Estatística Aplicada à Recursos Hídricos

Rachid Muleia

Universidade Eduardo Mondlane
Faculdade de Ciências
Departamento de Geologia

rachid.muleia@uem.mz
www.rachidmuleia.com

Julho 2023

Conteúdo

Bibliografia recomendada

- Reis, E. (2008). *Estatística Descritiva*. 7ªEdição.Edições Silabo

Introdução

Definição

Pode-se definir Estatística como sendo o ramo da matemática aplicada que fornece métodos para a recolha, organização, descrição, análise e interpretação de dados para melhor se tirar decisões sobre um dado fenómeno ou acontecimento.

- A estatística é de uma larga importância uma vez que é necessária para a compreensão de fenómenos que ocorrem em diversas áreas do conhecimento (Ex: Biologia, Agricultura, Ciências Sociais, Engenharia, etc.
- No estudo da Estatística, o termo *"estatística"* é utilizado para referir dois conceitos diferentes, conforme se utiliza no singular ou no plural.

Introdução

Informações históricas

- A origem da palavra Estatística está associada à palavra latina STATUS (Estado).
- Há indícios de que 3000 anos A.C. já se faziam censos na Babilónia, China e Egito e até mesmo o 4ºlivro do Velho Testamento.
    - *"Fazei o recenseamento de toda a comunidade dos filhos de Israel, de acordo com suas casas patriarcais: todos aqueles que têm de vinte anos para cima, aptos para o servir no exército de Israel"*:Números 26:2
- Na época do Imperador César Augusto, saiu um edito para que se fizesse o censo em todo o Império Romano. Por isso Maria e José teriam viajado para Belém.

Introdução

Conceitos básicos

A Estatística está dividida em duas partes, nomeadamente: *Estatística Descritiva* e *Estatística Inferencial.*

1. **Estatística Descritiva**: Consiste na recolha de dados númericos através da criação de instrumentos adequados: tabelas, gráficos e indicadores númericos.
   - Univariada: resumir ou descrever o conjunto de dados referentes a uma única variável.
   - Bivariada ou multivariada: relação entre duas ou mais variáveis
2. **Estatística Inferencial**: consiste de procedimentos para fazer generalizações sobre as características de uma população a partir da informação contida na amostra.

Introdução

Conceitos básicos

Os métodos de Inferência Estatística permitem:

1. estimar as características desconhecidas de uma população (por exemplo, a proporção de consumidores que preferem uma dada marca de detergentes)
2. testar se determinadas hipóteses sobre essas características desconhecidas são plausíveis (Ex, se a afirmação de um vendedor de que os resultados de imagem da marca que vende são superiores aos de outras marcas concorrentes).

Introdução

População, Amostra e Unidade Estatística

1. **População:** conjunto de unidades individuais, que podem ser pessoas, animais ou objectos com uma ou mais características em cumum na qual o pesquisador tem interesse em analisar.
2. **Amostra:** o subconjunto retirado da população, que se supõe ser representativo de todas as características da mesma, sobre o qual será feito o estudo, com o objectivo de serem tiradas conclusões válidas sobre a população.
3. **Unidade estatística:** cada elemento da população sobre a qual incide o experimento/estudo.

Introdução

População, Amostra e Unidade Estatística

Na estatística, na maioria das vezes usa-se a amostra e não a população para se estudar um determinada fenómeno. *Mas porquê estudar a amostra e não a população no seu todo?*

- custo alto para obter informação da população toda
- tempo muito longo para obter informação da população toda
- algumas vezes impossível, por exemplo, estudo de poluição atmosférica

Introdução

Exemplos que nos levam a estudar a amostra

- Quando se cozinha um certo molho não é necessário provar todo molho para saber se o molho está bom ou não.
- Assim como, quando se vai ao hospital para se fazer um exame de sangue, o técnico de laboratório não precisa tirar todo sangue para saber se o paciente esta enfermo ou não.

Introdução

Conceitos básicos:Variável, Dados, Parâmetro e Estimativa

1. **Variável**: toda característica que é observada em uma unidade experimental.
2. **Dados**: Informações obtidas em uma unidade experimental (numérica ou não).
3. **Parâmetro**: Valor que resume informação relativa a uma variável na população.
4. **Estimativa**: é uma estatística que é utilizada para inferir um valor de um parâmetro desconhecido em um modelo estatístico.

Introdução

Exemplo 1: população, amostra, unidade estatística

- 144 distritos

- 7 Variaveis:

1- superficie; 2- populacao; 3- temp. media mensal; 4- temp. maxima absoluta; 5- temp. minima absoluta; 6-humidade relative e 7- precipitacao media mensal

Governo de Mocamb. pretende desenvolver uma pesquisa a nivel nacional com objectivo de estimar a precipitaca media mensal do pais. Com base neste objectivo uma amostra de 10 distritos foi seleccionada aleatoriamente.

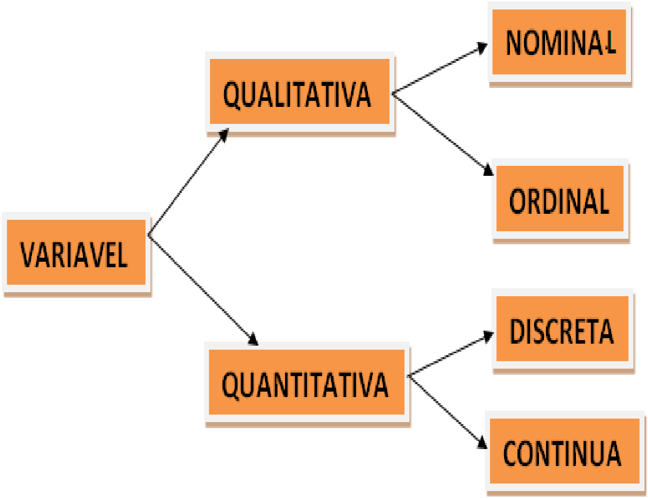
a) Qual seria a populacao?

b) Qual seria a amostra?

c) O que é unidade estatistica?

Introdução

Conceitos básicos: variáveis e classificação



Introdução

Conceitos básicos: variáveis e classificação

1. **Variáveis qualitativas** – quando os valores de uma determinada característica em estudo são expressos por atributos, qualidades ou ainda por categorias.
   - As *variáveis qualitativas nominais*- são caracterizadas por dados que se apresentam apenas sob aspecto qualitativo.
   - As *variáveis qualitativas ordinais* - são caracterizadas por categorias que apresentam uma ordenação natural.
2. **Variável quantitativa** – quando seus valores são expressos em números, os quais podem ser obtidos através de uma contagem ou mensuração
   - *Variáveis quantitativas Discretas*-quando só pode assumir valores pertencentes a um conjunto enumerável.
   - *Variáveis quantitativas Continuas* -quando pode assumir, teoricamente um valor qualquer dentro de um intervalo.

Introdução

Exemplo 2: variáveis e classificação

- **Variáveis qualitativas nominais**: sexo, raça e resultado de um teste
- **Variáveis qualitativas ordinais**: Nível escolaridade e conceito de qualidade.
- **Variáveis quantitativas Discretas**: idade dos alunos de uma escola, números de filhos de um determinado casal, numero de parafusos defeituosos numa linha de produção.
- **Variáveis quantitativas Continuas**: altura, peso, comprimento, salario de operários de uma determinada empresa

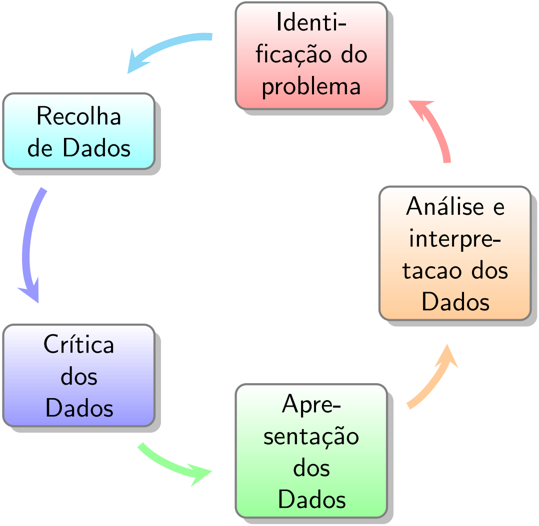
Introdução

Conceitos básicos: Observação

Observação

O facto de uma variável ser expressa por números não significa que ela seja necessariamente quantitativa, por que a classificação da variável depende do modo como foi medida, e não do modo como se manifesta. Por exemplo, para a variável peso de um lutador de boxe, se for anotado o peso marcado na balança, a variável é quantitativa continua; por outro lado, se esse peso for classificado segundo as categorias do boxe, a variável é quantitativa ordinal

As etapas do método estatístico



Etapas do método estatístico

Identificação do problema

- A aplicação do método estatístico depende da identificação clara do problema, isto é, do que se pretende resolver.
- É preciso que se tenha em mente o métdo que será usado para a resolução do problema
- Visto que a escolha do método depende do tipo de informação que será recolhida.
- A identificação clara do problema também possibilita saber que tipo de informação se deve recolher.
- Uma identificação incorreta do problema torna todas etapas seguintes inúteis.

Etapas do método estatístico

Recolha de Dados

Após a identificação do problema a fase a seguir consiste na recolha de dados. O processo de recolha de dados deve garantir o seguinte:

- Não perca tempo recolhendo algo que não vai usar, isto é recolha apenas informação necessária.
- Recolha informação que lhe ajudará a resolver o seu problema.
- Ademais, recolha dados tão completos quanto possíveis

No processo de recolha de dados, temos duas dois tipos de dados:

1. **Dados primários**: são aqueles que são especificamente recolhidos para o estudo em questão.
2. **Dados secundários**: aqueles que já foram recolhidos, tabulados, ordenados e, muitas vezes, até analisados, com propósitos outros ao de atender as necessidades da pesquisa em andamento, e que estão catalogados a disposição dos interessados.

Etapas do método estatístico

Recolha de Dados

- Todos os dados de inquéritos feitos directamente a uma população são dados primários
- Podemos ainda considerar como dados primários todos os dados disponíveis nas estatísticas publicadas pelo INE(Instituto Nacional de Estatística)
    - *Numero de nascimentos, casamentos e óbitos de dada região no ano de 1991.*
    - *O numero de desempregados em um determinado sector de actividade económica.*
- Dados Secundários será por exemplo uma estimativa da esperança de vida à nascença da população moçambicana no ano 2016 com base nos valores observados nos últimos 10 anos ou a taxa de inflação para o ano 2018.

Etapas do método estatistico

Crítica dos dados

Um dos primeiros e mais importantes passos em qualquer tarefa de processamento de dados é verificar se os seus valores de dados estão correctos ou, no mínimo, estão de acordo com alguns um conjunto de regras. Por exemplo:

- A variável género espera-se que tome dois valores apenas;
- A altura em centímetros espera-se que esteja dentro de limites razoáveis.

Não basta so recolher os dados, quer sejam dados primários ou secundários é preciso fazer uma limpeza antes da análise. Esta etapa consiste em:

- Identificar dados incompletos ou incorrectos;
- Eliminar erros capazes de produzir decisões erróneas.

Etapas do método estatístico

Apresentação dos dados

Apos a recolha e crítica dos dados, convém organizar os dados de maneira prática e racional, para um melhor entendimento do fenómeno que se pretende estudar.

Aqui começa o principal objectivo da Estatística Descritiva.

- *Criar instrumentos necessários para classificar e apresentar conjunto de dados númericos de tal forma que a informacao neles contida seja apreendida facilmente.*
- *Existem vários instrumentos para a apresentação de dados: Tabelas, gráficos, medidas de tendência central, etc*

Etapas do método estatśtico

Análsie e interpretação dos resultados

- Após a apresentação dos dados é preciso interpretar os dados
- A interpretacao estará mais tanto facilitada quando se estiverem escolhido, em etapas anteriores, instrumentos mais apropriados e análise do tipo de dados recolhido.
- Nesta etapa por vezes chega-se a conclusões enviesadas, que podem ser de forma propositada ou não.
  - *O enviesamento propositado tem a vista a satisfazer alguns fins, tais como políticos. Daí que podemos ter duas informações diferentes para um mesmo problema reportados pois duas organizações diferentes.*

Apresentação dos dados

Apresentação dos dados: Tabelas e Gráficos

Tabelas

- O sucesso na utilização de dados estatísticos depende em grande parte do modo como os dados são apresentados.
- Quando nos deparamos com grandes massas de dados não classificados torna-se difícil classifica-los
- É preciso proceder um trabalho prévio de ordenação

Apresentação dos dados: Tabelas e Gráficos

Tabelas

Quadro 1: Preços de venda (Euros) de 112 habitações tipologia T2 em Telheiras, 2007

184 800	186 400	172 800	186 000	164 000	210 000	154 000
165 600	144 800	148 000	188 000	170 000	180 000	167 200
157 000	191 600	146 400	170 000	172 000	175 800	148 000
176 800	164 000	213 000	176 000	187 400	182 000	159 200
219 000	212 800	173 600	181 200	168 000	178 000	196 000
164 400	180 000	148 000	164 000	169 000	183 000	165 000
172 000	141 600	140 000	164 000	173 600	140 000	184 000
139 800	159 800	179 600	184 000	171 600	159 000	167 200
175 200	199 200	178 400	171 800	190 000	165 600	140 000
192 400	212 000	183 600	167 600	172 000	209 000	203 200
220 000	224 000	197 000	156 000	152 000	178 000	148 000
222 000	206 000	172 400	165 800	176 000	196 000	209 200
196 400	149 600	148 800	164 000	164 200	170 600	172 000
161 600	182 000	188 000	185 200	198 000	188 000	200 800
152 800	200 000	164 400	186 000	166 000	190 000	172 000
195 000	225 000	172 000	174 400	215 600	158 000	184 000

Apresentação dos dados: Tabelas e Gráficos

Tabelas

Definição

- **Tabela**: é um quadro que resume um conjunto de observações.

As tabelas assim como os como gráficos devem apresentar três partes :o cabeçalho, o corpo e o rodapé.

- o **cabeçalho** deve dar-nos a informação sobre os dados, em que consiste e a que ser referem(lugar e época).
- o **corpo** é apresentado por colunas e sobcolunas dentro das quais se apresentam os dados.
- No **rodapé** para além da identificação da fonte dos dados, poderão ainda incluir-se quaisquer fontes pertinentes.

Apresentação dos dados: Tabelas e Gráficos

Exemplo: Apresentação de dados em Tabelas

Quadro 3: *Indicadores demográficos, NUTS III, 2005*

Indicadores	Taxa bruta de natalidade	Taxa bruta de mortalidade	Taxa bruta de nupcialidade	Taxa bruta de divórcio	Taxa de fecundidade geral	Taxa de fecundidade na adolescência	Índice sintético de fecundidade
	‰						Nº
Norte	10,0	8,7	5,0	1,9	38,4	16,5	1,3
Centro	9,1	11,6	4,4	2,0	38,3	15,0	1,3
Lisboa	11,7	9,5	4,3	2,7	48,0	22,2	1,6
Alentejo	9,0	14,4	4,0	2,0	40,4	21,6	1,4
Algarve	12,0	11,7	4,0	2,1	50,9	24,8	1,7
Portugal*	10,4	10,2	4,6	2,1	41,8	19,0	1,4

*Dados referentes apenas ao Continente.

Fonte: INE (2007), *Estatísticas Demográficas* – 2005.

Apresentação dos dados: Tabelas e Gráficos

Gráficos

Uma imagem vale mais que mil palavras. Confúcio

A representação gráfica tem por finalidade dar uma ideia, a mais representação gráfica possível dos resultados obtidos permitindo chegar a conclusões rápidas sobre a evolução do estudo ou sobre relação entre diferentes valores apresentados.

Apresentação dos dados: Tabelas e Gráficos

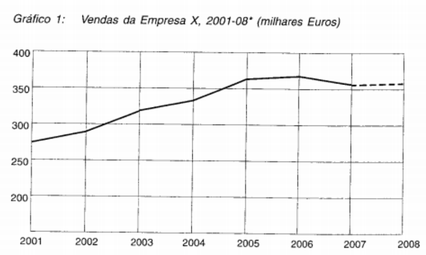
Gráfico de Linhas

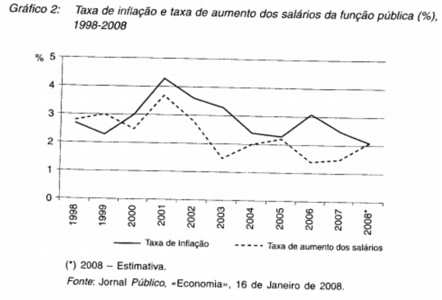
Gráfico de linha

Um gráfico de linhas exibe uma série como um conjunto de pontos conectado por uma única linha. As linhas de gráfico são usadas para representar grandes quantidades de dados que ocorrem em um período de tempo contínuo.

