XX.—On the Theory of Statistical Regression.

By **M. S. Bartlett,** B.A., B.Sc. (Queens' College, Cambridge). *Communicated by* J. WISHART, M.A., D.Sc.

(MS. received May 12, 1933. Read June 5, 1933.)

1. The product moment distribution in the general case of p normal variates, obtained in 1928 (1), and again in 1933 (2), has been awaiting further analysis. Some indication has already been given (Wishart, 1928) that new results might be expected from it; in the particular case of two variates obtained previously by Fisher (3), it has been used to deduce the distributions of the correlation coefficient (3), co-variance (4), and regression coefficient (5). In the general case, it has been used by Wilks (6) to furnish a proof of Fisher's distribution of the multiple correlation coefficient (7), and also in connection with his idea of a generalized variance (8). Further analysis appears to be most fruitful in studying statistical regression in general. It is shown in Part I of this paper that the product moment distribution can be split up into a chain of independent factors. Most of the known distributions related to regression or partial correlation are simply obtained, in a manner which clearly indicates the relations they bear to one another; the distribution of a partial regression coefficient of any order is also readily derived.

In Part II it is pointed out that the assumption of a normal system is not altogether necessary for some of the distributions to hold. A distribution which may be regarded as a further generalization of the product moment distribution, being the generalized partial product moment distribution, is obtained *ab initio*, in order to show what are the minimum assumptions about normality necessary for the various distributions obtained in I.

A note should be made here on the notation used. This follows Yule (9) with regard to the symbols for partial correlations, partial variates, etc. Further, the convention that Greek and English letters are to be used for true and estimated values respectively is employed as far as possible not only for parameters but also for variates. Thus if ξ denotes the value of a variate measured from its true mean, the estimate of ξ given by $\xi - \Sigma\xi/n$ (where Σ, unless otherwise specified, will always denote summation over the n observations in a sample) will be denoted provisionally by x (usually this has been called $x - \bar{x}$). We may write further, for interdependent variates ξ_1, ξ_2, ξ_3, assuming the regressions to be linear,

$$\begin{cases} \xi_{2\cdot1} = \xi_2 - \beta_{21}\xi_1, \\ \xi_{3\cdot21} = \xi_3 - \beta_{32\cdot1}\xi_2 - \beta_{31\cdot2}\xi_1, \end{cases}$$

where β_{21} is the true simple regression coefficient of ξ_2 on ξ_1, $\beta_{32\cdot1}$ the true partial regression coefficient of ξ_3 on ξ_2 for constant ξ_1, etc. Similarly

$$\begin{cases} x_{2\cdot1} = x_2 - b_{21}x_1, \\ x_{3\cdot21} = x_3 - b_{32\cdot1}x_2 - b_{31\cdot2}x_1, \end{cases}$$

where b_{21}, $b_{32\cdot1}$, etc., are the corresponding estimated regression coefficients from the sample.

The product sums $c_{\mu\nu}$ obtained from the sample are defined by the equation

$$c_{\mu\nu} = \Sigma(x_\mu x_\nu), \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and correspondingly we write

$$\gamma_{\mu\nu} = \Sigma(\xi_\mu \xi_\nu). \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

As an extension of this notation, we write further

$$\begin{cases} c_{\mu\nu\cdot1} = \Sigma(x_{\mu\cdot1}x_{\nu\cdot1}), \\ \gamma_{\mu\nu\cdot1} = \Sigma(\xi_{\mu\cdot1}\xi_{\nu\cdot1}), \text{ etc.} \end{cases}$$

I. An Analysis of the Product Moment Distribution.

2. The distribution of p normal variates will be written

$$\mathrm{U}_p(\xi_\mu) \equiv \pi^{-\frac{1}{2}p} \,|\, \mathrm{A} \,|^{\frac{1}{2}} e^{-\mathrm{A}(\xi,\,\xi)} \prod_{\mu=1}^{p} d\xi_\mu, \quad . \quad . \quad . \quad (3)$$

where $\mathrm{A}(\xi, \xi)$ is the positive definite quadratic form of the ξ_μ, with matrix $\mathrm{A} \equiv (a_\mu) = (\frac{1}{2}\Delta_{\mu\nu}/\sigma_\mu\sigma_\nu\Delta)$, Δ being the determinant of correlations $\rho_{\mu\nu}$, and $\Delta_{\mu\nu}$ the co-factor of $\rho_{\mu\nu}$ in Δ. The determinant of A is $|\,\mathrm{A}\,|$. Similarly we write $\mathrm{C} \equiv (c_{\mu\nu})$, $|\,\mathrm{C}\,| \equiv |\,c_{\mu\nu}\,|$.

The distribution of product moments is more conveniently for the present purpose regarded as a distribution of product sums $c_{\mu\nu}$. It has been shown to be independent of the distribution of the means, and the complete distribution may be written

$$\mathrm{U}_p(u_\mu)\mathrm{V}_p(c_{\mu\nu}, n-1), \quad . \quad . \quad . \quad . \quad (4)$$

where $u_\mu = \Sigma\xi_\mu/\sqrt{n}$, and

$$\mathrm{V}_p(c_{\mu\nu}, n-1) \equiv \frac{|\,\mathrm{A}\,|^{\frac{1}{2}(n-1)}\,|\,\mathrm{C}\,|^{\frac{1}{2}(n-p-2)} \exp\left(-\sum_{\mu,\,\nu=1}^{p} a_{\mu\nu}c_{\mu\nu}\right)}{\pi^{\frac{1}{4}p(p-1)} \prod_{r=1}^{p} \Gamma\frac{1}{2}(n-r)} \prod_{1 \leq \mu \leq \nu \leq p} dc_{\nu\mu}. \quad (5)$$

We write the distribution (5) $V_p(c_{\mu\nu}, n-1)$ to show that one degree of freedom has been lost by using the variates x_μ in place of ξ_μ, the distribution of $\gamma_{\mu\nu}$ being given by $V_p(\gamma_{\mu\nu}, n)$.

The complete distribution (4), obtained originally (1) by geometrical methods, has recently been obtained by another method (2), as the result of a certain multiple integral. The fact that this integral was evaluated (10) as a repeated integral has suggested that $V_p(c_{\mu\nu})$ can be split up into a product of independent distributions. This analysis of (5) is carried out in the case of two and three variates, and in the general case of p variates.

3. *Two Variates.*—In the case of two variates, we have

$$V_2(c_{\mu\nu}, n-1) \equiv \frac{|\,C\,|^{\frac{1}{2}(n-4)} e^{-\frac{1}{2(1-\rho_{12}^2)}\left(\frac{c_{11}}{\sigma_1^2} - \frac{2\rho_{12}c_{12}}{\sigma_1\sigma_2} + \frac{c_{22}}{\sigma_2^2}\right)}}{\pi^{\frac{1}{2}}\left[2\sigma_1\sigma_2\sqrt{\left(1-\rho_{12}^2\right)}\right]^{n-1}\Gamma\frac{1}{2}(n-1)\Gamma\frac{1}{2}(n-2)} dc_{11}dc_{12}dc_{22}. \tag{6}$$

Change the variates c_{11}, c_{12}, c_{22} to new variates c_{11}, $c_{22\cdot1}$, $u_{2\cdot1}$, where

$$c_{22\cdot1} = \Sigma x_{2\cdot1}^2 = \Sigma(x_2 - b_{21}x_1)^2 = |\,C\,|/c_{11},$$

since

$$b_{21} = \Sigma(x_2x_1)/\Sigma x_1^2,$$

and

$$u_{2\cdot1} = \sqrt{c_{11}}(b_{21} - \beta_{21}).$$

Then

$$dc_{11}dc_{22\cdot1}du_{2\cdot1} = dc_{11}dc_{12}dc_{22}/\sqrt{c_{11}},$$

and

$$\frac{1}{\left(1-\rho_{12}^2\right)}\left(\frac{c_{11}}{\sigma_1^2} - \frac{2\rho_{12}c_{12}}{\sigma_1\sigma_2} + \frac{c_{22}}{\sigma_2^2}\right) = \frac{c_{11}}{\sigma_1^2} + \frac{\Sigma(x_2 - \beta_{21}x_1)^2}{\sigma_2^2\left(1-\rho_{12}^2\right)}$$

$$= \frac{c_{11}}{\sigma_1^2} + \frac{c_{22\cdot1}}{\sigma_{2\cdot1}^2} + \frac{u_{2\cdot1}^2}{\sigma_{2\cdot1}^2},$$

where $\sigma_{2\cdot1}$ is the standard deviation of $\xi_{2\cdot1}$. Hence

$$V_2(c_{\mu\nu}, n-1) = \left[\frac{c_{11}^{\frac{1}{2}(n-3)} e^{-\frac{1}{2}c_{11}/\sigma_1^2} dc_{11}}{(2\sigma_1^2)^{\frac{1}{2}(n-1)}\Gamma\frac{1}{2}(n-1)}\right]\left[\frac{c_{22\cdot1}^{\frac{1}{2}(n-4)} e^{-\frac{1}{2}c_{22\cdot1}/\sigma_{2\cdot1}^2} dc_{22\cdot1}}{\left(2\sigma_{2\cdot1}^2\right)^{\frac{1}{2}(n-2)}\Gamma\frac{1}{2}(n-2)}\right]\left[\frac{e^{-\frac{1}{2}u_{2\cdot1}^2/\sigma_{2\cdot1}^2} du_{2\cdot1}}{\left(2\pi\sigma_{2\cdot1}^2\right)^{\frac{1}{2}}}\right], \tag{7}$$

$$= V_1(c_{11}, n-1)V_1(c_{22\cdot1}, n-2)U_1(u_{2\cdot1}).$$

We thus obtain the following results.

