

Elementos de Estatística (EST001-B)

Lupércio F. Bessegato e Marcel T. Vieira

Exercícios de Revisão nº 1 - Análise de Dados

Exercício 1:

A pressão mínima de injeção (psi) em amostras de moldagem por injeção de milho de alta amilose foi determinada para oito amostras diferentes (pressões mais altas correspondem a maior dificuldade de processamento), resultando nas observações a seguir:

15,0 13,0 18,0 14,5 12,0 11,0 8,9 8,0

- Determine os valores da média amostral e da mediana amostral. Compare-os.
- Em quanto a menor observação da amostra (atualmente 8,0) pode ser aumentada sem afetar o valor da mediana amostral?
- Suponha que desejemos que os valores da média e da mediana amostrais sejam expressos em quilo Pascal (kPa) em vez de psi. É necessário mudar as unidades de cada observação ou os valores calculados na parte (a) podem ser usados diretamente? Dica: 1 kPa \approx 6,89 psi.
- Quais são os valores dos quartis e qual é o valor da amplitude interquartílica?
- Faça um resumo dos cinco números (mínimo, máximo, mediana, 1º e 3º quartis).
- Construa um *boxplot* com base no resumo dos cinco números e comente suas características.
- Quão grande ou pequena deve ser uma observação para se qualificar como um *outlier*?

Exercício 2:

O artigo “*A thin-Film Oxygen Uptake Test for the Evaluation of Automotive Crankcase Lubricants*” (Lubric. Engr., 1984, p. 75-83) informou os seguintes dados sobre tempo de oxidação-indução (min) para diversos óleos comerciais:

87 103 130 160 180 195 132 145 211 105 145
153 152 138 87 99 93 119 129

- Calcule $\sum x_i$ e $\sum x_i^2$.
- Use os valores calculados em (a) para obter a variância amostral e o desvio padrão amostral.
- Se as observações fossem especificadas em horas, quais seriam os valores resultantes para a variância e para o desvio padrão amostrais? Qual a relação com os resultados obtidos em (a)?
- Determine o primeiro, o segundo e o terceiro quartis.
- Calcule a amplitude interquartílica.
- Se os dois maiores valores da amostra, 4,33 e 4,52, fossem 5,33 e 5,52, como a amplitude interquartil seria afetada? Explique.

Exercício 3:

O artigo “*A thin-Film Oxygen Uptake Test for the Evaluation of Automotive Crankcase Lubricants*” (Lubric. Engr., 1984, p. 75-83) informou os seguintes dados sobre tempo de oxidação-indução (min) para diversos óleos comerciais:

87 103 130 160 180 195 132 145 211 105 145
153 152 138 87 99 93 119 129

Elementos de Estatística (EST001-B)

Lupércio F. Bessegato e Marcel T. Vieira

- Calcule $\sum x_i$ e $\sum x_i^2$.
- Use os valores calculados em(a) para obter a variância amostral e o desvio padrão amostral.
- Se as observações fossem especificadas em horas, quais seriam os valores resultantes para a variância e para o desvio padrão amostrais? Qual a relação com os resultados obtidos em (a)?
- Determine o primeiro, o segundo e o terceiro quartis.
- Calcule a amplitude interquartílica.
- Se os dois maiores valores da amostra, 4,33 e 4,52, fossem 5,33 e 5,52, como a amplitude interquartil seria afetada? Explique.

Exercício 4:

Costumamos recorrer às séries temporais para ver o efeito de uma mudança social ou de uma nova política. Deseja-se analisar os dados sobre mortes causadas por veículos motorizados nos Estados Unidos, no período de 1960 à 1994. Como número de mortes tende a aumentar na medida em que os motoristas percorrem maior quilometragem, recorreremos à taxa de mortalidade, que é o número de mortes por 100 milhões de quilômetros percorridos.

- Analise o gráfico temporal dessas taxas de mortalidade (Figura 1). Descreva o padrão geral dos dados.
- Em 1974, o limite nacional de velocidade baixou para 55 milhas/hora, em uma tentativa de economizar gasolina após a guerra de 1973 no Oriente Médio. Em meados da década de 80, a maioria dos estados elevou para 65 milhas/hora o limite nas rodovias interestaduais. Alguns alegaram que o limite menor contribuiu para salvar vidas. O efeito dessa redução do limite de velocidade entre 1974 e meados de 1980 é visível em seu gráfico?
- Tem sentido fazer um histograma dessas 18 taxas de mortalidade?
- A impressão que um gráfico temporal dá depende das escalas usadas nos dois eixos. Analise a Figura 2 e comente sobre diferenças ou similaridades entre os dois gráficos. Qual sua conclusão?

