LAB 3C: Random Sampling
Response Sheet

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

Learning by sampling

(1) Write down two reasons why getting *everyone* in Los Angeles to fill out the survey would be difficult. Also, write a sentence why the DWP might consider using a sample of households instead.

Loading a population

(2) Load the cdc data into R and fill in the blanks to take a *convenience* sample of the first 50 people in the data:

```
s1 <- slice(_____, 1:_____)
```

(3) Why do you think we call this method a *convenience* sample?

Comparing your convenience sample

(4) Using your convenience sample, write and run code creating a bargraph for the number of people in each grade.

(5) Do you think the distribution of grade for your sample would look similar when compared to the whole cdc data?

(6) Which groups of people do you think are over- or under-represented in your convenience sample? Why?

Name:_____   Date:_____

LAB 3C: Random Sampling
Response Sheet

(7) Write and run code creating a bargraph for grade using the `cdc` data.

(8) Compare the distributions of the `cdc` data and your convenience sample and write down how they differ.

Using randomness

(9) Fill in the blanks below to create a sample by randomly selecting 50 people in the `cdc` data, without replacement. Call this new sample `s2`:

_____ <- sample(_____, size = _____, replace = _____)

(10) Write a sentence that explains why you think the distribution of grade for this *random sample* will look more or less similar to the distribution from the whole `cdc` data.

(11) Write and run code creating a bargraph for grade based on this *random sample* to check your prediction.

Increasing sample size

(12) Write and run code creating bargraphs for grade based on each of the following sample sizes: 10, 100, 1,000, 10,000.

LAB 3C: Random Sampling
Response Sheet

(13) How do the distributions change as the size of the sample increases? Why do you think this occurs?

(14) Write and run code to `tally()` the proportion of `grades` for your *convenience* sample and all your *random* samples.

(15) Which set of proportions looks most similar to the proportions of the population?

Lessons learned

(16) Write down a reason why estimates based on *convenience* samples might not improve even as sample size increases.