§10.2 Multinomial Experiments: Goodness of Fit

Characteristics of Multinomial Experiment

1. Fixed # of trials
2. Trials are independent
3. Outcomes are classified into one of n≥2 categories
4. Probability of categories remain constant

If workers are really sick when they take a sick day then there should be no difference in the frequency of absences given the day of the week. Do you think that the workers really only call in sick when they are sick? Test the claim, that workers only call in sick when they are sick, at the alpha = 0.01 level using the following information from a sample of 100 workers.

Day	Mon	Tues	Wed	Thurs	Fri
# Absent	27	19	22	20	12

§10.3 Contingency Tables: Independence & Homogeneity

In this section we will deal with 2 way tables – row and columns are used to classify 2 characteristics of the data. Our two goals will be to see if the row and column dat is independent or to see if different populations have different characteristics (for instance men vs. women).

Focus #1 – Independence

Assumptions

1. Random sample
2. Row and column variables are independent under the null and dependent under the alternative hypotheses
3. Expected frequency is >= 5 (observed can be anything)
4. Expected values are the row total multiplied by the column total divided by the grand total (this is the probability of a row times the probability of a column – what we would expect if the rows and columns were independent)
5. Chi-Squared Test Statistic with degrees of freedom (row-1)(column-1)
6. Test statistic looks just like for Goodness of Fit

Example: Recall the data from §3.3 of the marital status and whether or not a person smokes. Let's see if smoking and marital status are independent at the alpha 0.05 level.

	Married	Divorced	Single	
Smokes	54	38	11	103
Not Smoke	146	62	39	247
	200	100	50	350

You may recall there were instances when the probability from a contingency table of an intersection was extremely close to the probability if you incorrectly assumed independence – well that's because the information may well have been independent! However, that assumption can't be made when the information given is listed in a two-way table!

Focus #2 – Homogeneity (Gender influence is a popular homogeneity test)

H_0: Proportions are the same for populations
H_A: Proportions are different for the populations

The critical value, expected values and critical values are exactly the same for this test. It is just the question that we are asking that makes the difference. The focus is on the fact that the population can be broken into subpopulations and that may make a difference in proportions. If there are differences in proportions due to differing subpopulations, then there is dependence – refer back to the test for dependence.

Example: The claim has been made that males do better in math courses than females. At the alpha 0.01 level test this claim.

	Low	Average	High	
Male	15	60	25	100
Female	10	35	5	50
	25	95	30	150