Lesson 7: Do Diets Really Work? (continued #2)

CDT I.M. Smrt

Background

Recall that in Lesson 5 and 6 we investigated the impact diet (Vegetarian, Keto, and None) has on weight loss (kg lost).

```
library(tidyverse)
library(pwr) # see https://www.statmethods.net/stats/power.html for good tutorial.

## Load data
dietdat <- read_csv('https://raw.githubusercontent.com/jkstarling/MA376/main/WeightLoss.csv')

# glimpse(dietdat)
## change data types appropriately.
dietdat <- dietdat %>% mutate(diet = as.factor(diet)) %>%
  mutate(exercise = as.factor(exercise))
```

(1) What was our null and alternative hypotheses (in words and symbols, assuming we are going to conduct ANOVA)?

(2) For the diet study, what is a Type I error/Type I error rate? Type II error/Type II error rate?

(3) What is statistical power?

(4) What elements of a study can affect statistical power?

(5) Focusing on the F-statistic, we can explore the impact of how the elements will affect statistical power.

$$F = \left[\frac{R^2}{1 - R^2}\right] \times \left[\frac{n - \#\text{ of groups}}{\#\text{ of groups - 1}}\right]$$

Using the code from Lesson 6:

```
# contrasts(dietdat$diet) <- contr.sum
# multi.lm <- lm(xxx~xxx, data=dietdat)
# summary(xxx)
# anova(xxx)
```

We saw that we strong evidence to reject the null hypothesis ('rather large' F-statistic and 'rather small' p-value).

Let's see what happens when we reduce the number of observations.

```
dietdat2 <- dietdat %>% group_by(diet, exercise)  %>%  slice_head(n=10)
```

(6) Create a multiple-means model using 'dietdat2' data. Calculate the F statistic and p-value. Use the name 'multi.lm2'. Explain the results.

(7) Let's see what happens when we reduce the number of groups by combining Vegetarian and Keto diets.

Create a multiple-means model using 'dietdat3'. Calculate the F statistic and p-value. Use the name 'multi.lm3'. Explain the results.

```
# dietdat3 <- dietdat %>% mutate(kv=fct_collapse(diet, KV = c("Keto", "Vegetarian")))
# xxx
```

(8) Let's see what happens when we reduce the number of groups by combining Vegetarian and None diets (not recommended... why?).

Create a multiple-means model using 'dietdat4'. Calculate the F statistic and p-value. Use the name 'multi.lm4'. Explain the results.

```
dietdat4 <- dietdat %>% mutate(nv=fct_collapse(diet, NV = c("None", "Vegetarian")))
```

(9) Now we will go over HW #1.6.11 to show how calculate power in R:

Suppose you are testing to see whether, on average, heart rates measured in beats per minute (bpm), after jumping for 30 seconds are higher than after sitting for 30 seconds. There are 20 students from your statistics class who have volunteered to participate. You randomly assign 10 to the jumping treatment and 10 to the sitting treatment. You may assume that the SD of heart rates within each condition is 8 beats per minute.

a. *Use the Comparing Two Populations (we will use R) applet with both means set to 80. Use a significance level of Which of the following is the best rejection region for detecting that the population mean heart rate after jumping is higher than after sitting?*

How do we account for this without the applet? ANS: Use the standardized t-statistic:

$$t^*_{df} = \frac{\overline{\Delta} - \mu_0}{s/\sqrt{n}}$$

or rewriting it (with $\mu_0 = 0$):

$$\overline{\Delta} = t^*_{df} \cdot \frac{s}{\sqrt{n}}$$

```
# qt(xxx) * xxx / sqrt(xxx)
```

b. If you find a difference in the mean heart rate of 7 bpm, what will you conclude about the two treatments?

c. If you find a difference in the mean heart rate of 4 bpm, what will you conclude about the two treatments?

d. What is the probability of making a Type I error?

e. What is the probability of rejecting the null hypothesis if the difference in mean heartbeats is 5 bpm?

```
# n_students <- xxx
# sd <- xxx
# del <- xxx
# pwr.t.test(n=n_students, d=del/sd, sig.level = 0.05, alternative = "greater")
```

f. What is the probability of rejecting the null hypothesis if the difference in mean heartbeats is 10 bpm?

g. (Fill out the table... find the power for bpm = 5, 10, 15, 20, 25)