

Introduction to Statistics

Professor Timothy Halliday

August 24, 2009

Course Info

- Office hours: MW 8:00 to 9:15

Course Info

- Office hours: MW 8:00 to 9:15
- Text Book: Elementary Statistics 11th Edition

Course Info

- Office hours: MW 8:00 to 9:15
- Text Book: Elementary Statistics 11th Edition
- URL: www2.hawaii.edu/~halliday

Course Info

- Office hours: MW 8:00 to 9:15
- Text Book: Elementary Statistics 11th Edition
- URL: www2.hawaii.edu/~halliday
- E-Mail: halliday@hawaii.edu

Course Info

- Office hours: MW 8:00 to 9:15
- Text Book: Elementary Statistics 11th Edition
- URL: www2.hawaii.edu/~halliday
- E-Mail: halliday@hawaii.edu
- TA: Huixia Wang (huixia@hawaii.edu)

Course Info

- Syllabus is online.

Course Info

- Syllabus is online.
- The calendar might change.

Course Info

- Syllabus is online.
- The calendar might change.
- Fridays will be reserved for quizzes and problem sets.

Course Info

- Syllabus is online.
- The calendar might change.
- Fridays will be reserved for quizzes and problem sets.
- Problem sets will not be graded but their solutions will be covered by the TA.

Course Info

- Syllabus is online.
- The calendar might change.
- Fridays will be reserved for quizzes and problem sets.
- Problem sets will not be graded but their solutions will be covered by the TA.
- Quizzes will be based in large part on the PS.

Course Info

- Syllabus is online.
- The calendar might change.
- Fridays will be reserved for quizzes and problem sets.
- Problem sets will not be graded but their solutions will be covered by the TA.
- Quizzes will be based in large part on the PS.
- If you do the PS's you will do well. If you do not, you will not.

Course Info

- Emphasis of course is twofold

Course Info

- Emphasis of course is twofold
 - Conceptual and Theoretical

Course Info

- Emphasis of course is twofold
 - Conceptual and Theoretical
 - Monday and Wednesday Lectures

Course Info

- Emphasis of course is twofold
 - Conceptual and Theoretical
 - Monday and Wednesday Lectures
 - Professor

Course Info

- Emphasis of course is twofold
 - Conceptual and Theoretical
 - Monday and Wednesday Lectures
 - Professor
 - Problem Solving

Course Info

- Emphasis of course is twofold
 - Conceptual and Theoretical
 - Monday and Wednesday Lectures
 - Professor
 - Problem Solving
 - Friday Lectures

Course Info

- Emphasis of course is twofold
 - Conceptual and Theoretical
 - Monday and Wednesday Lectures
 - Professor
 - Problem Solving
 - Friday Lectures
 - TA

Course Info

- Emphasis of course is twofold
 - Conceptual and Theoretical
 - Monday and Wednesday Lectures
 - Professor
 - Problem Solving
 - Friday Lectures
 - TA
- Always e-mail the TA a question before e-mailing me.

Grading

- 60% of the grade is quizzes.

Grading

- 60% of the grade is quizzes.
- 40% of the grade is the final.

Grading

- 60% of the grade is quizzes.
- 40% of the grade is the final.
- If you do not come to the Friday classes, you will probably fail.

Grading

- 60% of the grade is quizzes.
- 40% of the grade is the final.
- If you do not come to the Friday classes, you will probably fail.
- 2 lowest quizzes will be dropped.

What is Statistics?

Statistics is a branch of mathematics devoted to the collection, organization and interpretation of data with the intent of making *inferences* about *population parameters* using *random sampling*

What is Statistics?

- In plain English, what does this mean?

What is Statistics?

- In plain English, what does this mean?
- First, we must define a population.

What is Statistics?

- In plain English, what does this mean?
- First, we must define a population.
- Let's use the population of the United States - that is to say all 300 Million of us.

What is Statistics?

- In plain English, what does this mean?
- First, we must define a population.
- Let's use the population of the United States - that is to say all 300 Million of us.
- Next, we must have a parameter - or a population characteristic - in mind.

What is Statistics?

- In plain English, what does this mean?
- First, we must define a population.
- Let's use the population of the United States - that is to say all 300 Million of us.
- Next, we must have a parameter - or a population characteristic - in mind.
- Let's use the unemployment rate - the percentage of the labor force that is currently unemployed.

What is Statistics?

- How do we compute this?

What is Statistics?

- How do we compute this?
- One way is to survey all Americans in the labor force and inquire about their employment status - this is called a **Census**.

