ON THE CORRELATION BETWEEN CERTAIN AVERAGES FROM SMALL SAMPLES*

By

ALLEN T. CRAIG

1. *Introduction.* It is well known that no correlation exists between the arithmetic mean and standard deviation of samples drawn at random from a normal universe. However, there seems to be in the literature no treatment of the correlation between other averages either for normal or non-normal universes. In the present paper, a few simple theorems are established which make possible the determination of the type of regression of the median on the arithmetic mean, of the range on the median, and of the range on the arithmetic mean. In case the regression is linear, the coefficient of correlation may be computed.

We shall understand a probability function $f(x)$ of a real variable x to be, for all values of x on a range of R a single-valued, non-negative, continuous function with $\int_R f(x)\,dx = 1$. Then $\int_a^b f(x)\,dx$ is the probability that a value of x chosen at random lies in the interval (a, b) where a and b are in R and $a < b$; and $f(x)\,dx$ is, to within infinitesimals of higher order, the probability that a value of x chosen at random lies in the interval $(x, x+dx)$. It will prove convenient to classify probability functions according as R is the range $(-\infty, \infty)$, $(0, \infty)$, or $(0, k)$, $k>0$. In accord with this classification,[1] we shall refer to probability functions as of the first, second, and third kinds respectively. In a similar manner, we define a probability function $F(x,y)$ of two independent variables.

*Presented to the American Mathematical Society, Dec. 28, 1931.

[1] Cf. L. Bachelier, Calcul des Probabilités, p. 155.

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

2. The correlation between the arithmetic mean $\bar{x}$ and the range W.

Theorem I. *Let $f(x)$ be the probability function of the variable x. Let $F_1(\bar{x}, W)$ be that of the arithmetic mean $\bar{x}$ and the range W in samples of three independent values of x. If $f(x)$ is a probability function of the first kind, then*

$$F_1(\bar{x}, W) = 18 \int_{\bar{x}+\frac{W}{3}}^{\bar{x}+\frac{2W}{3}} f(x_1) f(x_1 - W) f(3\bar{x} - 2x_1 + W) dx_1 .$$

Proof. Let x_1, x_2, x_3, be the three observed values of x. Write

$$x_1 + x_2 + x_3 = 3\bar{x},$$
$$x_1 \qquad - x_3 = W,$$
$$x_3 \leq x_2 \leq x_1 .$$

For $\bar{x}$ assigned, $-\infty < \bar{x} < \infty$, and W assigned, $0 \leq W < \infty$ we must have

$$\bar{x} + \frac{W}{3} \leq x_1 \leq \bar{x} + \frac{2W}{3},$$
$$x_3 = x_1 - W,$$
$$x_2 = 3\bar{x} - x_1 - x_3 .$$

If we consider all possible arrangements of x_1, x_2, x_3, we have

$$F_1(\bar{x}, W) d\bar{x}\, dW = 6 \int_{\bar{x}+\frac{W}{3}}^{\bar{x}+\frac{2W}{3}} f(x_1) f(x_2) f(x_3) dx_1\, dx_2\, dx_3 .$$

Let

$$x_1 = x_1$$
$$x_2 = 3\bar{x} - x_1 - x_3 ,$$
$$x_3 = x_1 - W.$$

The absolute value of the Jacobin is 3. Hence the theorem.

In the case of samples of four independent items x_1, x_2, x_3, x_4, the probability function $F_1(\bar{x}, W)$ is given by

$$F_1(\bar{x}, W) = 48\int_{\bar{x}+\frac{W}{4}}^{\bar{x}+\frac{W}{2}}\int_{4\bar{x}-3x_1+W}^{x_1} f(x_1)f(x_2)f(4\bar{x}-2x_1-x_2+W)f(x_1-W)\,dx_2\,dx_1$$

$$+48\int_{\bar{x}+\frac{W}{2}}^{\bar{x}+\frac{3W}{4}}\int_{x_1-W}^{4\bar{x}-3x_1+2W} f(x_1)f(x_2)f(4\bar{x}-2x_1-x_2+W)f(x_1-W)\,dx_2\,dx_1.$$

We note that the probability function is made up of the sum of two parts depending on whether x_1 is in the interval $(\bar{x}+\frac{W}{4}, \bar{x}+\frac{W}{2})$ or in the interval $(\bar{x}+\frac{W}{2}, \bar{x}+\frac{3W}{4})$. Moreover, it may be of interest to note the overlapping of the ranges of integration of x_2. To prove that $F_1(\bar{x}, W)$ is given as stated, we take

$$x_1 + x_2 + x_3 + x_4 = 4\bar{x},$$
(1) $$x_4 \le x_3, \quad x_2 \le x_1,$$
$$x_1 - x_4 = W.$$

