvidos

Olá caro(a) aluno(a),

Nesta aula, recordaremos, através de exercícios resolvidos, as principais informações das últimas quatro aulas. Elas envolverão desde a introdução dos cálculos das probabilidades (com os conceitos probabilísticos do tipo espaço amostral, eventos e os principais tipos de eventos), até os resultados advindos da definição de probabilidade e os conceitos de estatística descritiva. Faremos exercícios sobre os modelos de distribuições discretas e contínuas de probabilidade, e sobre as séries estatísticas e estatística descritiva.

Então, vamos fazer uma revisão?

Objetivos

- Recordar, através de situações-problema, os principais teoremas e modelos de distribuições probabilísticas discretas e contínuas
- Diferenciar os principais tipos séries estatísticas
- Usar situações para organizar dados em distribuições de frequência
- Calcular e interpretar as principais medidas estatísticas de posição e de dispersão

TÓPICO 1

Exercício sobre cálculos de probabilidade

OBJETIVOS

- Resolver problemas que envolvam os principais teoremas e conceitos de probabilidade
- Revisar problemas que diferenciem a variável aleatória discreta da variável aleatória contínua

Daremos início à resolução dos problemas. Neste tópico, você recordará a teoria dos cálculos das probabilidades e consequentemente os conceitos, como eventos, teorema do produto de Bayes, variáveis aleatórias discretas e contínuas. Vamos analisar esses exercícios?

EXERCÍCIO RESOLVIDO 1

Em um determinado bairro residencial, 65% dos lares têm TV por assinatura, 72% têm internet e 40% assinam TV e internet. Se selecionarmos um lar aleatoriamente, qual a probabilidade de este lar assinar:

- a. Pelo menos uma das duas opções.
- b. Exatamente uma das opções.
- c. Nenhuma das opções.

ATENÇÃO!



Veja que aqui usamos o teorema apresentado no tópico 1 da aula 1 em que se falava:

Se A e B são dois eventos quaisquer, então:

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

VAMOS ÀS SOLUÇÕES:

Iremos transformar as informações do problema em eventos. Assim, devemos ter os seguintes eventos:

T: "assinam TV".

I: "assinam internet".

$T \cap I$: "assinam TV e internet".

Com as seguintes probabilidades fornecidas no enunciado da questão, temos $P(T)=0,65$, $P(I)=0,72$ e $P(T \cap I)=0,4$. Então podemos encontrar as probabilidades desejadas da seguinte maneira:

- a. A probabilidade pedida “pelo menos uma das duas opções” tem como interpretação ocorrer somente uma opção ou ocorrer somente a outra opção ou ocorrer as duas opções. O que na verdade se resumiria na possibilidade de ocorrer a união dos dois eventos. Então:

$$P(T \cup I) = P(T) + P(I) - P(T \cap I) = 0,65 + 0,72 - 0,4 = 0,92$$

- b. Neste item, a probabilidade pedida “exatamente uma das opções” pode ser interpretada como sendo ocorre a 1ª opção e não ocorre a 2ª, ou ocorre a 2ª e não ocorre a 1ª. Em termos simbólicos, poderíamos escrever assim:

$$P(T \cup I^c) = P(T) + P(I^c) - P(T \cap I^c), \text{ em que } I^c \text{ é o complemento do evento.}$$

Como os dois eventos T e I são mutuamente exclusivos, então não existe intersecção entre os dois eventos. Assim a probabilidade se reduziria assim $P(T \cup I^c) = P(T) + P(I^c) = 0,65 + 0,28 = 0,93$

- c. Para a resolução deste item, quando temos “nenhuma das opções”, sabemos que está se pedindo a probabilidade de não ocorrer a 1ª

opção e não ocorrer a 2ª opção, ou seja, queremos $P(T^c \cap I^c)$. Vimos, pelos Teoremas 1 e 3 da Aula 1, que

$$P(T^c \cap I^c) = P(T \cup I)^c = 1 - P(T \cup I) = 1 - 0,92 = 0,08.$$



ATENÇÃO!

Lembre-se de que definimos na aula 1 (tópico 1) que dois eventos são mutuamente exclusivos quando não podem ocorrer simultaneamente. Assim, se A e B são mutuamente exclusivos, então $A \cap B = \emptyset$.

EXERCÍCIO RESOLVIDO 2

Um laboratório do IFCE tem 24 computadores e 4 impressoras. O departamento técnico do Instituto só pode realizar serviços de manutenção em cinco equipamentos de cada setor por semana. Em uma determinada semana, um técnico pegará aleatoriamente os cinco equipamentos do laboratório para realizar os serviços necessários, logo qual a probabilidade de que:

GUARDE BEM ISSO!



Você lembra que a definição de probabilidade é muito clara quando diz que a probabilidade de um evento A ocorrer dentro de um universo é calculado pela seguinte fórmula matemática:

$$P(A) = \frac{\text{nº de casos do evento A}}{\text{nº total de casos}}.$$

- a. O técnico leve três computadores.
- b. O técnico leve todas as impressoras do laboratório.

IREMOS ENTÃO ÀS SOLUÇÕES:

a. Iremos formular o evento A: “o técnico leva três computadores”. Então, aplicando a definição de probabilidade, teríamos:

$$P(A) = \frac{\text{nº de casos do evento A}}{\text{nº total de casos}} = \frac{n(A)}{n(S)}.$$

Sendo $n(A) \Rightarrow$ nº de casos do evento A, será obtido a partir do seguinte questionamento: se o técnico leva 3 computadores é porque ele vai levar 2 impressoras. Então, de quantas maneiras o técnico poderá selecionar 3 entre os 24 computadores e 2 entre as 4 impressoras? Lembrando a análise combinatória, teremos

$$C_{24,3} = \frac{24!}{3!(24-3)!} = \frac{24!}{3!21!} = \frac{24 \cdot 23 \cdot 22 \cdot 21!}{3 \cdot 2 \cdot 1 \cdot 21!} = 2024. \text{ E ainda}$$

$$C_{4,2} = \frac{4!}{2!(4-2)!} = \frac{4!}{2!2!} = \frac{4 \cdot 3 \cdot 2!}{2 \cdot 1 \cdot 2!} = 6.$$

E se ainda temos $n(S) \Rightarrow$ nº total de casos do espaço amostral S será obtido a partir do questionamento a seguir: de quantas maneiras o técnico poderá selecionar os 5 equipamentos dentre os 28 existentes no laboratório?

$$C_{28,5} = \frac{28!}{5!(28-5)!} = \frac{28!}{5!23!} = \frac{28 \cdot 27 \cdot 26 \cdot 25 \cdot 24 \cdot 23!}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \cdot 23!} = 98280$$

$$\text{Assim, a } P(A) = \frac{C_{24,3} \times C_{4,2}}{C_{28,5}} = \frac{2024 \times 6}{98280} = 0,12$$
$$P(B) = \frac{C_{4,4} \times C_{24,1}}{C_{28,5}} = \frac{1 \times 24}{98280} = 0,00024$$

Formulemos então o evento B: “o técnico leva as 4 impressoras”. Então, aplicando a definição de probabilidade de um evento, teremos:

$$P(B) = \frac{\text{nº de casos do evento B}}{\text{nº total de casos}} = \frac{n(B)}{n(S)}$$
$$n(S) = 98280$$

Sendo $n(B) \Rightarrow$ nº de casos do evento B, será obtido a partir do seguinte questionamento: se o técnico leva 4 impressoras é porque ele vai levar 1 computador. Então, de quantas maneiras o técnico poderá selecionar 4 entre as 4 impressoras e 1 entre os 24 computadores? Lembrando a análise combinatória, teremos:

$$C_{4,4} = \frac{4!}{4!(4-4)!} = \frac{4!}{4!0!} = 1, \text{ e ainda}$$

$$C_{24,1} = \frac{24!}{1!(24-1)!} = \frac{24!}{1!23!} = \frac{24 \cdot 23!}{1 \cdot 23!} = 24.$$

Devemos lembrar que $n(S)$ é o mesmo do item a), visto que estamos no mesmo experimento, assim $n(S) = 98280$.

$$P(B) = \frac{C_{4,4} \times C_{24,1}}{C_{28,5}} = \frac{1 \times 24}{98280} = 0,00024$$

A seguir você verá, respectivamente, dois exercícios que tratem do teorema do produto e do teorema de Bayes.

EXERCÍCIO RESOLVIDO 3

Em um determinado encontro de estudantes de Matemática e Estatística, havia 12 de Matemática que não tiveram reprovação durante sua vida acadêmica e 7 de Estatística que tiveram alguma reprovação. Ainda havia 4 de Estatística que não tiveram reprovação durante sua vida acadêmica e 6 de Matemática que tiveram alguma reprovação. Selecionou-se um estudante desse encontro. Se o estudante é de Estatística, qual a probabilidade de que ele nunca tenha sofrido reprovação?

VAMOS À SOLUÇÃO:

Inicialmente formularemos eventos para facilitar a compreensão do problema:

M: “o estudante é de Matemática”.

F: “o estudante é de Estatística”.

S: “o estudante sofreu reprovação”.

NS: “o estudante não sofreu reprovação”.

Note que a probabilidade é solicitada sobre alguma condição, portanto podemos representar a probabilidade pedida da seguinte forma:

$$P(NS/E) = \frac{P(NS \cap E)}{P(E)}.$$

Notamos ainda que, como os dois eventos NS e E não são independentes, então: $P(NS \cap E)$ pelo teorema do produto, pode ser escrito assim:

$$P(NS \cap E) = P(NS) \times P(E / NS) = \frac{1}{2} \times \frac{4}{11} = \frac{4}{22}$$

Agora pense no cálculo da probabilidade do estudante ser de estatística. Neste caso, levamos em consideração todas as possibilidades de selecionar aleatoriamente um estudante de estatística. Assim, teremos

$P(E) = P(S \cap E) \cup P(NS \cap E) = P(S \cap E) + P(NS \cap E)$. Também pelo teorema do produto, seguiremos:

$$= P(S) \times P(E / S) + P(NS) \times P(E / NS) = \frac{1}{2} \times \frac{7}{11} + \frac{1}{2} \times \frac{4}{11} = \frac{11}{22}.$$

$$\text{Assim, } P(NS / E) = \frac{\frac{4}{22}}{\frac{11}{22}} = \frac{4}{11}.$$

EXERCÍCIO RESOLVIDO 4

Estima-se que 85% das crianças até 5 anos de um município foram tratadas com uma vacina contra uma determinada doença. A probabilidade de uma criança ficar curada dessa doença é de 1 em 26, caso não faça um tratamento para curar dessa doença. Porém, esta probabilidade será de 1 em 2, usando o tratamento. Se uma criança desse município portadora dessa doença ficar curada, qual será a probabilidade desta criança ter tomado a vacina?

Solução:

Vamos iniciar formulando os seguintes eventos:

T: “a criança é tratada com a vacina”.

T^c: “a criança não é tratada com a vacina”.

C: “a criança fica curada”.

Lembrando a probabilidade condicional e do Teorema de Bayes, podemos encontrar a probabilidade solicitada da seguinte forma:

$$P(T / C) = \frac{P(T \cap C)}{P(C)} = \frac{P(T)P(C / T)}{P(C)} = \frac{0,85 \times 0,5}{P(C)}.$$

Assim, $P(C)$ a probabilidade de a criança ser curada é dada por

$$\begin{aligned} P(C) &= P(C \cap T) + P(C \cap T^c) \\ &= P(T)P(C / T) + P(T^c)P(C / T^c) = 0,85 \times \frac{1}{2} + 0,15 \times \frac{1}{26} = 0,43. \end{aligned}$$

ATENÇÃO!

Os conceitos de probabilidade condicional e o Teorema de Bayes podem ser vistos no tópico 1 da aula 1. Se houver qualquer dúvida, retorne a essa aula para melhor compreensão desses conceitos.

O exercício resolvido 5 tratará dos dois conceitos estudados na aula 1 – tópico 1: distribuição de probabilidade e função repartição.

EXERCÍCIO RESOLVIDO 5

Um ponto de vendas tem sete impressoras similares, das quais três apresentam algum defeito. Se uma pessoa chegar a este ponto de vendas para fazer uma compra aleatória de duas dessas impressoras, então:

- Determine a distribuição de probabilidade para o nº de defeitos.
- Encontre a função repartição.

VAMOS À SOLUÇÃO.

Primeiro, vamos considerar a variável aleatória X , cujos valores x serão as possíveis impressoras com defeito pela pessoa. Então x poderá ser 0, 1, e 2. Então:

$$f(0) = P(X=0) = \frac{C_{3,0}C_{4,2}}{C_{7,2}} = \frac{\frac{3!}{0!3!} \times \frac{4!}{2!2!}}{\frac{7!}{2!5!}} = \frac{6}{21}$$

$$f(1) = P(X=1) = \frac{C_{3,1}C_{4,1}}{C_{7,2}} = \frac{\frac{3!}{1!2!} \times \frac{4!}{1!3!}}{\frac{7!}{2!5!}} = \frac{12}{21}$$

$$f(2) = P(X=2) = \frac{C_{3,2}C_{4,0}}{C_{7,2}} = \frac{\frac{3!}{2!1!} \times \frac{4!}{0!4!}}{\frac{7!}{2!5!}} = \frac{3}{21}$$

Assim, teríamos a seguinte distribuição de probabilidade de X :

x	0	1	2
$f(x)$	$\frac{6}{21}$	$\frac{12}{21}$	$\frac{3}{21}$

A função repartição será dada da seguinte forma:

- $F(0) = P(X \leq 0) = P(X=0) = \frac{6}{21}$.
- $F(1) = P(X \leq 1) = P(X=0) + P(X=1) = \frac{6}{21} + \frac{12}{21} = \frac{18}{21}$.
- $F(2) = P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = \frac{6}{21} + \frac{12}{21} + \frac{3}{21} = 1$.

O último exercício resolvido desse tópico tratará do conceito de função densidade e função distribuição de probabilidade. Se tiver dúvidas, retorne à aula 1 para estudar esses conceitos.

EXERCÍCIO RESOLVIDO 6

$$\text{Seja } f(X) = \begin{cases} kx & \text{se } 0 < x \leq 2 \\ 0 & \text{se } x \leq 0 \text{ ou } x > 2. \end{cases}$$

Determine

- O valor de k para que $f(X)$ seja uma função densidade de probabilidade.
- A função distribuição de probabilidade dada por $P(0 \leq X < \frac{2}{3})$.

VAMOS RESPONDER ESSE PROBLEMA?

Para que a função seja uma f.d.p, $\int_{-\infty}^{+\infty} f(x)dx = 1$, teremos

$$\int_0^2 kx dx = 1 \Rightarrow k \left(\frac{x^2}{2} \right)_0^2 = 1 \Leftrightarrow k = \frac{1}{2}.$$

A função distribuição de probabilidade será

$$P(0 \leq X < \frac{2}{3}) = \int_0^{\frac{2}{3}} \frac{x}{2} dx = \frac{x^2}{4} \Big|_0^{\frac{2}{3}} = \frac{1}{9}.$$

Caro aluno, neste primeiro tópico, recordamos, através de alguns exercícios resolvidos, conceitos e teoremas da teoria de probabilidade vista na aula 1. Esperamos que você tenha aproveitado esta revisão para compreender melhor este assunto tão importante. Continuaremos nosso repasso no próximo tópico. Até lá!

TÓPICO 2

Exercícios: modelos de distribuições probabilísticas

OBJETIVOS

- Rever os principais modelos de distribuições discretas para resolver situações probabilísticas diversas
- Aplicar a distribuição uniforme em problema de probabilidade que envolva variável contínua
- Transformar a distribuição normal em uma distribuição normal padronizada, para resolver situações probabilísticas que envolvam variáveis contínuas

Agora, aluno, analisaremos os exercícios resolvidos relacionados aos modelos de distribuições discretas e contínuas de probabilidade. Eles servem para solucionar problemas relacionados a experimentos probabilísticos associados tanto a variáveis discretas como a variáveis contínuas. Teremos aqui, no total, cinco exercícios resolvidos para você estudar e se aprofundar nesse assunto.

EXERCÍCIO RESOLVIDO 1

Um levantamento mostra que 73% das pessoas que concluem o curso de professor-tutor, chegam a atuar na função antes de 1 ano da conclusão do curso.

- Determine a probabilidade de que oito pessoas que estão concluindo o curso de professor-tutor, três atuem como tal antes de 1 ano após a conclusão deste curso.
- Calcule a probabilidade de que, num grupo de quatro concludentes do curso observados, mais da metade venha a atuar como professor-tutor.



SAIBA MAIS!

Esperamos que você pesquise exercícios em outras fontes de estudo, pois Matemática aplicada a qualquer área do conhecimento, em especial Estatística e Probabilidade, deve ser um estudo constante, além da prática de exercícios. Indicamos a você o site Só Matemática (<http://www.somatematica.com.br/estat/basica/exercicios.php>), no qual poderá encontrar mais exercícios relacionados ao assunto de Estatística.

VAMOS ÀS RESPOSTAS!

Notamos que o problema envolve as seguintes características:

1º) Existem dois possíveis resultados a todo instante: a pessoa concludente do curso pode conseguir ou não atuar na área no prazo estipulado.

2º) A probabilidade de a pessoa conseguir atuar é sempre constante conforme informa o problema.

3º) Uma pessoa conseguir atuar não dependerá de que outra tenha conseguido, visto que as tentativas são independentes.

4º) A probabilidade é pedida sempre observando certo número de pessoas.

Assim, estamos diante de um problema que seguem todas as características do modelo de distribuição binomial. Então vamos utilizá-lo para buscar as respostas desejadas.

a. Seja a variável aleatória

X: “nº de pessoas que irão atuar antes de 1 ano do término do curso”. Assim teremos que encontrar a no item $P(X=3)=?$. Aplicando o modelo binomial visto no item 1.1 da aula-2, teremos

ATENÇÃO!

Como você já deve ter visto na aula 2, sobre distribuições discretas e contínuas de probabilidade, a distribuição binomial depende diretamente dos resultados advindos da probabilidade de um evento ocorrer com n repetições, em k tentativas. Lembrando-nos aqui da seguinte fórmula matemática:

$$P(X=k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}.$$

$$P(X=k) = \binom{n}{k} \cdot p^k \cdot q^{n-k} \Rightarrow P(X=3) = \binom{8}{3} \cdot 0,73^3 \times 0,27^5 = 0,031.$$

b. Neste item, é solicitada a seguinte probabilidade:

$$\begin{aligned} P(X \geq 3) &= P(X=3) + P(X=4) = \binom{8}{3} \cdot 0,73^3 \times 0,27^5 + \\ &+ \binom{8}{4} \cdot 0,73^4 \times 0,27^4 = 0,031 + 0,106 = 0,137. \end{aligned}$$

Agora vamos trabalhar e relembrar os conceitos de distribuição hipergeométrica e de Poisson que se seguirão nos dois exercícios resolvidos a seguir.

EXERCÍCIO RESOLVIDO 2

Iremos formar um comitê de 3 alunos selecionados de um total de 5 estudantes de Matemática e 4 de Estatística. Qual a probabilidade de este comitê.

- a. ser formado somente por estatístico.
- b. conter dois matemáticos.

VAMOS BUSCAR AS RESPOSTAS?

- a. Observe agora que este problema tem as seguintes características:
 - 1º) Existem dois possíveis resultados a todo instante (o estudante é de Estatística ou não).
 - 2º) A probabilidade de um estudante ser de Estatística vai variando a todo instante em que vamos selecionando um aluno para compor o comitê.
 - 3º) O 2º estudante selecionado ser de Estatística dependerá do 1º selecionado ter sido também de Estatística, pois assim a chance poderá ou não diminuir. Ou seja, o que ocorre numa etapa de seleção depende do que ocorreu numa etapa anterior, logo influenciará numa etapa seguinte.
 - 4º) A probabilidade é pedida sempre observando certo n° de estudantes. Assim, estamos diante de um problema que segue todas as características do modelo de distribuição hipergeométrica. Então vamos utilizá-lo para buscar a resposta pedida.
- Primeiro definiremos a seguinte variável aleatória que irá contar o número desejado de sucesso:
- X: “n° de estudantes de Estatística”

$$P(X=k) = \frac{\binom{nx}{k} \cdot \binom{N-nx}{n-k}}{\binom{N}{n}} \Rightarrow P(X=3) = \frac{\frac{4!}{3!1!} \times \frac{5!}{0!5!}}{\frac{9!}{3!6!}} = 0,048.$$

- b. Neste item, o problema envolve as mesmas características. Apenas vamos definir outra variável aleatória X: “n° de estudantes de matemática”. Assim, teremos

$$P(X=2) = \frac{\binom{5}{2} \cdot \binom{9-5}{3-2}}{\binom{9}{3}} = \frac{\binom{5}{2} \cdot \binom{4}{1}}{\binom{9}{3}} = \frac{\frac{5!}{2!3!} \times \frac{4!}{1!3!}}{\frac{9!}{3!6!}} = 0,476.$$

EXERCÍCIO RESOLVIDO 3

Sabendo-se que, a cada duas horas, cinquenta alunos acessam o sistema virtual Moodle durante um determinado curso da EaD, calcule a probabilidade de que, em 10 minutos observados, no máximo dois alunos acessem o sistema virtual?

Solução:

Aluno, você deve observar que o problema traz uma característica particular. É solicitada uma probabilidade de ocorrência para um determinado tempo, e nos é informada uma quantidade de ocorrência para outro tempo, ou seja, é fornecido um coeficiente de proporcionalidade. Desta forma, podemos chegar à solução do problema através da distribuição de Poisson. Vejamos:

A variável aleatória X será definida por

X : “nº de acesso ao sistema virtual”

$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$. Para aplicarmos a distribuição de Poisson, devemos encontrar o coeficiente de proporcionalidade λ . Então faremos

$$P(X \leq 2) = \frac{\ell^{-4,2} \times 4,2^0}{0!} + \frac{\ell^{-4,2} \times 4,2^1}{1!} + \frac{\ell^{-4,2} \times 4,2^2}{2!} = 0,21.$$

EXERCÍCIO RESOLVIDO 4

ATENÇÃO!

O conceito de distribuição uniforme se encontra na aula 2 – tópico 2 que trata da definição deste conceito e de outros como distribuição normal e distribuição normal padrão.

Escolhendo-se um valor aleatoriamente do intervalo $[0, 3]$, qual a probabilidade de que esse valor esteja entre 1 e 2?

VAMOS ENCONTRAR A RESPOSTA:

$$P(1 \leq X \leq 2) = ?$$

Sabemos pela distribuição uniforme que

$$\begin{cases} f(x) = \frac{1}{b-a}, & \text{se } a \leq x \leq b; \\ 0, & \text{caso contrário.} \end{cases}$$

$$\text{Assim, } f(x) = \begin{cases} \frac{1}{3-0}, & \text{se } 0 \leq x \leq 3 \\ 0, & \text{se } x < 0 \text{ ou } x > 3. \end{cases} = \begin{cases} \frac{1}{3}, & \text{se } 0 \leq x \leq 3 \\ 0, & \text{se } x < 0 \text{ ou } x > 3 \end{cases} \Leftrightarrow$$

$$P(1 \leq X \leq 2) = \int_1^2 \frac{1}{3} dx = \frac{1}{3}(x)_1^2 = \frac{1}{3}.$$

EXERCÍCIO RESOLVIDO 5

Sabendo-se que as baterias de uma determinada marca para celulares têm suas durações distribuídas normalmente com vida média de dois anos e desvio padrão de 80 dias, responda:

- Qual a probabilidade de que uma determinada bateria desta mesma marca tenha vida útil maior do que 780 dias?
- O fabricante desta marca de bateria oferece uma garantia de um ano. Se 6 mil baterias são fabricadas mensalmente, quantas deverão ser trocadas mensalmente?

VAMOS À SOLUÇÃO!

Como estamos estudando a variável contínua, tempo de duração, que está normalmente distribuída, iremos aplicar a distribuição normal de probabilidade. Então, por meio da curva normal:

$P(X > 780) = ?$, através da padronização da curva pelo escore-Z da distribuição normal padronizada, teremos que

$$Z = \frac{X - \mu}{\sigma} = \frac{780 - 730}{80} = 0,63. \text{ Assim:}$$

$P(X > 780) = P(Z > 0,63) = 1 - P(Z \leq 0,63)$. Fazendo uso da tabela da normal padrão, encontramos que $P(Z \leq 0,63) = 0,74$. Então, teremos $1 - 0,74 = 0,26$.

Como a garantia fornecida pelo fabricante é de 1 ano (365 dias), ou seja, a troca das baterias que apresentarem falhas nesse período seria dada por:

$$P(X < 365) = ?$$

$$Z = \frac{X - \mu}{\sigma} = \frac{365 - 730}{80} = -0,19$$

$$P(X < 365) = P(Z < -0,19) = 0,5 - 0,0753 = 0,4247.$$

Realizamos, nesse tópico, aplicações de todos os modelos de distribuições discretas e contínuas de probabilidade para resolvermos diversas situações probabilísticas. Agora, aluno, daremos início às aplicações da estatística descritiva. Primeiramente iremos recordar o estudo das séries estatísticas. Vamos lá, então?

TÓPICO 3

Exercício: séries estatísticas

OBJETIVOS

- Recordar a forma correta de representar as séries através de tabelas
- Diferenciar os principais tipos de séries estatísticas através de exercícios
- Lembrar a forma correta de escolher os gráficos para realizar as representações das séries

Aluno, iniciaremos este tópico com as aplicações dos principais modelos de séries, para posteriormente representarmos estes modelos através de tabelas e gráficos, assunto que foi estudado na aula 3. Aqui, veremos exercícios que nos lembrem como diferenciar os principais tipos de séries estatísticas e como representá-las através de tabelas ou gráficos da melhor forma possível.

EXERCÍCIO RESOLVIDO 1

Crie uma situação que envolva o nosso curso a distância. Ela deve ser identificada como uma série estatística do tipo categórica-geográfica.

EXEMPLO:

Você deve estar lembrado que o conceito da série estatística geográfica ou de localização foi definido como uma série em que, na sua representação, só ocorre variação na localidade da ocorrência. Logo, neste tipo de série, os dados serão agrupados, segundo localidades distintas da ocorrência. Notaremos também que a época da ocorrência e o fenômeno ocorrido ficarão fixos (Aula 3 – tópico 1).

