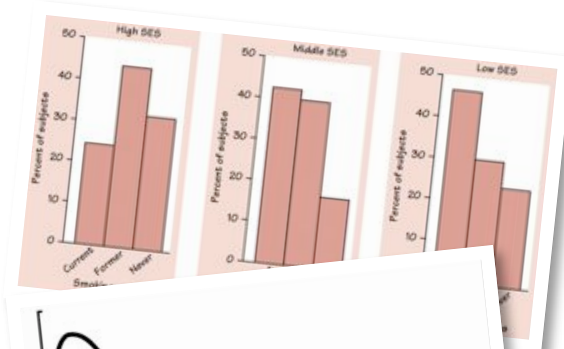
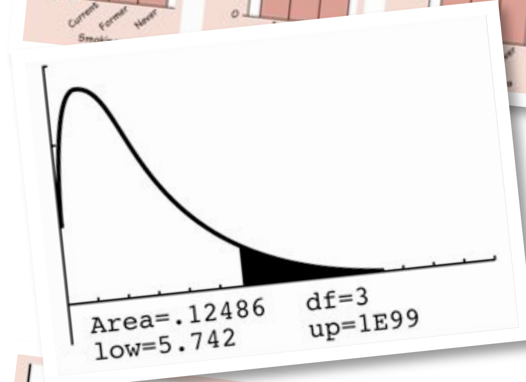
Chapter 13

Chi-Square Procedures

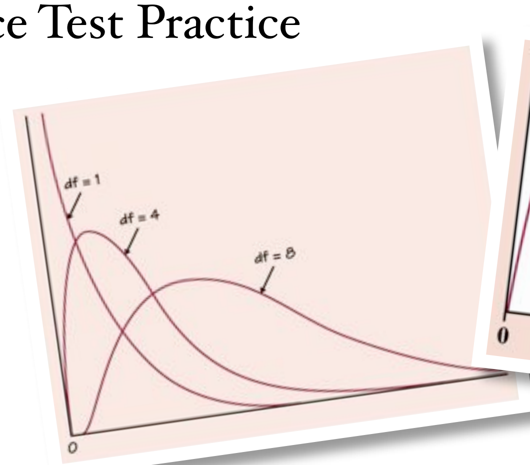
In the previous chapters, we discussed inference procedures for means and proportions. In some cases, we want to examine the distribution of proportions for a population or determine whether the distribution of one variable has been influenced by another. Chi-square procedures help us in these situations.

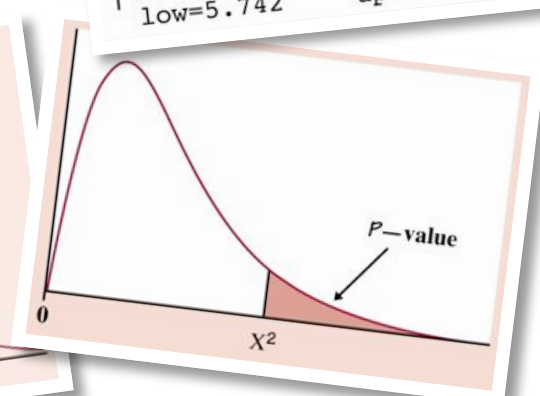




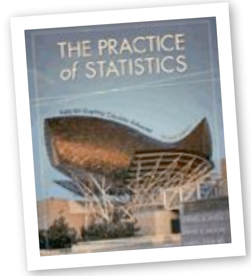
Chi-Square Procedures:

- 13.1 Test for Goodness of Fit
- 13.2 Inference for Two-Way Tables
- Significance Test Practice





AP Statistics Chapter 13: Chi-Square Procedures

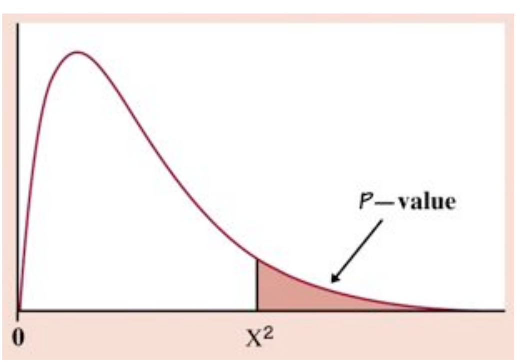


"A statistical analysis, properly conducted, is a delicate dissection of uncertainties, a surgery of suppositions."
~M.J. Moroney

Tentative Lesson Guide					
Date		**Stats**	**Lesson**	**Assignment**	**Done**
Mon	3/19	13.1	Goodness of Fit	Rd 727-743 Do 1-4, 10-13	
Tues	3/20	13.2	Test for Homogeneity	Rd 744-756 Do 14, 16-18	
Wed	3/21	13.2	Test for Independence	Rd 757-766 Do 19, 25-29	
Thu	3/22	**Rev**	**Review**	Do 31-35, 39	
Fri	3/23	**Ex**	**Exam Chapter 13**	**Online Quiz Due**	

Note:
The purpose of this guide is to help you organize your studies for this chapter. The schedule and assignments may change slightly.

