Hypothesis tests with more than 2 classes - $\chi^2$ test

Testing for independence - $\chi^2$ test

Comments on tests of significance
Ho - based on Ho, could the data we've observed be due to chance?

Rolling a die.

Is the die fair?

Does the distribution of the # of times each face comes up correspond to that predicted by Ho?

---

Roll 60 times

Is the #6's compatible with a fair die?

\[ Z = \frac{\text{observed} - \text{expected}}{\text{SE}} = \frac{\text{obs} - 10}{\sqrt{60 \cdot (1 - 0) \cdot \frac{1}{6} \cdot \frac{5}{6}}} \]

\[ = \frac{\text{obs} - 10}{2.9} \]

exped #6's to be between 4 and 14.
How do we use data about the number of times each of the six faces appeared?

Need a statistic that measures how far away the observed data is from the expected values.

<table>
<thead>
<tr>
<th>Value</th>
<th>Observed Frequency</th>
<th>Expected Frequency based on H₀</th>
<th>60 rolls</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

\[ \chi^2 = \text{chi-squared statistic} = \sum \left( \frac{\text{observed frequency} - \text{expected frequency}}{\text{expected frequency}} \right)^2 \]
\[
\frac{(4-10)^2}{10} + \frac{(6-10)^2}{10} + \frac{(17-10)^2}{10} + \frac{(16-10)^2}{10} \\
+ \frac{(8-10)^2}{10} + \frac{(9-10)^2}{10} = \frac{142}{10} = 14.2.
\]

\(\chi^2\) statistic is 14.2.

Under Ho (that the die is fair), what's the chance of getting a \(\chi^2\) value of 14.2 or greater?

- Look it up on a \(\chi^2\) curve.

How to determine the degree of freedom (DoF)

For a fully specified model

\[\# \text{DoF} = \# \text{terms in the sum} - 1\]

To compute \(\chi^2\)

\[\# \text{d.of.} = 6 - 1 = 5\]

P-value for \(\chi^2 = 14.2\) with 5d.o.f. \(\approx 1\%\).

Reject Ho at 5% level.
Chi-squared; 5 degrees-of-freedom

From table, p-value is slightly larger than 1%.

Chance of $x^2 > 14$ is in this area.
Summary.

a) basic data - observations.
b) chance model.
c) frequency table - observed + expected frequencies.
d) $\chi^2$ - statistic.
e) #d.o.f. - #terms in sum - 1. for $\chi^2$
f) observed significance level - % area under $\chi^2$ curve with #dof to right of observed $\chi^2$ value.
Testing Independence:

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right handed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left handed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambidextrous</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>here</th>
<th>not here</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fresh</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Soph</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>Junior</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>Senior</td>
<td>14</td>
<td></td>
</tr>
</tbody>
</table>

This is a measurement of the entire population of AMC 5 students.

Population

Sample.
Sample size: 2237.

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT handed</td>
<td>934</td>
<td>1070</td>
<td>2004</td>
</tr>
<tr>
<td>Left handed</td>
<td>113</td>
<td>92</td>
<td>205</td>
</tr>
<tr>
<td>Amb.</td>
<td>20</td>
<td>8</td>
<td>28</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1067</td>
<td>1170</td>
<td>2237</td>
</tr>
</tbody>
</table>

**Ho:** RT/Left/Amb. is independent of sex.

**Ho:** % of RH amongst men is same as % of RH amongst women, and similarly for LH/Amb.

**H1:** % differ between men + women.

χ² statistic: need expected values under Ho.
<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>RH</td>
<td>2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LH</td>
<td>205</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Am</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1067</td>
<td>170</td>
<td>2235</td>
</tr>
</tbody>
</table>

Independence - For each race, the same fraction of the 2004 RH people, split up between men and women, in the same proportion as the total (2235) splits up.

The fraction of men overall is \( \frac{1067}{2235} \)

We would expect this fraction of the 2004 RH people to be men.

\[ \frac{1067}{2235} \times 2004 \text{ RH Men.} \]
In general: expected value in a cell of the table
\[
\frac{(\text{row total}) \times (\text{column total})}{\text{table total}}
\]

Table of expected values

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1M</td>
<td>956</td>
<td>1048</td>
</tr>
<tr>
<td></td>
<td>934</td>
<td>1070</td>
</tr>
<tr>
<td>L1L</td>
<td>98</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>113</td>
<td>92</td>
</tr>
<tr>
<td>Amb.</td>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>4</td>
</tr>
</tbody>
</table>

\[\chi^2\text{-statistic} = \sum \left( \frac{\text{observed freq.} - \text{expected freq.}}{\text{expected freq.}} \right)^2\]
\[
\chi^2 = \frac{(956 - 934)^2}{956} + \frac{(1048 - 1070)^2}{1048} + \frac{(98 - 113)^2}{98} + \frac{(107 - 92)^2}{107} + \frac{(13 - 20)^2}{13} + \frac{(15 - 8)^2}{15}
\]

\[\chi^2 = 12.\]

**How many degrees of freedom?**

The box model is not fully specified
- don't know the true %'s in the box
- lose 1 dof for each parameter we estimate.

