

Análise exploratória III

Análises bivariadas

Prof. Me. Lineu Alberto Cavazani de Freitas

CE003 – Estatística II

Departamento de Estatística
Laboratório de Estatística e Geoinformação



Análise exploratória bivariada

- ▶ Em alguns casos podemos estar interessados na análise de duas variáveis simultaneamente.
- ▶ O objetivo é investigar a relação de **associação** entre as variáveis.
- ▶ Tabelas, gráficos e coeficientes específicos para relação entre variáveis podem ser usados.
- ▶ Tal como nas análises univariadas, as escolhas dependem dos tipos de variáveis.
- ▶ Considerando variáveis aos pares, as combinações podem ser:
 - ▶ Qualitativa x qualitativa.
 - ▶ Quantitativa x quantitativa.
 - ▶ Quantitativa x qualitativa.



Análise bivariada para variáveis qualitativas

Análise bivariada para variáveis qualitativas

- ▶ Neste tipo de situação avaliamos a **frequência** de observações para cada **combinação** de níveis das duas variáveis.
- ▶ Podem ser usadas **tabelas de frequências cruzadas**, também chamadas de **tabelas de dupla entrada**.
- ▶ Também é possível representar as frequências por meio de **recursos gráficos**.

Tabelas de frequências cruzadas

- ▶ As linhas dizem respeito aos níveis de uma variável.
- ▶ As colunas aos níveis da outra variável.
- ▶ As células mostram as frequências (absolutas ou relativas).
- ▶ As margens mostram as frequências marginais (de apenas uma das duas variáveis).
- ▶ No caso de frequências relativas podem ser usados os totais linha, coluna ou o total geral.

Tabelas de frequências cruzadas

Tabela 1. Tabela de dupla entrada para...

	f	g	h	i	j	Total
a	21	12	19	15	22	89
b	16	21	16	28	30	111
c	15	15	22	19	24	95
d	17	22	21	20	27	107
e	20	31	18	17	12	98
Total	89	101	96	99	115	500

Tabelas de frequências cruzadas

Tabela 2. Tabela de dupla entrada total linha...

	f	g	h	i	j	Total
a	0.24	0.13	0.21	0.17	0.25	1
b	0.14	0.19	0.14	0.25	0.27	1
c	0.16	0.16	0.23	0.20	0.25	1
d	0.16	0.21	0.20	0.19	0.25	1
e	0.20	0.32	0.18	0.17	0.12	1
Total	0.90	1.00	0.97	0.98	1.14	5

Tabelas de frequências cruzadas

Tabela 3. Tabela de dupla entrada total coluna...

	f	g	h	i	j	Total
a	0.24	0.12	0.20	0.15	0.19	0.90
b	0.18	0.21	0.17	0.28	0.26	1.10
c	0.17	0.15	0.23	0.19	0.21	0.95
d	0.19	0.22	0.22	0.20	0.23	1.06
e	0.22	0.31	0.19	0.17	0.10	1.00
Total	1.00	1.00	1.00	1.00	1.00	5.00

Tabelas de frequências cruzadas

Tabela 4. Tabela de dupla entrada total geral...

	f	g	h	i	j	Total
a	0.04	0.02	0.04	0.03	0.04	0.18
b	0.03	0.04	0.03	0.06	0.06	0.22
c	0.03	0.03	0.04	0.04	0.05	0.19
d	0.03	0.04	0.04	0.04	0.05	0.21
e	0.04	0.06	0.04	0.03	0.02	0.20
Total	0.18	0.20	0.19	0.20	0.23	1.00

Análise bivariada para variáveis qualitativas

- ▶ As frequências cruzadas podem ser representadas por meio de gráficos.
- ▶ Variações de gráficos de barras são as opções mais comuns.
- ▶ As possibilidades podem usar as frequências absolutas, relativas e permitem comparar a composição das variáveis.
- ▶ Gráficos para frequência para duas variáveis qualitativas:
 - ▶ Gráficos de barras lado a lado.
 - ▶ Gráfico de barras empilhadas.
 - ▶ Gráficos de barras empilhadas relativo.

