Math 361

Day 2

Traffic Fatalities and Federal Speed Limit Laws – Inv. A

Announcements

- HW 1 and Quiz 1 are due on Wednesday, October 9th by 7pm.
- Complete initial course survey in Canvas by 5pm on Tuesday, October 8th as part of HW 1.
- Buy the workbook by Rossman and Chance following the link on my website.
- Use the RossmanChance glossary or Quizlet to learn statistical definitions for the quiz.

Learning Objectives – Day 2

1. State **the five steps in the statistical process**

2. Identify **the observational units and variable in a given scenario.**

3. Classify **the type of a variable as binary, categorial or quantitative.**

4. Calculate **the mean, median, and standard deviation of a dataset**, by hand or with an applet.

Purple = a statistical term whose definition you should memorize (use the glossary!)

What is “statistics”?

Statistics = using data (pieces of information) to answer a research question

Five steps:

1. Have a research question
2. Make a plan to collect and analyze data
3. Gather data
4. Analyze data
5. Draw an appropriate conclusion

1. State the five steps of the statistical process

Just saying “data” is a bit vague…

Observational Units – the people or objects that information is collected from.

Variable – the piece of information being collected from each observational unit.

When we say “dataset” we’ll usually mean both the ID of the observational units and the values of the variable(s) collected.

Examples

Question: do most statistics students prefer pizza to hamburgers?

Question: What's the average number of miles driven of cars on campus?

3. **Identify the observational units and variable in a given scenario.**

Types of Variables

Variable: prefer pizza over hamburger

Value: yes or no

Variable: miles driven of a car

Value: any positive number

Variable: hometown

Value: any town name

3. Classify the type of a variable as binary, categorial or quantitative/numerical.

Types of Variables

Binary – variable can have only two possible values

Categorical – variable can have any number of values, but the values have no order or distance (i.e. it doesn't matter sense to add two values of the variable)

Quantitative – variable has values that are numbers (i.e. it does make sense to add two values together)

Investigation A, page 4

Today, we'll analyze a single, numerical **variable**, US traffic fatalities per year.

1. Have a research question:

Did a federal speed limit law reduce traffic fatalities?

2. Make a *plan* to collect and analyze data

3. Gather data

Wikipedia: list of motor vehicle deaths in the US by year

4. Analyze the data

   parts b, d, e, g and h

5. Draw an appropriate conclusion

Read the "study conclusions" box on page 9 and the discussion on page 10.

Activity – Inv. A

- Form small groups
- Try at least parts b, d, e, g and h
- Be prepared to discuss your findings with the class in 10-15 minutes.

It's okay if you don't know the "right" answer – remember that learning is more effective if you guess before being told an answer

Tools for describing a single *numerical* variable

Graph

- Dotplot (part h)
- Histogram

Numerical Summaries (box on page 7)

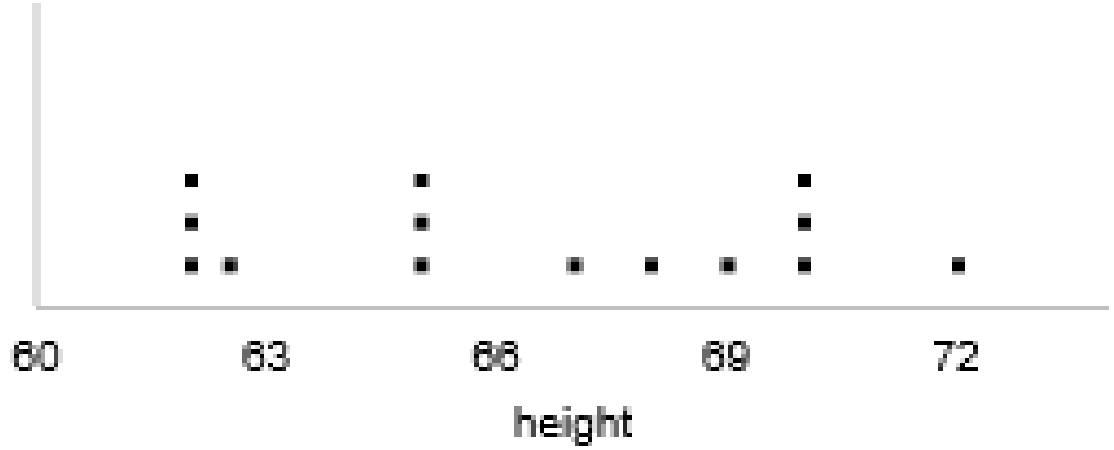
- Mean = average
- Median = middle
- Standard Deviation

mode = most popular

Dotplot – each value in the dataset is represented by a dot above a horizontal axis

