Data Collection and Sampling Techniques; Experimental Design

1.3 and 1.4

Jacob Ayers

MAT 110

Lesson #2

Objectives

- Identify sampling techniques
- Explain the difference between observational studies and experimental studies
- Expalin how statistics can be used or misused

Data Collection

We use data in many ways:

- determine whether we're meeting goals
- determine needs of a population
- learn about consumers of a product
- gain insight into people's perception of political events

Data Collection

We use data in many ways:

- determine whether we're meeting goals
- determine needs of a population
- learn about consumers of a product
- gain insight into people's perception of political events

In order to use data, we must first collect it.

There are a variety of ways of collecting data.

Data Collection

Often, data is collected using a survey.

Telephone surveys are surveys that are conducted over the phone.

Data Collection

Often, data is collected using a survey.

Telephone surveys are surveys that are conducted over the phone.

Pros:

- Relatively cheap
- More candid responses (not face-to-face)

Data Collection

Often, data is collected using a survey.

Telephone surveys are surveys that are conducted over the phone.

Pros:

- Relatively cheap
- More candid responses (not face-to-face)

Cons:

- Not everyone can/will be reached by phone
- Easy for interviewer to influence person being surveyed with tone of voice

Data Collection

Mailed questionnaires are surveys that are sent through the mail.

Data Collection

Mailed questionnaires are surveys that are sent through the mail.

Pros:

- Even cheaper than telephone surveys
- Easily cover wide geographic areas
- Respondents can remain anonymous, which leads to more candid responses

Data Collection

Mailed questionnaires are surveys that are sent through the mail.

Pros:

- Even cheaper than telephone surveys
- Easily cover wide geographic areas
- Respondents can remain anonymous, which leads to more candid responses

Cons:

- Low response rate
- Inappropriate answers to questions (anonymity)
- Not everyone will be able to read/understand the questions

Data Collection

Personal interviews are surveys in which people are interviewed.

Data Collection

Personal interviews are surveys in which people are interviewed.

Pros:

- In-depth responses

Cons:

- Costly
- Difficult to cover wide geographic area
- Potential for bias in selection of respondents

Data Collection

Which method is best?

Data Collection

Which method is best?

Researchers must weigh cost and effectiveness when they choose how to collect data - there are no hard-and-fast rules.

Data Collection

Which method is best?

Researchers must weigh cost and effectiveness when they choose how to collect data - there are no hard-and-fast rules.

If you need to reach a lot of people, you might choose either a mailed questionnaire or a telephone survey because they are cheaper than interviewing people personally.

If you want in-depth answers, a personal interview may be the best choice.

Sampling Techniques

How we select our sample is very important - we want to ensure that everyone in the population has an equal chance of being selected.

Example: conducting a survey on a street corner from 8-noon would exclude people who work during that time. The information we obtain is likely to be biased.

Sampling Techniques

How we select our sample is very important - we want to ensure that everyone in the population has an equal chance of being selected.

Example: conducting a survey on a street corner from 8-noon would exclude people who work during that time. The information we obtain is likely to be biased.

There are four main ways that researchers obtain unbiased samples.

Sampling Techniques

Random Sampling

A *random sample* is a sample in which the members are selected either by chance methods or using random numbers

One way to do a random sample: put everyone's name in a hat, then draw x names from the hat.

Sampling Techniques

Random Sampling

A *random sample* is a sample in which the members are selected either by chance methods or using random numbers

One way to do a random sample: put everyone's name in a hat, then draw x names from the hat.

This method has a tendency to be biased, since it's hard to thoroughly mix up the names.

Sampling Techniques

Random Sampling

A *random sample* is a sample in which the members are selected either by chance methods or using random numbers

One way to do a random sample: put everyone's name in a hat, then draw x names from the hat.

This method has a tendency to be biased, since it's hard to thoroughly mix up the names.

Better way: Random number generator on a computer or a calculator.

Systematic Sampling

A *systematic sample* is a sample obtained by selecting every kth member of the population.

Example: say there are 1000 members of the population and we want a sample size of 50.