Ainda temos o conceito de série categórica ou específica em que aquela em que a variação só ocorre no próprio fenômeno ocorrido. Ou seja, nesta série estatística, a época e o local da ocorrência permanecerão fixos, enquanto os dados que se referem ao fenômeno ocorrido serão agrupados de acordo com a modalidade da ocorrência do fenômeno (Aula 3 – tópico 1).

Sabemos que, podemos combinar as séries estatísticas. E como o enunciado pede para fazer uma identificação das séries categórica-geográfica. Então teremos a seguinte situação:

CLASSIFICAÇÃO DAS NOTAS	POLOS				
	A	B	C	D	E
BAIXA
MÉDIA
ALTA

Veja que na série existiriam alunos para as três categorias de notas, por isso ela é categórica. E também existem resultados para mais de uma localidade (polos), por isso ela também se classifica como Geográfica.

EXERCÍCIO RESOLVIDO 2

Realize alteração na série estatística anterior, de tal forma que passemos a ter uma série que também assuma uma característica de uma série cronológica.

Solução:

CLASSIFICAÇÃO DAS NOTAS	POLOS				
	A	B	C	D	E
	2009 2010	2009 2010	2009 2010	2009 2010	2009 2010
BAIXA
MÉDIA
ALTA

Note que, agora além da variação das categorias das notas e da localidade destas, também temos na série resultados para mais de um período, ou seja, também temos uma variação de época.

EXERCÍCIO RESOLVIDO 3

Simule valores para a série estatística anterior e escolha um gráfico para em seguida representá-la.
Vamos lá!

Nº DE ALUNOS POR CLASSIFICAÇÃO DAS NOTAS EM CADA POLO
(2009-2010)

CLASSIFICAÇÃO DAS NOTAS	POLOS				
	A	B	C	D	E
	2009-2010	2009-2010	2009-2010	2009-2010	2009-2010
BAIXA	49 21	25 30	08 15	04 02	10 05
MÉDIA	25 12	06 12	13 06	28 20	19 09
ALTA	10 05	02 09	12 10	07 02	06 14

Como se trata de uma série mista, usaremos o gráfico de coluna para representação mostrado na Figura 1.

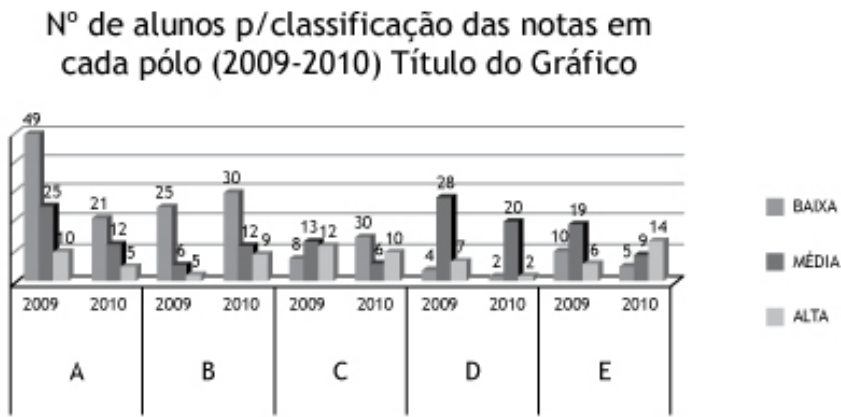


Figura 1: Gráfico de coluna representando os dados da tabela

EXERCÍCIO RESOLVIDO 4

Verifique, na série estatística abaixo, quais foram os erros cometidos nas representações.

Solução:

Nº DE VENDAS DO PRODUTO - X (2007-2010)

ANOS	Nº DE VENDAS
2006	450
2007	543
2008	602
2009	716
2010	512

Nº DE VENDAS DO PRODUTO - X (2006-2010)

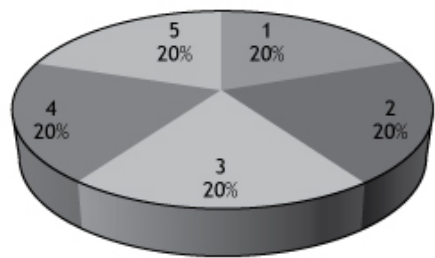


Figura 2: Gráfico representando os dados da tabela

Verificamos que o título está incompleto tanto no quadro (visto que as tabelas não devem conter o fechamento nas laterais), como no gráfico. E ainda, por se tratar de uma série cronológica, o gráfico correto seria um gráfico em linha.

Neste tópico, recordamos a forma correta de representarmos as séries estatísticas, bem como as diferenças entre os principais tipos de séries. Os exercícios resolvidos foram direcionados a gráficos e tabelas, nos quais trabalhamos com dados fictícios. No próximo e último tópico, iremos abordar exercícios resolvidos que envolvam o assunto de estatística descritiva.

TÓPICO 4

Exercício: estatística descritiva

OBJETIVOS

- Lembrar, através de aplicações, a forma correta de representarmos dados quantitativos em tabelas de distribuições de frequências
- Aplicar problemas que envolvam as principais medidas de posição e de dispersão

Nesse tópico, você verá a exposição de exercícios resolvidos que envolvam conceitos como tipos de variáveis, medidas de posição e de dispersão que irão descrever algum fenômeno estudado.

Então vamos aos exercícios.

EXERCÍCIO RESOLVIDO 1

Simule 35 valores referentes ao nº de vendas realizadas em um determinado estabelecimento comercial durante 40 dias observados, e em seguida:

- a. Organize os dados em uma tabela de distribuição de frequência.
- b. Encontre o nº médio de vendas realizadas.
- c. Interprete o resultado da mediana.

Faremos da seguinte forma:

- a. Supondo que os 40 valores fossem os abaixo informados:

2, 5, 4, 1, 3, 2, 2, 7, 3, 5, 6, 2, 7, 2, 3, 1, 7, 5, 2, 6
5, 7, 4, 5, 3, 2, 5, 4, 1, 2, 4, 6, 7, 5, 4, 2, 1, 2, 1, 4.

Assim, poderemos dispor da seguinte forma:

Nº de vendas	Nº de dias(Fi)	Fac	Fi
1	5	5	0,125
2	10	15	0,250
3	4	19	0,100
4	6	25	0,150
5	7	32	0,175
6	3	35	0,075
7	5	40	0,125
Σ	35		1

- b. Iremos encontrar o nº médio de vendas da seguinte forma:

$$\bar{X} = \frac{\sum_{i=1}^n X_i F_i}{n}.$$

Na distribuição de frequência, como a variável é discreta, chamaremos os valores da variável de X_i , então iremos inserir uma nova coluna na tabela de distribuição:

Nº de vendas(X_i)	Nº de dias(F_i)	$X_i F_i$	f_i
1	5	5	0,143
2	10	20	0,286
3	4	12	0,114
4	6	24	0,171
5	7	35	0,200
6	3	18	0,085
7	5	35	
Σ	40	149	1

$$\bar{X} = \frac{\sum_{i=1}^n X_i F_i}{n} = \frac{149}{35} = 4,2$$

- c. Como a variável é discreta e observamos um n° impar de observações, a mediana será o elemento que assuma a ordem $\frac{n+1}{2}$, ou seja, $\frac{35+1}{2} = 18^\circ$ o qual pela coluna das frequências acumuladas será igual a 3, logo podemos dizer que em 50% dos dias observados foram realizadas no máximo 3 vendas.

EXERCÍCIO RESOLVIDO 2

Tomando o exercício anteriormente resolvido como base, criamos esta nova situação: qual seria o n° de vendas máximo para 80% dos dias?

Solução:

Aluno, como queremos um ponto que irá limitar um percentual, iremos aplicar a separatriz. Para isso, deveremos calcular $P_{80} = ?$

Usaremos a seguinte relação:

$$X = (n-1) \times \frac{p}{100} + 1 = (40-1) \times \frac{80}{100} + 1 = 32,2 \cong 32^\circ = 5$$

EXERCÍCIO RESOLVIDO 3

Supondo os dados do exercício 1 como sendo referentes às taxas de juros cobradas por 35 instituições financeiras, organize os dados em uma distribuição de frequência. Depois interprete o resultado do sexagésimo terceiro percentil.

VAMOS RESOLVER?

Os dados são os seguintes:

2, 5, 4, 1, 3, 2, 2, 7, 3, 5, 6, 2, 7, 2, 3, 1, 7, 5, 2, 6
5, 7, 4, 5, 3, 2, 5, 4, 1, 2, 4, 6, 7, 5, 4, 2, 1, 2, 1, 4.

Como a variável Taxa de juros é contínua, portanto deveremos primeiro organizar os dados em uma tabela de frequência. Para isso, deveremos encontrar a amplitude que definirá os intervalos de classes. Usaremos o 1° método visto na aula-4 da seguinte forma: $n = 40 > 25 \Rightarrow k \cong \sqrt{40} \cong 7$.

$h \cong A_t \div k \Rightarrow A_t = 7 - 1 = 6 \Leftrightarrow h \cong 6 \div 7 \cong 1$. Assim, poderemos realizar a representação:

Taxas	Nº de instituições (Fi)	Fac
1 2	5	5
2 3	10	15
3 4	4	19
4 5	6	25
5 6	7	32
6 7	3	35
7 8	5	40
Σ	40	

Agora poderemos encontrar o sexagésimo terceiro percentil, assim:

$$P_i = \ell + \frac{\left(\frac{63n}{100} - \sum f\right) \times h}{FP63}.$$

Iremos primeiramente identificar a classe percentilica, através do elemento

$$\frac{63n}{100} = \frac{63 \times 40}{100} = 25,2. \text{ Pela coluna da Fac, vimos que esse elemento se}$$

encontra na classe $5 | 6$.

Então agora iremos identificar na tabela de distribuição de frequência todos os elementos que compõem a fórmula. Assim, teríamos

$$P63 = 5 + \frac{(25,2 - 25) \times 1}{7} = 5,03.$$

Podemos dizer que 63% das instituições cobram no máximo 5,03% de juros ou ainda que 37% das instituições possuem as maiores taxas de juros e cobram no mínimo 5,03% de juros.

EXERCÍCIO RESOLVIDO 4

Seja a seguinte distribuição de frequência das velocidades médias apresentadas por 50 veículos em certo percurso observado:

Velocidade	Fi
60 65	3
65 70	15
70 75	20
75 80	10
80 85	2
Σ	50

- Qual a velocidade média apresentada pela maioria dos veículos?
- As velocidades ocorreram com uma alta dispersão?



ATENÇÃO!

Os conceitos de mediana, moda e separatrizes foram estudados no tópico 2 da aula 4. Caso tenha alguma dúvida, retorne a essa aula e estude com mais atenção esses conceitos.

BOM, CARO ALUNO, VAMOS ÀS SOLUÇÕES:

- Vimos durante o estudo das principais medidas de posição que por definição a moda é o elemento que mais ocorre. Neste exemplo, a velocidade apresentada pela maioria é encontrada da seguinte maneira:

$$Mo = \ell + \frac{\Delta_1}{\Delta_1 + \Delta_2} xh$$

Primeiro vamos identificar a classe modal, para depois encontrarmos os elementos que

compõem a fórmula. Assim, a classe modal é a de maior frequência

70 | 75. Então teríamos $Mo = \ell + \frac{\Delta_1}{\Delta_1 + \Delta_2} xh = 70 + \frac{5}{5 + 10} \times 5 = 71,7$, em que

$\ell \Rightarrow$ limite inferior da classe

$\Delta_1 \Rightarrow 20 - 15 = 05$

$\Delta_2 \Rightarrow 20 - 10 = 10$

$h \Rightarrow$ Diferença entre os limites da classe (75-70=5)

Para classificarmos a dispersão de uma distribuição, faz-se necessário encontrar o coeficiente de variação (CV), então escreveremos

Velocidade	Fi	Xi	XiFi
60 65	3	62,5	187,5
65 70	15	67,5	1012,5
70 75	20	72,5	145,0
75 80	10	77,5	775
80 85	2	82,5	165
Σ	50		3590

$C.V = \frac{S}{\bar{X}} \cdot 100$, em que

$$\bar{X} = \frac{\sum_{i=1}^n XiFi}{n} = \frac{3590}{50} = 71,8$$

$$S = \sqrt{S^2}, \text{ em que } S^2 = \frac{1}{n-1} \left[\sum_{i=1}^N Xi^2Fi - \frac{(\sum_{i=1}^N XiFi)^2}{n} \right] = \frac{1}{49} \left[5850,5 - \frac{(496)^2}{50} \right] = 18,98$$

$$S = \sqrt{18,98} = 4,4. \text{ Assim, } CV = \frac{4,4}{71,8} = 6,13\%.$$

Como $CV \leq 10\%$, temos uma baixa dispersão das velocidades médias dos veículos.

EXERCÍCIO RESOLVIDO 5

Ainda com base nos dados do exercício anterior, e sabendo-se que outro grupo de 30 veículos apresentou um desvio-médio de 1,2, pergunta-se: qual dos dois grupos de veículos melhor representa a sua velocidade média?

VAMOS À RESPOSTA:

Vimos no tópico 3 da aula 4 que, quanto menor o resultado do desvio-médio, melhor será a representação de seu resultado médio. Com base nisso, vamos calcular o desvio-médio do grupo dos 50 veículos:

Velocidade	Fi	Xi	$ Xi - X \times Fi$
60 ┆ 65	3	62,5	27,9
65 ┆ 70	15	67,5	64,5
70 ┆ 75	20	72,5	14
75 ┆ 80	10	77,5	57
80 ┆ 85	2	82,5	21,4
Σ	50		184,8

$$DM = \sum_{i=1}^N \frac{|Xi - \bar{X}|}{N} . Fi$$

$$= \frac{184,8}{50} = 3,7 .$$

Assim, como $1,2 < 3,7$, podemos dizer que o grupo de 30 veículos melhor representa a sua velocidade média.

Bom, caro aluno, com esta aplicação, recordamos de toda parte da organização de dados em tabelas de distribuição de frequência e também dos objetivos das principais medidas de posição e de dispersão.

Assim, com esta aula 5, fizemos um resumo das quatro primeiras aulas que envolveram desde a introdução aos cálculos de probabilidades, até um resumo da estatística descritiva, tais como a organização das informações em tabelas de distribuições, os cálculos de medidas e as análises destas medidas.

Na próxima aula, daremos início ao estudo da inferência estatística que irá inferir sobre parâmetros populacionais, tomando por base as noções de probabilidade e os conhecimentos da estatística descritiva.

AULA 6

Correlação e regressão

Olá aluno(a),

Até este momento, estudamos os cálculos das probabilidades e da estatística descritiva, o qual irá contribuir bastante para uma melhor compreensão do estudo da inferência estatística.

Iniciaremos, nesta aula, o estudo da inferência estatística – com o apoio da estatística descritiva e das noções de ocorrências probabilística – o qual realiza estimativas e testa resultados para parâmetros populacionais com base em levantamentos amostrais. Começaremos pelo estudo da correlação e regressão entre variáveis quantitativas.

Então vamos iniciar a aula conhecendo o coeficiente de correlação de Pearson e analisando o seu resultado.

Objetivos

- Conhecer o coeficiente de correlação de Pearson
- Compreender o comportamento linear entre duas variáveis
- Encontrar o modelo de regressão linear simples, utilizado para estimar valores para uma variável dependente em função de uma variável independente

TÓPICO 1

Correlação

OBJETIVOS

- Interpretar o resultado do coeficiente de correlação de Pearson
- Analisar o coeficiente de determinação



SAIBA MAIS!

Sabia que Karl Pearson nasceu em 27 de março de 1857 em Londres? Ele foi um matemático que contribuiu para o desenvolvimento da Estatística. Mais detalhes da vida desse matemático você pode encontrar no site da Universidade Federal de Campina Grande: <http://www.dec.ufcg.edu.br/biografias/KarlPear.html>.

Veremos, neste tópico, como obter e analisar o resultado do coeficiente de Pearson.

Logo depois, iremos analisar o coeficiente de determinação. Neste momento, deverá ficar claro que a variável independente X explicará o comportamento de uma variável dependente Y .

Então, vamos ao estudo?

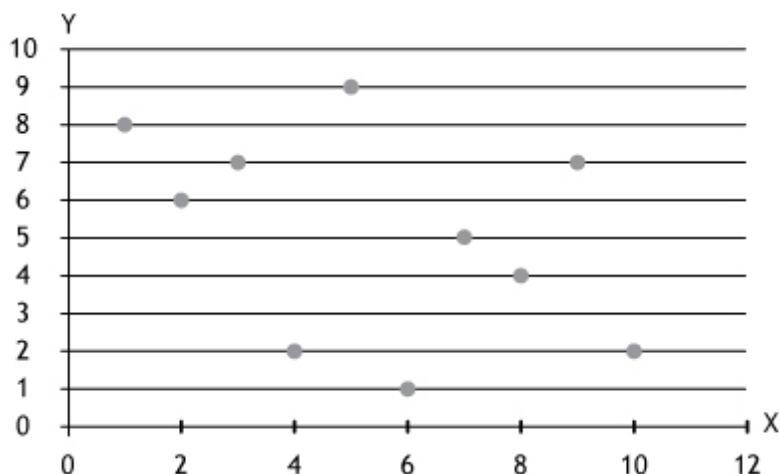
1.1 COEFICIENTE DE CORRELAÇÃO DE PEARSON

Definição: Trata-se de um coeficiente que avalia o grau de relação linear entre duas variáveis quantitativas. Através do coeficiente de correlação de Pearson (r), poderemos ver o quanto duas variáveis quantitativas X e Y estarão correlacionadas.

O coeficiente de correlação estará sempre entre -1 e 1, ou seja, $-1 \leq r \leq 1$. Quando r se aproxima de 1, dizemos que existe uma forte correlação positiva, porém, quando r se aproxima de -1, existe uma forte correlação negativa.

Assim, poderemos ter as seguintes situações:

1º CASO:

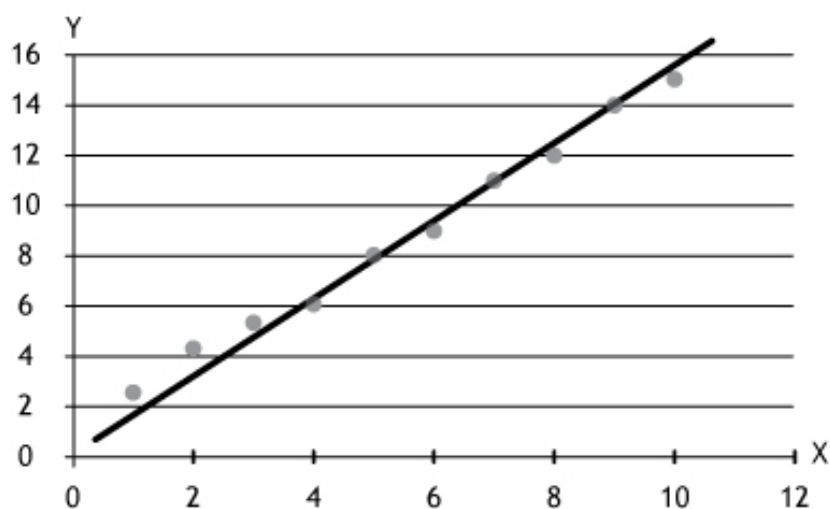


$r = 0$ (Não existe correlação)

Notemos, através dos comportamentos de cada ponto obtido no gráfico que não existe relação linear entre as duas variáveis X e Y.

Veja agora situações em que duas variáveis estariam correlacionadas.

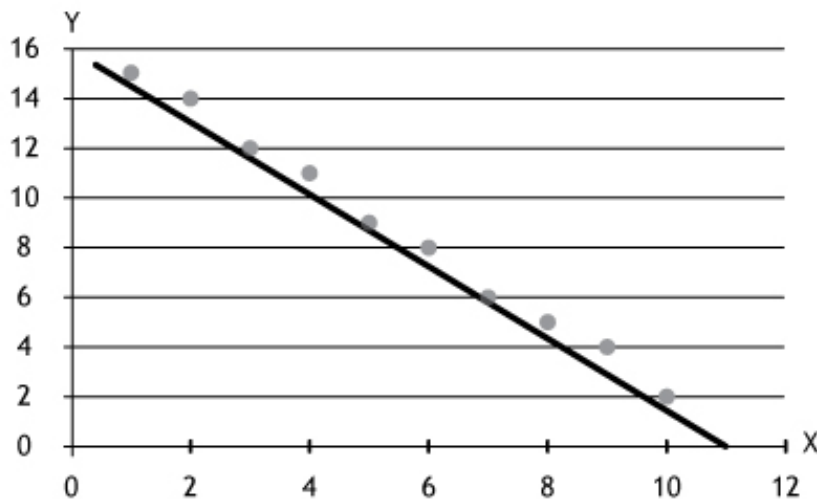
2º CASO:



$r > 0$ (Existe uma correlação positiva)

Neste caso, temos uma forte correlação positiva, ou seja, à medida que a variável X cresce, a variável Y também crescerá, ocorrendo então uma linearidade positiva entre as duas variáveis.

3º CASO:



$r < 0$ (Existe uma correlação negativa)

Enquanto a variável X cresce, notamos que a variável Y estará decrescendo, mostrando assim uma tendência a uma linearidade negativa entre as duas variáveis.

A seguir, você verá o cálculo do coeficiente de correlação de Pearson que será realizado da seguinte maneira:

Iremos inicialmente definir as duas variáveis X e Y. Sendo que

X \Rightarrow é a variável independente.

Y \Rightarrow é a variável dependente. Assim, teremos

$$r = \frac{S_{x,y}}{\sqrt{S_{x,x} \cdot S_{y,y}}}, \text{ em que}$$

$$S_{x,y} = \sum XY - \frac{\sum X \cdot \sum Y}{N}$$

$$S_{x,x} = \sum X^2 - \frac{(\sum X)^2}{N}$$

$$S_{y,y} = \sum Y^2 - \frac{(\sum Y)^2}{N}$$

Vamos à resolução de um exercício.

EXERCÍCIO RESOLVIDO 1

Encontre o coeficiente de correlação de Pearson, com base nos dados abaixo referentes à quantidade de salários ganhos e ao tempo de estudo em anos em uma amostra de 12 pessoas:

Tempo de estudo (X)	Nº de salários (Y)	XY	X ²	Y ²
4	2	8	16	4
0	1	0	0	1
13	8	104	169	64
2	1	2	4	1
2	1	2	4	1
9	5	45	81	25
15	9	135	225	81
11	7	77	121	49
8	4	32	64	16
10	6	60	100	36
5	2	10	25	4
20	10	200	400	100
99	56	675	1209	382

Solução:

Primeiramente, temos de chamar de X a coluna da variável independente e chamar de Y a coluna da variável dependente. Posteriormente, iremos encontrar os somatórios que serão necessários para encontrarmos o coeficiente de Pearson abaixo. São eles

$$\sum X, \sum Y, \sum XY, \sum X^2 \text{ e } \sum Y^2$$

Agora substituir os valores dos somatórios nas fórmulas abaixo:

$$r = \frac{S_{X,Y}}{\sqrt{S_{X,X} \cdot S_{Y,Y}}}$$

Sabemos que

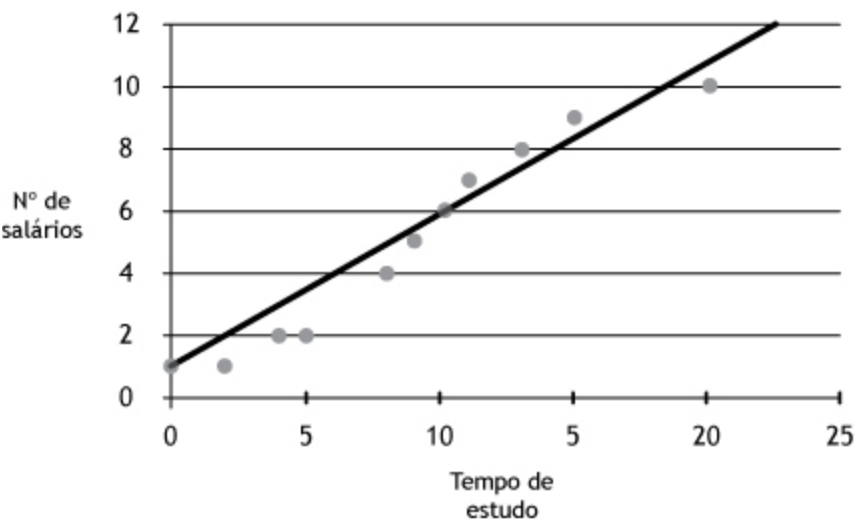
$$S_{X,Y} = \sum XY - \frac{\sum X \cdot \sum Y}{N} = 675 - \frac{99 \times 56}{12} = 213$$

$$S_{X,X} = \sum X^2 - \frac{(\sum X)^2}{N} = 1209 - \frac{(99)^2}{12} = 392,25$$

$$S_{Y,Y} = \sum Y^2 - \frac{(\sum Y)^2}{N} = 382 - \frac{(56)^2}{12} = 120,7. \text{ Assim, teremos:}$$

$$r = \frac{S_{X,Y}}{\sqrt{S_{X,X} \cdot S_{Y,Y}}} = \frac{213}{\sqrt{392,25 \times 120,7}} = 0,98.$$

Então, podemos dizer que as variáveis estão fortemente correlacionadas de forma positiva, ou seja, quanto maior o tempo de estudo das pessoas, é de se esperar que maior seja o nº de salários delas. Tal comportamento pode ser visto no gráfico abaixo:



EXERCÍCIO RESOLVIDO 2

Agora vejamos se existe correlação entre o nº de erros cometidos por 12 alunos em uma prova e o tempo de estudo(h) que cada aluno disponibiliza diariamente.

Tempo de estudo(h) (X)	Nº de erros (Y)	XY	X²	Y²
2	7	14	4	49
1	8	8	1	64
5	4	20	25	16
2	6	12	4	36
3	5	15	9	25
6	1	6	36	1
1	9	9	1	81
2	7	14	4	49
3	4	12	9	16
4	5	20	16	25
7	0	0	49	0
1	10	10	1	100
37	66	140	159	462

Solução:

Analogamente, após encontrarmos os resultados dos somatórios necessários para efetuarmos os cálculos do coeficiente de Pearson. Faremos isso da seguinte maneira:

$$S_{x,y} = \sum XY - \frac{\sum X \cdot \sum Y}{N} = 140 - \frac{37 \times 66}{12} = -63,5$$

$$S_{x,x} = \sum X^2 - \frac{(\sum X)^2}{N} = 159 - \frac{(37)^2}{12} = 44,9$$

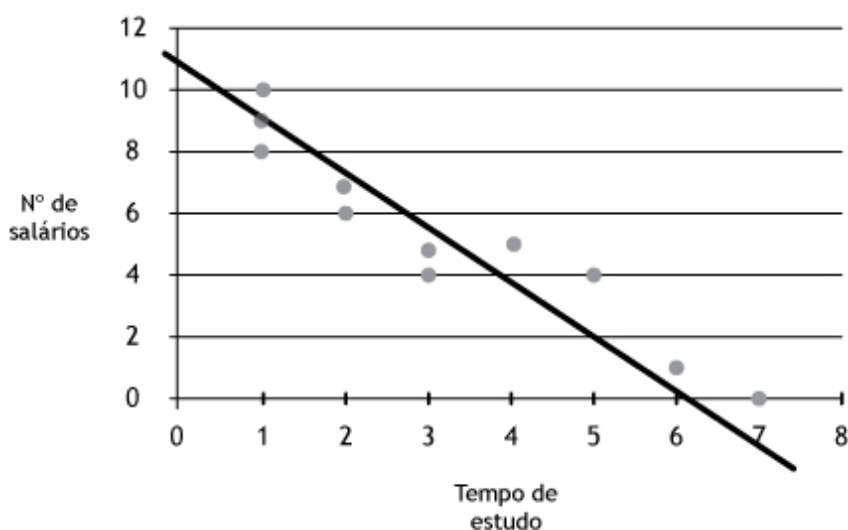
$$S_{y,y} = \sum Y^2 - \frac{(\sum Y)^2}{N} = 462 - \frac{(66)^2}{12} = 99.$$

Desta forma, caro aluno(a), teremos o seguinte resultado final do coeficiente de Pearson:

$$r = \frac{S_{x,y}}{\sqrt{S_{x,x} \cdot S_{y,y}}} = \frac{-63,5}{\sqrt{44,9 \times 99}} = -0,95.$$

Notamos, neste caso, que ocorreu uma forte correlação negativa, ou seja, quanto menor foi o tempo de estudo dedicado pelos alunos diariamente, maior foram os números de erros cometidos por eles na prova. Percebeu a diferença para o exercício anterior?

Vejamos então através do comportamento gráfico a seguir:



1.2 COEFICIENTE DE DETERMINAÇÃO

Também podemos encontrar, a partir do coeficiente de correlação r , o coeficiente de determinação r^2 , o qual irá nos mostrar o quanto uma variável independente X explicará do comportamento de uma variável dependente Y .

Dessa forma, voltemos ao exercício resolvido 2 da correlação entre

TEMPO DE ESTUDO DIÁRIO

x

Nº DE ERROS COMETIDOS NA PROVA

Você lembra que, nesse exercício, foi encontrado um coeficiente de correlação de Pearson r igual a $-0,95$. Então sabemos que o coeficiente de determinação $r^2 = (-0,95)^2 = 0,90 = 90\%$. Logo podemos dizer que a variável independente tempo de estudo diário explica em 90% a variação dos resultados da variável dependente nº de erros na prova.

Neste tópico, conhecemos as possíveis correlações que existem entre duas variáveis quantitativas. Daremos início agora ao estudo do modelo de regressão, o qual servirá para estimarmos valores de variáveis dependentes em função de valores de variáveis independentes.

TÓPICO 2

Regressão

OBJETIVOS

- Rever os principais modelos de distribuições discretas para resolver situações probabilísticas diversas
- Aplicar a distribuição uniforme em problema de probabilidade que envolva variável contínua
- Transformar a distribuição normal em uma distribuição normal padronizada, para resolver situações probabilísticas que envolvam variáveis contínuas

Você compreenderá nesse tópico a construção da Equação de Regressão Linear Simples (E.R.L.S), a partir dos seus conhecimentos adquiridos no tópico anterior.

Então? Vamos lá?

MODELO DE EQUAÇÃO DE REGRESSÃO LINEAR SIMPLES

Iremos encontrar um modelo de equação de regressão linear simples que possa ser usado para estimar valores de variáveis dependentes Y , através de resultados de variáveis independentes X . O modelo de E.R.L.S será descrito da seguinte maneira:

$Y = \alpha + \beta X$, em que

$\left\{ \begin{array}{l} \alpha \Rightarrow \text{será a constante da equação, ou seja, o intercepto da reta no eixo } Y. \\ \beta \Rightarrow \text{será o coeficiente angular que mostrará a inclinação da reta definindo} \end{array} \right.$

se a correlação entre as duas variáveis quantitativas ocorre de forma positiva ou negativa.

Veja que o modelo de E.R.L.S acima seria aplicado em caso de estarmos trabalhando com todos os resultados populacionais possíveis para duas variáveis quantitativas. Estimando-se um modelo de E.R.L.S com base em dados amostrais, teríamos a seguinte equação estimada:

$\bar{Y} = a + bX$, em que

$\bar{Y} \Rightarrow$ será o estimador de Y .

$a \Rightarrow$ será o estimador de α .

$b \Rightarrow$ será o estimador de β .

Inicialmente, devemos encontrar, com base na amostra levantada, os resultados de a e b . O procedimento é o seguinte:

$$a = \bar{Y} - b\bar{X} \text{ e } b = \frac{S_{x,y}}{S_{x,x}}.$$

Ainda $\bar{X} \Rightarrow$ é a média da variável X, ou seja, $\bar{X} = \frac{\sum X}{n}$.

$\bar{Y} \Rightarrow$ é a média da variável Y, ou seja, $\bar{Y} = \frac{\sum Y}{n}$.

Vamos realizar uma aplicação para compreender melhor como funciona.

EXERCÍCIO RESOLVIDO 3

Sejam as seguintes notas obtidas em uma avaliação final e o tempo de estudo (h) por semana disponível fora da escola para uma amostra de 15 alunos:

TEMPO DE ESTUDO(h) P/SEMANA	NOTAS
3	1
5	3
9	6
11	8
10	7
6	4
6	5
8	6
4	2
6	4
9	7
11	8,5
9	5
5	4
7	5

- Encontre o modelo de E.R.L.S
- Estime a nota para um aluno que disponibiliza 14 h semanais para estudar
- Mostre através de um gráfico de dispersão o intercepto do eixo Y.

Vamos à solução desse problema!

SOLUÇÃO DO ITEM A

Sabemos que a E.R.L.S é dada por $\bar{Y} = a + bX$. Assim, vamos precisar dos seguintes resultados: $\sum X$, $\sum Y$, $\sum XY$, $\sum X^2$ e. Devemos inicialmente definir qual variável é X e qual variável é Y. Se lembrarmos que Y é a variável dependente e X é a variável independente, então, no nosso exercício, definiremos

Y = nota

X = tempo de estudo semanal

Podemos agora encontrar os nossos somatórios:

TEMPO DE ESTUDO(h) P/SEMANA (X)	NOTAS (Y)	XY	Y ²	X ²
3	1	3	1	9
5	3	15	9	25
9	6	54	36	81
11	8	88	64	121
10	7	70	49	100
6	4	24	16	36
6	5	30	25	36
8	6	48	36	64
4	2	8	4	16
6	4	24	16	36
9	7	63	49	81
11	8,5	93,3	72,25	121
9	5	45	25	81
5	4	20	16	25
7	5	35	25	49
109	75,5	620,3	443,25	881

Agora podemos substituir os somatórios para calcularmos os valores de a e b, através de suas fórmulas:

$$b = \frac{S_{x,y}}{S_{x,x}}$$

$$S_{x,y} = \sum XY - \frac{\sum X \cdot \sum Y}{N} = 620,3 - \frac{109 \times 75,5}{15} = 71,7$$

$$S_{x,x} = \sum X^2 - \frac{(\sum X)^2}{N} = 881 - \frac{(109)^2}{15} = 88,9$$

Assim, teremos

$$b = \frac{S_{x,y}}{S_{x,x}} = \frac{71,7}{88,9} = 0,81$$

$$\bar{X} = \frac{\sum X}{n} = \frac{109}{15} = 7,3$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{75,5}{15} = 5,03$$

$$a = \bar{Y} - b\bar{X} = 5,03 - 0,81 \times 7,3 = -0,88.$$

Poderemos então definir a nossa E.R.L.S: $\bar{Y} = a + bX \Rightarrow \bar{Y} = -0,88 + 0,81X$.

SOLUÇÃO DO ITEM B

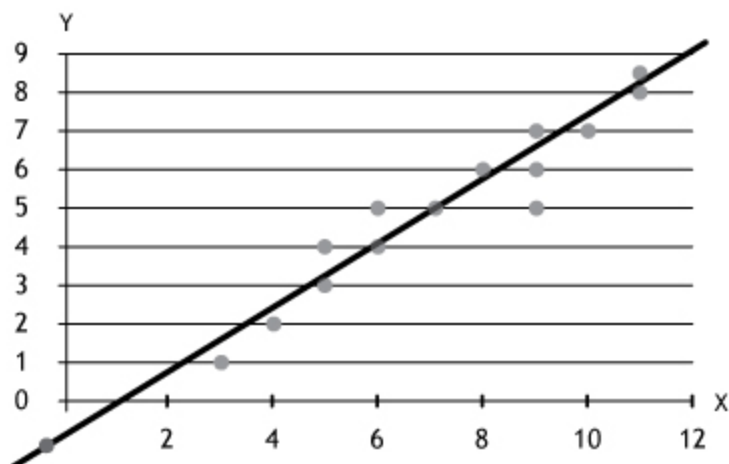
Para estimarmos a nota de um aluno que dedique 14h por semana ao estudo, devemos substituir na E.R.L.S estimada o valor de X por 14. Então teremos

$$\bar{Y} = -0,88 + 0,81X, \text{ ou seja,}$$

$$\bar{Y} = -0,88 + 0,81 \times 14 \cong 10.$$

SOLUÇÃO DO ITEM C

Aqui é pedido o intercepto do eixo Y. Assim primeiro devemos marcar no gráfico de dispersão os pontos correspondentes (X,Y).



Através da E.R.L.S vista no item a), podemos notar que o intercepto da equação dado pela (constante - a), quando o valor da variável independente $X = 0$, é exatamente -0,88.

Neste tópico, você aprendeu como construir um modelo de equação de regressão linear simples e como utilizá-lo como ferramenta da inferência estatística.

Depois desse estudo, finalizamos a nossa aula seis sobre correlação e regressão. Daremos continuidade ao estudo da inferência na próxima aula, conhecendo a estimação de parâmetros populacionais desconhecidos.



EXERCÍCIOS DE APROFUNDAMENTO!

1. Calcule e interprete o coeficiente de correlação entre as notas de 8 estudantes em duas disciplinas, selecionadas de uma turma aleatoriamente e mostradas a seguir:

Matemática	96,5	4	5,5	10	9	8,5	7	
Química	10	7	4,5	5	9	9,5	9,5	8

- Com base nos dados da questão anterior
 - Interprete o resultado do coeficiente de determinação
 - Encontre a E.R.L.S estimada
- Seja a seguinte amostra de 11 aplicações, referente a investidores, com seus lucros obtidos e seus respectivos tempos de aplicações, estime o lucro obtido por um investidor com três anos e meio de aplicação.

JUROS (R\$)	5,7	8,9	3,5	9,1	4,6	5,4	10,9	11,3	14,8	4,1	16,1
TEMPO DE APLICAÇÃO (anos)	4	6	1	7	5	4	8	9	11	2	12

- No exercício anterior, troque os valores da variável dependente de tal maneira que a correlação passe a ser negativa
- Mostre em forma de um gráfico e explique os resultados do intercepto dos coeficientes a e b da E.R.L.S do exercício 4.

AULA 7

Estimações

Olá aluno(a),

Na última aula, você viu e estudou as possíveis relações que existem entre duas variáveis quantitativas, através da correlação e da regressão, está lembrado? Agora daremos continuidade ao estudo da inferência estatística, realizando um dos objetivos básicos: a estimação de parâmetros populacionais desconhecidos.

Iniciaremos com as estimações de médias, logo depois prosseguiremos com as estimações de diferenças entre duas médias e finalizaremos com as estimações de proporções.

Então, vamos à aula?

Objetivos

- Conhecer a estimação da média populacional e a estimação das diferenças entre duas médias populacionais através de dados amostrais
- Aplicar a estimação de proporções populacionais desconhecidas

TÓPICO 1

Estimações de médias populares

OBJETIVO

- Investigar levantamentos amostrais, para realizarmos estimações de médias populacionais desconhecidas com certa confiabilidade

Iremos, com base em levantamentos amostrais, realizar estimações de médias populacionais desconhecidas, através de intervalos que irão nos proporcionar estas médias estimadas.

1.1 INTERVALO DE CONFIANÇA PARA MÉDIA

Uma maneira de expressarmos a precisão da estimação é mostrarmos os limites com os quais incluirão o verdadeiro valor do parâmetro populacional. Chamamos esses limites de **limites de confiança**, os quais irão determinar a formação do **intervalo de confiança**. Na sequência, iremos aprender como construir estes intervalos e realizaremos alguns exercícios de aplicação para estes intervalos.

Para compreendermos bem este assunto, devemos ter muito claro em mente que a aplicação do intervalo de confiança adequado se faz a partir de dois critérios:

1º CRITÉRIO:

Quando conhecemos o valor da variância populacional σ^2 , iremos aplicar o seguinte intervalo de confiança para estimação:

$$\bar{X} \pm Z_{\alpha} \sigma_x, \text{ em que}$$

$\bar{X} \Rightarrow$ É a média da amostra.

$Z_{\alpha} \Rightarrow$ É o valor de Z da tabela da normal padrão (estudado na aula 2 dessa disciplina), cujo valor de probabilidade é o mais próximo de $1 - \frac{\alpha}{2}$, no qual α é nível de significância.

Como nós estamos realizando estimação para um parâmetro populacional através de dados amostrais, temos sempre uma margem probabilística de confiança para os limites de estimação do intervalo de estimação. Assim, chamaremos de nível de significância α o complemento desta confiabilidade.

Por exemplo: se estivermos estimando com uma confiabilidade de 95%, o valor do nível de significância α será de 5%. Logo, neste caso, teríamos que $1 - \frac{\alpha}{2} = 1 - \frac{0,05}{2} = 0,975$.

Então, olhando este valor na tabela da normal padrão, veremos que o valor de $Z_{\alpha} = 1,96$.



ATENÇÃO!

Quando não soubermos o tamanho da população N, temos de considerar que a amostra não irá ultrapassar 5% da população.

$\sigma_x \Rightarrow$ É o desvio padrão estimado para a média, que será obtido da seguinte forma:

Quando $(n \leq 0,05N)$, ou seja, quando o tamanho da amostra que está sendo trabalhada for menor do que 5% do tamanho da população, teremos que $\sigma_x = \frac{\sigma}{\sqrt{n}}$, em que σ é o desvio padrão populacional

Quando $(n > 0,05N)$, ou seja, quando o tamanho da amostra for maior do que 5% do tamanho da população, teremos que $\sigma_x = \frac{\sigma}{\sqrt{n}} \left(\frac{N-n}{N-1} \right)$.

Vamos então a uma aplicação deste primeiro critério:

EXERCÍCIO RESOLVIDO 1

Desejando estimar o nº médio semanal de acesso ao sistema virtual Moodle dos alunos da EaD do curso de Matemática, selecionamos uma amostra de 40

alunos e verificamos que estes tiveram um acesso médio de 19,4 acessos ao sistema. Sabendo-se que os acessos semanais de todos os alunos deste curso ocorrem com uma variância de 2,2, realize a estimação desejada com uma confiabilidade de 90%.

Solução:

Primeiramente, vamos retirar do problema os dados fornecidos:

- Tamanho da amostra $n = 40$
- Média da amostra $\bar{X} = 19,4$
- Variância populacional $\sigma^2 = 2,2$
- Nível de significância $\alpha = 10\% = 0,1$.

Agora, vamos aplicá-los no intervalo de confiança de estimação da média que é dado por

$$\bar{X} \pm Z_{\alpha} \sigma_x^-$$

$$Z_{\alpha} = ?$$

Primeiramente calcularemos a relação $1 - \frac{\alpha}{2} = 1 - \frac{0,1}{2} = 0,95$. Depois iremos procurar na tabela da normal padrão (aula2) o valor de Z, cuja probabilidade seja a mais próxima de 0,95. Assim teremos $1 - \frac{\alpha}{2} = 1 - \frac{0,04}{2} = 0,98$



ATENÇÃO!

Os parâmetros que aparecem no texto \bar{X} , σ^2 e σ são medidas estatísticas que você, caro aluno(a), conheceu na Aula 4.

Tabela da Distribuição Normal Padrão

P(Z<z)	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633

Fonte: www.pucrs.br/format/rossana/psicologia/tabela_normal.pdf

Então, vimos que o valor de Z correspondente é 1,65, logo $Z_{\alpha} = 1,65$.

Na sequência, vamos encontrar o desvio padrão estimado para média σ_x . Considerando que a amostra não ultrapassa 5% do tamanho da população, teremos que $\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{2,2}}{\sqrt{40}} = 0,23$.

Agora vamos substituir os valores no intervalo de estimação $\bar{X} \pm Z_{\alpha} \sigma_x$
 $\Rightarrow 19,4 \pm 1,65 \times 0,23 = 19,4 \pm 0,38$. Então subtraindo e depois somando 0,38 de 19,4, teremos os seguintes limites do intervalo de estimação:

[19,02 até 19,78].

Assim, podemos concluir com 90% de confiança que o nº médio de acesso semanal ao sistema virtual *Moodle* é de no mínimo 19,02 e de no máximo 19,78.

2º CRITÉRIO:

Quando não conhecemos o valor da variância populacional σ^2 e tivermos uma amostra considerada pequena ($n < 30$), iremos aplicar o seguinte intervalo de confiança para estimação:

$\bar{X} \pm T_{\alpha(n-1)} S_x$, em que

$\bar{X} \Rightarrow$ É a média amostral

$T_{\alpha(n-1)} \Rightarrow$ É o valor de T da tabela da distribuição T-Student, em que

α é o nível de significância

$(n-1)$ é o grau de liberdade a ser visto na tabela

$S_x \Rightarrow$ É o valor do desvio padrão estimado da média quando não conhecemos o desvio padrão populacional σ , e segue o mesmo critério do σ_x .

Assim, quando ($n \leq 0,05N$), ou seja, o tamanho da amostra que está sendo trabalhada for menor do que 5% do tamanho da população, teremos que $S_x = \frac{S}{\sqrt{n}}$, na qual S é o desvio padrão populacional

Quando ($n > 0,05N$), ou seja, o tamanho da amostra for maior do que 5% do tamanho da população, teremos que $S_x = \frac{S}{\sqrt{n}} \left(\frac{N-n}{N-1} \right)$.

Vamos então a uma aplicação desse critério.

EXERCÍCIO RESOLVIDO 2

Suponha agora que estamos interessados em estimar o nº médio semanal de acesso ao sistema virtual *Moodle* dos alunos da EaD do curso de Matemática com

uma confiabilidade de 90% e que, para isso, tivéssemos apenas uma amostra de 26 alunos acusando um acesso médio ao sistema de 18,7 e um desvio padrão do nº de acesso semanal de 1,35. Assim, realizemos a nova estimação intervalar.

Solução:

Primeiro, retiremos as informações do novo problema:

Nível de significância $\alpha = 0,1$

Tamanho da amostra $n = 26$

Média amostral $\bar{X} = 18,7$

Desvio padrão amostral $S = 1,35$.

Assim, teremos $\bar{X} \pm T_{\alpha(n-1)}S_x^-$, em que

$$T_{\alpha(n-1)} = T_{0,1(26-1)} = T_{0,1(25)}.$$

Logo, poderemos encontrar na tabela da distribuição T- de Student abaixo

que $T_{0,1(25)} = 1,708$

Tabela t (student)

gl/ α	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
01	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
02	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
03	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,541	12,924
04	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
05	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
06	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
07	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,365	3,499	5,408
08	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
09	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792

23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,726
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,856	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,856	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
i	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Fonte: <http://www.somatematica.com.br/estat/tabelat.php>

Agora, vamos encontrar o desvio padrão estimado para média S_x^- . Seguindo o mesmo critério de determinação de σ_x^- , encontraremos

$$S_x^- = \frac{S}{\sqrt{n}} = \frac{1,35}{\sqrt{26}} = 0,265.$$

Substituindo os valores encontrados no intervalo de estimação para média, teremos

$$\bar{X} \pm T_{\alpha(n-1)} S_x^- = 18,7 \pm 1,708 \times 0,265 = 18,7 \pm 0,4 = [18,3 \text{ até } 19,1].$$

Assim, podemos concluir com 90% de confiança que o nº médio de acesso semanal ao sistema virtual Moodle é de no mínimo 18,3 e de no máximo 19,1.

Então, conseguiu entender ?

Agora, iremos estimar o quanto poderá variar a diferença entre duas médias. Vamos fazer isso juntos?

1.2 INTERVALO DE CONFIANÇA PARA DIFERENÇA ENTRE DUAS MÉDIAS

Neste subtópico, estimaremos a diferença entre as médias de duas populações. Semelhante à estimação de média simples, encontramos aqui também dois critérios:

1º CRITÉRIO

As variâncias populacionais σ_1^2 e σ_2^2 são conhecidas.

Para realizarmos a estimação da diferença entre as médias de duas populações, devemos selecionar duas amostras n_1 e n_2 destas populações. Assim, faremos a estimação intervalar, usando o seguinte intervalo de confiança:

$$\bar{X}_1 - \bar{X}_2 \pm Z_{\alpha} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \text{ em que}$$

$\overline{X}_1 \Rightarrow$ Média da amostra-1.

$\overline{X}_2 \Rightarrow$ Média da amostra-2.

$Z_\alpha \Rightarrow$ É o valor de Z da tabela normal padrão.

$\sigma_1^2 \Rightarrow$ É a variância da população-1.

$\sigma_2^2 \Rightarrow$ É a variância da população-2.

$n_1 \Rightarrow$ É o tamanho da amostra-1.

$n_2 \Rightarrow$ É o tamanho da amostra-2.

Vejamos o exercício resolvido abaixo para compreender melhor esse conceito.

EXERCÍCIO RESOLVIDO 3

Um professor afirma que as notas médias em sua disciplina, em duas instituições, são praticamente iguais. Para verificar se esta afirmação procede, um pesquisador seleciona uma amostra de 35 alunos de uma instituição e outra de 42 alunos da outra instituição. Ele encontrou que as notas médias amostrais são 7,3 e 7,5 respectivamente.

Realize a estimação da diferença entre as duas médias das populações de alunos, sabendo que as duas instituições apresentaram o desvio padrão das notas dos alunos que cursaram esta disciplina de 0,8 e 0,6 respectivamente. Adote uma confiabilidade de 96%

Solução:

Inicialmente, vamos retirar as informações do problema. Temos as seguintes informações:

$$n_1 = 35$$

$$n_2 = 42$$

$$\overline{X}_1 = 7,3$$

$$\overline{X}_2 = 7,5$$

$$\sigma_1^2 = (0,8)^2 = 0,64$$

$$\sigma_2^2 = (0,6)^2 = 0,36$$

$$\alpha = 0,04 .$$

Agora, vamos substituir o intervalo de confiança a ser calculado da seguinte maneira:

$$\overline{X}_1 - \overline{X}_2 \pm Z_\alpha \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} .$$

Calculando separadamente as três partes, teremos

$$(\overline{X_1} - \overline{X_2}) = 7,3 - 7,5 = -0,2$$

$$Z_\alpha = ?$$

Obs: para encontrarmos Z_α na tabela da normal padronizada, teremos primeiro que encontrar $1 - \frac{\alpha}{2} = 1 - \frac{0,04}{2} = 0,98$. Assim, na tabela da normal padronizada o valor de Z_α , será dado por $Z_\alpha = 2,05$

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{0,64}{35} + \frac{0,36}{42}} = 0,179.$$

$$\sigma_1^2 = (0,08)^2 = 0,64$$

$$\sigma_2^2 = (0,06)^2 = 0,36$$

Logo, o intervalo estimado para a diferença entre as duas médias será dado por $[-0,6 \text{ até } 0,2]$, considerando que a nota média de uma turma só pode assumir valores positivos, então o intervalo se resumiria a $[0 \text{ até } 0,2]$, ou seja, a diferença máxima entre as notas médias das duas instituições é de no máximo dois décimos. Sendo assim, podemos concluir, com uma confiabilidade de 96%, que o pesquisador não tem razão suficiente para suspeitar da informação do professor, visto que, com a confiabilidade da estimação, podemos perceber que praticamente não existe diferença significativa entre as médias das duas instituições.

2º CRITÉRIO

Se as variâncias populacionais σ_1^2 e σ_2^2 são desconhecidas, usaremos o seguinte intervalo de confiança para estimar a diferença entre duas médias populacionais:

$$\overline{X_1} - \overline{X_2} \pm t_\alpha(\varphi) \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, \text{ em que}$$

$$\overline{X_1} \Rightarrow \text{Média da amostra-1}$$

$$\overline{X_2} \Rightarrow \text{Média da amostra-2}$$

$$t_\alpha(\varphi) \Rightarrow \text{É o valor de T da tabela da}$$

distribuição T de Student com

φ graus de liberdade, em que

$$\varphi = \frac{(S_1^2 / n_1 + S_2^2 / n_2)^2}{\left[(S_1^2 / n_1)^2 / (n_1 - 1) \right] + \left[(S_2^2 / n_2)^2 / (n_2 - 1) \right]}$$

ATENÇÃO!

O valor de φ representa uma estimativa dos graus de liberdade, que não será sempre um número inteiro. Então devemos arredondar para baixo atingindo o inteiro mais próximo, para atingir a confiança desejada.

