The Annals of Statistics
1980, Vol. 8, No. 3, 687–692

MAXIMUM LIKELIHOOD ESTIMATES FOR A BIVARIATE NORMAL DISTRIBUTION WITH MISSING DATA[1]

BY RAM C. DAHIYA AND RAMESH M. KORWAR

Old Dominion University and University of Massachusetts

The maximum likelihood estimators (m.l.e.) are obtained for the parameters of a bivariate normal distribution with equal variances when some of the observations are missing on one of the variables. The likelihood equation for estimating ρ, the correlation coefficient, may have multiple roots but a result proved here provides a unique root which is the m.l.e. of ρ. The problem of estimating the difference δ of the two means is also considered and it is shown that the m.l.e. of δ is unbiased.

1. Introduction. Recently, several authors have investigated the problem of estimation and testing the difference of the means in the case of incomplete samples from bivariate normal distributions. Maximum likelihood estimators (m.l.e.) for the means and covariance matrix were given by Anderson (1957) for the general case. Mehta and Gurland (1969a) consider the problem of testing the equality of the two means in the special case when the two variances are the same. Morrison (1972, 1973), and Lin and Stivers (1975) have also considered this special case and have provided different test statistics. Maximum likelihood estimation of the correlation coefficient, ρ, however, becomes complicated in this special case and all these studies have utilized simple and intuitive estimators of ρ, based only on the complete pairs of observations in their test statistic. The problem of estimating the difference of two means has been further investigated by Mehta and Gurland (1969b), Morrison (1971), Lin (1971), and Mehta and Swamy (1973). Here we obtain maximum likelihood estimators of all the parameters in this special case of equal variances. Obtaining an m.l.e. for ρ involves solving a cubic equation, but we prove that there is a unique real root of this equation which always provides the m.l.e., and this makes it easy to obtain this root numerically. The asymptotic covariance matrix of the m.l.e.'s is also obtained and it is proved that the m.l.e. for the difference of the two means is an unbiased estimator.

2. The maximum likelihood estimates. Let us consider the incomplete bivariate sample

$$\begin{array}{l} x_1, \cdots, x_n, \quad x_{n+1}, \cdots, x_N \\ y_1, \cdots, y_n \end{array}$$

Received June 1978; revised February 1979.

[1]This research was partially supported by the Air Force Office of Scientific Research, AFSC under Grant AFOSR 78-3491.

AMS 1970 *subject classifications.* Primary 62F10; secondary 62H99.

Key words and phrases. Bivariate normal distribution, difference of two means, maximum likelihood estimation, missing data, unbiased estimators, uniqueness of maximum likelihood estimators.

from a bivariate normal distribution with mean vector (μ_1, μ_2), and a covariance matrix with common variance σ^2 and correlation coefficient ρ. It may be noted that (x_i, y_i), $i = 1, \cdots, n$, are paired observations. The likelihood function can be written as

$$L(\mu_1, \mu_2, \sigma, \rho) = (2\pi\sigma^2)^{-(N+n)/2}(1-\rho^2)^{-n/2} \times \exp\left(-\frac{\Sigma_{i=1}^N (x_i-\mu_1)^2}{2\sigma^2} - \frac{1}{2\sigma^2(1-\rho^2)}\Sigma_1^n\{y_i - \mu_2 - \rho(x_i - \mu_1)\}^2\right). \tag{2.1}$$

It can be shown that the m.l.e.'s are given by the solution of the following four equations:

$$\hat{\mu}_1 = \bar{x}^*, \qquad \hat{\mu}_2 = \bar{y} - \hat{\rho}(\bar{x} - \bar{x}^*), \tag{2.2}$$

$$\hat{\rho} = S_{12}/\left[N\hat{\sigma}^2 - \left(S_1^{*2} - S_1^2\right)\right], \tag{2.3}$$

$$\hat{\sigma}^2 = \left(S_1^{*2} - S_1^2 + \frac{S_1^2 + S_2^2 - 2\hat{\rho}S_{12}}{(1-\hat{\rho}^2)}\right) \Big/ (N+n),$$

where

$$\bar{x}^* = \Sigma_1^N x_i/N, \quad \bar{x} = \Sigma_1^n x_i/n, \quad \bar{y} = \Sigma_1^n y_i/n, \quad S_1^2 = \Sigma_1^n(x_i - \bar{x})^2,$$
$$S_2^2 = \Sigma_1^n(y_i - \bar{y})^2, \quad S_1^{*2} = \Sigma_1^N(x_i - \bar{x}^*)^2, \quad \text{and} \quad S_{12} = \Sigma_1^n(x_i - \bar{x})(y_i - \bar{y}). \tag{2.4}$$

The two equations in (2.3), however, need to be solved numerically and there may be multiple roots. This makes the problem of obtaining the m.l.e.'s very complicated. On eliminating $\hat{\sigma}^2$ from the two equations in (2.3), we get a cubic equation in $\hat{\rho}$ given by:

$$f(\hat{\rho}) = n\left(S_1^{*2} - S_1^2\right)\hat{\rho}^3 - (N-n)S_{12}\hat{\rho}^2 + \left[N\left(S_1^2 + S_2^2\right) - n\left(S_1^{*2} - S_1^2\right)\right]\hat{\rho} - (N+n)S_{12} = 0. \tag{2.5}$$

This equation may have three real roots and the likelihood function will have to be evaluated at all three in order to determine the m.l.e. for ρ. This makes the computations of the m.l.e. very cumbersome. In the following theorem, we prove that among the roots of (2.5) there is exactly one root that is of the same sign as S_{12}. This unique root is the m.l.e. $\hat{\rho}$ of ρ.

THEOREM 1. *The cubic equation* (2.5) *has exactly one root in* $[-1, 1]$ *having the same sign as* S_{12}. *This real root is the unique* m.l.e. $\hat{\rho}$ *of* ρ.

PROOF. In order to prove the first part of the theorem, we find the m.l.e.'s of μ_1, μ_2, and σ for a fixed ρ and evaluate the likelihood function at these estimators, which is given by

$$h(\rho) = L(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}, \rho) = A(1-\rho^2)^{N/2} / \left((1-\rho^2)\left(S_1^{*2} - S_1^2\right) + S_1^2 + S_2^2 - 2\rho S_{12}\right)^{(N+n)/2}, \tag{2.6}$$

where A is a constant which does not involve ρ. If $S_{12} > 0$, we have

$$h(\rho) > h(-\rho), \qquad \rho > 0,$$

which implies that the maximum of $h(\rho)$ would occur at a positive ρ. A similar result is true for $S_{12} < 0$. Hence the m.l.e. of ρ will be of the same sign as S_{12}.

Now we consider the cubic equation (2.5) and prove that it has at least one root in $[-1, 1]$ which is of the same sign as S_{12}. Evaluating $f(\hat{\rho})$ at $\hat{\rho} = -1, 0, 1$, we have

$$\begin{aligned} f(-1) &= -N(S_1^2 + S_2^2 + 2S_{12}) < 0, \\ f(0) &= -(N + n)S_{12}, \\ f(1) &= N(S_1^2 + S_2^2 - 2S_{12}) > 0. \end{aligned} \tag{2.7}$$

If $S_{12} > 0$, then it is clear from (2.7) that there are an odd number of real roots of $f(\hat{\rho}) = 0$ lying in $[0, 1]$. A similar result is true for $S_{12} < 0$.

Finally, in order to prove that there is a unique real root in $[-1, 1]$ which is of the same sign as S_{12}, we consider two different cases.

CASE I. $N(S_1^2 + S_2^2) - n(S_1^{*2} - S_1^2) > 0$.

For $S_{12} > 0$, we observe from the change of signs of the coefficients in $f(-\hat{\rho})$ that, by Descartes' rule, there is no negative root of $f(\hat{\rho}) = 0$. Furthermore, from (2.5), it follows that

$$(\hat{\rho}_1 + \hat{\rho}_2 + \hat{\rho}_3)/(\hat{\rho}_1\hat{\rho}_2\hat{\rho}_3) = (N - n)/(N + n) < 1, \tag{2.8}$$

where $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3$ are the three roots of $f(\hat{\rho}) = 0$. Since there is no negative root, (2.8) rules out the possibility of more than one real root in $[0, 1]$. Hence there is a unique real root in $[0, 1]$ for $S_{12} > 0$. Similarly, it can be shown that there is a unique real root in $[-1, 0]$ for $S_{12} < 0$.

