

Estatística Descritiva

Julienne Borges

Gráficos

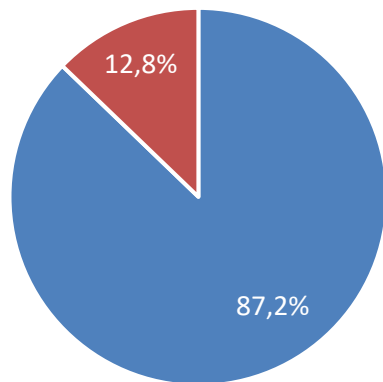
“Enquanto as tabelas interagem com nosso sistema verbal, os gráficos interagem com nosso sistema visual, que é mais rápido no processamento de informações. Isso significa que um gráfico bem projetado normalmente comunicará a informação mais rápido que uma tabela bem projetada. “

(Knafllic, 2019, p.53)

Gráfico de setores

- É um gráfico muito comum para representar distribuições de frequências de variáveis qualitativas.
- É particularmente útil quando o número de categorias não é grande e as categorias não obedecem a alguma ordem específica.
- **Vantagem:** todas as informações contidas na tabela de frequências podem ser transportadas para o gráfico.

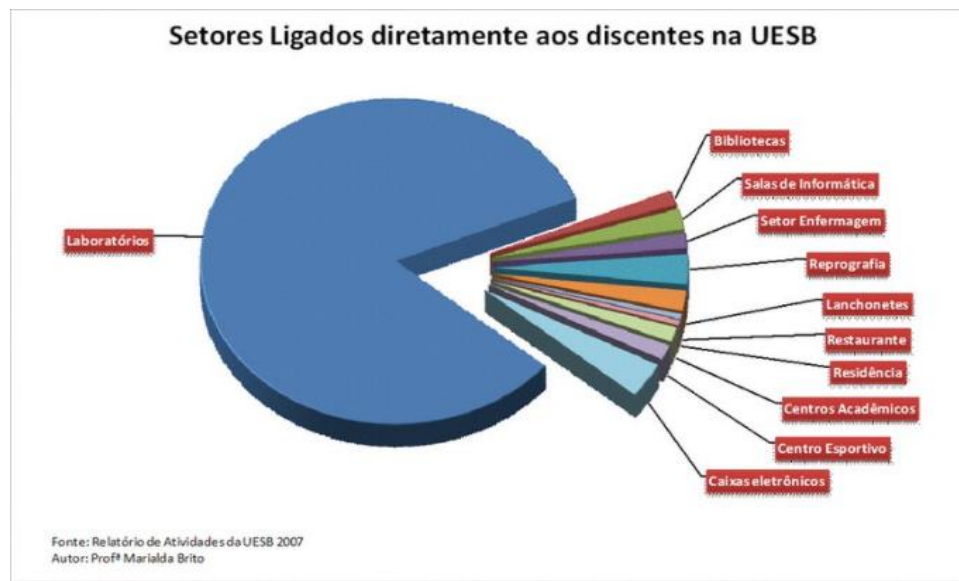
Gráfico de setores - Exemplo



■ Dentro das especificações ■ Fora das especificações

Figura 1 – Distribuição de peças produzidas por determinado setor de uma empresa de acordo com a qualidade

Gráfico de setores - Exemplo



Disponível em: <https://www.flickr.com/photos/mapasgraficos/3295827490/>

Qual(is) problema(s) vocês observam nesse gráfico?

*Uma regra importante sobre visualização de dados: não use 3D!
Não serve para nada e pode prejudicar muito, pois distorce a percepção visual
das categorias.*

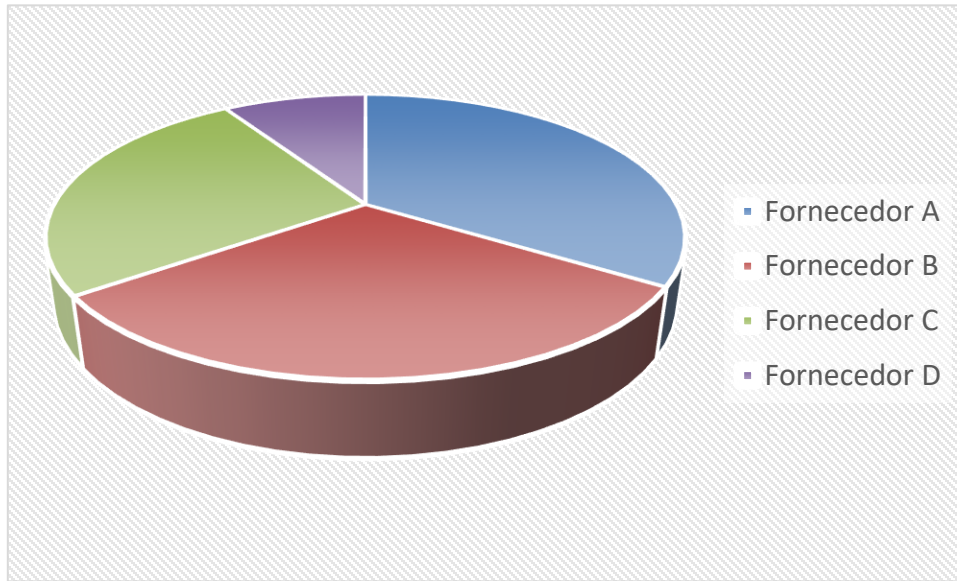
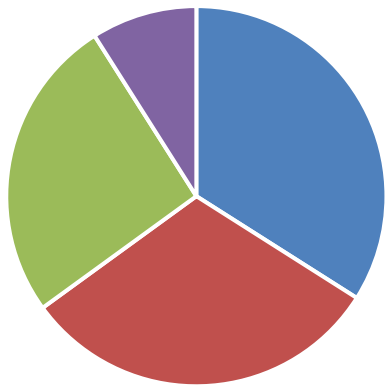
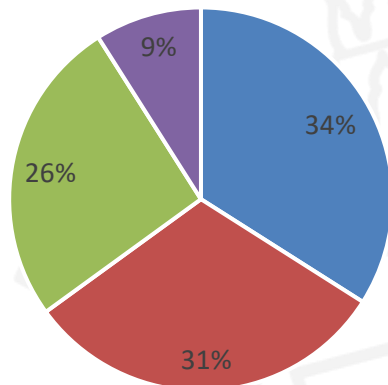


Figura 2 – Participação no mercado de fornecedores

Qual é o maior fornecedor de acordo com este visual?



■ Fornecedor A ■ Fornecedor B
■ Fornecedor C ■ Fornecedor D



■ Fornecedor A ■ Fornecedor B ■ Fornecedor C ■ Fornecedor D

Mesmo quando eliminamos o efeito 3D, não conseguimos atribuir corretamente valores quantitativos no espaço bidimensional. Quando as categorias têm tamanhos parecidos, é particularmente difícil dizer qual é a maior.

Gráfico de barras ou colunas

“As vezes os gráficos de barras são evitados por serem comuns. Isso é um erro. Os gráficos de barras devem ser aproveitados porque são comuns (...)”

(Knafllic, 2019, p.59)

Gráfico de barras ou colunas

- Utiliza o plano cartesiano com os valores ou categorias da variável no eixo das abcissas e as frequências ou porcentagens no eixo das ordenadas.
- Para cada valor ou categoria da variável desenha-se uma barra com altura correspondendo à sua frequência ou porcentagem.
- Esse tipo de gráfico se adapta bem à variáveis qualitativas (ou categóricas) ou quantitativas discretas.

