

---

## Midterm 1 Review

1. The Johnsons have 5 children. Their ages are: 3, 6, 7, 11, 18.

What is the mean age?

What is the standard deviation of the ages?

2. A class has 10 students. The scores on the latest test, in order were:

88 79 81 81 101 82 83 84 90 79

What is the median score?

What is Q3?

What is the IQR?

3. Consider the following two-way table of 49 men and women cell phone types.

	Male	Female
Iphone	13	10
Other Phone	10	16

What percent of all the people surveyed are men and have iPhones?

What percent of women have iPhones?

What percent of iPhone users are women?

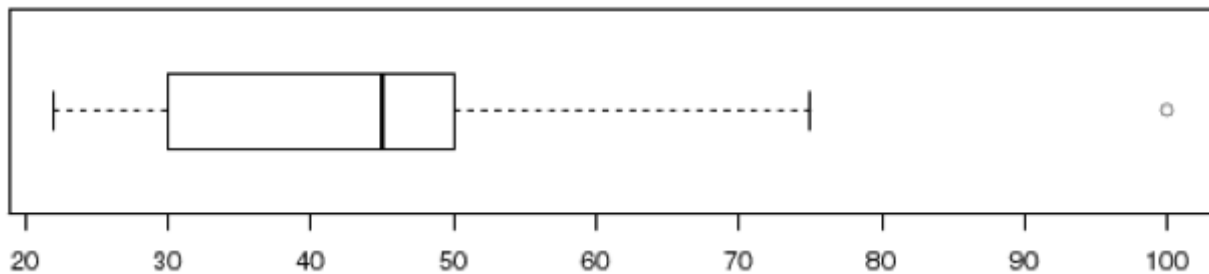
Is there evidence that gender is associated with iPhone usage?

4. Researchers conducted a study to see if there is a relationship between drinking wine and a lower risk of heart disease. They find people willing to participate in the study and monitor the people's drinking habits and their heart health. The researchers find that people who regularly drink wine generally have better heart health.

Was this an experiment or an observational study?

Does this mean that drinking wine is good for your health?

5. Consider the following boxplot:



What is the median?

What is Q1?

What is the IQR?

6. Michael played on his high school's soccer team and basketball team last year. He scored 18 goals throughout the soccer season and made 28 baskets throughout the basketball season. The following summarize the total number of goals and baskets by each of his team members throughout the season.

	Mean	Standard Deviation
Soccer Goals	10	4
Basketball Baskets	25	5

What is Michael's z-score in soccer?

Based on these statistics alone, is Michael a better soccer player or basketball player?

- 
7. A tire manufacturer believes that the treadlife of its snow tires can be described as unimodal, symmetric, and mound-shaped with a mean of 30,000 miles and standard deviation of 3,000 miles. What do we know about the IQR of this distribution?
- (a) The IQR is between 3,000 and 6,000 miles.
  - (b) The IQR is larger than 6,000 miles.
  - (c) It does not make sense to calculate the IQR because the distribution is unimodal and symmetric
  - (d) The IQR is between 27,000 and 33,000 miles.
8. Which plot could be used to examine the association between two categorical variables?
- (a) Segmented Bar Chart
  - (b) Histogram
  - (c) Side-by-Side boxplot
  - (d) Scatterplot
9. The temperature in June in Los Angeles is unimodal, symmetric, and mound-shaped with a mean of 77 degrees and standard deviation of 5 degrees.

What percent of days are colder than 87 degrees?

Between what two temperatures do 68% of the days in Los Angeles fall between?

We recorded a temperature of 59 degrees on June 5th. Is this unusual. Justify your answer?

10. A study of breakfast cereals looked at the potassium content (in milligrams) and fiber content (in grams) of 77 different cereals. A linear model is appropriate and produced the following equation:

$$\widehat{Potassium} = 35 + 25 * Fiber$$

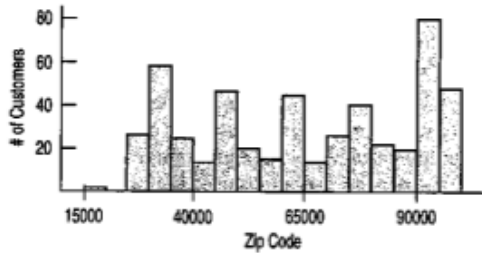
A cereal has 3 grams of fiber in a serving. How much potassium do we predict it to have?

The cereal actually has 99mg of potassium. What is the residual?

The intercept says that a cereal with 0 grams of fiber has how much potassium?

