

August, 1965

SOME NEW RESULTS ON THE DISTRIBUTION OF THE  
SAMPLE CORRELATION COEFFICIENT

Harold Ruben

Technical Report No. 59

# SOME NEW RESULTS ON THE DISTRIBUTION OF THE SAMPLE CORRELATION COEFFICIENT\*

BY HAROLD RUBEN

University of Minnesota

## 1. INTRODUCTION AND SUMMARY

The correlation coefficient  $r$  in samples of size  $n > 2$  from a non-singular bivariate normal population with correlation coefficient  $\rho$  can be represented in the form (using an obvious generic notation)

$$\tilde{r} = \frac{\xi + \tilde{\rho} \chi_{n-1}}{\chi_{n-2}},$$

where  $\tilde{r} = r/(1-r^2)^{\frac{1}{2}}$ ,  $\tilde{\rho} = \rho/(1-\rho^2)^{\frac{1}{2}}$ ,  $\xi$  is a standardized normal variate and  $\xi$ ,  $\chi_{n-1}$ ,  $\chi_{n-2}$  are independent. (An equality sign between two variates, here and subsequently, is to be interpreted as meaning that the variates have a common distribution.) Approximation of the  $\chi$ -variates in the representation, using Fisher's normalization of  $\chi^2$ , results in the approximation that

$$\frac{(n - \frac{5}{2})^{\frac{1}{2}} \tilde{r} - (n - \frac{3}{2})^{\frac{1}{2}} \tilde{\rho}}{(1 + \frac{1}{2} (\tilde{r}^2 + \tilde{\rho}^2))^{\frac{1}{2}}}$$

is distributed as a standardized normal variate. High accuracy of the latter approximation is indicated by theoretical considerations and corroborated by some computations. These suggest that the approximation is generally more precise than that based on the Fisherian  $z$  (the inverse hyperbolic tangent) transform, corrected for bias or otherwise, and is about as precise as Hotelling's refinement of  $z$ . Further, when  $n$  is sufficiently large for  $r$  to be approximated effectively by a normal variate, it is far superior to that approximation.

The above representation of  $r$  has independent value and interest in that it throws some light on various earlier results, and can also be used to yield additional

---

\*This research was sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research under Contract AF-AFOSR-885-65.

approximations and expansion for the probability integral of  $r$ , such as a high-accuracy Cornish-Fisher expansion and an approximation, valid for high  $n$  and  $\rho$ , based on  $F$  with  $n - 1$  and  $n - 2$  degrees of freedom. Finally, we may note that the representation provides an exceedingly simple derivation<sup>†</sup> of the exact distribution of  $r$  in normal samples.

## 2. A SIMPLE REPRESENTATION OF $r$

Denote the means, standard deviations and correlation coefficient of an arbitrary (not necessarily normal) non-singular, bivariate population by  $\mu_x, \mu_y, \sigma_x, \sigma_y, \rho$ .

Let

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2\}^{\frac{1}{2}}} \quad (1)$$

be the sample correlation coefficient based on a random sample  $((x_1, y_1), \dots, (x_n, y_n))$  of size  $n > 2$  ( $\sum \equiv \sum_1^n, \bar{x} = \sum x_i/n, \bar{y} = \sum y_i/n$ ). Defining the transformed correlations  $\tilde{\rho}, \tilde{r}$  by

$$\tilde{\rho} = \rho/(1-\rho^2)^{\frac{1}{2}}, \quad \tilde{r} = r/(1-r^2)^{\frac{1}{2}} \quad (2)$$

and the variates  $u, v, w$  by

$$u = \frac{b - \beta}{\sigma_y(1-\rho^2)^{\frac{1}{2}}/\{\sum(x_i - \bar{x})^2\}^{\frac{1}{2}}}, \quad (3)$$

$$v = \left( \frac{\sum(x_i - \bar{x})^2}{\sigma_x^2} \right)^{\frac{1}{2}}, \quad (4)$$

$$w = \frac{[\sum(y_i - \bar{y} - b(x_i - \bar{x}))^2]^{\frac{1}{2}}}{\sigma_y(1-\rho^2)^{\frac{1}{2}}}, \quad (5)$$

<sup>†</sup> I have used essentially this derivation (which appears to be much simpler than previous derivations) in class over the past few years. This experience suggests that the derivation should not present any difficulty to mathematics students in their first year of statistics.

where  $\beta = \rho\sigma_y/\sigma_x$ ,  $b = \Sigma(x_i - \bar{x})(y_i - \bar{y})/\Sigma(x_i - \bar{x})^2$  ( $\beta$ ,  $b$  are the slopes of the mean-square linear regression of  $y$  on  $x$  in the population and sample, respectively), we have the purely algebraic result

$$r = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{[\Sigma(x_i - \bar{x})^2 \Sigma(y_i - \bar{y})^2 - (\Sigma(x_i - \bar{x})(y_i - \bar{y}))^2]^{\frac{1}{2}}} = \frac{u + \tilde{\rho} v}{w} \quad (6)$$

Defining further

$$\epsilon_i^* = \frac{y_i - \mu_y - \beta(x_i - \mu_x)}{\sigma_y(1-\rho^2)^{\frac{1}{2}}} \quad (7)$$

( $\epsilon_i^*$  is the residual of  $y_i$  from the population mean-square regression, standardized by the factor  $\sigma_y(1-\rho^2)^{\frac{1}{2}}$ ), we also have the decomposition

$$\Sigma \epsilon_i^{*2} = (\Sigma \epsilon_i^*/\sqrt{n})^2 + u^2 + w^2 \quad (8)$$

where it will be noted that

$$u = \sum \frac{(x_i - \bar{x})}{\{\Sigma(x_i - \bar{x})^2\}^{\frac{1}{2}}} \epsilon_i^* \quad (9)$$

(8) and (9) are again purely algebraic in character.

For a normal population,  $u$  is a standardized normal variate (call this  $\xi$ ),  $v$  is a  $\chi_{n-1}$ ,  $w$  is a  $\chi_{n-2}$ , and  $u, v, w$  are independent; that is, (6) becomes

$$\tilde{r} = \frac{\xi + \tilde{\rho} \chi_{n-1}}{\chi_{n-2}} \quad (10)$$

in which the three variates on the right of (10) are independent. (Conditionally on fixed  $x_1, \dots, x_n$ , the  $\epsilon_i^*$  are independent standardized normal variates, and so therefore are also an arbitrary set of  $\eta_i$  defined by an orthogonal transformation of the  $\epsilon_i^*$  with  $\eta_1 = \Sigma \epsilon_i^*/\sqrt{n}$ ,  $\eta_2 = \Sigma(x_i - \bar{x})\epsilon_i^*/\{\Sigma(x_i - \bar{x})^2\}^{\frac{1}{2}}$ . Consequently,

conditionally on fixed  $x_1, \dots, x_n$ ,  $u(= \eta_2)$  and  $w(= \sqrt{\frac{1}{3} \sum_{i=1}^n \xi_i^2} = \sqrt{\frac{1}{3} \sum_{i=1}^n \eta_i^2})$  are independent and distributed as a  $\mathcal{N}(0,1)$  and a  $\chi_{n-2}$ . The latter conditional distribution does not involve the  $x_i$ , and furthermore  $v$  is distributed as a  $\chi_{n-1}$ , thus justifying the assertion prior to (10).)

It is of interest to relate (10) to previous results.

(a) Distribution of  $r$  in normal samples. For  $\rho = 0$ , (10) reduces to

$$\frac{(n-2)^{\frac{1}{2}} r}{(1-r^2)^{\frac{1}{2}}} = \frac{\xi}{\chi_{n-2}/(n-2)^{\frac{1}{2}}},$$

or  $r$  is distributed as  $t$  with  $n-2$  degrees of freedom. For general  $\rho$ , the distribution follows almost as simply. Thus according to (6) and (10),  $\hat{r}$ , conditionally on fixed  $v$  and  $w$ , is normal with mean  $\tilde{\rho} v/w$  and standard deviation  $1/w$ , the conditional frequency function being, then,

$$w\varphi(\tilde{r}w - \tilde{\rho}v),$$

where 
$$\varphi(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2} x^2}$$

is the standardized normal density function, so that the (unconditional) frequency function of  $\hat{r}$  is

$$\int_0^\infty \int_0^\infty w\varphi(\tilde{r}w - \tilde{\rho}v) f_{n-2}(w) f_{n-1}(v) dv dw,$$

where  $f_v(x) = \left\{ 2^{\frac{1}{2}} (v-2) \Gamma\left(\frac{v}{2}\right) \right\}^{-1} e^{-\frac{1}{2} x^2} x^{v-1} \quad (x > 0)$

is the frequency of a  $\chi_v$ . Correspondingly, the frequency function of  $r$

$(d\hat{r}/dr = (1-r^2)^{-3/2})$  is

$$(1-r^2)^{-3/2} \int_0^\infty \int_0^\infty w\varphi(\tilde{r}w - \tilde{\rho}v) f_{n-2}(w) f_{n-1}(v) dv dw = \frac{(1-\rho^2)^{\frac{1}{2}(n-1)} (1-r^2)^{\frac{1}{2}(n-4)}}{(n-3)! \pi} \times \int_0^\infty \int_0^\infty (v'w')^{n-2} e^{-\frac{1}{2}[v'^2 - 2\rho r v'w' + w'^2]} dv' dw',$$

on using the substitutions  $v' = v/(1-\rho^2)^{\frac{1}{2}}$ ,  $w' = w/(1-r^2)^{\frac{1}{2}}$  and simplifying the normalizing constant by the duplication formula

$$2^{n-3} \Gamma\left(\frac{n-1}{2}\right) \Gamma\left(\frac{n+2}{2}\right) = \Gamma(n+2) \sqrt{\pi} .$$