Gráfico de barras ou colunas - Exemplo

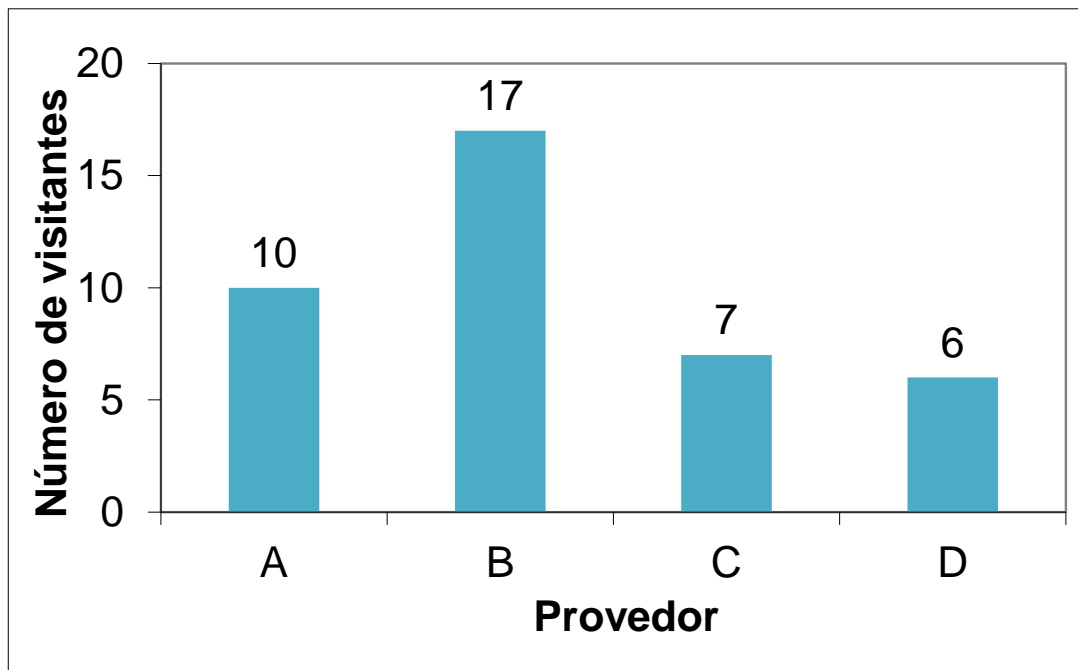


Figura 2 – Distribuição do provedor utilizado pelo visitante de um determinado *site*.

Gráfico de barras ou colunas – Exemplo 2

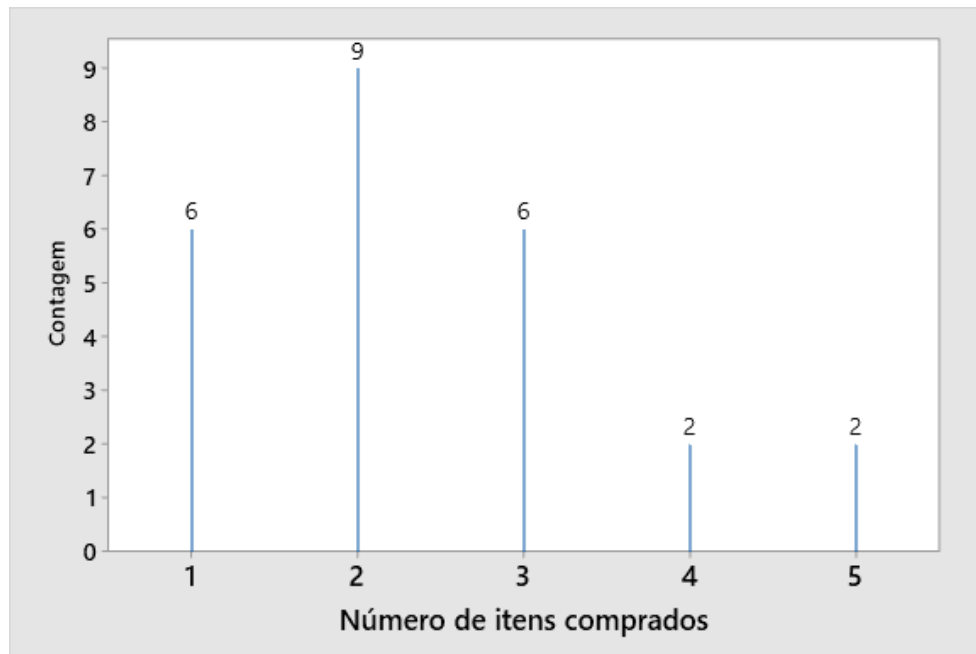


Figura 3 – Distribuição número de itens comprados por 25 clientes de uma loja.

Atenção à escala!

Quando for construir gráficos de barras o valor zero deve ser o ponto de início dos eixos, caso isto não seja feito, ocorre uma distorção no gráfico.

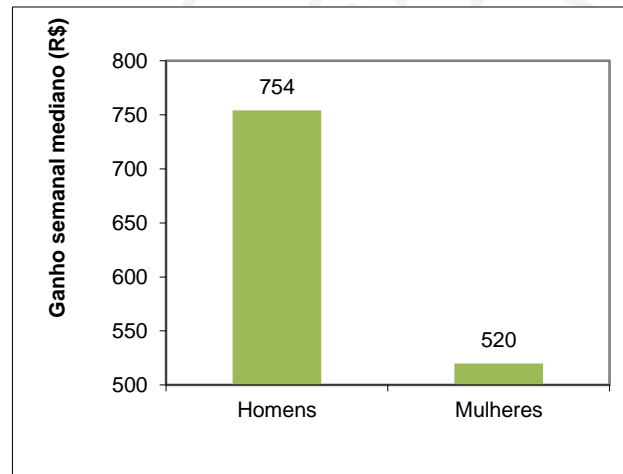
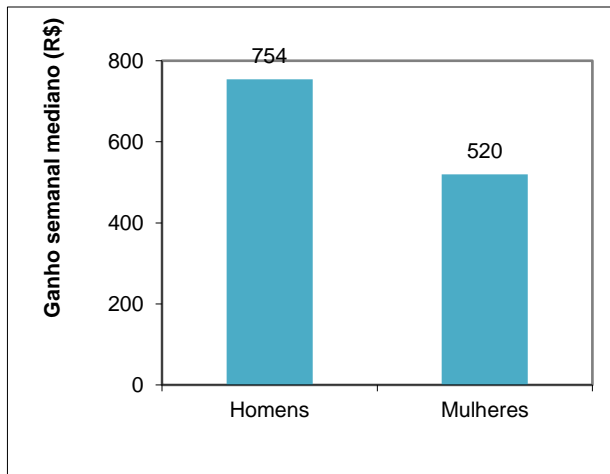
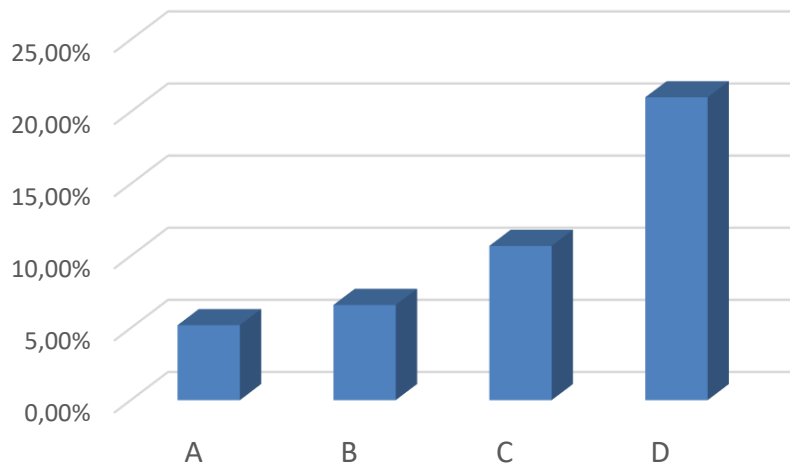


Gráfico de barras x Efeito 3D



Percebam a altura das barras em relação ao eixo y.

