Module 1: Introduction

Jon Starkweather, PhD
jonathan.starkweather@unt.edu
Consultant
Research and **S**tatistical **S**upport

Introduction to Statistics for the Social Sciences

RSS
Research and Statistical Support

The RSS short courses

The Research and Statistical Support (RSS) office at the University of North Texas hosts a number of “Short Courses”. A list of them is available at:

`http://www.unt.edu/rss/Instructional.htm`

RSS
Research and Statistical Support

Outline

RSS
Research and Statistical Support

Outline

RSS
Research and Statistical Support

Outline

RSS
Research and Statistical Support

Outline

RSS
Research and Statistical Support

Seeking Truth

- Is the Earth round or flat; why do you believe one rather than the other?

Seeking Truth

- Is the Earth round or flat; why do you believe one rather than the other?
- Authority...Do you believe what I tell you simply because I have a PhD?

RSS
Research and Statistical Support

Seeking Truth

- Is the Earth round or flat; why do you believe one rather than the other?
- Authority...Do you believe what I tell you simply because I have a PhD?
  - I certainly hope not; you will learn a great deal more if you investigate.

RSS
Research and Statistical Support

Seeking Truth

- Is the Earth round or flat; why do you believe one rather than the other?
- Authority...Do you believe what I tell you simply because I have a PhD?
  - I certainly hope not; you will learn a great deal more if you investigate.
- Epistemology

RSS
Research and Statistical Support

Seeking Truth

- Is the Earth round or flat; why do you believe one rather than the other?
- Authority...Do you believe what I tell you simply because I have a PhD?
  - I certainly hope not; you will learn a great deal more if you investigate.
- Epistemology
  - The study of the nature and grounds of knowledge.

RSS
Research and Statistical Support

Seeking Truth

- Is the Earth round or flat; why do you believe one rather than the other?
- Authority...Do you believe what I tell you simply because I have a PhD?
  - I certainly hope not; you will learn a great deal more if you investigate.
- Epistemology
  - The study of the nature and grounds of knowledge.
- Science

RSS
Research and Statistical Support

Seeking Truth

- Is the Earth round or flat; why do you believe one rather than the other?
- Authority...Do you believe what I tell you simply because I have a PhD?
  - I certainly hope not; you will learn a great deal more if you investigate.
- Epistemology
  - The study of the nature and grounds of knowledge.
- Science
  - Investigation.

RSS
Research and Statistical Support

Seeking Truth

- Is the Earth round or flat; why do you believe one rather than the other?
- Authority...Do you believe what I tell you simply because I have a PhD?
  - I certainly hope not; you will learn a great deal more if you investigate.
- Epistemology
  - The study of the nature and grounds of knowledge.
- Science
  - Investigation.
  - Seek out answers through rigorous pursuit of valid facts and reliable phenomena.

RSS
Research and Statistical Support

Science

Why should we believe what science (or scientists) say?

- Peer Review

RSS
Research and Statistical Support

Science

Why should we believe what science (or scientists) say?

- Peer Review
  - Theories are believed because, the scientific community accepts the evidence for them (or not, if not).

RSS
Research and Statistical Support

Science

Why should we believe what science (or scientists) say?

- Peer Review
  - Theories are believed because, the scientific community accepts the evidence for them (or not, if not).
- Empiricism: Based on evidence.

RSS
Research and Statistical Support

Science

Why should we believe what science (or scientists) say?

- Peer Review
  - Theories are believed because, the scientific community accepts the evidence for them (or not, if not).
- Empiricism: Based on evidence.
  - Scientific Method: A formal method for developing and testing theories.

RSS
Research and Statistical Support

Science

Why should we believe what science (or scientists) say?

- Peer Review
  - Theories are believed because, the scientific community accepts the evidence for them (or not, if not).
- Empiricism: Based on evidence.
  - Scientific Method: A formal method for developing and testing theories.
- How do we achieve the **Goals of Science**:

RSS
Research and Statistical Support

Science

Why should we believe what science (or scientists) say?

- Peer Review
  - Theories are believed because, the scientific community accepts the evidence for them (or not, if not).
- Empiricism: Based on evidence.
  - Scientific Method: A formal method for developing and testing theories.
- How do we achieve the **Goals of Science**:
  - Observation yields Description (1st Goal).

RSS
Research and Statistical Support

Science

Why should we believe what science (or scientists) say?

- Peer Review
  - Theories are believed because, the scientific community accepts the evidence for them (or not, if not).
- Empiricism: Based on evidence.
  - Scientific Method: A formal method for developing and testing theories.
- How do we achieve the **Goals of Science**:
  - Observation yields Description (1st Goal).
  - Experimentation yields Explanation (2nd Goal)

RSS
Research and Statistical Support

Science

Why should we believe what science (or scientists) say?

- Peer Review
  - Theories are believed because, the scientific community accepts the evidence for them (or not, if not).
- Empiricism: Based on evidence.
  - Scientific Method: A formal method for developing and testing theories.
- How do we achieve the **Goals of Science**:
  - Observation yields Description (1st Goal).
  - Experimentation yields Explanation (2nd Goal)
  - Modeling (e.g., regression) yields Prediction (3rd Goal)

RSS
Research and Statistical Support

Science

Why should we believe what science (or scientists) say?

- Peer Review
  - Theories are believed because, the scientific community accepts the evidence for them (or not, if not).
- Empiricism: Based on evidence.
  - Scientific Method: A formal method for developing and testing theories.
- How do we achieve the **Goals of Science**:
  - Observation yields Description (1st Goal).
  - Experimentation yields Explanation (2nd Goal)
  - Modeling (e.g., regression) yields Prediction (3rd Goal)
- When predictions are confirmed, evidence is born for belief in a theory.

RSS
Research and Statistical Support

Science

Why should we believe what science (or scientists) say?

