

# M140

---

## TMA 01

## 2020J

---

Covers Units 1 and 2

Cut-off date 18 November 2020

---

Please read the Assessment Guide on the module website before beginning work on this TMA. You can submit your TMA either by post or electronically using the University's online TMA/EMA service.

This TMA is marked out of 50. Your overall score for this TMA will be the sum of your marks for each question. Note, however, that because the University requires TMAs to be marked out of 100, the mark returned to you by the University will actually be out of 100 (i.e. twice the total of marks on your TMA script).

The marks allocated to each part of each question are indicated in brackets in the margin.

Guidance about how to answer TMA questions is given in Subsection 7.2 of Unit 1.

Note that the Minitab files that you require for this assignment should be downloaded from the 'Assessment' area on the module website.

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

If you have a disability that makes it difficult for you to attempt any of these questions, then please contact your Student Support Team or your tutor for advice.

*You should be able to answer Questions 1 to 3 after you have studied Unit 1.  
You will need Minitab to answer Questions 1, 2 and 3.*

*You should be able to answer Questions 4 to 6 after you have studied Unit 2.*

**Question 1** (Unit 1) – 8 marks

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

The data used in this question are about the numbers of students on three of the five most popular modules at the OU. The variable **ModuleName** contains the name of the module, and the variables **2015/16**, **2016/17**, **2017/18**, **2018/19** contain the numbers of students at the start of each of the modules in each of the academic years 2015/16 to 2018/19. The variable **MedianNumber** contains the median of the number of students on each module. These data are given in the Minitab worksheet **favourites.mwx**. Run Minitab and open this worksheet.

The mean number of students on a module over the four years, using the variables mentioned above, is

$$\text{mean number} = (2015/16 + 2016/17 + 2017/18 + 2018/19)/4.$$

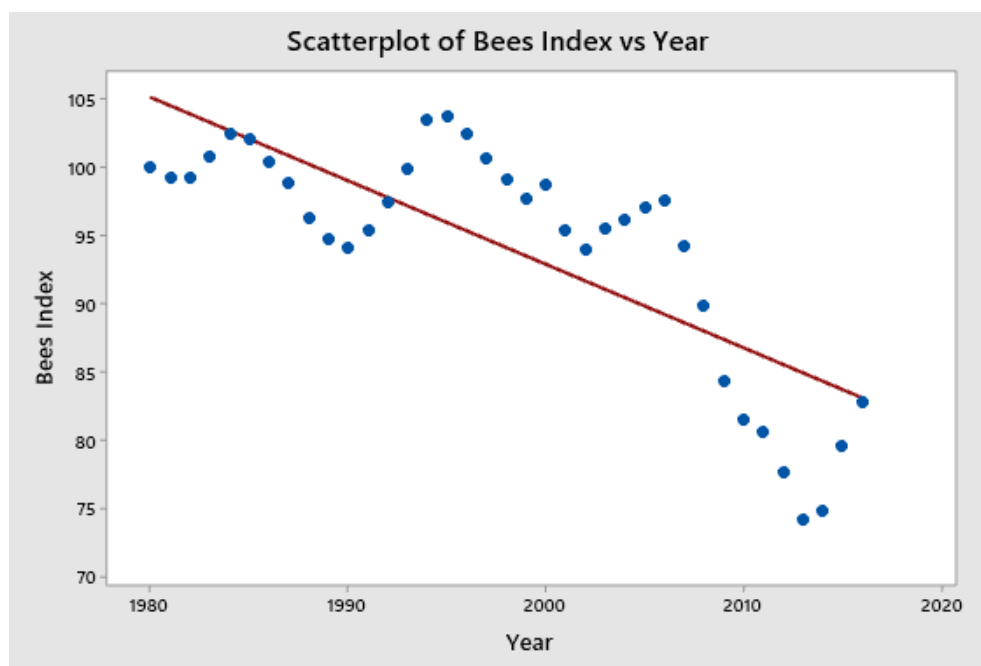
- (a) Calculate by hand the median number of students on DD102 and show all your working in your answer, clearly explaining your method. (You can check that you have the same answer as in the Minitab worksheet.) [2]
- (b) Use Minitab to produce a column of values, headed **MeanNumber**, containing the mean numbers of students on each module. Include this column of values in your answer. (You might use the **Row Statistics** option from the **Calc** menu in Minitab.) [2]
- (c) Use Minitab to calculate a column of values, headed **Diff**, containing the differences between the mean and the median numbers of students on each module. Use Minitab to round the differences to the nearest whole number, i.e. zero decimal places, and include this column of values in your answer. [2]
- (d) The range of the number of students on MST124 over the period is 1208. Use Minitab to calculate the corresponding ranges for DE100 and for DD102. What does the comparison of the ranges tell you about the numbers of students on DE100? What functions did you use in Minitab to calculate the ranges? [2]

