Multinomial Probability Distributions

Goodness of Fit tests are hypothesis tests that determine whether or not a categorical variable has a particular probability distribution across its categories.

Goodness-of-Fit tests employ the concept of a *multinomial probability distribution*, which is a probability distribution that gives the probabilities of three or more categorical outcomes. A multinomial probability distribution for *k* categorical outcomes is simply a list of probabilities, one per category of the variable, like this:

* Category probabilities () must always be between 0 and 1; that is, . Therefore, if probabilities are expressed as percentages in a problem, those percentages must be divided by 100 to get probabilities.
* In a properly formulated multinomial probability distribution, the sum of all of the category probabilities is 1.

*Example 1:*

Suppose you were studying a population of cats at an animal shelter. The variable you are interested in is color. The cats at the shelter come in three colors: tabby, black, and calico. 25% of the cats are tabby, 60% of the cats are black, and 15% of the cats are calico. Construct the multinomial probability distribution of color for this population.

|  |  |  |
| --- | --- | --- |
|  | Category | Probability |
| 1 | Tabby |  |
| 2 | Black |  |
| 3 | Calico |  |

A multinomial probability distribution can be used to calculate *expected frequencies*. *Expected frequencies* are the number of outcomes per category that we expect to find in a sample drawn from a population with the given distribution. To calculate the expected frequencies for a sample of size , use the following equation:

where

NOTE: Expected frequencies often come out with decimals. You should NOT round them to whole numbers. Keep them at four decimals unless they come out even to fewer decimal places.

*Example 2:*

Consider the multinomial probability distribution from Example 1. What are the expected frequencies for a random sample of size 50 drawn from this population? In other words, how many cats in the sample are expected to be tabby, how many black, and how many calico, theoretically?

|  |  |  |  |
| --- | --- | --- | --- |
|  | Categoryi | Probability | Expected  Frequency |
| 1 | Tabby |  |  |
| 2 | Black |  |  |
| 3 | Calico |  |  |

*Example 3:*

a) A given variable has eight categories. What is the multinomial probability distribution for this variable if the probability is the same for all eight categories?

b) What is the multinomial probability distribution if the variable has five categories and the probability is the same for all five categories?