**Goodness-of-Fit Test for a Single Categorical Variable**

Goodness-of-Fit tests test whether or not a categorical variable fits a particular multinomial probability distribution. A multinomial probability distribution gives the probabilities for three or more categorical outcomes of a given variable. The multinomial probability distribution for a variable with categorical outcomes is:

where

1. **Formulating the Hypotheses**: There is only one set of hypotheses for goodness-of-fit tests, and it incorporates a multinomial probability distribution. You **must fill in the probabilities** for each problem.

|  |
| --- |
| **Hypotheses for Goodness-of-Fit Tests** |
|  |
| **Answers questions about:** |
| If the population distribution of a given categorical variable fits  or not. |
| NOTE: |

1. **The Test Statistic:**

Here is the idea behind Goodness of Fit tests: if the null hypothesis is true, then the multinomial probability distribution in truly describes the population. This means that the observed frequencies from the sample should be the same as (or very close to) the expected frequencies calculated from the probability distribution. (For example, if 25% of the population fits into Category A, then about 25% of the sample should fit into Category A.) Large differences between the observed frequencies and the expected frequencies are evidence that the underlying population is NOT distributed according to the hypothetical multinomial probability distribution.

Some variation between observed and expected frequencies will occur due to random chance, so we need to use our hypothesis testing techniques to see if the differences we see are *statistically significant.* The sets a threshold at which we determine the sample distribution to be different enough from what we would expect under the null hypothesis, to contradict the null hypothesis.

The size of the difference between the observed and expected frequencies is quantified in our test statistic. When the test statistic is large enough to reject the null, we can conclude that the population is not distributed according to the hypothetical multinomial probability distribution.

The test statistic for a Goodness-of-Fit Test is:

1. **Deciding whether or not to Reject :**

Only large discrepancies between the observed and expected frequencies constitute evidence against the null hypothesis, so all Goodness-of-Fit tests are **upper tail tests.** (If the test statistic was in the lower tail, that would mean there were only small discrepancies between observed and expected).

|  |  |
| --- | --- |
| **When to Reject in a Goodness-of-Fit Test** | |
|  | **Always an Upper Tail Test:** |
| **p-value approach:** | Calculate the upper tail of  If the then reject and accept  If the , then do not reject . is unsupported. |
| **Critical Value: Approach** | Look up the UT Critical Value of , which is  If then reject and accept .  If , then do not reject is unsupported. |
| NOTES:   1. is a Test Statistic 2. is an Upper Tail Critical Value 3. is based on degrees of freedom. The degrees of freedom in Goodness-of-Fit Tests is | |

1. **Interpreting the test:**

(Note: This explanation of interpretation holds for ALL hypothesis tests.) We start every hypothesis test with a question about the parameter of interest, so we must end every hypothesis test with the answer to that question. In other words, we must *interpret* the conclusion of our test in terms of the original question.

Remember: in hypothesis testing you can never prove the null hypothesis. You can only prove the alternative hypothesis: when you reject the null and accept the alternative, then at your given level of significance you may conclude that is true. If you do not reject then you must conclude that is unsupported by the evidence. This gives us a clear guideline for how to *interpret* hypothesis tests: ***always look to the alternative hypothesis!***

In all that follows, you must substitute the actual words and numbers from your hypothesis test for the symbols. Notice that each interpretation simply states the alternative hypothesis in words, and says either that we can or cannot conclude that it is true.

|  |  |
| --- | --- |
| **How to Interpret a Goodness-of-Fit Test:** | |
| **When you:** | **Interpretation:** |
| **Reject** | At the significance level, we can conclude that the probability distribution of [*the variable*] is not . |
| **Do not reject** | At the significance level, we cannot conclude that the probability distribution of [*the variable*] is different from . |
| NOTES:   1. When is rejected, you can then look at the distribution of outcomes in the sample for information about the true population distribution. 2. Comparing the expected frequencies to the observed frequencies is often helpful to learn about how the actual distribution differs from the distribution. | |

**Assumptions Underlying These Hypothesis Tests**

All hypothesis tests use sampling distributions to determine the probability of sample statistics. In order for us to be confident that our choice of sampling distribution for any given test really is the way the sample statistic is distributed, certain assumptions must be met. If the assumptions are not met – that is, if any given assumption is not true – then we cannot rely on the results of the hypothesis tests. They may mislead us, give us the wrong answers, and cause us to draw the wrong conclusions.

For Goodness-of-Fit Tests, the only assumption that must be satisfied is that the expected frequency for each category must be greater than five:

*Exercise:* A bike shop sells bikes with frames made from four different materials: carbon fiber, aluminum, steel, and titanium. In the past, the share of sales by type of frame was 17% carbon fiber, 53% aluminum, 12% steel, and 18% titanium. The bike shop decided to use a Goodness-of-Fit test at the  **significance level** to see whether the probability of sales by frame type has changed from what it was in the past. In a random sample of 250 bike sales, the following frequencies were observed:

|  |  |  |
| --- | --- | --- |
| **Category** | **Frame Type** | **Number of bikes sold** |
| 1 | Carbon Fiber | 66 |
| 2 | Aluminum | 128 |
| 3 | Steel | 16 |
| 4 | Titanium | 40 |

Start by formulating the null and alternative hypotheses:

From the null hypothesis, fill in the Hypothesized Probability column.

From the sample data, fill in the Observed Frequency column.

For each row, the Expected Frequency is

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Frame material** | **Category** | **Hypothesized**  **Probability** | **Observed Frequency** | **Expected Frequency** | | **Difference Squared/**  **Expected Frequency** |
|  |  |  |  |  | |  |
| Carbon Fiber | 1 |  |  |  | |  |
| Aluminum | 2 |  |  |  | |  |
| Steel | 3 |  |  |  | |  |
| Titanium | 4 |  |  |  | |  |
|  |  |  | | |  | | |

The sum of the values in the last column is the test statistic. The test statistic and the critical value have *k* – 1 degrees of freedom, where *k* is the number of categories.

Now you are ready to complete the hypothesis test.

(left blank for hypothesis test)

Because we have determined that the distribution has changed from what it was in the past, the bike shop now wants more details. Which categories have seen the biggest changes? What adjustments should be made to inventory?