Tutorial 11

1. This question with steps below will help you to use **simulation** to illustrate how Central Limit Theorem (CLT) works.

   Consider the income of a population which follow an exponential distribution with mean λ. A random sample of size n is collected from this population to estimate λ. Let $\bar{x}$ denote the sample mean. Theoretically, CLT says: for this study, when sample size n is large enough then $\bar{x}$ will approximately follow a normal distribution with mean λ and standard deviation (sd) $\lambda/\sqrt{n}$.

   Steps below will help you to illustrate the CLT by simulation.

   Let's assume $\lambda = 5000$. Write R code for each question below:

   (a) Generate N samples, each sample has size n where $N = 100$, $n = 30$. Derive $\bar{x}$ for each sample and derive the sd of $\bar{x}$ from these samples.

   Hint: to generate a set of n values that follows an exponential distribution with mean λ, we use command: `rexp(n, rate = 1/λ)`.

   (b) Plot histogram of these $\bar{x}$. Does histogram have a bell curve ressembling a normal distribution? You can check the shape and also can use the rule of thumb (about 95% of points lie within 2 sd from the mean) to check.

   (c) Repeat 1b with same $N = 1000$ but $n = 100$. Does the histogram ressemble a normal distribution (compare the histogram with the previous one in 1b).

   (d) Repeat 1b with same $N = 1000$ but $n = 7$. Does the histogram ressemble a normal distribution? Give your comment about the effect of sample size n to the approximation of $\bar{x}$ distribution to a normal distribution.

   (e) Repeat the above question with $N = 50$ and $n = 100$. Compared to part (c), what do you observe about the distribution of $\bar{x}$ when $N = 50, n = 100$ and when $N = 1000, n = 100$? Hence conclude about the role of N in the approximation of $\bar{x}$ distribution to a normal distribution.

2. Repeat question above in Python.