Introdução
Estatística descritiva

Autor do documento original: Prof. Mário Escudeiro

Introdução

Objecto da Estatística

Recolha, compilação, análise e interpretação da informação

Recolha da informação	$\longrightarrow$	**Amostragem**
Descrição, classificação e apresentação da informação	$\longrightarrow$	**Estatística Descritiva**
Interpretação da informação	$\longrightarrow$	**Inferência Estatística**

Amostragem

O processo de amostragem deve ser objectivo e não tendencioso pelo que o melhor critério é o da amostragem aleatória. Dessa forma, garante-se que todos os elementos da população têm igual hipótese de ser integrados na amostra.

Estatística Descritiva

***Síntese e representação* de forma compreensível da informação contida num conjunto de dados *(construção de tabelas, gráficos ou cálculo de medidas centrais, de dispersão ou outras)*.**

Inferência Estatística

A partir de um conjunto limitado de dados (*amostra*), pretende-se caracterizar o todo a partir dos quais os dados foram obtidos (*população*).

População

Caracteriza-se pelo grupo inteiro de objectos (unidades) dos quais se pretende obter a informação.

Unidade

Qualquer membro individual da população.

Amostra

Uma parte ou subconjunto da população usada para obter informação acerca do todo.

Variável

Uma característica de uma unidade, que será medida a partir da unidade da amostra.

DGI
2019

Escalas de Representação

dados qualitativos

escala nominal	Dados classificados por categorias não ordenadas *ex. género, cor do cabelo, cor dos olhos*
escala ordinal	Dados classificados por categorias ordenadas *ex. notas num teste: mau, medíocre, suf, bom e muito bom*
escala de intervalo	Dados expressos numa escala numérica com origem arbitrária *ex. intervalo de temperatura*
escala absoluta	Dados expressos numa escala numérica com origem fixa *ex. peso expresso em kg*

dados quantitativos

Tipo de dados

Qualitativos → *Discretos*

Quantitativos → *Discretos*, *Contínuos*

DGI
2019

1.1. Dados Qualitativos

1.1.1. *Representação tabular/gráfica dos dados*

A descrição das amostras faz-se com recurso a ***tabelas de frequência, a diagramas de barras e a diagramas circulares***.

Cálculo de Frequências

F_k	*frequência absoluta da categoria* k
$F'_k = \sum_{k=1}^{K} F_k$	*frequência absoluta acumulada*
$f_k = \frac{F_k}{n}$	*frequência relativa da categoria* k
$f'_k = \sum_{k=1}^{K} f_k = \frac{F'_k}{n}$	*frequência relativa acumulada*

Exemplo

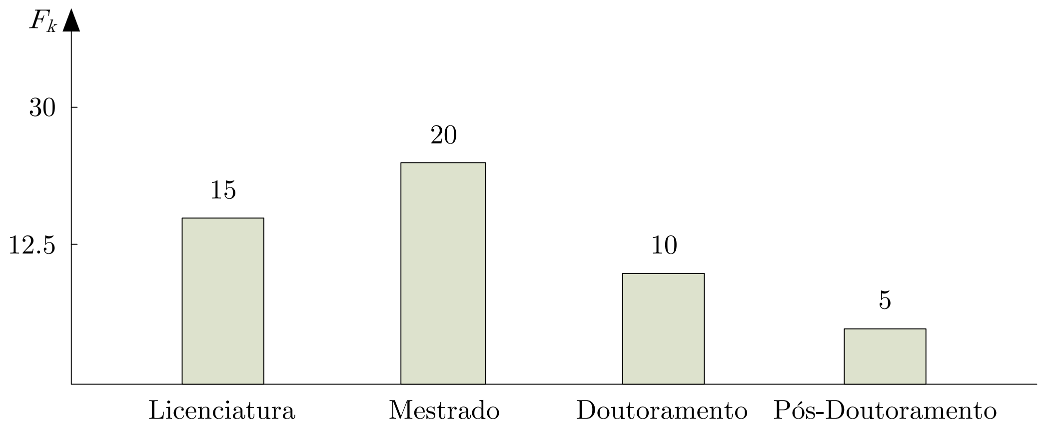
Numa amostra constituída por 50 professores de uma instituição de Ensino Superior constatou-se que 15 tinham como habilitação o grau de Licenciado, 20 o grau de Mestre, 10 o grau de Doutor e 5 encontravam-se em trabalhos de Pós-Doutoramento.

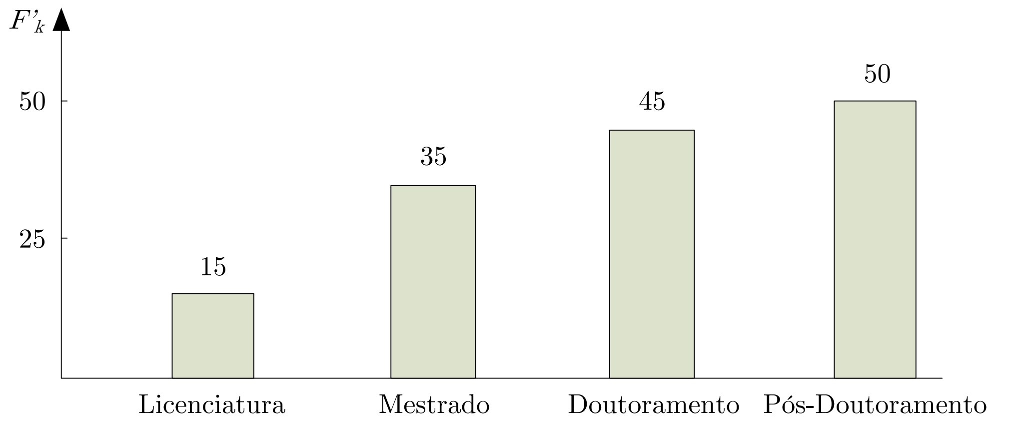
- A *variável grau académico* é uma variável *qualitativa* e pode ser representada numa *escala ordinal*.

i) *Tabela de frequências*

Grau académico	F_k	F'_k	f_k	f'_k
Licenciatura	15	15	0.3	0.3
Mestrado	20	35	0.4	0.7
Doutoramento	10	45	0.2	0.9
Pós-Doutoramento	5	50	0.1	1.0

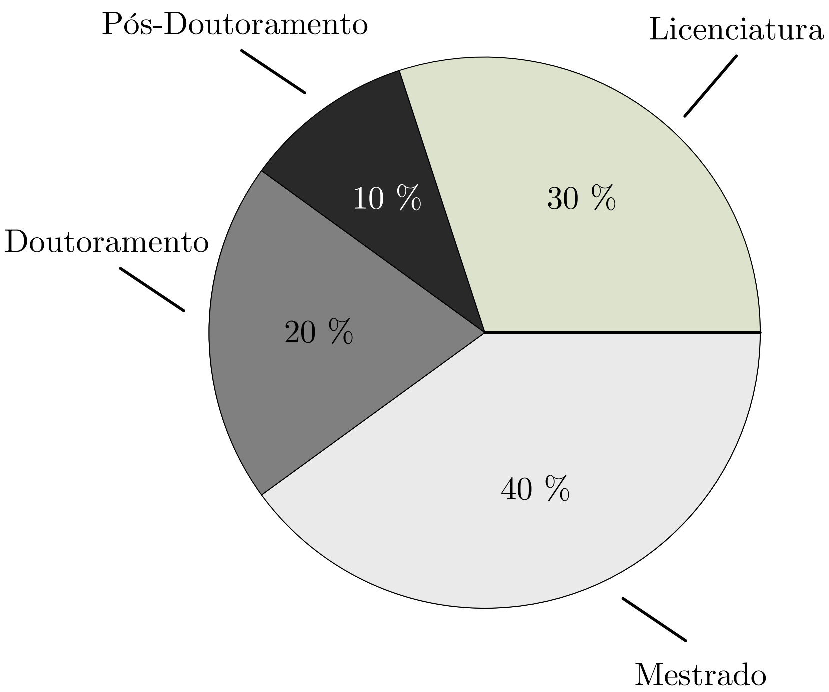
ii) *Diagramas de barras*





DGI
2019

iii) *Diagrama circular*



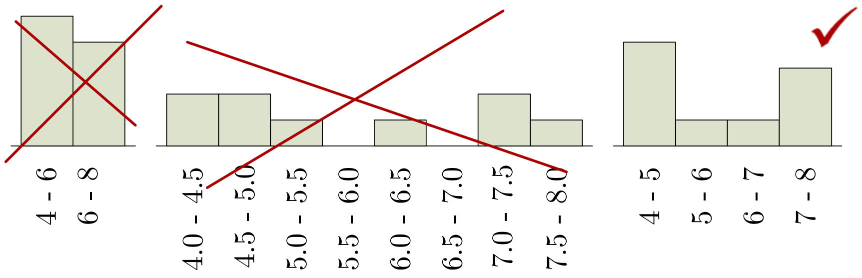
1.2. Dados Quantitativos

1.2.1. Representação tabular/gráfica dos dados

Construção de Histogramas

Denomina-se de histograma a representação gráfica dos dados em que se marcam as classes no eixo horizontal, as frequências no eixo vertical e se usam barras de área proporcional à frequência da classe correspondente.

- ***O número de intervalos ou classes* considerado *não deve ser demasiado pequeno* para não esconder a variabilidade, *nem demasiado grande*, para se poder evidenciar a regularidade.**



Regra prática* - *nº classes $\approx \sqrt{n}$ *(classes de igual amplitude)*

DGI
2019

Exemplo

Considere as classificações obtidas por 49 alunos num exame de Estatística:

9.2	5.2	10.8	10.1
8.3	13.2	9.1	13.5
9.4	12.5	8.4	11.1
8.5	9	11.3	11.9
6.2	8.1	9.5	13.3
14	5.7	7.7	6.8
13.8	6.7	8.3	12.9
9.2	7.4	9.4	6.7
9.9	7.8	$\boxed{4.3}$	10
9.8	12.3	9.8	8.5
10.3	8.1	6.5	
$\boxed{15}$	7.1	11.7	
12.9	13.6	11.2	

máximo

mínimo

Construa a tabela de frequências, o histograma e o respectivo polígono de frequências.

Classe		F_i	f_i
[0, 1 [		0	0
[1, 2 [		0	0
[2, 3 [		0	0
[3, 4 [		0	0
[4, 5 [	|	1	0.02
[5, 6 [	||	2	0.041
[6, 7 [	|||||	5	0.102
[7, 8 [	||||	4	0.082
[8, 9 [	|||||||	7	0.143
[9, 10 [	||||||||||	10	0.204
[10, 11 [	||||	4	0.082
[11, 12 [	|||||	5	0.102
[12, 13 [	||||	4	0.082
[13, 14 [	|||||	5	0.102
[14, 15 [	|	1	0.02
[15, 16 [	|	1	0.02
[16, 17 [		0	0
[17, 18 [		0	0
[18, 19 [		0	0
[19, 20]		0	0
	$\sum$	49	1

Definição das classes do histograma

$$\min\{x_1, x_2, \ldots, x_{49}\} = 4.3$$
$$\max\{x_1, x_2, \ldots, x_{49}\} = 15.0 \longrightarrow A = 15 - 4.3 = 10.7$$

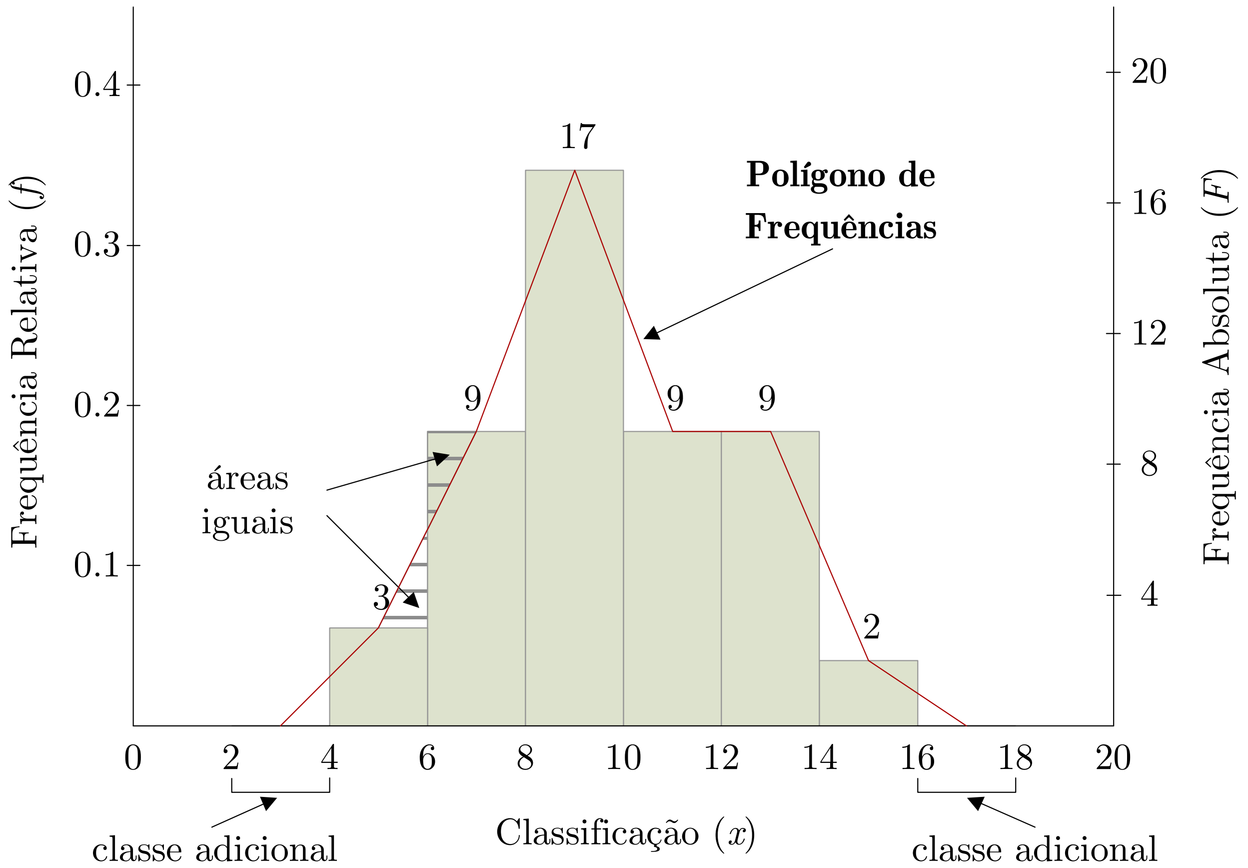
Utilização de números redondos

$$\text{n}^{\text{o}} \text{ classes} \approx \sqrt{n} = \sqrt{49} = 7$$
$$A_k = 10.7/7 = 1.529$$
$$\rightarrow \left.\begin{array}{l} \text{amplitude das classes} = 2 \\ \text{valor mínimo} = 4 \end{array}\right\} 6 \; classes$$

Classe		F_k	F'_k	f_k	f'_k
[4, 6 [	|||	**3**	**3**	**0.061**	**0.061**
[6, 8 [	|||||||||	**9**	**12**	**0.184**	**0.245**
[8, 10 [	|||||||||||||||||	**17**	**29**	**0.347**	**0.592**
[10, 12 [	|||||||||	**9**	**38**	**0.184**	**0.776**
[12, 14 [	|||||||||	**9**	**47**	**0.184**	**0.959**
[14, 16]	||	**2**	**49**	**0.041**	**1**
	$\sum$	**49**		**1**	

DGI
2019

Histograma e Polígono de Frequências



1.2.2. Estatísticas de Localização

1.2.2.1 Média Amostral $(\bar{x})$

- A média toma um *valor* que é *central* em relação aos dados que constituem a amostra.

$$\sum_{i=1}^{n}(x_i - \overline{x}) = 0$$

- A média *minimiza a soma dos erros quadráticos* dos dados.

$$SEQ = \sum_{i=1}^{n}(x_i - c)^2$$

Expressões de Cálculo da Média Amostral

$\overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$	*dados não agrupados*
$\overline{x} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{K} F_k \times x_k$ ou $\overline{x} = \displaystyle\sum_{k=1}^{K} f_k \times x_k$	*dados discretos agrupados*
$\overline{x} \approx \dfrac{1}{n}\displaystyle\sum_{k=1}^{K} F_k \times M_k$ ou $\overline{x} \approx \displaystyle\sum_{k=1}^{K} f_k \times M_k$	*dados contínuos agrupados*

Exemplo

Calcule a classificação média obtida pelos 49 alunos no exame de Estatística.

i) Dados não agrupados

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{49} \times 476 = 9.7143$$

ii) Dados agrupados

$$\overline{x} \approx \frac{1}{n}\sum_{k=1}^{K} F_k \times M_k = \frac{1}{49}\left(3 \times 5 + 9 \times 7 + 17 \times 9 + 9 \times 11 + 9 \times 13 + 2 \times 15\right) =$$

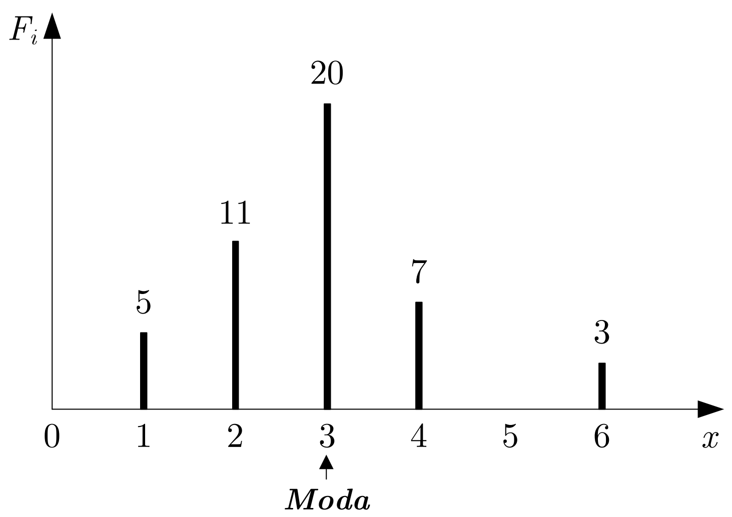
$$= 9.7347$$

2.2.2.2. *Moda (Mod)*

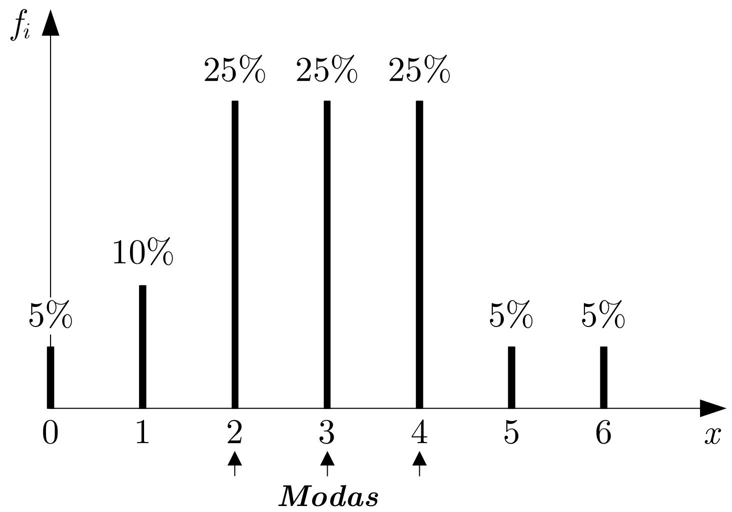
Valor mais comum de um conjunto de observações.

- ***Pode não existir*** **(conjunto amodal) e se existir *pode não ser única.***

i) Dados Discretos



(Distribuição unimodal)

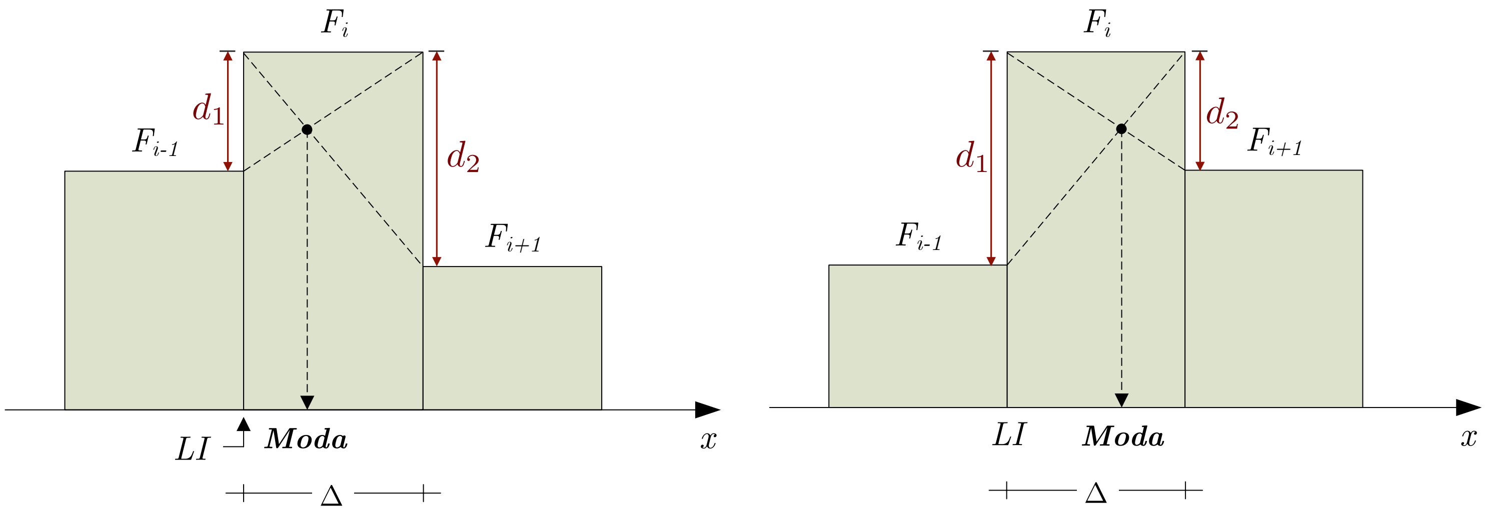


(Distribuição multimodal)

ii) Dados Contínuos

- Identificação da classe (ou classes) a que corresponde a maior frequência - Classe modal.

$$Moda = LI + \frac{d_1}{d_1 + d_2} \times \Delta$$



(aproxima a moda à classe adjacente de maior frequência)

DGI
2019

1.2.2.3. *Mediana (Med)*

A ideia é "partir ao meio" o conjunto dos valores observados.

- Inicia-se o cálculo com a *ordenação dos dados* por ordem crescente ou decrescente formando o vector: $(x_{1:n}, x_{2:n}, \ldots, x_{n:n})$

Dados Discretos

i) Dados não agrupados

- ***Se n for ímpar*, a mediana toma o valor do dado que, nesse vector, ocupa a posição central:**

$$Med = x_{(n+1)/2}$$

Exemplo

Calcule a mediana para o seguinte conjunto de valores: $\{6,\ 5,\ 8,\ 8,\ 4,\ 3,\ 10\}$

- ***Se n for par***, **a mediana toma o valor médio dos dois termos cuja localização no vector se aproxima mais da posição central:**

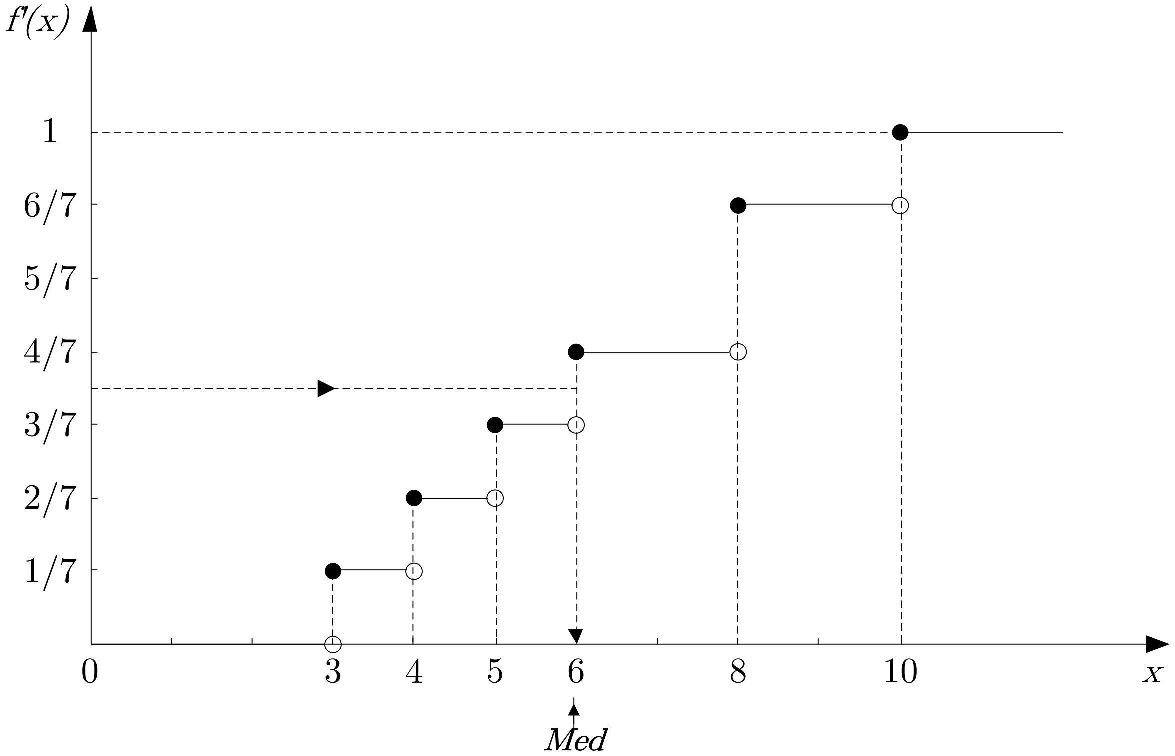
$$Med = \frac{x_{n/2} + x_{n/2+1}}{2}$$

Exemplo

Calcule a mediana do seguinte conjunto de valores: $\{1,\ 3,\ 3,\ 5,\ 7,\ 8\}$

Dados agrupados

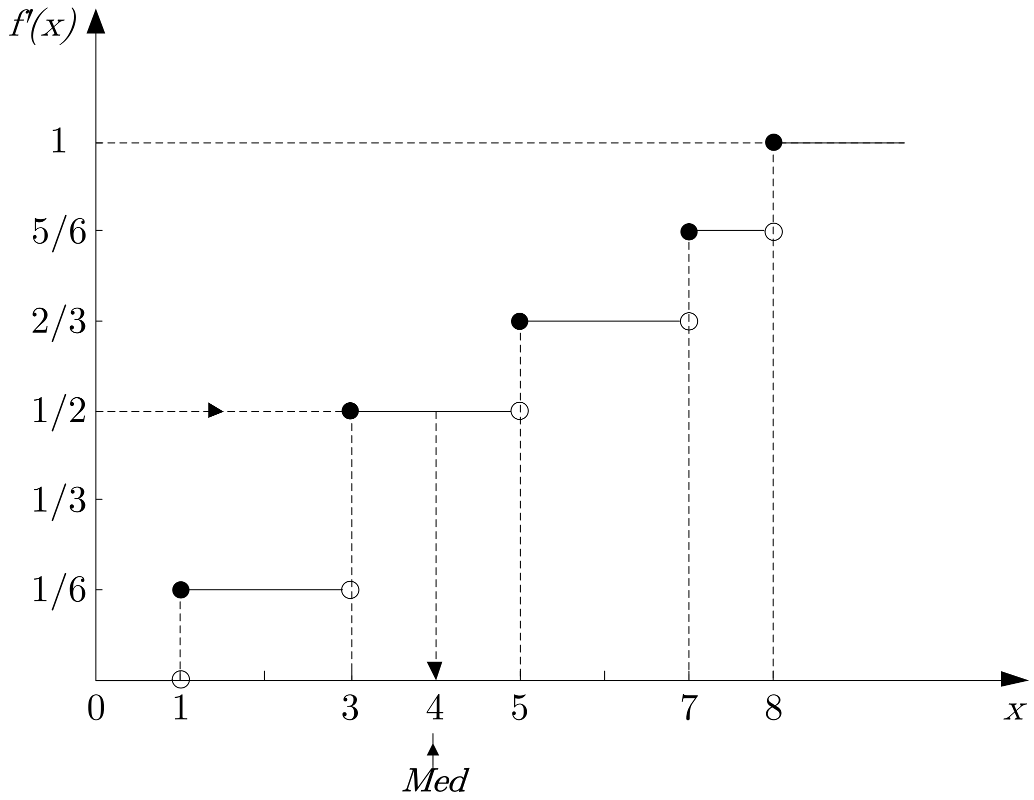
$$\{3,\ 4,\ 5,\ 6,\ 8,\ 8,\ 10\}$$



DGI
2019

Dados agrupados

$$\{1,\ 3,\ 3,\ 5,\ 7,\ 8\}$$



Dados Contínuos

- Para dados contínuos agrupados, começa por fazer-se a *identificação da classe mediana*.

