THE RELATION BETWEEN THE MEANS AND VARIANCES, MEANS SQUARED AND VARIANCES IN SAMPLES FROM COMBINATIONS OF NORMAL POPULATIONS

By

G. A. BAKER

The distributions of the means and variances of samples from the combinations of normal populations have been discussed in a previous paper.[1] It is known that if the sampled population is not normal the means and variances of samples are not independent.

The present discussion aims to give some idea of the relation between the means and the variances, means squared and variances of samples from a population that is the combination of normal populations. To this end the case of samples of two from such populations is rather completely investigated. Also empirical random sampling results for two special populations are presented.

Suppose that a population is represented by

$$(1)\qquad f(x)=\frac{1}{1+k}\left[\frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}+\frac{k}{\sigma\sqrt{2\pi}}\,e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}\right].$$

[1]"Random Sampling from Non-Homogeneous Populations," Metron, Vol. VIII, No. 3 (1930), pp. 1-21.

If a method used by Karl Pearson[1] is followed, the probability of

$$x_1 \text{ in } dx_1 \text{ is } f(x_1)dx_1 \, ,$$
$$x_2 \text{ in } dx_2 \text{ is } f(x_2)dx_2$$

and the probability of the concurrence of these two events is

$$f(x_1)\,f(x_2)\,dx_1\,dx_2 \tag{2}$$

which may be written

$$\frac{1}{(1+k)^2 2\pi}\Big[e^{-\frac{1}{2}[x_1^2+x_2^2]}+\frac{k^2}{\sigma^2}\,e^{-\frac{1}{2\sigma^2}[(x_1-m)^2+(x_2-m)^2]} \tag{3}$$
$$+\frac{k}{\sigma}\Big\{e^{-\frac{1}{2}\left[x_1^2+\frac{(x_2-m)^2}{\sigma^2}\right]}+e^{-\frac{1}{2}\left[x_2^2+\frac{(x_1-m)^2}{\sigma^2}\right]}\Big\}\Big]\,dx_1dx_2.$$

Now

$$x=\frac{1}{2}(x_1+x_2)$$

$$\Sigma^2=\frac{1}{2}\left[(x_1-x)^2+(x_2-x)^2\right] \, .$$

Whence

$$\begin{cases} x_1=-\Sigma+x \\ x_2=\Sigma+x \, . \end{cases} \tag{4}$$

Also $dx_1\,dx_2$ may be replaced[2] by

[1]Appendix to Papers by "Student" and R. A. Fisher, Biometrika, Vol. XIX (1925), p. 522.

[2]R. A. Fisher: "Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population," Biometrika, Vol. X (1915), p. 507.

$$c\, d\Sigma\, dx \tag{5}$$

In virtue of (5), (4) and (3), (6) is obtained.

$$\frac{1}{(1+k)^2 2\pi}\left[e^{-\frac{1}{2}[2x^2+2\Sigma^2]} + \frac{k^2}{\sigma^2}\, e^{-\frac{1}{2\sigma^2}[2\Sigma^2+2(x-m)^2]}\right.$$

$$+\frac{k}{\sigma}\left\{e^{-\frac{1}{2}\left[(-\Sigma+x)^2+\frac{(\Sigma+x-m)^2}{\sigma^2}\right]}\right. \tag{6}$$

$$\left.\left.+\, e^{-\frac{1}{2}\left[(\Sigma+x)^2+\frac{(-\Sigma+x-m)^2}{\sigma^2}\right]}\right\}\right].$$

This is the correlation surface for the means and standard deviations of samples of two drawn from (1). To get the correlation surface of the means and variances write

$$\Sigma^2 = u$$

$$d\Sigma = \frac{du}{2\sqrt{u}}$$

Then

$$F(x,u) = \frac{1}{(1+k)^2 2\pi}\left[\frac{e^{-\frac{1}{2}[2x^2+2u]}}{2\sqrt{u}} + \frac{k^2}{2\sqrt{u}\,\sigma^2}\, e^{-\frac{1}{2\sigma^2}[2u+2(x-m)^2]}\right.$$

$$+\frac{k}{\sigma}\left\{\frac{e^{-\frac{1}{2}\left[(\sqrt{u}+x)^2+\frac{(-\sqrt{u}+x-m)^2}{\sigma^2}\right]}}{2\sqrt{u}}\right. \tag{7}$$

$$\left.\left.+\,\frac{e^{-\frac{1}{2}\left[(-\sqrt{u}+x)^2+\frac{(\sqrt{u}+x-m)^2}{\sigma^2}\right]}}{2\sqrt{u}}\right\}\right]$$

is the desired surface.

