Chi Square Tests of Significance

There are three types of Chi Square tests:
1. Goodness of Fit
2. Independence
3. Test for Homogeneity of Proportions

Goodness of Fit

When an analyst attempts to fit a statistical model to observed data, he or she may wonder how well the model actually reflects the data. How "close" are the observed values to those which would be expected under the fitted model? One statistical test that addresses this issue is the chi-square goodness of fit test.

The formula for the test statistic for Chi Square is

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected}$$

The larger the Chi Square value the more likely the model does not fit because the observed is far away from the expected values.

Example:

In a recent year, at the 6 pm time slot, tv channels 2, 3, 4 and 5 captured the entire audience with 30%, 25%, 20%, and 25% respectively. The tv executives think these numbers are no longer correct. During the first week of the next season they decide to poll 500 viewers.

Step 1: Conditions
1. Randomization. SRS.
2. The expected values must all be greater than or equal to 5.

Step 2: Hypotheses
H_0: The television audience is distributed over channels 2, 3, 4, and 5 with percentages 30, 25, 30, and 25% respectively.
H_a: The audience distribution is not 30%, 25%, 30% and 25% respectively.

Step 3: Observed table (will be given)

	2	3	4	5
Observed number	139	138	112	111

Step 4: Expected table. What do we expect the numbers to be if the null hypothesis is true?

	2	3	4	5
Expected number	.30(500)=150	.25(500)=125	.20(500)-100	.25(500)=125

Step 5: Calculate Chi Square

$$\chi^2 = \frac{(139-150)^2}{150} + \frac{(138-125)^2}{125} + \frac{(112-100)^2}{100} + \frac{(111-125)^2}{125} = 5.167$$

Step 6: Calculate p-value. Because of the positive nature of Chi Square it will always be positive….so we are always looking at a right tailed test.

Degrees of freedom = n – 1 = 3

$\chi^2 cdf(5.167, 1E99, 3)$ = P($\chi^2 > 5.167) = .1600$

Step 6: Conclusion
With this large of a p-value, there is insignificant evidence to reject the null hypothesis. There is not sufficient evidence that the viewer preferences have changed.

Example 2: You try…

A grocery store manager wishes to determine whether a certain product will sell equally well in any of five locations in the store. Five displays are set up, one in each location, and the resulting numbers of the product sold are noted.

	1	2	3	4	5
Observed number	43	29	52	34	48

Is there significant evidence to conclude that location makes a difference? Test at both the 5% and 10% level of significance.

Homework:

Chi-Square Goodness of Fit Test

1. A researcher wanted to detemine whether pedestrian deaths were unformly distributed over the days of the week. She randomly selected 300 pedestrain deaths, recorded the day of the week on which the death occurred, and obtained the following results (data is from Insurance Institute of Highway Safety).

Day of week	Observed Frequency	Expected Frequency
Sunday	39	
Monday	40	
Tuesday	30	
Wednesday	40	
Thursday	41	
Friday	49	
Saturday	61	

Are pedestrain fatalities equally likely on the days of the week?

H_0 : Pedestrain fatalities occur with equal frequencies on each day of the week
H_A : Pedestrain fatalities do not occur with equal frequencies on each day of the week

2. Lacking a table of random digits, Dan decided to use the last digits in a set of five-place tables of logarithms. The number of times each integer occurred in a sample of size $n = 50$ is given below. If the collection of last digits is random, then we would expect each integer to occur, on average, equally often.

Integer	0	1	2	3	4	5	6	7	8	9
Frequency	8	6	7	3	5	7	1	4	6	3

Do the observed frequencies meet this expectation? Perform the appropriate test and report your findings.