Cycle 2 (R LAB)

1. The file 'bulbs.dat' contains life in years of a sample of 1000 bulbs randomly chosen from a total of 100000 bulbs produced by a company. Find an approximate probability distribution of the life of bulbs produced by the company. Draw a density plot of the distribution. Check your model with a quantile-quantile plot. What percentage of bulbs produced by the company will have life more than one year ?

2. The file pois.dat contains data of no. of accidents in a city for 365 days. Fit a Poisson distribution and calculate the theoretical frequencies. Also compare the theoretical and computed frequencies using a bar plot

3. The **Central limit theorem** states that the probability distribution of the means of random samples from a population will approach a normal distribution as the sample size increases. Write an R program to demonstrate Central Limit Theorem with the the following steps

   - Generate 1000 random numbers which are uniformly distributed between 0 and 1. This acts as the population.
   - Write a function to select n samples from this population and find its mean.
   - Choose 1000 random samples of size n from the population and plot the histogram of the sample means for $n = 2, 10, 20$ and 50. Compare with the distribution of the population.

4. If U_1 and U_2 are two independent random variables uniformly distributed between 0 and 1 find the density function of $U_1 + U_2$ using simulation. (You may generate 10000 realisations of U_1 and U_2, and plot the density function/histogram of $U_1 + U_2$ to identify the distribution of sum)

   What happens to the distribution if you plot sum of several such random variables ? Support your answer with plots of sums of 5,10 and 20 random variables.

5. Show by simulation that a binomial distribution $B(n, p)$ approaches a Poisson distribution with parameter $\lambda = np$ for small p and large n.

You may compare the density graphs of both distributions for $p = 0.1$ and $n = 5, 10, 20, 50$

6. Show by simulation that a binomial distribution $B(n, p)$ approaches a normal distribution $N(np, \sqrt{npq})$ for large n. You may compare the density graphs of both distributions for $p = 0.3$ and $n = 5, 10, 20, 50$

7. Plot the density graph of the random variable $Y = \sin(X)$ where X is uniformly distributed between 0 and 2π.

8. Let X be a random variable that is uniformly distributed in $(0, 1)$. Form a new random variable Y by rounding X to th nearest integer. Find the approximate probability distribution of the round-off error $Z = X - Y$ by generating 10000 samples of Z and plotting the histogram.

9. Using R generate a large number of samples from a normal distribution with mean $\mu = 20$ and standard deviation $\sigma = 4$. Let $x_1, x_2, \ldots x_n$ be the samples you generated. Compute each of the following "mean" values;

   (a) sample mean $SM = \frac{1}{n}\sum_{i=1}^{n} x_i$

   (b) geometric mean $GM = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$

   (c) harmonic mean $HM = \left(\frac{1}{n}\sum_{i=1}^{n} \frac{1}{x_i}\right)^{-1}$

   (d) root mean square $RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$

   Which of these estimates is closer to the true mean ?

10. Show that the mean of the samples drawn from a population converges to the population mean using the following steps. Generate n samples of Gaussian random variables with mean 3 and SD 1. Plot the sample mean against the number of samples for $n = 10$ to $n = 100$.