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

1. 5 number summary, IQR, boxplots, and Geysers.
2. Comparing two means with simulations, and the bicycling example.
3. Comparing two means with theory based t test, and breastfeeding and intelligence example.
4. Exams.

Read through ch7.

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

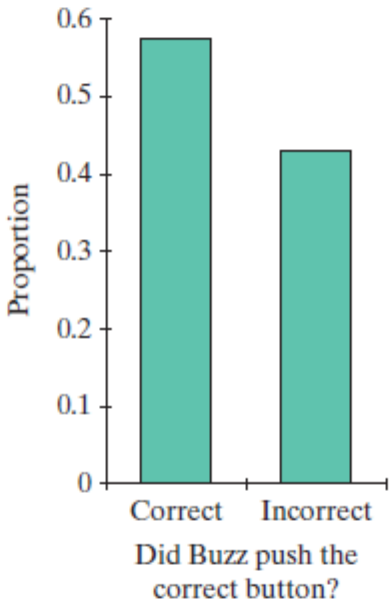
Quantitative vs. Categorical Variables

- Categorical
  - Values for which arithmetic does not make sense.
  - Gender, ethnicity, eye color…

- Quantitative
  - Adding or subtracting values makes sense.
  - Age, height, weight, distance, time…

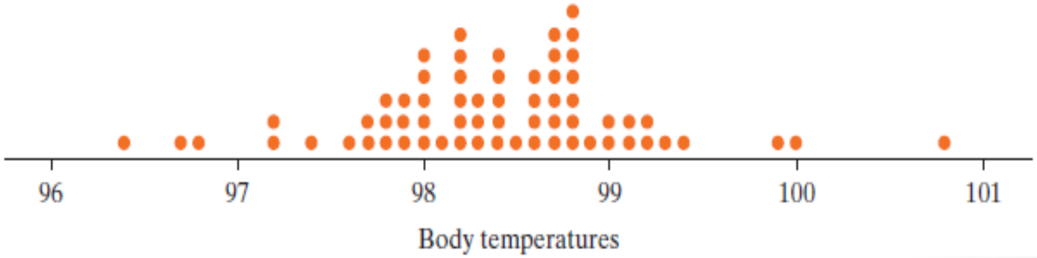
Graphs for a Single Variable

