

Univariate and Bivariate Tests

BUS 735: Business Decision Making and Research

Goals

- 1 Be able to distinguish different types of data and prescribe appropriate statistical methods.
- 2 Conduct a number of hypothesis tests using methods appropriate for questions involving only one or two variables.



Learning Outcomes

- LO1: Be able to construct and test hypotheses using a variety of bivariate statistical methods to compare characteristics between two populations.
- LO6: Be able to use standard computer packages such as R to conduct the quantitative analyses described in the learning objectives above.



Agenda

Learning Objective	Active Learning Activity
Be able to distinguish different types of data.	Lecture / Discussion
Learn and conduct hypothesis tests on single variables.	Learn by doing: work together on examples using R.
Learn and conduct hypothesis tests for differences between two variables.	Learn by doing: work together on examples using R.

Types of Data

- Nominal data: consists of categories that cannot be ordered in a meaningful way.
- Ordinal data: order is meaningful, but not the distances between data values.
 - Excellent, Very good, Good, Poor, Very poor.
- Interval data: order is meaningful, *and* distances are meaningful. However, there is *no natural zero*.
 - Examples: temperature, time.
- Ratio data: order, differences, and zero are all meaningful.
 - Examples: weight, prices, speed.

Types of Data

- Nominal data: consists of categories that cannot be ordered in a meaningful way.
- Ordinal data: order is meaningful, but not the distances between data values.
 - Excellent, Very good, Good, Poor, Very poor.
- Interval data: order is meaningful, *and* distances are meaningful. However, there is *no natural zero*.
 - Examples: temperature, time.
- Ratio data: order, differences, and zero are all meaningful.
 - Examples: weight, prices, speed.

Types of Data

- Nominal data: consists of categories that cannot be ordered in a meaningful way.
- Ordinal data: order is meaningful, but not the distances between data values.
 - Excellent, Very good, Good, Poor, Very poor.
- Interval data: order is meaningful, *and* distances are meaningful. However, there is *no natural zero*.
 - Examples: temperature, time.
- Ratio data: order, differences, and zero are all meaningful.
 - Examples: weight, prices, speed.

Types of Data

- Nominal data: consists of categories that cannot be ordered in a meaningful way.
- Ordinal data: order is meaningful, but not the distances between data values.
 - Excellent, Very good, Good, Poor, Very poor.
- Interval data: order is meaningful, *and* distances are meaningful. However, there is *no natural zero*.
 - Examples: temperature, time.
- Ratio data: order, differences, and zero are all meaningful.
 - Examples: weight, prices, speed.

Types of Data

- Nominal data: consists of categories that cannot be ordered in a meaningful way.
- Ordinal data: order is meaningful, but not the distances between data values.
 - Excellent, Very good, Good, Poor, Very poor.
- Interval data: order is meaningful, *and* distances are meaningful. However, there is *no natural zero*.
 - Examples: temperature, time.
- Ratio data: order, differences, and zero are all meaningful.
 - Examples: weight, prices, speed.

Types of Data

- Nominal data: consists of categories that cannot be ordered in a meaningful way.
- Ordinal data: order is meaningful, but not the distances between data values.
 - Excellent, Very good, Good, Poor, Very poor.
- Interval data: order is meaningful, *and* distances are meaningful. However, there is *no natural zero*.
 - Examples: temperature, time.
- Ratio data: order, differences, and zero are all meaningful.
 - Examples: weight, prices, speed.

Types of Data

- Nominal data: consists of categories that cannot be ordered in a meaningful way.
- Ordinal data: order is meaningful, but not the distances between data values.
 - Excellent, Very good, Good, Poor, Very poor.
- Interval data: order is meaningful, *and* distances are meaningful. However, there is *no natural zero*.
 - Examples: temperature, time.
- Ratio data: order, differences, and zero are all meaningful.
 - Examples: weight, prices, speed.

Types of Tests

- Different types of data require different statistical methods.
- Why? With interval data and below, operations like addition, subtraction, multiplication, and division are *meaningless!*
- Parametric statistics:
 - Typically take advantage of central limit theorem (imposes requirements on probability distributions)
 - Appropriate only for interval and ratio data.
 - More **powerful** than nonparametric methods.
- Nonparametric statistics:
 - Do not require assumptions concerning the probability distribution for the population.
 - There are many methods appropriate for ordinal data, some methods appropriate for nominal data.
 - Computations typically make use of *ranks* instead of actual data.

Types of Tests

- Different types of data require different statistical methods.
- Why? With interval data and below, operations like addition, subtraction, multiplication, and division are *meaningless!*
- Parametric statistics:
 - Typically take advantage of central limit theorem (imposes requirements on probability distributions)
 - Appropriate only for interval and ratio data.
 - More powerful than nonparametric methods.
- Nonparametric statistics:
 - Do not require assumptions concerning the probability distribution for the population.
 - There are many methods appropriate for ordinal data, some methods appropriate for nominal data.
 - Computations typically make use of *ranks* instead of actual data.

