PG_MKT & Business Technologies
Models and Decision in Business Analytics

Agenda

01

Conceitos Estatistica Descritiva

02

Introdução à Teoria da Probabilidade

03

Distribuições Discretas Univariadas

EXECUTIVE EDUCATION

Sessão Síncrona nº 1 – 7 de Setembro de 2021 (21h00-23h00)

Revisões da estatistica descritiva

- Conceito de estatistica
- Classificação dos dados
- Medidas de estatistica descritiva de tendência central
- Medidas de estatistica descritiva de dispersão

Introdução à Probabilidade

- Teroria Elementar de Probabilidade
- Variáveis aleatórias, distribuição de probabilidade
- Distribuições conjuntas de probabilidade

Distribuições Discretas Univariadas

- Distribuição Binomial
- Distribuição Hipergeométrica
- Distribuição de Poisson

01

Conceitos de Estatística Descritiva

Conceito de Estatística

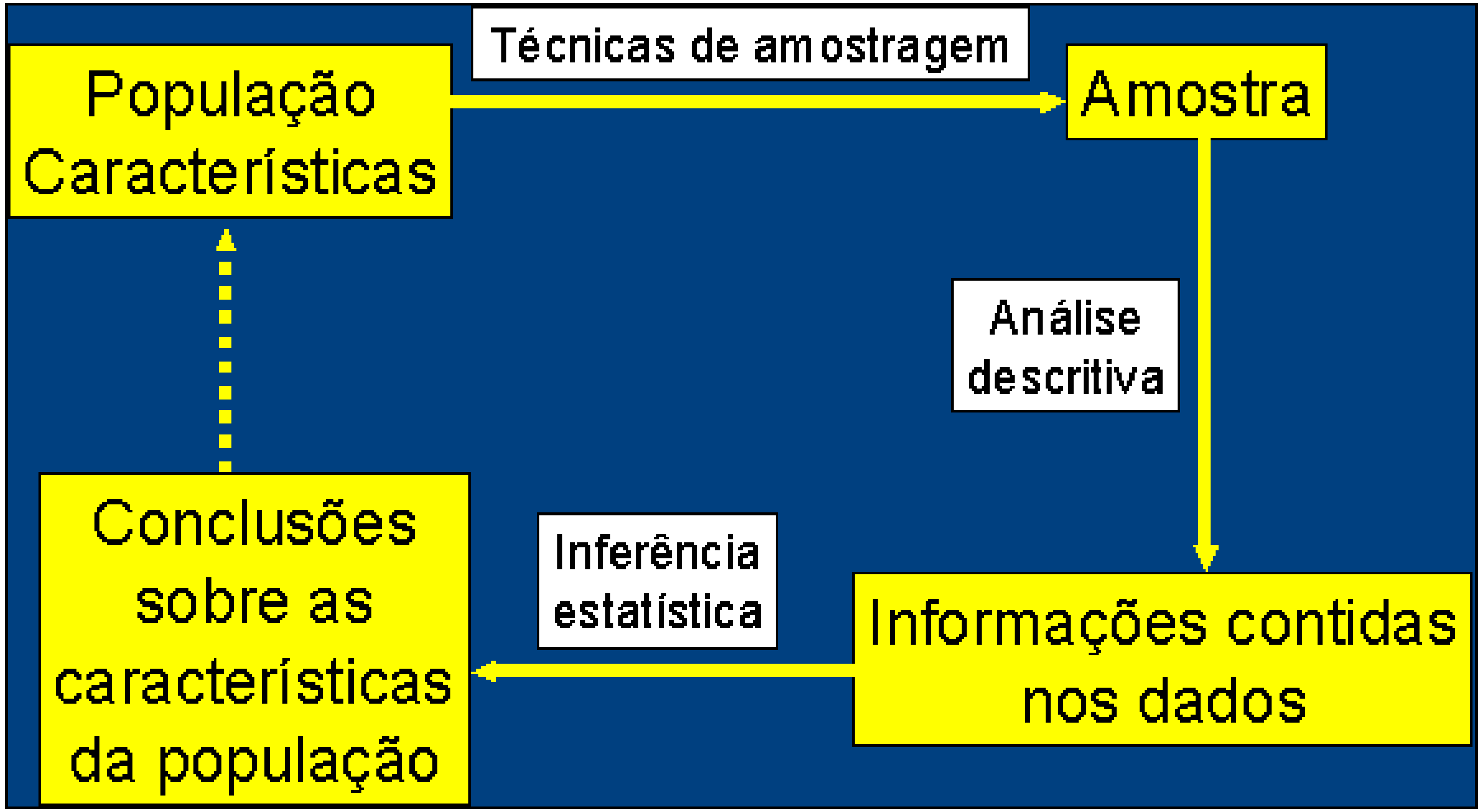
A estatística é um conjunto de técnicas que permite de forma sistemática: organizar, descrever, analisar e interpretar dados oriundos de estudos ou experiências, realizados em qualquer área do conhecimento.

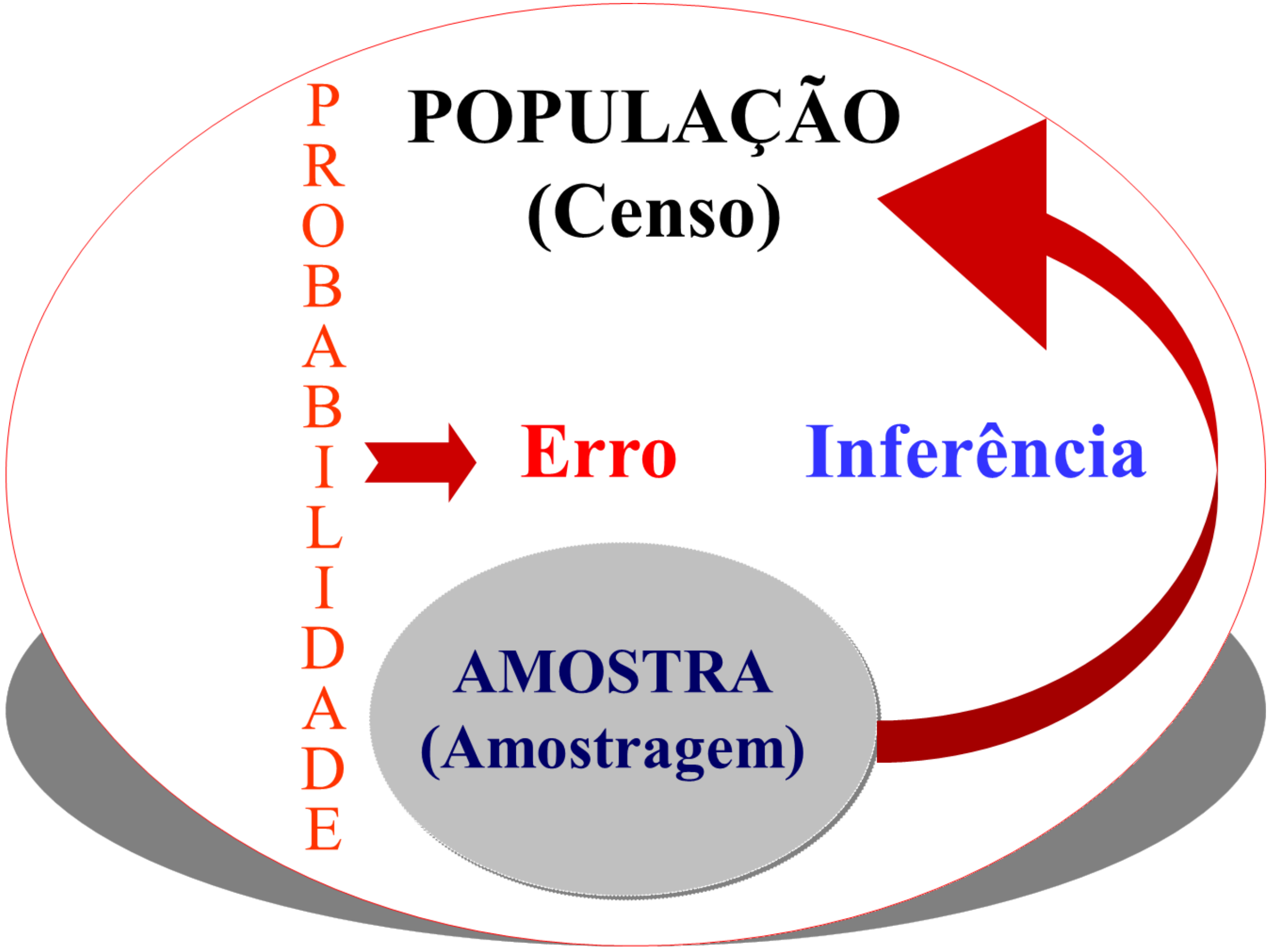
1 - Estatística Descritiva

2 - Probabilidade

3 - Inferência estatística

Funcionamento das diversas componentes da estatística





CENSO

Um levantamento efetuado sobre toda uma população é denominado de levantamento censitário ou simplesmente censo.

Tipologias de variáveis

Variável Qualquer característica associada a uma população

Classificação de variáveis		
Qualitativa	Nominal	sexo, cor dos olhos
	Ordinal	Classe social, grau de instrução
Quantitativa	Contínua	Peso, altura, salario
	Discreta	Número de filhos, número de carros

Variáveis Quantitativas

MEDIDAS DE POSIÇÃO OU TENDÊNCIA CENTRAL:

- Moda, Média, Mediana.

MEDIDAS DE DISPERSÃO:

- Amplitude, Intervalo-Interquartil, Quartís, Variância, Desvio Padrão, Coeficiente de Variação.

MÉDIA: Numa amostra de n observações, $x_1, x_2, \ldots, x_n$

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum x_i}{n}$$

Ex:2,5,3,7,8

Média = [(2+5+3+7+8)/5]=5

Se os dados estiverem agrupados (k valores distintos)

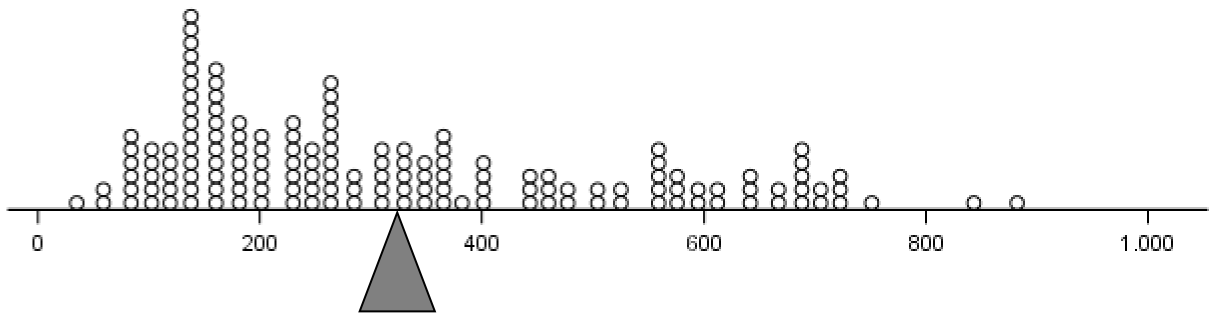
$$\bar{x} = \frac{x_1^* f_1 + x_2^* f_2 + \ldots + x_n^* f_n}{n} = \frac{\sum_{i=1}^{k} x_i^* f_i}{n} = \frac{\sum x_i^* f_i}{n}$$

Função Excel

AVERAGE (number1, number2, ...)

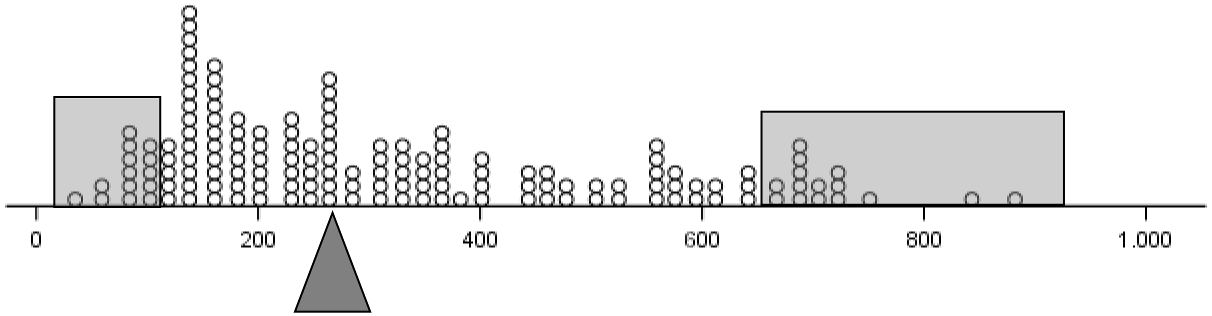
onde f_i designa a frequência absoluta de x_i^* (ou a frequência absoluta da classe com marca x_i^* no caso de dados agrupados em classes)

- A média pode ser pensada como o centro de massa dos valores das observações, ie, o ponto de equilibrio após dispormos as observações sobre uma régua.



Pontos afastados ou erros nas observações podem afastar a média da maior parte das observações.

- Uma média “aparada” não é mais do que uma “mistura” entre os conceitos de média e mediana por forma a combinar as qualidades de ambas.
- Uma média “aparada” é uma média que é calculada excluindo uma certa quantidade de observações em cada extremo da amostra (tipicos outliers)



MODA

É o valor (ou atributo) que ocorre com maior frequência, isto é, que se repete mais vezes.

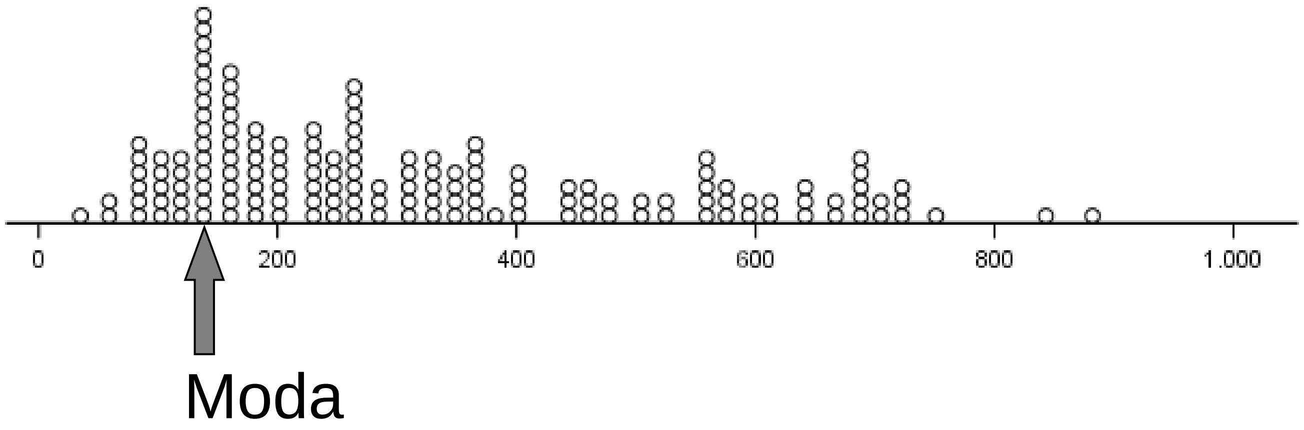
Função Excel

MODE.SNGL (number1, number2, ...)

