

Panel 1

Last Time

```

    graph TD
      variables --- cont
      variables --- discrete
      variables --- numeric
      variables --- indep
      variables --- depend
      cont --- ord1[ord.]
      cont --- nom1[nom.]
      numeric --- ord2[ord.]
      numeric --- nom2[nom.]
    
```

Random Sample : unbiased sample

Frequency Distribution: count + probability of them

1

Panel 2

1. Which year did you enter Seton Hall University?	<input type="text"/>	discrete; numerical
2. Which year do you expect to graduate from Seton Hall University?	<input type="text"/>	discrete; numerical
3. Are you male or female?	<input type="radio"/> Male <input type="radio"/> Female	nom.
4. What is your age in years?	<input type="text"/>	num + cont.
5. What is your weight in pounds (enter number only)?	<input type="text"/>	num + cont.
6. What is your height in inches (enter number only)?	<input type="text"/>	num + cont.
7. How many CD's do you own (approximately)?	<input type="text"/>	num. + discrete
8. Are you left or right handed?	<input type="radio"/> Left Handed <input type="radio"/> Right Handed <input type="radio"/> Ambixtertrous	nom.

2

Panel 3

The BBC asked viewers to call in with their favorite song. 7500 people called, over half liked "Imagine". Does this represent a random sample?

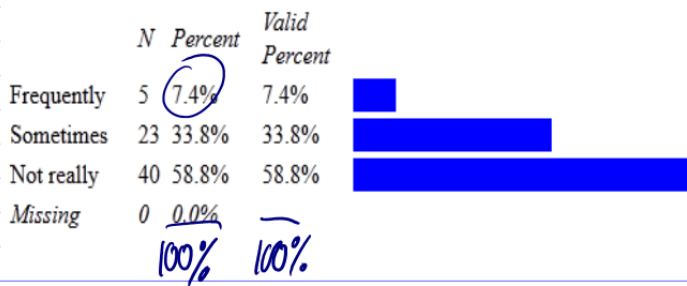
No!

3

Panel 4

A survey results in the following data:

11. Do you play video games on your computer?



Explain.

	Male / Female	comb	freq.	valid %
	1 2			
female	1	1	33.3%	50
male	1	1	33.3%	50
missing	1	1	33.3%	

4

Panel 5

Survey results in

Hair	freq	count	relid %
Brown	40%	40	$\frac{40}{100}$
Blond	5%	5	$\frac{5}{100}$
Red	5%	5	$\frac{5}{100}$
Black	40%	40	$\frac{40}{100} = 40.4\%$
unint	10%	10	X
	↑↑	100	100

How many people  
 a) are blond + red:  $\frac{10}{100}$   
 b) are not black:  $\frac{30}{100} = 30.0\%$

5

Panel 6

Name: \_\_\_\_\_

Quiz #1:

① Below is a sample survey. How many variables does this survey define? Classify all variables as numerical, ordinal, or nominal.

**Sample Survey**

Enter your name: \_\_\_\_\_

What is your cholesterol level: \_\_\_\_\_

How often do you visit a doctor:  Often  
 Sometimes  
 Rarely  
 Never

② Restaurant A collects voluntary "satisfaction" cards, restaurant B asks every 5<sup>th</sup> guest their opinion. Which one is closer to a random sample?

6

Panel 7

③ A survey about voter opinion results in the following frequency distribution:

a) What is the winning prob.?	very liberal	0.05
	liberal	0.20
	neutral	0.25
	conservative	<input type="text"/>
b) How many people, in %, are liberal or very liberal?	very conservative	0.20

7

Panel 8

Frequency distribution can be done graphically.

Other types of charts are Bar Charts, Pie Charts

apply to categorical variables

For numeric variables use Histograms

Age Group	Relative Frequency
10-20	0.02
20-30	0.16
30-40	0.18
40-50	0.19
50-60	0.18
60-70	0.14
70-80	0.08
80-90	0.05

8

Panel 9

Creates a Histogram manually

Data: 20, 32, 70, 80, 43, 41, 58, 26, 79, 33, 75, 96, 90, 22

① Find largest: 90  
smallest: 20

② Pick # of bins 5  
(categories, 5-10)

③ Find range = max - min = 60

④ Compute  $\frac{\text{range}}{\# \text{ bins}} = \text{bin width} = \frac{60}{5} = 12$   
↑  
round up

20-32	4
32-44	5
44-56	1
56-68	1
68-80	3

⑤ Make a table

Panel 10

Numeric Data Representation: Measures of Central Tendency

1.) Mean (average): add all values, divide by total number of numbers  
(sample size)

$$\text{mean} = \frac{1}{N} (x_1 + x_2 + x_3 + \dots + x_N) = \frac{\sum x_i}{n}$$

*sigma* (stands for sum)

Two symbols are used for mean:

$\mu$  = population mean  
 $\bar{x}$  = sample mean  
| x-bar

Ex: Data: 2, 3, 5, 4, 1, 5

$$\bar{x} = \frac{1}{6} (2+3+5+4+1+5) = \frac{1}{6} \cdot 20 = \frac{10}{3}$$

$$\bar{x} = 3.33$$

(if data represents a sample)

Panel 11

② Median is the middle value, i.e. that value s.t. 50% of data is less than it

A.) Must sort data, start with smallest

B.) Pick # in the middle if possible

Ex: 2, 3, 5, 4, 1, 5  $\Rightarrow$  1, 2, 3, 4, 5, 5

if  $N$  is even, we take the middle between closest numbers.

median = 3.5

Ex: 1, 3, 7  $\Rightarrow$  median = 3  
 1, 3, 7, 8  $\Rightarrow$  median =  $\frac{3+7}{2} = 5$

11

Panel 12

③ Mode: most frequent value

Data: 2, 3, 5, 4, 1, 5  $\Rightarrow$  mode = 5

Data: 2, 3, 5, 4, 2, 1, 5  $\Rightarrow$  mode = 2, 5

Why are there 3 measures of central tendency?

For numeric data: mean, median, mode makes sense

Ordinal data: ~~mean~~, median + mode makes sense

Nominal data: ~~mean~~, ~~median~~, mode makes sense

12