Apresentação dos dados: Tabelas e Gráficos

Gráfico de Linhas





- O gráfico de linhas geralmente é usado para representar variás quantitativas que se observam ao longo do tempo.
- Gráficos de linhas são ideais para exibir tendências ao longo do tempo.

Apresentação dos dados: Tabelas e Gráficos

Gráfico de barras

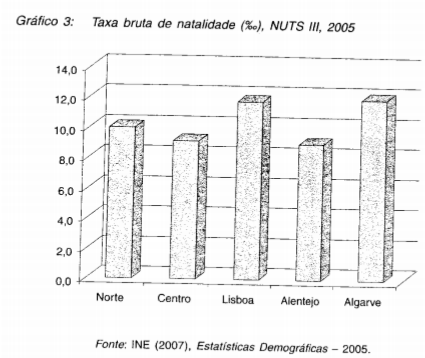
Um gráfico de barras é uma forma de resumir um conjunto de dados categóricos (variável qualitativa).

- Mostra os dados utilizando um número de barras com mesma largura.
- Cada uma das barras representa uma categoria particular
- A altura de cada barra é proporcional a uma agregação específica.

O gráfico de barras constrói-se colocando os valores da variável em observação no eixo horizontal e as respectivas frequências (absolutas ou relativas) no eixo vertical.

Apresentação dos dados: Tabelas e Gráficos

Gráfico de barras

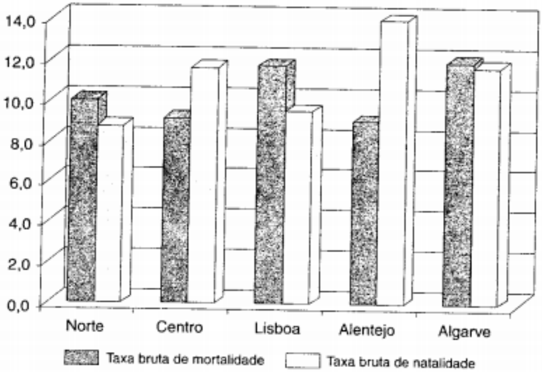


Os gráficos de barras também nos permitem fazer comparações múltiplas de duas ou mais variáveis

Apresentação dos dados: Tabelas e Gráficos

Gráfico de barras

Gráfico 4: *Taxas brutas de natalidade e mortalidade (‰), NUTS III*



Fonte: INE (2007), *Estatísticas Demográficas* – 2005.

Apresentação dos dados: Tabelas e Gráficos

Exercício: Gráfico de barras

Construa um gráfico de barras para o seguinte conjunto de dados:

Número de zonas verdes em cada bairro	*Percentagem de bairros*
1	2
2	4
3	20
4	50
5	18
6	5
≥ 7	1

Apresentação dos dados: Tabelas e Gráficos

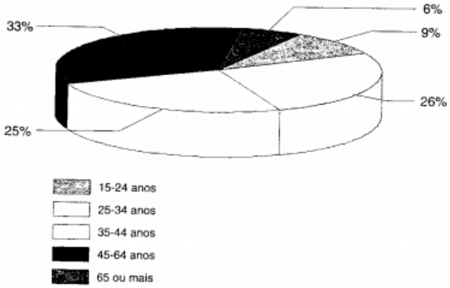
Gráfico de sectores ou diagrama circular

- O gráfico de sectores consiste na representação gráfica de um circulo, divido em sectores.
- É utilizando quando se pretende comparar cada parte com o total.
- Para se construir divide-se o circulo em sectores, cujas áreas são proporcionais aos valores das partes.

Apresentação dos dados: Tabelas e Gráficos

Gráfico de sectores ou diagrama circular

Gráfico 6: *Distribuição da população activa por grupo etário, 3º trimestre de 2007*



Fonte: INE, *Estatísticas do Emprego*, 3º trimestre de 2007.

Apresentação dos dados: Tabelas e Gráficos

Exercício: Gráfico de sectores

Uma empresa tem seis estabelecimentos comerciais. As rendas (em Euros) de cada estabelecimento durante o ano de 2007 foram 6 000, 20 000, 11 000, 18 000, 23 000, 22 000.

— Desenhe um gráfico de sectores para ilustrar estes dados.

— Que conclusões retira do gráfico? Será o mais adequado para este tipo de dados?

— Poderia utilizar um quadro de distribuição de frequências?

Distribuição de frequências

X_i	Frequências absolutas F_i
X_1	n_1
X_2	n_2
X_3	n_3
.	.
X_p	n_p

Distribuição de frequências

Quadro 5: *Distribuição do número de assoalhadas em 110 habitações vendidas em Telheiras durante o ano de 2006*

Nº de assoalhadas por habitação		Nº de habitações vendidas
1		0
2	~~////~~ ~~////~~ /	11
3	~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ //	52
4	~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ ///	28
5	~~////~~ ~~////~~ ~~////~~ ////	19
	Total	110

- Frequência absoluta simples: numero de vezes que cada modalidade/ou observação da variável se repete na amostra ou na população.
- Frequência relativa(f_{r_i}) de um valor da variável é dada por $\frac{f_i}{n}$, isto é o número de vezes que esse valor ocorre relativamente ao total.
- As frequências acumuladas são a soma do número(frequências absolutas acumuladas) ou proporções (frequências acumuladas relativas) de ocorrências para valores inferiores ou iguais ao valor dado.

Distribuição de frequências

Exemplo : distribuição de frequências para variável quantitativa discreta

- A distribuição de frequências do número de erros cometidos por uma datilógrafa apresenta-se da seguinte maneira.

Tabela 1: Distribuição de frequências do número de erros

X_i	f_i	f_{r_i}	F_i	F_{r_i}
0	10	0.10	10	0.10
1	15	0.15	25	0.25
2	25	0.25	50	0.50
3	40	0.40	90	0.90
4	10	0.10	100	1.00
Total	100	1.00		

Distribuição de frequências

Exemplo : distribuição de frequências para variável quantitativa discreta

- A tabela mostra que 10 paginas não tinham nenhum erro;
- 40 paginas foram encontradas com 3 erros;
- As frequências acumuladas nos mostram que 50% do total das páginas datilografadas tinham no máximo 2 erros.

Distribuição de frequências

Distribuição de frequências para variáveis quantitativas

- As variaveis continuas por poderem tomar um número infinito de não-numerável de valores,

Não existe uma fórmula exacta para o cálculo do número de classes. O bom senso diz-nos que não deverá ser um número muito grande para que não introduza irregularidades que poderão não existir na população, mas também não deverá ser muito pequeno para que não haja perda de informação. A selecção do número de classes ou intervalos não constitui nenhum método rigoroso e científico, nem existe nenhum método de selecção que possa ser considerado o mais correcto. Uma vez que o objectivo principal de uma distribuição de frequências é encontrar classes com algum significado e utilidade, é de evitar um número muito pequeno ou muito elevado de intervalos.

Figura 1: Pedaço extraído de Reis(2008)

Distribuição de frequências

Distribuição de frequências para variáveis quantitativas

Embora não exista uma formula cientifica exacta para o calculo de numero de classes, o processo de construção de classes deve obedecer algumas regras:

- Em geral o número de classes deverá estar compreendido entre 4 e 14.
- Nenhuma classe deverá ter frequência núla.
- As classes deverão ter, sempre que possível, amplitudes iguais;
- Os pontos médios da classe deverão ser de cálculo fácil.
- As classes abertas deverão sempre ser evitadas, embora nem sempre é possível.
- Os limites da classe são definidos de modo a que cada valor é incluido numa só classe.