- A classe mediana é aquela onde as frequências acumuladas passam de um valor inferior para um valor superior a metade dos dados ($n/2$ ou 0.5)

$$Med = LI + \frac{n/2 - F'_{i-1}}{F_i} \times \Delta$$

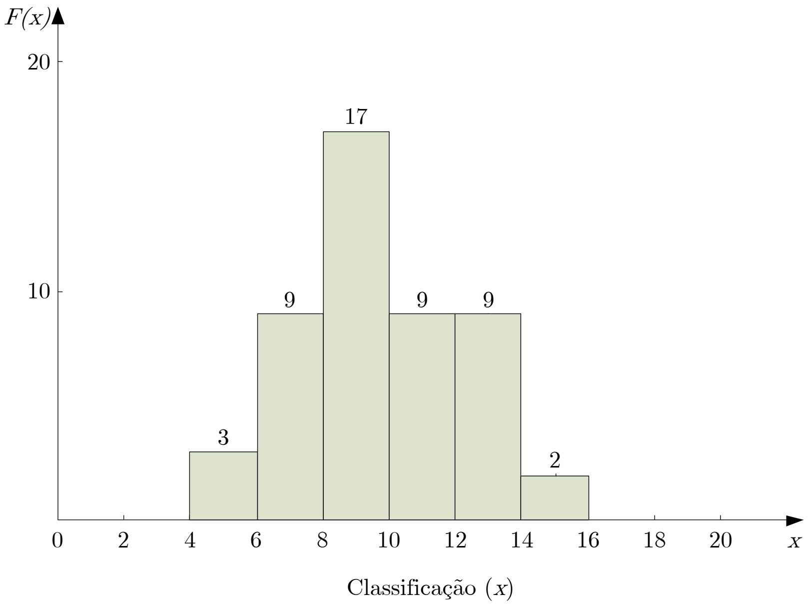
(frequências absolutas)

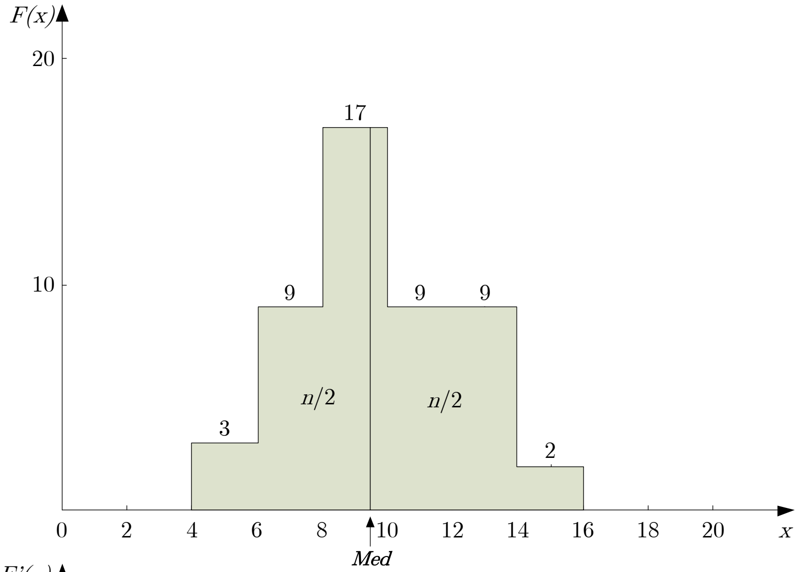
$$Med = LI + \frac{0.5 - f'_{i-1}}{f_i} \times \Delta$$

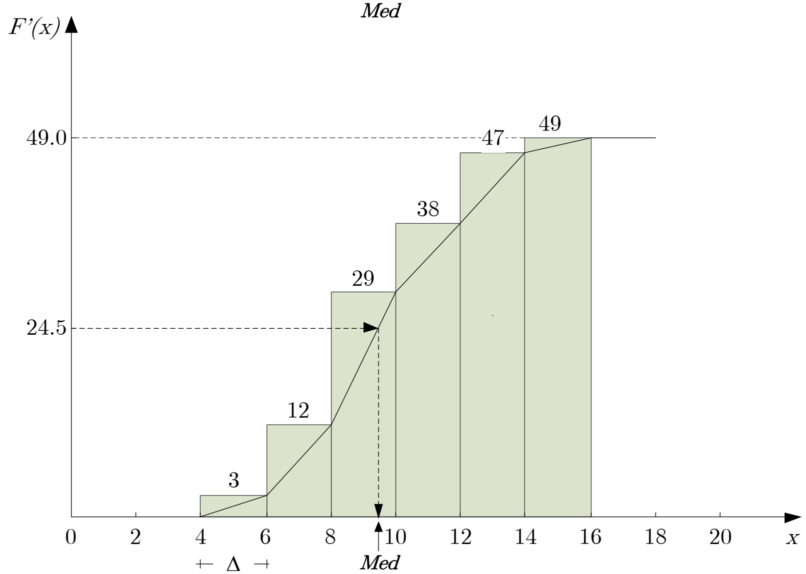
(frequências relativas)

Sendo:

LI	-	limite inferior da classe mediana
F'_i e f'_i	-	frequências absoluta e relativa acumuladas da classe mediana
F'_{i-1} e f'_{i-1}	-	frequências absoluta e relativa acumuladas da classe que precede a classe mediana
F_i e f_i	-	frequências absoluta e relativa da classe mediana
Δ	-	amplitude das classes







Cálculo do valor da mediana da classificação obtida pelos 49 alunos no exame de Estatística

$$Med = LI + \frac{n/2 - F'_{i-1}}{F'_i - F'_{i-1}} \times \Delta =$$

$$= 8 + \frac{49/2 - 12}{29 - 12} \times 2 = 9.471$$

1.2.2.4. Quantis

Quantis

Percentis	*Decis*	*Quartis*
(p_k)	(d_k)	(q_k)
$\left(\dfrac{k}{100}\right)$	$\left(\dfrac{k}{10}\right)$	$\left(\dfrac{k}{4}\right)$
$k = 1, 2, ..., 99$	$k = 1, 2, ..., 9$	$k = 1, 2, 3$

- $Med = p_{50} = d_5 = q_2$
- $p_{10} = d_1, p_{20} = d_2, \ldots, p_{90} = d_9$
- $p_{25} = q_1, p_{50} = q_2, \ldots, p_{75} = q_3$

1.2.3. Estatísticas de dispersão

1.2.3.1. Amplitude do Intervalo de Variação

(estatística básica de dispersão)

$$A = \max\{x_1, x_2, \ldots, x_n\} - \min\{x_1, x_2, \ldots, x_n\}$$

1.2.3.2. Variância amostral $\left(s^2\right)$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

(dados não agrupados)

$$s^2 = \frac{n}{n-1}\sum_{k=1}^{K} f_k \times (x_k - \overline{x})^2$$

(dados discretos agrupados)

$$s^2 = \frac{n}{n-1}\sum_{k=1}^{K} f_k \times (M_k - \overline{x})^2$$

(dados contínuos agrupados)

1.4.3.3 Desvio Padrão (s)

$$s = \sqrt{s^2}$$

1.4.4. Outras Estatísticas

1.4.4.1 Momentos

Momentos na Origem de ordem 1, 2, ...

$m'_i = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} x_j^i$	*dados não agrupados*
$m'_i = \displaystyle\sum_{k=1}^{K} f_k \times x_k^i$	*dados discretos agrupados*
$m'_i \approx \displaystyle\sum_{k=1}^{K} f_k \times M_k^i$	*dados contínuos agrupados*

Momentos Centrados de ordem 1, 2, ...

$m_i = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^i$	*dados não agrupados*
$m_i = \sum_{k=1}^{K} f_k \times (x_k - \bar{x})^i$	*dados discretos agrupados*
$m_i \approx \sum_{k=1}^{K} f_k \times (M_k - \bar{x})^i$	*dados contínuos agrupados*

Estimadores não enviesados dos momentos centrados

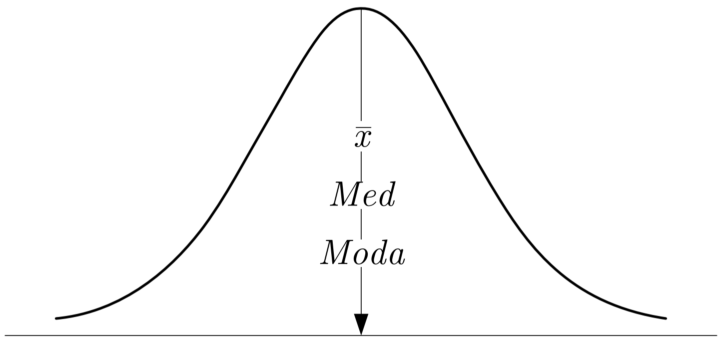
$k_2 = s^2 = \dfrac{n}{n-1} \times m_2$	2ª ordem
$k_3 = \dfrac{n^2}{(n-1)(n-2)} \times m_3$	3ª ordem
$k_4 = \dfrac{n \times (n^2 - 2n + 3)}{(n-1)(n-2)(n-3)} \times m_4 - \dfrac{3n \times (2n-3)}{(n-1)(n-2)(n-3)} \times m_2^2$	4ª ordem

1.4.4.2. ***Coeficiente de Assimetria***

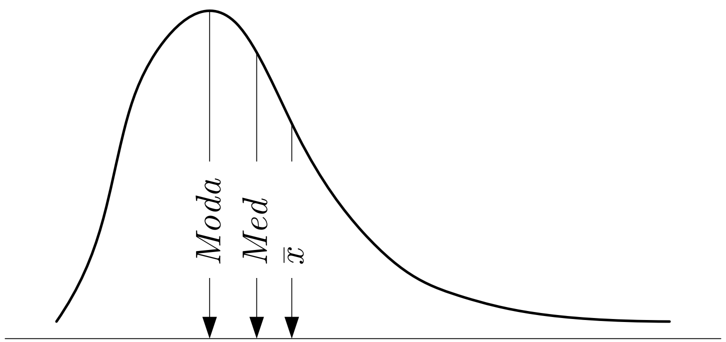
$CA = \dfrac{m_3}{m_2^{3/2}}$	*enviesado*
$CA' = \dfrac{k_3}{k_2^{3/2}}$	*enviesado (menos que o anterior)*
$CA'' = \dfrac{\sqrt{n \times (n-1)}}{n-2} \times CA$	*não enviesado para populações normais*

$$\begin{cases} \text{Distribuição simétrica} \quad \rightarrow \; C_A = 0 \\ \text{Distribuição assimétrica} \begin{cases} \text{assimétrica à esquerda} \quad \rightarrow \; C_A < 0 \\ \text{assimétrica à direita} \quad \rightarrow \; C_A > 0 \end{cases} \end{cases}$$

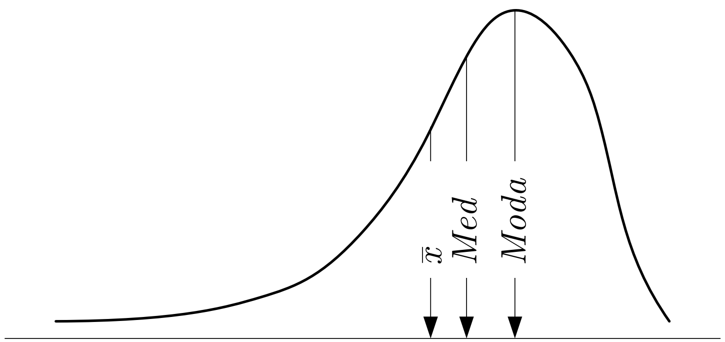
Posição da Média, Moda e Mediana



(distribuição simétrica)



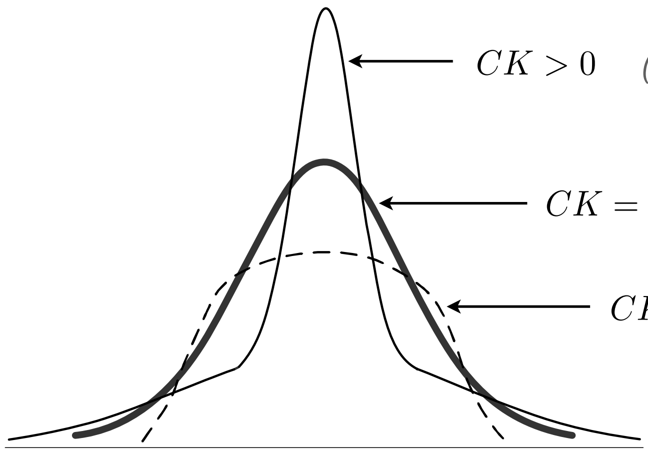
(distribuição assimétrica à direita)



(distribuição assimétrica à esquerda)

1.4.4.3. Coeficiente de Achatamento ou de Kurtosis

$CK = \dfrac{m_4}{m_2^2} - 3$	*enviesado*
$CK' = \dfrac{k_4}{k_2^2} - 3$	*enviesado (menos que o anterior)*
$CK'' = \dfrac{n-1}{(n-2)(n-3)} \times [(n+1)CK + 6]$	*não enviesado para populações normais*



Exemplo

Calcule os valores dos coeficientes de assimetria e de achatamento relativos à classificação obtida pelos 49 alunos no exame de Estatística.

- *Considerando os dados não agrupados:*

$$m_2 = 6.52980$$

$$m_3 = 1.86658$$

$$m_4 = 95.39174$$

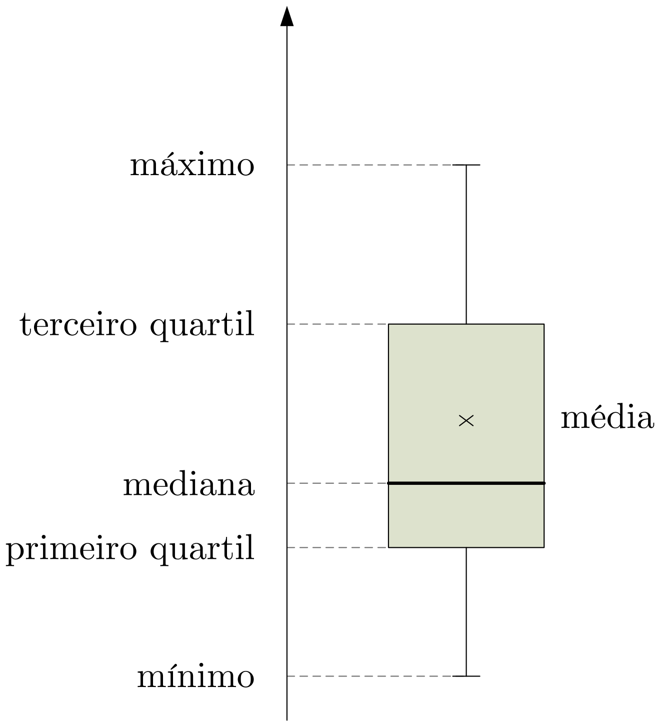
$$CA = 0.11187$$

$$CK = -0.7628$$

DGI
2019

1.4.5. Representação Gráfica de Estatísticas

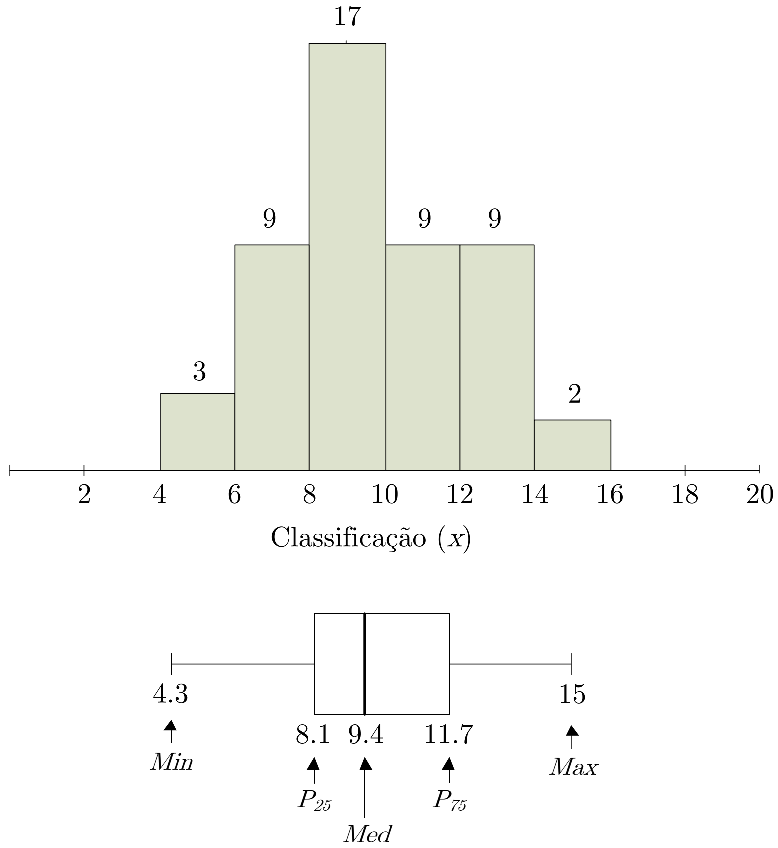
1.4.5.1. Box & Whisker Plots



DGI
2019

Exemplo

Construção do *box-plot* relativo à classificação obtida pelos 49 alunos no exame de estatística



1.5. Caracterização de Amostras Bivariadas

Uma amostra bivariada é constituída por pares ordenados (x,y), sendo x e y, atributos de um mesmo objecto

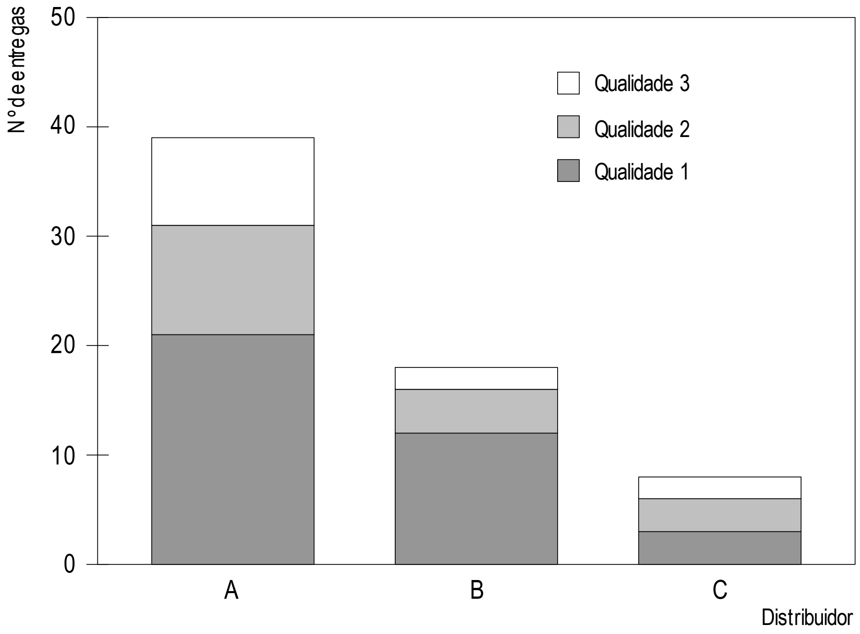
1.5.1. *Dados Qualitativos*

As formas mais usuais de caracterizar amostras bivariadas envolvem o recurso a ***tabelas de informação cruzada e a diagramas de barras sobrepostas***.

Tabela de Informação Cruzada

	Qualidade das Entregas			
Distribuidor	1	2	3	Total da Linha
A	21	10	8	39
	32.3%	15.4%	12.3%	60.0%
	53.8%	25.6%	20.5%	100.0%
	58.3%	58.8%	66.7%	-
B	12	4	2	18
	18.5%	6.2%	3.1%	27.7%
	66.7%	22.2%	11.1%	100.0%
	33.3%	23.5%	16.7%	-
C	3	3	2	8
	4.6%	4.6%	3.1%	12.3%
	37.5%	37.5%	25.0%	100.0%
	8.3%	17.6%	16.7%	-
	36	17	12	65
Total da Coluna	55.4%	26.2%	18.5%	100.0%
	100.0%	100.0%	100.0%	-

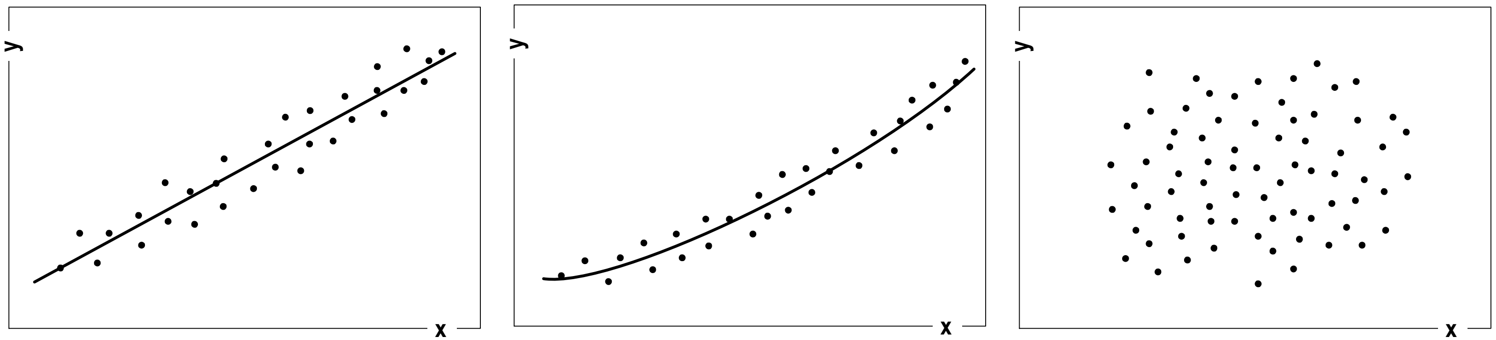
Diagrama de Barras Sobrepostas



1.5.2. *Dados Quantitativos*

- A melhor forma de caracterizar a relação entre as duas variáveis é a sua representação conjunta num sistema de eixos ortogonais - diagrama de dispersão

- A construção do diagrama de dispersão apresenta uma dupla função: Ajuda a determinar se existe alguma relação entre as variáveis e, caso exista, permite identificar a equação mais apropriada para a descrever.



(relação linear) *(relação não linear)* *(ausência de relação)*

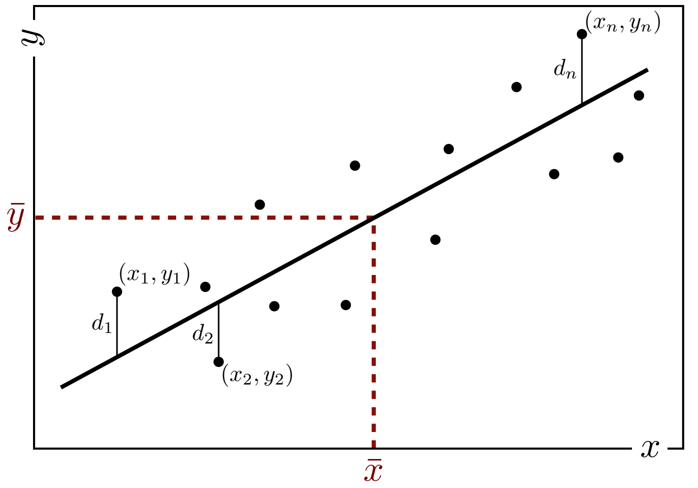
1.5.2.1. *Modelo de Regressão Linear*

$$y_i = f(\theta_1, \theta_2, \ldots, \theta_k) = f(a, b) = a + b \cdot x_i$$

sendo: y - variável dependente, x - variável independente , a - ordenada na origem

Método dos Mínimos Quadrados

- Ajustamento de uma *relação linear* aos dados observados que *minimiza o somatório do quadrado das distâncias entre os valores observados e o modelo ajustado.*



$$d_1^2 + d_2^2 + \ldots + d_n^2 = \text{mínimo}$$

$$SEQ = SEQ\left(a,b\right) = \sum_{i=1}^{N} {e_i}^2 = \sum_{i=1}^{N} \left[y_i - \left(a + b\cdot x_i\right)\right]^2$$

minimizar:

$$\begin{cases} \dfrac{\partial SEQ\left(a,b\right)}{\partial a} = 0 & \rightarrow \quad a = \dfrac{\frac{1}{N}\sum\limits_{i=1}^{N} {x_i}^2 \cdot \sum\limits_{i=1}^{N} y_i - \frac{1}{N}\sum\limits_{i=1}^{N} x_i \cdot \sum\limits_{i=1}^{N} x_i \cdot y_i}{\sum\limits_{i=1}^{N} {x_i}^2 - \frac{1}{N}\cdot\left(\sum\limits_{i=1}^{N} x_i\right)^2} \\ \dfrac{\partial SEQ\left(a,b\right)}{\partial b} = 0 & \rightarrow \quad b = \dfrac{\sum\limits_{i=1}^{N} x_i \cdot y_i - \frac{1}{N}\sum\limits_{i=1}^{N} x_i \cdot \sum\limits_{i=1}^{N} y_i}{\sum\limits_{i=1}^{N} {x_i}^2 - \frac{1}{N}\cdot\left(\sum\limits_{i=1}^{N} x_i\right)^2} \end{cases}$$

DGI
2019

Coeficiente de Correlação Amostral (r)

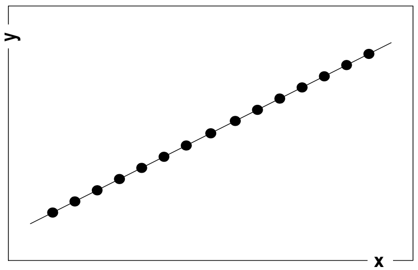
- Medida do ***grau de relacionamento*** linear entre os dados de uma amostra bivariada.

$$r = \frac{S_{xy}}{S_x \cdot S_y} = \frac{\sum_{i=1}^{N} x_i \cdot y_i - \frac{1}{N} \sum_{i=1}^{N} x_i \cdot \sum_{i=1}^{N} y_i}{\sqrt{\left[\sum_{i=1}^{N} {x_i}^2 - \frac{1}{N}\left(\sum_{i=1}^{N} x_i\right)^2\right] \cdot \left[\sum_{i=1}^{N} {y_i}^2 - \frac{1}{N}\left(\sum_{i=1}^{N} y_i\right)^2\right]}}$$

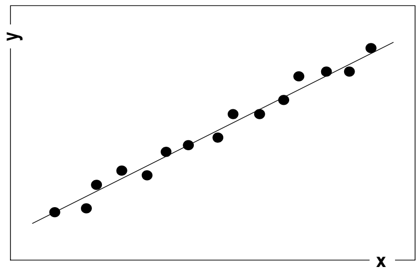
sendo: $S_{xy} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x}) \cdot (y_i - \bar{y})$

$$S^2{}_x = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2 \qquad S^2{}_y = \frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{y})^2$$

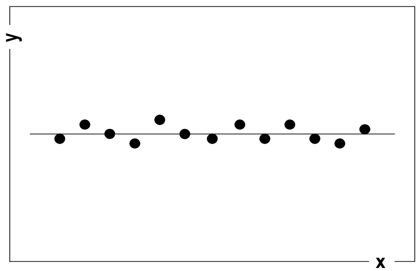
Variação do Coeficiente de Correlação Amostral



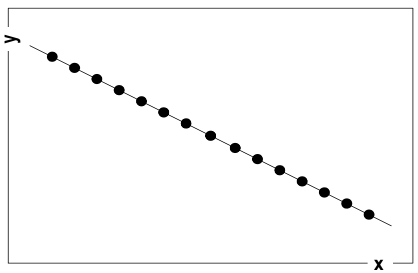
$r = 1$



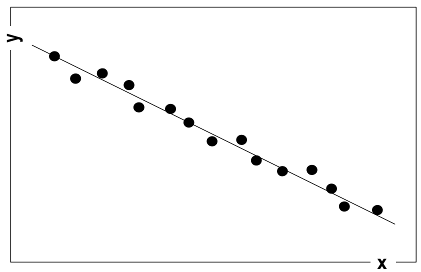
$0 < r < 1$



$r = 0$



$r = -1$



$0 > r > -1$

Coeficiente de Determinação (r^2)

- Traduz a *proporção da variação da variável dependente (Y)* que é *explicada pela variação da variável independente (X)*, através do modelo de regressão.

  – *Variação do Coeficiente de Determinação*

$$0 \leq r^2 \leq 1$$

Exemplo

Os dados representados a seguir referem-se ao volume de vendas mensais, expresso em unidades monetárias, de um artigo comercializado por uma empresa. Os valores representados referem-se aos últimos 12 meses.

mês	1	2	3	4	5	6	7	8	9	10	11	12
vol. vendas	3.7	4.1	5.9	6.6	6.4	7.2	8	8.8	9.4	9.3	9.7	9.9

Com base nos dados disponíveis determine:

a) A média e a variância amostral do volume de vendas mensal.

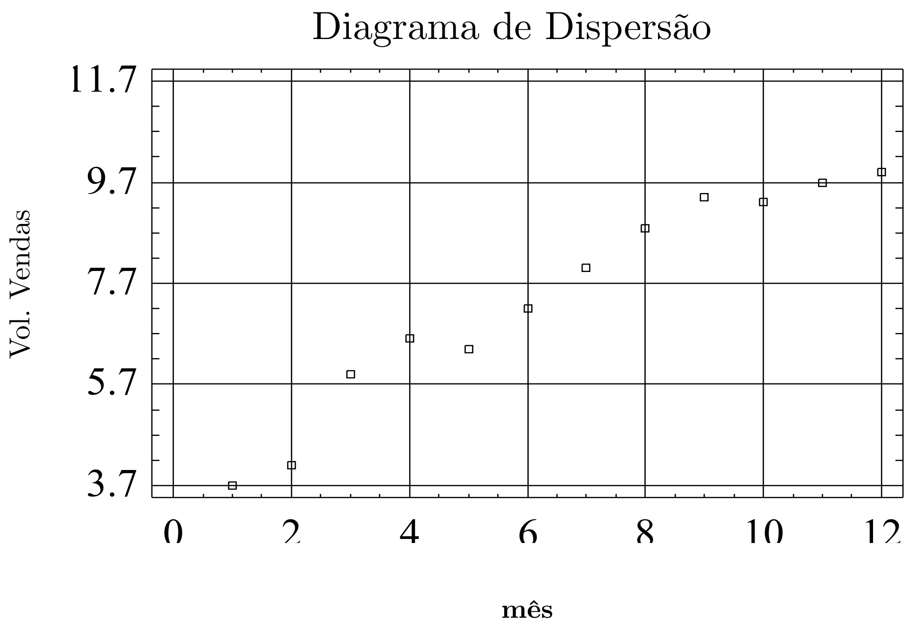
b) A correlação amostral entre os meses e o volume de vendas e, ainda, a previsão do volume de vendas para o próximo mês.

Resolução

a) $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{3.7+4.1+...+9.9}{12} = 7.4167$

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2 = \frac{1}{12-1}\cdot\left((3.7-7.4167)^2+\ldots+(9.9-7.4167)^2\right) = 4.5433$$

b)

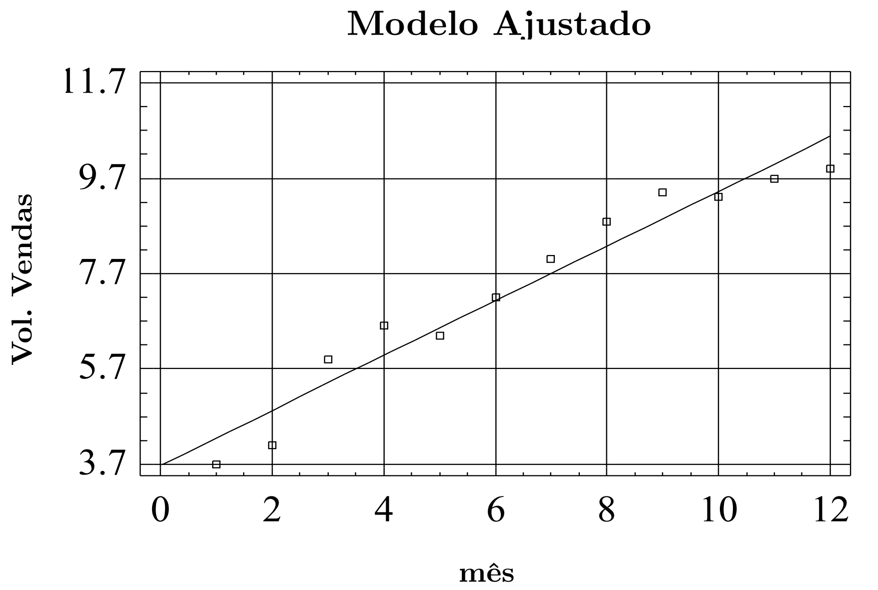


Parâmetros do modelo:

$$a = \frac{\frac{1}{N}\sum_{i=1}^{N} {x_i}^2 \cdot \sum_{i=1}^{N} y_i - \frac{1}{N}\sum_{i=1}^{N} x_i \cdot \sum_{i=1}^{N} x_i \cdot y_i}{\sum_{i=1}^{N} {x_i}^2 - \frac{1}{N}\cdot\left(\sum_{i=1}^{N} x_i\right)^2} = 3.6803$$

$$b = \frac{\sum_{i=1}^{N} x_i \cdot y_i - \frac{1}{N}\sum_{i=1}^{N} x_i \cdot \sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} {x_i}^2 - \frac{1}{N}\cdot\left(\sum_{i=1}^{N} x_i\right)^2} = 0.5748$$

$$y = 0.5748 \cdot x + 3.6803$$



$$r = \frac{\sum_{i=1}^{N} x_i \cdot y_i - \frac{1}{N}\sum_{i=1}^{N} x_i \cdot \sum_{i=1}^{N} y_i}{\sqrt{\left[\sum_{i=1}^{N} {x_i}^2 - \frac{1}{N}\left(\sum_{i=1}^{N} x_i\right)^2\right]\cdot\left[\sum_{i=1}^{N} {y_i}^2 - \frac{1}{N}\left(\sum_{i=1}^{N} y_i\right)^2\right]}} = 0.9723 \quad \rightarrow \quad r^2 = 0.9455$$

Previsão para o 13º mês (x = 13): $y = 0.5748 \times 13 + 3.6803 = 11.153$

2.1. Introdução

Frequentemente deparamo-nos com proposições em relação às quais ***não podemos tomar uma posição certa*** (falsa ou verdadeira). Em vez disso tomamos uma ***posição intermédia de incerteza*** dizendo que a proposição é mais ou menos provável.

A *teoria das Probabilidades* é um ***conjunto de ferramentas matemáticas*** que que têm como primeiro objectivo ***quantificar essa incerteza*** sendo, por isso, aplicada ao estudo de fenómenos (ou experiências) não determinísticos ou aleatórios.

