

Disciplina: 5170 - Probabilidade e Estatística

Aula 05

Prof. George Lucas M. Pezzott
glmpezzott2@uem.br

Departamento de Estatística - UEM
Sala: 222 - Bloco: E-90

Segunda-feira - 19:30 ~ 21:10 - D67 - Sala 108

Sexta-feira - 21:20 ~ 23:00 - D67 - Sala 108

Tabelas e gráficos

- Facilitar a leitura dos dados.
- Compreensão dos dados.
- Obter informações sobre a população.
- Fazer comparações.

Organização dos Dados

Representação dos Dados

- Representação na forma de tabelas:
 - Tabela de Frequência Simples;
 - Tabela de Frequência por classes.

- **Provedor usado pelos visitantes do site www.exemplo.com no dia 14/04/2019 (50 observações), em ordem de coleta:**

A	A	A	A	A	A	C	B	A	A
A	A	A	B	A	A	A	A	A	A
B	B	B	A	A	A	B	C	B	B
A	A	B	A	A	B	B	B	B	C
A	A	B	A	A	A	B	B	A	A

Organização dos Dados

Representação Tabular

Tabela: Distribuição de frequência do provedor usado pelo visitante do site www.exemplo.com no dia 14/04/2019.

Provedor	Frequências Simples		Frequências Acumuladas	
	Absoluta	Relativa	Absoluta	Relativa
	f_i	$fr_i(\%)$	F_i	$F_{r_i}(\%)$
A	31	62	31	62
B	16	32	47	94
C	2	4	49	98
D	1	2	50	100
Total	50	1,00	-	-

Fonte: Elaborada pelo professor

- **Exemplo: Tempos (em segundos) para carga de um aplicativo num sistema compartilhado (60 observações):**

12	62	46	36	54	66	67	63	59	22
64	15	17	14	47	31	40	43	58	21
31	11	61	25	24	60	33	50	67	36
29	64	28	28	65	40	68	58	51	61
36	53	11	70	27	13	20	39	54	25
15	62	52	68	64	12	18	55	34	34

Organização dos Dados

Representação Tabular

Tabela: Distribuição de frequências dos tempos (em segundos) para carga de um aplicativo num sistema compartilhado.

Tempos (Classes) c_i	Frequências				Ponto Médio Pm_i
	Absoluta f_i	Relativa fr_i	Acumulada F_i	Acum. Relativa Fr_i	
11 ┤ 21	11	0,1833	11	0,1833	16
21 ┤ 31	9	0,1500	20	0,3333	26
31 ┤ 41	11	0,1833	31	0,5166	36
41 ┤ 51	4	0,0667	35	0,5833	46
51 ┤ 61	10	0,1667	45	0,7500	56
61 ┤ 71	15	0,2500	60	1,0000	66
Total	60	1,0000	-	-	-

Fonte: Elaborada pelo autor

Principais Representações Gráficas

- **Gráficos para variáveis qualitativas:** barras, colunas e setores;
- **Gráficos para variáveis quantitativas:** linhas, histograma, polígono de frequências, polígonos de frequências acumuladas, boxplot e diagrama de dispersão.

Organização dos Dados

Representação Tabular

Tabela: Distribuição de frequência do provedor usado pelo visitante do site www.exemplo.com no dia 14/04/2019.

Provedor	Frequências Simples		Frequências Acumuladas	
	Absoluta	Relativa	Absoluta	Relativa
	f_i	$fr_i(\%)$	F_i	$F_{r_i}(\%)$
A	31	62	31	62
B	16	32	47	94
C	2	4	49	98
D	1	2	50	100
Total	50	1,00	-	-

Fonte: Elaborada pelo professor

Gráfico de colunas

No gráfico de colunas, as barras são **verticais** e separadas com tamanho proporcional a frequência observada de cada classe.

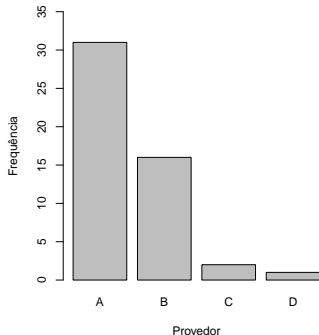


Figura: Gráfico de colunas para distribuição de frequência do provedor usado pelo visitante do site.

Gráfico de barras

No gráfico de barras, as barras são **horizontais** e separadas com tamanho proporcional a frequência observada de cada classe.

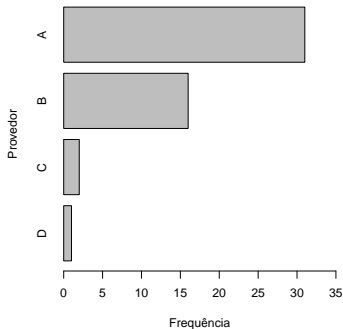


Figura: Gráfico de barras para distribuição de frequência do provedor usado pelo visitante do site.

