Journal of Statistical Planning and Inference 14 (1986) 263-268
North-Holland

OPTIMAL TESTS FOR THE CORRELATION COEFFICIENT IN A SYMMETRIC MULTIVARIATE NORMAL POPULATION

D.V. GOKHALE

Department of Statistics, University of California, Riverside, CA 92521, USA

Ashis SEN GUPTA

Department of Statistics, University of Wisconsin, Madison, WI 53706, USA and Indian Statistical Institute, Calcutta, India

Received 4 December 1984; revised manuscript received 21 October 1985
Recommended by P.S. Puri

Abstract: Consider a symmetric multivariate normal distribution (Rao, 1973; p. 196) with intraclass correlation coefficient ϱ. This paper gives optimal tests for H_0: $\varrho=0$ against H_1: $\varrho>0$, when some or none of the maginal parameters are known. The tests are locally most powerful similar and are unbiased whatever be the alternative value of ϱ.

AMS Subject Classification: 62H10.

Key words: Intraclass correlation coefficient; Symmetric multivariate normal distributions; Locally most powerful similar tests; Unbiased tests.

1. Notation and introduction

Let $g(x\,|\,m,\sigma^2,\varrho)$ denote the k-variate normal density given by

$$g(x\,|\,m,\sigma^2,\varrho)=(2\pi\sigma^2)^{-k/2}|\Sigma_\varrho|^{-1/2}\exp\{-(x-M)^{\mathrm{T}}\Sigma_\varrho^{-1}(x-M)/2\sigma^2\} \qquad (1)$$

where $M=(m,m,\dots,m)^{\mathrm{T}}$, $\sigma^2>0$, $\Sigma_\varrho=((\varrho+(1-\varrho)\delta_{ij}))$, δ_{ij} being the Kronecker delta and $-(k-1)^{-1}<\varrho<1$. This density has been extensively used as a model in multivariate analysis (e.g., see Gleser and Olkin (1969), Rao (1973), Kshirsagar (1972), etc.) Let $X_1,X_2,\dots,X_n$ be a random sample from the above distribution. In this article we derive locally most powerful similar and unbiased tests of H_0: $\varrho=0$ against H_1: $\varrho>0$ for the cases when some or none of m and σ^2 are known.

We consider one-sided alternatives H_1: $\varrho>0$ (the results for $\varrho>0$ follow analogously). We restrict to one-sided alternatives because in most practical situations the sign of ϱ is known. Also, such a restriction allows us to obtain locally most powerful similar tests of $\varrho=0$ which are globally unbiased, i.e., their unbiasedness holds for every fixed ϱ, $0<\varrho<1$.

To complete the notation, let $X_s=(X_{s1},X_{s2},\dots,X_{sk})^{\mathrm{T}}$, X_{si} denoting the i-th component of the vector X_s, $\bar{X}$ denote the sample mean vector; $\bar{\bar{X}}$ be the sample grand mean $\bar{\bar{X}}=(\sum_{s=1}^{n}\sum_{i=1}^{k}X_{si})/nk$, and $\bar{X}_s$ be defined by $\bar{X}_s=(\sum_{i=1}^{k}X_{si})/k$, $s=1,2,\dots,n$. Let

$$W=\sum_{s=1}^{n}\sum_{i=1}^{k}(X_{si}-\bar{X}_s)^2, \qquad B=k\sum_{s=1}^{n}(\bar{X}_s-\bar{\bar{X}})^2$$

and

$$T=\sum_{s=1}^{n}\sum_{i=1}^{k}(X_{si}-\bar{\bar{X}})^2=B+W.$$

Note that $W/(1-\varrho)\sigma^2$ is distributed as $\chi^2_{n(k-1)}$, $B/\{1+(k-1)\varrho\}\sigma^2$ is distributed as χ^2_{n-1} and W and B are independent (Rao, 1973).

2. Locally most powerful similar and unbiased tests

The structure of the locally most powerful similar test (Spjøtvoll, 1968) in each of the cases mentioned in the previous section is based on the inequality

$$\sum_{s=1}^{n}\left[\frac{\partial}{\partial\varrho}\ln g(x_s\,|\,m,\sigma^2,\varrho)\right]_{\varrho=0}\geq c(t) \tag{2}$$

where $c(t)$ *generically* denotes a constant depending on a fixed value t of the sufficient statistic in each case. Verification of regularity conditions in our problem is straightforward though tedious. Unbiasedness of these tests is proved separately for each case.

