REGRESSION ESTIMATION FOR BIVARIATE NORMAL DISTRIBUTIONS

CHIEN-PAI HAN

(Received Feb. 22, 1971)

Summary

In estimating the mean μ_y of one variable in a bivariate normal distribution, the experimenter can use the other variable, x, as an auxiliary variable to increase precision. In particular, if μ_x is known, he can use the regression estimator. When μ_x is unknown, a preliminary test can be performed and the estimator will be made to depend on the result of the preliminary test. The bias and mean square error of the preliminary test estimator are obtained and the relative efficiency is are discussed.

1. Introduction

In estimating the population mean of a variable y, the precision of the estimator may be increased by the use of an auxiliary variable x which is correlated with y. When the relationship between y and x is a straight line, a linear regression estimator can be constructed. We assume that x and y have a bivariate normal distribution with means μ_x and μ_y, variances σ_x^2 and σ_y^2 respectively and correlation coefficient ρ. Suppose a random sample of size n is taken and a regression estimator of μ_y is used; it is well known that the variance of the regression estimator is $(1-\rho^2)\sigma_y^2/n$. When ρ is large, there is a substantial reduction in variance. The construction of a linear regression estimator of μ_y requires the knowledge of μ_x. In case μ_x is unknown, we may perform a preliminary test for the hypothesis $H_0 : \mu_x=\mu_0$, where μ_0 is some constant that the investigator believes that the population mean μ_x should be. If H_0 is accepted, we use the regression estimator; otherwise, we use the ordinary estimator $\bar{y}$. We shall call this estimator the preliminary test estimator. This can arise in the following situation. The experimenter wishes to estimate the mean yield per acre of a certain crop. It is known that yield is highly correlated with the moisture content in soil. Hence the moisture content can be used as an auxiliary variable. The experimenter usually does not know the population mean value of the moisture content; but from the amount of rainfall reported by weather bureau (or other sources of information such as underground water, manual watering, type of soil, etc.), he believes that the population mean should be μ_0. After the sample is obtained, the experimenter can perform a preliminary test of the hypothesis that the population mean of the auxiliary variable is μ_0. If the hypothesis is not significant, the regression estimator is used; otherwise he simply uses the ordinary sample mean. Another example is given in Dempster ([3], p. 327). Although Example 8.1 in Dempster's book considers a trivariate normal distribution and interval estimation, the idea of utilizing regression estimator is the same as in this paper. Papers in the area of preliminary test estimation include Bancroft [1], Bennett [2], Han and Bancroft [4], Kale and Bancroft [5], Kitagawa [6], Mosteller [7], and others.

This paper will derive the bias and mean square error (MSE) of the preliminary test estimator when the covariance matrix is known or unknown. Relative efficiency of the preliminary test estimator to the usual estimator $\bar{y}$ is studied and tables are computed for selecting the level of significance of the preliminary test.

2. Bias and mean square error when the covariance matrix is known

Suppose (x_i, y_i), $i=1, 2, \cdots, n$ is a random sample from the bivariate normal distribution with means μ_x, μ_y, variances σ_x^2, σ_y^2 and correlation coefficient ρ. In this section we assume that μ_x and μ_y are unknown; σ_x^2, σ_y^2 and ρ are known. Without loss of generality we may let $\sigma_x^2=\sigma_y^2=1$. If μ_x is known, the regression estimator is $\bar{y}+\rho(\mu_x-\bar{x})$ where $\bar{x}$ and $\bar{y}$ are the sample means. The variance of the estimator is reduced to $(1-\rho^2)/n$. When μ_x is unknown, the estimator of μ_y is made to depend on the outcome of the preliminary test for $H_0 : \mu_x=0$ (here without loss of generality we let $\mu_0=0$). Hence

$$\bar{y}^* = \begin{cases} \bar{y}-\rho\bar{x} & \text{if } |\sqrt{n}\,\bar{x}|\leqq z_\alpha \\ \bar{y} & \text{if } |\sqrt{n}\,\bar{x}|> z_\alpha , \end{cases} \tag{2.1}$$

where $\bar{x}$ and $\bar{y}$ are the sample means and z_α is the $100(1-\alpha/2)\%$ point of the standard normal distribution.

