CERTAIN SYMMETRICAL PROPERTIES OF UNBIASED ESTIMATES OF VARIANCE AND COVARIANCE*

By K. R. Nair

1. Introductory

For a random sample of n observations $x_1, \ldots x_n$ drawn from *any* statistical population having standard deviation σ, an unbiased estimate of the variance σ^2 is obtained by dividing sum of squares of the n deviations from the sample mean $\bar{x}$ by $(n-1)$ instead of by n. Denoting this estimate by S^2, we have

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/(n-1). \tag{1}$$

It can easily be shown (see *e.g.*, Kendall, 1945) that the *population* variance σ^2 is equal to *half* the mean square of all possible variate differences. This property holds good for the *sample* variance S^2 as well, because of the identity

$$n \sum_{i=1}^{n} (x_i - \bar{x})^2 \equiv \sum_{i=1}^{n} \sum_{j>i} (x_i - x_j)^2 \tag{2}$$

whence

$$S^2 = \sum_{i=1}^{n} \sum_{j>i} (x_i - x_j)^2/\{n(n-1)\} \tag{3}$$

Similarly, the covariance of a bivariate sample $(x_1, y_1) \ldots (x_n, y_n)$ can be written

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})/(n-1) \equiv \sum_{i=1}^{n} \sum_{j>i} (x_i - x_j)(y_i - y_j)/\{n(n-1)\} \tag{4}$$

The correlation coefficient for x and y is

$$r = \frac{\sum_{1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{1}^{n} (x_i - \bar{x})^2 \sum_{1}^{n} (y_i - \bar{y})^2}} = \frac{\Sigma\Sigma (x_i - x_j)(y_i - y_j)}{\sqrt{\Sigma\Sigma (x_i - x_j)^2 \Sigma\Sigma (y_i - y_j)^2}} \tag{5}$$

Since any identity that holds good for variance will give a corresponding identity for the covariance we shall in the rest of the paper discuss only identities for variance to save space.

Cochran (1946) recently made use of form (3) for the variance to study the relative accuracy of systematic and random sampling in field experimental work.

---

* Part of a thesis approved for the Ph.D. degree of London University.

The identity (2) is a special case of a more general identity

$$\sum_{i=1}^{n} (p_i) \sum_{i=1}^{n} \{p_i (x_i - \bar{x})^2\} \equiv \sum_{i=1}^{n} \sum_{j>i} p_i p_j (x_i - x_j)^2, \tag{6}$$

where p_i is any real number subject to the condition that $\sum_1^n (p_i) \neq 0$; and $\bar{x} = \sum_1^n (p_i \cdot x_i)/\sum_1^n (p_i)$.

There are two applications of (6) which are of some interest. Firstly, if we have a grouped frequency distribution with f_i observations in the ith class interval centred at x_i and if $N = \sum_1^n (f_i)$ the unbiased estimate of the variance of this distribution obtained by using the degrees of freedom, $(N - 1)$ can be written

$$S^2 = \frac{1}{N(N-1)} \sum_{i=1}^{n} \sum_{j>i} f_i f_j (x_i - x_j)^2 \tag{7}$$

whereas the second moment will be

$$m_2 = \frac{1}{N^2} \sum_{i=1}^{n} \sum_{j>i} f_i f_j (x_i - x_j)^2 \tag{8}$$

It is interesting to compare (7) and (8) with the two formulæ for Gini's mean difference given by Kendall (1945). Adapting his summation limits to fit in with ours, we have either

$$\triangle_1 = \frac{2}{N(N-1)} \sum_{i=1}^{n} \sum_{j>i} f_i f_j \mid x_i - x_j \mid \tag{9}$$

or

$$\triangle_1 = \frac{2}{N^2} \sum_{i=1}^{n} \sum_{j \geqslant i} f_i f_j \mid x_i - x_j \mid \tag{10}$$

Kendall calls (9) the mean difference 'without repetition' and (10) the mean difference 'with repetition'.

The point that is usually put forward in favour of Gini's mean difference is that it is independent of the centre of location, being dependent only on the spread of the variate values among themselves. It is now clear that this property is shared by the second moment and variance of the sample.

Another application of (6) arises in the analysis of variance for k samples discussed elsewhere by the author (1943). If x_{il} is the lth observation in the ith sample ($i = 1, \ldots k$; $l = 1, \ldots n_i$), $\bar{x}_i$ mean of the ith sample, $\bar{x}$ the general mean and $N = \sum_1^k n_i$, the variance 'between samples' is

$$V_b = \sum_1^k n_i (\bar{x}_i - \bar{x})^2/(k-1) \tag{11}$$

and variance 'within samples' is

$$V_w = \sum_{i=1}^{k} \sum_{l=1}^{n_i} (x_{il} - \bar{x}_i)^2/(N - k). \tag{12}$$

It is assumed that the k samples come from normal populations having same unknown standard deviation σ.

To test the hypothesis that the ith and jth population means are the same, we use a t-test

$$t_{ij} = (\bar{x}_i - \bar{x}_j) \Big/ \sqrt{V_w \left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \tag{13}$$

with $(N - k)$ degrees of freedom.

Using (6) it can be shown that the variance ratio used to test whether all the k samples come from the same normal population (that is, whether the sample means significantly differ among themselves), is a weighted average of the ${}^kC_2\ t_{ij}^2$ values. Thus

$$V_b/V_w \equiv \sum_{i=1}^{k} \sum_{j>i} (n_i + n_j)\ t_{ij}^2 \Big/ \sum_{i=1}^{k} \sum_{j>i} (n_i + n_j) \tag{14}$$

I do not know whether any special significance can be attached to (14). My own object in proving it in the 1943 paper was this. Iyer (1937) took the unweighted average of the ${}^kC_2\ t_{ij}^2$'s and finding it different from the variance ratio V_b/V_w felt that the former would be the proper criterion for discriminating between the sample means. By proving the property (14) it was possible to show that no such test as proposed by Iyer was necessary.

A two-dimensional extension of the identity (2) was given by Husain (1943). Let the n observations be arranged in a two-way table having p rows and q columns, so that $n = pq$. Let x_{ij} be the observation in the ith row and jth column. Let $\bar{x}_{i.}$, $\bar{x}_{.j}$ be the mean of x in the ith row, jth column respectively and $\bar{x}_{..}$ the general mean. The following identity can be proved

$$pq \sum_{i=1}^{p} \sum_{j=1}^{q} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2$$

$$\equiv \sum_{i=1}^{p} \sum_{i'>i} \sum_{j=1}^{q} \sum_{j'>j} (x_{ij} + x_{i'j'} - x_{ij'} - x_{i'j})^2 \tag{15}$$

Each term on the right-hand side of (15) is the square of a diagonal difference and there are ${}^pC_2 . {}^qC_2$ such terms. The variance due to interaction between rows and columns is therefore one-fourth the mean square of all diagonal differences. The estimate of interaction variance obtained by using the right-hand side of (15) is very much similar to the form in which "Student" (1923) first derived the 'error' variance for chess-board trials.

Similar identities can be derived for higher order interactions, but they become very complex. The purpose of this paper is to show that the identities (2) and (15) are only special cases of two general classes of identities associated with the universe of sample permutations. The general identity to which (2) belongs can be applied to a problem of Pitman (1937).

2. Identities Associated with the Universe of Sample Permutations

For convenience, we shall assume that $x_1, \ldots x_n$ used in (2) are in ascending order of magnitude. Let $l_1, \ldots l_n$ be any real numbers satisfying the conditions

$$\left.\begin{array}{l} l_1 + \ldots + l_n = 0 \\ l_1^2 + \ldots + l_n^2 = 1 \\ l_1 \leqslant \ldots \leqslant l_n \end{array}\right\} \qquad (16)$$

If we write the n^2 cross-products of l and x in the form of the square given in Fig. 1 and pick up n cells, one from each row and

	l_1	.	.	.	.	l_n
x_1	$l_1 x_1$					$l_n x_1$
.						
.						
x_n	$l_1 x_n$					$l_n x_n$

Fig. 1

each column and add up the lx products for these cells, we get a linear function with n terms

$$u_k = \sum_{i,\, j = 1}^{n} (l_i\, x_j) \quad (i \text{ and } j \text{ take each of the values } 1, \ldots n \text{ once and once only}) \qquad (17)$$

which we shall call a general linear *contrast* of $x_1, \ldots x_n$ for the given $_1, \ldots l_n$. The word 'contrast' is used because u_k can be expressed as a linear function of differences $x_i - x_j$ in virtue of the fact that $\sum\limits_1^n (l_i) = 0$.