Gráfico de setores

No gráfico de setores (pizza), o ângulo da “fatia” de uma classe é proporcional a sua frequência relativa. O uso é recomendada quando não há ordem específica nas respostas e o número de categorias não é muito grande (geralmente menor que 5).

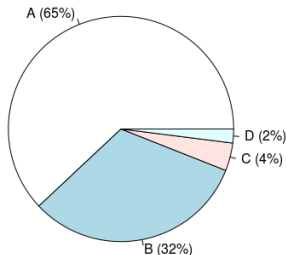


Figura: Gráfico de setores para distribuição de frequência do provedor usado pelo visitante do site.

Gráficos de linhas

- **Gráfico de linhas:** adequado para dados ordenados ao longo do tempo. Descreve como as mensurações da variável de interesse modificam-se ao longo do tempo. O eixo das abscissas representa os instantes em que a mensuração foi realizada e o eixo das ordenadas representa os valores observados.

Organização dos Dados

Representação Gráfica

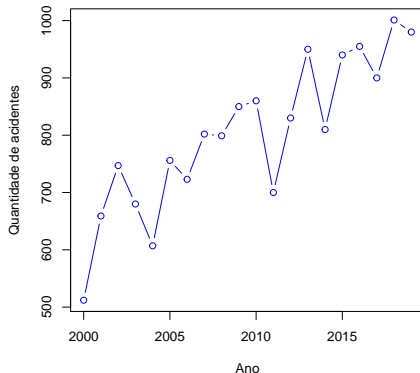
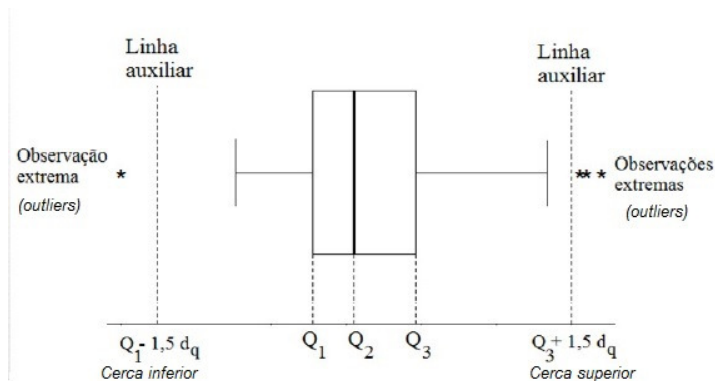


Figura: Comportamento da Quantidade de acidentes de trabalho no Brasil entre 2006 e 2019 (dados fictícios)

- **Gráfico de caixa (*box-plot*):**

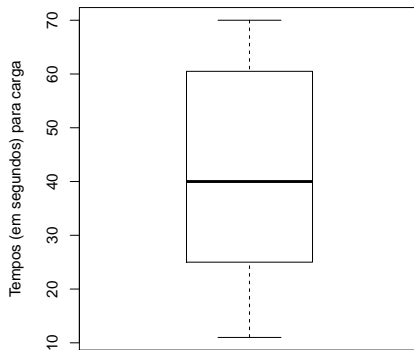
- Obter Q_1 , Q_2 (mediana) e Q_3 .
- Calcular a amplitude interquartil: $d_Q = Q_3 - Q_1$
- Calcular as cercas para identificação de *outliers*:
 $1,5(Q_3 - Q_1) = 1,5d_Q$
Cerca Inferior: $Q_1 - 1,5d_Q$
Cerca Superior: $Q_3 + 1,5d_Q$
- Obter o primeiro menor valor dos dados acima da cerca inferior.
- Obter o primeiro maior valor dos dados abaixo da cerca superior.

Ilustração:



Organização dos Dados

Representação Gráfica



Aplicativo A

Figura: *Box-plot* dos tempos das cargas dos dispositivos.

Organização dos Dados

Representação Gráfica

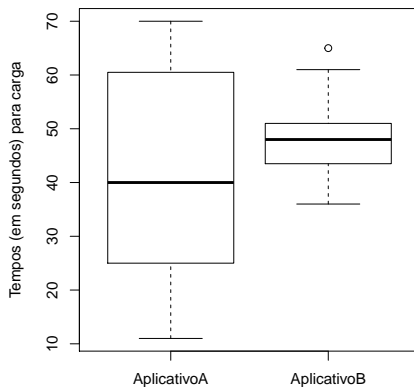


Figura: *Box-plot* dos tempos das cargas dos dispositivos.

- **Histograma:** é a forma mais usual de apresentação de distribuições de frequências de variáveis contínuas.

Tabela: Distribuição de frequências dos tempos (em segundos) para carga de um aplicativo num sistema compartilhado.

