29

J. Coll. Sci.,King Saud Univ., **13** (2), pp. 239-345 (1982)

Estimation of the Multiple Correlation Coefficient

Adnan M. Awad

* *Department of Statistics, Yarmouk University, Irbid, Jordan*

This paper provides a discussion of the estimatores of the multiple correlation coefficient given in the literature. It also gives some improvements to these estimators. New estimators have been suggested.

Key words. Estimation, Bayes estimates, multiple correlation, noncentral distributions.

Let $\underline{X} = \binom{x_1}{\underline{x}_2}\begin{matrix}1\\ p-1\end{matrix}$ be a pxl random vector having the N (μ, Σ) distribution. The population multiple correlation coefficient ρ is the maximum correlation between X_1 and a linear function of the p−1 variables in $\underline{X}_2$. Let R be the sample multiple correlation coefficient based on a sample of N = n + 1 observations on $\underline{X}$. Fisher (1928) estimated ρ^2 by the biased estimator.

$$\hat{\rho}^2_{-1} = R^2 - (p - 1)(1 - R^2)/(n - p + 1).$$

Ifram (1970) obtained the same estimator using the method of conditional moments. This estimator may be negative when R^2 is very close to zero.

The sample multiple correlation coefficient R is the maximum likelihood estimate of ρ, see *e.g.* Anderson (1958, PP. 86 − 96). Wishart 1931) estimated $\hat{\rho}^2$ by $\hat{\rho}^2_0 = R^2$. It is clear that $\hat{\rho}^2_0$ is an improvement to $\hat{\rho}^2_{-1}$ since $0 \leq \rho^2_0 \leq 1$ and var $(\hat{\rho}^2_0)$ < var $(\hat{\rho}^2_{-1})$ However, in the average, ρ^2_0 seems to overestimate ρ^2 when ρ^2 is very close to zero, since

$$(p - 1)/(N - 1) \leq E(R^2) \leq 1$$

The UMVU estimate of ρ^2 was obtained by Olikn and Pratt (1958) as

$$g(R^2) = 1 - (n - 2)(1 - R^2)\,{}_2F_1(1, 1; 1/2(n - p + 3); 1 - R^2)/(n - p + 1),$$

where ${}_2F_1$ is a hypergeometric function. For the definition of hypergeometric functions see Ifram (1965). This estimate has the defect that when R^2 is small it becomes negative.

This brings us to the question; How should a multiple correlation coefficient be estimated? Should one be estimating P^2 rather than P? Would it be better to transform and try to estimate $P^2/(1 - P^2)$? This paper provides some new estimates of P^2 and $P^2/(1 - P^2)$. Section 2 gives a comparison between the method of moment estimates and the Ifram method of conditional moments. Section 3 applies an approximation to the density of R^2 and then an estimate for P^2 is obtained through that density. A Bayes estimate is given in Section 4.

Method of Moments

Set $\theta = P^2/(1 - P^2)$ and $X = R^2/(1 - R^2)$. It is known that

$$E(R^2) = 1 - (n - p + 1)(1 - P^2)\,{}_2F_1(1, 1; 1/2(n + 2); P^2)/n.$$

So the method of moments does not lead to an «explicit» estimator. To get rid of this difficulty, Ifram (1970) has suggested a conditional method of moments estimate. His procedure depends on the fact that R^2 is a negative binomial mixture of beta variables, and it leads to the estimators.

$$\hat{P}_1^2 = nR^2/(n - p + 1) - (p - 1)/(n - p + 1) = \hat{P}_1^2$$

and

$$\hat{\theta}_1 = (n - p + 1)X / n - (p - 1) / n.$$

Now, we will show that $\hat{\theta}_1$ is a biased estimate and we will give another estimator $\hat{\theta}_2$ which is unbiased and has smaller variance than $\hat{\theta}_1$. Thus $\hat{\theta}_2$ may be an improvement to $\hat{\theta}_1$.