The locus of mean u's for given x's is

$$u=\frac{e^{-x^2}+k^2\sigma e^{-\frac{(x-m)^2}{\sigma^2}}+\frac{4\sqrt{2}k}{(\sigma^2+1)^{\frac{3}{2}}}e^{-\frac{2(x-\frac{m}{2})^2}{\sigma^2+1}}\left[\sigma^2+\frac{\{(\sigma^2-1)x+m\}^2}{\sigma^2+1}\right]}{e^{-x^2}+\frac{k^2}{\sigma}e^{-\frac{(x-m)^2}{\sigma^2}}+\frac{2\sqrt{2}k}{\sqrt{\sigma^2+1}}e^{-\frac{2(x-\frac{m}{2})^2}{\sigma^2+1}}}. \tag{8}$$

The locus of the mean x's for given u's is

$$x=\frac{\frac{mk^2}{\sigma}e^{-\frac{u}{\sigma^2}}+\frac{2\sqrt{2}k}{\sqrt{\sigma^2+1}}\left[\{(\sigma^2-1)\sqrt{u}+m\}e^{-\frac{2(\sqrt{u}-\frac{m}{2})^2}{\sigma^2+1}}-\{(\sigma^2-1)\sqrt{u}-m\}e^{-\frac{2(\sqrt{u}+\frac{m}{2})^2}{\sigma^2+1}}\right]}{e^{-u}+\frac{k^2}{\sigma}e^{-\frac{u}{\sigma^2}}+\frac{2\sqrt{2}k}{\sqrt{\sigma^2+1}}\left\{e^{-\frac{2(\sqrt{u}-\frac{m}{2})^2}{\sigma^2+1}}+e^{-\frac{2(-\sqrt{u}-\frac{m}{2})^2}{\sigma^2+1}}\right\}}. \tag{9}$$

The correlation surface for the means squared (= z) and variances is

$$\begin{aligned}\psi(u,z)=\frac{1}{(1+k)^2 2\pi}\Bigg[&\frac{e^{-\frac{1}{2}[2z+2u]}}{4\sqrt{u}\sqrt{z}}+\frac{k^2 e^{-\frac{1}{2\sigma^2}[2u+2(\sqrt{z}-m)^2]}}{\sigma^2 4\sqrt{u}\sqrt{z}}\\ &+\frac{k}{\sigma}\Bigg\{\frac{e^{-\frac{1}{2}\left[(\sqrt{u}+\sqrt{z})^2+\frac{(-\sqrt{u}+\sqrt{z}-m)^2}{\sigma^2}\right]}}{4\sqrt{u}\sqrt{z}}\\ &+\frac{e^{-\frac{1}{2}\left[(-\sqrt{u}+\sqrt{z})^2+\frac{(\sqrt{u}+\sqrt{z}-m)^2}{\sigma^2}\right]}}{4\sqrt{u}\sqrt{z}}\Bigg\}\Bigg].\end{aligned} \tag{10}$$

The locus of the mean u 's for given z 's is

$$(11)\quad u=\frac{e^{-z}+k^2\sigma e^{-\frac{(\sqrt{z}-m)^2}{\sigma^2}}+\frac{4\sqrt{2}k}{(\sigma^2+1)^{\frac{3}{2}}}e^{-\frac{2(\sqrt{z}-\frac{m}{2})^2}{\sigma^2+1}}\left[\sigma^2+\frac{\{(\sigma^2-1)\sqrt{z}+m\}^2}{\sigma^2+1}\right]}{e^{-z}+\frac{k^2}{\sigma}e^{-\frac{(\sqrt{z}-m)^2}{\sigma^2}}+\frac{2\sqrt{2}k}{\sqrt{\sigma^2+1}}e^{-\frac{2(\sqrt{z}-\frac{m}{2})^2}{\sigma^2+1}}}$$

The locus of the mean z 's for given u's is

$$z=\frac{1}{e^{-u}+\frac{k^2e^{-\frac{u}{\sigma^2}}}{\sigma}+\frac{2\sqrt{2}k}{\sqrt{\sigma^2+1}}\left\{e^{-\frac{2(\sqrt{u}-\frac{m}{2})^2}{\sigma^2-1}}+e^{-\frac{2(\sqrt{u}+\frac{m}{2})^2}{\sigma^2+1}}\right\}}$$

(12) multiplied by

$$\Bigg[e^{-u}+\frac{k^2}{\sigma}(\sigma^2+m^2)e^{-\frac{u}{\sigma^2}}$$

$$+\frac{4\sqrt{2}k}{(\sigma^2+1)^{\frac{3}{2}}}\Bigg\{\left(\sigma^2+\frac{[(\sigma^2-1)\sqrt{n}-m]^2}{\sigma^2+1}\right)e^{-\frac{2(\sqrt{u}-\frac{m}{2})^2}{\sigma^2+1}}$$

$$+\left(\sigma^2+\frac{[(\sigma^2-1)\sqrt{u}+m]^2}{\sigma^2+1}\right)e^{-\frac{2(\sqrt{u}+\frac{m}{2})^2}{\sigma^2+1}}\Bigg\}\Bigg]$$

By expanding the denominators of (8), (9), (11), and (12) by the multinomial theorem, it can be shown that each of these loci is essentially parabolic, $\sigma^2-1\neq 0$. They are subject to an exponential influence at the beginning of the range of the independent variable, which influence rapidly diminishes as the independent variable takes on higher values.