(i) $V_2(c_{\mu\nu})$ has been split up into three independent distributions. The particular function $u_{2\cdot1} = \sqrt{c_{11}}(b_{21} - \beta_{21})$ is normally distributed with standard deviation $\sigma_{2\cdot1}$. Further, while c_{11} is distributed like γ_{11}, but with one less degree of freedom (see, for example (2), p. 4), $c_{22\cdot1}$ is distributed like $\gamma_{22\cdot1}$, but with two less degrees of freedom. We may in fact consider

$$V_1(c_{22\cdot1}, n-2)U_1(u_{2\cdot1}),$$

write $c_{22\cdot 1}+u_{2\cdot 1}^2=y$, $u_{2\cdot 1}^2=yz$, say, and integrate for z. We then have the distribution of $y=\Sigma(x_2-\beta_{21}x_1)^2$ given by

$$\mathrm{V}_1(y, n-1)\int_0^1 \frac{\Gamma\frac{1}{2}(n-1)}{\pi^{\frac{1}{2}}\Gamma\frac{1}{2}(n-2)}(1-z)^{\frac{1}{2}(n-4)}z^{-\frac{1}{2}}dz=\mathrm{V}_1(y, n-1),$$

showing that the only effect of the substitution of b_{21} for β_{21} is that the distribution loses one degree of freedom. Our estimate of $\sigma_{2\cdot 1}^2$ is thus

$$s_{2\cdot 1}^2=c_{22\cdot 1}/(n-2).$$

(ii) If we consider

$$\mathrm{V}_1(c_{11}, n-1)\mathrm{U}_1(u_{2\cdot 1}),$$

write $u_{2\cdot 1}=\lambda\sqrt{c_{11}}$, and integrate out for c_{11}, we have $f(\lambda)d\lambda$, say,

$$=\int_0^\infty \frac{c_{11}^{\frac{1}{2}(n-)2}e^{-\frac{1}{2}c_{11}\left(\frac{1}{\sigma_1^2}+\frac{\lambda^2}{\sigma_{2\cdot 1}^2}\right)}d\lambda dc_{11}}{(2\sigma_1^2)^{\frac{1}{2}(n-1)}(2\pi\sigma_{2\cdot 1}^2)^{\frac{1}{2}}\Gamma\frac{1}{2}(n-1)}$$

$$=\frac{\Gamma\frac{1}{2}n}{\pi^{\frac{1}{2}}\Gamma\frac{1}{2}(n-1)}\frac{\sigma_1}{\sigma_{2\cdot 1}}\left(1+\frac{\lambda^2\sigma_1^2}{\sigma_{2\cdot 1}^2}\right)^{-\frac{1}{2}n}d\lambda, \quad . \quad . \quad . \quad (8)$$

or writing $\lambda=b_{21}-\beta_{21}$, and $\sigma_{2\cdot 1}^2=\sigma_2^2(1-\rho_{12}^2)$, we have finally for the distribution of the simple regression coefficient b_{21},

$$f(b_{21})db_{21}=\frac{(1-\rho_{12}^2)^{\frac{1}{2}(n-1)}\Gamma\frac{1}{2}n}{\pi^{\frac{1}{2}}\Gamma\frac{1}{2}(n-1)}\frac{\sigma_1}{\sigma_2}\left(1-2\rho_{12}\frac{\sigma_1}{\sigma_2}b_{21}+\frac{\sigma_1^2}{\sigma_2^2}b_{21}^2\right)^{-\frac{1}{2}n}db_{21}. \quad . \quad (9)$$

This distribution was first obtained independently by Romanovsky and Pearson (5).

(iii) If instead we again consider

$$\mathrm{V}_1(c_{22\cdot 1}, n-2)\mathrm{U}_1(u_{2\cdot 1}),$$

this being likely to lead to a more useful result, since $c_{22\cdot 1}$ and $u_{2\cdot 1}$ are both related to the same variate $\xi_{2\cdot 1}$, we have a normal variate $u_{2\cdot 1}$ with standard deviation $\sigma_{2\cdot 1}$, of which our independent estimate is $s_{2\cdot 1}=\sqrt{\{c_{22\cdot 1}/(n-2)\}}$.

It at once follows that the distribution of

$$t=u_{2\cdot 1}/s_{2\cdot 1}$$

is given by the well-known "t distribution," with $n-2$ degrees of freedom. This enables us to test the significance of an estimate b_{21} from any hypothetical value β_{21}. This result should be compared with that obtained by Fisher, who has shown (11) that exactly the same test is applicable whatever the distribution of ξ_1, provided ξ_2 is normal for each ξ_1, and we can suppose the set of values ξ_1 in the sample fixed from sample to sample. Any function of ξ_1 and ξ_2, provided it is a linear function of

ξ_2, is then, of course, normally distributed—for example, the regression coefficient b_{21} itself. In II the question of assumptions is reconsidered, and these two apparently distinct problems are shown to be special cases, the only condition necessary for the test to hold being the normality of $\xi_{2\cdot 1}$.

If we consider the particular case when β_{21} or ρ_{21} is zero, we have

$$t = r_{12}\sqrt{(n-2)}/\sqrt{(1-r_{12}^2)},$$

where r_{12} is our estimate of ρ_{12}. This explains why this particular function of r_{12} is distributed in a "t distribution," the test of significance of a correlation from zero being more fundamentally the test of significance of a regression from zero.

4. *Three Variates.*—We have in the case of three variates

$$V_3(c_{\mu\nu}, n-1) \equiv \frac{|C|^{\frac{1}{2}(n-5)} \exp\left\{-\frac{1}{2\Delta}\sum_{\mu,\nu=1}^{3}\frac{c_{\mu\nu}\Delta_{\mu\nu}}{\sigma_\mu\sigma_\nu}\right\} dc_{11}dc_{12}\ldots dc_{33}}{\pi^{\frac{3}{2}}(2^3\sigma_1{}^2\sigma_2{}^2\sigma_3{}^2\Delta)^{\frac{1}{2}(n-1)}\Gamma\frac{1}{2}(n-1)\Gamma\frac{1}{2}(n-2)\Gamma\frac{1}{2}(n-3)}. \qquad (10)$$

First write

$$\begin{cases} c_{22\cdot 1} = \Sigma x_{2\cdot 1}^2 = \Sigma(x_2 - b_{21}x_1)^2 = C_{33}/c_{11}, \\ c_{23\cdot 1} = \Sigma(x_{2\cdot 1}x_{3\cdot 1}) = \Sigma(x_2 - b_{21}x_1)(x_3 - b_{31}x_1) = -C_{23}/c_{11}, \\ c_{33\cdot 1} = \Sigma x_{3\cdot 1}^2 = \Sigma(x_3 - b_{31}x_1)^2 = C_{22}/c_{11}, \end{cases}$$

and

$$u_{2\cdot 1} = \sqrt{c_{11}}(b_{21} - \beta_{21}), \qquad u_{3\cdot 1} = \sqrt{c_{11}}(b_{31} - \beta_{31}).$$

Then

$$dc_{11}dc_{12}\ldots dc_{33} = c_{11}du_{2\cdot 1}du_{3\cdot 1}dc_{11}dc_{22\cdot 1}dc_{23\cdot 1}dc_{33\cdot 1},$$

and

$$\sum_{\mu,\nu=1}^{3}\frac{c_{\mu\nu}\Delta_{\mu\nu}}{\sigma_\mu\sigma_\nu\Delta} = \frac{c_{11}}{\sigma_1{}^2} + \sum_{\mu,\nu=2}^{3}\left(\frac{\Delta_{\mu\nu}}{\sigma_\mu\sigma_\nu\Delta}\Sigma(x_\mu - \beta_{\mu 1}x_1)(x_\nu - \beta_{\nu 1}x_1)\right)$$
$$= \frac{c_{11}}{\sigma_1{}^2} + \sum_{\mu,\nu=2}^{3}\left(\frac{\Delta_{\mu\nu}}{\sigma_\mu\sigma_\nu\Delta}[c_{\mu\nu\cdot 1} + u_{\mu\cdot 1}u_{\nu\cdot 1}]\right).$$

For this identity depends on the coefficients of c_{12}, c_{13}, and c_{11} being respectively equal on each side. For the first two we must have

$$\Delta_{12} + \rho_{12}\Delta_{22} + \rho_{13}\Delta_{32} = 0,$$
$$\Delta_{13} + \rho_{13}\Delta_{33} + \rho_{12}\Delta_{23} = 0,$$

which (since $\rho_{11} = 1$) are obviously true. Multiply the first of these by ρ_{12}, the second by ρ_{13}, and add; we get

$$\Delta - \Delta_{11} + \rho_{12}^2\Delta_{22} + 2\rho_{12}\rho_{13}\Delta_{23} + \rho_{13}^2\Delta_{33} = 0,$$

which is the identity required for the coefficient of c_{11}. We have also *

$$|C| = c_{11}\begin{vmatrix} c_{22\cdot 1} & c_{23\cdot 1} \\ c_{23\cdot 1} & c_{33\cdot 1} \end{vmatrix}.$$

* *Cf.* Ingham (10), p. 5.

It readily follows that

$$V_3(c_{\mu\nu},\ n-1)=V_1(c_{11},\ n-1)V_2(c_{\mu\nu\cdot 1},\ n-2)U_2(u_{\mu\cdot 1}), \quad . \quad . \quad . \qquad (11)$$

where μ, ν on the right-hand side can take the values 2 and 3.

Since $-\Delta_{32}/\sqrt{(\Delta_{22}\Delta_{33})}=\rho_{32\cdot 1}$—the correlation between $\xi_{3\cdot 1}$ and $\xi_{2\cdot 1}$, or the partial correlation between ξ_3 and ξ_2—and we have also $\Delta=(1-\rho^2_{32\cdot 1})\Delta_{22}\Delta_{33}$, we may write

$$V_2(c_{\mu\nu\cdot 1},\ n-2)=\frac{|\,c_{\mu\nu\cdot 1}\,|^{\frac{1}{2}(n-5)}e^{-\frac{1}{2(1-\rho^2_{32\cdot 1})}\left(\frac{c_{22\cdot 1}}{\sigma^2_{2\cdot 1}}-\frac{2\rho_{32\cdot 1}c_{32\cdot 1}}{\sigma_{2\cdot 1}\sigma_{3\cdot 1}}+\frac{c_{33\cdot 1}}{\sigma^2_{3\cdot 1}}\right)}}{\pi^{\frac{1}{2}}\left[2\sigma_{2\cdot 1}\sigma_{3\cdot 1}\sqrt{\left(1-\rho^2_{32\cdot 1}\right)}\right]^{n-2}\Gamma\frac{1}{2}(n-2)\Gamma\frac{1}{2}(n-3)}dc_{22\cdot 1}dc_{32\cdot 1}dc_{33\cdot 1}. \qquad (12)$$

If we compare this result with (6), we see that the quantities $c_{22\cdot 1}$, $c_{32\cdot 1}$, $c_{33\cdot 1}$ are jointly distributed exactly like c_{22}, c_{23}, c_{33}, but with one less degree of freedom. We may thus define partial variances and co-variance by the equations

$$\begin{cases} v_{22\cdot 1}=s^2_{2\cdot 1}=\Sigma x^2_{2\cdot 1}/(n-2) \\ v_{32\cdot 1}=r_{32\cdot 1}s_{2\cdot 1}s_{3\cdot 1}=\Sigma(x_{3\cdot 1}x_{2\cdot 1})/(n-2) \\ v_{33\cdot 1}=s^2_{3\cdot 1}=\Sigma x^2_{3\cdot 1}/(n-2). \end{cases}$$

We thus have the further results that not only, as we saw in the case of two variates, is $v_{22\cdot 1}$ distributed in an "s^2 distribution" with $n-2$ degrees of freedom, but also

(iv) the partial co-variance $v_{32\cdot 1}$ will be distributed in the Bessel function distribution (4) and (12) obtained for v_{32} or v_{21}, and the partial correlation coefficient $r_{32\cdot 1}$ will be distributed exactly like a simple correlation coefficient r_{32} or r_{21}, each, however, with one less degree of freedom. Fisher has shown this for the important case of the partial correlation coefficient by means of a geometrical argument (13).