Categorical



Bar Graph

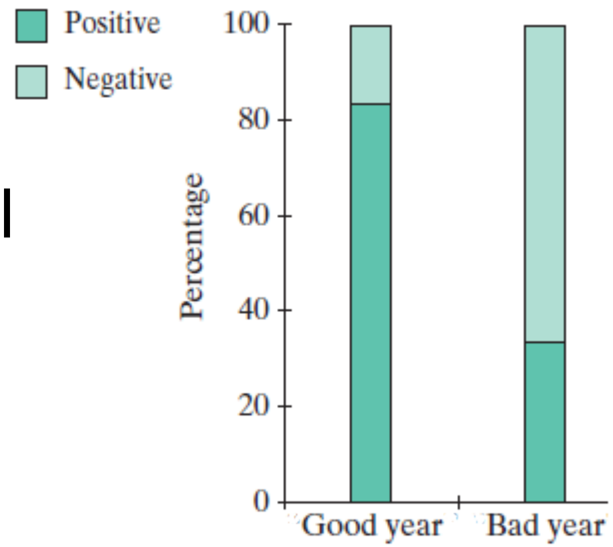
Quantitative



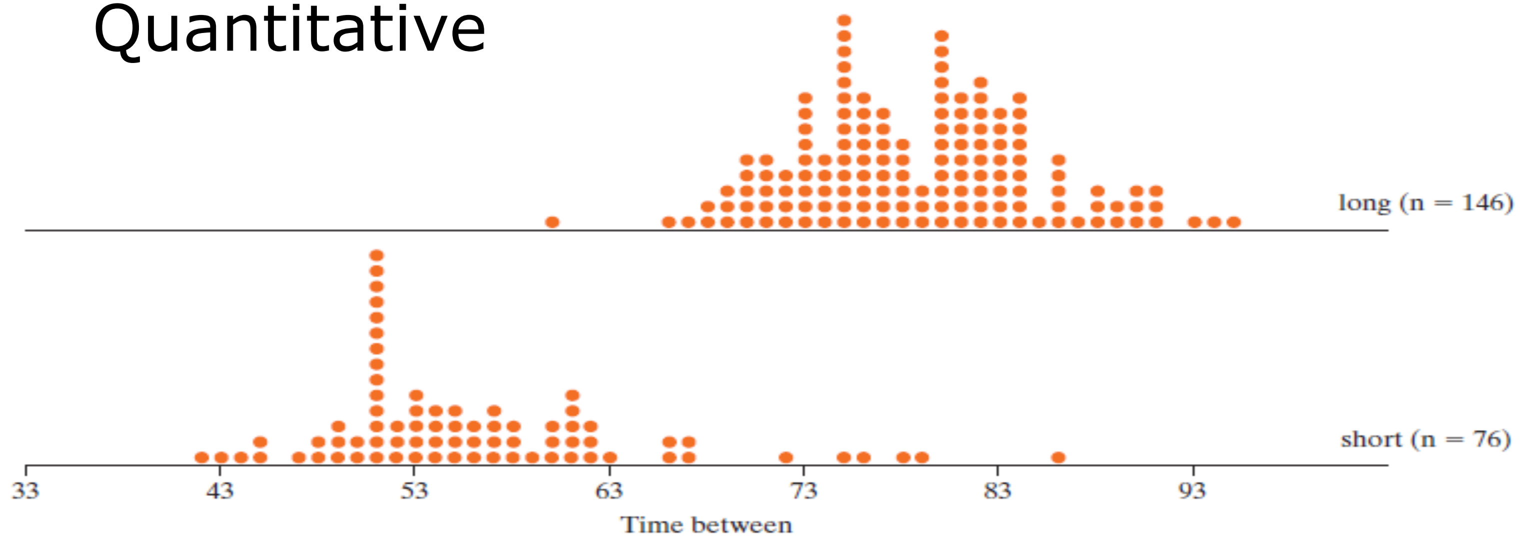
Dot Plot

Comparing Two Groups Graphically

Categorical



Quantitative



Notation Check

Statistics

- $\bar{x}$ Sample mean
- $\hat{p}$ Sample proportion.

Parameters

- μ Population mean
- π Population proportion or probability.

Statistics summarize a sample and parameters summarize a population

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.
- IQR(x,type=2) ## 5.5.
- IQR(x,type=1) ## 6. Can you see why?

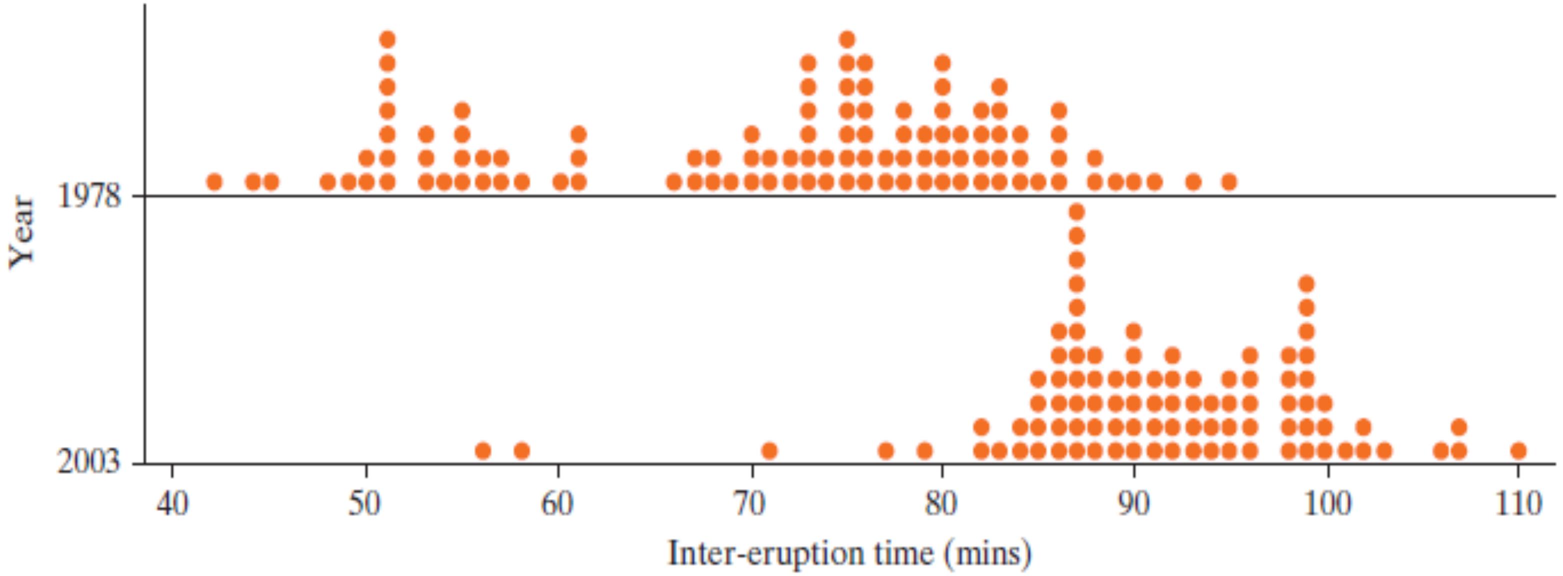
- For example, suppose data are 1, 3, 7, 7, 8, 9, 12, 14.
- M = 7.5, 25^{th} percentile = 5, 75^{th} percentile = 10.5. IQR = 5.5.

