DISTRIBUTIONS IN STRATIFIED SAMPLING

By Paul H. Anderson

University of Illinois

1. **Introduction.** In this paper, distributions of means and standard deviations will be derived for random and stratified samples. It is not necessary to define random sampling here, for one may find it defined in any elementary text. If before drawing the sample from a population π, it is divided into several strata $\pi_1, \pi_2, \cdots, \pi_s$, and the sample Σ is composed of s partial samples $\Sigma_1, \Sigma_2, \cdots, \Sigma_s$ each drawn with or without replacement from the strata; and if the sizes m_i of the partial samples are proportionate to the sizes M_i or corresponding strata, i.e., $m_i = kM_i$, then the sample which is obtained in this manner is a stratified sample. When the sizes of the partial samples are not proportionate to the sizes of the corresponding strata, the distributions of means and standard deviations will differ from the distributions obtained when the sizes of the partial samples are proportionate to the sizes of the corresponding strata. This will be shown in the sections that follow.

The distributions of means and standard deviations from well-known populations for stratified and random samples will be derived and compared, as to scatter and symmetry. It should be remembered even though stratification has little to recommend its use, in some cases, over random sampling, the impossibility of obtaining random samples makes its use necessary. Since most of the problems with which the practical statistician is confronted are of the kind which make random sampling difficult or even impossible, stratified sampling is being investigated by many research workers.

2. **The distribution of means and standard deviations for samples of two drawn from any population having a continuous frequency function.** Let $f(x)$ be a continuous frequency function whose mean is zero, and for $a \leq x \leq b$, let $f(x) > 0$, elsewhere let $f(x) = 0$. We select a sample of two elements (x_1, x_2) which can be represented by a point in a square of side $b - a$, as point P in Fig. 1. It is well known that the probability of getting a sample point in the element of area $dx_1\, dx_2$ is $f(x_1)f(x_2)\, dx_1\, dx_2$. The probability of getting a value of $\bar{x}$ (mean) less than the value of the mean represented by a point on the line RT (Fig. 1) whose equation is $x_1 + x_2 = 2\bar{x}$, is given by

$$\int_a^{2\bar{x}-a} dx_1 \int_a^{2\bar{x}-x_1} dx_2 f(x_1)f(x_2). \tag{1}$$

The distribution of $\bar{x} \leq \frac{1}{2}(a + b)$ is

$$2\int_a^{2\bar{x}-a} f(x_1)f(2\bar{x} - x_1)\, dx_1, \tag{2}$$

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

which is obtained by differentiating (1) with respect to $\bar{x}$. For all values $\bar{x} \geq \frac{1}{2}(a + b)$, we must use another equation which we shall now derive similarly. The probability of obtaining a mean less than the mean of any point on $R'T'$ (Fig. 1) is

$$1 - \int_{2\bar{x}-b}^{b} dx_1 \int_{2\bar{x}-x_1}^{b} dx_2 f(x_1)f(x_2).$$

Differentiating this expression, we obtain

$$2\int_{2\bar{x}-b}^{b} f(x_1)f(2\bar{x} - x_1)\, dx_1. \tag{3}$$

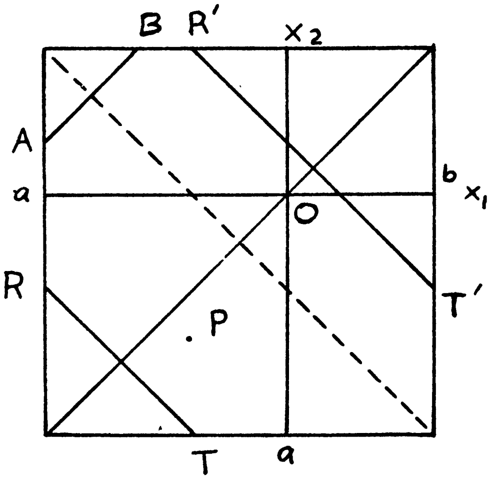

FIG. 1

The distribution of means is given by (2) and (3). Let us apply the theorem to the rectangular population

$$f(x) \begin{array}{l} = 1, \quad -\frac{1}{2} \leq x \leq \frac{1}{2}, \quad a = -\frac{1}{2}, \quad b = \frac{1}{2}, \\ = 0 \text{ elsewhere.} \end{array}$$

Substituting in (2) and (3) respectively, the results obtained are

$$g(\bar{x}) \begin{array}{ll} = 2(1 + 2\bar{x}), & \text{for } \bar{x} \leq 0, \\ = 2(1 - 2\bar{x}), & \text{for } \bar{x} \geq 0. \end{array}$$

J. O. Irwin [1] and Philip Hall [2] obtained these results also but by different methods. However, the distribution of $2\bar{x}$ was known to Laplace and other earlier writers.

