MA256 Lesson 12 - Paired Data (7.1-7.3)

Review of methods so far...

	1-Categorical	1-Quantitative	2-Categorical	1-Cat/1-Quant
parameter				
statistic				
test				
SE				
stand. stat				
valid. conds.				
CI				

Q1) Are the examples above (and the ones we have considered throughout the semester) representative of *independent groups designs* or *paired designs*? Why?

Q2) What is a *paired design*? What is the difference between a *paired design using matching* and a *paired design using repeated measures*?

Q3) What are a few benefits from using a *paired design*?

Q4) Using the ACFT as an example, explain how you would set up 1) a *paired design using matching* and 2) a *paired design using repeated measures.*

Theory-based approach for paired samples

Parameter/statistic:

H_0 :

H_a :

Strength of Evidence:

Confidence Interval:

Braking Reaction Time and Social Media

Researchers at Arizona State University (McNabb & Gray, 2016) explored the effects on driving with various types of cell phone use. In particular, they were interested in comparing the effects between text-based media and picture-based media. They had their subjects use a driving simulator and requested that they stay two seconds behind the car in front of them. The car in front traveled between 55 and 65 mph and was programmed to come to a complete stop eight times during the simulation. One of the variables they measured was the reaction time for the subjects to brake when the car in front of them stopped. They measured this when the subjects were instructed to scroll through Facebook messages that consisted of just text. They also measured this when the subjects were instructed to scroll through Instagram pictures that did not contain any text. The order of the conditions was randomized. The results, in seconds, are in the file `BrakeReactionTime`. We want to decide whether there is a difference in the average braking reaction time when looking at text (Facebook) on your phone while driving and looking at pictures (Instagram) on your phone while driving.

```
# library(tidyverse)
# braking <- read.table("https://www.isi-stats.com/isi/data/chap7/AP/BrakeReactionTime.txt",
#                        header=TRUE, stringsAsFactors = TRUE)
# braking_long <- braking %>% gather(key = "socialmedia", value = "time", Facebook, Instagram)
```

1a. *Conduct a two-sample t-test*, comparing the difference in mean braking times. What are the explanatory and response variables in this study and what types of variables are they (categorical or quantitative)?

1b. State the hypotheses either in words or using appropriate symbols.

1c. Plot histograms for each of the social media types.

```
# braking_long %>% ggplot(aes(x=XXXXX, fill=XXXXX)) + geom_histogram(binwidth = 0.15, position = "identity
```

1d. Assume validity conditions are met. Calculate the appropriate standardized statistic, p-value, and 95% confidence interval. What is your conclusion?

```
# mean.fb <- XXXXX
# mean.insta <- XXXXX
# sd.fb <- XXXXX
# sd.insta <- XXXXX
# n.fb <- XXXXX
# n.insta <- XXXXX
# n <- XXXXX
#
# SE <- XXXXX
# mystat <- XXXXX
# t <- XXXXX
#
# print(paste0("Our standardized stat is : ", round(t,4)))
# #CI
#
# pval <- XXXXX
#
#
# print(paste0("Our pvalue is : ", round(pval,4)))
# #CI
# multiplier <- XXXXX
# print(paste0("Our CI is: (", round(mystat - multiplier * SE,3),", ", round(mystat + multiplier * SE,3),
#
# print("XXXXXXXXXX Put your conclusion here XXXXXXXXXX")
```

1e. What is an issue with the analysis we did above? Can we do better?

1f. Conduct a paired t-test. State the hypotheses either in words or using appropriate symbols.

1g. Assume validity conditions are met. Calculate the appropriate standardized statistic, p-value, and 95% confidence interval. What is your conclusion?

```
# braking$delta <- XXXXX - XXXXX
#
# xbar_del <- XXXXX
# sd_del <- XXXXX
# n_del <- XXXXX
# SE <- XXXXX
#
# t.pair <- XXXXX
# print(paste0("Our std. stat is ", round(t.pair, 3), ". The mean difference in braking reaction time betw
# using Facebook and Instagram is ", round(xbar_del, 3)," seconds in the sample."))
#
# p.pair <- XXXXX
# print(paste0("Our p-value is ", round(p.pair, 3)))
#
# multiplier <- XXXXX
# print(paste0("Our CI is: (", round(xbar_del - multiplier * SE,3),", ", round(xbar_del + multiplier * SE,
```

1h. Based on the confidence interval, is there strong evidence that there is, on average, a difference in the average breaking reaction time between looking at text (Facebook) and looking at pictures (Instagram) on the phone? Explain why or why not.

