

Please note that you should round your answers to an appropriate level of accuracy.

Question 1 (Unit 3) – 9 marks

The Minitab file that you require for this question should be downloaded from the 'Assessment' area on the module website.

The Broadcasters' Audience Research Board (BARB) publishes weekly statistics on average weekly UK TV set viewing hours.

The Minitab file ukTV 20j.mwx is based on the data for the years 2017 and 2018. The column labelled Week gives the week number from 0 to 51. The column labelled 2017 contains the UK average weekly TV viewing hours for 2017, and the column labelled 2018 contains the viewing hours for 2018. Run Minitab and open this worksheet.

- (a) Produce horizontal boxplots of the data for the viewing figures for 2017 and 2018 on the same diagram using Minitab. Include informative labels for the horizontal and vertical axes as well as an informative title for the boxplots. Include a copy of your plots in your answer. [3]
- (b) On the basis of the boxplots obtained in part (a), are the viewing figures for 2018 left-skew, right-skew or neither? Explain your answer. [3]
- (c) Using the two boxplots you obtained in part (a), compare the viewing figures for 2017 and 2018. [3]

Question 2 (Unit 3) – 4 marks

The Department for Transport (DfT) publishes annual statistics on the number of cars and vans per household in England. In 2017, the DfT carried out a sample survey of 6846 households. The figures in Table 1 are based on the results of that survey.

Table 1 Numbers of households
by number of cars or vans

Number of cars or vans	Number of households
0	1634
1	2801
2	2411
Total	6846

Using Table 1, calculate by hand to two decimal places – that is, not using Minitab – the mean and standard deviation of the number of cars or vans in these households.

You should show each step of your calculations including:

- the batch size
- the sum of the data values
- the sum of the squares of the data values
- the mean
- the sum of the squares of the deviations
- the variance
- the standard deviation. [4]

Question 3 (Unit 3) – 4 marks

In Unit 3, the annual change in real earnings is explored. In this question we look at more recent trends in real earnings. Average weekly earnings are published regularly by the Office for National Statistics, and these include indices for different sectors of the economy. Table 2 shows recent values of the Consumer Prices Index (CPI) and the Average Weekly Earnings (AWE) index for the Finance and Business Services sector of the economy. The CPI figures are available from

www.ons.gov.uk/economy/inflationandpriceindices/datasets/consumerpriceindices. The AWE figures are available from www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/averageweeklyearningsearn01.

Table 2 Values of the CPI and the AWE index for the Finance and Business Services sector

Table 2 Values of the CPI and the AWE index for the Finance and Business Services sector

	CPI		AWE Index	
	2018	2019	2018	2019
January	104.4	106.3	170.2	178.7
February	104.9	106.8	171.4	177.2
March	105.0	107.0	173.2	176.9
April	105.4	107.6	171.0	179.7
May	105.8	107.9	171.3	180.5
June	105.8	107.9	172.4	181.0
July	105.8	107.9	174.3	181.3
August	106.5	108.4	174.3	182.0
September	106.6	108.5	174.3	183.2
October	106.7	108.3	177.0	183.1
November	107.0	108.5	176.8	183.9
December	107.1	108.5	177.2	183.5

----- the same

CPI AWE Index
2018 2019 2018 2019

January 104.4 106.3 170.2 178.7
February 104.9 106.8 171.4 177.2
March 105.0 107.0 173.2 176.9
April 105.4 107.6 171.0 179.7
May 105.8 107.9 171.3 180.5
June 105.8 107.9 172.4 181.0
July 105.8 107.9 174.3 181.3
August 106.5 108.4 174.3 182.0
September 106.6 108.5 174.3 183.2
October 106.7 108.3 177.0 183.1
November 107.0 108.5 176.8 183.9
December 107.1 108.5 177.2 183.5

- (a) For each of November 2019 and December 2019, calculate, as a percentage, the real earnings for that month compared with one year earlier. Give your answer to one decimal place. [2]
- (b) On the basis of your calculations for part (a), were workers in the Finance and Business Services sector of the economy better or worse off at the end of 2019 than they were a year earlier? Explain your reasoning. [2] page 3 of 6

Question 4 (Unit 4) – 8 marks

This question concerns systematic random sampling from the 'Sampling Department staff list' given in Table 1 of Unit 4 (Section 2). Suppose you want to interview roughly one-fifteenth of the 86 staff. You are going to choose a systematic random sample from that list using the random number table provided in the appendix to Unit 4, using row 58.

- (a) Use the pairs of digits in row 58 (from left to right) of the random number table to identify which of the first 15 people listed in the table you should use as the starting point for your systematic sample. Write down the label and name of the person you will start with. [1]
- (b) Select the rest of the sample and write down the labels of the people selected in the order that they were selected. [1]
- (c) List the names of all the sampled individuals in the order in which they were chosen, along with their gender and occupation. [3]
- (d) Comment on the representativeness of the sample with respect to gender, assuming that the full staff list in Table 1 of Unit 4 is the population. [3]

Question 5 (Unit 4) – 8 marks

A staff survey asked for people's opinions about the value for money of the food served in the staff restaurant. Their responses are given in Table 3.

