Cambridge University Press
978-0-521-12412-6 - Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples
F. N. David
Frontmatter
More information

I have endeavoured to indicate tonight in broad outline the early history of correlation which has now a most extensive literature. It is a long step from Francis Galton's 'reversion' in sweet peas seeds to the full theory of multiple correlation, which we now know to be identical with the spherical trigonometry of high-dimensioned space, the total correlation coefficients being the cosines of the edges of the polyhedra and the partial correlation coefficients the cosines of the polyhedral angles. But to find the correlation of the health of a child with the number of people per room while you render neutral its age, the health of its parents, the wages of its father, and the habits of its mother, is no less vital a problem than Galton's correlation of character in parent and offspring. It requires indeed more mathematics, but the mathematics are not there for the joy of the analyst but because they are essential to the solution. It is the transition from the mill as pestle & mortar to the mill with steam driven steel grain crushing rollers. But the inventor of milling was the person who bruised grain between two stones, and Galton was the man who discovered the highway across this new country with what he aptly terms "its easy descents to different goals".

Karl Pearson.

Facsimile of the last pages of the MS. of Karl Pearson's paper on the History of Correlation read before the Society of Biometricians and Mathematical Statisticians, June 14, 1920 (see *Biometrika*, XIII, p. 45)

TABLES
OF
THE ORDINATES AND PROBABILITY INTEGRAL OF THE DISTRIBUTION OF THE CORRELATION COEFFICIENT IN SMALL SAMPLES

BY

F. N. DAVID

ISSUED BY THE *BIOMETRIKA* OFFICE
UNIVERSITY COLLEGE, LONDON

CAMBRIDGE
AT THE UNIVERSITY PRESS
1954

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521124126

First published 1938
Reprinted 1954
This digitally printed version 2009

A catalogue record for this publication is available from the British Library

ISBN 978-0-521-07169-7 Hardback
ISBN 978-0-521-12412-6 Paperback

CONTENTS

EDITOR'S PREFACE

The coefficient of correlation has played an important part during the last fifty years in the development of the theory of mathematical statistics and its applications. In a field where many have worked three names may be linked specially with the development of the theory of r: those of Francis Galton, Karl Pearson and R. A. Fisher. Galton's pioneer work may seem to us easy now, but it is just such simple first steps which are often the most difficult to make and yet the most far-reaching in their consequences. It was Pearson who was largely responsible for the development of the theory of correlation from these first foundations, and for demonstrating beyond question that this new concept "brought psychology, anthropology, medicine and sociology into the field of mathematical treatment". Finally, at almost his first venture into the field of statistical theory, Fisher deduced the probability distribution of r in samples from a normal population, and at the same time drew attention to that valuable conception by which a sample may be represented as a point in multiple space.

The importance of Fisher's 1914 result was at once appreciated by Pearson, who, with his characteristic eagerness to put theory into numbers, was already early in 1915 planning the "Cooperative Study", which presented in tables of ordinates, in frequency constants and in photographs of models the varied forms assumed by the distribution of r. To improve on the tabled ordinates had probably long been Pearson's plan, but it was not until 1931 that he suggested to Miss David the computation of the tables of the probability integral now published in this volume. In the meantime Fisher had suggested the very useful logarithmic transformation of r, which enables the probability integral to be obtained from tables of the normal function with an accuracy sufficient for most common purposes. Nevertheless, in Pearson's view, however useful an approximation might be, it was desirable that for a statistical measure as important as the coefficient of correlation there should be available fundamental tables which would enable the probability integral to be determined with a considerable degree of mathematical accuracy for any size of sample, n, and any population correlation, ρ.

While accepting this view the present editor is well aware that the ideal of a mathematically accurate, all-embracing table, has not been attained. Karl Pearson's hope that it would be possible to interpolate accurately up to $n = 1600$ from a framework of tables at $n = 25$, 50, 100, 200, 400, 800 and 1600 was not fulfilled, and it has been necessary to rest content with a more restricted objective.

Miss David has loyally completed a difficult task whose solution in the region of high n and ρ proved increasingly elusive. She has made two special contributions of her own: (1) a scheme of charts based on the tables which, for the range they cover, provide a more comprehensive and useful picture of the relation between r, ρ and n than is elsewhere available; (2) an introduction in which she has tried, with, I believe, no little measure of success, to link together a number of illustrative examples with a simple statement of the principles by which the theory of probability may be used as a guide in drawing inferences from observation.

Recent developments of theory, coming in response to a need for the solution of new types of problem, have laid emphasis on a technique which deals with regression rather than correlation coefficients and considers the apportionment of the covariance or product sum, $\Sigma(x-\bar{x})(y-\bar{y})$, into parts that may be associated with different factors determining variation. But however important these new conceptions may be, the product-moment coefficient of correlation is likely always to have an essential part to play in the application of statistical method. When two variables are approximately normally correlated—and there are good reasons for supposing that there is considerable latitude in the stringency of this approximation—the coefficient by itself provides a completely adequate measure of the intensity of association. Lying between -1 and $+1$ and being independent of any units of measurement, in fact a pure measure of correlation, r has a direct and simple appeal. In providing tables and charts which deal with the relationship between r, ρ and n we are not, I feel certain, merely toying with an interesting historical relic of the past.

E. S. PEARSON

DEPARTMENT OF STATISTICS
UNIVERSITY COLLEGE
LONDON
February 1938

Editorial Note to 1954 *re-issue*. We are indebted to Dr A. K. Gayen for pointing out certain errors in R. A. Fisher's original formulae (7), (8) and (9) (p. viii) and for computing the consequent corrections to the columns headed Approximation I and II in Table VIII (p. xxxii).

INTRODUCTION

SECTION I. INTRODUCTORY

Throughout the following pages it will be assumed that we are dealing with two correlated random variables x and y for which the joint probability law is

$$p(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-\xi_1)^2}{\sigma_1^2} - \frac{2\rho(x-\xi_1)(y-\xi_2)}{\sigma_1\sigma_2} + \frac{(y-\xi_2)^2}{\sigma_2^2}\right\}}. \qquad \text{......(1)}$$

The probability laws of x and y thus each follow the normal distribution, i.e.

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{1}{2\sigma_1^2}(x-\xi_1)^2}; \quad p(y) = \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{1}{2\sigma_2^2}(y-\xi_2)^2}. \qquad \text{......(2)}$$

We shall be concerned with the relation between the correlation coefficient

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}},^* \qquad \text{......(3)}$$

calculated from a sample of size n, randomly drawn from an infinite population represented by the normal bivariate distribution law (1), and the correlation coefficient, ρ, of that population. The tables which follow give the ordinates and areas of the curves representing the sampling distribution of r for differing values of n and ρ.

The form of the distribution of r for $\rho = 0$ and any n was first given by "Student" (1), while for the general distribution for any n and ρ we are indebted to R. A. Fisher (2). "Student's" results were obtained by empirical means, but R. A. Fisher's proof, dependent upon geometrical argument and analogy, is mathematically rigorous.†

Distribution of r for any n, $\rho = 0$.

$$p(r \mid n, \rho = 0) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-2}{2}\right)}(1-r^2)^{\frac{n-4}{2}}. \qquad \text{......(4)}$$

Distribution of r for any n and any ρ.

$$p(r \mid n, \rho) = \frac{(1-\rho^2)^{\frac{n-1}{2}}}{\pi(n-3)!}(1-r^2)^{\frac{n-4}{2}} \frac{d^{n-2}}{d(r\rho)^{n-2}}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2 r^2}}\right). \qquad \text{......(5)}$$

Among other papers on the form of the distribution we may notice one by H. E. Soper (3) and one published from the Department of Applied Statistics entitled "A Cooperative Study" (4). In the latter paper tables were included of the ordinates of the distribution of r for given values of n and ρ. From these tables it was possible to calculate by quadrature the area under the curves and therefore the probability integral of r.

* $\bar{x}$ and $\bar{y}$ denote the sample means, i.e. $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

† An alternative proof of the general distribution is given in the Appendix. While it follows closely the lines of Professor Fisher's proof the distribution is reached by means of algebraic transformations. It is hoped that this form will be of use to those who are unable to visualize a solid figure. For an elegant proof using characteristic functions see S. Kullback, *Annals of Mathematical Statistics*, v, 4 (1934), pp. 263–305.

This method of quadrature was the only one by which a precise measure of the significance of r could be obtained until the publication of an important paper by R. A. Fisher (5), which contains the now well-known transformation of r referred to below.

Fisher showed that by a suitable transformation of r and ρ the distribution curves of r could be transformed approximately into normal curves.

Write $$z' = \tfrac{1}{2}\log\left(\frac{1+r}{1-r}\right); \quad \zeta = \tfrac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right). \qquad \ldots\ldots(6)$$

From $p(z' \mid n, \zeta)$ we obtain the moments in series:

$$\text{Mean }(z') = \zeta + \frac{\rho}{2(n-1)}\left\{1 + \frac{5+\rho^2}{4(n-1)} + \ldots\right\}, \qquad \ldots\ldots(7)$$

$$\sigma^2_{z'} = \frac{1}{n-1}\left\{1 + \frac{4-\rho^2}{2(n-1)} + \frac{22-6\rho^2-3\rho^4}{6(n-1)^2} + \ldots\right\}, \qquad \ldots\ldots(8)$$

$$\left.\begin{aligned} \beta_1 &= \frac{\rho^6}{(n-1)^3} + \ldots, \\ \beta_2 &= 3 + \frac{2}{n-1} + \frac{4+2\rho^2-3\rho^4}{(n-1)^2} + \ldots \end{aligned}\right\}. \qquad \ldots\ldots(9)$$

It is seen that, provided n be of reasonable size, z' is approximately normally distributed. Therefore instead of calculating the areas of the distribution curves of r, the problem can be put into terms of the area of the normal curve. This transformation is simple and, as will be seen later, it gives accurate results over the whole range of values of ρ and r.

The idea of constructing tables of the probability integral of r was suggested to the present writer by Karl Pearson in 1931. It was his desire to complete the series of extensive tables, calculated in his laboratories, associated with the fundamental tests of statistical sampling theory.* It is true that R. A. Fisher's z'-transformation is sufficient for most practical purposes, and is very simple to apply when n is large enough for us to take the expectation of z' as ζ and the standard error as $1/\sqrt{n-3}$. Nevertheless it is the function of a basic table, such as that given here, to form a standard against which the adequacy of approximations may be judged. This table should also be of some permanent value in providing a point of departure for the construction of useful working tools such as the charts which have been included in this volume.

SECTION II. CONSTRUCTION OF THE TABLES

Separate tables are given for sizes of sample from $n = 3$ to $n = 25$. For each size of sample ten distributions of r have been tabulated for values of $\rho = \cdot 0, \cdot 1, \cdot 2, \ldots \cdot 9$, each probability integral and ordinate being accurate to five decimal places. Tables are also given for $n = 50$, 100, 200 and 400 and a method of logarithmic interpolation is employed to obtain any intervening n. It was originally intended to include also tables for size of sample $n = 800$ and $n = 1600$. A small portion of each of these tables was calculated, but the addition of these extra values did not improve the results of logarithmic interpolation. In view of the labour involved (no ordinates of the curves had been calculated previously), it was decided to omit them.

When these tables were begun some time was spent in searching for a suitable method or methods by which the probability integral might be obtained. It would be superfluous to state these in detail, but it may be mentioned that the fitting of Pearson curves to the ordinates already tabled was carried out. This method had to be rejected, since the results obtained did not reach the required standard of accuracy. In 1932 F. Garwood (6) gave exact formulae for the probability integrals, but his results were expressed in terms of the ordinates, which themselves were only tabulated to five decimal places, and therefore the

* Most notable among these are the Tables of the Incomplete Gamma- and the Incomplete Beta-Functions.

areas calculated from his formulae had also to be rejected. Quadrature was the only method which gave accuracy, and it proved the simplest to use. A suitable quadrature formula was found and this was used throughout.

Case $\rho = 0$.

The probability law of r for any n, when ρ is zero, may be written

$$p(r \mid n, \rho = 0) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-2}{2}\right)}(1-r^2)^{\frac{n-4}{2}}. \qquad \text{......(10)}$$

Write $(1+r) = 2\omega$; then

$$\frac{\int_{-1}^{r} p(r \mid n, \rho = 0)\,dr}{\int_{-1}^{+1} p(r \mid n, \rho = 0)\,dr} = I_{\frac{1}{2}(1+r)}\left(\frac{n-2}{2}, \frac{n-2}{2}\right). \qquad \text{......(11)}$$

Or, alternatively, write $(1-r^2) = z$; then

$$\frac{\int_{-1}^{r} p(r \mid n, \rho = 0)\,dr}{\int_{-1}^{+1} p(r \mid n, \rho = 0)\,dr} = \begin{cases} \frac{1}{2}I_{1-r^2}\left(\frac{n-2}{2}, \frac{1}{2}\right) & r-\text{ve}, \\ 1-\frac{1}{2}I_{1-r^2}\left(\frac{n-2}{2}, \frac{1}{2}\right) & r+\text{ve}. \end{cases} \qquad \text{......(12)}$$

It is seen that the probability integral of (10) can be transformed into either of two incomplete B-function ratios. Quadrature was used to obtain the probability integral and the results were checked from the B-function tables.

Case $0 < \rho < +1$.

The method of quadrature used throughout to construct the tables is that due to Gregory (see for e.g. (7)). If $z_0, z_1, \dots z_p$ represent the ordinates of the curve, $f(x)$, at equal distances h apart, and Δ has its usual meaning as a difference symbol, then

$$\frac{1}{h}\int_{z_0}^{z_p} f(x)\,dx = [\tfrac{1}{2}z_0 + z_1 + z_2 + \dots + z_{p-2} + z_{p-1} + \tfrac{1}{2}z_p] - \tfrac{1}{12}[\Delta z_{p-1} - \Delta z_0] - \tfrac{1}{24}[\Delta^2 z_{p-2} + \Delta^2 z_0]$$
$$- \tfrac{19}{720}[\Delta^3 z_{p-3} - \Delta^3 z_0] - \tfrac{3}{160}[\Delta^4 z_{p-4} + \Delta^4 z_0] - \tfrac{863}{60480}[\Delta^5 z_{p-5} - \Delta^5 z_0] - \tfrac{275}{24192}[\Delta^6 z_{p-6} + \Delta^6 z_0] - \dots \text{etc.} \qquad \text{......(13)}$$

Formula (13) as quoted appears to entail much laborious calculation. If, however, we replace the differences by ordinates, i.e. if we write

$$\Delta z_{p-1} = z_p - z_{p-1}, \quad \Delta^2 z_{p-2} = z_p + z_{p-2} - 2z_{p-1},$$

and so on, then the formula is at once simplified and can be written in such a way that a minimum of calculation is necessary.

Gregory's formula (*i*) (up to and including sixth differences).

$$\begin{aligned} \frac{1}{h}\int_{z_0}^{z_p} f(x)\,dx = {} & [\tfrac{1}{2}z_0 + z_1 + \dots + z_{p-1} + \tfrac{1}{2}z_p] && + 0{\cdot}471{,}429\ [z_{p-3} + z_3] \\ & - 0{\cdot}195{,}776\ [z_p + z_0] && - 0{\cdot}260{,}607\ [z_{p-4} + z_4] \\ & + 0{\cdot}460{,}384\ [z_{p-1} + z_1] && + 0{\cdot}082{,}474\ [z_{p-5} + z_5] \\ & - 0{\cdot}546{,}536\ [z_{p-2} + z_2] && - 0{\cdot}011{,}367\ [z_{p-6} + z_6] \end{aligned} \qquad \text{......(14 a)}$$

Gregory's formula (*ii*) (up to and including eighth differences).

$$\frac{1}{h}\int_{z_0}^{z_p} f(x)\,dx = [\tfrac{1}{2}z_0+z_1+\dots+z_{p-1}+\tfrac{1}{2}z_p] \qquad -1\cdot140{,}564\ [z_{p-4}+z_4]$$
$$-0\cdot213{,}025^{-}[z_p+z_0] \qquad +0\cdot720{,}944\ [z_{p-5}+z_5]$$
$$+0\cdot589{,}020\ [z_{p-1}+z_1] \qquad -0\cdot297{,}855^{-}[z_{p-6}+z_6] \qquad \dots\dots(14b)$$
$$-0\cdot964{,}015^{-}[z_{p-2}+z_2] \qquad +0\cdot072{,}497\ [z_{p-7}+z_7]$$
$$+1\cdot240{,}890\ [z_{p-3}+z_3] \qquad -0\cdot007{,}893\ [z_{p-8}+z_8]$$

The simplicity of either of these formulae can be seen at a glance. Given the ordinates of the distribution of r for a fixed ρ and n, we first ascertain the chordal areas from z_0 to each succeeding ordinate. The corrective term involving $z_0, z_1, \dots z_6$, or $z_0, z_1, \dots z_8$, is next calculated, this being constant for the given distribution. There is now left only one calculation for each probability integral required. Either formula (14a) or (14b) was used throughout. The aim has been to provide tables of the probability integral which would be correct to five decimal places and it is hoped that this has been achieved. It is possible that in some cases the figure in the fifth decimal place may be one unit wrong. If this occurs it will be due to the fact that the ordinates from which the probability integral is obtained are themselves only correct to five decimal places.

The tables of ordinates published in "A Cooperative Study" were the foundation of the present tables. The values of the probability integral corresponding to these tables of ordinates were completed in 1934, but when examples came to be worked out from them, it was found that the differences were too large to allow of accurate interpolation. Accordingly more ordinates, and therefore more values of the probability integral, were calculated as it was found they were needed, and this accounts for the somewhat "lopsided" character of the table. Not all the ordinates which were calculated have been included. It is felt that the tables of ordinates as printed will give a sufficiently reasonable idea of the shape of the curves, and that interpolation between the ordinates is adequate for all practical purposes.

Calculation of Ordinates

Two methods were adopted for the calculation of the ordinates which were interpolated between those already published in "A Cooperative Study". It is known that

$$y_3 = \frac{1-\rho^2}{\pi\sqrt{1-r^2}}\left(\frac{1}{1-\rho^2r^2}+\frac{\rho r\arccos(-\rho r)}{(1-\rho^2r^2)^{\frac{3}{2}}}\right),$$

$$y_4 = \frac{(1-\rho^2)^{\frac{3}{2}}}{\pi}\left(\frac{3\rho r}{(1-\rho^2r^2)^2}+\frac{(1+2\rho^2r^2)\arccos(-\rho r)}{(1-\rho^2r^2)^{\frac{5}{2}}}\right),$$

where y_{n_1} is the ordinate of the curve $p(r \mid n, \rho)$ for $n = n_1$.

Hence by repeated application of the recurrence formula

$$y_{n+2} = \frac{2n-1}{n-1}\kappa_1 y_{n+1} + \frac{n-1}{n-2}\kappa_2 y_n,$$

where $$\kappa_1 = \frac{\rho r\sqrt{1-\rho^2}\sqrt{1-r^2}}{1-\rho^2r^2}, \qquad \kappa_2 = \frac{(1-\rho^2)(1-r^2)}{1-\rho^2r^2},$$

it was possible to obtain the ordinates for succeeding values of n, ranging from 5 to 25, for a given r and ρ. Where isolated ordinates were required the following formula was used:

$$y_n = \frac{n-2}{\sqrt{n-1}}(1-\rho^2)^{\frac{1}{2}}\chi(\rho, r)\left[1+\frac{\phi_1}{n-1}+\frac{\phi_2}{(n-1)^2}+\frac{\phi_3}{(n-1)^3}+\frac{\phi_4}{(n-1)^4}+\dots\right],$$

where
$$\log \chi(\rho, r) = -(n-1)\log \chi_1 - \log \chi_2,$$
$$\chi_1 = \frac{1-\rho r}{\{(1-\rho^2)(1-r^2)\}^{\frac{1}{2}}}, \qquad \chi_2 = \frac{\sqrt{2\pi}\{(1-\rho^2)(1-r^2)\}^{\frac{1}{4}}}{(1-\rho r)^{\frac{1}{2}}},$$
and
$$\phi_1 = \frac{r\rho+2}{8}, \qquad \phi_2 = \frac{(3r\rho+2)^2}{128}, \qquad \phi_3 = \frac{5\{15(r\rho)^3+18(r\rho)^2-4(r\rho)-8\}}{1024},$$
$$\phi_4 = \frac{3675(r\rho)^4+4200(r\rho)^3-2520(r\rho)^2-3360(r\rho)-336}{32768}.$$

This last formula is only approximate, and was originally intended to be used for $n > 25$, but by checking against the true results obtained from the recurrence formula, it was found to be reasonably accurate, even for n as low as 10.

Checks

The differences of each table of the probability integral were found on completion, and this process was sufficient to detect any gross errors which were made. A closer check presented and still presents more difficulty. In the main, panel-area formulae were used to check differences between successive tabled areas.

Panel Area between z_{-1} and z_0 (5 ordinates).

$$\int_{z_{-1}}^{z_0} f(x)\,dx = \frac{h}{720}[-19z_{-2}+346z_{-1}+456z_0-74z_1+11z_2]. \qquad \ldots\ldots(15a)$$

Panel Area between z_0 and z_1 (6 ordinates).

$$\int_{z_0}^{z_1} f(x)\,dx = \frac{h}{1440}[475z_0+1427z_1-798z_2+482z_3-173z_4+27z_5]. \qquad \ldots\ldots(15b)$$

Panel Area between z_0 and z_1 (8 ordinates).

$$\int_{z_0}^{z_1} f(x)\,dx = \frac{h}{120960}[36799z_0+139849z_1-121797z_2+123133z_3-88547z_4+41499z_5-11351z_6+1375z_7]. \qquad \ldots\ldots(15c)$$

In cases where there was disagreement between the results obtained from formulae (14) and formulae (15), it was usually found to be because formulae (15) did not include enough differences, and the result obtained from (14) was allowed to stand, after being carefully checked by both formulae (14). It is, however, too much to hope that tables of this size will be completely free from error, and the writer will be grateful for notification when any errors are found.

SECTION III. INTERPOLATION. METHODS AND ILLUSTRATION

For very many purposes modern statistical method calls for little more knowledge of a probability integral than a tabulation of the various "significance levels", and for those who wish only to use these levels, charts are published at the end of this volume. Sometimes, however, it is necessary to obtain a probability by interpolation, so the interpolation formulae which have proved most accurate will be stated briefly. These formulae are general adaptations of those already given by Karl Pearson (8). The calculation is straightforward but in most cases somewhat laborious.

Suppose that for a given sample of size n, it is desired to find the probability integral $z_{\theta\chi}$, for given ρ_χ and r_θ. It is simplest to set out the ordinates used in a diagram (see Fig. 1).

Using first differences.

$$z_{\theta\chi} = \phi\psi z_{0,0} + \phi\chi z_{0,1} + \theta\psi z_{1,0} + \theta\chi z_{1,1}. \qquad \ldots\ldots(16)$$

Up to and including third differences.

$$z_{\theta\chi} = \{1+\tfrac{1}{2}(\theta\phi+\chi\psi)\}\{\phi(\psi z_{0,0}+\chi z_{0,1})+\theta(\psi z_{1,0}+\chi z_{1,1})\}$$
$$-\tfrac{1}{6}\theta\phi\{(1+\phi)(\psi z_{-1,0}+\chi z_{-1,1})+(1+\theta)(\psi z_{2,0}+\chi z_{2,1})\}$$
$$-\tfrac{1}{6}\chi\psi\{(1+\psi)(\phi z_{0,-1}+\theta z_{1,-1})+(1+\chi)(\phi z_{0,2}+\theta z_{1,2})\}. \quad \text{......(17)}$$

Illustration.

Given $n = 20$, $\rho = \cdot 277$, $r = \cdot 185$, what is the value, by interpolation, of

$$\int_{-1}^{+\cdot 185} p(r \mid n=20,\ \rho = \cdot 277)\,dr\,? \quad \text{......(18)}$$

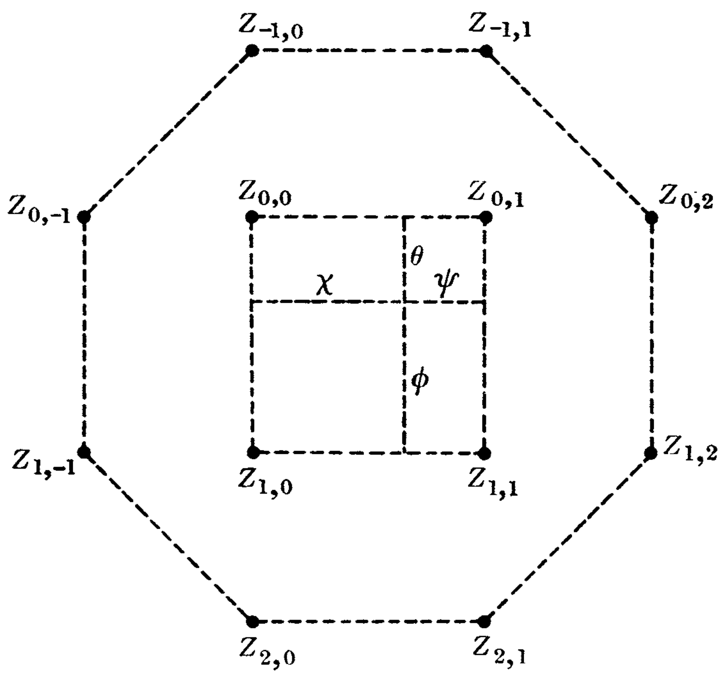

Fig. 1.

In order to test the accuracy of the interpolation formulae (16) and (17), ordinates of the distribution of r for $n = 20$ and $\rho = \cdot 277$ were calculated, and (18) found by quadrature. This was found to be ·33007. The values of z required for interpolation are:

	$\rho = \cdot 1$	$\rho = \cdot 2$	$\rho = \cdot 3$	$\rho = \cdot 4$
$r = \cdot 10$		·32570	·18188	
$r = \cdot 15$	·57957	·40536	·24280	·11854
$r = \cdot 20$	·66144	·49060	·31482	·16617
$r = \cdot 25$		·57806	·39668	

Here $\theta = \cdot 7$ and $\chi = \cdot 77$. Applying (16) we see that

$$\int_{-1}^{+\cdot 185} p(r \mid n=20,\ \rho = \cdot 277)\,dr = \cdot 336785,$$

and using (17),

$$\int_{-1}^{+\cdot 185} p(r \mid n=20,\ \rho = \cdot 277)\,dr = \cdot 33021.$$

Linear interpolation as in (16) is seen to be a rough approximation, the interpolated value differing from the true value in the third decimal place; but the result obtained from (17) is reasonable, there being a difference of only ·00014 between the true and interpolated values.

Formula (17) is for use in the middle of the table. We may quote one further formula, which will be of use when interpolating on the border of the table. For this, the scheme will be:

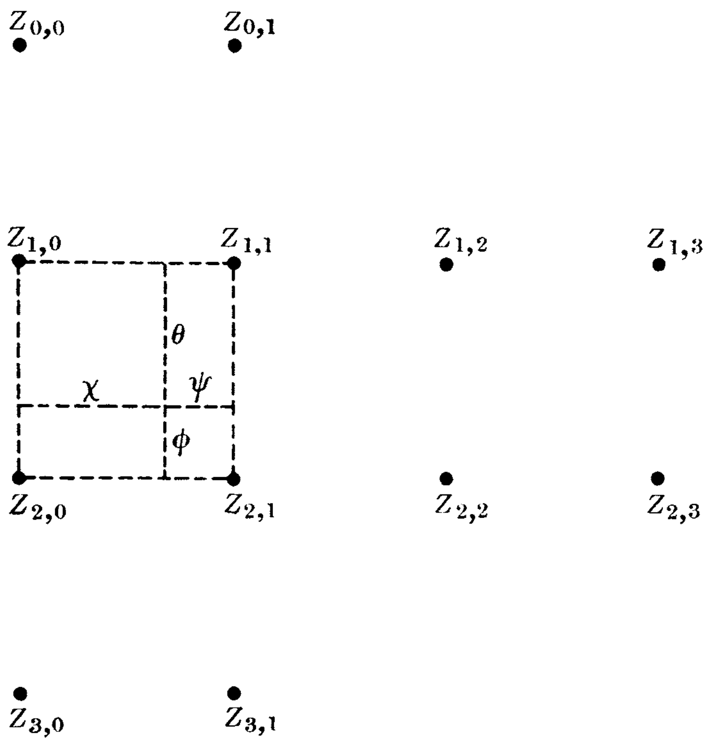

Fig. 2.

and $z_{\theta\chi}$ is obtained by substitution in the following:

$$\begin{aligned} z_{\theta\chi} = {} & \phi\{1+\tfrac{1}{3}\theta(1+\phi)-\tfrac{1}{6}\theta(1+\theta)\}\{\psi z_{0,0}+\chi z_{1,1}\}+\theta\{1+\tfrac{1}{3}\phi(1+\theta)-\tfrac{1}{6}\phi(1+\phi)\}\{\psi z_{2,0}+\chi z_{2,1}\} \\ & -\tfrac{1}{6}\chi\psi\{(4+\psi)[\phi z_{1,0}+\theta z_{2,0}]-3(3+\psi)[\phi z_{1,1}+\theta z_{2,1}]\}-\tfrac{1}{6}\chi\psi\{3(2+\psi)[\phi z_{1,2}+\theta z_{2,2}]-(1+\psi)[\phi z_{1,3}+\theta z_{2,3}]\} \\ & -\tfrac{1}{6}\theta\phi\{(1+\phi)[\psi z_{0,0}+\chi z_{0,1}]+(1+\theta)[\psi z_{3,0}+\chi z_{3,1}]\}. \qquad \ldots\ldots(19) \end{aligned}$$

Using (19) on the same figures as before, the tabulated values required are:

·32570	·18188		
·40536	·24280	·11854	·04383
·49060	·31482	·16617	·06696
·57806	·39668		

and we get
$$\int_{-1}^{+\cdot185} p(r \mid n=20, \rho=\cdot277)\,dr = \cdot32979,$$
differing from the true value by ·00028.

These three formulae should be enough for interpolation into all parts of the table where the specific size of sample is given.

Possibility of Graphical Interpolation

Formulae (17) and (19), while giving accurate results, often need laborious calculation in application. Graphical interpolation is possible if the probability integral is required to be correct to two units only. Let us consider the possibility of finding (18) graphically. Plot r along the abscissa, as ordinates take the corresponding values of the probability integral, and plot the two curves $\rho = \cdot 2$ and $\rho = \cdot 3$, for size of sample $n = 20$. Provided these be drawn on a reasonable scale it is possible to read off the probability integral

Diagram showing Graphical Interpolation for size of sample $n = 20$

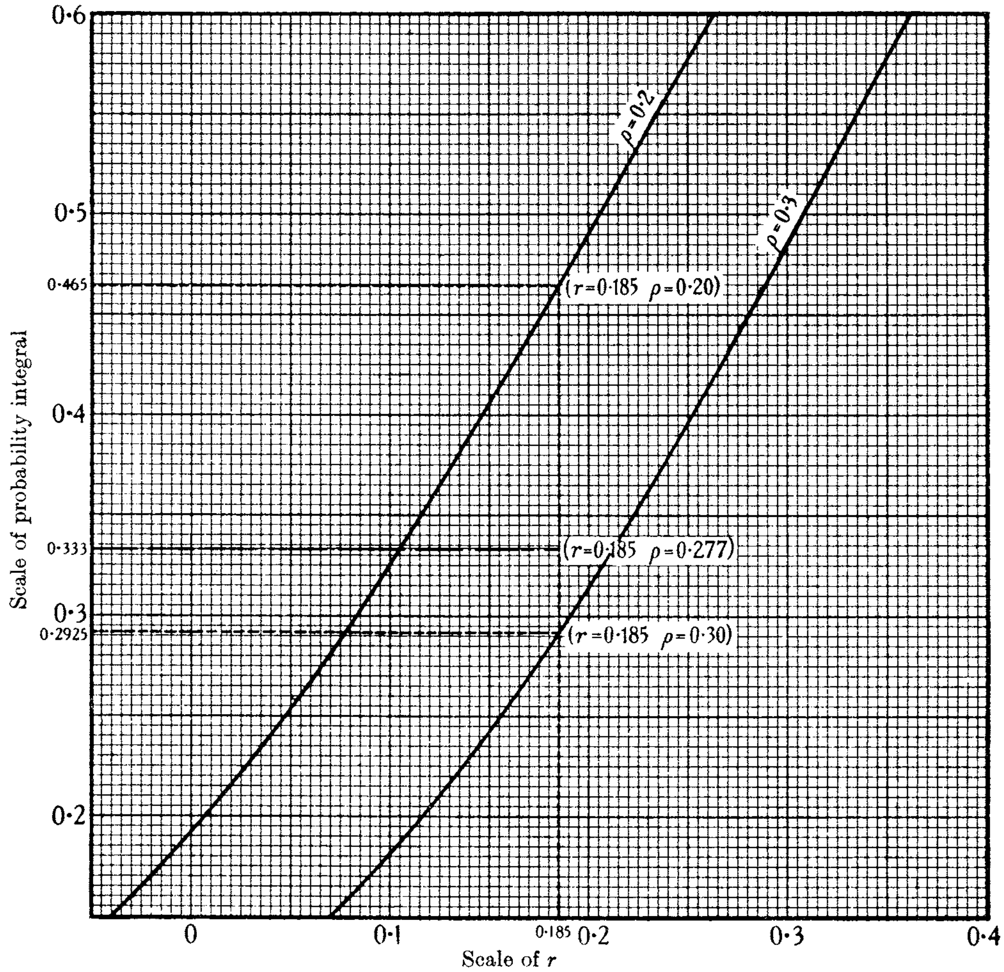

Fig. 3.

N.B. $\cdot 465 - \cdot 2925 = \cdot 1725$, $\cdot 1725 \times \cdot 23 = \cdot 0397$, $\cdot 0397 + \cdot 2925 = \cdot 3332$.

for $\rho = \cdot 277$ and $r = \cdot 185$. The probability integral obtained by this method was ·333, which is slightly more accurate than the result obtained by linear interpolation by formula (16). This result is sufficiently good if a rough guide is all that is required, and hence it is suggested that this method be used when a quick approximation to the probability integral is desired.

Logarithmic Interpolation for $n > 25$

The interpolation formulae discussed previously in this section are chiefly applicable to the first part of the table, where the size of sample is from $n = 3$ to $n = 25$. The second part of the table gives the ordinates and areas for samples of sizes 50, 100, 200 and 400. We shall now consider interpolation for the probability integral in this second part.