Term by term integration after expansion of  $\exp(\rho r v' w')$  as a power series in  $(\rho r v' w')$  yields

$$\frac{2^{n-3} (1-\rho^2)^{\frac{1}{2}(n-1)} (1-r^2)^{\frac{1}{2}(n-4)}}{(n-3)! \pi} \sum_{j=0}^{\infty} \Gamma^2\left(\frac{n-1+j}{2}\right) \frac{(2\rho r)^j}{j!} \quad (11)$$

for the frequency function of  $r$  (see, e.g., Cramér, 1946, p. 398), a result useful in the evaluations of the moments of  $r$ . On the other hand, from the easily proved result†

$$\int_0^{\infty} \int_0^{\infty} e^{-\frac{1}{2}(v'^2 - 2\rho r v' w' + w'^2)} dv' dw' = \frac{\cos^{-1}(-\rho r)}{(1-\rho^2 r^2)^{\frac{1}{2}}} ,$$

the frequency function can be expressed alternatively as

$$\frac{(1-\rho^2)^{\frac{1}{2}(n-1)} (1-r^2)^{\frac{1}{2}(n-4)}}{(n-3)! \pi} \frac{\partial^{n-2}}{\partial(\rho r)^{n-2}} \left( \frac{\cos^{-1}(-\rho r)}{(1-\rho^2 r^2)^{\frac{1}{2}}} \right) , \quad (12)$$

which is Fisher's original form (1915).

(b) Probability of a positive correlation in normal samples. From (10),

$$\begin{aligned} P(r > 0) &= P(\xi + \hat{\rho} \chi_{n-1} > 0) \\ &= P\left(\frac{\xi}{\chi_{n-1}/(n-1)^{\frac{1}{2}}} > -(n-1)^{\frac{1}{2}} \hat{\rho}\right) \\ &= P(t_{n-1} > -(n-1)^{\frac{1}{2}} \rho) . \end{aligned} \quad (13)$$

---

E.g., by transforming  $v', w'$  to polar coordinates, or by a linear transformation which converts  $v'^2 - 2\rho r v' w' + w'^2$  to a sum of squares. The result is essentially Sheppard's celebrated formula (see, e.g., Kendall and Stuart, Vol. 1, 1961, p. 351) for the probability of the positive quadrant in a bivariate normal distribution.

Formula (13) has been given by Armsen† (1955) and Ruben (1963).

(c) Moments of  $\tilde{r}$  in normal samples. A general formula for the moments of  $\hat{r}$  has been given by Harley (1957), who also evaluated explicitly the first four moments. (These four moments had already been determined by Fisher in his 1915 paper.) Formula (10) makes the determination, in explicit form, of all moments of  $\hat{r}$  a trivial matter. Thus, for  $m < n - 2$ ,

$$\begin{aligned} E(\tilde{r}^m) &= E(\chi_{n-2}^{-m}) E(\xi + \tilde{\rho} \chi_{n-1})^m \\ &= E(\chi_{n-2}^{-m}) \sum_{j=0}^m \binom{m}{j} E(\xi^j) E(\chi_{n-1}^{m-j}) \tilde{\rho}^{m-j} . \end{aligned}$$

After simplification

$$E(\tilde{r}^m) = \frac{1}{\langle n-m-2 \rangle_{\frac{1}{2}(m+1)}} \sum_{k=0}^{\frac{1}{2}(m-1)} \langle 1 \rangle_k \langle n-2 \rangle_{\frac{1}{2}(m+1)-k} \binom{m}{2k} \tilde{\rho}^{m-2k} \quad (m=1,3,\dots), \quad (14a)$$

$$E(\tilde{r}^m) = \frac{1}{\langle n-m-2 \rangle_{\frac{1}{2}m}} \sum_{k=0}^{\frac{1}{2}m} \langle 1 \rangle_k \langle n-1 \rangle_{\frac{1}{2}m-k} \binom{m}{2k} \tilde{\rho}^{m-2k} \quad (m=2,4,\dots) \quad (14b)$$

$m$ th degree polynomials in  $\tilde{\rho}$ . Here  $\langle a \rangle_0 = 1$ ,  $\langle a \rangle_j = a(a+2)(a+4)\dots(a+2j-2)$  for  $j = 1, 2, \dots$ .

(d) Approximate identification of  $\tilde{r}$  with non-central  $t$ . With the aim of supplementing tables of non-central  $t$  by means of David's tables (1958) of the probability integral of  $r$ , Harley (1957) approximated  $(n-2)r/(1-r^2)^{\frac{1}{2}}$  to a multiple of a non-central  $t$ , the multiple and non-centrality parameter being determined in an ad hoc manner by moment considerations. Insight into the nature of the approximation can be obtained from (10), as follows. Approximate  $\chi_{n-1}$  by a normal variate with mean  $(n-3/2)^{\frac{1}{2}}$

† Dr. Linhart kindly informed me of Dr. Armsen's paper after publication of my own note.

THE UNIVERSITY OF CHICAGO

PHYSICS DEPARTMENT

PHYSICS 350

PHYSICS 350

PHYSICS 350



and variance  $1/2$  (Fisher's normalization of  $\chi^2$ ). The numerator of (10) is then distributed approximately as a normal variate with mean

$$\mu = (n - \frac{3}{2})^{\frac{1}{2}} \tilde{\rho}$$

and variance

$$\sigma^2 = 1 + \frac{1}{2} \tilde{\rho}^2 .$$

Accordingly,  $(n-2)^{\frac{1}{2}} \tilde{r}$  is approximated by

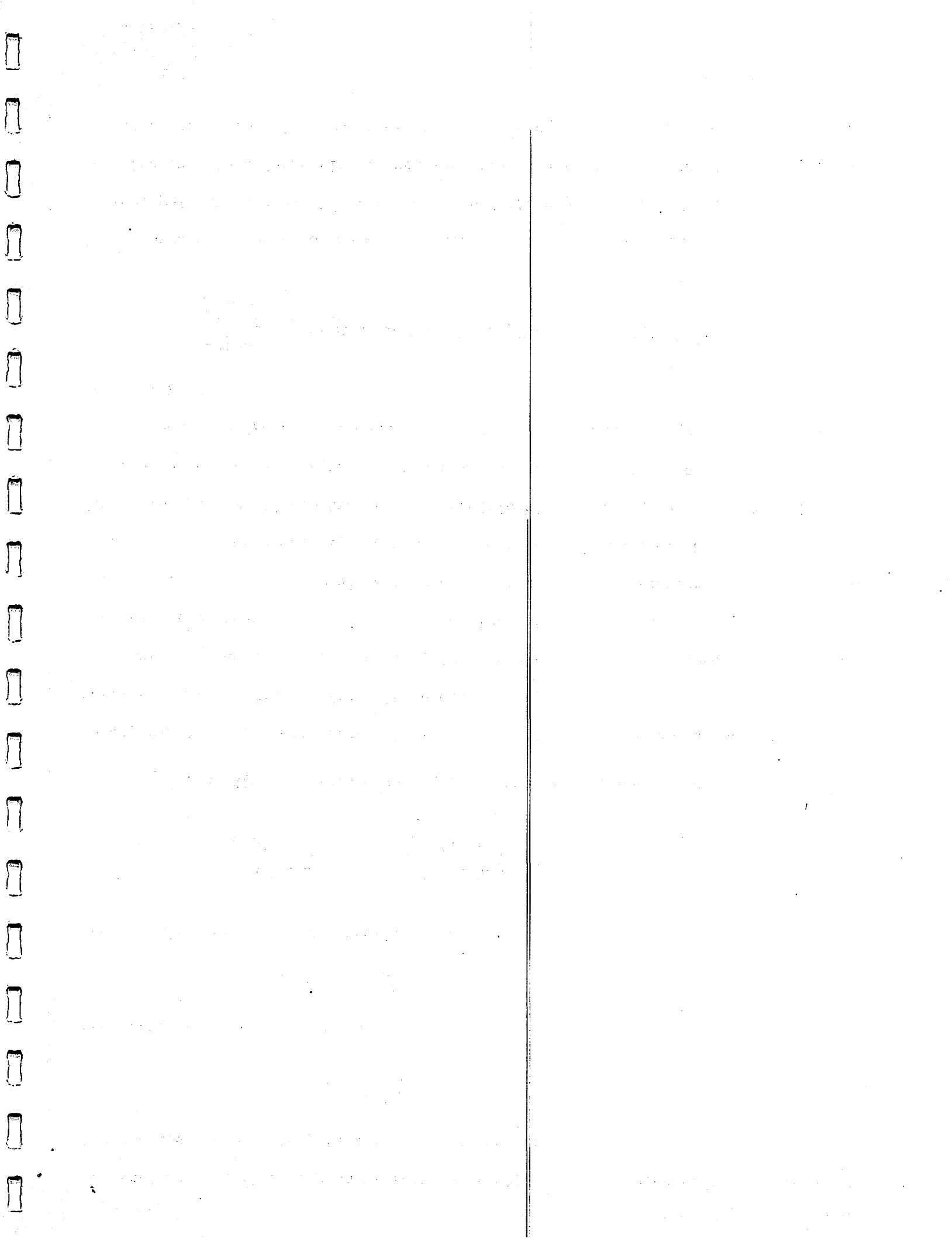
$$\frac{\mu + \sigma \eta}{\chi_{n-2}/(n-2)^{\frac{1}{2}}} = \sigma \frac{\eta + \mu/\sigma}{\chi_{n-2}/(n-2)^{\frac{1}{2}}} ,$$

where  $\eta$  is a  $N(0,1)$  variate and  $\eta, \chi_{n-2}$  are independent; that is,  $(n-2)^{\frac{1}{2}} \tilde{r}/\sigma$  is approximated by a non-central  $t$  with  $n-2$  degrees of freedom and non-centrality parameter  $\mu/\sigma$ . This is Harley's result.