The expectation of $\bar{y}^*$ is

$$\begin{aligned} \mathrm{E}(\bar{y}^*) &= \mathrm{E}\{\bar{y}-\rho\bar{x}/|\sqrt{n}\,\bar{x}|\leqq z_\alpha\}\,\mathrm{P}\{|\sqrt{n}\,\bar{x}|\leqq z_\alpha\} \\ &\quad +\mathrm{E}\{\bar{y}/|\sqrt{n}\,\bar{x}|> z_\alpha\}\,\mathrm{P}\{|\sqrt{n}\,\bar{x}|> z_\alpha\} \\ &= \mathrm{E}(\bar{y})-\rho\mathrm{E}\{\bar{x}/|\sqrt{n}\,\bar{x}|\leqq z_\alpha\}\,\mathrm{P}\{|\sqrt{n}\,\bar{x}|\leqq z_\alpha\}\,. \end{aligned} \tag{2.2}$$

Since $\mathrm{E}(\bar{y})=\mu_y$, therefore the second term is the bias of $\bar{y}^*$. Denoting the bias by B and using the fact that $\bar{x}$ is distributed as $N(\mu_x, 1/n)$, we obtain

$$(2.3)\quad \begin{aligned} B &= -\rho\, \mathrm{E}\{\bar{x}/|\sqrt{n}\,\bar{x}|\leqq z_\alpha\}\, \mathrm{P}\{|\sqrt{n}\,\bar{x}|\leqq z_\alpha\} \\ &= \int_{-z_\alpha/\sqrt{n}}^{z_\alpha/\sqrt{n}} \frac{-\sqrt{n}\,\bar{x}\rho}{\sqrt{2\pi}} \exp\left\{-\frac{n}{2}(\bar{x}-\mu_x)^2\right\} d\bar{x} \\ &= \frac{\sqrt{2}\rho}{\sqrt{\pi n}} \exp\{-(z_\alpha^2+a^2)/2\} \sinh(az_\alpha) - \frac{\rho a}{\sqrt{n}}\{\Phi(z_\alpha-a)-\Phi(-z_\alpha-a)\}\ , \end{aligned}$$

where $a=\sqrt{n}\mu_x$ and $\Phi(x)$ is the cumulative distribution function of $N(0, 1)$. When $\alpha=0$, $B=-\rho\mu_x$ which is the bias of always using the regression estimator. On the other hand, when $\alpha=1$, the estimator is $\bar{y}$ and $B=0$.

It is easily seen that the bias changes sign when ρ or a changes sign. Therefore, we need only to study the bias when a and ρ are positive. It is noted that $\sqrt{n}B$ is a function of a, ρ and α. Table 1 gives the values of $-\sqrt{n}B$ for some choices of ρ and α. The bias is zero when $\mu_x=0$, i.e., when the hypothesis H_0 is true.

Table 1 Values of $-\sqrt{n}B$

a	$\alpha=.05$			$\alpha=.10$			$\alpha=.20$			$\alpha=.50$		
	ρ			ρ			ρ			ρ		
	.1	.5	.9	.1	.5	.9	.1	.5	.9	.1	.5	.9
0	0	0	0	0	0	0	0	0	0	0	0	0
0.5	.034	.171	.308	.026	.131	.236	.016	.080	.144	.003	.016	.029
1.0	.058	.292	.525	.042	.212	.382	.025	.123	.221	.005	.023	.041
1.5	.066	.329	.592	.044	.222	.399	.024	.118	.212	.004	.019	.035
2.0	.057	.285	.513	.035	.174	.313	.017	.083	.149	.002	.012	.021
2.5	.039	.196	.353	.021	.107	.193	.009	.045	.080	.001	.005	.009
3.0	.022	.108	.194	.010	.052	.093	.004	.019	.034	.000	.002	.003

For fixed n, α and ρ, the magnitude of bias first increases then decreases to zero as μ_x increases. In general, the magnitude of bias is an increasing function of ρ and a decreasing function of α.

