Consider the system

$$\begin{aligned}
y_1 &= \alpha + \beta x_1 + \varepsilon_1 \\
y_2 &= \alpha + \beta x_2 + \varepsilon_2 \\
&\cdots\cdots\cdots \\
&\cdots\cdots\cdots \\
y_N &= \alpha + \beta x_N + \varepsilon_N
\end{aligned}$$

or in matrix form

$$y = X\beta^* + \varepsilon$$

where y is $N \times 1$, X is $N \times 2$, β is 2×1, and ε is $N \times 1$.

K-Variable Linear Model

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ 1 & x_N \end{bmatrix}, \ \beta^* = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}.$$

Good statistical practice requires inclusion of the column of ones.

Consider the general model

$$y = X\beta \ + \varepsilon$$

Convention: y is $N \times 1$, X is $N \times K$, β is $K \times 1$, and ε is $N \times 1$.

$$X = \begin{bmatrix} 1 & x_{21} \ldots . x_{K1} \\ 1 & x_{22} \ldots . x_{K2} \\ . & . \quad \ldots . \quad . \\ 1 & x_{2N} \ldots . x_{Kn} \end{bmatrix}, \ \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_K \end{bmatrix}.$$

A typical row looks like:
$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + ... + \beta_K x_{Ki} + \varepsilon_i$

The Least Squares Method:
First Assumption: $Ey = X\beta$
$$S(b) = (y - Xb)'(y - Xb)$$
$$= y'y - 2b'X'y + b'X'Xb$$

Normal Equations
$X'X\hat{\beta} - X'y = 0$
These equations *always* have a solution. (Clear from geometry to come)
If $X'X$ is invertible
$\hat{\beta} = (X'X)^{-1}X'y.$

More on the Linear Model

Proposition: $\hat{\beta}$ is a minimizer.
Proof: Let b be any other K-vector.

$$(y - Xb)'(y - Xb)$$
$$= (y - X\hat{\beta} + X(\hat{\beta} - b))'(y - X\hat{\beta} + X(\hat{\beta} - b))$$
$$= (y - X\hat{\beta})'(y - X\hat{\beta}) + (\hat{\beta} - b)'X'X(\hat{\beta} - b)$$
$$\geq (y - X\hat{\beta})(y - X\hat{\beta}). \text{ (Why?)}$$

Definition: $e = y - X\hat{\beta}$ is the vector of residuals.

Note: $Ee = 0$ and $X'e = 0$.

Proposition: The LS estimator is unbiased.

Proof: $E\hat{\beta} = E[(X'X)^{-1}X'y]$
$$= E[(X'X)^{-1}X'(X\beta + \varepsilon)] = \beta$$

Geometry of Least Squares

Consider $y = X\beta + \varepsilon$ with

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Definition: The space spanned by matrix X is the vector space which consists of all linear combinations of the column vectors of X.

Definition: $X(X'X)^{-1}X'y$ is the orthogonal projection of y to the space spanned by X.

Proposition: e is perpendicular to X, i.e., $X'e = 0$.

Proof:

$$e = y - X\hat{\beta} = y - X(X'X)^{-1}X'y$$
$$e = (I - X(X'X)^{-1}X')y$$
$$\Rightarrow X'e = (X' - X')y = 0. \blacksquare$$

Geometry of Least Squares (cont'd)

Thus the equation $y = X\hat{\beta} + e$ gives y as the sum of a vector in $R[X]$ and a vector in $N[X']$.

Common (friendly) projection matrices:

1. The matrix which projects to the space orthogonal to the space spanned by X (i.e. to $N[X']$ is

$$M = I - X(X'X)^{-1}X'.$$

Note: $e = My$. If X is full column rank, M has rank $(N - K)$.