Table 3 Value for money of food served in the staff restaurant

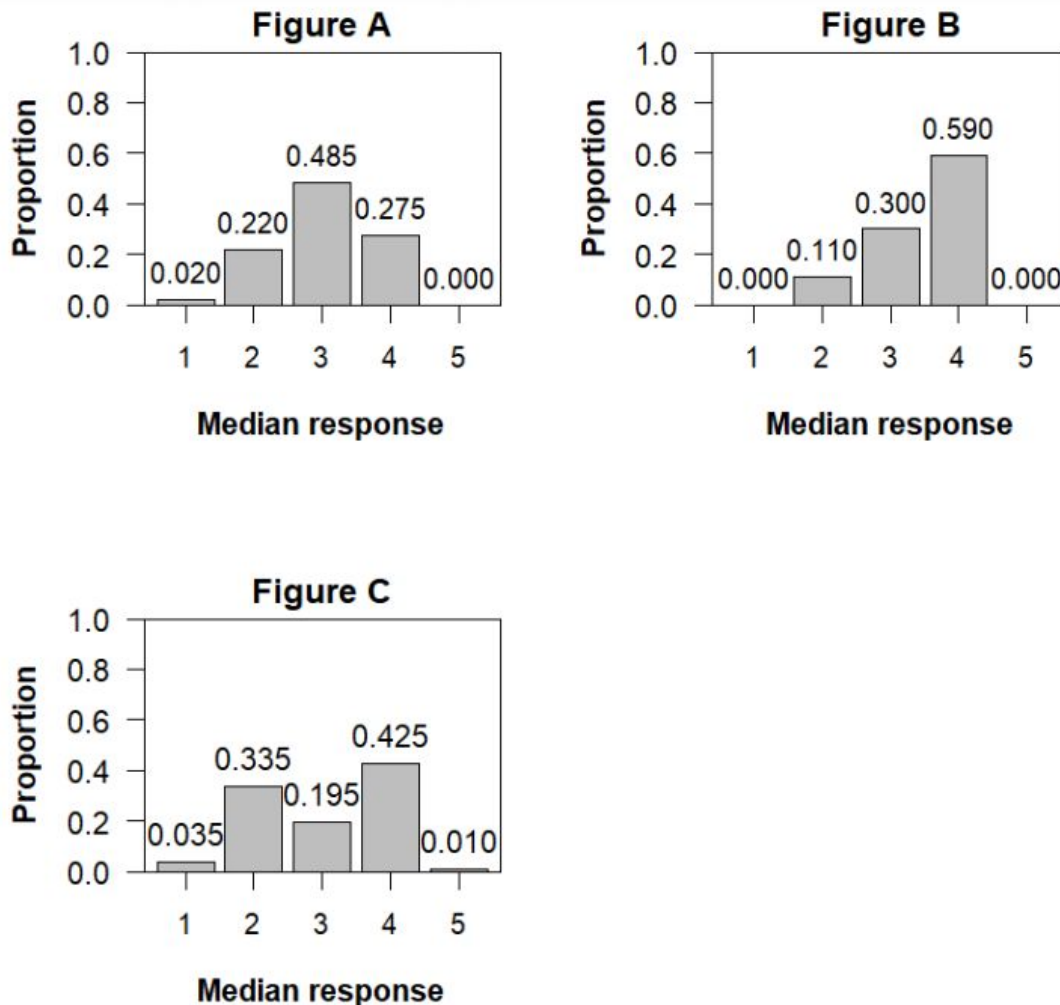
Response	Rating	Number
Very good value	1	36
Good value	2	71
Neutral	3	30
Poor value	4	121
Very poor value	5	23

(a) What is the median response for this population? Show your working. [2]

(b) Figures A, B and C below each show a distribution of the sample median, but they are based on sample medians for samples of different sizes from different populations.

- Sample 1 is the distribution of the sample median for samples of size 23 from the population given in Table 3.
- Sample 2 is the distribution of the sample median for samples of size 3 from the population given in Table 3.
- Sample 3 is the distribution of the sample median for samples of size 13 from a population different from that given in Table 3.

