



Ministério da Educação
Universidade Federal do Piauí
Campus Senador Helvídio Nunes de Barros
Coordenação de Matemática

Estatística Descritiva

Professor: Josimar Vasconcelos

Contato: josimar@ufpi.edu.br

www.ufpi.br/prof_josimar

prof-josimar.blogspot.com.br

22 de outubro de 2012





Distribuições de Frequências





👉 Distribuições de frequências

- A distribuição de frequência é a série estatística em que os dados são agrupados em classes com o objetivo de facilitar ao analista o seu estudo.





➤ Dados brutos

São os dados apresentados desordenadamente, da forma como foram coletados. Por exemplo,

Tabela: Informações de questionário estudantil

ID	TURMA	Sexo	IDADE	...	OPINIÃO DA TV
1	A	F	17	...	R
2	A	F	18	...	R
⋮	⋮	⋮	⋮	⋮	⋮
50	B	M	18	...	B





➡ Rol

São dados apresentados em ordem crescente. Por exemplo,

44,00	47,00	47,00	47,40	48,00	49,00	49,00	49,20	50,00
50,00	51,60	52,00	52,00	52,50	54,00	54,50	54,50	55,00
55,00	55,00	55,00
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
...	95,00





👉 Frequência Relativa simples

É o número de vezes que o elemento aparece no conjunto de dados, ou o número de elementos pertencentes a uma classe ou categoria.

Note que,

$$\sum_{i=1}^k f_i = f_1 + f_2 + f_3 + \dots + f_k = n,$$

em que n é o número de observações do conjunto de dados e k é o número de valores ou níveis que X assume.





➤ Frequência Relativa simples

Tomando como exemplo a variável idade, temos:

$$\sum_{i=1}^8 f_i = 9 + 22 + 7 + \dots + 2 = 50,$$

onde 50 é o número de observações da variável idade.





👉 Distribuição de frequências por valores

É o arranjo dos valores e suas respectivas frequências, ou seja, é uma tabela onde os valores da variável aparecem individualmente.

Por exemplo,

Tabela: Local do título da tabela de distribuição de frequência.

X_i	f_i	f_r	$f_r\%$
X_1	f_1	f_1/n	$f_{r1} \cdot 100$
\vdots	\vdots	\vdots	\vdots
X_k	f_k	f_k/n	$f_{rk} \cdot 100$
Σ	n	1,00	100

Obs.: utiliza-se essa tab. quando temos as variáveis qualit. nominal, ordinal e quant. discreta.





👉 Exemplo

Tabela: Distribuição de frequência para variável sexo.

Sexo	f	f_r	$f_r\%$
M	37	0,74	74,00
F	13	0,26	26,00
Σ	50	1,00	100,00





Note que, para variáveis que possuem ordenação natural (qualitativas ordinais e quantitativas discretas), faz sentido incluirmos colunas contendo as frequências acumuladas para baixo e para cima. Assim, temos o seguinte exemplo:



👉 Exemplo

Tabela: Distribuição de frequência para variável tolerância ao cigarro.

Toler	f	f_r	$f_r\%$	$F \downarrow$	$F \downarrow\%$	$F \uparrow$	$F \uparrow\%$
I	10	0,20	20,00	10	20,00	50	100,00
P	21	0,42	42,00	31	62,00	40	80,00
M	19	0,38	38,00	50	100,00	19	38,00
Σ	50	1,00	100,00	—	—	—	—





E para variável quantitativa
discreta ou contínua?





E para variável quantitativa discreta ou contínua?



Tabela: Dist. de freq. por classes para a variável peso.

Pesos	f	f_r	$f_r\%$	$F \downarrow$	$F \downarrow\%$	$F \uparrow$	$F \uparrow\%$
44, 00 – 51, 21	10	0,20	20,00	10	20,00	50	100,00
51, 21 – 58, 42	18	0,36	36,00	28	56,00	40	80,00
58, 42 – 65, 63	8	0,16	16,00	36	72,00	22	44,00
65, 63 – 72, 84	5	0,10	10,00	41	82,00	14	28,00
72, 84 – 80, 05	3	0,06	6,00	44	88,00	9	18,00
80, 05 – 87, 26	4	0,08	8,00	48	96,00	6	12,00
87, 26 – 94, 47	1	0,02	2,00	49	98,00	2	4,00
94, 47 – 101, 68	1	0,02	2,00	50	100,00	1	2,00
Σ	50	1,00	100,00	–	–	–	–

Fonte: Marcos Magalhães (2004).