Systematic Sampling

A *systematic sample* is a sample obtained by selecting every kth member of the population.

Example: say there are 1000 members of the population and we want a sample size of 50.

We would select every $\frac{1000}{50} = 20$ members. The first member (between 1 and 20) would be selected at random, and then every 20th member after would be selected.

Sampling Techniques

Stratfied Sampling

A *stratified sample* is a sample obtained by dividing the population into subgroups based on some characteristic, then selecting subjects at random from each subgroup.

Samples within each subgroup are selected using random sampling; there can be many subgroups.

Example: Say I own ten restaurants and I want to get a feel for the difference in service at each location.

Sampling Techniques

Stratfied Sampling

A *stratified sample* is a sample obtained by dividing the population into subgroups based on some characteristic, then selecting subjects at random from each subgroup.

Samples within each subgroup are selected using random sampling; there can be many subgroups.

Example: Say I own ten restaurants and I want to get a feel for the difference in service at each location.

I could randomly select 25 customers from each restaurant to generate a stratified sample.

Cluster Sampling

A *cluster sample* is a sample obtained by dividing the population into subgroups (as in stratified sampling), then selecting all members of one or more of the clusters as the members of the sample.

Example: Say I own ten restaurants and I want to know people's thoughts on the menu.

Sampling Techniques

Cluster Sampling

A *cluster sample* is a sample obtained by dividing the population into subgroups (as in stratified sampling), then selecting all members of one or more of the clusters as the members of the sample.

Example: Say I own ten restaurants and I want to know people's thoughts on the menu.

I could randomly select two of the restaurants and survey all the customers who walk in to generate a cluster sample.

Sampling Techniques

Determine the type of sampling that was used in each case:

To check the accuracy of a machine filling coffee cups, every fifth cup is selected and weighed.

To determine how long people exercise, a researcher interviews 5 people from a yoga class, 5 people from a weightlifting class, 5 people from an aerobics class, and 5 people from a swimming class.

Sampling Techniques

Determine the type of sampling that was used in each case:

To check the accuracy of a machine filling coffee cups, every fifth cup is selected and weighed.

Since every fifth cup is measured, this is a systematic sample.

To determine how long people exercise, a researcher interviews 5 people from a yoga class, 5 people from a weightlifting class, 5 people from an aerobics class, and 5 people from a swimming class.

Sampling Techniques

Determine the type of sampling that was used in each case:

To check the accuracy of a machine filling coffee cups, every fifth cup is selected and weighed.

Since every fifth cup is measured, this is a systematic sample.

To determine how long people exercise, a researcher interviews 5 people from a yoga class, 5 people from a weightlifting class, 5 people from an aerobics class, and 5 people from a swimming class.

Since members of the sample are selected from various subgroups, this is a stratified sample.

Sampling Techniques

Determine the type of sampling that was used in each case:

In a large school district, a researcher numbers all of the full-time teachers and randomly selects 30 to be interviewed.

For 15 minutes, all customers entering a specific Walmart store on a specific day are asked how many miles from the store they live.

Sampling Techniques

Determine the type of sampling that was used in each case:

In a large school district, a researcher numbers all of the full-time teachers and randomly selects 30 to be interviewed.

Since the sample is obtained at random, it is a random sample.

For 15 minutes, all customers entering a specific Walmart store on a specific day are asked how many miles from the store they live.

Sampling Techniques

Determine the type of sampling that was used in each case:

In a large school district, a researcher numbers all of the full-time teachers and randomly selects 30 to be interviewed.

Since the sample is obtained at random, it is a random sample.

For 15 minutes, all customers entering a specific Walmart store on a specific day are asked how many miles from the store they live.

Since all members of the sample come from the same subgroup, this is a cluster sample.

Sampling Techniques

Which method is best?

Sampling Techniques

Which method is best?

It depends on the population and the researcher's needs.

Random sample: can be used any time the population can be numbered

Sampling Techniques

Which method is best?

It depends on the population and the researcher's needs.