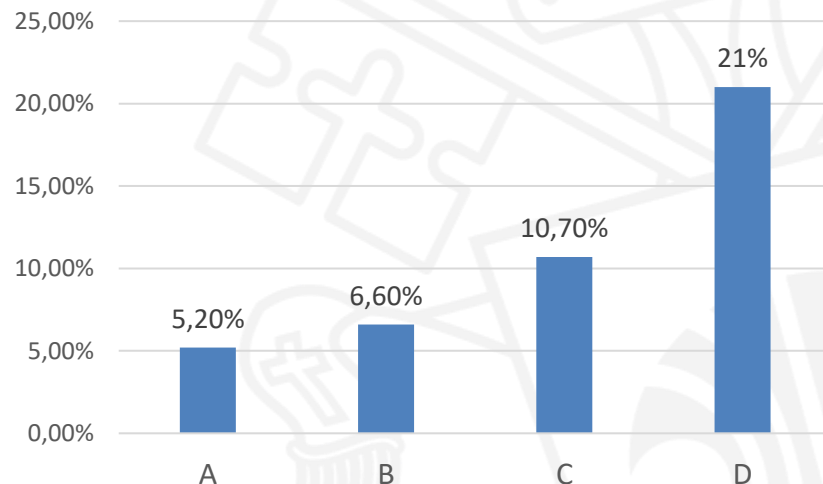
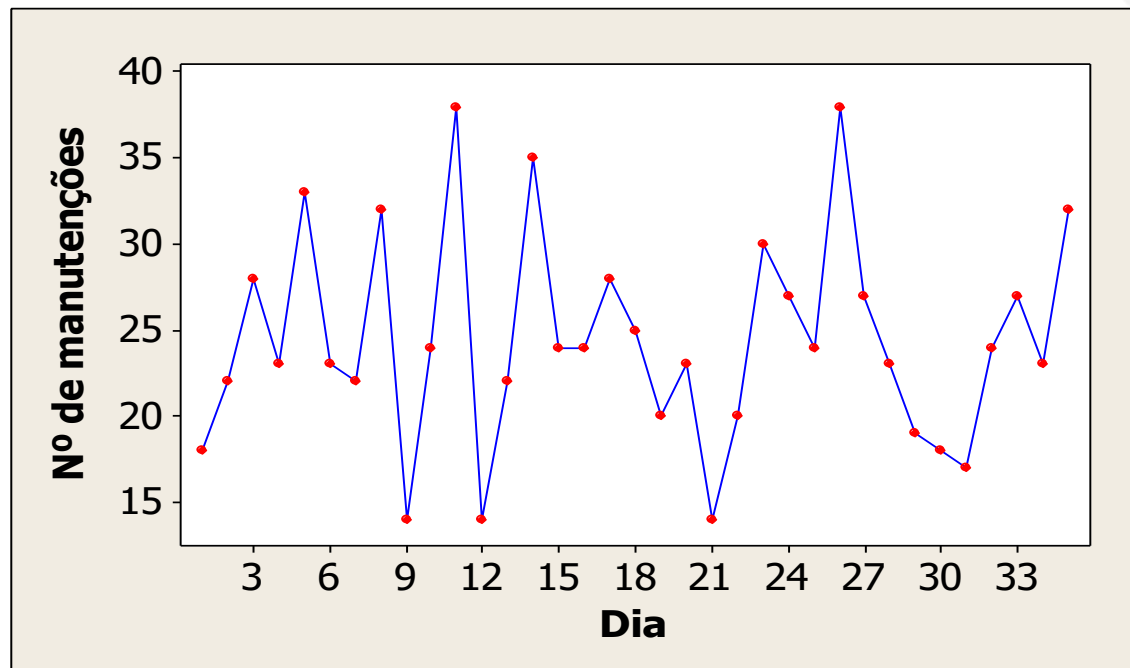


Gráfico em 2D em que as barras estão acompanhadas de seus respectivos valores.

Gráfico de linhas

- Sua construção é muito semelhante à do gráfico de barras porém utiliza-se linhas interligadas ao invés de barras para representar a frequência ou a porcentagem de cada valor ou categoria da variável estudada.
- É particularmente indicado para retratar séries temporais, ou seja, dados relativos a variáveis observadas anualmente, mensalmente, trimestralmente, de hora em hora, diariamente, etc.

Gráfico de linhas – Exemplo 1



Tendência?
Sazonalidade?

Figura 3 – Número de manutenções diárias em equipamentos durante um período de 5 semanas, em uma empresa automobilística.

Gráfico de linhas – Exemplo 2

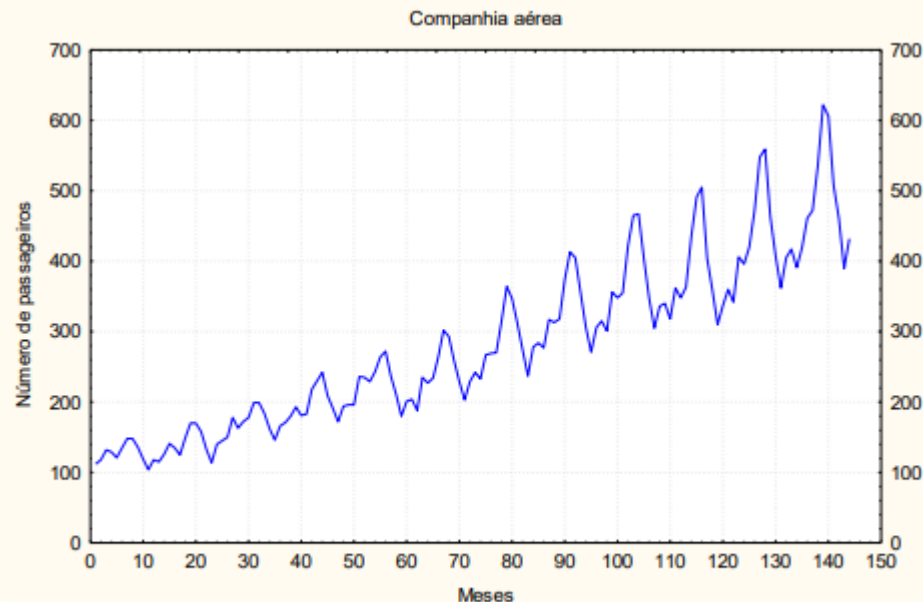


Figura 1 - Número de passageiros transportados

Tendência?
Sazonalidade?

Que padrões vocês
conseguem identificar?

