Assignment 1 – Statistical Methods (MATH 1068) SP2 2021

Instructions

- This assignment is worth 15% of your final grade and is due at 5pm Friday 9th of April.
- Submission is online and through the learnonline website.
- Assignments will be marked and returned online.
- Minitab output will be required in some sections, to avoid losing any information when uploaded. The file submitted should be a PDF document.
- The marks for each question are displayed next to the question.
- It is important that you follow any instructions or guidance in the questions, such as "Use Minitab" where required.
- The assignment consists of a total of **75 marks**.
- **Any late submission will attract a penalty of 10% off the maximum marks available per day. The cut-off time is <u>5pm </u>each day.**

Question 1 (28 Marks)

Forbes Data: The Forbes Global 2000 is a comprehensive list of the world's biggest and most powerful companies, as measured by a composite ranking for sales, profits, assets, and market value. All the values are in US dollars and the specific list is for the year, 2004.

You will need the MINITAB worksheet called *Forbes2000.MTW* for this question, which you can download from the **Data files** folder within the course website.

For full marks, ensure that appropriate axis labels and meaningful titles are included with all your graphical displays for this question.

(a) **(2 marks)** Use MINITAB to produce a histogram of the Market value.

(b) **(2 marks)** Use MINITAB to produce Descriptive Statistics for the Market value.

(c) **(2 marks)** Use MINITAB to produce a boxplot describing the different Market Values. Show the boxplot horizontally.

(d) **(4 marks)** Using your output from (a) to (c), comment on the shape of the distribution of the Forbes data. In particular, briefly describe:

- The shape of the distribution (skewed or symmetric);
- Modality: whether there is one peak, or multiple peaks, in the distribution.
- Whether there appear to be any outliers.

(e) **(4 marks)** Which measures of central tendency and dispersion would you use to numerically summarise and compare the market value for different companies? For full

marks, ensure you justify your choice of measures, and interpret the corresponding values.

(f) **(2 marks)** Use MINITAB to plot the Normality curve for the market value variable and briefly describe the plot and whether the data follows a Normal Distribution.

(g) **(6 marks)** Repeat the exploratory analysis on the log(marketvalue) in the data set. This includes calculating the descriptive statistics, producing a box plot, explaining the type of distribution and if there are any outliers present. Note that this column is a transformation of the data set by the function, Y=log(X) where X is the *marketvalue* (You are **not** required to transform the data).

(h) **(6 marks)** Assuming the data set, log(marketvalue) is approximately Normally distributed, use your MINITAB output to calculate the Market value ($ Billions) of the top 2.5% of companies (2 decimal places)?

Question 2 (24 marks)

Parking Tickets? The City of Adelaide enjoys dishing out parking tickets even if you get back to your car a minute late. The council is so impressed with their work that they have released the parking fine data to you outlining the Date, Location and Number of Minutes Overtime. We are interested in a certain street, Hailfax Street, where you park every day, and there are no parking meters, only 2 hour parking limits.

You will need the MINITAB worksheet called *parkingfines.MTW* for this question, which you can download from the Data files folder within the course website.

For full marks, ensure you use correct statistical notation when expressing probabilities. Also show calculations and MINITAB output that supports your answers and finish off each result with a conclusion / comment.

(a) **(5 marks)** For the entire month of January, what is the average number of minutes overtime until the fine gets issued? Find the probability that the council issues between 1 to 10 minutes overtime. For full marks, provide a correct probability statement and include your MINITAB output.

(b) **(4 marks)** Find the probability that for a randomly chosen day, a fine is issued for being less than 1 minute overtime on Hailfax Street. For full marks, provide a correct probability statement and show all working out. Do not use MINITAB to compute the probability.

(c) **(4 marks)** Suppose you plan to observe the average number of minutes overtime in parking fines for January every year. By using a random sample of 61 fines of the data and the mean of the sample as a point estimate, state the shape of the distribution for the means generated by this sample (give reasons). State the parameter values for the mean and standard deviation for this distribution.

(d) **(5 marks)** Find the probability that the mean number of minutes overtime is greater than 20 minutes every January. For full marks, provide a correct probability statement and include your MINITAB output.

(e) **(6 marks)** What would be the mean number of fines per day every January at the lowest 5% chance of getting a fine?

For full marks, provide a correct probability statement, full working out and include verification of your solution using MINITAB output.

Question 3 (24 marks)

Analysing data and confidence intervals: A medical experiment was conducted on the time between pulses along nerves. The data in the file *Nerve Pulses* gives the time taken for a pulse to travel through a nerve. There are 400 observations recorded and the units of measurement is 1/50 of a second.

You will need the MINITAB worksheet called *nervepulses.MTW* for this question, which you can download from the ***Data files*** folder within the course website.

For full marks, ensure you include the appropriate steps and MINITAB output that supports your answers.

(a) **(2 marks)** Use MINITAB to produce a histogram of the pulse times.

(b) **(4 marks)** Use MINITAB to produce Descriptive Statistics for the pulse times. State and compare the difference between the median and mean of the times.

(c) **(2 marks)** Comment on the shape of the distribution and if there are any outliers present.

(d) **(8 marks)** Calculate and interpret the 95% confidence interval for the population mean based on the sample data. All calculations should be done *manually* **without** using Minitab, however, use MINITAB to visualise the results.

(e) **(3 marks)** A well-known medical doctor claims the average pulse time is 10 1/50 seconds. Use part (d) to discuss whether the doctor's claim is valid and what conclusions can be made about the two confidence intervals. Explain briefly.

(f) **(4 marks)** As a further study conducted to analyse the true population mean of nerve pulses, load the dataset *meansamplepulses.MTW* into MINITAB. This dataset is comprised of the means of samples of size 10, 20 and 30 from the *nervepulses* dataset. Use MINITAB to produce Descriptive Statistics and a histogram of the sample means. Compare and discuss the means, distributions for the three samples and the mean of the data *nervepulses.MTW*