We should notice that $r_{32\cdot 1}$ is defined above as

$$r_{32\cdot 1}=\Sigma(x_{3\cdot 1}x_{2\cdot 1})/\sqrt{\left[\Sigma\left(x^2_{2\cdot 1}\right)\Sigma\left(x^2_{3\cdot 1}\right)\right]},$$

but this may be written

$$r_{32\cdot 1}=(r_{32}-r_{31}r_{21})/\sqrt{\left[\left(1-r^2_{31}\right)\left(1-r^2_{21}\right)\right]}.$$

To reduce (11) further, we now treat $V_2(c_{\mu\nu\cdot 1},\ n-2)$ exactly as we did $V_2(c_{\mu\nu},\ n-1)$ in the case of two variates. That is, write

$$\begin{aligned} c_{33\cdot 12}&=(c_{33\cdot 1}c_{22\cdot 1}-c^2_{23\cdot 1})/c_{22\cdot 1}=|\,C\,|/C_{33} \\ &=\Sigma x^2_{3\cdot 12}=\Sigma(x_3-b_{32\cdot 1}x_2-b_{31\cdot 2}x_1)^2, \\ u_{3\cdot 21}&=\sqrt{c_{22\cdot 1}}(b_{32\cdot 1}-\beta_{32\cdot 1}). \end{aligned}$$

We then have

$$\begin{aligned} V_3(c_{\mu\nu},\ n-1)=&\,V_1(c_{11},\ n-1)V_1(c_{22\cdot 1},\ n-2)V_1(c_{33\cdot 12},\ n-3) \\ &\times U_2(u_{2\cdot 1},\ u_{3\cdot 1})U_1(u_{3\cdot 21}). \quad . \quad . \quad . \quad . \quad . \end{aligned} \qquad (13)$$

Besides furnishing again all the results obtained for two variates, the case of three variates now gives the following.

(v) The quantity $c_{33\cdot 12}$ is distributed like $\gamma_{33\cdot 12}$, but with three less degrees of freedom. We may thus write $c_{33\cdot 12}=(n-3)v_{33\cdot 12}$, where $v_{33\cdot 12}$ is our estimate of the variance of $\xi_{3\cdot 12}$, that is, of ξ_3 from the regression plane

$$\xi_3-\beta_{32\cdot 1}\xi_2-\beta_{31\cdot 2}\xi_1=0.$$

Further, the quantities $u_{2\cdot 1}$, $u_{3\cdot 1}$, $u_{3\cdot 21}$ are normally distributed. Notice that since the correlation between $u_{2\cdot 1}$ and $u_{3\cdot 1}$ is $\rho_{32\cdot 1}$, we can write $w_1=u_{3\cdot 1}-\beta_{32\cdot 1}u_{2\cdot 1}=\sqrt{c_{11}}(b_{31}-\beta_{32\cdot 1}b_{21}-\beta_{31\cdot 2})$, $w_2=u_{3\cdot 21}$, and obtain three independent normal variates $u_{2\cdot 1}$, w_1, and w_2; moreover, w_1 and w_2 have the same standard deviation $\sigma_{3\cdot 21}$.

(vi) If we consider

$$V_1(c_{22\cdot 1},\ n-2)U_1(u_{3\cdot 21}),$$

we may obtain the distribution of the partial regression coefficient $b_{32\cdot 1}$, exactly as we found the distribution of the simple regression coefficient b_{21} by considering

$$V_1(c_{11},\ n-1)U_1(u_{2\cdot 1}).$$

The only point to notice is that one degree of freedom has been lost, so that if we write $u_{3\cdot 21}=\lambda\sqrt{c_{22\cdot 1}}$, and integrate out for $c_{22\cdot 1}$, we shall have

$$f(\lambda)d\lambda=\frac{\Gamma\frac{1}{2}(n-1)}{\pi\Gamma\frac{1}{2}(n-2)}\frac{\sigma_{2\cdot 1}}{\sigma_{3\cdot 21}}\left(1+\frac{\lambda^2\sigma^2_{2\cdot 1}}{\sigma^2_{3\cdot 21}}\right)^{-\frac{1}{2}(n-1)}d\lambda \quad . \quad . \quad (14)$$

analogously to (8), where here $\lambda=b_{32\cdot 1}-\beta_{32\cdot 1}$. If we write $\sigma^2_{3\cdot 21}=\sigma^2_{3\cdot 1}(1-\rho^2_{32\cdot 1})$, we have the result analogous to (9),

$$f(b_{32\cdot 1})db_{32\cdot 1}=\frac{(1-\rho^2_{32\cdot 1})^{\frac{1}{2}(n-2)}\Gamma\frac{1}{2}(n-1)}{\pi^{\frac{1}{2}}\Gamma\frac{1}{2}(n-2)}\frac{\sigma_{2\cdot 1}}{\sigma_{3\cdot 1}}\left(1-2\rho_{32\cdot 1}\frac{\sigma_{2\cdot 1}}{\sigma_{3\cdot 1}}b_{32\cdot 1}+\frac{\sigma^2_{2\cdot 1}}{\sigma^2_{3\cdot 1}}b^2_{32\cdot 1}\right)^{(n-1)}db_{32\cdot 1} \quad (15)$$

Similarly we obtain the distribution of $b_{31\cdot 2}$ by interchanging the suffixes 1 and 2, but the two distributions will not, of course, be independent.

(vii) Like the distribution of b_{21}, this distribution is of little value for testing the significance of regression coefficients, since it contains unknowns besides $\beta_{32\cdot 1}$ It is therefore more useful to consider

$$V_1(c_{33\cdot 12},\ n-3)U_1(u_{3\cdot 21}),$$

since now we have simply the normal variate $u_{3\cdot 21}$ with standard deviation $\sigma_{3\cdot 21}$, of which our independent estimate $s_{3\cdot 21}$ is given by

$$s^2_{3\cdot 21}=c_{33\cdot 12}/(n-3).$$

We may thus write

$$t=u_{3\cdot 21}/s_{3\cdot 21},$$

where t follows the "t distribution" with $n-3$ degrees of freedom. This enables us to test the significance of $b_{32\cdot1}$ from any hypothetical value $\beta_{32\cdot1}$, since $\beta_{32\cdot1}$ is the only unknown that occurs.

When $\beta_{32\cdot1}$ or $\rho_{32\cdot1}$ is supposed zero, we have

$$t=r_{32\cdot1}\sqrt{(n-3)}/\sqrt{(1-r^2_{32\cdot1})},$$

and we can test the significance of a partial correlation from zero.

(viii) We had not only $w_2=u_{3\cdot21}$, but also $w_1=u_{3\cdot1}-\beta_{32\cdot1}u_{2\cdot1}$ normally and independently distributed with standard deviation $\sigma_{3\cdot21}$, both being further independent of the distribution of $s^2_{3\cdot21}$. It follows that if we write

$$\begin{aligned}2s^2&=w_1{}^2+w_2{}^2\\&=(b_{31\cdot2}-2\beta_{31\cdot2})c_{13}+(b_{32\cdot1}-2\beta_{32\cdot1})c_{32}\\&\quad+\beta^2_{31\cdot2}c_{11}+2\beta_{32\cdot1}\beta_{31\cdot2}c_{12}+\beta^2_{32\cdot1}c_{22},\end{aligned} \qquad (16)$$

then $z=\frac{1}{2}\log(s^2/s^2_{3\cdot21})$ is distributed in Fisher's "z distribution" with $s^2_{3\cdot21}$ having $n-3$ degrees of freedom, and s^2 two degrees of freedom. This enables us to make a single test for the significance of $b_{32\cdot1}$ and $b_{31\cdot2}$ from hypothetical values $\beta_{32\cdot1}$ and $\beta_{31\cdot2}$. It is clearly a problem in the analysis of variance (see Fisher (14)).

If we suppose $\beta_{32\cdot1}=\beta_{31\cdot2}=0$, we have simply

$$2s^2=b_{32\cdot1}c_{32}+b_{31\cdot2}c_{31},$$

or since R^2, the estimated multiple correlation coefficient of ξ_3 with ξ_1 and ξ_2 is given by

$$R^2=(b_{32\cdot1}c_{32}+b_{31\cdot2}c_{31})/c_{33},$$

we have

$$2s^2/(n-3)s^2_{3\cdot21}=R^2/(1-R^2).$$

Thus, just as the test of significance of a partial regression coefficient $b_{32\cdot1}$ from zero becomes identical with the test of significance of the corresponding partial correlation coefficient $r_{32\cdot1}$ from zero, so the joint test of significance of $b_{32\cdot1}$ and $b_{31\cdot2}$ from zero must become identical with the test whether there is any significant multiple correlation.

5. *The General Case of* p *Variates.*—To obtain the various distributions above in their general form, we consider finally

$$V_p(c_{\mu\nu},\ n-1).$$

Write

$$\begin{cases}c_{\mu\nu\cdot1}=\Sigma(x_{\mu\cdot1}x_{\nu\cdot1})=c_{\mu\nu}-c_{\mu1}c_{\nu1}/c_{11},\\u_{\mu\cdot1}=\sqrt{c_{11}}(b_{\mu1}-\beta_{\mu1}),\end{cases}$$

where the convention is adopted that μ, ν can take all values 1 to p except those occurring explicitly—*e.g.* μ cannot take the value 1 in $c_{\mu\nu\cdot1}$.

Then

$$\Pi\, dc_{\mu\nu} = c_{11}^{\frac{1}{2}p}\Pi\, dc_{\mu\nu\cdot 1}\Pi\, du_{\mu\cdot 1},$$

and

$$\sum_{\mu,\,\nu=1}^{p}\frac{c_{\mu\nu}\Delta_{\mu\nu}}{\sigma_\mu\sigma_\nu\Delta} = \frac{c_{11}}{\sigma_1^{\,2}} + \sum_{\mu,\,\nu=2}^{p}\left(\frac{\Delta_{\mu\nu}}{\sigma_\mu\sigma_\nu\Delta}\Sigma(x_\mu-\beta_{\mu 1}x_1)(x_\nu-\beta_{\nu 1}x_1)\right)$$

$$= \frac{c_{11}}{\sigma_1^{\,2}} + \sum_{\mu,\,\nu=2}^{p}\left(\frac{\Delta_{\mu\nu}}{\sigma_\mu\sigma_\nu\Delta}[c_{\mu\nu\cdot 1}+u_{\mu\cdot 1}u_{\nu\cdot 1}]\right);$$

for similarly to the case of three variates, we have

$$\sum_{r=1}^{p}\rho_{1r}\Delta_{sr} = 0 \quad (s \neq 1),$$

and multiplying this equation by ρ_{1s}, and adding for all $s \neq 1$, we have

$$\Delta - \Delta_{11} + \sum_{r,\,s=2}^{p}\rho_{1r}\rho_{1s}\Delta_{sr} = 0,$$

and these identities are the equations required for the coefficients of $c_{1\mu}$ and c_{11}. We have also

$$|\,C\,| = c_{11}\,|\,c_{\mu\nu\cdot 1}\,|.$$

It follows that

$$V_p(c_{\mu\nu},\, n-1) = V_1(c_{11},\, n-1)V_{p-1}(c_{\mu\nu\cdot 1},\, n-2)U_{p-1}(u_{\mu\cdot 1}).$$

Similarly

$$V_{p-1}(c_{\mu\nu\cdot 1},\, n-2) = V_1(c_{22\cdot 1},\, n-2)V_{p-2}(c_{\mu\nu\cdot 12},\, n-3)U_{p-2}(u_{\mu\cdot 21}),$$

$$\vdots$$

$$V_3(c_{\mu\nu\cdot 1\,\ldots\,p-3},\, n-p+2) = V_1(c_{p-2,\,p-2\cdot 1\,\ldots\,p-3},\, n-p+2)V_2(c_{\mu\nu\cdot 1\,\ldots\,p-2},\, n-p+1) \times U_2(u_{\mu\cdot p-2,\,\ldots\,1}).$$

We may deduce from the distribution

$$V_2(c_{\mu\nu\cdot 1\,\ldots\,p-2},\, n-p+1) \quad . \quad . \quad . \quad . \quad (17)$$

results corresponding to those obtained from (12) in the case of three variates. That is, we may write

$$v_{\mu\nu\cdot 1\,\ldots\,p-2} = c_{\mu\nu\cdot 1\,\ldots\,p-2}/(n-p+1),$$

and obtain that

(ix) the distribution of the partial co-variance $v_{p,\,p-1\cdot 1\,\ldots\,p-2}$ is like that of a simple co-variance, and the distribution of the partial correlation coefficient $r_{p,\,p-1\cdot 1\,\ldots\,p-2}$ like that of a simple correlation coefficient, each distribution, however, with only $n-p+1$ degrees of freedom.

We now write further

$$V_2(c_{\mu\nu\cdot 1\,\ldots\,p-2},\, n-p+1) = V_1(c_{p-1,\,p-1\cdot 1\,\ldots\,p-2},\, n-p+1)V_1(c_{pp\cdot 1\,\ldots\,p-1},\, n-p) \times U_1(u_{p\cdot p-1,\,\ldots\,1}),$$

and hence have

$$V_p(c_{\mu\nu},\, n-1) = V_1(c_{11},\, n-1)\prod_{r=1}^{p-1}\{V_1(c_{\mu\nu\cdot 1\,\ldots\,r},\, n-r-1)U_{p-r}(u_{\mu\cdot r\,\ldots\,1})\}. \quad (18)$$

We may therefore deduce in the general case the following results.

(x) The quantity $c_{pp\cdot 1\ldots p-1}$ is distributed like $\gamma_{pp\cdot 1\ .\ p-1}$, but with p less degrees of freedom. We thus write

$$v_{pp\cdot 1\ldots p-1}=c_{pp\cdot 1\ldots p-1}/(n-p)$$

as our estimate of the variance of ξ_p from its regression "plane."

Further, the quantities $u_{\mu\cdot r\ldots 1}$ are all normally distributed. We may, moreover, write

$$\begin{cases} w_1 = u_{p\cdot 1}-\beta_{p2\cdot 13\ldots p-1}u_{2\cdot 1} \cdots\cdots -\beta_{p,\,p-1\cdot 1\ldots p-2}u_{p-1\cdot 1} \\ w_2 = u_{p\cdot 21}-\beta_{p3\cdot 124\ldots p-1}u_{3\cdot 21} \cdots\cdots -\beta_{p,\,p-1\cdot 1\ldots p-2}u_{p-1\cdot 21} \\ \vdots \\ w_{p-1}=u_{p\cdot p-1,\ldots 1}, \end{cases}$$

and obtain $p-1$ independent normal variates each with the same standard deviation. They may alternatively be written

$$\begin{cases} w_1 = \sqrt{c_{11}}(b_{p1}-\beta_{p,\,p-1\cdot 1\ldots p-2}b_{p-1,\,1} \cdots\cdots -\beta_{p1\cdot 2\ldots p-1}) \\ w_2 = \sqrt{c_{22\cdot 1}}(b_{p2\cdot 1}-\beta_{p,\,p-1\cdot 1\ldots p-2}b_{p-1,\,2\cdot 1} \cdots\cdots -\beta_{p2\cdot 13\ldots p-1}) \\ \vdots \\ w_{p-1}=\sqrt{c_{p-1,\,p-1\cdot 1\ldots p-2}}(b_{p,\,p-1\cdot 1\ldots p-2}-\beta_{p,\,p-1\cdot 1\ldots p-2}), \end{cases} \qquad (19)$$

showing that they contain only the unknowns $\beta_{p1\cdot 2\ldots p-1}, \ldots\ldots,$ $\beta_{p,\,p-1\cdot 1\ldots p-2}$.

(xi) The distribution of the partial regression coefficient

$$b\equiv b_{p,\,p-1\cdot 1\ldots p-2}$$

is obtained exactly as that of b_{21}, $b_{32\cdot 1}$, etc. We consider

$$V_1(c_{p-1,\,p-1\cdot 1\ldots p-2},\ n-p+1)U_1(u_{p\cdot p-1,\ldots 1}),$$

and obtain

$$f(\lambda)d\lambda=\frac{\Gamma\frac{1}{2}(n-p+2)}{\pi^{\frac{1}{2}}\Gamma\frac{1}{2}(n-p+1)}\,\frac{\sigma_{p-1\cdot 1\ldots p-2}}{\sigma_{p\cdot 1\ldots p-1}}\left(1+\frac{\lambda^2\sigma^2_{p-1\cdot 1\ldots p-2}}{\sigma^2_{p\cdot 1\ldots p-1}}\right)^{-\frac{1}{2}(n-p+2)}d\lambda, \qquad (20)$$

where

$$\lambda=b-\beta_{p,p-1\cdot 1\ldots p-2},$$

or

$$f(b)db=\frac{(1-\rho^2)^{\frac{1}{2}(n-p+1)}\Gamma\frac{1}{2}(n-p+2)}{\pi^{\frac{1}{2}}\Gamma\frac{1}{2}(n-p+1)}\,\frac{\sigma'}{\sigma}\left(1-2\rho\frac{\sigma'}{\sigma}b+\left[\frac{\sigma'}{\sigma}b\right]^2\right)^{-\frac{1}{2}(n-p+2)}db, \qquad (21)$$

where $\rho=\rho_{p,\,p-1\cdot 1\ldots p-2}$, $\sigma=\sigma_{p\cdot 1\ldots p-2}$, and $\sigma'=\sigma_{p-1\cdot 1\ldots p-2}$.

(xii) As before, if we consider instead

$$V_1(c_{pp\cdot 1\ldots p-1},\ n-p)U_1(u_{p\cdot p-1,\ldots 1}),$$

we may write

$$t=u_{p\cdot p-1,\ldots 1}/s_{p\cdot p-1,\ldots 1},$$

where $s^2_{p\cdot p-1,\ldots 1}=v_{pp\cdot p-1,\ldots 1}=c_{pp\cdot 1\ldots p-1}/(n-\rho)$, and $u_{p\cdot p-1,\ldots 1}$ $(\equiv w_{p-1})$ is given in (19).

This is therefore the test in the general case for the significance of the partial regression coefficient b from any hypothetical value β. When we put $\beta=0$, we have

$$t=r\sqrt{(n-p)}/\sqrt{(1-r^2)},$$

where r is the estimated partial correlation coefficient between ξ_p and ξ_{p-1} when $\xi_1 \ldots \xi_{p-2}$ are eliminated, and we have therefore a test of significance of r from zero.

(xiii) If we write

$$\begin{aligned}(p-1)s^2 &= w_1^2+w_2^2+\ldots w_{p-1}^2\\ &= \sum_{r=1}^{p-1}(b_{pr\cdot 1\ldots r-1,\,r+1,\ldots p-1}-2\beta_{pr\cdot 1\ldots r-1,\,r+1,\ldots p-1})c_{pr}\\ &\qquad + \sum_{r,\,s=1}^{p-1}\beta_{pr\cdot 1\ldots r-1,\,r+1,\ldots p-1}\beta_{ps\cdot 1\ldots s-1,\,s+1,\ldots p-1}c_{rs},\end{aligned} \qquad (22)$$

then $z=\frac{1}{2}\log(s^2/s^2_{p\cdot 1\ldots p-1})$ is distributed in Fisher's z distribution with $s^2_{p\cdot 1\ldots p-1}$ having $n-p$ degrees of freedom, and s^2 $p-1$ degrees of freedom. This gives, therefore, a single test for the significance of $b_{p1\cdot 2\ldots p-1},\ldots, b_{p,\,p-1\cdot 1\ldots p-2}$ from hypothetical values $\beta_{p1\cdot 2\ldots p-1},\ldots,\beta_{p,\,p-1\cdot 1\ldots p-2}$.

If we suppose these values all zero, we have simply

$$\begin{aligned}(p-1)s^2 &= \sum_{r=1}^{p-1} b_{pr\cdot 1\ldots r-1,\,r+1,\ldots p-1}c_{pr}\\ &= c_{pp}R^2,\end{aligned}$$

where R^2 is the estimated multiple correlation coefficient of ξ_p with $\xi_1, \ldots \xi_{p-1}$; and

$$(p-1)s^2/(n-p)s^2_{p\cdot 1}\ldots_{p-1}=R^2/(1-R^2)=v,$$

say.

The distribution of $v=n_1s_1^2/n_2s_2^2$, where s_1^2 and s_2^2 are independent estimates of a variance σ^2, with n_1 and n_2 degrees of freedom respectively, has been shown by Fisher to be

$$f(v)dv=\frac{\Gamma\frac{1}{2}(n_1+n_2)}{\Gamma\frac{1}{2}n_1\Gamma\frac{1}{2}n_2}\frac{v^{\frac{1}{2}n_1-1}}{(1+v)^{\frac{1}{2}(n_1+n_2)}}dv,$$

from which the z distribution is derived.

Write $n_1=p-1$, $n_2=n-p$, and $v=R^2/(1-R^2)$, and we obtain the well-known distribution (15) of R^2 when there is no real multiple correlation,

$$g(R^2)d(R^2)=\frac{\Gamma\frac{1}{2}(n-1)}{\Gamma\frac{1}{2}(p-1)\Gamma\frac{1}{2}(n-p)}(R^2)^{\frac{1}{2}(p-3)}(1-R^2)^{\frac{1}{2}(n-p-2)}d(R^2). \qquad (23)$$

The joint test of significance of the coefficients $b_{p1\cdot 2\ldots p-1}, \ldots, b_{p,\,p-1\cdot 1\ldots p-2}$ from zero thus becomes identical with the test of the multiple correlation being significant.

II. On the Assumption of Normality.

6. So far the notation has been used that ξ denotes the deviation of a variate from its true mean, and x the deviation from its estimated mean. This was convenient, since it allowed the estimate of $\xi_{2\cdot 1}$ to be written $x_{2\cdot 1}$, etc.

It may now be remarked, however, that if x is given its usual interpretation as the value of an observation, *e.g.*

$$x_1 = \xi_1 + m_1,$$

where m_1 is the true mean of x_1, the notation can be made quite complete by defining a quantity $x_0 \equiv 1$, and writing now $x_{1\cdot 0}$ as our estimate of ξ_1 ($\xi_{1\cdot 0} \equiv \xi_1$, since ξ_1 is already measured from its true mean). For we had previously the relations

$$\begin{cases} x_{\mu\cdot\upsilon} = x_\mu - x_\upsilon \Sigma(x_\mu x_\upsilon)/\Sigma x_\upsilon^2 \\ c_{\mu\mu\cdot\upsilon} = c_{\mu\mu} - c_{\mu\upsilon}^2/c_{\upsilon\upsilon} \\ u_{\mu\cdot\upsilon} = \sqrt{c_{\upsilon\upsilon}}(b_{\mu\upsilon} - \beta_{\mu\upsilon}), \end{cases}$$

and analogously we now have

$$\begin{cases} x_{1\cdot 0} = x_1 - \Sigma x_1/n = x_1 - \bar{x}_1 \\ c_{11\cdot 0} = c_{11} - (\Sigma x_1)^2/n = \Sigma(x_1 - \bar{x}_1)^2 \\ u_{1\cdot 0} = \sqrt{n}(\bar{x}_1 - m_1), \end{cases}$$

since

$$\Sigma(x_1 x_0) = \Sigma x_1 = n\bar{x}_1, \qquad \Sigma x_0^2 = n.$$

We may now write $c_{22\cdot 01}$ instead of $c_{22\cdot 1}$, etc., and in these expressions we shall now have the number of digits referring to variates eliminated giving the number of degrees of freedom lost. The joint distribution of the means and the product sums given by equation (4) will now be written

$$W_{p,\,0} \equiv U_p(u_{\mu\cdot 0})V_p(c_{\mu\upsilon\cdot 0},\, n-1). \qquad . \quad . \quad . \quad (24)$$

7. In the first part of this paper, the distributions arising out of the problem of regression were conveniently derived from the distribution $W_{p,\,0}$, which holds only on the assumption that the system of variates ξ_μ is normal. We have seen, however (using now the complete distribution $W_{p,\,0}$ given by (24), which includes the distribution of the means), that in the case of two variates we may write

$$W_{2,0} = \{V_1(c_{11\cdot 0},\, n-1)U_1(u_{1\cdot 0})\}\{V_1(c_{22\cdot 01},\, n-2)U_1(u_{2\cdot 0} - \beta_{21}u_{1\cdot 0})U_1(u_{2\cdot 10})\}, \quad (25)$$

where the variance of $u_{1\cdot 0}$ is σ_1^2, estimated by $c_{11\cdot 0}/(n-1)$, and the variance of $u_{2\cdot 0} - \beta_{21}u_{1\cdot 0}$ or $u_{2\cdot 10}$ is $\sigma_{2\cdot 1}^2$, estimated by $c_{22\cdot 01}/(n-2)$.

The original normal distribution $U_2(\xi_\mu)$ can of course be written

$$U_2(\xi_\mu) = U_1(\xi_1)U_1(\xi_2 - \beta_{21}\xi_1),$$

and equation (25) suggests that $W_{2,\,0}$, which depends on the normality of ξ_1 and ξ_2, can also be split up into two factors which depend respectively on the normality of ξ_1 and $\xi_{2\cdot 1}$. If we assume for the moment this to be true, we may observe that any distribution or test derived from both factors, such as the distribution of the regression coefficient, of the correlation coefficient, or of the co-variance, will therefore depend on the normality of ξ_1 and $\xi_{2\cdot 1}$, *i.e.* of ξ_1 and ξ_2. On the other hand, any distribution or test *derived from the second factor alone*, such as the distribution of the partial variance, or the test of significance of the regression coefficient, will depend simply on the normality of the partial or residual variate $\xi_{2\cdot 1}$.

For three variates we may write

$$W_{3,\,0} = \{V_1(c_{11\cdot 0},\, n-1)U_1(v_{1\cdot 0})\}\{V_2(c_{\mu\nu\cdot 01},\, n-2)U_2(u_{\mu\cdot 0} - \beta_{\mu 1}u_{1\cdot 0})U_2(u_{\mu\cdot 10})\}, \quad (26)$$

where we may expect the first factor to depend on ξ_1 being normal, the second on $\xi_{2\cdot 1}$, $\xi_{3\cdot 1}$ being normal, and then for the second factor we may write further

$$\{V_1(c_{22\cdot 01},\, n-2)U_1(u_{2\cdot 0} - \beta_{21}u_{1\cdot 0})U_1(u_{2\cdot 1})\}$$
$$\times\{V_1(c_{33\cdot 012},\, n-3)U_1(u_{3\cdot 0} - \beta_{31\cdot 2}u_{1\cdot 0} - \beta_{32\cdot 1}u_{2\cdot 0})U_1(u_{3\cdot 10} - \beta_{32\cdot 1}u_{2\cdot 10})U_1(u_{3\cdot 210})\}, \quad (27)$$

where in this expression we may expect the last complete factor to depend simply on $\xi_{3\cdot 21}$ being normal.

In order to give a formal proof of these results, which makes no assumptions that are not necessary, it is proposed to establish *ab initio* a general distribution which is an extension of the distribution $W_{p,\,0}$ obtained by Wishart, being the corresponding partial distribution when κ of the variates are eliminated. It will be written

$$W_{p,\,\kappa}.$$

The method is to find the moment-generating function of the quantities whose distribution we wish to find. This method was used in a recent paper (2) to obtain the distribution $W_{p,\,0}$, and it is there more fully explained. The work below gives perhaps the most general result obtainable in this way.

8. First, however, it is necessary to give a very brief discussion of moment-generating functions and their properties. The most fundamental definition of the moment-generating function of a quantity ϕ, which is some function of the pn observations in a sample in p variates ξ_μ, is perhaps

$$M(t) = E(e^{it\phi}), \quad (28)$$

where E denotes mathematical expectation—that is, $e^{it\phi}$ is to be averaged for all possible values of ϕ. Thus the moment of the rth order, which may be written $E(\phi^r)$, is the coefficient of $(it)^r/r!$ in the expansion of $M(t)$. We write it so that $M(t)$ is finite for all real t.

If $\phi = x_1 + x_2$, where x_1 and x_2 are any two independent quantities, we have

$$E(e^{it(x_1+x_2)}) = E(e^{itx_1} \,.\, e^{itx_2})$$

$$= E(e^{itx_1})E(e^{itx_2})$$

or

$$M_{x_1+x_2} = M_{x_1} \,.\, M_{x_2}.$$

For n independent quantities $x_1 \ldots x_n$, we have similarly

$$M_{\Sigma x} = \Pi M_{x_r}. \qquad (29)$$

In particular, if $x_1 \ldots x_n$ are observations referring to the same variate x, we have

$$M_{\Sigma x} = M_x{}^n.$$

These results hold whether $x_1 \ldots x_n$ are continuous variates or not, and the definition (28) thus renders obvious in all cases the property of moment-generating functions given in (29), which, since this may be written

$$K_{\Sigma x} = \log M_{\Sigma x} = \Sigma K_x,$$

is sometimes called the additive property of semi-invariants.

If $\phi = x$, one of the observations, and x is a continuous variate with distribution $f(x)$, we have

$$E(e^{itx}) = \int_{-\infty}^{\infty} e^{itx} f(x)\,dx,$$

and similarly we have in general, when ϕ is continuous,

$$M(t) = E(e^{it\phi}) = \int_{-\infty}^{\infty} e^{it\phi} F(\phi)\,d\phi;$$

this equation is important for determining $F(\phi)$, the distribution of ϕ, when we know $M(t)$. In order to find $M(t)$, we must, however, average $e^{it\phi}$ in terms of the original observations, since its value is clearly the same whether we average for values of ϕ or for values of the observations of which ϕ is a function, and while we do not yet know the chance of a particular ϕ arising, given by $F(\phi)d\phi$, we do know the chance of particular values of the observations arising.

Notice, if $\phi = \phi(x_1, x_2)$, say,

$$E(e^{it\phi}) = E_{x_1}E_{x_2}(e^{it\phi}),$$

or symbolically

$$E = E_{x_1}E_{x_2},$$

where E_{x_1} denotes averaging with respect to x_1, etc., and if in particular $E_{x_2}(e^{it\phi})$ were independent of x_1, we should have simply

$$E(e^{it\phi}) = E_{x_2}(e^{it\phi}).$$

The above discussion deals only with the moment-generating function of a single quantity ϕ. For the present purpose we require the joint moment-generating function of q quantities ϕ_r, but this may be defined in exactly the same way by

$$M(t_r) = E\left(\exp \sum_{r=1}^{q} it_r\phi_r\right), \quad . \quad . \quad . \quad . \quad (30)$$

and all the corresponding properties hold.

9. We suppose we have two sets of variates $\xi_1 \ldots \xi_\kappa, \eta_{\kappa+1} \ldots \eta_p$, where the variates $\eta_\mu(\mu = \kappa + 1 \ldots p)$ are independent of the variates $\xi_m(m = 1 \ldots \kappa)$, and constitute a normal system

$$U_p(\eta_\mu) = \pi^{-\frac{1}{2}p} \mid A \mid^{\frac{1}{2}} e^{-A(\eta, \eta)} \prod_{\mu=\kappa+1}^{p} d\eta_\mu \quad . \quad . \quad . \quad . \quad (31)$$

The sample contains n sets of values of the p variates ξ_m, η_μ. The n sets of values of η_μ are of course assumed independent. No assumption is made, however, about the variates ξ_m.

The estimated variates $x_{1\cdot 0}$, etc., corresponding to ξ_1, etc., are as before. We consider the quantities

$$w_{\mu,m} = \Sigma\lambda_m\eta_\mu,$$

where

$$\lambda_m = \begin{cases} x_0/\sqrt{\Sigma x_0{}^2}, & (m = 0) \\ x_{m\cdot 0 \ldots m-1}/\sqrt{\Sigma x^2{}_{m\cdot 0 \ldots m-1}}, & (m = 1 \ldots \kappa); \end{cases}$$

and

$$\Sigma_{\mu\nu} = \Sigma\eta_\mu\eta_\nu, \quad (\mu, \nu = \kappa + 1 \ldots p).$$

The quantity x_0 we may subsequently put equal to unity.

The joint moment-generating function of $w_{\mu,m}$, $\Sigma_{\mu\nu}$ will be evaluated by averaging first with respect to the η, and then with respect to the ξ. Since the quantities $w_{\mu,m}$, $\Sigma_{\mu\nu}$ are each sums of n quantities $\lambda_m\eta_\mu$, $\eta_\mu\eta_\nu$ independent with respect to the η, we may write

$$M(2t_{\mu,m}, t_{\mu\nu}) = E_\xi \Pi E_\eta\left(\exp\left\{2i \sum_{\mu=\kappa+1}^{p} \tau_\mu\eta_\mu + iT(\eta, \eta)\right\}\right), \quad . \quad (32)$$

where Π denotes a product of n factors corresponding to the n sets of observations in the sample, $T \equiv (t_{\mu\nu})$, and

$$\tau_\mu = \sum_{m=0}^{\kappa} \lambda_m t_{\mu,m}.$$

Now we have from (31),

$$E_\eta\left(\exp\left\{2i\sum_{\mu=\kappa+1}^{p}\tau_\mu\eta_\mu+iT(\eta,\eta)\right\}\right)$$
$$=\int\ldots\int_{-\infty}^{\infty}\pi^{-\frac{1}{2}p}|A|^{\frac{1}{2}}\exp\left\{2i\sum_{\mu=\kappa+1}^{p}\tau_\mu\eta_\mu-(A-iT)(\eta,\eta)\right\}\prod_{\mu=\kappa+1}^{p}d\eta_\mu.$$

Integrating * this expression, we therefore have from (32)

$$M=E_\xi\Pi(\{|A|/|A-iT|\}^{\frac{1}{2}}e^{-B(\tau,\tau)}),$$

where $B\equiv(b_{\mu\nu})$ is the reciprocal of $(A-iT)$; that is,

$$M=\{|A|/|A-iT|\}^{\frac{1}{2}n}E_\xi(e^{-\Sigma B(\tau,\tau)}).$$

Now, since †

$$\Sigma x_{m\cdot 0\ldots m-1}x_{n\cdot 0\ldots n-1}=0$$

if

$$n\neq m,$$

$$\Sigma\lambda_m\lambda_n=0$$

if

$$n\neq m,$$

and also obviously from the definition of λ_m,

$$\Sigma\lambda_m{}^2=1;$$

hence we have

$$\Sigma B(\tau,\tau)=\sum_{\mu,\nu=\kappa+1}^{p}b_{\mu\nu}(\Sigma\tau_\mu\tau_\nu)$$
$$=\sum_{\mu,\nu=\kappa+1}^{p}b_{\mu\nu}\left(\sum_{m,n=0}^{\kappa}t_{\mu,m}t_{\nu,n}\Sigma\lambda_m\lambda_n\right)$$
$$=\sum_{m=0}^{\kappa}B(t_m,t_m).$$

Now this expression is independent of the ξ; hence averaging with respect to ξ does not affect the result, and we have finally

$$M=\{|A|/|A-iT|\}^{\frac{1}{2}n}\exp\left\{-\sum_{m=0}^{\kappa}B(t_m,t_m)\right\}. \quad . \quad . \quad (33)$$

Now since, if $F(w_{\mu,m},\Sigma_{\mu\nu})$ represents the joint distribution of $w_{\mu,m}$, $\Sigma_{\mu\nu}$, we may write

$$M=\int\ldots\int_{-\infty}^{\infty}\exp\left\{2i\sum_{\mu=\kappa+1}^{p}\sum_{m=0}^{\kappa}t_{\mu,m}w_{\mu,m}+i\sum_{\mu,\nu=\kappa+1}^{p}t_{\mu\nu}\Sigma_{\mu\nu}\right\}F\,dw\,d\Sigma,$$

where

$$dw\equiv\prod_{\mu=\kappa+1}^{p}\prod_{m=0}^{\kappa}dw_{\mu,m},\qquad d\Sigma\equiv\prod_{\kappa+1\leqslant\mu\leqslant\nu\leqslant p}d\Sigma_{\mu\nu},$$

we have by a generalized form of Fourier's Integral Theorem (see (2)),

$$F=(2\pi)^{-(p-\kappa)}\pi^{-\frac{1}{4}(p-\kappa)(p+\kappa+1)}\int\ldots\int_{-\infty}^{\infty}\exp\left\{-2i\sum_{\mu=\kappa+1}^{p}\sum_{m=0}^{\kappa}t_{\mu,m}w_{\mu,m}-i\sum_{\mu,\nu=\kappa+1}^{p}t_{\mu\nu}\Sigma_{\mu\nu}\right\}$$
$$\times M\prod_{m=0}^{\kappa}dt_m\,dt,$$

* *Cf.* Wishart and Bartlett (2), p. 2, equation (5).

† See Yule (9), p. 183.

where

$$dt_m \equiv \prod_{\mu=\kappa+1}^{p} dt_{\mu,m}, \qquad dt \equiv \prod_{\kappa+1 \leqslant \mu \leqslant \nu \leqslant p} dt_{\mu\nu},$$

and M is given by (33). Integrate * successively with respect to the $t_m (m = 0, 1, \ldots \kappa)$, and we have

$$F = \int \ldots \int_{-\infty}^{\infty} \frac{|A|^{\frac{1}{2}n} \exp\left\{-i \sum_{\mu,\nu=\kappa+1}^{p} t_{\mu\nu}\Sigma_{\mu\nu} + \sum_{m=0}^{\kappa} B^{-1}(w_m, w_m)\right\}}{(2\pi)^{-(p-\kappa)}\pi^{-\frac{1}{2}p(p-\kappa)} |A - iT|^{\frac{1}{2}n} |B|^{\frac{1}{2}\kappa}} dt,$$

or since

$$B^{-1}(w_m, w_m) = A(w_m, w_m) - iT(w_m, w_m),$$

and

$$|B|^{-\frac{1}{2}\kappa} = |A - iT|^{\frac{1}{2}\kappa},$$

we have

$$F = \prod_{m=0}^{\kappa} \{\pi^{-\frac{1}{2}(p-\kappa)} |A|^{\frac{1}{2}} e^{-A(w_m, w_m)}\} . \int \ldots \int_{-\infty}^{\infty} \frac{|A|^{\frac{1}{2}(n-\kappa-1)} \exp\left\{-i \sum_{\mu,\nu=\kappa+1}^{p} t_{\mu\nu}\Sigma_{\mu\nu}'\right\}}{(2\pi)^{(p-\kappa)}\pi^{\frac{1}{2}(p-\kappa)(p-\kappa-1)} |A - iT|^{\frac{1}{2}(n-\kappa-1)}} dt,$$

where

$$\Sigma_{\mu\nu}' = \Sigma_{\mu\nu} - \sum_{m=0}^{\kappa} w_{\mu,m} w_{\nu,m}.$$

The integral remaining is exactly similar to the integral obtained in the paper quoted (2) to obtain $V_p(c_{\mu\nu\cdot 0}, n-1)$, but with $n-k-1$ instead of $n-1$. It has been evaluated by Ingham (10), and using his result, changing the variates from $w_{\mu,m}$, $\Sigma_{\mu\nu}$ to $w_{\mu,m}$, $\Sigma_{\mu\nu}'$, and inserting the differentials, we have

$$F dw d\Sigma' = \prod_{m=0}^{\kappa} \{U_{p-\kappa}(w_{\mu,m})\} V_{p-\kappa}(\Sigma_{\mu\nu}', n-\kappa-1).$$

To interpret this result, we simply write

$$\eta_\mu = \xi_{\mu\cdot 1 \ldots \kappa},$$

so that in a set of variates $\xi_1 \ldots \xi_p$ we have assumed only that the $p-\kappa$ partial variates $\xi_{\mu\cdot 1 \ldots \kappa}$ are normal.

The quantities $w_{\mu,m}$ are readily expressible in terms of the observations $x_1 \ldots x_p$. They are in fact analogous to the $w_1 \ldots w_{p+1}$ of equation (19), except for $w_{\mu,0}$, which are functions of the means—which were not considered in I. They may be written

$$\begin{cases} w_{\mu,0} = \sqrt{n}(\bar{x}_\mu - m_\mu - \beta_{\mu\kappa\cdot 1 \ldots \kappa-1}[\bar{x}_\kappa - m_\kappa] \ldots - \beta_{\mu 1\cdot 2 \ldots \kappa}[\bar{x}_1 - m_1]) \\ w_{\mu,1} = \sqrt{c_{11\cdot 0}}(b_{\mu 1} - \beta_{\mu\kappa\cdot 1 \ldots \kappa-1} b_{\kappa 1} \ldots - \beta_{\mu 1\cdot 2 \ldots \kappa}) \\ w_{\mu,2} = \sqrt{c_{22\cdot 01}}(b_{\mu 2\cdot 1} - \beta_{\mu\kappa\cdot 1 \ldots \kappa-1} b_{\kappa 2\cdot 1} \ldots - \beta_{\mu 2\cdot 13 \ldots \kappa}) \\ \vdots \\ w_{\mu,\kappa} = \sqrt{c_{\kappa\kappa\cdot 0 \ldots \kappa-1}}(b_{\mu\kappa\cdot 1 \ldots \kappa-1} - \beta_{\mu\kappa\cdot 1 \ldots \kappa-1}) \end{cases} \qquad (34)$$

* *Cf.* Wishart and Bartlett, *loc. cit.*

Just as $x_{\mu\cdot 0}$, $b_{\mu 1}$, etc. refer to the variates $\xi_1 \ldots \xi_p$, let $x_{\mu\cdot 0}'$, $b_{\mu 1}'$, etc. refer to the variates $\xi_1 \ldots \xi_{\kappa+1}, \eta_\kappa \ldots \eta_p$. Then we have

$$\begin{aligned}\Sigma\eta_\mu\eta_\nu - \sum_{m=0}^{\kappa} w_{\mu, m}w_{\nu, m} &= \Sigma x_{\mu\cdot 0}'x_{\nu\cdot 0}' - \sum_{m=1}^{\kappa} w_{\mu, m}w_{\nu,} \\ &= \Sigma(x_{\mu 0}' - b_{\mu 1}'x_{1\cdot 0})(x_{\nu\cdot 0}' - b_{\nu 1}'x_{1\cdot 0}) - \sum_{m=2}^{\kappa} w_{\mu, m}w_{\nu, m} \\ &= \Sigma(x_{\mu\cdot 01}'x_{\nu\cdot 01}') - \sum_{m=2}^{\kappa} w_{\mu, m}w_{\nu, n} \\ &= \ldots \\ &= \Sigma x_{\mu\cdot 01}'{}_{\ldots\kappa} x_{\nu\cdot 01}'{}_{\ldots\kappa}.\end{aligned}$$

But

$$\begin{aligned}x_{\mu\cdot 0}'{}_{\ldots\kappa} &= x_{\mu\cdot 0\ldots\kappa} - \beta_{\mu\kappa\cdot 1\ldots\kappa-1}x_{\kappa\cdot 0\ldots\kappa} \ldots - \beta_{\mu 1\cdot 2\ldots\kappa}x_{1\cdot 0\ldots\kappa} \\ &= x_{\mu\cdot 0\ldots}{}^{\kappa},\end{aligned}$$

since by definition $x_{\kappa\cdot 0\ldots\kappa}$, etc. $=0$. Hence

$$\Sigma_{\mu\nu}' = c_{\mu\nu\cdot 0\ldots\kappa}.$$

Thus we may write the distribution finally

$$W_{p,\kappa} \equiv \prod_{m=0}^{\kappa} \{U_{p-\kappa}(w_{\mu, m})\} V_{p-\kappa}(c_{\mu\nu\cdot 0\ldots\kappa}, n-\kappa-1). \quad . \quad . \quad (35)$$

It is suggested that the set of distributions represented by $W_{p,\kappa}$ is the more fundamental, at any rate for the study of regression or partial correlation, for though we saw in I that it may be deduced from Wishart's distribution, it is now seen to be true under more general conditions—namely, that the $p-\kappa$ variates $\xi_{\mu\cdot 1\ldots\kappa}$ are normal—and Wishart's distribution may alternatively be obtained from (35) by writing $\kappa=0$. For the present purpose, the most important cases are given by $\kappa=p-2$, and $p-1$.

10. Thus, for $p=2$, $\kappa=0$ gives Fisher's distribution for two variates. When $\kappa=1$, we have

$$W_{2,1} = U_1(w_0)U_1(w_1)V_1(c_{22\cdot 01}, n-2), \quad . \quad . \quad . \quad (36)$$

where

$$\begin{cases} w_0 \equiv w_{2,0} = \sqrt{n}(\bar{x}_2 - m_2 - \beta_{21}[\bar{x}_1 - m_1]) \\ w_1 \equiv w_{2,1} = \sqrt{c_{11\cdot 0}}(b_{21} - \beta_{21}). \end{cases}$$

This requires only that $\xi_{2\cdot 1}$ exists and is normal (this condition implies that the regression of ξ_2 on ξ_1 is linear, and that the variance of ξ_2 for each ξ_1 is constant).

Thus the quantity w_1 is normally distributed whatever the distribution of ξ_1, whether the observations ξ_1 can be regarded as fixed from sample to sample or not, or whether they are specially selected. The reason for this is that w_1, regarded as a function of $\xi_{2\cdot 1}$, is so weighted that the variance is independent of ξ_1. Since the infinite population of ξ_1

and $\xi_{2\cdot 1}$ can be split up into an infinite number of sub-populations for which the ξ_1 are fixed while $\xi_{2\cdot 1}$ vary (for $\xi_{2\cdot 1}$ and ξ_1 are assumed independent), the distribution of w_1 will be the sum of an infinite number of distributions each of which is normal, with the same variance and mean. Consequently, the distribution of w_1 will be similarly distributed, however ξ_1 varies.

In the case where ξ_1 is fixed from sample to sample the problem is simplified, since it is not so essential to consider the particular function w_1. Any linear function of $\xi_{2\cdot 1}$ will then be normal, but for any other function the variance will of course depend on ξ_1. There is no doubt that the condition that the set ξ_1 may be supposed fixed is commonly met with in practice, and this case was the one considered by Fisher (11), although he seems to suggest in conclusion that his test holds under somewhat wider conditions than he assumed.

It is important to notice that the test of significance of the mean $\bar{x}_2$ obtainable from $W_{2,1}$ is *only* valid when the set ξ_1 is fixed, for then $\bar{x}_1 \equiv m_1$, and w_0 becomes $\sqrt{n}(\bar{x}_2 - m_2)$. Otherwise, as for a random sample in two normal variates, we should have to consider $W_{1,0}$ for the variate ξ_2, given by $U_1(\sqrt{n}[\bar{x}_2 - m_2])V_1(c_{22\cdot 0}, n-1)$, with the corresponding assumption that ξ_2 is normal. The "t test" of significance of b_{21} from β_{21} is, however, valid, with no restrictions on ξ_1. Moreover, the special case of this, the test of significance of b_{21} or r_{21} from zero, is clearly to be regarded as a special case of the test of regression, not of correlation, for fewer assumptions are involved in the test of regression than in the test of correlation. Thus if we regarded the distribution of the correlation coefficient r_{21} when ρ_{21} is zero as a special case of the general distribution when $\rho_{21} \neq 0$, we might have supposed that ξ_2, ξ_1 must both be normal.

The simplest way to consider this distribution is to use Fisher's geometrical methods (3), and consider the chance of the two *radii vectores* representing in n dimensions the sample of ξ_1 and ξ_2 making an angle θ with each other when they are independent of each other—since $r_{21} = \cos\theta$. Clearly, however restricted the *radius vector* of ξ_1, say, is, if any direction of the *radius vector* of ξ_2 is equally likely, then the angle θ will be perfectly random, and the distribution of r_{21} when ρ_{21} is zero follows. The condition that any direction of the radius vector is equally likely is the condition of the normal law (compare Maxwell's proof of the normal law in the dynamical theory of gases). Hence the only condition for the distribution to hold is seen to be that ξ_2, say, is a normal variate (and the n observations ξ_2, of course, independent of each other).

In the case of three variates, put $p=3$, $\kappa=1$, and we have

$$W_{3,1} = U_2(w_{\mu,0})U_2(w_{\mu,1})V_2(c_{\mu\nu\cdot 01}, n-2). \qquad (37)$$

From $V_2(c_{\mu\nu\cdot01}, n-2)$ is obtained the distribution of the partial correlation coefficient $r_{32\cdot1}$, and this therefore depends only on $\xi_{3\cdot1}$ and $\xi_{2\cdot1}$ being normal, as has been pointed out by Fisher ((14), p. 163).

If instead we put $\kappa=2$, we have

$$W_{3,\,2}=U_1(w_0)U_1(w_1)U_1(w_2)V_1(c_{33\cdot012}, n-3), \quad . \quad . \quad (38)$$

where

$$\left\{\begin{aligned} w_0 &\equiv w_{3,\,0}=\sqrt{n}(\bar{x}_3-m_3-\beta_{32\cdot1}[\bar{x}_2-m_2]-\beta_{31\cdot2}[\bar{x}_1-m_1]) \\ w_1 &\equiv w_{3,\,1}=\sqrt{c_{11\cdot0}}(b_{31}-\beta_{32\cdot1}b_{21}-\beta_{31\cdot2}) \\ w_2 &= w_{3,\,2}=\sqrt{c_{22\cdot01}}(b_{32\cdot1}-\beta_{32\cdot1}), \end{aligned}\right.$$

provided only that $\xi_{3\cdot21}$ exists and is normal.

The test of the partial regression coefficient $b_{32\cdot1}$ from $\beta_{32\cdot1}$ thus depends only on this condition. Similarly for that of $b_{31\cdot2}$ from $\beta_{31\cdot2}$. This applies also to the joint test of regression using the "z distribution," and to the special case of this latter test when $\beta_{31\cdot2}$ and $\beta_{32\cdot1}$ are put equal to zero, when it coincides with the test of a significant multiple correlation.

The condition that $\xi_{3\cdot21}$ exists implies that the partial regressions with ξ_2 and ξ_1 are linear, since

$$\xi_{3\cdot21}=\xi_3-\beta_{32\cdot1}\xi_2-\beta_{31\cdot2}\xi_1,$$

but since we have seen that the distributions of ξ_2 and ξ_1 are immaterial, we may suppose them to be any required functions of the original observations, or what is more usual in practice if we have no prior knowledge of these functions, may consider the partial regressions with ξ_1, ξ_1^2, etc. We may, for example, by putting $\xi_2\equiv\xi_1^2$, interpret the above equation as representing the parabolic regression of a variate ξ_3 on ξ_1. Thus curvilinear regression and curvilinear partial regression provide no further theoretical difficulties. When a curved regression line is being fitted, it is often convenient to consider the partial regressions not with ξ_1, ξ_1^2, . . ., but with some orthogonal set such as $\xi_{1\cdot0}$, $\xi_{2\cdot01}$, . . ., where $\xi_2\equiv\xi_1^2$, etc.; the regression line may then be fitted term by term. This procedure is practically most important when the values of ξ_1 are separated by equidistant intervals.

11. The test of significance of $\bar{x}_1-m_1$ has been extended by Fisher to include the significance of a difference $\bar{x}_1-\bar{x}_1'$, where $\bar{x}_1$ and $\bar{x}_1'$ are the means from two independent samples S and S′. We may now regard the variate

$$w_{1,\,0}=\sqrt{n}(\bar{x}_1-m_1)$$

as a special case of the variates $w_{1,\,0}$, $w_{2,\,1}$, etc.; one, moreover, which is especially simple, as it is a function of x_1 and x_0, where the variate x_0 is not only fixed from sample to sample, but can take only the value unity. Thus since x_0 is fixed, it is not necessary to eliminate it to eliminate m_1, and we can not only consider

$$w_{1,0} - w_{1,0}' = \sqrt{n}(\bar{x}_1 - \bar{x}_1') \quad . \quad . \quad . \quad . \quad (39)$$

when the samples are equal in size, but

$$w_{1,0}/\sqrt{n} - w_{1,0}'/\sqrt{n'} = (\bar{x}_1 - \bar{x}_1'), \quad . \quad . \quad . \quad (40)$$

say, if they are unequal.

In the general case of the significance of any $w \sim w'$, the condition that the "independent variates" should be fixed from sample to sample for the test to be strictly applicable immediately becomes apparent.

Thus in the case of two variates, we only have

$$u \equiv w_{2,1} - w_{2,1}' = \sqrt{c_{11\cdot0}}(b_{21} - b_{21}') \quad . \quad . \quad . \quad (41)$$

provided that $\sqrt{c_{11\cdot0}'} = \sqrt{c_{11\cdot0}}$, that is, the samples are equal in size, and the same set of ξ_1 is contained in each sample. We may then consider the normal variate in (41), with its corresponding estimated variance

$$v = \left(c^2_{22\cdot01} + (c_{22\cdot01}')^2\right)/2(n-2),$$

and test the significance of the difference between b_{21} and b_{21}' by means of $t = u/\sqrt{v}$, which will follow the "t distribution" with $2(n-2)$ degrees of freedom.

If the samples are unequal in size, but the ξ_1 are fixed, we can weight $w_{2,1}$ and $w_{2,1}'$ in order to eliminate β_{21}, similarly to (40).

For three variates, when ξ_1 and ξ_2 are supposed fixed from sample to sample, we have similarly to (41) for two equal samples,

$$u_2 \equiv w_{3,2} - w_{3,2}' = \sqrt{c_{22\cdot01}}(b_{32\cdot1} - b_{32\cdot1}'). \quad . \quad . \quad . \quad (42)$$

Further, under the same conditions,

$$u_1 \equiv w_{3,1} - w_{3,1}' = \sqrt{c_{11\cdot0}}(b_{31} - b_{31}'), \quad . \quad . \quad . \quad (43)$$

and we have not only from (42) a normal variate by which we can test the significance of the difference between $b_{32\cdot1}$ and $b_{32\cdot1}'$, but from (43) a more sensitive test of significance between b_{31} and b_{31}' if ξ_2 appreciably affects ξ_3 and ξ_1. Again, since

$$2v_1 \equiv u_1^2 + u_2^2 = (c_{32\cdot0} - c_{32\cdot0}')(b_{32\cdot1} - b_{32\cdot1}') + (c_{31\cdot0} - c_{31\cdot0}')(b_{31\cdot2} - b_{31\cdot2}'), \quad . \quad (44)$$

we may test the joint significance between $b_{32\cdot1}$, $b_{31\cdot2}$, and $b_{32\cdot1}'$, $b_{31\cdot2}'$ respectively, by means of the "z distribution."

Analogous tests hold for three variates if the samples are unequal, provided ξ_1 and ξ_2 are still fixed for each sample; similarly for any number of variates.

12. To sum up the results of II, the following distributions depend simply on the existence and normality of the two partial or residual variates $\xi_{p\cdot1\ldots p-2}$, $\xi_{p-1\cdot1\ldots p-2}$:

(*a*) The distribution of the partial correlation coefficient $r_{p,\, p-1 \cdot 1 \ldots p-2}$.

(*b*) The distribution of the partial co-variance $v_{p,\, p-1 \cdot 1 \ldots p-2}$.

(*c*) The distribution of the partial regression coefficient $b_{p,\, p-1 \cdot 1 \ldots p-2}$.

On the other hand, the following distributions and tests of regression depend only on the existence and normality of the one residual, $\xi_{p \cdot 1 \ldots p-1}$:

(*d*) The distribution of the partial variance $v_{pp \cdot 1 \ldots p-1}$.

(*e*) The distribution of the normal and independent variates (defined in (34)), $w_{p,\, 0} \ldots w_{p,\, p-1}$.

(*f*) The test of significance of $b_{p,\, p-1 \cdot 1 \ldots p-2} \sim \beta_{p,\, p-1 \cdot 1 \ldots p-2}$—including the particular case when $\rho_{p,\, p-1 \cdot 1 \ldots p-2}$ assumed zero.

(*g*) The joint test of significance of $b_{pr \cdot 1 \ldots r-1,\, r+1 \ldots p-1} \sim \beta_{pr \cdot 1 \ldots r-1,\, r+1 \ldots p-1} (r = 1 \ldots p-1)$—including the particular case when no real multiple correlation is assumed.

(*h*) In the problem of two samples, provided we can regard the set of variates $\xi_1, \ldots \xi_{p-1}$ as fixed, the test of significance of $b_{p,\, p-1 \cdot 1 \ldots p-2} \sim b_{p}',_{\, p-1 \cdot 1 \ldots p-2}$, or the joint test of significance of $b_{pr \cdot 1 \ldots r-1,\, r+1 \ldots p-1} \sim b_{pr \cdot 1}' \ldots_{r-1,\, r+1 \ldots p-1} (r = 1 \ldots p-1)$.

Regression is seen to be a wider concept than correlation; for fewer assumptions are involved in the more important distributions and tests, which are moreover simpler.

Though all correlation tests—when we are concerned only with finding whether there is any correlation at all—reduce to regression tests, a distinction should obviously be made between problems of regression in general, and those where the idea of correlation may be usefully employed. In the case of two variates, for example, since the distribution of the correlation coefficient when ρ_{21} is not zero requires a random sample of two normal variates ξ_2 and ξ_1, a correlation coefficient has most meaning in a sample of this kind.

No mention has so far been made of the test of goodness of fit of regression lines (see Fisher (16)), since this depends on our having an array of values of ξ_2 for each value of ξ_1. The test is, however, simply another application of the analysis of variance, the "z distribution" being used to test, say, whether we have satisfactorily obtained a normal variate $\xi_{2 \cdot 1}$ by assuming $\xi_{2 \cdot 1} = \xi_2 - \beta_{21} \xi_1$ and fitting a straight line. The variation within arrays is compared with the variation of the means of the arrays from the fitted regression line. The test is thus strictly possible only when several values of ξ_2 correspond to each ξ_1, but in practice when this condition is not fulfilled we can sometimes group the values of ξ_2 provided the grouping is sufficiently fine. Alternatively, we could fit the regression line first, and afterwards group the deviations from this line; we should then, however, as an approximation have to neglect the number of degrees of freedom lost in fitting, and assume the fitted regression line to be the true (linear) regression line.

An exact theoretical test for the goodness of fit of a straight line would be to fit the regression line, and then group coarsely, making an estimate of $\sigma^2_{2\cdot 1}$ by finding $c_{22\cdot 01}$ *for each group*. That is, if there were a groups, finding

$$(n-2a)v_1 = \sum_{r=1}^{a} (c_{22\cdot 01})_r,$$

$$2(a-1)v_2 = c_{22\cdot 01} - \sum_{r=1}^{a} (c_{22\cdot 01})_r,$$

and comparing the independent estimates of $\sigma^2_{2\cdot 1}$ given by v_1 and v_2. Even if we were testing the fit of a curved regression line, it would probably be sufficient to assume the regression in any group to be linear. But the awkwardness of the above test makes its practical importance limited; especially as the test of goodness of fit is supplementary to the specific tests of significance of regression coefficients—tests which are usually more sensitive.

Thus it might happen that a regression coefficient b_{21} is significant, although the goodness of fit seemed adequate when β_{21} was assumed zero; this is because in the test of significance of b_{21} the variance due to the linear regression is isolated as a single square. This particular criticism of the test of goodness of fit is general, of course, and applies also to the case where there are arrays of values of ξ_2 for each ξ_1.

In conclusion, we may perhaps recall that the calculation of all regression coefficients is a problem in the theory of least squares, and consequently their accuracy depends only on the assumptions necessary for this theory. We had, for example, the normal variate

$$w_{2,1} = \sqrt{c_{11\cdot 0}}(b_{21} - \beta_{21}) = \Sigma\lambda_1\xi_{2\cdot 1},$$

where

$$\lambda_1 = x_{1\cdot 0}/\sqrt{\Sigma x^2_{1\cdot 0}},$$

whether the values of ξ_1 are fixed or not, provided $\xi_{2\cdot 1}$ is normal. It is clear, however, that $w_{2,1}$ is approximately normal whatever the distribution of $\xi_{2\cdot 1}$, provided no values of $\xi_{2\cdot 1}$ (or λ_1) are abnormal, since a linear function of n variates goes to normality as n becomes large.

Hence it is likely that

$$\sqrt{c_{11\cdot 0}} \mid b_{21} - \beta_{21} \mid < 2\sigma_{2\cdot 1},$$

and for a reasonably sized sample, so that $\sqrt{c_{11\cdot 0}}$ is reasonably large, we may conclude that the discrepancy between b_{21} and β_{21} is likely to be small. But the exact test of significance of $b_{21}-\beta_{21}$ must, for any finite sample, involve the assumption that $\xi_{2\cdot 1}$ is normal.

I should like to express my thanks to Dr J. Wishart, to whom I owe the suggestion that a more systematic and complete derivation might be possible of the various distributions and tests associated with regression than has perhaps hitherto been given. I am also indebted to Dr Wishart for advice and criticism while this paper was being prepared.

REFERENCES TO LITERATURE.

1. WISHART, J., 1928. *Biometrika*, vol. xx, A, pp. 32–52.
2. WISHART, J., and BARTLETT, M. S., 1933. *Proc. Camb. Phil. Soc.*, vol. xxix, pp. 260–270.
3. FISHER, R. A., 1915. *Biometrika*, vol. x, pp. 507–521.
4. PEARSON, K., JEFFERY, G. B., and ELDERTON, E. M., 1929. *Biometrika*, vol. xxi, pp. 164–193.
5. ROMANOVSKY, V., 1926. *Bulletin de l'Academie des Sciences de l'U.R.S.S.*, p. 646.
   PEARSON, K., 1926. *Proc. Roy. Soc.*, A, vol. cxii, pp. 1–14.
6. WILKS, S. S., 1932. *Annals Math. Statistics*, pp. 196–203.
7. FISHER, R. A., 1928. *Proc. Roy. Soc.*, A, vol. cxxi, pp. 654–673.
8. WILKS, S. S., 1932. *Biometrika*, vol. xxiv, pp. 471–494.
9. YULE, G. U., 1907. *Proc. Roy. Soc.*, A, vol. lxxix, pp. 182–193.
10. INGHAM, A. E., 1933. *Proc. Camb. Phil. Soc.*, vol. xxix, pp. 271–276.
11. FISHER, R. A., 1925. *Metron*, vol. v, No. 3, pp. 90–104.
12. WISHART, J., and BARTLETT, M. S., 1932. *Proc. Camb. Phil. Soc.*, vol. xxviii, pp. 455–459.
13. FISHER, R. A., 1924. *Metron*, vol. iii, Nos. 3–4, pp. 329–333.
14. FISHER, R. A., 1930. *Statistical Methods for Research Workers* (3rd ed.).
15. FISHER, R. A., 1924. *Phil. Trans.*, B, vol. ccxiii, pp. 89–142.
16. FISHER, R. A., 1922. *Journ. Roy. Stat. Soc.*, vol. lxxxv, pp. 597–612.

(*Issued separately September* 29, 1933.)