

Gabarito para Exercício 1 da Lista 06 (ref. ao assunto “População e Amostra”).

Exc.1 (a). Na 1-a classe, que estende-se de 3 (incluído) até 5 (excluído) (os limites da classe foram definidos no enunciado), encontram-se 9 observações (a menor delas é 3,1 e a maior é 4,9). A altura do “prédio” correspondente, h_1 , calcula-se a partir da relação

$$h_1 \times (\text{a amplitude da classe}) = \text{a frequência relativa}$$

cuja expressão numérica é assim:

$$h_1 \times 2 = \frac{9}{120}$$

Na derivação da expressão numérica usamos dois fatos óbvios: (i) que 2 é a amplitude da classe (pois $5 - 3 = 2$), e (ii) que a quantidade total de observações é 120, o que acarreta o valor $9/120$ para a frequência relativa das observações contidas na classe tratada. O resultado é

$$h_1 = \frac{9}{120}/2 \approx 0,0375$$

A 2-a classe (definida pelos limites 5 (incluído) e 8 (excluído)) contém 65 observações. Sua amplitude é $8 - 5 = 3$ e sua frequência relativa é $\frac{65}{120}$. Consequentemente, a altura do prédio de histograma correspondente à 2-a classe é

$$h_2 = \frac{65}{120}/3 \approx 0,180$$

A 3-a classe (definida pelos limites 8 (incluído) e 10 (incluído também)) contém 46 observações. Sua amplitude é $10 - 8 = 2$ e a frequência relativa é $\frac{46}{120}$. Consequentemente, a altura do prédio de histograma correspondente à 3-a classe é

$$h_3 = \frac{46}{120}/2 \approx 0,192$$

O desenho do histograma, cujos valores foram calculados acima, está à esquerda na Figura 1. Gostaria de chamar sua atenção ao fato que as escalas dos eixos não são obrigadas a serem iguais. Essa desigualdade permitida foi empregada por mim no desenho com o intuito de fazê-lo parecer legível.

Exc.1 (b). Antes de mais nada, devo corrigir um erro do enunciado: está faltando a classe $[5, 6)$. Com essa correção, podemos dizer que o problema é o que se chama “construção de histograma por classes de amplitudes iguais”. Observe que a amplitude comum é 1; esse “1” aparecerá nos denominadores dos cálculos que darão as alturas dos prédios do histograma.

As frequências absolutas (ou, em outras palavras, as quantidades de observações do conjunto) que se encontram em cada classe são assim:

$$a_1 = 3, a_2 = 6, a_3 = 13, a_4 = 25, a_5 = 27, a_6 = 34, a_7 = 12$$

As correspondentes frequências relativas (f) dão-se pela fórmula (recorde que 120 é a quantidade total de observações do conjunto)

$$f_i = \frac{a_i}{120}$$

e as alturas (h) dos correspondetes “prédios” do histograma dão-se pela fórmula

$$h_i = \frac{f_i}{\text{amplitude comum das classes}} = \frac{f_i}{1} = f_i$$

(já que no presente caso, cada amplitude é 1). Fazendo as contas, temos:

$$h_1 = 0,025, h_2 = 0,05, h_3 = 0,108, h_4 = 0,208, h_5 = 0,225, h_6 = 0,283, h_7 = 0,1$$

O histograma correspondente está na Figura 1.

Eu não vou comentar sobre as características do conjunto de dados sugeridas pelo histograma agora construído, assim como não vou comparar esse histograma com o que fiz no item (a). Também, não vou fazer esse tipo de pergunta nas provas do curso.

