Please note that the workshop is aimed to be a brief introduction to the topic and this PowerPoint is primarily designed to support the flow of the workshop. It cannot be seen as either an exclusive or exhaustive resource on the statistical concepts which are introduced in this course. You are encouraged to refer to peer-reviewed books or papers that are listed throughout the presentation.

It is acknowledged that a number of slides have been adapted from presentations produced by the previous statistical consultant (Kylie Lange) and a colleague with whom I worked with in the past (Dr Kelvin Gregory).

SPSS / PASW / IBM SPSS

• In late 2009 SPSS Inc. was taken over by IBM Company and the software changed its official name twice over the period of one year. From SPSS it was relabelled to PASW (Predictive Analytics Software) and later to IBM SPSS. Consequently, there may be books, online resources, etc. that use either of those different names but in fact refer to the same software.

• SPSS – Statistical Package for the Social Sciences

• PASW – Predictive Analytics Software

• IBM SPSS Statistics

Levels of Measurement and Measurement Scales

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio Data</td>
<td>Differences between measurements, true zero exists</td>
<td>Height, Age, Weekly Food Spending</td>
</tr>
<tr>
<td>Interval Data</td>
<td>Differences between measurements but no true zero</td>
<td>Temperature in Celsius, Standardized exam score</td>
</tr>
<tr>
<td>Ordinal Data</td>
<td>Ordered Categories (rankings, order, or scaling)</td>
<td>Service quality rating, Student letter grades</td>
</tr>
<tr>
<td>Nominal Data</td>
<td>Categories (no ordering or direction)</td>
<td>Marital status, Type of car owned, Gender/Gender</td>
</tr>
</tbody>
</table>
Selection of statistical methods

Example 1

Example 2

Example 3
Similar ones in other resources …

Exercise 1
Comparisons of Column Proportions (z-test)

- Please open - PISA_2000_Part1b.sav

Simplified data from PISA 2000 Study – Few countries selected
(The Programme for International Students Assessment)

[http://www.pisa.oecd.org](http://www.pisa.oecd.org)

Comparisons of Column Proportions (z-test)


Contingency Tables

- Useful in situations involving multiple population proportions
- Used to classify sample observations according to two or more characteristics
- Also called a cross-classification table.

Contingency Table Example

Left-Handed vs. Gender
Dominant Hand: Left vs. Right
Gender: Male vs. Female

- 2 categories for each variable, so called a 2 x 2 table
- Suppose we examine a sample of size 300
Sample results organized in a contingency table:

<table>
<thead>
<tr>
<th>Gender</th>
<th>Hand Preference</th>
<th>Left</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>12</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>24</td>
<td>156</td>
</tr>
</tbody>
</table>

36 264 300

120 Females, 12 were left handed
180 Males, 24 were left handed

Sample size = n = 300:

The Chi-Square Test Statistic

The Chi-square test statistic is:

\[ \chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} \]

where:

- \( f_o \) = observed frequency in a particular cell
- \( f_e \) = expected frequency in a particular cell if \( H_0 \) is true

\( \chi^2 \) for the 2 x 2 case has 1 degree of freedom

(Assumed: each cell in the contingency table has expected frequency of at least 5)

Computing the Average Proportion

The average proportion is:

\[ p = \frac{X_1 + X_2}{n_1 + n_2} = \frac{X}{n} \]

Here:

\[ \frac{12 + 24}{120 + 180} = \frac{36}{300} = 0.12 \]

i.e., the proportion of left handers overall is 0.12, that is, 12%

Finding Expected Frequencies

- To obtain the expected frequency for left handed females, multiply the average proportion left handed (\( \bar{p} \)) by the total number of females
- To obtain the expected frequency for left handed males, multiply the average proportion left handed (\( \bar{p} \)) by the total number of males

If the two proportions are equal, then

\[ P(\text{Left Handed | Female}) = P(\text{Left Handed | Male}) = 0.12 \]

i.e., we would expect \( 0.12(120) = 14.4 \) females to be left handed

\( 0.12(180) = 21.6 \) males to be left handed

\( \chi^2 \) Test for the Difference Between Two Proportions

\( H_0: \pi_1 = \pi_2 \) (Proportion of females who are left handed is equal to the proportion of males who are left handed)

