MA256 Lesson 13 - 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	$\hat{\pi}$	$\bar{x}$	$\hat{p}_1 - \hat{p}_2$	$\bar{x}_1 - \bar{x}_2$
test	one-prop z-test	one sample t-test	two-sample z-test	two-sample t-test
sd	$\sqrt{\frac{\pi(1-\pi)}{n}}$	$s/\sqrt{n}$	$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
stand. stat	$\frac{\hat{p}-\pi}{sd}$	$\frac{\bar{x}-\mu_0}{sd}$	$\frac{\hat{p}_1-\hat{p}_2-\mu_0}{sd}$	$\frac{\bar{x}_1-\bar{x}_2-\mu_0}{sd}$
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	$\hat{p} \pm z^*_{(1-\alpha/2)} \cdot sd$	$\bar{x} \pm t^*_{(1-\alpha/2,n-1)} \cdot sd$	$\hat{p}_1 - \hat{p}_2 \pm z^*_{(1-\alpha/2)} \cdot sd$	$\bar{x}_1 - \bar{x}_2 \pm t^*_{(1-\alpha/2,n-2)} \cdot sd$

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. pd using matching - the pairs come from matching similar individuals to create groups of two. pd 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: $t = \frac{\bar{x}_d - \mu_0}{s/\sqrt{n}}$ reject H_0 for t "more extreme" using the guidlines for appropriate significance level.

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

Rounding First Base

Imagine you are at the plate in baseball and have hit a hard line drive. You want to try to stretch your hit from a single to a double. Does the path that you take to "round" first base make much of a difference? For example, is it better to take a "narrow angle" or a "wide angle" around first base? This example is based on an actual study reported in a master's thesis by W. F. Woodward in 1970. In Woodward's study, he used a stopwatch to time 22 different runners going from a spot 35 feet past home to a spot 15 feet before second. He had each runner use each method, with a rest period in between. Here are the summary statistics on the difference in times (narrow - wide) for the 22 players:

Difference	Sample Size	Sample Average (sec)	Sample SD (sec)
narrow-wide	22	0.075	0.088

a. Use an appropriate theory-based method to investigate whether the choice of angle affects the running time. Be sure to find and report a standardized statistic and p-value. Interpret the value of the standardized statistic in the context of the study.

```
n <- 22
xdbar <- 0.075
se <- 0.088 / sqrt(n)
t <- xdbar / se
pval <- 2 * (1 - pt(t, n-1))
c(t, pval)
```

```
## [1] 3.9975134317 0.0006535864
```

b. Use an appropriate theory-based method to find and report a 95% confidence interval for the parameter of interest.

```
multiplier <- qt(0.975, n-1)
c(xdbar - multiplier * se, xdbar + multiplier * se)
```

```
## [1] 0.03598299 0.11401701
```

c. Provide a full (four-part) conclusion, including comments on significance, estimation, causation, and generalization.

We have strong evidence that the average difference in running time comparing narrow to wide angle is not 0, with longer running times for the narrow angle (on average between 0.036 and 0.114 seconds longer). This result suggests a cause-and-effect relationship between running time and choice of angle but does not necessarily generalize to all baseball players because little is known about how representative the baseball player samples are of the running characteristics of all players.

Filtering Water in Cameroon

Students and professors in the Nursing and Engineering Departments at Hope College went to the central African country of Cameroon to help improve drinking water quality and community health in rural communities by installing water filters in or near homes in one village. The families living in this village had no electricity, had no water distribution system, and got their water from streams. The filters they installed contained a diffuser plate, fine sand, coarse sand, and gravel. Using the new filters, family members were expected to gather their water and filter it before drinking, instead of drinking directly from the stream as was common practice.

Students working on this project examined the quality of the filters by looking at many different variables, including general observations, filter observations, microbiology observations, household practice observations, user perceptions, and water source observations. It should be noted, when making inferences, the water filters in this dataset should be treated as a sample of all filters that could be constructed if this pilot project were expanded to other villages. Thus, inference is to an as-yet-unbuilt, larger population of filters.

STEP 1: Ask a research question.

