**Mathematical Forum, Vol. II, No. 1, March 1979, pp. 25-31**

# ESTIMATION OF A CORRELATION COEFFICIENT FROM DOUBLE SAMPLES USING A GUESS VALUE

**S. R. SRIVASTAVA & D. C. SINGH**

Deptt. of Mathematics and Statistics

Banaras Hindu University, Varanasi, ( INDIA )

( Received on 1.9.78 )

ABSTRACT

In the present paper an attempt has been made to estimate the correlation coefficient $\rho$ of a bivariate normal population from double samples when a guess value, say $\rho_0$ of $\rho$ is available. The properties of the proposed estimator have been studied and recommendations regarding its use have been given.

## 1. *Introduction*

Katti (1962) considered an estimation procedure for the mean $\mu$ of a normal population $N(\mu, \sigma^2)$, $\sigma^2$ known from double samples when a guess value $\mu_0$ of the population mean $\mu$ was available. Arnold and Al-Bayyatti (1970) modified the procedure given by Katti. They have considered regions $R_1$ and $R_2$ in the sample space and a weighted estimator when the first sample mean belongs to these regions.

Given two normal populations with correlation coeffieients $\rho_1$ and $\rho_2$ Srivastava and Baneroft (1967) have given a rule of procedure for estimating $\rho_2$ from one or possibly two samples from the populations when it is suspected that $\rho_1$ may be equal to $\rho_2$. The object of this paper is to estimate the correlation coefficient $\rho$ of a bivariate normal population from one or possibly two samples when a prior value of $\rho$, say $\rho_0$ is given. We have defined the proposed estimator of the population correlation coefficient in Section 2 and have studied its properties in later Sections. An attempt regarding the use of the proposed estimator have also been made.

## 2. *The estimator*

Let $r_i$ $(i=1,2)$ denote a sample correlation coefficients computed from first and second samples of sizes $n_1$ and $n_2$ $(n_1>3,\ n_2>3)$ respectively. Let $z_i$ $(i=1, 2)$ denote the Fisher $z$-transformation

$$z_i=\tfrac{1}{2}\log_e\left(\frac{1+r_i}{1-r_i}\right),\quad i=1, 2$$

$$\text{and }\ \xi=\tfrac{1}{2}\log_e\left(\frac{1+\rho}{1-\rho}\right),\quad \xi_i=\tfrac{1}{2}\log_e\left(\frac{1+\rho_i}{1-\rho_i}\right),\quad i=0, 1$$

It is known that for $n$ sufficiently large $z_i \sim N\left(\xi, \frac{1}{n_{i-3}}\right)$, $i=0, 1$.

The proposed estimator $\hat{\xi}_p$ of $\xi$ is defined as follows. Take a sample of size $n_1$ from the population, compute $z_1$, If $z_1 \epsilon R$ our estimate for $\xi$ is $kz_1+(1-k)\xi_0$, $0 \leqslant k \leqslant 1$. If $z_1 \epsilon \bar{R}$, we obtain a second sample of size $n_2$ and take $z_p = \frac{(n-3)z_1+(n_2-3)z_2}{n_1+n_2-6}$ as an estimate of $\xi$. Mathematically,

$$\hat{\xi}_p = \begin{cases} kz_1+(1-K)\xi_0 & \text{if } z_1 \epsilon R \\ z_p & \text{if } z_1 \epsilon \bar{R} \end{cases}$$

The region $R$ is to be determined in three ways :

(i) By the intersection of $MSE\ (z_p)$ with $MSE\ [kz_p+(1-k)\xi_0]$ at the points $\xi = \xi_0 \pm \sqrt{\frac{1+k}{(1+k)(n_1+n_2-6)}}$. Call this region $R_1$. In this case $k \neq 1$.

(ii) Choose $R$ in such a way that $MSE\ (\hat{\xi}_p/\xi=\xi_0)$ is minimum This leads to region $R_2$ given by

$$R_2 = \left[\xi_0 - \sqrt{\frac{n_2-3}{k^2(n_1+n_2-6)^2-(n_1-3)^2}},\ \xi_0 + \sqrt{\frac{n_2-3}{k^2(n_1+n_2-6)^2-(n_1-3^2)}}\right]$$

n the case $0 \leqslant \frac{n_1-3}{n_1+n_2-6} \leqslant K \leqslant 1$.