Wijsman (1959) has shown that X is a gamma mixture of noncentral F', namely:

$$X \equiv F'_{p-1,n-p+1}(Y)$$

where Y is a gamma random variable with parameters $(1/2\,n, 2\theta)$. Hence

$$\begin{aligned} E(X) &= E(E(F'_{p-1,n-p+1}(Y) \mid Y)) \\ &= E((p - 1)(1 + Y/(p - 1))/(n - p - 1)) \\ &= (p - 1)(1 + n\theta/(p - 1))/(n - p - 1). \end{aligned}$$

So, the method of moments estimate for θ and P^2 are

$$\hat{\theta}_2 = (n - p - 1)X/n - (p - 1)/n.$$

$$\hat{P}_2^2 = 1 - n/((n - p - 1)X + (n - p + 1)).$$

Note that $\hat{\theta}_1 = \hat{\theta}_2 + 2X/n$, $E(\hat{\theta}_2) = \hat{\theta}$, and

$\operatorname{var}(\hat{\theta}_2)/\operatorname{var}(\hat{\theta}_1) = ((n - p - 1)/(n - p + 1))^2 < 1$. Hence $\hat{\theta}_1$ is a biased estimator and θ_2 is an unbiased estimator. Moreover $\hat{\theta}_2$ has smaller variance than $\hat{\theta}_1$.

It is clear that $\hat{\theta}_2$ may be negative. When $\hat{\theta}_2$ is positive with high probability? It will be shown that it is so when $n \geq 64$.

Now,

$$\begin{aligned}\text{var}(\hat{\theta}_2) &= ((n-p-1)/n)^2 \text{ var}(X)\\ &= ((n-p-1)/n)^2 \{E(\text{var}(F'_{p-1,n-p+1}(Y) \mid Y))\\ &\quad + \text{var}(E(F'_{p-1,n-p+1}(Y) \mid Y)\}\\ &= 2(p-1)(n-2)(n-p-3)^{-1} n^{-2} E(1 + 2Y/(p-1)\\ &\quad + Y^2/(n-2)(p-1))\\ &\quad + n^{-2}(p-1)^2 \text{ var}(1 + Y/(p-1))\\ &= 2(n-2)(n-p-3)^{-1} n^{-1}((p-1)/n + 2\theta + (2n-p-1)\theta^2/(n-2)).\end{aligned}$$

Note that var $(\hat{\theta}_2) \xrightarrow{A.S.} O$ as $n \to \infty$ and n var $(\hat{\theta}_2)$ converges to $4\theta(1+\theta)$ as $n \to \infty$ Hence $\hat{\theta}_2$ is a consistent estimate for θ and roughly speaking one may use the approximation

$$(\text{var}(\hat{\theta}_2))^{-1/2}(\hat{\theta}_2 - \theta) \xrightarrow{d} N(O,1)$$

to get the $P(\hat{\theta}_2 > 0) \doteqdot 1$ when $n \gtrsim 64$. So we recommend $\hat{\theta}_2$ when $n \gtrsim 64$.

Finally, $\hat{\theta}_1$ could be used to get its corresponding unbiased estimate, namely;

$$\hat{\theta}_3 = (n-p-1)X/n - (p-1)(n-p-3)(n-p+1)^{-1} n^{-1}.$$

Note that var $(\hat{\theta}_3)$ = var $(\hat{\theta}_2)$, $\hat{\theta}_2 - \hat{\theta}_3 = -4(p-1)/n$ and both $\hat{\theta}_2$ and $\hat{\theta}_3$ may be negative when X is very close to zero. Hence, in this case $\hat{\theta}_3$ is «better» than $\hat{\theta}_2$ when n is small in the sense that it is closer to the parameter θ.

Estimation through Approximations

Fisher (1928) has shown that as $n \to \infty$, $(n-p)R^2 = \chi_p^2((n-p)\rho^2)$. Let $\tau = (n-p)\rho^2$ and $Y = (n-p)R^2$. Applying the method of moments one gets the estimator $\hat{\rho}^2 = R^2 - p/(n-p)$.

