Question 1 – Rise of the machines

A semiconductor manufacturer produces devices used as central processing units in personal computers. The speed of the devices (in megahertz) is important because it determines the price that the manufacturer can charge for the devices. The files (6-42.csv) contains measurements on 120 devices. Construct the following plots for this data and comment on any important features that you notice.

(a) Histogram

(b) Boxplot

(c) Empirical cumulative distribution function

Question 2 – Samples

Consider the same data set as in Question 1, compute:

(a) the sample mean,

(b) the sample standard deviation,

(c) and the sample median.

Question 3 – The thickest rod

Eight measurements were made on the inside diameter of forged piston rings in an automobile engine. The data (in millimetres) is:

74.004, 73.999, 74.021, 74.001, 74.006, 74.002, 74.005.

Use R to construct a `qnorm` plot of the piston ring diameter data. Does it seem reasonable to assume that piston ring diameter is normally distributed? How about if you remove a single observation that is potentially an outlier?

Question 4 – A non-flat earth

In 1789, Henry Cavendish estimated the density of the Earth by using a torsion balance. His 29 measurements are in the file (*6-122.csv*), expressed as a multiple of the density of water.

(a) Calculate the sample mean, sample standard deviation, and median of the Cavendish density data

(b) Construct a `qnorm` plot of the data. Comment on the plot. Does there seem to be a "low" outlier in the data?

(c) Would the sample median be a better estimate of the density of the earth than the sample mean? Why?

Question 5 – Choice of Sample Size

The Charpy V-notch (CVN) technique measures impact energy. Assume that the impact energy is normally distributed with $\sigma = 1$ Joules. How many specimens must be tested to ensure that the error between the sample mean and the true mean is at most 0.5 with a confidence of 95%?

Question 6 – Using the CLT

Consider the following random variables

$$V \sim \text{Exp}(2) \quad \text{(exponential distribution)}$$
$$W \sim \text{G}(0.4) \quad \text{(geometric distribution)}$$

(a) What is the mean and variance of V and W?

Consider now,

$$S_n = \sum_{i=1}^{n} X_i,$$

where X_i is either V_i or W_i (distributed as V or W) and different X_i are assumed independent.

(b) What is the mean of S_i? Answer this separately for V and W.

(c) What is the variance of S_i? Answer this separately for V and W.

Question 7 – The CLT with simulation

Let X_i be independent and exponentially distributed with parameter 10, that is $X_i \sim \text{Exp}(10)$, for $i = 1, \dots, n$. Define

$$\tilde{Z}_n = \frac{S_n - E(S_n)}{\sqrt{var(S_n)}}.$$

(a) What distribution has $\tilde{Z}$ for large n? (*Hint: Use the Central Limit Theorem.*)

(b) Generate Monte Carlo estimates of $P\left(\left|\tilde{Z}_n\right| > 2.0\right)$ using no less than 10^6 generations of $\tilde{Z}_n$ for every n. Compare your results to $P(|Z| > 2.0)$ taken from a normal distribution table, where Z is a standard normal variable. Do this for $n = 2, 10, 20$, document your results and explain them.

Question 8 – Sample Mean

Suppose that a sample of size $n = 20$ is selected at random from a normal population with mean 100 and standard deviation 8. Let $\bar{X}$ be the sample mean.

(a) Calculate $P(98 \leq \bar{X} \leq 102)$.

(b) Find x such that $P\left(\left|\bar{X} - 100\right| > x\right) = 0.01$