- **Amplitude total (At):** é a diferença entre o maior e o menor valor observado da variável em estudo. Ou seja,

$$AT = X_{\max} - X_{\min}.$$





- **Amplitude total (At):** é a diferença entre o maior e o menor valor observado da variável em estudo. Ou seja,

$$AT = X_{\max} - X_{\min}.$$

- Para variável peso temos, $AT = 51$.





- **Número de classes (c):** corresponde à quantidade de classes, nas quais serão agrupados os elementos do rol.
 - a) $c = 5$, para $n \leq 25$ e $c \cong \sqrt{n}$, para $n > 25$. Em que, n é o número de observações.
 - b) Regra de *Sturges*: $c \cong 1 + 3,3 \log_{10} n$, onde n é o número de observações.





- Podemos utilizar os dados anteriores como exemplo.

a) $n = 50 > 25$, então, $c \cong \sqrt{50} \cong 7,07$

b) $c \cong 1 + 3,3 \log_{10} 50 \cong 6,61$





- **Amplitude das classes (h):** distância entre o mínimo e o máximo das classes. Devemos, em geral, construir classes de mesma amplitude, a qual pode ser obtida através da expressão:

$$h \cong \frac{AT}{c}.$$





- **Amplitude das classes (h):** distância entre o mínimo e o máximo das classes. Devemos, em geral, construir classes de mesma amplitude, a qual pode ser obtida através da expressão:

$$h \cong \frac{AT}{c}.$$

- Exemplo anterior, $h \cong 7,21$.





- **Amplitude das classes (h):** distância entre o mínimo e o máximo das classes. Devemos, em geral, construir classes de mesma amplitude, a qual pode ser obtida através da expressão:

$$h \cong \frac{AT}{c}.$$

- Exemplo anterior, $h \cong 7,21$.

Em algumas situações o número de classes pode não corresponder ao número calculado.





• Limite de classes - LI e LS

1. LI $|$ LS: considera valores entre LI e LS, incluindo LI e LS.
2. LI $|$ LS: considera valores entre LI e LS, incluindo LI e excluindo LS.
3. LI $-$ LS: considera valores entre LI e LS, excluindo LI e incluindo LS.
4. LI $-$ LS: não determina claramente se LI e LS devem ser considerados ou não.





Notação utilizada para tabela

- f ; frequência absoluta (número de ocorrências de cada classe);
- f_r ; frequência relativa;
- $f_r\%$; frequência relativa percentual;
- $F \downarrow$; frequência absoluta acumulada “abaixo de”;
- $F \downarrow\%$; frequência percentual acumulada “abaixo de”;
- $F \uparrow$; frequência absoluta acumulada “acima de”;
- $F \uparrow\%$; frequência percentual acumulada “acima de”.





Roteiro para a elaboração de uma dist. de frequências por classes

- Construção do Rol;
- Determinação da Amplitude Total (AT);
- Determinação do Número de Classes (c);
- Determinação da Amplitude das Classes (h);
- Determinação dos limites das classes (LI e LS);
- Construção da tabela de frequências ou a dist. de frequência.





Representação Tabular





Representação Tabular

Consiste em dispor os dados em linhas e colunas, distribuídas de modo ordenado, segundo algumas regras práticas.

- As tabelas devem apresentar:





Representação Tabular

Consiste em dispor os dados em linhas e colunas, distribuídas de modo ordenado, segundo algumas regras práticas.

- As tabelas devem apresentar:
 - Título \Rightarrow o quê? onde? quando?;





Representação Tabular

Consiste em dispor os dados em linhas e colunas, distribuídas de modo ordenado, segundo algumas regras práticas.

- As tabelas devem apresentar:
 - Título \Rightarrow o quê? onde? quando?;
 - Cabeçalho \Rightarrow especifica o conteúdo das colunas;





Representação Tabular

Consiste em dispor os dados em linhas e colunas, distribuídas de modo ordenado, segundo algumas regras práticas.

