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-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).

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.

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.

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 | | | | |
| Bags |  | **Orange** | **Not orange** | **Total** |
| **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.

Risk ratio: a ratio of risk\_var1\_with\_condition to risk\_var2\_with\_condition. Calculations dependent on the sample size and risk factor.

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

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

[Vassar Stats](http://vassarstats.net/odds2x2.html): A tool used to perform contingency table analysis.