Basic Data Analysis using OpenEpi

Zuber D. Mulla
Professor of OB/GYN
Assistant Dean for Faculty Development

February 5, 2020
After this session, the learner will be able to perform the following using OpenEpi

- A chi-square test
- A two-sample $t$-test
- A sample size calculation for a study that will compare two proportions (such as two risks)
OpenEpi

- Open source license
- Development was supported in part by a grant from the Bill and Melinda Gates Foundation to the Emory University Rollins School of Public Health
- www.openepi.com
Form of the dependent and independent variables

Statistical test or model that you will use
Review of Hypothesis Testing

- Null hypothesis ($H_0$) and alternative hypothesis ($H_A$)

- $H_0$ is usually the *hypothesis of no difference*

- The null hypothesis is the straw man
$H_0$: $RR = 1$

$H_A$: $RR \neq 1$
Two-Tailed Test

$H_0: \ RR = 1$

$H_A: \ RR \neq 1$

One-Tailed Test

$H_0: \ RR = 1$

$H_A: \ RR > 1$
<table>
<thead>
<tr>
<th></th>
<th>Ill</th>
<th>Well</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Not exposed</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>
Typically the independent variable \((X)\) is on the side. The figure illustrates a 2x2 contingency table where the independent variable \((X)\) is divided into two categories: Exposed and Not exposed. The dependent variables are divided into Ill and Well categories. The table shows:

- Exposed and Ill: A
- Exposed and Well: B
- Not Exposed and Ill: C
- Not Exposed and Well: D
Typically the dependent variable ($Y$) is on top.
Rows are horizontal
Columns are vertical
**What is a cell?** It’s the intersection of a row and column.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>
OpenEpi provides statistics for counts and measurements in disease with exact confidence limits, matched pair and person-time analysis, random numbers, sensitivity, specificity and other evaluation statistics, dose-response, and links to other useful sites.

OpenEpi is free and open source software for epidemiologic statistics, downloaded and run without a web connection. A server is not needed, and HTML, and should be compatible with recent Linux, Mac, and Windows systems. (If you are seeing this, your browser settings are allowing JavaScript, which is required to display the in-browser programs.)

Test results are provided for each module so that you can judge the importance of the problems and check important results with software from more than one source when available.

The programs have an open source license and can be downloaded for free. Components from other sources have licensing statements in the documentation. The complete source code is available in full text at OpenSource.org/licenses. OpenEpi development is supported by a grant from the Bill and Melinda Gates Foundation to Harvard University, Boston, MA, USA.
OpenEpi provides statistics for counts and measurements in disease risk with exact confidence limits, matched pair and person-time analysis, random numbers, sensitivity, specificity and other evaluation statistics, dose-response, and links to other useful sites.

OpenEpi is free and open source software for epidemiologic study and can be downloaded and run without a web connection. A server is not needed and should be compatible with recent Linux, Mac, and Windows system. (If you are seeing this, your browser settings are allowing websites to run in browsers of many iPhone and Android cellphones.

Test results are provided for each module so that you can judge and check important results with software from more than one source if provided.

The programs have an open source license and can be downloaded and components from other sources have licensing statements in the documentation. They are available in full text at OpenSource.org/licenses. OpenEpi developed by the Bill and Melinda Gates Foundation and Emory University.
OpenEpi: Question 1

- Two-sample $t$-test
- What is the null hypothesis?
OpenEpi: Question 1

- Use the “t test” module under “Continuous Variables”

- Click on the “Enter” tab and enter the required information

- Click “Calculate”
OpenEpi provides statistics for counts and measurements in disease with exact confidence limits, matched pair and person-time analysis, random numbers, sensitivity, specificity and other evaluation statistics, dose-response, and links to other useful sites.

OpenEpi is free and **open source** software for epidemiologic statistics downloadable and run without a web connection. A server is not needed and HTML, and should be compatible with recent Linux, Mac, and Windows system. (If you are seeing this, your browser settings are allowing you to use browsers of many iPhone and Android cellphones.

Test results are provided for each module so that you can judge and check important results with software from more than one source is provided.

The programs have an open source license and can be downloaded and components from other sources have licensing statements in the source available in full text at [OpenSource.org/licenses](http://www.opensource.org/licenses). OpenEpi developed by the Biomedical Health Research Foundation at Texas University.
OpenEpi: Question 2

- Use the “Two by Two Table” calculator/module under “Counts”

- Click on the “Enter” tab and enter the four cell values

- Click “Calculate”
Is there a relationship between tx and the subject’s sex?

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tx A</td>
<td>33</td>
<td>68</td>
</tr>
<tr>
<td>Tx B</td>
<td>34</td>
<td>65</td>
</tr>
</tbody>
</table>

Total: 101 (Tx A) and 99 (Tx B)
Tx status and subject’s sex

• The null hypothesis for the chi-square test of homogeneity is that the two tx groups are homogenous with respect to the prevalence of female subjects.

• Let alpha be 0.05.

• If the chi-square test $P$-value is 0.05 or less, then we reject the null hypothesis.
OpenEpi provides statistics for counts and measurements in disease with exact confidence limits, matched pair and person-time analysis, random numbers, sensitivity, specificity and other evaluation statistics, dose-response, and links to other useful sites.

OpenEpi is free and open source software for epidemiologic studies and can be downloaded and run without a web connection. A server is not needed and HTML, and should be compatible with recent Linux, Mac, and Windows system. (If you are seeing this, your browser settings are allowing the downloading of many iPhone and Android cellphones)

Test results are provided for each module so that you can judge, check important results with software from more than one source is provided.

The programs have an open source license and can be downloaded and components from other sources have licensing statements in the documentation. Available in full text at OpenSource.org/licenses. OpenEpi developed by the Bill and Melinda Gates Foundation, the Harvard University, and others.
OpenEpi: Questions 6 through 8

- Click on “Sample Size”
- Click on “Cohort/RCT”
- Click on the “Enter” tab
- Enter information & click “Calculate”