Test of Significance: Categorical Frequency Data

Dr. McKinley Thomas
Associate Professor
Department of Community Medicine
Mercer University

Utilized to investigate whether distributions of categorical variables differ from one another

FYI:
- Categorical [COUNTING]
- Continuous [MEASURING]

\[ \chi^2 \]

One Variable \( \chi^2 \)
- One variable (observed) compared to another variable (expected)
- Test of Goodness of Fit Test

Two Variable \( \chi^2 \)
- Two variables compared in relation to a null value
- Test of independence
You are interested in understanding if how patients report pain levels (categorical data) is any different in your post-surgery study than what would be expected ‘normally.’

You are interested in understanding if how patients report pain levels (categorical data) is any different in your post-surgery study than what would be expected ‘normally.’

### Goodness of Fit

<table>
<thead>
<tr>
<th>Category</th>
<th>Observed Frequency</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Pain</td>
<td>35</td>
<td>25</td>
</tr>
<tr>
<td>Moderate</td>
<td>30</td>
<td>25</td>
</tr>
<tr>
<td>Severe</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>Unbearable</td>
<td>15</td>
<td>25</td>
</tr>
<tr>
<td>TOTAL</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

### Analysis

- Calculations give us
  - df = (r-1)(c-1)
  - $\chi^2 = 10$, 3df
  - Check the table...
  - 10 > 7.82
    - Reject the null...

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

You are interested in comparing survival among participants who were randomly assigned to either a treatment or non-treatment groups.

\[ E_{ij} = \frac{(R \text{ marginal}_i \times C \text{ marginal}_j)}{N} \]

<table>
<thead>
<tr>
<th></th>
<th>Dead</th>
<th>Alive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treated</td>
<td>46 (40.97)</td>
<td>71 (76.02)</td>
<td>117</td>
</tr>
<tr>
<td>Untreated</td>
<td>37 (40.03)</td>
<td>83 (77.97)</td>
<td>120</td>
</tr>
<tr>
<td>Total</td>
<td>88</td>
<td>154</td>
<td>237</td>
</tr>
</tbody>
</table>

Analysis

- Calculations give us
  - \( df = (r-1)(c-1) \)
  - \( \chi^2 = 1.87, 1 \text{df} \)
  - Check the table...
  - \( 1.87 < 3.84 \)
  - Do not reject the null...

df

- Number of categories or classes being tested minus 1.
- May also be thought of as opportunities for change.
  - For example, if five random samples are drawn from a given population, there are four opportunities for change, or four degrees of freedom.
To calculate the $s^2$ of a random sample, we first calculate the $\bar{X}$ of that sample and then compute the sum of the several squared deviations from that mean.

While there will be $n$ such squared deviations only $(n - 1)$ of them are, in fact, free to assume any value whatsoever.

This is because the final squared deviation from the mean must include the one value of $X$ such that the sum of all the $X$s divided by $n$ will equal the obtained mean of the sample.