$S_1^2 \Rightarrow$ É a variância da amostra-1

$S_2^2 \Rightarrow$ É a variância da população-2

$n_1 \Rightarrow$ É o tamanho da amostra-1

$n_2 \Rightarrow$ É o tamanho da amostra-2

EXERCÍCIO RESOLVIDO 4

Suponha, no exercício resolvido 3, que a amostra de 35 alunos tenha apresentado uma média de 7,7 e um desvio padrão das notas de 1,4 e que a amostra de 42 alunos encontrou uma média de 7,2 e um desvio padrão de 1,2. Vamos verificar com uma confiabilidade de 95% a afirmação do professor de que não existe diferença entre as notas médias de sua disciplina nas duas instituições.

Solução:

como os desvios padrões informados se referem às duas amostras e não às duas populações, consideramos as variâncias populacionais desconhecidas, logo iremos aplicar o seguinte intervalo de estimação:

$$\overline{X}_1 - \overline{X}_2 \pm t_\alpha(\varphi) \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Vamos encontrar as três partes separadamente, ok?

$$\overline{X}_1 - \overline{X}_2 \Rightarrow (7,7 - 7,2) = 0,5$$

$$t_\alpha(\varphi).$$

Como vimos:

$$\begin{aligned}\varphi &= \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{\left[(S_1^2/n_1)^2/(n_1-1)\right] - \left[(S_2^2/n_2)^2/(n_2-1)\right]} \\ &= \frac{(1,4^2/35 + 1,2^2/42)^2}{\left[(1,4^2/35)^2/(35-1)\right] - \left[(1,2^2/42)^2/(42-1)\right]} \\ &= 1,2 \cong 1\end{aligned}$$

Então teremos de verificar na tabela o valor de $t_{0,05}(1) = 12,706$

$$\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \Rightarrow \sqrt{\frac{1,4^2}{35} + \frac{1,2^2}{42}} = 0,3.$$

Podemos agora substituir os três valores no intervalo de estimação da seguinte forma:

$$\overline{X}_1 - \overline{X}_2 \pm t_{\alpha}(\varphi) \times \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \Rightarrow 0,5 \pm 12,706 \times 0,3 = 0,5 \pm 3,81 = [-3,31 \text{ até } 4,31]$$

Lembrando que a nota média de uma turma só pode assumir valores positivos, então o intervalo se resumiria a $(0 \text{ até } 4,3)$. O pesquisador neste caso teria razão suficiente para achar que a informação do professor de que não existe diferença significativa entre as notas médias da disciplina para as duas instituições não é verdadeira.

Agora, iremos realizar estimações de proporções (porcentagens) populacionais desconhecidas, através de levantamentos amostrais, que será o nosso tema do próximo tópico.

TÓPICO 2

Estimações de proporções populacionais

OBJETIVO

- Investigar levantamentos amostrais a fim de estimarmos proporções populacionais desconhecidas com certa confiabilidade

Neste tópico, aprenderemos a estimar proporções populacionais desconhecidas, através de intervalos de confiança, tomando por base levantamentos amostrais. Vamos à aula.

2.1 INTERVALO DE CONFIANÇA PARA PROPORÇÃO

Como geralmente não conseguimos investigar uma população completa, iremos, com base em levantamentos amostrais, encontrar os limites de um intervalo que irão incluir o verdadeiro valor da porcentagem populacional desconhecida.

Para tanto, aplicaremos o seguinte intervalo de confiança para a estimação:

$p \pm Z_{\alpha} S_p$, em que

$p \Rightarrow$ É a proporção amostral

$$p = \frac{x}{n}.$$

Saiba que

$x \Rightarrow$ É o n° de casos na amostra

$n \Rightarrow$ É o tamanho da amostra

$Z_{\alpha} \Rightarrow$ É o valor de Z da tabela normal padrão

$S_p \Rightarrow$ É o desvio estimado para proporção, obtido por

$$S_p = \sqrt{\frac{pq}{n}} \text{ (para populações infinitas) ou}$$

$$S_p = \sqrt{\frac{pq}{n} \left(\frac{N-n}{N-1} \right)} \text{ (para populações finitas)}$$

$$\text{Ainda: } q = 1 - p$$

Então, vamos ver como funciona esta estimação de proporções populacionais desconhecidas a partir do próximo exercício resolvido:

EXERCÍCIO RESOLVIDO 5

Em uma amostra de 400 famílias de uma cidade-X, verifica-se que 128 utilizam da assinatura do serviço de internet. Determine o intervalo de confiança de 90% para a verdadeira proporção de famílias desta cidade que utilizam do serviço de internet.

Solução:

$$p \pm Z_\alpha S_p.$$

Primeiro vamos encontrar os três valores que compõem o intervalo, ok?

$$p = \frac{x}{n} = \frac{128}{400} = 0,32$$

$$Z_\alpha = ?$$

$$\text{Como vimos no tópico 1, devemos primeiro calcular } 1 - \frac{\alpha}{2} = 1 - \frac{0,1}{2} = 0,95.$$

Agora iremos procurar na tabela normal padrão (visto no tópico passado) a probabilidade mais próxima a 0,95, depois verificaremos que esta probabilidade é assumida pelo o valor $Z_\alpha = 1,65$

Como não temos uma população definida das famílias da cidade-X, iremos considerar a população como infinita, assim o desvio estimado para proporção será dado por $S_p = \sqrt{\frac{pq}{n}}$.

$$S_p = \sqrt{\frac{0,32 \times 0,68}{400}} = 0,023$$

Vamos então substituir os três valores no intervalo de estimação:

$$p \pm Z_\alpha S_p = 0,32 \pm 1,65 \times 0,023 = 0,32 \pm 0,04 = [0,28 \text{ até } 0,36], \text{ ou seja, a proporção}$$

de famílias do município-X que possuem serviço de internet, estimada com 90% de confiança, é de no mínimo 28% e de no máximo 36% das famílias do município.

Vejamos outro exemplo:

EXERCÍCIO RESOLVIDO 6

Em uma amostra de 310 escolas de uma cidade que possui um total de 1030 escolas, verificamos que 97 delas contam exclusivamente com professores graduados ou pós-graduados. Determine o intervalo de confiança de 95% para a real proporção de escolas desta cidade, que possuem o seu quadro contando somente com professores com as referidas qualificações.

Solução:

Como a estimação é de proporção, usaremos o referido intervalo de estimação

$$p \pm Z_{\alpha} S_p.$$

Primeiro vamos encontrar os três valores que compõem o referido intervalo:

$$p = \frac{x}{n} = \frac{97}{310} = 0,31$$

$$Z_{\alpha} = ?$$

Como vimos no tópico1, devemos primeiro calcular $1 - \frac{\alpha}{2} = 1 - \frac{0,05}{2} = 0,975$.

Agora iremos procurar na tabela normal padrão (visto no tópico passado) a probabilidade mais próxima a 0,975, então poderemos observar que a referida probabilidade é assumida pelo o valor $Z_{\alpha} = 1,96$.

Como temos uma população definida de 1030 escolas na cidade, iremos considerar a população como finita, assim o desvio estimado para proporção será

$$\text{dado por } S_p = \sqrt{\frac{pq}{n} \left(\frac{N-n}{N-1} \right)} = \sqrt{\frac{0,31 \times 0,69}{310} \left(\frac{1030-310}{1029} \right)} = 0,022.$$

Vamos então substituir os três valores no intervalo de estimação:

$p \pm Z_{\alpha} S_p = 0,31 \pm 1,96 \times 0,022 = 0,31 \pm 0,04 = [0,27 \text{ até } 0,35]$, ou seja, a proporção de escolas da cidade que contam com o quadro de professores

possuidores de graduação ou de pós-graduação, estimada com uma confiabilidade de 95%, é de no mínimo 27% e de no máximo 35% das escolas da cidade.

Espero que tenha compreendido!

Caro aluno, nesta aula, você conheceu e aplicou os mais importantes tipos de estimações de parâmetros populacionais desconhecidos. Na aula 8, iremos prosseguir com o estudo da inferência estatística, analisando os mais importantes testes estatísticos não-paramétricos. Então, até lá e bons estudos!



EXERCÍCIOS DE APROFUNDAMENTO!

1. Em uma população normal de uma variável X com variância populacional igual a 16, retiramos uma amostra de 27 observações, obtendo-se uma média de 8,03. Determine o intervalo de confiança para a média populacional μ , adotando um nível de significância de 90%.
2. Considere que a variância informada no exercício anterior tenha sido referente as 27 observações. Encontre o novo intervalo de confiança ao nível de 95%.
3. A duração de um equipamento de determinada marca ocorre com um desvio padrão de 1,2 anos. Sabendo-se que foram amostrados 120 equipamentos desta marca, obtendo-se uma média no tempo de duração de 13 anos, estime com 99% de confiabilidade o real tempo de duração dos equipamentos desta marca.
4. Sabendo que uma amostra de 16 pacientes que fizeram tratamento com o medicamento-A apresentou uma média de 12 dias e uma variância de 1,9 dias como período para recuperação; e que outra amostra de 19 pacientes que fizeram tratamento com o medicamento-B apresentou uma média de 15 dias e uma variância de 2,3 dias, estime com 95% de confiança a diferença nas médias dos tempos de recuperação para os dois medicamentos.
5. Calcule um intervalo de confiança de 97% para a proporção de itens produzidos com defeito por uma máquina, sabendo-se que uma amostra de 150 itens produzidos por esta máquina apresentou 13 itens com defeitos.

AULA 8

Testes estatísticos não-paramétricos

Olá aluno(a),

Em nossa última aula, vimos como realizar as estimações de parâmetros populacionais desconhecidos. Você lembra? Nesta aula, iremos dar continuidade ao estudo da inferência estatística, conhecendo os principais testes estatísticos não-paramétricos. Daremos início com o teste da discrepância. Teremos depois uma sequência com o teste da associação e o teste de Wilcoxon e finalizaremos nossa oitava aula com o teste da correlação.

Então, vamos à aula?

Objetivo

- Conhecer e aplicar os testes da discrepância, de Wilcoxon, da correlação e da associação entre duas variáveis

TÓPICO 1

Testes não-paramétricos para uma variável

OBJETIVO

- Testar diferentes ocorrências não-paramétricas para um determinado levantamento de uma variável quantitativa

Com base em levantamentos amostrais, iremos testar com certa confiabilidade, diferentes ocorrências não-paramétricas para um determinado levantamento de uma variável quantitativa.

1.1 INTRODUÇÃO - PASSOS PARA REALIZAÇÃO DE TESTES ESTATÍSTICOS

Antes de iniciarmos os testes estatísticos, vale ressaltar que podemos executar qualquer teste de significância a partir de uma sequência de cinco passos:

1º passo: formular as hipóteses do teste (haverá sempre duas hipóteses):

$$\begin{cases} H_0 : \text{A hipótese nula} \\ H_1 : \text{A hipótese alternativa, que contradirá sempre hipótese nula} \end{cases}$$

2º passo: fixar o nível de significância α . Trata-se do percentual que complementa a confiabilidade, estudado na aula passada. Depois iremos escolher a Variável do Teste (VT) e veremos que todo teste tem a sua variável a ser trabalhada.

3º passo: encontrar as Regiões Críticas e de Aceitação (RC e RA), a partir de onde nós iremos aceitar ou não o que defende a hipótese nula. Pois cada teste terá a sua curva com áreas críticas e áreas de aceitação para a hipótese H_0 formulada.

4º passo: calcular o valor da variável do teste com base nos levantamentos de dados.

5º passo: concluir o teste, comparando o resultado do cálculo da VT do 4º passo, com as regiões (RC e RA) do 3º passo, para aceitarmos ou não a hipótese H_0 .

Agora você irá conhecer o 1º teste.

1.2 TESTE DA DISCREPÂNCIA

Sejam E_1, E_2, \dots, E_k as k-etapas de um levantamento ou os k-eventos de um experimento e ainda:

- Fo_1, Fo_2, \dots, Fo_k as frequências observadas para as k-etapas ou k-eventos.
- Fe_1, Fe_2, \dots, Fe_k as frequências esperadas para as k-etapas ou k-eventos.

Estamos interessados em testar se existem diferenças significantes entre as frequências observadas e as frequências esperadas. Assim, vamos conhecer os cinco passos para execução do teste. Vamos a eles:

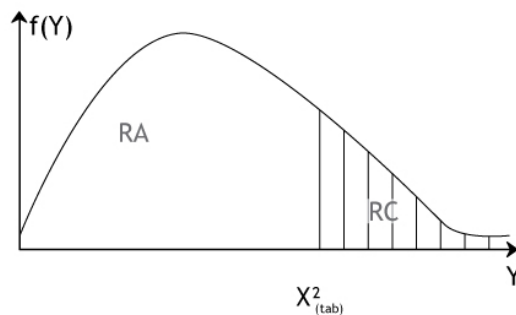
1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : \text{Não existe discrepância.} \\ H_1 : \text{Existe discrepância.} \end{cases}$$

2º passo: fixar o nível de significância α e escolher para variável do teste uma $\chi^2_{\alpha}(k-1)$, ou seja, o valor de uma distribuição contínua de probabilidade do qui-quadrado, que será encontrado na tabela de distribuição do qui-quadrado, na qual

$k \Rightarrow$ n° de etapas de um levantamento ou n° de eventos de um experimento

3º passo: encontrar as regiões crítica e de aceitação (RC e RA), as quais estarão inclusas na curva da distribuição qui-quadrado.



4º passo: calcular o valor da VT, através da seguinte relação:

$$\chi^2_{(cal)} = \sum_{i=1}^k \frac{(Fo_i - Fe_i)^2}{Fe_i}, \text{ em que}$$

$Fo_i \Rightarrow$ são as frequências observadas.

$Fe_i \Rightarrow$ são as frequências esperadas.



ATENÇÃO!

Fique atento para a seguinte informação: $\chi^2_{(tab)}$
 \Rightarrow É o valor tabelado da distribuição qui-quadrado.

5º passo: concluir comparando o resultado encontrado no 4º passo com o gráfico do 3º passo, para verificar se o resultado calculado cai na região RA, ou na região RC, e consequentemente aceitarmos ou rejeitarmos o que defende a hipótese H_0 , ou seja, iremos concluir se existe ou não discrepância significativa entre o que se observa e o que se espera.

Vamos então a uma aplicação?

EXERCÍCIO RESOLVIDO 1

Sejam os seguintes números médios de vendas em um estabelecimento comercial durante o 1º semestre de um determinado ano:

MESES	JAN	FEV	MAR	ABR	MAI	JUN
Nº MÉDIO DE VENDAS	38	41	49	39	52	21

Teste, ao nível de 90%, se existe discrepância significativa no nº de vendas no estabelecimento durante os meses do 1º semestre.

Solução:

Primeiramente, aluno, vamos pensar da seguinte forma, podemos chamar os números médios de vendas informados de Fo_i , que são as frequências observadas para as i -etapas. Assim, se não existisse discrepância entre os números de vendas, era de se esperar que todos os meses a quantidade de vendas fosse a mesma. Então poderíamos ter uma nova tabela de dados:

MESES	JAN	FEV	MAR	ABR	MAI	JUN
Fo_i	38	41	49	39	52	21
Fe_i	40	40	40	40	40	40

1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : \text{O número de vendas é praticamente o mesmo.} \\ H_1 : \text{O número de vendas é discrepante.} \end{cases}$$

2º passo: fixar o nível de significância $\alpha = 0,1$, visto que a confiabilidade do teste é de 90%. Assim iremos tomar para variável do teste (V.T) uma $\chi^2_{\alpha}(k-1)$, no nosso exemplo, teremos $\chi^2_{0,1}(6-1) = \chi^2_{0,1}(5)$. Assim, aluno, buscaremos esse resultado na tabela de distribuição do qui-quadrado, abaixo:

ATENÇÃO!

As Fe_i foram o resultado do nº médio das frequências observadas Fo_i .

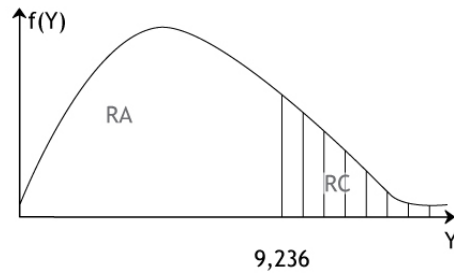
Agora vamos organizar os passos do teste.

Valores críticos (unilaterais à esquerda) da distribuição Qui-Quadrado

$$P(\chi^2 \text{ com } n \text{ graus de liberdade} \geq \text{valor tabelado}) = \alpha$$

	0,995	0,99	0,975	0,95	0,9	0,1	0,05	0,025	0,01	0,005
1	0,000	0,000	0,001	0,004	0,016	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	9,236	11,070	12,832	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	2,833	12,017	14,067	16,013	18,475	20,278
8	1,344	1,647	2,180	2,733	3,490	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	4,168	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	4,865	15,987	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	5,578	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	6,304	18,549	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	7,041	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	7,790	21,064	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	8,547	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	9,312	23,542	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	10,085	24,769	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	10,865	25,989	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	11,651	27,204	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	12,443	28,412	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	13,240	29,615	32,671	35,479	38,932	41,401
22	8,643	9,542	10,982	12,338	14,041	30,813	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	14,848	32,007	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	15,659	33,196	36,415	39,364	42,980	45,558
25	10,520	11,524	13,120	14,611	16,473	34,382	37,652	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	17,292	35,563	38,885	41,923	45,642	48,290
27	11,808	12,878	14,573	16,151	18,114	36,741	40,113	43,195	46,963	49,645
28	12,461	13,565	15,308	16,928	18,939	37,916	41,337	44,461	48,278	50,994
29	13,121	14,256	16,047	17,708	19,768	39,087	42,557	45,722	49,588	52,335
30	13,787	14,953	16,791	18,493	20,599	40,256	43,773	46,979	50,892	53,672
31	14,458	15,655	17,539	19,281	21,434	41,422	44,985	48,232	52,191	55,002
32	15,134	16,362	18,291	20,072	22,271	42,585	46,194	49,480	53,486	56,328
33	15,815	17,073	19,047	20,867	23,110	43,745	47,400	50,725	54,775	57,648
34	16,501	17,789	19,806	21,664	23,952	44,903	48,602	51,966	56,061	58,964
35	17,192	18,509	20,569	22,465	24,797	46,059	49,802	53,203	57,342	60,275
36	17,887	19,233	21,336	23,269	25,643	47,212	50,998	54,437	58,619	61,581
37	18,586	19,960	22,106	24,075	26,492	48,363	52,192	55,668	59,893	62,883
38	19,289	20,691	22,878	24,884	27,343	49,513	53,384	56,895	61,162	64,181
39	19,996	21,426	23,654	25,695	28,196	50,660	54,572	58,120	62,428	65,475
40	20,707	22,164	24,433	26,509	29,051	51,805	55,758	59,342	63,691	66,766
41	21,421	22,906	25,215	27,326	29,907	52,949	56,942	60,561	64,950	68,053
42	22,138	23,650	25,999	28,144	30,765	54,090	58,124	61,777	66,206	69,336
43	22,860	24,398	26,785	28,965	31,625	55,230	59,304	62,990	67,459	70,616
44	23,584	25,148	27,575	29,787	32,487	56,369	60,481	64,201	68,710	71,892
45	24,311	25,901	28,366	30,612	33,350	57,505	61,656	65,410	69,957	73,166
46	25,041	26,657	29,160	31,439	34,215	58,641	62,830	66,616	71,201	74,437
47	25,775	27,416	29,956	32,268	35,081	59,774	64,001	67,821	72,443	75,704
48	26,511	28,177	30,754	33,098	35,949	60,907	65,171	69,023	73,683	76,969
49	27,249	28,941	31,555	33,930	36,818	62,038	66,339	70,222	74,919	78,231
50	27,991	29,707	32,357	34,764	37,689	63,167	67,505	71,420	76,154	79,490

3º passo: teremos então a seguinte curva da distribuição do qui-quadrado



4º passo: agora iremos calcular o valor da VT. Neste teste, ela será dada através da seguinte relação:

$$\chi^2_{(cal)} = \sum_{i=1}^k \frac{(Fo_i - Fe_i)^2}{Fe_i}$$

$$= \frac{(38 - 40)^2}{40} + \frac{(41 - 40)^2}{40} + \dots + \frac{(21 - 40)^2}{40} = \frac{592}{40} = 14,8, \text{ em que}$$

$Fo_i \Rightarrow$ são as frequências observadas.

$Fe_i \Rightarrow$ são as frequências esperadas.

5º passo: quando comparamos o resultado do cálculo da V.T com as regiões (R.A e R.C), verificamos que o valor da variável do teste qui-quadrado calculada cai na região crítica, ou seja, rejeitaremos a hipótese H_0 , concluindo com uma confiabilidade de 90% que podemos afirmar que o nº de vendas são discrepantes durante os meses observados.

No item seguinte, iremos conhecer o nosso próximo teste não paramétrico. Vamos lá?

1.3 TESTE DE WILCOXON

Neste teste estatístico, compararemos os resultados de uma variável quantitativa em dois momentos, antes e depois de certa interferência para esta variável. Então vamos conhecer os cinco passos para execução desse teste:

1º passo: formular as hipóteses do teste.

$$\begin{cases} H_0 : \text{Não existe mudança significativa entre os dois momentos.} \\ H_1 : \text{Existe mudança significativa entre os dois momentos.} \end{cases}$$

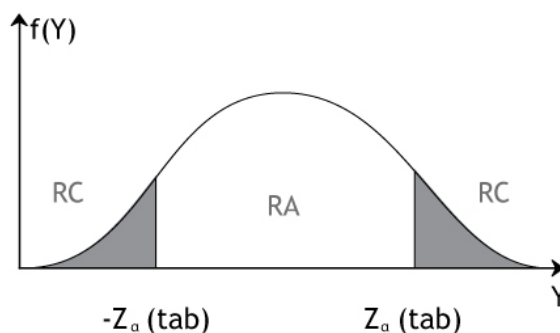


ATENÇÃO!

Lembre-se de que a tabela da normal padrão foi visto na aula 2. Se tiver dúvidas retorne a aula.

2º passo: fixar o nível de significância α e escolher para variável do teste uma normal padronizada Z_α , cujo valor será encontrado da mesma forma que vimos na aula-7, em que $Z_\alpha \Rightarrow$ É o valor de Z da tabela da normal padrão, cujo valor de probabilidade será o mais próximo de $1 - \frac{\alpha}{2}$.

3º passo: encontrar as regiões crítica e de aceitação (RC e RA), as quais estarão inclusas na curva da distribuição normal padronizada.



Observação: $Z_\alpha(tab) \Rightarrow$ É o valor tabelado da distribuição normal padronizada.

4º passo: calcular o valor da VT. Neste teste, ela será dada através da seguinte relação:

$$Z_{(cal)} = \frac{T - \mu(t)}{\sigma(t)}, \text{ em que } T \Rightarrow \text{ É a menor soma dos postos de mesmo sinal}$$

$$\mu(t) = \frac{n(n+1)}{4}$$

$$\sigma(t) = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

5º passo: iremos, neste último passo, concluir da mesma forma do teste anterior, comparando o resultado encontrado no 4º passo com o gráfico do 3º passo, para verificar se o resultado calculado cai em região RA (de aceitação) ou em região RC (crítica), para aceitarmos ou não a hipótese H_0 , ou seja, observaremos se houve mudança significativa ou não nos valores da variável.

EXERCÍCIO RESOLVIDO 2

Sejam as notas de uma turma antes e depois de uma nova técnica de ensino adotada. Deseja-se testar com uma confiabilidade de 98% se houve mudança significativa nas notas após o uso da nova técnica.

NOTAS (ANTES)	NOTAS (DEPOIS)
8,5	9
7	9
6,8	7,2
9,5	9
10	8
5	6
6,5	6,6
7,2	9,5
7	10
5	5
8	7
7	9,5
9	9,5
9,5	10
10	8,5
5	4,5
8,4	9,5
7	10
8	10
7,5	8

Solução:

Vamos iniciar desenvolvendo os passos do teste.

1º passo: formulação das hipóteses

$$\begin{cases} H_0 : \text{As notas não mudaram.} \\ H_1 : \text{As notas mudaram de forma significativa.} \end{cases}$$

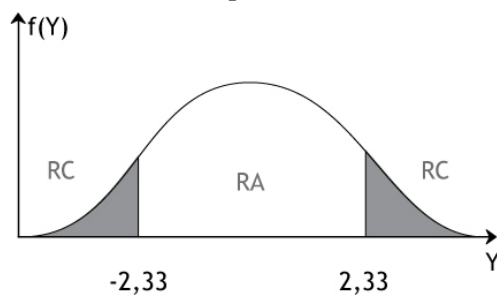
2º passo: fixaremos um nível de significância $\alpha = 0,02$, visto que a confiabilidade pedida no teste foi de 98%. Iremos escolher para variável do teste uma normal padronizada Z_α , que será obtido da seguinte maneira: primeiro iremos calcular o valor de probabilidade dado por $1 - \frac{\alpha}{2} = 1 - \frac{0,02}{2} = 0,99$. Depois, procuraremos o valor de Z_α , que assuma a probabilidade mais próxima deste valor, conforme o resultado abaixo:

Tabela da Distribuição Normal Padrão

P(Z<z)	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936

O valor de Z_{α} , será dado por $Z_{\alpha} = 2,33$

3º passo: colocaremos o valor de Z_{α} encontrado na tabela normal padrão acima na curva da distribuição normal padronizada.



4º passo: com base no levantamento amostral das 20 notas obtidas antes e depois da nova técnica de ensino, calcularemos o valor da V.T do teste de Wilcoxon

$$Z_{(cal)} = \frac{T - \mu(t)}{\sigma(t)}.$$

Vejamos como funciona:

$T \Rightarrow$ é a menor soma dos postos de mesmo sinal. Para encontrarmos o valor de T , iremos preencher os dois postos, o positivo e o negativo. Iniciaremos fazendo a diferença entre o antes e o depois:

NOTAS (ANTES)	NOTAS (DEPOIS)	Di	Postos(-)	Postos(+)
8,5	9	-0,5	10,5°	
7	9	-2	5,5°	
6,8	7,2	-0,4	13°	
9,5	9	0,5		
10	8	2		
5	6	-1	8°	4,5°
6,5	6,6	-0,1	14°	1°
7,2	9,5	-2,3	4°	
6	10	-4	1°	
5	5	0		
8	7	1		
7	9,5	-2,5	3°	
9	9,5	-0,5	10,5°	3°
9,5	10	-0,5	10,5°	
10	8,5	1,5		
5	4,5	0,5		
8,4	9,5	-1,1	7°	2°
7	10	-3	2°	4,5°
8	10	-2	5,5°	
7,5	8	-0,5	10,5°	
			105	15

Note que o preenchimento dos postos (posições) deve ser feito dos maiores para os menores. Por exemplo, o 1° posto negativo refere-se à diferença ($di = -4$) que é a maior diferença negativa, enquanto o 1° posto positivo se refere à diferença ($di = 2$), que é a maior diferença positiva. Quando houver diferenças iguais, devemos tirar a média entre as posições. Perceba que a diferença ($di=0,5$) aparece duas vezes e, pela ordem dos postos positivos, uma seria a quarta maior e a outra a quinta, porém, como são iguais, tiramos a média entre 4 e 5, e atribuímos a posição 4,5° as duas.