From (1) it readily follows that

(2) $$2x_1 + x_2 + x_3 = 4\bar{x} + W.$$

For assigned values of $\bar{x}$ and W, the upper limit on x_1 is found from (2) by taking $x_2 = x_3 = x_4 = x_1 - W$. Thus $x_1 = \bar{x} + \frac{3W}{4}$. Similarly, the lower limit on x_1 is found from (2) by taking $x_2 = x_3 = x_1$. Thus $x_1 = \bar{x} + \frac{W}{4}$. But x_2 may not always be as large as x_1 for all values of x_1. This may be seen by taking $x_2 = x_1$ and $x_3 = x_4 = x_1 - W$ in (2). This leads to $x_1 = \bar{x} + \frac{W}{2}$. Thus, for $\bar{x} + \frac{W}{4} \le x_1 \le \bar{x} + \frac{W}{2}$, we see that x_1 is the upper limit on x_2. To determine the lower limit on x_2 for this region of variation of x_1, we select x_2 as near $x_4 = x_1 - W$ as is possible without causing x_3 to exceed x_1. But $x_3 = 4\bar{x} - 2x_1 - x_2 + W$. At most, then $4\bar{x} - 2x_1 - x_2 + W = x_1$ or $x_2 = 4\bar{x} - 3\bar{x}_1 + W$. Thus we have established the limits of integration used in the first part of the sum of which $F_1(\bar{x}, W)$ consists. A similar argument shows if $\bar{x} + \frac{W}{2} \le x_1 \le \bar{x} + \frac{3W}{4}$, that

$$x_1 - W \le x_2 \le 4\bar{x} - 3x_1 + 2W.$$

If $f(x)$ is a probability function of the second kind, we observe in samples of three independent items x_1, x_2, x_3, for $\bar{x}$ assigned, that $0 \le W \le 3\bar{x}$. If $0 \le W \le 3\bar{x}/2$, we have

$$\bar{x} + \frac{W}{3} \le x_1 \le \bar{x} + \frac{2W}{3},$$
$$x_2 = 3\bar{x} - 2x_1 + W,$$
$$x_3 = x_1 - W,$$

and if $\frac{3\bar{x}}{2} \le W \le 3\bar{x}$, we have

$$W \le x_1 \le \bar{x} + \frac{2W}{3},$$
$$x_2 = 3\bar{x} - 2x_1 + W,$$
$$x_3 = x_1 - W.$$

Accordingly,

$$F_1(\bar{x}, W) = 18\int_{\bar{x}+\frac{W}{3}}^{\bar{x}+\frac{2W}{3}} f(x_1)f(x_1 - W)f(3\bar{x} - 2x_1 + W)dx_1, \quad 0 \le W \le \frac{3\bar{x}}{2},$$
$$= 18\int_{W}^{\bar{x}+\frac{2W}{3}} f(x_1)f(x_1 - W)f(3\bar{x} - 2x_1 + W)dx_1, \quad \frac{3\bar{x}}{2} \le W \le 3\bar{x}:$$

In samples of four independent items x_1, x_2, x_3, x_4, drawn from a universe characterized by a law of probability of this kind, we find

$$F_1(\bar{x}, W) = 48\int_{\bar{x}+\frac{W}{4}}^{\bar{x}+\frac{W}{2}}\int_{4\bar{x}-3x_1+W}^{x_1} f(x_1)f(x_2)f(4\bar{x} - 2x_1 - x_2 + W)f(x_1 - W)dx_2\,dx_1$$
$$+48\int_{\bar{x}+\frac{W}{2}}^{\bar{x}+\frac{3W}{4}}\int_{x_1-W}^{4\bar{x}-3x_1+2W} f(x_1)f(x_2)f(4\bar{x} - 2x_1 - x_2 + W)f(x_1 - W)dx_2\,dx_1,$$
$$0 \le W \le \frac{4\bar{x}}{3},$$
$$= 48\int_{W}^{\bar{x}+\frac{W}{2}}\int_{4\bar{x}-3x_1+W}^{x_1} f(x_1)f(x_2)f(4\bar{x} - 2x_1 - x_2 + W)f(x_1 - W)dx_2\,dx_1$$
$$+48\int_{\bar{x}+\frac{W}{2}}^{\bar{x}+\frac{3W}{4}}\int_{x_1-W}^{4\bar{x}-3x_1+2W} f(x_1)f(x_2)f(4\bar{x} - 2x_1 - x_2 + W)f(x_1 - W)dx_2\,dx_1,$$
$$\frac{4\bar{x}}{3} \le W \le 2\bar{x},$$
$$= 48\int_{W}^{\bar{x}+\frac{3W}{4}}\int_{x_1-W}^{4\bar{x}-3x_1+2W} f(x_1)f(x_2)f(4\bar{x} - 2x_1 - x_2 + W)f(x_1 - W)dx_2\,dx_1.$$
$$2\bar{x} \le W \le 4\bar{x}.$$