Types of Tests

- Different types of data require different statistical methods.
- Why? With interval data and below, operations like addition, subtraction, multiplication, and division are *meaningless!*
- Parametric statistics:
 - Typically take advantage of central limit theorem (imposes requirements on probability distributions)
 - Appropriate only for interval and ratio data.
 - More **powerful** than nonparametric methods.
- Nonparametric statistics:
 - Do not require assumptions concerning the probability distribution for the population.
 - There are many methods appropriate for ordinal data, some methods appropriate for nominal data.
 - Computations typically make use of *ranks* instead of actual data.

Types of Tests

- Different types of data require different statistical methods.
- Why? With interval data and below, operations like addition, subtraction, multiplication, and division are *meaningless!*
- Parametric statistics:
 - Typically take advantage of central limit theorem (imposes requirements on probability distributions)
 - Appropriate only for interval and ratio data.
 - More **powerful** than nonparametric methods.
- Nonparametric statistics:
 - Do not require assumptions concerning the probability distribution for the population.
 - There are many methods appropriate for ordinal data, some methods appropriate for nominal data.
 - Computations typically make use of *ranks* instead of actual data.

Types of Tests

- Different types of data require different statistical methods.
- Why? With interval data and below, operations like addition, subtraction, multiplication, and division are *meaningless!*
- Parametric statistics:
 - Typically take advantage of central limit theorem (imposes requirements on probability distributions)
 - Appropriate only for interval and ratio data.
 - More **powerful** than nonparametric methods.
- Nonparametric statistics:
 - Do not require assumptions concerning the probability distribution for the population.
 - There are many methods appropriate for ordinal data, some methods appropriate for nominal data.
 - Computations typically make use of *ranks* instead of actual data.

Types of Tests

- Different types of data require different statistical methods.
- Why? With interval data and below, operations like addition, subtraction, multiplication, and division are *meaningless!*
- Parametric statistics:
 - Typically take advantage of central limit theorem (imposes requirements on probability distributions)
 - Appropriate only for interval and ratio data.
 - More **powerful** than nonparametric methods.
- Nonparametric statistics:
 - Do not require assumptions concerning the probability distribution for the population.
 - There are many methods appropriate for ordinal data, some methods appropriate for nominal data.
 - Computations typically make use of *ranks* instead of actual data.

Types of Tests

- Different types of data require different statistical methods.
- Why? With interval data and below, operations like addition, subtraction, multiplication, and division are *meaningless!*
- Parametric statistics:
 - Typically take advantage of central limit theorem (imposes requirements on probability distributions)
 - Appropriate only for interval and ratio data.
 - More **powerful** than nonparametric methods.
- Nonparametric statistics:
 - Do not require assumptions concerning the probability distribution for the population.
 - There are many methods appropriate for ordinal data, some methods appropriate for nominal data.
 - Computations typically make use of *ranks* instead of actual data.

Types of Tests

- Different types of data require different statistical methods.
- Why? With interval data and below, operations like addition, subtraction, multiplication, and division are *meaningless!*
- Parametric statistics:
 - Typically take advantage of central limit theorem (imposes requirements on probability distributions)
 - Appropriate only for interval and ratio data.
 - More **powerful** than nonparametric methods.
- Nonparametric statistics:
 - Do not require assumptions concerning the probability distribution for the population.
 - There are many methods appropriate for ordinal data, some methods appropriate for nominal data.
 - Computations typically make use of *ranks* instead of actual data.

Types of Tests

- Different types of data require different statistical methods.
- Why? With interval data and below, operations like addition, subtraction, multiplication, and division are *meaningless!*
- Parametric statistics:
 - Typically take advantage of central limit theorem (imposes requirements on probability distributions)
 - Appropriate only for interval and ratio data.
 - More **powerful** than nonparametric methods.
- Nonparametric statistics:
 - Do not require assumptions concerning the probability distribution for the population.
 - There are many methods appropriate for ordinal data, some methods appropriate for nominal data.
 - Computations typically make use of *ranks* instead of actual data.

Types of Tests

- Different types of data require different statistical methods.
- Why? With interval data and below, operations like addition, subtraction, multiplication, and division are *meaningless!*
- Parametric statistics:
 - Typically take advantage of central limit theorem (imposes requirements on probability distributions)
 - Appropriate only for interval and ratio data.
 - More **powerful** than nonparametric methods.
- Nonparametric statistics:
 - Do not require assumptions concerning the probability distribution for the population.
 - There are many methods appropriate for ordinal data, some methods appropriate for nominal data.
 - Computations typically make use of *ranks* instead of actual data.

