Lesson 12 Overview

- Lesson 11 covered two inference methods for categorical data from 2 groups
  - Confidence Intervals for the difference of two proportions
  - Two sample z-test of equality of two proportions
- Lesson 12 covers three inference methods for categorical data
  - Chi-square test for comparisons between 2 categorical variables
  - Fisher’s exact test
  - McNemar Chi-square test for paired categorical data

Chi-square test

The Chi-square test can be used for two applications

1. The Chi-square test can also be used to test for independence between two variables
   - The null hypothesis for this test is that the variables are independent (i.e. that there is no statistical association).
   - The alternative hypothesis is that there is a statistical relationship or association between the two variables.

2. The Chi-square test can be used to test for equality of proportions between two or more groups.
   - The null hypothesis for this test is that the 2 proportions are equal.
   - The alternative hypothesis is that the proportions are not equal (test for a difference in either direction)

Contingency Tables

- Setting: Let $X_1$ and $X_2$ denote categorical variables, $X_1$ having $I$ levels and $X_2$ having $J$ levels. There are $IJ$ possible combinations of classifications.

<table>
<thead>
<tr>
<th>Level=1</th>
<th>Level=2</th>
<th>Level=J</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level=1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level=2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level=J</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- When the cells contain frequencies of outcomes, the table is called a contingency table.

Chi-square Test: Testing for Independence

Step 1: Hypothesis (always two-sided):
   - $H_0$: Independent
   - $H_A$: Not independent

Step 2: Calculate the test statistic:
   $$ \chi^2 = \sum \frac{(o_i - e_i)^2}{e_i} - J \text{ with df} = (I-1)(J-1) $$

Step 3: Calculate the p-value
   - $p$-value = $P(\chi^2 > \chi^2)$ <- value 2-sided

Step 4: Draw a conclusion
   - $p$-value < $\alpha$: reject independence
   - $p$-value > $\alpha$: do not reject independence

Chi-square Test: Testing for Independence

Racial differences and cardiac arrest.

An Example

In a large mid-western city, the association in the incidence of cardiac arrest and subsequent survival was studied in 6117 cases of non-traumatic, out of hospital cardiac arrest.

During a 12 month period, fewer than 1% of African-Americans survived an arrest-to-hospital discharge, compared to 2.6% of Caucasians.
Chi-square Test: Testing for Independence
Racial differences and cardiac arrest
Survival to Discharge

<table>
<thead>
<tr>
<th>Race</th>
<th>YES</th>
<th>NO</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caucasian</td>
<td>84</td>
<td>3123</td>
<td>3207</td>
</tr>
<tr>
<td>African-American</td>
<td>24</td>
<td>2886</td>
<td>3110</td>
</tr>
<tr>
<td>Total</td>
<td>108</td>
<td>6009</td>
<td>6117</td>
</tr>
</tbody>
</table>

Chi-square Test: Testing for Independence
Scientific Hypothesis:
An association exists between race (African-American/Caucasian) and survival to hospital discharge (Yes/No) in cases of non-traumatic out-of-hospital cardiac arrest.

Statistical Hypothesis:
$H_0$: Race and survival to hospital discharge are independent in cases of non-traumatic out-of-hospital cardiac arrest.

$H_A$: Race and survival to hospital discharge are not independent in cases of non-traumatic out-of-hospital cardiac arrest.

1. Obtain a random sample of $n$ independent observations (the selection of one observation does not influence the selection of any other).
2. Observations are classified subsequently according to cells formed by the intersection rows and columns in a contingency table.
   - Rows ($r$) consist of mutually exclusive categories of one variable.
   - Columns ($c$) consist of mutually exclusive categories of the other variable.
3. The frequency of observations in each cell is determined along with marginal totals.
4. Expected frequencies are calculated under the null hypothesis of independence (no association) and compared to observed frequencies.
   Recall: $A$ and $B$ are independent if:
   $$P(A \text{ and } B) = P(A) \times P(B)$$
5. Use the Chi-square ($\chi^2$) test statistic to observe the difference between the observed and expected frequencies.

$\chi^2$ Distribution
- The probabilities associated with the chi-square distribution are in appendix D.
- The table is set up in the same way as the $t$-distribution.
- The chi-square distribution with 1 df is the same as the square of the $Z$ distribution.
- Since the distribution only takes on positive values all the probability is in the right-tail.

