HW 11 – Math 58B

your name here

2026-04-16

```r
library(tidyverse)
library(mosaic)
library(praise)
```

Assignment Summary (Goals)

- ANOVA test of means

Q1. Robustness of ANOVA

Use the following applet to assess the robustness of the ANOVA procedure:

http://shiny.calpoly.sh/ANOVA_robust/

a. Simulate samples of size 20 from populations with equal means and standard deviations 6, 6, and 6. What is your Type I error rate?

b. Simulate samples of size 20 from populations with equal means and (varying) standard deviations 4, 6, and 8. Now what is your Type I error rate?

c. Simulate samples of size 20 from populations with equal means and (varying) standard deviations 1, 6, and 11. Now what is your Type I error rate?

d. Do the error rates you found in a, b, and c vary by sample size (when sample sizes are equal)? (That is, if the samples are still equal but much larger or much smaller, do you see the same variability in error rates?)

e. Next repeat your simulation study (same three versions) with sample sizes $n_1 = 10, n_2 = 20$, and $n_3 = 30$. How do results differ?

f. Finally, repeat the above (both $n_1 = n_2 = n_3 = 20$ and $n_1 = 10, n_2 = 20, n_3 = 30$) simulation studies, but specify population means to be -3, 0, and 3, so that you study the power of the test under different conditions. (Remember that if the means are not all equal, then H_A will be true. So if you reject H_0 you will be measuring power and not the type I error rate.) How does the power change with equal/unequal standard deviations and/or equal/unequal sample sizes? Explain.

Q2. Crash Tests

The National Transportation Safety Administration conducts crash tests on automobiles. The file crash.txt contains data on automobile crash tests in which stock automobiles are crashed into a wall at 35 miles per hour with dummies in the driver and front passenger seat (as reported by the Data and Story Library (DASL) web site, http://lib.stat.cmu.edu/DASL/Datafiles/Crash.html). Response variables are measurements of injury extent on head (column 5), chest (column 6), left leg (column 7), and right leg (column 8). Explanatory variables include whether the dummy was on the driver or passenger side (column 9), protective devices in the car (column 10), number of doors on the car (column 11: 2, 4, or other), year of make (column 12), and size of car (column 14).

```
crashdata <- read_csv("https://pomona.box.com/shared/static/h3kzeh66rkpalq73svepejzwwfjbi4dd
```

(a) Produce boxplots of head injury measurements by number of doors. Write a few sentences comparing and contrasting the distributions of these measurements among the three groups. (Comment on shape, center, spread, unusual observations, and any other features of interest. Pay particular attention to the question of whether the extent of head injury seems to differ among the three groups.)

(b) In addition to the boxplots, produce histograms of the head injury measurements for each of the "number of doors" categories. Do the data suggest that each of the population distributions of head injury measurements are normally distributed? Explain.

After making the histogram, you might add the following layer to break the histogram apart for each of the different types of doors:

```
facet_grid( ~Doors)
```

(c) Apply a log transformation to the head injury measurements (use the data wrangling function `mutate()`). Then examine boxplots or histogram of this transformed variable by number of doors. Do these distributions appear to be roughly normally distributed?

(d) Does the technical condition which states each group should have approximately equal population standard deviations appear to be satisfied on the transformed data? Explain.

(e) Conduct an ANOVA on these transformed data. Report the **hypotheses** along with the value of the **F statistic** and **p-value**. Summarize your **conclusions** about whether the data provide evidence that the extent of head injury varies among vehicles with different numbers of doors.

Your R code will likely look something like this:

```
lm(response ~ explanatory, data = yourdataset) |>
  anova()
```

```
praise()
```

```
[1] "You are glorious!"
```