In order to find the MSE of $\bar{y}^*$, we note that

$$(2.4)\quad \mathrm{MSE}(\bar{y}^*) = \mathrm{E}(\bar{y}^{*2}) - [\mathrm{E}(\bar{y}^*)]^2 + B^2$$

$\mathrm{E}(\bar{y}^{*2})$ can be written as

$$(2.5)\quad \begin{aligned} \mathrm{E}(\bar{y}^{*2}) &= \mathrm{E}\{(\bar{y}-\rho\bar{x})^2/|\sqrt{n}\,\bar{x}|\leqq z_\alpha\}\, \mathrm{P}\{|\sqrt{n}\,\bar{x}|\leqq z_\alpha\} \\ &\quad + \mathrm{E}\{\bar{y}^2/\sqrt{n}\,\bar{x}|>z_\alpha\}\, \mathrm{P}\{|\sqrt{n}\,\bar{x}|>z_\alpha\} \\ &= \mathrm{E}(\bar{y}^2) - 2\rho\, \mathrm{E}\{\bar{x}\bar{y}/|\sqrt{n}\,\bar{x}|\leqq z_\alpha\}\, \mathrm{P}\{|\sqrt{n}\,\bar{x}|\leqq z_\alpha\} \\ &\quad + \rho^2\, \mathrm{E}\{\bar{x}^2/\sqrt{n}\,\bar{x}|\leqq z_\alpha\}\, \mathrm{P}\{|\sqrt{n}\,\bar{x}|\leqq z_\alpha\}\ . \end{aligned}$$

It is easily seen that $\mathrm{E}(\bar{y}^2) = 1/n + \mu_y^2$. Since $\bar{x}$ and $\bar{y}$ have a bivariate normal distribution, the last two terms on the right hand side of (2.5) can be evaluated directly. Let $f(\bar{x}, \bar{y})$ be the bivariate normal density of $\bar{x}$ and $\bar{y}$, we have

$$\mathrm{E}\{\bar{x}\bar{y}/|\sqrt{n}\,\bar{x}|\leqq z_\alpha\}\ \mathrm{P}\{|\sqrt{n}\,\bar{x}|\leqq z_\alpha\}=\int_{-z_\alpha/\sqrt{n}}^{z_\alpha/\sqrt{n}}\int_{-\infty}^{\infty}\bar{x}\bar{y}f(\bar{x},\bar{y})d\bar{y}d\bar{x}\ .$$

After some algebra, this integral is found to be

$$\text{(2.6)}\qquad -\frac{\sqrt{2}}{\sqrt{\pi n}}\exp\{-(z_\alpha^2+a^2)/2\}\left[\frac{\rho z_\alpha}{\sqrt{n}}\cosh(az_\alpha)+\mu_y\sinh(az_\alpha)\right] +\left(\frac{\rho}{n}+\mu_x\mu_y\right)[\Phi(z_\alpha-a)-\Phi(-z_\alpha-a)]\ .$$

Similarly we find

$$\text{(2.7)}\qquad \mathrm{E}\{\bar{x}^2/|\sqrt{n}\,\bar{x}|\leqq z_\alpha\}\ \mathrm{P}\{|\sqrt{n}\,\bar{x}|\leqq z_\alpha\} =-\frac{\sqrt{2}}{\sqrt{\pi n}}\exp\{-(z_\alpha^2+a^2)/2\}\left[\frac{z_\alpha}{\sqrt{n}}\cosh(az_\alpha)+\frac{a}{\sqrt{n}}\sinh(az_\alpha)\right] +\frac{1}{n}(1+a^2)[\Phi(z_\alpha-a)-\Phi(-z_\alpha-a)]\ .$$