From Fig. 1, it is seen that the probability of obtaining a value of S (standard deviation), less than the value of S on AB whose equation is $x_2 - x_1 = 2\,S$ is

$$1 - 2 \int_a^{b-2S} dx_1 \int_{2S+x_1}^{b} dx_2 f(x_1) f(x_2).$$

Upon differentiating this expression with respect to S, we obtain

$$h(S) = 4 \int_a^{b-2S} f(x) f(2S + x)\, dx. \tag{4}$$

For the rectangular population $h(S) = 4(1 - 2S)$. This result agrees with that found by P. R. Rider [3].

3. Sampling from a rectangular population. Let the rectangular population be $f(x) = 1$, for $0 \leq x \leq 1$, elsewhere $f(x) = 0$. From this population we select a stratified sample of two elements which is chosen so that $0 \leq x_1 \leq \frac{1}{2}$ and $\frac{1}{2} \leq x_2 \leq 1$. The probability of obtaining a mean less than the mean of any point on the line $R'T'$ (Fig. 2) whose equation is $x_1 + x_2 = 2\bar{x}$, is

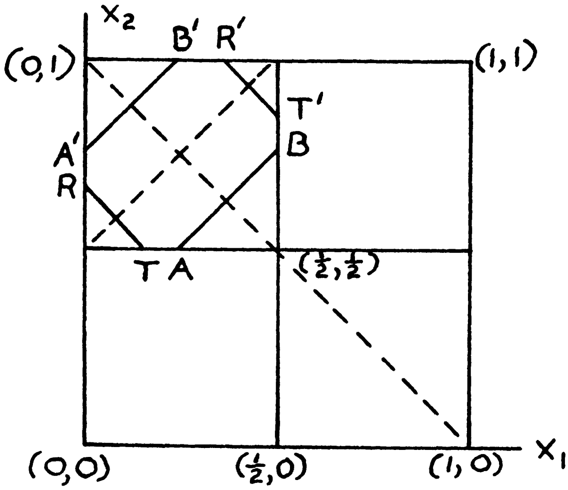

FIG. 2

$$4 \int_0^{2\bar{x}-\frac{1}{2}} dx_1 \int_0^{2\bar{x}-x_1} dx_2 = 4(2\bar{x}^2 - \bar{x} + \tfrac{1}{8}).$$

Similarly, the probability of obtaining a mean less than the mean of any point on $R'T'$ (Fig. 2) whose equation is $x_1 + x_2 = 2\bar{x}$, is

$$1 - 4 \int_{2\bar{x}-1}^{\frac{1}{2}} dx_1 \int_{2\bar{x}-x_1}^{1} dx_2 = 12\bar{x} - 8\bar{x}^2 - \tfrac{7}{2}.$$

Differentiating the right-hand side of the above two equations with respect to $\bar{x}$, we get the distribution of means of stratified samples of two elements from a rectangular distribution function to be

$$(5) \qquad g(\bar{x}) \begin{array}{ll} = 16\bar{x} - 4, & \text{for } \frac{1}{4} \leq \bar{x} \leq \frac{1}{2}, \\ = 4 - 4\bar{x}, & \text{for } \frac{1}{2} \leq \bar{x} \leq \frac{3}{4}. \end{array}$$

The distribution of means for random samples of two elements from the same rectangular population is

$$(6) \qquad g(\bar{x}) \begin{array}{ll} = 4\bar{x}, & \text{for } 0 \leq \bar{x} \leq \frac{1}{2}, \\ = 4 - 4\bar{x}, & \text{for } \frac{1}{2} \leq \bar{x} \leq 1. \end{array}$$

Upon examining (5) and (6) we see that:

