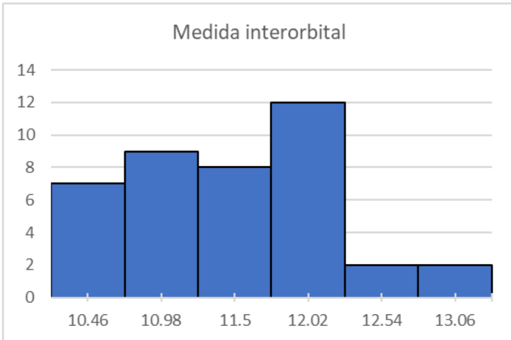
39. Seja X – v.a. (população) que representa a medida interorbital (em mm) de um pombo doméstico escolhido ao acaso. X pode tomar valores reais não negativos, ou seja, uma infinidade não contável de valores distintos, pelo que X é uma variável contínua.

a) A representação adequada de dados contínuos é o histograma.
$n = 40$: menor k tal que $2^k \geq 40 \Rightarrow \underline{k = 6}$ ($2^5 = 32$ (< 40) e $2^6 = 64$ (> 40))
$x_{(1)} = 10.2$, $x_{(40)} = 13.3 \Rightarrow r = x_{(40)} - x_{(1)} = 3.1$
$r/k = 3.1/6 = 0.51(6) \rightarrow \underline{h = 0.52}$
(note-se que, sendo r/k uma dízima infinita, é necessário arredondá-la por excesso para se garantir que todos os valores da amostra são incluídos nalguma das classes; por outro lado, como as observações são apresentadas com uma casa decimal, convém fazer um arredondamento pequeno, bastando para tal considerar duas casas decimais para h)
Tabela de frequências simples para os dados classificados:

Classe C_j	m_j	n_j
[10.20, 10.72[	10.46	7
[10.72, 11.24[	10.98	9
[11.24, 11.76[	11.50	8
[11.76, 12.28[	12.02	12
[12.28, 12.80[	12.54	2
[12.80, 13.32[	13.06	2
		40

Gráfico do histograma:



b) Média: $\overline{x} = \frac{1}{40}\sum_{i=1}^{40} x_i = 459.2/40 = \underline{11.48}$
Mediana: n = 40 é par $\Rightarrow Q_{1/2} = (x_{(20)} + x_{(21)})/2 = (11.6 + 11.6)/2 = \underline{11.6}$
1º Quartil: n/4 = 10 é inteiro $\Rightarrow Q_{1/4} = (x_{(10)} + x_{(11)})/2 = (10.9 + 11)/2 = \underline{10.95}$
3º Quartil: 3n/4 = 30 é inteiro $\Rightarrow Q_{3/4} = (x_{(30)} + x_{(31)})/2 = (11.9 + 11.9)/2 = \underline{11.9}$
Variância: $s^2 = \frac{1}{39}\left(\sum_{i=1}^{40} x_i^2 - 40\,\overline{x}^2\right) = (5290.28 - 40 \times 11.48^2)/39 \approx 0.4786$
Desvio-padrão: $s = \sqrt{s^2} \approx \sqrt{0.4786} \approx \underline{0.69}$

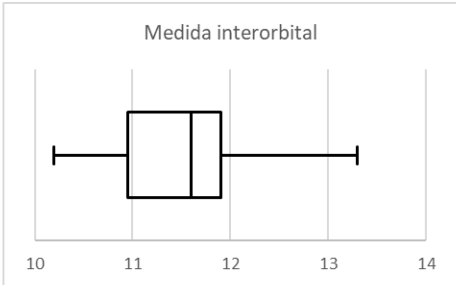
c) Para construir a *boxplot* (ou *box-and-whiskers* ou caixa-com-bigodes, em português) é necessário calcular as barreiras, ou limites, de *outliers* – Limite Inferior (LI) e Limite Superior (LS) – bem como os correspondentes pontos adjacentes – Adjacente Inferior (AI) e Adjacente Superior (AS).
Sendo H a amplitude inter-quartis, i.e., $H = Q_{3/4} - Q_{1/4}$, tem-se:
$$LI = Q_{1/4} - 1.5\,H \;\text{ e }\; LS = Q_{3/4} + 1.5\,H\,.$$
As observações que estejam fora do intervalo [LI, LS] designam-se por *outliers* na amostra. Também é comum ser mais específico dizendo-se que uma observação que seja menor do que LI é um *outlier* inferior e uma observação que seja maior do que LS é um *outlier* superior.

Para os dados observados, tem-se:
$H = Q_{3/4} - Q_{1/4} = 11.9 - 10.95 = 0.95$; $1.5H = 1.425$
$LI = Q_{1/4} - 1.5\,H = 10.95 - 1.425 = 9.53 \rightarrow 9.53 < x_{(1)} \Rightarrow$ Não existem *outliers* inferiores
$LS = Q_{3/4} + 1.5\,H = 11.9 + 1.425 = 13.33 \rightarrow 13.33 > x_{(40)} \Rightarrow$ Não existem *outliers* superiores

Finalmente é necessário calcular os pontos AI e AS. O ponto AI é a menor observação na amostra que não é inferior a LI, ou seja, é o menor valor na amostra que pertence ao intervalo [LI, LS]. O ponto AS é a maior observação na amostra que não é superior a LS, ou seja, é o maior valor na amostra que pertence ao intervalo [LI, LS].