(iii) Test the hypothesis $H_p$ ; $\xi_1 = \xi_0$ at the level $\alpha$ and take the region $R_3$ given by

$$R_3 = \left[\xi_0 - \frac{z_\alpha}{\sqrt{n_1-3}},\ \xi_0 + \frac{z_\alpha}{\sqrt{n_1-3}}\right]$$

Where $z_\alpha$ is $\left(1-\frac{\alpha}{2}\right)$ 100% probability point of a standard normal variate It may be noted that the region in each case have the form

$$R = [\xi_0 - a,\ \xi_0 + a], \text{ where } a\text{'s are constants.}$$

3. *Mean square error and the relative efficiency of the estimator*

Since the mean square error of an estimator involves the square of the bias in its expression we have not derived and discussed the expected value of the proposed estimator. The mean square error for $\hat{\xi}_p$ is given by

$$MSE\,(\hat{\xi}_p/\xi,R)=\int\limits_{z_1\epsilon R}[k(z_1-\xi_0)+\xi_0-\xi]^2\,f(z_1)dz_1+\int\limits_{z_1\epsilon \bar{R}}\frac{(n_1-3)^2(z_1-\xi)^2+(n_2-3)}{(n_1+n_2-6)^2}f(z_1)dz_1$$

$$=k^2\int\limits_{\xi_0-a}^{\xi_0+a}(z_1-\xi_0)^2f(z_1)dz_1+(1-k)^2(\xi_0-\xi)^2\int\limits_{\xi_0-a}^{\xi_0+a}f(z_1)dz_1$$

$$+2k(1-k)(\xi_0-\xi)\times\int\limits_{\xi_0-a}^{\xi_0+a}(z_1-\xi)f(z_1)\,dz_1$$

$$+\left(\frac{n_1-3}{n_1+n_2-6}\right)^2\left[\int\limits_{-\infty}^{\xi_0-a}(z_1-\xi)^2\,f(z_1)\,dz_1+\int\limits_{\xi_0+a}^{\infty}(z_1-\xi)^2\,f(z_1)\,dz_1\right]$$

$$+\frac{n_2-3}{(n_1+n_2-6)^2}\left[\int\limits_{-\infty}^{\xi_0-a}f(z_1)\,dz_1+\int\limits_{\xi_0+a}^{\infty}f(z_1)\,dz_1\right]. \qquad (3{\cdot}1)$$

If we integrate and simplify (3.1), we get

$$MSE\,(\hat{\xi}_p/\xi,\,R)=\frac{1}{n_1-3}\Big[\Big(k^2-\frac{1}{1+u}+\delta^2(1-k)^2\Big)\Big(\Phi(\delta+a\sqrt{n_1-3})$$
$$-\Phi(\delta-a\sqrt{n_1-3})\Big)+\Big(k^2-\Big(\frac{1}{1+u}\Big)^2\Big)\Big((\delta-a\sqrt{n_1-3})$$
$$\phi(\delta-a\sqrt{n_1-3})-(\delta+a\sqrt{n_1-3})\phi(\delta+a\sqrt{n_1-3})$$
$$+2k(1-k)\delta(\phi(\delta-a\sqrt{n_1-3})-\phi(\delta+a\sqrt{n_1-3}))+\frac{1}{1+u}\Big],$$

where $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x}e^{-\frac{1}{2}t^2}\,dt,\ \phi(x)=\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}x^2}$

and $\delta=\sqrt{n_1-3}\,(\xi_0-\xi),\ u=\frac{n_2-3}{n_1-3}.$

The relative efficiency of the estimator $\hat{\xi}_p$ with respect to $z_p$ is given by

