Inverse of a Square Matrix

For an $N \times N$ square matrix $\mathbf{A}$, the *inverse* of $\mathbf{A}$, $\mathbf{A}^{-1}$, exists if and only if $\mathbf{A}$ is of full rank, i.e., if and only if no column of $\mathbf{A}$ is a linear combination of the others. $\mathbf{A}^{-1}$ is the unique matrix that satisfies

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I} \tag{1}$$

Common Terminology

If a square matrix **A** has an inverse, we say that **A** is “invertible,” “nonsingular,” and “of full rank.”

If the transpose of a matrix is its inverse, and vice versa, i.e.,

$$\mathbf{A}\mathbf{A}' = \mathbf{A}'\mathbf{A} = \mathbf{I} \tag{2}$$

the matrix **A** is *orthogonal*.

Properties of a Matrix Inverse

$\mathbf{A}^{-1}$ has the following properties:

1. $(\mathbf{A}')^{-1} = \left(\mathbf{A}^{-1}\right)'$
2. If $\mathbf{A} = \mathbf{A}'$, then $\mathbf{A}^{-1} = \left(\mathbf{A}^{-1}\right)'$

3. The inverse of the product of several invertible square matrices is the product of their inverses in reverse order. For example

$$(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$$

4. For nonzero scalar c,

$$(c\mathbf{A})^{-1} = \frac{1}{c}\mathbf{A}^{-1}$$

5. For diagonal matrix $\mathbf{D}$, $\mathbf{D}^{-1}$ is a diagonal matrix with diagonal elements equal to the reciprocal of the corresponding diagonal elements of $\mathbf{D}$.

Consider a symmetric matrix **A.** We adopt the following definitions.

Bilinear Form

$$\mathbf{b}'\mathbf{A}\mathbf{c}$$

Quadratic Form

$$\mathbf{b}'\mathbf{A}\mathbf{b}$$

Positive Definite Symmetric Matrix A

$$\mathbf{b}'\mathbf{A}\mathbf{b} > 0 \quad \text{for all non-null } \mathbf{b}$$

Linear Independence

A set of vectors $\mathbf{x}_i$ is *linearly independent* if no linear combination of them is **0**.

Rank of a Matrix

The *rank* of a matrix is the maximum number of linearly independent rows and/or columns in the matrix.

Column Space Projectors

In homework 2, we saw that, for a vector **x**, the operator $\mathbf{P_x} = \mathbf{x}\left(\mathbf{x}'\mathbf{x}\right)^{-1}\mathbf{x}'$ and $\mathbf{Q_x} = \mathbf{I} - \mathbf{P_x}$ have a number of key properties, most of which trace back to the following:

$$\mathbf{P_x} = \mathbf{P_x}' = \mathbf{P_x}^2 \tag{1}$$

$$\mathbf{Q_x} = \mathbf{Q_x}' = \mathbf{Q_x}^2 \tag{2}$$

$$\mathbf{P_x}\mathbf{Q_x} = \mathbf{0} \tag{3}$$

$$\mathbf{P_x}\mathbf{x} = \mathbf{x},\ \mathbf{Q_x}\mathbf{x} = \mathbf{0} \tag{4}$$

The key point of the homework assignment is that $\mathbf{P_x}$ and $\mathbf{Q_x}$ can be used to decompose a vector $\mathbf{y}$ into two component vectors that are orthogonal to each other, with one component collinear with $\mathbf{x}$ and the other orthogonal to it.

Specifically, for any $\mathbf{y}$, define

$$\hat{\mathbf{y}} = \mathbf{P_x y}, \ \mathbf{e} = \mathbf{Q_x y} \tag{5}$$

Clearly, $\hat{\mathbf{y}}$ is collinear with $\mathbf{x}$, since

$$\mathbf{P_x y} = \mathbf{x}\left(\mathbf{x}'\mathbf{x}\right)^{-1}\mathbf{x}'\mathbf{y} = \mathbf{x}b \tag{6}$$

with

$$b = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}} \tag{7}$$

It also follows that

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{e} \tag{8}$$

since

$$\begin{aligned} \hat{\mathbf{y}} + \mathbf{e} &= \mathbf{P_x y} + \mathbf{Q_x y} \\ &= \mathbf{P_x y} + \left(\mathbf{I} - \mathbf{P_x y}\right) \\ &= (\mathbf{P_x} + \mathbf{I} - \mathbf{P_x})\mathbf{y} \\ &= \mathbf{I y} = \mathbf{y} \end{aligned} \tag{9}$$

and that

$$\mathbf{e}'\hat{\mathbf{y}} = 0 \tag{10}$$

since

$$\begin{aligned} \mathbf{e}'\hat{\mathbf{y}} &= \mathbf{y}'\mathbf{Q_x}'\mathbf{P_x}\mathbf{y} \\ &= \mathbf{y}'\mathbf{Q_x}\mathbf{P_x}\mathbf{y} \qquad (11) \\ &= \mathbf{y}'\mathbf{0}\mathbf{y} = 0 \end{aligned}$$

Now consider an **X** of full column rank with more than one column. Similar results to the preceding ones can be established, as follows:

First, define the *column space of* **X**, Sp(**X**), as the set of all linear combinations of the columns of **X**. Now define

$$\mathbf{P_X} = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'} \tag{12}$$

and

$$\mathbf{Q_X} = \mathbf{I} - \mathbf{P_X} \tag{13}$$

Now for any *matrix* **Y**, the columns of

$$\hat{\mathbf{Y}} = \mathbf{P_X}\mathbf{Y} \tag{14}$$

are in Sp(**X**), since

$$\begin{aligned}\hat{\mathbf{Y}} &= \mathbf{X}\left\{\left(\mathbf{X'X}\right)^{-1}\mathbf{X'Y}\right\} \\ &= \mathbf{XB}\end{aligned} \tag{15}$$

Moreover, as before, we can define $\mathbf{E} = \mathbf{Q}_{\mathbf{x}}\mathbf{Y}$ and establish results analogous to those in Equations (8)–(11).

The above results are central in linear regression.

The Algebra of Linear Regression

Suppose you have a single criterion, with *N* scores in the vector **y**. You wish to predict them as linear functions of *N* scores on *p* predictors in the matrix **X**. Assuming that all scores are in deviation score form, we have

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e} = \hat{\mathbf{y}} + \mathbf{e} \tag{16}$$

The *least squares criterion* says choose **b** to minimize the sum of squared errors $\mathbf{e}'\mathbf{e}$.

It is well known that the least squares criterion is satisfied by selecting

$$\mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \tag{17}$$

From our previous results on projectors, we see that this automatically implies that $\hat{\mathbf{y}}$ and $\mathbf{e}$ are uncorrelated, and from standard covariance algebra, we can immediately deduce that

$$S_y^2 = S_{\hat{y}}^2 + S_e^2 \tag{18}$$

The Multiple Correlation Coefficient

The *multiple correlation coefficient* is the correlation between scores in $\mathbf{y}$ and $\hat{\mathbf{y}}$. This correlation satisfies

$$r^2 = \frac{S_{\hat{y}}^2}{S_y^2} \tag{19}$$

Note that

$$\begin{aligned} S_{\hat{y}}^2 &= \mathbf{b}'\mathbf{S_{XX}b} = \mathbf{s_{yX}S_{XX}^{-1}S_{XX}S_{XX}^{-1}s_{Xy}} \\ &= \mathbf{s_{yX}S_{XX}^{-1}s_{Xy}} \end{aligned} \tag{20}$$

If all observed variables are changed to standard score form, correlations remain unchanged, but variances become unity, and it is easy to see that

$$r^2 = \mathbf{r_{yX}} \mathbf{R_{XX}^{-1}} \mathbf{r_{Xy}} \qquad (21)$$

Furthermore, if the predictors are uncorrelated, then $\mathbf{b} = \mathbf{r_{Xy}}$ and $r^2 = \mathbf{b}'\mathbf{b}$.

Regression as a Model – Fixed Regressors

We can conceptualize multiple regression as a statistical model in a variety of ways. For example, suppose we consider the scores in **X** to be fixed, and the criterion scores in **y** to be random variables, according to the following model:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e} \tag{22}$$

with the scores in **e** independent, identically distributed random variables with means of zero and variances of σ_e^2. **b** contains *population parameters*. Note that it immediately follows that

$$E(\mathbf{y}) = \mathbf{X}\mathbf{b} \tag{23}$$

Suppose that we add a column of 1's to the far left of **X** and the remaining columns in $\mathbf{X}_1$ contain the predictor scores in deviation score form. We can then write

$$\mathbf{y} = \begin{bmatrix} \mathbf{1} & \mathbf{X_1} \end{bmatrix} \begin{bmatrix} b_0 \\ \mathbf{b_1} \end{bmatrix} + \mathbf{e} \tag{24}$$

If we *estimate* **b** with the least squares principle, we obtain

$$\hat{\mathbf{b}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \tag{25}$$

One may show that

$$\hat{b}_0 = \overline{Y} \tag{26}$$

and

$$\hat{\mathbf{b}}_{\mathbf{1}} = \left(\mathbf{X}_1{}'\mathbf{X}_1\right)^{-1}\mathbf{X}_1{}'\mathbf{y} \tag{27}$$

which is why many discussions begin by re-expressing **y** in deviation score form, then dropping the intercept term b_0. In any case, it is easy to show that $\hat{\mathbf{b}}$ in Equation (25) is an unbiased estimator of $\mathbf{b}$. (C.P.)

It also follows in straightforward fashion (C.P.) that the covariance matrix for the elements of $\hat{\mathbf{b}}$ is

$$\mathrm{Var}\left(\hat{\mathbf{b}}\right) = \sigma_e^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1} \tag{28}$$

Random Multivariate Regression Model

Suppose we treat the predictors in **x** to be random variables, rather than lists of scores, and the variables in **y** to be one or more criterion variable.

We can rewrite the model as

$$\mathbf{y} = \mathbf{B}'\mathbf{x} + \mathbf{e} \tag{29}$$

In the population, the least squares criterion minimizes $\mathrm{Tr}\left\{E\left(\mathbf{ee}'\right)\right\}$, and

$$\mathbf{B} = \mathbf{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{\Sigma}_{\mathbf{xy}} \tag{30}$$

From the above, we may show (C.P.) that

$$\mathrm{Var}\left(\hat{\mathbf{y}}\right) = \mathrm{Var}\left(\mathbf{B}'\mathbf{x}\right) = \mathbf{\Sigma}_{\mathbf{yx}}\mathbf{\Sigma}_{\mathbf{xx}}^{-1}\mathbf{\Sigma}_{\mathbf{xy}} \tag{31}$$

$$\mathrm{Var}\left(\hat{\mathbf{y}}\right) = \mathrm{Cov}\left(\mathbf{y}, \hat{\mathbf{y}}\right) \tag{32}$$

$$\mathrm{Cov}\left(\hat{\mathbf{y}}, \mathbf{e}\right) = \mathbf{0} \tag{33}$$