Math507M: Statistics for Medical and Public Health Researchers

Instructor: Nan LIN
nlin@wustl.edu

Department of Mathematics, College of Arts and Sciences
Division of Biostatistics, School of Medicine
Washington University in St. Louis

Acknowledgement: These slides are made based on course materials previously used by Jeff Gill and Han Liang Gan.
Background

This chapter gives additional procedures for comparing two groups:

<table>
<thead>
<tr>
<th>Level of Measurement</th>
<th>Continuous</th>
<th>Ordinal</th>
<th>Nominal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paired Data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent (Unpaired) Data</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

under weaker assumptions:

- different distributional assumptions due smaller sample sizes providing a weaker justification for normally distributed statistics (CLT),
- less precise levels of measurement than last chapter (eg. above table),
- and some experimental settings that impose more heterogeneity on sample comparisons.
Running Example


- **Objectives:** To establish the relative cost effectiveness of community leg ulcer clinics that use four layer compression bandaging versus usual care provided by district nurses.

- **Design:** Randomised controlled trial with 1 year of follow up.

- **Setting:** Eight community based research clinics in four trusts in Trent.

- **Subjects:** 233 patients with venous leg ulcers allocated at random to intervention (120) or control (113) group.

- **Interventions:** Weekly treatment with four layer bandaging in a leg ulcer clinic (clinic group) or usual care at home by the district nursing service (control group).
Outcome Data:
- time to complete ulcer healing in weeks
- ulcer status (healed or unhealed) at 3 months and 12 months
- total ulcer free weeks
- “health-related quality of life (HRQoL) at 0, 3, 12 months, measured from 0 (very poor health) to 100 (very good health)
- some others not considered in the textbook.

Standard Thresholds: “To have an 80% chance of detecting as significant (at the 5% level) an increase in healing from 50% to 70%, 206 patients were required.”

Procedures: “We used t tests or Mann-Whitney tests for continuous measurements, \( \chi^2 \) tests for categorical data, the Kaplan-Meier method to calculate the time from recruitment for the initial leg ulcer to heal, and the log rank test to compare the healing times of both groups.”

Summary: Table 1 at http://www.bmj.com/content/316/7143/1487
Paired Data

- Paired data come from non-independent groupings.
- Two main types:
  - **Matched Cases Data**: a case-control study that “matches” similar cases together for comparison, the same case measured at different points in time.
  - **Cross-Over Trial Data**: subjects get of different treatments (or exposures) at different points in time, typically in RCT settings.
- Key statistical issue: *the variability is less than with independent comparisons.*
Standard Tests For Paired Samples

- **Continuous Data:**
  - Normally distributed data: paired Students-$t$ test.
  - Otherwise: Wilcoxon Signed Rank (matched pairs) test.

- **Ordinal Data:**
  - sign test or Wilcoxon Signed Rank (matched pairs) test.

- **Nominal Data:**
  - Two categories: McNemar’s Test.
Paired Students-\(t\) Test Setup

- **Group 1:** \(x_{11}, x_{12}, \ldots, x_{1n}\).
- **Group 2:** \(x_{21}, x_{22}, \ldots, x_{2n}\).
- **Pair these:** \((x_{11}, x_{21}), (x_{12}, x_{22}), \ldots, (x_{1n}, x_{2n})\).
- **Define the paired difference:**
  \[ d_1 = x_{11} - x_{21}, d_2 = x_{12} - x_{22}, \ldots, d_n = x_{1n} - x_{2n}. \]
- **\(H_0\):** \(\delta = 0\).
- **\(H_A\):** \(\delta \neq 0\) (most common), \(H_A: \delta > 0\) or \(H_A: \delta < 0\).
- **Assumptions:**
  - \(d_i \sim \mathcal{N}(\delta, \sigma^2_\delta)\),
  - \(d_i\) independent of \(d_j, \forall i \neq j\).
Paired Students-\(t\) Test Steps

- Obtain: \(d_1, d_2, \ldots, d_n\) from the data.
- Calculate the mean and standard deviation of these values:

\[
\bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i \\
\sigma_d = \frac{1}{n-1} \sqrt{\sum_{i=1}^{n} (d_i - \bar{d})^2}.
\]  

(1)

- Use the latter to get the standard error of the mean difference:

\[
SE(\bar{d}) = \frac{\sigma_d}{\sqrt{n}}.
\]

(2)

- The test statistic is distributed Students-\(t\) with \(df = n - 1\):

\[
t = \frac{\bar{d} - 0}{SE(\bar{d})}
\]

