MONASH

WP 4/89

MONASH

UNIVERSITY

AUSTRALIA

NULL DISTRIBUTION OF THE SMALL SAMPLE MEAN CORRELATION COEFFICIENT: AN APPLICATION TO MEDICAL RESEARCH

Muhammad I. Bhatti

Working Paper No. 4/89

May 1989

DEPARTMENT OF ECONOMETRICS

ISSN 1032-3813

ISBN 0 86746 791 6

NULL DISTRIBUTION OF THE SMALL SAMPLE MEAN CORRELATION COEFFICIENT: AN APPLICATION TO MEDICAL RESEARCH

Muhammad I. Bhatti

Working Paper No. 4/89

May 1989

DEPARTMENT OF ECONOMETRICS, FACULTY OF ECONOMICS AND POLITICS

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

NULL DISTRIBUTION OF THE SMALL SAMPLE MEAN CORRELATION COEFFICIENT: AN APPLICATION TO MEDICAL RESEARCH

Muhammad I. Bhatti

In this note we outlined the procedure to obtain upper tail percentage points of the distribution of the average of k independent sample correlation coefficient (r) each of which is based on n pairs of observations. Null distribution of $\bar{r}$ demonstrates its application to medical research.

Key words: Average correlation coefficient, Bessel function, characteristics function, critical values, Fubini's theorem, medical research.

1. INTRODUCTION

The α, β and δ cells of the pancreatic islets of Langerhans produce respectively the hormones glucagon, insulin and somatostatin which play central roles in the control of blood glucose concentration [Cooperstein and Dudley (1981, Chapter 18)]. Direct and indirect interactions of these cells can be seen by Bondy and Rosenberg (1977) and Kuhl, Faber, Hornnes and Jansen (1978), Gerich (1977), etc. In biomedical sciences biostatisticians are asked to assess the significance of average correlation coefficient ($\bar{r}$) between various sets of above-mentioned cells and determine its critical values. The details of such experiments are given in Bhatti (1987), Bergons, Tam, Molner, Rajotte and McGregor (1986). This note is an attempt to provide the distribution of $\bar{r}$ and a table of such critical values for small sample size. These critical points are determined by using modern availability of computing power, with the hope of better and accurate results. In Section 2 of this note the derivation of the distribution of mean correlation coefficient ($\bar{r}$) is given. Section 3 demonstrates the table of critical values for selected values of K_1 (the number of correlations in average), N (the number of points from which each correlation is calculated), and α' (the upper tail area). If one is interested in different or higher values of K_1, N and α' he can modify APL programs given at the end of the Appendix or he can obtain the extended version of the APL programs from the author on request.

2. DERIVATION OF THE DISTRIBUTION OF $\bar{r}$

The purpose of this section is to obtain upper tail percentage points of the distribution of the average of k independent sample correlation coefficients (r) each of which is based on n pairs of independent observations. According to Hogg and Craig (1978, p.302) the null (ρ=0, where ρ is the population correlations coefficient) p.d.f. of r based on n pairs of observations is

$$g(r) = \begin{cases} \frac{\Gamma[(n-1)/2]}{\Gamma(1/2)\Gamma[(n-2)/2]} (1-r^2)^{(n-4)/2} & , \text{ for } -1 < r < 1 \\ 0 & , \text{ otherwise .} \end{cases}$$

where

$$r = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{[\Sigma(x_i-\bar{x})^2\Sigma(y_i-\bar{y})^2]^{1/2}} .$$

The moment-generating function of the null distribution of r is therefore:

$$M_r(t) = E(e^{tr}) = \frac{\Gamma[(n-1)/2]}{\Gamma(1/2)\Gamma[(n-2)/2]} \int_{-1}^{1} e^{tr}(1-r^2)^{\frac{n-3}{2} - \frac{1}{2}} dr .$$

$$M_r(t) = E(e^{tr}) = \Gamma[(n-1)/2]\ 2^{\frac{n-3}{2}}\ t^{-(\frac{n-3}{2})}\ I_{(\frac{n-3}{2})}(t), \text{ if } n > 2$$