A. The stratified sample means are more stable than the random means.

B. The random sample means and the stratified sample means are both distributed symmetrically.

C. The range of the random means is twice the range of the stratified means.

It remains now to find the distributions of the standard deviations for samples of two elements where one element is selected from each half of the population. All points on *AB* (Fig. 2) have the same standard deviation. Furthermore the equation of AB is $x_2 - x_1 = 2S$. The probability of obtaining a standard deviation less than the standard deviation of any point on AB (Fig. 2) is

$$4 \int_{\frac{1}{2}}^{\frac{1}{2}-2S} dx_1 \int_{2S+x_1}^{\frac{1}{2}} dx_2 = 8S^2.$$

Furthermore, the probability of getting a standard deviation less than the standard deviation on the line *A'B'* (Fig. 2) of which the equation is $x_2 - x_1 = 2S$, is

$$1 - 4 \int_{1-2S}^{0} dx_1 \int_{1}^{2S+x_1} dx_2 = -1 - 8S^2 + 8S.$$

Differentiation of the right-hand side of the above two equations with respect to S yields the distribution of standard deviations of stratified samples of two elements from a rectangular distribution function to be

$$(7) \qquad h(S) \begin{array}{ll} = 16S, & \text{for } 0 \leq S \leq \frac{1}{4}, \\ = 8 - 16S, & \text{for } \frac{1}{4} \leq S \leq \frac{1}{2}. \end{array}$$

The distribution of the standard deviations for random samples of two elements is

$$(8) \qquad h(S) = 4(1 - 2S), \qquad \text{for } 0 \leq S \leq \tfrac{1}{2}.$$

From (7) and (8) it is easily seen that:

A. The range of the standard deviations for stratified and random samples is the same.

B. The distribution of standard deviations for random samples of two elements is skewed, but the distribution of the standard deviations for stratified samples of two elements is symmetrical.

If we take a random sample of two elements from the rectangular population on the interval $-\frac{1}{2} \leq x \leq \frac{1}{2}$, then Student's ratio $t = \sqrt{2}\bar{x}/\sqrt{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2}$ will have the distribution

$$F(t) \begin{array}{ll} = 1/2(t-1)^2 & \text{for } t \leq 0, \\ = 1/2(t+1)^2 & \text{for } t \geq 0. \end{array}$$

This result was obtained by Laderman [7] and others. According to the reasoning used by Laderman, the probability of getting a value of t less than the value on OS is (for stratified samples of two elements)

$$4 \int_{-\frac{1}{2}}^{0} dx_1 \int_{0}^{x_1(t+1)/(t-1)} dx_2 = -\tfrac{1}{2}(t+1)/(t-1).$$

When $t \geq 0$, the probability of obtaining a value of t greater than the value on OS is

$$4 \int_{0}^{\frac{1}{2}} dx_2 \int_{x_2(t-1)/(t+1)}^{0} dx_2 .$$

It follows easily that the probability of getting a value of t less than the value represented on OS is for stratified samples equal to

$$1 - 4 \int_{0}^{\frac{1}{2}} dx_2 \int_{x_2(t-1)/(t+1)}^{0} dx_1 = 1 + \tfrac{1}{2}(t-1)/(t+1).$$

Differentiating the right-hand side of the first and third above equations with respect to t, we find the distribution of Student's ratio for stratified samples of two elements from a rectangular population to be

$$F(t) \begin{array}{ll} = 1/(t-1)^2, & \text{for } -1 \leq t \leq 0, \\ = 1/(t+1)^2, & \text{for } \quad 0 \leq t \leq 1. \end{array}$$

Comparing the random sample and stratified sample distributions of t, we find that

A. The stratified t's are more stable than the random t's.

B. Both distributions are symmetrical.

C. The range for the stratified t's is $-1 \leq t \leq +1$, while the range for the t's obtained from random samples is $-\infty \leq t \leq +\infty$.