$T \Rightarrow$ é a menor soma dos postos. Vimos no exercício que em um posto ocorreu a soma 105 e no outro posto ocorreu a soma 15. Como a menor soma será o valor de T , então teremos que $T = 15$.

Vamos agora encontrar o valor da média estimada $\mu(t)$, que será dada por

$$\mu(t) = \frac{n(n+1)}{4} = \frac{19 \times 20}{4} = 95.$$

Por fim, iremos encontrar o desvio estimado dado por

$$\sigma(t) = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{19 \times 20 \times 21}{24}} = 18,2.$$



ATENÇÃO!

Tínhamos 20 alunos na amostra do nosso problema, mas verificamos que um desses alunos apresentou resultados da nota antes e nota depois iguais, logo este aluno ficará excluído da amostra. Por isso, passamos a trabalhar com a amostra $n=19$.

Agora podemos finalizar o cálculo da V.T, substituindo os três valores encontrados:

$$Z_{(cal)} = \frac{T - \mu(t)}{\sigma(t)} = \frac{15 - 95}{18,2} = -4,4$$

5º passo: comparando o valor do cálculo da VT com as regiões (RC e RA) do 3º passo, notamos que este resultado se encontra na região crítica da curva, assim rejeitaremos a hipótese H_0 . Desta forma, com uma confiabilidade de

98% (a qual foi adotada no teste), podemos concluir que houve uma mudança significativa nas notas após a nova técnica de ensino adotada. Como o valor do cálculo da V.T cai na RC inferior da curva, conclui-se que a mudança ocorrida foi para mais, verificando-se uma maior diferença negativa entre o antes e o depois, ou seja, as notas depois da técnica de ensino aumentaram mais do que diminuíram.

Neste tópico, vimos dois importantes testes não-paramétricos que envolviam a apuração de uma variável. No próximo tópico, veremos dois outros testes não-paramétricos que irão envolver relações entre duas variáveis. Vamos então ao próximo tópico.

TÓPICO 2

Testes não-paramétricos para duas variáveis

OBJETIVO

- Testar, com base em levantamentos amostrais de certa confiabilidade, diferentes ocorrências não-paramétricas para um determinado levantamento que envolva a relação entre duas variáveis

Caro aluno, neste tópico, primeiramente, você irá aprender como aplicar um teste estatístico da correlação. Depois aprenderá o teste estatístico da associação entre duas variáveis. Vamos aos testes?

2.1 TESTE DA CORRELAÇÃO

Conforme foi visto na aula 6, através do coeficiente de correlação, podemos verificar o quanto duas variáveis quantitativas estão correlacionadas. Podemos também, com base neste coeficiente amostral, aplicar um teste para verificar com certa confiabilidade se existe ou não esta correlação.

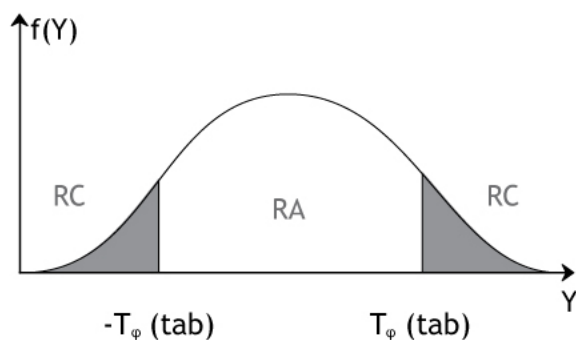
Para tanto, teremos que aplicar os cinco passos do teste a seguir:

1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : \text{Não existe correlação entre as duas variáveis.} \\ H_1 : \text{Existe correlação entre as duas variáveis.} \end{cases}$$

2º passo: fixar o nível de significância α e escolher para variável do teste uma $T_\alpha(n-2)$, ou seja, o valor de uma distribuição-T de Student que será encontrado na tabela da distribuição T.

3º passo: encontrar as regiões crítica e de aceitação(RC e RA), as quais estarão incluídas na curva da distribuição T de Student



ATENÇÃO!

$T_{\alpha}(tab) \Rightarrow$ é o mesmo valor tabelado da distribuição T de Student, que vimos na aula anterior sobre estimação.

4º passo: calcular o valor da VT. Neste teste, calcularemos da seguinte maneira:

$$T_{cal} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}, \text{ em que}$$

$r \Rightarrow$ é o coeficiente de correlação amostral

$n \Rightarrow$ é o tamanho da amostra

5º passo: concluiremos este passo da mesma maneira que fizemos nos testes anteriores, comparando o resultado encontrado no 4º passo com o gráfico do 3º passo, para verificar se o resultado calculado cai na região RA ou na região RC, para sabermos se aceitamos ou rejeitamos a hipótese H_0 , ou seja, se existe ou não correlação entre as duas variáveis quantitativas.

EXERCÍCIO RESOLVIDO 1

Com base na tabela abaixo, do exercício resolvido 1 da aula 6:

Tempo de estudo (X)	Nº de salários (Y)	XY	X ²	Y ²
4	2	8	16	4
0	1	0	0	1
13	8	104	169	64
2	1	2	4	1
2	1	2	4	1
9	5	45	81	25
15	9	135	225	81
11	7	77	121	49
8	4	32	64	16
10	6	60	100	36
5	2	10	25	4
20	10	200	400	100

Teste, ao nível de 95% de confiança, se realmente existe correlação entre o nº de salários ganhos e o tempo de estudo das pessoas.

Solução:

Vamos aplicar os cinco passos do teste.

1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : \text{O nº de salários não está correlacionado com o tempo de estudo das pessoas.} \\ H_1 : \text{O nº de salários está correlacionado com o tempo de estudo das pessoas.} \end{cases}$$

2º passo: fixando o nível de significância $\alpha = 0,05$ e escolhendo a V.T. do teste da correlação como sendo uma $T_\alpha(n-2) = T_{0,05}(12-2) = T_{0,05}(10) = 2,228$, iremos encontrar esse valor na tabela da distribuição T de Student abaixo:

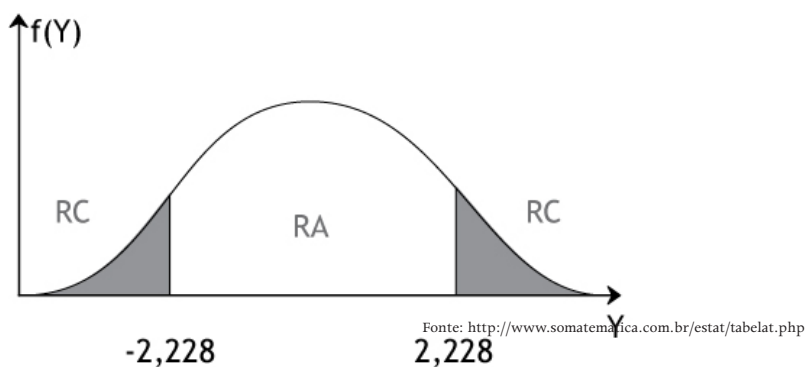
Tabela t (student)

gl/α	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
01	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
02	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
03	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,541	12,924
04	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
05	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
06	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
07	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,965	3,499	5,408
08	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
09	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318

Fonte: <http://www.somatematica.com.br/estat/tabelat.php>

Vimos na tabela da distribuição T que o valor tabelado é dado por $T_{0,05}(10) = 2,228$

3º passo: definir a curva da distribuição que será dada por



4º passo: calcular a V.T (variável do teste) com base nos dados amostrais levantados.

Como vimos na aula6, o coeficiente de correlação r foi calculado da seguinte forma:

$$r = \frac{S_{X,Y}}{\sqrt{S_{X,X} \cdot S_{Y,Y}}}$$

Então com base nos resultados encontrados na tabela de somatórios do nosso exercício, encontramos que

$$S_{X,Y} = \sum XY - \frac{\sum X \cdot \sum Y}{N} = 675 - \frac{99 \times 56}{12} = 213$$

$$S_{X,X} = \sum X^2 - \frac{(\sum X)^2}{N} = 1209 - \frac{(99)^2}{12} = 392,25$$

$$S_{Y,Y} = \sum Y^2 - \frac{(\sum Y)^2}{N} = 382 - \frac{(56)^2}{12} = 120,7.$$

Assim, tivemos

$$r = \frac{S_{X,Y}}{\sqrt{S_{X,X} \cdot S_{Y,Y}}} = \frac{213}{\sqrt{392,25 \times 120,7}} = 0,98.$$

Então, substituindo este valor em nossa fórmula do T_{cal} , teremos

$$T_{cal} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,98\sqrt{12-2}}{\sqrt{1-0,98^2}} = 15,5$$

5º passo: comparando o resultado do cálculo da V.T com a curva da distribuição do 3º passo, e verificamos que o resultado se encontra na região crítica, rejeitaremos a hipótese H_0 . Assim, com uma confiabilidade de 95%, podemos concluir que realmente existe correlação entre o nº de salários ganhos e o tempo de estudo das pessoas.

2.2 TESTE DA ASSOCIAÇÃO

Outro importante teste não-paramétrico que envolve duas variáveis é o teste da associação. Com ele, poderemos verificar se existe relação de dependência entre duas variáveis, sejam elas qualitativas ou quantitativas, adotando também, com base no levantamento amostral, certo nível de confiabilidade para o teste.

Para tanto, aluno, seguiremos os cinco passos do teste:

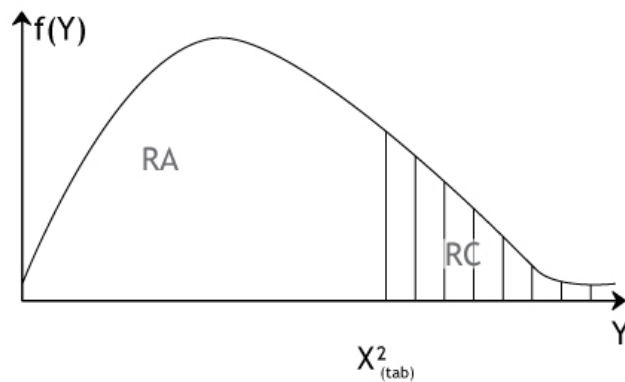
1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : \text{não existe associação entre as duas variáveis} \\ H_1 : \text{existe associação entre as duas variáveis} \end{cases}$$

2º passo: fixar o nível de significância α e escolher para variável do teste uma $\chi^2_{\alpha}(L-1)(C-1)$, ou seja, o valor de uma distribuição contínua do qui-quadrado, que será encontrado na tabela da distribuição qui-quadrado, conforme vimos no teste da discrepância, em que

- $L \Rightarrow$ É o nº de linhas da tabela de contingência, a qual irá dispor os valores das duas variáveis.
- $C \Rightarrow$ É o nº de colunas da tabela de contingência. (veremos no ex. de aplicação).

3º passo: encontrar as regiões: crítica e de aceitação (RC e RA), as quais estarão incluídas na curva da distribuição qui-quadrado.



Observe que $\chi^2_{(tab)} \Rightarrow$ É o valor tabelado da distribuição qui-quadrado.

4º passo: calcular o valor da V.T. Neste teste calcularemos da seguinte maneira:

$$\chi^2_{(cal)} = \sum_{i=1}^L \sum_{j=1}^C \frac{(Fo_{ij} - Fe_{ij})^2}{Fe_{ij}}, \text{ em que}$$

$Fo_{i,j} \Rightarrow$ São as frequências observadas para cada linha- i e coluna- j

$Fe_{i,j} \Rightarrow$ São as frequências esperadas para cada linha- i e coluna- j

Cada $Fe_{i,j}$ será obtida da seguinte forma $\frac{(\sum \text{linha} - i) \times (\sum \text{coluna} - j)}{n}$.

5º passo: comparar o resultado encontrado no 4º passo com o gráfico do 3º passo, para verificarmos se iremos aceitar ou não a hipótese H_0 , ou seja, se aceitaremos ou não que as duas variáveis do teste estão associadas.

EXERCÍCIO RESOLVIDO 2

Com base no seguinte levantamento amostral, referente à preferência por determinada disciplina e o sexo do aluno, teste com 90% de confiança se existe associação entre a disciplina preferida e o sexo do aluno:

DISCIPLINA	SEXO		TOTAL
	Masculino	Feminino	
Matemática	18	11	29
Português	13	14	27
Ciências	8	16	24
TOTAL	39	41	80

Solução:

1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : \text{a preferência pela disciplina não depende do sexo do aluno} \\ H_1 : \text{a preferência pela disciplina depende do sexo do aluno} \end{cases}$$

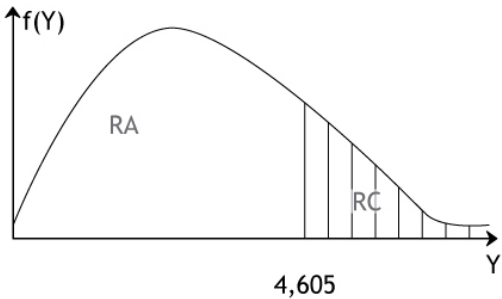
2º passo: como a confiabilidade pedida no teste é de 90%, então a V.T do teste será dada por $\chi^2_{\alpha}(L-1)(C-1) = \chi^2_{0,1}(3-1)(2-1) = \chi^2_{0,1}(2)(1) = \chi^2_{0,1}(2)$. Olhando na tabela da distribuição qui-quadrado abaixo, temos que

Valores críticos (unilaterais à esquerda) da distribuição Qui-Quadrado
 $P(X^2 \text{ com } n \text{ graus de liberdade} \geq \text{valor tabelado}) = \alpha$

	0,995	0,99	0,975	0,95	0,9	0,1	0,05	0,025	0,01	0,005
1	0,000	0,000	0,001	0,004	0,016	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	6,251	7,815	9,348	11,345	12,838

$$\chi^2_{0,1}(2) = 4,605$$

3º passo: formamos agora a curva da distribuição do qui-quadrado, destacando as regiões crítica e de aceitação.



4º passo: com base no levantamento amostral, vamos calcular a variável do teste (V.T)

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

Note ainda que os dados levantados referem-se às frequências observadas $Fo_{i,j}$.

Assim temos

DISCIPLINA	SEXO		TOTAL
	Masculino	Feminino	
Matemática	18	11	29
Português	13	14	27
Ciências	08	16	24
TOTAL	39	41	80

Então, a partir das frequências observadas, iremos encontrar as frequências esperadas, lembrando que cada $Fe_{i,j}$ será obtida da seguinte forma:

$$Fe_{i,j} = \frac{(\sum \text{linha} - i) \times (\sum \text{coluna} - j)}{n}$$

$$Fe_{i,j} = \begin{bmatrix} \frac{29 \times 39}{80} & \frac{29 \times 41}{80} \\ \frac{27 \times 39}{80} & \frac{27 \times 41}{80} \\ \frac{24 \times 39}{80} & \frac{24 \times 41}{80} \end{bmatrix} = \begin{bmatrix} 14,1 & 14,9 \\ 13,2 & 13,8 \\ 11,7 & 12,3 \end{bmatrix}$$

Agora voltemos à fórmula do cálculo da VT, $\chi^2_{(cal)} = \sum_{i=1}^L \sum_{j=1}^c \frac{(Fo_{ij} - Fe_{ij})^2}{Fe_{ij}}$.

Então teremos

$$\chi^2_{(cal)} = \frac{(18-14,1)^2}{14,1} + \frac{(11-14,9)^2}{14,9} + \frac{(13-13,2)^2}{13,2} + \frac{(14-13,8)^2}{13,8} + \frac{(8-11,7)^2}{11,7} + \frac{(16-12,3)^2}{12,3}$$

$$= 4,4$$

5º passo: comparando o resultado do $\chi^2_{(cal)}$ com as regiões (RA e RC) da curva do qui-quadrado do 3º passo, vimos que o resultado do $\chi^2_{(cal)}$ encontra-se

na região de RA. Sendo assim, aceitaremos a hipótese H_0 , ou seja, concluímos com 90% de confiabilidade que a preferência pela disciplina não depende do sexo do aluno.

Você aprendeu, nesta aula, os principais testes estatísticos não-paramétricos envolvendo uma e duas variáveis. Na próxima aula, iremos aplicar os principais testes que irão envolver parâmetros estatísticos, tais como médias, variâncias e proporções. Mantenha o entusiasmo! Até a próxima aula!



ATIVIDADES DE APROFUNDAMENTO!

1. Crie uma situação que você possa aplicar o teste estatístico de Wilcoxon, para uma amostra de tamanho 10
2. Sabendo-se que duas amostras de tamanho 20 cada uma, referentes a duas variáveis quantitativas apresentam um coeficiente de correlação de -0,93 entre elas. Teste, ao nível de 90% de confiança, se realmente existe correlação entre as duas variáveis
3. Em 60 lançamentos de um dado, obtivemos 15 vezes o resultado um, 12 vezes o resultado dois, 8 vezes o resultado três, 7 vezes o resultado quatro, 10 vezes o resultado cinco e 8 vezes o resultado seis. Teste ao nível de 95%, se podemos considerar o dado equilibrado

AULA 9

Testes estatísticos não-paramétricos

Olá aluno(a),

Vimos na aula passada as aplicações dos principais testes estatísticos não-paramétricos, está lembrado? Nesta aula, iremos continuar aplicando importantes testes, porém estes envolverão parâmetros estatísticos, denominados testes paramétricos. Iniciaremos com o teste que envolve um dos parâmetros estatísticos mais utilizados, o teste da média. Prosseguiremos com o teste da proporção e finalizaremos com dois testes, o que tratará da diferença entre duas médias e, em seguida, o teste para diferença entre duas proporções.

Então, vamos à aula?

Objetivos

- Conhecer os testes populacionais da média e da proporção aplicados em amostras com certa confiabilidade
- Aplicar os testes para diferenças entre duas médias e duas proporções populacionais, através de levantamentos amostrais

TÓPICO 1

Testes estatísticos envolvendo um único parâmetro populacional

OBJETIVO

- Testar se determinados parâmetros populacionais assumem ou não determinados valores, através de um levantamento amostral, realizado com certa confiabilidade

Antes de começarmos a nossa aula, é importante lembrar que, para realizarmos qualquer teste estatístico, devemos seguir a organização dos cinco passos do teste, conforme vimos na aula anterior.

1º passo: formular as hipóteses do teste (teremos sempre duas hipóteses):

$$\begin{cases} H_0 : \text{A hipótese nula (hipótese a ser testada)} \\ H_1 : \text{A hipótese alternativa (hipótese que contradiz a hipótese nula)} \end{cases}$$

2º passo: fixar o nível de significância α . Você lembra? Trata-se do percentual complementar da confiabilidade que vimos nas aulas 7 e 8.

Depois teremos de definir a variável do teste (V.T), pois, como vimos, todo teste tem a sua variável a ser trabalhada

3º passo: encontrar as regiões crítica e de aceitação (RC e RA), a partir de onde nós iremos aceitar ou não o que defende a hipótese nula.

4º passo: calcular o valor da VT, com base na amostra levantada.

5º passo: comparar o resultado do cálculo da V.T encontrado no 4º passo, com as regiões (RC e RA) do 3º passo, para podermos aceitar como verdadeiro o que defende a hipótese H_0 .

Vamos conhecer o 1º teste?

1.1 TESTE DA MÉDIA

Neste teste, iremos verificar, com base em resultados amostrais, se as médias populacionais assumem determinados valores. Para tanto, teremos de seguir os cinco passos do teste:

1º passo: formular as hipóteses do teste

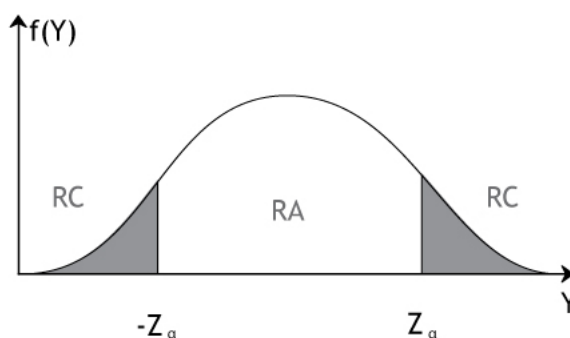
$$\begin{cases} H_0 : \mu = \mu_0 ; \mu > \mu_0 ; \mu < \mu_0 \\ H_1 : \mu \neq \mu_0 ; \mu \leq \mu_0 ; \mu \geq \mu_0 \end{cases}$$

2º passo: fixar o nível de significância α e escolher a variável do teste (VT) a partir do seguinte critério:

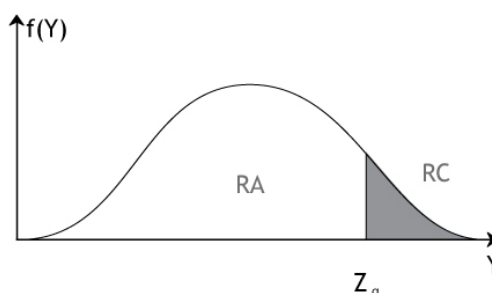
- Se o tamanho da amostra n for inferior a 30 e conhecermos o desvio-padrão populacional σ , então a VT = Z_α . Se não conhecermos o desvio-padrão populacional σ , então o VT = $T_\alpha(n-2)$
- Se $n \geq 30$, então sempre a VT = Z_α .

3º passo: encontrar as regiões crítica e de aceitação (RC e RA). Dependendo das hipóteses formuladas no 1º passo e da variável do teste a ser trabalhada, poderemos ter uma das seis curvas seguintes:

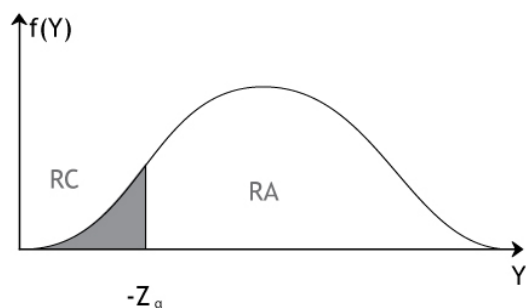
1ª) Quando estivermos com a hipótese $H_1 : \mu \neq \mu_0$, usando a VT = Z_α , teremos



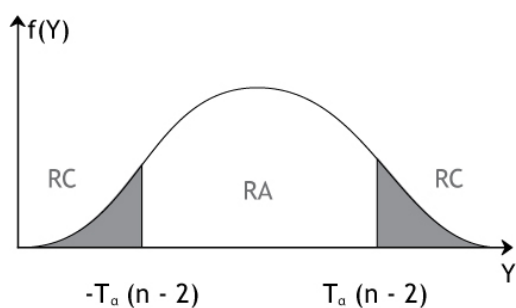
2ª) Quando estivermos com a hipótese $H_1 : \mu \geq \mu_0$, usando a VT = Z_α , teremos



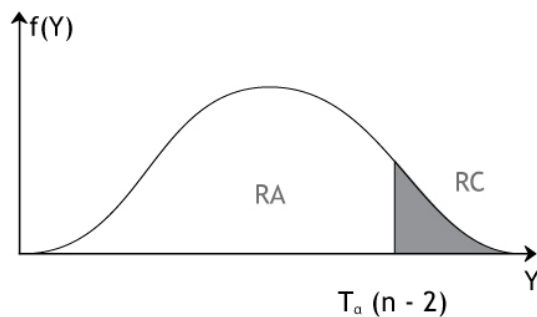
3ª) Quando estivermos com a hipótese $H_1 : \mu \leq \mu_0$, usando a VT = Z_α , teremos



4ª) Quando estivermos com a hipótese $H_1 : \mu \neq \mu_0$, usando a VT = $T_\alpha(n-2)$, teremos



5ª) Quando estivermos com a hipótese $H_1 : \mu \geq \mu_0$, usando a VT = $T_\alpha(n-2)$, teremos

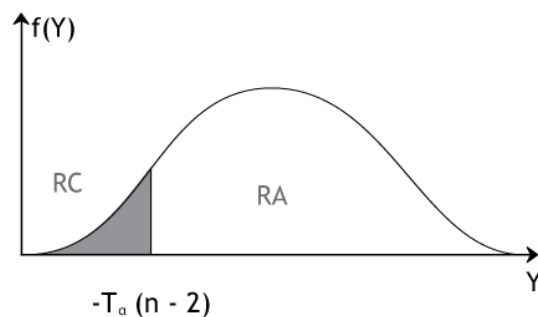


6ª) Quando estivermos com a hipótese $H_1 : \mu \leq \mu_0$, usando a VT = $T_\alpha(n-2)$, teremos

ATENÇÃO!

$Z_\alpha \Rightarrow$ É o valor tabelado da distribuição normal padrão.

$T_\alpha(n-2) \Rightarrow$ É o valor tabelado da distribuição T-Student.



4º passo: calcular o valor da V.T de acordo com os dados amostrais a partir do seguinte critério:

- Se o tamanho da amostra n for inferior a 30 e não conhecermos o desvio-padrão populacional σ , então teremos

$$T_{cal} = \frac{\bar{X} - \mu}{S_x^-}$$

- Se o tamanho da amostra $n \geq 30$

- conhecendo o desvio-padrão populacional σ , teremos

$$Z_{cal} = \frac{\bar{X} - \mu}{\sigma_x^-}$$

- não conhecendo o desvio-padrão populacional σ , teremos

$$Z_{cal} = \frac{\bar{X} - \mu}{S_x^-}$$

5º passo: comparar o resultado encontrado no 4º passo com o gráfico trabalhado no 3º passo, para verificar se o resultado calculado cai na região RA, ou na região RC, e consequentemente aceitarmos ou rejeitarmos o que defende a hipótese H_0 .