Gráficos de barras lado a lado

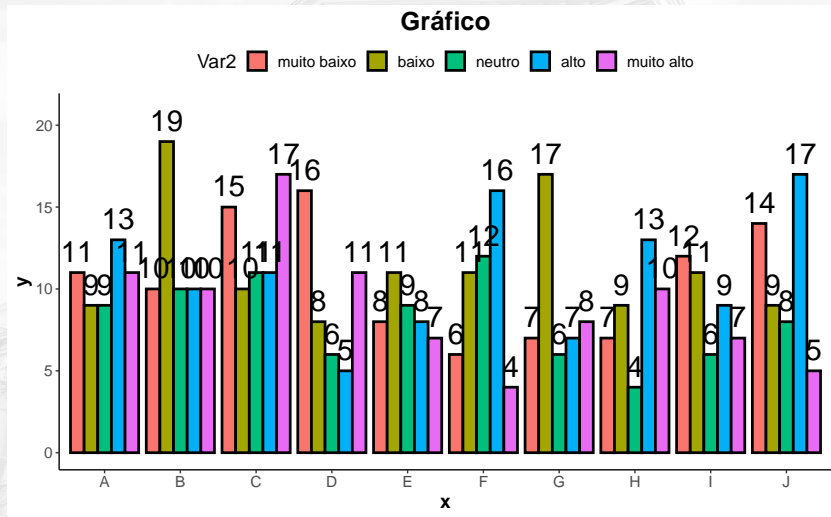


Figura 1. Gráfico de barras lado a lado...

Gráficos de barras empilhadas

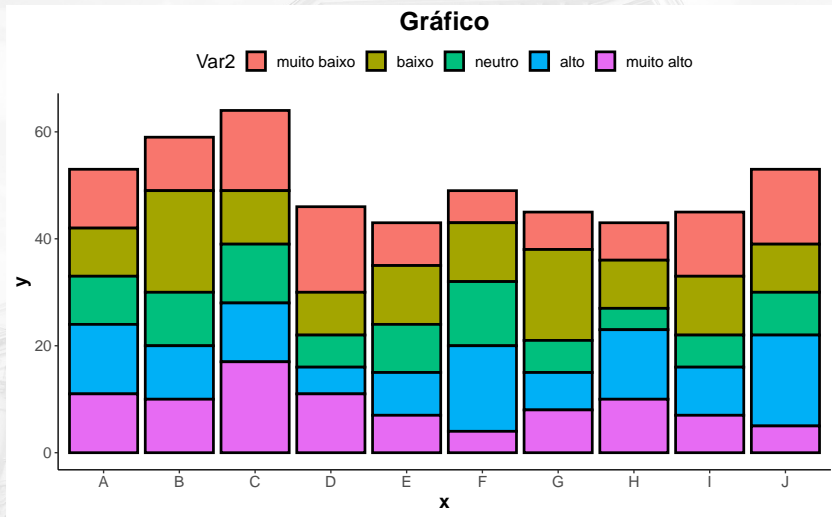


Figura 2. Gráfico de barras empilhadas...

Gráficos de barras empilhadas relativo

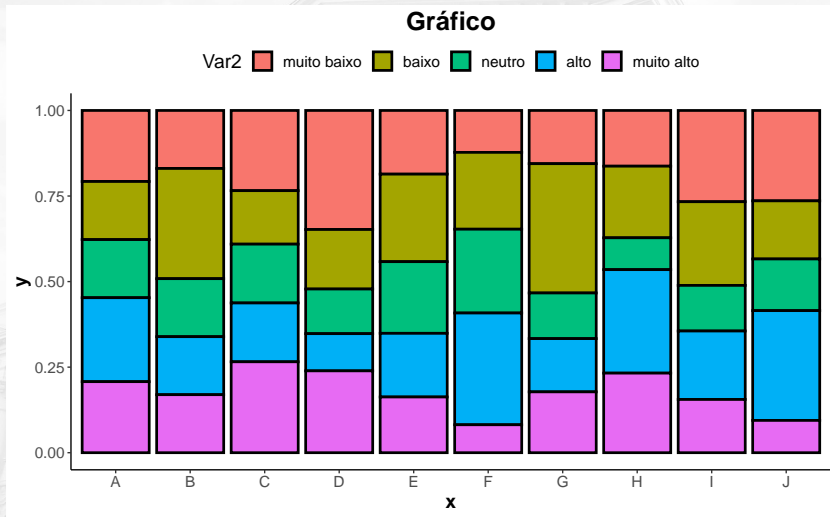


Figura 3. Gráfico de barras empilhadas relativo...