By means of a different method, distributions of means of stratified samples will be obtained. Let (A) and (B) be rectangular populations $f(x)$, $f(y)$ respectively, with positive values on the interval 0, 1. From the rectangular population (A) select a stratified sample of two elements x_1 and x_2 such that $0 \leq x_1 \leq \frac{1}{2}$, $\frac{1}{2} \leq x_2 \leq 1$. Then the probability of getting a sample point in the element of area $dx_1\, dx_2$ is $4\, dx_1\, dx_2$. Now let $y_1 = 2x_1$ (change of unit of measurement), $y_2 = 2x_2 - 1$ (change of unit of measurement and translation). Then $4\, dx_1\, dx_2 = dy_1\, dy_2$. We have also that $0 \leq y_1 \leq 1$, $0 \leq y_2 \leq 1$. With respect to the distribution of the means, a stratified sample of two elements from (A) is the same as a random sample of two elements from (B). Now the means for random samples of two elements from (B) have the distribution $g(\bar{y})$ which is really expression (6) with $\bar{y}$ substituted for $\bar{x}$. Furthermore, we have $\bar{y} = \frac{1}{2}(y_1 + y_2) = \frac{1}{2}(2x_1 + 2x_2 - 1) = 2\bar{x} - \frac{1}{2}$. Hence it follows readily that $g(\bar{x}) = 16\bar{x} - 4$ for $\frac{1}{4} \leq \bar{x} \leq \frac{1}{2}$, $g(\bar{x}) = 12 - 16\bar{x}$ for $\frac{1}{2} \leq \bar{x} \leq \frac{3}{4}$.

From the rectangular population (A), take a stratified sample of three elements $0 \leq x_1 \leq \frac{1}{3}, \frac{1}{3} \leq x_2 \leq \frac{2}{3}, \frac{2}{3} \leq x_3 \leq 1$. The sample points will all lie in a cube within the unit cube. Then the probability of getting a sample point in the element of volume $dx_1\, dx_2\, dx_3$ is $27\, dx_1\, dx_2\, dx_3$. Now let $y_1 = 3x_1$, $y_2 = 3x_2 - 1$, $y_3 = 3x_3 - 2$. Therefore $0 \leq y_i \leq 1$, for $i = 1, 2, 3$. Furthermore, $dy_1\, dy_2\, dy_3 = 27\, dx_1\, dx_2\, dx_3$. With respect to the distribution of the means, a stratified sample of three elements from (A) is the same as a random sample of three elements from (B). Now the means for random sample of three elements from (B) have the distribution

$$g(\bar{y}) = \begin{cases} 27\bar{y}^2/2, & \text{for } 0 \leq \bar{y} \leq \frac{1}{3}, \\ 9(6\bar{y} - 6\bar{y}^2 - 1)/2, & \text{for } \frac{1}{3} \leq \bar{y} \leq \frac{2}{3}, \\ 27(1 - \bar{y})^2/2, & \text{for } \frac{2}{3} \leq \bar{y} \leq 1. \end{cases}$$

We have also $\bar{y} = 3\bar{x} - 1$. For $\bar{y} = 0, \frac{1}{3}, \frac{2}{3}, 1$, $\bar{x} = \frac{1}{3}, \frac{4}{9}, \frac{5}{9}, \frac{2}{3}$, respectively. Hence

$$g(\bar{x}) = \begin{cases} 81(3\bar{x} - 1)^2/2, & \text{for } \frac{1}{3} \leq \bar{x} \leq \frac{4}{9}, \\ 27(54\bar{x} - 54\bar{x}^2 - 13)/2, & \text{for } \frac{4}{9} \leq \bar{x} \leq \frac{5}{9}, \\ 81(2 - 3\bar{x})^2/2, & \text{for } \frac{5}{9} \leq \bar{x} \leq \frac{2}{3}. \end{cases}$$

Thus we have found the distribution of the means for stratified samples of three elements when one element is selected from each third of the population.

From the rectangular population (A), take a stratified sample of four elements $0 \leq x_1 \leq \frac{1}{4}, \frac{1}{4} \leq x_2 \leq \frac{1}{2}, \frac{1}{2} \leq x_3 \leq \frac{3}{4}, \frac{3}{4} \leq x_4 \leq 1$. Again, a stratified sample of four elements from (A) (with respect to the distribution of means) is the same as a random sample of four elements from (B). The means for random samples of four elements from (B) have the distribution:

$$\text{(C)} \qquad g(\bar{y}) = \begin{cases} 128\bar{y}^3/3, & \text{for } 0 \leq \bar{y} \leq \frac{1}{4}, \\ 8[1 - 24(\bar{y} - \frac{1}{2})^2 - 48(\bar{y} - \frac{1}{2})^3]/3, & \text{for } \frac{1}{4} \leq \bar{y} \leq \frac{1}{2}, \\ 8[1 - 24(\bar{y} - \frac{1}{2})^2 + 48(\bar{y} - \frac{1}{2})^3]/3, & \text{for } \frac{1}{2} \leq \bar{y} \leq \frac{3}{4}, \\ 128(1 - \bar{y})^3/3, & \text{for } \frac{3}{4} \leq \bar{y} \leq 1. \end{cases}$$

