Since $\binom{1^r}{2^{\frac{1}{2}r}}$ is the number of ways r units can be grouped in pairs when r is even and since 0 is the number of ways r units can be grouped in pairs where r is odd, it follows that the rth standard moment of the normal curve is the number of ways in which r units can be grouped in pairs.

This development is of interest in that it makes possible the tracing of the value $\binom{1^r}{2^{\frac{1}{2}r}}$ back through the various stages of the development to the coefficient of $(2^{\frac{1}{2}r})$ in the power product expansion of the multinomial theorem.

REFERENCES

[1] H. C. CARVER, "Frequency Curves," Chapter VII of *Handbook of Mathematical Statistics*.

[2] H. C. CARVER, "Fundamentals of the theory of sampling," *Annals of Math. Stat.*, Vol. 1 (1930), pp. 101–121.

[3] P. S. DWYER, "Combined Expansions of Products of Symmetric Power Sums and of Sums of Symmetric Power Products with Application to Sampling," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 1–47, 97–132.

UNIVERSITY OF MICHIGAN,
ANN ARBOR, MICHIGAN

ON A METHOD OF SAMPLING[1]

BY E. G. OLDS

It is recorded that Diogenes fared forth with a lantern in his search for an honest man. History does not tell us how many dishonest men he encountered before he found the first honest one but, judging from the fact that he took his lantern, apparently he expected to have a long search. The general problem of sampling inspection, of which the above is a special case, can be stated as follows:

Given a lot, of size m, containing s items of a specified kind. If items are to be drawn without replacement until i of the s items have been drawn, how many drawings, on the average, will be necessary?

Uspensky[2] has solved a problem concerning balls in an urn, from which the answer to the above question can be obtained for the special case $i = 1$. For the general case, the distribution for the number n of the drawing in which the ith specified item appears, is given by terms of the series:

$$\nu_0' = \sum_{n=1}^{m-s+i} \frac{C_{n-1,i-1}C_{m-n,s-i}}{C_{m,s}} = \sum_{n=0}^{\infty} \frac{C_{n-1,i-1}C_{m-n,s-i}}{C_{m,s}}, \tag{1}$$

[1] Presented to The Institute of Mathematical Statistics, Dec. 27, 1938, at Detroit, Mich., as part of a paper, entitled "Remarks on two methods of sampling inspection."

[2] J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill, New York, 1937, p. 178.

where the first symbol indicates the number of ways of choosing $i - 1$ of the specified items to fill the first $n - 1$ places, the second symbol indicates the number of ways of disposing of $s - 1$ specified items in the last $m - n$ places, and the denominator gives the number of ways that the s items can be scattered through the lot. In order to get the average number of draws we multiply ν_0' by n and sum. Then we have

$$(2)\quad \nu_1' = \sum_{n=0}^{\infty} \frac{nC_{n-1,i-1}C_{m-n,s-1}}{C_{m,s}} = \frac{i(m+1)}{s+1}\sum_{n=0}^{\infty}\frac{C_{n,i}C_{m-n,s-1}}{C_{m+1,s+1}} = \frac{i(m+1)}{s+1}.$$

Example 1. On a table of 200 bargain shirts there are 5 which have a 15 in. neckband and 35 in. sleeves. How many shirts must be examined, on the average, to find two of the desired kind?

Solution. For this case, $m = 100$, $s = 5$, $i = 2$. Therefore $\bar{n} = [2(201)] \div 6 = 67$. Thus, an average of 67 shirts must be examined.

Suppose μ_K represents the Kth moment about the mean, ν_K the Kth moment about the origin, and ν_K' the moment relation given by

$$(3)\quad \nu_K' = (\nu_1 + K - 1)^{(K)},$$

where $(\nu_1 + K - 1)^{(K)}$ represents the result of expanding $(\nu + K - 1)^{(K)}$ and changing the exponent of ν to the corresponding subscript. (For example, $\nu_3' = (\nu_1 + 2)^{(3)} = \nu_3 + 3\nu_2 + 2\nu_1$.) It is easy to derive the recurrence relation

$$(4)\quad \nu_K' = \frac{(i+K-1)(m+K)}{s+K}\nu_{K-1}'.$$

From this result the computation of the moments about the mean is theoretically direct. Actually the results do not seem to be very compact. The variance is given by

$$(5)\quad \mu_2 = \frac{(m+1)(m-s)}{(s+1)^2(s+2)}[i(s+1) - i^2].$$

In case s is unknown and n is known for a particular value of i, we may estimate s, $\left(\text{or rather } \frac{1}{s+1}\right)$, by using the relation, $n = \frac{i(m+1)}{s+1}$. Then

$$(6)\quad \frac{1}{s+1}\text{ est.} = \frac{n}{i(m+1)},$$

and the variance, using this estimate, is given by

$$(7)\quad \text{Variance of}\left(\frac{1}{s+1}\right)\text{est.} = \frac{n}{n+i(m+1)}\cdot\frac{1}{i(m+1)}\left[\frac{n}{i} - 1\right]\left[1 - \frac{n}{m+1}\right].$$

Example 2. In order to check a box of 144 screws, screws are drawn until 10 good screws are obtained. In a particular case only 10 drawings were necessary. Estimate the number of good screws in the lot.

