

Estatística Básica

Prof. Fernando de Souza Bastos
fernando.bastos@ufv.br

Instituto de Ciências Exatas e Tecnológicas
Universidade Federal de Viçosa
Campus UFV - Florestal



Sumário

- 1 Associação entre Variáveis Quantitativas
- 2 Associação entre Variáveis Qualitativas e Quantitativas

Tabela: Anos de Serviço (X) versus
Nº de Clientes (Y)

Agente	X	Y
A	2	48
B	4	56
C	5	64
D	6	60
E	6	65
F	6	63
G	7	67
H	8	70
I	8	71
J	10	72

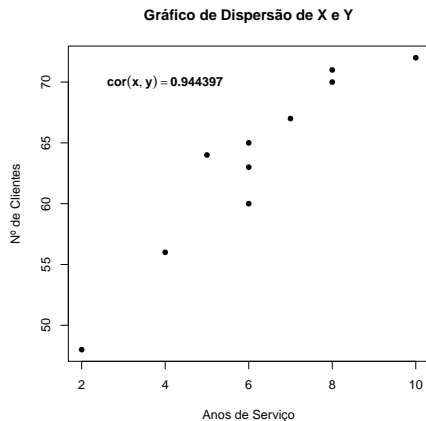
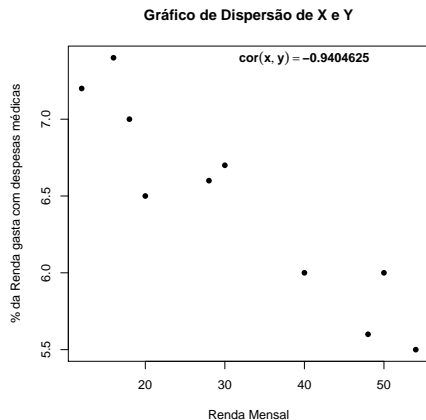


Tabela: Renda bruta mensal (X) e porcentagem da renda gasta em saúde (Y).

Família	X	Y
A	12	7.2
B	16	7.4
C	18	7.0
D	20	6.5
E	28	6.6
F	30	6.7
G	40	6.0
H	48	5.6
I	50	6.0
J	54	5.5



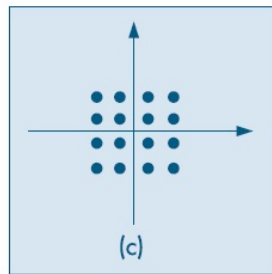
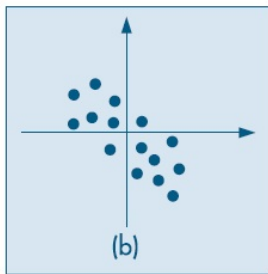
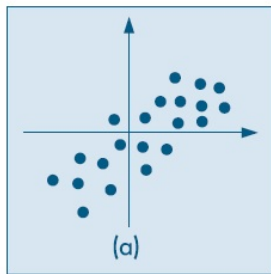


Figura: Morettin and Bussab (2009)

