Computação

Estatística e Probabilidade

Francisco de Assis Amaral Bastos

Fortaleza - Ceará

Ed UECE

2015

Química
Ciências Biológicas
Artes Plásticas
Computação
Física
Matemática
Pedagogia

Editora Filiada à

ASSOCIAÇÃO BRASILEIRA DAS EDITORAS UNIVERSITÁRIAS

Editora da Universidade Estadual do Ceará – EdUECE
Av. Dr. Silas Munguba, 1700 – Campus do Itaperi – Reitoria – Fortaleza – Ceará
CEP: 60714-903 – Fone: (85) 3101-9893
Internet: www.uece.br – E-mail: eduece@uece.br

Secretaria de Apoio às Tecnologias Educacionais
Fone: (85) 3101-9962

Sumário

Apresentação

Este livro destina-se, aos alunos do Curso de Licenciatura em Computação ofertado pelo Centro de Ciências e Tecnologia da Universidade Estadual do Ceará, modalidade a distância, dentro do sistema Universidade Aberta do Brasil. No entanto, também pode ser utilizado por outros cursos, de licenciatura ou bacharelado, que tenham disciplinas de Estatística e Probabilidade em suas grades curriculares. Seu objetivo principal é apresentar aos alunos os conceitos fundamentais de Probabilidade e Inferência Estatística, habilitando-os à análise de dados em trabalhos acadêmicos ou científicos. Além da teoria, são dados exemplos de aplicação após cada assunto exposto, bem como propostos exercícios para avaliação dos conhecimentos adquiridos. Ao longo do texto são apresentadas em destaque observações e informações importantes, para melhor compreensão dos temas e complementação dos conhecimentos necessários. Também, com o intuito de “forçar” o aluno a resolver problemas para fixação dos assuntos abordados, é sugerida a solução de exercícios de forma complementar aos exemplos dados, bem como são postos desafios que, inclusive, necessitam do conhecimento de assuntos vistos em outras disciplinas do curso, como Fundamentos do Cálculo. Para facilitar a consulta a conhecimentos necessários ao aprendizado de Probabilidade, no final do livro são apresentados dois apêndices, com o resumo de Teoria dos Conjuntos e Técnicas de Contagem. Em outro apêndice estão as fórmulas utilizadas no livro, para consulta rápida pelos alunos, quando necessário. Um destaque é que todos os exercícios propostos estão resolvidos, para auxílio no caso de dúvidas na solução dos mesmos. Vale ressaltar aos alunos que isto não os exime da responsabilidade de resolvê-los, muito pelo contrário, cabe a cada um a percepção de como melhor utilizar esta facilidade. O livro é composto de quatro partes. Na Parte 1 são apresentados os conceitos de Probabilidade, ferramenta essencial para a compreensão dos processos vistos nas partes seguintes. Na Parte 2 estão os conceitos de Variáveis Aleatórias, onde são definidos modelos matemáticos, chamados de Modelos Probabilísticos, que generalizam o comportamento dessas variáveis, permitindo o cálculo de probabilidades em situações específicas que se enquadrem dentro de um caso genérico. A Parte 3 aborda a Inferência Estatística, imprescindível no trabalho científico no qual se pretende validar, probabilisticamente, hipóteses a respeito de parâmetros populacionais, este assunto depende intrinsicamente dos conhecimentos sobre Probabilidade e Variáveis Aleatórias, vistos nas partes anteriores. Por fim, a Parte 4 trata da aplicação de uma técnica bastante aplicada que é a análise do relacionamento funcional, de causa e efeito, entre duas ou mais variáveis e a estimação de um modelo matemático que permita prever o valor de uma variável em função de outra, para isto são discutidos o Coeficiente de Correlação e a Análise de Regressão.

O Autor

Parte I

Probabilidade

Capítulo 1

O Ambiente das Probabilidades

Introdução

No dia-a-dia, frequentemente nos deparamos com situações em que precisamos tomar decisões baseadas intuitivamente em probabilidades. Por exemplo, ao sair de casa é necessário avaliar se a probabilidade de chover é grande ou pequena para decidir-se quanto à roupa que será utilizada. Estas situações podem representar casos em que as consequências do erro na decisão assumida podem ser de extrema gravidade ou mesmo sem qualquer importância. Para o cidadão comum, apenas o conhecimento intuitivo das probabilidades é suficiente. No entanto, quando estão envolvidos problemas mais complexos, em qualquer área de conhecimento, é necessária uma formalização para a determinação das probabilidades, criando um embasamento teórico a ser aplicado, em especial, em trabalhos científicos.

O cálculo das probabilidades tem sua origem nos jogos de azar como, por exemplo, o lançamento de uma moeda. A princípio, caso haja evidências de que a moeda é "honesta", intuitivamente podemos considerar que as chances de ocorrência de cara de cara e coroa são iguais a 50%. Um problema surge caso haja desconfiança de que a moeda não seja "honesta", que é a forma de se determinar essas probabilidades, sendo este, então, o grande objetivo do estudo do Cálculo das Probabilidades.

Um fato importante que deve ser ressaltado é que, embora o cálculo das probabilidades tenha surgido com o estudo dos jogos de azar, ele se aplica a várias situações do cotidiano que têm comportamentos análogos a diversos jogos. Veja que o caso do lançamento de uma moeda, que possui apenas dois resultados possíveis, cara ou coroa, se encaixa em várias situações semelhantes, como, por exemplo, o processo de fabricação de peças que pode gerar peças boas ou defeituosas. Assim, a fundamentação teórica que vale para o lançamento de uma moeda também é aplicável à fabricação de peças num processo produtivo.

1 Experimentos Aleatórios

EXPERIMENTOS ALEATÓRIOS OU PROBABILÍSTICOS são aqueles que, ao serem repetidos em condições semelhantes, podem apresentar resultados distintos. Esta característica é o que fundamenta o estudo da Probabilidade, na medida em que, ao não sabermos o que exatamente irá acontecer, mas sim o que poderá acontecer, há o interesse em se mensurar a possibilidade de se obter determinados resultados. São exemplos de experimentos aleatórios:

E_1: Jogar um dado e anotar o número mostrado na face superior
E_2: Jogar uma moeda até que apareça uma cara e anotar a sequência de caras (C) e coroas (K) que aparecem
E_3: Testar uma lâmpada e anotar o tempo decorrido (em horas) até queimar

2 Espaço Amostral

O conjunto S de todos os resultados possíveis de um experimento aleatório é chamado de ESPAÇO AMOSTRAL. Para os experimentos dados acima, temos:

E_1: $S_1=\{1,2,3,4,5,6\}$
E_2: $S_2=\{C,KC,KKC,KKKC,...\}$
E_3: $S_3=\{t \in R \mid t \geq 0\}$

Os espaços amostrais podem ser classificados como:

- DISCRETOS – Quando os resultados do experimento são pontuais: S_1 e S_2
- CONTÍNUOS – Quando os resultados do experimento podem assumir quaisquer valores dentro de um intervalo ou união de intervalos reais: S_3

Necessariamente, os espaços contínuos são INFINITOS, enquanto que os discretos podem ser FINITOS ou INFINITOS ENUMERÁVEIS.

$S_1=\{1,2,3,4,5,6\}$	FINITO	DISCRETO
$S_2=\{C,KC,KKC,KKKC...\}$	FINITO ENUMERÁVEL	
$S_3=\{t \in R \mid t \geq 0\}$	INFINITO	CONTÍNUO

3 Eventos

No contexto deste livro, um EVENTO é qualquer subconjunto de S. Na prática, um evento corresponde ao subconjunto formado pelos resultados do experimento que atendem a determinadas características de interesse. Por exemplo, no lançamento de um dado uma única vez (experimento E_1 visto acima), se o interesse é a obtenção de um número par, o evento correspondente será o subconjunto $A=\{2,4,6\}$. Quando lançamos uma moeda até aparecer uma cara (experimento E_2), podemos estar interessados em saber qual a probabilidade de serem necessários no máximo 3 lançamentos para se obter cara, assim, $B=\{C,KC,KKC\}$. Ao se adquirir um determinado tipo de lâmpada (experimento E_3), pode-se estar interessado em saber qual a probabilidade de que a mesma dure pelo menos 500 horas, neste caso o evento é o subconjunto $C=\{t\epsilon R \mid t\geq 500\}$.

Se na realização de um experimento ocorrer um resultado x pertencente a um subconjunto A (EVENTO), então diz-se que o evento A ocorreu, caso contrário, A não ocorreu. Por exemplo, se lançarmos um dado e o resultado obtido é o número 6, como $6\in\{2,4,6\}$ então A (nº par) ocorreu.

Assim, podemos enunciar o seguinte:

1) Se A é subconjunto de S, então A é um EVENTO;
2) $\varnothing$ (conjunto vazio) é dito EVENTO IMPOSSÍVEL;
3) S (espaço amostral) é dito EVENTO CERTEZA.

Para espaços amostrais discretos, temos:

1) Se A é um conjunto unitário ele é dito um EVENTO SIMPLES;
2) Se A é um conjunto não unitário ele é dito EVENTO COMPOSTO, formado pela união de eventos simples;

Capítulo 2

Probabilidade

1 Probabilidade

1.1 Função Probabilidade

Observação

Os conceitos que serão vistos a seguir são baseados em espaços amostrais finitos, para melhor compreensão. Mas todos eles são válidos para espaços infinitos. Na Parte 2 será abordado como calcular probabilidades para o caso infinito, com a introdução do conceito de Variáveis Aleatórias.

Na realização de um experimento E, a probabilidade de ocorrência de um evento qualquer A é uma função P, aplicada aos eventos associados ao espaço amostral S, satisfazendo às seguintes condições:

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

2) $P(S) = 1$

3) $P\left(\bigcup_{i \in N^*} A_i\right) = \sum_{i \in N^*} P(A_i)$, se $A_i \cap A_j = \phi$ para $\forall i \neq j$

Observações

- P é uma função de conjuntos, P: $\wp(S) \rightarrow [0,1]$, onde $\wp(S)$ são as partes de S, ou seja, todos os subconjuntos (EVENTOS) de S;
- Se, para os conjuntos A_1, A_2,...,A_n, não existir qualquer interseção entre eles, ou seja, $A_i \cap A_j = \phi$ para $\forall i \neq j$ (condição 3), eles são ditos MUTUAMENTE EXCLUSIVOS.
- Na prática, percebe-se que a frequência relativa (ou percentual) satisfaz a essas condições, ou seja, calcular probabilidades é calcular percentuais.

Importante

Veja que trabalhar com probabilidade é trabalhar com operações de conjuntos. Portanto, é muito importante fazer uma revisão sobre o assunto, que pode ser encontrado no APÊNDICE 2

EXEMPLO 1.2.1.1[1]

Suponha que na fabricação de certo tipo de peça, ela pode apresentar 0,1,2 ou 3 defeitos, com probabilidades $\frac{4}{10}$, $\frac{3}{10}$, $\frac{2}{10}$ e $\frac{1}{10}$, respectivamente:

$$S = \{0,1,2,3\}$$

Eventos de S: $\varnothing,\{0\},\{1\},\{2\},\{3\},\{0,1\},\{0,2\},\{0,3\},\{1,2\},\{1,3\},\{2,3\}$
$\{0,1,2\},\{0,1,3\},\{0,2,3\},\{1,2,3\},\{0,1,2,3\}$

$P(\varnothing)=0 \quad P(S)=P(\{0,1,2,3\})=1$

$$A_1 = \{0\}; P(A_1) = \frac{4}{10} \qquad A_2 = \{1\}; P(A_2) = \frac{3}{10} \qquad A_3 = \{2\}; P(A_3) = \frac{2}{10}$$

$$A_4 = \{3\}; P(A_4) = \frac{1}{10} \qquad A_5 = \{1,2\}; P(A_5) = P(A_2) + P(A_3) = \frac{3}{10} + \frac{2}{10} = \frac{5}{10}$$

$$A_6 = \{1,2,3\}; P(A_6) = P(A_5) + P(A_4) = \frac{5}{10} + \frac{1}{10} = \frac{6}{10}$$

ou $P(A_6) = P(A_2) + P(A_3) + P(A_4) = \frac{3}{10} + \frac{2}{10} + \frac{1}{10} = \frac{6}{10}$

Observações

Veja que:

- $P(S) = P(\{0,1,2,3\}) = P(\{0\}) + P(\{1\}) + P(\{2\}) + P(\{3\}) = \frac{4}{10} + \frac{3}{10} + \frac{2}{10} + \frac{1}{10} = 1$
- $S=S\cup\varnothing \Rightarrow P(S)=P(S)+P(\varnothing) \Rightarrow 1=1+P(\varnothing) \Rightarrow P(\varnothing)=0$
- A soma das probabilidades dos eventos simples é igual a 1
- Conhecendo-se as probabilidades dos eventos simples, determinam-se as probabilidades de quaisquer eventos compostos

Numa situação prática, para este exemplo, podemos imaginar que em um processo de fabricação, de determinado fabricante, de certo tipo de peça, 40% $\left(\frac{4}{10}\right)$ delas não apresentam defeitos, 30% $\left(\frac{3}{10}\right)$ apresentam 1 defeito, 20% $\left(\frac{2}{10}\right)$ apresentam 2 defeitos e 10% $\left(\frac{1}{10}\right)$ apresentam 3 defeitos. Assim, um consumidor ao comprar uma dessas peças, a probabilidade de que ela não apresente defeito (A_1) é $\frac{4}{10}$ e de que ela seja defeituosa (A_6) é $\frac{6}{10}$. Outra interpretação importante é que, se o consumidor comprar várias peças, de uma única vez ou ao longo do tempo, espera-se que 40% delas não apresentem defeitos e 60% apresentem pelo menos um defeito.

1.2 Propriedades das probabilidades

1) $P(\phi)=0$

2) $P(\overline{A})=1-P(A)$, onde $\overline{A}$ é o complementar de A

3) $P\left(\bigcup_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i)$, se $A_i \cap A_j = \phi$ para $\forall i \neq j$

4) $P(A_1 \cup A_2)=P(A_1)+P(A_2)-P(A_1 \cap A_2)$

5) $P\left(\bigcup_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i)-\sum_{i<j=2}^{n} P\left(A_i \cap A_j\right)+$

$$+\sum_{i<j<r=3}^{n} P\left(A_i \cap A_j \cap A_r\right)-\ldots+(-1)^{n-1} P\left(A_1 \cap A_2 \cap \ldots \cap A_n\right)$$

Importante

A terceira propriedade enunciada acima consiste na **REGRA DA SOMA**, ou seja, a **PROBABILIDADE DA UNIÃO** de eventos é a **SOMA DAS PROBABILIDADES** de cada evento, observando-se que, caso haja alguma interseção entre eles, elas devem ser consideradas (somadas ou subtraídas) nesta soma, situação da quarta e quinta propriedades.

EXEMPLO 1.2.1.2[1]

Considere uma roleta com os números $1,2,3,4,5,6,7,8,9,10,11$ e 12, cujas probabilidades de serem obtidos são, respectivamente:

$$P(1)=\frac{1}{78},\ P(2)=\frac{2}{78},\ P(3)=\frac{3}{78},\ P(4)=\frac{4}{78},\ P(5)=\frac{5}{78},\ P(6)=\frac{6}{78},P(7)=\frac{7}{78},$$
$$P(8)=\frac{8}{78},\ P(9)=\frac{9}{78},\ P(10)=\frac{10}{78},\ P(11)=\frac{11}{78} \text{ e } P(12)=\frac{12}{78}$$

No experimento que consiste em rodar a roleta uma única vez, sejam os eventos:

A: o nº é par B: o nº é ímpar C: o nº é primo D: o nº é múltiplo de 4
E: o nº é múltiplo de 3, menor do que 12 e não primo

$S=\{1,2,3,4,5,6,7,8,9,10,11,12\}$

A={2,4,6,8,10,12}

$$P(A)=P(2)+P(4)+P(6)+P(8)+P(10)+P(12)=\frac{2}{78}+\frac{4}{78}+\frac{6}{78}+\frac{8}{78}+$$

$$+\frac{10}{78}+\frac{12}{78}=\frac{42}{78}$$

B={1,3,5,7,9,11}

$$P(B)=P(1)+P(3)+P(5)+P(7)+P(9)+P(11)=$$

$$=\frac{1}{78}+\frac{3}{78}+\frac{5}{78}+\frac{7}{78}+\frac{9}{78}+\frac{11}{78}=\frac{36}{78}$$

Observe que A e B são complementares, então verifica-se que $P(A)=1-P(B)=1-\frac{36}{78}=\frac{42}{78}$, bem como $P(B)=1-P(A)=1-\frac{42}{78}=\frac{36}{78}$.

$$C=\{2,3,5,7,11\}\Rightarrow P(C)=P(2)+P(3)+P(5)+P(7)+P(11)=$$

$$=\frac{2}{78}+\frac{3}{78}+\frac{5}{78}+\frac{7}{78}+\frac{11}{78}=\frac{28}{78}$$

$$D=\{4,8,12\}\Rightarrow P(D)=P(4)+P(8)+P(12)=\frac{4}{78}+\frac{8}{78}+\frac{12}{78}=\frac{24}{78}$$

$$E=\{6,9\}\Rightarrow P(E)=P(6)+P(9)=\frac{6}{78}+\frac{9}{78}=\frac{15}{78}$$

Como C, D e E são mutuamente exclusivos, não há qualquer interseção entre eles, $C\cap D=\phi$, $C\cap E=\phi$ e $D\cap E=\phi$, temos que:

$$P(C\cup D\cup E)=P(C)+P(D)+P(E)=\frac{28}{78}+\frac{24}{78}+\frac{15}{78}=\frac{67}{78}$$

Como, de fato:

$$C\cup D\cup E=\{2,3,4,5,6,7,8,9,11,12\}$$

$$P(C\cup D\cup E)=P(2)+P(3)+P(4)+P(5)+P(6)+P(7)+P(8)+P(9)+$$

$$+P(11)+P(2)=\frac{2}{78}+\frac{3}{78}+\frac{4}{78}+\frac{5}{78}+\frac{6}{78}+\frac{7}{78}+\frac{8}{78}+$$

$$+\frac{9}{78}+\frac{11}{78}+\frac{12}{78}=\frac{67}{78}$$

Como B e C não são mutuamente exclusivos, pois existe interseção entre eles, $B\cap C\neq\phi$, temos que:

$$B \cap C = \{3,5,7,11\} \Rightarrow P(B \cap C) = \frac{3}{78} + \frac{5}{78} + \frac{7}{78} + \frac{11}{78} = \frac{26}{78}$$

$$P(B \cup C) = P(B) + P(C) - P(B \cap C) = \frac{36}{78} + \frac{28}{78} - \frac{26}{78} = \frac{38}{78}$$

Como, de fato:

$$B \cup C = \{1,2,3,5,7,9,11\}$$

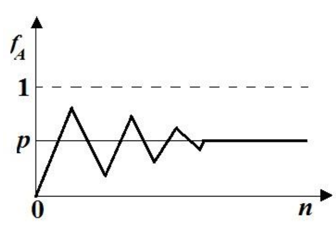
$$P(B \cup C) = P(1) + P(2) + P(3) + P(5) + P(7) + P(9) + P(11) = \frac{1}{78} + \frac{2}{78} + \frac{3}{78} + \frac{5}{78} +$$

$$+ \frac{7}{78} + \frac{9}{78} + \frac{11}{78} = \frac{38}{78}$$

2 Determinação das probabilidades

2.1 Frequência Relativa

Avaliando a definição de probabilidade, é fácil verificar que a frequência relativa (ou o percentual) atende às condições impostas. Desta forma, a probabilidade de ocorrência de um determinado evento A pode ser determinada (aproximada) a partir da realização do experimento uma certa quantidade de vezes n. Na medida que o experimento vai sendo repetido, a frequência relativa do evento A, $f_A = \frac{n_A}{n}$, onde n_A é a quantidade de vezes que A ocorreu nas n repetições, tende a se estabilizar em um valor p, que pode ser considerado como uma aproximação da probabilidade desejada. É evidente que, quanto maior a quantidade de repetições do experimento, melhor será a aproximação. Isto pode ser representado pela expressão e figura abaixo.



$$\lim_{n \to \infty} f_A = p$$

Importante

- Para experimentos com espaços amostrais infinitos (contínuos), o procedimento descrito acima é o que deve ser aplicado. Naturalmente, na maioria das situações práticas é inviável a realização repetida do experimento para se determinar probabilidades, nestes casos recorre-se a ocorrências passadas para estimação dessas probabilidades.
- Para espaços infinitos enumeráveis também deve ser aplicado este procedimento. Mas, se o experimento for composto pela realização de outros experimentos, como é o caso do experimento E_2, apresentado como exemplo na definição de experimento aleatório, onde o lançamento de uma moeda até a obtenção de uma cara é um experimento composto do experimento que é o lançamento de uma moeda uma única vez, com os resultados possíveis C ou K. Caso as probabilidades de ocorrência de C ou K forem previamente conhecidas elas determinam as probabilidades das sequências obtidas no evento composto. Isto será visto mais adiante, com a REGRA DO PRODUTO.

2.2 Espaços Amostrais Finitos

Seja E um experimento aleatório com uma quantidade finita n de resultados possíveis, ou seja, seu espaço amostral $S=\{a_1,a_2,\ldots,a_n\}$ é finito. Se P é uma probabilidade, então:

$$1)\ P(\{a_i\}) = p_i \qquad 0 \le p_i \le 1$$

$$2)\ \sum_{I=1}^{n} p_i = 1$$

Observações

- O resultado (1) acima indica que cada evento simples $\{a_i\}$ tem probabilidade p_i de ocorrer
- O item (2) resulta de que $P(S)=1$ (por definição):

$$S=\{a_1\}\cup\{a_2\}\cup\ldots\cup\{a_n\}\Rightarrow P(S)=P(\{a_1\})+P(\{a_2\})+\ldots+P(\{a_n\})\Rightarrow 1=\sum P(\{a_i\})$$

Importante

Para os espaços finitos em que seus eventos simples não tenham a mesma probabilidade de ocorrer, as probabilidades desses eventos devem ser calculadas pelo método da frequência relativa ou, a exemplo dos espaços infinitos enumeráveis, pela Regra do Produto se o experimento for composto pela realização de outros experimentos com probabilidades previamente conhecidas.

EXEMPLO 1.2.2.2[1]

Retornando ao exemplo 1.2.1.1[1], da fabricação de peças, vimos que:

$$S=\{0,1,2,3\}$$
$$P(\{0\})=\frac{4}{10};0\le P(\{0\})\le 1$$
$$P(\{1\})=\frac{3}{10};0\le P(\{1\})\le 1$$
$$P(\{2\})=\frac{2}{10};0\le P(\{2\})\le 1$$
$$P(\{3\})=\frac{1}{10};0\le P(\{3\})\le 1$$
$$P(\{0\})+P(\{1\})+P(\{2\})+P(\{3\})=\frac{4}{10}+\frac{3}{10}+\frac{2}{10}+\frac{1}{10}=\frac{10}{10}=1$$

3.3 Espaços Amostrais Finitos Equiprováveis

Seja E é um experimento com espaço amostral finito, $S=\{a_1,a_2,\dots,a_n\}$, e seus eventos simples, $\{a_1\}$, $\{a_2\}$,..., $\{a_n\}$, têm a mesma probabilidade de ocorrer, então ele é dito equiprovável e:

1) $P(\{a_i\})=p=\frac{1}{n}$

2) Se A é um evento composto, com k elementos, então $P(A)=\frac{k}{n}$

O resultado em (2) é comumente enunciado de uma das seguintes formas:

$$P(A)=\frac{\text{número de elementos de } A}{\text{número de elementos de } S};P(A)=\frac{\text{número de casos favoráveis a } A}{\text{número de casos possíveis}};P(A)=\frac{n(A)}{n(S)}$$

Observações

- Este resultado permite concluir que, em experimentos com espaços amostrais finitos equiprováveis, a probabilidade de ocorrência de determinado evento A é igual à proporção (ou percentual) de elementos pertencentes ao evento A em relação ao total de elementos do espaço amostral S.
- Esta interpretação, reforçada pelo método da frequência relativa, pode ser estendida para qualquer situação, ou seja, a probabilidade de ocorrência de um evento A é dada pela proporção favorável a A, em relação a todas as possibilidades. Por exemplo, se lançarmos aleatoriamente um objeto sobre uma superfície com 30% de sua área pintada de vermelho, a probabilidade de que o objeto atinja a cor vermelha é 0,30.

Importante

Observe que, em muitas situações, no caso de espaços amostrais finitos equiprováveis, é necessária a determinação da quantidade de elementos do espaço amostral e dos eventos. Assim, é importante uma revisão nas TÉCNICAS DE CONTAGEM, especialmente ANÁLISE COMBINATÓRIA, cujo conteúdo é encontrado no APÊNDICE 3

EXEMPLO 1.2.3.3[1]

Experimento: lançar um dado e observar a face superior

Considerando S equiprovável, temos:

$$S = \{1,2,3,4,5,6\}; n(S) = 6$$

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}$$

$$A \equiv n^{o}\ par \Rightarrow A = \{2,4,6\}; n(A) = 3; P(A) = \frac{n(A)}{n(S)} = \frac{3}{6} = \frac{1}{2}$$

EXEMPLO 1.2.3.3[2]

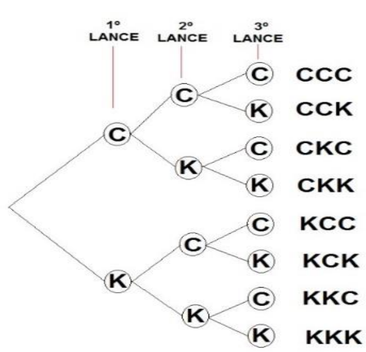
Experimento: lançar uma moeda 3 vezes e observar a sequência obtida de caras (C) e coroas (K):
Considerando S equiprovável, temos:

$$S = \{KKK, KKC, KCK, CKK, KCC, CKC, CCK, CCC\}; n(S) = 8$$

$$P(\{KKK\}) = P(\{KKC\}) = \ldots = P(\{CCC\}) = \frac{1}{8}$$

$$A \equiv 1\ cara \Rightarrow A = \{KKC, KCK, CKK\}; n(A) = 3; P(A) = \frac{n(A)}{n(S)} = \frac{3}{8}$$

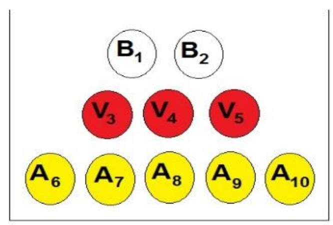
Uma outra maneira de representarmos o espaço amostral de um experimento é através de uma Árvore de Possibilidades. Para o exemplo 1.2.3.3[2], temos:



EXEMPLO 1.2.3.3[3]

Uma urna contém 10 bolas, sendo 2 brancas, 3 vermelhas e 5 amarelas.
Experimento: retirar 6 bolas da urna, simultaneamente.
Seja o evento $A \equiv$retirar exatamente 1 bola branca, 2 vermelhas e 3 amarelas

Ao contrário dos exemplos anteriores, aqui fica quase impraticável a enumeração dos elementos do espaço amostral e do evento em questão, face à quantidade dos mesmos. Recorreremos, então, à Análise Combinatória. Veja que, como a retirada é simultânea, não existe uma ordem de retirada nem é possível que uma mesma bola seja retirada mais de uma vez, ou seja, não há repetição. Assim será utilizada a COMBINAÇÃO SIMPLES, para a determinação dessas quantidades.



O espaço amostral S é composto por todas as sextuplas que podem ser formadas (não ordenadas e sem repetição) a partir das bolas existentes na urna:

$$n(S) = C_{10,6} = \frac{10!}{6!4!} = 210$$

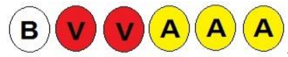
O evento A é formado por todas as sextuplas que podem ser formadas (sem ordem e sem repetição) a partir das bolas existentes na urna, contendo exatamente 1 bola branca, 2 vermelhas e 3 amarelas:

$$n(A) = C_{2,1} \times C_{3,2} \times C_{5,3} = 2 \times 3 \times 10 = 60 \Rightarrow P(A) = \frac{n(A)}{n(S)} = \frac{60}{210} = \frac{2}{7}$$

EXEMPLO 1.2.3.3[4]

Vamos considerar a mesma situação do exemplo 1.2.3.3[3], mas que a retirada das bolas seja feita de forma sucessiva, sem reposição. Ou seja, existe uma ordem na retirada das bolas, mas não é possível que uma bola seja retirada mais de uma vez. Assim para a contagem vamos utilizar ARRANJO SIMPLES. Então:

$$n(S) = A_{10,6} = \frac{10!}{4!} = 151.200$$

Uma sequência possível de ser obtida é 

, cuja quantidade é:

$$A_{2,1} \times A_{3,2} \times A_{5,3} = 2 \times 3 \times 60 = 360$$

Mas existem várias sequências possíveis, já que a retirada é ordenada, com 1 bola branca, 2 vermelhas e 3 amarelas, cada uma delas com 360 possibilidades. Para determinarmos todas essas sequências possíveis vamos utilizar a PERMUTAÇÃO DE ELEMENTOS REPETIDOS, aqui o B repetido 1 vez, o V, 2 vezes e o A, 3 vezes:

$$P_6^{1,2,3} = \frac{6!}{1!2!3!} = 60$$

Então:

$$n(A) = 60 \times 360 = 21.600 \Rightarrow P(A) = \frac{21.600}{151.200} = \frac{1}{7}$$

Capítulo 3

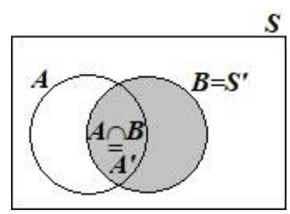
Probabilidade Condicional e Independência

1 Probabilidade Condicional (condicionada)

Para entendermos o conceito de probabilidade condicional, vamos considerar a seguinte situação: Num jogo que consiste no lançamento de um dado uma única vez, um apostador ganha determinada quantia se o número obtido for par, $A=\{2,4,6\}$. Assim, independente de qualquer condição, a probabilidade do apostador ganhar é $\frac{1}{2}$. Agora, se após o lançamento do dado for dada a informação de que o número obtido é maior do que 3, $B=\{4,5,6\}$, a probabilidade do apostador ganhar passa a ser $\frac{2}{3}$, ou seja, a probabilidade de A condicionada a B é $\frac{2}{3}$. Podemos, então, interpretar a probabilidade condicional da seguinte forma, considerando-se espaços amostrais finitos equiprováveis:

Como se sabe que B ocorreu, o espaço amostral reduz-se a $S' = B$. Portanto, $n(S')$ = $n(B)$. Em S', os casos favoráveis a A são os elementos de $A' = A \cap B$, então $n(A') = n(A \cap B)$, de tal forma que:

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



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

Define-se, então, a probabilidade condicional como:

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

EXEMPLO 1.3.1[1]

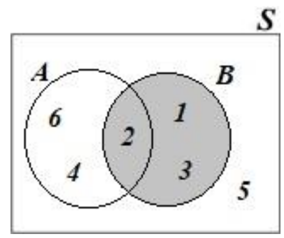
Um dado foi lançado e verificou-se que a face obtida apresenta um nº menor do que 4, qual a probabilidade desse nº ser par ?

$S = \{1,2,3,4,5,6\}$

$A = \{2,4,6\} \quad B = \{1,2,3\} \quad A \cap B = \{2\}$

$P(B) = \frac{3}{6} \quad P(A \cap B) = \frac{1}{6}$

$P(A|B) = \frac{1/6}{3/6} = \frac{1}{3}$



2 Eventos Independentes

Considerando ainda o jogo que consiste no lançamento de um dado uma única vez, no qual um apostador ganha determinada quantia se o número obtido for par, A=$\{2,4,6\}$, vimos que, independentemente de qualquer condição, a probabilidade do apostador ganhar é $\frac{1}{2}$. Dada a informação de que o número obtido é maior do que 3, B=$\{4,5,6\}$, a probabilidade do apostador ganhar passou a ser $\frac{2}{3}$, ou seja, a probabilidade de A mudou pelo fato de B ter ocorrido.

Mas, se a informação dada for de que o número obtido é maior ou igual a 3, B=$\{3,4,5,6\}$, a probabilidade do apostador ganhar não muda, continuando igual a $\frac{1}{2}\left(\frac{2}{4}\right)$. Neste caso, diz-se que os eventos A e B são independentes.

Assim, Se A e B são independentes, temos que:

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

Então, se e A e B são dois eventos, diz-se que eles são mutuamente independentes se e somente se:

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

Consequências:

1) A e B são $MI \Leftrightarrow P(A|B) = P(A)$ e $P(B|A) = P(B)$

2) Se $A \cap B = \phi$ então A e B são $MI \Leftrightarrow P(A) = 0$ ou $P(B) = 0$

Observações

- O resultado encontrado em (1) mostra que se A e B são independentes, a ocorrência de B não afeta a probabilidade de ocorrência de A. Ou seja, a proporção de elementos de A em S (espaço amostral) é a mesma proporção de elementos de A em B. Em outras palavras, é indiferente (independente) trabalharmos em S ou em B, pois a probabilidade de ocorrência de A é a mesma em ambos os casos, com o mesmo valendo para B em relação a A.
- Já o resultado encontrado em (2), chama a atenção para um erro comum que é, ao se dizer que dois eventos, A e B, são independentes, considerá-los exclusivos, ou seja, $A \cap B = \phi$, ou vice e versa.

Para exemplificar a primeira observação acima, suponha a seguinte situação em relação à preferência de 200 usuários, em relação a três provedores de Internet:

SEXO	PROVEDOR			Total
	I	II	III	
Masculino (M)	20	50	10	80
Feminino (F)	30	40	50	120
Total	50	90	60	200

Se considerarmos todos os usuários, o percentual que prefere o provedor I é de 25% $\left(\frac{50}{200}\right)$. Se considerarmos apenas os homens a preferência pelo provedor I continua 25% $\left(\frac{20}{80}\right)$, ou seja, $P(I|M) = P(I) = 0{,}25$ então I e M são independentes. O mesmo não acontece com o provedor II, cuja preferência considerando-se todos os usuários é de 45% $\left(\frac{90}{200}\right)$, mas de 62,5% $\left(\frac{50}{80}\right)$ entre os homens, ou seja $P(II|M) \neq P(II)$ então II e M não são independentes.

EXEMPLO 1.3.2[1]

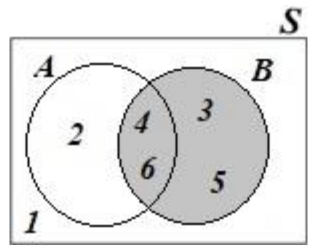
Um dado foi lançado e verificou-se que a face obtida apresenta um nº maior que 2, qual a probabilidade desse nº ser par ?

$S = \{1,2,3,4,5,6\}$

$A = \{2,4,6\} \quad B = \{3,4,5,6\} \quad A \cap B = \{4,6\}$

$P(A) = \frac{3}{6} = \frac{1}{2} \quad P(B) = \frac{4}{6} = \frac{2}{3} \quad P(A \cap B) = \frac{2}{6} = \frac{1}{3}$

$P(A|B) = \frac{1/3}{2/3} = \frac{1}{2} = P(A)$



Conclui-se então que A e B são independentes. Observe que a definição de eventos independentes é atendida:

$$P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{2}{3} = \frac{2}{6} = \frac{1}{3} = P(A \cap B)$$

EXEMPLO 1.3.2[2]

Voltando ao problema das peças, exemplo 1.2.1.1[1], que apresentam $0,1,2$, ou 3 defeitos com probabilidades $\frac{1}{10}$, $\frac{2}{10}$, $\frac{3}{10}$ e $\frac{4}{10}$ respectivamente, se um cliente compra uma dessas peças, a probabilidade de que ela apresente no máximo 2 defeitos é $\frac{6}{10}$. Mas, se o vendedor informar ao cliente que está vendendo a peça por um preço menor por ela ser defeituosa a probabilidade de que a peça apresente no máximo 2 defeitos passa a ser $\frac{5}{9}$. Portanto estes dois eventos não são independentes:

$S=\{0,1,2,3\}$

$A \equiv$ a peça apresenta no máximo 2 defeitos; $A=\{0,1,2\}$; $P(A) = \frac{1}{10} + \frac{2}{10} + \frac{3}{10} = \frac{6}{10}$

$B \equiv$ a peça é defeituosa; $B=\{1,2,3\}$; $P(B) = \frac{2}{10} + \frac{3}{10} + \frac{4}{10} = \frac{9}{10}$

$A \cap B = \{1,2\}; P(A \cap B) = \frac{2}{10} + \frac{3}{10} = \frac{5}{10}$ A∩B={1,2}; P(A∩B)=2/10+3/10=5/10

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{5/10}{9/10} = \frac{5}{9} \neq P(A)$$

Verificando pela definição de eventos independentes:

$P(A) \cdot P(B) = \frac{6}{10} \cdot \frac{9}{10} = \frac{27}{50} \neq P(A \cap B)$

A definição de eventos independentes pode ser estendida para mais de dois eventos, A_1, A_2, ... , A_n. Neste caso, os n eventos são ditos MUTUAMENTE INDEPENDENTES se e somente se as probabilidades de todas as interseções possíveis (2 a 2, 3 a 3,...) são iguais ao produto das probabilidades dos eventos incluídos na interseção:

$$
\begin{aligned}
&P\left(A_i \cap A_j\right) = P(A_i)P(A_j), i < j, j = 2,3,\ldots,n \\
&P\left(A_i \cap A_j \cap A_k\right) = P(A_i)P(A_j)P(A_k), i < j < k, k = 3,4,\ldots,n \\
&P\left(A_i \cap A_j \cap A_k \cap A_l\right) = P(A_i)P(A_j)P(A_k)P(A_l), i < j < k < l, l = 4,5,\ldots,n \\
&\ldots \\
&P\left(A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_n\right) = P(A_1)P(A_2)P(A_3)\ldots P(A_n)
\end{aligned}
$$

OBSERVAÇÃO: ao todo são 2^n-n-1 operações a serem verificadas

Por exemplo, no caso de 4 eventos, A_1, A_2, A_3 e A_4, para que eles sejam mutuamente independentes devemos ter:

$$
\begin{aligned}
&P(A_1 \cap A_2) = P(A_1)P(A_2) \quad && P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) \\
&P(A_1 \cap A_3) = P(A_1)P(A_3) \quad && P(A_1 \cap A_2 \cap A_4) = P(A_1)P(A_2)P(A_4) \\
&P(A_1 \cap A_4) = P(A_1)P(A_4) \quad && P(A_1 \cap A_3 \cap A_4) = P(A_1)P(A_3)P(A_4) \\
&P(A_2 \cap A_3) = P(A_2)P(A_3) \quad && P(A_2 \cap A_3 \cap A_4) = P(A_2)P(A_3)P(A_4) \\
&P(A_2 \cap A_4) = P(A_2)P(A_4) \quad && P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1)P(A_2)P(A_3)P(A_4) \\
&P(A_3 \cap A_4) = P(A_3)P(A_4) &&
\end{aligned}
$$

$n = 4 \Rightarrow 2^4 - 4 - 1 = 11$ operações

EXEMPLO 1.3.2[3]

Uma urna contem 3 bolas vermelhas (v_1,v_2 e v_3) e 4 bolas brancas (b_4,b_5,b_6 e b_7). Seja o experimento que consiste na retirada de uma amostra ordenada com reposição de 4 bolas da urna.

Sejam os eventos: $A_1 \equiv$ a primeira bola retirada é branca, $A_2 \equiv$ a segunda bola retirada é branca, $A_3 \equiv$ a terceira bola retirada é branca e $A_4 \equiv$ a quarta bola retirada é branca. Mostre que A_1,A_2,A_3 e A_4 são MUTUAMENTE INDEPENDENTES.

$$n(S) = A^*_{7,4} = 7^4 = 2.401$$

$A_1 \equiv b\bullet\bullet\bullet$ (o símbolo • significa que a bola pode ser de qualquer cor, banca ou vermelha)

$$n(A_1) = A^*_{4,1} \times A^*_{7,3} = 4 \times 7^3 = 4 \times 343 = 1.372 \Rightarrow P(A_1) = \frac{1.372}{2.401} = \frac{4}{7}$$

De forma análoga encontramos $P(A_2) = \frac{4}{7}; P(A_3) = \frac{4}{7}; P(A_4) = \frac{4}{7}$

$A_1 \cap A_2 \equiv bb\bullet\bullet$

$$n(A_1 \cap A_2) = A^*_{4,2} \times A^*_{7,2} = 4^2 \times 7^2 = 16 \times 49 = 784 \Rightarrow P(A_1 \cap A_2) = \frac{784}{2.401} = \frac{16}{49}$$

De forma análoga encontramos

$$P(A_1 \cap A_3) = \frac{16}{49}; P(A_1 \cap A_4) = \frac{16}{49}; P(A_2 \cap A_3) = \frac{16}{49}; P(A_2 \cap A_4) =$$
$$= \frac{16}{49}; P(A_3 \cap A_4) = \frac{16}{49}$$

Verifica-se, então, que

$$P(A_i \cap A_j) = P(A_i)P(A_j) = \frac{4}{7} \times \frac{4}{7} = \frac{16}{49}; i < j; j = 2,3,4$$

$A_1 \cap A_2 \cap A_3 \equiv bbb\bullet$

$$n(A_1 \cap A_2 \cap A_3) = A^*_{4,3} \times A^*_{7,1} = 4^3 \times 7^1 = 64 \times 7 = 448 \Rightarrow P(A_1 \cap A_2 \cap A_3) = \frac{448}{2.401} = \frac{64}{343}$$

De forma análoga encontramos

$$P(A_1 \cap A_2 \cap A_4) = \frac{64}{343}; P(A_1 \cap A_3 \cap A_4) = \frac{64}{343}; P(A_2 \cap A_3 \cap A_4) = \frac{64}{343}$$

Verifica-se, então, que

$$P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k) = \frac{4}{7} \times \frac{4}{7} \times \frac{4}{7} = \frac{64}{343}; i < j < k, k = 3,4$$

$A_1 \cap A_2 \cap A_3 \cap A_4 \equiv bbbb$

$$n(A_1 \cap A_2 \cap A_3 \cap A_4) = A^*_{4,4} = 256 \Rightarrow P(A_1 \cap A_2 \cap A_3 \cap A_4) = \frac{256}{2.401}$$

Verifica-se, então, que

$$P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1)P(A_2)P(A_3)P(A_4) = \frac{4}{7} \times \frac{4}{7} \times \frac{4}{7} \times \frac{4}{7} = \frac{256}{2.401}$$

Assim, como todas as 11 operações foram satisfeitas, concluímos que A_1, A_2, A_3 e A_4 são MUTUAMENTE INDEPENDENTES.

A partir da expressão da probabilidade condicional, $P(A_2|A_1)=\dfrac{P(A_1\cap A_2)}{P(A_1)}$, chegamos ao resultado:

$$P(A_1\cap A_2)=P(A_1)P(A_2|A_1)$$

Estendendo para n eventos, tem-se:

$$P(A_1\cap A_2\cap\ldots\cap A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1\cap A_2)\ldots P(A_n|A_1\cap A_2\cap\ldots\cap A_{n-1})$$

Se os eventos forem MUTUAMENTE INDEPENDENTES, passamos a ter:

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

e

$$P(A_1\cap A_2\cap\ldots\cap A_n)=P(A_1)P(A_2)P(A_3)\ldots P(A_n)$$

Importante

Os resultados enunciados acima constituem o que se denomina REGRA DO PRODUTO, ou seja, a probabilidade da interseção de dois ou mais eventos é igual ao produto das probabilidades de ocorrência desses eventos, sendo que a probabilidade seguinte é sempre condicionada ao que ocorreu anteriormente.

No Capítulo 2, ítem 3, desta partee, foi destacado como importante que, para experimentos com espaços amostrais infinitos enumeráveis ou finitos não equiprováveis, as probabilidades de seus eventos poderiam ser calculadas a partir da regra do produto, desde que esses experimentos sejam compostos por um outro experimento, que chamaremos de original, e que sejam conhecidas as probabilidades dos eventos simples do experimento original. O exemplo a seguir ilustra este fato.

EXEMPLO 1.3.2[4]

Desafio

Determine uma "fórmula" para calcular a probabilidade de que sejam necessários exatamante x lançamentos, veja que x=1,2,3,4,... (neste exemplo x = 4)

Uma moeda, que apresenta cara duas vezes mais do que coroa, é lançada até que apareça uma cara (C). Qual a probabilidade de que sejam necessários extamente 4 lançamentos ?

Aqui, o experimento original é o lançamento de uma moeda uma única vez, com $P(C)$ e $P(K)$ dadas por:

$$P(C)=2P(K);P(C)+P(K)=1\Rightarrow 2P(K)+P(K)=1\Rightarrow P(K)=\frac{1}{3};P(C)=\frac{2}{3}$$

$$S=\{C,KC,KKC,KKKC,KKKKC,KKKKKC,\ldots\}$$

A_1="K no primeiro lançamento"
A_2="K no segundo lançamento"
A_3="K no terceiro lançamento"
A_4="C no quarto lançamento"

$$A = A_1 \cap A_2 \cap A_3 \cap A_4 = \{KKKC\}$$

Como A_1, A_2, A_3 e A_4 são mutuamente independentes, ou seja, as probabilidades de cara e coroa não se alteram a cada lançamento, temos:

$$P(A) = P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1)P(A_2)P(A_3)P(A_4) =$$
$$= P(K)P(K)P(K)P(C) = \frac{1}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{2}{3} = \frac{2}{81}$$

No exemplo a seguir aliamos a árvore de possibiidades com a regra do produto para o cáculo de probabilidades.

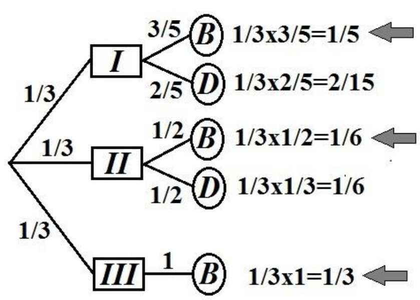
EXEMPLO 1.3.2[5]

Em um balcão existem 3 gavetas I, II e III. Na gaveta I estão 5 peças, sendo 3 boas e 2 defeituosas, na gaveta II são 8 peças, 4 boas e 4 defeituosas, já a gaveta III contém 10 peças, todas boas. Um técnico seleciona ao acaso uma gaveta e dela retira, também ao acaso, uma peça. Qual a probabilidade dessa peça ser boa ?

Sejam os eventos:

$I \equiv$ "gaveta I" ; $II \equiv$"gaveta II ; $III \equiv$ "gaveta III" ; $B \equiv$ "peça boa" ; $D \equiv$ "peça defeituosa"

$$P(I) = \frac{1}{3}; P(II) = \frac{1}{3}; P(III) = \frac{1}{3}; P(B\,|\,I) = \frac{3}{5}$$
$$; P(B\,|\,II) = \frac{1}{2}; P(B\,|\,III) = 1$$
$$B = (I \cap B) \cup (II \cap B) \cup (III \cap B)$$
$$P(B) = P(I \cap B) + P(II \cap B) + P(III \cap B) =$$
$$= P(I)P(B\,|\,I) + P(II)PB\,|\,II) + P(III)P(B\,|\,III) =$$
$$= \frac{1}{3}\cdot\frac{3}{5} + \frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot 1 = \frac{1}{5} + \frac{1}{6} + \frac{1}{3} = \frac{7}{10}$$



$P(B)$=1/5+1/6+1/3=7/10

Sugestão

Represente o exemplo anterior (1.3.2[4]) através de uma árvore de possibilidades

EXEMPLO 1.3.2[6]

Considere uma urna com 10 bolas, sendo 4 brancas (b) e 6 vermelhas (v). Dessa urna são retiradas, sucessivamente, 5 bolas. Qual a probabilidade de que a sequência das cores retiradas seja $bvvbv$?

Definamos os eventos: A_1≡bola branca na 1ª retirada, A_2≡bola vermelha na 2ª retirada, A_3≡bola vermelha na 3ª retirada, A_4≡bola branca na 4ª retirada e A_5≡bola vermelha na 5ª retirada.

O que se deseja é calcular $P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5)$

1º caso

Se as bolas forem retiradas sem reposição, os eventos A_1, A_2, A_3, A_4 e A_5 não são mutuamente independentes, pois, após a retirada de uma bola, na retirada seguinte as probabilidades de retirada de uma determinada cor são diferentes das anteriores, em função da diminuição da quantidade de bolas na urna. Então:

$$P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) = P(A_1) \times P(A_2|A_1) \times P(A_3|A_1 \cap A_2) \times P(A_4|A_1 \cap A_2 \cap A_3) \times$$
$$\times P(A_5|A_1 \cap A_2 \cap A_3 \cap A_4) == \frac{4}{10} \cdot \frac{6}{9} \cdot \frac{5}{8} \cdot \frac{3}{7} \cdot \frac{4}{6} = \frac{1}{21}$$

2º caso

Se as bolas forem retiradas com reposição, os eventos A_1, A_2, A_3, A_4 e A_5 são mutuamente independentes, pois, como após a retirada de uma bola a mesma é recolocada na urna, na retirada seguinte as probabilidades de retirada de uma determinada cor são iguais às anteriores, pois a quantidade de bolas na urna permanece a mesma. Então:

$$P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) = P(A_1)P(A_2)P(A_3)P(A_4)P(A_5) = \frac{4}{10} \cdot \frac{6}{10} \cdot \frac{6}{10} \cdot \frac{4}{10} \cdot \frac{6}{10} =$$
$$= \frac{108}{3.125}$$

Capítulo 4

Teorema da Probabilidade Total e Teorema de Bayes

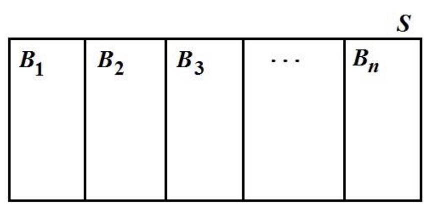
1 Partição do Espaço Amostral

Seja S o espaço amostral associado a um experimento E. Se $B_1, B_2, \ldots, B_n$ são eventos satisfazendo a:

$$1)\ \bigcup_{i=1}^{n} B_i = S$$

$$2)\ B_i \cap B_j = \phi \quad \forall\ i \neq j$$

$$3)\ P(B_i) > 0 \quad \forall\ i = 1,2,\ldots,n$$



Então os eventos B_i formam uma PARTIÇÃO do espaço amostral S.

EXEMPLO 1.4.1[1]

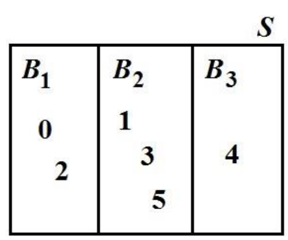
$$S = (0,1,2,3,4,5\}$$

$$B_1 = \{0,2\} \quad B_2 = \{1,3,5\} \quad B_3 = \{4\}$$

$$B_1 \cup B_2 \cup B_3 = \{0,1,2,3,4,5\} = S$$

$$P(B_1) = \frac{2}{6} > 0 \quad P(B_2) = \frac{3}{6} > 0 \quad P(B_3) = \frac{1}{6} > 0$$

$$B_1 \cap B_2 = \phi \quad B_1 \cap B_3 = \phi \quad B_2 \cap B_3 = \phi$$

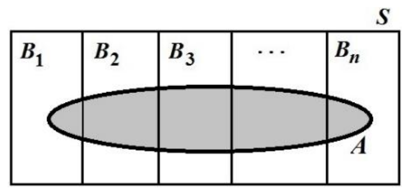


Na prática podemos ter situações como, por exemplo, a partição de uma população de acordo com o sexo B_1≡Masculino, B_2≡Feminino, ou as peças de determinado tipo podem ser separadas por fabricante, B_1≡Fabricante A, B_2≡Fabricante B, B_3≡Fabricante C e B_4≡Outros fabricantes.

2 Teorema da Probabilidade Total

Se B_1, B_2,...,B_n são uma partição de um espaço amostral S e A um evento qualquer de S, então:

$$P(A) = \sum_{i=1}^{n} P(A|B_i)\cdot P(B_i)$$



EXEMPLO 1.4.2[1]

Uma urna contém 60 bolas, numeradas de 1 a 60. Uma bola é retirada, aleatoriamente, desta urna. Qual a probabilidade do número da bola retirada ser primo ?

Sejam os eventos:

B_1≡o nº da bola retirada é menor ou igual a 10
B_2≡o nº da bola retirada é maior do que 10 e menor ou igual a 30
B_3≡o nº da bola retirada é maior que 30 e menor ou igual a 60
A≡o nº da bola retirada é primo.

$$S = \{1,2,3,\ldots,60\}; n(S) = 60$$

$$B_1 = \{1,2,3,\ldots,10\}; n(B_1) = 10; P(B_1) = \frac{10}{60} = \frac{1}{6}$$

$$B_2 = \{11,12,13,\ldots,30\}; n(B_2) = 20; P(B_2) = \frac{20}{60} = \frac{1}{3}$$

$$B_3 = \{31,32,33,\ldots,60\}; n(B_3) = 30; P(B_3) = \frac{30}{60} = \frac{1}{2}$$

$$A = \{2,3,5,7,11,13,17,19,23,29,31,37,41,43,47,49,53,59\}; n(A) = 18$$

$$P(A) = \frac{18}{60} = \frac{3}{10}$$

$$B_1 \cap A = \{2,3,5,7\}; n(B_1 \cap A) = 4; P(B_1 \cap A) = \frac{4}{60}; P(A|B_1) = \frac{4}{10} = \frac{2}{5}$$

$$B_2 \cap A = \{11,13,17,19,23,29\}; n(B_2 \cap A) = 6; P(B_2 \cap A) = \frac{6}{60}$$

$$P(A|B_2) = \frac{6}{20} = \frac{3}{10}$$

$$B_3 \cap A = \{31,37,41,43,47,49,53,59\}; n(B_3 \cap A) = 8$$

$$P(B_3 \cap A) = \frac{8}{60}; P(A|B_3) = \frac{8}{30} = \frac{4}{15}$$

Veja que a probabilidade de A já está determinada acima mas, aplicando o Teorema da Probabilidade Total, temos:

$$P(A)=P(B_1)P(A\,|\,B_1)+P(B_2)P(A\,|\,B_2)+P(B_3)P(A\,|\,B_3)=$$
$$=\frac{1}{6}\cdot\frac{2}{5}+\frac{1}{3}\cdot\frac{3}{10}+\frac{1}{2}\cdot\frac{4}{15}=\frac{1}{15}+\frac{1}{10}+\frac{2}{15}=\frac{45}{150}=\frac{3}{10}$$

EXEMPLO 1.4.2[2]

25% de determinado tipo de peça são produzidas pelo fabricante A, 20% pelo B, 45% pelo C e 10% por outros (O) fabricantes. 2%, 4%, 3% e 5% das peças produzidas por A, B, C e O, respectivamente, são defeituosas. Um consumidor, ao comprar uma dessas peças, qual a probabilidade dele adquirir uma peça defeituosa (d) ?

$$P(A)=0{,}25 \quad P(d\,|\,A)=0{,}02$$
$$P(B)=0{,}20 \quad P(d\,|\,B)=0{,}04$$
$$P(C)=0{,}45 \quad P(d\,|\,C)=0{,}03$$
$$P(O)=0{,}10 \quad P(d\,|\,O)=0{,}05$$
$$P(d)=P(A)P(d\,|\,A)+P(B)P(d\,|\,B)+P(C)P(d\,|\,C)+P(O)P(d\,|\,O)=$$
$$=0{,}25\times0{,}02+0{,}20\times0{,}04+0{,}45\times0{,}03+0{,}10\times0{,}05=$$
$$=0{,}005+0{,}008+0{,}0135+0{,}005=0{,}0315=3{,}15\%$$

Sugestão

Calule $P(b)$, onde b é peça boa, e interprete o resultado

Na prática, temos que $3{,}15\%$ desse tipo de peça são defeituosas.

3 Teorema de Bayes

Seja S um espaço amostral associado a um experimento E. Se B_1, B_2, ..., B_n são uma partição de S e A um evento, com $P(A)>0$, então:

$$P(B_j|A)=\frac{P(A|B_j)\cdot P(B_j)}{\sum_{i=1}^{n}P(A|B_i)\cdot P(B_i)}$$

EXEMPLO 1.4.3[1]

Considerando o exemplo 1.4.2[1], da urna com 60 bolas, se a bola retirada apresentar número primo, qual a probabilidade de o número seja menor ou igual a 10 ?

O que se está desejando é a probabilidade $P(B_1 \mid A)$, que é dada por:

$$P(B_1 \mid A) = \frac{P(B_1)P(A \mid B_1)}{P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2) + P(B_3)P(A \mid B_3)} =$$

$$= \frac{\frac{1}{6} \cdot \frac{2}{5}}{\frac{1}{6} \cdot \frac{2}{5} + \frac{1}{3} \cdot \frac{3}{10} + \frac{1}{2} \cdot \frac{4}{15}} = \frac{1/15}{3/10} = \frac{10}{45}$$

EXEMPLO 1.4.3[2]

Considerando, agora, o exemplo 1.4.2[2], se a peça comprada é defeituosa, qual a probabilidade de ter sido produzida pelo fabricante B ?

O que se deseja é $P(B \mid d)$:

Sugestão

Calule $P(A|d)$, $P(C|d)$, $P(O|d)$, $P(A|b)$, $P(B|b)$, $P(C|b)$ e $P(O|b)$, onde b é peça boa e interprete os resultados

$$P(B \mid d) = \frac{P(B)P(d \mid B)}{P(A)P(d \mid A) + P(B)P(d \mid B) + P(C)P(d \mid C) + P(O)P(d \mid O)} =$$

$$= \frac{0{,}20 \times 0{,}04}{0{,}25 \times 0{,}02 + 0{,}20 \times 0{,}04 + 0{,}45 \times 0{,}03 + 0{,}10 \times 0{,}05} =$$

$$= \frac{0{,}008}{0{,}005 + 0{,}008 + 0{,}0135 + 0{,}005} = \frac{0{,}008}{0{,}0315} = 0{,}254 = 25{,}4\%$$

Na prática isto siginfica que se forem separadas todas as peças defeituosas, 25,4% delas são do fabricante B.

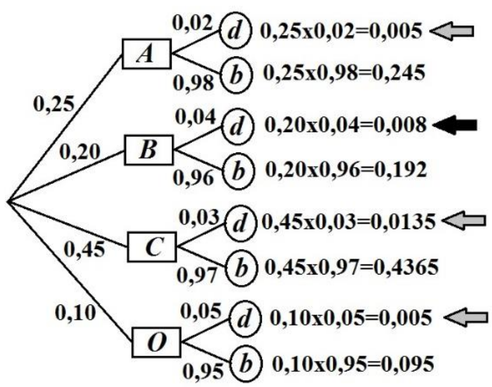
A situação dos exemplos 1.4.2[2] e 1.4.3[2] pode ser representada pela seguinte tabela, supondo uma produção de 10.000 peças (veja que pode ser utilizada qualquer quantidade de peças, basta obedecer aos percentuais informados).

	FABRICANTE A	FABRICANTE B	FABRICANTE C	OUTROS (O)	TOTAL
DEFEITUOSAS (d)	50	80	135	50	315
NÃO DEFEITUOSAS (b)	2.450	1.020	4.365	950	9.685
TOTAL	2.500	2.000	4.500	1.000	10.000

As probabilidades pedidas são:

$$P(d) = \frac{315}{10.000} = 0{,}0315 = 3{,}15\% \qquad P(B\,|\,d) = \frac{80}{315} = 0{,}254 = 25{,}4\%$$

Utilizando a árvore de possibilidades:



$P(d) = 0{,}005+0{,}008+0{,}0135+0{,}005 = 0{,}0315$

$P(B|d) = 0{,}008/0{,}0315 = 0{,}254$

Sugestão

Representе o caso das bolas (exemplos 1.4.2[1] e 1.4.3[1]) através de uma tabela e da árvore de possibilidades e verifique os resultados obtidos. Veja que, neste caso, as quantidades já estão definidas.

Atividades de avaliação

1) Sejam A e B eventos associados a um experimento E. Para as afirmações seguintes, dê o significado em teoria dos conjuntos: (a) ao menos A ou B ocorrem; (b) ambos A e B ocorrem; (c) nem A nem B ocorrem; (d) A ocorre e B não ocorre; (e) exatamente um de A ou B ocorre; (f) não mais que um de A ou B ocorre; (g) se A ocorre, B ocorre; (h) o evento A ou B; (i) o evento A e B; (j) A e B mutuamente exclusivos
2) Sejam A,B e C eventos associados a um experimento E. Para as afirmações seguintes, dê o significado em teoria dos conjuntos: (a) Se A ocorre então B não ocorre; (b) Nenhum dos resultados A, B e C ocorre; (c) Somente A ocorre; (d) Ao menos um de A,B,C ocorre; (e) Exatamente um deles ocorre; (f) Não mais que um ocorre; (g) Ao menos dois ocorrem; (h) Exatamente dois deles ocorrem; (i) Não mais que dois ocorrem; (j) A e C ocorrem e B não ocorre; (k) Todos ocorrem
3) Uma urna contém 3 bolas brancas, b_1, b_2 e b_3, e duas bolas pretas, p_4 e p_5 . Uma amostra de extensão 3, com reposição (sem reposição) é retirada da urna. Supondo S equiprovável, qual a probabilidade de: a) a 1ª bola ser branca b) a 2ª bola ser branca c) a 3ª bola ser branca
4) Uma urna contém M bolas, sendo M_w brancas e M-M_w pretas. Uma amostra de extensão n é retirada, com reposição (sem reposição), da urna. Supondo o espaço equiprovável, e o evento A_j = a j-ésima bola é branca, mostre que $P(A_j) = M_w/M$.
5) Uma urna contém 5 bolas , sendo 2 vermelhas e o restante amarelas. Desta urna retiramos 3 bolas, em extrações sucessivas sem reposição. Se A, B e C são eventos dados por A: exatamente 1 bola vermelha, B: pelo menos 1 bola amarela, C: pelo menos 1 bola amarela e 1 vermelha, determine $P(A)$, $P(B)$ e $P(C)$ supondo S equiprovável.
6) Uma urna contém 5 bolas, sendo 2 vermelhas e 3 amarelas. Desta urna são retiradas 4 bolas simultaneamente. Se A, B, C e D são os eventos A: 2 bolas vermelhas e 2 amarelas, B: 1 bola vermelha e 3 amarelas, C: 1 amarela e 3 vermelhas e D: pelo menos 1 vermelha e 1 amarela, determine $P(A)$, $P(B)$, $P(C)$ e $P(D)$ supondo S equiprovável
7) Uma urna contém 5 bolas, sendo 2 vermelhas e 3 amarelas. Desta urna são retiradas, com reposição, uma amostra não ordenada de extensão 4. Se A, B, C, D, E e F são os eventos A: 2 bolas vermelhas e 2 amarelas, B: 1 bola vermelha e 3 amarelas, C: 1 amarela e 3 vermelhas, D: pelo menos 1 vermelha e 1 amarela, E: 4 vermelhas e F: 4 amarelas, determine $P(A)$, $P(B)$, $P(C)$, $P(D)$, $P(E)$ e $P(F)$, supondo S equiprovável.

8) Uma urna contém 5 bolas, sendo 2 vermelhas e 3 amarelas. Desta urna são retiradas, com reposição, uma amostra ordenada de extensão 4. Se A, B, C, D, E e F são os eventos A: 2 bolas vermelhas e 2 amarelas, B: 1 bola vermelha e 3 amarelas, C: 1 amarela e 3 vermelhas, D: pelo menos 1 vermelha e 1 amarela, E: 4 vermelhas e F: 4 amarelas, determine $P(A)$, $P(B)$, $P(C)$, $P(D)$, $P(E)$ e $P(F)$, supondo S equiprovável.

9) Em uma sala, 10 pessoas estão usando emblemas numerados de 1 a 10. Três pessoas são escolhidas ao acaso e convidadas a sair da sala simultaneamente. Determine a probabilidade de que: a) o menor número de emblema seja 5? b) o maior número de emblema seja 5?

10) [MEYER] Dez fichas numeradas de 1 a 10 são misturadas em uma urna. Duas fichas (x_1,x_2) são extraídas da urna sucessivamente e sem reposição. Qual a probabilidade de que: (a) $x_1+x_2=10$ (b) $x_1 \leq x_2$ dado que $x_1+x_2=10$

11) [MEYER] Dentre 6 números positivos e 8 negativos, escolhem-se ao acaso 4 números (com reposição) e multiplicam-se esses números. Qual a probabilidade de que o produto seja positivo?

12) [MEYER] Um lote contém n peças das quais se sabe serem r defeituosas. Se a ordem de inspeção das peças se fizer ao acaso, qual a probabilidade de que a peça inspecionada em k-ésimo lugar ($k \geq r$) seja a última peça defeituosa contida no lote?

13) [MEYER] Uma urna contem N objetos. São retirados ao acaso, com reposição, n ($n \leq N$) objetos da urna. Qual a probabilidade de que nenhum objeto seja escolhido mais de uma vez?

14) Considere um conjunto com N objetos. Serão formados n grupos (N é múltiplo de n) com $k=N/n$ objetos cada um. Se, para a composição dos grupos, os objetos são selecionados aleatoriamente, qual a probabilidade de que r, $r \leq k$, objetos específicos pertençam a um mesmo grupo?

15) [MORGADO] Uma loteria é composta de N números e apenas um prêmio, ou seja, dos N números é sorteado apenas um, que corresponde ao bilhete premiado. Um apostador (A) compra n bilhetes em apenas uma extração enquanto que outro apostador (B) compra um único bilhete em n extrações diferentes. Qual dos dois tem maior probabilidade de ganhar ? (considere válida a seguinte desigualdade: $\left(1-\frac{1}{N}\right)^n \geq 1-\frac{n}{N}$).

