

DQR 401- Personal
Finance/Statistical Analysis

Study



Open Access 10

Topic 20

Sampling and Surveys

Definitions

- **Data:** observations that has been collected
- **Statistics** (2 definitions -singular/plural)
 - The science of collecting, organizing , and interpreting data
 - Data are observations that have been collected
 - The data that describe or summarize something

Example: Of the twenty students in class, 8 are male and 12 are female.

Statistics (plural): are the numbers

 - 20 students
 - 8 males
 - 12 females

Types of Statistics

- Descriptive Statistics

- Consists of organizing and summarizing data that has been collected

- The type of data you collect will determine what descriptive statistics can be applied

- Tables, graphs, charts, summary measures

- Inferential Statistics

- Methods that take sampling information, extends them to the population, measuring the reliability of the result

Types of Statistical Study

- Two Basic Types of Statistical Study

1. **Observational Study**

- We observe and measure a specific characteristic of the sample, we do not modify the subject in any way.

2. **Experimental Design**

- We apply a treatment to some or all of the sample and then try to measure the effect of the treatment.

Determine which type of study is given:

1. A sample of fish is taken from a lake to measure the effect of pollution from a nearby factory on the fish.

Observational Study

2. A clinic gives a drug to a group of ten patients and a placebo to another group of ten patients to find out if the drug has an effect on the patients' illness.

Experimental Design

Experimental Designs

- Treatment and Control Groups (selected randomly)
 - Treatment Group
 - The group of sample members who receive the treatment being tested.
 - Control Group
 - The group of sample members who do not receive the treatment being tested.
- Placebo/Placebo Effect
 - Placebo
 - Lacks the active ingredients of the treatment being tested.
 - Placebo Effect
 - Occurs when patients improve simply because they believe they are receiving a useful treatment.

Experimental Designs

- Blinding in Experiments
 - Ways we can try to eliminate the placebo effect
 - intended to increase the reliability of the experiment
 - Single-Blind
 - The participants in the study do not know whether they are in the control or treatment group, only the experimenters know.
 - Double-Blind
 - Neither the participants nor the experimenters in the study know who belongs to the control or treatment groups.

Seeking Correlation

- Relationship between Variables

- In observational studies and experimental designs, we often try to determine if a relationship exists between two variables.

- The **response variable** (dependent variable) is the variable whose value can be explained by, or is determined by, the value of the **explanatory variable** (independent variable).

- Is there a relationship between the variables? If yes, which is the response and explanatory variables?

- Height in inches and Head Circumference?

Explanatory - Height Response - Head Circumference

- Weight of a car and gas mileage?

Explanatory - Weight Response - Gas Mileage

- Shoe size and IQ? **No Relationship**

Definitions

- Population
 - The complete collection of all people or things to be studied.
- Sample
 - Is a subset of the population from which the data are actually obtained
- Population Parameters
 - Specific characteristics of the population that a statistical study is designed to estimate
- Sample Statistics
 - Numbers or observations that summarize raw data

Population and Sample

- Describe the population, sample, population parameters, and sample statistics
 - In order to determine the GPA of students at Lynn, a poll of 500 students to find the average GPA for these students.

Population:

- population of all students at Lynn

Sample:

- sample consists of 500 students selected

Parameter:

- average GPA of all students at Lynn

Statistic:

- average GPA of the students in the sample at Lynn

Confounding Variables

Watch Out for Confounding Variables

- **Confounding variables** are not intended to be part of a study
 - Can be difficult to interpret results properly
 - Factors that may influence a study's results (not measured)

Determine Confounding Variables in a Statistical Study

- A study is conducted to determine if there is a relationship between the number of hours studying for an exam and the exam score. List some possible confounding variables.
 - When was the studying done?
 - Done over several days or completed in one day (weeks ago)
 - Did you study in the morning or late at night
 - Health on the day of the exam
 - Amount of sleep the night before the exam

Choosing a Sample

Common Sampling Methods

- Simple random sampling
 - Choose items in such a way that every sample of given size has same chance of being selected -random number generator
- Systematic Sampling
 - Use a simple system to choose the sample, selecting every n th item from the population
- Convenience Sampling
 - Use a sample that is easy to collect, people that are easy to find and collect
- Stratified Sampling
 - Use when we are concerned about differences among subgroups, or **strata** within the population

Sampling Methods

- Identify the type of sampling method used. Do you think the sample is representative of the population?
 1. You survey student in your dorm by choosing the closest 10 rooms to yours.
convenience sampling - Yes, if randomly assigned to rooms
 2. I take a sample of individuals from class by selecting every 5th person in entering class.
systematic sampling - Yes, if randomly enter rooms
 3. I take a sample of individuals from class by selecting 5 women and 5 men to find average age of each gender.
Stratified sampling - Yes, if randomly select women/men
 4. I assign a number to each student in class and use a random number generator to select sample.
simple random sampling - Yes, best chance for representative sample

Bias in Sampling

A statistical study suffers from *bias* if its design or conduct tends to favor certain results.

- samples not representative of the population. We say that the sample is *biased*.

- Determine if the sample would be biased.

1. To determine the average age of students at Lynn University, we took a sample of 150 incoming freshman.

Biased - not representative of the population

2. To estimate the average income of a student at Lynn, we polled 500 students randomly selected from the registration office.

Not Biased - representative of the population

Bias in Statistical Studies

- Bias is Sampling Procedures

Look for Bias in the Sample (2 common forms)

1. Selection Bias

- occurs when researchers selects a sample in such a way that is **unrepresentative** of the population.
 - Usually select members that will give desirable results

2. Participation Bias (Non-Response Bias)

- Occurs in surveys and polls; arises when people **choose** whether they want to participate
 - People are more likely to participate if they feel strongly about an issue
 - Opinions may not be representative of the entire population
 - » Those with little attachment to an issue are not likely to participate in the study

Selection vs. Participation Bias

Analyze the following:

- Determine whether the study suffers from possible Selection Bias, Participation Bias or Both
 - A study is conducted where listeners of a sports radio show were asked to call in and vote which team will win Super bowl.

Possible Non-response Bias

- A poll of students is designed to estimate what percent are involved in a relationship with another Lynn student. A random sample of 100 freshman are selected.

Possible Selection Bias

- A study is conducted on American citizens to determine the public opinion of their home town. A sample of residents from New York City is selected and asked if they would like to participate in the survey.

Possible Selection Bias and Non-response Bias

Response Bias

- Response Bias

Consider the Setting and Wording in Surveys

- The setting or wording may produce inaccurate or dishonest responses
 - Surveys concerning sensitive results (personal questions) can produce dishonest results
 - A study was done at Lynn to determine the amount alcohol consumption by the student body. The survey requires you give your name and the results will be available to parents.
 - Bias can affect the outcome of a study (wording of question)
 - Do you agree with the record setting excess spending by the government this year (Deficit topping 1 trillion in first 9 months of 2009 for the first time ever)?
 - Do you feel the government should be spending money to combat the recession and economic problems (contributing to deficit).