Geyser Eruptions

Example 6.1

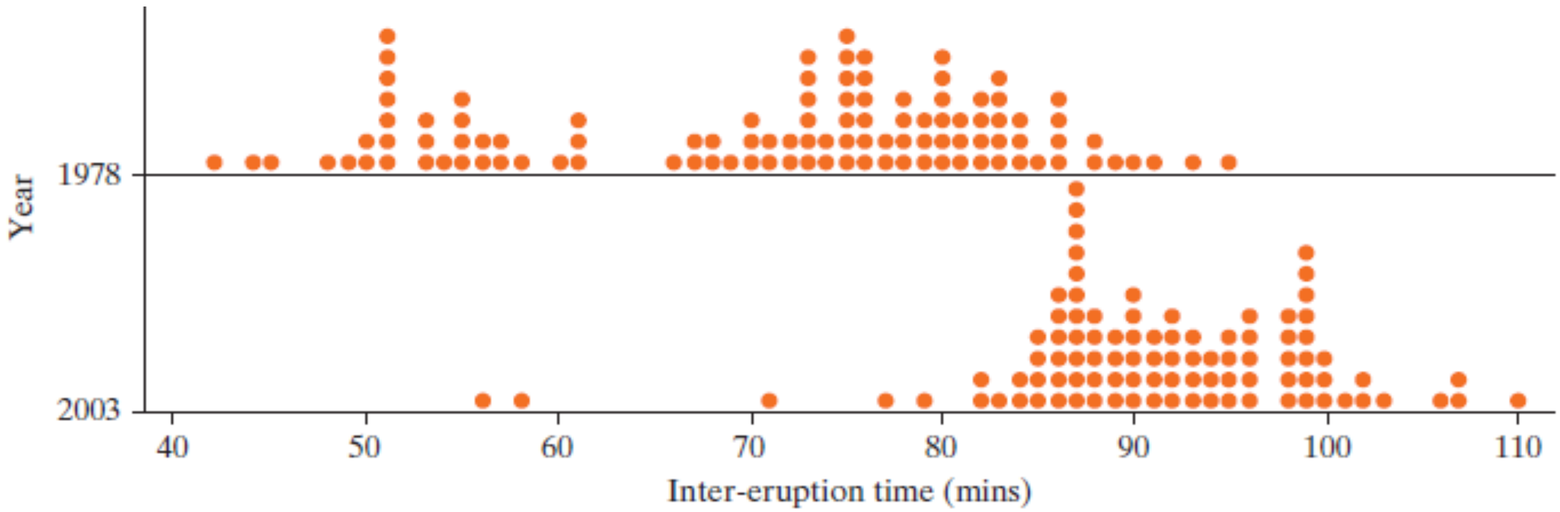
Old Faithful Inter-Eruption Times

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



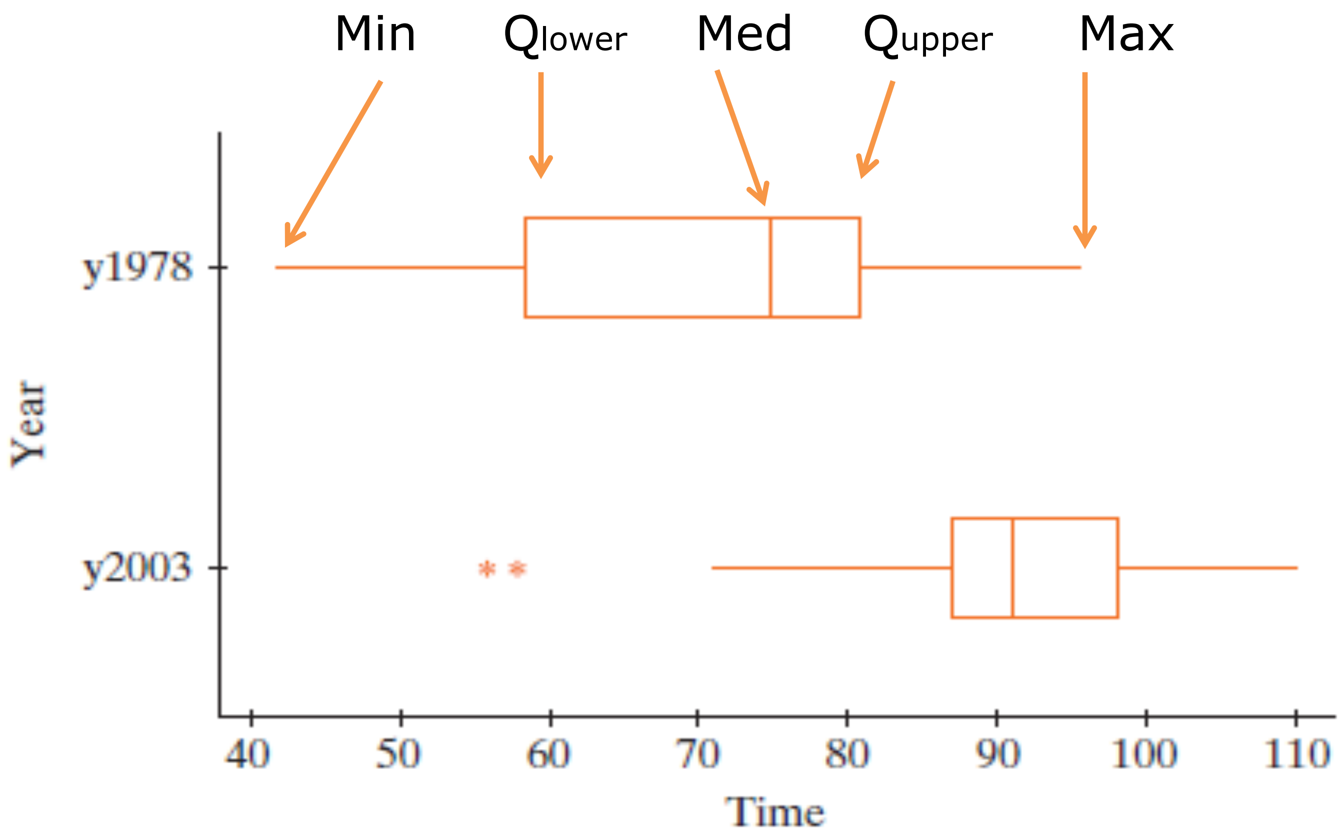