Tempos (Classes) c_i	Frequências				Ponto Médio Pm_i
	Absoluta f_i	Relativa fr_i	Acumulada F_i	Acum. Relativa Fr_i	
11 ┤ 21	11	0,1833	11	0,1833	16
21 ┤ 31	9	0,1500	20	0,3333	26
31 ┤ 41	11	0,1833	31	0,5166	36
41 ┤ 51	4	0,0667	35	0,5833	46
51 ┤ 61	10	0,1667	45	0,7500	56
61 ┤ 71	15	0,2500	60	1,0000	66
Total	60	1,0000	-	-	-

Fonte: Elaborada pelo autor

Organização dos Dados

Representação Gráfica

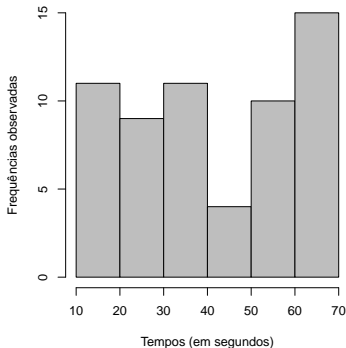


Figura: Distribuição de frequência dos tempos de carga de um certo aplicativo. Apresentação em histograma de frequência.

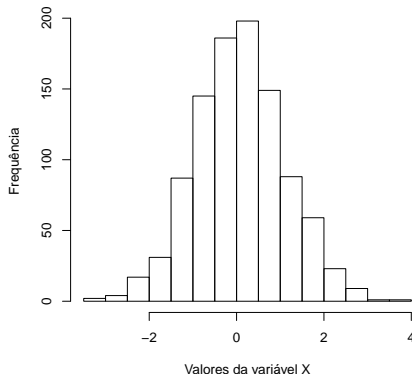


Figura: Distribuição de frequência da variável X. Apresentação em histograma de frequência.

Organização dos Dados

Representação Gráfica

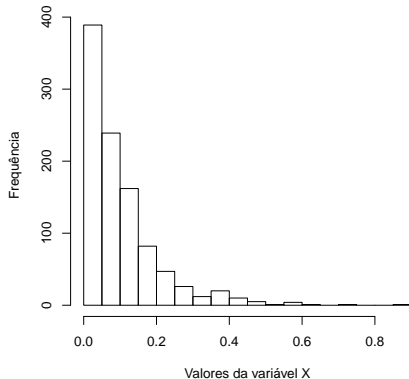


Figura: Distribuição de frequência da variável X. Apresentação em histograma de frequência.

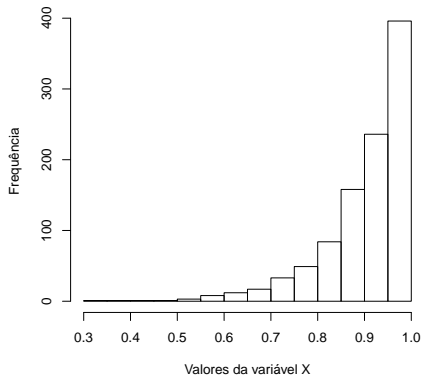


Figura: Distribuição de frequência da variável X. Apresentação em histograma de frequência.

- **Histograma:** São retângulos justapostos, feitos sobre as classes da variável em estudo. A área de cada retângulo é igual à frequência observada da correspondente classe.
 - Eixo x: as classes.
 - Eixo y: as frequências absolutas de observações de cada classe.

Em alguns casos, o eixo y pode representar as *densidades de frequências*:

$$d_i = \frac{fr_i}{LS_i - LI_i} = \frac{fr_i}{Ac_i}$$

e seu uso é fundamental quando as classes não têm a mesma amplitude.

- **Histograma:** São retângulos justapostos, feitos sobre as classes da variável em estudo. A área de cada retângulo é igual à frequência observada da correspondente classe.
 - Eixo x: as classes.
 - Eixo y: as frequências absolutas de observações de cada classe.

Em alguns casos, o eixo y pode representar as *densidades de frequências*:

$$d_i = \frac{fr_i}{LS_i - LI_i} = \frac{fr_i}{Ac_i}$$

e seu uso é fundamental quando as classes não têm a mesma amplitude.

Organização dos Dados

Representação Gráfica

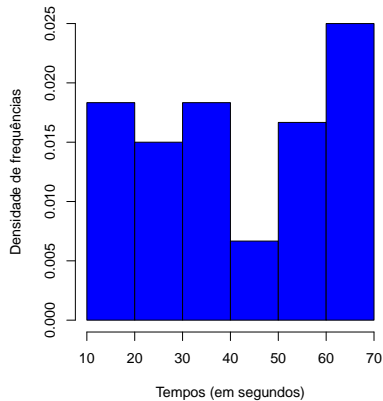
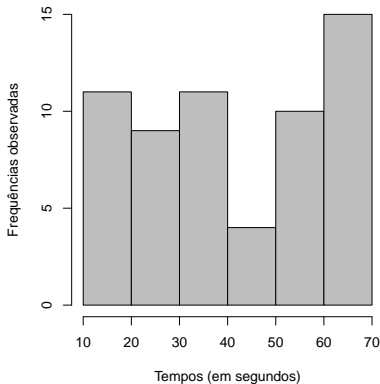


Figura: Distribuição de frequência dos tempos de carga de um certo aplicativo. Apresentação em histograma de frequência e densidade de frequências.