Since $\bar{y} = 4\bar{x} - \frac{3}{2}$, we have for $\bar{y} = 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$ respectively, $\bar{x} = \frac{3}{8}, \frac{7}{16}, \frac{1}{2}, \frac{9}{16}, \frac{10}{16}$. Hence

$$g(\bar{x}) = \begin{cases} 512(4\bar{x} - \frac{3}{2})^3/3, & \text{for } \frac{3}{8} \leq \bar{x} \leq \frac{7}{16}, \\ 32[1 - 24(4\bar{x} - 2)^2 - 48(4\bar{x} - 2)^3]/3, & \text{for } \frac{7}{16} \leq \bar{x} \leq \frac{1}{2}, \\ 32[1 - 24(4\bar{x} - 2)^2 + 48(4\bar{x} - 2)^3]/3, & \text{for } \frac{1}{2} \leq \bar{x} \leq \frac{9}{16}, \\ 512(1 - 4\bar{x} + \frac{3}{2})^3/3, & \text{for } \frac{9}{16} \leq \bar{x} \leq \frac{10}{16}. \end{cases}$$

This is the distribution of the means for stratified samples of four elements (one element from each quartile). We can extend this to stratified samples of size n where one element is selected from each stratum and there are n strata. As n increases, we note that

A. The range of the means decreases.

B. The scatter of the means decreases.

C. The number of arcs in the distribution of the means increases.

Take the stratified sample of four elements (two elements from each half), $0 \leq x_1 \leq \frac{1}{2}, 0 \leq x_2 \leq \frac{1}{2}, \frac{1}{2} \leq x_3 \leq 1, \frac{1}{2} \leq x_4 \leq 1$. With respect to the distribution of the means, a stratified sample of four elements from (A) is the same as a random sample of four elements from (B). Now the means for random samples of four elements from (B) have the distribution (C). Furthermore $\bar{y} = 2\bar{x} - \frac{1}{2}$, $d\bar{y} = 2\, d\bar{x}$, and for $\bar{y} = 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$, $\bar{x} = \frac{1}{4}, \frac{3}{8}, \frac{1}{2}, \frac{5}{8}, \frac{3}{4}$. Thus

$$g(\bar{x}) = \begin{cases} 256(2\bar{x} - \frac{1}{2})^3/3, & \text{for } \frac{1}{4} \leq \bar{x} \leq \frac{3}{8}, \\ 16[1 - 24(2\bar{x} - 1)^2 - 48(2\bar{x} - 1)^3]/3, & \text{for } \frac{3}{8} \leq \bar{x} \leq \frac{1}{2}, \\ 16[1 - 24(2\bar{x} - 1)^2 + 48(2\bar{x} - 1)^3]/3, & \text{for } \frac{1}{2} \leq \bar{x} \leq \frac{5}{8}, \\ 256(\frac{3}{2} - 2\bar{x})^3/3, & \text{for } \frac{5}{8} \leq \bar{x} \leq \frac{3}{4}. \end{cases}$$

Hence the distribution of the means for stratified sample of four elements (two elements from each half) has been found.

If we take a stratified sample of six elements (three elements from each half) we find that the graph of the distribution function of the means will consist of six arcs; the range will be $\frac{1}{4} \leq \bar{x} \leq \frac{3}{4}$. Thus we see that as we take more elements from each half, the distribution becomes smoother. The number of arcs in the distribution of the means also increases. The range of the means remains the same but scatter decreases as we take more elements from each half of the population (A).

The results so far obtained are true for the rectangular population which is symmetric. In order to make further comparisons in the distributions of means and standard deviations for stratified and random samples, let us now consider a skewed distribution.

