

Noções de Probabilidade e Estatística

Marcos N. Magalhães e Antonio Carlos P. Lima

Modificações da 7a. edição (2010)

As alterações feitas na **7a. edição (2010)** em relação à **6a. edição- 3a. reimpressão (2008)** são indicadas neste documento. As principais modificações dessa edição foram feitas nos Capítulos 1 e 6, com a introdução de novos conceitos e exemplos. No Capítulo 1 foram acrescentadas 12 páginas e, no Capítulo 6, 4 páginas. Dessa forma, os Capítulos 2, 3, 4 e 5 tiveram a numeração das páginas aumentada de 12 e, nos Capítulos 7, 8 e 9, o aumento foi de 16.

Capítulo 1

Como houve modificação na paginação, estamos indicando de forma geral as alterações:

- Após a Figura 1.6, fazemos uma discussão sobre o uso da frequência absoluta ao invés da densidade de frequência em histogramas.
- O conceito de *box-plot* foi modificado e é descrito no final desse arquivo.
- Introduzimos o conceito de assimetria e uma forma de quantificá-la, o coeficiente de Bowley.
- Introduzimos o gráfico *Q-Q plot* para a comparação de conjuntos de dados

Alguns dos antigos exemplos foram modificados e seis novos foram incluídos.

Modificações em exercícios do Capítulo 1:

Seção 1.2 (antiga página 20)

Exercício 3 - novo item:

- c. Construa o *box-plot* utilizando os dados brutos.

Exercício 4 - redação alterada do item (c):

- c. Usando as informações disponíveis, obtenha o *box-plot*. Indique as suposições feitas.

Exercício 5 - redação alterada e inclusão de item:

- c. Com base nos histogramas obtidos em (b) e em suposições adequadas, obtenha as medidas necessárias para construir um *box-plot* para o peso de adultos em cada região. Usando essa representação gráfica, faça uma rediscussão do item (a).
- d. Utilizando a mediana e os quartis obtidos em (c), determine o coeficiente de assimetria de Bowley para o peso de adultos em cada região. Os valores desses coeficientes estão de acordo com os padrões observados em (b) e (c)?

(Novo exercício) 6. Em um estudo clínico dois anestésicos estão sendo avaliados. Cada um dos medicamentos *Alpha* e *Beta* foram aplicados em grupos diferentes de 18 cobaias, sendo que todas elas tinham aproximadamente as mesmas características físicas, incluindo peso e idade. O tempo para o completo efeito dos anestésicos foi medido, em segundos, para cada animal. Os resultados, após ordenação, são apresentados a seguir.

Medicamento *Alpha*: 24, 24, 24, 25, 25, 26, 26, 27, 28, 29, 30, 30, 30, 31, 31, 32, 32 e 33.

Medicamento *Beta*: 19, 19, 19, 20, 22, 25, 26, 26, 27, 29, 29, 31, 34, 34, 37, 40, 41 e 42.

- a. Construa um *box-plot* para o tempo de efeito de cada anestésico e comente as diferenças encontradas.
- b. Para o tempo de efeito de cada medicamento, determine o coeficiente de assimetria de Bowley e comente os resultados.
- c. Compare o desempenho dos anestésicos usando um *Q-Q plot*.

Seção 1.4 (antiga página 23)

Exercício 9 - redação alterada do item (c):

- c. Determine os 1º, 2º e 3º quartis.

Exercício 10 - redação alterada do item (d):

- d. Construa o *box-plot* a partir da tabela de dados brutos.

Exercício 12 - novo item:

- e. Usando a tabela de dados brutos construa o *box-plot*.

Exercício 13 - novo item:

- d.** Construa o *box-plot* utilizando os dados brutos.
- Exercício 14 - redação alterada do item (d):
- d.** Utilizando o histograma, que porcentagem dos clubes marcaram mais de 38 gols?
- Exercício 15 - redação alterada do item (b):
- b.** Fazendo suposições adequadas, obtenha o *box-plot*.
- Exercício 16 - redação alterada do item (b):
- b.** Com as devidas suposições, construa o *box-plot*.
- Exercício 20 - alteração e novo item:
- d.** Faça um *Q-Q plot* comparando as notas de português e matemática.
- e.** Comente sobre a afirmação: os aprovados são melhores em português do que em matemática.
- Exercício 24 - alteração e novo item:
- b.** Médicos afirmam que o grupo dos falso-positivos é mais jovem do que o dos falso-negativos. O que você diria a respeito? Justifique sua resposta baseando-se em gráficos e tabelas de frequência.
- c.** Compare, com auxílio do *Q-Q plot*, a glicose dos pacientes falso-positivos com a dos falso-negativos.
- Exercício 25 - alteração e novo item:
- e.** Considere a variável Total e compare os blocos usando um *Q-Q plot*.
- f.** Baseando-se nos itens anteriores, você diria que existem diferenças nas áreas dos apartamentos dos blocos A e B? Em caso positivo, qual(is) cômodo(s) apresenta(m) o problema?
- g.** Explore descritivamente os dados referentes a problemas estruturais (rachaduras e infiltrações). Com a informação contida na variável Andar divida os apartamentos em três categorias dependendo do andar onde se encontra: baixo, médio e alto. Estude a ocorrência de rachaduras e infiltrações para cada categoria.
- Exercício 26 - novo item:
- e.** Compare, usando um *Q-Q plot*, o comportamento da variável Temposp nos bairros Jardim Raposo e Jardim d'Abril.

