Math 244 Lecture Notes

Chapter 24: Chi-Squared Tests for Homogeneity and Independence

Overview: Last class, we examined a distribution of values using a χ^2 Goodness of Fit test. This time, we will expand on that to test Homogeneity and Independence.

We are using the χ^2-distribution to test claims involving expected and observed counts. For the purposes of this class, we will only be using it for hypthesis tests. For Goodness of Fit, this looked like...

- STEP I: $H_0 : p_1 = \#_1, ..., p_n = \#_n$ and H_A :At least one proportion is different.
- STEP II: Our model is
$$\chi^2 = \Sigma \frac{(Obs - Exp)^2}{Exp}$$
where $Exp = np_i$. The assumptions are
  (a) Random
  (b) Independence/10% Rule for each trial
  (c) Large n: $np_i \geq 5$. We need five expected counts for each category. NOTE: No restriction on the observed.
  (d) Countable data.
- STEP III: Find the P-Value.
χ^2 is always a right-tailed test, so $P - Val = \chi^2 cdf(Low\, \chi^2, \infty, df)$.
- STEP IV: Conclusion.

Also, we determined how rare individual categories were using standardized residuals. These acted like Z-scores.
$$St.Res. = \frac{Obs - Exp}{\sqrt{Exp}}$$

Example 1. According to Marvel Comics, 10% of all Ms. Marvel covers are "variants". If 200 Ms. Marvel comics are sampled and 15 variant covers are found, what can we say about Marvel's claim for the distribution?

Independence Review: Before moving on to new topics, let's revisit the idea of independent variables from MTH243.

Two random events, A and B, are considered to be independent if any of the following are true:

- $P(A) = P(A|B)$.
- The outcome of A does not affect the likelihood of an outcome for B.
- $P(A \text{ and } B) = P(A)P(B)$.
- The column/rows are proportional. That is, the columns will have similar ratios.

Example 2. Cat Adoption Team in Sherwood has 60 cats. Of those, 20 are kittens, 24 are male, and 10 are male kittens. Are the events of "male" and "kitten" independent?

	Male	Not Male
Kitten		
Not Kitten		

Example 3. Determine if events A and B are independent in the following table.

	A	Not A	**Total**
B	3	12	**15**
Not B	13	52	**65**
Total	**16**	**64**	**80**

Example 4. Fill in the following table assuming that A and B are independent. You might get decimals. That's okay.

	A	Not A	**Total**
B			**25**
Not B			**75**
Total	**20**	**80**	**100**

Assuming that two events are independent, we can find $EXP_{A \text{ and } B} =$

Degrees of Freedom: Given row/column totals, how many cells do we need to fill in before we can solve for the rest by adding/subtracting?

Given a $r \times c$ matrix, we find the df by $df =$

Now that we have a way to find expected counts and the degrees of freedom, we can test for independence using χ^2.

Example 5. Cat Adoption Team in Sherwood has 60 cats. Of those, 20 are kittens, 24 are male, and 10 are male kittens. Are the events of "male" and "kitten" independent?

	Male	Not Male
Kitten		
Not Kitten		

STEP I:

STEP II:

STEP III:

STEP IV:

Example 6. What is the standardized residual for Male Kitten?

This same process can be used to check if two variables are related (Independent) or if multiple column distributions appear the same (Homogeneity).

A test for **Goodness of Fit** compares one distribution for a variable to a claimed distribution.
A test for **Independence** compares two or more variables to see if they are related. Typically, the information all comes from one large sample and it is sorted accordingly.
A test for **Homogeneity** compares one variable for several subpopulations. Typically, the information is collected separately for each subpopulation.

Example 7. Determine what test should be used.

(a) We want to see if there is a relationship between gender and health care coverage (Covered/Not Covered).

(b) The Division Dean wants to compare the final grade distributions for three different professors to see if they look the same.

(c) Laura at OHSU wants to check if there is a relationship between BP (high/low) and diabetes (yes/no).

(d) A trading card company claims that 5% of cards are rare, 20% are uncommon, and 75% are common. We want to check the claim.

Example 8. Let's do one of these in detail: The Division Dean wants to compare the grade distribution for two different instructors. For Mr. Wherry, he passed 50 students, failed 20, and 10 withdrew. For Mrs. Mushet, she passed 80 students, failed 10, and 30 withdrew. Do the distributions appear the same?

Verifying our answers in the calculator takes two processes. We first must type in our matrix. Then, we must perform the test.

Creating a Matrix: If you have a TI-83, TI-84, hit [2nd]→[Matrix] and [Edit] Matrix A. You will need to state the Rows then Columns when determining the size.

If you have a TI-89, go into [Apps]→[Data/Matrix]. Select the option for “Matrix” and name it something you’ll remember. State the number of Rows then Columns. Then, type in your data.

Checking H-Test with Calculator: In our calculator, we use “Chi-SquaredTest” (not GoF) to check our hypothesis test. This is found in either [Stat]→[Tests] on the TI-83/84 OR [Stat/List]→[F6:Tests] on the TI-89. Type in your relevant information and you are good to go!

Calc P-Value:

Options, Options, Options: We now have multiple ways of doing problems involving proportions and counts.

Example 9. Test the claim that the probability of winning a contest by Coca Cola is 1 in 4 if we actually win 10 times in 60 trials. Do this twice: Once with χ^2 GoF and again with 1-propZtest.

Example 10. Hal and Carol want to test if there is a difference in the proportion of times that they are late to work. Hal was late 50 times in 200 days. In contrast, Carol was late 40 times in 200 days. Do this twice: Once with χ^2 Homogeneity and again with 2-propZtest.

OBSERVATIONS: