8.	p-value: df = 3	Exact p-value: .1511	Reconciliation with Critical Value of Rejection:
9.	Recommended Decision:	Regarding significance: Not Sig	Regarding H_0: Fail to Reject
10.	Interpretation: (English)	if the proportions are like the TA's, then the professors distribution happens 15.11% of the time, This is Not Significant. It appears the professors distribution is similar to the TA's.	

Inference for two-way tables

The two-sample z procedures of Chapter 13 allow us to compare the proportions of successes in two groups (either two populations or two treatment groups in an experiment). What if we want to compare more than two groups? We need a new statistical test. The new test starts by presenting the data as a two-way table. Two-way tables have more general uses than comparing the proportions of successes in several groups. As we saw in Chapter 4, they can be used to describe relationships between any two categorical variables. The same test that compares several proportions also tests whether the row and column variables are related in any two-way table. We will start with the problem of comparing several proportions.

Example 6: Does background music influence wine purchases?

Conditional distributions

Market researchers know that background music can influence the mood and purchasing behavior of customers. One study in a supermarket in Northern Ireland compared three treatments: no music, French accordion music, and Italian string music. Under each condition, the researchers recorded the numbers of bottles of French, Italian, and other wine purchased.[5] Here is a table that summarizes the data:

	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

The conditional distributions of types of wine sold for each kind of music being played and the marginal distribution of the types of wine sold are shown in the following table:

Column percents for wine and music

Wine	Music: None	French	Italian	Total
French	35.7	52.0	35.7	40.7
Italian	13.1	1.3	22.6	12.8
Other	51.2	46.7	41.7	46.5
Total	100.0	100.0	100.0	100.0

There appears to be an association between the music played and the type of wine that customers buy. Sales of Italian wine are very low when French music is playing but are higher when Italian music or no music is playing. French wine is popular in this market, selling well under all music conditions but notably better when French music is playing.

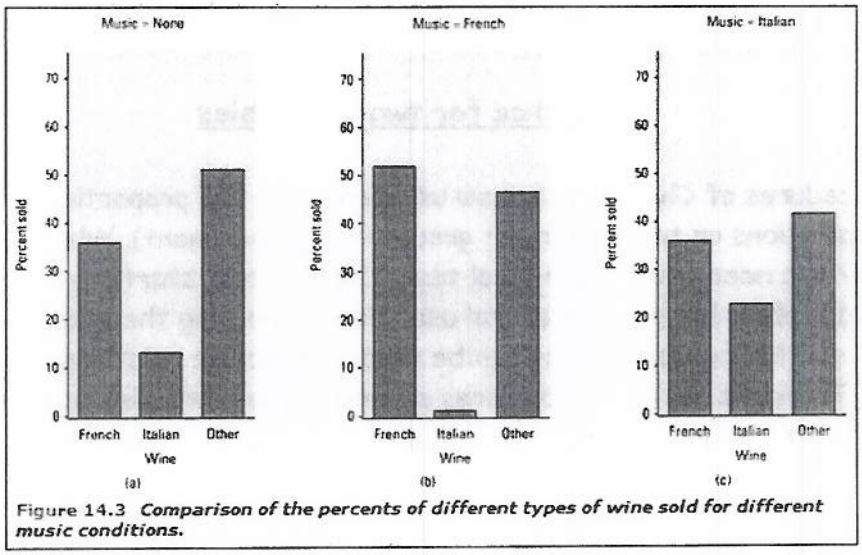


Figure 14.3 ***Comparison of the percents of different types of wine sold for different music conditions.***

Another way to look at these data is to examine the row percents. These fix a type of wine and compare its sales when different types of music are playing. Figure 14.4 on the next page displays these results.

We see that more French wine is sold when French music is playing. Similarly for Italian wine. The negative effect of French music on sales of Italian wine is dramatic.

First, we summarize the observed relation between the music being played and the type of wine purchased. The researchers expected that music would influence sales, so music type is the explanatory variable and the type of wine purchased is the response variable. In general, the clearest way to describe this kind of relationship is to compare the conditional distributions of the response variable for each value of the explanatory variable. So we will compare the column percents that give the conditional distribution of purchases for each type of music played.

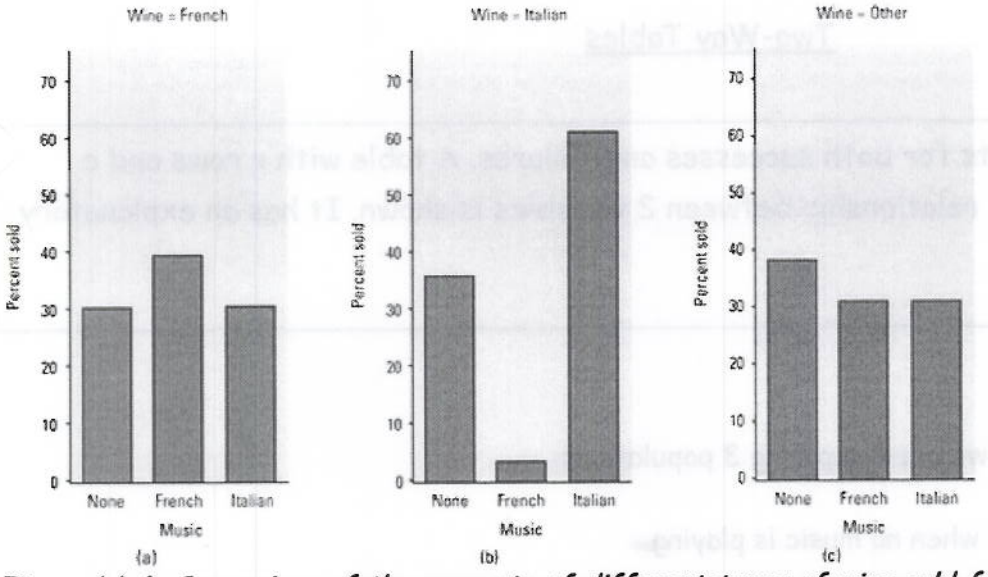


Figure 14.4 ***Comparison of the percents of different types of wine sold for different music conditions.***

To compare the three population distributions we might use chi-square goodness of fit procedures several times:

- Test H_0: the distribution of wine types for no music is the same as the distribution of wine types for French music.
- Test H_0: the distribution of wine types for no music is the same as the distribution of wine types for Italian music.
- Test H_0: the distribution of wine types for French music is the same as the distribution of wine types for Italian music.

The weakness of doing three tests is that we get *three* results, one for each test alone. That doesn't tell us how likely it is that three sample distributions are as different as these are if the corresponding population distributions are the same. It may be that the sample distributions are significantly different if we look at just two groups, but not significantly different if we look at two other groups. We can't safely compare many parameters by doing tests or confidence intervals for two parameters at a time.

The problem of how to do many comparisons at once with some overall measure of confidence in all our conclusions is common in statistics. This is the problem of *multiple comparisons*. Statistical methods for dealing with multiple comparisons usually have two parts:

1. An *overall* test to see if there is good evidence of any differences among the parameters that we want to compare.
2. A detailed *follow-up analysis* to decide which of the parameters differ and to estimate how large the differences are.

The overall test is one with which we are familiar—the chi-square test—but in this new setting it will be used for comparing several population proportions. The follow-up analysis can be quite elaborate.

Two-Way Tables

Two-Way Tables: Gives counts for both successes and failures. A table with r rows and c columns is a r x c table. The relationship between 2 variables is shown. It has an explanatory and a response variable.

Example 7: In our wine example, we are comparing 3 populations:

Population 1: bottles of wine sold when no music is playing

Population 2: bottles of wine sold when French music is playing

Population 3: bottles of wine sold when Italian music is playing

We have three independent samples, of sizes 84, 75, and 84, with a separate sample being taken from each population. The null hypothesis for the chi-square test is

H_0: The proportions of each wine type sold are the same in all three populations.

The parameters of the model are the proportions of the three types of wine that would be sold in each of the three environments. There are three proportions (for French wine, Italian wine, and other wine) for each.

Computing Expected cell counts

Expected Cell counts:

$$\frac{row\ total\ x\ column\ total}{n}$$

Example 8: Find the expected counts for the table from the wine study.

	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

Wine	None	French	Italian	Total
French	34.22	30.56	34.22	99
Italian	10.72	9.57	10.72	31.01
Other	39.06	34.88	39.06	113
Total	84	75.01	84	243

Note that although any count of bottles of wine sold must be a whole number, an expected count need not be.

We can check our work by adding the expected counts to obtain the row and column totals, as in the table. These should be the same as those in the table of observed counts, except for small roundoff errors, such as 113.001 rather than 113 for the total number of bottles of other wine sold.

Example 9: It's hard for smokers to quit. Perhaps prescribing a drug to fight depression will work as well as the usual nicotine patch. Perhaps combining the patch and the drug will work better than either treatment alone. Here are data from a randomized, double-blind trial that compared four treatments.[6] A "success" means that the subject did not smoke for a year following the beginning of the study.

Treatment	Subjects	Successes
Nicotine patch	244	40
Drug	244	74
Patch plus drug	245	87
Placebo	160	25

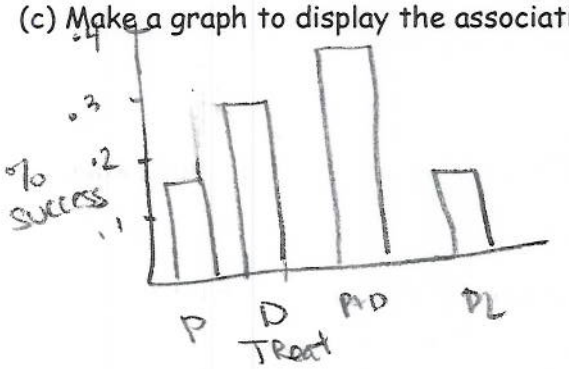
(a) Summarize these data in a two-way table.

Treat	Success	Failure	total
Patch	40	204	244
Drug	74	170	244
Patch + Drug	87	158	245
Placebo	25	135	160
	226	667	893

(b) Calculate the proportion of subjects who refrain from smoking in each of the four treatment groups.

	$\hat{p}$
P	.16
D	.30
P+D	.36
PL	.16

(c) Make a graph to display the association. Describe what you see.



Patch + Drug has highest success rate

(d) Explain in words what the null hypothesis H_0: $p_1 = p_2 = p_3 = p_4$ says about subjects' smoking habits.

All treatments have the same proportion of subjects quit smoking

(e) Find the expected counts if H_0 is true, and display them in a two-way table similar to the table of observed counts.

	S	F
P	61.75	182.25
D	61.75	182.25
P+D	62.00	183.00
PL	40.49	119.51

(f) Compare the tables of observed and expected counts. Explain how the comparison expresses the same association you see in (b) and (c).

The observed success is lower than expected for all categories except Patch + Drug. It appears this is the most effective treatment.

Chi-Square Test for homogeneity of Populations

Chi-Square Statistic

The **chi-square statistic** is a measure of how far the observed counts in a two-way table are from the expected counts. The formula for the statistic is

$$X^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$$

The sum is over all $r \times c$ cells in the table.

The X^2 Statistic and its P-value

As in the test for goodness of fit, you should think of the chi-square statistic X^2 as a measure of the distance of the observed counts from the expected counts. Like any distance, it is always zero or positive, and it is zero only when the observed counts are exactly equal to the expected counts. Large values of X^2 are evidence against H_0 because they say that the observed counts are far from what we would expect if H_0 were true. Although the alternative hypothesis H_a is ***many-sided***, the chi-square test is one-sided because any violation of H_0 tends to produce a large value of X^2. Small values of X^2 are not evidence against H_0.

The same chi-square procedure that we used to test goodness of fit allows us to compare the distribution of proportions in several populations, provided that we take *separate and independent random samples* from each population.

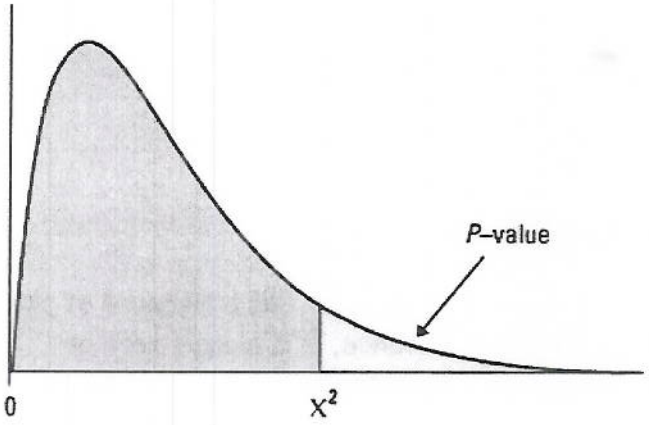
Chi-Square Test for Homogeneity of Populations

Select independent SRSs from each of c populations. Classify each individual in a sample according to a categorical response variable with r possible values. There are c different sets of proportions to be compared, one for each population.

The null hypothesis is that the distribution of the response variable is the same in all c populations. The alternative hypothesis says that these c distributions are not all the same.

If H_0 is true, the chi-square statistic X^2 has approximately a χ^2 distribution with $(r - 1)(c - 1)$ degrees of freedom (df).

The P-value for the chi-square test is the area to the right of X^2 under the chi-square density curve with $(r - 1)(c - 1)$ degrees of freedom.



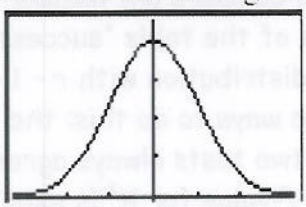
The chi-square test, like the z procedure for comparing two proportions, is an approximate method that becomes more accurate as the counts in the cells of the table get larger. Fortunately, the approximation is accurate for quite modest counts. Here is a practical guideline.[11]

You can safely use the chi-square test with critical values from the chi-square distribution when no more than 20% of the expected counts are less than 5 and all individual expected counts are 1 or greater. In particular, all four expected counts in a 2 × 2 table should be 5 or greater.

When to use a Chi-Square Test:

You can safely use the chi-square test with critical values from the chi-square distribution when no more than 20% of the expected counts are less than 5 and all individual expected counts are 1 or greater. In particular, all four expected counts in a 2 by 2 table should be 5 or greater.

Example 10: Perform the chi-square test on the association between music and wine.

1.	Parameter of Interest:	Population: amount of wine sold w/ various types of music the proportion of types of wine sold	
2.	Choice of test:	Chi-Squared for Homogeneity of Populations	
3.	Check of conditions:	Statement of necessary conditions: all expected cell counts are ≥ 1 & all are greater than 5 Willing to call each sample an SRS	Verification of satisfaction:
4.	Null Hypothesis:	H_0: (English) The distribution of wine types are the same in all populations H_0: (symbols)	
5.	Alternative Hypothesis:	H_a: (English) The distribution is not the same for all types of music H_a: (symbols)	
6.	Test Statistic:	Formula: $\chi^2 = 18.27$	Value:
7.	Test: Level of Significance $\alpha =$ ____	Sketch of sampling distribution assuming that H_0 is true.  Identify the location of the test statistic in the sketch and shade the appropriate region for the *p*-value	

8.	*p*-value: d.f. = (3-1)(3-1) = 4	Exact *p*-value: .0011	Reconciliation with Critical Value of Rejection:
9.	Recommended Decision:	Regarding significance: Sig	Regarding H_0: Reject
10.	Interpretation: (English)	It appears the populations have cliff distributions as the type of music played has a significant effect on wave heights	

One last warning. The test confirms only that there is some relationship. The percents we have compared describe the nature of the relationship. **The chi-square test does not in itself tell us what population our conclusion describes.** If the study was done in one market on a Saturday, the results may apply only to Saturday shoppers at this market. The researchers may invoke their understanding of consumer behavior to argue that their findings apply more generally, but that is beyond the scope of the statistical analysis.

The chi-square Test and the z test

We can use the chi-square test to compare any number of proportions. If we are comparing r proportions and make the columns of the table "success" and "failure," the counts form an $r \times 2$ table. *P*-values come from the chi-square distribution with $r - 1$ degrees of freedom. If $r = 2$, we are comparing just two proportions. We have two ways to do this: the z test and the chi-square test with 1 degree of freedom for a 2×2 table. These two tests always agree. In fact, the chi-square statistic X^2 is just the square of the z statistic, and the *P*-value for X^2 is exactly the same as the two-sided *P*-value for z. We recommend using the z test to compare two proportions because it gives you the choice of a one-sided test and is related to a confidence interval for $p_1 - p_2$.

Example 11: In example 9, you began to analyze data on the effectiveness of several treatments designed to help smokers quit.

(a) Starting from the table of expected counts, find the 8 components of the chi-square statistic and then the statistic X^2 itself.

7.66 + 2.43 + 16.08 + 5.93 + 2.60 + .82 + 3.42 + 2.01

$\chi^2 = 34.94$ $\quad d.f. = (4-1)(2-1) = 3$

(b) Find the *P*-value for the test. Explain what it tells you.

P-value ≈ .0000001 Samples like this are very unlikely to happen very strong evidence against the null

(c) Which term contributes the most to X^2? Does this surprise you?

Patch + Drug and Successes. No - we can see that this group had more than the expected # of Successes

(d) What conclusion would you draw from this study?

Reject the null. Treatment is strongly associated with success. The patch with the drug was the most successful

(e) Perform the chi-square test on your calculator. Are your results the same?

yes!

Chi-Square Test of Association/Independence

Two-way tables can arise in several ways. The music and wine study in is an experiment that compared three music treatments using separate and independent samples. Each group is a sample from a separate population corresponding to a separate treatment. The study design fixes the size of each sample in advance, and the data record which of three outcomes (types of wine purchased) occurred for each category of the explanatory variable (types of music). The null hypothesis of "no difference among the treatments" takes the form of "equal proportions among the three populations." Example 12 illustrates a different setting for a two-way table.

Example 12: Two-way table

Many popular businesses are franchises—think of McDonald's. The owner of a local franchise benefits from the brand recognition, national advertising, and detailed guidelines provided by the franchise chain. In return, he or she pays fees to the franchise firm and agrees to follow its policies. The relationship between the local entrepreneur and the franchise firm is spelled out in a detailed contract.

One clause that the contract may or may not contain is the entrepreneur's right to an exclusive territory. This means that the new outlet will be the only representative of the franchise in a specified territory and will not have to compete with other outlets of the same chain. How does the presence of

an exclusive-territory clause in the contract relate to the survival of the business? A study designed to address this question collected data from a sample of 170 new franchise firms.[14]

Two categorical variables were measured for each firm. First, the firm was classified as successful or not based on whether or not it was still franchising as of a certain date. Second, the contract each firm offered to franchisees was classified according to whether or not there was an exclusive-territory clause. Here are the count data, arranged in a two-way table:

Observed numbers of firms

	Exclusive Territory		
Success	**Yes**	**No**	**Total**
Yes	108	15	123
No	34	13	47
Total	142	28	170

The Chi-Square Test of Association/Independence

Use the **chi-square test of association/independence** to test the null hypothesis

H_0: There is no association between two categorical variables.

when you have a two-way table from a single SRS, with each individual classified according to both of two categorical variables.

Example 13: What are the expected counts for the two-way table above?

Exc

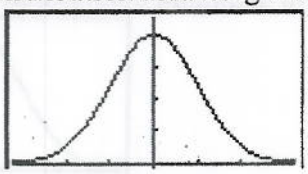
Success	Y	N
Y	102.74	20.26
N	39.26	7.74

Example 14: Are the variables above independent? Run a chi-square test.

Hypothesis Testing Control Form **Page** _____ **Problem** _____

Problem description (if not from book).

Pop: New franchis Firms

1.	Parameter of Interest:		
2.	Choice of test:	Chi-square for Ind	
3.	Check of conditions:	Statement of necessary conditions: Single SRS of 170 Rest. all expected counts over 1	Verification of satisfaction: ✓ overs
4.	Null Hypothesis:	H_0: (English) Exclusive Territory & Success are Ind H_0: (symbols)	
5.	Alternative Hypothesis:	H_a: (English) Exclusive Territory and Success H_a: (symbols) are dependent	
6.	Test Statistic: $\chi^2 =$	Formula: .27+.70+1.37+3.57=5.91	Value:
7.	Test: Level of Significance $\alpha =$ _____ df = 1·1 = 1	Sketch of sampling distribution assuming that H_0 is true.  Identify the location of the test statistic in the sketch and shade the appropriate region for the *p*-value	
8.	*p*-value:	Exact *p*-value: .015	Reconciliation with Critical Value of Rejection:
9.	Recommended Decision:	Regarding significance: Sig	Regarding H_0: Reject
10.	Interpretation: (English)	Success & Exclusive territory seem to be dependent	

Note: Can't say one causes the other!