Exercício 5:

Os peixes-boi são animais marinhos, grandes e dóceis, que vivem ao largo da costa de uma localidade distante. Muitos deles são mortos ou feridos por barcos a motor. São analisados dados sobre os registros de barco a motor (em milhares) e o número de peixes-boi mortos por esses barcos nessa localidade, nos anos de 1977 a 1990.

- A Figura 3 apresenta um diagrama de dispersão desses dados. Descreva a forma e a direção da relação (é razoável descrever o padrão global por uma linha reta? a associação é positiva ou negativa?).
- O quadrado da correlação (r^2) é a fração da variação nos valores da resposta (y) que é explicada pela regressão de mínimos quadrados de y em x . Verifica-se que as variáveis: barcos a motor registrados e peixes-boi mortos apresentam uma correlação $r = 0,941$. Que fração da variação nas mortes dos peixes-boi pode ser explicada pelo número de barcos registrados? Pode-se prever com exatidão o número de peixes-boi mortos a partir do número de registros de barcos?
- A reta de regressão dos dados é

The regression equation is Peixes mortos = - 41,4 + 0,125 Barcos.
--

Elementos de Estatística (EST001-B)

Lupércio F. Bessegato e Marcel T. Vieira

Preveja o número de peixes-boi que serão mortos em um ano, quando houver 716.000 barcos a motor registrados.

- d. Suponha que, em um futuro distante, haja 2 milhões de barcos registrados na localidade em questão. Com o auxílio da reta de regressão, preveja o número de peixes-boi mortos. Você julga que essa predição é confiável? Por quê?
- e. Eis mais quatro anos de dados sobre os peixes-boi, no período de 1991 a 1994:

1991	716	53	1993	716	35
1992	716	38	1994	735	49

Esses pontos foram acrescentados no diagrama de dispersão. Sabe-se que o governo local tomou medidas mais severas para proteger os peixes-boi durante esses anos. Analise o novo diagrama de dispersão, apresentado na Figura 4. Você vê alguma evidência de resultados satisfatórios dessas medidas? Por quê?

- f. No item (c) fizemos uma previsão das mortes de peixes-boi em um ano com 716.000 barcos a motor registrados. Na verdade os registros de barcos a motor permaneceram durante os próximos três anos. Compare a média das mortes de peixes-boi nesses anos com sua previsão no item (c). Qual o erro de sua previsão?

Exercício 6:

As partículas trazidas pelo ar, como poeira e fuligem, são parte importante na poluição do ar. Para medir a poluição por essas partículas, um motor a vácuo extrai ar através de um filtro durante 24 horas. Pesa o filtro no começo e no fim do período. O aumento de peso é uma medida da concentração de partículas no ar. Em um estudo de poluição do ar, tomaram-se medidas a cada 6 dias, com instrumentos idênticos, no centro de uma pequena cidade e em uma localidade rural situada a 16 km a sudoeste da cidade. Como os ventos costumam soprar do oeste, suspeita-se que os resultados rurais seja, em geral, inferiores aos da cidade, mas que os resultados da cidade possam ser preditos com base nos resultados rurais. Foram efetuadas leituras nos níveis de partículas (em gramas), nas duas localidades, a cada seis dias, durante um período de sete meses.

- a. Queremos usar o nível rural de partículas para predizer o nível na cidade no mesmo dia. Quais é a variável explicativa e a variável resposta? A Figura 5 apresenta um diagrama de dispersão da relação entre esses dados. O gráfico sugere que a aplicação da reta de regressão de mínimos quadrados para a predição dará resultados aproximadamente corretos no intervalo de valores dos dados levantados?
- b. O coeficiente de correlação dos dados é $r = 0,975$. Qual a percentagem da variação observada nos níveis de poluição na cidade que é explicada por esse relacionamento linear?
- c. A reta de regressão dos dados é

The regression equation is Cidade = - 2,58 + 1,09 Rural.

Interprete a inclinação? Qual é sua unidade de medida? É razoável interpretar o intercepto no contexto da questão?

- d. No décimo quarto dia da série, a leitura rural foi 88, mas não se dispunha do resultado para a cidade. Qual é sua estimativa para esse resultado naquela data?

