Name:____________________________________ Date:________________

LAB 2I: R's Normal Distribution Alphabet
Response Sheet

Directions: Record your responses to the lab questions in the spaces provided.

Where we're headed

Get set up

(1) Start by loading the `titanic` data. Write and run code calculating the `mean` age of people in the data but `shuffle` their `survival` status 500 times.

(2) After creating `shfls`, write and run code using `mutate` to add a new variable to the dataset.

(3) Finally, write and run code calculating the `mean` and `sd` of the `diff` variable.

Is it normal?

(4) Is the distribution close to normal? Explain how you determined this. Describe the center and spread of the distribution.

(5) Compute and write down the mean difference in the age of the *actual* survivors and the *actual* non-survivors.

Name:____ Date:__________________

LAB 2I: R's Normal Distribution Alphabet
Response Sheet

Using the normal model

(6) Draw a sketch of a normal curve. Label the mean age difference, based on your shuffles, and the actual age difference of survivors minus non-survivors from the actual data. Then, shade in the area, under the normal curve, that is *smaller* than the actual difference.

(7) Fill in the blanks to calculate the probability of an even smaller difference occurring than our actual difference using a normal model.

```
pnorm(___________, mean = diff_mean, sd = ___________)
```

Extreme probabilities

(8) If you wanted to instead calculate the probability that the difference would be larger than the one observed, we could run (fill in the blanks):

```
1 - pnorm(___________, mean = diff_mean, sd = ___________)
```

Simulating normal draws

(9) Fill in the blanks in the following two lines of code to simulate 100 heights of randomly chosen men. Assume the mean height is 67 inches and the `standard deviation` is 3 inches.

```
draws <- rnorm(___________, mean = ___________, sd = ___________)
```

(10) Fill in the blank below to plot your simulated heights with a `histogram`.

```
histogram(draws, fit = ___________)
```

P's and Q's

(11) How tall can a man be and still be in the shortest 25% of heights if the mean height is 67 inches with a standard deviation of 3 inches?

Name:____ Date:__________________

LAB 2I: R's Normal Distribution Alphabet
Response Sheet

On your own

Using the `titanic` data:

(12) Were women on the Titanic typically younger than men? Write and run code using a histogram, 500 random shuffles and a normal model to answer the question.

Using the `cdc` data:

(13) Using 500 random shuffles and a normal model, how much taller would the typical male have to be than the typical female in order for the difference to be in the upper 1% by chance alone?

(14) How can we use this value to justify the claim that the average `Male` in our data is taller than the average `Female`?