Finally, consider $f(x)$ to be a probability function of the third kind. In samples of three independent items x_1, x_2, x_3, for $0 \le \bar{x} \le k/3$, we obtain $0 \le W \le 3\bar{x}$; for $k/3 \le \bar{x} \le 2k/3$, we obtain $0 \le W \le k$; for $2k/3 \le \bar{x} \le k$, we obtain $0 \le W \le 3(k-\bar{x})$. It is fairly easy to see that for $\bar{x}$ and W assigned as indicated, the following regions of selection of x_1 are valid:

for $0 \le \bar{x} \le k/2$ and $0 \le W \le 3\bar{x}/2$,
or for $k/2 \le \bar{x} \le k$ and $0 \le W \le 3(k-\bar{x})/2$, then
$\bar{x} + W/3 \le x_1 \le \bar{x} + 2W/3$;
for $0 \le \bar{x} \le k/3$ and $3\bar{x}/2 \le W \le 3\bar{x}$,
or for $k/3 \le \bar{x} \le k/2$ and $3\bar{x}/2 \le W \le 3(k-\bar{x})/2$, then
$W \le x_1 \le \bar{x} + 2W/3$;
for $2k/3 \le \bar{x} \le k$ and $3(k-\bar{x})/2 \le W \le 3(k-\bar{x})$
or for $k/2 \le \bar{x} \le 2k/3$ and $3(k-\bar{x})/2 \le W \le 3\bar{x}/2$, then
$\bar{x} + W/3 \le x_1 \le k$;
for $k/3 \le \bar{x} \le k/2$ and $3(k-\bar{x})/2 \le W \le k$,
or for $k/2 \le \bar{x} \le 2k/3$ and $3\bar{x}/2 \le W \le k$, then $W \le x_1 \le k$.

Thus,

$$F_1(\bar{x}, W) = 18\int_{\bar{x}+\frac{W}{3}}^{\bar{x}+\frac{2W}{3}} f(x_1)\, f(x_1 - W)\, f(3\bar{x} - 2x_1 + W)\, dx_1,$$

$$= 18\int_{W}^{\bar{x}+\frac{2W}{3}} f(x_1)\, f(x_1 - W)\, f(3\bar{x} - 2x_1 + W)\, dx_1,$$

$$= 18\int_{\bar{x}+\frac{W}{3}}^{k} f(x_1)\, f(x_1 - W)\, f(3\bar{x} - 2x_1 + W)\, dx_1,$$

$$= 18\int_{W}^{k} f(x_1)\, f(x_1 - W)\, f(3\bar{x} - 2x_1 + W)\, dx_1,$$

over those regions of the $\bar{x}W$-plane indicated above.

In case of samples of four independent items x_1, x_2, x_3, x_4, drawn from a universe characterized by a probability function of the third kind, for $0 \le \bar{x} \le k/4$, we obtain $0 \le W \le 4\bar{x}$;

for $k/4 \leq \bar{x} \leq 3k/4$, we obtain $0 \leq W \leq k$; for $3k/4 \leq \bar{x} \leq k$, we obtain $0 \leq W \leq 4(k-\bar{x})$. Let us denote as follows the regions of the $\bar{x}W$-plane bounded by the given lines:

$$(A)\begin{cases} \bar{x} = 0 \\ W = \frac{4\bar{x}}{3} \\ W = \frac{4(k-\bar{x})}{3} \end{cases} \qquad (E)\begin{cases} W = \frac{4\bar{x}}{3} \\ W = 2(k-\bar{x}) \\ W = 4(k-\bar{x}) \end{cases}$$

$$(B)\begin{cases} W = \frac{4(k-\bar{x})}{3} \\ W = 2\bar{x} \\ W = \frac{4\bar{x}}{3} \end{cases} \qquad (F)\begin{cases} W = \frac{4\bar{x}}{3} \\ W = 2\bar{x} \\ W = \frac{4(k-\bar{x})}{3} \\ W = 2(k-\bar{x}) \end{cases}$$

$$(C)\begin{cases} W = 2\bar{x} \\ W = 4\bar{x} \\ W = \frac{4(k-\bar{x})}{3} \end{cases} \qquad (G)\begin{cases} W = k \\ W = 2\bar{x} \\ W = \frac{4(k-\bar{x})}{3} \end{cases}$$

$$(D)\begin{cases} W = \frac{4\bar{x}}{3} \\ W = 2(k-\bar{x}) \\ W = \frac{4(k-\bar{x})}{3} \end{cases} \qquad (H)\begin{cases} W = k \\ W = \frac{4\bar{x}}{3} \\ W = 2(k-\bar{x}) \end{cases}$$

Further, let

$$\theta = f(x_1)f(x_2)f(x_1-W)f(4\bar{x}-2x_1-x_2+W),$$

and let

$$\int_a^b\int_c^d \theta\, dx_2\, dx_1 = \begin{pmatrix} b & d \\ a & c \end{pmatrix} \theta.$$

