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

Bring a PENCIL or PEN and CALCULATOR and any books or notes you want to the midterm and final.
My office hour today will be 1-2 instead of the usual 12-1.
No lecture or office hour Mon Sep3 Labor Day.
HW3 is due Wed and is 4.CE.10, 5.3.28, 6.1.17, and 6.3.14.
In 5.3.28d, use the theory-based formula. You do not need to use an applet.
The midterm will be on ch1-6.
http://www.stat.ucla.edu/~frederic/13/sum18 .

1. Five number summary and IQR.
2. Comparing two means and bicycling to work example.
3. t-test for comparing to means.
4. Strength of evidence.
5. t versus normal, and when to use what formula.
6. Causation and prediction.
7. CIs and tests.
8. Review list.
9. Example problems.

1. Five number summary, IQR, and geysers.

6.1: Comparing Two Groups: Quantitative Response
6.2: Comparing Two Means: Simulation-Based Approach
6.3: Comparing Two Means: Theory-Based Approach

Exploring Quantitative Data

Section 6.1

Quartiles

- Suppose 25% of the observations lie below a certain value x. Then x is called the ***lower quartile*** (or 25th percentile).
- Similarly, if 25% of the observations are greater than x, then x is called the ***upper quartile*** (or 75th percentile).
- The lower quartile can be calculated by finding the median, and then determining the median of the values below the overall median. Similarly the upper quartile is median$\{x_i : x_i >$ overall median$\}$.

IQR and Five-Number Summary

- The difference between the quartiles is called the ***inter-quartile range*** (IQR), another measure of variability along with standard deviation.
- The ***five-number summary*** for the distribution of a quantitative variable consists of the minimum, lower quartile, median, upper quartile, and maximum.
- Technically the IQR is not the interval (25th percentile, 75th percentile), but the difference 75th percentile – 25th .
- Different software use different conventions, but we will use the convention that, if there is a range of possible quantiles, you take the middle of that range.
- For example, suppose data are 1, 3, 7, 7, 8, 9, 12, 14.
- M = 7.5, 25th percentile = 5, 75th percentile = 10.5. IQR = 5.5.

IQR and Five-Number Summary

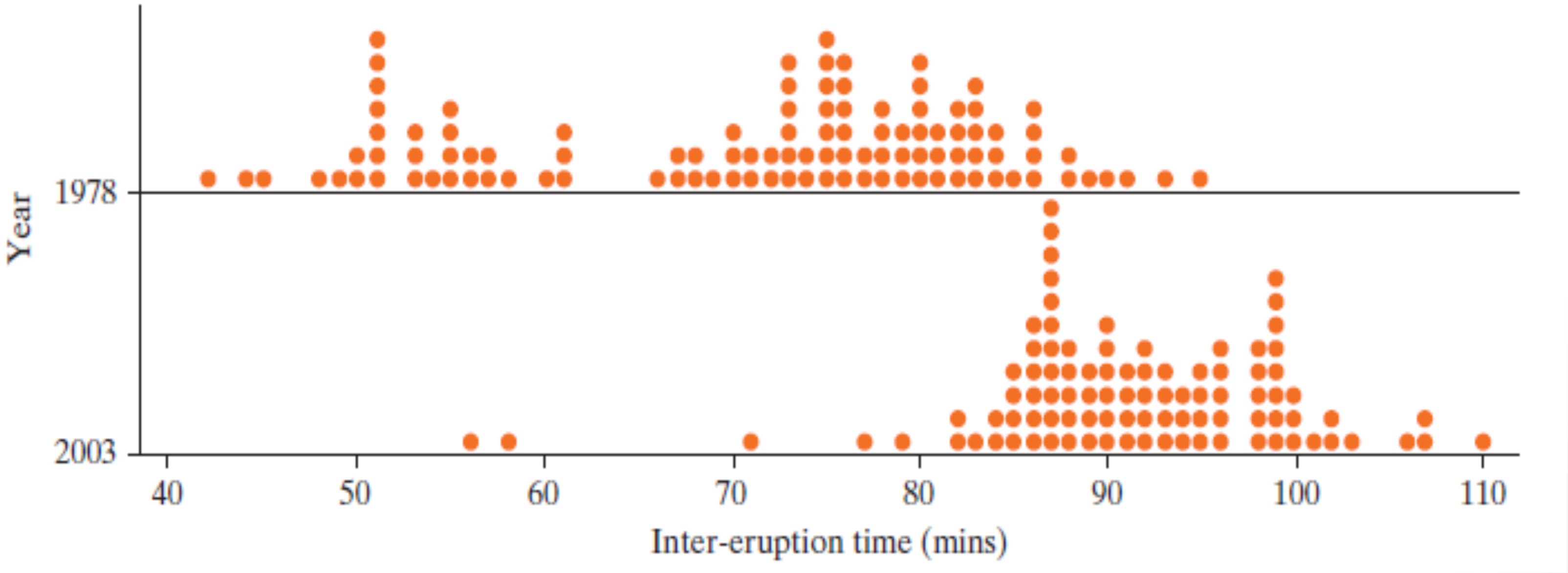
- For medians and quartiles, we will use the convention, if there is a range of possibilities, take the middle of the range.
- In R, this is type = 2. type = 1 means take the minimum.
- x = c(1, 3, 7, 7, 8, 9, 12, 14)
- quantile(x,.25, type=2) ## 5.5
- IQR(x,type=2) ## 5.5
- IQR(x,type=1) ## 6. Can you see why?