Para os dados observados, tem-se:
$LI < x_{(1)} \Rightarrow AI = x_{(1)} = 10.2$
$LS > x_{(40)} \Rightarrow AS = x_{(40)} = 13.3$

Representação dos dados em *boxplot*:

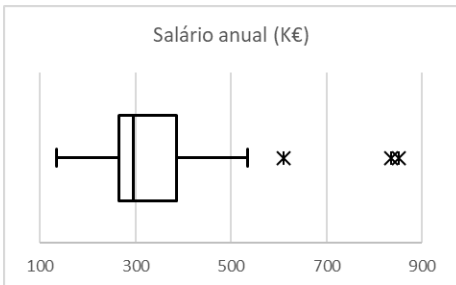


40. Seja X – v.a. (população) que representa a o salário, em milhares de Euros, de um jogador, escolhido ao acaso, na liga de futebol profissional europeia em questão.

a) Média: $\overline{x} = \frac{1}{50}\sum_{i=1}^{50} x_i = 16925/50 = 338.5$
Mediana: n = 50 é par $\Rightarrow Q_{1/2} = (x_{(25)} + x_{(26)})/2 = 295 + 295)/2 = 295$
Variância: $s^2 = \frac{1}{49}\left(\sum_{i=1}^{49} x_i^2 - 50\,\overline{x}^2\right) = (6772225 - 50\times338.5^2)/49 \approx 21288.01$
1º Quartil: n/4 = 12.5 é não inteiro $\Rightarrow Q_{1/4} = x_{(13)} = 265$
3º Quartil: 3n/4 = 37.5 é não inteiro $\Rightarrow Q_{3/4} = x_{(38)} = 385$

b) $H = Q_{3/4} - Q_{1/4} = 385 - 265 = 120$; $1.5H = 180$
$LI = Q_{1/4} - 1.5\,H = 265 - 180 = 85 \;\rightarrow\; 85 < x_{(1)} \Rightarrow$ Não existem *outliers* inferiores
$LS = Q_{3/4} + 1.5\,H = 385 + 180 = 565 \;\rightarrow\; x_{(48)} > 565 \Rightarrow$ $x_{(48)} = 610$, $x_{(49)} = 835$ e $x_{(50)} = 850$ são candidatos a *outliers*
(note-se que um *outlier* na amostra diz-se um candidato a *outlier* na população)
$LI < x_{(1)} \Rightarrow AI = x_{(1)} = 135$
$x_{(47)} < LS < x_{(48)} \Rightarrow AS = x_{(47)} = 535$

Representação dos dados em *boxplot*:



41. Sejam:
T_1 – v.a. (população) que representa o tempo de espera de um cliente escolhido ao acaso na zona Z_1
T_2 – v.a. (população) que representa o tempo de espera de um cliente escolhido ao acaso na zona Z_2
Uma vez que T_1 e T_2 representam tempos, são ambas variáveis contínuas. Além disso, dizem respeito a zonas de atendimento distintas, pelo que são variáveis independentes.
Sendo n_i a dimensão da amostra de tempos de T_i, i = 1, 2, tem-se que $n_1 = n_2 = 12$, ou seja as amostras de tempos de T_1 e T_2 têm a mesma dimensão, n = 12. Com tão poucas observações não se devem construir histogramas; contudo, podem construir-se *boxplots* e, como o objectivo é comparar as duas zonas de atendimento, devem construir-se *boxplots* paralelas, ou seja, representadas no mesmo referencial.

Medianas: n = 12 é par $\Rightarrow Q_{1/2} = (x_{(6)} + x_{(7)})/2$;
1º.s Quartis: n/4 = 3 é inteiro $\Rightarrow Q_{1/4} = (x_{(3)} + x_{(4)})/2$
3º.s Quartis: 3n/4 = 9 é inteiro $\Rightarrow Q_{3/4} = (x_{(9)} + x_{(10)})/2$
$H = Q_{3/4} - Q_{1/4}$; $LI = Q_{1/4} - 1.5\ H$; $LS = Q_{3/4} + 1.5\ H$

	Z_1	Z_2
$Q_{1/4}$:	(4.9 + 5.1)/2 = 5.0	(4.1 + 4.5)/2 = 4.3
$Q_{1/2}$:	(5.5 + 5.7)/2 = 5.6	(5.8 + 5.8)/2 = 5.8
$Q_{3/4}$:	(5.8 + 5.8)/2 = 5.8	(6.0 + 6.6)/2 = 6.3
H:	5.8 – 5.0 = 0.8	6.3 – 4.3 = 2.0
1.5 H:	1.2	3.0
LI:	5.0 – 1.2 = 3.8	4.3 – 3.0 = 1.3
AI:	$t_{1,(1)} = 4.8$	$t_{2,(1)} = 2.0$
LS:	5.8 + 1.2 = 7.0	6.3 + 3.0 = 9.3
AS:	$t_{1,(12)} = 6.5$	$t_{2,(10)} = 6.6$
Outliers:	Não existem	$t_{2,(11)} = 9.9$ e $t_{2,(12)} = 10.6$

Representação das duas amostras em *boxplots* paralelas: