Stat 13, Intro. to Statistical Methods for the Life and Health Sciences.

1. Get midterms back.
2. Comparing two means and biking to work example.
3. Paired data.

Read chapter 7.

HW3 is due Wed, Feb26, 1159pm. 4.CE.10, 5.3.28, 6.1.17, and 6.3.14.
On 5.3.28d, use the theory-based formula. You do not need to use an applet.

The course website is http://www.stat.ucla.edu/~frederic/13/W25 .

1. Get midterms back.
Please be quiet until all the midterms have been returned.

Mean was 70%.
SD was 21%.
90-100 = A range,
80-90 = B range,
70-80 = C range,
60-70 = D range,
below 60 = F.

2. Comparing Two Means: Simulation-Based Approach and bicycling to work example.

Section 6.2

Similar to proportions.

- We will be comparing means, much the same way we compared two proportions using randomization techniques.
- The difference here is that the response variable is quantitative (the explanatory variable is still binary though). So if cards are used to develop a null distribution, numbers go on the cards instead of words.

Bicycling to Work

Example 6.2

Bicycling to Work

- Does bicycle weight affect commute time?
- British Medical Journal (2010) presented the results of a randomized experiment done by Jeremy Groves, who wanted to know if bicycle weight affected his commute to work.
- For 56 days (January to July) Groves tossed a coin to decide if he would bike the 27 miles to work on his carbon frame bike (20.9lbs) or steel frame bicycle (29.75lbs).
- He recorded the commute time for each trip.

Bicycling to Work

- What are the observational units?
  - Each trip to work on the 56 different days.
- What are the explanatory and response variables?
  - Explanatory is which bike Groves rode (categorical – binary)
  - Response variable is his commute time (quantitative)

Bicycling to Work

- **Null hypothesis:** Commute time is not affected by which bike is used.
- **Alternative hypothesis:** Commute time is affected by which bike is used.

Bicycling to Work

- In chapter 5 we used the difference in proportions of "successes" between the two groups.
- Now we will compare the difference in averages between the two groups.
- The parameters of interest are:
  - μ_{carbon} = Long term average commute time with carbon framed bike
  - μ_{steel} = Long term average commute time with steel framed bike.

Bicycling to Work

- μ is the population mean. It is a parameter.
- Using the symbols μ_{carbon} and μ_{steel}, we can restate the hypotheses.

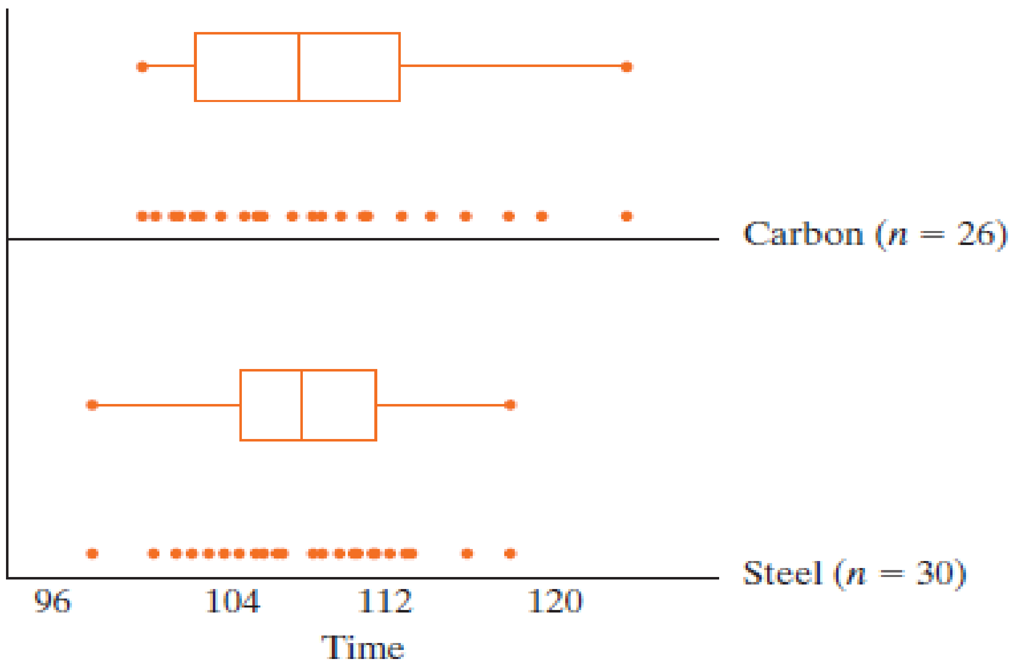
- **H_0:** $\mu_{carbon} = \mu_{steel}$
- **H_a:** $\mu_{carbon} \neq \mu_{steel}$.

