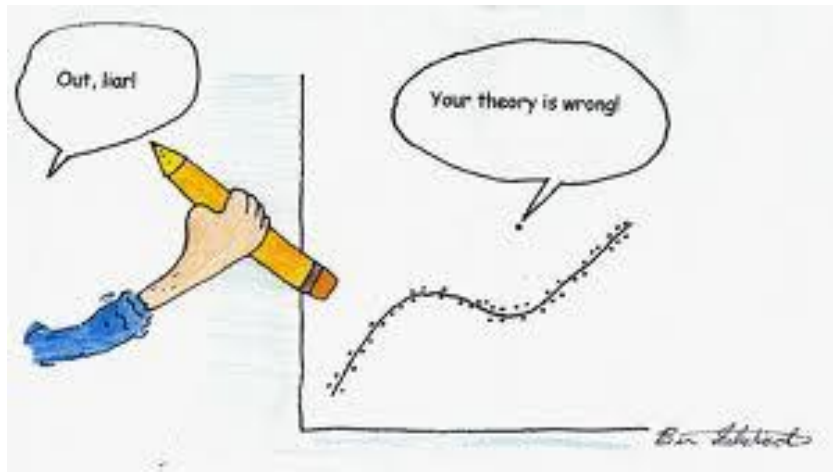


Chapter 1: Introduction to Statistics



Name _____

Period _____

1.1 Review and Preview

Data:

→ Ex: measurements, genders, survey responses

Statistics:

Common & Important Goal of Statistics:

Population:

→ Ex: scores, people, measurements, etc.

Census:

Sample:

*It is critical to obtain sample data that are representative of the population being studied

→ The data must be collected in an appropriate way (ie: random sampling); otherwise the data is useless.

→ Ex: If you surveyed your graduating class by asking them to write their annual income & mail it back to you, the responses won't be representative of the population.

Ex: The Gallup corporation collected data from 1013 adults in the United States. Results showed that 66% of the respondents worried about identity theft.

→ What is the population?

→ What is the sample?

→ What was the objective of this study?

1.2 Statistical and Critical Thinking

Key Elements in a Statistical Study:

Prepare Data:

1. Context
 - What does the data mean? What is the goal of the study?
2. Source of the Data
 - Are the data from a source with a special interest so that there is pressure to obtain results that are favorable to the source?
3. Sampling Method
 - Were the data collected in a way that is unbiased?

Analyze Data:

4. Graph the Data
5. Explore the Data
 - Were there any outliers? How is the data distributed? Are there missing data? What important statistics summarize the data?
6. Apply Statistical Methods
 - Use technology to obtain results

Form Conclusions:

7. Statistical Significance
 - Do the results have statistical significance?
 - Achieved in a study when we get a result that is very unlikely to occur by chance
 - Do the results have practical significance?
 - Common sense may suggest that the finding does not make enough of a difference to justify its use or to be practical

Ex: Determine whether there is a relationship between IQ score and brain volume.

IQ	96	87	101	103	127	96	88	85	97	124
Brain Volume (cu. cm)	1005	1035	1281	1051	1034	1079	1104	1439	1029	1160

Context: The data in the table consists of measured IQ scores and measured brain volumes from 10 different subjects suggesting a possible hypothesis: People with larger brains tend to have higher IQ scores.

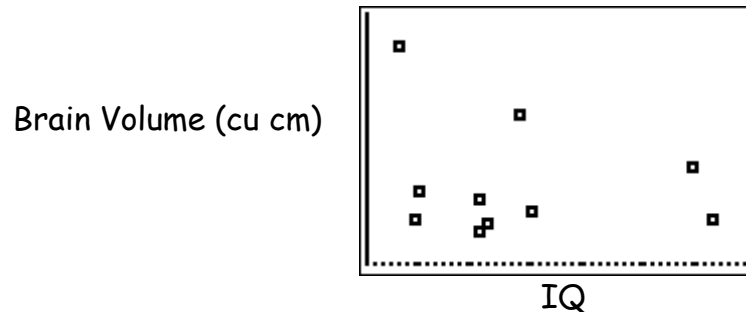
Source of the Data: "Brain Size, Head Size, and IQ in Monozygotic Twins" article written by M.J. Tramo, W.C., Loftus, T.A. Stukel, J.B. Weaver, and M.S. Gazziniga

- Researchers are from reputable medical schools & hospitals
- Would not gain by putting a spin on the results

→ When physicians who conduct clinical experiments of the efficacy of drugs receive funding from drug companies, they have an incentive to obtain favorable results

Sampling Method: Subjects were recruited and paid for their participation, all were between the ages of 24 and 43, all had at least a high school education, and no subjects medical history included neurologic or psychiatric disease.

Graph the Data:



Explore the Data: One outlier of an IQ of 85 and a brain volume of 1439, there is no missing data, the data does not appear to follow any type of pattern between IQ or brain volume—certainly not increasing.