Fonte: <https://www.inf.ufsc.br/~marcelo.menezes.reis/Cap4.pdf>

Gráfico de linhas – Exemplo 2

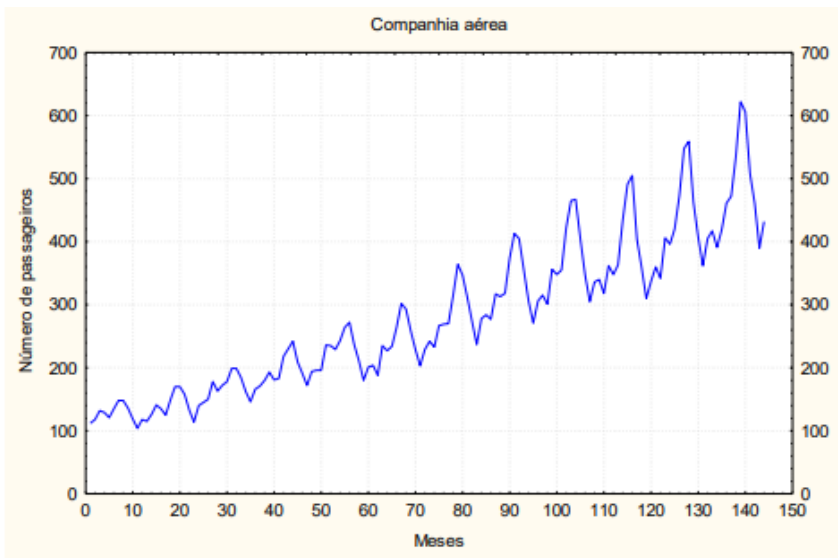


Figura 1 - Número de passageiros transportados

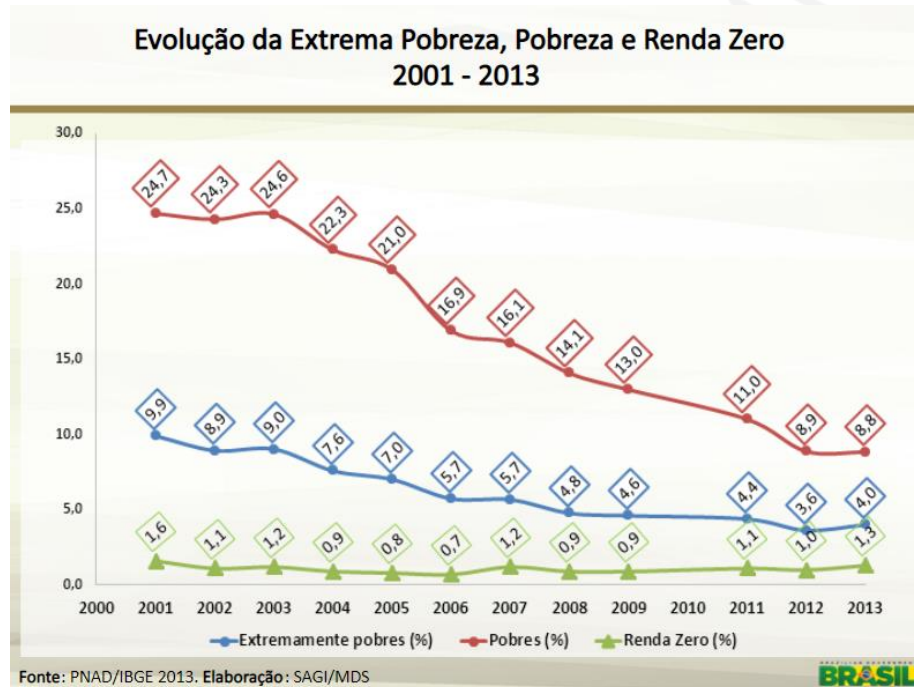
Fonte: <https://www.inf.ufsc.br/~marcelo.menezes.reis/Cap4.pdf>

Dois padrões podem ser identificados:

- 1) Há uma tendência crescente no número de passageiros transportados.
- 2) Há uma sucessão regular de "picos e vales" no número de passageiros transportados, isso deve ser causado pelas oscilações devido a feriados, períodos de férias escolares, etc., que estão geralmente relacionados às estações do ano, e que se repetem todo ano (com maior ou menor intensidade).

Tais padrões poderiam ser incorporados a um modelo estatístico, possibilitando fazer previsões que auxiliarão na tomada de decisões.

Gráfico de linhas – Exemplo 2



Histograma

- É a forma mais usual de apresentação de distribuições de frequências de variáveis quantitativas distribuídas em intervalos de valores (classes).
- É construído utilizando o plano cartesiano em que o eixo X é representado pelas faixas de valores da variável e o eixo Y representa as frequências ou as porcentagens associadas a cada uma das faixas de valores.

Histograma – Exemplo 1

Observe a tabela que deu origem ao histograma exibido na Figura 5:

Tabela 1 - Distribuição de freqüência do diâmetro (em centímetros) de peças fabricadas por uma indústria:

Diâmetro (em cm)	Freqüência	Porcentagem	Freqüência acumulada	Porcentagem acumulada
1,810 ┤ 1,822	7	14,0	7	14,0
1,822 ┤ 1,834	14	28,0	21	42,0
1,834 ┤ 1,846	18	36,0	39	78,0
1,846 ┤ 1,858	7	14,0	46	92,0
1,858 ┤ 1,870	4	8,0	50	100,0
Total	50	100,0	-	-

Histograma – Exemplo 1

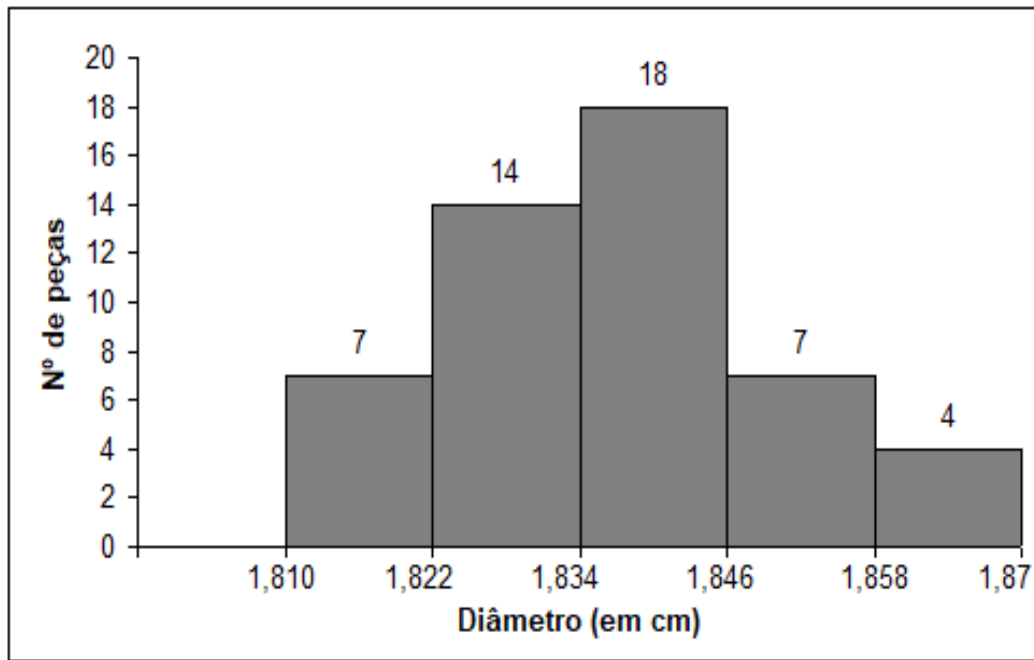


Figura 5 – Histograma para o diâmetro (em centímetros) das peças produzidas por uma indústria.

Histograma – Exemplo 2

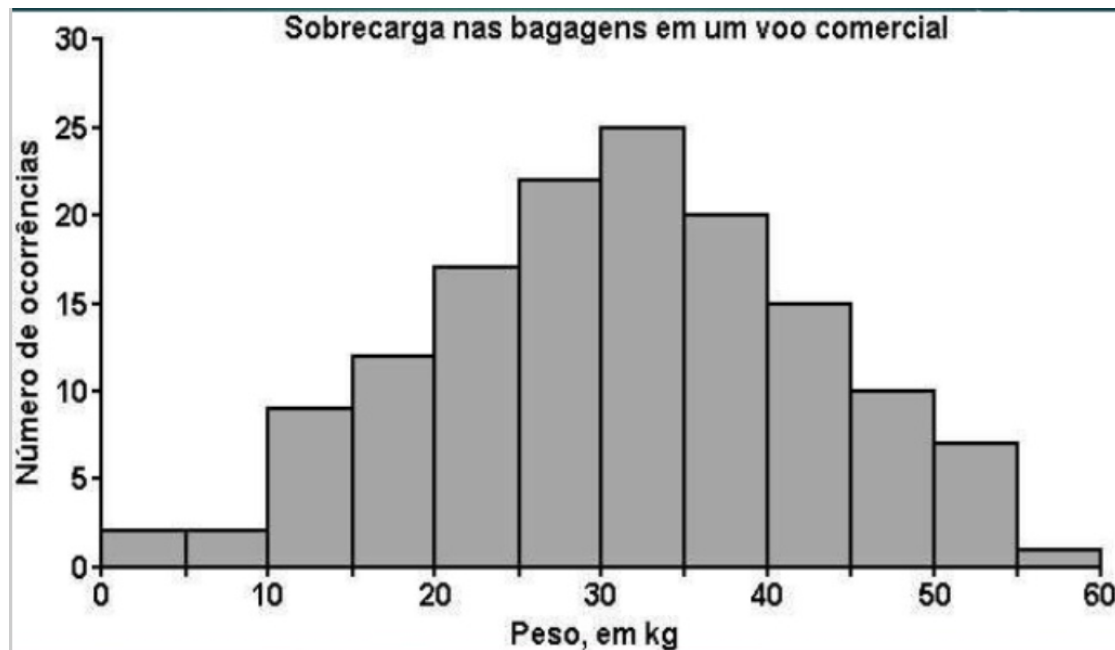


Figura 6 – Histograma para a sobrecarga (em kg) nas bagagens de um voo comercial.