Variável qualitativa

Ex: 4,5,4,6,5,8,4,4
Mo = 4

- A moda é o valor mais frequente de uma amostra.
- Ao contrário do que acontece com a mediana e a média, uma amostra pode possuir mais do que uma moda.



Algumas observações sobre a Moda:

- A moda é a única medida de localização central que pode ser utilizada para dados numa escala nominal;
- A moda pode não ter significado, especialmente em dados de natureza contínua ou em dados discretos com poucas observações repetidas;
- Quando os dados estão agrupados em classes podemos falar da classe modal, ou seja, da classe com maior frequência.

Mediana

A mediana é o valor da variável que ocupa a posição central de um conjunto de n dados ordenados. O valor que divide os dados em duas partes iguais.

Posição da mediana: **(n+1)/2**

Ex: 2,5,3,7,8

Dados ordenados: 2,3,5,7,8 => (5+1)/2=3
=> Md = 5

Função Excel

MEDIAN (number1, number2, ...)

Média Vs Mediana

Quando é que a mediana é uma medida melhor do que a média?

Imagine uma empresa pequena com 7 trabalhadores com os seguintes salários:

28.000€
33.000€
33.000€
34.000€
37.000€
40.000€
400.000€

Qual o salário tipico nesta empresa?

Média= 86.000€

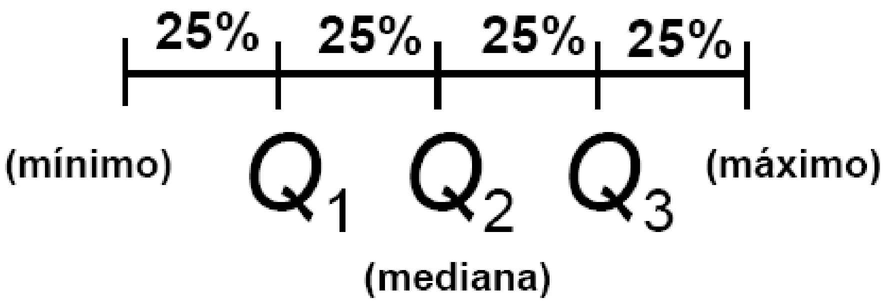
Mediana= 34.000€

Medidas de dispersão

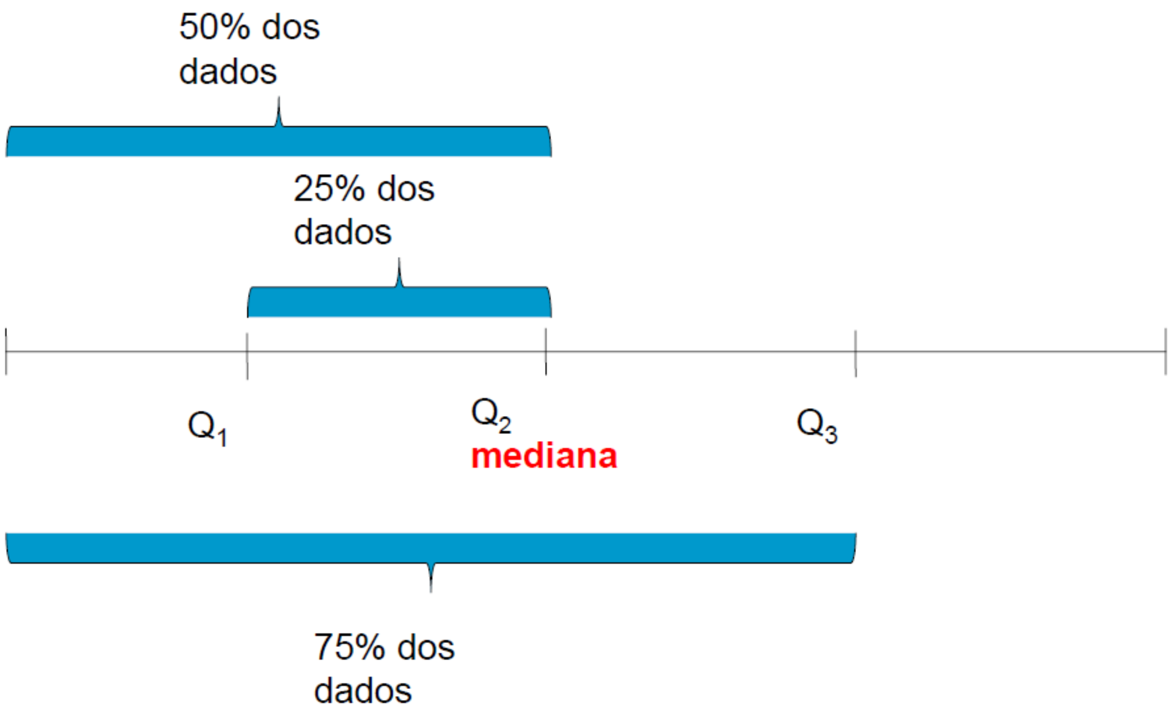
- Localização relativa:
  - Mínimo;
  - Máximo;
  - Quartil;
  - Percentil.

- **Quartís** – são os valores (Q_1, Q_2 e Q_3) que dividem a amostra, depois de ordenada, em quatro partes iguais (ou o mais iguais possível).

- Q2 coincide com a mediana.



- Os Quartis dividem os dados em quatro partes iguais;
- Primeiro Quartil: 25% dos valores são menores ou iguais a Q_1, e 75% são maiores ou iguais a Q_1.
- Terceiro Quartil: 25% dos valores são maiores ou iguais a Q_3, e 75% dos valores são menores ou iguais a Q_3.



$$Q_1 = x_{\left(\frac{n+1}{4}\right)} + 0{,}75_{\left(x_{\left(\frac{n+1}{4}+1\right)} - x_{\left(\frac{n+1}{4}\right)}\right)}$$

$$Q_3 = x_{\left(\frac{3n+3}{4}\right)} + 0{,}25_{\left(x_{\left(\frac{3n+3}{4}+1\right)} - x_{\left(\frac{3n+3}{4}\right)}\right)}$$

Onde:

- n: tamanho da amostra;
- x:

O Segundo Quartil Q_2= Mediana.

Exemplo: Calculo de Quartís

Considere a seguinte amostra de 20 dados:

109	125	126	130	130	131	132	133	133	133	133	134	134	140	140	142	144	144	145	152

De acordo com a metodologia apresentada, o Q1 e Q3 são nas posições 5,75 e 15,25 respetivamente.

Calculo dos Q1 e Q2:

$Q_1 = 130 + 0,75 \times (131 - 130)$	e	$Q_3 = 140 + 0,25 \times (142 - 140)$
$= 130,75$		$= 140,5$

Medidas de Dispersão:

- Amplitude;
- distância inter-quartil;
- Variância;
- desvio padrão;
- coeficiente de variação.

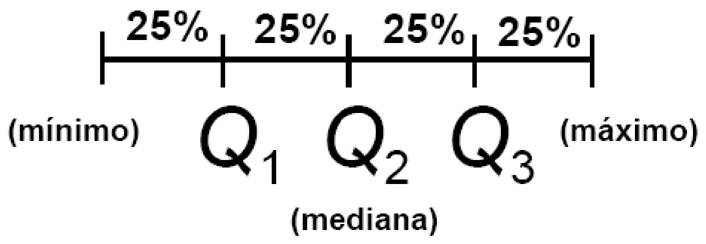
- A amplitude de uma amostra é a diferença entre o máximo e o mínimo.

  Exemplo: Na amostra 1.2; 1.7; 2.1; 2.2; 2.3 a amplitude é .

$$2.3 - 1.2 = 1.1$$

Distância inter-quartil (IQR)

- É a diferença entre o 3º e o 1º quartís, $Q_3 - Q_1$.
- No intervalo que vai de Q_1 a Q_3 encontram-se 50% das observações (correspondem às obervações mais centrais).



Função Excel

QUARTILE-INC (array, quart)

IQR= QUARTILE.INC (array, 3) – QUARTILE.INC (array, 1)

Exemplo: Intervalo-Interquartil

É a diferença entre o terceiro quartil e o primeiro quartil, ou seja, IQR= Q3-Q1

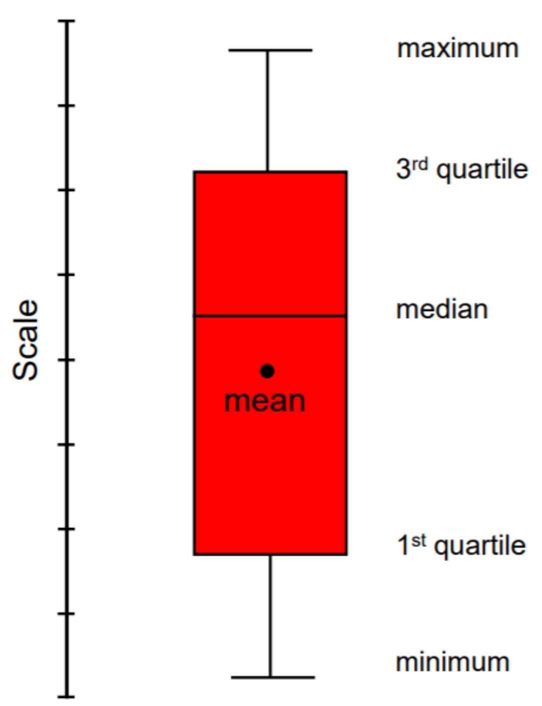
Dados: 15,5,3,8,10,2,7,11,12

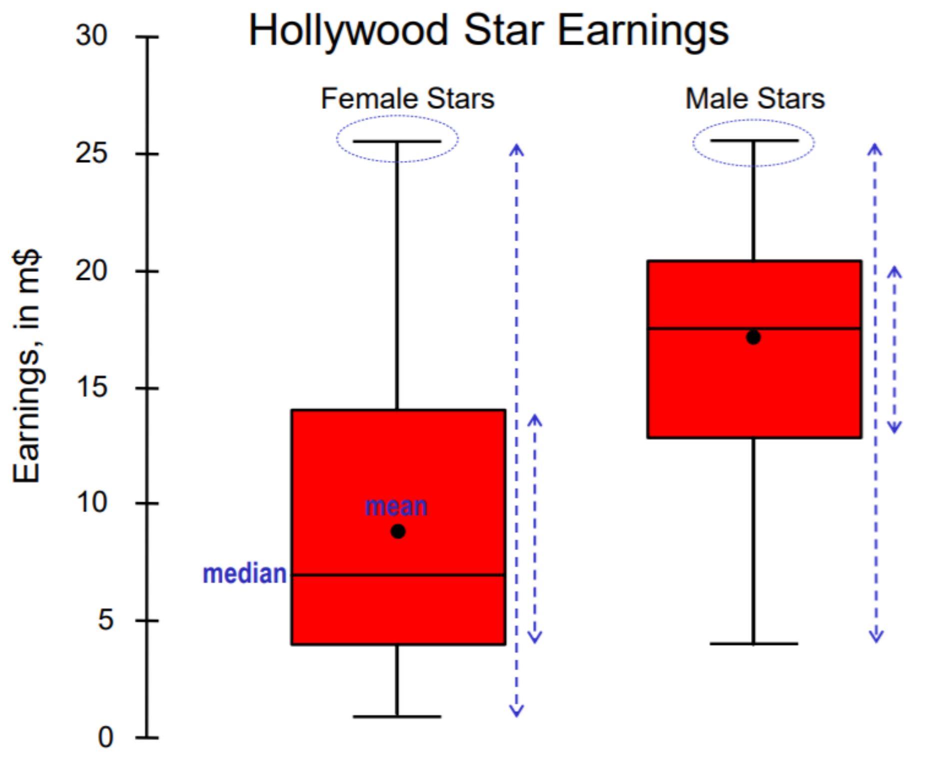
Q1=4,5 e Q3=11,25

d =Q3-Q1=4,9-2,05=2,85

Max,Min,Q1,Q3,Q2: importantes para se ter uma boa ideia da forma da distribuição dos dados (simétrica ou assimétrica).

Representação BOX-PLOT





IPAM EXECUTIVE EDUCATION

Variância

- É a média dos quadrados dos desvios das observações em relação à média do conjunto de dados (população):

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

- Habitualmente considera-se uma versão corrigida da variância quando se trata de uma amostra:

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

- Se tomarmos a raiz quadrada da variância obtemos o desvio padrão que também é uma medida de dispersão e vem na mesma unidade das observacões.

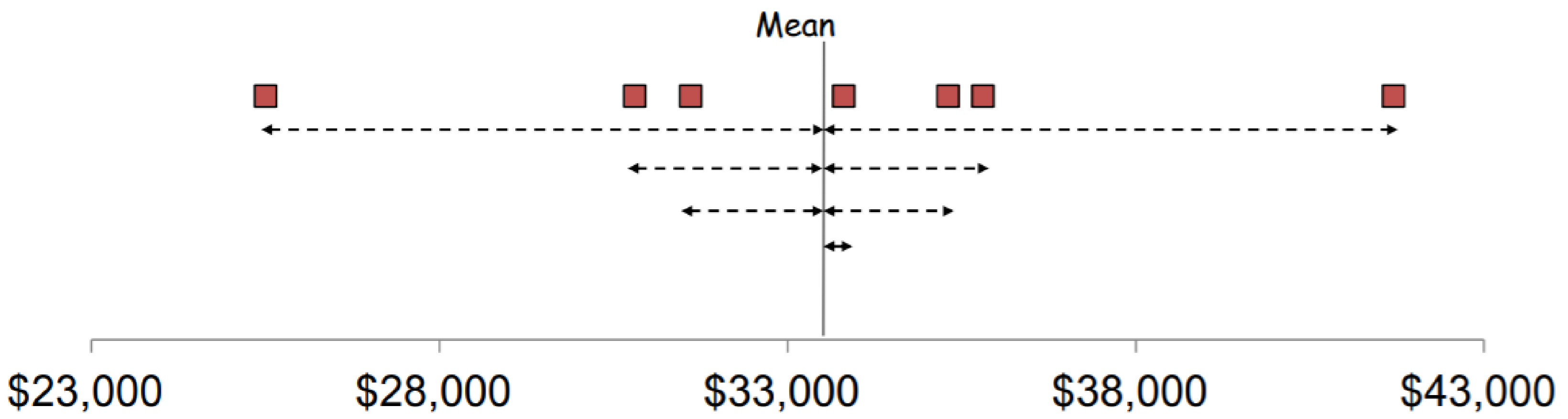
$$s_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \qquad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- Nos programas de estatística e nas máquinas de calcular o que aparece são as versões corrigidas da variância e do desvio padrão.
- O desvio padrão e a variância podem ser fortemente afectados por erros ou observações muito afastadas.