It is then not difficult to verify that

$$F_1(\bar{x},W) = 48\left[\begin{pmatrix} \bar{x}+\frac{W}{2} & x_1 \\ \bar{x}+\frac{W}{4} & 4\bar{x}-3x_1+W \end{pmatrix}\theta + \begin{pmatrix} \bar{x}+\frac{3W}{4} & 4\bar{x}-3x_1+2W \\ \bar{x}+\frac{W}{2} & x_1-W \end{pmatrix}\theta\right], \quad (A)$$

$$= 48\left[\begin{pmatrix} \bar{x}+\frac{W}{2} & x_1 \\ W & 4\bar{x}-3x_1+W \end{pmatrix}\theta + \begin{pmatrix} \bar{x}+\frac{3W}{4} & 4\bar{x}-3x_1+2W \\ \bar{x}+\frac{W}{2} & x_1-W \end{pmatrix}\theta\right], \quad (B)$$

$$= 48\left[\begin{pmatrix} \bar{x}+\frac{3W}{4} & 4\bar{x}-3x_1+W \\ W & x_1-W \end{pmatrix}\theta\right], \quad (C)$$

$$= 48\left[\begin{pmatrix} \bar{x}+\frac{W}{2} & x_1 \\ \bar{x}+\frac{W}{4} & 4\bar{x}-3x_1+W \end{pmatrix}\theta + \begin{pmatrix} k & 4\bar{x}-3x_1+2W \\ \bar{x}+\frac{W}{2} & x_1-W \end{pmatrix}\right], \quad (D)$$

$$= 48\left[\begin{pmatrix} k & x_1 \\ \bar{x}+\frac{W}{4} & 4\bar{x}-3x_1+W \end{pmatrix}\right]\theta \;, \tag{E}$$

$$= 48\left[\begin{pmatrix} \bar{x}+\frac{W}{2} & x_1 \\ W & 4\bar{x}-3x_1+W \end{pmatrix}\right]\theta \;, \tag{F}$$

$$= 48\left[\begin{pmatrix} k & 4\bar{x}-3x_1+2W \\ W & x_1-W \end{pmatrix}\right]\theta \;, \tag{G}$$

$$= 48\left[\begin{pmatrix} k & x_1 \\ W & 4\bar{x}-3x_1+W \end{pmatrix}\right]\theta \;. \tag{H}$$

As illustrations of these theorems, let us find the correlation between the range and the mean for universes of specified types.

Example 1. Let $f(x) = e^{-x}, \quad 0 \leq x < \infty.$
For samples of three items, we have

$$F_1(\bar{x}, W) = 6We^{-3\bar{x}}, \quad 0 \leq W \leq \frac{3\bar{x}}{2},$$
$$= 18\left(\bar{x} - \frac{W}{3}\right)e^{-3\bar{x}}, \quad \frac{3\bar{x}}{2} \leq W \leq 3\bar{x}.$$

The distributions of the marginal totals of W and $\bar{x}$ are obtained by integrating $F_1(\bar{x}, W)$ with regard to $\bar{x}$ and W respectively. We readily find

$$\varphi(\bar{x}) = \frac{27\bar{x}^2}{2}e^{-3\bar{x}}, \qquad 0 \leq \bar{x} < \infty,$$

and

$$\psi(W) = 2e^{-2W}(e^{W}-1), \quad 0 \leq W < \infty,$$

as previously given by the writer.[2] For $\bar{x}$ assigned, the mean of the array of W is $\overline{W}_{\bar{x}} = \frac{3\bar{x}}{2}$. Thus the regression of W on $\bar{x}$ is linear and $r = \frac{\sqrt{15}}{5}$.

---

[2] American Journal of Mathematics, Vol. 54 (1932), pp. 359, 366.

Example 2. Let $f(x) = 1/k, \quad 0 \le x \le k.$

For samples of three items, we have

$$F_1(\bar{x}, W) = \frac{6W}{k^3},$$

$$= \frac{18}{k^3}\left(\bar{x} - \frac{W}{3}\right),$$

$$= \frac{18}{k^3}\left(k - \bar{x} - \frac{W}{3}\right),$$

$$= \frac{18}{k^3}(k - W)$$

over those regions of the $\bar{x}W$-plane indicated above. The marginal totals[3] are distributed in accord with

$$\varphi(\bar{x}) = \frac{27\bar{x}^2}{2k^3}, \qquad 0 \le \bar{x} \le \frac{k}{3},$$

$$= \frac{9}{2k^3}\left[-6\bar{x}^2 + 6k\bar{x} - k^2\right], \quad \frac{k}{3} \le \bar{x} \le \frac{2k}{3},$$

$$= \frac{27}{2k^3}(k - \bar{x})^2, \quad \frac{2k}{3} \le \bar{x} \le k,$$

and

$$\psi(W) = \frac{6W}{k^3}(k - W), \quad 0 \le W \le k.$$

We readily find

$$\overline{W}_{\bar{x}} = \frac{3\bar{x}}{2}, \quad 0 \le \bar{x} \le \frac{k}{3},$$

---

[3]Cf. H. L. Rietz, On a Certain Law of Probability of Laplace, Proc. Int. Math. Congress, Toronto (1924), pp. 795-799.