Bicycling to Work

Remember:

- The hypotheses are about the longterm association between commute time and bike used, not just his 56 trips.
- Hypotheses are always about populations or processes, not the sample data.

Bicycling to Work



	Sample size	Sample mean	Sample SD
Carbon frame	26	108.34 min	6.25 min
Steel frame	30	107.81 min	4.89 min

Bicycling to Work

- The sample mean was higher for the carbon framed bike.
- Does this indicate the bike is better?
- Or could a higher average just come from the random assignment? Perhaps the carbon frame bike was randomly assigned to days where traffic was heavier or weather slowed down Dr. Groves on his way to work?

Bicycling to Work

- **Statistic:**
- The observed difference in average commute times

$\bar{x}_{\text{carbon}} - \bar{x}_{\text{steel}}$ = 108.34 - 107.81

= 0.53 minutes

Bicycling to Work

Simulation:

- We can imagine simulating this study with index cards.
  - Write all 56 times on 56 cards.
- Shuffle all 56 cards and randomly redistribute into two stacks:
  - One with 26 cards (representing the times for the carbon-frame bike)
  - Another 30 cards (representing the times for the steel-frame bike)

Bicycling to Work

Simulation (continued):

- Shuffling assumes the null hypothesis of no association between commute time and bike
- After shuffling we calculate the difference in the average times between the two stacks of cards.
- Repeat this many times to develop a null distribution

Carbon Frame

116	114	119	123	113
111	113	106	118	109
103	103	104	112	110
101	102	100	102	107
105	103	111	106	102
108				

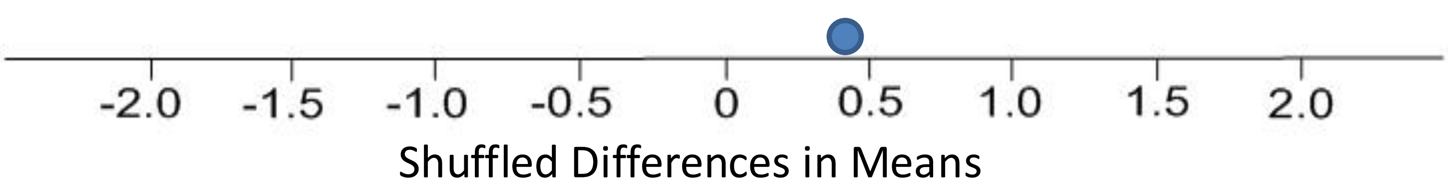
mean = 108.27

Steel Frame

116	116	109	118	113
110	113	104	113	105
111	111	110	105	106
103	102	98	109	108
102	112	101	106	102
105	105	106	107	106

mean = 107.87

108.27 – 107.87 = 0.40



Carbon Frame

116	114	119	123	113
111	113	106	118	109
103	103	104	112	110
101	102	100	102	107
105	103	111	106	102
108				