Old Faithful Inter-Eruption Times

	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

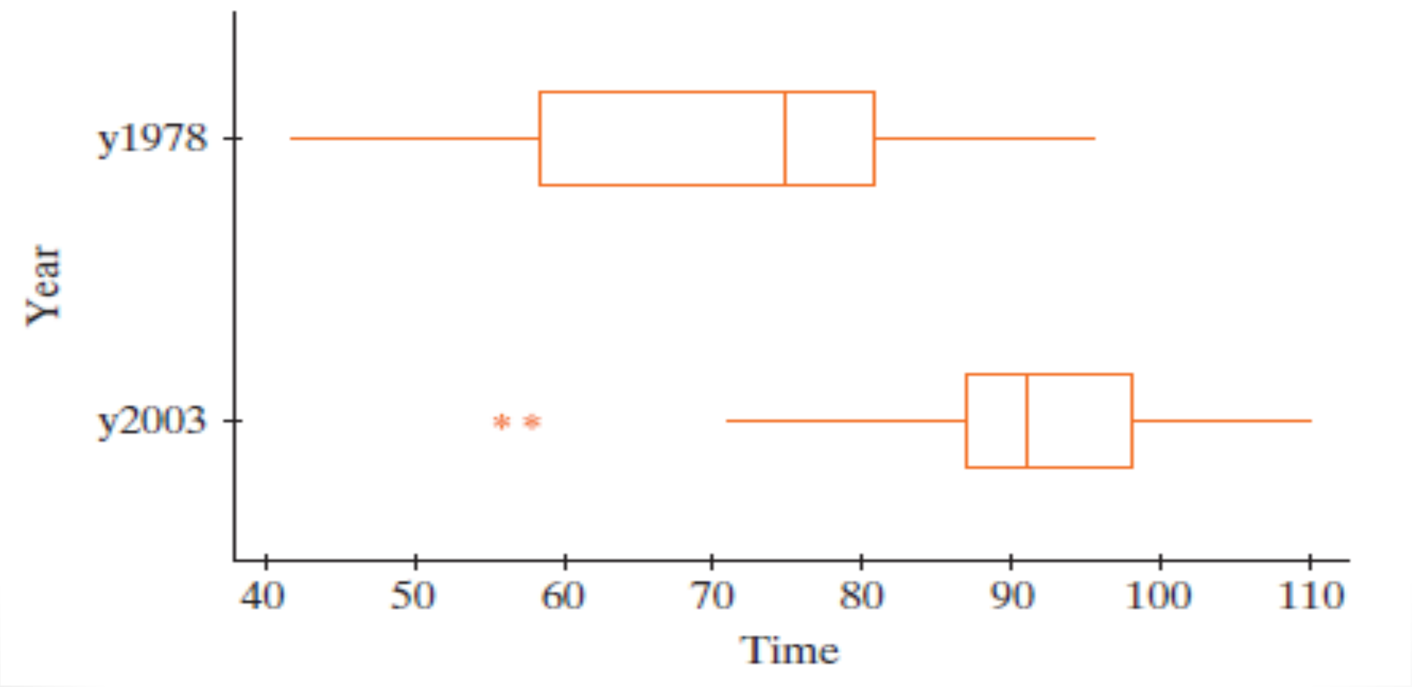
Boxplots



Min
Qlower
Med
Qupper
Max
y1978
y2003
Year
40
50
60
70
80
90
100
110
Time

Boxplots (Outliers)

- A data value that is more than 1.5 × IQR above the upper quartile or 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.

