Stats1 Chapter 2: Measures of Data

Measures of Spread

Quartiles – which item?

You need to be able to find the **quartiles** of both listed data and of grouped data. The rule is exactly the same as for the median.

Listed data

Items	n	Position of LQ & UQ	LQ & UQ
1,4,7,9,10	5	2nd & 4th	4 & 9
4,9,10,15	4	1st/2nd & 3rd/4th	6.5 & 12.5
2,4,5,7,8,9,11	7	2nd & 6th	4 & 9
1,2,3,5,6,9,9,10,11,12	10	3rd and 8th	3 & 10

Can you think of a rule to find the position of the LQ/UQ given n?

To find the position of the LQ/UQ for listed data, find $\frac{1}{4}n$ or $\frac{3}{4}n$ then as before:
- If a decimal, round up.
- If whole, use halfway between this item and the one after.

Grouped data

IQ of L6Ms2 (q)	Frequency (f)
$80 \leq q < 90$	7
$90 \leq q < 100$	5
$100 \leq q < 120$	3
$120 \leq q < 200$	2

Position to use for LQ:

4.25

To find the LQ and UQ of grouped data, find $\frac{1}{4}n$ and $\frac{3}{4}n$, then use linear interpolation.

Again, **DO NOT** round this value.

Percentiles

The LQ, median and UQ give you 25%, 50% and 75% along the data respectively. But we can have any percentage you like!

$n = 43$

Item to use for 57th percentile?

$\mathbf{43 \times 0.57 = 24.51}$

You will always find these for grouped data in an exam, **so never round this position**.

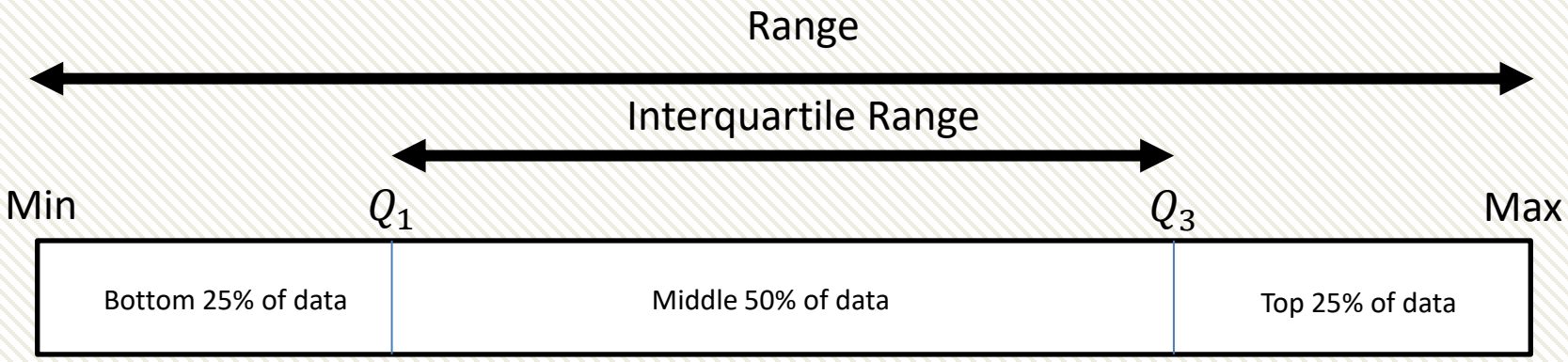
Notation:

Lower Quartile: Q_1 Median: Q_2

Upper Quartile: Q_3 57th Percentile: P_{57}

Measures of Spread

The interquartile range and interpercentile range are examples of **measures of spread.**



Interquartile Range = Upper Quartile − Lower Quartile

Why might we favour the interquartile range over the range?
Because it gives us the spread of the data excluding the extreme values at either end.

We can control this further by having for example the "*10th to 90th interpercentile range*", which would be $P_{90} - P_{10}$. This would typically be symmetrical about the median, so that we could interpret this as "*the range of the data with the most extreme 10% of values at either end excluded*".
The 10th percentile is also known as the 1st **decile** (D_1), and similarly $P_{90} = D_9$.

Test Your Understanding

These are the same as the 'Test Your Understanding' questions on your supplementary sheet from before.

Age of relic (years)	Frequency
0-1000	24
1001-1500	29
1501-1700	12
1701-2000	35

Item for Q_1:

?

$Q_1 =$?

$Q_3 =$?

$IQR =$?

Shark length (cm)	Frequency
$40 \leq x < 100$	17
$100 \leq x < 300$	5
$300 \leq x < 600$	8
$600 \leq x < 1000$	11

Item to use for P_{10}: ?

