MA256 Lesson 12 - Paired Data (7.1-7.3)

Review of methods so far...

	1-Categorical Ch 3.2	1-Quantitative Ch. 3.3	2-Categorical Ch. 5	1-Cat/1-Quant Ch. 6
parameter	π	μ	$\pi_1 - \pi_2$	$\mu_1 - \mu_2$
statistic	$\hat{p}$	$\bar{x}$	$\hat{p}_1 - \hat{p}_2$	$\bar{x}_1 - \bar{x}_2$
test	one-prop z-test	one sample t-test	two-sample z-test	two-sample t-test
SE	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$s/\sqrt{n}$	$\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$	$\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$
stand. stat	$z = \frac{\hat{p}-\pi}{SE}$	$t = \frac{\bar{x}-\mu_0}{SE}$	$z = \frac{\hat{p}_1-\hat{p}_2-\pi_0}{SE}$	$t = \frac{\bar{x}_1-\bar{x}_2-\mu_0}{SE}$
valid. conds.	10 succ/10 fails	Symmetric distr. OR $\geq$ 20 obs. and sample dist no strongly skewed	10 obs in each box of a 2-way table	$\geq$ 20 obs. in each and not strongly skewed
CI	$\hat{p} \pm z^*_{(1-\alpha/2)} \cdot SE$	$\bar{x} \pm t^*_{(1-\alpha/2,n-1)} \cdot SE$	$\hat{p}_1 - \hat{p}_2 \pm z^*_{(1-\alpha/2)} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	$\bar{x}_1 - \bar{x}_2 \pm t^*_{(1-\alpha/2,n-2)} \cdot SE$

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

We have assumed independence between each observation with the studies we have looked at so far. There were no systematic connections relating individuals in one group to individuals in another group.

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

A paired design has response values that come in pairs, with one response value in the pair for Group 1 and the other for Group 2. Paired Design using matching - the pairs come from matching similar individuals to create groups of two. Paired Design using repeated measures - the pairs come from measuring the same individual twice, once under each condition

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

Since we are accounting for variability in the observational/experimental unit, we can have more powerful tests and narrower confidence intervals.

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.*

PD-matching: Two people/items/etc. become independent study groups when they are paired due to having nearly identical characteristics. Example: use some method to identify pairs such as similar body types, GPA, other physical test (IOCT for example) and have one of them take the ACFT after a training regimen and the other take the ACFT with a different training regimen or no training regimen.
PD-repeated measures: Each observational unit (person/item, etc.) participates is measured in both conditions. Example: Measure ACFT performance before a formal required ACFT training regimen and again afterwards for each individual/participant.

Theory-based approach for paired samples

Parameter/statistic: μ_d and $\bar{x}_d (= \frac{1}{n}\sum_{i=1}^{n}(x_{i2} - x_{i1})$

$H_0 : \mu_d = \mu_0$, where μ_0 is a number (typically 0)

$H_a : \mu_d \neq \mu_0$

Strength of Evidence: Calculate t statistic: $t = \frac{\bar{x}_d - \mu_0}{s_d/\sqrt{n}}$ reject H_0 for t "more extreme" using the guidlines for appropriate significance level.

Confidence Interval: $\bar{x}_d \pm t^*_{(1-\alpha/2, n-1)} \times s_d/\sqrt{n}$ reject H_0 if μ_0 is NOT in CI.

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)?

The explanatory variable is the type of media being looked at, Facebook or Instagram (categorical), and the response variable is the brake reaction time (quantitative).

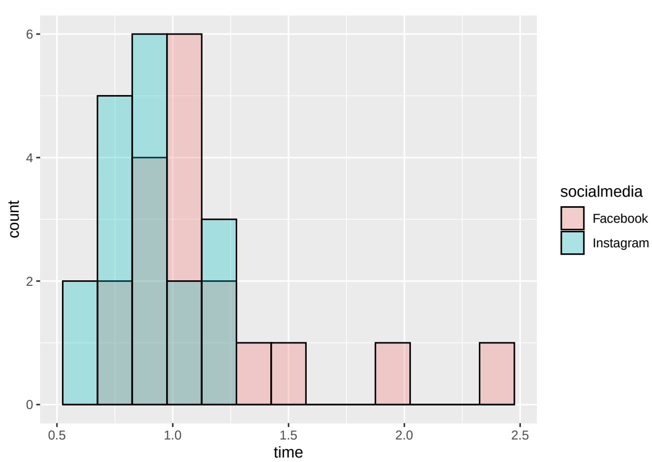
1b. State the hypotheses either in words or using appropriate symbols.