Therefore by (2.4) we obtain that

$$\mathrm{MSE}\,(\bar{y}^*)=\frac{1}{n}[1+f(a)]\ ,$$

where

$$\text{(2.8)}\qquad f(a)=\frac{\sqrt{2}\,\rho^2}{\sqrt{\pi}}\exp\{-(z_\alpha^2+a^2)\}[z_\alpha\cosh(az_\alpha)-a\sinh(az_\alpha)] -\rho^2(1-a^2)[\Phi(z_\alpha-a)-\Phi(-z_\alpha-a)]\ .$$

When $\alpha=0$, $\mathrm{MSE}\,(\bar{y}^*)=1/n-\rho^2(1/n-\mu_x^2)$ which is the MSE of the regression estimator. When $\alpha=1$, $\mathrm{MSE}\,(\bar{y}^*)=1/n$ which is the variance of $\bar{y}$.

3. Relative efficiency of $\bar{y}^*$

In practice, the experimenter wants to select an estimator for μ_y with the smallest bias and MSE. Since bias is a part of MSE, it is reasonable to consider only the MSE. The relative efficiency of $\bar{y}^*$ to the usual estimator $\bar{y}$ is defined to be

$$\text{(3.1)}\qquad e=\frac{1}{\mathrm{MSE}\,(\bar{y}^*)}\Big/\frac{1}{\mathrm{MSE}\,(\bar{y})}=\frac{1}{1+f(a)}\ ,$$

where $f(a)$ is given in (2.8). e is a function of a and α. The experimenter has prior information that μ_x is close to 0 but the exact value of μ_x is unknown. He would like to select an estimator, or the level of the preliminary test, such that the relative efficiency is the largest when $\mu_x=0$ and is at least as large as e_0 when $\mu_x\neq 0$. With such a selection procedure, the experimenter has a guaranteed protection in the sense that the relative efficiency of $\bar{y}^*$ is no less than e_0. This selection procedure was first recommended by Han and Bancroft [4].

Table 2 gives the values of e_0 for $\rho=0.1$ (0.1) 0.9, the corresponding α level to use and the maximum relative efficiency e^* which occurs at $\mu_x=0$. For given values of ρ and e_0, the experimenter enters the table to select the α level to use for the preliminary test. If the null hypothesis is true, the relative efficiency of $\bar{y}^*$ can be as large as e^*. It is noted that $f(a)$ in (2.8) is a symmetric function of ρ and therefore Table 2 can also be used for negative values of ρ.

Table 2 Relative efficiency of $\bar{y}^*$

α^* \ ρ		.1	.2	.3	.4	.5	.6	.7	.8	.9
.50	e^*	1.00	1.00	1.01	1.01	1.02	1.03	1.04	1.05	1.06
	e_0	1.00	1.00	.99	.99	.98	.97	.96	.95	.93
.40	e^*	1.00	1.01	1.01	1.02	1.03	1.05	1.07	1.09	1.12
	e_0	1.00	.99	.99	.97	.96	.94	.93	.91	.88
.30	e^*	1.00	1.01	1.02	1.04	1.06	1.08	1.12	1.16	1.21
	e_0	1.00	.99	.97	.96	.93	.91	.88	.85	.81
.20	e^*	1.00	1.01	1.03	1.06	1.10	1.14	1.21	1.29	1.40
	e_0	1.00	.98	.96	.93	.89	.85	.80	.76	.71
.10	e^*	1.01	1.02	1.05	1.10	1.16	1.25	1.38	1.56	1.83
	e_0	.99	.96	.92	.87	.81	.74	.68	.62	.56
.05	e^*	1.01	1.03	1.07	1.13	1.22	1.35	1.55	1.86	2.40
	e_0	.99	.94	.88	.81	.73	.66	.58	.52	.46

From Table 2, we observe that when ρ is small, say 0.2 or less, the relative efficiencies are close to unity. This is because the estimator has no significant change for small values of ρ. Hence, there is little difference whether one selects $\alpha=.05$ or .50. When ρ becomes large, we expect that the regression estimator is better and the relative efficiencies fluctuate. If the experimenter decides that the relative efficiency should be at least .80, then with the selection procedure given above, he would use $\alpha=.05$ for the preliminary test when $\rho=.4$. With such a choice, the relative efficiency of $\bar{y}^*$ is at least .81 and can be as large as 1.13 when the null hypothesis of the preliminary test is true. If $\rho=.9$, he would use $\alpha=.30$, then the guaranteed relative efficiency is .81 and it can be as high as 1.21.