Apply Statistical Methods: Correlation coefficient was found to be -0.22 using the TI graphing calculator indicating a very weak, negative correlation

Conclusion: There is not enough data or evidence to conclude there is a relationship between IQ and brain volume.

Pitfalls of Analyzing Data:

→ Voluntary response sample (self-selected sample):

- ◆ Ex: polls conducted through the internet, mail, or telephone
- ◆ Ex: readers of Newsweek magazine are asked to respond online to the question: "Will you still use Napster if you have to pay a fee?"
- ◆ Conclusions can only be made about those who respond, not the population

→ Correlation & Causality:

- ◆ Ex: wealth & IQ

→ Reported Results:

- ◆ Ex: If you measure people's weights, they will probably differ from the weights the same people would tell you

→ Small samples:

- ◆ Ex: The Children's Defense Fund published "Out of School in America" where it reported that among high school students in this region, 67% were suspended at least 3 times

- ◆ Just because a sample is large, doesn't mean it was collected in an appropriate way

→ Loaded Questions:

- ◆ Ex: Even though it's the poor who receive welfare, only 19% agreed that too little money is being spent when the word "welfare" was used, but 63% agreed when it was worded "assistance to the poor"

→ Order of Questions:

- ◆ Ex: When asked what contributes more to air pollution—traffic or pollution, more people blamed which ever was presented first.

→ Nonresponse:

- ◆ High refusal rate because of persistent telemarketers
- ◆ Sugging: sales pitch that initially sounds like an opinion poll

→ Missing Data:

- ◆ At random
- ◆ Due to special factors

→ Precise Numbers:

- ◆ It would be better to say there the number of adults in the U.S. is about 240 million, rather than stating that there are 241,472,385

→ Percentages:

- ◆ Percent means "divided by 100"
- ◆ To find the percent of a number:
--Find 6% of 1200
- ◆ To change a fraction to a percent:
--Express $\frac{3}{4}$ as a percent
- ◆ To change a decimal to a percent:
--Express 0.25 as a percent
- ◆ To change a percent to a decimal:
--Express 85% as a decimal
- ◆ Ex: If you take 100% of a quantity, you take it all.

1.3 Types of Data

Parameter:

→ Ex: Of the total 3,250 pedestrian pushbuttons in NYC, 77% do not work.

Statistic:

→ Ex: Out of 877 executives surveyed, 45% said they wouldn't hire someone with a typo on their job application.

Ex: In a Harris poll, 2320 adults in the US were surveyed about body piercings and 5% of respondents said that they had a body piercing but not on the face. Based on the latest available data there are 241,472,385 adults in the US.

Parameter:

Statistic:

Categorical Data (also known as qualitative or attribute data):

→ Ex: eyecolors or genders of professional athletes

→ Categorical data as numbers: The numbers 12, 74, 77, 76, 73, 78, 88, 19, 9, 23 and 25 were sewn on the jerseys of the starting offense for the New Orleans Saints (these numbers are substitutes for names and do not measure or count anything)

Quantitative Data:

→ Ex: heights or weights of supermodels

→ Use the appropriate unit of measure

→ Three types of quantitative data:

1. Finite Discrete Data:

-- Ex: # of eggs a hen laid

2. Infinite Discrete Data:

--Ex: you could roll a die

3. Continuous (or Numerical) Data:

-- ex: amount of milk from a cow

Another Way to Classify Data—Using 4 Levels of Measurement:

1. Nominal:

→ Lack ordering & shouldn't be used for calculations

→ Ex: survey responses (yes, no, undecided), Social Security Numbers (are substitutes for names & do not count or measure anything)

2. Ordinal:

→ Ex: ranking vacation destinations 1 to 5

3. Interval:

→ Ex: temperatures

→ Ex: the years 1000, 2008, 1776, 1492

4. Ratio:

- Ex: prices of college textbooks
- Ex: weights (in carats) or diamond engagement rings
- zero starting points makes ratios meaningful

1.4 Collecting Sample Data

*If sample data are not collected in an appropriate way, the data may be so utterly useless that no amount of statistical torturing can salvage them.

Observational Study:

Ex: Gallup poll

Experiment:

→ Experimental Units: subjects in experiments

-- ex: clinical trial of the drug Lipitor

→ Experiments are often better than observational studies because they typically reduce the chance of having the results affected by some variable that is not part of the study (lurking variable)

--In an observational study, you may incorrectly conclude that ice cream causes drowning. When you notice temperature as the lurking variable and perform an experiment, where one group got ice cream and the other did not, you'd notice that ice cream consumption has no effect on drowning.

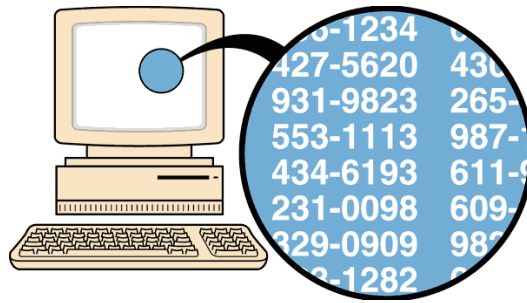
Common Methods of Sampling:

*It is essential to select the sample of subjects in such a way that the sample is likely to be representative of the larger population

❖ Random sample:

→ selected by many different methods

→ must be planned & carried out carefully



❖ Simple random sample (of size n):

Ex: A teacher selects 10 students by rolling a die and selecting the corresponding row (6 rows of 10 students)

→ random sample?

→ simple random sample?

Ex #2: Consider a box with 100 marbles. You reach in and select 13 marbles. Is this a random sample or a simple random sample?

Ex #3: For the Presidential election, you select a random sample of all voting precincts in your state and then interview all the voters as they leave the polling place. Is this a random sample or a simple random sample?

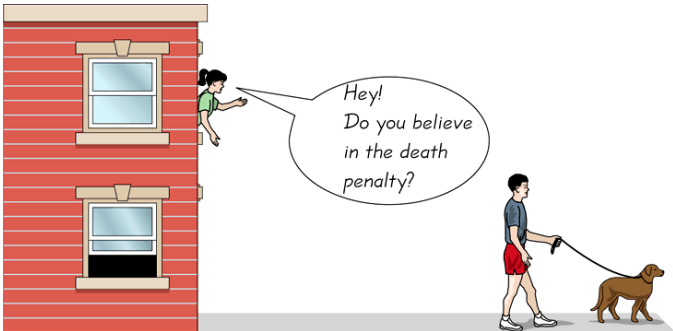
❖ Systematic Sampling:

→ Ex: every 3rd person who enters



❖ Convenience Sampling:

→ Ex: sample only those in this class

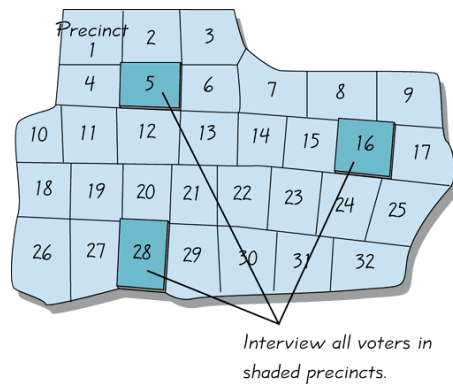


❖ Stratified Sampling:



❖ Cluster Sampling:

→ Cluster sampling uses all members from a sample of clusters; whereas, stratified sampling uses a sample of members from all subgroups (& often reduces variation in the results).



❖ Multistage Sampling:

→ Very complicated sampling design but it is more practical & less expensive than a simpler design

→ Used by the US Census Bureau

3 Types of Observational Studies:

1. Retrospective Study (also called case-control study):

→ collect data about the resulting characteristic that is of concern

→ Ex: collect data about a group of drivers who died in car crashes & a group of drivers who did not die in car crashes

2. Cross-Sectional Study:

3. Prospective Study (also called longitudinal or cohort):

→ Follow groups with a potentially causative factor & those without it

→ Ex: a group of drivers who use cell phones & a group that does not

Three Factors in Designing Experiments:

1. Use randomization
2. Use replication
3. Control the effects of variables

Randomization:

→ Uses chance as a way to create two groups that are similar (flipping a coin)

→ Still possible for randomization to result in unbiased samples, especially if sample sizes are very small

Replication:

→ Replication is effective when we have enough subjects to recognize the differences resulting from different treatments.

Controlling Effects of Variables:

→ Confounding:

◆ Results of the experiment may be ruined because of confounding

◆ Try to plan the experiment so confounding doesn't occur

→ Blinding:

◆ Placebo effect:

- ◆ Blinding:

- ◆ Double blind:

→ Completely Randomized Experimental Design:

- ◆ Ex: flip a coin to determine which trees got fertilizer

→ Randomized Block Design:

- ◆ Form blocks or groups of subjects with similar characteristics, then randomly assign treatments to the subjects within each block

- ◆ Ex: testing fertilizer on a block of trees in dry soil & a block in moist soil because the moisture content of the soil can affect tree growth

- ◆ Ex: make sure some trees in the moist soil have fertilizer & others don't so that you can determine if it was the fertilizer or moisture that was effective (watering, temperature, & sunlight must also be controlled)

→ Matched Pairs Design:

Ex: Before/After: You can collect measurements from subjects before and after some treatment

Ex: Twins: One twin was given Crest and the other was given another toothpaste to test Crest

→ Rigorously Controlled Design:

- ◆ Ex: you are testing the effectiveness of a drug that lowers blood pressure. If the placebo group contains a 30-year old overweight, male, smoker who drinks heavily & consumes an abundance of salt & fat, the treatment group should also include a person with these characteristics

Sampling Error:

- Error that comes from chance simple fluctuations

Nonsampling Error:

- Ex: selecting a biased sample; using a defective measure instrument; copying the data incorrectly

Nonrandom Sampling Error: