

Solutions to Exercises 1

1. Data types:

- (a) Categorical
- (b) Continuous
- (c) Ordinal
- (d) Discrete

2. (a) Systematic sampling.

- (b) Quasi-random.

3. Relative frequency table for data on number of calls:

Calls made	Tally	Frequency	Relative Frequency %
0		5	25
1		6	30
2		3	15
3		3	15
4		2	10
5		1	5
Totals		20	100

4. Relative frequency table for length of call data:

Length of call	Tally	Frequency	Relative Frequency %
$250 \leq \text{length} < 270$		1	4
$270 \leq \text{length} < 290$		6	24
$290 \leq \text{length} < 310$		9	36
$310 \leq \text{length} < 330$		7	28
$330 \leq \text{length} < 350$		2	8
Totals		25	100

Solutions to Exercises 2

1. The stem and leaf plot is

8		1	2											
8		5	7	8	9									
9		2	3	3	4									
9		5	5	6	6	6	7	7	9	9				
10		0	0	0	0	0	1	2	2	2	3	3	4	4
10		5	6	6	6	6	6	7	8	9				
11		0	2	3	3	3								
11		5	6											
12														
12		8												

Stem Leaf

$n = 50$, stem unit = 1kg, leaf unit = 0.1kg.

Comment: The distribution is roughly symmetric around a value of about 10kg. Most of the potatoes weigh between about 8 and 12kg. There are no obvious outliers.

2. A histogram is more appropriate, because the data are continuous.
3. The frequency polygons show a clear increase in sales due to the advertising campaign (the “After” polygon is shifted to the right), suggesting that the campaign has been a success.

Solutions to Exercises 3

1. (a) Let the weight of sack i be x_i . The sum of the weights is

$$\sum_{i=1}^{50} x_i = 505.8 \text{ kg.}$$

Therefore the sample mean is

$$\bar{x} = \frac{1}{50} \sum_{i=1}^{50} x_i = \frac{505.8}{50} = 10.116 \text{ kg.}$$

- (b) The observations into increasing order.

8.1	8.2	8.5	8.7	8.8	8.9	9.2	9.3	9.3	9.4
9.5	9.5	9.6	9.6	9.6	9.7	9.7	9.9	9.9	10.0
10.0	10.0	10.0	10.0	10.1	10.2	10.2	10.2	10.3	10.3
10.4	10.4	10.4	10.5	10.6	10.6	10.6	10.6	10.6	10.7
10.8	10.9	11.0	11.2	11.3	11.3	11.3	11.5	11.6	12.8

As there are $n = 50$ observations, the median M is the $(50 + 1)/2 = 25\frac{1}{2}$ th smallest observation, that is, half way between the 25th and 26th smallest observations, which are 10.1 and 10.2, respectively:

$$M = 10.1 + \frac{1}{2}(10.2 - 10.1) = \frac{10.1 + 10.2}{2} = 10.15 \text{ kg.}$$

- (c) The range of the data is the difference between the largest and smallest observations. Here the minimum and maximum values are $\min = 8.1$ kg and $\max = 12.8$ kg and so the range is

$$\text{Range} = \max - \min = 12.8 - 8.1 = 4.7 \text{ kg.}$$

- (d) The interquartile range is the difference between the upper and lower quartiles. As there are $n = 50$ observations, the lower quartile is the $(50 + 1)/4 = 12\frac{3}{4}$ th smallest observation, that is, three quarters of the way between the 12th and 13th smallest observations, which are 9.5 and 9.6, respectively:

$$Q_1 = 9.5 + \frac{3}{4} \times (9.6 - 9.5) = 9.575 \text{ kg.}$$

Similarly, the upper quartile is the $3(50 + 1)/4 = 38\frac{1}{4}$ th smallest observation, that is, a quarter of the way between the 38th and 39th smallest observations. As both of these observations are 10.6 kg, we have

$$Q_3 = 10.6 + \frac{1}{4}(10.6 - 10.6) = 10.6 \text{ kg.}$$

Therefore, the interquartile range is

$$\text{IQR} = Q_3 - Q_1 = 10.6 - 9.575 = 1.025 \text{ kg.}$$

- (e) The sample standard deviation is best calculated using your calculator in statistics mode. However we can also do the calculation the long way. We find $\sum_{i=1}^{50} x_i^2 = 5157.54$. Therefore the sample variance is

