Gabarito da Lista 1

Exercício 1

a) Quantitativa contínua
b) Qualitativa ordinal
c) Quantitativa discreta
d) Quantitativa contínua
e) Quantitativa contínua
f) Quantitativa discreta
g) Qualitativa nominal
h) Qualitativa nominal

Exercício 2

O enunciado nos deu que $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$.

a)

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \overline{x}) &= x_1 - \overline{x} + x_2 - \overline{x} + \ldots + x_n - \overline{x} \\
&= \sum_{i=1}^{n} x_i - n\overline{x} \\
&= \sum_{i=1}^{n} x_i - n\frac{\sum_{i=1}^{n} x_i}{n} \\
&= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i \\
&= 0
\end{aligned} \tag{1}$$

b)

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \overline{x})^2 &= \sum_{i=1}^{n}(x_i^2 - 2x_i\overline{x} + \overline{x}^2) \\
&= (x_1^2 - 2x_1\overline{x} + \overline{x}^2) + (x_2^2 - 2x_2\overline{x} + \overline{x}^2) + \ldots + (x_n^2 - 2x_n\overline{x} + \overline{x}^2) \\
&= \sum_{i=1}^{n} x_i^2 - 2\overline{x}\sum_{i=1}^{n} x_i + n\overline{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2\overline{x}n\overline{x} + n\overline{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2n\overline{x}^2 + n\overline{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - n\overline{x}^2
\end{aligned} \tag{2}$$

Exercício 3

Suponha uma variável X com n observações $(X_1, X_2, ..., X_n)$ com média $\overline{x}$, mediana MD(X) e desvio padrão DP(X).

a) Multiplicando cada observação por 2 é como se tivéssemos uma nova variável Y $(2X_1, 2X_2, ..., 2X_n)$.

Mediana: A mediana será a mesma observação mas com o seu valor multiplicado por 2, MD(Y) = 2MD(X).

Média: $\overline{y} = \frac{\sum_{i=1}^{n} 2x_i}{n} = 2\frac{\sum_{i=1}^{n} x_i}{n} = 2\overline{x}$

Desvio Padrão:

$$\begin{aligned} DP(Y) &= \sqrt{\frac{\sum_{i=1}^{n}(2x_i - \overline{y})^2}{n}} \\ &= \sqrt{\frac{\sum_{i=1}^{n}(2x_i - 2\overline{x})^2}{n}} \\ &= \sqrt{\frac{\sum_{i=1}^{n}[2(x_i - \overline{x})]^2}{n}} \\ &= \sqrt{4\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} \\ &= 2\sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} \\ &= 2DP(X) \end{aligned} \tag{3}$$

b) Somando 10 unidades em cada observação temos a variável Y $(X_1 + 10, X_2 + 10, ..., X_n + 10)$

Mediana: A mediana será a mesma observação mas com o seu valor mais 10 unidades, MD(Y) = MD(X) + 10.

Média: $\overline{y} = \frac{\sum_{i=1}^{n} x_i + 10}{n} = \frac{\sum_{i=1}^{n} x_i + n10}{n} = \frac{\sum_{i=1}^{n} x_i}{n} + 10 = \overline{x} + 10$

Desvio Padrão:

$$\begin{aligned} DP(Y) &= \sqrt{\sum_{i=1}^{n}(x_i + 10 - \overline{y})^2)n} \\ &= \sqrt{\frac{\sum_{i=1}^{n}(x_i + 10 - (\overline{x} + 10))^2}{n}} \\ &= \sqrt{\frac{\sum_{i=1}^{n}(x_i + 10 - \overline{x} - 10)^2}{n}} \\ &= \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} \\ &= DP(X) \end{aligned} \tag{4}$$

c) Subtraindo-se a média geral $\overline{x}$ de cada observação temos a variável Y $(X_1 - \overline{x}, X_2 - \overline{x}, ..., X_n - \overline{x})$

Mediana: A mediana será a mesma observação mas com o seu valor menos $\overline{x}$ unidades, MD(Y) = MD(X) - $\overline{x}$.

Média: $\overline{y} = \frac{\sum_{i=1}^{n} x_i - \overline{x}}{n} = \frac{\sum_{i=1}^{n} x_i - n\overline{x}}{n} = \frac{\sum_{i=1}^{n} x_i}{n} - \overline{x} = \overline{x} - \overline{x} = 0$

Desvio Padrão:

$$
\begin{aligned}
DP(Y) &= \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x} - \overline{y})^2}{n}} \\
&= \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x} - 0)^2}{n}} \\
&= \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} \\
&= DP(X) \qquad (5)
\end{aligned}
$$

d) Subtraindo-se a média geral $\overline{x}$ e dividindo pelo desvio padrão DP(X) de cada observação temos a variável Y $(\frac{X_1-\overline{x}}{DP(X)}, \frac{X_2-\overline{x}}{DP(X)}, ..., \frac{X_n-\overline{x}}{DP(X)})$

Mediana: A mediana será a mesma observação mas com o seu valor menos $\overline{x}$ e dividido por DP(X), $MD(Y) = \frac{MD(X)-\overline{x}}{DP(X)}$.