\( H_1: \pi_1 \neq \pi_2 \) (The two proportions are not the same – Hand preference is not independent of gender)

- If \( H_0 \) is true, then the proportion of left-handed females should be the same as the proportion of left-handed males
- The two proportions above should be the same as the proportion of left-handed people overall
Observed vs. Expected Frequencies

<table>
<thead>
<tr>
<th>Gender</th>
<th>Hand Preference</th>
<th>Left</th>
<th>Right</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>Observed = 12</td>
<td>12</td>
<td>108</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>Expected = 14.4</td>
<td></td>
<td>105.6</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>Observed = 24</td>
<td>36</td>
<td>264</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>Expected = 21.6</td>
<td>180</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The Chi-Square Test Statistic

\[
\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}
\]

Here, \( \chi^2 = 0.7576 < \chi^2_{0.05} = 3.841 \), so we do not reject \( H_0 \) and conclude that there is not sufficient evidence that the two proportions are different at \( \alpha = 0.05 \).

Exercise 2

- Please open - PISA_2000_Part1b_AUSTRALIA.sav

Simplified data from PISA 2000 Study – Few countries selected
(The Programme for International Students Assessment)

http://www.pisa.oecd.org

Exercise 3

- Please start a new data file.

<table>
<thead>
<tr>
<th>Sex · Q3</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet · Q21d</td>
<td>Count</td>
<td>Count</td>
</tr>
<tr>
<td>Yes</td>
<td>92</td>
<td>97</td>
</tr>
<tr>
<td>No</td>
<td>51</td>
<td>44</td>
</tr>
</tbody>
</table>

\( \chi^2 \) Test of Independence

- Similar to the \( \chi^2 \) test for equality of more than two proportions, but extends the concept to contingency tables with \( r \) rows and \( c \) columns

\( H_0 \): The two categorical variables are independent (i.e., there is no relationship between them)

\( H_1 \): The two categorical variables are dependent (i.e., there is a relationship between them)
**\( \chi^2 \) Test of Independence**

The Chi-square test statistic is:

\[
\chi^2 = \sum \left( \frac{f_o - f_e}{f_e} \right)^2
\]

- where:
  - \( f_o \): observed frequency in a particular cell of the \( r \times c \) table
  - \( f_e \): expected frequency in a particular cell if \( H_0 \) is true
- \( \chi^2 \) for the \( r \times c \) case has \((r-1)(c-1)\) degrees of freedom

(Assumed: each cell in the contingency table has expected frequency of at least 1)

**Expected Cell Frequencies**

- Expected cell frequencies:
  
  \[
  f_e = \frac{\text{row total} \times \text{column total}}{n}
  \]

  Where:
  - row total = sum of all frequencies in the row
  - column total = sum of all frequencies in the column
  - \( n \) = overall sample size

**Decision Rule**

- The decision rule is

\[
\text{If } \chi^2 > \chi^2_U, \text{ reject } H_0, \text{ otherwise, do not reject } H_0
\]

Where \( \chi^2_U \) is from the chi-squared distribution with \((r-1)(c-1)\) degrees of freedom

**Example**

- The hypothesis to be tested is:

  \( H_0: \) Meal plan and class standing are independent (i.e., there is no relationship between them)

  \( H_1: \) Meal plan and class standing are dependent (i.e., there is a relationship between them)

**Example:**

<table>
<thead>
<tr>
<th>Class</th>
<th>Number of meals per week</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20/week</td>
<td>10/week</td>
</tr>
<tr>
<td>Fresh</td>
<td>24</td>
<td>32</td>
</tr>
<tr>
<td>Soph</td>
<td>22</td>
<td>26</td>
</tr>
<tr>
<td>Junior</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>Senior</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td>70</td>
<td>88</td>
</tr>
</tbody>
</table>

**Example:**

<table>
<thead>
<tr>
<th>Class</th>
<th>Number of meals per week</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20/week</td>
<td>10/week</td>
</tr>
<tr>
<td>Fresh</td>
<td>24.5</td>
<td>30.8</td>
</tr>
<tr>
<td>Soph</td>
<td>21.0</td>
<td>26.4</td>
</tr>
<tr>
<td>Junior</td>
<td>10.5</td>
<td>13.2</td>
</tr>
<tr>
<td>Senior</td>
<td>14.0</td>
<td>17.6</td>
</tr>
<tr>
<td>Total</td>
<td>70</td>
<td>88</td>
</tr>
</tbody>
</table>

