



SME0803 Visualização e Exploração de Dados

Associação entre variável quantitativa e qualitativa

Prof. Cibeles Russo

cibele@icmc.usp.br

Baseado em

Murteira, B. J. F., Análise Exploratória de Dados. McGraw-Hill, Lisboa, 1993.

Notas de aula de Análise Exploratória de Dados do Mário de Castro, ICMC-USP, 2010.

Associação entre variável quantitativa e qualitativa

Seja $x \in x_1, \dots, x_k, 1 < k \leq n$, uma variável qualitativa e y uma variável quantitativa.

Dados observados: n pares de valores (x_j, y_j) , sendo que

$x_j \in x_1, \dots, x_k, j = 1, \dots, n$.

É muito comum o interesse na relação causal unilateral $x \rightarrow y$.

Associação entre variável quantitativa e qualitativa

Seja $x \in x_1, \dots, x_k, 1 < k \leq n$, uma variável qualitativa e y uma variável quantitativa.

Dados observados: n pares de valores (x_j, y_j) , sendo que

$x_j \in x_1, \dots, x_k, j = 1, \dots, n$.

É muito comum o interesse na relação causal unilateral $x \rightarrow y$.

Representação dos dados:

- medidas resumo e gráficos de y para cada nível de x ,
- gráficos de médias,
- boxplots y para cada nível de x ,
- gráficos de violino de y para cada nível de x .

Associação entre variável quantitativa e qualitativa

Cada nível x_i ocorre f_i vezes (frequência). Para cada nível x_i calculamos a variância s_i^2 dos valores y_j para os quais $x_j = x_i$, $j = 1, \dots, n$ e $i = 1, \dots, k$.

Média ponderada das variâncias:

$$\bar{s}^2 = \frac{\sum_{i=1}^k f_i s_i^2}{\sum_{i=1}^k f_i} = \frac{\sum_{i=1}^k f_i s_i^2}{n}$$

Variância de y: $s^2 = \frac{1}{n-1} \sum_{j=1}^n (y_j - \bar{y})^2$

Obs: podemos ter $s_i = 0$ mas $s^2 > 0$.

Associação entre variável quantitativa e qualitativa

Ganho na variância: $s^2 - \bar{s}^2$

Ganho relativo na variância:

$$R^2 = \frac{s^2 - \bar{s}^2}{s^2} = 1 - \frac{\bar{s}^2}{s^2}, 0 \leq R^2 \leq 1.$$

Quanto maior R^2 , mais forte a associação entre x para y .

Quanto maior R^2 , maior o poder de explicação de x para y (em termos de variabilidade).

Associação entre variável quantitativa e qualitativa

Considere novamente os dados da Companhia MB (Bussab & Morettin, 2009). Como investigar a associação entre salário e grau de instrução dos funcionários?

O cálculo de $R^2 = 0.37$ aponta para uma possível associação entre grau de instrução e salário.

Vamos considerar medidas descritivas e representações gráficas.

Associação entre variável quantitativa e qualitativa

Tabela: Medidas-resumo para o salário em diferentes níveis de grau de instrução.

Grau de instrução	Salário	
	média	desvio-padrão
ensino fundamental	7.84	2.96
ensino médio	11.53	3.72
ensino superior	16.48	4.50

Associação entre variável quantitativa e qualitativa

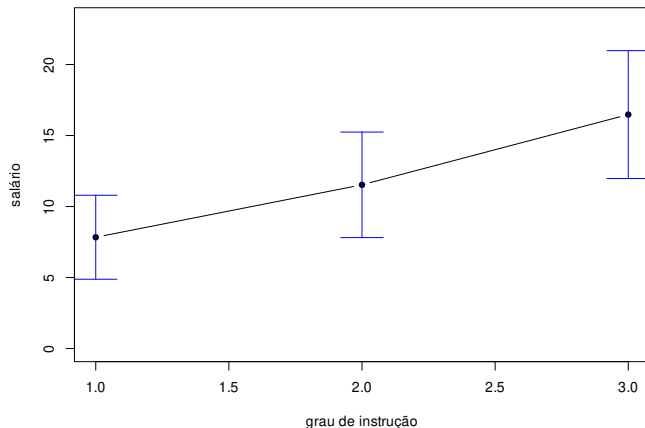


Figura: Gráfico de médias e desvios-padrão de salário em diferentes níveis de grau de instrução.

Associação entre variável quantitativa e qualitativa

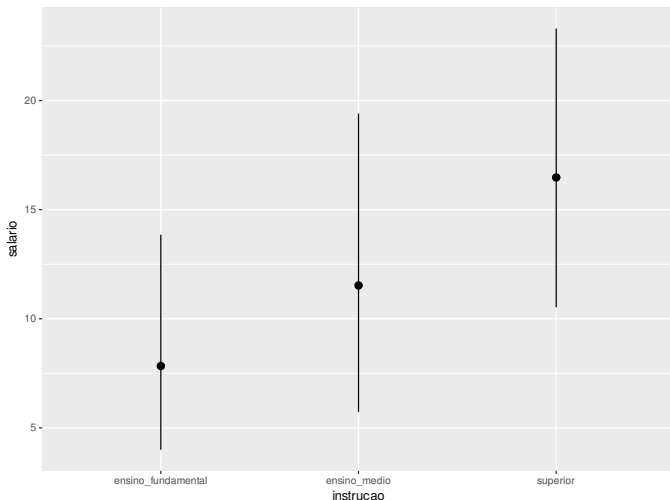


Figura: Gráfico de médias, mínimo e máximo de salário em diferentes níveis de grau de instrução.

Associação entre variável quantitativa e qualitativa

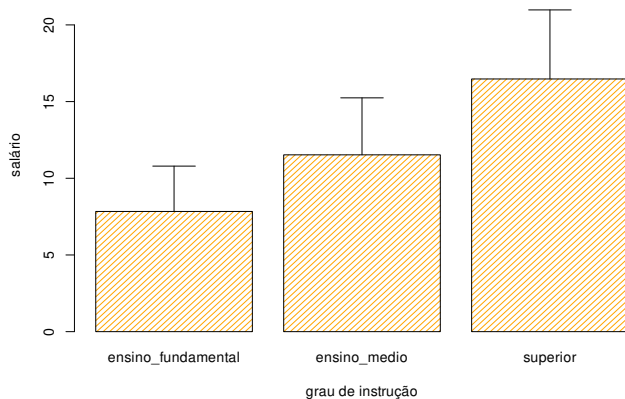


Figura: Gráfico de médias e desvios-padrão de salário em diferentes níveis de grau de instrução.

Associação entre variável quantitativa e qualitativa

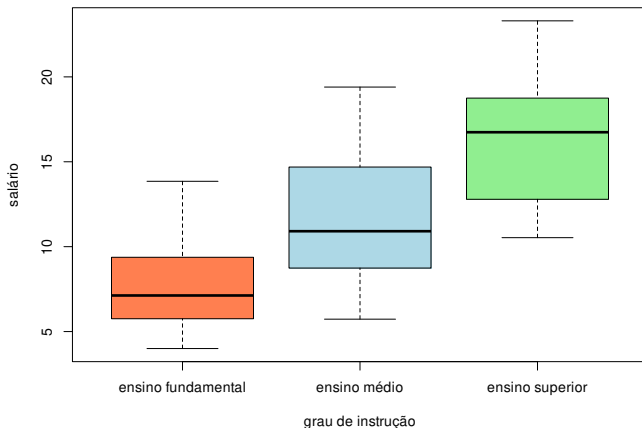


Figura: Boxplots para salário em diferentes níveis de grau de instrução.

Associação entre variável quantitativa e qualitativa

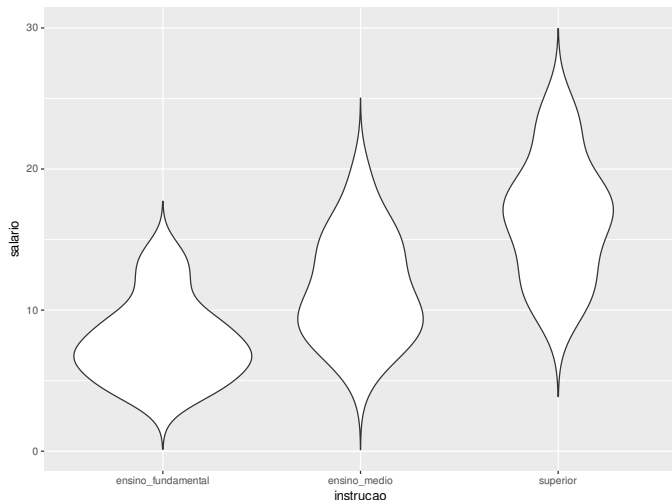


Figura: Gráficos de violino para salário em diferentes níveis de grau de instrução.

Associação entre variável quantitativa e qualitativa

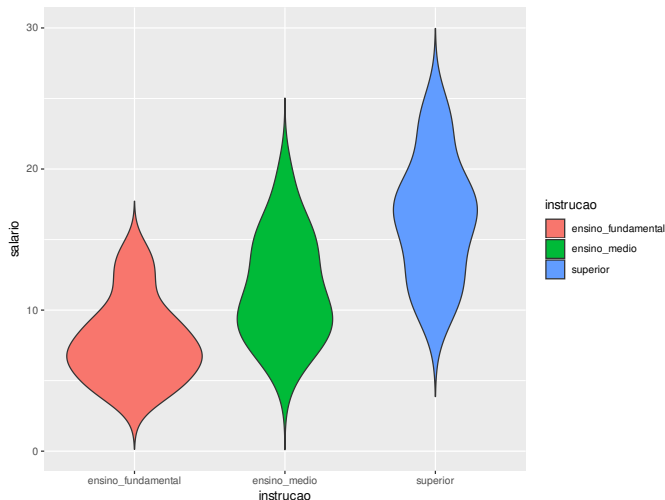


Figura: Gráficos de violino para salário em diferentes níveis de grau de instrução.

Associação entre variável quantitativa e qualitativa

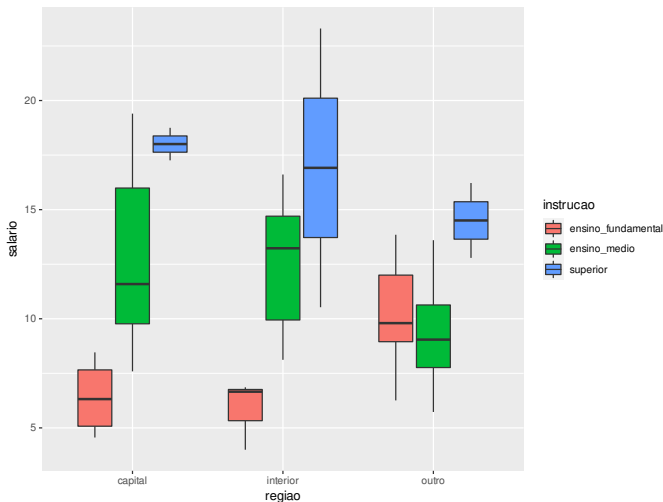


Figura: Boxplots para salário em diferentes níveis de grau de instrução condicionado a região.