Ex: Suppose we observe the heights of 14 OIT statistics students in inches:

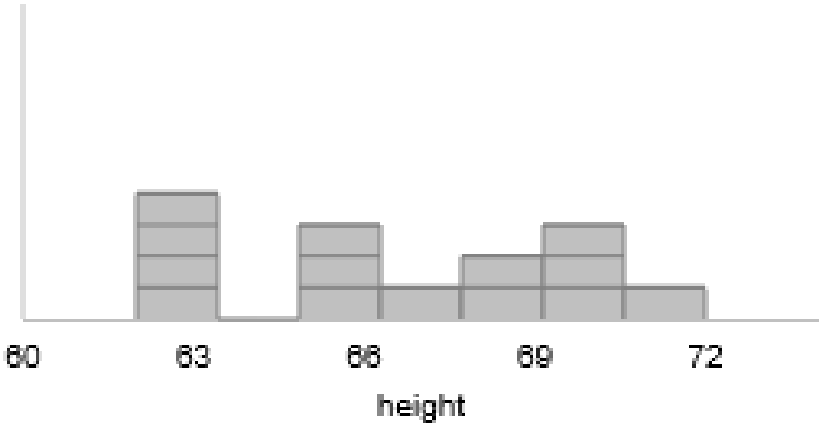
62, 62, 62, 62.5, 65, 65, 65, 67, 68, 69, 70, 70, 70, and 72



Histogram – dataset is “binned” and the height of a “bar” represents # of data points in a bin

Ex: Suppose we observe the heights of 14 OIT statistics students in inches:

