Based on the influencers in the two samples, can we conclude that the mean heights of the two populations are different?

Is the difference in the sample means consistent with the population means being the same, and chance error in the sample means due to sampling variability?

H₀: mean height of population 1 = mean height of population 2.

H₁: mean heights of the 2 populations differ.
A test statistic measures the size of the observed difference in the sample means in terms of the size of the difference we would expect (if \( H_0 \) were true, and the variability is just sampling variability).

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

\[ \text{SE difference} = \sqrt{\left( \text{SE}_1 \right)^2 + \left( \text{SE}_2 \right)^2} \]

\[ \text{SE}_1 \]

The p-value is the probability of observing a test statistic as large as the one we have, or larger, if \( H_0 \) is true. (\( \approx \) shaded area.)
If $p < 0.05$ we reject $H_0$, and say that the result is statistically significant.

If $p > 0.05$ we say that we do not have evidence to reject $H_0$. 
What happens if there are more than 2 classes?

We have a hypothesis that specifies how frequently each of the classes should appear in the population.

How do we test the data in a sample to see if it is/is not compatible with the hypothesis?

\[
\begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
\end{array}
\]

Is a die fair?

Does the distribution of the number of times each face comes up correspond to that predicted by the null hypothesis?

[Can't test each number separately as they are not independent]
<table>
<thead>
<tr>
<th>value</th>
<th>observed frequency</th>
<th>expected frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>60</strong></td>
<td></td>
</tr>
</tbody>
</table>

Need a measure of how far the observed frequencies are from the expected frequencies.

\[
X^2 = \text{sum of } \frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}}
\]

(intuitively this has the right sort of properties)
\[ \chi^2 = \frac{(4-10)^2}{10} + \frac{(6-10)^2}{10} + \frac{(17-10)^2}{10} + \frac{(16-10)^2}{10} \]

\[ + \frac{(8-10)^2}{10} + \frac{(9-10)^2}{10} \]

\[ = \frac{142}{10} = 14.2. \]

Now, under Ho (fair die), what is the chance of obtaining data that give a \( \chi^2 \) value \( \geq 14.2 \)?

Is it possible to get a value like this just from sampling variability?

The probability of getting different values for the \( \chi^2 \) statistic depends on how many categories there are.

\( \chi^2 \) curves are parameterized by the number of degrees of freedom.
Chi-squared, 5 degrees-of-freedom
How to determine the number of DoF?

Ho had specified the frequencies of each of the 6 possible outcomes.

However, knowing the frequencies of 5 of the outcomes + total # rolls, fixes the frequency of the 6th.

In a fully specified model

\[
\text{DoF} = \# \text{ terms in the sum for } \chi^2 - 1
\]

In this case \# DoF = 6-1 = 5

The p-value (chance that, if Ho is true, the \( \chi^2 \) value computed from the data is this or larger than the one we have) is given by the area to the right of \( \chi^2 = 14.2 \) on the graph of \( \chi^2 \) with 5 DoF.
Look these p-values up in a table...

From the table, the p-value is between 5% and 10% (i.e., close to 1%).

So: chance of getting $X^2 = 14.2$ or larger for 60 rolls of a fair die is $< 1%$.

 Reject the null hypothesis that the data came from a fair die.

Note:
- Model is fully specified.
- Expected frequency in each row of the table is large ($\geq 5$).
<table>
<thead>
<tr>
<th>Day of week</th>
<th>observed</th>
<th>expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>16852</td>
<td>16916</td>
</tr>
<tr>
<td>T</td>
<td>16553</td>
<td>16916</td>
</tr>
<tr>
<td>3</td>
<td>16490</td>
<td>16916</td>
</tr>
<tr>
<td>Th</td>
<td>17299</td>
<td>16916</td>
</tr>
<tr>
<td>Fr</td>
<td>16348</td>
<td>16916</td>
</tr>
<tr>
<td>Sa</td>
<td>17019</td>
<td>16916</td>
</tr>
<tr>
<td>Sun</td>
<td>17753</td>
<td>16916</td>
</tr>
</tbody>
</table>

Total 118,415

\[
\chi^2 = \sum \frac{( \text{observed freq.} - \text{expected freq.})^2}{\text{expected freq.}} = 94.0
\]

\[6 \text{ Dof.}\]

p-value \ll 0.001

Reject Ho that earthquakes have no preference for particular days of the week.
Does anyone believe this?

Total # earthquake is large

# large earthquakes is small

\[ \text{by sampling variability, there will be some days with more of them, and some with fewer.} \]

After a large earthquake, lots of aftershocks.

\[ \text{large earthquakes} \]
\[ \text{aftershocks} \]
\[ \text{lots of small earthquakes} \]

\[ \text{non-uniform distribution.} \]
What's the point here?

Rejecting $H_0$ is the start of a process of thinking a little more carefully about what the process actually is that generated the data.
Testing Independence

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>F</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>amb.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Are handedness and sex independent?

Is having facial hair independent of the department of the hospital that you work in?

Ho: independence.

What are the expected values predicted by Ho?
Ho: handedness and sex are independent.

% of RH amongst men equals % RH amongst women etc.

H_1: distribution of handedness differs between men and women.

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>26</td>
<td>42</td>
</tr>
<tr>
<td>L</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>F</td>
</tr>
<tr>
<td>-----------</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>R+ handed</td>
<td>26</td>
<td>42</td>
</tr>
<tr>
<td>not R+ handed</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>52</td>
</tr>
</tbody>
</table>

**Independence**: The fraction of men and women on each row should be the same.

\[ \frac{68}{83} \] is the fraction of all people who are R\+ handed.

Since we have 31 men, we would expect \[ 31 \times \frac{68}{83} \] R\+ handed men.

We have 52 women, we would expect \[ 52 \times \frac{68}{83} \] R\+ handed women.
In general:

\[
\text{expected value} = \frac{\text{row total} \times \text{column total}}{\text{table total}}
\]

\[
\begin{array}{cc|ccc}
 & M & F \\
\hline
RT & 68 \times 31 & 68 \times 52 & 25.4 & 42.6 \\
\hline
\text{not RT} & 15 \times 31 & 15 \times 52 & 5.6 & 9.4 \\
\end{array}
\]

\[\chi^2 = \sum \frac{(\text{observed freq} - \text{expected freq})^2}{\text{expected freq}}\]

\[= \frac{(26 - 25.4)^2}{25.4} + \frac{(42 - 42.6)^2}{42.6} + \frac{(5 - 5.6)^2}{5.6} + \frac{(10 - 9.4)^2}{9.4}\]

\[= 0.125\]

How many degrees of freedom?
Don't know the true \% of each ticket - estimating them from the data.

- lose 1 Dof for each parameter we estimate.

Table with m rows and n columns has $(m-1)(n-1)$ dof.

In this example $m = 2 \Rightarrow (2-1)(2-1) = 1 \text{ dof}$.

From the table, the p-value is a little bit less than 90\%.

Do not reject the null hypothesis that sex and handedness are independent.