J. O. Irwin, On the Frequency Distributions of Means, etc., Biometrika, Vol. 19 (1927), pp. 225-239.

P. Hall, The Distribution of Means for Samples of Size N, Biometrika, Vol. 19 (1927), pp. 240-245.

J. Neyman and E. S. Pearson, On the Use and Distribution of Certain Test Criteria, Biometrika, Vol. 20 (1928), p. 210.

$$= \frac{5k^3 - 27k^2\bar{x} + 27k\bar{x}^2}{6k^2 - 36k\bar{x} + 36\bar{x}^2}, \quad \frac{k}{3} \le \bar{x} \le \frac{2k}{3},$$

$$= \frac{3}{2}(k - \bar{x}), \qquad \frac{2k}{3} \le \bar{x} \le k.$$

Thus the regression curve of W on $\bar{x}$ is continuous, but the regression is non-linear for $\frac{k}{3} \le \bar{x} \le \frac{2k}{3}$.

3. The correlation between the arithmetic mean $\bar{x}$ and the median ξ.

Theorem II. *Let $f(x)$ be the probability function of the variable x. Let $F_2(\bar{x}, \xi)$ be that of the arithmetic mean $\bar{x}$ and the median ξ in samples of three independent values of x. If $f(x)$ is a probability function of the first kind, then*

$$F_2(\bar{x}, \xi) = 18 f(\xi) \int_{3\bar{x}-2\xi}^{\infty} f(x_1) f(3\bar{x} - \xi - x_1) dx_1, \quad \xi \le \bar{x},$$

$$= 18 f(\xi) \int_{\xi}^{\infty} f(x_1) f(3\bar{x} - \xi - x_1) dx_1, \quad \bar{x} \le \xi.$$

Proof. Let x_1, x_2, x_3, be the three observed values of x. Write

$$x_1 + x_2 + x_3 = 3\bar{x},$$
$$x_2 = \xi$$
$$x_3 \le x_2 \le x_1$$

For $\bar{x}$ and ξ assigned, $\xi \le \bar{x}$, we must have

$$3\bar{x} - 2\xi \le x_1 < \infty$$
$$x_2 = \xi$$
$$x_3 = 3\bar{x} - \xi - x_1,$$

and for $\bar{x} \le \xi$,

$$\xi \le x_1 < \infty$$
$$x_2 = \xi$$
$$x_3 = 3\bar{x} - \xi - x_1.$$

If we consider all possible arrangements of x_1, x_2, x_3, we have

$$F_2'(\bar{x},\xi)\,d\bar{x}\,d\xi = 6f(\xi)\,d\xi\int_{3\bar{x}-2\xi}^{\infty} f(x_1)\,f(x_3)\,dx_1\,dx_3, \qquad \xi \le \bar{x},$$

$$= 6f(\xi)\,d\xi\int_{\xi}^{\infty} f(x_1)\,f(x_3)\,dx_1\,dx_3, \qquad \bar{x} \le \xi.$$

The change of variable $x_3 = 3\bar{x} - \xi - x_1$ establishes the theorem.

In case of samples of five independent items x_1, x_2, x_3, x_4, x_5, the probability function $F_2'(\bar{x},\xi)$ is given by

$$F_2'(\bar{x},\xi) = 150f(\xi)\int_{\xi}^{5\bar{x}-4\xi}\int_{5\bar{x}-3\xi-x_1}^{\infty}\int_{5\bar{x}-2\xi-x_1-x_2}^{\xi} f(x_1)f(x_2)f(x_4)f(5\bar{x}-\xi-x_1-x_2-x_4)\,dx_4\,dx_2\,dx_1$$

$$+150f(\xi)\int_{5\bar{x}-4\xi}^{\infty}\int_{\xi}^{\infty}\int_{5\bar{x}-2\xi-x_1-x_2}^{\xi} f(x_1)f(x_2)f(x_4)f(5\bar{x}-\xi-x_1-x_2-x_4)\,dx_4\,dx_2\,dx_1, \quad \xi \le \bar{x},$$

$$= 150f(\xi)\int_{\xi}^{\infty}\int_{\xi}^{\infty}\int_{5\bar{x}-2\xi-x_1-x_2}^{\xi} f(x_1)f(x_2)f(x_4)f(5\bar{x}-\xi-x_1-x_2-x_4)\,dx_4\,dx_2\,dx_1, \quad \bar{x} \le \xi.$$

This follows immediately from the fact that for $\bar{x}$ and ξ assigned, $\xi \le \bar{x}$, we may have either

$$\begin{aligned}
\xi \le x_1 &\le 5\bar{x} - 4\xi,\\
5\bar{x} - 3\xi - x_1 \le x_2 &< \infty,\\
5\bar{x} - 2\xi - x_1 - x_2 \le x_4 &\le \xi,\\
x_3 &= \xi,\\
x_5 &= 5\bar{x} - \xi - x_1 - x_2 - x_4,
\end{aligned}$$

or

$$\begin{aligned}
5\bar{x} - 4\xi \le x_1 &< \infty,\\
\xi \le x_2 &< \infty,\\
5\bar{x} - 2\xi - x_1 - x_2 \le x_4 &\le \xi\\
x_3 &= \xi\\
x_5 &= 5\bar{x} - \xi - x_1 - x_2 - x_4,
\end{aligned}$$

and for $\bar{x} \leq \xi$, we must have

$$\begin{aligned} \xi \leq x_1 &< \infty \\ \xi \leq x_2 &< \infty, \\ 5\bar{x} - 2\xi - x_1 - x_2 \leq x_4 &\leq \xi, \\ x_3 &= \xi \\ x_5 &= 5\bar{x} - \xi - x_1 - x_2 - x_4. \end{aligned}$$

If $f(x)$ is a probability function of the second kind, it is clear that $0 \leq \xi \leq \frac{3\bar{x}}{2}$ in samples of three items. Then

$$F_2(\bar{x},\xi) = 18 f(\xi) \int_{3\bar{x}-2\xi}^{3\bar{x}-\xi} f(x_1) f(3\bar{x}-\xi-x_1)\,dx_1, \qquad 0 \leq \xi \leq \bar{x},$$

$$= 18 f(\xi) \int_{\xi}^{3\bar{x}-\xi} f(x_1) f(3\bar{x}-\xi-x_1)\,dx_1, \qquad \bar{x} \leq \xi \leq \frac{3\bar{x}}{2}.$$

In case of samples of five independent items drawn at random from a universe characterized by a probability function of the second kind, $F_2(\bar{x},\xi)$ can best be expressed in a form employing the notation used previously. Thus we write

$$\Phi = f(x_1) f(x_2) f(x_4) f(5\bar{x} - \xi - x_1 - x_2 - x_4),$$

$$u_{ij} = 5\bar{x} - i\xi - x_1 - x_2 - \cdots\cdots - x_j,$$

and

$$\int_a^b \int_c^d \int_e^f \Phi\, dx_4\, dx_2\, dx_1 = \begin{pmatrix} b & d & f \\ a & c & e \end{pmatrix} \Phi.$$

Then

$$\begin{aligned} F_2(\bar{x},\xi) = 150 f(\xi) \Bigg[&\begin{pmatrix} u_{40} & u_{21} & \xi \\ \xi & u_{31} & u_{22} \end{pmatrix} \Phi + \begin{pmatrix} u_{40} & u_{11} & u_{12} \\ \xi & u_{21} & 0 \end{pmatrix} \Phi \\ &+ \begin{pmatrix} u_{30} & u_{21} & \xi \\ u_{40} & \xi & u_{22} \end{pmatrix} \Phi + \begin{pmatrix} u_{30} & u_{11} & u_{12} \\ u_{40} & u_{21} & 0 \end{pmatrix} \Phi \\ &+ \begin{pmatrix} u_{20} & u_{11} & u_{12} \\ u_{30} & \xi & 0 \end{pmatrix} \Phi \Bigg], \qquad 0 \leq \xi \leq \bar{x}, \end{aligned}$$

$$= 150 f(\xi)\left[\begin{pmatrix} u_{30} & u_{21} & \xi \\ \xi & \xi & u_{22} \end{pmatrix}\Phi + \begin{pmatrix} u_{30} & u_{11} & u_{12} \\ \xi & u_{21} & 0 \end{pmatrix}\Phi\right.$$

$$\left. + \begin{pmatrix} u_{20} & u_{11} & u_{12} \\ u_{30} & \xi & 0 \end{pmatrix}\Phi\right], \quad \bar{x} \leq \xi \leq \frac{5\bar{x}}{4},$$

$$= 150 f(\xi)\left[\begin{pmatrix} u_{20} & u_{11} & u_{12} \\ \xi & \xi & 0 \end{pmatrix}\Phi\right], \quad \frac{5\bar{x}}{4} \leq \xi \leq \frac{5\bar{x}}{3}.$$

Finally, consider $f(x)$ to be a probability function of the third kind. In samples of three independent items, for $0 \leq \bar{x} \leq k/3$, we obtain $0 \leq \xi \leq 3\bar{x}/2$; for $k/3 \leq \bar{x} \leq 2k/3$, we obtain $(3\bar{x}-k)/2 \leq \xi \leq 3\bar{x}/2$; for $2k/3 \leq \bar{x} \leq k$, we obtain $(3\bar{x}-k)/2 \leq \xi \leq k$. It is not difficult to verify for $\bar{x}$ and ξ assigned as indicated, the following regions of selection of x_1 are valid:

for $0 \leq \bar{x} \leq k/3$ and $0 \leq \xi \leq \bar{x}$,
or for $k/3 \leq \bar{x} \leq k/2$ and $3\bar{x} - k \leq \xi \leq \bar{x}$, then
$3\bar{x} - 2\xi \leq x_1 \leq 3\bar{x} - \xi$;
for $k/3 \leq \bar{x} \leq k/2$ and $(3\bar{x}-k)/2 \leq \xi \leq 3\bar{x} - k$,
or for $k/2 \leq \bar{x} \leq k$ and $(3\bar{x}-k)/2 \leq \xi \leq \bar{x}$, then
$3\bar{x} - 2\xi \leq x_1 \leq k$;
for $0 \leq \bar{x} \leq k/2$ and $\bar{x} \leq \xi \leq 3\bar{x}/2$,
or for $k/2 \leq \bar{x} \leq 2k/3$ and $3\bar{x} - k \leq \xi \leq 3\bar{x}/2$, then
$\xi \leq x_1 \leq 3\bar{x} - \xi$;
for $k/2 \leq \bar{x} \leq 2k/3$ and $\bar{x} \leq \xi \leq 3\bar{x} - k$,
or for $2k/3 \leq \bar{x} \leq k$ and $\bar{x} \leq \xi \leq k$, then $\xi \leq x_1 \leq k$.

Thus

$$F_2(\bar{x}, \xi) = 18 f(\xi) \int_{3\bar{x}-2\xi}^{3\bar{x}-\xi} f(x_1) f(3\bar{x} - \xi - x_1)\, dx_1,$$

$$= 18 f(\xi) \int_{3\bar{x}-2\xi}^{k} f(x_1) f(3\bar{x} - \xi - x_1)\, dx_1,$$

$$= 18 f(\xi) \int_{\xi}^{3\bar{x}-\xi} f(x_1) f(3\bar{x}-\xi-x_1)\, dx_1,$$

$$= 18 f(\xi) \int_{\xi}^{k} f(x_1) f(3\bar{x}-\xi-x_1)\, dx_1,$$

over those regions of the $\bar{x}\xi$-plane as indicated above.

With samples of five items, the correlation surface is defined in so many parts that we shall not take the space necessary to consider it.

As illustrations of these theorems, we shall find the correlation between the median and the mean for universes of specified types.

Example 1. Let $f(x) = e^{-x}, \quad 0 \le x < \infty.$
For samples of three items, we have

$$F_2(\bar{x},\xi) = 18\xi e^{-3\bar{x}}, \quad 0 \le \xi \le \bar{x},$$
$$= 18(3\bar{x}-2\xi) e^{-3\bar{x}}, \quad \bar{x} \le \xi \le \frac{3\bar{x}}{2}$$

The distribution function of the marginal totals of ξ is given by[4]

$$\phi(\xi) = 6e^{-2\xi}(1-e^{-\xi}), \quad 0 \le \xi < \infty.$$

For $\bar{x}$ assigned, the mean of the array of ξ is

$$\bar{\xi}_{\bar{x}} = \frac{5\bar{x}}{6}.$$

Thus the regression of ξ on $\bar{x}$ is linear and $r = \frac{5\sqrt{267}}{89}$.

Example 2. Let $f(x) = \frac{1}{k}, \quad 0 \le x \le k.$
For samples of three items, we have

$$F_2(\bar{x},\xi) = \frac{18\xi}{k^3}$$

$$= \frac{18}{k^3}(k-3\bar{x}+2\xi),$$

[4]Cf. American Journal of Mathematics, Vol. 54 (1932), p. 364.

$$= \frac{18}{k^3}(3\bar{x}-2\xi),$$

$$= \frac{18}{k^3}(3k-\xi),$$

over those regions of the $\bar{x}\xi$-plane indicated above. The distribution function of the marginal totals of ξ is given by[5]

$$\phi(\xi) = \frac{6\xi}{k^3}(k-\xi), \qquad 0 \le \xi \le k.$$

We find

$$\begin{aligned} \bar{\xi}_{\bar{x}} &= \frac{5\bar{x}}{6}, & 0 \le \bar{x} \le \frac{k}{3}, \\ &= \frac{5\bar{x}^3-(3\bar{x}-k)^3}{6\bar{x}^2-2(3\bar{x}-k)^2}, & \frac{k}{3} \le \bar{x} \le \frac{2k}{3}, \\ &= \frac{(5\bar{x}+k)}{6}, & \frac{2k}{3} \le \bar{x} \le k. \end{aligned}$$

Thus the regression curve of ξ on $\bar{x}$ is continuous but the regression is non-linear for $\frac{k}{3} \le \bar{x} \le \frac{2k}{3}$.