4. **Sampling from a skewed population.** Let us consider the population $f(x) = 2x, 0 \leq x \leq 1, f(x) = 0$ elsewhere. If we take random samples of two elements from this population, the points represented by each sample will lie inside the unit square. For random samples of two elements from this population the distribution of means will consist of two cubics $g(\bar{x}) = 32\bar{x}^3/3$, for $0 \leq \bar{x} \leq \frac{1}{2}$, $g(\bar{x}) = 16(3\bar{x} - 2\bar{x}^3 - 1)/3$, for $\frac{1}{2} \leq \bar{x} \leq 1$. Furthermore, the distribution of the standard deviation for random sample of two elements is a cubic: $h(S) = 16(4S^3 - 3S + 1)/3$, for $0 \leq S \leq \frac{1}{2}$. Now we consider the distribution of means for stratified samples of two elements when one element is selected from the range $(0 \leq x_1 \leq \frac{1}{2})$ which comprises one fourth of the total

population. The other element is selected from the range ($\frac{1}{2} \leq x_2 \leq 1$) which constitutes three quarters of the total population. By use of the geometric method the distribution of the stratified means is found to be

$$g(\bar{x}) \begin{array}{ll} = 16(32\bar{x}^3 - 6\bar{x} + 1)/9, & \text{for } \frac{1}{4} \leq \bar{x} \leq \frac{1}{2}, \\ = 16(30\bar{x} - 9 - 32\bar{x}^3)/9, & \text{for } \frac{1}{2} \leq \bar{x} \leq \frac{3}{4}. \end{array}$$

The range of the stratified means is less, and the distribution is more nearly symmetrical than it is for the random means as may be seen by comparing the graphs of the two distribution functions. Thus we see that stratification gives the means greater stability. The distribution of the standard deviations of the stratified samples of two elements is:

$$h(S) \begin{array}{ll} = 64(3S - 8S^3)/9, & \text{for } 0 \leq S \leq \frac{1}{4}, \\ = 128(4S^3 - 3S + 1)/9, & \text{for } \frac{1}{4} \leq S \leq \frac{1}{2}. \end{array}$$

Upon comparing the distributions of the standard deviations for random and stratified samples, we observe that the random case yields a single cubic whereas the stratified case yields two cubics. The distribution obtained for the stratified case is more symmetrical than it is for the random case as may be seen by sketching the graphs of the two distribution functions. The range for both distributions is the same.

5. **Sampling from a normal population.** We shall consider a normal population F having the frequency function $e^{-\frac{1}{2}x^2}/\sqrt{2\pi}$, $(-\infty \leq x \leq \infty)$; and the ith moment about the mean will be μ_i. Divide this population into two equal parts F_1 and F_2 such that the frequency function of F_1 is $2e^{-\frac{1}{2}x^2}/\sqrt{2\pi}$, $(-\infty \leq x \leq 0)$, and the frequency function of F_2 is $2e^{-\frac{1}{2}x^2}/\sqrt{2\pi}$, $(0 \leq x \leq \infty)$. The ith moment of F_1 about the origin will be m_{i1}, while the ith moment about its mean will be μ_{i1}; the corresponding ith moments for F_2 will be m_{i2} and μ_{i2} respectively. In what follows M_i' will be the ith moment about the origin of the distribution sought, while M_i will be the ith moment about the mean. Furthermore, the constants β_1, β_2, κ, S_κ (measure of skewness) which will be used here are defined in Elderton [8]. Finally, $E[f(x)]$ will be the expected value of $f(x)$.

If we take a random sample of n elements $x_1, x_2, \cdots, x_n$ from F_1 and a random sample of n elements $x_{n+1}, \cdots, x_{2n}$ from F_2, the $2n$ elements $x_1, \cdots, x_n, x_{n+1}, \cdots, x_{2n}$ will be a stratified sample from the population F. Let $\bar{x}_1 = (1/n) \sum_{i=1}^{n} x_i$, $\bar{x}_2 = (1/n) \sum_{i=n+1}^{2n} x_i$, and $\bar{x} = \frac{1}{2}(\bar{x}_1 + \bar{x}_2)$; then $\bar{x}$ will be the mean of the stratified sample. By using Tchouproff's [6] formulae and expected values, we obtain the following values:

