Neuendorf

The Chi-Square (χ^2) Statistic

The chi-square (χ^2) statistic allows us to assess the relationship between two variables that are assumed to be measured at the nominal level. (This means that if the statistic is used on two variables measured at, say, the ordinal level, their ordinality is not taken into account by the chi-square.) This statistic is nonparametric—that is, it does not test whether a relationship exists in a population. The chi-square looks at the difference between an observed two-variable frequency distribution and the chance, or "expected," distribution (i.e., "O" vs. "E"). The formula is:

$$\chi^2 = \sum\left(\frac{(O-E)^2}{E}\right)$$

df = (r-1) x (c-1)
where
r = # of rows in table
c = # of columns in table

Suppose we find the following frequencies in a table relating gender (assuming a sample of just two genders) to one's subscription to the *New York Times*:

Table 1: OBSERVED FREQUENCIES

	F	M	
Subscribers	15	10	25
Non-subs.	12	18	30
	27	28	55

Given no information about the contents of the table, but only information about each of the two variables independently (i.e., F=27, M=28, SUB=25, NO SUB=30), we can calculate expected frequencies:

Table 2: MARGINALS

	F	M	
Subscribers			25
Non-subs.			30
	27	28	55

Table 3: CALCULATION OF EXPECTED FREQUENCIES*

	F	M	
Subscribers	$\frac{(27x25)}{55}$	$\frac{(28x25)}{55}$	25
Non-subs.	$\frac{(27x30)}{55}$	$\frac{(28x30)}{55}$	30
	27	28	55

Table 4: EXPECTED FREQUENCIES

	F	M	
Subscribers	12.3	12.7	25
Non-subs.	14.7	15.3	30
	27	28	55

Now we may calculate chi-square:

$$\chi^2 = \frac{(15-12.3)^2}{12.3} + \frac{(10-12.7)}{12.7} + \frac{(12-14.7)^2}{14.7} + \frac{(18-15.3)^2}{15.3}$$

$$= .59 + .57 + .50 + .48$$

$$\chi^2 = 2.14$$

To test the significance of this figure, we look in a chi-square table. Select a desired level of significance (e.g., $p = .05$). Calculate the degrees of freedom for the test:
$df = (2\text{-}1)\,(2\text{-}1) = 1$

So, under $p = .05$, $df = 1$, the critical value from a chi-square table is 3.84146. Since our chi-square of 2.41 is smaller than this, it is <u>not</u> statistically significant. Our observed frequency distribution for the two variables taken together does not differ substantially from chance. In our sample, gender is not significantly related to *New York Times* subscription.

NOTE: If a significant relationship <u>had</u> been found, we would state it like this, for example: "In our sample, males and females differ significantly in their likelihood of subscribing to the *New York Times*, such that females are more likely to subscribe. This gender difference is strong enough to warrant the conclusion that it is not due to chance."

* -- Notice that each cell entry is a simplification of a basic probability formula using the "multiplication rule":

$$\frac{25}{55} \quad \text{x} \quad \frac{27}{55} \quad \text{x} \quad 55 \quad = \quad \frac{27 \text{ x } 25}{55} \quad = \quad 12.3$$

(prob. of subscriber x prob. of female x n = expected value for cell)

1/25