

Estatística Descritiva II

Bacharelado em Economia - FEA - Noturno

1^o Semestre 2014

Sumário

- 1 Objetivos da Aula
- 2 Fluência da Fala
- 3 Dados Agrupados
- 4 Histograma
- 5 Boxplot
- 6 Gráfico por Barras
- 7 Gráfico por Setores

Objetivos da Aula

Objetivos da Aula

Esta aula será motivada com um conjunto de dados reais que será inicialmente analisado através de algumas medidas resumo.

Posteriormente, discutiremos a forma de **dados agrupados** bem como a representação desses dados através de **histogramas**. Adaptações de algumas medidas resumo para dados agrupados serão discutidas. Finalmente, alguns gráficos mais conhecidos, tais como **boxplot**, **gráfico de barras** e **gráfico de setores**, serão apresentados.

Sumário

- 1 Objetivos da Aula
- 2 Fluência da Fala**
- 3 Dados Agrupados
- 4 Histograma
- 5 Boxplot
- 6 Gráfico por Barras
- 7 Gráfico por Setores

Fluência da Fala

Descrição do Projeto

- Dados do projeto **Perfil Evolutivo da Fluência da Fala de Falantes do Português Brasileiro**
- Estudo realizado pela Faculdade de Medicina - USP e pela Faculdade de Filosofia, Letras e Ciências Humanas - USP
- Projeto desenvolvido em 2006 referente a tese de doutorado
- Análise Estatística realizada pelo Centro de Estatística Aplicada (CEA) do IME-USP (**Dados CEA0P16**)

Fluência da Fala

Objetivo do Projeto

Avaliar o perfil da fluência da fala de acordo com a idade, gênero e grau de escolaridade.

Fluência da Fala

Objetivo do Projeto

Avaliar o perfil da fluência da fala de acordo com a idade, gênero e grau de escolaridade.

Amostra

A amostra consistiu de 594 indivíduos residentes na cidade de São Paulo com idade entre 2 e 99 anos.

Fluência da Fala

Descrição do Experimento

Foram obtidas de cada indivíduo amostras de fala auto-expressiva. O indivíduo era apresentado a uma figura e orientado a discorrer sobre a mesma durante um tempo mínimo de 3 minutos e máximo de 6 minutos. Para crianças de 2 e 3 anos, as amostras foram obtidas com a colaboração dos pais.

Fluência da Fala

Algumas Variáveis do Estudo

- Gênero (1:feminino e 2:masculino)
- Idade (em anos)
- Grau de escolaridade (pré-escola a superior)
- Fluxo de palavras por minuto (FPM)
- Fluxo de sílabas por minuto (FSM)
- Número de interjeições durante o discurso (INTERJ)
- Número de palavras não terminadas durante o discurso (PNT)
- Número de pausas durante o discurso (PAUSA).

Fluência da Fala

Variáveis Qualitativas

- Nominal: gênero
- Ordinal: grau de escolaridade

Fluência da Fala

Variáveis Qualitativas

- Nominal: gênero
- Ordinal: grau de escolaridade

Variáveis Quantitativas

- Contínua: idade, fluxo de palavras por minuto e fluxo de sílabas por minuto
- Discreta: número de interjeições durante o discurso, número de palavras não terminadas durante o discurso e número de pausas durante o discurso

Fluência da Fala

Medidas Resumo

Variável	Média	D. Padrão	CV
FPM	98,76	29,94	30%
FSM	179,29	54,71	30%
INTERJ	4,36	4,41	101%

Fluência da Fala

Medidas Resumo

Variável	Média	D. Padrão	CV
FPM	98,76	29,94	30%
FSM	179,29	54,71	30%
INTERJ	4,36	4,41	101%

- o fluxo médio de sílabas por minuto é quase que o dobro do fluxo médio de palavras por minuto
- a variável com maior dispersão em relação à média é número de interjeições
- fluxo de sílabas e fluxo de palavras apresentam dispersão em relação à média praticamente iguais

Fluência da Fala

Medidas Resumo

Variável	n	min	Q1	Md	Q3	max
FPM	594	28,7	78,05	96,94	117,98	209,09
FSM	594	53,8	142,80	176,47	214,29	364,64
INTERJ	594	0	1	3	6	25

Fluência da Fala

Medidas Resumo

Variável	n	min	Q1	Md	Q3	max
FPM	594	28,7	78,05	96,94	117,98	209,09
FSM	594	53,8	142,80	176,47	214,29	364,64
INTERJ	594	0	1	3	6	25

