Module 11: Nominal and Ordinal Variable Analysis

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

UNT UNIVERSITY OF NORTH TEXAS
Discover the power of ideas.

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`

Outline

1. Chi-square test
   - One-way Classification Tables
   - Multi-way Contingency Tables
   - Effect Size
   - Kappa

Outline

1. Chi-square test
   - One-way Classification Tables
   - Multi-way Contingency Tables
   - Effect Size
   - Kappa

2. Wilcoxon's Ranks tests
   - Wilcoxon's Rank-Sum Test
   - Wilcoxon's Matched-Pairs Signed-Ranks Test

RSS
Research and Statistical Support

Outline

1. Chi-square test
   - One-way Classification Tables
   - Multi-way Contingency Tables
   - Effect Size
   - Kappa
2. Wilcoxon's Ranks tests
   - Wilcoxon's Rank-Sum Test
   - Wilcoxon's Matched-Pairs Signed-Ranks Test
3. Kruskal-Wallis One-way ANOVA

Outline

RSS
Research and Statistical Support

Parametric statistics

- All of the previous modules dealt with *Parametric* statistics.

Parametric statistics

- All of the previous modules dealt with *Parametric* statistics.
  - Concerned with population values (i.e. parameters).

Parametric statistics

- All of the previous modules dealt with *Parametric* statistics.
  - Concerned with population values (i.e. parameters).
  - Require Interval and/or ratio scaled variables.

RSS
Research and Statistical Support

Parametric statistics

- All of the previous modules dealt with *Parametric* statistics.
  - Concerned with population values (i.e. parameters).
  - Require Interval and/or ratio scaled variables.
  - Assumptions about population distributions.

RSS
Research and Statistical Support

Parametric statistics

- All of the previous modules dealt with *Parametric* statistics.
  - Concerned with population values (i.e. parameters).
  - Require Interval and/or ratio scaled variables.
  - Assumptions about population distributions.
- This module (11) concerns itself with *Nonparametric* statistics.

Nonparametric statistics

- Most Nonparametrics are still concerned with populations, but the hypotheses are not formally stated using population values.

Nonparametric statistics

- Most Nonparametrics are still concerned with populations, but the hypotheses are not formally stated using population values.
    - Nominal or ordinal scaled variables.

Nonparametric statistics

- Most Nonparametrics are still concerned with populations, but the hypotheses are not formally stated using population values.
  - Nominal or ordinal scaled variables.
  - Few if any assumptions.

Nonparametric statistics

- Most Nonparametrics are still concerned with populations, but the hypotheses are not formally stated using population values.
  - Nominal or ordinal scaled variables.
  - Few if any assumptions.
  - Sometimes called distribution-free tests because, they do not make assumptions about a population distribution.

RSS
Research and Statistical Support

Nonparametric statistics

- Most Nonparametrics are still concerned with populations, but the hypotheses are not formally stated using population values.
  - Nominal or ordinal scaled variables.
  - Few if any assumptions.
  - Sometimes called distribution-free tests because, they do not make assumptions about a population distribution.
- Unfortunately, nonparametric tests tend to have less power or sensitivity to detect significance than their parametric partners.

Chi-square test introduction

- The chi-square test has two forms.

Chi-square test introduction

- The chi-square test has two forms.
  - Chi-square **Goodness-of-Fit** which tests whether or not the sample data fit the hypothesized population proportions.

Chi-square test introduction

- The chi-square test has two forms.
  - Chi-square **Goodness-of-Fit** which tests whether or not the sample data fit the hypothesized population proportions.
  - Chi-square **Test of Independence** which tests for the presence or absence of a relationship between two variables.

Chi-square test introduction

- The chi-square test has two forms.
  - Chi-square **Goodness-of-Fit** which tests whether or not the sample data fit the hypothesized population proportions.
  - Chi-square **Test of Independence** which tests for the presence or absence of a relationship between two variables.
- Both Chi-square tests use the same formula and are based on the distribution of Chi-square.

Chi-square test introduction

- The chi-square test has two forms.
  - Chi-square **Goodness-of-Fit** which tests whether or not the sample data fit the hypothesized population proportions.
  - Chi-square **Test of Independence** which tests for the presence or absence of a relationship between two variables.
- Both Chi-square tests use the same formula and are based on the distribution of Chi-square.
  - Symbol: χ^2

RSS
Research and Statistical Support

Chi-square distribution

- The χ^2 distribution is not normal (i.e. normally distributed); it is positively skewed.

RSS
Research and Statistical Support

Chi-square distribution

- The χ^2 distribution is not normal (i.e. normally distributed); it is positively skewed.

`http://www.medcalc.be/manual/chi-square-table.php`

RSS
Research and Statistical Support

Chi-square distribution

- The χ^2 distribution is not normal (i.e. normally distributed); it is positively skewed.

http://www.medcalc.be/manual/chi-square-table.php

- Notice in the table linked above, the sixth column corresponds to a significance level of 0.05 where:

RSS
Research and Statistical Support

Chi-square distribution

- The χ^2 distribution is not normal (i.e. normally distributed); it is positively skewed.

http://www.medcalc.be/manual/chi-square-table.php

- Notice in the table linked above, the sixth column corresponds to a significance level of 0.05 where:
    - The first column is degrees of freedom (*df*)

Core of Chi-square

- The core idea of any chi-square test is the comparison of **Observed** versus **Expected** frequencies.

Core of Chi-square

- The core idea of any chi-square test is the comparison of **Observed** versus **Expected** frequencies.
- The general formula is:

Core of Chi-square

- The core idea of any chi-square test is the comparison of **Observed** versus **Expected** frequencies.
- The general formula is:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Core of Chi-square

- The core idea of any chi-square test is the comparison of **Observed** versus **Expected** frequencies.
- The general formula is:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

- Where O is the observed frequency, E is the expected frequency.

Core of Chi-square

- The core idea of any chi-square test is the comparison of **Observed** versus **Expected** frequencies.
- The general formula is:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

- Where O is the observed frequency, E is the expected frequency.
  - E is the frequency expected *if the null hypothesis were true*.

RSS
Research and Statistical Support

One-way Classification Table Example

- The One-way Chi-square test is the *Goodness-of-fit* test.

One-way Classification Table Example

- The One-way Chi-square test is the *Goodness-of-fit* test.
- Say we randomly picked 100 students walking into the University Administration building.

RSS
Research and Statistical Support

One-way Classification Table Example

- The One-way Chi-square test is the *Goodness-of-fit* test.
- Say we randomly picked 100 students walking into the University Administration building.
- We would expect, because they were picked at random, that an equal number of those students would be Freshman, Sophomore, Junior, and Senior levels.

One-way Classification Table Example

- The One-way Chi-square test is the *Goodness-of-fit* test.
- Say we randomly picked 100 students walking into the University Administration building.
- We would expect, because they were picked at random, that an equal number of those students would be Freshman, Sophomore, Junior, and Senior levels.
  - We would expect 25 Freshmen, 25 Sophomores, 25 Juniors, and 25 Seniors.

One-way Classification Table Example

- The One-way Chi-square test is the *Goodness-of-fit* test.
- Say we randomly picked 100 students walking into the University Administration building.
- We would expect, because they were picked at random, that an equal number of those students would be Freshman, Sophomore, Junior, and Senior levels.
  - We would expect 25 Freshmen, 25 Sophomores, 25 Juniors, and 25 Seniors.
  - The null hypothesis would be: $H_0 : E = O$

One-way Classification Table Example

- The One-way Chi-square test is the *Goodness-of-fit* test.
- Say we randomly picked 100 students walking into the University Administration building.
- We would expect, because they were picked at random, that an equal number of those students would be Freshman, Sophomore, Junior, and Senior levels.
  - We would expect 25 Freshmen, 25 Sophomores, 25 Juniors, and 25 Seniors.
  - The null hypothesis would be: $H_0 : E = O$
  - The alternative hypothesis: $H_1 : E \neq O$

One-way Classification Table Example

- The One-way Chi-square test is the *Goodness-of-fit* test.
- Say we randomly picked 100 students walking into the University Administration building.
- We would expect, because they were picked at random, that an equal number of those students would be Freshman, Sophomore, Junior, and Senior levels.
  - We would expect 25 Freshmen, 25 Sophomores, 25 Juniors, and 25 Seniors.
  - The null hypothesis would be: $H_0 : E = O$
  - The alternative hypothesis: $H_1 : E \neq O$
- Instead, we found: 32 Freshmen, 28 Sophomores, 23 Juniors, and 17 Seniors.