Agora você vai acompanhar todos os passos do teste, através de um exemplo de aplicação.

EXERCÍCIO RESOLVIDO 1

Um professor afirma que a nota média de suas turmas que concluem a disciplina de estatística nunca são inferiores a 7,5. Para verificarmos se a informação

do professor é verdadeira, resolvemos testá-la, selecionando uma amostra de 35 turmas das quais o referido professor lecionou. Encontramos uma nota média e um desvio padrão para a disciplina de estatística de 7,2 e 0,9 respectivamente. Com base nesses resultados amostrais, vamos testar, com 90% de confiabilidade, se o professor está ou não correto em sua afirmação.

Solução:

Primeiramente, aluno, vamos organizar o teste em seus cinco passos:

1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : \mu \geq 7,5 \\ H_1 : \mu < 7,5 \end{cases}$$

2º passo: fixar o nível de significância $\alpha = 0,1$, visto que a confiabilidade do teste é de 90%. Assim, de acordo com o critério do 2º passo do teste, como o tamanho da amostra trabalhada $n=35>30$, então iremos tomar para variável do teste (V.T), uma Z_α .

Desta forma, como já vimos em aulas anteriores, para encontrarmos Z_α , iremos primeiro calcular o valor da relação $1 - \frac{\alpha}{2} = 1 - \frac{0,1}{2} = 0,95$. Na tabela da distribuição normal padrão abaixo, veremos que o valor de Z_α que assume esse resultado probabilístico é de $Z_\alpha = 1,65$.

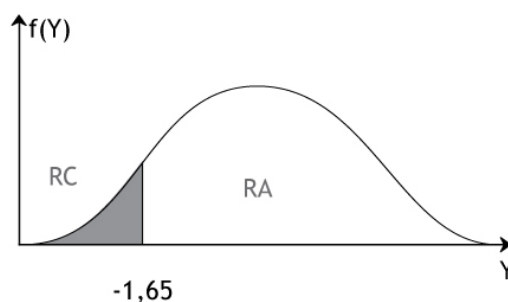
Tabela da Distribuição Normal Padrão

P(Z<z)	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441

1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

Fonte: www.pucrs.br/format/rossana/psicologia/tabela_normal.pdf

3º passo: definir as regiões crítica e de aceitação do teste de acordo com as hipóteses formuladas no 1º passo:



4º passo: como na nossa amostra $n=35 > 30$ e não conhecemos o desvio-padrão populacional, e sim o amostral, usaremos a seguinte metodologia de cálculo:

$$Z_{cal} = \frac{\bar{X} - \mu}{\frac{S_x}{\sqrt{35}}} = \frac{7,2 - 7,5}{\frac{0,9}{\sqrt{35}}} = -2$$

5º passo: comparando o resultado do cálculo da VT = - 2 encontrado no 4º passo, com as regiões (R.A e R.C) do 3º passo, podemos notar que se encontra na região crítica, ou seja, rejeitamos a hipótese H_0 . Portanto, com a confiabilidade de 90% que foi adotada no teste, não temos razões suficientes para acreditar na informação do professor.

EXERCÍCIO RESOLVIDO 2

Suponha no exercício anterior que a amostra selecionada para o teste fosse de 27 turmas. Vejamos o que mudaria para neste novo teste.

Solução: Vamos novamente aos cinco passos

1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : \mu \geq 7,5 \\ H_1 : \mu < 7,5 \end{cases}$$

2º passo: fixar o nível de significância $\alpha = 0,1$, visto que a confiabilidade do teste é de 90%, então de acordo com o critério do 2º passo do teste, como

o tamanho da amostra trabalhada $n=27<30$. Logo iremos tomar para variável do teste(V.T), uma $T_{\alpha}(n-2)$. Desta forma, como já vimos em aulas anteriores, $T_{0,1}(27-2)=T_{0,1}(25)=2,06$, conforme nos mostra a tabela da distribuição T-Student abaixo.

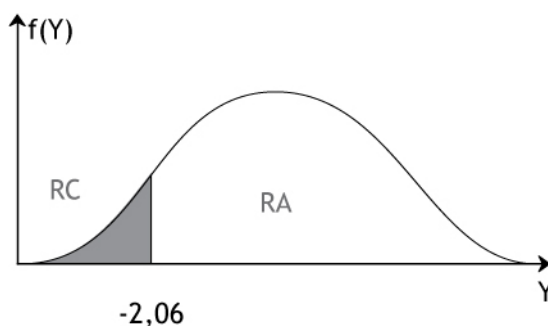
Como vocês puderam notar, o valor 2,06 foi encontrado no cruzamento do grau 25, com o nível de significância de 5% e não de 10%, conforme o problema. Isso ocorre porque esta tabela é para curva de duas caldas críticas. A curva desse problema será de apenas uma calda, conforme veremos no 3º passo. Por isso, é necessário olhar na metade do valor da significância.

TABELA T (STUDENT)

gl/ α	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
01	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
02	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
03	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,541	12,924
04	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
05	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
06	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
07	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,365	3,499	5,408
08	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
09	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,726
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,856	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,856	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646

Fonte: <http://www.somatematica.com.br/estat/tabelat.php>

3º passo: definir as regiões crítica e de aceitação do teste de acordo com as hipóteses formuladas no 1º passo:



4º passo: como na nossa amostra $n=27 < 30$ e não conhecemos o desvio-padrão populacional, e sim o amostral, usaremos a seguinte metodologia de cálculo:

$$T_{cal} = \frac{\bar{X} - \mu}{\frac{S_x}{\sqrt{27}}} = \frac{7,2 - 7,5}{\frac{0,9}{\sqrt{27}}} = -1,2$$

5º passo: comparando o resultado do cálculo da VT = - 1,2 encontrado no 4º passo, com as regiões (R.A e R.C) do 3º passo, podemos notar que se encontra na região de aceitação, concluindo-se que aceitamos a hipótese H_0 . Ou seja, com a confiabilidade de 90% que foi adotada no teste, não temos razões suficientes para desacreditar na informação do professor.

Agora iremos conhecer o nosso segundo teste paramétrico. Vamos lá?

1.2 TESTE DA PROPORÇÃO

Neste teste, iremos observar, através do nosso levantamento amostral, quando as proporções populacionais assumem determinados resultados. Para isso, teremos de acompanhar os cinco passos do teste:

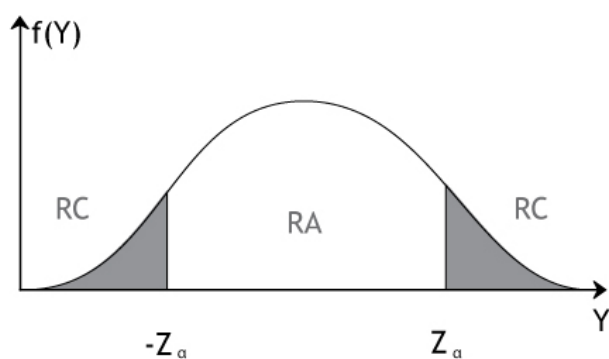
1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : P = P_0; P > P_0; P < P_0 \\ H_1 : P \neq P_0; P \leq P_0; P \geq P_0 \end{cases}$$

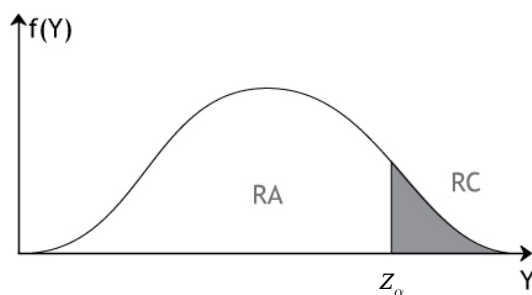
2º passo: fixar o nível de significância α e definir que a variável do teste (VT), independente dos dados amostrais, sempre será uma normal padronizada Z_α .

3º passo: encontrar as regiões crítica e de aceitação (RC e RA). Assim, de acordo com as hipóteses formuladas no 1º passo, podemos ter

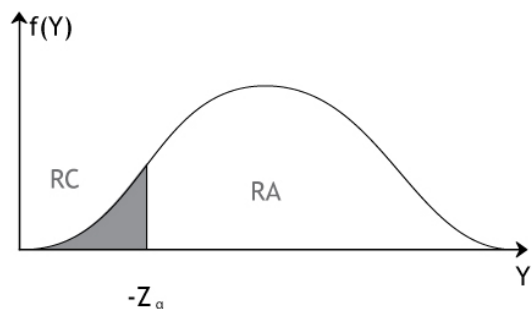
1ª) Quando estivermos com a hipótese $H_1 : P \neq P_0$, teremos



2ª) Quando estivermos com a hipótese $H_1 : P \geq P_0$, teremos



3ª) Quando estivermos com a hipótese $H_1 : P \leq P_0$, teremos



4º passo: calcular o valor da V.T. de acordo com os dados amostrais, através da seguinte relação

$$Z_{cal} = \frac{f - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{0,19 - 0,23}{\sqrt{\frac{0,23 \times 0,77}{1000}}} = -3,01$$

na qual devemos saber que

$$\begin{cases} f \Rightarrow \text{É a proporção encontrada na amostra} \\ P_0 \Rightarrow \text{É o valor a qual a hipótese } H_0 \text{ está se referindo no teste} \\ n \Rightarrow \text{É o tamanho da amostra} \end{cases}$$

5º passo: concluir o teste, comparando o resultado encontrado no 4º passo com o gráfico trabalhado no 3º passo, para verificar se o resultado calculado cai na região RA, ou na região RC, e consequentemente aceitarmos ou rejeitarmos o que defende a hipótese H_0 .

Agora você aluno irá acompanhar todos os passos do teste, através de um exemplo de aplicação.

EXERCÍCIO RESOLVIDO 2

A Secretaria de Educação de uma cidade afirma que a proporção estudantes que têm acesso ao ensino superior acima dos 40 anos de idade é de 0,23. Teste com uma confiabilidade de 95% se esta afirmação é verdadeira, sabendo-se que, numa amostra de 1000 estudantes que tiveram acesso ao ensino superior na cidade, verificou-se que 190 tinham mais de 40 anos.

Solução:

Vamos acompanhar os cinco passos do teste, ok!

1º passo: formular das hipóteses

$$\begin{cases} H_0 : P_0 = 0,23 \\ H_1 : P_0 \neq 0,23 \end{cases}$$

2º passo: fixar o nível de significância $\alpha = 0,05$, visto que a confiabilidade pedida no teste foi de 95% e como a variável para o teste da proporção será sempre uma normal padronizada Z_α . Encontraremos o valor da relação $= 1 - \frac{0,05}{2} = 0,975$. Assim, iremos procurar o valor de Z_α , que assuma a probabilidade mais próxima deste valor, conforme os resultados da tabela abaixo:

Tabela da Distribuição Normal Padrão

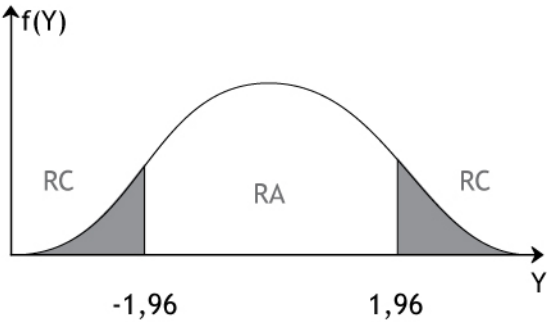
$P(Z < z)$	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141

0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

Fonte: www.pucrs.br/format/rossana/psicologia/tabela_normal.pdf

O valor de Z_{α} , será dado por $Z_{\alpha} = 1,96$

3º passo: colocar o valor de Z_{α} encontrado na tabela normal padrão acima na curva da distribuição normal padronizada



4º passo: calcular o valor da V.T, com base no resultado amostral apurado, da seguinte forma:

$$Z_{cal} = \frac{f - P_0}{\sqrt{\frac{P_0(1 - P_0)}{n}}} = \frac{0,19 - 0,23}{\sqrt{\frac{0,23 \times 0,77}{1000}}} = -3,01$$

5º passo: comparando o valor do cálculo da V.T, com as regiões (RC e RA) do 3º passo, notamos que cai numa região crítica, assim rejeitaremos a hipótese

H_0 , ou seja, com uma confiabilidade de 95% que adotamos em nosso problema, podemos concluir que a proporção de estudantes que têm acesso ao ensino superior com mais de 40 anos não é exatamente de 23%.

Neste tópico, vimos dois importantes testes paramétricos que envolveram a apuração de um parâmetro. No próximo tópico, veremos dois outros testes paramétricos que envolverão relações entre dois parâmetros.

Vamos lá então?

TÓPICO 2

Testes estatísticos envolvendo dois parâmetros populacionais

OBJETIVO

- Testar o tipo de desigualdade que existe entre dois parâmetros populacionais, através de um levantamento amostral, realizado com certa confiabilidade

É válido lembrar que, para realizarmos qualquer teste estatístico, devemos seguir a organização dos cinco passos do teste, conforme vimos no tópico anterior. Então vamos aos testes!

2.1 TESTE DE SIGNIFICÂNCIA PARA IGUALDADE ENTRE DUAS MÉDIAS

Neste teste, iremos verificar, com base em resultados amostrais, se existem diferenças significantes entre duas médias populacionais. Para tanto, teremos que seguir os cinco passos do teste:

1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : \mu_1 = \mu_2 \text{ ou } \mu_1 - \mu_2 = d \\ H_1 : \mu_1 \neq \mu_2 \text{ ou } \mu_1 - \mu_2 \neq d \end{cases}$$

Observação: d é uma diferença que admitimos entre as médias

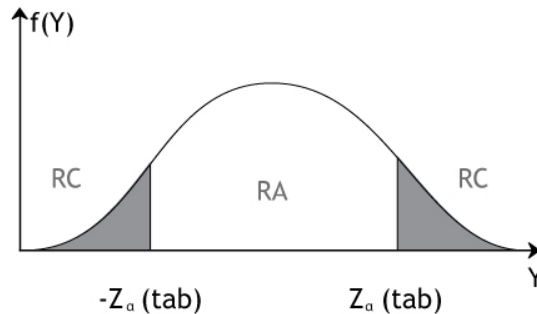
2º passo: fixar o nível de significância α e escolher a variável do teste da seguinte maneira:

- Se σ (desvio padrão populacional) for conhecido, teremos para V.T (variável do teste) uma normal padronizada Z_α

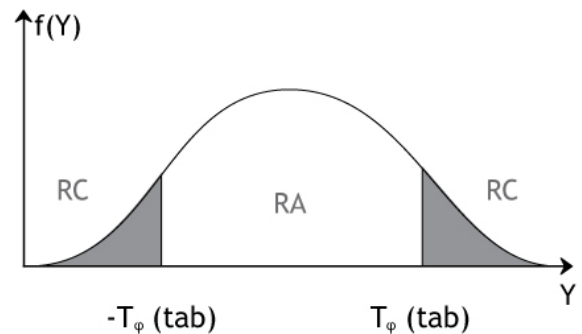
- Se σ for desconhecido, teremos para V.T uma distribuição T_φ , em que $\varphi = n_1 + n_2 - 2$

3º passo: encontrar as regiões crítica e de aceitação (RC e RA), as quais estarão inclusas na curva da distribuição adequada

1ª) Quando σ for conhecido, usaremos



2ª) Quando σ for desconhecido, usaremos



ATENÇÃO!

$Z_\alpha(tab) \Rightarrow$ É o valor tabelado da distribuição normal padrão

$T_\alpha(tab) \Rightarrow$ É o valor tabelado da distribuição T de Student

4º passo: calcular o valor desta variável do teste, dependendo da V.T do 2º passo, da seguinte forma:

$$Z_{cal} = \frac{(\bar{X}_1 - \bar{X}_2) - d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$\text{ou } T_{cal} = \frac{(\bar{X}_1 - \bar{X}_2) - d}{S_c \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}} \text{ com } S_c = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

\overline{X}_1 e $\overline{X}_2 \Rightarrow$ São as médias amostrais

σ_1^2 e $\sigma_2^2 \Rightarrow$ São as variâncias populacionais

n_1 e $n_2 \Rightarrow$ São os tamanhos das duas amostras

S_1^2 e $S_2^2 \Rightarrow$ São as variâncias amostrais

5º passo: concluiremos este passo da mesma maneira fizemos nos testes anteriores, comparando o resultado encontrado no 4º passo com o gráfico do 3º passo, para verificar se o resultado calculado cai na região RA, ou na região RC, e consequentemente aceitarmos ou rejeitarmos a hipótese H_0 , ou seja, iremos concluir se existe ou não igualdade entre as duas médias.

EXERCÍCIO RESOLVIDO 1

Um professor afirma que não existe diferença significativa entre as médias de duas turmas das suas disciplinas de estatística. Duvidando da veracidade desta informação, e resolvendo testar ao nível de 98%, selecionamos uma amostra de 33 alunos da turma-1 que apresentou uma nota média de 6,9 e outra de 38 alunos da turma-2, a qual já apresentou uma nota média de 7,4. Sabendo-se que as variâncias das duas turmas são respectivamente 1,1 e 0,8, realize o teste desejado.

Solução:

Vamos seguir o teste passo a passo

1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : \mu_1 = \mu_2 \text{ ou } \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 \neq \mu_2 \text{ ou } \mu_1 - \mu_2 \neq 0 \end{cases}$$

2º passo: fixar o nível de significância $\alpha = 0,02$ e, como no problema os desvios padrões populacionais σ são conhecidos, pois aprendemos em aulas anteriores que $\sigma = \sqrt{\sigma^2}$, então temos que $\sigma_1 = \sqrt{1,1} = 1,05$ e $\sigma_2 = \sqrt{0,8} = 0,89$. Iremos escolher para variável do teste uma normal padronizada Z_α . Encontraremos o valor da relação $= 1 - \frac{0,02}{2} = 0,99$. Assim, procuraremos o valor de Z_α que assuma a probabilidade mais próxima deste valor e conforme os resultados da tabela abaixo:

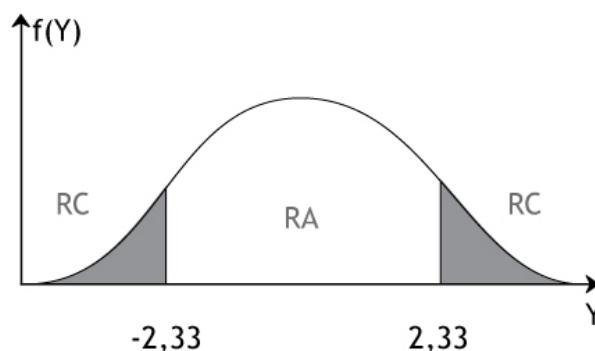
Tabela da Distribuição Normal Padrão

P(Z<z)	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

Fonte: www.pucrs.br/format/rossana/psicologia/tabela_normal.pdf

O valor de Z_{α} , será dado por: $Z_{\alpha} = 2,33$

3º passo: de acordo com as regras vistas no 3º passo, podemos definir a curva da distribuição da seguinte forma:



4º passo: de acordo com a nossa V.T trabalhada, e com base nos dados amostrais levantados, vamos calcular a V.T (variável do teste).

$$Z_{cal} = \frac{(\bar{X}_1 - \bar{X}_2) - d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(6,9 - 7,4) - 0}{\sqrt{\frac{1,1^2}{33} + \frac{0,8^2}{38}}} = -2,16$$

5º passo: concluir, comparando o resultado do cálculo da V.T, com a curva da distribuição do 3º passo, e verificar que o resultado se encontra na região de aceitação, ou seja, aceitaremos a hipótese H_0 . Assim com uma confiabilidade de 98%, não temos razões suficientes para duvidar da informação do professor.

EXERCÍCIO RESOLVIDO 2

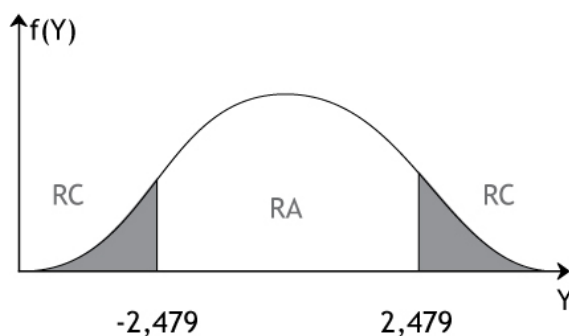
Suponha que, no exercício anterior, não fossem informadas as variâncias das duas turmas, e sim as variâncias das duas amostras de alunos. Uma de 20 alunos de uma turma que apresentou variância de 1,4, e outra de 18 alunos da outra turma que apresentou uma variância de 1,9. Então, iríamos aplicar o teste da seguinte forma:

1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : \mu_1 = \mu_2 \text{ ou } \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 \neq \mu_2 \text{ ou } \mu_1 - \mu_2 \neq 0 \end{cases}$$

2º passo: fixar o nível de significância $\alpha = 0,02$. Como no problema os desvios padrões populacionais σ seriam desconhecidos, então escolheremos para variável do teste, T_φ , em que $\varphi = n_1 + n_2 - 2$, ou seja, $T_\alpha(n_1 + n_2 - 2) = T_{0,02}(15 + 13 - 2) = T_{0,02}(26) = 2,479$. Conforme vimos na tabela da distribuição T-Student abaixo:

3º passo: de acordo com as regras vistas no 3º passo, podemos definir a curva da distribuição da seguinte forma:



4º passo: de acordo com a nossa V.T trabalhada, e com base nos dados amostrais levantados, vamos calcular a V.T (variável do teste).

$$T_{cal} = \frac{(\bar{X}_1 - \bar{X}_2) - d}{S_c \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}} = \frac{(6,9 - 7,4) - 0}{1,12 \sqrt{\frac{20 + 18}{20 \cdot 18}}} = -1,37$$

Onde:

$$S_c = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(20 - 1)1,4 + (18 - 1)1,9}{20 + 18 - 2}} = 1,12.$$

5º passo: podemos concluir, comparando o resultado do cálculo da V.T = -1,37, com a curva da distribuição do 3º passo. Assim verificamos que o resultado se encontra na região de aceitação, ou seja, aceitaremos que não existe diferença significativa entre as médias das duas turmas.

Agora você vai conhecer o segundo teste estatístico paramétrico que envolve dois parâmetros.

2.2 TESTE DE SIGNIFICÂNCIA PARA IGUALDADE ENTRE DUAS PROPORÇÕES

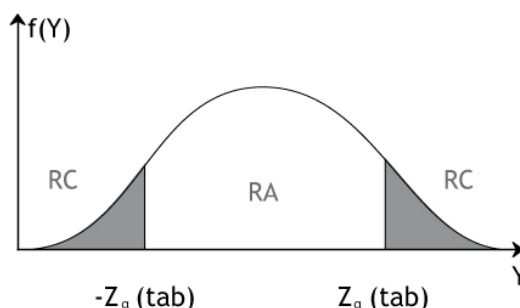
Neste teste, iremos, através de levantamentos amostrais, verificar se existem diferenças significantes entre duas médias populacionais. Para tanto, teremos que seguir os cinco passos do teste:

1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$$

2º passo: fixar o nível de significância α e pré-estabelecer que a variável do teste será uma normal padronizada Z_α

3º passo: construir as regiões crítica e de aceitação (RC e RA), com o auxílio da tabela normal padronizada.



$Z_\alpha (\text{tab}) \Rightarrow$ É o valor tabelado da distribuição normal padrão.

4º passo: calcular o valor desta variável do teste da seguinte forma:

$$Z_{cal} = \frac{f_1 - f_2}{\sqrt{p_*(1 - p_*)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \text{ em que}$$

f_1 e $f_2 \Rightarrow$ São as proporções amostrais, obtidas por

$$f_1 = \frac{x_1}{n_1} \text{ e } f_2 = \frac{x_2}{n_2}$$

p_* \Rightarrow É o estimador comum a p_1 e p_2 , que é obtido por $p_* = \frac{x_1 + x_2}{n_1 + n_2}$

n_1 e $n_2 \Rightarrow$ São os tamanhos das duas amostras

x_1 e $x_2 \Rightarrow$ São os números de casos de cada amostra

5º passo: concluiremos este passo da mesma maneira dos testes anteriores, comparando o resultado encontrado no 4º passo com o gráfico do 3º passo, para verificar se o resultado calculado cai na região RA ou na região RC para aceitarmos ou rejeitarmos a hipótese H_0 . Ou seja, iremos concluir se existe ou não igualdade entre as duas proporções populacionais.

Para que você, aluno, possa fixar melhor o funcionamento do teste, vamos acompanhar a sequência dos cinco passos do teste, através de um exercício resolvido.

EXERCÍCIO RESOLVIDO 3

Em uma amostra de 32 alunos do sexo masculino do curso da EaD de Matemática, verificou-se que 19 deles acessam o ambiente virtual Moodle diariamente. Em outra amostra de 40 estudantes do sexo feminino do mesmo curso, notamos que 28 deles também acessam diariamente o ambiente. Com uma confiabilidade de 95%, teste se existe diferença significativa entre as proporções de estudantes dos dois sexos que acessam o ambiente diariamente.

Solução:

Vamos organizar os passos do teste:

1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$$

2º passo: fixar o nível de significância $\alpha = 0,05$, visto que a confiabilidade pedida no teste foi de 95% e, como a variável para este teste é sempre uma normal padronizada Z_α , iremos encontrar o valor da relação $= 1 - \frac{0,05}{2} = 0,975$. Assim, procurando na tabela da distribuição normal padrão, qual valor de Z_α , que assume a probabilidade mais próxima a este resultado?