CASE II. $N(S_1^2 + S_2^2) - n(S_1^{*2} - S_1^2) < 0$.

If $S_{12} > 0$, it follows from the change of signs of coefficients in $f(\hat{\rho})$ and by Descartes' rule that there is at most one positive root of $f(\hat{\rho}) = 0$. Since we have already shown that there is at least one root which is of the same sign as S_{12}, it follows that there is a unique root of $f(\hat{\rho}) = 0$ in the interval $[0, 1]$. A similar argument holds for $S_{12} < 0$. □

It may be pointed out that Theorem 1 simplifies the computation of the m.l.e. of ρ considerably because one does not have to worry about multiple roots of the cubic equation and the likelihood does not have to be evaluated for choosing the proper m.l.e. We have tried solving the cubic by the Newton-Raphson method with $u = 2S_{12}/(S_1^2 + S_2^2)$ as the starting value and have found that this numerical procedure converges after three or four iterations for most of the examples considered. Once $\hat{\rho}$ is obtained, the m.l.e.'s of μ_2 and σ^2 can be easily computed from the expressions (2.2) and (2.3).

Let $\theta = (\mu_1, \mu_2, \rho, \sigma^2)'$ and $\hat{\theta}$ be the m.l.e. of θ. It is straightforward to obtain the information matrix $I(\theta)$ and thus derive the asymptotic covariance matrix of $\hat{\theta}$, given by

$$(2.9) \qquad I^{-1}(\theta) = \frac{\sigma^2}{n}\begin{bmatrix} \lambda & \lambda\rho & 0 & 0 \\ \lambda\rho & 1-(1-\lambda)\rho^2 & 0 & 0 \\ 0 & 0 & a_{11} & a_{12} \\ 0 & 0 & a_{21} & a_{22} \end{bmatrix},$$

where

$$\lambda = n/N, \quad a_{11} = (1-\rho^2)^3/D\sigma^2, \quad a_{12} = a_{21} = 2\rho\lambda\sigma^2(1-\rho^2)/D,$$
$$a_{22} = 2\lambda\sigma^2(1-2\rho^2-\rho^4)/D,$$

and where

$$D = 1 - 2\rho^2(1+\lambda) - \rho^4(1-2\lambda).$$

It is obvious from (2.9) that $\hat{\mu}_1$ and $\hat{\mu}_2$ are asymptotically independent of $\hat{\rho}$ and $\hat{\sigma}^2$.

3. Estimation of difference of means. The problem of estimating $\delta = \mu_1 - \mu_2$ has been investigated by Mehta and Gurland (1969b): an estimator, $Z(u)$, of δ is proposed which is given by

$$(3.1) \qquad Z(u) = \bar{x}^* - \bar{y} + u(\bar{x} - \bar{x}^*),$$

where

$$(3.2) \qquad u = 2S_{12}/\left(S_1^2 + S_2^2\right).$$

Lin (1971) considers two other estimators $Z(v)$ and $Z(w)$, where

$$(3.3) \qquad v = S_{12}/S_1^2, \quad \text{and} \quad w = S_{12}/S_2^2.$$

It may be pointed out that u, v, and w are estimators of ρ. Furthermore, $Z(v)$ is the m.l.e. of δ in the general case when the variances of the two populations are not assumed to be the same.

In the special case of equal variances, the m.l.e. of δ is given by:

$$(3.4) \qquad \hat{\delta} = \bar{x}^* - \bar{y} + \hat{\rho}(\bar{x} - \bar{x}^*).$$

Since all the investigations, loc. cit., regarding the tests for $\delta = 0$ assume equal variances of the two populations, it would be of interest to utilize $\hat{\delta}$ as the test statistic. Morrison (1973) has obtained a test statistic, for testing $\delta = 0$, on replacing ρ by u in the likelihood ratio test statistic derived for a fixed ρ. Since u makes use of only complete pairs, the estimator $\hat{\rho}$ of ρ could be utilized instead in this test statistic.

Now we prove that $\hat{\delta}$ is an unbiased estimator of δ. It may be pointed out that $\bar{x}^*$ is not distributed independent of $\hat{\rho}$ and hence the direct conditional expectation of $\hat{\delta}$ for a given $\hat{\rho}$ will not work here in proving this result.