Table m rows and n columns there are \((m-1) \times (n-1)\) d.o.f.

\[(3-1) \times (2-1) = 2\ d.o.f.\]

\[\chi^2 = 12,\ 2\ d.o.f.\]

From table, p-value is < 1%.

Reject H0 at 1% level (highly statistically significant).
A few comments on significance tests.

1. 1% and 5% - "highly statistically significant" 
    "statistically significant"

   These are conventional but arbitrary.

   Don't obsess over the difference between 4.9% and 5.1%.

2. Recall what a P-value is.

   - The chance of observing data as extreme or more, assuming Ho is true.

   [Note: the probability that Ho is true.]

   - If do enough tests, eventually you will get a statistically significant result.

   Decide how you will analyse the data. 
   (Which tests, which hypotheses)
   before you collect the data.
One sided or two sided tests.

Is this a fair coin?

deviations in either direction
(too many heads or too many tails)
provide evidence against $H_0$

$z$-statistic = \frac{\text{observed} - \text{expected}}{SE}$

counting heads - too many $H_1$, $z > 0$. $T$

counting tails - too many $T$, $z < 0$. $T$

P-value is sum
of the two
areas.

61 H out of 100 throws.

$z = \frac{61 - 50}{\sqrt{100 \times \frac{1}{2} \times \frac{1}{2}}} = \frac{61 - 50}{5} = 2.2$. $z$

Both shaded areas at correspond to $z$-values
more extreme than 2.2.

Both give evidence against $H_0$. $T$

Two-tailed test.
How to decide if it's a 1-tailed or 2-tailed test that's needed?

- depends on $H_1$ - alternative hypothesis.

One-tailed - if $H_1$ specifies in which direction (large/smaller) the data should be, compared with $H_0$.

Two-tailed - if $H_1$ says "different" without specifying in which direction.

By the difference between "statistically significant" and "important":

$$Z = \frac{\text{observed} - \text{expected}}{SE}.$$

$$SE\% = \frac{SD_{box}}{\text{# draws}} \times 100\%$$

as # draws increases SE decreases

$Z$ increases.

$p$-value decreases.

For very large samples, even very small differences between observed & expected can be claimed to be clinically meaningful.
Example.

Reading Scores. Rural 2.5
city 2.6

SD = 10
both samples size 2500

Two sample Z-test.

\[ Z = \frac{\text{observed diff} - \text{expected diff}}{\text{SE diff}} \]

H0: average score is the same for the 2 groups.

\[ Z = \frac{1 - 0}{0.28} = \frac{1}{0.28} = 3.5 \]

SEave = \( \frac{10}{\sqrt{2500}} = 0.2 \)

SEdiff = \( \sqrt{0.2^2 + 0.2^2} = 0.28 \)

P-value < 1%

But: what does a difference in 1 point on the reading test actually mean?

- partial understanding of 1 word in 40
- not very important.
Could something else give the effect we've observed?

or - does the box model for the accurately represent reality?

\[
\begin{array}{cccccccc}
\rule{0.5cm}{0.15cm} & \rule{0.5cm}{0.15cm} & \rule{0.5cm}{0.15cm} & \rule{0.5cm}{0.15cm} & \rule{0.5cm}{0.15cm} & \rule{0.5cm}{0.15cm} & \rule{0.5cm}{0.15cm} & \rule{0.5cm}{0.15cm} \\
\end{array}
\]

ESP - try to make it come up 6.

720 rolls.
143 6's.

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

\[
= 2.3
\]

\[
\text{SE} = \sqrt{720 \times \frac{1}{6} \times \frac{5}{6}} = 10
\]

\[
\leq 5\% \text{ (actually 1\%).}
\]

Reject H0 that die is fair.

Is ESP real?

Remember: rejected H0 - on the basis of the box model for a fair die, we would not expect this many 6's.

But: there are many possible alternative hypotheses and the significance test doesn't tell you which one is true.

(eq: die is biased).