- 50% dos indivíduos falaram até 3 interjeições
- o fluxo de sílabas de 75% dos indivíduos foi igual ou menor a 214,29 sílabas por minuto
- 25% dos entrevistados tiveram um fluxo de palavras menor ou igual a 78,05 palavras por minuto

Fluência da Fala

Medidas Resumo: FPM

Gênero	Média	D. Padrão	CV
Feminino	99,34	29,69	30%
Masculino	97,95	30,33	31%

Fluência da Fala

Medidas Resumo: FPM

Gênero	Média	D. Padrão	CV
Feminino	99,34	29,69	30%
Masculino	97,95	30,33	31%

- as mulheres têm valor médio para FPM ligeiramente superior aos homens
- as variabilidades em relação às médias são praticamente as mesmas

Fluência da Fala

Medidas Resumo: FPM

Gênero	n	min	Q1	Md	Q3	max
Feminino	349	28,7	79,52	98,70	118,93	209,09
Masculino	245	34,3	76,20	96,39	117,80	181,62

Fluência da Fala

Medidas Resumo: FPM

Gênero	n	min	Q1	Md	Q3	max
Feminino	349	28,7	79,52	98,70	118,93	209,09
Masculino	245	34,3	76,20	96,39	117,80	181,62

- as mulheres têm valor mediano para FPM ligeiramente superior aos homens
- o maior e o menor valor ocorrem no grupo feminino, porém o grupo masculino tem uma variabilidade interquartil maior do que o grupo feminino

Sumário

- 1 Objetivos da Aula
- 2 Fluência da Fala
- 3 Dados Agrupados**
- 4 Histograma
- 5 Boxplot
- 6 Gráfico por Barras
- 7 Gráfico por Setores

Fluência da Fala

Número de Interjeições

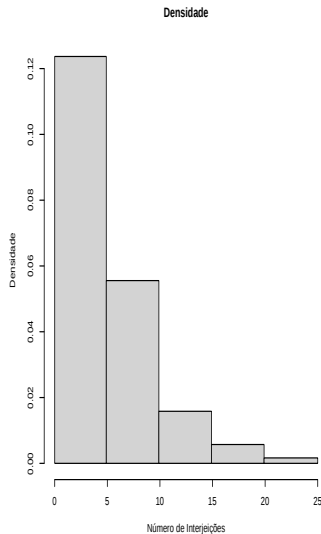
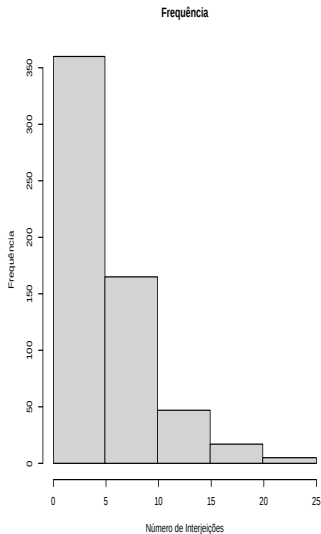
Classe	n_i	f_i	%
0 - 5	360	0,606	60,6
5 - 10	165	0,278	27,8
10 - 15	47	0,079	7,9
15 - 20	17	0,029	2,9
20 - 25	5	0,008	0,8
Total	594	1,000	100

em que n_i e f_i correspondem, respectivamente, à frequência absoluta e à frequência relativa da i -ésima classe, $f_i = n_i/n$.

Sumário

- 1 Objetivos da Aula
- 2 Fluência da Fala
- 3 Dados Agrupados
- 4 Histograma**
- 5 Boxplot
- 6 Gráfico por Barras
- 7 Gráfico por Setores

Bases Iguais



Histograma por Densidade

Definição

No histograma por densidade a área da classe corresponde à frequência relativa da classe

$$f = \Delta \times h,$$

em que Δ é a base e h a altura da classe.

