LETTERS TO THE EDITOR

A NOTE ON THE PROBLEM OF *k* SAMPLES

A familiar problem in analytic statistics is to test the hypothesis whether k samples of sizes $n_1, n_2, \cdots, n_k$ have been drawn at random from the same unknown Normal Universe. In general, the k samples could have come from k different Normal Universes with means $\mu_1, \mu_2, \cdots, \mu_k$ and standard deviations $\sigma_1, \sigma_2, \cdots, \sigma_k$. The most common hypothesis tested is whether, given that $\sigma_1 = \sigma_2 = \cdots = \sigma_k = \sigma$, we can infer that $\mu_1 = \mu_2 = \cdots = \mu_k = \mu$, where σ and μ are unknown. The 'statistic' which has been considered appropriate for this problem is Fisher's Ratio of Variances, say V_b/V_w where

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

$$V_w = \sum_{i:1}^{k} \sum_{j:1}^{n_i} (x_{ij} - \bar{x}_i)^2 / (N-k);$$

$$N = \sum_1^k (n_i) \text{ and } \bar{x} = \sum_1^k (n_i \bar{x}_i) / N.$$

When there are only two samples ($k = 2$) we have

$$V_b / V_w = (\bar{x}_1 - \bar{x}_2)^2 / V_w \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

which is distributed like "Student's" t^2 with $n_1 + n_2 - 2$ degrees of freedom. Fisher has also shown that when $k > 2$, we may compare the differences between any two sample means, by means of the same t-test, calculating V_w from within all the k samples, with $N-k$ degrees of freedom. Thus

$$t^2_{ij} = (\bar{x}_i = \bar{x}_j)^2 / V_w \left(\frac{1}{n_i} + \frac{1}{n_j}\right).$$

The total number of such tests of pair comparisons is $k(k-1)/2$. Thus the single hypothesis $\mu_1 = \mu_2 = \cdots \mu_k$ is being broken up into $k(k-1)/2$ separate tests involving the hypothesis $\mu_i = \mu_j$ for the particular pair of ith and jth samples.

The following identity is easy to prove

$$\Sigma n_i (\bar{x}_i - \bar{x})^2 \equiv \Sigma \Sigma n_i n_j (\bar{x}_i - \bar{x}_j)^2 / N.$$

Therefore

$$V_b/V_w = \Sigma\Sigma n_i n_j (\bar{x}_i - \bar{x}_j)^2 / N (k-1) V_w$$
$$= \Sigma\Sigma (n_i + n_j) t_{ij}^2 / \Sigma\Sigma (n_i + n_j).$$

The ratio of variances is therefore only a weighted mean of the $k(k-1)/2$ different values of t^2. This result brings out clearly the connection between the two tests of significance.

It may be recalled that a paper by P. V. Krishna Iyer[1] and also certain notes he[2] and the author[3,4] published in this journal some time ago dealt with the same problem of k samples. Mr. Krishna Iyer took the unweighted mean of all the values of t^2 and finding it different from the ratio of variances felt that the former would be the proper criterion for discriminating between the sample means. The form he obtained for the distribution of the unweighted mean of all the values of t^2 was found to be erroneous. In fact the true distribution does not come out in a suitable form for exact tests of significance. The relationship between the ratio of variances and the weighted mean of the different values of t^2, established above would indicate that there is no point in trying to develop a test based on the unweighted mean.

Statistical Laboratory, K. R. NAIR.
Presidency College,
Calcutta,
March 10, 1943.

1. P. V. Krishna Iyer, *Proc. Ind. Acad. Sci.*, 1937, **5**, 528. 2. —, *Curr. Sci.*, 1938, **6**, 392. 3. K. R. Nair, *Ibid.*, 1937, **6**, 290. 4. —, *Ibid.*, 1938, **7**, 21.

AN EXTENSION OF THE NEWTONIAN RELATION

$$\sum_{r=0}^{w-1} (-)^r a_r s_{w-r} + (-)^w w a_w = 0^*$$

THE relations given below may be considered as extensions of the Newtonian relation mentioned above :—

$$G(pq) = \sum_{m=1}^{p-1}\sum_{m=1}^{q-1} (-)^{s+m+2} a_{s+m} G(p-m, q-s) + q\sum_{m=1}^{p-1} (-)^{q+m+2} a_{q+m} G(p-m) + p\sum_{s=1}^{q-1} (-)^{p+s+2} a_{p+s} G(q-s) + \overline{p+q}\,(-)^{p+q+2} a_{p+q}.$$

$$G(p\,q\,r) = \sum_{t=1}^{p-1}\sum_{m=1}^{q-1}\sum_{s=1}^{r-1} (-)^{t+m+s+3} a_{t+m+s}.\ G(p-t, q-m, r-s) + r\sum_{t=1}^{p-1}\sum_{m=1}^{q-1} (-)^{r+m+t+3} a_{t+m+r}.\ G(p-t, q-m) + q\sum_{t=1}^{p-1}\sum_{s=1}^{r-1} (-)^{t+q+s+3} a_{t+q+s}.\ G(p-t, r-s) + p\sum_{m=1}^{q-1}\sum_{s=1}^{r-1} (-)^{p+m+s+3} a_{p+m+s}.\ G(q-m, r-s) + \overline{p+q}\sum_{s=1}^{r-1} (-)^{p+q+s+3} a_{p+q+s} G(r-s) + \overline{q+r}\sum_{t=1}^{p-1} (-)^{t+q+r+3} a_{t+q+r} G(p-t) + \overline{p+r}\sum_{m=1}^{q-1} (-)^{p+m+r+3} a_{p+m+r} G(q-m) + (-)^{p+q+r+3}\, 2!\ \overline{p+q+r}\, a_{p+q+r}.$$

It may be mentioned that when $p-t = q-m = r-s$, the coefficient of a_{t+m+s} will be $(-)^{t+m+s+3}\,3!$, in place of $(-)^{t+m+s+3}$. Similarly when any two of the terms $p-t$, $q-m$ and $r-s$ are equal, the usual coefficient given above will have to be multiplied by 2!

Similar relation can be given for $G(p, p_2 \ldots p_s)$. For economy in space, the coefficient of a typical term alone is given.

$(\alpha-1)!\ \overline{p_k+p_l+p_m \cdots \alpha \text{ terms}}\ \Sigma\,\Sigma \ldots\ s-\alpha$ times $(-)^{p_k+p_l+p_m \cdots + \Sigma r_t + s}\, a_{p_k+p_l+p_m+\cdots \Sigma r_t}$. $G(p_1-r_1, p_2-r_2, \ldots p_{k-}-r_{k-1}, p_{k+1}-r_{k+1}, \ldots p_{l-1}-r_{l-1}, p_{l+1}-r_{l+1}, \ldots)$. As mentioned before, the coefficient will have to be multiplied by π_1', π_2', etc., if π_1', π_2', etc., terms of p_1-r_1, p_2-r_2, $\ldots p_s-r_s$ happen to be equal.

The importance of these relations lies in the fact that they show the structure of the methods of evaluations of the Bipartitional Functions G a (P, Q) and G h (P, Q) by distributions in *plano* dealt with by Dr. Sukhatme. It is not possible in the short compass of this note to bring out these relations. It is hoped to do this in detail later on.

Imperial Agricultural Research Institute, P. V. KRISHNA IYER.
New Delhi,
March 15, 1943.

* The notations used in this note are the same as those of Macmahon in his book *Combinatory Analysis* and Sukhatme in his paper entitled "On Bipartitional Functions," *Phil. Trans. Roy. Soc.*, London, 1938, **237**, 375–409.

NEW FORMULÆ FOR THE DETERMINATION OF MAGNETIC QUANTITIES H, *l*, M AND *m*

FROM the equations $T = 2\pi\sqrt{I/MH}$ and $H \tan\theta = 2Md/(d^2-l^2)^2$, we may obtain from two positions of the deflection magnetometer,

$$H = \frac{\pi\sqrt{8I}}{T}\ \frac{\sqrt{d_1/\tan\theta_1} - \sqrt{d_2/\tan\theta_2}}{d_1^2 - d_2^2}.$$

This equation gives better results for H than the equations given by Worsnop and Flint[1] and by Hutchinson.[2] The value of l (half the magnetic length) is eliminated in the above equation and hence no error arises due to its uncertainty. Attention may be drawn to this correct method of calculating H, since it is not usually given in College text-books.

Gordon College,
Rawalpindi, Punjab, V. R. SINGAL.
February 26, 1943.

1. *Advanced Practical Physics*, 1927, p. 441. 2. *Advanced Text-Book of Electricity and Magnetism*, 1935, p. 122.