$$M' = E(\bar{x}) = \tfrac{1}{2}(m_{11} + m_{12}) = 0,$$

$$M_2 = E(\bar{x}^2) = (\mu_{21} + \mu_{22})/4n = (1 - m_{12}^2)/2n,$$

$$M_3 = E(\bar{x}^3) = (\mu_{31} + \mu_{32})/8n^2 = 0,$$

$$M_4 = E(\bar{x}^4) = [\mu_{41} + \mu_{42} + 3n(\mu_{21} + \mu_{22})^2 - 3(\mu_{21}^2 + \mu_{22}^2)]/16n^3,$$

$$\beta_1 = M_3^2/M_2^3 = 0, \qquad \beta_2 = M_4/M_2^2 = 3 + 4(\pi - 3)/n(\pi - 2)^2.$$

From these constants, we see that the variance of the stratified means is $(1 - 2/\pi)/2n$, but the variance of random means of $2n$ elements is $1/2n$ as is well-known. Thus it is obvious that the scatter of the stratified means is less than the scatter of the random means. Furthermore, the stratified means are distributed symmetrically since $M_3 = 0$. Observing β_2, we notice that the distribution of the stratified means is slightly more peaked than normal. Since it is well known that random means from a normal population are normally distributed, the differences between the two distributions are easy to see. As $n \rightarrow \infty$, $\beta_2 \rightarrow 3$, so it is reasonably likely that the stratified means tend to be normally distributed as the size of the sample increases.

If we select a random sample (x, y) of two elements from the normal population F, then the variance (S^2) will be:

$$S^2 = \tfrac{1}{2}(x^2 + y^2) - (x + y)^2/4 = (x - y)^2/4.$$

The method of expectations gives us the following values:

$$M_2 = \tfrac{1}{2}, \qquad M_3 = 1, \qquad M_4 = \tfrac{15}{4},$$

$$\beta_1 = 8, \qquad \beta_2 = 15, \qquad S_\kappa = \frac{\sqrt{\beta_1}(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)} = \sqrt{2}.$$

Therefore, the skewness of this distribution as measured by Elderton's formula is 1.414. For a stratified sample, where we select x from F_1 and y from F_2, the second, third, and fourth central moments of S^2 are:

$$M_2 = (\pi^2 + 2\pi + 2)/2\pi^2,$$

$$M_3 = (4\pi^3 + 7\pi^2 - 12\pi + 8)/4\pi^3,$$

$$M_4 = (15\pi^4 + 30\pi^3 - 40\pi^2 + 24\pi - 12)/4\pi^4.$$

It follows easily that $\beta_1 = 4.71084$, $\beta_2 = 10.28489$, $\kappa = 19.4$, $2\beta_2 - 3\beta_1 - 6 = .438324$, $S_\kappa = 1.02$. For samples of two elements, the stratified samples yield a distribution for the variance which is less skewed than the corresponding distribution of the variances for random samples. The variances for random samples of two elements are distributed as a Type III curve, while the variance for stratified samples of two elements is either a Type III or a Type VI curve. The difference between the random case and the stratified case as seen from this point of view is not clear cut.

It is interesting to see what sort of bias is introduced by taking n elements of the sample from F_1 and by taking $2n$ elements of the sample from F_2. Under these circumstances, the complete sample will contain $3n$ elements, and the mean of the sample will be $\bar{x} = \sum_{i=1}^{3n} x_i/3n = (\bar{x}_1 + 2\bar{x}_2)/3$. As before, the central moments and the β's are found to be:

$$M_2 = (\mu_{21} + 2\mu_{22})/9n = \mu_{22}/3n, \qquad M_3 = (\mu_{31} + 2\mu_{32})/27n^2 = \mu_{32}/9n^2,$$

$$M_4 = [\mu_{42} - 3\mu_{22}^2 + 9n\mu_{22}^2]/27n^3,$$

$$\beta_1 = \mu_{32}^2/3n\mu_{22}^3, \qquad \beta_2 = \mu_{42}/3n\mu_{22}^2 - 1/n + 3.$$

We notice first that the means are not symmetrically distributed for small values of n since $\beta_1 \not\fallingdotseq 0$, but as $n \to \infty$, $\beta_1 \to 0$, so the means tend to be symmetrically distributed. It is evident also that $\beta_2 \to 3$ with increasing n; consequently, the bias which is present for small values of n tends to disappear as n increases. Incorrect proportioning of the sizes of the partial samples in stratified sampling introduces an error into the results whose magnitude decreases with an increase in n.