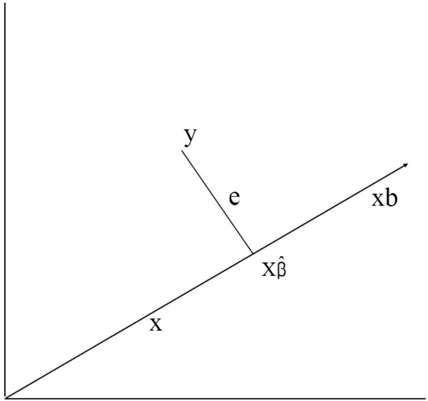
2. The matrix which projects to the space spanned by X is

$$I - M = X(X'X)^{-1}X'.$$

Note: $\hat{y} = y - e = y - My = (I - M)y$. If X is full column rank, $(I - M)$ has rank K.

Example in R^2

$$y_i = x_i\beta + \varepsilon_i \qquad i = 1, 2$$



What is the case of singular $X'X$?

Properties of projection matrices

1. Projection matrices are *idempotent*.
I.G. $(I - M)(I - M) = (I - M)$.

Proof: $(I - M)(I - M)$
$$= (X(X'X)^{-1}X')(X(X'X)^{-1}X')$$
$$= X(X'X)^{-1}X' = (I - M) \blacksquare$$

2. Idempotent matrices have eigenvalues equal to zero or one.

Proof: Consider the characteristic equation
$Mz = \lambda z \Rightarrow M^2 z = M\lambda z = \lambda^2 z$.
Since M is idempotent, $M^2 z = Mz$.
Thus, $\lambda^2 z = \lambda z$, which implies that λ is either 0 or 1. $\blacksquare$

Properties of projection matrices

3. The number of nonzero eigenvalues of a matrix is equal to its rank.
$\Rightarrow$ For idempotent matrices, trace $=$ rank.

More assumptions to the K-variable linear model:

Second assumption: $V(y) = V(\varepsilon) = \sigma^2 I_N$ where y and ε are N-vectors. With this assumption, we can obtain the sampling variance of $\hat{\beta}$.

Proposition: $V(\hat{\beta}) = \sigma^2 (X'X)^{-1}$

Proof:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon \end{aligned}$$

hence

$$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$$

cont'd

$$\begin{aligned}\mathrm{V}(\widehat{\beta}) &= \mathrm{E}[\widehat{\beta} - \mathrm{E}(\widehat{\beta})][\widehat{\beta} - \mathrm{E}(\widehat{\beta})]' = \mathrm{E}(\widehat{\beta} - \beta)(\widehat{\beta} - \beta)' \\ &= \mathrm{E}(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1} = (X'X)^{-1}X'\mathrm{E}(\varepsilon\varepsilon')X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}\end{aligned}$$

Comment: Since $\widehat{\beta}$ is an unbiased estimator of β, we have that the Mean Square Error of $\widehat{\beta}$ is

$$\mathrm{MSE}(\widehat{\beta}) = \mathrm{E}(\widehat{\beta} - \beta)(\widehat{\beta} - \beta)' = \mathrm{V}(\widehat{\beta}).$$

(Gauss-Markov) *Theorem*: $\widehat{\beta}$ is the best linear unbiased estimator (BLUE) of β, where "best" means with the minimum MSE (in the matrix sense).

Comment: An implication of the GM theorem is that for any $K-$vector $c \neq 0$ we have that $c'\widehat{\beta}$ is BLUE for $c'\beta$.

Gauss-Markov Theorem - cont'd

Proof: Consider an alternative linear unbiased estimator of β, say b. Since b is linear we have $b = A'y$. Substituting y with $X\beta + \varepsilon$ we get

$$b = A'X\beta + A'\varepsilon$$

Since b is unbiased, the equation below implies that

$$A'X = I_K$$

Hence

$$\mathrm{MSE}(b) = \mathrm{V}(b) = \mathrm{E}(A'\varepsilon\varepsilon'A) = A'\mathrm{E}(\varepsilon\varepsilon')A = \sigma^2 A'A$$

Finally, we can write

$$\begin{aligned} \mathrm{V}(b) - \mathrm{V}(\widehat{\beta}) &= \sigma^2[A'A - (X'X)^{-1}] = \sigma^2[A'A - \underbrace{A'X}_{I_K}(X'X)^{-1}\underbrace{X'A}_{I_K}] \\ &= \sigma^2 A'[I_N - X(X'X)^{-1}X']A = \sigma^2 A'MA = \sigma^2 A'M'MA \geq 0 \blacksquare \end{aligned}$$