RSS Research and Statistical Support

One-way Classification Table Example

- The One-way Chi-square test is the *Goodness-of-fit* test.
- Say we randomly picked 100 students walking into the University Administration building.
- We would expect, because they were picked at random, that an equal number of those students would be Freshman, Sophomore, Junior, and Senior levels.
  - We would expect 25 Freshmen, 25 Sophomores, 25 Juniors, and 25 Seniors.
  - The null hypothesis would be: $H_0 : E = O$
  - The alternative hypothesis: $H_1 : E \neq O$
- Instead, we found: 32 Freshmen, 28 Sophomores, 23 Juniors, and 17 Seniors.
- This study design constitutes a one-way classification table because, there is only one variable (class level) with multiple categories.

The One-way Classification Table

	Freshmen	Sophomore	Junior	Senior
Observed	32	28	23	17
Expected	25	25	25	25

The One-way Classification Table

	Freshmen	Sophomore	Junior	Senior
Observed	32	28	23	17
Expected	25	25	25	25

- Degrees of Freedom (*df*) is the number of Categories or Columns minus 1.

RSS
Research and Statistical Support

The One-way Classification Table

	Freshmen	Sophomore	Junior	Senior
Observed	32	28	23	17
Expected	25	25	25	25

- Degrees of Freedom (*df*) is the number of Categories or Columns minus 1.
- $df = C - 1 = 4 - 1 = 3$

http://www.medcalc.be/manual/chi-square-table.php

RSS Research and Statistical Support

Calculate Chi-square

- Using the formula from above,

Calculate Chi-square

- Using the formula from above,

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Research and Statistical Support

Calculate Chi-square

- Using the formula from above,

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$\chi^2 = \frac{(32-25)^2}{25} + \frac{(28-25)^2}{25} + \frac{(23-25)^2}{25} + \frac{(17-25)^2}{25} = 5.04$$

Calculate Chi-square

- Using the formula from above,

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$\chi^2 = \frac{(32-25)^2}{25} + \frac{(28-25)^2}{25} + \frac{(23-25)^2}{25} + \frac{(17-25)^2}{25} = 5.04$$

- And since $\chi^2_{calc} = 5.04 < 7.815 = \chi^2_{crit}$ we fail to reject the null hypothesis and conclude that this sample does not indicate a significant difference between the observed and expected frequencies of class level.

RSS
Research and Statistical Support

Multi-way Chi-square

- When we have more than one categorical variable, we call the chi-square test a test of *Independence*.

Multi-way Chi-square

- When we have more than one categorical variable, we call the chi-square test a test of *Independence*.
  - Are the cells of the table *Independent* of one another, or is there some relationship occurring among them.

RSS
Research and Statistical Support

Multi-way Chi-square

- When we have more than one categorical variable, we call the chi-square test a test of *Independence*.
  - Are the cells of the table *Independent* of one another, or is there some relationship occurring among them.
- In the one-way example above, we called the table a *classification table* because we were classifying frequencies on one variable.

Multi-way Chi-square

- When we have more than one categorical variable, we call the chi-square test a test of *Independence*.
  - Are the cells of the table *Independent* of one another, or is there some relationship occurring among them.
- In the one-way example above, we called the table a *classification table* because we were classifying frequencies on one variable.
- In the multi-way situation, we call the table a *contingency table* because, the frequencies of one variable are contingent upon another (or more than one) variable.

RSS
Research and Statistical Support

A Two-way Example

- Suppose we wondered about the gender frequency of students entering the UNT Administration building from above?

A Two-way Example

- Suppose we wondered about the gender frequency of students entering the UNT Administration building from above?
- A 2 X 4 design (Gender by Class Level).

A Two-way Example

- Suppose we wondered about the gender frequency of students entering the UNT Administration building from above?
- A 2 X 4 design (Gender by Class Level).

	Class		Level		
Gender	Freshmen	Sophomore	Junior	Senior	Total
Male	32	28	23	17	100
Female	28	29	20	15	92
Total	60	57	43	32	192

RSS
Research and Statistical Support

Expected Frequencies in a Two-way Design

- In the one-way design, expected frequencies were simply even proportions; but here, with a more complex design, we must calculate the expected frequencies which are *contingent* upon two variables.

Expected Frequencies in a Two-way Design

- In the one-way design, expected frequencies were simply even proportions; but here, with a more complex design, we must calculate the expected frequencies which are *contingent* upon two variables.
- The basic equation for calculating the Expected frequencies is:

Expected Frequencies in a Two-way Design

- In the one-way design, expected frequencies were simply even proportions; but here, with a more complex design, we must calculate the expected frequencies which are *contingent* upon two variables.
- The basic equation for calculating the Expected frequencies is:

$$E_{ij} = \frac{R_i C_j}{n_t}$$

Expected Frequencies in a Two-way Design

- In the one-way design, expected frequencies were simply even proportions; but here, with a more complex design, we must calculate the expected frequencies which are *contingent* upon two variables.
- The basic equation for calculating the Expected frequencies is:

$$E_{ij} = \frac{R_i C_j}{n_t}$$

- Where E_{ij} is a particular cell, R_i is the row total, C_j is the column total, and n_t is the total number of individuals (or cases).

RSS
Research and Statistical Support

Expected Frequencies for the current example

- For the current example, we have the following Expected frequencies for each cell:

Expected Frequencies for the current example

- For the current example, we have the following Expected frequencies for each cell:

$$E_{11} = \frac{100 \times 60}{192} \quad E_{12} = \frac{100 \times 57}{192} \quad E_{13} = \frac{100 \times 43}{192} \quad E_{14} = \frac{100 \times 32}{192}$$

$$E_{21} = \frac{92 \times 60}{192} \quad E_{22} = \frac{92 \times 57}{192} \quad E_{23} = \frac{92 \times 43}{192} \quad E_{24} = \frac{92 \times 32}{192}$$

Expected Frequencies for the current example

- For the current example, we have the following Expected frequencies for each cell:

$$E_{11} = \frac{100 \times 60}{192} \quad E_{12} = \frac{100 \times 57}{192} \quad E_{13} = \frac{100 \times 43}{192} \quad E_{14} = \frac{100 \times 32}{192}$$

$$E_{21} = \frac{92 \times 60}{192} \quad E_{22} = \frac{92 \times 57}{192} \quad E_{23} = \frac{92 \times 43}{192} \quad E_{24} = \frac{92 \times 32}{192}$$

- Which leads to:

$$E_{11} = 31.25 \quad E_{12} = 29.69 \quad E_{13} = 22.40 \quad E_{14} = 16.67$$

$$E_{21} = 28.75 \quad E_{22} = 27.32 \quad E_{23} = 20.60 \quad E_{24} = 15.33$$

RSS
Research and Statistical Support

Table with Expected Frequencies

- Here we have the Expected Frequencies for each cell, listed in parentheses.

	Class		Level		
Gender	Freshmen	Sophomore	Junior	Senior	Total
Male	32(31.25)	28(29.69)	23(22.40)	17(16.67)	100
Female	28(28.75)	29(27.32)	20(20.60)	15(15.33)	92
Total	60	57	43	32	192

- Of course, you can not have 31.25 persons (frequencies), so you could round to the nearest whole number.

RSS
Research and Statistical Support

Calculating χ^2 for the two-way example

- Recall the formula for $\chi^2 = \sum \frac{(O-E)^2}{E} =$

$$\frac{(32-32.25)^2}{31.25} + \frac{(28-29.69)^2}{29.69} + \frac{(23-22.40)^2}{22.40} + \frac{(17-16.67)^2}{16.67} +$$

$$\frac{(28-28.75)^2}{28.75} + \frac{(29-27.32)^2}{27.32} + \frac{(20-20.60)^2}{20.60} + \frac{(15-15.33)^2}{15.33} =$$

$$\chi^2 = 0.286$$

RSS Research and Statistical Support

Degrees of Freedom in Two-way designs

- Recall, earlier we said $df = C - 1$

Degrees of Freedom in Two-way designs

- Recall, earlier we said $df = C - 1$
- In the Two-way situation, we have rows *and* columns.