What is Statistics?

- How do we compute this?
- One way is to survey all Americans in the labor force and inquire about their employment status - this is called a **Census**.
- What are the potential issues here?

What is Statistics?

- How do we compute this?
- One way is to survey all Americans in the labor force and inquire about their employment status - this is called a **Census**.
- What are the potential issues here?
 - It is expensive!!!!

What is Statistics?

- How do we compute this?
- One way is to survey all Americans in the labor force and inquire about their employment status - this is called a **Census**.
- What are the potential issues here?
 - It is expensive!!!!
- A solution is to survey a smaller subset of the population.

What is Statistics?

- How do we compute this?
- One way is to survey all Americans in the labor force and inquire about their employment status - this is called a **Census**.
- What are the potential issues here?
 - It is expensive!!!!
- A solution is to survey a smaller subset of the population.
- Desirable characteristic of the survey

What is Statistics?

- How do we compute this?
- One way is to survey all Americans in the labor force and inquire about their employment status - this is called a **Census**.
- What are the potential issues here?
 - It is expensive!!!!
- A solution is to survey a smaller subset of the population.
- Desirable characteristic of the survey
 - Must be representative.

What is Statistics?

- How do we compute this?
- One way is to survey all Americans in the labor force and inquire about their employment status - this is called a **Census**.
- What are the potential issues here?
 - It is expensive!!!!
- A solution is to survey a smaller subset of the population.
- Desirable characteristic of the survey
 - Must be representative.
 - Must use Random Sampling.

What is Statistics?

- The unemployment rate in the sample is a statistic.

What is Statistics?

- The unemployment rate in the sample is a statistic.
- A statistic is a random function of the data.

What is Statistics?

- The unemployment rate in the sample is a statistic.
- A statistic is a random function of the data.
- We use the statistic to make an inference about the population parameter.

What is Statistics?

- The unemployment rate in the sample is a statistic.
- A statistic is a random function of the data.
- We use the statistic to make an inference about the population parameter.
- The statistic will not equal the population parameter.

What is Statistics?

- The unemployment rate in the sample is a statistic.
- A statistic is a random function of the data.
- We use the statistic to make an inference about the population parameter.
- The statistic will not equal the population parameter.
 - **Sampling Error**

What is Statistics?

- The unemployment rate in the sample is a statistic.
- A statistic is a random function of the data.
- We use the statistic to make an inference about the population parameter.
- The statistic will not equal the population parameter.
 - **Sampling Error**
 - Because of this, we will learn how to make statements like, "With probability equal to 95%, the true unemployment rate is between 7.8% and 8.1%."

Sampling

- There are a variety of ways to sample from a population.

Sampling

- There are a variety of ways to sample from a population.
 - **Random Sampling**

Sampling

- There are a variety of ways to sample from a population.
 - **Random Sampling**
 - Each individual has an equal chance of being selected.

Sampling

- There are a variety of ways to sample from a population.
 - **Random Sampling**
 - Each individual has an equal chance of being selected.
 - **Stratified Sample**

Sampling

- There are a variety of ways to sample from a population.
 - **Random Sampling**
 - Each individual has an equal chance of being selected.
 - **Stratified Sample**
 - Divide the population into *strata* (e.g. races, HIV +, HIV-, etc.) and draw a random sample within each strata

Sampling

- There are a variety of ways to sample from a population.
 - **Random Sampling**
 - Each individual has an equal chance of being selected.
 - **Stratified Sample**
 - Divide the population into *strata* (e.g. races, HIV +, HIV-, etc.) and draw a random sample within each strata
 - **Cluster Sample**

Sampling

- There are a variety of ways to sample from a population.
 - **Random Sampling**
 - Each individual has an equal chance of being selected.
 - **Stratified Sample**
 - Divide the population into *strata* (e.g. races, HIV +, HIV-, etc.) and draw a random sample within each strata
 - **Cluster Sample**
 - Divide population into *clusters* (e.g. villages, neighborhoods, etc.), randomly select the clusters and then sample all the people within the cluster

Sampling

- There are a variety of ways to sample from a population.
 - **Random Sampling**
 - Each individual has an equal chance of being selected.
 - **Stratified Sample**
 - Divide the population into *strata* (e.g. races, HIV +, HIV-, etc.) and draw a random sample within each strata
 - **Cluster Sample**
 - Divide population into *clusters* (e.g. villages, neighborhoods, etc.), randomly select the clusters and then sample all the people within the cluster
- Random sampling is the most straight-forward.

Sampling

- There are a variety of ways to sample from a population.
 - **Random Sampling**
 - Each individual has an equal chance of being selected.
 - **Stratified Sample**
 - Divide the population into *strata* (e.g. races, HIV +, HIV-, etc.) and draw a random sample within each strata
 - **Cluster Sample**
 - Divide population into *clusters* (e.g. villages, neighborhoods, etc.), randomly select the clusters and then sample all the people within the cluster
- Random sampling is the most straight-forward.
- Other sampling schemes require adjustment to conduct inference.

Sampling

- When is sampling used?

Sampling

- When is sampling used?
- Most economic data come from random samples.

Sampling

- When is sampling used?
- Most economic data come from random samples.
 - Unemployment rates

Sampling

- When is sampling used?
- Most economic data come from random samples.
 - Unemployment rates
 - Rates of uninsured

Sampling

- When is sampling used?
- Most economic data come from random samples.
 - Unemployment rates
 - Rates of uninsured
 - Inequality Measures

Sampling

- When is sampling used?
- Most economic data come from random samples.
 - Unemployment rates
 - Rates of uninsured
 - Inequality Measures
 - Poverty Measures

Sampling

- When is sampling used?
- Most economic data come from random samples.
 - Unemployment rates
 - Rates of uninsured
 - Inequality Measures
 - Poverty Measures
- Sampling is also used in other disciplines

Sampling

- When is sampling used?
- Most economic data come from random samples.
 - Unemployment rates
 - Rates of uninsured
 - Inequality Measures
 - Poverty Measures
- Sampling is also used in other disciplines
 - Astronomy

Sampling

- When is sampling used?
- Most economic data come from random samples.
 - Unemployment rates
 - Rates of uninsured
 - Inequality Measures
 - Poverty Measures
- Sampling is also used in other disciplines
 - Astronomy
 - Medicine/Public Health

Sampling

- When is sampling used?
- Most economic data come from random samples.
 - Unemployment rates
 - Rates of uninsured
 - Inequality Measures
 - Poverty Measures
- Sampling is also used in other disciplines
 - Astronomy
 - Medicine/Public Health
 - Political Science (Polling)

Sampling

- When is sampling used?
- Most economic data come from random samples.
 - Unemployment rates
 - Rates of uninsured
 - Inequality Measures
 - Poverty Measures
- Sampling is also used in other disciplines
 - Astronomy
 - Medicine/Public Health
 - Political Science (Polling)
 - Demography

Statistical Abuses

- Correlation is not causality

Statistical Abuses

- Correlation is not causality
 - Europeans have universal coverage; Americans do not.

Statistical Abuses

- Correlation is not causality
 - Europeans have universal coverage; Americans do not.
 - Europeans live longer than Americans.

Statistical Abuses

- Correlation is not causality
 - Europeans have universal coverage; Americans do not.
 - Europeans live longer than Americans.
 - Americans do not live longer because they do not have universal coverage.

Statistical Abuses

- Correlation is not causality
 - Europeans have universal coverage; Americans do not.
 - Europeans live longer than Americans.
 - Americans do not live longer because they do not have universal coverage.
 - There are other factors at play e.g. obesity and other public health issues that are beyond the scope of the medical system.

Statistical Abuses

- Correlation is not causality
 - Europeans have universal coverage; Americans do not.
 - Europeans live longer than Americans.
 - Americans do not live longer because they do not have universal coverage.
 - There are other factors at play e.g. obesity and other public health issues that are beyond the scope of the medical system.
- Statistics adds a veneer of legitimacy to bad science.

Statistical Abuses

- Correlation is not causality
 - Europeans have universal coverage; Americans do not.
 - Europeans live longer than Americans.
 - Americans do not live longer because they do not have universal coverage.
 - There are other factors at play e.g. obesity and other public health issues that are beyond the scope of the medical system.
- Statistics adds a veneer of legitimacy to bad science.
 - Some claimed that the MMR vaccine causes autism.

Statistical Abuses

- Correlation is not causality
 - Europeans have universal coverage; Americans do not.
 - Europeans live longer than Americans.
 - Americans do not live longer because they do not have universal coverage.
 - There are other factors at play e.g. obesity and other public health issues that are beyond the scope of the medical system.
- Statistics adds a veneer of legitimacy to bad science.
 - Some claimed that the MMR vaccine causes autism.
 - Many parents did not vaccinate their children because of this (debunked) claim.