Random sample: can be used any time the population can be numbered

Systematic Sample: selects subjects in ordered population; fast and convenient when population is easily numbered

Sampling Techniques

Which method is best?

It depends on the population and the researcher's needs.

Random sample: can be used any time the population can be numbered

Systematic Sample: selects subjects in ordered population; fast and convenient when population is easily numbered

Stratified Sample: useful when you want to determine differences between subgroups and on assembly lines

Sampling Techniques

Which method is best?

It depends on the population and the researcher's needs.

Random sample: can be used any time the population can be numbered

Systematic Sample: selects subjects in ordered population; fast and convenient when population is easily numbered

Stratified Sample: useful when you want to determine differences between subgroups and on assembly lines

Cluster Sample: can be done quickly and cheaply, but one cluster may not represent the entire population

Sampling Techniques

No sample is a perfect representation of the population from which it was chosen.

There will always be some degree of error in the results.

Sampling Techniques

No sample is a perfect representation of the population from which it was chosen.

There will always be some degree of error in the results.

Definition

Sampling error is the difference obtained from a sample and the results obtained from the population.

Sampling Techniques

No sample is a perfect representation of the population from which it was chosen.

There will always be some degree of error in the results.

Definition

Sampling error is the difference obtained from a sample and the results obtained from the population.

Example: Say you are president of a college whose student population is 60% female. If I obtain a sample that is 63% female, then the difference of 3% is due to sampling error.

Sampling Techniques

No sample is a perfect representation of the population from which it was chosen.

There will always be some degree of error in the results.

Definition

Sampling error is the difference obtained from a sample and the results obtained from the population.

Example: Say you are president of a college whose student population is 60% female. If I obtain a sample that is 63% female, then the difference of 3% is due to sampling error.

Usually, the degree of sampling error is unknown.

Sampling Techniques

In addition to sampling error, there is nonsampling error.

Definition

A *nonsampling error* occurs when the data are obtained erroneously or the sample is biased.

Sampling Techniques

In addition to sampling error, there is nonsampling error.

Definition

A *nonsampling error* occurs when the data are obtained erroneously or the sample is biased.

Examples of nonsampling error:

- measuring temperature on a defective thermometer
- researcher makes an error in recording a value

Experimental Design

There are two types of study:

Experimental Design

There are two types of study:

Observational studies are studies in which the researcher simply watches what is happening or what has happened in the past and tries to draw conclusions based on these observations.

Experimental Design

There are two types of study:

Observational studies are studies in which the researcher simply watches what is happening or what has happened in the past and tries to draw conclusions based on these observations.

Experimental studies are studies in which the researcher manipulates one variable and tries to determine how that manipulation affects other variables.

Observational Studies

There are three main types of observational study:

- Cross-sectional studies - data all gathered at one time

Observational Studies

There are three main types of observational study:

- Cross-sectional studies - data all gathered at one time
- Retrospective studies - data gathered from past records

Observational Studies

There are three main types of observational study:

- Cross-sectional studies - data all gathered at one time
- Retrospective studies - data gathered from past records
- Longitudinal studies - data gathered over time (past and present)

Observational Studies

Advantages of observational studies:

- occurs in natural setting - people won't be influenced by the researcher
- can be done in situations where performing an experiment would be unethical/dangerous (e.g. studying suicide rate)
- can be done using variables that the researcher is unable to manipulate (e.g. smokers vs. non-smokers)

Observational Studies

Disadvantages of observational studies:

- unable to determine direct cause-effect relationship between variables
- can be costly and time-consuming
- when gathering data from outside sources, the accuracy of your results is dependent upon the accuracy of their data

Experimental Studies

Example: divide students into two groups (randomly assign to each group). Each person does as many push-ups as possible.

Experimental Studies

Example: divide students into two groups (randomly assign to each group). Each person does as many push-ups as possible.

One group is given instructions to try to increase the number of push-ups they can do by 10% each day, while the other is simply told to do their best.

Experimental Studies

Example: divide students into two groups (randomly assign to each group). Each person does as many push-ups as possible.

