Estatística Descritiva

Material desenvolvido no projeto *Elaboração de material didático para o ensino da Estatística na UFES*. Autor principal: Prof. Dr. Alessandro José Queiroz Sarnaglia.
Apoio: Programa de aprimoramento e desenvolvimento do ensino (PRÓ-ENSINO).

Estrutura

Estrutura

Visão geral sobre estatística

Considere as seguintes afirmações que foram retiradas de publicações recentes:

- "As pessoas que comem três porções diárias de grãos integrais têm risco de sofrer problemas cardíacos reduzido em 37%";
- "Dos 1500 danos à espinha dorsal em menores de idade, 70% resultam de acidentes de carro e 68 dos feridos não estavam usando o cinto de segurança";
- "Espera-se que a produção americana de carvão sofra uma redução de 3.1% em 2007".

Baseiam-se na coleta de **observações**.

Visão geral sobre estatística

Considere as seguintes afirmações que foram retiradas de publicações recentes:

- "As pessoas que comem três porções diárias de grãos integrais têm risco de sofrer problemas cardíacos reduzido em 37%";
- "Dos 1500 danos à espinha dorsal em menores de idade, 70% resultam de acidentes de carro e 68 dos feridos não estavam usando o cinto de segurança";
- "Espera-se que a produção americana de carvão sofra uma redução de 3.1% em 2007".

Baseiam-se na coleta de **observações**.

Definição

Observações *consistem de informações coletadas através de medições, questionários ou contagens.*

Definição

A ***Estatística*** *é a ciência que coleta, organiza, analisa e interpreta conjunto de observações.*

Definição

*A **Estatística** é a ciência que coleta, organiza, analisa e interpreta conjunto de observações.*

Os conjuntos de dados estatísticos (observações) podem formar:

- uma população;
- ou uma amostra.

Definição

*A **Estatística** é a ciência que coleta, organiza, analisa e interpreta conjunto de observações.*

Os conjuntos de dados estatísticos (observações) podem formar:

- uma população;
- ou uma amostra.

Definição

*Uma **população** é o conjunto formado por todas observações de interesse.*

Definição

*A **Estatística** é a ciência que coleta, organiza, analisa e interpreta conjunto de observações.*

Os conjuntos de dados estatísticos (observações) podem formar:

- uma população;
- ou uma amostra.

Definição

*Uma **população** é o conjunto formado por todas observações de interesse.*

Interesse: descrever alguma característica de toda a população através de UMA medida.

Definição

*A **Estatística** é a ciência que coleta, organiza, analisa e interpreta conjunto de observações.*

Os conjuntos de dados estatísticos (observações) podem formar:

- uma população;
- ou uma amostra.

Definição

*Uma **população** é o conjunto formado por todas observações de interesse.*

Interesse: descrever alguma característica de toda a população através de UMA medida.

Definição

*Um quantitativo numérico utilizado para descrever alguma característica populacional é denominado **parâmetro**.*

Experimentos:

1. Um lote com N lâmpadas será inspecionado. A i-ésima lâmpada queima no instante T_i, $i = 1, \ldots, N$. Seja $\mu = \frac{1}{N} \sum_i T_i$ a média desses tempos de queima. Temos que μ é um parâmetro;

Experimentos:

1. Um lote com N lâmpadas será inspecionado. A i-ésima lâmpada queima no instante T_i, $i = 1, \ldots, N$. Seja $\mu = \frac{1}{N} \sum_i T_i$ a média desses tempos de queima. Temos que μ é um parâmetro;
2. Suponha que o Brasil tenha N trabalhadores com carteira assinada. Sejam $R_1, \ldots, R_N$ as rendas desses trabalhadores. A renda média desses indivíduos $\mu = \frac{1}{N} \sum_i R_i$ é um parâmetro populacional;

Experimentos:

1. Um lote com N lâmpadas será inspecionado. A i-ésima lâmpada queima no instante T_i, $i = 1, \ldots, N$. Seja $\mu = \frac{1}{N}\sum_i T_i$ a média desses tempos de queima. Temos que μ é um parâmetro;
2. Suponha que o Brasil tenha N trabalhadores com carteira assinada. Sejam $R_1, \ldots, R_N$ as rendas desses trabalhadores. A renda média desses indivíduos $\mu = \frac{1}{N}\sum_i R_i$ é um parâmetro populacional;
3. Os tempos de vida das pessoas nascidas em uma cidade no ano de 2000 serão observados. A média desses tempos é um parâmetro;

Experimentos:

1. Um lote com N lâmpadas será inspecionado. A i-ésima lâmpada queima no instante T_i, $i = 1, \ldots, N$. Seja $\mu = \frac{1}{N}\sum_i T_i$ a média desses tempos de queima. Temos que μ é um parâmetro;
2. Suponha que o Brasil tenha N trabalhadores com carteira assinada. Sejam $R_1, \ldots, R_N$ as rendas desses trabalhadores. A renda média desses indivíduos $\mu = \frac{1}{N}\sum_i R_i$ é um parâmetro populacional;
3. Os tempos de vida das pessoas nascidas em uma cidade no ano de 2000 serão observados. A média desses tempos é um parâmetro;
4. A proporção dos dias chuvosos em uma determinada localidade será investigada. Essa proporção é um parâmetro.

Experimentos:

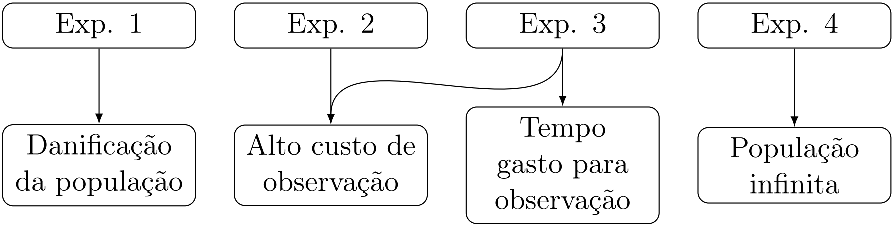
1. Um lote com N lâmpadas será inspecionado. A i-ésima lâmpada queima no instante T_i, $i = 1, \ldots, N$. Seja $\mu = \frac{1}{N}\sum_i T_i$ a média desses tempos de queima. Temos que μ é um parâmetro;
2. Suponha que o Brasil tenha N trabalhadores com carteira assinada. Sejam $R_1, \ldots, R_N$ as rendas desses trabalhadores. A renda média desses indivíduos $\mu = \frac{1}{N}\sum_i R_i$ é um parâmetro populacional;
3. Os tempos de vida das pessoas nascidas em uma cidade no ano de 2000 serão observados. A média desses tempos é um parâmetro;
4. A proporção dos dias chuvosos em uma determinada localidade será investigada. Essa proporção é um parâmetro.

Definição

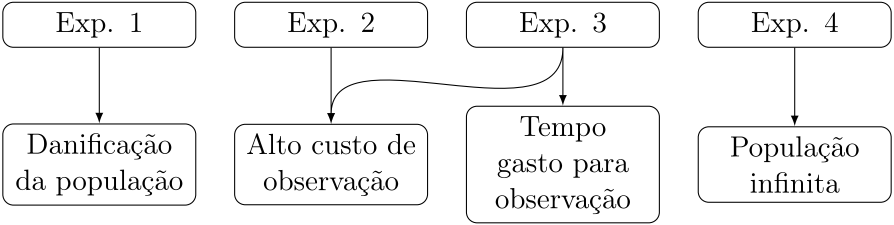
Um ***censo*** *é a coleta completa das observações de uma população.*

Problemas de um censo:

Problemas de um censo:

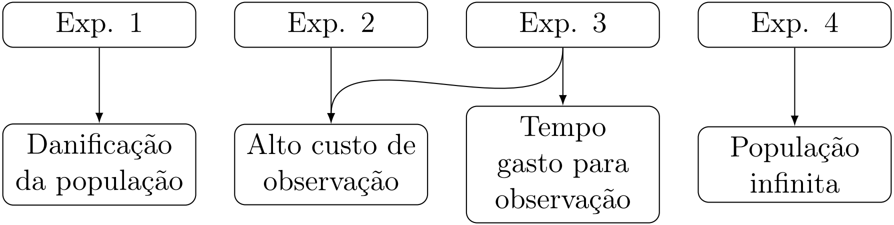


Problemas de um censo:



Alternativa: utilizar apenas uma parcela da população para tirar conclusões sobre a mesma.

Problemas de um censo:

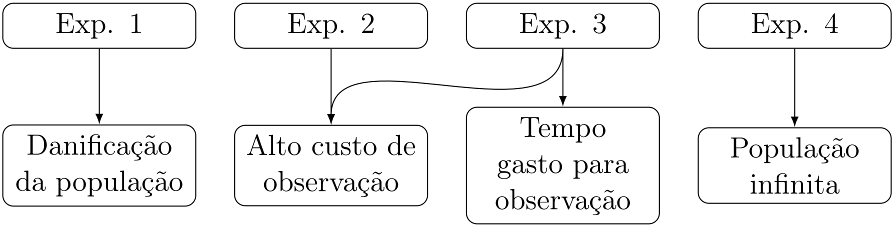


Alternativa: utilizar apenas uma parcela da população para tirar conclusões sobre a mesma.

Definição

Qualquer subconjunto de uma população é denominado de ***amostra****.*

Problemas de um censo:



Alternativa: utilizar apenas uma parcela da população para tirar conclusões sobre a mesma.

Definição

*Qualquer subconjunto de uma população é denominado de **amostra**.*

Definição

*Um quantitativo numérico utilizado para descrever alguma característica da amostra é denominado de **estatística**.*

Quando uma estatística é utilizada para aproximar o valor de um parâmetro, ela também é denominada de **estimador**.

Desvantagem:

amostra "ruim"

Desvantagem:

amostra “ruim” $\rightarrow$ Estimador fornecerá uma aproximação “ruim”.

Desvantagem:

amostra “ruim” $\rightarrow$ Estimador fornecerá uma aproximação “ruim”.

Solução: amostras “representativas” da população;

Desvantagem:

amostra “ruim” $\rightarrow$ Estimador fornecerá uma aproximação “ruim”.

Solução: amostras “representativas” da população;
Como? Principalmente, através de **amostragem aleatória simples**.

Desvantagem:

amostra “ruim” $\rightarrow$ Estimador fornecerá uma aproximação “ruim”.

Solução: amostras “representativas” da população;
Como? Principalmente, através de **amostragem aleatória simples**.

Definição

*A **amostragem aleatória simples** seleciona uma amostra de tamanho n de forma que **TODAS** as possíveis amostras de tamanho n tenham a mesma chance de serem selecionadas.*

Outras técnicas aleatórias: amostragem aleatória estratificada, por conglomerados, entre outras.

Desvantagem:

amostra “ruim” $\rightarrow$ Estimador fornecerá uma aproximação “ruim”.

Solução: amostras “representativas” da população;
Como? Principalmente, através de **amostragem aleatória simples**.

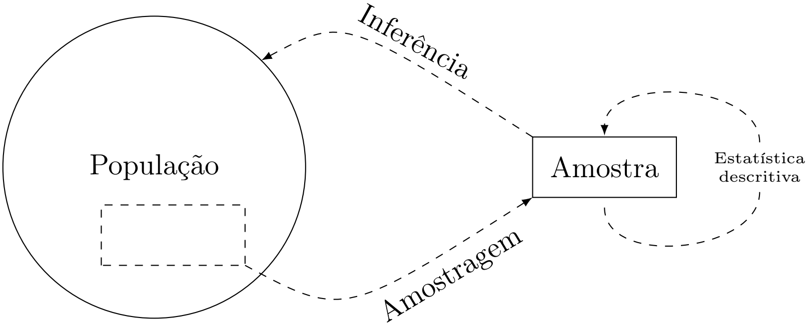
Definição

*A **amostragem aleatória simples** seleciona uma amostra de tamanho n de forma que **TODAS** as possíveis amostras de tamanho n tenham a mesma chance de serem selecionadas.*

Outras técnicas aleatórias: amostragem aleatória estratificada, por conglomerados, entre outras.
Estatística divide-se basicamente em:

- Estatística Descritiva - Técnicas para a organização, o resumo e a apresentação das observações. Se utilizada em uma amostra essas técnicas apenas descrevem **os dados em questão**;
- Inferência Estatística - Conjunto de técnicas que se baseia fortemente na teoria de probabilidade para extrapolar as informações de uma amostra para toda a população.

Representamos um estudo estatístico por meio do seguinte esquema:



Estrutura

Estrutura

Motivação

- Fato: uma pesquisa $\rightarrow$ grande quantidade de dados;
- Problema: nenhuma conclusão sem o correto tratamento dos dados;
- Solução: **estatística descritiva**.

Motivação

- Fato: uma pesquisa $\rightarrow$ grande quantidade de dados;
- Problema: nenhuma conclusão sem o correto tratamento dos dados;
- Solução: **estatística descritiva**.

Objetivos:

- resumo;
- visualização;
- e descrição.

Estatística Descritiva

Motivação

- Fato: uma pesquisa $\rightarrow$ grande quantidade de dados;
- Problema: nenhuma conclusão sem o correto tratamento dos dados;
- Solução: **estatística descritiva**.

Objetivos:

- resumo;
- visualização;
- e descrição.

Interesse: alguma característica (numérica, ou não) que TODOS elementos da amostra possuem.

Motivação

- Fato: uma pesquisa $\rightarrow$ grande quantidade de dados;
- Problema: nenhuma conclusão sem o correto tratamento dos dados;
- Solução: **estatística descritiva**.

Objetivos:

- resumo;
- visualização;
- e descrição.

Interesse: alguma característica (numérica, ou não) que TODOS elementos da amostra possuem.

Definição

*Uma **variável** é qualquer característica numérica, ou não, compartilhada por TODOS indivíduos de uma população.*

As variáveis são classificadas de acordo com os valores que assumem. Elas podem ser:

- qualitativas (ou categóricas), se seus valores são categorias, qualidades ou atributos dos indivíduos. Ainda se subdividem em:

As variáveis são classificadas de acordo com os valores que assumem. Elas podem ser:

- qualitativas (ou categóricas), se seus valores são categorias, qualidades ou atributos dos indivíduos. Ainda se subdividem em:
  - ordinais - quando essas categorias admitem alguma **ordenação** lógica;
  - nominais - quando esses atributos não podem ser ordenados lógicamente;

As variáveis são classificadas de acordo com os valores que assumem. Elas podem ser:

- qualitativas (ou categóricas), se seus valores são categorias, qualidades ou atributos dos indivíduos. Ainda se subdividem em:
  - ordinais - quando essas categorias admitem alguma **ordenação** lógica;
  - nominais - quando esses atributos não podem ser ordenados lógicamente;
- quantitativas, se seus valores são números, geralmente resultados de contagens ou medições. Ainda se subdividem em:

As variáveis são classificadas de acordo com os valores que assumem. Elas podem ser:

- qualitativas (ou categóricas), se seus valores são categorias, qualidades ou atributos dos indivíduos. Ainda se subdividem em:
    - ordinais - quando essas categorias admitem alguma **ordenação** lógica;
    - nominais - quando esses atributos não podem ser ordenados lógicamente;
- quantitativas, se seus valores são números, geralmente resultados de contagens ou medições. Ainda se subdividem em:
    - discretas - seus possíveis valores formam um conjunto finito, ou infinito **enumerável**;
    - contínuas - seus possíveis valores formam um conjunto não enumerável.

As variáveis são classificadas de acordo com os valores que assumem. Elas podem ser:

- qualitativas (ou categóricas), se seus valores são categorias, qualidades ou atributos dos indivíduos. Ainda se subdividem em:
    - ordinais - quando essas categorias admitem alguma **ordenação** lógica;
    - nominais - quando esses atributos não podem ser ordenados lógicamente;
- quantitativas, se seus valores são números, geralmente resultados de contagens ou medições. Ainda se subdividem em:
    - discretas - seus possíveis valores formam um conjunto finito, ou infinito **enumerável**;
    - contínuas - seus possíveis valores formam um conjunto não enumerável.

Definição

*Um conjunto é **enumerável** se é possível estabelecer uma bijeção entre seus elementos e um subconjunto de $\mathbb{Z}$.*

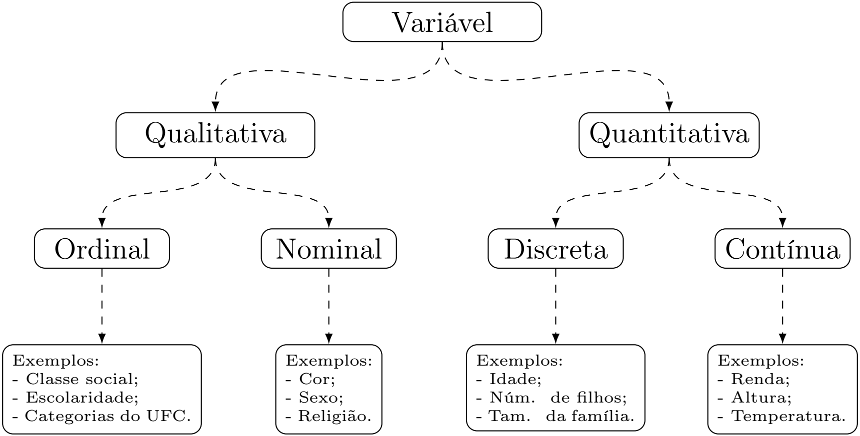
Comentário

Informalmente, um conjunto é enumerável quando os seus elementos podem ser contados. Assim, por exemplo, um intervalo do tipo $(a, b]$, *com* $a < b$, *é um conjunto não enumerável.*

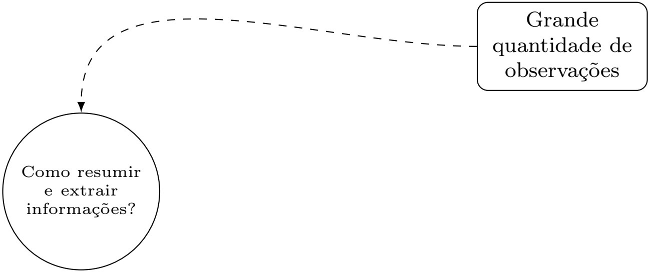
Comentário

Informalmente, um conjunto é enumerável quando os seus elementos podem ser contados. Assim, por exemplo, um intervalo do tipo $(a, b]$, com $a < b$, é um conjunto não enumerável.

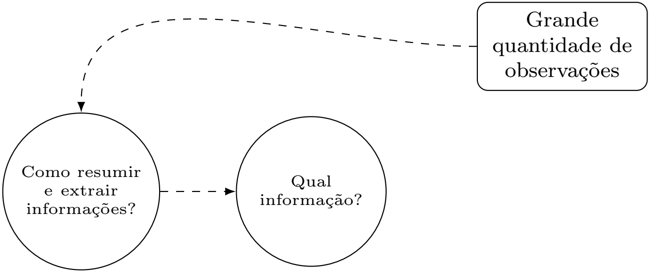
Esquematicamente:



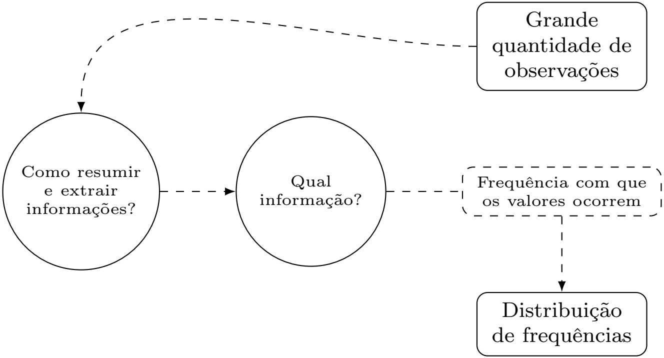
Quando utilizar a distribuição de frequências?



Quando utilizar a distribuição de frequências?



Quando utilizar a distribuição de frequências?



Estrutura

Distribuição de frequências

Seja X uma variável. Suponha que $x_1, x_2, \ldots, x_n$ são observações de X.

Seja X uma variável. Suponha que $x_1, x_2, \ldots, x_n$ são observações de X. As observações $x_1, x_2, \ldots, x_n$ podem ser os valores de X para a população completa ou apenas uma amostra de X. Agora, considere que:

Distribuição de frequências

Seja X uma variável. Suponha que $x_1, x_2, \ldots, x_n$ são observações de X. As observações $x_1, x_2, \ldots, x_n$ podem ser os valores de X para a população completa ou apenas uma amostra de X. Agora, considere que:

- dos valores $x_1, x_2, \ldots, x_n$, apenas $k \leq n$ são diferentes;

Distribuição de frequências

Seja X uma variável. Suponha que $x_1, x_2, \ldots, x_n$ são observações de X. As observações $x_1, x_2, \ldots, x_n$ podem ser os valores de X para a população completa ou apenas uma amostra de X. Agora, considere que:

- dos valores $x_1, x_2, \ldots, x_n$, apenas $k \leq n$ são diferentes;
- o restante são apenas repetições desses valores;

Distribuição de frequências

Seja X uma variável. Suponha que $x_1, x_2, \ldots, x_n$ são observações de X. As observações $x_1, x_2, \ldots, x_n$ podem ser os valores de X para a população completa ou apenas uma amostra de X. Agora, considere que:

- dos valores $x_1, x_2, \ldots, x_n$, apenas $k \leq n$ são diferentes;
- o restante são apenas repetições desses valores;
- sejam $x_1^*, x_2^*, \ldots, x_k^*$ as observações diferentes;

Distribuição de frequências

Seja X uma variável. Suponha que $x_1, x_2, \ldots, x_n$ são observações de X. As observações $x_1, x_2, \ldots, x_n$ podem ser os valores de X para a população completa ou apenas uma amostra de X. Agora, considere que:

- dos valores $x_1, x_2, \ldots, x_n$, apenas $k \leq n$ são diferentes;
- o restante são apenas repetições desses valores;
- sejam $x_1^*, x_2^*, \ldots, x_k^*$ as observações diferentes;
- suponha que o valor x_i^* se repetiu n_i vezes, $i = 1, 2, \ldots, k$.

Distribuição de frequências

Seja X uma variável. Suponha que $x_1, x_2, \dots, x_n$ são observações de X. As observações $x_1, x_2, \dots, x_n$ podem ser os valores de X para a população completa ou apenas uma amostra de X. Agora, considere que:

- dos valores $x_1, x_2, \dots, x_n$, apenas $k \leq n$ são diferentes;
- o restante são apenas repetições desses valores;
- sejam $x_1^*, x_2^*, \dots, x_k^*$ as observações diferentes;
- suponha que o valor x_i^* se repetiu n_i vezes, $i = 1, 2, \dots, k$.

Definição

A distribuição de frequências de X em $x_1, x_2, \dots, x_n$ é dada por

X	x_1^*	x_2^*	$\dots$	x_k^*	*Total*
Freq.	n_1	n_2	$\dots$	n_k	$\sum_{i=1}^{k} n_i = n$

Comentário

Ressaltamos que:

Comentário

Ressaltamos que:

- *se X for ordinal ou discreta, devemos dispor os x_i^* de forma que*

$$x_1^* < x_2^* < \cdots < x_k^*;$$

Comentário

Ressaltamos que:

- *se X for ordinal ou discreta, devemos dispor os x_i^* de forma que*

$$x_1^* < x_2^* < \ldots < x_k^*;$$

- *a distribuição de frequências fornece um resumo considerável dos dados;*

Comentário

Ressaltamos que:

- *se* X *for ordinal ou discreta, devemos dispor os* x_i^* *de forma que*

$$x_1^* < x_2^* < \ldots < x_k^*;$$

- *a distribuição de frequências fornece um resumo considerável dos dados;*
- *não há perda de informação ao utilizar essa técnica neste caso;*

Comentário

Ressaltamos que:

- *se X for ordinal ou discreta, devemos dispor os x_i^* de forma que*

$$x_1^* < x_2^* < \ldots < x_k^*;$$

- *a distribuição de frequências fornece um resumo considerável dos dados;*
- *não há perda de informação ao utilizar essa técnica neste caso;*
- *a frequência com que os valores de X ocorrem fica evidente;*

Comentário

Ressaltamos que:

- *se X for ordinal ou discreta, devemos dispor os x_i^* de forma que*

$$x_1^* < x_2^* < \ldots < x_k^*;$$

- *a distribuição de frequências fornece um resumo considerável dos dados;*
- *não há perda de informação ao utilizar essa técnica neste caso;*
- *a frequência com que os valores de X ocorrem fica evidente;*
- *utiliza-se esse tipo de distribuição de frequências no caso discreto quando $k \ll n$.*

Exemplo 1

Seja $X =$ "Tipo de música preferida". Neste caso $X \in \{p, r, s\}$, onde $p =$ pagode, $r =$ rock e $s =$ sertanejo. Suponha que $n = 40$ pessoas foram entrevistadas e o valor de X para cada uma delas foi verificado. Os dados são

$$
\begin{array}{cccccccccc}
s & p & p & p & s & p & s & s & r & p \\
s & r & p & s & r & p & s & s & p & p \\
p & s & r & s & s & p & p & p & s & s \\
p & r & s & s & r & s & p & p & p & s
\end{array}
$$

Exemplo 1

Seja X = "Tipo de música preferida". Neste caso $X \in \{p, r, s\}$, onde p = pagode, r = rock e s = sertanejo. Suponha que $n = 40$ pessoas foram entrevistadas e o valor de X para cada uma delas foi verificado. Os dados são

$$\begin{array}{cccccccccc}
s & \color{red}{p} & \color{red}{p} & \color{red}{p} & s & \color{red}{p} & s & s & r & \color{red}{p} \\
s & r & \color{red}{p} & s & r & \color{red}{p} & s & s & \color{red}{p} & \color{red}{p} \\
\color{red}{p} & s & r & s & s & \color{red}{p} & \color{red}{p} & \color{red}{p} & s & s \\
\color{red}{p} & r & s & s & r & s & \color{red}{p} & \color{red}{p} & \color{red}{p} & s
\end{array}$$

Temos que $n_p = 17$.

Exemplo 1

Seja $X =$ "Tipo de música preferida". Neste caso $X \in \{p, r, s\}$, onde $p =$ pagode, $r =$ rock e $s =$ sertanejo. Suponha que $n = 40$ pessoas foram entrevistadas e o valor de X para cada uma delas foi verificado. Os dados são

$$
\begin{array}{cccccccccc}
s & p & p & p & s & p & s & s & r & p \\
s & r & p & s & r & p & s & s & p & p \\
p & s & r & s & s & p & p & p & s & s \\
p & r & s & s & r & s & p & p & p & s
\end{array}
$$

Temos que $n_r = 6$.

Exemplo 1

Seja $X =$ "Tipo de música preferida". Neste caso $X \in \{p, r, s\}$, onde $p =$ pagode, $r =$ rock e $s =$ sertanejo. Suponha que $n = 40$ pessoas foram entrevistadas e o valor de X para cada uma delas foi verificado. Os dados são

$$
\begin{array}{cccccccccc}
\color{red}{s} & p & p & p & \color{red}{s} & p & \color{red}{s} & \color{red}{s} & r & p \\
\color{red}{s} & r & p & \color{red}{s} & r & p & \color{red}{s} & \color{red}{s} & p & p \\
p & \color{red}{s} & r & \color{red}{s} & \color{red}{s} & p & p & p & \color{red}{s} & \color{red}{s} \\
p & r & \color{red}{s} & \color{red}{s} & r & \color{red}{s} & p & p & p & \color{red}{s}
\end{array}
$$

Temos que $n_s = 17$.

Exemplo 1

Seja $X =$ "Tipo de música preferida". Neste caso $X \in \{p, r, s\}$, onde $p =$ pagode, $r =$ rock e $s =$ sertanejo. Suponha que $n = 40$ pessoas foram entrevistadas e o valor de X para cada uma delas foi verificado. Os dados são

$$\begin{array}{cccccccccc} s & p & p & p & s & p & s & s & r & p \\ s & r & p & s & r & p & s & s & p & p \\ p & s & r & s & s & p & p & p & s & s \\ p & r & s & s & r & s & p & p & p & s \end{array}$$

Assim, a distribuição de frequências de X é

X	p	r	s	Total
Freq.	17	6	17	40

Exemplo 2

O RH de uma empresa com 600 funcionários deseja fazer um levantamento com respeito a escolaridade e número de filhos (variáveis E e N, respectivamente) dos mesmos. Uma amostra de $n = 30$ funcionários forneceu as seguintes observações:

i	E_i	N_i	i	E_i	N_i	i	E_i	N_i
1	f	3	11	f	2	21	f	2
2	s	2	12	s	0	22	s	2
3	f	2	13	f	2	23	f	1
4	m	1	14	s	1	24	f	2
5	s	1	15	f	3	25	f	2
6	m	1	16	s	0	26	s	1
7	f	2	17	m	1	27	s	2
8	s	3	18	f	1	28	s	1
9	m	1	19	m	1	29	f	3
10	m	2	20	m	2	30	m	3

f = “fundamental”, m = “médio” e s = “superior”

Exemplo 2 - Variável E

O RH de uma empresa com 600 funcionários deseja fazer um levantamento com respeito a escolaridade e número de filhos (variáveis E e N, respectivamente) dos mesmos. Uma amostra de $n = 30$ funcionários forneceu as seguintes observações:

i	E_i	N_i	i	E_i	N_i	i	E_i	N_i
1	f	3	11	f	2	21	f	2
2	s	2	12	s	0	22	s	2
3	f	2	13	f	2	23	f	1
4	m	1	14	s	1	24	f	2
5	s	1	15	f	3	25	f	2
6	m	1	16	s	0	26	s	1
7	f	2	17	m	1	27	s	2
8	s	3	18	f	1	28	s	1
9	m	1	19	m	1	29	f	3
10	m	2	20	m	2	30	m	3

$n_f = 12$, $n_m = 8$ e $n_s = 10$

Exemplo 2 - Variável N

O RH de uma empresa com 600 funcionários deseja fazer um levantamento com respeito a escolaridade e número de filhos (variáveis E e N, respectivamente) dos mesmos. Uma amostra de $n = 30$ funcionários forneceu as seguintes observações:

i	E_i	N_i	i	E_i	N_i	i	E_i	N_i
1	f	3	11	f	2	21	f	2
2	s	2	12	s	0	22	s	2
3	f	2	13	f	2	23	f	1
4	m	1	14	s	1	24	f	2
5	s	1	15	f	3	25	f	2
6	m	1	16	s	0	26	s	1
7	f	2	17	m	1	27	s	2
8	s	3	18	f	1	28	s	1
9	m	1	19	m	1	29	f	3
10	m	2	20	m	2	30	m	3

$n_0 = 2$, $n_1 = 11$, $n_2 = 12$ e $n_3 = 5$

Assim, as distribuições de frequências de E e N são dadas por

E	f	m	s	Total
Freq.	12	8	10	30

e

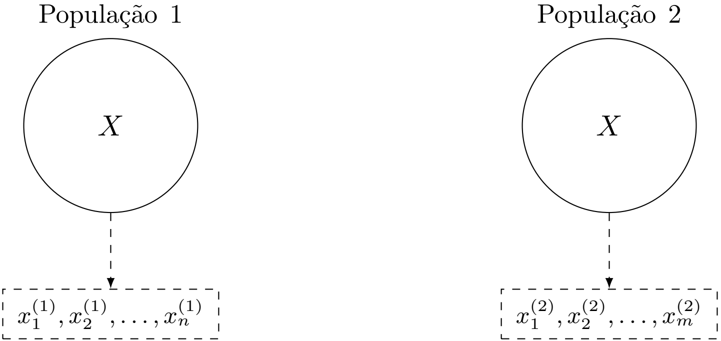
N	0	1	2	3	Total
Freq.	2	11	12	5	30

Distribuição de frequências relativas

Considere o seguinte esquema:

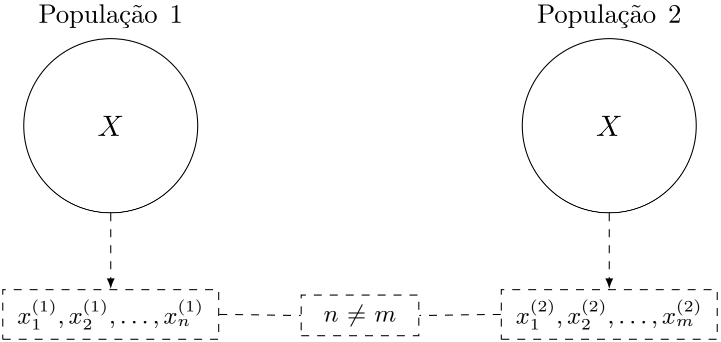
Distribuição de frequências relativas

Considere o seguinte esquema:



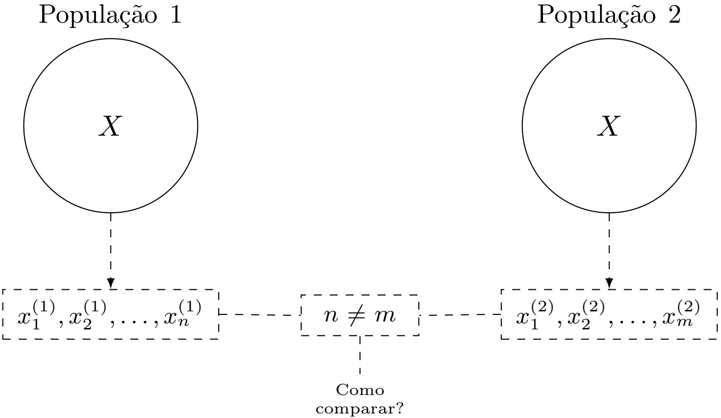
Distribuição de frequências relativas

Considere o seguinte esquema:



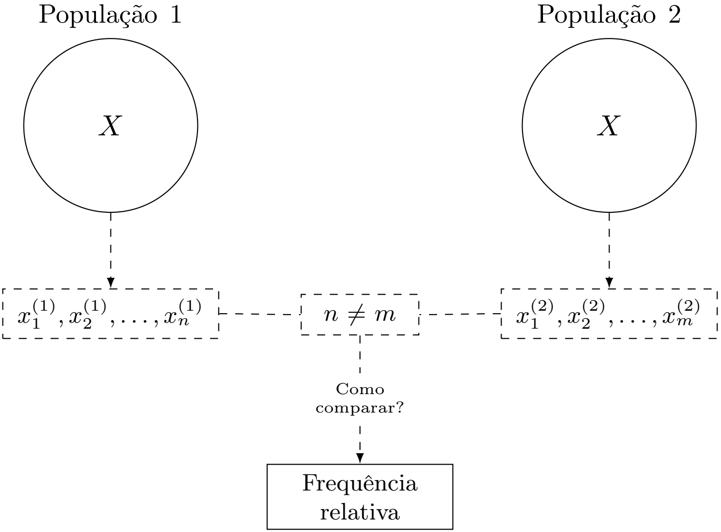
Distribuição de frequências relativas

Considere o seguinte esquema:



Distribuição de frequências relativas

Considere o seguinte esquema:



Distribuição de frequências relativas - Definição

Definição

Suponha X uma variável com a seguinte distribuição de frequências:

X	x_1^*	x_2^*	$\ldots$	x_k^*	*Total*
Freq.	n_1	n_2	$\ldots$	n_k	$\sum_{i=1}^k n_i = n$

A frequência relativa do i-ésimo valor (x_i^*) é definida por $f_i = \frac{n_i}{n}$. A distribuição de frequências relativas é dada por

X	x_1^*	x_2^*	$\ldots$	x_k^*	*Total*
Freq.	f_1	f_2	$\ldots$	f_k	$\sum_{i=1}^k f_i = 1$

Exemplo 2 - cont.