Função Excel

STDEV.P(number1, number2,…) - população
STDEV.S (number1, number2,...) - amostra

$$\text{Desvio Padrão} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(difference_i)^2}$$



Coeficiente de variação

- É a razão entre o desvio padrão e a média, ***v* = *s* / *x***.

- Trata-se de uma medida relativa de dispersão e por isso não tem uma unidade especifica (vem em %).

Algumas Observações sobre o CV

- É uma medida de dispersão relativa;
- Elimina o efeito da magnitude dos dados;
- Exprime a variabilidade em relação à média;
- É útil para comparar duas ou mais variáveis.

Coeficiente de variação

Se **CV <= 30%** os dados são considerados homogéneos
Se **CV >= 30%** os dados são considerados heterogéneos

Dados homogéneos: concentração aceitável e pode-se realizar comparações entre médias e outros indicadores estatísticos.

Dados heterogéneos: dispersão considerável e pode comprometer comparações entre médias e outros indicadores estatísticos

Exemplo do CV
Altura e peso de alunos

	Média	Desvio padrão	Coeficiente de variação
Altura	1,143m	0,063m	5,5%
Peso	50Kg	6kg	12%

Conclusão:

Em relação à interpretação das médias, os alunos são, aproximadamente, duas vezes mais dispersos quanto ao peso do que quanto a altura

ORGANIZAÇÃO E REPRESENTAÇÃO DOS DADOS

Uma das formas de organizar e resumir a informação contida em dados observados é por meio de tabela de frequências e gráficos.

Tabela de freqüência: relaciona categorias (ou classes) de valores, juntamente com contagem (frequências) do número de valores que se enquadram em cada categoria ou classe.

Variáveis qualitativas: Podemos construir uma tabela de frequência que agrupem os dados por categoria de classificação e a sua representação gráfica é mediante gráfico de barras ou gráfico circular.

Tabela de frequência

Exemplo:

Considere a variável grau de Instrução

(Variável qualitativa)

Grau de instrução	Contagem	f_i	f_{r_i}	$f_{r_i}\%$
1º Grau		12	0,3333	33,3%
2º Grau		18	0,5000	50 %
Superior		6	0,1667	16.7%
total		n=36	1,0000	100%

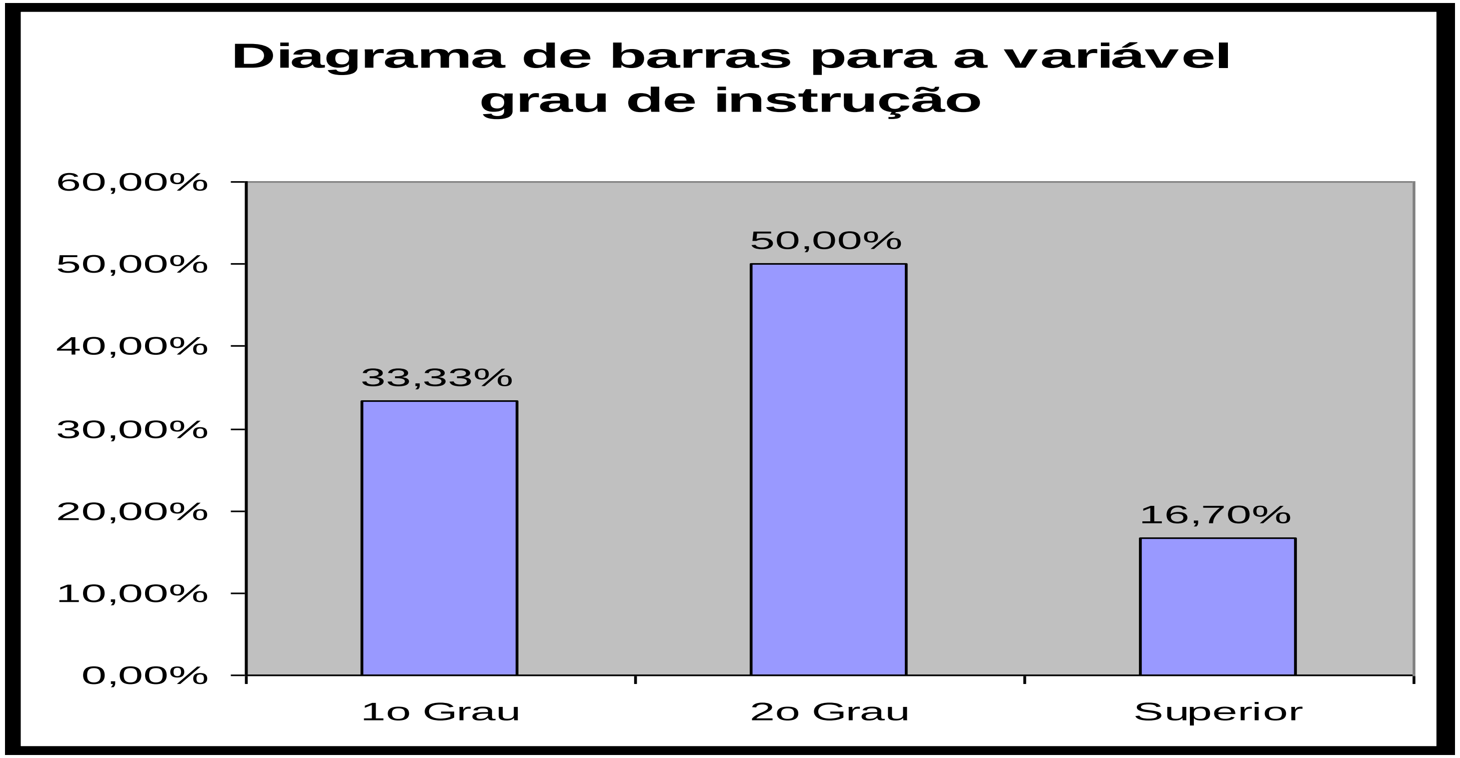
f_i :**Frequência absoluta** da categoria i (número de indivíduos que pertencem à categoria i

$$f_{r_i} = \frac{f_i}{n}$$: **Frequência relativa** da categoria i

$$f_{r_i}\% = f_{r_i} * 100\%$$: **Frequência relativa percentual** da categoria i

Representação gráfica de variáveis qualitativas

- Gráfico de Barras
- Gráfico circular



Exemplo de variável Quantitativa discreta:

Organizam-se mediante tabelas de frequências e a representação gráfica é mediante gráfico de barras.

Exemplo:

Considere a variável número de filhos.

Tabela: Distribuição de frequências de funcionários da empresa segundo o número de filhos:

i	Número de filhos (X_i)	Número de funcionários (f_i)	% de funcionários (f_{ri})
1	0	4	20%
2	1	5	25%
3	2	7	35%
4	3	3	15%
5	5	1	5%
total		20	100%

Determinação das medidas de posição e medidas de dispersão para variáveis quantitativas discretas agrupados em tabela de freqüências:

Média:

$$\overline{X} = \frac{X_1 f_1 + X_2 f_2 + \cdots + X_k f_k}{n} = \frac{\sum_{i=1}^{k} X_i f_i}{n}$$

Exemplo: Considere a tabela anterior e determine a média de filhos dos funcionários.

$$\overline{X} = \frac{0\times 4 + 1\times 5 + 2\times 7 + 3\times 3 + 5\times 1}{20} = \frac{33}{20} = 1{,}65$$

Mediana:

Dados ordenados:

0 0 0 0 1 1 1 1 1 2 2 2 2 2 2 2 3 3 3 5 => (20+1)/2=10,5 => Md = (2+2) /2=2

Variância:

$$S^2 = \frac{(X_1 - \overline{X})^2 f_1 + (X_2 - \overline{X})^2 f_2 + \cdots + (X_k - \overline{X})^2 f_k}{n-1} = \frac{\sum_{i=1}^{k} (X_i - \overline{X})^2 f_i}{n-1}$$

Cálculo da variância para os dados da tabela (filhos funcionários)

$$S^2 = \frac{4(0-1{,}65)^2 + 5(1-1{,}65)^2 + 7(2-1{,}65)^2 + 3(3-1{,}65)^2 + (5-1{,}65)^2}{19}$$

$$= \frac{16{,}3125}{19} = 0{,}858553$$

Desvio padrão:

$$S = \sqrt{S^2} = \sqrt{0{,}858553} = 0{,}927$$

Variáveis Quantitativas continuas:

Os seus valores podem ser qualquer número real e ainda geralmente existe um grande número de valores diferentes. Como proceder para construir uma tabela de frequência nestes casos?

A alternativa consiste em construir classes ou intervalos de valores e contar o número de ocorrências em cada intervalo.

Procedimento para construção de tabelas de frequência para variáveis contínuas:

1. Escolha o número de intervalos de classe (k)
2. Identifique o menor valor (MIN) e o valor máximo (MAX) dos dados.
3. Calcule a amplitude dos dados (A): **A=MAX –MIN**
4. Calcule o comprimento de cada intervalo de classe (h): $\boxed{h = \frac{A}{k}}$
5. Arredonde o valor de h de forma que seja obtido um número conveniente.
6. Obtenha os limites de cada intervalo de classe.

$$\begin{aligned}&\text{PRIMEIRO INTERVALO:}\\&\text{Limite inferior: } \mathrm{LI}_1 = MIN\\&\text{Limite superior: } \mathrm{LS}_1 = LI_1 + h\end{aligned}$$

EXECUTIVE EDUCATION

- Ponto médio de cada intervalo de classe:

$$X_{i}' = \frac{LS_i + LI_i}{2}$$

- Contagem dos dados pertencentes a cada intervalo.
- Frequências absolutas de cada intervalo de classe.
- Frequências relativas de cada intervalo de classe.
- Frequências acumuladas absolutas de cada intervalo de classe.

$$F_i = f_1 + f_2 + \cdots + f_i = \sum_{j=1}^{i} f_j$$

- Frequências acumuladas relativa de cada intervalo de classe.

$$F_{r_i} = f_{r_1} + f_{r_2} + \cdots + f_{r_i} = \sum_{j=1}^{i} f_{r_j}\,;\ ou\ F_{r_i} = \frac{F_i}{n}$$

Exemplo:
Considere a variável salário de uma empresa comercializadora de produtos de informática.

Procedimento:
1. Considere k=5.
2. MIN=4; MAX=23,30.
3. A=MAX-MIN=23,30-4=19,30
4. h=19,3/5=3,86
5. h≈3,9
6. Cálculo dos limites de cada intervalo:

PRIMEIRO INTERVALO
$$LI_1 = 4$$
$$LS_1 = 4 + 3{,}9 = 7{,}9$$
SEGUNDO INTERVALO
$$LI_2 = 7{,}9$$
$$LS_2 = 7{,}9 + 3{,}9 = 11{,}8$$

Os demais limites dos intervalos foram gerados seguindo o procedimento anterior.

Ponto médio: $X'_1 = \frac{(4+7,9)}{2} = 5,95; \quad X'_2 = \frac{(7,9+11,8)}{2} = 9,85.....$

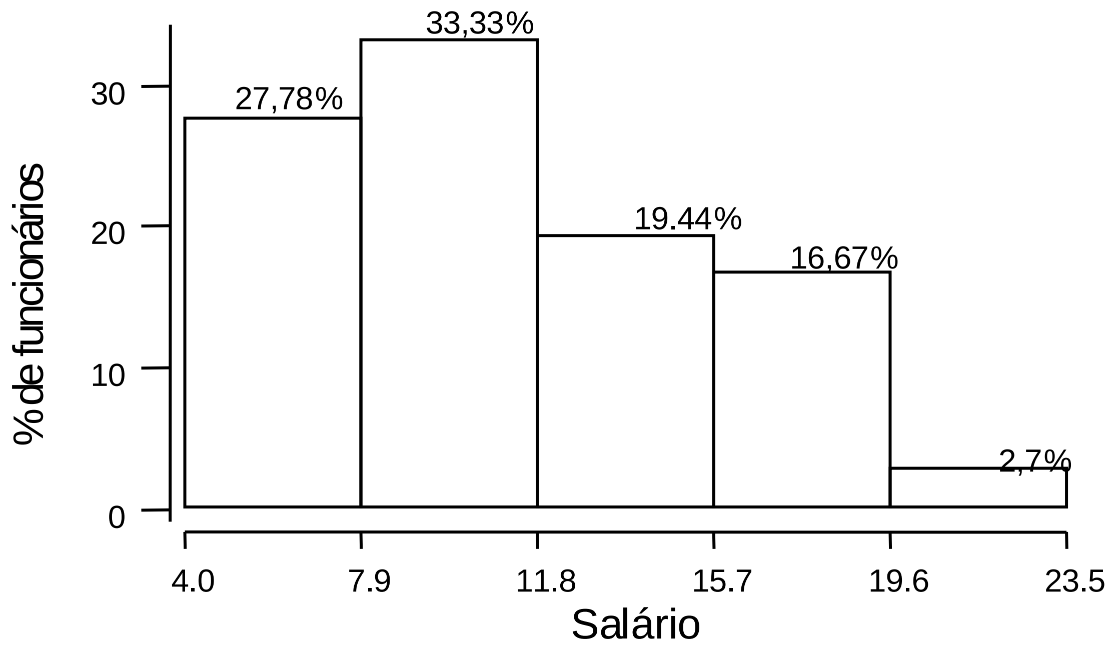
De forma similar obtém-se os outros pontos médios.