$$e\,(\hat{\xi}_p/\xi,\,R)=\frac{MSE\,(z_p)}{MSE\,(\hat{\xi}_p)}$$

$$=\Big[1+(1+u)\Big\{\Big(k^2-\frac{1}{1+u}+\delta^2(1-k)^2\Big)\Big(\Phi(\delta+a\sqrt{n_1-3})-\Phi(\delta-a\sqrt{n_1-3})\Big)$$
$$+\Big(k^2-\Big(\frac{1}{1+u}\Big)^2\Big)\Big((\delta-a\sqrt{n_1-3})\phi(\delta-a\sqrt{n_1-3})-(\delta+a\sqrt{n_1-3})\phi(\delta+a\sqrt{n_1-3})$$
$$+2k(1-k)\delta(\phi(\delta-a\sqrt{n_1-3})-\phi(\delta+a\sqrt{n_1-3})\Big\}\Big]^{-1}.$$

If the initial estimate $\xi_0$ is equal to the true value $\xi$, then the relative efficiency of the estimator $\hat{\xi}_p$ with respect to $z_p$ is given by

$$e(\hat{\xi}_p/\xi_0,\ R=\Big[1+(1+u)\Big\{\Big(k^2-\frac{1}{1+u}\Big)\Big(\Phi(a\sqrt{n_1-3})-\Phi(-a\sqrt{n_1-3})\Big)$$
$$-2\Big(k^2-\Big(\frac{1}{1+u}\Big)^2\Big)(a\sqrt{n_1-3})\ \phi(a\sqrt{n_1-3})\Big\}\Big]^{-1}.$$

*Special Cases*

For $k=0$,

$$MSE\ (\hat{\xi}_p/\xi,\ R)=MSE\ (z_p)+\frac{1}{n_1-3}\Big[\Big(\delta^2-\frac{1}{1+u}\Big)\Big(\Phi(\delta+a\sqrt{n_1-3})$$
$$-\Phi(\delta-a\sqrt{n_1-3})\Big)-\Big(\frac{1}{1+u}\Big)^2(\delta-a\sqrt{n_1-3})\ \phi(\delta-a\sqrt{n_1-3})$$
$$+\Big(\frac{1}{1+u}\Big)^2(\delta+a\sqrt{n_1-3})\ \phi(\delta+a\sqrt{n_1-3})\Big]. \qquad (3.11)$$

and for $k=1$,

$$MSE(\hat{\xi}_p/\xi,\ R)=MSE\ (z_1)-\frac{1}{n_1-3}\Big[\Big(1-\frac{1}{1+u}\Big)\Big(1-\Phi(\delta+a\sqrt{n_1-3})\Big)$$
$$+\Big(1-\frac{1}{1+u}\Big)\Big(\Phi(\delta-a\sqrt{n_1-3})\Big)$$
$$-\Big(1-\Big(\frac{1}{1+u}\Big)^2\Big)\Big((\delta-a\sqrt{n_1-3})\phi(\delta-a\sqrt{n_1-3})\Big)$$
$$+\Big(1-\Big(\frac{1}{1+u}\Big)^2\Big)\Big((\delta+a\sqrt{n_1-3})\ \phi(\delta+a\sqrt{n_1-3})\Big)\Big] \qquad (3.1.2)$$

We now prove the following results :

Result I. For $k=0$ and $\sqrt{\frac{1}{1+u}}\leqslant\delta\leqslant a\sqrt{n_1-3}$, $MSE\ (\hat{\xi}_p/\xi,\ R)>MSE(z_p)$

Proof. It follows easily from the equation (3.1.1).

Result II. For $k=1$ and $\delta\leqslant a\sqrt{n_1-3}$, $MSE\ (\hat{\xi}_p/\xi,\ R)<MSE\ (z_1)$.

Proof. It follows directly from the equation (3.1.2).

4. *Recommendations*

In order to make recommendation we have calculated the relative efficiency, $e(\hat{\xi}_p/\xi, R)$ for the regions $R_1$, $R_2$, $\delta=0(\cdot5)2\cdot5$ and $k=0(\cdot4)1\cdot0$ and $u=2, 3$. For $R_2$ we have taken the same set as before with $\alpha=\cdot05$ ( see appendix ). From these tables we observe that the relative efficiency in general decreases as $u$ increases. Whereas the region $R_3$ is to be used for small values of $k$ and $\delta$, the region $R_2$ should be used for small values of $\delta$ and large values of $k$. There is not much to recommend for $R_1$.

