

## S4-1 Quantiles and Spread

- quantiles and ogives
- five-number summaries, box plots
- inter-quartile range
- standard deviation

[Summary](#) [Learn](#) [Solve](#) [Revise](#) [Answers](#)

---

---

### Summary

---

---

If a set of data is arranged from smallest to largest, then the  $n$ th percentile is the value  $n\%$  of the way along the line. The  $n$ th decile is  $n$  tenths of the way along and the  $n$ th quartile is  $n$  quarters of the way along.

An ogive is a graph of percentile number against variable value.

A five-number summary is a summary of a data set giving an indication of typical values and of spread. It consists of the five quartiles  $Q_0$ ,  $Q_1$ ,  $Q_2$ ,  $Q_3$  and  $Q_4$ .

The inter-quartile range is a measure of the spread of a data set equal to  $Q_3 - Q_1$ .

Outliers are data points very different from the rest of the set. A number is considered an outlier if it is more than 1.5 times the IQR below  $Q_1$  or 1.5 times the IQR above  $Q_3$ .

A box plot is a graphical representation of data based on the five-number summary.

The standard deviation is the most commonly used measure of spread and is the square root of the mean of the squares of the deviations of the data from the mean.

---

---

### Learn

---

---

#### Quantiles and Ogives

If a set of measurements is ordered from smallest to biggest, a quantile is an indicator of how far along the set a measurement lies. Three types of quantiles are commonly used. They are percentiles, deciles and quartiles.

#### *Percentiles*

Many mothers take their babies to a clinic for a regular check-up. The clinic makes various measurements like mass, length, head circumference etc.

Rosemary took her baby girl for a check-up at age 6 months. She was told that the baby's head circumference was 42.9 cm. This didn't mean much to Rosemary: she didn't know if it was normal, large or small.

The nurse told her that 42.9 cm was at the 97<sup>th</sup> **percentile**, meaning that her baby's head circumference was bigger than 97% of baby girls that age and smaller than 3%. This meant a lot more. Rosemary could tell now that her baby had an unusually large head.



An unusually large head can be a sign of a problem like hydrocephalus, where there is too much fluid pressure around the brain causing the skull to grow larger than it should. However, she was told that it is only a serious worry above the 98<sup>th</sup> percentile. Also, Rosemary and the baby's father were both big people, the baby's length was at the 90<sup>th</sup> percentile and she showed no other signs of problems. So she was told not to worry, but that they would keep an eye on it at future check-ups.

Percentiles are useful for seeing how a measurement compares to the normal range. For 6-month-old girls, selected percentiles are shown in this table.

Percentile	3	10	25	50	75	90	97
Head circumference (cm)	38.0	38.8	39.7	40.5	41.4	42.1	42.9

So 3% of 6-month-old girls have a circumference less than 38.0 cm, 10% have a circumference less than 38.8 cm and so on. Think about what this table is saying until you have a good feel for it.

Another way of thinking about percentiles is this: if you lined up all the 6-month-old girls in the world in order from smallest to largest head, the baby at the 10<sup>th</sup> percentile would be 10% of the way along the line, the baby at the 60<sup>th</sup> percentile would be 60% of the way along the line and the baby at the 100<sup>th</sup> percentile would be right at the end of the line. You can probably see that the 50<sup>th</sup> percentile is also the median.

## Practice

Q1 Use the table above to answer the following questions about the head circumference of 6-month-old girls.