Single Mean T-Test

6 / 18

- Test whether the population mean is equal or different to some value.
- Uses the sample mean its statistic.
- Parametric test that depends on results from Central Limit Theorem.
- Hypotheses
 - Null: The population mean is equal to some specified value.
 - Alternative: The population mean is [greater/less/different] than the value in the null.

Single Mean T-Test

6 / 18

- Test whether the population mean is equal or different to some value.
- Uses the sample mean its statistic.
- Parametric test that depends on results from Central Limit Theorem.
- Hypotheses
 - Null: The population mean is equal to some specified value.
 - Alternative: The population mean is [greater/less/different] than the value in the null.

Single Mean T-Test

6/ 18

- Test whether the population mean is equal or different to some value.
- Uses the sample mean its statistic.
- Parametric test that depends on results from Central Limit Theorem.
- Hypotheses
 - Null: The population mean is equal to some specified value.
 - Alternative: The population mean is [greater/less/different] than the value in the null.

Single Mean T-Test

6 / 18

- Test whether the population mean is equal or different to some value.
- Uses the sample mean its statistic.
- Parametric test that depends on results from Central Limit Theorem.
- Hypotheses
 - Null: The population mean is equal to some specified value.
 - Alternative: The population mean is [greater/less/different] than the value in the null.

Single Mean T-Test

6 / 18

- Test whether the population mean is equal or different to some value.
- Uses the sample mean its statistic.
- Parametric test that depends on results from Central Limit Theorem.
- Hypotheses
 - Null: The population mean is equal to some specified value.
 - Alternative: The population mean is [greater/less/different] than the value in the null.

Single Mean T-Test

- Test whether the population mean is equal or different to some value.
- Uses the sample mean its statistic.
- Parametric test that depends on results from Central Limit Theorem.
- Hypotheses
 - Null: The population mean is equal to some specified value.
 - Alternative: The population mean is [greater/less/different] than the value in the null.

Example: Average Hourly Earnings

Dataset: Current Population Survey from 2004 that includes data on average hourly earnings, marital status, gender, and age for thousands of people.

<http://murraylax.org/datasets/cps04.csv>

Answer the following questions:

- 1 Report the mean average hourly earnings in the sample.
- 2 Construct a 95% confidence interval estimate for the average hourly earnings.
- 3 Test the hypothesis that average hourly earnings is greater than \$16.50.
- 4 Test the hypothesis that average hourly earnings *is different than* \$16.50.

Example: Average Hourly Earnings

Dataset: Current Population Survey from 2004 that includes data on average hourly earnings, marital status, gender, and age for thousands of people.

<http://murraylax.org/datasets/cps04.csv>

Answer the following questions:

- 1 Report the mean average hourly earnings in the sample.
- 2 Construct a 95% confidence interval estimate for the average hourly earnings.
- 3 Test the hypothesis that average hourly earnings is greater than \$16.50.
- 4 Test the hypothesis that average hourly earnings *is different than* \$16.50.

Example: Average Hourly Earnings

Dataset: Current Population Survey from 2004 that includes data on average hourly earnings, marital status, gender, and age for thousands of people.

<http://murraylax.org/datasets/cps04.csv>

Answer the following questions:

- 1 Report the mean average hourly earnings in the sample.
- 2 Construct a 95% confidence interval estimate for the average hourly earnings.
- 3 Test the hypothesis that average hourly earnings is greater than \$16.50.
- 4 Test the hypothesis that average hourly earnings *is different than* \$16.50.

Single Median Nonparametric Test

Why not perform a t-test on the mean?

- Ordinal data: Cannot compute sample means (they are meaningless), only median is meaningful.
- Small sample size and you are not sure the population is not normal.

Hypothesis test appropriate for medians:

Single-sample Wilcoxon Signed Rank Test

Single Median Nonparametric Test

Why not perform a t-test on the mean?

- Ordinal data: Cannot compute sample means (they are meaningless), only median is meaningful.
- Small sample size and you are not sure the population is not normal.

Hypothesis test appropriate for medians:

Single-sample Wilcoxon Signed Rank Test

Single Median Nonparametric Test

Why not perform a t-test on the mean?

- Ordinal data: Cannot compute sample means (they are meaningless), only median is meaningful.
- Small sample size and you are not sure the population is not normal.

Hypothesis test appropriate for medians:

Single-sample Wilcoxon Signed Rank Test

Single Median Nonparametric Test

Why not perform a t-test on the mean?

- Ordinal data: Cannot compute sample means (they are meaningless), only median is meaningful.
- Small sample size and you are not sure the population is not normal.

Hypothesis test appropriate for medians:

Single-sample Wilcoxon Signed Rank Test

Single Median Nonparametric Test

Why not perform a t-test on the mean?

- Ordinal data: Cannot compute sample means (they are meaningless), only median is meaningful.
- Small sample size and you are not sure the population is not normal.

Hypothesis test appropriate for medians:

Single-sample Wilcoxon Signed Rank Test

Wilcoxon Signed Rank Test

Hypotheses:

- Null: The population is centered around the null specified value.
- Alternative: The population is centered around a value different from the null specified value.

Sample estimates:

- Sample median (middle number)
- Interpolated median: for ordinal data with limited number of outcomes, this takes into account the percentage of the data that is *strictly below* versus *strictly above* the median.

Wilcoxon Signed Rank Test

Hypotheses:

- Null: The population is centered around the null specified value.
- Alternative: The population is centered around a value different from the null specified value.

Sample estimates:

- Sample median (middle number)
- Interpolated median: for ordinal data with limited number of outcomes, this takes into account the percentage of the data that is *strictly below* versus *strictly above* the median.

Wilcoxon Signed Rank Test

Hypotheses:

- Null: The population is centered around the null specified value.
- Alternative: The population is centered around a value different from the null specified value.

Sample estimates:

- Sample median (middle number)
- Interpolated median: for ordinal data with limited number of outcomes, this takes into account the percentage of the data that is *strictly below* versus *strictly above* the median.

Wilcoxon Signed Rank Test

Hypotheses:

- Null: The population is centered around the null specified value.
- Alternative: The population is centered around a value different from the null specified value.

Sample estimates:

- Sample median (middle number)
- Interpolated median: for ordinal data with limited number of outcomes, this takes into account the percentage of the data that is *strictly below* versus *strictly above* the median.

Wilcoxon Signed Rank Test

Hypotheses:

- Null: The population is centered around the null specified value.
- Alternative: The population is centered around a value different from the null specified value.

Sample estimates:

- Sample median (middle number)
- Interpolated median: for ordinal data with limited number of outcomes, this takes into account the percentage of the data that is *strictly below* versus *strictly above* the median.

Wilcoxon Signed Rank Test

Hypotheses:

- Null: The population is centered around the null specified value.
- Alternative: The population is centered around a value different from the null specified value.

Sample estimates:

- Sample median (middle number)
- Interpolated median: for ordinal data with limited number of outcomes, this takes into account the percentage of the data that is *strictly below* versus *strictly above* the median.

Example: Attitudes Grade School Kids

- Dataset: 438 students in grades 4 through 6 were sampled from three school districts in Michigan. Students ranked from 1 (most important) to 4 (least important) how important grades, sports, being good looking, and having lots of money were to each of them.
- Dataset
<http://murraylax.org/datasets/gradschools.csv>.
- Answer some of these questions:
 - ① Report the median and interpolated median for how important grades are to students.
 - ② Report a 95% confidence interval for the median.
 - ③ Is the median importance for grades is greater than 3?
 - ④ Is the median importance for money less than 3?

Example: Attitudes Grade School Kids

- Dataset: 438 students in grades 4 through 6 were sampled from three school districts in Michigan. Students ranked from 1 (most important) to 4 (least important) how important grades, sports, being good looking, and having lots of money were to each of them.
- Dataset
<http://murraylax.org/datasets/gradschools.csv>.
- Answer some of these questions:
 - ① Report the median and interpolated median for how important grades are to students.
 - ② Report a 95% confidence interval for the median.
 - ③ Is the median importance for grades is greater than 3?
 - ④ Is the median importance for money less than 3?

Example: Attitudes Grade School Kids

- Dataset: 438 students in grades 4 through 6 were sampled from three school districts in Michigan. Students ranked from 1 (most important) to 4 (least important) how important grades, sports, being good looking, and having lots of money were to each of them.
- Dataset
<http://murraylax.org/datasets/gradschools.csv>.
- Answer some of these questions:
 - 1 Report the median and interpolated median for how important grades are to students.
 - 2 Report a 95% confidence interval for the median.
 - 3 Is the median importance for grades is greater than 3?
 - 4 Is the median importance for money less than 3?

Example: Attitudes Grade School Kids

- Dataset: 438 students in grades 4 through 6 were sampled from three school districts in Michigan. Students ranked from 1 (most important) to 4 (least important) how important grades, sports, being good looking, and having lots of money were to each of them.
- Dataset
<http://murraylax.org/datasets/gradschools.csv>.
- Answer some of these questions:
 - 1 Report the median and interpolated median for how important grades are to students.
 - 2 Report a 95% confidence interval for the median.
 - 3 Is the median importance for grades is greater than 3?
 - 4 Is the median importance for money less than 3?

Difference in Means (Independent Samples)

- Suppose you want to know whether the mean from one population is larger than the mean for another.
- Statistic: Difference in the sample means ($\bar{x}_1 - \bar{x}_2$).
- Hypotheses:
 - Null: the difference between the two means is zero.
 - Alternative: the difference is [above/below/not equal] to zero.

Difference in Means (Independent Samples)

- Suppose you want to know whether the mean from one population is larger than the mean for another.
- Statistic: Difference in the sample means ($\bar{x}_1 - \bar{x}_2$).
- Hypotheses:
 - Null: the difference between the two means is zero.
 - Alternative: the difference is [above/below/not equal] to zero.