Exemplos de Aplicação:

- **Estatística Indutiva** ***(ex. estimação por intervalo e teste de hipóteses)***
- **Estudos de Fiabilidade** ***(ex. previsão de vida de equipamentos)***
- **Mecânica Quântica** ***(ex. descrição do movimento de partículas)***
- **Optimização** ***(ex. processos de Markov e filas de espera)***
- **Actividades Recreativas** ***(ex. totoloto, jogo de dados, etc.)***

DGI
2019

2.2. Conceitos Fundamentais

2.2.1. *Experiência Aleatória*

Uma experiência aleatória é um ***processo*** (ou um conjunto de circunstâncias) ***sujeito à influência de factores causais***, conduzindo a resultados incertos.

Características Fundamentais

- Pode *repetir-se indefinidamente* sob condições essencialmente inalteradas;
- Cada vez que se realiza a experiência obtém-se um *resultado individual imprevisível*, sendo, no entanto, possível descrever o conjunto de todos os possíveis resultados da experiência;
- Ao longo de um *grande número de repetições* da experiência, os *resultados* são irregulares ou mesmo caóticos quando considerados individualmente, mas, quando considerados em conjunto *evidenciam regularidades* ou padrões bem definidos.

2.2.2. *Espaço Amostral*

Dada uma experiência aleatória, denomina-se por espaço amostral ou espaço de resultados (S) o ***conjunto*** fundamental formado por ***todos os resultados que é possível obter quando se efectua a experiência***.

- Os espaços amostrais podem ser ***discretos ou contínuos***, conforme os seus elementos sejam numeráveis ou não. Os espaços amostrais discretos podem ser ***finitos ou infinitos***.

$S = \{1, 2, 3, 4, 5, 6\}$	*(espaço amostral discreto e finito)*
$S = \{1, 2, 3, \ldots\}$	*(espaço amostral discreto e infinito)*
$S = [0;\ 500]$	*(espaço amostral contínuo)*

- O espaço amostral de uma experiência aleatória ***depende de forma como esta é avaliada***.

2.2.3. Acontecimentos ou Eventos

Na realização de uma experiência aleatória podemos estar interessados na eventualidade de ocorrência de um determinado acontecimento. Um acontecimento está associado a uma experiência aleatória, constituindo um ***sub-conjunto do espaço amostral*** (S) associado a essa experiência, ou seja, um conjunto de resultados possíveis.

2.2.3.1. Acontecimentos Simples

Contêm *exactamente um* dos resultados possíveis de uma experiência aleatória.

2.2.3.2. Acontecimentos Compostos

Contêm *mais do que um* dos resultados possíveis de uma experiência aleatória.

Exemplo

Lançamento de uma moeda ao ar três vezes. Se o resultado for avaliado pela sequência de caras *(H)* e coroas *(T)*, o espaço amostral é constituído por 8 resultados.

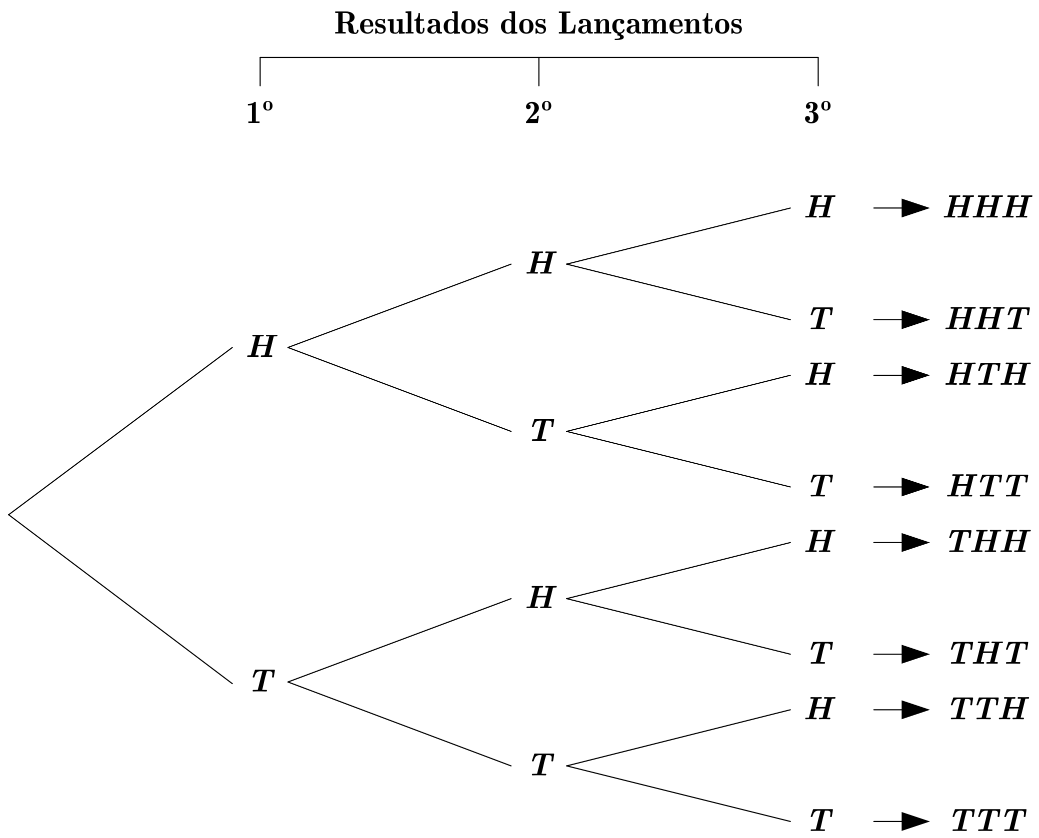


Diagrama em árvore

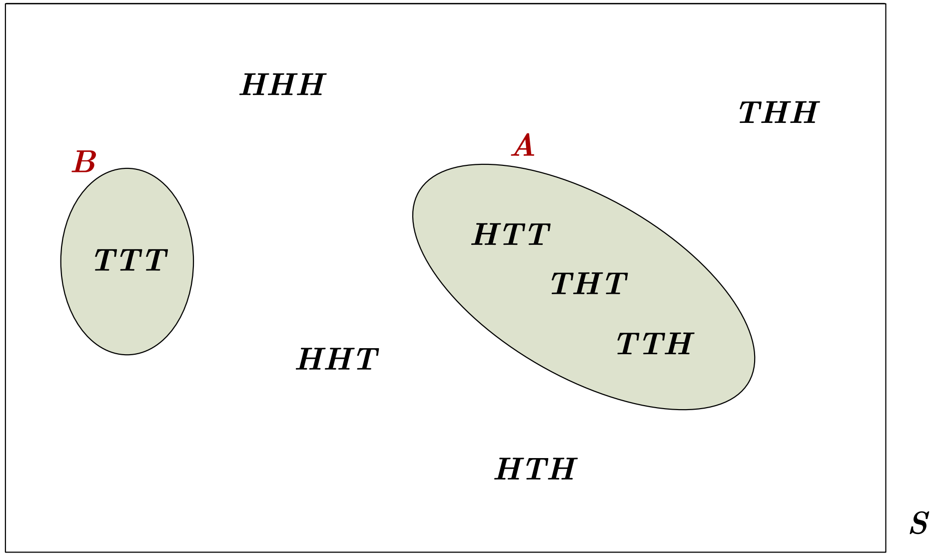


Diagrama de Venn

Considerando que:

A – Representa o acontecimento *"saída de duas coroas"*

B – Representa o acontecimento *"saída de três coroas"*

→ **A – acontecimento composto**

B – acontecimento simples

DGI
2019

2.2.3.3. Acontecimento certo

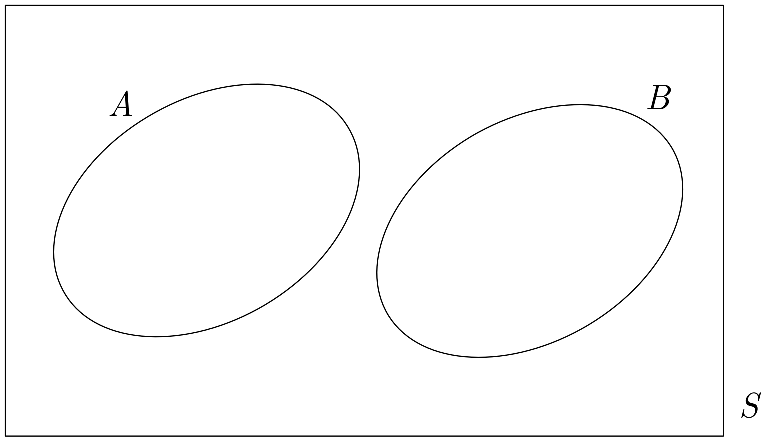
Contem *todos os resultados possíveis* de uma experiência aleatória.

2.2.3.4. Acontecimento impossível

Não contem *nenhum dos resultados possíveis* de uma experiência aleatória.

2.2.3.5. Acontecimentos incompatíveis

Dois acontecimentos dizem-se *mutuamente exclusivos ou incompatíveis* se eles não puderem ocorrer simultaneamente, isto é, se $A \cap B = \emptyset$



Exemplo: *Lançamento de um dado*

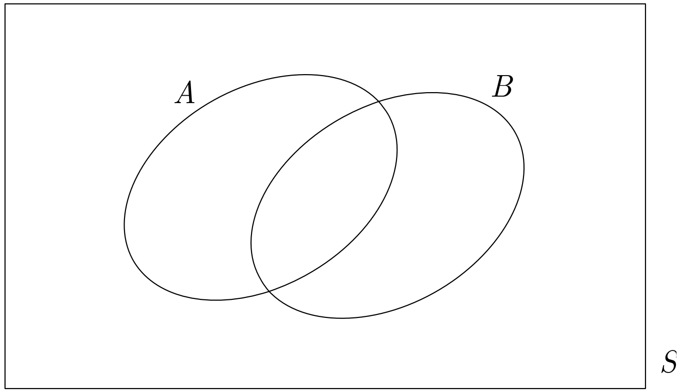
A - saída de face par

B - saída de face ímpar

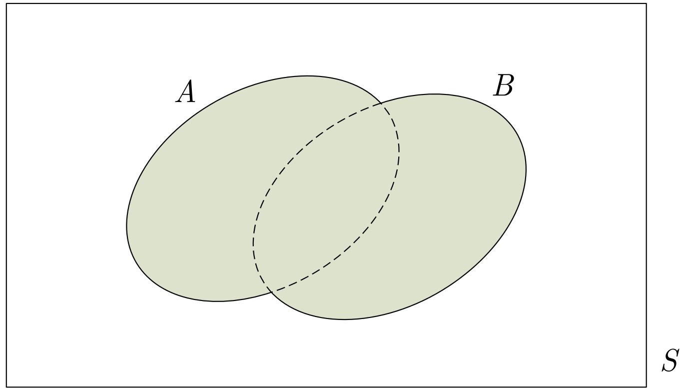
2.2.3.6. *Operações entre Acontecimentos*

Como os acontecimentos correspondem a subconjuntos do espaço amostral, podem aplicar-se as operações lógicas de reunião, intersecção, diferença e complementaridade.

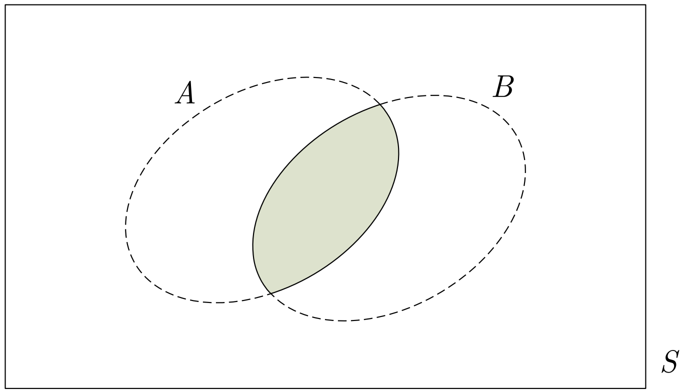
Operações Lógicas entre Conjuntos



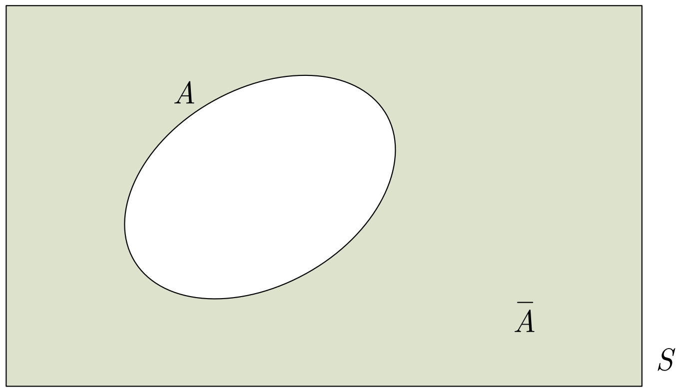
$A,\ B$ e S



Reunião $(A \cup B)$



Intersecção $(A \cap B)$



Complementar de A $(\bar{A})$

Propriedades das operações sobre conjuntos

$A \cup B = B \cup A$	*(comutatividade)*
$A \cap B = B \cap A$	"
$(A \cup B) \cup C = A \cup (B \cup C)$	*(associatividade)*
$(A \cap B) \cap C = A \cap (B \cap C)$	"
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	*(distributividade)*
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	"
$\overline{A \cup B} = \overline{A} \cap \overline{B}$	*(Leis de De Morgan)*
$\overline{A \cap B} = \overline{A} \cup \overline{B}$	"
$A \cup \emptyset = A$	*(elemento neutro)*
$A \cap S = A$	"
$A \cup S = S$	*(elemento absorvente)*
$A \cap \emptyset = \emptyset$	"

2.2.4. Conceito de Probabilidade

2.2.4.1. *Definição Clássica (Lei de Laplace)*

Se um acontecimento pode ocorrer de N_A maneiras diferentes, num total de N maneiras possíveis (N resultados mutuamente exclusivos e equiprováveis) então a probabilidade do acontecimento A é dada por:

$$P(A) = \frac{N_A}{N}$$

Propriedades

P1: $0 \leq P(A) \leq 1$

P2: $P(S) = 1$

P3: $P(A \cup B) = P(A) + P(B)$ *(A e B mutuamente exclusivos)*

DGI
2019

2.2.4.2. Definição Frequencista

Se após N repetições de uma experiência aleatória, se observarem N_A ocorrências do acontecimento A, então a probabilidade de ocorrência do acontecimento A é dada por:

$$P(A) = \lim_{N \to \infty} \frac{N_A}{N}$$

Propriedades

P1: $0 \leq P(A) \leq 1$

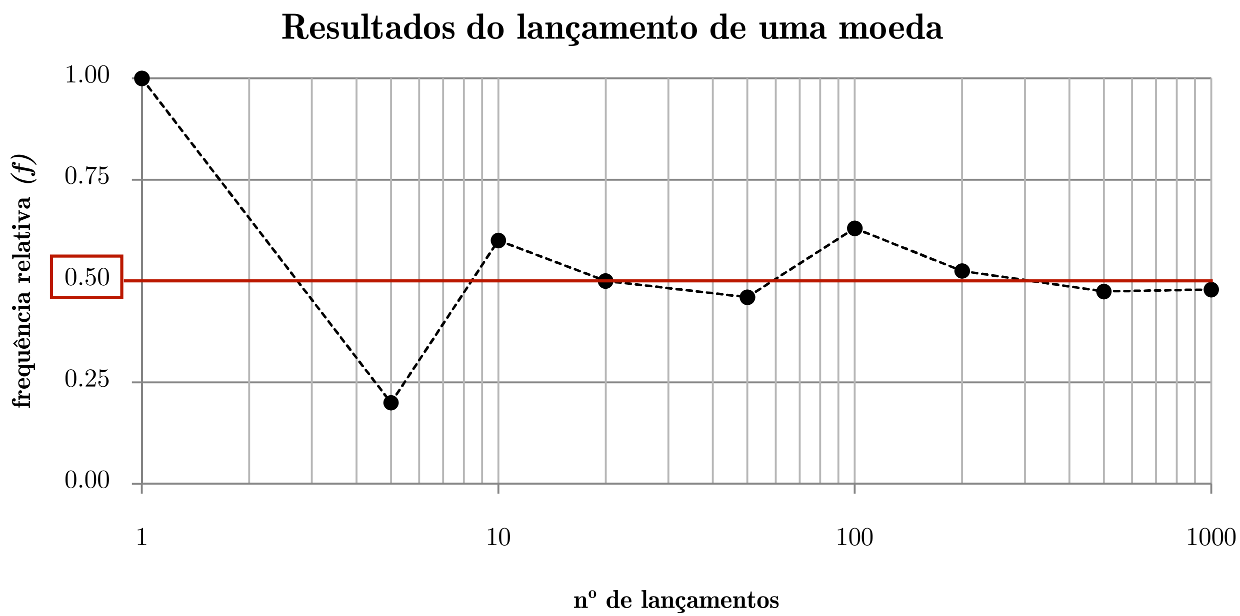
P2: $P(S) = 1$

P3: $P(A \cup B) = P(A) + P(B)$ *(A e B mutuamente exclusivos)*

Exemplo

O quadro abaixo traduz os resultados obtidos do lançamento de uma moeda ao ar. Qual a probabilidade de "sair face"?

nº de lançamentos	1	5	10	20	50	100	200	500	1000
nº de "faces"	1	1	6	10	23	63	105	237	479
frequência relativa	1	0.2	0.6	0.5	0.46	0.63	0.525	0.474	0.479



2.2.4.3. Definição Axiomática

$P(A)$ é a *probabilidade de ocorrência* do acontecimento A desde que sejam *satisfeitos os seguintes axiomas*:

Axioma 1: para *qualquer acontecimento* A (*i.e.*, qualquer subconjunto de um espaço amostral S), a probabilidade desse acontecimento satisfaz a relação: $0 \leq P(A) \leq 1$

Axioma 2: A *probabilidade* associada ao *acontecimento certo* é:

$$P(S) = 1$$

Axioma 3: Se *dois acontecimentos* A e B forem *mutuamente exclusivos*, então:

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

Com base nos axiomas anteriores podem deduzir-se as seguintes propriedades:

P1: Para qualquer acontecimento A:

$$P(A) + P(\bar{A}) = 1$$

P2: Acontecimento impossível:

$$P(\emptyset) = 0$$

P3: Para quaisquer acontecimentos A e B:

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

2.2.5. Probabilidade Condicional

$P(A|B)$ é a probabilidade de ocorrer o acontecimento A dado que ocorreu (ou ocorrerá) o acontecimento B. Como B ocorreu (ou ocorrerá), B *passa a ser o novo espaço amostral*, que vem substituir o espaço original S.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0$$

ou

$$P(A \cap B) = P(A|B) \times P(B)$$

Exemplo

Lançamento de um dado honesto.

A – saída de face 4

B – saída de face par

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

2.2.6. Acontecimentos Independentes

Se $P\left(A|B\right)=P\left(A\right)$, isto é, se *a probabilidade de ocorrência de* $\boldsymbol{A}$ *não é afectada pela ocorrência, ou não de* $\boldsymbol{B}$, dizemos que $\boldsymbol{A}$ e $\boldsymbol{B}$ são acontecimentos independentes.

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

Se $A_1,A_2,...,A_n$ são independentes então eles devem ser independentes dois a dois:

$$P(A_i\cap A_j)=P(A_i)\times P(A_j),\quad i\neq j$$

três a três:

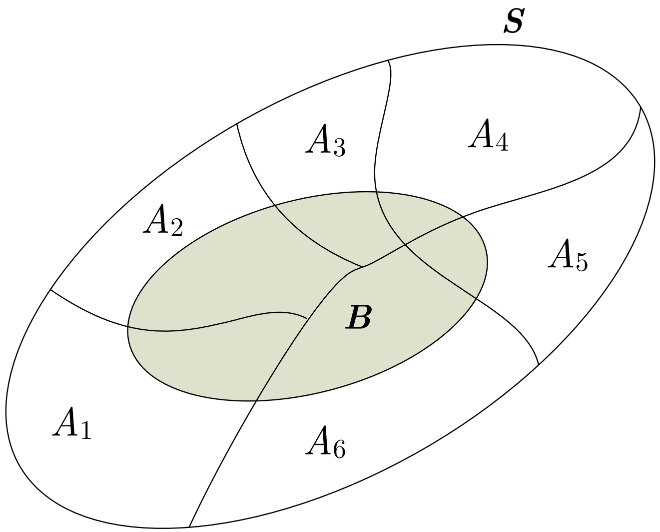
$$P(A_i\cap A_j\cap A_k)=P(A_i)\times P(A_j)\times P(A_k),\quad i\neq j,i\neq k,j\neq k$$

$(\ldots)$

$$P\left(\bigcap_{n=1}^{N}A_n\right)=\prod_{n=1}^{N}P\left(A_n\right)$$

2.2.7. Teorema de Bayes

Sejam $A_1, A_2, \ldots, A_n$ acontecimentos mutuamente exclusivos que constituem uma partição do espaço amostral S.



$$P(A_k|B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(A_k) \times P(B|A_k)}{\sum_{k=1}^{n} P(A_k) \times P(B|A_k)}$$

DGI
2019

Exemplo

Admita-se que num determinado país, 1% da população sofre de tuberculose e ainda que:

- Para uma pessoa que tenha, de facto, contraído a doença, uma microrradiografia tem um resultado positivo (detecta a tuberculose) em 95% dos casos e,
- Para uma pessoa não tuberculosa, essa percentagem é de apenas 0.5%.

Pretende saber-se qual a probabilidade de uma pessoa a quem a microrradiografia tenha dado resultado positivo estar tuberculosa.

2.3. Análise Combinatória

2.3.1. *Princípio fundamental da contagem*

Numa sequência de n selecções/acontecimentos independentes em que cada selecção/acontecimento tem $n_1, n_2, \ldots, n_n$ resultados possíveis, o número total de sequências diferentes é dado por $n_1 \times n_2 \times \ldots \times n_n$.

Notas

As selecções/acontecimentos dizem-se independentes se os resultados obtidos não condicionarem o resultado das novas selecções/acontecimentos.

O número de elementos de um conjunto U denomina-se de Cardinal e representa-se por $\#(U)$.

Exemplo

Calcule o número total de números com três dígitos que podem ser formados se se admitir que:

a) O dígito zero não faz parte da classe das centenas e que não é permitida a repetição de qualquer dígito.

R:

$D = \{0,1,2,3,4,5,6,7,8,9\}$

dezenas

Número (N): $\boxed{\mathsf{c}}\boxed{\mathsf{d}}\boxed{\mathsf{u}}$

centenas *unidades*

$\mathsf{c} = D \setminus \{0\} \quad \rightarrow \quad \#(\mathsf{c}) = 9$

$\mathsf{d} = D \setminus \mathsf{c} \quad \rightarrow \quad \#(\mathsf{d}) = 9$

$\mathsf{u} = D \setminus \{\mathsf{c},\mathsf{d}\} \quad \rightarrow \quad \#(\mathsf{u}) = 8$

$\longrightarrow \#(\mathsf{N}) = \#(\mathsf{c}) \times \#(\mathsf{d}) \times \#(\mathsf{u}) = 648$

b) O dígito zero não faz parte da classe das centenas e que é permitida a repetição de qualquer dígito.

R:

$\mathsf{c} = D \setminus \{0\} \quad \rightarrow \quad \#(\mathsf{c}) = 9$

$\mathsf{d} = D \quad \rightarrow \quad \#(\mathsf{d}) = 10$

$\mathsf{u} = D \quad \rightarrow \quad \#(\mathsf{u}) = 10$

$\longrightarrow \#(\mathsf{N}) = \#(\mathsf{c}) \times \#(\mathsf{d}) \times \#(\mathsf{u}) = 900$

2.3.2. *Permutações Simples*

Designam-se por permutações de n elementos distintos às sequências constituídas por aqueles elementos e que diferem umas das outras apenas pela ordem pela qual eles se dispõem.

$$\mathrm{P}_n = \mathrm{A}_n^n = n!$$

Nota

Factorial de um número natural: $n! = \prod_{k=1}^{n} k, \quad k > 0$

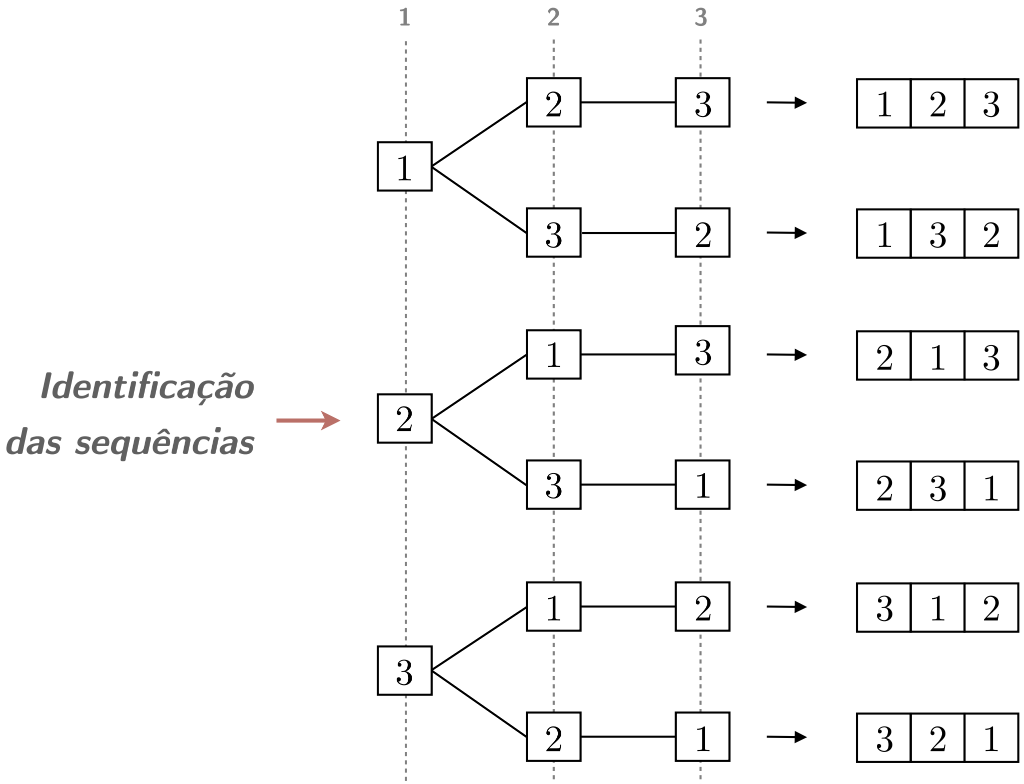
por definição: $0! = 1$

DGI
2019

Exemplo

$U = \{1, 2, 3\} \quad \rightarrow \quad n = 3$

Permutações sem repetição de U: $\mathrm{P}_3 = 3! = 3 \times 2 \times 1 = 6$



DGI

2019

2.3.3. Permutações Circulares

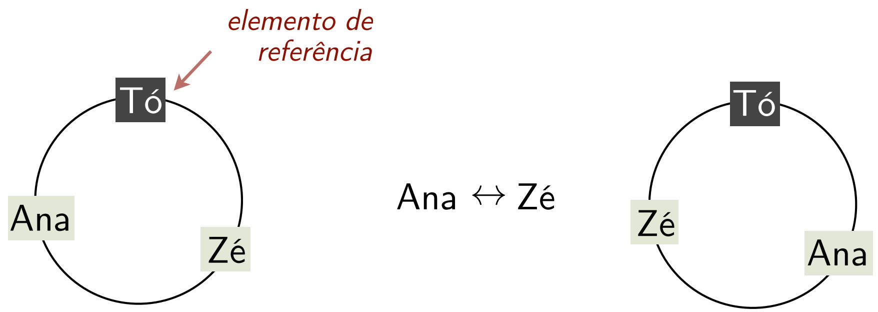
Quando os elementos estão arranjados em círculo é necessário escolher um deles para servir de elemento de referência aos restantes. Assim, duas permutações circulares são diferentes se a sequência dos restantes elementos, no sentido horário, é diferente.

$$\# = (n-1)!$$

Exemplo

De quantas maneiras um grupo de 3 pessoas se pode dispor em torno de uma mesa redonda?