$H_0 : \mu_1 - \mu_2 = 0$. The reaction time for braking when looking at Facebook or Instagram is the same, on average.

$H_a : \mu_1 - \mu_2 \neq 0$. The reaction time for braking when looking at Facebook or Instagram is not the same, on average.

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

```
braking_long %>% ggplot(aes(x=time, fill=socialmedia)) + geom_histogram(binwidth = 0.15, position = "ident
```



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

```r
mean.fb <- mean(braking$Facebook)
mean.insta <- mean(braking$Instagram)
sd.fb <- sd(braking$Facebook)
sd.insta <- sd(braking$Instagram)
n.fb <- length(braking$Facebook)
n.insta <- n.fb
n <- n.fb + n.insta

SE <- sqrt((sd.fb^2 / n.fb) + (sd.insta^2 / n.insta))
mystat <- mean.fb - mean.insta
t <- (mystat) / SE

print(paste0("Our standardized stat is : ", round(t,4)))
```

```
## [1] "Our standardized stat is : 2.5925"
```

```r
#CI

pval <- 2*(1-pt(abs(t), n - 2))


print(paste0("Our pvalue is : ", round(pval,4)))
```

```
## [1] "Our pvalue is : 0.0139"
```

```r
#CI
multiplier <- pt(0.975, n - 2)
print(paste0("Our CI is: (", round(mystat - multiplier * SE,3),", ", round(mystat + multiplier * SE,3), ")
```

```
## [1] "Our CI is: (0.195, 0.379)"
```

```r
print("We have strong evidence to reject the null hypothesis, and state that we have evidence in different
```

```
## [1] "We have strong evidence to reject the null hypothesis, and state that we have evidence in differen
```

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

One issue is that there is dependence between the first measure and the second measure since it is the same subject/person.

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

$H_0 : \mu_d = 0$. There difference in the reaction time for braking when looking at Facebook or Instagram is zero, on average.

$H_a : \mu_d \neq 0$. The reaction time for braking when looking at Facebook or Instagram is not the same, on average.

Let μ_d represent the mean difference in braking reaction time between Facebook and Instagram; $H_0 : \mu_d = 0$; $H_a : \mu_d \neq 0$.

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

```r
braking$delta <- braking$Facebook - braking$Instagram

xbar_del <- mean(braking$delta)
sd_del <- sd(braking$delta)
n_del <- length(braking$delta)
```

```r
SE <- sd_del / sqrt(n_del)

t.pair <- xbar_del / SE
print(paste0("Our std. stat is ", round(t.pair, 3), ". The mean difference in braking reaction time betwee
using Facebook and Instagram is ", round(xbar_del, 3)," seconds in the sample."))
```

```
## [1] "Our std. stat is 2.966. The mean difference in braking reaction time between\nusing Facebook and I
```

```r
p.pair <- 2*(1-pt(abs(t.pair), n_del -1))
print(paste0("Our p-value is ", round(p.pair, 3)))
```

```
## [1] "Our p-value is 0.009"
```

```r
multiplier <- pt(0.975, n_del - 1)
print(paste0("Our CI is: (", round(xbar_del - multiplier * SE,3),", ", round(xbar_del + multiplier * SE,3)
```

```
## [1] "Our CI is: (0.207, 0.367)"
```

We have very strong evidence that there is a difference in average braking reaction times between drivers who are looking at Facebook on their phones compared to drivers who are looking at Instagram on their phones, with the longer reaction times, on average, for those looking at Facebook. We also see that we have stronger evidence than when we just used the two-sample t-test.

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.

Yes, because the entire interval is positive (or doesn't contain 0) there is strong evidence that the average braking reaction time is higher when drivers are looking at Facebook on their phones compared to when they are looking at Instagram on their phones.