The probability relations in general between means and variances, means squared and variances will be expected to approximate those for the case of samples of two, because of the following considerations. Suppose that n (the number in the sample) is large.[1] When a large proportion of the sample comes from the first component, the first term of (7) with 2 in the numerator of the exponent replaced by n and with $u^{-\frac{1}{2}}$ replaced by $u^{\frac{n-3}{2}}$ will be an approximation to the surface of the means and variances. Similarly, when a large proportion of the sample comes from the second component, the second term of (7) with 2 in the numerator of the exponent replaced by n and with $u^{-\frac{1}{2}}$ replaced by $u^{\frac{n-3}{2}}$ will be an approximation to the surface of the means and variances. When about equal proportions of the sample come from each component, the last term of (7) with $\frac{2}{2}$ in the numerator of each exponent replaced by $\frac{n}{2}$ and with $u^{-\frac{1}{2}}$ replaced by $u^{\frac{n-3}{2}}$ will be an approximation to the surface of the means and variances. Or, all together, (7) with the mentioned changes in the exponents of the terms, with proper weighting of the terms, and with $u^{-\frac{1}{2}}$ replaced by $u^{\frac{n-3}{2}}$ is a proportionate approximation to the distribution of the means and variances of samples drawn from a population represented by (1). Further, increasing n will not influence relations (8), (9), (11), and (12) as approximations for the general case except the exponential term, if it is assumed that the denominators are expanded and then multiplied by the numerators, for $\frac{1}{n}$ occurs to the same power in the numerators and denominators.

---

[1]Note: The effect of k and of the binomial coefficients is roughly as follows. If the $n+1$ terms denoting s from the first component of (1) and $n-s$ from the second component are divided into thirds, then, if l_1, l_2, l_3 are the exponents of k in the middle terms,

$$l_1 = \frac{n-2}{6}, \quad l_2 = \frac{3n}{6}, \quad l_3 = \frac{5n+2}{6}$$

or approximately, since n is large and since only a proportionate expression is desired $l_1' = 0$, $l_2' = \frac{n}{3}$ $l_3 = \frac{2n}{3}$

or the exponents of k of the middle terms of the three sections above are $\frac{n}{3}$ times the exponents of k in (7). The effect of increasing n because of the binomial coefficients is to weight the middle section of the possible surfaces to a much greater extent than the extreme sections, so that with n very large the last term of (7) with 2 replaced by n becomes an approximation to the desired surface.

From (8), (9), (11), and (12) it is clear that the parameters of the sampled population have great influence on the regression relations considered. It should be borne in mind in this connection that many flattened and skewed, as well as bimodal, distributions can be adequately represented by combinations of normal populations. Also, results (8), (9), (11), and (12) can be extended to the sums and differences of any number of normal curves, subject to the condition that the resultant is always positive.

In 1925, Dr. Neyman[1] gave the correlation coefficient between the deviations of the means of samples from the mean of the sampled population and the variances of these samples for samples of n drawn at random from an infinite uni-variate population in terms of the betas of the sampled population as

$$(13) \qquad \rho' = \frac{\sqrt{n-1}\,\sqrt{\beta_1}}{\sqrt{(n-1)\beta_2 - n + 3}} \, .$$

Similarly, the correlation coefficient between the deviations squared of the means of samples from the mean of the sampled population and the variances is

$$(14) \qquad R' = \frac{\sqrt{n-1}\,(\beta_2 - 3)}{\sqrt{(\beta_2 + 2n - 3)\left[(n-1)\beta_2 - n + 3\right]}} \, .$$

Under certain very special conditions the statement of ρ' and R' may give an adequate idea of the regression relation between the means and variances, means squared and variances of samples from a population represented by (1). In general the mere statement of these coefficients will not give any useful

[1] J. Splawa-Neyman: "Contributions to the Theory of Small Samples Drawn from a Finite Population," Biometrika, Vol. XVII (1925), pp. 472-479.

notion of the actual probability relations. This is true because: (a) the regression relations between means and variances, means squared and variances of samples from a population represented by (1) are essentially parabolic, as shown for samples of two and as seems probable for larger samples; (b) the frequency arrays may vary markedly in dispersion, in skewness, and in other characteristics.

To illustrate these remarks, samples of four were drawn from two special populations by throwing dice.