Análise bivariada para variáveis quantitativas

Análise bivariada para variáveis quantitativas

- ▶ Buscamos identificar **padrões** e **tendências** na análise das duas variáveis.
 - ▶ A medida que os valores de uma variável aumentam, a outra reduz?
 - ▶ A medida que os valores de uma variável aumentam, a outra aumenta?
 - ▶ A medida que os valores de uma variável aumentam, a outra se mantém estável?
- ▶ As principais técnicas são o **coeficiente de correlação** e o **diagrama de dispersão**.
 - ▶ O coeficiente é uma métrica que avalia a associação linear entre um par de variáveis numéricas.
 - ▶ O diagrama é um gráfico de pares ordenados.

Coefficiente de correlação linear de Pearson

- ▶ Usado para determinar se existe **relação linear** entre variáveis quantitativas.

- ▶ Assume valores entre -1 e 1.
- ▶ Se o valor é maior 0, então existe uma associação linear **positiva**.
- ▶ Se o valor é menor que 0, então existe uma associação linear **negativa**.
- ▶ Se o valor é igual a 0, então **não existe** uma associação linear.

▶ **CORRELAÇÃO NÃO IMPLICA EM CAUSALIDADE.**

- ▶ O fato de existir uma correlação linear, seja positiva ou negativa, não implica que uma variável possui real influência nos desfechos da outra.
- ▶ Causalidade causa correlação, mas correlação não implica em causalidade.

Covariância e correlação

- ▶ A covariância entre duas variáveis Y_1 e Y_2 é dada por:

$$\text{Cov}(y_1, y_2) = \frac{1}{n-1} \sum_{i=1}^n (y_{1i} - \bar{y}_1) \cdot (y_{2i} - \bar{y}_2).$$

- ▶ A partir da covariância podemos obter a correlação, que padroniza a medida pelas variâncias, fazendo com que, independente das variáveis, sempre seja um valor entre -1 e 1.

$$r = \frac{\sum_{i=1}^n (y_{1i} - \bar{y}_1) \cdot (y_{2i} - \bar{y}_2)}{\sqrt{\sum_{i=1}^n (y_{1i} - \bar{y}_1)^2} \cdot \sqrt{\sum_{i=1}^n (y_{2i} - \bar{y}_2)^2}} = \frac{\text{Cov}(y_1, y_2)}{\sqrt{V(y_1) \cdot V(y_2)}}.$$

Outros tipos de correlação

- ▶ A correlação de Pearson não serve para descrever associações que não sejam lineares.
- ▶ Existem outros tipos de correlação que servem inclusive para variáveis de outros tipos.
- ▶ Alguns exemplos são:
 - ▶ Correlação de Spearman.
 - ▶ Correlação de Kendall.
 - ▶ Ponto-bisserial.

Diagrama de dispersão

- ▶ O diagrama de dispersão é a ferramenta favorita para visualizar duas variáveis quantitativas.
- ▶ Em um eixo são representados os valores de uma variável.
- ▶ No outro eixo os valores de uma segunda variável.
- ▶ Os pares ordenados são representados por pontos.

