Bull. Korean Math. Soc.
Vol. 13, No. 2, 1976.

A NOTE ON THE PROOF OF THE RELATION $\bar{X}'(X'X)^{-1}\bar{X}=1/n$

BY HEE-KYUN CHOI

1. Introduction.

It is well known that in the multiple regression model

$$Y_i=B_1+B_2X_{2i}+\cdots\cdots+B_kX_{ki}+u_i \qquad i=1,\cdots\cdots,n \tag{1}$$

$$E(u_i)=0 \quad \text{for all } i \tag{2}$$

$$E(u_iu_j)=\begin{cases}\sigma^2 & i=j:\ i,j=1,\cdots\cdots,n\\ 0 & i\neq j:\ i,j=1,\cdots\cdots,n\end{cases} \tag{3}$$

the error variance of predicting the expected value of Y associated with X_0 is given by

$$\mathrm{Var}(\hat{Y}_0-E(Y_0))=\sigma^2\ (X_0'(X'X)^{-1}\ X_0) \tag{4}$$

where $\hat{Y}_0$ is the prediction of Y at X_0 and

$$X_0=\begin{pmatrix}1\\ X_{20}\\ X_{30}\\ \vdots\\ X_{k0}\end{pmatrix},\qquad X=\begin{pmatrix}1 & X_{21}\cdots X_{k1}\\ 1 & X_{22}\cdots X_{k2}\\ \vdots & \vdots \quad\ \vdots\\ 1 & X_{2n}\cdots X_{kn}\end{pmatrix}.$$

The matrix X on n sample observations on the k independent variables with $X_{1i}=1$ has rank $k<n$ and the vector X_0 is the particular values of the independent variables to predict Y.

Desalvo [1] has shown that error variance at the sample mean of X is given by

$$\mathrm{Var}(\hat{Y}_0-E(Y_0))_{X_0=\bar{X}}=\sigma^2(\bar{X}'(X'X)^{-1}\bar{X})=\sigma^2/n \tag{5}$$

as proving the relation

$$\bar{X}'(X'X)^{-1}\bar{X}=1/n \tag{6}$$

without statistical property as will be shown below.

<A summary of Desalvo's proof>

$$\bar{X}'(X'X)^{-1}\bar{X}=(1/|X'X|)\ \bar{X}'C\bar{X} \quad (C \text{ is the cofactor matrix})$$
$$=(1/(|X'X|n^2))\sum_{r=1}^{k}\sum_{j=1}^{k}(\sum_{i=1}^{n}X_{ri})(\sum_{i=1}^{n}X_{ji})c_{rj}. \tag{7}$$

When $r=1$ in the right side of (7)

$$\sum_{i=1}^{n}X_{1i}\sum_{j=1}^{k}(\sum_{i=1}^{n}X_{ji})c_{1j}=n|X'X| \tag{8}$$

while $r\neq 1$,

$$\sum_{i=1}^{n}X_{ri}\sum_{j=1}^{k}(\sum_{i=1}^{n}X_{ji})\ c_{rj}=0,$$

hence

$$\bar{X}'(X'X)^{-1}\bar{X}=1/n.$$

Now we will show a simpler method for the proof of the relation (6) with such statistical property as the least-square method.

2. An alternative proof.

A consequence of the least-squares fit is that the sum of the residuals is zero.

That is

$$\sum_{i=1}^{n}e_i=\mathbf{1}'(I_n-X(X'X)^{-1}X')U=\sum_{i=1}^{n}a_i\ u_i=0, \tag{9}$$

where u_i and a_i are the ith component of the disturbance vector U and A' $(=I_n-X\ (X'X)^{-1}X')$ respectively and $\mathbf{1}$ denotes the vector of n unities.

Then it is clear that $u_1, u_2, \cdots\cdots, u_n$ are linearly independent and every $a_i=0$ by (3) and (9) respectively. i. e. $A'=0$.

Since $I_n-X(X'X)^{-1}X'$ is symmetric and idempotent [2] and $\bar{X}=\frac{X'\mathbf{1}}{n}$ we obtain that

$$0=\mathbf{1}'\ (I_n-X(X'X)^{-1}X')\ (I_n-(X(X'X)^{-1}X')'\mathbf{1}$$
$$=\mathbf{1}'\ (I_n-X(X'X)^{-1}X')\mathbf{1}$$
$$=\mathbf{1}'I_n\ \mathbf{1}-\mathbf{1}'\ X(X'X)^{-1}X'\mathbf{1}$$
$$=n-n\bar{X}\ (X'X)^{-1}\bar{X}n$$

hence

$$\bar{X}'(X'X)^{-1}\bar{X}=1/n.$$

This final result is the same as that of Desalvo.

References

[1] Joseph S. Desalvo, *"Standard Error of Forecast in Multiple Regression; Proof of a useful result"* The American Statistician, Vol. 25, No. 4, (October, 1971) pp. 32—35.

[2] J. Johnston, *"Econometric Method"* 2nd Edition, McGraw—Hill Kogakusha, Ltd. Tokyo, 1972, pp. 128—130.

Ajou Institute of Technology