Suppose that a population is represented by

$$(15)\quad f(x)=\frac{1}{1+k}\left[\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x+m_1)^2}+\frac{k}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(x-m_2)^2}{\sigma^2}}\right].$$

The first four moments of $f(x)$ about its mean are

$$u_0=1,$$

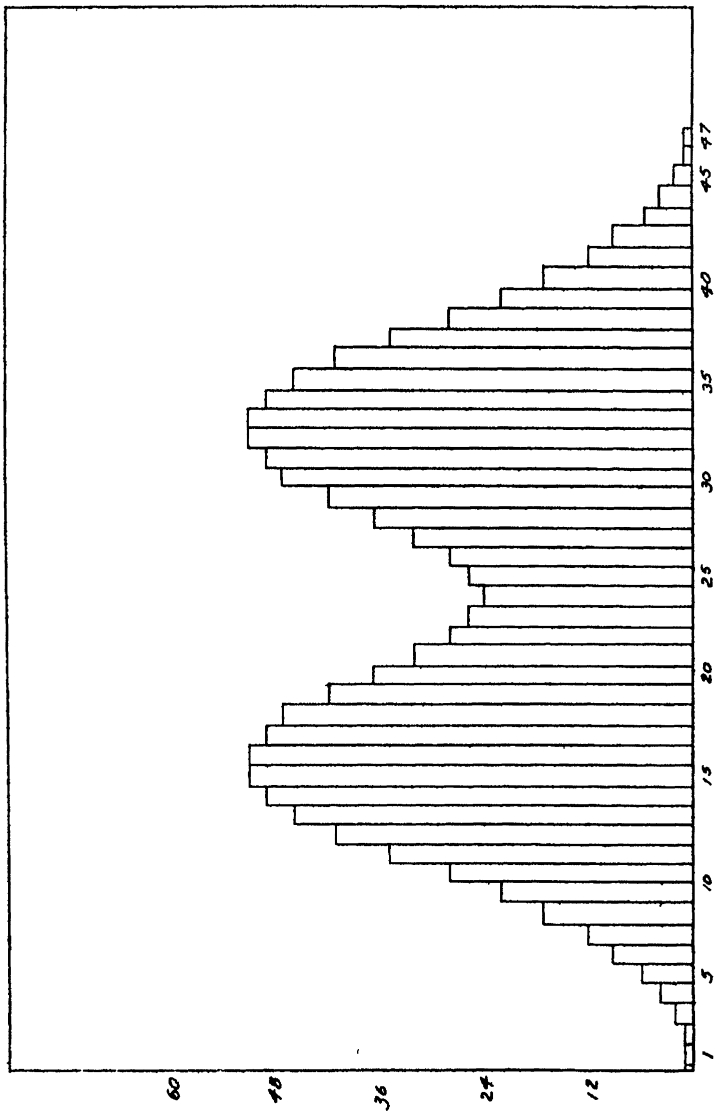
$$u_1=\frac{(-m_1+km_2)}{1+k}=0,$$

$$u_2=\frac{[1+m_1^2+k(\sigma^2+m_2^2)]}{1+k},$$

$$u_3=\frac{-3m_1-m_1^3+k(3m_2\sigma^2+m_2^3)}{1+k},$$

$$u_4=\frac{3+6m_1^2+m_1^4+k(3\sigma^4+6m_2^2\sigma^2+m_2^4)}{1+k}.$$

CHART A



Population I, from Which the 1038 Samples of Four of Tables I and II were drawn.

TABLE I

Correlation Table Showing the Relation Between the Means and Variances of Samples of Four from Population I

Means of Samples	Variances of Samples																	
	0 to 15	15 to 30	30 to 45	45 to 60	60 to 75	75 to 90	90 to 105	105 to 120	120 to 135	135 to 150	150 to 165	165 to 180	180 to 195	195 to 210	210 to 225	225 to 240	240 to 255	T'tls
15 to 17			1															1
13 to 15	1	1																2
11 to 13	5	1	1	2														9
9 to 11	13	8	0	4		2	2											29
7 to 9	11	15	10	5	7	2	6	7			1		1					65
5 to 7	4	13	16	22	10	7	11	7	1	2								93
3 to 5	12	13	12	12	25	12	5	10	4	2	2	3	1	1			1	115
1 to 3	4	8	18	18	20	16	32	14	8	5	3	3	2	1		1		153
– 1 to 1	5	4	7	34	21	24	27	29	8	10	7	5	1		2			184
– 3 to – 1	5	12	14	14	16	18	21	13	7	7	3	2		1	1	1		135
– 5 to – 3	2	17	8	18	12	18	17	10	6	1	5	1	1	1	1			118
– 7 to – 5	6	6	12	3	14	9	3	6	4	1			1					65
– 9 to – 7	11	6	9	3	9	2	1	2	2									45
–11 to – 9	3	6	3				1											13
–13 to –11	4	1	1		1													7
–15 to –13	1	1	1		1													4
–17 to –13																		
Totals	87	112	113	135	136	110	126	98	40	28	21	14	7	4	4	2	1	1038

Whence

$$(16)\qquad \beta_1 = \frac{(1+k)[-3m_1 - m_1^3 + k(3m_2\sigma^2 + m_2^3)]^2}{[1+m_1^2 + k(\sigma^2 + m_2^2)]^3},$$

$$(17)\qquad \beta_2 = \frac{(1+k)[3+6m_1^2 + m_1^4 + k(3\sigma^4 + 6m_2^2\sigma^2 + m_2^4)]}{[1+m_1^2 + k(\sigma^2 + m_2^2)]^2}.$$

Thus, for any special population of the form (15), R' and ρ' can be easily calculated.