Gráfico de Pontos

Quando os dados consistem em um pequeno conjunto de números, estes podem ser representados traçando-se uma reta com uma escala que abranja todas as mensurações observadas e grafando-se as respectivas frequências como pontos acima da reta.

Gráfico de Pontos - Exemplo

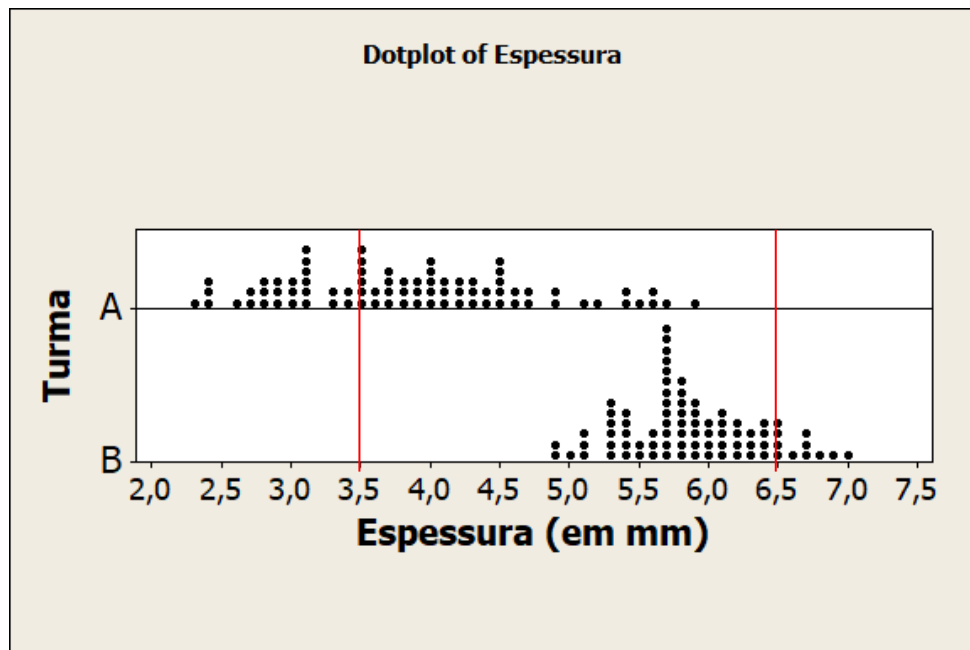


Figura 7 – Diagrama de pontos para a espessura (em mm) de azulejos produzidos por duas turmas de uma linha de produção.

Limites de especificação:
[3,5;6,5]

Boxplot, Diagrama em caixa, Gráfico de Bigodes...

O *Boxplot* é um gráfico para representar dados quantitativos sendo conveniente para revelar medidas de tendência central, dispersão, distribuição dos dados e a presença de *outliers* (*valores discrepantes*).

Boxplot ou Diagrama em caixa

A construção de um diagrama em caixa exige que tenhamos o valor mínimo, o primeiro quartil (Q_1), a mediana (ou segundo quartil Q_2), o terceiro quartil (Q_3), o valor máximo e a distância interquartílica (DIQ) \rightarrow AIQ

A distância interquartílica (DIQ) é obtida pela distância entre o terceiro e o primeiro quartil: $DIQ = Q_3 - Q_1$.

Boxplot ou Diagrama em caixa

Para identificar a presença de valores discrepantes na amostra deve-se, primeiramente, calcular as seguintes medidas:

$$Q_1 - 1,5 \times DIQ \quad \text{e} \quad Q_3 + 1,5 \times DIQ$$

Assim:

- se houver, na amostra, valores inferiores a $(Q_1 - 1,5 \times DIQ)$, eles serão considerados valores discrepantes.
- se houver, na amostra, valores superiores a $(Q_3 + 1,5 \times DIQ)$, eles serão considerados valores discrepantes.

Boxplot - Exemplo 1

Considere a seguinte amostra de idades (em anos) de estudantes universitários:

18	19	21	21	21	22
22	22	23	23	24	27

Vamos construir o boxplot para tais dados e iniciamos calculando as medidas necessárias:

Cálculo dos quartis:

- 1º quartil → P25

Posição do primeiro quartil na amostra: $L = \left(\frac{25}{100}\right) \cdot 12 = 3$

O primeiro quartil será a média dos valores que ocupam as posições 3 e 4 na amostra, ou seja, $Q_1 = 21$.

- 2º quartil → P50

Posição do segundo quartil na amostra: $L = \left(\frac{50}{100}\right) \cdot 12 = 6$

O segundo quartil será a média dos valores que ocupam as posições 6 e 7 na amostra, ou seja, $Q_2 = 22$.

- 3º quartil → P75

Posição do terceiro quartil na amostra: $L = \left(\frac{75}{100}\right) \cdot 12 = 9$

O terceiro quartil será a média dos valores que ocupam as posições 9 e 10 na amostra, ou seja, $Q_3 = 23$.

- Distância interquartílica (DIQ):

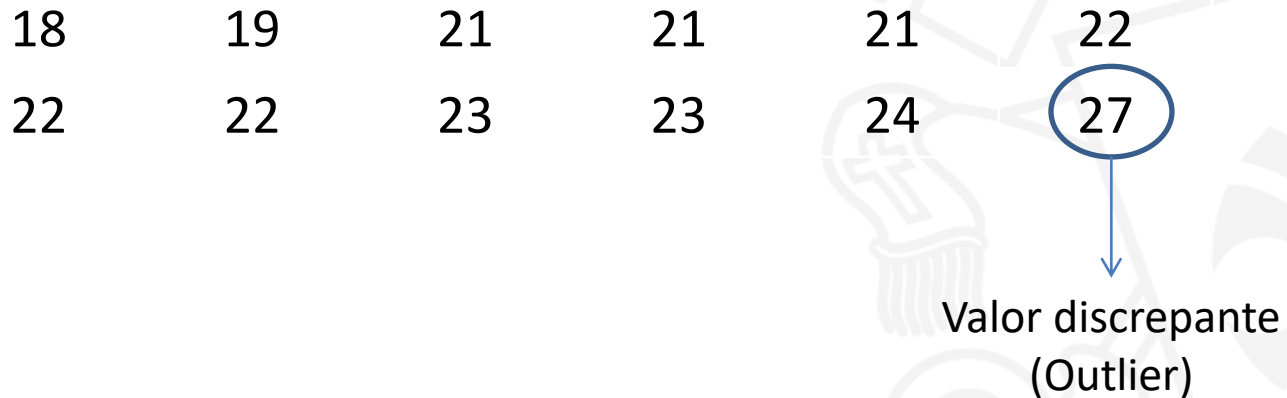
$$DIQ = Q_3 - Q_1 = 23 - 21 = 2.$$

- Limites de detecção de valores discrepantes:

