Goodness of Fit Theory

- Concepts for Chi Square Tests
  - 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 also think of it as #categories-1 for variable X multiplied by #cat.-1 for Var. Y
        - 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 ≥ 5
- Goodness of Fit Testing:
  - When we have known/old data that we are testing against with 1 data set for a single variable
  - You should be given:
    - k – number of categories
    - A proportion/percentage for each category
    - A large sample size (n)
    - Level of significance (α)
    - The observed counts from the experiment
  - The hypotheses statements:
    - Null

      The proportion distributions are correct.

    - Alternative

      At least 1 of the proportion distributions are incorrect.
  - Formulas

- Expected counts

$$E_i = n \cdot p_i$$

- Degrees of freedom

$$df = \#categories - 1 = k - 1$$

- Test Statistic

$$TS = \sum \left(\frac{(O_i - E_i)^2}{E_i} \right)$$

- P-value

$$p = \chi^2cdf(TS, E99, df)$$

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

At α% level of significance, there is sufficient evidence to say that at least 1 of the proportion distributions are incorrect.

  - Failing to Reject

At α% level of significance, there is insufficient evidence to say that at least 1 of the proportion distributions are incorrect.

- Calculator Trick (X^2 GOF-Test)
  - How to get to the function

STAT --> TESTS --> D: X^2GOF-Test

  - What you need
    - Observed counts in a list
    - Expected counts in a list

make sure counts are in the same order for both lists!

    - Degrees of freedom
  - What you will get (that is relevant)

- Test Statistic (χ^2)
- P-value (p)
- Degrees of freedom (again)