62, 62, 62, 62.5, 65, 65, 65, 67, 68, 69, 70, 70, 70, and 72



Numerical Summaries

Mean = average of **n** data points

$$\frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

Interpretation: a "typical" observation

$$\frac{62+62+62+62.5+65+65+65+67+68+69+70+70+70+72}{14} = 66.393$$

2. Calculate the mean, median, and standard deviation of a dataset, by hand or with an applet.

Numerical Summaries

Median = Value so that 50% of the data is above and 50% of the data is below

Interpretation: a "typical" observation

62, 62, 62, 62.5, 65, 65, **65, 67**, 68, 69, 70, 70, 70, and 72

(**65+67**)/2=66

2. Calculate the mean, median, and standard deviation of a dataset, by hand or with an applet.

Numerical Summaries

Standard deviation=

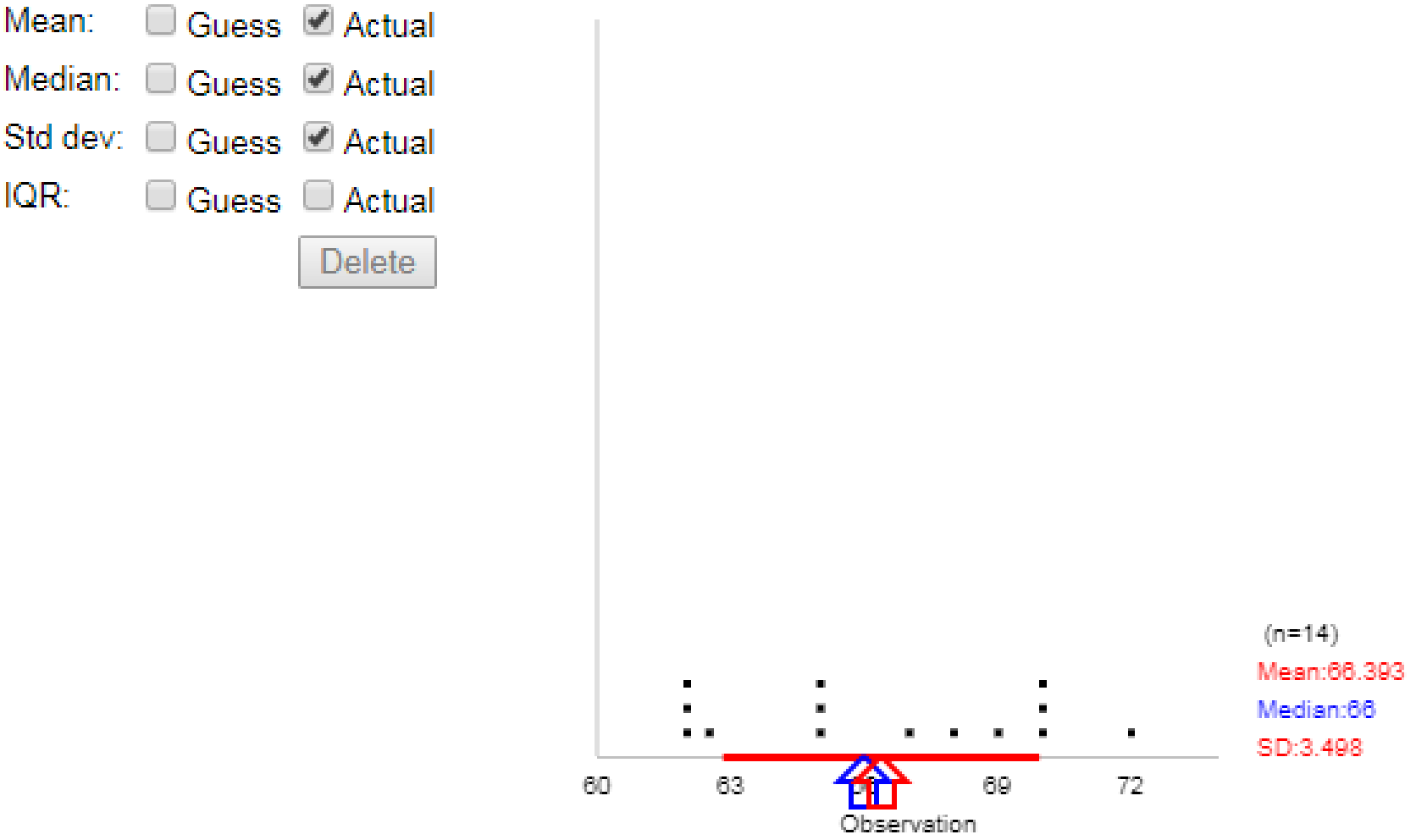
$$\sqrt{\frac{(x_1-mean)^2+(x_2-mean)^2+\cdots+(x_n-mean)^2}{n-1}}$$

Interpretation: a "typical" deviation from the mean

$$\sqrt{\frac{(62-66.393)^2+(62-66.393)^2+\cdots+(72-66.393)^2}{14-1}}=3.498$$

2. Calculate the mean, median, and standard deviation of a dataset, by hand or with an applet.

Using the Descriptive Statistics Applet



2. Calculate the mean, median, and standard deviation of a dataset, by hand or with an applet.

Describing ***center, spread*** and ***shape*** of a distribution

Numerical summaries of ***center***:

Mean (average) or **Median** (middle value) give us an idea of a "typical value" of the variable.

