where F_2 is the upper and F_1 the lower critical value of the analysis of variance distribution with $p_u - 1$ and $N - \sum_{u=1}^{r} p_u + r - 1$ degrees of freedom. In case of a single criterion of classification the confidence limits (8) are identical with those given in my previous paper.

THE FREQUENCY DISTRIBUTION OF A GENERAL MATCHING PROBLEM

By T. N. E. Greville

Bureau of the Census

1. Introduction. This paper considers the matching of two decks of cards of arbitrary composition, and the complete frequency distribution of correct matchings is obtained, thus solving a problem proposed by Stevens.[1] It is also shown that the results can be interpreted in terms of a contingency table.

Generalizing a problem considered by Greenwood,[2] let us consider the matching of two decks of cards consisting of t distinct kinds, all the cards of each kind being identical. The first or "call" deck will be composed of i_1 cards of the first kind, i_2 of the second, etc., such that

$$i_1 + i_2 + i_3 + \cdots + i_t = n;$$

and the second or "target" deck will contain j_1 cards of the first kind, j_2 of the second, etc., such that

$$j_1 + j_2 + \cdots + j_t = n.$$

Any of the i's or j's may be zero. It is desired to calculate, for a given arrangement of the "call" deck, the number of possible arrangements of the "target" deck which will produce exactly r matchings between them ($r = 0, 1, 2, \cdots, n$). It is clear that these frequencies are independent of the arrangement of the call deck. For convenience the call deck may be thought of as arranged so that all the cards of the first kind come first, followed by all those of the second kind, and so on.

2. Formulae for the frequencies. Let us consider the number of arrangements of the target deck which will match the cards in the k_1th, k_2th, $\cdots$, k_sth positions in the call deck, regardless of whether or not matchings occur elsewhere. Let the cards in these s positions in the call deck consist of c_1 of the first kind, c_2 of the second, etc. Then:

$$c_1 + c_2 + \cdots + c_t = s.$$

The number of such arrangements of the target deck is

$$\frac{(n-s)!}{\prod_{h=1}^{t} (j_h - c_h)!}. \tag{1}$$

---

[1] W. L. Stevens, *Annals of Eugenics*, Vol. 8 (1937), pp. 238–244.

[2] J. A. Greenwood, *Annals of Math. Stat.*, Vol. 9 (1938), pp. 56–59.

For fixed values of the c's, the s specified positions may be selected in

$$(2) \qquad \prod_{h=1}^{t} \frac{i_h!}{c_h!(i_h - c_h)!}$$

ways.

Consider now the expression

$$(3) \qquad V_s = \sum \frac{(n-s)! \prod_{h=1}^{t} i_h!}{\prod_{h=1}^{t} c_h!(i_h - c_h)!(j_h - c_h)!}$$

obtained by summing the product of (1) and (2) over all sets of values of the numbers $c_1, c_2 \cdots, c_t$ satisfying the conditions:

$$0 \leqq c_h \leqq i_h, \qquad c_h \leqq j_h, \qquad \text{and} \qquad \sum_{h=1}^{t} c_h = s.$$

Let W_s denote the number of arrangements of the target deck which result in exactly s matchings. Then it is evident that V_s exceeds W_s, since the former includes those arrangements which give more than s matchings, and these, moreover, are counted more than once. Consider an arrangement which produces u matchings, where $u > s$. Such an arrangement will be counted once in V_s for every set of s matchings which can be selected from the total of u—that is ${}^{u}C_s$ times. In other words,

$$V_r = W_r + {}^{r+1}C_r W_{r+1} + {}^{r+2}C_r W_{r+2} + \cdots + {}^{n}C_r W_n.$$

It has been shown[3] that the solution of these equations is

$$(4) \qquad W_r = V_r - {}^{r+1}C_r V_{r+1} + {}^{r+2}C_r V_{r+2} - \cdots + (-1)^{n-r}\, {}^{n}C_r V_n.$$

3. Computation of the frequencies. Equations (3) and (4) apparently give the solution of the problem, but in practice the labor of carrying out the summation indicated in (3) would often be very great. However, (3) may be rewritten in the form

$$(5) \qquad V_s = \frac{(n-s)!}{\prod_{h=1}^{t} j_h!} H_s,$$

where

$$H_s = \sum \left\{ \prod_{h=1}^{t} \frac{i_h!\, j_h!}{c_h!\,(i_h - c_h)!\,(j_h - c_h)!} \right\}.$$

[3] H. GEIRINGER, *Annals of Math. Stat.*, Vol. 9 (1938), p. 262.

It will be seen that H_s is the coefficient of x^s in the product

$$\prod_{h=1}^{t} \left\{ \sum_{k=0}^{i'_h} \frac{i_h!\, j_h!\, x^k}{k!\,(i_h - k)!\,(j_h - k)!} \right\}, \tag{6}$$

where i'_h denotes the smaller of i_h and j_h. The factor $\prod_{h=1}^{t} j_h!$ was included in H_s in order to make the coefficients in the polynomials of (6) always integers.

Equation (4) may now be written in the form

$$W_r = \sum_{s=r}^{n} (-1)^{s-r}\, {}^sC_r \frac{(n-s)!}{\prod_{h=1}^{t} j_h!} H_s,$$

or

$$W_r = \frac{1}{r!} \sum_{s=r}^{n} \frac{(-1)^{s-r}}{(s-r)!} \frac{s!\,(n-s)!}{\prod_{h=1}^{t} j_h!} H_s, \tag{7}$$

a form which lends itself to actual computation.

