

## Math 3307 Chapter 1 and 2 video script plus Popper 1 and 2

Welcome to the first of the videos. This one is over Chapter 1 and Chapter 2. The Popper Questions are scattered throughout the text here and the answer form, called an EMCF, is up in CourseWare. I'm going to call it "Popper 1 and 2" throughout.

These two chapters are an introduction to **Statistics**...with a whole lot of vocabulary. In order to learn it all, I recommend typing up a dictionary in alphabetical order or making flash cards. Most of my successful students do this.

A list of some vocabulary words is at the end of the script.

Most everybody has an opinion about statistics. Let's look at our book's definition and some additional definitions from the internet.

Statistics is "the practice of collecting and analyzing numerical data in large quantities especially for the purpose of inferring proportions in a whole from those in a representative sample." First internet definition up. Google

Two points: Collecting and Analyzing plus Inferring Proportions

"Statistics is a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data." Merriam Webster

Two points: Collect/analyze plus Interpret/present

Investopedia by Mitchell Grant:

"Statistics is a form of mathematical analysis that uses quantified models, representation and synopses for a given set of experimental data or real-life studies. Statistics studies methodologies to gather, review, analyze and draw conclusions from data."

Two points: Gather/review/calculate plus Analyze/draw conclusions

Statistics: gathering and analyzing data, inferring the truth, making descisions, drawing conclusions. Doing things mathematically

All of this is true and we'll do all of it over the course of the semester. At the beginning we will collect and analyze. We'll look at data presentation, too. Then we'll take a little break and review some Probability. Then we'll get to inferring the facts and making decisions at the end of the semester.

**Descriptive statistics** is what the first half of the course is about. We will collect data and calculate certain measures and display data. There is no decision making in descriptive statistics, it's all factual. Lists of statistics with labels. Then we will shift gears and begin with **Inferential statistics**. We will estimate where the true statistic is and make important decisions in the second half of the semester. We will use data to make choices and decisions about the unknowable and we will quantify exactly how wrong we could be!

### Popper 1 and 2 Question 1

Statistics includes

- A. Analyzing
- B. Calculating
- C. Inferring
- D. All of the above
- E. None of the above

### Chapter 1 and 2 Essay #1

---

Write two paragraphs about your definition of Statistics, in your own words.

Turn in under Assigments in CourseWare.

Now for more vocabulary:

**Variable:** Ah, the “x” in an algebra problem...almost. We do have “X” and it does take on various values. It’s important to know the Domain...we need to be able to state all the possible values the variable can take on. So it’s not an unknown OFTEN. Sometimes variables are not numbers; we call these Qualitative variables. The numbers are Quantitative variables.

**Data and Data Set:** The numbers we will collect. The data set is the set of all numbers we collected.

**Categorical Data** can be gathered into specific groups and reported in charts, tables, and histograms by group name or type.

An example of categorical data: Suppose I want to survey the whole class for eye color. I’d set up a collection mechanism (an online survey for example). And my variable would be qualitative: eye color. I’d let X take on the values:

{blue, gray, hazel, green, brown, violet}.

I’d get everybody to report one of those colors. And I’d add up how many of each color the students in this class had. Then I’d make a table and report the number and the percent of each color.

### **Popper 1 and 2 Question 2**

The variable X here is eye color. It is

- A. Qualitative
- B. Quantitative

For example:

Blue eyes	5 students	11%	Note %s add to 100%
Gray eyes	1 student	02%	
Hazel eyes	4 students	09%	
Green eyes	3 students	06%	
Brown eyes	31 students	70%	
Violet eyes	1 student	02%	

The data is how many students with each color. How many is the **raw data** and the percents are the statistics.

The **Population** is the students in this class. If for example there are 45 students. Then the population has 45 members. If I had chosen a random **Sample** of 12 students to poll, then my population would number 45 and my sample would number 12. The Population is everything of interest for the study and the Sample is usually a subset of the population.