(3)

under the null assumption.
Pairing the Same Case At Two Points In Time

- **Two Groups HRQoL:**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>n</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero Months (baseline)</td>
<td>66.3</td>
<td>36</td>
<td>18.8</td>
<td>3.1</td>
</tr>
<tr>
<td>Three Months</td>
<td>58.9</td>
<td>36</td>
<td>22.0</td>
<td>3.7</td>
</tr>
</tbody>
</table>

- **Students-\(t\) Test For \( \bar{d} \):**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>95% CI</th>
<th>t</th>
<th>df</th>
<th>2*qt(0.025,df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline-3 Months</td>
<td>7.3</td>
<td>16.5</td>
<td>2.8</td>
<td>[1.7:12.9]</td>
<td>2.661</td>
<td>35</td>
<td>2*qt(0.025,35)</td>
</tr>
</tbody>
</table>

- The confidence interval here is calculated by:

\[
[\bar{d} \pm t_{df,\alpha} \times SE(\bar{d})] = [7.3 \pm 2.03 \times 2.8] = [1.7 : 12.9] \tag{4}
\]

where:

\[
qt(0.025,df=35,lower.tail=FALSE)
\]

# [1] 2.030108
Paired Sample Tests of Center, Example in R

Effectiveness of AZT treatment for HIV: 300mg versus 600mg; outcome is p24 antigen level in pg/ml, (a viral protein that is detectable early after infection); only 10 pairs (20 cases).

```r
hist(y-x)
```

![Histogram of y - x](image)
t.test(x, y, var.equal=TRUE, alternative = "two.sided",
    paired = TRUE, conf.level = 0.95)

##
## Paired t-test
##
## data: x and y
## t = -2.1437, df = 9, p-value = 0.06066
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -22.4024507 0.6024507
## sample estimates:
## mean of the differences
## -10.9
This is the same as

```r
t.test(x-y, alternative = "two.sided", conf.level = 0.95)
```

```r
##
## One Sample t-test
##
## data:  x - y
## t = -2.1437, df = 9, p-value = 0.06066
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -22.4024507 0.6024507
## sample estimates:
##  mean of x
##  -10.9
```
Wilcoxon Signed Rank Test: For ordinal or non-normal data

- Consider ordinal data as “ranks.”
- Ranks are resistant to outliers.
- Removes information about distributional shape (therefore sometimes called nonparametric).
- Group 1: \( x_{11}, x_{12}, \ldots, x_{1n} \).
- Group 2: \( x_{21}, x_{22}, \ldots, x_{2n} \).
- Pair these: \((x_{11}, x_{21}), (x_{12}, x_{22}), \ldots, (x_{1n}, x_{2n})\).
- Define the paired difference:
  \[ d_1 = x_{11} - x_{21}, \ d_2 = x_{12} - x_{22}, \ldots, d_n = x_{1n} - x_{2n}. \]
- \( H_0 \): median(\( \delta \)) = 0.
- \( H_A \): median(\( \delta \)) \( \neq \) 0, (most common), \( H_A \): median(\( \delta \)) > 0, or \( H_A \): median(\( \delta \)) < 0.
- Assumptions: \( d_i \sim \text{symmetric}, \ \forall i, d_i \text{ independent of } d_j, i \neq j \).
Wilcoxon Signed Rank Test Steps

- From the data obtain:
  - the absolute value of the paired differences: $|d_1|, |d_2|, \ldots, |d_n|$ and
  - the sign of the paired differences: $\text{sign}(d_1), \text{sign}(d_2), \ldots, \text{sign}(d_n)$.

- Rank the absolute differences, discarding values equal to zero and averaging ties, calling these $R_1, R_2, \ldots, R_{n'}$, where the new $n' < n$ from the discards.

- Calculate:

\[
T^+ = \left| \sum_{i=1}^{n'} \text{sign}(d_i) \times R_i \right| = \left| \sum_{i=1}^{n'} \text{signed rank of } d_i \right|. \tag{5}
\]

- The test statistic is:

\[
z = \frac{T^+ - \frac{n'(n'+1)}{4}}{\sqrt{\frac{n'(n'+1)(2n'+1)}{24}}}, \tag{6}
\]