R:



$\longrightarrow \quad \# = (3-1)! = 2$

2.3.4. *Permutações com Repetição*

Se tivermos n elementos e cada um deles se repetir $n_1, n_2, \ldots, n_r$ vezes, o número de permutações é dado por:

$$\mathrm{P}_{n_1,n_2,\ldots,n_r} = \frac{n!}{n_1! \times n_2! \times \ldots \times n_r!}, \quad n_1 + n_2 + \ldots + n_r = n$$

Exemplo

Quantas palavras se podem formar com as letras que compõem a palavra ESTATISTICA?
(as palavras podem ou não ter significado)

R:

$$\text{ESTATISTICA} \left\{ \begin{array}{ll} \text{E} & (\times 1) \\ \text{S} & (\times 2) \\ \text{T} & (\times 3) \\ \text{A} & (\times 2) \\ \text{I} & (\times 2) \\ \text{C} & (\times 1) \end{array} \right. \quad \rightarrow \quad \mathrm{P}_{1,2,3,2,2,1} = \frac{11!}{1! \times 2! \times 3! \times 2! \times 2! \times 1!} = 831\,600$$

DGI
2019

2.3.5. *Arranjos Simples*

Designam-se por arranjos de n elementos distintos tomados p a p os agrupamentos constituídos por p daqueles elementos e que diferem uns dos outros quer pelos elementos que neles figuram quer pela ordem na qual eles se dispõem.

$$\mathrm{A}_p^n = n \times (n-1) \times (n-2) \times \ldots \times (n-p+1) = \frac{n!}{(n-p)!}, \quad n \geq p$$

Exemplo

$U = \{1, 2, 3\} \quad \rightarrow \quad n = 3$

número de Arranjos de U, 2 a 2: $\quad \mathrm{A}_p^n = \mathrm{A}_2^3 = \dfrac{3!}{(3-2)!} = 6$

sequências: $\boxed{1\,|\,2}$ $\boxed{1\,|\,3}$ $\boxed{2\,|\,1}$ $\boxed{2\,|\,3}$ $\boxed{3\,|\,1}$ $\boxed{3\,|\,2}$

DGI
2019

2.3.6. Combinações

Designam-se por combinações de n elementos distintos tomados p a p os agrupamentos constituídos por p daqueles elementos que diferem uns dos outros pelos elementos que neles figuram, independentemente da ordem pela qual se dispõem.

$$\mathrm{C}_p^n = \frac{n!}{p!\,(n-p)!} = \frac{\mathrm{A}_p^n}{p!}, \quad n \geq p$$

Exemplo

$U = \{1, 2, 3\} \quad \rightarrow \quad n = 3$

número de Combinações de U, 2 *a* 2: $\quad \mathrm{C}_p^n = \mathrm{C}_2^3 = \dfrac{3!}{2!\,(3-2)!} = 3$

sequências: $\boxed{1\,|\,2}$ $\boxed{1\,|\,3}$ $\boxed{2\,|\,3}$

Folha de Cálculo

1. *Factorial de um Número Natural*

Designação Port[Ing]: FACTORIAL [FACT]

Definição: FACT(*número ou endereço de célula*)

2. *Arranjos e Permutações Simples*

Designação Port[Ing]: PERMUTAR [PERMUT]

Definição: PERMUT(*número ou endereço de célula* - n, *número ou endereço de célula* - p)

3. *Combinações*

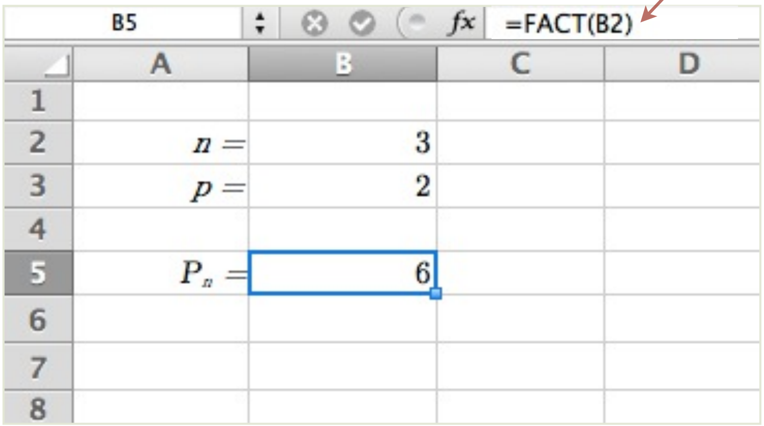
Designação Port[Ing]: COMBIN[COMBIN]

Definição: COMBIN(*número ou endereço de célula* - n, *número ou endereço de célula* - p)

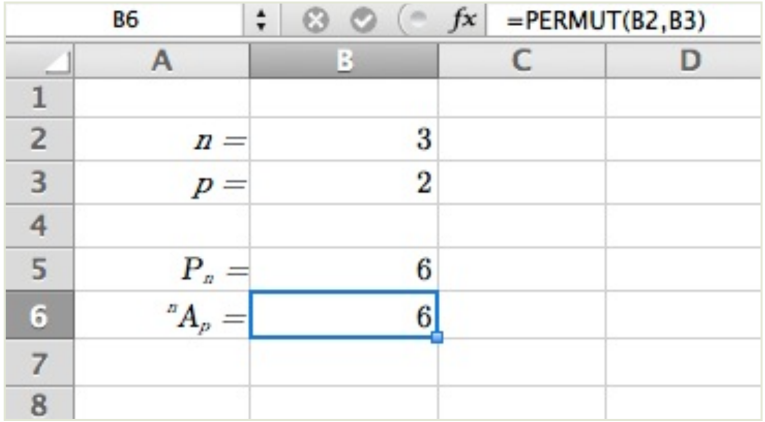
Folha de Cálculo

$U = \{1, 2, 3\}$

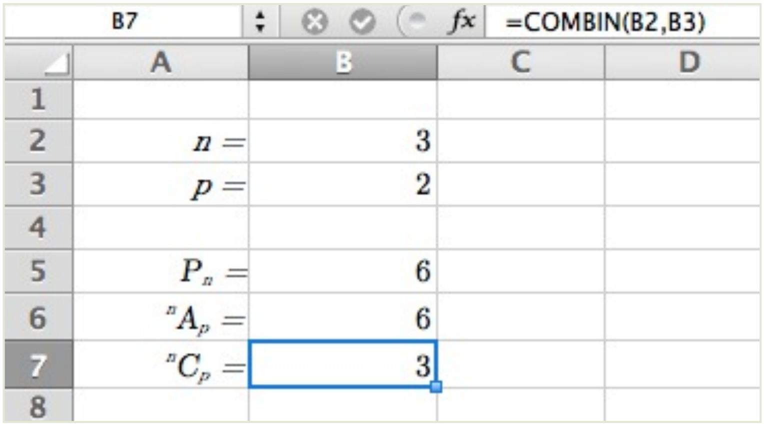
alternativa



Permutações Simples



Arranjos Simples

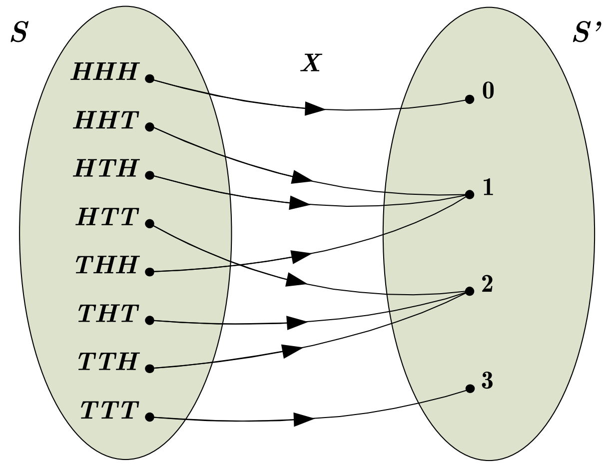


Combinações

3.1. Variáveis Aleatórias

Variáveis que expressam os resultados de uma experiência aleatória.

- Uma variável aleatória é o resultado de uma **aplicação de um espaço amostral (S) num conjunto de chegada qualquer (S')**. **A cada resultado de S corresponde um único valor no conjunto de chegada** (variável aleatória), a diferentes valores de S pode corresponder o mesmo valor no conjunto de chegada.



X - nº de "caras" em 3 lançamentos de uma moeda

- Denotam-se, em geral, por uma **letra maiúscula** $(X,\ Y,\ Z,\ \ldots)$.
- As variáveis aleatórias expressas numa *escala quantitativa* classificam-se em **discretas ou contínuas, consoante o conjunto de valores que possam tomar seja numerável ou não numerável**. As variáveis do *primeiro tipo* encontram-se normalmente associadas a *contagens e as do segundo a medidas*.
- Em algumas situações, conjunto de valores que a variável aleatória toma *pode confundir-se* com o espaço amostral – S.

Exemplo

Experiência aleatória – medição do peso de uma pessoa

Espaço amostral – conjunto de todos os pesos

Variável aleatória – peso

- Noutras situações, **o conjunto de valores** que a variável aleatória toma **distingue-se claramente dos resultados**, correspondendo a uma transformação destes.

Exemplo

Experiência aleatória – três lançamentos de uma moeda equilibrada

Espaço amostral $-S = \{HHH,\ HHT,\ HTH,\ HTT,\ THH, THT,\ TTH,\ TTT\}$

Variável aleatória – número de "caras"

3.2. Distribuições de Variáveis Discretas

3.2.1. *Função de Probabilidade*

Seja Y uma variável discreta. Chama-se função de Probabilidade da variável Y e designa-se por $p(y)$ a **função que associa a cada valor de y a probabilidade de Y ser igual a** y.

$$p(y) = P(Y = y)$$

Dos axiomas de probabilidade resulta que:

P1: $0 \leq p(y) \leq 1$

P2: $\sum_{y} p(y) = 1$

3.2.2. Função Distribuição de Probabilidade

Chama-se Função Distribuição de Probabilidade de Y e designa-se por $F(y)$ **a função que associa a cada valor de y a probabilidade de Y ser menor ou igual a y.**

$$F(y) = P(Y \leq y) = \sum_{Y \leq y} p(y)$$

Como resultado da sua definição, a Função Distribuição de Probabilidade satisfaz as seguintes propriedades:

P1: $0 \leq F(y) \leq 1$

P2: $P(a < Y \leq b) = F(b) - F(a)$

As Funções de Probabilidade e de Distribuição de Probabilidade podem representar-se através de tabelas ou diagramas de barras.

Exemplo

A variável Y traduz o número de dias com precipitação igual ou superior a 10 mm registados no mês de Agosto em Bragança, sendo conhecida a respectiva função de probabilidade $p(y)$.

y	$p(y)$
0	k
1	$2k$
2	$5k$
3	$5k$
4	$5k$
5	k
6	k
$\sum$	1

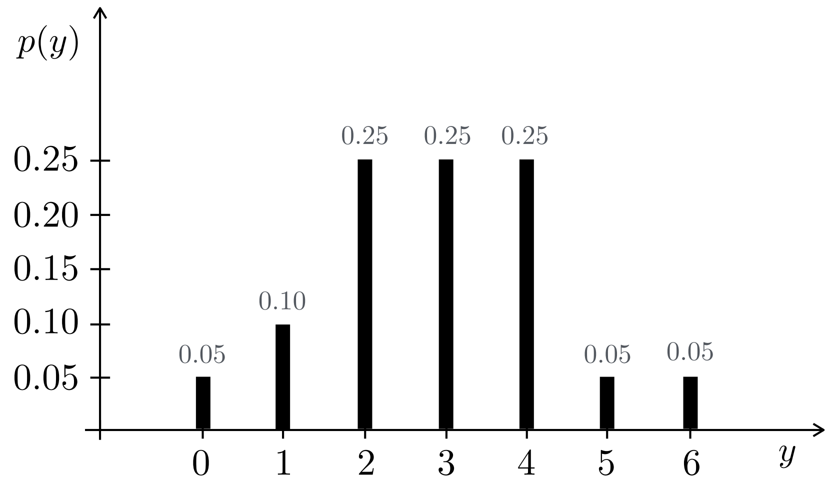
$$\sum_y p(y) = 1 \Leftrightarrow$$

$$\Leftrightarrow k + 2k + 5k + 5k + 5k + k + k = 1 \Leftrightarrow$$

$$\Leftrightarrow k = 0.05 \longrightarrow$$

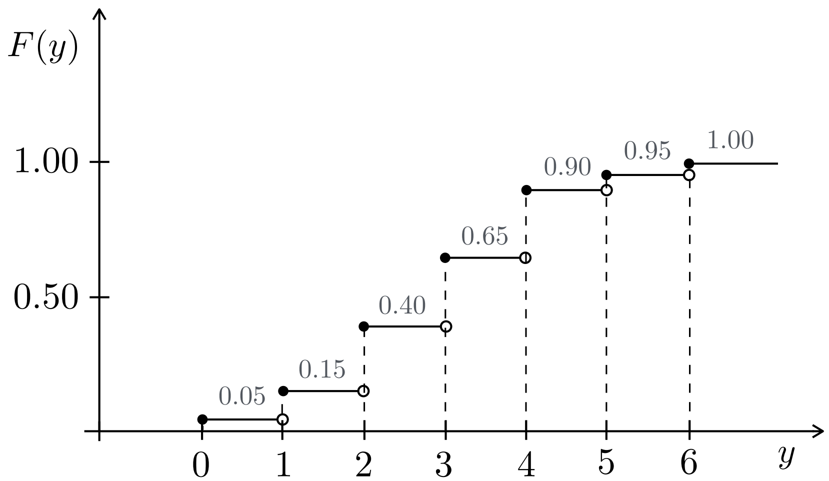
y	$p(y)$	$F(y)$
0	0.05	0.05
1	0.1	0.15
2	0.25	0.4
3	0.25	0.65
4	0.25	0.9
5	0.05	0.95
6	0.05	1
$\sum$	1	

Função de Probabilidade



$$p\left(y\right)=\begin{cases}0.05, & y=0\\ 0.10, & y=1\\ 0.25, & y=2\\ 0.25, & y=3\\ 0.25, & y=4\\ 0.05, & y=5\\ 0.05, & y=6\end{cases}$$

Função Distribuição de Probabilidade



$$F\left(y\right)=\begin{cases}0.00, & y<0\\ 0.05, & 0\leq y<1\\ 0.15, & 1\leq y<2\\ 0.40, & 2\leq y<3\\ 0.65, & 3\leq y<4\\ 0.90, & 4\leq y<5\\ 0.95, & 5\leq y<6\\ 1.00, & y\geq 6\end{cases}$$

3.3. Distribuições de Variáveis Contínuas

3.3.1. Função Densidade de Probabilidade

Seja X uma **variável aleatória que toma valores num espaço amostral contínuo**.

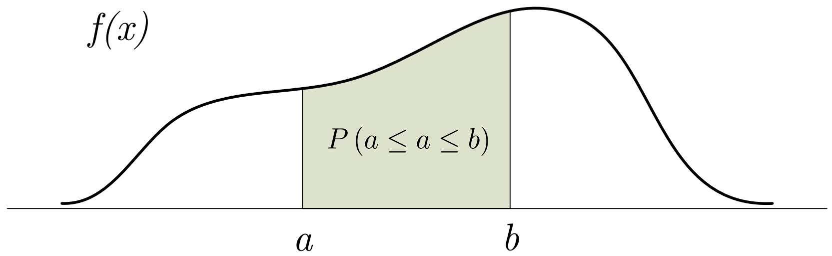
Chama-se Função Densidade de Probabilidade da variável X e designa-se por $f(x)$ uma função tal que:

$$P\left(x \in [x, x + dx]\right) = f(x)\, dx$$

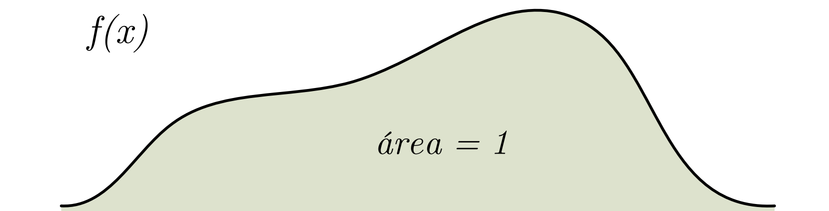
Dos axiomas da probabilidade e da definição de $f(x)$ resulta que:

P1: $f(x) \geq 0$

P2: $P(a \leq x \leq b) = P(a < x < b) = \int\limits_a^b f(x)\, dx$



P3: $\int\limits_{-\infty}^{+\infty} f(x)\, dx = 1$



3.3.2. Função Distribuição de Probabilidade

Chama-se Função Distribuição de Probabilidade de X e designa-se por $F(x)$ a função que associa a cada valor de x a probabilidade de X ser menor ou igual a x.

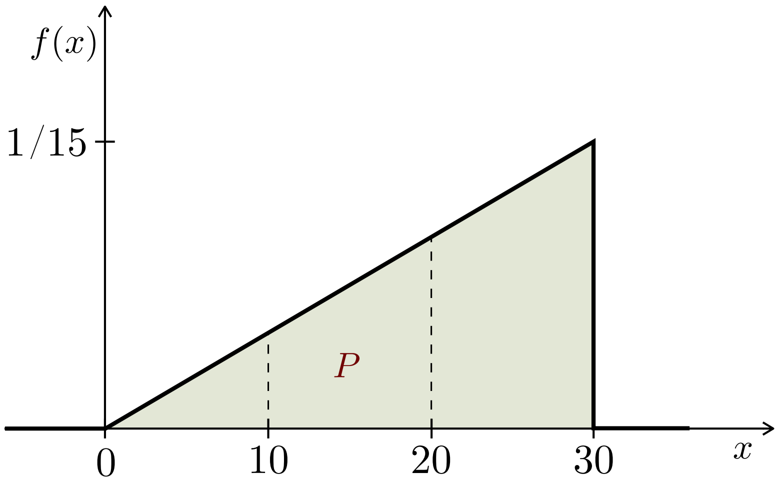
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\ du$$

Da definição de $F(x)$ resulta que:

P1: $\lim\limits_{x \to +\infty} F(x) = 1$

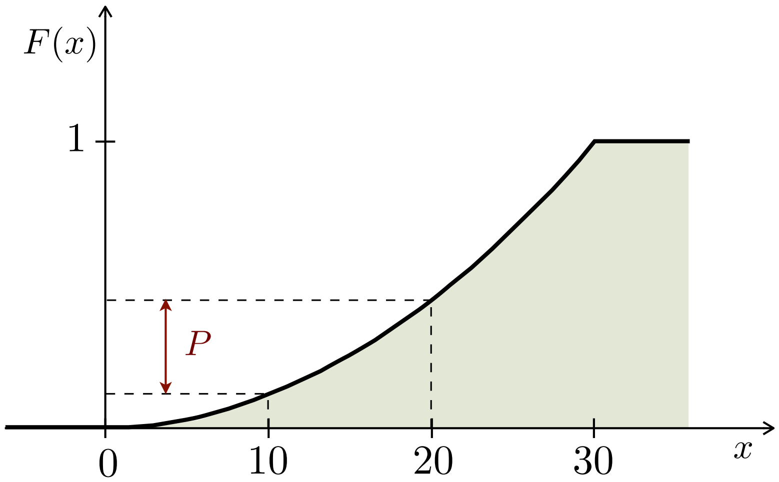
P2: $\dfrac{dF(x)}{dx} = f(x)$

Exemplo



$$f(x) = \begin{cases} 0, & x < 0 \\ \dfrac{1}{450}x, & 0 \leq x \leq 30 \\ 0, & x > 30 \end{cases}$$

$$P = P(10 \leq x \leq 20)$$

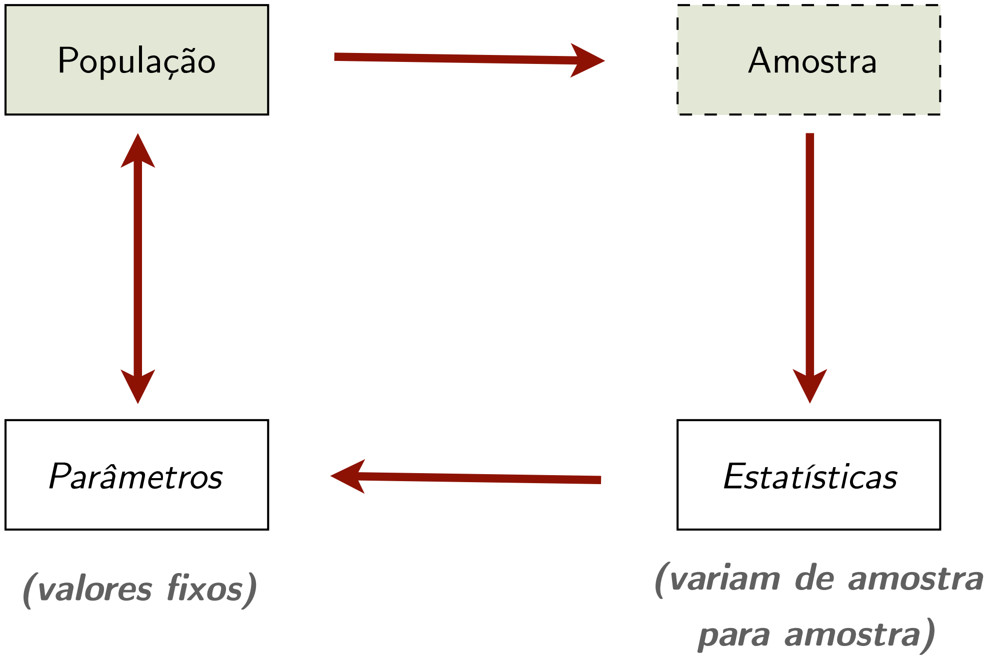


$$F(x) = \begin{cases} 0, & x < 0 \\ \dfrac{1}{900}x^2, & 0 \leq x \leq 30 \\ 1, & x > 30 \end{cases}$$

DGI

2019

3.4. Parâmetros das Distribuições



DGI
2019

3.4.1. Parâmetros de Localização

	Variáveis Discretas	Variáveis Contínuas
Valor esperado	$\mu_Y = E(Y) = \sum_{y} y \times p\,(y)$	$\mu_X = E(X) = \int_{-\infty}^{+\infty} x \times f\,(x)\,dx$
Mediana	$F\,[Y = \min\,(Y)] \geq 0.5$ $\eta_Y = \min(Y)$	$\int_{-\infty}^{\eta_X} f\,(x)\;dx = 0.5$
Moda	ξ_Y: valor(es) de Y para o(s) qual(is) $p(y)$ é máximo.	ξ_X:valor(es) de X para o(s) qual(is) $f(x)$ é máximo.

DGI
2019

3.4.2. Parâmetros de Dispersão

	Variáveis Discretas	Variáveis Contínuas
Variância*	$\sigma^2{}_Y = VAR(Y) = \sum_Y (y - \mu_Y)^2 \times p(y)$	$\sigma^2{}_X = VAR(X) = \int_{-\infty}^{+\infty} (x - \mu_X)^2 \times f(x)\, dx$
Desvio Padrão	$\sigma_Y = \sqrt{\sigma^2{}_Y}$	$\sigma_X = \sqrt{\sigma^2{}_X}$

∗ Em alternativa podem utilizar-se as seguintes expressões simplificadas no cálculo da variância de variáveis discretas e contínuas:

$$\sigma^2{}_Y = \sum_Y y^2 \times p(y) - \mu_Y{}^2 \quad \text{e} \quad \sigma^2{}_X = \int_{-\infty}^{+\infty} x^2 \times f(x)\, dx - \mu_X{}^2$$

3.4.3. Outros Parâmetros

	Variáveis Discretas	Variáveis Contínuas
Momentos na Origem	$\mu'_i = \sum_y y^i \times p(y)$	$\mu'_i = \int_{-\infty}^{+\infty} x^i \times f(x)\, dx$
Momentos Centrados	$\mu_i = \sum_y (y - \mu_Y)^i \times p(y)$	$\mu_i = \int_{-\infty}^{+\infty} (x - \mu_X)^i \times f(x)\, dx$
Coeficiente de Assimetria	$\tau = \frac{\sum_Y (y - \mu_Y)^3 \times p(y)}{\sigma_Y{}^3} = \frac{\mu_3}{\sigma^3}$	$\tau = \frac{\int_{-\infty}^{+\infty} (x - \mu_X)^3 \times f(x)\, dx}{\sigma_X{}^3} = \frac{\mu_3}{\sigma^3}$
Coeficiente de Achatamento	$\kappa = \frac{\sum_Y (y - \mu_Y)^4 \times p(y)}{\sigma_Y{}^4} - 3 = \frac{\mu_4}{\sigma^4} - 3$	$\kappa = \frac{\int_{-\infty}^{+\infty} (x - \mu_X)^4 \times f(x)\, dx}{\sigma_X{}^4} - 3 = \frac{\mu_4}{\sigma^4} - 3$

Relações de interdependência entre momentos centrados (μ_i) e momentos na origem (μ'_i)

$$\mu_2 = \mu'_2 - (\mu'_2)^2$$

$$\mu_3 = \mu'_3 - 3 \times \mu'_2 \times \mu'_1 + 2 \times (\mu'_1)^3$$

$$\mu_4 = \mu'_4 - 4 \times \mu'_3 \times \mu'_1 + 6 \times \mu'_2 \times (\mu'_1)^2 - 3 \times (\mu'_1)^4$$

3.5. Variáveis Transformadas

3.5.1. *Transformação Linear*

Seja Z uma variável qualquer (discreta ou contínua) à qual se aplica uma transformação linear.

$W = a + b \times Z$	$\mathrm{E}(W) = a + b \times E(Z)$
	$\mathrm{VAR}(W) = b^2 \times VAR(Z)$

Exemplo

Calcule o valor esperado e a variância da variável Y, obtida a partir de uma outra variável X, de acordo com a seguinte expressão: $Y = 2X + 4$

sendo: $\mu_X = 4 \quad e \quad {\sigma^2}_X = 0.5$

R: $\mu_Y = 12 \quad e \quad {\sigma^2}_Y = 2$

3.5.2. *Transformação Não Linear*

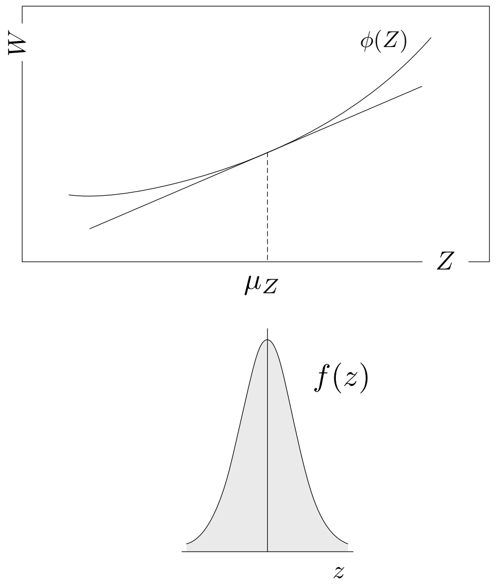
Seja Z uma variável qualquer (discreta ou contínua) à qual se aplica uma transformação não linear.

$W = \phi(Z)$	$E(W) = E[\phi(Z)] \approx \phi[E(Z)]$
$\phi(Z)$ função não linear	$VAR(W) = VAR[\phi(Z)] \approx \left(\frac{d\phi}{dz}\|_{Z=\mu_Z}\right)^2 \times VAR(Z)$

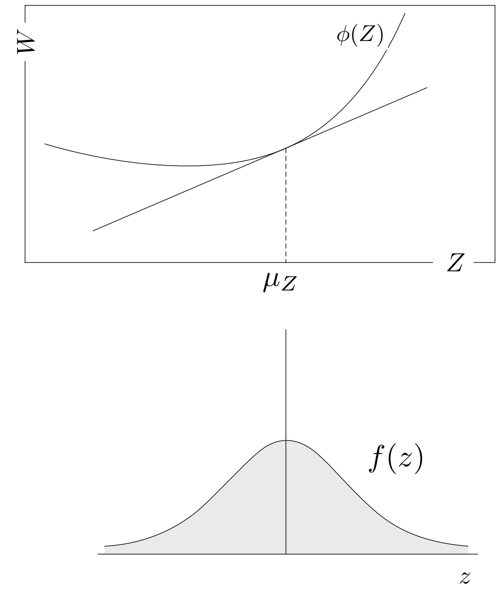
Desenvolvimento em Série de Taylor da função $\phi(Z)$ em torno de $E(Z)$:

$$\phi(Z) \approx \phi[E(Z)] + (Z - E(Z)) \times \left[\frac{\partial \phi}{\partial Z}\right]_{Z=E(Z)}$$

A aproximação será tanto *melhor* quanto *menor* for a *curvatura da função* $\phi(Z)$ *em torno de* μ_Z e quanto *menor for a dispersão da distribuição da variável* Z.



(aproximação linear mais adequada)



(aproximação linear menos adequada)

Exemplo

Considere a variável X com $\mu_X = 4 \quad e \quad \sigma^2{}_X = 0.5$. Calcule o valor esperado e a variância de: $T = 2X^2$

R: $\mu_T = 32 \quad e \quad \sigma^2{}_T = 128$

3.6. Combinação Linear de Variáveis Independentes

Considere-se a variável W definida através de uma combinação linear de duas variáveis independentes X e Y, contínuas ou discretas:

$$W = a + b \times X + c \times Y$$

Os parâmetros da sua distribuição de probabilidade podem ser calculados de duas formas distintas:

- Diretamente a partir da sua distribuição de probabilidade;
- Com base nas distribuições de probabilidade das variáveis que lhe deram origem e na combinação linear dessas variáveis.

$$\mu_W = a + b \times \mu_X + c \times \mu_Y$$

$$\sigma_W^2 = b^2 \times \sigma_X^2 + c^2 \times \sigma_Y^2$$

Generalização

Seja W uma variável aleatória definida a partir de uma combinação linear de N variáveis independentes

$$W = a + b_1 \times X_1 + b_2 \times X_2 + \ldots + b_N \times X_N$$

Os parâmetros da sua distribuição de probabilidade são dados por:

$$\mu_W = a + b_1 \times \mu_{X_1} + b_2 \times \mu_{X_2} + \ldots + b_N \times \mu_{X_N}$$

$$\mu_W = b_1^2 \times \sigma^2_{X_1} + b_2^2 \times \sigma^2_{X_2} + \ldots + b_N^2 \times \sigma^2_{X_N}$$

DGI
2019

3.7. Geração de Observações Aleatórias com recurso à Técnica de Monte Carlo

3.7.1. Introdução

A geração de observações aleatórias envolve duas operações a realizar sequencialmente:

(1) *Geração de números aleatórios* seguindo uma distribuição uniforme $\mathrm{U}(0,\,1)$

- *Método da "semente"*
- *Recurso a tabelas de números aleatórios*

(2) *Transformação daqueles números* noutros igualmente aleatórios, seguindo a distribuição em causa.

3.7.2. Geração de observações que envolvem Variáveis Discretas

- Gerar n números aleatórios Z_i $(i = 1, 2, \ldots, n)$ distribuídos segundo $U(0, 1)$;
- Dividir o intervalo $[0, 1]$ em sub intervalos de dimensões proporcionais aos valores de $p(x)$;
- Fazer corresponder a cada valor gerado Z_i, o valor X_i em conformidade com o intervalo a que Z_i pertença.

Exemplo

Pretende-se gerar uma amostra aleatória de dimensão $n = 8$ a partir de uma população com distribuição definida através da função de probabilidade seguinte:

x	$p(x)$
0	0.25
1	0.50
2	0.25

3.7.3. Geração de observações que envolvem Variáveis Contínuas

- Gerar *n* números aleatórios Z_i $(i = 1, 2, \ldots, n)$ distribuídos segundo $\mathrm{U}(0, 1)$;
- Transformação dos valores recorrendo à inversa de $F(x)$.

Exemplo

Pretende-se gerar uma amostra aleatória de dimensão $n = 3$ a partir de uma população com distribuição exponencial negativa $(\beta = 2)$

4.1. Definição de Distribuição Conjunta

Para um ***mesmo espaço amostral*** (S) ***podem definir-se várias variáveis aleatórias****.* As diferentes variáveis aleatórias possuem ***características próprias que as individualizam*** mas ***podem também apresentar características de natureza conjunta que merecem ser identificadas.***

4.2. Função de Probabilidade Conjunta de Duas Variáveis Discretas

$$p(x,y) = PROB\,(X = x \,\wedge\, Y = y)$$

- $\displaystyle\sum_{X}\sum_{Y} p(x,y) = 1$

DGI
2019

4.3. Função de Probabilidade Conjunta de Duas Variáveis Contínuas

$$f(x,y) = PROB\left\{X \in [x,\ x+dx]\,,\ Y \in [y,\ y+dy]\right\}$$

A probabilidade de um par de valores $(X,\ Y)$ se situar no domínio D é dada por:

$$PROB\left\{(X,Y) \in D\right\} = \int\int_D f(x,y)\,dx\,dy$$

- $\int\int_S f(x,y)\,dx\,dy = 1$

DGI
2019

4.4. Distribuições Marginais

Quando as distribuições $p_X(x)$ e $p_Y(y)$ são obtidas a partir de $p(x,y)$, designam-se por *distribuições marginais*.

4.4.1. *Função de Probabilidade Marginal de Duas Variáveis Discretas*

$$p_X(x) = \sum_Y p(x,y)$$

$$p_Y(y) = \sum_X p(x,y)$$

4.4.2. *Função de Probabilidade Marginal de Duas Variáveis Contínuas*

$$f_X(x) = \int_Y f(x,y)\,dy$$

$$f_Y(y) = \int_X f(x,y)\,dx$$

DGI 2019

Exemplo

Considere duas variáveis aleatórias X e Y cuja função de probabilidade conjunta $p(x,y)$ se apresenta no quadro seguinte:

$X \backslash Y$	1	2	3
2	$2k$	0.18	0.18
3	0.16	0.12	k

1. Determine o valor de k?

$$\sum_x \sum_y p(x,y) = 1 \quad \Leftrightarrow \quad k = 0.12$$

		Y			
		1	2	3	$p_X(x)$
X	2	0.24	0.18	0.18	0.60
	3	0.16	0.12	0.12	0.40
$p_Y(y)$		0.40	0.30	0.30	**1.00**

DGI
2019

Exemplo (cont.)

2. Deduza a função de probabilidade de Y.

$$p_Y(y) = \sum_X p(x, y)$$

$\boldsymbol{y}$	$\boldsymbol{p(y)}$
1	0.4
2	0.3
3	0.3
$\sum$	1

DGI
2019

4.5. Distribuições Condicionais

4.5.1. *Função de Probabilidade Condicional de duas Variáveis Discretas*

$$p_{X|Y}(x|y) = \frac{p(x,y)}{P_Y(y)}$$

$$p_{Y|X}(y|x) = \frac{p(x,y)}{P_X(x)}$$

4.5.2. *Função de Probabilidade Condicional de duas Variáveis Contínuas*

$$f_{X|Y}(x|y) = \frac{f(x,y)}{\int_X f(x,y)\,dx}$$

$$f_{Y|X}(y|x) = \frac{f(x,y)}{\int_Y f(x,y)\,dy}$$

DGI
2019

4.6. Independência

Sejam X e Y duas variáveis aleatórias discretas. Tais variáveis dizem-se ***independentes se e só se, para todos os valores de X e de Y, os acontecimentos***:

$$X = x \quad \text{e} \quad Y = y \quad \text{forem independentes}$$

Tal implica que:

$$PROB\,(X = x|Y = y) = PROB\,(X = x) \quad ,\forall_{x,y}$$

$$PROB\,(Y = y|X = x) = PROB\,(Y = y) \quad ,\forall_{x,y}$$

donde

$$p_{X|Y}(x|y) = \frac{p(x,y)}{P_Y(y)} = p_X(x), \quad \forall_{x,y}$$

$$p_{Y|X}(y|x) = \frac{p(x,y)}{P_X(x)} = p_Y(y), \quad \forall_{x,y}$$

ou

$$p(x,y) = p_X(x) \times p_Y(y), \quad \forall_{x,y}$$

O conceito de independência pode estender-se ao caso de variáveis contínuas. Se X e Y forem *duas variáveis contínuas*, elas são *independentes* se e só se:

$$\forall_{x,y} : f(x,y) = f_X(x) \times f_Y(y)$$

Exemplo (cont.)

3. Verifique se as variáveis são independentes.

As variáveis são independentes se:

$$\forall_{x,y}\ p(x,y) = p(x) \times p(y)$$

		Y			
		1	2	3	$p_X(x)$
X	2	0.24	0.18	0.18	0.6
	3	0.16	0.12	0.12	0.4
$p_Y(y)$		0.4	0.3	0.3	**1**

$0.24 = 0.6 \times 0.4$ ✓	$0.18 = 0.6 \times 0.3$ ✓	$0.18 = 0.6 \times 0.3$ ✓
$0.16 = 0.4 \times 0.4$ ✓	$0.12 = 0.4 \times 0.3$ ✓	$0.12 = 0.4 \times 0.3$ ✓

$\rightarrow$ são independentes

4.7. Covariância e Correlação

Se duas variáveis são independentes então, entre elas, não existe qualquer relação.

É de esperar que as seguintes variáveis sejam independentes:

- *Cor dos olhos dos professores do IPB*
- *Disciplinas que leccionam*
- No que diz respeito às variáveis altura e peso dos professores, é de esperar que elas não sejam independentes. Embora possa haver excepções, é de esperar que o peso tenda a aumentar com o aumento da altura.

Existem **várias possibilidades de relacionamento** entre o peso e a altura que satisfaçam a condição enunciada (relação linear, quadrática, exponencial, etc.). De todas elas a *mais simples é a linear, que pode ser utilizada como aproximação das restantes.* **A Covariância e a Correlação são medidas do grau de relacionamento linear entre duas variáveis.**

4.7.1. Covariância

$$\sigma_{xy} = \sum_X \sum_Y (x-\mu_X)(y-\mu_Y) \times p(x,y) \qquad \textbf{(variáveis discretas)}$$

$$\sigma_{xy} = \int\int (x-\mu_X)(y-\mu_Y) \times f(x,y)\, dx\, dy \qquad \textbf{(variáveis contínuas)}$$

De uma forma geral, tanto para as variáveis discretas ou contínuas verifica-se que: $\sigma_{xy} = \mathrm{E}\left[(x-\mu_X)(y-\mu_Y)\right]$

4.7.2. Correlação

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \frac{\mathrm{E}\left[(x-\mu_X)(y-\mu_Y)\right]}{\sqrt{\mathrm{E}\left[(x-\mu_X)^2\right] \times \mathrm{E}\left[(y-\mu_Y)^2\right]}}$$

DGI

2019

A correlação pode ser interpretada como a covariância das variáveis padronizadas:

$$X^* = \frac{x - \mu_X}{\sigma_X} \quad \text{e} \quad Y^* = \frac{y - \mu_Y}{\sigma_Y}$$

É de esperar que as seguintes variáveis sejam independentes:

- O coeficiente de correlação ρ_{XY} toma valores compreendidos entre -1 e 1

$$-1 \leq \rho_{xy} \leq 1$$

ρ_{xy}	**Interpretação**
1	Relacionamento linear perfeito *(directo)*
0	Ausência de relação linear
-1	Relacionamento linear perfeito *(inverso)*

Se $\varrho_{XY} = 0$ não se pode concluir que X e Y sejam independentes. $\rho_{XY} = 0$ significa apenas que não existe uma relação linear entre X e Y. Poderão existir relações não lineares entre elas.

Covariância positiva: quando uma das variáveis se desvia significativamente do seu valor esperado a outra também tenderá a desviar-se no mesmo sentido. Tal acarretará um *aumento da dispersão da soma das duas variáveis.*

Covariância negativa: Os desvios das duas variáveis serão mais frequentemente de sentido contrário, implicando uma *diminuição da dispersão da sua soma*.

DGI
2019

4.8. Valor Esperado e Variância de uma Função de várias Variáveis Aleatórias

Considerem-se duas variáveis aleatórias quaisquer, discretas ou contínuas (X e Y)

e defina-se uma nova variável aleatória Z como função das anteriores

$$Z = g(X,\ Y)$$

Considere-se o caso: $Z = g(X,\ Y) = X + Y$

Neste caso, o *valor esperado e a variância de Z* vêm dados por:

$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y) = \mu_X + \mu_Y$$

$$\mathrm{VAR}(X+Y) = \mathrm{VAR}(X) + 2 \times \mathrm{COV}(X,Y) + \mathrm{VAR}(Y) = {\sigma^2}_X + 2 \times \sigma_{XY} + {\sigma^2}_Y$$

Exemplo (cont.)

4. Considere uma nova variável $Z = X\text{-}Y$. Calcule a sua variância e o seu valor esperado

$$Z = X - Y$$

$$E(Z) = E(X) - E(Y) = \mu_X - \mu_Y = 0.5$$

$$VAR(Z) = VAR(X - Y) = VAR(X) + 2 \times COV(X, Y) + VAR(Y) =$$

X, Y independentes $\rightarrow$ $= VAR(X) + VAR(Y) =$

$$= 0.24 + 0.69 = 0.93$$

Quando X e Y forem não correlacionadas (*independentes*):

$$\mathrm{VAR}(X+Y) = \mathrm{VAR}(X) + \mathrm{VAR}(Y) = {\sigma^2}_X + {\sigma^2}_Y$$

Quando X e Y forem *perfeitamente correlacionadas*:

$$\frac{\sigma_{XY}}{\sigma_X \times \sigma_Y} = 1 \quad \Rightarrow \quad \sigma_{XY} = \sigma_X \times \sigma_Y$$

Considere-se agora o caso em que a variável $Z = g(X, Y)$ é uma *combinação linear de* X *e* Y*:*

$$Z = g(X, Y) = a \times X + b \times Y$$

Neste caso, o *valor esperado* e a *variância* de *Z* vêm dados por:

$$\mathrm{E}(a \times X + b \times Y) = a \times \mathrm{E}(X) + b \times \mathrm{E}(Y) = a \times \mu_X + b \times \mu_Y$$

$$\begin{aligned}\mathrm{VAR}(a \times X + b \times Y) &= a^2 \times \mathrm{VAR}(X) + 2 \times a \times b \times \mathrm{COV}(X, Y) + b^2 \times \mathrm{VAR}(Y) = \\ &= a^2 \times {\sigma^2}_X + 2 \times a \times b \times \sigma_{XY} + b^2 \times {\sigma^2}_Y\end{aligned}$$

Considere-se finalmente o caso em que Z é uma função não linear em X e/ou em Y. Neste caso é possível obter *aproximações de $E(Z)$ e $Var(Z)$ a partir de uma aproximação linear de $g(X,\ Y)$ em torno de (μ_X, μ_Y):*

$$g(X,\ Y) \approx g(\mu_X, \mu_Y) + (X - \mu_X) \cdot \left.\frac{\partial g}{\partial X}\right|_{\substack{X=\mu_X \\ Y=\mu_Y}} + (Y - \mu_Y) \cdot \left.\frac{\partial g}{\partial Y}\right|_{\substack{X=\mu_X \\ Y=\mu_Y}}$$

Neste caso o *valor esperado* e a *variância* de Z vêm dados por:

$$\mathrm{E}\left[g(X,Y)\right] \approx g\left[\mathrm{E}(X), \mathrm{E}(Y)\right] = g(\mu_X, \mu_Y)$$

$$\mathrm{VAR}\left[g(X,Y)\right] \approx \left(\left.\frac{\partial g}{\partial X}\right|_{\substack{X=\mu_X \\ Y=\mu_Y}}\right)^2 \times \sigma^2{}_X +$$

$$+ 2 \times \left(\left.\frac{\partial g}{\partial X}\right|_{\substack{X=\mu_X \\ Y=\mu_Y}}\right)\left(\left.\frac{\partial g}{\partial Y}\right|_{\substack{X=\mu_X \\ Y=\mu_Y}} +\right) \times \sigma_{XY} + \left(\left.\frac{\partial g}{\partial Y}\right|_{\substack{X=\mu_X \\ Y=\mu_Y}}\right)^2 \times \sigma^2{}_Y$$

5.1. Distribuição Binomial [$\mathrm{B}(n, p)$]

5.1.1. Introdução

Os processos mais simples de recolha de dados consistem na ***contagem do número de vezes que um determinado acontecimento ocorre*** ao longo de uma série de experiências (ex. contagem de peças defeituosas num posto de inspecção).

Frequentemente tais experiências apresentam as seguintes características:

- A cada experiência tem apenas *dois resultados possíveis* (sucesso ou insucesso).
- A *probabilidade de ocorrência* de cada resultado mantém-se *inalterada* de experiência para experiência
- Os *resultados* associados a diferentes experiências são *independentes* (a probabilidade de sucesso p não é afectada pelo possível conhecimento dos resultados obtidos até então).

Experiências nestas condições dizem-se ***experiências de Bernoulli***.

Uma variável aleatória binomial Y representa o número de vezes que, no decurso de n experiências de Bernoulli, ocorre um dos resultados possíveis. A variável aleatória Y pode tomar os valores 0, 1, 2, ..., n.

$$Y \rightsquigarrow \mathrm{B}(n, p)$$

- Trata-se de uma distribuição muito utilizada em amostragem, em situações em que se conhece a dimensão da amostra e se sabe quantas vezes um acontecimento ocorreu.
- A Distribuição Binomial está para as distribuições discretas de probabilidade, assim como a distribuição normal está para as distribuições contínuas.

5.1.2. *Função de Probabilidade*

$$p(y) = \binom{n}{y} \times p^y \times q^{(n-y)} = C_y^n \times p^y \times q^{(n-y)}$$

$p(y)$ representa a probabilidade de, em n tentativas, se obterem y sucessos e $(n$ - $y)$ insucessos.

Exemplo

De uma linha de produção, retiram-se 3 peças (cada uma de um turno diferente). Admitindo que, para todos os turnos, a percentagem de peças defeituosas é de 10%, calcular a probabilidade de entre as três peças recolhidas haver 0, 1, 2 ou 3 peças defeituosas.

sendo: Y– número de peças defeituosas de entre as três seleccionadas

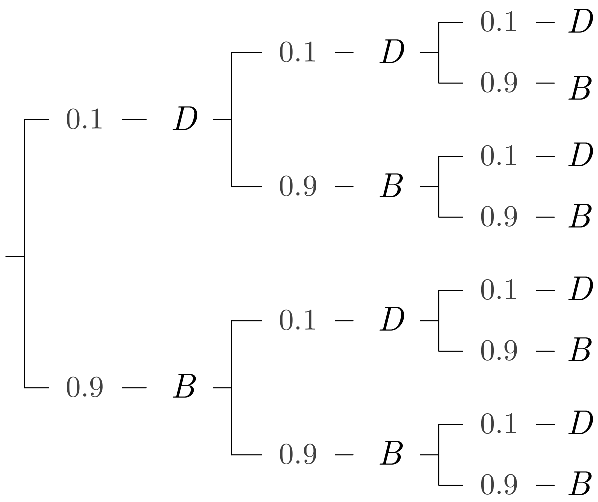
p – probabilidade de saída de peça defeituosa *(sucesso)*

q – probabilidade de saída de peça não defeituosa

representação

dos acontecimentos

(diagrama em árvore)



Função de Probabilidade

$y=0$	$p(0)=1\times(0.1)^0\times(0.9)^3=0.7290$
$y=1$	$p(1)=3\times(0.1)^1\times(0.9)^2=0.2430$
$y=2$	$p(2)=3\times(0.1)^2\times(0.9)^1=0.0270$
$y=3$	$p(3)=1\times(0.1)^3\times(0.9)^0=0.0010$

$$p(y)=\mathrm{C}_y^n\times p^y\times q^{(n-y)}$$

$$\downarrow$$

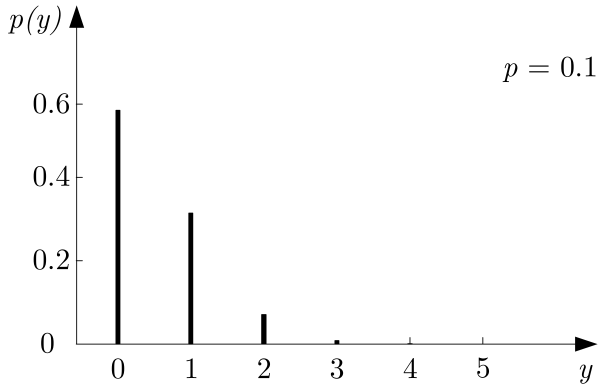
$$Y \rightsquigarrow B(n=3, p=0.1)$$

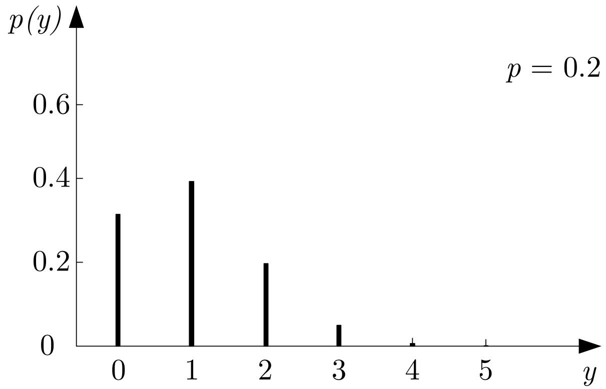
5.1.3. *Simetria*

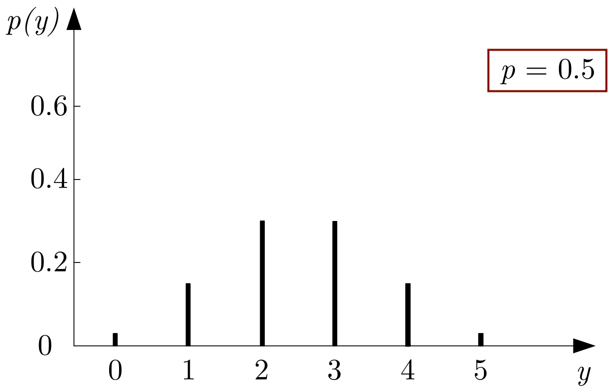
Considere-se a distribuição associada à variável aleatória Y:

$$Y \to B(n = 5,\ p)$$

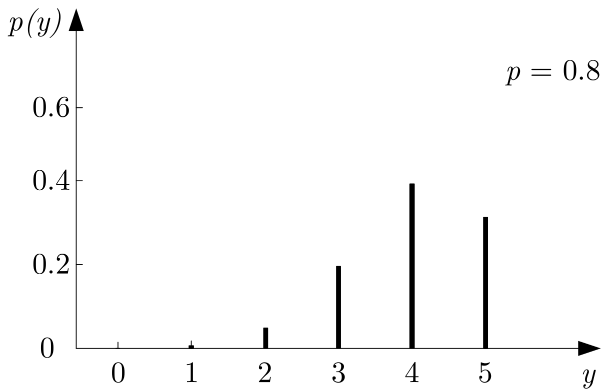
y	$p(y)$				
	$p = 0.1$	$p = 0.2$	$p = 0.5$	$p = 0.8$	$p = 0.9$
0	0.5905	0.3277	0.0313	0.0003	0
1	0.3281	0.4096	0.1563	0.0064	0.0005
2	0.0729	0.2048	0.3125	0.0512	0.0081
3	0.0081	0.0512	0.3125	0.2048	0.0729
4	0.0005	0.0064	0.1563	0.4096	0.3281
5	0	0.0003	0.0313	0.3277	0.5905
$\sum$	1	1	1	1	1

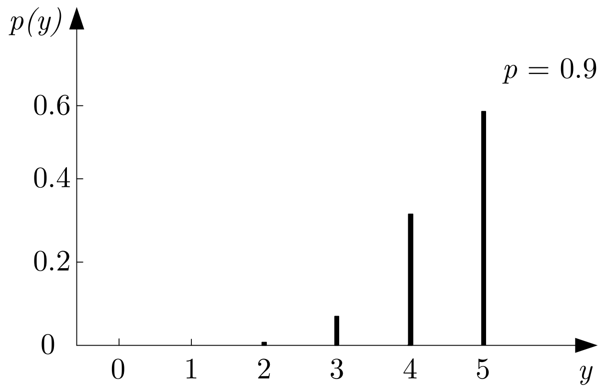






$p\left(y\right)$ é simétrica se $p = 0.5$





5.1.4. Parâmetros

Valor esperado	$\mu = n \times p$
Variância	$\sigma^2 = n \times p \times q$
Desvio Padrão	$\sigma = \sqrt{n \times p \times q}$

DGI
2019

5.1.5. *Valores Tabelados*

Os valores tabelados correspondem às probabilidades

$$p(y) = \mathrm{C}_y^n \times p^y \times q^{(n-y)}$$

Exemplo

$Y \rightsquigarrow B(n = 3,\ p = 0.1)$

$n = 2$

y		p	**0.05 (0.95)**	**0.10 (0.90)**	**0.15 (0.85)**	**0.20 (0.80)**	**0.25 (0.75)**	**0.30 (0.70)**	**0.35 (0.65)**	**0.40 (0.60)**	**0.45 (0.55)**	**0.50 (0.50)**
0	**(2)**		.9025	.8100	.7225	.6400	.5625	.4900	.4225	.3600	.3025	.2500
1	**(1)**		.0950	.1800	.2550	.3200	.3750	.4200	.4550	.4800	.4950	.5000
2	**(0)**		.0025	.0100	.0225	.0400	.0625	.0900	.1225	.1600	.2025	.2500

$\boxed{n = 3}$

y		p	**0.05 (0.95)**	**0.10 (0.90)**	**0.15 (0.85)**	**0.20 (0.80)**	**0.25 (0.75)**	**0.30 (0.70)**	**0.35 (0.65)**	**0.40 (0.60)**	**0.45 (0.55)**	**0.50 (0.50)**
0	**(3)**		.8574	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664	.1250
1	**(2)**		.1354	.2430	.3251	.3840	.4219	.4410	.4436	.4320	.4084	.3750
2	**(1)**		.0071	.0270	.0574	.0960	.1406	.1890	.2389	.2880	.3341	.3750
3	**(0)**		.0001	.0010	.0034	.0080	.0156	.0270	.0429	.0640	.0911	.1250

5.2. Distribuição Binomial Negativa [$\mathrm{BN}(r,p)$]

5.2.1. *Introdução*

Esta distribuição permite descrever o comportamento de uma variável aleatória que, embora associada à *repetição de experiências de Bernoulli*, ***envolve um processo de contagem distinto daquele que é descrito pela distribuição Binomial***.

Numa *sequência infinita de experiências de Bernoulli*, a variável Y seguirá uma distribuição Binomial Negativa se representar ***o número de insucessos até ocorrer o r-ésimo sucesso (inclusive)***.

$$Y \rightsquigarrow \mathrm{BN}(r,p)$$

5.2.2. *Função de Probabilidade*

$$p(y) = \mathrm{C}_y^{y+r-1} \times p^r \times q^y$$

$p(y)$ representa a probabilidade de ocorrerem y insucessos até ocorrer o r-ésimo sucesso.

Definição da distribuição Binomial Negativa a partir da distribuição Binomial

Y – número de insucessos até ao r-ésimo sucesso

p – probabilidade de sucesso

Número total de tentativas $= Y + r$

$$\underbrace{\boxed{}\boxed{}\boxed{\;\cdots\;}\boxed{}\boxed{}}_{y+r-1} + \underset{\substack{1 \\ \text{sucesso}}}{\boxed{}}$$

$$\begin{pmatrix} r-1 & \text{sucessos} \\ y & \text{insucessos} \end{pmatrix}$$

$\Rightarrow$

$$p(y) = \left(C_y^{y+r-1} \times p^{r-1} \times q^y\right) \times p$$

$$= C_y^{y+r-1} \times p^r \times q^y$$

5.2.3. Parâmetros

Valor esperado	$\mu = \dfrac{r \times q}{p}$
Variância	$\sigma^2 = \dfrac{r \times q}{p^2}$
Desvio Padrão	$\sigma = \sqrt{\dfrac{r \times q}{p^2}}$

5.2.4. Caso Particular - Distribuição Geométrica ($r = 1$) $[\ \mathrm{G}(p)\ ]$

Y representa o número de experiências de Bernoulli realizadas até à ocorrência do primeiro sucesso (inclusive).

$$Y \rightsquigarrow \mathrm{G}(p)$$

5.2.4.1. Função de Probabilidade

$$\underbrace{\boxed{}\boxed{}\boxed{\ \cdots\ }\boxed{}\boxed{}}_{y \text{ insucessos}} + \underset{\substack{1 \\ \text{sucesso}}}{\boxed{}} \quad \rightarrow \quad p(y) = q^y \times p = p \times q^y$$

5.2.4.2. Parâmetros

Valor esperado	$\mu = \dfrac{q}{p}$
Variância	$\sigma^2 = \dfrac{q}{p^2}$

5.3. Distribuição Hipergeométrica $[\ \mathrm{H}(M \times p, M \times q, n)\ ]$

5.3.1. *Introdução*

Considere-se uma ***população finita constituída por M elementos de dois tipos*** (ex. peças defeituosas e peças boas). Sejam p e $q = 1 - p$ as proporções nas quais os dois tipos de elementos figuram na população. Nestas condições teremos:

$M \times p$ elementos de um tipo (ex. peças defeituosas)

$M \times q$ elementos do outro tipo (ex. peças não defeituosas)

Admitindo que se retiram sucessivamente e sem reposição, n elementos da população.

- Uma variável aleatória Hipergeométrica Y representa o número de vezes que, no decurso das n experiências sucessivas, ocorre um dos dois resultados possíveis. Y pode tomar os valores 0, 1, 2, ..., n.

A diferença fundamental entre esta situação e a que estava na base da distribuição Binomial é que ***as experiências não são de Bernoulli***. De facto dado que a população é finita e que se tiram elementos sem reposição, os resultado de cada experiência passará a estar dependente dos resultados obtidos nas experiências anteriores. Dessa forma, ***os resultados das sucessivas experiências deixam de ser independentes***.

5.3.2. Função de Probabilidade

$$p(y) = \frac{\mathrm{C}_{y}^{M\times p} \times \mathrm{C}_{n-y}^{M\times q}}{\mathrm{C}_{n}^{M}}$$

$p(y)$ representa a probabilidade de, em n tentativas, se obterem y sucessos e n-y insucessos.

No conjunto dos $\boldsymbol{n}$ resultados, o *número de casos possíveis* é:

$$\mathrm{C}_n^M$$

Dessas combinações, o número daquelas que contêm exactamente $\boldsymbol{y}$ elementos do primeiro tipo e consequentemente, $\boldsymbol{n}$ **-** $\boldsymbol{y}$ elementos do segundo tipo é:

$$\mathrm{C}_y^{M\times p} \times \mathrm{C}_{n-y}^{M\times q}$$

Dado que ***as diferentes combinações de elementos da população são equiprováveis***, a função de probabilidade $\boldsymbol{p(y)}$ vem dada por:

$$p(y) = \frac{\text{n}^{\text{o}} \text{ casos favoráveis}}{\text{n}^{\text{o}} \text{ casos possíveis}} = \frac{\mathrm{C}_y^{M\times p} \times \mathrm{C}_{n-y}^{M\times q}}{\mathrm{C}_n^M}$$

5.3.3. Parâmetros

Valor esperado	$\mu = n \times p$
Variância	$\sigma^2 = n \times p \times q \times \dfrac{M-n}{M-1}$
Desvio Padrão	$\sigma = \sqrt{n \times p \times q \times \dfrac{M-n}{M-1}}$

5.3.4. Relação entre as Distribuições Hipergeométrica e Binomial

Quando $M >> n$, o facto de não haver reposição dos elementos retirados ***não afecta substancialmente as probabilidades*** associadas aos resultados das experiências subsequentes. Dessa forma:

$$\mathrm{H}(M \times p, M \times q, n) \approx \mathrm{B}(n, p)$$

Critério de aproximação (regra prática): $M \geq 10 \times n$

Exemplo

y	$p(y)$			
	H(10,90,10)	B(10,0.1)	H(10,90,5)	B(5,0.1)
0	0.330	0.349	0.584	0.590
1	0.408	0.387	0.339	0.328
2	0.202	0.194	0.070	0.073
3	0.052	0.057	0.006	0.008
4	0.008	0.011	0.000	0.000
5	0.001	0.002	0.000	0.000
$\sum$	1.000	1.000	1.000	1.000
μ	1.000	1.000	0.500	0.500
σ^2	0.818	0.900	0.432	0.450

5.4. Distribuição de Poisson [$\mathrm{P}(\lambda)$]

5.4.1. *Introdução*

A distribuição de Poisson tem na sua génese um processo de Poisson e permite descrever certos tipos de fenómenos ou ***acontecimentos cuja ocorrência se repete no tempo ou no espaço***.

Exemplos:

- O número de golos marcados num jogo de futebol
- O número de avarias de uma máquina durante um dia de trabalho
- O número de erros ortográficos existentes num livro
- O número de defeitos de isolamento ao longo de um cabo eléctrico.

Para que a distribuição de Poisson tenha aplicação é suficiente que se verifiquem as seguintes condições:

- Os acontecimentos ocorrem um a um e não em grupos.
- A probabilidade de um acontecimento ocorrer num intervalo (de tempo ou de espaço) infinitamente pequeno é proporcional ao comprimento do intervalo e é independente do número de ocorrências verificadas em intervalos anteriores (***processo sem memória***).

Uma variável aleatória de Poisson Y representa o ***número ocorrências de um dado evento num intervalo temporal ou numa região espacial***. Y pode tomar os valores

$$\{0, 1, 2, 3, \ldots\} = \mathbb{N}_0$$

5.4.2. Função de Probabilidade

$$p(y) = \frac{e^{-\lambda} \times \lambda^y}{y!}$$

$\boldsymbol{p(y)}$ representa a probabilidade de se verificarem $\boldsymbol{y}$ ocorrências de um dado evento no intervalo de espaço/tempo considerado

5.4.3. Parâmetros

Valor esperado	$\mu = \lambda$
Variância	$\sigma^2 = \lambda$
Desvio Padrão	$\sigma = \sqrt{\lambda}$

DGI
2019

5.4.4. *Valores Tabelados*

Os valores correspondem às probabilidades acumuladas: $F(y) = \sum_{u=0}^{y} \frac{e^{-\lambda} \times \lambda^u}{u!}$

λ

y ↓	*0.10*	*0.20*	*0.30*	*0.40*	*0.50*	*0.60*	*0.70*	*0.80*	*0.90*	*1.00*
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	0.3679
1	0.9953	0.9825	0.9631	0.9384	0.9098	0.8781	0.8442	0.8088	0.7725	0.7358
2	0.9998	0.9989	0.9964	0.9921	0.9856	0.9769	0.9659	0.9526	0.9371	0.9197
3	1.0000	0.9999	0.9997	0.9992	0.9982	0.9966	0.9942	0.9909	0.9865	0.9810
4	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9986	0.9977	0.9963
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9994
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999

Exemplo

$\lambda = 1$

$P(y=3) = ?$

$$P(y=3) = \frac{e^{-\lambda} \times \lambda^y}{y!} = \frac{e^{-1} \times 1^3}{3!} = 0.0613$$ *(da função de probabilidade)*

$$P(y=3) = F(3) - F(2) = 0.9810 - 0.9197 = 0.0613$$ *(da tabela)*

5.4.5. Relação entre a Distribuição de Poisson e a Distribuição Binomial

Na prática, interessa aproximar a distribuição Binomial pela de Poisson pois *o cálculo da função de probabilidade desta última é mais fácil*.

Critério de aproximação (regra prática):

- A *dimensão* da amostra é *elevada* $(n \geq 20)$
- A *distribuição* binomial é *assimétrica* $(n \times p < 7)$

Se a distribuição Binomial for simétrica, torna-se mais prático aproximá-la à distribuição normal (*distribuição simétrica*).

Se uma distribuição Hipergeométrica poder ser aproximada por uma distribuição Binomial e esta, por uma distribuição de Poisson, então deve recorrer-se a essa última distribuição para calcular $p(y)$.

Exemplo

y	$p(y)$			
	B(10,0.1)	B(50,0.02)	B(100,0.01)	P(1)
0	0.349	0.364	0.366	0.368
1	0.387	0.372	0.370	0.368
2	0.194	0.186	0.185	0.184
3	0.057	0.061	0.061	0.061
4	0.011	0.015	0.015	0.015
5	0.002	0.003	0.003	0.003

6.1. Distribuição Uniforme [$\mathrm{U}(a,b)$]

6.1.1. *Introdução*

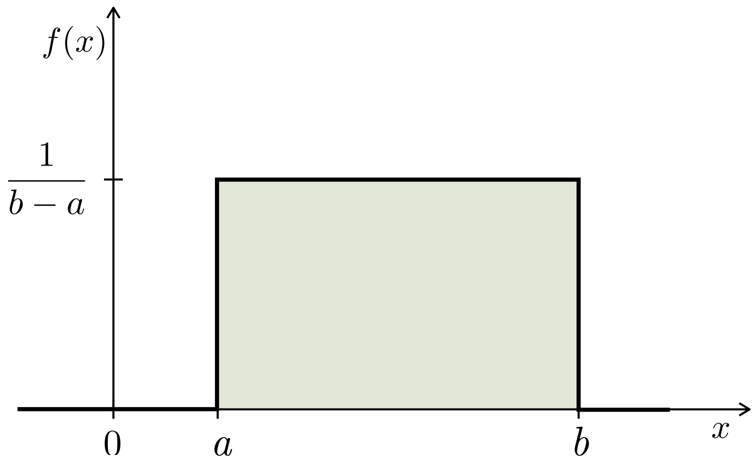
- *A distribuição Uniforme é a mais simples* das distribuições contínuas e uma das mais importantes.
- Utiliza-se para representar uma quantidade que varia aleatoriamente num intervalo $[a,\ b]$ e cuja probabilidade de tomar valores num qualquer sub intervalo de $[a,\ b]$ é proporcional ao seu comprimento.

Uma variável aleatória contínua X segue uma distribuição Uniforme quando a sua ***função densidade de probabilidade é constante (positiva) dentro de um intervalo finito [a, b] e nula fora desse intervalo***.