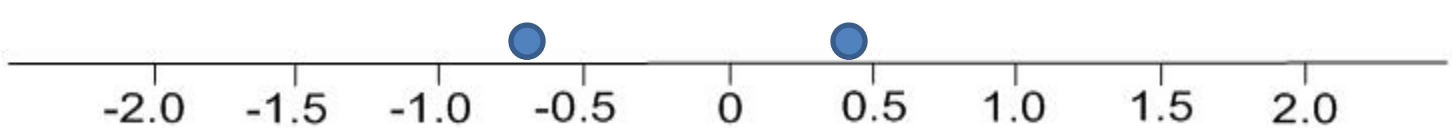
mean = 107.69

Steel Frame

116	116	109	118	113
110	113	104	113	105
111	111	110	105	106
103	102	98	109	108
102	112	101	106	102
105	105	106	107	106

mean = 108.37

107.69 – 108.37 = -0.68



Shuffled Differences in Means

Carbon Frame

116	114	119	123	113
111	113	106	118	109
103	103	104	112	110
101	102	100	102	107
105	103	111	106	102
108				

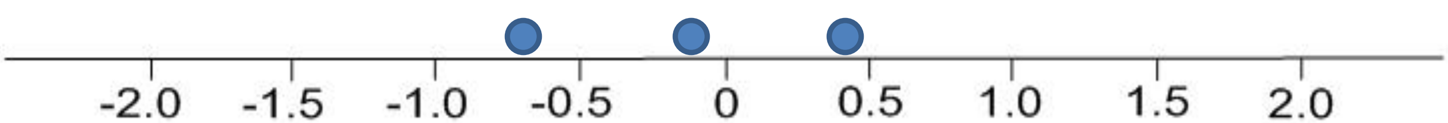
mean = 107.97

Steel Frame

116	116	109	118	113
110	113	104	113	105
111	111	110	105	106
103	102	98	109	108
102	112	101	106	102
105	105	106	107	106

mean = 108.13

107.97 – 108.13 = -0.16

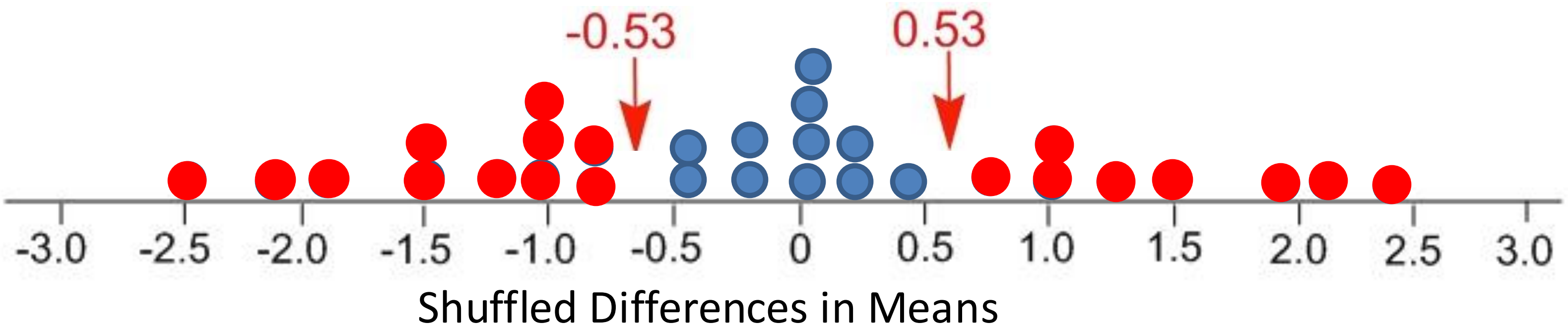


Shuffled Differences in Means

More Simulations

-2.50 -2.11 -1.93 -1.53 -1.51
-1.20 -1.21 -1.10 -1.11
-0.98 -0.64 -0.31 0.02 -0.52
2.53 1.90 0.71
0.44 0.38 0.22
1.89 0.13
1.46 0.81 0.55 1.50
1.79

Nineteen of our 30 simulated statistics were as or more extreme than our observed difference in means of 0.53, hence our estimated p-value for this null distribution is 19/30 = 0.63.



