

The Wayback Machine - <https://web.archive.org/web/20130307064041/http://www.amstat.org/publications...>

# A Novel Method of Proof With an Application to Regression

Juliet Popper Shaffer  
University of California

Yung-Pin Chen  
Smith College

*Journal of Statistics Education* v.7, n.3 (1999)

Copyright (c) 1999 by Juliet Popper Shaffer and Yung-Pin Chen, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

---

**Key Words:** Best linear unbiased estimation; Least squares estimation; Unbiased estimation; Variance estimation.

## Abstract

A useful way of approaching a statistical problem is to consider whether the addition of some missing information would transform the problem into a standard form with a known solution. The EM algorithm ([Dempster, Laird, and Rubin 1977](#)), for example, makes use of this approach to simplify computation. Occasionally it turns out that knowledge of the missing values is not necessary to apply the standard approach. In such cases the following simple logical argument shows that any optimality properties of the standard approach in the full-information situation generalize immediately to the approach in the original limited-information situation: If any better estimate were available in the limited-information situation, it would also be available in the full-information situation, which would contradict the optimality of the original estimator. This approach then provides a simple proof of optimality, and often leads directly to a simple derivation of other properties of the solution. The approach can be taught to graduate students and theoretically-inclined undergraduates. Its application to the elementary proof of a result in linear regression, and some extensions, are described in this paper. The resulting derivations provide more insight into some equivalences among models as well as proofs simpler than the standard ones.

## 1. Introduction

1 Assume the linear regression model

(1)

$$\mathbf{y}_+ = \mathbf{X}_+ \boldsymbol{\beta}_+ + \boldsymbol{\epsilon},$$

where  $\mathbf{y}_+$  is an  $n \times 1$  random vector of observations on the dependent variable in the  $n$  experimental units,  $\mathbf{X}_+$  is a fixed, known  $n \times (p+1)$  matrix with all elements of the first column equal to one and of rank  $p+1$ ,  $\boldsymbol{\beta}_+ = (\beta_0, \dots, \beta_p)'$  is a fixed, unknown  $(p+1)$ -vector of coefficients, and  $\boldsymbol{\epsilon}$  is a random  $n \times 1$  vector with expected value zero and covariance matrix  $\sigma^2 \mathbf{I}$ . Under model (1), the best linear unbiased estimator (*BLUE*) of the vector  $\boldsymbol{\beta}_+$  is

(2)

$$\hat{\boldsymbol{\beta}}_+ = (\mathbf{X}_+' \mathbf{X}_+)^{-1} \mathbf{X}_+' \mathbf{y}_+.$$

2 Let  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$  be the vector of coefficients excluding the constant additive term. A well-known fact, stated in most regression textbooks, is that the expression for estimating  $\boldsymbol{\beta}$  can be obtained from the inversion of a  $p \times p$  rather than a  $(p+1) \times (p+1)$  matrix by expressing  $\mathbf{X}_+$  and  $\mathbf{y}_+$  in deviation form (i.e., subtracting column means), and using an expression of the same form as (2).

3 Let  $\bar{y}_+$  be the mean of the  $n$  components of  $\mathbf{y}_+$ , and let  $\bar{\mathbf{X}}_+$  be the  $1 \times (p+1)$  row vector with elements  $\bar{x}_{+, (0)}, \bar{x}_{+, (1)}, \dots, \bar{x}_{+, (p)}$ , where  $\bar{x}_{+, (i)}$  is the mean of the  $n$  components of column  $i$  of the matrix  $\mathbf{X}_+$ . Let  $\bar{\mathbf{1}}$  be an  $n \times 1$  vector of ones. Then  $\hat{\boldsymbol{\beta}}$ , with elements 2 to  $(p+1)$  of  $\hat{\boldsymbol{\beta}}_+$ , can be expressed as

(3)

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y},$$

where  $\mathbf{X}$  is the  $n \times p$  matrix consisting of columns 2 to  $(p+1)$  of  $(\mathbf{X}_+ - \bar{\mathbf{1}} \bar{\mathbf{X}}_+)$  (the first column of  $(\mathbf{X}_+ - \bar{\mathbf{1}} \bar{\mathbf{X}}_+)$  is identically zero and is dropped) and  $\mathbf{y} = (\mathbf{y}_+ - \bar{\mathbf{1}} \bar{y}_+)$ . The proof of this result is not immediately obvious because the covariance matrix of  $\mathbf{y}$ , given  $\mathbf{X}$ , is no longer  $\sigma^2 \mathbf{I}$  but is

(4)

$$\boldsymbol{\Sigma}_y = \sigma^2 [\mathbf{I} - (1/n) \mathbf{J}]$$

where  $\mathbf{J}$  is an  $n \times n$  matrix with all elements equal to 1. [Draper and Smith \(1998, pp. 27-28\)](#) and [Daniel and Wood \(1980, pp. 13-14\)](#) state the result without proof. [Sen and Srivastava \(1990, pp. 42 and 146\)](#) ask for the

proof in a problem, and suggest using a generalized inverse (since the covariance matrix (4) is singular). [Arnold \(1981\)](#), [Graybill \(1976\)](#), [Searle \(1971\)](#), and [Seber \(1977\)](#) have relatively long proofs involving partitioning of matrices.

4 This paper presents a proof that involves only elementary facts that are usually presented in a first introduction to regression. The method of proof involves consideration of missing information that, if known, would put the equation to be solved into a standard regression form.

## 2. Proof of the Equivalence of Raw Form and Deviation Form Solutions

5 Under Model [\(1\)](#), noting that

$$\mathbf{\bar{1}} \bar{y}_+ = \mathbf{\bar{1}} (\bar{\mathbf{X}}_+ \boldsymbol{\beta}_+ + \bar{\epsilon}), \quad (5)$$

where  $\bar{\epsilon}$  is the mean of the  $n$  components of  $\boldsymbol{\epsilon}$ , and subtracting (5) from (1) yields

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\epsilon} - \mathbf{\bar{1}} \bar{\epsilon}, \quad (6)$$

where  $\mathbf{X}$  is the  $n \times p$  matrix defined below [Equation \(3\)](#). Suppose  $\bar{\epsilon}$  were known. Adding  $\mathbf{\bar{1}} \bar{\epsilon}$  to both sides of (6) yields the equation

$$\mathbf{y}^* = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (7)$$

where  $\mathbf{y}^* = \mathbf{y} + \mathbf{\bar{1}} \bar{\epsilon}$ . Note that  $\mathbf{E}(\mathbf{y}^*) = \mathbf{X} \boldsymbol{\beta}$  and  $\boldsymbol{\Sigma}_{\mathbf{y}^*} = \sigma^2 \mathbf{I}$ , so (7) is in the form (1), but with  $\mathbf{y}_+$  and  $\mathbf{X}_+$  replaced by  $\mathbf{y}^*$  and  $\mathbf{X}$ , the latter in deviation form. Thus, the BLUE of  $\boldsymbol{\beta}$  is

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}^* = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' (\mathbf{y} + \mathbf{\bar{1}} \bar{\epsilon}) = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}. \quad (8)$$

The last equality follows because, although  $\bar{\epsilon}$ , which appears in (8), is unknown, each element  $i$  of the vector  $\mathbf{X}' (\mathbf{\bar{1}} \bar{\epsilon})$  is in the form  $\bar{\epsilon} (\sum_{i=1}^n x_{ij}) = 0$ ,  $j = 1, \dots, p$ , since  $\mathbf{X}$  is in deviation form. Then  $\tilde{\boldsymbol{\beta}}$  is the BLUE of  $\boldsymbol{\beta}$ , as is  $\hat{\boldsymbol{\beta}}_+$  omitting the first element. Therefore,  $\hat{\boldsymbol{\beta}}_+$ , without the intercept element, and  $\tilde{\boldsymbol{\beta}}$  must be equivalent.

### 3. Extensions

6 Note that, instead of adding  $(\mathbf{1}\bar{\epsilon})$  to both sides of (6) in [Section 2](#), the realization of any random variable independent of  $\epsilon$  with mean zero and variance  $\sigma^2/n$  could be added to each observation, and the proof would proceed in the same way. By combining the method of Section 2 with the addition of an external variable independent of  $\epsilon$ , a further equivalence can be demonstrated.

*Theorem:* Consider a model of the form (1) but in which the covariance matrix of  $\epsilon$  is

$$\Sigma_{\epsilon} = \sigma^2[(1 - \rho)\mathbf{I} + \rho\mathbf{J}], \quad -1/(n - 1) \leq \rho \leq 1. \quad (9)$$

The BLUE of  $\beta$  in the model (1) but with the covariance matrix of  $\epsilon$  given by Equation (9) is the same as the BLUE of  $\beta$  in the model (1) with the covariance matrix of  $\epsilon = \sigma^2\mathbf{I}$  (note that  $\beta$  is  $\beta_+$  without the intercept term). Furthermore, the covariance matrix of the BLUE of  $\beta$  with error covariance matrix (9) is  $(1 - \rho)$  times as great as the covariance matrix of the BLUE of  $\beta$  with error covariance matrix  $\sigma^2\mathbf{I}$ .

*Proof:* Three cases will be considered separately.

- i.  $\rho = 0$ . In this case, the model reduces to (1) with error covariance matrix  $\sigma^2\mathbf{I}$ , so no proof is needed.
- ii.  $\rho < 0$ . As in [Section 2](#), convert the model (1) into the form (7) by subtracting the expression for  $\mathbf{1}\bar{y}_+$  from (1) and adding  $\mathbf{1}\bar{\epsilon}$  to both sides. The covariance matrix of  $\mathbf{y}^*$  is unchanged from that of  $\mathbf{y}_+$  by this transformation; i.e., it is given by (9). Now add a random vector  $\mathbf{1}\delta$  to both sides of (7), where  $\delta$  is the realization of a real-valued random variable independent of  $\epsilon$  and with expected value zero and variance  $\sigma^2|\rho|$ . The equation (7) becomes

$$\mathbf{y}^* + \mathbf{1}\delta = \mathbf{X}\beta + \epsilon + \mathbf{1}\delta \quad (10)$$

or

$$\mathbf{z} = \mathbf{X}\beta + \eta \quad (11)$$

where  $\mathbf{z} = \mathbf{y}^* + \mathbf{1}\delta$  and  $\boldsymbol{\eta} = \boldsymbol{\epsilon} + \mathbf{1}\delta$ . The covariance matrix of  $\boldsymbol{\eta}$  is  $\sigma^2(1 - \rho)\mathbf{I} = \tau^2\mathbf{I}$ , where  $\tau^2 = (1 - \rho)\sigma^2$ , and therefore Model (11) is in the same form as Model (7), with a covariance matrix  $(1 - \rho)$  times as large, so the theorem follows from the previous one. A proof along these lines was used in [Shaffer \(1981, p. 609\)](#).

- iii.  $\rho > 0$ . The proof extends to the case  $\rho > 0$  by first converting the model to (7) with  $\mathbf{X}$  in deviation form as in (ii). Now express  $\boldsymbol{\epsilon}$  as  $\boldsymbol{\eta} + \mathbf{1}\delta$ , where  $\boldsymbol{\eta}$  is a random  $n \times 1$  vector with mean zero and covariance matrix  $\sigma^2(1 - \rho)\mathbf{I}$ , and  $\delta$  is a real-valued random variable with mean zero and variance  $\sigma^2\rho$ . Subtract  $\mathbf{1}\delta$  from both sides of (7), so that the resulting model is in the same form as Model (7), with a covariance matrix  $(1 - \rho)$  times as large, as in (ii), but now with  $\mathbf{z} = \mathbf{y}^* - \mathbf{1}\delta$ . The proof proceeds as in (ii).

7 The proof for  $\rho > 0$  requires a further comment. Up to now, there has been no reference to the distribution of the errors, except for expected values and variances. As is well known, the results above are distribution-free, requiring only that  $\sigma^2$  be finite. However, the proof above for  $\rho > 0$ , in which a random vector is expressed as the sum of two others, requires some distributional assumption, since not all random vectors can be expressed in that form.

8 Assume two models A and B with the same regression coefficients but different error covariance matrices. Because the BLUE, given any model, has a fixed form, depending only on the covariance matrix of the errors and not on other features of the error distribution (of course with means of errors equal to zero), it follows that if A and B (with the specified error covariance matrices above) have the same BLUEs under any error distribution, they have the same BLUEs for all error distributions. Therefore, it can be assumed, without loss of generality, that the error distributions are normal. In that case, the proof of the theorem for  $\rho > 0$  can be carried out, since a normal random vector with the given covariance matrix can be decomposed as required in the proof.

9 Note that another proof, in the same spirit, that combines negative and positive values of  $\rho$  can be obtained by transforming the model into the deviation form (6), and then adding the realization of a single independent random variable with variance  $\sigma^2(1 - \rho)/n$  to each observation. The proof using this alternative approach appears to be more compact but involves somewhat more matrix manipulation than the previous one. Both proofs start with the deviation form (6). In order to determine the variance of the single independent variable value to be added to each observation using this alternative proof, it is necessary to derive the covariance matrix of  $(\boldsymbol{\epsilon} - \mathbf{1}\bar{\epsilon})$ , which requires a fair amount of calculation because of the nonzero covariances in (9). In

the original proof given, by starting with (6) and adding  $(\mathbf{1}\bar{\epsilon})$  to both sides, it is immediately clear that the equation is in the form (7) with the covariance matrix unchanged from the original form (9). The single

further step, required in both proofs, involves adding (or subtracting) the realization of a random variable, independent of  $\epsilon$ , to each observation. The only knowledge needed for this latter step is that the variance of that random variable is added to every term in the covariance matrix; this is easy to show. However, the alternative proof, although somewhat more complex than the first one given, is nonetheless simpler than proofs presently available in the literature.