The total number of contrasts of the type (17) is $n!$. They form a population which may be called a universe of sample permutations for linear contrasts of $x_1, \ldots x_n$ associated with a given $l_1 \ldots l_n$. It can easily be shown that the largest contrast in this universe is

$\overset{n}{\underset{1}{\Sigma}}(l_i x_i)$. Since the l's of (16) can be chosen in an infinite number of ways, there is an infinite number of universes each consisting of $n!$ contrasts. When some of the l's are equal, the total number of distinct values for u_k will be only a fraction of $n!$, say $(n!)/m$ as each distinct u_k repeats m times. The value of m can be calculated from the usual rules of permutation. At any rate, the moments of the universe of contrasts will be independent of m, and we may proceed to calculate the first two moments, assuming that there are $n!$ values of u_k.

It is easily seen that

$$\overset{n!}{\underset{k=1}{S}} (u_k) = 0 \tag{18}$$

$$\overset{n!}{\underset{k=1}{S}} (u_k^2) = (n-1)! \overset{n}{\underset{1}{\Sigma}} (l_i^2) \overset{n}{\underset{1}{\Sigma}} (x_i^2) + 4\,(n-2)! \overset{n}{\underset{i=1}{\Sigma}} \underset{j>i}{\Sigma} (l_i l_j) \overset{n}{\underset{i=1}{\Sigma}} \underset{j>i}{\Sigma} (x_i x_j)$$

$$= \frac{n!}{(n-1)} \overset{n}{\underset{1}{\Sigma}} (x_i - \bar{x})^2 \tag{19}$$

from which it follows that

$$\mu_1'(u_k) = 0 \tag{20}$$

$$\mu_2(u_k) = \overset{n}{\underset{1}{\Sigma}} (x_i - \bar{x})^2/(n-1) = S^2. \tag{21}$$

That μ_1' will be zero could easily have been anticipated. But the second moment of the universe of u_k becoming independent of $l_1, \ldots l_n$ and equal to the variance of the sample $x_1, \ldots x_n$ was not such an obvious result.

We shall consider two special universes with different sets of values for $l_1, \ldots l_n$.

(i) Let

$$l_1 = \ldots = l_{n-1} = -\frac{1}{\sqrt{n(n-1)}};\ l_n = \sqrt{\frac{(n-1)}{n}}. \tag{22}$$

Since $(n-1)$ of the l's are equal, the number of distinct members in the universe is $\dfrac{n!}{(n-1)!}$ or n. They are

$$u_1 = \frac{1}{\sqrt{n(n-1)}}\Big[(n-1)\,x_n - x_{n-1} \ldots - x_1\Big] = \sqrt{\frac{n}{(n-1)}}\,(x_n - \bar{x})$$

$$u_2 = \frac{1}{\sqrt{n(n-1)}}\Big[-x_n + (n-1)\,x_{n-1} - \ldots - x_1\Big]$$

$$= \sqrt{\frac{n}{(n-1)}}\,(x_{n-1} - \bar{x})$$

$$\vdots$$

$$u_n = \frac{1}{\sqrt{n(n-1)}}\Big[-x_n - x_{n-1} \ldots + (n-1)\,x_1\Big]$$

$$= \sqrt{\frac{n}{(n-1)}}\,(x_1 - \bar{x}) \tag{23}$$

Substituting in (21), we get the second moment or the mean square of $u_1, \ldots u_n$

$$\frac{1}{n}\sum_{k=1}^{n}(u_k^2) = S^2. \tag{24}$$

which we obviously know to be true from the values of $u_1, \ldots u_n$ given in (23).

What (24) points out is that instead of saying that an unbiased estimate of variance is obtained by dividing the sum of squares of n deviations $x_i - \bar{x}$ by $(n - 1)$, we could say that it is the *mean square* of the n *contrasts* $\sqrt{\frac{n}{(n-1)}}.(x_i - \bar{x})$.

The largest linear contrast $\sum_1^n (l_i x_i)$ in this universe of n is u_1.

(ii) As another example, we shall derive (2) and (3) as a special case of (21).