THEOREM 2. *The estimator $\hat{\delta}$ is unbiased for δ.*

PROOF. We have

$$E(\hat{\delta}) = E(\bar{x}^*) - E(\bar{y}) + E\hat{\rho}(\bar{x} - \bar{x}^*)$$
$$= \delta + E\hat{\rho}(\bar{x} - \bar{x}^*),$$

and we prove below that

$$E\hat{\rho}(\bar{x} - \bar{x}^*) = 0.$$

Now $(\bar{x}^*, \bar{x}, \bar{y})$ is independent of $(S_1^2, S_2^2, S_{12}, S_1^{**2})$, where

$$\bar{x}^{**} = \Sigma_{n+1}^N x_i / (N - n), \quad S_1^{**2} = \Sigma_{n+1}^N (x_i - \bar{x}^{**})^2.$$

Also note that

$$S_1^{*2} = S_1^2 + S_1^{**2} + n(\bar{x} - \bar{x}^*)^2 + (N - n)(\bar{x}^{**} - \bar{x}^*)^2$$

and

$$(\bar{x} - \bar{x}^*)^2 = (N - n)^2(\bar{x} - \bar{x}^{**})^2 / N^2,$$
$$(\bar{x}^{**} - \bar{x}^*)^2 = n^2(\bar{x} - \bar{x}^{**})^2 / N^2.$$

Thus

$$E[\hat{\rho}(\bar{x} - \bar{x}^*)] = ((N - n)/N)E[\hat{\rho}(\bar{x} - \bar{x}^{**})]$$
$$= ((N - n)/N)E\left[E\left\{\hat{\rho}(\bar{x} - \bar{x}^{**})|S_1^2, S_2^2, S_{12}, S_1^{**2}, (\bar{x} - \bar{x}^{**})^2\right\}\right]$$
$$= ((N - n)/N)E\left[\hat{\rho}E\left\{(\bar{x} - \bar{x}^{**})|(\bar{x} - \bar{x}^{**})^2\right\}\right],$$

where we made use of the fact that

(i) $\hat{\rho}$ is a function of S_1^2, S_2^2, S_{12}, S_1^{**2}, and $(\bar{x} - \bar{x}^{**})^2$,

(ii) $\bar{x} - \bar{x}^{**}$ is independent of $(S_1^2, S_2^2, S_{12}, S_1^{**2})$.

Hence

$$E\hat{\rho}(\bar{x} - \bar{x}^*) = \frac{N - n}{N} E[\hat{\rho}E(Z||Z|)],$$

where Z is a normal random variable with mean zero. But $E(Z||Z|) = 0$. Hence the result. ▯

REFERENCES

[1] ANDERSON, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *J. Amer. Statist. Assoc.* **52** 200–204.

[2] LIN, PI-ERH (1971). Estimation procedures for difference of means with missing data. *J. Amer. Statist. Assoc.* **66** 634–636.

[3] LIN, PI-ERH and STIVERS, L. E. (1975). Testing for equality of means with incomplete data on one variable: A Monte Carlo study. *J. Amer. Statist. Assoc.* **70** 190–193.

[4] MEHTA, J. S. and GURLAND, J. (1969a). Testing equality of means in the presence of correlation. *Biometrika* **56** 119–126.

[5] MEHTA, J. S. and GURLAND, J. (1969b). Some properties and an application of a statistic arising in testing correlation. *Ann. Math. Statist.* **40** 1736–1745.

[6] MEHTA, J. S. and GURLAND, J. (1973). A test for equality of means in the presence of correlation and missing values. *Biometrika* **60** 211–213.

[7] Mehta, J. S. and Swamy, P. A. V. B. (1973). Bayesian analysis of a bivariate normal distribution with incomplete observations. *J. Amer. Statist. Assoc.* **68** 922–927.

[8] Morrison, D. F. (1972). The analysis of a single sample of repeated measurements. *Biometrics* **28** 55–71.

[9] Morrison, D. F. (1973). A test for equality of means of correlated variates with missing data of one response. *Biometrika* **60** 101–105.

Department of Mathematical Science
Old Dominion University
Norfolk, Virginia 23508

Mathematics and Statistics
University of Massachusetts
Amherst, Massachusetts 01003