Difference in Means (Independent Samples)

- Suppose you want to know whether the mean from one population is larger than the mean for another.
- Statistic: Difference in the sample means ($\bar{x}_1 - \bar{x}_2$).
- Hypotheses:
 - Null: the difference between the two means is zero.
 - Alternative: the difference is [above/below/not equal] to zero.

Difference in Means (Independent Samples)

- Suppose you want to know whether the mean from one population is larger than the mean for another.
- Statistic: Difference in the sample means ($\bar{x}_1 - \bar{x}_2$).
- Hypotheses:
 - Null: the difference between the two means is zero.
 - Alternative: the difference is [above/below/not equal] to zero.

Difference in Means (Independent Samples)

- Suppose you want to know whether the mean from one population is larger than the mean for another.
- Statistic: Difference in the sample means ($\bar{x}_1 - \bar{x}_2$).
- Hypotheses:
 - Null: the difference between the two means is zero.
 - Alternative: the difference is [above/below/not equal] to zero.

Example: Average Hourly Earnings

Dataset: Current Population Survey from 2004 that includes data on average hourly earnings, marital status, gender, and age for thousands of people.

<http://murraylax.org/datasets/cps04.csv>

Answer the following questions:

- 1 What is the average hourly earnings for males versus females?
- 2 Estimate a 95% confidence interval for the difference in average hourly earnings between males and females.
- 3 Test the hypothesis that men and women earn have *different average hourly earnings*
- 4 Test the hypothesis that men earn on average *more than \$2.00* per hour above women.

Example: Average Hourly Earnings

Dataset: Current Population Survey from 2004 that includes data on average hourly earnings, marital status, gender, and age for thousands of people.

<http://murraylax.org/datasets/cps04.csv>

Answer the following questions:

- 1 What is the average hourly earnings for males versus females?
- 2 Estimate a 95% confidence interval for the difference in average hourly earnings between males and females.
- 3 Test the hypothesis that men and women earn have *different average hourly earnings*
- 4 Test the hypothesis that men earn on average *more than \$2.00* per hour above women.

Example: Average Hourly Earnings

Dataset: Current Population Survey from 2004 that includes data on average hourly earnings, marital status, gender, and age for thousands of people.

<http://murraylax.org/datasets/cps04.csv>

Answer the following questions:

- 1 What is the average hourly earnings for males versus females?
- 2 Estimate a 95% confidence interval for the difference in average hourly earnings between males and females.
- 3 Test the hypothesis that men and women earn have *different average hourly earnings*
- 4 Test the hypothesis that men earn on average *more than \$2.00* per hour above women.

Nonparametric Tests for Differences in Medians

- Mann-Whitney U test: nonparametric test to determine difference in *medians*.
- Assumptions:
 - Samples are independent
 - The underlying distributions have the same shape (i.e. only the location of the distribution is different).
(violating this assumption does not severely change the sampling distribution of the Mann-Whitney U test)
- Null hypothesis: medians for the two populations are the same.
- Alternative hypotheses: medians for the two populations are different.

Nonparametric Tests for Differences in Medians

- Mann-Whitney U test: nonparametric test to determine difference in *medians*.
- Assumptions:
 - Samples are independent
 - The underlying distributions have the same shape (i.e. only the location of the distribution is different).
(violating this assumption does not severely change the sampling distribution of the Mann-Whitney U test)
- Null hypothesis: medians for the two populations are the same.
- Alternative hypotheses: medians for the two populations are different.

Nonparametric Tests for Differences in Medians

- Mann-Whitney U test: nonparametric test to determine difference in *medians*.
- Assumptions:
 - Samples are independent
 - The underlying distributions have the same shape (i.e. only the location of the distribution is different).
(violating this assumption does not severely change the sampling distribution of the Mann-Whitney U test)
- Null hypothesis: medians for the two populations are the same.
- Alternative hypotheses: medians for the two populations are different.

Nonparametric Tests for Differences in Medians

- Mann-Whitney U test: nonparametric test to determine difference in *medians*.
- Assumptions:
 - Samples are independent
 - The underlying distributions have the same shape (i.e. only the location of the distribution is different).
(violating this assumption does not severely change the sampling distribution of the Mann-Whitney U test)
- Null hypothesis: medians for the two populations are the same.
- Alternative hypotheses: medians for the two populations are different.

Nonparametric Tests for Differences in Medians

- Mann-Whitney U test: nonparametric test to determine difference in *medians*.
- Assumptions:
 - Samples are independent
 - The underlying distributions have the same shape (i.e. only the location of the distribution is different).
(violating this assumption does not severely change the sampling distribution of the Mann-Whitney U test)
- Null hypothesis: medians for the two populations are the same.
- Alternative hypotheses: medians for the two populations are different.

