Vol.4 No.2
Spring 1993

A BAYESIAN APPROACH TO COMPUTING MISSING REGRESSOR VALUES

A.H.W. Yousef, M.S. Abu-Salih and M.A. Ali

Yarmouk University, Irbid, Jordan

Abstract

In this article, Lindley's measure of average information is used to measure the information contained in incomplete observations on the vector of unknown regression coefficients [9]. This measure of information may be used to compute the missing regressor values.

Introduction

A problem frequently occurring in statistical practice is that of dealing with multiple regression when some of the values of the explanatory variables are missing. Procedures frequently used to deal with this problem [see 1,7] are: (i) complete case method, (ii) zero-order regression method and (iii) first-order regression method. A brief description of these approaches may be found in Donner and Rosner [4] and Hill and Ziemer [7].

Properties of these procedures have been compared extensively, mainly through simulation studies by Beale and Little [2] and Donner and Rosner [4]. Haitovsky [5], and Heiberger [6] have found that for multinormal distributions the method of maximum likelihood is preferable to the complete case method. Donner [4] studied the relative effectiveness of these methods by considering a linear regression model having two regressor variables. He concluded that the zero-order regression method is relatively effective for estimating the coefficients of incompletely observed variables when the correlations involving these variables are weak and the proportion of missing observations is fairly high. The first-order regression method is most effective for estimating the coefficient of the completely observed variable. Hill and Ziemer [7] investigated the performance of some common procedures for replacing missing regressor values under varying conditions of multicollinearity. Their analytical and numerical results indicate that the zero-order regression method is preferable to other procedures given ill-conditioned designs. In addition, they stated that incomplete sample observations should not be thrown away under conditions of extreme multicollinearity. Afifi and Elashoff [1] suggested that the parameters of a regression model be estimated using all the data of a sample.

Shannon [11] derived an expression that quantifies the amount of information produced by a source. Shannon's work motivated a number of information measures in statistics, [8,9,12]. In this article, Lindley's measure of average information is used to measure the information contained in the incomplete observations [9]. This measure of information can be used to compute the missing regressor values.

Keywords: Information; Missing regressor values; Bayes' approach

Results and Discussion

Loss of Information Due to Incomplete Observations

We consider a two-stage normal regression model

$$f(Y|\beta) \sim N(X\beta, \sigma^2 I_n), \text{ and } h_1(\beta) \sim N(\mu, \sigma^2 I_p), \quad (2.1)$$

where $Y(nx1)$ is a vector of observations, $X(nxp)$ is a known matrix with rank $(x) = p<n$, $\beta(px1)$ is a vector of unknown regression coefficients, σ^2 is an unknown positive constant, I_n is the nxn identity matrix, $f(Y|\beta)$ is the conditional distribution of Y given β, $h_1(\beta)$ denotes the prior distribution of β, $\mu(px1)$ is a vector of unknown values and I_p is the (pxp) identity matrix.

Lindley's [9] average amount of information in Y about β with the prior $h_1(\beta)$ and posterior $h_2(\beta|y)$ is defined by:

$$J(Y;\beta) = E_Y [I(h_2(\beta|y)) - I(h_1(\beta))] \quad (2.2)$$

where $I(h) = \int h(z) \ln h(z)\, dz$ for all z such that $h(z) > 0$. Alternatively, (2.2) may be written as:

$$\begin{aligned} J(Y;B) &= \iint f(y,\beta) \ln[f(y,\beta)/g(s))\, h_1(\beta)]\, dy\, d\beta \\ &= \iint f(y|\beta)\, h_1(\beta) \ln [f(y|\beta)]\, dy\, d\beta \\ &\quad - \iint f(y|\beta)\, h_1(\beta) \ln[g(y)]\, dy\, d\beta \\ &= E_{h_1} E_f [\ln f(y|\beta)] - E_{h_1} E_f [\ln g(y)] \end{aligned} \tag{2.3}$$

where $f(y,\beta) = f(y|\beta)\, h_1(\beta)$, $f(y|\beta)$ is the conditional distribution of Y given β and $g(y)$ is the marginal distribution of Y. E_{h_1} and E_f denote the expectation operators with respect to $h_1(\beta)$ and $f(y|\beta)$, respectively. Observe that for our setup in (2.1) we have

$$g(y) \sim N(X\mu, \sigma^2 V),\ V = (In + XX') \tag{2.4}$$

We partition Y and X as follows:

$$Y' = (Y_1' : Y_2') \text{ and } X' = (X_c' : X_o') \tag{2.5}$$

where $X_c((n-t)xp)$ is the design matrix, $Y_1((n-t)x1)$ is the vector of observations corresponding to the complete observations and $X_o(txp)$ represents the rows in X corresponding to the vector of incomplete observations $Y_2(tx1)$.

