Chapter 11: Goodness-of-Fit and Contingency Tables

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

The M&M company claims that the distribution of plain M&M candies in a bag is 23% blue, 23% orange, 15% green, 12% red, 15% yellow, and 12% brown. Even though this is their claim, do you think this represents the true proportions of color distribution in all of the M&M bags? ***How would we check?***

H_0: _____

H_1: _____

What to Compare and How to Compare It

Expected Counts (E_i)	Observed Counts (O_i)
The number in each category we would expect to see if H_0 is true * Two ways of calculating the expected counts: 1. If the expected counts are EQUAL, then $E = $ —— 2. If the expected counts are not equal, then calculate using $E_i = \mu_i = $ _____ where $i = 1, 2 ... k$	Observe how many in your sample are in each category. *This information will be given.*

*If the observation and experiment counts are "close", then _____ H_0

* If the observation and experiment counts "far apart" (BIG DIFFERENCE), then _____ H_0

Ex: Finding the expected counts.
 (a) A single die is rolled 45 times with the following results. Assuming that the die is fair and all outcomes are equally likely, find the expected frequency E for each empty cell.

Outcome	1	2	3	4	5	6
Observed O_i	13	6	12	9	3	2
Expected E_i						

 (b) Jon works as an usher at a theatre. The theatre has 1000 seats that are accessed through five entrances. Each guest should use the entrance that's marked on their ticket. Entrances A and B should each have 30% of the guests using these entrances. Entrance C should have 20% of the guests using its entrance. Entrances D and E should each have 10% of the guests using these entrances. Find the expected frequency for each E for each entrance.

Entrance	A	B	C	D	E
Observed O_i	398	202	205	87	108
Expected E_i					

Steps for Hypothesis Test for Goodness-of-Fit

What to Find…	*Check Requirements*
● Number of categories, k ● Expected Counts, E_i	● The data has to be randomly selected. ● The sample data consist of frequency counts for each of the different categories. ● For each category, the expected frequency is at least 5.

Step 1: Hypotheses H_0: $p_1 = p_2 = \ldots = p_k$ (all outcomes are equally likely) H_1: at least one of the probabilities is different from the others H_0: $p_1 = \#,\ p_2 = \# \ldots p_k = \#$ H_1: at least one of the probabilities is different from the claimed distribution	*Step 2: Level of Significance*

Step 3: Test Statistic

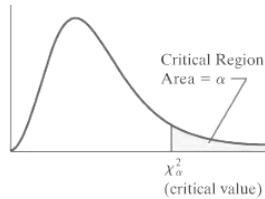$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

ALWAYS _____-TAILED TEST!

Step 4: Find a Critical Value or P-Value to check using either the Critical Value or P-value method.

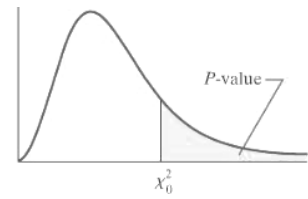*Step 5: Make a decision and draw a conclusion.*

CRITICAL REGION METHOD	P-VALUE METHOD
* Table A-4 $df = k - 1$ Critical Region Area = α χ^2_α (critical value) $\begin{cases} \textit{Reject } H_0 \ \sim \ \text{if } \chi^{2*} \text{ lies in the critical region} \\ \textit{Fail to Reject } H_0 \ \sim \ \text{if } \chi^{2*} \text{ doesn't lie in the critical region} \end{cases}$	$df = k - 1$ P-value χ^2_0 $\begin{cases} \textit{Reject } H_0 \ \sim \ \text{if } P\text{-value} \leq \alpha \\ \textit{Fail to Reject } H_0 \ \sim \ \text{if } P\text{-value} > \alpha \end{cases}$

GRAPHING CALCULATOR (TI-83 OR 84)

Instructions:

STAT $\Rightarrow$ TESTS $\Rightarrow$ D: $\chi^2 GOF$ -Test

Compare the observed O values to the corresponding expected E values.