Let

$$l_1 = -\frac{1}{\sqrt{2}},\ l_2 = \ldots = l_{n-1} = 0,\ l_n = \frac{1}{\sqrt{2}}. \tag{25}$$

Since $(n - 2)$ of the l's are equal, the number of distinct linear contrasts in the universe is $\frac{n!}{(n-2)!}$ or $n(n - 1)$. Each of them is of the type $\frac{1}{\sqrt{2}}(x_i - x_j)$. Using (21), we get

$$\frac{1}{n(n-1)}\sum_{\substack{i=1 \\ (i \neq j)}}^{n}\sum_{j=1}^{n}\tfrac{1}{2}(x_i - x_j)^2 = S^2 \tag{26}$$

Since $x_i - x_j$ and $x_j - x_i$ have the same square, (25) may be re-written

$$\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j>1}(x_i - x_j)^2 = S^2 \tag{27}$$

which is same as the result (3) proved directly from the identity (2).

The largest contrast $\sum_1^n (l_i x_i)$ in this universe is $\frac{1}{\sqrt{2}}(x_n - x_1)$ or $\frac{1}{\sqrt{2}}$ (range).

Coming now to the generalisation of the identity (15) for interaction variance, we shall assume a set of pq quantities l_{hk} satisfying the conditions

$$\left.\begin{aligned} \sum_{k=1}^{q}(l_{hk}) = \sum_{h=1}^{p}(l_{hk}) = 0 \\ \sum_{h=1}^{p}\sum_{k=1}^{q}(l_{hk}^2) = 1 \end{aligned}\right\} \tag{28}$$

The values l_{hk} and x_{ij} can be written side by side in the form of two $p \times q$ tables as shown in Fig. 2.

	1	.	. .	k	. .	q	Mean
1	l_{11}			l_{1k}		l_{1q}	0
.							
h	l_{h1}			l_{hk}		l_{hq}	0
.							
p	l_{p1}			l_{pk}		l_{pq}	0
Mean	0			0		0	0

	1	. .	j	. .	q	Mean
1	x_{11}		x_{1j}		x_{1q}	$\bar{x}_{1\cdot}$
.						
i	x_{i1}		x_{ij}		x_{iq}	$\bar{x}_{i\cdot}$
.						
p	x_{p1}		x_{pj}		x_{pq}	$\bar{x}_{p\cdot}$
Mean	$\bar{x}_{\cdot 1}$		$\bar{x}_{\cdot j}$		$\bar{x}_{\cdot q}$	$\bar{x}_{\cdot\cdot}$

FIG. 2

If we superimpose the x-table on the l-table, multiply the l and x of each cell and add up the products for all the pq cells we get a general linear contrast $\Sigma\,(l_{hk}x_{ij})$ which is a function only of diagonal differences of the type $(x_{ij} + x_{i'j'} - x_{ij'} - x_{i'j})$ and hence may be called the general 'interaction contrast'. By interchanging the rows among themselves and the columns among themselves of either the x-table or the l-table, we will get $p!\,q!$ different composite squares giving us a universe of $p!\,q!$ interaction contrasts. If some of the rows or columns are identical among themselves the number of distinct composite squares will be less than $p!\,q!$. This will not, however, affect the moments of the universe of contrasts and hence for calculating them we shall assume there are $p!\,q!$ distinct contrasts. It can be shown that the first moment vanishes and that the second moment or the mean square of the $p!\,q!$ interaction contrasts of the type $\Sigma\, l_{hk}\, x_{ij}$ is equal to the interaction variance

$$\frac{1}{(p-1)(q-1)} \sum_{i=1}^{p} \sum_{j=1}^{q} (x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x}_{\cdot\cdot})^2 \qquad (29)$$

As an example, let us consider the following table of values for l_{hk}. There is a common factor $\sqrt{\dfrac{(p-1)(q-1)}{pq}}$ in the l's and so for convenience it has been written outside the table.