which is asymptotically standard normal under $H_0$. 
<table>
<thead>
<tr>
<th>Subj.</th>
<th>X_A</th>
<th>X_B</th>
<th>original X_A−X_B</th>
<th>absolute X_A−X_B</th>
<th>rank of absolute X_A−X_B</th>
<th>signed rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>78</td>
<td>78</td>
<td>0</td>
<td>0</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>3</td>
<td>64</td>
<td>62</td>
<td>+2</td>
<td>2</td>
<td>1</td>
<td>+1</td>
</tr>
<tr>
<td>4</td>
<td>45</td>
<td>48</td>
<td>−3</td>
<td>3</td>
<td>2</td>
<td>−2</td>
</tr>
<tr>
<td>5</td>
<td>64</td>
<td>68</td>
<td>−4</td>
<td>4</td>
<td>3.5</td>
<td>−3.5</td>
</tr>
<tr>
<td>6</td>
<td>52</td>
<td>56</td>
<td>−4</td>
<td>4</td>
<td>3.5</td>
<td>−3.5</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>25</td>
<td>+5</td>
<td>5</td>
<td>5</td>
<td>+5</td>
</tr>
<tr>
<td>8</td>
<td>50</td>
<td>44</td>
<td>+6</td>
<td>6</td>
<td>6</td>
<td>+6</td>
</tr>
<tr>
<td>9</td>
<td>64</td>
<td>56</td>
<td>+8</td>
<td>8</td>
<td>7</td>
<td>+7</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>40</td>
<td>+10</td>
<td>10</td>
<td>8.5</td>
<td>+8.5</td>
</tr>
<tr>
<td>11</td>
<td>78</td>
<td>68</td>
<td>+10</td>
<td>10</td>
<td>8.5</td>
<td>+8.5</td>
</tr>
<tr>
<td>12</td>
<td>22</td>
<td>36</td>
<td>−14</td>
<td>14</td>
<td>10</td>
<td>−10</td>
</tr>
<tr>
<td>13</td>
<td>84</td>
<td>68</td>
<td>+16</td>
<td>16</td>
<td>11</td>
<td>+11</td>
</tr>
<tr>
<td>14</td>
<td>40</td>
<td>20</td>
<td>+20</td>
<td>20</td>
<td>12</td>
<td>+12</td>
</tr>
<tr>
<td>15</td>
<td>90</td>
<td>58</td>
<td>+32</td>
<td>32</td>
<td>13</td>
<td>+13</td>
</tr>
<tr>
<td>16</td>
<td>72</td>
<td>32</td>
<td>+40</td>
<td>40</td>
<td>14</td>
<td>+14</td>
</tr>
</tbody>
</table>

W = 67.0  
N = 14
Wilcoxon signed rank test in R

Consider comparing two groups measuring tumour size (grams) in female Buffalo rats:

```r
A <- c(5.81, 1.02, 1.14, 2.13, 2.57, 1.13, 1.06, 1.28, 3.22, 2.79)
B <- c(1.59, 2.71, 6.64, 4.66, 1.15, 1.28, 5.74, 3.23, 1.27, 1.39)
wilcox.test(A,B,paired = TRUE,exact=FALSE,correct=FALSE)
```

```r
##
## Wilcoxon signed rank test
##
## data:  A and B  
## V = 19, p-value = 0.3863
## alternative hypothesis: true location shift is not equal to 0
```
Continuity correction

The “continuity correction” is often considered necessary when using a continuous distribution to approximate a discrete distribution. In this case, it adds 0.5 in the direction of the mean to each value before the sum, to help with small sample size.

```
wilcox.test(A,B,paired = TRUE,exact=FALSE, correct = TRUE)
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data:  A and B
## V = 19, p-value = 0.4148
## alternative hypothesis: true location shift is not equal to 0
```
For small samples, more appropriately, one should use the EXACT option, which avoids the normal approximation.

```r
wilcox.test(A,B,paired = TRUE,exact=TRUE)
```

```
##
## Wilcoxon signed rank test

## data:  A and B
## V = 19, p-value = 0.4316
## alternative hypothesis: true location shift is not equal to 0
```
Comparing Two Paired Groups, Nominal Outcome

Brown et al (1987) performed a matched case-control study on all cases of testicular cancer of young men in the DC area from 1/1/76 to 6/30/86 (page 137 in the book).

Control: men in the same hospital as the cases diagnosed with malignancy other than cancer of the genital tract, similar in age (±2 years), race, vital status, and year of diagnosis.

Key explanatory variable of interest: undescended testes at birth.

The analysis focuses on the discordant pairs: no undescended testes and cancer, undescended testes and no cancer:

<table>
<thead>
<tr>
<th>Undescended</th>
<th>Testicular Cancer</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>4</td>
<td>11</td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>No</td>
<td>3</td>
<td>241</td>
<td></td>
<td>244</td>
</tr>
<tr>
<td>Total</td>
<td>7</td>
<td>252</td>
<td></td>
<td>259</td>
</tr>
</tbody>
</table>

The book takes the $b$ and $c$ values from the $2 \times 2$ table and relabels them as $f$ and $g$, “discordant pairs.”