Os and Es are close	Os and Es are far apart
Small χ^2 value, large P-value	Large χ^2 value, small P-value
χ^2 here	χ^2 here
Fail to reject H_0	Reject H_0
Good fit with assumed distribution	Not a good fit with assumed distribution

"If the P is low, the null must go."
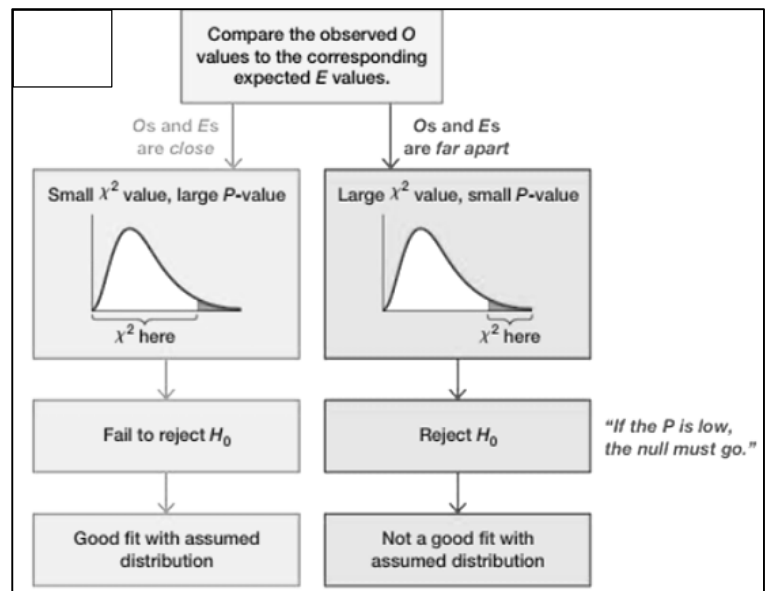
Ex: The M&M company claims that the distribution of plain M&M candies in a bag is 23% blue, 23% orange, 15% green, 12% red, 15% yellow, and 12% brown. Suppose we took a simple random sample of 400 M&Ms from the populations of all M&Ms. The results are shown below:

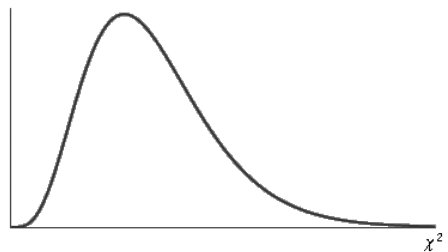| COLOR	Blue	Orange	Green	Red	Yellow	Brown
FREQUENCY	53	66	38	96	88	59
EXPECTED						

Find E_i =

Is the proportion of each color different than the claim of the M&M's manufacturer?

Null and Alternative Hypothesis

Test Statistic

P-value/Critical Region

Decision about Null Hypothesis

Conclusion

Ex: A company sells their products exclusively by mail. The company's management wants to find out if the number of orders received at the company's office on each of the five days of the week is the same. The company took a random sample of 400 orders received during a four-week period. The following table lists the frequency distribution for these orders by the day of the week.
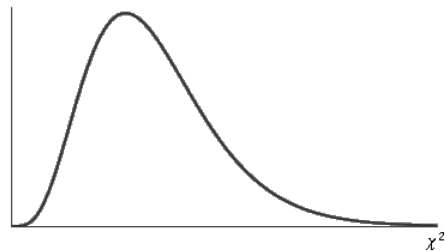
	Monday	Tuesday	Wednesday	Thursday	Friday
Number of Orders	92	71	65	83	89
Expected Number					

Test the claim that the orders are evenly distributed over the five days of the week. Use $\alpha = .025$

Null and Alternative Hypothesis

Test Statistic

P-value/Critical Region

x^2

Decision about Null Hypothesis

Conclusion