Chapter 13
Section 2
Inference for two-way tables

Example 13.4 Treating Cocaine Addiction

Chronic users of cocaine need the drug to feel pleasure. Perhaps giving them a medication that fights depression will help them stay off cocaine.

A three year study compared an antidepressant called ***desipramine*** with ***lithium*** (a standard treatment for cocaine addiction) and a placebo.

The subjects were 72 chronic users randomly assigned to each treatment.

Here are the counts and proportions of the subjects who avoided relapse into cocaine use during the study.

Group	Treatment	Subjects	**No relapse**	Proportion
1	Desipramine	24	14	0.583
2	Lithium	24	6	0.250
3	Placebo	24	4	0.167

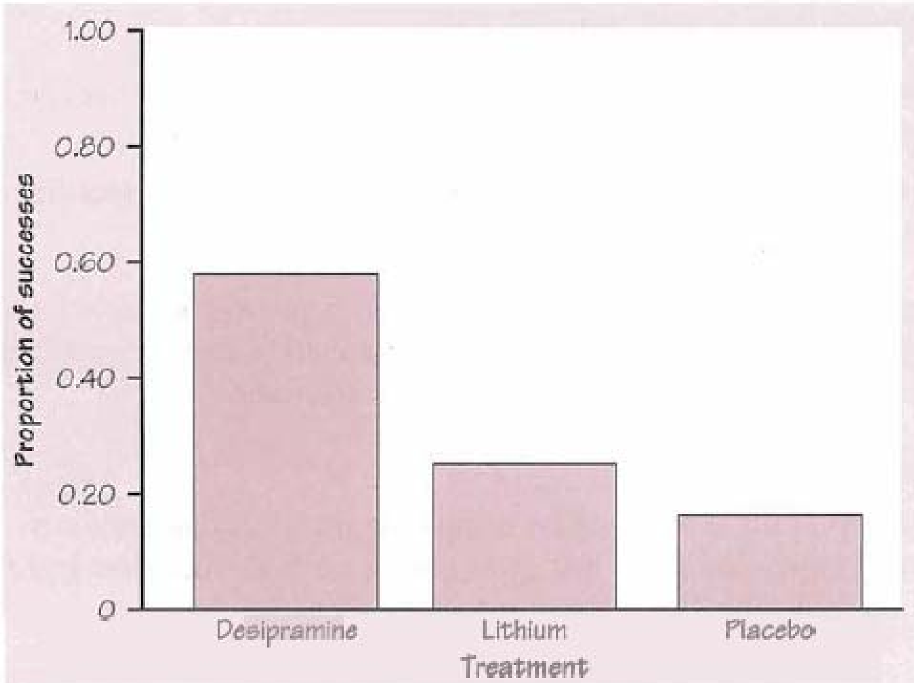


FIGURE 13.3 Bar graph comparing the success rates of three treatments for cocaine addiction.

<u>The problem with multiple comparisons</u>

Test

$H_0 : \hat{p}_1 = \hat{p}_2$ to see if the success rate of desipramine differs from that of lithium

$H_0 : \hat{p}_1 = \hat{p}_3$ to see if desipramine differs from a placebo

$H_0 : \hat{p}_2 = \hat{p}_3$ to see if lithium differs from a placebo

The weakness of doing three tests is that we get three P-values, one for each test alone. That does not tell us how likely it is that three sample proportions are spread apart as far as these are.

It may be that $\hat{p}_1 = 0.583$ and $\hat{p}_3 = 0.167$ are significantly different if we look at just two groups, but not significantly different if we know that they are the smallest and largest proportions in three groups.

As we look at more groups, we expect the gap between the smallest and largest sample proportion to get larger.

Think of comparing the tallest and the shortest person in larger and large groups of people.

We cannot safely compare many parameters by doing tests or confidence intervals for two parameters at a time.

Statistical methods for dealing with many comparisons with some overall measure of confidence 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.

	Relapse	
	NO	YES
Desipramine	14	10
Lithium	6	18
Placebo	4	20

- 3 x 2 table (row x columns table)
- relationship between two categories
- explanatory is the treatment
- response is success (no relapse), failure (relapse)
- 6 combinations, each count occupies a cell of the table

Expected Count

$$H_0 : \hat{p}_1 = \hat{p}_2 = \hat{p}_3$$

H_a : not all p_1, p_2 and p_3 are equal

The **expected count** in any cell of a two-way table when H_0 is true is

$$\text{expected count} = \frac{\text{row total x column total}}{\text{table total}}$$

Relapse

	NO	YES	Total
Desipramine	14	10	24
Lithium	6	18	24
Placebo	4	20	24
Total	24	48	72

The proportion of relapses among all 72 subject is

$$\frac{\text{count of relapses}}{\text{table total}} = \frac{\text{column 2 total}}{\text{table total}} = \frac{48}{72} = \frac{2}{3}$$

Then the expected count for the cell for Desipramine and relapse is:

$$\frac{\text{row 1 total} \times \text{column 2 total}}{\text{table total}} = \frac{(24)(48)}{72}$$

Calculating the remaining cells the same way, we can compare the expected and the observed.

	Observed		Expected	
	NO	YES	NO	YES
Desipramine	14	10	8	16
Lithium	6	18	8	16
Placebo	4	20	8	16

Because 2/3 of all subjects relapsed, we expect 2/3 of the subjects in each group to relapse if there are no differences among the treatments. (H_0)

The 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 chi-square statistics is a sum of terms
- One for each cell in the table

In this example, 14 subjects in the desipramine group succeeded in avoiding a relapse. The expected count for this cell is 8. The component of the chi-square form this cell is

$$\frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}} = \frac{(14-8)^2}{8} = \frac{36}{8} = 4.5$$

- As in the test for goodness of fit, you should think of the chi-square statistic χ^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 χ^2 are evidence against the 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 is many sided, the chi-square test is one-sided because any violation of H_0 tends to produce a large value of χ^2.

- Small values of χ^2 are not evidence against the H_0.

- The chi-square procedure allows us to compare the distribution of proportions in several populations, provided that

    a) we take separate and independent samples from each population.

    b) The samples are SRSs from each of ***c*** populations.

    c) Classify each individual in a sample according to a categorical response variable with ***r*** possible values.

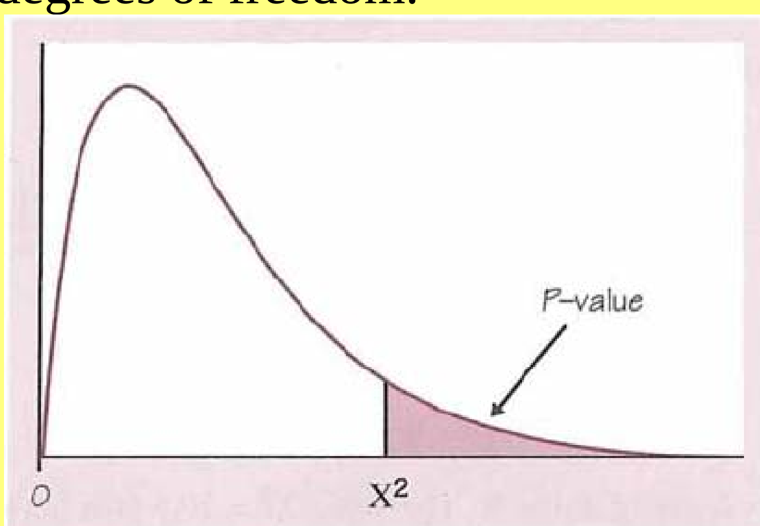
    d) There are ***c*** different sets of proportions to be compared, one for each population.

e) The null hypothesis is that the distribution of the response variable is the same in all ***c*** populations.

f) The alternative hypothesis says that these ***c*** distributions are not all the same.

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

h) The P-value for the chi-square test is the are to the right of χ^2 under the chi-square density curve with ***df*** degrees of freedom.



i) 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.

j) In particular, all four expecte4d counts in a 2 x 2 table should be 5 or greater.

Is Desipramine effective in treating cocaine addition?

Step 1: Identify populations of interest.
State hypotheses in words and symbols.