The null hypothesis is that the discordant pairs are equal, $H_0 : f = g$. 
Comparing Two Paired Groups, Nominal Outcome

- McNemar’s test statistic is given by:

\[
\chi^2_{\text{McNemar}} = \frac{(f - g)^2}{f + g} = \frac{(11 - 3)^2}{11 + 3} = 4.5714,
\]

which has \(df = 1\) since it comes from a 2 \(\times\) 2 table (number of rows minus one times number of columns minus one). And reject \(H_0\) when the test statistic is large.

\[
pchisq(4.5714, df=1, lower.tail=FALSE) \quad \# p-value
\]

\[
\# [1] 0.03250999
\]

- A Yate’s-style correction for small samples uses:

\[
\chi^2_{\text{McNemar}} = \frac{(|f - g| - 1)^2}{f + g},
\]
Chi-square distribution
McNemar test in R

```r
x <- matrix(c(4, 3, 11, 241), 2, 2)
mcnemar.test(x, correct=TRUE)
```

```
##
## McNemar's Chi-squared test with continuity correction
##
## data: x
## McNemar's chi-squared = 3.5, df = 1, p-value = 0.06137
```
Paired Samples With Nominal Outcome

- Now let’s ask about the exact p-value of McNemar’s test.
- Returning to the Brown et al data, we start with the table:

\[
\begin{array}{cc}
(i) & \\
4 & 11 \\
3 & 241 \\
\end{array}
\]

- We then identify 3 more tables that make the off-diagonal values more discordant:

\[
\begin{array}{ccc}
(ii) & (iii) & (iv) \\
4 & 12 & 4 & 13 & 4 & 14 \\
2 & 241 & 1 & 241 & 0 & 241 \\
\end{array}
\]

- For the original table and the three more discordant tables, calculate:

\[
p = \frac{(f + g)!}{f!g!} \left( \frac{1}{2} \right)^{f+g}.
\]
Paired Samples With Nominal Outcome

- For the 4 tables, the probabilities are:

\[ p(i) = \frac{14!}{11!3!} \left( \frac{1}{2} \right)^{11+3} = 0.022217 \]

\[ p(i) = \frac{14!}{12!2!} \left( \frac{1}{2} \right)^{11+3} = 0.005554 \]

\[ p(i) = \frac{14!}{13!1!} \left( \frac{1}{2} \right)^{11+3} = 0.000854 \]

\[ p(i) = \frac{14!}{14!0!} \left( \frac{1}{2} \right)^{11+3} = 0.000061 \]

- The total 0.028686 is then the p-value for a one-sided alternative.

- To get the p-value for the two-sided alternative, one doubles it and it becomes 0.057372 (the book rounds this to 0.06).
Exact p-value of McNemar’s test

```r
library(exact2x2)
mcnemar.exact(x)
```

```
##
## Exact McNemar test (with central confidence intervals)
##
## data:  x
## b = 11, c = 3, p-value = 0.05737
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.9685982 20.4687696
## sample estimates:
## odds ratio
## 3.666667
```
Standard Tests For Independent (Unpaired) Samples

- **Continuous Data:**
  - Normally distributed data, modest samples: independent samples Students-\( t \) test.
  - Non-normally distributed data, modest samples: Wilcoxon (as before) or Mann-Whitney \( U \) test (generally preferred).

- **Ordinal Data:**
  - Mann-Whitney \( U \) test for difference.
  - Chi-Squared test for trend.

- **Nominal Data:**
  - Two Categories: proportions Chi-Square test, add Yates correction if all expected frequencies are not greater than five.
  - More Than Two Categories: Chi-Square cross-tabs test, collapse categories if all expected frequencies are not greater than five.
Independent Samples With Continuous Outcome, Students-\( t \) Test Setup

- **Group 1**: \( x_{11}, x_{12}, \ldots, x_{1n_1} \).
- **Group 2**: \( x_{21}, x_{22}, \ldots, x_{2n_2} \).
- \( H_0 : \delta = 0 \).
- \( H_A : \delta \neq 0 \) (most common), \( H_A : \delta > 0 \) or \( H_A : \delta < 0 \).
- **Assumptions**:
  - variables are continuous and independent,
  - the group standard deviations are similar (one no more than about twice the other),
  - the population data for each group is distributed normal.
- **Restriction**: sample sizes are modest.
- **We are now switching from a normal distribution to a students-\( t \) distribution for the test statistic.**
Independent Samples With Continuous Outcome, Students-\( t \) Test Steps

- From the two samples get \( \bar{x}_1, \bar{x}_2, s_1, s_2 \), based on the sample sizes \( n_1 \) and \( n_2 \).

