VARIANCE OF A GENERAL MATCHING PROBLEM*

By Joseph A. Greenwood

Let us match two decks of cards: (A) composed of t distinct groups of s identical symbols each, and (B) a target deck composed of i_1 symbols of the first kind, i_2 of the second, etc., such that

$$i_1 + i_2 + \cdots + i_t = st = n. \tag{1}$$

It is not necessary that all the i's be different from zero.

(a) *Forming the Relative Frequency Table.* The first part of the paper is concerned with forming a 2x2-way table showing the relative frequencies of hits and misses of all pairs of cards in the target deck. The notation $\overset{i}{0}$ indicates a miss at the ith card of the target deck, $\overset{i}{1}$ a hit. $\overset{i}{0} = j$ indicates a miss at the ith card, with the matching card identical to the jth target card.

Case I. *ith and jth target cards the same symbol.*

i	j	Theoretical freq.	Weighted freq.		
If 0 then —[1]	0	$n - s - 1$	$(t - 1)(n - s - 1)$	2.1	
0	1	s	$(t - 1)\, s = n - s$	2.2	(2)
1	0	$n - s$	$n - s$	2.3	
1	1	$s - 1$	$s - 1$	2.4	
			Total $= t(n - 1)$		

But $\overset{i}{0}$ occurs in $(t - 1)/t$ of the events. Thus we must weight 2.1 and 2.2 with a factor $(t - 1)$, giving the last column in (2).

Case II. *ith and jth target cards different*

i	j	Theoretical freq.	Weighted freq.		
If $0 = j$ then —	0	$n - s$	$n - s$	3.1	
$0 = j$	1	$s - 1$	$s - 1$	3.2	
$0 \neq j$	0	$n - s - 1$	$(n - s - 1)(t - 2)$	3.3	(3)
$0 \neq j$	1	s	$s(t - 2)$	3.4	
1	0	$n - s - 1$	$n - s - 1$	3.5	
1	1	s	s	3.6	
			Total $= t(n - 1)$		

* Presented to the American Mathematical Society, September 9, 1937.

[1] Read, 'then out of $n - 1$ times'.

But $\overset{i}{0} = j$ occurs in $1/(t - 1)$ of all events $\overset{i}{0}$, and $\overset{i}{1}$ occurs in $1/t$ of all events $\overset{i}{1} + \overset{i}{0}$. Therefore entries 3.3 and 3.4 must be weighted with the factor $(t - 2)$, and then entries 3.1, 3.2, 3.3 and 3.4 must be weighted with the factor $(t - 1)$. It is important that the totals of the two parts to be weighted be equal before the weighting factors are applied. This gives rise to the last column in table (3).

Now the number of ways the ith card can be like the jth card of the target deck is[2]

$$\alpha_1 = \sum_{j=1}^{t} \binom{i_j}{2}$$

The number of ways they can be unequal is

$$\alpha_2 = \sum_{u<v}^{1,\cdots,t} (i_u i_v) = \binom{n}{2} - \alpha_1. \tag{4}$$

Since the totals of the last columns of the two tables are equal we weight the entries of their last columns with α_1 and α_2, respectively. So, combining 3.1, 3.3 and 3.2, 3.4 we form α_1 times (2) + α_2 times (3) to give the new table

i	j	Relative frequencies
0	0	$(n - s - 1)(t - 1)\alpha_1 + [(t - 1)(n - s - 1) + 1]\alpha_2$
0	1	$(n - s)\alpha_1 + (n - s - 1)\alpha_2$
1	0	$(n - s)\alpha_1 + (n - s - 1)\alpha_2$
1	1	$(s - 1)\alpha_1 + s\alpha_2$..