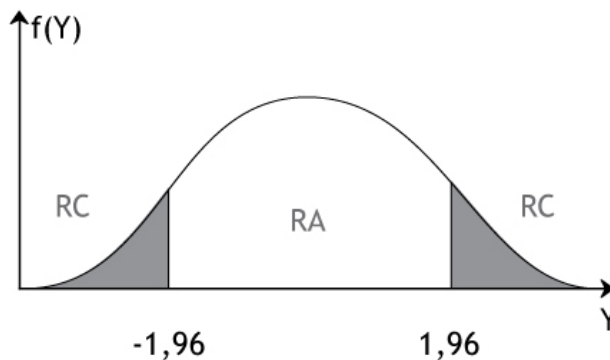
Tabela da Distribuição Normal Padrão

P(Z<z)	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

Fonte: www.pucrs.br/format/rossana/psicologia/tabela_normal.pdf

Encontraremos que $Z_{\alpha} = 1,96$

3º passo: colocar o valor de Z_{α} encontrado na tabela normal padrão acima na curva da distribuição normal padronizada



4º passo: aplicar a metodologia de cálculo da variável do teste (V.T), com base nas apurações amostrais do problema.

$$f_1 = \frac{x_1}{n_1} = \frac{19}{32} = 0,59$$

$$f_2 = \frac{x_2}{n_2} = \frac{28}{40} = 0,70$$

$$p_* = \frac{x_1 + x_2}{n_1 + n_2} = \frac{19 + 28}{32 + 40} = \frac{47}{72} = 0,65$$

Assim, teremos

$$Z_{cal} = \frac{f_1 - f_2}{\sqrt{p_*(1-p_*)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0,59 - 0,70}{\sqrt{0,65 \times 0,35 \times \left(\frac{1}{32} + \frac{1}{40}\right)}} = -0,97$$

5º passo: comparando o resultado do cálculo da V.T = -0,97 com a curva da distribuição do 3º passo, verificaremos que o resultado se encontra na região de aceitação, ou seja, podemos aceitar com 95% de confiança que não existe diferença significativa entre as proporções de estudantes do sexo masculino e do sexo feminino da EaD (curso de Matemática), que acessam o ambiente Moodle diariamente.

Nesta aula, você aprendeu como podemos, com base em levantamentos amostrais, aplicar os principais testes estatísticos paramétricos que envolvem um ou dois parâmetros. Na próxima aula, aplicaremos uma miscelânea de exercícios, os quais irão revisar toda a parte de inferência estatística vista nas quatro últimas aulas. Então, até a próxima aula!



ATIVIDADES DE APROFUNDAMENTO!

1. Se em um levantamento amostral oriundo de uma população de tamanho 500 temos uma média para uma variável igual a 8,2 e um desvio padrão de 0,9, teste ao nível de confiança de 90% se a média dessa população é superior a 7, sabendo-se que o tamanho da amostra trabalhada foi de 62.
2. No exercício anterior, suponha que o tamanho da amostra foi de 23. Agora repita o teste com uma confiabilidade de 95%.
3. Sabendo-se que, numa amostra de 150 lâmpadas, 131 delas têm no mínimo 1300hs de vida útil, teste com 97% de confiabilidade se a real proporção de lâmpadas da mesma marca que não atingem este tempo de vida útil, é de 15%
4. Com base na 1ª questão, e sabendo-se que num outro levantamento amostral para uma outra população, tivemos $\bar{X} = 7,5$; $S^2 = 1,2$; $n = 62$, adote uma confiabilidade de 95% e teste se existe diferença significativa entre as médias das duas populações.

AULA 10

Miscelânea de exercícios resolvidos

Olá caro aluno(a),

Nesta aula, iremos recordar, através de exercícios resolvidos, as principais informações sobre a inferência estatística, englobando a correlação e regressão, os principais tipos de estimações e os principais tipos de testes estatísticos paramétricos e não paramétricos, os quais foram os assuntos vistos em nossas últimas quatro aulas.

Então, vamos à revisão?

Objetivos

- Recordar, através de novas aplicações, como encontrar um coeficiente de correlação entre duas variáveis e como construir uma equação de regressão linear simples
- Realizar estimações de parâmetros populacionais desconhecidos através de situações problemas criadas
- Conhecer os principais tipos de testes estatísticos, tanto paramétricos como não paramétricos

TÓPICO 1

Exercícios sobre correlação e regressão

OBJETIVOS

- Encontrar e interpretar os coeficientes de correlação e de determinação
- Construir modelos de regressão linear simples e compreender suas utilizações

Vamos então iniciar as resoluções das situações-problema. Neste tópico, você recordará como se verifica quando duas variáveis quantitativas estão ou não correlacionadas e como usar as equações de regressão linear simples para estimar valores de uma variável dependente em função de uma variável independente.

EXERCÍCIO RESOLVIDO 1

Seja tempo diário de estudo em horas e o nº de reprovações de uma amostra de 11 estudantes em suas disciplinas cursadas estabelecidos conforme o quadro abaixo:

Horas de estudo semanal fora da escola	2	1	2	5	1	2	3	6	7	6	10
Nº de reprovações	1	4	3	4	3	2	6	1	3	4	0

a) Interprete o resultado do Coeficiente de Determinação

b) Estime o nº de reprovações de um estudante que estuda 8 horas semanais fora da escola

VAMOS ÀS SOLUÇÕES!

a) Primeiro devemos lembrar que o coeficiente de determinação será obtido por r^2 . Deveremos primeiro encontrar o coeficiente de correlação r . Para tanto, iremos definir quem é a variável independente X e a variável dependente Y .

No exemplo, você pode notar que o N° de reprovações será a variável dependente, enquanto as Horas de estudo semanal fora da escola serão a variável independente.

$$r = \frac{S_{x,y}}{\sqrt{S_{x,x} \cdot S_{y,y}}}, \text{ em que}$$

$$S_{x,y} = \sum XY - \frac{\sum X \cdot \sum Y}{N} = 128 - \frac{45 \times 33}{11} = -7$$

$$S_{x,x} = \sum X^2 - \frac{(\sum X)^2}{N} = 269 - \frac{45^2}{11} = 84,9$$

$$S_{y,y} = \sum Y^2 - \frac{(\sum Y)^2}{N} = 121 - \frac{33^2}{11} = 22$$

Então, vamos encontrar os somatórios:

Horas de estudo semanal fora da escola (X)	N° de reprovações (Y)	XY	X ²	Y ²
2	1	2	4	1
1	4	4	1	16
2	3	6	4	9
5	4	20	25	16
1	3	3	1	9
2	2	4	4	4
3	6	18	9	36
6	1	6	36	1
7	3	21	49	9
6	4	24	36	16
10	2	20	100	4
45	33	128	269	121

$$\text{Logo teremos } r = \frac{S_{x,y}}{\sqrt{S_{x,x} \cdot S_{y,y}}} = \frac{-7}{\sqrt{84,9 \times 22}} = -0,16$$

$$\Rightarrow r^2 = (-0,16)^2 = 0,0256.$$

Ou seja, podemos concluir, através do coeficiente de determinação, que o n° de horas de estudo semanal fora da escola explica em aproximadamente 3% o n° de reprovações

b) Para realizarmos a estimação de um resultado da variável dependente, devemos agora construir a E.R.L.S, a qual é dada da seguinte forma:

$$\begin{aligned}\bar{Y} &= a + bX \Rightarrow \bar{Y} = 6,3 - 0,8X \\ \bar{Y} &= 6,3 - 0,8X \Rightarrow \bar{Y} = 6,3 - 0,8 \times 8 \cong 0\end{aligned}, \text{ a partir da qual sabemos que}$$

$$a = \bar{Y} - b\bar{X} \quad \text{e} \quad b = \frac{S_{X,Y}}{S_{X,X}}.$$

$$\text{Assim, teremos } b = \frac{S_{X,Y}}{S_{X,X}} = \frac{-7}{84,9} = -0,08.$$

$$\text{Como sabemos que } \bar{X} = \frac{\sum X}{n} = \frac{45}{11} = 4,09 \quad \text{e} \quad \bar{Y} = \frac{\sum Y}{n} = \frac{33}{11} = 3.$$

$$\text{Podemos agora encontrar } a = \bar{Y} - b\bar{X} = 3 - (-0,8) \times 4,09 = 6,3.$$

$$\text{Agora você vai construir a E.R.L.S dada por } \bar{Y} = a + bX \Rightarrow \bar{Y} = 6,3 - 0,8X.$$

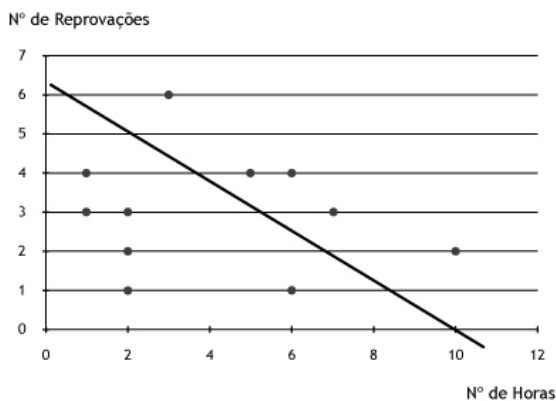
Com a equação construída, podemos estimar o valor de \bar{Y} substituindo o valor de X na equação da seguinte maneira:

$\bar{Y} = 6,3 - 0,8X \Rightarrow \bar{Y} = 6,3 - 0,8 \times 8 \cong 0$. Ou seja, estima-se que, se um aluno estudar semanalmente 8 h além do horário escolar, ele não sofrerá reprovação.

EXERCÍCIO RESOLVIDO 2

Com base nos dados da questão anterior, construa o gráfico de dispersão entre as duas variáveis X e Y , explicando o seu comportamento.

Solução:



Podemos notar, por meio do gráfico de dispersão, que existe um fraco comportamento linear negativo entre as duas variáveis. Isso pode ser explicado através do coeficiente de correlação r encontrado na questão anterior que foi de $r = -0,16$.

Veja que, nessa seção, tratamos de assuntos que foram trabalhados na aula 6 dessa disciplina. Qualquer dúvida em relação à resolução dessas questões, retorne à aula para entender melhor os conceitos já estudados. Você estará mais amadurecido frente a esses conceitos já estudados e pode vir a entendê-los melhor. No próximo tópico, serão explanados exercícios que tratem sobre estimações.

TÓPICO 2

Exercícios sobre estimações

OBJETIVOS

- Estimar parâmetros populacionais desconhecidos (médias e proporções) com base em levantamentos amostrais
- Encontrar os intervalos de confiança para estimar as diferenças entre duas médias e também as diferenças entre duas proporções

Vamos dar início a este segundo tópico recordando, através de exercícios aplicativos, como devemos estimar médias e proporções populacionais. Posteriormente, aplicaremos exercícios para estimarmos as diferenças entre dois resultados destes parâmetros populacionais desconhecidos.

EXERCÍCIO RESOLVIDO 1

Para estimarmos o número médio de acesso semanal dos alunos de uma instituição de ensino à biblioteca, selecionamos uma amostra de 42 semanas, obtendo-se uma média de 1425 alunos e um desvio padrão de 107 alunos. Com base neste resultado amostral, vamos adotar uma confiabilidade de 90% e realizar a estimação.

Vamos à solução!

Primeiro vamos organizar as informações do problema:

A amostra n é 42

A média amostral \bar{X} é 1425

O desvio-padrão amostral S é 107

O nível de significância α , que é o complemento da confiabilidade do teste, é 10% ou 0,1. Desta forma, como vimos na aula 7, estamos em uma situação em que a nossa amostra é maior que 30 e o nosso desvio-padrão populacional (σ) é desconhecido, assim teremos de usar o seguinte intervalo de estimação: $\bar{X} \pm Z_{\alpha} S_x$. Como faremos para encontrar $Z_{\alpha} = ?$

Teremos de calcular a relação $1 - \frac{\alpha}{2} = 1 - \frac{0,1}{2} = 0,95$.: e posteriormente procurar na tabela da normal padrão:

Tabela da Distribuição Normal Padrão

P(Z<z)	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

Fonte: www.pucrs.br/format/rossana/psicologia/tabela_normal.pdf

Então, vimos que o valor de Z correspondente é 1,65, logo $Z_{\alpha} = 1,65$.

Agora vamos encontrar o desvio-padrão estimado para média S_x . Conforme vimos na aula 7, iremos considerar que a amostra não ultrapassa 5% do tamanho da população, assim teremos $S_x = \frac{S}{\sqrt{n}} = \frac{107}{\sqrt{27}} = 20,6$.

Desta maneira, teremos apenas de substituir os três valores encontrados, conforme será mostrado abaixo:

$$\bar{X} \pm Z_{\alpha} S_x \Rightarrow 1425 \pm 1,65 \times 16,5 \Rightarrow 1425 \pm 27,2 \Rightarrow [1397,8 \text{ até } 1452,2].$$

Ou seja, com uma confiabilidade de 90%, poderíamos dizer que o nº médio semanal de estudantes que frequentam a biblioteca da instituição é de no mínimo 1398 e de no máximo 1452.

EXERCÍCIO RESOLVIDO 2

Suponha, no exercício anterior, que a amostra trabalhada fosse de apenas 27 semanas. Assim, estime o novo nº médio semanal de estudantes que frequentam a biblioteca.

Solução:

Perceba você que agora estaríamos em uma situação que a amostra seria menor que 30, pois $n = 27$. Desta maneira, devemos trabalhar com o seguinte intervalo de estimação $\bar{X} \pm T_{\alpha(n-1)}S_{\bar{x}} \Rightarrow 1425 \pm 1,706 \times 20,6 \Rightarrow 1425 \pm 35,1 \Rightarrow [1389,9 \text{ até } 1460,1]$, no qual $T_{\alpha(n-1)} = T_{0,1(27-1)} = T_{0,1(26)}$.

Com o auxílio da tabela da distribuição T-Student abaixo, encontraremos:

Tabela t (student)

gl/ α	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
01	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
02	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
03	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,541	12,924
04	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
05	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
06	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
07	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,365	3,499	5,408
08	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
09	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792

23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,726
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,856	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,856	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646

Fonte: <http://www.somatematica.com.br/estat/tabelat.php>

$T_{0,1(26)} = 1,706$. Agora só precisamos encontrar o novo valor de S_x^- , que será dado por $S_x^- = \frac{S}{\sqrt{n}} = \frac{107}{\sqrt{27}} = 20,6$.

Assim já podemos encontrar o intervalo de confiança para estimarmos a nova média, conforme mostramos abaixo:

$$\begin{aligned}\bar{X} \pm T_{\alpha(n-1)} S_x^- &\Rightarrow 1425 \pm 1,706 \times 20,6 \Rightarrow \\ 1425 \pm 35,1 &\Rightarrow [1389,9 \text{ até } 1460,1]\end{aligned}$$

Logo o novo n° médio de estudantes que frequentam a biblioteca por semana é de no mínimo 1390 e no máximo 1460.

EXERCÍCIO RESOLVIDO 3

Sabendo-se que, numa amostra de 80 canetas de certa marca, 6 apresentaram algum tipo de defeito na qualidade de suas tintas. Deseja-se estimar com 93% de confiabilidade a real proporção de canetas que apresentam os mesmos defeitos encontrados na amostra.

VAMOS À SOLUÇÃO!

Vimos na aula 7 que, para estimarmos as proporções populacionais, devemos usar o seguinte intervalo de estimação:

$$p \pm Z_\alpha S_p = 0,075 \pm 0,39 \times 0,03 = 0,075 \pm 0,01 = [0,065 \text{ até } 0,085] = [6,5\% \text{ até } 8,5\%] F_e.$$

Em que $p \Rightarrow$ proporção encontrada em nossa amostra, que em nosso exemplo será

$$\text{obtida por } p = \frac{6}{80} = 0,075$$

$$Z_\alpha = ?$$

Como vimos no exercício 1, devemos primeiro calcular $1 - \frac{\alpha}{2} = 1 - \frac{0,07}{2} = 0,65$.

Agora, procurando na tabela normal padrão abaixo a probabilidade mais próxima a 0,65, verificamos que $Z_{\alpha} = 0,39$.

Tabela da Distribuição Normal Padrão

P(Z<z)	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

Fonte: www.pucri.br/format/rossana/psicologia/tabela_normal.pdf

$$S_p = ?$$

Como vimos na aula 7, podemos considerar a população deste nosso problema como infinita, visto que não temos a população de canetas produzidas da marca especificada no problema. Sendo assim, teremos que

$$S_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0,075 \times 0,925}{80}} = 0,03$$

Então agora podemos substituir os valores no intervalo de estimação da seguinte forma:

$$p \pm Z_{\alpha} S_p = 0,075 \pm 0,39 \times 0,03 = 0,075 \pm 0,01 = [0,065 \text{ até } 0,085] = [6,5\% \text{ até } 8,5\%]$$

Logo a proporção de canetas da referida marca que apresentam algum tipo de defeito é de no mínimo 6,5% e de no máximo 8,5%.

Veja que, nesse tópico 2, tratamos mais uma vez de conceitos como Intervalo de confiança, em que podemos ver as principais aplicações referentes a esse assunto. No próximo tópico, você poderá recordar conceitos de testes não-paramétricos. Assunto estudado na aula 8. Então, vamos nessa?

TÓPICO 3

Exercícios sobre testes não-paramétricos

OBJETIVOS

- Testar, com certa confiabilidade, diferentes ocorrências não-paramétricas para um determinado levantamento de uma variável quantitativa, tomando por base levantamentos amostrais
- Verificar diferentes ocorrências não-paramétricas para um determinado levantamento que envolva a relação entre duas variáveis

Inicaremos recordando, através de exercícios resolvidos, como aplicar os testes não-paramétricos da discrepância e de Wilcoxon.

EXERCÍCIO RESOLVIDO 1

Observando os seguintes percentuais de desistências dos alunos matriculados em um curso de nível superior, apurados num determinado ano:

BIMESTRES	1ºBIM	2ºBIM	3ºBIM	4ºBIM	5ºBIM	6ºBIM
% DE DESISTÊNCIAS	12	8	7	5	7	15

Teste, ao nível de 95% de confiabilidade, se existe discrepância significativa ao longo dos bimestres.

VEJAMOS A SOLUÇÃO:

Primeiro, conforme vimos na aula 8, você deve organizar a tabela das frequências observadas Fo_i e das frequências esperadas Fe_i para as i -etapas.

BIMESTRES	1ºBIM	2ºBIM	3ºBIM	4ºBIM	5ºBIM	6ºBIM
Fo_i	12	8	7	5	7	9
Fe_i	9	9	9	9	9	9

Agora já podemos organizar os cinco passos do teste, conforme aprendemos na aula 8. Vamos a eles:

1º passo: formular as hipóteses do teste

$$\left\{ \begin{array}{l} H_0 : \text{O n}^\circ \text{ de desistências é praticamente o mesmo durante todos os} \\ \text{bimestres} \\ H_1 : \text{O n}^\circ \text{ de desistências é discrepante} \end{array} \right.$$



ATENÇÃO!

As Fe_i foram o resultado do nº médio das frequências observadas Fo_i .

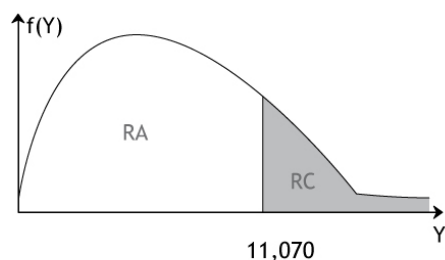
2º passo: fixando o nível de significância $\alpha = 0,05$, visto que a confiabilidade do teste é de 905%, então iremos tomar para variável do teste (V.T) uma $\chi^2_{\alpha}(k-1)$. Dessa forma, conforme as informações de nosso problema, teremos $\chi^2_{0,05}(6-1) = \chi^2_{0,05}(5)$. Você irá procurar este resultado na tabela de distribuição do qui-quadrado, abaixo:

Valores críticos (unilaterais à esquerda) da distribuição Qui-Quadrado $P(\chi^2 \text{ com } n \text{ graus de liberdade} \geq \text{valor tabelado}) = \alpha$

	0,995	0,99	0,975	0,95	0,9	0,1	0,05	0,025	0,01	0,005
1	0,000	0,000	0,001	0,004	0,016	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	9,236	11,070	12,832	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	2,833	12,017	14,067	16,013	18,475	20,278
8	1,344	1,647	2,180	2,733	3,490	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	4,168	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	4,865	15,987	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	5,578	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	6,304	18,549	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	7,041	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	7,790	21,064	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	8,547	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	9,312	23,542	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	10,085	24,769	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	10,865	25,989	28,869	31,526	34,805	37,156

19	6,844	7,633	8,907	10,117	11,651	27,204	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	12,443	28,412	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	13,240	29,615	32,671	35,479	38,932	41,401
22	8,643	9,542	10,982	12,338	14,041	30,813	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	14,848	32,007	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	15,659	33,196	36,415	39,364	42,980	45,558
25	10,520	11,524	13,120	14,611	16,473	34,382	37,652	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	17,292	35,563	38,885	41,923	45,642	48,290
27	11,808	12,878	14,573	16,151	18,114	36,741	40,113	43,195	46,963	49,645
28	12,461	13,565	15,308	16,928	18,939	37,916	41,337	44,461	48,278	50,994
29	13,121	14,256	16,047	17,708	19,768	39,087	42,557	45,722	49,588	52,335
30	13,787	14,953	16,791	18,493	20,599	40,256	43,773	46,979	50,892	53,672
31	14,458	15,655	17,539	19,281	21,434	41,422	44,985	48,232	52,191	55,002
32	15,134	16,362	18,291	20,072	22,271	42,585	46,194	49,480	53,486	56,328
33	15,815	17,073	19,047	20,867	23,110	43,745	47,400	50,725	54,775	57,648
34	16,501	17,789	19,806	21,664	23,952	44,903	48,602	51,966	56,061	58,964
35	17,192	18,509	20,569	22,465	24,797	46,059	49,802	53,203	57,342	60,275
36	17,887	19,233	21,336	23,269	25,643	47,212	50,998	54,437	58,619	61,581
37	18,586	19,960	22,106	24,075	26,492	48,363	52,192	55,668	59,893	62,883
38	19,289	20,691	22,878	24,884	27,343	49,513	53,384	56,895	61,162	64,181
39	19,996	21,426	23,654	25,695	28,196	50,660	54,572	58,120	62,428	65,475
40	20,707	22,164	24,433	26,509	29,051	51,805	55,758	59,342	63,691	66,766
41	21,421	22,906	25,215	27,326	29,907	52,949	56,942	60,561	64,950	68,053
42	22,138	23,650	25,999	28,144	30,765	54,090	58,124	61,777	66,206	69,336
43	22,860	24,398	26,785	28,965	31,625	55,230	59,304	62,990	67,459	70,616
44	23,584	25,148	27,575	29,787	32,487	56,369	60,481	64,201	68,710	71,892
45	24,311	25,901	28,366	30,612	33,350	57,505	61,656	65,410	69,957	73,166
46	25,041	26,657	29,160	31,439	34,215	58,641	62,830	66,616	71,201	74,437
47	25,775	27,416	29,956	32,268	35,081	59,774	64,001	67,821	72,443	75,704
48	26,511	28,177	30,754	33,098	35,949	60,907	65,171	69,023	73,683	76,969
49	27,249	28,941	31,555	33,930	36,818	62,038	66,339	70,222	74,919	78,231
50	27,991	29,707	32,357	34,764	37,689	63,167	67,505	71,420	76,154	79,490

3º passo: iremos agora construir a curva da distribuição do qui-quadrado



4º passo: agora, com base em nosso levantamento, vamos calcular o valor da V.T. que, neste teste, conforme vimos, será feito da seguinte forma:

$$\chi^2_{(cal)} = \sum_{i=1}^k \frac{(Fo_i - Fe_i)^2}{Fe_i} = \frac{(12-9)^2}{9} + \frac{(8-9)^2}{9} + \dots + \frac{(9-9)^2}{9} = \frac{34}{9} = 3,8 \text{ em}$$

que

$Fo_i \Rightarrow$ São as frequências observadas

$Fe_i \Rightarrow$ São as frequências esperada

5º passo: quando comparamos o resultado do cálculo da V.T. com as regiões (R.A e R.C) encontradas no 3º passo, verificamos que o valor da variável do teste qui-quadrado calculada se encontra na região de aceitação(R.A), ou seja, aceitaremos a hipótese H_0 , concluindo assim com uma confiabilidade de 95%, que podemos afirmar não haver discrepâncias significantes entre o nº de desistências entre os alunos matriculados na disciplina.

EXERCÍCIO RESOLVIDO 2

Sejam os números de acessos semanais de 15 alunos de um determinado polo ao sistema *Moodle*, antes e depois de um encontro presencial com a participação do professore-formador e do conteudista da disciplina. Aplique um teste, ao nível de 94% de confiança, para saber se houve mudança no número de acesso semanal após o encontro.

ACESSOS (ANTES)	ACESSOS (DEPOIS)
3	6
4	6
3	3
8	6
7	15
13	12
10	14
7	8
8	8
4	1
10	11
18	27
16	22
12	15
11	11

Solução:

Iniciaremos desenvolvendo os passos do teste.

1º passo: formulação das hipóteses

$$\left\{ \begin{array}{l} H_0 : \text{Não houve mudança significativa} \\ H_1 : \text{Houve mudança significativa} \end{array} \right.$$

2º passo: fixaremos um nível de significância $\alpha=0,06$, visto que a confiabilidade pedida no teste foi de 94%. Agora você escolherá para V.T uma normal padronizada Z_α .Conforme vimos anteriormente, temos que primeiro calcular o valor de probabilidade dado por $1-\frac{\alpha}{2}=1-\frac{0,06}{2}=0,97$ e depois procurar o valor de Z_α , que assuma a probabilidade mais próxima deste valor, conforme o resultado abaixo:

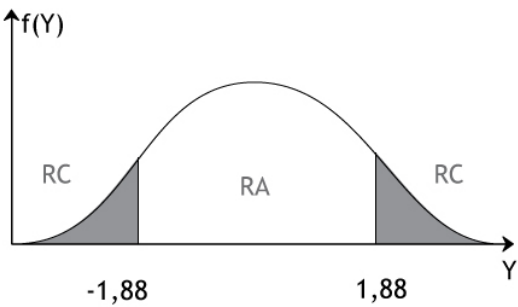
Tabela da Distribuição Normal Padrão

P(Z<z)	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

Fonte: www.pucri.br/format/rossana/psicologia/tabela_normal.pdf

O valor de Z_α será dado por $Z_\alpha = 1,88$

3º passo: colocaremos o valor de Z_{α} encontrado na tabela normal padrão acima na curva da distribuição normal padronizada.



4º passo: com base no levantamento amostral dos 15 alunos, calcularemos o valor da V.T do teste de Wilcoxon $Z_{(cal)} = \frac{T - \mu(t)}{\sigma(t)}$.