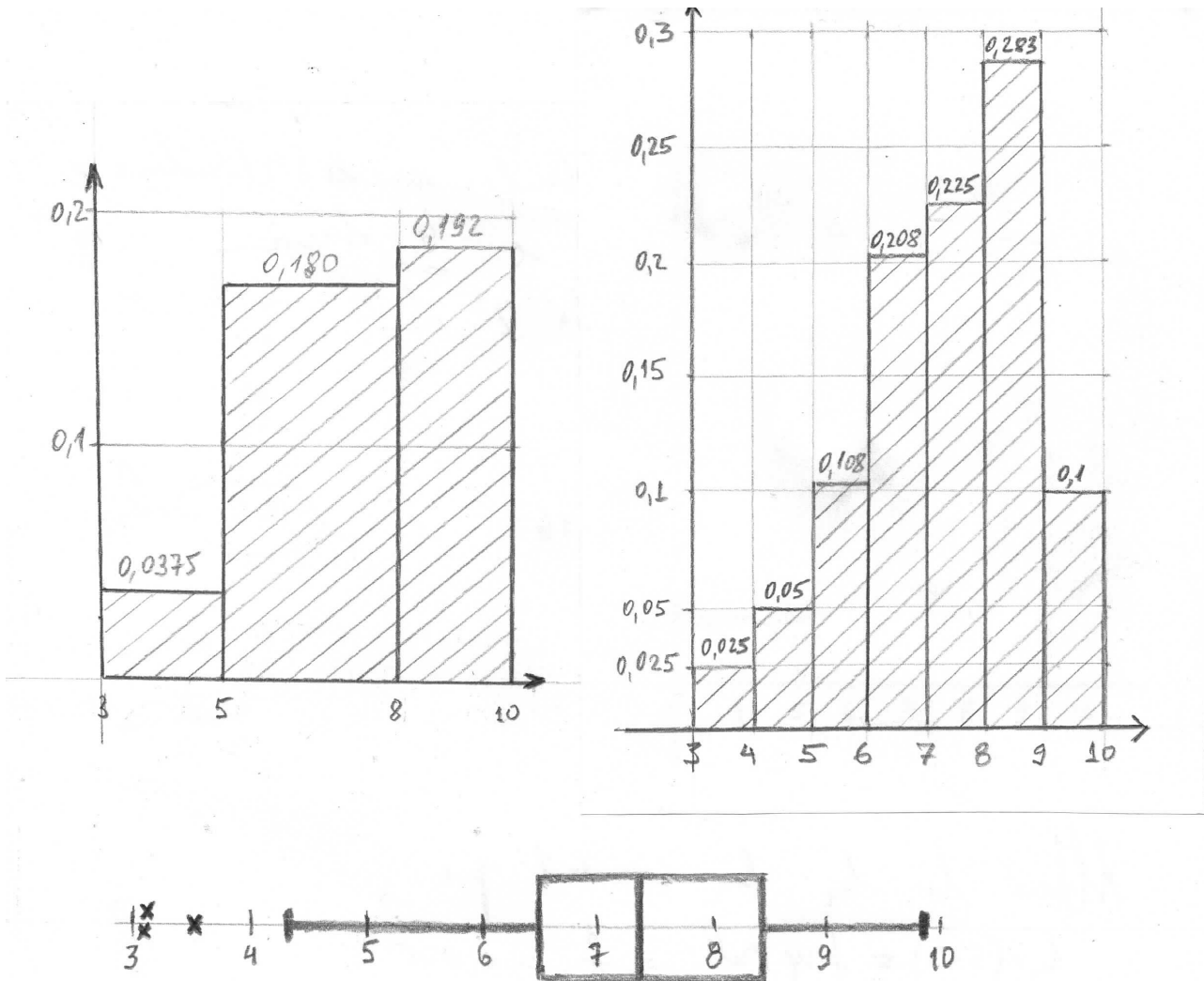


Figura 1

Exc.1 (c). O primeiro passo na construção de Box-Plot é o cálculo auxiliar que determina os números ordinais das observações que serão usadas para definir os valores de Q_1 , Q_2 e Q_3 (o primeiro, o segundo e o terceiro quartis). O cálculo usa o número total de observações de conjunto, que é 120 no presente caso ($n = 120$).

O cálculo auxiliar para a determinação do valor de Q_1 é:

$$\frac{1}{4}(n + 1) = \frac{120 + 1}{4} = 30,25$$

o que significa que

$$Q_1 = \text{a média simples entre a 30-a e 31-a observação do conjunto ordenado} = \frac{6,4 + 6,5}{2} = 6,45$$

(Observe que o conjunto já veio ordenado; se não viesse assim, teria a necessidade de ordená-lo antes de fazer os cálculos.)

O cálculo auxiliar para a determinação do valor de $Q2$ é:

$$\frac{1}{2}(n+1) = \frac{120+1}{2} = 60,5$$

o que significa que

$$Q2 = \text{a média simples entre a 60-a e 61-a observação do conjunto ordenado} = \frac{7,3 + 7,3}{2} = 7,3$$

O cálculo auxiliar para a determinação do valor de $Q3$ é:

$$\frac{3}{4}(n+1) = \frac{3(120+1)}{4} = 90,75$$

o que significa que

$$Q3 = \text{a média simples entre a 90-a e 91-a observação do conjunto ordenado} = \frac{8,3 + 8,4}{2} = 8,35$$

Observe que nos cálculos de valores de $Q1, Q2, Q3$ eu divido por 2 pois estou calculando média, enquanto que nos cálculos auxiliares meus multiplicadores são $\frac{1}{4}, \frac{1}{2}$ e $\frac{3}{4}$, valores de advêm da regra do cálculo auxiliar. Em outras palavras, esses $\frac{1}{4}, \frac{1}{2}$ e $\frac{3}{4}$ não têm nada a ver com pesos de médias ponderadas ou qualquer coisa semelhante.

Agora calculamos um valor auxiliar

$$1,5(Q3 - Q1) = 1,5(8,35 - 6,45) = 2,85$$

e usamos-o para definir L e R que serão empregados na construção de Box-Plot de acordo com as regras formuladas no parágrafo seguinte.

$$L = Q1 - 1,5(Q3 - Q1) = 6,45 - 2,85 = 3,6 \text{ e } R = Q3 + 1,5(Q3 - Q1) = 8,35 + 2,85 = 11,2$$

Qualquer observação menor que L será marcada como “outlier” (valor atípico), e o mínimo marcado no Box-Plot será a observação mínima entre as que estão não menor que L (no nosso caso, é 4,3). Os valores atípicos são 3.1 (marcado duas vezes) e 3.5. O papel de R obedece a lógica semelhante: qualquer observação acima de R será marcada como “outlier”, e o máximo do conjunto será definido como a maior observação que não ultrapassa R . No presente caso, R é maior que qualquer observação, é por isso que não haverá valores atípicos do lado direito do desenho, e o máximo adquirirá o valor da maior observação do conjunto.

Vale recordar que L e R não estarão marcados no desenho de Box-Plot. O desenho só mostra os valores atípicos, o mínimo, os valores de $Q1, Q2$ e $Q3$ e o máximo (vale realçar que o mínimo e o máximo mostrados por Box-Plot não sempre coincidem com o mínimo e o máximo do conjunto de observações; eles são determinados de acordo com o procedimento descrito no parágrafo acima). Todo o Box-Plot está na Figura 1. É uma tradição não marcar os valores numéricos dos elementos de Box-Plot, mas não é erro se você marcar os valores ao lado. É que a escala do eixo, cuja presença é obrigatória, permite descobrir esses valores caso houver necessidade.