Tabela: Distribuição de frequências da variável salário.

i	*Intervalos de classe*	*Ponto médio (X'_i)*	*Freqüência Absoluta (f_i)*	*Freqüência Relativa* (f_{r_i})	*Freqüência Acumulada Absoluta (F_i)*	*Freqüência Acumulada Relativa* (F_{r_i})
1	4,0 |-- 7,9	5,95	10	0,277778	10	0,277778
2	7,9 |-- 11,8	9,85	12	0,333333	22	0,611111
3	11,8 |-- 15,7	13,75	7	0,194444	29	0,805556
4	15,7 |-- 19,6	17,65	6	0,166667	35	0,972222
5	19,6 |-- 23,5	21,55	1	0,027778	36	1
	Total		36	1,000000		

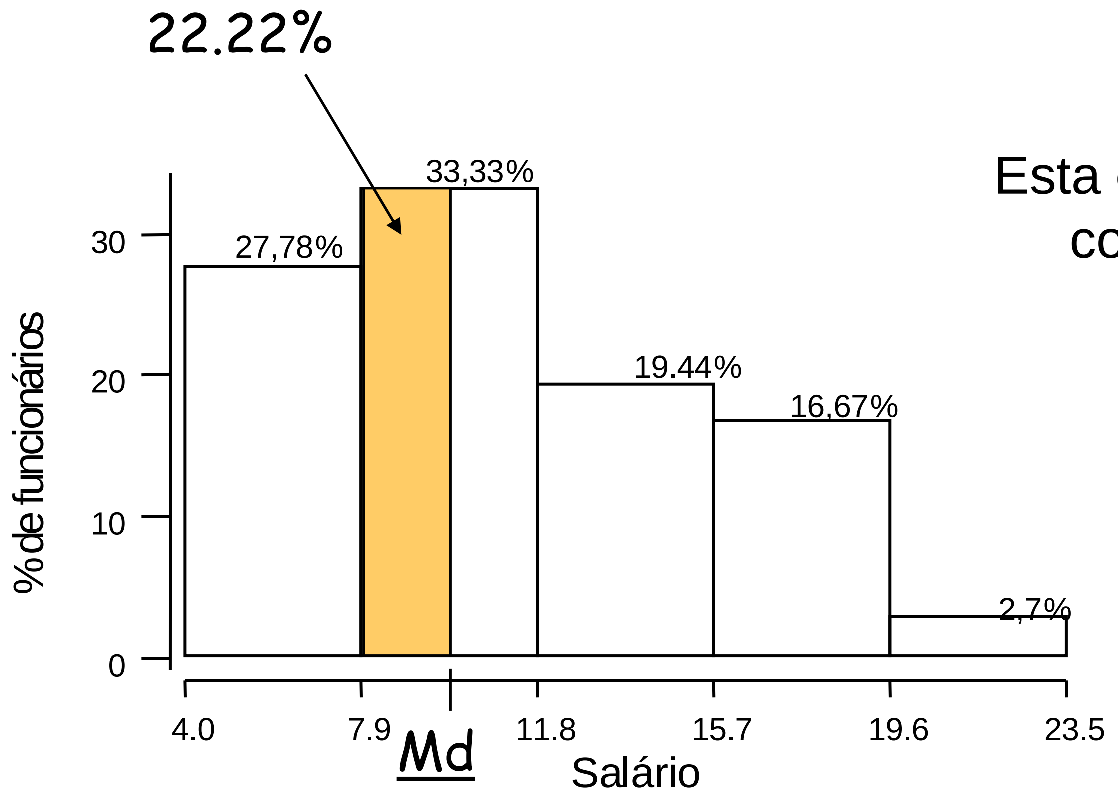
Representação gráfica:

Histograma de frequências relativas (em %) para a variável salário:



Útil para encontrar os percentis: Exemplo Q2 ou Md

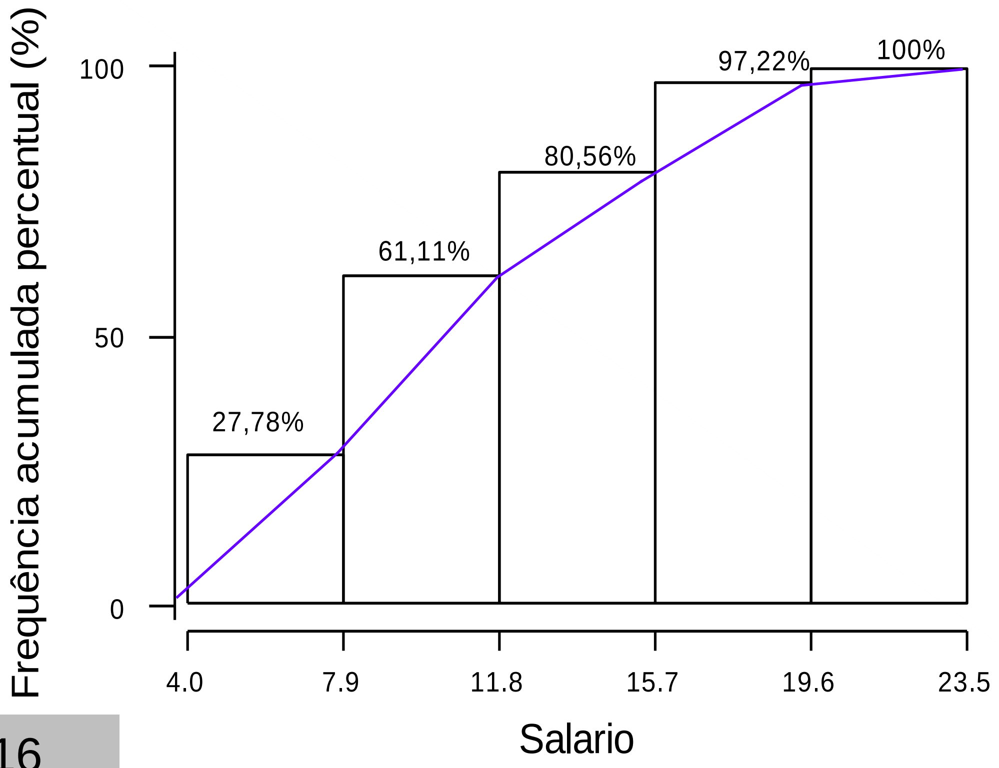
$$\frac{11{,}8-7{,}9}{33{,}33\%}=\frac{Md-7{,}9}{22{,}22}\Rightarrow Md=10{,}5$$



Esta é uma distribuição com assimetria à direita

Histograma de freqüência acumulada relativa (em %)

61% dos empregados tem salário inferior a 12 salarios mínimos



19% possuim salário superior a 16 salários mínimos

Medidas de posição e medidas de dispersão para variáveis contínuas agrupadas em tabela de freqüências.

Média: $$\overline{X} = \frac{X_1' f_1 + X_2' f_2 + \cdots X_k' f_k}{n} = \frac{\sum_{i=1}^{k} X_i' f_i}{n}$$

Exemplo: Considere a tabela anterior

$$\overline{X} = \frac{5{,}95 \times 10 + 9{,}85 \times 12 + 13{,}75 \times 7 + 17{,}65 \times 6 + 21{,}55 \times 1}{36}$$

$$= \frac{401{,}4}{35} = \quad 11{,}15$$

Se calculamos a média para dados não agrupados apresentadas anteriormente resulta:

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_{36}}{36} = \frac{4 + 4{,}36 + \cdots + 23{,}30}{36} = 11{,}122$$

Este resultado difere do valor obtido anteriormente. Porque?

- **Moda (mo):** $$mo = LI_i + \left(\frac{d_1}{d_1 + d_2}\right) \times h$$

i : Classe modal (é aquela classe que tem maior frequência absoluta (f_i))

LI_i : é o limite inferior da classe modal.

$d_1 = f_i - f_{i-1}$

$d_2 = f_i - f_{i+1}$

h : comprimento do intervalo de classe.

Exemplo: Considere a tabela com os salários:

Já que, $f_2 = 12 > f_j \quad j \neq 2 \Rightarrow$ i =2, é a classe modal

$$mo = LI_2 + \left(\frac{d_1}{d_1 + d_2}\right) \times h = 7{,}9 + \left(\frac{12-10}{(12-10)+(12-7)}\right) \times 3{,}9 = 9{,}014$$

Mediana (Md)

$$Md = LI_i + \left(\frac{0{,}5n - F_{i-1}}{f_i} \right) \times h$$

i : é a classe médiana (é o intervalo de classe onde a coluna dos F_i na TDF superou o 50% dos dados)

LI_i : Limite inferior da classe mediana.

$F_{i\text{-}1}$: é a frequência acumulada absoluta da classe anterior a classe mediana

f_i : frequência absoluta da classe mediana.

h : comprimento do intervalo de classe.

Exemplo: Considere a tabela dos salários:

Já que, $F_2 = 22 > n/2 \Rightarrow$ i =2, é a classe mediana

$$Md = LI_2 + \left(\frac{0{,}5n - F_1}{f_1} \right) \times h = 7{,}9 + \left(\frac{18-10}{12} \right) \times 3{,}9 = 8{,}55$$

Variância:

$$S^2 = \frac{\sum_{i=1}^{k} f_i \left(X_i^{'} - \bar{X}\right)^2}{n-1}$$

Exemplo: Considere a tabela 2.2. Vimos que $\bar{X} = 11{,}15$

i	*Intervalos de classe*	X'_i	f_i	$f_i\left(X_i^{'} - \bar{X}\right)^2$
1	4,0 |-- 7,9	5,95	10	270,40
2	7,9 |-- 11,8	9,85	12	20,28
3	11,8 |-- 15,7	13,75	7	47,32
4	15,7 |-- 19,6	17,65	6	253,50
5	19,6 |-- 23,5	21,55	1	108,16
	Total		36	699,66

$$S^2 = \frac{\sum_{i=1}^{5} f_i \left(X_i^{'} - \bar{X}\right)^2}{36-1} = \frac{699{,}66}{35} = 19{,}99029 \Rightarrow \quad \mathrm{S} = 4{,}47105\ (\text{Desvio Padrão})$$

Distribuições de Frequência

Distribuição de Frequência
Variável discreta

- Frequência – é a quantidade de vezes que a variável (evento) ocorre. Por outras palavras, é a quantidade de vezes em que o atributo de qualidade aparece. Exemplo: Experimento avaliação do Transporte: (bom) (regular) (ruim).
- Intervalo de Classe de uma variável discreta: São os atributos variáveis de qualidade atribuídos à variável. Os intervalos são os próprios atributos. Exemplo: o atributo (bom) é em si mesmo uma classe da variação qualitativa.
- Limites de Classe – Não existem limites de classe.

Distribuição de Frequência
Variável Contínua

- Frequência – é a quantidade de vezes que a variável (evento) ocorre . Em outras palavras: a frequência é nº de vezes em que a variável assume um certo valor.

- Intervalo de Classe : É obtido dividindo o conjunto de valores assumidos pela variável em intervalos de classe e indicando a frequência dos valores observados para cada intervalo.

- Limites de Classe – É o intervalo entre o valor máximo e mínimo de uma variável dentro de uma classe. A cada intervalo de classe estão associados os limites de classe (valores extremos) e o ponto médio.

- Frequência Absoluta - valor total das observações
- Frequência Relativa - Valor percentual das observações
- Frequência Acumulada - Somatória das frequências de todos intervalos

Modalidade Esportiva	Frequencia Relativa	Frequência Absoluta	Frequência Acumulada	Frequência Acumulada
Futebol	35%	70	70	35%
Vôlei	25%	50	120	60%
Basquete	20%	40	160	80%
Natação	10%	20	180	90%
Tênis	8%	15	195	98%
Ciclismo	2%	5	200	100%
	100%	200	200	100%

Histograma

- Histograma: Gráfico das distribuições das frequências de uma variável. As frequências dos fenómenos são proporcionais à superfície de cada retângulo que as representam. Para intervalos de mesma amplitude as frequências serão proporcionais às alturas.

- Gráfico de Barras (Histograma) – Gráfico de retângulos, diagrama de colunas.

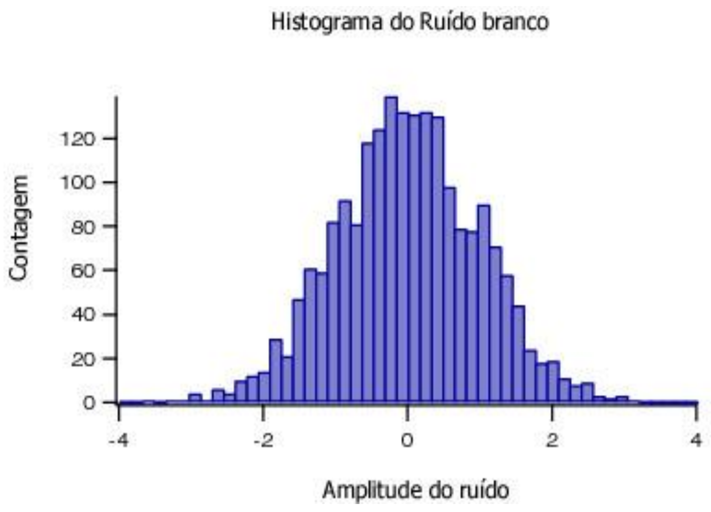
Histograma

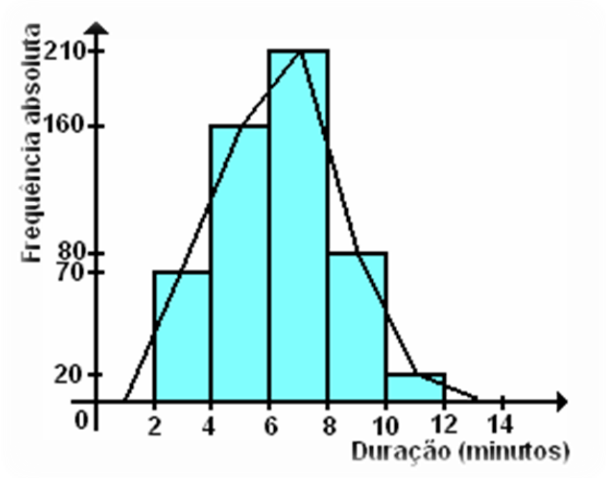
- Distribuições Simétricas e Assimétricas - Os histogramas podem apresentar distribuição simétricas ou assimétricas.

- Polígono de Frequências – Unindo-se os valores médios dos intervalos de classe, o histograma se transforma num *polígono de frequências*. Pode-se então comparar este polígono com uma curva teórica (Normal).

Representação gráfica de uma distribuição de frequências por meio de retângulos justapostos.