Solution. Here $m = 144$, $i = 10$, $n = 10$. The estimate for s is obtained from $\left(\frac{1}{s+1}\right)$ est. $= \frac{10}{10(145)} = \frac{1}{145}$ and, as might be expected, the conclusion is that all the screws are good. Furthermore the variance of the estimated quantity is zero.

It is obvious that the number of draws necessary to obtain any particular number of specified items is correlated with the numbers of draws for lesser numbers of items. To investigate this, let us suppose that n_j represents the number of draws to obtain exactly j specified items and that $x_j = n_j - n_{j-1}$. It follows immediately from our previous results, that

$$E(x_1) = E(x_2) = E(x_3) = \cdots = \frac{m+1}{s+1}. \tag{8}$$

This result could be obtained from the fact that, corresponding to any arrangement of the lot for which $x_a = a$ and $x_b = b$, there is another arrangement where $x_a = b$ and $x_b = a$, formed by moving $a - b$ of the non-specified items from the first group to the second. From this fact we see, also, that

$$E(x_1^2) = E(x_2^2) = E(x_3^2) = \cdots . \tag{9}$$

But $x_1 = n_1$ and $\sigma_{n_1}^2 = \frac{(m+1)(m-s)}{(s+1)^2(s+2)}[s+1-1] = ds$.

Therefore,

$$\sigma_{x_1}^2 = \sigma_{x_2}^2 = \cdots = ds. \tag{10}$$

But, from our previous formula we have

$$\sigma_{n_2}^2 = d(2s-2), \qquad \sigma_{n_3}^2 = d(3s-6), \text{ etc.}$$

Since $n_2 = x_1 + x_2$, it follows that

$$\sigma_{n_2}^2 = \sigma_{x_1}^2 + 2r_{x_1,x_2}\sigma_{x_1}\sigma_{x_2} + \sigma_{x_2}^2$$

where r_{x_1,x_2} is the correlation between x_1 and x_2. Therefore,

$$r_{x_1,x_2} = -1/s. \tag{11}$$

Also, since $x_1 = n_2 - x_2$, it follows that

$$r_{n_2,x_2} = \sqrt{\frac{s-1}{2s}}. \tag{12}$$

Likewise, from $x_2 = n_2 - x_1$, we get

$$r_{n_2,x_1} = \sqrt{\frac{s-1}{2s}}. \tag{13}$$

Finally, we obtain the three general results

$$r_{n_i,x_{i+1}} = -\sqrt{\frac{i}{s(s-i+1)}}, \tag{14}$$

$$r_{n_i,x_i} = \sqrt{\frac{s-i+1}{si}}, \tag{15}$$

$$r_{n_{i+1},n_i} = \sqrt{\frac{i(s-i)}{(i+1)(s-i+1)}}. \tag{16}$$

Example 3. The cards of a deck are turned one by one until two aces have appeared. The second ace appears when the 36th card is turned. How many more cards should one expect to have to turn to find a third ace?

Solution. Here $m = 52$, $s = 4$, $i = 2$, $n_2 = 36$.

Then $\bar{n}_2 = 2\cdot\frac{53}{5}$, $\bar{x}_3 = \frac{53}{5}$, and $r_{n_2,x_3} = -\sqrt{\frac{2}{4(4-2+1)}} = -\frac{\sqrt{6}}{6}$. Also $\sigma_{x_3} = \sqrt{4d}$ and $\sigma_{n_2} = \sqrt{6d}$. Since $\frac{x_3 - \bar{x}_3}{\sigma_{x_3}} = r_{n_2,x_3}\frac{(n_2 - \bar{n}_2)}{\sigma_{n_2}}$, we have

$$x_3 = \frac{53}{5} - \frac{2}{\sqrt{6}}\cdot\frac{\sqrt{6}}{6}\left(36 - \frac{106}{5}\right) = \frac{17}{3}.$$

Of course this result could have been obtained more directly by noting that there were two aces left among the 16 remaining cards.

Conclusion. The results given in this note might be useful when it is necessary to estimate the number of items to be drawn in order to secure a desired number of a particular type, such as may be the case in obtaining a sample with previously defined characteristics. Also the note disproves such intuitive notions as the one that when looking for a desired record, one is most likely to have to search the whole pile to find it. As far as methods of sampling inspection are concerned, the one implied in this note has little to recommend it.

Carnegie Institute of Technology,
Pittsburgh, Pa.

RANK CORRELATION WHEN THERE ARE EQUAL VARIATES[1]

By Max A. Woodbury

If there is given a set of number pairs

$$(X_1, Y_1), (X_2, Y_2), \cdots, (X_N, Y_N), \tag{1}$$

we may assign to each variate its "rank" (i.e. one more than the number of corresponding variates in the set greater than the given variate). In this way there is obtained a set of pairs of ranks

$$(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N). \tag{2}$$

[1] Presented at the fall meeting, Mich. section of the Math. Assn. of America, Nov. 18, 1939, Kalamazoo College.