Degrees of Freedom in Two-way designs

- Recall, earlier we said $df = C - 1$
- In the Two-way situation, we have rows *and* columns.
- So, $df = (R - 1)(C - 1)$

Degrees of Freedom in Two-way designs

- Recall, earlier we said $df = C - 1$
- In the Two-way situation, we have rows *and* columns.
- So, $df = (R - 1)(C - 1)$
  - Where R = the number of rows and C = the number of columns.

Degrees of Freedom in Two-way designs

- Recall, earlier we said $df = C - 1$
- In the Two-way situation, we have rows *and* columns.
- So, $df = (R - 1)(C - 1)$
  - Where R = the number of rows and C = the number of columns.
- For the current example:

Degrees of Freedom in Two-way designs

- Recall, earlier we said $df = C - 1$
- In the Two-way situation, we have rows *and* columns.
- So, $df = (R - 1)(C - 1)$
  - Where R = the number of rows and C = the number of columns.
- For the current example:
  - $df = (2 - 1)(4 - 1) = 3$

  http://www.medcalc.be/manual/chi-square-table.php

Degrees of Freedom in Two-way designs

- Recall, earlier we said $df = C - 1$
- In the Two-way situation, we have rows *and* columns.
- So, $df = (R - 1)(C - 1)$
  - Where R = the number of rows and C = the number of columns.
- For the current example:
  - $df = (2 - 1)(4 - 1) = 3$

  http://www.medcalc.be/manual/chi-square-table.php
- So, our $\chi^2_{crit} = 7.815$ is the same.

RSS
Research and Statistical Support

Two-way Example Results

- So our $\chi^2_{calc} = 0.286 < 7.185 = \chi^2_{crit}$ we fail to reject the null hypothesis and we conclude that there was not a relationship between Gender and Class Level.

RSS
Research and Statistical Support

Two-way Example Results

- So our $\chi^2_{calc} = 0.286 < 7.185 = \chi^2_{crit}$ we fail to reject the null hypothesis and we conclude that there was not a relationship between Gender and Class Level.
- Stated another way, the two variables were not independent of one another.

Two-way Example Results

- So our $\chi^2_{calc} = 0.286 < 7.185 = \chi^2_{crit}$ we fail to reject the null hypothesis and we conclude that there was not a relationship between Gender and Class Level.
- Stated another way, the two variables were not independent of one another.
- Stated still another way, the Observed frequencies for each cell did not differ significantly from the Expected frequencies.

Two-way Example Results

- So our $\chi^2_{calc} = 0.286 < 7.185 = \chi^2_{crit}$ we fail to reject the null hypothesis and we conclude that there was not a relationship between Gender and Class Level.
- Stated another way, the two variables were not independent of one another.
- Stated still another way, the Observed frequencies for each cell did not differ significantly from the Expected frequencies.
- Like with correlation, chi-square is very sensitive to sample size.

RSS
Research and Statistical Support

Two-way Example Results

- So our $\chi^2_{calc} = 0.286 < 7.185 = \chi^2_{crit}$ we fail to reject the null hypothesis and we conclude that there was not a relationship between Gender and Class Level.
- Stated another way, the two variables were not independent of one another.
- Stated still another way, the Observed frequencies for each cell did not differ significantly from the Expected frequencies.
- Like with correlation, chi-square is very sensitive to sample size.
  - If given a large enough sample, any chi-square analysis will be significant.

RSS
Research and Statistical Support

Odds as Effect Size

- We can calculate the **odds** of cell membership as a measure of effect size which allows us to go beyond the simple hypothesis testing context.

Odds as Effect Size

- We can calculate the **odds** of cell membership as a measure of effect size which allows us to go beyond the simple hypothesis testing context.
- In order to calculate the odds for a given cell, we must identify the cell in our question.

RSS
Research and Statistical Support

Odds as Effect Size

- We can calculate the **odds** of cell membership as a measure of effect size which allows us to go beyond the simple hypothesis testing context.
- In order to calculate the odds for a given cell, we must identify the cell in our question.
  - For example, if you are a Male entering the UNT Administration building, what are the odds you are a Freshman?

Odds as Effect Size

- We can calculate the **odds** of cell membership as a measure of effect size which allows us to go beyond the simple hypothesis testing context.
- In order to calculate the odds for a given cell, we must identify the cell in our question.
  - For example, if you are a Male entering the UNT Administration building, what are the odds you are a Freshman?
  - To answer that question, simply divide the number of Freshmen by the number of *not Freshmen* for the Male row.

RSS
Research and Statistical Support

Odds as Effect Size

- We can calculate the **odds** of cell membership as a measure of effect size which allows us to go beyond the simple hypothesis testing context.
- In order to calculate the odds for a given cell, we must identify the cell in our question.
  - For example, if you are a Male entering the UNT Administration building, what are the odds you are a Freshman?
  - To answer that question, simply divide the number of Freshmen by the number of *not Freshmen* for the Male row.
  - Odds of a male also being a Freshman: $\frac{32}{68} = 0.4706$ or nearly 50/50 odds.

RSS
Research and Statistical Support

Odds as Effect Size

- We can calculate the **odds** of cell membership as a measure of effect size which allows us to go beyond the simple hypothesis testing context.
- In order to calculate the odds for a given cell, we must identify the cell in our question.
  - For example, if you are a Male entering the UNT Administration building, what are the odds you are a Freshman?
  - To answer that question, simply divide the number of Freshmen by the number of *not Freshmen* for the Male row.
  - Odds of a male also being a Freshman: $\frac{32}{68} = 0.4706$ or nearly 50/50 odds.
- Stated another way: there is a 47.06% chance a male entering the building is also a Freshman.

RSS
Research and Statistical Support

Phi as 2 X 2 Contingency Effect Size

- When in the 2 X 2 situation, Phi can be used to measure the association of the two variables.

RSS
Research and Statistical Support

Phi as 2 X 2 Contingency Effect Size

- When in the 2 X 2 situation, Phi can be used to measure the association of the two variables.
- Symbol: ϕ

Phi as 2 X 2 Contingency Effect Size

- When in the 2 X 2 situation, Phi can be used to measure the association of the two variables.
- Symbol: ϕ
- Calculation:

RSS
Research and Statistical Support

Phi as 2 X 2 Contingency Effect Size

- When in the 2 X 2 situation, Phi can be used to measure the association of the two variables.
- Symbol: ϕ
- Calculation:

$$\phi = \sqrt{\frac{\chi^2}{n_t}}$$

Phi as 2 X 2 Contingency Effect Size

- When in the 2 X 2 situation, Phi can be used to measure the association of the two variables.
- Symbol: ϕ
- Calculation:

$$\phi = \sqrt{\frac{\chi^2}{n_t}}$$

- The resulting number will be a correlation coefficient and is interpreted as such.

Phi as 2 X 2 Contingency Effect Size

- When in the 2 X 2 situation, Phi can be used to measure the association of the two variables.
- Symbol: ϕ
- Calculation:

$$\phi = \sqrt{\frac{\chi^2}{n_t}}$$

- The resulting number will be a correlation coefficient and is interpreted as such.
- Of course, it is limited to the 2 X 2 situation only.

Cramer's V

- Cramer's V is used as an analog to Phi but for contingency tables larger than 2 X 2.

Cramer's V

- Cramer's V is used as an analog to Phi but for contingency tables larger than 2 X 2.
- The forumula is:

Cramer's V

- Cramer's V is used as an analog to Phi but for contingency tables larger than 2 X 2.
- The forumula is:

$$V = \sqrt{\frac{\chi^2}{n_t(k-1)}}$$

RSS
Research and Statistical Support

Cramer's V

- Cramer's V is used as an analog to Phi but for contingency tables larger than 2 X 2.
- The forumula is:

$$V = \sqrt{\frac{\chi^2}{n_t(k-1)}}$$

- Where k is the smaller of: number of rows or number of columns.

RSS
Research and Statistical Support

Cramer's V

- Cramer's V is used as an analog to Phi but for contingency tables larger than 2 X 2.
- The forumula is:

$$V = \sqrt{\frac{\chi^2}{n_t(k-1)}}$$

- Where k is the smaller of: number of rows or number of columns.
- A note of caution regarding Phi and Cramer's V. Interpreting a correlation among two strictly categorical variables is essentially meaningless.

Cramer's V

- Cramer's V is used as an analog to Phi but for contingency tables larger than 2 X 2.
- The forumula is:

$$V = \sqrt{\frac{\chi^2}{n_t(k-1)}}$$

- Where k is the smaller of: number of rows or number of columns.
- A note of caution regarding Phi and Cramer's V. Interpreting a correlation among two strictly categorical variables is essentially meaningless.
  - What does it mean to say that class standing level and gender are (or are not) correlated at .60?

Cramer's V

- Cramer's V is used as an analog to Phi but for contingency tables larger than 2 X 2.
- The forumula is:

$$V = \sqrt{\frac{\chi^2}{n_t(k-1)}}$$

- Where k is the smaller of: number of rows or number of columns.
- A note of caution regarding Phi and Cramer's V. Interpreting a correlation among two strictly categorical variables is essentially meaningless.
  - What does it mean to say that class standing level and gender are (or are not) correlated at .60?
  - NOT MUCH!

RSS
Research and Statistical Support

Cohen's kappa

- Cohen's kappa is a measure of *agreement*.

Cohen's kappa

- Cohen's kappa is a measure of *agreement*.
- Suppose we have two tenured faculty members rate 28 graduate students' teaching effectiveness.

Cohen's kappa

- Cohen's kappa is a measure of *agreement*.
- Suppose we have two tenured faculty members rate 28 graduate students' teaching effectiveness.
  - Not Effective, Effective, Highly Effective

Cohen's kappa

- Cohen's kappa is a measure of *agreement*.
- Suppose we have two tenured faculty members rate 28 graduate students' teaching effectiveness.
  - Not Effective, Effective, Highly Effective
- It would be beneficial to know if both faculty agree on the ratings; or to what extent do they agree or disagree.

Cohen's kappa

- Cohen's kappa is a measure of *agreement*.
- Suppose we have two tenured faculty members rate 28 graduate students' teaching effectiveness.
  - Not Effective, Effective, Highly Effective
- It would be beneficial to know if both faculty agree on the ratings; or to what extent do they agree or disagree.
- One could simply calculate the percentage of agreement, but that measure does not take into account the random chance of agreement.

Cohen’s kappa

- Cohen’s kappa is a measure of *agreement*.
- Suppose we have two tenured faculty members rate 28 graduate students’ teaching effectiveness.
  - Not Effective, Effective, Highly Effective
- It would be beneficial to know if both faculty agree on the ratings; or to what extent do they agree or disagree.
- One could simply calculate the percentage of agreement, but that measure does not take into account the random chance of agreement.
- Cohen’s kappa corrects this deficiency.

Agreement Data

NE = Not Effective, E = Effective, HE = Highly Effective.

	Faculty 1			
Faculty 2	NE	E	HE	Total
NE	4	0	0	4
E	0	5	1	6
HE	0	3	15	18
Total	4	8	16	28

RSS
Research and Statistical Support

Percentage of Agreement and Random Chance

- Of the 28 graduate students, 24 were rated the same by both faculty (add along the diagonal).

Percentage of Agreement and Random Chance

- Of the 28 graduate students, 24 were rated the same by both faculty (add along the diagonal).
    - This means, $24/28 = .8571$ or 85.71% agreement.

Percentage of Agreement and Random Chance

- Of the 28 graduate students, 24 were rated the same by both faculty (add along the diagonal).
  - This means, $24/28 = .8571$ or 85.71% agreement.
- However, consider the following:

RSS
Research and Statistical Support

Percentage of Agreement and Random Chance

- Of the 28 graduate students, 24 were rated the same by both faculty (add along the diagonal).
  - This means, $24/28 = .8571$ or 85.71% agreement.
- However, consider the following:
  - The probability of 'Effective' for Faculty 1 is 8/28 = .2857.

RSS Research and Statistical Support

Percentage of Agreement and Random Chance

- Of the 28 graduate students, 24 were rated the same by both faculty (add along the diagonal).
  - This means, $24/28 = .8571$ or 85.71% agreement.
- However, consider the following:
  - The probability of 'Effective' for Faculty 1 is 8/28 = .2857.
  - The probability of 'Effective' for Faculty 2 is 6/28 = .2143.

Percentage of Agreement and Random Chance

- Of the 28 graduate students, 24 were rated the same by both faculty (add along the diagonal).
  - This means, $24/28 = .8571$ or 85.71% agreement.
- However, consider the following:
  - The probability of 'Effective' for Faculty 1 is 8/28 = .2857.
  - The probability of 'Effective' for Faculty 2 is 6/28 = .2143.
  - So, the probability of both faculty agreeing on 'Effective' for one student is .2857*.2143 = .0612.

RSS
Research and Statistical Support

Percentage of Agreement and Random Chance

- Of the 28 graduate students, 24 were rated the same by both faculty (add along the diagonal).
    - This means, $24/28 = .8571$ or 85.71% agreement.
- However, consider the following:
    - The probability of 'Effective' for Faculty 1 is 8/28 = .2857.
    - The probability of 'Effective' for Faculty 2 is 6/28 = .2143.
    - So, the probability of both faculty agreeing on 'Effective' for one student is .2857*.2143 = .0612.
    - Which is not a lot, but across all 28 students, we can expect .0612*28 = 1.71 agreements just by random chance.

RSS
Research and Statistical Support

Calculate kappa

- Calculating kappa is similar to calculating the usual χ^2.

Calculate kappa

- Calculating kappa is similar to calculating the usual χ^2.
- The equation for kappa (κ) is:

Calculate kappa

- Calculating kappa is similar to calculating the usual χ^2.
- The equation for kappa (κ) is:

$$\kappa = \frac{\sum f_o - \sum f_e}{n_t - \sum f_e}$$

Calculate kappa

- Calculating kappa is similar to calculating the usual χ^2.
- The equation for kappa (κ) is:

$$\kappa = \frac{\sum f_o - \sum f_e}{n_t - \sum f_e}$$

- Where f_o is the observed frequencies on the diagonal and f_e is the expected frequencies on the diagonal.

Calculating the Expected Frequencies

- Use the same formula from earlier to calculate the Expected Frequencies:

Calculating the Expected Frequencies

- Use the same formula from earlier to calculate the Expected Frequencies:

$$E_{ij} = \frac{R_i C_j}{n_t}$$

Calculating the Expected Frequencies

- Use the same formula from earlier to calculate the Expected Frequencies:

$$E_{ij} = \frac{R_i C_j}{n_t}$$

- For Not Effective (NE): (4*4)/28 = .571

RSS Research and Statistical Support

Calculating the Expected Frequencies

- Use the same formula from earlier to calculate the Expected Frequencies:

$$E_{ij} = \frac{R_i C_j}{n_t}$$

- For Not Effective (NE): (4*4)/28 = .571
- For Effective (E): (6*8)/28 = 1.714

RSS Research and Statistical Support

Calculating the Expected Frequencies

- Use the same formula from earlier to calculate the Expected Frequencies:

$$E_{ij} = \frac{R_i C_j}{n_t}$$

- For Not Effective (NE): (4*4)/28 = .571
- For Effective (E): (6*8)/28 = 1.714
- For Highly Effective (HE): (18*16)/28 = 10.286

Calculating the Expected Frequencies

- Use the same formula from earlier to calculate the Expected Frequencies:

$$E_{ij} = \frac{R_i C_j}{n_t}$$

- For Not Effective (NE): (4*4)/28 = .571
- For Effective (E): (6*8)/28 = 1.714
- For Highly Effective (HE): (18*16)/28 = 10.286
- Then, sum them to get f_e = 12.571

RSS Research and Statistical Support

Calculating the Observed Frequencies

- Simply add up the observed frequencies to get f_o

Calculating the Observed Frequencies

- Simply add up the observed frequencies to get f_o

$$4 + 5 + 15 = 24$$

- Now we can calculate kappa.

