Statistics 300: Elementary Statistics

Section 11-2

Chapter 11 concerns the analysis of statistics that are "counts" in "categories"

Section 11-2 concerns "counts" in "categories" where each data value falls in one and only one category.

Chapter 11-2

- Two names; same procedure
- Multinomial Tests
- Goodness-of-Fit Tests

Multinomial Tests

- Binomial models had two possible outcomes, or categories, for each trial
- Multinomial models have three or more possible outcomes, or categories, for each trial

Multinomial Tests

- As with Binomial models, there is a probability that each trial will fall in each category
- The sum of the probabilities of all the categories must equal 1

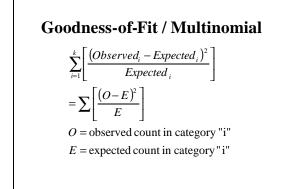
Goodness-of-Fit Tests

- "Goodness-of-Fit" is an idea that can be applied in to situations other than Multinomial models
- In this case, a good fit means that the relative frequency of the data in each category is close to the hypothesized probability

Goodness-of-Fit / Multinomial

- Compare the counts in each category to the number expected for each category
- Test statistic with "k" categories

 $\sum_{i=1}^{k} \left[\frac{(Observed_{i} - Expected_{i})^{2}}{Expected_{i}} \right]$



Goodness-of-Fit / Multinomial

- Observed counts come from the data
- Expected counts come from the hypothesis
- If H₀: is correct, the test statistic should follow a chi-square distribution with k-1 degrees of freedom

$$\sum_{i=1}^{k} \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

Goodness-of-Fit / Multinomial

- Two general types of problems that specify how "expected" counts should be done
- All categories have equal proportions
 - Expected counts are all the same $% \left({{{\mathbf{F}}_{\mathbf{F}}}^{T}} \right)$
 - Expected count = (1/k)*N
 - N = total of observed counts
- Each category has a specified proportion
 - $\mathbf{p}_i = \mathbf{proportion}$ for category "i", and $\sum p_i = 1$
 - (Expected Count)_i = p_i*N
 - N = total of observed counts

Multinomial / Goodness of Fit

- All tests are "right tailed" tests
- Why?
- Because when the test statistic is close to zero, the data are in agreement with the null hypothesis
- The null hypothesis is only rejected when the test statistic value is large, i.e., in the right tail critical region