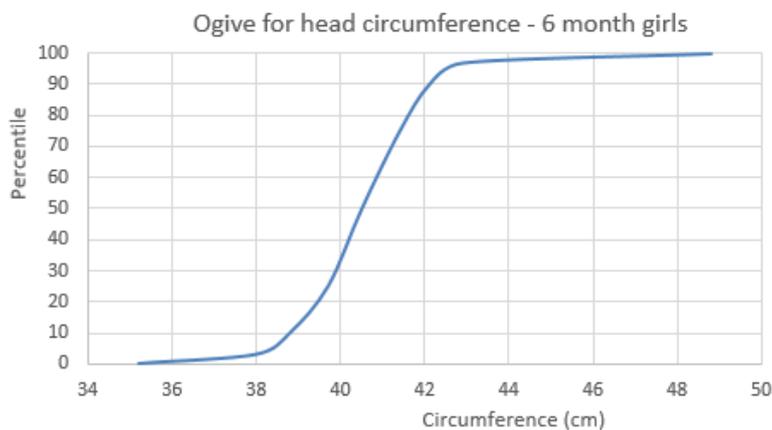
(a) Alice is at the 10<sup>th</sup> percentile; Shani is at the 50<sup>th</sup> percentile. Who has

the larger head?

- (b) What is the circumference of Alice's head?
- (c) What is the circumference of Sharni's head?
- (d) Marcia's head circumference is 41.4 cm. What percentile is that?
- (e) Is Marcia's head larger or smaller than average?
- (f) It is possible to interpolate between the percentiles given in the table.  
For example, a head circumference of 40.0 cm would be about the 30<sup>th</sup> or 31<sup>st</sup> percentile. What would be the percentile for a 41.9 cm head?
- (g) What would be the percentile for a 38.2 cm head?
- (h) What would be the percentile for a 43.1 cm head?
- (i) What circumference is the 20<sup>th</sup> percentile?
- (j) What circumference is the median?
- (k) Katie's head is larger than 65% of 6-month girls. What is her head circumference?
- (l) Jean's head is smaller than 13% of 6-month girls. What is her percentile?
- (m) What is Jean's head circumference?

## Ogives

Another way of presenting the information in the table above is by using a graph called an **ogive**. An ogive shows the percentile for each circumference.

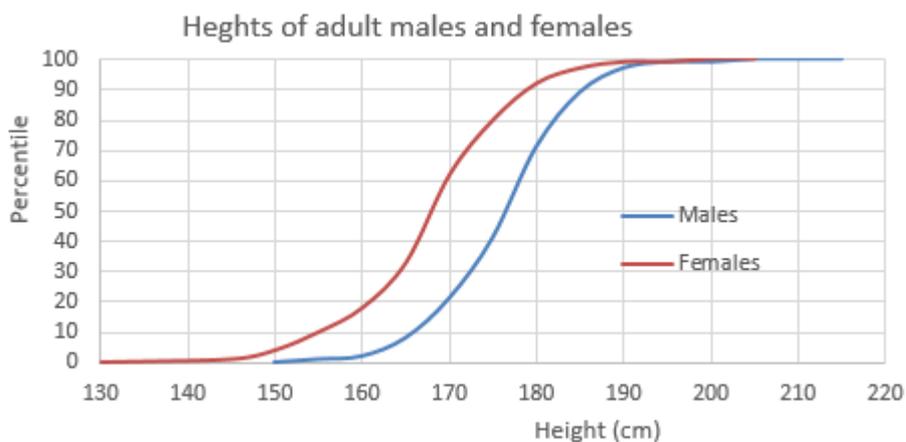


Using an ogive, we can read off the percentile for any circumference or the circumference for any percentile. Make sure you have the hang of how this graph works and what it shows us.

## Practice

- Q2 Use the ogive above to answer the following questions for a 6-month-old girl.
- What circumference is the 60<sup>th</sup> percentile?
  - What circumference is the 0<sup>th</sup> percentile?
  - What percentile is a circumference of 39.2 cm?
  - What percentile is a circumference of 45.9 cm?
  - What is the median circumference?

The graph below is the ogives for heights of adult males and females. If you have finished growing or have a good idea how tall you will grow to, you might like to see which percentile you are on.