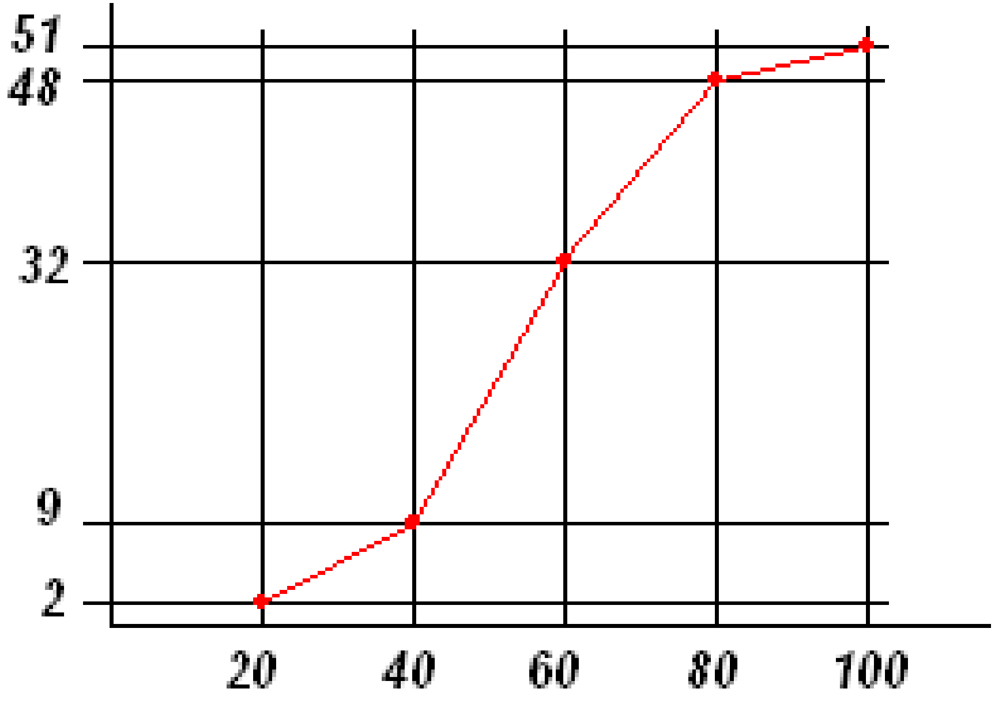
A distribuição de frequência é o método mais útil para descrever resultados obtidos com respeito a uma variável.





Poligono de Frequências acumuladas

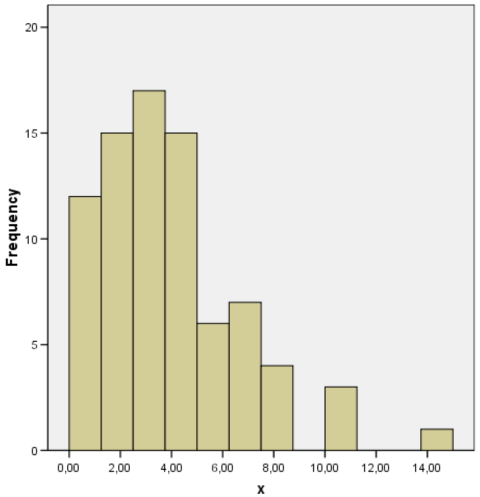
Notas	Nº de Alunos
0 |-- 20	2
20 |-- 40	7
40 |-- 60	23
60 |-- 80	16
80 |-- 100	3
Total	**51**



- Uma distribuição possui assimetria positiva (alternativamente negativa) quando existe uma concentração de valores na zona de valores mais reduzidos (alternativamente elevados) da amostra.

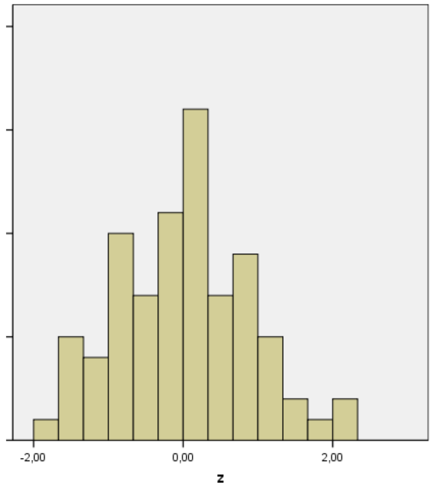
Assimetria positiva

Coef.ass. >0



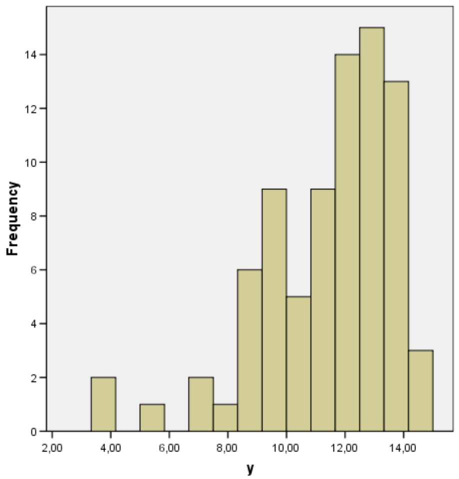
Quase simetria

Coef.ass. ~ 0



Assimetria negativa

Coef.ass. <0



Coeficiente de assimetria

- É uma medida que assume o valor zero quando a distribuição de frequências da amostra é completamente simétrica e assume valores diferentes de zero (positivos ou negativos) quando a distribuição não é simétrica.

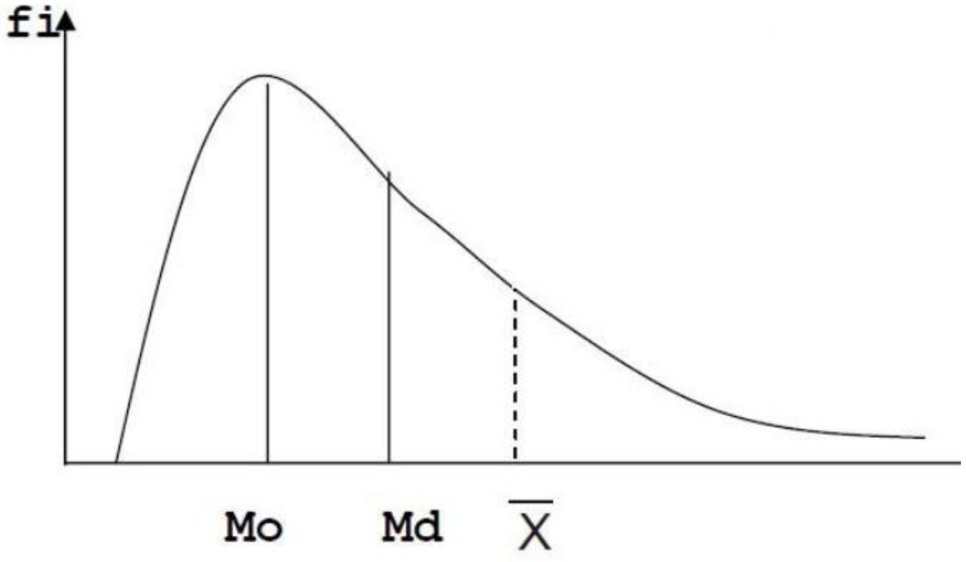
- NOTA: Atenção que numa amostra é quase impossível observar simetria pura. Por isso o coeficiente de assimetria assume valores quase sempre diferentes de zero. Para termos uma ideia se a assimetria é relevante devemos comparar o valor do coeficiente com o erro associado. Se o coeficiente não exceder 2 ou 3 vezes o erro, o seu valor não será muito relevante, especialmente quando queremos extrapolar para a população.

Assimetria positiva:	média > mediana > moda
Assimetria negativa:	média < mediana < moda
Simetria pura:	média = mediana = moda
Simetria aproximada:	média ~ mediana ~ moda

- A assimetria representa a concentração dos valores em um dos extremos da distribuição;

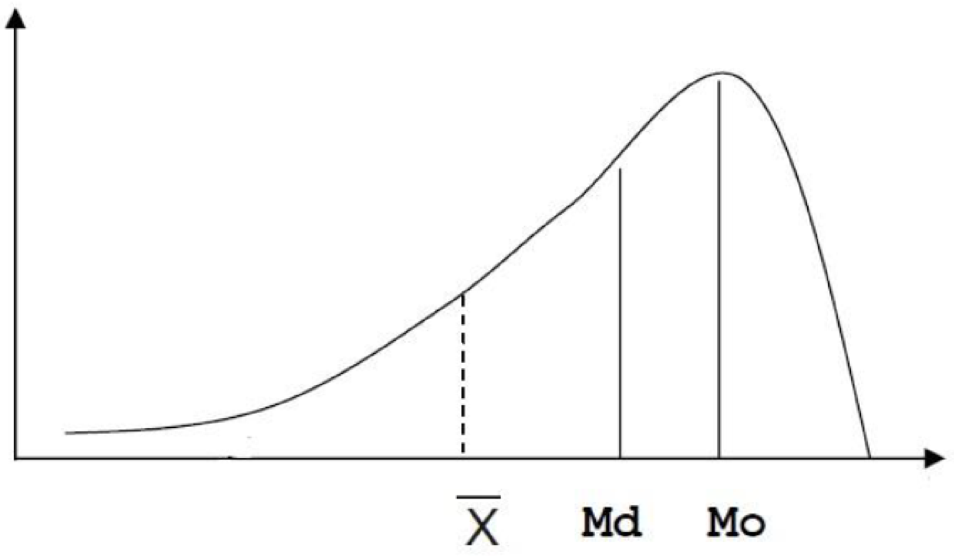
- Neste caso, a distribuição se comporta de forma simétrica, onde $\bar{x} = Moda = Mediana$.

- Assimetria positiva;



- $\bar{x} > Mediana > Moda.$

- Assimetria negativa;



- $\bar{x} < Mediana < Moda.$

- O coeficiente de assimetria pode ser calculado da seguinte maneira:

$$Assimetria = \frac{\bar{x} - Mo}{s}$$

Onde:

- $\bar{x}$: média amostral;
- Mo: moda;
- s: desvio padrão.

- Nas distribuições considera-se:

  Simétrica:

  $|As| < 0{,}15;$

  Assimétrica moderada:

  $0{,}15 \leq |As| < 1;$

  Fortemente Assimétrica:

  $|As| \geq 1.$

- Caso trate-se de uma amostra amodal, o coeficiente de assimetria pode ser determinado utilizando a mediana:

$$Assimetria = \frac{3 * (\bar{x} - Md)}{s}$$

Onde:

- Md: mediana;
- $\bar{x}$: média amostral;
- s: desvio padrão.

- Nas distribuições considera-se:

  Simétrica:

  $|As| < 0{,}15;$

  Assimétrica moderada:

  $0{,}15 \leq |As| < 1;$

  Fortemente Assimétrica:

  $|As| \geq 1.$

Curtose

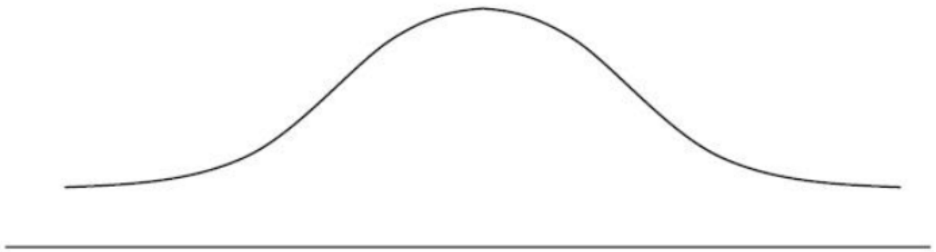
- A curtose representa o grau de achatamento da distribuição;
- É determinado por:

$$k = \frac{1}{2} * \frac{(Q_3 - Q_1)}{P_{90} - P_{10}}$$

Onde:

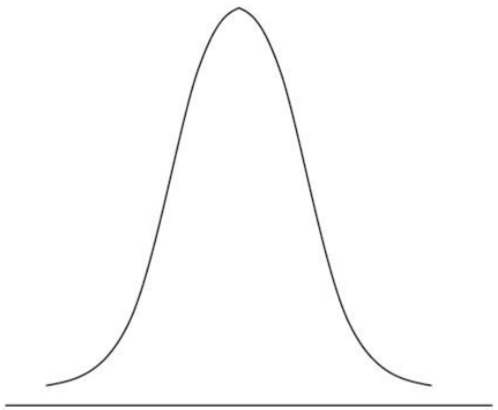
- Q_3 e Q_1: 3º e 1º quartis;
- P_{90} e P_{10}: 10º e 90º percentis.

- O coeficiente de curtose para a distribuição normal é 0,263. A curva de distribuição normal é denominada Mesocúrtica.

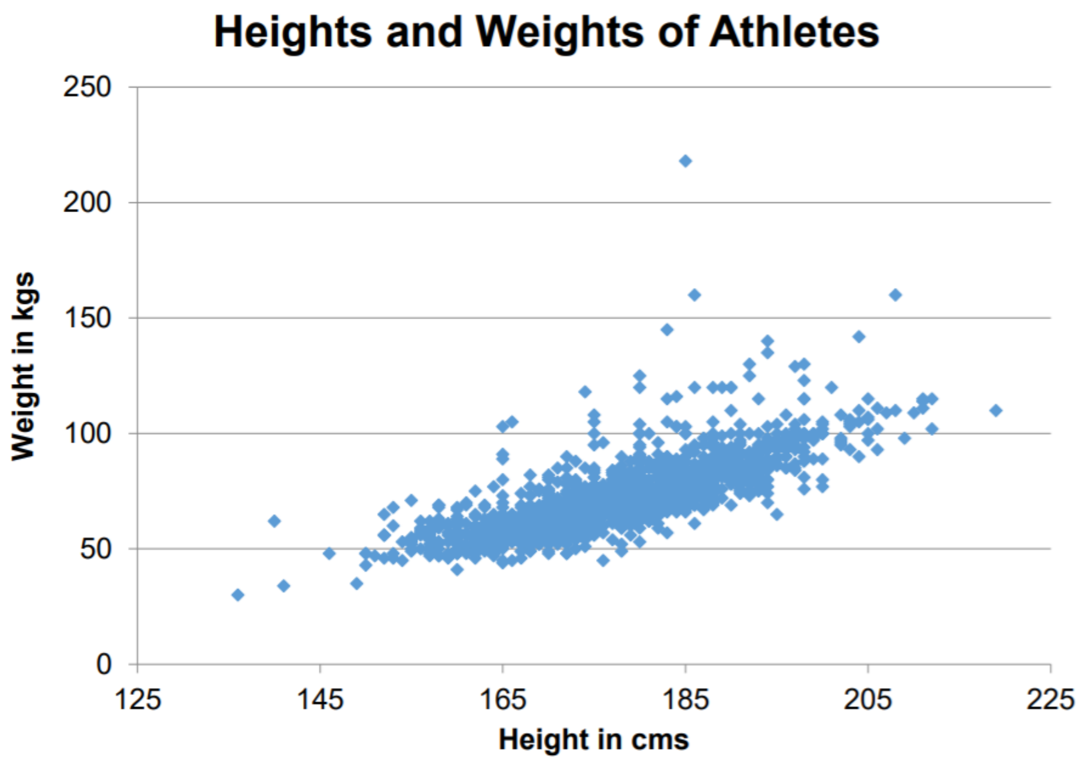


- Quando o coeficiente de curtose $k > 0{,}263$, a distribuição é mais achatada, denominada Platicurtica.