Now,

$$\frac{\partial}{\partial\varrho}\ln g(x_s\,|\,m,\sigma^2,\varrho)=-\frac{1}{2}\left[(\det\Sigma_\varrho^{-1})\frac{\partial}{\partial\varrho}(\det\Sigma_\varrho)\right.$$
$$\left.+\frac{1}{\sigma^2}\frac{\partial}{\partial\varrho}\{(x_s-M)^{\mathrm{T}}\Sigma_\varrho^{-1}(x_s-M)\}\right], \tag{3}$$

$$\frac{\partial}{\partial\varrho}(\det\Sigma_\varrho)=-k(k-1)\varrho(1-\varrho)^{k-2}, \tag{4}$$

$$\frac{\partial}{\partial\varrho}\{(x_s-M)^{\mathrm{T}}\Sigma_\varrho^{-1}(x_s-M)\}=\left[\{1+(k-1)\varrho\}^2\sum_{i=1}^{k}(x_{si}-m)^2\right.$$
$$\left.-\{1+(k-1)\varrho^2\}\left\{\sum_{i=1}^{k}(x_{si}-m)\right\}^2\right]$$
$$+\{1+(k-1)\varrho\}(1-\varrho)^2. \tag{5}$$

When $\varrho=0$, (4) equals zero and (5) equals $\sum_{s=1}^{n}\sum\sum_{i\neq j=1}^{k}(x_{si}-m)(x_{sj}-m)$. Hence

(2) reduces to

$$\sum_{s=1}^{n} \sum\sum_{i\neq j=1}^{k} (x_{si}-m)(x_{sj}-m) \geq c(t). \tag{6}$$

Case (i): m known, σ^2 unknown.

Assume without loss of generality that $m=0$ and note that $\sum_{s=1}^{n} \sum_{i=1}^{k} x_{si}^2$ is a complete sufficient statistic for σ^2 under H_0. The inequality (6) for the locally most powerful similar test can be written as

$$T_1 = \frac{\sum_{s=1}^{n} (\sum_{i=1}^{k} x_{si})^2}{\sum_{s=1}^{n} \sum_{i=1}^{k} x_{si}^2} \geq c\left(\sum_{s=1}^{n} \sum_{i=1}^{k} x_{si}^2\right).$$

Since the distribution of T_1 is independent of σ^2, T_1 is stochastically independent of the complete sufficient statistic $\sum_{s=1}^{n} \sum_{i=1}^{k} x_{si}^2$. Hence the critical region can be written as

$$T_1 \geq c. \tag{7}$$

Let $Z = \sum_{s=1}^{n}(\sum_{i=1}^{k} x_{si})^2$. Then $T_1 = Z/\{W+(Z/k)\}$ and (7) reduces to

$$T_1' = \frac{Z/nk}{W/n(k-1)} \geq \frac{c(k-1)}{(k-c)} = c', \quad \text{say.}$$

Thus T_1' has the F distribution with n and $n(k-1)$ degrees of freedom under the null hypothesis.

To prove unbiasedness of the test (7) note that

$$\begin{aligned}\Pr[T_1 \geq c \mid H_1] &= \Pr\left[\frac{Z/k}{W/(k-1)} \geq c' \,\middle|\, H_1\right] \\ &= \Pr\left[F_{n;\,n(k-1)} \geq \frac{c'(1-\varrho)}{\{1+(k-1)\varrho\}}\right] \\ &\geq \Pr[F_{n;\,n(k-1)} \geq c'] = \alpha,\end{aligned}$$

since $\varrho \geq 0$.

Case (ii): m unknown, σ^2 known.

We can assume without loss of generality that $\sigma^2 = 1$. The complete sufficient statistic for m under H_0 is $\bar{\bar{X}}$. Using

$$x_{si} - m = x_{si} - \bar{\bar{X}} + \bar{\bar{X}} - m,$$

the criterion (6) reduces to

$$T_2 = (k-1)B - W \geq c(\bar{\bar{X}}).$$

The distribution of T_2 is independent of m; therefore, T_2 is stochastically independent of $\bar{\bar{X}}$. Hence the critical region is given by

$$T_2 \geq c. \tag{8}$$

Statistic T_2 is a linear combination of two independent χ^2 random variables, for which the distribution theory is developed and a computationally convenient form provided in Sen Gupta (1982). Note that T_2 can also be written as $kB-T$.

