General Notes

- You will submit a minimum of two files, the core files must conform to the following naming conventions (including capitalization and underscores). 123456789 is a placeholder, please replace these nine digits with your nine-digit Bruin ID. The files you must submit are:
  1. *123456789_stats102c_hw3.Rmd*: Your markdown file which generates the output file of your submission.
  2. *123456789_stats102c_hw3.html/pdf*: Your output file, either a PDF or an HTML file depending on the output you choose to generate.
  3. *Included image files:* If you answer your questions with images files, you must upload them to this portal as well, or your Rmd file will not knit.
  4. Please place all of your Rmd (and image) file(s) into a single folder named 123456789_stats102c_hw3 and compress the folder into 123456789_stats102c_hw3.zip.
  5. You will submit two files; one html/pdf file (123456789_stats102c_hw3.html/pdf) and one compressed file (123456789_stats102c_hw3.zip).

  If you fail to submit any of the required core files you will receive ZERO points for the assignment. If you submit any files which do not conform to the specified naming convention, you will receive (at most) half credit for the assignment.

- **Your .Rmd file must knit**. If your .Rmd file does not knit you will receive (at most) half credit for the assignment.
  The two most common reason files fail to knit are because of workspace/directory structure issues and missing include files. To remedy the first, ensure all of the file paths in your document are relative paths pointing at the current working directory. To remedy the second, simply make sure you upload any and all files you source or include in your .Rmd file.

- Your coding should adhere to the tidyverse style guide: https://style.tidyverse.org/.

NOTE: *Everything* you need to do this assignment is here, in your class notes, or was covered in discussion or lecture.

- Please **DO NOT** look for solutions online.
- Please **DO NOT** collaborate with anyone inside (or outside) of this class.
- Please work **INDEPENDENTLY** on this assignment.
- **EVERYTHING** you submit **MUST** be 100% your, original, work. Any student suspected of plagiarizing, in whole or in part, any portion of this assignment, will be **immediately** referred to the Dean of Student's office without warning.

Problem 1: Given the 4-dimensional multivariate normal distribution with mean vector

$$\mu^T = (2, 1.5, 3, 1)$$

and covariance matrix

$$\Sigma = \begin{pmatrix} 2.8 & 0 & 0.2 & 2 \\ 0 & 1.7 & 2 & 0 \\ 0.2 & 2 & 3.6 & -1.2 \\ 2 & 0 & -1.2 & 3 \end{pmatrix}$$

(a) Generate 1000 random observations from this multivariate normal distribution using the Choleski factorization method[1].

(b) Draw an array of scatter plots for each pair of variables and examine if they agree with the parameters. (You may use `pairs` in R)

Problem 2: Write R code to standardize an d-dimensional multivariate normal sample **X** with the known Σ and sample size n, where d and n can be arbitrary integer numbers.

(a) Derive an algorithm for standardizing a multivariate normal sample.

(b) Implement your algorithm in R.

(c) Use the generated data from Problem 1 to verify your algorithm.

Problem 3: Given X is a continuous random variable from the density $f(x)$. Let $\theta = \int g(x)f(x)dx = E[g(X)]$. Suppose we draw iid samples $X_1, ..., X_m$ from $f(x)$. Let $\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m} g(X_i)$.

(a) Prove $E[\hat{\theta}] = \theta$

(b) Prove $Var[\hat{\theta}] = Var[g(X)]/m$, and specify how to estimate $Var[g(X)]$.

(c) Specify how to construct 99% confidence interval of θ using central limit theorem.

(d) Suppose $f(x)$ is the exponential density with the rate, 1/3. Write a function (**mc2()**) to calculate a Monte Carlo estimate of $E[\sqrt{X}]$.

(e) Construct the 95% confidence interval of $E[\sqrt{X}]$. Repeat your function 1000 times, how often the confidence interval capture the true value of $E[\sqrt{X}]$.

Problem 4: Find the air-conditioning data set **aircondit** from the **boot** package. The data includes the 12 time intervals in hours between successive failures of the air-conditioning equipment. Assume that the time intervals between failures follow an exponential distribution with the hazard rate λ. Please use bootstrap to estimate the bias and standard error of $\hat{\lambda}_{MLE}$.

[1] You may use **chol()** function

Problem 5: Suppose X is a random variable from Beta($\alpha = 3, \beta = 2$).

(a) Write R code to compute the Monte Carlo estimator of the CDF.

(b) Write R code using the "hit-or-miss" approach to estimate the CDF.

(c) Compare your estimates with the outputs of the `pbeta` function in R for $x = 0.1, 0.2, ..., 0.9$.