Fact Sheet – Goodness-of-Fit (12.1)

A goodness-of-fit test is an inferential procedure used to determine whether a frequency distribution follows a specific (probability) distribution.

Example: A 6-sided die is fair, i.e. $p_1 = p_2 = \ldots = p_6 = \frac{1}{6}$.

Example: The distribution of plain M&M candies in a bag is 13% brown, 14% yellow, 13% red, 20% orange, 24% blue, and 16% green.

Conditions

To perform a Goodness-of-Fit test for *k* categories, the following two conditions must be met.

- Each expected frequency must be at least 1: $E_i \geq 1$ for all *i* = 1 to *k*.
- No more than 20% of all expected frequencies are less than 5.

You can check the conditions by computing the expected frequencies once you have set up the null hypothesis H_0. The expected frequency for each category can be computed by multiplying the sample size (*n*) by the claimed proportion (p_i) for that category: $E_i = n \cdot p_i$

Hypothesis Test

Step 1

The null hypothesis will be of the form $p_1 = \#,\ p_2 = \#,\ \ldots, p_k = \#$.

H_1 will always be "At least one proportion is different than claimed."

Step 3

The test statistic is $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$, where O_i is the observed sample frequency for category *i* and E_i is the computed expected frequency for category *i*.

Just write "Goodness-of-Fit", rather than writing the test statistic.

Step 4

To compute the test statistic and P-value using StatCrunch …

1. Enter the observed counts in the first column.
Enter the expected counts in the second column.
Name the columns observed and expected.
2. Select **Stat**, highlight **Goodness-of-fit**, then highlight **Chi-Square test**.
3. Select the column that contains the observed counts and select the column that contains the expected counts. Click Calculate.