6. **Sampling from a population $y = \phi(x)$.** Suppose we have a well-behaved frequency function $\phi(x)$ of which the first four moments are finite. Furthermore, it will be required that $\phi(x)$ be continuous and Riemann-integrable. Divide the total x-axis into K parts $I_1, I_2, \cdots, I_k$ with the separating points $\alpha_1, \alpha_2, \cdots, \alpha_{k-1}$ in such manner that $\int_{-\infty}^{\alpha_1} \phi(x)\,dx = \cdots = \int_{\alpha_{k-1}}^{\infty} \phi(x)\,dx = 1/K$. In this section, we extend some of the definitions of the last section; μ_{it} will be the ith moment about the mean of the tth part I_t, and m_{it} will be the ith moment about the origin of the tth part I_t. Take a sample of Kn elements from this population so that n elements are drawn from each part. The mean of this sample will be $\bar{x} = \sum_{i=1}^{Kn} x_i/Kn$. We write this as $\bar{x} = \sum_{i=1}^{K} \bar{x}_i/K$, where $\bar{x}_i = \sum_{ni-n+1}^{ni} x_i/n$. It follows easily then that:

$$M_2 = \sum_{i=1}^{K} \mu_{2i}/K^2 n, \qquad M_3 = \sum_{i=1}^{K} \mu_{3i}/K^3 n^2,$$

$$M_4 = \left[\sum_{i=1}^{K} \mu_{4i} + 3n\left(\sum_{i=1}^{K} \mu_{2i}\right)^2 - 3\sum_{i=1}^{K} \mu_{2i}^2\right] \Big/ K^4 n^3,$$

as $n \to \infty$, $\beta_1 \to 0$, and $\beta_2 \to 3$. Therefore, it is evident that if we divide a population into K equal parts and take a sample of Kn elements (n elements from each part), the distribution of the means probably tends to normal as the number of elements in the sample increases.

7. **Summary.** Distributions of means for stratified samples have been obtained for the rectangular population which is symmetric and also for a triangular population which can be considered an example of a J-shaped population. For

both populations, the means obtained from stratified samples show less variability than the means of random samples. The stratified sample means obtained from the skewed-population exhibit less skewness than do the random sample means obtained from the same population.

The effect of stratification in sampling upon the distribution of the standard deviations is to make the distribution more symmetric. This is true for the three populations investigated.

For stratified samples from the rectangular population Student's ratio is much more stable than it is for random samples of the same size.

Thus it is evident that stratified samples possess advantages over random samples of a nature that makes stratified samples worthy of use in research work where it is easy to obtain them.

In conclusion, the author is grateful to Professor A. R. Crathorne for suggesting the problem of this paper and guiding it to its conclusion.

REFERENCES

[1] J. O. Irwin, "On the frequency distribution of the means of samples from a population having any law of frequency with finite moments, with special reference to Pearson's Type II," *Biometrika*, Vol. 19 (1927), pp. 225–239.

[2] Philip Hall, "The distribution of means for samples of size N drawn from a population in which the variate takes values between 0 and 1, all such values being equally probable," *Biometrika*, Vol. 21 (1927), pp. 240–244.

[3] Paul R. Rider, "On the distribution of the ratio of the mean to standard deviation in small samples from non-normal universes," *Biometrika*, Vol. 21 (1929), pp. 124–143.

[4] J. Neyman, "On the two different aspects of the representative method," *Roy. Stat. Soc. Jour.*, Vol. 97 (1934), pp. 558–625.

[5] A. E. R. Church, "On the means and squared standard deviations," *Biometrika*, Vol. 8 (1926), pp. 321–394.

[6] A. A. Tchouproff, "On the mathematical expectation of the moments of frequency distributions," *Biometrika*, Vol. 12 (1918–19), pp. 140–169, 185–210.

[7] Jack Laderman, "The distribution of Student's ratio for samples of two items drawn from non-normal universes," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 376–380.

[8] W. Palin Elderton, *Frequency Curves and Correlation*, London: Charles and Edwin Layton, 1927 (Second Edition), pp. 239.