- Calculate the mean difference:
  \[
  \bar{d} = \bar{x}_1 - \bar{x}_2. \quad (8)
  \]

- Calculate the pooled standard deviation:
  \[
  SD_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \quad (9)
  \]

- Use this to get the standard error of the mean difference:
  \[
  SE(\bar{d}) = SD_{pooled} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}. \quad (10)
  \]

- The test statistic is distributed Students-\( t \) with \( df = n_1 + n_2 - 2 \) under the null:
  \[
  t = \frac{\bar{d} - 0}{SE(\bar{d})} \quad (11)
  \]
### Independent Samples With Continuous Outcome, Students-$t$ Test Example

- **Leg Ulcer-Free Time in Weeks:**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>$n$</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention Group</td>
<td>20.1</td>
<td>120</td>
<td>18.5</td>
<td>1.7</td>
</tr>
<tr>
<td>Control Group</td>
<td>14.2</td>
<td>113</td>
<td>17.6</td>
<td>1.7</td>
</tr>
</tbody>
</table>

- **Students-$t$ Test for a two-sided alternative:**

<table>
<thead>
<tr>
<th></th>
<th>$t$</th>
<th>df</th>
<th>pvalue</th>
<th>Mean Diff</th>
<th>SE Diff</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference in Weeks</td>
<td>2.485</td>
<td>231</td>
<td>0.017883</td>
<td>5.9</td>
<td>2.4</td>
<td>[1.2:10.5]</td>
</tr>
</tbody>
</table>

pvalue is given by

$2*(1-\text{pt}(2.485, df=231))$

```r
#> [1] 0.0136645
```

- **The confidence interval here is calculated by:**

$$\bar{d} \pm t_{df,\alpha} \times SE(\bar{d}) = [5.9 \pm 1.97 \times 2.4] = [1.2:10.5]$$

```
qt(0.025, df=231, lower.tail=FALSE)
```

```r
#> [1] 1.970287
```
Independent Samples With Continuous Outcome, No Normal Data Assumption

- For this case we prefer the Mann-Whitney Test (also confusingly called a two-sample Wilcoxon test).
- Start: combine the two groups and rank the full sample in decreasing order (again tied values get the mean of the two adjacent ranks), to produce a statistic $W$ (to be described).
- The test statistic is distributed standard normal:

$$z = \frac{W - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1n_2(n_1+n_2+1)}{12}}}.$$  

(12)

- Some software uses slight variations on this with better small sample properties.
- $H_0 : \delta = 0$.
- $H_A : \delta \neq 0$ (most common), $H_A : \delta > 0$ or $H_A : \delta < 0$.
- Assumptions: groups are independent, cases can be ordered.
Independent Samples With Continuous Outcome, No Normal Data Assumption

Steps for a small sample Mann-Whitney test:

- select the group with smaller total rank (for convenience),
- order both groups together,
- for each observation in this group count the number of times it is lower than the other group,
- $W$ is the sum.

Buffalo Rats Example:

A: 5.8 1.0 1.1 2.1 2.5 1.1 1.0 1.2 3.2 2.7
B: 1.5 2.7 6.6 4.6 1.1 1.2 5.7 3.2 1.2 1.3
Independent Samples With Continuous Outcome, No Normal Data Assumption

A <- c(5.8, 1.0, 1.1, 2.1, 2.5, 1.1, 1.0, 1.2, 3.2, 2.7)
B <- c(1.5, 2.7, 6.6, 4.6, 1.1, 1.2, 5.7, 3.2, 1.2, 1.3)
rats <- data.frame(y = c(A,B),
                   group = c(rep('A',length(A)),rep('B',length(B))))
wilcox.test(y~group, data=rats)

##
## Wilcoxon rank sum test with continuity correction
##
## data:  y by group
## W = 34, p-value = 0.2394
## alternative hypothesis: true location shift is not equal to 0
Independent Samples With Nominal Outcome

- **Approach 1: Difference In Proportions:**
  - $H_0: \pi_{\text{Intervention}} - \pi_{\text{Control}} = 0$ versus
  - $H_A: \pi_{\text{Intervention}} - \pi_{\text{Control}} \neq 0$.
  - The test assumes that there is a common underlying population proportion, $\pi$, estimated by:
    \[
    \hat{\pi} = \frac{n_1 \hat{\pi}_1 + n_2 \hat{\pi}_2}{n_1 + n_2}
    \]  
    (13)
  
  with standard error estimated by:
  \[
  SE(\hat{\pi}_1 - \hat{\pi}_2) = \sqrt{\hat{\pi}(1 - \hat{\pi}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}.
  \]  
  (14)
  