Geyser Eruptions

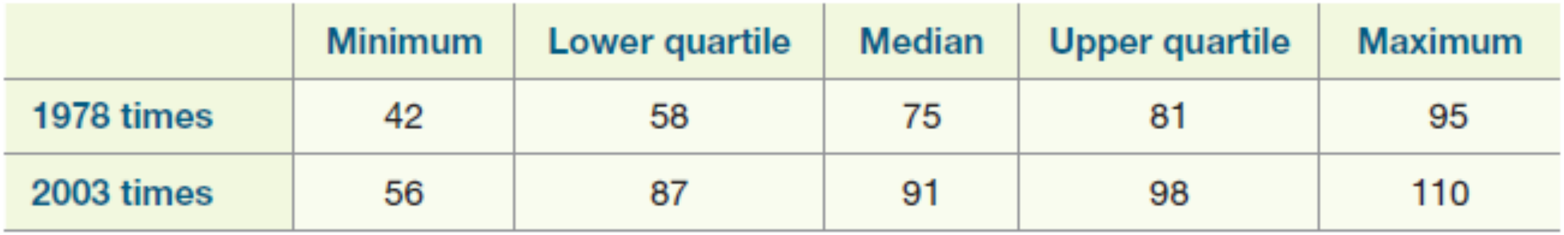
Example 6.1

Old Faithful Inter-Eruption Times

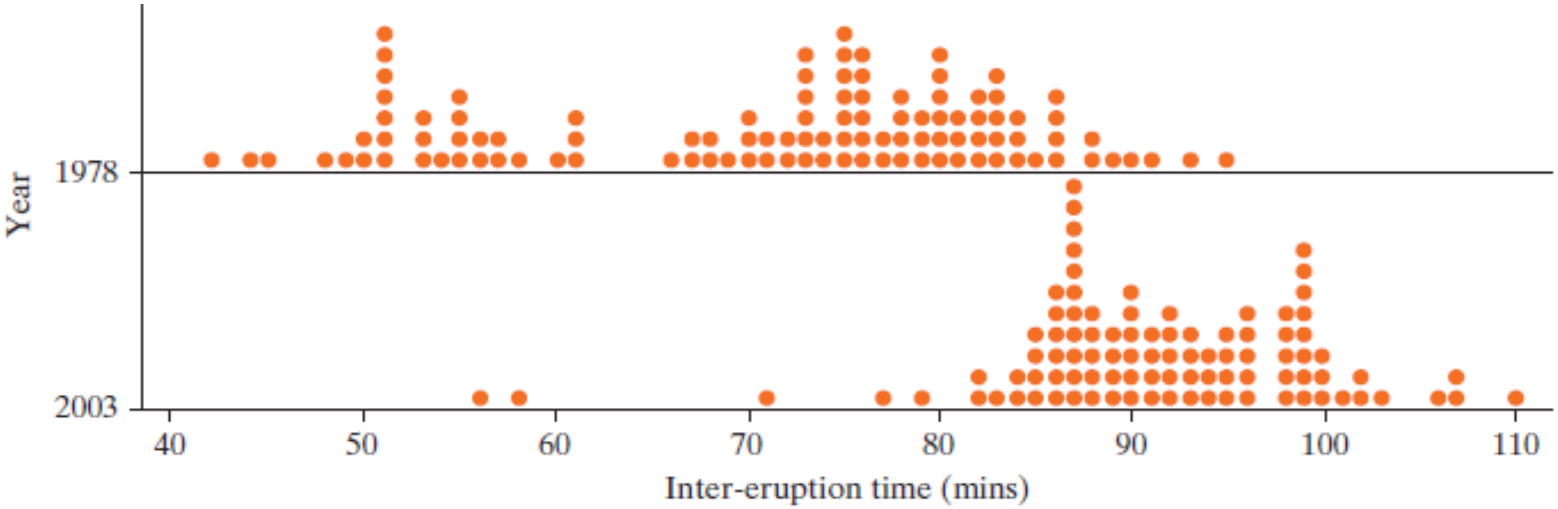
- How do the five-number summary and IQR differ for inter-eruption times between 1978 and 2003?



Old Faithful Inter-Eruption Times

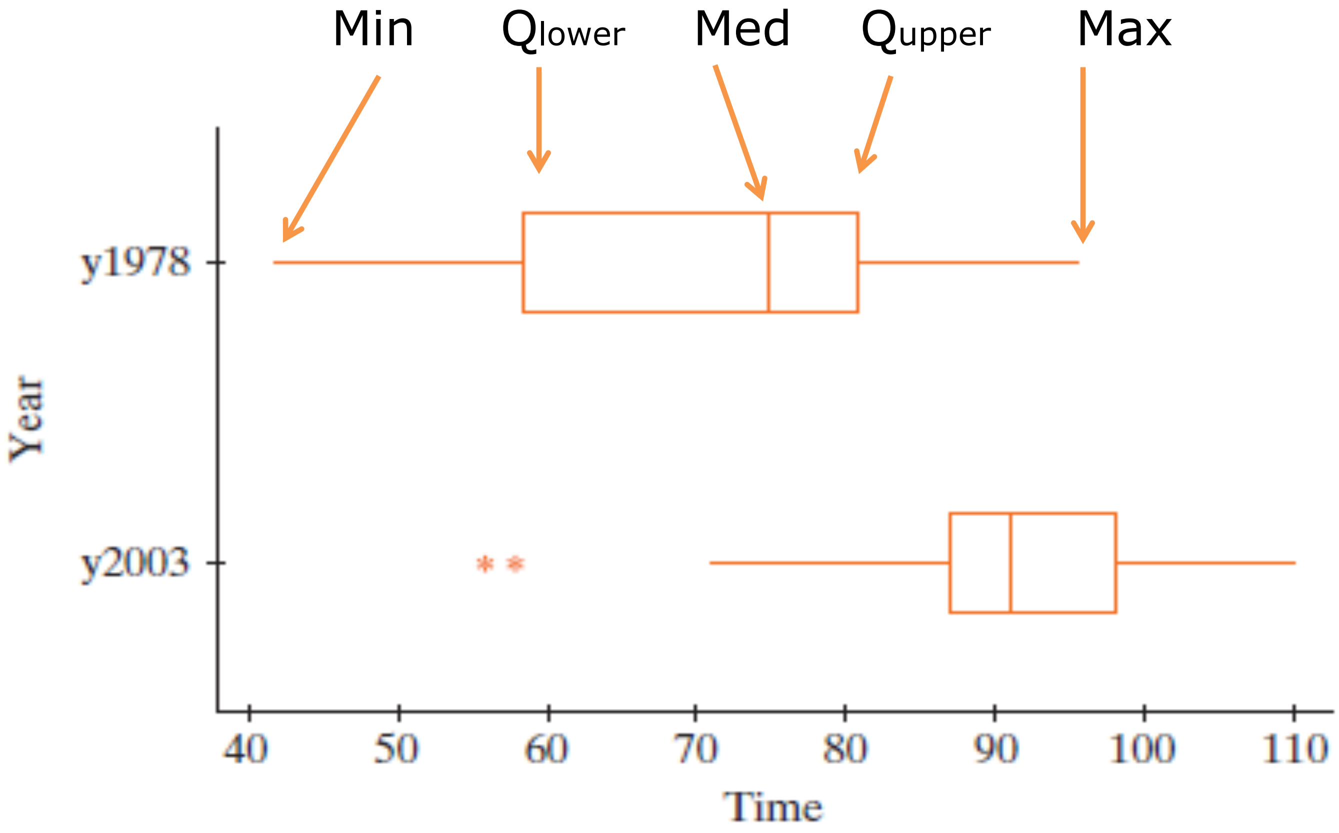


	Minimum	Lower quartile	Median	Upper quartile	Maximum
1978 times	42	58	75	81	95
2003 times	56	87	91	98	110



