Due time: 11PM, Wednesday August 21

Writing and submission:

- Write your answers on A4 paper. Page 1 should only have your student number, the subject code and the subject name. Write on one side of each sheet only. Each question should start on a new page. The question number must be written at the top of each page.
- Put the pages in question order and all the same way up. Use a scanning app to scan all pages to PDF - do not take photos with your phone camera ! Scan directly from above. Crop pages to A4.
- Submit your scanned assignment as a single PDF file. Gradescope will give you an opportunity to match questions with pages in your submission - please take the time to do this as it will make marking more efficient. Once submitted, carefully review the submission in Gradescope. Scan again and resubmit if necessary.

Exercise 1 (6 marks): Find a graph representing a data set from a reputable source. This can be from an academic journal, a government report, a reputable news website, or a professional organisation. Ensure the graph is complex enough to allow for a detailed analysis. You can use a graph that shows a good example or a graph that demonstrates a bad example.

1. Provide context in no more than 100 words: Briefly describe the graph you have chosen. Include the source, the main topic or subject of the graph, and its intended message or purpose.
2. Using the five principles of good graphic design that you learned in Module 1, evaluate the graph and write a detailed commentary addressing each principle (Total number of words 250 for all the questions below):
   - Is the data shown clearly in the graph and why?
   - Does the graphic use good alignment on common scales for quantities to be compared and why?
   - Does the graph use simplicity in design while successfully conveying the message and why?
   - Does the graph have good visual encoding and why?
   - Does the graph use standard forms?

Exercise 2 (12 marks): Let $X_1, X_2, \ldots, X_m$ be a random sample from a Binomial distribution $Binomial(n, \sqrt{p})$.

1. Compute the maximum likelihood estimator for p. (Please note that $\sqrt{p}) \geq 0$). Please show all your workings.
2. For the particular case when $X_1 = 1, X_2 = 3, X_3 = 3$ are the realisations of a random sample from a $Binomial(n = 5, \sqrt{p})$, what is the maximum likelihood estimate for the parameter p? Please justify your answer.

Exercise 3 (12 marks): Given a random sample of size n from a distribution with density function given by

$$f(x) = \theta \left(\frac{1}{x}\right)^{\theta+1}, \; x \geq 1$$

(in what follows we assume that $\theta \geq 1$). Hint: $\int_1^\infty x^{-\theta}\, dx = \frac{1}{\theta-1}$.

1. Using the method of moments, find an estimator for θ. Please show all your workings.

2. We have now two samples drawn from a the distribution:

   - Sample 1: $X_1 = 1.333$, $X_2 = 1.577$, $X_3 = 1.523$, $X_4 = 1.645$, $X_5 = 1.447$, $X_6 = 1.684$, $X_7 = 1.529$, $X_8 = 1.556$, $X_9 = 1.414$.
   - Sample 2: $X_1 = 1.333$, $X_2 = 1.333$, $X_3 = 1.523$, $X_4 = 1.333$, $X_5 = 1.333$, $X_6 = 1.333$, $X_7 = 1.333$, $X_8 = 1.333$, $X_9 = 1.333$.

   In addition to the variance, what summary statistics could we use to compare the variability between these samples? Compute the summary statistics for those two samples and (by hand), draw two plots displaying the variability of the summary statistics. Please show all your workings. What can you say about the variability between these two samples? Hint : Consider the following definition of the sample median and how you can use it to find the summary statistics in this question:

$$\hat{m} = \begin{cases} x_{((n+1)/2)} & \text{when } n \text{ is odd} \\ \frac{1}{2}\left(x_{(n/2)} + x_{((n/2)+1)}\right) & \text{when } n \text{ is even.} \end{cases}$$

Exercise 4 (18 marks) : An Australian Olympic athlete has been training with a tennis ball launcher whose speed follows a Normal distribution $N(\mu, \sigma^2 = 25 \text{ km/h}^2)$. Two independent random samples are taken from the machine: $X_1, X_2, \ldots, X_6$ and $Y_1, Y_2, \ldots, Y_6$.

1. Are the estimators $T_1 = \bar{X}$, $T_2 = \bar{Y}$, and $T_3 = \sum_{i=1}^{6} Y_i$ unbiased estimators of μ? Please justify your answer and show all your workings.

2. Given the estimator $T_4 = aT_1 + (1-a)T_2$, compute its mean square error (MSE). Please show all your workings.

3. What is the sampling distribution of T_4? Please justify your answer and show all your workings.

Exercise 5 (6 marks): Use R to generate a dataset of 1000 random numbers from a skewed distribution. Create a histogram of the data and a QQ plot comparing the quantiles of your skewed dataset to the quantiles of a standard normal distribution. Submit the QQ plot and include the R code used to generate the plot. Explain the resulting curvature of the points in the QQ plot. What does this tell you about the relationship between the skewed distribution and the normal distribution?