Goodness of Fit Tests

Last Time: Can generalize the binomial distribution to examine the distribution of outcomes across $k$ categories (“multinomial distribution”)

- The null hypothesis specifies a probability for each category and the probabilities must sum to one. The alternative hypothesis just says the null is not true.
- Need a statistic to measure the overall discrepancy between our observed counts and our expected counts under the null hypothesis.
  - General idea: compare the observed counts (or proportions) to the expected counts (or proportions)
    \[ \text{MAD} = \frac{\sum |O_i - E_i|}{k}, \frac{\sum (O_i - E_i)^2}{E_i} = \frac{\sum (\hat{p}_i - \pi_{i0})^2}{\pi_{i0} / n^2} \]
  - Ideally: standardized
    \[ \chi^2 = \frac{\sum (O_i - E_i)^2}{E_i} = \frac{\sum ((\hat{p}_i - \pi_{i0}) / \sqrt{\pi_{i0} / n})^2}{\pi_{i0} / n} \]
- To get a p-value, need to simulate random samples assuming the null hypothesis is true, see how often we get a statistic at least as extreme (larger) than the observed

Example: Recall my claim about the color distribution of M&M candies and your sample results.

<table>
<thead>
<tr>
<th>Blue</th>
<th>Brown</th>
<th>Green</th>
<th>Orange</th>
<th>Red</th>
<th>Yellow</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
</tr>
</tbody>
</table>

(a) Open the Goodness of Fit applet from Lecture Notes page and enter your observed counts and then press Use Table.

(b) Check the Show Table box (expand?).

(c) Use the Statistic pull-down menu to select Max-Min. Report and verify the value of the Max – Min statistic for your sample data. Does it make sense to use this statistic here?

(d) Does the bar graph provide preliminary evidence against the null hypothesis?

(e) Check the Show Sampling Options box. Enter the hypothesized probabilities (to 4 decimal places), separated by commas. Press Sample. Select the Plot radio button. Are the simulated proportions the same as the expected probabilities (denoted by red lines)? Are they the same as the observed proportions? Which are they closer to?

(f) Press Sample a few more times to see the sample to sample variation in the results. Then specify the Number of Samples to end up with 1000 total. Describe the distribution of the max – min statistic. Is it symmetric? Is this what you would expect? Is it centered at zero?

(g) Use the distribution in (f) to approximate a p-value (explain what you are finding).
(h) Now calculate (by hand) the MAD statistic (mean absolute difference between observed and expected counts), dividing by the number of categories.

(i) Examine the null distribution for the MAD statistic (use the Statistic pull-down menu on the left). Describe the behavior of the MAD’s null distribution. What does the MAD approximate for the p-value?

(j) Repeat for the chi-square ($\chi^2$) statistic (sum of observed minus expected counts squared divided by expected counts). How does this statistic behave? What does this statistic approximate for the p-value? Which distribution was the most “well-behaved” (i.e. might follow a mathematical model)?

With a multinomial process when the sample size is large (the expected cell counts are all at least 5), the distribution of the chi-square statistic is well modelled by a **chi-square distribution**. This distribution is skewed to the right and is parameterized by “degrees of freedom.” For the goodness-of-fit test, the degrees of freedom is $k$ (the number of categories) minus one ($k - 1$), which controls the amount of skewness in the distribution and mean = $df$.

(k) Check the Overlay Chi-square distribution box. How well does the blue curve approximate your simulated null distribution? How well does the p-value from the chi-square distribution agree with the p-value from the simulation? Are the conditions for the chi-square approximation met here?

(l) Mars Inc claims that for Crispy M&Ms the colors are equally distributed. How many students in class today found a small p-value? Is this what you might expect? Explain. What does this tell you about the power of the chi-square test?

(m) The Mars Corporation also claims the color probabilities for plain M&Ms to be:

<table>
<thead>
<tr>
<th></th>
<th>Blue</th>
<th>Brown</th>
<th>Green</th>
<th>Orange</th>
<th>Red</th>
<th>Yellow</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.24</td>
<td>0.13</td>
<td>0.16</td>
<td>0.20</td>
<td>0.13</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Carry out a test of significance to determine whether your sample provides convincing evidence that your sample did not come from a process with these probabilities. (State the hypotheses, making sure the parameter definitions are clear, report the chi-square statistic, and approximate a p-value, explaining how you have done so. State your final conclusion in context.)

$H_0: \pi_1 = \pi_2 = \ldots = \pi_6$

$H_a$: at least one $\pi$ does not match

$\chi^2$ test statistic, $df = 5$