Chi-square distributions and critical values for 1 df, 4 df and 20 df

For $\chi^2$-distribution with 20 df, the critical value at $\alpha = 0.05$ is 31.4

Critical value for $\alpha = 0.05$ and $\chi^2$-distribution with 1 df is 3.84

Critical value for $\alpha = 0.05$ and $\chi^2$-distribution with 4 df is 9.49

Since the Chi-square distribution is always positive, the rejection region is only in the right tail.
How to Identify the critical value

- The rejection region of the Chi-square test is the upper tail so there is only one critical value.
- First calculate the df to identify the correct Chi-square distribution.
  - For a $2 \times 2$ table, there are $(2-1)(2-1) = 1$ df.
- Use the CHIINV function to find the critical value.
- General formula: $=\text{CHIINV}(\alpha, df)$.

State the conclusion

- The p-value for $P(\chi^2 > X^2) = \text{CHIDIST}(X^2, df)$.
- Reject the null hypothesis by either the rejection region method or the p-value method.
  - $X^2 > \text{Critical Value}$
  - or
  - $p\text{-value} < \alpha$.

Chi-square Test: Testing for Independence

Calculating expected frequencies

<table>
<thead>
<tr>
<th>Survival to Discharge</th>
<th>YES</th>
<th>NO</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caucasian</td>
<td>84</td>
<td>3123</td>
<td>3207</td>
</tr>
<tr>
<td>56.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African-American</td>
<td>24</td>
<td>2886</td>
<td>2910</td>
</tr>
<tr>
<td>56.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>108</td>
<td>6009</td>
<td>6117</td>
</tr>
</tbody>
</table>

Under the assumption of independence:

$P(\text{YES and Caucasian}) = P(\text{YES}) \times P(\text{Caucasian})$

$= \frac{108}{6117} \times \frac{3207}{6117} = 0.009256$

Expected cell count $e_{ij} = 0.009256 \times 6117 = 56.62$

Chi-square Test: Testing for Independence

Calculating expected frequencies

<table>
<thead>
<tr>
<th>Survival to Discharge</th>
<th>YES</th>
<th>NO</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caucasian</td>
<td>84</td>
<td>3123</td>
<td>3207</td>
</tr>
<tr>
<td>56.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African-American</td>
<td>24</td>
<td>2886</td>
<td>2910</td>
</tr>
<tr>
<td>51.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>108</td>
<td>6009</td>
<td>6117</td>
</tr>
</tbody>
</table>

Expected Cell Counts = (Marginal Row total * Marginal Column Total) / $n$

Rule of Thumb: Check to see if expected frequencies are > 2

No more than 20% of cells with expected frequencies < 5

Chi-square Test: Testing for Independence

Step 1: Hypothesis (always two-sided):

$H_0$: Independent (Race/Survival)

$H_A$: Not independent

Step 2: Calculate the test statistic:

$p-value = P(\chi^2 > 28.42) = \text{CHIDIST}(28.42, 1) < 0.001$

Step 4: Draw a conclusion

$p\text{-value} < \alpha \implies$ reject independence.

A significant association exists between race and survival to hospital discharge in cases of non-traumatic out-of-hospital cardiac arrest.
Chi-square Test: Testing for Equality or Homogeneity of Proportions

Testing for equality or homogeneity of proportions – examines differences between two or more independent proportions.

In chi-square test for independence, we examine the cross-classification of a single sample of observations on two qualitative variables.

The chi-square test can also be used for problems involving two or more independent populations.

Example

Patients with evolving myocardial infarction were assigned independent and randomly to one of four thrombolytic treatments, and then followed to determine 30 day mortality.

Are these four treatment populations equal with respect to 30-day mortality?

<table>
<thead>
<tr>
<th>30 day outcome</th>
<th>Streptokinase and SC Heparin</th>
<th>Streptokinase and IV Heparin</th>
<th>Accelerated t-PA and IV Heparin</th>
<th>Accelerated t-PA Streptokinase with IV Heparin</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survived</td>
<td>9091</td>
<td>9609</td>
<td>9605</td>
<td>37997</td>
<td></td>
</tr>
<tr>
<td>Died</td>
<td>705</td>
<td>768</td>
<td>652</td>
<td>723</td>
<td>2848</td>
</tr>
<tr>
<td>Total</td>
<td>9796</td>
<td>10377</td>
<td>10344</td>
<td>10328</td>
<td>40845</td>
</tr>
</tbody>
</table>

Under the assumption of independence:

\[
P(\text{Streptokinase and SC Heparin} \text{ and } \text{Survival}) = \frac{9796 \times 40845}{37997} = 0.223
\]

Expected cell count = 0.223 * 40845 = 9112.95

Chi-square Test: Testing for Equality or Homogeneity of Proportions

Step 1: Hypothesis (always two-sided):

\[H_0: \text{The four treatment options are homogeneous with respect to 30 day survival.}\]

\[H_A: \text{The four treatment options are not homogeneous with respect to 30 day survival.}\]

Step 2: Calculate the test statistic:

\[X^2 = \sum \frac{(e_{ij} - o_{ij})^2}{e_{ij}} - \chi^2 \text{ with } df = (I-1)(J-1)\]

Step 3: Calculate the p-value = P(X^2 > X^2)

Step 4: Draw a conclusion

p-value < α reject independence
p-value > α do not reject independence

Chi-square Test: Testing for Equality or Homogeneity of Proportions

Step 1: Hypothesis (always two-sided):

\[H_0: \text{The four treatment options are homogeneous with respect to 30 day survival.}\]

\[H_A: \text{The four treatment options are not homogeneous with respect to 30 day survival.}\]

Step 2: Calculate the test statistic:

\[X^2 = \sum \frac{(e_{ij} - o_{ij})^2}{e_{ij}} = 10.85 \text{ with } df = (2-1)(4-1) = 3\]
Chi-square Test:
Testing for Equality or Homogeneity of Proportions
Step 3: Calculate the p-value
\[ \text{p-value} = P(\chi^2 > 10.85) = \text{chidist}(10.85, 3) = 0.013 \]
Step 4: Draw a conclusion
- The four treatment groups are not equal with respect to 30 day mortality.
- The largest relative departure from expected was noted in patients receiving accelerated t-PA and IV heparin, with fewer patients than expected dying.

Chi-Square Testing

**Independence**
- Ho: two classification criteria are independent
- Ha: two classification criteria are not independent.

**Requirements:**
- One sample selected randomly from a defined population.
- Observations cross-classified into two nominal criteria.
- Conclusions phrased in terms of independence of the two classifications.

**Equality**
- Ho: populations are homogeneous with regard to one classification criterion.
- Ha: populations are not homogeneous with regard to one classification criterion.

**Requirements:**
- Two or more samples are selected from two or more populations.
- Observations are classified on one nominal criterion.
- Conclusions phrased with regard to homogeneity or equality of treatment populations.

Chi-square test in EXCEL or online calculator
- There is no Excel function or Data Analysis Tool for the Chi-square test.
- For the Excel examples, you’ll need to calculate the expected cell frequencies from the observed marginal totals and then calculate the statistic from the observed and expected frequencies.
- This website will calculate the Chi-square statistic and p-value for data in a 2 X 2 table. Enter the cell counts in the table. Choose the Chi-square test without Yate’s correction to obtain the same results as in the example www.graphpad.com/quickcalcs/contingency1.cfm

Chi-Square Testing: Rules of Thumb
- All expected frequencies should be equal to or greater than 2. (observed frequencies can be less than 2).
- No more than 20% of the cells should have expected frequencies of less than 5.
- What if these rules of thumb are violated?

Small Expected Frequencies
- Chi-square test is an approximate method.
- The chi-square distribution is an idealized mathematical model.
- In reality, the statistics used in the chi-square test are qualitative (have discrete values and not continuous).
- For 2 X 2 tables, use Fisher’s Exact Test (i.e. \( P(x=k) \sim B(n, p) \)) if your expected frequencies are less than 2. (Section 6.6)

Fisher’s Exact Test: Description
- The Fisher’s exact test calculates the exact probability of the table of observed cell frequencies given the following assumptions:
  - The null hypothesis of independence is true
  - The marginal totals of the observed table are fixed
- Calculation of the probability of the observed cell frequencies uses the factorial mathematical operation.
  - Factorial is notated by ! which means multiply the number by all integers smaller than the number
- Example: \( 7! = 7*6*5*4*3*2*1 = 5040 \)
**Fisher's Exact Test: Calculation**

If margins of a table are fixed, the exact probability of a table with cells $a, b, c, d$ and marginal totals $(a+b), (c+d), (a+c), (b+d) = (a + b)! \times (c + d)! \times (a + c)! \times (b + d)! / n! \times a! \times b! \times c! \times d!$

**Fisher's Exact Test: Calculation Example**

The exact probability of this table = $9! \times 9! \times 13! \times 5! / 18! \times 1! \times 8! \times 4! \times 5! = 136080 / 1028160 = 0.132$

**Probability for all possible tables with the same marginal totals**

- The following slide shows the 6 possible tables for the observed marginal totals: 9, 9, 5, 13. The probability of each table is also given.
- The observed table is Table II
- The p-value for the Fisher's exact test is calculated by summing all probabilities less than or equal to the probability of the observed table.
- The probability is smallest for the tables (tables I and VI) that are least likely to occur by chance if the null hypothesis of independence is true.