#### Population/Sample

Population parameter – the true value for a property we are measuring

Population statistics – numbers calculated for the whole population

Sample – a subset of the population of interest

Sample statistic – a number calculated from the sample subset members that is supposed to estimate the parameter of interest.

Often it's not realistic to get population statistics. Sample statistics are, hopefully, a good estimate of the true statistic. How good matters.

### Focus on understanding:

---

A local school district would like to conduct a survey to estimate the percentage of the registered voters in the district who would support a school bond levy (tax). To determine the level of support, the school board surveys 1,000 registered voters from their district. What are:

The population *everyone in the district*

Population parameters *would probably be 18 or older, eligible to vote*

The sample *1000 registered voters*

The variable(s)  *$\bar{X}$  yes or no*

Raw data *how many support "yes"; don't support "no"*

Sample statistics *100% binomial % yes % no*

---

### Sampling Techniques

pages 4 – 7

OYO

*onyourown*

Simple random sampling

Random sample

Systematic sampling, Convenience sampling, Cluster sampling

Stratified sampling

## **Bias in Data Collection**      **page 9**

IMPORTANT to know about or discover!

Bias: “Prejudice in favor or against one thing, person, or group compared with another, usually in a way considered to be unfair” Google Dictionary

Often cannot be avoided. But if you know about it, you can often compensate.

### **Chapter 1 and 2 Essay #2**

---

Television stations, radio stations, and newspapers often predict the winners of important elections long before the votes are counted. They make these predictions based on polls.

- A     What factors might cause a prediction to be inaccurate?
  
- B     Political parties often conduct their own pre-election polls to find out what voters think about their campaign and their candidates. How might a political party bias such a poll?

Keep your answer to one page front side only, please.

## **Chapter 2 Organizing and Displaying Data**

---

### **2.1 Displaying Categorical Data**

---

#### Frequency and Relative Frequency Tables

See the table with the eye colors above, Frequency is the number, Relative Frequency is the percent column.

**Dot diagrams:**

(line plots – page 33)

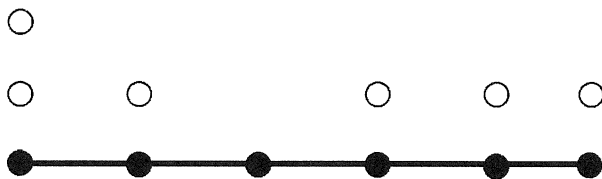
These summarize data visually and quickly. Put one dot for each observation. Note that you don't need to sort the data to make a dot diagram.

Dot diagrams are also useful with qualitative or categorical data.

For example:

If I toss a die 6 times and get: 1 4 5 6 1 2

I'd put a horizontal line down and mark off the 6 possible numbers and then put a dot above each recorded value:



**Popper 1 and 2 Question 3**

Which of the following are properties of dot diagrams?

- A. Quick and easy
- B. Shows frequency
- C. Shows percent of population
- D. A and B only
- E. Takes a lot of time

### Dot Diagram Homework question one.

2 1 5 0 1 3 2 0 7 1 3 4 2 4 1 2 2 5 1 3 4  
3 1 1 0 2 4 1 1 3 2 3 5 2 2 4 4 0 3 1 4 0

This data summarizes the number of times per week that a small regional airport with 44 flights per day that there are delayed takeoffs. These are all single digits.

Make a dot diagram and analyze the data completely. Show your diagram and identify the highest frequency with an arrow.

### Bar Graphs and Circle graphs

---

More data display. Bar diagrams show frequency or relative frequency on a pair of axes usually with frequency on the y axis. Circle graphs (also called pie charts) are particularly useful for relative frequency data like percents. Histograms and bar charts are related.

Example:

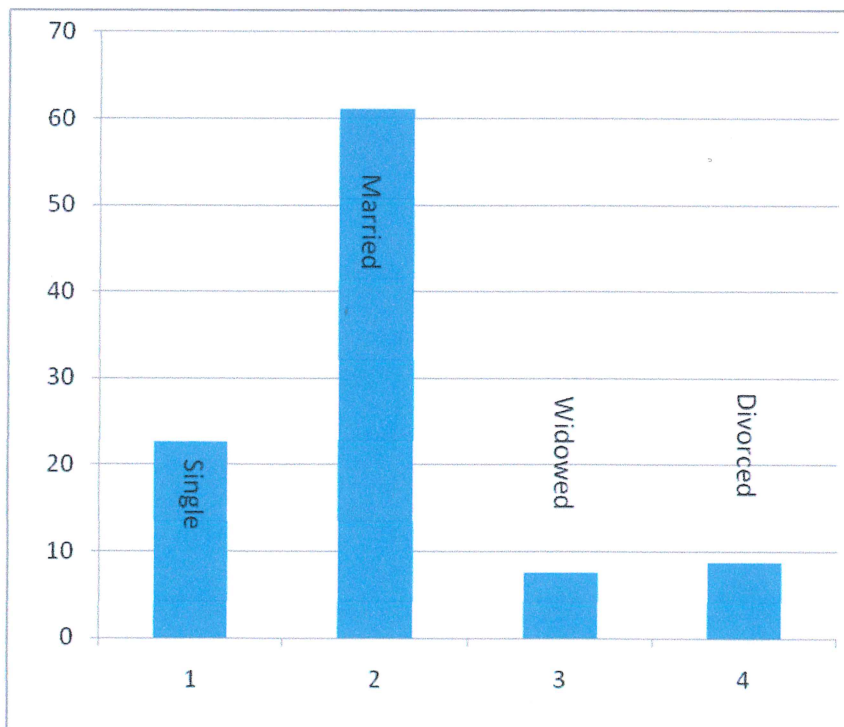
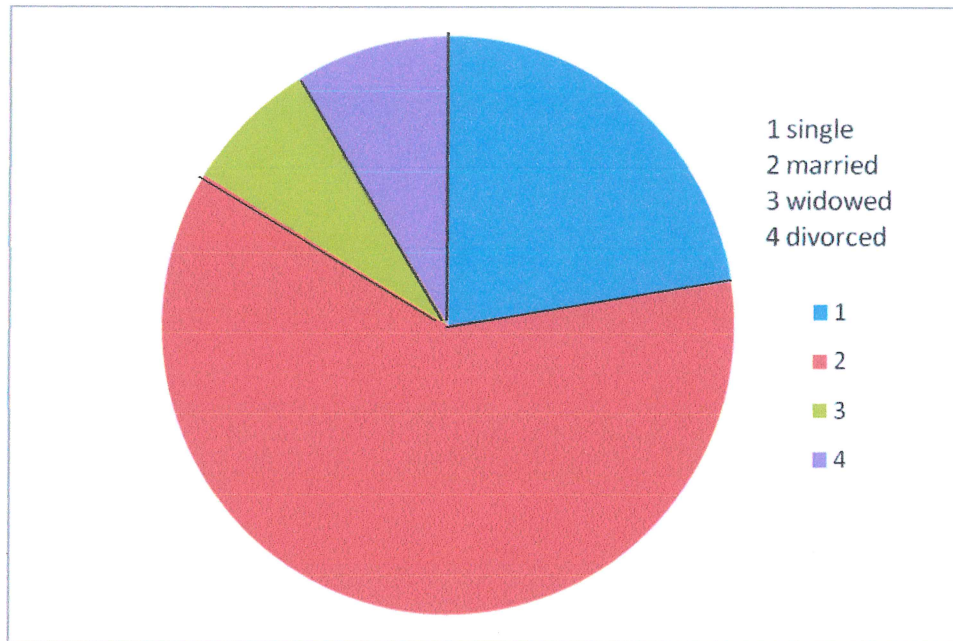
Here is a distribution of information about Americans aged 18 or older from the 1990's

Marital status	Count In Millions	Percent
Single	41.8	22.6
Married	113.3	61.1
Widowed	13.9	7.5
Divorced	16.3	8.8



There are a couple of ways to display this information graphically. One is a histogram or bar chart and another is a pie chart or circle graph.