Nonparametric Tests for Differences in Medians

- Mann-Whitney U test: nonparametric test to determine difference in *medians*.
- Assumptions:
 - Samples are independent
 - The underlying distributions have the same shape (i.e. only the location of the distribution is different).
(violating this assumption does not severely change the sampling distribution of the Mann-Whitney U test)
- Null hypothesis: medians for the two populations are the same.
- Alternative hypotheses: medians for the two populations are different.

Example: Grade School Kids' Attitudes

- Dataset: 438 students in grades 4 through 6 were sampled from three school districts in Michigan. Students ranked from 1 (most important) to 4 (least important) how important grades, sports, being good looking, and having lots of money were to each of them.
- Dataset
<http://murraylax.org/datasets/gradschools.csv>.
- Answer these questions:
 - ① Report the median and interpolated median for how important grades are for boys versus girls.
 - ② Is the median importance for grades different for boys versus girls?

Example: Grade School Kids' Attitudes

- Dataset: 438 students in grades 4 through 6 were sampled from three school districts in Michigan. Students ranked from 1 (most important) to 4 (least important) how important grades, sports, being good looking, and having lots of money were to each of them.
- Dataset
<http://murraylax.org/datasets/gradschools.csv>.
- Answer these questions:
 - ① Report the median and interpolated median for how important grades are for boys versus girls.
 - ② Is the median importance for grades different for boys versus girls?

Example: Grade School Kids' Attitudes

- Dataset: 438 students in grades 4 through 6 were sampled from three school districts in Michigan. Students ranked from 1 (most important) to 4 (least important) how important grades, sports, being good looking, and having lots of money were to each of them.
- Dataset
<http://murraylax.org/datasets/gradschools.csv>.
- Answer these questions:
 - 1 Report the median and interpolated median for how important grades are for boys versus girls.
 - 2 Is the median importance for grades different for boys versus girls?

Example: Grade School Kids' Attitudes

- Dataset: 438 students in grades 4 through 6 were sampled from three school districts in Michigan. Students ranked from 1 (most important) to 4 (least important) how important grades, sports, being good looking, and having lots of money were to each of them.
- Dataset
<http://murraylax.org/datasets/gradschools.csv>.
- Answer these questions:
 - ① Report the median and interpolated median for how important grades are for boys versus girls.
 - ② Is the median importance for grades different for boys versus girls?

Dependent Samples - Paired Samples

- Use a **paired sampled test** if the two samples have the same individuals or sampling units.
- Many examples include before/after tests for differences:
 - The Biggest Loser: Compare the weight of people on the show before the season begins and one year after the show concludes.
 - Training session: Are workers more productive 6 months after they attended some training session versus before the training session.
- Examples besides before/after tests for differences:
 - Do students spend more time studying than watching TV?
 - Does the unemployment rate for White/Caucasian differ from the unemployment rate for African Americans (sampling unit = U.S. state).
- These are *dependent samples*, because you have the *same sampling units* in each group.

Dependent Samples - Paired Samples

- Use a **paired sampled test** if the two samples have the same individuals or sampling units.
- Many examples include before/after tests for differences:
 - The Biggest Loser: Compare the weight of people on the show before the season begins and one year after the show concludes.
 - Training session: Are workers more productive 6 months after they attended some training session versus before the training session.
- Examples besides before/after tests for differences:
 - Do students spend more time studying than watching TV?
 - Does the unemployment rate for White/Caucasian differ from the unemployment rate for African Americans (sampling unit = U.S. state).
- These are *dependent samples*, because you have the *same sampling units* in each group.

Dependent Samples - Paired Samples

- Use a **paired sampled test** if the two samples have the same individuals or sampling units.
- Many examples include before/after tests for differences:
 - The Biggest Loser: Compare the weight of people on the show before the season begins and one year after the show concludes.
 - Training session: Are workers more productive 6 months after they attended some training session versus before the training session.
- Examples besides before/after tests for differences:
 - Do students spend more time studying than watching TV?
 - Does the unemployment rate for White/Caucasian differ from the unemployment rate for African Americans (sampling unit = U.S. state).
- These are *dependent samples*, because you have the *same sampling units* in each group.

Dependent Samples - Paired Samples

- Use a **paired sampled test** if the two samples have the same individuals or sampling units.
- Many examples include before/after tests for differences:
 - The Biggest Loser: Compare the weight of people on the show before the season begins and one year after the show concludes.
 - Training session: Are workers more productive 6 months after they attended some training session versus before the training session.
- Examples besides before/after tests for differences:
 - Do students spend more time studying than watching TV?
 - Does the unemployment rate for White/Caucasian differ from the unemployment rate for African Americans (sampling unit = U.S. state).
- These are *dependent samples*, because you have the *same sampling units* in each group.