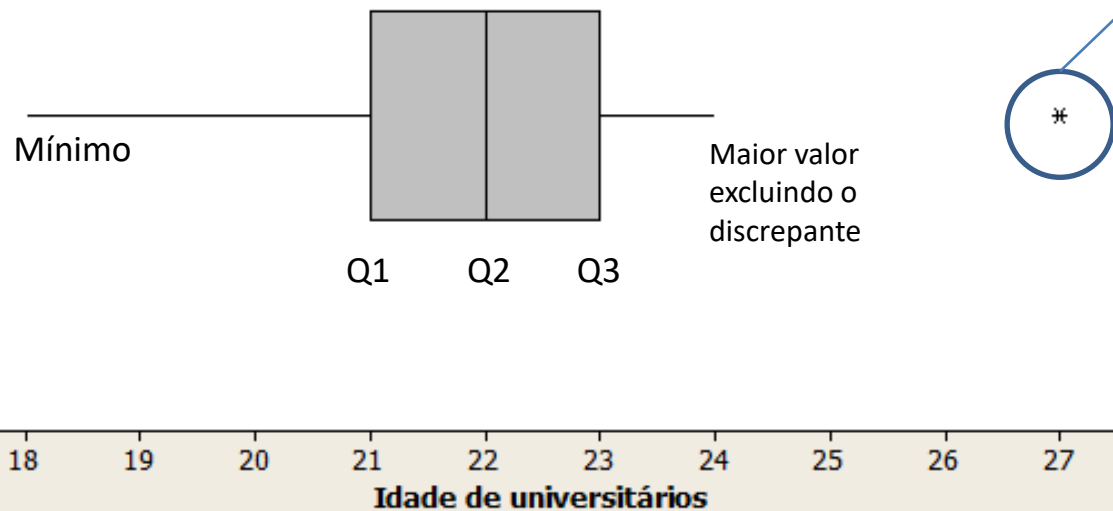
$$Q_1 - 1,5 \times DIQ = 21 - 1,5 \times 2 = 18$$

$$Q_3 + 1,5 \times DIQ = 23 + 1,5 \times 2 = 26$$

Todos os valores presentes na amostra que estejam abaixo de 18 ou acima de 26 são considerados discrepantes e virão marcados com * no boxplot.



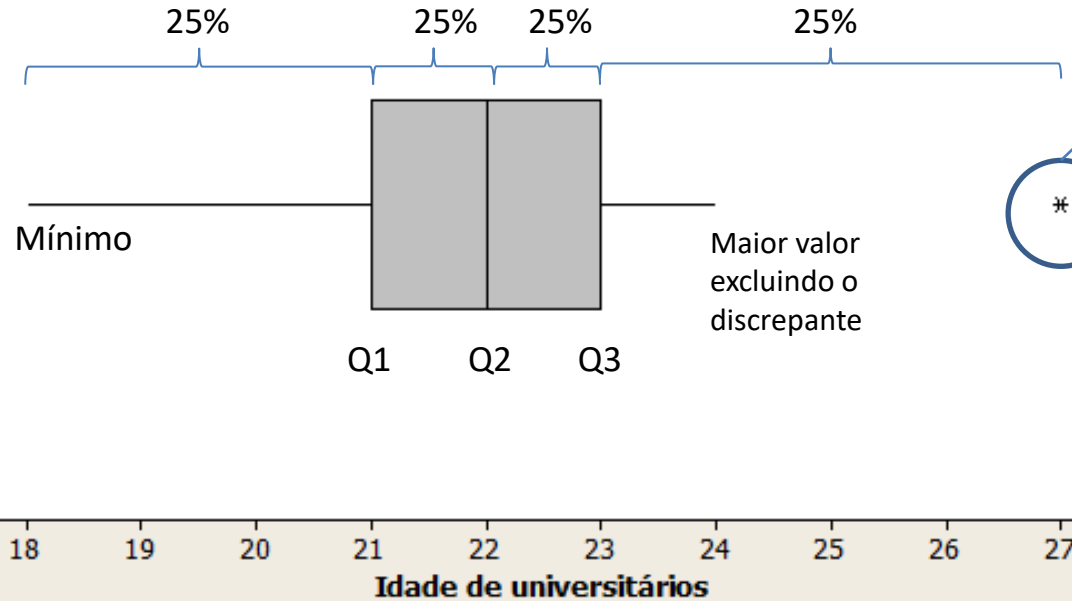
Boxplot of Idade de universitários



Valor
discrepante

Figura 9 – Boxplot para a idade (em anos) de estudantes universitários.

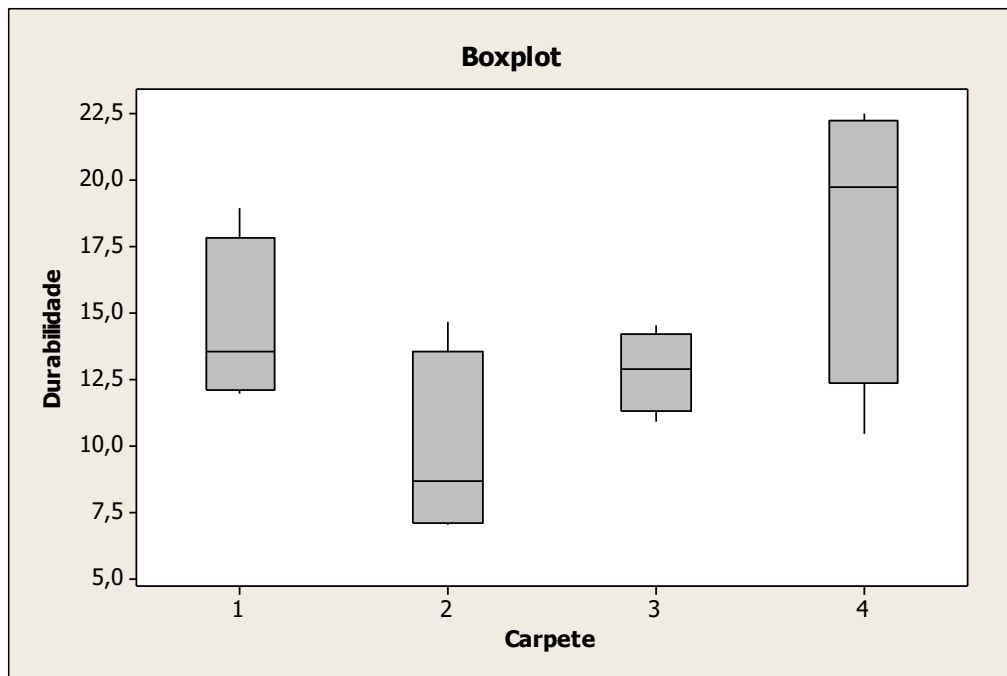
Boxplot of Idade de universitários



Valor
discrepante

Figura 9 – Boxplot para a idade (em anos) de estudantes universitários.

