

# Outliers

There is no agreed definition of outliers, and any exam question requiring their use will say specifically how they are to be calculated

## Common Definitions

### Median and IQR

Outliers more than  $1.5 \times \text{IQR}$  above upper quartile

$$> Q_3 + 1.5(Q_3 - Q_1)$$

or less than  $1.5 \times \text{IQR}$  below lower quartile

$$< Q_1 - 1.5(Q_3 - Q_1)$$

Generally  $> Q_3 + k(Q_3 - Q_1)$

$$< Q_1 - k(Q_3 - Q_1)$$

---

### Mean and Standard Deviation

Outliers more than 2 standard deviations away from the mean

$$> \bar{x} + 2 \text{ s.d}$$

$$< \bar{x} - 2 \text{ s.d}$$

or  $> \mu + 2\sigma$

$$< \mu - 2\sigma$$

I have not see any variation on this definition

**Exercise 3A**

1 Some data is collected.  $Q_1 = 46$  and  $Q_3 = 68$ .

A value greater than  $Q_3 + 1.5 \times (Q_3 - Q_1)$  or smaller than  $Q_1 - 1.5 \times (Q_3 - Q_1)$  is defined as an outlier.

Work out whether the following are outliers using this rule:

- a 7      b 88      c 105

42 
$$IQR = Q_3 - Q_1 = 68 - 46 = 22$$

$$\begin{aligned} \text{Outliers} &> Q_3 + 1.5IQR \\ &> 68 + 1.5 \times 22 = 101 \\ &< Q_1 - 1.5IQR \\ &< 46 - 1.5 \times 22 = 13 \end{aligned}$$

- a)  $7 < 13$       so 7 is an outlier  
 $13 < 88 < 101$       so 88 is not an outlier  
 $105 > 101$       so 105 is an outlier

2 The masses of male and female turtles are given in grams. For males, the lower quartile was 400 g and the upper quartile was 580 g. For females, the lower quartile was 260 g and the upper quartile was 340 g.

An outlier is an observation that falls either  $1 \times$  (interquartile range) above the upper quartile or  $1 \times$  (interquartile range) below the lower quartile.

a Which of these male turtle masses would be outliers?

- 400 g    260 g    550 g    640 g

b Which of these female turtle masses would be outliers?

- 170 g    300 g    340 g    440 g

c What is the largest mass a male turtle can be without being an outlier?

**Hint** The definition of an outlier here is different from that in question 1. You will be told which rule to use in the exam.

Males       $Q_1 = 400g$        $Q_3 = 580g$

Females       $Q_1 = 260g$        $Q_3 = 340g$

a) Males 
$$\text{Outlier} > Q_3 + IQR = 580 + 180 = 760$$

$$\text{outlier} < Q_1 - IQR = 400 - 180 = 220$$

a) outliers 400g 260g 550g 640g  
none of these

Females 
$$\text{Outlier} > Q_3 + IQR = 340 + 80 = 420$$

$$\text{outlier} < Q_1 - IQR = 260 - 80 = 180$$

b) outliers 170g 300g 340g 440g

c) male can be up to 760g without being an outlier

3 The masses of arctic foxes are found and the mean mass was 6.1 kg. The variance was 4.2. An outlier is an observation which lies  $\pm 2$  standard deviations from the mean.

a Which of these arctic fox masses are outliers?

2.4 kg      10.1 kg      3.7 kg      11.5 kg

b What are the smallest and largest masses that an arctic fox can be without being an outlier?

$$\mu = 6.1 \text{ kg} \quad \sigma = \sqrt{4.2}$$

$$\begin{aligned} \text{Outlier} &> \mu + 2\sigma = 6.1 + 2\sqrt{4.2} \\ &= 10.2 \text{ kg} \end{aligned}$$

$$\begin{aligned} \text{Outlier} &< \mu - 2\sigma = 6.1 - 2\sqrt{4.2} \\ &= 2.00 \text{ kg} \end{aligned}$$

a) outliers 2.4 kg 10.1 kg 3.7 kg **11.5 kg**

b) smallest without being an outlier 2.00 kg  
largest without being an outlier 10.2 kg

4 The ages of nine people at a children's birthday party are recorded.  $\Sigma x = 92$  and  $\Sigma x^2 = 1428$ .

a Calculate the mean and standard deviation of the ages. (3 marks)

An outlier is an observation which lies  $\pm 2$  standard deviations from the mean.

One of the ages is recorded as 30.

b State, with a reason, whether this is an outlier. (2 marks)

c Suggest a reason why this age could be a legitimate data value. (1 mark)

d Given that all nine people were children, clean the data and recalculate the mean and standard deviation. (3 marks)

#### Problem-solving

After you clean the data you will need to find the new values for  $n$ ,  $\Sigma x$  and  $\Sigma x^2$ .

$$\text{Mean} = \frac{\Sigma x}{n} \quad \text{s.d} = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$$

$$a) \quad \bar{x} = \frac{92}{9} = 10.2 \quad \text{to 3 s.f.}$$

$$\text{s.d.} = \sqrt{\frac{1428}{9} - \left(\frac{92}{9}\right)^2} = 7.36 \quad \text{to 3 s.f.}$$

---

$$b) \quad 10.2 + 2(7.36) = 24.92$$

$30 > 24.92$  so 30 an outlier

---

c) Supervising adult at party

---

$$d) \quad n \rightarrow 8 \quad \sum x \rightarrow 92 - 30 = 62$$

$$\sum x^2 \rightarrow 1428 - 30^2 = 528$$

$$\bar{x} = \frac{62}{8} = 7.75$$

$$\text{s.d.} = \sqrt{\frac{528}{8} - 7.75^2} = 2.44$$

to 3 s.f.

---