There are several research questions we will ask in this investigation. The first one is: On average, is there a significant difference in the E. coli counts between the water that has just been filtered and water that is sitting in the bottom of a filter after it was filtered the previous day? The dataset we will use contains results from 14 water filters each giving E. coli counts (per 100 mL) on the first day and the second day after the water was filtered.

STEP 2: Design a study and collect data.

1. What are the observational units?

The observational units are the filters.

2. What variables are measured/recorded on each observational unit?

Escherichia coli count on just filtered water and E.coli count on water filtered the day before

3. Identify the role of these variables (explanatory, response). Classify the variables in this study as categorical or quantitative.

These variables are quantitative, but we can also consider the difference in E. coli counts for each day as a quantitative response variable.

4. Are the samples of the first E. coli count independent or dependent of the samples of the second E. coli count? Explain. Based on your answer, is this an independent samples or paired design?

The samples of the first E. coli count are dependent of the samples of the second E. coli count because they came from the same filter. If the filter is functioning well, we would expect both E.coli counts to be low. If the filter is not functioning well, we would expect both E. coli counts to be higher. So we will perform a matched pairs test for the mean difference.

5. Could the sample size of 14 be large enough to give a valid p-value from a theory-based test of significance?

The validity conditions are questionable with a sample size of 14. It might be reasonable to use the theory-based t-test if the distribution of the differences is reasonably symmetric.

6. State the null and alternative hypotheses to be investigated with this study in symbols or in words.

$H_0 : \mu_d = 0$ and $H_a : \mu_d \neq 0$ where μ_d is the long-run average of the differences (Day 1 – Day 2) of E. coli counts.

STEP 3: Explore the data.

7. What are the average E. coli counts for each day? Did the E. coli in the sample increase or decrease on average from Day 1 to Day 2? Explain. Also give the standard deviations for the E. coli counts for each day.

```
library(tidyverse)
ecoli.time <- read.csv("https://raw.githubusercontent.com/jkstarling/MA256/main/data/Ecoli_time-1.csv")

ecoli.time %>%
  summarise(
    mean1=mean(Day_1),
    sd1= sd(Day_1),
    mean2=mean(Day_2),
    sd2= sd(Day_2)
  )
```

```
##      mean1      sd1    mean2      sd2
## 1 43.47857 64.18452 94.78571 82.45575
```

8. What is the average of the difference (Day 1 – Day 2) in E. coli counts? Does the sign of this average correspond to your answer of increasing or decreasing between Day 1 and Day 2 in #7? Explain. Also give the standard deviation for the differences in E. coli counts.

```
ecoli.time <- ecoli.time %>%
  mutate(Difference= Day_1 -Day_2)

ecoli.time %>%
  summarise(
    mean = mean(Difference),
    s = sd(Difference),
    n = n()
  )
```

```
##        mean        s  n
## 1 -51.30714 58.68702 14
```

See R output.

STEP 4: Draw inferences beyond the data.

9. Carry out an appropriate test of significance to see whether, on average, there is a genuine difference between the first E. coli count and the second E. coli count. Report your p-value and a 95% confidence interval.

```
xbar <- -51.63
n <- 14
se <- 58.77 / sqrt(n)
null <- 0
t <- (xbar-null) / se
pvalue <- 2*(1-pt(abs(t),n-1))
pvalue
```

```
## [1] 0.005892376
```

```
multiplier <- qt(.975, n-1)
CI <- c(xbar - multiplier * se, xbar + multiplier * se)
CI
```

```
## [1] -85.56279 -17.69721
```

See R output.

STEP 5: Formulate conclusions.

10. Based on your p-value and confidence interval, what conclusions can you draw from this test?

There is strong evidence against the null and in support of the alternative that there is a genuine difference between the mean E.coli on Day 1 and Day 2. On average there is between 17.697/100 mL and 85.563/100 mL more E. coli on Day 2 than on Day 1.

11. Can generalizations be made to a larger population? Is there a cause-effect relationship between the variables?

We can generalize to filters yet to be made as this is being treated as pilot data. We cannot draw a cause-and-effect conclusion as this was an observational study and there are many possible confounding variables.

STEP 6: Look back and ahead.

12. Discuss the study design. Why does pairing make sense here? What would you do differently to improve upon this study? What further research would you propose based on your findings from this study?

Researchers were wondering whether there was a difference in E. coli counts from just filtered water compared to water filtered the day before. Data were gathered on 14 filters in a village in Cameroon. The average E. coli count of just filtered water was 43.48/100 mL of water and 94.79/100 mL of water for the water filtered the day before. A matched pairs test was run on the data and resulted in a p-value of 0, which offered very strong evidence against the null and in support of the alternative that there is a difference in the average E. coli counts between Day 1 and Day 2 of filtered water. We can generalize these results to the population of filters yet to be built but are not able to draw any causal inference as this was an observational study, not a randomized experiment. Further research could be done on newer filters, also looking at more than 1-day-old water.

12-REDUX. What if we did not use a paired test? Conduct a different test, but compare the difference in the measurements from Day 1 and Day 2. What is our conclusion? (assume validity conditions are met for a theory based test) Are there any issues with conducting this test?

```
mysummary <- ecoli.time %>% summarise(mean1 = mean(Day_1), mean2 = mean(Day_2), sd1 = sd(Day_1), sd2 = sd(
mysummary
```

```
##      mean1    mean2      sd1      sd2
## 1 43.47857 94.78571 64.18452 82.45575
```

```
xbar1 <- mysummary[1,1]
xbar2 <- mysummary[1,2]
s1 <- mysummary[1,3]
s2 <- mysummary[1,4]
n1 <- 14
n2 <- 14
se <- sqrt(s1 / n1 + s2 / n2)
null <- 0
statistic <- xbar1  - xbar2
t <- (statistic - null) / se
n <- n1 + n2
pvalue <- pt(t, n - 2)
pvalue
```

```
## [1] 3.506649e-15
```

We see that we have stronger evidence against the null hypothesis. One issue is that we did not account for the dependence among the observations from Day 1 to Day 2. (and our validity conditions aren't met.)

PART II

As mentioned earlier, sand is used in the filter. Different filters had different amounts of sand. The students split the filters into two groups: those that had more than 2 inches of sand and those that had less. We will explore whether this difference results in different E. coli counts when filtering the water. In this sample, we have results from 19 filters, 14 with more than 2 inches of sand and 5 with less than 2 inches of sand.

13. Identify the explanatory and response variables in this study. Classify the variables in this study as categorical or quantitative.

Explanatory variable is amount of sand and response is E. coli counts. Amount of sand is categorical (>2 inches or <2 inches) and E. coli counts is a quantitative variable.

14. What are the average E. coli counts for each sand level? Are the counts higher or lower with the higher level of sand in the filter? Also give the standard deviations for the E. coli counts for each sand level.

```
ecoli.sand <- read.csv("https://raw.githubusercontent.com/jkstarling/MA256/main/data/Ecoli_sand-1.csv")

ecoli.sand %>%
  group_by(sand_level) %>%
  summarise(
    mean_sand=mean(day_1),
    sd_sand= sd(day_1),
    n=n())
```

```
## # A tibble: 2 x 4
##   sand_level mean_sand sd_sand     n
##   <chr>          <dbl>   <dbl> <int>
## 1 above           38.8    70.0    14
## 2 below           80.6    96.8     5
```

See R output above.

15. Are the samples of the first E. coli count independent or dependent of the samples of the second E. coli count? Explain.

Independent. The filters that have >2 inches of sand are different filters than those with <2 inches of sand. Independent two-sample test for means.

16. Is the sample large enough to use a theory-based test?

No, the sample size is not large enough to conduct a theory-based test.

17. (Assume you can conduct a theory-based test and...) Carry out an appropriate test of significance to see whether on average there is a difference in the E. coli counts between filters with more than 2 inches of sand and those with less. Report your p-value, a 95% confidence interval, and your conclusion.

```
xbar_above <- 38.83571
xbar_below <- 80.600
s_above <- 70.006
s_below <- 96.83388
n_above <- 14
n_below <- 5
se <- sqrt(s_above^2/n_above + s_below^2/n_below)
null <- 0
```

```
statistic <- xbar_above - xbar_below
t <- (statistic-null) / se

n <- n_above+n_below
pvaluesand <- 2*(1-pt(abs(t), n-2)); pvaluesand
```

```
## [1] 0.3883366
```

```
multiplier <- qt(.975,n-2)
CIsand <- c(statistic - multiplier * se, statistic + multiplier * se)
c(t, pvaluesand, CIsand)
```

```
## [1]   -0.8853176    0.3883366 -141.2934857   57.7649057
```

We have weak evidence against the null and so conclude it is plausible that the average E. coli counts are the same for filters with >2 inches of sand and filters with <2 inches of sand.

PART III

Let's look at one last research question. As a general rule, the flow rate of a water filter in good working condition should be around 1,000 mL/min. Let's test to see whether these water filters had an average flow rate that is significantly different than 1,000 mL/min.

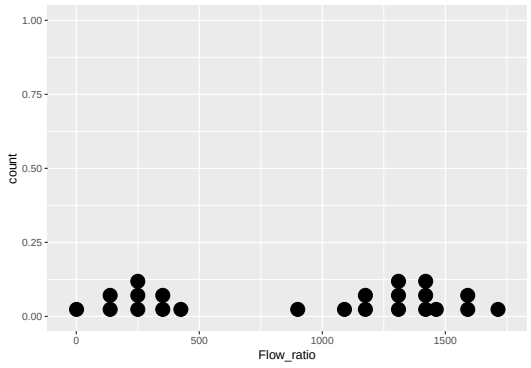
18. List some similarities and differences between this research question and our first research question looking at the average difference in first and second E. coli counts.

Similarities: Both are looking at a quantitative response variable, and both are a test on a single mean.
Differences: The first question was a matched pairs analysis on two dependent measures; this is comparing the quantitative response to a parameter value that makes contextual sense.

19. Create a dotplot of the flow rates. Include average and standard deviation for the flow rate. What do you observe?

```
ecoli.flow <- read.csv("https://raw.githubusercontent.com/jkstarling/MA256/main/data/Ecoli_flow-1.csv")

ecoli.flow %>%
  ggplot(aes(x=Flow_ratio))+geom_dotplot()
```




```
ecoli.flow %>%
  summarise(
    meanflow = mean(Flow_ratio),
    sflow = sd(Flow_ratio),
    nflow = n()
  )
```

```
##   meanflow    sflow nflow
## 1 913.5652 582.8859    23
```

Dotplot of the flow rates shows two clumps of data.

20. Carry out an appropriate test of significance to see whether the Cameroon water filters (population) tend to flow at a rate different than 1,000 mL/min, or is an average of 1,000 mL/min plausible? State your hypotheses, 95% confidence interval, p-value, and conclusions.

```
xbar <- 913.6
s <- 582.9          # sample standard deviation
n <- 23           # sample size
se <- s / sqrt(n)  # standard error of the null distribution
null <- 1000
t <- (xbar-null) / se
pvalue_flow <- 2*(1-pt(abs(t), n-1))

multiplier <- qt(.975, n-1) # 95% CI
CI <- c(xbar-multiplier*se, xbar+multiplier*se)
c(t, pvalue_flow, CI)
```

```
## [1]   -0.7108592    0.4846403  661.5351380 1165.6648620
```

We have weak evidence against the null and so it is plausible that on average the filters are flowing at 1000

21. You should have found that 1,000 mL/min is a plausible value for average flow rate for the population of all similar water filters. Does this mean that all the filters have flow rates of about 1,000 mL/min? Is testing a single mean here really telling us what we need to know about these filters? Why or why not?

Recalling what the dotplot looked like when we explored our data, it appears that roughly half of the filters are filtering too quickly and the other half are filtering too slowly. This does make the average where we want it, but none of the filters individually are doing a good job of filtering, at least from a flow standpoint. Exploring the data is a very important part of the six-step method.