- Peer Review
  - Theories are believed because, the scientific community accepts the evidence for them (or not, if not).
- Empiricism: Based on evidence.
  - Scientific Method: A formal method for developing and testing theories.
- How do we achieve the **Goals of Science**:
  - Observation yields Description (1st Goal).
  - Experimentation yields Explanation (2nd Goal)
  - Modeling (e.g., regression) yields Prediction (3rd Goal)
- When predictions are confirmed, evidence is born for belief in a theory.
- When predictions fail, evidence is born for rejecting a theory.

RSS
Research and Statistical Support

What is statistics?

That's a very good, simple question; unfortunately, it does not have a simple answer.

What is statistics?

That's a very good, simple question; unfortunately, it does not have a simple answer.

- For a free overview consider the first two sections (definition & Scope) of the Wiki entry:

  `http://en.wikipedia.org/wiki/Statistics`

RSS
Research and Statistical Support

What is statistics?

That's a very good, simple question; unfortunately, it does not have a simple answer.

- For a free overview consider the first two sections (definition & Scope) of the Wiki entry:

  `http://en.wikipedia.org/wiki/Statistics`

- The reason there is no quick answer is because statistics has grown to include a great many subdivisions (some of which are downright hostile toward each other).

RSS
Research and Statistical Support

What is statistics?

That's a very good, simple question; unfortunately, it does not have a simple answer.

- For a free overview consider the first two sections (definition & Scope) of the Wiki entry:

  `http://en.wikipedia.org/wiki/Statistics`

- The reason there is no quick answer is because statistics has grown to include a great many subdivisions (some of which are downright hostile toward each other).
- As you will see; for our purposes, statistics will be very narrowly defined here.

RSS
Research and Statistical Support

Divisions within statistics

There are many divisions, just to give you an idea; consider these few:

- Theoretical and Applied

Divisions within statistics

There are many divisions, just to give you an idea; consider these few:

- Theoretical and Applied
  - This is very similar to saying; Mathematical and virtually everything else.

Divisions within statistics

There are many divisions, just to give you an idea; consider these few:

- Theoretical and Applied
  - This is very similar to saying; Mathematical and virtually everything else.
  - Mathematical statistics is the orthodoxy; we in the social sciences tend to be less rigid with regard to many topics and ways we go about conducting statistics (e.g., measurement, assumptions, etc.).

RSS
Research and Statistical Support

Divisions within statistics

There are many divisions, just to give you an idea; consider these few:

- Theoretical and Applied
  - This is very similar to saying; Mathematical and virtually everything else.
  - Mathematical statistics is the orthodoxy; we in the social sciences tend to be less rigid with regard to many topics and ways we go about conducting statistics (e.g., measurement, assumptions, etc.).
- Frequentist and Bayesian (more on this stuff later, RE: Probability)

RSS
Research and Statistical Support

Divisions within statistics

There are many divisions, just to give you an idea; consider these few:

- Theoretical and Applied
  - This is very similar to saying; Mathematical and virtually everything else.
  - Mathematical statistics is the orthodoxy; we in the social sciences tend to be less rigid with regard to many topics and ways we go about conducting statistics (e.g., measurement, assumptions, etc.).
- Frequentist and Bayesian (more on this stuff later, RE: Probability)
  - Within the Frequentist paradigm: Fisherian vs. Neyman & Pearson interpretation.

RSS
Research and Statistical Support

Divisions within statistics

There are many divisions, just to give you an idea; consider these few:

- Theoretical and Applied
  - This is very similar to saying; Mathematical and virtually everything else.
  - Mathematical statistics is the orthodoxy; we in the social sciences tend to be less rigid with regard to many topics and ways we go about conducting statistics (e.g., measurement, assumptions, etc.).
- Frequentist and Bayesian (more on this stuff later, RE: Probability)
  - Within the Frequentist paradigm: Fisherian vs. Neyman & Pearson interpretation.
  - We will stay within the Frequentist paradigm: $p(D|H)$.

RSS
Research and Statistical Support

Divisions within statistics

There are many divisions, just to give you an idea; consider these few:

- Theoretical and Applied
  - This is very similar to saying; Mathematical and virtually everything else.
  - Mathematical statistics is the orthodoxy; we in the social sciences tend to be less rigid with regard to many topics and ways we go about conducting statistics (e.g., measurement, assumptions, etc.).
- Frequentist and Bayesian (more on this stuff later, RE: Probability)
  - Within the Frequentist paradigm: Fisherian vs. Neyman & Pearson interpretation.
  - We will stay within the Frequentist paradigm: $p(D|H)$.
  - The Bayesian paradigm: $p(H|D)$.

RSS
Research and Statistical Support

Key terms 101

There will be several slides of key terms you should expect to know by heart. Do not simply memorize them, get to know them intimately[1].

[1] Memorization is often necessary but **not** sufficient for leaning or mastery to take place.

RSS Research and Statistical Support

Key terms 101

There will be several slides of key terms you should expect to know by heart. Do not simply memorize them, get to know them intimately[1].

- **Population**: The entire set of individuals the analyst is interested in studying.
  - Typically, Greek letters are used as symbols for population values; which themselves are called *population parameters*.

[1] Memorization is often necessary but **not** sufficient for leaning or mastery to take place.

RSS Research and Statistical Support

Key terms 101

There will be several slides of key terms you should expect to know by heart. Do not simply memorize them, get to know them intimately[1].

- **Population**: The entire set of individuals the analyst is interested in studying.
  - Typically, Greek letters are used as symbols for population values; which themselves are called *population parameters*.
- **Sample**: The subset of individuals, from the population, which the analyst can analyze.
  - Typically, italicized English letters are used as symbols for sample values; which themselves are called *sample statistics*.

[1] Memorization is often necessary but **not** sufficient for leaning or mastery to take place.

RSS Research and Statistical Support

Key terms 101

There will be several slides of key terms you should expect to know by heart. Do not simply memorize them, get to know them intimately[1].

- **Population**: The entire set of individuals the analyst is interested in studying.
  - Typically, Greek letters are used as symbols for population values; which themselves are called *population parameters*.
- **Sample**: The subset of individuals, from the population, which the analyst can analyze.
  - Typically, italicized English letters are used as symbols for sample values; which themselves are called *sample statistics*.