**Question 2** (Unit 1) – 9 marks

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

The data used in this question relate to a measure of UK biodiversity. Insects are important pollinators and each year the number of 1 km grid squares in which insect pollinators are found is recorded. An ‘occupancy index’ is calculated from these data and scaled so that for the first year (1980) the index is 100. (Data source: [jncc.defra.gov.uk](http://jncc.defra.gov.uk) Indicator D1c.) These data are given in the Minitab worksheet **bees\_hoverflies.mwx**. In the worksheet, for each year, the variable **Bees Index** contains the occupancy indicator for wild bees and the **Hoverflies Index** contains the occupancy index for hoverflies. Run Minitab and open this worksheet.

Figure 1 shows the scatterplot of **Bees Index** against **Year** from 1980 to 2016, with a straight-line model fitted through the points.



*Figure 1*

- (a) By looking at the data file in the Minitab worksheet **bees\_hoverflies.mwx**, in which year was the **Bees Index** highest and what was the value of **Hoverflies Index** in that year? [2]
- (b) Describe two main features that are evident at first glance of the scatterplot for the **Bees Index** data against **Year** over this period. [2]
- (c) Comment in general terms on whether the scatter of points around the straight-line model has changed over the period, and whether the straight-line model provides an adequate representation of the data over this period. Give a reason for your answer to the latter part of the question. [2]
- (d) Use Minitab to produce a scatterplot of **Hoverflies Index** on the vertical axis against **Year** on the horizontal axis. Do not attempt to add a straight line to these data. Include a copy of your scatterplot in your answer. Describe briefly what the plot tells you, including whether or not you think that a straight-line model would be an appropriate thing to contemplate for these data. [3]

**Question 3** (Unit 1) – 8 marks

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

The National Survey of Young People’s Well-Being, 2010 collected information on how happy young people reported they felt. The data are available at [beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=7899](http://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=7899). The Minitab worksheet **well\_being.mwx** is an extract of the data for 90 schools, and contains the (modified) school identifier and the average score of pupils, on an overall happiness scale. The happiness scale runs from 0, which represents very unhappy, through to 10, which is very happy. Open this worksheet in Minitab.

- (a) Use Minitab to produce the default stemplot of the average happiness score of pupils (i.e. the stemplot produced when the **Increment** field is left blank and the **Trim outliers** option is not selected). Include a copy of your plot as your answer. [1]
- (b) Briefly describe the shape of the stemplot that you produced in part (a). Your answer should consider whether the distribution is unimodal, bimodal or multimodal, whether it is symmetric, right-skew or left-skew, and whether or not there are outliers. Justify your conclusions. [6]
- (c) Use the stemplot that you obtained in part (a) to find the median average happiness score of pupils. Remember to show your working. [1]

**Question 4** (Unit 2) – 11 marks

*Do not use Minitab to answer this question.*

The Higher Education Statistics Authority (HESA) publishes the median salaries of full-time first degree leavers who were employed in the UK ([www.hesa.ac.uk/news/28-06-2018/sfr250-higher-education-leaver-statistics](http://www.hesa.ac.uk/news/28-06-2018/sfr250-higher-education-leaver-statistics)). The stemplot in Figure 2 was obtained from the data for 2016/17.

Stem-and-Leaf Display: Salary

Stem-and-leaf of Salary N = 36  
Leaf Unit = 100

```

 9   16  000055555
16   17  0000055
17   18  0
(2)  19  00
17   20  000
14   21  00
12   22  000555
 6   23
 6   24  00
 4   25  0
 3   26  5
 2   27
 2   28
 2   29
 2   30  0
 1   31  0
```

$n = 36$  16|0 represents £16 000

*Figure 2*

- (a) Use Figure 2 to calculate the median, upper quartile and lower quartile of the data. Show all your working. [3]
- (b) What other values, in addition to those that you have calculated in part (a), would you include in a five-figure summary? Give their values as part of your answer and give one advantage a five-figure summary has over a stem-and-leaf plot. [3]
- (c) Calculate the range and the interquartile range for the data shown in Figure 2. Explain which you think is the more suitable measure of spread. [3]
- (d) Why do you think HESA publish median salaries and not mean salaries? (You might draw on evidence from the stemplot to support your answer.) [2]

**Question 5** (Unit 2) – 6 marks

The Open University has developed some statistical models that can generate information on how likely registered students are to complete all the assignments for a given module. This information is usually summarised as a number somewhere on a scale from 0 to 1 – which in this question will be called the ‘likelihood’ – with 0 meaning very unlikely to complete, and 1 meaning very likely to complete. Figure 3 is a boxplot showing this information for all the students registered in a geographical area for M140 in a previous year.