- 1978 IQR = 81 – 58 = 23
- 2003 IQR = 98 – 87 = 11

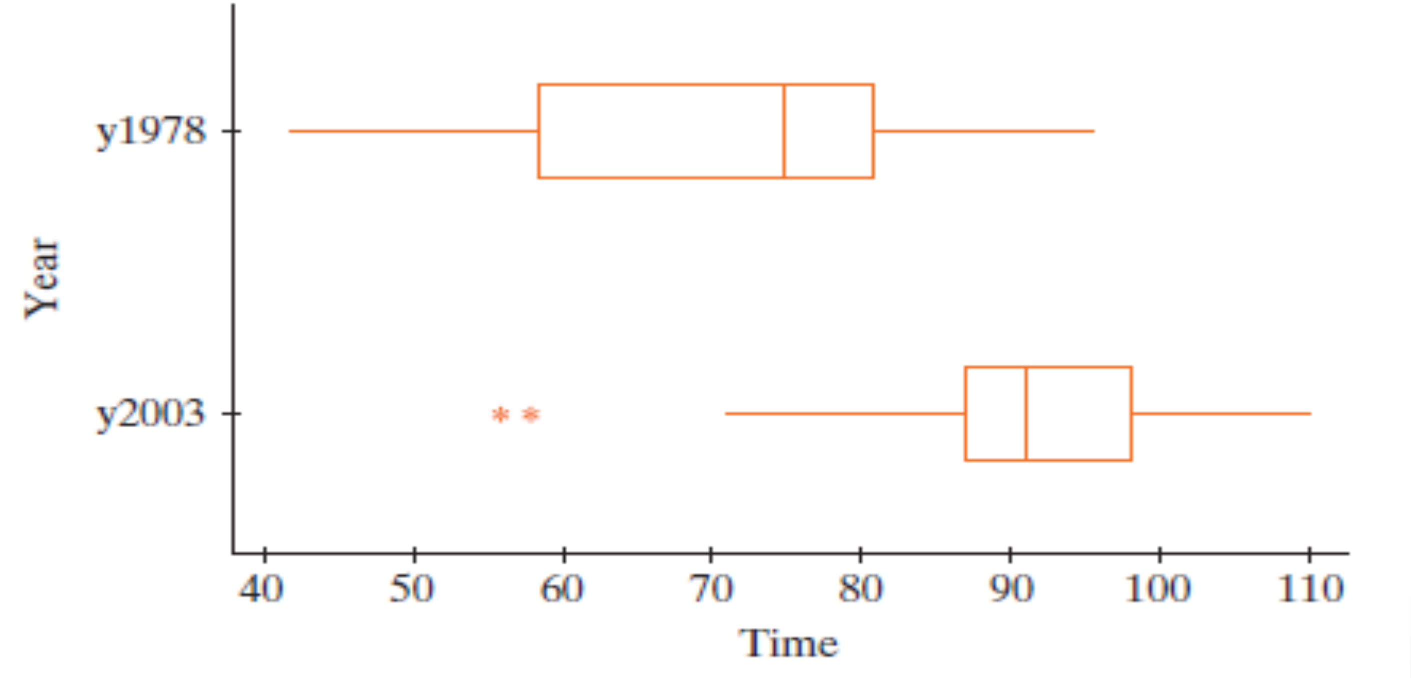
Boxplots



Min Q_{lower} Med Q_{upper} Max

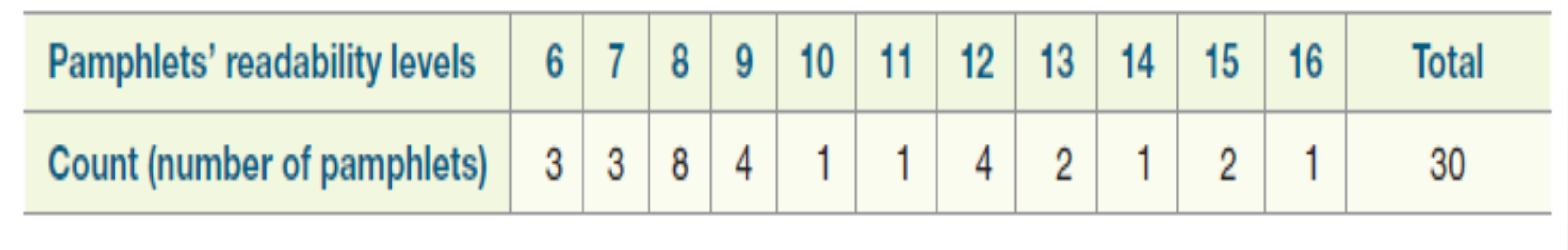
Boxplots (Outliers)

- A data value that is more than 1.5 × IQR above the upper quartile or 1.5 IQR below the lower quartile is considered an outlier.
- When these occur, the whiskers on a boxplot extend out to the farthest value not considered an outlier and outliers are represented by a dot or an asterisk.



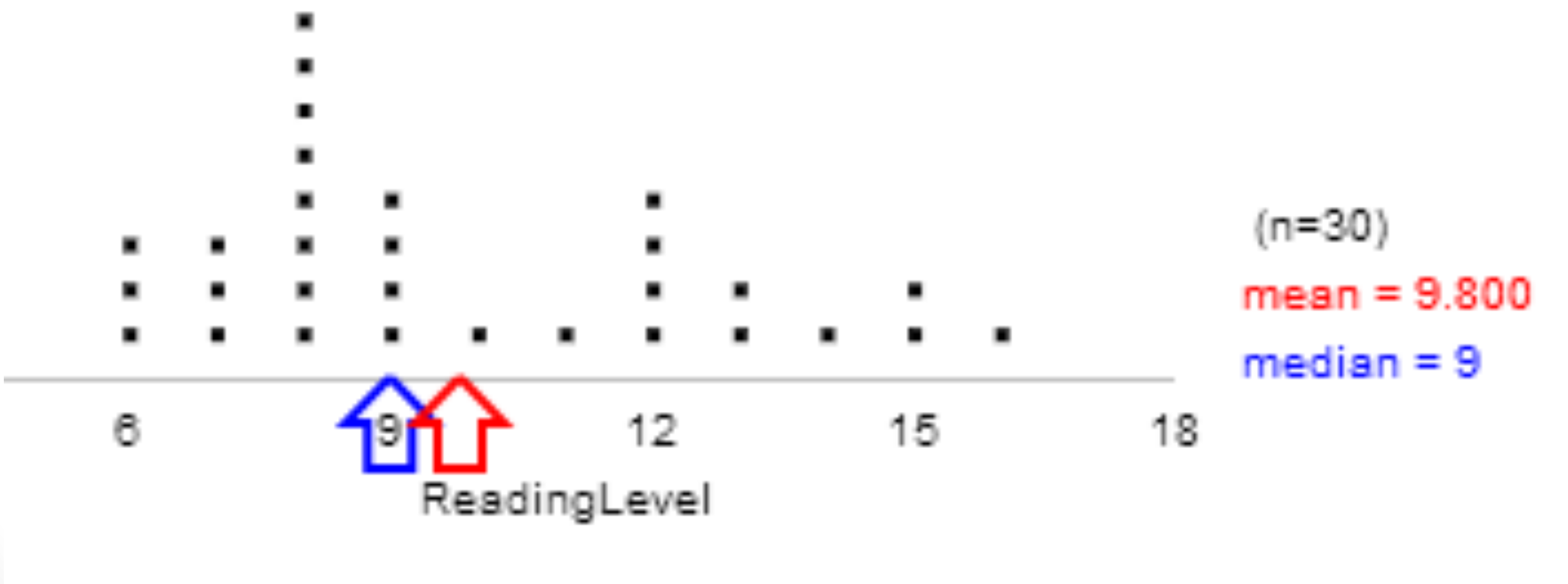
Cancer Pamphlet Reading Levels

- Short et al. (1995) compared reading levels of cancer patients and readability levels of cancer pamphlets. What is the:
  - Median reading level?
  - Mean reading level?
- Are the data skewed one way or the other?



Pamphlets' readability levels	6	7	8	9	10	11	12	13	14	15	16	Total
Count (number of pamphlets)	3	3	8	4	1	1	4	2	1	2	1	30

- Skewed a bit to the right
- Mean to the right of median



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.