Histograma por Densidade

Número de Interjeições

Classe	Δ_i	f_i	h_i
0 - 5	5	0,606	0,1212
5 - 10	5	0,278	0,0556
10 - 15	5	0,079	0,0158
15 - 20	5	0,029	0,0058
20 - 25	5	0,008	0,0016

Histograma por Densidade

Número de Interjeições

Classe	Δ_i	f_i	h_i
0 - 5	5	0,606	0,1212
5 - 10	5	0,278	0,0556
10 - 15	5	0,079	0,0158
15 - 20	5	0,029	0,0058
20 - 25	5	0,008	0,0016

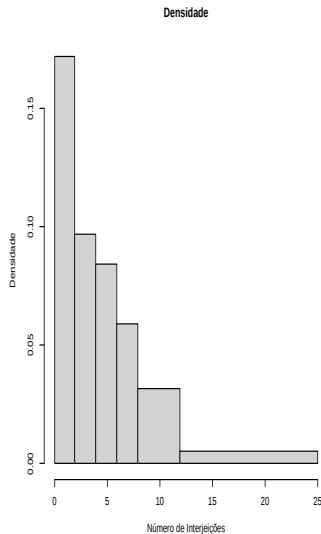
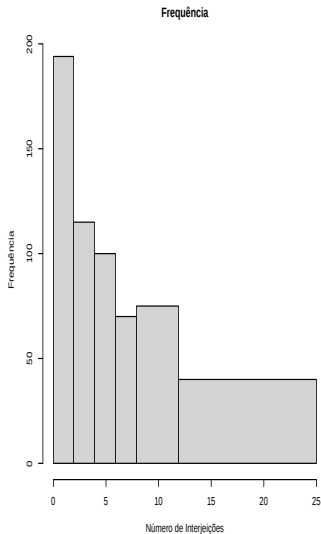
Note que a densidade é dada por $h = \frac{f}{\Delta}$.

Bases Desiguais

Número de Interjeições

Classe	n_i	f_i	%
0 - 2	194	0,327	32,7
2 - 4	115	0,194	19,4
4 - 6	100	0,168	16,8
6 - 8	70	0,118	11,8
8 - 12	75	0,126	12,6
12 - 25	40	0,067	6,7
Total	594	1,000	100

Bases Desiguais

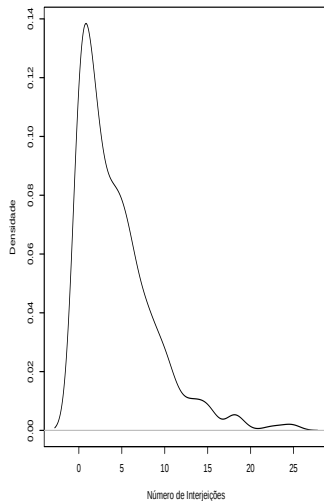
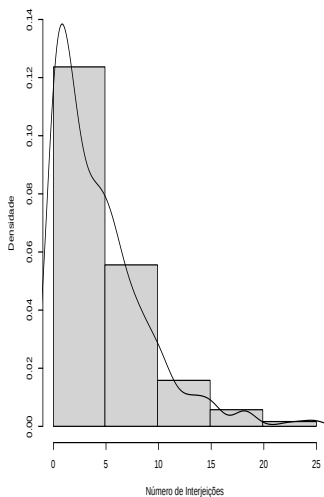


Histograma

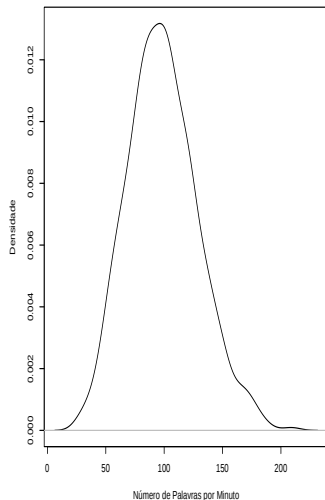
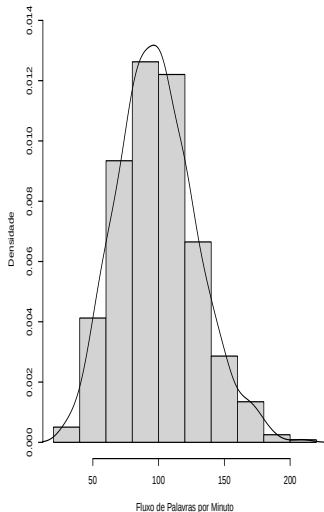
Comentários

- quando as bases são iguais os histogramas pelos métodos da frequência e da densidade apresentam a mesma forma, portanto qualquer um dos dois métodos poderá ser aplicado
- quando as bases são desiguais o histograma pelo método da frequência pode apresentar distorções na forma da distribuição, conforme pode ser notado pelo histograma da frequência do número de interjeições. Assim, neste caso, somente o método da densidade deve ser utilizado

Aproximação por Densidade: INTERJ



Aproximação por Densidade: FPM



Dados Agrupados

Média Aproximada

No caso de dados agrupados podemos obter uma média aproximada

$$\begin{aligned}\bar{x}^* &= \frac{n_1 x_1^* + \cdots + n_k x_k^*}{n} \\ &= f_1 x_1^* + \cdots + f_k x_k^* \\ &= \sum_{i=1}^k f_i x_i^*,\end{aligned}$$

em que k é o número de classes, n_i é o número de indivíduos da i -ésima classe, x_i^* é o ponto médio da i -ésima classe e $f_i = n_i/n$.