$P_{10} =$?

$P_{90} =$?

10th to 90th interpercentile range:

?

Test Your Understanding

These are the same as the 'Test Your Understanding' questions on your supplementary sheet from before.

Age of relic (years)	Frequency
0-1000	24
1001-1500	29
1501-1700	12
1701-2000	35

Item for Q_1:
25th

$$Q_1 = 1000.5 + \left(\frac{1}{29} \times 500\right)$$
$$= 1017.74 \text{ years}$$
$$Q_3 = 1700.5 + \left(\frac{10}{35} \times 300\right)$$
$$= 1786.21 \text{ years}$$
$$IQR = 768.47$$

Shark length (cm)	Frequency
$40 \leq x < 100$	17
$100 \leq x < 300$	5
$300 \leq x < 600$	8
$600 \leq x < 1000$	11

Item to use for P_{10}: $\mathbf{41 \times 0.1 = 4.1}$

$$P_{10} = \mathbf{40} + \left(\frac{\mathbf{4.1}}{\mathbf{17}} \times \mathbf{60}\right) = \mathbf{54.47}$$
$$P_{90} = \mathbf{600} + \left(\frac{\mathbf{6.9}}{\mathbf{11}} \times \mathbf{400}\right) = \mathbf{850.91}$$

10th to 90th interpercentile range:
$\mathbf{850.91 - 54.47 = 796.44}$

Supplementary Exercise 2

Q1) May 2013 Q4 (continued)

The following table summarises the times, t minutes to the nearest minute, recorded for a group of students to complete an exam.

Time (minutes) t	11 – 20	21 – 25	26 – 30	31 – 35	36 – 45	46 – 60
Number of students f	62	88	16	13	11	10

(c) Show that the estimated value of the lower quartile is 18.6 to 3 significant figures. **(1)**

(d) Estimate the interquartile range of this distribution. **(2)**

?

Q2) June 2005 Q2

The following table summarises the distances, to the nearest km, that 134 examiners travelled to attend a meeting in London.

Distance (km)	Number of examiners
41–45	4
46–50	19
51–60	53
61–70	37
71–90	15
91–150	6

(c) Use interpolation to estimate the median Q_2, the lower quartile Q_1, and the upper quartile Q_3 of these data.

?

Supplementary Exercise 2

Q1) May 2013 Q4 (continued)

The following table summarises the times, t minutes to the nearest minute, recorded for a group of students to complete an exam.

Time (minutes) t	11 – 20	21 – 25	26 – 30	31 – 35	36 – 45	46 – 60
Number of students f	62	88	16	13	11	10

(c) Show that the estimated value of the lower quartile is 18.6 to 3 significant figures. **(1)**

(d) Estimate the interquartile range of this distribution. **(2)**

(c) $Q_1 = 10.5 + \frac{(50/50.25)}{62} \times 10 [= 18.56]$ (*) ($n+1$ gives 18.604…) B1 cso **(1)**

(d) $Q_3 = 25.5$ (Use of $n+1$ gives 25.734…) B1
IQR = 6.9 (Use of $n+1$ gives 7.1) B1 ft **(2)**

Q2) June 2005 Q2

The following table summarises the distances, to the nearest km, that 134 examiners travelled to attend a meeting in London.

Distance (km)	Number of examiners
41–45	4
46–50	19
51–60	53
61–70	37
71–90	15
91–150	6

(c) Use interpolation to estimate the median Q_2, the lower quartile Q_1, and the upper quartile Q_3 of these data.

$Q_2 = 50.5 + \frac{(67-23)}{53} \times 10 = 58.8$ awrt 58.8/58.9 M1 A1

$Q_1 = 52.48$; $Q_3 = 67.12$ awrt 52.5/52.6 67.1/67.3 A1 A1

Special case : no working B1 B1 B1 (≡ A's on the epen)

Supplementary Exercise 2

Q3)

The ages of 300 houses in a village are recorded given the following table of results.

Age a (years)	Number of houses
$0 \leq a < 20$	36
$20 \leq a < 40$	92
$40 \leq a < 60$	74
$60 \leq a < 100$	39
$100 \leq a < 200$	14
$200 \leq a < 300$	27
$300 \leq a < 500$	18

Use linear interpolation to estimate the lower quartile, upper quartile and hence the interquartile range.

?

Q4)

A cyber-café recorded how long each user stayed during one day giving the following results.