Capítulo 6

Neste capítulo, introduzimos, na Seção 6.2, os conceitos de *gráfico de probabilidade* e de *gráfico de probabilidade Normal*. Dois novos exemplos foram também incluídos.

Modificações em exercícios do Capítulo 6:

Seção 6.2 (antiga página 193)

Novo exercício

- 10.** Para a variável Peso na Tabela 1.1, use gráficos de probabilidade para discutir a conveniência do ajuste dos modelos $U[40, 100]$ e $N(60, 100)$ aos dados.

Seção 6.3 (antiga página 195)

Exercício 35 - novo item:

- c.** Para os dados de cada bloco, calcule a média e o desvio padrão para a variável Sala. Em seguida, use esses valores para construir gráficos de probabilidade Normal. Comente.C

Exercício 36 - alteração de redação no item (b):

- b.** Construa um gráfico de probabilidade Normal supondo um modelo $N(19,140)$. Você acha que essa distribuição é adequada para esses dados?

Exercício 37 - alteração de redação nos itens (b) e (c):

- b.** Para cada região de procedência, construa um histograma para a variável Temposp. Compare os gráficos. Existe diferença entre eles?
- c.** Foi proposta a Uniforme contínua $[1, 10]$ para modelar a variável Resid. Qual é a sua opinião, com base em um gráfico de probabilidade?

Capítulo 7

Mudança de redação no Exercício 4 da Seção 7.4:

4. Uma amostra em 100 cidades brasileiras, de até 20 mil habitantes, indicou que o valor médio da hora-aula para os professores do ensino fundamental em escolas municipais é de R\$ 10. Obtenha um intervalo de confiança para o valor médio nacional da hora aula em cidades do tipo mencionado. Baseado em estudos anteriores, o desvio padrão é assumido ser igual a R\$ 2. Use $\gamma = 0,95$ como coeficiente de confiança.

Capítulo 8

Mudança de redação no item (c) do Exercício 42 da Seção 8.6:

- c. Construa uma nova variável com a informação da área total (isto é, obtenha uma variável que é a soma das áreas de cada cômodo) e faça gráficos de probabilidade Normal considerando cada bloco. O modelo Normal lhe parece adequado para descrever o comportamento dessa nova variável?

Capítulo 9

Os gráficos *box-plot* foram contruídos usando a nova forma. Alteração de redação nos itens (a) e (b) do Exercício 29 da Seção 9.6:

- a. Dependendo de cada uma das três categorias de andar criadas, construa um gráfico *box-plot* para as áreas da sala e discuta se a localização interfere no tamanho da sala.
b. Construa gráficos de probabilidade Normal para as medidas de área da sala, para cada um dos grupos formados (andar baixo, intermediário e alto). Discuta se o modelo Normal é adequado.

Apêndice B- Respostas dos Exercícios

Lembrando que gráficos e respostas de exercícios pares de fim de capítulo não são apresentados, indicamos abaixo as respostas aos itens modificados e/ou incluídos.

Seção 1.2:

- 3c. mín. = 1,50; $Q_1 = 3,40$; $Q_2 = 6,30$; $Q_3 = 8,95$; máx. = 10,7. Sem extremos.
4c. Supor mín. = 0 e máx. = 10. Do histograma, $Q_1 = 2,48$; $Q_2 = 4,00$ e $Q_3 = 5,56$. Sem extremos.
5c. Supor as observações espalhadas de modo uniforme nas faixas. Região A: mín. = 20, $Q_1 = 45,00$; $Q_2 = 52,86$; $Q_3 = 60,40$ e máx. = 90. O intervalo de corte é [29,76; 83,50] e precisamos avaliar a existência de pontos extremos. Na primeira faixa vão 8 observações em intervalo de 2,5; começando em 20 e terminando em 37,5 (40 não é da faixa). Na última faixa, colocamos 9 observações, a cada 2,5 iniciando em 70 e terminando em 90. Os bigodes inferior (27,50) e superior (82,50) determinam 4 pontos extremos na parte inferior e 3 na parte superior. Região B: mínimo = 40, $Q_1 = 73,25$; $Q_2 = 80,57$; $Q_3 = 87,76$ e máximo = 110. Proceda como na Região A e obtenha os bigodes inferior e superior respectivamente iguais a 52,00 e 109,26; teremos 6 pontos extremos na parte inferior e 2 na parte superior. A construção dos *box-plots* confirma o mencionado no item (a). A Região B tem medidas superiores às da Região A.
5d. $g_b(A) = -0,021$; $g_b(B) = -0,009$. Baixa assimetria de acordo com os gráficos obtidos.
6a. *Box-plot* sem valores extremos mas com diferentes variabilidades. Alpha: mín. = 24,0; $Q_1 = 25,0$; $Q_2 = 28,5$; $Q_3 = 31,0$; máx. = 33,0; Beta: mín. = 19,0; $Q_1 = 22,0$; $Q_2 = 28,0$; $Q_3 = 34,0$; máx. = 42,0.
6b. $g_b(Alpha) = -0,167$; $g_b(Beta) = 0,000$. Alpha ligeiramente assimétrico.
6c. O gráfico dos pares ordenados sugere comportamento diferente nas caudas.