Estimation of Variance

Proposition: $s^2 = e'e/(N-K)$ is an unbiased estimator for σ^2.

$$\begin{aligned} \textit{Proof}: \quad & e = y - X\hat{\beta} = My = M\varepsilon \Rightarrow \\ & e'e = \varepsilon' M\varepsilon \end{aligned}$$

$$\begin{aligned} Ee'e &= E\varepsilon' M\varepsilon = E \; tr \; \varepsilon' M\varepsilon \quad (Why?) \\ &= tr \; E\varepsilon' M\varepsilon = tr \; EM\varepsilon\varepsilon' \quad (important \; trick) \\ &= tr \; M \; E\varepsilon\varepsilon' = \sigma^2 \; tr \; M = \sigma^2(N-K) \end{aligned}$$

$\Rightarrow s^2 = e'e/(N-K)$ is unbiased for σ^2. ■

Fit: Does the Regression Model Explain the Data?

We will need the useful idempotent matrix
$A = I - 1(1'1)^{-1}1' = I - 11'/N$ which sweeps out means.

Here 1 is an N-vector of ones.

Note that $AM = M$ when X contains a constant term.

Definition: The correlation coefficient in the K-variable case is
$R^2 =$ (Sum of squares due to X)/(Total sum of squares)
$= 1 - (e'e/y'Ay)$.

More Fit

Using A, $y'Ay =: \sum_{i=1}^{N}(y_i - \bar{y})^2$

$$y'Ay = (Ay)'(Ay) = (A\hat{y} + Ae)'(A\hat{y} + Ae)$$
$$= \hat{y}'A\hat{y} + e'Ae \text{ since } \hat{y}'e = 0$$

Thus, $y'Ay = \hat{y}'A\hat{y} + e'e$ since $Ae = e$.

Scaling yields:

$$1 = \frac{\hat{y}'A\hat{y}}{y'Ay} + \frac{e'e}{y'Ay}$$

What are the two terms of this splitup?

More Fit

R^2 gives the fraction of variation explained by X:

$$R^2 = 1 - (e'e/y'Ay).$$

Note: The adjusted squared correlation coefficient is given by

$$\bar{R}^2 = 1 - \frac{e'e/(N-K)}{y'Ay/(N-1)}$$

(*Why might this be preferable*?)

Reporting

Always report characteristics of the sample, i.e. means, standard deviations, anything unusual or surprising, how the data set is collected and how the sample is selected.

Report $\hat{\beta}$ and standard errors (*not* t-statistics).

The usual format is

$$\begin{array}{c} \hat{\beta} \\ (s.e. \text{ of } \hat{\beta}) \end{array}$$

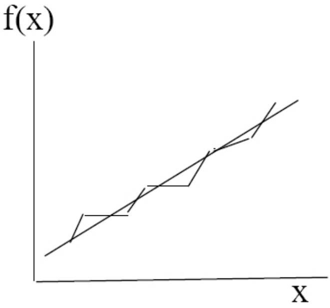
Specify S^2 or σ^2_{ML}.
Report N and R^2.

Plots are important. For example, predicted vs. actual values or predicted and actual values over time in time series studies should be presented.

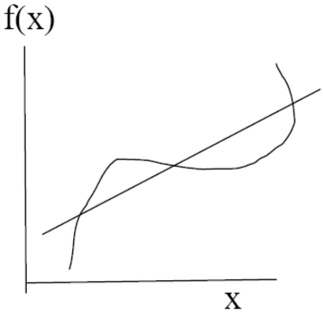
Comments on Linearity

Consider the following argument: Economic functions don't change suddenly. Therefore they are continuous. Thus they are differentiable and hence nearly linear by Taylor's Theorem.

This argument is false (but irrelevant).



Continuous, not diff, but well-approximated by a line.



Continuous, diff, and not well-approximated…