Dependent Samples - Paired Samples

- Use a **paired sampled test** if the two samples have the same individuals or sampling units.
- Many examples include before/after tests for differences:
 - The Biggest Loser: Compare the weight of people on the show before the season begins and one year after the show concludes.
 - Training session: Are workers more productive 6 months after they attended some training session versus before the training session.
- Examples besides before/after tests for differences:
 - Do students spend more time studying than watching TV?
 - Does the unemployment rate for White/Caucasian differ from the unemployment rate for African Americans (sampling unit = U.S. state).
- These are *dependent samples*, because you have the *same sampling units* in each group.

Dependent Samples - Paired Samples

- Use a **paired sampled test** if the two samples have the same individuals or sampling units.
- Many examples include before/after tests for differences:
 - The Biggest Loser: Compare the weight of people on the show before the season begins and one year after the show concludes.
 - Training session: Are workers more productive 6 months after they attended some training session versus before the training session.
- Examples besides before/after tests for differences:
 - Do students spend more time studying than watching TV?
 - Does the unemployment rate for White/Caucasian differ from the unemployment rate for African Americans (sampling unit = U.S. state).
- These are *dependent samples*, because you have the *same sampling units* in each group.

Dependent Samples - Paired Samples

- Use a **paired sampled test** if the two samples have the same individuals or sampling units.
- Many examples include before/after tests for differences:
 - The Biggest Loser: Compare the weight of people on the show before the season begins and one year after the show concludes.
 - Training session: Are workers more productive 6 months after they attended some training session versus before the training session.
- Examples besides before/after tests for differences:
 - Do students spend more time studying than watching TV?
 - Does the unemployment rate for White/Caucasian differ from the unemployment rate for African Americans (sampling unit = U.S. state).
- These are *dependent samples*, because you have the *same sampling units* in each group.

Dependent Samples - Paired Samples

- Use a **paired sampled test** if the two samples have the same individuals or sampling units.
- Many examples include before/after tests for differences:
 - The Biggest Loser: Compare the weight of people on the show before the season begins and one year after the show concludes.
 - Training session: Are workers more productive 6 months after they attended some training session versus before the training session.
- Examples besides before/after tests for differences:
 - Do students spend more time studying than watching TV?
 - Does the unemployment rate for White/Caucasian differ from the unemployment rate for African Americans (sampling unit = U.S. state).
- These are *dependent samples*, because you have the *same sampling units* in each group.

Parametric and Nonparametric Paired Samples Tests 16/ 18

Paired-samples difference in means t-test

- Appropriate for interval or ratio data
- Appropriate when assumptions of CLT are met

Wilcoxon-Signed Rank Test for Paired-samples

- Tests for a difference in medians (center of distribution)
- Appropriate for ordinal, interval, or ratio data
- Appropriate when assumptions of CLT are met

Parametric and Nonparametric Paired Samples Tests 16/ 18

Paired-samples difference in means t-test

- Appropriate for interval or ratio data
- Appropriate when assumptions of CLT are met

Wilcoxon-Signed Rank Test for Paired-samples

- Tests for a difference in medians (center of distribution)
- Appropriate for ordinal, interval, or ratio data
- Appropriate when assumptions of CLT are met

Parametric and Nonparametric Paired Samples Tests 16/ 18

Paired-samples difference in means t-test

- Appropriate for interval or ratio data
- Appropriate when assumptions of CLT are met

Wilcoxon-Signed Rank Test for Paired-samples

- Tests for a difference in medians (center of distribution)
- Appropriate for ordinal, interval, or ratio data
- Appropriate when assumptions of CLT are met

Parametric and Nonparametric Paired Samples Tests 16/ 18

Paired-samples difference in means t-test

- Appropriate for interval or ratio data
- Appropriate when assumptions of CLT are met

Wilcoxon-Signed Rank Test for Paired-samples

- Tests for a difference in medians (center of distribution)
- Appropriate for ordinal, interval, or ratio data
- Appropriate when assumptions of CLT are met

Parametric and Nonparametric Paired Samples Tests 16/ 18

Paired-samples difference in means t-test

- Appropriate for interval or ratio data
- Appropriate when assumptions of CLT are met

Wilcoxon-Signed Rank Test for Paired-samples

- Tests for a difference in medians (center of distribution)
- Appropriate for ordinal, interval, or ratio data
- Appropriate when assumptions of CLT are met

Parametric and Nonparametric Paired Samples Tests 16/ 18

Paired-samples difference in means t-test

- Appropriate for interval or ratio data
- Appropriate when assumptions of CLT are met

Wilcoxon-Signed Rank Test for Paired-samples

- Tests for a difference in medians (center of distribution)
- Appropriate for ordinal, interval, or ratio data
- Appropriate when assumptions of CLT are met

Parametric and Nonparametric Paired Samples Tests 16/ 18

Paired-samples difference in means t-test

- Appropriate for interval or ratio data
- Appropriate when assumptions of CLT are met