Seção 1.4:

- 9c. $Q_1 = 1,82$; $Q_2 = 3,65$ e $Q_3 = 7,13$.
13c. $Q_2 = 7,51$. d. mín. = 6,10; $Q_1 = 7,30$; $Q_2 = 7,50$; $Q_3 = 7,65$; máx. = 8,30. Pontos extremos: 2 na parte superior e 5 na inferior.
15b. mín. = 100, $Q_1 = 116$, $Q_2 = 125$, $Q_3 = 134$, máx. = 160; sem extremos.

Seção 6.2:

10. Pelos gráficos, conclui-se que a Normal é mais adequada.

Seção 7.4:

4. [9,608; 10,392].

Box-plot

(Conforme apresentado na 7a. edição às páginas 20 e 21)

Uma representação gráfica envolvendo quartis é o *box-plot* ou *gráfico de caixa*, que permite visualizar diversos aspectos da distribuição dos dados, tais como posição, variabilidade, assimetria e mesmo a ocorrência de valores atípicos. Note que, para obter o *box-plot* conforme será definido, precisamos usar informações da tabela de dados brutos.

Para a construção do *box-plot*, definimos um retângulo ("caixa") em que a aresta inferior coincide com o primeiro quartil e a superior, com o terceiro quartil. A mediana é representada por um traço no interior do retângulo. Segmentos de reta, denominados *bigodes* por alguns autores, são incluídos no *box-plot*, partindo dos primeiro e terceiro quartis e terminando em valores definidos a seguir.

O intervalo $[Q_1; Q_3]$ contém 50% das observações centrais e dá uma ideia de quão dispersos são os valores observados. A amplitude desse intervalo, $IQ = Q_3 - Q_1$, recebe o nome de *intervalo interquartil*. Alguns autores consideram que, para conjuntos de dados "bem comportados", a maior parte das observações se situa no intervalo $[Q_1 - 1,5IQ; Q_3 + 1,5IQ]$, sendo que dados fora desse intervalo são potenciais valores *atípicos*, pois apresentariam um padrão distinto do esperado para a maior parte das unidades experimentais. Por essa razão, os limites desse intervalo serão denominados *pontos de corte*. Os segmentos partindo dos primeiro e terceiro quartis são limitados pelos valores mínimo e máximo dentro do intervalo acima descrito. Valores abaixo ou acima desses limites são representados por asteriscos e denominados *valores extremos*. A Figura 1.7 ilustra os elementos de um *box-plot*.

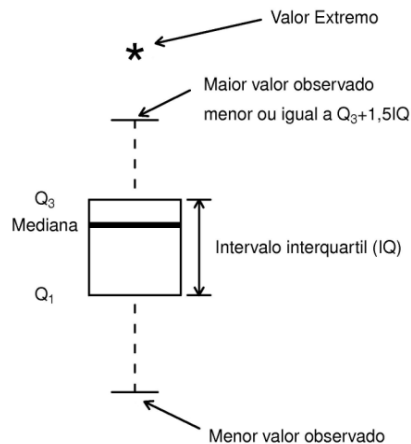


Figura 1.7. Elementos de um box-plot.

Na figura apresentada existe um valor extremo na parte superior do gráfico, mas não na parte inferior. O segmento superior vai do terceiro quartil até o maior valor que é menor ou igual a $Q_3 + 1,5IQ$; na parte inferior, o segmento vai de Q_1 até o menor valor observado que seja superior a $Q_1 - 1,5IQ$.

Como comentário adicional, note que os valores extremos são observações que podem ser atípicas para o fenômeno que gera os dados e, se esse for o caso, serão denominados *valores aberrantes* ou *outlier*. Esses dados devem ser investigados com maior cuidado pois muitas vezes devem-se a erros de digitação ou de registro dos valores quando se realiza a coleta dos dados. Entretanto, vale notar que para distribuições assimétricas é comum o aparecimento de valores extremos, como será discutido adiante.

(Segue o texto no livro, com exemplos de construção do box-plot)