Using the estimates of the noncentrality parameter of $\chi_p^2((n-p)\rho^2)$ given by Perlman and Rasmussen (1975) we may estimate ρ^2 by.

$$\hat{\rho}_b^2 = R^2 - p(n-p)^{-1} + (b(n-p)^{-2}(R^2)^{-1}$$

if $p \geq 5$ and be is some constant such that $0 < b < 4(p-4)$.

To compare $\hat{\rho}_b^2$ with the other estimates we need Theorem 1 P. 464 given by Parlman and Rasmussen (1975) which states that: Let Y be a $\chi_p^2(\tau)$, $\tau_b = Y - p + bY^{-1}$, $p \geq 5$, and $O < b < 4(p-4)$ then for every fixed $\tau \geq 0$,

$$E(Y - p - \tau)^2 > E(\hat{\tau}_b - \tau)^2.$$

Apply this theorem with $Y = (n-p)R^2$ and $\tau = (n-p)\rho^2$ to get

$$E(\hat{\rho}^2 - \rho^2)^2 > E(\hat{\rho}_b^2 - \rho^2)^2.$$

Hence $\hat{P}_b^{\,2}$ is an improvement to $\hat{P}^2$.

Note that $\hat{p}_b^{\,2}$ may be greater than one or negative. So there is a need to find when $\hat{P}_b^{\,2}$ lies between 0 and 1. It can be shown that this occures when.

$$\frac{n}{2(n-p)} - \frac{(n^2-4b)^{1/2}}{2(n-p)} \leqslant R^2 \leq \frac{p}{2(n-p)} - \frac{(p^2-4b)^{1/2}}{2(n-p)}$$

or

$$\frac{p}{2(n-p)} + \frac{p^2-4b}{2(n-p)} \leq R^2.$$

So, it is clear that this estimate is not a good one, since it may be used only in a small range of the possible values or R^2.

Bayes Estimation

Unfortunately, the method of moments does not lead to a positive estimate for θ or for p^2. We could get rid of this difficulty by using a Bayes method.

The Bayes approach assumes that p^2 is a random variable with some prior distribution q (p^2). Then the Bayes estimate $\hat{p}_B^{\,2}$ is the one which minimizes the posterior mean square error (see Wasan (1970), P. 185). Moreover it is known that $0 \leq p^2 \leq 1$, hence $0 \leqslant E(P^2 \mid R^2) \leq 1$ a.s. So the Bayes estimate is an improvement to the other estimates.

It is known that (see Anderson (1958), P. 95) the density of R^2 given p^2 is

$$f(R^2|P^2) = \frac{(1-p^2)^{\frac{n}{2}} \, (R^2)^{\frac{p-3}{2}}}{B\left(\frac{n-p+1}{2}, \frac{p-1}{2}\right)} (1-R^2)^{\frac{n-p-1}{2}} \cdot {}_2F_1\left(\frac{n}{2}, \frac{n}{2}; \frac{p-1}{2}; P^2 R^2\right).$$

Set $\varphi = p^2$ and $r = R^2$. Assume φ has a beta prior distribution

$q(\varphi) = \varphi^{\alpha-1}(1-\varphi)^{B-1}/B(\alpha, \beta)$.

Then the posterior mean of p^2 given r is

$$E(\varphi|r) = \int_0^1 \varphi\, \xi(\varphi)\, f(r|\varphi)\, d\varphi \Big/ \int_0^1 \xi(\varphi)\, f(r|\varphi)\, d\varphi$$

$$= \frac{\alpha \;\; {}_3F_2\left(\frac{n}{2}, \frac{n}{2}\ \alpha+1\ \frac{p-1}{2}, \alpha+\beta+1+\frac{n}{2}; r\right)}{\alpha+\beta+\frac{n}{2} \;\; {}_3F_2\left(\frac{n}{2}, \frac{n}{2}, \alpha; \frac{p-1}{2}, \alpha+\beta+\frac{n}{2}; r\right)}$$

Unfortunately this estimate $\hat{p}_B^{\,2} = E(\varphi | r)$ has a complicated expression, so there is a need to approximate it.