1	$-\frac{1}{(q-1)}$		$-\frac{1}{(q-1)}$
$-\frac{1}{(p-1)}$	$\frac{1}{(p-1)(q-1)}$		$\frac{1}{(p-1)(q-1)}$
⋮			
$-\frac{1}{(p-1)}$	$\frac{1}{(p-1)(q-1)}$		$\frac{1}{(p-1)(q-1)}$

$\times \sqrt{\frac{(p-1)(q-1)}{pq}}$

Since $(p-1)$ of the rows and $(q-1)$ of the columns of the above table are identical, the total number of interaction contrasts is pq. Taking the first of them, got by superimposing the x-table of Fig. 2 on this l-table, we get

$$u_{11} = \sqrt{\frac{(p-1)(q-1)}{pq}} \left[x_{11.} - \frac{1}{(q-1)} \sum_{j=2}^{q} x_{1i} - \frac{1}{(p-1)} \sum_{i=2}^{p} x_{i1} + \frac{1}{(p-1)(q-1)} \sum_{i=2}^{p} \sum_{i=2}^{q} x_{ij} \right]$$

$$= \sqrt{\frac{pq}{(p-1)(q-1)}} (x_{11} - \bar{x}_{1.} - \bar{x}_{.1} + \bar{x}_{..}) \tag{30}$$

A typical contrast can therefore be written

$$u_{ij} = \sqrt{\frac{pq}{(p-1)(q-1)}} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}) \begin{pmatrix} i = 1, \ldots p \\ j = 1, \ldots q \end{pmatrix} \tag{31}$$

The usual expression for interaction variance (29) can therefore be derived as a special case from the general universe of interaction contrasts.

As another example, let us assume that

$$\left. \begin{aligned} & l_{11} = l_{22} = -l_{21} = -l_{12} = \tfrac{1}{2}; \\ \text{and} \quad & l_{hk} = 0 \text{ when } h = 3, \ldots p \text{ and } k = 3, \ldots q. \end{aligned} \right\} \tag{32}$$

The interaction contrasts now take the form

$$u_k = \tfrac{1}{2} (x_{ij} + x_{i'j'} - x_{ij'} - x_{i'j}) \tag{33}$$

and their total number is $\frac{1}{2} pq (p-1)(q-1)$. They form $\frac{1}{4} pq (p-1)(q-1)$ couplets which have same magnitude but are of opposite sign. The total number of distinct u_k^2 is therefore $\frac{1}{4} pq (p-1)(q-1)$ and their mean square

$$\frac{1}{pq(p-1)(q-1)} \sum_{i=1}^{p} \sum_{i'>i} \sum_{j=1}^{q} \sum_{j'>j} (x_{ij} + x_{i'j'} - x_{ij'} - x_{i'j})^2 \tag{34}$$

should be equal to the interaction variance (29), thereby leading us to the identity (15).

3. Analogy with Pitman's Problem

If $(x_1, y_1) \ldots (x_n, y_n)$ is a sample of paired observations of two variates x and y which are independently distributed, Pitman (1937) devised a test for the significance of the correlation coefficient r_{xy} based on its distribution in the universe of sample permutations generated by pairing each x with every y in turn once. For this universe of $n!$ values of r, he found that, whatever be the values of x and y

$$E(r) = 0\,;\; E(r^2) = 1/(n-1) \tag{35}$$

and that under certain conditions depending on x and y

$$E(r^3) \to 0\,;\; E(r^4) \to \frac{3}{(n-1)(n+1)}. \tag{36}$$

A proof that $E(r^2) = 1/(n-1)$ follows from (19) by substituting $l_i = \frac{y_i - \bar{y}}{\sqrt{\Sigma (y_i - \bar{y})^2}}$. The conditions $l_i \leqslant l_{i+1}$ and $x_i \leqslant x_{i+1}$ are not necessary for this proof.

4. Upper Limit of the Largest Contrast

The conditions $l_i \leqslant l_{i+1}$ and $x_i \leqslant x_{i+1}$ were imposed only for the reason that the largest of the $n!$ contrasts could then conveniently be written $\sum_1^n (l_i x_i)$. From the analogy with Pitman's problem we find that for given l and x, the upper limit of the largest contrast is less than $S\sqrt{(n-1)}$, becoming equal to $S\sqrt{(n-1)}$ only when there is perfect correlation between l and x, or when l_i is proportional to $x_i - \bar{x}$, that is

$$l_i = \frac{x_i - \bar{x}}{\sqrt{\Sigma (x_i - \bar{x})^2}} \tag{37}$$

The largest contrast in that case reduces simply to

$$\sum_1^n (l_i x_i) = \frac{\sum_1^n (x_i - \bar{x})^2}{\sqrt{\sum_1^n (x_i - \bar{x})^2}} = S\sqrt{(n-1)} \tag{38}$$

These results are of some use when $l_1, \ldots l_n$ are given and the sample variance S^2 is fixed, but not the individual observations $x_1, \ldots x_n$. Thus, if l's have the values given in (22), the largest contrast is u_1 of (23) or $\sqrt{\frac{n}{(n-1)}}$ times the extreme deviate $(x_n - \bar{x})$. The problem is to find the upper limit of the extreme deviate when the sample variance S^2 is fixed. The upper limit is obviously $(n - 1)\, S/\sqrt{n}$, and we can also obtain the additional information that this happens when $x_1, \ldots x_n$ in the particular sample which provides this upper limit satisfies the equations

$$\frac{x_1 - \bar{x}}{-1} = \frac{x_2 - \bar{x}}{-1} = \ldots = \frac{x_n - \bar{x}}{(n-1)}, \tag{39}$$

that is, when $x_1 = \ldots = x_{n-1}$.

Pearson and Chandrasekar (1936) made use of this property of the extreme deviate while examining a test criterion for the rejection of outlying observations. Tang (1938) applied it to problems in analysis of variance of designed experiments.

If l's have the value given in (25), the largest contrast becomes $\frac{1}{\sqrt{2}}$ (range). The upper limit of the range in samples of n with a fixed S is therefore $\sqrt{2(n-1)}\,.S$ which it attains when $x_1, \ldots x_n$ satisfy the conditions

$$\left.\begin{array}{c} -(x_1 - \bar{x}) = (x_n - \bar{x}) \\ (x_2 - \bar{x}) = \ldots = (x_{n-1} - \bar{x}) = 0 \end{array}\right\} \tag{40}$$

In other words the range reaches its upper limit $\sqrt{2(n-1)}\,.S$ when the intermediate observations $x_2, \ldots x_{n-1}$ are all concentrated at the mid-point $\frac{1}{2}(x_1 + x_n)$ of the range.

Lastly, let us take

$$l_i = \frac{2\sqrt{3}}{\sqrt{n(n^2-1)}}\left(i - \frac{n+1}{2}\right) \tag{41}$$

The largest contrast $\sum_1^n (l_i\, x_i)$ becomes proportional to Gini's mean difference g for ungrouped data, *viz.*,

$$g = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j>i} |\, x_i - x_j \,|$$
$$= \frac{4}{n(n-1)} \sum_{i=1}^{n} \left\{\left(i - \frac{n+1}{2}\right) x_i\right\} \tag{42}$$

The upper limit of g in samples of n with a fixed S is therefore $\sqrt{\frac{4(n+1)}{3n}}\,.\,S$ and this limit is reached when $x_i - \bar{x}$ is proportional to $i - \bar{i}$ or when the x's are at equal intervals. It is interesting to note that, in general, g is proportional to the linear regression of the *magnitude* of x on the *rank* of x.

References

Cochran, W. G. .. "Relative accuracy of systematic and stratified random samples for a certain class of populations," *Ann. Math. Stat.*, 1946, **17**, 164.

Husain, Q. M. .. "A note on interaction," *Sankhya*, 1943, **6**, 321.

Iyer, P. V. K. .. *Proc. Ind. Acad. Sci.*, 1937, **5**, 528.

Kendall, M. G. .. *The Advanced Theory of Statistics*, Vol. I, 1945, second edition, Charles Griffin & Co., Ltd. (see Sections 2.19 and 2.20).

Nair, K. R. .. "A note on the problem of *k* samples," *Curr. Sci.*, 1943, **12**, 112.

Pearson, E. S., and Chandrasekar, C. "The efficiency of statistical tools and a criterion for the rejection of outlying observations," *Biometrika*, 1936, **28**, 308.

Pitman, E. J. G. .. "Significance tests which may be applied to samples from any populations. II. The correlation coefficient test," *Suppl. J. R. Stat. Soc.*, 1937, **4**, 225.

"Student" .. "On testing varieties of cereals," *Biometrika*, 1923, **15**, 271.

Tang, P. C. .. "The power function of the analysis of variance tests with tables and illustrations for their use," *Stat. Res. Memoirs*, 1938, **2**, 126.