Suponha que todos 600 funcionários foram questionados e que as distribuições de frequências de E e N neste caso são:

E	f	m	s	Total
Freq.	230	180	190	600

e

N	0	1	2	3	Total
Freq.	30	230	235	105	600

Exemplo 2 - cont.

As distribuições de frequências relativas de E e N para a amostra e para a população são dadas, respectivamente, por

E	f	m	s	Total
Freq.	0.4	0.27	0.33	1

E	f	m	s	Total
Freq.	0.38	0.3	0.32	1

e

N	0	1	2	3	Total
Freq.	0.07	0.37	0.4	0.16	1

N	0	1	2	3	Total
Freq.	0.05	0.38	0.39	0.18	1

Estrutura

Distribuição de frequências

Neste caso: $x_1, \ldots, x_n$ tem muitas (ou todas) observações diferentes; o método anterior nem resume e nem extrai informação dos dados.

Distribuição de frequências

Neste caso: $x_1, \ldots, x_n$ tem muitas (ou todas) observações diferentes; o método anterior nem resume e nem extrai informação dos dados.
Alternativa - Agrupar valores próximos em intervalos. **Como?**

Distribuição de frequências

Neste caso: $x_1, \ldots, x_n$ tem muitas (ou todas) observações diferentes; o método anterior nem resume e nem extrai informação dos dados.

Alternativa - Agrupar valores próximos em intervalos. **Como?**

1. Escolha $a_0 < a_k$ tais que $x_1, \ldots, x_n \in [a_0, a_k]$;

Neste caso: $x_1, \ldots, x_n$ tem muitas (ou todas) observações diferentes; o método anterior nem resume e nem extrai informação dos dados.

Alternativa - Agrupar valores próximos em intervalos. **Como?**

1. Escolha $a_0 < a_k$ tais que $x_1, \ldots, x_n \in [a_0, a_k]$;
2. Fixe uma partição $I_1 = [a_0, a_1]$, $I_i = (a_{i-1}, a_i]$, $i = 2, \ldots, k$, isto é,

$$\bigcup_{i=1}^{k} I_i = [a_0, a_k] \text{ e } I_i \cap I_j = \emptyset, i \neq j;$$

Distribuição de frequências

Neste caso: $x_1, \ldots, x_n$ tem muitas (ou todas) observações diferentes; o método anterior nem resume e nem extrai informação dos dados.

Alternativa - Agrupar valores próximos em intervalos. **Como?**

1. Escolha $a_0 < a_k$ tais que $x_1, \ldots, x_n \in [a_0, a_k]$;
2. Fixe uma partição $I_1 = [a_0, a_1]$, $I_i = (a_{i-1}, a_i]$, $i = 2, \ldots, k$, isto é,
$$\bigcup_{i=1}^{k} I_i = [a_0, a_k] \text{ e } I_i \cap I_j = \emptyset, i \neq j;$$
3. A frequência n_i é o número de observações do intervalo I_i.

Distribuição de frequências

Neste caso: $x_1, \ldots, x_n$ tem muitas (ou todas) observações diferentes; o método anterior nem resume e nem extrai informação dos dados.

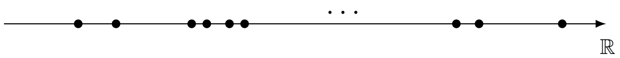
Alternativa - Agrupar valores próximos em intervalos. **Como?**

1. Escolha $a_0 < a_k$ tais que $x_1, \ldots, x_n \in [a_0, a_k]$;
2. Fixe uma partição $I_1 = [a_0, a_1]$, $I_i = (a_{i-1}, a_i]$, $i = 2, \ldots, k$, isto é,

$$\bigcup_{i=1}^{k} I_i = [a_0, a_k] \text{ e } I_i \cap I_j = \emptyset, i \neq j;$$

3. A frequência n_i é o número de observações do intervalo I_i.

Esquematicamente:

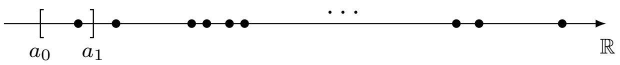


Distribuição de frequências

Neste caso: $x_1, \ldots, x_n$ tem muitas (ou todas) observações diferentes; o método anterior nem resume e nem extrai informação dos dados.
Alternativa - Agrupar valores próximos em intervalos. **Como?**

1. Escolha $a_0 < a_k$ tais que $x_1, \ldots, x_n \in [a_0, a_k]$;
2. Fixe uma partição $I_1 = [a_0, a_1]$, $I_i = (a_{i-1}, a_i]$, $i = 2, \ldots, k$, isto é,
$$\bigcup_{i=1}^{k} I_i = [a_0, a_k] \text{ e } I_i \cap I_j = \emptyset, i \neq j;$$
3. A frequência n_i é o número de observações do intervalo I_i.

Esquematicamente:



Distribuição de frequências

Neste caso: $x_1, \ldots, x_n$ tem muitas (ou todas) observações diferentes; o método anterior nem resume e nem extrai informação dos dados.

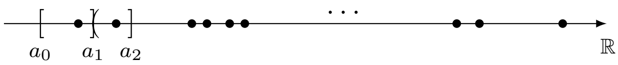
Alternativa - Agrupar valores próximos em intervalos. **Como?**

1. Escolha $a_0 < a_k$ tais que $x_1, \ldots, x_n \in [a_0, a_k]$;
2. Fixe uma partição $I_1 = [a_0, a_1]$, $I_i = (a_{i-1}, a_i]$, $i = 2, \ldots, k$, isto é,

$$\bigcup_{i=1}^{k} I_i = [a_0, a_k] \text{ e } I_i \cap I_j = \emptyset, i \neq j;$$

3. A frequência n_i é o número de observações do intervalo I_i.

Esquematicamente:



Distribuição de frequências

Neste caso: $x_1, \ldots, x_n$ tem muitas (ou todas) observações diferentes; o método anterior nem resume e nem extrai informação dos dados.

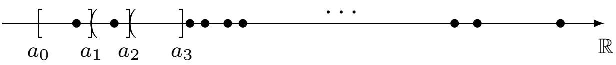
Alternativa - Agrupar valores próximos em intervalos. **Como?**

1. Escolha $a_0 < a_k$ tais que $x_1, \ldots, x_n \in [a_0, a_k]$;
2. Fixe uma partição $I_1 = [a_0, a_1]$, $I_i = (a_{i-1}, a_i]$, $i = 2, \ldots, k$, isto é,

$$\bigcup_{i=1}^{k} I_i = [a_0, a_k] \text{ e } I_i \cap I_j = \emptyset, i \neq j;$$

3. A frequência n_i é o número de observações do intervalo I_i.

Esquematicamente:



Distribuição de frequências

Neste caso: $x_1, \ldots, x_n$ tem muitas (ou todas) observações diferentes; o método anterior nem resume e nem extrai informação dos dados.

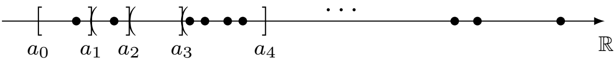
Alternativa - Agrupar valores próximos em intervalos. **Como?**

1. Escolha $a_0 < a_k$ tais que $x_1, \ldots, x_n \in [a_0, a_k]$;
2. Fixe uma partição $I_1 = [a_0, a_1]$, $I_i = (a_{i-1}, a_i]$, $i = 2, \ldots, k$, isto é,

$$\bigcup_{i=1}^{k} I_i = [a_0, a_k] \text{ e } I_i \cap I_j = \emptyset, i \neq j;$$

3. A frequência n_i é o número de observações do intervalo I_i.

Esquematicamente:



Distribuição de frequências

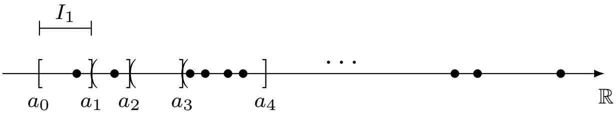
Neste caso: $x_1, \ldots, x_n$ tem muitas (ou todas) observações diferentes; o método anterior nem resume e nem extrai informação dos dados.
Alternativa - Agrupar valores próximos em intervalos. **Como?**

1. Escolha $a_0 < a_k$ tais que $x_1, \ldots, x_n \in [a_0, a_k]$;
2. Fixe uma partição $I_1 = [a_0, a_1]$, $I_i = (a_{i-1}, a_i]$, $i = 2, \ldots, k$, isto é,

$$\bigcup_{i=1}^{k} I_i = [a_0, a_k] \text{ e } I_i \cap I_j = \emptyset, i \neq j;$$

3. A frequência n_i é o número de observações do intervalo I_i.

Esquematicamente:



Distribuição de frequências

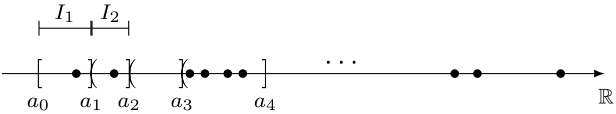
Neste caso: $x_1, \ldots, x_n$ tem muitas (ou todas) observações diferentes; o método anterior nem resume e nem extrai informação dos dados.
Alternativa - Agrupar valores próximos em intervalos. **Como?**

1. Escolha $a_0 < a_k$ tais que $x_1, \ldots, x_n \in [a_0, a_k]$;
2. Fixe uma partição $I_1 = [a_0, a_1]$, $I_i = (a_{i-1}, a_i]$, $i = 2, \ldots, k$, isto é,

$$\bigcup_{i=1}^{k} I_i = [a_0, a_k] \text{ e } I_i \cap I_j = \emptyset, i \neq j;$$

3. A frequência n_i é o número de observações do intervalo I_i.

Esquematicamente:

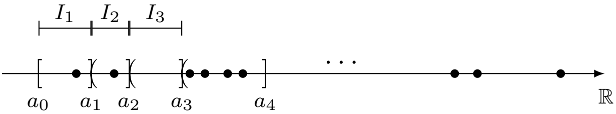


Distribuição de frequências

Neste caso: $x_1, \ldots, x_n$ tem muitas (ou todas) observações diferentes; o método anterior nem resume e nem extrai informação dos dados.
Alternativa - Agrupar valores próximos em intervalos. **Como?**

1. Escolha $a_0 < a_k$ tais que $x_1, \ldots, x_n \in [a_0, a_k]$;
2. Fixe uma partição $I_1 = [a_0, a_1]$, $I_i = (a_{i-1}, a_i]$, $i = 2, \ldots, k$, isto é,
$$\bigcup_{i=1}^{k} I_i = [a_0, a_k] \text{ e } I_i \cap I_j = \emptyset, i \neq j;$$
3. A frequência n_i é o número de observações do intervalo I_i.

Esquematicamente:



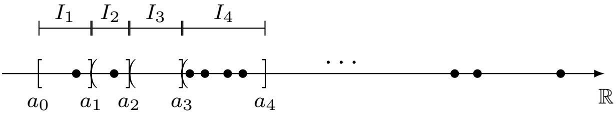
Distribuição de frequências

Neste caso: $x_1, \ldots, x_n$ tem muitas (ou todas) observações diferentes; o método anterior nem resume e nem extrai informação dos dados.

Alternativa - Agrupar valores próximos em intervalos. **Como?**

1. Escolha $a_0 < a_k$ tais que $x_1, \ldots, x_n \in [a_0, a_k]$;
2. Fixe uma partição $I_1 = [a_0, a_1]$, $I_i = (a_{i-1}, a_i]$, $i = 2, \ldots, k$, isto é,
$$\bigcup_{i=1}^{k} I_i = [a_0, a_k] \text{ e } I_i \cap I_j = \emptyset, i \neq j;$$
3. A frequência n_i é o número de observações do intervalo I_i.

Esquematicamente:



Distribuição de frequências

Neste caso: $x_1, \ldots, x_n$ tem muitas (ou todas) observações diferentes; o método anterior nem resume e nem extrai informação dos dados.

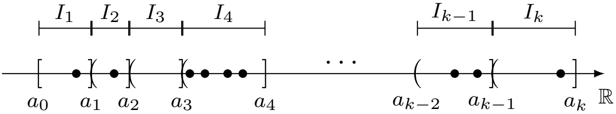
Alternativa - Agrupar valores próximos em intervalos. **Como?**

1. Escolha $a_0 < a_k$ tais que $x_1, \ldots, x_n \in [a_0, a_k]$;
2. Fixe uma partição $I_1 = [a_0, a_1]$, $I_i = (a_{i-1}, a_i]$, $i = 2, \ldots, k$, isto é,

$$\bigcup_{i=1}^{k} I_i = [a_0, a_k] \text{ e } I_i \cap I_j = \emptyset, i \neq j;$$

3. A frequência n_i é o número de observações do intervalo I_i.

Esquematicamente:



Neste caso: $x_1, \ldots, x_n$ tem muitas (ou todas) observações diferentes; o método anterior nem resume e nem extrai informação dos dados.

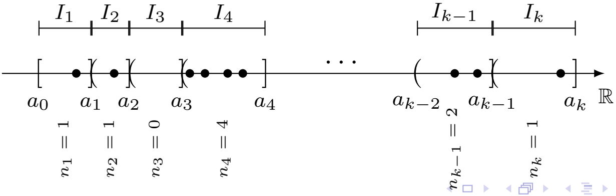
Alternativa - Agrupar valores próximos em intervalos. **Como?**

1. Escolha $a_0 < a_k$ tais que $x_1, \ldots, x_n \in [a_0, a_k]$;
2. Fixe uma partição $I_1 = [a_0, a_1]$, $I_i = (a_{i-1}, a_i]$, $i = 2, \ldots, k$, isto é,

$$\bigcup_{i=1}^{k} I_i = [a_0, a_k] \text{ e } I_i \cap I_j = \emptyset, i \neq j;$$

3. A frequência n_i é o número de observações do intervalo I_i.

Esquematicamente:



Definição

A distribuição de frequências dos dados $x_1, \dots, x_n$ *é definida por*

X	$[a_0, a_1]$	$(a_1, a_2]$	$\dots$	$(a_{k-1}, a_k]$	*Total*
Freq.	n_1	n_2	$\dots$	n_k	$\sum_{i=1}^{k} n_i = n$

Amostras de tamanhos diferentes são comparadas por meio de suas frequências relativas.

Definição

A distribuição de frequências relativas de $x_1, \dots, x_n$ *é definida por*

X	$[a_0, a_1]$	$(a_1, a_2]$	$\dots$	$(a_{k-1}, a_k]$	*Total*
Freq.	f_1	f_2	$\dots$	f_k	$\sum_{i=1}^{k} f_i = 1$

onde $f_i = n_i/n$, $i = 1, 2, \dots, k$.

Problema:

intervalo de grande amplitude

Problema

Problema:

intervalo de grande amplitude $\rightarrow$ possível frequência relativa alta.

Problema

Problema:

intervalo de grande amplitude $\rightarrow$ possível frequência relativa alta.

Alternativa 1: estabelecer intervalos de mesma amplitude.

Problema

Problema:

intervalo de grande amplitude $\rightarrow$ possível frequência relativa alta.

Alternativa 1: estabelecer intervalos de mesma amplitude. **Como?**

Problema

Problema:

intervalo de grande amplitude $\rightarrow$ possível frequência relativa alta.

Alternativa 1: estabelecer intervalos de mesma amplitude. **Como?**

- fixe a quantidade k de intervalos;

Problema

Problema:

intervalo de grande amplitude $\to$ possível frequência relativa alta.

Alternativa 1: estabelecer intervalos de mesma amplitude. **Como?**

- fixe a quantidade k de intervalos;
- tome $a_0 = \min(x_1, x_2, \ldots, x_n)$ e $a_k = \max(x_1, x_2, \ldots, x_n)$;

Problema:

intervalo de grande amplitude $\rightarrow$ possível frequência relativa alta.

Alternativa 1: estabelecer intervalos de mesma amplitude. **Como?**

- fixe a quantidade k de intervalos;
- tome $a_0 = \min(x_1, x_2, \ldots, x_n)$ e $a_k = \max(x_1, x_2, \ldots, x_n)$;
- obtenha a amplitude dos intervalos por $h = \frac{a_k - a_0}{k}$;

Problema

Problema:

intervalo de grande amplitude $\rightarrow$ possível frequência relativa alta.

Alternativa 1: estabelecer intervalos de mesma amplitude. **Como?**

- fixe a quantidade k de intervalos;
- tome $a_0 = \min(x_1, x_2, \ldots, x_n)$ e $a_k = \max(x_1, x_2, \ldots, x_n)$;
- obtenha a amplitude dos intervalos por $h = \frac{a_k - a_0}{k}$;
- calcule o limite superior do i-ésimo intervalo por

$$a_i = a_{i-1} + h = a_0 + ih, i = 1, 2, \ldots, k.$$

Problema

Problema:

intervalo de grande amplitude → possível frequência relativa alta.

Alternativa 1: estabelecer intervalos de mesma amplitude. **Como?**

- fixe a quantidade k de intervalos;
- tome $a_0 = \min(x_1, x_2, \ldots, x_n)$ e $a_k = \max(x_1, x_2, \ldots, x_n)$;
- obtenha a amplitude dos intervalos por $h = \frac{a_k - a_0}{k}$;
- calcule o limite superior do i-ésimo intervalo por

$$a_i = a_{i-1} + h = a_0 + ih, i = 1, 2, \ldots, k.$$

Comentário

1. $\uparrow k \Rightarrow \downarrow$ *Perda de informação,* $\downarrow$ *Resumo dos dados;*
2. $\downarrow k \Rightarrow \uparrow$ *Perda de informação,* $\uparrow$ *Resumo dos dados;*
3. *Sturges (1926) sugere a utilização de* $k = 1 + 3.322 \log_{10} n$.

Problema

Alternativa 2: utilizar a frequência relativa por unidade de medida da variável em estudo. Isto é, a **densidade**.

Problema

Alternativa 2: utilizar a frequência relativa por unidade de medida da variável em estudo. Isto é, a **densidade**.

Definição

Um intervalo I_i *com frequência relativa* f_i *tem densidade* $d_i = \frac{f_i}{h_i}$, *onde* h_i *é a amplitude deste intervalo*

Problema

Alternativa 2: utilizar a frequência relativa por unidade de medida da variável em estudo. Isto é, a **densidade**.

Definição

Um intervalo I_i com frequência relativa f_i tem densidade $d_i = \frac{f_i}{h_i}$, onde h_i é a amplitude deste intervalo

Comentário

Assim:

- *a densidade não é influenciada pela amplitude do intervalo;*

Problema

Alternativa 2: utilizar a frequência relativa por unidade de medida da variável em estudo. Isto é, a **densidade**.

Definição

Um intervalo I_i com frequência relativa f_i tem densidade $d_i = \frac{f_i}{h_i}$, onde h_i é a amplitude deste intervalo

Comentário

Assim:

- *a densidade não é influenciada pela amplitude do intervalo;*
- *quanto maior (menor) a densidade maior (menor) a frequência por unidade de medida da variável.*

Exemplo 3

Um determinado fabricante de baterias de carro deseja determinar a vida útil (em anos) de seus produtos. Para isso, a vida útil de $n = 40$ baterias foram observadas e os dados são apresentados abaixo.

2.0	4.1	3.5	4.5	3.2	3.7	3.0	2.6
3.4	1.7	3.1	3.3	3.8	3.1	4.7	3.7
1.9	4.3	3.4	3.6	2.9	3.3	3.9	3.1
3.3	3.1	3.7	4.4	3.2	4.1	1.9	3.4
4.7	3.8	3.2	2.6	3.9	3.0	4.2	3.5

Exemplo 3

Primeiramente, calculemos a distribuição de frequências para os intervalos

$$I_1 = [1.5, 2.0], I_2 = (2.0, 2.5], \ldots, I_7 = (4.5, 5.0].$$

Exemplo 3

Primeiramente, calculemos a distribuição de frequências para os intervalos

$$I_1 = [1.5, 2.0], I_2 = (2.0, 2.5], \ldots, I_7 = (4.5, 5.0].$$

Temos que a distribuição de frequências é dada por

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$	Total
Freq.	4	0	5	15	9	6	1	40

Exemplo 3

Primeiramente, calculemos a distribuição de frequências para os intervalos

$$I_1 = [1.5, 2.0], I_2 = (2.0, 2.5], \ldots, I_7 = (4.5, 5.0].$$

Temos que a distribuição de frequências é dada por

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$	Total
Freq.	4	0	5	15	9	6	1	40

A distribuição de frequências relativas é dada por

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$	Total
Freq.	0.100	0.000	0.125	0.375	0.225	0.150	0.025	1.000

Exemplo 3

Primeiramente, calculemos a distribuição de frequências para os intervalos

$$I_1 = [1.5, 2.0], I_2 = (2.0, 2.5], \ldots, I_7 = (4.5, 5.0].$$

Temos que a distribuição de frequências é dada por

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$	Total
Freq.	4	0	5	15	9	6	1	40

A distribuição de frequências relativas é dada por

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$	Total
Freq.	0.100	0.000	0.125	0.375	0.225	0.150	0.025	1.000

Comentário

- *O intervalo* $(2.0, 2.5]$ *tem frequência relativa* $f_2 = 0$*;*
- *A idéia é unir* $[1.5, 2.0]$ *e* $(2.0, 2.5]$ *para que não haja nenhum intervalo com frequência* 0*.*

Exemplo 3

Devemos utilizar as densidades. Neste caso, temos que:

X	$[1.5, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Freq.	0.100	0.250	0.750	0.450	0.300	0.050

Exemplo 3

Devemos utilizar as densidades. Neste caso, temos que:

X	$[1.5, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Freq.	0.100	0.250	0.750	0.450	0.300	0.050

Podemos também determinar a distribuição de frequências com intervalos iguais.

1. Tomemos $k = 1 + 3.322 \log_{10} 40 = 6.322 \approx 6$;
2. Temos que
$$a_0 = \min(x_1, \ldots, x_n) = 1.7 \quad \text{e} \quad a_k = a_6 = \max(x_1, \ldots, x_n) = 4.7;$$
3. A amplitude dos intervalos seria $h = \frac{a_k - a_0}{k} = \frac{4.7-1.7}{6} = \frac{3.0}{6} = 0.5$.

Exemplo 3

Devemos utilizar as densidades. Neste caso, temos que:

X	$[1.5, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Freq.	0.100	0.250	0.750	0.450	0.300	0.050

Podemos também determinar a distribuição de frequências com intervalos iguais.

1. Tomemos $k = 1 + 3.322 \log_{10} 40 = 6.322 \approx 6$;
2. Temos que
$$a_0 = \min(x_1, \ldots, x_n) = 1.7 \quad \text{e} \quad a_k = a_6 = \max(x_1, \ldots, x_n) = 4.7;$$
3. A amplitude dos intervalos seria $h = \frac{a_k - a_0}{k} = \frac{4.7-1.7}{6} = \frac{3.0}{6} = 0.5$.

A distribuição de frequências neste caso é dada por

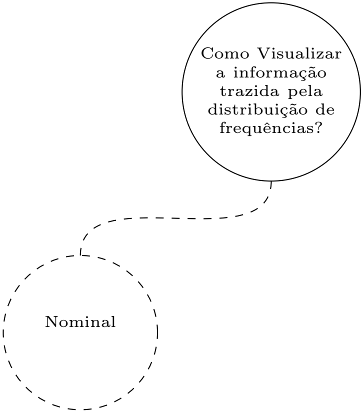
X	$[1.7, 2.2]$	$(2.2, 2.7]$	$(2.7, 3.2]$	$(3.2, 3.7]$	$(3.7, 4.2]$	$(4.2, 4.7]$	Total
Freq.	4	2	10	13	7	4	40

Estrutura

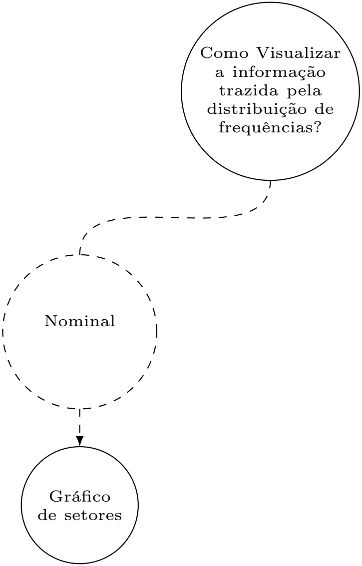
Gráficos

Como Visualizar a informação trazida pela distribuição de frequências?

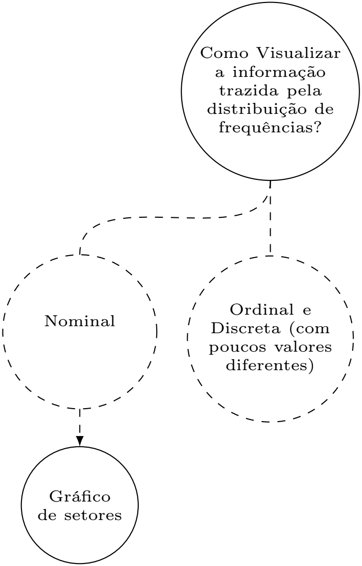
Gráficos


Como Visualizar a informação trazida pela distribuição de frequências?
Nominal

Gráficos

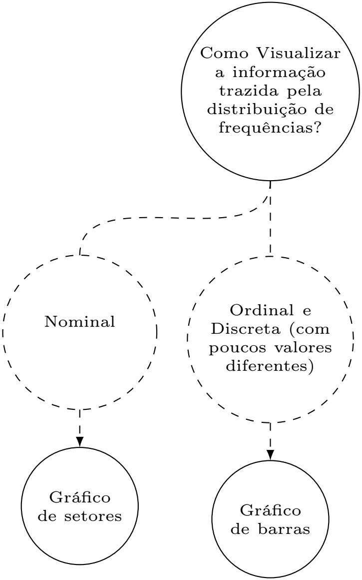

Como Visualizar a informação trazida pela distribuição de frequências?
Nominal
Gráfico de setores

Gráficos

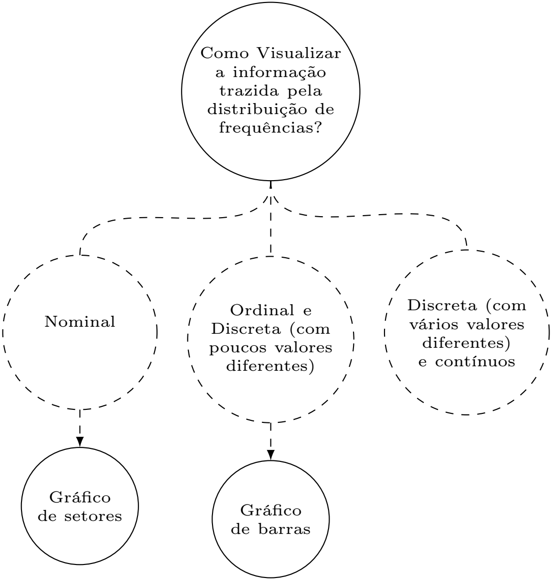


Como Visualizar a informação trazida pela distribuição de frequências?
Nominal
Ordinal e Discreta (com poucos valores diferentes)
Gráfico de setores

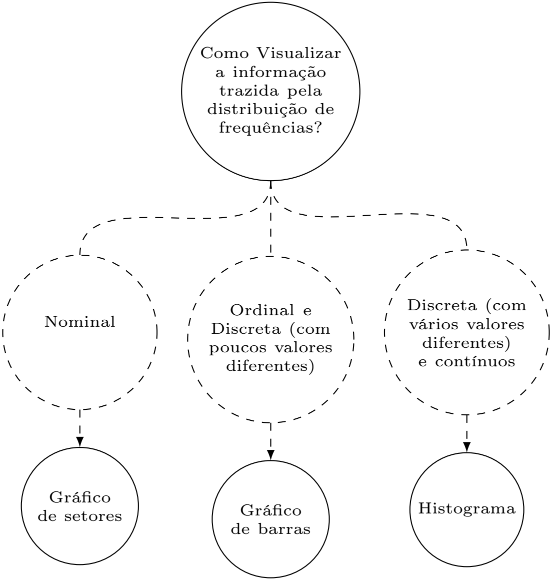
Gráficos


Como Visualizar a informação trazida pela distribuição de frequências?
Nominal
Ordinal e Discreta (com poucos valores diferentes)
Gráfico de setores
Gráfico de barras

Gráficos


Como Visualizar a informação trazida pela distribuição de frequências?
Nominal
Ordinal e Discreta (com poucos valores diferentes)
Discreta (com vários valores diferentes) e contínuos
Gráfico de setores
Gráfico de barras

Gráficos


Como Visualizar a informação trazida pela distribuição de frequências?
Nominal
Ordinal e Discreta (com poucos valores diferentes)
Discreta (com vários valores diferentes) e contínuos
Gráfico de setores
Gráfico de barras
Histograma

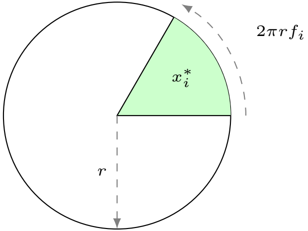
Estrutura

Gráfico de setores

É construido da seguinte maneira:

1. suponha que os valores diferentes da variável são $x_1^*, \ldots, x_k^*$;
2. sejam $f_1, \ldots, f_k$ as frequências relativas desses valores;
3. desenhe um círculo de raio arbitrário;
4. representamos no círculo desenhado o valor x_i^* através de um arco de circunferência de ângulo proporcional a sua frequência relativa. Isto é, um arco de comprimento $2\pi r f_i$, onde r é o raio arbitrário.

Esquematicamente:

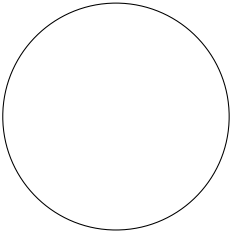


Exemplo 1 - cont.

Relembrando, a distribuição de frequências relativas da variável X = "Tipo de música preferida" é dada por

X	p	r	s	Total
Freq.	0.425	0.150	0.425	1.000

Assim, o gráfico de setores de X para a amostra observada no Exemplo 1 é dado abaixo.

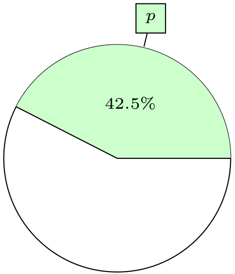


Exemplo 1 - cont.

Relembrando, a distribuição de frequências relativas da variável X = "Tipo de música preferida" é dada por

X	p	r	s	Total
Freq.	0.425	0.150	0.425	1.000

Assim, o gráfico de setores de X para a amostra observada no Exemplo 1 é dado abaixo.

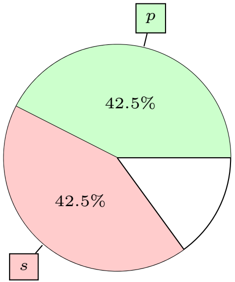


Exemplo 1 - cont.

Relembrando, a distribuição de frequências relativas da variável X = "Tipo de música preferida" é dada por

X	p	r	s	Total
Freq.	0.425	0.150	0.425	1.000

Assim, o gráfico de setores de X para a amostra observada no Exemplo 1 é dado abaixo.

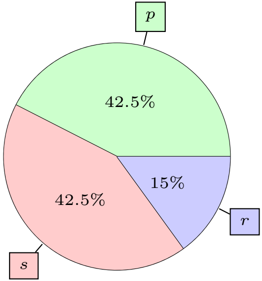


Exemplo 1 - cont.

Relembrando, a distribuição de frequências relativas da variável X = "Tipo de música preferida" é dada por

X	p	r	s	Total
Freq.	0.425	0.150	0.425	1.000

Assim, o gráfico de setores de X para a amostra observada no Exemplo 1 é dado abaixo.



Estrutura

É construido da seguinte maneira:

1. suponha que os valores diferentes da variável são $x_1^* < \cdots < x_k^*$;
2. sejam $f_1, \ldots, f_k$ as frequências relativas desses valores;
3. desenhe o eixo cartesiano;
4. no eixo das abscissas, em cima dos valores $x_1^*, \ldots, x_k^*$, posicione retângulos (barras);
5. as bases de todos retângulos devem ter o mesmo comprimento escolhido arbitrariamente;
6. a altura de um retângulo é dada pela frequência relativa do valor de X que é representado por ele. Isto é, a altura da barra que representa o valor x_i^* é dada por f_i.

