
S1: Processing data, variance and standard deviation

Past Paper Questions 2006 -
2013

Name:

- 4 The time, x seconds, spent by each of a random sample of 100 customers at an automatic teller machine (ATM) is recorded. The times are summarised in the table.

Time (seconds)	Number of customers
$20 < x \leq 30$	2
$30 < x \leq 40$	7
$40 < x \leq 60$	18
$60 < x \leq 80$	27
$80 < x \leq 100$	23
$100 < x \leq 120$	13
$120 < x \leq 150$	7
$150 < x \leq 180$	3
Total	100

- (a) Calculate estimates for the mean and standard deviation of the time spent at the ATM by a customer. *(4 marks)*
- (b) The mean time spent at the ATM by a random sample of 36 customers is denoted by \bar{Y} .
- (i) State why the distribution of \bar{Y} is approximately normal. *(1 mark)*
- (ii) Write down estimated values for the mean and standard error of \bar{Y} . *(2 marks)*
- (iii) Hence estimate the probability that \bar{Y} is less than $1\frac{1}{2}$ minutes. *(3 marks)*

- 1 The times, in seconds, taken by 20 people to solve a simple numerical puzzle were

17 19 22 26 28 31 34 36 38 39
41 42 43 47 50 51 53 55 57 58

- (a) Calculate the mean and the standard deviation of these times. *(3 marks)*
- (b) In fact, 23 people solved the puzzle. However, 3 of them failed to solve it within the allotted time of 60 seconds.
- Calculate the median and the interquartile range of the times taken by all 23 people. *(4 marks)*
- (c) For the times taken by all 23 people, explain why:
- (i) the mode is **not** an appropriate numerical measure;
- (ii) the range is **not** an appropriate numerical measure. *(2 marks)*

4 A library allows each member to have up to 15 books on loan at any one time.

The table shows the numbers of books currently on loan to a random sample of 95 members of the library.

Number of books on loan	0	1	2	3	4	5–9	10–14	15
Number of members	4	13	24	17	15	11	5	6

(a) For these data:

(i) state values for the mode and range; (2 marks)

(ii) determine values for the median and interquartile range; (4 marks)

(iii) calculate estimates of the mean and standard deviation. (4 marks)

(b) Making reference to your answers to part (a), give a reason for preferring:

(i) the median and interquartile range to the mean and standard deviation for summarising the given data; (1 mark)

(ii) the mean and standard deviation to the mode and range for summarising the given data. (1 mark)

1 The average maximum monthly temperatures, u degrees Fahrenheit, and the average minimum monthly temperatures, v degrees Fahrenheit, in New York City are as follows.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Maximum (u)	39	40	48	61	71	81	85	83	77	67	54	41
Minimum (v)	26	27	34	44	53	63	68	66	60	51	41	30

(a) (i) Calculate, to one decimal place, the mean and the standard deviation of the 12 values of the average **maximum** monthly temperature. (2 marks)

(ii) For comparative purposes with a UK city, it was necessary to convert the temperatures from degrees Fahrenheit ($^{\circ}\text{F}$) to degrees Celsius ($^{\circ}\text{C}$). The formula used to convert f $^{\circ}\text{F}$ to c $^{\circ}\text{C}$ is:

$$c = \frac{5}{9}(f - 32)$$

Use this formula and your answers in part (a)(i) to calculate, in $^{\circ}\text{C}$, the mean and the standard deviation of the 12 values of the average maximum monthly temperature. (3 marks)

(b) The value of the product moment correlation coefficient, r_{uv} , between the above 12 values of u and v is 0.997, correct to three decimal places.

State, giving a reason, the corresponding value of r_{xy} , where x and y are the exact equivalent temperatures in $^{\circ}\text{C}$ of u and v respectively. (2 marks)

- 6 For each of the Premiership football seasons 2004/05 and 2005/06, a count is made of the number of goals scored in each of the 380 matches. The results are shown in the table.

Number of goals scored in a match	Number of matches	
	2004/05	2005/06
0	30	32
1	79	82
2	99	95
3	68	78
4	60	48
5	24	30
6	11	9
7	6	6
8	2	0
9	1	0
Total	380	380

- (a) For the number of goals scored in a match during the **2004/05** season:
- determine the median and the interquartile range; *(4 marks)*
 - calculate the mean and the standard deviation. *(4 marks)*
- (b) Two statistics students, Jole and Katie, independently analyse the data on the number of goals scored in a match during the **2005/06** season.
- Jole determines correctly that the median is 2 and that the interquartile range is also 2.
 - Katie calculates correctly, to two decimal places, that the mean is 2.48 and that the standard deviation is 1.59.
- Use your answers from part (a), together with Jole's and Katie's results, to compare briefly the two seasons with regard to the average and the spread of the number of goals scored in a match. *(2 marks)*
 - Jole claims that Katie's results must be wrong as 95% of values always lie within 2 standard deviations of the mean and $(2.48 - 2 \times 1.59) < 0$ which is nonsense.

Explain why Jole's claim is incorrect. (You are not expected to confirm Katie's results.) *(2 marks)*

4 The runs scored by a cricketer in 11 innings during the 2006 season were as follows.

47 63 0 28 40 51 a 77 0 13 35

The exact value of a was unknown but it was greater than 100.

- (a) Calculate the median and the interquartile range of these 11 values. *(4 marks)*
- (b) Give a reason why, for these 11 values:
- (i) the mode is **not** an appropriate measure of average;
- (ii) the range is **not** an appropriate measure of spread. *(2 marks)*

1 Ms N Parker always reads the columns of announcements in her local weekly newspaper. During each week of 2008, she notes the number of births announced. Her results are summarised in the table.