More on “Individuals” and the analyst’s role on the next slide.

[1] Memorization is often necessary but **not** sufficient for leaning or mastery to take place.

RSS Research and Statistical Support

Key terms 101 continued

Keep in mind; The analyst defines “population”, “sample”, and “individuals” at the very beginning of a study.

Key terms 101 continued

Keep in mind; The analyst defines “population”, “sample”, and “individuals” at the very beginning of a study.

- “individuals” can refer to virtually any type of individuals:

Key terms 101 continued

Keep in mind; The analyst defines “population”, “sample”, and “individuals” at the very beginning of a study.

- “individuals” can refer to virtually any type of individuals:
  - persons, mice, horses, dogs, students, classrooms, departments, districts, colleges, states, countries, cars, drugs, etc.

RSS
Research and Statistical Support

Key terms 101 continued

Keep in mind; The analyst defines “population”, “sample”, and “individuals” at the very beginning of a study.

- “individuals” can refer to virtually any type of individuals:
  - persons, mice, horses, dogs, students, classrooms, departments, districts, colleges, states, countries, cars, drugs, etc.

One common distinction used when describing a population or a sample is:

RSS
Research and Statistical Support

Key terms 101 continued

Keep in mind; The analyst defines “population”, “sample”, and “individuals” at the very beginning of a study.

- “individuals” can refer to virtually any type of individuals:
  - persons, mice, horses, dogs, students, classrooms, departments, districts, colleges, states, countries, cars, drugs, etc.

One common distinction used when describing a population or a sample is:

- *subjects* tends to refer to non-humans, while *participants* tends to refer to humans.

RSS
Research and Statistical Support

Key terms 101 continued

Keep in mind; The analyst defines “population”, “sample”, and “individuals” at the very beginning of a study.

- “individuals” can refer to virtually any type of individuals:
  - persons, mice, horses, dogs, students, classrooms, departments, districts, colleges, states, countries, cars, drugs, etc.

One common distinction used when describing a population or a sample is:

- *subjects* tends to refer to non-humans, while *participants* tends to refer to humans.

Did you catch our narrow definition of “statistics” on the previous slide? If not; don’t worry, it will come up again and again.

RSS
Research and Statistical Support

Key terms 101 continued

Population values (parameters) are typically unknown.

Key terms 101 continued

Population values (parameters) are typically unknown.

- We do not know the yearly salary of all professors; because, it would be impractical to collect that information from **all** professors (e.g., the information would likely be out of date by the time we finished collecting it).

Key terms 101 continued

Population values (parameters) are typically unknown.

- We do not know the yearly salary of all professors; because, it would be impractical to collect that information from **all** professors (e.g., the information would likely be out of date by the time we finished collecting it).
- Here, the population is understood to be all professors;

RSS
Research and Statistical Support

Key terms 101 continued

Population values (parameters) are typically unknown.

- We do not know the yearly salary of all professors; because, it would be impractical to collect that information from **all** professors (e.g., the information would likely be out of date by the time we finished collecting it).
- Here, the population is understood to be all professors;
  - Across **all** colleges, universities, countries...etc. That's a heck of a lot of folks!

RSS
Research and Statistical Support

Key terms 101 continued

Population values (parameters) are typically unknown.

- We do not know the yearly salary of all professors; because, it would be impractical to collect that information from **all** professors (e.g., the information would likely be out of date by the time we finished collecting it).
- Here, the population is understood to be all professors;
  - Across **all** colleges, universities, countries...etc. That's a heck of a lot of folks!
- Perhaps we should more carefully and narrowly define our population:

RSS
Research and Statistical Support

Key terms 101 continued

Population values (parameters) are typically unknown.

- We do not know the yearly salary of all professors; because, it would be impractical to collect that information from **all** professors (e.g., the information would likely be out of date by the time we finished collecting it).
- Here, the population is understood to be all professors;
  - Across **all** colleges, universities, countries...etc. That's a heck of a lot of folks!
- Perhaps we should more carefully and narrowly define our population:
  - Tenured professors teaching at public universities in the continental United States.

RSS
Research and Statistical Support

Key terms 101 continued

Because it would be impractical to assess our population directly, we tend to conduct analyses on samples with the understanding that our sample is representative of the defined population.

Key terms 101 continued

Because it would be impractical to assess our population directly, we tend to conduct analyses on samples with the understanding that our sample is representative of the defined population.

- We use a few terms to describe the representativeness of a sample:

Key terms 101 continued

Because it would be impractical to assess our population directly, we tend to conduct analyses on samples with the understanding that our sample is representative of the defined population.

- We use a few terms to describe the representativeness of a sample:
  - **Generalizability** is used interchangeably with representativeness; do our sample results generalize to the population?

RSS
Research and Statistical Support

Key terms 101 continued

Because it would be impractical to assess our population directly, we tend to conduct analyses on samples with the understanding that our sample is representative of the defined population.

- We use a few terms to describe the representativeness of a sample:
  - **Generalizability** is used interchangeably with representativeness; do our sample results generalize to the population?
  - **External Validity** refers to whether or not our sample is valid for drawing conclusions about the population.

RSS
Research and Statistical Support

Key terms 101 continued

Because it would be impractical to assess our population directly, we tend to conduct analyses on samples with the understanding that our sample is representative of the defined population.

- We use a few terms to describe the representativeness of a sample:
  - **Generalizability** is used interchangeably with representativeness; do our sample results generalize to the population?
  - **External Validity** refers to whether or not our sample is valid for drawing conclusions about the population.
- All three terms are used to describe how well a sample result can be applied to the larger population from which it came.

RSS
Research and Statistical Support

Key terms 101 continued

A **convenience sample** is one which was readily at hand or convenient to collect. Convenience samples tend to be less **generalizable** or less **externally valid**. Meaning, convenience samples tend to be less representative of their populations.