4. The correlation between the median ξ and the range W.

Theorem III. *Let $f(x)$ be the probability function of the variable x. Let $F_3(\xi, W)$ be that of the median ξ and the range W in samples of $2m+1$ independent values of x. If $f(x)$ is a probability function of the first kind, then*

$$F_3(\xi, W) = \frac{(2m+1)!}{[(m-1)!]^2} f(\xi) \int_{\xi}^{\xi+W} f(x_1) f(x_1-W) \left[\int_{\xi}^{x_1} f(t)\,dt\right]^{m-1} \left[\int_{x_1-W}^{\xi} f(t)\,dt\right]^{m-1} dx_1.$$

Proof. We have

$$x_1 - x_{2m+1} = W,$$

$$x_{m+1} = \xi,$$

$$\xi \le x_2, \cdots\cdots, x_m \le x_1,$$

$$x_1 - W \le x_{m+1}, \cdots\cdots, x_{2m} \le \xi.$$

[5]Cf. P. R. Rider, On the Distribution of the Ratio of Mean to Standard Deviation, etc., Biometrika, Vol. 21 (1929), pp. 136-137.

Hence the theorem.

If $f(x)$ is a probability function of the second kind, then

$$F_3(\xi,W)=\frac{(2m+1)!}{[(m-1)!]^2}f(\xi)\int_{\xi}^{\xi+W}f(x_1)f(x_1-W)\left[\int_{\xi}^{x_1}f(t)dt\right]^{m-1}\left[\int_{x_1-W}^{\xi}f(t)dt\right]^{m-1}dx_1,\quad W\le\xi,$$

$$=\frac{(2m+1)!}{[(m-1)!]^2}f(\xi)\int_{W}^{\xi+W}f(x_1)f(x_1-W)\left[\int_{\xi}^{x_1}f(t)dt\right]^{m-1}\left[\int_{x_1-W}^{\xi}f(t)dt\right]^{m-1}dx_1,\quad \xi\le W.$$

Finally, consider $f(x)$ to be a probability function of the third kind. We observe for $0\le\xi\le k$, that $0\le W\le k$. For assigned values of ξ and W, the following regions of selection of x_1 are obvious:
for $0\le\xi\le k/2$, and $0\le W\le\xi$,
or for $k/2\le\xi\le k$ and $0\le W\le k-\xi$, then $\xi\le x_1\le\xi+W$;
for $0\le\xi\le k/2$ and $\xi\le W\le k-\xi$, then $W\le x_1\le\xi+W$;
for $0\le\xi\le k/2$, and $k-\xi\le W\le k$,
or for $k/2\le\xi\le k$ and $\xi\le W\le k$, then $W\le x_1\le k$;
for $k/2\le\xi\le k$ and $k-\xi\le W\le\xi$, then $\xi\le x_1\le k$.
If we write

$$\psi=f(x_1)f(x_1-W)\left[\int_{\xi}^{x_1}f(t)dt\right]^{m-1}\left[\int_{x_1-W}^{\xi}f(t)dt\right]^{m-1},$$

we have

$$F_3(\xi,W)=\frac{(2m+1)!}{[(m-1)!]^2}f(\xi)\int_{\xi}^{\xi+W}\psi\,dx_1,$$

$$=\frac{(2m+1)!}{[(m-1)!]^2}f(\xi)\int_{W}^{\xi+W}\psi\,dx_1,$$

$$=\frac{(2m+1)!}{[(m-1)!]^2}f(\xi)\int_{W}^{k}\psi\,dx_1,$$

$$=\frac{(2m+1)!}{[(m-1)!]^2}f(\xi)\int_{\xi}^{k}\psi\,dx_1,$$

over those regions of the ξW-plane previously indicated.

We shall consider two simple examples.

Example 1. Let $f(x) = e^{-x}, \quad 0 \le x < \infty.$
With samples of three items,

$$F_3(\xi, W) = 3e^{-3\xi}(e^{W} - e^{-W}), \qquad W \le \xi,$$
$$= 3e^{-W-\xi}(1 - e^{-2\xi}), \qquad \xi \le W.$$

The regression is readily shown to be non-linear.

Example 2. Let $f(x) = \frac{1}{k}, \quad 0 \le x \le k.$
With samples of three items,

$$F_3(\xi, W) = \frac{6W}{k^3},$$
$$= \frac{6\xi}{k^3},$$
$$= \frac{6}{k^3}(k - W)$$
$$= \frac{6}{k^3}(k - \xi),$$

over those regions of the ξW-plane which have been previously given. The mean of the array of W corresponding to an assigned ξ is $\overline{W}_\xi = \frac{k}{2}$. Accordingly, there is no correlation between the median and the range in samples of three items drawn from this universe.

It is easy to employ the type of argument used in establishing Theorem III to obtain the probability function of the median and lower quartile. Thus, if $f(x)$ is a probability function of the second kind and $F_4(\xi, \eta)$ is the probability function of the median ξ and the lower quartile η in samples of $4m+1$ items, then

$$F_4(\xi, \eta) = \frac{(4m+1)!}{(2m)!\,m!\,(m-1)!} f(\xi) f(\eta) \left[\int_\xi^\infty f(t)\,dt\right]^{2m} \left[\int_0^\eta f(t)\,dt\right]^{m} \cdot \left[\int_\eta^\xi f(t)\,dt\right]^{m-1}, \qquad \eta \le \xi.$$

Allen T Craig