# 2-Way Hypothesis Testing Theory

- Concepts for Chi Square Tests Refresher
  - The properties of a Chi-Square distribution:
    - It is right-skewed and therefore, non-negative
    - The degrees of freedom vary depending on type of test:
      - One-way Goodness of fit: #categories - 1
      - Two-way: (#rows - 1)(#columns - 1) ← can think #cat.-1 for variable 1 x #cat.-1 for variable 2
        - For Independence and Homogeneity
      - The mean of this distribution also equals the degrees of freedom
  - The observed counts ($O$) are the number of observations that fall into each category while the expected counts ($E$) are the number of observations we think will fall into each category.
  - Sample size is large when each expected count $\geq 5$
- Test of Independence Specific:
  - When we have 2 different variable data sets that we need to see if there is any association
  - The hypothesis Statements:
    - Null: The variables are independent.
    - Alternative: The variables are dependent on each other.
  - Interpretation
    - Rejecting

      At [$\alpha$]% level of significance, we have sufficient evidence to say that there is a relationship between [the 2 variables in question].

    - Failing to Reject

      At [$\alpha$]% level of significance, we have insufficient evidence to say that there is a relationship between [the 2 variables in question].

- Test of Homogeneity Specific:

- When we have 2 different variable data sets that we need to see if the distributions are equal
- The hypothesis Statements:
  - Null: The distributions are homogeneous.
  - Alternative: The distributions are not homogeneous.
- Interpretation
  - Rejecting

At [α]% level of significance, we have sufficient evidence to say that the distribution of variables is not homogeneous.

  - Failing to Reject

At [α]% level of significance, we have insufficient evidence to say that the distribution of variables is not homogeneous.

- What applies to both Independence and Homogeneity
  - You should be given:
    - A sample size (n) from a single, overall population
      - Must have ALL expected counts $\geq 5$
    - Two categorical variables of interest
      - $C_X$ – X categories that represents the rows
      - $C_Y$ – Y categories that represents the columns
    - A level of significance ($\alpha$)
  - Formulas
    - Expected counts

$$E_{i,j} = \frac{(i^{th} \text{ row sum})(j^{th} \text{ column sum})}{n}$$

    - Degrees of Freedom

$$df = (\#\text{rows} - 1)(\#\text{columns} - 1)$$

    - Test Statistic

$$TS = \sum \left( \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \right)$$

- P-value

$$p = \chi^2 cdf(TS, Eqq, df)$$

- Making a Decision
  - Rejecting ($\alpha > p$)
  - Failing to reject ($\alpha < p$)

- Calculator Trick
  - Part 1: The Matrix
    - How you get to the function

2nd → $x^{-1}$ → Edit → Pick an empty matrix

    - What you need to know
      - The matrix is set up via rows x columns
      - You only enter the observed counts
  - Part 2: X2-Test
    - How you get to the function

Stat → Tests → C: $\chi^2$-Test

    - What you need to know
      - Before starting this function, you only have a matrix of observed counts
      - Just choose an EMPTY matrix for the expected counts
    - What you get from the function
      - Expected counts will fill the empty matrix
      - Test statistic ($\chi^2$)
      - p-value
      - Degrees of Freedom