Consider Lagrange's mid-point formula for the graduation of four ordinates at equal distance, h, apart (9). Suppose the four ordinates to be z_0, z_1, z_2, z_3. Then if z_x, the required ordinate, be situated, as shown in Fig. 4, at a distance hx from the foot of the first ordinate, z_0,

$$\begin{aligned} z_x = z_0 + \tfrac{1}{6}\{&x^3[(z_3-z_2)-2(z_2-z_1)+(z_1-z_0)] \\ &-3x^2[(z_3-z_2)-3(z_2-z_1)+2(z_1-z_0)] \\ &+x[2(z_3-z_2)-7(z_2-z_1)+11(z_1-z_0)]\}. \qquad \ldots\ldots(20) \end{aligned}$$

We may use this formula in order to construct tables for large values of n. The method is as follows. Using a logarithmic scale for n it is seen that the probability integral is tabled at five equidistant values of this new variable, corresponding to an argument interval of log 2, i.e. at

$$\log 25, \quad \log 25 + \log 2, \quad \log 25 + 2\log 2, \quad \log 25 + 3\log 2, \quad \log 25 + 4\log 2.$$

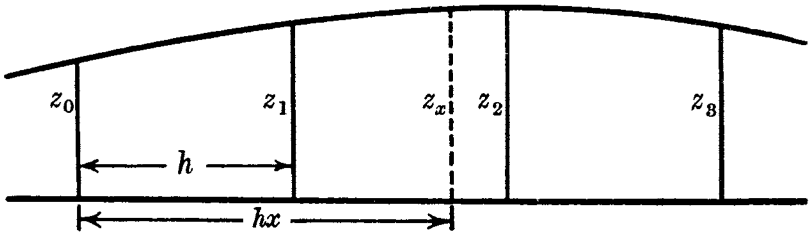

Fig. 4.

If then, for example, it is wished to interpolate for $n = 160$, the values of z_0, z_1, z_2, z_3 to be inserted in (20) will be the values of the probability integral at $n = 50$, 100, 200 and 400, and

$$x = (\log 160 - \log 50)/\log 2 = 1{\cdot}678{,}072.$$

Both for illustration and as a test of the accuracy of (20), the probability integral for $\rho = {\cdot}8$ and $n = 160$ is given below for several values of r. The exact result obtained by quadrature from ordinates calculated from the formulae of p. x is also given.

Table I. *Probability integral for* $\rho = {\cdot}8$

r \ n	50	100	200	400	160 (by four-point)	160 (exact)
·60	·00237	·00005⁻	—	—	·00000	·00000
·65	·01146	·00067	—	—	·00000	·00003
·70	·04945	·01023	·00054	—	·00168	·00172
·75	·17732	·10000	·03643	·00582	·05422	·05388
·80	·47693	·48387	·48864	·49200	·48729	·48729
·85	·84835⁺	·93506	·98547	·99909	·97323	·97402
·90	·99374	·99986	1·00000	1·00000	1·00000	1·00000

It will be seen that the agreement between the exact values and the results of applying (20) is fairly good. These results are again used in Section V, where they are compared with the results of applying the z'-transformation.

If the probability integral is required for a value of ρ lying between the tabled values, it will be necessary to construct two such tables as that given above for the tabled values of ρ lying immediately above and immediately below the desired value. These obtained, the method of graphical interpolation suggested on p. xiv will give the required value correct to two decimal places. The writer is of the opinion that the increase in accuracy would be very little if four tables such as the above were calculated and interpolation by second differences employed.

It may be that the probability integral for a given ρ, r and large n will be wanted quickly and approximately. To this end diagrams, numbered I–X, have been constructed and will be found at the end of this introduction. A separate diagram is given for each value of ρ which is tabled. n is plotted along the abscissa and curves are drawn of the probability integral for given values of r. Hence for a given n and ρ, the values of the probability integral for r at an argument of ·05 can be read off immediately. For a value of ρ lying between the tabled values the probability integral for the ρ lying immediately above and immediately below may be found, and the method of p. xiv employed.

SECTION IV. USE OF THE TABLES AND ILLUSTRATIONS

In whatever field he is working the applied mathematician is concerned with bridging a gap between a conceptual mathematical model and the data of his experience. Thus the mathematical statistician has need to consider how a precise but abstract theory of probability may be employed most usefully to draw inferences from observation. There may be differences of opinion as to the best methods of answering some of the questions discussed below, but there should be general agreement on the importance of defining with precision the terms of the questions asked and the principles adopted in answering them. For this reason it has seemed well to introduce the illustrations of the use of tables and charts given below with a somewhat formal statement of the guiding principles which the writer has followed in their solution. At the same time no claim is made that the types of problem illustrated are exhaustive, nor that the approach to their solution is unique; the problems discussed might certainly have been formulated in a different way, the same tables or charts being used in their solution.

The problems of practical statistics which call for the introduction of the theory of probability will almost always be found at the root to be concerned with the relation between what may be termed the collective character* or characters of a sample, and the collective character or characters of the population from which the sample has been randomly drawn. It is rarely possible to determine a collective character directly from a knowledge of the population, for in most cases the populations studied are either infinite or very large, and even if it were possible the question would arise as to whether we are justified in spending much time and labour in so doing. In practice we take our randomly drawn sample, and use it to obtain information about the population we are studying. Provided the sample is of reasonable size we may do this with a fair degree of accuracy.

Thus if the collective character under consideration in the population is the coefficient of correlation, ρ, between two variable characteristics x and y, we may wish to obtain answers to such questions as the following:

1. Are the observed data in a sample consistent with the hypothesis that in the population

(i) $\rho \geqslant \rho_0$, (ii) $\rho \leqslant \rho_0$, (iii) $\rho = \rho_0$,

where ρ_0 is some specified value?

2. How may the observed data in a sample be used to the best advantage in order to calculate limits ρ_b and ρ_a, such that the statements

(i) $\rho_b \leqslant \rho \leqslant +1$, (ii) $-1 \leqslant \rho \leqslant \rho_a$, (iii) $\rho_b \leqslant \rho \leqslant \rho_a$,

regarding the unknown value of ρ in the population sampled may be made with given degrees of confidence?

3. Suppose we have k ($k \geqslant 2$) independent randomly drawn samples. Are the observed data consistent with the hypothesis that in the k populations sampled the coefficients of correlation, $\rho_1, \rho_2, \ldots \rho_k$, are all equal to

(i) a specified value, ρ_0,

(ii) a common but unspecified value, ρ?

* As far as the writer is aware, the term "collective character" was first used by J. Neyman in his paper "On two different aspects of the Representative Method", *J. Roy. Statist. Soc.* XCVII (1934), 561. He obtained it by translating a Russian phrase. Applied respectively to the sample and the population it is used instead of the terms "statistic" and "parameter".

Questions 1 and 3 are concerned with the testing of a statistical hypothesis. We ask whether the data are consistent with a specified hypothesis or not; if we decide to say "No", we risk one form of error, that of rejecting the hypothesis when it is true. In any precise test it should be possible to fix what is commonly called the significance level so as to control this risk of error, say ϵ, at some prescribed figure, e.g. $\epsilon = \cdot 01$ or $\cdot 05$. If, on the other hand, we say "Yes, the data appear consistent with the hypothesis", it may nevertheless happen that the hypothesis tested is false, and that through the inadequacy of the data we have failed to detect the fact that some alternative hypothesis is true. These two types of error cannot altogether be avoided in testing a statistical hypothesis; if the test is arranged to reduce the risk of the first it will increase that of the second, and *vice versa*. Some illustration of this is given below. Question 2 leads to the problem of interval estimation. Here again two considerations will be taken into account; for example, in question 2 (iii) we must consider both (*a*) the risk that the interval ρ_b to ρ_a fails to cover the unknown value of ρ in the population sampled, and (*b*) the breadth of the interval ρ_b to ρ_a, which for the given risk we should like to be as narrow as possible. In this connexion the definition of narrowness will have to be discussed.

It is necessary to emphasize again that in the work which follows we assume that the samples have been randomly drawn from some population in which the variates x and y under consideration follow the normal bivariate distribution given in equation (1) above. That the results will be approximately true even when the distribution is far removed from normal is suggested by the empirical work of E. S. Pearson* and others(10, 11), but so far it has not been established mathematically.

Question 1

A sample consisting of n pairs of observations is available.

(i) Is $\rho \geqslant \rho_0$? Here the admissible hypotheses alternative to that tested are that $\rho < \rho_0$.

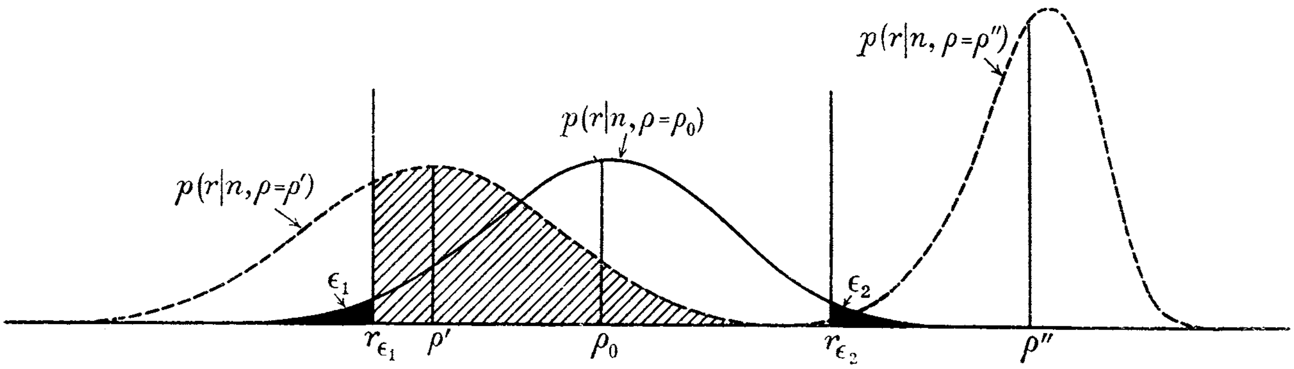

Fig. 5.

Fig. 5 shows hypothetical sampling distributions of r for three values of ρ, and size of sample equal to n. We want to test if $\rho \geqslant \rho_0$ and, bearing in mind the alternative hypotheses, we may suggest the following rule:

Reject the hypothesis tested if $r < r_{\epsilon_1}$, where

$$\epsilon_1 = \int_{-1}^{r_{\epsilon_1}} p(r \mid n, \rho_0)\, dr. \qquad \text{......(21)}$$

Accordingly, if $\rho > \rho_0$, e.g. equals ρ'', then the chance of rejecting this hypothesis when it is true will be $< \epsilon_1$. If $\rho = \rho_0$ the chance of rejection when the hypothesis is true is exactly equal to ϵ_1. On the other hand,

* E. S. Pearson, *Biometrika*, XXI (1929), 356–60: "The results suggest that the normal bivariate surface can be mutilated and distorted to a remarkable degree without affecting the frequency distribution of r."

if in fact $\rho = \rho'$ and the hypothesis tested were false, the risk of failing to detect this is measured by the shaded area in Fig. 5 under $p(r \mid n, \rho = \rho')$ to the right of the ordinate at r_{ϵ_1}. Clearly in the case illustrated, if $\rho = \rho'$, we should be almost as likely as not to fail to detect the fact that the hypothesis $\rho \geqslant \rho_0$ was false.

(ii) Is $\rho \leqslant \rho_0$? The admissible alternative hypotheses will be that $\rho > \rho_0$, and in a way similar to the above we may set up the following rule:

Reject H_0, the hypothesis tested, if $r > r_{\epsilon_2}$, where

$$\epsilon_2 = \int_{r_{\epsilon_2}}^{+1} p(r \mid n, \rho_0)\, dr. \qquad \text{......(22)}$$

The four charts provided at the end of this volume can be used to obtain the limits r_{ϵ_1} and r_{ϵ_2}, for values of ϵ_1 and ϵ_2 equal to ·005, ·01, ·025 and ·05, and for varying values of n and ρ. If it is desired to take a different probability level from those given, it will be possible to find the necessary figures from the main tables.

(iii) Is $\rho = \rho_0$? Here the admissible alternative hypotheses to that tested will be that $-1 < \rho < \rho_0$ (e.g. $\rho = \rho'$), and $\rho_0 < \rho < +1$ (e.g. $\rho = \rho''$). The test we may set up will be a combination of those used in questions 1 (i) and 1 (ii). We may postulate as our rule:

Reject the hypothesis tested if $r > r_{\epsilon_2}$ or $r < r_{\epsilon_1}$, where ϵ_1 and ϵ_2 are defined as above and

$$\epsilon_1 = \epsilon_2 = \tfrac{1}{2}\epsilon, \qquad \text{......(23)}$$

where ϵ will be the chance of rejecting the hypothesis tested when it is true. It is easily seen that this risk of rejection will be the same if ϵ_1 is not equal to ϵ_2 so long as

$$\epsilon_1 + \epsilon_2 = \epsilon. \qquad \text{......(24)}$$

Neyman and Pearson (12) have shown that in certain cases if we take

$$\epsilon_1 = \epsilon_2,$$

we shall be less likely to reject the hypothesis tested when it is false, than when it is true. A test leading to such consequences they have termed biased.* Such a situation would arise if in Fig. 5, for any curves of the system $p(r \mid n, \rho)$ having ρ either below or above ρ_0, the proportional area included between ordinates at r_{ϵ_1} and r_{ϵ_2} was greater than $1-\epsilon$. It was decided to test whether the rejection limits obtained from the distribution of r by taking equal tail areas were biased. Following the procedure of Neyman and Pearson an unbiased test may be obtained by solving for r_{ϵ_1} and r_{ϵ_2} from the two following equations:

$$\int_{r_{\epsilon_1}}^{r_{\epsilon_2}} p(r \mid n, \rho)\, dr = 1 - \epsilon, \qquad \text{......(25)}$$

$$\frac{d}{d\rho}\int_{r_{\epsilon_1}}^{r_{\epsilon_2}} p(r \mid n, \rho)\, dr = 0. \qquad \text{......(26)}$$

This solution has been investigated and a note on unbiased limits for r has already been published (13). We may state here the conclusion which was reached. It was found that the rejection limits obtained by taking unbiased limits for r differed very little from the limits which were obtained by taking equal tail areas from the r-distribution, and that for all practical purposes these two sets of limits could be regarded as coincident. This result is not surprising when it is remembered that by R. A. Fisher's z'-transformation the distribution curves of r are transformed approximately into a series of normal curves. There is a slight error introduced by this transformation, and this is the reason why the unbiased limits and the equal tail-area limits are not quite coincident, but this error is usually so small as to be negligible.

* The word "biased" throughout the next few pages will be used in Neyman and Pearson's sense.

Illustration. Question 1. *Case* (*i*).

In the production of a certain aluminium die-casting previous experience over a long period of time had shown a manufacturer that two measurements of quality, namely tensile strength (x), measured in pounds per square inch, and hardness (y), measured in terms of Rockwell's E, were both approximately normally distributed. Since the determination of x involves the destruction of the casting, it is desired to use the character y as a measure of strength in its place, and for this purpose it is considered essential that the correlation, ρ, between the value of x and y in the same specimen should be at least as high as $+\cdot 80$. Tests of x and y on 25 specimens are available, giving a sample correlation, r, of $+\cdot 641$. Should it be concluded that the correlation between the characteristics in this type of casting is insufficient to justify the use of y in predicting strength?

In statistical terminology we see that it is necessary to test the hypothesis that $\rho \geqslant +\cdot 80$, the admissible alternative hypotheses being that $\rho < +\cdot 80$.

Using the tables we find that $P\{r \leqslant \cdot 641 \mid n = 25, \rho = \cdot 80\} = \cdot 04584$.* From Charts I, II, III, IV, respectively, we see that using the

·05 limit we reject the hypothesis tested if			r is less than		·65,
·025	,,	,,	r	,,	·605,
·01	,,	,,	r	,,	·55,
·005	,,	,,	r	,,	·51.

It is therefore seen that, if in the sampled material the correlation between the two characters was ·80 or more, so small a value of r as that observed would be expected to occur through chance sampling fluctuations less than once in 20 times, when testing 25 specimens. On the assumption that the specimens tested are a random selection from the material, the manufacturer would feel very doubtful whether the correlation between x and y was high enough for his purpose, although he might be well advised to examine a further sample before rejecting the hypothesis, $\rho \geqslant +\cdot 80$.

Illustration. Question 1. *Case* (*iii*).

Let us suppose that in dealing with the same material as in Case (i) the manufacturer had found from past experience that $\rho = +\cdot 63$, and as a routine test for control of quality he proposed that in the future random samples of 20 specimens should be drawn from each batch of several hundred castings, x and y measured, and the correlation between them calculated. What control limits r_1 and r_2 should he specify in order that he may be reasonably certain of detecting whether ρ for the batch had altered appreciably from $+\cdot 63$?

We shall assume (i) that the batch is so large that we may regard the sample as being drawn from an infinite population, (ii) that within the batch the quality of the material is homogeneous. By making r_1 and r_2 close to $\rho_0 = +\cdot 63$, the manufacturer would reduce the risk of passing material for which ρ was much greater or much less than $+\cdot 63$. This he might do if he were trying to establish a rigorous control. But if r_1 and r_2 are too close to $\rho_0 = +\cdot 63$, he would run a second risk in that he would often reject material which was really satisfactory. This would be possible owing to the wide variation in r for samples of 20. Therefore, before setting up his two limits r_1 and r_2 he must decide upon the risk he is willing to undertake. We shall suppose that he is content with the control if the limits chosen entail a risk of rejecting satisfactory material 5 times in 100, i.e. 5 times in 100 he will run the risk of rejecting the hypothesis $\rho = \rho_0 = +\cdot 63$, when it is actually true.

We obtain the limits r_1 and r_2 from Chart II, which shows that

$$P\{(\cdot 275 \leqslant r \leqslant \cdot 845) \mid n = 20, \rho = +\cdot 63\} = \cdot 05.$$

The two limits for r that he should therefore set up will be $r_1 = +\cdot 275$, $r_2 = +\cdot 845$.

* This useful shorthand is often found in statistical papers. Put into words it means "the probability that r is less than or equal to ·641, given that $n = 25$ and $\rho = \cdot 80$, is equal to ·04584".

Question 2. To determine a confidence interval ρ_b to ρ_a for ρ.

Certain aspects of the problem of estimation have recently been advanced considerably by the work of R. A. Fisher (14) on fiducial probability and J. Neyman (15) on interval estimation. The procedure and terminology developed by the latter will be followed, but from the practical point of view there is no substantial difference between the results reached by either approach. For the sake of clarity the problem has been divided into three classes.

Case (i). To determine ρ_b so that we may make the statement $\rho_b \leqslant \rho \leqslant +1$ with a given degree of confidence. If we turn to Chart I, with which is associated a confidence coefficient of ·90, and as an illustration consider the lower of the two curves marked $n = 10$, we know that, whatever be the value of ρ in the sampled population, we shall expect in repeated sampling to find that 5 out of 100 samples of 10 give a value of r falling below this curve. Thus if we consider the set of points (r, ρ), which we may meet when drawing samples of 10 from normal bivariate populations, we should expect 5 per cent. of these points to lie below the lower curve, and 95 per cent. above it. It follows that if in general, in a given sample of size $n = n_0$,

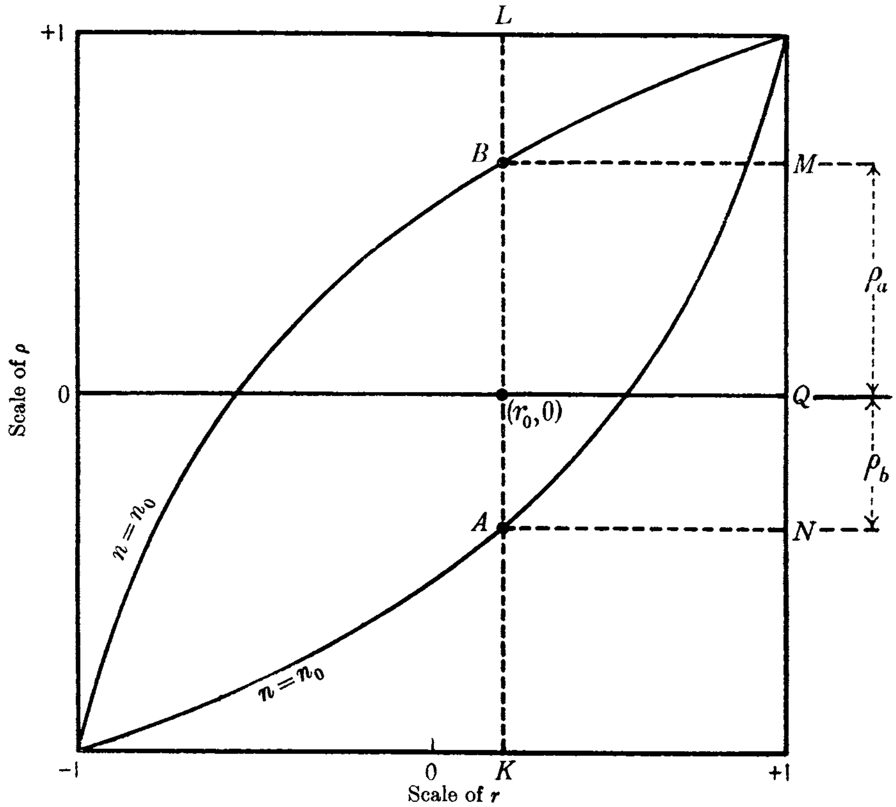

Fig. 6.

we find the sample correlation coefficient $r = r_0$, we may adopt the following procedure to determine the limit ρ_b:

Plot the point $(r_0, \rho = 0)$ and draw a line parallel to the axis of ρ through this point. Suppose this line cuts the lower curve for $n = n_0$ at the point A. Draw the line AN perpendicular to the axis of ρ, cutting the axis in the point N. Then $QN = \rho_b$, and we may make the statement

$$\rho_b \leqslant \rho \leqslant +1 \qquad \text{......(27)}$$

with a degree of confidence measured by the confidence coefficient of ·95. We may express this by saying that 95 times in 100 we shall expect the interval ρ_b to $+1{\cdot}00$ to cover the true population value. The confidence coefficient represents the probability that the statement (27) will be correct if we follow this procedure. If Charts II, III and IV are used the confidence coefficients will be ·975, ·99 and ·995, respectively.

Case (ii). To determine ρ_a so that we may make the statement $-1 \leqslant \rho \leqslant \rho_a$ with a given degree of confidence. In case (ii), given r_0 as in case (i), continue the line through $(r_0, \rho = 0)$, which is parallel to the axis of ρ, until it cuts the upper of the two curves for $n = n_0$. Suppose the point of intersection to be B, and draw a line BM, perpendicular to the axis of ρ, and cutting this axis in the point M. Then $MQ = \rho_a$ and we may associate with the statement

$$-1 \leqslant \rho \leqslant \rho_a \qquad \ldots\ldots(28)$$

a confidence coefficient of ·95 if Chart I is used.

Case (iii). To determine ρ_a and ρ_b so that we may make the statement $\rho_b \leqslant \rho \leqslant \rho_a$ with a given degree of confidence. Using Chart I, and the pair of curves for $n = n_0$, where n_0 is the specified size of sample, we know that of the set of points (r, ρ) that may be met in our statistical experience when randomly drawing samples of size n_0, 90 per cent. will fall within the lozenge-shaped belt between the two curves, 5 per cent. above the upper curve, and 5 per cent. below the lower curve. Hence if for an observed value of $r = r_0$ in a sample we determine, as in Fig. 6, the points A and B, M and N, and therefore ρ_a and ρ_b, where $\rho_b = NQ$ and $\rho_a = MQ$, we may associate with the statement

$$\rho_b \leqslant \rho \leqslant \rho_a \qquad \ldots\ldots(29)$$

a confidence coefficient of ·90. Intervals with confidence coefficients of ·95, ·98 and ·99 can similarly be obtained from Charts II, III and IV, respectively.

In the present type of problem it is clear that we could obtain an infinite variety of belts provided that (24) holds. We should have the same risk of error in making statement (29) if the two curves of Fig. 6 were based on those values of r_{ϵ_1} and r_{ϵ_2} obtained by a consideration of (21), (22) and (24) instead of those obtained from (21), (22) and (23). For instance, a belt with confidence coefficient ·90 could be obtained by taking $\epsilon_1 = {\cdot}02$, $\epsilon_2 = {\cdot}08$, as well as with $\epsilon_1 = \epsilon_2 = {\cdot}05$.

One such type of belt has been discussed in the note on unbiased limits for r, to which we have already referred (13). It has been shown by J. Neyman (15) that, in the case of certain skew distributions, such "unbiased" belts have definite theoretical advantages over those obtained by taking equal tail areas. In the case of the r-distribution, however, evidence given in the paper (13) suggests that the equal tail-area belt and the unbiased belt may be considered as coincident for all practical purposes.

Illustration.

The width of span (x) and length of forearm (y) of 20 males have been measured, and the correlation between these two variates is found to be $+{\cdot}550$. Assuming that width of span and length of forearm are both approximately normally distributed, what interval will cover the correlation coefficient between x and y in the population?

The sample correlation coefficient is $+{\cdot}550$. Using Chart I and the pair of curves for $n = 20$, we see that we may make the statement

$${\cdot}21 \leqslant \rho \leqslant {\cdot}76,$$

with a degree of confidence measured by a confidence coefficient of ·90. This is equivalent to saying that in repeated sampling nine times out of ten we expect the interval ·21 to ·76 to cover the true population value.

We may decide that the risk of the interval failing to cover the true value once in ten times is too great. Accordingly we would turn to Chart II and make the statement

$$P\{{\cdot}160 \leqslant \rho \leqslant {\cdot}785\} = {\cdot}95.$$

Here the risk of failing to cover the true population value by the interval ·160 to ·785 is reduced to ·05, but it should be noticed that in reducing the risk of error we have increased the breadth of the interval.

We should naturally like the interval for ρ to be as narrow as possible, but we can make this interval narrow only by increasing the risk of being wrong. It is therefore necessary to balance the consequences entailed by a wrong decision against the advantages of the narrow interval.

The sample we are discussing was in fact randomly drawn from a population of 10,000 males. The correlation between width of span and length of forearm for this population was found to be ·758.

Question 3

In attempting to use any single comprehensive criterion to test hypotheses regarding the values of unknown parameters in more than one population, we find that little progress has so far been made in the development of such tests except in the case where the sample estimates of the unknown population parameters are normally distributed. If, for example, we have samples from two populations with correlations ρ_1 and ρ_2, the classical method of testing the hypothesis that ρ_1 equals ρ_2 would be to calculate the ratio of the difference between the sample correlations to an estimate of the standard error of this difference, i.e.

$$\frac{r_1 - r_2}{\sqrt{\frac{(1-r_1^2)^2}{n_1-1} + \frac{(1-r_2^2)^2}{n_2-1}}}, \qquad \text{......(30)}$$

and refer this ratio to the normal probability scale. This procedure is adequate when dealing with large samples, provided the hypothetical common value of $|\rho|$ is not too near unity. It may however be extremely inaccurate in other cases. The sampling distribution of (30) is unknown, and its form would be extraordinarily difficult to determine, since r_1 and r_2 follow different non-normal distributions, depending on the value of n_1 and n_2 and that of the unknown common ρ. Further, even if the sampling distribution were obtainable, the test might be less efficient in detecting true differences between ρ_1 and ρ_2 than other tests which might be devised.

R. A. Fisher's z'-transformation has the great practical advantage that it provides, instead of r, a quantity z' which is approximately normally distributed, and whose standard error is practically independent of ρ. Thus when we have samples from several populations, we may test hypotheses regarding the population ρ's by applying to the sample z''s the appropriate tests from "normal theory". Since, however, the distribution is not precisely normal or independent of ρ, some approximation is entailed by such a procedure and it is therefore important to note that for one type of problem, such as Question 3 (i), the present tables provide an exact test.

Question 3 (*i*)

k samples, containing respectively $n_1, n_2, \dots n_k$ observations of two correlated variables x and y, have been drawn from k normal bivariate populations, with unknown correlation coefficients $\rho_1, \rho_2, \dots \rho_k$. Are the data consistent with the hypothesis that

$$\rho_1 = \rho_2 = \dots = \rho_k = \rho_0, \qquad \text{......(31)}$$

where ρ_0 is some specified value?

In order to determine the best form of test to use, it is necessary to decide upon the kind of alternatives to (31) that appear most likely, having regard to general considerations of the type of problem dealt with. If the ρ's are not all equal to ρ_0 we might have grounds for thinking (*a*) that they would have some other common value differing from ρ_0, (*b*) that they might be unequal but all $>\rho_0$ (or $<\rho_0$), (*c*) that they might assume any different and unequal values whatsoever. In practice it is evident that we cannot be *certain* of the class of admissible hypotheses alternate to those tested; nevertheless we shall generally have a fairly clear idea of the form of departure from (31) which it is most important for us not to overlook. Consequently we shall prefer to use that test which is most likely to detect such a departure if it exists.

Consider first the alternative (*c*) mentioned above. Since ρ_0 is specified in (31), for each sample we may obtain the probability integral of r, given n, for $\rho = \rho_0$. For the tth sample ($t = 1, 2, \ldots k$), define π_t as follows:

$$\left.\begin{aligned} \pi_t &= 2\int_{-1}^{r_t} p\,(r \mid n = n_t,\ \rho = \rho_0)\,dr, \text{ if } r_t \leqslant \text{median value of } r^* \\ \pi_t &= 2\int_{r_t}^{+1} p\,(r \mid n = n_t,\ \rho = \rho_0)\,dr, \text{ if } r_t > \text{median value of } r \end{aligned}\right\}. \qquad \ldots\ldots(32)$$

Hence $0 \leqslant \pi_t \leqslant 1$. Further, if the hypothesis tested is true, the k values of π_t will be independent and each equally likely to assume any value between 0 and 1. We may now follow the suggestion of R. A. Fisher (16) and Karl Pearson (17) and take

$$P = \pi_1 \pi_2 \ldots \pi_k \qquad \ldots\ldots(33)$$

as a criterion to test the hypothesis (31). It will be seen that P has a maximum value of unity and tends to zero as the values of r_t diverge from ρ_0.

As Fisher has stated, and as follows also from Karl Pearson's work, if the hypothesis tested be true

$$\chi^2 = -2\log_e(P) \qquad \ldots\ldots(34)$$

is distributed as in the standard χ^2 distribution

$$p\,(\chi^2) = \frac{1}{2^{f/2}\,\Gamma\left(\frac{f}{2}\right)}(\chi^2)^{f/2-1}\,e^{-\frac{1}{2}\chi^2} \qquad \ldots\ldots(35)$$

with degrees of freedom, f, equal to $2k$. Thus a simple test of the hypothesis (31) is available, provided that the values of the π_t's can be calculated. The present tables and charts make such calculations possible.

In considering alternative (*b*), we see that the hypothesis tested is that $\rho_t \leqslant \rho_0$, and the admissible alternative hypotheses will assume that there is at least one population, though which one cannot be specified, for which $\rho_t > \rho_0$. In this case we define π_t as

$$\pi_t = \int_{r_t}^{+1} p\,(r \mid n = n_t,\ \rho = \rho_0)\,dr, \qquad \ldots\ldots(36)$$

so that $0 \leqslant \pi_t \leqslant 1$ ($t = 1, 2, \ldots k$), and if the hypothesis tested is true the π_t's will have the same properties as before. Thus the criterion P will be as in (33) and will be related by the transformation (34) to the χ^2 distribution with $2k$ degrees of freedom. It will be noted that P will now approach zero as the differences $r_t - \rho_0$ increase, i.e. as it becomes less likely that the hypothesis $\rho_t \leqslant \rho_0$ is true.

Karl Pearson suggested a slightly different procedure to test the significance of (33), the hypothesis of alternative (*b*). He wrote

$$P\{\lambda \geqslant \lambda_n\} = P_{\lambda_n} = I\left(n-1,\ -\frac{\log_{10}\lambda_n}{\sqrt{n}\log_{10}e}\right), \qquad \ldots\ldots(37)$$

where $I\,(p,\,u)$ is the function given in *Tables of the Incomplete Gamma-Function*, λ_n is the criterion (33) and n is the number of samples which are tested. Tables of P_{λ_n} were calculated for different values of $-\log_{10}\lambda_n$ and n, and may therefore be used to test hypothesis (31). This process is, however, essentially the same as that proposed by R. A. Fisher. Since

$$1 - P\{\chi^2 > \chi_0^2\} = 1 - P_{\chi^2}\ (\text{say}) = I\left(\tfrac{1}{2}(N-3),\ \frac{\frac{1}{2}\chi_0^2}{\sqrt{\frac{1}{2}(N-1)}}\right), \qquad \ldots\ldots(38)$$

if we calculate P_{χ^2} for $N = 2n+1$ and $\chi^2 = -\dfrac{2\log_{10}\lambda_n}{\log_{10}e}$, then from (37) and (38)

$$P\{\lambda \geqslant \lambda_n\} = 1 - P_{\chi^2}. \qquad \ldots\ldots(39)$$

* For very small samples median r will of course differ somewhat from ρ_0.

It should be understood that these two tests are precise, in the sense that the sampling distribution of the criterion used is known if H_0, the hypothesis tested, be true, so that the risk of rejecting H_0 when true is exactly controlled. How far such tests are biased, in the sense of Neyman and Pearson's terminology, or whether more powerful tests could be found which would be more likely to detect departures of the ρ's from ρ_0, are theoretical problems requiring further investigation.

Example I. The following example is quoted by Tippett (18) from a paper by Tschepourkowsky (1905). The table contains the correlation coefficient between cephalic index and upper face form for samples of skulls belonging to thirteen races. It is desired to test the hypothesis that there is no association between cephalic index and upper face form. Accordingly, in our terminology we shall test the hypothesis that

$$\rho_1 = \rho_2 = \ldots = \rho_{13} = 0, \qquad \ldots\ldots(40)$$

where alternatively the ρ's have any other values between -1 and $+1$.

Table II

Race	Number of skulls measured	Correlation coefficient	$\int_{-1}^{r_t} p(r \mid n, \rho = 0)$	π_t
Australians	66	+·089	·761	·478
Negroes	77	+·182	·946	·108
Duke of York Islanders	53	−·093	·255	·510
Malays	60	−·185	·079	·158
Fijians	32	+·217	·883	·334
Papuans	39	−·255	·060	·120
Polynesians	44	+·002	·505	·990
Alfourous	19	−·302	·104⁵	·209
Micronesians	32	−·251	·083	·166
Copts	34	−·147	·203	·406
Etruscans	47	−·021	·445	·890
Europeans	80	−·198	·039	·078
Ancient Thebans	152	−·067	·207⁵	·415

The values of π_t were calculated directly from the *Tables of the Incomplete Beta-Function*, by means of the relations expressed in (11) and (12). Similar results would have been obtained by application of the Lagrangian interpolation formula to the tables of r.

Here we see that

$$\log P = \log \prod_{t=1}^{13} \pi_t = -7{\cdot}174{,}045$$

$$\chi^2 = -2\log_e P = 33{\cdot}0376, \qquad f = 2k = 26,$$

$$f = 26, \qquad \chi^2 = 35{\cdot}563, \; P_{\chi^2} = {\cdot}10,$$

$$f = 26, \qquad \chi^2 = 31{\cdot}795, \; P_{\chi^2} = {\cdot}20,$$

and hence the probability of getting a larger χ^2 than the one we have obtained will lie between ·1 and ·2.

Tippett approaches the example in another way. Using equations (6) and (8) of Fisher's z'-transformation he finds the quantity

$$\chi^2 = \sum_{t=1}^{13} (n_t - 3)(z'_t)^2 = 17{\cdot}26,$$

and refers this to the χ^2 tables with degrees of freedom $f = k = 13$. The tables give

$$f = 13, \qquad \chi^2 = 19{\cdot}812, \; P_{\chi^2} = {\cdot}10,$$

$$f = 13, \qquad \chi^2 = 16{\cdot}985, \; P_{\chi^2} = {\cdot}20.$$

Since here he is testing the hypothesis $\rho = 0$, the error introduced by the transformation is very slight. It is seen that the result is comparable with the previous one obtained by using the probability integral of r.

The results of either of these tests would therefore give us no clear grounds for rejecting the hypothesis that there is no association between cephalic index and upper face form.

Example II. Matuszewski and Supinska (19) carried out a series of experiments with streptococcus. They measured on samples of different types of bacteria the rate of increase per hour of the number of bacteria, and the amount of acid in 10^{-10} mg. produced by one cell in one hour. The results of their experiments were given in tabular form, and rough plotting seemed to suggest that the assumption of normality would be justifiable. The present writer worked out a series of correlation coefficients, which are given below in Table III.

We shall test the hypothesis* that

$$\rho_1 = \rho_2 = \rho_3 = \dots = \rho_{11} = 0, \qquad \text{......(41)}$$

i.e. that there is no association between rate of increase of the bacteria and the amount of acid produced by one cell, the alternative hypotheses being that the ρ's may have any other values between $+1$ and -1. Using the tables of r we obtain $\int_{-1}^{r} p(r \mid n, \rho)\,dr$ for each sample and hence π_t, using (32).

Table III

Type of bacteria		Number of experiments with the same culture	Correlations between rate of increase and amount of acid	$\int_{-1}^{r} p(r)\,dr$	π_t	$\log_{10} \pi_t$
Streptococcus Lactis	3	6	$-\cdot3945^-$	$\cdot2195^-$	$\cdot4389$	$\bar{1}\cdot64237$
,, ,,	4	6	$+\cdot6268$	$\cdot9085^+$	$\cdot1830$	$\bar{1}\cdot26245$
,, ,,	5	7	$+\cdot8276$	$\cdot9892$	$\cdot0215^+$	$\bar{2}\cdot33244$
Streptococcus Cremoris	6	6	$-\cdot1973$	$\cdot3539$	$\cdot7078$	$\bar{1}\cdot84991$
Streptococcus Lactis	7	6	$+\cdot5015^+$	$\cdot8446$	$\cdot3108$	$\bar{1}\cdot49248$
,, ,,	8	6	$-\cdot4498$	$\cdot1854$	$\cdot3708$	$\bar{1}\cdot56914$
,, ,,	9	5	$-\cdot0878$	$\cdot5557$	$\cdot8886$	$\bar{1}\cdot94871$
,, ,,	10	5	$-\cdot6396$	$\cdot1226$	$\cdot2452$	$\bar{1}\cdot38952$
,, ,,	11	6	$-\cdot0167$	$\cdot4875^-$	$\cdot9749$	$\bar{1}\cdot98896$
,, ,,	12	5	$-\cdot3717$	$\cdot2690$	$\cdot5379$	$\bar{1}\cdot73070$
,, ,,	13	5	$-\cdot4953$	$\cdot1981$	$\cdot3961$	$\bar{1}\cdot59780$

$$\chi^2 = -2\log_e P = 23\cdot926, \qquad f = 2k = 22,$$

$$P_{\chi^2} = \cdot35.$$

An alternative method of testing hypothesis (41) is that used by Tippett in the previous example. We shall note further on in the text that a rough approximation to equation (8) is to assume that z' is approximately normally distributed with standard deviation equal to $1/\sqrt{n-3}$. The quantity $z'\sqrt{n-3}$ will therefore be normally distributed with unit standard deviation in populations where ρ is zero. Hence if we have k samples and consider the expression

$$\chi^2 = \sum_{t=1}^{k} (n_t - 3) z_t'^2,$$

we see that this will be distributed as χ^2 with k degrees of freedom. The alternative method by which we proceed is therefore clear. Converting each r_t of Table III to z_t' by means of the relation (6), we finally obtain

$$\chi^2 = \sum_{t=1}^{11} (n_t - 3) z_t'^2 = 11\cdot5416, \qquad f = k = 11,$$

$$P_{\chi^2} = \cdot40.$$

* It would, of course, be wrong to consider that the truth of any hypothesis tested is proved or disproved when it is based on such scanty data. The only inference we may draw would be that the hypothesis may be true, or alternatively, may be false, but that further experimentation would be necessary to confirm it.

This result is comparable with that of the other method. In either method we therefore find no reason to reject the hypothesis that there is no association between the rate of increase of the bacteria and the amount of acid produced by one cell in one hour.

Question 3 (*ii*)

This type of problem differs from the previous one in that now ρ_0 is not specified and the hypothesis to be tested is that

$$\rho_1 = \rho_2 = \ldots = \rho_k. \qquad \ldots\ldots(42)$$

In such cases no exact test is known, but two lines of procedure are possible:

(*a*) If k is large enough, we may obtain from the sample correlation coefficients, r_t, some form of weighted estimate, say r_0, of the unknown hypothetical common ρ_0, and using this for ρ_0 apply the same methods as were described in dealing with Question 3 (i).

In a recent paper Karl Pearson (17) suggested using as r_0 an approximation to the maximum likelihood estimate of the unknown ρ_0. The method of approximation was as follows: He calculated the weighted mean of the first four powers of the sample correlation coefficients, i.e.

$$\mu_v = \frac{\sum_{t=1}^{n} n_t r_t^v}{N}, \quad \text{where} \quad N = \sum_{t=1}^{n} n_t, \qquad \ldots\ldots(43)$$

and substituted them in the following equations:

$$\left.\begin{aligned} \rho_1 &= \mu_1 \\ \rho_2 &= \mu_1 + \rho_1(\mu_2 - \rho_1^2) \\ \rho_3 &= \mu_1 + \rho_2(\mu_2 - \rho_2^2) + \rho_1^2(\mu_3 - \rho_1^3) \\ \rho_4 &= \mu_1 + \rho_3(\mu_2 - \rho_3^2) + \rho_2^2(\mu_3 - \rho_2^3) + \rho_1^3(\mu_4 - \rho_1^4) \end{aligned}\right\}. \qquad \ldots\ldots(44)$$

ρ_4 was his final approximation to the common ρ. The process could be extended to ρ_5 and ρ_6 and so on, but moments higher than μ_4 would then have to be calculated. This procedure has been tried on several examples and found to give quite a reasonable value for ρ.*

Since r_0 is a function of $r_1, r_2, \ldots r_k$, the expressions π_t of (32) or (36) will not now be independent, and consequently the test based on the transformation

$$\chi^2 = -2\log_e(P)$$

will no longer be accurate, in the sense that in repeated sampling this χ^2 would not follow the distribution (35).

(*b*) As an alternative R. A. Fisher's z'-transformation may be used. If it may be assumed that when (42) is true, z'_t is distributed normally about a common but unknown mean for all values of t, with standard deviation $1/\sqrt{n_t - 3}$,† then

$$\chi^2 = \sum_{t=1}^{k} (n_t - 3)(z'_t - \bar{z}')^2, \qquad \ldots\ldots(45)$$

where

$$\bar{z}' = \frac{\sum_{t=1}^{k} (n_t - 3) z'_t}{\sum_{t=1}^{k} (n_t - 3)}, \qquad \ldots\ldots(46)$$

* It is probable that this procedure was in Karl Pearson's mind when he wrote that the present tables would "largely assist the investigator to determine whether a series of correlation coefficients of samples may be assumed to have a common origin". See also *Biometrika*, xxv, 395.

† This approximation for the standard deviation is derived from equation (8). If we neglect the terms containing ρ and higher powers of ρ we get

$$\sigma_{z'}^2 = \frac{3n^2 + 8}{3(n-1)^2},$$

which is approximately equal to $1/(n-3)$.

will be distributed as χ^2 with degrees of freedom $f = k-1$. Unless all the n_t are equal, it will be seen from (7) that the expected values of the z_t' are not precisely the same, so that an approximation is involved in this test also.

Example I. The following example is taken from a recent paper by E. S. Pearson and S. S. Wilks (20).

Table IV. *Racial Correlation Coefficients for equal small samples of* 20 *taken from* 30 *races*

r_t	$\int_{-1}^{r_t} p(r\mid n,\rho)\,dr$	π_t	$\log \pi_t$	r_t	$\int_{-1}^{r_t} p(r\mid n,\rho)\,dr$	π_t	$\log \pi_t$	r_t	$\int_{-1}^{r_t} p(r\mid n,\rho)\,dr$	π_t	$\log \pi_t$
+·097	·2085	·4170	$\bar{1}$·62014	+·219	·3853	·7706	$\bar{1}$·58580	+·178	·3100	·6200	$\bar{1}$·79239
+·198	·3507	·7014	$\bar{1}$·84597	−·152	·0331	·0662	$\bar{2}$·82086	+·763	·9979	·0042	$\bar{3}$·62325
+·576	·9418	·1164	$\bar{1}$·06595	+·319	·5637	·8726	$\bar{1}$·94082	+·101	·2127	·4254	$\bar{1}$·62880
−·015	·1008	·2016	$\bar{1}$·30449	+·310	·5473	·9054	$\bar{1}$·95684	+·449	·7877	·4246	$\bar{1}$·62798
+·173	·3115	·6230	$\bar{1}$·79449	+·019	·1277	·2554	$\bar{1}$·40722	+·245	·4300	·8600	$\bar{1}$·93450
+·764	·9980	·0040	$\bar{3}$·60206	+·445	·7816	·4368	$\bar{1}$·64028	+·360	·6385	·7230	$\bar{1}$·85914
−·037	·0858	·1716	$\bar{1}$·23452	+·410	·7256	·5488	$\bar{1}$·73941	+·592	·9460	·1080	$\bar{1}$·03342
+·667	·9823	·0354	$\bar{2}$·54900	[+·946	1·0000	·0000	$<\bar{6}$]*	−·515	·0003	·0006	$\bar{4}$·77815
+·014	·1234	·2468	$\bar{1}$·39235	+·018	·1268	·2536	$\bar{1}$·40415	+·023	·1311	·2622	$\bar{1}$·41863
−·112	·0472	·0944	$\bar{2}$·97497	+·160	·2921	·5842	$\bar{1}$·76656	+·259	·4458	·8916	$\bar{1}$·95017

Samples of 20 skulls are randomly drawn from each of 30 different races, and the correlation between head length and head breadth is calculated for each sample. The question which we may ask is: Are the data consistent with the hypothesis

$$\rho_1 = \rho_2 = \ldots = \rho_{30} = \rho, \qquad \ldots\ldots(47)$$

where ρ is not specified?

We assume that if the hypothesis (47) is not true the ρ's may assume any different and unequal values whatever. Since ρ is not specified it is necessary to obtain an estimate of the common correlation coefficient, r_0, from the data. Various methods may be devised. Here Karl Pearson's maximum likelihood estimate is used, and to this end equations (44) are employed. Successive approximations give

$$\rho_1 = \cdot 2490, \quad \rho_2 = \cdot 2726, \quad \rho_3 = \cdot 2761, \quad \rho_4 = \cdot 2774.$$

These results seem to suggest that we may well take $r_0 = \cdot 277$. Using the table for $n = 20$ we obtain by interpolation the π_t's of equation (32),

$$\log_{10} P = -20\cdot 7077,^*$$

and hence from (34)

$$\chi^2 = -2\log_e(P) = 95\cdot 362, \qquad f = 58,^\dagger$$

$$P_{\chi^2} < \cdot 0001.$$

We therefore reject the hypothesis (47) and decide that it is most unlikely that each of the 30 races have the same correlation between head length and head breadth.

E. S. Pearson and S. S. Wilks also decided to reject (47) but they employed a different procedure. Using (46), they found

$$\bar{z}' = \cdot 24913.$$

Hence using (45)

$$\chi^2 = \sum_{t=1}^{30} (n_t - 3)(z_t' - \bar{z}')^2 = 96\cdot 01, \qquad f = k - 1 = 29,$$

$$P_{\chi^2} < \cdot 000{,}030.$$

We see that using either procedure we should reject hypothesis (47).

* The value $+\cdot 946$ was used to obtain the maximum likelihood estimate of ρ, but was omitted from the calculations which follow, since the purpose of the example is purely illustrative. It is clear that had this coefficient been included, the effect would have been to increase the value of χ^2, and hence make the hypothesis (47) even more improbable.

† It is not obvious what the exact degrees of freedom will be in this case. We know that if we are testing the hypothesis $\rho = \rho_0$, where ρ_0 is some fixed value, the degrees of freedom will be $2k$. Without further theoretical work it is not possible to say what the effect will be on the number of degrees of freedom of calculating the weighted means of the first four powers of the r's. The writer is aware that $2k$ is not correct in this case but offers it as an approximation until the problem is solved correctly.

Karl Pearson, using his P_{λ_n} test, found that in only 3·7 per cent. of cases would a more improbable result than the one he reached be obtained, and he therefore decided also to reject hypothesis (47). It is of interest to note why his result differs so greatly from those obtained by the two previous methods.

Pearson defines his π_t as

$$\pi_t = \int_{-1}^{r_t} p(r \mid n, \rho)\, dr,$$

and his criterion is

$$P = \prod_{t=1}^{k} \pi_t.$$

If we consider the class of admissible alternative hypotheses we see that here Pearson is really testing the hypothesis $\rho \geqslant r_0 = \cdot 277$ with the alternative that at least one $\rho_t < \cdot 277$. This is equivalent to the alternative (*b*) discussed under Question 3 (i), while the two previous methods are equivalent to the alternative (*c*), where we assume that the ρ_t's may take any values between minus unity and plus unity. The discrepancy between the results of Karl Pearson and of E. S. Pearson and S. S. Wilks is therefore not important, because we see that they are really using methods designed to test different hypotheses.

In the case of two samples of size n_1 and n_2, respectively, it would be quite unjustifiable to attempt to estimate a common ρ for the two populations, and to carry out the first procedure suggested on p. xxiii above, i.e. by calculating a π_1 and π_2 based on the estimate of ρ. The second procedure involving the use of the z'-transformation may however still be used. Since the degrees of freedom for χ^2 are now $k-1 = 1$, the second method of procedure reduces to the following:

Calculate

$$z'_t = \tfrac{1}{2} \log \frac{1+r_t}{1-r_t}$$

for $t = 1, 2$, and test the hypothesis

$$\rho_1 = \rho_2 \qquad \text{......(48)}$$

by finding the ratio

$$\frac{z'_1 - z'_2}{\sqrt{\dfrac{1}{n_1-3} + \dfrac{1}{n_2-3}}} \qquad \text{......(49)}$$

and referring this to the normal probability scale. When we are testing hypothesis (48) we assume the admissible alternative hypotheses will be that $\rho_1 > \rho_2$ or $\rho_1 < \rho_2$, and hence we should consider both tail areas of the normal curve. If, on the other hand, we are testing the hypothesis

$$\rho_1 \geqslant \rho_2 \qquad \text{......(50)}$$

the admissible alternative hypothesis will be that $\rho_1 < \rho_2$, and we should therefore only concern ourselves with one tail area.

E. S. Pearson has suggested the following rough test, when dealing with small samples, involving the use of Chart I, without the need for any transformation of variables. The rule to be adopted is as follows:

Using the observed values of r_1 and r_2 read off from the appropriate curves for n_1 and n_2 in Chart I the quantities ρ_{a_1} and ρ_{b_1}, ρ_{a_2} and ρ_{b_2} as shown in Fig. 7 below. Here we must distinguish between the two hypotheses. If we are testing hypothesis (48) then the rule will be: Reject the hypothesis, $\rho_1 = \rho_2$, if

$$\rho_{b_2} > \rho_{a_1} \quad \text{or} \quad \rho_{b_1} > \rho_{a_2}. \qquad \text{......(51)}$$

The risk of rejecting hypothesis (48) when it is true will be approximately ·02, provided n_1 and n_2 are not too different. If we are testing hypothesis (50) then the rule will be: Reject the hypothesis, $\rho_1 \geqslant \rho_2$, if

$$\rho_{b_2} > \rho_{a_1}. \qquad \text{......(52)}$$

The first kind of error, i.e. the risk of rejecting hypothesis (50) when it is true, will be ·01. The basis of this rule will be discussed later.

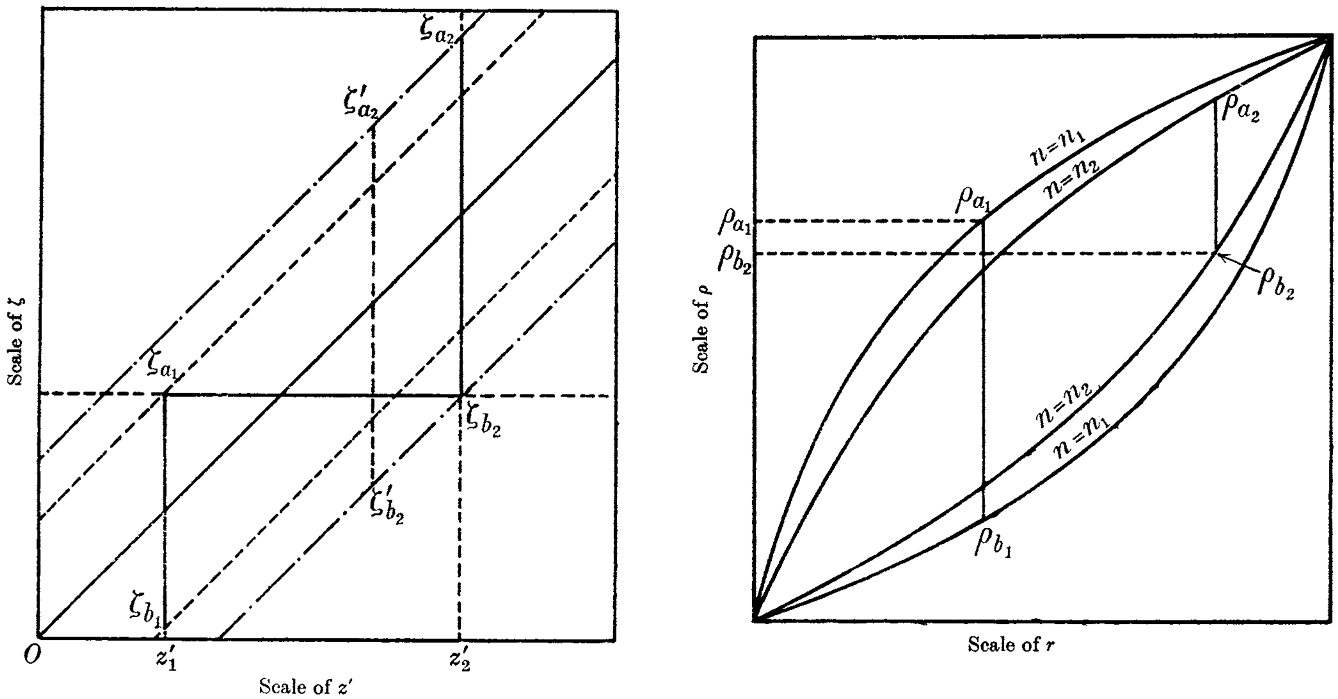

Fig. 7.

Example I. Suppose we consider two of the types of bacteria, i.e. Streptococcus Lactis No. 5 and Streptococcus Cremoris No. 6, given in the example on p. xxv. We may ask the question: Could the correlation between k, the increase per hour of the number of bacteria, and the amount of acid, b, produced by one cell in one hour, have a common value for both populations of bacteria?

Table V

Type of bacteria	Number of experiments with the same culture	Correlation between k and b
Streptococcus Lactis 5	7	$r_1 = +\cdot 8276$
Streptococcus Cremoris 6	6	$r_2 = -\cdot 1973$

We are testing the hypothesis

$$\rho_1 = \rho_2,$$

and our alternative hypotheses will be that $\rho_1 > \rho_2$ or $\rho_1 < \rho_2$. Turning to Chart I we see that $\rho_{a_1} = +\cdot 95$, $\rho_{b_1} = +\cdot 33$; $\rho_{a_2} = +\cdot 56$, $\rho_{b_2} = -\cdot 755$, giving

$$\rho_{b_1} < \rho_{a_2}.$$

There does not therefore seem to be any ground for rejecting the hypothesis that the correlation between the two variables is the same for both types of bacteria.

This is confirmed by an application of the z'-test, which gives the probability of a larger difference between the r's to be ·071. It would, of course, be ridiculous to accept the hypothesis as proved on such scanty data. The results of our analysis would lead us to accept the hypothesis as proved only if it was confirmed by further experimentation.

Example II. A series of measurements on skulls of different Swiss races are given in "Les Crânes Valaisans".*

* *Crania Helvetica*, I. "Les Crânes Valaisans de la Vallée du Rhône", par Eugene Pittard.

Random samples of 10 were picked out from the Biel and Sierre series, and the coefficient of correlation between maximum length and maximum breadth of the skulls was worked out:

$$\text{Correlation for the Biel series} = r_1 = +{\cdot}777,$$

$$\text{Correlation for the Sierre series} = r_2 = -{\cdot}352.$$

We ask the question: Is it possible that the correlation between maximum length and breadth of the skull can be the same for both series? We have no *a priori* knowledge which might lead us to test the hypothesis $\rho_1 \geqslant \rho_2$, and we must therefore test the hypothesis

$$\rho_1 = \rho_2,$$

and assume that if the hypothesis tested is not true, then ρ_1 may be $> \rho_2$ or $< \rho_2$.

Chart I shows us that the lines $r = r_1$ and $r = r_2$ cut the pair of curves $n = 10$ at the points

$$({\cdot}777, {\cdot}92)\ ({\cdot}777, {\cdot}38) \quad \text{and} \quad (-{\cdot}352, +{\cdot}23)\ (-{\cdot}352, -{\cdot}73).$$

We see that

$$\rho_{b_1} > \rho_{a_2}.$$

We should therefore decide to reject the hypothesis $\rho_1 = \rho_2$. The z'-test gives the probability of a greater difference between the r's to be ·0086. This confirms the result of our approximate test.

Theoretical Basis of Rule given

The position may be understood most clearly by comparing in Fig. 7 the confidence belt for (r, ρ) and that for (z', ζ). For rough purposes it may be supposed that z' is normally distributed about ζ with standard deviation $1/\sqrt{n-3}$. Thus the confidence belt for a given n is bounded by the two parallel lines

$$z' = \zeta \pm \psi_0/\sqrt{n-3}. \qquad \text{......(53)}$$

The values of ψ_0 for the (z', ζ) charts corresponding to Charts I, II, III and IV, respectively, would have the values shown in the table below.

Table VI

Chart:	I	II	III	IV
ψ_0	1·645	1·960	2·326	2·576
ψ_0'	2·326	2·772	3·289	3·693
Risk of error in using rule	·02	·006	·001	·0002

Corresponding to the observed r_1 and r_2 we have z_1' and z_2' and find from the belt ζ_{a_1}, ζ_{b_1} and ζ_{a_2}, ζ_{b_2}. Suppose we now follow the rule:

Reject the hypothesis (48) that $\zeta_1 = \zeta_2$, if either

$$\zeta_{b_2} > \zeta_{a_1} \quad \text{or} \quad \zeta_{b_1} > \zeta_{a_2}. \qquad \text{......(54)}$$

This is equivalent to the rule (51) expressed above in terms of the ρ's. It is also seen from the diagram to be equivalent to rejecting the hypothesis tested if

$$|z_2' - z_1'| > \psi_0/\sqrt{n_1 - 3} + \psi_0/\sqrt{n_2 - 3}. \qquad \text{......(55)}$$

But provided n_1 and n_2 be of reasonable size and not too different

$$\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}} \simeq \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{n_1 - 3}} + \frac{1}{\sqrt{n_2 - 3}}\right). \qquad \text{......(56)}$$