Polígono de frequência

- O polígono de frequências consiste da união de todos os segmentos de reta que ligam os pontos médios dos intervalos de classe.
- Os seguimentos de reta devem iniciar e terminar no eixo $y=0$.
- Assim como o histograma, o polígono de frequência serve para visualizar a forma da distribuição da variável estudada.

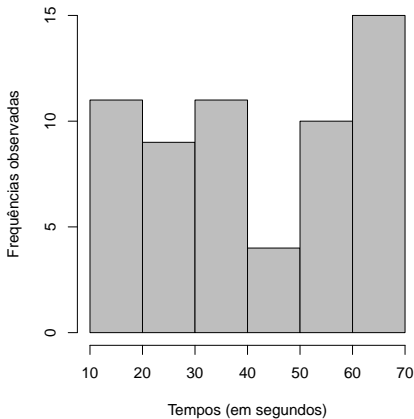


Figura: Distribuição de frequência dos tempos de carga de um certo aplicativo. Apresentação em histograma de frequência.

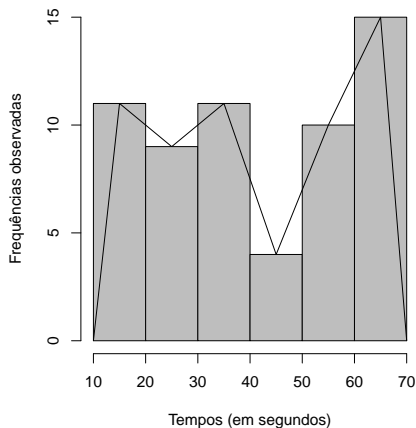


Figura: Distribuição de frequência dos tempos de carga de um certo aplicativo. Apresentação em histograma e polígono de frequência.

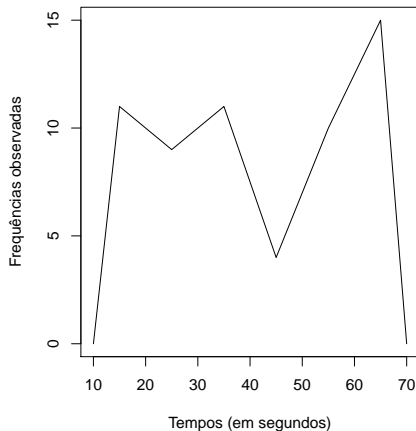


Figura: Distribuição de frequência dos tempos de carga de um certo aplicativo. Apresentação em polígono de frequência.

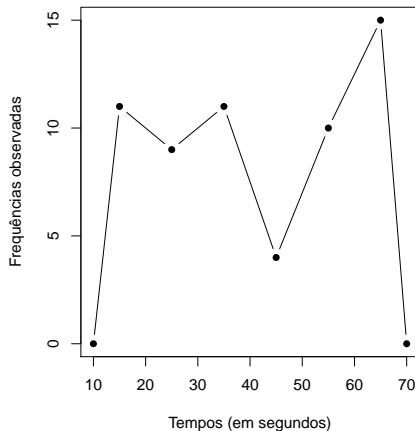


Figura: Distribuição de frequência dos tempos de carga de um certo aplicativo. Apresentação em polígono de frequência.

Polígono de frequências acumuladas

- O polígono de frequências acumuladas (absoluta, relativa ou porcentagem) obtém-se de um histograma de frequências acumuladas. Trata-se de uma curva poligonal ascendente.
- Através do polígono de porcentagens acumuladas pode-se estimar os percentis da distribuição, isto é, conhecer os valores da variável que estão acima ou abaixo de uma determinada porcentagem de interesse.

Polígono de frequências acumuladas

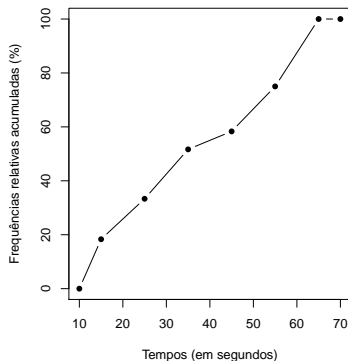


Figura: Distribuição de frequência relativas (%) dos tempos de carga de um certo aplicativo. Apresentação em polígono de frequência acumulada.