Key Concepts and Terminology

Chi-Square Analysis: Non-parametric analysis method used in analyzing categorical data with no continuous dependent variable. It is very sensitive to large sample sizes thus increasing power but also prone to Type I errors.

Counts: A measure used in counting within each category. Just a simple count of say 10 Orange M&Ms in a 1.74 Oz bag made up of 56 M&Ms.

Proportions: The ratio of counts relative the sample size (n). It can be expressed as a percentage.

Chi-Square Test: One tailed test reported as a two tailed test.

Goodness of Fit Test: Compare the observed distribution to the expected distribution (equal vs unequal proportions).

Test of Association: A test to determine association between two variables or categories (OR sample (my bag) vs population (class bag)). A comparison of observed responses to expected responses in a truly independent scenario for the variables involved (odds ratio == 1).

Chi-Square Statistic:

$$\chi^2 = \sum \frac{(\text{Observed Freq} - \text{Expected Freq})^2}{\mathit{Expected\ Frequency}}$$

$$\chi^2 = \sum_{i=0}^{\infty} \frac{(O - E)^2}{E}$$

Key Concepts and Terminology

Chi-Square Distribution: Positive or right skewed non-normal distribution with increasing symmetry as the degrees of freedom increase (DF++); unique because it begins at zero. The PDF is dependent on the Chi-square value and DF (constant).

$$\chi^2 = \sum_{i=I}^{r} Y_i^2 \text{ , mean}=0\text{, variance}=I\text{, } r = DF$$

$$P_r(x) = \frac{x^{\frac{r}{2}-I} e^{-\frac{x}{2}}}{\Gamma\left(\frac{1}{2}r\right) 2^{\frac{r}{2}}}$$

$$D_r(\chi^2) = \int_0^{\chi^2} \frac{t^{\frac{r}{2}-I} e^{-\frac{t}{2}} \, dt}{\Gamma\left(\frac{1}{2}r\right) 2^{\frac{r}{2}}}$$

$$P\left(\frac{1}{2}r, \frac{1}{2}\chi^2\right) = 1 - \int_0^{\chi^2} \frac{t^{\frac{r}{2}-I} \, e^{-\frac{t}{2}} \, dt}{\Gamma\left(\frac{1}{2}r\right) 2^{\frac{r}{2}}}$$

Goodness of Fit DF: # of Categories – 1

Test of Association DF: (Nrow-1) * (Ncol-1)

Phi-stat: A measure of the strength of association following a Chi-sq test of association analysis.

Key Concepts and Terminology

Fisher's exact p: Used as an alternative to the Chi-sq test of association if one or more of the cell counts in the contingency table is <5.

$$Fisher's\ exact\ p = \frac{((a+b)!\,(c+d)!\,(a+c)!\,(b+d)!)}{a!\,b!\,c!\,d!\,N!}$$

Yate's correction (1934): Subtraction of 0.5 from the absolute value difference of O-E frequencies. The outcome reduces the calculated Chi-sq stat.

Contingency table: A table used to represent a Chi-square test of association and also used when dealing with treatment-control setups like in clinical trials.

		Color		
		Orange	Not orange	Total
Bags	Mybag	2	54	56
	Class	440	1581	2021
	Total	442	1635	2077

Odds Ratio: a ratio of odds_var1_with_condition to odds_var2_with_condition. Unlike the risk ratio, calculations here are dependent on outcomes.

$$Odds\ Ratio = \frac{\frac{a=64}{b=166}}{\frac{c=178}{d=277}} = .5999$$

Risk ratio: a ratio of risk_var1_with_condition to risk_var2_with_condition. Calculations dependent on the sample size and risk factor.

Key Concepts and Terminology

OR confidence interval: If 1 falls outside the OR CI then the OR is significantly different from 1 (using alpha = 0.05).

$$OR\_CI = e^{In(OR) \pm 1.96 * \sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}}$$

Fisher's test in R: Use of a matrix is needed to accomplish this endeavor.

Vassar Stats: A tool used to perform contingency table analysis.