Where

$$I_{(\frac{n-3}{2})}(t) = \frac{(t/2)^{\frac{n-3}{2}}}{(\frac{n-3}{2})!} \left[1 + \frac{t^2}{2(n-1)} + \frac{t^4}{2^3(n-1)(n+1)} + \ldots\right]$$

is the modified Bessel function of the second kind of order (n-3)/2. The corresponding characteristic function (c.f.) can be obtained by replacing t with it in $M_r(t)$. Also, note that since the p.d.f. of the distribution of r is symmetric, the c.f. will be real. Thus,

$$\phi_r(t) = M_r(it) = \Gamma[(n-1)/2]\ 2^{(\frac{n-3}{2})}\ (it)^{-\frac{1}{2}(n-3)}\ I_{(\frac{n-3}{2})}(it) .$$

Expressing $\phi_r(t)$ in terms of the Bessel function $J_{(\frac{n-3}{2})}(t)$, and by using the relation

$$I_v(z) = i^{-v} J_v(iz) \quad ,$$

we get the c.f. as,

$$\phi_r(t) = \Gamma[(n-1)/2]2^{(\frac{n-3}{2})} t^{-\frac{1}{2}(n-3)} J_{(\frac{n-3}{2})}(t) \; .$$

Thus, the c.f. of the distribution of the mean $\bar{r}$ of k independent values of r (i.e., the c.f. of the distribution of $\bar{r}$) is given by

$$\phi_{\bar{r}}(t) = M_{\bar{r}}(it) = [M_r(it/k)]^k$$

$$\phi_{\bar{r}}(t) = \left[\Gamma[(n-1)/2]\; 2^{(\frac{n-3}{2})} (t/k)^{-\frac{1}{2}(n-3)} J_{(\frac{n-3}{2})}(t/k)\right]^k \; .$$

Note, that $\phi_{\bar{r}}(t)$ is an even function of t, since $\phi_{\bar{r}}(t) = \phi_{\bar{r}}(-t)$ and thus the distribution of $\bar{r}$ is symmetric. Using the inversion formula for characteristic function, see Feller (1968), we can obtain the p.d.f. of the distribution of $\bar{r}$ as,

$$f(\bar{r}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it\bar{r}} \phi_{\bar{r}}(t)dt$$

$$= \frac{1}{\pi} \int_{0}^{\infty} \cos(t\bar{r})\phi_{\bar{r}}(t)dt \; .$$

The upper tail area above r_0 is,

$$P(\bar{r} > r_0) = \frac{1}{2} - P(0 \le \bar{r} \le r_0) = \frac{1}{2} - \int_0^{r_0} f(\bar{r})d\bar{r}$$

$$P(\bar{r} > r_0) = \frac{1}{2} - \int_0^{r_0} \left\{\frac{1}{\pi} \int_0^{\infty} \cos(t\bar{r})\phi_{\bar{r}}(t)dt\right\}d\bar{r} \ .$$

Using Fubini's theorem,

$$P(\bar{r} > r_0) = \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \phi_{\bar{r}}(t)\left[\int_0^{r_0} \cos(t\bar{r})d\bar{r}\right]dt$$

$$= \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \phi_{\bar{r}}(t)\ \frac{1}{t} \sin(tr_0)dt \quad .$$

An APL program to evaluate the critical values is given in Appendix I.

3. COMPUTATION OF CRITICAL VALUES

The results of these computed critical values are given in Table 1, for selected values of N, K_1 and α'. Where N is the number of points from which each correlation is calculated (3, 4, 5, ..., 10), K_1 is the number of correlations in average (1, 2, 3, ..., 10) and α' is the upper tail area (0.05, 0.025, 0.01, 0.005).

TABLE OF CRITICAL VALUES FOR SELECTED VALUES OF N, K_1 AND α

N	K_1 / α'	1	2	3	4	5	6	7	8	9	10
3	.05	.988	.8458	.6775	.5852	.5227	.4766	.4409	.4123	.3887	.3686
	.025	.997	.9211	.7952	.6836	.6133	.5610	.5200	.4868	.4592	.4360
	.01	.9995	.9761	.8828	.7891	.7119	.6533	.6074	.5703	.5391	.5117
	.005	.9998	.9956	.9297	.8535	.7734	.7129	.6641	.625	.5908	.5625
4	.05	.9121	.6821	.5518	.4768	.4260	.3884	.3594	.3362	.3169	.3005
	.025	.9511	.7739	.6455	.5591	.5015	.4583	.4246	.3975	.3750	.3557
	.01	.9870	.8608	.7383	.6484	.5840	.5356	.4971	.4663	.4404	.4179
	.005	.9999	.9082	.7988	.7051	.6367	.5859	.5449	.5117	.4834	.4590
5	.05	.8030	.5881	.4773	.4126	.3688	.3364	.3114	.2913	.2745	.2604
	.025	.8823	.6802	.5591	.4853	.4351	.3977	.3684	.3449	.3254	.3088
	.01	.9744	.7656	.6465	.5654	.5088	.4663	.4331	.4058	.3833	.3643
	.005	.9743	.8164	.7012	.6172	.5566	.5117	.4756	.4463	.4219	.4004
6	.05	.7292	.5239	.4263	.3689	.3297	.3008	.2784	.2605	.2455	.2329
	.025	.8120	.6094	.5005	.4346	.3894	.3559	.3298	.3086	.2913	.2764
	.01	.9409	.6982	.5811	.5078	.4570	.4185	.3882	.3638	.3433	.3262
	.005	.9547	.7500	.6328	.5557	.5010	.4590	.4268	.4004	.3779	.3594
7	.05	.6694	.4775	.3889	.3365	.3009	.2745	.2541	.2377	.2241	.2125
	.025	.7588	.5571	.4575	.3972	.3560	.3252	.3013	.2820	.2659	.2524
	.01	.8447	.6416	.5332	.4648	.4180	.3828	.3550	.3325	.3140	.2979
	.005	.8887	.6924	.5810	.5098	.4590	.4209	.3906	.3662	.3457	.3286
8	.05	.6216	.4416	.3599	.3115	.2786	.2451	.2352	.2200	.2074	.1968
	.025	.7065	.5166	.4238	.3682	.3296	.3013	.2791	.2611	.2463	.2336
	.01	.7910	.5977	.4951	.4316	.3877	.3545	.3291	.3081	.2910	.2761
	.005	.8389	.6475	.5410	.4746	.4258	.3906	.3623	.3398	.3208	.3047
9	.05	.5884	.4128	.3367	.2913	.2605	.2377	.2200	.2058	.1940	.1841
	.025	.6680	.4839	.3970	.3445	.3086	.2817	.2610	.2444	.2305	.2186
	.01	.7529	.5615	.4644	.4048	.3633	.3320	.3081	.2886	.2725	.2583
	.005	.8008	.6094	.5078	.4443	.3994	.3662	.3398	.3184	.3008	.2852
10	.05	.5493	.3889	.3173	.2747	.2455	.2241	.2074	.1940	.1829	.1735
	.025	.6328	.4568	.3745	.3250	.2910	.2659	.2462	.2305	.2173	.2062
	.01	.7178	.5313	.4373	.3818	.3428	.3135	.2905	.2722	.2568	.2437
	.005	.7666	.5781	.4373	.4199	.3789	.3457	.3203	.3003	.2832	.2695

ACKNOWLEDGEMENT

This research was supported by the National Research Council of Canada and the Australian Research Council.

I am grateful to Professor Maxwell L. King and Mr Wu Xiao Ping, Monash University, Australia, and to Professor John R. McGregor, University of Alberta, Edmonton, Canada, for encouragement and helpful comments on early drafts of this paper.

REFERENCES

Bergins, R., Tam, Y.K., Molnar, G.D., Rajotte, R.V., Weiczorek, K.R., McGregor, J.R. and Fawrett, D.M., (1986): Pharmicokindic Approach to the Estimation of Hepatic Removal of Insulin. Pancreas (Vol. 1, No. 6), pp. 544-549.

Bhatti, M.I. (1987): The Derivation and Application to Medical Research of the Null Distribution of the Small Sample Mean Correlation Coefficient, unpublished M.Sc. project, University of Alberta, Edmonton, Canada.

Bondy, P.K., and Rosenberg, L.E., (1980): Metabolic Control and Disease (8th Ed.). W.B. Saunders Company.

Cooperstein, S.J. and Dudley, W. (ed.) (1981): The Islets of Langerhans, Academic Press, John Wiley and Sons, Inc., New York.

Feller, W. (1968): An Introduction to Probability Theory and Its Application, Vol. 1, (Chapter XV, 3rd ed.)

Gerich, J.E. (1977): Sonatostatin. Its Possible Role in Carbohydrate Homeostasis and the Treatment of Diabetes Mellitus. Arch. Inter.

Hogg, R.V. and Craig, A.T. (1978): Introduction to Mathematical Statistics, 4th ed. The Macmillan Co., New York.

Kuhl, C., Faber, O.K., Hornnes, P. and Jensen, S.L. (1978): C-Peptide Metabolism and the Liver. Diabetes (27, suppl. 1), 197.

APPENDIX-I

APL - PROGRAM USED TO CALCULATE CRITICAL VALUES OF THE DISTRIBUTION OF THE AVERAGE SAMPLE CORRELATION.

The program consists of the following input:

(1) $N_1 = 100$; the number of subintervals in the numerical integration.

(2) N; Number of points from which each correlation is calculated (3,4,5,6,...,10).

(3) K_1; Number of correlations in average (1,2,3,...,10).

(4) S.G; Upper tail area (0.05, 0.025, 0.01, 0.005).

The output consists of the requisite critical values given below:

CRITICAL VALUES FOR DISTRIBUTION OF AVERAGE CORRELATION

```
          ∇RBAR[⎕]∇
          ∇ RBAR

[1]       X1←(0, N1)÷N1
[2]       X2←(X1[1+ N1]+X1[ N1])÷2
[3]       K←((01)*0.5)x(!((N-3)÷2))÷(!((N-4)÷2))
[4]       12←0
[5]       A1←0
[6]       X3←A1+X2
[7]       BV←N1ρ0
[8]       L←1
[9]       U←X3[L]
[10]      BV[L]←Kx+/(1÷N1)x(20(Ux(20(0X2))))x((100x2)*(N-3)
[11]      L←L+1
[12]      →(9x(L≤N1))+(13x(L>N1))
[13]      I3←+/((10RxK1xX3)÷X3)x(BV*K1)x(1÷N1)
[14]      I2←I2+I3
[15]      A1←A1+1
[16]      J1←((1÷01)xI3)*2
[17]      L1←10*¯8
[18]      →(6x(J1≥L1)x(A1<100))+(19x((J1<L1)+(A1≥100)))
[19]      I1←0.5-(1÷01)xI2
          ∇


          ∇RNDF[⎕]∇
          ∇ RNDF
[1]       C←((LI1x10000)+(((10000xI1)-(L10000xI1))>0.5))÷10000
          ∇


          ∇RSRCH[⎕]∇
          ∇ RSRCH
[1]       A←0
[2]       B←1
[3]       R←(A+B)÷2
[4]       R
[5]       RBAR
[6]       RNDF
[7]       →(8x(C>SG))+(10x(C<SG))+(12x(C=SG))
[8]       A←R
[9]       →3
[10]      B←R
[11]      →3
[12]      R
          ∇
```

MONASH UNIVERSITY

DEPARTMENT OF ECONOMETRICS

WORKING PAPERS

1988

1/88 John Preston and Esme Preston, "Two Papers on Simple Word Processors for the PC".

2/88 Esme Preston, "Multilingual and Mathematical Text Processing".

3/88 Esme Preston and John Preston, "The Accountant's PC: Getting Started".

4/88 Maxwell L. King, "Towards a Theory of Point Optimal Testing".

5/88 Ralph D. Snyder, "Statistical Foundations of Exponential Smoothing".

6/88 Grant H. Hillier, "On the Interpretation of Exact Results for Structural Equation Estimators".

7/88 Kuldeep Kumar, "Some Recent Developments in Non-Linear Time Series Modelling.

8/88 Maxwell L. King, "Testing for Fourth-Order Autocorrelation in Regression Disturbances When First-Order Autocorrelation is Present".

9/88 Ralph D. Snyder, "Kalman Filtering with Partially Diffuse Initial Conditions".

10/88 Asraul Hoque, "Indirect Rational Expectations and Estimation in a Single Equation Macro Model".

11/88 Asraul Hoque, "Efficiency of OLS Relative to C-O for Trended x and Positive Autocorrelation Coefficient".

12/88 Russel J. Cooper & Keith R. McLaren, "Regular Alternatives to the Almost Ideal Demand System".

13/88 Maxwell L. King, "The Power of Student's t Test: Can a Non-Similar Test Do Better?".

14/88 John Hamilton, "On-Line Management of Time Series Databases: Database Retrieval Program DBR".