Diagrama de dispersão

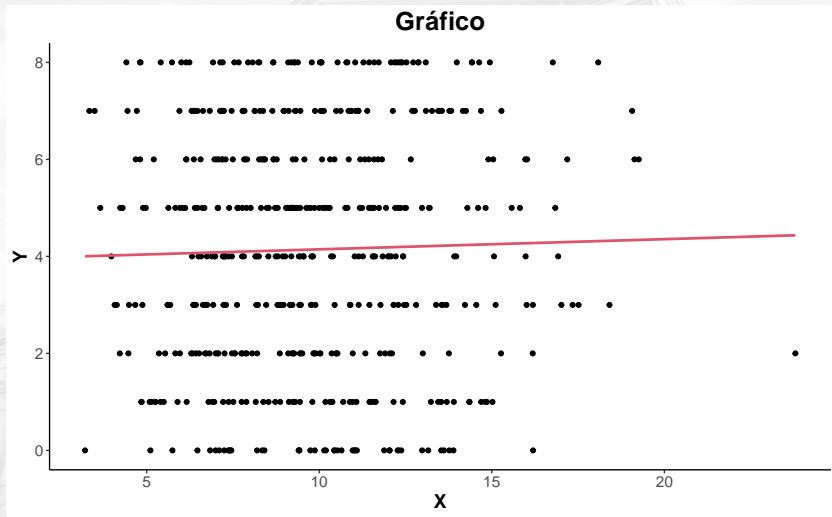


Figura 4. Diagrama de dispersão...

Interpretação gráfica

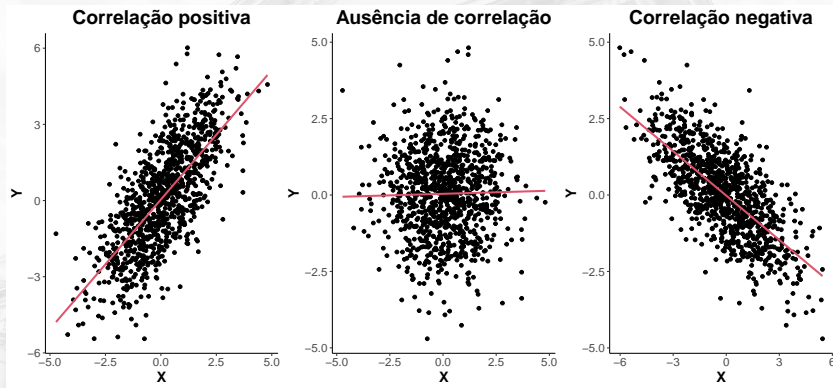


Figura 5. Diagrama de dispersão...

Covariância, correlação e diagrama de dispersão

Exemplo

- Considere as variáveis peso (Y_1) e altura (Y_2) de um conjunto de 10 indivíduos.

Y_1 : 60,09; 57,97; 54,12; 70,76; 59,74; 50,41; 58,19; 65,35; 71,18; 54,76

Y_2 : 1,54; 1,62; 1,52; 1,76; 1,63; 1,52; 1,65; 1,67; 1,66; 1,57

- $\overline{Y_1} = 60,26$; $\overline{Y_2} = 1,61$.
- $Var(Y_1) = 47,8$; $Var(Y_2) = 0,006$.
- Obtenha a covariância, coeficiente de correlação e o diagrama de dispersão.

Covariância, correlação e diagrama de dispersão

Exemplo

$$\text{Cov}(y_1, y_2) = \frac{1}{10 - 1} \{[(60,09 - 60,26) \cdot (1,54 - 1,61)] + \dots + [(57,76 - 60,26) \cdot (1,57 - 1,61)]\}$$

$$\text{Cov}(y_1, y_2) = 0,44$$

$$r = \frac{0,44}{\sqrt{47,8 \cdot 0,006}} = 0,82$$

Covariância, correlação e diagrama de dispersão

Exemplo - digrama de dispersão

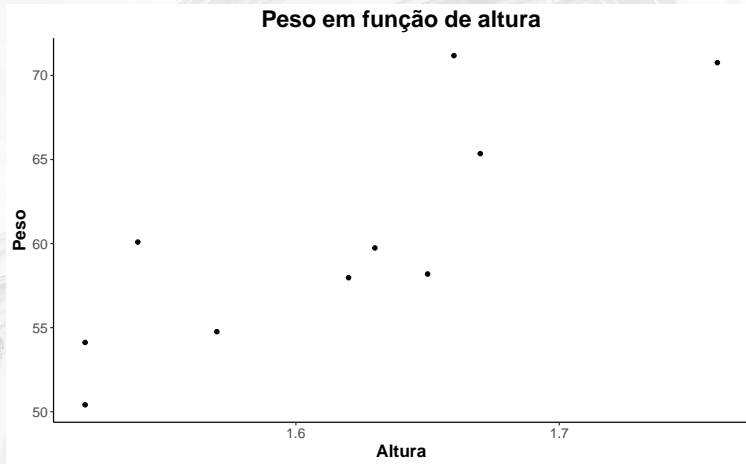


Figura 6. Diagrama de dispersão para peso e altura.

Covariância, correlação e diagrama de dispersão

Exemplo - digrama de dispersão

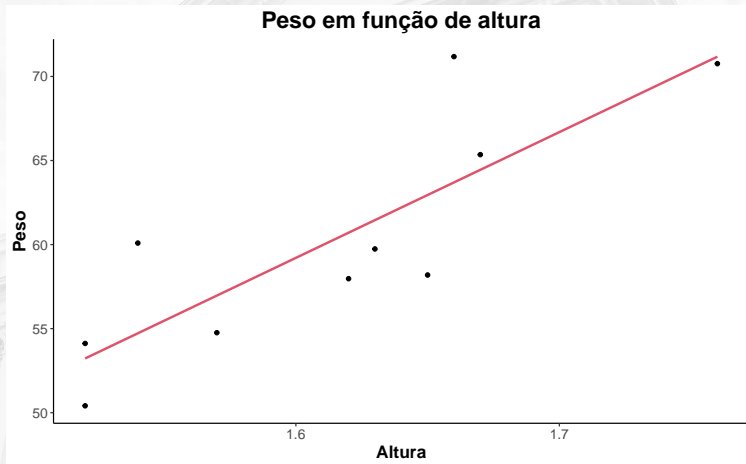
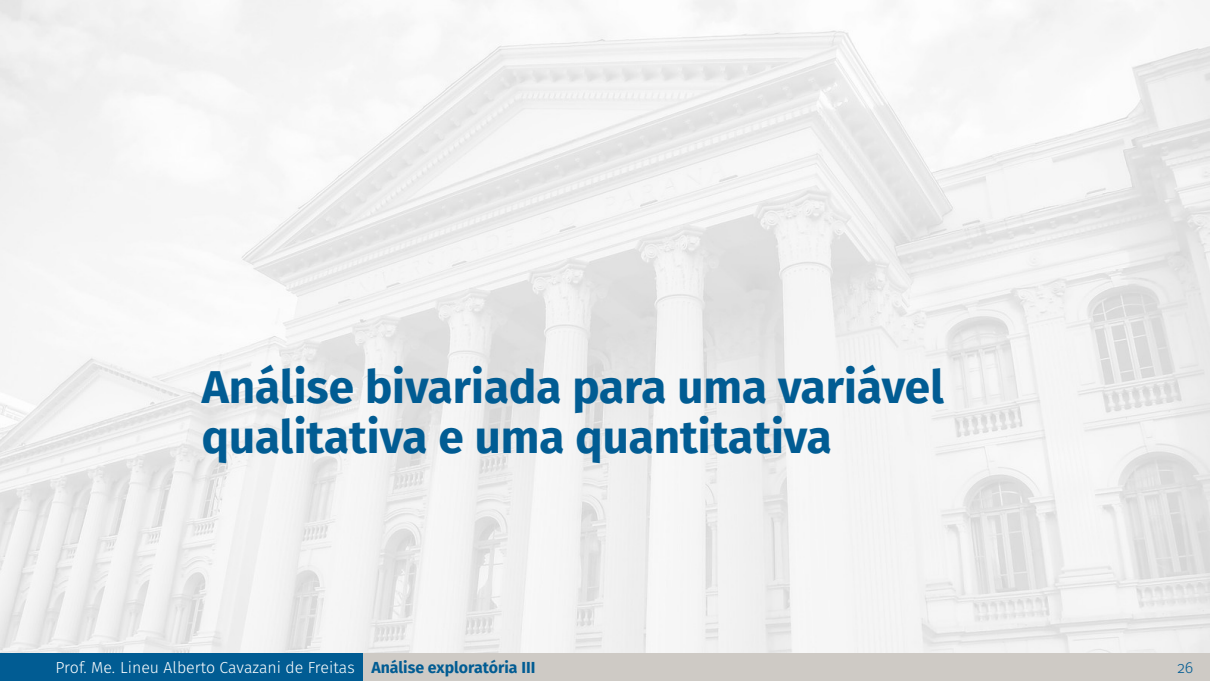