- Quando o coeficiente de curtose $k < 0{,}263$, a curva é mais alongada, denominada Leptocurtica.



Estatística descritiva vista entre duas variáveis



Estatística descritiva vista entre duas variáveis: Covariância

Quando se consideram pares de variáveis (X,Y) deixamos de estar interessados em explorar isoladamente apenas cada uma das variáveis. O objetivo passa a ser o estudo da variação conjunta dessas variáveis, tentando pôr em evidência as "relações" eventualmente existentes entre elas. Para além dos indicadores individuais que já estudamos até aqui, podemos analisar a relação estatistica entre duas variáveis numa amostra bivariada.

Através da análise da **Covariância**:

$$cov(x,y)=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1}=\frac{n\sum_{i=1}^{n}x_i y_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n(n-1)}$$

Função Excel

Covariance.s (range, range2)

Estatística descritiva vista entre duas variáveis: Correlação

$$\text{Correlation} = \frac{Covariance(X,Y)}{Stdev(X)Stdev(Y)}$$

O intervalo da correlação está entre -1 e +1

Função Excel

correl (range1, range2)

02

Introdução à Teoria da Probabilidade

Teoria das Probabilidades

A teoria das probabilidades procura estimar as possibilidades de ocorrer um determinado acontecimento.

É um ramo da matemática que cria, elabora e pesquisa modelos para estudar acontecimentos ou fenómenos aleatórios.

Um exemplo clássico de acontecimento aleatório é o lançamento de um dado ou de uma moeda. No caso do dado, os resultados possíveis são os números de 1 a 6. No caso da moeda, cara ou coroa. A partir daí, podemos definir espaço amostral:

Espaço Amostral

<ins>Experimento aleatório</ins>: É um acontecimento que pode apresentar resultados diferentes, quando repetido nas mesmas condições.

<ins>Espaço amostral</ins>: É o conjunto de todos os resultados possíveis de um acontecimento aleatório. Indicamos o espaço amostral por S.

Um Exemplo

Vamos analisar alguns fenómenos aleatórios. Os dados dos experimentos são os comuns, de 6 faces.

1) Lançamento de um dado e registro do resultado

Conjunto de todos os resultados possíveis: {1, 2, 3, 4, 5, 6}

Um dos subconjuntos dele é {1, 3, 5}, que pode ser identificado por "ocorrer número ímpar no lançamento de um dado".

- **espaço amostral:** S = {1, 2, 3, 4, 5, 6}
- **evento E:** "ocorrer número ímpar no lançamento de um dado" → A = {1, 3, 5}

Eventos: certo, impossível

Evento certo: Ocorre quando um evento coincide com o espaço amostral.

Evento impossível: Ocorre quando um evento é vazio.

Probabilidade de Ocorrer um Evento

$$P(A) = \frac{\text{número de elementos de A}}{\text{número de elementos de } \Omega} \Rightarrow P(A) = \frac{n(A)}{n(\Omega)}$$

NOTA: P(A) possui um intervalo fixo: $0 \leq P(A) \leq 1$. Quando P(A) = 0, o evento é impossível. Quando P(A) = 1, o evento é certo.

Exemplo: Consideremos o experimento Aleatório do lançamento de um moeda perfeita. Calcule a probabilidade de sair cara.

Espaço amostral: $\Omega = \{\text{cara, coroa}\} \Rightarrow n(\Omega) = 2$

Evento A: $A = \{\text{cara}\} \Rightarrow n(A) = 1$

Temos P(A) = 50%

Exemplo:

No lançamento simultâneo de 3 moedas perfeitas distinguíveis, qual é a probabilidade de serem obtidas:

a) Pelo menos 2 caras?

b) Exatamente 2 caras?

C = cara K = coroa

$\Omega = \{CCC, CCK, CKC, CKK, KCC, KCK, KKC, KKK\} \Rightarrow$

$n(\Omega) = 8$

$A = \{CCC, CCK, CKC, KCC\} \Rightarrow n(A) = 4$

Exemplo:

a) A = {CCC, CCK, CKC, KCC} $\Rightarrow$ n(A) = 4

b) B = {CCK, CKC, KCC} $\Rightarrow$ n(B) = 3

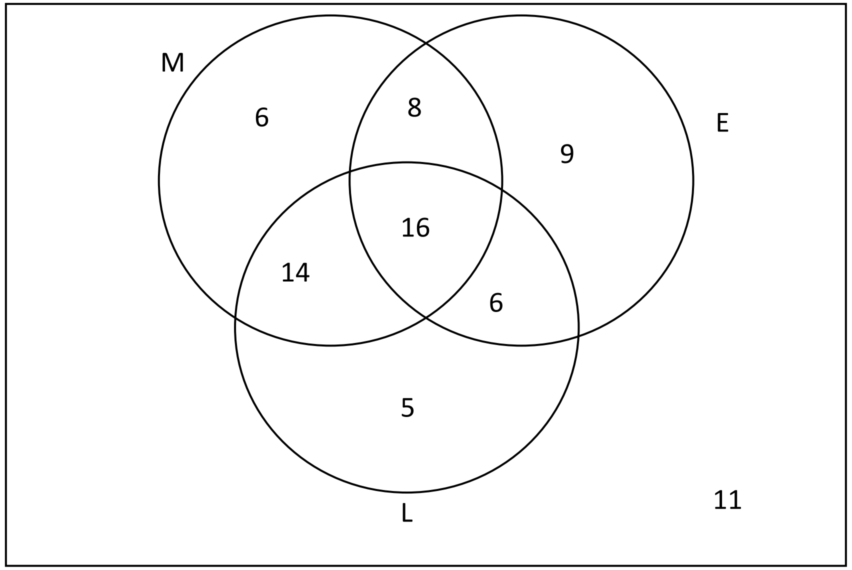
$$P(B) = \frac{3}{8} = 0{,}375 = 37{,}5$$

Exemplo:

Num grupo de 75 jovens, 16 gostam de música, desporto e leitura; 24 gostam de música e desporto; 30 gostam de música e leitura; 22 gostam de desporto e leitura; 6 gostam somente de música; 9 gostam somente de desporto e 5 gostam somente de leitura. Calcule a probabilidade de escolher, ao acaso, um dos seguintes jovens:

a) ele gostar de música;

b) ele não gostar de nenhuma dessas atividades.



$n(\Omega) = 75$

gostam de música: 6 + 8 + 16 + 14 = **44**

não gostam de nenhuma dessas atividades:

75 – (6 + 9 + 5 + 8 + 6 + 14 + 16) = 75 – 64 = **11**

a) a probabilidade de gostar de música:

$$P(A) = \frac{n(A)}{n(\Omega)} = \frac{44}{75} \cong 58\%$$

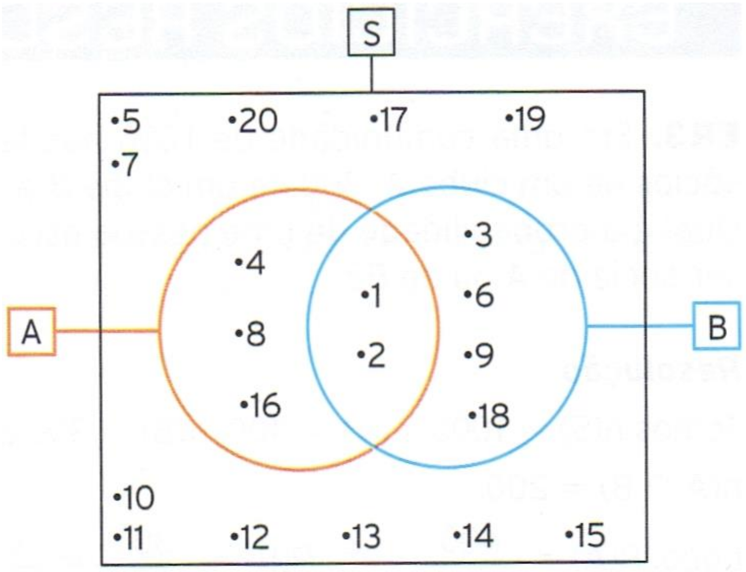
b) probabilidade de não gostar de nenhuma dessas atividades:

$$P(B) = \frac{n(B)}{n(\Omega)} = \frac{11}{75} \cong 14\%$$

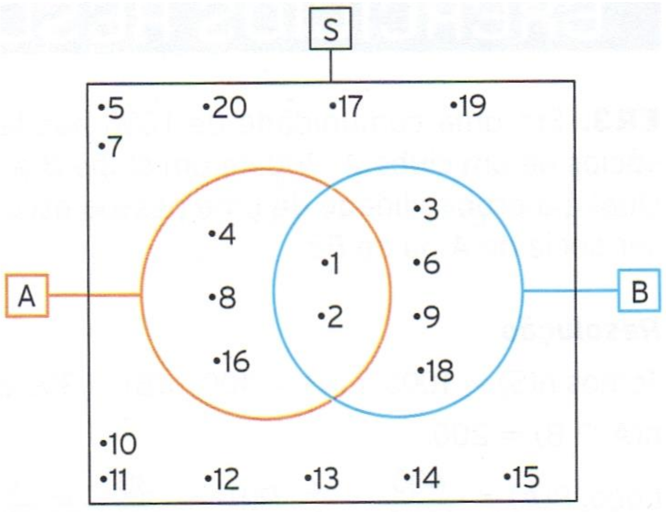
Probabilidade de União de Eventos

Vamos retirar uma bola de uma urna que contém 20 bolas numeradas de 1 a 20 e considerar os eventos *A* - “obtenção de divisor de 16” e *B* - “obtenção de divisor de 18”. Temos então:

- S = {1, 2, 3, ..., 20}
- n(S) = 20
- A = {1, 2, 4, 8, 16} , n(A) = 5
- B = {1, 2, 3, 6, 9, 18}, n(B) = 6



Probabilidade de União de Eventos



Note que existem elementos que satisfazem:

- apenas o evento *A*: 4, 8, 16
- apenas o evento *B*: 3, 6, 9, 18
- o evento *A* e o evento *B*: 1, 2
- o evento *A* ou o evento *B*: 4, 8, 16, 3, 6, 9, 18, 1, 2

Sabemos que {1, 2} = A ∩ B e {4, 8, 16, 3, 6, 9, 18, 1, 2} = A ∪ B.
Podemos então dizer que:

- a ocorrência do *evento A* ***e*** *do evento B* é dada por A ∩ B
- a ocorrência do *evento A* ***ou*** *do evento B* é dada por A ∪ B

Probabilidade de União de Eventos

Vamos calcular P(A), P(B), P(A ⋂ B), P(A ⋃ B):

$$P(A) = \frac{n(A)}{n(S)} \Rightarrow P(A) = \frac{5}{20}$$

$$P(B) = \frac{n(B)}{n(S)} \Rightarrow P(B) = \frac{6}{20}$$

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

$$\Rightarrow P(A \cap B) = \frac{2}{20}$$

$$P(A \cup B) = \frac{n(A \cup B)}{n(S)}$$

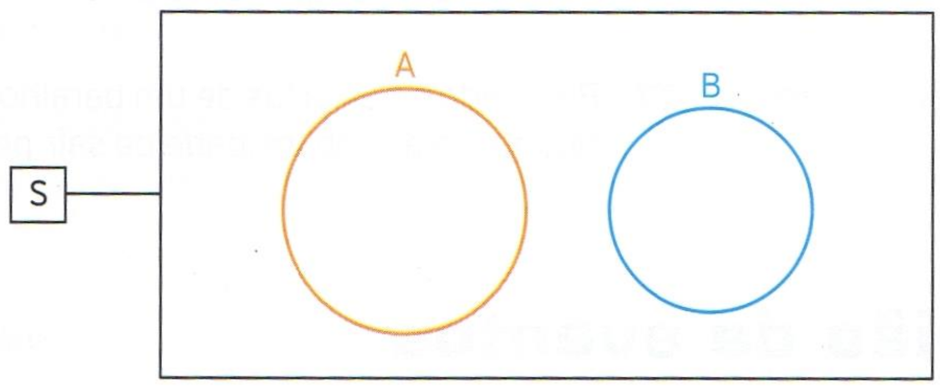
$$\Rightarrow P(A \cup B) = \frac{9}{20}$$

Esses resultados mostram que P(A ⋃ B) = P(A) + P(B) – P(A ⋂ B).

Probabilidade Eventos mutuamente exclusivos

Se *A* e *B* são conjuntos disjuntos, isto é, $A \cap B = \emptyset$, os eventos *A* e *B* são ditos **mutuamente exclusivos**.

Nesse caso, como $n(A \cap B) = 0$ e $P(A \cap B) = 0$, vem $P(A \cup B) = P(A) + P(B)$.



Exercício:

Ao retirar uma carta de um baralho de 52 cartas, qual é a probabilidade de que essa carta seja vermelha ou um ás?

$n(\Omega) = 52$

Evento A: a carta é vermelha $\Rightarrow$ $n(A) = 26$

Evento B: a carta é ás $\Rightarrow$ $n(B) = 4$

$n(A \cap B) = 2$

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

$$P(A \cup B) = \frac{26}{52} + \frac{4}{52} - \frac{2}{52}$$

$$P(A \cup B) = \frac{28}{52}$$

$$P(A \cup B) = \frac{7}{13} \cong 53{,}8\%$$

Probabilidade de Evento Complementar

A probabilidade de não ocorrer o evento A é a probabilidade de ocorrer o evento complementar de A, representado por .

Nessas condições, temos :

$$A \cup \overline{A} = \Omega \text{ e } A \cap \overline{A} = \varnothing$$

Então,

$$P(\Omega) = P(A \cup \overline{A})$$

$$1 = P(A) + P(\overline{A}) \qquad P(\overline{A}) = 1 - P(A)$$

Eventos Independentes e Probabilidade Condicionada

A e B são independentes se a ocorrência de um deles não altera a probabilidade de ocorrência do outro. Formalmente:

$$P(A \mid B) = P(A)$$

Pela expressão anterior, se A e B são independentes: $P(A \cap B) = P(A).P(B)$

Nota: neste caso $A \cap B$ denota a possibilidade de ocorrência simultânea dos dois eventos

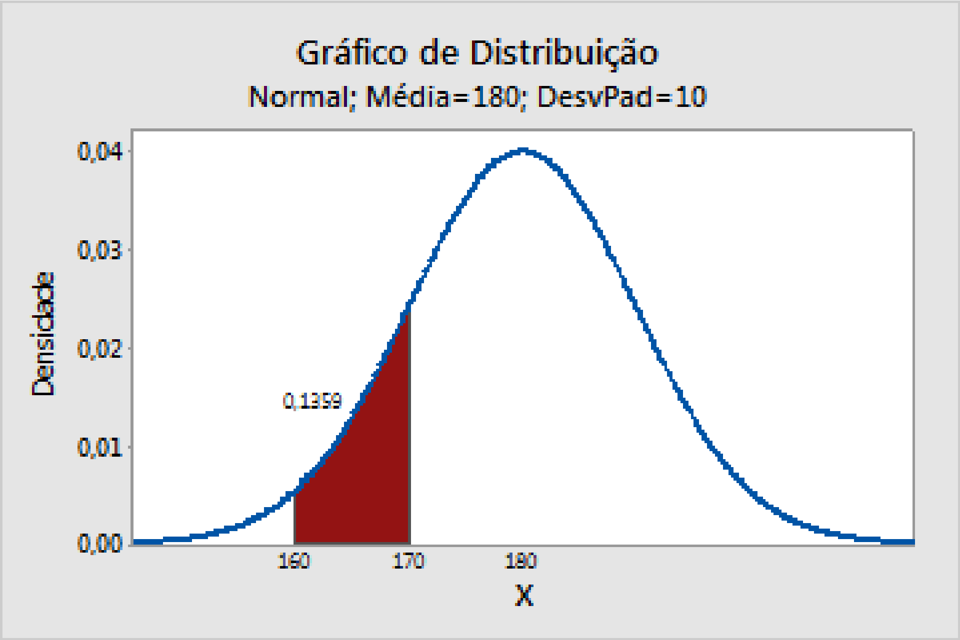
Probabilidade de um evento A, dado que aconteceu um outro evento B

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

03

Distribuições

Discretas Univariadas

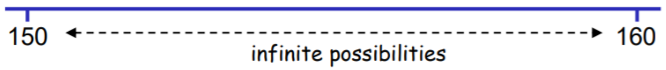


Uma *distribuição de probabilidade* é um modelo matemático que relaciona um certo valor da variável em estudo com a sua probabilidade de ocorrência.

Há dois tipos de distribuição de probabilidade:

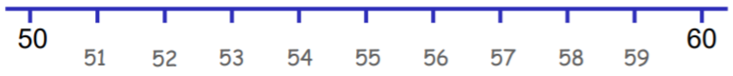
1. Distribuições Contínuas: Quando a variável que está sendo medida é expressa em uma escala continua:

(heights of men and women)



2. Distribuições Discretas: Quando a variável que está sendo medida só pode assumir certos valores, como por exemplo os valores inteiros:

Test of Discreteness
(number of students in a class)



Variável Discreta

Função de probabilidade: É a função que atribui a cada valor x_i da variável discreta X a sua probabilidade de ocorrência e pode ser representada pela tabela:

x	x_1	x_2	...	x_n
$P(X=x)$	$P(X=x_1)$	$P(X=x_2)$	...	$P(X=x_n)$

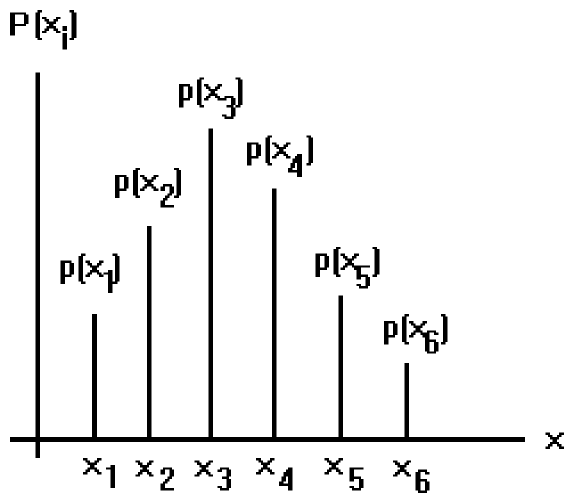
Uma função de probabilidade deve satisfazer as seguintes condições:

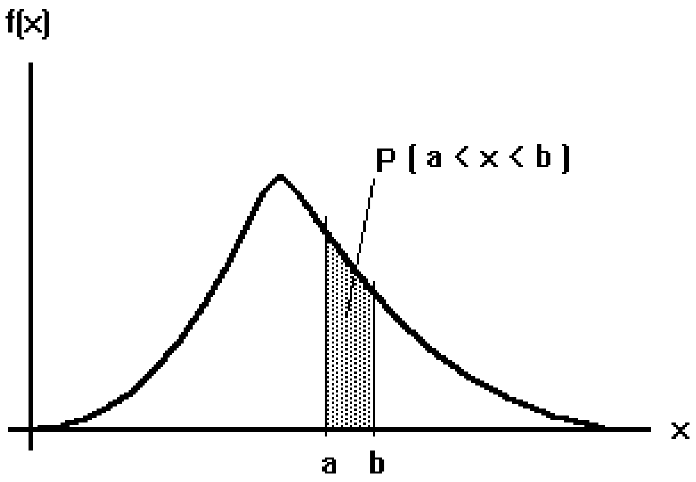
$$0 \leq P(X = x_i) \leq 1 \quad \text{e} \quad \sum_{i=1}^{n} P(X = x_i) = 1$$

No caso de distribuições discretas, a probabilidade de que a variável X assuma um valor específico x_o é dada por: $P(X = x_o) = P(x_o)$

No caso de variáveis contínuas, as probabilidades são especificadas em termos de intervalos, pois a probabilidade associada a um número específico é zero.

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





Média

Valor Esperado (média): Dada a variável X, assumindo os valores x_1, x_2, ..., x_n, chamamos de *valor médio,* ou *valor esperado,* ou *esperança matemática* de X o valor

$$E(X) = x_1 \times P(X = x_1) + \ldots + x_n \times P(X = x_n) = \sum_{i=1}^{n} x_i \times P(X = x_i)$$

Notação: $\mu = E(X)$

Variância

É o valor esperado da variável $(X - E(X))^2$, ou seja, se X assume os valores $x_1, x_2, \ldots, x_n$, então:

$$\mathrm{Var}(X) = \sum_{i=1}^{n} [x_i - E(X)]^2 \times P(X = x_i)$$

Notação: $\sigma^2 = \mathrm{Var}(X)$.

Desvio Padrão: É definido como a raiz quadrada positiva da variância, isto é,

$$\mathrm{DP}(X) = \sqrt{\mathrm{Var}(X)}.$$

Notação: $\sigma = \mathrm{DP}(X)$.

Função de Distribuição Acumulada

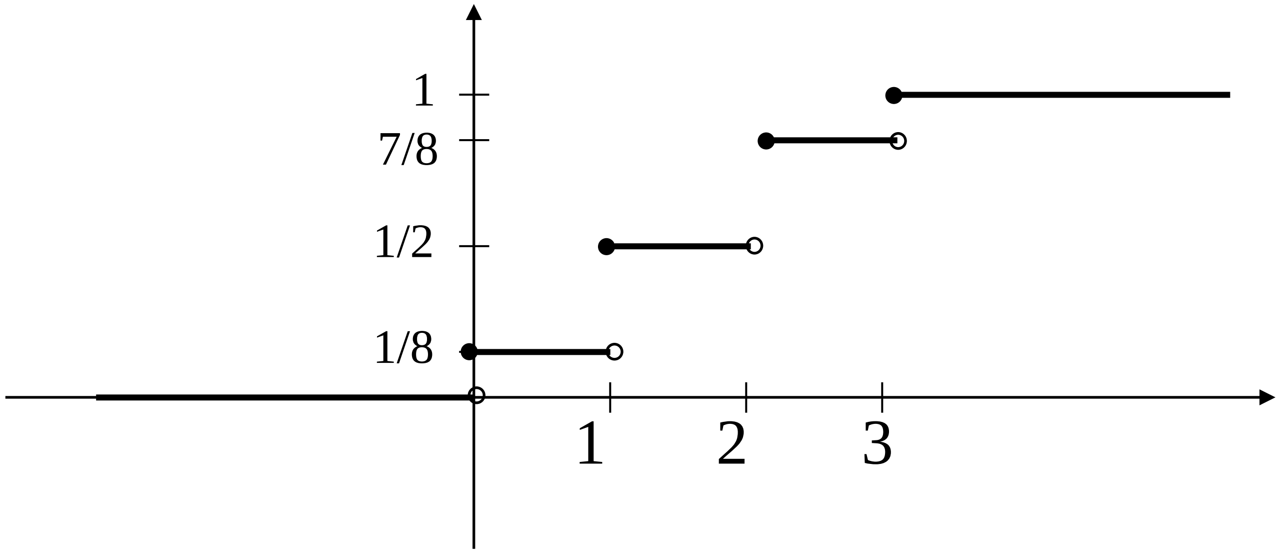
- A função de distribuição acumulada de uma variável aleatória X é a função F_X: **R**$\rightarrow$**R** definida por

$$F_X(x) = P(X \leq x)$$

Função de Distribuição Acumulada

Exemplo

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



Se $x < 0$: $P(X \leq x) = 0$

Se $0 \leq x < 1$: $P(X \leq x) = P(X=0) = 1/8$

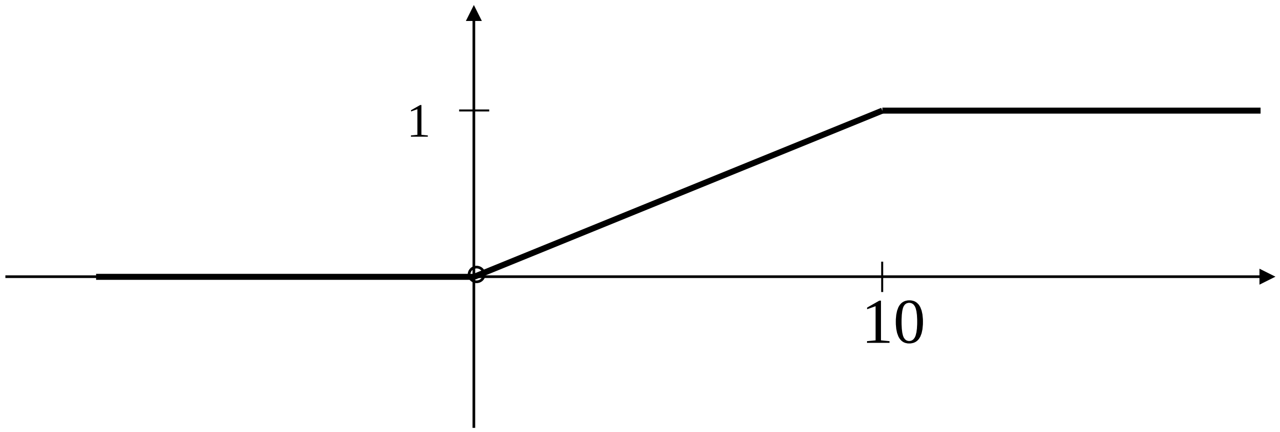
Se $1 \leq x < 2$: $P(X \leq x) = P(X=0 \text{ ou } X=1) = 1/8 + 3/8 = 1/2$

Função de Distribuição Acumulada

- Roleta numerada continuamente de 0 a 10

  X = prêmio ganho

$$P(X \leq x) = \begin{cases} 0, \text{ se } x < 0 \\ x/10, \text{ se } 0 \leq x \leq 10 \\ 1, \text{ se } x > 10 \end{cases}$$



Principais Distribuições Discretas

- Distribuição de Bernoulli
- Distribuição Binomial
- Distribuição Hipergeométrica
- Distribuição de Poisson

Distribuição de Bernoulli

Alguns exemplos:

- Lançar uma moeda e observar se ocorre cara ou coroa;
- Lançar um dado e observar se ocorre seis ou não;
- Numa linha de produção, observar se um item, tomado ao acaso, é defeituoso ou não defeituoso;
- Verificar se um servidor de intranet está ativo ou não ativo;
- O resultado de um exame médico para detecção de uma doença é positivo ou negativo;
- Um paciente submetido a um tratamento, durante um período de tempo fixo, cura-se ou não da doença;
- Um entrevistado concorda ou não com uma afirmação feita.

Distribuição de Bernoulli

Denominamos sucesso e fracasso aos dois resultados possíveis em cada caso.

O ensaio de Bernoulli é caracterizado por uma variável aleatória X, definida por X=1, se sucesso; X=0, se fracasso.

A função de probabilidade de X (Distribuição de Bernoulli) é dada por:

x	***p(x)***
0	1-p
1	p
Total	1

Onde p= P{Sucesso}

"X ~ Bernoulli (p)" indica uma variável com *distribuição de Bernoulli* com parâmetro p, isto é,

$$X = \begin{cases} \mathbf{1}, \text{ se ocorrer "sucesso"} \\ \mathbf{0}, \text{ se ocorrer "fracasso"} \end{cases}$$

e sua função de probabilidade pode ser representada pela tabela:

X	1	0
$P(X=x)$	p	$1 - p$

Temos que esperança e variância são dadas por:

$$E(X) = p,$$
$$\mathrm{Var}(X) = p(1 - p).$$

Repetições independentes de um ensaio de Bernoulli, com a mesma probabilidade de ocorrência de "sucesso", dão origem ao modelo de probabilidade binomial.

A distribuição binomial é adequada para descrever situações em que os resultados de uma variável aleatória podem ser agrupados em apenas duas classes ou categorias.

As categorias devem ser mutuamente excludentes, de forma que não haja dúvidas na classificação do resultado da variável nas categorias, devendo ser exaustivas, de forma que não seja possível nenhum outro resultado diferente das categorias.

Por exemplo, um produto manufaturado pode ser classificado como perfeito ou defeituoso, a resposta de um questionário pode ser verdadeira ou falsa, as chamadas telefónicas podem ser locais ou interurbanas.

Condições de aplicação da distribuição Binomial:

- são feitas *n* repetições do experimento, onde n é uma constante;
- há apenas dois resultados possíveis em cada repetição, denominados sucesso e falha;
- a probabilidade de sucesso *(p)* e de falha *(1- p)* permanecem constante em todas as repetições;
- as repetições são independentes, ou seja, o resultado de uma repetição não é influenciado por outros resultados.

