

# Stats 598z: Midterm exam 1

---

**Important:**

Write your name and PUID on all sheets, and include the number of sheets

There are 8 questions, each for 5 points (but not all equally easy)

Attempt all questions, and when appropriate include a brief justification of your answer

Don't spend time polishing your answers as the main idea is more important

---

- `mtrx1 <- matrix( c(1,2,3), nrow = 3, ncol = 4) .` What is `mtrx1`? Write it out.
  - `mtrx2 <- matrix( c(1,2,3) + c(4,5), nrow = 2, ncol = 4) .` What is `mtrx2`? Write it out.
  - What is the outcome of running `matrix(1:3, nrow=1) + matrix(1:3,ncol=1)`?
- `my_vec` is a length-1000 vector of integers.
  - You want to check if 1 and 2 ever occur in successive positions. Write a few lines of R code to do this. You will get partial credit if you use loops, so try to use vectorization
  - Write a few lines of R code to return the indices of all 1's that are followed by 2's
- Let `my_mat` and `my_df` respectively be a matrix and a dataframe with  $n$  rows and  $m$  columns.
  - What does `nrow(my_mat)` return?
  - What does `length(my_mat)` return?
  - What does `nrow(my_df)` return?
  - What does `length(my_df)` return?
- Recall that `rnorm(n)` returns  $n$  Gaussian variables. You first sample 100 such Gaussian variables.
  - Having done this, you want to leave all samples less than 5 untouched. On the other hand, you want to replace each sample greater than 5 with a new sample from `rnorm`. You don't care what the value of the replacement is. Give a few lines to R to do this.
  - Suppose instead you want to keep repeating part (a) till all samples are less than 5. Provide a few lines of R to do this
- Suppose you run the lines

```
x <- FALSE
my_true <- function() {
  x <- TRUE
  return(x)
}
```

  - What is the result of running `x || my_true()`? What is the value of `x` after running this?
  - What is the result of running `x || {x <- my_true()}?` What is the value of `x` after running this?
  - What is the result and output from running `x || my_true() || {print("Hello")}`?
  - Explain when you would use `condition1 & condition2` and when `condition1 && condition2` briefly.

6. (a) Explain in a sentence what an aesthetic mapping (or `aes`) is in the context of `ggplot`.
- (b) Suppose `daily_temp` is a dataframe with columns `time` and `temperature`. Your friend runs the command
- ```
ggplot() + geom_point(daily_temp, aes(x=time, y = temperature, color = "blue"))
```
- but complains that the color is wrong. Explain what might be wrong and how to fix it.
- (c) Suppose you wanted to plot `temperature` vs `time` as above, but wanted all observations larger than 100 degrees to be one color, and all below to be another. Give a few lines of R to do this.

7. (Two unrelated subquestions)

- (a) You first run the following lines in R:

```
'+' <- function(x,y) {
  print("Hello")
  return(x+y)
}
```

What is the result of `1 + 2`?

- i. 3
- ii. "Hello"
- 3
- iii. 3
- "Hello"
- iv. "Hello"
- "Hello"
- "Hello"
- ... (repeat)

Explain why

- (b) Explain briefly why one would perform cross-validation, and the steps involved in  $k$ -fold cross-validation.
8. (a) You are given a vector `vec`. You want to create a new vector `pair_mean` whose first element is the mean of first element of `vec` and the last element of `vec`, whose second element is the mean of the second element of `vec` and the second last element, etc. Depending on whether `vec` has odd or even length, the last element of `pair_mean` is either the middle element of `vec` or the mean of the middle two elements of `vec`. Write a few lines of R to do this.
- (b) Now suppose the length of `vec` is a multiple of 10. Create a new vector whose first element is the mean of the first 10 elements of `vec`, whose second element is the mean of the next 10 elements etc.