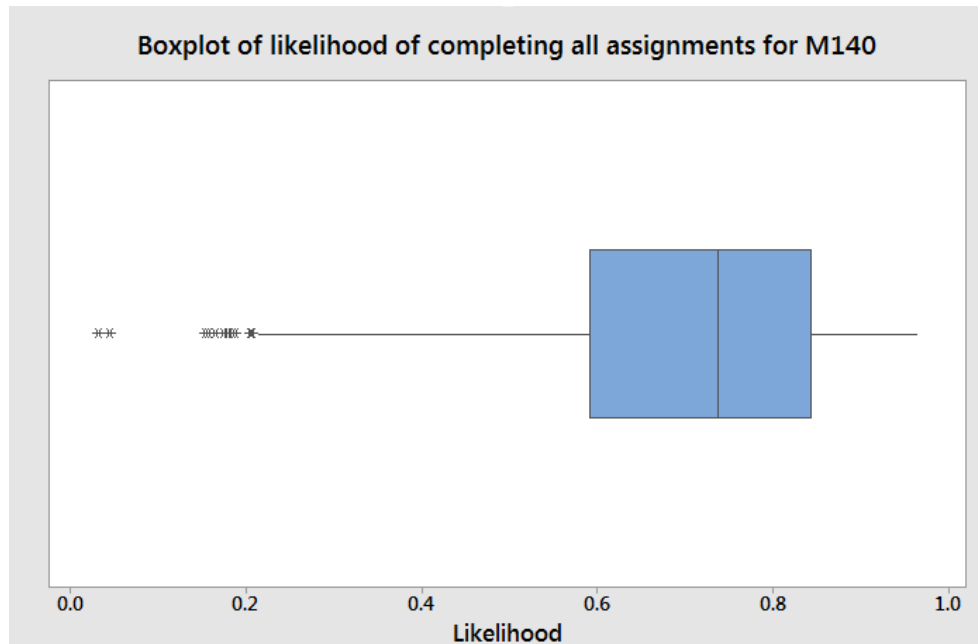


Figure 3

- (a) Is the distribution of this batch of data left or right skew? Justify your answer. [2]
- (b) The mean value of the likelihood in Figure 3 was correctly calculated as 0.62 and the median value correctly calculated as 0.68. It was realised that the data for two students had been included in error and, after excluding their data, the mean and median were both correctly recalculated. The new value of the mean was 0.61 and the median remained 0.68. How could the mean value have changed but not the median? Make two valid points to answer this part of the question. [2]
- (c) The statistical model that generates the likelihood of completing all assignments on M140 for each student uses data available at the start of the module that reflect each student’s motivation, opportunity, determination, resilience and interest in study. Where would you place yourself – in terms of the median, quartiles and whiskers – on the boxplot? Briefly justify your answer. [2]

**Question 6** (Unit 2) – 8 marks

The Office for National Statistics publishes the Consumer Prices Index (CPI) ([www.ons.gov.uk/economy/inflationandpriceindices](http://www.ons.gov.uk/economy/inflationandpriceindices)). The monthly values of the CPI, for the first four months of 2019, 2018 and 2017, are listed in Table 1.

**Table 1** CPI values

|          | CPI 2019 | CPI 2018 | CPI 2017 |
|----------|----------|----------|----------|
| January  | 106.4    | 104.4    | 101.4    |
| February | 106.8    | 104.9    | 102.1    |
| March    | 107.0    | 105.0    | 102.9    |
| April    | 107.6    | 105.4    | 102.9    |

- (a) Use the values of the CPI in Table 1 to calculate the annual inflation rates, based on the CPI, for each of the months January 2019, January 2018, April 2019 and April 2018, as a percentage rounded to two decimal places. (The base date is January 2015.) [2]
- (b) Comment briefly on
- The January 2019 inflation rate compared to the January 2018 rate.
  - The April 2019 inflation rate compared to the April 2018 rate.
  - How the inflation rate in January 2019 compared with that in April 2019, relative to the same comparison in 2018.
- (There is no need to try to justify any changes that you observe.) [3]
- (c) A CPI index-linked pension was £600 per month in February 2018. How much should the pension be per month in February 2019? [1]
- (d) Based on the CPI, calculate the purchasing power (in pence) of the pound in March 2019 compared to two years earlier. [2]
-