Theorem 1.

If $f_1(y_1|\beta) \sim N(X_c\beta, \sigma^2 I_{(n-t)})$, $f_2(y_2|\beta) \sim N(X_o\beta, \sigma^2 I_t)$, and $h_1(\beta) \sim N(\mu, \sigma^2 I_p)$ then $g_1(y_1)$ the marginal distribution of Y_1 and $g_2(y_2)$ the marginal distribution of Y_2 are also multivariate normal:

i. $g_1(y_1) \sim N(X_c\mu, \sigma^2 V_1)$.

ii. $g_2(y_2) \sim N(X_o\mu, \sigma^2 V_2)$, $h_2(\beta|y_2) \sim N(\mu_*, \sigma^2 S)$

$$\text{where } S = (I_p + X_o'X_o)^{-1},\ \mu_* = S(\mu + X_o'y_2) \tag{2.6}$$

$$V_1 = (I_{(n-t)} + X_cX_c') \text{ and } V_2 = (I_t + X_oX_o') \tag{2.7}$$

See Lindley [9 p. 114].

Let $f_1(y_1|\beta)$ and $g_1(y_1)$ denote the conditional distribution of Y_1 given β and the marginal distribution of Y_1, respectively. Then the mean amount of information in Y_1 about β is given by:

$$J(Y_1;\beta) = E_{h_1}E_{f_1} [\ln f_1(y_1|\beta)] - E_{h_1}E_{f_1} [\ln g_1(y_1)] \tag{2.8}$$

where E_{f_1} denotes the expectation operator with respect to $f_1(y_1|\beta)$. We define a measure of loss of information due to the incomplete vector of observations Y_2 by

$$L(X_o;X_c) = J(Y;\beta) - J(Y_1;\beta), \tag{2.9}$$

where $J(Y;\beta)$ and $J(Y_1;\beta)$ are given in (2.3) and (2.8), respectively. Now, the problem in hand is to evaluate $L(X_o;X_c)$.

The following results are needed to compute $J(Y;\beta)$ $J(Y_1;\beta)$ and $J(Y_2;\beta)$.

Lemma 1.

Let $f(y|\beta) \sim N(X\beta, \sigma^2 I_n)$

$h_1(\beta) \sim N(\beta, \sigma^2 I_p)$, $\quad g(y) \sim N(X\beta, \sigma^2 V)$,

$f_1(y_1|\beta) \sim N(X_c\beta, \sigma^2 I_{(n-t)})$, $\quad g_1(y_1) \sim N(X_c\mu, \sigma^2 V_1)$

$f_2(y_2|\beta) \sim N(X_o\beta, \sigma^2 I_t)$, $\quad g_2(y_2) \sim N(X_o\mu, \sigma^2 V_2)$

Then

i. $E_{h_1}E_f [\ln f(y|\beta)] = -\frac{n}{2} [\ln 2\pi + \ln \sigma^2 + 1]$

ii. $E_{h_1}E_f [\ln g(y)] = -\frac{n}{2} [\ln 2\pi + \ln \sigma^2 + \frac{1}{n} \ln|V| + 1]$

iii. $E_{h_1}E_{f_1} [\ln f_1(y_1|\beta)] = -\frac{n-t}{2} [\ln 2\pi + \ln \sigma^2 + 1]$

iv. $E_{h_1}E_{f_1} [\ln g_1(y_1)] = -\frac{n-t}{2} [\ln 2\pi + \ln \sigma^2 + \{\frac{1}{n-t} \ln|V_1|\} + 1]$

v. $E_{h_1}E_{f_2} [\ln f_2(y_2|\beta)] = -\frac{t}{2} [\ln 2\pi + \ln \sigma^2 + 1]$

vi. $E_{h_1}E_{f_2} [\ln g_2(y_2)] = -\frac{t}{2} [\ln 2\pi + \ln \sigma^2 + \frac{1}{t} \ln|V_2| + 1]$

where V, V_1 and V_2 are given in (2.4) and (2.7), respectively.