**Set of 6 possible tables with marginal totals: 9,9,5,13**

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>b</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>c</td>
<td>5</td>
<td>8</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>d</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>13</td>
<td>18</td>
</tr>
<tr>
<td>a+b</td>
<td>9</td>
<td>16</td>
<td>20</td>
<td>13</td>
<td>16</td>
<td>23</td>
</tr>
<tr>
<td>c+d</td>
<td>9</td>
<td>13</td>
<td>14</td>
<td>9</td>
<td>13</td>
<td>18</td>
</tr>
<tr>
<td>a+c</td>
<td>15</td>
<td>13</td>
<td>12</td>
<td>18</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>b+d</td>
<td>13</td>
<td>15</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>18</td>
</tr>
</tbody>
</table>

**Fisher's Exact Test: p-value**

The observed table (Table II) has probability $= 0.132$

P-value for the Fisher's exact test = $Pr(\text{Table II}) + Pr(\text{Table V}) + Pr(\text{Table I}) + Pr(\text{Table VI}) = 0.132 + 0.132 + 0.0147 + 0.0147 = 0.293$

**Conclusion of Fisher's Exact test**

- At significance level 0.05, the null hypothesis of independence is not rejected because the p-value of 0.294 > 0.05.
- Looking back at the probabilities for each of the 6 tables, only Tables I and VI would result in a significant Fisher’s exact test result: $p = 2 \times 0.0147 = 0.0294$ for either of these tables.
- This makes sense, intuitively, because these tables are least likely to occur by chance if the null hypothesis is true.
Fisher’s Exact test Calculator for a 2x2 table

This website will calculate the Fisher’s exact test p-value after you enter the cell counts for a 2 x 2 contingency table.

Use the p-value for the two-sided test.

www.graphpad.com/quickcalcs/contingency1.cfm

Tests for Categorical Data

- To compare proportions between two groups or to test for independence between two categorical variables, use the Chi-square test
- If more than 20% of the expected cell frequencies < 5, use the Fisher’s exact test
- When categorical data are paired, the McNemar test is the appropriate test.

Comparing Proportions with Paired data

- When data are paired and the outcome of interest is a proportion, the McNemar Test is used to evaluate hypotheses about the data.
  - Developed by Quinn McNemar in 1947
  - Sometimes called the McNemar Chi-square test because the test statistic has a Chi-square distribution
- The McNemar test is only used for paired nominal data.
  - Use the Chi-square test for independence when nominal data are collected from independent groups.

Examples of Paired Data for Proportions

- Pair-Matched data can come from
  - Case-control studies where each case has a matching control (matched on age, gender, race, etc.)
  - Twins studies – the matched pairs are twins.
  - Before - After data
    - the outcome is presence (+) or absence (-) of some characteristic measured on the same individual at two time points.

Summarizing the Data

- Like the Chi-square test, data need to be arranged in a contingency table before calculating the McNemar statistic
- The table will always be 2 X 2 but the cell frequencies are numbers of ‘pairs’ not numbers of individuals
- Examples for setting up the tables are in the following slides for
  - Case – Control paired data
  - Twins paired data: one exposed and one unexposed
  - Before – After paired data

Pair-Matched Data for Case-Control Study: outcome is exposure to some risk factor

<table>
<thead>
<tr>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exposed</td>
</tr>
<tr>
<td>Exposed</td>
<td>a</td>
</tr>
<tr>
<td>Unexposed</td>
<td>c</td>
</tr>
</tbody>
</table>

- a - number of case-control pairs where both are exposed
- b - number of case-control pairs where the case is exposed and the control is unexposed
- c - number of case-control pairs where the case is unexposed and the control is exposed
- d - number of case-control pairs where both are unexposed

The counts in the table for a case-control study are numbers of pairs not numbers of individuals.
Paired Data for Before-After counts

- The data set-up is slightly different when we are looking at ‘Before-After’ counts of some characteristic of interest.
- For this data, each subject is measured twice for the presence or absence of the characteristic: before and after an intervention.
- The ‘pairs’ are not two paired individuals but two measurements on the same individual.
- The outcome is binary: each subject is classified as + (characteristic present) or – (characteristic absent) at each time point.

Paired Data for Before-After Counts

<table>
<thead>
<tr>
<th>Before treatment</th>
<th>After treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>a</td>
</tr>
<tr>
<td>+</td>
<td>b</td>
</tr>
<tr>
<td>-</td>
<td>c</td>
</tr>
<tr>
<td>-</td>
<td>d</td>
</tr>
</tbody>
</table>

- a - number of subjects with characteristic present both before and after treatment
- b - number of subjects where characteristic is present before but not after
- c - number of subjects where characteristic is present after but not before
- d - number of subjects with the characteristic absent both before and after treatment.