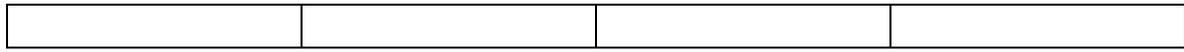
## Deciles and Quartiles

Deciles are used a lot less than percentiles. Whereas percentiles divide up the data into 100 parts, deciles divide it up into 10 parts. So the 1<sup>st</sup> decile is the 10<sup>th</sup> percentile, the second decile is the 20<sup>th</sup> percentile and so on.

Quartiles are commonly used though. Quartiles divide the data up into 4 quarters. So the first quartile is a quarter of the way through the data if arranged in order or the 25<sup>th</sup> percentile, the second quartile is half way through the data or the 50<sup>th</sup> percentile (or median), the third quartile is the 75<sup>th</sup> percentile and the fourth quartile is the 100<sup>th</sup> percentile or highest value. The 0<sup>th</sup> quartile is the lowest value.

Quartiles are often called  $Q_0$ ,  $Q_1$ ,  $Q_2$ ,  $Q_3$  and  $Q_4$ .

Other names for them are MIN or MinX, lower quartile, median or Med, upper quartile and MAX or MaxX.



$Q_0$	$Q_1$	$Q_2$	$Q_3$	$Q_4$
zeroth quartile	first quartile	second quartile	third quartile	fourth quartile
MIN MinX	lower quartile	median Med	upper quartile	MAX MaxX

## Practice

- Q3 Use the adult heights ogive above to answer the following questions.
- What is the 4<sup>th</sup> decile height for women?
  - What is the 9<sup>th</sup> decile height for men?
  - What is the first quartile height for women?
  - What is the  $Q_4$  height for men?
  - What is the median height for women?
  - What percentile is the same as the upper quartile?

## Five-Number Summaries

For many purposes, the five quartiles provide a good summary of a set of data, giving an indication of typical values and of the spread (how much variation there is or how spread out it would be if plotted along a number line).

The set of five quartiles is called a **five-number summary**. The five number summary for female adult heights from the ogive above would be: 130, 163, 168, 173, 205.

## Practice

- Q4 Give the five-number summaries for:
- The heights of adult males from the ogive above,
  - The head circumferences of 6-month-old girls from the earlier ogive.
- Q5 Estimate:
- the five-number summary for the heights of people of your age and sex in centimetres,
  - The five number summary for the masses of adults in kilograms.



$Q_3$  is the median of the second half of the data if there is an even number of data, or the median of the second half excluding the middle one if there is an odd number of data.

The five-number summary given above was actually worked out this way.

### Practice

- Q6 Calculate the five-number summaries for the following data sets by hand.
- (a) 12, 14, 15, 15, 17, 18, 20, 21, 21, 25
  - (b) 32, 35, 36, 40, 41, 44, 49, 52, 60
  - (c) 9, 17, 23, 16, 25, 11, 16

### Inter-quartile Range

The inter-quartile range is a commonly-used measure of the spread of data. The inter-quartile range (or IQR) is  $Q_3 - Q_1$ , the difference between the lower and upper quartiles. The inter-quartile range is, for some purpose, a better indicator of spread than the range because it takes into account all the data rather than just two numbers. It is less susceptible than the range to individual numbers which are very different from the majority of the data.

### Practice

- Q7 Find the interquartile range of the data sets in question P6. This won't take long as you have already worked out the five-number summaries.

### Outliers

Sometimes a data set includes a number which is very different from the other data. For instance, if the following are the heights in centimetres of the main cast in a musical:

172, 179, 168, 181, 82, 177, 184, 170,

the 82 would be considered very different from the others. Such numbers are called outliers because they lie outside the normal range.

Outliers often occur because of errors in the data collection or copying. In the case above someone may have left the initial 1 off the height. But they can sometimes be real: there may have been a small child in the musical.

A number is generally considered to be an outlier if it is more than 1.5 times the IQR below Q1 or more than 1.5 times the IQR above Q3.