RSS
Research and Statistical Support

Key terms 101 continued

A **convenience sample** is one which was readily at hand or convenient to collect. Convenience samples tend to be less **generalizable** or less **externally valid**. Meaning, convenience samples tend to be less representative of their populations.

Example

```
Tenured professors at UNT
```

RSS
Research and Statistical Support

Key terms 101 continued

A **convenience sample** is one which was readily at hand or convenient to collect. Convenience samples tend to be less **generalizable** or less **externally valid**. Meaning, convenience samples tend to be less representative of their populations.

Example

```
Tenured professors at UNT
```

A **random sample** is one in which each individual (in the population) has an equal chance of being chosen (for inclusion in the sample). Random samples tend to be more representative of their populations.

RSS
Research and Statistical Support

Key terms 101 continued

A **convenience sample** is one which was readily at hand or convenient to collect. Convenience samples tend to be less **generalizable** or less **externally valid**. Meaning, convenience samples tend to be less representative of their populations.

Example

```
Tenured professors at UNT
```

A **random sample** is one in which each individual (in the population) has an equal chance of being chosen (for inclusion in the sample). Random samples tend to be more representative of their populations.

Example

```
Randomly selected tenured professors from
public universities in the continental United
States.
```

The importance of samples and populations.

There are two types of statistics we will deal with here.

The importance of samples and populations.

There are two types of statistics we will deal with here.

- **Descriptive Statistics** are used to summarize and make understandable–to describe–a group of numbers from a sample.

RSS
Research and Statistical Support

The importance of samples and populations.

There are two types of statistics we will deal with here.

- **Descriptive Statistics** are used to summarize and make understandable–to describe–a group of numbers from a sample.
- **Inferential Statistics** are used to draw conclusions and make inferences based on the numbers from a sample. These inferences may extend beyond the study which produced them.

RSS
Research and Statistical Support

The importance of samples and populations.

There are two types of statistics we will deal with here.

- **Descriptive Statistics** are used to summarize and make understandable–to describe–a group of numbers from a sample.
- **Inferential Statistics** are used to draw conclusions and make inferences based on the numbers from a sample. These inferences may extend beyond the study which produced them.

Recall the goals of science; Description, Explanation, and Prediction.

- Descriptive statistics will be a small, but essential, part of what we do here. Description is only the 1st goal of science.

RSS
Research and Statistical Support

The importance of samples and populations.

There are two types of statistics we will deal with here.

- **Descriptive Statistics** are used to summarize and make understandable–to describe–a group of numbers from a sample.
- **Inferential Statistics** are used to draw conclusions and make inferences based on the numbers from a sample. These inferences may extend beyond the study which produced them.

Recall the goals of science; Description, Explanation, and Prediction.

- Descriptive statistics will be a small, but essential, part of what we do here. Description is only the 1st goal of science.
- Inferential statistics will be the majority of what we do here; allowing us to explain and predict; the 2nd and 3rd goals of science.

RSS
Research and Statistical Support

Basic Concepts

- **Variable**: Condition, phenomena, or characteristic that can have different values; virtually anything can be a variable and there are many specific kinds of variables in research design and statistical analysis.

RSS
Research and Statistical Support

Basic Concepts

- **Variable**: Condition, phenomena, or characteristic that can have different values; virtually anything can be a variable and there are many specific kinds of variables in research design and statistical analysis.
- **Value**: Number indicating magnitude or category of a variable.

RSS
Research and Statistical Support

Basic Concepts

- **Variable**: Condition, phenomena, or characteristic that can have different values; virtually anything can be a variable and there are many specific kinds of variables in research design and statistical analysis.
- **Value**: Number indicating magnitude or category of a variable.
- **Score**: A particular individual's value on a variable.

RSS
Research and Statistical Support

Types of Variables: Variable Characteristics

Some variables are specifically named because of the characteristics they contain.

Types of Variables: Variable Characteristics

Some variables are specifically named because of the characteristics they contain.

- **Discrete** or **categorical** variables have a limited or finite number of possible values.

Types of Variables: Variable Characteristics

Some variables are specifically named because of the characteristics they contain.

- **Discrete** or **categorical** variables have a limited or finite number of possible values. Gender

Types of Variables: Variable Characteristics

Some variables are specifically named because of the characteristics they contain.

- **Discrete** or **categorical** variables have a limited or finite number of possible values. Gender
- **Continuous** variables theoretically have an infinite number of possible values.

RSS
Research and Statistical Support

Types of Variables: Variable Characteristics

Some variables are specifically named because of the characteristics they contain.

- **Discrete** or **categorical** variables have a limited or finite number of possible values. Gender
- **Continuous** variables theoretically have an infinite number of possible values. 100ths of a second reaction time

RSS
Research and Statistical Support

Types of Variables: Variable Characteristics

Some variables are specifically named because of the characteristics they contain.

- **Discrete** or **categorical** variables have a limited or finite number of possible values. Gender
- **Continuous** variables theoretically have an infinite number of possible values. 100ths of a second reaction time
- Often in the social sciences, variables are treated as continuous when we realize they truly are not capable of having an infinite number of possible values (e.g., human height).

RSS
Research and Statistical Support

Types of Variables: Variable Uses

Some variables are specifically named because of the way they are used in a study or analysis.

Types of Variables: Variable Uses

Some variables are specifically named because of the way they are used in a study or analysis.

- **Independent variables** are those which are manipulated by the researcher.

Types of Variables: Variable Uses

Some variables are specifically named because of the way they are used in a study or analysis.

- **Independent variables** are those which are manipulated by the researcher.
  - Also called: input variables or predictor variables.

RSS
Research and Statistical Support

Types of Variables: Variable Uses

Some variables are specifically named because of the way they are used in a study or analysis.

- **Independent variables** are those which are manipulated by the researcher.
  - Also called: input variables or predictor variables.
  - Sometimes, the term grouping variable is used for independent variables which are not truly manipulated by the researcher, but are the input variables (e.g., gender).

RSS
Research and Statistical Support

Types of Variables: Variable Uses

Some variables are specifically named because of the way they are used in a study or analysis.

- **Independent variables** are those which are manipulated by the researcher.
  - Also called: input variables or predictor variables.
  - Sometimes, the term grouping variable is used for independent variables which are not truly manipulated by the researcher, but are the input variables (e.g., gender).
- **Dependent variables** are those used to measure change in the independent variable(s).

RSS
Research and Statistical Support

Types of Variables: Variable Uses

Some variables are specifically named because of the way they are used in a study or analysis.

- **Independent variables** are those which are manipulated by the researcher.
  - Also called: input variables or predictor variables.
  - Sometimes, the term grouping variable is used for independent variables which are not truly manipulated by the researcher, but are the input variables (e.g., gender).
- **Dependent variables** are those used to measure change in the independent variable(s).
  - Also called: output variables or outcome variables.

RSS
Research and Statistical Support

Types of Variables: Variable Values

Variables are also classified by the content of what they express. This is often true in the context of statistical software.

Types of Variables: Variable Values

Variables are also classified by the content of what they express. This is often true in the context of statistical software.

- **Numeric** variables are expressed with numbers; meaning their values are numbers (e.g., age in years).

RSS
Research and Statistical Support

Types of Variables: Variable Values

Variables are also classified by the content of what they express. This is often true in the context of statistical software.

- **Numeric** variables are expressed with numbers; meaning their values are numbers (e.g., age in years).
- **String** variables are typically expressed in letters or words (e.g., birth state).

RSS
Research and Statistical Support

Types of Variables: Variable Values

Variables are also classified by the content of what they express. This is often true in the context of statistical software.

- **Numeric** variables are expressed with numbers; meaning their values are numbers (e.g., age in years).
- **String** variables are typically expressed in letters or words (e.g., birth state).
- **Currency** variables are expressed in a particular monetary type (e.g., Euros, Dollars, etc.)

RSS
Research and Statistical Support

Types of Variables: Variable Values

Variables are also classified by the content of what they express. This is often true in the context of statistical software.

- **Numeric** variables are expressed with numbers; meaning their values are numbers (e.g., age in years).
- **String** variables are typically expressed in letters or words (e.g., birth state).
- **Currency** variables are expressed in a particular monetary type (e.g., Euros, Dollars, etc.)
- **Date** variables can be expressed in a variety of ways (e.g., 09/14/2010, 14Sep10, 14/09/10, etc.).

RSS
Research and Statistical Support

Types of Variables: Variable Values

Variables are also classified by the content of what they express. This is often true in the context of statistical software.

- **Numeric** variables are expressed with numbers; meaning their values are numbers (e.g., age in years).
- **String** variables are typically expressed in letters or words (e.g., birth state).
- **Currency** variables are expressed in a particular monetary type (e.g., Euros, Dollars, etc.)
- **Date** variables can be expressed in a variety of ways (e.g., 09/14/2010, 14Sep10, 14/09/10, etc.).

You can see how some of these might overlap (e.g., a numeric variable might also be a currency variable).

RSS
Research and Statistical Support

Measurement Scales

There are four **Measurement Scales** or *levels of measurement*.

Measurement Scales

There are four **Measurement Scales** or *levels of measurement*.

- Nominal: Naming things with numbers.

Measurement Scales

There are four **Measurement Scales** or *levels of measurement*.

- Nominal: Naming things with numbers.
- Ordinal: Numbers have sequential meaning.

RSS
Research and Statistical Support

Measurement Scales

There are four **Measurement Scales** or *levels of measurement*.

- Nominal: Naming things with numbers.
- Ordinal: Numbers have sequential meaning.
- Interval: Distances between units are equal.

RSS
Research and Statistical Support

Measurement Scales

There are four **Measurement Scales** or *levels of measurement*.

- Nominal: Naming things with numbers.
- Ordinal: Numbers have sequential meaning.
- Interval: Distances between units are equal.
- Ratio: there is a 'true zero' indicating an absence of the variable.

RSS
Research and Statistical Support

Nominal Scale

The nominal scale simply uses numbers to name objects.

Nominal Scale

The nominal scale simply uses numbers to name objects.

- Thoroughbreds' race numbers have no purpose other than identifying each horse in a race.

Nominal Scale

The nominal scale simply uses numbers to name objects.

- Thoroughbreds' race numbers have no purpose other than identifying each horse in a race.

This can be applied to several types of variables already mentioned using **coding**.

Nominal Scale

The nominal scale simply uses numbers to name objects.

- Thoroughbreds' race numbers have no purpose other than identifying each horse in a race.

This can be applied to several types of variables already mentioned using **coding**.

- Categorical variables such as Birth State or Class Standing can be coded by assigning values (*numbers*) to value labels (*categories*):

RSS
Research and Statistical Support

Nominal Scale

The nominal scale simply uses numbers to name objects.

- Thoroughbreds' race numbers have no purpose other than identifying each horse in a race.

This can be applied to several types of variables already mentioned using **coding**.

- Categorical variables such as Birth State or Class Standing can be coded by assigning values (*numbers*) to value labels (*categories*):
  - 1 = Alabama, 2 = Alaska, ... 50 = Wyoming.

RSS
Research and Statistical Support

Nominal Scale

The nominal scale simply uses numbers to name objects.

- Thoroughbreds' race numbers have no purpose other than identifying each horse in a race.

This can be applied to several types of variables already mentioned using **coding**.

- Categorical variables such as Birth State or Class Standing can be coded by assigning values (*numbers*) to value labels (*categories*):
  - 1 = Alabama, 2 = Alaska, ... 50 = Wyoming.
  - 1 = Freshman, 2 = Sophomore, 3 = Junior, 4 = Senior.

RSS
Research and Statistical Support

Nominal Scale

The nominal scale simply uses numbers to name objects.

- Thoroughbreds' race numbers have no purpose other than identifying each horse in a race.

