The χ^2 (Chi-Square) Tests for Categorical Variables

	One-way Table	*Two-Way Table*	
The Tests ➲	**Goodness of Fit**	**Homogeneity**	**Independence**
Use it when you have...	**one sample** from **one population**, and **one categorical** variable with r categories	c **samples** from c **populations**, and **one categorical** variable with r categories	**one sample** from **one population**, and **two categorical** variables with r and c categories
To answer the question...	Is the specified distribution of the variable correct?	Is the distribution of the variable the same for all the populations?	Is there an association between the two variables?
In your calculator...	Enter the **observed counts** in one **list**, and the **expected counts** in another **list**.	Enter the **observed counts** in an $r \times c$ **matrix.**	
To calculate expected counts... *(Do NOT round to a whole number!)*	Expected count = expected proportion $\times$sample size	Expected count = row total $\times$column total $\div$ total sample size (Note: your calculator will create an expected matrix when you run the test)	
degrees of freedom	$df = \#categories - 1$	$df = (\#rows - 1) \cdot (\#columns - 1)$	
(Note: Do NOT include the total row or total column when counting "#rows" and "#columns" – just count the rows and columns of observed data)			

<table>
<tr><th colspan="4">Summary of the χ^2 (Chi-Square) Tests</th></tr>
<tr><th></th><th>Goodness of Fit</th><th>Homogeneity</th><th>Independence</th></tr>
<tr><td>State</td><td>H_0: The stated distribution of <describe the variable> is as specified.
H_a: The stated distribution of <describe the variable> is not as specified.

$\alpha = \cdots$ (select a significance level)</td><td>H_0: There is no difference in the true distributions of <describe the variable> for <describe the populations>.
H_a: There is a difference in the true distributions of <describe the variable> for <describe the populations>.

$\alpha = \cdots$ (select a significance level)</td><td>H_0: There is no association between <describe the two variables> for <describe the population>.
H_a: There is an association between <describe the two variables> for <describe the population>.

$\alpha = \cdots$ (select a significance level)</td></tr>
<tr><td>Plan</td><td>• Random: the data come from a well-designed random sample or randomized experiment
• 10%: When sampling without replacement, check that $n \le \frac{1}{10}N$
• Large Counts: All expected counts are at least 5</td><td>• Random: the data come independent random samples or from the groups in a randomized experiment
• 10%: When sampling without replacement, check that $n \le \frac{1}{10}N$
• Large Counts: All expected counts are at least 5</td><td>• Random: the data come from a well-designed random sample or randomized experiment
• 10%: When sampling without replacement, check that $n \le \frac{1}{10}N$
• Large Counts: All expected counts are at least 5</td></tr>
<tr><td rowspan="2">Do</td><td>Expected count = expected proportion $\times$sample size

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$
If H_0 is true, the χ^2 statistic has a χ^2 distribution with $df = \#categories - 1$.</td><td colspan="2">Expected count = row total $\times$column total $\div$ table total

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$
If H_0 is true, the χ^2 statistic has a χ^2 distribution with $df = (\#rows - 1)\cdot(\#columns - 1)$</td></tr>
<tr><td colspan="3">P-Value is area to the right of χ^2,that is,
$P - value = P(\chi^2 > \text{the calculated test statistic})$</td></tr>
<tr><td>Conclude</td><td colspan="3">If $P - value < \alpha$, Reject H_0, else Fail to reject H_0. <State in context.>
If significant, consider a follow-up analysis. Identify the largest contributor(s) to the test statistic and indicate if the observed values were fewer or more than expected.</td></tr>
</table>