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).

```r
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.

The data should be paired because the same chimpanzees are being compared under two conditions-the music condition and the no music condition.

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?

```r
music.mean <- mean(chimp.social$Music)
nomusic.mean <- mean(chimp.social$NoMusic)
c(music.mean, nomusic.mean)
```

```
## [1]  8.00000 16.33333
```

The mean for the no music condition (16.333) was larger than for the music condition (8.000).

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

```
chimp.social$delta <- chimp.social$NoMusic - chimp.social$Music

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

```
## [1] 8.333333
```

$x_d = 8.333$

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

```
sd_d <- sd(chimp.social$delta)
n.d <- length(chimp.social$delta)
SE <- sd_d / sqrt(n.d)

t.pair <- xbar_d / SE
print(paste0("Our std. stat is ", round(t.pair, 3)))
```

```
## [1] "Our std. stat is 5.054"
```

```
p.pair <- 2*(1-pt(abs(t.pair), n.d -1))
print(paste0("Our p-value is ", round(p.pair, 3)))
```

```
## [1] "Our p-value is 0"
```

```
multiplier <- pt(0.975, n.d - 1)
print(paste0("Our CI is: (", round(xbar_d - multiplier * SE,3),", ", round(xbar_d + multiplier * SE,3), ")
```

```
## [1] "Our CI is: (6.967, 9.699)"
```

$p - value < 0.0001$

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

Because the p-value is less than 0.01, we have very strong evidence that, on average, there is a difference in the number of incidences of social behavior among such chimpanzees between when music is present and when it is not. When music is not present, on average, more social incidences occur than when it is present.

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?

\sol{The mean aggression rate for the music condition (0.777) was higher than the no music condition (0.257).

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

```
chimp.agg$delta <- chimp.agg$NoMusic - chimp.agg$Music

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

```
## [1] -0.5202947
```

$x_d = -0.520$

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

```
sd_d <- sd(chimp.agg$delta)
n.d <- length(chimp.agg$delta)
SE <- sd_d / sqrt(n.d)

t.pair <- xbar_d / SE
print(paste0("Our std. stat is ", round(t.pair, 3)))
```

```
## [1] "Our std. stat is -1.724"
```

```
p.pair <- 2*(1-pt(abs(t.pair), n.d -1))
print(paste0("Our p-value is ", round(p.pair, 3)))
```

```
## [1] "Our p-value is 0.115"
```

```
multiplier <- pt(0.975, n.d - 1)
print(paste0("Our CI is: (", round(xbar_d - multiplier * SE,3),", ", round(xbar_d + multiplier * SE,3), ")
```

```
## [1] "Our CI is: (-0.769, -0.272)"
```

p-value ≈ 0.12

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

Because the p-value is greater than 0.10, we do not have strong evidence that, on average, there is a difference in the aggression rates of such chimpanzees between when music is present and when it is not.

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).

```r
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.

The data should be paired because the researchers measured the sweat rate from the tattooed area as well as the sweat rate from the nontattooed area on the opposite side of the same person's body.

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

```r
print(mean(tats$NonTatttoo))
```

```
## [1] 0.9351818
```

```r
print(mean(tats$Tattoo))
```

```
## [1] 0.922
```

The mean sweat rate for the nontattooed area (0.935) was larger than for the tattooed area (0.922).

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

```r
tats$delta <- tats$Tattoo - tats$NonTatttoo

xbar_d <- mean(tats$delta)
print(xbar_d)
```

```
## [1] -0.01318182
```

$\bar{x}_d = -0.013$

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

```r
sd_d <- sd(tats$delta)
n.d <- length(tats$delta)
SE <- sd_d / sqrt(n.d)

t.pair <- xbar_d / SE
print(paste0("Our std. stat is ", round(t.pair, 3)))
```

```
## [1] "Our std. stat is -0.4"
```

```r
p.pair <- 2*(1-pt(abs(t.pair), n.d -1))
print(paste0("Our p-value is ", round(p.pair, 3)))
```

```
## [1] "Our p-value is 0.693"
```

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

Because the p-value is larger than 0.10, we do not have strong evidence that, on average, there is a difference in the sweat rates between tattooed areas and nontattooed areas.