11.1 Chi Square Tests

One way tables: displays the distribution of a single categorical variable for individuals in a sample.

Ex:

color	Blue	Orange	Green	Yellow	Red	Brown	Total
count	48	50	52	42	35	40	267

Remember what m&m's said the proportion should be:

Blue: .24 orange: .2 Green = .16 Yellow = .14 Red = .13

Brown = .13

To do a significance test involving proportion, we would have to test:

$H_0: P_{Blue} = .24$ $H_0: P_{orange} = .2$

$H_a: P_{Blue} \neq .24$. - - - - -

distribution

Instead, we can take all colors into account using chi square test for goodness of fit.

In a chi square test, the null hypothesis is a claim about the distribution of a single variable and the alternative hypothesis is that it does not have that distribution.

Ex: H_0: The color distribution for m&m's is correct.
H_a: The color distribution for m&m's is not correct.

In a chi square test, we compare <u>observed counts</u> to <u>expected counts</u>.

The more observed counts vary from the expected counts, the more evidence we have against H_0.

Ex: Determine the expected counts for our sample of m&m's.

color	Blue	Orange	Green	Yellow	Red	Brown
expected count	(.24)(267) 64.08	53.4	42.72	37.38	34.71	34.71

Ex: How far off are we?

Blue: 16.08
Orange: 3.4
Green: 9.28
Yellow: 4.62
Red: .29
Brown: 5.29

The Chi Square statistic, χ^2, is a measure of how far off the observed values are from the expected counts.

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$ where the sum is over all possible values of a categorical variable.

Ex: Determine χ^2 for our m&m data

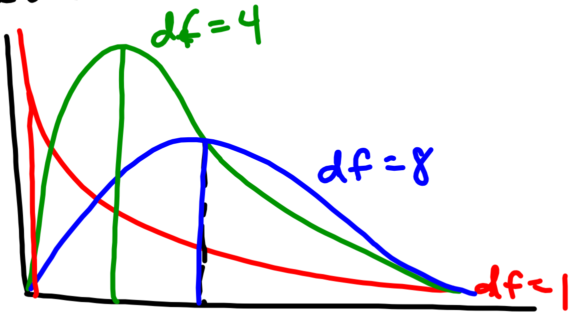
$$\chi^2 = \frac{(48-64.08)^2}{64.08} + \frac{(50-53.4)^2}{53.4} + \frac{(52-42.72)^2}{42.72} + \frac{(42-37.38)^2}{37.38} + \frac{(35-34.71)^2}{34.71}$$

$$+ \frac{(40-34.71)^2}{34.71}$$

$$\chi^2 = 7.65$$

Chi Square Distribution

Family of density curves that take only positive values and is skewed right. It has degrees of freedom df = # of categories − 1



Characteristics of the Chi Square distribution

1. mean is equal to the degrees of freedom
2. For df>2 the mode is at df-2
3. good estimator when all expected counts are at least 5.

Ex: Find the p value for our χ^2 from our M&M example.

Chi Square Test For Goodness of Fit

Conditions

① Random: Data comes from a random sample or randomized experiment.

② 10% condition: if no replacement in sampling, $n \leq \frac{1}{10}N$

③ Large counts: all expected counts are at least 5.

Suppose conditions are met. To determine whether a categorical variable has a specified distribution in a population of interest perform a test of:

H_0: stated distribution of a categorical variable in the population of interest is correct

H_a: stated distribution of a categorical variable in the population of interest is not correct.

Start by finding the expected counts for each category assuming H_0 is true. Then calculate:

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

where the sum is over all k different categories

The p value is the area to the right of χ^2 under the density curve of the chi square distribution with $df = k - 1$

Check your understanding p. 691

H_0: The distribution of eye color and wing type are as predicted

H_a: The distribution of eye color and wing type are not predicted

$\alpha = .01$

Conditions:

Large counts: all expected values are greater than 5
($\frac{9}{16} \cdot 200 = 112.5, 37.5, 37.5, 12.5$)

Random: Random sample

10% condition: 200 is less than $\frac{1}{10}$ of all fruit flies

We will perform the χ^2 test for goodness of fit.

On our calculator, we will do the χ^2 GOF Test using observed counts of 99, 49, 42, 10 and expected counts of 112.5, 37.5, 37.5, 12.5

$\chi^2 = 6.18$

p value = .1029

df = 3

We fail to reject the H_0 since $.1029 > .01$ meaning there is not enough to show the distribution of eye color and wing type is different than predicted.