

Question 1

- a) All individuals in the Irish public.
- b) The 80 individuals contacted.
- c) $n = 80$.
- d) Parameter: the true proportion of individuals in favour of the policy. $p = \text{unknown}$.
Statistic: the proportion of individuals in favour of the policy in our sample. $\hat{p} = \frac{23}{80} = 0.2875$.
- e) The variable “in favour” (yes / no) was collected from each individual \Rightarrow categorical data.

- f) Only individuals in her local area were contacted. This is not representative of all individuals in Ireland as individuals in other areas may have a different opinion.
- g) She believes that the true proportion is less than 0.3. Certainly 0.2875 is less than 0.3. However, as mentioned above, the sample may not represent the target population. Even if the sample *did* represent the target population, we need to carry out a formal statistical test to answer her question (later in the course).

Question 2

- a) Lines of code - Numeric discrete.
- b) Time taken - Numeric continuous.
- c) Experience level - Categorical.
- d) File size - Numeric continuous.

- e) Does it work? - Categorical.
- f) Age - Numeric discrete.
- g) Gender - Categorical.

Question 3

We need to sort the data for the median and IQR.
We also need to sum the numbers and the squared numbers for mean and standard deviation.

Therefore:

	1	2	3	4	5	6	7	8	9	10	Σ
x_i	0	1	1	2	2	3	4	4	5	8	30
x_i^2	0	1	1	4	4	9	16	16	25	64	140

a)
$$\bar{x} = \frac{\sum x_i}{n} = \frac{30}{10} = 3.$$

c)
$$s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1} = \frac{140 - 10(3^2)}{9}$$

$$= \frac{140 - 10(9)}{9}$$

$$= \frac{50}{9} = 5.56.$$

$$\Rightarrow s = \sqrt{s^2} = \sqrt{5.56} = 2.36.$$

	Position	Value
Q_1	$\frac{n+1}{4} = \frac{11}{4} = 2.75$	$\frac{1+1}{2} = 1$
Q_2	$\frac{n+1}{4} \times 2 = 2.75(2) = 5.5$	$\frac{2+3}{2} = 2.5$
Q_3	$\frac{n+1}{4} \times 3 = 2.75(3) = 8.25$	$\frac{4+5}{2} = 4.5$

b) The median is $Q_2 = 2.5$.

d) $IQR = Q_3 - Q_1 = 4.5 - 1 = 3.5$.

Question 4

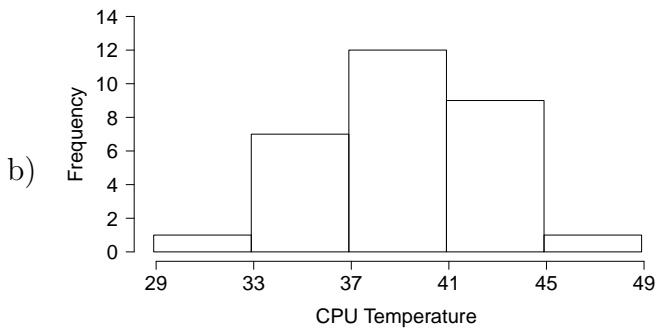
a) $\text{width} = \frac{\max(x) - \min(x)}{\# \text{classes}} = \frac{47.9 - 29.7}{5} = \frac{18.2}{5} = 3.64.$

Always round this up \Rightarrow width = 4.

Starting at 29 \Rightarrow 29–33, 33–37, 37–41, 41–45, 45–49.

Thus, the frequency table is:

Class	Freq.	R. Freq.
29 – 32.9	1	$\frac{1}{30} = 0.033$
33 – 36.9	7	$\frac{7}{30} = 0.233$
37 – 40.9	12	$\frac{12}{30} = 0.4$
41 – 44.9	9	$\frac{9}{30} = 0.3$
45 – 48.9	1	$\frac{1}{30} = 0.033$
Total:	$n = 30$	$\frac{30}{30} = 1.0$



	Position	Value
Q_1	$\frac{n+1}{4} = \frac{31}{4} = 7.75$	$\frac{36.7+36.8}{2} = \mathbf{36.75}$
Q_2	$\frac{n+1}{4} \times 2 = 7.75(2) = 15.5$	$\frac{39.3+39.6}{2} = \mathbf{39.45}$
Q_3	$\frac{n+1}{4} \times 3 = 7.75(3) = 23.25$	$\frac{41.4+41.6}{2} = \mathbf{41.5}$

c) The median is $Q_2 = 39.45$.

d) The quartiles are $Q_1 = 36.75$, $Q_2 = 39.45$ and $Q_3 = 41.5$.

e) $IQR = Q_3 - Q_1 = 41.5 - 36.75 = 4.75$.

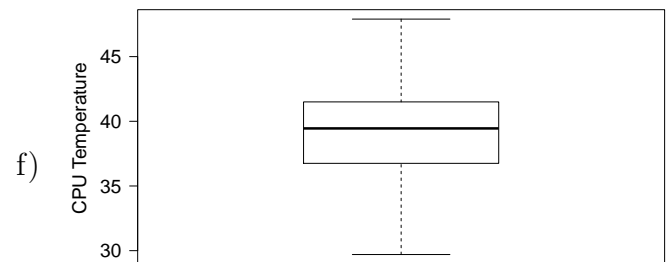
$$LF = Q_1 - 1.5IQR$$

$$= 36.75 - 1.5(4.75) = 29.625.$$

$$UF = Q_3 + 1.5IQR$$

$$= 41.5 + 1.5(4.75) = 48.625$$

There are no numbers less than 29.625 or greater than 48.625 \Rightarrow no outliers.



g) Based on the histogram and boxplot, we see that the data is symmetrical.