This can be applied to several types of variables already mentioned using **coding**.

- Categorical variables such as Birth State or Class Standing can be coded by assigning values (*numbers*) to value labels (*categories*):
  - 1 = Alabama, 2 = Alaska, ... 50 = Wyoming.
  - 1 = Freshman, 2 = Sophomore, 3 = Junior, 4 = Senior.
- It is important to note how the numbers (or codes) are assigned. Those above, could just as easily be:

RSS
Research and Statistical Support

Nominal Scale

The nominal scale simply uses numbers to name objects.

- Thoroughbreds' race numbers have no purpose other than identifying each horse in a race.

This can be applied to several types of variables already mentioned using **coding**.

- Categorical variables such as Birth State or Class Standing can be coded by assigning values (*numbers*) to value labels (*categories*):
  - 1 = Alabama, 2 = Alaska, ... 50 = Wyoming.
  - 1 = Freshman, 2 = Sophomore, 3 = Junior, 4 = Senior.
- It is important to note how the numbers (or codes) are assigned. Those above, could just as easily be:
  - 4 = Freshman, 3 = Sophomore, 2 = Junior, 1 = Senior.

RSS
Research and Statistical Support

Ordinal Scale

The ordinal scale has the added property of sequence. The numbers' sequential order has meaning.

Ordinal Scale

The ordinal scale has the added property of sequence. The numbers' sequential order has meaning.

- Finishing positions of the horses in a race or rats in a maze.

RSS
Research and Statistical Support

Ordinal Scale

The ordinal scale has the added property of sequence. The numbers' sequential order has meaning.

- Finishing positions of the horses in a race or rats in a maze.
- Birth order among siblings.

RSS
Research and Statistical Support

Ordinal Scale

The ordinal scale has the added property of sequence. The numbers' sequential order has meaning.

- Finishing positions of the horses in a race or rats in a maze.
- Birth order among siblings.
- Preference ratings of similar products

RSS
Research and Statistical Support

Ordinal Scale

The ordinal scale has the added property of sequence. The numbers' sequential order has meaning.

- Finishing positions of the horses in a race or rats in a maze.
- Birth order among siblings.
- Preference ratings of similar products
  - I prefer Diet Coke (1) over Diet Pepsi (2) and both over Diet Mountain Dew (3).

RSS
Research and Statistical Support

Ordinal Scale

The ordinal scale has the added property of sequence. The numbers' sequential order has meaning.

- Finishing positions of the horses in a race or rats in a maze.
- Birth order among siblings.
- Preference ratings of similar products
  - I prefer Diet Coke (1) over Diet Pepsi (2) and both over Diet Mountain Dew (3).
- As with Nominal, it is important to note how the numbers (e.g., ratings) are assigned. Those above, could be:

RSS
Research and Statistical Support

Ordinal Scale

The ordinal scale has the added property of sequence. The numbers' sequential order has meaning.

- Finishing positions of the horses in a race or rats in a maze.
- Birth order among siblings.
- Preference ratings of similar products
  - I prefer Diet Coke (1) over Diet Pepsi (2) and both over Diet Mountain Dew (3).
- As with Nominal, it is important to note how the numbers (e.g., ratings) are assigned. Those above, could be:
  - I prefer Diet Coke (3) over Diet Pepsi (2) and both over Diet Mountain Dew (1).

RSS
Research and Statistical Support

Interval Scale

Interval scale has the additional property of equal intervals between units of measure.

Interval Scale

Interval scale has the additional property of equal intervals between units of measure.

- Time of day or clock time.

Interval Scale

Interval scale has the additional property of equal intervals between units of measure.

- Time of day or clock time.
- The numbers identify objects or points, convey sequence, and there are equal intervals between the units.

RSS
Research and Statistical Support

Interval Scale

Interval scale has the additional property of equal intervals between units of measure.

- Time of day or clock time.
- The numbers identify objects or points, convey sequence, and there are equal intervals between the units.
  - The interval between 1 o'clock and 2 o'clock is the same as between 4 o'clock and 5 o'clock (accept on Fridays).

RSS
Research and Statistical Support

Ratio Scale

Ratio scale has the additional property of a true zero point; a zero which represents an absence of magnitude on the variable.

Ratio Scale

Ratio scale has the additional property of a true zero point; a zero which represents an absence of magnitude on the variable.

- U.S. pounds of weight

Ratio Scale

Ratio scale has the additional property of a true zero point; a zero which represents an absence of magnitude on the variable.

- U.S. pounds of weight
  - An individual (human, non-human, or object) can not weight negative 140 lbs.

RSS
Research and Statistical Support

Ratio Scale

Ratio scale has the additional property of a true zero point; a zero which represents an absence of magnitude on the variable.

- U.S. pounds of weight
  - An individual (human, non-human, or object) can not weight negative 140 lbs.
- Kelvin temperature.

RSS
Research and Statistical Support

Ratio Scale

Ratio scale has the additional property of a true zero point; a zero which represents an absence of magnitude on the variable.

- U.S. pounds of weight
  - An individual (human, non-human, or object) can not weight negative 140 lbs.
- Kelvin temperature.
  - Zero degrees Kelvin means literally an absence of temperature.

RSS
Research and Statistical Support

Objective vs. Subjective Measurement

The “accept on Fridays” comment above is worth remembering. because it highlights another consideration in measurement: **Objective** vs. **Subjective**.

Objective vs. Subjective Measurement

The “accept on Fridays” comment above is worth remembering. because it highlights another consideration in measurement: **Objective** vs. **Subjective**.

- **Objective**: The clock is an example of objective measurement; if it is working correctly, it has no variance when measuring the interval between 4 o’clock and 5 o’clock on multiple days.

RSS
Research and Statistical Support

Objective vs. Subjective Measurement

The “accept on Fridays” comment above is worth remembering. because it highlights another consideration in measurement: **Objective** vs. **Subjective**.