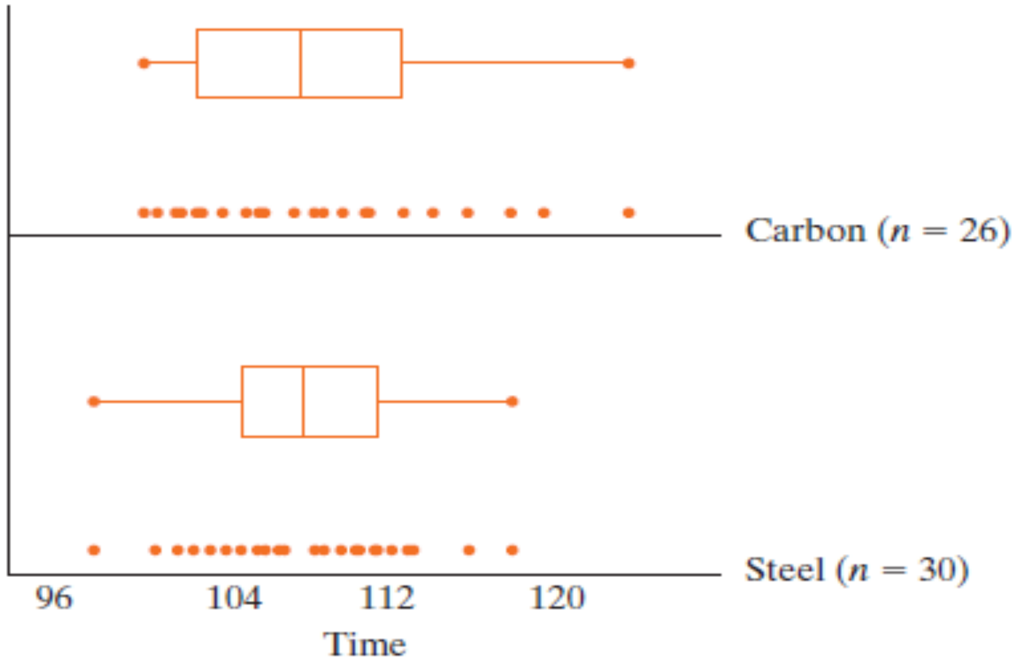
- $\mathbf{H_0}$: $\mu_{carbon} = \mu_{steel}$
- $\mathbf{H_a}$: $\mu_{carbon} \neq \mu_{steel}$.

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 average and variability for commute time was higher for the carbon frame bike
- Does this indicate a tendency?
- 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

- The 3S Strategy

Statistic:

- Choose a statistic:
- The observed difference in average commute times

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

$$= 0.53 \text{ minutes}$$

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
- Let’s see what this looks like

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				

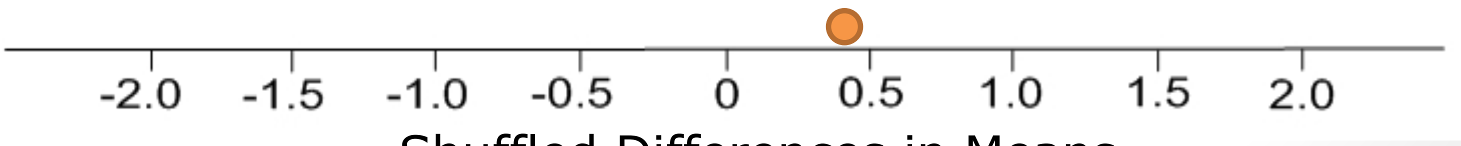
mean = 108.27

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



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				

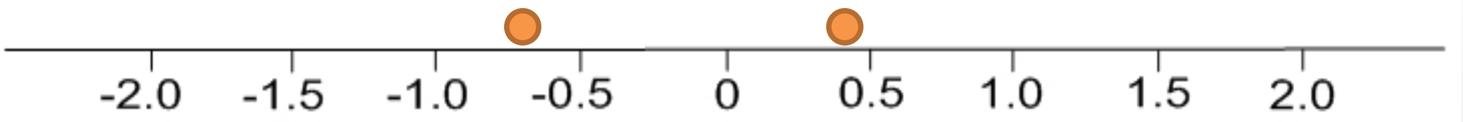
mean = 107.69

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				

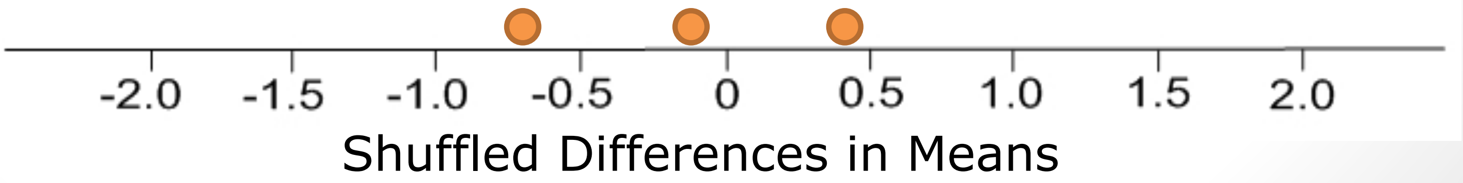
mean = 107.97

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

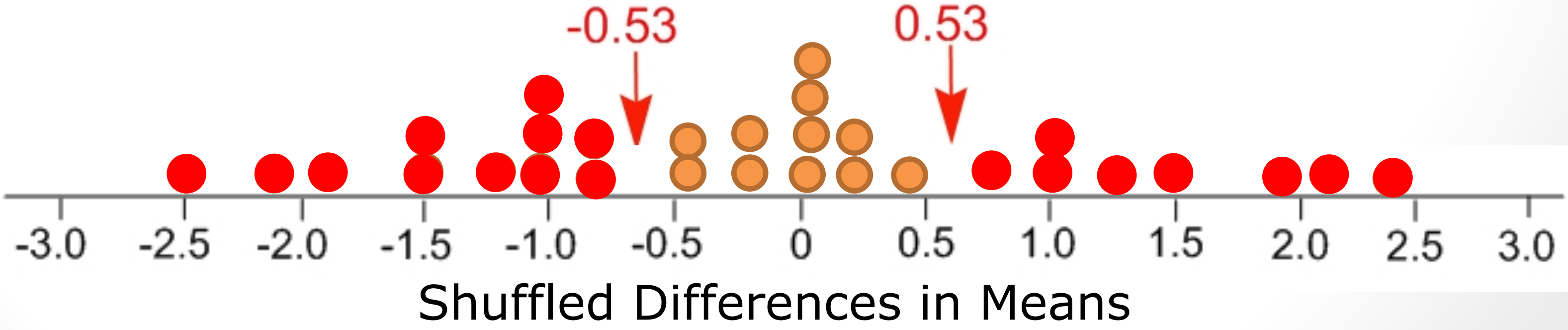
107.97 – 108.13 = -0.16



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 lighter bike to 26 days and the heavier to 30 days, a difference as extreme as 0.53 minutes or more would occur in about 72% of the repetitions.
- Therefore, we do not have strong evidence that the commute times for the two bikes will differ in the long run. The difference observed by Dr. Groves is not statistically significant.

Bicycling to Work

- Have we proven that the bike Groves chooses is not associated with commute time? (Can we conclude the null?)
  - No, a large p-value is not “strong evidence that the null hypothesis is true.”
  - It suggests that the null hypothesis is plausible
  - There could be a small long-term difference. But there also could be no 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 ± 1.96(1.47)= 0.53 ± 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

Scope of conclusions

- Can we generalize our conclusion to a larger population?
- Two key questions.
    - Was the sample randomly obtained and representative of the overall population of interest?
    - Was this an experiment? Were the observational units randomly assigned to treatments?

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.

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.

3. t-test, and breastfeeding and intelligence example.

Example 6.3

Breastfeeding and Intelligence

- A 1999 study in *Pediatrics* examined if children who were breastfed during infancy differed from bottle-fed.
- 323 children recruited at birth in 1980-81 from four Western Michigan hospitals.
- Researchers deemed the participants representative of the community in social class, maternal education, age, marital status, and sex of infant.
- Children were followed-up at age 4 and assessed using the General Cognitive Index (GCI)
  - A measure of the child's intellectual functioning
- Researchers surveyed parents and recorded if the child had been breastfed during infancy.

Breastfeeding and Intelligence

- Explanatory and response variables.
  - **Explanatory variable:** Whether the baby was breastfed. (Categorical)
  - **Response variable:** Baby’s GCI at age 4. (Quantitative)

- Is this an experiment or an observational study?
- Can cause-and-effect conclusions be drawn in this study?