4. Bias and mean square error when the covariance matrix is unknown

When σ_x^2, σ_y^2 and ρ are unknown, the preliminary test estimator becomes

$$\bar{y}' = \begin{cases} \bar{y} - \dfrac{S_{xy}}{S_x}\bar{x} & \text{if } |t| \leqq t_\alpha \\ \bar{y} & \text{if } |t| > t_\alpha , \end{cases} \tag{4.1}$$

where

$$t = \frac{\bar{x}}{\sqrt{S_x/(n(n-1))}} , \qquad S_x = \sum_{i=1}^{n} (x_i - \bar{x})^2 ,$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) , \qquad S_y = \sum_{i=1}^{n} (y_i - \bar{y})^2 ,$$

and t_α is the $100(1-\alpha/2)\%$ point of t distribution with $n-1$ degrees of freedom.

The expected value of $\bar{y}'$ is

$$\begin{aligned} \mathrm{E}\,(\bar{y}') &= \mathrm{E}\left[\bar{y} - \frac{S_{xy}}{S_x}\bar{x}/|t| \leqq t_\alpha\right] \mathrm{P}\,(|t| \leqq t_\alpha) \\ &\quad + \mathrm{E}\,[\bar{y}/|t| > t_\alpha]\, \mathrm{P}\,(|t| > t_\alpha) \\ &= \mu_y - \mathrm{E}\left[\frac{S_{xy}}{S_x}\bar{x}/|t| \leqq t_\alpha\right] \mathrm{P}\,(|t| \leqq t_\alpha) . \end{aligned} \tag{4.2}$$

Therefore, the second term is the bias. Let $f(\bar{x})$ be the normal density of $\bar{x}$ and $g(S_x, S_{xy}, S_y)$ be the density of S_x, S_y and S_{xy} which have a Wishart distribution, then

$$\begin{aligned} &\mathrm{E}\left[\frac{S_{xy}}{S_x}\bar{x}/|t| \leqq t_\alpha\right] \mathrm{P}\,(|t| \leqq t_\alpha) \\ &= \int \cdots \int_{|t| \leqq t_\alpha} \frac{S_{xy}}{S_x}\bar{x} f(\bar{x}) g(S_x, S_{xy}, S_y)\, d\bar{x}\, dS_x\, dS_{xy}\, dS_y . \end{aligned}$$

Following Rao [8], we make the following transformation.

$$W_1 = \frac{S_x}{\sigma_x^2} , \qquad W_2 = \frac{S_{xy} - \beta S_x}{\sigma_{y\cdot x}\sqrt{S_x}} , \qquad W_3 = \frac{1}{\sigma_{y\cdot x}^2}(S_y - S_{xy}^2/S_x) \tag{4.3}$$

where $\beta = \rho\sigma_y/\sigma_x$ and $\sigma_{y\cdot x}^2 = \sigma_y^2(1-\rho^2)$. After making the substitution and integrating out W_3, W_2 and W_1 successively, we obtain

$$\frac{\text{Bias}}{\sigma_y} = -\sum_{i=0}^{(n-1)/2-1} H_i \rho \mu'_{2i+1} \tag{4.4}$$

where μ'_k is the kth moment of $N(\mu_x/c\sigma_x, n/c)$ and

$$H_i = \frac{1}{i!}\left[\frac{n(n-1)}{2t_\alpha^2}\right]^i \frac{1}{c} \exp\left\{-\frac{n(n-1)}{2(t_\alpha^2+n-1)}\left(\frac{\mu_x}{\sigma_x}\right)^2\right\} \tag{4.5}$$

where $c = 1 + (n-1)/t_\alpha^2$. Here we assume that n is odd and $\geqq 3$, so that $(n-1)/2$ is an integer.