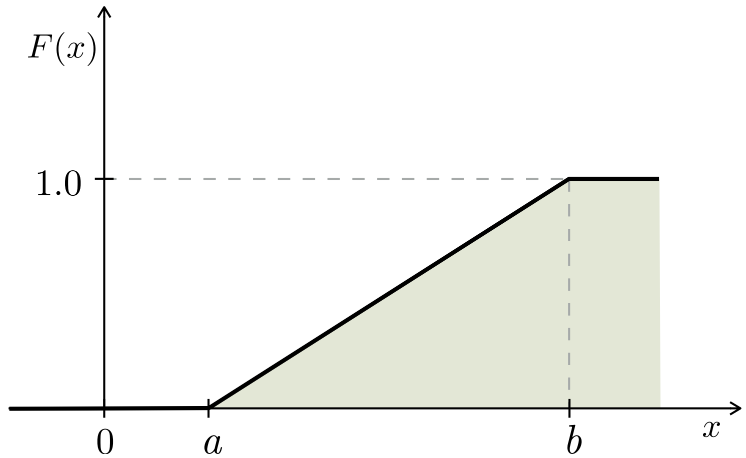
$$X \rightsquigarrow U(a,\ b) \qquad\qquad f(x) = \begin{cases} 0, & x < a \\ k, & a \le x \le b \\ 0, & x > b \end{cases}$$

6.1.2. Função Densidade e Distribuição de Probabilidade

$$f(x) = \begin{cases} 0, & x < a \\ \dfrac{1}{b-a}, & a \le x \le b \\ 0, & x > b \end{cases}$$



$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ 1, & x > b \end{cases}$$



6.1.3. Parâmetros

Valor esperado	$\mu = \dfrac{a+b}{2}$
Variância	$\sigma^2 = \dfrac{(b-a)^2}{12}$

Exemplo

Suponha que selecciona um número no pertencente ao intervalo $[5,\ 7]$. Determine a probabilidade de esse número se encontrar compreendido entre:

(a) 6.3 e 6.95

(b) 6.5 e 7.5

6.2. Distribuição Exponencial Negativa [$\mathrm{EN}(\beta)$]

6.2.1. Introdução

Existe uma relação estreita entre a distribuição Exponencial Negativa e a distribuição de Poisson

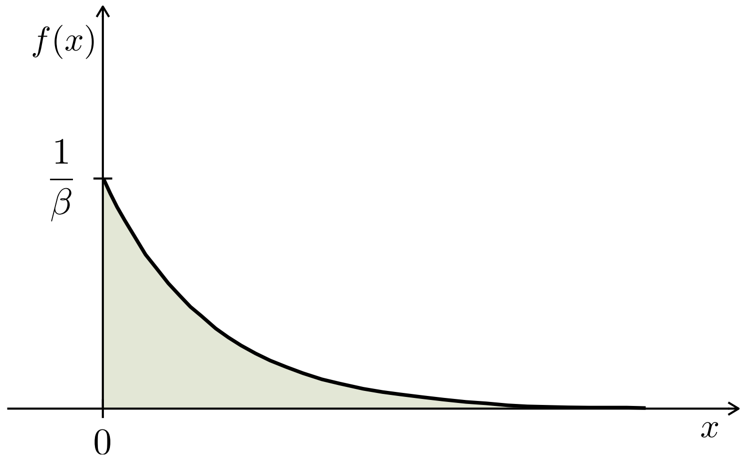
- Admitindo que λ é o parâmetro da distribuição de Poisson que ***descreve o número de ocorrências por unidade de tempo (ou de espaço)***, *a variável tempo (ou distância) entre ocorrências sucessivas segue uma distribuição Exponencial Negativa* de parâmetro $\beta = 1/\lambda$.

$$X \rightsquigarrow \mathrm{EN}(\beta)$$

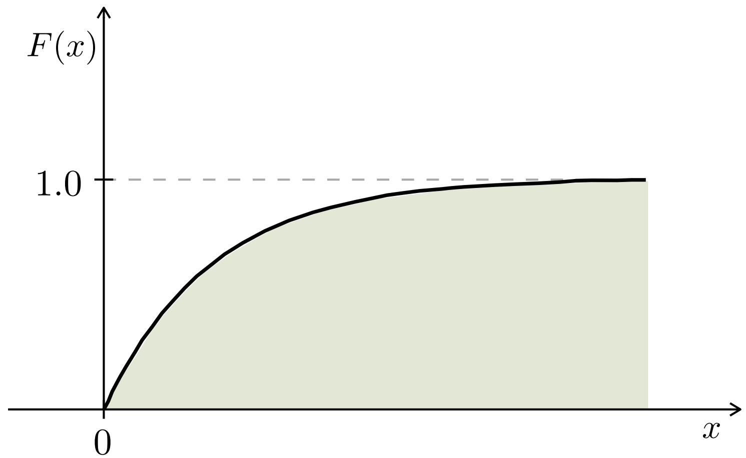
- Trata-se de uma das distribuições contínuas mais importantes e é muito utilizada no estudo de filas de espera e de fiabilidade de sistemas complexos

6.2.2. Função Densidade e Distribuição de Probabilidade

$$f(x) = \frac{1}{\beta} \times e^{-\frac{x}{\beta}}$$



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



DGI

2019

6.2.3. Parâmetros

Valor esperado	$\mu = \beta = \dfrac{1}{\lambda}$
Variância	$\sigma^2 = \beta^2 = \left(\dfrac{1}{\lambda}\right)^2$

Exemplo

O tempo entre duas chamadas consecutivas que chegam a uma central telefónica segue uma distribuição exponencial cuja média é de 2 segundos.

(a) Qual a probabilidade de a central não receber chamadas nos próximos 5 segundos?

(b) Qual a probabilidade de a central receber pelo menos 2 chamadas nos próximos 5 segundos?

6.3. Distribuição Normal $\left[\, \mathrm{N}(\mu, \sigma^2) \,\right]$

6.3.1. *Introdução*

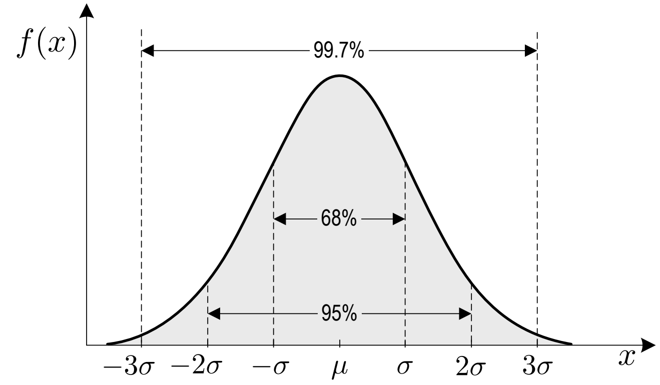
A distribuição Normal é, pela sua forma e pelas suas propriedades, aquela que mais frequentemente é utilizada para *descrever fenómenos físicos contínuos* sendo muito utilizada, como veremos em *Inferência estatística*.

A distribuição Normal fica completamente especificada através:

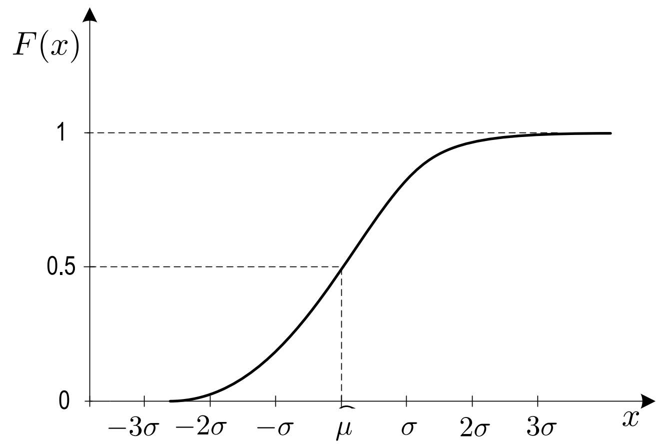
- Do seu valor esperado, μ
- Da sua variância, σ^2
- Da sua função densidade de probabilidade, $f(x)$

6.3.2. Função Densidade e Distribuição de Probabilidade

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



$$F(x) = \int_{-\infty}^{x} f(u)\,du$$



A função distribuição de probabilidade não é integrável analiticamente, só podendo ser definida pela via da integração numérica.

6.3.3. *Distribuição Normal Reduzida* [$Z(0,1)$]

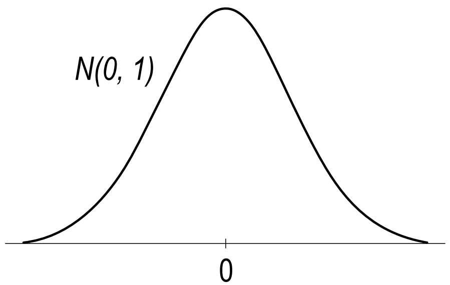
Se uma variável X segue uma distribuição $N(\mu,\ \sigma^2)$ então a variável

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

Segue uma distribuição Normal com valor esperado nulo e variância unitária

$$Z \rightsquigarrow N\left(\mu = 0,\ \sigma^2 = 1\right)$$

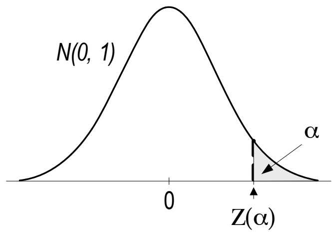
A passagem de uma variável normal para uma variável normal reduzida pode ser entendida como uma ***dupla transformação***, correspondendo à translação da origem de μ para 0 e à mudança de escala executada através do divisor σ.



$$f(x) = \frac{1}{\sqrt{2\pi \times \sigma^2}} \times e^{-\frac{z^2}{2}}$$

6.3.4. ***Valores Tabelados***

Os valores tabelados correspondem à área assinalada na Figura:



$$P\left[Z \geq z(\alpha)\right] = \alpha$$

$Z(\alpha) = a + b$

a ↓ b →	***0.00***	***0.01***	***0.02***	***0.03***	***0.04***	***0.05***	***0.06***	***0.07***	***0.08***	***0.09***
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559

Exemplo

Considere a variável aleatória $X \rightsquigarrow N\left(\mu = 100,\ \sigma^2 = 100\right)$

Determine:

(a) $P\left(X > 95\right)$

(b) $P\left(95 < X < 110\right)$

DGI
2019

6.3.5. Relação entre a Distribuição Normal e a Distribuição Binomial

Uma distribuição Binomial $\mathbf{B}(\boldsymbol{n},\ \boldsymbol{p})$ na qual:

- A *dimensão* da amostra $(\boldsymbol{n})$ é e*levada* $(n \geq 20)$
- $\boldsymbol{p}$ tome um valor suficientemente grande para que a *distribuição seja simétrica* $(n \times p \geq 7)$

Pode ser aproximada por uma distribuição Normal com:

$$\mathrm{B}\,(n,\ p) \approx \mathrm{N}\left(\mu = n \times p,\ \sigma^2 = n \times p \times q\right)$$

6.3.6. *Aproximação entre as Diferentes Distribuições*

$$
\begin{tikzcd}[column sep=huge, row sep=huge]
\mathrm{H}\left(M \times p, M \times q, n\right) \arrow[r, "M \geq 10 \times n"] \arrow[dr, "\begin{array}{l} n \geq 20 \\ n \times p \geq 7 \end{array}"'] & \mathrm{B}\left(n,\ p\right) \arrow[r, "n \geq 20", "n \times p < 7"'] \arrow[d, "\begin{array}{l} n \geq 20 \\ n \times p \geq 7 \end{array}"] & \mathrm{P}\left(\lambda\right) \arrow[dl, "\lambda \geq 30"] \\
& \mathrm{N}\left(\mu,\ \sigma^2\right) &
\end{tikzcd}
$$

Hipergeométrica	Binomial	Poisson
$\mu = n \times p$	$\mu = n \times p$	$\mu = \lambda$
$\sigma^2 = n \times p \times q \times \dfrac{M-n}{M-1}$	$\sigma^2 = n \times p \times q$	$\sigma^2 = \lambda$

DGI
2019

6.4. Distribuição Qui-Quadrado $\left[\, \chi^2\,(GL) \,\right]$

6.4.1. Definição

Considerem-se GL variáveis aleatórias, Z_i ,mutuamente independentes, seguindo todas elas a distribuição normal padronizada:

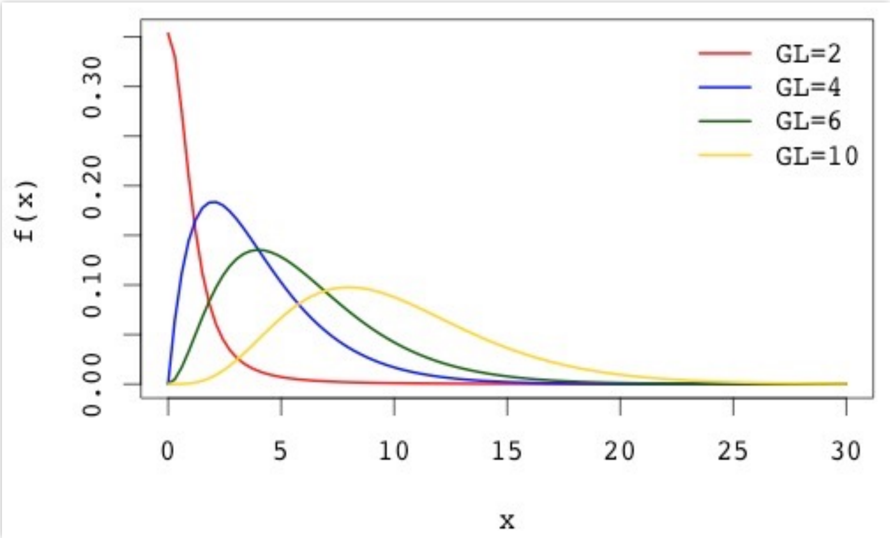
$$Z_i \rightsquigarrow \mathrm{IN}(0,1), \quad i = 1, ..., GL$$

A variável aleatória

$$X = \sum_{i=1}^{GL} \mathrm{Z}_i^2 \rightsquigarrow \chi^2_{GL}$$

com: $\mu = GL$

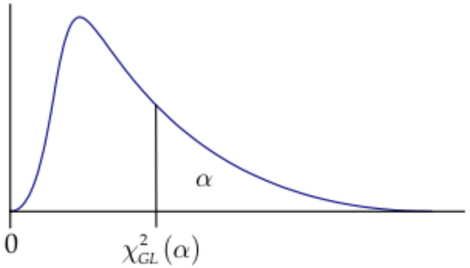
$\sigma^2 = 2 \times GL$



6.4.2. *Valores Tabelados da Distribuição Qui-Quadrado*

Valores Críticos de Distribuições χ^2_{GL}

Os valores tabelados correspondem a $\chi^2_{GL}(\alpha)$, tais que, $P\left(\chi^2_{GL} \geq \chi^2_{GL}(\alpha)\right) = \alpha$.



GL	α												
	0.995	**0.990**	**0.975**	**0.950**	**0.900**	**0.750**	**0.500**	**0.250**	**0.100**	**0.050**	**0.025**	**0.010**	**0.005**
1	0.000	0.000	0.001	0.004	0.016	0.102	0.455	1.323	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	0.575	1.386	2.773	4.605	5.991	7.378	9.210	10.60
3	0.072	0.115	0.216	0.352	0.584	1.213	2.366	4.108	6.251	7.815	9.348	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	1.923	3.357	5.385	7.779	9.488	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.610	2.675	4.351	6.626	9.236	11.07	12.83	15.09	16.75
6	0.676	0.872	1.237	1.635	2.204	3.455	5.348	7.841	10.64	12.59	14.45	16.81	18.55
7	0.989	1.239	1.690	2.167	2.833	4.255	6.346	9.037	12.02	14.07	16.01	18.48	20.28
8	1.344	1.646	2.180	2.733	3.490	5.071	7.344	10.22	13.36	15.51	17.53	20.09	21.95
9	1.735	2.088	2.700	3.325	4.168	5.899	8.343	11.39	14.68	16.92	19.02	21.67	23.59
10	2.156	2.558	3.247	3.940	4.865	6.737	9.342	12.55	15.99	18.31	20.48	23.21	25.19
11	2.603	3.053	3.816	4.575	5.578	7.584	10.34	13.70	17.28	19.68	21.92	24.72	26.76
80	51.17	53.54	57.15	60.39	64.28	71.14	79.33	88.13	96.58	101.9	106.6	112.3	116.3
90	59.20	61.75	65.65	69.13	73.29	80.62	89.33	98.65	107.6	113.1	118.1	124.1	128.3
100	67.33	70.06	74.22	77.93	82.36	90.13	99.33	109.1	118.5	124.3	129.6	135.8	140.2

Extrato daTabela de Valores Críticos da Distribuição Qui-Quadrado (Tabelas: pág. 16)

6.5. Distribuição *t-Student* $[\,\mathrm{t}\,(GL)\,]$

6.5.1. *Definição*

Sejam Z e V, duas variáveis aleatórias independentes e:

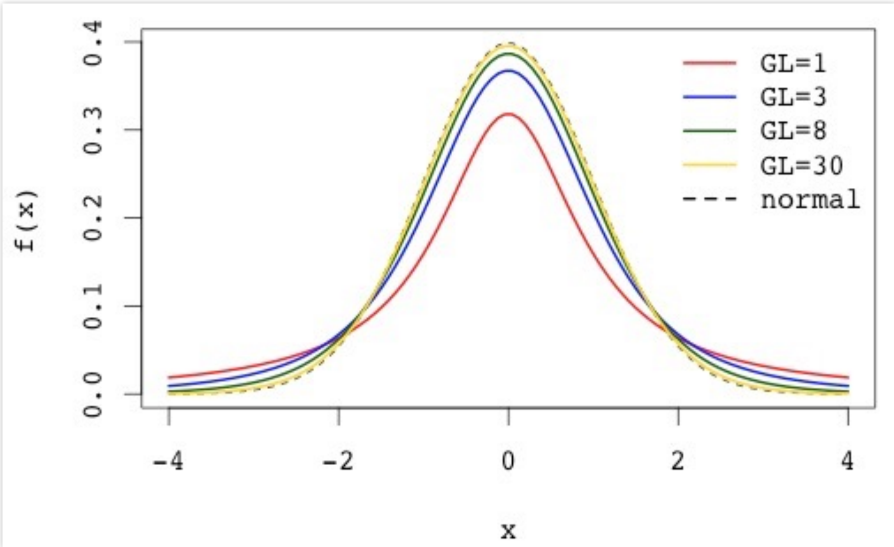
$$Z \rightsquigarrow \mathrm{N}(0,1) \quad \text{e} \quad V \rightsquigarrow \chi^2_{GL}$$

A variável aleatória

$$X = \frac{Z}{\sqrt{V/GL}} \rightsquigarrow t_{GL}$$

com: $\mu = 0$

$$\sigma^2 = \frac{GL}{GL-2}$$

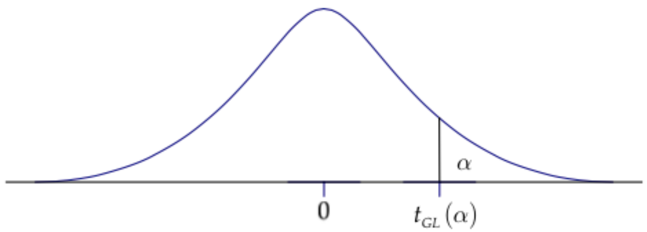


DGI
2019

6.5.2. *Valores Tabelados da Distribuição t-Student*

Valores Críticos de Distribuições t_{GL}

Os valores tabelados correspondem a $t_{GL}(\alpha)$, tais que, $P\left(t_{GL} \geq t_{GL}(\alpha)\right) = \alpha$.



GL \ α	0.25	0.20	0.15	0.10	0.05	0.025	0.010	0.005	0.001	0.0005
1	1.0000	1.3764	1.9626	3.0777	6.3138	12.706	31.821	63.657	318.31	636.62
2	0.8165	1.0607	1.3862	1.8856	2.9200	4.3027	6.9646	9.9248	22.327	31.599
3	0.7649	0.9785	1.2498	1.6377	2.3534	3.1824	4.5407	5.8409	10.215	12.924
4	0.7407	0.9410	1.1896	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732	8.6103
5	0.7267	0.9195	1.1558	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934	6.8688
6	0.7176	0.9057	1.1342	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076	5.9588
7	0.7111	0.8960	1.1192	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853	5.4079
8	0.7064	0.8889	1.1081	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0413
9	0.7027	0.8834	1.0997	1.3830	1.8331	2.2622	2.8214	3.2498	4.2968	4.7809
10	0.6998	0.8791	1.0931	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5869
11	0.6974	0.8755	1.0877	1.3634	1.7959	2.2010	2.7181	3.1058	4.0247	4.4370
60	0.6786	0.8477	1.0455	1.2958	1.6706	2.0003	2.3901	2.6603	3.2317	3.4602
120	0.6765	0.8446	1.0409	1.2886	1.6577	1.9799	2.3578	2.6174	3.1595	3.3735
∞	0.6745	0.8416	1.0364	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902	3.2905

Extrato da Tabela de Valores Críticos da Distribuição t_{GL} (Tabelas: pág. 16)

6.6. Distribuição F-*Fisher* $[\,\mathrm{F}\,(GL_1, GL_2)\,]$

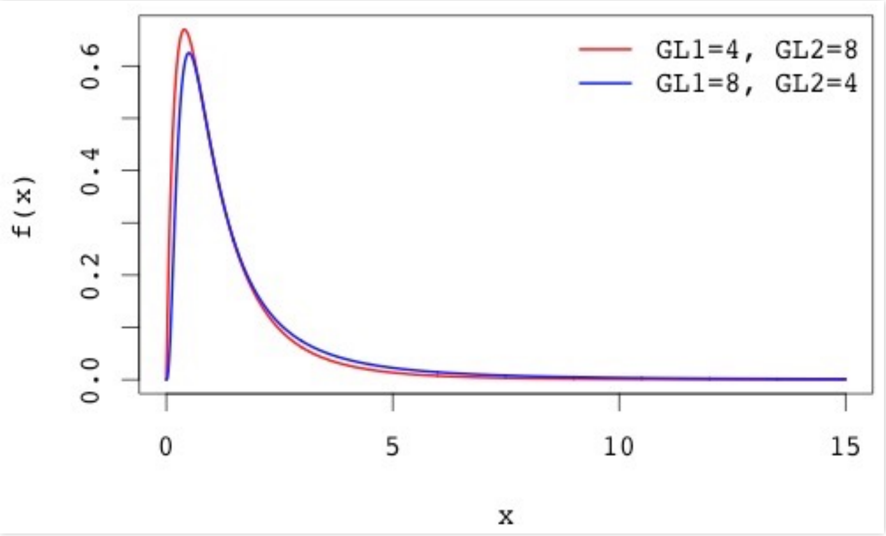
6.6.1. *Definição*

Sejam X_1 e X_2, duas variáveis aleatórias independentes, caracterizadas por distribuições do Qui-Quadrado:

$$X_1 \rightsquigarrow \chi^2_{GL_1} \quad \text{e} \quad X_2 \rightsquigarrow \chi^2_{GL_2}$$

A variável aleatória

$$X = \frac{X_1/GL_1}{X_2/GL_2} \rightsquigarrow \mathrm{F}_{GL_1,GL_2}$$



com: $\mu = GL_2 GL_2 - 2, \quad GL_2 > 2$

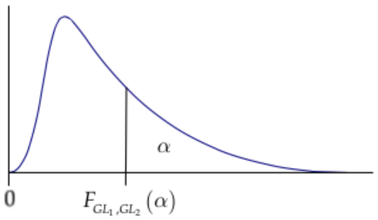
$$\sigma^2 = \frac{2 \times GL_2 \times (GL_1 + GL_2 - 2)}{GL_1 \times (GL_2 - 2)^2 \times (GL_2 - 4)}, \quad GL_2 > 4$$

DGI
2019

6.6.2. Valores Tableados da Distribuição F-Fisher

Valores Críticos de Distribuições $F_{GL1,GL2}$

Os valores tabelados correspondem a $F_{GL_1,GL_2}(\alpha)$, tais que, $P\left(F_{GL_1,GL_2} \geq F_{GL_1,GL_2}(\alpha)\right) = \alpha$.



A seguinte equivalência poderá ser útil:

$$F_{GL_1,GL_2}(\alpha) = \frac{1}{F_{GL_2,GL_1}(1-\alpha)}$$

α = 10%

GL₂	GL₁																						
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**14**	**16**	**18**	**20**	**25**	**30**	**40**	**50**	**100**	**150**	**200**
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.47	60.71	61.07	61.35	61.57	61.74	62.05	62.26	62.53	62.69	63.01	63.11	63.17
2	8.526	9.000	9.162	9.243	9.293	9.326	9.349	9.367	9.381	9.392	9.401	9.408	9.420	9.429	9.436	9.441	9.451	9.458	9.466	9.471	9.481	9.485	9.486
3	5.538	5.462	5.391	5.343	5.309	5.285	5.266	5.252	5.240	5.230	5.222	5.216	5.205	5.196	5.190	5.184	5.175	5.168	5.160	5.155	5.144	5.141	5.139
4	4.545	4.325	4.191	4.107	4.051	4.010	3.979	3.955	3.936	3.920	3.907	3.896	3.878	3.864	3.853	3.844	3.828	3.817	3.804	3.795	3.778	3.772	3.769
5	4.060	3.780	3.619	3.520	3.453	3.405	3.368	3.339	3.316	3.297	3.282	3.268	3.247	3.230	3.217	3.207	3.187	3.174	3.157	3.147	3.126	3.119	3.116
6	3.776	3.463	3.289	3.181	3.108	3.055	3.014	2.983	2.958	2.937	2.920	2.905	2.881	2.863	2.848	2.836	2.815	2.800	2.781	2.770	2.746	2.738	2.734
7	3.589	3.257	3.074	2.961	2.883	2.827	2.785	2.752	2.725	2.703	2.684	2.668	2.643	2.623	2.607	2.595	2.571	2.555	2.535	2.523	2.497	2.488	2.484
8	3.458	3.113	2.924	2.806	2.726	2.668	2.624	2.589	2.561	2.538	2.519	2.502	2.475	2.455	2.438	2.425	2.400	2.383	2.361	2.348	2.321	2.312	2.307
9	3.360	3.006	2.813	2.693	2.611	2.551	2.505	2.469	2.440	2.416	2.396	2.379	2.351	2.329	2.312	2.298	2.272	2.255	2.232	2.218	2.189	2.179	2.174
10	3.285	2.924	2.728	2.605	2.522	2.461	2.414	2.377	2.347	2.323	2.302	2.284	2.255	2.233	2.215	2.201	2.174	2.155	2.132	2.117	2.087	2.077	2.071
11	3.225	2.860	2.660	2.536	2.451	2.389	2.342	2.304	2.274	2.248	2.227	2.209	2.179	2.156	2.138	2.123	2.095	2.076	2.052	2.036	2.005	1.994	1.989
25	2.918	2.528	2.317	2.184	2.092	2.024	1.971	1.929	1.895	1.866	1.841	1.820	1.785	1.758	1.736	1.718	1.683	1.659	1.627	1.607	1.565	1.549	1.542
26	2.909	2.519	2.307	2.174	2.082	2.014	1.961	1.919	1.884	1.855	1.830	1.809	1.774	1.747	1.724	1.706	1.671	1.647	1.615	1.594	1.551	1.536	1.528
27	2.901	2.511	2.299	2.165	2.073	2.005	1.952	1.909	1.874	1.845	1.820	1.799	1.764	1.736	1.714	1.695	1.660	1.636	1.603	1.583	1.539	1.523	1.515
28	2.894	2.503	2.291	2.157	2.064	1.996	1.943	1.900	1.865	1.836	1.811	1.790	1.754	1.726	1.704	1.685	1.650	1.625	1.592	1.572	1.528	1.512	1.504
29	2.887	2.495	2.283	2.149	2.057	1.988	1.935	1.892	1.857	1.827	1.802	1.781	1.745	1.717	1.695	1.676	1.640	1.616	1.583	1.562	1.517	1.501	1.493
30	2.881	2.489	2.276	2.142	2.049	1.980	1.927	1.884	1.849	1.819	1.794	1.773	1.737	1.709	1.686	1.667	1.632	1.606	1.573	1.552	1.507	1.491	1.482

Extrato daTabela de Valores Críticos da Distribuição F-Fisher (Tabelas: págs. 19 e 20)

7.1. População, Amostra e Inferência Estatística

O objectivo fundamental da estatística é inferir as características de uma população, a partir de resultados observados numa ou mais amostras extraídas dessa população.

- ***O processo de obtenção/extracção de amostras é designado de amostragem.***

Exemplo

Tirar conclusões sobre as alturas (ou pesos) dos estudantes do IPB *(população)* observando apenas um grupo de 100 estudantes *(amostra)* seleccionados na população.

7.2. Amostragem Aleatória

Para se poderem tirar ***conclusões representativas*** acerca de uma população a partir da análise de uma amostra, a ***recolha da amostra a utilizar deve ser feita de acordo com certas regras***.

Considere-se uma população constituída por M elementos (ex. alunos do IPB) e a variável aleatória X que representa uma característica da população a estudar (ex. altura).

- Para obter uma amostra de tamanho $\boldsymbol{n}$ $(n < M)$ deve escolher-se aleatoriamente um indivíduo da população. Dessa forma, esse indivíduo (altura) pode tomar qualquer valor entre todas as alturas possíveis.

x_1 – altura do 1º indivíduo seleccionado $\rightarrow$ valor da variável aleatória X_1
x_2 – altura do 2º indivíduo seleccionado $\rightarrow$ valor da variável aleatória X_2
x_3 – altura do 3º indivíduo seleccionado $\rightarrow$ valor da variável aleatória X_3
$\vdots$ $\vdots$
x_n – altura do n indivíduo seleccionado $\rightarrow$ valor da variável aleatória X_n

- ***O processo de amostragem diz-se aleatório*** (e as amostras aleatórias) ***quando as funções de probabilidade de X_1, X_2, ..., X_n*** *são idênticas à função* $p(x)$*:*

$$\forall_X : p(x_1) = p(x_2) = \ldots = p(x_n) = p(x)$$

- A justificação básica para para se utilizar a ***amostragem aleatória*** é assegurar que *as* ***inferências*** feitas a partir dos dados amostrais ***não são distorcidas por um enviezamento na selecção***. Tal enviezamento existirá sempre que ocorrer uma tendência sistemática para ***sobrerepresentar ou subrepresentar*** uma parte da população.

Como obter uma amostra aleatória de dimensão n?

(1) Numerar de 1 até M todos os elementos da população;

(2) Colocar bolas numeradas de 1 até M numa urna;

(3) Retirar uma bola, seleccionar o elemento da população com o número correspondente e registar a sua altura;

(4) Repor a bola na urna;

(5) Repetir os passos (3) e (4) até terem sido registadas as alturas de n elementos.

- Quando a *amostragem é feita com reposição as variáveis* X_1, X_2, ..., X_n *são independentes.*

- Um processo de amostragem aleatória com estas características diz-se *aleatório simples*, verificando-se:

$$p_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = p(x_1) \times p(x_2) \times \ldots \times p(x_n)$$

- Quando $M \to \infty$ (e a dimensão da amostra, n, se mantém finita), a dependência entre as variáveis X_1, X_2, ..., X_n tende a desaparecer. ***Quando a população é infinita, é indiferente realizar a amostragem aleatória com ou sem reposição*** (a amostragem é aleatória simples em qualquer dos casos).

7.3. Distribuições Amostrais

parâmetros populacionais $\neq$ estatísticas amostrais

Para uma dada população e uma variável aleatória sobre ela definida, ***os parâmetros da distribuição correspondente são fixos enquanto que as medidas estatísticas variam de amostra para amostra***.

- Uma ***estatística amostral*** calculada a partir de X_1, X_2, ..., X_n é uma função de várias variáveis aleatórias e é, portanto, ***ela própria uma variável aleatória***. ***A distribuição de probabilidade de uma estatística amostral designa-se distribuição amostral da estatística***.

- Naturalmente, podemos calcular, para uma distribuição amostral, a média, a variância, desvio padrão, etc.