- **Objective**: The clock is an example of objective measurement; if it is working correctly, it has no variance when measuring the interval between 4 o’clock and 5 o’clock on multiple days.
- **Subjective**: Our perception tends to be more subjective; our perception of the interval between 4 o’clock and 5 o’clock varies depending on the day.

RSS
Research and Statistical Support

Objective vs. Subjective Measurement

The “accept on Fridays” comment above is worth remembering. because it highlights another consideration in measurement: **Objective** vs. **Subjective**.

- **Objective**: The clock is an example of objective measurement; if it is working correctly, it has no variance when measuring the interval between 4 o’clock and 5 o’clock on multiple days.
- **Subjective**: Our perception tends to be more subjective; our perception of the interval between 4 o’clock and 5 o’clock varies depending on the day.
  - On Fridays, we perceive a longer interval between 4 and 5 o’clock compared to other days of the week because, we are looking forward to the weekend.

RSS
Research and Statistical Support

Direct vs. Indirect Measurement

Often in social science, we are interested in studying things we can not physically touch or observe.

Direct vs. Indirect Measurement

Often in social science, we are interested in studying things we can not physically touch or observe.

- **Direct** measurement is possible with observable, often called *manifest*, variables. Generally these are physical things, but not always.

RSS
Research and Statistical Support

Direct vs. Indirect Measurement

Often in social science, we are interested in studying things we can not physically touch or observe.

- **Direct** measurement is possible with observable, often called *manifest*, variables. Generally these are physical things, but not always.
  - Human Height, weight, hair length, eye color, skin color (or more precisely, pigmentation).

RSS
Research and Statistical Support

Direct vs. Indirect Measurement

Often in social science, we are interested in studying things we can not physically touch or observe.

- **Direct** measurement is possible with observable, often called *manifest*, variables. Generally these are physical things, but not always.
  - Human Height, weight, hair length, eye color, skin color (or more precisely, pigmentation).
  - Generally less measurement error.

RSS
Research and Statistical Support

Direct vs. Indirect Measurement

Often in social science, we are interested in studying things we can not physically touch or observe.

- **Direct** measurement is possible with observable, often called *manifest*, variables. Generally these are physical things, but not always.
  - Human Height, weight, hair length, eye color, skin color (or more precisely, pigmentation).
  - Generally less measurement error.
- **Indirect** measurement is generally done with unobservable, often called *latent*, variables.

RSS
Research and Statistical Support

Direct vs. Indirect Measurement

Often in social science, we are interested in studying things we can not physically touch or observe.

- **Direct** measurement is possible with observable, often called *manifest*, variables. Generally these are physical things, but not always.
  - Human Height, weight, hair length, eye color, skin color (or more precisely, pigmentation).
  - Generally less measurement error.
- **Indirect** measurement is generally done with unobservable, often called *latent*, variables.
  - Love, Extroversion, etc.

RSS
Research and Statistical Support

Direct vs. Indirect Measurement

Often in social science, we are interested in studying things we can not physically touch or observe.

- **Direct** measurement is possible with observable, often called *manifest*, variables. Generally these are physical things, but not always.
  - Human Height, weight, hair length, eye color, skin color (or more precisely, pigmentation).
  - Generally less measurement error.
- **Indirect** measurement is generally done with unobservable, often called *latent*, variables.
  - Love, Extroversion, etc.
  - Generally more measurement error.

RSS
Research and Statistical Support

Direct vs. Indirect Measurement

Often in social science, we are interested in studying things we can not physically touch or observe.

- **Direct** measurement is possible with observable, often called *manifest*, variables. Generally these are physical things, but not always.
  - Human Height, weight, hair length, eye color, skin color (or more precisely, pigmentation).
  - Generally less measurement error.
- **Indirect** measurement is generally done with unobservable, often called *latent*, variables.
  - Love, Extroversion, etc.
  - Generally more measurement error.

How would you define sadness? Drunkenness? Success in College?

RSS
Research and Statistical Support

Operational Definitions

Operation Definitions allow us to define variables with measurement. Think quantitatively. What is the quantity of this characteristic, phenomena, feature, behavior, emotion, etc.? Defining a variable operationally means defining it in such a way that description and observation are not the only benefits, but measurement as well.

RSS
Research and Statistical Support

Operational Definitions

Operation Definitions allow us to define variables with measurement. Think quantitatively. What is the quantity of this characteristic, phenomena, feature, behavior, emotion, etc.? Defining a variable operationally means defining it in such a way that description and observation are not the only benefits, but measurement as well.

- Sadness

RSS
Research and Statistical Support

Operational Definitions

Operation Definitions allow us to define variables with measurement. Think quantitatively. What is the quantity of this characteristic, phenomena, feature, behavior, emotion, etc.? Defining a variable operationally means defining it in such a way that description and observation are not the only benefits, but measurement as well.

- Sadness
  - Number of crying episodes in 5 days.

RSS
Research and Statistical Support

Operational Definitions

Operation Definitions allow us to define variables with measurement. Think quantitatively. What is the quantity of this characteristic, phenomena, feature, behavior, emotion, etc.? Defining a variable operationally means defining it in such a way that description and observation are not the only benefits, but measurement as well.

- Sadness
  - Number of crying episodes in 5 days.
  - Self-rating of intensity of crying episodes.

RSS
Research and Statistical Support

Operational Definitions

Operation Definitions allow us to define variables with measurement. Think quantitatively. What is the quantity of this characteristic, phenomena, feature, behavior, emotion, etc.? Defining a variable operationally means defining it in such a way that description and observation are not the only benefits, but measurement as well.

- Sadness
  - Number of crying episodes in 5 days.
  - Self-rating of intensity of crying episodes.
- Drunkenness

RSS
Research and Statistical Support

Operational Definitions

Operation Definitions allow us to define variables with measurement. Think quantitatively. What is the quantity of this characteristic, phenomena, feature, behavior, emotion, etc.? Defining a variable operationally means defining it in such a way that description and observation are not the only benefits, but measurement as well.