Number of births	1	2	3	4	5	6	7	8
Number of weeks	1	2	9	13	7	13	6	1

- (a) Calculate the mean, median and modes of these data. *(5 marks)*
- (b) State, with a reason, which of the three measures of average in part (a) you consider to be the **least** appropriate for summarising the number of births. *(2 marks)*

- 5** A survey of all the households on an estate is undertaken to provide information on the number of children per household.

The results, for the 99 households with children, are shown in the table.

Number of children (x)	1	2	3	4	5	6	7
Number of households (f)	14	35	25	13	9	2	1

- (a)** For these 99 households, calculate values for:
- (i)** the median and the interquartile range; *(3 marks)*
 - (ii)** the mean and the standard deviation. *(3 marks)*
- (b)** In fact, 163 households were surveyed, of which 64 contained no children.
- (i)** For all 163 households, calculate the value for the mean number of children per household. *(2 marks)*
 - (ii)** State whether the value for the standard deviation, when calculated for all 163 households, will be smaller than, the same as, or greater than that calculated in part **(a)(ii)**. *(1 mark)*
 - (iii)** It is claimed that, for all 163 households on the estate, the number of children per household may be modelled approximately by a normal distribution.

Comment, with justification, on this claim. Your comment should refer to a fact other than the discrete nature of the data. *(2 marks)*

January 2010

- 2** Lizzie, the receptionist at a dental practice, was asked to keep a weekly record of the number of patients who failed to turn up for an appointment. Her records for the first 15 weeks were as follows.

20 26 32 a 37 14 27 34 15 18 b 25 37 29 25

Unfortunately, Lizzie forgot to record the actual values for two of the 15 weeks, so she recorded them as a and b . However, she did remember that $a < 10$ and that $b > 40$.

- (a)** Calculate the median and the interquartile range of these 15 values. *(4 marks)*
- (b)** Give a reason why, for these data:
 - (i)** the mode is **not** an appropriate measure of average;
 - (ii)** the standard deviation **cannot** be used as a measure of spread. *(2 marks)*
- (c)** Subsequent investigations revealed that the missing values were 8 and 43.

Calculate the mean and the standard deviation of the 15 values. *(2 marks)*

- 2** Before leaving for a tour of the UK during the summer of 2008, Eduardo was told that the UK price of a 1.5-litre bottle of spring water was about 50p.
- Whilst on his tour, Eduardo noted the prices, x pence, which he paid for 1.5-litre bottles of spring water from 12 retail outlets.
- He then subtracted 50p from each price and his resulting differences, in pence, were
- 18 -11 1 15 7 -1 17 -16 18 -3 0 9
- (a) (i)** Calculate the mean and the standard deviation of these differences. *(2 marks)*
- (ii)** Hence calculate the mean and the standard deviation of the prices, x pence, paid by Eduardo. *(2 marks)*
- (b)** Based on an exchange rate of €1.22 to £1, calculate, in euros, the mean and the standard deviation of the prices paid by Eduardo. *(3 marks)*

- 1** Giles, a keen gardener, rents a council allotment.
- During early April 2011, he planted 27 seed potatoes.
- When he harvested his potato crop during the following August, he counted the number of new potatoes that he obtained from each seed potato.
- He recorded his results as follows.
- | | | | | | | | |
|-------------------------------|----------|---|---|---|----|----|-----------|
| Number of new potatoes | ≤ 6 | 7 | 8 | 9 | 10 | 11 | ≥ 12 |
| Frequency | 2 | 2 | 1 | 4 | 8 | 6 | 4 |
- (a)** Calculate values for the median and the interquartile range of these data. *(3 marks)*
- (b)** Advise Giles on how to record his corresponding data for 2012 so that it would then be possible to calculate the mean number of new potatoes per seed potato. *(1 mark)*

- 1** The number of matches in each of a sample of 85 boxes is summarised in the table.

Number of matches	Number of boxes
Less than 239	1
239–243	1
244–246	2
247	3
248	4
249	6
250	10
251	13
252	16
253	20
254	5
255–259	3
More than 259	1
Total	85

- (a)** For these data:
- (i)** state the modal value; *(1 mark)*
- (ii)** determine values for the median and the interquartile range. *(3 marks)*
- (b)** Given that, on investigation, the 2 extreme values in the above table are 227 and 271 :
- (i)** calculate the range; *(1 mark)*
- (ii)** calculate estimates of the mean and the standard deviation. *(4 marks)*
- (c)** For the numbers of matches in the 85 boxes, suggest, with a reason, the most appropriate measure of spread. *(2 marks)*

2 Katy works as a clerical assistant for a small company. Each morning, she collects the company's post from a secure box in the nearby Royal Mail sorting office.

Katy's supervisor asks her to keep a daily record of the number of letters that she collects.

Her records for a period of 175 days are summarised in the table.



Daily number of letters (x)	Number of days (f)
0–9	5
10–19	16
20	23
21	27
22	31
23	34
24	16
25–29	10
30–34	5
35–39	3
40–49	4
50 or more	1
Total	175

- (a) For these data:
- (i) state the modal value; *(1 mark)*
 - (ii) determine values for the median and the interquartile range. *(3 marks)*
- (b) The most letters that Katy collected on any of the 175 days was 54. Calculate estimates of the mean and the standard deviation of the daily number of letters collected by Katy. *(4 marks)*
- (c) During the same period, a total of 280 letters was also delivered to the company by private courier firms.

Calculate an estimate of the mean daily number of **all** letters received by the company during the 175 days. *(2 marks)*

Temperatures and Maximum Weights

