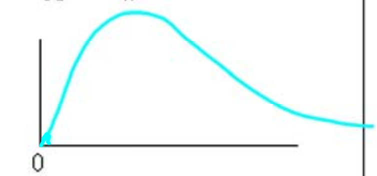
AP Stats Notes on Chi-Square Distributions (χ^2)

Chi-square (χ^2) distributions are a family of distributions that are distinguished from each other by their **degrees of freedom**.

Typical χ^2 Distribution:



- Each χ^2 distribution is **positive** and **skewed right**.
- Like other probability distributions, **the total area under the curve is equal to 1**.
- Each χ^2 distribution, regardless of degrees of freedom, **begins at 0, increases to a peak, and then approaches the x-axis asymptotically**. Skewed right
- As the number of degrees of freedom increases, χ^2 distributions become more **symmetric** and less **skewed right**.
- Two of the primary uses for χ^2 distributions are to test for **independence** and ~~**goodness of fit**~~.

χ^2-Test χ^2-GOF

Chi-square distributions can be used to test how well data fits with a **hypothesized distribution**, much like we do with 1-Proportion Z tests and 2-Proportion Z tests, but with **more than 2 proportions.** χ^2 distributions can also be used to test for **independence**. This is appropriate for multivariate data that is displayed in **two-way tables**. In either case, we use a process similar to what we used for hypothesis testing: we identify the hypothesized distribution:

- state our **null hypothesis** and **alternate hypothesis**
- determine the number of **degrees of freedom**
- check that the **technical conditions** are met for a test
- calculate a **test statistic** for the data that has been gathered
- calculate a **p-value** for the statistic, and
- reach a **conclusion** to either reject or fail to reject the null hypothesis.

Using χ^2 to test for Goodness of Fit

Suppose that the M&M/Mars Company claims that the distribution of colors in M&M candies is 30% brown, 20% each of yellow and red, and 10% each of orange, green, and blue. Suppose further that a class of

var → color of M&M's ⇒ categ.

Statistics students examines a sample of 100 M&M candies and gathers the following data:

sample (no decimal)

	Brown	Yellow	Red	Orange	Green	Blue	**Total**
Observed	38	25	15	12	8	2	100
Expected	30	20	20	10	10	10	100

same

.3(100) .2(100) → .10(100) →

can be decimals (do not round)

Complete the table by calculating the number of each color that would be expected if the distribution of candies matched the manufacturer's claims, and the total for that row.

Before we can conduct a test for goodness of fit, we need to state the **null and alternate hypotheses**:

H_0: The actual proportion of each color **matches** = the manufacturer's claims. (30% brown, 20% each of yellow and red, and 10% each of orange, green, and blue.)

H_a: The actual proportion of ~~each~~ at least one color **is different** ≠ **from** what the manufacturer claims.

The number of **degrees of freedom** is $\cancel{n}$ c − 1, or 6 − 1, which is 5. (cats.)

We must assume that these 100 M&M candies represent an SRS of all M&M candies. We must check that all of the expected values in the table are at least 1, which they are, and that no more than 20% are less than 5. Since they are all at least 5, we have verified the technical conditions.

exp ≥ 5

The test statistic is $\chi^2 = \Sigma \frac{(O-E)^2}{E}$. Filling in the appropriate values, we get

$$\chi^2 = \frac{(38-30)^2}{30} + \frac{(25-20)^2}{20} + \frac{(15-20)^2}{20} + \frac{(12-10)^2}{10} + \frac{(8-10)^2}{10} + \frac{(2-10)^2}{10}$$

$$= 2.13 + 1.25 + 1.25 + 0.4 + 0.4 + 6.4$$

$$\chi^2 = 11.83$$

Using the χ^2 table and the appropriate degrees of freedom (**5**), we determine that $0.025 < p < 0.05$. This represents enough evidence to **reject the** atleast one

null hypothesis at the 0.05 level, so the actual proportion of ~~each~~ color for M&M candies may be **different than the manufacturer claims**.