1.1 What is Statistics with answers

Where do we see statistics in the world around us?

- Weather forecasts, e.g., chance of rain
- Insurance, e.g., different insurance premiums for different risks
- Tests for approval of new drugs
- "Big Data" (http://en.wikipedia.org/wiki/Big_data)

Aug 30-10:37 AM

Why Study Statistics?

- Some jobs require statistics, e.g., actuaries in insurance companies.
- To write research papers more easily.
- To understand how data should be gathered, summarized, and analyzed to draw statistical conclusions.
- To describe variables and data.
- To model the uncertain.
- To make decision under uncertainty.
- To follow numerical arguments & understand statistical studies.
- To make accurate inferences and intelligent predictions based on incomplete information.

Take a chance on me!

Aug 30-10:38 AM

Career in Statistics

- Six analytics and data science jobs are included in Glassdoor's 50 best jobs In America for 2018
- Statistician is #4 on U.S. News list of The 25 Best Jobs of 2017 and #1 on both the lists of Best Business Jobs and Best STEM Jobs.
- Assessing Demand for PhD Statisticians and Biostatisticians
- Women flocking to statistics, the newly hot, high-tech field of data science, Washington Post, December 19, 2014
- For Today's Graduate, Just One Word: Statistics, August 5, 2009
- Statistical Analysis and Data Mining top "The 25 Hottest Skills That Got People Hired in 2014"
- Careercast names Statistician one of 2015 top jobs for millennials
- The Importance of Data Occupations in the U.S. Economy, 2015 U.S. Department of Commerce Report
- Statistician Projected As Top 10 Fastest-Growing Job for 2014-2024, New Projections for 2016-2026
- U.S. News Rates Statistician as Best Business Job

Aug 30-10:12 AM

How is statistics different from your other math classes?

Based on inductive rather than deductive reasoning.

- Deductive reasoning happens when a researcher works from the more general information to the more specific; relies on initial premise to be correct. In mathematics, deduction proves, counter-examples disprove
- Inductive reasoning works from specific observations to broader generalizations and theories.

Answers are not made with complete certainty.

Short essays may be required on tests.

Fewer homework problems -- quality, not quantity.

Aug 30-10:27 AM

Always be skeptical of statistical analysis

- Don't blindly accept numbers or findings.
- Think about the numbers and their source.
- Think about the procedures used to generate the data.

Aug 30-10:29 AM

1.1 What is Statistics

Essential Questions:

- What is statistics?
- What are the fundamental logic and tools of statistics?
- How does statistics function in real-world situations?

Focus Points:

- Identify variables in a statistical study.
- Distinguish between quantitative and qualitative variables.
- Identify populations and samples.
- Distinguish between parameters and statistics.
- Determine the level of measurement.
- Compare descriptive and inferential statistics.

Aug 30-10:18 AM

1.1 What is Statistics with answers

Statistics

Statistics is the science of collecting, analyzing and drawing conclusions from data.

Statistics is both the science of uncertainty and the technology of extracting information from data.

Aug 30-10:18 AM

Suppose we want to know the mean length of all fish in Memorial Pond?

Aug 30-10:19 AM

Suppose we want to know the mean length of all fish in Memorial Pond?

Is this a value that can be known?

How would you determine this?

What can we do to estimate this unknown population characteristic?

We can calculate the value from a sample!

How would you take a sample?
What would the sample look like?

- Likely, all different types of fish from Memorial Pond;
- Likely, fish from different locations, different depths, etc.;
- Likely, fish of different ages; different genders.

Aug 30-10:20 AM

Mean Length of fish in Memorial Pond...

Suppose you took a sample of fish from Memorial Pond.

How could the data be organized or summarized?

- Create a graph
- State a range of lengths
- Calculate the average length

Aug 30-10:20 AM

What if our sample comes from the ocean?

Aug 30-10:23 AM

Suppose you took a sample of fish from the ocean and the average length of the fish from the sample was 2 foot.

Could someone claim that average length of the fish in Memorial Pond was 2 foot?

No. Generalizations based on the results of a sample can be made back only to the population from which the sample came.

Aug 30-10:23 AM

1.1 What is Statistics with answers

Suppose 95% of the fish from Memorial Pond are between 3 inches and 2 feet long.
Suppose 95% of the fish from the ocean are between 1 foot and 4 feet long.

- Your friend tells you he caught a three foot fish. What can you say about where the fish came from?
- Your friend tells you he caught a 6 inch fish. What can you say about where the fish came from?
- Your friend tells you he caught an 18 inch fish. What can you say about where the fish came from?

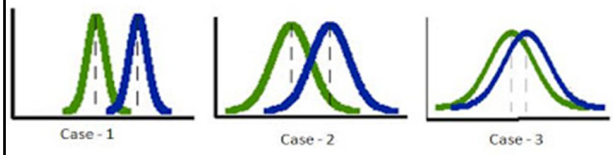
Aug 30-10:25 AM

Distribution

The distribution of data tells us what values the variables take and how often it takes the values.

Hypothesis Testing

Each claim (hypothesis) may have its own distribution.



Aug 30-10:25 AM

RECAP

Individuals are the people or object included in the study.

A variable is a characteristic of the individual to be measured or observed.

What was our individual? What was our variable?

A quantitative variable has a value or numerical measurement for which operations such as addition or averaging make sense.

A qualitative variable describes an individual by placing the individual into a category or group, such as male or female.

What type of variable did we analyze with our fish?
A population data, the data are from *every* individual of interest.

Obtaining information about an entire population is called a census.

A sample data, the data are from *only* some of the individuals of interest.

What type of data did we use with our fish?

Generalizations based on the results of a random sample can be made back only to the population from which the sample came. Be sure to sample from the population of interest.

A population parameter is a numerical measure that describes an aspect of a population.

A sample statistic is a numerical measure that describes an aspect of a sample.

Aug 30-10:27 AM

Example 1: Using Basic Terminology

Television station QUE wants to know the proportion of TV owners in Virginia who watch the station's new program at least once a week. The station asks a group of 1000 TV owners in Virginia if they watch the program at least once a week.

1. Identify the individuals of the study and the variable.
The individuals are the 1000 TV owners surveyed. The variable is the response "does" or "does not" watch the new program at least once a week.
2. Do the data comprise a sample? If so, what is the underlying population?
The data comprise a sample of the population of responses from all TV owners in Virginia.
3. Is the variable qualitative or quantitative?
Qualitative- the categories are the two possible responses, "does" or "does not" watch the program.
4. Identify a quantitative variable that might be of interest.
Age or income might be of interest.
5. Is the proportion of viewers in the sample who watch the new program at least once a week a statistic or a parameter?
Statistic- the proportion is computed from sample data.

Aug 30-10:36 AM

Example 1: Using Basic Terminology

Television station QUE wants to know the proportion of TV owners in Virginia who watch the station's new program at least once a week. The station asks a group of 1000 TV owners in Virginia if they watch the program at least once a week.

1. Identify the individuals of the study and the variable.
The individuals are the 1000 TV owners surveyed. The variable is the response "does" or "does not" watch the new program at least once a week.
2. Do the data comprise a sample If so, what is the underlying population?
The data comprise a sample of the population of responses from all TV owners in Virginia.
3. Is the variable qualitative or quantitative?
Qualitative- the categories are the two possible responses, "does" or "does not" watch the program.
4. Identify a quantitative variable that might be of interest.
Age or income might be of interest.
5. Is the proportion of viewers in the sample who watch the new program at least once a week a statistic or a parameter?
Statistic- the proportion is computed from sample data.

Aug 30-10:36 AM

Levels of Measurement

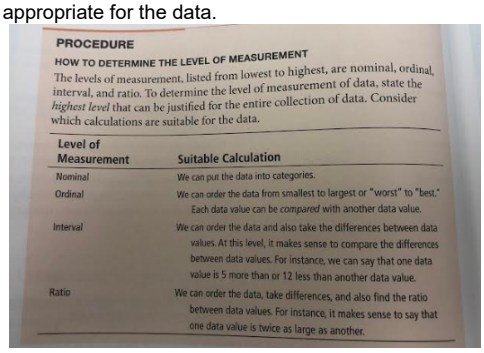
This is another way to classify data.

1. The nominal level of measurement applies to data that consist of names, labels, or categories. There are no implied criteria by which the data can be ordered from smallest to largest. qualitative
2. The ordinal level of measurement applies to data that can be arranged in order. However, differences between the data values either cannot be determined or are meaningless.
3. The interval level of measurement applies to data that can be arranged in order. In addition, differences between data values are meaningful.
4. The ratio level of measurement applies to data that can be arranged in order. In addition, both differences between data values and ratios of data values are meaningful. Data at the ratio level have a true zero.

Aug 30-5:02 PM

1.1 What is Statistics with answers

You should categorize the level of measurement by the *highest* level appropriate for the data.



PROCEDURE

HOW TO DETERMINE THE LEVEL OF MEASUREMENT

The levels of measurement, listed from lowest to highest, are nominal, ordinal, interval, and ratio. To determine the level of measurement of data, state the *highest level* that can be justified for the entire collection of data. Consider which calculations are suitable for the data.

Level of Measurement	Suitable Calculation
Nominal	We can put the data into categories.
Ordinal	We can order the data from smallest to largest or "worst" to "best." Each data value can be *compared* with another data value.
Interval	We can order the data and also take the differences between data values. At this level, it makes sense to compare the differences between data values. For instance, we can say that one data value is 5 more than or 12 less than another data value.
Ratio	We can order the data, take differences, and also find the ratio between data values. For instance, it makes sense to say that one data value is twice as large as another.

Aug 30-5:11 PM

Example 2: Levels of Measurement

The following describe different data associated with a state senator. For each data entry, indicate the corresponding *level of measurement.*

1. The senator's name is Sam Wilson.
2. The senator is 58 years old.
3. The years in which the senator was elected to the Senate are 2000, 2006, and 2012.
4. The senator's total taxable income last year was $878,314.
5. The senator surveyed his constituents regarding his proposed water protection bill. The choices for response were strong support, support, neutral, against, or strongly against.
6. The senator's marital status is "married".
7. A leading news magazine claims the senator is ranked seventh for his voting record on bills regarding public education.

Nominal Ordinal Interval Ratio

Aug 30-5:12 PM

Example 2: Levels of Measurement

The following describe different data associated with a state senator. For each data entry, indicate the corresponding *level of measurement.*

1. The senator's name is Sam Wilson. Nominal level.
2. The senator is 58 years old. Ratio level. Notice that age has a meaningful zero. It makes sense to give age ratios.
3. The years in which the senator was elected to the Senate are 2000, 2006, and 2012. Interval level. Dates can be ordered, and the difference between dates has meaning. However, ratios do not make sense. In addition, the year 0 does not mean "no time".
4. The senator's total taxable income last year was $878,314. Ratio level. It makes sense to say that the senator's income is 10 times that of someone earning $87,831.40
5. The senator surveyed his constituents regarding his proposed water protection bill. The choices for response were strong support, support, neutral, against, or strongly against. Ordinal level. The choices can be ordered, but there is no meaningful numerical difference between the two choices.
6. The senator's marital status is "married". Nominal Level.
7. A leading news magazine claims the senator is ranked seventh for his voting record on bills regarding public education. Ordinal level. Ranks can be ordered, but differences between ranks may vary in meaning.

Aug 30-5:12 PM

Descriptive statistics involves methods of organizing, picturing, and summarizing information from samples or populations.

Inferential statistics involves methods of using information from a sample to draw conclusions regarding the population.

Aug 30-5:26 PM

HW: pg. 10: 1, 5, 7, 9, 11, 13, 15

Answers are in the back. Check your answers. Tomorrow we will go over any questions from the homework.

Aug 30-5:28 PM

Beyond the Sea

Statistics mean never having to say you're certain!

I am a little fish

What's so sexy about math?

Aug 30-10:31 AM