### Pie chart



% on y axis. Categorical data

Why was it important to use the percentages and not the raw counts in both representations? To get it into a reasonably sized chart. Easier to compare too. Note the x axis is used for the data categories.

\*\*\*See page 24 for a useful summary of which type of representation to use when.

## 2.2 Displaying Quantitative Data

---

### Frequency and Relative Frequency Tables

The Rules page 26

Example:

Fifty candidates entering an astronaut training program were given a psychological profile test measuring bravery. NASA grouped the data to make it more compact.

Note that the scores are grouped into units of the SAME length. Why is this important?

Score in points	# of candidates
60 - 79	8
80 - 99	16
100 - 119	18
120 - 139	8

140 - 159	6
What about the extremes?	

An improvement on dot diagrams, stem and leaf plots work on data with many various measurements. It is fairly low tech and can be quickly done in a meeting or on the fly. I find them exceptionally useful in small classes ( $n < 50$ ) for a quick grade analysis.

The stems are the 10's and the leaves are the single digits in each day's total. It can be useful to organize the leaves in order, too.

Here is one of my classes, a final:

10	123	
09	45779	
08	327758	
07	459	
06	78	
BELOW	1111	④

Let's find the median – The grade in the middle:

count down: 101, 102, 103 and up:

the 4 belows 60s, and then 67, 68, 74, which one is the median? The second 87.

Turn the page sideways (anti clockwise)...note the resemblance to a dot diagram!  
What does this tell you about my class?

*lots of good grades! the median was a B+*

Note that in each case, there was somebody pretty close to the next level.

**Popper 1 and 2 Question 3**

How many students were in my class?

- A. Can't tell from the diagram
- B. 10
- C. 11
- D. 29
- E. 23

Sometimes if the data is unusually condensed, you might split the stems making more rows rather than fewer rows.

Here are some quiz grades out of 130 points:

112 114 114 116 118 119 120 121 122 123  
124 125 125 126 127 127 129

*nicely ordered  
is unusual... usually  
you have to do it  
yourself*

The best data presentation is to show 110 – 114, 115 – 119, 120 – 124, 125 – 129 rather than just 2 stems with LOOOOONG leaf lines:

11		244
11		689
12		01234
12		556779

Note that the stems are now both a hundreds and a tens digit! And note that they are ordered numbers!

Summarizing we have

Vocabulary:

Statistics, Inferential Statistics, Inferential Statistics. Variables – qualitative and quantitative. Data and data sets. Categorical Data. Population, Sample. Population parameter, sample statistic. Bias.

### **For turn-in and grading:**

#### **Chapter 1 and 2 Essay #1**

---

Write two paragraphs about your definition of Statistics, in your own words.

#### **Chapter 1 and 2 Essay #2**

---

Television stations, radio stations, and newspapers often predict the winners of important elections long before the votes are counted. They make these predictions based on polls.

- A What factors might cause a prediction to be inaccurate?
- B Political parties often conduct their own pre-election polls to find out what voters think about their campaign and their candidates. How might a political party bias such a poll?

**Popper 1 and 2: 3 questions**

#### **Homework from the book:**

Chapter 1: 2, 4, 8

Chapter 2: 2, 4, 6, 8

Plus one problem from the script (below)

**Dot Diagram Homework question one.**

2 1 5 0 1 3 2 0 7 1 3 4 2 4 1 2 2 5 1 3 4  
3 1 1 0 2 4 1 1 3 2 3 5 2 2 4 4 0 3 1 4 0

This data summarizes the number of times per week that a small regional airport with 44 flights per day that there are delayed takeoffs. These are all single digits.

Make a dot diagram and analyze the data completely. Show your diagram and identify the highest frequency with an arrow.