In order to make generalizations about an experimental sample, we need to figure out how big a part chance might have played in the outcome of the experiment.

We begin by assuming that our experimental results reflect only the random variations caused by assorted factors beyond our control. This assumption is called the null hypothesis. If our experiment is successful and our theory is true, we will be able to reject the null hypothesis by showing that chance variation is not a reasonable explanation for our results.

The chi-square ($\chi^2$) test allows you to compare a sample distribution (‘observed’ data) with a distribution derived from a null hypothesis (‘expected’ data), and decide whether your sample could reasonably be a random sample from the population.

**EXERCISE.** You have a theory that men are more likely than women to delete the /r/ in *market*. You test this theory by asking 25 men and 25 women where Market Avenue is. Five women and 10 men say, “You mean /ma:kit/ Street?” Everyone else says, “You mean /markit/ Street?”

1. Show the distribution of your observed data by filling in cells A1-B2 of Table (a) below.

2. What is your null hypothesis?

3. Do you think you’ll be able to reject the null hypothesis? (In other words, are these results significant?) Why or why not?

4. Suppose you ask 50 men and 50 women the same question. 38 of the men and 36 of the women delete the /r/ in *market*; everyone else pronounces the /r/. Do you think you’ll be able to reject the null hypothesis? (In other words, are these results significant?) Why or why not?

5. Fill in cells A3-C3 and C1-C2 of Table (a) (your marginals) and then calculate the percentages in D1-D3.

6. Copy your marginals and percentages to Table (b). This is where you’ll calculate your expected values. Imagine that there’s no relation between sex and /r/ pronunciation (your null hypothesis). You should be able to calculate, based on your totals and the percentages in row D, how many women and men are predicted to pronounce or delete /r/.

<table>
<thead>
<tr>
<th>Table (a): Observed</th>
<th>Table (b): Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1. /r/</td>
</tr>
<tr>
<td>A.</td>
<td></td>
</tr>
<tr>
<td>B.</td>
<td></td>
</tr>
<tr>
<td>C. Total</td>
<td></td>
</tr>
<tr>
<td>D. Percent</td>
<td></td>
</tr>
</tbody>
</table>

7. Once you have both tables filled in, you can calculate chi-square. To do this, just calculate $(observed-expected)^2/expected$ for each cell, and sum the total:

$$(A1)\_ \_ \_ + (A2)\_ \_ \_ + (B1)\_ \_ \_ + (B2)\_ \_ \_ = (\chi^2)\_ \_ \_$$
8. To figure out the probability of this chi-square statistic, we first need to establish our **degrees of freedom**. Degrees of freedom depend on the number of **categories** in each direction. In this case we have two columns (/r/ and no /r/) and two rows (women and men) – in other words, we have ___ categories in each direction. **Degrees of freedom (df) = (c-1)(r-1)**, where c is the number of categories in columns and r is the number categories in rows. We have ____ degree(s) of freedom.

9. Use the chi-square distribution table ([http://www.statsoftinc.com/textbook/sttable.html#chi](http://www.statsoftinc.com/textbook/sttable.html#chi)) to find the probability of your chi-square. The probability of your chi-square is between _____ and _____.

10. Are these results significant at the p < .05 level? ____  At the p < .01 level? _____

11. The higher the chi-square, the higher/lower the probability and the more/less significant the result.

12. Any time you have a 4-cell table like this one, you will have ___ degree(s) of freedom, and the minimum chi-square value yielding a probability lower than .05 will be _____.

---

**Sources:**