The bias in (4.4) decreases as n increases. It changes sign when ρ or μ_x changes sign. Therefore, we shall compute the bias for positive ρ and μ_x. Table 3 gives the bias for $n=9$ which may represent the behavior of the bias.

Table 3 Values of $-\text{Bias}/\sigma_y$, $n=9$

μ_x/σ_x	$\alpha=.05$			$\alpha=.10$			$\alpha=.20$			$\alpha=.50$		
	ρ			ρ			ρ			ρ		
	.1	.5	.9	.1	.5	.9	.1	.5	.9	.1	.5	.9
0	0	0	0	0	0	0	0	0	0	0	0	0
0.3	.018	.092	.166	.014	.069	.124	.008	.040	.073	.001	.007	.013
0.6	.029	.143	.257	.018	.091	.163	.009	.043	.077	.001	.005	.010
0.9	.023	.114	.206	.011	.056	.102	.004	.020	.035	0	.001	.003
1.2	.010	.051	.092	.004	.018	.033	.001	.004	.008	0	0	0
1.5	.002	.013	.023	.001	.003	.006	0	0	.001	0	0	0

In general, the magnitude of the bias first increases then decreases to zero as μ_x/σ_x increases; it is an increasing function of ρ and a decreasing function of α. Hence, the behavior of the bias is similar to the case when the covariance matrix is known.

To derive the mean square error, we first find $\mathrm{E}(\bar{y}'^2)$. This expectation is

$$
\begin{aligned}
(4.6)\qquad \mathrm{E}(\bar{y}'^2) &= \mathrm{E}\left[\left(\bar{y}-\frac{S_{xy}}{S_x}\bar{x}\right)^2\Big/|t|\leqq t_\alpha\right]\mathrm{P}(|t|\leqq t_\alpha) \\
&\quad +\mathrm{E}\left[\bar{y}^2/|t|>t_\alpha\right]\mathrm{P}(|t|>t_\alpha) \\
&= \mathrm{E}(\bar{y}^2)-2\,\mathrm{E}\left(\frac{S_{xy}}{S_x}\bar{x}\bar{y}/|t|\leqq t_\alpha\right)\mathrm{P}(|t|\leqq t_\alpha) \\
&\quad +\mathrm{E}\left(\frac{S_{xy}^2}{S_x^2}\bar{x}^2/|t|\leqq t_\alpha\right)\mathrm{P}(|t|\leqq t_\alpha)\,.
\end{aligned}
$$

Again, we use the transformation given in (4.3); and the last two terms on the right hand side of (4.6) are found to be, for odd values of n,

$$
\begin{aligned}
&\mathrm{E}\left(\frac{S_{xy}}{S_x}\bar{x}\bar{y}/|t|\leqq t_\alpha\right)\mathrm{P}(|t|\leqq t_\alpha) \\
&\quad =\sum_{i=0}^{(n-1)/2-1}\beta H_i[(\mu_y-\beta\mu_x)\sigma_x\mu'_{2i+1}+\beta\sigma_x^2\mu'_{2i+2}]\,. \\
&\mathrm{E}\left(\frac{S_{xy}^2}{S_x^2}\bar{x}^2/|t|\leqq t_\alpha\right)\mathrm{P}(|t|\leqq t_\alpha) \\
&\quad =\sum_{i=0}^{(n-1)/2-1}\sigma_x^2\beta^2H_i\mu'_{2i+2}+\frac{1}{n-3}\sigma_{y\cdot x}^2\sum_{i=1}^{(n-1)/2-2}H_i\mu'_{2i+2}\,.
\end{aligned}
$$

