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_hw1.Rmd*: Your markdown file which generates the output file of your submission.
  2. *123456789_stats102c_hw1.html/pdf*: Your output file, either a PDF or an HTML file depending on the output you choose to generate.

  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/.

- Any functions/classes you write should have the corresponding comments as the following format.

```
my_function <- function(x, y, ...){
#A short description of the function
#Args:
#x: Variable type and dimension
#y: Variable type and dimension
#Return:
#Variable type and dimension
Your codes begin here
}
```

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 $\mu^T = (4, 3, 2, 1)$ and covariance matrix

$$\Sigma = \begin{pmatrix} 3 & 0 & 2 & 2 \\ 0 & 1 & 1 & 0 \\ 2 & 1 & 9 & -2 \\ 2 & 0 & -2 & 4 \end{pmatrix} \tag{1}$$

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

(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 a function to standardize an d-dimensional multivariate normal sample **X** with the sample size n, where d and n can be arbitrary integer numbers.

(a) Derive the formula for standardizing a multivariate normal sample.

(b) Implement the function in R.

(c) Use the example from the last problem to check if the mean of the transformed sample is equal to zero and its covariance matrix is equal to the identity matrix.

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 copies $X_1, ..., X_m$ from $f(x)$. Let $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{m} g(X_i)$.

(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 95% confidence interval of θ using central limit theorem.

(d) Suppose $f(x)$ is the exponential density with the rate, 1/2. Write a function to calculate a Monte Carlo estimate of $E[X^2]$.

(e) Construct the 95% confidence interval of $E[X^2]$. Repeat your function 1000 times, how often the confidence interval capture the true value of $E[X^2]$.

Problem 4: Suppose X is a random variable from Beta(3, 4).

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

(b) Use the "hit-or-miss" approach to estimate the CDF.

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