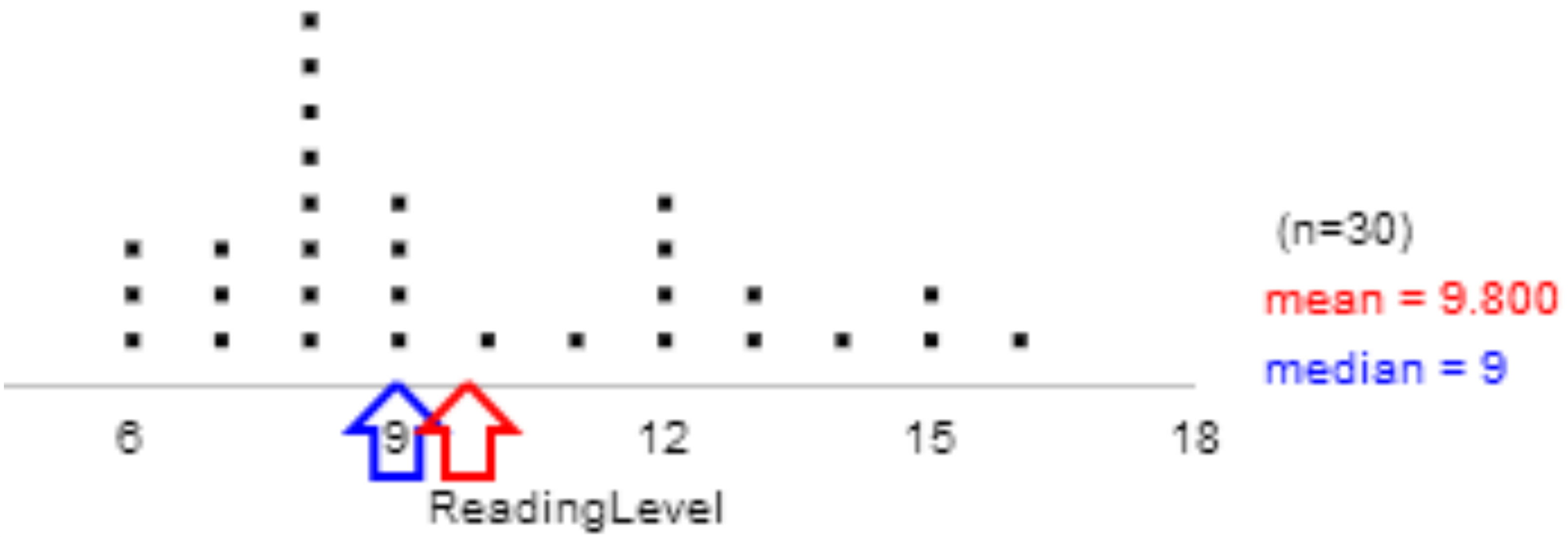


Cancer Pamphlet Reading Levels

- 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 > median



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

Section 6.2

Comparison with 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 in these examples. We will get to quantitative explanatory explanatory variables and response variables when we do correlation and regression later, in ch10.

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.

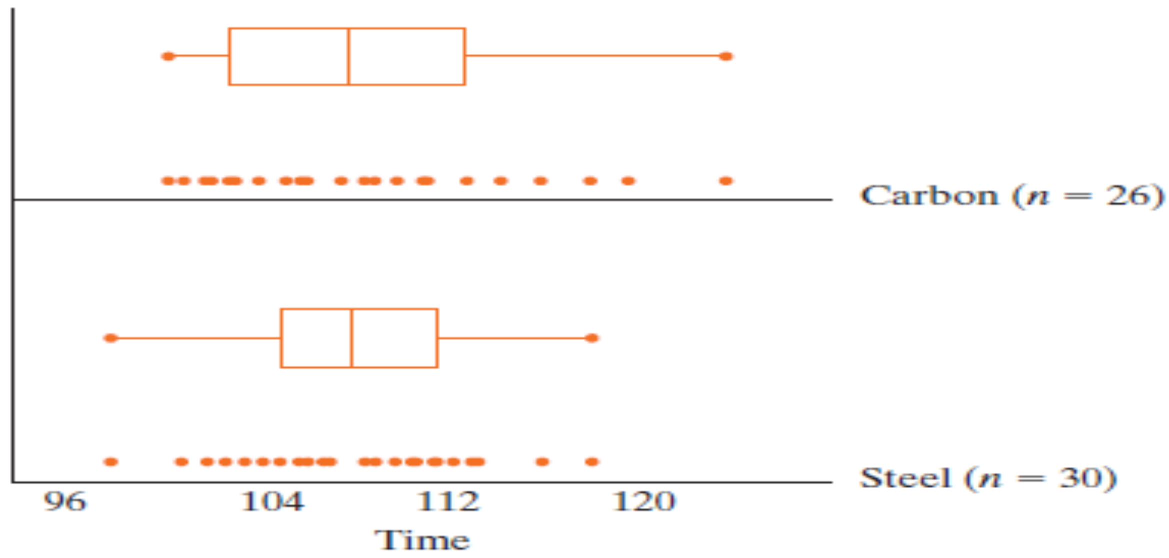
- **H_0:** $\mu_{carbon} = \mu_{steel}$
- **H_a:** $\mu_{carbon} \neq \mu_{steel}$.

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

The observed difference in average commute times

$$\bar{x}_{carbon} - \bar{x}_{steel} = 108.34 - 107.81$$
$$= 0.53 \text{ minutes}$$

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 relationship 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				