Bicycling to Work

- Using 1000 simulations, we obtain a p-value of 72%.
- What does this p-value mean?
- If mean commute times for the bikes are the same in the long run, and we repeated random assignment of the carbon bike to 26 days and the steel bike to 30 days, a mean difference as extreme as 0.53 minutes or more would occur in about 72% of the simulations.
- Therefore, we do not have strong evidence that the commute times for the two bikes will differ in the long run. The difference between bikes observed by Dr. Groves is not statistically significant.

Bicycling to Work

- Have we proven that the bikes are equivalent? (Can we conclude the null is true?)
  - No, a large p-value is not "strong evidence that the null hypothesis is true."
  - It suggests that the null hypothesis is consistent with the data.
  - There could be no long-term difference.
    But there also could be a small long-term difference.

Bicycling to Work

- Imagine we want to generate a 95% confidence interval for the long-run difference in average commuting time.
  - Sample difference in means ± 1.96×SE for the difference between the two means
- From simulations, the SE = standard deviation of the simulated differences between sample means = 1.47.
- $0.53 \pm 1.96(1.47) = 0.53 \pm 2.88$
- -2.35 to 3.41.
- What does this mean?

Bicycling to Work

- We are 95% confident that the true longterm difference (carbon – steel) in average commuting times is between -2.41 and 3.47 minutes.
- We are 95% confident the carbon framed bike is between 2.41 minutes faster and 3.47 minutes slower than the steel framed bike.
- Does it make sense that the interval contains 0, based on our p-value?

Bicycling to Work

- Was the sample representative of an overall population?
- What about the population of all days Dr. Groves might bike to work?
  - No, Groves commuted on consecutive days in this study and did not include all seasons.
- Was this an experiment? Were the observational units randomly assigned to treatments?
  - Yes, he flipped a coin for the bike.
  - We can probably draw cause-and-effect conclusions here.

Bicycling to Work

- We cannot generalize beyond Groves and his two bikes.
- A limitation is that this study is not ***double-blind.***
  - The researcher and the subject (which happened to be the same person here) were not blind to which treatment was being used.
  - Dr. Groves knew which bike he was riding, and this might have affected his state of mind or his choices while riding.

2. Paired Data.

Chapter 7

- The paired data sets in this chapter have one *pair* of quantitative response values for each obs. unit.
- This allows for a comparison where the other possible confounders are as similar as possible between the two groups.
- Paired data studies remove individual variability by looking at the difference score for each subject.
- Reducing variability in data improves inferences:
  - Narrower confidence intervals.
  - Smaller p-values when the null hypothesis is false.
  - Less influence from confounding factors.

- The main idea is to look at the difference between responses, and then analyze these differences the way we analyzed one variable previously.

Paired data and studying with music example.

Example 7.1

Studying with Music

- Many students study while listening to music.
- Does it hurt their ability to focus?
- In "Checking It Out: Does music interfere with studying?" Stanford Prof Clifford Nass claims the human brain listens to song lyrics with the same part that does word processing.
- Instrumental music is, for the most part, processed on the other side of the brain, and Nass claims that listening to instrumental music has virtually no interference on reading text.

Studying with Music

Consider the experimental designs:

Experiment A — Random assignment to 2 groups

- 27 students were randomly assigned to 1 of 2 groups:
  - One group listens to music with lyrics.
  - One group listens to music without lyrics.
- Students play a memorization game while listening to the particular music that they were assigned.

Studying with Music

Experiment B — Paired design using repeated measures

- All students play the memorization game twice:
  - Once while listening to music with lyrics
  - Once while listening to music without lyrics.