Referring again to (6) and (10), it is clear that  $(n-2)^{\frac{1}{2}} \tilde{r}$ , the square root of the regression mean-square ratio in an analysis of variance of the  $y_i$  ( $x_i$  fixed) used in testing  $\beta = 0$ , can be regarded as a weighted non-central  $t$  (rather than, as above, a weighted normal), thereby providing an alternative, though less direct, derivation of the distribution of  $r$ . Conditionally on given  $v$ ,  $(n-2)^{\frac{1}{2}} \tilde{r}$  is a non-central  $t$  with  $n-2$  degrees of freedom and non-centrality parameter  $\tilde{\rho} \tilde{v}$ , the corresponding conditional frequency function of  $\tilde{r}$  being then (see, e.g., Wilks, 1962, p. 247)

$$\frac{e^{-\frac{1}{2} \tilde{\rho}^2 v^2}}{\Gamma(\frac{n-2}{2}) \sqrt{\pi}} \sum_{j=0}^{\infty} \Gamma\left(\frac{n-1+j}{2}\right) \frac{(\tilde{\rho} v \sqrt{2})^j}{j!} (\tilde{r}^2)^{\frac{1}{2}j} (1 + \tilde{r}^2)^{-\frac{1}{2}(n-1+j)} .$$

Averaging with respect to the distribution of  $v$  and multiplying by  $(1-r^2)^{-3/2}$  yields again the series (11) for the (unconditional) frequency function of  $r$ . It will be noted that this derivation is not essentially distinct from the familiar derivation of the distribution of the multiple correlation coefficient  $R$  in normal samples (e.g.,

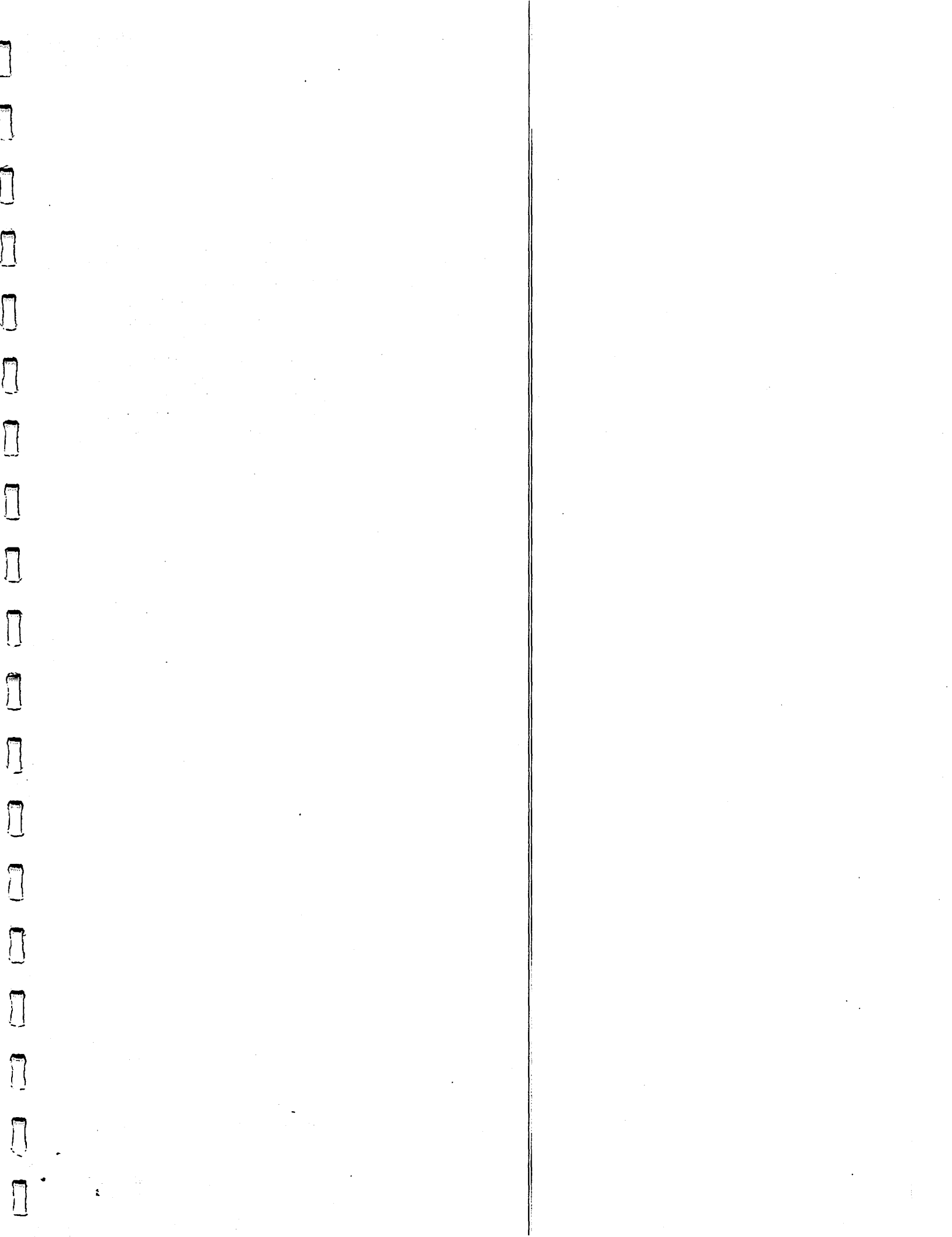


Anderson, 1958, pp. 92-94) consisting in weighting the conditional distribution of

the regression mean-square ratio  $\frac{R^2/(p-1)}{(1-R^2)/(n-p)}$  for given values of the p-1

'independent' variables with respect to the distribution of those variables.<sup>†</sup>

<sup>†</sup> The failure to carry this out for  $p = 2$  (except in the case  $\rho = 0$ , as in Anderson, 1958, pp. 61-64) provides an interesting historical reflection. At the time of appearance of Fisher's 1915 paper, when (in Hotelling's phrase) statisticians were enthusiastically exploring the universe with the correlation coefficient as their chief measuring instrument and the shift in emphasis from correlational to regression and analysis of variance concepts still lay well ahead in the future, 'regression' and conditional derivations (simple as these are) for the distribution of  $r$  could not be expected. However, there appears to be no reason (other than inertia) why such derivations should not be presented in textbooks and in class. On the other hand, the derivations do have an element of asymmetry which is absent in the corresponding derivations for  $p > 2$ .



### 3. A NEW APPROXIMATION FOR THE DISTRIBUTION OF $r$

From (10), for an arbitrary fixed value  $r_0$  ( $|r_0| < 1$ ) with  $\tilde{r}_0 = r_0 / (1 - r_0^2)^{\frac{1}{2}}$ ,

$$\begin{aligned} P(r \leq r_0) &= P(\tilde{r} \leq \tilde{r}_0) \\ &= P(\xi + \tilde{\rho} \chi_{n-1} - \tilde{r}_0 \chi_{n-2} \leq 0), \end{aligned} \quad (15)$$

the evaluation of the probability integral of  $r$  being then equivalent to the probability that the variate  $L$ , defined by

$$L = \xi + \tilde{\rho} \chi_{n-1} - \tilde{r}_0 \chi_{n-2} \quad (16)$$

is non-positive. (Cf. Kendall and Stuart, Vol. 1, 1961; p.299.) Since  $L$  is a linear combination of three independent variates, one of which ( $\xi$ ) is exactly normal and two ( $\chi_{n-1}$  and  $\chi_{n-2}$ ) are approximately normal, it is evident that  $L$  itself will be approximately normal, and one would further expect  $L$  to be more nearly normal than either of the two  $\chi$ -variates. This gains urgency from a consideration of the skewness and kurtosis of  $L$ .