Dados Agrupados

Variância Aproximada

No caso de dados agrupados podemos obter uma variância aproximada

$$\begin{aligned}
 s^{*2} &= \frac{n_1(x_1^* - \bar{x}^*)^2 + \cdots + n_k(x_k^* - \bar{x}^*)^2}{n} \\
 &= f_1(x_1^* - \bar{x}^*)^2 + \cdots + f_k(x_k^* - \bar{x}^*)^2 \\
 &= \sum_{i=1}^k f_i(x_i^* - \bar{x}^*)^2.
 \end{aligned}$$

Desvio Padrão Aproximado

O desvio padrão aproximado fica dado por

$$s^* = \sqrt{s^{*2}}.$$

Média Aproximada

Número de Interjeições

Classe	f_i	x_i^*
0 - 5	0,606	2,5
5 - 10	0,278	7,5
10 - 15	0,079	12,5
15 - 20	0,029	17,5
20 - 25	0,008	22,5

Média Aproximada

Número de Interjeições

Classe	f_i	x_i^*
0 - 5	0,606	2,5
5 - 10	0,278	7,5
10 - 15	0,079	12,5
15 - 20	0,029	17,5
20 - 25	0,008	22,5

Portanto, temos que $\bar{x}^* = 0,606 \times 2,5 + 0,278 \times 7,5 + 0,079 \times 12,5 + 0,029 \times 17,5 + 0,008 \times 22,5 = 5,275$.

Variância Aproximada

Número de Interjeições

Classe	x_i^*	f_i	$f_i(x_i^* - \bar{x}^*)^2$
0 - 5	2,5	0,606	4,667
5 - 10	7,5	0,278	1,376
10 - 15	12,5	0,079	4,124
15 - 20	17,5	0,029	4,334
20 - 25	22,5	0,008	2,374
Total		1,000	16,875

Variância Aproximada

Número de Interjeições

Classe	x_i^*	f_i	$f_i(x_i^* - \bar{x}^*)^2$
0 - 5	2,5	0,606	4,667
5 - 10	7,5	0,278	1,376
10 - 15	12,5	0,079	4,124
15 - 20	17,5	0,029	4,334
20 - 25	22,5	0,008	2,374
Total		1,000	16,875

Portanto, temos que $s^* = \sqrt{16,875} \cong 4,108$.

Sumário

- 1 Objetivos da Aula
- 2 Fluência da Fala
- 3 Dados Agrupados
- 4 Histograma
- 5 Boxplot**
- 6 Gráfico por Barras
- 7 Gráfico por Setores

Boxplot

Definição

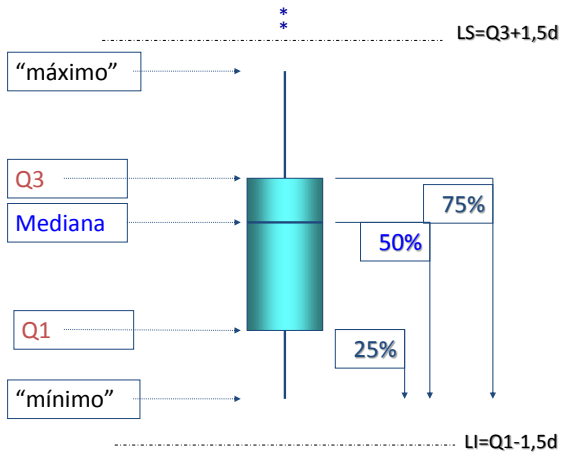
É um tipo de gráfico para descrever a distribuição dos dados que é construído com os quartis Q1 e Q3, a mediana e valores extremos. Foi criado por [John Tukey](#) nos anos 70.