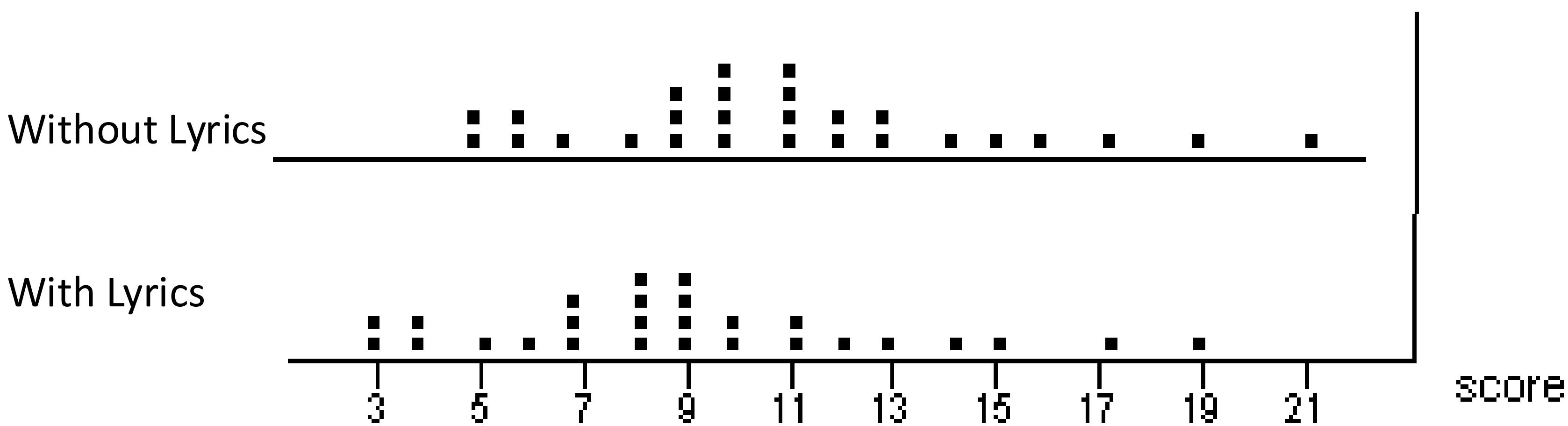
Experiment C — Paired design using matching

- Sometimes repeating something is impossible (like testing a surgical procedure) but we can still pair.
  - Test each student on memorization.
  - Match students up with similar scores and randomly:
    - Have one play the game while listening to music with lyrics and the other while listening to music without lyrics.

Studying with Music

We will focus on the repeated measures type of pairing.

- What if everyone could remember exactly 2 more words when they listened to a song without lyrics?
- Using Experiment A, there could be a lot of overlap between the two sets of scores and it would be difficult to detect a difference, as shown here.



Studying with Music

- Variability in people's memorization abilities may make it difficult to see differences between the songs in Experiment A.
- The paired design focuses on the *difference* in the number of words memorized, instead of the number of words memorized.
- **By looking at this difference, the variability in general memorization ability is taken away.**

Studying with Music

- In Experiment B, there would be no variability at all in our hypothetical example.
- While there is substantial variability in the number of words memorized between students, there would be no variability in the *difference in the number of words memorized.* All values would be exactly 2.
- Hence we would have extremely strong evidence of a difference in ability to memorize words between the two types of music.

Pairing and Random Assignment

Pairing often increases power, and makes it easier to detect statistical significance.

In our memorizing with or without lyrics example:

- If we see significant improvement in performance, is it attributable to the type of song?
- What about experience? Could that have made the difference?
- What is a better design?
  - Randomly assign each person to which song they hear first: with lyrics first, or without.
  - This cancels out an “experience” effect

Pairing and Observational Studies

You can often do matched pairs in observational studies, when you know the potential confounder ahead of time.

If you are studying whether the portacaval shunt decreases the risk of heart attack, you could match each patient getting the shunt with a patient of similar health not getting the shunt.

If you are studying whether lefthandedness causes death, and you want to account for age in the population, you could match each leftie with a rightie of the same age, and compare their ages at death.