STAT/Q SCI 403: Introduction to Resampling Method
Spring 2019
Homework 03

Instructions:

- You have to submit all your answers in a single PDF file generated by either LaTeX or *Rmarkdown*.
- You may use the LaTeXtemplate `HW_template.tex` to submit your answer.
- For questions using R, you have to attach your code in the PDF file. If the question ask you to plot something, you need to attach the plot in the PDF as well.
- If the question asks you to show a figure, the clarity of the figure will also be graded.
- The total score of this homework is 8 points.
- Questions with ♠ will be difficult questions.

Questions:

1. In this question, we will again analyze the `chickwts` dataset in R. This time we will focus on those chicken taking `linseed`. Namely, we will use those chicken whose `feed` is `linseed`. We have $n = 12$ chicken in this dataset. Some prior analysis suggested that chicken taking `linseed` will on average has a weight 240 gram. We will test this hypothesis. In other words, the null hypothesis is

$$H_0 : \text{chicken whose } \texttt{feed} \text{ is } \texttt{linseed} \text{ has an average weight 240.} \tag{1}$$

   And we will use t-test to test this hypothesis.

   (a) **(0.5 pt)** Under significance level $\alpha = 10\%$, can we reject the null hypothesis using t-test?

   (b) **(0.5 pt)** What are the mean and standard deviation of this 12 chicken whose `feed` is `linseed`?

   (c) **(1 pt)** Assume the actual distribution of those chicken taking `linseed` as their `feed` follows from a normal distribution with mean 220 and standard deviation 52, i.e., $N(220, 52^2)$. Use a size $N = 10,000$ Monte Carlo Simulation to find out the power of this t-test under significance level $\alpha = 10\%$? Hint: The power of a test is the probability of rejecting the H_0. Namely, you need to compute the probability of rejecting H_0 (1) using a t-test under significance level $\alpha = 10\%$ and the data is a size $n = 12$ random sample from $N(220, 52^2)$.

(d) **(1 pt)** (continue from the previous question) What will the power be if the sample size is $n = 12, 24, 36, 48, 60, 72, 84, 96$? Use a size $N = 10,000$ Monte Carlo Simulation for each case to find out. Moreover, use a figure to display the power versus sample size.

(e) **(1 pt)** (continue from the previous question) Assume that when the sample size $n = 20$, estimating the power using a size $N = 10,000$ Monte Carlo Simulation has a Monte Carlo Error 0.005 (the variance caused by N). What will the Monte Carlo Error be when we increase N to $1,000,000$?

2. In this question, we will study a very interesting result. Let $U_1, \cdots, U_{50} \sim \mathsf{Uni}[0,1]$ be 50 IID random variables from a uniform distribution over $[0,1]$. Let $M_{50} = \min\{U_1, \cdots, U_{50}\}$ be the minimum value of these 50 random numbers. We will study the distribution of M_{50}.

(a) **(1 pt)** Use a size $N = 10,000$ Monte Carlo Simulation to compute the mean of M_{50} and the standard deviation of M_{50}.

(b) **(1 pt)** It appears that the mean and standard deviation are roughly the same. This is a feature of the exponential distribution. Now use histogram to show that the distribution is roughly $\mathsf{Exp}(50)$, the exponential distribution with rate $\lambda = 50$. You need to attach the theoretical density curve of $\mathsf{Exp}(50)$ to this histogram. Note that the density of $\mathsf{Exp}(50)$ is

$$p_{\mathsf{Exp}(50)}(x) = 50 \cdot e^{-50x}$$

for $x \geq 0$ and 0 otherwise.

(c) **(1 pt)** Now change the sample size 50 to 100 and repeat compute the histogram of $M_{100} = \min\{U_1, \cdots, U_{100}\}$. Show the histogram and compare to the density curve of $\mathsf{Exp}(100)$. Does the distribution looks like $\mathsf{Exp}(100)$?

(d) **(1 pt; ♠)** Show that for n IID random variables $U_1, \cdots, U_n \sim \mathsf{Uni}[0,1]$, the random variable $M_n = \min\{U_1, \cdots, U_n\}$ converges in distribution in the following sense:

$$n \cdot M_n \xrightarrow{D} \mathsf{Exp}(1).$$