16) [MORGADO] Um prisioneiro possui 50 bolas brancas, 50 bolas pretas e 2 urnas iguais. O prisioneiro deve colocar, como preferir, as bolas nas duas urnas (nenhuma das urnas pode ficar vazia). As urnas serão embaralhadas e o prisioneiro, de olhos vendados, deverá escolher uma delas e, em seguida, retirar uma bola da urna selecionada. Se a bola retirada for branca, ele será libertado. Como o prisioneiro deve proceder, na distribuição das bolas nas urnas, para maximizar a probabilidade de ser libertado?

17) [MEYER] A urna 1 contém x bolas brancas e y bolas vermelhas. A urna 2 contém z bolas brancas e v vermelhas. Uma bola é escolhida ao acaso da urna 1 e colocada na urna 2. A seguir, uma bola é escolhida ao acaso da urna 2. Qual a probabilidade de que essa bola seja branca?
18) [MEYER] Suponha que A e B sejam eventos independentes associados a um experimento. Se a probabilidade de A ou B ocorrerem for igual a $0{,}6$ enquanto que a probabilidade de ocorrência de A for $0{,}4$, determine a probabilidade de ocorrência de B.
19) [MEYER] Sejam A e B dois eventos associados a um experimento. Suponha que $P(A)$=0,4 e $P(A \cup B)$=0,7. Seja $P(B)$=p. Para que valores de p: a) A e B são mutuamente exclusivos? b) A e B são independentes?
20) [MEYER] Suponha que um escritório possua 100 máquinas de calcular. Algumas dessas máquinas são elétricas (E), enquanto outras são manuais (M) e algumas são novas (N) e outras usadas (U):

	E	M	TOTAL
N	40	30	70
U	20	10	30
TOTAL	60	40	100

Uma pessoa entra no escritório, pega uma máquina ao acaso e descobre que é nova. Qual a probabilidade de que seja elétrica? Qual a probabilidade dela ser manual?

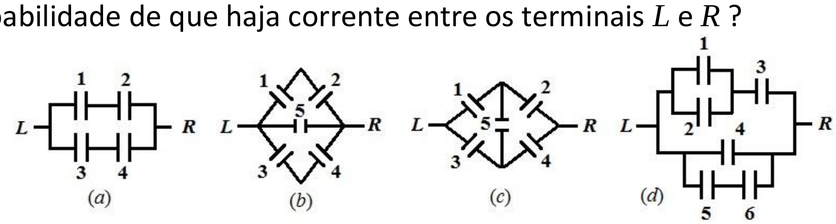
21) Numa certa faculdade, 25% dos estudantes foram reprovados em matemática, 15% reprovados em química, e 10% em matemática e em química. Um estudante é selecionado ao acaso.
    a) Se ele foi reprovado em química, qual a probabilidade de que tenha sido reprovado em matemática?
    b) Se ele foi reprovado em matemática, qual a probabilidade de que ele tenha sido reprovado em química?
    c) Qual a probabilidade de que ele tenha sido reprovado em matemática ou em química?
22) Numa certa faculdade, 4% dos homens e 1% das mulheres têm altura maior do que $1{,}80\text{m}$. Além disso, 60% dos estudantes são mulheres. Se um estudante é escolhido ao acaso e tem altura maior que $1{,}80\text{m}$, qual a probabilidade de que o estudante seja mulher?
23) Um homem possui duas moedas, uma comum (I) e outra de duas caras (II). Apanha uma moeda ao acaso e lança. Se cair cara, qual a probabilidade de que a moeda lançada tenha sido a de duas caras?

24) As máquinas A,B e C fabricam, respectivamente, 50%, 30% e 20% da produção total de certo artigo. A probabilidade de que a máquina A fabrique um artigo defeituoso é $0{,}03$, a B é $0{,}04$ e a C é $0{,}05$.
   a) qual a probabilidade de que um artigo tomado ao acaso seja defeituoso?
   b) Qual a probabilidade de que um artigo tomado ao acaso tenha sido fabricado pela máquina A, dado que é defeituoso?
   c) Qual a probabilidade de que um artigo tomado ao acaso tenha sido fabricado pela máquina B, dado que é defeituoso?
   d) Qual a probabilidade de que um artigo tomado ao acaso tenha sido fabricado pela máquina C, dado que é defeituoso?

25) Certa doença acha-se presente em cerca de 1 entre 1.000 pessoas. Um programa de provas deve ser executado mediante o emprego de um aparelho de investigação que dá uma leitura positiva com probabilidade de $0{,}99$ para uma pessoa doente e com probabilidade de $0{,}05$ para uma pessoa sadia. Qual a probabilidade de que uma pessoa escolhida ao acaso, cuja leitura é positiva, tenha efetivamente a doença ?

26) [MEYER] Seja o experimento E: lançar dois dados distinguíveis e observar as faces superiores. Sejam os eventos A: o primeiro dado mostra número par, B: o segundo dado mostra número ímpar e C: ambos mostram pares ou ambos mostram ímpares. Mostre que os eventos são 2 a 2 independentes mas não são mutuamente independentes.

27) [MEYER] A probabilidade de fechamento de cada relé nos circuitos abaixo é dada por p>0. Se todos os relés funcionam independentemente, qual é a probabilidade de que haja corrente entre os terminais L e R ?



28) [MEYER] Três componentes, C_1, C_2 e C_3 de um mecanismo são postos em série. Suponha que esses componentes sejam dispostos em ordem aleatória. Seja A o evento "C_2 está à direita de C_1" e B o evento "C_3 está à direita de C_1". São os eventos A e B mutuamente independentes ?

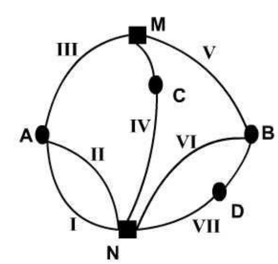
29) [XAVIER] Num certo grupo racial a probabilidade de uma pessoa do sexo masculino ser daltônica é p=$0{,}16$. São escolhidas ao acaso duas pessoas , x e y, do referido sexo e do referido grupo racial. Determine a probabilidade de ambos serem daltônicos sabendo-se que: a) x é daltônico b) Pelo menos 1 deles é daltônico

30) Um submarino dispara 5 torpedos em cadência rápida contra um navio. Cada torpedo tem probabilidade igual a $0{,}25$ de atingir o alvo. (a) Qual a probabilidade do navio receber pelo menos 1 impacto ? (b) exatamente 1 impacto ? (c) exatamente k impactos (k=$0{,}1{,}2{,}3{,}4{,}5$) ?

31) Para pessoas de $20,21$ e 22 anos, as probabilidades de morte durante um certo ano são iguais a p_1, p_2 e p_3 respectivamente. Se a,b e c são pessoas com $20,21$ e 22 anos respectivamente, pede-se para calcular:
   a) A probabilidade dos três sobreviverem nesse ano
   b) De todos terem morrido nesse ano
   c) De no máximo ter ocorrido uma morte nesse ano
   d) Pelo menos um deles ter morrido nesse ano
32) [MEYER] Uma montagem eletrônica é formada de dois subsistemas A e B. De procedimentos de ensaios anteriores, as seguintes probabilidades se admitem conhecidas: $P(A$ falhe)=0,20; $P(A$ e B falhem)=0,15 e $P(B$ falhe sozinho)=$0{,}15$. Calcule as seguintes probabilidades: (a) $P(A$ falhe dado que B tenha falhado) (b) $P(A$ falhe sozinho).
33) [MEYER] Duas máquinas, A e B, sendo operadas independentemente, podem ter alguns desarranjos cada dia. As probabilidades de desarranjo de cada máquina estão na tabela abaixo.

NÚMERO DE DESARRANJOS	0	1	2	3	4	5	6
MÁQUINA A	0,10	0,20	0,30	0,20	0,09	0,07	0,04
MÁQUINA B	0,30	0,10	0,10	0,10	0,10	0,15	0,15

Calcule as seguintes probabilidades: (a) A e B tenham o mesmo número de desarranjos (b) O número total de desarranjos seja menor que 4; menor que 5 (c) A tenha mais desarranjos que B (d) B tenha duas vezes mais desarranjos que A (e) B tenha 4 desarranjos, quando se saiba que B já tenha tido 2 desarranjos (*f*) O número mínimo de desarranjos das duas máquinas seja 3; seja menor que 3 (g) O número máximo de desarranjos das máquinas seja 3; seja maior que 3
34) [XAVIER] Um viajante deseja ir de uma cidade A a outra cidade B. As estradas existentes estão mostradas na figura abaixo. O viajante não deve passar mais de uma vez pelo mesmo ponto. Ao partir de A, o viajante pode escolher as estradas I, II e III com probabilidades $0{,}4$, $0{,}4$ e $0{,}2$ respectivamente. Chegando à encruzilhada M, vindo por III pode escolher as estradas IV ou V com probabilidades $0{,}3$ e $0{,}7$, respectivamente. Chegando à encruzilhada N, vindo de I ou II, pode escolher IV, VI ou VII com probabilidades $0{,}2$, $0{,}3$ e $0{,}5$, respectivamente. Porém, se ele chega a N por IV pode escolher VI ou VII comprobabilidades $0{,}4$ e $0{,}6$, respectivamente. (a) Se cada trajeto de A a B é um evento simples, descreva o espaço amostral correspondente. (b) Qual a probabilidade do viajante: (i) Passar por C? (ii) Passar por D? (iii) Passar por C e D? (iv) Passar por C ou D?



35) [XAVIER] Um emissor opera com um “alfabeto” de três símbolos, que representaremos por A_1, A_2 e A_3, sendo suas mensagens nada mais que sequencias desses símbolos. Ao ser emitido o símbolo A_i, a probabilidade de em seguida ser emitido o símbolo A_j é dada por $p_{i,j}$, essas probabilidades são chamadas de probabilidade de transição de A_i para A_j e a matriz $(p_{i,j})$ é a respectiva matriz de transição.

Se tivermos $\begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} = \begin{pmatrix} 0 & 1/3 & 2/3 \\ 1/3 & 1/3 & 1/3 \\ 2/9 & 0 & 7/9 \end{pmatrix}$ e houver probabilidades p_1=5/21, p_2=6/21 e p_3=10/21 de uma mensagem começar pelos símbolos A_1, A_2 e A_3, respectivamente, qual a probabilidade da mensagem começar: (a) por uma das sequencias A_1A_2 ou A_2A_3? (b) por uma sequencia da forma $A_iA_iA_i$?

Solução das atividades

1)

$a)\ x \in A \cup B \qquad b)\ x \in A \cap B \qquad c)\ x \in \overline{A} \cap \overline{B} \qquad d)\ x \in A \cap \overline{B}$

$e)\ x \in \left(A \cap \overline{B}\right) \cup \left(\overline{A} \cap B\right) \qquad f)\ x \in \overline{(A \cap B)} \qquad g)\ x \in A \Rightarrow x \in B \Rightarrow A \subseteq B$

$h)\ A \cup B \quad i)\ A \cap B \quad j)\ A \cap B = \phi$

2)

$a)\ x \in A \Rightarrow x \in \overline{B} \Rightarrow A \subseteq \overline{B} \qquad b)\ x \in \overline{A} \cap \overline{B} \cap \overline{C} \qquad c)\ x \in A \cap \overline{B} \cap \overline{C}$

$d)\ x \in A \cup B \cup C$

$e)\ x \in \left(A \cap \overline{B} \cap \overline{C}\right) \cup \left(B \cap \overline{A} \cap \overline{C}\right) \cup \left(C \cap \overline{A} \cap \overline{B}\right)$

$f)\ x \in \left(A \cap \overline{B} \cap \overline{C}\right) \cup \left(B \cap \overline{A} \cap \overline{C}\right) \cup \left(C \cap \overline{A} \cap \overline{B}\right) \cup \left(\overline{A} \cap \overline{B} \cap \overline{C}\right)$

$g)\ x \in \left(A \cap B \cap \overline{C}\right) \cup \left(A \cap \overline{B} \cap C\right) \cup \left(\overline{A} \cap B \cap C\right) \cup \left(A \cap B \cap C\right)$

$h)\ x \in \left(A \cap B \cap \overline{C}\right) \cup \left(A \cap C \cap \overline{B}\right) \cup \left(B \cap C \cap \overline{A}\right)$

$i)\ x \in \overline{(A \cap B \cap C)} \qquad j)\ x \in \left(A \cap C \cap \overline{B}\right) \qquad k)\ x \in \left(A \cap B \cap C\right)$

3)

COM REPOSIÇÃO:

$n(S) = A^{*}_{5,3} = 5^3$

$a)\ A \equiv b** \Rightarrow n(A) = A^{*}_{3,1} \times A^{*}_{5,2} \Rightarrow P(A) = \dfrac{3 \times 5^2}{5^3} = \dfrac{3}{5}$

$b)\ A \equiv *b* \Rightarrow \text{análogo ao item } (a) \Rightarrow P(A) = \dfrac{3}{5}$

$c)\ A \equiv **b \Rightarrow \text{análogo aos itens } (a)\ e\ (b) \Rightarrow P(A) = \dfrac{3}{5}$

SEM REPOSIÇÃO: Basta utilizar ARRANJO SIMPLES, os resultados serão os mesmos anteriores, todos iguais a $\dfrac{3}{5}$

4)

É uma generalização do exercício 3.

COM REPOSIÇÃO:

$$n(S) = A^*_{M,n}$$

$$n(A) = A^*_{M_w,1} \times A^*_{M,n-1} \Rightarrow P(A) = \frac{A^*_{M_w,1} \times A^*_{M,n-1}}{A^*_{M,n}} = \frac{M_w}{M}$$

SEM REPOSIÇÃO:

$$n(S) = A_{M,n}$$

$$n(A) = A_{M_w,1} \times A_{M-1,n-1} \Rightarrow P(A) = \frac{A_{M_w,1} \times A_{M-1,n-1}}{A_{M,n}} = \frac{M_w}{M}$$

OBS: faça o desenvolvimento dos arranjos e as simplificações necessárias para chegar ao resultado.

5)

População: 2 bolas vermelhas e 3 amarelas$\Rightarrow N$=5.

Amostra (<u>ordenada sem repetição</u>): $n = 3 \Rightarrow n(S) = A_{5,3} = \frac{5!}{2!} = 60$

$$A \equiv vaa \Rightarrow n(A) = A_{2,1} \times A_{3,2} \times P_3^{1,2} = \frac{2!}{1!} \times \frac{3!}{1!} \times \frac{3!}{1!2!} = 36 \Rightarrow P(A) = \frac{36}{60}$$

$$B \equiv avv \text{ ou } aav \text{ ou } aaa \Rightarrow n(B) = A_{3,1} \times A_{2,2} \times P_3^{1,2} + A_{3,2} \times A_{2,1} \times P_3^{2,1} + A_{3,3} =$$

$$= \frac{3!}{2!} \times \frac{2!}{0!} \times \frac{3!}{1!2!} + \frac{3!}{1!} \times \frac{2!}{1!} \times \frac{3!}{2!1!} + \frac{3!}{0!} = 18 + 36 + 6 = 60 \Rightarrow P(B) = \frac{60}{60} = 1$$

$$C \equiv avv \text{ ou } aav \Rightarrow n(C) = A_{3,1} \times A_{2,2} \times P_3^{1,2} + A_{3,2} \times A_{2,1} \times P_3^{2,1} =$$

$$= \frac{3!}{2!} \times \frac{2!}{0!} \times \frac{3!}{1!2!} + \frac{3!}{1!} \times \frac{2!}{1!} \times \frac{3!}{2!1!} = 18 + 36 = 54 \Rightarrow P(C) = \frac{54}{60}$$

6)

População: 2 bolas vermelhas e 3 amarelas$\Rightarrow N$=5

Amostra (<u>não ordenada sem repetição</u>): $n = 4 \Rightarrow n(S) = C_{5,4} = \frac{5!}{4!1!} = 5$

$$A \equiv vvaa \Rightarrow n(A) = C_{2,2} \times C_{3,2} = \frac{2!}{2!0!} \times \frac{3!}{2!1!} = 1 \times 3 = 3 \Rightarrow P(A) = \frac{3}{5}$$

$$B \equiv vaaa \Rightarrow n(B) = C_{2,1} \times C_{3,3} = \frac{2!}{1!1!} \times \frac{3!}{3!0!} = 2 \times 1 = 2 \Rightarrow P(B) = \frac{2}{5}$$

$$C \equiv avvv \Rightarrow C = \phi \Rightarrow P(C) = 0$$

$$D \equiv avvv \text{ ou } aavv \text{ ou } aaav \Rightarrow n(D) = 0 + C_{3,2} \times C_{2,2} + C_{3,3} \times C_{2,1} =$$

$$= \frac{3!}{2!1!} \times \frac{2!}{2!0!} + \frac{3!}{3!0!} \times \frac{2!}{1!1!} = 3 \times 1 + 1 \times 2 = 3 + 2 = 5 \Rightarrow P(D) = \frac{5}{5} = 1$$

7)

População: 2 bolas vermelhas e 3 amarelas$\Rightarrow N$=5.

Amostra (não ordenada com repetição): $n = 4 \Rightarrow n(S) = C^*_{5,4} = C_{8,4} = \frac{8!}{4!4!} = 70$

$$A \equiv vvaa \Rightarrow n(A) = C^*_{2,2} \times C^*_{3,2} = C_{3,2} \times C_{4,2} = \frac{3!}{2!1!} \times \frac{4!}{2!2!} = 3 \times 6 = 18 \Rightarrow P(A) = \frac{18}{70}$$

$$B \equiv vaaa \Rightarrow n(B) = C^*_{2,1} \times C^*_{3,3} = C_{2,1} \times C_{5,3} = \frac{2!}{1!1!} \times \frac{5!}{3!2!} = 2 \times 10 = 20 \Rightarrow P(B) = \frac{20}{70}$$

$$C \equiv avvv \Rightarrow n(C) = C^*_{3,1} \times C^*_{2,3} = C_{3,1} \times C_{4,3} = \frac{3!}{1!2!} \times \frac{4!}{3!1!} = 3 \times 4 = 12 \Rightarrow P(C) = \frac{12}{70}$$

$$D \equiv avvv \text{ ou } aavv \text{ ou } aaav \Rightarrow n(D) = C^*_{3,1} \times C^*_{2,3} + C^*_{3,2} \times C^*_{2,2} + C^*_{3,3} \times C^*_{2,1} =$$

$$= C_{3,1} \times C_{4,3} + C_{4,2} \times C_{3,2} + C_{5,3} \times C_{2,1} = \frac{3!}{1!2!} \times \frac{4!}{3!1!} + \frac{4!}{2!2!} \times \frac{3!}{2!1!} + \frac{5!}{3!2!} \times \frac{2!}{1!1!} = 3 \times 4 + 6 \times 3 + 10 \times 2 =$$

$$= 12 + 18 + 20 = 50 \Rightarrow P(D) = \frac{50}{70}$$

$$E \equiv vvvv \Rightarrow n(E) = C^*_{2,4} = C_{5,4} = \frac{5!}{4!1!} = 5 \Rightarrow P(E) = \frac{5}{70}$$

$$F \equiv aaaa \Rightarrow n(F) = C^*_{3,4} = C_{6,4} = \frac{6!}{4!2!} = 15 \Rightarrow P(E) = \frac{15}{70}$$

8)

População: 2 bolas vermelhas e 3 amarelas $\Rightarrow N$=5.

Amostra (ordenada com repetição): $n = 4 \Rightarrow n(S) = A^*_{5,4} = 5^4 = 625$

$$A \equiv vvaa \Rightarrow n(A) = A^*_{2,2} \times A^*_{3,2} \times P_4^{2,2} = 2^2 \times 3^2 \times 6 = 4 \times 9 \times 6 = 216 \Rightarrow P(A) = \frac{216}{625}$$

$$B \equiv vaaa \Rightarrow n(B) = A^*_{2,1} \times A^*_{3,3} \times P_4^{1,3} = 2^1 \times 3^3 \times 4 = 2 \times 27 \times 4 = 216 \Rightarrow P(B) = \frac{216}{625}$$

$$C \equiv avvv \Rightarrow n(C) = A^*_{3,1} \times A^*_{2,3} \times P_4^{1,3} = 3^1 \times 2^3 \times 4 = 3 \times 8 \times 4 = 96 \Rightarrow P(C) = \frac{96}{625}$$

$$D \equiv avvv \text{ ou } aavv \text{ ou } aaav \Rightarrow n(D) = A^*_{3,1} \times A^*_{2,3} \times P_4^{1,3} + A^*_{3,2} \times A^*_{2,2} \times P_4^{2,2} +$$

$$+ A^*_{3,3} \times A^*_{2,1} \times P_4^{3,1} == 3^1 \times 2^3 \times 4 + 3^2 \times 2^2 \times 6 + 3^3 \times 2^1 \times 4 = 96 + 216 + 216 = 528 \Rightarrow$$

$$\Rightarrow P(D) = \frac{528}{625}$$

$$E \equiv vvvv \Rightarrow n(E) = A^*_{2,4} = 2^4 = 16 \Rightarrow P(E) = \frac{16}{625}$$

$$F \equiv aaaa \Rightarrow n(F) = A^*_{3,4} = 3^4 = 81 \Rightarrow P(E) = \frac{81}{625}$$

9)

População: N=10

Amostra (não ordenada sem reposição): $n = 3 \Rightarrow n(S) = C_{10,3} = \frac{10!}{3!7!} = 120$

$$A \equiv \underline{[5]}\underline{[>5]}\underline{[>5]} \Rightarrow n(A) = 1 \times C_{5,2} = \frac{5!}{2!3!} = 10 \Rightarrow P(A) = \frac{10}{120}$$

$$B \equiv \underline{[5]}\underline{[<5]}\underline{[<5]} \Rightarrow n(B) = 1 \times C_{4,2} = \frac{4!}{2!2!} = 6 \Rightarrow P(B) = \frac{6}{120}$$

10)

População: N=10

Amostra (ordenada sem reposição): $n = 2 \Rightarrow n(S) = A_{10,2} = \frac{10!}{8!} = 90$

$a)$ $A = \{(x_1, x_2) | x_1 + x_2 = 10\} = \{(1{,}9); (9{,}1); (2{,}8); (8{,}2); (3{,}7); (7{,}3); (4{,}6); (6{,}4)\}$

$$\Rightarrow n(A) = 8 \Rightarrow P(A) = \frac{8}{90}$$

$b)$ $A = \{(x_1, x_2) | x_1 + x_2 = 10\} = \{(1{,}9); (9{,}1); (2{,}8); (8{,}2); (3{,}7); (7{,}3); (4{,}6); (6{,}4)\}$

$$\Rightarrow n(A) = 8 \Rightarrow P(A) = \frac{8}{90}$$

$B = \{(x_1, x_2) | x_1 \le x_2\}$ $\quad A \cap B = \{(x_1, x_2) | x_1 \le x_2 \wedge\wedge x_1 + x_2 = 10\} =$

$= \{(1{,}9); (2{,}8); (3{,}7); (4{,}6)\}$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{4}{90}}{\frac{8}{90}} = \frac{4}{8}$$

11)

População: N=14: 6 positivos (+) e 8 negativos (-)
Amostra <u>(ordenada com reposição)</u>: $n = 4 \Rightarrow n(S) = A^{*}_{14,4} = 14^4 = 38416$

$$A \equiv (+)(+)(+)(+) \text{ ou } (-)(-)(-)(-) \text{ ou } (+)(+)(-)(-)$$

$$\Rightarrow P(A) = \frac{6}{14}\times\frac{6}{14}\times\frac{6}{14}\times\frac{6}{14}+\frac{8}{14}\times\frac{8}{14}\times\frac{8}{14}\times\frac{8}{14}+$$

$$+\frac{6}{14}\times\frac{6}{14}\times\frac{8}{14}\times\frac{8}{14}\times P_4^{2,2} = \frac{6^4}{14^4}+\frac{8^4}{14^4}+\frac{6^2\times 8^2\times 6}{14^4} = \frac{19216}{38416} = \frac{1201}{2401}$$

12)

$$n(S) = A_{n,k} = \frac{n!}{(n-k)!}$$

$$A \equiv \underbrace{\underbrace{BBB\ldots B}_{k-r}\underbrace{DDD\ldots D}_{r-1}}_{k-1}D \Rightarrow n(A) = A_{n-r,k-r}\times A_{r,r-1}\times P_{k-1}^{k-r,r-1}$$

$$n(A) = \frac{(n-r)!}{[(n-r)-(k-r)]!}\times\frac{r!}{[r-(r-1)]!}\times\frac{(k-1)!}{(k-r)!(r-1)!} = \frac{r!(n-r)!(k-1)!}{(n-k)!(k-r)!(r-1)!}$$

$$P(A) = \frac{\dfrac{r!(n-r)!(k-1)!}{(n-k)!(k-r)!(r-1)!}}{\dfrac{n!}{(n-k)!}} = \frac{r!(n-r)!(k-1)!}{(n-k)!(k-r)!(r-1)!}\times\frac{(n-k)!}{n!} =$$

$$= \frac{r!(n-r)!}{n!}\times\frac{(k-1)!}{(r-1)![(k-1)-(r-1)]!}$$

$$P(A) = \frac{1}{\binom{n}{r}}\times\binom{k-1}{r-1} \Rightarrow P(A) = \frac{\binom{k-1}{r-1}}{\binom{n}{r}}$$

13)

$$n(S) = A^{*}_{N,n} = N^n$$

$$n(A) = A_{N,n} = \frac{N!}{(N-n)!}$$

$$P(A) = \frac{\dfrac{N!}{(N-n)!}}{N^n} = \frac{(N-1)!}{(N-n)!\,N^{n-1}}$$

14)

$$n(S)=\frac{\binom{N}{k}\binom{N-k}{k}\binom{N-2k}{k}\cdots\binom{N-(n-2)k}{k}\binom{k}{k}}{n!}=\frac{\binom{N}{k}\binom{N-k}{k}\binom{N-2k}{k}\cdots\binom{2k}{k}}{n!}$$

$$n(A)=\frac{\binom{N-r}{k-r}\binom{(N-r)-(k-r)}{k}\binom{(N-r)-[(k-r)+k]}{k}\cdots\binom{(N-r)-[(k-r)+(n-3)k]}{k}\binom{k}{k}}{(n-1)!}=$$

$$=\frac{\binom{N-r}{k-r}\binom{N-k}{k}\binom{N-2k}{k}\cdots\binom{2k}{k}}{(n-1)!}$$

$$P(A)=\frac{\binom{N-r}{k-r}\binom{N-k}{k}\binom{N-2k}{k}\cdots\binom{2k}{k}}{(n-1)!}\times\frac{n!}{\binom{N}{k}\binom{N-k}{k}\binom{N-2k}{k}\cdots\binom{2k}{k}}=\frac{n\times\binom{N-r}{k-r}}{\binom{N}{k}}$$

15)

Jogando n bilhetes em uma única extração tem-se:

$$n(S)=N$$
$$A\equiv A\,\text{ganhar}\Rightarrow n(A)=n$$
$$P(A)=\frac{n}{N}$$

Jogando n bilhetes em n extrações tem-se:

$$n(S)=N^n$$
$$B\equiv B\,\text{ganhar}\qquad \overline{B}\equiv B\,\text{não ganhar}\Rightarrow n(\overline{B})=(N-1)^n$$
$$P(\overline{B})=\frac{(N-1)^n}{N^n}=\left(\frac{N-1}{N}\right)^n=\left(1-\frac{1}{N}\right)^n\qquad P(B)=1-P(\overline{B})=1-\left(1-\frac{1}{N}\right)^n$$

Comparando as duas probabilidades:

Como $1-\frac{n}{N}\le\left(1-\frac{1}{N}\right)^n \Leftrightarrow \frac{n}{N}\ge 1-\left(1-\frac{1}{N}\right)^n$ tem-se que $P(A)\ge P(B)$

16)

Considere que o prisioneiro coloque k bolas, das quais a brancas, em uma urna e 100-k, das quais 50-a brancas, na outra. Sejam os eventos:

U_1≡o prisioneiro seleciona a urna com k bolas (a brancas)
U_2≡o prisioneiro escolhe a urna com 100-k bolas (50-a brancas)
B≡o prisioneiro seleciona uma bola branca
L≡prisioneiro é libertado

Então:

$$P(L) = P(U_1 \cap B) + P(U_2 \cap B)$$
$$P(L) = P(U_1)P(B|U_1) + P(U_2)P(B|U_2)$$
$$P(L) = \frac{1}{2} \times \frac{a}{k} + \frac{1}{2} \times \frac{50-k}{100-a}$$
$$P(L) = \frac{1}{2} \times \frac{50k + a(100-2k)}{k(100-k)}$$

Veja que $k \leq 50$, caso contrário $P(L)<0$.
Para $k=50$ temos que $P(L) = \frac{1}{2}$, independentemente do valor de a.
Para $k<50$ (U_1 com a menor quantidade de bolas), temos:
Para k fixo → $P(L)$ cresce se a cresce →$P(L)$ é máxima em $a=k$, pois $a \leq k$

$a = k \Rightarrow P(L) = \frac{1}{2} \times \frac{50k + k(100-2k)}{k(100-k)} = \frac{75-k}{100-k} = 1 - \frac{25}{100-k}$ que é máxima para k mínimo, ou seja, para $k=1$

Então, como para $P(L)$ ser máxima deve-se ter $a=k$ e $k=1$, conclui-se que a urna U_1 deve ter uma única bola e essa bola deve ser branca, ou seja, $k=a=1$, e a probabilidade máxima é:

$$P(L) = 1 - \frac{25}{99} = \frac{74}{99} \approx 0{,}75$$

17)

B_I≡ ”retirar bola branca da urna I”
V_I≡ ”retirar bola vermelha da urna I”
B_{II}≡ ”retirar bola branca da urna II”

$$B_{II} = (B_I \cap B_{II}) \cup (V_I \cap B_{II}) \Rightarrow P(B_{II}) = P(B_I \cap B_{II}) + P(V_I \cap B_{II}) =$$

$$= P(B_I)P(B_{II}|B_I) + P(V_I)P(B_{II}|V_I) = \frac{x}{x+y}\cdot\frac{z+1}{z+v+1} + \frac{y}{x+y}\cdot\frac{z}{z+v+1}$$

18)

$$P(A \cup B) = 0{,}6$$
$$P(A) = 0{,}4$$
$$A\ e\ B\ MI \Rightarrow P(A \cap B) = P(A)P(B)$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A)P(B) \Rightarrow$$
$$\Rightarrow 0{,}6 = 0{,}4 + P(B) - 0{,}4P(B) \Rightarrow 0{,}6 = 0{,}4 + 0{,}6P(B) \Rightarrow P(B) = \frac{0{,}2}{0{,}6} = \frac{1}{3}$$

19)

$$P(A) = 0{,}4 \quad P(A \cup B) = 0{,}7 \quad P(B) = p$$

$a)$ $A\ e\ B\ ME \Rightarrow A \cap B = \phi \Rightarrow P(A \cap B) = 0$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) \Rightarrow 0{,}7 = 0{,}4 + p \Rightarrow p = 0{,}3$$

$b)$ $A\ e\ B\ MI \Rightarrow P(A \cap B) = P(A)P(B)$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A)P(B)$$
$$\Rightarrow 0{,}7 = 0{,}4 + p - 0{,}4p \Rightarrow p = 0{,}5$$

20)

$a)$ $P(E|N) = \dfrac{P(E \cap N)}{P(N)} = \dfrac{40/100}{70/100} = \dfrac{40}{70} = \dfrac{4}{7}$

$b)$ $P(M|N) = \dfrac{P(M \cap N)}{P(N)} = \dfrac{30/100}{70/100} = \dfrac{30}{70} = \dfrac{3}{7}$

21)

$M \equiv$ reprovado em Matemática $\Rightarrow P(M) = 0{,}25$
$Q \equiv$ reprovado em Química $\Rightarrow P(Q) = 0{,}15$
$M \cap Q \equiv$ reprovado em Matemática e Química $\Rightarrow P(M \cap Q) = 0{,}10$

$$a)\ P(M|Q) = \frac{P(M \cap Q)}{P(Q)} = \frac{0{,}10}{0{,}15} = \frac{10}{15} = \frac{2}{3}$$

$$b)\ P(Q|M) = \frac{P(M \cap Q)}{P(M)} = \frac{0{,}10}{0{,}25} = \frac{10}{25} = \frac{2}{5}$$

$$c)\ P(M \cup Q) = P(M) + P(Q) - P(M \cap Q) = 0{,}25 + 0{,}15 - 0{,}10 = 0{,}30$$

22)

$M \equiv$ Mulher $\Rightarrow P(M) = 0{,}60$
$H \equiv$ Homem $\Rightarrow P(H) = 0{,}40$
$A \equiv$ altura maior que 1,80 $\Rightarrow P(A|M) = 0{,}01 \ \ P(A|H) = 0{,}04$

$$P(M|A) = \frac{P(M)P(A|M)}{P(M)P(A|M) + P(H)P(A|H)} = \frac{0{,}60 \times 0{,}01}{0{,}60 \times 0{,}01 + 0{,}40 \times 0{,}04}$$

$$= \frac{0{,}006}{0{,}006 + 0{,}016} = \frac{0{,}006}{0{,}022} = \frac{6}{22} = \frac{3}{11}$$

23)

$I \equiv$ moeda comum $\Rightarrow P(I)$=0,5
$II \equiv$ moeda com duas caras $\Rightarrow P(II)$=0,5
$C \equiv$ cara $\Rightarrow P(C|I) = 0{,}5 \ \ P(C|II) = 1$

$$P(II|C) = \frac{P(II)P(C|II)}{P(I)P(C|I) + P(II)P(C|II)} = \frac{0{,}5 \times 1}{0{,}5 \times 0{,}5 + 0{,}5 \times 1} = \frac{0{,}5}{0{,}75} = \frac{50}{75} = \frac{2}{3}$$

24)

$A \equiv$ peça fabricada pela máquina $A \Rightarrow P(A) = 0{,}50$
$B \equiv$ peça fabricada pela máquina $B \Rightarrow P(B) = 0{,}30$
$C \equiv$ peça fabricada pela máquina $C \Rightarrow P(C) = 0{,}20$
$D \equiv$ peça defeituosa $\Rightarrow P(D|A) = 0{,}03 \ \ P(D|B) = 0{,}04 \ \ P(D|C) = 0{,}05$

$$a)\ P(D) = P(A)P(D|A) + P(B)P(D|B) + P(C)P(D|C) = 0{,}50 \times 0{,}03 + 0{,}30 \times 0{,}04 + 0{,}20 \times 0{,}05$$
$$= 0{,}015 + 0{,}012 + 0{,}010 = 0{,}037$$

$$b)\ P(A|D) = \frac{P(A)P(D|A)}{P(A)P(D|A) + P(B)P(D|B) + P(C)P(D|C)} = \frac{0{,}015}{0{,}037} = \frac{15}{37} \approx 0{,}405$$

$$c)\ P(B|D) = \frac{P(B)P(D|B)}{P(A)P(D|A) + P(B)P(D|B) + P(C)P(D|C)} = \frac{0{,}012}{0{,}037} = \frac{12}{37} \approx 0{,}324$$

$$d)\ P(C|D) = \frac{P(C)P(D|C)}{P(A)P(D|A) + P(B)P(D|B) + P(C)P(D|C)} = \frac{0{,}010}{0{,}037} = \frac{10}{37} \approx 0{,}270$$

25)

$D \equiv$ pessoa doente $\Rightarrow P(D) = 0{,}001$
$S \equiv$ pessoa sadia $\Rightarrow P(S) = 0{,}999$
$P \equiv$ teste positivo $\Rightarrow$ $P(P|D) = 0{,}99$ $P(P|S) = 0{,}05$

$$P(D|P) = \frac{P(D)P(P|D)}{P(D)P(P|D) + P(S)P(P|S)} = \frac{0{,}001 \times 0{,}99}{0{,}001 \times 0{,}99 + 0{,}999 \times 0{,}05}$$
$$= \frac{0{,}00099}{0{,}00099 + 0{,}04995} = \frac{0{,}0099}{0{,}05094} \approx 0{,}0194$$

26)

$$n(S) = A^{*}_{6,2} = 6^2 = 36$$
$$A = \{(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),(6,1),$$
$$(6,2),(6,3),(6,4),(6,5),(6,6)\}$$
$$\Rightarrow n(A) = 18 \Rightarrow P(A) = \frac{18}{36} = \frac{1}{2}$$
$$B = \{(1,1),(2,1),(3,1),(4,1),(5,1),(6,1),(1,3),(2,3),(3,3),(4,3),(5,3),(6,3),(1,5),(2,5),$$
$$(3,5),(4,5),(5,5),(6,5)\}$$
$$\Rightarrow n(B) = 18 \Rightarrow P(B) = \frac{18}{36} = \frac{1}{2}$$
$$C = \{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6),(1,1),(1,3),(1,5),(3,1),$$
$$(3,3),(3,5),(5,1),(5,3),(5,5)\}$$
$$\Rightarrow n(C) = 18 \Rightarrow P(C) = \frac{18}{36} = \frac{1}{2}$$
$$A \cap B = \{(2,1),(4,1),(6,1),(2,3),(4,3),(6,3),(2,5),(4,5)(6,5)\} \Rightarrow n(A \cap B) = 9 \Rightarrow$$
$$\Rightarrow P(B) = \frac{9}{36} = \frac{1}{4}$$
$$A \cap C = \{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\} \Rightarrow n(A \cap C) = 9 \Rightarrow$$
$$\Rightarrow P(C) = \frac{9}{36} = \frac{1}{4}$$
$$B \cap C = \{(1,1),(1,3),(1,5),(3,1),(3,3),(3,5),(5,1),(5,3),(5,5)\} \Rightarrow n(B \cap C) = 9 \Rightarrow$$
$$\Rightarrow P(C) = \frac{9}{36} = \frac{1}{4}$$
$$A \cap B \cap C = \phi \Rightarrow P(A \cap B \cap C) = 0$$

$$\left.\begin{array}{l} P(A\cap B)=\dfrac{1}{4}=\dfrac{1}{2}\cdot\dfrac{1}{2}=P(A)P(B)\Rightarrow A,B\text{ são }MI \\ P(A\cap C)=\dfrac{1}{4}=\dfrac{1}{2}\cdot\dfrac{1}{2}=P(A)P(C)\Rightarrow A,C\text{ são }MI \\ P(B\cap C)=\dfrac{1}{4}=\dfrac{1}{2}\cdot\dfrac{1}{2}=P(B)P(C)\Rightarrow B,C\text{ são }MI \end{array}\right\}\Rightarrow A,B,C\text{ são 2 a 2 independentes}$$

$$P(A\cap B\cap C)=0\neq P(A)P(B)P(C)=\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{8}\Rightarrow A,B,C\text{ não são }MI$$

27)

$R_i\equiv$ o relé i está fechado $\Rightarrow P(R_i)=p$

$a)$ passa corrente se R_1 e R_2 estiverem fechados ou se R_3 e R_4 estiverem fechados:

$$P(\text{passar corrente})=P[(R_1\cap R_2)\cup(R_3\cap R_4)]$$
$$=P(R_1\cap R_2)+P(R_3\cap R_4)-P[(R_1\cap R_2)\cap(R_3\cap R_4)]$$
$$P(R_1)P(R_2)+P(R_3)P(R_4)-P(R_1)P(R_2)P(R_3)P(R_4)=p^2+p^2-p^4$$
$$=2p^2-p^4=p^2(2-p^2)$$

$b)$ $P[(R_1\cap R_2)\cup R_5\cup(R_3\cap R_4)]=\ldots$

$c)$ $P[(R_1\cap R_2)\cup(R_1\cap R_5\cap R_4)\cup(R_3\cap R_4)\cup(R_3\cap R_5\cap R_2)]=\ldots$

$d)$ $P[(R_1\cap R_3)\cup(R_2\cap R_3)\cup R_4\cup(R_5\cap R_6)]=\ldots$

28)

$$S=\{C_1C_2C_3,C_1C_3C_2,C_2C_1C_3,C_2C_3C_1,C_3C_1C_2,C_3C_2C_1\}\Rightarrow n(S)=6$$
$$A=\{C_1C_2C_3,C_1C_3C_2,C_3C_1C_2\}\Rightarrow n(A)=3\Rightarrow P(A)=\frac{3}{6}=\frac{1}{2}$$
$$B=\{C_1C_2C_3,C_1C_3C_2,C_2C_1C_3\}\Rightarrow n(B)=3\Rightarrow P(B)=\frac{3}{6}=\frac{1}{2}$$
$$A\cap B=\{C_1C_2C_3,C_1C_3C_2\}\Rightarrow n(A\cap B)=2\Rightarrow$$
$$\Rightarrow P(A\cap B)=\frac{2}{6}=\frac{1}{3}\neq P(A)P(B)\Rightarrow A,B\text{ não são }MI$$

29)

$X\equiv x$ é daltônico $\quad\quad$ $Y\equiv y$ é daltônico
$X\cap Y\equiv$ ambos são daltônicos $\quad$ $X\cup Y\equiv$ pelo menos um é daltônico

$P(X) = p$
$P(Y) = p$
$P(X \cap Y) = P(X)P(Y) = p^2$
$P(X \cup Y) = P(X) + P(Y) - P(X \cap Y) = p + p - p^2 = 2 - p^2 = p(2-p)$

$a)\ P(X \cap Y | X) = \dfrac{P(X \cap Y \cap X)}{P(X)} = \dfrac{P(X \cap Y)}{P(X)} = \dfrac{p^2}{p} = p = 0{,}16$

$b)\ P(X \cap Y | X \cup Y) = \dfrac{P[(X \cap Y) \cap (X \cup Y)]}{P(X \cup Y)} = \dfrac{P(X \cap Y)}{P(X \cup Y)} = \dfrac{p^2}{p(2-p)} = \dfrac{p}{2-p} = \dfrac{0{,}16}{1{,}84} \approx 0{,}087$

30)

$A \equiv \text{acertar} \Rightarrow P(A) = 0{,}25$
$E \equiv \text{errar} \Rightarrow P(E) = 1 - P(A) = 1 - 0{,}25 = 0{,}75$

$a)\ P(\text{pelo menos um impacto}) = 1 - P(\text{nenhum impacto})$
$P(\text{nenhum impacto}) = P(EEEEE) = 0{,}75 \times 0{,}75 \times 0{,}75 \times 0{,}75 \times 0{,}75 = 0{,}75^5$
$P(\text{pelo menos um impacto}) = 1 - 0{,}75^5$

$b)\ P(\text{exatamente um impacto}) = P(AEEEE \text{ ou } EAEEE \text{ ou } EEAEE \text{ ou } EEEAE \text{ ou } EEEEA) =$
$= P(AEEEE) + P(EAEEE) + P(EEAEE) + P(EEAE) + P(EEEA) = 5 \times 0{,}25 \times 0{,}75^4$

$c)\ \text{exatamente } k \text{ impactos} \equiv \underbrace{\underbrace{AA\ldots A}_{k}\underbrace{EE\ldots E}_{5-k}}_{5} \Rightarrow P(\text{exatamente } k \text{ impactos}) =$

$= 0{,}25^k \times 0{,}75^{5-k} \times P_5^{k,5-k} \Rightarrow P(\text{exatamente } k \text{ impactos}) =$
$= \begin{pmatrix} 5 \\ k \end{pmatrix} \times 0{,}25^k \times 0{,}75^{5-k},\ k = 0{,}1{,}2{,}3{,}4{,}5$

31)

$A \equiv a$ morre $\Rightarrow P(A) = p_1$
$B \equiv b$ morre $\Rightarrow P(B) = p_2$
$C \equiv c$ morre $\Rightarrow P(C) = p_3$

$a)\ P(\bar{A} \cap \bar{B} \cap \bar{C}) = P(\bar{A})P(\bar{B})P(\bar{C}) = (1-p_1)(1-p_2)(1-p_3)$
$b)\ P(A \cap B \cap C) = P(A)P(B)P(C) = p_1 p_2 p_3$
$c)\ P(\bar{A} \cap \bar{B} \cap \bar{C}) + P(A \cap \bar{B} \cap \bar{C}) + P(\bar{A} \cap B \cap \bar{C}) + P(\bar{A} \cap \bar{B} \cap C) =$
$= (1-p_1)(1-p_2)(1-p_3) + p_1(1-p_2)(1-p_3) + (1-p_1)p_2(1-p_3) + (1-p_1)(1-p_2)p_3$
$d)\ P(A \cup B \cup C) = 1 - P(\bar{A} \cap \bar{B} \cap \bar{C}) = 1 - (1-p_1)(1-p_2)(1-p_3)$

32)

Sejam os eventos:

$A \equiv A$ falhar
$B \equiv B$ falhar
C=A-$B \equiv A$ falhar sozinho
D=B-$A \equiv B$ falhar sozinho
$P(A)$=0,20 $P(A \cap B)$=0,15 $P(D)$=0,15

$$a)\ B = (B - A) \cup (A \cap B) = D \cup (A \cap B) \Rightarrow P(B) = P(D) + P(A \cap B) =$$
$$= 0{,}15 + 0{,}15 = 0{,}30$$
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0{,}15}{0{,}30} = 0{,}50$$
$$b)\ A = (A - B) \cup (A \cap B) = C \cup (A \cap B) \Rightarrow P(A) = P(C) + P(A \cap B) \Rightarrow$$
$$\Rightarrow P(C) = P(A) - P(A \cap C) = 0{,}20 - 0{,}15 = 0{,}05$$

33)

Seja $P_A(a)$ a probabilidade de que A tenha a desarranjos, $P_B(b)$ a de que a B tenha b desarranjos e $P(a,b)$ a probabilidade de que A tenha a desarranjos e B tenha b desarranjos, conjuntamente. Como as máquinas funcionam independentemente, tem-se $P(a,b)$=$P_A(a)P_B(b)$. A tabela abaixo, chamada de Distribuição Conjunta de Probabilidades, apresenta, em suas células, essas probabilidades.

		A							$P_B(b)$
		0	1	2	3	4	5	6	
B	0	0,03	0,06	0,09	0,06	0,023	0,021	0,012	0,3
	1	0,01	0,02	0,03	0,02	0,009	0,009	0,004	0,1
	2	0,01	0,02	0,03	0,02	0,009	0,007	0,004	0,1
	3	0,01	0,02	0,03	0,02	0,009	0,007	0,004	0,1
	4	0,01	0,02	0,03	0,02	0,009	0,007	0,004	0,1
	5	0,015	0,03	0,045	0,03	0,0135	0,0105	0,006	0,15
	6	0,015	0,03	0,045	0,03	0,0135	0,0105	0,006	0,15
$P_A(a)$		0,1	0,2	0,3	0,2	0,09	0,07	0,04	1

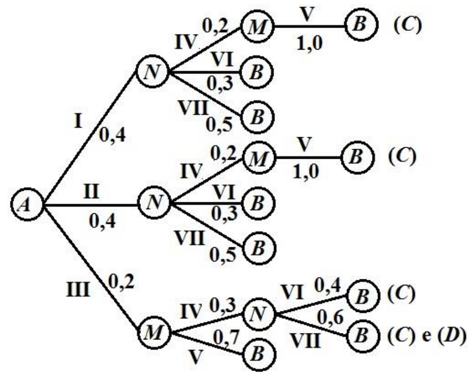
Para determinação das probabilidades pedidas, basta identificar na tabela os pares (a,b) que satisfazem às condições especificadas e somar as probabilidades $P(a,b)$ correspondentes. Veja a solução do item *a* e, para os demais, é só seguir o mesmo procedimento.

$$a)\ A=\{(a,b)|a=b\}=\{(0,0),(1,1),\ldots,(6,6)\}\Rightarrow P(A)=P(0,0)+P(1,1)+\ldots+P(6,6)=$$
$$=P_A(0)P_B(0)+P_A(1)P_B(1)+\ldots+P_A(6)P_B(6)=0{,}1\times0{,}3+0{,}2\times0{,}1+\ldots+0{,}04\times0{,}15=$$
$$=0{,}03+0{,}02+\ldots+0{,}006=0{,}1255$$

		A							$P_B(b)$
		0	1	2	3	4	5	6	
B	0	0,03	0,06	0,09	0,06	0,023	0,021	0,012	0,3
	1	0,01	0,02	0,03	0,02	0,009	0,009	0,004	0,1
	2	0,01	0,02	0,03	0,02	0,009	0,007	0,004	0,1
	3	0,01	0,02	0,03	0,02	0,009	0,007	0,004	0,1
	4	0,01	0,02	0,03	0,02	0,009	0,007	0,004	0,1
	5	0,015	0,03	0,045	0,03	0,0135	0,0105	0,006	0,15
	6	0,015	0,03	0,045	0,03	0,0135	0,0105	0,006	0,15
$P_A(a)$		0,1	0,2	0,3	0,2	0,09	0,07	0,04	1

$$b)\ A_1=\{(a,b)|a+b<4\}$$
$$A_2=\{(a,b)|a+b<5\}$$
$$c)\ A=\{(a,b)|a>b\}$$
$$d)\ A=\{(a,b)|b=2a\}$$
$$e)\ A_1=\{(a,b)|b\geq 2)\}$$
$$A_2=\{(a,b)|b=4)\}$$
$$A_1\cap A_2=A_2\Rightarrow P(A_2|A_1)=\frac{P(A_2)}{P(A_1)}$$
$$f)\ A_1=\{(a,b)|\min(a,b)=3\}$$
$$A_2=\{(a,b)|\min(a,b)<3\}$$
$$g)\ A_1=\{(a,b)|\max(a,b)=3\}$$
$$A_2=\{(a,b)|\max(a,b)>3\}$$

34)



$a)\ S=\{(I,IV,V);(I,VI);(I,VII);(II;IV;V);(II;VI);(II,VII);(III,IV,VI);$
$(III,IV,VII);(III,V)\}$

$b)\ C\equiv$"o viajante passa por C" $\qquad D\equiv$"o viajante passa por D"

$$P(C)=0{,}4\times 0{,}2+0{,}4\times 0{,}2+0{,}2\times 0{,}3\times 0{,}4+0{,}2\times 0{,}3\times 0{,}6=0{,}22$$

$$P(D)=0{,}4\times 0{,}5+0{,}4\times 0{,}5+0{,}2\times 0{,}3\times 0{,}6=0{,}436$$

$$P(C\cap D)=0{,}2\times 0{,}3\times 0{,}6=0{,}036$$

$$P(C\cup D)=P(C)+P(D)-P(C\cap D)=0{,}22+0{,}436-0{,}036=0{,}62$$

35)

$$a)\ P(A_1A_2 \text{ ou } A_2A_3)=P(A_1A_2)+P(A_2A_3)=p_1p_{1,2}+p_2p_{2,3}=\frac{5}{21}\cdot\frac{1}{3}+\frac{6}{21}\cdot\frac{1}{3}=\frac{11}{63}$$

$$b)\ P(A_iA_iA_i)=P(A_1A_1A_1 \text{ ou } A_2A_2A_2 \text{ ou } A_3A_3A_3)=P(A_1A_1A_1)+P(A_2A_2A_2)+P(A_3A_3A_3)=$$

$$=p_1p_{1,1}p_{1,1}+p_2p_{2,2}p_{2,2}+p_3p_{3,3}p_{3,3}=\frac{5}{21}\cdot 0^2+\frac{6}{21}\cdot\left(\frac{1}{3}\right)^2+\frac{10}{21}\cdot\left(\frac{7}{9}\right)^2=\frac{6}{189}+\frac{490}{1701}=\frac{544}{1701}$$

Parte 2

Variáveis Aleatórias

Capítulo 1

Variáveis Aleatórias

Introdução

No planejamento de qualquer atividade, diversas variáveis estão envolvidas e seus “comportamentos” têm influência decisiva neste planejamento. Por exemplo, ao se dimensionar uma central de atendimento, a quantidade de usuários (cliente) que procuram o serviço e o tempo que o(a) atendente (servidor) necessita para o atendimento são fundamentais. É evidente que, dependendo do tipo de serviço que será oferecido, essas variáveis têm quantidade média de demanda e tempo médio de atendimento diferentes. Mas, dentro de pressupostos razoáveis, podemos considerar que elas têm o mesmo modelo de comportamento (para a mesma variável), independentemente do tipo de serviço oferecido. Assim, passa ser interesse do planejador conhecer que tipo de modelo genérico tem o serviço de uma central de atendimento, restando a ele, então, a determinação dos valores dos parâmetros, por exemplo a média, que determinam o modelo para situações específicas.

Os comportamentos dessas variáveis, chamadas de VARIÁVEIS ALEATÓRIAS, são dados em termos de probabilidade e os modelos que os representam são chamados de MODELOS PROBABILÍSTICOS e permitem que se determinem as probabilidades de que elas (variáveis) assumam determinados valores de interesse como, por exemplo, a probabilidade de que o tempo de atendimento seja superior a 5 minutos. Retornando ao caso do lançamento de uma moeda podemos determinar um modelo probabilístico que represente qualquer situação semelhante, tendo como parâmetros a quantidade de lançamentos e a probabilidade de se obter cara em um único lançamento. Assim, de forma semelhante, no caso da fabricação de peças, conhecendo-se a probabilidade de que uma peça fabricada seja defeituosa, podemos calcular, por exemplo num lote de 100 peças, qual a probabilidade de que no máximo 10 sejam defeituosas.

1 Definição

Formalmente, vamos definir a Variável Aleatória como uma função X que associa os elementos do espaço amostral de um experimento aleatório a um número real, ou seja:

$$X : S \to R$$

Uma variável aleatória pode ser discreta, quando seu espaço amostral (possíveis valores) S_X é discreto (finito ou finito enumerável) ou contínua, quando seu espaço amostral S_X é contínuo (infinito).

Importante

- Se o espaço amostral S do experimento que dá origem à variável aleatória for discreto, o espaço amostral da variável também será discreto;

- Se S for contínuo, S_X poderá ser contínuo ou discreto. Por exemplo, ao se testar uma lâmpada e anotar o tempo até que ela queime, S é contínuo. Se X é a variável que associa o tempo obtido a ele próprio, então S_X também é contínuo, correspondendo ao próprio S, ao passo que, se classificarmos uma lâmpada como defeituosa se ela durar menos que 100 horas e como boa se durar 100 horas ou mais, podemos associar os tempos inferiores a 100 ao valor 0 (defeituosa) e os tempos maiores ou iguais a 100 ao valor 1 (boa). Neste caso, S_X é discreto.

2 Variáveis Aleatórias Discretas

Se X é uma variável aleatória com espaço amostral finito, S_X=$\{x_1,x_2,\ldots,x_n\}$, ou infinito enumerável, S_X=$\{x_1,x_2,\ldots,x_n,\ldots\}$, então X é do tipo discreto, com:

$$1)\ 0 \le P_X(X = x_i) \le 1$$

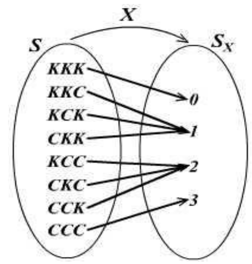
$$2)\ \sum_{i=1}^{\infty} P_X(X = x_i) = 1$$

Onde $P_X(X = x_i)$ é a probabilidade de que a variável X assuma determinado valor x_i.

EXEMPLO 2.1.2[1]

Seja o experimento que consiste em lançar uma moeda 3 vezes e X a variável aleatória que dá a quantidade de caras obtidas nos 3 lançamentos.

$C \equiv$ cara, $K \equiv$ coroa
S = {$KKK,KKC,KCK,CKK,KCC,CKC,CCK,CCC$}
$X(KKK) = 0$ $X(KKC) = X(KCK) = X(CKK) = 1$
$X(KCC) = X(CKC) = X(CCK) = 2$ $X(CCC) = 3$
$S_X = \{0,1,2,3\}$



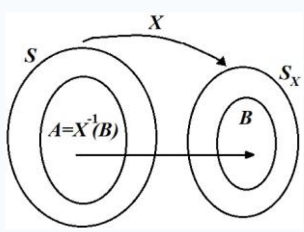
A probabilidade de que a variável aleatória X assuma algum valor é dada pela probabilidade dos eventos de S (espaço amostral do experimento que originou X) associados ao resultado desejado de X.

No exemplo 2.1.2[1], considerando S equiprovável, as probabilidades de se obter exatamente 0 caras, 1 cara, 2 caras e 3 caras são dadas por:

$$P_X(X=0) = P(kkk)\frac{1}{8}; P_X(X=1) = P(CKK, KCK, KKC) = \frac{3}{8};$$
$$P_X(X=2) = P(CCK, CKC, KCC) = \frac{3}{8}; P_X(X=3) = P(CCC) = \frac{1}{8}$$
$$0 \le P_X(X=x) \le 1, x = 0,1,2,3$$
$$\sum_{i=1}^{4} P_X(X=x_i) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$$

Observações

- Se B é um evento de S_X (espaço amostral da variável aleatória), sua imagem inversa, denotada por $X^{-1}(B)$ é o evento A de S (espaço amostral do experimento aleatório) que tem seus elementos associados a B, no exemplo apresentado (2.1.2[1]), o evento de S_X, $B\equiv$"obter no máximo 1 cara", B={0,1}, tem como imagem inversa o evento de S, A={KKK,KKC,KCK,CKK}, ou seja, X^{-1}({0,1})={KKK,KKC,KCK,CKK};
- Assim, na observação de uma variável aleatória, a probabilidade de que ocorra um evento B, ou seja, seu valor x pertença ao subconjunto B, $P_X(x \in B)$, é dada pela probabilidade de sua imagem inversa no experimento que originou X: $P_X(x \in B)=P(X^{-1}(B))$ ou, simplesmente, $P_X(B)=P(X^{-1}(B))$.

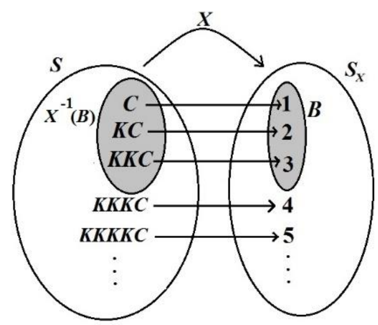


EXEMPLO 2.1.2[2]

Seja o experimento $E \equiv$ lançar uma moeda até se obter uma cara (C). Qual a probabilidade de serem necessários no máximo 3 lançamentos para obtenção de uma cara ?

$X \equiv$ quantidade de lançamentos até se obter uma cara

$$B = \{1,2,3\}; X^{-1}(B) = \{C, KC, KKC\}$$
$$P(X_X \le 3) = P(B) = P(X^{-1}(B)) =$$
$$= P(C, KC, KKC) = P(C) + P(KC) + P(KKC) =$$
$$= \frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{7}{8}$$



2.1 Distribuição de Probabilidade

No caso discreto, a Distribuição de Probabilidade de uma variável aleatória X é uma tabela que associa os possíveis valores de X às suas respectivas probabilidades. No lançamento da moeda 3 vezes, com $X \equiv$ quantidade de caras (exemplo 2.1.2[1]), temos (a partir deste ponto vamos utilizar a notação $P(X=x)$ no lugar de $P_X(X=x)$):

x_i	$P(X=x_i)$
0	1/8
1	3/8
2	3/8
3	1/8
SOMA	1

EXEMPLO 2.1.2.1[1]

Construa a Distribuição de Probabilidade do exemplo 2.1.2[2] e verifique que ela atende às condições para que seja realmente uma Distribuição de Probabilidade.

x_i	$P(X=x_i)$
1	1/2
2	1/4
3	1/8
4	1/16
...	...
SOMA	1

1ª condição: $P(X=x)=\left(\frac{1}{2}\right)^x=\frac{1}{2^x}, x=1,2,\ldots \Rightarrow 0 \le P(X=x) \le 1$ (satisfeita)

2ª condição: $\sum_{i=1}^{\infty} P(X=x_i)=\underbrace{\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\ldots}_{PG\ com\ a_1=\frac{1}{2}\ e\ q=\frac{1}{2}}=\frac{1/2}{1-1/2}=1$ (satisfeita)

2.2 Função Densidade de Probabilidade – $f(x)$

A Função Densidade de Probabilidade é um modelo matemático, chamado de MODELO PROBABILÍSTICO que representa o "comportamento" das variáveis aleatórias. No caso discreto, é uma função que, se aplicada a um valor x_i , resulta na probabilidade da variável aleatória X assumir este valor:

$$f: R \to R$$

$$f(x)=\begin{cases} P(X=x) & , x=x_1, x_2, \ldots, x_n, \ldots \\ 0 & , x \neq x_1, x_2, \ldots, x_n, \ldots \end{cases}$$

$$1)\ 0 \le f(x) \le 1$$

$$2)\ \sum_{i=1}^{\infty} f(x_i)=1$$

EXEMPLO 2.1.2.2[1]

No lançamento da moeda 3 vezes com $X \equiv$ quantidade de caras (exemplo 2.1.2[1]) , temos:

$$f(x) = \begin{cases} \binom{3}{x} \cdot \frac{1}{8} & x = 0,1,2,3 \\ 0 & x \neq 0,1,2,3 \end{cases}$$

$$f(0) = \binom{3}{0} \cdot \frac{1}{8} = \frac{3!}{0!(3-0)!} \cdot \frac{1}{8} = 1 \cdot \frac{1}{8} = \frac{1}{8} = P(X = 0)$$

$$f(1) = \binom{3}{1} \cdot \frac{1}{8} = \frac{3!}{1!(3-1)!} \cdot \frac{1}{8} = 3 \cdot \frac{1}{8} = \frac{3}{8} = P(X = 1)$$

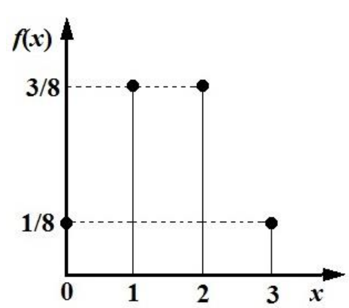
$$f(2) = \binom{3}{2} \cdot \frac{1}{8} = \frac{3!}{2!(3-2)!} \cdot \frac{1}{8} = 3 \cdot \frac{1}{8} = \frac{3}{8} = P(X = 2)$$

$$f(3) = \binom{3}{3} \cdot \frac{1}{8} = \frac{3!}{3!(3-3)!} \cdot \frac{1}{8} = 1 \cdot \frac{1}{8} = \frac{1}{8} = P(X = 3)$$

Importante

A Função Densidade de Probabilidade é definida para todos os Reais, mesmo para variáveis discretas. Veja que, no caso do lançamento da moeda 3 vezes com $X \equiv$ número de caras (exemplo 2.1.2[1]), a probabilidade de que X assuma um valor diferente de $0,1,2$ ou 3 é zero, por exemplo, a probabilidade de aparecerem 5 caras é zero, bem como a de aparecerem $2,5$ caras: $f(5)=0$ e $f(2,5)=0$.

Gráfico de $f(x)$:



As funções densidade de probabilidade são extremamente úteis, para generalizar o comportamento de variáveis aleatórias que seguem um determinado padrão. Por exemplo, consideremos o lançamento de uma moeda n vezes. Podemos, então, determinar a função densidade de probabilidade para este caso genérico (n lançamentos) e, a partir dela, calcularmos as probabilidades da variável $X \equiv$quantidade de caras, para qualquer número de lançamentos da moeda.

No caso de 3 lançamentos (exemplo 2.12.1[1]), a expressão $\binom{3}{x} \cdot \frac{1}{8}$ pode ser escrita como $\binom{3}{x} \cdot \left(\frac{1}{2}\right)^3$, onde 3 é a quantidade de lançamentos e $\frac{1}{2}$ é a probabilidade de se obter cara em um único lançamento da moeda. Por analogia, se a quantidade de lançamentos for n, o 3 será substituido por n e a função densidade passa a ser:

$$f(x) = \begin{cases} \binom{n}{x} \cdot \left(\frac{1}{2}\right)^n & x = 0,1,2,\ldots,n \\ 0 & x \neq 0,1,2,\ldots,n \end{cases}$$

Assim, no lançamento de uma moeda 10 vezes, a probabilidade de obtermos 4 caras será dada por:

$$n = 10; P(X = 4) = f(4) = \binom{10}{4} \cdot \left(\frac{1}{2}\right)^{10} = \frac{210}{1024}$$

Agora, suponha que a probabilidade de se obter cara num único lançamento de uma moeda seja $\frac{2}{3}$ e não $\frac{1}{2}$. Qual a expressão de $f(x)$ para n lançamentos dessa moeda ?

A expressão $\binom{n}{x} \cdot \left(\frac{1}{2}\right)^n$ pode ser escrita como $\binom{n}{x} \cdot \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{n-x}$, onde o primeiro $\frac{1}{2}$ é a probabilidade de cara, resultado de interesse, e o segundo é a probabilidade de coroa, veja que essas probabilidades são complementares. Assim, se a probabilidade de cara agora for $\frac{2}{3}$, a de coroa será $\frac{1}{3}$, e a expressão para $f(x)$, em n lançamentos, por analogia, passa a ser:

$$f(x) = \begin{cases} \binom{n}{x} \cdot \left(\frac{2}{3}\right)^x \cdot \left(\frac{1}{3}\right)^{n-x} & x = 0,1,2,\ldots,n \\ 0 & x \neq 0,1,2,\ldots,n \end{cases}$$

Enfim, a situação acima pode ser generalizada para o caso de n repetições independentes de um experimento com apenas dois resultados possíveis, $S \equiv$sucesso e $F \equiv$fracasso, com probabilidades $P(S)=p$ e $P(F)=1-p=q$. Assim, a variável aleatória $X \equiv$quantidade de sucessos nas n repetições tem Função Densidade de Probabilidade dada por:

$$f(x)=\begin{cases}\dbinom{n}{x}p^{x}q^{n-x} & x=0,1,2,\ldots,n\\ 0 & x\neq 0,1,2,\ldots,n\end{cases}$$

A variável aleatória descrita acima é dita ter DISTRIBUIÇÃO DE PROBABILIDADE BINOMIAL com parâmetros n e p, denotando-se por $X \sim B(n,p)$. Mais adiante este modelo probabilístico será visto detalhadamente.

> **Observação**
>
> Os parâmetros de um modelo probabilístico são os elementos da função cujos valores são necessários e suficientes para se determinar a expressão específica para uma situação particular. No caso do modelo Binomial, é preciso conhecer os valores de n e p, veja que o valor de q é determinado por p, pois são complementares. Portanto, n e p são seus parâmetros.