Elementos de Estatística (EST001-B)

Lupércio F. Bessegato e Marcel T. Vieira

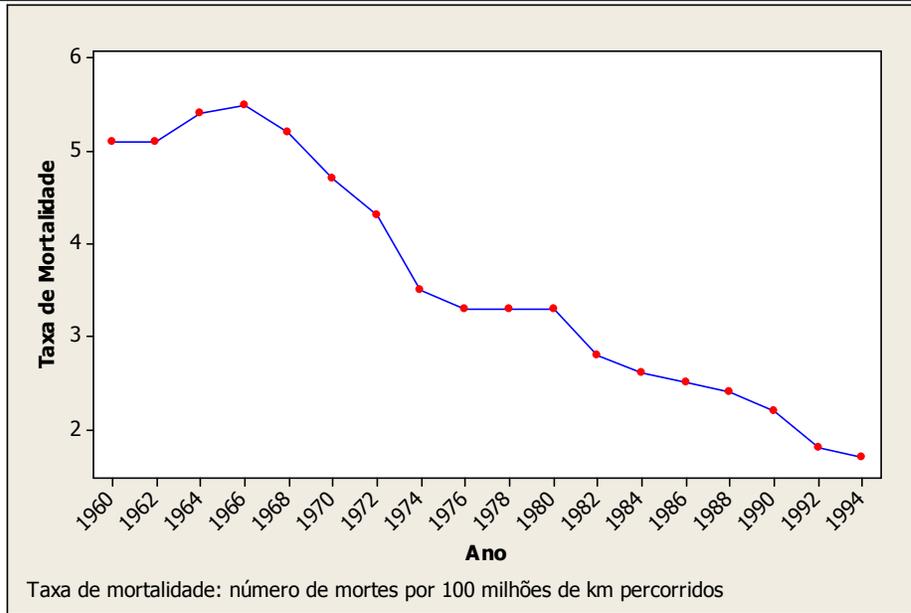


Figura 1 - Gráfico temporal da taxa de mortalidade - 1960/1994.

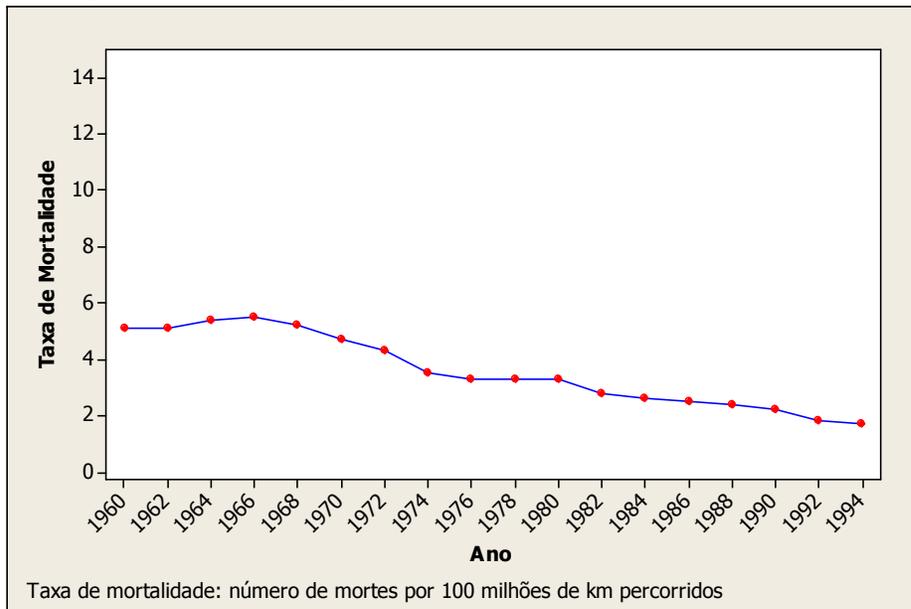


Figura 2 - Gráfico temporal da taxa de mortalidade - 1960/1994.

Elementos de Estatística (EST001-B)