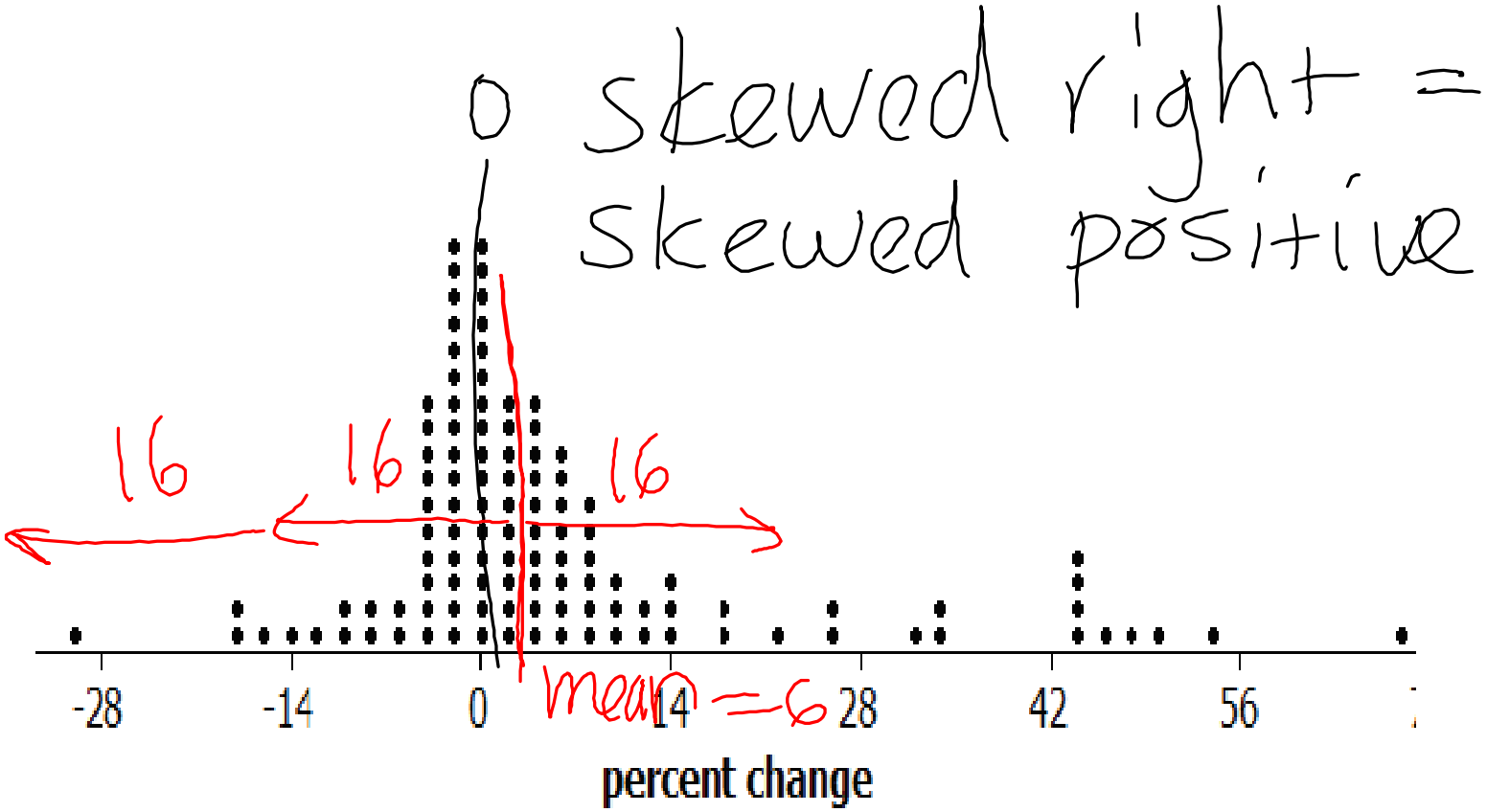
Numerical summary of ***spread***:

Standard deviation gives us an idea of a "typical deviation" from the mean of the variable, i.e. how much *variation* there is.

Words to describe ***Shape*****:**

- Symmetric, skewed right or skewed left
- Bell shaped (one hill), bimodal (two hills) , uniform (rectangle)….

Visualizing Center, Shape and Spread



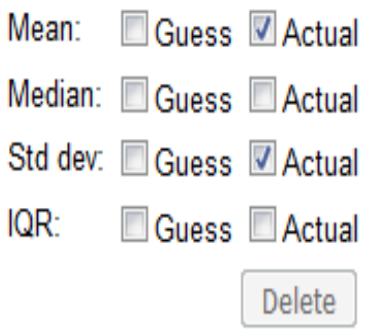
3. Describe the *center*, *shape* and *spread* of a distribution, from a dotplot

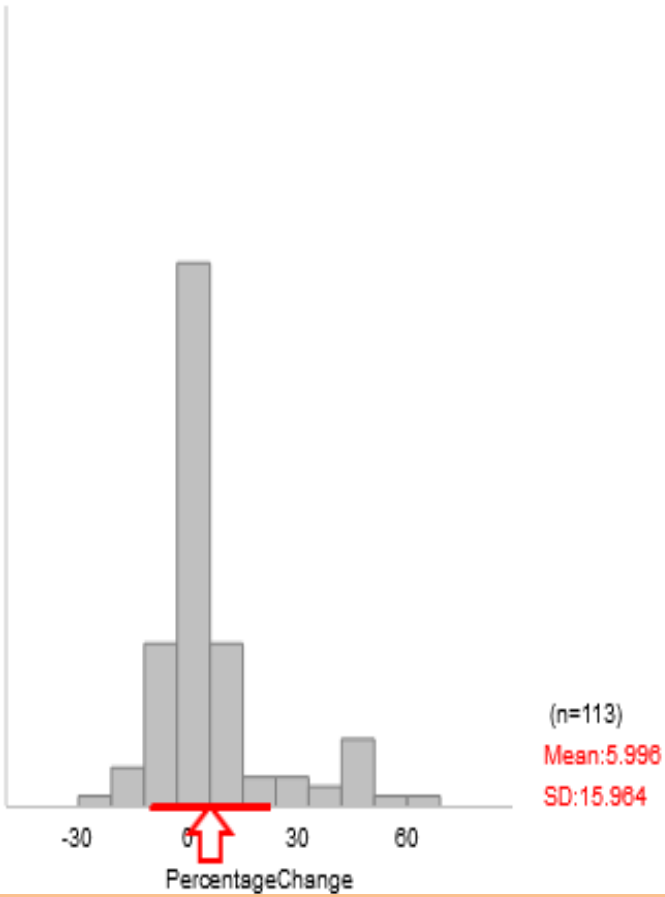
Describing the distribution of percent changes in traffic fatalities