The  $m$ th cumulant of  $L$ ,  $\kappa_m(L)$ , is related to the corresponding cumulants of  $\xi$ ,  $\chi_{n-1}$  and  $\chi_{n-2}$  by

$$\kappa_m(L) = \kappa_m(\xi) + \tilde{\rho}^m \kappa_m(\chi_{n-1}) + (-\tilde{r}_0)^m \kappa_m(\chi_{n-2}). \quad (17)$$

On using

$$\kappa_3(\chi_v) \sim \frac{1}{4v^{\frac{1}{2}}}, \quad \kappa_4(\chi_v) \sim \frac{3}{16v^2},$$

one obtains for the standardized third and fourth cumulants of  $L$ ,

$$\gamma_1(L) \equiv \frac{\kappa_3(L)}{\{\kappa_2(L)\}^{\frac{3}{2}}} = \frac{\tilde{\rho}^3 - \tilde{r}_0^3}{\{1 + \frac{1}{2}(\tilde{\rho}^2 + \tilde{r}_0^2)\}^{\frac{3}{2}}} \frac{1}{4n^{\frac{1}{2}}} + o\left(\frac{1}{n}\right),$$

$$\gamma_2(L) \equiv \frac{\kappa_4(L)}{\{\kappa_2(L)\}^2} = \frac{\tilde{\rho}^4 + \tilde{r}_0^4}{\{1 + \frac{1}{2}(\tilde{\rho}^2 + \tilde{r}_0^2)\}^2} \frac{3}{16n^2} + o\left(\frac{1}{n^2}\right).$$

1870  
1871  
1872  
1873  
1874  
1875  
1876  
1877  
1878  
1879  
1880  
1881  
1882  
1883  
1884  
1885  
1886  
1887  
1888  
1889  
1890  
1891  
1892  
1893  
1894  
1895  
1896  
1897  
1898  
1899  
1900

1901  
1902  
1903  
1904  
1905  
1906  
1907  
1908  
1909  
1910  
1911  
1912  
1913  
1914  
1915  
1916  
1917  
1918  
1919  
1920  
1921  
1922  
1923  
1924  
1925  
1926  
1927  
1928  
1929  
1930

Asymptotically, then

$$|r_1(L)| \leq 2^{\frac{3}{2}} \frac{|\tilde{\rho}|^3 + |\tilde{r}_0|^3}{(|\tilde{\rho}|^2 + |\tilde{r}_0|^2)^{\frac{3}{2}}} \frac{1}{4n^{\frac{1}{2}}} \leq \frac{1}{(2n)^{\frac{1}{2}}} = r_1(\chi_n), \quad (18)$$

$$r_2(L) \leq 2^2 \frac{|\tilde{\rho}|^4 + |\tilde{r}_0|^4}{(|\tilde{\rho}|^2 + |\tilde{r}_0|^2)^2} \frac{3}{16n^2} \leq \frac{3}{4n^2} = r_2(\chi_n). \quad (19)$$

In view of (18) and (19), we shall approximate  $L$  by a normal variate. By Fisher's normalization of  $\chi^2$ ,  $\chi_{n-1}$  and  $\chi_{n-2}$  are approximately normal with means  $(n - \frac{3}{2})^{\frac{1}{2}}$ ,  $(n - \frac{5}{2})^{\frac{1}{2}}$  and variances  $\frac{1}{2}$ . Hence  $L$  is approximately normal with mean

$$(n - \frac{3}{2})\tilde{\rho} - (n - \frac{5}{2})\tilde{r}_0$$

and variance

$$1 + \frac{1}{2}(\tilde{\rho}^2 + \tilde{r}_0^2).$$

Accordingly,

$$\begin{aligned} P(r \leq r_0) &\equiv P(L \leq 0) \\ &\equiv \Phi \left[ \frac{(n - \frac{5}{2})^{\frac{1}{2}} \tilde{r}_0 - (n - \frac{3}{2})^{\frac{1}{2}} \tilde{\rho}}{\{1 + \frac{1}{2}(\tilde{r}_0^2 + \tilde{\rho}^2)\}^{\frac{1}{2}}} \right], \quad (20) \end{aligned}$$

where  $\Phi$  denotes the standardized cumulative normal distribution function; that is, replacing  $\tilde{r}_0$  by  $\tilde{r}$ ,

$$g_n(r, \rho) \equiv \frac{(n - \frac{5}{2})^{\frac{1}{2}} \tilde{r} - (n - \frac{3}{2})^{\frac{1}{2}} \tilde{\rho}}{\{1 + \frac{1}{2}(\tilde{r}^2 + \tilde{\rho}^2)\}^{\frac{1}{2}}} \quad (21)$$

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100



is approximately distributed as a standardized normal variate<sup>†</sup>. This result has the following interesting interpretation.  $r$  is asymptotically normal with mean  $\rho$  and variance  $(1 - \rho^2)^2/n$ . Hence  $\tilde{r} (= r/(1-r^2)^{\frac{1}{2}})$  is likewise asymptotically normal with mean  $\tilde{\rho} (= \rho/(1 - \rho^2)^{\frac{1}{2}})$  and variance

$$\left( \frac{d}{dr} \frac{r}{(1 - r^2)^{\frac{1}{2}}} \right)_{r = \rho}^2 \frac{(1 - \rho^2)^2}{n} = \frac{1}{n(1 - \rho^2)} = \frac{1 + \tilde{\rho}^2}{n}.$$

The corresponding estimate of variances is  $(1 + \tilde{r}^2)/n$ . An approximate standard error of  $\tilde{r}$ , based on averaging the last two variances, is

$$\left[ \frac{1}{n} \left\{ 1 + \frac{1}{2}(\tilde{r}^2 + \tilde{\rho}^2) \right\} \right]^{\frac{1}{2}},$$

so that

$$\frac{n^{\frac{1}{2}}(\tilde{r} - \tilde{\rho})}{\left\{ 1 + \frac{1}{2}(\tilde{r}^2 + \tilde{\rho}^2) \right\}^{\frac{1}{2}}},$$

a variate which in large samples is not essentially distinct from  $g_n(r, \rho)$ , is asymptotically normal with zero mean and unit standard derivation.

The efficacy of the proposed approximation based on  $g_n(r, \rho)$  (formula (20)) can be judged from the accompanying tables, in which  $C$ ,  $F_1$ ,  $F_2$ ,  $H$ ,  $N$  have the following meaning:

$C = 10^5 \times$  error when  $g_n(r, \rho)$  is approximated by a standardized normal variate,

$F_1 = 10^5 \times$  error when  $z = \frac{1}{2} \log \frac{1+r}{1-r}$  is approximated by a normal variate with mean

$\zeta = \frac{1}{2} \log \frac{1+\rho}{1-\rho}$  and variance  $\frac{1}{n-3}$  (Fisher's normalization of  $r$ ),

<sup>†</sup>This result can also be obtained directly from (10) with the aid of the following result due to Geary (1930). If  $q$  denotes the ratio of two independent normal variates with means  $\mu_1, \mu_2$  and standard derivations  $\sigma_1, \sigma_2$ , then  $(\mu_1 - \mu_2 q)/(\sigma_1^2 + \sigma_2^2 q^2)^{\frac{1}{2}}$  is approximately normal with zero mean and unit standard derivation, provided the range of the denominator variate is effectively positive. The required result follows from (10) on approximating  $\chi_{n-1}$  and  $\chi_{n-2}$  by normal variates as above. (Here  $q = \tilde{r}$ ,

$$\mu_1 = (n - \frac{3}{2})^{\frac{1}{2}} \tilde{\rho}, \mu_2 = (n - \frac{5}{2})^{\frac{1}{2}} \tilde{r}, \sigma_1^2 = 1 + \frac{1}{2}\tilde{\rho}^2, \sigma_2^2 = \frac{1}{2}.)$$

the following... [faded text]

the following... [faded text]

the following... [faded text]

[faded text]

the following... [faded text]

[faded text]

the following... [faded text]

[faded text]

[faded text]

[faded text]

[faded text]

[faded text]

[faded text]

$F_2 = 10^5 \times$  error when  $z$  is approximated by a normal variate with mean  $\zeta + \frac{\rho}{2(n-1)}$  and variance  $\frac{1}{n-3}$  (Fisher's normalization of  $r$ , corrected for bias),  
 $H = 10^5 \times$  error when  $z - \frac{3z+r}{n}$  is approximated by a normal variate with mean  $\zeta - \frac{3\zeta+\rho}{n}$  and variance  $\frac{1}{n-1}$  (Hotelling's<sup>†</sup> approximation, 1953),  
 $N = 10^5 \times$  error when  $r$  is approximated by a normal variate with mean  $\rho$  and variance  $(1 - \rho^2)^2/n$ ,  
 $P(r) \cong 10^5 \times$  actual value of probability integral.

The values of  $n$ ,  $r$ ,  $\rho$  in the tables are those selected by Miss David (1938; pp. xxxii - xxxiii) in her own investigation of the accuracy of the bias-corrected  $z$ -transformation and the approximation which consists in replacing  $r$  by a normal variate.

The tables suggest the following conclusions.

- (i) The proposed approximation is much superior to that based on  $z$  (good as this is);
- (ii) it is still superior to the approximation based on the bias-corrected  $z$  over the entire range of distribution, except possibly at the tails;
- (iii) it is of about the same order of accuracy as that based on Hotelling's refinement of  $z$ ;
- (iv) it is much superior to the approximation which consists in regarding  $r$  itself as normally distributed (when  $n$  is sufficiently large for the latter approximation to be at all effective).

---

<sup>†</sup> The transformation  $z^* = z - (3z+r)/n$  was obtained by Hotelling for the purpose of improving on the stabilization of variance induced by the inverse hyperbolic tangent transformation, and not primarily for improving on the  $z$ -normalization. However (as Hotelling conjectured and as the present calculations indicate), improved normalization does in fact generally appear to result from the use of  $z^*$ .



*[Faint, illegible text, possibly bleed-through from the reverse side of the page.]*

*[Faint, illegible text, possibly bleed-through from the reverse side of the page.]*

#### 4. CONCLUDING REMARKS

Equations (10) and (15) give rise to various additional approximations for the distribution of  $r$ . Although an investigation into the relative merits and demerits of these approximations hardly seems worthwhile, two of the approximations are of sufficient interest and importance to warrant some mention.

- (i) The variate  $|\xi|$  is stochastically negligible relative to the variate  $|\tilde{\rho}| \chi_{n-1}$  if  $n|\tilde{\rho}| \gg 1$ . This suggests that under the latter condition (10) can be adequately replaced by the relation

$$\tilde{r} = \tilde{\rho} \frac{\chi_{n-1}}{\chi_{n-2}},$$

$$\text{i.e., } \frac{\tilde{r}}{\tilde{\rho}} = \left(\frac{n-1}{n-2}\right)^{\frac{1}{2}} F_{n-1, n-2}^{\frac{1}{2}} \quad (|n\tilde{\rho}| \gg 1), \quad (22)$$

or

$$\log \frac{\tilde{r}}{\tilde{\rho}} = \log \left\{ \left(\frac{n-1}{n-2}\right)^{\frac{1}{2}} \right\} + z_{n-1, n-2} \quad (|n\tilde{\rho}| \gg 1), \quad (23)$$

where  $z_{n-1, n-2}$  is a variate distributed as Fisher's  $z$  (half the natural logarithm of a variance ratio) with  $n-1$  and  $n-2$  degrees of freedom ( $r$  and  $\rho$  are to be of the sign in (23)). An equivalent statement is that

$$\frac{1}{1 + \left(\frac{\tilde{r}}{\tilde{\rho}}\right)^2} = \beta_{\frac{1}{2}(n-2), \frac{1}{2}(n-1)} \quad (|n\tilde{\rho}| \gg 1), \quad (24)$$

where the variate of the right of (24) is a beta-variate of the first kind with parameters  $\frac{1}{2}(n-2)$  and  $\frac{1}{2}(n-1)$ . Spot calculations indicate that this approximation gives remarkable accuracy. It should be remarked that the calculation of the probability integral of  $r$  for high  $n$  and  $\rho$  has proved in the past to be most difficult (see E. S. Pearson's preface to Miss David's tables), and that the approximation mentioned here performs best precisely under such conditions.

1. The first part of the document discusses the importance of maintaining accurate records of all transactions. It emphasizes that proper record-keeping is essential for the integrity of the financial system and for the ability to detect and prevent fraud.

2. The second part of the document outlines the various methods used to collect and analyze data. It describes the use of statistical techniques to identify trends and anomalies in the data, and the importance of using reliable sources of information.

3. The third part of the document discusses the role of the auditor in the process. It explains that the auditor's primary responsibility is to provide an independent and objective assessment of the financial statements. This involves a thorough review of the records and a comparison of the results with the applicable accounting standards.

4. The fourth part of the document discusses the importance of transparency and accountability in the financial system. It argues that the public has a right to know how their money is being spent, and that this information should be made available in a clear and accessible format.

5. The fifth part of the document discusses the role of the government in the financial system. It explains that the government has a responsibility to ensure that the financial system is fair and equitable, and that it is able to provide the services that are needed by the public.

6. The sixth part of the document discusses the importance of education and training in the financial system. It argues that the public needs to be educated about the various risks and opportunities associated with financial transactions, and that this education should be provided in a clear and accessible format.

7. The seventh part of the document discusses the importance of innovation and technology in the financial system. It explains that the use of new technologies can help to improve the efficiency and effectiveness of the financial system, and that this should be encouraged and supported.

8. The eighth part of the document discusses the importance of international cooperation in the financial system. It argues that the financial system is a global system, and that it is essential for countries to work together to ensure its stability and integrity.

9. The ninth part of the document discusses the importance of the legal system in the financial system. It explains that the legal system provides the framework for the financial system, and that it is essential for the system to be able to enforce the rules and regulations that govern it.

10. The tenth part of the document discusses the importance of the financial system in the overall economy. It explains that the financial system is the lifeblood of the economy, and that it is essential for the economy to be able to grow and prosper.

11. The eleventh part of the document discusses the importance of the financial system in the overall economy. It explains that the financial system is the lifeblood of the economy, and that it is essential for the economy to be able to grow and prosper.

12. The twelfth part of the document discusses the importance of the financial system in the overall economy. It explains that the financial system is the lifeblood of the economy, and that it is essential for the economy to be able to grow and prosper.

13. The thirteenth part of the document discusses the importance of the financial system in the overall economy. It explains that the financial system is the lifeblood of the economy, and that it is essential for the economy to be able to grow and prosper.

14. The fourteenth part of the document discusses the importance of the financial system in the overall economy. It explains that the financial system is the lifeblood of the economy, and that it is essential for the economy to be able to grow and prosper.

15. The fifteenth part of the document discusses the importance of the financial system in the overall economy. It explains that the financial system is the lifeblood of the economy, and that it is essential for the economy to be able to grow and prosper.

16. The sixteenth part of the document discusses the importance of the financial system in the overall economy. It explains that the financial system is the lifeblood of the economy, and that it is essential for the economy to be able to grow and prosper.

17. The seventeenth part of the document discusses the importance of the financial system in the overall economy. It explains that the financial system is the lifeblood of the economy, and that it is essential for the economy to be able to grow and prosper.

18. The eighteenth part of the document discusses the importance of the financial system in the overall economy. It explains that the financial system is the lifeblood of the economy, and that it is essential for the economy to be able to grow and prosper.

19. The nineteenth part of the document discusses the importance of the financial system in the overall economy. It explains that the financial system is the lifeblood of the economy, and that it is essential for the economy to be able to grow and prosper.

20. The twentieth part of the document discusses the importance of the financial system in the overall economy. It explains that the financial system is the lifeblood of the economy, and that it is essential for the economy to be able to grow and prosper.

For  $\rho = 0$ ,  $(n-2)^{\frac{1}{2}}\tilde{r}$  is distributed exactly as  $t$  with  $n - 2$  degrees of freedom. From general continuity considerations,  $(n-2)^{\frac{1}{2}}\tilde{r}$  is then distributed approximately as  $t$  with  $n - 2$  degrees of freedom if  $|\rho|$  is small. The result of the previous paragraph provides an interesting contrast in that it asserts that the distribution of  $\tilde{r}$  is to be referred to an  $F$ (with approximately equal degrees of freedom) if  $|\rho|$  is large.

(ii) We have seen in Section 3 that the rather crude Fisherian normalization of  $\chi^2$  already produces high accuracy with the aid of (15), involving the computation of  $P(L \leq 0)$ . It is clear that extremely high accuracy can be obtained if the latter probability is evaluated by means of a Cornish-Fisher asymptotic expansion and that such accuracy can be obtained even if only the first correcting term in the expansion is used. The main term in the expansion is  $\Phi(-\mu_L/\sigma_L)$ , where  $\mu_L$ ,  $\sigma_L$  are the mean and standard deviation of  $L$ . This does not differ materially from (19). [(19) is obtained from  $\Phi(-\mu_L/\sigma_L)$  by approximating  $\mu_L$  and  $\sigma_L$ , in that the mean and variance of a  $\chi_\nu$  are taken as  $(\nu - \frac{1}{2})^{\frac{1}{2}}$  and  $\frac{1}{2}$ .] The correcting terms, which are in powers of  $n^{-\frac{1}{2}}$ , involve the cumulants of  $L$ , and these are easily computed from (17). (Bounds to the error induced by termination after a given number of correcting terms can also be computed.)

The robustness of  $r$  (previously studied by Gayen, 1951) can also be similarly investigated through the relationship

$$\rho(r \leq r_0) = P(u + \tilde{\rho}v - \tilde{r}_0 w \leq 0), \quad (25)$$

where  $u$ ,  $v$ ,  $w$  are defined by (3), (4) and (5). The variate

$$L = u + \tilde{\rho}v - \tilde{r}_0 w \quad (26)$$

is a symmetric function of the observations whose asymptotic moments can be determined from Fisher's theory of  $k$ -statistics, and the Cornish-Fisher expansion for  $P(L \leq 0)$  thereby determined (cf. David and Johnson, 1951). The variates  $u$ ,  $v$ ,  $w$ , being functions of central moments, are asymptotically jointly normal, whence  $L$  is itself asymptotically normal. The leading term in the expansion is again  $\Phi(-\mu_L/\sigma_L)$ .

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes that proper record-keeping is essential for ensuring transparency and accountability in financial operations.

2. The second part of the document outlines the various methods and tools used to collect and analyze data. It highlights the need for consistent data collection procedures and the use of advanced analytical techniques to derive meaningful insights from the data.

3. The third part of the document focuses on the role of technology in modern data management. It discusses how cloud-based solutions and data integration tools have revolutionized the way organizations handle their information, enabling faster processing and easier access to data.

4. The fourth part of the document addresses the challenges associated with data security and privacy. It stresses the importance of implementing robust security measures to protect sensitive information from unauthorized access and data breaches.

5. The fifth part of the document explores the ethical implications of data collection and analysis. It discusses the need for transparency in data practices and the importance of obtaining informed consent from individuals whose data is being collected.

6. The sixth part of the document provides a detailed overview of the data lifecycle, from initial data collection to final reporting and archiving. It outlines the key stages and the responsibilities of different roles involved in the process.

7. The seventh part of the document discusses the importance of data quality and the steps taken to ensure high standards of accuracy and reliability. It highlights the role of data cleaning and validation in maintaining the integrity of the data.

8. The eighth part of the document focuses on the integration of data from various sources and the challenges of ensuring consistency and compatibility across different systems and formats.

9. The ninth part of the document discusses the role of data in decision-making and the importance of providing timely and accurate information to support strategic planning and operational decisions.

10. The tenth part of the document provides a summary of the key findings and recommendations from the study. It emphasizes the need for continuous improvement in data management practices and the importance of staying up-to-date with the latest trends and technologies in the field.

11. The eleventh part of the document discusses the future of data management and the emerging trends that will shape the industry. It highlights the growing importance of artificial intelligence and machine learning in data analysis and the potential for new data sources and collection methods.