EXEMPLO 2.1.2.2[2]

Retornando ao exemplo 1.2.1[1], no caso em que as probabilidades de que certo tipo de peça tem probabilidades $\frac{4}{10}$, $\frac{3}{10}$, $\frac{2}{10}$ e $\frac{1}{10}$ de apresentar $0,1,2$ ou 3 defeitos. Seja X a variável aleatória que dá a quantidade de defeitos de uma peça comprada (selecionada) ao acaso. Determine a Distribuição de Probabilidade e a Função Densidade de Probabilidade de X. Qual a probabilidade de que uma peça apresente pelo menos um defeito ?

DISTRIBUIÇÃO DE PROBABILIDADE:

x_i	$P(X=x_i)$
0	4/10
1	3/10
2	2/10
3	1/10
SOMA	1

FUNÇÃO DENSIDADE DE PROBABILIDADE:

$$f(x)=\begin{cases}\dfrac{4-x}{10}, x=0,1,2,3\\ 0, x\neq 0,1,2,3\end{cases}$$

Sugestão

Construa o gráfico de $f(x)$

PROBABILIDADE DE PELO MENOS UM DEFEITO:

$$P(X \geq 1) = f(1) + f(2) + f(3) = \frac{3}{10} + \frac{2}{10} + \frac{1}{10} = \frac{6}{10}$$

Ou, pela propriedade do complementar

$$P(X \geq 1) = 1 - P(X = 0) = 1 - f(0) = 1 - \frac{4}{10} = \frac{6}{10}$$

EXEMPLO 2.1.2.2[3]

Uma urna contém n bolas numeradas de 1 a n. Duas bolas, (a_1,a_2), são retiradas ao acaso da urna, sucessivamente e com reposição. Seja X a variável que dá a soma dos números das bolas retiradas. Considerando o espaço amostral equiprovável, determine a Distribuição de Probabilidade e a Função Densidade de Probabilidade de X.

Os resultados possíveis do experimento, e as respectivas somas são:

POSSÍVEIS RESULTADOS

(1,1)	(2,1)	(3,1)	(4,1)	...	(n,1)
(1,2)	(2,2)	(3,2)	(4,2)	...	(n,2)
(1,3)	(2,3)	(3,3)	(4,3)	...	(n,3)
...	...	...	...	...	...
(1,n-2)	(2,n-2)	(3,n-2)	(4,n-2)	...	(n,n-2)
(1,n-1)	(2,n-1)	(3,n-1)	(4,n-1)	...	(n,n-1)
(1,n)	(2,n)	(3,n)	(4,n)	...	(n,n)

SOMAS

2	3	4	5	...	n+1
3	4	5	6	...	n+2
4	5	6	7	...	n+3
...	...	...	...	...	...
n-1	N	n+1	n+2	...	2n-2
n	n+1	n+2	n+3	...	2n-1
n+1	n+2	n+3	n+4	...	2n

DISTRIBUIÇÃO DE PROBABILIDADE:

A quantidade de resultados possíveis do experimento é n^2, e os valores possíveis da variável X são 2,3,4,...,n,n+1,n+2, ... 2n. Observe que a soma 2 aparece 1 vez, a 3 aparece 2 vezes e assim até a soma n-1 que aparece n-2 vezes, a soma n aparece n-1 vezes e a soma n+1 aparece n vezes. A partir de então, a soma n+2 aparece n-1 vezes, a soma n+3 aparece n-2 vezes, até a soma 2n que aparece 1 vez. Assim, a distribuição de probabilidade de X será:

Desafio

Verfifique que a soma de $P(X=x)=1$

x	$P_X(X=x)$
2	$1/n^2$
3	$2/n^2$
4	$3/n^2$
...	...
n-1	$(n-2)/n^2$

x	$P_X(X=x)$
n	$(n-1)/n^2$
n+1	n/n^2
n+2	$(n-1)/n^2$
...	...
2n-2	$3/n^2$

x	$P_X(X=x)$
2n-1	$2/n^2$
2n	$1/n^2$
SOMA	1

FUNÇÃO DENSIDADE DE PROBABILIDADE:

Para os valores de 2 a n+1 é fácil verificar que o numerador de $f(x)$ é x-1. Para os valores de n+2 a 2n, basta escolher dois pares quaisquer, por exemplo $(n+2,n-1)$ e $(n+3,n-2)$, e determinar a reta que passa por esses pontos. Assim, a Função Densidade de Probabilidade de X será dada por:

Sugestões

- Estipule um valor para n e verifique a validade da Distribuição de Probabilidade e $f(x)$ encontradas e construa o gráfico de $f(x)$
- Determine a Função Densidade de Probabilidade da variável do exemplo 2.1.2.1[1] e construa seu gráfico

$$f(x)=\begin{cases}\dfrac{x-1}{n^2}, x=2,3,4,\ldots n, n+1\\ \dfrac{-x+2n+1}{n^2}, x=n+2, n+3,\ldots,2n-1,2n\\ 0, x\neq 2,3,4,\ldots,2n\end{cases}$$

2.3 Função distribuição acumulada de probabilidade – $F(x)$

A Função Distribuição Acumulada de Probabilidade, em ambos os casos, discreto ou contínuo, é uma função que, se aplicada a um valor x, fornece a probabilidade da variável aleatória X assumir um valor menor ou igual a x:

$$F: R \to R$$
$$F(x)=P(X\leq x)$$

No caso discreto temos:

$$F(x)=\sum_{x_i\leq x} P(X=x_i)$$

EXEMPLO 2.1.2.3[1]

No exemplo 2.1.2[1], referente ao lançamento de uma moeda 3 vezes, com $X \equiv$quantidade de caras, temos:

x_i	$P(X=x_i)$	$P(X \leq x_i)$
0	1/8	1/8
1	3/8	4/8
2	3/8	7/8
3	1/8	1
SOMA	1	-

$$F(x) = \sum_{x_i \leq x} P(X = x_i) = \sum_{x_i \leq x} f(x_i) = \sum_{x_i \leq x} \binom{3}{x_i} \times \frac{1}{8}$$

$$F(0) = P(X \leq 0) = P(X = 0) = \frac{1}{8}$$

$$F(1) = P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{1}{8} + \frac{3}{8} = \frac{4}{8}$$

$$F(2) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}$$

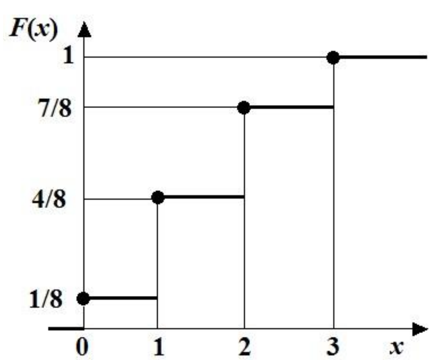
$$F(3) = P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) =$$

$$= \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$$

Importante

Da mesma forma que a Função Densidade, a Função de Distribuição também é definida para todos os reais. Assim, se $x<0 \Rightarrow F(x)=0$, se $0 \leq x<1 \Rightarrow F(x)=1/8$, se $1 \leq x<2 \Rightarrow F(x)=4/8$, se $2 \leq x<3 \Rightarrow F(x)=7/8$ e se $x \geq 3 \Rightarrow F(x)=1$. Por exemplo, $F(-1)=0$, $F(0{,}8)=1/8$, $F(1{,}5)=4/8$, $F(2{,}1)=7/8$ e $F(5)=1$

Gráfico de $F(x)$:



Sugestão

Determine a Função Distribuição Acumulada de Probabilidade das variáveis dos exemplos 2.1.2..2[2] e 2.1.2.2[3] e construa seus gráficos

2.4 Esperança Matemática ou Valor Esperado – $E(X)$

No caso discreto, a Esperança Matemática de uma variável aleatória X é definida como:

$$E(X)=\sum_{i=1}^{\infty}x_iP(X=x_i)$$

A esperança matemática de uma função de X, $g(X)$, é dada por:

$$E(g(X))=\sum_{i=1}^{\infty}g(x_i)P(X=x_i)$$

EXEMPLO 2.1.2.4[1]

Para o lançamento da moeda 3 vezes, com X≡quantidade de caras (exemplo 2.1.2[1]), temos:

$$E(X)=\sum_{i=1}^{4}x_iP(X=x_i)=0\cdot\frac{1}{8}+1\cdot\frac{3}{8}+2\cdot\frac{3}{8}+3\cdot\frac{1}{8}=\frac{12}{8}=\frac{3}{2}$$
$$E(X^2)=\sum_{i=1}^{4}x_i^2P(X=x_i)=0^2\cdot\frac{1}{8}+1^2\cdot\frac{3}{8}+2^2\cdot\frac{3}{8}+3^2\cdot\frac{1}{8}=\frac{24}{8}=3$$

Sugestão

Calcule a Esperança Matemática das variáveis dos exemplos 2.1.2.2[2] e 2.1.2.2[3]

Observações

- É fácil verificar que a Esperança Matemática é a média aritmética dos valores da variável X, ponderados por suas probabilidades e determina a média dos valores observados de X em repetidas realizações do experimento. Por exemplo, a quantidade de pessoas que acessam uma determinada página na internet em um mês é uma variável aleatória e sua Esperança Matemática é a média mensal de pessoas que acessam a página, ao longo do tempo;
- A unidade da Esperança Matemática é a mesma da variável. Por exemplo, se a variável é dada em centímetros (cm), a Esperança Matemática também é dada em cm.

EXEMPLO 2.1.2.4[2]

Retornando ao exemplo 1.2.1[1], no caso em que as probabilidades de que certo tipo de peça tem probabilidades $\frac{4}{10}$, $\frac{3}{10}$, $\frac{2}{10}$ e $\frac{1}{10}$ de apresentar $0{,}1{,}2$ ou 3 defeitos. Qual a média de defeitos por peça fabricada ? O fabricante vende uma peça por $R\$50{,}00$, mas devolve ao comprador $R\$2{,}00$ por cada defeito apresentado. Qual a receita líquida esperada pelo fabricante, por peça ?

MÉDIA DE DEFEITOS:

$E(X) = 0\times\frac{4}{10} + 1\times\frac{3}{10} + 2\times\frac{2}{10} + 3\times\frac{1}{10} = 1$ defeito por peça

RECEITA LÍQUIDA ESPERADA:

Distribuição de Probabilidade da Receita Líquida R

r=50-2x	$P(R=r)$
50-2x0=50	$P(R=50)=P(X=0)=4/10$
50-2x1=48	$P(R=48)=P(X=1)=3/10$
50-2x2=46	$P(R=46)=P(X=2)=2/10$
50-2x3=44	$P(R=44)=P(X=3)=1/10$

$E(R) = 50\times\frac{4}{10} + 48\times\frac{3}{10} + 46\times\frac{2}{10} + 44\times\frac{1}{10} = \frac{200}{10} + \frac{144}{10} + \frac{92}{10} + \frac{44}{10} = \frac{480}{10} = R\$48{,}00$ por peça

2.5 Variância – $V(X)$

A variância de uma variável aleatória, tanto no caso discreto como no contínuo, é definida como:

$$V(X) = E[X - E(X)]^2 \quad \text{ou} \quad V(X) = E(X^2) - \{E(X)\}^2$$

Observações

- Lembrando que a Esperança Matemática é a média dos valores da variável aleatória e analisando a definição de Variância, verificamos que $V(X)$ é a média das distâncias (ao quadrado) dos valores de X em relação à sua média $E(X)$;
- O enunciado acima mostra que a Variância é uma medida indicativa do grau de dispersão (variabilidade) dos valores da variável em torno de sua média, ou seja, quanto mais dispersos forem os valores (grandes distâncias em relação à média), maior será a Variância;
- À raiz quadrada da variância, dá-se o nome de Desvio Padrão, denotado por σ_X. Também é comum denotar a Variância por σ_X^2 e a Esperança por μ_X.
- A unidade da variância é a unidade da variável ao quadrado e a do desvio padrão é a mesma da variável. Por exemplo, se a variável é dada em centímetros (cm), a variância é dada em cm^2 e o desvio padrão é dado em cm.

EXEMPLO 2.1.2.5[1]

No lançamento da moeda 3 vezes, com $X \equiv$quantidade de caras (exemplo 2.1.2[1]), temos:

$$\mu_X = E(X) = \frac{3}{2}$$

$$\mu_{X^2} = E(X^2) = 3$$

$$\sigma_X^2 = V(X) = E(X^2) - [E(X)]^2 = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4}$$

$$\sigma_X = \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2}$$

EXEMPLO 2.1.2.5[2]

Considerando o exemplo 2.1.2.4[2], qual a variância e o desvio padrão da quantidade de defeitos e da receita líquida ?

$$E(X) = 1$$

$$E(X^2) = 0^2 \times \frac{4}{10} + 1^2 \times \frac{3}{10} + 2^2 \times \frac{2}{10} + 3^2 \times \frac{1}{10} = 2$$

$$V(X) = E(X^2) - \{E(X)\}^2 = 2 - 1^2 = 1 \qquad \sigma_X = 1$$

$$E(R) = 48$$

$$E(R^2) = 50^2 \times \frac{4}{10} + 48^2 \times \frac{3}{10} + 46^2 \times \frac{2}{10} + 44^2 \times \frac{1}{10} = \frac{10000}{10} + \frac{6912}{10} + \frac{4232}{10} + \frac{1936}{10} =$$

$$= \frac{23080}{10} = 2308$$

$$V(R) = 2308 - 48^2 = 2308 - 2304 = 4{,}00\,(R\$^2) \qquad \sigma_R = R\$2{,}00$$

3 Variáveis Aleatórias Contínuas

Diz-se que uma variável aleatória X é contínua se seu espaço amostral S_X é um intervalo ou a união de intervalos reais.

3.1 Função densidade de probabilidade – $f(x)$

Se X é uma variável aleatória contínua, sua Função Densidade de Probabilidade é uma função satisfazendo a:

1) $f(x) \geq 0 \ \forall x \in S_X$

2) $f(x) = 0 \ \forall x \notin S_X$

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

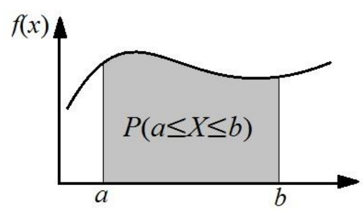
Observações

- A condição 3 indica que a área total sob $f(x)$ é igual a 1, o que é equivalente ao caso discreto, no qual a soma de todas as probabilidades $P(X=x)$ também é igual a 1.
- Para variáveis discretas, o cálculo de probabilidades e de suas medidas características (esperança e variância) são feitos através de somas. Para as variáveis contínuas esses cálculos são através de áreas (integrais)
- Diferente do caso discreto, a função densidade de uma variável contínua aplicada a um valor x não resulta em $P(X=x)$. No caso contínuo $P(X=x)$=0

Para variáveis contínuas, as probabilidades serão determinadas por áreas sob $f(x)$, lembrando que a área total é igual a 1 (condição 3), assim, a área sob a curva em determinado intervalo $[a,b]$ é a probabilidade de que X assuma um valor dentro deste intervalo:

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

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



Importante

Como consequências do enunciado acima, tem-se:

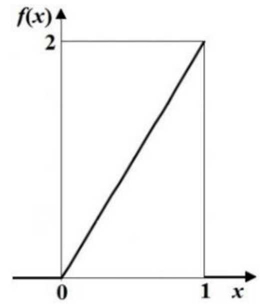
1) $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$

2) $P(X = c) = P(c \leq X \leq c) = \int_c^c f(x)dx = 0$

EXEMPLO 2.1.3.1[1]

O diâmetro (cm) de certo tipo de cabo é uma variável aleatória com Função Densidade de Probabilidade dada por $f(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & x \notin [0,1] \end{cases}$. Construa o gráfico de $f(x)$. Verifique que $f(x)$ realmente é uma Função Densidade de probabilidade. Qual a probabilidade de que um cabo selecionado (comprado) ao acaso tenha diâmetro entre $0{,}25cm$ e $0{,}75cm$?

GRÁFICO DE $f(x)$



VERIFICAÇÃO DE QUE $f(x)$ É UMA DENSIDADE DE PROBABILIDADE

1ª condição

Para $x<0$ ou $x>1 \Rightarrow f(x)=0$ (definição da função). Para $x \in [0,1] \Rightarrow x \geq 0 \Rightarrow 2x \geq 0$ (satisfeita)

2ª condição

Pela integral: $\int_{-\infty}^{+\infty} f(x)dx = \int_0^1 2xdx = \left.\frac{2x^2}{2}\right|_0^1 = 1$.

Pela área da figura(triângulo): $\frac{1}{2} \times 1 \times 2 = 1$ (satisfeita)

PROBABILIDADE DE $0{,}25 \leq X \leq 0{,}75$

Pela integral: $P\left(\frac{1}{4} \leq X \leq \frac{3}{4}\right) = \int_{1/4}^{3/4} 2xdx = \left.\frac{2x^2}{2}\right|_{1/4}^{3/4} = \frac{1}{2}$

Pela área da figura(trapézio): $P\left(\frac{1}{4} \leq X \leq \frac{3}{4}\right) = \dfrac{\frac{1}{2}\left(\frac{3}{2}+\frac{1}{2}\right)}{2} = \frac{1}{2}$

Importante

A probabilidade $P(0{,}25 \leq X \leq 0{,}75) = 1/2$, na prática, corresponde a afirmar que cerca de 50% das peças fabricadas têm diâmetro entre $0{,}25cm$ e $0{,}75cm$

3.2 Função Distribuição Acumulada de Probabilidade – $F(x)$

Da mesma forma que no caso discreto, a Função de Distribuição Acumulada de uma variável contínua é dada por $F(x)=P(X\leq x)$, então:

$$F(x)=P(X\leq x)=\int_{-\infty}^{x} f(t)dt$$

OBS: $f(t)$ tem a mesma expressão de $f(x)$

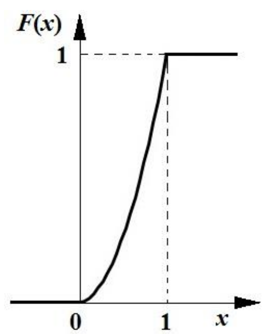
EXEMPLO 2.1.3.2[1]

Retornado ao exemplo 2.1.3.1[1], onde a variável aleatória X (diâmetro de um tipo de cabo) tem função densidade de probabilidade dada por $f(x)=\begin{cases}2x & x\in[0,1]\\ 0 & x\notin[0,1]\end{cases}$, sua função de distribuição acumulada será dada por:

$$x<0 \Rightarrow F(x)=\int_{-\infty}^{x} 0dt=0$$

$$0\leq x\leq 1 \Rightarrow F(x)=\int_{-\infty}^{x} 0dt+\int_{0}^{x} 2tdt=0+x^2=x^2$$

$$x>1 \quad F(x)=\int_{-\infty}^{0} 0dt+\int_{0}^{1} 2tdt+\int_{1}^{x} 0dt=0+1+0=1$$



Então: $$F(x)=\begin{cases}0 & x<0\\ x^2 & 0\leq x\leq 1\\ 1 & x>1\end{cases}$$

Observação

Se X é uma variável aleatória, discreta ou contínua, e $F(x)$ a função de distribuição de X, então:

$$1)\ 0\leq F(x)\leq 1 \quad \forall x\in R$$

$$2)\ se\ x<y \quad F(x)\leq F(y)$$

$$3)\ \lim_{x\to -\infty} F(x)=0\ e\ \lim_{x\to\infty} F(x)=1$$

Importante

- A probabilidade $P(a\leq X\leq b)$ pode ser calculada a partir da Função Distribuição Acumulada de Probabilidade: $P(a\leq X\leq b)=F(b)-F(a)$
- Assim, uma vez determinada $F(x)$, não é mais necessária a operação de calcular a integral de $f(x)$ toda vez que se for calcular determinadas probabilidades

EXEMPLO 2.1.3.2[2]

Ainda utilizando o exemplo 2.1.3.1[1], onde a variável aleatória X (diâmetro de um tipo de cabo) tem função densidade de probabilidade dada por $f(x)=\begin{cases} 2x & x\in[0,1] \\ 0 & x\notin[0,1] \end{cases}$, a probabilidade de que um cabo selecionado ao acaso tenha diâmetro entre $0{,}25cm$ e $0{,}75cm$ pode ser calculada da seguinte forma:

$$F(x)=\begin{cases} 0 & x<0 \\ x^2 & 0\le x\le 1 \\ 1 & x>1 \end{cases} \quad P\left(\frac{1}{4}\le X\le\frac{3}{4}\right)=F\left(\frac{3}{4}\right)-F\left(\frac{1}{4}\right)=\left(\frac{3}{4}\right)^2-\left(\frac{1}{4}\right)^2=\frac{9}{16}-\frac{1}{16}=\frac{1}{2}$$

3.3 Esperança Matemática e Variância

Para as variáveis contínuas, a esperança e variância, cujos significados já foram abordados no caso discreto, são determinadas pelas expressões seguintes:

DESAFIO

Mostre que

$E[X-E(X)]^2 =$
$=E(X^2)-[E(X)]^2$

$$E(X)=\int_{-\infty}^{+\infty} xf(x)dx$$

$$E[g(X)]=\int_{-\infty}^{+\infty} g(x)\cdot f(x)dx$$

$$V(X)=E[X-E(X)]^2 \quad \text{ou} \quad V(X)=E(X^2)-[E(X)]^2$$

EXEMPLO 2.1.3.3[1]

Retornado ao exemplo 2.1.3.1[1], no qual a variável aleatória X (diâmetro de um tipo de cabo) tem função densidade de probabilidade dada por $f(x)=\begin{cases} 2x & x\in[0,1] \\ 0 & x\notin[0,1] \end{cases}$, a esperança e a variância são dadas por:

$$E(X)=\int_{-\infty}^{+\infty} xf(x)dx=\int_0^1 x2xdx=2\int_0^1 x^2dx=\left.\frac{2x^3}{3}\right|_0^1=\frac{2}{3}=0{,}67cm$$

$$E(X^2)=\int_{-\infty}^{+\infty} x^2 f(x)dx=\int_0^1 x^2\,2xdx=2\int_0^1 x^3dx=\left.\frac{2x^4}{4}\right|_0^1=\frac{1}{2}\Rightarrow V(X)=\frac{1}{2}-\left(\frac{2}{3}\right)^2=$$

$$=\frac{1}{18}=0{,}06cm^2$$

Importante

A interpretação de $E(X)$ é que, no processo de produção deste tipo de cabo, o diâmetro médio dos mesmos é de $0{,}67cm$

4 Propriedades da Esperança Matemática e Variância

Se X é uma variável aleatória, discreta ou contínua, e c uma constante, as seguintes propriedades são verificadas:

ESPERANÇA	VARIÂNCIA
1) $E(X \pm c) = E(X) \pm c$	1) $V(X \pm c) = V(X)$
2) $E(cX) = cE(X)$	2) $V(cX) = c^2 V(X)$
3) $E(c) = c$	3) $V(c) = 0$

Desafio

Demonstre as propriedades da Esperança e da Variância

EXEMPLO 2.1.4[1]

Seja X uma variável aleatória contínua com Função Densidade de Probabilidade dada por:

$$f(x) = \begin{cases} 1+x & -1 \le x \le 0 \\ 1-x & 0 \le x \le 1 \\ 0 & x \notin [-1,1] \end{cases}$$

Calcule $E(X)$, $E(X^2)$, $E(X+2)$ e $E(2X^2)$

$$E(X) = \int_{-1}^{0} x(1+x)dx + \int_{0}^{1} x(1-x)dx = 0 \qquad E(X+2) = E(X) + 2 = 0 + 2 = 2$$

$$E(X^2) = \int_{-1}^{0} x^2(1+x)dx + \int_{0}^{1} x^2(1-x)dx = \frac{1}{6} \qquad E(2X^2) = 2E(X^2) = 2.\frac{1}{6} = \frac{1}{3}$$

Calcule $V(X)$, $V(X+2)$ e $V(2X)$

$$\left.\begin{aligned} E(X) &= 0 \\ E(X^2) &= \frac{1}{6} \end{aligned}\right\} \Rightarrow V(X) = \frac{1}{6} - 0^2 = \frac{1}{6}$$

$$V(X+2) = V(X) = \frac{1}{6}$$

$$V(2X) = 2^2 V(X) = 4 \cdot \frac{1}{6} = \frac{2}{3}$$

EXEMPLO 2.1.2.4[2]

Retornando aos exemplos 2.1.2.4[2] e 2.1.2.5[2], no caso em que as probabilidades de que certo tipo de peça tem probabilidades $\frac{4}{10}$, $\frac{3}{10}$, $\frac{2}{10}$ e $\frac{1}{10}$ de apresentar $0{,}1{,}2$ ou 3 defeitos e fabricante vende uma peça por $R\$50{,}00$, mas devolve ao comprador $R\$2{,}00$ por cada defeito apresentado. A receita líquida esperada e a variância podem ser calculadas da seguinte forma:

$E(X)=1$ e $V(X)=1$

A receita líquida pode ser escrita como: $R=50-2X$

Então

$$E(R)=E(50-2X)=E(\underbrace{-2X}_{Y}+\underbrace{50}_{c})=E(-2X)+E(50)=-2E(X)+50=$$
$$=-2\times 1+50=48$$
$$V(R)=V(50-2X)=V(\underbrace{-2X}_{Y}+\underbrace{50}_{c})=V(-2X)=(-2)^2V(X)=4\times 1=4$$

EXEMPLO 2.1.4[3]

Retornando ao exemplo 2.1.3.3[1], no qual a variável aleatória X (diâmetro de um tipo de cabo) tem função densidade de probabilidade dada por $f(x)=\begin{cases}2x & x\in[0,1]\\ 0 & x\notin[0,1]\end{cases}$, considere que o custo de manutenção deste cabo é dado por $C=C_1+C_2X$. A Esperança e Variância do custo serão dadas por:

$$E(X)=\frac{2}{3}=0{,}67cm \qquad V(X)=\frac{1}{18}=0{,}06cm^2$$
$$V(C)=V(C_1+C_2X)=V(\underbrace{C_2X}_{Y}+\underbrace{C_1}_{c})=V(C_2X)=C_2^2\times V(X)=\frac{C_2^2}{18}$$
$$E(C)=E(C_1+C_2X)=E(\underbrace{C_2X}_{Y}+\underbrace{C_1}_{c})=E(C_2X)+E(C_1)=C_2E(X)+C_1=\frac{2C_2}{3}+C_1$$

Outras duas propriedades muito importantes da Esperança e Variância, envolvendo n variáveis aleatórias, X_1, X_2,..., X_n são as seguintes:

- $E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n)$ para quaisquer X_1, X_2,..., X_n
- $V(X_1 + X_2 + \ldots + X_n) = V(X_1) + V(X_2) + \ldots + V(X_n)$ se X_1, X_2,..., X_n são independentes

Capítulo 2

Algumas Variáveis Aleatórias Discretas

1 Binomial – $B(n,p)$

Seja E_1 um experimento com apenas dois resultados possíveis, S (sucesso) e F (fracasso), tal que $P(S)=p$ e $P(F)= 1\text{-}p = q$. Seja E o experimento que consiste em repetir n vezes, independentemente, o experimento E_1. Seja X a variável aleatória que dá a quantidade de sucessos obtidos nas n repetições (E_1 é conhecido como Experimento de Bernoulli). Então:

$$f(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & x = 0,1,2,3,\ldots,n \\ 0 & x \neq 0,1,2,3,\ldots,n \end{cases}$$

$$E(X) = np \qquad V(X) = npq$$

EXEMPLO 2.2.1[1]

Sabe-se que uma determinada moeda apresenta cara 3 vezes mais frequentemente do que coroa. Essa moeda é lançada 5 vezes. Seja X o número de caras que aparecem. Estabeleça a densidade de probabilidade discreta de X e calcule a probabilidade de que o número de caras seja maior ou igual a 2 e menor ou igual a 4. Calcule $E(X)$, $V(X)$ e σ.

$$P(C) = \frac{3}{4} \qquad P(K) = \frac{1}{4}$$

$$f(x) = \begin{cases} \binom{5}{x}\left(\frac{3}{4}\right)^x\left(\frac{1}{4}\right)^{3-x} & , x = 0,1,2,3,5 \\ 0 & , x \neq 0,1,2,3,5 \end{cases}$$

Desafio

- Determine $f(x)$ (parta de uma sequência específica de sucessos e fracassos)
- Mostre que $f(x)$ satisfaz às duas condições da definição de Função Densidade de Probabilidade (para mostrar que $\Sigma f(x)=1$ use o Binômio de Newton)
- Determine $E(X)$ e $V(X)$ (calcule a Esperança e Variância do Experimento de Bernoulli e depois use as propriedades da Esperança e Variância para a soma de n variáveis aleatórias independentes)

$$P(2 \le X \le 4) = f(2) + f(3) + f(4) = \binom{5}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^3 + \binom{5}{3}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^2 +$$

$$+\binom{5}{4}\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^1 = 10\cdot\frac{9}{16}\cdot\frac{1}{64} + 10\cdot\frac{27}{64}\cdot\frac{1}{16} + 5\cdot\frac{81}{256}\cdot\frac{1}{4} =$$

$$= \frac{90+270+405}{1024} = \frac{765}{1024}$$

$$E(X) = 5\cdot\frac{3}{4} = \frac{15}{4} \qquad V(X) = 5\cdot\frac{3}{4}\cdot\frac{1}{4} = \frac{15}{16} \qquad \sigma = \sqrt{\frac{15}{16}} = \frac{\sqrt{15}}{4}$$

2 Geométrica Discreta – *Geométrica*(p)

Seja E_1 um experimento cujos resultados sejam sucesso S ou fracasso F, tal que $P(S)=p$ e $P(F)=q=1-p$. Seja E o experimento constituído da repetição, independentemente, de E_1 até que ocorra 1 sucesso. Seja X a variável aleatória que dá a quantidade de repetições necessárias para se obter 1 sucesso. Então:

$$f(x) = \begin{cases} pq^{x-1} & x = 1,2,3,\ldots \\ 0 & x \ne 1,2,3,\ldots \end{cases}$$

$$E(X) = \frac{1}{p} \qquad V(X) = \frac{q}{p^2}$$

DESAFIO

Determine $f(x)$ (parta de uma sequência específica de sucessos e fracassos)

EXEMPLO 2.2.2[1]

Um dado é lançado até que apareça um 6. Qual a probabilidade de que no máximo 6 lançamentos sejam necessários ? Calcule $E(X)$, $V(X)$ e σ.

$$p = \frac{1}{6} \qquad q = 1-\frac{1}{6} = \frac{5}{6} \qquad f(x) = \begin{cases} \frac{1}{6}\cdot\left(\frac{5}{6}\right)^{x-1} & x = 1,2,\ldots \\ 0 & x \ne 1,2,\ldots \end{cases}$$

$$P(X \le 6) = \frac{1}{6}\left[\left(\frac{5}{6}\right)^0 + \left(\frac{5}{6}\right)^1 + \left(\frac{5}{6}\right)^2 + \left(\frac{5}{6}\right)^3 + \left(\frac{5}{6}\right)^4 + \left(\frac{5}{6}\right)^5\right]$$

$$E(X) = \frac{1}{1/6} = 6 \qquad V(X) = \frac{5/6}{(1/6)^2} = \frac{5}{6}\cdot 36 = 30 \qquad \sigma = \sqrt{30}$$

3 Binomial Negativa (Pascal) – *Pascal*(r,p)

Suponhamos que um experimento E_1 tenha como resultados sucesso (S) ou fracasso (F), com $P(S)=p$ e $P(F)=1$-$p=q$. Seja E o experimento que consiste em repetir o experimento E_1, independentemente, até que r sucessos ocorram. Seja X a variável que dá a quantidade de repetições necessárias até que ocorram r sucessos. Então:

Desafio

Determine $f(x)$ (parta de uma sequência específica de sucessos e fracassos)

$$f(x)=\begin{cases}\binom{x-1}{r-1}p^r q^{x-r} & , \quad x=r,r+1,r+2,\ldots\\ 0 & , \quad x\neq r,r+1,r+2,\ldots\end{cases}$$

$$E(X)=r\cdot\frac{1}{p} \qquad V(X)=r\cdot\frac{q}{p^2}$$

EXEMPLO 2.2.3[1]

Um dado é lançado até que o número 6 apareça 3 vezes. Qual a probabilidade de que exatamente 6 lançamentos sejam necessários ? Calcule $E(X)$, $V(X)$ e σ.

$$r=3 \quad p=\frac{1}{6} \quad q=1\text{-}\frac{1}{6}=\frac{5}{6} \quad f(x)=\begin{cases}\binom{x-1}{3-1}\cdot\left(\frac{1}{6}\right)^3\cdot\left(\frac{5}{6}\right)^{x-3} & x=3,4,5,\ldots\\ 0 & x\neq 3,4,5,\ldots\end{cases}$$

$$\Rightarrow f(x)=\begin{cases}\binom{x-1}{2}\cdot\frac{1}{216}\cdot\left(\frac{5}{6}\right)^{x-3} & x=3,4,5,\ldots\\ 0 & x\neq 3,4,5,\ldots\end{cases}$$

$$P(X=6)=\binom{6-1}{2}\cdot\frac{1}{216}\cdot\left(\frac{5}{6}\right)^{6-3}=\binom{5}{2}\cdot\frac{1}{216}\cdot\left(\frac{5}{6}\right)^{3}=10\cdot\frac{1}{216}\cdot\frac{125}{216}=\frac{1250}{46656}$$

$$E(X)=3\cdot\frac{1}{1/6}=18 \qquad V(X)=3\cdot\frac{5/6}{\left(1/6\right)^2}=3\cdot\frac{5}{6}\cdot 36=90 \qquad \sigma_X=\sqrt{90}$$

4 Hipergeométrica – *Hipergeométrica*(N,R,n)

Suponha que se tenha N objetos, dos quais R são de um tipo 1 e N-R de um tipo 2. Uma amostra não ordenada de extensão $n\leq N$ é retirada ao acaso, sem reposição. Seja X a variável aleatória que dá a quantidade de objetos do tipo 1 na amostra. Então:

$$f(x) = \begin{cases} \dfrac{\dbinom{R}{x}\dbinom{N-R}{n-x}}{\dbinom{N}{n}} & x = 0,1,2,\ldots,n \\ 0 & x \neq 0,1,2,\ldots,n \end{cases}$$

$$E(X) = np \quad V(X) = npq \cdot \frac{N-n}{N-1} \quad p = \frac{R}{N} \quad q = 1 - p = \frac{N-R}{N}$$

Observações

- Veja que se a amostra for retirada sucessivamente (ordenada) com reposição, resulta numa Distribuição Binomial com parâmetros n e p=R/N, pois a cada retirada, a probabilidade de se retirar um objeto do Tipo 1 permanece a mesma
- A amostra com reposição tem a propriedade de tornar a população (objetos na urna) infinita assim, com n fixo, se N tende a infinito a expressão $\frac{N-n}{N-1}$ tende a 1 e $V(X)$ passa a ser npq, que é a variância da $B(n,p)$

EXEMPLO 2.2.4[1]

Uma urna contém 15 bolas, sendo 10 brancas e 5 vermelhas. Uma amostra de extensão 4, não ordenada e sem reposição, é retirada ao acaso da urna. Seja X a variável aleatória que dá o número de bolas vermelhas na amostra. Determine a densidade f de X. Calcule $P(X=3)$, $E(X)$, $V(X)$ e σ.

$$N=15 \quad R=5 \quad n=4$$

$$f(x) = \begin{cases} \dfrac{\dbinom{10}{x}\dbinom{5}{4-x}}{\dbinom{15}{4}} & x = 0,1,2,3,4 \\ 0 & x \neq 0,1,2,3,4 \end{cases} \qquad P(X=3) = \frac{\dbinom{10}{3}\dbinom{5}{1}}{\dbinom{15}{4}} = \frac{40}{91}$$

$$E(X) = 4 \cdot \frac{5}{15} = \frac{1}{3}$$

$$V(X) = 4 \cdot \frac{5}{15} \cdot \frac{10}{15} \cdot \frac{15-4}{15-1} = \frac{8}{9} \cdot \frac{11}{14} = \frac{88}{126} = \frac{44}{63} \qquad \sigma_X = \sqrt{\frac{44}{63}}$$

5 Poisson – $Poisson(\lambda)$

Seja X uma variável aleatória discreta assumindo os valores $0,1,2,...,n,...$, se sua densidade de probabilidade for dada por:

$$f(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!} & , \ x = 0,1,2,\ldots,n,\ldots \\ 0 & , \ x \neq 0,1,2,\ldots,n,\ldots \end{cases}$$

$$E(X) = \lambda \qquad V(X) = \lambda$$

Então X é dita ter distribuição de Poisson com parâmetro λ>0.

EXEMPLO 2.2.5[1]

A média anual de acidentes em um determinado cruzamento no horário de pico é igual a 3. Se essa quantidade X tem distribuição de Poisson, qual a probabilidade de, num determinado ano, acontecer pelo menos um acidente no horário de pico ?

$$\lambda = 3 \qquad f(x) = \begin{cases} \dfrac{e^{-3} 3^x}{x!} & , \ x = 0,1,2,\ldots,n,\ldots \\ 0 & , \ x \neq 0,1,2,\ldots,n,\ldots \end{cases}$$

$$E(X) = 3 \qquad V(X) = 3$$

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{e^{-3} 3^0}{0!} = 1 - e^{-3} = 1 - 0{,}0498 = 0{,}9502$$

5.1 Binomial x Poisson

Suponha uma variável aleatória $B(n,p)$, com n grande e p pequeno. Por exemplo, n=1.000 e p=0,0001. A probabilidade de que X seja igual a 5 será dada por:

$$f(x) = \begin{cases} \binom{1000}{x} 0{,}0001^x 0{,}9999^{1000-x} & , \ x = 0,1,2,\ldots,1000 \\ 0 & , \ x \neq 0,1,2,\ldots,1000 \end{cases}$$

$$P(X = 5) = f(5) = \binom{1000}{5} 0{,}0001^5 0{,}9999^{995}$$

Observa-se que o cálculo das probabilidades nesta situação torna-se bastante difícil, que pode ser contornado aproximando-se a Distribuição Binomial pela de Poisson, ou seja, quando n é grande e p pequeno, as probabilidades da Binomial, $B(n,p)$, correspondente podem ser aproximadas por uma distribuição de Poisson com parâmetro $\lambda=np$:

$$f(x)=\begin{cases}\dfrac{(np)^x e^{-np}}{x!} & , \ x=0,1,2,\ldots,n,\ldots\\ 0 & , \ x\neq 0,1,2,\ldots,n,\ldots\end{cases}$$

$$E(X)=np \qquad V(X)=np$$

A solução do problema acima pode, então, ser aproximada por:

$$\lambda=np=1000\times 0{,}0001=0{,}1$$

$$P(X=5)=f(5)=\frac{0{,}1^5 e^{-0{,}1}}{5!}=\frac{0{,}00001\times 0{,}90484}{120}=0{,}0000000754$$

Observações

- Pelo fato de que a aproximação da Binomial pela Poisson é possível quando p é pequeno e n grande, a Distribuição de Poisson também é conhecida como a distribuição dos eventos raros, ou seja, o evento em questão tem probabilidade (p) pequena de ocorrer, sendo necessária uma grande quantidade de repetições do experimento (n) para que ele ocorra.
- Uma boa aproximação é obtida quando $n\geq 30$ e np ou $nq<5$.
- Veja que se p (probabilidade de sucesso) for grande, a probabilidade de se obter uma determinada quantidade de sucessos (x) deve ser calculada pela equivalente quantidade de fracassos (x sucessos $\equiv$ n-x fracassos), com λ passando a ser nq.

EXEMPLO 2.2.5.1[1]

Ao decolar de um porta-aviões, determinado tipo de avião tem probabilidade muito pequena, p=0,0002, de se perder por queda no mar. Qual a probabilidade de dois ou mais acidentes dessa natureza em n=500 decolagens?

$$\lambda=np=500\times 0{,}0002=0{,}1 \qquad P(X=x)=\frac{0{,}1^x e^{-0{,}1}}{x!} \ , \ x=0,1,2,\ldots$$

$$P(X\geq 2)=1-P(X<2)=1-\left[P(X=0)+P(X=1)\right]=$$

$$=1-\left(\frac{0{,}1^0 e^{-0{,}1}}{0!}+\frac{0{,}1^1 e^{-0{,}1}}{1!}\right)=0{,}0045$$

EXEMPLO 2.2.5.1[2]

A probabilidade de que uma ligação telefônica seja bem sucedida é 0,99. Em 1000 tentativas de ligações, qual a probabilidade de que mais de 997 sejam bem sucedidas ?

$$X \equiv \text{sucessos} \Rightarrow Y = n - X \equiv \text{fracassos}$$

$$p = 0{,}99; q = 0{,}01 \Rightarrow \lambda = nq = 1000 \times 0{,}01 = 10 \qquad P(Y = y) = \frac{10^x e^{-10}}{x!} \;, y = 0{,}1{,}2{,}\ldots$$

$$P(X > 997) = P(X = 998) + P(X = 999) + P(X = 1000) = P(Y = 2) + P(Y = 1) +$$

$$+ P(Y = 0) = \left(\frac{10^2 e^{-10}}{2!} + \frac{10^1 e^{-10}}{1!} + \frac{10^0 e^{-10}}{0!} \right) = 0{,}002769$$

5.2 Processo de Poisson

Muitos fenômenos podem ser vistos como uma quantidade de ocorrências de um evento dentro de um determinado intervalo, ou espaço, contínuo. Se o intervalo for o tempo, podemos ter, por exemplo, a quantidade de chamadas que chegam a determinada central telefônica, a quantidade de acidentes de trânsito em determinado cruzamento, etc. Se o espaço contínuo for o volume de água em um reservatório de uma cidade, as ocorrências podem ser a quantidade de bactérias dentro deste volume. Podemos ter também a quantidade de falhas numa peça de um cabo elétrico, aqui o espaço contínuo é o comprimento da peça. Um fenômeno com as mesmas características dos descritos acima é chamado de Processo de Poisson, desde que satisfaça às seguintes condições:

1) Se X_t é a variável aleatória que dá a quantidade de ocorrências de um evento no intervalo (espaço) contínuo t, e $X_{\Delta t}$ é a variável que dá a quantidade de ocorrências em um intervalo pequeno Δt de t, então a probabilidade de x ocorrências no intervalo Δt é diretamente proporcional a Δt, ou seja, $P(X_{\Delta t}=x)=\alpha\Delta t$
2) A probabilidade de duas ou mais ocorrências em um mesmo intervalo pequeno Δt de t é desprezível, ou seja, $P(X_{\Delta t}\geq 2)\sim 0$
3) Se Δt_1 e Δt_2 forem dois intervalos pequenos de t não superpostos, então a ocorrência ou a não ocorrência de um evento em Δt_1 não exercerão influência sobre a ocorrência ou não ocorrência de um evento em Δt_2, ou seja, as variáveis $X_{\Delta t_1}$ e $X_{\Delta t_2}$ são independentes.

Assim, a probabilidade de ocorrência de x eventos em um intervalo (espaço) contínuo de comprimento t é dada por:

$$P(X_t = x) = \frac{(\alpha t)^x e^{-\alpha t}}{x!}, x = 0,1,2,\ldots$$

que é a Distribuição de Poisson com parâmetro $\lambda=\alpha t$, onde α é a TAXA DE OCORRÊNCIA por unidade do espaço considerado.

EXEMPLO 2.2.5.2[1]

Uma adutora apresenta vazamentos a uma taxa de 2 por 100 km. Considerando que a quantidade de vazamentos segue um Processo de Poisson, qual a probabilidade de ocorrer no máximo um vazamento num trecho de 500 km ?

$$\alpha = \frac{2}{100} \Rightarrow \alpha t = \frac{2}{100} \cdot 500 = 10$$

$$P(X \le 1) = P(X = 0) + P(X = 1) = \frac{10^0 e^{-10}}{0!} + \frac{10^1 e^{-10}}{1!} =$$

$$= e^{-10} + 10e^{-10} = 11e^{-10} = 0{,}000499$$

EXEMPLO 2.2.5.2[2]

Determinado sistema apresenta falhas a uma taxa de 5 por hora. Se o sistema for “ligado” às $7{:}00$, qual a probabilidade de ocorram 5 falhas até às $9{:}30$, sendo apenas 1 falha até às $7{:}40$?

Fazendo X_1≡quantidade de falhas das $7{:}00$ às $7{:}40$ (40 minutos) e X_2≡quantidade de falhas das $7{:}40$ às $9{:}30$ $(110$ minutos), deseja-se calcular $P(X_1$=1 e X_2=4), como X_1e X_2 são independentes (condição 3 do Processo de Poisson), temos, trabalhando com minutos:

$$\alpha = \frac{5}{60} \Rightarrow \lambda_1 = \frac{5}{60} \times 40 = \frac{10}{3}; \lambda_2 = \frac{5}{60} \times 110 = \frac{55}{6}$$

$$P(X_1 = 1) = \frac{(10/3)^1 e^{-10/3}}{1!} = 0{,}1189 \qquad P(X_2 = 4) = \frac{(55/6)^4 e^{-55/6}}{4!} = 0{,}0307$$

$$P(X_1 = 1 \text{ e } X_2 = 4) = P(X_1 = 1) \times P(X_2 = 4) = 0{,}1189 \times 0{,}0307 = 0{,}00365$$

Capítulo 3

Algumas Variáveis Aleatórias Contínuas

1 Normal – $N(\mu,\sigma^2)$

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

$$E(X) = \mu \qquad V(X) = \sigma^2$$

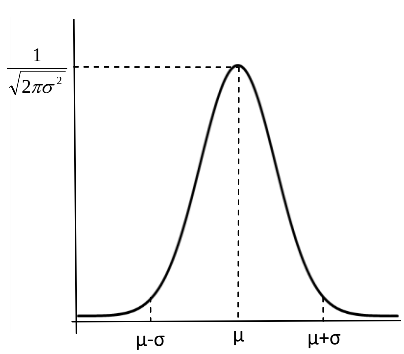
A Função de Distribuição Acumulada de Probabilidade de X, $F(x)$, não tem uma expressão analítica definida, mas seus valores podem ser obtidos a partir da Distribuição Normal Padrão, denotada por Z, cuja Função Distribuição de Probabilidade Acumulada, denotada por $\Phi(z)$, é aproximada utilizando-se a expansão de Taylor pra a função erro. A Distribuição Normal tem as seguintes características:

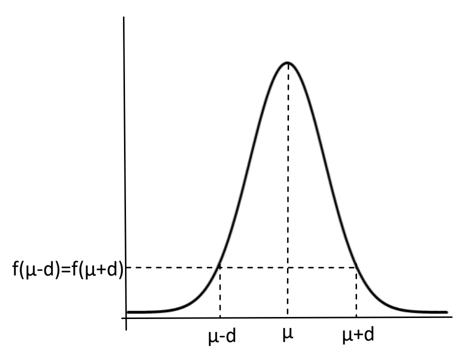
Desafio

Determine o máximo de $f(x)$ e seus pontos de inflexão

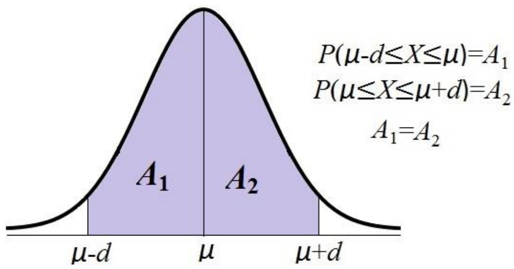
1) Os limites de $f(x)$ quando x tende a $+\infty$ e $-\infty$ são iguais a zero.
2) $f(x)$ é simétrica em torno de μ, ou seja, $f(\mu + d)$ = $f(\mu - d)$
3) O valor máximo de $f(x)$ ocorre em $x = \mu$
4) Os pontos de inflexão de $f(x)$ ocorrem em $x_1 = \mu - \sigma$ e $x_2 = \mu + \sigma$

Os gráficos seguintes ilustram estas propriedades.



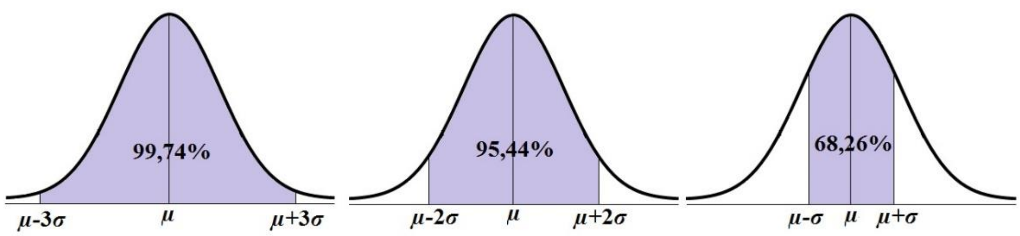


A característica 2 garante que as áreas sob $f(x)$ entre μ-d e μ+d são iguais, ou seja, em termos de probabilidade, $P(\mu-d \leq X \leq \mu) = P(\mu \leq X \leq \mu+d)$. A figura seguinte ilustra este resultado.



Além dessas características, a variabilidade da Distribuição Normal garante o seguinte:

1) 99,74% dos elementos da distribuição estão entre $\mu - 3\sigma$ e $\mu + 3\sigma$
2) 95,44% dos elementos estão entre $\mu - 2\sigma$ e $\mu + 2\sigma$
3) 68,26% dos elementos estão entre $\mu - \sigma$ e $\mu + \sigma$



EXEMPLO 2.3.1[1]

O comprimento, em milímetros de determinado tipo de objeto tem distribuição normal com média 5*mm* e desvio padrão 1*mm*, então podemos dizer que, dos objetos produzidos:

- 99,74% deles têm comprimento entre 5-3x1=2*mm* e 5+3x1=8*mm*;
- 95,44% deles têm comprimento entre 5-2x1=3*mm* e 5+2x1=7*mm*;
- 68,26% deles têm comprimento *entre* 5-1=4*mm* e 5+1=6*m*m.

1.1 Normal Padrão (reduzida) – $N(0,1)$

Desafio

Mostre que $E(Z)=0$ e $V(Z)=1$ (aplique as propriedades da Esperança e da Variância)

Se X tem Distribuição Normal $N(\mu,\sigma^2)$, então $Z = \dfrac{X-\mu}{\sigma}$ tem distribuição Normal com média 0 e variância 1, ou seja $Z \sim N(0,1)$.

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad -\infty < z < +\infty \qquad F(z) = P(Z \le z) = \Phi(z) = \int_{-\infty}^{z} f(z)dz$$

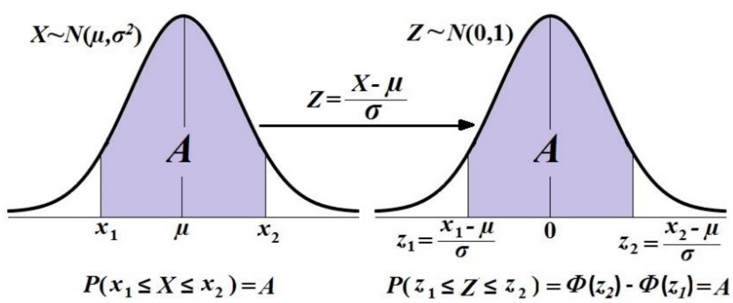
$$E(Z) = 0 \qquad V(Z) = 1$$

Alguns valores de $\Phi(z)$, para $z \geq 0$, estão tabelados no Anexo 1.

As probabilidades de uma variável aleatória X com distribuição $N(\mu,\sigma^2)$ podem ser determinadas pela transformação de X em Z:

$$P(x_1 \le X \le x_2) = P\left(\frac{x_1-\mu}{\sigma} \le Z \le \frac{x_2-\mu}{\sigma}\right) = P(z_1 \le Z \le z_2) = \Phi(z_2) - \Phi(z_1)$$

Graficamente:

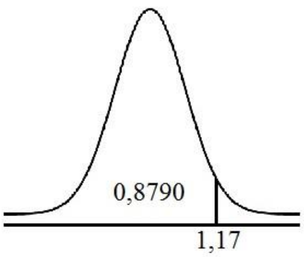


Os valores de $\Phi(z)$ estão disponíveis na tabela do Anexo 1.

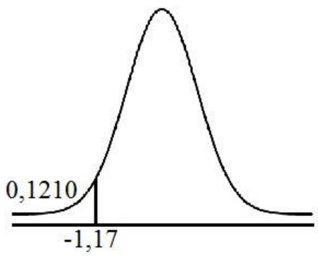
Veja como utilizar a tabela:

- No corpo da tabela estão os valores de $\Phi(z)$, ou seja $P(Z \leq z)$ para $z \geq 0$
- Na primeira coluna (coluna indicadora) estão os valores de z com uma casa decimal
- Na primeira linha (cabeçalho) estão os valores da segunda casa decimal de z

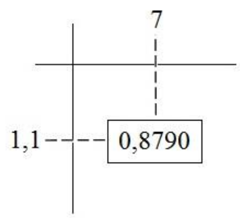
- Para encontrarmos a probabilidade para um determinado z, procuramos incialmente na coluna indicadora seu valor com uma casa decimal, em seguida procuramos no cabeçalho o correspondente à sua segunda casa decimal. A probabilidade desejada é encontrada no cruzamento entre a linha e coluna correspondentes
- Para valores negativos de z, usa-se a simetria da Distribuição Normal e a propriedade dos eventos complementares: $P(Z \leq -z) = 1 - P(Z \leq z)$



$P(Z \leq 1{,}17) = 0{,}8790$



$P(Z \leq -1{,}17) = 1 - 0{,}8790 = 0{,}1210$



EXEMPLO 2.3.1.1[1]

O processo de fabricação de determinado objeto apresenta variações de tal forma que seu comprimento final é uma variável aleatória com distribuição Normal com média $\mu = 10\ cm$ e variância $\sigma^2 = 4\ cm^2$. Qual a probabilidade de que um objeto escolhido ao acaso tenha comprimento entre $6\ cm$ e $12\ cm$? Maior do que $11\ cm$? menor do que $5\ cm$?

$$P(6 \leq X \leq 12) = P\left(\frac{6\text{-}10}{2} \leq Z \leq \frac{12-10}{2}\right) = P(-2 \leq Z \leq 1) = \Phi(1) - \Phi(-2) =$$

$$= \Phi(1) \text{-} [1\text{-}\Phi(2)] = 0{,}8413\text{-}(1\text{-}0{,}9772) = 0{,}8413\text{-}0{,}0228 = 0{,}8185$$

$$P(X > 11) = P\left(Z > \frac{11\text{-}10}{2}\right) = P(Z > 0{,}5) = 1 - P(Z \leq 0{,}5) = 1 - \Phi(0{,}5) =$$

$$= 1 - 0{,}6915 = 0{,}3085$$

$$P(X < 5) = P\left(Z < \frac{5\text{-}10}{2}\right) = P(Z < 2{,}5) = P(Z > 2{,}5) = 1 - P(Z \leq 2{,}5) = 1 - \Phi(2{,}5) =$$

$$= 1 - 0{,}9938 = 0{,}0062$$

1.2 BINOMIAL x NORMAL

Para uma distribuição $B(n,p)$, quando n é grande e p próximo a $0{,}5$, suas probabilidades podem ser aproximadas por uma distribuição Normal com média $\mu=np$ e variância $\sigma^2=npq$. Ou seja:

$$\lim_{\substack{n\to\infty \\ p\to\frac{1}{2}}} B(n,p) = N(np, npq)$$

EXEMPLO 2.3.1.2[1]

No lançamento de uma moeda 500 vezes, qual a probabilidade de que sejam obtidas pelo menos 270 caras ?

A variável original é $B\left(500,\frac{1}{2}\right)$, deseja-se calcular $P(X\geq 270)=1-P(X<270)=1-P(X\leq 269)$, aproximando-se pela distribuição normal tem-se:

$$n=500 \quad p=\frac{1}{2}$$

$$np = 500\cdot\frac{1}{2}=250 \quad npq=500\cdot\frac{1}{2}\cdot\frac{1}{2}=125$$

$$B\left(500,\frac{1}{2}\right)\to N(250{,}125)$$

$$P(X\leq 269)=\Phi\left(\frac{269-250}{\sqrt{125}}\right)=\Phi\left(-\frac{19}{\sqrt{125}}\right)=1-\Phi\left(\frac{19}{\sqrt{125}}\right)=$$

$$=1-\Phi(1{,}70)=1-0{,}9554=0{,}0446$$

2 Uniforme – $U(a,b)$

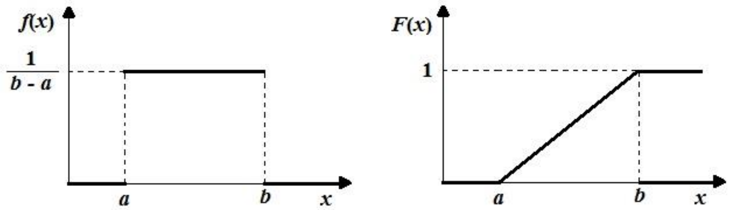
Desafio

Mostre que $f(x)$ satizfaz às duas condições da definição de Função Densidade de Probabilidade. Determine $F(x)$, $E(X)$ e $V(X)$

$$f(x)=\begin{cases}\dfrac{1}{b-a} & x\in[a,b]\\ 0 & x\notin[a,b]\end{cases} \qquad F(x)=\frac{x-a}{b-a}$$

$$E(X)=\frac{a+b}{2} \qquad V(X)=\frac{(b-a)^2}{12}$$

Gráficos de $f(x)$ e $F(x)$:



EXEMPLO 2.3.2[1]

O peso de determinado objeto é uniformemente distribuído em [3,5]. Determine $f(x)$ e $F(x)$. Calcule $P(3{,}5 \leq X \leq 4{,}5)$, $E(X)$, $V(X)$ e σ.

$$f(x) = \begin{cases} \dfrac{1}{5\text{-}3} & x \in [3{,}5] \\ 0 & x \notin [3{,}5] \end{cases} \Rightarrow f(x) = \begin{cases} \dfrac{1}{2} & x \in [3{,}5] \\ 0 & x \notin [3{,}5] \end{cases}$$

$$F(x) = \frac{x-3}{5-3} \Rightarrow F(x) = \begin{cases} 0, x < 3 \\ \dfrac{x-3}{2}, x \in [3{,}5] \\ 1, x > 5 \end{cases}$$

$$P(3{,}5 \leq X \leq 4{,}5) = F(4{,}5) - F(3{,}5) = \frac{4{,}5-3}{2} - \frac{3{,}5-3}{2} = \frac{1{,}5-0{,}5}{2} = \frac{1}{2} = 0{,}5$$

$$E(X) = \frac{3+5}{2} = \frac{8}{2} = 4 \quad V(X) = \frac{(5\text{-}3)^2}{12} = \frac{4}{12} = \frac{1}{3} \qquad \sigma = \sqrt{\frac{1}{3}} = \frac{1}{\sqrt{3}}$$

3 Exponencial – $Exp(\lambda)$

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & x > 0\ ,\ \lambda > 0 \\ 0 & x \leq 0 \end{cases}$$

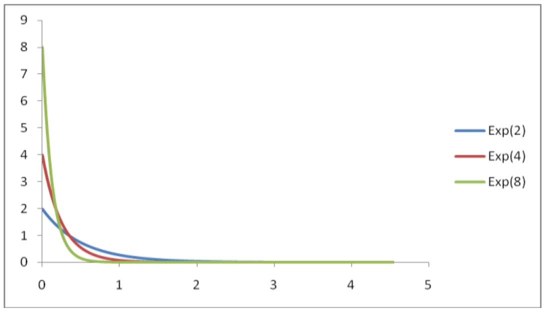
$$F(x) = 1 - e^{-\lambda x}$$

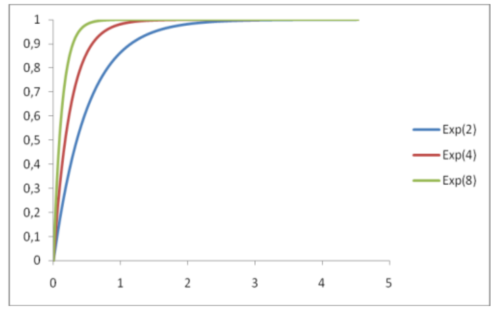
$$E(X) = \frac{1}{\lambda} \qquad V(X) = \frac{1}{\lambda^2}$$

Desafio

Mostre que $f(x)$ satizfaz às duas condições da definição de Função Densidade de Probabilidade. Determine $F(x)$, $E(X)$ e $V(X)$

GRÁFICOS DE $f(x)$ e $F(x)$ para λ=2; λ=4 e λ=8:





EXEMPLO 2.3.3[1]

Suponha que o tempo (min) das chamadas telefônicas de uma determinada cidade tenha distribuição exponencial com parâmetro λ=0,5. Determine $f(x)$ e $F(x)$. Calcule a probabilidade de que uma ligação dure entre 1 min e 3 min. Calcule $E(X)$ e $V(X)$.

$$f(x) = \begin{cases} 0{,}5e^{-0{,}5x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

$$F(x) = 1 - e^{-\lambda x} = 1 - e^{-0{,}5x}$$

$$P(1 \leq X \leq 3) = F(3)\text{-}F(1) = \left(1\text{-}e^{-0{,}5\cdot 3}\right) - \left(1 - e^{-0{,}5\cdot 1}\right) = e^{-0{,}5} - e^{-1{,}5} =$$

$$= 0{,}6065 - 0{,}2231 = 0{,}3834 = 38{,}34\%$$

$$E(X) = \frac{1}{0{,}5} = 2\min \text{ (tempo médio de conversação)} \quad V(X) = \frac{1}{0{,}5^2} = 4\min^2$$

3.1 Relação Exponencial x Poisson

Se X_t é um Processo de Poisson, já vimos que a probabilidade de ocorrência de x eventos em um intervalo (espaço) contínuo de tamanho t é dada pela Distribuição de Poisson:

$P(X_t = x) = \dfrac{(\alpha t)^x e^{-\alpha t}}{x!}$, $x = 0,1,2,...$, onde α é a taxa de ocorrências por unidade de t.

Seja T a variável aleatória que dá a "distância" a partir de um ponto qualquer até a primeira ocorrência do evento.

Para facilitar o entendimento, vamos considerar que T é o tempo de espera decorrido até a primeira ocorrência de um evento a partir de um instante qualquer. Assim, podemos denotar a probabilidade de que este tempo de espera seja superior a t por $P(T>t)$ e a probabilidade de que seja inferior ou igual a t por $P(T \leq t)=1-P(T>t)$.

É fácil perceber que a afirmação de que o tempo de espera até a primeira ocorrência do evento é superior a t é equivalente a afirmar que não existe qualquer ocorrência do evento no intervalo t, ou seja, $P(T>t)=P(X_t=0)$, então:

$P(T \leq t)=1-P(T>t)=1-P(X_t=0)=1-\dfrac{(\alpha t)^0 e^{-\alpha t}}{0!}=1-e^{-\alpha t}$ que é a Função de Distribuição Acumulada de uma variável aleatória Exponencial com parâmetro α, cuja derivada é igual a $\alpha e^{-\alpha t}$, sua Função Densidade de Probabilidade.

EXEMPLO 2.3.3.1[1]

As mensagens destinadas a um determinado centro de atendimento chegam a uma taxa de 5 por minuto. Qual a probabilidade de que o tempo de espera para o recebimento de uma mensagem seja inferior a 3 *min* ?

$$\alpha = \frac{1}{5} \text{ (mensagens por minuto)} \quad P(T \leq 3)=1-e^{-\frac{1}{5}\cdot 3}=1-e^{-\frac{3}{5}}=1-0{,}5488=0{,}4512$$

EXEMPLO 2.3.3.1[2]

Uma adutora apresenta vazamentos a uma taxa de 2 por 100 *km*. Considerando que a quantidade de vazamentos segue um Processo de Poisson, qual a probabilidade de percorrermos no mínimo 500 *Km* dessa adutora até encontrarmos um vazamento ?

$$\alpha = \frac{2}{100} = 0{,}02 \quad \text{(vazamentos por km)}$$
$$P(T > 500) = e^{-0{,}02\times 500} = e^{-10} = 0{,}0000454$$

Atividades de avaliação

1) Uma variável aleatória X tem a seguinte distribuição de probabilidade:

x	$P(X=x)$
1	3/a
2	5/a
3	7/a
4	9/a
5	11/a

a) determine o valor de a
b) determine $f(x)$ e $F(x)$
c) esboce os gráficos de $f(x)$ e $F(x)$
d) calcule $E(X)$ e $V(X)$

2) Uma urna contém 1 bola marcada com o nº 1, 2 com o nº 2, e assim sucessivamente até n bolas marcadas com o nº n. Uma bola é retirada dessa urna. Seja X a variável aleatória que associa a bola retirada ao seu número. Determine a função densidade de probabilidade $f(x)$ e a função de distribuição acumulada de probabilidade $F(x)$ e esboce seus gráficos. Determine $E(X)$ e $V(X)$. Para $n=6$, calcule $P(X=3)$, $P(X\leq 2)$, $P(2\leq X\leq 4)$, $E(X)$ e $V(X)$.
3) Suponha que uma urna contenha n bolas numeradas de 1 a n. Duas bolas são retiradas sucessivamente, com reposição, da urna. Se X=máximo(x_1,x_2), determine $f(x)$, $F(x)$, $E(X)$ e $V(X)$.
4) [XAVIER] Seja (a_{ij}) uma matriz real (todos os seus termos são números reais), com n linhas e m colunas. Dois jogadores, A e B, participam de um jogo que consiste no seguinte: o jogador A seleciona uma linha da matriz escolhendo ao acaso um número de 1 a n enquanto B seleciona uma coluna escolhendo ao acaso um número de 1 a m. Se i e j forem as escolhas de A e B, respectivamente, B pagará a A uma quantia equivalente a a_{ij} se $a_{ij}>0$, caso $a_{ij}<0$ então será A que pagará a B o equivalente a a_{ij}, se $a_{ij}=0$ não haverá pagamento. Seja X a variável aleatória que dá o ganho do jogador A, que será negativo se B ganhar. Considerando que as escolhas feitas por A e B são independentes, determine o ganho esperado do jogador A, nos seguintes casos:
   a) Se $n=m=2$, p é a probabilidade de A selecionar $i=1$, p' a probabilidade de B selecionar $j=1$ e $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$
   b) Se m e n são quaisquer, sendo p_i a probabilidade de A escolher o número i e p'_j a probabilidade de B escolher o número j
5) [XAVIER] Considerando a situação do item (a) do problema 4, verificar que, no caso de $p'>1/2$, tem-se: (a) $E(X)=0$ para $p=1/2$; (b) $E(X)$ assume valor máximo $2p'-1$, para $p=1$

6) Num jogo que consiste no lançamento de um dado 3 vezes, um apostador paga R$ $2{,}00$ para entrar no jogo mas ganha R$ $3{,}00$ por cada nº 6 que aparece. Qual o ganho esperado desse apostador?
7) [MEYER] Um exame falso-verdadeiro é constituído de 10 questões. Qual a probabilidade de se obter 70% ou mais de respostas certas ? Exatamente 70% ? Determine o valor esperado e a variância da quantidade de questões certas.
8) [MEYER] Um exame é constituído de 20 questões, cada questão com 5 itens, sendo 4 errados e 1 certo. Qual a probabilidade de se acertar 60% ou mais das questões ? Determine o valor esperado e a variância da quantidade de questões certas.
9) Qual a probabilidade de se encontrar 2 ou menos peças defeituosas em um pacote de 5 peças retiradas, sem reposição, de um lote contendo 6 peças boas e 4 com defeito ? Determine o valor esperado e a variância da quantidade de peças defeituosas para pacotes de 5 peças.
10) [adaptado de MEYER] A probabilidade de um acesso bem sucedido a determinada página na Internet é $0{,}8$. Supondo que tentativas de acesso sejam feitas até que tenha ocorrido 1 acesso bem sucedido:
    a) Qual a probabilidade de que exatamente 6 tentativas sejam necessárias ?
    b) Qual a probabilidade de que menos de 6 tentativas sejam necessárias ?
    c) Qual a probabilidade de que mais de 3 tentativas sejam necessárias ?
    d) Qual a probabilidade de que mais de 1 tentativa sejam necessárias ?
    e) Qual a quantidade esperada de tentativas, e a variância ?
11) Resolva o problema anterior para que tenham ocorrido 3 acessos bem sucedidos.
12) [adaptado de MEYER] Seja X uma variável aleatória com função densidade de probabilidade dada por:

$$f(x) = \begin{cases} ax & 0 \le x \le 1 \\ a & 1 \le x \le 2 \\ -ax + 3a & 2 \le x \le 3 \\ 0 & x \notin [0,3] \end{cases}$$

    a) Determine o valor de a para que $f(x)$ seja realmente uma função densidade de probabilidade
    b) Esboce o gráfico de $f(x)$
    c) Determine $F(x)$ e esboce seu gráfico
    d) Calcule $P(1/2 \le X \le 5/2)$, $P(1/2 \le X < 5/2)$, $P(1/2 < X \le 5/2)$ e $P(1/2 < X < 5/2)$
    e) Calcule $E(X)$ e $V(X)$

13) Seja X uma variável aleatória Binomial, $B(4,1/2)$
    a) Determine $f(x)$ e esboce seu gráfico
    b) Determine $F(x)$ e esboce o seu gráfico
    c) Calcule $P(1<X\le 2)$; $P(1\le X\le 2)$; $P(1<X<2)$ e $P(1\le X<2)$
    d) Calcule $E(X)$ e $V(X)$
14) O tempo de duração, em mil horas, de determinado equipamento é uma variável aleatória com função densidade de probabilidade dada por:

$$f(x)=\begin{cases} k(x-1) & , \ 1\leq x<2 \\ k(3-x) & , \ 2\leq x\leq 3 \\ 0 & , \ x\notin[1,3] \end{cases}$$

a) Determine o valor de k, $F(x)$, $E(X)$ e $V(X)$. Esboce os gráficos de $f(x)$ e $F(x)$.
b) O lucro líquido (do fabicante) por equipamento é dado por $L = C_1 - C_2 X$. Qual o lucro líquido esperado do fabricante, por equipamento ?
c) Determinado comprador ganha, por equipamento, R\$5,00 se o mesmo durar entre 1,5 e 2,5 mil horas e perde R\$2,00 caso contrário. Qual o ganho esperado do comprador, por equipamento ?

15) [MEYER] Na produção de petróleo a temperatura de distribuição T (ºC) é decisiva na determinação da qualidade do produto final. Suponha que T seja uma variável aleatória uniformemente distribuída em [150,300]. Admita-se que produzir um galão de petróleo custa C_1 dólares. Se o óleo for destilado à uma temperatura menor do que 200^oC o produto é conhecido como NAFTA e se vende por C_2 dólares por galão. Se o óleo for destilado a uma temperatura maior, do que 200^oC o produto é denominado ÓLEO REFINADO DESTILADO e se vende por C_3 dólares por galão. Determine o lucro líquido esperado por galão.

16) [adaptado de XAVIER] A probabilidade de que um pacote de mensagens enviado a partir de certo dispositivo não chegue ao seu destino é p. Suponha que duas mensagens devem ser enviadas ao custo de c_1 e c_2 respectivamente. Deve-se decidir em enviar as mensagens em um único pacote ou em pacotes separados. Três critérios de decisão podem ser adotados: (I) escolher a forma de envio para a qual o valor esperado da perda seja mínimo; (II) escolher a forma de envio que tenha maior probabilidade de chegada de ambas as mensagens e (III) escolher a forma de envio para a qual a probabilidade de chegada de pelo menos uma das mensagens é maior. Qual deve ser a decisão, para cada um dos critérios?

17) O diâmetro de um cabo elétrico é normalmente distribuído com média 0,8 e variância 0,0004.
a) Qual a probabilidade de que o diâmetro de um cabo escolhido ao acaso ultrapasse 0,81?
b) Se o diâmetro do cabo diferir da média em mais de 0,02 ele é considerado defeituoso. Qual a probabilidade de se encontrar um cabo defeituoso?
c) Se o cabo for defeituoso, o fabricante perde R\$0,01. Se seu dâmetro estiver entre a média e 0,78, o fabricante ganha R\$0,05 e se estiver entre a média e 0,82 ele ganha R\$0,10. Qual o ganho esperado por cabo?

d) Qual deve ser o diâmetro de um eletroduto de modo que 95% dos cabos passem pelo mesmo ?
e) Num lote de 5 cabos, qual a probabilidade de que exatamente 3 sejam defeituosos ?

18) A altura, em metros, das pessoas que frequentam determinado restaurante tem distribuição normal $N(1{,}60;0{,}04)$. Qual deve ser a altura da porta do restaurante para que 90% das pessoas não tenham que se abaixar para ultrapassá-la?
19) A probabilidade de que um certo tipo de peça seja defeituosa é $0{,}40$. Num lote de 10 peças, qual a probabilidade de que no máximo uma tenha defeito ? (utilize as distribuições Binomial e Normal e compare os resultados)
20) A duração das chamadas telefônicas de determinada central tem distribuição exponencial com parâmetro λ. Se a duração média dessas chamadas é de 3 min, qual a probabilidade de que em 5 chamadas pelo menos uma dure menos de $1{,}5$ min ?
21) Um número X é escolhido ao acaso no intervalo $[1{,}4]$. Calcular:
    a) A probabilidade de que o número escolhido esteja entre 2 e 3
    b) Entre $0{,}5$ e $2{,}5$
    c) seja exatamente o 2
    d) Calcule a esperança e a variância de X
22) [MEYER] A incidência anual de poliomielite, num determinado país, é de 3 casos por 100.000 pessoas. Supondo esse dado fidedigno, qual a probabilidade de ocorrerem no máximo 5 casos anuais, numa cidade com 60.000 habitantes?
23) [MEYER] Uma certa liga é formada pela união da mistura em fusão de dois metais. A liga resultante contém uma certa porcentagem X de chumbo que pode ser considerada uma variável aleatória com função densidade de probabilidade dada por: $f(x)=\begin{cases}\frac{3}{5}\cdot 10^{-5}\cdot x(100-x) & 0\le x\le 100\\ 0 & x\notin[0{,}100]\end{cases}$. Suponha que P, o lucro líquido obtido pela venda dessa liga (por libra) seja a seguinte função da porcentagem contida: $P=C_1+C_2X$. Calcule o lucro líquido esperado (por libra).
24) [MEYER] Suponha que um dispositivo eletrônico tenha uma duração de vida X (em unidades de 1.000 horas), a qual é considerada uma variável aleatória contínua com função densidade de probabilidade dada por $f(x)=\begin{cases}e^{-x} & x>0\\ 0 & x\le 0\end{cases}$. Suponha que o custo de fabricação de um desses dispositivos seja $\text{US\$}2{,}00$. O fabricante vende a peça por $\text{US\$}5{,}00$ mas garante o reembolso total se $X\le 0{,}9$. Qual será o lucro esperado, por peça fabricada ?

25) [XAVIER] Considere a implantação de uma central telefônica em um escritório com n telefones individuais e uma certa quantidade de linhas que permitam ligações externas. Se a probabilidade de que cada um desses telefones se encontre ocupando uma linha externa num momento qualquer for igual a p, qual o número mínimo de linhas externas que a central deve ter para se ter uma probabilidade não inferior a $0{,}95$ de haver, a qualquer momento, linha externa disponível para uma pessoa que necessite fazer uma ligação externa?

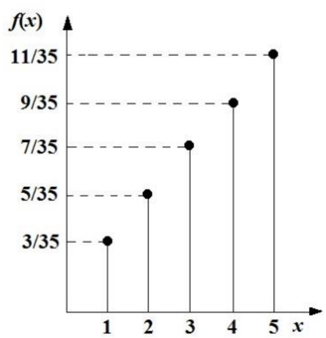
Solução das atividades

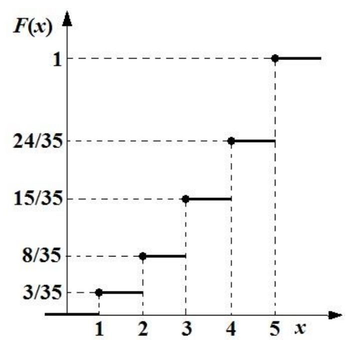
1)

$$a)\ \frac{3}{a}+\frac{5}{a}+\frac{7}{a}+\frac{9}{a}+\frac{11}{a}=1 \Rightarrow a=35$$

$$b)\ f(x)=\begin{cases}\dfrac{2x+1}{35}, x=1,2,3,4,5\\ 0, x\neq 1,2,3,4,5\end{cases} \qquad F(x)=\begin{cases}0, x<1\\ \displaystyle\sum_{x_i\le x}\dfrac{2x_i+1}{35}, 1\le x\le 5, x_i=1,2,3,4,5\\ 1, x>5\end{cases}$$

$c)$





$$d)\ E(X)=1\times\frac{3}{35}+2\times\frac{5}{35}+3\times\frac{7}{35}+4\times\frac{9}{35}+5\times\frac{11}{35}=\frac{125}{35}=\frac{25}{7}$$

$$E(X^2)=1^2\times\frac{3}{35}+2^2\times\frac{5}{35}+3^2\times\frac{7}{35}+4^2\times\frac{9}{35}+5^2\times\frac{11}{35}=\frac{505}{35}=\frac{101}{7}$$

$$V(X)=\frac{101}{7}-\left(\frac{25}{7}\right)^2=\frac{707-625}{49}=\frac{82}{49}$$

2)

$$n(S)=1+2+3+\ldots+n=\frac{n(n+1)}{2}$$

$$P(X=x)=\frac{x}{\frac{n(n+1)}{2}}=\frac{2x}{n(n+1)} \Rightarrow f(x)=\begin{cases}\dfrac{2x}{n(n+1)}, x=1,2,\ldots,n\\ 0, x\neq 1,2,\ldots,n\end{cases}$$

$$F(x)=\begin{cases}0, x<1\\ \displaystyle\sum_{x_i\leq x}\frac{2x_i}{n(n+1)}, 1\leq x\leq n\,,\, x_i=1,2,\ldots,n\\ 1, x>n\end{cases}$$

OS GRÁFICOS FICAM A CARGO DO ALUNO

$$E(X)=\sum_{i=1}^{n}x_iP(X=x_i)=\sum_{i=1}^{n}\left(x_i\times\frac{2x_i}{n(n+1)}\right)=\frac{2}{n(n+1)}\sum_{i=1}^{n}x_i^2=\frac{2}{n(n+1)}\sum_{k=1}^{n}k^2$$

$$=\frac{2}{n(n+1)}\times\frac{n(n+1)(2n+1)}{6}=\frac{2n+1}{3}$$

$$E(X^2)=\sum_{i=1}^{n}x_i^2P(X=x_i)=\sum_{i=1}^{n}\left(x_i^2\times\frac{2x_i}{n(n+1)}\right)=\frac{2}{n(n+1)}\sum_{i=1}^{n}x_i^3=\frac{2}{n(n+1)}\sum_{k=1}^{n}k^3$$

$$=\frac{2}{n(n+1)}\times\left(\frac{n(n+1)}{2}\right)^2=\frac{n(n+1)}{2}$$

$$V(X)=\frac{n(n+1)}{2}-\left(\frac{2n+1}{3}\right)^2=\frac{n(n+1)}{2}-\frac{(2n+1)^2}{9}=\frac{9n(n+1)-2(2n+1)^2}{18}$$

$$n=6\Rightarrow f(x)=\begin{cases}\dfrac{x}{21}, x=1,2,3,4,5,6\\ 0, x\neq 1,2,3,4,5,6\end{cases}$$

$$P(X=3)=\frac{3}{21}=\frac{1}{7}$$

$$P(X\leq 2)=P(X=1)+P(X=2)=\frac{1}{21}+\frac{2}{21}=\frac{3}{21}=\frac{1}{7}$$

$$P(2\leq X\leq 4)=P(X=2)+P(X=3)+P(X=4)=\frac{2}{21}+\frac{3}{21}+\frac{4}{21}=\frac{9}{21}=\frac{3}{7}$$

$$E(X)=\frac{2n+1}{3}=\frac{2\times 6+1}{3}=\frac{13}{3}$$

$$V(X)=\frac{9n(n+1)-2(2n+1)^2}{18}=\frac{9\times 6\times(6+1)-2\times(2\times 6+1)^2}{18}=\frac{40}{18}=\frac{20}{9}$$

3)

$$n(S)=A_n^2 \Rightarrow n(S)=n^2$$
$$A=\{(1,x);(2,x);\ldots;(x-1,x);(x,1);(x,2);\ldots;(x,x-1);(x,x)\} \Rightarrow n(A)=2(x-1)+1$$
$$=2x-1 \Rightarrow P(A)=P(X=x)=\frac{2x-1}{n^2}$$

$$f(x)=\begin{cases}\dfrac{2x-1}{n^2},\ x=1,2,3,\ldots,n\\ 0,\ x\neq 1,2,3,\ldots,n\end{cases}\qquad F(x)=\begin{cases}0,\ x<1\\ \displaystyle\sum_{x_i\prec x}\frac{2x_i-1}{n^2},1\leq x\leq n,\ x_i=1,2,\ldots,n\\ 1,\ x>n\end{cases}$$

$$E(X)=\sum_{i=1}^{n}\left(x_i\times\frac{2x_i-1}{n^2}\right)=\frac{1}{n^2}\left(2\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}x_i\right)=\frac{2}{n^2}\sum_{i=1}^{n}x_i^2-\frac{\frac{n(n+1)}{2}}{n^2}=$$
$$=\frac{2}{n^2}\sum_{i=1}^{n}x_i^2-\frac{n+1}{2n}=\frac{2}{n^2}\cdot\frac{n(n+1)(2n+1)}{6}-\frac{n+1}{2n}==\frac{(n+1)(2n+1)}{3n}-\frac{n+1}{2n}=$$
$$=\frac{2(n+1)(2n+1)-3(n+1)}{6n}=\frac{(n+1)(4n+2-3)}{6n}=\frac{(n+1)(4n-1)}{6n}$$
$$E(X^2)=\sum_{i=1}^{n}\left(x_i^2\times\frac{2x_i-1}{n^2}\right)=\frac{1}{n^2}\left(2\sum_{i=1}^{n}x_i^3-\sum_{i=1}^{n}x_i^2\right)=$$
$$=\frac{1}{n^2}\left[2\times\left(\frac{n(n+1)}{2}\right)^2-\frac{n(n+1)(2n+1)}{6}\right]=\frac{1}{n^2}\left[\frac{n^2(n+1)^2}{2}-\frac{n(n+1)(2n+1)}{6}\right]=$$
$$=\frac{(n+1)^2}{2}-\frac{(n+1)(2n+1)}{6n}=\frac{6n(n+1)^2-2(n+1)(2n+1)}{12n}$$

$$V(X)=\frac{6n(n+1)^2-2(n+1)(2n+1)}{12n}-\left(\frac{(n+1)(4n-1)}{6n}\right)^2$$
$$=\frac{3n\left[6n(n+1)^2-2(n+1)(2n+1)\right]-(n+1)^2(4n-1)^2}{36n^2}=$$
$$=\frac{18n^2(n+1)^2-6n(n+1)(2n+1)-(n+1)^2(4n-1)^2}{36n^2}=$$
$$=\frac{(n+1)^2[18n^2-(4n-1)^2]-6n(n+1)(2n+1)}{36n^2}$$

4)

$a)\ P(A\,\text{escolher}\,1) = p \Rightarrow P(A\,\text{escolher}\,2) = 1 - p$

$P(B\,\text{escolher}\,1) = p' \Rightarrow P(B\,\text{escolher}\,2) = 1 - p'$

$P(X = 1) = pp' + (1-p)(1-p')$

$P(X = -1) = (1-p)p' + p(1-p')$

$E(X) = 1 \times P(X = 1) + (-1) \times P(X = -1) = [pp' + (1-p)(1-p')] -$

$-[(1-p)p' + p(1-p')] = 4pp' - 2(p + p') + 1$

$b)\ P(A\,\text{escolher}\,i) = p_i$

$P(B\,\text{escolher}\,j) = p'_j$

$E(X) = \sum_{i=1}^{m} \sum_{j=1}^{n} p_i p'_j a_{ij}$

5)

$E(X) = p(4p' - 2) + (1 - 2p')$

$E(X) = ap + b, a = 4p' - 2 > 0 \text{ e } b = 1 - 2p' < 0, \text{pois } p' > \frac{1}{2}, \text{por hipótese}$

$i)\ E(X) = 0 \text{ para } p = -\frac{b}{a} = -\frac{1-2p'}{4p'-2} = \frac{1}{2}$

$ii)\ a > 0 \Rightarrow E(X)$ cresce com p e atinge seu máximo em $p = 1$ (valor máximo de p) $\Rightarrow$

$\Rightarrow \max(E(X)) = (4p' - 2) + (1 - 2p') = 2p' - 1$

6)

$X \equiv$ quantidade de números 6 no lançamento do dado 3 vezes

$X \sim B\left(3, \frac{1}{6}\right) \Rightarrow E(X) = 3 \times \frac{1}{6} = \frac{1}{2}$

$G \equiv$ ganho do jogador $\Rightarrow G = 3X - 2 \Rightarrow E(3X) - E(2) = E(G) = 3E(X) - 2 =$

$= 3 \times \frac{1}{2} - 2 = -0{,}5$

7)

DISTRIBUIÇÃO BINOMIAL COM PARÂMETROS n=10 E p=1/2

$$X \sim B\left(10, \frac{1}{2}\right)$$

$$f(x) = P(X = x) = \begin{cases} \binom{10}{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{10-x} & , x = 0,1,2,\ldots,10 \\ 0 & , x \neq 0,1,2,\ldots,10 \end{cases}$$

$$f(x) = \begin{cases} \binom{10}{x} \times \frac{1}{2^{10}} & , x = 0,1,2,\ldots,10 \\ 0 & , x \neq 0,1,2,\ldots,10 \end{cases}$$

$$P(X \geq 7) = \sum_{x=7}^{10} P(X = x) = \frac{1}{2^{10}} \times \left[\binom{10}{7} + \binom{10}{8} + \binom{10}{9} + \binom{10}{10}\right] =$$

$$= \frac{120 + 45 + 10 + 1}{2^{10}} = \frac{176}{1024} \approx 0{,}1719$$

$$E(X) = np = 10 \times \frac{1}{2} = 5 \qquad V(X) = npq = 5 \times \frac{1}{2} = 2{,}5$$

8)

DISTRIBUIÇÃO BINOMIAL COM PARÂMETROS n=20 E p=0,2=1/5

$$X \sim B\left(20, \frac{1}{5}\right) \qquad f(x) = P(X = x) = \begin{cases} \binom{20}{x}\left(\frac{1}{5}\right)^{x}\left(\frac{4}{5}\right)^{20-x} & , x = 0,1,2,\ldots,20 \\ 0, x \neq 0,1,2,\ldots,20 \end{cases}$$

$$P(X \geq 12) = \sum_{x=12}^{20} P(X = x) = \sum_{x=12}^{20} \binom{20}{x}\left(\frac{1}{5}\right)^{x}\left(\frac{4}{5}\right)^{20-x} \approx 0{,}000102$$

$$E(X) = np = 20 \times \frac{1}{5} = 4 \qquad V(X) = npq = 4 \times \frac{4}{5} = 3{,}2$$

9)

DISTRIBUIÇÃO HIPERGEOMÉTRICA COM PARÂMETROS N=10 E R=4

$$N=10 \quad R=4 \quad n=5 \qquad f(x)=\begin{cases} \dfrac{\binom{4}{x}\binom{6}{5-x}}{\binom{10}{5}} & x=0,1,2,3,4,5 \\ 0 & x \neq 0,1,2,3,4,5 \end{cases}$$

$$P(X \geq 2)=P(X=0)+P(X=1)+P(X=2)=\frac{\binom{4}{0}\binom{6}{5}}{\binom{10}{5}}+\frac{\binom{4}{1}\binom{6}{4}}{\binom{10}{5}}+\frac{\binom{4}{2}\binom{6}{3}}{\binom{10}{5}}$$

$$=\frac{1\times 6+4\times 15+6\times 20}{252}=\frac{186}{252}=\frac{31}{42}$$

$$n=5 \quad p=\frac{4}{10} \quad q=\frac{6}{10}$$

$$E(X)=np=5\times\frac{4}{10}=2$$

$$V(X)=npq\cdot\frac{N-n}{N-1}=5\times\frac{4}{10}\times\frac{6}{10}\times\frac{10\text{-}5}{10\text{-}1}=\frac{6}{9}=\frac{2}{3}$$

10)

DISTRIBUIÇÃO GEOMÉTRICA COM PARÂMETRO p=0,8

$$f(x)=\begin{cases} 0{,}8\times 0{,}2^{x-1} & x=1,2,3,\ldots \\ 0 & x \neq 1,2,3,\ldots \end{cases}$$

$$a)\ P(X=6)=f(6)=0{,}8\times 0{,}2^{5}=0{,}8\times 0{,}0032=0{,}000256$$

$$b)\ P(X<6)=f(1)+f(2)+f(3)+f(4)+f(5)=$$
$$=0{,}8\times\left[0{,}2^{0}+0{,}2^{1}+0{,}2^{2}+0{,}2^{3}+0{,}2^{4}\right]=0{,}8\times 1{,}2486=0{,}99968$$

$$c)\ P(X>3)=1-P(X\leq 3)=1-\left[f(1)+f(2)+f(3)\right]=$$
$$=1-0{,}8\times\left[0{,}2^{0}+0{,}2^{1}+0{,}2^{2}\right]=1-0{,}8\times 1{,}24=1-0{,}992=0{,}008$$

$$d)\ P(X>1)=1-P(X=1)=1-f(1)=1-0{,}8\times 0{,}2^{0}=1-0{,}8=0{,}2$$

$$e)\ E(X)=\frac{1}{p}=\frac{1}{0{,}8}=1{,}25 \qquad V(X)=\frac{q}{p^2}=\frac{0{,}2}{0{,}8^2}=\frac{0{,}2}{0{,}64}=0{,}3125$$

11)

DISTRIBUIÇÃO DE PASCAL COM PARÂMETROS *p*=0,8 E *R*=3

$$f(x)=\begin{cases}\binom{x-1}{2}0{,}8^3 0{,}2^{x-3} & ,\ x=3,4,5\ldots\\ 0 & ,\ x\neq 3,4,5,\ldots\end{cases}$$

$$a)\ P(X=6)=f(6)=\binom{5}{2}0{,}8^3\times 0{,}2^3=10\times 0{,}512\times 0{,}008=0{,}04096$$

$$b)\ P(X<6)=f(3)+f(4)+f(5)=0{,}8^3\times\left[\binom{2}{2}0{,}2^0+\binom{3}{2}0{,}2^1+\binom{4}{2}0{,}2^2\right]$$

$$=0{,}512\times(1+0{,}6+0{,}24)=0{,}512\times 1{,}84=0{,}94208$$

$$c)\ P(X>3)=1-f(3)=1-\binom{2}{2}0{,}8^3 0{,}2^0=1-0{,}8^3=1-0{,}512=0{,}488$$

$$d)\ P(X>1)=1$$

$$e)\ E(X)=\frac{R}{p}=\frac{3}{0{,}8}=3{,}75 \qquad V(X)=R\times\frac{q}{p^2}=3\times\frac{0{,}2}{0{,}8^2}=\frac{0{,}6}{0{,}64}=0{,}9375$$

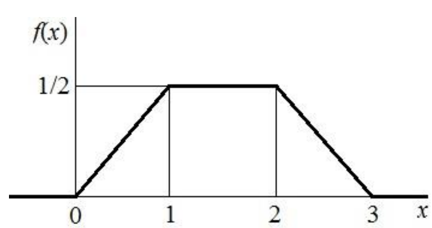
12)

$$a)\ \int_0^1 ax\,dx+\int_1^2 a\,dx+\int_2^3(-ax+3a)\,dx=1\Rightarrow a\times\frac{x^2}{2}\Big|_0^1+a\times x\Big|_1^2-a\times\frac{x^2}{2}\Big|_2^3+3a\times x\Big|_2^3=$$

$$=1\Rightarrow\frac{a}{2}+a-\frac{5a}{2}+3a=1\Rightarrow a=\frac{1}{2}$$

$b)$

$$f(x)=\begin{cases}\frac{x}{2} & 0\le x<1\\ \frac{1}{2} & 1\le x<2\\ -\frac{x}{2}+\frac{3}{2} & 2\le x\le 3\\ 0 & x\notin[0,3]\end{cases}$$



$$c)\ \ x<0 \Rightarrow F(x)=\int_{-\infty}^{x} f(t)dt=\int_{-\infty}^{x} 0dt=0$$

$$0\le x<1 \Rightarrow F(x)=\int_{-\infty}^{0} f(t)dt+\int_{0}^{x} f(t)dt=\int_{-\infty}^{0} 0dt+\int_{0}^{x}\frac{t}{2}dt=0+\frac{x^2}{4}=\frac{x^2}{4}$$

$$1\le x<2 \Rightarrow F(x)=\int_{-\infty}^{0} f(t)dt+\int_{0}^{1} f(t)dt+\int_{1}^{x} f(t)dt=\int_{-\infty}^{0} 0dt+\int_{0}^{1}\frac{t}{2}dt+\int_{1}^{x}\frac{1}{2}dt$$

$$=0+\frac{1}{4}+\frac{x-1}{2}=\frac{x}{2}-\frac{1}{4}$$

$$2\le x<3 \Rightarrow F(x)=\int_{-\infty}^{0} f(t)dt+\int_{0}^{1} f(t)dt+\int_{1}^{2} f(t)dt+\int_{2}^{x} f(t)dt$$

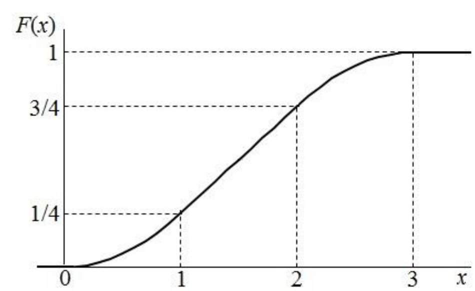
$$=\int_{-\infty}^{0} 0dt+\int_{0}^{1}\frac{t}{2}dt+\int_{1}^{2}\frac{1}{2}dt+\int_{2}^{x}\left(-\frac{t}{2}+\frac{3}{2}\right)dt=0+\frac{1}{4}+\frac{1}{2}+\left(-\frac{x^2}{4}+\frac{3x}{2}-2\right)=$$

$$=-\frac{x^2}{4}+\frac{3x}{2}-\frac{5}{4}$$

$$x\ge 3 \Rightarrow F(x)=\int_{-\infty}^{0} f(t)dt+\int_{0}^{1} f(t)dt+\int_{1}^{2} f(t)dt+\int_{2}^{3} f(t)dt+\int_{3}^{x} f(t)dt$$

$$=\int_{-\infty}^{0} f(t)dt+\int_{0}^{3} f(t)dt+\int_{3}^{x} f(t)dt=0+1+0=1$$

$$F(x)=\begin{cases} 0 & x<0 \\ \dfrac{x^2}{4} & 0\le x<1 \\ \dfrac{x}{2}-\dfrac{1}{4} & 1\le x<2 \\ -\dfrac{x^2}{4}+\dfrac{3x}{2}-\dfrac{5}{4} & 2\le x\le 3 \\ 1 & x>3 \end{cases}$$



$$d)\ \ P\left(\frac{1}{2}\le X\le\frac{5}{2}\right)=P\left(\frac{1}{2}\le X<\frac{5}{2}\right)=P\left(\frac{1}{2}<X\le\frac{5}{2}\right)=P\left(\frac{1}{2}<X<\frac{5}{2}\right)$$

$$=F\left(\frac{5}{2}\right)-F\left(\frac{1}{2}\right)=\left(-\frac{\left(\frac{5}{2}\right)^2}{4}+\frac{3\times\frac{5}{2}}{2}-\frac{5}{4}\right)-\frac{\left(\frac{1}{2}\right)^2}{4}=\frac{7}{8}$$

$$e)\ \ E(X)=\int_{-\infty}^{+\infty} xf(x)dx=\int_0^1 x\cdot\frac{x}{2}dx+\int_1^2 x\cdot\frac{1}{2}dx+\int_2^3 x\cdot\left(-\frac{x}{2}+\frac{3}{2}\right)dx$$

$$=\int_0^1\frac{x^2}{2}dx+\int_1^2\frac{x}{2}dx-\int_2^3\frac{x^2}{2}dx+\int_2^3\frac{3x}{2}dx=\left.\frac{x^3}{6}\right|_0^1+\left.\frac{x^2}{4}\right|_1^2-\left.\frac{x^3}{6}\right|_2^3+\left.\frac{3x^2}{2}\right|_2^3=$$

$$=\frac{1}{6}+\frac{3}{4}-\frac{19}{6}+\frac{15}{4}=\frac{3}{2}$$

$$E(X^2)=\int_{-\infty}^{+\infty} x^2 f(x)dx=\int_0^1 x^2\cdot\frac{x}{2}dx+\int_1^2 x^2\cdot\frac{1}{2}dx+\int_2^3 x^2\cdot\left(-\frac{x}{2}+\frac{3}{2}\right)dx=$$

$$=\int_0^1\frac{x^3}{2}dx+\int_1^2\frac{x^2}{2}dx-\int_2^3\frac{x^3}{2}dx+\int_2^3\frac{3x^2}{2}dx=\left.\frac{x^4}{8}\right|_0^1+\left.\frac{x^3}{6}\right|_1^2-\left.\frac{x^4}{8}\right|_2^3+\left.\frac{3x^3}{6}\right|_2^3$$

$$=\frac{1}{8}+\frac{7}{6}-\frac{65}{8}+\frac{57}{6}=\frac{128}{48}=\frac{8}{3}\Rightarrow V(X)=E(X^2)-[E(X)]^2=\frac{8}{3}-\left(\frac{3}{2}\right)^2=$$

$$=\frac{8}{3}-\frac{9}{4}=\frac{5}{12}$$

13)

$a)$

$$f(x)=\begin{cases}\binom{4}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{4-x} & ,\ x=0,1,2,3,4\\ 0 & ,\ x\neq 0,1,2,3,4\end{cases}$$

$$f(x)=\begin{cases}\binom{4}{x}\frac{1}{16} & ,\ x=0,1,2,3,4\\ 0 & ,\ x\neq 0,1,2,3,4\end{cases}$$

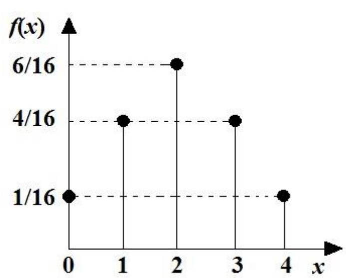
$$f(0)=\binom{4}{0}\frac{1}{16}=\frac{1}{16}$$

$$f(1)=\binom{4}{1}\frac{1}{16}=\frac{4}{16}$$

$$f(2)=\binom{4}{2}\frac{1}{16}=\frac{6}{16}$$

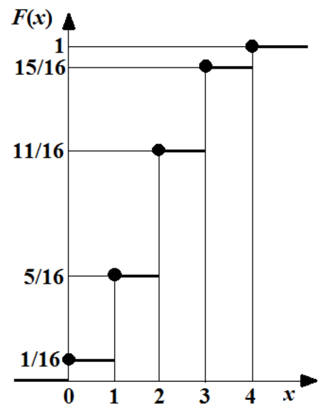
$$f(3)=\binom{4}{3}\frac{1}{16}=\frac{4}{16}$$

$$f(4)=\binom{4}{4}\frac{1}{16}=\frac{1}{16}$$



$b)$

$$F(x)=\begin{cases}0 & , x<0\\ \sum_{x_i \le x}\binom{4}{x_i}\frac{1}{16} & , 0\le x\le 4\ ,\ x_i=0,1,2,3,4\\ 1 & , x>4\end{cases}$$



$$F(0)=\frac{1}{16}$$

$$F(1)=\frac{1}{16}+\frac{4}{16}=\frac{5}{16}$$

$$F(2)=\frac{1}{16}+\frac{4}{16}+\frac{6}{16}=\frac{11}{16}$$

$$F(3)=\frac{1}{16}+\frac{4}{16}+\frac{6}{16}+\frac{4}{16}=\frac{15}{16}$$

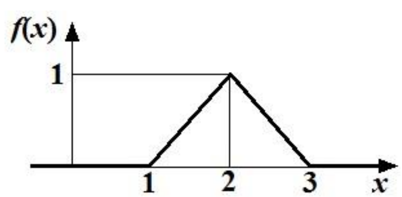
$$F(4)=\frac{1}{16}+\frac{4}{16}+\frac{6}{16}+\frac{4}{16}+\frac{1}{16}=\frac{16}{16}=1$$

$c)\ P(1<X\le 2)=f(2)=\frac{6}{16}$ $\qquad P(1\le X\le 2)=f(1)+f(2)=\frac{4}{16}+\frac{6}{16}=\frac{10}{16}$

$P(1<X<2)=0$ $\qquad P(1<X\le 2)=f(2)=\frac{6}{16}$

$d)\ \ E(X)=np=4\times\frac{1}{2}=2$ $\qquad V(X)=npq=4\times\frac{1}{2}\times\frac{1}{2}=1$

14)



$a)\ \int_1^2 k(x-1)dx+\int_2^3 k(3-x)dx=1\Rightarrow k=1$

$$\Rightarrow f(x)=\begin{cases}x-1 & ,\ 1\le x<2\\ 3-1 & ,\ 2\le x\le 3\\ 0 & ,\ x\notin[1,3]\end{cases}$$

$$x < 1 \Rightarrow F(x) = 0$$

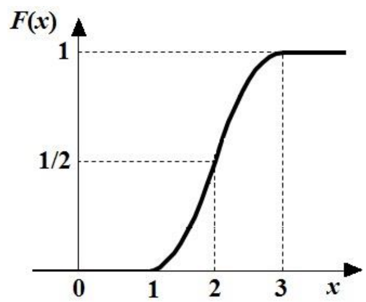
$$1 \le x < 2 \Rightarrow F(x) = \int_1^x (t-1)dt \Rightarrow F(x) = \frac{x^2}{2} - x + \frac{1}{2}$$

$$2 \le x < 3 \Rightarrow F(x) = \int_1^2 (t-1)dt + \int_2^x (3-t)dt \Rightarrow F(x) = \frac{1}{2} + \left(-\frac{x^2}{2} + 3x - 4\right) =$$

$$= -\frac{x^2}{2} + 3x - \frac{7}{2}$$

$$x \ge 3 \Rightarrow F(x) = 1$$

$$F(x) = \begin{cases} 0 & , \ x < 1 \\ \frac{x^2}{2} - x + \frac{1}{2} & , \ 1 \le x < 2 \\ -\frac{x^2}{2} + 3x - \frac{7}{2} & , \ 2 \le x < 3 \\ 1 & , \ x \ge 3 \end{cases}$$



$$E(X) = \int_1^2 x(x-1)dx + \int_2^3 x(3-x)dx = 2$$

$$E(X^2) = \int_1^2 x^2(x-1)dx + \int_2^3 x^2(3-x)dx = \frac{50}{12} \Rightarrow V(X) = \frac{50}{12} - 2^2 = \frac{1}{6}$$

$$b) \ E(L) = E[C_1 - C_2 X] = E(C_1) - E(C_2 X) = C_1 - C_2 E(X) \Rightarrow$$
$$\Rightarrow E(L) = C_1 - 2C_2$$

$$c) \ P\left(\frac{3}{2} \le X \le \frac{5}{2}\right) = F\left(\frac{5}{2}\right) - F\left(\frac{3}{2}\right) = \frac{7}{8}$$

$$P(G = 5) = P\left(\frac{3}{2} \le X \le \frac{5}{2}\right) = \frac{7}{8} \qquad P(X = -2) = 1 - p(X = 5) = \frac{1}{8}$$

$$E(G) = 5 \times \frac{7}{8} - 2 \times \frac{1}{8} = R\$4{,}00$$

15)

T≡Temperatura

$$f(t)=\begin{cases}\dfrac{1}{300-150}, t\in(150{,}300)\\ 0, t\notin(150{,}300)\end{cases} \Rightarrow f(t)=\begin{cases}\dfrac{1}{150}, t\in(150{,}300)\\ 0, t\notin(150{,}300)\end{cases} \Rightarrow$$

$$\Rightarrow F(t)=\begin{cases}0, t\le 150\\ \dfrac{t-150}{150}, t\in(150{,}300)\\ 1, t\ge 300\end{cases}$$

L≡Lucro líquido:

$$T<200 \Rightarrow L=C_2-C_1 \Rightarrow P_L(L=C_2-C_1)=P_T(T<200)=P_T(T\le 200)=F_T(200)$$
$$T\ge 200 \Rightarrow L=C_3-C_1 \Rightarrow P_L(L=C_3-C_1)=P_T(T\ge 200)=1-P_T(T<200)$$
$$=1-P_T(T\le 200)=1-F_T(200)$$
$$F_T(200)=\frac{300-200}{150}=\frac{100}{150}\Rightarrow 1-F_T(200)=\frac{50}{150}$$
$$E(L)=(C_2-C_1)P_L(L=C_2-C_1)+(C_3-C_1)P_L(L=C_3-C_1)=$$
$$=(C_2-C_1)\times\frac{50}{150}+(C_3-C_1)\times\frac{100}{150}=\frac{C_2+2C_3}{3}-C_1$$

16)

C_0≡nenhuma mensagem se extravia
C_1≡apenas a mensagem de custo c_1 se extravia
C_2≡apenas a mensagem de custo c_2 se extravia
C_3≡ambas as mensagens se extraviam
C_4≡pelo menos uma mensagem não se extravia

MÉTODO 1 (apenas 1 pacote) – X_1

$P_1(C_0)=1-p$ → perda=0 $\quad$ $P_1(C_1)=0$ → perda=c_1
$P_1(C_2)=0$ → perda=c_2 $\quad$ $P_1(C_3)=p$ → perda=c_1+c_2
$P_1(C_4)=1-P_1(C_3)=1-p$

$E(X_1)=0(1-p)+p(c_1+c_2)=p(c_1+c_2)$

MÉTODO 2 (2 pacotes) – X_2

$P_2(C_0)=(1\text{-}p)(1\text{-}p)=(1\text{-}p)^2$ → perda=0
$P_2(C_1)=(1\text{-}p)p$ → perda=c_1
$P_2(C_2)=p(1\text{-}p)$ → perda=c_2
$P_2(C_3)=pp=p^2$ → perda=c_1+c_2
$P_2(C_4)=1\text{-}P_2(C_3)=1\text{-}p^2$

$E(X_2)=0(1\text{-}p)^2+(1\text{-}p)pc_1+p(1\text{-}p)c_2+p^2(c_1+c_2)=p(1\text{-}p)(c_1+c_2)+p^2(c_1+c_2)=(c_1+c_2)[p(1\text{-}p)+p^2]=p(c_1+c_2)$

CRITÉRIO I (menor valor esperado):

Como $E(X_1)=E(X_2)$ é indiferente a escolha do método de envio

CRITÉRIO II (maior probabilidade de ambas as mensagens não se extraviarem):

Como $P_1(C_3)>P_2(C_3)$ deve-se escolher o método 1

CRITÉRIO III (maior probabilidade de pelo menos uma mensagem não se extraviar):

Como $P_2(C_4)>P_1(C_4)$ deve-se escolher o método 2

17)

$$\left.\begin{array}{l}\mu = 0{,}8\\ \sigma^2 = 0{,}0004 \Rightarrow \sigma = 0{,}02\end{array}\right\} \Rightarrow X \sim N(0{,}8;0{,}0004)$$

$$a)\ \ P(X > 0{,}81) = P\left(Z > \frac{0{,}81-0{,}8}{0{,}02}\right) = 1 - P(Z \le 0{,}5) = 1 - \Phi(0{,}5) =$$
$$= 1 - 0{,}6915 = 0{,}3085$$

$$b)\ \ P(|X-0{,}8| > 0{,}02) = P(X-0{,}8 < -0{,}02 \text{ ou } X-0{,}8 > 0{,}02) = P\left(Z < \frac{0{,}78-0{,}8}{0{,}02}\right) +$$
$$+ P\left(Z > \frac{0{,}82-0{,}8}{0{,}02}\right) = \Phi(-1) + [1-\Phi(1)] = [1-\Phi(1)] + [1-\Phi(1)] = 2[1-\Phi(1)] =$$
$$= 2(1-0{,}8413) = 2 \times 0{,}1587 = 0{,}3174$$

$c)\ G \equiv$ ganho do fabricante

$P(G = -0{,}01) = 0{,}3174$

$P(G = 0{,}05) = P(0{,}78 \leq X \leq 0{,}8) = \dfrac{(1-0{,}3174)}{2} = \dfrac{0{,}6826}{2} = 0{,}3413$

$P(G = 0{,}10) = P(0{,}8 \leq X \leq 0{,}82) = 0{,}3413$

$E(G) = -0{,}01 \times 0{,}3174 + 0{,}05 \times 0{,}3413 + 0{,}10 \times 0{,}3413 = R\$0{,}04821$

$d)\ P(X \leq x) = 0{,}95 \Rightarrow P\left(Z \leq \dfrac{x-0{,}8}{0{,}02}\right) = 0{,}95 \Rightarrow \Phi\left(\dfrac{x-0{,}8}{0{,}02}\right) = 0{,}95 \Rightarrow \dfrac{x-0{,}8}{0{,}02}$

$= 1{,}64 \Rightarrow x = 0{,}8328 \text{ cm}$

$e)\ Y \equiv$ quantidade de cabos defeituosos no lote

$Y \sim B(n, p) \quad n = 5 \quad p = 0{,}3174$

$P(Y = 3) = \dbinom{5}{3} 0{,}3174^3 \times 0{,}6826^2 = 0{,}148989$

18)

$X \sim N(1{,}60; 0{,}04)$

$\mu = 1{,}60 \quad \sigma^2 = 0{,}04 \Rightarrow \sigma = 0{,}2$

$P(X \leq x) = 0{,}90 \Rightarrow P\left(Z \leq \dfrac{x-1{,}60}{0{,}2}\right) = 0{,}90 \Rightarrow \Phi\left(\dfrac{x-1{,}60}{0{,}2}\right) = 0{,}90 \Rightarrow$

$\Rightarrow \dfrac{x-1{,}60}{0{,}2} = 1{,}28 \Rightarrow x = 1{,}856 \text{ m}$

19)

BINOMIAL

$X \sim B(10; 0{,}4) \quad n = 10 \quad p = 0{,}4$

$P(X \leq 1) = P(X = 0) + P(X = 1) = \dbinom{10}{0} 0{,}4^0 0{,}6^{10} + \dbinom{10}{1} 0{,}4^1 0{,}6^9$

$= 0{,}0060466176 + 0{,}0403107840 = 0{,}0464$

NORMAL

$\mu = np = 10 \times 0{,}4 = 4 \quad \sigma^2 = npq = 4 \times 0{,}6 = 2{,}4 \Rightarrow X \sim N(4; 2{,}4)$

$\sigma = 1{,}54919333848297$

$$P(X \le 1) = P\left(Z \le \frac{1-4}{1{,}5491933384\,8297}\right) = P(Z \le -1{,}936491673) =$$

$$= \Phi(-1{,}94) = 1 - \Phi(1{,}94) = 1 - 0{,}9738 = 0{,}0262$$

20)

$$X \equiv \text{tempo}$$

$$X \sim Exp(\lambda) \quad E(X) = \frac{1}{\lambda} = 3 \Rightarrow \lambda = \frac{1}{3}$$

$$f(x) = \begin{cases} \frac{1}{3} e^{-\frac{x}{3}}, x \ge 0 \\ 0, x < 0 \end{cases} \qquad F(x) = 1 - e^{-\frac{x}{3}}, x \ge 0$$

$$P(X < 1{,}5) = P(X \le 1{,}5) = F(1{,}5) = 1 - e^{-\frac{1{,}5}{3}} = 1 - 0{,}60653066 = 0{,}39346934$$

21)

$$X \sim \text{Uniforme}(1{,}4)$$

$$f(x) = \begin{cases} \frac{1}{3}, x \in [1{,}4] \\ 0, x \notin [1{,}4] \end{cases} \qquad F(x) = \begin{cases} 0, x < 1 \\ \frac{x-1}{3}, x \in [1{,}4] \\ 1, x > 4 \end{cases}$$

$$a) \; P(2 \le X \le 3) = F(3) - F(2) = \frac{3-1}{3} - \frac{2-1}{3} = \frac{2}{3} - \frac{1}{3} = \frac{1}{3}$$

$$b) \; P(0{,}5 \le X \le 2{,}5) = P(X \le 2{,}5) - P(X \le 0{,}5) = F(2) - F(0{,}5) =$$

$$= \frac{2{,}5-1}{3} - 0 = \frac{\frac{3}{2}}{3} = \frac{3}{6} = \frac{1}{2}$$

$$c) \; P(X = 2) = 0$$

$$d) \; E(X) = \frac{1+4}{2} = 2{,}5 \qquad V(X) = \frac{(4-1)^2}{12} = \frac{9}{12} = \frac{3}{4}$$

22)

$$\left.\begin{array}{l} n = 60000 \\ p = \dfrac{3}{100000} \end{array}\right\} \Rightarrow \lambda = np = \frac{60000 \times 3}{100000} = 1{,}8$$

$$P(X = x) = \frac{1{,}8^x e^{-1{,}8}}{x!}, x = 0,1,2,\ldots$$

$$P(X \le 5) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) =$$

$$= e^{-1{,}8}\left(\frac{1{,}8^0}{0!} + \frac{1{,}8^1}{1!} + \frac{1{,}8^2}{2!} + \frac{1{,}8^3}{3!} + \frac{1{,}8^4}{4!} + \frac{1{,}8^5}{5!}\right) = 0{,}989622$$

23)

$$f(x) = \begin{cases} \dfrac{3}{5} \cdot 10^{-5} \cdot x(100 - x) & 0 \le \mathrm{X} \le 100 \\ 0 & \mathrm{x} \notin [0,100] \end{cases}$$

$$E(X) = \int_0^{100} x \cdot \frac{3}{5} \cdot 10^{-5} \cdot x(100 - x) = \frac{3}{5} \cdot 10^{-5}\left(100\int_0^{100} x^2 dx - \int_0^{100} x^3 dx\right) =$$

$$= \frac{3}{5} \cdot 10^{-5}\left(100 \cdot \left.\frac{x^3}{3}\right|_0^{100} - \left.\frac{x^4}{4}\right|_0^{100}\right) = \frac{3}{5} \cdot 10^{-5}\left(100 \cdot \frac{100^3}{3} - \frac{100^4}{4}\right) =$$

$$= \frac{3}{5} \cdot 10^{-5}\left(\frac{10^8}{3} - \frac{10^8}{4}\right) = \frac{3}{5} \cdot 10^3\left(\frac{1}{3} - \frac{1}{4}\right) = \frac{3}{5} \cdot 10^3 \cdot \frac{4-3}{12} = \frac{10^3}{20} = 50$$

$$E(P) = E(C_1 + C_2 X) = E(C_1) + E(C_2 X) = C_1 + C_2 E(X) = C_1 + 50 C_2$$

24)

$$f(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & x \le 0 \end{cases} \Rightarrow X \sim \mathrm{Exp}(1) \Rightarrow F(x) = 1 - e^{-x}$$

$$P(X \le 0{,}9) = F(0{,}9) = 1 - e^{-0{,}9} = 0{,}59343034 \Rightarrow P(X > 0{,}9) =$$

$$= 1 - 0{,}59343034 = 0{,}40656966$$

$$X \le 0{,}9 \Rightarrow L = -2 \qquad X > 0{,}9 \Rightarrow L = 5 - 2 = 3$$

$$E(L) = -2 \times 0{,}59343034 + 3 \times 0{,}40656966 = US\$0{,}0328$$

25)

Dada uma pessoa que necessite fazer uma ligação externa, designemos por X a variável aleatória igual à quantidade de outras pessoas ocupando linha externa, assim, X tem distribuição Binomial $B(n\text{-}1\ ,p)$:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x},\ x = 0,1,2,3,\ldots,n-1$$

Se x_0 é a quantidade de linhas externas existentes, a probabilidade de que, a determinado instante, se tenha uma linha externa desocupada é dada por:

$$P(X \leq x_0 - 1) = \sum_{x=0}^{x_0-1} \binom{n}{x} p^x (1-p)^{n-x}$$

Como $F(x)=P(X\leq x)$ é crescente com $\lim\limits_{x\to+\infty} F(x) = 1$ existirá um x_0 tal que $F(x_0\text{-}1)\geq 0{,}95$, que é o número mínimo de linhas procurado.

Obs: dependendo dos valores de n-1 e p, pode-se utilizar a aproximação pela Distribuição de Poisson ou Normal.

Para n=51 e p=0,10 temos:

x	$P(X\leq x)$	x	$P(X\leq x)$
0	0,00515	7	0,87785
1	0,03379	8	0,94213
2	0,11173	9	0,97546
3	0,25029	10	0,99065
4	0,43120	11	0,99678
5	0,61612	12	0,99900
6	0,77023	13	$\cong 1$

Veja que se tem $P(X\leq x)\geq 0{,}95$ em x=9, ou seja, x_0-1=9, então a quantidade (mínima) de linhas externas desejada é x_0=10

Parte 3

Inferência Estatística

Capítulo 1

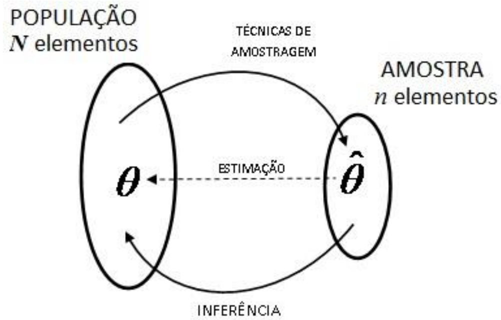
Medidas Descritivas

Introdução

A INFERÊNCIA ESTATÍSTICA cuida da estimação e testes de hipóteses feitas a respeito de parâmetros populacionais. Os parâmetros populacionais são medidas que caracterizam a população, como a média a variância e a proporção. Por exemplo, o provedor de determinada página da Internet, para melhor dimensionar seu sistema de atendimento e aprimorar a qualidade do mesmo, pode estar interessado em determinar (estimar) o tempo médio (média) de acesso de seus usuários bem como a variabilidade (variância) desse tempo, ou mesmo a incidência (proporção) de acessos mal sucedidos. Aqui está caracterizado um caso de ESTIMAÇÃO.

Por outro lado, esse mesmo provedor necessita comprar certo tipo de componente para seus equipamentos, que devem atender a determinadas condições, como o tempo médio de vida, que não deve ser inferior a um valor específico. Certo fornecedor afirma que seu produto atende a essa especificação. O provedor deve, então, verificar (testar) a validade dessa afirmação (hipótese), a fim de decidir se compra ou não o componente desse fornecedor. Aqui está caracterizado um caso de TESTE DE HIPÓTESES.

Em ambos os casos, para fazermos inferências a respeito de um parâmetro populacional, genericamente denotado por θ, o procedimento consiste em retirar uma amostra da população e, a partir dessa amostra, utilizarmos um estimador, genericamente denotado por $\hat{\theta}$, para estimar o valor de θ. Denotaremos a quantidade de elementos (tamanho) da população por N e da amostra por n. A figura abaixo ilustra este procedimento.



1 O que são Medidas Descritivas

As medidas descritivas têm o objetivo de descrever o comportamento das populações em relação às suas diversas características (variáveis), como, por exemplo, o tempo de duração da utilização do computador pelos alunos de uma escola, bem como se este tempo varia muito, de aluno para aluno. Outra variável que pode ser citada, de grande interesse, é a quantidade de alunos que utilizam o computador. As medidas descritivas podem ser classificadas em duas categorias básicas: medidas de posição e de variação.

As medidas de posição, como a termo sugere, ocupa uma posição entre os valores mínimo e máximo da variável. Além disso, esse tipo de medida, ao ser comparada entre duas populações, indica qual delas está em uma posição superior (ou inferior) em relação à outra. Veja que, para um grupo de alunos cujos valores mínimo e máximo do tempo de utilização do computador são $2h$ e $5h$, respectivamente, a média aritmética, que será definida posteriormente, é uma medida que está posicionada entre esses valores como, por exemplo, $4h$.

As medidas de variação indicam o grau de variabilidade dos valores da variável. Observe que os valores mínimo e máximo da variável já fornece uma primeira informação sobre a variabilidade da mesma. A diferença entre estes dois valores é chamada de amplitude total. Se outro grupo de alunos tem um mínimo de $3h$ e máximo de $5h$ com média de $4h$ de utilização, percebemos que ele está na mesma posição (médias iguais) do primeiro grupo, mas tem uma variabilidade menor, analisando a amplitude total.

Importante

Uma única medida descritiva não é suficiente para a análise do comportamento das populações. Pelo menos deve-se utilizar uma medida de posição associada a uma de variação.

2 Medidas Descritivas e Inferência

Nos trabalhos estatísticos, as medidas descritivas têm papel fundamental na inferência sobre as variáveis das populações. A Inferência Estatística é um dos seus ramos, que cuida da estimação e testes de hipóteses sobre parâmetros populacionais, esta inferência é realizada a partir de amostras, ou seja, a partir de uma parte da população (amostra) tiramos conclusões sobre a mesma. Assim, as medidas descritivas são calculadas a partir de amostras e, de acordo com condições previamente impostas no planejamento do trabalho, são consideradas como estimativas válidas dos respectivos parâmetros populacionais.

Observação

No contexto deste livro serão vistas apenas a média aritmética e a proporção (medidas de posição), e a variância (medida de variação). Outras medidas podem ser encontradas na bibliografia indicada.

3 Média Aritmética

A média aritmética de um conjunto de dados é determinada pelo quociente entre a soma dos elementos desse conjunto e a quantidade desses elementos. Para uma amostra com n elementos, $x_1, x_2, x_3, \ldots, x_n$, sua média aritmética, denominada média amostral, pode ser expressada pela seguinte fórmula:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

DESAFIO

Mostre que

$$\sum_{i=1}^{n} \left(x_i - \overline{X}\right) = 0$$

EXEMPLO 3.1.3[1]

Uma escola gastou nos 5 primeiros meses do ano 300,400,200,300 e 500 reais, respectivamente, na manutenção de seu laboratório de informática. O gasto médio com esse serviço pode ser calculado através da média aritmética:

$$\overline{X} = \frac{\sum_{i=1}^{5} x_i}{5} = \frac{300 + 400 + 200 + 300 + 500}{5} = \frac{1700}{5} = R\$340{,}00$$

Veja que a média corresponde a um valor unitário, ou seja, o que se fez foi determinar um valor igual (unitário) para cada manutenção mensal realizada, não alterando o total gasto:

x_1	x_2	x_3	x_4	x_5	TOTAL
300	400	200	300	500	1700
340	340	340	340	340	1700

Este fato permite uma aplicação importante que é calcular (estimar) totais baseados na média aritmética.

Supondo que as condições do laboratório da escola não mudem durante o ano, os cinco meses utilizados são uma amostra daquilo que ocorrerá durante todo o ano, então, podemos inferir que durante o ano (12 meses), a escola deverá gastar ao todo $R\$4.080,00$:

$$\text{TOTAL GASTO} = 12 \times 340 = \text{R\$}4.080,00$$

Esse resultado foi obtido através de uma variação da fórmula da média:

$$\sum_{i=1}^{n} x_i = n \times \overline{X}$$

Outra variação bastante utilizada, para se encontrar a quantidade de elementos do grupo, é:

$$n = \frac{\sum_{i=1}^{n} x_i}{\overline{X}}$$

4 Proporção

A proporção dá a relação entre a quantidade de elementos de um grupo que pertencem a certa categoria, atendendo a determinada característica, e a quantidade total de elementos deste grupo. A proporção amostral é dada pela seguinte expressão:

$$\hat{p} = \frac{n_A}{n}$$

Onde n_A é a quantidade de elementos da amostra que pertencem à categoria A e n é a quantidade total de elementos da amostra.

EXEMPLO 3.1.4[1]

Uma escola gastou nos 5 primeiros meses do ano $300,400,200,300$ e 500 reais, respectivamente, na manutenção de seu laboratório de informática. A proporção de meses com gastos acima de $R\$300,00$ é dada por:

$$\hat{p} = \frac{2}{5} = 0,40$$

Multiplicando-se a proporção por 100, temos o percentual de meses com gasto superior a $R\$300,00$:

$$0,40\text{x}100 = 40\%$$

A proporção amostral permite estimar a quantidade de elementos da população que atende a determinada característica. Para um determinado ano, a quantidade de meses com gasto superior a $R\$300,00$ é:

$$0,40\text{x}12=4,8\cong 5 \text{ meses}$$

5 Variância e Desvio Padrão

Para efeitos de inferência, a variância de uma amostra com n elementos, $x_1, x_2, x_3, \ldots, x_n$, sua variância é determinada pelo quociente entre a soma dos quadrados das diferenças entre seus elementos e a média da amostra e a quantidade desses elementos menos um, denominada variância amostral, e é expressada pela seguinte fórmula:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n-1} \quad \text{ou} \quad S^2 = \frac{\sum_{i=1}^{n} x_i^2}{n-1} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$$

Desafio

Mostre que

$\sum_{i=1}^{n}(x_i - a)^2$

é mínima quando

$a = \overline{X}$

O desvio padrão amostral é dado pela raiz quadrada da variância amostral:

$$S = \sqrt{S^2}$$

Observações

- A rigor, o denominador da fórmula da variância é a quantidade n de elementos. Para a variância amostral, que é utilizada na inferência, usa-se n-1 para torná-la um estimador não viciado da variância populacional, como será visto posteriormente;
- Observe que a variância mede, em média, o quanto os elementos do conjunto de dados se afastam da média do conjunto, ou seja, quanto mais esses elementos se afastarem da média maior será a variância ou, em outras palavras, mais heterogêneo é o grupo;
- A unidade da variância é a unidade da variável ao quadrado, enquanto que a unidade do desvio padrão é a própria unidade da variável.

EXEMPLO 3.1.5[1]

Uma escola gastou nos 5 primeiros meses do ano $300,400,200,300$ e 500 reais, respectivamente, na manutenção de seu laboratório de informática. A variância e desvio padrão amostrais dos gastos com esse serviço são dados por:

$$\overline{X} = 340$$

$$S^2 = \frac{(300-340)^2 + (400-340)^2 + \ldots + (500-340)^2}{5-1} = 13.000\ R\$^2$$

$$S = \sqrt{13.000} = R\$114,02$$

Capítulo 2

Distribuições Amostrais de Probabilidade

1 Conceitos iniciais de Estimadores e Estimativas

Para as discussões seguintes, vamos adiantar as definições de ESTIMADOR e ESTIMATIVA, que serão detalhadas posteriormente no Capítulo 3.

- ESTIMADOR ($\hat{\theta}$) – é uma função dos elementos da amostra, que tem o objetivo de estimar o valor do parâmetro populacional θ
- ESTIMATIVA – é o valor numérico do estimador

Por exemplo, para estimarmos o valor da média populacional, denotada por μ_X, podemos utilizar a média aritmética da amostra, denotada por $\overline{X}$, então:

PARÂMETRO POPULACIONAL - $\theta \equiv \mu_X$

ESTIMADOR - $\hat{\theta} \equiv \overline{X} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

Se a amostra for 2,3,4,1,2,3,2,5,5,3

ESTIMATIVA: $\overline{X} = \dfrac{\sum_{i=1}^{10} x_i}{10} = \dfrac{2+3+4+1+2+3+2+5+5+3}{10} = \dfrac{30}{10} = 3$

É importante observar que a inferência envolve um EXPERIMENTO ALEATÓRIO, que é a retirada de uma amostra da população. Desta forma, a variável populacional de interesse é uma VARIÁVEL ALEATÓRIA, geralmente denotada por X, com determinada distribuição de probabilidade (modelo probabilístico), não necessariamente conhecida, com esperança (média) $E(X) = \mu_X$ e variância $V(X) = \sigma_X^2$. Já o estimador, que também é uma variável aleatória, pois seu valor varia dependendo da amostra retirada, deve, necessariamente, ter sua distribuição de probabilidade conhecida, com esperança(média) $E(\hat{\theta}) = \mu_{\hat{\theta}}$ e variância $V(\hat{\theta}) = \sigma_{\hat{\theta}}^2$. A necessidade de se conhecer a distribuição de probabilidade do estimador deve-se ao fato de que, ao realizarmos uma estimativa a partir de uma amostra, estamos sujeitos a um ERRO DE ESTIMATIVA, que é a diferença (para mais ou para menos) entre o valor real do parâmetro (desconhecido) e sua estimativa: $e = \hat{\theta} - \theta$. Assim, todo o processo de inferência é baseado na probabilidade, chamada de NÍVEL DE CONFIANÇA, denotado por 1-α, de que o erro de estimativa esteja dentro de uma margem aceitável d:

$$P\left(|\,\theta - \hat{\theta}\,| \le d\right) = 1 - \alpha$$

Como vimos acima, os estimadores são variáveis aleatórias e suas distribuições de probabilidade são chamadas de DISTRIBUIÇÕES AMOSTRAIS.

2 Distribuição Amostral da Média - $\overline{X}$

Para amostras retiradas de populações com distribuição de probabilidade Normal a distribuição de probabilidade de $\overline{X}$ é normal, independentemente do tamanho da amostra (n). Caso a população não seja Normal ou seja desconhecida, a distribuição de $\overline{X}$ será normal para grandes amostras. Ou seja:

$$\overline{X} \to N\left(\mu_{\overline{X}}, \sigma_{\overline{X}}^2\right) \text{ se } \begin{cases} X \to N(\mu_X, \sigma_X^2) \\ \text{ou} \\ n \text{ grande} \end{cases}$$

Com

$$\mu_{\overline{X}} = \mu_X \quad \text{e} \quad \sigma_{\overline{X}}^2 = \begin{cases} \dfrac{\sigma_X^2}{n} \quad \text{(população infinita)} \\ \text{ou} \\ \dfrac{\sigma_X^2}{n} \cdot \dfrac{N-n}{N-1} \quad \text{(população finita)} \end{cases}$$

Onde μ_X e σ_X^2 são, respectivamente, a média e variância populacionais.

Observações

- Veja que na expressão de $\sigma_{\overline{X}}$ para populações finitas, o fator $\dfrac{N-n}{N-1}$, para n fixo, tende a 1 quando N tende a infinito, resultando na expressão de $\sigma_{\overline{X}}$ para população infinita
- Na prática, a amostra pode ser considerada grande se $n \geq 30$
- Em decorrência da primeira observação, também na prática, pode-se considerar a população infinita se $\dfrac{n}{N} \leq 0{,}05 \ \ (5\%)$

EXEMPLO 3.2.2[1]

O tempo de acesso dos usuários de determinada página na Internet tem média 170 min e variância 400 min^2. Observada uma amostra de 100 usuários: (a) qual a probabilidade de que a média amostral seja superior a 175, considerando a população infinita ? (b) Considerando que a população é finita, com 500 usuários, qual será essa probabilidade ? (c) E se a população for de 10.000 usuários ?