- Sadness
  - Number of crying episodes in 5 days.
  - Self-rating of intensity of crying episodes.
- Drunkenness
  - Number of slurred words in a 5 minute conversation.

RSS
Research and Statistical Support

Operational Definitions

Operation Definitions allow us to define variables with measurement. Think quantitatively. What is the quantity of this characteristic, phenomena, feature, behavior, emotion, etc.? Defining a variable operationally means defining it in such a way that description and observation are not the only benefits, but measurement as well.

- Sadness
  - Number of crying episodes in 5 days.
  - Self-rating of intensity of crying episodes.
- Drunkenness
  - Number of slurred words in a 5 minute conversation.
  - Blood Alcohol Level (ML/L).

RSS
Research and Statistical Support

Operational Definitions

Operation Definitions allow us to define variables with measurement. Think quantitatively. What is the quantity of this characteristic, phenomena, feature, behavior, emotion, etc.? Defining a variable operationally means defining it in such a way that description and observation are not the only benefits, but measurement as well.

- Sadness
  - Number of crying episodes in 5 days.
  - Self-rating of intensity of crying episodes.
- Drunkenness
  - Number of slurred words in a 5 minute conversation.
  - Blood Alcohol Level (ML/L).
- Success in College

RSS
Research and Statistical Support

Operational Definitions

Operation Definitions allow us to define variables with measurement. Think quantitatively. What is the quantity of this characteristic, phenomena, feature, behavior, emotion, etc.? Defining a variable operationally means defining it in such a way that description and observation are not the only benefits, but measurement as well.

- Sadness
  - Number of crying episodes in 5 days.
  - Self-rating of intensity of crying episodes.
- Drunkenness
  - Number of slurred words in a 5 minute conversation.
  - Blood Alcohol Level (ML/L).
- Success in College
  - Graduate or not

RSS
Research and Statistical Support

Operational Definitions

Operation Definitions allow us to define variables with measurement. Think quantitatively. What is the quantity of this characteristic, phenomena, feature, behavior, emotion, etc.? Defining a variable operationally means defining it in such a way that description and observation are not the only benefits, but measurement as well.

- Sadness
  - Number of crying episodes in 5 days.
  - Self-rating of intensity of crying episodes.
- Drunkenness
  - Number of slurred words in a 5 minute conversation.
  - Blood Alcohol Level (ML/L).
- Success in College
  - Graduate or not
  - Number of A's

RSS
Research and Statistical Support

Operational Definitions

Operation Definitions allow us to define variables with measurement. Think quantitatively. What is the quantity of this characteristic, phenomena, feature, behavior, emotion, etc.? Defining a variable operationally means defining it in such a way that description and observation are not the only benefits, but measurement as well.

- Sadness
  - Number of crying episodes in 5 days.
  - Self-rating of intensity of crying episodes.
- Drunkenness
  - Number of slurred words in a 5 minute conversation.
  - Blood Alcohol Level (ML/L).
- Success in College
  - Graduate or not
  - Number of A's
  - Grade Point Average

RSS
Research and Statistical Support

Summary of Module 1

Module 1 covered the following topics:

- The search for truth and the tenants, as well as goals, of science.

Summary of Module 1

Module 1 covered the following topics:

- The search for truth and the tenants, as well as goals, of science.
- Some core terms, such as the distinction and importance of populations and samples.

RSS
Research and Statistical Support

Summary of Module 1

Module 1 covered the following topics:

- The search for truth and the tenants, as well as goals, of science.
- Some core terms, such as the distinction and importance of populations and samples.
- Other key terms related to the definition of variables.

RSS
Research and Statistical Support

Summary of Module 1

Module 1 covered the following topics:

- The search for truth and the tenants, as well as goals, of science.
- Some core terms, such as the distinction and importance of populations and samples.
- Other key terms related to the definition of variables.
- Some principles of measurement.

RSS
Research and Statistical Support

Summary of Module 1

Module 1 covered the following topics:

- The search for truth and the tenants, as well as goals, of science.
- Some core terms, such as the distinction and importance of populations and samples.
- Other key terms related to the definition of variables.
- Some principles of measurement.
  - Nominal Scales

RSS
Research and Statistical Support

Summary of Module 1

Module 1 covered the following topics:

- The search for truth and the tenants, as well as goals, of science.
- Some core terms, such as the distinction and importance of populations and samples.
- Other key terms related to the definition of variables.
- Some principles of measurement.
  - Nominal Scales
  - Ordinal Scales

RSS
Research and Statistical Support

Summary of Module 1

Module 1 covered the following topics:

- The search for truth and the tenants, as well as goals, of science.
- Some core terms, such as the distinction and importance of populations and samples.
- Other key terms related to the definition of variables.
- Some principles of measurement.
  - Nominal Scales
  - Ordinal Scales
  - Interval Scales

RSS
Research and Statistical Support

Summary of Module 1

Module 1 covered the following topics:

- The search for truth and the tenants, as well as goals, of science.
- Some core terms, such as the distinction and importance of populations and samples.
- Other key terms related to the definition of variables.
- Some principles of measurement.
  - Nominal Scales
  - Ordinal Scales
  - Interval Scales
  - Ration Scales

RSS
Research and Statistical Support

Summary of Module 1

Module 1 covered the following topics:

- The search for truth and the tenants, as well as goals, of science.
- Some core terms, such as the distinction and importance of populations and samples.
- Other key terms related to the definition of variables.
- Some principles of measurement.
  - Nominal Scales
  - Ordinal Scales
  - Interval Scales
  - Ration Scales
  - Additional considerations of measurement/definition of variables.

RSS
Research and Statistical Support

This concludes Module 1

Next time Module 2.

- Next time we'll begin covering how to display data.
- Until next time; have a nice day.

These slides initially created on: September 13, 2010
These slides last updated on: September 23, 2010

- The bottom date shown is the date this Adobe.pdf file was created; LaTeX[2] has a command for automatically inserting the date of a document's creation.

RSS
Research and Statistical Support

[2] This document was created in LaTeX using the Beamer package