$\chi^2$ - test for independence

Comments on significance tests

Quiz
<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>W</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>55.7</td>
<td>87.3</td>
<td>143</td>
</tr>
<tr>
<td>L</td>
<td>4.3</td>
<td>6.7</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>94</td>
<td>154</td>
</tr>
</tbody>
</table>

Expected values

\[ \% \text{ of people who are RH} = \frac{143}{154} \times 100 = 93\% \]

\( H_0: \) 93\% of men should be RH
\( H_0: \) 93\% of women should be RH

Men: \( \frac{93}{100} \times 60 = 55.7 \)

Women: \( \frac{93}{100} \times 94 = 87.3 \)

\[ \text{Expected value} = \frac{\text{row total} \times \text{column total}}{\text{table total}} \]
### Testing Independence

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right handed</td>
<td>54</td>
<td>89</td>
<td>143</td>
</tr>
<tr>
<td>Left handed</td>
<td>6</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>Ambidextrous</td>
<td>60</td>
<td>94</td>
<td>154</td>
</tr>
</tbody>
</table>

Are "handedness" and sex independent?

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>90%</td>
<td>95%</td>
</tr>
<tr>
<td>L</td>
<td>10%</td>
<td>5%</td>
</tr>
</tbody>
</table>

Is this data consistent with a chance model that says that sex and handedness are independent?
Box Model.

Q: how many tickets of each type?

$H_0$: handedness + sex are independent

$\frac{\% \text{ of RH amongst men}}{\% \text{ of RH amongst women}}$ is same

$H_1$: $\% \text{ of RH differs bet. men + women.}$

Compute the expected values under $H_0$. 
\[ \chi^2 = \text{sum} \left( \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \right) \]

\[ = \frac{(54 - 55.7)^2}{55.7} + \frac{(89 - 87.3)^2}{87.3} \]

\[ + \frac{(6 - 4.3)^2}{4.3} + \frac{(5 - 6.7)^2}{6.7} \]

\[ = 1.19 \]

Also need the number of degrees of freedom.

The model this time is not fully specified - we don't know the true % of each ticket in the box - we estimate them from the data.

Contingency table with m rows and n cols, has \((m-1) \times (n-1)\) d.o.f.
A CHI-SQUARE TABLE

The chi-square curve, with degrees of freedom shown along the left of the table, is shown in the body of the table.

<table>
<thead>
<tr>
<th>Degrees of freedom</th>
<th>99%</th>
<th>95%</th>
<th>90%</th>
<th>70%</th>
<th>50%</th>
<th>30%</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00016</td>
<td>0.0039</td>
<td>0.016</td>
<td>0.15</td>
<td>0.46</td>
<td>1.07</td>
<td>2.71</td>
<td>3.84</td>
<td>6.64</td>
</tr>
<tr>
<td>2</td>
<td>0.020</td>
<td>0.10</td>
<td>0.21</td>
<td>0.71</td>
<td>1.39</td>
<td>2.41</td>
<td>4.60</td>
<td>5.99</td>
<td>9.21</td>
</tr>
<tr>
<td>3</td>
<td>0.12</td>
<td>0.35</td>
<td>0.58</td>
<td>1.42</td>
<td>2.37</td>
<td>3.67</td>
<td>6.25</td>
<td>7.82</td>
<td>11.34</td>
</tr>
<tr>
<td>4</td>
<td>0.30</td>
<td>0.71</td>
<td>1.06</td>
<td>2.20</td>
<td>3.36</td>
<td>4.88</td>
<td>7.78</td>
<td>9.49</td>
<td>13.28</td>
</tr>
<tr>
<td>5</td>
<td>0.55</td>
<td>1.14</td>
<td>1.61</td>
<td>3.00</td>
<td>4.35</td>
<td>6.06</td>
<td>9.24</td>
<td>11.07</td>
<td>15.09</td>
</tr>
<tr>
<td>6</td>
<td>0.87</td>
<td>1.64</td>
<td>2.20</td>
<td>3.83</td>
<td>5.35</td>
<td>7.23</td>
<td>10.65</td>
<td>12.59</td>
<td>16.81</td>
</tr>
<tr>
<td>7</td>
<td>1.24</td>
<td>2.17</td>
<td>2.83</td>
<td>4.67</td>
<td>6.35</td>
<td>8.38</td>
<td>12.02</td>
<td>14.07</td>
<td>18.48</td>
</tr>
<tr>
<td>8</td>
<td>1.65</td>
<td>2.73</td>
<td>3.49</td>
<td>5.33</td>
<td>7.34</td>
<td>9.52</td>
<td>13.36</td>
<td>15.51</td>
<td>20.09</td>
</tr>
<tr>
<td>10</td>
<td>2.56</td>
<td>3.94</td>
<td>4.86</td>
<td>7.27</td>
<td>9.34</td>
<td>11.78</td>
<td>15.99</td>
<td>18.31</td>
<td>23.21</td>
</tr>
<tr>
<td>11</td>
<td>3.05</td>
<td>4.58</td>
<td>5.58</td>
<td>8.15</td>
<td>10.34</td>
<td>12.90</td>
<td>17.28</td>
<td>19.68</td>
<td>24.73</td>
</tr>
<tr>
<td>12</td>
<td>3.57</td>
<td>5.23</td>
<td>6.30</td>
<td>9.03</td>
<td>11.34</td>
<td>14.01</td>
<td>18.55</td>
<td>21.03</td>
<td>26.22</td>
</tr>
<tr>
<td>13</td>
<td>4.11</td>
<td>5.89</td>
<td>7.04</td>
<td>9.93</td>
<td>12.34</td>
<td>15.12</td>
<td>19.81</td>
<td>22.36</td>
<td>27.69</td>
</tr>
<tr>
<td>14</td>
<td>4.66</td>
<td>6.57</td>
<td>7.79</td>
<td>10.82</td>
<td>13.34</td>
<td>16.22</td>
<td>21.06</td>
<td>23.69</td>
<td>29.14</td>
</tr>
<tr>
<td>15</td>
<td>5.23</td>
<td>7.26</td>
<td>8.55</td>
<td>11.72</td>
<td>14.34</td>
<td>17.32</td>
<td>22.31</td>
<td>25.00</td>
<td>30.58</td>
</tr>
<tr>
<td>16</td>
<td>5.81</td>
<td>7.96</td>
<td>9.31</td>
<td>12.62</td>
<td>15.34</td>
<td>18.42</td>
<td>23.54</td>
<td>26.30</td>
<td>32.00</td>
</tr>
<tr>
<td>17</td>
<td>6.41</td>
<td>8.67</td>
<td>10.09</td>
<td>13.53</td>
<td>16.34</td>
<td>19.51</td>
<td>24.77</td>
<td>27.59</td>
<td>33.41</td>
</tr>
<tr>
<td>18</td>
<td>7.00</td>
<td>9.39</td>
<td>10.87</td>
<td>14.44</td>
<td>17.34</td>
<td>20.60</td>
<td>25.99</td>
<td>28.87</td>
<td>34.81</td>
</tr>
<tr>
<td>19</td>
<td>7.63</td>
<td>10.12</td>
<td>11.65</td>
<td>15.35</td>
<td>18.34</td>
<td>21.69</td>
<td>27.20</td>
<td>30.14</td>
<td>36.19</td>
</tr>
<tr>
<td>20</td>
<td>8.26</td>
<td>10.85</td>
<td>12.44</td>
<td>16.27</td>
<td>19.34</td>
<td>22.78</td>
<td>28.41</td>
<td>31.41</td>
<td>37.57</td>
</tr>
</tbody>
</table>

In this case \((2-1) \times (2-1) = 1\) dof.

\[
\chi^2 = 1.19.
\]

**P-value**: what's the prob. of having data that gives \(\chi^2 \geq 1.19\) with 1 dof.

From the table, the P-value is between 30\% and 10\% (nearer to 30\% than 10\%).

P-value of \(\leq 25\%\) does not give significant evidence against \(H_0\). Cannot reject \(H_0\).

Conclude that handedness and sex are independent.
Another Example:

Random sample of 314 Hermit crab shells. Each classified as occupied or empty.
Is shell species independent of occupied/empty?
(Do hermit crabs prefer a particular species of shell?)

<table>
<thead>
<tr>
<th>Species</th>
<th>Occupied</th>
<th>Empty</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austrocochlea</td>
<td>67</td>
<td>52</td>
<td>89</td>
</tr>
<tr>
<td>Bembicium</td>
<td>10</td>
<td>30</td>
<td>51</td>
</tr>
<tr>
<td>C. iritida</td>
<td>125</td>
<td>101</td>
<td>174</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>182</strong></td>
<td><strong>132</strong></td>
<td><strong>314</strong></td>
</tr>
</tbody>
</table>

Expected values based on Ho: species and occupied/empty are independent

\[
EV = \frac{(\text{row total}) \times (\text{col total})}{\text{table total}}
\]

\[
X^2 = \sum \frac{(\text{obs} - \text{exp})^2}{\text{exp}}
\]

In red on table above
\[ \begin{align*}
&= \frac{(47 - 52)^2}{52} + \frac{(42 - 37)^2}{37} \\
&\quad + \frac{(10 - 30)^2}{30} + \frac{(41 - 21)^2}{21} \\
&\quad + \frac{(125 - 101)^2}{101} + \frac{(49 - 73)^2}{73} \\
&= 47.1
\end{align*} \]

\# d.o.f. = (\# rows - 1) \times (\# cols - 1) \\
&= (3 - 1) \times (2 - 1) \\
&= 2.