Samples of four were drawn from a population approximately represented by

$$(18)\qquad f_1(x) = 648\left[\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x+1.7)^2} + \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-1.7)^2}\right]$$

The actual sampled population is shown in Chart A and is hereinafter called Population I.

Table I shows the distribution of 1038 samples of four drawn from Population I with respect to the observed values of the means and the variances. The arrays for constant values of the variances are at first distinctly bimodal, gradually becoming unimodal. Chart I shows the means of arrays of Table I with the regression lines as calculated without correction for groupings. It is apparent that the locus of the mean variances for a given value of the means diverges a great deal from a straight line. This regression relation looks as though it was a normal curve, which is what would be expected from (8) with $\sigma^2 - 1 = 0$. The theoretical and actual correlation coefficients for this and three subsequent tables are compared in Table V and the constants of the marginal distributions of Tables I to IV are presented in Table VI.

If the deviations of the means of the samples of Table I from the mean of Population I are squared, Table II results. Chart II shows the means of arrays and regression lines of Table II. The regression lines are very poor fits to the means of the arrays which are, apparently, exponential loci.

Table III shows the distribution of 1058 samples of four drawn from a population approximately represented by

$$(19) \quad f_2(x) = 972\left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x+.8)^2} + \frac{1}{3\sqrt{2\pi}} e^{-\frac{1}{2}(x-2.4)^2}\right]$$

with respect to the observed values of the means and variances of the samples. The actual sampled population is presented in Chart B and is hereinafter called Population II. Chart III shows the means of arrays and regression lines of Table III. This chart resembles Chart I in that the locus of the mean variances for given values of the means is so obviously non-linear. Also, a glance at Table III is sufficient to see that the arrays vary markedly in skewness.

Table IV shows the relation between the means squared and variances of samples of four from Population II. Chart IV shows the means of arrays and regression lines for Table IV. In this case the regression relations seem to be fairly near linear, and the frequency distributions of the arrays do not change strikingly.

CHART I

The Means of Arrays and Regression Lines of the Means and Variances of Samples from Population I