Boxplot

Construção

- o corpo do gráfico consiste de uma caixa que é construída do quartil Q1 até o quartil Q3
- dentro da caixa uma linha que corresponde à mediana é traçada
- duas linhas horizontais LI e LS, respectivamente, abaixo e acima dos quartis Q1 e Q3 são traçadas, em que $LI = Q1 - 1,5d$ e $LS = Q3 + 1,5d$ com $d = Q3 - Q1$
- linhas verticais são traçadas da caixa até o maior ou menor valor antes da linha horizontal correspondente
- observações que estão acima ou abaixo das linhas horizontais são destacadas

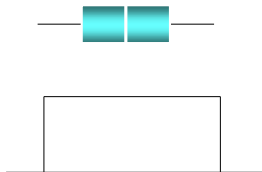
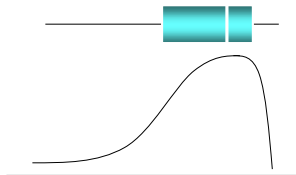
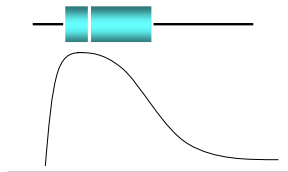
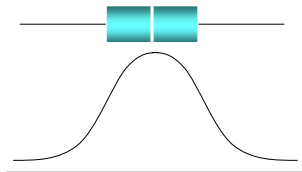
Ilustração



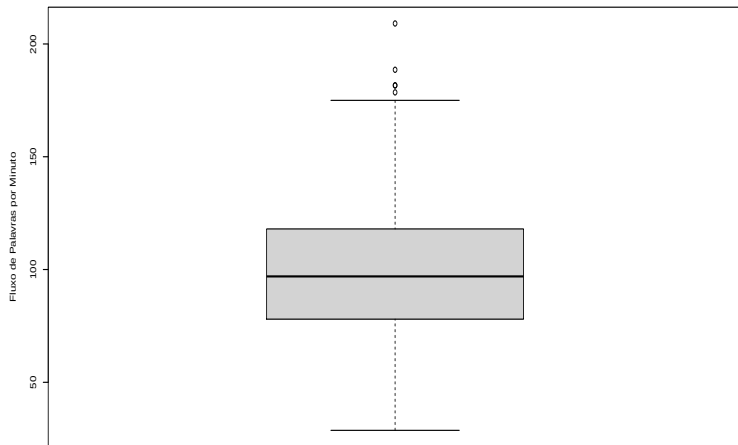
"máximo" é o maior valor menor do que LS

