Goodness-of-Fit Test

A **goodness-of-fit test** is used to test the hypothesis that an observed frequency distribution fits (or conforms to) some claimed distribution.

Notation for the goodness-of-fit test (Please note that the notation varies depending on the text)

O represents the observed frequency of an outcome

E represents the expected frequency of an outcome

k represents the number of different categories or outcomes

n represents the total number of trials

To use the goodness-of-fit test there are several requirements that must be met.

- The data have been randomly selected
- The sample data consist of frequency counts for each of the different categories
- The sample data come from a multinomial experiment
- For each category the expected frequency is at least five.

Test Statistic

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$df = k-1$$

The goodness-of-fit test is a hypothesis test. There are seven steps for a hypothesis test.

1. State the null hypothesis
2. State the alternative hypothesis
3. State the level of significance
4. State the test statistic
5. Calculate
6. Statistical Conclusion
7. Experimental Conclusion

Example

Suppose you are interested in finding out if there is a difference among diet soda drinkers in Phoenix, AZ. A study was conducted in which 200 randomly selected diet soda drinkers in Phoenix were ask to taste three leading brands of diet soda and then asked which they prefer. The results are as follows.

Brand	A	B	C
	60	75	65

Assume the experiment was preformed correctly. Determine if there is a different preference for any one brand of diet soda at a level of significance of 0.05.

Step 1: Null Hypothesis

$$H_0: \text{p}_A = p_B = p_c$$

Step 2: Alternative Hypothesis

$$H_A: \text{at least one probability is different}$$

Step 3: Level of Significance

$$\alpha = 0.05$$

Step 4: Test Statistic

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$
$$df = k-1$$

Step 5: Calculations

We need to find the expected frequency for each possible outcome. As the null hypothesis states that the probability of each diet soda is equal, then the number of people who prefer each diet soda should be equal. There are three diet sodas and 200 people.

$$E = \frac{200}{3} = 66.67$$

Plug this into the test statistic to get the observed value for the test statistic.

$$\chi^2_{obs} = \sum \frac{(O-E)^2}{E} = \frac{(60-66.67)^2}{66.67} + \frac{(75-66.67)^2}{66.67} + \frac{(65-66.67)^2}{66.67}$$

$$\chi^2_{obs} = .667 + 1.041 + .042$$

$$\chi^2_{obs} = 1.75$$

$$df = k - 1 = 3 - 1 = 2$$

Find the critical value for the Chi-Square distribution at a level of significance of 0.05 and degrees of freedom equal to 2. To do this, look at the table of critical values for the chi-square distribution. Look at the row for degrees of freedom equal to 2 and alpha equal to 0.05.

$$\chi^2_{crit} = 5.991$$

Step 6: Statistical Conclusion

To get your statistical conclusion, compare the critical value to the observed value of the test statistic. If the observed test statistic is greater than the level of significance then reject the null hypothesis.

Since the observed test statistic is, $\chi^2_{obs} = 1.75$ is less than the critical value $\chi^2_{crit} = 5.991$, then we will fail to reject the null hypothesis at a level of significance of 0.05.

Step 7: Experimental Conclusion

There is not enough evidence to indicate that there is a significant difference between the preferences of diet soda at a level of significance of 0.05.