- As tabelas devem apresentar:
 - Título \Rightarrow o quê? onde? quando?;
 - Cabeçalho \Rightarrow especifica o conteúdo das colunas;
 - Coluna indicadora \Rightarrow especifica o conteúdo das linhas;





Representação Tabular

Consiste em dispor os dados em linhas e colunas, distribuídas de modo ordenado, segundo algumas regras práticas.

- As tabelas devem apresentar:
 - Título \Rightarrow o quê? onde? quando?;
 - Cabeçalho \Rightarrow especifica o conteúdo das colunas;
 - Coluna indicadora \Rightarrow especifica o conteúdo das linhas;
 - Corpo \Rightarrow caselas onde são registrados os dados;





Representação Tabular

Consiste em dispor os dados em linhas e colunas, distribuídas de modo ordenado, segundo algumas regras práticas.

- As tabelas devem apresentar:
 - Título \Rightarrow o quê? onde? quando?;
 - Cabeçalho \Rightarrow especifica o conteúdo das colunas;
 - Coluna indicadora \Rightarrow especifica o conteúdo das linhas;
 - Corpo \Rightarrow caselas onde são registrados os dados;
 - Rodapé \Rightarrow notas e identificação da fonte dos dados.



Exemplo

PRODUÇÃO DE CAFÉ ← TÍTULO
BRASIL — 1978-82

ANOS	PRODUÇÃO (1.000 t)
1978	2.535
1979	2.666
1980	2.122
1981	3.750
1982	2.007

RODAPÉ → FONTE: IBGE.

Diagram Labels:

- TÍTULO:** PRODUÇÃO DE CAFÉ
- CABEÇALHO:** ANOS, PRODUÇÃO (1.000 t)
- COLUNA INDICADORA:** ANOS
- COLUNA NUMÉRICA:** PRODUÇÃO (1.000 t)
- CASA OU CÉLULA:** 2.535
- LINHAS:** 1979, 1980, 1981, 1982
- CORPO:** 1978, 1979, 1980, 1981, 1982
- RODAPÉ:** FONTE: IBGE.





Representação Gráfica





Representação Gráfica

Tem por finalidade uma melhor visualização do conteúdo das tabelas, expondo, sempre que possível, as mesmas informações nelas contidas.

- Os tipos mais usados de gráficos são:



Representação Gráfica

Tem por finalidade uma melhor visualização do conteúdo das tabelas, expondo, sempre que possível, as mesmas informações nelas contidas.

- Os tipos mais usados de gráficos são:
 - de linhas e de superfície simples e em faixa;



Representação Gráfica

Tem por finalidade uma melhor visualização do conteúdo das tabelas, expondo, sempre que possível, as mesmas informações nelas contidas.

- Os tipos mais usados de gráficos são:
 - de linhas e de superfície simples e em faixa;
 - de colunas ou barras simples, remontadas ou superpostas;





Representação Gráfica

Tem por finalidade uma melhor visualização do conteúdo das tabelas, expondo, sempre que possível, as mesmas informações nelas contidas.

- Os tipos mais usados de gráficos são:
 - de linhas e de superfície simples e em faixa;
 - de colunas ou barras simples, remontadas ou superpostas;
 - de setores em círculo (pizza);





Representação Gráfica

Tem por finalidade uma melhor visualização do conteúdo das tabelas, expondo, sempre que possível, as mesmas informações nelas contidas.

- Os tipos mais usados de gráficos são:
 - de linhas e de superfície simples e em faixa;
 - de colunas ou barras simples, remontadas ou superpostas;
 - de setores em círculo (pizza);
 - Box plot;





Representação Gráfica

Tem por finalidade uma melhor visualização do conteúdo das tabelas, expondo, sempre que possível, as mesmas informações nelas contidas.