Gráfico de barras

Esquematicamente:

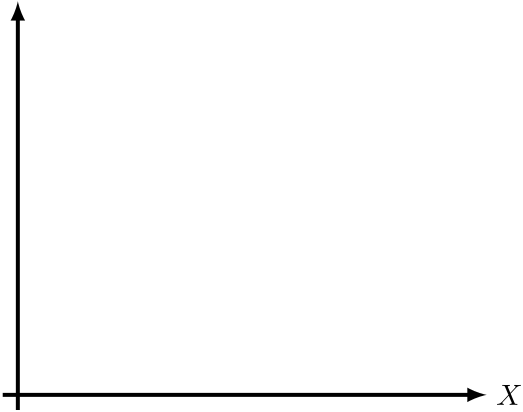


Gráfico de barras

Esquematicamente:

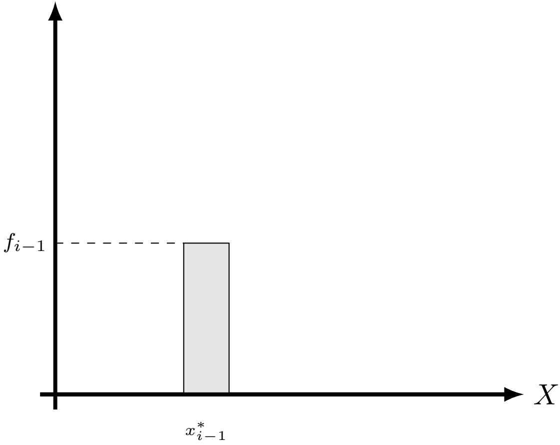


Gráfico de barras

Esquematicamente:

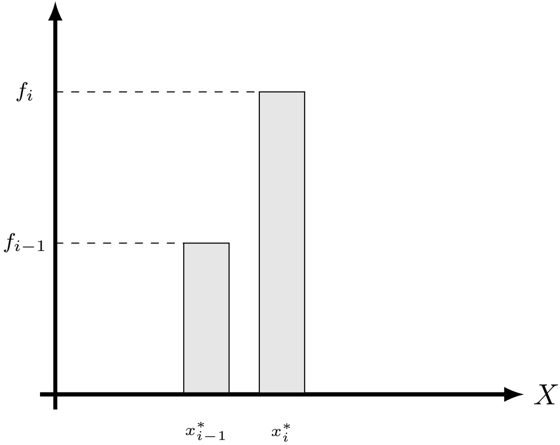


Gráfico de barras

Esquematicamente:

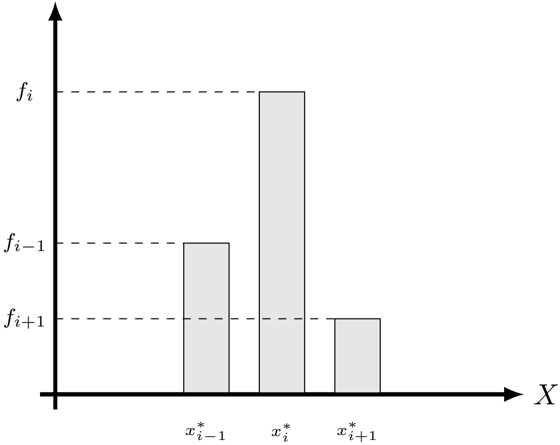
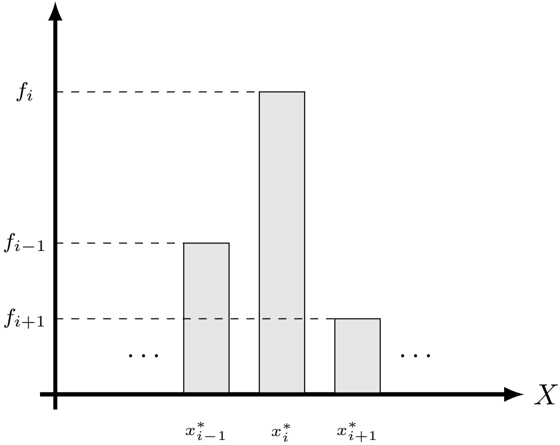


Gráfico de barras

Esquematicamente:

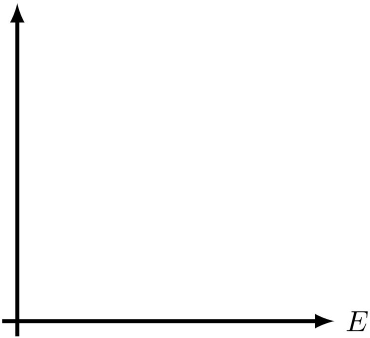


Exemplo 2 - cont.

Relembrando a distribuição de frequências relativas de E:

E	f	m	s	Total
Freq.	0.4	0.27	0.33	1

O gráfico de barras para a variável E nesta amostra é dado abaixo.

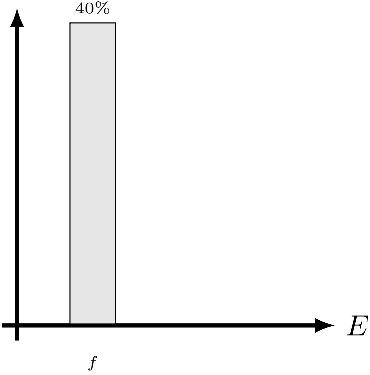


Exemplo 2 - cont.

Relembrando a distribuição de frequências relativas de E:

E	f	m	s	Total
Freq.	0.4	0.27	0.33	1

O gráfico de barras para a variável E nesta amostra é dado abaixo.

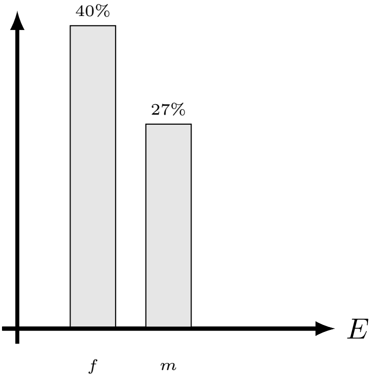


Exemplo 2 - cont.

Relembrando a distribuição de frequências relativas de E:

E	f	m	s	Total
Freq.	0.4	0.27	0.33	1

O gráfico de barras para a variável E nesta amostra é dado abaixo.

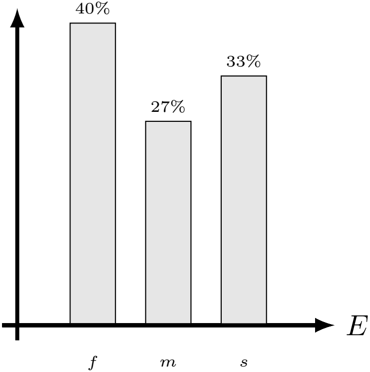


Exemplo 2 - cont.

Relembrando a distribuição de frequências relativas de E:

E	f	m	s	Total
Freq.	0.4	0.27	0.33	1

O gráfico de barras para a variável E nesta amostra é dado abaixo.



Exemplo 2 - cont.

Comentário

Em um gráfico de barras, um valor que teve frequência 0 *deve ser representado.*

Suponha que a variável número de filhos (N) do Exemplo 2 tivesse a seguinte distribuição de frequências relativas:

N	0	1	2	3	Total
Freq.	0.27	0.57	0.00	0.16	1

O gráfico de barras neste caso é dado abaixo.

Exemplo 2 - cont.

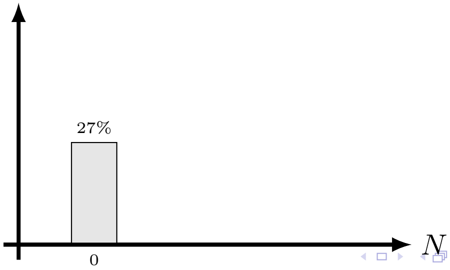
Comentário

Em um gráfico de barras, um valor que teve frequência 0 *deve ser representado.*

Suponha que a variável número de filhos (N) do Exemplo 2 tivesse a seguinte distribuição de frequências relativas:

N	0	1	2	3	Total
Freq.	0.27	0.57	0.00	0.16	1

O gráfico de barras neste caso é dado abaixo.



Exemplo 2 - cont.

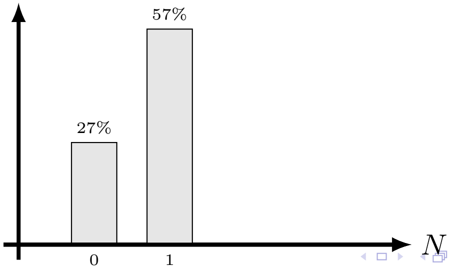
Comentário

Em um gráfico de barras, um valor que teve frequência 0 *deve ser representado.*

Suponha que a variável número de filhos (N) do Exemplo 2 tivesse a seguinte distribuição de frequências relativas:

N	0	1	2	3	Total
Freq.	0.27	0.57	0.00	0.16	1

O gráfico de barras neste caso é dado abaixo.



Exemplo 2 - cont.

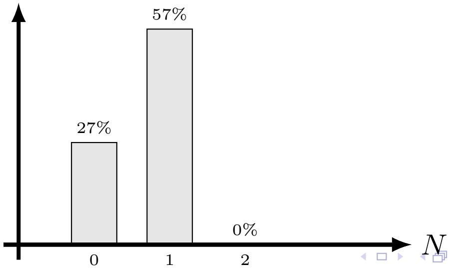
Comentário

Em um gráfico de barras, um valor que teve frequência 0 *deve ser representado.*

Suponha que a variável número de filhos (N) do Exemplo 2 tivesse a seguinte distribuição de frequências relativas:

N	0	1	2	3	Total
Freq.	0.27	0.57	0.00	0.16	1

O gráfico de barras neste caso é dado abaixo.



Exemplo 2 - cont.

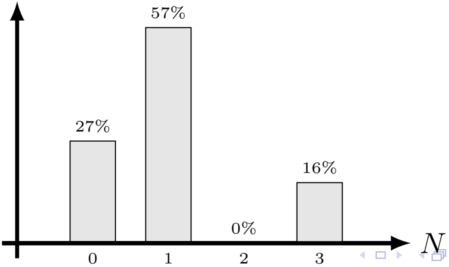
Comentário

Em um gráfico de barras, um valor que teve frequência 0 *deve ser representado.*

Suponha que a variável número de filhos (N) do Exemplo 2 tivesse a seguinte distribuição de frequências relativas:

N	0	1	2	3	Total
Freq.	0.27	0.57	0.00	0.16	1

O gráfico de barras neste caso é dado abaixo.



Estrutura

Histograma

É construido da seguinte maneira:

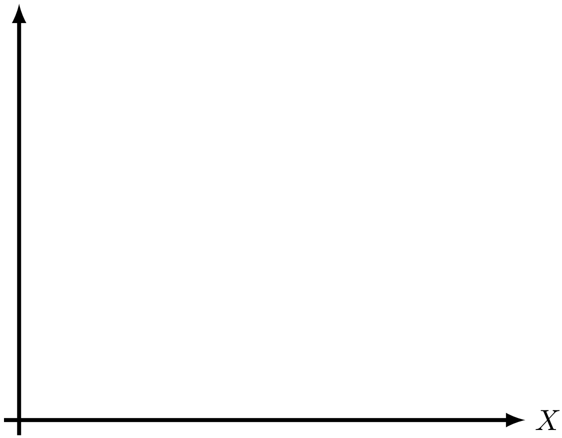
1. suponha que a variável X forneceu a seguinte distribuição de densidades:

X	$[a_0, a_1]$	$(a_1, a_2]$	$\dots$	$(a_{k-1}, a_k]$
Freq.	d_1	d_2	$\dots$	d_k

2. desenhe o eixo cartesiano;
3. no eixo das abscissas, em cima do intervalo $I_i = (a_{i-1}, a_i]$ posicione um retângulos;
4. a base desse retângulo deve corresponder ao intervalo I_i. Portanto, **não deve haver espaço entre as barras**;
5. a altura desse retângulo é dada pela densidade do i-ésimo intervalo, d_i.

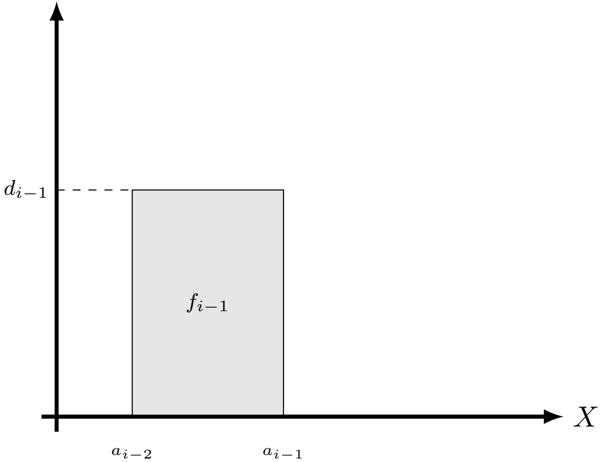
Histograma

Esquematicamente:



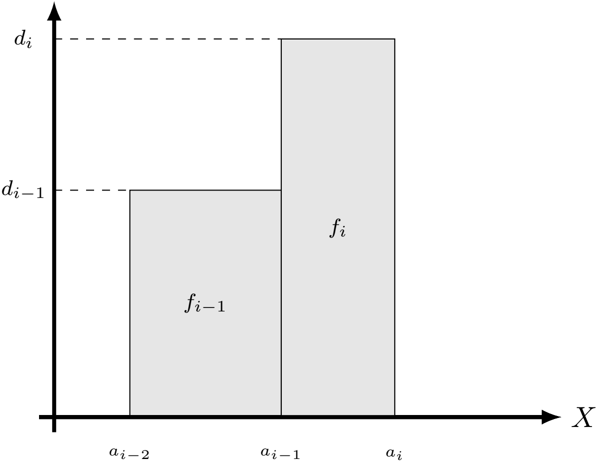
Histograma

Esquematicamente:



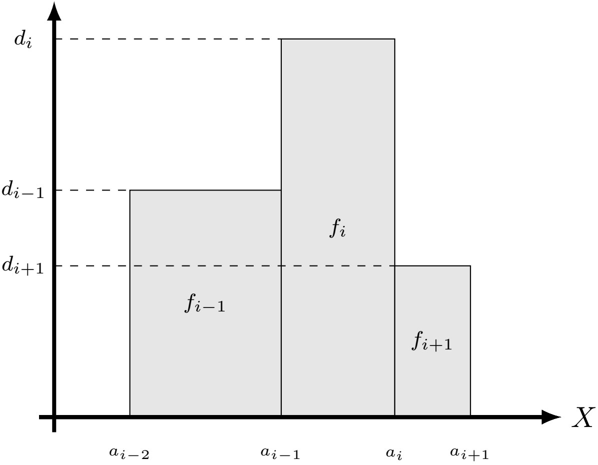
Histograma

Esquematicamente:



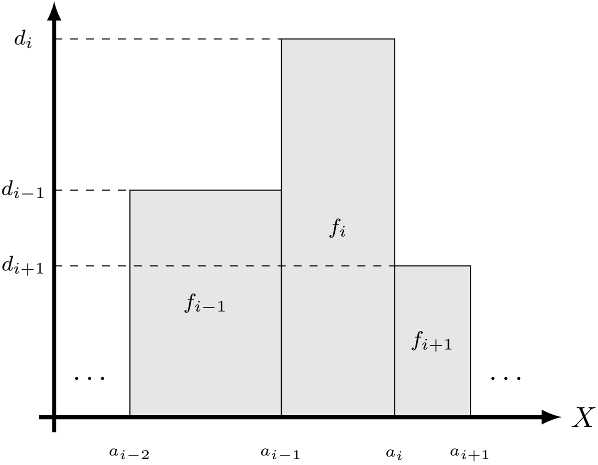
Histograma

Esquematicamente:



Histograma

Esquematicamente:



Exemplo 3 - cont.

Relembrando a distribuição de frequências relativas do Exemplo 3.

X	[1.5, 2.0]	(2.0, 2.5]	(2.5, 3.0]	(3.0, 3.5]	(3.5, 4.0]	(4.0, 4.5]	(4.5, 5.0]	Total
Freq.	0.100	0.000	0.125	0.375	0.225	0.150	0.025	1.000

Assim, a distribuição de densidades é dada abaixo.

X	[1.5, 2.0]	(2.0, 2.5]	(2.5, 3.0]	(3.0, 3.5]	(3.5, 4.0]	(4.0, 4.5]	(4.5, 5.0]
Freq.	0.200	0.000	0.250	0.750	0.450	0.300	0.050

O histograma por sua vez é apresentado abaixo.

Exemplo 3 - cont.

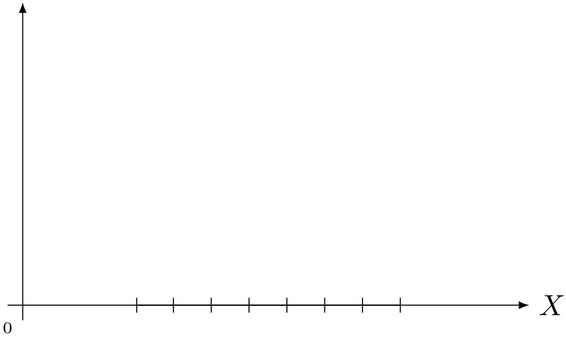
Relembrando a distribuição de frequências relativas do Exemplo 3.

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$	Total
Freq.	0.100	0.000	0.125	0.375	0.225	0.150	0.025	1.000

Assim, a distribuição de densidades é dada abaixo.

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Freq.	0.200	0.000	0.250	0.750	0.450	0.300	0.050

O histograma por sua vez é apresentado abaixo.



Exemplo 3 - cont.