11. Consider the following plot and questions

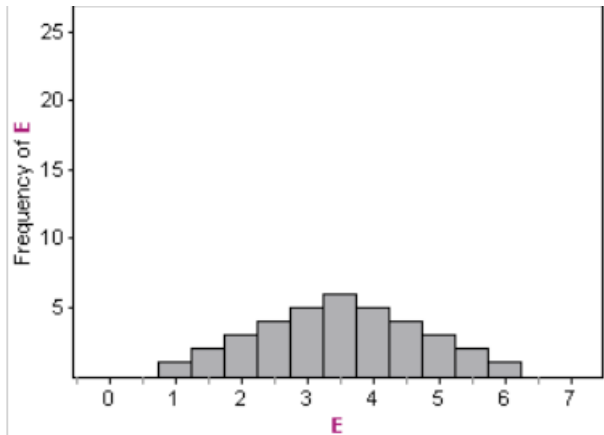
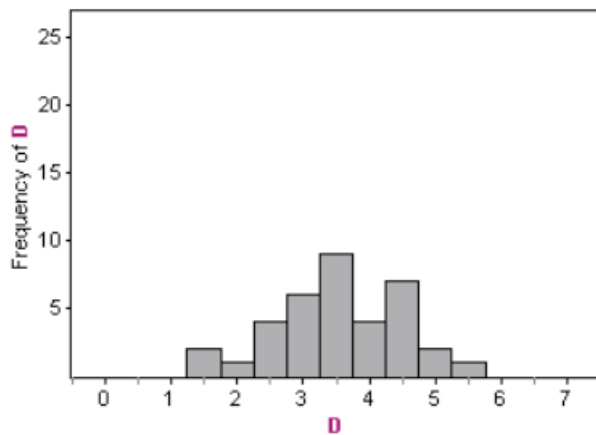
**56. Zip codes revisited.** Here are some summary statistics to go with the histogram of the zip codes of 500 customers from the Holes-R-Us Internet Jewelry Salon that we saw in Exercise 49:



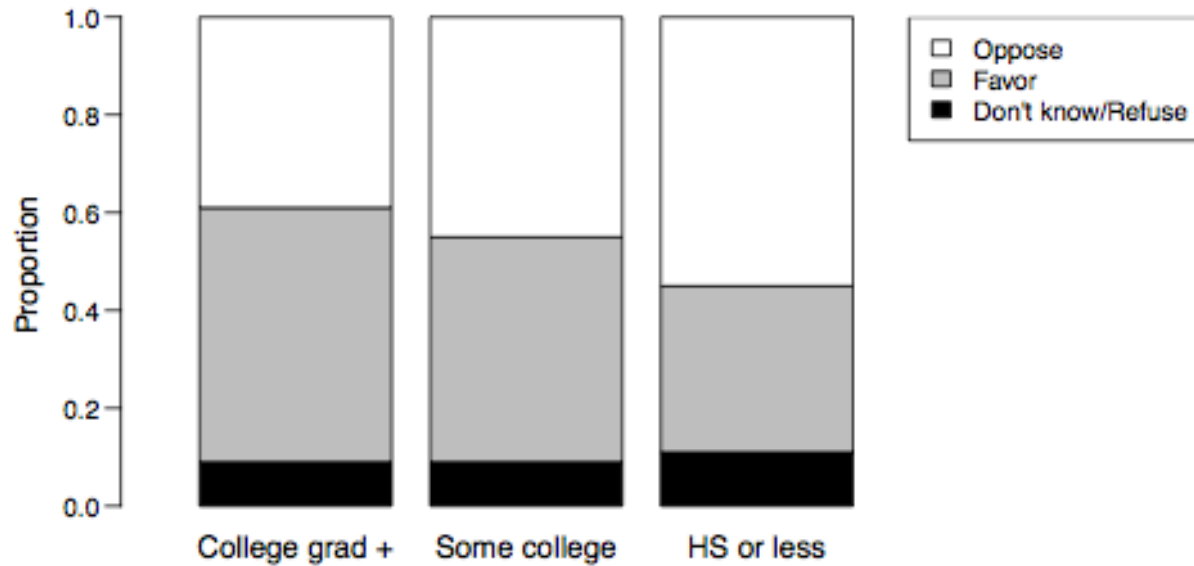
Count	500
Mean	64,970.0
StdDev	23,523.0
Median	64,871
IQR	44,183
Q1	46,050
Q3	90,233

- Is the mean or median a better summary of the center of the zip code distribution? Why?
- Is the standard deviation or the IQR a better summary of the spread? Why?
- What can these statistics tell you about the company's sales?

12. Which has more spread?

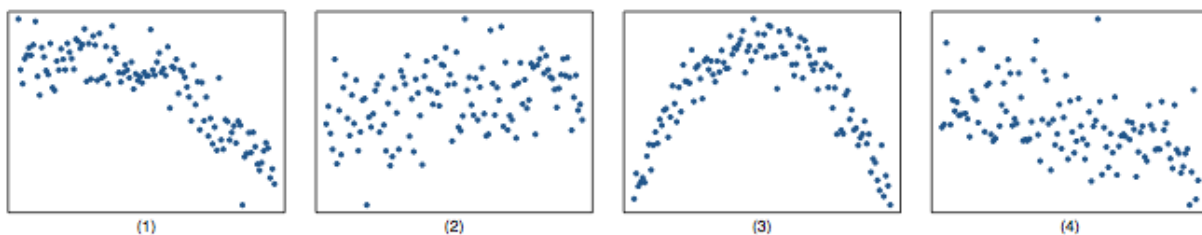


13. A Pew Research poll from August/September 2010 asked respondents “Do you favor or oppose allowing gays and lesbians to marry legally?” The survey also collected information on the education level of these respondents. The segmented relative frequency bar chart below shows the distribution of responses based on education level.