12. The twelfth part of the document provides a detailed overview of the data governance framework, including the roles and responsibilities of different stakeholders and the key principles that guide the management of data.

13. The thirteenth part of the document discusses the importance of data literacy and the need for organizations to invest in training and education to ensure that their employees have the skills and knowledge to effectively manage and analyze data.

14. The fourteenth part of the document provides a detailed overview of the data security and privacy regulations that apply to organizations, including the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA).

15. The fifteenth part of the document discusses the role of data in the digital economy and the importance of data as a key asset for organizations. It highlights the need for organizations to develop a data-driven culture and to invest in the infrastructure and talent needed to support their data initiatives.

16. The sixteenth part of the document provides a detailed overview of the data integration and interoperability challenges that organizations face and the solutions that are available to address these challenges.

17. The seventeenth part of the document discusses the importance of data in the public sector and the role of data in improving government services and decision-making. It highlights the need for transparency and accountability in the use of public data and the importance of ensuring that data is used in a responsible and ethical manner.

18. The eighteenth part of the document provides a detailed overview of the data management and analytics tools and platforms that are available to organizations, including cloud-based solutions and open-source software.

19. The nineteenth part of the document discusses the importance of data in the healthcare industry and the role of data in improving patient care and outcomes. It highlights the need for data integration and interoperability in the healthcare sector and the importance of ensuring that data is used in a secure and ethical manner.

20. The twentieth part of the document provides a detailed overview of the data management and analytics trends and forecasts for the future, including the growing importance of real-time data and the use of advanced analytics techniques.



We remark in conclusion that in comparing  $g_n(r, \rho)$  (defined in (21)) with Fisher's z-transform of  $r$  from the point of view of statistical applications, rather than that (as in Section 3) of accuracy, it is clear that  $g_n(r, \rho)$  lacks the flexibility of  $z$  arising from its variance-stabilization property:  $g_n(r, \rho)$  cannot, for example, be used to test for the homogeneity of correlations in independent sets of data (unless, of course, one approximates further by substituting an estimate for the common unknown correlation). On the other hand, it can profitably be used (just as can the z-transform, but more accurately) to test the more restricted hypothesis that the population correlations have a specified common value (see pp. xxiv - xxv of Miss David's tables for examples when the latter is zero) by referring

$$\{g_{n_1}(r_1, \rho_0)\}^2 + \dots + \{g_{n_k}(r_k, \rho_0)\}^2 \quad (27)$$

to  $\chi^2$  with  $k$  degrees of freedom ( $\rho_0$  is the postulated common value of the population correlations,  $r_i$  is the correlation computed from the  $i$ th set of data,  $n_i$  = size of  $i$ th set). Another and very obvious advantage is that only tables of the normal distribution are required for probability integrals and percentage points of  $r$ . In particular, tests and confidence intervals relating to  $\rho$  and based on  $g_n(r, \rho)$  are easily constructed in the usual manner. Thus a two-sided confidence interval for  $\tilde{\rho}$  with coverage probability  $1 - \alpha$ , arising from the inversion of the inequality

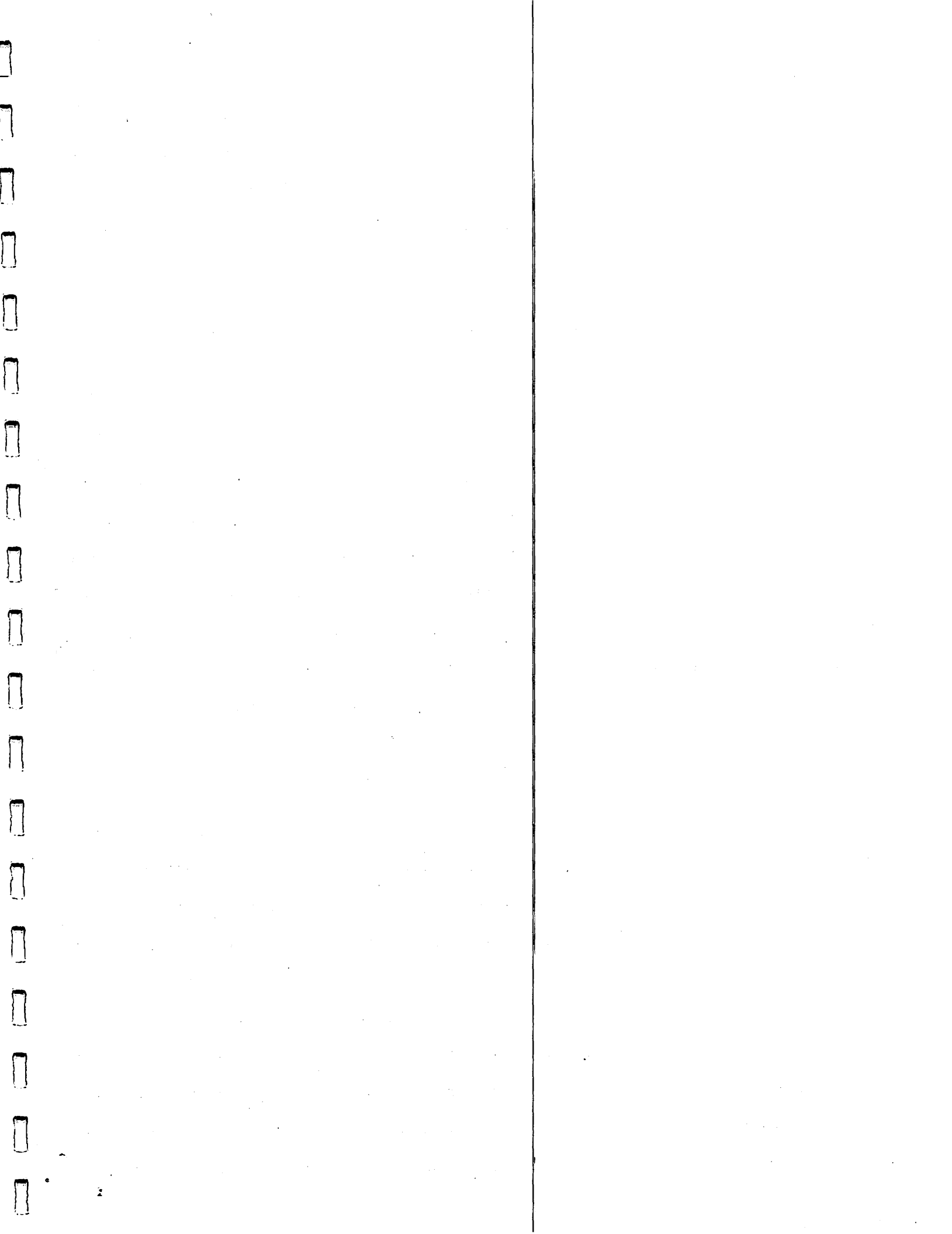
$|g_n(r, \rho)| < d_{\alpha/2}$  ( $d_{\alpha/2}$  is the two sided 100% point of a standard normal deviate) has limits

$$(2n-3-d_{\alpha/2}^2)^{-1} [ \{ (2n-5)(2n-3) \}^{\frac{1}{2}} r \pm \{ (4n-6-d_{\alpha/2}^2) + (4n-8-2d_{\alpha/2}^2) r^2 \}^{\frac{1}{2}} d_{\alpha/2} ], \quad (28)$$

a rough approximation being

$$\frac{r}{(1-r^2)^{\frac{1}{2}}} \pm d_{\alpha/2} \frac{1}{n^{\frac{1}{2}}} \frac{1}{(1-r^2)^{\frac{1}{2}}} \quad (29)$$

Denoting the confidence limits for  $\tilde{\rho}$  by  $c_1, c_2$ , the corresponding limits for  $\rho$  are  $c_1/(1+c_1^2)^{\frac{1}{2}}, c_2/(1+c_2^2)^{\frac{1}{2}}$ .



## REFERENCES

- ANDERSON, T. W. (1958). An Introduction to Multivariate Statistical Analysis. New York: Wiley.
- ARMSÉN, P. (1956). Note on a simple identity concerned with particular values of the distribution functions of correlation coefficients in bivariate normal populations. Jour. Nat. Inst. Personnel Res., 6, 175-6.
- CRAMÉR, H. (1946). Mathematical Methods of Statistics. Princeton University Press.
- DAVID, F. N. (1938). Tables of the Correlation Coefficient. Cambridge University Press.
- DAVID, F. N. and JOHNSON, N.L. (1951). The effect of non-normality on the power function of the F-test. Biometrika, 38, 43-57.
- FISHER, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika, 10, 507-21.
- GAYEN, A.K. (1951). The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes. Biometrika, 38, 219-47.
- GEARY, R. C. (1930). The frequency distribution of the quotient of two normal variables. Jour. Roy. Statist. Soc., 93, 442-46.
- HARLEY, B. I. (1957). Relation between the distributions of non-central t and of a transformed correlation coefficient. Biometrika, 44, 219-24.
- HOTELLING, H. (1953). New light on the correlation coefficient and its transforms. Jour. Roy. Statist. Soc. B, 15, 193-225.
- KENDALL, M. G. & STUART, A. (1961). The Advanced Theory of Statistics, Vol. 2. London: Charles Griffin.
- RUBEN, H. (1963). Probability of a positive sample correlation. Ann. Math. Statist., 34, 694 (abstract).
- WILKS, S. S. (1962). Mathematical Statistics. New York: Wiley.