Breastfeeding and Intelligence

- **Null hypothesis:** There is no relationship between breastfeeding during infancy and GCI at age 4.
- **Alternative hypothesis:** There is a relationship between breastfeeding during infancy and GCI at age 4.

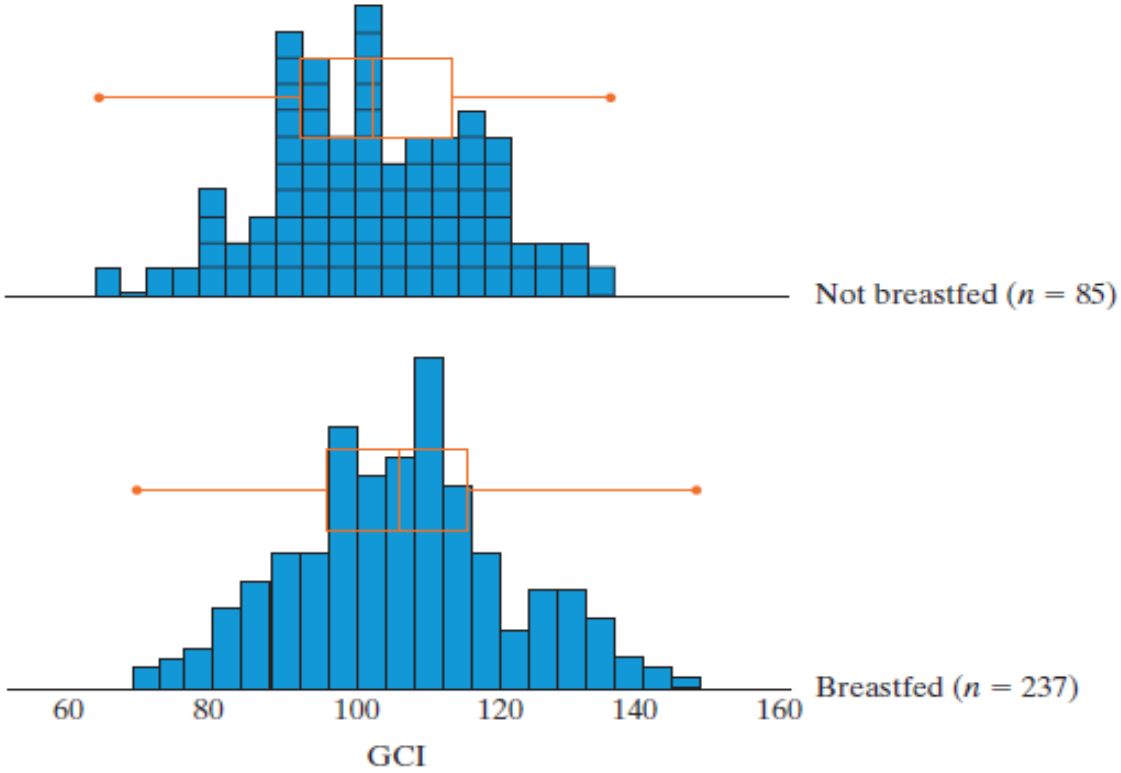
Breastfeeding and Intelligence

  - $\mu_{breastfed}$ = Average GCI at age 4 for breastfed children
  - μ_{not} = Average GCI at age 4 for children not breastfed

- **H_0:** $\mu_{breastfed} = \mu_{not}$
- **H_a:** $\mu_{breastfed} \neq \mu_{not}$

Breastfeeding and Intelligence

Group	Sample size, n	Sample mean	Sample SD
Breastfed	237	105.3	14.5
Not BF	85	100.9	14.0



Breastfeeding and Intelligence

The difference in means was 4.4.

- If breastfeeding is not related to GCI at age 4:
  - Is it **possible** a difference this large could happen by chance alone? Yes
  - Is it **plausible (believable, fairly likely)** a difference this large could happen by chance alone?
    - We can investigate this with simulations.
    - Alternatively, we can use a formula, or what your book calls a theory-based method.

T-statistic

- To use theory-based methods when comparing multiple means, the t-statistic is often used. Here the sample sizes are large, but if they were small and the populations were normal, the t-test would be more appropriate than the z-test.
- the t-statistic is again simply the number of standard errors our statistic is above or below the mean under the null hypothesis.
- $t = \frac{statistic - hypothesized\ value\ under\ Ho}{SE} = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
- Here, t = $\frac{(105.3 - 100.9) - 0}{\sqrt{(\frac{14.5^2}{237} + \frac{14.0^2}{85})}} = 2.46.$
- p-value ~ 1.4 or 1.5%. [2 * (1-pnorm(2.46))], or use pt.

Breastfeeding and Intelligence

Meaning of the p-value:

- If breastfeeding were not related to GCI at age 4, then the probability of observing a difference of 4.4 or more or -4.4 or less just by chance is about 1.4%.

- A 95% CI can also be obtained using the t-distribution. The SE is $\sqrt{(\frac{14.5^2}{237} + \frac{14.0^2}{85})}$ = 1.79. So the margin of error is multiplier x SE.

Breastfeeding and Intelligence

- The SE is $\sqrt{(\frac{14.5^2}{237} + \frac{14.0^2}{85})}$ = 1.79. The margin of error is multiplier x SE.
- The multiplier should technically be obtained using the t distribution, but for large sample sizes you get almost the same multiplier with t and normal. Use 1.96 for a 95% CI to get 4.40 +/- 1.96 x 1.79 = 4.40 +/- 3.51 = (0.89, 7.91).
- The book uses 2 instead of 1.96, and the applet uses 1.9756 from the t-distribution. Just use 1.96 for 95% CIs for this class.

Breastfeeding and Intelligence

- We have strong evidence against the null hypothesis and can conclude the association between breastfeeding and intelligence here is statistically significant.
- Breastfed babies have statistically significantly higher average GCI scores at age 4.
- We can see this in both the small p-value (0.015) and the confidence interval that says the mean GCI for breastfed babies is 0.89 to 7.91 points higher than that for non-breastfed babies.

Breastfeeding and Intelligence

- To what larger population(s) would you be comfortable generalizing these results?
  - The participants were all children born in Western Michigan.
  - This limits the population to whom we can generalize these results.

Breastfeeding and Intelligence

- Can you conclude that breastfeeding improves average GCI at age 4?
  - No. The study was not a randomized experiment.
  - We cannot conclude a cause-and-effect relationship.
- There might be alternative explanations for the significant difference in average GCI values.
- What might some confounding factors be?