Figura 7. Diagrama de dispersão para peso e altura com linha de tendência linear.



Análise bivariada para uma variável qualitativa e uma quantitativa

Análise bivariada para uma variável qualitativa e uma quantitativa

- ▶ Neste caso estamos interessados em avaliar se os valores da variável numérica estão associados com os níveis da variável categórica.
- ▶ Podemos usar medidas descritivas para os valores dentro de cada um dos níveis da variável categórica.
- ▶ Para representar graficamente esta situação podemos criar um boxplot da variável numérica para cada nível do fator de interesse.

Tabela de medidas descritivas para níveis de um fator

Tabela 5. Tabela...

nominal	Média	Mediana	Desvio padrão
A	9.84	9.45	2.85
B	9.53	9.37	3.28
C	9.22	8.91	2.83
D	9.65	9.64	2.48
E	9.97	9.79	2.99
F	10.05	9.20	3.48
G	9.96	10.35	2.91
H	10.44	10.44	3.71
I	9.00	9.04	3.29
J	9.97	9.32	3.41

Box-plot para níveis de um fator

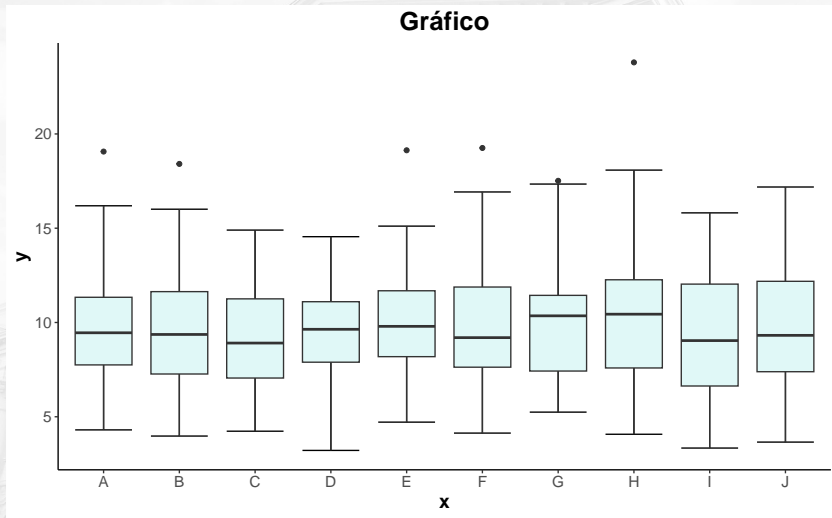


Figura 8. box-plot para



Outros tipos de gráficos e análises

Outros tipos de gráficos e análises

- ▶ Vimos as alternativas usuais para representação e análise de variáveis quantitativas e qualitativas.
- ▶ Contudo existem diversas situações particulares que exigem análises específicas.
- ▶ Algumas casos são: mapas, séries temporais, gráficos de perfil, nuvens de palavras.
- ▶ Também é possível trabalhar com gráficos que representam mais de duas variáveis ao mesmo tempo.
- ▶ Outra possibilidade é combinar gráficos.

O que foi visto:

- ▶ Análises bivariadas.
 - ▶ Qualitativa x qualitativa.
 - ▶ Quantitativa x quantitativa.
 - ▶ Quantitativa x qualitativa.

Próximos assuntos:

- ▶ Introdução à probabilidades.