- ***O conceito de distribuição amostral de estatísticas será abordado considerando apenas o caso de estas serem obtidas com base em amostragens aleatórias.***

7.3.1. Distribuição da Média Amostral

Valor esperado	$E\left(\bar{X}\right) = \mu_{\bar{X}} = \mu_X$	
Variância	$VAR\left(\bar{X}\right) = \sigma^2_{\bar{X}} = \dfrac{\sigma^2_X}{n}$	*(população infinita ou finita com reposição)*
	$VAR\left(\bar{X}\right) = \sigma^2_{\bar{X}} = \dfrac{\sigma^2_X}{n} \times \left(\dfrac{M-n}{M-1}\right)$	*(população finita, sem reposição)*

Justificação:

$$E\left(\bar{X}\right) = E\left(\frac{1}{n} \times \sum_{i=1}^{n} X_i\right) = \frac{1}{n} \times E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n} \times E\left(X_1 + X_2 + \ldots + X_n\right) =$$

$$= \frac{1}{n} \times \left[E\left(X_1\right) + E\left(X_2\right) + \ldots + E\left(X_n\right)\right] = \frac{1}{n} \times \sum_{i=1}^{n} E\left(X\right) = \frac{1}{n} \times \left[n \times E\left(X\right)\right] = E\left(X\right)$$

$$Var\left(\bar{X}\right) = Var\left(\frac{1}{n} \times \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \times Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \times Var\left(X_1 + X_2 + \ldots + X_n\right) =$$

$$= \frac{1}{n^2} \times \left[Var\left(X_1\right) + Var\left(X_2\right) + \ldots + Var\left(X_n\right)\right] =$$

$$= \frac{1}{n^2} \times \sum_{i=1}^{n} Var\left(X\right) = \frac{1}{n^2} \times \left[n \times Var\left(X\right)\right] = \frac{Var\left(X\right)}{n}$$ *(população infinita ou finita com reposição)*

Exemplo

Considere-se uma população com quatro elementos aos quais se associam os seguintes valores da variável aleatória X: $\{2,\ 4,\ 6,\ 6\}$

(a) Qual a função de probabilidade $p(x)$?

(b) Defina a função de probabilidade de $\bar{X}$ e a média amostral, calculadas para amostras aleatórias de dimensão 2 (amostras sem reposição).

Forma da Distribuição da Média Amostral quando a População é Normal

Quando a população é normal, a distribuição da média amostral também é normal.

(1) A média amostral é uma combinação linear das variáveis X_i.

(2) Se a população for normal, então

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

é uma combinação linear de variáveis normais.

(3) Qualquer combinação linear de variáveis normais é, ela própria, uma variável normal.

- Sendo a *população normal*, presume, em rigor, que a *população é infinita*. Nesta situação:

$$\bar{X} \to N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$$ *população infinita*

- Se se admitir que a *população é finita mas de grande dimensão*, com uma *distribuição aproximadamente normal*, e que a *amostra é também de grande dimensão*, então na definição de $\bar{X}$ deve-se entrar em *consideração com o factor de redução*, vindo:

$$\bar{X} \to N\left(\mu_X, \frac{\sigma_X^2}{n} \times \frac{M-n}{M-1}\right)$$ *população finita e amostra de grande dimensão*

7.4. Teorema do Limite Central

(1) Sejam X_1, X_2, ..., X_n ***variáveis independentes com distribuições idênticas*** (com valor esperado μ_X e variância σ_X^2 finitos).

- Qualquer que seja a forma da distribuição comum aos X_i's, a distribuição da variável soma

$$S = \sum_{i=1}^{n} X_i$$

tende para uma distribuição normal à medida que $\boldsymbol{n}$ aumenta.

- *Qualquer que seja a distribuição de uma população* (com valor esperado μ_X e variância σ_X^2 finitos), então *a distribuição da média amostral*

$$\bar{X} = \frac{1}{n} \times S$$

tende para uma distribuição normal quando a dimensão da amostra, $\boldsymbol{n}$, tende para infinito.

(presupõe-se que a amostragem é aleatória simples, para que as variáveis X_i sejam independentes)

(2) As condições apresentadas em (1) podem ser relaxadas de duas formas:

- As variáveis X_1, X_2, ..., X_n ***podem ter distribuições diferentes*** umas das outras (desde que se verifique a condição de que ***a contribuição da variância de cada uma delas para a variância da soma seja relativamente pequena***).

- As variáveis X_1, X_2, ..., X_n ***podem ser dependentes desde que a correlação seja fraca***.

Questão: Quando é que o valor de n é suficientemente grande para que a distribuição Normal seja efectivamente uma boa aproximação da média amostral?

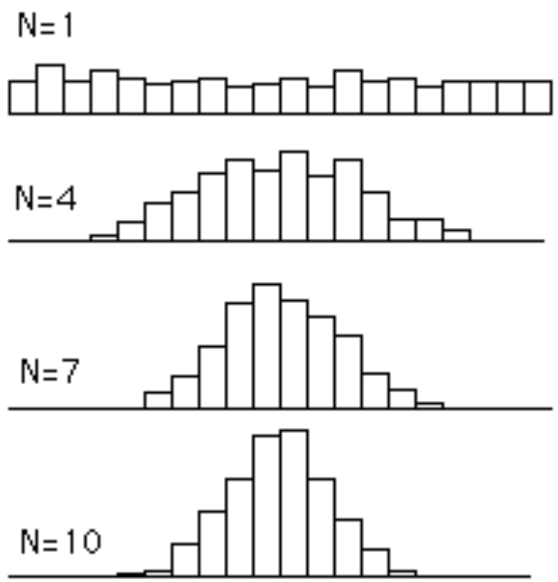
R: Para um dado valor de n, ***o rigor da aproximação depende da forma da distribuição original sendo, em particular, tanto menor quanto maior for a assimetria desta***.

Regra Prática:	$n \geq 10$	*se a distribuição original for simétrica*
	$n \geq 50$	*se a distribuição original for muito assimétrica*

DGI
2019

Teorema do Limite Central

Para uma qualquer população (caracterizada por uma qualquer distribuição) com variância finita, a distribuição da média amostral calculada com base numa amostra aleatória simples tende para uma distribuição normal, à medida que a dimensão da amostra aumenta.



Exemplo

O custo de transporte de um lote de determinada matéria-prima depende da localização do fornecedor. Sabe-se que o custo é caracterizado por uma distribuição normal com valor médio igual a 575 u.m. e desvio padrão igual a 100 u.m..

a) Se forem adquiridos 10 lotes da referida matéria-prima a diferentes fornecedores, qual a média e a variância do custo médio desses lotes? Qual a probabilidade desse custo ultrapassar 6000 u.m.?

b) Determine a probabilidade do custo médio dos 10 lotes adquiridos ser superior a 600 u.m..

8.1. Estimadores

No âmbito da inferência estatística, ao calcular uma estatística (por exemplo uma média amostral) o objectivo passa a ser o de ***caracterizar*** um conjunto mais vasto de dados: a ***população a partir da qual a amostra foi retirada***.

- A ***forma mais directa*** de se proceder a uma inferência estatística corresponde à ***estimação pontual de parâmetros***.
- ***Designa-se por estimador (pontual) de um parâmetro uma estatística cujo valor constitui uma estimativa do parâmetro em causa.***

Exemplos

A média amostral $\left(\bar{X}\right)$ é um estimador do valor esperado (μ_X) da população

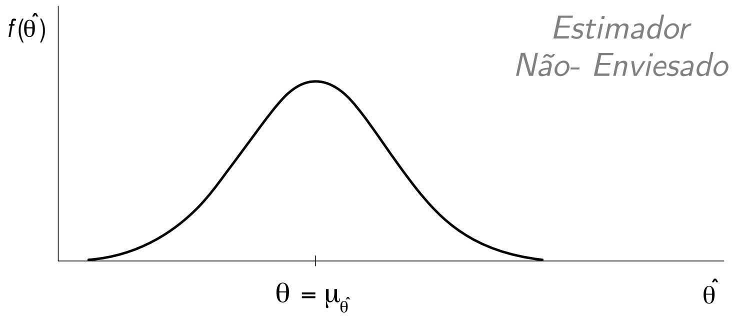
O erro quadrático médio (EQM_X) é um estimador da variância $\left(\sigma_X^2\right)$ da população

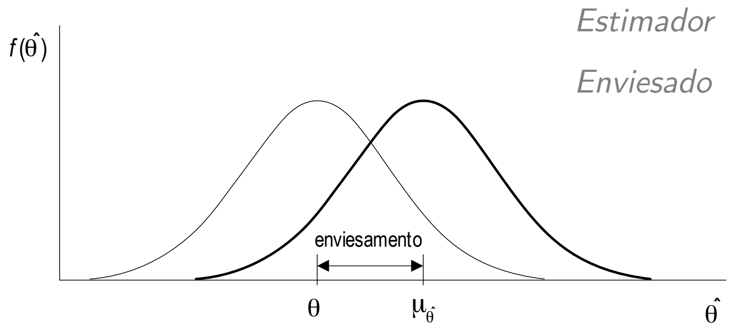
8.2. Propriedades Desejáveis dos Estimadores

8.2.1. ***Enviesamento***

Seja θ *um parâmetro qualquer de uma população e* $\hat{\theta}$ *um estimador desse parâmetro.*

- *O estimador diz-se não* ***enviesado quando o seu valor esperado coincidir com o valor do parâmetro***.





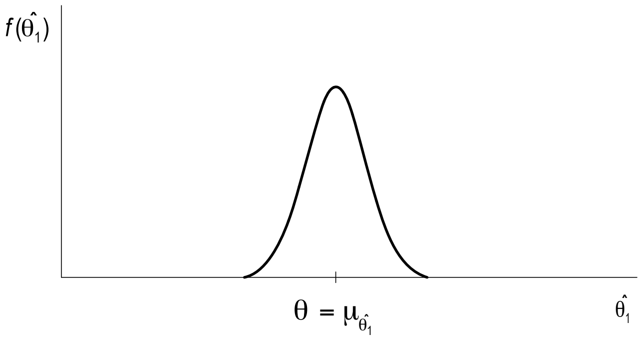
$$\mathrm{E}\left[\hat{\theta}\right] = \theta \qquad\qquad \mathrm{Enviesamento} = \mathrm{E}\left[\hat{\theta}\right] - \theta$$

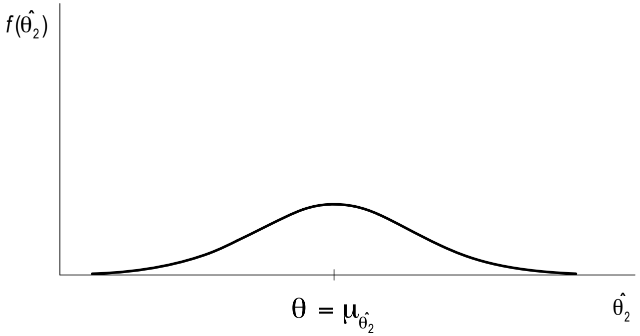
- *O estimador será enviesado se a diferença* $\mathrm{E}\left(\hat{\theta}\right) - \theta$ *for maior ou menor do que zero*. Tal diferença designa-se por enviesamento.

DGI
2019

8.2.2. ***Eficiência***

Considere-se o caso de dois estimadores não enviesados $(\hat{\theta}_1,\ \hat{\theta}_2)$ de um mesmo parâmetro θ. Admita-se que, para amostras de dimensão n, as variâncias daqueles estimadores são $\sigma^2{}_1$ e $\sigma^2{}_2$, com $\sigma^2{}_1 < \sigma^2{}_2$.





- Neste caso, o estimador $\hat{\theta}_1$ é melhor, a sua distribuição é mais concentrada em torno de θ ou seja tem uma menor variância – $\hat{\theta}_1$ é pois um estimador mais eficiente.

$$\text{Eficiência relativa entre } \hat{\theta}_1 \text{ e } \hat{\theta}_1 = \frac{VAR(\hat{\theta}_1)}{VAR(\hat{\theta}_2)}$$

- *Quando **dois estimadores são não enviesados, o melhor estimador é aquele que tem menor variância.***

DGI

2019

Exemplo

Com base numa amostra de duas observações, considere os dois estimadores seguintes para a média:

$$U = \frac{1}{2}X_1 + \frac{1}{2}X_2$$

$$V = \frac{1}{3}X_1 + \frac{2}{3}X_2$$

(a) U e V são estimadores enviesados?

(b) Qual o estimador mais eficiente?

Generalize-se agora a comparação entre estimadores, admitindo que estes podem ser enviesados.

A medida da eficiência do estimador $\hat{\theta}$ será o desvio quadrático médio (DQM), dado por:

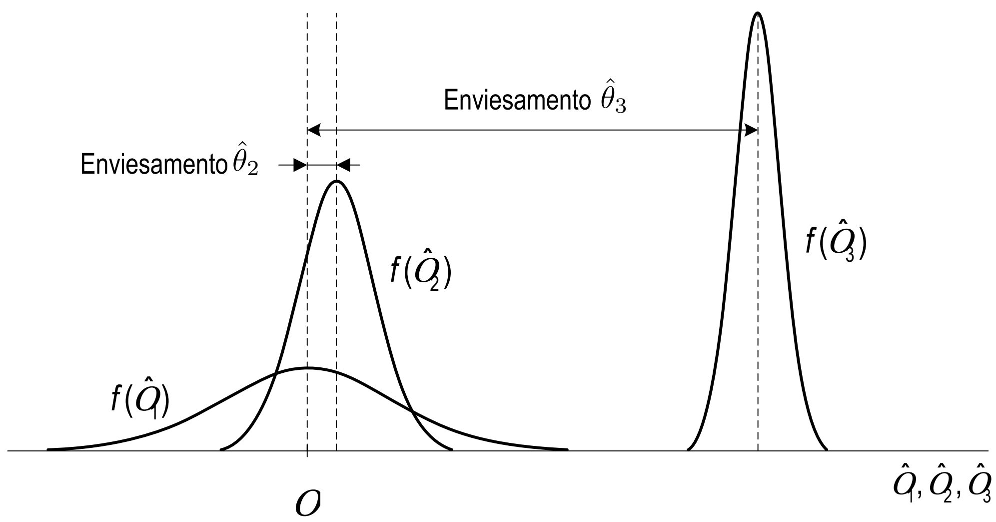
$$DQM = E\left[(\hat{\theta} - \theta)^2\right] = \sigma_{\hat{\theta}}^2 + (\text{enviesamento})^2$$

- Para quaisquer dois estimadores, enviesados ou não enviesados:

$$\text{Eficiencia relativa de } \theta_1 \text{e } \theta_2 = \frac{DQM(\theta_1)}{DQM(\theta_2)}$$

DGI
2019

Considere-se o exemplo da figura. Dos três estimadores $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$, o de menor variância $\hat{\theta}_3$ está longe de ser o melhor, dado que é muito mais enviesado do que os outros dois. O melhor também não será o que apresenta enviesamento nulo $(\hat{\theta}_1)$, dado que tem uma variância muito grande. Restará possivelmente $\hat{\theta}_2$ como o melhor compromisso.



- Se existir um estimador que seja mais eficiente do que qualquer outro então diz-se que o estimador é eficiente (ou absolutamente eficiente).

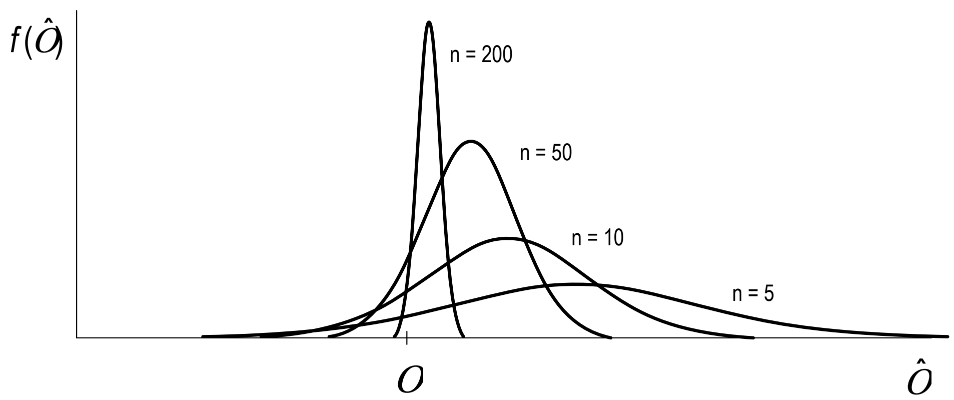
8.2.3. Consistência

Seja $\hat{\theta}$ um estimador do parâmetro θ e n a dimensão da amostra com base na qual $\hat{\theta}$ é calculado.

- O estimador $\hat{\theta}$ diz-se consistente quando, para qualquer valor positivo ∂, se verifica que:

$$\lim_{n \to \infty} PROB\left(\left|\hat{\theta} - \theta\right| < \partial\right) = 1$$

Isto é, quando $n \to \infty$ *o estimador consistente concentra-se perfeitamente sobre o seu alvo* (o valor do parâmetro estimado).



8.3. Métodos de Estimação

8.3.1. Método da Máxima Verosimilhança

Considere-se uma população com função densidade que contenha um parâmetro populacional, θ, a ser estimado por meio de determinada estatística. A função densidade pode neste caso ser denotada por $f(x,\ \theta)$. Admitindo que haja *n observações independentes*, $X_1,\ X_2,\ \ldots,\ X_n$, a função densidade conjunta para essas observações é:

$$L = f(x_1,\ \theta)\ f(x_2,\ \theta)\ \ldots\ f(x_n,\ \theta) \quad ; \quad (L \equiv verosimilhanca)$$

A máxima verosimilhança é obtida, derivando L em ordem a θ e igualando a zero:

$$\frac{dL}{d\theta} = 0$$

- Para o caso de vários parâmetros (k parâmetros), deverão ser tomadas as várias derivadas parciais em relação a cada um deles:

$$L = f(x_1,\ \theta_1)\ f(x_2,\ \theta_2)\ \ldots\ f(x_n,\ \theta_k) \qquad \frac{\partial L}{\partial \theta_i} = 0 \quad ; \quad (i = 1,\ 2,\ \ldots,\ k)$$

Na generalidade dos casos, os estimadores de máxima verosimilhança são dos melhores que se podem obter, apresentando as seguintes características:

(1) Em geral são consistentes;

(2) Embora nem sempre sejam não enviesados e eficientes para amostras pequenas, tendem a possuir aquelas propriedades à medida que a dimensão da amostra cresce;

(3) As suas distribuições aproximam-se de distribuições normais quando a dimensão da amostra tende para infinito.

DGI
2019

8.3.2. *Método da Estimação Linear com Variância mínima*

Para que o ***método da máxima verosimilhança possa ser aplicado, há que especificar previamente o tipo de distribuição da população. No método da estimação linear com variância mínima tal especificação não é necessária***.

- O método consiste em ***combinar linearmente um conjunto de estimadores não-enviesados com variância mínima***.

No caso geral de uma *combinação linear de* k *estimadores*, obtem-se:

$$\hat{\theta} = \sum_{i=1}^{k} a_i \times \hat{\theta}_i \quad \text{com} \quad a_i = \frac{1/\sigma_i^2}{\sum_{j=1}^{k} \left(1/\sigma_j^2\right)}$$

9.1. Conceito de Intervalo de Confiança

Através dos métodos de ***estimação pontual***, faz-se associar a cada parâmetro de uma população, ***um número isolado***, que constitui uma estimativa do parâmetro.

A grande limitação dos métodos de estimação pontual é que ***não fornecem qualquer informação relativa ao rigor ou à confiança das estimativas que através deles são obtidas***.

A ***limitação*** é ***ultrapassada recorrendo aos métodos de estimação por intervalo***. O conceito envolvido em tais métodos é o ***Intervalo de Confiança***.

DGI
2019

Exemplo

Admita-se que a altura dos alunos do IPB segue uma distribuição normal com parâmetros $\mu = 1.69$ m e $\sigma^2 = (0.06 \text{ m})^2$.

Para amostras de dimensão n, a média amostral $\bar{X}$, segue uma distribuição normal:

$$\bar{X} \to \mathrm{N}\left(\mu,\ \sigma^2/n\right)$$

e

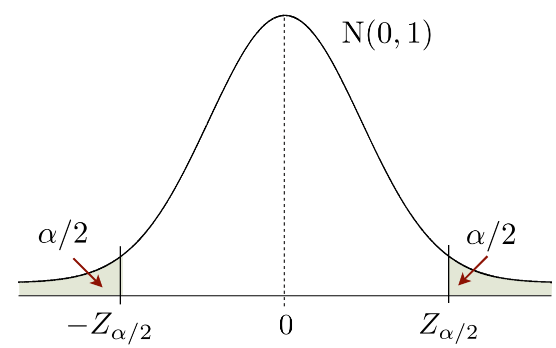
$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad \to \quad \mathrm{N}(0,\ 1)$$

Definindo $Z_{\alpha/2}$ como o valor para o qual:

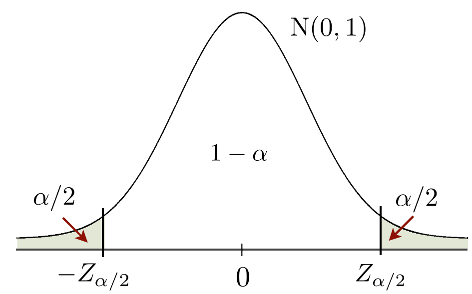
$$P\left(Z > Z_{\alpha/2}\right) = \alpha/2$$

e

$$P\left(Z < -Z_{\alpha/2}\right) = \alpha/2$$



vem: $P\left(-Z_{\alpha/2} < Z < Z_{\alpha/2}\right) = 1-\alpha$



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

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

Reescrevendo a desigualdade:

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

De acordo com a expressão, *o intervalo*: $\bar{X} \pm Z_{\alpha/2} \times \dfrac{\sigma}{\sqrt{n}}$

Incluirá, com a probabilidade $(1-\alpha)$ *o valor de* μ *(isto é, o parâmetro).* ***Tal intervalo designa-se por intervalo de confiança de*** $\boldsymbol{\mu}$ ***a*** $(1-\alpha)\times 100\%$***. Os seus extremos designam-se por limites de confiança.***

Se para uma amostra de dimensão $n = 25$, se obtivesse $\bar{X} = 1.69$, definir o intervalo de confiança a 95%.

Limite Inferior: $\bar{X} - Z_{0.025} \times \dfrac{\sigma}{\sqrt{n}} = 1.69 - 1.96 \times \dfrac{0.06}{5} = 1.666$

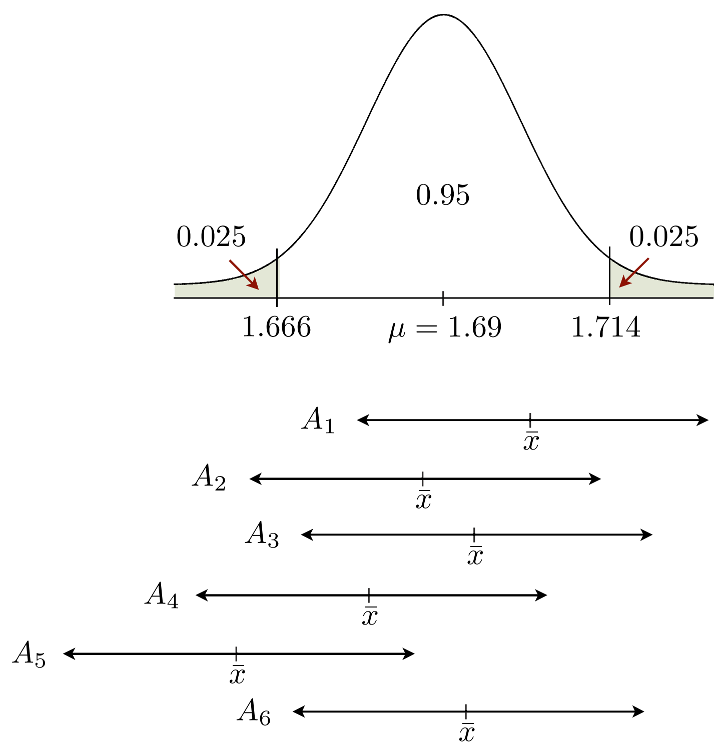
Limite Superior: $\bar{X} + Z_{0.025} \times \dfrac{\sigma}{\sqrt{n}} = 1.69 + 1.96 \times \dfrac{0.06}{5} = 1.714$

Intervalo de Confiança: $[1.666, 1.714]$ ou $1.666 < \mu < 1.714$

Nota

Neste caso admitiu-se que σ^2 (variância da população) era conhecida. Na prática é normalmente conhecida uma estimativa.

O valor de μ é fixo (trata-se de um parâmetro da população a estimar). ***O que varia de amostra para amostra é o intervalo especificado*** *(por depender de $\bar{X}$).*



- ***Numa amostra em cada vinte, o intervalo não conterá o valor do parâmetro estimado.***

DGI
2019

Simetria do Intervalo de Confiança relativamente ao valor da Estimativa Pontual

O intervalo de confiança de μ a $(1-\alpha)\cdot 100\%$:

$$\bar{X} \pm Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$$

é um intervalo simétrico em relação a $\bar{X}$

- Para quaisquer valores α_1, α_2 que satisfaçam a condição $\alpha = \alpha_1 + \alpha_2$, os intervalos de confiança: $\left[\bar{X} - Z_{\alpha 1} \times \frac{\sigma}{\sqrt{n}},\ \bar{X} + Z_{\alpha 2} \times \frac{\sigma}{\sqrt{n}}\right]$

  são intervalos de confiança a $(1-\alpha)\cdot 100\%$.

- Quando a estatística a partir da qual se definem os intervalos tem uma distribuição unimodal simétrica(como sucede com $\bar{X}$, quando a população é normal), então o *intervalo é simétrico* em relação à estatística (isto é, o que se obtém quando $\alpha_1 = \alpha_2 = \alpha/2$) *é o de menor amplitude e portanto o mais interessante*.

Considerando novamente a amostra de dimensão $n = 25$ com $\overline{X} = 1.69$, defina os intervalos de confiança a 95% para as três situações seguintes:

(1) $\alpha_1 = \alpha_2 = 0.025$

$[1.666,\ 1.714]$ $\qquad$ $\text{Amplitude} = 1.714 - 1.666 = 0.048$

(2) $\alpha_1 = 0.01,\ \alpha_2 = 0.04$

$[1.662,\ 1.711]$ $\qquad$ $\text{Amplitude} = 1.711 - 1.662 = 0.049\ (> 0.048)$

(3) $\alpha_1 = 0.005,\ \alpha_2 = 0.045$

$[1.659,\ 1.710]$ $\qquad$ $\text{Amplitude} = 1.710 - 1.659 = 0.051\ (> 0.048)$

Especificação de um Intervalo de Confiança

A especificação de um intervalo de confiança implica o conhecimento simultâneo de:

(1) Um estimador do parâmetro em causa;

(2) A sua distribuição amostral;

(3) Um valor particular do estimador (*isto é, uma estimativa pontual*).

DGI
2019

9.2. Intervalo de Confiança para o Valor Esperado

9.2.1. Amostra de Grande Dimensão, População Qualquer

Neste caso, de acordo com o Teorema do Limite Central, a média amostral segue aproximadamente uma distribuição normal.

$$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \to \mathrm{N}(0,\ 1) \qquad I.C. \text{ a } (1-\alpha)\times 100\% \quad \to \quad \bar{x} \pm z(\alpha/2) \times \frac{\sigma}{\sqrt{n}}$$

Como a dimensão da amostra é grande, o erro de estimação é desprezável, vindo $\sigma \approx s$

com: $$s = \sqrt{\frac{1}{n-1}\sum_{n=1}^{N}(X_n - \bar{X})^2}$$

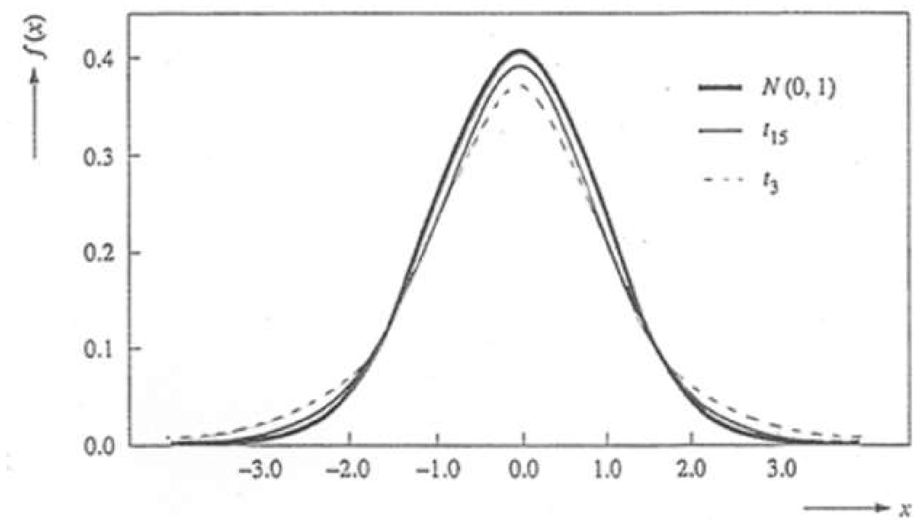
9.2.2. Amostra de Pequena Dimensão, População Normal

Como a dimensão da *amostra é pequena*, já *não é válida a aproximação* $\sigma \approx s$

$$\frac{\bar{x}-\mu}{s/\sqrt{n}} \to t_{n-1} \qquad I.C.\text{a}(1-\alpha)\times 100\% \quad \to \quad \bar{x} \pm t_{n-1}(\alpha/2) \times \frac{s}{\sqrt{n}}$$

9.2.2.1. Distribuição Normal versus t-Student

A utilização da *distribuição* t-**Student leva à definição de Intervalos de *Confiança de maior amplitude*** que os obtidos (inapropriadamente?) através da distribuição Normal.



- ***A distribuição t-Student apresenta maior variância em relação à distribuição Normal***.

  Isso significa que, se os valores de z estão a ser usados quando se deveria usar t, o *Intervalo de Confiança* determinado apresentará uma *amplitude inferior* à que deveria ter.

Intervalo de confiança para o valor esperado a 90% . Dimensão da amostra , $n = 5$

$$t_4(\alpha/2 = 0.05): \quad \bar{x} \pm 2.1318 \times s \Big/ \sqrt{5}$$

$$Z(\alpha/2 = 0.05): \quad \bar{x} \pm 1.6449 \times s \Big/ \sqrt{5}$$

$$t_4(\alpha/2 = ?) = 1.6449 \quad \rightarrow \quad \alpha/2 = 0.088$$
$$\rightarrow \quad \alpha = 0.176$$
$$\rightarrow \quad (1-\alpha) \times 100\% = 82.4\%$$

Na realidade, usando inadvertidamente a distribuição Normal, definimos um intervalo de confiança a 82.4% em vez de 90%.

	Nível de Significância desejada do I.C. para populações Normais		
$\boldsymbol{n}$	90%	95%	99%
5	82.4	87.8	93.8
10	86.5	91.8	97.0
15	87.8	92.9	97.8
20	88.3	93.5	98.1
25	88.7	93.8	98.3
30	88.9	94.0	98.4
50	89.3	94.4	98.6

Os procedimentos propostos anteriormente cobrem todas as situações de obtenção de intervalos de confiança para o valor esperado, excepto quando a amostra é pequena e a população não é normal. Neste caso, seria necessário recorrer à mediana amostral (fora do objectivo da disciplina).

Que procedimento deve ser seguido?

	População	
	Normal	*Não Normal*
$n \leq 30$	t de Student (Normal, se σ conhecido)	Mediana
$n > 30$	t de Student (Normal, se σ conhecido) 2ª escolha: Normal	Mediana 2ª escolha: Normal

9.3. Intervalo de Confiança para a Proporção Binomial

9.3.1. *Amostra de Grande Dimensão*

Considere-se uma ***população dicotómica, constituída por elementos de dois tipos***.

O valor p, que corresponde à ***proporção de elementos de um dos dois tipos designa-se por Proporção Binomial***.

Y – nº de ocorrências de elementos de um dado tipo contidos na amostra.

$$Y \to \mathrm{B}(n,\ p)$$

- De acordo com o ***Teorema do Limite Central, para valores de n suficientemente elevados***:

$$Y \approx \mathrm{N}(n \times p,\ n \times p \times q)$$

como: $$p \approx \hat{p} = \frac{y}{n} \to \mathrm{N}\left(p, \frac{p \times q}{n}\right)$$

dessa forma:

$$\frac{\hat{p}-p}{\sqrt{\frac{p\times q}{n}}} \rightarrow \mathrm{N}(0,1) \qquad I.C.\mathrm{a}(1-\alpha)\times 100\% \quad \rightarrow \quad \hat{p}\pm z(\alpha/2)\sqrt{\frac{\hat{p}\times\hat{q}}{n}}$$

9.3.2. *Amostra de Pequena Dimensão (solução exacta para* $n \leq 30$ *)*

Seja Y a variável aleatória representando *o* ***número de sucessos*** numa amostra aleatória de tamanho n.