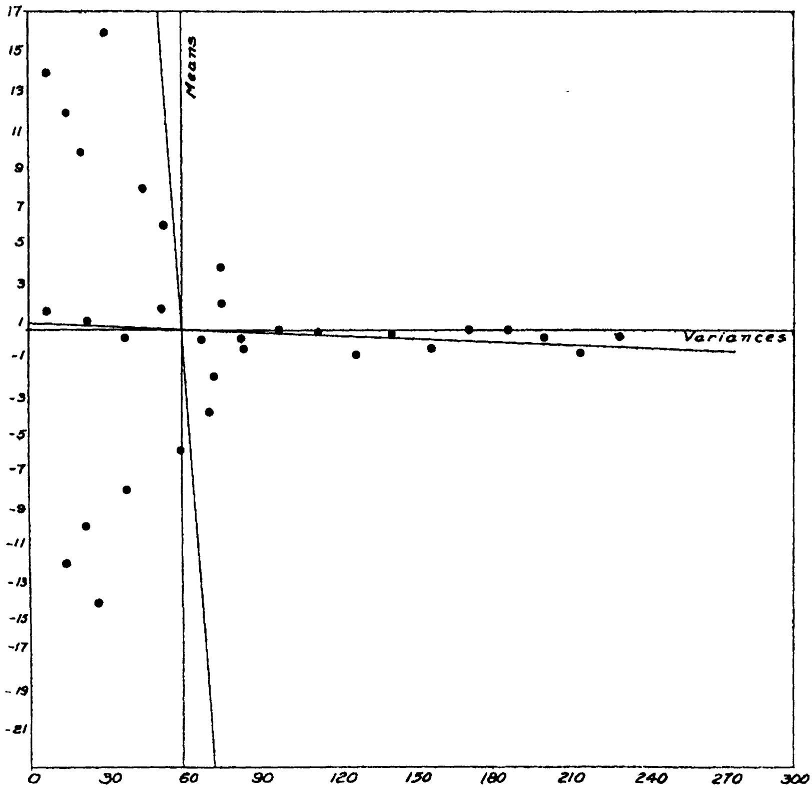


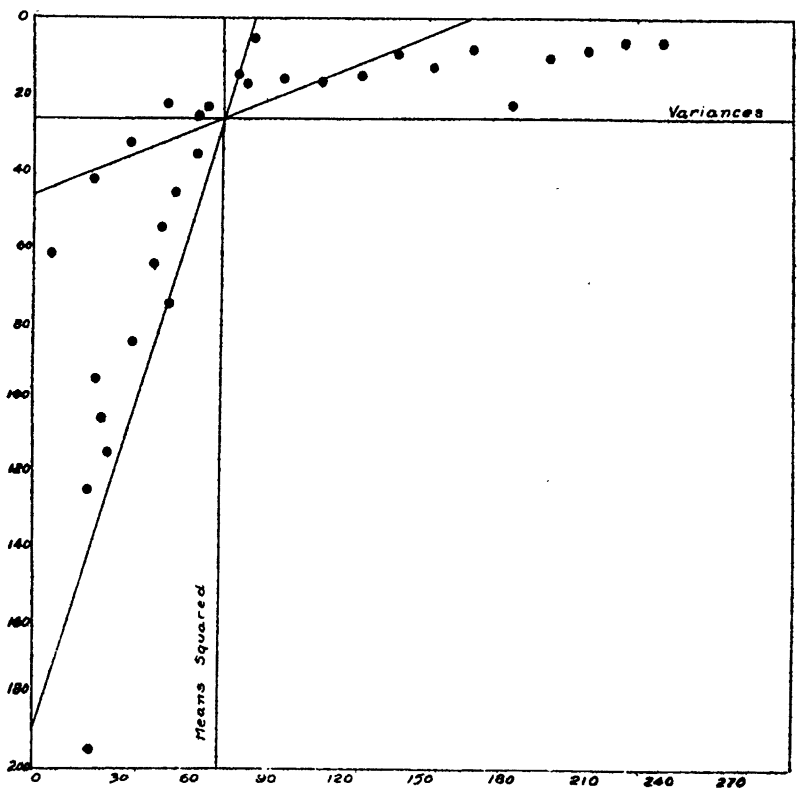
TABLE II

Correlation Table Showing the Relation Between the Means Squared and Variances of Samples of Four from Population I

Means Squared of Samples	*Variances of Samples*																	
	0 to 15	15 to 30	30 to 45	45 to 60	60 to 75	75 to 90	90 to 105	105 to 120	120 to 135	135 to 150	150 to 165	165 to 180	180 to 195	195 to 210	210 to 225	225 to 240	240 to 255	T'ls
0 to 10	13	25	42	64	59	58	83	56	25	22	12	10	3	2	3	2	1	480
10 to 20	7	21	10	18	24	19	14	15	6	3	8	4	2	2				154
20 to 30	5	14	13	21	16	15	11	9	2									106
30 to 40	4	9	13	12	17	9	6	5	2	3								80
40 to 50	7	7	11	6	5	4	2	5	3				1					51
50 to 60	7	5	6	2	5			4	2									31
60 to 70	10	6	6	4	6	2	5	2										41
70 to 80	6	6	3	2	1	1	2	2			1		1					25
80 to 90	5	3	1		2	2	1											14
90 to 100	5	3	2	2														12
100 to 110	4	6	1	2			1											14
110 to 120	3	1	1				1											6
120 to 130			1															2
130 to 140	5	2	2	1	1													11
140 to 150	2	1																3
150 to 160																		
160 to 170	1																	1
170 to 180		2	1															3
180 to 190	2																	2
190 to 200																		
220 to 230				1														1
250 to 260		1																1
Totals	87	112	113	135	136	110	126	98	40	28	21	14	7	4	4	2	1	1038

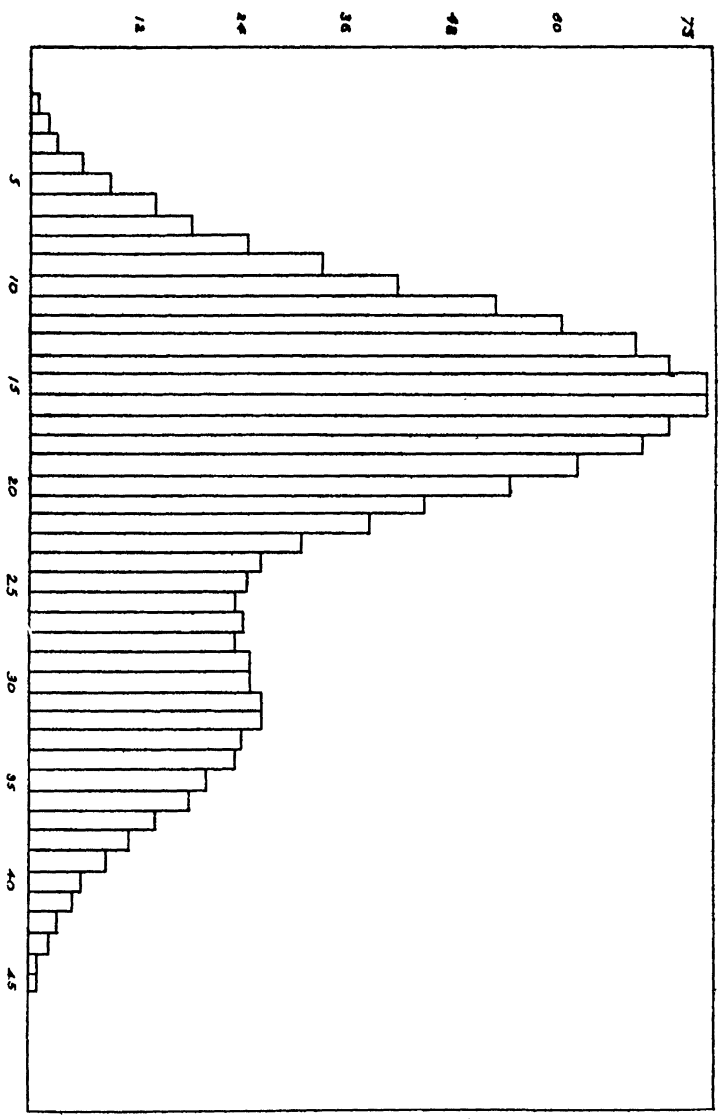
CHART II

The Means of Arrays and Regression Lines of the Means Squared and Variances of Samples from Population I



NOTE: The last thirteen class intervals of the means squared are grouped into one group.

CHART B



Population II, from Which the 1058 Samples of Four of Tables III and IV Were Drawn

TABLE III

Correlation Table Showing the Relation Between the Means and Variances of Samples of Four from Population II

Means of Samples	Variances of Samples																	
	0 to 15	15 to 30	30 to 45	45 to 60	60 to 75	75 to 90	90 to 105	105 to 120	120 to 135	135 to 150	150 to 165	165 to 180	180 to 195	195 to 210	210 to 225	225 to 240	240 to 255	T'ls
15 to 17	1		1															2
13 to 15		2	1															3
11 to 13			2				1	1										4
9 to 11	2	2	3		5	1	0	4					1					18
7 to 9	3	3	10	4	6	7	2	4	5	4		1						49
5 to 7	5	3	13	17	6	10	4	6	10	2	2	4			1			83
3 to 5	4	11	12	18	24	12	18	9	2	6	3	2						121
1 to 3	9	13	21	22	19	27	7	7	5	7	6		1		1	1		146
1 to 1	24	26	29	24	16	13	15	8	8	1	4	3	1		1			173
− 3 to − 1	45	36	26	22	14	14	11	4	5	1	1	1	1			1		182
− 5 to − 3	59	36	21	15	10	8			1	2								152
− 7 to − 5	36	29	7	7	2	1	1	1			1							85
− 9 to − 7	14	9	10		1													34
−11 to − 9	3	1																4
−13 to −11	1	1																2
−15 to −13																		
−17 to −15																		
Totals	206	172	156	129	103	93	59	44	36	23	17	11	4		3	2		1058

CHART III

The Means of Arrays and Regression Lines of the Means and Variances of Samples from Population II

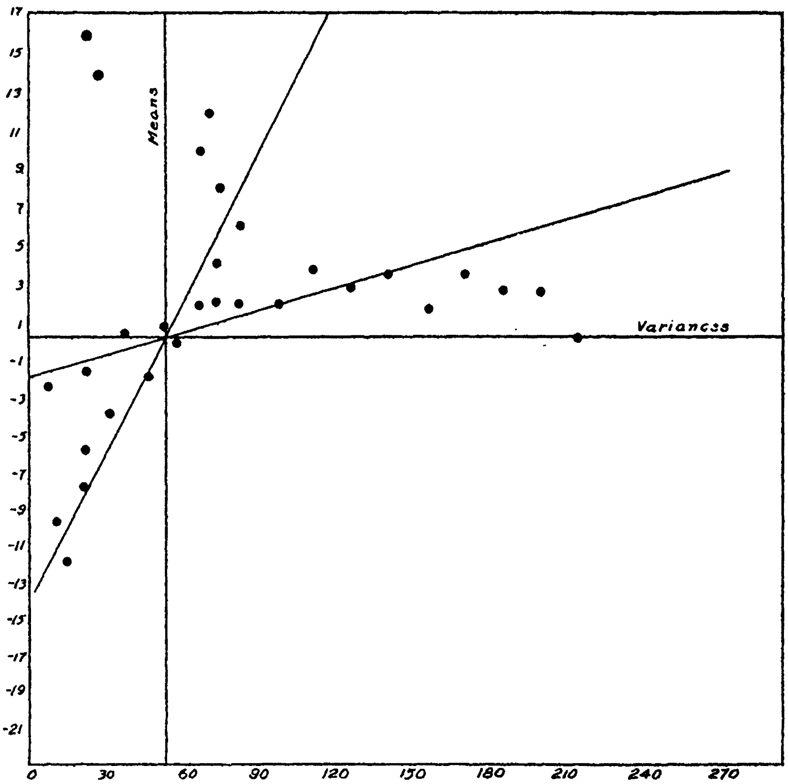


TABLE IV

Correlation Table Showing the Relation Between the Means Squared and Variances of Samples of Four from Population II