Music and Chimpanze Soothing

You may have heard the phrase "Music soothes the savage beast," which is actually a misquote of a line of poetry by William Congreve. Researchers (Wallace et al., 2017) examined what effect music had on captive chimpanzees. In particular, they compared the number of incidences of social behavior described as playing, grooming another chimpanzee, or being groomed by another chimpanzee in an environment when music was playing versus the same chimpanzee behavior when no music was playing. We want to test to see whether, on average, there is difference in the number of incidences of social behavior with the chimpanzees between when music is present and when it is not. Use the data from the file `MusicSocial` to answer (a)-(e).

```
# chimp.social <- read.table("https://www.isi-stats.com/isi/data/chap7/MusicSocial.txt",
#                            header=TRUE, stringsAsFactors = TRUE)
```

2a. Explain why these data should be considered paired.

2b. What are the mean number of social behaviors for each condition (with music vs. without music)? Which condition resulted in more social behaviors on average?

```
# music.mean <- XXXXX
# nomusic.mean <- XXXXX
# c(music.mean, nomusic.mean)
```

2c. What is the average difference in the number of social behaviors (music – no music)?

```
# chimp.social$delta <- XXXXX
#
# xbar_d <- XXXXX
# print(xbar_d)
```

2d. Assume validity conditions are met. Calculate the appropriate standardized statistic, p-value, and 95% confidence interval.

```
# sd_d <- XXXXX
# n.d <- XXXXX
# SE <- XXXXX
#
# t.pair <- XXXXX
# print(paste0("Our std. stat is ", round(t.pair, 3)))
#
# p.pair <- XXXXX
# print(paste0("Our p-value is ", round(p.pair, 3)))
#
# multiplier <- XXXXX
# print(paste0("Our CI is: (", round(xbar_d - multiplier * SE,3),", ", round(xbar_d + multiplier * SE,3),
```

2e. State a conclusion in the context of the study.

Music and Chimpanze Aggression

Consider the background of the previous question. The researchers also kept track of the total number of aggressive events each chimpanzee displayed and divided these by the total time that an individual was present in the condition. This gives the rates of aggression per individual per hour in the music condition and in the no music condition. Only the 11 chimpanzees that showed any aggression were included in this analysis. We would like to investigate whether there was a significant difference in the aggression rates between the music and no music conditions. The data can be found in the file `MusicAggression` to answer (a)–(d).

```
# library(tidyverse)
# chimp.agg <- read.table("https://www.isi-stats.com/isi/data/chap7/MusicAggression.txt",
#                        header=TRUE, stringsAsFactors = TRUE)
```

3a. What are the mean rates of aggression for each condition (music and no music)? Which condition resulted in the higher aggression rate on average?

3b. What is the average difference in the number of rates of aggression (music – no music)?

```
# chimp.agg$delta <- XXXXX
#
# xbar_d <- XXXXX
# print(xbar_d)
```

3c. Assume validity conditions are met. Calculate the appropriate standardized statistic, p-value, and 95% confidence interval.

```
# sd_d <- XXXXX
# n.d <- XXXXX
# SE <- XXXXX
#
# t.pair <- XXXXX
# print(paste0("Our std. stat is ", round(t.pair, 3)))
#
# p.pair <- XXXXX
# print(paste0("Our p-value is ", round(p.pair, 3)))
#
# multiplier <- XXXXX
# print(paste0("Our CI is: (", round(xbar_d - multiplier * SE,3),", ", round(xbar_d + multiplier * SE,3),
```

3d. State a conclusion in the context of the study.

Tattoos and sweat rates

Researchers in Australia (Rogers et al., 2019) investigated whether tattoos affect sweat rates when exercising. They recruited 22 subjects, each with a tattoo that was more than 2 months old, larger than 11.4 cm2, and more than 50% shaded. The participants cycled on a stationary bike for 20 minutes, and the researchers measured the sweat rate from the tattooed area as well as the sweat rate from the non-tattooed area on the opposite side of the person's body. (For example, if the tattoo was on a person's right forearm, the researchers would also measure the sweat rate from the person's left forearm.) Is there a significant difference in the sweat rates between a tattooed area and a non-tattooed area? Use the results in mg/cm2 per minute found in the file `Tattoo` to answer (a)-(e).

```
# library(tidyverse)
# tats <- read.table("https://www.isi-stats.com/isi/data/chap7/Tattoo.txt",
#                    header=TRUE, stringsAsFactors = TRUE)
```

4a. Explain why these data should be considered paired.

4b. What is the mean sweat rates for each condition? Which condition resulted in a higher average sweat rate?

4c. What is the average difference in sweat rates (tattoo – no tattoo)?

4d. Assume validity conditions are met. Calculate the appropriate standardized statistic and p-value.

4e. State a conclusion in the context of the study.