Giving your reasons, say which figure (A, B or C) relates to which sample (1, 2 or 3). [6]



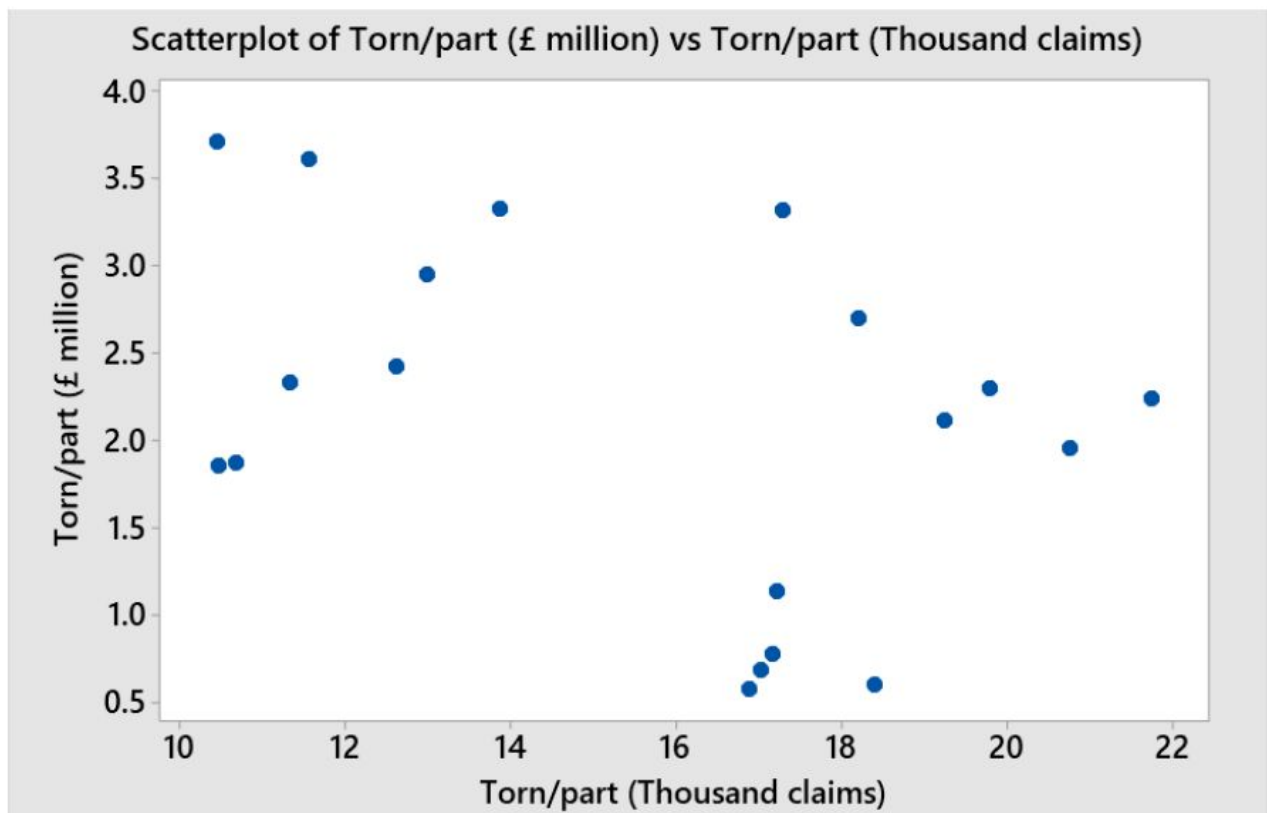
Question 6 (Unit 5) – 17 marks

The Minitab file that you require for this question should be downloaded from the 'Assessment' area on the module website.

The Bank of England publishes statistics about the numbers of claims they receive for damaged bank notes and how many British pounds they exchange for those damaged notes (www.bankofengland.co.uk/statistics/banknote). The claims are categorised by the type of damage. You will analyse the data for torn notes in this question. The data for the years 2000 to 2018 are in the Minitab worksheet `banknotes-torn.mwx`. The first column is the year, followed by a column for the value of new notes (in millions of pounds) that were paid out in exchange for claims for torn notes, and a column for the number of claims (in thousands) concerning torn notes.

Figure D below is a scatterplot of how many British pounds were exchanged each year for torn notes (in £ million) against the number of claims each year for torn notes (in thousands).

Figure D



(a) Interpret the scatterplot in Figure D giving reasons for your answers and using the following checklist for interpreting scatterplots.

- (i) Is the relationship positive, negative or neither? [2]
- (ii) Is the relationship linear or non-linear? [2]
- (iii) Is the relationship strong or weak? [2]
- (iv) Are there potential outliers? If so, which points are they? [2]

(b) Use Minitab to fit a least squares regression line to the data in banknotes-torn.mwx. You should treat Torn (£ million) as the response variable, y , and Torn (Thousand claims) as the explanatory variable, x . Include the appropriate Minitab output to show the regression equation in your answer. What is the equation of the least squares regression line in terms of x and y ?

[2]

- (c) Produce a residual plot using Minitab. Include a copy of your plot in your answer. Comment on what the residual plot tells you. [4]
- (d) You should use the equation of the least squares regression line from part (b) (and hence your calculator rather than Minitab) to answer this part.
- If 12.5 thousand claims for torn notes are submitted in 2019, estimate how many million British pounds the Bank of England will have to pay out for those claims. [2]
- (e) If the number of claims fell to 4.9 thousand in 2020, would it be appropriate to use the equation of the least squares regression line from part (b) to estimate how many British pounds the Bank of England would have to pay out for those claims? Why, or why not? [1]