Como a distribuição do tempo de acesso é desconhecida mas a amostra é grande (n>30), pode-se considerar que a média amostral, $\overline{X}$, tem distribuição Normal:

ítem (a): $P(\overline{X} > 175)$ para população infinita

$$\mu_X = 170 \quad \sigma_X^2 = 400 \Rightarrow \sigma_X = \sqrt{400} = 20 \quad n = 100$$

$$\mu_{\overline{X}} = \mu_X = 170 \quad \sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n} = \frac{400}{100} = 4 \Rightarrow \sigma_{\overline{X}} = 2$$

$$\overline{X} \to N(170,4)$$

$$P(\overline{X} > 175) = P\left(Z > \frac{175-170}{2}\right) = P(Z > 2{,}5) = 1 - P(Z \leq 2{,}5)$$

$$= 1 - \Phi(2{,}5) = 1 - 0{,}9938 = 0{,}0062$$

ítem (b): $P(\overline{X} > 175)$ para população finita (N=500)

$$N = 500 \quad \mu_X = 170 \quad \sigma_X^2 = 400 \Rightarrow \sigma_X = \sqrt{400} = 20 \quad n = 100$$

$$\mu_{\overline{X}} = \mu_X = 170 \quad \sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n} \cdot \frac{N-n}{N-1} = \frac{400}{100} \cdot \frac{500-100}{500-1} = 4 \times 0{,}8 = 3{,}2 \Rightarrow \sigma_{\overline{X}} = 1{,}79$$

$$\overline{X} \to N(170;3{,}2) \quad P\left(\overline{X} > 175\right) = P\left(Z > \frac{175-170}{1{,}79}\right) = P\left(Z > 2{,}79\right) = 1 - P(Z \le 2{,}79)$$
$$= 1 - \Phi(2{,}79) = 1 - 0{,}9974 = 0{,}0026$$

ítem (c): $P\left(\overline{X} > 175\right)$ para população finita (N=10.000)

$$N = 10.000 \quad \mu_X = 170 \quad \sigma_X^2 = 400 \Rightarrow \sigma_X = \sqrt{400} = 20 \quad n = 100$$
$$\mu_{\overline{X}} = \mu_X = 170 \quad \sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n} \cdot \frac{N-n}{N-1} = \frac{400}{100} \cdot \frac{10.000-100}{10.000-1} = 4 \times 0{,}99 = 3{,}96 \Rightarrow \sigma_{\overline{X}} = 1{,}99$$
$$\overline{X} \to N(170;3{,}96) \quad P\left(\overline{X} > 175\right) = P\left(Z > \frac{175-170}{1{,}99}\right) = P\left(Z > 2{,}51\right) = 1 - P(Z \le 2{,}51)$$
$$= 1 - \Phi(2{,}51) = 1 - 0{,}9940 = 0{,}0060$$

Veja que para a população finita igual a 10.000 tem-se $\frac{n}{N} = \frac{100}{10000} = 0{,}01 < 5\%$, podendo então ser considerada infinita. Compare os resultados dos itens (a) e (c).

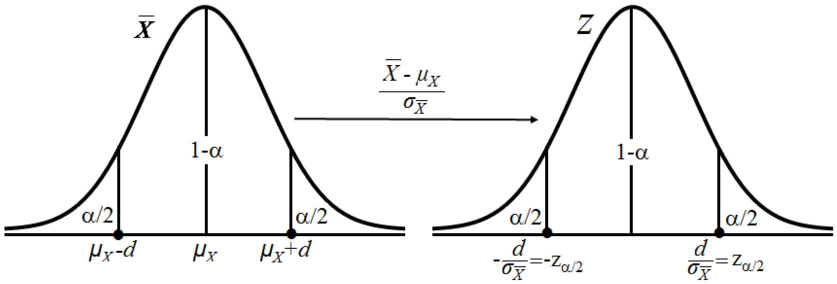
2.1 Cálculo do tamanho da amostra para estimar a média populacional

Em geral, ao se planejar uma pesquisa por amosragem, inicialmente define-se um um erro de estimativa d, considerado aceitável, associado a um determinado nível de confiança 1-α (probabilidade de se cometer no máximo o erro d): $P\left(|\hat{\theta} - \theta| \le d\right) = 1 - \alpha$.

Fazendo $\hat{\theta} = \overline{X}$ e $\theta = \mu_X$, e considerando-se população infinita, tem-se:

$$P\left(\left|\overline{X} - \mu_X\right| \le d\right) = 1 - \alpha \Rightarrow P\left(-d \le \overline{X} - \mu_X \le d\right) = 1 - \alpha$$
$$\Rightarrow P\left(-d \le \overline{X} - \mu_{\overline{X}} \le d\right) = 1 - \alpha \Rightarrow P\left(-\frac{d}{\sigma_{\overline{X}}} \le \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} \le \frac{d}{\sigma_{\overline{X}}}\right) = 1 - \alpha$$
$$\Rightarrow P\left(-\frac{d}{\sigma_{\overline{X}}} \le Z \le \frac{d}{\sigma_{\overline{X}}}\right) = 1 - \alpha \Rightarrow \frac{d}{\sigma_{\overline{X}}} = z_{\alpha/2} \Rightarrow d = z_{\alpha/2} \cdot \sigma_{\overline{X}} \Rightarrow d = z_{\alpha/2} \cdot \frac{\sigma_X}{\sqrt{n}}$$

Então: $$\boxed{n = \frac{z_{\alpha/2}^2 \cdot \sigma_X^2}{d^2}}$$



Para populações finitas, no desenvolvimento acima substitui-se $\sigma_{\bar{X}}$ por $\frac{\sigma_X}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$, obtendo-se:

$$n = \frac{N \cdot z_{\alpha/2}^2 \cdot \sigma_X^2}{z_{\alpha/2}^2 \cdot \sigma_X^2 + (N-1) \cdot d^2}$$

Dividindo-se o numerador e denominador da expressão acima por Nd^2, passaremos a ter:

$$n = \frac{\dfrac{N \cdot z_{\alpha/2}^2 \cdot \sigma_X^2}{N \cdot d^2}}{\dfrac{z_{\alpha/2}^2 \cdot \sigma_X^2}{N \cdot d^2} + \dfrac{(N-1) \cdot d^2}{N \cdot d^2}} = \frac{\dfrac{z_{\alpha/2}^2 \cdot \sigma_X^2}{d^2}}{\dfrac{\dfrac{z_{\alpha/2}^2 \cdot \sigma_X^2}{d^2}}{N} + \dfrac{(N-1)}{N}}$$

Fazendo $\frac{z_{\alpha/2}^2 \cdot \sigma_X^2}{d^2} = n_0$, que é o tamanho da amostra para população infinita, podemos escrever $n = \dfrac{n_0}{\dfrac{(N-1)}{N} + \dfrac{n_0}{N}} \cong \dfrac{n_0}{1 + \dfrac{n_0}{N}}$

Assim, na prática, mesmo a população sendo conhecidamente finita, com N elementos, primeiramente se calcula o tamanho da amostra considerando-a infinita, ou seja, calcula-se n_0 e, em seguida, verifica-se a relação de n_0 com N e corrige-se o tamanho da amostra, se necessário, da seguinte forma:

$$n = \begin{cases} n_0 & \text{, se } \dfrac{n_0}{N} \leq 0{,}05 \\ \dfrac{n_0}{1 + \dfrac{n_0}{N}} & \text{, caso contrário} \end{cases}$$

EXEMPLO 3.2.2.1[1]

Qual o tamanho da amostra para se estimar a média do tempo de acesso dos usuários de determinada página na Internet, que tem variância de 400 min^2, com erro de estimativa máximo de 2 min, com 90% de confiança: (a) considerando a população infinita? (b) se a população for finita, com 500 usuários? (c) se a população for finita, com 10.000 usuários? (d) Para população infinita, qual o erro de estimativa se a amostra for de 100 usuários ?

ítem (a): tamanho da amostra para população infinita

$$\sigma_X^2 = 400$$
$$d = 2 \quad 1-\alpha = 0{,}90 \Rightarrow z_{\alpha/2} = 1{,}64$$
$$n = \frac{z_{\alpha/2}^2 \cdot \sigma_X^2}{d^2} = \frac{1{,}64^2 \cdot 400}{2^2} \cong 269$$

ítem (b): tamanho da amostra para população finita (N=500)

como $\frac{n_0}{N} = \frac{269}{500} = 0{,}54 > 0{,}05$ a população é considerada, realmente, finita

$$N = 500$$
$$n_0 = 269$$
$$n = \frac{n_0}{1+\frac{n_0}{N}} = \frac{269}{1+\frac{269}{500}} \cong 175$$

ítem (c): tamanho da amostra para população finita (N=10.000)

como $\frac{n_0}{N} = \frac{269}{10.000} = 0{,}03 < 0{,}05$ a população pode ser considerada infinita
então $n = n_0 = 269$

ítem (d): erro se a amostra for de 100 usuários, para população infinita

$$100 = \frac{1{,}64^2 \cdot 400}{d^2} \Rightarrow d = \sqrt{\frac{1{,}64^2 \cdot 400}{100}} \cong 3{,}28$$

3 Distribuição amostral da proporção - $\hat{p}$

Para amostras grandes ($n \geq 30$), a distribuição de probabilidade de $\hat{p}$ é normal, com média $\mu_{\hat{p}}$ e variância $\sigma^2_{\hat{p}}$:

$$\hat{p} \to N\left(\mu_{\hat{p}}, \sigma^2_{\hat{p}}\right)$$

Com

$$\mu_{\hat{p}} = p \quad \text{e} \quad \sigma^2_{\hat{p}} = \begin{cases} \dfrac{pq}{n} & \text{(população infinita)} \\ \dfrac{pq}{n} \cdot \dfrac{N-n}{N-1} & \text{(população finita)} \end{cases}$$

Onde p é a proporção populacional e $q = 1 - p$

Observação

No caso da proporção amostral, se um elemento pertence à categoria A, atribui-se a ele o valor 1, caso contrário o valor 0, assim, a proporção amostral também pode ser calculada por $\hat{p} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ (média amostral)

EXEMPLO 3.2.3[1]

O percentual de acessos mal sucedidos a determinada página da Internet é de 5%. Se forem observados 100 acessos: (a) qual probabilidade de que no máximo 3% deles sejam mal sucedidos? (b) qual será essa probabilidade se a população for finita, com 500 acessos? (c) e se for de 10.000 acessos?

Como a amostra é grande (n>30), pode-se considerar que a proporção amostral, $\hat{p}$, tem distribuição Normal:

ítem (a): $P(\hat{p} \leq 0{,}03)$ para população infinita

$$p = 0{,}05 \quad n = 100 \quad \mu_{\hat{p}} = p = 0{,}05$$

$$\sigma^2_{\hat{p}} = \frac{pq}{n} = \frac{0{,}05 \times 0{,}95}{100} = 0{,}00475 \Rightarrow \sigma_{\hat{p}} = 0{,}022 \quad \hat{p} \to N(0{,}05; 0{,}00475)$$

$$P(\hat{p} \le 0{,}03) = P\left(Z \le \frac{0{,}03 - 0{,}05}{0{,}022}\right) = P(Z \le -0{,}91) = 1 - P(Z \le 0{,}91)$$
$$= 1 - \Phi(0{,}91) = 1 - 0{,}8186 = 0{,}1814$$

ítem (b): $P(\hat{p} \le 0{,}03)$ para população finita (N=500)

$$N = 500 \quad p = 0{,}05 \quad n = 100 \quad \mu_{\hat{p}} = p = 0{,}05$$
$$\sigma^2_{\hat{p}} = \frac{pq}{n} \cdot \frac{500 - 100}{500 - 1} = \frac{0{,}05 \times 0{,}95}{100} \cdot \frac{400}{499} = 0{,}000381 \Rightarrow \sigma_{\hat{p}} = 0{,}0195$$
$$\hat{p} \to N(0{,}05; 0{,}000381) \quad P(\hat{p} \le 0{,}03) = P\left(Z \le \frac{0{,}03 - 0{,}05}{0{,}0195}\right) =$$
$$= P(Z \le -1{,}03) = 1 - P(Z \le 1{,}03) = 1 - \Phi(1{,}03) = 1 - 0{,}8485 = 0{,}1515$$

ítem (c): $P(\hat{p} \le 0{,}03)$ para população finita (N=10.000)

$$N = 10.000 \quad p = 0{,}05 \quad n = 100 \quad \mu_{\hat{p}} = p = 0{,}05$$
$$\sigma^2_{\hat{p}} = \frac{pq}{n} \cdot \frac{10.000 - 100}{10.000 - 1} = \frac{0{,}05 \times 0{,}95}{100} \cdot \frac{9.900}{9.999} = 0{,}00047 \Rightarrow \sigma_{\hat{p}} = 0{,}0217$$
$$\hat{p} \to N(0{,}05; 0{,}00047) \quad P(\hat{p} \le 0{,}03) = P\left(Z \le \frac{0{,}03 - 0{,}05}{0{,}0217}\right) =$$
$$P(Z \le -0{,}92) = 1 - P(Z \le 0{,}92) == 1 - \Phi(0{,}92) = 1 - 0{,}8212 = 0{,}1788$$

Veja que para a população finita igual a 10.000 tem-se $\frac{n}{N} = \frac{100}{10000} = 0{,}01 < 5\%$, podendo então ser considerada infinita. Compare os resultados dos itens (a) e (c).

3.1 Cálculo do tamanho da amostra para estimar a proporção populacional

Procedendo-se de forma análoga ao caso da média, tem-se:

$$P(|\hat{p} - p| \le d) = 1 - \alpha \Rightarrow P(-d \le \hat{p} - p \le d) = 1 - \alpha$$
$$\Rightarrow P(-d \le \hat{p} - \mu_{\hat{p}} \le d) = 1 - \alpha \Rightarrow P\left(-\frac{d}{\sigma_{\hat{p}}} \le \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} \le \frac{d}{\sigma_{\hat{p}}}\right) = 1 - \alpha$$
$$\Rightarrow P\left(-\frac{d}{\sigma_{\hat{p}}} \le Z \le \frac{d}{\sigma_{\hat{p}}}\right) = 1 - \alpha \Rightarrow \frac{d}{\sigma_{\hat{p}}} = z_{\alpha/2} \Rightarrow d = z_{\alpha/2} \cdot \sigma_{\hat{p}} \Rightarrow d = z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}$$

Então $n = \dfrac{z_{\alpha/2}^2 \cdot pq}{d^2}$ para populações infinitas e $n = \dfrac{N \cdot z_{\alpha/2}^2 \cdot pq}{z_{\alpha/2}^2 \cdot pq + (N-1) \cdot d^2}$ para populações finitas.

O procedimento prático para determinação do tamanho da amostra permanece o mesmo:

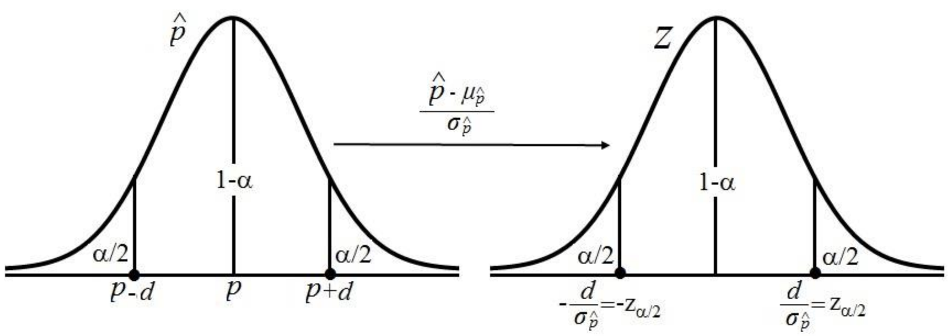
Desafio

Mostre que n é máximo quando p=0,5 (encontre o máximo de

$$f(p) = \frac{z_{\alpha/2}^2 \cdot pq}{d^2}$$

, lembre que q=1-p)

$$n = \begin{cases} n_0 & ,\text{se } \dfrac{n_0}{N} \le 0{,}05 \\ \dfrac{n_0}{1 + \dfrac{n_0}{N}} & ,\text{caso contrário} \end{cases}$$ onde $$n_0 = \frac{z_{\alpha/2}^2 \cdot pq}{d^2}$$



Observações

- como não se conhece o valor de p, que é o objeto da estimativa, ele deve ser substituído por uma estimativa inicial;
- caso se deseje obter o tamanho máximo de amostra possível deve-se substituí-lo por 0,5.

EXEMPLO 3.2.3.1[1]

Qual o tamanho da amostra para se estimar a proporção de acessos mal sucedidos a determinada página na Internet, com erro máximo de estimativa de 0,02 e 95% de confiança, tendo-se a informação de que a proporção poulacional está em torno de 10%: (a) considerando a população infinita? (b) se a população for finita, com 10.000 acessos ? (c) se a população for finita, com 20.000 acessos ? (d) Para população infinita, qual o erro de estimativa se a amostra for do 100 acessos ? (e) para a população infinita, qual é o tamanho máximo da amostra ?

ítem (a): tamanho da amostra para população infinita

$$p = 0,10$$
$$d = 0,02 \quad 1-\alpha = 0,95 \Rightarrow z_{\alpha/2} = 1,96$$
$$n = \frac{z_{\alpha/2}^2 \cdot p \cdot q}{d^2} = \frac{1,96^2 \cdot 0,10 \cdot 0,90}{0,02^2} \cong 864$$

ítem (b)

$$N = 10.000$$
$$n_0 = 864$$
$$\frac{n_0}{N} = \frac{864}{10.000} = 0,086 > 0,05 \Rightarrow \text{população finita}$$
$$n = \frac{n_0}{1+\frac{n_0}{N}} = \frac{864}{1+\frac{864}{10.000}} \cong 248$$

ítem (c): tamanho da amostra para população finita (N=20.000)

como $\frac{n_0}{N} = \frac{864}{20.000} = 0,04 < 0,05$ a população pode ser considerada infinita
então $n = n_0 = 864$

ítem (d): erro de estimativa se a amostra for de 100 acessos, para população infinita

$$100 = \frac{1,96^2 \cdot 0,10 \cdot 0,90}{d^2} \Rightarrow d = \sqrt{\frac{1,96^2 \cdot 0,10 \cdot 0,90}{100}} \cong 0,06$$

ítem (e): amostra máxima, para população infinita

$$n = \frac{z_{\alpha/2}^2 \cdot 0,5 \cdot 0,5}{d^2} = \frac{1,96^2 \cdot 0,5 \cdot 0,5}{0,02^2} \cong 2.401$$

Capítulo 3

Estimação

1 O que é Estimação

Estimação é o processo de se ESTIMAR, a partir de uma amostra, o valor de um determinado parâmetro populacional, genericamente denotado por θ. Os parâmetros populacionais são medidas que caracterizam a população, como a média, a proporção e a variância. No Capítulo 1 vimos que as medidas descritivas calculadas a partir de amostras são utilizadas para realizar estas estimativas. Mas apenas o cálculo dessas medidas não são suficientes para se ter uma estimativa "confiável". Como visto no Capítulo 2, elas são obtidas por um processo aleatório (amostra), portanto sujeitas a erro, sendo necessário que as associemos a uma margem de erro tolerável e à probabilidade de que elas estejam dentro desta margem.

2 Estimador

Um estimador, genericamente denotado por $\hat{\theta}$, é uma função dos elementos da amostra, que tem o objetivo de estimar o valor do parâmetro populacional θ.

Para uma amostra de tamanho *n*, $x_1, x_2, x_3, \ldots, x_n$, temos os seguintes estimadores:

- $\overline{X} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ (média amostral) é um estimador de μ_X (média populacional)

- $\hat{p} = \dfrac{n_A}{n}$ (proporção amostral) é um estimador de *p* (proporção populacional), onde n_A é a quantidade de elementos, na amostra, pertencentes à categoria *A*, para a qual se deseja estimar a proporção *p*

- $S^2 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n-1}$ (variância amostral) é um estimador de σ_X^2 (variância populacional)

Para um mesmo parâmetro podem ser definidos diferentes estimadores.

Por exemplo, para se estimar a média populacional, podemos utilizar $\overline{X} = \dfrac{\sum_{i=1}^{n} x_i}{n}$, como mostrado acima, ou $M = \dfrac{x_{\max} + x_{\min}}{2}$, a questão é determinar qual deles é o "melhor". Existem algumas propriedades dos estimadores que permitem esta determinação, aqui veremos apenas uma delas, que é o conceito de ESTIMADOR NÃO VICIADO: $\hat{\theta}$ é um estimador não viciado de θ se $E(\hat{\theta}) = \theta$.

Observação

$E(\hat{\theta}) = \theta$ Significa que, se retirarmos várias amostras de mesmo tamanho, a média das estimativas feitas com $\hat{\theta}$ é o próprio valor de θ

Desafio

Mostre que $\overline{X}$ e $\hat{p}$ são não viciados (use a propriedade de $E(X)$ para a soma de variáveis aleatórias independentes)

No Capítulo 2 (Distribuições Amostrais) já vimos que $\overline{X}$ e $\hat{p}$ são não viciados, ou seja:

- $E(\overline{X}) = \mu_X$
- $E(\hat{p}) = p$

3 Estimativas

3.1 Estimativa por ponto

É o valor numérico do estimador.

EXEMPLO 3.3.3.1[1]

Uma amostra de 36 ligações telefônicas apresentou os seguintes resultados em relação ao tempo de duração (T) em minutos e presença de ruídos (R; 1=sim,0=não). Quais as estimativas por ponto do tempo médio de duração e da proporção de ligações com ruído ?

T	2	4	2	3	3	3	5	5	8	2	4	3	3	3	8	5	4	7
R	0	0	1	1	0	1	0	0	0	0	0	1	1	0	0	0	0	1

T	3	3	3	2	4	4	5	5	6	6	6	8	9	6	6	7	7	5
R	0	0	1	0	0	0	0	1	1	1	0	0	0	0	0	1	1	0

$$\overline{X} = \frac{2+4+2+3+\ldots+7+7+5}{36} = \frac{169}{36} = 4{,}7 \text{ min}$$

$$\hat{p} = \frac{0+0+1+1+\ldots+1+1+0}{36} = \frac{12}{36} = 0{,}33 = 33\%$$

3.2 Estimativa por intervalo (Intervalo de Confiança)

A estimativa por intervalo é feita atribuindo-se uma margem de erro à estimativa por ponto. Denotando-se a estimativa por intervalo por IC (intervalo de confiança), ela será dada por:

$$IC \equiv \left(\hat{\theta} - d_1, \hat{\theta} + d_2\right)$$

Na expressão acima, d_1 e d_2 são as margens de erro atribuídas à estimativa por ponto, elas serão iguais quando a distribuição de probabilidade de $\hat{\theta}$ for simétrica, neste caso a expressão passa a ser:

$$IC \equiv \left(\hat{\theta} - d, \hat{\theta} + d\right)$$

No exemplo acima, se for atribuída uma margem de erro de $0{,}5$ unidades para a estimativa da média e de $0{,}02$ (2 pontos percentuais) para a estimativa da proporção, os intervalos de confiança serão:

Para a média: $IC \equiv \left(4{,}7 - 0{,}5; 4{,}7 + 0{,}5\right) \equiv (4{,}2; 5{,}2)$

Para a proporção:

$$IC \equiv \left(0{,}33 - 0{,}02; 0{,}33 + 0{,}02\right) \equiv (0{,}31; 0{,}35)$$

Observação

As margens de erro para média e proporção são iguais em ambos os lados porque, como visto anteriormente, $\overline{X}$ e $\hat{p}$ têm Distribuição Normal, que é simétrica.

Na construção de intervalos de confiança a determinação da margem de erro é feita a partir da probabilidade (nível de confiança), denotada por $1\text{-}\alpha$, de que o erro de estimativa esteja dentro desta margem, no caso de distribuições simétricas, considerando-se um erro máximo de estimativa igual a d, teremos:

$$P\left(|\,\hat{\theta}-\theta\,|\leq d\right)=1-\alpha$$

4 Intervalo de Confiança para a Média

Já vimos que para populações Normais ou amostras grandes ($n\geq 30$), a distribuição de probabilidade de $\overline{X}$ é normal, com média $\mu_{\overline{X}}$ e variância $\sigma^2_{\overline{X}}$:

$$\overline{X}\rightarrow N\left(\mu_{\overline{X}},\sigma^2_{\overline{X}}\right), \text{ com } \quad \mu_{\overline{X}}=\mu_X \quad \text{e} \quad \sigma^2_{\overline{X}}=\begin{cases}\dfrac{\sigma^2_X}{n} & \text{(população infinita)}\\[2ex] \dfrac{\sigma^2_X}{n}\cdot\dfrac{N-n}{N-1} & \text{(população finita)}\end{cases}.$$

Então $Z=\dfrac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{X}}}\rightarrow N(0,1)$.

Como $\mu_{\overline{X}}=\mu_X$ podemos escrever

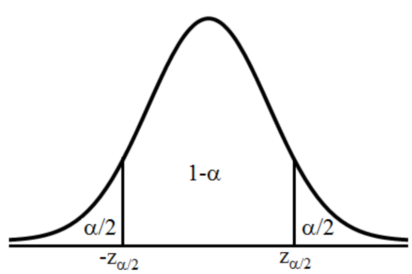
$$P\left(-z_{\alpha/2}\leq\frac{\overline{X}-\mu_X}{\sigma_{\overline{X}}}\leq z_{\alpha/2}\right)=1-\alpha\,.$$

Isolando μ_X chegamos a

$$P\left(\overline{X}-z_{\alpha/2}\cdot\sigma_{\overline{X}}\leq\mu_X\leq\overline{X}+z_{\alpha/2}\cdot\sigma_{\overline{X}}\right)=1-\alpha\,.$$

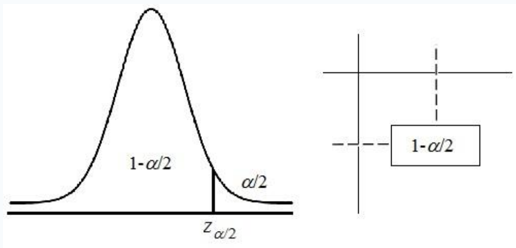
Então, o Intervalo de Confiança para a média populacional, com 1-α de confiança será dado por:

$$IC_{\mu_X}\equiv\begin{cases}\left(\overline{X}-z_{\alpha/2}\cdot\dfrac{\sigma_X}{\sqrt{n}}\,,\,\overline{X}+z_{\alpha/2}\cdot\dfrac{\sigma_X}{\sqrt{n}}\right) & \text{(população infinita)}\\[2ex] \left(\overline{X}-z_{\alpha/2}\cdot\dfrac{\sigma_X}{\sqrt{n}}\cdot\sqrt{\dfrac{N-n}{N-1}}\,,\,\overline{X}+z_{\alpha/2}\cdot\dfrac{\sigma_X}{\sqrt{n}}\cdot\sqrt{\dfrac{N-n}{N-1}}\right) & \text{(população finita)}\end{cases}$$



Observação

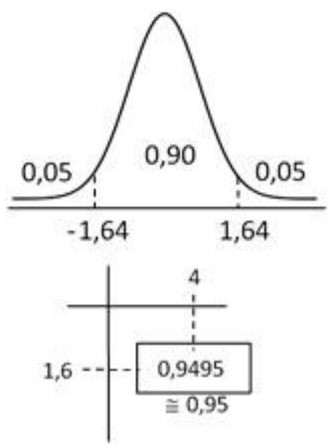
Para determinação de $z_{\alpha/2}$ procuramos no corpo da tabela (anexo I) o valor mais próximo de $1-\alpha/2$, em seguida encontramos o valor de z correspondente:



EXEMPLO 3.3.4[1]

A variância do tempo de acesso dos usuários de determinada página na Internet é de $400\ min^2$. Para estimar o tempo médio de acesso, foram observados 100 usuários, cujo tempo médio amostral foi de $175\ min$. Qual a estimativa por intervalo para a média, com 90% de confiança ?

$$\sigma_X^2 = 400 \Rightarrow \sigma_X = \sqrt{400} = 20$$
$$n = 100 \quad \overline{X} = 175 \quad 1-\alpha = 0{,}90 \Rightarrow z_{\alpha/2} = 1{,}64$$
$$IC_{\mu_X} \equiv \left(175 - 1{,}64 \cdot \frac{20}{\sqrt{100}}\,;175 + 1{,}64 \cdot \frac{20}{\sqrt{100}}\right)$$
$$IC_{\mu_X} \equiv \left(175 - 1{,}64 \cdot \frac{20}{\sqrt{100}}\,;175 + 1{,}64 \cdot \frac{20}{\sqrt{100}}\right)$$
$$IC_{\mu_X} \equiv (175 - 3{,}38\,;175 + 3{,}38)$$
$$IC_{\mu_X} \equiv (171{,}62\,;178{,}38)$$



Conclusões:

- A margem de erro d é de $\pm 3{,}38\ min$
- A probabilidade (confiança) de que a média da população esteja entre $171{,}62\ min$ e $178{,}38 min$ ou seja, dentro da margem de erro, é de 90%

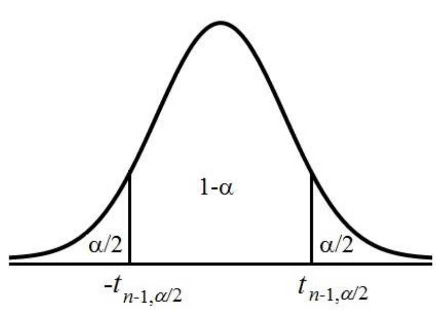
4.1 Intervalo de Confiança para a Média quando σ_X^2 é desconhecida

Observe que no intervalo de confiança para a média visto acima, é necessário o conhecimento da variância populacional σ_X^2. Caso ela seja desconhecida, deve ser substituída por sua estimativa, que é a variância amostral, dada por $S^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}$ ou $S^2 = \dfrac{\sum_{i=1}^{n} x_i^2}{n-1} - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$. Demostra-se que S^2 é um estimador não viciado de σ_X^2.

Com isto, a expressão $\dfrac{\bar{X} - \mu_{\bar{X}}}{\frac{\sigma_X}{\sqrt{n}}}$ no desenvolvimento de $P\left(\left|\bar{X} - \mu_X\right| \le d\right) = 1 - \alpha$ passa a ser $\dfrac{\bar{X} - \mu_{\bar{X}}}{\frac{S}{\sqrt{n}}}$, onde $S = \sqrt{S^2}$, que não tem distribuição Normal (Z) e sim distribuição t-Student que também é simétrica e tem, como a Normal, a forma de sino. Os valores da distribuição t (veja Anexo 2), para um mesmo nível de confiança, dependem do tamanho da amostra (n) e são denotados por t_{n-1}, onde n-1 é chamado de graus de liberdade. Assim, $\dfrac{\bar{X} - \mu_{\bar{X}}}{\frac{S}{\sqrt{n}}}$ é dita ter distribuição t-Student com n-1 graus de liberdade.

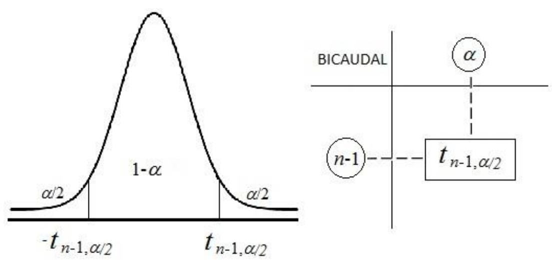
Neste caso, o intervalo de confiança para a média passa a ser:

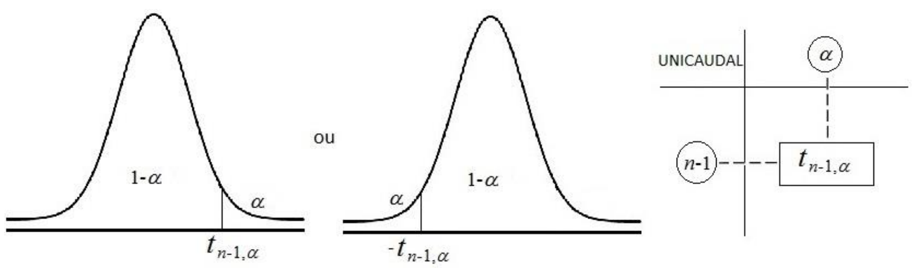
$$IC_{\mu_X} \equiv \begin{cases} \left(\bar{X} - t_{n-1,\alpha/2} \cdot \dfrac{S}{\sqrt{n}}, \bar{X} + t_{n-1,\alpha/2} \cdot \dfrac{S}{\sqrt{n}}\right) & \text{(população infinita)} \\ \left(\bar{X} - t_{n-1,\alpha/2} \cdot \dfrac{S}{\sqrt{n}} \cdot \sqrt{\dfrac{N-n}{N-1}}, \bar{X} + t_{n-1,\alpha/2} \cdot \dfrac{S}{\sqrt{n}} \cdot \sqrt{\dfrac{N-n}{N-1}}\right) & \text{(população finita)} \end{cases}$$



Veja como utilizar a tabela (Anexo 2) da distribuição t-Student:

- No corpo da tabela estão os valores de t
- Na coluna indicadora estão os valores dos graus de liberdade $GL=n\text{-}1$
- No cabeçalho estão os valores de α
    - Quando a situação envolve os dois lados da distribuição, caso dos intervalos de confiança e testes de hipóteses bilaterais, procura-se α na linha BICAUDAL
    - Quando o problema envolve apenas um dos lados da distribuição, caso dos testes de hipóteses unilaterais, procura-se α na linha UNICAUDAL





EXEMPLO 3.3.4.1[1]

Para estimar o tempo médio de acesso a determida página da Internet, foram observados 49 usuários, cujo tempo médio amostral foi de 175 min e variância amostral de 400 min^2. Qual a estimativa por intervalo para a média, com 90% de confiança ?

$$n = 49 \Rightarrow GL = 48$$

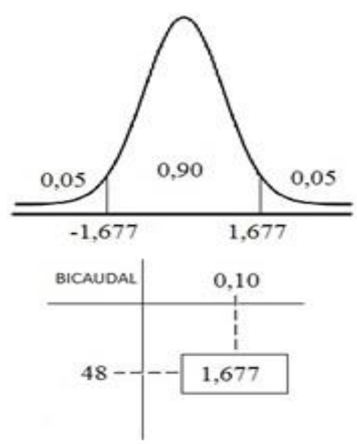
$$\overline{X} = 175 \qquad S^2 = 400 \Rightarrow S = \sqrt{400} = 20$$

$$1 - \alpha = 0{,}90 \Rightarrow \alpha = 0{,}10 \Rightarrow t_{48;0{,}05} = 1{,}677$$

$$IC_{\mu_X} \equiv \left(175 - 1{,}677 \cdot \frac{20}{\sqrt{49}}\,;175 + 1{,}677 \cdot \frac{20}{\sqrt{49}}\right)$$

$$IC_{\mu_X} \equiv (175 - 4{,}79\,;175 + 4{,}79)$$

$$IC_{\mu_X} \equiv (166{,}83\,;179{,}79)$$



Conclusões:

- A margem de erro d é de $\pm 4{,}79\ min$
- A probabilidade (confiança) de que a média da população esteja entre $166{,}83$ min e $179{,}79$ min ou seja, dentro da margem de erro, é de 90%

EXEMPLO 3.3.4.1[2]

Considerando os dados do exemplo 3.3.3.1[1], qual a estimativa por intervalo do tempo médio de duração com 95% de confiança ?

$$n = 36 \Rightarrow GL = 35$$

$$\overline{X} = \frac{2+4+2+3+\ldots+7+7+5}{36} = \frac{169}{36} = 4{,}7$$

$$\sum x = 169; \sum x^2 = 927 \Rightarrow S^2 = \frac{927}{35} - \frac{169^2}{36 \times 35} \Rightarrow$$

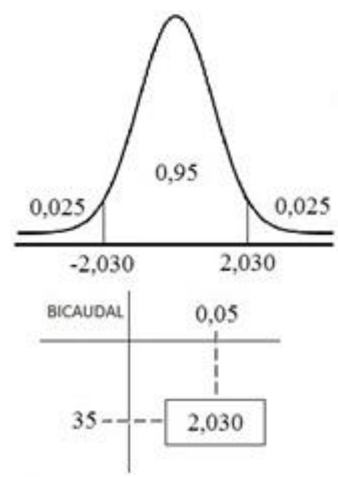
$$\Rightarrow S^2 = 3{,}82 \Rightarrow S = \sqrt{3{,}82} = 1{,}95$$

$$1 - \alpha = 0{,}95 \Rightarrow \alpha = 0{,}05 \Rightarrow t_{36;0,025} = 2{,}030$$

$$IC_{\mu_X} \equiv \left(4{,}7 - 2{,}030 \cdot \frac{1{,}95}{\sqrt{36}} ; 4{,}7 + 2{,}030 \cdot \frac{1{,}95}{\sqrt{36}} \right)$$

$$IC_{\mu_X} \equiv (4{,}7 - 0{,}66 ; 4{,}7 + 0{,}66)$$

$$IC_{\mu_X} \equiv (4{,}04 \ ; \ 5{,}36)$$



Conclusões:

- A margem de erro d é de $\pm 0{,}66\ min$
- A probabilidade (confiança) de que a média da população esteja entre 4,04 min e 5,36 min ou seja, dentro da margem de erro, é de 95%

5 Intervalo de Confiança para a Proporção

Também já foi visto que para amostras grandes ($n \geq 30$), a distribuição de probabilidade de $\hat{p}$ é normal, com média $\mu_{\hat{p}}$ e variância $\sigma^2_{\hat{p}}$:

$$\hat{p} \rightarrow N\left(\mu_{\hat{p}}, \sigma^2_{\hat{p}}\right), \text{com} \quad \mu_{\hat{p}} = p \quad \text{e} \quad \sigma^2_{\bar{X}} = \begin{cases} \dfrac{pq}{n} & \text{(população infinita)} \\ \dfrac{pq}{n} \cdot \dfrac{N-n}{N-1} & \text{(população finita)} \end{cases}, \text{onde}$$

p é a proporção populacional e $q = 1 - p$

Também foi visto que $d = z_{\alpha/2} \cdot \sigma_{\hat{p}}$, assim, o Intervalo de Confiança para a proporção populacional, com 1-α de confiança é dado por:

$$IC_p \equiv \begin{cases} \left(\hat{p} - z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}\hat{q}}{n}} \,,\, \hat{p} + z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}\hat{q}}{n}} \right) & \text{(população infinita)} \\ \left(\hat{p} - z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}\hat{q}}{n} \cdot \dfrac{N-n}{N-1}} \,,\, \hat{p} + z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}\hat{q}}{n} \cdot \dfrac{N-n}{N-1}} \right) & \text{(população finita)} \end{cases}$$

Importante

Como a proporção populacional não é conhecida, pois é o objeto da estimativa, nas expressões de $\sigma^2_{\hat{p}}$ p e q são substituídos por suas estimativas $\hat{p}$ e $\hat{q}$, respectivamente, onde $\hat{q} = 1 - \hat{p}$.

EXEMPLO 3.3.3.5[1]

Uma amostra de 100 componentes apresentou 10% de defeituosos, qual a estimativa por intervalo para a proporção populacional, com 95% de confiança?

$$n = 100 \quad \hat{p} = 0{,}10 \quad \hat{q} = 1 - 0{,}10 = 0{,}90 \quad 1 - \alpha = 0{,}95 \Rightarrow z_{\alpha/2} = 1{,}96$$

$$IC_p \equiv \left(0{,}10 - 1{,}96 \cdot \sqrt{\frac{0{,}10 \times 0{,}90}{100}} \,;\, 0{,}10 + 1{,}96 \cdot \sqrt{\frac{0{,}10 \times 0{,}90}{100}} \right)$$

$$IC_p \equiv (0{,}10 - 1{,}96 \cdot 0{,}03 \,;\, 0{,}10 + 1{,}96 \cdot 0{,}03) \quad IC_p \equiv (0{,}10 - 0{,}06 \,;\, 0{,}10 + 0{,}06)$$

$$IC_p \equiv (0{,}04 \,;\, 0{,}16)$$

Conclusões:

- A margem de erro d é de $\pm 0{,}06$ (seis pontos percentuais)
- A probabilidade (confiança) de que a proporção na população esteja entre $0{,}04$ e $0{,}16$ ou seja, dentro da margem de erro, é de 95%

EXEMPLO 3.3.3.5[2]

Considerando os dados do exemplo 3.3.3.1[1], qual a estimativa por intervalo da proporção de chamadas com ruído, com 90% de confiança ?

$$\hat{p} = \frac{0+0+1+1+\ldots+1+1+0}{36} = \frac{12}{36} = 0{,}33 = 33\% \quad \hat{q} = 1 - 0{,}33 = 0{,}67$$

$$1 - \alpha = 0{,}90 \Rightarrow \alpha = 0{,}10 \Rightarrow z_{0,05} = 1{,}64$$

$$IC_p \equiv \left(0{,}33 - 1{,}64 \cdot \sqrt{\frac{0{,}33 \times 0{,}67}{36}} \,;\, 0{,}33 + 1{,}64 \cdot \sqrt{\frac{0{,}33 \times 0{,}67}{36}} \right)$$

$$IC_p \equiv (0{,}33 - 0{,}13;\, 0{,}33 + 0{,}13) \qquad IC_p \equiv (0{,}20;0{,}46)$$

Conclusões:

- A margem de erro d é de $\pm 0{,}13$ pontos percentuais
- A probabilidade (confiança) de que a proporção de chamadas com ruído, na população, esteja entre 0,20 e 0,46, ou seja, dentro da margem de erro, é de 90%

Capítulo 4

Testes de Hipóteses

1 O que são Testes de Hipóteses

Os testes de hipóteses são procedimentos que têm por objetivo verificar (testar) a veracidade de afirmativas (hipóteses) feitas a respeito de parâmetros populacionais. Assim, uma hipótese H_0, chamada de HIPÓTESE NULA, de que determinado parâmetro populacional θ tem valor θ_0 (H_0: $\theta = \theta_0$), pode ser testada a partir de sua estimativa, através de um estimador $\hat{\theta}$. Retirando-se uma amostra da população, a estimativa de θ (valor de $\hat{\theta}$) é calculada e comparada com seu valor hipotético θ_0, dependendo da proximidade entre $\hat{\theta}$ e θ_0, decide-se em aceitar ou rejeitar a hipótese nula, com uma certa probabilidade de erro. Note que aqui também se está trabalhando com experimentos aleatórios (amostras), de tal forma que o mecanismo de construção/aplicação dos testes de hipóteses é semelhante à construção de intervalos de confiança.

2 Tipos de erros

Os possívei erros que podem ser cometidos no teste de uma hipótese estão descritos no quadro abaixo.

DECISÃO TOMADA	CONDIÇÃO (desconhecida) DE H_0	
	VERDADEIRA	FALSA
ACEITA-SE H_0	decisão correta	**ERRO TIPO *II***
REJEITA-SE H_0	**ERRO TIPO *I***	decisão correta

A probabilidade de se cometer um erro TIPO I é denotada por α e chamada de NÍVEL DE SIGNIFICÂNCIA do teste. Todo o processo de construção/aplicação do teste é baseado nesta probabilidade:

$$
\begin{aligned}
&P(\text{ERRO TIPO } I) = P(\text{REJEITAR } H_0 \mid H_0 \text{ É VERDADEIRA}) \\
&= P(\text{REJEITAR } H_0 \mid \theta = \theta_0) = \alpha
\end{aligned}
$$

3 Tipos de testes

Para o teste de parâmetros populacionais podem ser consideradas as seguintes situações:

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases} \quad \text{(teste bilateral)}$$

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases} \quad \text{(teste unilateral)}$$

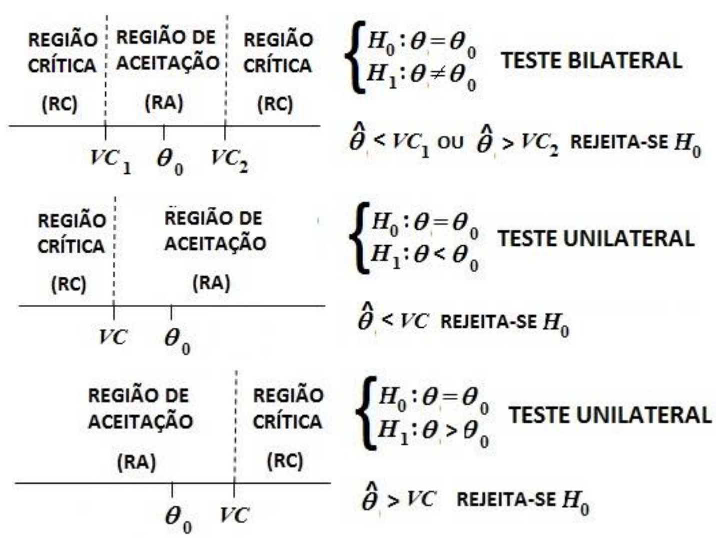
$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases} \quad \text{(teste unilateral)}$$

Observações

- Para os testes unilaterais, as hipóteses nulas podem ser substituídas, quando necessário, por desigualdades, $\geq$ e $\leq$, respectivamente
- Isto acontece quando se afirma que o parâmetro populacional deve ser no mínimo (máximo) igual a um determinado valor como, por exemplo, o percentual de determinada vitamina em um medicamento

3.1 Mecanismo do teste

Como foi dito inicialmente, a partir de uma amostra de tamanho n calcula-se a estimativa de θ (valor de $\hat{\theta}$) e compara-se com θ_0, se $\hat{\theta}$ estiver próximo de θ_0, então aceita-se a hipótese nula. Para verificar a proximidade de $\hat{\theta}$ com θ_0 são determinados valores de referência, denominados VALORES CRÍTICOS, desta forma, dependendo da posição do valor da estimativa em relação aos valores críticos, aceita-se ou não H_0. Os procedimentos estão descritos nas figuras abaixo.



Os valores críticos são determinados a partir do nível de significância (α) do teste, ou seja, a partir do desenvolvimento de $P(\text{rejeitar } H_0 \mid H_0 \text{ é verdade}) = \alpha$:

$$P(\text{rejeitar } H_0 \mid H_0 \text{ é verdade}) = P(\text{rejeitar } H_0 \mid \theta = \theta_0) = P(\hat{\theta} \in RC \mid \theta = \theta_0) = \alpha$$

4 Testes de Hipóteses para a Média

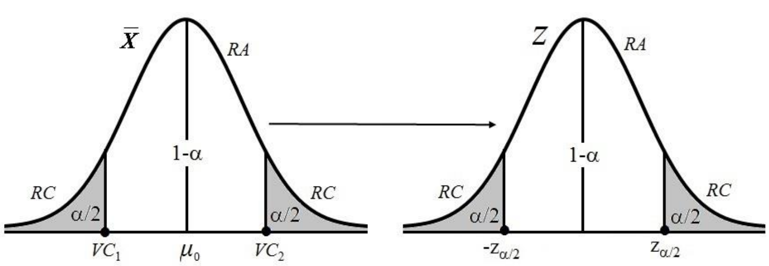
4.1 Teste Bilateral

$$\begin{cases} H_0 : \mu_X = \mu_0 \\ H_1 : \mu_X \neq \mu_0 \end{cases}$$

$$P(\text{rejeitar } H_0 \mid H_0 \text{ é verdade}) = P(\text{rejeitar } H_0 \mid \mu_X = \mu_0) = P(\overline{X} \in RC \mid \mu_X = \mu_0) =$$
$$= P(\overline{X} \leq VC_1 \text{ ou } \overline{X} \geq VC_2) = 1 - P(VC_1 < \overline{X} < VC_2) = \alpha \Rightarrow P(VC_1 < \overline{X} < VC_2) = 1 - \alpha$$

$$P\left(\frac{VC_1 - \mu_0}{\sigma_{\overline{X}}} < \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} < \frac{VC_2 - \mu_0}{\sigma_{\overline{X}}}\right) = 1 - \alpha \Rightarrow \begin{cases} \dfrac{VC_1 - \mu_0}{\sigma_{\overline{X}}} = -z_{\alpha/2} \\ \dfrac{VC_2 - \mu_0}{\sigma_{\overline{X}}} = z_{\alpha/2} \end{cases} \Rightarrow \begin{cases} VC_1 = \mu_0 - z_{\alpha/2} \cdot \sigma_{\overline{X}} \\ VC_2 = \mu_0 + z_{\alpha/2} \cdot \sigma_{\overline{X}} \end{cases}$$

Assim, se $\overline{X} \leq \mu_0 - z_{\alpha/2} \cdot \sigma_{\overline{X}}$ ou $\overline{X} \geq \mu_0 + z_{\alpha/2} \cdot \sigma_{\overline{X}}$ rejeita-se H_0. Onde, como já visto, $\sigma_{\overline{X}} = \dfrac{\sigma_X}{\sqrt{n}}$ para populações infinitas e $\sigma_{\overline{X}} = \dfrac{\sigma_X}{\sqrt{n}} \cdot \sqrt{\dfrac{N-n}{N-1}}$ para populaçoes finitas. Observe que rejeitar H_0 para $\overline{X} \leq \mu_0 - z_{\alpha/2} \cdot \sigma_{\overline{X}}$ ou $\overline{X} \geq \mu_0 + z_{\alpha/2} \cdot \sigma_{\overline{X}}$ é equivalente a rejeitar H_0 para $\dfrac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} \leq -z_{\alpha/2}$ ou $\dfrac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} \geq z_{\alpha/2}$. A expressão $Z_c = \dfrac{\overline{X} - \mu_0}{\sigma_{\overline{X}}}$ é chamada de ESTATÍSTICA DE PROVA e $z_{\alpha/2}$, chamado de *z* tabelado, é o VALOR CRÍTICO de Z_c. Assim, se $|Z_c| \geq z_\alpha$ rejeita-se H_0.

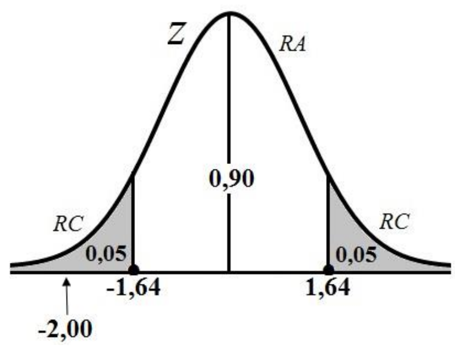


EXEMPLO 3.4.4.1[1]

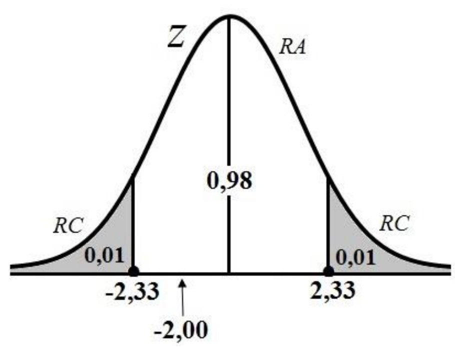
A variância do tempo de acesso dos usuários de determinada página na Internet é de 400 min^2. Para testar a hipótese de que o tempo médio de acesso é de 179 *min*, contra a alternativa de que é diferente, foram observados 100 usuários, cujo tempo médio foi de 175 *min*. Teste a hipótese aos níveis de significância de 10% e 2%.

$$\begin{cases} H_0 : \mu_X = 179 \\ H_1 : \mu_X \neq 179 \end{cases}$$

$$\alpha = 0{,}10 \Rightarrow z_{\alpha/2} = 1{,}64$$
$$\sigma_X^2 = 400 \Rightarrow \sigma_X = \sqrt{400} = 20$$
$$n = 100 \Rightarrow \sigma_{\bar{X}} = \frac{20}{\sqrt{100}} = 2 \qquad \bar{X} = 175$$
$$Z_c = \frac{175 - 179}{2} = -2 \Rightarrow |Z_c| > z_{\alpha/2} \Rightarrow$$
$$\Rightarrow \text{rejeita - se } H_0$$



$$\alpha = 0{,}02 \Rightarrow z_{\alpha/2} = 2{,}33$$
$$\sigma_X^2 = 400 \Rightarrow \sigma_X = \sqrt{400} = 20$$
$$n = 100 \Rightarrow \sigma_{\bar{X}} = \frac{20}{\sqrt{100}} = 2 \qquad \bar{X} = 175$$
$$Z_c = \frac{175 - 179}{2} = -2 \Rightarrow |Z_c| < z_{\alpha/2} \Rightarrow$$
$$\Rightarrow \text{aceita - se } H_0$$



4.2 Testes Unilaterais

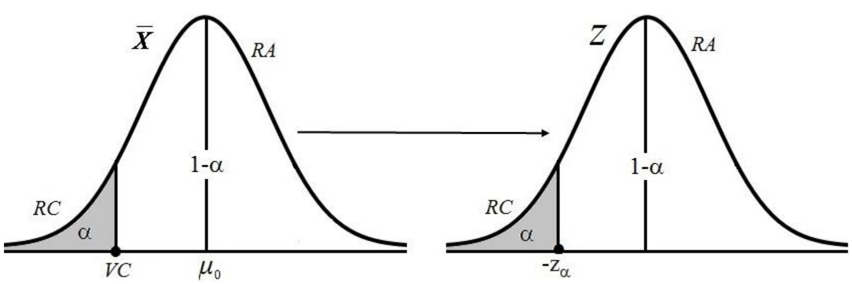
1º caso

$$\begin{cases} H_0 : \mu_X = \mu_0 \\ H_1 : \mu_X < \mu_0 \end{cases}$$

$$P(\text{rejeitar } H_0 \mid H_0 \text{ é verdade}) = P(\text{rejeitar } H_0 \mid \mu_X = \mu_0) = P(\overline{X} \in RC \mid \mu_X = \mu_0)$$
$$= P(\overline{X} \le VC) = \alpha \Rightarrow P\left(\frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} \le \frac{VC - \mu_0}{\sigma_{\overline{X}}}\right) = \alpha \Rightarrow \frac{VC - \mu_0}{\sigma_{\overline{X}}} = -z_\alpha \Rightarrow$$
$$\Rightarrow VC = \mu_0 - z_\alpha \cdot \sigma_{\overline{X}}$$

Assim, se $\overline{X} \le \mu_0 - z_\alpha \cdot \sigma_{\overline{X}}$ rejeita-se H_0, caso contrário, se $\overline{X} > \mu_0 - z_\alpha \cdot \sigma_{\overline{X}}$ aceita-se H_0. E o procedimento equivalente, utilizando a ESTATÍSTICA DE PROVA Z_c, será:

$Z_c = \dfrac{\overline{X} - \mu_0}{\sigma_{\overline{X}}}$, se $Z_c \le -z_{\alpha/2}$, ou seja, se $|Z_c| \ge z_\alpha$, rejeita-se H_0.

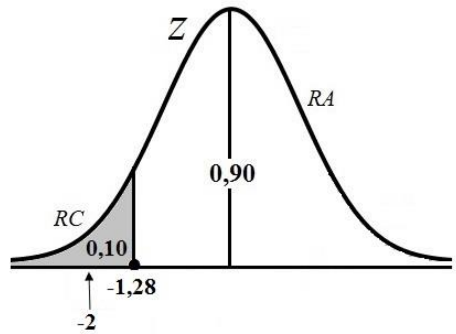


EXEMPLO 3.4.4.2[1]

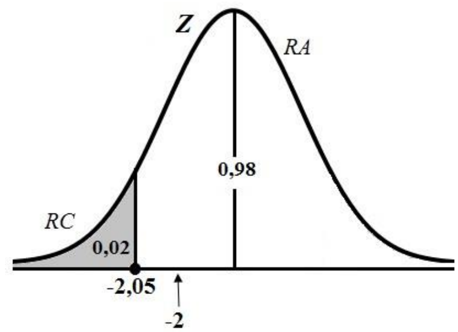
A variância do tempo de acesso dos usuários de determinada página na Internet é de 400 min^2. Para testar a hipótese de que o tempo médio de acesso é de 179 *min*, contra a alternativa de que é menor, foram observados 100 usuários, cujo tempo médio foi de 175 *min*. Teste a hipótese aos níveis de significância de 10% e 2%.

$$\begin{cases} H_0 : \mu_X = 179 \\ H_1 : \mu_X < 179 \end{cases}$$

$$\alpha = 0{,}10 \Rightarrow z_\alpha = 1{,}28$$
$$\sigma_X^2 = 400 \Rightarrow \sigma_X = \sqrt{400} = 20$$
$$n = 100 \Rightarrow \sigma_{\overline{X}} = \frac{20}{\sqrt{100}} = 2 \qquad \overline{X} = 175$$
$$Z_c = \frac{175-179}{2} = -2 \Rightarrow |Z_c| > z_\alpha \Rightarrow$$
$$\Rightarrow \text{rejeita - se } H_0$$



$$\alpha = 0{,}02 \Rightarrow z_\alpha = 2{,}05$$
$$\sigma_X^2 = 400 \Rightarrow \sigma_X = \sqrt{400} = 20$$
$$n = 100 \Rightarrow \sigma_{\overline{X}} = \frac{20}{\sqrt{100}} = 2 \qquad \overline{X} = 175$$
$$Z_c = \frac{175-179}{2} = -2 \Rightarrow |Z_c| < z_\alpha \Rightarrow$$
$$\Rightarrow \text{aceita - se } H_0$$



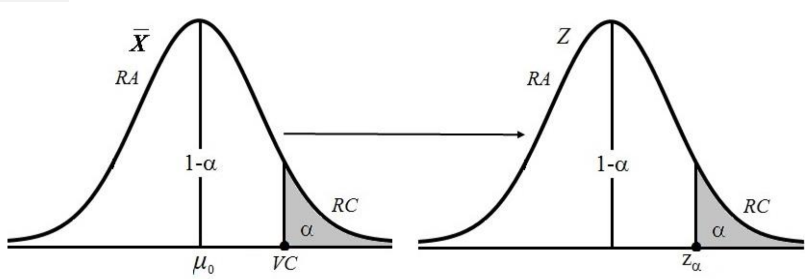
2º caso

$$\begin{cases} H_0 : \mu_X = \mu_0 \\ H_1 : \mu_X > \mu_0 \end{cases}$$

$$P(\text{rejeitar } H_0 \mid H_0 \text{ é verdade}) = P(\text{rejeitar } H_0 \mid \mu_X = \mu_0) = P(\overline{X} \in RC \mid \mu_X = \mu_0)$$
$$= P(\overline{X} \geq VC) = \alpha \Rightarrow P\left(\frac{\overline{X}-\mu_0}{\sigma_{\overline{X}}} \geq \frac{VC-\mu_0}{\sigma_{\overline{X}}}\right) = \alpha \Rightarrow \frac{VC-\mu_0}{\sigma_{\overline{X}}} = z_\alpha \Rightarrow$$
$$\Rightarrow VC = \mu_0 + z_\alpha \cdot \sigma_{\overline{X}}$$

Assim, se $\overline{X} \geq \mu_0 + z_\alpha \cdot \sigma_{\overline{X}}$ rejeita-se H_0, caso contrário, se $\overline{X} < \mu_0 + z_\alpha \cdot \sigma_{\overline{X}}$ aceita-se H_0. E o procedimento equivalente, utilizando a ESTATÍSTICA DE PROVA Z_c, será:

$Z_c = \dfrac{\overline{X}-\mu_0}{\sigma_{\overline{X}}}$, se $Z_c \geq z_{\alpha/2}$, ou seja, se $|Z_c| \geq z_\alpha$, rejeita-se H_0.

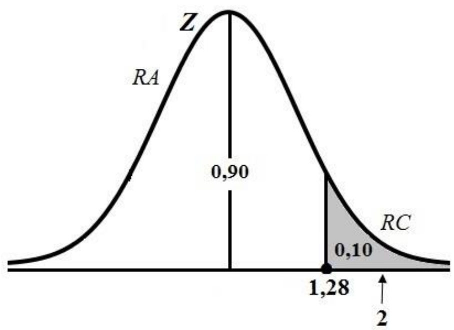


EXEMPLO 3.4.4.2[2]

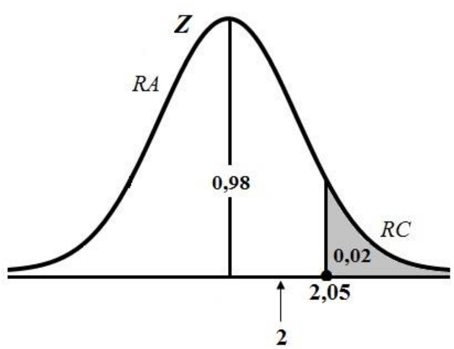
A variância do tempo de acesso dos usuários de determinada página na Internet é de 400 min^2. Para testar a hipótese de que o tempo médio de acesso é de 179 min, contra a alternativa de que é maior, foram observados 100 usuários, cujo tempo médio foi de 183 min. Teste a hipótese aos níveis de significância de 10% e 2%.

$$\begin{cases} H_0 : \mu_X = 179 \\ H_1 : \mu_X > 179 \end{cases}$$

$$\alpha = 0{,}10 \Rightarrow z_\alpha = 1{,}28$$
$$\sigma_X^2 = 400 \Rightarrow \sigma_X = \sqrt{400} = 20$$
$$n = 100 \Rightarrow \sigma_{\overline{X}} = \frac{20}{\sqrt{100}} = 2 \qquad \overline{X} = 183$$
$$Z_c = \frac{183 - 179}{2} = 2 \Rightarrow |Z_c| > z_\alpha \Rightarrow$$
$$\Rightarrow \text{ rejeita - se } H_0$$



$$\alpha = 0{,}02 \Rightarrow z_\alpha = 2{,}05$$
$$\sigma_X^2 = 400 \Rightarrow \sigma_X = \sqrt{400} = 20$$
$$n = 100 \Rightarrow \sigma_{\overline{X}} = \frac{20}{\sqrt{100}} = 2 \qquad \overline{X} = 183$$
$$Z_c = \frac{183 - 179}{2} = 2 \Rightarrow |Z_c| < z_\alpha \Rightarrow$$
$$\Rightarrow \text{ aceita - se } H_0$$



5 Testes de Hipóteses para a Média quando σ_X^2 é desconhecida

Como visto na parte sobre intervalos de confiança, quando não se conhece a variância populacional, σ_X^2, ela deve ser substituída por sua estimativa S^2 e a distribuição de probabilidade a ser utilizada é a t-Student. Assim, as estatísticas de prova passam a ser as seguintes.

5.1 Teste Bilateral

$$\begin{cases} H_0 : \mu_X = \mu_0 \\ H_1 : \mu_X \neq \mu_0 \end{cases}$$

As estatísticas de prova são $T_c = \dfrac{\overline{X} - \mu_0}{\dfrac{S}{\sqrt{n}}}$ para populações infinitas e $T_c = \dfrac{\overline{X} - \mu_0}{\dfrac{S}{\sqrt{n}} \cdot \sqrt{\dfrac{N-n}{N-1}}}$ para populações finitas, se $|T_c| \geq t_{n-1,\alpha/2}$, rejeita-se H_0.

EXEMPLO 3.4.5.1[1]

Para testar a hipótese de que o tempo médio de acesso é de 179 *min*, contra a alternativa de que é diferente, foram observados 81 usuários, cujo tempo médio foi de 175 *min* e variância 400 *min*². Teste a hipótese aos níveis de significância de 10% e 2%.

$$\begin{cases} H_0 : \mu_X = 179 \\ H_1 : \mu_X \neq 179 \end{cases}$$

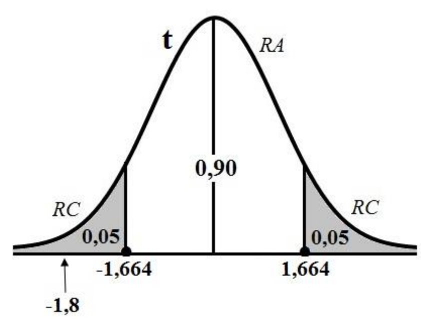
$GL = 81 - 1 = 80 \qquad \alpha = 0{,}10 \Rightarrow t_{n-1,\alpha/2} = 1{,}664$

$S^2 = 400 \Rightarrow S = \sqrt{400} = 20$

$n = 36 \Rightarrow \sigma_{\overline{X}} = \dfrac{20}{\sqrt{81}} \qquad \overline{X} = 175$

$T_c = \dfrac{175 - 179}{\dfrac{20}{9}} = -1{,}8 \Rightarrow |T_c| > t_{n-1,\alpha/2} \Rightarrow$

$\Rightarrow$ rejeita - se H_0



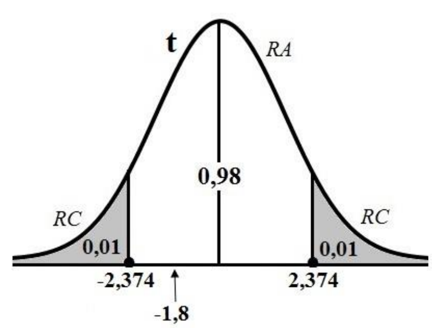
$GL = 81 - 1 = 80 \qquad \alpha = 0{,}02 \Rightarrow t_{n-1,\alpha/2} = 2{,}374$

$S^2 = 400 \Rightarrow S = \sqrt{400} = 20$

$n = 81 \Rightarrow \sigma_{\overline{X}} = \dfrac{20}{\sqrt{81}} = \qquad \overline{X} = 175$

$T_c = \dfrac{175 - 179}{\dfrac{20}{9}} = -1{,}8 \Rightarrow |T_c| < t_{n-1,\alpha/2} \Rightarrow$

$\Rightarrow$ aceita - se H_0



5.2 Testes Unilaterais

1º caso

$$\begin{cases} H_0 : \mu_X = \mu_0 \\ H_1 : \mu_X < \mu_0 \end{cases}$$

As estatísticas de prova são $T_c = \dfrac{\overline{X} - \mu_0}{\dfrac{S}{\sqrt{n}}}$ para populações infinitas e $T_c = \dfrac{\overline{X} - \mu_0}{\dfrac{S}{\sqrt{n}} \cdot \sqrt{\dfrac{N-n}{N-1}}}$ para populações finitas, se $|T_c| \geq t_{n-1,\alpha/2}$, rejeita-se H_0.

