In this paper we estimate the population mean \( \bar{Y} \) from a simple random sample \( (y_1, y_2, \ldots, y_n) \) of size \( n \). Suppose that the variable \( y_i \) is related to an auxiliary variable, say \( x_i \), where \( \bar{X} \) is known. It is known that the linear regression estimator \( \hat{Y}_n = \bar{Y} + b(\bar{X} - \bar{x}) \) is unbiased provided that the target population is essentially infinite and the paired observations \( (x_i, y_i) \) follow the usual linear model.

Situations arise, however, in which while the assumption of an essentially infinite population is plausible, the model linking the variables \( x_i \) and \( y_j \) is of a quadratic form. Hence we propose the parabolic model

\[
Y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + e_i, \quad i = 1, 2, \ldots, n \tag{1.1}
\]

Our assumptions are as follows: (a) The underlying population is very large, i.e., essentially infinite; (b) \( \bar{X} \) is known; (c) \( \beta_1, \beta_2, \beta_3 \) are unknown parameters; (d) \( e_1, e_2, \ldots, e_n \) are independent random variables and conditional expectations \( \mathbb{E}(e_i \mid x_i) = \mathbb{E}(e_i) = 0, \mathbb{E}(e_i x_i) = 0 \) for \( i = 1, 2, \ldots, n \); and finally (e) \( x_1, \ldots, x_n \) are observable without error. Taking expectations of the model in (1.1) over the entire population, we note that

\[
\bar{Y} = \beta_1 + \beta_2 \bar{X} + \beta_3 (\bar{X}^2 + s_x^2), \tag{1.2}
\]

Using the method of least square estimation, one can find unbiased estimates for \( \beta_i \); however since population data on \( x \) are not generally available, \( \mathbb{E}(x^2) \) has to be estimated. Therefore, the parabolic estimator \( \hat{Y}_p \) of \( \bar{Y} \) is proposed as

\[
\hat{Y}_p = \hat{\beta}_1 + \hat{\beta}_2 \bar{x} + \hat{\beta}_3 \left( \bar{x}^2 + s_x^2 \right), \tag{1.3}
\]

where \( s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2, \left( \bar{x}^2 + s_x^2 \right) \) is an unbiased estimator of \( \mathbb{E}(x^2) \), and \( \hat{\beta}_1 \) is the least square estimator of \( \beta_1 \) for \( i = 1, 2, 3 \). The purpose of this paper is to show that \( \hat{Y}_p \) is an unbiased estimator of the population mean \( \bar{Y} \); to compute the exact variance \( V(\hat{Y}_p) \) of the parabolic estimator; and to determine an unbiased estimator \( v(\hat{Y}_p) \) for \( V(\hat{Y}_p) \).

Let

\[
A = \text{col} (y_1, \ldots, y_n), \quad B = \text{col} (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3), \quad F = \text{col} (e_1, \ldots, e_n), \quad \text{and}
\]

\[
P = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}.
\]

Then the parabolic model (1.1) becomes

\[
A = PB + F, \quad \text{or} \quad F = A - PB. \tag{2.1}
\]

Since \( \sum_{i=1}^{n} e_i^2 = \mathbb{E}(F) = (A - PB)'(A - PB) \), we are then to determine a vector \( \hat{B} = \text{col} (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) \) that minimizes \( \sum_{i=1}^{n} e_i^2 \). Using the least square methods in regression analysis [2], we find that a solution of \( \hat{B} \) is given by \( \hat{B} = (P'P)^{-1}P'A \) and the conditional expectation of \( \hat{B} \), given \( x_1 \), is

\[
E_1(\hat{B}) = B. \quad \text{Since the conditional expectation} \quad E_1(A) \text{ of } A \text{ is } E_1(A) = PB, \text{ hence the unconditional} \quad E(B) = E(E_1(\hat{B})) = B. \text{ Therefore } \hat{B} = (P'P)^{-1}P'A \text{ is an unbiased estimator of } B.
\]

### 3. EXPECTATION AND VARIANCE OF \( \hat{Y}_p \)

Consider the expression (1.3) for \( \hat{Y}_p \). Taking conditional expectations of both sides, given \( x_1 \), we get

\[
E_1(\hat{Y}_p) = \beta_1 + \beta_2 \bar{x} + \beta_3 (\bar{x}^2 + s_x^2).
\]

Therefore

\[
E(\hat{Y}_p) = E[E_1(\hat{Y}_p)] = \beta_1 + \beta_2 \bar{x} + \beta_3 (\bar{x}^2 + s_x^2) = \bar{Y},
\]

which shows that \( \hat{Y}_p \) is unbiased for \( \bar{Y} \).

In order to compute the exact variance \( V(\hat{Y}_p) \), we first note that the conditional dispersion

\[
V(\hat{Y}_p) = E[V(\hat{Y}_p)] = E[V_1(\hat{Y}_p)] = \bar{Y} - \beta_1 - \beta_2 \bar{x} - \beta_3 (\bar{x}^2 + s_x^2).
\]

394
matrix $D_1$ of $\hat{\beta}$, given $x_1$, is obtained as follows.

$$D_1(P'PB) = D_1(P'P)$$

$$= P'D_1(A)P = P'\sigma_e^2 IP.$$ 

Then

$$(P'P)D_1(\hat{\beta}) = \sigma_e^2 I$$

and

$$D_1(\hat{\beta}) = \sigma_e^2 (P'P)^{-1}.$$ 

If we let $L = \text{col}(1, \bar{x}, \bar{x}^2 + s^2_x)$, we get

$$\bar{y}_p = L'\hat{\beta} = \hat{\beta}'L$$

and

$$V_1(L'\hat{\beta}) = L'D_1(\hat{\beta})L$$

$$= L'\sigma_e^2 (P'P)^{-1}L.$$ 

Hence

$$V_1(\bar{y}_p) = \sigma_e^2 L' (P'P)^{-1}L.$$ 

Finally, the unconditional variance becomes

$$V(\bar{y}_p) = E[V_1(\bar{y}_p)] + V[E[\bar{y}_p]]$$

$$= \sigma_e^2 E_x[L'(P'P)^{-1}L] + V(\bar{\beta}_1 + \bar{\beta}_2 \bar{x} + \beta_3 \bar{x}^2 + \beta_3 s^2_x).$$ 

Hence

$$V(\bar{y}_p) = \sigma_e^2 E_x[L'(P'P)^{-1}L]$$

$$+ \beta_3^2 V(s^2_x).$$ 

4. ESTIMATION OF $V(\bar{y}_p)$

Let

$$R_0^2 = (A - P\hat{\beta})' (A - P\hat{\beta}),$$

which reduces to

$$R_0^2 = A'[I - P(P'P)^{-1}P']A,$$

and it can be shown [2] that

$$E_1(R_0^2) = (n-3)\sigma_e^2.$$ 

Moreover,

$$E(R_0^2) = (n-3)\sigma_e^2.$$ 

Therefore, an unbiased estimator $\sigma_e^2$ of $\sigma_e^2$ is

$$\sigma_e^2 = \frac{R_0^2}{(n-3)}.$$ 

Next, we let

$$\beta_3^2 = \beta_3^2 - v(\beta_3^2),$$

where $v(\beta_3)$ represents an unbiased estimator of $V_1(\beta_3^2)$. Since $D_1(\hat{\beta}) = \sigma_e^2 (P'P)^{-1}$ we have

$$\hat{\beta}_3^2 = \frac{V_1(\beta_3^2)}{\sigma_e^2} \approx \sigma_e^2 [c_{11}].$$

Then the element $v_{33}$ in the $(3,3)$ position of $D_1(\hat{\beta})$ is

$$v_{33} = \frac{\sigma_e^2 c_{33}}{\sigma_e^2} = \frac{c_{33} R_0^2}{(n-3)}.$$ 

We now note that

$$\beta_3^2 = \beta_3^2 - c_{33} R_0^2/(n-3)$$

and

$$E_1(\beta_3^2) = E_1(\beta_3^2) - \sigma_e^2 c_{33}$$

$$= E_1(\beta_3^2) - V_1(\beta_3^2)$$

$$= [E_1(\beta_3^2)]^2 - \beta_3^2,$$

which establishes that

$$\beta_3^2 = \beta_3^2 - c_{33} R_0^2/(n-3)$$

is an unbiased estimator of $\beta_3^2$.

To find an unbiased estimator of $V(s^2_x)$, we proceed as follows. Since

$$s^2_x = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \frac{n}{n-1} \bar{m}_2,$$

where $\bar{m}_2$ is the second moment about the sample mean, we can write the variance of $s^2_x$ in terms of the cumulants of $x$, [3].

$$V(s^2_x) = \frac{n}{n-1} \sum_{i=1}^{4} \frac{1}{n-1} V(m_{2i})$$

$$= \frac{n}{n-1} \kappa_4 + \frac{2\kappa_2^2}{n-1},$$

(4.3)

where $\kappa_r$ is the rth cumulant of $x$. Using the theory of symmetric functions [3], we recall that the $k$-statistic of order $k_r$ which is a function of the observations, is an unbiased estimator of $\kappa_r$ for all integers $r$. Moreover, the $i$-statistic (or polykay) $\bar{s}_{22}$ has the property $E(\bar{s}_{22}) = \kappa_2 \cdot \kappa_2 = \kappa_2^2$, where $\bar{s}_{22}$ is also a function of observations only. Hence, if we let

$$v(s^2_x) = \frac{k_4}{n} + 2\bar{s}_{22}/(n-1),$$

(4.4)
we get, from the tables of \( k \)-statistics,

\[
\hat{V}(s_x^2) = \frac{1}{n+1} \left[ \frac{n-1}{n} k_4 + 2k_2^2 \right],
\]

(4.5)

where \( k_2 = f(t_1, t_2) \), \( k_4 = g(t_1, t_2, t_3, t_4) \)

and \( t_j = \frac{\sum x_j}{n} \).

Finally, it follows that

\[
\hat{V}(\hat{y}_p^2) = \frac{1}{n-3} \left[ R_0^2 \left( (p^p)^{-1}L \right) \right. \\
+ \left( \frac{2}{n-3} c_{33} \right) \hat{V}(s_x^2)
\]

(4.6)

is an unbiased estimator of \( V(\hat{y}_p^2) \) determined by the observed sample values \( y_1, \ldots, y_n \) and \( x_1, \ldots, x_n \).

REFERENCES


*This problem was initially suggested by the late H.O. Hartly.