Média: $\overline{y} = \frac{\sum_{i=1}^{n} \frac{x_i-\overline{x}}{DP(X)}}{n} = \frac{\sum_{i=1}^{n} x_i-\overline{x}}{nDP(X)} = 0$, pois vimos que $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$

Desvio padrão:

$$
\begin{aligned}
DP(Y) &= \sqrt{\frac{\sum_{i=1}^{n}(\frac{x_i-\overline{x}}{DP(X)} - \overline{y})^2}{n}} \\
&= \sqrt{\frac{\sum_{i=1}^{n}(\frac{x_i-\overline{x}}{DP(X)} - 0)^2}{n}} \\
&= \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n(DP(X))^2}} \\
&= \frac{1}{DP(X)}\sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} \\
&= \frac{1}{DP(X)}DP(X) \\
&= 1 \qquad (6)
\end{aligned}
$$

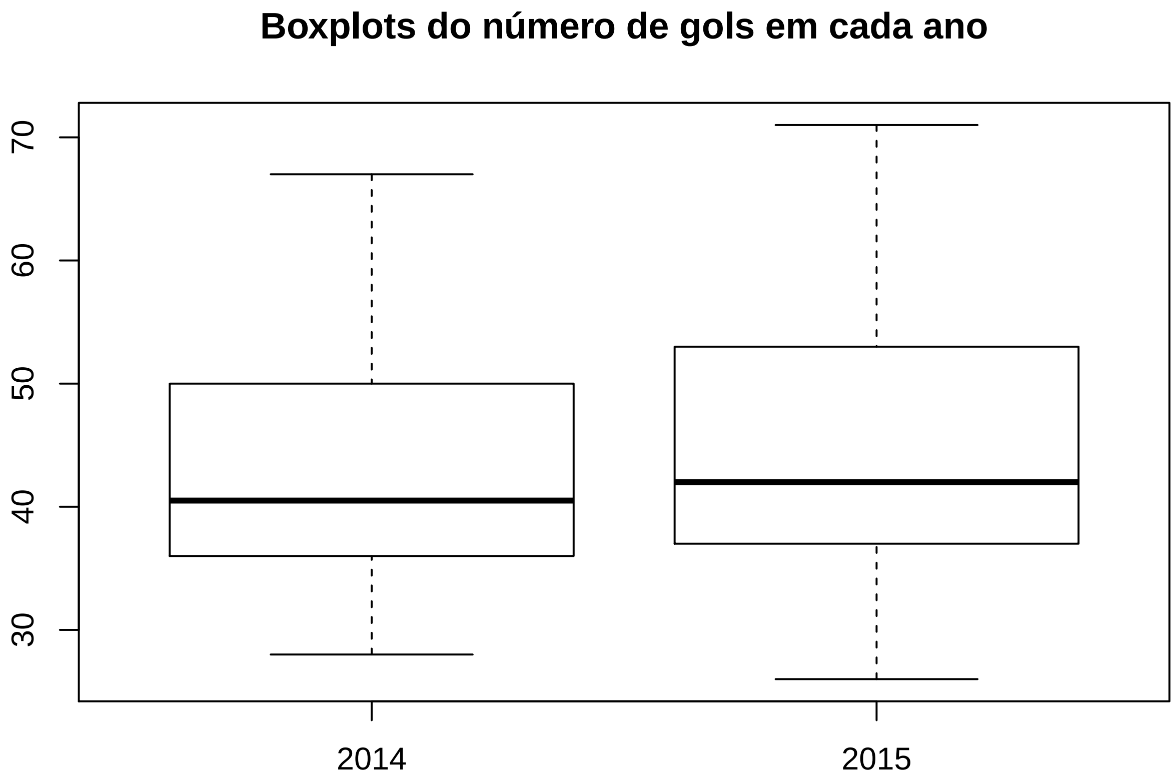
Exercício 4

Observação: esta é apenas **uma** das formas de se obter os valores dos quartis. Neste caso estou utilizando o software R para a resolução do exercício

a)

Ano	Mínimo	Q_1	Mediana	Q_3	Máximo	Dist. interquartil
2014	28	36	40,5	49,5	67	13,5
2015	26	37,5	42	53	71	15,5

b) Para a construção do boxplot, não há nada de errado quando é feito pelo R, porém se observarmos as escalas temos que o limite superior e o limite inferior correspondem ao máximo e o mínimo respectivamente observados nos dois anos. Veja abaixo:



Porém o máximo e mínimo observados não necessariamente serão o LS e o LI. Para sabermos os valores verdadeiros de **LS** e **LI** devemos utilizar as fórmulas $LS = Q3 + 1,5(Q_3 - Q_1)$ e $LI = Q1 - 1,5(Q_3 - Q_1)$. Em 2014 temos $LI = 15,75$ e $LS = 69,75$. Em 2015 temos $LI = 14,25$ e $LS + 76,25$.

Interpretação:

Observando os dois boxplots temos que a mediana de gols foi maior em 2015 e o mesmo occoreu com a distância interquartil. Isso mostra que houve uma maior variabilidade do número de gols em 2015. Percebe-se também que a variação do número de gols é maior dos times que estão acima da mediana nos dois anos. Por último, em ambos os anos não houve a presença de *outliers*.