We want to compare the proportions of cocaine addicts who do not relapse in the populations of patients treated with desipramine (p_1), lithium (p_2), and placebo (p_3).

H_0: $p_1 = p_2 = p_3$.The proportions of cocaine addicts who avoid relapse are the same.
H: Not all three of the proportions are equal.

Step 2:Choose the appropriate inference procedure and verify conditions for its use.

To use the chi-square test for homogeneity of populations:

- The data must come from independent SRSs
- All expected cell counts are greater than 1, and no more than 20% are less than 5.

Step 3:Carry out the procedure.

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(14-8)^2}{8} + \frac{(10-16)^2}{16} + \frac{(6-8)^2}{8} + \frac{(18-16)^2}{16} + \frac{(4-8)^2}{8} + \frac{(20-16)^2}{16} = 10.5$$

$$\chi^2 = 4.50 + 2.25 + 0.50 + 0.25 + 2.00 + 1.00 = 10.50$$

df = (r-1)(c-1)= (3-1)(2-1)=2

Table C χ^2 critical values

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73

Step 4:Interpret your results in context.

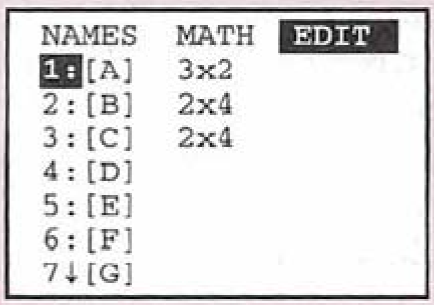
Since the P-value is less than 0.01, the differences among the three proportions are statistically significant at the α=0.01 level. We would reject the null hypothesis.

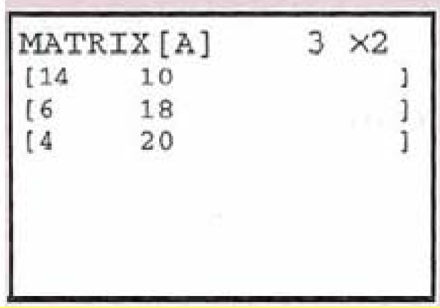
Using the calculator

- Enter the observed counts in the matrix [A].

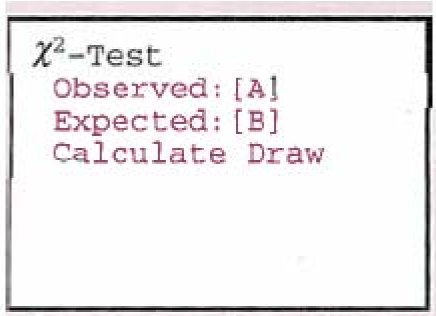
TI-83

- Press 2nd x^{-1}(MATRIX), arrow to EDIT, choose 1:[A].



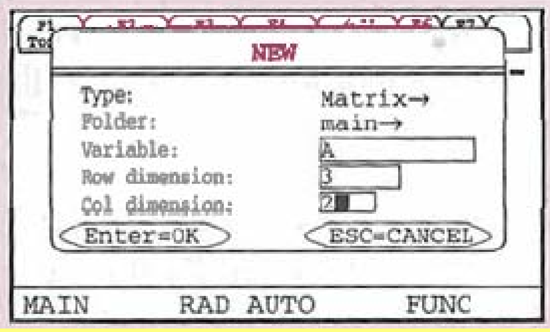


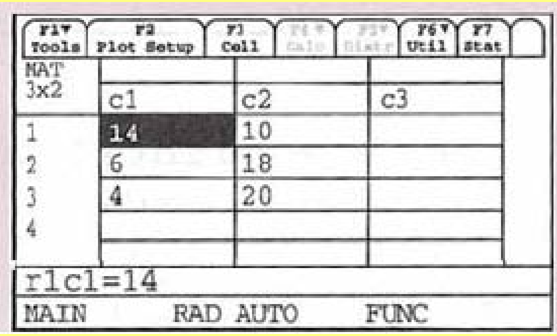
- Press STAT, arrow to TESTS, and choose F1 χ^2 Test. . . .



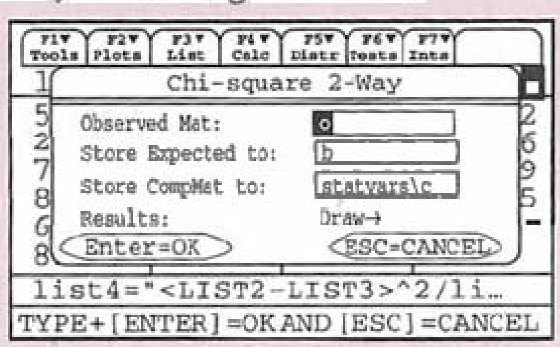
TI-89

- Press APPS, select 6:Data/Matrix Editor, and then 3:New. . . .
- Adjust your settings to match those shown.

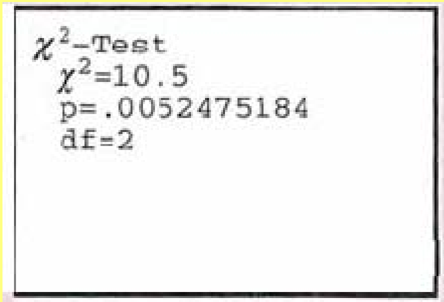


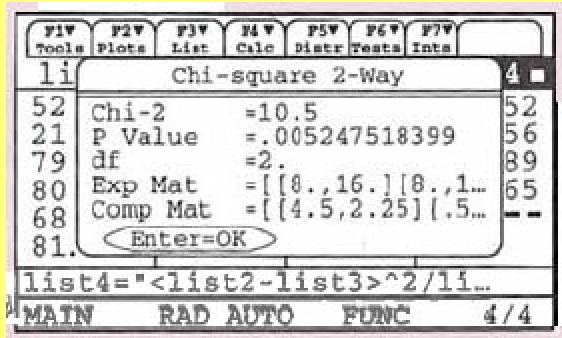


- In the Statistics/List Editor, press 2nd F1 ([F6]) C: and choose 8:Chi2 2-way. . . .
- Adjust your settings as shown.

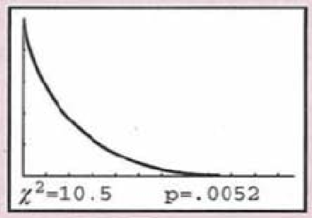


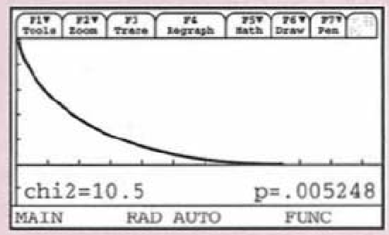
- Choose "Calculate" or "Draw" to carry out the test. If you choose "Calculate," you should get these results:





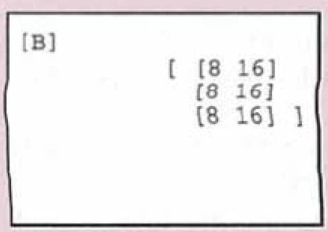
If you specify "Draw," the χ^2 curve with 2 degrees of freedom will be drawn, the critical area in the tail will be shaded, and the *P*-value will be displayed.

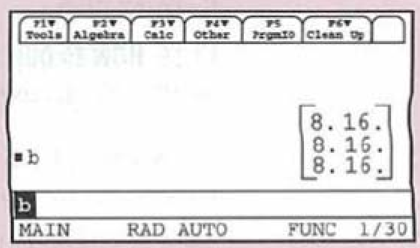




If you want to see the expected counts, simply ask for a display of the matrix [B].

- Press 2nd x^{-1} (MATRIX), and choose 2:[B].
- Press 2nd – (Var-LINK) and choose *b*.





Follow-up analysis

There is no prescription on how to do a formal follow-up analysis, but you should look at the data to see what specific effects they suggest.

A final look at the cocaine study:

There were significant differences among the proportions of successes for three treatments for cocaine addiction. We can see the specific differences in three different ways.

The sample proportions are:

$$\hat{p}_1 = 0.583 \quad \hat{p}_2 = 0.250 \quad \hat{p}_3 = 0.167$$

These suggest that the major difference between them is that desipramine has a much higher success rate than either the lithium or a placebo. That is the effect that the study hoped to find.

Next, compare the observed and expected counts:

- Treatment 1(desipramine) has more success and less failures than we would expect if all three treatments had the same success rate in the population.

- The other treatments had fewer successes and more failures than expected.

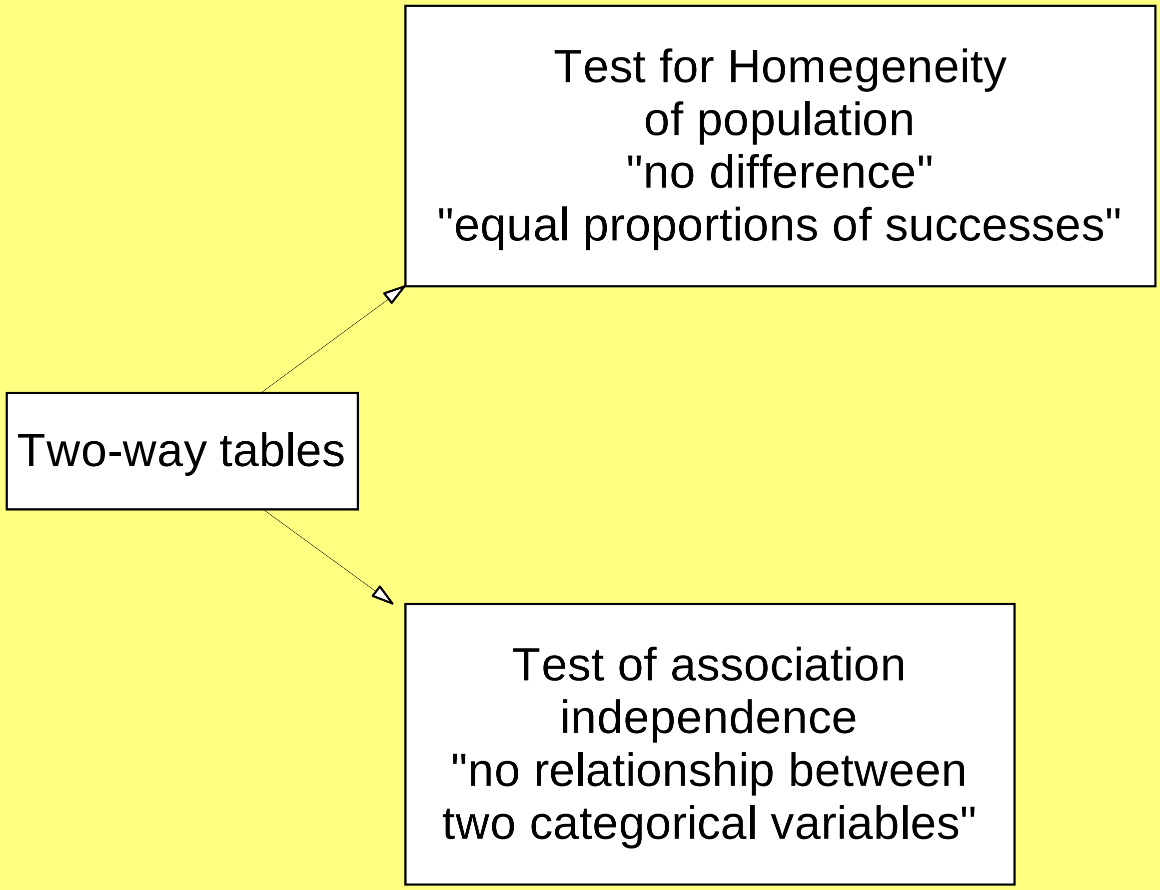
- Looking at the "distances" between observed and expected counts that are added to get χ^2:

$$\chi^2 = 4.50 + 2.25 + 0.50 + 0.25 + 2.00 + 1.00 = 10.50$$

- The arrangement of these "*components*" of χ^2 is the same as the 3 x 2 arrangement on the table. The largest components show which cells contribute the most to the overall distance χ^2.