**Example:**

Expected cell frequencies if \( H_0 \) is true:

**Example:**

Expected cell frequencies if \( H_0 \) is true:

**Example for one cell:**

\[
f_e = \frac{\text{row total} \times \text{column total}}{n}
\]

\[
= \frac{30 \times 70}{200} = 10.5
\]
Example: The Test Statistic

- The test statistic value is:

\[ \chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} \]
\[ = \frac{(24 - 24.5)^2}{24.5} + \frac{(32 - 30.8)^2}{30.8} + \cdots + \frac{(10 - 8.4)^2}{8.4} = 0.709 \]

\( \chi^2_o = 12.592 \) for \( \alpha = 0.05 \) from the chi-squared distribution with \((4 - 1)(3 - 1) = 6\) degrees of freedom

Example: Decision and Interpretation

The test statistic is \( \chi^2 = 0.709 \), \( \chi^2_o = 12.592 \) with 6 d.f. = 12.592

Decision Rule:
- If \( \chi^2 > 12.592 \), reject \( H_0 \), otherwise, do not reject \( H_0 \)

Here, \( \chi^2 = 0.709 < \chi^2_o = 12.592 \), so do not reject \( H_0 \)

Conclusion: there is not sufficient evidence that meal plan and class standing are related at \( \alpha = 0.05 \)

Fisher Exact Test of Significance

- "Fisher’s exact test directly computes \( p \), the probability of getting a table as strong as the observed table or stronger. This requires computing Fisher's for the given table and all stronger tables, then summing the separate \( p \)'s to get the total probability of a table that strong or stronger." Garson 2011

For more details see
http://faculty.chass.ncsu.edu/garson/PA765/fisher.htm

Exercise 4

- Please open -
PISA_2000_Part1b_AUSTRALIA_SMALL.sav

Simplified data from PISA 2000 Study – Few countries selected
(The Programme for International Students Assessment)

http://www.pisa.oecd.org

McNemar Test (Related Samples)

- Used to determine if there is a difference between proportions of two related samples
- Uses a test statistic the follows the normal distribution

<table>
<thead>
<tr>
<th>Condition 1</th>
<th>Yes</th>
<th>No</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>A</td>
<td>B</td>
<td>A+B</td>
</tr>
<tr>
<td>No</td>
<td>C</td>
<td>D</td>
<td>C+D</td>
</tr>
<tr>
<td>Totals</td>
<td>A+C</td>
<td>B+D</td>
<td>n</td>
</tr>
</tbody>
</table>

McNemar Test (Related Samples)
McNemar Test (Related Samples)

• The sample proportions of interest are
  \[ p_1 = \frac{A+B}{n} \]  
  proportion of respondents who answer yes to condition 1
  \[ p_2 = \frac{A+C}{n} \]  
  proportion of respondents who answer yes to condition 2

• Test
  \[ H_0: \pi_1 = \pi_2 \]  
  (the two population proportions are equal)
  \[ H_1: \pi_1 \neq \pi_2 \]  
  (the two population proportions are not equal)

\[ Z = \frac{B-C}{\sqrt{B+C}} \]

where the test statistic \( Z \) is approximately normally distributed

Cochran's Q Test

• Cochran's Q tests whether the percentages (proportions) of a given variable are the same across multiple dependent samples. It extends the McNemar test beyond two related samples.

Exercise 5

• Please open
  – Related_Samples.sav

Data from

Chi-Square Goodness-of-Fit Test

• Does sample data conform to a hypothesized distribution?
  • Examples:
    • Are technical support calls equal across all days of the week? (i.e., do calls follow a uniform distribution?)
    • Do measurements from a production process follow a normal distribution?
  • Sample data for 10 days per day of week:
    
    | Day       | Sum of calls for this day: |
    |-----------|-----------------------------|
    | Monday    | 290                         |
    | Tuesday   | 250                         |
    | Wednesday | 238                         |
    | Thursday  | 257                         |
    | Friday    | 265                         |
    | Saturday  | 230                         |
    | Sunday    | 192                         |
    |           | Σ = 1722                    |
Logic of Goodness-of-Fit Test