- Os tipos mais usados de gráficos são:
 - de linhas e de superfície simples e em faixa;
 - de colunas ou barras simples, remontadas ou superpostas;
 - de setores em círculo (pizza);
 - Box plot;
 - Histogramas;



Gráfico de Linhas

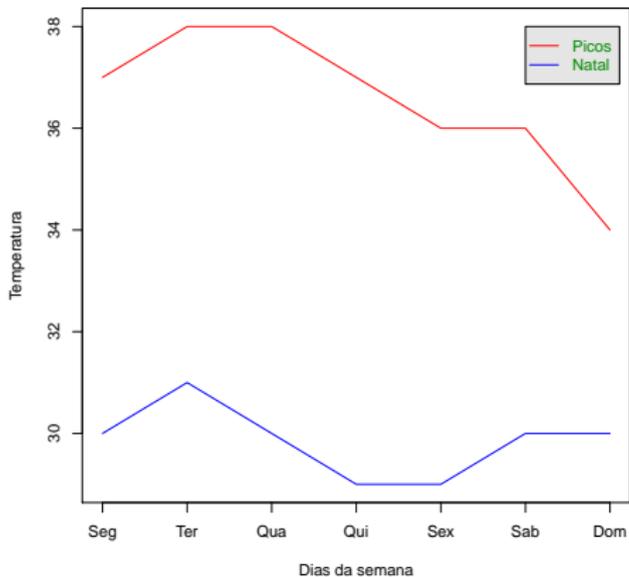
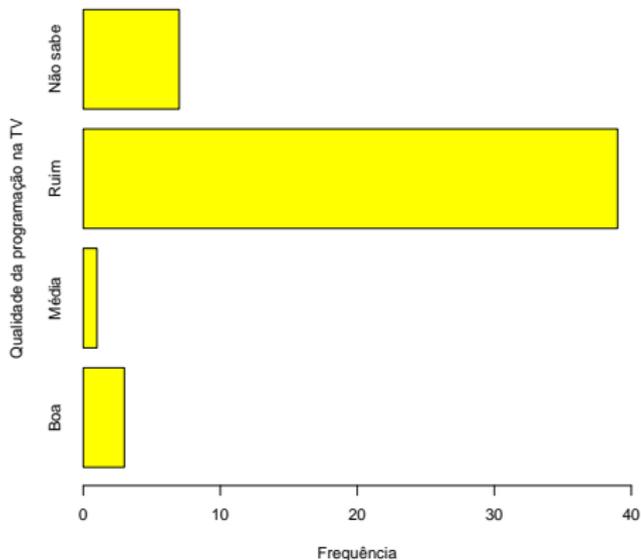


Figura: Gráfico de linhas para a variável temperatura.



Gráfico de Barras

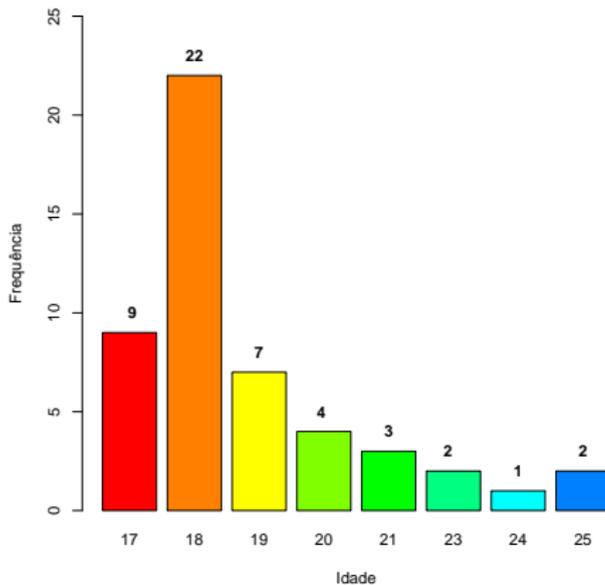


Fonte: Gráfico de barras para a variável opinião a respeito da qualidade da programação na TV.





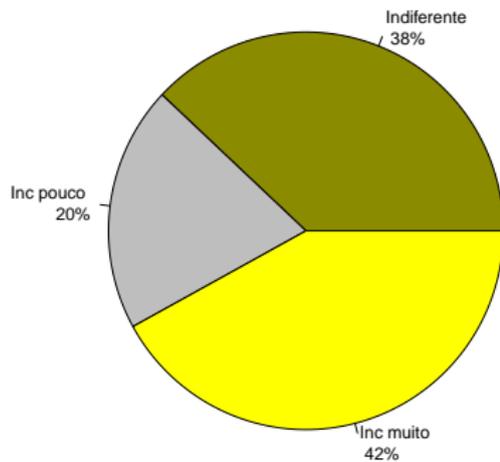
Gráfico de Colunas



Fonte: Gráfico de colunas para a variável idade.