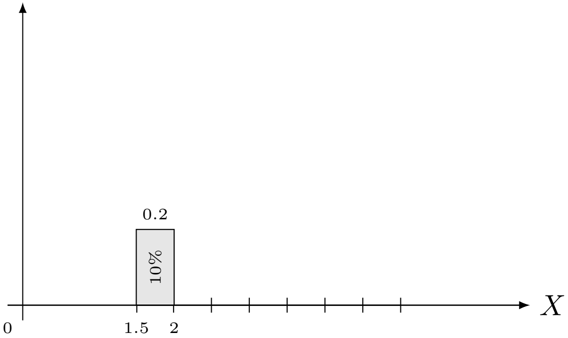
Relembrando a distribuição de frequências relativas do Exemplo 3.

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$	Total
Freq.	0.100	0.000	0.125	0.375	0.225	0.150	0.025	1.000

Assim, a distribuição de densidades é dada abaixo.

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Freq.	0.200	0.000	0.250	0.750	0.450	0.300	0.050

O histograma por sua vez é apresentado abaixo.



Exemplo 3 - cont.

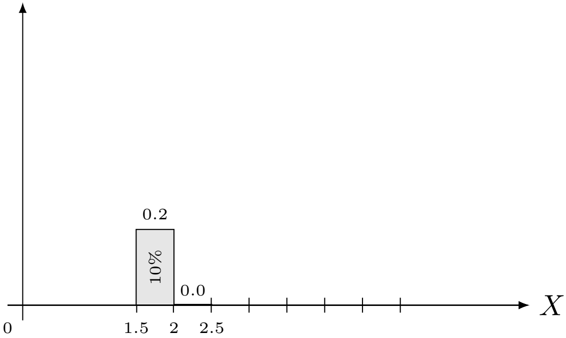
Relembrando a distribuição de frequências relativas do Exemplo 3.

X	[1.5, 2.0]	(2.0, 2.5]	(2.5, 3.0]	(3.0, 3.5]	(3.5, 4.0]	(4.0, 4.5]	(4.5, 5.0]	Total
Freq.	0.100	0.000	0.125	0.375	0.225	0.150	0.025	1.000

Assim, a distribuição de densidades é dada abaixo.

X	[1.5, 2.0]	(2.0, 2.5]	(2.5, 3.0]	(3.0, 3.5]	(3.5, 4.0]	(4.0, 4.5]	(4.5, 5.0]
Freq.	0.200	0.000	0.250	0.750	0.450	0.300	0.050

O histograma por sua vez é apresentado abaixo.



Exemplo 3 - cont.

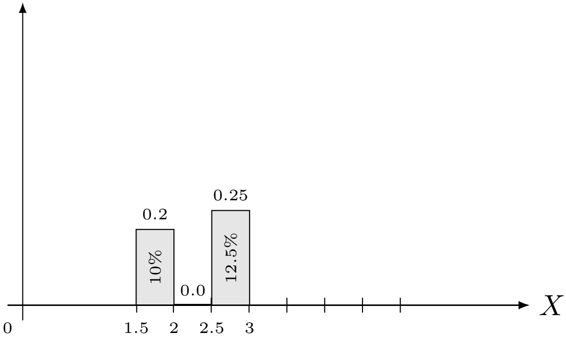
Relembrando a distribuição de frequências relativas do Exemplo 3.

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$	Total
Freq.	0.100	0.000	0.125	0.375	0.225	0.150	0.025	1.000

Assim, a distribuição de densidades é dada abaixo.

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Freq.	0.200	0.000	0.250	0.750	0.450	0.300	0.050

O histograma por sua vez é apresentado abaixo.



Exemplo 3 - cont.

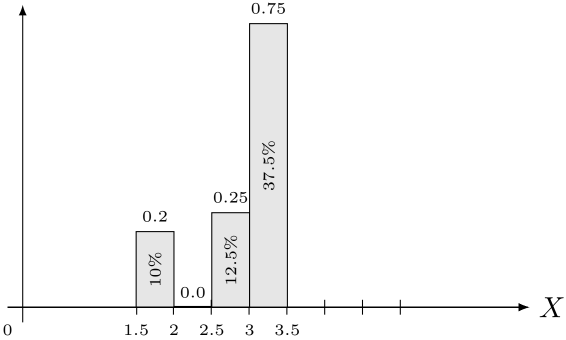
Relembrando a distribuição de frequências relativas do Exemplo 3.

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$	Total
Freq.	0.100	0.000	0.125	0.375	0.225	0.150	0.025	1.000

Assim, a distribuição de densidades é dada abaixo.

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Freq.	0.200	0.000	0.250	0.750	0.450	0.300	0.050

O histograma por sua vez é apresentado abaixo.



Exemplo 3 - cont.

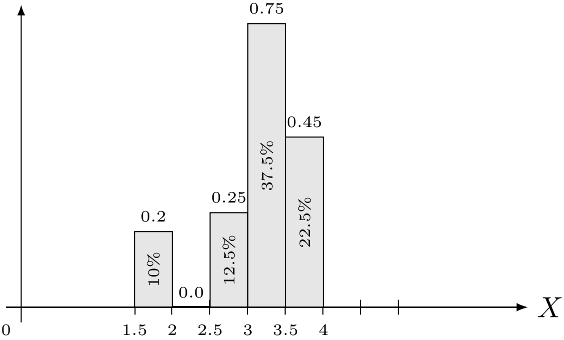
Relembrando a distribuição de frequências relativas do Exemplo 3.

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$	Total
Freq.	0.100	0.000	0.125	0.375	0.225	0.150	0.025	1.000

Assim, a distribuição de densidades é dada abaixo.

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Freq.	0.200	0.000	0.250	0.750	0.450	0.300	0.050

O histograma por sua vez é apresentado abaixo.



Exemplo 3 - cont.

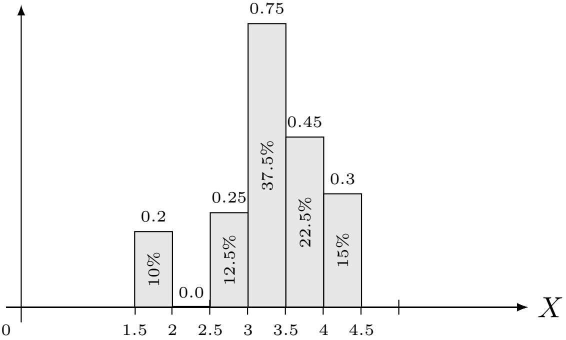
Relembrando a distribuição de frequências relativas do Exemplo 3.

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$	Total
Freq.	0.100	0.000	0.125	0.375	0.225	0.150	0.025	1.000

Assim, a distribuição de densidades é dada abaixo.

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Freq.	0.200	0.000	0.250	0.750	0.450	0.300	0.050

O histograma por sua vez é apresentado abaixo.



Exemplo 3 - cont.

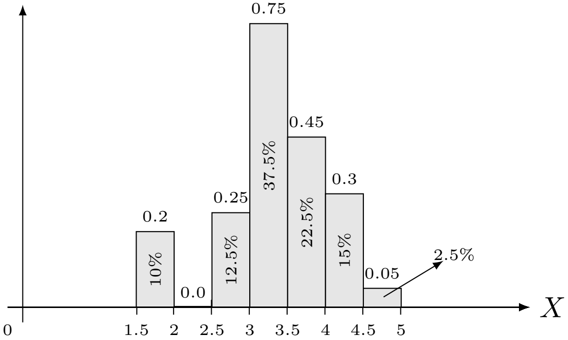
Relembrando a distribuição de frequências relativas do Exemplo 3.

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$	Total
Freq.	0.100	0.000	0.125	0.375	0.225	0.150	0.025	1.000

Assim, a distribuição de densidades é dada abaixo.

X	$[1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Freq.	0.200	0.000	0.250	0.750	0.450	0.300	0.050

O histograma por sua vez é apresentado abaixo.



Exemplo 3 - cont.

Relembrando, quando os dois primeiros intervalos foram unidos a distribuição de densidades é dada abaixo.

X	$[1.5, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Dens.	0.100	0.250	0.750	0.450	0.300	0.050

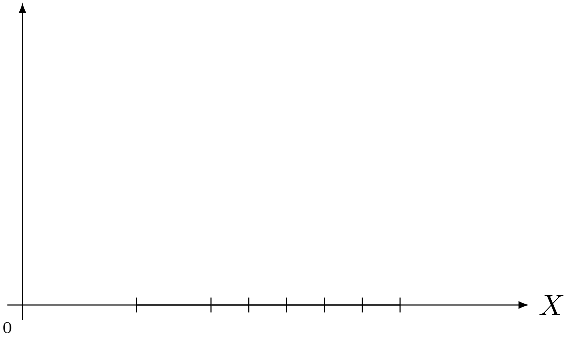
Neste caso, o histograma é dado abaixo. Note que não há mais intervalos de frequência 0.

Exemplo 3 - cont.

Relembrando, quando os dois primeiros intervalos foram unidos a distribuição de densidades é dada abaixo.

X	$[1.5, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Dens.	0.100	0.250	0.750	0.450	0.300	0.050

Neste caso, o histograma é dado abaixo. Note que não há mais intervalos de frequência 0.

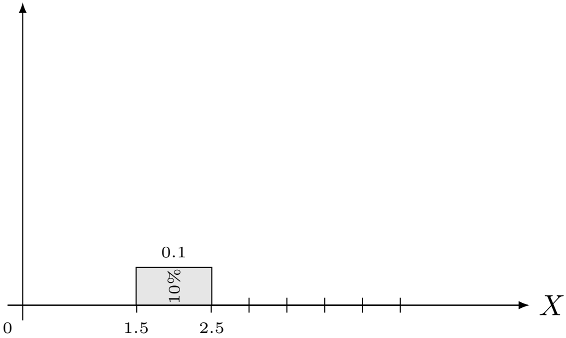


Exemplo 3 - cont.

Relembrando, quando os dois primeiros intervalos foram unidos a distribuição de densidades é dada abaixo.

X	$[1.5, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Dens.	0.100	0.250	0.750	0.450	0.300	0.050

Neste caso, o histograma é dado abaixo. Note que não há mais intervalos de frequência 0.

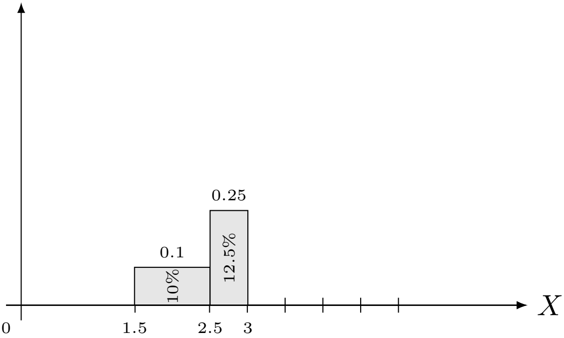


Exemplo 3 - cont.

Relembrando, quando os dois primeiros intervalos foram unidos a distribuição de densidades é dada abaixo.

X	$[1.5, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Dens.	0.100	0.250	0.750	0.450	0.300	0.050

Neste caso, o histograma é dado abaixo. Note que não há mais intervalos de frequência 0.

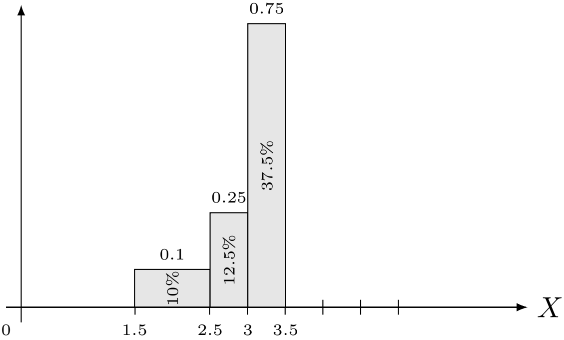


Exemplo 3 - cont.

Relembrando, quando os dois primeiros intervalos foram unidos a distribuição de densidades é dada abaixo.

X	$[1.5, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Dens.	0.100	0.250	0.750	0.450	0.300	0.050

Neste caso, o histograma é dado abaixo. Note que não há mais intervalos de frequência 0.

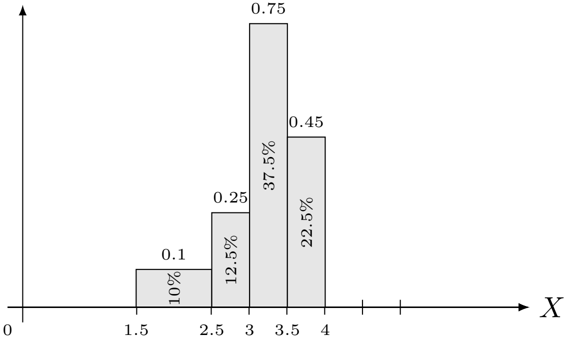


Exemplo 3 - cont.

Relembrando, quando os dois primeiros intervalos foram unidos a distribuição de densidades é dada abaixo.

X	$[1.5, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Dens.	0.100	0.250	0.750	0.450	0.300	0.050

Neste caso, o histograma é dado abaixo. Note que não há mais intervalos de frequência 0.

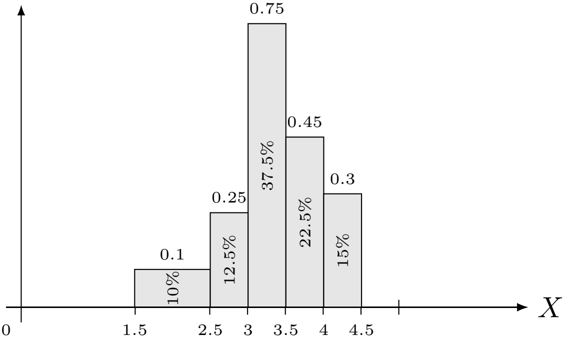


Exemplo 3 - cont.

Relembrando, quando os dois primeiros intervalos foram unidos a distribuição de densidades é dada abaixo.

X	$[1.5, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Dens.	0.100	0.250	0.750	0.450	0.300	0.050

Neste caso, o histograma é dado abaixo. Note que não há mais intervalos de frequência 0.

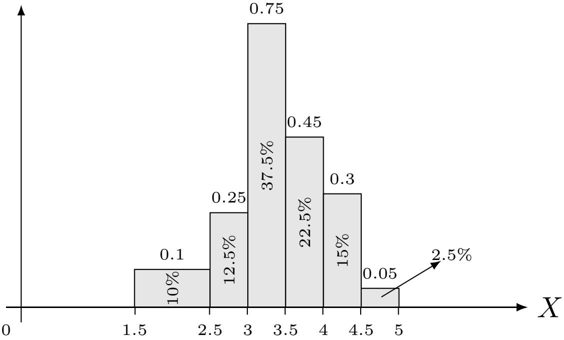


Exemplo 3 - cont.

Relembrando, quando os dois primeiros intervalos foram unidos a distribuição de densidades é dada abaixo.

X	$[1.5, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Dens.	0.100	0.250	0.750	0.450	0.300	0.050

Neste caso, o histograma é dado abaixo. Note que não há mais intervalos de frequência 0.



Estrutura

Medidas Resumo

Principal objetivo: expressar alguma característica dos elementos da amostra.

Medidas Resumo

Principal objetivo: expressar alguma característica dos elementos da amostra.

Nos principais casos são de interesse as seguintes características dos dados:

- centro; e
- dispersão.

Medidas Resumo

Principal objetivo: expressar alguma característica dos elementos da amostra.

Nos principais casos são de interesse as seguintes características dos dados:

- centro; e
- dispersão.

Vantagem: um único número para:

- resumir os dados; e
- representar características importantes.

Medidas Resumo

Principal objetivo: expressar alguma característica dos elementos da amostra.

Nos principais casos são de interesse as seguintes características dos dados:

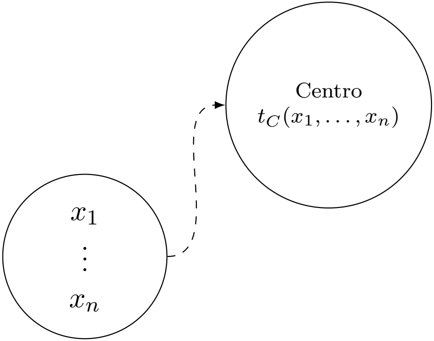
- centro; e
- dispersão.

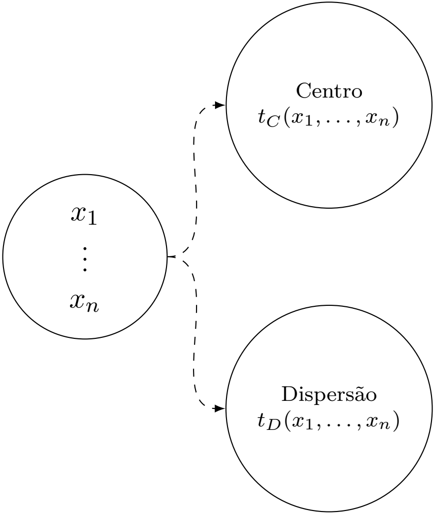
Vantagem: um único número para:

- resumir os dados; e
- representar características importantes.