  - The test statistic is therefore:
    \[
    z = \frac{\hat{\pi}_1 - \hat{\pi}_2}{SE(\hat{\pi}_1 - \hat{\pi}_2)}
    \]  
    (15)
  
  where tail values imply a difference in the two proportions.
Independent Samples With Nominal Outcome

Example of *leg ulcers healed at 3 months*:

\[
p_1 = \frac{\text{number healed under intervention}}{\text{number in intervention}} = \frac{22}{120} = 0.183
\]

\[
p_2 = \frac{\text{number healed under control}}{\text{number in control}} = \frac{17}{113} = 0.150
\]

\[
p = \frac{(120 \cdot 0.183) + (113 \cdot 0.150)}{120 + 113} = 0.167
\]

\[
SE(p_1 - p_2) = \sqrt{0.167(1 - 0.167) \left(\frac{1}{120} + \frac{1}{113}\right)} = 0.049
\]

\[
z = \frac{0.183 - 0.150}{0.049} = 0.673
\]

\[2*(1-pnorm(0.673)) \quad \# pvalue
\]

## [1] 0.5009473
Independent Samples With Nominal Outcome

- A confidence interval without the (restrictive) assumption of a single underlying population means that we can use the following SE:

\[
SE(\hat{\pi}_1 - \hat{\pi}_2) = \sqrt{\frac{\hat{\pi}_1(1 - \hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1 - \hat{\pi}_2)}{n_2}}
\]

\[
= \sqrt{\frac{0.183 \cdot 0.817}{120} + \frac{0.150 \cdot 0.85}{113}} = 0.0487
\]

- The 95% CI for the difference of proportions is now:

\[
(p_1 - p_2) \pm [z_\alpha \times SE(p_1 - p_2)] = (0.183 - 0.150) \pm [1.95 \times 0.0487] = [-0.062452 : 0.128452]
\]
Independent Samples With Nominal Outcome

- Approach 2: \( r \times c \) Contingency Tables:
  - \( H_0 \): no association/relationship between row and column variables versus \( H_A \): there is a relationship between levels of row variables and levels of column variables.
  - The test assumes that (1) two independent unordered variables, (2) at least 80% of expected cell counts have more than five values, and (3) no zero expected cell counts.
  - Expected cell counts:
    \[
    E_{ij} = \frac{\text{Row Total}(R_i) \times \text{Column Total}(C_j)}{N}.
    \]
  - Calculate the difference between observed and expected for each cell: \( O_{ij} - E_{ij} \).
  - The test statistic is:
    \[
    \chi^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}},
    \]
    which has degrees of freedom equal to \( df = (r - 1)(c - 1) \).
Independent Samples With Nominal Outcome

- With small samples the distributional assumption is suspect, and Yate’s Correction is designed to improve this:
  \[ \chi^2_Y = \sum_{ij} \frac{(|O_{ij} - E_{ij}| - 0.5)^2}{E_{ij}}. \]
- As a reminder, the ulcer table is given by:

<table>
<thead>
<tr>
<th></th>
<th>Healed</th>
<th>Not Healed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>(a = 22)</td>
<td>(b = 98)</td>
<td>(m = 120)</td>
</tr>
<tr>
<td>Control</td>
<td>(c = 17)</td>
<td>(d = 96)</td>
<td>(n = 113)</td>
</tr>
<tr>
<td>Total</td>
<td>(r = 39)</td>
<td>(s = 194)</td>
<td>(N = 233)</td>
</tr>
</tbody>
</table>
### Independent Samples With Nominal Outcome

**Table of leg ulcers healed at 3 months:**

<table>
<thead>
<tr>
<th></th>
<th>( O )</th>
<th>( E )</th>
<th>( O - E )</th>
<th>( (O - E)^2 / E )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention/healed</td>
<td>22</td>
<td>20.086</td>
<td>1.914</td>
<td>0.1824</td>
</tr>
<tr>
<td>Intervention/not healed</td>
<td>98</td>
<td>99.914</td>
<td>1.914</td>
<td>0.0367</td>
</tr>
<tr>
<td>Control/healed</td>
<td>17</td>
<td>18.914</td>
<td>-1.914</td>
<td>0.1937</td>
</tr>
<tr>
<td>Control/not healed</td>
<td>96</td>
<td>94.086</td>
<td>-1.914</td>
<td>0.0389</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>233</td>
<td>233</td>
<td>0</td>
<td>0.4518</td>
</tr>
</tbody>
</table>
```r
exd <- obs <- c(22, 17, 98, 96)
diff <- obs - exd
chi <- diff^2 / exd
( chi.sq <- sum(chi) )
```

## [1] 0.4517515
obs <- matrix(c(22,17,98,96), ncol=2)
chisq.test(obs, correct=F)

##
## Pearson's Chi-squared test
##
## data: obs
## X-squared = 0.4518, df = 1, p-value = 0.5015
Independent Samples With Nominal Outcome, Fisher’s Exact Test I

- Approach 3: Fisher’s Exact Test, used for (mostly $2 \times 2$) tables with small samples where Yate’s correction is not sufficient.

- Consider the standard table setup:

<table>
<thead>
<tr>
<th>Factor A</th>
<th>Present</th>
<th>Absent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>$a$</td>
<td>$b$</td>
<td>$m$</td>
</tr>
<tr>
<td>Absent</td>
<td>$c$</td>
<td>$d$</td>
<td>$n$</td>
</tr>
<tr>
<td>Total</td>
<td>$r$</td>
<td>$s$</td>
<td>$N$</td>
</tr>
</tbody>
</table>

- The hypergeometric analysis asks, given all marginals fixed, what is the exact probability of observing these cell values:

$$p(\text{table}) = \frac{m!n!r!s!}{N!a!b!c!d!}$$
Next we do this calculation for *every* table with the same marginal values but giving the same or higher levels of association, add the probabilities, and double this for a two-sided test.
Independent Samples With Nominal Outcome, Fisher’s Exact Test

Consider data on deaths 6 months after fractured neck of femur in a specialized orthopaedic ward (Factor A) and a general ward (Factor B).

So: \[ a = 2, b = 2, c = 18, d = 14, m = 8, m = 8, n = 32, r = 20, s = 20, N = 40 \] (page 142).
Independent Samples With Nominal Outcome, Fisher’s Exact Test

- This gives the table:

<table>
<thead>
<tr>
<th></th>
<th>Orthopaedic Ward</th>
<th>General Ward</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>$a = 2$</td>
<td>$b = 6$</td>
<td>$m = 8$</td>
</tr>
<tr>
<td>Absent</td>
<td>$c = 18$</td>
<td>$d = 14$</td>
<td>$n = 32$</td>
</tr>
<tr>
<td>Total</td>
<td>$r = 20$</td>
<td>$s = 20$</td>
<td>$N = 40$</td>
</tr>
</tbody>
</table>

- Applying the formula:

\[
p(\text{table}) = \frac{m!n!r!s!}{N!a!b!c!d!} = \frac{8!32!20!20!}{40!2!6!18!14!} = 0.0957601 \quad (19)
\]

Calculated in R by:

```r
(gamma(9)*gamma(33)*gamma(21)*gamma(21))/
(gamma(41)*gamma(3)*gamma(7)*gamma(19)*gamma(15))
```

```r
## [1] 0.0957601
```
Independent Samples With Nominal Outcome, Fisher’s Exact Test

- There are two more tables (page 142) with the same marginal values that give stronger associations, and these have \( p(\text{table}) \) values of 0.020160 and 0.001638:

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>6</th>
<th>1</th>
<th>7</th>
<th>0</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18</td>
<td>14</td>
<td>19</td>
<td>13</td>
<td>20</td>
<td>12</td>
</tr>
</tbody>
</table>

with odds ratios of \( \frac{bc}{ad} = 10.231 \), and \( \frac{bc}{ad} = \infty \).

- For some reason the book sets up this table backwards, hence \( \frac{bc}{ad} \) instead of \( \frac{ad}{bc} \).

- So the summed probability is:

\[
0.0957601 + 0.020160 + 0.001638 = 0.1175581.
\]

- The probability of interest is double this for a two-sided test:

\[
2 \times 0.1175581 = 0.2351162.
\]
Independent Samples With Nominal Outcome, Fisher’s Exact Test

```r
ortho.df <- data.frame(c(2,18,6,14),expand.grid(1:2,1:2))
names(ortho.df) <- c("counts","death","ward")
( ortho.x <- xtabs(counts ~ death + ward, data=ortho.df) )

## ward
## death 1 2
## 1 2 6
## 2 18 14

fisher.test(ortho.x)
```

```
## Fisher's Exact Test for Count Data
##
data:  ortho.x
## p-value = 0.2351
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.02308176 1.80076397
```
Independent Samples With Nominal Outcome, Fisher’s Exact Test

An alternative implementation in R:

```r
( ortho.mat <- matrix(c(2,18,6,14), nrow=2, 
                      dimnames = list(c("Death", "No Death"), c("Ward A", "Ward B"))) 

## Ward A Ward B 
## Death 2 6 
## No Death 18 14

fisher.test(ortho.mat)
```

```
##
## Fisher's Exact Test for Count Data
##
## data:  ortho.mat
## p-value = 0.2351
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.02308176 1.80076397
## sample estimates:
## odds ratio
## 0.2680041
```
Crosstabs For a Trend in $2 \times c$ Tables

- Suppose we have a nominal/dichotomous variable crossed with an ordinal variable (such as pain index or cancer stage).
- The regular chi-square test is not efficient since it does not account for the ordering of the second variable.

The book gives the following contrived example on page 143:

<table>
<thead>
<tr>
<th>Outcome of Trial</th>
<th>Worse</th>
<th>Same</th>
<th>Slightly better</th>
<th>Moderately better</th>
<th>Much better</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Score ($x_i$)</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Treatment A ($a_i$)</td>
<td>11</td>
<td>53</td>
<td>42</td>
<td>27</td>
<td>11</td>
</tr>
<tr>
<td>Treatment B</td>
<td>1</td>
<td>13</td>
<td>16</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>Total ($n_i$)</td>
<td>12</td>
<td>66</td>
<td>58</td>
<td>42</td>
<td>18</td>
</tr>
<tr>
<td>$p_i = a_i/n_i$</td>
<td>0.9167</td>
<td>0.8030</td>
<td>0.7241</td>
<td>0.6429</td>
<td>0.6111</td>
</tr>
<tr>
<td>$q_i = 1 - a_i/n_i$</td>
<td>0.0833</td>
<td>0.1970</td>
<td>0.2759</td>
<td>0.3571</td>
<td>0.3889</td>
</tr>
</tbody>
</table>
Crosstabs for a Trend in $2 \times c$ Tables

- First calculate the treatment score variability:

$$T_{xp} = \sum_{i=1}^{c} n_i (p_i - \bar{p})(x_i - \bar{x}) = \sum_{i=1}^{c} a_i x_i - \frac{1}{N} \left( \sum_{i=1}^{c} a_i \sum_{i=1}^{c} n_i x_i \right).$$

(20)

where $c$ is the number of columns.

- Then calculate the score variability:

$$T_{xx} = \sum_{i=1}^{c} n_i x_i^2 - \frac{1}{N} \left( \sum_{i=1}^{c} (n_i x_i)^2 \right).$$

(21)

- This gives the test statistic for the trend:

$$X_{\text{trend}}^2 = \frac{T_{xp}^2}{T_{xx}} \times \frac{1}{\bar{p}} \times \frac{1}{\bar{q}},$$

(22)

which has $df = 1$. 


Crosstabs for a Trend in $2 \times c$ Tables

```r
a <- c(11, 53, 42, 27, 11)
n <- c(12, 66, 58, 42, 18)
x <- -2:2
p <- a/n; q <- 1-p
Txp <- sum(n*(p-mean(p))*(x-mean(x)))
Txx <- sum(n*(x^2)) - (sum(n*x))^2/196
(X.trend <- Txp^2/(Txx*mean(p)*mean(q)))

## [1] 6.699793

pchisq(X.trend, df=1, lower.tail=FALSE)

## [1] 0.009642408
```
Related Rate Models

- Accounts for occurrences, maximum possible events, time.
- Note that the binomial does not account for repeat events on the same unit.
- A key problem is that units may differ in size: disease counts are higher in bigger cities, etc.
- Focus on rate:

\[
\text{Rate} = \frac{\#\text{events}}{\text{unit}} = \frac{\text{occurrences}}{\text{possibilities}}
\]

- Cellular abnormalities from gamma radiation:
  - ca is the count of chromosomal abnormalities,
  - cells is the number (in hundreds) of exposed cells,
  - doseamt = dose amount,
  - doserate = rate of application.
Chi-Square Testing of Rate Models

Generally we use the more developed tools in R:

```r
library(faraway)
data(dicentric)
round(xtabs(ca/cells ~ doseamt + doserate, dicentric),2)
```

```
# doserate
doseamt  0.1  0.25  0.5  1  1.5  2  2.5  3  4
## 1   0.05  0.05  0.07  0.07  0.06  0.07  0.07  0.07
## 2.5 0.16  0.28  0.29  0.32  0.38  0.41  0.41  0.37  0.44
## 5   0.48  0.82  0.90  0.88  1.23  1.32  1.34  1.24  1.43
```

```r
summary(xtabs(ca/cells ~ doseamt + doserate, dicentric))
```

```
# Call: xtabs(formula = ca/cells ~ doseamt + doserate, data = dicentric)
# Number of cases in table: 13.27542
# Number of factors: 2
# Test for independence of all factors:
# Chisq = 0.04019, df = 16, p-value = 1
# Chi-squared approximation may be incorrect
```