- The largest component by far is for the top left cell in the table: desipramine has more successes than would be expected.

- All three ways of examining the data point to the same conclusion: desipramine works better than the other treatments.



The Chi-square test for association/independence

<u>Example 13.9</u>

In a study of heart disease in male federal employees, researchers classified 356 volunteer subjects according to their socioeconomic status (SES) and their smoking habits. There were three categories of SES: high, middle, and low. Individuals were asked whether they were current smokers, former smokers, or had never smoked, producing three categories for smoking habits as well. Here is the two-way table that summarizes the data:

Observed counts for smoking and SES

Smoking	High	Middle	Low	Total
	SES			
Current	51	22	43	116
Former	92	21	28	141
Never	68	9	22	99
Total	211	52	93	356

This is a 3 x 3 table of two categorical variables from the same population.

H_0 : there is no association between SES and smoking habits

> **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 relationship 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.

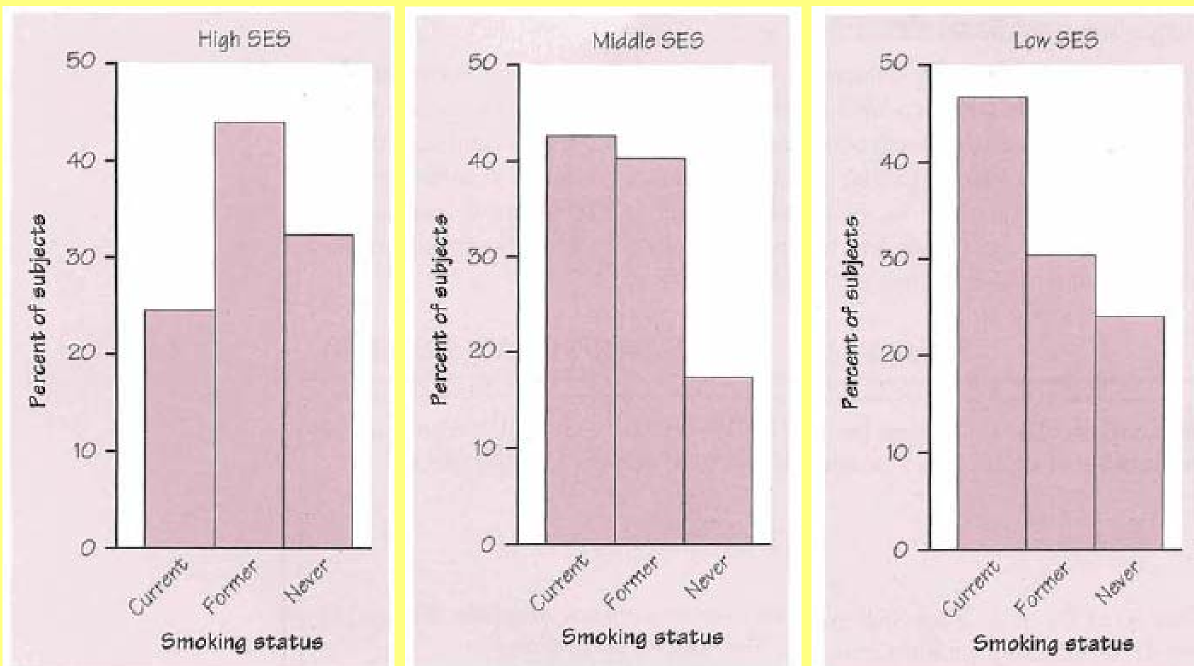
Computing conditional distributions

1. Analysis of data by computing descriptive statistics by categories.
2. SES is the explanatory variable and smoking is the response variable.

$$\frac{51}{211} = 0.242 \qquad \frac{92}{211} = 0.436 \qquad \frac{68}{211} = 0.322$$

Column percents for smoking and SES

	SES		
Smoking	High	Middle	Low
Current	24.2	42.3	46.2
Former	43.6	40.4	30.1
Never	32.2	17.3	23.7
Total	100.0	100.0	100.0



The percent of current smokers decreases as SES increases from low to middle to high. Relatively speaking, few high-SES subjects smoke.

The percent of former smokers increases as SES increases, suggesting that higher-SES smokers were more likely to quit.

The percent of people who never smoked is highest in the high-SES group, but the middle-SES group has a somewhat lower percentage than the low-SES group, but the middle-SES group has a somewhat lower percentage than the low-SES group.

Overall, the column percents suggest that there is a negative association between smoking and SES, higher-SES people tend to smoke less.

The *chi-square test of association/independence assess* whether this observed association is statistically significant. That is,

Is the SES- smoking relationship in the sample sufficiently strong for use to conclude that it is due to a relationship between these two variables in the underlying population and not merely to chance?

Computing expected cell counts

$$\text{expected count} = \frac{\text{row total} \times \text{column total}}{n}$$

$$\frac{211 \times 116}{356} = 68.75 \qquad \frac{211 \times 141}{356} = 83.57 \qquad \frac{211 \times 99}{356} = 58.68$$

Expected counts for smoking and SES

	SES			
Smoking	High	Middle	Low	Total
Current	68.75	16.94	30.30	115.99
Former	83.57	20.60	36.83	141.00
Never	58.68	14.46	25.86	99.00
Total	211.00	52.00	92.99	355.99

Performing the χ^2 test for Association/Independence

Step 1: *State the hypotheses*

H_0 : smoking and SES are independent

H smoking and SES are dependent

You can also say

H_0 : there is no association between smoking and SES

H there is an association between smoking and SES

Step 2: *Choose an inference procedure and verify conditions.*

All expected cell counts are at least 1 and no more than 20% of them are less than 5.

Step 3: *Carry out the test*

$$
\begin{aligned}
X^2 &= \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \\
&= \frac{(51-68.75)^2}{68.75} + \frac{(22-16.94)^2}{16.94} + \frac{(43-30.30)^2}{30.30} \\
&\quad + \frac{(92-83.57)^2}{83.57} + \frac{(21-20.60)^2}{20.60} + \frac{(28-36.83)^2}{36.83} \\
&\quad + \frac{(68-58.68)^2}{58.68} + \frac{(9-14.46)^2}{14.46} + \frac{(22-25.86)^2}{25.86} \\
&= 4.583 + 1.511 + 5.323 + 0.850 + 0.008 + 2.117 + 1.480 + 2.062 + 0.576 \\
&= 18.51
\end{aligned}
$$

Because there are r=3 smoking categories and c=3 SES groups, the degrees of freedom for this statistics are

$$(r-1)(c-1) = (3-1)(3-1) = 4$$

The test statistic χ^2 has a $\chi^2(4)$ distribution. You can find it in table E in your text book or table C in your formula packet.

Table C χ^2 critical values

df	Tail probability p .25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47	20.00
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51	22.11

The calculated value χ^2 =18.51 lies between upper critical points corresponding to probabilities 0.001 and 0.0005. The P-value is therefore between 0.001 and 0.0005.

Step 4: *Interpret your results in context.*
There is strong evidence (P-value < 0.001) of an association between smoking and SES in the population of male federal employees.

A note: This association does not show that SES causes smoking behavior.

Remarks:
How can you distinguish the two type of chi-square test for two-way tables?

a. In the test of association/independence, there is a single sample from an single population. The individuals in the sample are classified according to two categorical variables.
b. For the test of homogeneity of populations, there is a sample from each of two or more populations. Each individual is classified based on a single categorical variable.

The chi-square test and the z test

i. We can use the chi-square test to compare any number of proportions.
ii. If we are comparing r proportions and make the columns of the table "success" and "failure", the counts form an r x 2 table and the P-values come from the chi-square distribution with r-1 ***df***.

iii. If r=2, we are comparing two proportions. We have two ways to do this: the **z** test from Section 12.2 and the chi-square test with 1 degree of freedom for a 2 x 2 table.

iv. These two tests always agree. In fact, the chi-square statistic χ^2 is just the square of the **z** statistic, and the P-value for χ^2 is exactly the same as the two-sided P-value for **z**.

v. ***Recommendation:*** use 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 $p_1 - p_2$.