Calculating kappa

- Recall, kappa (κ) is:

Calculating kappa

- Recall, kappa (κ) is:

$$\kappa = \frac{\sum f_o - \sum f_e}{n_t - \sum f_e}$$

Calculating kappa

- Recall, kappa (κ) is:

$$\kappa = \frac{\sum f_o - \sum f_e}{n_t - \sum f_e}$$

- So, for the current example:

Research and Statistical Support

Calculating kappa

- Recall, kappa (κ) is:

$$\kappa = \frac{\sum f_o - \sum f_e}{n_t - \sum f_e}$$

- So, for the current example:

$$\kappa = \frac{\sum f_o - \sum f_e}{n_t - \sum f_e} = \frac{24 - 12.571}{28 - 12.571} = .7407$$

RSS Research and Statistical Support

Calculating kappa

- Recall, kappa (κ) is:

$$\kappa = \frac{\sum f_o - \sum f_e}{n_t - \sum f_e}$$

- So, for the current example:

$$\kappa = \frac{\sum f_o - \sum f_e}{n_t - \sum f_e} = \frac{24 - 12.571}{28 - 12.571} = .7407$$

- So, agreement is really lower than the 85.71% from above; after accounting for chance it is 74.07%.

Wilcoxon's Rank-Sum test

- Wilcoxon's Rank-Sum test is a non-parametric replacement for the Independent Samples *t* test.

Wilcoxon's Rank-Sum test

- Wilcoxon's Rank-Sum test is a non-parametric replacement for the Independent Samples *t* test.
- When data do not conform to the assumptions of the *t* test, Wilcoxon's Rank-Sum test is an appropriate alternative.

Wilcoxon's Rank-Sum test

- Wilcoxon's Rank-Sum test is a non-parametric replacement for the Independent Samples *t* test.
- When data do not conform to the assumptions of the *t* test, Wilcoxon's Rank-Sum test is an appropriate alternative.
- However, as mentioned previously, non-parametric tests tend to have less power than their parametric companions.

Wilcoxon's Rank-Sum test

- Wilcoxon's Rank-Sum test is a non-parametric replacement for the Independent Samples *t* test.
- When data do not conform to the assumptions of the *t* test, Wilcoxon's Rank-Sum test is an appropriate alternative.
- However, as mentioned previously, non-parametric tests tend to have less power than their parametric companions.
    - The Rank-Sum test has less power than the Independent Samples *t* test.

RSS
Research and Statistical Support

Wilcoxon's Rank-Sum test

- Wilcoxon's Rank-Sum test is a non-parametric replacement for the Independent Samples *t* test.
- When data do not conform to the assumptions of the *t* test, Wilcoxon's Rank-Sum test is an appropriate alternative.
- However, as mentioned previously, non-parametric tests tend to have less power than their parametric companions.
  - The Rank-Sum test has less power than the Independent Samples *t* test.
- The general idea of the Rank-Sum test is to test whether two samples originated with the same population, similar to the Independent Samples *t* test.

Wilcoxon's Rank-Sum test

- Wilcoxon's Rank-Sum test is a non-parametric replacement for the Independent Samples *t* test.
- When data do not conform to the assumptions of the *t* test, Wilcoxon's Rank-Sum test is an appropriate alternative.
- However, as mentioned previously, non-parametric tests tend to have less power than their parametric companions.
  - The Rank-Sum test has less power than the Independent Samples *t* test.
- The general idea of the Rank-Sum test is to test whether two samples originated with the same population, similar to the Independent Samples *t* test.
  - However, it is not specifically tied to *mean* differences, but rather; differences in central tendency.

RSS
Research and Statistical Support

Ranked Sums

- If we rank the scores of two groups from lowest to highest, then sum the groups' ranked scores...

Ranked Sums

- If we rank the scores of two groups from lowest to highest, then sum the groups' ranked scores...
- We would expect, **if the groups are different**, to find the sum of one group to be smaller than the sum of the other group.

RSS
Research and Statistical Support

Ranked Sums

- If we rank the scores of two groups from lowest to highest, then sum the groups' ranked scores...
- We would expect, **if the groups are different**, to find the sum of one group to be smaller than the sum of the other group.
- As a significance test, we take the sum of the ranks for the smaller group and compare it to a tabled value to determine if the groups are significantly different.

Ranked Sums

- If we rank the scores of two groups from lowest to highest, then sum the groups' ranked scores...
- We would expect, **if the groups are different**, to find the sum of one group to be smaller than the sum of the other group.
- As a significance test, we take the sum of the ranks for the smaller group and compare it to a tabled value to determine if the groups are significantly different.
  - If the groups are equal size, then use the smaller of the two ranked sums.

RSS
Research and Statistical Support

Ranked Sums

- If we rank the scores of two groups from lowest to highest, then sum the groups' ranked scores...
- We would expect, **if the groups are different**, to find the sum of one group to be smaller than the sum of the other group.
- As a significance test, we take the sum of the ranks for the smaller group and compare it to a tabled value to determine if the groups are significantly different.
  - If the groups are equal size, then use the smaller of the two ranked sums.

http://www.unt.edu/rss/class/Jon/ISSS_SC/Module011/ws_tables/

RSS Research and Statistical Support

Small Example

- Say we have two groups' Driving Anger scores; a group of police officers and a group of taxi drivers.

Small Example

- Say we have two groups' Driving Anger scores; a group of police officers and a group of taxi drivers.
- The police officers' scores are: 8, 15, 12, 10, 13

Small Example

- Say we have two groups' Driving Anger scores; a group of police officers and a group of taxi drivers.
- The police officers' scores are: 8, 15, 12, 10, 13
- The taxi drivers' scores are: 27, 28, 19, 17, 26, 28

Small Example

- Say we have two groups' Driving Anger scores; a group of police officers and a group of taxi drivers.
- The police officers' scores are: 8, 15, 12, 10, 13
- The taxi drivers' scores are: 27, 28, 19, 17, 26, 28
- We would expect police officers to have a lower level of Driving Anger than the Taxi drivers.

RSS
Research and Statistical Support

Small Example

- Say we have two groups' Driving Anger scores; a group of police officers and a group of taxi drivers.
- The police officers' scores are: 8, 15, 12, 10, 13
- The taxi drivers' scores are: 27, 28, 19, 17, 26, 28
- We would expect police officers to have a lower level of Driving Anger than the Taxi drivers.
  - One-tailed test: police officers $<$ taxi drivers.

RSS Research and Statistical Support

Small Example

- Say we have two groups' Driving Anger scores; a group of police officers and a group of taxi drivers.
- The police officers' scores are: 8, 15, 12, 10, 13
- The taxi drivers' scores are: 27, 28, 19, 17, 26, 28
- We would expect police officers to have a lower level of Driving Anger than the Taxi drivers.
  - One-tailed test: police officers $<$ taxi drivers.
- To test this we will first rank all the scores.

RSS
Research and Statistical Support

Ranked Data

	Raw Scores	Rank
Police	8	1
Officers	15	5
	12	3
	10	2
	13	4
Taxi	27	9
Drivers	28	10.5
	19	7
	17	6
	26	8
	28	10.5

Tied scores get tied ranks half-way between the two whole number ranks they would occupy if sequential.

RSS Research and Statistical Support

Calculate W_s

- Sum the Ranks for the **smaller** group, the police officers: $\sum R_s = 1 + 5 + 3 + 2 + 4 = 15$

RSS
Research and Statistical Support

Calculate W_s

- Sum the Ranks for the **smaller** group, the police officers: $\sum R_s = 1 + 5 + 3 + 2 + 4 = 15$
- Look in the table for the critical value of W_s with a significance level of 0.05 and:

RSS
Research and Statistical Support

Calculate W_s

- Sum the Ranks for the **smaller** group, the police officers: $\sum R_s = 1 + 5 + 3 + 2 + 4 = 15$
- Look in the table for the critical value of W_s with a significance level of 0.05 and:
  - n_1 = smaller group = 5
  - n_2 = larger group = 6

RSS Research and Statistical Support

Calculate W_s

- Sum the Ranks for the **smaller** group, the police officers: $\sum R_s = 1 + 5 + 3 + 2 + 4 = 15$
- Look in the table for the critical value of W_s with a significance level of 0.05 and:
  - n_1 = smaller group = 5
  - n_2 = larger group = 6

http://www.unt.edu/rss/class/Jon/ISSS_SC/Module011/ws_tables/

RSS Research and Statistical Support

Calculate W_s

- Sum the Ranks for the **smaller** group, the police officers: $\sum R_s = 1 + 5 + 3 + 2 + 4 = 15$
- Look in the table for the critical value of W_s with a significance level of 0.05 and:
  - n_1 = smaller group = 5
  - n_2 = larger group = 6

http://www.unt.edu/rss/class/Jon/ISSS_SC/Module011/ws_tables/

- Since our calculated $W_s = 15 < 20 = W_s$ critical value; we reject the null hypothesis and conclude that the two groups *are* significantly different.

RSS
Research and Statistical Support

Caution

- It is important to note that the table of W_s displays Critical Lower-Tail Values of W_s, where $n_1 \leq n_2$.

Caution

- It is important to note that the table of W_s displays Critical Lower-Tail Values of W_s, where $n_1 \leq n_2$.
  - The calculated W_s needs to be less than or equal to the critical value in order to reject the null (i.e. find a significant difference).

RSS Research and Statistical Support

Caution

- It is important to note that the table of W_s displays Critical Lower-Tail Values of W_s, where $n_1 \leq n_2$.
  - The calculated W_s needs to be less than or equal to the critical value in order to reject the null (i.e. find a significant difference).
- If we wanted to test if the *Upper-Tail* was significant (i.e. hypothesize that the taxi drivers tend to score significantly *higher* than the police officers) we would need to calculate W_s'

RSS Research and Statistical Support

Caution

- It is important to note that the table of W_s displays Critical Lower-Tail Values of W_s, where $n_1 \leq n_2$.
  - The calculated W_s needs to be less than or equal to the critical value in order to reject the null (i.e. find a significant difference).
- If we wanted to test if the *Upper-Tail* was significant (i.e. hypothesize that the taxi drivers tend to score significantly *higher* than the police officers) we would need to calculate W_s'

$$W_s' = 2\overline{W} - W_s$$

RSS Research and Statistical Support

Caution

- It is important to note that the table of W_s displays Critical Lower-Tail Values of W_s, where $n_1 \leq n_2$.
  - The calculated W_s needs to be less than or equal to the critical value in order to reject the null (i.e. find a significant difference).
- If we wanted to test if the *Upper-Tail* was significant (i.e. hypothesize that the taxi drivers tend to score significantly *higher* than the police officers) we would need to calculate W_s'

$$W_s' = 2\overline{W} - W_s$$

- where $2\overline{W} = n_1\left(n_1 + n_2 + 1\right)$

Caution

- It is important to note that the table of W_s displays Critical Lower-Tail Values of W_s, where $n_1 \leq n_2$.
  - The calculated W_s needs to be less than or equal to the critical value in order to reject the null (i.e. find a significant difference).
- If we wanted to test if the *Upper-Tail* was significant (i.e. hypothesize that the taxi drivers tend to score significantly *higher* than the police officers) we would need to calculate W_s'

$$W_s' = 2\overline{W} - W_s$$

- where $2\overline{W} = n_1 (n_1 + n_2 + 1)$
  - Notice, the table provides $2\overline{W}$ in the right most column.

RSS
Research and Statistical Support

Caution

- It is important to note that the table of W_s displays Critical Lower-Tail Values of W_s, where $n_1 \leq n_2$.
  - The calculated W_s needs to be less than or equal to the critical value in order to reject the null (i.e. find a significant difference).
- If we wanted to test if the *Upper-Tail* was significant (i.e. hypothesize that the taxi drivers tend to score significantly *higher* than the police officers) we would need to calculate W_s'

$$W_s' = 2\overline{W} - W_s$$

- where $2\overline{W} = n_1 (n_1 + n_2 + 1)$
  - Notice, the table provides $2\overline{W}$ in the right most column.
- Then, if W_s' is larger than the critical value, we would reject the null and conclude that the taxi drivers scored significantly higher on the Driving Anger scale.

RSS Research and Statistical Support

Normal Approximation

- Notice the tables of W_s are only useful when n_1 and n_2 are less than or equal to 25.

Normal Approximation

- Notice the tables of W_s are only useful when n_1 and n_2 are less than or equal to 25.
- For larger samples, the distribution of W_s approaches normal; which means we can calculate a z score for them.

Normal Approximation

- Notice the tables of W_s are only useful when n_1 and n_2 are less than or equal to 25.
- For larger samples, the distribution of W_s approaches normal; which means we can calculate a z score for them.
  - The mean of the distribution of W_s is: $\frac{n_1(n_1+n_2+1)}{2}$

Normal Approximation

- Notice the tables of W_s are only useful when n_1 and n_2 are less than or equal to 25.
- For larger samples, the distribution of W_s approaches normal; which means we can calculate a z score for them.
  - The mean of the distribution of W_s is: $\frac{n_1(n_1+n_2+1)}{2}$
  - And the standard deviation of the distribution of W_s is: $\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}$

RSS
Research and Statistical Support

Normal Approximation

- Notice the tables of W_s are only useful when n_1 and n_2 are less than or equal to 25.
- For larger samples, the distribution of W_s approaches normal; which means we can calculate a z score for them.
  - The mean of the distribution of W_s is: $\frac{n_1(n_1+n_2+1)}{2}$
  - And the standard deviation of the distribution of W_s is: $\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}$
- So, the z score is calculated using:

$$z = \frac{statistic - mean}{standard\ deviation} = \frac{W_s - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}}$$

RSS
Research and Statistical Support

For the current example

- For our current example: police officers vs. taxi drivers we calculate a z score of -2.738.

RSS
Research and Statistical Support

For the current example

- For our current example: police officers vs. taxi drivers we calculate a z score of -2.738.

$$z = \frac{statistic - mean}{standard\ \ deviation} = \frac{W_s - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}} = \frac{15 - \frac{5(5+6+1)}{2}}{\sqrt{\frac{5(6)(5+6+1)}{12}}} = -2.738$$

RSS Research and Statistical Support

For the current example

- For our current example: police officers vs. taxi drivers we calculate a z score of -2.738.

$$z = \frac{statistic - mean}{standard\ \ deviation} = \frac{W_s - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}} = \frac{15 - \frac{5(5+6+1)}{2}}{\sqrt{\frac{5(6)(5+6+1)}{12}}} = -2.738$$

- So we could say the police officers scored significantly lower than the taxi drivers because a critical z value of -1.64 corresponds to a one-tailed test of z at 0.05 (negative because we hypothesized the police would be **lower**).

`http://www.mathsisfun.com/data/standard-normal-distribution-table.html`

RSS
Research and Statistical Support

Wilcoxon's Matched-Pairs Signed-Ranks test

- The Wilcoxon's Matched-Pairs Signed-Ranks test is an appropriate alternative to the Dependent Samples *t* test.

Wilcoxon's Matched-Pairs Signed-Ranks test

- The Wilcoxon's Matched-Pairs Signed-Ranks test is an appropriate alternative to the Dependent Samples *t* test.
- It is used when the assumptions for the Dependent Samples *t* test can not be met.

RSS
Research and Statistical Support

Wilcoxon's Matched-Pairs Signed-Ranks test

- The Wilcoxon's Matched-Pairs Signed-Ranks test is an appropriate alternative to the Dependent Samples *t* test.
- It is used when the assumptions for the Dependent Samples *t* test can not be met.
- Specifically it is used to determine if a significant difference exists among two related sets of scores.

RSS
Research and Statistical Support

Wilcoxon's Matched-Pairs Signed-Ranks test

- The Wilcoxon's Matched-Pairs Signed-Ranks test is an appropriate alternative to the Dependent Samples *t* test.
- It is used when the assumptions for the Dependent Samples *t* test can not be met.
- Specifically it is used to determine if a significant difference exists among two related sets of scores.
  - e.g., pretest to post test.

RSS
Research and Statistical Support

Wilcoxon's Matched-Pairs Signed-Ranks test

- The Wilcoxon's Matched-Pairs Signed-Ranks test is an appropriate alternative to the Dependent Samples *t* test.
- It is used when the assumptions for the Dependent Samples *t* test can not be met.
- Specifically it is used to determine if a significant difference exists among two related sets of scores.
  - e.g., pretest to post test.
- Like the previous Wilcoxon test, this one works with ranks and the sum of ranks.

RSS
Research and Statistical Support

Quick example

- Suppose were were interested in documenting the effectiveness of Xanax as an anti-anxiety treatment.

RSS
Research and Statistical Support

Quick example

- Suppose were were interested in documenting the effectiveness of Xanax as an anti-anxiety treatment.
- We gather 10 individuals who meet the diagnostic criteria for general anxiety and measure their symptoms with a standard anxiety survey.

Quick example

- Suppose were were interested in documenting the effectiveness of Xanax as an anti-anxiety treatment.
- We gather 10 individuals who meet the diagnostic criteria for general anxiety and measure their symptoms with a standard anxiety survey.
- Then, we administer a protocol of Xanax for two weeks and follow that with another measure of their symptoms on the anxiety survey.

RSS
Research and Statistical Support

Quick example

- Suppose were were interested in documenting the effectiveness of Xanax as an anti-anxiety treatment.
- We gather 10 individuals who meet the diagnostic criteria for general anxiety and measure their symptoms with a standard anxiety survey.
- Then, we administer a protocol of Xanax for two weeks and follow that with another measure of their symptoms on the anxiety survey.
- We would expect the post test scores to be lower than the pretest scores.

Quick example

- Suppose were were interested in documenting the effectiveness of Xanax as an anti-anxiety treatment.
- We gather 10 individuals who meet the diagnostic criteria for general anxiety and measure their symptoms with a standard anxiety survey.
- Then, we administer a protocol of Xanax for two weeks and follow that with another measure of their symptoms on the anxiety survey.
- We would expect the post test scores to be lower than the pretest scores.
    - One-tailed test, lower end

Example Data

Pre	Post	Difference	Rank of difference	Signed Rank
15	8	7	2.5	2.5
18	10	8	4.5	4.5
17	8	9	6.5	6.5
19	11	8	4.5	4.5
20	13	7	2.5	2.5
22	12	10	8.5	8.5
16	18	-2	1	-1
24	12	12	10	10
23	14	9	6.5	6.5
21	11	10	8.5	8.5

$$T+ = \sum positive\ ranks = 54$$
$$T- = \sum negative\ ranks = -1$$

T_{calc} and T_{crit}

- Once we have the sum of both positive and negative difference ranks, we compute T_{calc} which is simply the smaller of the two in absolute value.

RSS
Research and Statistical Support

T_{calc} and T_{crit}

- Once we have the sum of both positive and negative difference ranks, we compute T_{calc} which is simply the smaller of the two in absolute value.
  - Since $T- = -1$ is smaller in absolute value than $T+ = 54$, then $T_{calc} = 1$ (the absolute value of the smaller rank sum).

T_{calc} and T_{crit}

- Once we have the sum of both positive and negative difference ranks, we compute T_{calc} which is simply the smaller of the two in absolute value.
  - Since $T- = -1$ is smaller in absolute value than $T+ = 54$, then $T_{calc} = 1$ (the absolute value of the smaller rank sum).
- To find the critical value (T_{crit}), we use the number of participants or cases ($n = 10$) and look in the T distribution table, specifically the column with a significance level of 0.05 (the table linked below has only one column: the 0.05 values are listed).

RSS
Research and Statistical Support

T_{calc} and T_{crit}

- Once we have the sum of both positive and negative difference ranks, we compute T_{calc} which is simply the smaller of the two in absolute value.
  - Since $T- = -1$ is smaller in absolute value than $T+ = 54$, then $T_{calc} = 1$ (the absolute value of the smaller rank sum).
- To find the critical value (T_{crit}), we use the number of participants or cases ($n = 10$) and look in the T distribution table, specifically the column with a significance level of 0.05 (the table linked below has only one column: the 0.05 values are listed).
  - As before, all the values in the table are for one-tailed tests.

RSS
Research and Statistical Support

T_{calc} and T_{crit}

- Once we have the sum of both positive and negative difference ranks, we compute T_{calc} which is simply the smaller of the two in absolute value.
  - Since $T- = -1$ is smaller in absolute value than $T+ = 54$, then $T_{calc} = 1$ (the absolute value of the smaller rank sum).
- To find the critical value (T_{crit}), we use the number of participants or cases ($n = 10$) and look in the T distribution table, specifically the column with a significance level of 0.05 (the table linked below has only one column: the 0.05 values are listed).
  - As before, all the values in the table are for one-tailed tests.

http://comp9.psych.cornell.edu/Darlington/wilcoxon/wilcox5.htm

RSS
Research and Statistical Support

T_{calc} versus T_{crit}

http://comp9.psych.cornell.edu/Darlington/wilcoxon/wilcox51.htm

- So, the T_{crit} (labeled S in the table linked above), for $n = 10$ would be 10 (with exact significance level at 0.04199) or we could use 11 (with an exact significance level of 0.05273).

Research and Statistical Support

T_{calc} versus T_{crit}

http://comp9.psych.cornell.edu/Darlington/wilcoxon/wilcox51.htm

- So, the T_{crit} (labeled S in the table linked above), for $n = 10$ would be 10 (with exact significance level at 0.04199) or we could use 11 (with an exact significance level of 0.05273).
  - Remember, because we are dealing with ranks, T_{crit} must be a discrete number.

T_{calc} versus T_{crit}

http://comp9.psych.cornell.edu/Darlington/wilcoxon/wilcox51.htm

- So, the T_{crit} (labeled S in the table linked above), for $n = 10$ would be 10 (with exact significance level at 0.04199) or we could use 11 (with an exact significance level of 0.05273).
  - Remember, because we are dealing with ranks, T_{crit} must be a discrete number.
- So, since $T_{calc} = 1 < 10 = T_{crit}$ we reject the null hypothesis and conclude that the post-test scores were significantly lower than the pretest scores.

RSS Research and Statistical Support

From T to z

- When sample sizes are greater than 50, we can conduct a z test with our ranked sums T.

From T to z

- When sample sizes are greater than 50, we can conduct a z test with our ranked sums T.
- The distribution of T is approximately normal when $n > 50$ with a mean of: $\frac{n(n+1)}{4}$

RSS
Research and Statistical Support

From T to z

- When sample sizes are greater than 50, we can conduct a z test with our ranked sums T.
- The distribution of T is approximately normal when $n > 50$ with a mean of: $\frac{n(n+1)}{4}$
- And a standard deviation of: $\sqrt{\frac{n(n+1)(2n+1)}{24}}$

Research and Statistical Support

From T to z

- When sample sizes are greater than 50, we can conduct a z test with our ranked sums T.
- The distribution of T is approximately normal when $n > 50$ with a mean of: $\frac{n(n+1)}{4}$
- And a standard deviation of: $\sqrt{\frac{n(n+1)(2n+1)}{24}}$
- All of which gives us what we need to compute z:

RSS
Research and Statistical Support

From T to z

- When sample sizes are greater than 50, we can conduct a z test with our ranked sums T.
- The distribution of T is approximately normal when $n > 50$ with a mean of: $\frac{n(n+1)}{4}$
- And a standard deviation of: $\sqrt{\frac{n(n+1)(2n+1)}{24}}$
- All of which gives us what we need to compute z:

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

RSS
Research and Statistical Support

Current Example applied to z test

- Our current example has $T = 1$ and $n = 10$ so;

Current Example applied to z test

- Our current example has $T = 1$ and $n = 10$ so;

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} = \frac{1 - \frac{10(10+1)}{4}}{\sqrt{\frac{10(10+1)(2*10+1)}{24}}} = -2.701$$

RSS
Research and Statistical Support

Current Example applied to z test

- Our current example has $T = 1$ and $n = 10$ so;

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} = \frac{1 - \frac{10(10+1)}{4}}{\sqrt{\frac{10(10+1)(2*10+1)}{24}}} = -2.701$$

- Clearly, our z calculated value is more extreme than a critical value of -1.64 (one-tailed, 0.05 significance).

Current Example applied to z test

- Our current example has $T = 1$ and $n = 10$ so;

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} = \frac{1 - \frac{10(10+1)}{4}}{\sqrt{\frac{10(10+1)(2*10+1)}{24}}} = -2.701$$

- Clearly, our z calculated value is more extreme than a critical value of -1.64 (one-tailed, 0.05 significance).

  http://www.mathsisfun.com/data/standard-normal-distribution-table.html

RSS Research and Statistical Support

Kruskal-Wallis One-way ANOVA

- The Kruskal-Wallis test is a nonparametric replacement for the One-way ANOVA when the assumptions of One-way ANOVA are not met.

RSS
Research and Statistical Support

Kruskal-Wallis One-way ANOVA

- The Kruskal-Wallis test is a nonparametric replacement for the One-way ANOVA when the assumptions of One-way ANOVA are not met.
- The Kruskal-Wallis test is a direct extension of the Wilcoxon's Rank-Sum test for independent groups.

Kruskal-Wallis One-way ANOVA

- The Kruskal-Wallis test is a nonparametric replacement for the One-way ANOVA when the assumptions of One-way ANOVA are not met.
- The Kruskal-Wallis test is a direct extension of the Wilcoxon's Rank-Sum test for independent groups.
  - Both are based on the sums of ranks.

RSS
Research and Statistical Support

Kruskal-Wallis One-way ANOVA

- The Kruskal-Wallis test is a nonparametric replacement for the One-way ANOVA when the assumptions of One-way ANOVA are not met.
- The Kruskal-Wallis test is a direct extension of the Wilcoxon's Rank-Sum test for independent groups.
  - Both are based on the sums of ranks.
- As with the Wilcoxon's Rank-Sum test we again rank all of the scores (regardless of group membership) and then sum the ranks for each group.

RSS
Research and Statistical Support

Omnibus test of differences

- The Kruskal-Wallis test is used to identify differences in central tendency among more than 2 groups.

Omnibus test of differences

- The Kruskal-Wallis test is used to identify differences in central tendency among more than 2 groups.
- As with the One-way ANOVA, the Kruskal-Wallis test can only tell us if there is a significant difference among the central tendencies of the groups; it does not tell us *where* the group differences are located.

Omnibus test of differences

- The Kruskal-Wallis test is used to identify differences in central tendency among more than 2 groups.
- As with the One-way ANOVA, the Kruskal-Wallis test can only tell us if there is a significant difference among the central tendencies of the groups; it does not tell us *where* the group differences are located.
  - Secondary analysis, such as the Wilcoxon's Rank-Sum test would be necessary (much like conducting post-hoc testing in the ANOVA situation).

Compute H

- To calculate the Kruskal-Wallis test; compute H

Compute H

- To calculate the Kruskal-Wallis test; compute H

$$H = \left[\frac{12}{n_t(n_t+1)}\right] * \sum \frac{R_i^2}{n_i} - 3\left(n_t + 1\right)$$

Compute H

- To calculate the Kruskal-Wallis test; compute H

$$H = \left[\frac{12}{n_t(n_t+1)}\right] * \sum \frac{R_i^2}{n_i} - 3\,(n_t + 1)$$

- where n_t is the total number of participants, R_i is the sum of the ranks in group i, and n_i is the number of participants in group i.

Compute H

- To calculate the Kruskal-Wallis test; compute H

$$H = \left[\frac{12}{n_t(n_t+1)}\right] * \sum \frac{R_i^2}{n_i} - 3\left(n_t + 1\right)$$

- where n_t is the total number of participants, R_i is the sum of the ranks in group i, and n_i is the number of participants in group i.
- The comparison distribution is the chi-square distribution with $df = k - 1$ where k is the number of groups.

RSS
Research and Statistical Support

Quick Example

- Suppose we added limousine drivers to our earlier example comparing driving anger among police officers and taxi drivers.

Quick Example

- Suppose we added limousine drivers to our earlier example comparing driving anger among police officers and taxi drivers.

Police		Taxi		Limousine	
Score	Rank	Score	Rank	Score	Rank
8	1	27	13	16	9
15	7.5	28	14.5	15	7.5
12	3	19	11	14	6
10	2	17	10	13	4.5
13	4.5	26	12		
		28	14.5		

Tied scores get tied ranks half-way between the two whole number ranks they would occupy if sequential.

Calculate H

- First, we need the sum of each rank (R_i) and the number of participants in each group (n_i).

Calculate H

- First, we need the sum of each rank (R_i) and the number of participants in each group (n_i).
  $R_1 = 1 + 7.5 + 3 + 2 + 4.5 = 18$ and $n_1 = 5$
  $R_2 = 13 + 14.5 + 11 + 10 + 12 + 14.5 = 75$ and $n_2 = 6$
  $R_3 = 9 + 7.5 + 6 + 4.5 = 27$ and $n_3 = 4$

RSS
Research and Statistical Support

Calculate H

- First, we need the sum of each rank (R_i) and the number of participants in each group (n_i).
  $R_1 = 1 + 7.5 + 3 + 2 + 4.5 = 18$ and $n_1 = 5$
  $R_2 = 13 + 14.5 + 11 + 10 + 12 + 14.5 = 75$ and $n_2 = 6$
  $R_3 = 9 + 7.5 + 6 + 4.5 = 27$ and $n_3 = 4$
- Then we can calculate H

RSS
Research and Statistical Support

Calculate H

- First, we need the sum of each rank (R_i) and the number of participants in each group (n_i).
  $R_1 = 1 + 7.5 + 3 + 2 + 4.5 = 18$ and $n_1 = 5$
  $R_2 = 13 + 14.5 + 11 + 10 + 12 + 14.5 = 75$ and $n_2 = 6$
  $R_3 = 9 + 7.5 + 6 + 4.5 = 27$ and $n_3 = 4$
- Then we can calculate H

$$H = \left[\frac{12}{n_t(n_t+1)}\right] * \sum \frac{R_i^2}{n_i} - 3\,(n_t + 1) =$$

$$\left[\frac{12}{15(15+1)}\right] * \left[\frac{18^2}{5} + \frac{75^2}{6} + \frac{27^2}{4}\right] - 3\,(15 + 1) = 11.2275$$

Compare and make a decision

- Our H_{calc} (which is really a χ^2 value) is 11.2275.

Compare and make a decision

- Our H_{calc} (which is really a χ^2 value) is 11.2275.
- We have $df = k - 1 = 3 - 1 = 2$ which with 0.05 significance level, yields a critical value of 5.991.

  http://www.medcalc.be/manual/chi-square-table.php

Compare and make a decision

- Our H_{calc} (which is really a χ^2 value) is 11.2275.
- We have $df = k - 1 = 3 - 1 = 2$ which with 0.05 significance level, yields a critical value of 5.991.

  http://www.medcalc.be/manual/chi-square-table.php
- So, since $H_{calc} = 11.2275 > 5.991 = \chi^2_{crit}$ we reject the null hypothesis and conclude there was a significant difference in driving anger among the three groups.

RSS
Research and Statistical Support

Compare and make a decision

- Our H_{calc} (which is really a χ^2 value) is 11.2275.
- We have $df = k - 1 = 3 - 1 = 2$ which with 0.05 significance level, yields a critical value of 5.991.

  http://www.medcalc.be/manual/chi-square-table.php
- So, since $H_{calc} = 11.2275 > 5.991 = \chi^2_{crit}$ we reject the null hypothesis and conclude there was a significant difference in driving anger among the three groups.
  - Secondary analysis, such as the Wilcoxon's Rank-Sum test would be necessary to determine where the differences were among each group (much like conducting post-hoc testing in the ANOVA situation).

RSS
Research and Statistical Support

Summary of Module 11

Module 11 covered the following topics:

Summary of Module 11

Module 11 covered the following topics:

- Chi-square tests

Summary of Module 11

Module 11 covered the following topics:

- Chi-square tests
- Wilcoxon's Rank-Sum test

Summary of Module 11

Module 11 covered the following topics:

- Chi-square tests
- Wilcoxon's Rank-Sum test
- Wilcoson's Matched-Pairs Signed-Ranks test

Summary of Module 11

Module 11 covered the following topics:

- Chi-square tests
- Wilcoxon's Rank-Sum test
- Wilcoson's Matched-Pairs Signed-Ranks test
- Kruskal-Wallis One-Way ANOVA

RSS
Research and Statistical Support

This concludes Module 11

- Until next time; have a nice day.

These slides initially created on: October 28, 2010
These slides last updated on: November 2, 2010

- The bottom date shown is the date this Adobe.pdf file was created; LaTeX[1] has a command for automatically inserting the date of a document's creation.

RSS
Research and Statistical Support

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