- O intervalo de confiança é determinado da seguinte forma:

Limite Inferior: $$P(Y \geq \hat{y}) \rightarrow \sum_{y=\hat{y}}^{n} \mathrm{C}_y^n \times p_i^y \times (1-p_i)^{n-y} = \alpha/2$$

Limite Superior: $$P(Y \leq \hat{y}) \rightarrow \sum_{y=0}^{\hat{y}} \mathrm{C}_y^n \times p_s^y \times (1-p_s)^{n-y} = \alpha/2$$

Nota: Em alternativa pode utilizar-se a *tabela A6* (R. L. Iman & W. J. Conover)

Exemplo

$n = 20$

$\hat{y} = 12$

$\text{Estimativa Pontual}: \quad \hat{p} = \hat{y}/n = 0.6 \quad \rightarrow \quad \text{Intervalo de Confiança:} \ [0.361, 0.809]$

- Note-se que o ***intervalo de confiança não é simétrico*** em relação à estimativa pontual. Note-se também que o ***intervalo de confiança é bastante amplo,*** pelo que é pouco informativo. Como a ***amplitude do intervalo de confiança diminui com o aumento da dimensão da amostra, este exemplo ilustra a necessidade de amostras de dimensão elevada quando se pretende estimar a proporção binomial***.

9.4. Intervalo de Confiança para a Variância de uma População Normal

Se, de uma população normal forem retiradas amostras de dimensão n, com variância amostral s^2:

$$\frac{(n-1)\times s^2}{\sigma^2} \rightarrow \chi^2_{n-1}$$

$$I.C.\text{a}(1-\alpha)\times 100\% \quad \rightarrow \quad \frac{(n-1)\times s^2}{\chi^2_{n-1}(\alpha/2)} \leq \sigma^2 \leq \frac{(n-1)\times s^2}{\chi^2_{n-1}(1-\alpha/2)}$$

Parâmetro	Condições	Distribuição	Intervalo de Confiança	Notas
Valor Esperado μ	Amostra de grande dimensão. População qualquer.	$\dfrac{\overline{x}-\mu}{\sigma/\sqrt{n}} \to N(0,1)$	$\overline{x} \pm z(\alpha/2)\cdot\dfrac{\sigma}{\sqrt{n}}$	Estimador do desvio padrão: $\sigma \approx s$ (1)
	Amostra de pequena dimensão. População Normal.	$\dfrac{\overline{x}-\mu}{s/\sqrt{n}} \to t_{n-1}$	$\overline{x} \pm t_{n-1}(\alpha/2)\cdot\dfrac{s}{\sqrt{n}}$	Não é válida a aproximação: $\sigma \approx s$
Proporção Binomial p	Amostra de grande dimensão. População de Bernoulli. (2)	$\dfrac{\hat{p}-p}{\sqrt{\dfrac{p\cdot q}{n}}} \to N(0,1)$	$\hat{p} \pm z(\alpha/2)\cdot\sqrt{\dfrac{\hat{p}\cdot\hat{q}}{n}}$ (3)	Estimador da proporção binomial: $p \approx \hat{p} = \dfrac{y}{n}$
Variância σ^2	População Normal.	$\dfrac{(n-1)\cdot s^2}{\sigma^2} \to \chi^2_{n-1}$	$\dfrac{(n-1)\cdot s^2}{\chi^2_{n-1}(\alpha/2)} \leq \sigma^2 \leq \dfrac{(n-1)\cdot s^2}{\chi^2_{n-1}(1-\alpha/2)}$	

(1) O desvio padrão σ, sendo desconhecido, é estimado através do erro padrão $s = \sqrt{\dfrac{1}{n-1}\cdot\sum_{i=1}^{n}(x_i-\overline{x})^2}$

(2) No caso de amostras de pequena dimensão é necessário recorrer à solução exacta através da distribuição binomial.

(3) Se a população é finita e a amostragem é efectuada sem reposição, a aproximação da distribuição de Y pela normal continua válida se $n \geq 20,\, n\cdot p \geq 7$ e desde que a dimensão da população M seja grande e a dimensão da amostra, n, muito menor que M. Em rigor, nesta situação, a estimativa do erro padrão de Y/n deverá ter em conta o factor de redução $(M-n)/(M-1)$

(4) A igualdade das variâncias deverá ser testada recorrendo ao teste à razão das variâncias (Teste F).

Parâmetro	Condições		Distribuição	Intervalo de Confiança	Notas
Diferença de valores esperados $\mu_A - \mu_B$ (4)	Amostras independentes de grande dimensão. População qualquer.	Variâncias diferentes	$\dfrac{(\overline{x}_A - \overline{x}_B) - (\mu_A - \mu_B)}{\sqrt{\dfrac{\sigma_A^2}{n_A} + \dfrac{\sigma_B^2}{n_B}}} \rightarrow N(0,1)$	$(\overline{x}_A - \overline{x}_B) \pm z(\alpha/2) \cdot \sqrt{\dfrac{\sigma_A^2}{n_A} + \dfrac{\sigma_B^2}{n_B}}$	Estimadores dos desvios padrão: $\sigma_A^2 \approx s_A^2,\ \sigma_B^2 \approx s_B^2$
		Variâncias iguais	$\dfrac{(\overline{x}_A - \overline{x}_B) - (\mu_A - \mu_B)}{\sigma \cdot \sqrt{\dfrac{1}{n_A} + \dfrac{1}{n_B}}} \rightarrow N(0,1)$	$(\overline{x}_A - \overline{x}_B) \pm z(\alpha/2) \cdot \sigma \cdot \sqrt{\dfrac{1}{n_A} + \dfrac{1}{n_B}}$	$\sigma_A^2 = \sigma_B^2 = \sigma^2 \approx s^2$ $s^2 = \dfrac{(n_A - 1) \cdot s_A^2 + (n_B - 1) \cdot s_B^2}{n_A + n_B - 2}$
	Amostras independentes de pequena dimensão. Populações Normais.	Variâncias diferentes	$\dfrac{(\overline{x}_A - \overline{x}_B) - (\mu_A - \mu_B)}{\sqrt{\dfrac{s_A^2}{n_A} + \dfrac{s_B^2}{n_B}}} \rightarrow t_{GL}$	$(\overline{x}_A - \overline{x}_B) \pm t_{GL}(\alpha/2) \cdot \sqrt{\dfrac{s_A^2}{n_A} + \dfrac{s_B^2}{n_B}}$	$GL = \dfrac{\left(s_A^2/n_A + s_B^2/n_B\right)^2}{\dfrac{\left(s_A^2/n_A\right)^2}{n_A - 1} + \dfrac{\left(s_B^2/n_B\right)^2}{n_B - 1}}$
		Variâncias iguais	$\dfrac{(\overline{x}_A - \overline{x}_B) - (\mu_A - \mu_B)}{s \cdot \sqrt{\dfrac{1}{n_A} + \dfrac{1}{n_B}}} \rightarrow t_{GL}$	$(\overline{x}_A - \overline{x}_B) \pm t_{GL}(\alpha/2) \cdot s \cdot \sqrt{\dfrac{1}{n_A} + \dfrac{1}{n_B}}$	$GL = n_A + n_B - 2$ $s^2 = \dfrac{(n_A - 1) \cdot s_A^2 + (n_B - 1) \cdot s_B^2}{n_A + n_B - 2}$
Diferença de proporções binomiais $p_A - p_B$	Amostras independentes de grande dimensão. Populações de Bernoulli.		$\dfrac{(\hat{p}_A - \hat{p}_B) - (p_A - p_B)}{\sqrt{\dfrac{p_A \cdot q_A}{n_A} + \dfrac{p_B \cdot q_B}{n_B}}} \rightarrow N(0,1)$	$(\hat{p}_A - \hat{p}_B) \pm z(\alpha/2) \cdot \sqrt{\dfrac{\hat{p}_A \cdot \hat{q}_A}{n_A} + \dfrac{\hat{p}_B \cdot \hat{q}_B}{n_B}}$	Estimadores das proporções binomiais: $\hat{p}_A = \dfrac{y_A}{n_A},\ \hat{p}_B = \dfrac{y_B}{n_B}$
Razão de variâncias σ_A^2/σ_B^2	Populações Normais.		$\dfrac{s_A^2/\sigma_A^2}{s_B^2/\sigma_B^2} \rightarrow F_{GL_1,GL_2}$	$\dfrac{s_A^2/s_B^2}{F_{GL_1,GL_2}(\alpha/2)} \leq \dfrac{\sigma_A^2}{\sigma_B^2} \leq \dfrac{s_A^2/s_B^2}{F_{GL_1,GL_2}(1-\alpha/2)}$	$GL_1 = n_A - 1,\ GL_2 = n_B - 1$ $F_{GL_1,GL_2}(1-\alpha/2) = \dfrac{1}{F_{GL_2,GL_1}(\alpha/2)}$

DGI
2019

9.5. Amostras Dependentes

Intervalo de Confiança a (1-α)·100% para a diferença dos valores esperados, $\mu_1 - \mu_2$, quando as *amostras não podem ser consideradas independentes*: ***observações emparelhadas***.

- Define-se uma nova variável aleatória:

$$D = X_1 - X_2$$

Os valores da variável D, são considerados como constituindo uma amostra e o Intervalo de Confiança determina-se como na situação em que existe apenas uma amostra (Intervalo de Confiança para o valor esperado).

Exemplo

Numa experiência industrial, uma certa tarefa foi realizada por 10 operários, primeiro de acordo com o método A e em seguida de acordo com o método B. Os dados que se seguem fornecem os resultados obtidos (tempos em minutos)

operário	1	2	3	4	5	6	7	8	9	10
método A	5.4	5.6	5.8	6.2	4.9	4.9	5.9	6.1	5.3	5.9
método B	5.2	5.7	5.3	5.9	4.9	4.8	5.7	6.0	5.3	5.7
Diferença	-0.2	0.1	-0.5	-0.3	0	-0.1	-0.2	-0.1	0	-0.2

9.6. Dimensionamento de Amostras

Qual o interesse em dimensionar convenientemente uma amostra?

- Recolher e tratar uma ***amostra grande demais*** para os resultados que se pretendem obter constitui um ***desperdício de recursos***.
- É destituída de interesse uma amostra cuja ***dimensão não seja suficiente*** para se poderem tirar *conclusões* ***relevantes***.

O ***dimensionamento correcto*** de uma amostra com base na qual se pretende estimar um ou mais parâmetros de uma população passa pela ***identificação de qual a fiabilidade desejada para a(s) estimativa(s)***.

DGI
2019

9.6.1. Dimensionamento de Amostras para o Valor Esperado

População Normal com σ^2 conhecida.

$$\bar{x} \pm z(\alpha/2) \times \frac{\sigma}{\sqrt{n}}$$

Objectivo: Qual deverá ser a dimensão da amostra a recolher para que a amplitude do intervalo de confiança para o valor esperado não exceda ω?

$$\omega = 2 \times z(\alpha/2) \times \frac{\sigma}{\sqrt{n}} \quad \rightarrow \quad n = 4 \cdot \sigma^2 \times \left[\frac{z(\alpha/2)}{\omega}\right]^2$$

9.6.1. *Dimensionamento de Amostras para o a Proporção Binomial*

Considerando que a dimensão da amostra a retirar é grande (n *grande*):

$$\hat{p} \pm z(\alpha/2) \times \sqrt{\frac{\hat{p} \times \hat{q}}{n}}$$

$$\omega = 2 \times z(\alpha/2) \times \sqrt{\frac{\hat{p} \times \hat{q}}{n}} \quad \rightarrow \quad n = 4 \times \hat{p} \times \hat{q} \times \left[\frac{z(\alpha/2)}{\omega}\right]^2$$

A expressão anterior depende de $\hat{p}$ e $\hat{q} = 1 - \hat{p}$ que são geralmente desconhecidos antes da obtenção da amostra. Nestes casos, utiliza-se o valor mais conservativo: $\hat{p} = 1/2$

$$n = \left[\frac{z(\alpha/2)}{\omega}\right]^2$$

Nota: $\phi = \hat{p} \times \hat{q} = \hat{p} \times (1 - \hat{p})$

$$\frac{d\phi}{d\hat{p}} = (1 - \hat{p}) - \hat{p} = 1 - 2 \times \hat{p} \qquad \frac{d\phi}{d\hat{p}} = 0 \rightarrow \hat{p} = 1/2 \qquad \frac{d^2\phi}{d\hat{p}^2} = -2 < 0 \Rightarrow \text{máximo}$$

10.1. Introdução

Os procedimentos apresentados nos dois ***capítulos anteriores permitem definir estimativas (pontuais ou por intervalo) de parâmetros populacionais***.

Será agora apresentado um outro procedimento de inferência estatística – ***O Teste de Hipóteses – cujo objectivo fundamental é o de verificar se os dados amostrais são ou não compatíveis com determinadas populações ou com valores previamente fixados dos correspondentes parâmetros populacionais***.

O resultado do Teste corresponde inevitavelmente a uma de ***duas respostas possíveis para aquela questão - afirmativa ou negativa. Em ambos os casos corre-se o risco de errar. Uma das características do Teste de Hipóteses é, justamente, a de permitir controlar ou minimizar tal risco***.

DGI
2019

10.2. Metodologia do Teste de Hipóteses

(1) *Definição das Hipóteses;*

(2) *Identificação da Estatística de Teste (ET) e caracterização da sua distribuição amostral;*

(3) *Definição da Regra de Decisão, com especificação do Nível de Significância do Teste (α);*

(4) *Cálculo do valor da ET e Tomada de Decisão.*

10.2.1. Definição da Hipóteses

Uma Hipótese é uma conjectura acerca de uma ou mais populações.

O Teste de Hipóteses é ***utilizado para refutar hipóteses*** e não para provar a veracidade de uma hipótese.

- O objectivo é o de mostrar que uma ***hipótese é insustentável porque ocorrerá com uma probabilidade demasiadamente baixa***.

O Teste de Hipóteses é constituído por ***duas hipóteses***:

- *a hipótese nula* (H_0) *e a hipótese alternativa* (H_1).

A *estratégia* básica seguida na metodologia do Teste de Hipóteses consiste em ***tentar suportar a validade de H_1, uma vez provada a inverosimilhança de H_0***. Por outras palavras, conseguindo-se mostrar que, com elevada probabilidade, a hipótese nula é falsa, fica assim corroborada a validade da hipótese alternativa.

Se, pelo contrário, não se puder rejeitar H_0, a hipótese H_1 não será reforçada pelo Teste.

Notas Importantes

(1) ***A hipótese alternativa contem sempre uma desigualdade*** $(>$ *ou* $<)$ ***ou uma não igualdade*** $(\neq)$***, mas nunca uma igualdade*** $(=)$.

(2) ***A hipótese nula é considerada verdadeira ao longo do procedimento de teste até ao momento em que haja evidência estatística clara apontando em sentido contrário***. Neste caso (isto é, quando se rejeitar H_0), aceita-se como válida a hipótese alternativa (visto que se admite que esta é a hipótese complementar de H_0).

Notas Importantes *(cont.)*

(3) ***A hipótese nula contem sempre uma igualdade***. Mesmo quando na hipótese nula faz sentido figurar o sinal ≥ ou ≤, o Teste é efectuado considerando apenas a situação em que H_0 mais se aproxima de H_1, ou seja, supondo que é verdadeira a afirmação de H_0 que corresponde à igualdade (se se provar que H_0 é falsa, então também serão falsas todas as outras que se afastam mais de H_1).

(4) *Quando* H_1 ***contiver uma desigualdade*** ($>$ ou $<$), o ***Teste diz-se unilateral*** (à direita para o sinal $>$ e à esquerda para o sinal $<$). *Quando* H_1 ***contiver uma não igualdade*** ($\neq$), o ***Teste diz-se bilateral***.

10.2.2. Identificação da Estatística de Teste e Caracterização da sua Distribuição Amostral

A Estatística de Teste é utilizada para verificar a plausibilidade da hipótese nula.

- Para que tal estatística possa cumprir a sua função ***é necessário conhecer a sua distribuição amostral quando se admitir que é verdadeira a hipótese nula***.

10.2.3. Definição da Regra de Decisão com Especificação do Nível de Significância do Teste

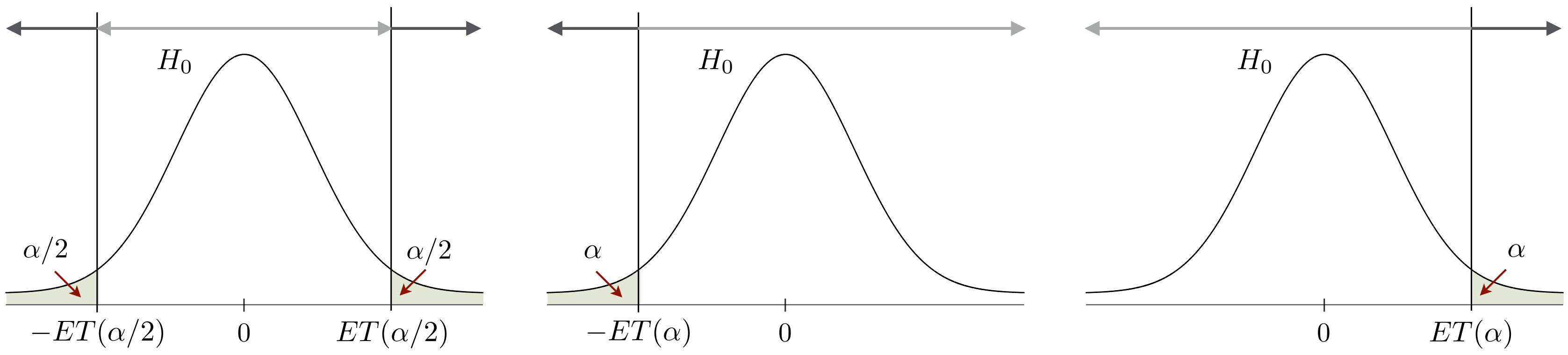
A decisão de rejeitar ou não a hipótese nula fundamenta-se no valor que a Estatística de Teste toma.

- ***A distribuição da Estatística de Teste é especificada na suposição de que H_0 é verdadeira***. Assim, se o valor da ET for muito improvável e, pelo contrário, este valor for razoavelmente provável quando se verificar a hipótese alternativa, então H_0 deverá ser rejeitada em favor de H_1.

DGI
2019

Para que esta decisão possa ser tomada de uma forma controlada, é conveniente que, à partida se ***fixe o valor a partir do qual se considera improvável a validade da hipótese nula***. Tal fixação corresponde à definição da Regra de Decisão do teste. A formalização desta regra passa pela *especificação de uma região de rejeição*.

Distribuição da Estatística de Teste quando H_0 é Verdadeira



[Teste Bilateral] **[Teste Unilateral à Esquerda]** **[Teste Unilateral à Direita]**

Região de rejeição de H_0
Região de não rejeição de H_0

Nível de Significância do Teste ou Erro de Tipo I

A probabilidade de, no caso de H_0 ser verdadeira, o valor da ET pertencer à região de rejeição, designa-se por Nível de significância do Teste (α).

O ***Nível de Significância*** representa, então, a probabilidade (ou o risco) de se ***incorrer no erro de rejeitar H_0 quando esta é, de facto, verdadeira – Erro Tipo I***

Cálculo do Valor da Estatística de Teste e Tomada de Decisão

A última fase do Teste de Hipóteses corresponde ao cálculo do valor da Estatística de Teste e, face ao valor obtido, à aplicação da Regra de Decisão

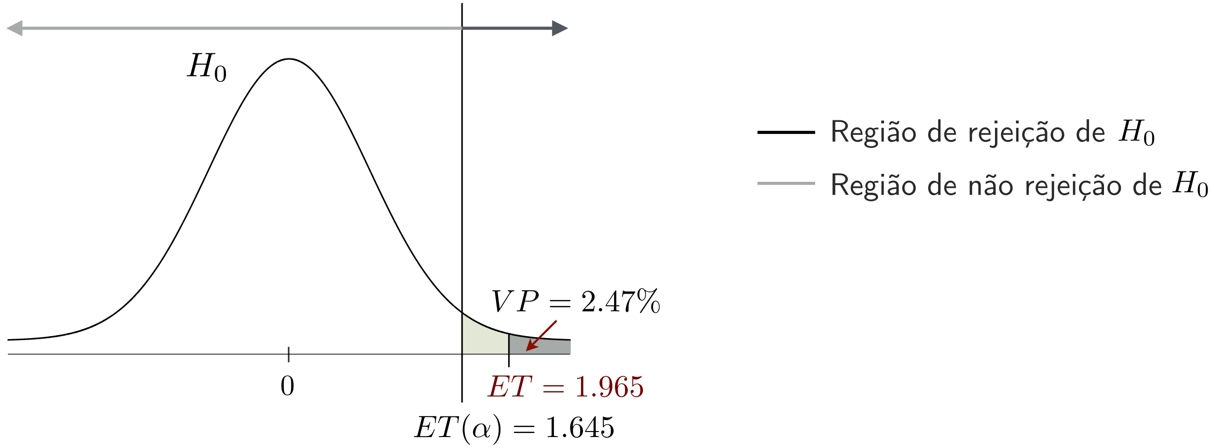
- Se a Estatística de Teste pertencer à região de rejeição, $ET > ET(\alpha)$, *então* H_0 ***é rejeitada***;
- Se a Estatística de Teste não se localizar na região de rejeição, $ET < ET(\alpha)$, *então* H_0 ***não é rejeitada e o resultado do teste diz-se inconclusivo***.

Valor de Prova

De acordo com o procedimento anteriormente descrito para o teste de hipóteses, no final toma-se a decisão de rejeitar ou de não rejeitar da hipótese nula. Esta dicotomia é, na realidade artificial, na medida em que:

- A fixação de um nível de significância é francamente arbitrária;
- Os dados amostrais podem contradizer a hipótese nula em maior ou menor grau.

O Valor de Prova (VP), definido como correspondendo à probabilidade da ET tomar um valor igual ou mais extremo do que aquele que de facto é observado, constitui uma medida do grau com que os dados amostrais contradizem a hipótese nula.

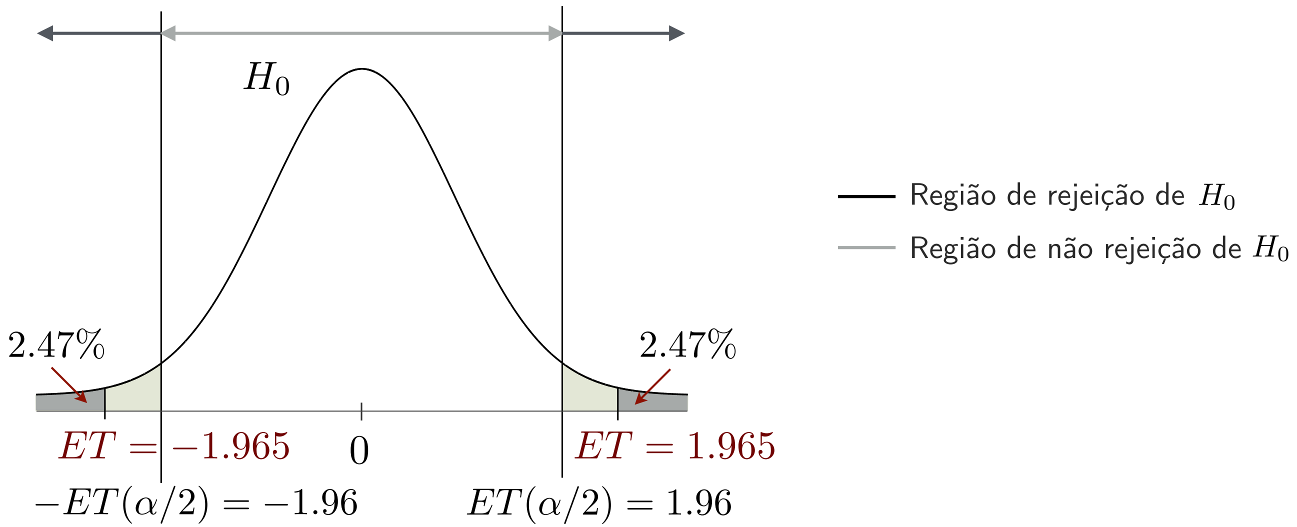


Note-se que, tal como a ET, o ***Valor de Prova é calculado admitindo que H_0 é verdadeira***.

Como é evidente, ***quanto menor for o Valor de Prova, maior será o grau com que a hipótese nula é contradita***.

- ***Dada a relevância da informação contida no Valor de Prova, é recomendável a sua inclusão explícita nos resultados de qualquer Teste de Hipóteses.*** Por exemplo, muito mais esclarecedor do que dizer que uma hipótese nula foi rejeitada ao nível de significância de 5% é afirmar que isso sucedeu e que o Valor de Prova foi, suponha-se, de 0.3%.

Quando o ***Teste é Bilateral***, no cálculo do Valor de Prova devem ***tomar-se em consideração ambas as caudas da distribuição da Estatística de Teste***.



$$VP = 2 \times 2.47\% = 4.94\%$$

Erro Tipo II e Potência do Teste

Foi dito anteriormente que existe o risco de se ***rejeitar*** H_0 quando a hipótese é, de facto ***verdadeira – Nível de Significância (α) ou Erro Tipo I***.

Existe ainda a possibilidade de se cometer um outro tipo de erro no teste de hipóteses: o que corresponde a ***não se rejeitar*** H_0 ***quando a hipótese alternativa,*** H_1***, é verdadeira***. Este erro é designado por *Erro de Tipo II* e a probabilidade de nele se incorrer é denotada por β.

	H_0 Verdadeira	H_0 Falsa
H_0 rejeitada	Erro do Tipo I (probabilidade α)	Potência do Teste (probabilidade 1-β)
H_0 não rejeitada	Decisão acertada (probabilidade 1-α)	Erro do Tipo II (probabilidade β)

A diferença ***1* - *β*** traduz a probabilidade de se *rejeitar* H_0 quando ela é, de facto, falsa.

Atendendo ao seu significado – ***a probabilidade de, correctamente, rejeitar uma hipótese nula falsa – tal diferença designa-se por Potência do Teste***.

Resumindo:

$$P(\text{erro Tipo I}) = P\left(\text{rejeitar } H_0 | H_0 \text{ verdadeira}\right) = P\left(ET \in \text{ZR} | H_0 \text{ verdadeira}\right) = \alpha$$

$$P(\text{erro Tipo II}) = P\left(\text{não rejeitar } H_0 | H_0 \text{ falsa}\right) = P\left(ET \notin \text{ZR} | H_0 \text{ falsa}\right) = \beta$$

$$\text{Potência do Teste} = P\left(\text{rejeitar } H_0 | H_0 \text{ falsa}\right) = P\left(ET \in \text{ZR} | H_0 \text{ falsa}\right) = 1 - \beta$$

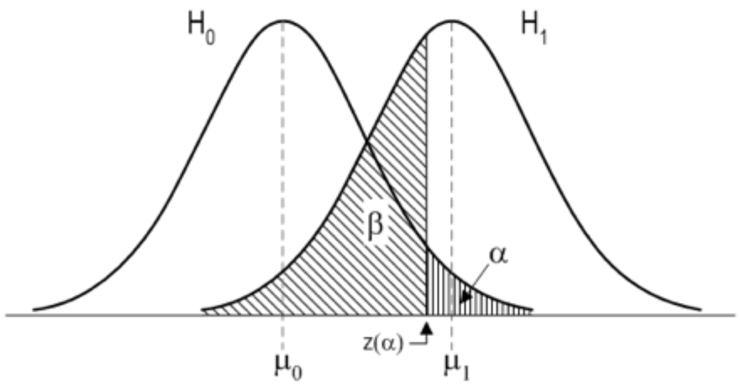
Dessa forma, a determinação da probabilidade de se cometer um Erro de Tipo II passa por:

(1) Determinar em que condições amostrais H_0 não é rejeitada, considerando o teste que estamos a efectuar (*unilateral ou bilateral*);

(2) Determinar β (probabilidade de não rejeitarmos H_0 quando H_0 é falsa), considerando o valor verdadeiro e as condições amostrais de não rejeição obtidas no ponto anterior.

Exemplo

$$\mathrm{H}_0 : \mu = \mu_0$$
$$\mathrm{H}_1 : \mu > \mu_0$$



$X \rightarrow \mathrm{N}(\mu, \sigma^2)$ *com variância conhecida*

$$ET = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

(1) H_0 não rejeitada se: $ET < z(\alpha) \quad \rightarrow \quad \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < z(\alpha) \Leftrightarrow \bar{X} < \mu_0 + z(\alpha) \times \dfrac{\sigma}{\sqrt{n}}$

(2) $\beta = P(\text{não rejeitar } H_0 | H_0 \text{ falsa}) = P\left(\bar{X} < \mu_0 + z(\alpha) \times \dfrac{\sigma}{\sqrt{n}} | \mu = \mu_1\right) =$

$$= P\left(z < \frac{\mu_0 + z(\alpha) \times \frac{\sigma}{\sqrt{n}} - \mu_1}{\sigma/\sqrt{n}}\right) =$$

$$= P\left(z < z(\alpha) - \frac{(\mu_1 - \mu_0)}{\sigma/\sqrt{n}}\right)$$

DGI
2019

Erro de Tipo I versus Erro do Tipo II

Seria desejável que tanto α como β fossem o mais pequenos possível. No entanto, se se mantiverem constantes os dados amostrais com base nos quais se calcula a estatística de teste:

- *ao diminuir α, aumenta β e, inversamente, ao diminuir β, aumenta α.*

É importante notar que ***só existe uma forma de diminuir um dos riscos α ou β sem aumentar o outro: para o conseguir, há que aumentar o número de dados amostrais com base nos quais é calculada a Estatística de Teste***.

Registe-se, no entanto, que as ***amostras exageradamente grandes podem, no extremo, conduzir à rejeição, por excesso de potência, de uma hipótese nula que, embora não sendo exactamente verdadeira, seja aproximadamente verdadeira***.

Por outras palavras, com amostras de dimensões exageradas, o erro padrão da estatística de teste é tão reduzido que qualquer diferença insignificante entre H_0 e H_1 será detectada, implicando a rejeição de H_0.

10.3. Dimensionamento de Amostras

Como vimos, existem dois tipos de erros que se podem cometer num teste de hipóteses:

- o erro de rejeitar H_0 quando esta é verdadeira (Erro do tipo I; probabilidade α) e
- o erro de não rejeitar H_0 quando esta é falsa (Erro do Tipo II; probabilidade β).

É possível (por exemplo, para o caso de amostras de grande dimensão) determinar qual a dimensão da amostra a recolher se pretendemos controlar simultaneamente as probabilidades de incorreremos nos Erros de Tipo I e II.

10.4. Relação entre Teste de Hipóteses e Intervalos de Confiança

A relação fundamental que existe entre os testes de hipóteses e os intervalos de confiança pode ser enunciada nos termos seguintes:

Uma hipótese nula, H_0: $\theta = \theta_0$, pode ser rejeitada a um nível de significância α se, e só se, o intervalo de confiança de θ a $(1-\alpha)\cdot 100\%$ não incluir o valor de θ_0.

Note-se que a condição, impõe que o *intervalo de confiança seja compatível com a natureza de H_1*, ou seja, que para testes bilaterais se construam intervalos de confiança bilaterais e, para testes unilaterais (num sentido), se construam intervalos de confiança unilaterais (no mesmo sentido).

A implicação essencial desta relação é a de que se pode proceder ao Teste de Hipóteses recorrendo a Intervalos de Confiança.