10 [McElroy \(1967\)](#) proved that the error covariance matrix (9) is a necessary and sufficient condition for (2) to be the BLUE of  $\beta_+$  in Model (1), for  $-1/(n-1) < \rho < 1$ . McElroy's result is more general than the result in this theorem in that it includes the estimate of  $\beta_0$  and includes necessity as well as sufficiency of (9).

On the other hand, the result in this theorem is more general than McElroy's result in that the singular matrices resulting when  $\rho = -1/(n-1)$  and when  $\rho = 1$  are included in the range of  $\rho$ .

11 An interesting insight follows from considering the case  $\rho = 1$ . In this case, there is (with probability one) a single realization  $\epsilon$  of a random variable added to the expected value for each observation. Then, in the form (6), each element of  $\mathbf{y}$  is exactly equal to its expected value, so the covariance matrix of  $\mathbf{y}$  is the null matrix, and  $\beta$  can be calculated exactly.

## 4. Conclusion

12 The development in [Section 2](#) provides a simple proof that estimators of the coefficient vector in a standard linear model, excluding the intercept term, can be obtained by expressing the sample values of the predictors and predicted variable in deviation form and using the standard equation for the estimator of  $\beta$  in a model without an intercept. From a geometric point of view, note that the regression plane goes through the point  $(\bar{x}_{+, (1)}, \dots, \bar{x}_{+, (p)}, \bar{y}_+)$ , and centering shifts the plane to go through the origin but doesn't change the slopes. (Of course the solution in this paper is for a model with an intercept, and should not be confused with the solution for a model that assumes an intercept of zero, for which the estimators and their properties are different.) The proof proceeds by noting that the addition of an unknown value ( $\bar{\epsilon}$ ) would put the relevant expressions into standard form. The same approach, with the addition of an external random vector, leads to a simple proof that the BLUE of  $\beta$  when the error covariance matrix is of the form  $\sigma^2[(1-\rho)\mathbf{I} + \rho\mathbf{J}]$  is the same as when it is of the standard form  $\sigma^2\mathbf{I}$ , and yields a simple expression for the comparative variances of the estimators. It follows also from the proof that the results hold even if the covariance matrix of  $\mathbf{y}_+$  is singular. Neither extensive matrix manipulation nor knowledge of generalized inverses is necessary. The results generalize under appropriate conditions to the correlation model ([Arnold 1981](#), [Shaffer 1991](#)), in which the elements of  $\mathbf{X}$  are realizations of random variables.

## Acknowledgments

We would like to thank Katherine Halvorsen for her valuable comments which helped to improve the

presentation of this paper.

---

## References

- Arnold, S. F. (1981), *The Theory of Linear Models and Multivariate Analysis*, New York: Wiley-Interscience.
- Daniel, C. and Wood, F. S. (1980), *Fitting Equations to Data: Computer Analysis of Multifactor Data for Scientists and Engineers* (2nd ed.), New York: Wiley-Interscience.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society, Ser. B*, 39, 1-22.
- Draper, N. R., and Smith, H. (1998), *Applied Regression Analysis* (3rd ed.), New York: Wiley.
- Graybill, F. A. (1976), *Theory and Application of the Linear Model*, North Scituate, MA: Duxbury Press.
- McElroy, F. W. (1967), "A Necessary and Sufficient Condition That Ordinary Least Squares Estimators Be Best Linear Unbiased," *Journal of the American Statistical Association*, 62, 1302-1304.
- Searle, S. R. (1971). *Linear Models*, New York: Wiley.
- Seber, G. A. F. (1977), *Linear Regression Analysis*, New York: Wiley.
- Sen, A., and Srivastava, M. (1990), *Regression Analysis: Theory, Methods, and Applications*, New York: Springer-Verlag.
- Shaffer, J. P. (1981), "The Analysis of Variance Mixed Model With Allocated Observations: Application to Repeated Measurement Designs," *Journal of the American Statistical Association*, 76, 607-611.
- (1991), "The Gauss-Markov Theorem and Random Regressors," *The American Statistician*, 45, 269-273.
- 

Juliet Popper Shaffer  
University of California  
Department of Statistics  
367 Evans Hall #3860  
Berkeley, CA 94720-3860

*shaffer@stat.berkeley.edu*

Yung-Pin Chen  
Department of Mathematics  
Smith College  
Northampton, MA 01063

*ypchen@neal.smith.edu*

---

[JSE Homepage](#) | [Subscription Information](#) | [Current Issue](#) | [JSE Archive \(1993-1998\)](#) | [Data Archive](#) | [Index](#) | [Search JSE](#) |  
[JSE Information Service](#) | [Editorial Board](#) | [Information for Authors](#) | [Contact JSE](#) | [ASA Publications](#)