Length of stay (minutes)	Number of houses
$0 \leq l < 30$	15
$30 \leq l < 60$	31
$60 \leq l < 90$	32
$90 \leq l < 120$	23
$120 \leq l < 240$	17
$240 \leq l < 360$	2

Use linear interpolation to estimate:

a) The lower quartile.
b) The upper quartile.
c) The 90th percentile.

?

Q5)

Distance (to the nearest mile)	Number of commuters
0 – 9	10
10 – 19	19
20 – 29	43
30 – 39	25
40 – 49	8
50 – 59	6
60 – 69	5
70 – 79	3
80 – 89	1

Find the interquartile range for the distance travelled by commuters.

?

Supplementary Exercise 2

Q3)

The ages of 300 houses in a village are recorded given the following table of results.

Age a (years)	Number of houses
$0 \leq a < 20$	36
$20 \leq a < 40$	92
$40 \leq a < 60$	74
$60 \leq a < 100$	39
$100 \leq a < 200$	14
$200 \leq a < 300$	27
$300 \leq a < 500$	18

Use linear interpolation to estimate the lower quartile, upper quartile and hence the interquartile range.

$$Q_1 = \mathbf{20} + \left(\frac{\mathbf{39}}{\mathbf{92}} \times \mathbf{20}\right) = \mathbf{28.5}$$

$$Q_3 = \mathbf{60} + \left(\frac{\mathbf{23}}{\mathbf{39}} \times \mathbf{40}\right) = \mathbf{83.6}$$

$$IQR = \mathbf{55.1}$$

Q4)

A cyber-café recorded how long each user stayed during one day giving the following results.

Length of stay (minutes)	Number of houses
$0 \leq l < 30$	15
$30 \leq l < 60$	31
$60 \leq l < 90$	32
$90 \leq l < 120$	23
$120 \leq l < 240$	17
$240 \leq l < 360$	2

Use linear interpolation to estimate:

a) The lower quartile.
b) The upper quartile.
c) The 90th percentile.

$$Q_1 = \mathbf{30} + \left(\frac{\mathbf{15}}{\mathbf{31}} \times \mathbf{30}\right) = \mathbf{44.5}$$

$$Q_3 = \mathbf{90} + \left(\frac{\mathbf{12}}{\mathbf{23}} \times \mathbf{30}\right) = \mathbf{105.7}$$

$$P_{90} = \mathbf{120} + \left(\frac{\mathbf{7}}{\mathbf{17}} \times \mathbf{120}\right) = \mathbf{169.4}$$

Q5)

Distance (to the nearest mile)	Number of commuters
0 – 9	10
10 – 19	19
20 – 29	43
30 – 39	25
40 – 49	8
50 – 59	6
60 – 69	5
70 – 79	3
80 – 89	1

Find the interquartile range for the distance travelled by commuters.

$$Q_1 = \mathbf{19.5} + \left(\frac{\mathbf{1}}{\mathbf{43}} \times \mathbf{10}\right) = \mathbf{19.7}$$

$$Q_3 = \mathbf{29.5} + \left(\frac{\mathbf{18}}{\mathbf{25}} \times \mathbf{10}\right) = \mathbf{36.7}$$

$$IQR = \mathbf{17.0}$$

Exercises 2.3

Pearson Statistics & Mechanics Year 1/AS
Pages 10-11

Homework Exercise

1 From the large data set, the daily mean pressure (hPa) during the last 16 days of July 2015 in Perth is recorded. The data is given below:

1024	1022	1021	1013	1009	1018	1017	1024
1027	1029	1031	1025	1017	1019	1017	1014

a Find the median pressure for that period.

b Find the lower and upper quartiles.

2 Rachel records the number of CDs in the collections of students in her year. The results are in the table below.

Number of CDs	35	36	37	38	39
Frequency	3	17	29	34	12

Find Q_1, Q_2 and Q_3.

Hint This an ungrouped frequency table so you do not need to use interpolation. Use the rules for finding the median and quartiles of **discrete** data.

3 A hotel is worried about the reliability of its lift. It keeps a weekly record of the number of times it breaks down over a period of 26 weeks. The data collected is summarised in the table opposite.

Use interpolation to estimate the median number of breakdowns.

Number of breakdowns	**Frequency**
0–1	18
2–3	7
4–5	1

(2 marks)

4 The weights of 31 Jersey cows were recorded to the nearest kilogram. The weights are shown in the table.

Weight of cattle (kg)	300–349	350–399	400–449	450–499	500–549
Frequency	3	6	10	7	5

a Find an estimate for the median weight.

b Find the lower quartile, Q_1.

c Find the upper quartile, Q_3.

d Interpret the meaning of the value you have found for the upper quartile in part **c**.

Homework Exercise

5 A roadside assistance firm kept a record over a week of the amount of time, in minutes, people were kept waiting for assistance. The times are shown below.

Time waiting, t (minutes)	$20 \leqslant t < 30$	$30 \leqslant t < 40$	$40 \leqslant t < 50$	$50 \leqslant t < 60$	$60 \leqslant t < 70$
Frequency	6	10	18	13	2

a Find an estimate for the mean wait time. **(1 mark)**

b Calculate the 65th percentile. **(2 marks)**

The firm writes the following statement for an advertisement:

> Only 10% of our customers have to wait longer than 56 minutes.

c By calculating a suitable percentile, comment on the validity of this claim. **(3 marks)**

6 The table shows the recorded wingspans, in metres, of 100 endangered Californian condors.

Wingspan, w (m)	$1.0 \leqslant w < 1.5$	$1.5 \leqslant w < 2.0$	$2.0 \leqslant w < 2.5$	$2.5 \leqslant w < 3.0$	$3.0 \leqslant w$
Frequency	4	20	37	28	11

a Estimate the 80th percentile and interpret the value. **(3 marks)**

b State why it is not possible to estimate the 90th percentile. **(1 marks)**

Homework Exercise

7 The lengths of a number of slow worms were measured, to the nearest mm.
The results are shown in the table.

Lengths of slow worms (mm)	Frequency
125–139	4
140–154	4
155–169	2
170–184	7
185–199	20
200–214	24
215–229	10

a Work out how many slow worms were measured.

b Estimate the interquartile range for the lengths of the slow worms.

c Calculate an estimate for the mean length of slow worms.

d Estimate the number of slow worms whose length is more than one interquartile range above the mean.

Problem-solving

For part **d**, work out $\bar{x}$ + IQR, and determine which class interval it falls in. Then use proportion to work out how many slow worms from that class interval you need to include in your estimate.

8 The table shows the monthly income for workers in a factory.

Monthly income, x (£)	$900 \leqslant x < 1000$	$1000 \leqslant x < 1100$	$1100 \leqslant x < 1200$	$1200 \leqslant x < 1300$
Frequency	3	24	28	15

a Calculate the 34% to 66% interpercentile range. **(3 marks)**

b Estimate the number of data values that fall within this range. **(2 marks)**

Homework Exercise

9 A train travelled from Lancaster to Preston. The times, to the nearest minute, it took for the journey were recorded over a certain period. The times are shown in the table.

Time for journey (minutes)	15–16	17–18	19–20	21–22
Frequency	5	10	35	10

a Calculate the 5% to 95% interpercentile range. **(3 marks)**

b Estimate the number of data values that fall within this range. **(1 mark)**

10 From the large data set, the daily mean temperature (°C) for Leeming during the first 10 days of June 1987 is given below:

14.3 12.7 12.4 10.9 9.4 13.2 12.1 10.3 10.3 10.6

a Calculate the median and interquartile range. **(2 marks)**

The median daily mean temperature for Leeming during the first 10 days of May 1987 was 9.9 °C and the interquartile range was 3.9 °C.

b Compare the data for May with the data for June. **(2 marks)**

The 10% to 90% interpercentile range for the daily mean temperature for Leeming during July 1987 was 5.4 °C.

c Estimate the number of days in July 1987 on which the daily mean temperature fell within this range. **(1 mark)**

Homework Answers

1 **a** 1020 hPa **b** $Q_1 = 1017$ hPa, $Q_3 = 1024.5$ hPa

2 Median 37, $Q_1 = 37$, $Q_3 = 38$

3 1.08

4 **a** 432 kg **b** 389 kg **c** 480 kg

d Three-quarters of the cows weigh 480 kg or less.

5 **a** 44.0 minutes **b** 48.8 minutes

c 90th percentile = 57.8 minutes so 10% of customers have to wait longer 57.8 minutes, not 56 minutes as stated by the firm.

6 **a** 2.84 m. 80% of condors have a wingspan of less than 2.84 m.

b The 90th percentile is in the $3.0 \leqslant w$ class. There is no upper boundary for this class, so it is not possible to estimate the 90th percentile.

7 **a** 71 **b** 24.6 **c** 193.1 mm **d** 7

8 **a** £81.87 **b** 22

9 **a** 6.2 minutes **b** 54

10 **a** Median 11.5 °C, $Q_1 = 10.3$ °C, $Q_3 = 12.7$ °C, IQR = 2.4 °C

b On average, the temperature was higher in June than in May (higher median). The temperature was more variable in May than June (higher IQR).

c 24 days