Principal desvantagem de medidas resumo: **perda de informação!**

$$x_1$$
$$\vdots$$
$$x_n$$


x_1
$\vdots$
x_n
Centro
$t_C(x_1, \ldots, x_n)$


Centro
$t_C(x_1, \ldots, x_n)$
x_1
$\vdots$
x_n
Dispersão
$t_D(x_1, \ldots, x_n)$

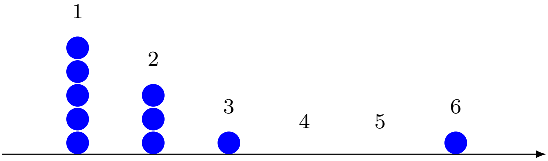
Estrutura

Medidas de tendência central

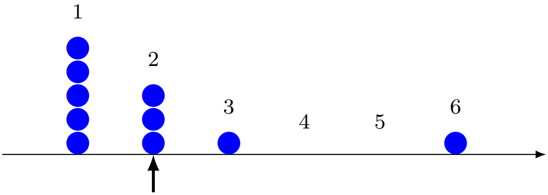
- Buscam representar um valor "típico" dos dados em questão;
- O que seria valor "típico"?
- Principais:
  - ponto de equilíbrio dos dados;
  - ponto central dos dados;
  - ponto de alta frequência;
- Diferentes medidas de tendência central:
  - média;
  - mediana;
  - moda.

Estrutura

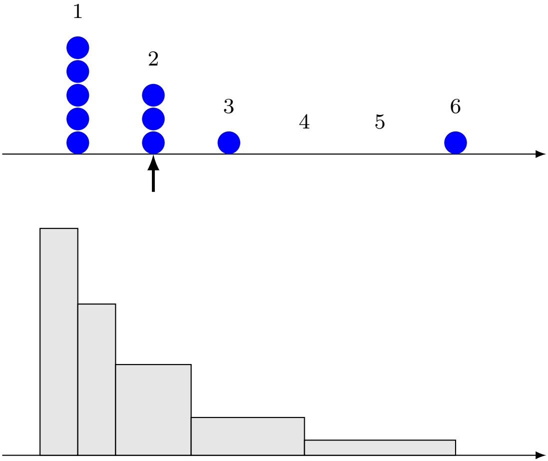
Média



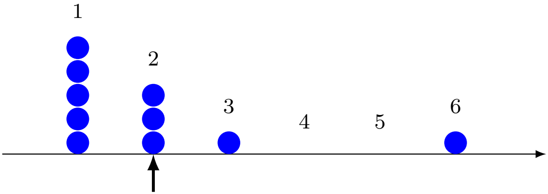
Média

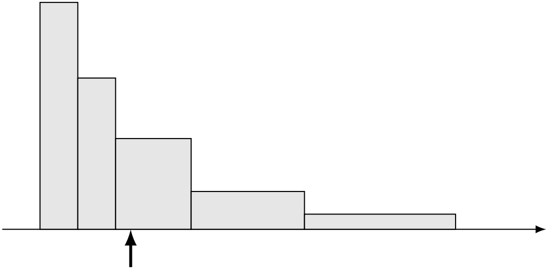


Média



Média





Cálculo da média

Definição

Dadas as observações $x_1, \ldots, x_n$, a média é definida por

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

Suponha que os valores são apresentados na forma da seguinte distribuição de frequências:

X	x_1^*	x_2^*	$\ldots$	x_k^*	Total
Freq.	n_1	n_2	$\ldots$	n_k	$\sum_{i=1}^{k} n_i = n$

A média pode ser calculada como

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_j x_j^* = \sum_{i=1}^{k} f_j x_j^*,$$

onde $f_j = \frac{n_j}{n}$ é a frequência relativa do j-ésimo valor.

Suponha que as observações são dadas na forma da seguinte distribuição de frequências intervalar:

X	$[a_0, a_1]$	$(a_1, a_2]$	$\ldots$	$(a_{k-1}, a_k]$	Total
Freq.	n_1	n_2	$\ldots$	n_k	$\sum_{i=1}^{k} n_i = n$

Cálculo da média

Suponha que as observações são dadas na forma da seguinte distribuição de frequências intervalar:

X	$[a_0, a_1]$	$(a_1, a_2]$	$\ldots$	$(a_{k-1}, a_k]$	Total
Freq.	n_1	n_2	$\ldots$	n_k	$\sum_{i=1}^{k} n_i = n$

Problema: desconhecemos os valores observados.

Cálculo da média

Suponha que as observações são dadas na forma da seguinte distribuição de frequências intervalar:

X	$[a_0, a_1]$	$(a_1, a_2]$	$\dots$	$(a_{k-1}, a_k]$	Total
Freq.	n_1	n_2	$\dots$	n_k	$\sum_{i=1}^{k} n_i = n$

Problema: desconhecemos os valores observados.
Para $i = 1, \dots, k$, devemos **aproximar** os valores do intervalo $(a_{i-1}, a_i]$ pelo ponto médio $x_i^m = \frac{a_{i-i}+a_i}{2}$ desse intervalo. Agora, a média é aproximada por

$$\bar{x} \approx \frac{1}{n} \sum_{i=1}^{k} n_j x_j^m = \sum_{i=1}^{k} f_j x_j^m,$$

onde $f_j = \frac{n_j}{n}$ é a frequência relativa do j-ésimo intervalo.

Características

Comentário

A média possui as seguintes características:

- *tem propriedades "boas";*
- *é influenciada por valores atípicos;*
- *não recomendada em dados assimétricos;*
- *só é calculada em variáveis quantitativas;*
- *no caso intervalar* $\downarrow k \Rightarrow$ *aproximação ruim para a média.*

Estrutura

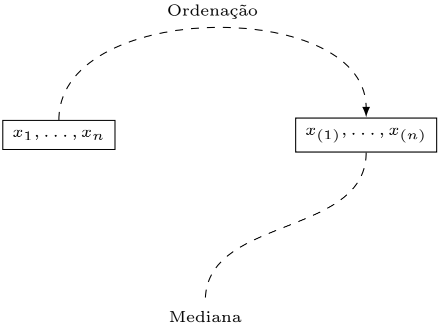
Mediana

$\boxed{x_1, \ldots, x_n}$

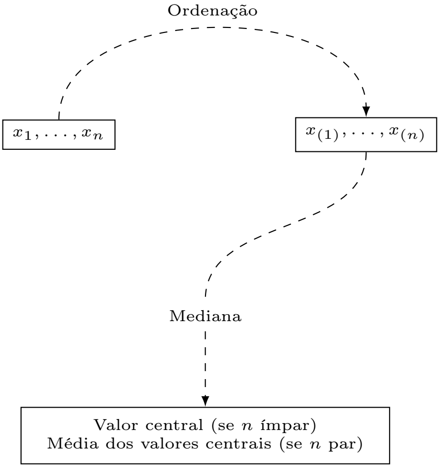
Mediana

$$\boxed{x_1, \ldots, x_n} \xrightarrow{\text{Ordenação}} \boxed{x_{(1)}, \ldots, x_{(n)}}$$

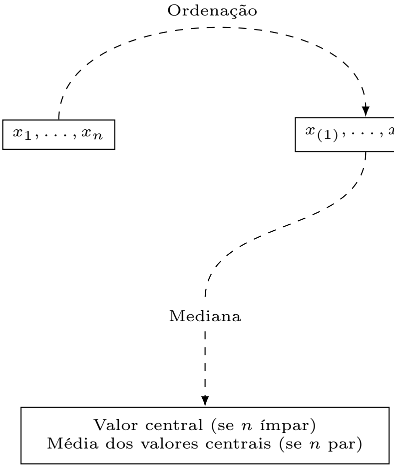
Mediana



Mediana


Ordenação
$x_1, \ldots, x_n$
$x_{(1)}, \ldots, x_{(n)}$
Mediana
Valor central (se n ímpar)
Média dos valores centrais (se n par)

Mediana



A Mediana divide os dados de forma que 50% deles são menores ou iguais e 50% deles são maiores ou iguais que a mediana.

Definição

Dadas as observações $x_1, \ldots, x_n$*, sejam* $x_{(1)}, \ldots, x_{(n)}$ *esses valores ordenados, isto é,* $x_{(1)} \leq \cdots \leq x_{(n)}$*. A mediana é definida por*

$$Med = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{se } n \text{ é ímpar;} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{se } n \text{ é par.} \end{cases}$$

Problema: No caso intervalar, devemos aproximar o valor da mediana.

Definição

Dadas as observações $x_1, \ldots, x_n$, sejam $x_{(1)}, \ldots, x_{(n)}$ esses valores ordenados, isto é, $x_{(1)} \leq \cdots \leq x_{(n)}$. A mediana é definida por

$$Med = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{se } n \text{ é ímpar;} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{se } n \text{ é par.} \end{cases}$$

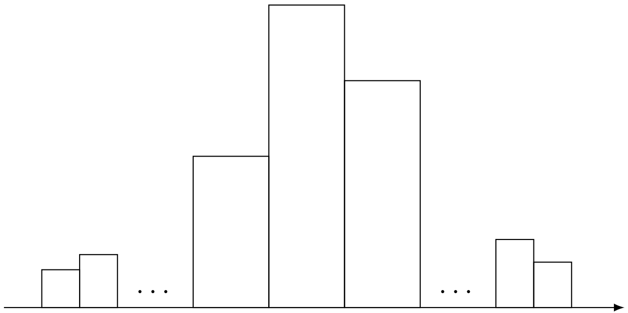
Problema: No caso intervalar, devemos aproximar o valor da mediana. **Como?** Suponha a distribuição de frequências relativas abaixo:

X	$[a_0, a_1]$	$(a_1, a_2]$	$\cdots$	$(a_{k-1}, a_k]$	Total
Freq.	f_1	f_2	$\cdots$	f_k	$\sum_{i=1}^{k} f_i = 1$
Freq. acum.	$F_1 = f_1$	$F_2 = F_1 + f_2$	$\cdots$	$F_k = F_{k-1} + f_k = 1$	$\times$

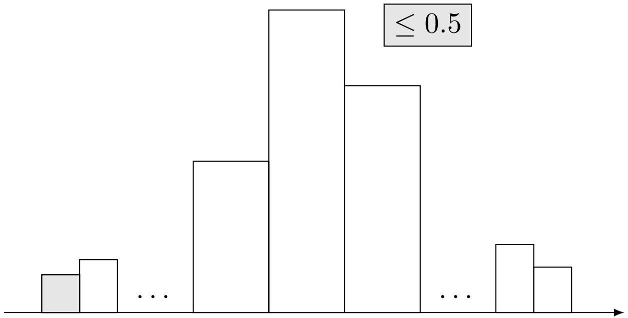
Denominamos $F_1 = f_1$ e $F_i = F_{i-1} + f_i$, $i = 2, \ldots, k$, de frequências acumuladas. Seja I_l o primeiro intervalo tal que $F_l \geq 0.5$. Isto é, l é o menor valor em $\{1, \ldots, k\}$ tal que $F_l \geq 0.5$. Devemos aproximar a mediana por

$$Med \approx a_{l-1} + \frac{0.5 - F_{l-1}}{d_l}.$$

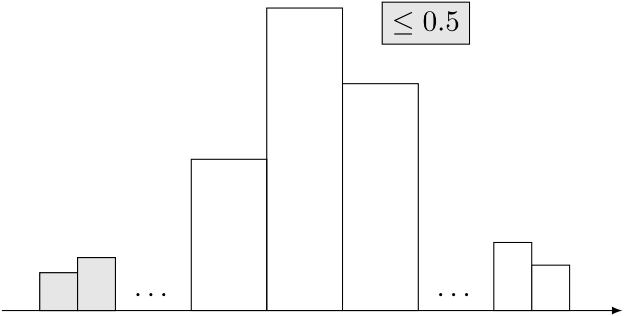
Esquematicamente:



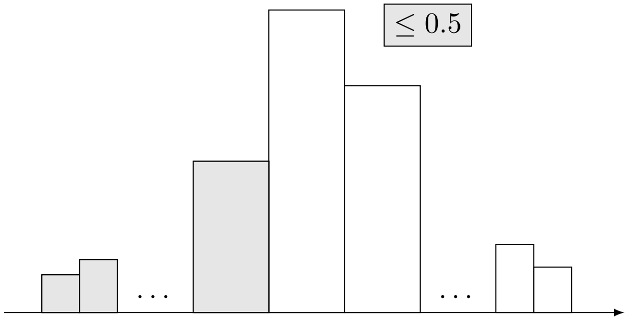
Esquematicamente:



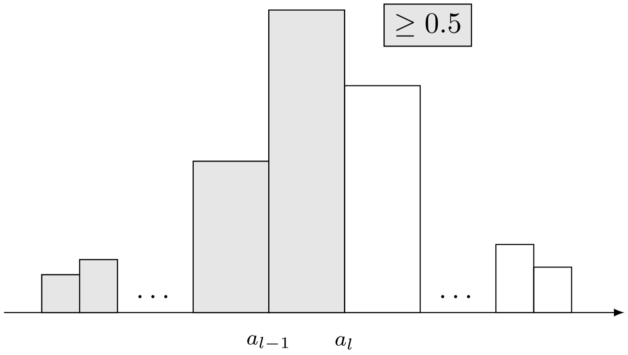
Esquematicamente:



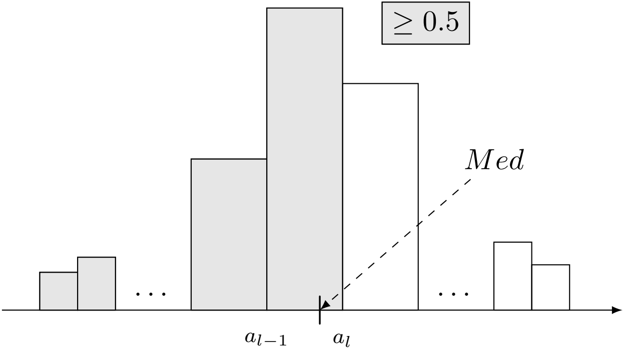
Esquematicamente:



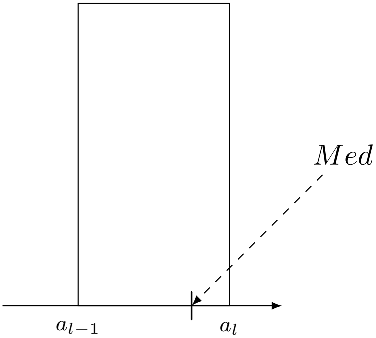
Esquematicamente:



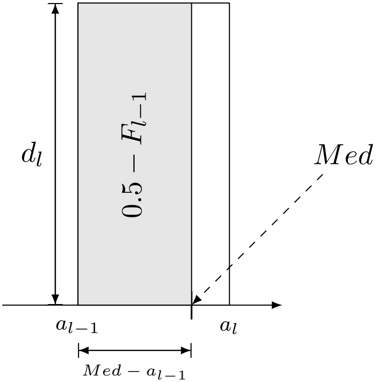
Esquematicamente:



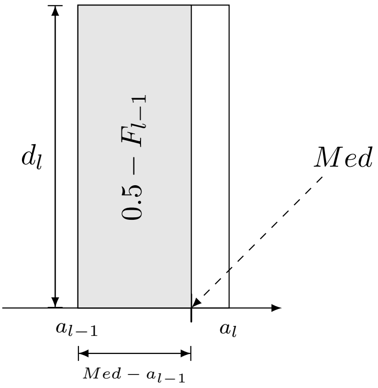
Olhando o retângulo de base $(a_{l-1}, a_l]$, temos o seguinte gráfico.



Olhando o retângulo de base $(a_{l-1}, a_l]$, temos o seguinte gráfico.



Olhando o retângulo de base $(a_{l-1}, a_l]$, temos o seguinte gráfico.



Assim, temos que

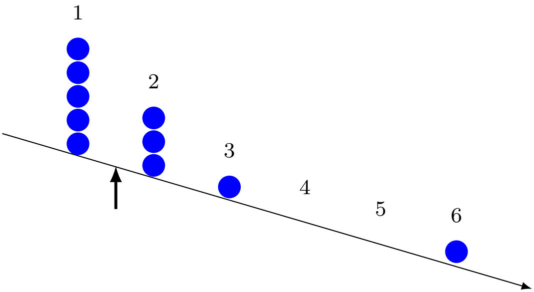
$$0.5 - F_{l-1} = d_l \times (Med - a_{l-1}) \Rightarrow Med \approx a_{l-1} + \frac{0.5 - F_{l-1}}{d_l}.$$

Comentário

- *sem interpretação física;*
- *depende apenas da posição e não do valor;*
- *menos influência de dados atípicos;*
- *pode ser usada em variáveis qualit. ordinais.*

Comentário

- *sem interpretação física;*
- *depende apenas da posição e não do valor;*
- *menos influência de dados atípicos;*
- *pode ser usada em variáveis qualit. ordinais.*



Estrutura

Moda

Definição

Dadas as observações $x_1, \ldots, x_n$*, se* $x_1^*, \ldots, x_k^*$ *denotarem os* k *valores diferentes, a moda é dada pelo valor com maior frequência.*

Comentário

Um conjunto de dados pode ser amodal, unimodal, bimodal, $\ldots$.

Moda

Definição

Dadas as observações $x_1, \ldots, x_n$*, se* $x_1^*, \ldots, x_k^*$ *denotarem os* k *valores diferentes, a moda é dada pelo valor com maior frequência.*

Comentário

Um conjunto de dados pode ser amodal, unimodal, bimodal, $\ldots$.

Problema: em variáveis contínuas, frequentemente, observamos poucos valores repetidos. Assim, na maioria dos casos, esse tipo de dado é amodal.

Moda

Definição

Dadas as observações $x_1, \ldots, x_n$, *se* $x_1^*, \ldots, x_k^*$ *denotarem os* k *valores diferentes, a moda é dada pelo valor com maior frequência.*

Comentário

Um conjunto de dados pode ser amodal, unimodal, bimodal, ...

Problema: em variáveis contínuas, frequentemente, observamos poucos valores repetidos. Assim, na maioria dos casos, esse tipo de dado é amodal.

Alternativa: utilizar a classe com maior densidade, ou **classe modal**.

Moda

Definição

Dadas as observações $x_1, \ldots, x_n$, *se* $x_1^*, \ldots, x_k^*$ *denotarem os* k *valores diferentes, a moda é dada pelo valor com maior frequência.*

Comentário

Um conjunto de dados pode ser amodal, unimodal, bimodal, $\ldots$.

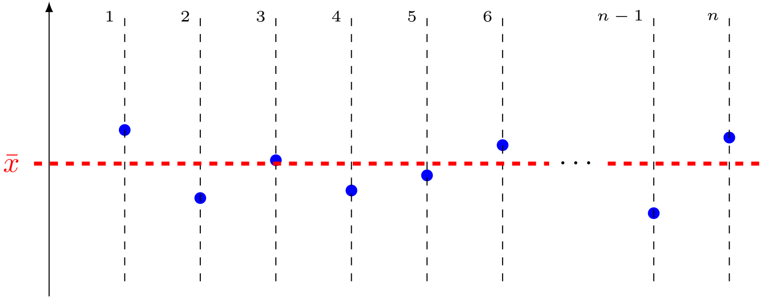
Problema: em variáveis contínuas, frequentemente, observamos poucos valores repetidos. Assim, na maioria dos casos, esse tipo de dado é amodal.
Alternativa: utilizar a classe com maior densidade, ou **classe modal**.
Aproximação: assim, a moda pode ser aproximada pelo valor médio da classe modal.

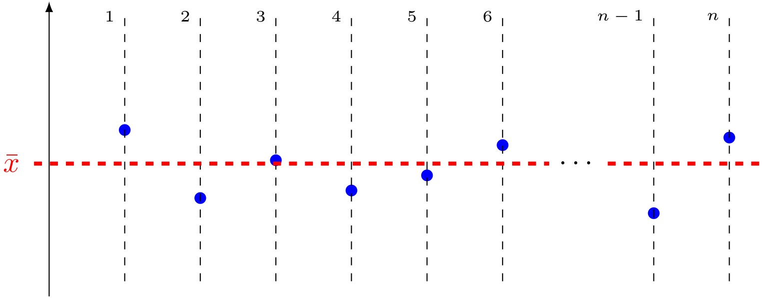
Comentário

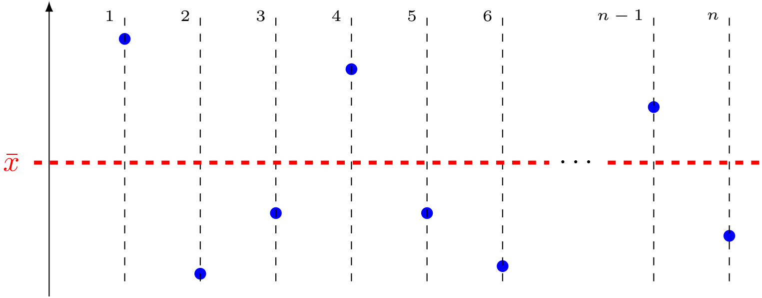
A moda de um conjunto de dados:

- *representa o(s) valor(es) mais provável(eis);*
- *é muito indicada em dados multimodais;*
- *não é afetada por dados atípicos;*
- *pode ser usada em variáveis qualitativas.*

Estrutura


1
2
3
4
5
6
$n-1$
n
$\bar{x}$


1
2
3
4
5
6
n − 1
n
x̄
· · ·


1
2
3
4
5
6
n − 1
n
x̄
· · ·

Medidas de dispersão

Principal objetivo:

- uma medida para representar quão disperso os dados estão.

Medidas de dispersão

Principal objetivo:

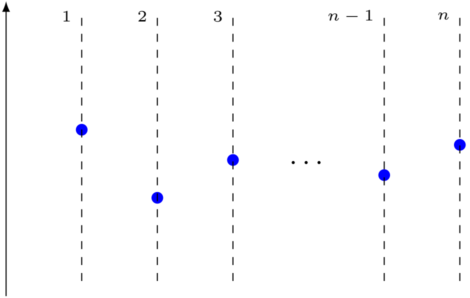
- uma medida para representar quão disperso os dados estão.

Quatro medidas apresentadas:

- amplitude amostral;
- variância;
- desvio-padrão;
- coeficiente de variação.

Estrutura

Amplitude



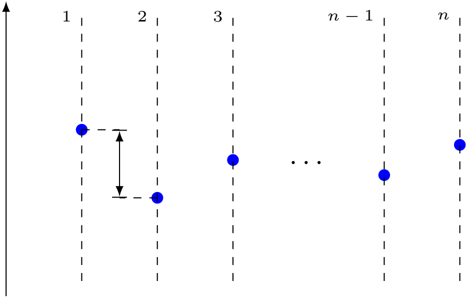
Elementos da amostra: $x_1, \ldots, x_n$.
Elementos ordenados: $x_{(1)}, \ldots, x_{(n)}$.

Definição

A amplitude da amostra é definida por

$$Amp = x_{(n)} - x_{(1)}.$$

Amplitude



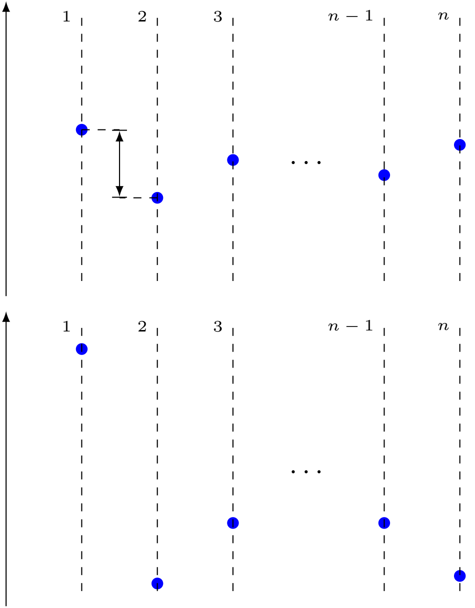
Elementos da amostra: $x_1, \ldots, x_n$.
Elementos ordenados: $x_{(1)}, \ldots, x_{(n)}$.

Definição

A amplitude da amostra é definida por

$$Amp = x_{(n)} - x_{(1)}.$$

Amplitude



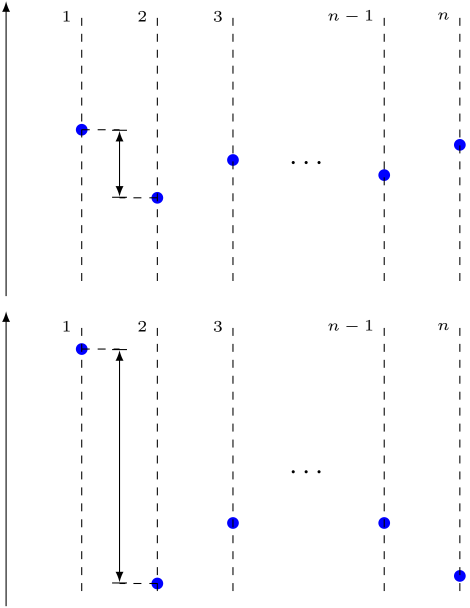
Elementos da amostra: $x_1, \ldots, x_n$.
Elementos ordenados: $x_{(1)}, \ldots, x_{(n)}$.

Definição

A amplitude da amostra é definida por

$$Amp = x_{(n)} - x_{(1)}.$$

Amplitude



Elementos da amostra: $x_1, \ldots, x_n$.

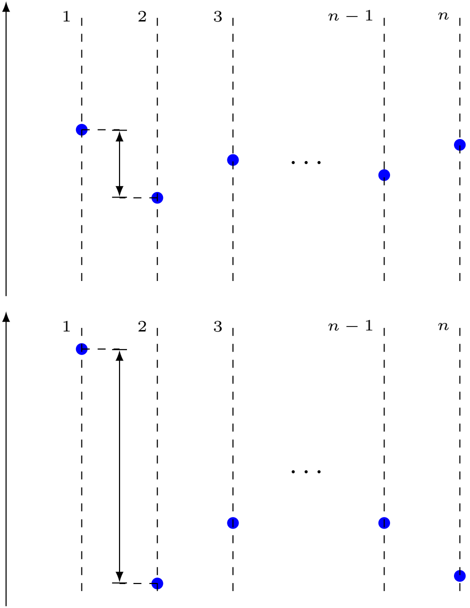
Elementos ordenados: $x_{(1)}, \ldots, x_{(n)}$.

Definição

A amplitude da amostra é definida por

$$Amp = x_{(n)} - x_{(1)}.$$

Amplitude



Elementos da amostra: $x_1, \ldots, x_n$.
Elementos ordenados: $x_{(1)}, \ldots, x_{(n)}$.

Definição

A amplitude da amostra é definida por

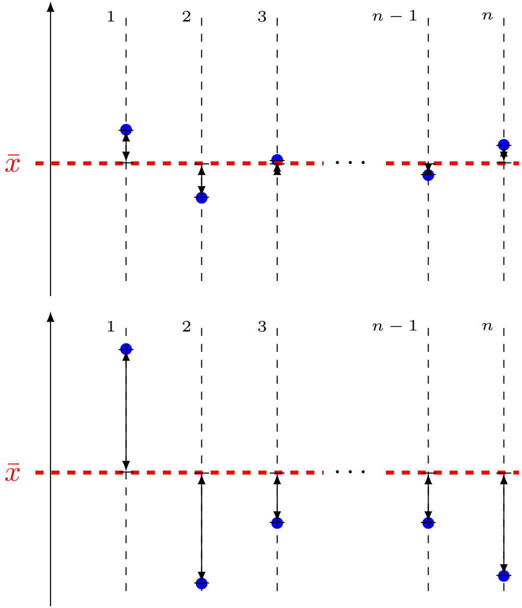
$$Amp = x_{(n)} - x_{(1)}.$$

Comentários:

- cálculo rápido;
- fácil interpretação;
- alto impacto de dados atípicos.

Estrutura

Variância



Idéia: Todos dados expressarem a dispersão.

Definição

Variância populacional:

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

Variância amostral:

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

Comentário

Alguns comentários:

- *mais robusta a dados atípicos;*

Comentário

Alguns comentários:

- *mais robusta a dados atípicos;*
- *se os dados, por exemplo, são expressos em cm a variância é em* cm^2*;*

Comentário

Alguns comentários:

- *mais robusta a dados atípicos;*
- *se os dados, por exemplo, são expressos em cm a variância é em* cm^2*;*
- *é possível mostrar que* $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$*;*

Comentário

Alguns comentários:

- *mais robusta a dados atípicos;*
- *se os dados, por exemplo, são expressos em cm a variância é em cm^2;*
- *é possível mostrar que* $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$;
- *portanto* $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$;

Comentário

Alguns comentários:

- *mais robusta a dados atípicos;*
- *se os dados, por exemplo, são expressos em cm a variância é em* cm^2*;*
- *é possível mostrar que* $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$*;*
- *portanto* $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$*;*
- *e* $s^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\bar{x}^2$*;*

Comentário

Alguns comentários:

- *mais robusta a dados atípicos;*
- *se os dados, por exemplo, são expressos em cm a variância é em cm*^{2}*;*
- *é possível mostrar que* $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$*;*
- *portanto* $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$*;*
- *e* $s^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\bar{x}^2$*;*
- *para dados apresentados em frequências intervalares, temos que*

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i^m)^2 - \bar{x}^2$$

e que

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

onde x_i^m *é o ponto médio do i-ésimo intervalo.*

Estrutura

Desvio-padrão

Para remediar o fato de a variância ser expressa na unidade de medida da variável ao quadrado, podemos calcular o **desvio-padrão**.

Definição

O desvio-padrão populacional é dado por

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

O desvio-padrão amostral é dado por

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

A grande vantagem do desvio-padrão é que ele é expresso na mesma unidade de medida dos dados.

Estrutura

Coeficiente de variação

Problema: As medidas de dispersão apresentadas são influenciadas pela "grandeza" da variável estudada.

Coeficiente de variação

Problema: As medidas de dispersão apresentadas são influenciadas pela "grandeza" da variável estudada.
É comum então representar o desvio-padrão como percentual da média.
Denomina-se essa medida de coeficiente de variação.

Coeficiente de variação

Problema: As medidas de dispersão apresentadas são influenciadas pela "grandeza" da variável estudada.
É comum então representar o desvio-padrão como percentual da média.
Denomina-se essa medida de coeficiente de variação.

Definição

O coeficiente de variação populacional é dado por

$$CV = \frac{\sigma}{\mu},$$

onde μ é a média populacional. O coeficiente de variação amostral é dado por

$$cv = \frac{s}{\bar{x}},$$

onde $\bar{x}$ é a média amostral.

Coeficiente de variação

Problema: As medidas de dispersão apresentadas são influenciadas pela "grandeza" da variável estudada.
É comum então representar o desvio-padrão como percentual da média. Denomina-se essa medida de coeficiente de variação.

Definição

O coeficiente de variação populacional é dado por

$$CV = \frac{\sigma}{\mu},$$

onde μ é a média populacional. O coeficiente de variação amostral é dado por

$$cv = \frac{s}{\bar{x}},$$

onde $\bar{x}$ é a média amostral.

Vantagem: Duas populações com médias muito diferentes podem ter suas disperções comparadas através do coeficiente de variação.

Exemplo 2 - cont.

Retornando ao Exemplo 2. Temos que a distribuição de frequências é dada por

N	0	1	2	3	Total
Freq.	2	11	12	5	30

Assim, temos que

$$\bar{x} = \frac{2 \cdot 0 + 11 \cdot 1 + 12 \cdot 2 + 5 \cdot 3}{30} \approx 1.67,$$

$$Med = \frac{x_{(15)} + x_{(16)}}{2} = \frac{2+2}{2} = 2,$$

e

$$Mod = 2.$$

Retornando ao Exemplo 2. Temos que a distribuição de frequências é dada por

N	0	1	2	3	Total
Freq.	2	11	12	5	30

Assim, temos que

$$Amp = x_{(30)} - x_{(1)} = 3 - 0 = 3,$$

$$s^2 \approx \frac{11 \cdot 1^2 + 12 \cdot 2^2 + 5 \cdot 3^2}{29} - \frac{30 \cdot 1.67^2}{29} = \frac{104}{29} - \frac{83.667}{29} \approx 0.701,$$

$$s \approx \sqrt{0.701} \approx 0.837$$

e

$$cv = \frac{s}{\bar{x}} = \frac{0.837}{1.67} \approx 0.501 = 50.1\%.$$

Exemplo 3 - cont.

Retornando ao Exemplo 3. Temos que

X	$[1.5, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Pt. Méd.	2.00	2.75	3.25	3.75	4.25	4.75
Freq.	4	5	15	9	6	1
Freq.	0.100	0.125	0.375	0.225	0.150	0.025
F_i	0.100	0.225	0.600	0.825	0.975	1.000
Dens.	0.100	0.250	0.750	0.450	0.300	0.050

Assim, temos que

$$\bar{x} \approx 2 \cdot 0.1 + 2.75 \cdot 0.125 + \cdots + 4.75 \cdot 0.025 \approx 3.363$$

$$Med = a_{l-1} + \frac{0.5 - F_{l-1}}{d_l} = 3 + \frac{0.5 - 0.225}{0.750} \approx 3.367,$$

e

$$Classe\ Modal = (3.0, 3.5] \Rightarrow Mod \approx \frac{3 + 3.5}{2} = 3.25.$$

Exemplo 3 - cont.

Retornando ao Exemplo 3. Temos que

X	$[1.5, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$
Pt. Méd.	2.00	2.75	3.25	3.75	4.25	4.75
Freq.	4	5	15	9	6	1
Freq.	0.100	0.125	0.375	0.225	0.150	0.025
F_i	0.100	0.225	0.600	0.825	0.975	1.000
Dens.	0.100	0.250	0.750	0.450	0.300	0.050

Assim, temos que

$$Amp = x_{(40)} - x_{(1)} = 4.7 - 1.7 = 3$$

$$s^2 \approx \frac{4 \cdot 2^2 + 5 \cdot 2.75^2 + \cdots + 1 \cdot 4.75^2}{39} - \frac{40 \cdot 3.363^2}{39} = \frac{469.75}{39} - \frac{452.391}{39} \approx 0.445,$$

$$s \approx \sqrt{0.445} \approx 0.667$$

e

$$cv = \frac{s}{\bar{x}} = \frac{0.667}{3.363} \approx 0.198 = 19.8\%.$$

FIM!