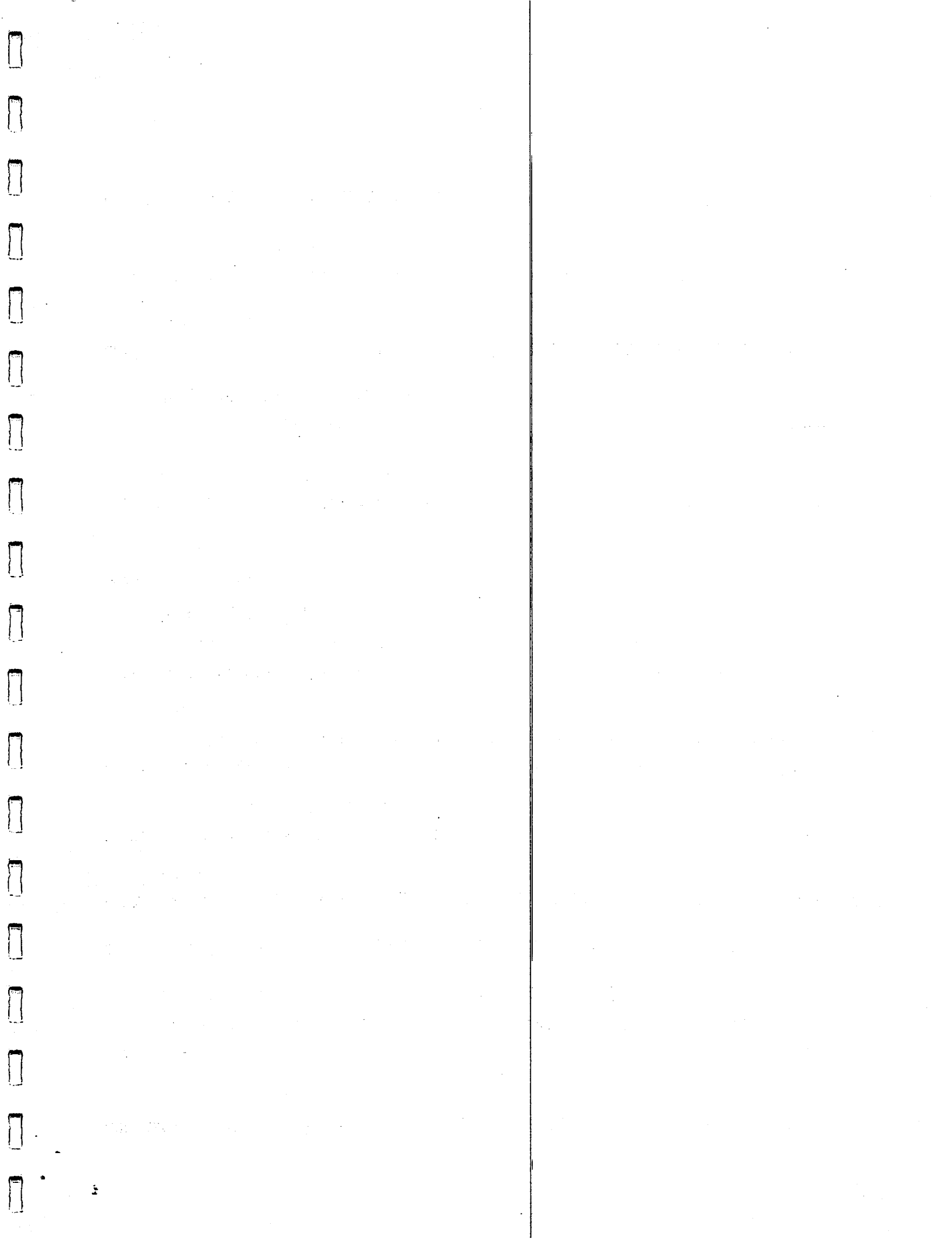


Table 1. Accuracy of approximations to probability integral of r for n = 11,  $\rho = 0.5$

r	P(r)	C	F <sub>1</sub>	F <sub>2</sub>	H
-.25	995	45	147	-48	-180
-.20	1414	52	257	-16	-194
-.15	1974	59	405	35	-197
-.10	2710	67	597	108	-188
-.05	3666	74	836	204	-163
.00	4893	80	1120	322	-124
.05	6449	83	1447	457	- 71
.80	93270	-346	717	-174	204
.825	95517	-387	580	- 56	278
.85	97267	-400	454	44	307
.875	98526	-377	332	103	281
.90	99335	-313	213	110	207

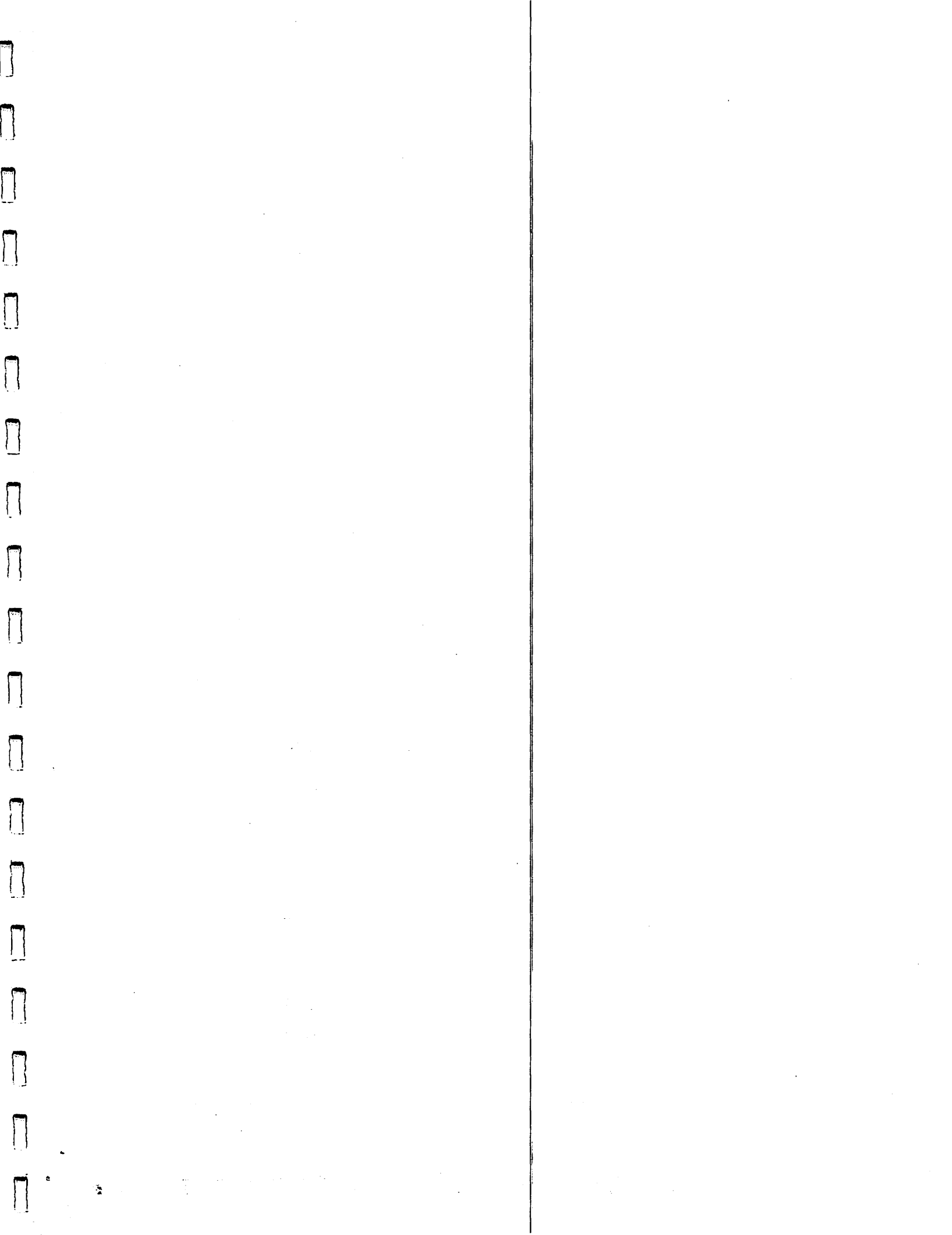


Table 2. Accuracy of approximations to probability integral of r for n = 11,  $\rho = 0.9$

r	P(r)	C	F <sub>1</sub>	F <sub>2</sub>	H
.60	881	306	497	107	-120
.625	1165	340	694	164	-123
.65	1553	372	882	240	-122
.675	2086	402	1165	342	-112
.70	2826	427	1528	475	- 90
.725	3864	441	1989	644	- 54
.75	5335	438	2561	820	- 2
.80	10461	377	4071	1360	165
.81	12047	333	4398	1441	181
.82	13873	300	4744	1528	210
.83	15998	263	5086	1600	232
.84	18470	223	5413	1650	246
.85	21342	186	5714	1672	250
.86	24679	150	5966	1651	235
.87	28545	119	6151	1579	201
.88	33006	97	6244	1449	145
.89	38127	83	6217	1249	65
.90	43957	79	6043	979	- 36
.91	50517	77	5698	646	-147
.92	57776	64	5168	271	-250
.93	65614	19	4461	98	-313
.94	73783	- 86	3611	-395	-301
.95	81851	-266	2691	-532	-179
.97	94987	-669	1040	-179	259
.975	97083	-677	723	- 34	305
.98	98571	-606	450	64	279
.985	99470	-455	227	87	184