"mínimo" é o menor valor maior do que LI

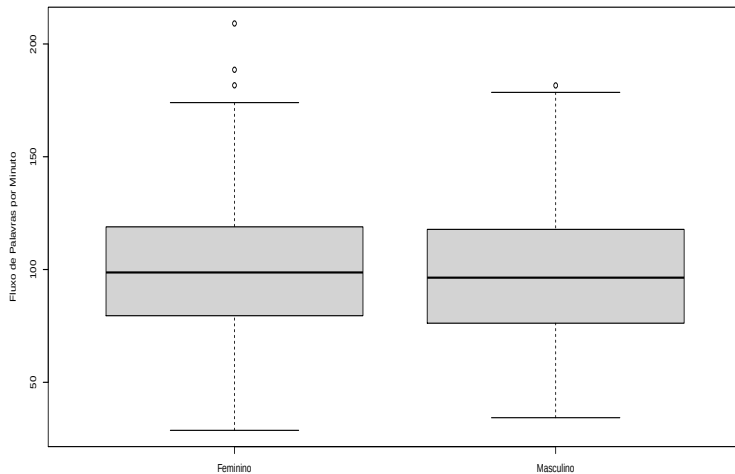
Formas da Distribuição



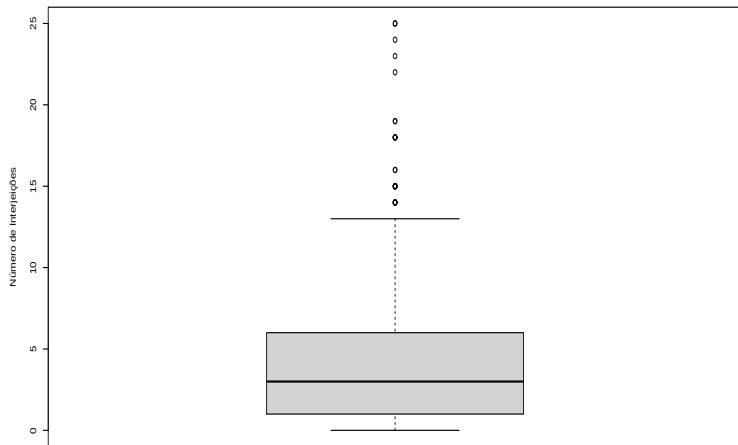
Boxplot: FPM



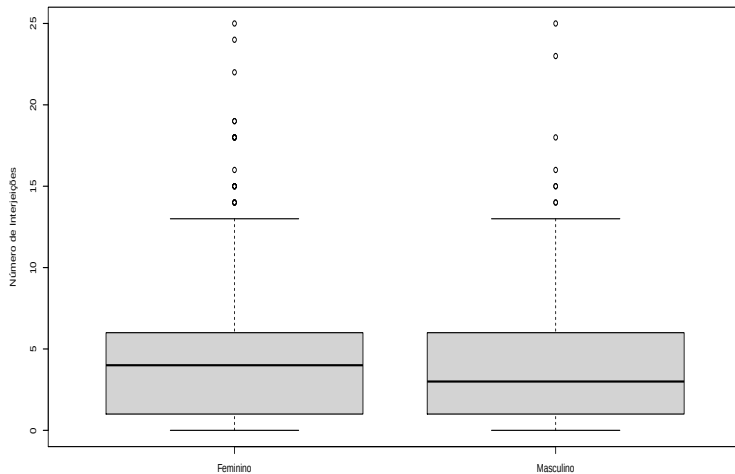
Boxplot: FPM por Sexo



Boxplot: INTERJ



Boxplot: INTERJ por Sexo



Sumário

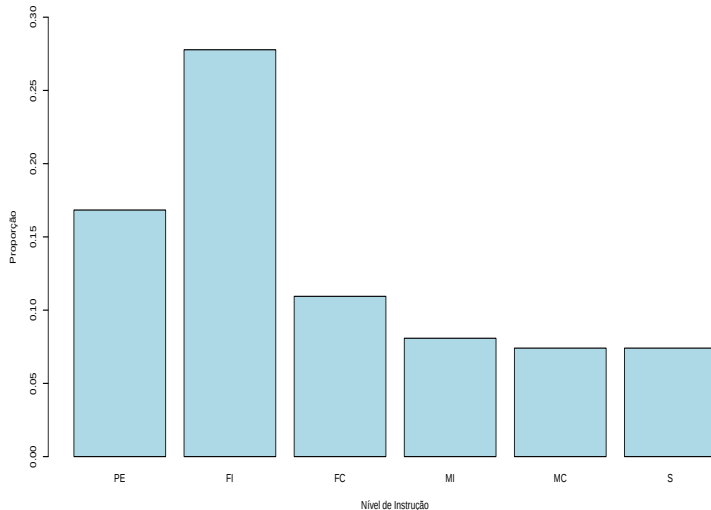
- 1 Objetivos da Aula
- 2 Fluência da Fala
- 3 Dados Agrupados
- 4 Histograma
- 5 Boxplot
- 6 Gráfico por Barras**
- 7 Gráfico por Setores

Gráfico por Barras

Definição

Os gráficos de barras são muito usados para comparar os níveis ou atributos de variáveis qualitativas. As barras podem aparecer na vertical ou na horizontal, quando também são chamadas de colunas. Seja na horizontal ou na vertical, quanto maior o comprimento de uma barra, maior o valor que representa. Quanto menor o comprimento de uma barra menos valor ela tem. A não ser que esteja representando números negativos, esta regra é aplicável.

Gráfico por Barras: Nível de Instrução



Sumário

- 1 Objetivos da Aula
- 2 Fluência da Fala
- 3 Dados Agrupados
- 4 Histograma
- 5 Boxplot
- 6 Gráfico por Barras
- 7 Gráfico por Setores**

Gráfico por Setores

Definição

Os gráficos por setores são representados por círculos divididos proporcionalmente de acordo com os níveis ou atributos de variáveis qualitativas. Os valores são expressos em números ou em porcentagem.

Gráfico por Setores: Gênero

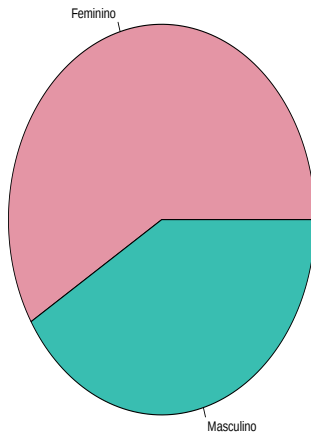


Gráfico por Setores: Nível de Instrução