## References

1. Arnold, J. C. and Al-Bayyatti, H. A. (1970), On double stage estimation of the mean using prior knowledge, Biometrics, 26, p. 787-800.

2. Katti, S. K. (1962), Use of some a priori knowledge in the estimation of mean from double samples, Biometrics, 18, p. 139-147.

3. Srivastava, S. R. and Bancroft, T. A. (1967), Inferences concerning a population correlation coefficient from one or possibly two samples subsequent to a preliminary test of significance, Jour. of Roy, Stat. Society B, 29, p. 282-291.

## APPENDIX

### Values of $e\ (\hat{\xi}_p/\xi,\ R)$

*Region $R_1$*

| $k \backslash \delta$ | $u=2$: 0 | .5 | 1.0 | 1.5 | 2.0 | 2.5 | $u=3$: 0 | .5 | 1.0 | 1.5 | 2.0 | 2.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1.441 | 1.007 | .638 | .553 | .575 | .700 | 1.418 | .943 | .578 | .487 | .535 | .671 |
| .2 | 1.573 | 1.026 | .636 | .494 | .523 | .643 | 1.537 | .944 | .542 | .443 | .481 | .615 |
| .4 | 1.658 | 1.038 | .609 | .562 | .471 | .576 | 1.606 | .935 | .511 | .403 | .429 | .548 |
| .6 | 1.421 | 1.051 | .665 | .509 | .489 | .572 | 1.380 | .991 | .560 | .434 | .436 | .528 |
| .8 | 0.740 | 0.701 | .594 | .475 | .408 | .409 | 0.688 | .617 | .476 | .373 | .336 | .301 |
| 1.0* | | | | | | | | | | | | |

*Region $R_2$*

| $k / \delta$ | $u=2$: .0 | .5 | 1.0 | 1.5 | 2.0 | 2.5 | $u=3$: 0 | .5 | 1.0 | 1.5 | 2.0 | 2.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| .4 | 2.114 | 1.249 | .572 | .332 | .236 | .192 | 1.870 | .970 | .438 | .291 | .263 | .306 |
| .6 | 1.445 | 1.060 | .678 | .610 | .476 | .645 | 1.424 | .951 | .570 | .458 | .436 | .595 |
| .8 | 1.283 | 0.968 | .655 | .556 | .590 | .707 | 1.282 | .903 | .642 | .487 | .534 | .585 |
| 1.0 | 1.209 | 0.964 | .697 | .532 | .657 | .772 | 1.213 | .940 | .626 | .547 | .598 | .729 |
| 0** | | | | | | | | | | | | |
| .2** | | | | | | | | | | | | |

*Region $R_3$ ; $\alpha = .05$*

| $k / \delta$ | $u=2$: 0 | .5 | 1.0 | 1.5 | 2.0 | 2.5 | $u=3$: 0 | .5 | 1.0 | 1.5 | 2.0 | 2.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 7.914 | 1.173 | .364 | .202 | .158 | .162 | 9.322 | 0.942 | .281 | .155 | .121 | .125 |
| .2 | 4.698 | 1.349 | .464 | .258 | .198 | .198 | 4.492 | 1.104 | .362 | .198 | .152 | .153 |
| .4 | 2.117 | 1.222 | .561 | .329 | .251 | .244 | 1.759 | 0.984 | .441 | .256 | .194 | .190 |
| .6 | 1.105 | 0.953 | .671 | .471 | .375 | .373 | 0.873 | 0.755 | .533 | .372 | .296 | .284 |
| .8 | 0.626 | 0.650 | .591 | .487 | .405 | .383 | .0.484 | 0.506 | .465 | .385 | .321 | .205 |
| 1.0 | 0.437 | 0.459 | .506 | .525 | .498 | .480 | 0.334 | 0.353 | .396 | .417 | .399 | .387 |

*For this value of $k$, region $R_1$ is not defined.
**For these values of $k$, region $R_2$ not defined.