Tabelas de Distribuição e Frequências

1. Elementos de uma tabela de distribuição de frequências:
   - frequências simples/absolutas (f_i)
   - Frequência absoluta acumulada (F_i)
   - frequência relativa absoluta $fr_i = \dfrac{f_i}{n}$
   - Frequência relativa acumulada $Fr_i = \dfrac{F_i}{n}$
2. Procedimentos para construção de uma tabela de distribuição de frequências
   - Determinar Amplitude Total (A_t): $At = X_{\text{max}} - X_{\text{min}}$
   - Determinar o número de classes (k): $k = \lceil \sqrt{n} \rceil$ ou Sturges: $k = \lceil 1 + 3.322 \log(n) \rceil$, se $n \leqslant 25$ $k = 5$
   - Determinar Amplitude do Intervalo de Classe (h): $h = \left\lceil \dfrac{At}{k} \right\rceil$
   - Definição das classes e suas respectivas frequências (veja o exemplo)

Exemplo 8: Tabelas de Distribuição de Frequências

Tabela 2: Dados brutos: Precipitação média mensal de 2012, Moç.

9.2	41	56.3	63.4	67	78.8	116.1	122.9	243.4
23.9	41	56.3	63.4	67	78.8	116.1	122.9	243.4
23.9	44.6	56.3	63.4	67	78.8	116.1	122.9	243.4
23.9	44.6	56.3	65.3	67	80.9	116.1	130.5	243.4
23.9	45.3	56.3	65.3	67	84.1	116.1	135.1	243.4
23.9	45.4	56.3	65.3	67	84.2	116.1	135.1	243.4
23.9	47.3	56.3	65.3	67	84.2	116.1	135.1	243.4
24.9	48.6	56.3	65.3	72.1	94.4	116.4	135.1	265
24.9	49.7	56.6	65.3	72.1	94.4	116.4	135.1	265
24.9	52.7	56.6	65.3	77	94.4	116.4	135.1	
25.2	52.7	60.4	65.3	77	94.4	116.4	135.1	
35.3	52.7	60.4	65.3	77	94.4	116.4	135.1	
35.3	52.7	60.4	67	77	94.4	116.4	135.1	
40.2	52.7	60.4	67	77	98.9	116.4	135.1	
40.2	55.8	60.4	67	78.4	107.8	118.4	135.1	

Exemplo 8: Tabelas de Distribuição de Frequências

1. $A_t = X_{\mathsf{max}} - X_{\mathsf{min}} = 265 - 9.2 = 255.8$
2. $k = \lceil \sqrt{137} \rceil = \lceil 11.70 \rceil = 12$
3. $h = \left\lceil \dfrac{A_T}{k} \right\rceil = \left\lceil \dfrac{255.8}{12} \right\rceil = 22$

Exemplo 8: Tabelas de Distribuição de Frequências

i	Classes	f_i	F_i	fr_i	Fr_i
1	9.2 ⊢ (9.2+ 22)=31.2	11	11	0.080292	0.080292
2	31.2 ⊢ (31.2+ 22)=53.2	21	32	0.153285	0.233577
3	53.2 ⊢ 75.2	41	73	0.29927	0.532847
4	75.2 ⊢ 97.2	20	93	0.145985	0.678832
5	97.2 ⊢ 119.2	18	111	0.131387	0.810219
6	119.2 ⊢ 141.2	16	127	0.116788	0.927007
7	141.2 ⊢ 163.2	0	127	0	0.927007
8	163.2 ⊢ 185.2	0	127	0	0.927007
9	185.2 ⊢ 207.2	0	127	0	0.927007
10	207.2 ⊢ 229.2	0	127	0	0.927007
11	229.2 ⊢ 251.2	8	135	0.058394	0.985401
12	251.2 ⊢ 273.2	2	137	0.014599	1
	total	137			

Representação gráfica das distribuições de frequências

Variável quantitativa discreta

- Para presentar a distribuição de frequências relativas para uma variável discreta utiliza-se o **gráfico de barras ou o diagrama diferencial.**

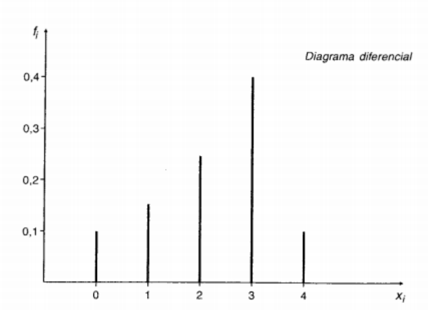


Figura 2: Distribuicao de frequencias para numero de erros

Representação gráfica das distribuições de frequências

Variável quantitativa contínua

- A representação gráfica adequada para este tipo de variáveis é o *histograma*.
- O *histograma* é um gráfico, formado por uma sucessão de rectângulos adjacentes, tendo cada um por base um intervalo de classe e uma área igual (ou proporcional) à frequência relativa (ou absoluta) dessa classe.
- Ao contrário do *gráfico de barras*,em que estas estão separadas e em que a altura de cada barra é o mais relevante, no histograma as barras (rectângulos) estão juntas e o que é importante é a área de cada uma.
- Considerando, então, para áreas das barras as frequências relativas, vemos que a área total ocupada pelo histograma é igual a 1 ou 100%

Representação gráfica das distribuições de frequências

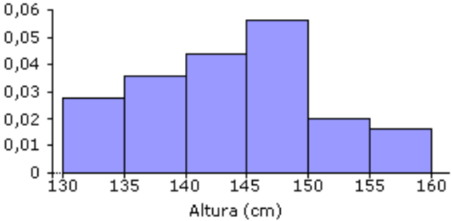
Variável quantitativa contínua: histograma

Tabela 3: Distribuição de frequências

Classes	f_i	f_r	$h = f_r/c$
$[130, 135[$	7	0.14	0.028
$[135, 140[$	9	0.18	0.036
$[140, 145[$	11	0.22	0.044
$[145, 150[$	14	0.28	0.056
$[150, 155[$	5	0.1	0.020
$[155, 160[$	4	0.08	0.016
Total	50	1	

Representação gráfica das distribuições de frequências

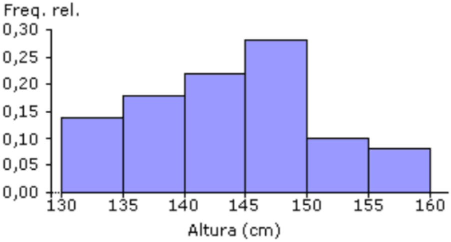
Variável quantitativa contínua: histograma



No histograma ao lado, a área do rectângulo mais à esquerda é igual a $5 \times 0{,}028 = 0{,}14$; a área do rectângulo seguinte é $5 \times 0{,}036 = 0{,}18$ e assim sucessivamente, donde a área total do histograma é igual a 1 (soma das frequências relativas).

Representação gráfica das distribuições de frequências

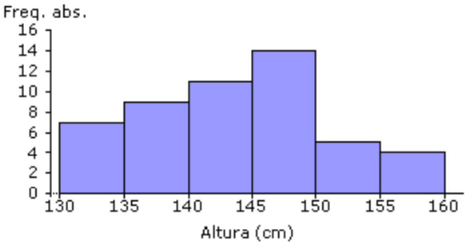
Variável quantitativa contínua: histograma



No histograma ao lado, a área do rectângulo mais à esquerda é igual a $5 \times 0{,}14$; a área do rectângulo seguinte é $5 \times 0{,}18$ e assim sucessivamente, donde a área total do histograma é igual a 5 ($=5\times$ 1 onde 1 é a soma das frequências relativas).

Representação gráfica das distribuições de frequências

Variável quantitativa contínua: histograma



No histograma ao lado, a área do rectângulo mais à esquerda é igual a 5×7; a área do rectângulo seguinte é 5×9 e assim sucessivamente, donde a área total do histograma é igual a 250 ($=5\times50$, onde 50 é a soma das frequências absolutas).

Representação gráfica das distribuições de frequências

Variável quantitativa contínua: histograma

- A imagem transmitida tem sempre o mesmo aspecto.
- As áreas dos rectângulos ou são iguais às frequências relativas, ou são proporcionais, com a mesma constante de proporcionalidade,que é igual à amplitude de classe no caso do segundo histograma ou à amplitude de classe vezes o número de dados, no caso do terceiro histograma
- O eixo vertical só serve como auxílio para a construção dos rectângulos, não transmitindo, no caso do histograma, qualquer informação relevante:

Representação gráfica das distribuições de frequências

Variável quantitativa contínua: histograma

Observação

- Não devemos perder de vista que o histograma representa os dados através das áreas das barras e não das alturas, o que constitui uma grande diferença relativamente ao gráfico de barras.
- Outra grande diferença é que no histograma as barras estão juntas, para transmitir a ideia de continuidade da variável em estudo, enquanto no gráfico de barras estas são separadas.