Breastfeeding and Intelligence

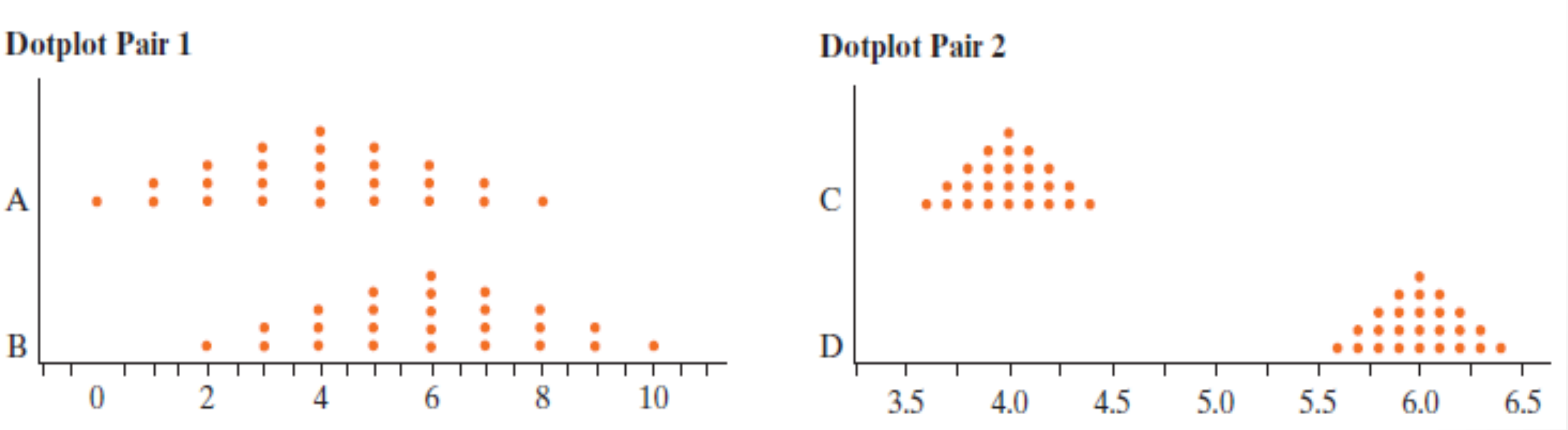
- Can you conclude that breastfeeding improves average GCI at age 4?
  - No. The study was not a randomized experiment.
  - We cannot conclude a cause-and-effect relationship.
- There might be alternative explanations for the significant difference in average GCI values.
  - Maybe better educated mothers are more likely to breastfeed their children
  - Maybe mothers that breastfeed spend more time with their children and interact with them more.
  - Some mothers who do not breastfeed are less healthy or their babies have weaker appetites and this might slow down development in general.

4. Strength of Evidence

- We already know:
    - As sample size increases, the strength of evidence increases.
    - Just as with proportions, as the sample means move farther apart, the strength of evidence increases.

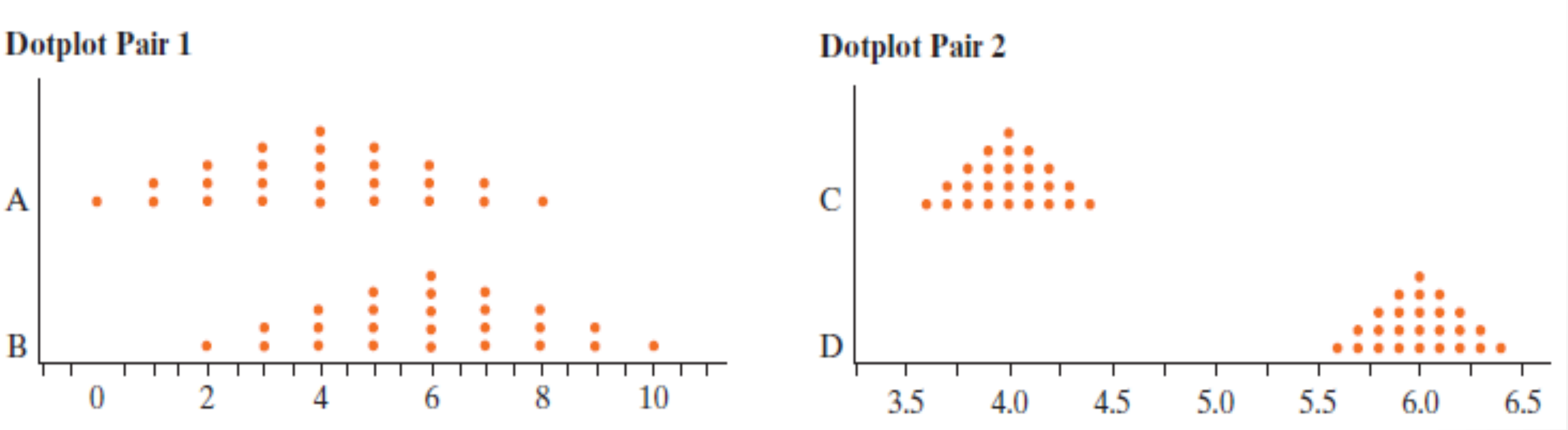
More Strength of Evidence

- If the means are the same distance apart, but the standard deviations change, then the strength of evidence changes too.
- Which gives stronger evidence against the null?



More Strength of Evidence

- If the means are the same distance apart, but the standard deviations change, then the strength of evidence changes too.
- Which gives stronger evidence against the null?



- Smaller SDs lead to stronger evidence against the null.

Effects on Width of Confidence Intervals

- Just as before:
    - As sample size increases, confidence interval widths tend to decrease.
    - As confidence level increases, confidence interval widths increase.
    - The difference in means will not affect the width (margin of error) but will affect the center of the CI.
- As we saw with a single mean, as the SDs of the samples increase, the width of the confidence interval will increase.

5. t versus normal and assumptions.

Why do we sometimes use the t distribution and sometimes the normal distribution in testing and confidence intervals?

The central limit theorem states that, for any iid random variables X_1, ..., X_n with mean μ and SD σ, $(\bar{x}$ - μ) ÷ (σ/√n) -> standard normal, as n -> ∞.

iid means independent and identically distributed, like draws from the same large population.

standard means mean 0 and SD 1.

5. t versus normal and assumptions.

CLT: $(\bar{x}$ - μ) ÷ (σ/√n) -> standard normal.

If Z is std. normal, then P(|Z| < 1.96) = 95%.

So, if n is large, then

P(|$(\bar{x}$ - μ) ÷ (σ/√n)| < 1.96) ~ 95%.

Mult. by (σ/√n) and get

P(|$\bar{x}$ - μ| < 1.96 σ/√n) ~ 95%.

P(μ − $\bar{x}$ is in the range 0 +/- 1.96 σ/√n) ~ 95%.

P(μ is in the range $\bar{x}$ +/- 1.96 σ/√n) ~ 95%.

This all assumes n is large. What if n is small?

5. t versus normal and assumptions.

CLT: $(\bar{x} - \mu) \div (\sigma/\sqrt{n})$ -> standard normal.

What about if n is small?

A property of the normal distribution is that the sum of independent normals is also normal, and from this it follows that if $X_1, \ldots, X_n$ are iid and normal, then $(\bar{x} - \mu) \div (\sigma/\sqrt{n})$ is standard normal.

So again P(μ is in the range $\bar{x}$ +/- 1.96 $\sigma/\sqrt{n}$) = 95%.

This assumes you know σ. What if σ is unknown?

5. t versus normal and assumptions.

Suppose X_1, ..., X_n are iid with mean μ and SD σ.

CLT: ($\bar{x}$ - μ) ÷ (σ/√n) ~ std. normal.

If X_1, ..., X_n are normal, then ($\bar{x}$ - μ)÷(σ/√n) is std. normal.

σ is the SD of the population from which X_1, ..., X_n are drawn. s is the SD of the sample, X_1, ..., X_n .

Gosset (1908) showed that replacing σ with s,

if X_1, ..., X_n are normal, then ($\bar{x}$ - μ)÷(s/√n) is t distributed.

So we need the multiplier from the t distribution.

5. t versus normal and assumptions.

To sum up,

if the observations are iid and n is large, then

P(μ is in the range $\bar{x}$ +/- 1.96 $\sigma/\sqrt{n}$) ~ 95%.

If the observations are iid and normal, then

P(μ is in the range $\bar{x}$ +/- 1.96 $\sigma/\sqrt{n}$) ~ 95%.

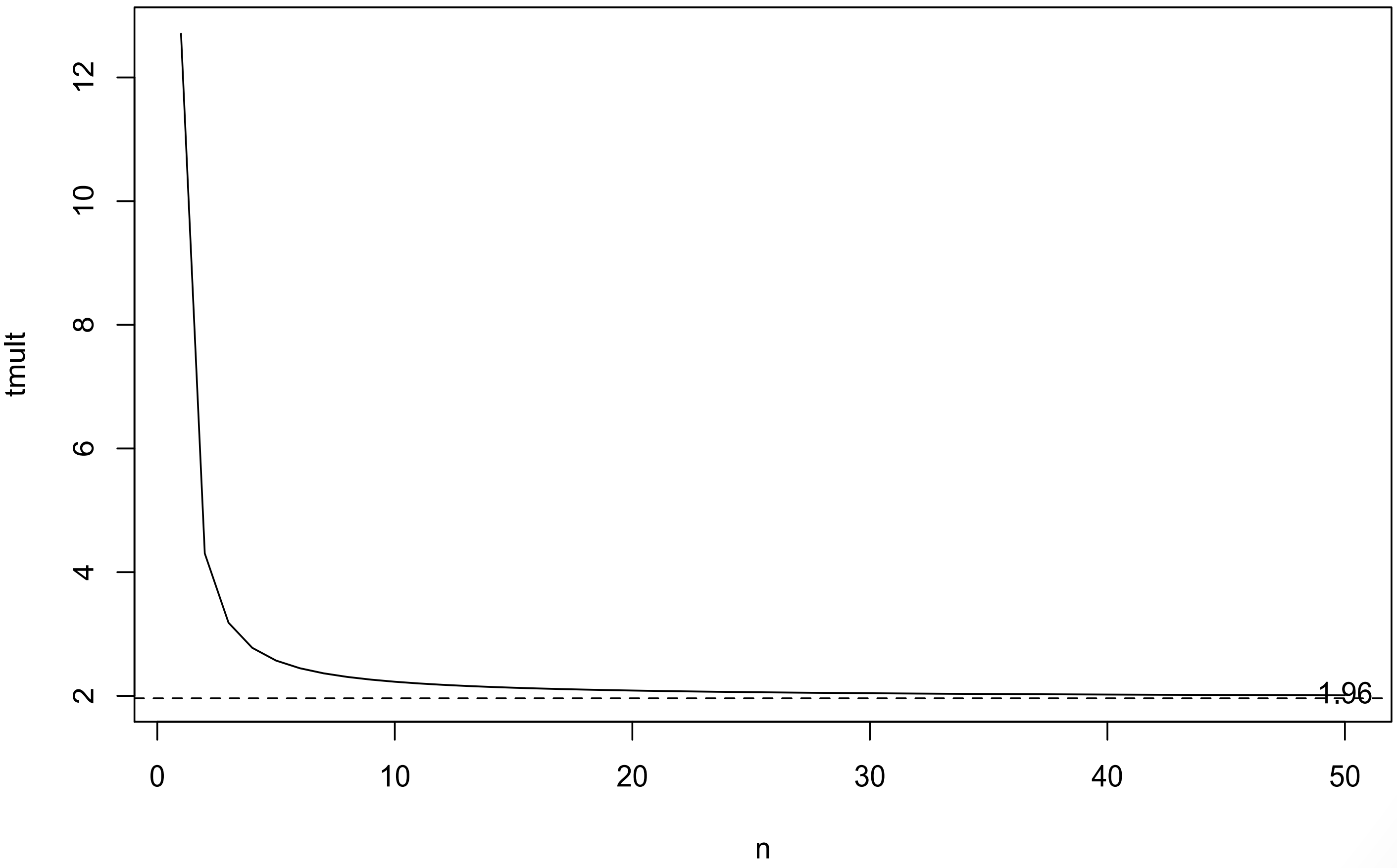
If the obs. are iid and normal and σ is unknown, then

P(μ is in the range $\bar{x}$ +/- t_{mult} $s/\sqrt{n}$) ~ 95%.

where t_{mult} is the multiplier from the t distribution.

This multiplier depends on n.

5. t versus normal and assumptions.



When to use which formula.

a. 1 sample numerical data, iid observations, want a 95% CI for μ.

- If n is large and σ is known, use $\bar{x}$ +/- 1.96 σ/√n.
- If n is small, draws are normal, and σ is known, use $\bar{x}$ +/- 1.96 σ/√n.
- If n is small, draws are normal, and σ is unknown, use $\bar{x}$ +/- t_{mult} s/√n.
- If n is large and σ is unknown, t_{mult} ~ 1.96, so we can use $\bar{x}$ +/- 1.96 s/√n.

n ≥ 30 is often considered large enough to use 1.96.

In practice, we typically do not know the draws are normal, but if the distribution looks roughly symmetrical without enormous outliers, the t formula may be reasonable.

b. 1 sample binary data, iid observations, want a 95% CI for π.

View the data as 0 or 1, so sample percentage p = $\bar{x}$, and

s = √[p(1-p)], σ = √[π(1–π)].

When to use which formula.

a. 1 sample numerical data, iid observations, want a 95% CI for μ.

- If n is large and σ is known, use $\bar{x}$ +/- 1.96 σ/√n.
- If n is small, draws are normal, and σ is known, use $\bar{x}$ +/- 1.96 σ/√n.
- If n is small, draws ~ normal, and σ is unknown, use $\bar{x}$ +/- t_{mult} s/√n.
- If n is large and σ is unknown, t_{mult} ~ 1.96, so we can use $\bar{x}$ +/- 1.96 s/√n.

b. 1 sample binary data, iid observations, want a 95% CI for π.

View the data as 0 or 1, so sample percentage p = $\bar{x}$, and

s = √[p(1-p)], σ = √[π(1−π)].

If n is large and π is unknown, use $\bar{x}$ +/- 1.96 s/√n.

Here large n means ≥ 10 of each type in the sample.

When to use which formula.

What if n is small and the draws are not normal, and you want a theory-based test or CI?

How should you find the t multiplier for a CI or a p-value using the t-statistic, when n is small?

These are questions outside the scope of this course, but some techniques have been developed, such as the bootstrap, which are sometimes useful in these situations.

When to use which formula.

c. Numerical data from 2 samples, iid observations, want a 95% CI for μ_1 - μ_2.

If n is large and σ is unknown, use $\bar{x}_1$ - $\bar{x}_2$+/- 1.96 $\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$.

As with one sample, if σ_1 is known, replace s_1 with σ_1, and the same for σ_2. And as with one sample, if σ_1 and σ_2 are unknown, the sample sizes are small, and the distributions are roughly normal, then use t_{mult} instead of 1.96. If the sample sizes are small, the distributions are normal, and σ_1 and σ_2 are known, then use 1.96.

d. Binary data from 2 samples, iid observations, want a 95% CI for π_1 - π_2.

same as in c above, with $p_1 = \bar{x}_1$, $s_1 = \sqrt{[p_1 (1-p_1)]}$, $\sigma_1 = \sqrt{[\pi_1 (1-\pi_1)]}$.

Large for binary data means sample has ≥ 10 of each type.

6. Causation and prediction.

Note that for prediction, you sometimes do not care about confounding factors.

* Forecasting wildfire activity using temperature.

Warmer weather may directly cause wildfires via increased ease of ignition, or due to confounding with people choosing to go camping in warmer weather. It does not really matter for the purpose of merely *predicting* how many wildfires will occur in the coming month.

* The same goes for predicting lifespan, or liver disease rates, etc., using smoking as a predictor variable.