In the musical example, the Q1 is 169, Q3 is 180, so the IQR is 11, 1.5 times the IQR is 16.5. So anything below 152.5 or above 196.5 would be an outlier. The 82 is therefore definitely an outlier.

## Practice

Q8 List any outliers in the following data sets:

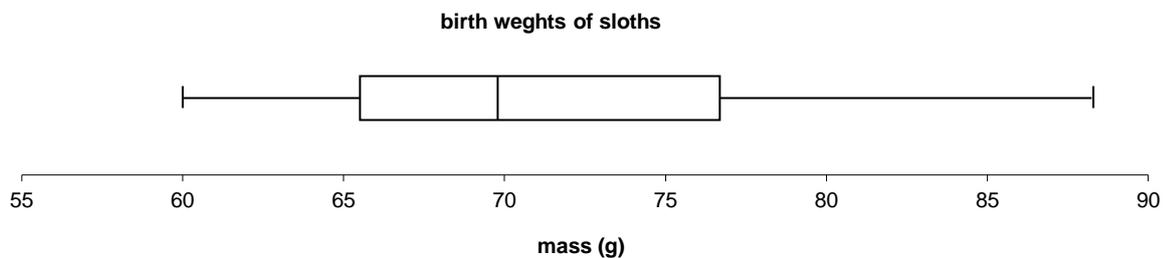
(a) 22, 24, 26, 26, 27, 28, 30, 31, 31, 35, 41

(b) 22, 34, 36, 39, 41, 43, 49, 53, 60

(c) 16, 17, 19, 16, 14, 16, 16, 17

## Box Plots

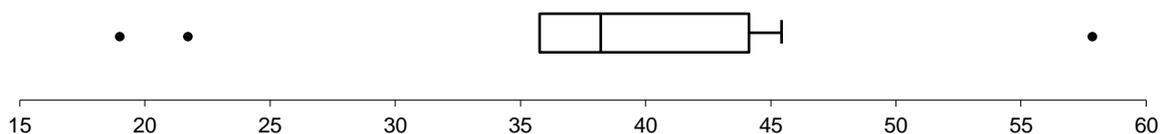
A box plot (often called a box-and-whisker plot) is a graphical representation of a 5-number summary. The box plot for the sloth data above would look like this.



To draw a box plot, draw a suitable scale – one that will cover all the numbers in the data. Then add five vertical lines corresponding to the 5 numbers of the 5-number summary, then add horizontal lines as shown to make the box and the whiskers.

An exception to this procedure is if there are outliers. Outliers are plotted as dots and the ends of the whiskers are then placed at the lowest and highest values that aren't outliers.

So for this data: 19, 22, 36, 37, 37, 39, 41, 44, 46, 58, the five number summary is 19, 36, 38, 44, 58, Q1 is 36, Q3 is 44 and so the IQR is 8. Anything below 24 or above 56 is an outlier. Thus 19, 22 and 58 are outliers and the ends of the whiskers will go at 36 and 44 and the plot will look like this:



## Parallel Box Plots

Box plots are often used to compare two sets of data. Suppose we also had data for the baby sloths born at the Juanita Gonzalez Sloth Farm. We could produce parallel box plots for the two farms like this.



This shows that the babies born at Juanita Gonzales are slightly more varied in mass and, on the whole, slightly heavier.

## Practice

Q9 Draw box plots for each of the following sets of data. You may use your calculator to get the 5-number summary if you wish. Don't forget to check for outliers.

(a) 6, 11, 12, 9, 4, 10, 12, 8, 15, 7, 8, 9

(b) 82, 57, 66, 61, 79, 70, 51, 68, 22, 72, 69

(c) 12.7, 11.9, 12.2, 12.4, 12.5, 11.9, 12.4, 12.6, 12.0, 12.5, 13.2

Q10 Draw parallel box plots to compare the number of days absent in Term 1 for Year 10 students with and without part-time jobs.

