WU #18

Math 58B, Spring 2023

Tuesday, April 4, 2023

Your Name: _____

Names of people you worked with: _____

1. What is your favorite movie?
2. Name one thing that is nice about **tidy** data.
3. Consider the research done in *The Lancet* "Transplantation of cardiac-committed mouse embryonic stem cells to infarcted sheep myocardium: a preclinical study". Does treatment using embryonic stem cells (ESCs) help improve heart function following a heart attack? Each of these sheep was randomly assigned to the ESC or control group, and the change in their hearts' pumping capacity was measured in the study.

```
library(openintro)
data(stem_cell)
stem_cell <- stem_cell %>%
  mutate(change = after - before)

stem_cell %>%
  group_by(trmt) %>%
  summarize(n_obs = n(), mean_chng = mean(change), sd_chng = sd(change))
```

```
## # A tibble: 2 x 4
##   trmt  n_obs mean_chng sd_chng
##   <fct> <int>     <dbl>   <dbl>
## 1 ctrl      9     -4.33    2.76
## 2 esc       9      3.5     5.17
```

Let's say you don't know any formulas. At all. How might you estimate how the difference in means varies (i.e., the SE) when:

- the null hypothesis is true
- the null hypothesis may or may not be true.

Solution:

Randomization Test:

If the null hypothesis is true, the treatments can be shuffled across the values of change (because the point is that the treatment does nothing!). With repeated shuffles (and then calculation of the differences in means), the SE can be estimated directly from the variability of the shuffled differences in means.

Bootstrapping:

If there is no assumption about the hypotheses, the data can be bootstrapped. That is, sample with replacement from each of the two groups (without shuffling). After each group is bootstrapped, the SE can be estimated directly from the variability of the bootstrapped differences in means.