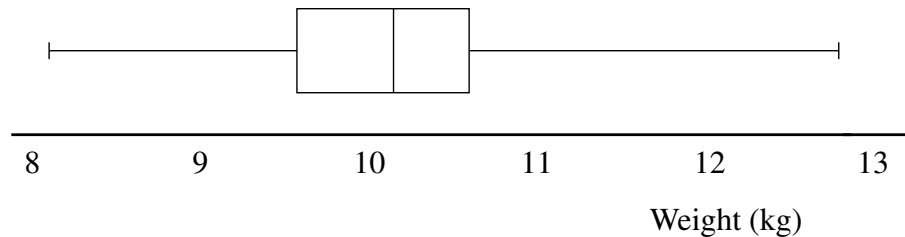
$$\begin{aligned} s^2 &= \frac{1}{n-1} \left\{ \sum_{i=1}^n x_i^2 - n\bar{x}^2 \right\} \\ &= \frac{1}{49} (5157.54 - 50 \times 10.116^2) \\ &= \frac{40.8672}{49} \\ &= 0.8340245 \text{ kg}^2. \end{aligned}$$

The sample standard deviation is

$$s = \sqrt{s^2} = \sqrt{0.8340245} \simeq 0.913 \text{ kg}.$$

Note that this is the value obtained on calculators using the s or $x\sigma_{n-1}$ button and not the σ or $x\sigma_n$ button.

- (f) The box plot is constructed from the minimum (8.1), lower quartile (9.575), median (10.15), upper quartile (10.6) and maximum (12.8), which were found earlier and looks like this:



Comment: The distribution of weights is roughly symmetric about a median of around 10 kg.

- (g) Grouping these data into a frequency table gives

Class	Class Interval	Frequency
1	$8.0 \leq x < 8.5$	2
2	$8.5 \leq x < 9.0$	4
3	$9.0 \leq x < 9.5$	4
4	$9.5 \leq x < 10.0$	9
5	$10.0 \leq x < 10.5$	14
6	$10.5 \leq x < 11.0$	9
7	$11.0 \leq x < 11.5$	5
8	$11.5 \leq x < 12.0$	2
9	$12.0 \leq x < 12.5$	0
10	$12.5 \leq x < 13.0$	1
Total		$n = 50$

(The class intervals and frequencies can be determined from the stem and leaf plot in Exercise 1 of Chapter 2.)

- (h) The modal class in the grouped frequency table is the class with the largest frequency, that is class 5, with $10.0 \leq x < 10.5$.
2. (a) Mark is correct. There is one very large cookie which weighs 51.3g; this might distort the sample mean and sample standard deviation (as their calculation involves all the data).
- (b) The median and inter-quartile range are more appropriate measures of location and spread for these data as they are more robust to outliers.

Ordered data: 22.4, 23.4, 24.9, 25.4, 25.6, 26.0, 26.3, 26.5, 27.1, 51.3

$n = 10$ so the median is the $(n + 1)/2 = 11/2 = 5.5$ th smallest observation, i.e. $25.6 + 0.5 \times (26.0 - 25.6) = 25.8g$.

Q_1 is the $(n + 1)/4 = 11/4 = 2.75$ th smallest observation, i.e. $23.4 + 0.75 \times (24.9 - 23.4) = 24.525g$.

Q_3 is the $3 \times (n + 1)/4 = 3 \times 11/4 = 8.25$ th smallest observation, i.e. $26.5 + 0.25 \times (27.1 - 26.5) = 26.65g$.

Therefore, the inter-quartile range is

$$IQR = Q_3 - Q_1 = 26.65 - 24.525 = 2.125g$$

3. Females have the more variable visit lengths as the standard deviation for females is greater than that for males.