(5)

Now using the entries from (5) form the 2x2-way table

		Total
$(t - 1)(n - s - 1)\alpha_1 + [(t - 1)(n - s - 1) + 1]\,\alpha_2$	$(n - s)\alpha_1 + (n - s - 1)\alpha_2$	$(tn - n - t + 1)(\alpha_1 + \alpha_2)$
$(n - s)\alpha_1 + (n - s - 1)\alpha_2$	$(s - 1)\alpha_1 + s\alpha_2$	$(n - 1)(\alpha_1 + \alpha_2)$
$(tn - n - t + 1)(\alpha_1 + \alpha_2)$	$(n - 1)(\alpha_1 + \alpha_2)$	$t(n - 1)(\alpha_1 + \alpha_2)$

(6)

[2] If $i_v < 2$ define $\binom{i_v}{2} = 0$.

(b) *Obtaining the Correlation, Variance and Maximal Conditions.* Substituting from table (6) into the formulas given by Yule[3] for δ and the coefficient of correlation r, we obtain the average correlation

$$r = \frac{\alpha_2 + (1 - t)\alpha_1}{(tn - n - t + 1)(\alpha_1 + \alpha_2)} = \frac{\binom{n}{2} - t\alpha_1}{\binom{n}{2}(tn - n - t + 1)}$$
$$= \frac{t\alpha_2 + (1 - t)\binom{n}{2}}{\binom{n}{2}(tn - n - t + 1)} \tag{7}$$

by (4).

We now give a proof that r is a maximum when $i_j = s$ $(j = 1, \cdots, t)$. From (7) it is sufficient to show that under the same conditions α_2 is a maximum.

Let $i_j = s + \delta_j$, then

$$\sum^{1,\cdots,t} \delta_j = 0 \qquad \text{by (1).} \tag{8}$$

$$\alpha_2 = \sum_{u<v}^{1,\cdots,t} (s + \delta_u)(s + \delta_v) = \binom{t}{2} s^2 + \sum_{u<v}^{1,\cdots,t} \delta_u \delta_v \quad \text{by (8).}$$

Assume some $\delta_u \neq 0$ and

$$\sum_{u<v}^{1,\cdots,t} \delta_u \delta_v \geqq 0. \tag{9}$$

Add

$$\sum^{1,\cdots,t} \delta_u^2 + \sum_{u<v}^{1,\cdots,t} \delta_u \delta_v$$

to both sides of (9). Then

$$\left(\sum^{1,\cdots,t} \delta_u\right)^2 \geqq \sum^{1,\cdots,t} \delta_u^2 + \sum_{u<v}^{1,\cdots,t} \delta_u \delta_v \tag{10}$$

or $0 \geqq$ a positive number. This necessarily implies the desired result.

---

[3] Yule, G. U. *An Introduction to the Theory of Statistics*, London: Griffin and Co., 1927, pp. 216–217. The table can be symbolized with

			Total
	a_1	a_2	a_3
	b_1	b_2	b_3
Total	c_1	c_2	c_3

$\delta = b_2 - (c_2 b_3 / c_3)$

He then gives $r = \delta c_3 / \sqrt{c_1 c_2 a_3 b_3}$, the correlation coefficient.

Yule[4] gives an expression for the variance in a situation which includes the present problem as a special case, to be

$$\sigma^2 = npq[1 + r(n - 1)], \tag{11}$$

where r is the average correlation between all pairs of variables. Substituting our result (7) in (11) with $p = 1/t$ gives the desired variance.

It is interesting to note that when $i_j = s$, $(j = 1, \cdots, t)$ r reduces to $1/(n - 1)^2$ giving

$$\sigma^2 = \frac{n^2(t - 1)}{t^2(n - 1)} = \frac{n}{n - 1}\sigma_b^2$$

where σ_b^2 is the variance of the binomial case.[5]

DUKE UNIVERSITY.

---

[4] Op. cit., p. 286.

[5] Concerning this special case see also Bartlett, M. S. *Properties of sufficiency and statistical tests.* Proc. Royal Soc. A. 1937, CLX, 268–282.

Olds, E. G. *A moment-generating function useful in certain matching problems.* Abstract No. 428, Bull. Amer. Math. Soc. 1937, XLIII, 779.