Representação gráfica das distribuições de frequências

Variável quantitativa contínua: histograma

- Caso as amplitudes das classes seja diferente é preciso normalizar a altura do histograma de mod que esta seja proporcional a area do mesmo.

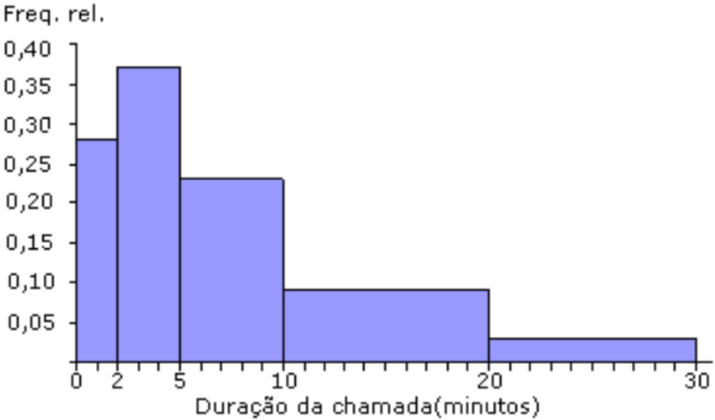
Tabela 4: Distribuição de frequências da duração das chamadas telefónicas

Classes	f_I	f_r
$[0, 2[$	28	0.28
$[2, 5[$	37	0.37
$[5, 10[$	23	0.23
$[10, 20[$	9	0.09
$[20, 30[$	3	0.03
	100	

Representação gráfica das distribuições de frequências

Variável quantitativa contínua: histograma

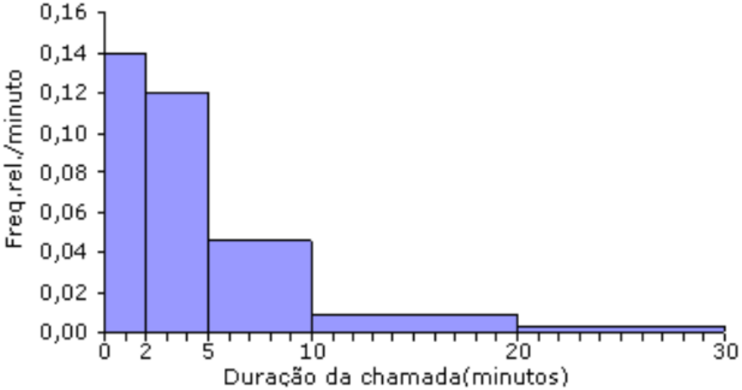
- Se usarmos as frequências relativas directamente, teremos uma representação errada do histograma



Representação gráfica das distribuições de frequências

Variável quantitativa contínua: histograma

- Considerando as alturas correctas, temos o seguinte histograma:



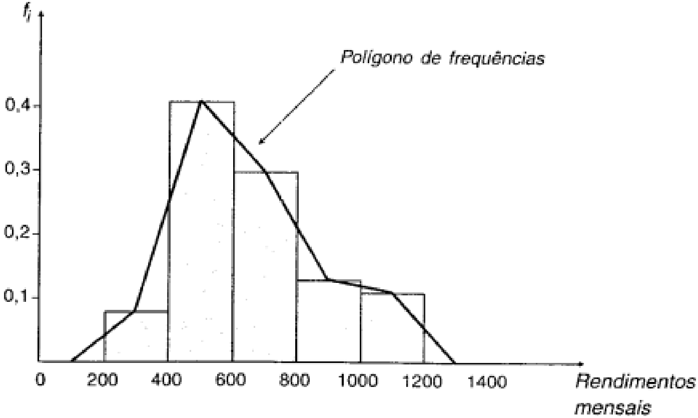
Representação gráfica das distribuições de frequências

Variável quantitativa contínua: histograma

- Uma forma alternativa de apresentar dados quantitativos contínuos é usando o *polígono de frequências*, que se obtém unindo os pontos médios dos topos dos rectângulos com segmentos de rectas.
- A área sobre o histograma deve ser igual a área sobre o polígono de frequências.

Representação gráfica das distribuições de frequências

Variável quantitativa contínua: histograma



Medidas de tendência central: Média

O objectivo das medidas de tendência central é descrever ou sumariar o conjunto de dados através de um valor apenas.

Média

É a medida de tendência central mais usada para descrever resumidamente uma distribuição de frequências. Todavia, há vários tipos de médias que são usados de acordo com cada objectivo.

1. Média Aritimética
2. Média Geométrica
3. Média Harmónica
4. Média Quadrática

O valor a escolher depende da característica dos dados

Média Aritmética

- **A média aritmética simples:**

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

- **A média aritmética ponderada:**
  - Dados não agrupados

$$\bar{x} = \frac{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}{n} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \tag{2}$$

Média aritimética

Exemplo 3: Dados não agrupados

Tabela 5: Preço das T-shirts em Euros

Modelo	Preço
Simples	4.9
Madona	9.9
Julio Iglesias	7.6
Springsteen	7.6
David Bowie	8.4
Maradona	7.8

$$\bar{X} = \frac{4.9 + 9.9 + 7.6 + 7.6 + 8.4 + 7.8}{6} = 7.70$$

Este valor é referente ao preço médio das T-shirts, mas se o quantidade de T-shirts vendida o valor do preç médio será diferente.

Média aritimética

Exemplo 4: Dados agrupados

Tabela 6: Média dos preços das T-shirts vendidas

Modelo	Preço do modelo X_i	Quantidade vendida f_i	Preço × Quantidade $X_i \cdot f_i$
Simples	4.90	15	73.50
Madona	9.90	5	49.50
Julio Iglesias	7.60	10	76.00
Springsteen	7.60	8	60.80
David Bowie	8.40	6	50.40
Maradona	7.80	6	46.80
Total		$\sum$ =50	$\sum$ = 357.00

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i \cdot f_i}{\sum_{i=1}^{n} f_i} = \frac{357.00}{50} = 7.14$$

Média Aritmética

Para dados agrupados em classe a média aritimética pode ser calculada com base na seguinte fórmula:

$$\bar{x} = x_0 + c\frac{\sum_{i=1}^{n} d' f_i}{n} \tag{3}$$

Onde,

- $d' = \dfrac{x_0 - x_i}{c}$
- x_0 será igual ao ponto médio da classe de maior frequência, se o número de classes for par, ou o ponto médio da classe intermédiaria, se o número de classes for ímpar.
- c será igual à amplitude do intervalo de classe.

Exemplo

Classes	f_i	a) F_i	Marca de classe x_i	$d'_i = \frac{x_i - 75}{20}$	$d'_i f_i$
[5 – 25[	4	4	15	$\frac{15-75}{20} = -3$	$-3*4=-12$
[25 – 45[	6	10	35	$\frac{35-75}{20} = --2$	$-2*6=-12$
[45 – 65[	14	24	55	$\frac{55-75}{20} = -1$	$-1*14=-14$
[65 – 85[	***26***	**50**	***75***	$\frac{75-75}{20} = 0$	$0*26=0$
[85 – 105[	14	64	95	$\frac{95-75}{20} = 1$	$1*14=14$
[105 – 125[	8	72	115	$\frac{115-75}{20} = 2$	$2*8=16$
[125 – 145[	6	78	135	$\frac{135-75}{20} = 3$	$3*6=18$
[145 – 165[	2	80	155	$\frac{155-75}{20} = 4$	$4*2=8$
	$\sum_{i=1}^{8} f_i = 80$				$\sum_{i=1}^{8} d'_i f_i = 18$

Aplicando a fórmula, a média será igual a: $\bar{x} = x_0 + c\frac{\sum_{i=1}^{n} d' f_i}{n} = 79.5$

Média aritimétia

Exemplo 6:método alternativo para dados agrupados em classe

Tabela 7: Calculo da média usando o ponto médio

Classes	f_i	F_i	Ponto médio X_i
$[5–25[$	4	4	15$=\dfrac{5+25}{2}$
$[25–45[$	6	10	35
$[45–65[$	14	24	55
$[65–85[$	26	50	75
$[85–105[$	14	64	95
$[105–125[$	8	72	115
$[125–145[$	6	78	135
$[145–165[$	2	80	155

$$\bar{X} = \frac{\sum_{i=1}^{8} X_i \cdot f_i}{\sum_{i=1}^{8} f_i} = \frac{6360}{80} = 79.5$$

Algumas propriedades da média aritimética

- A soma algébrica dos desvios tomados em relação a média é nula: $\sum_{i=1}^{n}(x_i - \bar{x})$
- Somando-se (ou subtraindo-se) uma constante (c) de todos os valores de uma variável, a média do conjunto fica aumentada (ou diminuída) dessa constante: $y_i = x_i + c \Rightarrow \bar{Y} = \bar{X} + c$
- Multiplicando-se (ou dividindo-se) todos os valores de uma variável por uma constante (c), a média do conjunto fica multiplicada (ou dividida) por essa constante: $y_i = x_i \times c \Rightarrow \bar{Y} = \bar{X} \times c$

Média Geométrica

Média geométrica

A média geométrica de n valores é definida, genericamente, como a raiz n-ésima do produto de todos eles.

1. Média geométrica simples

$$\sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n} \tag{4}$$

2. Média geométrica Ponderada

$$\sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot x_3^{f_3} \cdots x_n^{f_n}} \tag{5}$$

A média geométrica se aplica apenas a números positivos a fim de evitar o cálculo do produto de um número negativo que poderia resultar em números imaginários.

Aplicações da média geométrica

Aplicação

A média geométrica é mais apropriada que a média aritmética para descrever crescimentos proporcionais, tanto crescimento exponencial (proporção constante de crescimento) e crescimento variado.

Exemplo

- Vamos considerar um aumento sucessivo de salários de 15% no primeiro mês, 12% no segundo mês e 21% no terceiro mês.
- Vamos determinar a média geométrica dos dos aumentos.
- Primeiro devemos determinar as taxas percentuais.

$$\begin{bmatrix} 15\% \\ 12\% \\ 21\% \end{bmatrix} = \begin{bmatrix} 1.15 \\ 1.12 \\ 1.21 \end{bmatrix} \tag{6}$$

$$\sqrt[3]{1.15 \cdot 1.12 \cdot 1.21} = 1.1594$$

O valor 1.1594 corresponde a taxa média de 15.94%

Média harmónica

Média hármonica

- Em situaçoẽs onde a proporcionalidade inversa esteja presente é aconselhavel o cálculo da média harmónica. Por exemplo, quando estudamos fenómenos como a velocidade média, o custo médio de bens comprados com uma quantia fixa.
- A média harmónica é o inverso da média aritimética dos inversos dos valores observados.

O cálculo da média harmónica é com base na seguinte fórmula (Dados simples):

$$\bar{x}_M = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} \tag{7}$$

Média harmónica

A média harmónica para dados agrupados é calculada com base na seguinte fórmula:

$$\bar{x}_M = \frac{n}{f_1\frac{1}{x_1} + f_2\frac{1}{x_2} + \cdots + f_n\frac{1}{x_n}} = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=i}^{n} f_i\frac{1}{x_i}} \tag{8}$$

onde n e f_i representam as o numero total de observações e frequências a absolutas simples, respectivamente.

Exemplo: Média harmónica

- Um avião voa três distâncias iguais a uma velocidade de 300, 400, e 300 milhas/hora.
- Calcule a velocidade média com que o avião percorreu a distância total.
- Neste caso acoselha-se a utilização da média harmónica por se tratar de uma distância fixa percorrida em velocidades diferentes

$$\bar{x}_M = \frac{3}{\frac{1}{300} + \frac{1}{400} + \frac{1}{300}} = 327.27\text{milhas/hora}$$

Mediana

Mediana

- É definido como o valor da variável que ocupa a posição central na sucessão das observações ou na distribuição de frequências.
- O número de observações para valores que lhe são inferiores é igual ao número de observações para valores que lhe são superior.
- Antes de achar a mediana é preciso ordenar os dados em ordem crescente ou decrescente).

1. Caso N seja ímpar a mediana sera o elemento central(de ordem $\dfrac{N+1}{2}$
2. Se o N for par, será a média entre os elementos centrais (de ordem $\dfrac{N}{2}$ e $\dfrac{N+2}{2}$)

Exemplo: Mediana

Cálculo da mediana para valores discretos

X_i	F_i	cum F_i	
1	1	1	
2	3	4	
3	5	9	$\leftarrow$ contém o 6º elemento
4	2	11	
	$\sum = 11$		

Exemplo: Mediana

Cálculo da mediana para valores discretos

X_i	F_i	cum F_i
82	5	5
85	10	15
87	15	30
89	8	38
90	4	42
	$\sum = 42$	

- Como $N = 42$ é par, a mediana será a média entre os valores correspondentes aos elementos de ordem $\dfrac{N}{2}$ e $\dfrac{N+2}{2}$, ou seja $21\circ$ e $22\circ$

Mediana

Cálculo da mediana para dados agrupados em classe

Quando os dados estiverem agrupados em classe, o calculo usando a fórmula:

$$M_d = Li(Md) + \frac{\frac{N}{2} - F_{\text{ant}}}{f_{\text{Md}}} c \tag{9}$$

- $Li(Md)$- limite inferior da classe mediana.
- c- amplitude da classe mediana.
- F_{ant}-frequência acumulada até a classe anterior à classe mediana.
- f_{Md}-frequência simples da classe mediana

Exemplo: Mediana

Cálculo da mediana para dados agrupados em classe

Classes	F_i	cum F_i
35 - 45	5	5
45 - 55	12	17
55 - 65	18	35
65 - 75	14	49
75 - 85	6	55
85 - 95	3	58
	$\sum = 58$	

Exemplo: Mediana

Cálculo da mediana para dados agrupados em classe

- Calcula-se $\dfrac{N}{2} = 29$.
- Identificar a classe mediana $[55, 65[$
- Calcula-se a mediana utilizando a formula:

$$55 + \frac{29 - 17}{18} \times 10 = 61.67$$

Mediana

Características importantes da mediana

1. A mediana é fácil de calcular e de compreender.
2. Não é afectada por valores extremos.
3. É uma medida muito utilizada, sobretudo para distribuições fortemente assimétricas.
4. Para fins de inferência a estatística, a mediana não satisfaz propriedades de um bom estimador.

Moda

- A moda é o valor que ocorre com maior frequência em uma série de valores.
- O cálculo da médiana é fácil quando os dados estiverem ordenados.
- Podes entretanto, encontrar uma série de dados sem moda. Este tipo de distribuição de dados chama-se **amodal.**

Moda para dados não agrupados.

Quando se lida com valores não agrupados, a moda é facilmente reconhecida: basta, de acordo com a definição, procurar o valor que mais se repete.

A série de dados: $7, 8, 9, 10, 10, 10, 11, 12, 13, 15$; tem moda igual a 10

Moda

Cáculo da moda para dados agrupados em classe

Existem vários métodos de determinação da moda. Mas para o nosso estudo, iremos determinar a moda usando o método de King.

$$M_o = l + c\frac{f_{\text{post}}}{f_{\text{ant}} + f_{\text{post}}} \tag{10}$$

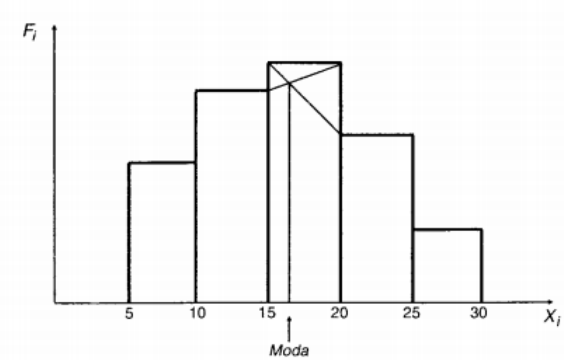
onde,

- l- limite inferior da classe modal.
- c -amplitude do intervalo de classe modal.
- f_{ant}-frequência simples da classe adjacente anterior à classe modal.
- f_{post}-frequência simples da classe posterior à classe modal

Moda

Determinação gráfica da moda

A moda poderá ser determinada graficamente, para tal é necessário construir o histograma da distribuição, identificar a classe modal e fazer a seguinte construção.



Outras medidas de tendência central: Média aparada

- A média aparada é calculado pela "descartar" uma certa percentagem de valores maiores e menores de um conjunto de dados.
- Por exemplo, a média aparada de 10% encontra-se, eliminando a 10% das observações menores e 10% das observações maiores, e depois calcula-se a média dos restantes valores.

Por exemplo, considere o seguinte conjunto de dados.

$$5, 4, 6, 7, 8, 10, 11, 0, 7, 18$$

- a média aparada de 10% será

$$\bar{X} = \frac{4+5+6+7+7+0+10+11}{8} = 7.25$$

- a média aparada de 20% será

$$\bar{X} = \frac{5+6+7+7+0+10}{6} = 7.1667$$

Média aparada

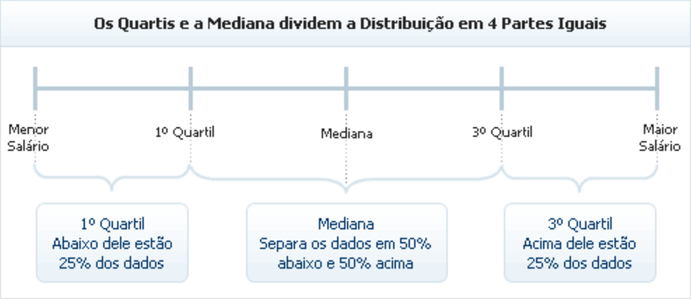
Características

- Diferentemente da média aritmética,a média aparada é menos sensível a valores extremos.
- Embora a média aparada seja menos sensível, ela não se assemelha a mediana que é praticamente insensível a valores extremos.
- A média aparada tem vantagem sobre a média, pois, para além de ser menos insensível a valores extremos, ela usa mais observações do que a mediana.

Medidas de separatiz:Quartis, Decis e Percentis

Quartis

- Os quartis são os valores da variável que dividem a distribuição de frequências em quatro partes iguais.



Medidas de separatiz:Quartis, Decis e Percentis

Quartis

- O primeiro quartil Q_1, será um valor da variável tal que o número de observações para valores inferiores será 25%, e superiores,75%.
- O segundo quartil Q_2 é igual a mediana. Este divide a distribuição em duas partes iguais.
- O terceiro quartil Q_3 será um valor da variável tal que a sua esquerda concentram-se 75% das observações e a sua direita as restantes 25%.

Medidas de separatiz:Quartis, Decis e Percentis

Quartis

- O cálculo dos quartis é similar ao cálculo da mediana.
- Ordene os dados em ordem crescente.
- Calcule o índice i, a posição do p-ésimo quartil sera:

$$i = \left(\frac{p}{4}\right)n$$

- Se i não é inteiro arredonde por excesso. O p-ésimo quartil será o valor na posição i
- Caso i seja inteiro, o p-ésimo quartil será a média aritmética dos valores na posição i e $i+1$.

Medidas de separatiz:Quartis, Decis e Percentis

Exemplo: Quartis

Tamanho dos sapatos X_i	Nº de pares vendidos F_i	cum F_i	f_i (%)	cum f_i (%)
35	30	30	0,65	0,65
35 ½	40	70	0,87	1,52
36	50	120	1,08	2,60
36 ½	150	270	3,26	5,86
37	300	570	6,51	12,37
Q_1 = 37 ½	600	1 170	13,02	25,39
38	950	2 120	20,61	46,00
Q_2 = 38 ½	820	2 940	17,79	63,79
Q_3 = 39	750	3 690	16,27	80,06
39 ½	440	4 130	9,55	89,61
40	250	4 380	5,42	95,03
40 ½	150	4 530	3,26	98,29
41	40	4 570	0,86	99,15
41 ½	39	4 609	0,85	100,00
	$\sum = 4609$		$\sum = 100{,}00$	

Medidas de separatiz:Quartis, Decis e Percentis

Decis e Percentis

- Os decis são valores da variável que dividem a distribuição em dez partes iguais enquanto que os percentis dividem em 100 partes iguais.
- O número de decis e de percentis será 9 e 99, respectivamente.

Medidas de separatiz:Quartis, Decis e Percentis

Dados agrupados em classe

- Caso os dados estejam agrupados em classe, o calculo de Quartis, Decis e Percentis é feito usando as seguintes fórmulas:

$$Q_p = l + c\frac{\frac{np}{4} - F_{ant}}{f_{Q_p}} \tag{11}$$

$$D_p = l + c\frac{\frac{np}{10} - F_{ant}}{f_{D_p}} \tag{12}$$

$$C_p = l + c\frac{\frac{np}{100} - F_{ant}}{f_{C_p}} \tag{13}$$

Medidas de separatiz:Quartis, Decis e Percentis

Exercício

Dias de internamento	*Número de doentes*
Menos de 5	48
5 – 10	33
10 – 20	27
20 – 30	18
30 – 45	15
Mais de 45	9
Total	150

a) Calcule o primeiro quartil e o percentil 25.

Medidas de dispersão absoluta

- *Podemos definir a estatística como uma ciência que lida com variabilidade*.
- Sem variabilidade não faz sentido a existência da estatística
- Por exemplo, se numa turma todos tivessem a mesma idade, não haveria necessidade de calcular a média .
- Se numa empresa todos tivessem o mesmo salário, não se perderia tempo em procurar uma medida que possa condensar a informação na distribuição de dados
- *As medidas de dispersão servem para verificar a representatividade das medidas de localização*
- Ademais, podemos ter dois conjuntos de dados com mesma média, desta forma não sendo suficiente a média para descrever o conjunto de dados.

Medidas de dispersão absoluta

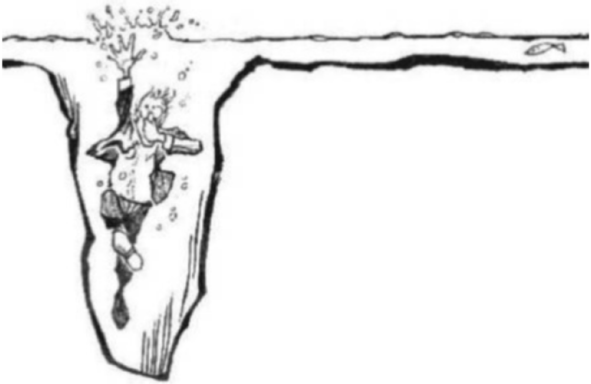


Figura 4: Estaticista se afogando num lado de profundidade média de 0.5 m.

Medidas de dispersão absoluta

x	20	20	20	20	20
y	15	10	20	25	30

- Ambas distribuições tem mesma média $\mu_x = \mu_y = 20$
- Porém, os valores de X não apresentam nenhuma variação em torno da média.

Medidas de dispersão absoluta

Amplitude Total

- A primeira medida de dispersão é a *amplitude total*
- É a diferença entre o valor máximo e mínimo da váriavel.

$$A_t = X_{max} - X_{min}$$

- Se os dados estiverem agrupados em classe, podemos calcular a amplitude da seguinte maneira:
  - $A_t =$Ponto médio da última classe – ponto médio da primeira classe.
  - $A_t =$Limite superior da última classe – limite inferior da primeira classe

Medidas de dispersão absoluta

Amplitude Total

- O uso da amplitude total tem como desvantagem o facto de ter em conta apenas os dois valores extremos que a variável toma.
- À parte isso, a amplitude total h́ insensível a valores intermédios.

Exemplo: Considere a amostra 4, 6, 1, 2, 3, 3

$$A_t = 1 - 6 = 4$$

Medidas de dispersão absoluta

Amplitude Interquartil

- Uma outra medida de dispersão é a *amplitude do intervalo interquartis* ou simplesmente *amplitude interquartil.*
- Esta é definida como a diferença entre o terceiro e o primeiro quartil
- Corresponde ao intervalo que engloba 50% das observações.

$$IQ = Q_3 - Q_1$$

- Tem a desvantagem de não ser influenciado por metade dos valores observados.
- Esta medida deve ser utilizada para distribuições assimétricas uma vez que não é afectada por valores extremos.

Medidas de dispersão absoluta

Exemplo: Amplitude Interquartil

Quadro 21: *Preços (por 100 Kg) das sementes de flores*

Variedade de flor	*Preço por 100 Kg* (Euros)
A	133
B	135
C	146
D	175
E	179
F	188
G	195
H	204
I	219
J	233
L	254

$$IQ = 219 - 146 = 73\text{euros}$$

- 50% das observações centrais apresentam uma variação de preço de 73 euros

Medidas de dispersão absoluta

Variância

- A variância é a mediada de variabilidade que utiliza todos os dados. A variância é baseada na diferença entre o valor de cada observação (x_i) e a média. A diferença entre cada valor de x_i e a media, é chamada de desvio em relação a média.