With 2, 0, 6, 3, 37, 1, 0, 11, 8, 3

Without 4, 6, 10, 5, 3, 7, 2, 5, 3, 8, 3, 5, 4, 3



## Standard Deviation

As we have seen, the range is a measure of spread. However, it reflects only two bits of data, the highest and lowest values.

The inter-quartile range is a better measure as it takes into account more of the data. However, numbers in the data can change without affecting the inter-quartile range.

The standard deviation is by far the most commonly used measure of spread. It is calculated using all the data and any change to any number will change the standard deviation.

Standard deviation is one of the statistics you get when you ask for the 1-variable statistics on your calculator. On most calculators, it is called  $\sigma n$ .  $\sigma$ , pronounced 'sigma' is a lower-case Greek s. It is the abbreviation generally used for standard deviation.

## Practice

Q11 Use your calculator to find the standard deviations of each of the following sets of data.

(a) 6, 11, 12, 9, 4, 10, 12, 8, 15, 7, 8, 9

(b) 82, 57, 66, 61, 79, 70, 51, 68, 22, 72, 69

(c) 12.7, 11.9, 12.2, 12.4, 12.5, 11.9, 12.4, 12.6, 12.0, 12.5, 13.2



Calculating the standard deviation by hand can be quite time consuming, especially for large data sets. We will show how it is done using a small data set: 11, 17, 9, 13, 15, 13.

Standard deviation is sometimes called *root mean square deviation*. If you remember this, it will tell you the steps you need to go through to calculate it. These are the steps:

1. Work out the mean of the data
2. Work out the deviation for each number. The deviation is how far the number is from the mean. Specifically, it is the number minus the mean
3. Square the deviations
4. Calculate the mean of the square deviations
5. take the square root of the mean square deviation

Hence the name *root mean square deviation*.

This gives us something like the average deviation of the data from the mean.

Using a table for the calculations is probably the easiest way.

$$\text{Mean} = (11 + 17 + 9 + 13 + 15 + 13) \div 6 = 13$$

Data	11	17	9	13	15	13
Deviation	-2	4	-4	0	2	0
Squared deviation	4	16	16	0	4	0

$$\text{Mean squared deviation} = (4 + 16 + 16 + 0 + 4 + 0) \div 6 = 6.67$$

$$\text{Root mean square deviation} = \sqrt{6.67} = 2.58$$

$$\text{Standard deviation} = 2.58.$$

## Practice

Q12 Find the standard deviations of the following set of data by hand.

3.5, 11.1, 8.7, 7.9, 6.0

## *Standard Deviations from Samples*

On the calculator,  $x\sigma n$  gives the standard deviation of the data you enter. But the calculator also gives another standard deviation,  $x\sigma n-1$ . This standard deviation is used if the data you enter is a sample of a larger population and if you want an estimate of the standard deviation of the whole population.

## *Variance*

Variance is a word for the square of the standard deviation, i.e. the value you get before the last step of taking the square root.

---

---

## Solve

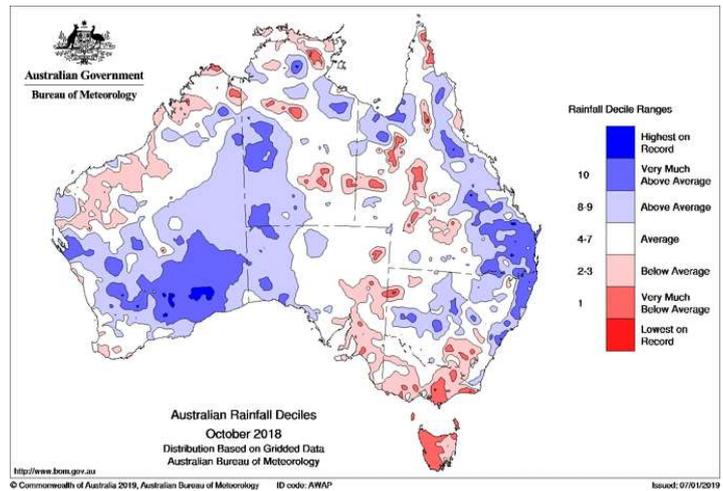
---

---

Q51 Use the ogives of adult heights above to work out what percentage of women are taller than the third quartile man.