- If calls are uniformly distributed, the 1722 calls would be expected to be equally divided across the 7 days:

\[
\frac{1722}{7} = 246 \text{ expected calls per day if uniform}
\]

- Chi-Square Goodness-of-Fit Test: test to see if the sample results are consistent with the expected results

Observed vs. Expected Frequencies

<table>
<thead>
<tr>
<th></th>
<th>Observed ( f_o )</th>
<th>Expected ( f_e )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>290</td>
<td>246</td>
</tr>
<tr>
<td>Tuesday</td>
<td>250</td>
<td>246</td>
</tr>
<tr>
<td>Wednesday</td>
<td>238</td>
<td>246</td>
</tr>
<tr>
<td>Thursday</td>
<td>257</td>
<td>246</td>
</tr>
<tr>
<td>Friday</td>
<td>265</td>
<td>246</td>
</tr>
<tr>
<td>Saturday</td>
<td>230</td>
<td>246</td>
</tr>
<tr>
<td>Sunday</td>
<td>192</td>
<td>246</td>
</tr>
<tr>
<td>TOTAL</td>
<td>1722</td>
<td>1722</td>
</tr>
</tbody>
</table>

Chi-Square Test Statistic

\[
\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}
\]

(\CHEDULE{where df = k – p – 1})

- The test statistic is
- \( k \) = number of categories
- \( f_o \) = observed frequency
- \( f_e \) = expected frequency
- \( p \) = number of parameters estimated from the data

The Rejection Region

- \( H_0: \) The distribution of calls is uniform over days of the week
- \( H_1: \) The distribution of calls is not uniform

\[
\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}
\]

- Reject \( H_0 \) if \( \chi^2 > \chi^2_{0.05} \)

\( k = 2 \) degrees of freedom, since \( p = 1 \) here (the mean was estimated)

- \( k = 2 \) degrees of freedom, since \( p = 1 \) here (the mean was estimated)

Exercise 6

- Please open
  - PISA_2000_Part1b_AUSTRALIA_SMALL.sav
SPSS – BOOKS (Hard copies)


SPSS – BOOKS (Online copies)

Hard copies and online versions
• Online versions
  • Chapter 8 in Marston, Louise. (2010). Introductory statistics for health and nursing using SPSS. Los Angeles: SAGE.
  • Chapters 8 & 14 in Larson-Hall, Jenifer. (2010). A guide to doing statistics in second language research using SPSS

SPSS – Help and Resources

• SPSS has a range of help options available
  - Topics
    • Used to find specific information
  - Tutorial
    • Find illustrated, step-by-step instructions for the basic features
  - Case studies
    • Hands-on examples of various types of statistical procedures
  - Statistics coach
    • To help you find the procedure you want to use
And manuals available online -

SPSS – Online tutorials and resources

(!!! Please keep in mind that usually online resources are not academically peer reviewed. Despite many of them being of high quality as well as being very useful from educational point of view, they shouldn’t be treated as a completely reliable and academically sound references)

- Statnotes: Topics in Multivariate Analysis, by G. David Garson
http://www.statisticalassociates.com/
- UCLA Institute for Digital Research and Education - SPSS Starter Kit
http://www.ats.ucla.edu/stat/spss/sk/default.htm
- Getting Started with SPSS for Windows by John Samuel, Indiana University
http://www.indiana.edu/~statmath/stat/spss/win/index.html
- Companion Website for the 3rd edition of Discovering Statistics Using SPSS by Andy Field
http://www.uk.sagepub.com/field3e/SPSSFlashmovieselect.htm
- SPSS for Windows and Amos tutorials by Information Technology Services, University of Texas
http://ssc.utexas.edu/software/software-tutorials#SPSS
- Journey in Survey Research by John Hall
http://surveyresearch.weebly.com/index.html

Archives of SPSSX-L@LISTSERV.UGA.EDU – List Serve that is endorsed by IBM SPSS
http://www.listserv.uga.edu/archives/spssx-l.html

Other forums
http://groups.google.com/group/comp.soft-sys.stat.spss/topics?gvc=2
http://www.spssforum.com/

THANK YOU

Please provide us with your feedback by completing the short survey.