Means Squared of Samples	*Variances of Samples*																
	0 to 15	15 to 30	30 to 45	45 to 60	60 to 75	75 to 90	90 to 105	105 to 120	120 to 135	135 to 150	150 to 165	165 to 180	180 to 195	195 to 210	210 to 225	225 to 240	240 to 255	T'ls
0 to 15	117	101	93	89	68	67	46	24	19	13	12	5	4		2	2		661
15 to 30	36	35	24	21	17	12	5	7	5	5	3	2						172
30 to 45	26	16	7	15	5	6	5	3	7		2	3			1			96
45 to 60	10	9	14	2	6	2	1	2	1	2	0	1						50
60 to 75	5	4	9	2	1	3	1	3	3	2								33
75 to 90	6	3	3		2	2		3	1	1								21
90 to 105	3		2		2	1		1					1					10
105 to 120	1	1			2													4
120 to 135	1																	2
135 to 150		1					1											2
150 to 165			2															2
165 to 180			1															1
180 to 195		1																1
195 to 210																		
210 to 225		1																1
225 to 240	1		1															2
240 to 255																		
Totals	206	172	156	129	103	93	59	44	36	23	17	11	4		3	2		1058

CHART IV

The Means of Arrays and Regression Lines of the Means Squared and Variances of Samples from Population II

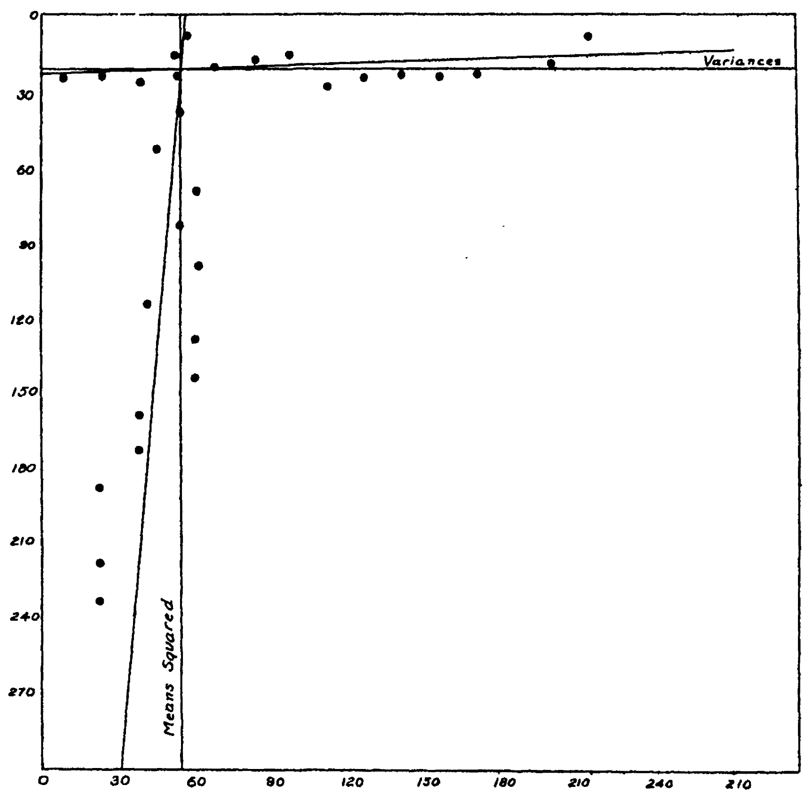


TABLE V

Correlation Coefficients of Tables I-IV

	Correlation-Coefficient	
Number of Table	Theoretical	Actual[1]
I	.00	-.05
II	-.34	-.37
III	.40	.37
IV	-.07	-.05

TABLE VI

Constants of the Marginal Distributions of Tables I-IV in Terms of Class Intervals

Marginal Distribution	Mean	Standard Deviation
Means of Samples from Population I	.252[2]	2.467
Variances of Samples from Population I	4.890[3]	2.900
Means Squared of Samples from Population I	3.591[3]	3.203
Means of Samples from Population II	.07	2.237
Variances of Samples from Population II	3.570[3]	2.854
Means Squared of Samples from Population II	1.408[3]	1.744

[1]Calculated without corrections for grouping.

[2]Is so far from zero because of the groupings employed. Many means were exactly odd integers. These were all put forward into higher classes, making the calculated mean too large.

[3]Origin taken at the beginning of the range.

From the results for the case of samples of two and from the results of empirical sampling, it seems clear that the simplest regression relation that is generally applicable to the means and variances, means squared and variances, of samples from populations which are the combinations of normal populations is parabolic. For small samples and for certain values of the parameters of the sampled population the regression relations may involve exponential terms that are quite important. As the size of the samples increases, it is expected that this exponential term will decrease in influence. It seems plausible that even with large samples the regression relation of means and variances, means squared and variances will remain essentially parabolic. It is not expected that the determination of a good approximation to the regression relations will serve to give an adequate notion of the probability relations of the means and variances, means squared and variances of samples from a population represented by (1), because the arrays may vary in number of modes, in skewness, in dispersion, and in other characteristics. For instance, surface (7) may be trimodal so that arrays may be bimodal or unimodal, and in such a case the arrays must vary markedly. Surfaces (7) and (10) with 2 replaced by n and with the terms suitably weighted are valuable approximations to the probability relations of the means and variances, means squared and variances of samples drawn from a population represented by (1).

G. A. Baker