Since $T_2' = T_2/nk(k-1)$ is distributed as $\{(k-1)\chi^2_{n-1} - \chi^2_{n(k-1)}\}/nk(k-1)$ under the (null) hypothesis, its percentage points can be obtained by numerical integration. These are given in Table 1 for $\alpha = 0.10$, 0.05 and 0.01, $n = 5(5)30$ and $k = 2(1)10$. For large samples, the normal approximation with mean $-(nk)^{-1}$ and variance $2(nk-k+1)/n^2k^2(k-1)$ may be used.

Table 1
Percentage points of T_2'

n	k	$\alpha=0.10$	0.05	0.01	n	k	$\alpha=0.10$	0.05	0.01
5	2	0.40437	0.58050	0.96600	20	2	0.25303	0.33646	0.50200
	3	0.23518	0.35160	0.60740		3	0.14721	0.20024	0.30700
	4	0.16560	0.25246	0.44380		4	0.10421	0.14313	0.22200
	5	0.12772	0.19698	0.34970		5	0.08073	0.11148	0.17400
	6	0.10392	0.16150	0.28860		6	0.06591	0.09133	0.14310
	7	0.08758	0.13687	0.24570		7	0.05570	0.07736	0.12160
	8	0.07567	0.11875	0.21390		8	0.04823	0.06711	0.10569
	9	0.06661	0.10486	0.18940		9	0.04253	0.05925	0.09347
	10	0.05947	0.09388	0.16999		10	0.03803	0.05305	0.08379
10	2	0.33110	0.45179	0.70100	25	2	0.23020	0.30446	0.45040
	3	0.19300	0.27138	0.43500		3	0.13384	0.18070	0.27420
	4	0.13650	0.19450	0.31620		4	0.09475	0.12905	0.19790
	5	0.10563	0.15169	0.24850		5	0.07341	0.10047	0.15500
	6	0.08617	0.12435	0.20480		6	0.05994	0.08229	0.12746
	7	0.07277	0.10538	0.17410		7	0.05066	0.06970	0.10820
	8	0.06298	0.09143	0.15150		8	0.04387	0.06045	0.09405
	9	0.05551	0.08075	0.13409		9	0.03869	0.05337	0.08316
	10	0.04963	0.07230	0.12026		10	0.03460	0.04778	0.07453
15	2	0.28440	0.38150	0.57700	30	2	0.21267	0.28020	0.41200
	3	0.16562	0.22790	0.35500		3	0.12357	0.16596	0.25000
	4	0.11722	0.16310	0.25730		4	0.08749	0.11844	0.18020
	5	0.09078	0.12710	0.20190		5	0.06779	0.09218	0.14107
	6	0.07410	0.10416	0.16620		6	0.05536	0.07549	0.11590
	7	0.06260	0.08825	0.14120		7	0.04678	0.06393	0.09840
	8	0.05420	0.07656	0.12280		8	0.04052	0.05544	0.08550
	9	0.04779	0.06761	0.10865		9	0.03573	0.04895	0.07558
	10	0.04273	0.06053	0.09740		10	0.03196	0.04382	0.06773

If $n=1$, i.e. if we have only one k-variate observation from (1), it is easy to see that (8) is equivalent to

$$W \leq c$$

where W is χ^2_{k-1} under H_0.

To prove unbiasedness of (8) note that

$$\begin{aligned}\Pr[T_2 \geq c] &= \Pr[(k-1)B - W \geq c \mid \mathrm{H}_1] \\ &= \Pr[\{1+(k-1)\varrho\}(k-1)\chi^2_{n-1} - (1-\varrho)\chi^2_{n(k-1)} \geq c] \\ &\geq \Pr[(k-1)\chi^2_{n-1} - \chi^2_{n(k-1)} \geq c] \\ &= \Pr[(k-1)B - W \geq c \mid \mathrm{H}_0] \\ &= \alpha.\end{aligned}$$

Case (iii): Both m and σ^2 unknown.

In this case, $\bar{\bar{X}}$ and $T = \sum_{s=1}^{n} \sum_{i=1}^{k} (x_{si} - \bar{x}_s)^2$ are complete and sufficient under H_0. The criterion (6) can be written as

$$T_3 = \frac{B}{T} \geq c(\bar{\bar{X}}, T).$$

Once again the distribution of T_3 is independent of m and σ^2 and hence T_3 is stochastically independent of $(\bar{\bar{X}}, T)$. Thus the locally most powerful similar test becomes

$$T_3 \geq c. \tag{9}$$

Note that, under H_0, B/T has a Beta distribution with parameters $\frac{1}{2}(n-1)$ and $\frac{1}{2}n(k-1)$, hence the critical value of T_3 can be easily obtained.

Unbiasedness is proved in a manner similar to case (i).

Case (iv): Both m and σ^2 known.

It can be assumed without loss of generality that $m=0$ and $\sigma^2=1$. From (6) the locally most powerful test is given by

$$T_4 = \sum_{s=1}^{n} \sum\sum_{i \neq j = 1}^{k} x_{si} x_{sj} \geq c. \tag{10}$$

Statistic T_4 gives the 'best natural unbiased estimator' of ϱ (Sen Gupta, 1982). Distribution of T_4 is expressible in terms of the well-tabulated Kummer function (Sen Gupta, 1982) or under H_0 as

$$\{(k-1)\chi^2_n - \chi^2_{n(k-1)}\}/nk(k-1).$$

Table 2 gives the percentage points of $T_4' = T_4/nk(k-1)$ for $\alpha = 0.10$, 0.05 and 0.01, $n = 5(5)30$ and $k = 2(1)10$, obtained as in case (ii). For large values of n, the distribution of T_4' can be approximated by the normal distribution with mean zero and variance $2/nk(k-1)$.

Unbiasedness is proved in a manner similar to case (ii).

Table 2
Percentage points of T_4'

n	k	$\alpha=0.10$	0.05	0.01	n	k	$\alpha=0.10$	0.05	0.01
5	2	0.54220	0.73200	1.14180	20	2	0.28200	0.36686	0.53520
	3	0.32866	0.45390	0.72560		3	0.16682	0.22080	0.32940
	4	0.23636	0.32977	0.53280		4	0.11903	0.15866	0.23890
	5	0.18466	0.25910	0.42120		5	0.09263	0.12396	0.18759
	6	0.15155	0.21345	0.34830		6	0.07586	0.10176	0.15448
	7	0.12852	0.18149	0.29700		7	0.06424	0.08632	0.13130
	8	0.11158	0.15786	0.25880		8	0.05571	0.07495	0.11420
	9	0.09858	0.13968	0.22940		9	0.04919	0.06624	0.10106
	10	0.08830	0.12526	0.20590		10	0.04403	0.05934	0.09060
10	2	0.39315	0.51819	0.77570	25	2	0.25300	0.32825	0.47600
	3	0.23508	0.31630	0.48520		3	0.14926	0.19679	0.29160
	4	0.16832	0.22846	0.35400		4	0.10640	0.14120	0.21100
	5	0.13122	0.17897	0.27890		5	0.08276	0.11023	0.16556
	6	0.10756	0.14716	0.23020		6	0.06775	0.09044	0.13625
	7	0.09114	0.12497	0.19600		7	0.05737	0.07670	0.11578
	8	0.07908	0.10860	0.17060		8	0.04975	0.06658	0.10067
	9	0.06985	0.09603	0.15110		9	0.04391	0.05883	0.08905
	10	0.06254	0.08607	0.13558		10	0.03931	0.05270	0.07983
15	2	0.32403	0.42340	0.62330	30	2	0.23145	0.29974	0.43300
	3	0.19247	0.25626	0.38620		3	0.13627	0.17915	0.26420
	4	0.13750	0.18450	0.28080		4	0.09706	0.12840	0.19090
	5	0.10708	0.14430	0.22080		5	0.07548	0.10018	0.14960
	6	0.08772	0.11853	0.18198		6	0.06178	0.08217	0.12309
	7	0.07430	0.10059	0.15480		7	0.05230	0.06966	0.10456
	8	0.06445	0.08738	0.13469		8	0.04535	0.06047	0.09089
	9	0.05691	0.07723	0.11920		9	0.04003	0.05342	0.08038
	10	0.05095	0.06920	0.10690		10	0.03583	0.04785	0.07205

Acknowledgement

The authors would like to thank the Editor-in-Chief, a member of the Editorial Board and the referee for a very careful reading of the manuscript and several suggestions which have greatly improved the presentation of the paper.

References

Gleser, L.J. and I. Olkin (1969). Testing for the equality of means, equality of variances and equality of covariances under restrictions upon the parameter space. *Ann. Inst. Statist. Math.* 21, 33–48.

Kshirsagar, A.M. (1972). *Multivariate Analysis.* Marcel and Dekker, New York.

Rao, C.R. (1973). *Linear Statistical Inference and Its Applications.* John Wiley & Sons, New York.

Sen Gupta, A. (1982). On tests for the equicorrelation coefficient and the generalized variance of a standard symmetric multivariate normal distribution. Technical Report 54, Department of Statistics, Stanford University.

Spjøtvoll, E. (1968). Most powerful tests for some non-exponential families. *Ann. Math. Statist.* 39, 772–784.