Note that

$$ {}_3F_2\left(\frac{n}{2}, \frac{n}{2}, \alpha; \frac{p-1}{2}, \alpha+\beta+\frac{n}{2}; R^2\right) = \sum_{k=0}^{\infty} a_k r^k $$

where

$$ a_k = \frac{\Gamma\left(\frac{n}{2}+K\right)\Gamma(\alpha+K)\,\Gamma\left(\frac{p-1}{2}\right)\Gamma\left(\alpha+\beta+\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma(\alpha)\,\Gamma\left(\frac{p-1}{2}+K\right)\Gamma\left(\alpha+\beta+\frac{n}{2}+K\right)K!} $$

Similarly

$$ {}_3F_2\left(\frac{n}{2}, \frac{n}{2}, \alpha+1; \frac{p-1}{2}, \alpha+1+\beta+\frac{n}{2}; R^2\right) = \sum_{k=0}^{\infty} b_k r^k $$

where b_k equales a_k with each α is replaced by $\alpha + 1$.

Hence

$$ \frac{\alpha+\beta+\frac{n}{2}}{\alpha} E(\varphi | r) = 1 + (b_1 - a_1)\, r + (b_2 - a_2 - a_1(b_1 - a_1))\, r^2 + \ldots.. $$

It can be shown that:

$$ b_k - a_k = a_k \frac{\left(\beta+\frac{n}{2}\right)k}{\alpha\left(\alpha+\beta+\frac{n}{2}+K\right)} $$

Hence

$$ \hat{P}^2_\beta = \frac{\alpha}{\alpha+\beta+\frac{n}{2}} + \frac{\alpha\left(\beta+\frac{n}{2}\right)\left(\frac{n}{2}\right)^2}{\frac{p-1}{2}\cdot\left(\alpha+\beta+\frac{n}{2}\right)^2\left(\alpha+\beta+1+\frac{n}{2}\right)} R^2 $$

and

$$ \mathrm{Var}\,(\hat{p}_B^{\,2}) \le \mathrm{Var}\,(R^2) \qquad \text{IF} \quad \alpha \le \frac{p-1}{2} $$

Acknowledgment

Thanks are due to the referee for his suggestions for improving this paper.

References

Anderson, T. W. (1958) *Introduction to Multivariate Analysis.* Wiley, New York.

Fisher, R.A. (1928) The general sampling distribution of the multiple correlation coefficient. *Proc. Roy. Soc. Ser. A;* **121**, 654 - 673.

Ifram, A.F. (1965) Hypergeometric functions in sequential analysis. *Ann. Math. Statist.* **36**, 1870-1872.

Ifram, A.F. (1970) On mixtures of distributions with applications to estimation. *JASA*, **65**, 749-754.

Olkin and **Pratt** (1958) Uniformly minimum variance unbiased estimate for the multiple correlation coefficient. *Ann. Math. Statist.* **29**, 201 - 208.

Perlman, M.D. and **Rasmussen, U.A.** (1975). Some remarks on estimating a noncentrality parameter. *Comm. Statist.*, **4** (5), 455 - 468.

Wasan, M.T. (1970) *Parametric Estimation.* Mc Graw - Hill. N.Y.

Wijsman, R.A. (1959) Applications of a certain representation of the Wishart matrix. *Ann. Math. Statist.*, **30**, 597 - 601.

Wishart, J. (1931) The mean and second moment coefficient of the multiple correlation coefficient in samples from a normal population. *Biometricka*, **22**, 361 - 553.

تقدير معامل الارتباط المتعدد

عدنان محمد عوض

دائرة الإحصاء، جامعة اليرموك، اربد، الأردن.

يناقش هذا البحث التقديرات المختلفة لمعامل الارتباط المتعدد التي تمت دراستها حتى الآن. كما أنه يقترح تقديرات جديدة أفضل من تلك التقديرات السابقة. وقد إستعملنا في هذا البحث ثلاث طرق لتقدير معامل الارتباط المتعدد. هي طريقة العزوم والعزوم المشروطة وطريقة التقدير التقريبي وطريقة بيز. وقد بينا أن طريقة بيز هي أفضل أسلوب يمكن استعماله لتقدير معامل الارتباط المتعدد.