Vejamos como funciona:

$T \Rightarrow$ é a menor soma dos postos de mesmo sinal. Para encontrarmos o valor de T , iremos preencher os dois postos, o positivo e o negativo. Iniciaremos fazendo a diferença entre o antes e o depois:

Nº DE ACESSOS (ANTES)	Nº DE ACESSOS (DEPOIS)	di	Postos(-)	Postos(+)
3	6	-3	5,5º	
4	6	-2	7º	
3	3	0		
8	6	2		2º
7	15	-8	2º	
13	12	1	8º	3º
10	14	-4	4º	
7	8	-1	8,5º	
8	8	0		
4	1	3		1º
10	11	-1	8,5º	
18	27	-9	1º	
16	22	-6	3º	
12	15	-3	5,5º	
11	11	0		
			53	5

Note que o preenchimento dos postos (posições) deve ser feito dos maiores para os menores. Por exemplo, o 1º posto negativo se refere à diferença (di = -9) que

é a maior diferença negativa, enquanto o 1º posto positivo se refere à diferença ($d_i = 3$), que é a maior diferença positiva. Quando houver diferenças iguais, devemos tirar a média entre as posições. Por exemplo, perceba que a diferença ($d_i = -1$) aparece duas vezes e, pela ordem dos postos positivos, uma seria a oitava maior e a outra a nona, porém, como são iguais, tiramos a média entre 8 e 9 e atribuímos a posição 8,5º as duas.

$T \Rightarrow$ é a menor soma dos postos. Vimos no exercício que em um posto ocorreu a soma 53 e no outro posto ocorreu a soma 5. Como a menor soma será o valor de T , então teremos que $T = 5$.

Vamos agora encontrar o valor da média estimada $\mu(t)$ que será dada por

$$\mu(t) = \frac{n(n+1)}{4} = \frac{12 \times 13}{4} = 39.$$

Por fim, encontraremos o desvio estimado

$$\begin{aligned} \text{dado por } \sigma(t) &= \sqrt{\frac{n(n+1)(2n+1)}{24}} \\ &= \sqrt{\frac{12 \times 13 \times 25}{24}} = 12,7. \end{aligned}$$

Agora podemos finalizar o cálculo da V.T, substituindo os três valores encontrados:

$$Z_{(cal)} = \frac{T - \mu(t)}{\sigma(t)} = \frac{5 - 39}{12,7} = -2,7$$

5º passo: comparando o valor do cálculo da VT com as regiões (RC e RA) do 3º passo, notamos que este resultado se encontra na região crítica da curva, assim rejeitaremos a hipótese H_0 . Desta forma, com uma confiabilidade de 94% (a qual foi a adotada no teste), podemos concluir que houve uma mudança significativa nas notas após a nova técnica de ensino adotada. Como o valor do cálculo da V.T cai na RC inferior da curva, conclui-se que a mudança ocorrida foi para mais, verificando-se uma maior diferença negativa entre o antes e o depois, ou seja, os números de acesso ao ambiente Moodle aumentaram mais do que diminuíram.



ATENÇÃO!

Tínhamos 15 alunos fazendo parte de nossa amostra no problema, mas verificamos que três desses alunos têm resultados do nº de acesso antes e depois iguais. Com isso, este aluno ficará excluído da amostra, portanto passamos a trabalhar com a amostra $n=12$.

Até agora, você recordou dos testes não-paramétricos os quais apuram comportamentos de uma variável. Daremos continuidade a eles com exercícios resolvidos que envolvem os testes não-paramétricos da correlação e da associação, os quais irão envolver relações entre duas variáveis. Vamos então prosseguir?

EXERCÍCIO RESOLVIDO 3

Com base na amostra abaixo, teste, ao nível de 98%, se existe correlação significativa entre o nº de acertos em um Quiz realizado pelo aluno de uma turma EaD e o nº de dias que este estudante passou sem acessar o ambiente virtual

Nº de dias (X)	Nº de acertos (Y)
5	5
7	4
5	6
3	7
0	9
1	9
2	8
2	9
6	5
8	3
0	10
1	10
2	8

VAMOS À NOSSA SOLUÇÃO:

Vamos aplicar os cinco passos do teste conforme visto na aula 8.

1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : \text{O nº de acertos não está correlacionado com o nº de dias sem acesso ao ambiente.} \\ H_1 : \text{O nº de acertos está correlacionado com o nº de dias sem acesso ao ambiente.} \end{cases}$$

2º passo: fixar o nível de significância $\alpha = 0,02$ e escolher a V.T do teste da correlação, que já sabemos ser uma $T_{\alpha}(n-2) = T_{0,02}(13-2) = T_{0,02}(11) = 6,965$. Iremos encontrar esse valor na tabela da distribuição T de Student abaixo:

Tabela t (student)

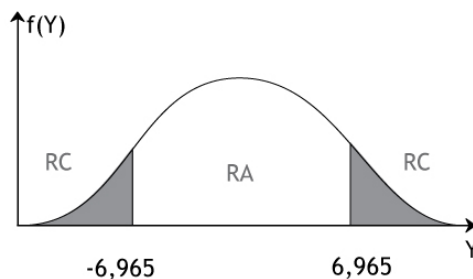
gl/ α	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
01	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
02	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
03	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,541	12,924
04	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
05	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
06	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
07	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,965	3,499	5,408
08	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
09	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,726
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,856	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,856	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646

Fonte: <http://www.somatematica.com.br/estat/tabelat.php>

Vimos, na tabela da distribuição T, que o valor tabelado é dado por

$$T_{0,02}(11) = 6,965$$

3º passo: definir a curva da distribuição que será dada por



4º passo: calcular a V.T (variável do teste) com base nos dados amostrais levantados.

Para tanto, teremos de aplicar a seguinte fórmula:

$$T_{cal} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = ?$$

Note que primeiro teremos de encontrar o coeficiente de correlação amostral de Pearson, o qual, conforme vimos na aula 6, será dado por $r = \frac{S_{x,y}}{\sqrt{S_{x,x} \cdot S_{y,y}}}$

Então, com base nos resultados encontrados na tabela de somatórios do nosso exercício, temos que

Nº de dias (X)	Nº de acertos (Y)	XY	X²	Y²
5	5	25	25	25
7	4	28	49	16
5	6	30	25	36
3	7	21	9	49
0	9	0	0	81
1	9	9	1	81
2	8	16	4	64
2	9	18	4	81
6	5	30	36	25
8	3	24	64	9
0	10	0	0	100
1	10	10	1	100
2	8	16	4	64
42	93	227	222	731

$$S_{x,y} = \sum XY - \frac{\sum X \cdot \sum Y}{N} = 227 - \frac{42 \times 93}{13} = -73,5$$

$$S_{x,x} = \sum X^2 - \frac{(\sum X)^2}{N} = 222 - \frac{(42)^2}{13} = 86,3$$

$$S_{y,y} = \sum Y^2 - \frac{(\sum Y)^2}{N} = 731 - \frac{(93)^2}{13} = 65,7.$$

$$\text{Assim, tivemos } r = \frac{S_{x,y}}{\sqrt{S_{x,x} \cdot S_{y,y}}} = \frac{-73,5}{\sqrt{86,3 \times 65,7}} = -0,98.$$

Então, substituindo este valor em nossa fórmula do T_{cal} , teremos

$$T_{cal} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,98\sqrt{13-2}}{\sqrt{1-0,98^2}} = 16,3$$

5º passo: comparando o resultado do cálculo da V.T com a curva da distribuição do 3º passo, verificamos que o resultado se encontra na região crítica, portanto rejeitaremos a hipótese H_0 . Assim, com uma confiabilidade de 98%, iremos concluir que existe sim uma correlação entre o nº de acertos no Quiz e o nº de dias que o aluno ficou sem acessar o ambiente.

EXERCÍCIO RESOLVIDO 4

Teste, ao nível de 97,5%, se existe relação de dependência entre a idade do estudante e a escolha pela área do seu curso de graduação. Tome por base os resultados apurados abaixo:

IDADE	ÁREAS			TOTAL
	A	B	C	
Menos de 20 anos	15	16	12	43
De 20 a 25 anos	8	10	6	24
De 26 a 31 anos	10	20	11	41
Mais de 31 anos	7	4	11	22
TOTAL	40	50	40	130

Solução:

Agora você terá apenas que seguir os cinco passos do teste da associação.

1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : \text{a escolha pela área do curso do aluno independe de sua idade} \\ H_1 : \text{a escolha pela área do curso do alunos depende da idade do aluno} \end{cases}$$

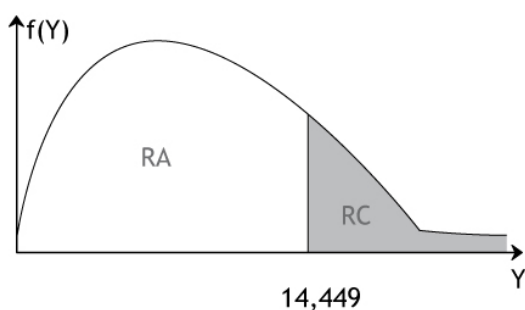
2º passo: encontrar a variável do teste ao nível de 97,5% de confiabilidade, ou seja, encontrar o valor do $\chi^2_{\alpha}(L-1)(C-1) = \chi^2_{0,025}(4-1)(3-1) = \chi^2_{0,025}(3)(2) = \chi^2_{0,025}(6)$. Olhando na tabela da distribuição qui-quadrado abaixo, temos que

Valores críticos (unilaterais à esquerda) da distribuição Qui-Quadrado P(
 χ^2 com n graus de liberdade \geq valor tabelado) = α

	0,995	0,99	0,975	0,95	0,9	0,1	0,05	0,025	0,01	0,005
1	0,000	0,000	0,001	0,004	0,016	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	9,236	11,070	12,832	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	2,833	12,017	14,067	16,013	18,475	20,278
8	1,344	1,647	2,180	2,733	3,490	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	4,168	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	4,865	15,987	18,307	20,483	23,209	25,188

$$\chi^2_{0,025}(6) = 14,449$$

3º passo: formar a curva da distribuição do qui-quadrado, destacando as regiões crítica e de aceitação.



4º passo: tomando por base a nossa amostra levantada, calcular a variável do teste (V.T)

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

Dessa forma, teremos as frequências observadas $Fo_{i,j}$.

IDADE	ÁREAS			TOTAL
	A	B	C	
Menos de 20 anos	15	16	12	43
De 20 a 25 anos	8	10	6	24
De 26 a 31 anos	10	20	11	41
Mais de 31 anos	7	4	11	22
TOTAL	40	50	40	130

Agora você irá, a partir das frequências observadas, encontrar as frequências esperadas, lembrando que cada $Fe_{i,j}$ (frequência esperada) será obtida da seguinte forma: $\frac{(\sum \text{linha} - i) \times (\sum \text{coluna} - j)}{n}$

Agora voltemos à fórmula do cálculo da VT, $\chi^2_{(cal)} = \sum_{i=1}^L \sum_{j=1}^c \frac{(Fo_{ij} - Fe_{ij})^2}{Fe_{ij}}$.

$$Fe_{i,j} = \begin{bmatrix} \frac{43 \times 40}{130} & \frac{43 \times 50}{130} & \frac{43 \times 40}{130} \\ \frac{24 \times 40}{130} & \frac{24 \times 50}{130} & \frac{24 \times 40}{130} \\ \frac{41 \times 40}{130} & \frac{41 \times 50}{130} & \frac{41 \times 40}{130} \\ \frac{22 \times 40}{130} & \frac{22 \times 50}{130} & \frac{22 \times 40}{130} \end{bmatrix} = \begin{bmatrix} 13,2 & 16,5 & 13,2 \\ 7,4 & 9,2 & 7,4 \\ 12,6 & 15,8 & 12,6 \\ 6,8 & 8,5 & 6,8 \end{bmatrix}$$

Assim teremos

$$\begin{aligned} \chi^2_{(cal)} = & \frac{(15-13,2)^2}{13,2} + \frac{(16-16,5)^2}{16,5} + \frac{(12-13,2)^2}{13,2} + \frac{(8-7,4)^2}{7,4} + \\ & + \frac{(10-9,2)^2}{9,2} + \frac{(6-7,4)^2}{7,4} + \frac{(10-12,6)^2}{12,6} + \frac{(20-15,8)^2}{15,8} + \\ & + \frac{(11-12,6)^2}{12,6} + \frac{(7-6,8)^2}{6,8} + \frac{(4-8,5)^2}{8,5} + \frac{(11-6,8)^2}{6,8} = 7,59 \end{aligned}$$

5º passo: comparar o resultado do $\chi^2_{(cal)}$ com as regiões (RA e RC) da curva do qui-quadrado do 3º passo. Vimos que o resultado do $\chi^2_{(cal)}$ se encontra na região de RA. Assim, aceitaremos a hipótese H_0 , ou seja, concluímos com 97,5% de confiabilidade que a escolha pela área do curso independe da idade do aluno.

Neste tópico, aplicamos exercícios envolvendo testes estatísticos não-paramétricos. Daremos a este assunto continuidade no último tópico, com exercícios que envolvam a aplicação dos principais testes estatísticos paramétricos.

TÓPICO 4

Exercícios sobre testes paramétricos

OBJETIVOS

- Testar, com certa confiabilidade, diferentes ocorrências paramétricas para um determinado levantamento de uma variável quantitativa, tomando por base levantamentos amostrais
- Testar, com certa confiabilidade, se existem diferenças significativas entre resultados de parâmetros, apurados em duas populações

Vamos iniciar recordando, através de exercícios resolvidos, como aplicar os testes da média e da proporção.

EXERCÍCIO RESOLVIDO 1

Se uma amostra de tamanho 25, selecionada de uma população de tamanho 400, acusou uma média de 231 e um desvio padrão 13 para uma determinada variável quantitativa. Com base nestes resultados, teste, adotando uma confiabilidade de 80%, se podemos aceitar que a média populacional para esta variável não ultrapassa a 240

Vamos à solução: inicialmente, vamos organizar o teste em seus cinco passos:

1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : \mu \leq 240 \\ H_1 : \mu > 240 \end{cases}$$

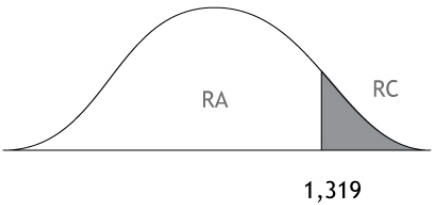
2º passo: fixar o nível de significância $\alpha = 0,20$, visto que a confiabilidade do teste é de 80%. Assim, de acordo com o critério do 2º passo, como o tamanho da amostra trabalhada $n=25 < 30$, iremos tomar para variável do teste (V.T), uma $T_{\alpha}(n-2)$. Desta forma, como já vimos em aulas anteriores, $T_{0,2}(25-2) = T_{0,2}(23) = 1,319$, conforme nos mostra a tabela da distribuição T-Student abaixo.

Tabela t (student)

gl/ α	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
01	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
02	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
03	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,541	12,924
04	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
05	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
06	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
07	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,365	3,499	5,408
08	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
09	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,726
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,856	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,856	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646

Fonte: <http://www.somatematica.com.br/estat/tabelat.php>

3º passo: definir as regiões críticas e de aceitação do teste de acordo com as hipóteses formuladas no 1º passo:



4º passo: como na nossa amostra $n=25 < 30$ e não conhecemos o desvio-padrão populacional, e sim o amostral, usaremos a seguinte metodologia de cálculo:

$$T_{cal} = \frac{\bar{X} - \mu}{S_x} = \frac{231 - 140}{\frac{13}{\sqrt{25}} \sqrt{\frac{400 - 25}{400 - 1}}} = -4,17$$

5º passo: comparando o resultado do cálculo da VT = -4,17 encontrado no 4º passo, com as regiões (R.A e R.C) do 3º passo, podemos notar que se encontra na região de aceitação, ou seja, aceitaremos a hipótese H_0 . Portanto, com a confiabilidade de 80% que foi adotada no teste, podemos afirmar que a média populacional não ultrapassa 240.

EXERCÍCIO RESOLVIDO 2

Um levantamento amostral realizado em um Município acusou que, de uma amostra de 450 jovens de 16 anos, apenas 108 já concluíram o ensino fundamental. Teste, com base neste levantamento, adotando um nível de confiabilidade de 90%, se podemos afirmar que a proporção de jovens na referida idade deste município, que ainda não concluíram o ensino fundamental é de 80% .

Solução:

Acompanhemos então os cinco passos do teste, ok?

1º passo: formular das hipóteses

$$\begin{cases} H_0 : P_0 = 80\% \\ H_1 : P_0 \neq 80\% \end{cases}$$

2º passo: fixar o nível de significância $\alpha = 0,10$, visto que a confiabilidade pedida no teste foi de 90% e como a variável para o teste da proporção será sempre uma normal padronizada Z_α . Encontraremos o valor da relação $= 1 - \frac{0,10}{2} = 0,95$. Assim, iremos procurar o valor de Z_α , que assuma a probabilidade mais próxima deste valor, conforme os resultados da tabela abaixo:

Tabela da Distribuição Normal Padrão

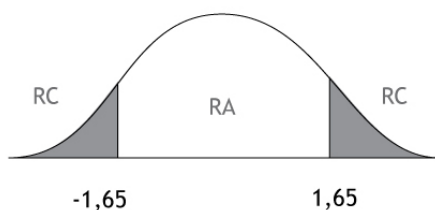
P(Z<z)	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517

0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

Fonte: www.pucrs.br/format/rossana/psicologia/tabela_normal.pdf

O valor de Z_{α} será dado por $Z_{\alpha} = 1,65$

3º passo: colocar o valor de Z_{α} encontrado na tabela normal padrão acima na curva da distribuição normal padronizada



4º passo: calcular o valor da V.T, com base no resultado amostral apurado da seguinte forma:

$$Z_{cal} = \frac{f - P_0}{\sqrt{\frac{P_0(1 - P_0)}{n}}} = \frac{0,76 - 0,80}{\sqrt{\frac{0,80 \times 0,20}{450}}} = -2,12, \text{ em que } f = \frac{342}{450} = 0,76$$

5º passo: comparando o valor do cálculo da V.T, com as regiões (RC e RA) do 3º passo, notamos que se encontra na região de aceitação da hipótese H_0 , ou seja, com uma confiabilidade de 90% que adotamos em nosso problema, podemos concluir que a proporção de jovens de 16 anos no Município que ainda não concluíram o ensino fundamental é de 80%.

EXERCÍCIO RESOLVIDO 3

Uma amostra de tamanho 33, oriunda de uma determinada população, apresenta uma média para uma variável igual a 8,2 e um desvio padrão de 0,9, enquanto outra amostra de tamanho 41, selecionada de uma outra população, apresentou uma média de 8,5 para a mesma variável, e uma variância de 1,8. Adote uma confiabilidade de 97% e teste se existe diferença significativa entre as médias das duas populações.

Solução:

Vamos seguir o teste, passo a passo:

1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : \mu_1 = \mu_2 \text{ ou } \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 \neq \mu_2 \text{ ou } \mu_1 - \mu_2 \neq 0, \text{ em que} \\ \mu_1 = \text{média populacional-1} \\ \mu_2 = \text{média populacional-2} \end{cases}$$

2º passo: fixar o nível de significância $\alpha = 0,03$ e, como no problema, os desvios padrões populacionais σ são conhecidos, pois $\sigma_1 = 0,9$, e como sabemos que $\sigma = \sqrt{\sigma^2}$, então temos que $\sigma_2 = \sqrt{1,8} = 1,3$. Iremos escolher para variável do teste uma normal padronizada Z_α . Encontraremos o valor da relação $= 1 - \frac{0,03}{2} = 0,985$. Assim, procuraremos o valor de Z_α que assuma a probabilidade mais próxima deste valor e conforme os resultados da tabela abaixo:

Tabela da Distribuição Normal Padrão

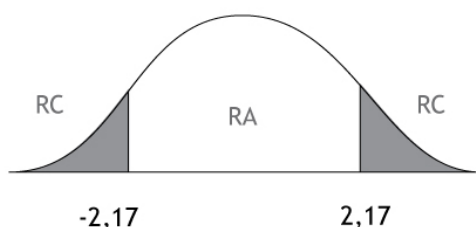
P(Z<z)	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177

1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

Fonte: www.pucrs.br/format/rossana/psicologia/tabela_normal.pdf

O valor de Z_α será então dado por $Z_\alpha = 2,17$

3º passo: de acordo com as regras vistas no 3º passo, podemos definir a curva da distribuição da seguinte forma:



4º passo: de acordo com a nossa V.T trabalhada, e com base nos dados amostrais levantados, iremos calcular a V.T (variável do teste) da seguinte forma:

$$Z_{cal} = \frac{(\bar{X}_1 - \bar{X}_2) - d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(8,2 - 8,5) - 0}{\sqrt{\frac{0,9^2}{33} + \frac{1,3^2}{41}}} = -1,15$$

5º passo: Comparando o resultado do cálculo da V.T, com a curva da distribuição do 3º passo, constatamos que o valor da V.T calculada se encontra na área de aceitação, ou seja, podemos concluir, com uma confiabilidade de 97%, a qual foi adotada em nosso problema, que não existe diferença significativa entre as médias das duas populações.

EXERCÍCIO RESOLVIDO 4

Supondo que um levantamento em dois dos nossos cursos da EaD e observamos que, em uma amostra de 36 alunos escolhidos aleatoriamente do curso-1, 22 postam suas tarefas até o limite das datas marcadas pelos tutores. Outra amostra de 39 alunos escolhidos aleatoriamente do curso-2, 23 postam suas tarefas com atraso. Com base nestas informações, teste, ao nível de 95% de confiança, se

existe diferença significativa entre as proporções de estudantes dos dois cursos que postam tarefas dentro dos prazos marcados pelos tutores

Solução:

Vamos então organizar os passos do teste:

1º passo: formular as hipóteses do teste

$$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$$

2º passo: fixar o nível de significância $\alpha = 0,05$, visto que a confiabilidade pedida no teste foi de 95% e, como a variável para este teste é sempre uma normal padronizada Z_α , iremos encontrar o valor da relação $= 1 - \frac{0,05}{2} = 0,975$. Assim, procure na tabela da distribuição normal padrão qual valor de Z_α que assume a probabilidade mais próxima a este resultado?

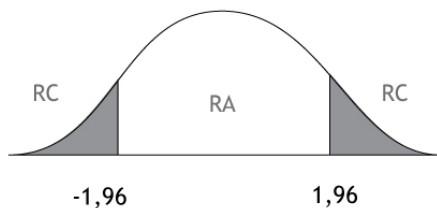
Tabela da Distribuição Normal Padrão

P(Z<z)	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

Fonte: www.pucrs.br/format/rossana/psicologia/tabela_normal.pdf

Encontraremos que será $Z_{\alpha} = 1,96$

3º passo: colocar o valor de Z_{α} encontrado na tabela normal padrão acima na curva da distribuição normal padronizada



4º passo: aplicar a metodologia de cálculo da variável do teste (V.T), com base nas apurações amostrais do problema.

$$f_1 = \frac{x_1}{n_1} = \frac{22}{36} = 0,61$$

$$f_2 = \frac{x_2}{n_2} = \frac{16}{39} = 0,41$$

$$p_* = \frac{x_1 + x_2}{n_1 + n_2} = \frac{22 + 16}{36 + 39} = \frac{38}{75} = 0,51$$

Assim, teremos

$$Z_{cal} = \frac{f_1 - f_2}{\sqrt{p_*(1-p_*)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0,61 - 0,41}{\sqrt{0,51 \times 0,49 \times \left(\frac{1}{36} + \frac{1}{39}\right)}} = 1,79$$

5º passo: comparando o resultado do cálculo da V.T = 1,79 com a curva da distribuição do 3º passo, verificaremos que o resultado se encontra na região de aceitação, ou seja, podemos aceitar, com 95% de confiança, que não existe diferença significativa entre as proporções de estudantes dos dois cursos que postam tarefas dentro dos prazos marcados pelos tutores

Bom, caro aluno, com a revisão dos testes paramétricos através das aplicações de exercícios, encerramos a nossa coletânea de exercícios para recordarmos as quatro últimas aulas: correlação e regressão; estimações; testes estatísticos não-paramétricos; e testes estatísticos paramétricos.

Desta forma, finalizamos o estudo da inferência estatística, concluindo assim o nosso material para disciplina de Estatística e Probabilidade, a qual foi composta de três partes: o estudo dos cálculos das probabilidades, o estudo da estatística descritiva e o estudo da inferência estatística.

Esperamos que aproveitem ao máximo todo o material elaborado e que obtenham sucesso no decorrer da disciplina, na conclusão de seu curso e no caminhar de sua vida.

REFERÊNCIAS

ANDERSON, David R; SWEENEY, Dennis J; WILLIAMS, Thomas A. **Estatística Aplicada à Administração e Economia**. 2. ed.

SPIEGEL, Murray R. **Probabilidade e Estatística**. coleção Schaum. Makron Books do Brasil, 1993.

HAZZAN, Samuel. **Fundamentos de matemática elementar**: combinatória, probabilidade. Vol 5. São Paulo: Atual, 1993.

MORETTIN, Luiz Gonzaga. **Estatística Básica**. Vol 1. 7. ed. São Paulo: Pearson Education do Brasil, 1999.

SIMON, J. Fonseca. **Curso de Estatística**. 5. ed. São Paulo: Pioneira Thomson Learning, 1995.

WALPOLE, Ronald E; MYERS, Raymond H; MYERS, Sharon L; YE Keying. **Probabilidade & Estatística para engenharia e ciências**. 8. ed.

CURRÍCULO

Paulo Maia Ferreira

Paulo Maia Ferreira é graduado em Estatística pelo Departamento de Estatística e Matemática Aplicada (DEMA) da Universidade Federal do Ceará (UFC), pós-graduado em Didática do Ensino Superior pela Universidade de Fortaleza (Unifor). Atuou como estatístico na Federação das Indústrias do Estado do Ceará (FIEC), como professor de graduação na Universidade de Fortaleza e na Universidade do Vale do Acaraú e de pós-graduação na Universidade Federal do Ceará em convênio com o Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE). Atualmente, é professor de Estatística e de Probabilidade no IFCE – campus Fortaleza.