Proof.

We observe that

$$\ln f(y|\beta) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (y-X\beta)' I_n (y-X\beta)$$

Therefore

$$E_f[\ln f(y|\beta)] = -\frac{n}{2} [\ln 2\pi + \ln \sigma^2 + \frac{n\sigma^2}{n\sigma^2}]$$

Taking expectation over $h_1(\beta)$ we obtain

$$E_{h_1}E_f[\ln f(y|\beta)] = -\frac{n}{2} [\ln 2\pi + \ln \sigma^2 + 1] \tag{2.10}$$

This completes the proof of part (i).

Next we have

$$\begin{aligned} \ln g(y) &= -\frac{n}{2} [\ln 2\pi + \ln \sigma^2 + \frac{1}{n} \ln |V|] \\ &\quad - \frac{1}{2\sigma^2} [(y-X\mu)' V^{-1} (y-X\mu)] \\ &= -\frac{n}{2} [\ln 2\pi + \ln \sigma^2 + \frac{1}{n} \ln |V|] \\ &\quad - \frac{1}{2\sigma^2} [\{(y-X\beta) + X(\beta-\mu)\}' V^{-1} \{(y-X\beta) + X(\beta-\mu)\}] \\ &= -\frac{n}{2} [\ln 2\pi + \ln \sigma^2 + \frac{1}{n} \ln |V|] \\ &\quad - \frac{1}{2\sigma^2} (y-X\beta)' V^{-1} (y-X\beta) - \frac{2}{2\sigma^2} (y-X\beta)' V^{-1} X(\beta-\mu) \end{aligned}$$

$$- \frac{2}{2\sigma^2} (\beta-\mu)'X'V^{-1} X (\beta-\mu).$$

Hence

$$E_f[\ln g(y)] = -\frac{n}{2}[\ln 2\pi + \ln\sigma^2 + \frac{1}{n}\ln|V|]$$

$$-\frac{1}{2\sigma^2} trV^{-1}\sigma^2 I_n - \frac{2}{2\sigma^2}(\beta-\mu)'X'V^{-1}X(\beta-\mu),$$

where tr(A) denotes the trace of a square matrix A. Taking expectation over $h_1(\beta)$ we obtain

$$E_{h_1}E_f[\ln g(y)] = -\frac{n}{2}[\ln 2\pi + \ln\sigma^2 + \frac{1}{n}\ln|V|] - \frac{1}{n}trV^{-1}$$

$$-\frac{1}{2\sigma^2} trXV^{-1}X\sigma^2 I_p$$

$$= -\frac{n}{2}[\ln 2\pi + \ln\sigma^2 + \frac{1}{n}\ln|V|] - \frac{1}{n}tr(V^{-1} + V^{-1}XX')$$

$$= -\frac{n}{2}[\ln 2\pi + \ln\sigma^2 + \frac{1}{n}\ln|V| + 1]$$

This completes the proof of part (ii) of the above lemma.

The proofs of part (iii) and part (v) are similar to the proof of part (i) while the proofs of part (iv) and part (vi) are similar to the proof of part (ii). Hence the proofs of these parts are omitted.

Now, under the hypotheses of Lemma 1, the amount of loss of information due to Y_2, the vector of incomplete observations, may be measured by

$$L(X_o, X_c) = J(Y;\beta) - J(Y_1;\beta). \quad (2.10)$$

Now,

$$J(Y;\beta) = E_{h_1}E_f[\ln f(y|\beta)] - E_{h_1}E_f[\ln g(y)]$$

$$= \frac{1}{2}\ln|V| \quad (2.11)$$

and

$$J(Y_1;\beta) = E_{h_1}E_{f_1}[\ln f(y_1|\beta)] - E_{h_1}E_f[\ln g_1(y_1)]$$

$$= \frac{1}{2}\ln|V_1| \quad (2.12)$$

Next, in accordance with the partition of X in (2.5) we partition V as follows:

$$V = \begin{bmatrix} V_1 & X_c'X_o' \\ X_o'X_c & V_2 \end{bmatrix} \quad (2.13)$$

where V_1 and V_2 are given in (2.7).