4. Factorial moments. The factorial moments of the frequency distribution of the number of matchings are easy to compute. Let m_s denote the sth factorial moment, so that

$$m_s = \frac{\sum_{r=s}^{n} r^{(s)} W_r}{\sum_{r=0}^{n} W_r}. \tag{8}$$

Substituting from (4)

$$\sum_{r=s}^{n} r^{(s)} W_r = \sum_{r=s}^{n} \left\{ r^{(s)} \sum_{u=r}^{n} (-1)^{u-r}\, {}^uC_r V_u \right\}.$$

Reversing the order of summation and simplifying,

$$\sum_{r=s}^{n} r^{(s)} W_r = \sum_{u=s}^{n} \left\{ u^{(s)} V_u \sum_{r=s}^{u} (-1)^{u-r}\, {}^{u-s}C_{r-s} \right\} = s!\, V_s.$$

Hence,

$$V_0 = \sum_{r=0}^{n} W_r = \frac{n!}{\prod_{h=1}^{t} j_h!}, \tag{9}$$

and from (5) and (8),

$$m_s = \frac{H_s}{{}^nC_s}. \tag{10}$$

5. Mean and variance. From (6)

$$H_1 = \sum_{h=1}^{t} i_h j_h \tag{11}$$

and

$$H_2 = \tfrac{1}{2}\sum_{h=1}^{t} i_h(i_h - 1)j_h(j_h - 1) + \sum_{\substack{h,k=1\\ h\neq k}}^{t} i_h i_k j_h j_k. \tag{12}$$

Hence the mean number of matchings is

$$m_1 = \frac{\sum_{h=1}^{t} i_h j_h}{n}. \tag{13}$$

The variance μ_2 is

$$m_2 + m_1 - m_1^2 = \frac{1}{n^2(n-1)}\left[n\sum_{h=1}^{t} i_h(i_h - 1)j_h(j_h - 1) + 2n\sum_{\substack{h,k=1\\ h<k}}^{t} i_h i_k j_h j_k \right.$$
$$\left. + n(n-1)\sum_{h=1}^{t} i_h j_h - (n-1)\left(\sum_{h=1}^{t} i_h j_h\right)^2\right],$$

or

$$\mu_2 = \frac{1}{n^2(n-1)}\left\{\left(\sum_{h=1}^{t} i_h j_h\right)^2 - n\sum_{h=1}^{t} i_h j_h(i_h + j_h) + n^2\sum_{h=1}^{t} i_h j_h\right\}. \tag{14}$$

In the special case $j_1 = j_2 = \cdots = j_t = j$, these formulae become

$$M_1 = j, \qquad \mu_2 = \frac{j}{n(n-1)}\left(n^2 - \sum_{h=1}^{t} i_h^2\right).$$

These formulae have previously been given by Stevens,[4] and those for the special case also by Greenwood. The maximal conditions for the variance, given by Greenwood for this particular case, apparently can not be put in a simple form for the general case.

6. Unequal decks. Suppose the call deck contains m cards, $m < n$, and is to be matched with m cards selected from the target deck. It can be assumed without loss of generality that the first m cards in any arrangement of the target deck are the ones to be used. The formulae of this paper can be applied to this

[4] W. L. Stevens, *Annals of Eugenics*, loc. cit., *Psychol. Review*, Vol. 46 (1939), pp. 142–150.

more general problem by the expedient of imagining $n - m$ blank cards to be added at the end of the call deck and regarding these as an additional kind. It is thus apparent that formulae (13) and (14) apply without modification to this altered situation.

7. Application to contingency table. Stevens[5] has considered the distribution of entries in a contingency table with fixed marginal totals, and has pointed out that the problem of matching two decks of cards may be dealt with from that standpoint. A contingency table classifies data into n columns and m rows, and we may consider the row as indicating the kind of card which occupies a given position in the call deck, the columns having the same function with respect to the target deck. Stevens defines a quantity c as the sum of entries in a prescribed set of cells, subject to the condition that no two cells of the set are in the same row or column, and mentions as unsolved the problem of the exact sampling distribution of c.

We now have at our disposal the machinery for solving this problem. Following Stevens's notation, let $a_1, a_2, \cdots, a_m$ denote the fixed row totals and $b_1, b_2, \cdots, b_n$ the fixed column totals, while x_{rs} denotes the frequency of the cell in the rth row and the sth column. Then, let $c = \sum_{h=1}^{l} x_{r_h s_h}$, where l does not exceed either m or n. Imagine two decks of N cards $\left(N = \sum_{h=1}^{m} a_h = \sum_{h=1}^{n} b_h\right)$, the first containing a_1 cards of one kind, a_2 of another, etc., and the second containing b_1 cards of one kind, b_2 of another, etc. Moreover, let the r_hth kind in the first deck and the s_hth kind in the second deck be the same kind ($h = 1, 2, \cdots, l$), the other kinds being all different. Evidently c is the number of matchings between the two decks. Hence, the methods of this paper can be used to obtain the distribution of c. The formulae we have obtained agree with those for the expected value and variance of c given by Stevens.

ON METHODS OF SOLVING NORMAL EQUATIONS

By Paul G. Hoel

University of California, Los Angeles

There seems to be considerable disagreement concerning what is the most satisfactory method of solving a set of normal equations. Since such information as errors of estimate and significance of results is usually desired in addition to the solution, in its broader aspects the problem is one of deciding what is the most satisfactory method of calculating the inverse of a symmetric matrix.

For equations with several unknowns some compact systematic method of

[5] W. L. Stevens, *Annals of Eugenics*, loc. cit.