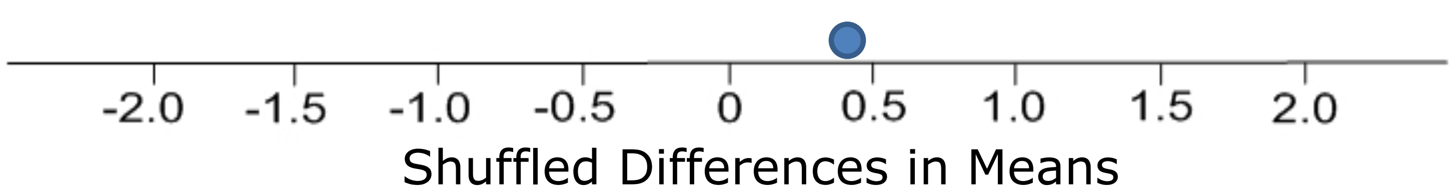
mean = 108.27

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



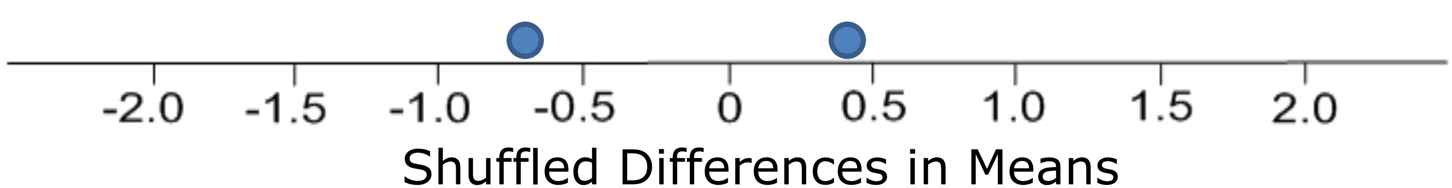
Carbon Frame				
116	114	119	123	113
111	113	106	118	109
103	103	104	112	110
101	102	100	102	107
105	103	111	106	102
108				

mean = 108.29 / 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 / 107.87

107.69 – 108.37 = -0.68



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				

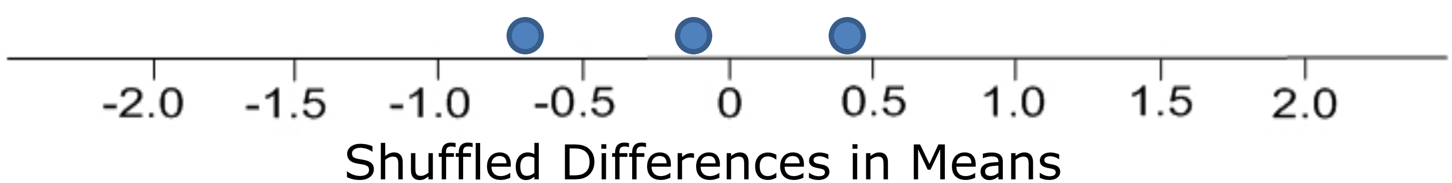
mean = 107.97

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

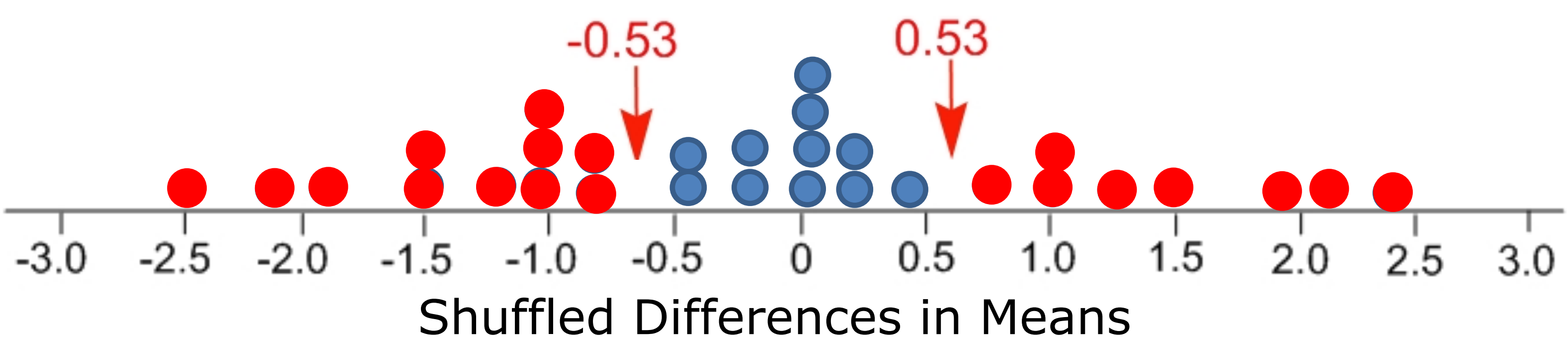
mean = 108.13

107.97 – 108.13 = -0.16



More Simulations

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 is true?
  - No, a large p-value is not "strong evidence that the null hypothesis is true."
  - The null hypothesis is consistent with the data.
  - 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 differences = 1.47. (The theory-based formula is $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ = 1.52.)
- Using 1.47, we get 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. The carbon framed bike seems to be between 2.41 minutes faster and 3.47 minutes slower than the steel framed bike.
- Note the interval contains 0.

