# Final Exam - MATH 141

## 12/13/2021

**Exam starts on 12/13/2021 9:00 AM and ends on 12/15/2021 at 9:00 AM (48 hours).**

**You must submit your midterm exam into Gradescope within the 48-hour window.**

**Instructions:**

- Please provide complete answers/solutions for each question/problem.

- If it involves mathematical computations, please provide your reasoning and/or detailed solutions.

- Please abide by the Reed College Honor Principle.

- This exam is a take-home exam, you can use all of the notes you have, textbook, lecture slides, homework solutions, lab solutions, the course website, and the internet. Group work and collaboration is encouraged. Use all of the resources you have and what you need to answer the exam questions. Each student must take responsibility and ownership of their work and submit their work individually.

- If you have any technical questions, please send them to the instructor as a direct message on Slack. For the sake of fairness, the instructor will not respond to conceptual or clarifying questions. Because we may be unable to answer your question during the time frame in which you are taking the exam, document the issue on the exam and then proceed to the next problem.

- If you can't figure out why a code chunk is preventing you from knitting the document, replace "r" at the top of the code chunk with "r eval = FALSE, echo = TRUE". The code will not be executed, but it will be printed in your pdf, earning you some partial credit.

- Please follow the general exam guidelines written in the course website.

- **Please save your work as pdf file(s), don't put your name in any part of the document, and submit it to the Gradescope page for this course. Your document upload will correspond to your name automatically in Gradescope.**

**R Libraries**

```r
# loading packages
library(tidyverse)
library(dplyr)
library(ggplot2)
library(gghighlight)
library(infer)
```

# I. Marble-ous urns! You got to urn it!

You happen to randomly meet a stranger who sells urns and it so happens that you need an urn for some reason but you don't have any money. The stranger offers you one of two urns they have but you have to choose which urn has more blue marbles. Suppose that the stranger tells you that one of the urns have equal amount of blue, green, red, and yellow marbles. The other urn has the same colored marbles with unknown distribution. You don't know how many total marbles are there within each urn but you know that it's finite. The stranger tells you that you can only take samples with replacement from each urn. This means that you draw one marble, record the marble color, put it back to the urn you draw the marble from, and repeat.

*Which urn has more blue marbles? Urn 1 or Urn 2?*

1. Load the `urns.csv` data set. This data set has two columns. The `urn` column is the label of the urn and the `observations` column is the outcome when you sampled marbles from each earn.

   Suppose that you plan on sampling one marble again without looking at the urns.

   a. Create a two-way table of the `urns.csv` data set. Use the `table` function in R to create a quick table. You don't have to make a nice looking table.
   b. What is the probability that the new marble came from urn 1 given that you have a blue marble?
   c. What is the probability that the new marble came from urn 2 given that you have a blue marble?
   d. Which one is more likely and what do these probabilities mean?

2. Given the `urns.csv` data set, perform a statistical method appropriate for this type of variables. Use either simulation or theoretical methods.

   a. What is the appropriate statistical method and what are the hypotheses statements?
   b. What is the test-statistic and the p-value? Provide justifications when using either simulation or theoretical method. Interpret the results in context.
   c. Is the data a convincing evidence to support that one of the urns has an imbalanced marble distribution? Why or why not? Explain.

3. Your claim is that there are more blue marbles in urn 2 than in urn 1 because you discovered that it is more likely to get a blue marble in urn 2 than in urn 1. You took data again in a different way. Load the `marbles.csv` data set. This data set has five columns. The `urn` column is the label of the urn. The `B`, `G`, `R`, and `Y` columns contains the number of independent draws you made until you observe the corresponding marble color. For example in column `B`, a 4 means you drew 3 non-blue marbles until you got 1 blue marble. These four columns are discrete numerical variables. For a given marble, a smaller draw average means a larger marble frequency in that urn.

   a. Produce a visualization to compare the blue marble distributions between urn 1 and urn 2. Describe your observations of your visualization.
   b. What are the two statistical method that are appropriate for this types of variables? Explain each method in context.
   c. What are hypotheses statements and their corresponding mathematical symbols for the first method?
   d. What are hypotheses statements and their corresponding mathematical symbols for the second method?

4. Given the `marbles.csv` data set, perform one of the statistical methods described from the questions in problem 3. Use either simulation or theoretical methods.

   a. What is the p-value and the confidence interval? Provide justifications when using either simulation or theoretical method. Interpret the results in context.
   b. Is the data a convincing evidence to support that urn 2 has more blue marbles than urn 1? Why or why not? Explain.
   c. What type of statistical error is possible based on your results? Explain and discuss how can the level of significance (or confidence level) affect your decisions whether you picked urn 1 or urn 2.

## II. Group Project Self-Assessment.

Please provide elaborate answers to the following questions regarding your contribution and experience in your group. These questions will count as your individual credit for your final project report.

1. Describe briefly what your project was about.

2. In a short paragraph, describe the particular contributions that you made in the final report.

3. Do you feel that the workload for the entire project was shared equally among all group members? If not, please elaborate.

4. What have you learned in this experience? What can you improve on yourself for the next time you are in a group project?

You made it through this exam and this class! Congratulations! You should be proud of yourself.

Learn from failures and make every incremental successes count!

Cheers! :-)