EXEMPLO 3.4.5.2[1]

Para testar a hipótese de que o tempo médio de acesso é de 179 *min*, contra a alternativa de que é menor, foram observados 81 usuários, cujo tempo médio foi de 175 *min* e variância 400 *min*². Teste a hipótese aos níveis de significância de 10% e 2%.

$$\begin{cases} H_0 : \mu_X = 179 \\ H_1 : \mu_X < 179 \end{cases}$$

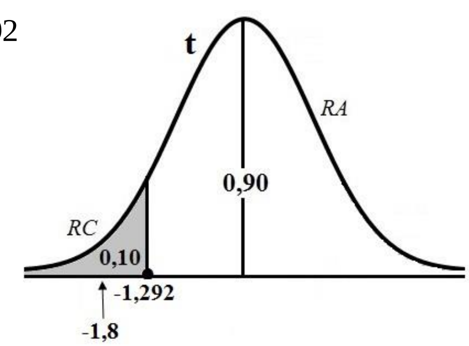
$$GL = 81 - 1 = 80 \qquad \alpha = 0{,}10 \Rightarrow t_{n-1,\alpha} = 1{,}292$$

$$S^2 = 400 \Rightarrow S = \sqrt{400} = 20$$

$$n = 81 \Rightarrow \sigma_{\overline{X}} = \frac{20}{\sqrt{81}} \qquad \overline{X} = 175$$

$$T_c = \frac{175 - 179}{\dfrac{20}{\sqrt{81}}} = -1{,}8 \Rightarrow |T_c| > t_{n-1,\alpha} \Rightarrow$$

$$\Rightarrow \text{rejeita - se } H_0$$



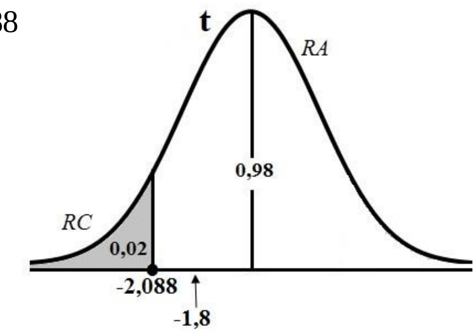
$GL = 81 - 1 = 80 \quad \alpha = 0{,}02 \Rightarrow t_{n-1,\alpha} = 2{,}088$

$S^2 = 400 \Rightarrow S = \sqrt{400} = 20$

$n = 81 \Rightarrow \sigma_{\overline{X}} = \dfrac{20}{\sqrt{81}} \quad \overline{X} = 175$

$T_c = \dfrac{175 - 179}{\dfrac{20}{\sqrt{81}}} = -1{,}8 \Rightarrow |T_c| < t_{n-1,\alpha} \Rightarrow$

$\Rightarrow$ aceita - se H_0



2º caso

$$\begin{cases} H_0 : \mu_X = \mu_0 \\ H_1 : \mu_X > \mu_0 \end{cases}$$

As estatísticas de prova são $T_c = \dfrac{\overline{X} - \mu_0}{\dfrac{S}{\sqrt{n}}}$ para populações infinitas e $T_c = \dfrac{\overline{X} - \mu_0}{\dfrac{S}{\sqrt{n}} \cdot \sqrt{\dfrac{N-n}{N-1}}}$ para populações finitas, se $|T_c| \geq t_{n-1,\alpha/2}$, rejeita-se H_0.

EXEMPLO 3.4.5.2[2]

Para testar a hipótese de que o tempo médio de acesso é de 179 min, contra a alternativa de que é maior, foram observados 81 usuários, cujo tempo médio foi de 183 *min* e variância 400 *min*². Teste a hipótese aos níveis de significância de 10% e 2%.

$$\begin{cases} H_0 : \mu_X = 179 \\ H_1 : \mu_X > 179 \end{cases}$$

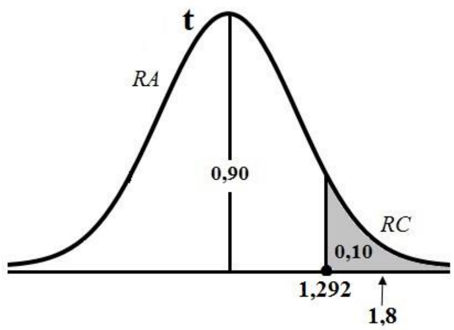
$$GL = 81 - 1 = 80 \qquad \alpha = 0{,}10 \Rightarrow t_{n-1,\alpha} = 1{,}292$$

$$S^2 = 400 \Rightarrow S = \sqrt{400} = 20$$

$$n = 81 \Rightarrow \sigma_{\overline{X}} = \frac{20}{\sqrt{81}} \qquad \overline{X} = 183$$

$$T_c = \frac{183 - 179}{\dfrac{20}{\sqrt{81}}} = 1{,}8 \Rightarrow |T_c| > t_{n-1,\alpha} \Rightarrow$$

$$\Rightarrow \text{rejeita - se } H_0$$



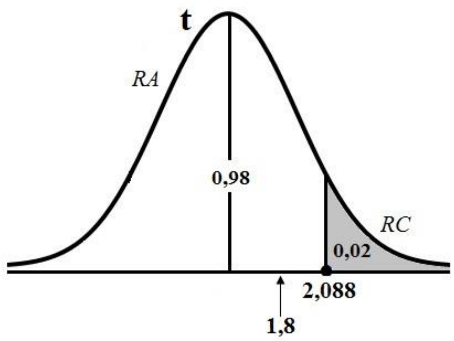
$$GL = 81 - 1 = 80 \qquad \alpha = 0{,}02 \Rightarrow t_{n-1,\alpha} = 2{,}088$$

$$S^2 = 400 \Rightarrow S = \sqrt{400} = 20$$

$$n = 81 \Rightarrow \sigma_{\overline{X}} = \frac{20}{\sqrt{81}} \qquad \overline{X} = 183$$

$$T_c = \frac{183 - 179}{\dfrac{20}{\sqrt{81}}} = 1{,}8 \Rightarrow |T_c| < t_{n-1,\alpha} \Rightarrow$$

$$\Rightarrow \text{aceita - se } H_0$$



6 Testes de Hipóteses para a Proporção

6.1 Teste Bilateral

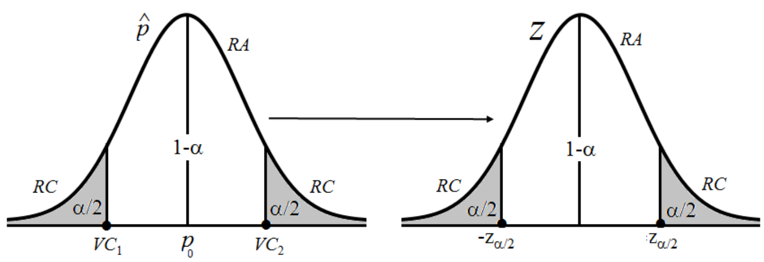
$$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}$$

$$P(\text{rejeitar } H_0 \mid H_0 \text{ é verdade}) = P(\text{rejeitar } H_0 \mid p = p_0) = P(\hat{p} \in RC \mid p = p_0) =$$
$$= P(\hat{p} \leq VC_1 \text{ ou } \hat{p} \geq VC_2 \mid p = p_0) = 1 - P(VC_1 < \hat{p} < VC_2 \mid p = p_0) = \alpha \Rightarrow$$
$$\Rightarrow P(VC_1 < \hat{p} < VC_2) = 1 - \alpha$$

$$\Rightarrow P\left(\frac{VC_1 - p_0}{\sigma_{\hat{p}}} < \frac{\hat{p} - \mu_0}{\sigma_{\hat{p}}} < \frac{VC_2 - p_0}{\sigma_{\hat{p}}}\right) = 1 - \alpha \Rightarrow \begin{cases} \dfrac{VC_1 - p_0}{\sigma_{\hat{p}}} = -z_{\alpha/2} \\ \dfrac{VC_2 - p_0}{\sigma_{\hat{p}}} = z_{\alpha/2} \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} VC_1 = p_0 - z_{\alpha/2} \cdot \sigma_{\hat{p}} \\ VC_2 = p_0 + z_{\alpha/2} \cdot \sigma_{\hat{p}} \end{cases}$$

Assim, se $\hat{p} \leq p_0 - z_{\alpha/2} \cdot \sigma_{\hat{p}}$ ou $\hat{p} \geq p_0 + z_{\alpha/2} \cdot \sigma_{\hat{p}}$ rejeita-se H_0. Onde, como já visto, $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$ para populações infinitas e $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n} \cdot \frac{N-n}{N-1}}$ para populaçoes finitas. Como p e q são desconhecidos eles são substituídos por seus valores da hipótese nula p_0 e q_0, respectivamente. Assim teremos as seguintes expressões para o desvio padrão de $\hat{p}$: $\sigma_{\hat{p}} = \sqrt{\frac{p_0 q_0}{n}}$ e $\sigma_{\hat{p}} = \sqrt{\frac{p_0 q_0}{n} \cdot \frac{N-n}{N-1}}$. Observe que rejeitar H_0 para $\hat{p} \leq p_0 - z_{\alpha/2} \cdot \sigma_{\hat{p}}$ ou $\hat{p} \geq p_0 + z_{\alpha/2} \cdot \sigma_{\hat{p}}$ é equivalente a rejeitar H_0 para $\frac{\hat{p} - p_0}{\sigma_{\hat{p}}} \leq -z_{\alpha/2}$ ou $\frac{\hat{p} - p_0}{\sigma_{\hat{p}}} \geq z_{\alpha/2}$. A expressão $Z_c = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}}$ é chamada de ESTATÍSTICA DE PROVA e $z_{\alpha/2}$ é o VALOR CRÍTICO de Z_c , chamado de z tabelado. Assim, se $|Z_c| \geq z_\alpha$ rejeita-se H_0.



EXEMPLO 3.4.6.1[1]

Para testar a hipótese de que o percentual de acessos mal sucedidos a determinada página da Internet é 10%, contra a alternativa de que é diferente, foram observados 100 acessos, dos quais 15 foram mal sucedidos. Teste a hipótese aos níveis de significância de 10% e 2%.

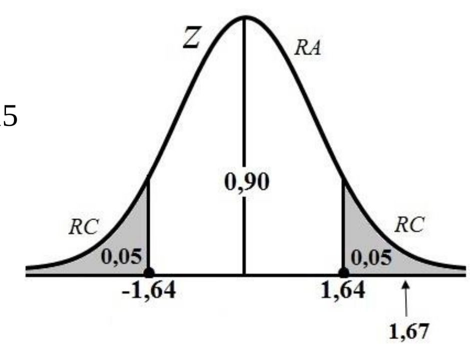
$$\begin{cases} H_0 : p = 0{,}10 \\ H_1 : p \neq 0{,}10 \end{cases}$$

$$\alpha = 0{,}10 \Rightarrow z_{\alpha/2} = 1{,}64$$

$$n = 100 \Rightarrow \sigma_{\hat{p}} = \sqrt{\frac{0{,}10 \times 0{,}90}{100}} = 0{,}03 \quad \hat{p} = 0{,}15$$

$$Z_c = \frac{0{,}15 - 0{,}10}{0{,}03} = 1{,}67 \Rightarrow |Z_c| > z_{\alpha/2} \Rightarrow$$

$$\Rightarrow \text{rejeita - se } H_0$$

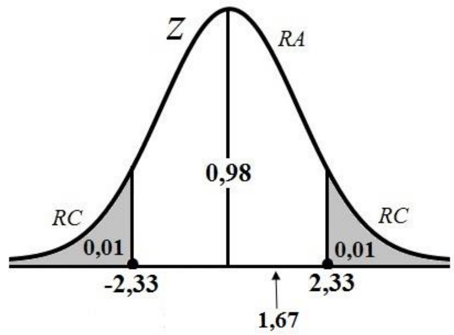


$$\alpha = 0{,}02 \Rightarrow z_{\alpha/2} = 2{,}33$$

$$n = 100 \Rightarrow \sigma_{\hat{p}} = \sqrt{\frac{0{,}10 \times 0{,}90}{100}} = 0{,}03 \quad \hat{p} = 0{,}15$$

$$Z_c = \frac{0{,}15 - 0{,}10}{0{,}03} = 1{,}67 \Rightarrow |Z_c| < z_{\alpha/2} \Rightarrow$$

$$\Rightarrow \text{aceita - se } H_0$$



6.2 Testes Unilaterais

1º caso

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p < p_0 \end{cases}$$

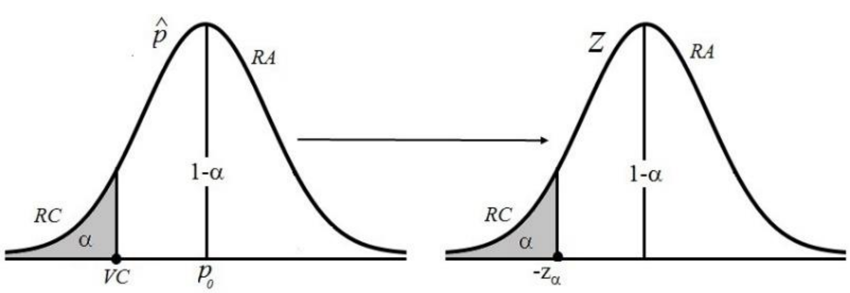
$$P(\text{rejeitar } H_0 \mid H_0 \text{ é verdade}) = P(\text{rejeitar } H_0 \mid p = p_0) = P(\hat{p} \in RC \mid p = p_0) =$$

$$= P(\hat{p} \le VC \mid p = p_0) = \alpha \Rightarrow P\left(\frac{\hat{p} - p_0}{\sigma_{\hat{p}}} \le \frac{VC - p_0}{\sigma_{\hat{p}}}\right) = \alpha \Rightarrow \frac{VC - p_0}{\sigma_{\hat{p}}} = -z_\alpha \Rightarrow$$

$$\Rightarrow VC = p_0 - z_\alpha \cdot \sigma_{\hat{p}}$$

Assim, se $\hat{p} \le p_0 - z_\alpha \cdot \sigma_{\hat{p}}$ rejeita-se H_0, caso contrário, se $\hat{p} > p_0 - z_\alpha \cdot \sigma_{\hat{p}}$ aceita-se H_0. E o procedimento equivalente, utilizando a ESTATÍSTICA DE PROVA Z_c, será:

$Z_c = \dfrac{\hat{p} - p_0}{\sigma_{\hat{p}}}$, se $Z_c \le -z_\alpha$, ou seja, se $|Z_c| \ge z_\alpha$, rejeita-se H_0.



EXEMPLO 3.4.6.2[1]

Para testar a hipótese de que o percentual de acessos mal sucedidos a determinada página da Internet é de 10%, contra a alternativa de que é menor, foram observados 100 acessos, dos quais 5 foram mal sucedidos. Teste a hipótese aos níveis de significância de 10% e 2%.

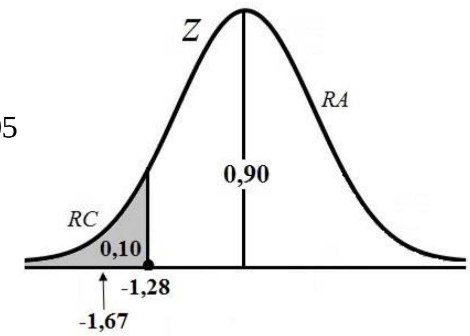
$$\begin{cases} H_0 : p = 0{,}10 \\ H_1 : p < 0{,}10 \end{cases}$$

$$\alpha = 0{,}10 \Rightarrow z_\alpha = 1{,}28$$

$$n = 100 \Rightarrow \sigma_{\hat{p}} = \sqrt{\frac{0{,}10 \times 0{,}90}{100}} = 0{,}03 \quad \hat{p} = 0{,}05$$

$$Z_c = \frac{0{,}05 - 0{,}10}{0{,}03} = -1{,}67 \Rightarrow |Z_c| > z_\alpha \Rightarrow$$

$$\Rightarrow \text{rejeita - se } H_0$$

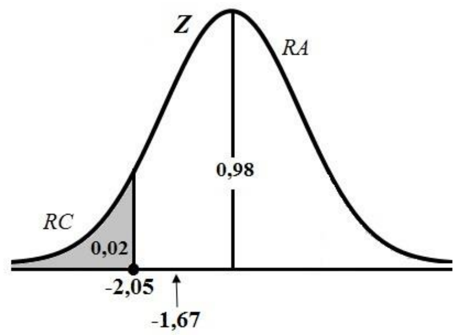


$$\alpha = 0{,}02 \Rightarrow z_\alpha = 2{,}05$$

$$n = 100 \Rightarrow \sigma_{\hat{p}} = \sqrt{\frac{0{,}10 \times 0{,}90}{100}} = 0{,}03 \quad \hat{p} = 0{,}05$$

$$Z_c = \frac{0{,}05 - 0{,}10}{0{,}03} = -1{,}67 \Rightarrow |Z_c| < z_\alpha \Rightarrow$$

$$\Rightarrow \text{aceita - se } H_0$$



2º caso

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p > p_0 \end{cases}$$

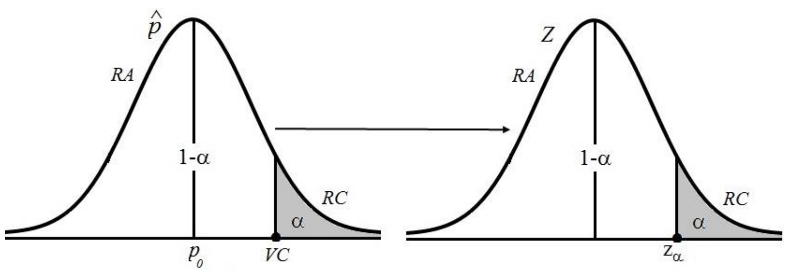
$$P(\text{rejeitar } H_0 \mid H_0 \text{ é verdade}) = P(\text{rejeitar } H_0 \mid p = p_0) = P(\hat{p} \in RC \mid p = p_0)$$

$$= P(\hat{p} \geq VC \mid p = p_0) = \alpha \Rightarrow P\left(\frac{\hat{p} - p_0}{\sigma_{\hat{p}}} \geq \frac{VC - p_0}{\sigma_{\hat{p}}}\right) = \alpha \Rightarrow \frac{VC - p_0}{\sigma_{\hat{p}}} = z_\alpha \Rightarrow$$

$$\Rightarrow VC = p_0 + z_\alpha \cdot \sigma_{\hat{p}}$$

Assim, se $\hat{p} \geq p_0 + z_\alpha \cdot \sigma_{\hat{p}}$ rejeita-se H_0, caso contrário, se $\hat{p} < p_0 + z_\alpha \cdot \sigma_{\hat{p}}$ aceita-se H_0. O procedimento equivalente, utilizando a ESTATÍSTICA DE PROVA Z_c, será:

$Z_c = \dfrac{\hat{p} - p_0}{\sigma_{\hat{p}}}$, se $Z_c \geq z_\alpha$, ou seja, se $|Z_c| \geq z_\alpha$, rejeita-se H_0.



EXEMPLO 3.4.6.2[2]

Para testar a hipótese de que o percentual de acessos mal sucedidos a determinada página da Internet é de 10%, contra a alternativa de que é maior, foram observados 100 acessos, dos quais 15 foram mal sucedidos. Teste a hipótese aos níveis de significância de 10% e 2%.

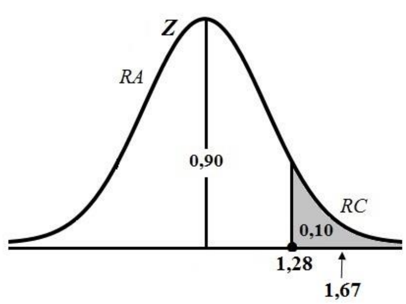
$$\begin{cases} H_0 : p = 0{,}10 \\ H_1 : p > 0{,}10 \end{cases}$$

$$\alpha = 0{,}10 \Rightarrow z_\alpha = 1{,}28$$

$$n = 100 \Rightarrow \sigma_{\hat{p}} = \sqrt{\frac{0{,}10 \times 0{,}90}{100}} = 0{,}03 \quad \hat{p} = 0{,}15$$

$$Z_c = \frac{0{,}15 - 0{,}10}{0{,}03} = 1{,}67 \Rightarrow |Z_c| > z_\alpha \Rightarrow$$

$$\Rightarrow \text{rejeita - se } H_0$$

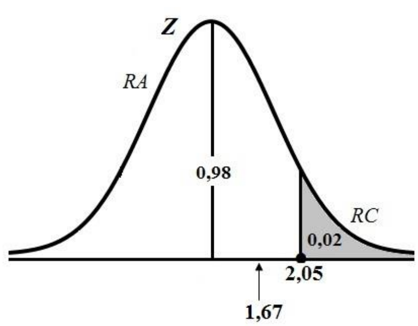


$$\alpha = 0{,}02 \Rightarrow z_\alpha = 2{,}05$$

$$n = 100 \Rightarrow \sigma_{\hat{p}} = \sqrt{\frac{0{,}10 \times 0{,}90}{100}} = 0{,}03 \quad \hat{p} = 0{,}15$$

$$Z_c = \frac{0{,}15 - 0{,}10}{0{,}03} = 1{,}67 \Rightarrow |Z_c| < z_\alpha \Rightarrow$$

$$\Rightarrow \text{aceita - se } H_0$$



Atividades de avaliação

1) Para avaliar o tempo (horas) mensal gasto na Internet e o principal uso da mesma, foi realizada uma pesquisa com uma amostra de 50 alunos de uma escola que tem um total de 800 alunos. O quadro abaixo mostra o resultado obtido. (a) calcule a média, a variância e o desvio padrão do tempo gasto com a Internet (b) qual o tempo total gasto na utilização da Internet pelos alunos da escola ? (c) qual o percentual de alunos que utilizam a Internet para pesquisa escolar ? (d) quantos alunos na escola utilizam a Internet para pesquisa escolar ?

	TEMPO	USO		TEMPO	USO		TEMPO	USO		TEMPO	USO		TEMPO	USO
1	120	PE	11	140	PE	21	140	PE	31	120	PE	41	120	PE
2	100	PE	12	160	OUT	22	100	PE	32	110	PE	42	110	PE
3	250	OUT	13	160	OUT	23	130	OUT	33	110	PE	43	130	OUT
4	180	OUT	14	170	OUT	24	160	OUT	34	120	OUT	44	120	OUT
5	100	OUT	15	120	PE	25	170	PE	35	130	PE	45	150	PE
6	100	PE	16	100	OUT	26	130	PE	36	170	PE	46	120	PE
7	130	PE	17	100	PE	27	130	OUT	37	180	OUT	47	130	PE
8	180	PE	18	160	PE	28	140	OUT	38	160	PE	48	120	OUT
9	100	OUT	19	140	PE	29	110	OUT	39	160	OUT	49	110	PE
10	110	PE	20	130	OUT	30	120	PE	40	110	PE	50	110	PE

PE = PESQUISA ESCOLAR OUT=OUTROS

2) [STEVENSON] Se se extrai uma amostra de uma distribuição normal, qual a probabilidade de que a média amostral esteja compreendida em cada um dos intervalos abaixo?
3) [adaptado de STEVENSON] A espessura média das baterias fabricadas por determinado processo é $0{,}2\ cm$ e o desvio padrão $0{,}01\ cm$. Para uma amostra de 36 bateiras, responda: (a) Que percentagem de médias amostrais estará no intervalo $0{,}2 \pm 0{,}004\ cm$? (b) Qual a probabilidade de se obter uma média amostral que se afaste por mais de $0{,}005\ cm$ da média do processo?
4) Se o tempo médio (por usuário) de utilização de uma determinada página da Internet é 24 minutos, com distribuição normal e desvio padrão de 3 minutos, qual é a probabilidade de que uma amostra aleatória de 100 usuários apresente tempo médio que difira por mais de 30 segundos da média da população?
5) Os pesos dos habitantes de certa cidade têm desvio padrão de $45Kg$. Tomada uma amostra de 225 habitantes, qual a probabilidade de a média amostral se

afastar por $7{,}5Kg$ ou mais da média de todas os 20.000 habitantes da cidade? E se a população consistisse de 2.000 habitantes? E se a população for considerada infinita?

6) [STEVENSON] Supondo uma amostra suficientemente grande, determine a percentagem de proporções amostrais que poderemos esperar nos seguintes intervalos: $a)\ p \pm 1{,}64\sigma_{\hat{p}}$ $b)\ p \pm 1{,}96\sigma_{\hat{p}}$ $c)\ p \pm 2{,}00\sigma_{\hat{p}}$ $d)\ p \pm 2{,}33\sigma_{\hat{p}}$
7) [STEVENSON] Se vamos extrair amostras de tamanho n=100 observações, em que a proporção populacional é 20%, que percentagem de proporções amostrais podemos esperar nos intervalos abaixo?
   a) 16% a 24% b) maior que 24% c) 12% a 28% d) menos de 12% ou mais de 28%
8) [STEVENSON] O Departamento de Compras de uma companhia rejeita rotineiramente remessas de peças se uma amostra aleatória de 100 peças, extraída de um lote com 10.000 peças, acusa 10 ou mais defeituosas. Determine a probabilidade de se ter um lote rejeitado caso ele tenha uma percentagem de peças defeituosas de: (a) 3% (b) 5% (c) 18% (d) 25%
9) Considerando-se desconhecida a média do exercício 2, responda:
   a) Qual o tamanho da amostra para se estimar a média populacional com margem de erro de $0{,}004$ cm e 90% de confiança?
   b) Para uma amostra de tamanho 100, com margem de erro de $0{,}005$ cm, qual o nível de confiança?
   c) Para uma amostra de tamanho 49, qual a margem de erro, com 80% de confiança?
10) Qual o tamanho da amostra para se estimar a proporção populacional, com margem de erro de 5% e 90% de confiança, nos seguintes casos:
    a) Se a proporção populacional for 70%
    b) Se a proporção populacional for desconhecida e se deseja que a amostra seja máxima.
    c) O mesmo para os itens (a) e (b), considerando que a população é finita, com 5.000 elementos.
11) Deseja-se estimar a média dos pesos de certa peça. Pelas especificações do produto, o desvio padrão dos pesos é 10 Kg. Calcule o tamanho da amostra para um erro de estimativa de $1{,}5$ Kg com 95% de confiança, nos seguintes casos: (a) considerando a população infinita e (b) considerando a população finita, com 600 elementos.
12) Deseja-se estimar a proporção de alunos de uma escola que usam o *Chat* disponibilizado pelo AVA (Ambiente Virtual de Aprendizagem) utilizado pela mesma. Calcule o tamanho da amostra para que o erro de estimativa seja de 2% com 99% de confiança, nos seguintes casos: (a) supondo-se que a proporção é de 30% e que a população (total de alunos da escola) é infinita; (b) supondo-se que a proporção é 30% e que a população é finita, com 10.000 alunos; (c) análogo ao item (a), para que a amostra seja a maior possível; (d) análogo ao item (b), para que a amostra seja a maior possível

13) O diâmetro de um cabo elétrico é normalmente distribuído com média $0{,}8$ e variância $0{,}0004$. (a) Qual a probabilidade de que uma amostra de 16 cabos tenha média superior a $0{,}807$? (b) Qual o tamanho da amostra para se estimar a média populacional com margem de erro de $0{,}001$ e 90% de confiança ? (c) Qual a margem de erro para uma amostra de 100 cabos, com 95% de confiança ? (d) Com que probabilidade (nível de confiança) a estimativa está dentro da margem de erro de $0{,}002$, numa amostra de 200 cabos ?
14) A proporção de peças defeituosas em determinado processo de fabricação é 10%. (a) Num lote (amostra) de 100 peças, qual a probabilidade de que o percentual de peças defeituosas não ultrapasse a proporção populacional em mais de $0{,}02$? (b) Qual o tamanho da amostra para se estimar a proporção de peças defeituosas com margem de erro de 5% com $90\ \%$ de confiança ? (c) Resolver o item (b) para que se tenha amostra máxima. (d) Para a amostra máxima, qual a margem de erro com 95% de confiança ?
15) Foram retiradas 25 peças da produção diária de uma máquina, encontrando-se um comprimento médio de $5{,}2\ mm$. Sabendo-se que os comprimentos dessas peças têm distribuição normal com desvio-padrão $1{,}2\ mm$, construa intervalos de confiança para a média aos níveis de 90% e 95% de confiança. Repita o exercício, considerando-se o desvio padrão da população desconhecido e que a amostra forneceu S=$1{,}2\ mm$.
16) De uma população normal com variância $1{,}96$, obteve-se a seguinte amostra (pesos em Kg): $25,\ 26,\ 26,\ 27,\ 28,\ 28,\ 29,\ 30,\ 25,\ 25,\ 28,\ 28,\ 30,\ 30,\ 25,\ 27$. Quais as estimativas por ponto e por intervalo, 98% de confiança, para a média populacional? Repita o problema considerando que a população é finita, com 100 elementos.
17) Quais as estimativas por ponto e por intervalo (98% de confiança) da proporção de elementos com peso superior a $26\ Kg$, na população do problema 16 ? Repita o problema considerando que a população é finita, com 100 elementos.
18) Foi realizada uma pesquisa com uma amostra de 300 alunos de um curso a distância sobre a qualidade do ambiente virtual de aprendizagem utilizado, 200 deles se mostraram insatisfeitos com a qualidade do mesmo. Quais as estimativas por ponto e por intervalo (90% de confiança) para a proporção de alunos insatisfeitos na população ?
19) Para as situações do problema 14, teste a hipótese de que o comprimento médio da população é igual a $4{,}5\ mm$, contra a alternativa de que é maior que $4{,}5mm$.
20) Para o exercício 15, teste a hipótese de que a média populacional é 28, contra a alternativa de que é diferente de 28.
21) Para o problema 15, teste a hipótese de que o percentual de pessoas com peso superior a $26\ Kg$, na população, é 64%, contra a alternativa de que é menor que 64%.

Solução das atividades

1)

$$a)\quad \overline{X} = \frac{\sum_{i=1}^{50} x_i}{50} = \frac{120+100+150+\ldots+110+110}{50} = \frac{6670}{50} = 133{,}40h$$

$$b)\quad T = 133{,}40 \times 800 = 106720h$$

$$c)\quad \hat{p} = \frac{n_A}{50} = \frac{30}{50} = 60\%$$

$$d)\quad N_A = \hat{p} \times N = \frac{30}{50} \times 800 = 480$$

2)

$$\overline{X} \to N\left(\mu_{\overline{X}}, \sigma_{\overline{X}}\right); \mu_{\overline{X}} = \mu_X$$

$$a)\quad P(\mu_X - 1{,}64\sigma_{\overline{X}} \le \overline{X} \le \mu_X + 1{,}64\sigma_{\overline{X}}) =$$

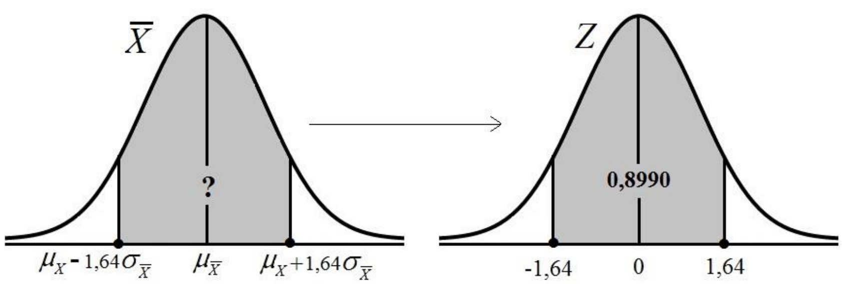
$$= P\left(\frac{(\mu_X - 1{,}64\sigma_{\overline{X}}) - \mu_{\overline{X}}}{\sigma_{\overline{X}}} \le Z \le \frac{(\mu_X + 1{,}64\sigma_{\overline{X}}) - \mu_{\overline{X}}}{\sigma_{\overline{X}}}\right) =$$

$$= P\left(\frac{(\mu_X - 1{,}64\sigma_{\overline{X}}) - \mu_X}{\sigma_{\overline{X}}} \le Z \le \frac{(\mu_X + 1{,}64\sigma_{\overline{X}}) - \mu_X}{\sigma_{\overline{X}}}\right) =$$

$$= P(-1{,}64 \le \overline{X} \le 1{,}64) =$$

$$= \Phi(1{,}64) - \Phi(-1{,}64) = \Phi(1{,}64) - [1 - \Phi(1{,}64)] = 2 \times \Phi(1{,}64) - 1 =$$

$$= 2 \times 0{,}9495 - 1 = 0{,}8990$$



Para os itens *b*, *c* e *d* o procedimento é análogo, fica a cargo do aluno.

3)

$$n = 36;\ \mu_X = 0{,}2;\ \sigma_X = 0{,}01 \Rightarrow \mu_{\overline{X}} = 0{,}2;\ \sigma_{\overline{X}} = \frac{0{,}01}{\sqrt{36}} = \frac{0{,}01}{6} \cong 0{,}002$$

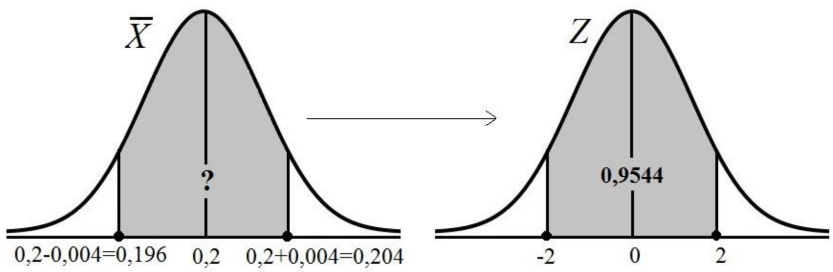
$$\overline{X} \to N(0{,}2;0{,}002)$$

$$a)\ P(0{,}2 - 0{,}004 \le \overline{X} \le 0{,}2 + 0{,}004) =$$

$$= P\left[\frac{(0{,}2 - 0{,}004) - 0{,}2}{0{,}002} \le Z \le \frac{(0{,}2 + 0{,}004) - 0{,}2}{0{,}002}\right] =$$

$$= P\left(-\frac{0{,}004}{0{,}002} \le Z \le \frac{0{,}004}{0{,}002}\right) = P(-2 \le Z \le 2) = \Phi(2) - \Phi(-2) =$$

$$= \Phi(2) - [1 - \Phi(2)] = 2 \times \Phi(2) - 1 = 2 \times 0{,}9772 - 1 = 1{,}9544 - 1 = 0{,}9544$$



$$b)\ P(|\overline{X} - \mu_X| \ge 0{,}005) = P(|\overline{X} - 0{,}2| \ge 0{,}005) =$$

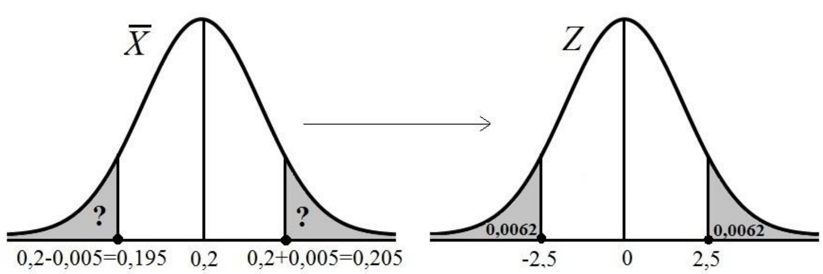
$$= P(\overline{X} - 0{,}2 \le -0{,}005 \vee \overline{X} - 0{,}2 \ge 0{,}005) =$$

$$= P(\overline{X} - 0{,}2 \le -0{,}005) + P(\overline{X} - 0{,}2 \ge 0{,}005) = P\left(\frac{\overline{X} - 0{,}2}{0{,}002} \le -\frac{0{,}005}{0{,}002}\right) +$$

$$+ P\left(\frac{\overline{X} - 0{,}2}{0{,}002} \ge \frac{0{,}005}{0{,}002}\right) = P(Z \le -2{,}5) + P(Z \ge 2{,}5) = [1 - P(Z \le 2{,}5)] +$$

$$+ [1 - P(Z \le 2{,}5)] = 2[1 - P(Z \le 2{,}5)] = 2[1 - \Phi(2{,}5)] = 2 \times (1 - 0{,}9938) =$$

$$= 2 \times 0{,}0062 = 0{,}0124$$



4)

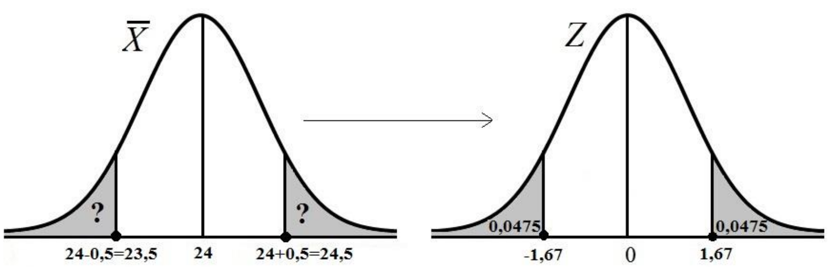
$$n = 100;\mu_X = 24;\sigma_X = 3 \Rightarrow \mu_{\overline{X}} = 24;\sigma_{\overline{X}} = \frac{3}{\sqrt{100}} = 0{,}3$$

$$P(|\,\overline{X} - \mu_X\,| \geq 0{,}5) = P(|\,\overline{X} - 24\,| \geq 0{,}5) = P(\overline{X} - 24 \leq -0{,}5 \vee \overline{X} - 24 \geq 0{,}5) =$$

$$P(\overline{X} - 24 \leq -0{,}5) + P(\overline{X} - 24 \geq 0{,}5) = P\left(\frac{\overline{X} - 24}{0{,}3} \leq -\frac{0{,}5}{0{,}3}\right) + P\left(\frac{\overline{X} - 24}{0{,}3} \geq \frac{0{,}5}{0{,}3}\right) =$$

$$= P(Z \leq -1{,}67) + P(Z \geq 1{,}67) = [1 - P(Z \leq 1{,}67)] + [1 - P(Z \leq 1{,}67)] =$$

$$= 2[1 - P(Z \leq 1{,}67)] = 2[1 - \Phi(1{,}67)] = 2 \times (1 - 0{,}9525) = 2 \times 0{,}0475 = 0{,}095$$



OBS: A partir do exercício seguinte serão omitidas as interpretações gráficas, que serão retomadas a partir do exercício 18.

5)

$$a)\ \ N = 20000; n = 225;\sigma_X = 45 \Rightarrow \sigma_{\overline{X}} = \frac{45}{\sqrt{225}} \cdot \sqrt{\frac{20000 - 225}{20000 - 1}} = 2{,}98$$

$$P(|\,\overline{X} - \mu_X\,| \geq 7{,}5) = P(|\,\overline{X} - \mu_{\overline{X}}\,| \geq 7{,}5) = P(\overline{X} - \mu_{\overline{X}} \leq -7{,}5 \vee \overline{X} - \mu_{\overline{X}} \geq 7{,}5) =$$

$$= P(\overline{X} - \mu_{\overline{X}} \leq -7{,}5) + P(\overline{X} - \mu_{\overline{X}} \geq 7{,}5) = P\left(\frac{\overline{X} - \mu_{\overline{X}}}{2{,}98} \leq -\frac{7{,}5}{2{,}98}\right) +$$

$$+ P\left(\frac{\overline{X} - \mu_{\overline{X}}}{2{,}98} \geq \frac{7{,}5}{2{,}98}\right) = P(Z \leq -2{,}52) + P(Z \geq 2{,}52) = [1 - P(Z \leq 2{,}52)] +$$

$$+ [1 - P(Z \leq 2{,}52)] = 2[1 - P(Z \leq 2{,}52)] = 2[1 - \Phi(2{,}52)] = 2 \times (1 - 0{,}9941) =$$

$$= 2 \times 0{,}0059 = 0{,}0118$$

$$b)\ N = 2000; n = 225; \sigma_X = 45 \Rightarrow \sigma_{\overline{X}} = \frac{45}{\sqrt{225}} \cdot \sqrt{\frac{2000-225}{2000-1}} = 2{,}83$$

$$P(|\overline{X} - \mu_X| \geq 7{,}5) = P(|\overline{X} - \mu_{\overline{X}}| \geq 7{,}5) = P(\overline{X} - \mu_{\overline{X}} \leq -7{,}5 \vee \overline{X} - \mu_{\overline{X}} \geq 7{,}5) =$$

$$= P(\overline{X} - \mu_{\overline{X}} \leq -7{,}5) + P(\overline{X} - \mu_{\overline{X}} \geq 7{,}5) = P\left(\frac{\overline{X} - \mu_{\overline{X}}}{2{,}83} \leq -\frac{7{,}5}{2{,}83}\right) +$$

$$+ P\left(\frac{\overline{X} - \mu_{\overline{X}}}{2{,}83} \geq \frac{7{,}5}{2{,}83}\right) = P(Z \leq -2{,}65) + P(Z \geq 2{,}65) = [1 - P(Z \leq 2{,}65)] +$$

$$+ [1 - P(Z \leq 2{,}65)] = 2[1 - P(Z \leq 2{,}65)] = 2[1 - \Phi(2{,}65)] = 2 \times (1 - 0{,}9960) =$$

$$= 2 \times 0{,}004 = 0{,}008$$

$$c)\ n = 225; \sigma_X = 45 \Rightarrow \sigma_{\overline{X}} = \frac{45}{\sqrt{225}} = 3$$

$$P(|\overline{X} - \mu_X| \geq 7{,}5) = P(|\overline{X} - \mu_{\overline{X}}| \geq 7{,}5) = P(\overline{X} - \mu_{\overline{X}} \leq -7{,}5 \vee \overline{X} - \mu_{\overline{X}} \geq 7{,}5) =$$

$$P(\overline{X} - \mu_{\overline{X}} \leq -7{,}5) + P(\overline{X} - \mu_{\overline{X}} \geq 7{,}5) = P\left(\frac{\overline{X} - \mu_{\overline{X}}}{3} \leq -\frac{7{,}5}{3}\right) +$$

$$+ P\left(\frac{\overline{X} - \mu_{\overline{X}}}{3} \geq \frac{7{,}5}{3}\right) = P(Z \leq -2{,}5) + P(Z \geq 2{,}5) = [1 - P(Z \leq 2{,}5)] +$$

$$+ [1 - P(Z \leq 2{,}5)] = 2[1 - P(Z \leq 2{,}5)] = 2[1 - \Phi(2{,}5)] = 2 \times (1 - 0{,}9938) =$$

$$= 2 \times 0{,}0062 = 0{,}0124$$

6)

Mesma solução e resultados do EXERCÍCIO 1, agora com $\hat{p} \to N(\mu_{\hat{p}}, \sigma_{\hat{p}}); \mu_{\hat{p}} = p$

7)

$$n = 100; p = 0{,}2 \Rightarrow \mu_{\hat{p}} = 0{,}2; \sigma^2_{\hat{p}} = \frac{0{,}2 \times 0{,}8}{100} = 0{,}0016; \sigma_{\hat{p}} = \sqrt{0{,}0016} = 0{,}04$$

$$\hat{p} \to N(0{,}2; 0{,}0016)$$

$$a)\ P(0{,}16 \le \hat{p} \le 0{,}24) = P\left(\frac{0{,}16-0{,}2}{0{,}04} \le Z \le \frac{0{,}24-0{,}2}{0{,}04}\right) = P(-1 \le Z \le 1) =$$

$$= \Phi(1) - \Phi(-1) == \Phi(1) - [1-\Phi(1)] = 2\Phi(1) - 1 = 2\times 0{,}8413 - 1 =$$

$$= 1{,}6826 - 1 = 0{,}6826 = 68{,}26\%$$

$$b)\ P(\hat{p} > 0{,}24) = 1 - P(\hat{p} \le 0{,}24) = 1 - P\left(Z \le \frac{0{,}24-0{,}2}{0{,}04}\right) = 1 - P(Z \le 1) = 1 - \Phi(1) =$$

$$= 1 - 0{,}8413 = 0{,}1587 = 15{,}87\%$$

$$c)\ P(0{,}12 \le \hat{p} \le 0{,}28) = P\left(\frac{0{,}1-0{,}2}{0{,}04} \le Z \le \frac{0{,}28-0{,}2}{0{,}04}\right) = P(-2{,}5 \le Z \le 2{,}0) =$$

$$= \Phi(2{,}0) - \Phi(-2{,}5) = \Phi(2) - [1-\Phi(2{,}5)] = 0{,}9772 - [1-0{,}9938] =$$

$$= 0{,}9772 - 0{,}0062 = 0{,}9710 = 97{,}10\%$$

$$d)\ P(\hat{p} < 0{,}12 \vee \hat{p} > 0{,}18) = P(\hat{p} < 0{,}12) + P(\hat{p} > 0{,}18) = P(\hat{p} \le 0{,}12) +$$

$$+ [1 - P(\hat{p} \le 0{,}18)] = P\left(Z \le \frac{0{,}12-0{,}2}{0{,}04}\right) + \left[1 - P\left(Z \le \frac{0{,}18-0{,}2}{0{,}04}\right)\right] =$$

$$= P(Z \le -2) + P(Z \le -0{,}5) = \Phi(-2) + \Phi(-0{,}5) = [1-\Phi(2)] + [1-\Phi(0{,}5)] =$$

$$= [1-0{,}9772] + [1-0{,}6915] = 0{,}0228 + 0{,}3085 = 0{,}3313 = 33{,}13\%$$

8)

$$N = 10000 \quad n = 100 \quad \hat{p} \to N(\mu_{\hat{p}}, \sigma^2_{\hat{p}})$$

$$\mu_{\hat{p}} = p$$

$$\sigma^2_{\hat{p}} = \frac{p\times q}{100}\cdot\frac{10000-100}{10000-1} = p\times q\times 0{,}0099 \Rightarrow \sigma_{\hat{p}} = \sqrt{p\times q\times 0{,}0099}$$

$$P(\text{rejeitar lote}) = P\left(\hat{p} \ge \frac{10}{100}\right) = P(\hat{p} \ge 0{,}10) = 1 - P(\hat{p} \le 0{,}10) =$$

$$= 1 - P\left(Z \le \frac{0{,}10-p}{\sqrt{p\times q\times 0{,}0099}}\right) \Rightarrow P(\text{rejeitar lote}) = 1 - \Phi\left(\frac{0{,}10-p}{\sqrt{p\times q\times 0{,}0099}}\right)$$

$$a)\ p = 0{,}03$$

$$P(\text{rejeitar lote}) = 1 - \Phi\left(\frac{0{,}10-0{,}03}{\sqrt{0{,}03\times 0{,}97\times 0{,}0099}}\right) = 1 - \Phi(4{,}12) \cong 0$$

$b)\ \ p = 0{,}05$

$$P(\text{rejeitar lote}) = 1 - \Phi\left(\frac{0{,}10 - 0{,}05}{\sqrt{0{,}05 \times 0{,}95 \times 0{,}0099}}\right) = 1 - \Phi(2{,}31) = 1 - 0{,}9896 = 0{,}0104$$

$c)\ \ p = 0{,}18$

$$P(\text{rejeitar lote}) = 1 - \Phi\left(\frac{0{,}10 - 0{,}18}{\sqrt{0{,}18 \times 0{,}82 \times 0{,}0099}}\right) = 1 - \Phi(-2{,}09) = 1 - [1 - \Phi(2{,}09)] =$$

$$= \Phi(2{,}09) = 0{,}9817$$

$d)\ \ p = 0{,}25$

$$P(\text{rejeitar lote}) = 1 - \Phi\left(\frac{0{,}10 - 0{,}25}{\sqrt{0{,}25 \times 0{,}75 \times 0{,}0099}}\right) = 1 - \Phi(-3{,}48) = 1 - [1 - \Phi(3{,}48)] =$$

$$= \Phi(3{,}48) \cong 1$$

9)

$a)\ \ \sigma_X = 0{,}01 \quad d = 0{,}004 \quad 1 - \alpha = 0{,}90 \Rightarrow z_{\alpha/2} = 1{,}64$

$$n = \frac{z_{\alpha/2}^2 \times \sigma_X^2}{d^2} = \frac{1{,}64^2 \times 0{,}01^2}{0{,}004^2} \cong 17$$

$b)\ \ \sigma_X = 0{,}01 \quad d = 0{,}002 \quad n = 100$

$$n = \frac{z_{\alpha/2}^2 \times \sigma_X^2}{d^2} \Rightarrow z_{\alpha/2}^2 = \frac{n \times d^2}{\sigma_X^2} \Rightarrow z_{\alpha/2} = \sqrt{\frac{n \times d^2}{\sigma_X^2}} = \sqrt{\frac{100 \times 0{,}002^2}{0{,}01^2}} = 2$$

$$1 - \alpha = \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) = \Phi(z_{\alpha/2}) - [1 - \Phi(z_{\alpha/2})] = 2\Phi(z_{\alpha/2}) - 1 =$$

$$= 2\Phi(2) - 1 = 2 \times 0{,}9772 - 1 = 0{,}9544 = 95{,}44\%$$

$c)\ \ \sigma_X = 0{,}01 \quad n = 49 \quad 1 - \alpha = 0{,}80 \Rightarrow z_{\alpha/2} \cong 1{,}28$

$$n = \frac{z_{\alpha/2}^2 \times \sigma_X^2}{d^2} \Rightarrow d^2 = \frac{z_{\alpha/2}^2 \times \sigma_X^2}{n} \Rightarrow d = \sqrt{\frac{z_{\alpha/2}^2 \times \sigma_X^2}{n}} = \sqrt{\frac{1{,}28^2 \times 0{,}01^2}{49}} = 0{,}002$$

10)

$a)\ \ d = 0{,}05 \quad 1 - \alpha = 0{,}90 \Rightarrow z_{\alpha/2} = 1{,}64$

$p = 0{,}70$

$$n = \frac{z_{\alpha/2}^2 \times p \times q}{d^2} = \frac{1{,}64^2 \times 0{,}7 \times 0{,}3}{0{,}05^2} \cong 226$$

$b)\ \ d = 0{,}05 \quad 1 - \alpha = 0{,}90 \Rightarrow z_{\alpha/2} = 1{,}64$

$p = 0{,}50$

$$n = \frac{z_{\alpha/2}^2 \times p \times q}{d^2} = \frac{1{,}64^2 \times 0{,}5 \times 0{,}5}{0{,}05^2} \cong 269$$

$c)\ N = 5000$

$(a)\ n_0 = 226 \qquad n = \dfrac{n_0}{1+\dfrac{n_0}{N}} = \dfrac{226}{1+\dfrac{226}{5000}} \cong 216$

$(b)\ n_0 = 269 \qquad n = \dfrac{n_0}{1+\dfrac{n_0}{N}} = \dfrac{269}{1+\dfrac{269}{5000}} \cong 255$

11)

$a)\ \sigma_X = 10 \quad d = 1{,}5 \quad 1-\alpha = 0{,}95 \Rightarrow z_{\alpha/2} = 1{,}96$

$$n_0 = \frac{z_{\alpha/2}^2 \times \sigma_X^2}{d^2} = \frac{1{,}96^2 \times 10^2}{1{,}5^2} \cong 171$$

$b)\ N = 600$

$$n = \frac{n_0}{1+\dfrac{n_0}{N}} = \frac{171}{1+\dfrac{171}{600}} \cong 133$$

12)

$d = 0{,}02$

$1-\alpha = 0{,}99 \Rightarrow z_{\alpha/2} \cong 2{,}57$

$a)\ p = 0{,}30 \qquad n = \dfrac{z_{\alpha/2}^2 \times p \times q}{d^2} = \dfrac{2{,}57^2 \times 0{,}30 \times 0{,}70}{0{,}02^2} \cong 3468$

$b)\ n_0 = 3468 \qquad n = \dfrac{n_0}{1+\dfrac{n_0}{N}} = \dfrac{3468}{1+\dfrac{3468}{10000}} \cong 2575$

$c)\ p = 0{,}50 \qquad n = \dfrac{z_{\alpha/2}^2 \times p \times q}{d^2} = \dfrac{2{,}57^2 \times 0{,}50 \times 0{,}50}{0{,}02^2} \cong 4128$

$d)\ n_0 = 4128 \qquad n = \dfrac{n_0}{1+\dfrac{n_0}{N}} = \dfrac{4128}{1+\dfrac{4128}{10000}} \cong 2922$

13)

$$X \to N(0{,}8;0{,}0004) \qquad \mu_X = 0{,}8 \quad \sigma_X^2 = 0{,}0004 \Rightarrow \sigma_X = 0{,}02$$

$$\overline{X} \to N\left(0{,}8;\frac{0{,}0004}{n}\right) \qquad \mu_{\overline{X}} = 0{,}8 \quad \sigma_{\overline{X}}^2 = \frac{0{,}0004}{n} \Rightarrow \sigma_{\overline{X}} = \frac{0{,}02}{\sqrt{n}}$$

$$a) \quad n = 16 \Rightarrow \sigma_{\overline{X}} = \frac{0{,}02}{\sqrt{16}} = 0{,}005$$

$$P(\overline{X} > 0{,}807) = 1 - P(\overline{X} \le 0{,}807) = 1 - P\left(Z \le \frac{0{,}807 - 0{,}8}{0{,}005}\right) =$$

$$= 1 - P(Z \le 1{,}4) = 1 - \Phi(1{,}4) = 1 - 0{,}9192 = 0{,}0808$$

$$b) \quad d = 0{,}001 \quad 1 - \alpha = 0{,}90 \Rightarrow z_{\alpha/2} = 1{,}64$$

$$n = \frac{z_{\alpha/2}^2 \times \sigma_X^2}{d^2} = \frac{1{,}64^2 \times 0{,}02^2}{0{,}001^2} \cong 1076$$

$$c) \quad n = 100 \quad 1 - \alpha = 0{,}95 \Rightarrow z_{\alpha/2} = 1{,}96$$

$$n = \frac{z_{\alpha/2}^2 \times \sigma_X^2}{d^2} \Rightarrow d = \sqrt{\frac{z_{\alpha/2}^2 \times \sigma_X^2}{n}} = \sqrt{\frac{1{,}96^2 \times 0{,}02^2}{100}} \cong 0{,}004$$

14)

$$a) \quad n = 100 \quad p = 0{,}10$$

$$\hat{p} \to N\left(0{,}10;\frac{0{,}10 \times 0{,}90}{100}\right) \quad \mu_{\hat{p}} = 0{,}10 \quad \sigma_{\hat{p}}^2 = 0{,}0009 \Rightarrow \sigma_{\hat{p}} = 0{,}03$$

$$P(|\,\hat{p} - p\,| > 0{,}02) = P\left(\left|\frac{\hat{p} - p}{0{,}03}\right| > \frac{0{,}02 - 0{,}10}{0{,}03}\right) = P(|\,Z\,| > 2{,}67) =$$

$$= P(Z < -2{,}67) + P(Z > 2{,}67) = \Phi(-2{,}67) + [1 - \Phi(2{,}67)] =$$

$$= [1 - \Phi(2{,}67)] + [1 - \Phi(2{,}67)] = 2[1 - \Phi(2{,}67)] = 2(1 - 0{,}9962) = 0{,}0076$$

$$b) \quad d = 0{,}05 \quad 1 - \alpha = 0{,}90 \Rightarrow z_{\alpha/2} = 1{,}64 \quad n = \frac{z_{\alpha/2}^2 \times p \times q}{d^2} = \frac{1{,}64^2 \times 0{,}10 \times 0{,}90}{0{,}05^2} \cong 97$$

$$c) \quad p = 0{,}5 \quad d = 0{,}05 \quad 1 - \alpha = 0{,}90 \Rightarrow z_{\alpha/2} = 1{,}64$$

$$n = \frac{z_{\alpha/2}^2 \times p \times q}{d^2} = \frac{1{,}64^2 \times 0{,}5 \times 0{,}5}{0{,}05^2} \cong 269$$

15)

$$a)\quad n = 25 \quad \overline{X} = 5{,}2 \quad \sigma_X = 1{,}2 \quad 1-\alpha = 0{,}90 \Rightarrow \alpha = 0{,}10 \Rightarrow z_{\alpha/2} = 1{,}64$$

$$IC_{\mu_X} \equiv \overline{X} \pm z_{\alpha/2} \times \sigma_{\overline{X}} = \overline{X} \pm z_{\alpha/2} \times \frac{\sigma_X}{\sqrt{n}} = 5{,}2 \pm 1{,}64 \times \frac{1{,}2}{\sqrt{25}} = 5{,}2 \pm 0{,}39 \Rightarrow$$

$$\Rightarrow IC_{\mu_X} \equiv (4{,}81;5{,}59)$$

Para 1-α=0,95 a solução é análoga, com $z_{\alpha/2}$=1,96

$$b)\quad n = 25 \Rightarrow GL = 24 \quad \overline{X} = 5{,}2 \quad S = 1{,}2 \quad 1-\alpha = 0{,}90 \Rightarrow \alpha = 0{,}10 \Rightarrow t_{n-1,\alpha/2} = 1{,}711$$

$$IC_{\mu_X} \equiv \overline{X} \pm t_{n-1,\alpha/2} \times \frac{S}{\sqrt{n}} = 5{,}2 \pm 1{,}711 \times \frac{1{,}2}{\sqrt{25}} = 5{,}2 \pm 0{,}41 \Rightarrow IC_{\mu_X} \equiv (4{,}79;5{,}61)$$

Para 1-α=0,95 a solução é análoga, com $t_{n-1,\alpha/2}$=2,064

16)

$$a)\quad n = 16$$

$$\sum x = 25 + 26 + 26 + \ldots + 25 + 27 = 437$$

$$\sum x^2 = 25^2 + 26^2 + 26^2 + \ldots + 25^2 + 27^2 = 11987$$

$$\overline{X} = \frac{\sum x}{n} = \frac{437}{16} = 27{,}31 \quad \text{(estimativa por ponto)}$$

$$S = \sqrt{\frac{\sum x^2}{n-1} - \frac{\left(\sum x\right)^2}{n(n-1)}} = \sqrt{\frac{11987}{15} - \frac{437^2}{16 \times 15}} = 1{,}85$$

$$1-\alpha = 0{,}98 \Rightarrow \alpha = 0{,}02 \Rightarrow t_{n-1,\alpha/2} = 2{,}602$$

$$IC_{\mu_X} \equiv \overline{X} \pm t_{n-1,\alpha/2} \times \frac{S}{\sqrt{n}} = 27{,}31 \pm 2{,}602 \times \frac{1{,}85}{\sqrt{16}} = 27{,}31 \pm 1{,}20 \Rightarrow$$

$$\Rightarrow IC_{\mu_X} \equiv (26{,}11;28{,}51)$$

$$b)\quad n = 16 \quad N = 100$$

$$\overline{X} = 27{,}31 \quad \text{(estimativa por ponto)} \quad S = 1{,}85$$

$$1-\alpha = 0{,}98 \Rightarrow \alpha = 0{,}02 \Rightarrow t_{n-1,\alpha/2} = 2{,}602$$

$$IC_{\mu_X} \equiv \overline{X} \pm t_{n-1,\alpha/2} \times \frac{S}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = 27{,}31 \pm 2{,}602 \times \frac{1{,}85}{\sqrt{16}} \cdot \sqrt{\frac{100-16}{100-1}} =$$

$$= 27{,}31 \pm 1{,}11 \Rightarrow \quad \Rightarrow IC_{\mu_X} \equiv (26{,}20;28{,}42)$$

17)

$$a)\quad n = 16 \quad \hat{p} = \frac{10}{16} = 0{,}625 \quad 1-\alpha = 0{,}98 \Rightarrow \alpha = 0{,}02 \Rightarrow z_{\alpha/2} = 2{,}33$$

$$IC_{\hat{p}} \equiv \hat{p} \pm z_{\alpha/2} \times \sqrt{\frac{\hat{p}\hat{q}}{n}} = 0{,}625 \pm 2{,}33 \times \sqrt{\frac{0{,}625 \times 0{,}375}{16}} = 0{,}635 \pm 0{,}282 \Rightarrow$$

$$\Rightarrow IC_{\hat{p}} \equiv (0{,}353;0{,}917)$$

$$b)\quad n = 16 \quad N = 100 \quad \hat{p} = 0{,}625 \quad 1-\alpha = 0{,}98 \Rightarrow \alpha = 0{,}02 \Rightarrow z_{\alpha/2} = 2{,}33$$

$$IC_{\hat{p}} \equiv \hat{p} \pm z_{\alpha/2} \times \sqrt{\frac{\hat{p}\hat{q}}{n} \cdot \frac{N-n}{N-1}} = 0{,}625 \pm 2{,}33 \times \sqrt{\frac{0{,}625 \times 0{,}375}{16} \cdot \frac{100-16}{100-1}} =$$

$$= 0{,}635 \pm 0{,}260 \Rightarrow IC_{\hat{p}} \equiv (0{,}375;0{,}895)$$

18)

$$n = 300 \quad \hat{p} = \frac{200}{300} = 0{,}67 \quad 1-\alpha = 0{,}90 \Rightarrow \alpha = 0{,}10 \Rightarrow z_{\alpha/2} = 1{,}64$$

$$IC_{\hat{p}} \equiv \hat{p} \pm z_{\alpha/2} \times \sqrt{\frac{\hat{p}\hat{q}}{n}} = 0{,}67 \pm 1{,}64 \times \sqrt{\frac{0{,}67 \times 0{,}33}{300}} = 0{,}67 \pm 0{,}04 \Rightarrow$$

$$\Rightarrow IC_{\hat{p}} \equiv (0{,}63;0{,}71)$$

19)

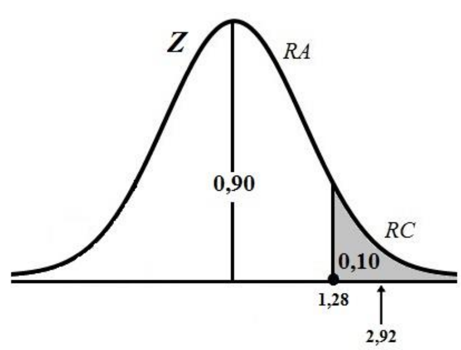
$$a)\quad \begin{cases} H_0 : \mu_X = 4{,}5 \\ H_1 : \mu_X > 4{,}5 \end{cases}$$

$$n = 25 \quad \overline{X} = 5{,}2 \quad \sigma_X = 1{,}2$$

$$1-\alpha = 0{,}90 \Rightarrow \alpha = 0{,}10 \Rightarrow z_{\alpha} = 1{,}28$$

$$Z_c = \frac{\overline{X} - \mu_0}{\frac{\sigma_X}{\sqrt{n}}} = \frac{5{,}2 - 4{,}5}{\frac{1{,}2}{\sqrt{25}}} = 2{,}92$$

$|Z_c| > z_{\alpha/2} \Rightarrow$ rejeita - se H_0



Para 1-α=0,95 a solução é análoga, com z_{α}=1,64

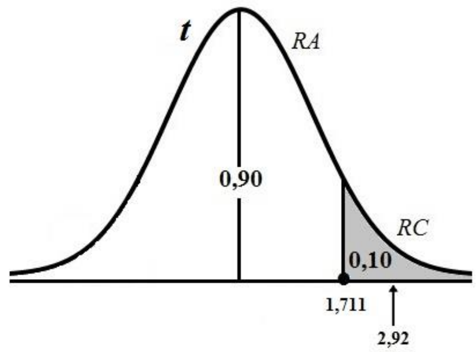
$$b)\ \begin{cases} H_0 : \mu_X = 4{,}5 \\ H_1 : \mu_X > 4{,}5 \end{cases}$$

$n = 25 \quad \overline{X} = 5{,}2 \quad S = 1{,}2$

$1-\alpha = 0{,}90 \Rightarrow \alpha = 0{,}10 \Rightarrow t_{n-1,\alpha} = 1{,}71$

$$T_c = \frac{\overline{X}-\mu_0}{\frac{S}{\sqrt{n}}} = \frac{5{,}2-4{,}5}{\frac{1{,}2}{\sqrt{25}}} = 2{,}92$$

$|T_c| > t_{n-1,\alpha} \Rightarrow$ rejeita - se H_0



Para 1-α=0,95 a solução é análoga, com $t_{n-1,\alpha/2}$=2,064

20)

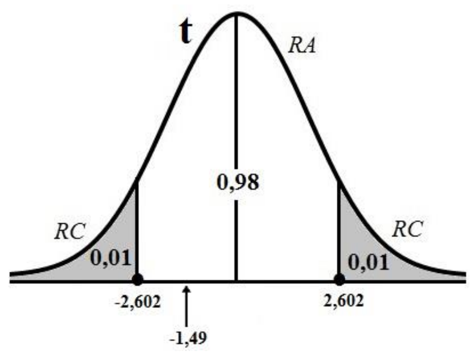
$$a)\ \begin{cases} H_0 : \mu_X = 28 \\ H_1 : \mu_X \neq 28 \end{cases}$$

$n = 16 \quad \overline{X} = 27{,}31 \quad S = 1{,}85$

$1-\alpha = 0{,}98 \Rightarrow \alpha = 0{,}02 \Rightarrow t_{n-1,\alpha/2} = 2{,}602$

$$T_c = \frac{\overline{X}-\mu_0}{\frac{S}{\sqrt{n}}} = \frac{27{,}31-28}{\frac{1{,}85}{\sqrt{16}}} = -1{,}49$$

$|T_c| < t_{n-1,\alpha/2} \Rightarrow$ aceita - se H_0



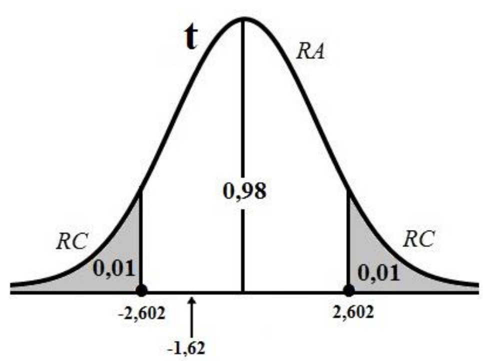
$$b)\ \begin{cases} H_0 : \mu_X = 28 \\ H_1 : \mu_X \neq 28 \end{cases}$$

$n = 16 \quad N = 100 \quad \overline{X} = 27{,}31 \quad S = 1{,}85$

$1-\alpha = 0{,}98 \Rightarrow \alpha = 0{,}02 \Rightarrow t_{n-1,\alpha/2} = 2{,}602$

$$T_c = \frac{\overline{X}-\mu_0}{\frac{S}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}} = \frac{27{,}31-28}{\frac{1{,}85}{\sqrt{16}} \cdot \sqrt{\frac{100-16}{100-1}}} = -1{,}62$$

$|T_c| < t_{n-1,\alpha/2} \Rightarrow$ aceita - se H_0



21)

$$\begin{cases} H_0 : p = 0{,}64 \\ H_1 : p < 0{,}64 \end{cases}$$

$$n = 16 \quad \hat{p} = \frac{10}{16} = 0{,}625 \quad 1-\alpha = 0{,}98 \Rightarrow \alpha = 0{,}02 \Rightarrow z_\alpha = 2{,}05$$

$$a)\ Z_c = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0{,}625 - 0{,}64}{\sqrt{\frac{0{,}64 \times 0{,}36}{16}}} = -0{,}125$$

$$|Z_c| < z_\alpha \Rightarrow \text{aceita - se } H_0$$

$$b)\ Z_c = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n} \cdot \frac{N-n}{N-1}}} = \frac{0{,}625 - 0{,}64}{\sqrt{\frac{0{,}64 \times 0{,}36}{16} \cdot \frac{100-16}{100-1}}} = -0{,}136$$

$$|Z_c| < z_\alpha \Rightarrow \text{aceita - se } H_0$$

Parte 4

Correlação e Regressão

Capítulo 1

Correlação

Introdução

No cotidiano é comum a análise simultânea de duas ou mais características de uma população como, por exemplo, o peso e altura das pessoas ou, ainda, a quantidade de memória de um sistema e o tempo de resposta do mesmo. Nestes casos, estão envolvidas duas variáveis aleatórias, X e Y, quem têm seus comportamentos probabilísticos individuais, já discutidos na Parte 2, ou seja, cada uma com sua respectiva função densidade e de distribuição acumulada de probabilidades, esperança matemática e variância. É de grande interesse verificar-se se há algum tipo de relação entre essas variáveis, o que pode ser feito através do que se chamam MEDIDAS DE ASSOCIAÇÃO, uma delas o COEFICIENTE DE CORRELAÇÃO. Verificada a existência de associação, a etapa seguinte é a "construção" de um modelo funcional ou MODELO DE PREVISÃO que a represente, permitindo a previsão de uma das variáveis (Y) a partir de valores da outra (X). No caso da existência de uma relação linear entre essas variáveis é utilizada a ANÁLISE DE REGRESSÃO LINEAR com esta finalidade.