Wilcoxon-Signed Rank Test for Paired-samples

- Tests for a difference in medians (center of distribution)
- Appropriate for ordinal, interval, or ratio data
- Appropriate when assumptions of CLT are met

Paired Samples Means: Motor Vehicle Fatalities

17 / 18

- Centers for Disease Control and Prevention (CDC) state-level data (50 obs) on motor vehicle fatalities by state, age, and sex
- Variables: Motor vehicle occupant fatality rate per 100,000 members of the population
 - Over all age groups
 - Individual age groups: 0-20, 21-34, 35-54, and 55+.
 - Male versus Female
- Dataset: <http://murraylax.org/datasets/vehiclegfatalities.csv>
- Answer the following questions:
 - Report the sample average mortality rate for age groups 21-34 and 35-54.
 - Report the difference in the average mortality rate for age groups 21-34 and 35-54.
 - Report a confidence interval for the difference above.
 - Do age groups 21-34 and 35-54 have different average mortality rates?

Paired Samples Means: Motor Vehicle Fatalities

17 / 18

- Centers for Disease Control and Prevention (CDC) state-level data (50 obs) on motor vehicle fatalities by state, age, and sex
- Variables: Motor vehicle occupant fatality rate per 100,000 members of the population
 - Over all age groups
 - Individual age groups: 0-20, 21-34, 35-54, and 55+.
 - Male versus Female
- Dataset: <http://murraylax.org/datasets/vehiclefatalities.csv>
- Answer the following questions:
 - Report the sample average mortality rate for age groups 21-34 and 35-54.
 - Report the difference in the average mortality rate for age groups 21-34 and 35-54.
 - Report a confidence interval for the difference above.
 - Do age groups 21-34 and 35-54 have different average mortality rates?

Paired Samples Means: Motor Vehicle Fatalities

17 / 18

- Centers for Disease Control and Prevention (CDC) state-level data (50 obs) on motor vehicle fatalities by state, age, and sex
- Variables: Motor vehicle occupant fatality rate per 100,000 members of the population
 - Over all age groups
 - Individual age groups: 0-20, 21-34, 35-54, and 55+.
 - Male versus Female
- Dataset: <http://murraylax.org/datasets/vehiclegfatalities.csv>
- Answer the following questions:
 - Report the sample average mortality rate for age groups 21-34 and 35-54.
 - Report the difference in the average mortality rate for age groups 21-34 and 35-54.
 - Report a confidence interval for the difference above.
 - Do age groups 21-34 and 35-54 have different average mortality rates?

Paired Samples Means: Motor Vehicle Fatalities

- Centers for Disease Control and Prevention (CDC) state-level data (50 obs) on motor vehicle fatalities by state, age, and sex
- Variables: Motor vehicle occupant fatality rate per 100,000 members of the population
 - Over all age groups
 - Individual age groups: 0-20, 21-34, 35-54, and 55+.
 - Male versus Female
- Dataset: <http://murraylax.org/datasets/vehiclegfatalities.csv>
- Answer the following questions:
 - Report the sample average mortality rate for age groups 21-34 and 35-54.
 - Report the difference in the average mortality rate for age groups 21-34 and 35-54.
 - Report a confidence interval for the difference above.
 - Do age groups 21-34 and 35-54 have different average mortality rates?

Paired-Samples Medians: Grade School Attitudes

18 / 18

- Dataset: 438 students in grades 4 through 6 were sampled from three school districts in Michigan. Students ranked from 1 (most important) to 4 (least important) how important grades, sports, being good looking, and having lots of money were to each of them.
- Dataset
<http://murraylax.org/datasets/gradschools.csv>.
- Answer these questions:
 - ① Report the median and interpolated median for how important are grades.
 - ② Report the median and interpolated median for how important is sports.
 - ③ Is the median importance for grades different that the median importance for sports?

Paired-Samples Medians: Grade School Attitudes

- Dataset: 438 students in grades 4 through 6 were sampled from three school districts in Michigan. Students ranked from 1 (most important) to 4 (least important) how important grades, sports, being good looking, and having lots of money were to each of them.
- Dataset
<http://murraylax.org/datasets/gradschools.csv>.
- Answer these questions:
 - Report the median and interpolated median for how important are grades.
 - Report the median and interpolated median for how important is sports.
 - Is the median importance for grades different that the median importance for sports?

Paired-Samples Medians: Grade School Attitudes

- Dataset: 438 students in grades 4 through 6 were sampled from three school districts in Michigan. Students ranked from 1 (most important) to 4 (least important) how important grades, sports, being good looking, and having lots of money were to each of them.
- Dataset
<http://murraylax.org/datasets/gradschools.csv>.
- Answer these questions:
 - 1 Report the median and interpolated median for how important are grades.
 - 2 Report the median and interpolated median for how important is sports.
 - 3 Is the median importance for grades different that the median importance for sports?