Lupércio F. Bessegato e Marcel T. Vieira

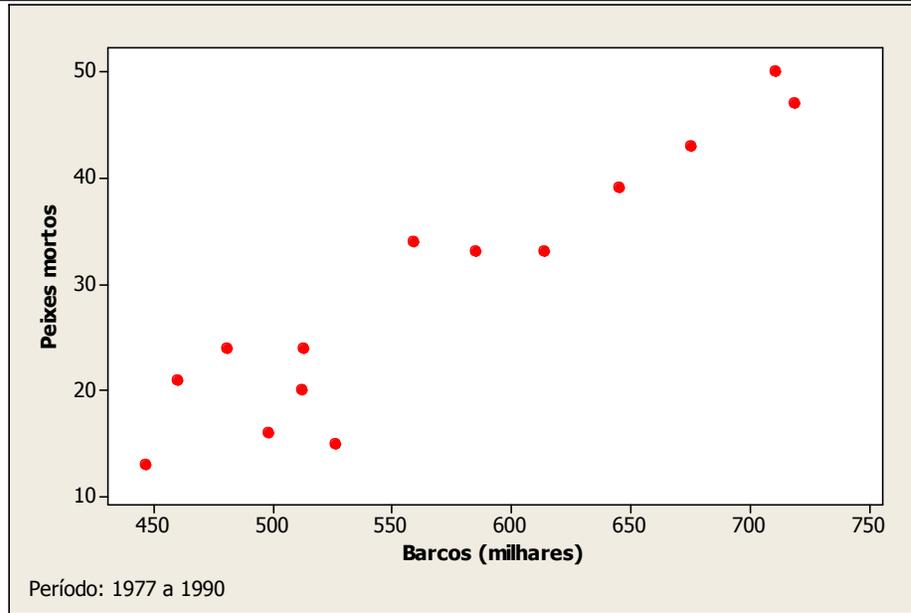


Figura 3 - Diagrama de dispersão de barcos a motor vs. peixes-boi mortos - 1977/1990.

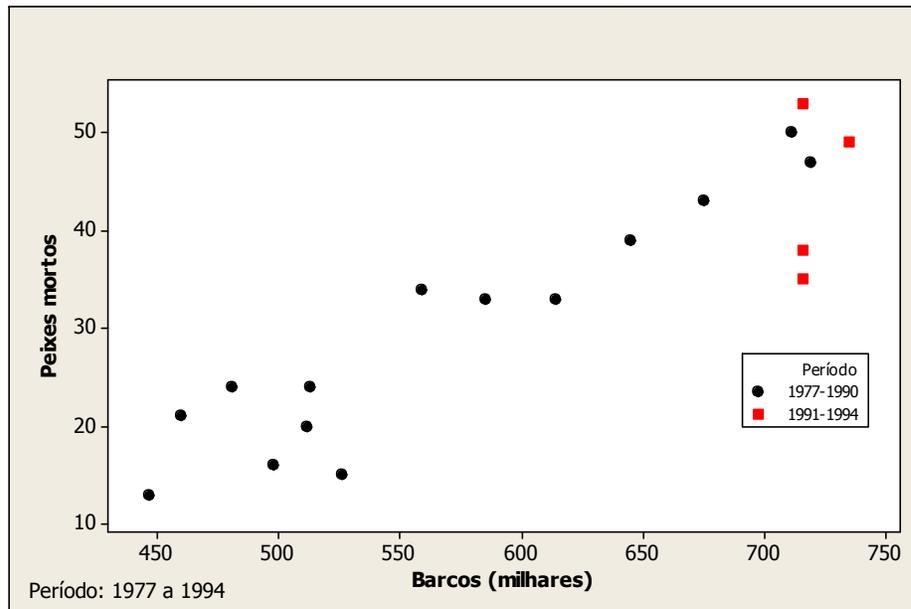


Figura 4 -Diagrama de dispersão de barcos a motor vs. peixes-boi mortos - 1977/1994.

Elementos de Estatística (EST001-B)

Lupércio F. Bessegato e Marcel T. Vieira

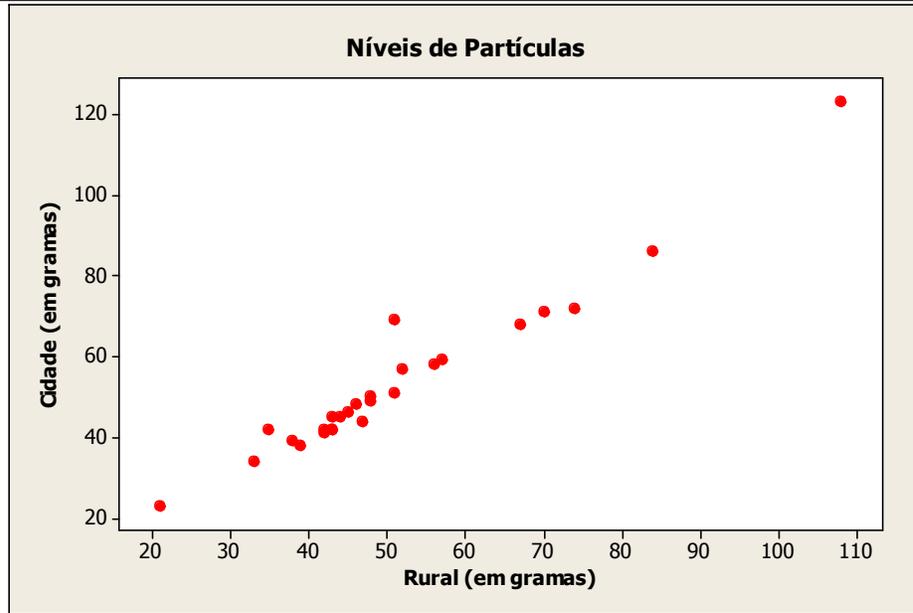


Figura 5 - Níveis de partículas (gramas) em duas localidades vizinhas.