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 could 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.

Breastfeeding and Intelligence

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.

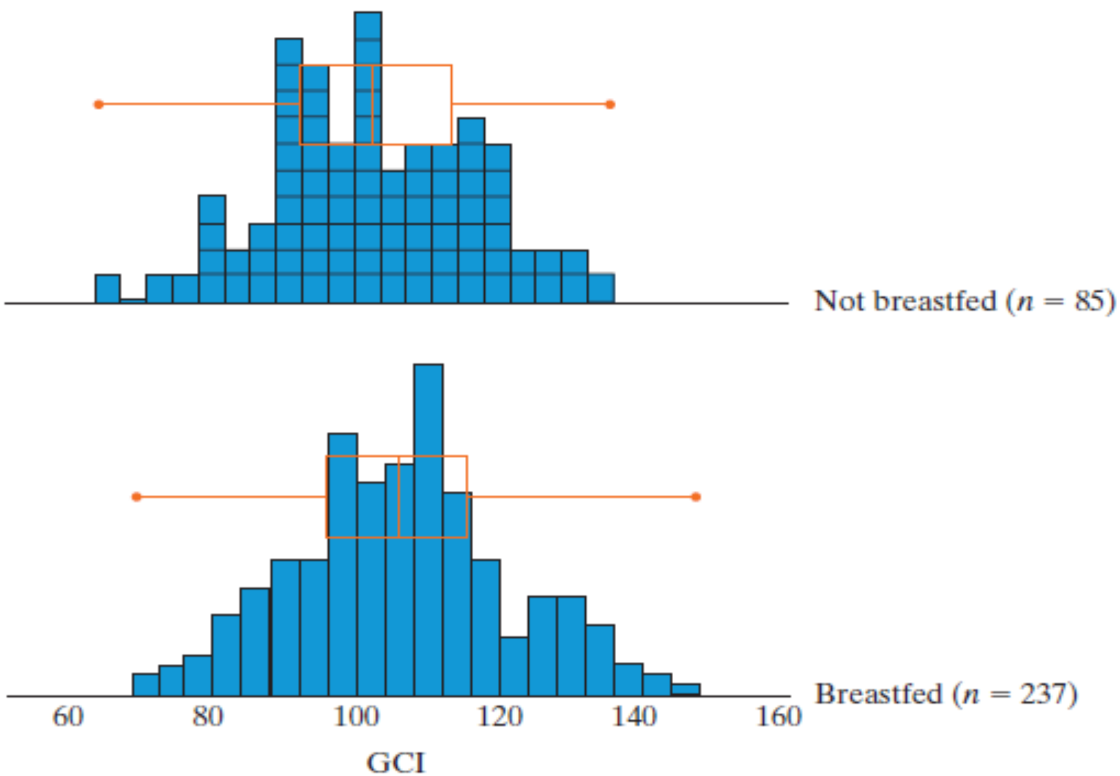
Breastfeeding and Intelligence

- $\mu_{breastfed}$ = 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 theory-based methods.

T-statistic

- If we can assume the draws are iid and the populations are normal, with unknown sds, then t-statistic is used.
- It is the number of standard deviations our statistic is above or below the mean under the null hypothesis.
- $t = \frac{statistic - hypothesized\ value}{SE} = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
- Here, t = $\frac{105.3 - 100.9}{\sqrt{(\frac{14.5^2}{237} + \frac{14.0^2}{85})}} = 2.46$. p-value ~ 1.4%.
- 2*pnorm(2.46,lower=F)] = 1.39%,
  2*pt(2.46,lower=F,df=320) = 1.44%. df = n_1+n_2-2 here.

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 $\sqrt{\left(\frac{14.5^2}{237} + \frac{14.0^2}{85}\right)}$ = 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 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.

4. Exams.

I will hand them out now in alphabetical order by last name and then we are done for the day. After you get your exam you may leave but please remain QUIET!

The mean was 80%, the median was 81%. The sd was 12%.

The grading scale is the usual,

96.7-100 = A+,

93.3-96.7 = A,

90-93.3 = A-,

86.7-90 = B+,

83.3-86.7 = B, etc.

I keep a record of your score, not the letter grade.

I do reward improvement on the final exam. I will not completely ignore your midterm, but I do reward improvement.