Gráfico de Dispersão

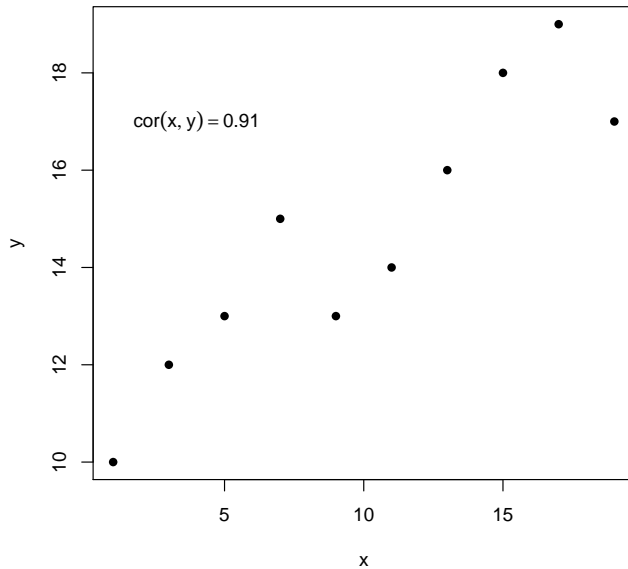


Gráfico de Dispersão

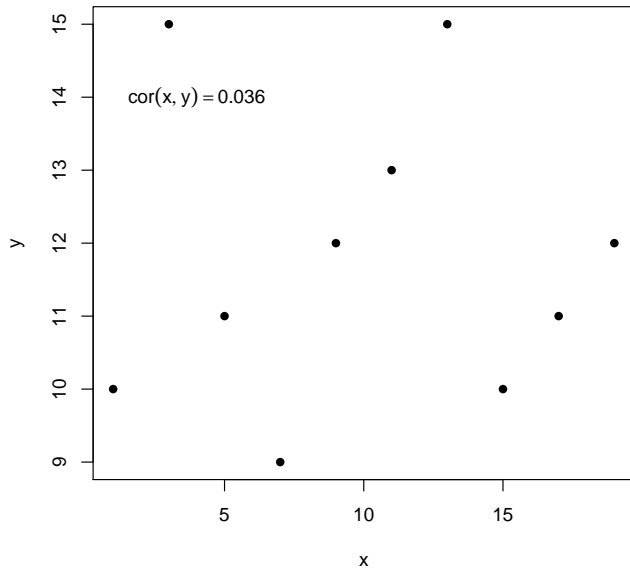


Gráfico de Dispersão

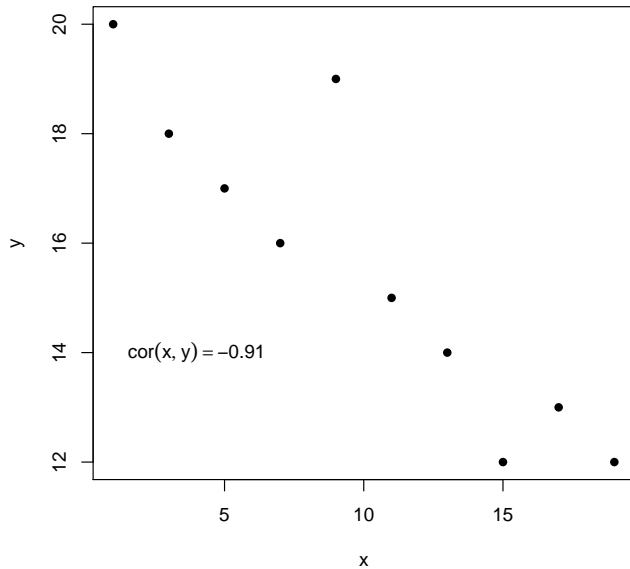


Gráfico de Dispersão

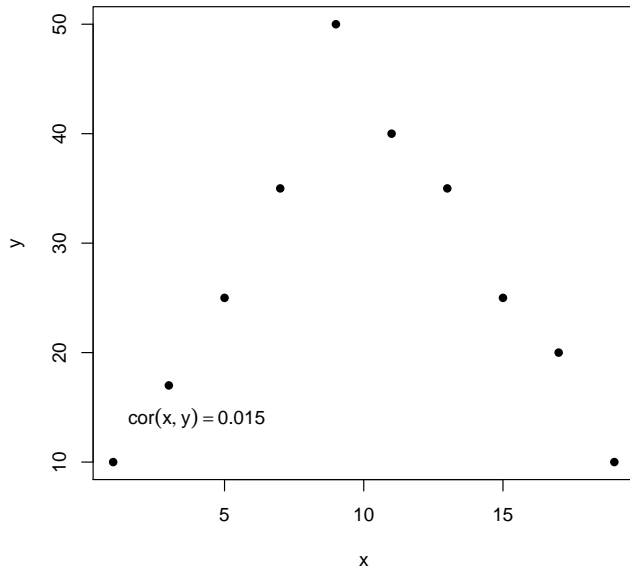


Tabela: Cálculo do coeficiente de correlação.

Agente	Anos	Clientes	$x - \bar{x}$	$y - \bar{y}$	$\frac{x - \bar{x}}{dp(x)} = z_x$	$\frac{y - \bar{y}}{dp(y)} = z_y$	$z_x \cdot z_y$
A	2	48	-3.7	-8.5	-1.54	-1.05	1.617
B	3	50	-2.7	-6.5	-1.12	-0.8	0.896
C	4	56	-1.7	-0.5	-0.71	-0.06	0.043
D	5	52	-0.7	-4.5	-0.29	-0.55	0.160
E	4	43	-1.7	-13.5	-0.71	-1.66	1.179
F	6	60	0.3	3.5	0.12	0.43	0.052
G	7	62	1.3	5.5	0.54	0.68	0.367
H	8	58	2.3	1.5	0.95	0.18	0.171
I	8	64	2.3	7.5	0.95	0.92	0.874
J	10	72	4.3	15.5	1.78	1.91	3.400
Total	57	565	0	0			8.759

$$Cor = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{dp(X)} \right) \left(\frac{y_i - \bar{y}}{dp(Y)} \right) = \frac{8.759}{10} = 0.8759$$

Não é difícil provar que o coeficiente de correlação satisfaz:

$$-1 \leq \text{cor}(X, Y) \leq 1$$

Não é difícil provar que o coeficiente de correlação satisfaz:

$$-1 \leq \text{cor}(X, Y) \leq 1$$

DEF: Dados n pares de valores $(x_1, y_1), \dots, (x_n, y_n)$, chamaremos de covariância entre as duas variáveis X e Y a igualdade:

$$\text{cov}(X, Y) = \sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{n}$$

Não é difícil provar que o coeficiente de correlação satisfaz:

$$-1 \leq \text{cor}(X, Y) \leq 1$$

DEF: Dados n pares de valores $(x_1, y_1), \dots, (x_n, y_n)$, chamaremos de co-variância entre as duas variáveis X e Y a igualdade:

$$\text{cov}(X, Y) = \sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{n}$$

Com a definição acima, o coeficiente de correlação pode ser escrito como:

$$\text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{dp(X)dp(Y)}$$

A covariância mede a relação linear entre duas variáveis. A covariância é semelhante à correlação entre duas variáveis, no entanto, elas diferem nas seguintes maneiras:

- Os coeficientes de correlação são padronizados. Assim, um relacionamento linear perfeito resulta em um coeficiente de correlação 1. A correlação mede tanto a força como a direção da relação linear entre duas variáveis.

A covariância mede a relação linear entre duas variáveis. A covariância é semelhante à correlação entre duas variáveis, no entanto, elas diferem nas seguintes maneiras:

- Os coeficientes de correlação são padronizados. Assim, um relacionamento linear perfeito resulta em um coeficiente de correlação 1. A correlação mede tanto a força como a direção da relação linear entre duas variáveis.
- Os valores de covariância não são padronizados. Como os dados não são padronizados, é difícil determinar a força da relação entre as variáveis.

Associação entre Variáveis Qualitativas e Quantitativas

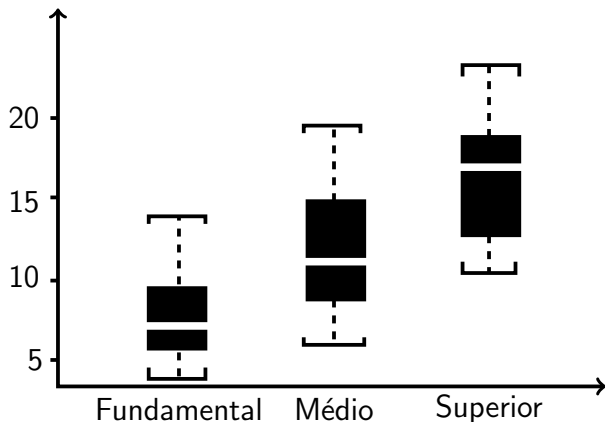


Figura: Box plots de salário segundo grau de instrução. Morettin and Bussab (2009)

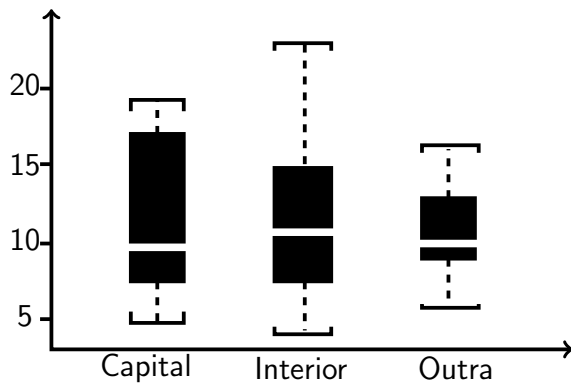


Figura: Box plots de salário segundo região de procedência. Morettin and Bussab (2009)

Referências

P. Morettin and W. Bussab. *Estatística básica*. Editora Saraiva, São Paulo, 6 edition, 2009.