

Panel 1

Stats: Cost Time

Grades/web site/, DyKnow/, StatCrunch

Statistics: making sense of data (Science of...)

Population, sample, descriptive + inferential stats

Parameters, statistics

Types of variables:

- categorical
 - ordinal
 - nominal
- quantitative (numeric)

Panel 2

Other categories for variables:

continuous: infinitely many possibilities (in every subinterval)

discrete: you can count # of choices.

Example

numeric: Age: continuous

numeric: Income: discrete

ordinal: Opinion about today's lecture, from 0 (sad) to 6 (super) *cont. but treated as discrete*

nominal: Who did you vote for: Obama McCain *discrete*

Panel 3

Dealing with Two Variables

Often deal with situation where we want to analyze two variables and their connections

Ex: Smoking vs. Lung Cancer

Often, one variable is independent, the other dependent. Independent one is often earlier in time.

Ex: Age vs. Party Affiliation
 Height vs. Weight
 Smoking vs. Lung Cancer

3

Panel 4

Random Sample

Want to pick a sample that represents a population adequately.

Def: Random Sample is a sample chosen in such a way that every other sample of the same size has the same probability to be selected

"unbiased" sample

4

Panel 5

Example: Find average income of NYC residents.

① Ask everyone. *impractical.*

② Select random sample of, say, 1000 NYC residents *biased*

① ~~Phone book. Pick a page, 1000 numbers from that point on.~~

② Use a computer to pick 1000 #'s from NYC phone book.
better

*stratified
random
sample*

pick subsamples in different areas

*difficult to
pick true
random sample*

5

Panel 6

Ex: Last election I saw on TV at 9:45 pm:

Channel X: Candidate A - 44% *could not*

Candidate B - 52% *happen*

Margin of error: 3% *for true*

Channel Y: Candidate A - 49% *random*

Candidate B - 48% *samples*

Margin of error - 2%

What is wrong with this picture?

6

Panel 7

Distribution: set of all possible values of a variable together with their frequencies (or probabilities)

Ex: Data: What's your gender: 1, 2, 1, 1, 2, 2, 2, 9
(1 = male, 2 = female, 9 = no answer) - Frequency Distribution

	count	frequency %	valid % (percentage)
(1) male	3	$\frac{3}{8} = 0.375 = 37.5\%$	$\frac{3}{7} = 42.86\%$
(2) female	4	$\frac{4}{8} = 50\%$	$\frac{4}{7} = 57.14\%$
(9) missing	1	$\frac{1}{8} = 12.5\%$	—
	8	100%	100%

7

Panel 8

Suppose you meet a random adult on the street. What is the chance that the person is male or female?

Use GSS survey!

Frequency table results for SEX:

SEX	Frequency	Relative Frequency	Percent
Female	1094	0.540781	54.078102
Male	929	0.45921898	45.921898

8

Panel 9

Example: A variable records religious affiliation.

Distribution is as follows

protestant	20%
catholic	15%
jewish	5%
muslem	10%
none	20%
other	30%
	<u>100</u>

What's the winning probability.

Panel 10

Analyse the highest degree in GSS data.

- a) how many have at least a BA ?
- b) how many have at most a HS degree ?

Frequency table results for HIGHEST DEGREE:

HIGHEST DEGREE	Frequency	Relative Frequency	Percent
0 - Less than HS	297	0.14688428	14.688427
1 - High School	1003	0.49604353	49.60435
2 - Junior College	173	0.085558854	8.555885
3 - Bachelor	355	0.17556874	17.556875
4 - Graduate	194	0.095944606	9.5944605

> 64.2%

> + 27%

Panel 11

To create a meaningful report, you

- a) need a good set of data, e.g. GSS
- b) need software to compute stats, e.g. StatCrunch + calculator
- c) need software to create a report: Word + ppt/parts

11

Panel 12

To summarize data:

Frequency distribution ✓

Bar charts

Pie charts

} categorical vars

Histograms

} interval vars.

Q7 on Board

12