Keep your homework organized and refer to this when you turn in your assignments at the end of the chapter.



Class Website:
Be sure to log on to the class website for notes, worksheets, links to our text companion site, etc.

http://web.mac.com/statsmonkey

Don't forget to take your online quiz!. Be sure to enter my email address correctly!
http://bcs.whfreeman.com/yates2e

My email address is:

jmmolesky@isd194.k12.mn.us

Chapter 13 Objectives and Skills:

These are the expectations for this chapter. You should be able to answer these questions and perform these tasks accurately and thoroughly. Although this is not an exhaustive review sheet, it gives a good idea of the "big picture" skills that you should have after completing this chapter. The more thoroughly and accurately you can complete these tasks, the better your preparation.

Chi-Square Distributions

- ☐ Describe Chi-Square Distributions
- ☐ Recognize when to use a Goodness of Fit Test, Test for Homogeneity, or Test for Independence

Goodness of Fit Test

- ☐ Use percents and bar graphs to compare hypothesized and actual distributions.
- ☐ Calculate expected counts.
- ☐ Calculate the chi-square statistic.
- ☐ Conduct a Goodness of Fit Test to determine if a population distribution is different from a specified distribution.

Calculator Skills

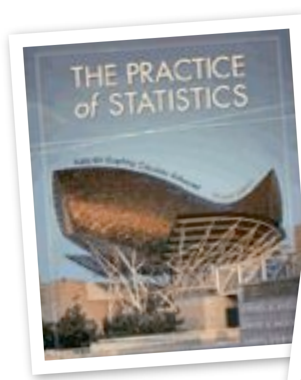
- ☐ Enter two-way table data into a matrix on the TI.
- ☐ Determine expected counts, chi-square statistic, and p-value from calculator output.
- ☐ Use calculator output to write a complete Chi-Square significance test.

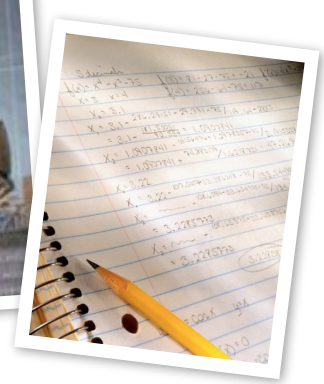
Test for Homogeneity

- ☐ Organize data into a two-way table.
- ☐ Use percents and bar graphs to compare distributions.
- ☐ Calculate expected counts.
- ☐ Calculate the chi-square statistic.
- ☐ Determine degrees of freedom.
- ☐ Perform a Test for Homogeneity to determine if the distribution of a categorical variable is the same in multiple populations.

Test for Independence

- ☐ Organize data into a two-way table.
- ☐ Use percents and bar graphs to compare distributions.
- ☐ Calculate expected counts.
- ☐ Calculate the chi-square statistic.
- ☐ Determine degrees of freedom.
- ☐ Perform a Test for Independence to determine if there is an association between two categorical variables.





13.1: Goodness of Fit Test

According to the m&m/Mars company, in 1995*"...the new mix of colors of m&m's plain chocolate candies will contain 30 percent browns, 20 percent yellows and reds, and 10 percent each of oranges, greens, and blues."* However, the mix of colors has been known to change every few years. Your task today is to determine whether or not the current mix of colors matches that of 1995. We want to see if there is sufficient evidence to reject the company's 1995 claim. To do this, we'll be introduced to a new type of test -- the **Chi-square Goodness of Fit Test**.

- Open a bag of milk chocolate m&m's and carefully count how many of each color are in the sample. (Or, use the data from your teacher's bag) Record the observed data in the "**observed**" row of the table below.
- Using the statement from the m&m/Mars company, determine how many of each color you expected to see. Note, you'll have to figure this out using the total number of m&m's in your (or your teacher's) sample bag. Enter these counts in the "**expected**" row below.

	Brown	Yellow	Red	Orange	Green	Blue	Total
Observed							
Expected							
$(O-E)^2/E$							

If your bag reflects the distribution advertised in 1995, there should be little difference between the observed and expected counts. To quantify the difference, we'll calculate a total which we'll call "Chi-Square" or X^2.

- For each color, perform this calculation: (observed - expected)^2 / expected. Enter each value in the last row of the table. Add up all of these "component" values to find X^2.
- If this total value is small, we have little evidence to suggest a difference in distributions. However, the larger X^2 gets, the more evidence we have to suggest the company's claim may no longer be applicable to bags of milk chocolate m&m's.

To determine the likelihood of observing a difference between observed and expected as extreme as the one we observed, we must look up the p-value on a Chi-square table. Chi-square distributions are skewed right and specified by degrees of freedom. In a Goodness of Fit test, the degrees of freedom equal one less than the number of categories.

X^2 Distribution

Find the p-value for our test by looking up X^2 for 5 degrees of freedom. Sketch the curve and observed X^2 below. Interpret the result in the context of the problem.

13.1: Goodness of Fit Test

A Goodness of Fit Test is used to determine whether a population has a certain hypothesized distribution. The null hypothesis is that the population proportions are equal to the hypothesized proportions. The alternative is that at least two of the proportions differ from the hypothesized proportions. If all expected counts are at least 1 and 80% of them are greater than 5, then X^2 = SUM (observed-expected)^2/expected has an approximately Chi-Square Distribution with df=(k-1).

Chi-Square Distributions

The chi-square distributions are a family of distributions that take on only positive values and are skewed to the right. A specific chi-square distribution is determined by its degrees of freedom.

Example: Accidents and Cell Phones

Are motor vehicle accidents due to cell phones equally likely on any given day of the week? A study of 699 drivers who were using a cell phone when they were involved in an accident examined this question. Use the data below to carry out a Goodness of Fit Test to determine whether or not there is significant evidence that accidents due to cell phone use are not equally likely on each day of the week.

	Sun	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**	**Total**
Observed	**20**	**133**	**126**	**159**	**136**	**113**	**12**	**699**
Expected								
(O-E)^2/E								

Hypotheses:

Ho:

Ha:

Conditions:

Sampling Distribution of X^2 :

Conclusion:

13.2: Chi-Square Test for Homogeneity

The two-sample procedures in Chapter 12 allow us to compare two proportions from two populations. What if we want to compare more than two proportions? We'll need a new test for that. If data is presented in a two-way table, we can compare multiple proportions by using a Chi-Square Test for Homogeneity.

Example: Does Background Music Influence Wine Purchases?

A study in a supermarket in Northern Ireland was conducted to determine whether or not the sales of wine changed relative to the type of background music that was played. Researchers recorded the amount and type of wine that was sold while Italian, French, and no music was played.

	Music			
Wine	**None**	**French**	**Italian**	**Total**
French	**30**	**39**	**30**	**99**
Italian	**11**	**1**	**19**	**31**
Other	**43**	**35**	**35**	**113**
Total	**84**	**75**	**84**	**243**

If music had no effect on the type of wine sold, we would expect to see similar distributions for each type of wine.

Sketch the three wine distributions and compare:

To compare the three population distributions, we must determine what counts we would *expect* to see if the three distributions were the same. To calculate the expected cell counts, we use the following formula...try to determine *why?*

$$\text{expected count} = \frac{(\text{row total}) \bullet (\text{column total})}{\textit{total}}$$

Calculate the expected counts for each cell and enter them in parentheses next to the observed counts.

To test the significance of the difference between the observed and expected counts, we must calculate a X^2 value. If this value is close to zero, there there is not much of a difference between the distributions. However, if this value is large, then we may have evidence that the distributions differ.

$X^2 = \sum \frac{(\textit{observed} - \text{expected})^2}{\text{expected}}$ over all cells in the table. Calculate this value. X^2 =_____________.

How likely was this observed difference? To calculate the p-value, we must look up our X^2 on the table. The degrees of freedom in a test for homogeneity is (row -1)(column - 1). P-value=______________.

Conclusion?

13.2: Chi-Square Test for Independence

In a sense, the Test for Homogeneity can be used to determine whether or not one categorical variable has an effect on another. If the goal of our analysis is to determine an association between two categorical variables, we call the test a Test for Independence. If one variable is affecting the other, then we would expect to see differences between the distributions of counts.

The null hypothesis vs the alternative in a test for independence is

Ho:There is no association between the two categorical variables
Ha: There is an association between the two categorical variables

Chi Square procedures can be used for a test of homogeneity or a test of independence if all expected counts are at least one and if 80% of the expected counts are greater than 5.

If these conditions are met, the distribution of X^2 will be Chi-Square with df=(r-1)(c-1).

Example: Smoking Habits - Students and Parents
How are the smoking habits of students and parents related? Does a parent's habits affect their child's smoking habits? Consider the following data from eight high schools in Arizona and perform a test for independence:

	Student Smokes	**Student Doesn't Smoke**	**Total**
Both Parents Smoke	**400**	**1380**	
One Parent Smokes	**416**	**1823**	
Neither Parent Smokes	**188**	**1168**	
Total			

Hypotheses:
Ho:
Ha:

Conditions:

Sampling Distribution of X^2 :

Conclusion: