A SIMPLE METHOD FOR CALCULATING MEAN SQUARE CONTINGENCY

By

ELMER B. ROYER
The Ohio State University

If we wish to test for a possible relationship between two variables which are not quantitatively measurable, but each of which has two or more categories, the usual procedure is to make a two-way table, giving the frequencies of all the possible combinations.

Assuming independence between the two variables, a second table is built, making the frequencies of each column proportional to the frequencies in the column of row totals. When this is done, each of the row frequencies is found to be proportional to the row of column totals.

The deviation of the actual frequency for a compartment as found in Table 1 from the expected frequency as found in Table 2, is squared and this square is divided by the expected frequency. These quotients are summed over the entire table, giving us Chi-square.

The calculation of Chi-square can be made much simpler by simplifying the formula.

The probability of the occurrence of two independent events is the product of their separate probabilities. Thus the probability of the joint occurrence of Category 3 of the first classification and Category d of the second classification is the probability of the occurrence Category 3 (which is taken to be the fraction of the total number of cases which fall in Category 3), times the probability of the occurrence of Category d. The expected frequency of the compartment is this product of separate probabilities, multiplied by the total number of cases. If we let f_a be the actual compartment frequency, f_e the expected compartment frequency,

f_r the total frequency for the row, f_c the total frequency for the column, and N the number of cases, we may write,

$$f_e = \frac{f_r}{N} \cdot \frac{f_c}{N} \cdot N = \frac{f_r f_c}{N} \qquad (1)$$

Also,

$$\chi^2 = \Sigma \frac{(f_a - f_e)^2}{f_e} = \Sigma \frac{f_a^2 - 2 f_a f_e + f_e^2}{f_e} = \Sigma \frac{f_a^2}{f_e} - 2\Sigma f_a + \Sigma f_e . \qquad (2)$$

Since the table must sum to N, whether we have filled it with actual frequencies, or with expected frequencies, (2) reduces to,

$$\chi^2 = \Sigma \frac{f_a^2}{f_e} - N$$

Substituting for f_e from (1),

$$\chi^2 = \Sigma \frac{f_a^2}{\frac{f_r f_c}{N}} - N = N\Sigma \frac{f_a^2}{f_r f_c} - N = N\left[\Sigma \frac{f_a^2}{f_r f_c} - 1\right] \qquad (3)$$

In order to illustrate our method, we shall compute Chi-square for Table 18 on page 86 of Fisher's *Statistical Methods for Research Workers*. Our computations are presented in the following table.

	Black Self	Black Piebald	Brown Self	Brown Piebald	Total
Coupling F_1 Males	88 7744 25.3902	82 6724 22.0459	75 5625 18.4426	60 3600 11.8033	305
F_1 Females	38 1444 11.7398	34 1156 9.3984	30 900 7.3171	21 441 3.5854	123
Repulsion F_1 Males	115 13225 31.6388	93 8649 20.6914	80 6400 15.3110	130 16900 40.4306	418
F_1 Females	96 9216 25.7430	88 7744 21.6313	95 9025 25.2095	79 6241 17.4330	358
Frequency Total	337	297	280	290	1204
Product Total	94.5118	73.7670	66.2802	73.2523	
Quotient	.280450	.248374	.236715	.252594	1.018133

The actual frequency is the first entry in each compartment. The square is read from a table and written directly beneath the frequency. The reciprocal of 305 is put into the keyboard of a calculator and multiplied in turn by each of the squares in the first row. The products make the third entries in the compartments.

These products are summed by columns and the sums divided by the frequency totals of the corresponding columns. These quotients are summed horizontally. This sum is $\Sigma \frac{f_a^2}{f_r f_c}$, and can be substituted in (3). For our example,

$$\chi^2 = 1204(1.018133 - 1.000000)$$
$$= 21.832$$

This answer agrees exactly with the answer obtained by Fisher in his Table 19, page 87. The advantages of this method are two-fold: (1) There is considerable saving of labor; (2) with the simplification of calculations, we have greatly reduced the danger of errors caused by dropping of decimal places.

Elmer B. Royer