Since $\mathrm{E}(\bar{y}^2)=\mu_y^2+\sigma_y^2/n$, we obtain by (2.4)

$$(4.7)\qquad \frac{\text{MSE}(\bar{y}')}{\sigma_y^2}=\frac{1}{n}+2\sum_{i=0}^{(n-1)/2-1} H_i\rho^2\frac{\mu_x}{\sigma_x}\mu'_{2i+1}-\sum_{i=0}^{(n-1)/2-1} H_i\rho^2\mu'_{2i+1} +\frac{1}{n-3}\sum_{i=0}^{(n-1)/2-2}(1-\rho^2)H_i\mu'_{2i+2}\,.$$

As a partial check, when $\alpha=0$, the regression estimator is always used and the mean square error becomes

$$\frac{\text{MSE}(\bar{y}')}{\sigma_y^2}=\frac{1-\rho^2}{n}+\rho^2\left(\frac{\mu_x}{\sigma_x}\right)^2+\frac{1}{n-3}(1-\rho^2)\left[\frac{1}{n}+\left(\frac{\mu_x}{\sigma_x}\right)^2\right].$$

When $\alpha=1$, we only use $\bar{y}$ to estimate μ_y and $\text{MSE}(\bar{y}')=\sigma_y^2/n$.

5. Relative efficiency of $\bar{y}'$

In order to evaluate the gain and loss of precision of the preliminary test estimator $\bar{y}'$, we consider the relative efficiency of $\bar{y}'$ to the usual estimator $\bar{y}$. This is defined to be

$$e'=\frac{1/\text{MSE}(\bar{y}')}{1/\text{MSE}(\bar{y})}=1\bigg/\left[n\frac{\text{MSE}(\bar{y}')}{\sigma_y^2}\right].$$

Since e' is symmetric about $\rho=0$ and $\mu_x=0$, so we only need to consider $\rho\geqq 0$ and $\mu_x\geqq 0$. The values of e' are calculated by using computer for selected values of n, ρ and α. After studying the graphs of e', we find that, in general, for fixed values of n, ρ and α, the relative efficiency has a maximum larger than unity at $\mu_x=0$. When μ_x increases, e' decreases first to a minimum and then increases to unity. This shows a typical behavior of a preliminary test estimator. This is, we expect to gain in precision when μ_x/σ_x is close to zero and lose for moderate values of μ_x/σ_x. When μ_x/σ_x tends to infinity, $\bar{y}'$ and $\bar{y}$ are asymptotically equivalent.

In order to consider the selection of the level α of the preliminary test, we compute, as in Section 3, the maximum relative efficiency e^* and the smallest relative efficiency e_0. Table 4 gives the value of e_0 for $n=5, 7, 9, 11, 15, 19$, the corresponding α level to use and e^*. Since ρ is unknown, we can compute the sample correlation coefficient as an estimate.

As Table 4 shows, when ρ is small and the sample size is small, one does not gain in precision by using the regression estimator. Therefore, if ρ is .3 or less and n is less than 10, the experimenter should always use $\bar{y}$ as the estimator. When ρ is large, say .5 and larger, he can determine the smallest relative efficient he wishes to have and select the α level from the table. Because the experimenter has prior information that the true value of μ_x is close to μ_0, consequently, the use of the preliminary test estimator results in higher precision.