From (2.13) we obtain

$$|V| = |V_1|\,|V_2 - X_o'X_c'V_1^{-1}X_c'X_o|$$

$$= |V_1|\,\left|\left[I_t + X_o'X_o' - X_o'X_c'V_1^{-1}X_c'X_o'\right]\right|$$

$$= |V_1|\,\left|I_t + X_o'(I_p - X_c'V_1^{-1}X_c')X_o'\right| \quad (2.14)$$

Using (2.14) in (2.11) we get

$$J(Y;\beta) = \frac{1}{2}\ln|V_1| + \frac{1}{2}\ln\left|I_t + X_o(I_p - X_c'V_1^{-1}X_c)X_o'\right|$$

$$= \frac{1}{2}\ln|V_1| + \frac{1}{2}\ln\left|I_t + X_o'S_1X_o'\right| \quad (2.15)$$

where $S_1' = I_p - X_c'V_1^{-1}X_c'$

Next, from (2.10), (2.12), (2.7) and (2.15) we obtain

$$L(X_o' \vdots X_c') = \frac{1}{2}\ln|V_1| + \frac{1}{2}\ln\left|I_t + X_o'S_1'X_o'\right| - \frac{1}{2}\ln|V_1|$$

$$= \frac{1}{2}\ln\left|I_t + X_o'S_1'X_o'\right|$$

Information Contained in Incomplete Observations

In this section, we want to compute the average amount of information contained in Y_2, the vector of incomplete observations. The average amount of information contained in Y_2 on β is given by

$$J(Y_2;\beta) = E_{h_1}E_{f_2}[\ln f_2(Y_2|\beta)] - E_{h_1}E_{f_2}[\ln g_2(Y_2)]$$

From parts (v) and (vi) of Lemma 1 we obtain

$$J(Y_2;\beta) = -\frac{t}{2}[\ln 2\pi + \ln\sigma^2 + 1] + \frac{t}{2}\left[\ln 2\pi + \ln\sigma^2 + \frac{1}{t}\ln|V_2| + 1\right]$$

$$= \frac{1}{2}\ln|V_2| = \frac{1}{2}\ln\left|I_t + X_o'X_o'\right|.$$

To compute the missing regressor values the experimenter may maximize $J(Y_2;\beta)$ for the choice of X_O. He may impose some restriction on the columns or rows of X_O.

References

1. Afifi, A.A. and Elashoff, R.M. Missing Observations in Multivariate Statistics I. Review of the Literature. *J. Amer. Statist. Assoc.* **61**, 595-604, (1966).
2. Beale, E.M.L. and Little, R.J.A. Missing Values in Multivariate Analysis. *J. Royal Statist. Soc.* Series B, **37**, 129-145, (1975).
3. Donner, A. The Relative Effectiveness of Procedures Commonly Used in Multiple Regression Analysis for Dealing with Missing Values. *The American Statistician* **36**, 378-381, (1982).
4. Donner, A. and Rosner, B. Missing Value Problems in Multiple Linear Regression with Two Independent Variables. *Commun. Statist. Theor. Meth.* **11**, 127-140, (1982).
5. Haitovsky, Y. Missing Data in Regression Analysis. *J. Royal Statist. Soc.* Series B, **30**, 60-82, (1968).
6. Heiberger, R.M. Regression with the Pairwise Present Covariance Matrix: A Dangerous Practice Technical Report No. 19, Department of Statistics, University of Pennsylvania. (1977).
7. Hill, R.C. and Ziemer, R.F. Missing Regressor Values Under Conditions of Multicollinearity. *Commun Statist. Theor. Meth.* 12(22), 2557-3573, (1983).
8. Kullback, S. and Leibler, R.A. On Information Sufficiency, *Ann. Math. Statist.* **22**, 79-86, (1951).
9. Lindley, D.V. On a Measure of Information Provided by an

Experiment. *Ann. Math. Statist.* **27**, 986-1005, (1956).
10. Lindley, D.V. Introduction to Probability and Statistics, Part 2, Inference, Cambridge University Press, London, (1965).
11. Shannon, C.E. A Mathematical Theory of Communications, *Bell Systems Techn. J.* **27**, 379-423, (1948).
12. Yousef, A.W. Analysis of Regression Data with Missing Regressor Values. Unpublised Master Thesis, Dept. of Statistics, Yarmouk University, (1990).