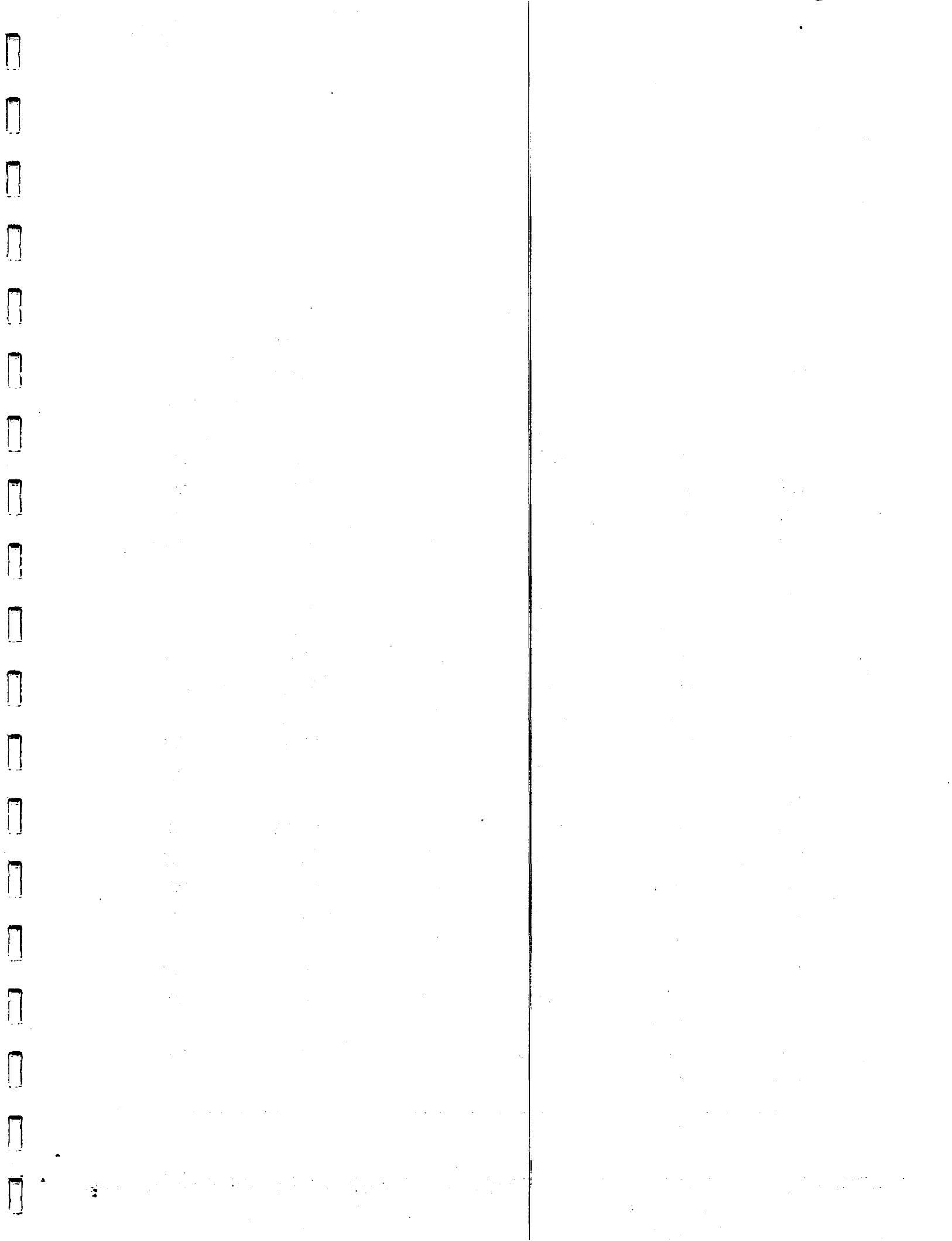




Table 3. Accuracy of approximations to probability integral of r for a = 160,  $\rho = 0.8$

r	P(r)	C	F <sub>1</sub>	F <sub>2</sub>	H
.70	172	7	16	- 2	- 4
.75	5388	22	381	26	4
.80	48729	2	1271	14	0
.85	97402	- 26	179	- 6	5

Table 4. Accuracy of approximations to probability integral of r for n = 400,  $\rho = 0$

r	P(r)	C	F <sub>1</sub>	H	N
- .16	66	0	- 1	- 1	- 4
- .14	251	0	- 2	- 2	- 7
- .12	817	0	- 3	- 3	-10
- .10	2282	0	- 2	- 3	- 7
- .08	5507	0	2	2	5
- .06	11559	0	8	7	23
- .04	21249	0	1	11	34
- .02	34503	0	8	8	27
.00	50000	0	0	0	0

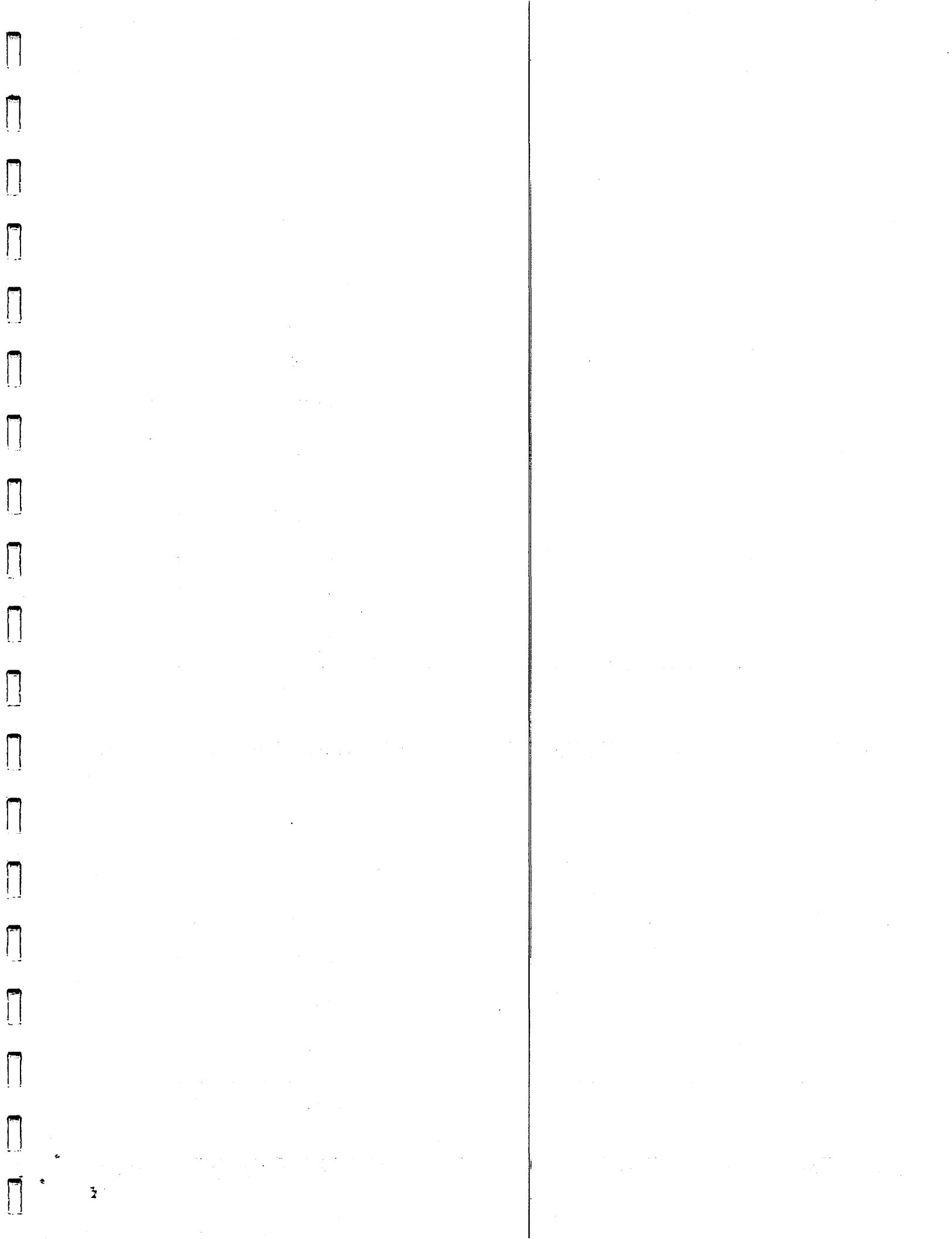


Table 5. Accuracy of approximations to probability integral of r for n = 400,  $\rho = 0.4$

$r^\dagger$	P(r)	C	F <sub>1</sub>	H	N
.22	4	- 1	- 1	- 1	- 4
.26	83	0	2	- 1	- 37
.30	1121	1	27	- 3	-220
.34	8132	0	157	3	-281
.38	31550	- 1	367	8	544
.42	60850	- 9	355	- 1	569
.46	92781	-24	110	-25	-342
.50	99341	23	44	27	-194
.54	99999	-16	- 15	-16	- 42

<sup>†</sup>In Miss David's tables each number in this column has a negative sign inadvertently attached to it.

Table 6. Accuracy of approximations to probability integral of r for n = 400,  $\rho = 0.9$

r	P(r)	C	F <sub>1</sub>	H	N
.85	1	0	0	0	- 1
.86	17	1	1	0	- 15
.87	261	4	17	- 3	-169
.88	2588	11	144	- 1	-658
.89	15254	7	561	10	63
.90	49100	0	900	0	1797
.91	86000	- 7	476	- 8	-278
.92	98941	- 9	62	4	-1670
.93	99988	0	2	1	- 67