Table 4 Relative efficiency of $\bar{y}'$

α \ ρ		$n=5$				$n=7$				$n=9$			
		.3	.5	.7	.9	.3	.5	.7	.9	.3	.5	.7	.9
.50	e^*	.99	1.01	1.04	1.08	1.00	1.01	1.04	1.07	1.00	1.01	1.04	1.07
	e_0	.98	.97	.95	.93	.99	.98	.96	.93	.99	.98	.96	.93
.40	e^*	.98	1.01	1.07	1.15	.99	1.02	1.07	1.14	1.00	1.03	1.07	1.13
	e_0	.97	.95	.91	.87	.97	.95	.92	.88	.98	.95	.92	.88
.30	e^*	.97	1.02	1.12	1.27	.99	1.04	1.12	1.25	1.00	1.04	1.12	1.24
	e_0	.94	.90	.85	.78	.95	.92	.86	.80	.96	.92	.86	.79
.20	e^*	.94	1.03	1.19	1.50	.97	1.05	1.20	1.47	.99	1.07	1.20	1.45
	e_0	.88	.82	.74	.65	.91	.85	.76	.67	.92	.85	.75	.66
.10	e^*	.90	1.02	1.29	1.99	.95	1.08	1.33	1.96	.98	1.10	1.35	1.93
	e_0	.76	.67	.56	.46	.82	.72	.60	.49	.84	.71	.58	.46
.05	e^*	.85	1.00	1.35	2.52	.93	1.09	1.44	2.54	.97	1.13	1.48	2.52
	e_0	.61	.51	.40	.31	.71	.59	.46	.35	.74	.58	.44	.33

α \ ρ		$n=11$				$n=15$				$n=19$			
		.3	.5	.7	.9	.3	.5	.7	.9	.3	.5	.7	.9
.50	e^*	1.00	1.01	1.04	1.07	1.00	1.02	1.04	1.07	1.00	1.02	1.04	1.07
	e_0	.99	.98	.96	.93	.99	.98	.96	.93	.99	.98	.95	.93
.40	e^*	1.00	1.03	1.07	1.13	1.00	1.03	1.07	1.13	1.01	1.03	1.07	1.12
	e_0	.98	.93	.92	.87	.98	.95	.91	.86	.98	.95.	91	.86
.30	e^*	1.00	1.04	1.12	1.24	1.01	1.05	1.12	1.23	1.01	1.05	1.12	1.23
	e_0	.96	.91	.85	.78	.96	.91	.84	.76	.96	.91	.83	.75
.20	e^*	1.00	1.07	1.20	1.44	1.01	1.08	1.21	1.43	1.01	1.08	1.21	1.42
	e_0	.92	.84	.74	.64	.92	.83	.72	.61	.92	.82	.70	.59
.10	e^*	1.00	1.12	1.36	1.91	1.01	1.13	1.37	1.89	1.02	1.14	1.37	1.88
	e_0	.84	.69	.55	.43	.83	.66	.51	.39	.82	.64	.48	.36
.05	e^*	.99	1.15	1.49	2.50	1.02	1.17	1.51	2.48	1.03	1.18	1.52	2.47
	e_0	.74	.56	.41	.30	.72	.51	.36	.25	.71	.48	.32	.23

The author wishes to thank Drs. William Kennedy and Eric West for their help in computing the tables. This research was partially supported by National Science Foundation Grant GP9046.

IOWA STATE UNIVERSITY

REFERENCES

[1] Bancroft, T. A. (1944). On biases in estimation due to the use of preliminary tests of significance, *Ann. Math. Statist.*, **15**, 190-204.

[2] Bennett, B. M. (1952). Estimation of means on the basis of preliminary tests of significance, *Ann. Inst. Statist. Math.*, **4**, 31-43.

[3] Dempster, A. P. (1969). *Elements of Continuous Multivariate Analysis*, Addison-Wesley, Reading, Massachusetts.

[4] Han, C. P. and Bancroft, T. A. (1968). On pooling means when variance is unknown, *Jour. Amer. Statist. Assoc.*, **63**, 1333-1342.

[5] Kale, B. K. and Bancroft, T. A. (1967). Inference for some incompletely specified models involving normal approximations to discrete data, *Biometrics*, **23**, 335-348.

[6] Kitagawa, Tosio (1963). Estimation after preliminary tests of significance, *Univ. Cal. Pub. in Statist.*, **3**, 147-186.

[7] Mosteller, Frederick (1948). On pooling data, *Jour. Amer. Statist. Assoc.*, **43**, 231-242.
[8] Rao, C. R. (1965). *Linear Statistical Inference and Its Application*, John Wiley, New York.