	Dados nao agrupados	Dados agrupados
População	$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{N}$	$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2 f_i}{N}$
Amostra	$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$	$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 f_i}{n-1}$

Medidas de dispersão absoluta

Variância

- No cáculo da variância podemos dividir a soma dos desvios por $n-1$ ou n, isto é

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} \tag{14}$$

ou

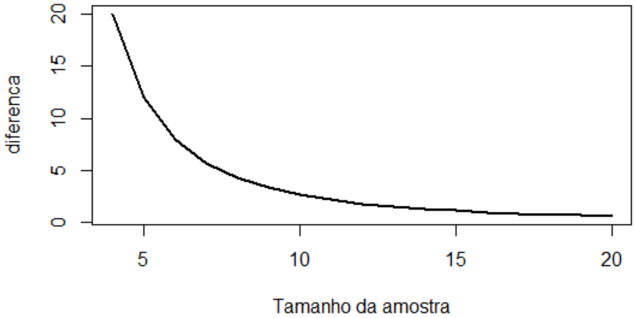
$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n} \tag{15}$$

- A fórmula 15 produz valores viciados para amostras pequenas.
- Para amostras grandes as duas fórmulas produzem valores aproximadamente iguais.

Medidas de dispersão absoluta

Variância

- A partir do gráfico podemos ver que a medida que o tamanho da amostra aumenta as diferenças entre as fórmulas 14 e 15 são minúsculas.



Medidas de dispersão absoluta

Desvio-padrão

- Uma vez que a variância é expressa em unidades ao quadrado o que torna difícil a sua interpretação, recorre-se ao desvio padrão que é definido como sendo a raiz quadrada positiva da variância:

$$\text{desvio-padrão amostral} \quad s = +\sqrt{s^2}$$

$$\text{desvio-padrão amostral} \quad \sigma = +\sqrt{\sigma^2}$$

Medidas de dispersão absoluta

Exemplo: Desvio-padrão

- Considere a seguinte amostra: 1, 3, 2, 2, 4

Temos que

$\sum_{i=1}^{n} x_i = 1 + 3 + 2 + 2 + 4 = 12$, $\sum_{i=1}^{n} x_i^2 = 1^2 + 3^2 + 2^2 + 2^2 + 4^2 = 34$

Daí que

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{34 - 12^2/5}{4} = 1.3$$

$$s = \sqrt{1.3} = 1.1042$$

Medidas de dispersão absoluta

Desvio-padrão

- É a medida de dispersão mais utilizada. As suas propriedades matemáticas tornam-no particularmente apropriado em situações de inferência estatística.
- O desvio-padrão é uma medida de dispersão afectada por todos valores, portanto, qualquer alteração nestes provoca uma alteração no primeiro.
- O seu valor pode ser fortemente influenciado por valores extremos. Por essa razão,a sua utilização é menos aconselhada em distribuições altamente assimétricas.

Medidas de dispersão relativa

Coeficiente de dispersão e variação

- Todas medidas de dispersão apresentadas até agora, são medidas de dispersão absoluta.
- Portanto, não são válidas para comparação da dispersão de duas distribuições, sobre tudo quando estas estão em unidades de medidas diferentes.
- o coeficiente de variação é uma medida relativa de dispersão que mede o grau de concentração de valores em torno da média, em valor percentual.

$$CV = \frac{s}{\bar{x}} \times 100$$

Medidas de dispersão relativa

Coeficiente de dispersão e variação

Eis algumas regras empíricas para a interpretação do coeficiente de variação:

Rule of thumb ou "regra do polegar"

- Se $CV < 15\%$ tem-se baixa dispersão.
- Se $15\% \leqslant CV < 30\%$ tem-se média dispersão.
- Se $CV \geqslant 30\%$tem-se elevada dispersão.

Medidas de dispersão relativa

Exemplo

Considere as seguintes estimativas referentes a duas amostras de 50 pneus de marcas diferentes, onde $\overline{X}_1$ é a média, em Km, de duração da primeira marca e $\overline{X}_2$ é o valor correspondente para a segunda marca e S_1 e S_2 são os respectivos desvios-padrão.

$$\bar{x}_1 = 50000\,\text{Km} \qquad s_1 = 12000\,\text{Km}$$

$$\bar{x}_2 = 30000\,\text{Km} \qquad s_2 = 8000\,\text{Km}$$

Em termos relativos, qual das marcas de pneus apresenta uma maior dispersão relativa da sua duração?

$$Cv_1 = \frac{s_1}{\bar{x}_1} \,.\, 100 = \frac{12000}{50000} \,.\, 100 = 24\%$$

$$Cv_2 = \frac{s_2}{\bar{x}_2} \,.\, 100 = \frac{8000}{30000} \,.\, 100 = 26{,}7\%$$

Quando se comparam os respectivos coeficientes de variação constata-se a maior dispersão relativa da segunda marca.

Medidas de Assimetria

- A forma mais simples de medir o grau de assimetria de uma distribuição consiste na comparação de três medidas de tendência central.
- Numa distribuição *simétrica* temos que $\bar{x} = M_o = M_e$
- Quando a média $\geqslant$ mediana$\geqslant$ moda, temos uma distribuição *assimétrica positiva*.
- No caso inverso- média $\leqslant$ mediana$\leqslant$ moda, temos uma distribuição *assimétrica negativa*

Medidas de Assimetria

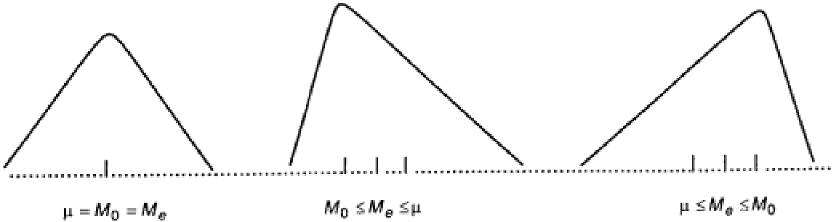


Figura 1: Distribuição simétrica

Figura 2: Assimetria positiva

Figura 3: Assimetria negativa

Medidas de Assimetria

Existem várias fórmulas para o cálculo do coeficiente de assimetria, dentre elas são úteis:

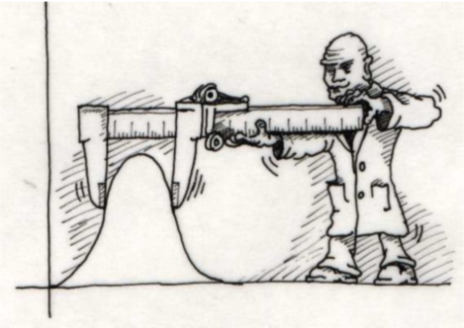
$$1^{\underline{o}}\text{Coeficiente de assimetria de Pearson}\, AS = \frac{\bar{x} - M_o}{s}$$

$$2^{\underline{o}}\text{Coeficiente de assimetria de Pearson}\, AS = \frac{Q_1 + Q_3 - 2M_e}{Q_3 - Q_1}$$

Rule of thumb

- Se $AS = 0$ diz-se que a distribuição é simétrica.
- Se $AS > 0$ diz-se que a distribuição é assimétrica positiva.
- $AS < 0$, diz-se que a distribuição é assimétrica negativa.

Medidas de achatamento ou curtose



Medidas de curtose medem o grau de achatamento de uma distribuição de frequências unimodal, medido em relação ao de uma distribuição normal (curva normal)

Medidas de achatamento ou curtose

- Na prática, as medidas de curtose medem o grau de concentração de valores da distribuição em torno do centro desta distribuição.
- Numa distribuição unimodal, quanto maior for a concentração de valores em torno do centro, menor será o valor da sua curtose.
- Uma distribuição de frequência é:
  - **Mesocúrtica**: quando apresenta uma grau de achatamento igual a da distribuição normal.
  - **Platicúrtica**: quando apresenta grau de achatamento maior que a da distribuição normal.
  - **Leptocúrtica**: quando apresenta grau de achatamento menor que a da distribuição normal.

Medidas de achatamento ou curtose

Para medir o grau de *kurtosis* pode ser utilizada a seguinte medida:

$$K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10}}$$

Rule of thumb

- $K = 0.263$ a distribuição de frequência é mesocúrtica.
- $K > 0.263$ a distribuição de frequência é platicúrtica
- $K < 0.263$ a distribuição de frequência é leptocúrtica

Medidas de achatamento ou curtose