"I gather, young man, that you wish to be a Member of Parliament. The first lesson you must learn is, when I call for statistics about the rate of infant mortality, what I want is proof that fewer babies died when I was Prime Minister than when anyone else was prime minister. That is a political statistic."

Winston Churchill

13.2 INFERENCE FOR TWO-WAY TABLES (Pages 717 - 738)

OVERVIEW: Sometimes one wants to compare proportions from two or more groups. One can use the chi-square test to do this. Basically, the null hypothesis states that all the population proportions are e qual. The alternate hypothesis is that they are not, which means that there appears to be a significant difference between at least two of the population proportions.

If the null hypothesis is true, the expected count for a cell in a two-way table is

expected count = (row total)(column total)/(table total).

The X^2 statistic is

X^2 = sum[(observed count -expected count)^2/(expected count)].

The **degrees of freedom** is (number of rows - 1)(number of columns - 1).

You can use X^2 when all expected cell counts are at least 1, and n o more than 20% are less than 5.

Example:

The table displays the number of students passed and failed by each of three instructors who teach Introductory Statistics at a specific university.

	Instructor #1	Instructor #2	Instructor #3	TOTALS
Passed	50	47	56	**153**
Failed	5	14	8	**27**
TOTALS	**55**	**61**	**64**	**180**

Suppose we wish to examine the null hypothesis, H_0, that the proportions of students failed by the three instructors are the same. The expected frequencies under H_0 are displayed in the lower right of each observed count cell. Remember that expected count =(row total)(column total)/(table total). For instance,

46.75 = (55)(153)/180.

Note also that all expected counts are more than5, and that degrees of freedom = (2-1)(3-1) = 2.

	Instructor #1	Instructor #2	Instructor #3	TOTALS
Passed	$50_{46.75}$	$47_{51.85}$	$56_{54.40}$	**153**
Failed	$5_{8.25}$	$14_{9.15}$	$8_{9.60}$	**27**
TOTALS	**55**	**61**	**64**	**180**

The calculation of X^2 yields

$x^2 = (50-46.75)^2/46.75 + (47-51.85)^2/51.85 + (56-54.40)^2/54.40 + (5-8.25)^2/8.25 + (14-9.15)^2/9.15 + (8-9.60)^2/9.60 = 4.84$

Using the X^2 table with 2 degrees of freedom, we cannot reject H_0 at the 5%level of significance, since the critical region is $x^2 > 5.99$, and our calculated value is not in this region. At the 10% level of significance, the critical region is $X^2 > 4.61$. We would reject H_0 at this level.

Using the TI-83 to calculate a P-value, we have $\boldsymbol{x^2}$ **cdf (**4.84,1E99,2) = .0889216175, or about 8.9%. Note that this "jives" with rejection of H_0 at the 10% level of significance, and non-rejection at the 5% level.

==========================

On page 735, the authors note that the chi-square test and the z-test "agree" when you have a 2-by-2 table situation. In this situation, the chi-square statistic is the square of the z-statistic, and the P-value for X^2 is the same for the two-sided P-value for z. What follows is an attempt to illustrate this fact using data from Example 12.13(page 683). Here is the data in tabular from with expected values calculated and displayed.

	Gemifibozoil	Placebo	TOTALS
Heart attack	$56_{70.36}$	$84_{69.64}$	140
No heart attack	$1995_{1980.64}$	$1946_{1960.36}$	3941
TOTALS	**2051**	**2030**	**4081**

On page 685, the calculated z-statistics is shown to be z = -2.47.

The calculated X^2 is

$X^2 = (56-70.36)^2/70.36 + (84-69.64)^2/69.64 + (1995-1980.64)^2/1980.64 + (1946-1960.36)^2/1960.36 = 6.101$.

Note that $(-2.47)^2 = 6.101$, illustrating that, in this situation, $X^2 = z^2$.

The degrees of freedom here is (2-1)(2-1) = 1. The P-value for X^2 is X^2cdf(6.101,1E99,1) = .0135105406, or about 1.35%.

On page 685, the authors give a 1-sided P-value(for the z-statistic) of .0068. Note that 2(.0068) = .0136, or about1.36%, further illustrating what the authors are saying in their comparison of X^2 and z on page 735.

RETURN TO TEXTBOOK HOME PAGE