P-value for \( \chi^2 = 47.1 \) with 2 d.o.f. is \(< 1\%\). 

very strong evidence against null hypothesis 
that species is independent of occupation status.
Comments on tests of significance

1. The 5% and 1% thresholds are arbitrary but conventional.
   Don’t obsess over the difference between 4.9 and 5.1.

2. Recall what a p-value is:
   It’s the chance of observing data as extreme or more assuming H0 is true.
   - if you do enough tests, eventually you will get a statistically significant result.

   Decide how you will analyse the data before you collect the data.

3. One-sided vs Two-sided test.
   eg is this a fair coin?
   does the bottle filling machine over-/under fill?
   - Deviations in both directions cast doubt on H0.
Example: tossing a coin.
both too many heads and too many tails give evidence against coin being fair

\[ Z = \frac{\text{obs} - \text{exp}}{\text{SE}} \]

lots of heads: \( Z > 0 \)

lots of tails: \( Z < 0 \)

61 heads in 100 tosses.

\[ \text{SE} = \sqrt{100 \times (1-0) \times \frac{1}{2} \times \frac{1}{2}} \]

\[ = 5 \]

\[ Z = \frac{61 - 50}{5} = 2.2. \]

\[ Z = -2.2 \quad Z = 2.2 \]

- P-value is the sum of the two areas.
- two-tailed test.
The difference between "statistically significant" and "important"

- a particular problem when sample sizes are large.

\[ Z = \frac{\text{obs} - \text{exp.}}{\text{SE}} \]

\[ \text{SE}_\% = \frac{\text{SD}_{\text{obs}}}{\sqrt{\text{# draws}}} \times 100. \]

As \# draws increases, \( \text{SE}_\% \) decreases.
As \( \text{SE}_\% \) decreases, \( Z \)-statistic increases.

Eventually, we will get a p-value that is considered statistically significant.

- does it mean that it is important?

\[ \text{e.g.: } \text{violence linked to childhood candy} \]
\[ \text{rural vs. urban reading test scores.} \]
Cancer clusters at phone masts

SEVEN clusters of cancer and other serious illnesses have been discovered around mobile phone masts, raising concerns over the technology’s potential impact on health.

Studies of the sites show high incidences of cancer, brain haemorrhages and high blood pressure within a radius of 400 yards of mobile phone masts.

One of the studies, in Warwickshire, showed a cluster of 31 cancers around a single street. A quarter of the 30 staff at a special school within sight of the 90ft high mast have developed tumours since 2000, while another quarter have suffered significant health problems.

The mast is being pulled down by the mobile phone operator O2 after the presentation of the evidence by local protesters. While rejecting any link to ill-health, O2 admitted the decision was "clearly rare and unusual".

Phone masts have provoked protests throughout Britain with thousands of people objecting each week to planning applications. There are about 47,000 masts in the UK.

Dr John Walker, a scientist who compiled the cluster studies with the help of local campaigners in Devon, Lincolnshire, Staffordshire and the West Midlands, said he was convinced they showed a potential link between the angle of the beam of radiation emitted from the masts’ antennae and illnesses discovered in local populations.

"Masts should be moved away from conurbations and schools and the power turned down," he said.

Some scientists already believe such a link exists and studies in other European countries suggest a rise in cancers close to masts. In 2005 Sir William Stewart, chairman of the Health Protection Agency, said he found four such studies to be of concern but that the health risk remained unproven.
Could something else give the effect we've observed?

- Tests are based on a box model
- If the box model doesn't reflect reality, the test statistic isn't valid.

| □ □ □ □ □ □ | roll a die, use ESP to make it come up 6.

720 rolls
143 6's.

\[ Z = \frac{143 - 120}{10} \]

\[ SE = \sqrt{\frac{720 \times (1-0) \times \frac{1}{6} \times \frac{5}{6}}{10}} = 10 \]

\[ P = 1\% \quad \text{reject } H_0 \text{ that die is unbiased.} \]

Is ESP real?

Is the die biased?

Significance test does not tell you anything about which of these alternate hypotheses is more likely.
There are lots of different SEs

Basic one is for sum of draws

\[ SE_{\text{sum}} = \sqrt{\# \text{draws} \times SD \text{ box}} \]

\[ SE_{\text{ave}} = \frac{SE_{\text{sum}}}{\# \text{draws}} = \frac{SD \text{ box}}{\sqrt{\# \text{draws}}} \]

\[ SE_{\%} = SE_{\text{ave}} \times 100 = \frac{SD \text{ box}}{\sqrt{\# \text{draws}}} \times 100 \]

When making a z-test, use SE corresponding to obs/expected.

- \( SE_{\text{sum}} \)
- \( SE_{\text{ave}} \)
- \( SE_{\%} \)