Boxplot - Exemplo 2



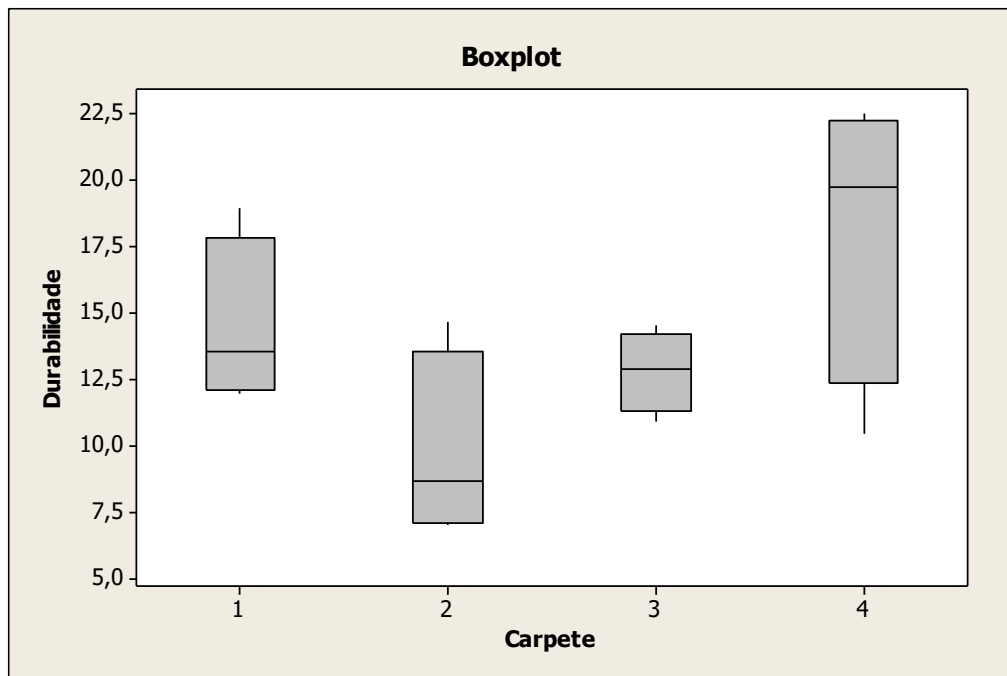
Perguntas:

Qual marca de carpete apresentou:

- Maior duração mediana?
- Maior variabilidade?
- Maior homogeneidade?

Figura 10 – Diagrama em caixa para a durabilidade de quatro marcas de carpete.

Boxplot - Exemplo 2



Perguntas:

Qual marca de carpete apresentou:

- Maior duração mediana?
Marca 4
- Maior variabilidade?
Marca 4
- Maior homogeneidade?
Marca 3

Gráfico de colunas agrupadas ou empilhadas

É utilizado quando os dados encontram-se representados em tabelas de contingência.

Gráfico de colunas agrupadas e empilhadas – Exemplo 1

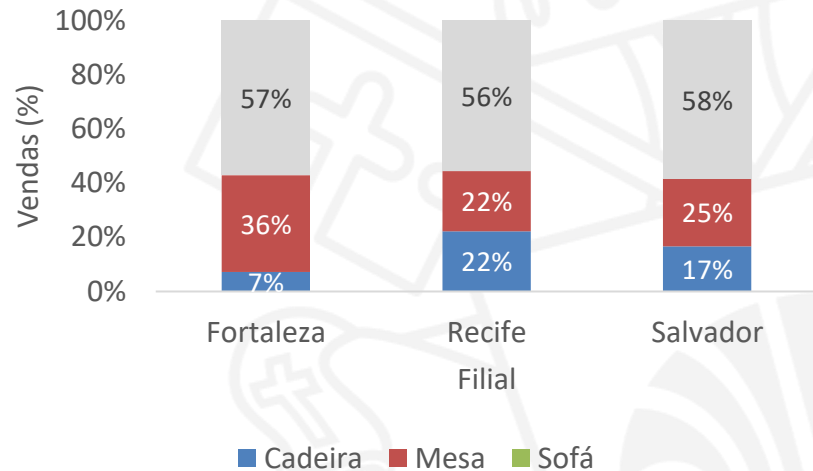
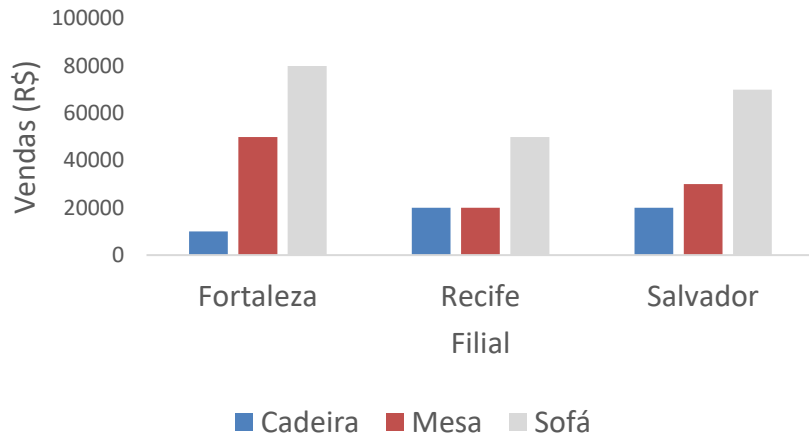
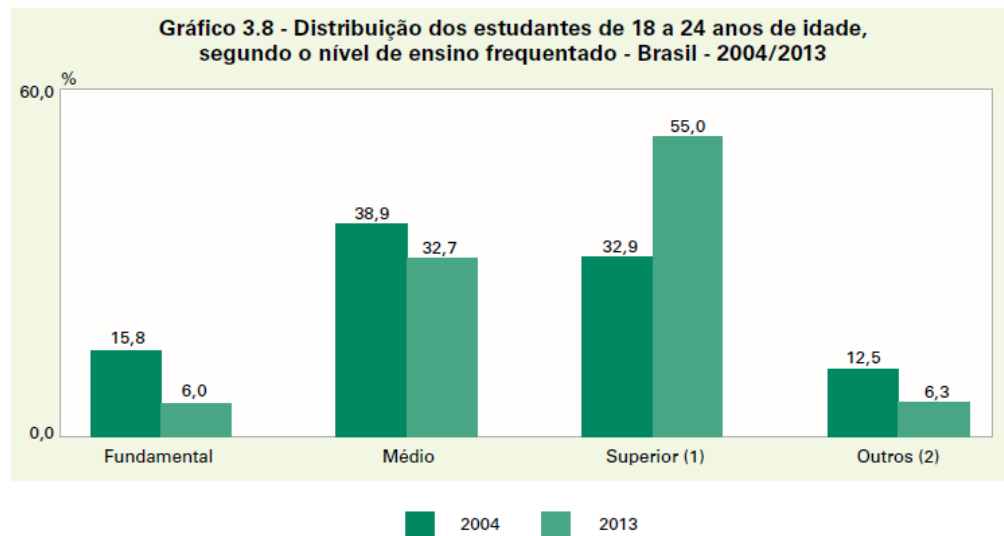


Figura 11 – Distribuição das vendas (em R\$) por filial e por produto na primeira quinzena do mês de janeiro.

Gráfico de colunas agrupadas – Exemplo 1



Fonte: IBGE, Pesquisa Nacional por Amostra de Domicílios 2004/2013.

(1) Inclusive mestrado e doutorado. (2) Pré-vestibular, supletivo e alfabetização de adultos.

Gráfico de colunas empilhadas – Exemplo 1

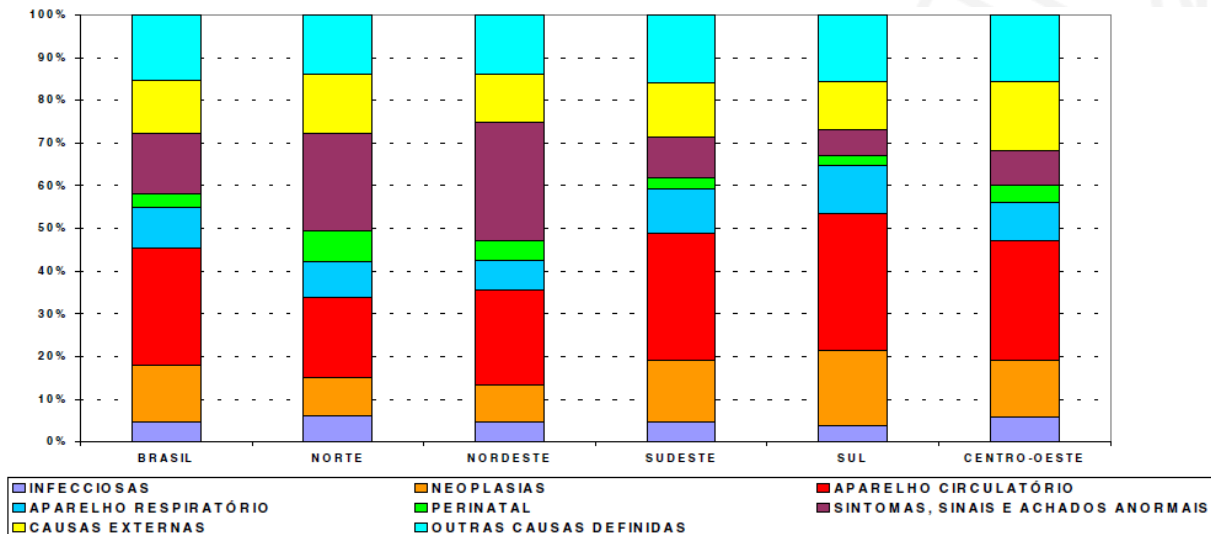


Figura 7 – Mortalidade proporcional por causas, segundo regiões, Brasil, 2001.