7. CIs and tests.

Suppose we are comparing blood pressures in a treatment group and a control group. We observe a difference of 10.2 mm, do a 2-sided test, and find a p-value of 3%.

Would the 95%-CI for the difference in blood pressures between the two groups contain zero?

CIs and tests.

Suppose we are comparing blood pressures in a treatment group and a control group. We observe a difference of 10.2 mm, do a 2-sided test, and find a p-value of 3%.

Would the 95%-CI for the difference in blood pressures between the two groups contain zero or not?

No. It would not contain zero.

For what confidence level would the CI just barely contain 0?

97%.

8. Review list.

1. Meaning of SD.
2. Parameters and statistics.
3. Z statistic for proportions.
4. Simulation and meaning of pvalues.
5. SE for proportions.
6. What influences pvalues.
7. CLT and validity conditions for tests.
8. 1-sided and 2-sided tests.
9. Reject the null vs. accept the alternative.
10. Sampling and bias.
11. Significance level.
12. Type I, type II errors, and power.
13. CIs for a proportion.
14. CIs for a mean.
15. Margin of error.
16. Practical significance.
17. Confounding.
18. Observational studies and experiments.
19. Random sampling and random assignment.
20. Two proportion CIs and testing.
21. IQR and 5 number summaries.
22. CIs for 2 means and testing.
23. Placebo effect, adherer bias, and nonresponse bias.
24. Prediction and causation.

9. Example problems.

NCIS was the top-rated tv show in 2014. It was 3rd in 2016 and is now 5th in 2017.

A study finds that in a certain city, people who watch NCIS are much more likely to die than people who do not watch NCIS. Can we conclude that NCIS is a dangerous tv show to watch?

Example problems.

NCIS was the top-rated tv show in 2014. It was 3rd in 2016 and is now 5th in 2017.

A study finds that in a certain city, people who watch NCIS are much more likely to die than people who do not watch NCIS. Can we conclude that NCIS is a dangerous tv show to watch?

No. Age is a confounding factor. The median age of a viewer is 61 years old.

- 1. Suppose the population of American adults has a mean systolic blood pressure of 120 mm Hg and an SD of 20 mm Hg. You take a simple random sample of 100 American adults. Which of the following is true?
- A typical adult's blood pressure would differ from 120 by about **20** mm Hg, and a typical sample of size 100 would have a sample mean that differs from 120 by about **2** mm Hg.
- A typical adult's blood pressure would differ from 120 by about **20** mm Hg, and a typical sample of size 100 would have a sample mean that differs from 120 by about **20** mm Hg.
- A typical adult's blood pressure would differ from 120 by about **2** mm Hg, and a typical sample of size 100 would have a sample mean that differs from 120 by about **0.2** mm Hg.
- A typical adult's blood pressure would differ from 120 by about **20** mm Hg, and a typical sample of size 100 would have a sample mean that differs from 120 by about **0.2** mm Hg.

- 1. Suppose the population of American adults has a mean systolic blood pressure of 120 mm Hg and an SD of 20 mm Hg. You take a simple random sample of 100 American adults. Which of the following is true?
- **A typical adult's blood pressure would differ from 120 by about 20 mm Hg, and a typical sample of size 100 would have a sample mean that differs from 120 by about 2 mm Hg.**
- A typical adult's blood pressure would differ from 120 by about **20** mm Hg, and a typical sample of size 100 would have a sample mean that differs from 120 by about **20** mm Hg.
- A typical adult's blood pressure would differ from 120 by about **2** mm Hg, and a typical sample of size 100 would have a sample mean that differs from 120 by about **0.2** mm Hg.
- A typical adult's blood pressure would differ from 120 by about **20** mm Hg, and a typical sample of size 100 would have a sample mean that differs from 120 by about **0.2** mm Hg.

EXAMPLE PROBLEMS.

- In the portacaval shunt example, why did the studies with historical controls find that the portacaval shunt seemed to be associated with lower death rates?
- a. Those getting the shunt smoked more.
- b. Those getting the shunt were healthier.
- c. Those getting the shunt were genetically predisposed to die younger.
- d. The explanatory variable is a confounding factor t-test with 95% central limit theorem.
- e. None of the above.

EXAMPLE PROBLEMS.

- In the portacaval shunt example, why did the studies with historical controls find that the portacaval shunt seemed to be associated with lower death rates?
- a. Those getting the shunt smoked more.
- **b. Those getting the shunt were healthier.**
- c. Those getting the shunt were genetically predisposed to die younger.
- d. The explanatory variable is a confounding factor t-test with 95% central limit theorem.
- e. None of the above.

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

a. Find a 95%-CI for how much less an average UCLA student's blood glucose level is than an average 2nd grader.

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

a. Find a 95%-CI for how much less an average UCLA student's blood glucose level is than an average 2nd grader.

$2.0 +/- 1.96 \surd(1.5^2/100 + 2.2^2/80) = 2.0 +/- 0.564.$

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

b. Is the difference observed between the mean blood glucose at UCLA and in 2nd grade statistically significant?

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

b. Is the difference observed between the mean blood glucose at UCLA and in 2nd grade statistically significant?

Yes. The 95%-CI does not come close to containing 0.

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

c. Is this an observational study or an experiment?

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

c. Is this an observational study or an experiment?

Observational study.

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

d. Does going to UCLA cause your blood glucose level to drop?

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

d. Does going to UCLA cause your blood glucose level to drop?

No. Age is a confounding factor. Young kids eat more candy.

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

e. The mean blood glucose level of all 43,301 UCLA students is a

parameter random variable t-test

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

e. The mean blood glucose level of all 43,301 UCLA students is a

parameter

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

f. If we took another sample of 100 UCLA students and 80 2nd graders, and used the difference in sample means to estimate the difference in population means, how much would it typically be off by?

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

f. If we took another sample of 100 UCLA students and 80 2nd graders, and used the difference in sample means to estimate the difference in population means, how much would it typically be off by? $SE = \sqrt{(1.5^2/100 + 2.2^2/80)} = .288$ mmol/L

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

g. How much does one UCLA student's blood glucose level typically differ from the mean of UCLA students?

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

g. How much does one UCLA student's blood glucose level typically differ from the mean of UCLA students?

1.5 mmoL/L.

Example problems.

- Researchers take a simple random sample of Californians and a simple random sample of Texans to see who does more exercise. They find that the Californians spend 2.5 hours per week exercising on average and the Texans spend 2.0 hours per week exercising on average. The researchers do a 2-sided test on the difference between the two means and find a p-value of 2.3%. Which of the following would be true of 90% and 95% confidence intervals for the weekly mean exercising time for Californians minus the mean exercising time for Texans?
- a. Both the 90% CI and the 95% CI will contain zero.
- b. Neither the 90% CI nor the 95% CI will contain zero.
- c. The 95% CI will not contain zero, but the 90% CI might contain zero.
- d. The 95% CI will contain zero, but the 90% CI might not contain zero.

Example problems.

- Researchers take a simple random sample of Californians and a simple random sample of Texans to see who does more exercise. They find that the Californians spend 2.5 hours per week exercising on average and the Texans spend 2.0 hours per week exercising on average. The researchers do a 2-sided test on the difference between the two means and find a p-value of 2.3%. Which of the following would be true of 90% and 95% confidence intervals for the weekly mean exercising time for Californians minus the mean exercising time for Texans?
- a. Both the 90% CI and the 95% CI will contain zero.
- **b. Neither the 90% CI nor the 95% CI will contain zero.**
- c. The 95% CI will not contain zero, but the 90% CI might contain zero.
- d. The 95% CI will contain zero, but the 90% CI might not contain zero.