Mesmo variáveis contínuas podem ser divididas em duas categorias.

Exemplo: a velocidade de um automóvel pode ser classificada como dentro ou fora do limite legal.

Geralmente, denomina-se as duas categorias como *sucesso* ou *falha*. Como as duas categorias são mutuamente excludentes e coletivamente exaustivas:

P (sucesso) + P(falha) = 1

Consequentemente, sabendo-se que, por exemplo, a probabilidade de sucesso é *P(sucesso) = 0,6*, a probabilidade de falha é *P(falha) = 1-0,6 = 0,4.*

A variável *X* correspondente ao número de sucessos em *n* ensaios de Bernoulli independentes e com mesma probabilidade *p* de sucesso tem *distribuição binomial com parâmetros n e p.*

A função de probabilidade é dada por:

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \quad , \quad k = 0, 1, \dots, n.$$

Notação: $X \sim B(n; p)$.

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Os parâmetros da distribuição Binomial são *n* e *p*.

A média e a variância são calculadas como:

$E(x) = np$

$Var(x) = np(1 - p)$

A distribuição Binomial é usada com frequência no controle de qualidade quando a amostragem é feita sobre uma população infinita ou muito grande.

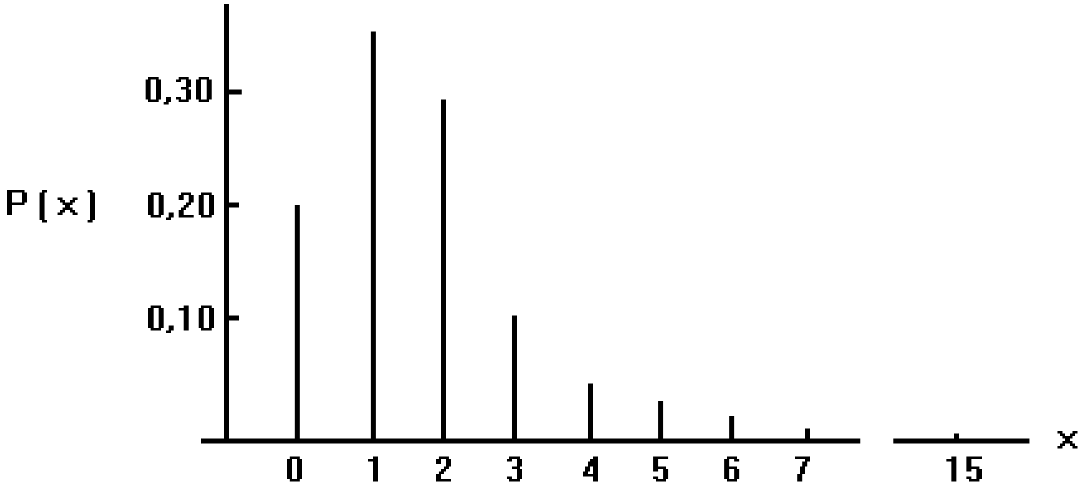
Nas aplicações de controle da qualidade, *x* em geral representa o número de defeituosos observados em uma amostra de *n* itens.

Condições

Por exemplo, se *p = 0,10* e *n = 15*, a probabilidade de obter *x* itens não conformes é calculada usando a equação da Binomial. Por exemplo, para *x=1*

$$\binom{15}{1} = \frac{15!}{1!(15-1)!} = 15$$

$$P(1) = \binom{15}{1} * 0{,}10^{\wedge}1 * (1-0{,}1)^{\wedge}(15-1) = 15*0{,}10*0{,}23=0{,}34$$



Exemplo:

Um dado equilibrado é lançado 3 vezes.
Qual é a probabilidade de se obter a *face 5* duas vezes?

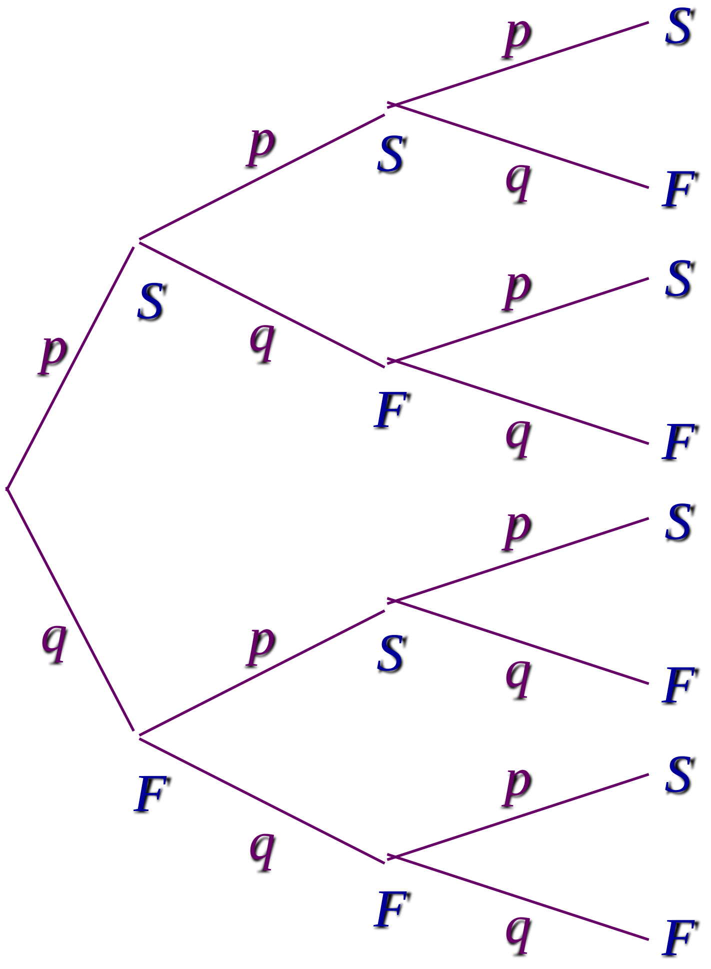
Denotamos,

S: "sucesso", ocorrer face 5;
F: "fracasso", não ocorrer face 5.

É fácil ver que $\boldsymbol{p} = P(\text{sucesso}) = 1/6$ e
$\boldsymbol{q = 1 - p} = P(\text{fracasso}) = 5/6$

$$\Omega = \{SSS, SSF, SFS, FSS, SFF, FSF, FFS, FFF\}$$

Estamos interessados no número total de sucessos que, no caso, é o *número de vezes que a face 5 é observada* nos 3 lançamentos do dado.



Ω	Prob	X
(SSS)	p^3	3
(SSF)	p^2q	2
(SFS)	p^2q	2
(SFF)	pq^2	1
(FSS)	p^2q	2
(FSF)	pq^2	1
(FFS)	pq^2	1
(FFF)	q^3	0

$\Rightarrow$

A função de probabilidade de X é dada por:

Probabilidades binomiais para n = 3 e $P(S) = p$

nº. de sucessos	probabilidades	p = 1/6
0	q^3	125/216=0,5787
1	$3pq^2$	75/216=0,3472
2	$3p^2q$	15/216=0,0694
3	p^3	1/216=0,0046

Podemos escrever essa função como

$$P(X = k) = \binom{3}{k} p^k q^{3-k} \quad , \quad k = 0, 1, 2, 3.$$

No exemplo, para $n = 3$ e $p = 1/6$, $P(X = 2) = 0{,}0694$.

Resultados

É especialmente importante interpretar os resultados. Os valores dizem-se pouco usuais se se encontrarem para além dos seguintes limites:

$$\text{Valores Máximos Usuais} = \mu + 2\,\sigma$$
$$\text{Valores Mínimos Usuais} = \mu - 2\,\sigma$$

Função Excel: BINOM.DIST

=BINOM.DIST(x, n, p, FALSE/TRUE)

=BINOM.DIST(x, n, p, FALSE/TRUE)

FALSE → P(successes = x)

TRUE → P(successes ≤ x)

Qual a probabilidade de ganhar 3 vezes num lançamento de um dado 10 vezes?

P(X=3) = BINOM.DIST(3, 10, 0.1667, FALSE) = 0.1551

Qual a probabilidade de ganhar no máximo 5 vezes num lançamento de um dado 10 vezes?

P(X≤5) = BINOM.DIST(5, 10, 0.1667, TRUE) = 0.9976

Distribuição Hipergeométrica

Considere o problema básico de inspeção por amostragem, em que observamos uma amostra de *n* itens de um lote com *N* itens, sendo *r* defeituosos. Avaliamos o número *X* de itens defeituosos na amostra. A variável aleatória *X* aparenta ser binomial, mas só é realmente binomial se:

- A seleção da amostra for aleatória (para garantir a mesma probabilidade p de sair item defeituoso em todos os ensaios);
- Com reposição (para garantir independência entre ensaios).

Distribuição Hipergeométrica

A segunda condição não costuma ser satisfeita na prática. Se a amostragem for aleatória, mas sem reposição, a distribuição de *X* é conhecida como hipergeométrica de parâmetros *N*, *n* e *r*.

A função de probabilidade de X é expressa por:

$$P(x) = \frac{\binom{r}{x} \cdot \binom{N-r}{n-x}}{\binom{N}{n}}$$

x = 0,1,....,min(r,n)

A média e a variância são calculadas como:

$$\mu = n \cdot p$$

$$\sigma^2 = n \cdot p \cdot (1 - p) \cdot \frac{N-n}{N-1}$$

Exemplo

Em problemas de controle de qualidade, suponha que num lote de N = 100 peças, r = 10 sejam defeituosas. Escolhendo n = 5 peças sem reposição,

a) a probabilidade de não se obter peças defeituosas é:

$$p_0 = \frac{\binom{10}{0}\binom{90}{5}}{\binom{100}{5}} = \frac{\binom{90}{5}}{\binom{100}{5}} \simeq 0,584$$

b) a probabilidade de se obter pelo menos uma defeituosa é:

$$p_1 + p_2 + \ldots + p_5 = 1 - p_0 \simeq 0,426$$

Distribuição de Poisson

A distribuição de Poisson é adequada para descrever situações onde existe uma probabilidade de ocorrência em um campo ou intervalo contínuo, geralmente tempo ou área.

Por exemplo, o nº de acidentes por mês, nº de defeitos por metro quadrado, nº de clientes atendidos por hora.

Nota-se que a variável aleatória é discreta (número de ocorrência), no entanto a unidade de medida é contínua (tempo, área).

Além disso, as falhas não são contáveis, pois não é possível contar o número de acidentes que não ocorreram, nem mesmo o número de defeitos que não ocorreram.

Condições de aplicação:

- o número de ocorrências durante qualquer intervalo depende somente da extensão do intervalo;
- as ocorrências ocorrem independentemente, ou seja, um excesso ou falta de ocorrências em algum intervalo não exerce efeito sobre o número de ocorrências em outro intervalo;
- a possibilidade de duas ou mais ocorrências acontecerem em um pequeno intervalo é muito pequena quando comparada à de uma única ocorrência.

A distribuição de Poissson fica completamente caracterizada por um único parâmetro λ que representa a taxa média de ocorrência por unidade de medida.

A equação para calcular a probabilidade de *x* ocorrências é dada por:

$$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \qquad x = 0, 1, \ldots, n$$

A média e a variância da distribuição de Poisson são:

$$\mu = \lambda$$

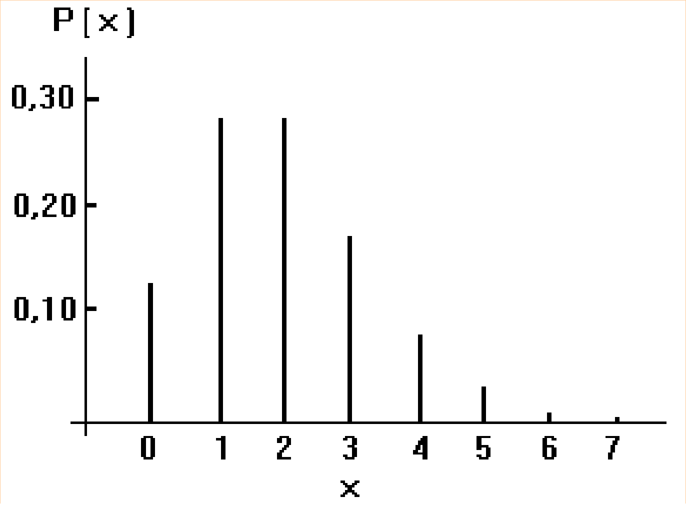
$$\sigma^2 = \lambda^2$$

Exemplo: A aplicação típica da distribuição de Poisson no controle da qualidade é como um modelo para o número de defeitos (não-conformidades) que ocorre por unidade de produto (por m^2, por volume ou por tempo, etc.).

O número de defeitos de pintura segue uma distribuição de Poisson com $\lambda = 2$.

Então, a probabilidade que uma peça apresente mais de 4 defeitos de pintura virá dada por:

$$1 - P\{X \leq 4\} = 1 - \sum_{x=0}^{4} \frac{e^{-2} 2^4}{4!} = 1 - 0{,}945 = 0{,}055 = 5{,}5\%$$



x	P(x)
0	0,135
1	0,270
2	0,270
3	0,180
4	0,090
5	0,036
6	0,012

Exemplo 2:

Um telefone recebe, em média, cinco chamadas por minuto. Supondo que a distribuição de Poisson seja adequada nessa situação, obter a probabilidade de que o telefone não receba chamadas durante um intervalo de um minuto.

$$P(N = 0) = \frac{5^0 e^{-5}}{0!} = e^{-5} = 0,0067$$

Função Excel

=POISSON.DIST(x, λ, FALSE / TRUE)

A local convenience store uses a Poisson distribution to approximate the number of customers arriving each hour during weeknights at the only checkout counter at the store. From past data it knows that on average 30.5 customers arrive per hour at the checkout counter during weeknights.

Q1: What is the probability that 25 customers arrive at this checkout counter in a given hour during a weeknight?

P(Customers = 25) = POISSON.DIST(25, 30.5, FALSE)
= 0.0469

Função Excel

=POISSON.DIST(x, λ, FALSE / TRUE)

Q2: What is the probability that fewer than 33 customers arrive at this checkout counter in a given hour during a weeknight?

$$\begin{aligned} P(\text{Customers} < 33) &= P(\text{Customers} \le 32) \\ &= \text{POISSON.DIST}(32, 30.5, \text{TRUE}) \\ &= 0.6511 \end{aligned}$$

Obrigado!

(luisflcosta@sapo.pt)