Center: mean is 5.995

Spread: SD=15.964

Shape: One hill, skewed right



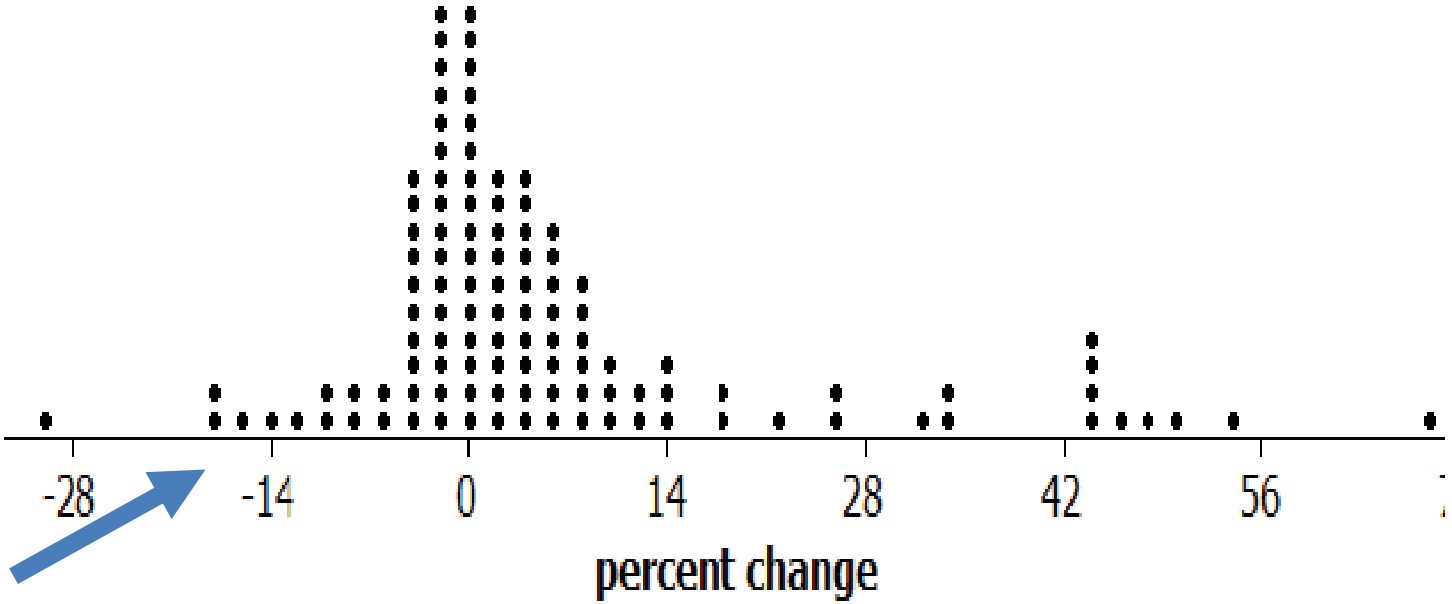


Most percent changes were clustered around 6% with a typical deviation from 6% of about 16%. There are a few extremely large percent changes.

3. Describe **the *center*, *shape* and *spread* of a distribution,** from a dotplot or histogram

Did traffic fatalities decrease after the Federal Speed Limit Law?

- You found the percent change in fatalities dropped by 17.14% after the law was passed.



Is 17.14% a **large** change compared to the changes between other years?