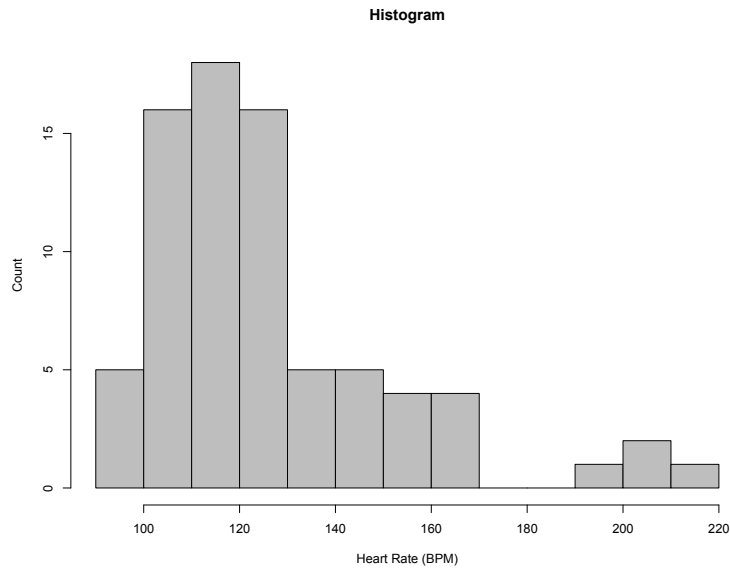


Is there an association between position on gay marriage and education level? Explain.

14. Estimate the correlation coefficient for each of the four scatterplots below.



15. The following histogram is of 77 athletes heart rates following a 5K race. Choose the statement that is the best description of the dataset.



- (a) Data is right skewed and the median and IQR best represent center and spread.
- (b) Data is right skewed and the mean and standard deviation best represent center and spread.
- (c) Data is left skewed and the mean and standard deviation best represent center and spread.
- (d) Data is left skewed and the median and IQR best represent center and spread.
16. Again, consider the histogram of 77 athletes heart rates following a 5K race from above. Estimate the mean. Provide a single number, not a range of values.
17. Are there any outliers?
18. A recent study, reported in the LA Times, found that the density of liquor stores in a neighborhood was strongly and positively correlated with the amount of crime. A scatterplot that plots crime rates against liquor store density shows a fairly linear association. This means that (choose the best answer):
- (a) Neighborhoods with an above average amount of liquor stores tended to have a below average amount of crime.
- (b) Neighborhoods with an above average amount of liquor stores tended to have an above average amount of crime.
- (c) If we could decrease the number of liquor stores in high-crime neighborhoods, the amount of crime would decline.
- (d) There are no conclusions that can be made.

Name:

Langholz Fall 2012

19. Data recorded from 29 rollercoasters about total length in feet and speed in mph was recorded from amusement parks all over the United States. The scatterplot shows a moderately strong, positive linear relationship. The corresponding  $R^2$  is 44.9%. Use the following information from the data recorded to write the regression equation for predicting speed from length.

	Mean	Standard Deviation
Speed	42.53	12.03
Length	2072.35	977.42

Interpret the slope in context

Interpret the intercept in context

Ghost Rider is 4,533 feet long with a maximum speed of 56 mph. Do we have a positive or negative residual?

20. To study the difference exercise makes on the grades of college students, a researcher desires to compare the GPA of students at randomly selected colleges across the United States. The researcher selects students at random and after interviewing them to find out who exercises and who does not, chose 644 students of each (exercisers and non-exercisers). The researcher made sure the two groups were similar in racial composition, gender, major, and ever every student had accumulated at least 120 units towards graduation.

What is the “treatment”?

What is the response variable or outcome of interest?

Is this an observational study or experiment?

21. A recent headline on BBC News, proclaimed, “Happiness wards off heart disease: Being happy and staying positive may help ward off heart disease, a study suggests.” The first paragraph states that researchers in the US monitored the health of 1700 people over 10 years, finding the most anxious and depressed were at the highest risk of heart disease.

What type of study is this?

Name a confounding factor?



22. 1. Match the following variables with the histograms and bar graphs given below. These data represent Stats 10 students at UCLA. [Hint: Think about how each variable should behave.]

- (a) height of students.
- (b) gender breakdown of students
- (c) the time it took students to get to their first class of the day
- (d) the number of hours of sleep students received last night
- (e) whether or not students graduated from high school in California
- (f) the number of piercings students have

