STATS 707
Computational Introduction to Statistics

Assignment One

Due by: 11am on Friday 15th March 2024

Important Notes:

- This assignment is to be solved using the *R Studio* statistical package. Type your answers in a *word* document, paste the *R Studio* outputs/figures into this document and convert this document into a PDF before submitting.
- Please ensure that you clearly identify which question and task each of your answers relates to.
- Make sure that all plots and R output is suitably labelled so that it is easily and correctly identifiable.
- Your attention is drawn to the policies regarding plagiarism and late submission which are described in the digital course outline.
- Maximum marks for each question are indicated in square brackets.
- **Screenshots of the accompanying R code should be pasted** along with outputs.

Question 1: [25 Marks]

Task 1:

About 30% of human twins are identical and rest are fraternal. Identical twins are necessarily the same sex – half are males and half are females. One-quarter of fraternal twins are both male, one-quarter are both female and one-half are mixed: one male, one female. You have just become parent of twins and are told they are both girls. Given this information, what is the probability that they are identical?

[10 Marks]

Task 2:

A coin is said to be unbiased if P(heads) = P(tails) = 0.5. Probability that a newly minted coin by a Government Mint is unbiased is 0.95. When an unbiased coin is tossed 100 times, the probability of getting 62 heads or more is 0.01 (*0.01048 rounded off to two decimals*) for an unbiased coin. As part of a quality control process at the Mint some of the newly minted coins are randomly selected and each coin is tossed 100 times. If a coin tosses heads 62 times or more, it is set aside as a possibly biased (faulty) coin. All such coins are to be melted down and re-minted, which has additional costs. A consultant has been hired to identify efficiencies in this process and see if the criteria of 62 heads or

more could possibly be further optimized. Since you work as a data scientist, the consultant asks you to calculate the probability that the coin is unbiased given that it tossed 62 or more heads out of 100 tosses. Find this probability. Assume (just for this example) that a biased coin always has a P(heads) =0.55.
Clearly define your events, state the formulae you are using and show your working.

[15 Marks]

Question 2: [35 Marks]

This question is about the Central Limit Theorem (CLT). Task 2 is to be implemented using R programming.

Task 1:
Explain in your own words what the CLT states. Explain also, the significance of it in practice.

[10 Marks]

Task 2:
Choose a non-normal probability distribution. Illustrate how CLT holds true **as sample size increases** by simulating repeatedly from this distribution of your choosing. Make sure to report the following.

a. State the exact distribution (including the exact parameter values) that you have chosen and plot its probability distribution. [3 Marks]
b. State the expected value (i.e. population mean) of this probability distribution. [2 Marks]
c. Use histogram and the summary statistic to illustrate that CLT works. Make sure that you clearly label each output so that we know exactly what you are reporting. Accompany a brief but clear explanation as to why you think that your output illustrates that the CLT works. [20 Marks]