Null hypotheses for Paired Nominal data

- The null hypothesis for case-control pair matched data is that the proportion of subjects exposed to the risk factor is equal for cases and controls.
- The null hypothesis for twin paired data is that the proportions with the event are equal for exposed and unexposed twins.
- The null hypothesis for before-after data is that the proportion of subjects with the characteristic (or event) is the same before and after treatment.

McNemar’s test

- For any of the paired data Null Hypotheses the following are true if the null hypothesis is true:
  \[ b = c \]
  \[ \frac{b}{b+c} = 0.5 \]
- Since cells ‘b’ and ‘c’ are the cells that identify a difference, only cells ‘b’ and ‘c’ are used to calculate the test statistic.
- Cells ‘b’ and ‘c’ are called the discordant cells because they represent pairs with a difference.
- Cells ‘a’ and ‘d’ are the concordant cells. These cells do not contribute any information about a difference between pairs or over time so they aren’t used to calculate the test statistic.

McNemar Statistic

- The McNemar’s Chi-square statistic is calculated using the counts in the ‘b’ and ‘c’ cells of the table:
  \[ \chi^2 = \frac{(b - c)^2}{b + c} \]
- Square the difference of (b-c) and divide by b+c.
- If the null hypothesis is true the McNemar Chi-square statistic = 0.

McNemar statistic distribution

- The sampling distribution of the McNemar statistic is a Chi-square distribution.
- Since the McNemar test is always done on data in a 2 X 2 table, the degrees of freedom for this statistic = 1.
- For a test with alpha = 0.05, the critical value for the McNemar statistic = 3.84.
  - The null hypothesis is not rejected if the McNemar statistic < 3.84.
  - The null hypothesis is rejected if the McNemar statistic > 3.84.
P-value for McNemar statistic

- You can find the p-value for the McNemar statistic using the CHIDIST function in Excel.
- Enter = CHIDIST(test statistic, 1) to obtain the p-value.
- If the test statistic is > 3.84, the p-value will be < 0.05 and the null hypothesis of equal proportions between pairs or over time will be rejected.

McNemar test Example

- In 1989 results of a twin study were published in Social Science and Medicine: Twins, smoking and mortality: a 12 year prospective study of smoking-discordant twin pairs. Kaprio J and Koskenvuo M.
- 22 pairs of twins were enrolled in the study. One of the twins smoked, the other didn’t. The twins were followed to see which twin died first.
  - For 17 pairs of twins: the smoking twin died first
  - For 5 pairs of twins: the nonsmoking twin died first

Data for Twin study in a table

<table>
<thead>
<tr>
<th>Non-smoking Twin</th>
<th>Died 1st</th>
<th>Died 2nd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking Twin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Died 1st</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>Died 2nd</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

McNemar test hypotheses

- $H_0$: The proportion of smoking twins who died first is equal to the proportion of nonsmoking twins who died first.
- $H_A$: The proportion of smoking twins who died first is not equal to the proportion of nonsmoking twins who died first.

- In this study that counts which twin dies first, all data are discordant and are in the ‘b’ and ‘c’ cells.

McNemar test

- Significance level of the test = 0.05
- Critical value for Chi-square distribution with 1 df = 3.84
- Calculate the test statistic
  \[ \chi^2 = \frac{(b - c)^2}{b + c} = \frac{(17 - 5)^2}{17 + 5} = 6.54 \]
- P-value = 0.01
  - =CHIDIST(6.54, 1)

Decision and Conclusion for Twin study

- Decision: The null hypothesis of equal proportions of first death for smoking and non-smoking twins is rejected.
  - By the rejection region method: 6.54 > 3.84
  - By the p-value method: 0.01 < 0.05
- Conclusion: A significantly different proportion of smoking twins died first compared to their non-smoking twin indicating a different risk of death associated with smoking (p = 0.01).
McNemar test in EXCEL and online calculator

There is no EXCEL function or Data Analysis Tool for the McNemar Chi-square test.

This website will calculate the McNemar test statistic and p-value

http://www.graphpad.com/quickcalcs/McNemar1.cfm

Readings and Assignments

- Reading
  - Chapter 6 pgs 149 – 153
  - Chapter 5 pgs 119-121
- Work through the Lesson 12 Practice Exercises
- Lesson 12 Excel Module
- Complete Homework 8 and submit by the due date