One group is given instructions to try to increase the number of push-ups they can do by 10% each day, while the other is simply told to do their best.

After four days, the test was performed again.

Experimental Studies

Example: divide students into two groups (randomly assign to each group). Each person does as many push-ups as possible.

One group is given instructions to try to increase the number of push-ups they can do by 10% each day, while the other is simply told to do their best.

After four days, the test was performed again.

Here, the researchers manipulated one variable (instructions given) and measured the change in another variable (improvement at doing push-ups).

Experimental Studies

The *independent variable* is the variable that the researcher manipulates.

Experimental Studies

The *independent variable* is the variable that the researcher manipulates.

The *dependent variable* is the variable that changes as a result of the manipulation.

Experimental Studies

The *independent variable* is the variable that the researcher manipulates.

The *dependent variable* is the variable that changes as a result of the manipulation.

In our previous example, the instructions given were the independent variable and the number of push-ups was the dependent variable.

Experimental Studies

In our example, the students were divided into two groups - this is something that happens in all experiments

Experimental Studies

In our example, the students were divided into two groups - this is something that happens in all experiments

The group that received special instructions is called the *treatment group*, while the group that did not is called the *control group*.

Experimental Studies

Advantages of experimental studies:

- researcher gets to decide who is selected and who goes in what group
- researcher gets to control the independent variable

Experimental Studies

Advantages of experimental studies:

- researcher gets to decide who is selected and who goes in what group
- researcher gets to control the independent variable

Example: Researcher can determine what dosage of a medicine the treatment group receives (control group receives a placebo)

Experimental Studies

Disadvantages of experimental studies:

- may occur in unnatural setting - results don't translate
- Hawthorne effect - subjects behave differently when they know they're part of an experiment
- Confounding variables - variables other than the independent variable may impact results
- Placebo effect - even members of control group may respond positively to "treatment"

Experimental Design

Researchers randomly assigned 10 people to each of three different groups. Group 1 was instructed to write an essay about the hassles in their lives. Group 2 was instructed to write an essay about circumstances that made them feel thankful. Group 3 was asked to write an essay about events they felt neutral about. After the exercise, they were given a questionnaire on their outlook on life. The researchers found that those who wrote about circumstances that made them feel thankful had a more optimistic outlook on life. The conclusion is that focusing on the positive makes you more optimistic about life in general.

Experimental Design

Was this an observational study or an experimental study?

What is the independent variable?

What is the dependent variable?

What may be a confounding variable?

Experimental Design

Was this an observational study or an experimental study?

The researcher had control over which type of essay each participant wrote, so this is an experimental study.

What is the independent variable?

What is the dependent variable?

What may be a confounding variable?

Experimental Design

Was this an observational study or an experimental study?

The researcher had control over which type of essay each participant wrote, so this is an experimental study.

What is the independent variable?

The independent variable is the type of essay written.

What is the dependent variable?

What may be a confounding variable?

Experimental Design

Was this an observational study or an experimental study?

The researcher had control over which type of essay each participant wrote, so this is an experimental study.

What is the independent variable?

The independent variable is the type of essay written.

What is the dependent variable?

The dependent variable is the result of the questionnaire.

What may be a confounding variable?

Experimental Design

Was this an observational study or an experimental study?

The researcher had control over which type of essay each participant wrote, so this is an experimental study.

What is the independent variable?

The independent variable is the type of essay written.

What is the dependent variable?

The dependent variable is the result of the questionnaire.

What may be a confounding variable?

There are many possible confounding variables. For example, income is not controlled for, and it could certainly have an impact on one's outlook on life.

Experimental Design

What can you say about the sample size?

Do you agree with the conclusion?

Experimental Design

What can you say about the sample size?

The sample size is $10 \times 3 = 30$.

Do you agree with the conclusion?

Experimental Design

What can you say about the sample size?

The sample size is $10 \times 3 = 30$.

Do you agree with the conclusion?

This answer is subjective.

Next Steps

- Complete Assignment #1
- Begin Module #2
  - Read 2.1 and 2.2
  - Watch Video Lesson #3

Thanks for watching!