1 Variáveis Aleatórias Bidimensionais

Considere duas variáveis aleatórias, X e Y, com densidades de probabilidade $f_X(x)$ e $f_Y(y)$, respectivamente. Então, a variável aleatória bidimensional (X,Y) terá função densidade de probabilidade conjunta denotada por $f_{xy}(x,y)$.

Importante

- $f_X(x)$ e $f_Y(y)$ são chamadas densidades marginais de X e Y;
- X e Y podem ser ambas do mesmo tipo (discreta ou contínua) ou de tipos diferentes;
- Quando ambas são discretas, (X,Y) é dita DISCRETA, $f_{XY}(x,y)$ pode ser escrita como $P_{XY}(x,y)$ e fornece $P_{XY}(X=x \text{ e } Y=y)$, a probabilidade de X e Y assumirem os valores x e y, simultaneamente;
- Quando ambas são contínuas, (X,Y) é dita CONTÍNUA, $f_{XY}(x,y)$ fornece $P_{XY}(X\in I_x \text{ e } Y\in I_y)$, a probabilidade de X e Y assumirem valores dentro dos intervalos reais I_x e I_y, simultaneamente.
- Como $f_{XY}(x,y)$ é uma densidade de probabilidade, tem-se que $f_{XY}(x,y)\geq 0$ e $\sum\sum P_{XY}(x_i, y_j) = 1$ e $\iint f_{XY}(x,y) = 1$, para os casos discreto e contínuo, respectivamente.

1.1 Esperança Matemática, Variância e Covariância

$\mu_X = E(X)$ - Esperança Matemática ou Média de X

$\mu_Y = E(Y)$ - Esperança Matemática ou Média de Y

$\sigma_X^2 = V(X) = E\left\{[X - E(X)]^2\right\} = E(X^2) - [E(X)]^2$ - Variância de X

$\sigma_X^2 = V(X) = E\left\{[X - E(X)]^2\right\} = E(X^2) - [E(X)]^2$ - Variância de Y

$\sigma_{XY} = COV(XY) = E\{[X - E(X)][Y - E(Y)]\} = E(XY) - E(X)E(Y)$ - Covariância de (X,Y)

Sendo:

CASO DISCRETO	CASO CONTÍNUO
$E(X) = \sum x f_X(x)$	$E(X) = \int x f_X(x)dx$
$E(Y) = \sum y f_Y(y)$	$E(Y) = \int y f_Y(y)dy$
$E(XY) = \sum_y \sum_x xy f_{XY}(x,y)$	$E(XY) = \int_y \int_x xy f_{XY}(x,y)dxdy$

Importante

- Duas variáveis, X e Y, são INDEPENDENTES se e somente se $f_{XY}(x,y)=f_X(x)f_Y(y)$;
- Demosntra-se que, se X e Y são independentes, então $E(XY) = E(X)E(Y)$, assim, $COV_{XY}= 0$;
- Por outro lado, a recíproca não é verdadeira, ou seja, se $COV_{XY}= 0$ não implica que as variáveis são independentes.

EXEMPLO 4.1.1.1[1]

O quadro abaixo mostra a distribuição de probabilidade conjunta das variáveis X e Y. Calcule $E(X)$, $V(X)$, $E(Y)$, $V(Y)$ e $COV(X,Y)$. Verifique se X e Y são independentes.

Y	X				
	0	1	2	3	SOMA
0	0,03	0,075	0,0225	0,0225	0,15
1	0,05	0,125	0,0375	0,0375	0,25
2	0,12	0,30	0,09	0,09	0,60
SOMA	0,20	0,50	0,15	0,15	1

$$E(X) = 0\times 0{,}20 + 1\times 0{,}50 + 2\times 0{,}15 + 3\times 0{,}15 = 1{,}25$$
$$E(X^2) = 0^2\times 0{,}20 + 1^2\times 0{,}50 + 2^2\times 0{,}15 + 3^2\times 0{,}15 = 2{,}45$$
$$V(X) = 2{,}45 - 1{,}25^2 = 0{,}8875$$

$$E(Y) = 0\times 0{,}15 + 1\times 0{,}25 + 2\times 0{,}60 = 1{,}45$$
$$E(Y^2) = 0^2\times 0{,}15 + 1^2\times 0{,}25 + 2^2\times 0{,}60 = 2{,}65$$
$$V(Y) = 2{,}65 - 1{,}45^2 = 0{,}5475$$

$$E(XY) = (0\times 0)\times 0{,}03 + (0\times 1)\times 0{,}05 + (0\times 2)\times 0{,}20 +$$
$$+ (1\times 0)\times 0{,}075 + (1\times 1)\times 0{,}125 + (1\times 2)\times 0{,}30 +$$
$$+ (2\times 0)\times 0{,}0225 + (2\times 1)\times 0{,}0375 + (2\times 2)\times 0{,}09 +$$
$$+ (3\times 0)\times 0{,}0225 + (3\times 1)\times 0{,}0375 + (3\times 2)\times 0{,}09 = 1{,}8125$$
$$COV(XY) = 1{,}8125 - 1{,}25\times 1{,}45 = 0$$

$$P_X(X=0)P_Y(Y=0) = 0{,}20\times 0{,}15 = 0{,}03 = P_{XY}(X=0, Y=0)$$
$$P_X(X=0)P_Y(Y=1) = 0{,}20\times 0{,}25 = 0{,}05 = P_{XY}(X=0, Y=1)$$
$$P_X(X=0)P_Y(Y=2) = 0{,}20\times 0{,}60 = 0{,}12 = P_{XY}(X=0, Y=2)$$

$$P_X(X=1)P_Y(Y=0)=0{,}50\times 0{,}15=0{,}075=P_{XY}(X=1,Y=0)$$
$$P_X(X=1)P_Y(Y=1)=0{,}50\times 0{,}25=0{,}125=P_{XY}(X=1,Y=1)$$
$$P_X(X=1)P_Y(Y=2)=0{,}50\times 0{,}60=0{,}30=P_{XY}(X=1,Y=2)$$
$$P_X(X=2)P_Y(Y=0)=0{,}15\times 0{,}15=0{,}0225=P_{XY}(X=2,Y=0)$$
$$P_X(X=2)P_Y(Y=1)=0{,}15\times 0{,}25=0{,}0375=P_{XY}(X=2,Y=1)$$
$$P_X(X=2)P_Y(Y=2)=0{,}15\times 0{,}60=0{,}09=P_{XY}(X=2,Y=2)$$
$$P_X(X=3)P_Y(Y=0)=0{,}15\times 0{,}15=0{,}0225=P_{XY}(X=3,Y=0)$$
$$P_X(X=3)P_Y(Y=1)=0{,}15\times 0{,}25=0{,}0375=P_{XY}(X=3,Y=1)$$
$$P_X(X=3)P_Y(Y=2)=0{,}15\times 0{,}60=0{,}09=P_{XY}(X=3,Y=2)$$

Como, para todo par (x,y), tem-se que $P_{XY}(x,y)=P_X(X=x)P_Y(Y=y)$, concluímos que X e Y são independentes.

1.2 Coeficiente de Correlação Linear

O Coeficiente de Correlação Linear é uma medida que indica o grau de associação LINEAR entre as variáveis X e Y, definido por:

$$\rho_{XY}=\frac{COV(XY)}{\sqrt{V(X)V(Y)}}=\frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

Demonstra-se que $-1\le\rho_{XY}\le 1$ e que se $\rho_{XY}^2=1$, ou seja, se $\rho_{XY}=\pm 1$ então Y é uma combinação linear de X, ou seja, $Y = AX + B$. Caso o coeficiente seja negativo isso indica que pequenos valores de X se associam a grandes valores de Y e grandes valores de X se associam a pequenos valores de Y (reta decrescente). Se positivo, pequenos valores X se associam a pequenos valores de Y e grandes valores X se associam a grandes valores de Y (reta crescente). Se $\rho_{XY}=0$ as variáveis são não correlacionadas linearmente.

Observações

- O coeficiente de correlação linear só existe se $V(X)\neq 0$ e $V(Y)\neq 0$;
- Se X e Y são independentes elas não são correlacionadas linearmente, ou seja, $\rho_{XY}=0$. Mas a recíproca não é verdadeira;
- $\rho_{XY}=0$ não significa que X e Y não tenham algum tipo e associação, elas podem, por exemplo, ter uma relação quadrática: $Y=X^2$

A partir deste ponto, utilizaremos apenas a expressão Coeficiente de Correlação.

2 Coeficiente de Correlação Amostral

2.1 Estimadores dos parâmetros populacionais

Dada uma amostra aleatória, (x_1,y_1), (x_2,y_2),..., (x_n,y_n), temos os seguintes estimadores:

$$S_X^2 = \frac{\sum (x-\overline{X})^2}{n-1} = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}$$ é o estimador da variância de X

$$S_Y^2 = \frac{\sum (y-\overline{Y})^2}{n-1} = \frac{\sum y^2 - \frac{(\sum y)^2}{n}}{n-1}$$ é o estimador da variância de Y

$$S_{XY} = \frac{\sum (x-\overline{X})(y-\overline{Y})}{n-1} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{n-1}$$ é o estimador da covariância de (X,Y)

$$r_{XY} = \frac{S_{XY}}{\sqrt{S_X^2 S_Y^2}} = \frac{S_{XY}}{S_X S_Y} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left\{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]\right\}}}$$ é o estimador do coeficiente de correlação entre X e Y

EXEMPLO 4.1.2.1[1]

A tabela seguinte apresenta os resultados obtidos a partir de uma amostra de 10 alunos de uma escola, referentes ao tempo (min) de uso da Internet por dia (X) e o escore obtido em um teste de qualificação para monitor do Laboratório de Informática (Y). Calcule as estimativas pontuais de $V(X)$, $V(Y)$, $COV(X,Y)$ e ρ_{XY}.

	1	2	3	4	5	6	7	8	9	10	SOMA
x	30	40	10	50	10	20	30	40	50	60	340
y	5	7	8	10	5	2	7	9	7	10	70
x^2	900	1600	100	2500	100	400	900	1600	2500	3600	14200
y^2	25	49	64	100	25	4	49	81	49	100	546
xy	150	280	80	500	50	40	210	360	350	600	2620

$$S_X^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1} = \frac{14200 - \frac{340^2}{10}}{9} = 293{,}33 \Rightarrow S_X = 17{,}13$$

$$S_Y^2 = \frac{\sum y^2 - \frac{\left(\sum y\right)^2}{n}}{n-1} = \frac{546 - \frac{70^2}{10}}{9} = 6{,}22 \Rightarrow S_Y = 2{,}49$$

$$S_{XY} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{n-1} = \frac{2620 - \frac{340 \times 70}{10}}{9} = 26{,}67$$

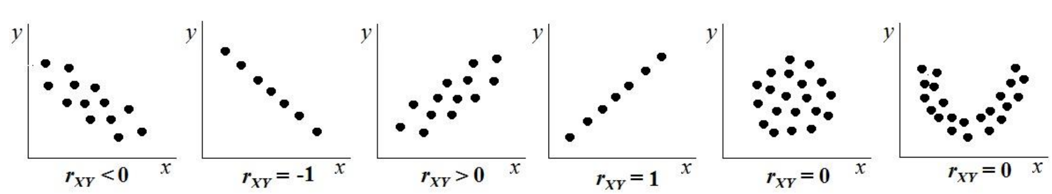
$$r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{26{,}67}{17{,}13 \times 2{,}49} = 0{,}62$$

2.2 Análise do Coeficiente de Correlação Amostral

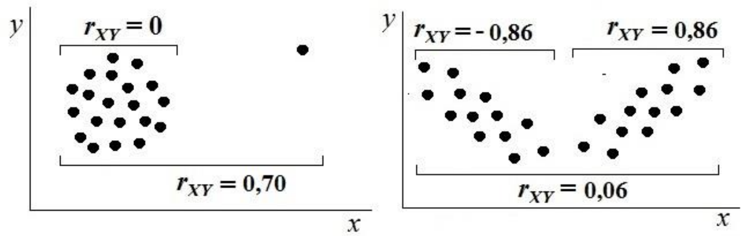
Seguindo o mesmo comportamento do coeficiente de correlação populacional, temos os seguintes resultados:

- $-1 \le r_{XY} \le 1$
- $r_{XY} < 0 \ \Rightarrow$ pequenos valores de X se associam a grandes valores de Y e grandes valores de X se associam a pequenos valores de Y (reta decrescente)
- $r_{XY} > 0 \ \Rightarrow$ pequenos valores de X se associam a pequenos valores de Y e grandes valores de X se associam a grandes valores de Y (reta crescente)
- $r_{XY} = -1 \ \Rightarrow$ correlação linear perfeita (decrescente) entre X e Y
- $r_{XY} = 1 \ \Rightarrow$ correlação linear perfeita (crescente) entre X e Y
- $r_{XY} = 0 \ \Rightarrow$ ausência de correlação linear entre X e Y
- se X e Y são independentes $\Rightarrow r_{XY} = 0$ (a recíproca não é verdadeira)

Uma ferramenta fundamental para a análise da correlação linear entre as variáveis X e Y é o DIAGRAMA DE DISPERSÃO, também conhecido como NUVEM DE PONTOS, que é o gráfico cartesiano dos pontos (x,y). As figuras abaixo caracterizam, através do referido diagrama, as situações descritas acima.

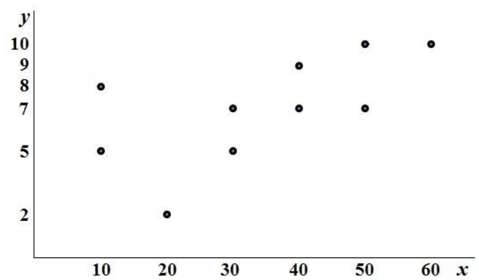


Antes de qualquer análise numérica de r_{XY} é fundamental a avaliação gráfica. Em muitas situações seu valor não é um bom indicador do grau de relação entre as variáveis. A presença de valores fora da nuvem de pontos (*outliers*) ou mesmo a existência de estratos detectados na amostra podem influenciar decisivamente no valor de r_{XY}. As figuras abaixo ilustram essas situações.



EXEMPLO 4.1.2.2[1]

Construa o diagrama de dispersão do exemplo 4.1.2.1[1].



3 Intervalo de Confiança para ρ_{XY}

Para a construção de intervalos de confiança e testes de hipóteses para ρ_{XY} é necessário o conhecimento da distribuição de probabilidade de seu estimador r_{XY}. Mas esta distribuição depende tanto de ρ_{XY} e de n (tamanho da amostra) e seu formato varia demasiadamente de acordo com o valor de ρ_{XY}, não sendo facilmente determinada. Assim, será utilizada a variável aleatória abaixo, que tem distribuição Normal.

$$R = \frac{1}{2} \cdot \ln\left(\frac{1 + r_{XY}}{1 - r_{XY}}\right) \qquad R \sim N(\mu_R, \sigma_R^2)$$

$$\mu_R = \frac{1}{2} \cdot \ln\left(\frac{1 + \rho_{XY}}{1 - \rho_{XY}}\right) \text{ e } \sigma_R^2 = \frac{1}{n-3} \Rightarrow \sigma_R = \frac{1}{\sqrt{n-3}}$$

Podemos então escrever $P\left(-z_{\alpha/2} \leq \frac{R-\mu_R}{\sigma_R} \leq z_{\alpha/2}\right) = 1-\alpha$. Substituindo-se R e μ_R por suas respectivas expressões e isolando ρ_{XY} (reveja na Parte 3 – Capítulo 3), chegamos aos seguintes resultados para o Intervalo de Confiança de ρ_{XY}:

$$IC_{\rho} : \left(\frac{e^{c_1}-1}{1+e^{c_1}}, \frac{e^{c_2}-1}{1+e^{c_2}}\right)$$

$$c_1 = 2(R - z_{\alpha/2} \cdot \sigma_R) \qquad c_2 = 2(R + z_{\alpha/2} \cdot \sigma_R)$$

Observações

- $1\text{-}\alpha$ é o Nível de Confiança da estimativa e $z_{\alpha/2}$ é a abcissa da Curva Normal Padrão, associada a $1\text{-}\alpha$;
- $\ln$ é o logaritmo na base natural (e ≅ 2,7183);

EXEMPLO 4.1.3[1]

Foram avaliados os desempenhos de 100 processadores, com relação ao tempo de utilização (X) e temperatura (Y), encontrando-se um coeficiente de correlação de $0{,}70$. Determine a estimativa por intervalo, com 95% de confiança, para o coeficiente de correlação populacional.

$$n = 100 \quad r_{XY} = 0{,}70 \quad 1-\alpha = 0{,}95 \Rightarrow z_{\alpha/2} = 1{,}96$$

$$R = 0{,}5\ln\left(\frac{1+0{,}7}{1-0{,}7}\right) = 0{,}8673 \qquad \sigma_R = \frac{1}{\sqrt{100-3}} = 0{,}1015$$

$$c_1 = 2(0{,}8673 - 1{,}96 \cdot 0{,}1015) = 1{,}3366$$

$$c_2 = 2(0{,}8673 + 1{,}96 \cdot 0{,}1015) = 2{,}1326$$

$$IC : \left(\frac{e^{1{,}3366}-1}{1+e^{1{,}3366}}, \frac{e^{2{,}1326}-1}{1+e^{2{,}1326}}\right) \Rightarrow IC : (0{,}5839; 0{,}7881)$$

Sugestão

Construa um intervalo de confiança para o EXEMPLO 4.1.2.1[1]

4 Testes de Hipóteses para ρ_{XY}

Os testes de significância para o coeficiente de correlação são os seguintes:

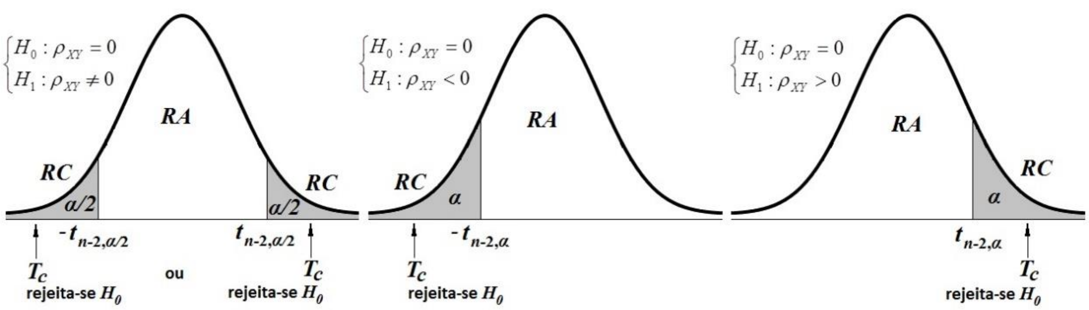
$$\begin{cases} H_0 : \rho_{XY} = 0 \\ H_1 : \rho_{XY} \neq 0 \end{cases} \quad \text{(teste bilateral)} \qquad \begin{cases} H_0 : \rho_{XY} = 0 \\ H_1 : \rho_{XY} < 0 \end{cases} \quad \text{(teste unilateral)} \qquad \begin{cases} H_0 : \rho_{XY} = 0 \\ H_1 : \rho_{XY} > 0 \end{cases} \quad \text{(teste unilateral)}$$

Já é conhecido que o critério de decisão dos testes de hipóteses (reveja na Parte 3 – Capítulo 4) é baseado no nível de significância do teste (α), que é a probabilidade de se cometer um erro do Tipo I, rejeitar H_0 quando ela é verdadeira. Sob a hipótese de que ρ_{XY} = 0 (H_0), a seguinte variável aleatória, função do coeficiente de correlação amostral r_{XY} , tem distribuição de probabilidade t de Student com n-2 graus de liberdade:

$$T = \frac{r_{XY}\sqrt{n-2}}{\sqrt{1-r_{XY}^2}}$$

Assim, esta será a estatística de prova, representada por T_C , e o critério de decisão é: se $|T_c| \geq t_{n-2,\alpha/2}$ (teste bilateral) ou $|T_c| \geq t_{n-2,\alpha}$ (teste unilateral) , rejeita-se H_0.

As figuras abaixo ilustram o procedimento.



EXEMPLO 4.1.4[1]

Considerando a situação do exemplo 4.3[1], teste a hipótese de que $\rho_{XY} = 0$, contra a hipótese de que $\rho_{XY} \neq 0$, ao nível de significância de 10%.

obs: como n-2=98 é grande e não se encontra na tabela (Anexo 2) vamos considerar ∞ graus de liberdade, veja que se fossem considerados 72 graus de liberdade, ou qualquer outro valor entre 72 e ∞ a decisão seria a mesma.

Sugestão

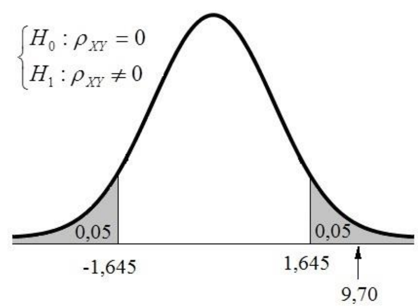
Para o EXEMPLO 4.1.2.1[1], teste a hipótese de que $\rho_{XY} = 0$ contra a alternativa de que $\rho_{XY} > 0$

$$n = 100 \qquad r_{XY} = 0{,}70$$

$$\alpha = 0{,}10 \Rightarrow t_{98;0,05} = 1{,}645$$

$$T_c = \frac{r_{XY}\sqrt{n-2}}{\sqrt{1-r_{XY}^2}} = \frac{0{,}70\sqrt{100-2}}{\sqrt{1-0{,}70^2}} = 9{,}70$$

$$|T_c| > t_{98;0,05} \Rightarrow \text{rejeita - se } H_0$$



Capítulo 2

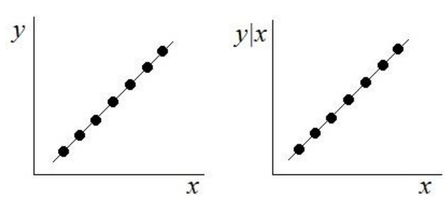
Regressão Linear Simples

1 A variável Aleatória $E(Y|X)$

Na seção 1 do capítulo 1 desta parte, foram discutidos aspectos da variável aleatória bidimensional (X,Y). Uma informação complementar que será acrescentada aqui é a existência da variável aleatória $E(Y|X)$, leia-se Esperança de Y dado X, cujos valores são dados por $E(Y|x)$, que é o valor esperado, ou média, da variável Y para um dado valor fixo x de X. O gráfico de $E(Y|X)$ é chamado de Curva de Regressão (da média) de Y em X. Caso a Curva de Regressão seja linear, a chamaremos de Reta de Regressão de Y em X. O mesmo vale para a variável $E(X|Y)$.

2 O problema da Regressão Linear

Foi visto que se $\rho_{XY} = \pm 1$, então Y é uma combinação linear de X, da forma $Y=A+BX$. Neste caso, a Curva de Regressão de Y em X também é linear, $E(Y|x)= \beta_0+\beta_1 x$, com $\beta_0=A$ e $\beta_1=B$. Isto significa que todos os pontos (x,y) e $(y|x,x)$ estão sobre a reta correspondente, como ilustrado na figura abaixo.

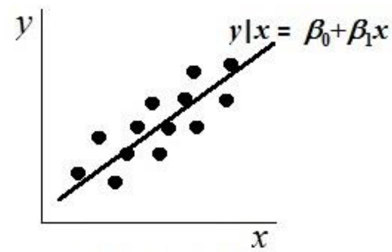


$\rho_{XY} = 1$

Observações

- Para simplificação da notação, usaremos a expressão $Y= \beta_0+\beta_1 x$, para representar a Reta de Regressão de Y em X, ressaltando que ela fornece a média da variável Y dado um valor fixo x de X;
- Este entendimento será muito importante no momento em que forem tratados os procedimentos de inferência para o modelo de regressão.

Caso contrário, a relação entre X e Y não será estritamente linear, o que não ocorre necessariamente com a curva de regressão de X em Y, que poderá ser uma reta, como ilustrado na figura abaixo.



$0 < \rho_{XY} < 1$

Na avaliação do comportamento de duas variáveis, caso haja algum indicativo de que elas tenham uma relação linear, o problema, então, consiste em encontrar a melhor aproximação da Curva de Regressão de Y em X a partir de uma amostra de pares (x,y). Lembramos que a ferramenta ideal para visualizar a possível relação entre X e Y é o Diagrama de Dispersão, as figuras apresentadas nesta seção são exemplos disto.

Importante

- O termo Linear Simples é utilizado pelo fato de existir apenas uma variável "preditora" X;
- Existem diversas situações em que a relação original das variáveis não é linear, mas que, através de uma tranformação simples, pode-se conduzir o caso para uma situação linear, por exemplo: $Z = aW^b \Rightarrow \underbrace{\ln Z}_{Y} = \underbrace{\ln a}_{b_0} + \underbrace{b}_{b_1}\underbrace{\ln W}_{X}$;
- O método para análise da Regressão Linear Simples pode ser estendido pra o caso de existirem mais de uma variável preditora, $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$, passando a se chamar de Regressão Linear Múltipla.

EXEMPLO 4.2.2[1]

Para a amostra abaixo, calcule o coeficiente de correlação entre X e Y, construa o diagrama de dispersão e faça o gráfico dos valores médios de Y dado X. Determine a equação que relaciona Y a X e a curva de regressão de Y em X

x	3	3	3	3	3	3	3	4	4	4	4	4	5	5	5	6	6	6	6	6
y	7	7	7	7	7	7	7	9	9	9	9	9	11	11	11	13	13	13	13	13

$$\sum x = 86 \quad \sum y = 192 \quad \sum x^2 = 398 \quad \sum y^2 = 1956 \quad \sum xy = 882$$

$$(\bar{y} \mid x=3)=\frac{7+7+7+7+7+7+7}{7}=7 \qquad (\bar{y} \mid x=4)=\frac{9+9+9+9+9}{5}=9$$

$$(\bar{y} \mid x=5)=\frac{11+11+11}{3}=11 \qquad (\bar{y} \mid x=6)=\frac{13+13+13+13+13}{5}=13$$

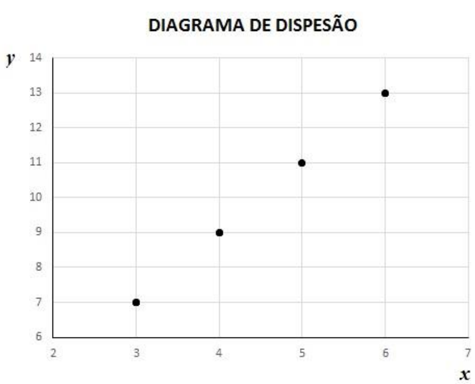
$$S_X^2=\frac{\sum x^2-\frac{\left(\sum x\right)^2}{n}}{n-1}=\frac{398-\frac{86^2}{20}}{19}=1{,}48 \Rightarrow S_X=1{,}22$$

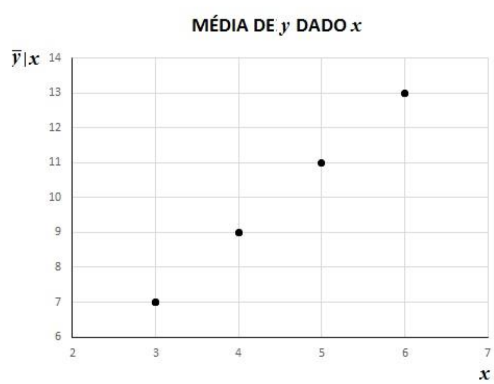
$$S_Y^2=\frac{\sum y^2-\frac{\left(\sum y\right)^2}{n}}{n-1}=\frac{1956-\frac{192^2}{20}}{19}=2{,}94 \Rightarrow S_Y=2{,}44$$

$$S_{XY}=\frac{\sum xy-\frac{\sum x\sum y}{n}}{n-1}=\frac{882-\frac{86\times 192}{20}}{19}=2{,}97$$

$$r_{XY}=\frac{S_{XY}}{S_X S_Y}=\frac{2{,}97}{1{,}22\times 2{,}44}=1$$

OBS: o valor encontrado foi 0,99716, por causa das aproximações dos valores utilizados, caso essas aproximações não fossem feitas o valor do coeficiente de correlação seria exatamente 1





Veja que no diagrama de dispersão aparecem apenas 4 pontos pois, como para cada x os valores de y são iguais, os pontos se sobrepõem, o que não é o caso do gráfico da regressão de y em x pois, neste caso, existem apenas 4 valores médios de y.

Tanto pelo valor do coeficiente de correlação quanto pela visualização dos gráficos temos que a relação entre y e x é linear, bem como a curva de regressão:

$$y=2x+1 \quad \text{e} \quad \bar{y} \mid x=2x+1$$

EXEMPLO 4.2.2[2]

Repetir o exemplo 4.2.2[1] para a amostra abaixo.

x	3	3	3	3	3	3	3	4	4	4	4	4	5	5	5	6	6	6	6	6
y	5	5,5	6	7	8	8,5	9	7,5	7	9,5	10	11	9,5	10,5	13	11	12,5	13	14	15

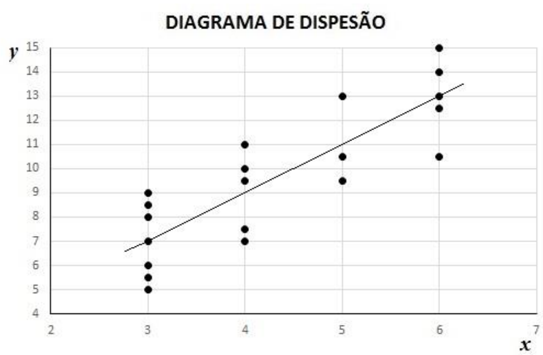
$$\sum x = 86 \quad \sum y = 192 \quad \sum x^2 = 398 \quad \sum y^2 = 2000 \quad \sum xy = 882$$

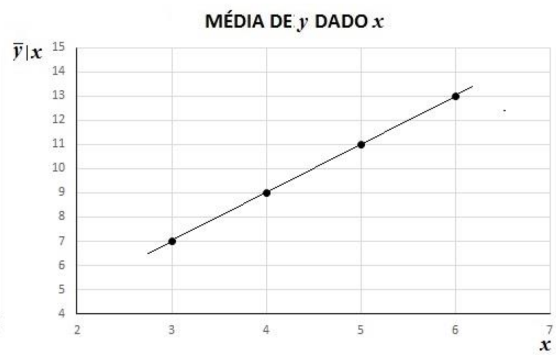
$$(\bar{y} \mid x = 3) = \frac{5 + 5{,}5 + 6 + 7 + 8 + 8{,}5 + 9}{7} = 7 \qquad (\bar{y} \mid x = 4) = \frac{7{,}5 + 7 + 9{,}5 + 10 + 11}{5} = 9$$

$$(\bar{y} \mid x = 5) = \frac{9{,}5 + 10{,}5 + 13}{3} = 11 \qquad (\bar{y} \mid x = 6) = \frac{11 + 12{,}5 + 13 + 14 + 15}{5} = 13$$

$$S_X^2 = 1{,}48 \Rightarrow S_X = 1{,}22 \qquad S_Y^2 = 8{,}25 \Rightarrow S_Y = 2{,}87$$

$$S_{XY} = 2{,}97 \;\; r_{XY} = 0{,}85$$





O diagrama de dispersão mostra que não existe uma relação perfeita entre X e Y, mas indica uma tendência linear crescente, o valor do coeficiente de correlação confirma isto, mas percebe-se que a curva de regressão é linear:

Observe que a linha de regressão, que também foi traçada no diagrama de dispersão, encontra-se "centralizada" na nuvem de pontos e é uma aproximação da relação entre X e Y e é utilizada para estimar a média da variável Y dado um valor x de X.

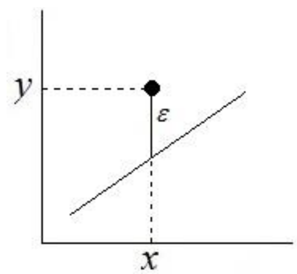
3 O método dos Mínimos Quadrados

Retirada uma amostra (x_1,y_1), (x_2,y_2),..., (x_n,y_n), nem todos os pontos estão sobre uma linha reta, assim, será acrescentado ao modelo de regressão um componente de erro, ε, que é a diferença entre o verdadeiro valor de y e aquele que seria se o mesmo estivesse sobre a reta. O modelo passa então a ser:

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

Assim, para a amostra de n observações existem, obviamente, n erros ε_i, correspondentes a cada um dos pares (x_i, y_i) obtidos. Portanto podemos escrever:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$



O método dos mínimos quadrados consiste em determinar estimativas dos parâmetros β_0 e β_1, b_0 e b_1 respectivamente, que minimizem a soma dos quadrados dos erros, determinando-se uma "reta estimada" ou "reta média" dada por:

$$\hat{y} = b_0 + b_1 x$$

Importante

- No modelo geral $Y = \beta_0 + \beta_1 x + \varepsilon$, o erro ε é uma variável aleatória normal com média 0 (zero) e variância desconhecida σ^2, ou seja $\varepsilon \sim N(0, \sigma^2)$;
- Também no modelo acima, x é um valor fixo da variável X;
- Detrminada a reta $\hat{y} = b_0 + b_1 x$, $\hat{y}$ é o valor esperado (média) da variável Y, dado um valor específico x da variável X.

O problema então é:

Dado $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, minimizar $S = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2$.

Observe que em S as variáveis são β_0 e β_1 . Calculando as derivadas parciais de S em relação a β_0 e β_1 e igualando-as a zero, chega-se ao seguinte sistema de equações, chamadas de equações normais, e respectiva solução, que fornece as expressões de b_0 e b_1 que são os estimadores desejados:

Desafio

Encontre as equações normais, b_0 e b_1

$$\begin{cases} \sum y - nb_0 - b_1 \sum x = 0 \\ \sum xy - b_0 \sum x - b_1 \sum x^2 = 0 \end{cases}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$b_1 = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}} = \frac{S_{XY}}{S_X^2} = r_{XY} \times \frac{S_Y}{S_X}$$

Desafio

Demonstre que

$$\frac{S_{XY}}{S_X^2} = r_{XY} \times \frac{S_Y}{S_X}$$

Observação

$S_{XY}, S_X^2, r_{XY}, S_Y$ e S_X são os estimadores da covariância , variância, desvio padrão e coeficiente de correlação, já vistos no Capítulo 1, Seção 2.1 desta Parte.

EXEMPLO 4.2.3[1]

Foram avaliados 30 computadores de Laboratórios de Informática de Escolas Públicas em relação ao tempo de uso (X) e custo de manutenção mensal (Y) dos mesmos. Os resultados estão no quadro abaixo. (a) construa o diagrama de dispersão; (b) calcule o coeficiente de correlação entre X e Y; (c) determine a reta de regressão e (d) estime o custo médio de manutenção para um computador com 5 anos de uso.

x	0,5	0,5	1	1	1,5	1,5	2	2	2	2,5	2,5	3	3	3,5	3,5	3,5	4	4	4,5	4,5
y	8	5	8	10	7	10	12	15	8	10	15	12	18	20	20	18	25	20	22	25

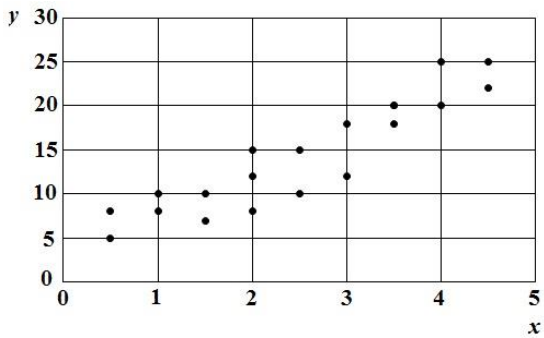
$$n = 20 \quad \sum x = 50{,}50 \quad \sum x^2 = 158{,}75 \quad \sum y = 288 \quad \sum y^2 = 4886 \quad \sum xy = 867$$

$$\bar{X} = \frac{50{,}50}{20} = 2{,}525 \qquad \bar{Y} = \frac{288}{20} = 14{,}4$$

$$S_X = \sqrt{\frac{158{,}75 - \frac{50{,}50^2}{20}}{19}} = 1{,}28 \qquad S_Y = \sqrt{\frac{4886 - \frac{288^2}{20}}{19}} = 6{,}24$$

$$S_{XY} = \frac{867 - \frac{50{,}50 \times 288}{20}}{19} = 7{,}36$$

DIAGARAMA DE DISPERSÃO



COEFICIENTE DE CORRELAÇÃO:

$$r_{XY} = \frac{7{,}36}{1{,}28 \times 6{,}24} = 0{,}9203$$

RETA DE REGRESSÃO:

$$b_1 = \frac{S_{XY}}{S_X^2} = \frac{7{,}36}{1{,}28^2} = 4{,}48 \text{ ou } b_1 = r_{XY} \times \frac{S_Y}{S_X} = 0{,}92 \times \frac{1{,}28}{6{,}24} = 4{,}48$$

$$b_0 = \overline{Y} - b_1 \overline{X} = 14{,}4 - 4{,}48 \times 2{,}525 = 3{,}1$$

$$\hat{y} = 3{,}1 + 4{,}48x$$

ESTIMATIVA DO CUSTO MÉDIO PARA 5 ANOS DE USO:

$$x = 5$$

$$\hat{y} = 3{,}1 + 4{,}48 \times 5 = 25{,}48$$

4 Coeficiente de Determinação x Coeficiente de Correlação

Uma das ferramentas utilizadas para analisar a eficiência do modelo de regressão é o que se denomina COEFICIENTE DE DETERMINAÇÃO, que é obtido a partir da seguinte identidade:

$$Y_i - \hat{Y}_i = Y_i - \bar{Y} - \left(\hat{Y}_i - \bar{Y}\right) \Rightarrow \left(Y_i - \hat{Y}_i\right)^2 = \left\{Y_i - \bar{Y} - \left(\hat{Y}_i - \bar{Y}\right)\right\}^2 \Rightarrow$$
$$\Rightarrow \sum \left(Y_i - \hat{Y}_i\right)^2 = \sum \left\{Y_i - \bar{Y} - \left(\hat{Y}_i - \bar{Y}\right)\right\}^2$$

Desenvolvendo a última igualdade, chega-se ao seguinte resultado:

$$\sum \left(Y_i - \bar{Y}_i\right)^2 = \sum \left(Y_i - \hat{Y}_i\right)^2 + \sum \left(\hat{Y}_i - \bar{Y}\right)^2$$

Observando os elementos envolvidos em cada soma de quadrados, este resultado pode ser expressado por:

SOMA DE QUADRADOS EM TORNO DA MÉDIA (**SQM**)	=	SOMA DE QUADRADOS EM TORNO DA REGRESSÃO (**SQE**)	+	SOMA DE QUADRADOS DEVIDO À REGRESSÃO (**SQR**)

Com a relação vista acima, temos que a variação de Y em torno de sua média (SQM), também chamada de soma de quadrado total, pode ser atribuída em parte à reta de regressão (SQR) e outra parte devido aos erros (SQE), pelo fato de que as observações reais não se situam sobre a reta de regressão. Se isto ocorresse, a soma de quadrados em torno da regressão (SQE) seria zero.

Portanto, quanto maior SQR for do que SQE, melhor será o ajustamento, pois, assim, a variação de Y em torno da média será melhor explicada pela reta de regressão. Ou, de forma equivalente, quanto mais próxima de 1 for a relação SQR/SQM, melhor será o ajustamento. Assim, o COEFICIENTE DE DETERMINAÇÃO é dado por:

$$R^2 = \frac{SQR}{SQM}$$ onde $$SQM = \sum y^2 - \frac{\left(\sum y\right)^2}{n}$$ e $$SQR = b_1\left(\sum xy - \frac{\sum x \sum y}{n}\right)$$

Importante

- Demonstra-se que, no caso do modelo de regressão linear simples, o quadrado do coeficiente de correlação linear é igual ao coeficiente de determinação, ou seja, $r_{XY}^2 = R^2$;
- Na Regressão Linear Múltipla, R^2 é o instrumento utilizado para uma verificação preliminar do grau de relacionamento linear entre a variável Y e as variáveis X, a exemplo de r_{XY} na regressão simples;
- r_{XY} e R^2 não têm o mesmo significado, o coeficiente de correlação indica o grau de associação linear entre X e Y, enquanto que o coeficiente de determinação avalia qual a parcela da variação total de Y é explicada pela reta de regressão estimada.

EXEMPLO 4.2.4[1]

Determine o coeficiente de determinação para o caso do exemplo 4.2.3[1], a partir de sua definição e compare com o coeficiente de correlação.

$$n = 20 \quad \sum y = 288 \quad \sum y^2 = 4886 \quad \sum xy = 867 \quad r_{XY} = 0{,}9203$$

$$SQM = 4886 - \frac{288^2}{20} = 738{,}8 \qquad SQR = 4{,}48 \times \left(867 - \frac{50{,}25 \times 288}{20}\right) = 625{,}66$$

$$R^2 = \frac{625{,}66}{738{,}8} = 0{,}8469 = 84{,}69\% \qquad r_{XY}^2 = 0{,}9203^2 = 0{,}8469 = R^2$$

5 Intervalos de Confiança e Testes de Hipóteses da Regressão

5.1 Intervalo de Confiança e Teste de Hipóteses para β_0

O estimador de β_0, b_0, é uma variável aleatória com Esperança e Variância dadas por:

$$E(b_0) = \beta_0 \quad \text{e} \quad V(b_0) = \sigma^2\left[\frac{1}{n} + \frac{\bar{X}^2}{(n-1)S_X^2}\right]$$

A expressão abaixo fornece um estimador não viciado de σ^2.

$$S^2 = \frac{SQE}{n-2}$$

Onde $SQE = SQM - SQR$ (reveja na seção 4 deste capítulo) e

$$SQM = \sum y^2 - \frac{\left(\sum y\right)^2}{n} \quad \text{e} \quad SQR = b_1\left(\sum xy - \frac{\sum x \sum y}{n}\right)$$

Assim, a seguinte variável tem distribuição t de Student com n-2 graus de liberdade.

$$T_0 = \frac{b_0 - \beta_0}{\sqrt{S^2\left[\frac{1}{n} + \frac{\overline{X}^2}{(n-1)S_X^2}\right]}}$$

O intervalo de confiança e a estatística de prova para H_0:β_0=$\beta_{0,0}$ são:

$$IC_{\beta_0} : \left(b_0 - t_{n-2,\alpha/2} \cdot \sqrt{S^2\left[\frac{1}{n} + \frac{\overline{X}^2}{(n-1)S_X^2}\right]}, b_0 + t_{n-2,\alpha/2} \cdot \sqrt{S^2\left[\frac{1}{n} + \frac{\overline{X}^2}{(n-1)S_X^2}\right]} \right)$$

$$T_{\beta_0} = \frac{b_0 - \beta_{0,0}}{\sqrt{S^2\left[\frac{1}{n} + \frac{\overline{X}^2}{(n-1)S_X^2}\right]}}$$

EXEMPLO 4.2.5.1[1]

Para a amostra seguinte determine o intervalo de confiança para o coeficiente linear da reta de regressão de Y em X e teste a hipótese de que ele é igual a 0 (zero), contra a hipótese de que é maior do que 0 (zero), use 1-α=0,95.

x	2,8	3,0	3,4	3,8	4,2	4,7	5,0	5,6
y	20,5	23,2	21,0	18,3	19,0	19,7	16,7	16,3

x	5,8	6,5	6,7	7,6	8,4	8,6	9,0	9,7
y	17,8	15,5	13,2	14,6	12,8	10,5	12,7	9,5

$$S_Y^2 = \frac{4502{,}01 - \frac{261{,}5^2}{16}}{15} = 15{,}21 \qquad S_{XY} = \frac{1424{,}71 - \frac{94{,}8 \times 261{,}5}{16}}{15} = -8{,}31$$

$$r_{XY} = \frac{-8{,}31}{\sqrt{5{,}03} \times \sqrt{15{,}21}} = -0{,}95 \qquad b_1 = -0{,}95 \times \frac{\sqrt{15{,}21}}{\sqrt{5{,}03}} = -1{,}65$$

$$n = 16 \quad \sum x = 94{,}8 \quad \sum x^2 = 637{,}08 \quad \sum y = 261{,}5 \quad \sum y^2 = 4502{,}01$$

$$\sum xy = 1424{,}71 \quad \overline{X} = \frac{94{,}8}{16} = 5{,}93 \quad \overline{Y} = \frac{261{,}5}{16} = 16{,}34 \quad S_X^2 = \frac{637{,}08 - \frac{94{,}8^2}{16}}{15} = 5{,}03$$

$$b_0 = 16{,}34 - (-1{,}65 \times 5{,}93) = 26{,}14 \quad SQM = 4502{,}01 - \frac{261{,}5^2}{16} = 228{,}12$$

$$SQR = -1{,}65 \times \left(1424{,}71 - \frac{94{,}8 \times 261{,}5}{16}\right) = 206{,}19 \quad SQE = 228{,}12 - 206{,}19 = 21{,}93$$

$$S^2 = \frac{21{,}93}{14} = 1{,}57$$

INTERVALO DE CONFIANÇA

$$n = 16 \quad 1 - \alpha = 0{,}95 \Rightarrow t_{14;0,025} = 2{,}145$$

$$IC_{\beta_0} : \left(26{,}14 - 2{,}145 \cdot \sqrt{1{,}57\left[\frac{1}{16} + \frac{5{,}93^2}{15 \times 5{,}03}\right]} ; 26{,}14 + 2{,}145 \cdot \sqrt{1{,}57\left[\frac{1}{16} + \frac{5{,}93^2}{15 \times 5{,}03}\right]} \right)$$

$$IC_{\beta_0} : (26{,}14 - 1{,}95 ; 26{,}14 + 1{,}95) \Rightarrow IC_{\beta_0} : (24{,}19 ; 28{,}09)$$

TESTE DE HIPÓTESES

$$n = 16 \quad 1 - \alpha = 0{,}95 \Rightarrow t_{14;0,025} = 2{,}145$$

$$\begin{cases} H_0 : \beta_0 = 0 \\ H_1 : \beta_0 \neq 0 \end{cases} \quad T_{\beta_0} = \frac{26{,}14 - 0}{\sqrt{1{,}57\left[\frac{1}{16} + \frac{5{,}93^2}{15 \times 5{,}03}\right]}} = 28{,}69$$

$$\left|T_{\beta_0}\right| > t_{14;0,025} \Rightarrow \text{rejeita - se } H_0$$

5.2 Intervalo de Confiança e Teste de Hipóteses para β_1

O estimador de β_1, b_1, é uma variável aleatória com Esperança e Variância dadas por:

$$E(b_1) = \beta_1 \quad \text{e} \quad V(b_1) = \frac{\sigma^2}{(n-1)S_X^2}$$

Assim, a seguinte variável tem distribuição *t-Student* com *n*-2 graus de liberdade, onde S^2 é o estimador de σ^2 (veja seção anterior):

$$T_{\beta_1} = \frac{b_1 - \beta_1}{\sqrt{\dfrac{S^2}{(n-1)S_X^2}}}$$

O intervalo de confiança e a estatística de prova para H_0:β_1=$\beta_{1,0}$ são:

$$IC_{\beta_1}:\left(b_1 - t_{n-1,\alpha/2}\cdot\sqrt{\frac{S^2}{(n-1)S_X^2}}\ , b_1 + t_{n-1,\alpha/2}\cdot\sqrt{\frac{S^2}{(n-1)S_X^2}}\right) \quad \text{e} \quad T_{\beta_1} = \frac{b_1 - \beta_{1,0}}{\sqrt{\dfrac{S^2}{(n-1)S_X^2}}}$$

Importante

No caso da Regressão Linear Múltipla, $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$, é utilizada a Análise de Variância, envolvendo as somas de quadrados vistas na seção sobre o Coeficiente de Determinação, na qual é aplicado o teste F (veja na bibliografia indicada) que verifica a adequação da reta ajustada, testando as hipóteses $H_0 : \beta_1 = \beta_2 = \ldots = \beta_n = 0$ e $H_1 : \text{pelo menos um } \beta_i \neq 0$). Na regressão simples o teste F é equivalente ao teste t apresentado.

EXEMPLO 4.2.5.2[1]

Considerando os dados do EXEMPLO 4.2.5.1[1], construa o intervalo de confiança para o coeficiente angular da reta de regressão e teste a hipótese de que ele é igual a 0 (zero) contra a hipótese alternativa de que é menor do que 0 (zero)

INTERVALO DE CONFIANÇA

$$n = 16 \quad 1-\alpha = 0{,}95 \Rightarrow t_{14;0,025} = 2{,}145$$

$$IC_{\beta_1}:\left(-1{,}65 - 2{,}145\cdot\sqrt{\frac{1{,}57}{15\times 5{,}03}}\ ;-1{,}65 + 2{,}145\cdot\sqrt{\frac{1{,}57}{15\times 5{,}03}}\right)$$

$$IC_{\beta_1}:(-1{,}65 - 0{,}31;-1{,}65 + 0{,}31) \Rightarrow IC_{\beta_0}:(\text{-}1{,}96\ ;-1{,}34)$$

TESTE DE HIPÓTESES

$n=16 \quad 1-\alpha=0{,}95 \Rightarrow \alpha=0{,}05 \Rightarrow t_{14;0,05}=1{,}761$

$$\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 < 0 \end{cases} \quad T_{\beta_1} = \frac{-1{,}65-0}{\sqrt{\dfrac{1{,}57}{15\times 5{,}03}}} = -11{,}438$$

$\left|T_{\beta_1}\right| > t_{14;0,025} \Rightarrow$ rejeita - se H_0

5.3 Intervalo de Confiança para o valor médio de Y dado x_k

Para um dado valor x_k, $\hat{y}_k = b_0 + b_1 x_k$ é um estimador do valor esperado (média) de Y dado o valor fixo x_k de X, ou seja, $\hat{y}_k$ é um estimador (não viciado) de $Y_k = E(Y \mid x_k)$, com média $\mu_{\hat{y}_k} = Y_k$ e variância $\sigma^2_{\hat{y}_k} = \sigma^2 \left[\dfrac{1}{n} + \dfrac{(x_k - \bar{X})^2}{(n-1)S_X^2} \right]$.

Assim, a variável abaixo tem distribuição *t-Student* com *n*-2 graus de liberdade.

$$T = \frac{\hat{y}_k - Y_k}{\sqrt{S^2 \left[\dfrac{1}{n} + \dfrac{(x_k - \bar{X})^2}{(n-1)S_X^2} \right]}}$$

A estimativa por intervalo, com $100(1-\alpha)\%$ de confiança, para $E(Y|x_k)$ é dada, então, por:

$$IC_{Y_k}: \left(\hat{y}_k - t_{n-2,\alpha/2} \times \sqrt{S^2 \left[\frac{1}{n} + \frac{(x_k - \bar{X})^2}{(n-1)S_X^2} \right]} ; \hat{y}_k + t_{n-2,\alpha/2} \times \sqrt{S^2 \left[\frac{1}{n} + \frac{(x_k - \bar{X})^2}{(n-1)S_X^2} \right]} \right)$$

Importante

Veja que o termo $\sqrt{S^2 \left[\frac{1}{n} + \frac{(x_k - \bar{X})^2}{(n-1)S_X^2} \right]}$ é mínimo quando $x_k = \bar{X}$, e cresce à medida que x_k se afasta de $\bar{X}$, em ambos os sentidos, isto significa que quanto mais afastado de $\bar{X}$ estiver o valor de X, menor será a precisão da estimativa do valor médio de Y dado o valor x_k.

EXEMPLO 4.2.5.3[1]

Para o exemplo 4.2.5.1[1], calcule os intervalos de confiança (1-α=0,95) para todos os pares observados, construa um gráfico com o diagrama de dispersão, a reta de regressão estimada e as curvas dos limites dos intervalos de confiança.

$$n=16 \quad \sum x=94{,}8 \quad \sum x^2=637{,}08 \quad \sum y=261{,}5 \quad \sum y^2=4502{,}01 \quad \sum xy=1424{,}71$$

$$\overline{X}=5{,}93 \quad \overline{Y}=16{,}34 \quad S_X^2=5{,}03 \quad S_Y^2=15{,}21 \quad S_{XY}=-8{,}31 \quad b_1=-1{,}65 \quad b_0=26{,}14$$

$$\hat{y}=26{,}14-1{,}65x \quad SQM=228{,}12 \quad SQR=206{,}19 \quad SQE=21{,}93 \quad S^2=1{,}57$$

$$n=16 \Rightarrow 16-2=14\,GL \quad 1-\alpha=0{,}95 \Rightarrow \alpha=0{,}05 \quad \Rightarrow \quad t_{14;0{,}025}=2{,}145$$

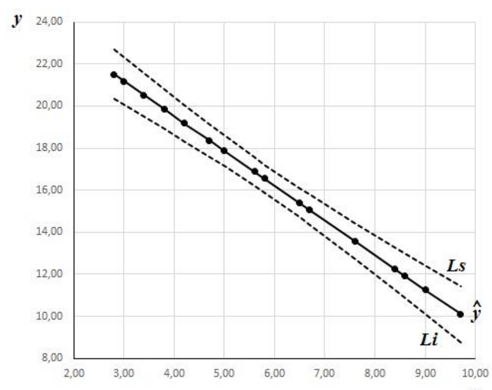
$$IC_{Y_1}:\left(\hat{y}_1-t_{n-2,\alpha/2}\times\sqrt{S^2\left[\frac{1}{n}+\frac{(x_1-\overline{X})^2}{(n-1)S_X^2}\right]};\hat{y}_1+t_{n-2,\alpha/2}\times\sqrt{S^2\left[\frac{1}{n}+\frac{(x_1-\overline{X})^2}{(n-1)S_X^2}\right]}\right)$$

$$IC_{Y_1}:\left(21{,}51-2{,}145\times\sqrt{1{,}57\left[\frac{1}{16}+\frac{(2{,}8-5{,}93)^2}{15\times 5{,}03}\right]};21{,}51+2{,}145\times\sqrt{1{,}57\left[\frac{1}{16}+\frac{(2{,}8-5{,}93)^2}{15\times 5{,}03}\right]}\right)$$

$$IC_{Y_1}:(21{,}51-1{,}18;21{,}51+1{,}18) \qquad IC_{Y_1}:(20{,}33;22{,}69)$$

Os demais intervalos são determinados analogamente.

x	y	$\hat{y}$	INTERVALO DE CONFIANÇA – 95%		
			erro	*Li*	*Ls*
2,8	20,50	21,51	1,18	20,34	22,69
3,0	23,20	21,18	1,13	20,05	22,31
3,4	21,00	20,52	1,03	19,49	21,55
3,8	18,30	19,86	0,94	18,92	20,80
4,2	19,00	19,20	0,86	18,34	20,05
4,7	19,70	18,37	0,77	17,60	19,14
5,0	16,70	17,87	0,73	17,14	18,60
5,6	16,30	16,88	0,68	16,20	17,56
5,8	17,80	16,55	0,67	15,88	17,22
6,5	15,50	15,39	0,69	14,70	16,09
6,7	13,20	15,06	0,71	14,35	15,77
7,6	14,60	13,57	0,85	12,73	14,42
8,4	12,80	12,25	1,02	11,23	13,27
8,6	10,50	11,92	1,07	10,85	12,99
9,0	12,70	11,26	1,16	10,09	12,42
9,7	9,50	10,10	1,35	8,75	11,45



Sugestão

Faça os cálculos para os outros intervalos de confiança

Observe na tabela, que está em ordem crescente de x, que o valor do erro no intervalo de confiança diminui à medida que x se aproxima de $\overline{X} = 5{,}93$ e passa a crescer a partir de então, indicando que quanto mais distante da média dos valores de X estiver o valor para o qual se pretende estimar Y, menor será a precisão da estimativa. As linhas pontilhadas do gráfico ilustram este fato.

5.4 Intervalo de Confiança para um valor individual de Y dado x_k

Caso se deseje estimar o valor individual Y_k de Y que seria obtido caso se fizesse uma nova observação x_k de X, por exemplo um valor futuro, o estimador continua sendo $\hat{y}_k = b_0 + b_1 x_k$, com $\mu_{\hat{y}_k} = Y_k$ mas a variância de $\hat{y}_k$ passa a ser $\sigma^2_{\hat{y}_k} = \sigma^2\left[1 + \frac{1}{n} + \frac{(x_k - \overline{X})^2}{(n-1)S_X^2}\right]$ e a variável T será:

$$T = \frac{\hat{y}_k - Y_k}{\sqrt{S^2\left[1 + \frac{1}{n} + \frac{(x_k - \overline{X})^2}{(n-1)S_X^2}\right]}}$$

E o Intervalo de Confiança será:

$$IC_{Y_k} : \hat{y}_k - t_{n-2,\alpha/2} \times \sqrt{S^2\left[1 + \frac{1}{n} + \frac{(x - \overline{X})^2}{(n-1)S_X^2}\right]} ; \hat{y}_k + t_{n-2,\alpha/2} \times \sqrt{S^2\left[1 + \frac{1}{n} + \frac{(x - \overline{X})^2}{(n-1)S_X^2}\right]}$$

Importante

Como no caso da seção anterior, o termo $\sqrt{S^2 1 + \left[\frac{1}{n} + \frac{(x_k - \overline{X})^2}{(n-1)S_X^2}\right]}$ é mínimo quando $x_k = \overline{X}$ e cresce à medida que x_k se afasta de $\overline{X}$, em ambos os sentidos, isto significa que quanto mais afastado de $\overline{X}$ estiver o novo valor de X para o qual se deseja estimar Y, menor será a precisão da estimativa.

EXEMPLO 4.2.5.4[1]

Para o exemplo 4.3.3[1], calcule as previsões do custo mensal de manutenção para computadores com 5, 8 e 10 anos de uso. Determine os intervalos de confiança para estas estimativas, com 98% de confiança.

$$n = 20 \quad \sum x = 50{,}50 \quad \sum x^2 = 158{,}75 \quad \sum y = 288 \quad \sum y^2 = 4886 \quad \sum xy = 867$$

$$\overline{X} = 2{,}525 \quad \overline{Y} = 14{,}4 \quad S_X^2 = 1{,}28 \quad S_Y^2 = 6{,}24 \quad S_{XY} = 7{,}36$$

$$b_0 = 3{,}1 \quad b_1 = 4{,}48 \quad \hat{y} = 3{,}1 + 4{,}48x$$

$$SQM = 738{,}8 \quad SQR = 625{,}66 \quad SQE = 113{,}14 \quad S^2 = 6{,}29$$

$$n = 20 \Rightarrow 18\,GL \quad 1-\alpha = 0{,}98 \Rightarrow \alpha = 0{,}02 \Rightarrow t_{18;0{,}01} = 2{,}552$$

$$x = 5 \Rightarrow \hat{y} = 3{,}1 + 4{,}48 \times 5 = 25{,}48$$

$$IC_{Y_k}: \left(\begin{array}{l} 25{,}48 - 2{,}552 \times \sqrt{6{,}29\left[1 + \dfrac{1}{20} + \dfrac{(5-2{,}525)^2}{19 \times 1{,}64}\right]}; \\ 25{,}48 + 2{,}552 \times \sqrt{6{,}29\left[1 + \dfrac{1}{20} + \dfrac{(5-2{,}525)^2}{19 \times 1{,}64}\right]} \end{array} \right)$$

$$IC_{Y_k}: (25{,}48 - 7{,}14; 25{,}48 + 7{,}14) \Rightarrow IC_{Y_k}: (18{,}34; 32{,}62)$$

Sugestão

- **Faça os** cálculos para os intervalos de confiança para x=8 e x=10 ;
- Repita o exemplo anterior (4.2.5.3[1]) para este exemplo.

x	$\hat{y}$	INTERVALO DE CONFIANÇA (1-α=0,98)		
		Li	*Ls*	*Erro*
5	25,48	18,34	32,62	7,14
8	38,90	29,83	47,97	9,07
10	47,85	37,07	58,63	10,78

Atividades de avaliação

1) Suponha que a variável aleatória bidimensional (X,Y) tenha a seguinte distribuição de probabilidade conjunta.

Y	X		
	-1	0	1
-1	1/12	1/6	1/12
0	1/6	0	1/6
1	1/12	1/6	1/12

a) Mostre que $E(XY)=E(X)E(Y)$ e, consequentemente, $\rho_{XY}=0$

b) Mostre que, embora não sejam correlacionadas (item *a*), X e Y não são independentes

2) Para a distribuição conjunta abaixo, mostre que X e Y são independentes e, consequentemente, são não correlacionadas.

Y	X		
	-1	0	1
-1	0,21	0,06	0,03
0	0,14	0,04	0,02
1	0,35	0,10	0,05

3) Uma pesquisa envolvendo uma amostra de 20 observações, resultou nos dados mostrados na tabela abaixo. Construa os diagramas de dispersão e calcule os coeficientes de correlação, para as variáveis (X,Y_1), (X,Y_2) e (X,Y_3).

OBS	X	Y_1	Y_2	Y_3
1	2,0	9,0	4,0	3,7
2	2,5	8,0	3,8	4,0
3	2,5	9,5	3,4	3,5
4	3,0	12,0	3,2	4,3
5	3,0	10,0	3,5	4,0
6	3,5	12,5	3,4	3,5
7	3,5	14,0	3,0	3,7
8	4,0	13,5	3,2	3,2
9	4,0	15,5	3,5	4,5
10	4,0	16,5	2,9	4,2

OBS	X	Y_1	Y_2	Y_3
11	4,0	14,0	2,7	3,5
12	4,5	15,0	3,0	4,0
13	5,0	16,0	2,6	3,3
14	5,5	18,0	2,8	4,0
15	5,5	19,5	2,5	3,6
16	6,0	18,0	2,3	4,2
17	6,0	22,0	2,5	3,8
18	6,5	21,5	2,4	3,6
19	6,5	19,5	2,7	4,0
20	7,0	25,0	2,2	3,8

4) Para as variáveis do exercício 3, construa os intervalos com 90% de confiança para os coeficientes de correlação populacionais.

5) Para as variáveis do exercício 3, faça o teste de que os coeficientes de correlação populacionais são iguais a 0 (zero) contra as seguintes hipóteses alternativas (use α=0,02): a) $\rho_{XY_1}>0$ b) $\rho_{XY_2}<0$ b) $\rho_{XY_3}\neq 0$

6) Para as variáveis do exercício 3 encontre as equações das retas de regressão e construa seus gráficos no mesmo plano do diagrama de dispersão. Calcule os coeficientes de determinação e compare-os com os coeficientes de correlação encontrados no exercício 4.

7) Para as retas de regressão encontradas no exercício 6, construa intervalos de confiança para os coeficientes lineares e angulares, use $1-\alpha=0,95$.

8) Para as retas de regressão encontradas no exercício 6, teste as hipóteses de que os coeficientes lineares e angulares são iguais a 0 (zero), contra as hipóteses de que são diferentes de 0 (zero), use α=0,05.

9) Para as variáveis do exercício 3 encontre os valores estimados de Y_i (i=1,2,3) e seus respectivos intervalos de confiança, use $1-\alpha=0,98$. Faça os gráficos das retas de regressão com as respectivas curvas dos intervalos de confiança.

10) A partir das retas encontradas no exercício 6 estime os valores de Y_i (i=1,2,3) para X=10 e determine seus respectivos intervalos de confiança, use $1-\alpha=0,90$.

Solução das atividades

1)

Y	X			
	-1	0	1	SOMA
-1	1/12	1/6	1/12	1/3
0	1/6	0	1/6	1/3
1	1/12	1/6	1/12	1/3
SOMA	1/3	1/3	1/3	1

$a)\ \ E(X) = -1\times\frac{1}{3} + 0\times\frac{1}{3} + 1\times\frac{1}{3} = 0 \qquad E(Y) = -1\times\frac{1}{3} + 0\times\frac{1}{3} + 1\times\frac{1}{3} = 0$

$$E(XY) = \left[\left(-1\times -1\times\frac{1}{12}\right) + \left(-1\times 0\times\frac{1}{6}\right) + \left(-1\times 1\times\frac{1}{12}\right)\right] +$$

$$+\left[\left(0\times -1\times\frac{1}{6}\right) + (0\times 0\times 0) + \left(0\times 1\times\frac{1}{6}\right)\right] + \left[\left(1\times -1\times\frac{1}{12}\right) + \left(1\times 0\times\frac{1}{6}\right) + \left(1\times 1\times\frac{1}{12}\right)\right] = 0$$

$$\rho_{XY} = \frac{E(XY) - E(X)E(Y)}{\sigma_X\sigma_Y} = \frac{0 - 0\times 0}{\sigma_X\sigma_Y} = 0$$

$b)\ \ P(X=-1)\times P(Y=-1) = \frac{1}{3}\times\frac{1}{3} = \frac{1}{9} \neq \frac{1}{12} = P(X=-1, Y=-1)$, então X e Y não são independentes, pois para que o sejam é necessário que $P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j)$ para todos os pares (x_i,y_j)

2)

Y	X			
	-1	0	1	SOMA
-1	0,21	0,06	0,03	0,30
0	0,14	0,04	0,02	0,20
1	0,35	0,10	0,05	0,50
SOMA	0,70	0,20	0,10	1

$a)\ \ P(X=-1)\times P(Y=-1) = 0{,}70\times 0{,}30 = 0{,}21 = P(X=-1, Y=-1)$

$P(X=-1)\times P(Y=0) = 0{,}70\times 0{,}20 = 0{,}14 = P(X=-1, Y=-1)$

...

$P(X=1)\times P(Y=1) = 0{,}10\times 0{,}50 = 0{,}05 = P(X=1, Y=1)$

Como $P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j)$ para todos os pares (x_i,y_j), X e Y são independentes.

$$b)\quad E(X) = -1\times 0{,}70 + 0\times 0{,}20 + 1\times 0{,}10 = -0{,}6$$

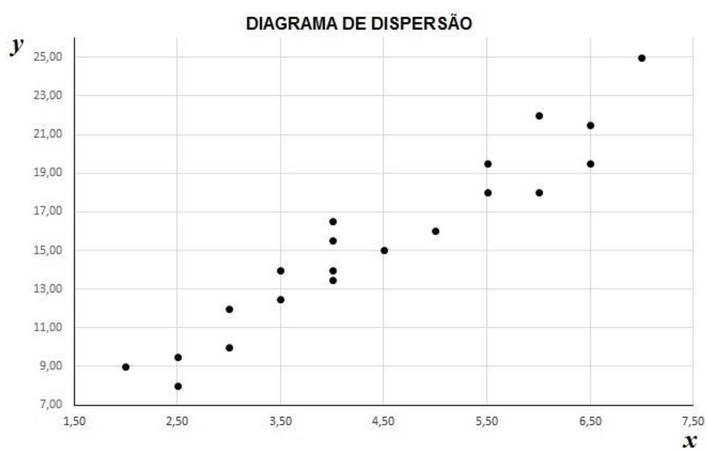
$$E(Y) = -1\times 0{,}3 + 0\times 0{,}20 + 1\times 0{,}50 = 0{,}2$$

$$E(XY) = \left[(-1\times -1\times 0{,}21) + (-1\times 0\times 0{,}14) + (-1\times 1\times 0{,}35)\right] +$$
$$+\left[(0\times -1\times 0{,}06) + (0\times 0\times 0{,}04) + (0\times 1\times 0{,}10)\right] +$$
$$+\left[(1\times -1\times 0{,}03) + (1\times 0\times 0{,}02) + (1\times 1\times 0{,}05)\right] = -0{,}12$$

$$\rho_{XY} = \frac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y} = \frac{-0{,}12 - (-0{,}6\times 0{,}2)}{\sigma_X \sigma_Y} = 0$$

3)

SOLUÇÃO PARA (X, Y_1)



$$n = 20 \quad \sum x = 88{,}5 \quad \sum x^2 = 434{,}25 \quad \sum y = 309$$

$$\sum y_1^2 = 5183 \quad \sum xy = 1493{,}75$$

$$\bar{X} = \frac{\sum x}{n} = \frac{88{,}5}{20} = 4{,}425 \qquad \bar{Y} = \frac{\sum y}{n} = \frac{309}{20} = 15{,}45$$

$$S_X = \sqrt{\frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1}} = \sqrt{\frac{434{,}25 - \frac{88{,}5^2}{20}}{19}} = 1{,}498$$

$$S_Y = \sqrt{\frac{\sum y^2 - \frac{\left(\sum y\right)^2}{n}}{n-1}} = \sqrt{\frac{5183 - \frac{309^2}{20}}{19}} = 4{,}639$$

$$S_{XY} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{n-1} = \frac{1493{,}75 - \frac{88{,}5 \times 309}{20}}{19} = 6{,}654$$

COEFICIENTE DE CORRELAÇÃO: $r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{6{,}654}{1{,}498 \times 4{,}639} = 0{,}9574$

As soluções para (X,Y_2) e (X,Y_3) são análogas, ficando a cargo do aluno.

4)

SOLUÇÃO PARA (X,Y_1)

$$n = 20 \quad 1-\alpha = 0{,}90 \Rightarrow z_{0,05} = 1{,}64 \quad r_{XY} = 0{,}9574$$

$$\sigma_R = \sqrt{\frac{1}{n-3}} = \sqrt{\frac{1}{20-3}} = 0{,}2425 \quad R = \frac{1}{2}\ln\left(\frac{1+r_{XY}}{1-r_{XY}}\right) = \frac{1}{2}\ln\left(\frac{1+0{,}9574}{1-0{,}9574}\right) = 1{,}914$$

$$c_1 = 2 \times (R - z_{\alpha/2} \times \sigma_R) = 2 \times (1{,}914\text{-}1{,}64 \times 0{,}2425) = 3{,}0325$$

$$c_2 = 2 \times (R + z_{\alpha/2} \times \sigma_R) = 2(1{,}914 + 1{,}64 \times 0{,}2425) = 4{,}6235$$

$$IC_\rho : \left(\frac{e^{c_1}-1}{1+e^{c_1}}; \frac{e^{c_2}-1}{1+e^{c_2}}\right) \Rightarrow IC_\rho : \left(\frac{e^{3,0325}-1}{1+e^{3,0325}}; \frac{e^{4,6235}-1}{1+e^{4,6235}}\right) \Rightarrow IC_\rho : (0{,}9080; 0{,}9806)$$

As soluções para (X,Y_2) e (X,Y_3) são análogas, ficando a cargo do aluno.

5)

SOLUÇÃO PARA (X,Y_1)

$$n = 20 \quad \alpha = 0{,}02 \Rightarrow t_{18;0,02} = 2{,}214 \quad r_{XY} = 0{,}9574$$

$$\begin{cases} H_0 : \rho_{XY} = 0 \\ H_0 : \rho_{XY} = 0 \end{cases}$$

$$T = \frac{r_{XY} \times \sqrt{n-2}}{\sqrt{1-r_{XY}^2}} = \frac{0{,}9574 \times \sqrt{20-2}}{\sqrt{1-0{,}9574^2}} = 14{,}07 \Rightarrow |T| > t_{18;0,02} \Rightarrow \text{rejeita -se } H_0$$

As soluções para (X,Y_2) e (X,Y_3) são análogas, ficando a cargo do aluno.

6)

SOLUÇÃO PARA (X,Y_1)

$n = 20 \quad \sum x = 88,5 \quad \sum x^2 = 434,25 \quad \sum y = 309 \quad \sum y^2 = 5183 \quad \sum xy = 1493,75$

$\overline{X} = 4,425 \quad \overline{Y} = 15,45 \quad S_X = 1,498 \quad S_Y = 4,639 \quad S_{XY} = 6,654 \quad r_{XY} = 0,9574$

$$b_1 = \frac{S_{XY}}{S_X^2} = \frac{6,654}{1,498^2} = 2,9651 \text{ ou } b_1 = r_{XY} \times \frac{S_Y}{S_X} = 0,9574 \times \frac{4,639}{1,498} = 2,9651$$

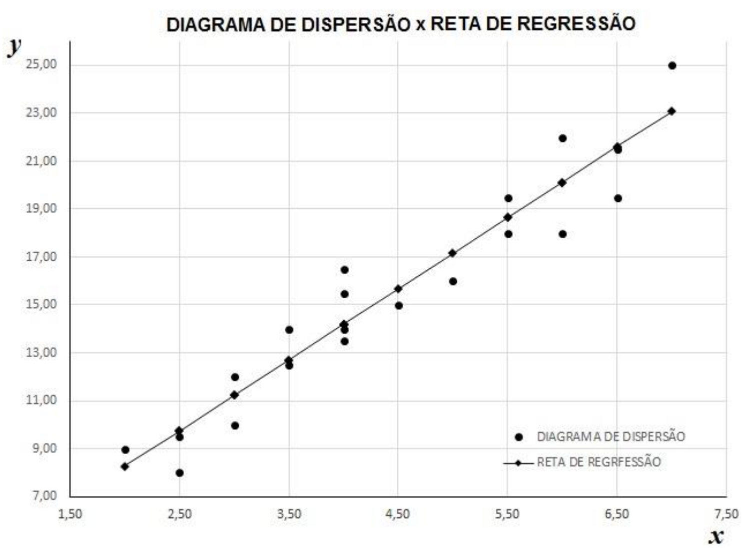
$$b_0 = \overline{Y} - b_1 \overline{X} = 15,45 - 1,9651 \times 4,425 = 2,3294$$

$$\hat{y} = 2,3294 + 2,9651x$$

$$SQM = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 5183 - \frac{309^2}{20} = 408,95$$

$$SQR = b_1 \left(\sum xy - \frac{\sum x \sum y}{n} \right) = 2,9651 \times \left(1493,75 - \frac{88,5 \times 309}{20} \right) = 374,8644$$

$$R^2 = \frac{SQR}{SQM} = 0,9167 \qquad r_{XY}^2 = 0,9574^2 = 0,9167 = R^2$$



As soluções para (X,Y_2) e (X,Y_3) são análogas, ficando a cargo do aluno.

7)

SOLUÇÃO PARA (X,Y_1)

$$n = 20 \quad \overline{X} = 4{,}425 \quad S_X = 1{,}498 \quad b_0 = 2{,}3294 \quad b_1 = 2{,}9651$$
$$SQM = 408{,}95 \quad SQR = 374{,}8644$$

$$SQE = SQM - SQR = 408{,}95 - 374{,}8644 = 34{,}0856$$
$$S^2 = \frac{SQE}{n-2} = \frac{34{,}0856}{20-2} = 1{,}8936$$

INTERVALO DE CONFIANÇA PARA O COEFICIENTE LINEAR (β_0)

$$1-\alpha = 0{,}95 \Rightarrow t_{18;0{,}025} = 2{,}101$$
$$IC_{\beta_0:}\left(b_0 - t_{n-2,\alpha/2} \times \sqrt{S^2\left[\frac{1}{n} + \frac{\overline{X}^2}{(n-1)S_X^2}\right]}\ , b_0 + t_{n-2,\alpha/2} \times \sqrt{S^2\left[\frac{1}{n} + \frac{\overline{X}^2}{(n-1)S_X^2}\right]}\right)$$
$$IC_{\beta_0:}\left(\begin{array}{l} 2{,}3294 - 2{,}101 \times \sqrt{1{,}8936 \times \left[\frac{1}{20} + \frac{4{,}425^2}{(20-1)1{,}498^2}\right]}\ ,; \\ 2{,}3294 + 2{,}101 \times \sqrt{1{,}8936 \times \left[\frac{1}{20} + \frac{4{,}425^2}{(20-1)1{,}498^2}\right]} \end{array}\right)$$
$$IC_{\beta_0:}(2{,}3294 - 2{,}0632; 2{,}3294 + 2{,}0632) \Rightarrow IC_{\beta_0:}(0{,}2662; 4{,}3926)$$

INTERVALO DE CONFIANÇA PARA O COEFICIENTE ANGULAR (β_1)

$$1-\alpha = 0{,}95 \Rightarrow t_{18;0{,}025} = 2{,}101$$
$$IC_{\beta_1:}\left(b_1 - t_{n-1,\alpha/2} \times \sqrt{\frac{S^2}{(n-1)S_X^2}}\ , b_1 + t_{n-1,\alpha/2} \times \sqrt{\frac{S^2}{(n-1)S_X^2}}\right)$$
$$IC_{\beta_1:}\left(2{,}9651 - 2{,}101 \times \sqrt{\frac{1{,}8936}{(20-1) \times 1{,}498^2}}\ , 2{,}9651 + 2{,}101 \times \sqrt{\frac{1{,}8936}{(20-1) \times 1{,}498^2}}\right)$$
$$IC_{\beta_1:}(2{,}9651 - 0{,}4428\ , 2{,}9651 + 0{,}4428) \Rightarrow IC_{\beta_1:}(2{,}5223; 3{,}4079)$$

As soluções para (X,Y_2) e (X,Y_3) são análogas, ficando a cargo do aluno.

8)

SOLUÇÃO PARA (X,Y_1)

$$n = 20 \quad \overline{X} = 4{,}425 \quad S_X = 1{,}498 \quad b_0 = 2{,}3294$$
$$b_1 = 2{,}9651 \quad SQM = 408{,}95 \quad SQR = 374{,}8644$$

$$SQE = SQM - SQR = 408{,}95 - 374{,}8644 = 34{,}0856$$
$$S^2 = \frac{SQE}{n-2} = \frac{34{,}0856}{20-2} = 1{,}8936$$

TESTE DE HIPÓTESES PARA O COEFICIENTE LINEAR (β_0)

$$\alpha = 0{,}05 \Rightarrow t_{18;0,025} = 2{,}101$$
$$\begin{cases} H_0 : \beta_0 = 0 \\ H_1 : \beta_0 \neq 0 \end{cases}$$
$$T_{\beta_0} = \frac{b_0 - \beta_{0,0}}{\sqrt{S^2\left[\frac{1}{n} + \frac{\overline{X}^2}{(n-1)S_X^2}\right]}} = \frac{2{,}3294 - 0}{\sqrt{1{,}8936 \times \left[\frac{1}{20} + \frac{4{,}425^2}{(20-1)\times 1{,}498^2}\right]}} = 2{,}372$$
$$\left|T_{\beta_0}\right| > t_{18;0,025} \Rightarrow \text{rejeita - se } H_0$$

TESTE DE HIPÓTESES PARA O COEFICIENTE ANGULAR (β_1)

$$\alpha = 0{,}05 \Rightarrow t_{18;0,025} = 2{,}101$$
$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$
$$T_{\beta_1} = \frac{b_1 - \beta_{1,0}}{\sqrt{\frac{S^2}{(n-1)S_X^2}}} = \frac{2{,}9651 - 0}{\sqrt{\frac{1{,}8936}{(20-1)\times 1{,}498^2}}} = 14{,}07$$
$$\left|T_{\beta_1}\right| > t_{18;0,025} \Rightarrow \text{rejeita - se } H_0$$

As soluções para (X,Y_2) e (X,Y_3) são análogas, ficando a cargo do aluno.

9)

SOLUÇÃO PARA (X, Y_1)

$n = 20 \quad \bar{X} = 4{,}425 \quad S_X = 1{,}498 \quad b_0 = 2{,}3294 \quad b_1 = 2{,}9651 \quad S^2 = 1{,}8936$

$$IC_{Y_k}: \left(\hat{y}_k - t_{n-2,\alpha/2} \times \sqrt{S^2\left[\frac{1}{n} + \frac{(x_k - \bar{X})^2}{(n-1)S_X^2}\right]}; \hat{y}_k + t_{n-2,\alpha/2} \times \sqrt{S^2\left[\frac{1}{n} + \frac{(x_k - \bar{X})^2}{(n-1)S_X^2}\right]} \right)$$

Para o primeiro par observado:

$x_1 = 2 \quad y_1 = 9 \quad \hat{y}_1 = 2{,}3294 + 2{,}9651 \times 2 = 8{,}26$

$1 - \alpha = 0{,}98 \Rightarrow t_{18;0{,}01} = 2{,}552$

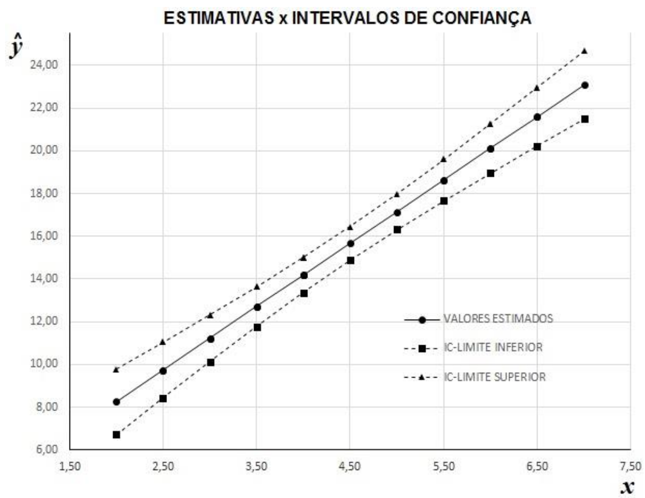
$$IC_{Y_1}: \left(\begin{array}{l} 8{,}26 - 2{,}552 \times \sqrt{1{,}8936 \times \left[\frac{1}{20} + \frac{(2-4{,}425)^2}{(20-1)\times 1{,}498^2}\right]}; \\ 8{,}26 + 2{,}552 \times \sqrt{1{,}8936 \times \left[\frac{1}{20} + \frac{(2-4{,}425)^2}{(20-1)\times 1{,}498^2}\right]} \end{array} \right)$$

$IC_{Y_1}: (8{,}26 - 1{,}52; 8{,}26 + 1{,}52) \Rightarrow IC_{Y_1}: (6{,}74; 9{,}78)$

Para o demais pares o procedimento é análogo, fica a cargo do aluno.

OBS	x	y	$\hat{y}$	INTERVALO DE CONFIANÇA		
				erro	LI	LS
1	2	9	8,26	1,52	6,74	9,78
2	2,5	8	9,74	1,3	8,44	11,04
3	2,5	9,5	9,74	1,3	8,44	11,04
4	3	12	11,22	1,1	10,13	12,32
5	3	10	11,22	1,1	10,13	12,32
6	3,5	12,5	12,71	0,93	11,78	13,64
7	3,5	14	12,71	0,93	11,78	13,64
8	4	13,5	14,19	0,82	13,37	15,01
9	4	15,5	14,19	0,82	13,37	15,01
10	4	16,5	14,19	0,82	13,37	15,01

OBS	x	y	$\hat{y}$	INTERVALO DE CONFIANÇA		
				erro	LI	LS
11	4	14	14,19	0,82	13,37	15,01
12	4,5	15	15,67	0,79	14,89	16,46
13	5	16	17,15	0,84	16,31	18
14	5,5	18	18,64	0,98	17,66	19,61
15	5,5	19,5	18,64	0,98	17,66	19,61
16	6	18	20,12	1,16	18,96	21,28
17	6	22	20,12	1,16	18,96	21,28
18	6,5	21,5	21,6	1,36	20,24	22,97
19	6,5	19,5	21,6	1,36	20,24	22,97
20	7	25	23,09	1,59	21,49	24,68



As soluções para (X,Y_2) e (X,Y_3) são análogas, ficando a cargo do aluno.

10)

SOLUÇÃO PARA (X,Y_1)

$$n=20 \quad \overline{X}=4{,}425 \quad S_X=1{,}498 \quad b_0=2{,}3294 \quad b_1=2{,}9651 \quad S^2=1{,}8936$$

$$IC_{Y_k}:\left(\hat{y}_k - t_{n-2,\alpha/2}\times\sqrt{S^2\left[1+\frac{1}{n}+\frac{(x-\overline{X})^2}{(n-1)S_X^2}\right]};\hat{y}_k + t_{n-2,\alpha/2}\times\sqrt{S^2\left[1+\frac{1}{n}+\frac{(x-\overline{X})^2}{(n-1)S_X^2}\right]}\right)$$

$$x=10 \quad \hat{y}=2{,}3294+2{,}9651\times 10=31{,}98$$

$$1-\alpha=0{,}90 \Rightarrow t_{18;0{,}05}=1{,}734$$

$$IC_Y:\left(31{,}98\pm 1{,}734\times\sqrt{1{,}8936\times\left[1+\frac{1}{20}+\frac{(10-4{,}425)^2}{(20-1)\times 1{,}498^2}\right]}\right)\Rightarrow IC_Y:\left(31{,}98\pm 3{,}18\right)$$

$$IC_Y:\left(28{,}80;35{,}16\right)$$

As soluções para (X,Y_2) e (X,Y_3) são análogas, ficando a cargo do aluno.

APÊNDICE 1

Formulário

Probabilidade

$P(A) = \dfrac{n(A)}{n(S)}$ (espaços amostrais equiprováveis, $n(A)$ e $n(S)$ número de elementos de A e S)

$P(\phi) = 0$ (ϕ é o conjunto vazio)

$P(\overline{A}) = 1 - P(A)$ ($\overline{A}$ é o complementar de A)

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j=2}^{n} P\left(A_i \cap A_j\right) + \sum_{i<j<r=3}^{n} P\left(A_i \cap A_j \cap A_r\right) - \ldots + (-1)^{n-1} P\left(A_1 \cap A_2 \cap \ldots \cap A_n\right)$$ (eventos quaisquer)

$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$ (eventos mutuamente exclusivos)

$P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ (probabilidade condicional)

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\ldots P(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1})$$

(enventos quaisquer)

$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2)P(A_3)\ldots P(A_n)$ (eventos mutuamente independentes)

$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)$ (Teorema da Probabilidade Total)

$P(B_j|A) = \dfrac{P(A|B_j) \cdot P(B_j)}{\sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)}$ (Teorema de Bayes)

Variáveis Aleatórias

$E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i)$ (X discreta)

$E(g(X)) = \sum_{i=1}^{\infty} g(x_i) P(X = x_i)$ (X discreta)

$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$ (X contínua)

$E(g(X)) = \int_{-\infty}^{+\infty} g(x) \cdot f(x) dx$ (X contínua)

$V(X)=E[X-E(X)]^2$ ou $V(X)=E(X^2)-\{E(X)\}^2$ (X qualquer)

$E(X\pm c)=E(X)\pm c$ (c constante e X qualquer)

$E(cX)=cE(X)$ (c constante e X qualquer)

$E(c)=c$ (c constante)

$V(X\pm c)=V(X)$ (c constante e X qualquer)

$V(cX)=c^2V(X)$ (c constante e X qualquer)

$V(c)=0$ (c constante)

$E(X_1+X_2+...+X_n)=E(X_1)+E(X_2)+...+E(X_n)$ (quaisquer X_1, X_2,..., X_n)

$V(X_1+X_2+...+X_n)=V(X_1)+V(X_2)+...+V(X_n)$ (X_1, X_2,..., X_n independentes)

$$f(x)=\begin{cases}\binom{n}{x}p^xq^{n-x} & x=0,1,2,3,\dots,n\\ 0 & x\neq 0,1,2,3,\dots,n\end{cases}$$ (Distribuição Binomial)

$E(X)=np \qquad V(X)=npq$

$$f(x)=\begin{cases}pq^{x-1} & x=1,2,3,\dots\\ 0 & x\neq 1,2,3,\dots\end{cases}$$ (Distribuição Geométrica)

$E(X)=\dfrac{1}{p} \qquad V(X)=\dfrac{q}{p^2}$

$$f(x)=\begin{cases}\binom{x-1}{R-1}p^Rq^{x-R} & , \quad x=R,R+1,R+2,\dots\\ 0 & , \quad x\neq R,R+1,R+2,\dots\end{cases}$$ (Distribuição Binomial Negativa

$E(X)=R\cdot\dfrac{1}{p} \qquad V(X)=R\cdot\dfrac{q}{p^2}$

ou Pascal)

$$f(x)=\begin{cases}\dfrac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}} & x=0,1,2,\dots,n\\ 0 & x\neq 0,1,2,\dots,n\end{cases}$$ (Distribuição

$E(X)=np \qquad V(X)=npq\cdot\dfrac{N-n}{N-1} \qquad p=\dfrac{R}{N} \qquad q=1-p=\dfrac{N-R}{N}$

Hipergeométrica)

$$f(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & , \ x = 0,1,2,\ldots,n,\ldots \\ 0 & , \ x \neq 0,1,2,\ldots,n,\ldots \end{cases}$$

(Distribuição de Poisson)

$$E(X) = \lambda \qquad V(X) = \lambda$$

$$f(x) = \begin{cases} \dfrac{1}{b-a} & x \in [a,b] \\ 0 & x \notin [a,b] \end{cases} \qquad F(x) = \frac{x-a}{b-a}$$

(Distribuição Uniforme)

$$E(X) = \frac{a+b}{2} \qquad V(X) = \frac{(b-a)^2}{12}$$

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & x > 0 \ , \lambda > 0 \\ 0 & x \leq 0 \end{cases}$$

$$F(x) = 1 - e^{-\lambda x}$$

(Distribuição Exponencial)

$$E(X) = \frac{1}{\lambda} \qquad V(X) = \frac{1}{\lambda^2}$$

Intervalos de Confiança

Para estimar a média populacional - μ_X

TAMANHO DA POPULAÇÃO	VARIÂNCIA POPULACIONAL - σ_X^2	
	CONHECIDA	DESCONHECIDA
INFINITA	$\bar{X} \pm z_{\alpha/2} \cdot \dfrac{\sigma_X}{\sqrt{n}}$	$\bar{X} \pm t_{n-1,\alpha/2} \cdot \dfrac{S}{\sqrt{n}}$
FINITA	$\bar{X} \pm z_{\alpha/2} \cdot \dfrac{\sigma_X}{\sqrt{n}} \cdot \sqrt{\dfrac{N-n}{N-1}}$	$\bar{X} \pm t_{n-1,\alpha/2} \cdot \dfrac{S}{\sqrt{n}} \cdot \sqrt{\dfrac{N-n}{N-1}}$

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}} \quad \text{ou} \quad S = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n-1} - \frac{\left(\sum_{i=1}^{n} x_i\right)}{n(n-1)}}$$

Para estimar a proporção populacional - p

POPULAÇÃO INFINITA	$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}\hat{q}}{n}}$
POPULAÇÃO FINITA	$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}\hat{q}}{n} \cdot \dfrac{N-n}{N-1}}$

$\hat{p} = \dfrac{n_A}{n} \quad \hat{q} = 1 - \hat{p}$

Testes de Hipóteses

Para a média populacional - μ_X

TAMANHO DA POPULAÇÃO	ESTATÍSTICA DE PROVA		CRITÉRIO DE DECISÃO	
	σ_X^2 CONHECIDA	σ_X^2 DESCONHECIDA	TESTE BILATERAL	TESTE UNILATERAL
INFINITA	$Z_c = \dfrac{\bar{X} - \mu_0}{\dfrac{\sigma_X}{\sqrt{n}}}$	$T_c = \dfrac{\bar{X} - \mu_0}{\dfrac{S}{\sqrt{n}}}$	$\lvert Z_c \rvert > z_{\alpha/2}$ $\lvert T_c \rvert > t_{n-1,\alpha/2}$ REJEITA-SE H_0	$\lvert Z_c \rvert > z_{\alpha}$ $\lvert T_c \rvert > t_{n-1,\alpha}$ REJEITA-SE H_0
FINITA	$Z_c = \dfrac{\bar{X} - \mu_0}{\dfrac{\sigma_X}{\sqrt{n}} \cdot \sqrt{\dfrac{N-n}{N-1}}}$	$T_c = \dfrac{\bar{X} - \mu_0}{\dfrac{S}{\sqrt{n}} \cdot \sqrt{\dfrac{N-n}{N-1}}}$		

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}} \quad \text{ou} \quad S = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n-1} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}}$$

Para a proporção populacional - p

TAMANHO DA POPULAÇÃO	ESTATÍSTICA DE PROVA	CRITÉRIO DE DECISÃO	
		TESTE BILATERAL	TESTE UNILATERAL
INFINITA	$Z_c = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}}$	$\lvert Z_c \rvert > z_{\alpha/2}$ REJEITA-SE H_0	$\lvert Z_c \rvert > z_{\alpha}$ REJEITA-SE H_0
FINITA	$Z_c = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n} \cdot \dfrac{N-n}{N-1}}}$		

$$\hat{p} = \frac{n_A}{n} \quad \hat{q} = 1 - \hat{p} \qquad q_0 = 1 - p_0$$

Correlação e Regressão

VARIÁVEIS ALEATÓRIAS BIDIOMENSIONAIS

$\mu_X = E(X)$ - Esperança Matemática ou Média de X

$\mu_Y = E(Y)$ - Esperança Matemática ou Média de Y

$\sigma_X^2 = V(X) = E\{[X - E(X)]^2\} = E(X^2) - [E(X)]^2$ - Variância de X

$\sigma_Y^2 = V(Y) = E\{[Y - E(Y)]^2\} = E(Y^2) - [E(Y)]^2$ - Variância de Y

$\sigma_{XY} = COV(XY) = E\{[X - E(X)][Y - E(Y)]\} = E(XY) - E(X)E(Y)$
- Covariância de (X,Y)

$\rho_{XY} = \dfrac{COV(XY)}{\sqrt{V(X)V(Y)}} = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}$ - Coeficiente de Correlação de (X,Y)

$S_X^2 = \dfrac{\sum (x - \overline{X})^2}{n-1} = \dfrac{\sum x^2 - \dfrac{\left(\sum x\right)^2}{n}}{n-1}$ é o estimador da variância de X

$S_Y^2 = \dfrac{\sum (y - \overline{Y})^2}{n-1} = \dfrac{\sum y^2 - \dfrac{\left(\sum y\right)^2}{n}}{n-1}$ é o estimador da variância de Y

$S_{XY} = \dfrac{\sum (x - \overline{X})(y - \overline{Y})}{n-1} = \dfrac{\sum xy - \dfrac{\sum x \sum y}{n}}{n-1}$ é o estimador da covariância de (X,Y)

$r_{XY} = \dfrac{S_{XY}}{\sqrt{S_X^2 S_Y^2}} = \dfrac{S_{XY}}{S_X S_Y} = \dfrac{\sum xy - \dfrac{\sum x \sum y}{n}}{\sqrt{\left\{\left[\sum x^2 - \dfrac{\left(\sum x\right)^2}{n}\right]\left[\sum y^2 - \dfrac{\left(\sum y\right)^2}{n}\right]\right\}}}$ é o estimador do coeficiente de correlação entre X e Y

$$IC_\rho : \left(\frac{e^{c_1}-1}{1+e^{c_1}}, \frac{e^{c_2}-1}{1+e^{c_2}} \right)$$ - intervalo de confiança para ρ_{XY}

$$c_1 = 2(R - z_{\alpha/2} \cdot \sigma_R) \qquad c_2 = 2(R + z_{\alpha/2} \cdot \sigma_R)$$

$$T = \frac{r_{XY}\sqrt{n-2}}{\sqrt{1-r_{XY}^2}}$$ - testes de hipóteses para ρ_{XY}

$$|T_c| \geq t_{n-2,\alpha/2} \Rightarrow \text{rejeita - se } H_0$$

Regressão Linear

$$\hat{y} = b_0 + b_1 x$$

$$b_0 = \bar{y} - b_1 \bar{x}$$ - método dos mínimos quadrados

$$b_1 = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{2}} = \frac{S_{XY}}{S_X^2} = r_{XY} \times \frac{S_Y}{S_X}$$

$$SQM = \sum y^2 - \frac{\left(\sum y\right)^2}{n}$$ - soma de quadrados total (em torno da média)

$$SQR = b_1 \left(\sum xy - \frac{\sum x \sum y}{n} \right)$$ - soma de quadrados devido à regressão

$SQE = SQM - SQR$ - soma de quadrados em torno da regressão (erro)

$$S^2 = \frac{SQE}{n-2}$$ - estimador da variância do erro

$$R^2 = \frac{SQR}{SQM}$$ - coeficiente de determinação

Intervalos de confiança para os parâmetros da regressão

PARÂMETRO	INTERVALO
b_0	$b_0 \pm t_{n-2,\alpha/2} \cdot \sqrt{S^2\left[\frac{1}{n}+\frac{\overline{X}^2}{(n-1)S_X^2}\right]}$
b_1	$b_1 \pm t_{n-1,\alpha/2} \cdot \sqrt{\frac{S^2}{(n-1)S_X^2}}$

$$\overline{X}=\frac{\sum_{i=1}^{n} x_i}{n} \qquad S^2=\frac{SQE}{n-2} \qquad S_X^2=\frac{\sum x^2-\frac{\left(\sum x\right)^2}{n}}{n-1}$$

Testes de hipóteses para os parâmetros da regressão

PARÂMETRO	ESTATÍSTICA DE PROVA	CRITÉRIO DE DECISÃO	
		TESTE BILATERAL	TESTE UNILATERAL
b_0	$T_{0C}=\frac{b_0-\beta_{0,0}}{\sqrt{S^2\left[\frac{1}{n}+\frac{\overline{X}^2}{(n-1)S_X^2}\right]}}$	$\lvert T_{0C}\rvert > t_{n-2;\alpha/2}$ REJEITA-SE H_0	$\lvert T_{0C}\rvert > t_{n-2,\alpha}$ REJEITA-SE H_0
b_1	$T_{1C}=\frac{b_1-\beta_{1,0}}{\sqrt{\frac{S^2}{(n-1)S_X^2}}}$		

$$\overline{X}=\frac{\sum_{i=1}^{n} x_i}{n} \qquad S^2=\frac{SQE}{n-2} \qquad S_X^2=\frac{\sum x^2-\frac{\left(\sum x\right)^2}{n}}{n-1}$$

Intervalos de confiança para as previsões

SITUAÇÃO	INTERVALO
ESTIMAR O VALOR MÉDIO DE Y DADO UM VALOR x	$\hat{y}_k \pm t_{n-2,\alpha/2} \times \sqrt{S^2\left[\frac{1}{n}+\frac{(x_k-\overline{X})^2}{(n-1)S_X^2}\right]}$
ESTIMAR UM VALOR INDIVIDUAL DE Y DADO UM VALOR x	$\hat{y}_k \pm t_{n-2,\alpha/2} \times \sqrt{S^2\left[1+\frac{1}{n}+\frac{(x_k-\overline{X})^2}{(n-1)S_X^2}\right]}$

$$\overline{X}=\frac{\sum_{i=1}^{n} x_i}{n} \qquad S^2=\frac{SQE}{n-2} \qquad S_X^2=\frac{\sum x^2-\frac{\left(\sum x\right)^2}{n}}{n-1}$$

APÊNDICE 2

Conjuntos

Noções primitivas

Na teoria dos conjuntos são consideradas noções primitivas (aceitas sem definição):

- Conjunto
- Elemento
- Pertinência entre elemento e conjunto

A noção matemática de conjunto é a mesma da linguagem comum: agrupamento, classe, coleção, etc.

Relação de pertinência

Se A é um conjunto e x um elemento que pertence a A então escreve-se $x \in A$, caso contrário escreve-se $x \notin A$.

Descrição de um conjunto

Existem duas maneiras para descrever um conjunto e seus elementos:

- Enumeração – enumeram-se, citam-se ou escrevem-se seus elementos: $A = \{2,3,5,7\}$
- Propriedade dos elementos – define-se uma propriedade característica P, de seus elementos: $A = \{x | x \text{ é natural primo menor do que } 11\}$

Conjunto unitário

É aquele que possui apenas um elemento: $A = \{2\}$

Conjunto vazio

É aquele que não possui elementos, denotado por ϕ. Em geral descrito por uma propriedade logicamente falsa: $A = \{x | x \neq x\} = \phi$.

Conjunto universo

Usualmente denotado por U ou Ω é o conjunto no qual se procura a solução, ou soluções, de um problema específico. Ao se descrever um conjunto através de determinada propriedade P, é fundamental definir o conjunto U, que restringirá o espaço de busca pelas soluções: $A=\{x \in U \mid x \text{ tem a propriedad e } P\}$

Conjuntos iguais

Os conjuntos A e B são iguais se todos os elementos de A pertencem a B e, reciprocamente, todo elemento de B pertence a A.

$$A=B \Leftrightarrow (\forall x)(x \in A \Leftrightarrow x \in B)$$

Subconjunto

Um conjunto A é subconjunto do conjunto B e, e somente se, todo elemento de A também pertence a B.

$$A \subset B \Leftrightarrow (\forall x)(x \in A \Rightarrow x \in B)$$

- O símbolo $\subset$ é denominado "sinal de inclusão"
- A notação $A \subset B$ indica que "A é subconjunto de B", "A está contido em B" ou "A é parte de B"
- A notação $A \not\subset B$ indica que "A não está contido em B"
- Também usa-se $B \supset A$ e lê-se "B contém A"

Se A, B e C são conjuntos quaisquer, valem as seguintes propriedades da inclusão:

- $\phi \subset A$
- $A \subset A$ (reflexiva)
- $A \subset B \wedge B \subset A \Rightarrow A=B$ (anti-simétrica)
- $A \subset B \wedge B \subset C \Rightarrow A \subset C$ (transitiva)

Conjunto das partes

O conjunto das partes de A é aquele formado por todos os subconjuntos de A, e denota-se por $\wp(A)$

- $\wp(A)$ é um conjunto cujos elementos também são conjuntos, neste caso, se B é subconjunto de A, é correta a notação $B \in \wp(A)$
- Se A tem n elementos, $\wp(A)$ tem 2^n elementos, isto é, a quantidade de subconjuntos de A é 2^n, incluindo o conjunto vazio

União (reunião)

A união de dois conjuntos *A* e *B*, denotada por $A \bigcup B$ é o conjunto formado pelos elementos que pertencem a *A* ou a *B*, isto é, pertencem a pelo menos um deles.

$$A \bigcup B = \{x | x \in A \vee x \in B\}$$

Se A, B e C são conjuntos quaisquer, valem as seguintes propriedades da união:

- $A \bigcup A = A$ (idempotente)
- $A \bigcup \phi = A$ (elemento neutro)
- $A \bigcup B = B \bigcup A$ (comutativa)
- $(A \bigcup B) \bigcup C = A \bigcup (B \bigcup C)$ (associativa)

Interseção

A interseção de dois conjuntos A e B, denotada por $A \bigcap B$ é o conjunto formado pelos elementos que pertencem a A e a B.

$$A \bigcap B = \{x | x \in A \wedge x \in B\}$$

- Se $A \bigcap B = \phi$ eles são ditos disjuntos, ou exclusivos
- Se para n conjuntos, $A_1, A_2, \ldots, A_n$, tem-se que eles são dois a dois disjuntos, ou seja, $(\forall i \neq j)(A_i \bigcap A_j = \phi)$, eles são ditos mutuamente exclusivos

Se A, B e C são conjuntos quaisquer, valem as seguintes propriedades da interseção:

- $A \bigcap A = A$ (idempotente)
- $A \bigcap U = A$ (elemento neutro)
- $A \bigcap B = B \bigcap A$ (comutativa)
- $(A \bigcap B) \bigcap C = A \bigcap (B \bigcap C)$ (associativa)

Propriedades envolvendo união e interseção

Se A, B e C são conjuntos quaisquer, valem as seguintes propriedades gerais:

- $A \cup (A \cap B) = A$
- $A \cap (A \cup B) = A$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (distributiva da união em relação à interseção)
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (distributiva da interseção em relação à união)

Diferença

A diferença entre dois conjuntos A e B, denotada por A-B é dada pelo conjunto formado pelos elementos de A que não pertencem a B.

$$A - B = \{x | x \in A \wedge x \notin B\}$$

Diferença simétrica

A diferença simétrica entre dois conjuntos A e B, denotada por $A \Delta B$ é dada por:

$$A \Delta B = (A - B) \cup (B - A)$$

Se A e B são conjuntos quaisquer, valem as seguintes propriedades para a diferença simétrica:

- $A \Delta \phi = A$
- $A \Delta A = \phi$
- $A \Delta B = B \Delta A$

Complementar

Se A e B são conjuntos com $A \subset B$, o complementar de A em relação a B, denotado por C_B^A, é o conjunto formado pelos elementos de B que não pertencem a A, ou seja:

$$C_B^A = B - A$$

- O complementar de um conjunto A em relação ao conjunto universo U pode ser denotado simplesmente por A^c ou $\overline{A}$

Se A e B são subconjuntos de um conjunto universo U, valem as seguintes propriedades:

- $\overline{A} \cap A = \phi$ e $\overline{A} \cup A = U$
- $\overline{U} = \phi$ e $\overline{\phi} = U$
- $\overline{\overline{A}} = A$
- $\overline{A \cap B} = \overline{A} \cup \overline{B}$
- $\overline{A \cup B} = \overline{A} \cap \overline{B}$

Veja que essas propriedades valem para A e B em relação a qualquer outro conjunto que não seja o universo.

Partição

Dados os conjuntos $\Omega, A_1, A_2, \ldots, A_n$, então $A_1, A_2, \ldots, A_n$ formam uma partição de Ω se:

- $A_i \neq \phi, i = 1, 2, \ldots, n$
- $\bigcup_{i=1}^{n} A_i = \Omega$
- $A_i \cap A_j = \phi, \forall i \neq j$

Observe que $A_1, A_2, \ldots, A_n$ são subconjuntos de A

Partição ordenada

Chama-se de partição ORDENADA de Ω a SEQUENCIA de conjuntos $(A_1, A_2, \ldots, A_n)$

Aqui, $(A_1, A_2, A_3 \ldots, A_n)$ e $(A_2, A_1, A_3 \ldots, A_n)$, por exemplo, são partições diferentes.

Partição não ordenada

Chama-se de partição NÃO ORDENADA de Ω a FAMÍLIA de conjuntos $\{A_1, A_2, \ldots, A_n\}$

Aqui, $\{A_1, A_2, A_3 \ldots, A_n\}$ e $\{A_2, A_1, A_3 \ldots, A_n\}$, por exemplo, são partições iguais.

Princípio da inclusão-exclusão

A quantidade de elementos de um conjunto A será denotada por $\#A$ ou $n(A)$.
Dados dois conjuntos, A e B, a quantidade de elementos da união dos mesmos é dada por:

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

Para três conjuntos, tem-se:

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

Denotando os conjuntos por A_1, A_2, A_3 podemos escrever:

$$n(A_1 \cup A_2 \cup A_3) = \sum_{i=1}^{3} n(A_i) - \sum_{i<j=2}^{3} (A_i \cap A_j) + n(A_1 \cap A_2 \cap A_3)$$

Por indução finita prova-se que:

$$n\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} n(A_i) - \sum_{i<j=2}^{n} n\left(A_i \cap A_j\right) + \sum_{i<j<r=3}^{n} n\left(A_i \cap A_j \cap A_r\right) - \ldots + \\ + (-1)^{n-1} n\left(A_1 \cap A_2 \cap \ldots \cap A_n\right)$$

APÊNDICE 3

Técnicas de Contagem

Permutações

Quantidade de maneiras que se pode ordenar um conjunto de objetos.

Permutações simples

Quantidade de maneiras que se pode ordenar *n* objetos distintos: $a_1, a_2, ..., a_n$

$$P_n = n!$$

EXEMPLO

Determinar a quantidade de todas as ordenações possíveis dos números 1,2 e *3*.
1,2,3 ; 1,3,2 ; 2,1,3 ; 2,3,1 ; 3,1,2 ; 3,2,1

$$P_3 = 3! = 1 \times 2 \times 3 = 6$$

Permutações caóticas

Quantidade de permutações dos objetos $(a_1, a_2, ..., a_n)$ nas quais nenhum deles ocupa sua posição original.

$$D_n = n!\sum_{i=0}^{n}\frac{(-1)^i}{i!} \quad \text{ou} \quad D_n = \text{inteiro mais próximo de } \frac{n!}{e}$$

EXEMPLO

Determinar a quantidade de permutações caóticas dos números 1,2 e 3.

2,3,1 ;e 3,1,2

$$D_3 = 3!\sum_{i=0}^{3}\frac{(-1)^i}{i!} = 6\times\left[\frac{(-1)^0}{0!}+\frac{(-1)^1}{1!}+\frac{(-1)^2}{2!}+\frac{(-1)^3}{3!}\right] = 6\times\left(\frac{1}{1}-\frac{1}{1}+\frac{1}{2}-\frac{1}{6}\right) = 6\times\frac{2}{6} = 2$$

Ou $\quad \frac{3!}{e} = 2{,}2\ldots \Rightarrow D_3 = 2$

Permutações de elementos repetidos (anagramas)

Quantidade de maneiras que se pode ordenar *n* elementos, nem todos distintos:

$$\underbrace{a_1 a_1,\ldots a_1}_{n_1},\underbrace{a_2,a_2,\ldots a_2}_{n_2},\ldots,\underbrace{a_k,a_k,\ldots a_k}_{n_k} \qquad n_1+n_2+\ldots+n_k=n$$

$$P_n^{n_1,n_2,\ldots,n_k}=\frac{n!}{n_1!n_2!\ldots n_k!}$$

EXEMPLO

Determinar a quantidade de anagramas que podem ser formados com a palavra ANA

ANA , AAN , NAA

$$P_3^{2,1}=\frac{3!}{2!1!}=\frac{1\times 2\times 3}{1\times 2\times 1}=3$$

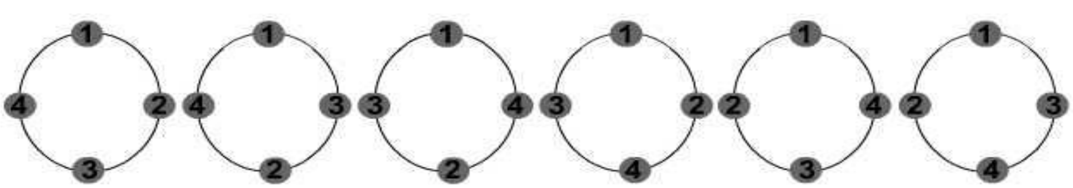
Permutações circulares

Quantidade de maneiras que se pode colocar n objetos distintos, $a_1,a_2,\ldots,a_n$, em n lugares (equiespaçados) em torno de um círculo, considerando-se equivalentes as disposições que possam coincidir por rotação.

$$PC_n=(n-1)!$$

EXEMPLO

Determinar a quantidade de maneiras que se pode dispor 4 pessoas em uma mesa redonda.



$$PC_4=3!=6$$

Arranjos

Quantidade de subgrupos ORDENADOS, com k elementos, que podem ser formados a partir de um grupo com n elementos distintos.

Arranjos simples

Quantidade de subgrupos ORDENADOS, com k elementos, não repetidos, que podem ser formados a partir de um grupo com n elementos distintos.

$$A_{n,k} = \frac{n!}{(n-k)!}$$

EXEMPLO

Dado o conjunto $\{a_1, a_2, a_3\}$, os subgrupos ordenados com 2 elementos não repetidos que podem ser formados são:

$$S = \{[a_1, a_2], [a_2, a_1], [a_1, a_3], [a_3, a_1], [a_2, a_3], [a_3, a_2]\} \qquad n(S) = A_{3,2} = \frac{3!}{(3-2)!} = 6$$

Arranjos com repetição (completos)

Quantidade de subgrupos ORDENADOS, com k elementos, podendo ser repetidos, que podem ser formados a partir de um grupo com n elementos distintos.

$$A^*_{n,k} = n^k$$

EXEMPLO

Dado o conjunto $\{a_1, a_2, a_3\}$, os subgrupos ordenados com 2 elementos, com repetição, que podem ser formados são:

$$S = \{[a_1, a_1], [a_1, a_2], [a_1, a_3], [a_2, a_1], [a_2, a_2], [a_2, a_3], [a_3, a_1], [a_3, a_2], [a_3, a_3]\}$$
$$n(S) = A^*_{3,2} = 3^2 = 9$$

Combinações

Quantidade de subgrupos NÃO ORDENADOS, com k elementos, que podem ser formados a partir de um grupo com n elementos distintos.

Combinações simples

Quantidade de subgrupos NÃO ORDENADOS, com k elementos, não repetidos, que podem ser formados a partir de um grupo com n elementos distintos.

$$C_{n,k} = \frac{n!}{k!(n-k)!}$$

EXEMPLO

Dado o conjunto $\{a_1, a_2, a_3\}$, os subgrupos não ordenados com 2 elementos não repetidos que podem ser formados são:

$$S = \{[a_1, a_2], [a_1, a_3], [a_2, a_3]\} \qquad n(S) = C_{3,2} = \frac{3!}{2!(3-2)!} = 3$$

Combinações com repetição (completas)

Quantidade de subgrupos ORDENADOS, com k elementos, podendo ser repetidos, que podem ser formados a partir de um grupo com n elementos distintos.

$$C^*_{n,k} = C_{n+k-1,k}$$

EXEMPLO

Dado o conjunto $\{a_1, a_2, a_3\}$, os subgrupos não ordenados com 2 elementos, com repetição, que podem ser formados são:

$$S = \{[a_1, a_1], [a_1, a_2], [a_1, a_3], [a_2, a_2], [a_2, a_3], [a_3, a_3]\}$$

$$n(S) = C^*_{3,2} = C_{3+2-1,2} = C_{4,2} = \frac{4!}{2!(4-2)!} = 6$$

Retirada de bolas de uma urna

Para contarmos de quantas maneiras podemos retirar uma amostra de n bolas de uma urna contendo N bolas distinguíveis temos as situações a seguir. Denotaremos esta quantidade por $n(A)$.

OBS: Amostra ordenada significa que a ordem de retirada das bolas é levada em consideração, por exemplo, o resultado (b_1,b_2) é diferente do resultado (b_2,b_1) devendo, portanto, os dois serem contados. O mesmo não acontece na amostra não ordenada, onde apenas um deles é contado. Na amostra com reposição as bolas retiradas são recolocadas na urna, podendo ser retiradas novamente.

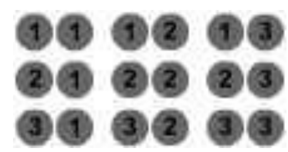
Amostra ordenada com reposição

$$n(A) = A^{*}_{N,n}$$

EXEMPLO

Retirar uma amostra de 2 bolas, ORDENADA COM REPOSIÇÃO, de uma urna contendo 3 bolas distinguíveis.

$$n(A) = A^{*}_{3,2} = 3^2 = 9$$



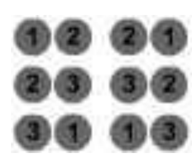
Amostra ordenada sem reposição

$$n(A) = A_{N,n}$$

EXEMPLO

Retirar uma amostra de 2 bolas, ORDENADA SEM REPOSIÇÃO, de uma urna contendo 3 bolas distinguíveis.

$$n(A) = A_{3,2} = \frac{3!}{(3-2)!} = 6$$



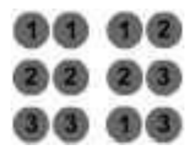
Amostra não ordenada com reposição

$$n(A) = C^*_{N,n}$$

EXEMPLO

Retirar uma amostra de 2 bolas, NÃO ORDENADA COM REPOSIÇÃO, de uma urna contendo 3 bolas distinguíveis.

$$n(A) = C^*_{3,2} = C_{3+2-1,2} = C_{4,2} = \frac{4!}{2!(4-2)!} = 6$$



Amostra não ordenada sem reposição

$$n(A) = C_{N,n}$$

EXEMPLO

Retirar uma amostra de 2 bolas, NÃO ORDENADA SEM REPOSIÇÃO, de uma urna contendo 3 bolas distinguíveis.

$$n(A) = C_{3,2} = \frac{3!}{2!(2-2)!} = 3$$

Distribuição de bolas em urnas

Para contarmos de quantas maneiras podemos distribuir n bolas em N urnas numeradas temos que considerar as seguintes situações. Denotaremos esta quantidade por $n(B)$.

OBS: Distribuição sem exclusão significa que uma urna pode não receber bolas ou receber uma ou mais bolas. Distribuição com exclusão significa que uma urna pode não receber bolas ou receber apenas uma bola.

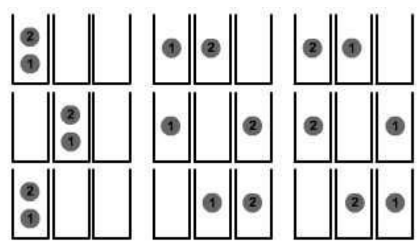
Bolas distinguíveis sem exclusão

$$n(B) = A^{*}_{N,n}$$

EXEMPLO

Distribuir 2 bolas DISTINGUÍVEIS em 3 urnas, SEM EXCLUSÃO.

$$n(B) = A^{*}_{3,2} = 3^2 = 9$$



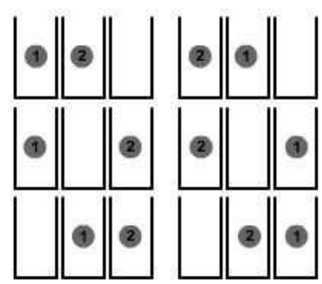
Bolas distinguíveis com exclusão

$$n(B) = A_{N,n}$$

EXEMPLO

Distribuir 2 bolas DISTINGUÍVEIS em 3 urnas, COM EXCLUSÃO.

$$n(B) = A_{3,2} = \frac{3!}{(3-2)!} = 6$$



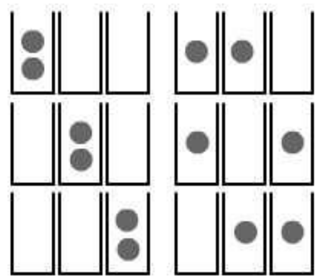
Bolas não distinguíveis sem exclusão

$$n(B) = C^{*}_{N,n}$$

EXEMPLO

Distribuir 2 bolas NÃO DISTINGUÍVEIS em 3 urnas, SEM EXCLUSÃO.

$$n(B) = C^*_{3,2} = C_{3+2-1,2} = C_{4,2} = \frac{4!}{2!(4-2)!} = 6$$



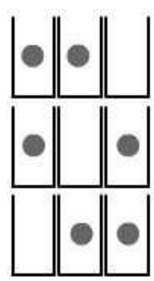
Bolas não distinguíveis com exclusão

$$n(B) = C_{N,n}$$

EXEMPLO

Distribuir 2 bolas NÃO DISTINGUÍVEIS em 3 urnas, COM EXCLUSÃO.

$$n(B) = C_{3,2} = \frac{3!}{2!(3-2)!} = 3$$



Princípio da inclusão-exclusão

Dados os conjuntos $S, A_1, A_2, ..., A_n$, sendo $A_1, A_2, ..., A_n$ subconjuntos de S e as seguintes somas:

$$S_0 = n(S)$$

$$S_1 = \sum_{i=1}^{n} n(A_i)$$

$$S_2 = \sum_{1 \le i < j \le n}^{n} n(A_i \cap A_j)$$

$$S_3 = \sum_{1 \le i < j < k \le n}^{n} n(A_i \cap A_j \cap A_k)$$

...

$$S_n = n(A_1 \cap A_2 \cap ... \cap A_n)$$

Observe que cada soma, a partir de S_1, tem $C_{n,k}, k = 1,2,...,n$ parcelas.
Então, tem-se:

a) A quantidade de elementos de S que pertencem a exatamente p ($p \le n$) dos subconjuntos $A_1, A_2, ..., A_n$ é:

$$a_p = \sum_{k=0}^{n-p} (-1)^k C_{p+k,k} S_{p+k}$$

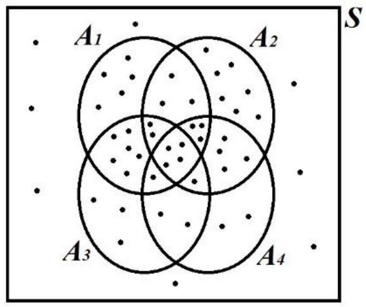
b) A quantidade de elementos de S que pertencem a pelo menos p ($p \le n$) dos subconjuntos $A_1, A_2, ..., A_n$ é:

$$b_p = \sum_{k=0}^{n-p} (-1)^k C_{p+k-1,k} S_{p+k}$$

c) A quantidade de elementos de S que pertencem a $A_1 \cup A_2 \cup .... \cup A_n$ é :

$$n(A_1 \cup A_2 \cup .... \cup A_n) = S_1 - S_2 + S_3 + ... + (-1)^{n-1} S_n$$

EXEMPLO



$$n = 4; n(S) = 50$$

$$n(A_1) = 24 \quad n(A_2) = 24 \quad n(A_3) = 19 \quad n(A_4) = 17$$

$$n(A_1 \bigcap A_2) = 12 \quad n(A_1 \bigcap A_3) = 13 \quad n(A_1 \bigcap A_4) = 8 \quad n(A_2 \bigcap A_3) = 7;$$

$$n(A_2 \bigcap A_4) = 12; n(A_3 \bigcap A_4) = 8$$

$$n(A_1 \bigcap A_2 \bigcap A_3) = 6 \quad n(A_1 \bigcap A_2 \bigcap A_4) = 7 \quad n(A_1 \bigcap A_3 \bigcap A_4) = 5$$

$$n(A_2 \bigcap A_3 \bigcap A_4) = 5$$

$$n(A_1 \bigcap A_2 \bigcap A_3 \bigcap A_4) = 4$$

$$S_0 = n(S) = 42$$

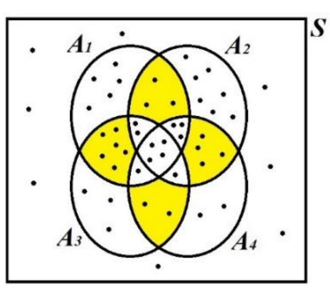
$$S_1 = \sum_{i=1}^{4} n(A_i) = 24 + 24 + 19 + 17 = 84$$

$$S_2 = \sum_{1 \le i < j \le 4}^{4} n(A_i \cap A_j) = 12 + 13 + 8 + 7 + 12 + 8 = 60$$

$$S_3 = \sum_{1 \le i < j < k \le 4}^{4} n(A_i \cap A_j \cap A_k) = 6 + 7 + 5 + 5 = 23$$

$$S_4 = n(A_1 \cap A_2 \cap .A_3 \bigcap A_4) = 4$$

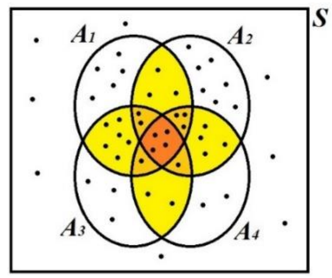
Quantidade de elementos que pertencem a exatamente 2 conjuntos:



$n = 4; p = 2$

$$a_2 = \sum_{k=0}^{4-2} (-1)^k C_{2+k,k} S_{2+k} = \sum_{k=0}^{2} (-1)^k C_{2+k,k} S_{2+k} = (-1)^0 C_{2+0,0} S_{2+0} + (-1)^1 C_{2+1,1} S_{2+1} + +$$
$$(-1)^2 C_{2+2,2} S_{2+2} = C_{2,0} S_2 - C_{3,1} S_3 + C_{4,2} S_4 = 1 \times 60 - 3 \times 23 + 6 \times 4 = 60 - 69 + 24 = 15$$

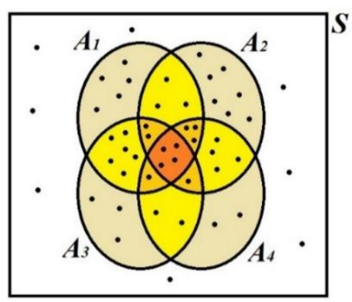
Quantidade de elementos que pertencem a pelo menos 2 conjuntos:



$n = 4; p = 2$

$$b_2 = \sum_{k=0}^{4-2} (-1)^k C_{2+k-1,k} S_{2+k} = \sum_{k=0}^{2} (-1)^k C_{1+k,k} S_{2+k} = (-1)^0 C_{1+0,0} S_{2+0} + (-1)^1 C_{1+1,1} S_{2+1} +$$
$$+ (-1)^2 C_{1+2,2} S_{2+2} = C_{1,0} S_2 - C_{2,1} S_3 + C_{3,2} S_4 = 1 \times 60 - 2 \times 23 + 3 \times 4 = 60 - 46 + 12 = 26$$

Quantidade de elementos que pertencem a pelo menos 1 conjunto (união dos conjuntos):



$n = 4$

$$n(A_1 \cup A_2 \cup A_3 \cup A_4) = S_1 - S_2 + S_3 - S_4 = 84 - 60 + 23 - 4 = 43$$

ANEXO 1

Distribuição $N(0,1)$
Valores de $\Phi(z)=P(Z \leq z)$ para $z \geq 0$

z	0	1	2	3	4	5	6	7	8	9
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

ANEXO 2

Distribuição *t-Student*

α					
UNICAUDAL	0,10	0,05	0,025	0,02	0,01
BICAUDAL	0,20	0,10	0,05	0,04	0,02
GL	VALORES DE *t*				
1	3,078	6,314	12,706	15,895	31,821
3	1,638	2,353	3,182	3,482	4,541
4	1,533	2,132	2,776	2,999	3,747
5	1,476	2,015	2,571	2,757	3,365
6	1,440	1,943	2,447	2,612	3,143
7	1,415	1,895	2,365	2,517	2,998
8	1,397	1,860	2,306	2,449	2,896
9	1,383	1,833	2,262	2,398	2,821
10	1,372	1,812	2,228	2,359	2,764
11	1,363	1,796	2,201	2,328	2,718
12	1,356	1,782	2,179	2,303	2,681
13	1,350	1,771	2,160	2,282	2,650
14	1,345	1,761	2,145	2,264	2,624
15	1,341	1,753	2,131	2,249	2,602
16	1,337	1,746	2,120	2,235	2,583
17	1,333	1,740	2,110	2,224	2,567
18	1,330	1,734	2,101	2,214	2,552
19	1,328	1,729	2,093	2,205	2,539
20	1,325	1,725	2,086	2,197	2,528
21	1,323	1,721	2,080	2,189	2,518
22	1,321	1,717	2,074	2,183	2,508
23	1,319	1,714	2,069	2,177	2,500
24	1,318	1,711	2,064	2,172	2,492
25	1,316	1,708	2,060	2,167	2,485
26	1,315	1,706	2,056	2,162	2,479
27	1,314	1,703	2,052	2,158	2,473
28	1,313	1,701	2,048	2,154	2,467
29	1,311	1,699	2,045	2,150	2,462
30	1,310	1,697	2,042	2,147	2,457
31	1,309	1,696	2,040	2,144	2,453
32	1,309	1,694	2,037	2,141	2,449
33	1,308	1,692	2,035	2,138	2,445
34	1,307	1,691	2,032	2,136	2,441
35	1,306	1,690	2,030	2,133	2,438
36	1,306	1,688	2,028	2,131	2,434
37	1,305	1,687	2,026	2,129	2,431

α					
UNICAUDAL	0,10	0,05	0,025	0,02	0,01
BICAUDAL	0,20	0,10	0,05	0,04	0,02
GL	VALORES DE *t*				
38	1,304	1,686	2,024	2,127	2,429
39	1,304	1,685	2,023	2,125	2,426
40	1,303	1,684	2,021	2,123	2,423
41	1,303	1,683	2,020	2,121	2,421
42	1,302	1,682	2,018	2,120	2,418
43	1,302	1,681	2,017	2,118	2,416
44	1,301	1,680	2,015	2,116	2,414
45	1,301	1,679	2,014	2,115	2,412
46	1,300	1,679	2,013	2,114	2,410
47	1,300	1,678	2,012	2,112	2,408
48	1,299	1,677	2,011	2,111	2,407
49	1,299	1,677	2,010	2,110	2,405
50	1,299	1,676	2,009	2,109	2,403
51	1,298	1,675	2,008	2,108	2,402
52	1,298	1,675	2,007	2,107	2,400
53	1,298	1,674	2,006	2,106	2,399
54	1,297	1,674	2,005	2,105	2,397
55	1,297	1,673	2,004	2,104	2,396
56	1,297	1,673	2,003	2,103	2,395
57	1,297	1,672	2,002	2,102	2,394
58	1,296	1,672	2,002	2,101	2,392
59	1,296	1,671	2,001	2,100	2,391
60	1,296	1,671	2,000	2,099	2,390
61	1,296	1,670	2,000	2,099	2,389
62	1,295	1,670	1,999	2,098	2,388
63	1,295	1,669	1,998	2,097	2,387
64	1,295	1,669	1,998	2,096	2,386
65	1,295	1,669	1,997	2,096	2,385
66	1,295	1,668	1,997	2,095	2,384
67	1,294	1,668	1,996	2,095	2,383
68	1,294	1,668	1,995	2,094	2,382
69	1,294	1,667	1,995	2,093	2,382
70	1,294	1,667	1,994	2,093	2,381
71	1,294	1,667	1,994	2,092	2,380
72	1,293	1,666	1,993	2,092	2,379
∞	1,282	1,645	1,960	2,236	2,576

Sobre o Autor

Francisco de Assis Amaral Bastos: Bacharel em Estatística pela Universidade Federal do Ceará (UFC), com título de Especialista em Métodos Quantitativos, pela Universidade de Fortaleza (UNIFOR) e Mestre em Computação Aplicada, na área de Sistemas de Apoio à Decisão, pela Universidade Estadual do Ceará (UECE). Atuou na Telecomunicações do Ceará S.A. (TELECEARÁ), no cargo de Estatístico, onde exerceu as funções de chefe da Divisão de Estudos de Demanda, Divisão de Estudos de Mercado e Divisão de Estatística. Professor aposentado da Universidade Estadual do Ceará lotado no Centro de Ciências e Tecnologia, Curso de Bacharelado em Ciência da Computação, onde, além das atividades acadêmicas nas áreas de estatística e probabilidade, exerceu as funções de Subchefe do Departamento de Estatística e Computação, vice Coordenador do Curso de Bacharelado em Ciência da Computação e Coordenador do curso de Licenciatura em Computação do programa Universidade Aberta do Brasil (UAB).

Bibliografia

- **Probabilidade - Aplicações à Estatística**, Paul L. Mayer, Ao Livro Técnico S.A.
- **Estatística Aplicada e Probabilidade para Engenheiros**, Douglas C. Montgomery, George S. Runger, LTC
- **Probabilidade**, Teresinha de Maria B. Sampaio Xavier, Airton Fontenele Sampaio Xavier, Coleção Universitária de Problemas, LTC
- **Análise Combinatória e Probabilidade**, Augusto Cesar de Oliveira Morgado et all, Coleção do Professor de Matemática, Sociedade Brasileira de Matemática
- **Curso de Estatística**, Jairo Simon da Fonseca, Gilberto de Andrade Martins, Editora Atlas
- **Estatística Básica**, Wilton O. Bussab, Pedro A. Morettin, Editora Atual
- **Estatística Básica Volume I** – Probabilidade, Luiz Gonzaga Morettin, Markron Books
- **Estatística Básica Volume II** – Inferência, Luiz Gonzaga Morettin, Markron Books
- **Estatística Elementar**, Paul G. Hoel, Editora Fundo de Cultura
- **Estatística Aplicada à Administração**, William J. Stevenson, Haraper & Row do Brasil