Q52 Use the ogives of adult heights to sketch a histogram of men's heights. Use intervals 150-160 cm, 160-170 cm etc.

Q53 Quantiles are sometimes used in a way different from the standard way described in this module: they are used to refer to ranges rather than points along the number line. For example, in the rainfall map to the right, the mid-blue colour labelled as the 10<sup>th</sup> decile actually refers to the range of rainfalls from D<sub>9</sub> (decile 9) to D<sub>10</sub>. Only the dark blue is actually at D<sub>10</sub>.

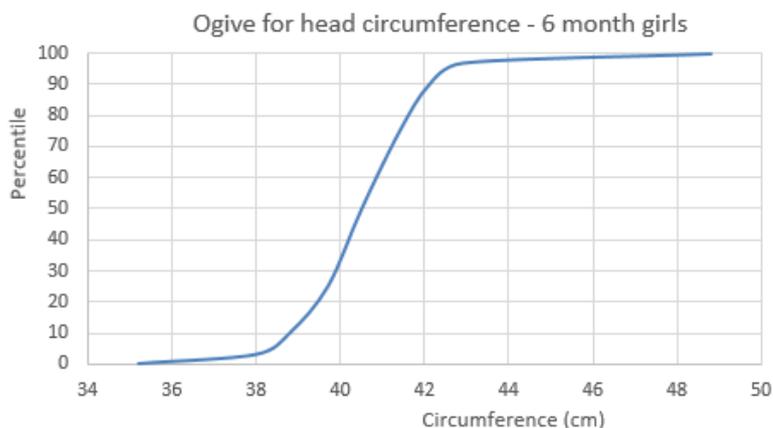


- (a) What range of deciles do the light pink areas refer to?  
 (b) What decile range do the darkest red areas refer to?

## Revise

### Revision Set 1

Q61 Use the ogive below to answer the following questions.



- (a) What circumference is the 60<sup>th</sup> percentile?  
 (b) What circumference is the 2<sup>nd</sup> decile?  
 (c) What circumference is the upper quartile?  
 (d) What percentile is a circumference of 40.0 cm?  
 (e) What decile is a circumference of 38.9 cm?  
 (f) Between which quartiles is a circumference of 38.4 cm?  
 (g) What is the median circumference?  
 (h) What is the largest circumference?  
 (i) Find the five-number summary.

- (j) Draw a box and whisker plot.
- (k) Find the inter-quartile range.

**Q62** The following are the numbers of times the patients in Ward 12 called for a nurse yesterday: 5, 2, 13, 4, 0, 1, 6, 5, 3, 0, 3, 8.

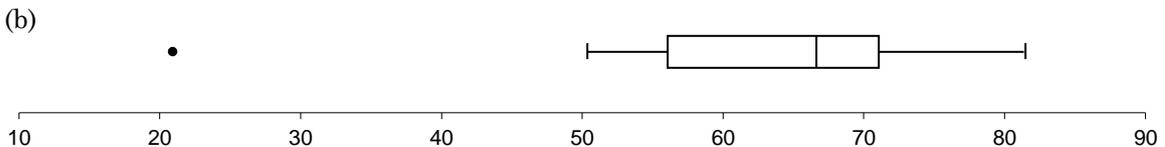
Use a calculator to find:

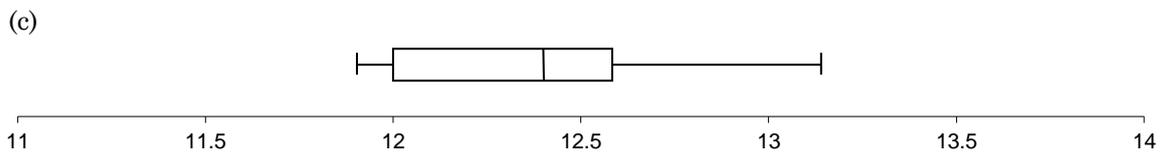
- (a) the mean
- (b) the five-number summary
- (c) the inter-quartile range
- (d) the standard deviation for the ward
- (e) the standard deviation for the whole hospital assuming that Ward 12 is a representative sample.