Fonte: DASIS/SVS – Ministério da Saúde.

Diagrama de dispersão

“Um diagrama de dispersão mostra a **relação entre duas variáveis quantitativas, medidas nos mesmos indivíduos**. Os valores de uma variável aparecem no eixo horizontal e os valores da outra variável aparecem no eixo vertical. Cada indivíduo nos dados aparece como o ponto no gráfico determinado pelos valores de ambas as variáveis para tal indivíduo”. (MOORE, 2011, p.77)

Diagrama de dispersão



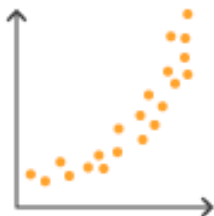
Quais semelhanças e qual a diferença você consegue identificar entre os três diagramas de dispersão?

Diagrama de dispersão

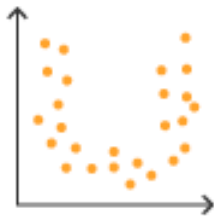


E agora?

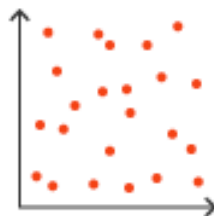
Diagrama de dispersão



Correlação não linear (Exponencial)



Correlação não linear (Quadrática)



Ausência de Correlação

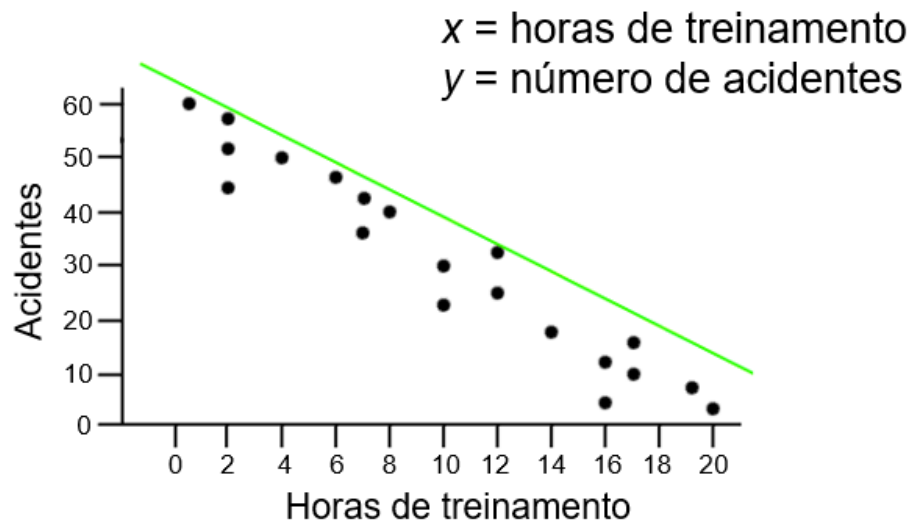
O que você pode dizer a respeito das correlações exibidas nesses três gráficos?

Diagrama de dispersão

Ao analisar um diagrama de dispersão busque as seguintes informações:

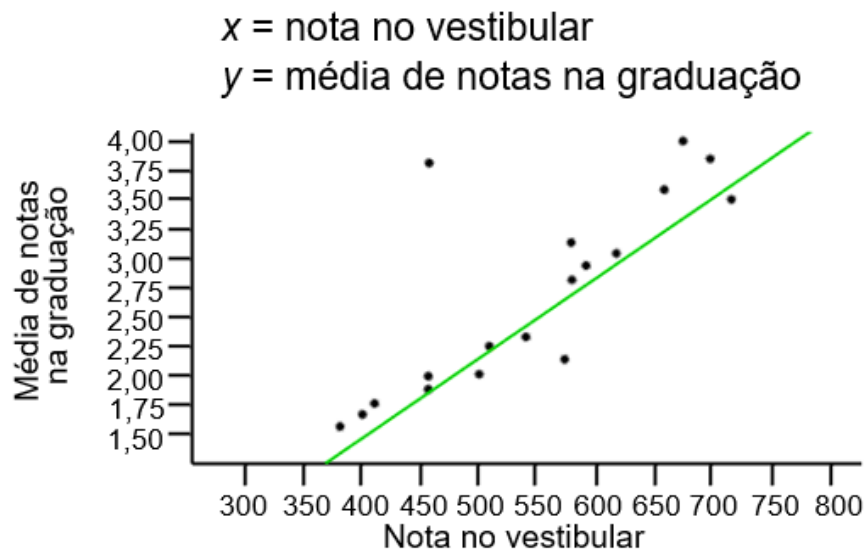
- ✓ *Existe uma correlação entre as duas variáveis?*
- ✓ *Essa correlação é linear ou não?*
- ✓ *Essa correlação é positiva ou negativa?*
- ✓ *Existem valores discrepantes (outliers)?*

Diagrama de dispersão – Exemplo 1



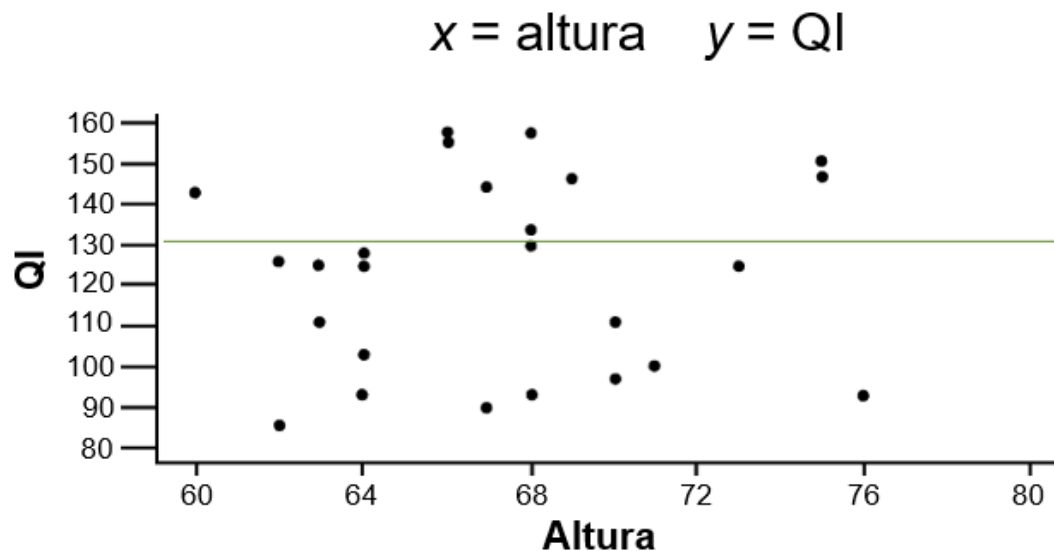
Correlação negativa: à medida
que x cresce, y decresce.

Diagrama de dispersão – Exemplo 2



Correlação positiva: à medida que x cresce, y cresce também.

Diagrama de dispersão – Exemplo 3



Não há correlação linear.

Qual é o gráfico correto para a
minha situação?



PUC Minas
Virtual