Gráfico de Setores

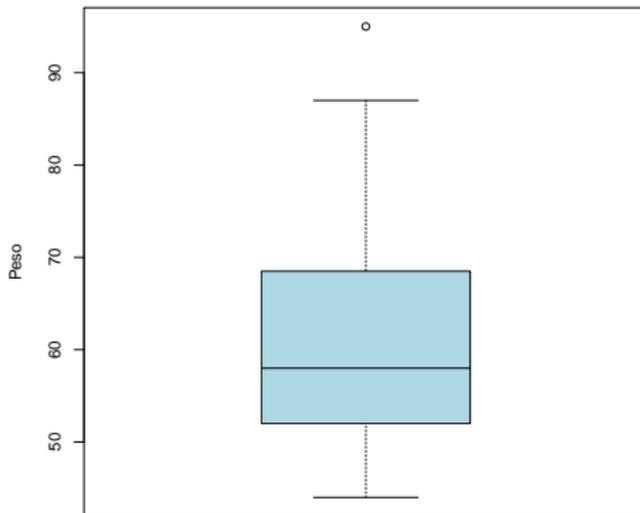


Fonte: Gráfico de setor para a variável tolerância ao cigarro.





Box plot

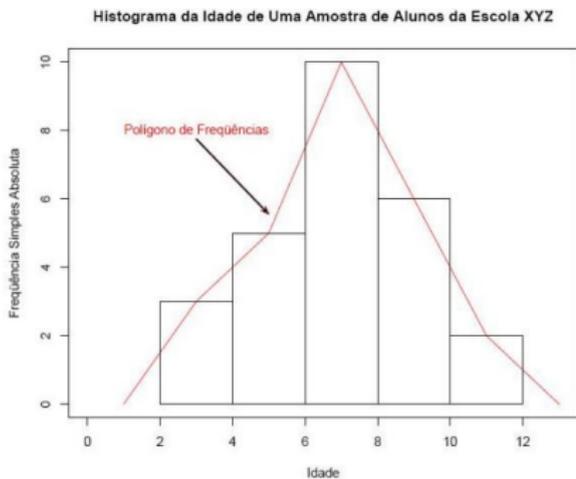


Fonte: Box plot para a variável peso.





Histograma e Polígono de Frequência

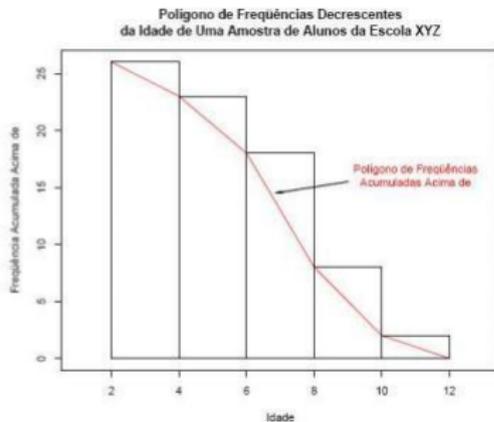
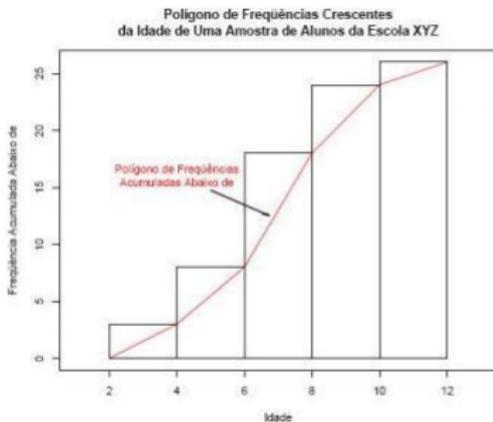


Fonte: Dados fictícios





Polígono de Freqüências Acumuladas - Ogiva de Galton



Fonte: Dados fictícios