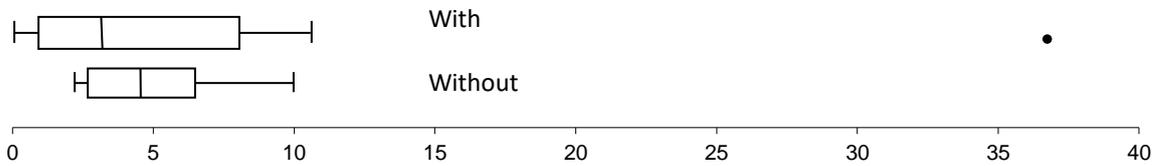
## Answers

- Q1** (a) Sharni (b) 38.8 cm (c) 40.5 cm (d) 75<sup>th</sup>  
 (e) larger (f) ~84<sup>th</sup> (~ means approximately)  
 (g) ~5<sup>th</sup> (h) ~98<sup>th</sup> (i) ~39.4 cm (j) 40.5 cm  
 (k) ~41 cm (l) 87<sup>th</sup> (m) ~42.0 cm
- Q2** (a) ~40.9 cm (b) ~35 cm (c) ~15<sup>th</sup> (d) ~99<sup>th</sup>  
 (e) ~40.5 cm
- Q3** (a) ~166 cm (b) ~185 cm (c) ~163 cm (d) ~215 cm  
 (e) ~168 cm (f) 75<sup>th</sup>
- Q4** (a) 150, 172, 176, 181, 215 (approx.) (b) 35.0, 39.7, 40.6, 41.4, 49.0 (approx.)
- Q5** (a) Depends on your age and sex  
 (b) Maybe something like 38, 59, 70, 92, 225
- Q6** (a) 12, 15, 17.5, 21, 25 (b) 32, 35.5, 41, 50.5, 60 (c) 9, 11, 16, 23, 25
- Q7** (a) 6 (b) 15 (c) 12
- Q8** (a) 41 (b) none (c) 14, 19
- Q9** (a)





Q10 Days absent for students with and without part-time jobs

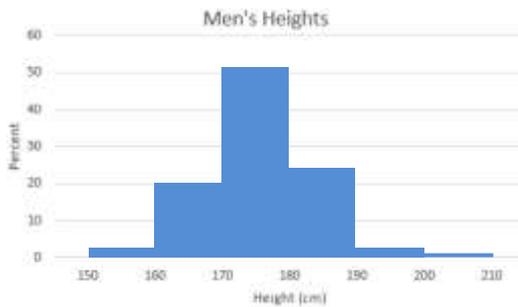


Q11 (a) 2.86 (b) 15.61 (c) 2.17

Q12 2.56

Q51 ~7%

Q52



Q53 (a) 2 to 4 (b)  $D_0$  (no range)

Q61 (a) ~40.8 cm (b) ~39.4 cm (c) ~41.4 cm (d) ~33<sup>rd</sup>  
 (e) ~1<sup>st</sup> (f)  $Q_0$  and  $Q_1$  (g) ~40.6 cm (h) ~49.0 cm  
 (i) Approx 35.0, 39.7, 40.6, 41.3, 49.0  
 (j)



Q62 (a) 4.17 (b) 0, 1.5, 3.5, 5.5, 13 (c) 4.0  
 (d) 3.53 (e) 3.69