POL 213 – Spring 2024

Quantitative Analysis in Political Science II

Lecture 2

Lauren Peritz

U.C. Davis

lperitz@ucdavis.edu

April 11, 2024

Table of Contents

Least Squares Estimator

Last time, we derived the least squares estimator for a regression of Y on a single variable X.

$$Y_i = A + BX_i + Ei$$

The best linear fit is obtained by finding the A and B parameters that minimize the squared residuals, summed over all observations $\sum E_i^2$.

Let's do this numerically with simulated data. Then we can compare it to our "canned" regression estimators.

Least Squares Estimator in R

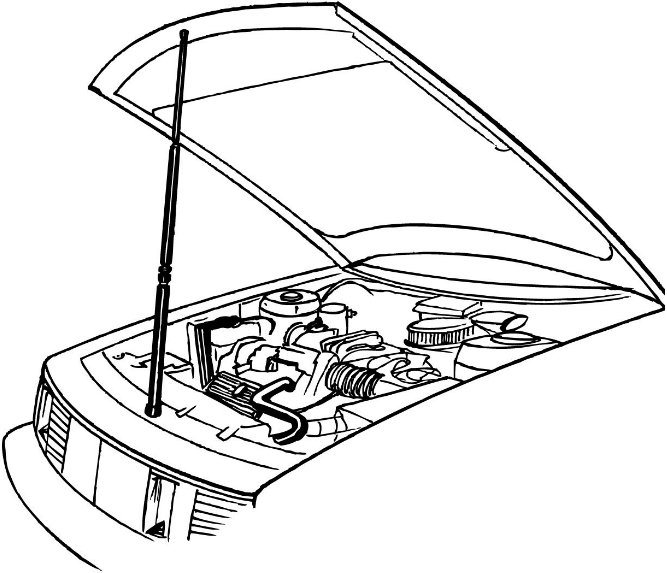
Use the R file saved to the Canvas webpage:[1]

OLS_optimization.r

[1] Thanks to Chris Hare for sharing this code.

Table of Contents

Linear Regression:
Properties of the Least Squares Estimator
(Source: Fox Chapter 6)



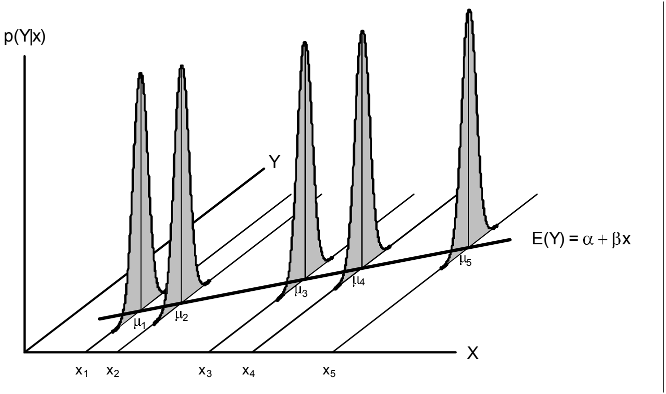
Statistical Inference for Regression

Recall the linear population regression function is:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

- The coefficients α and β are the population regression parameters to be estimated.
- The error ε_i represents the aggregated, omitted causes of Y, whether it's other explanatory variables that have been omitted, measurement error in Y, or inherently random components.

The key assumptions of the linear regression model concern the behavior of the errors, or equivalently: the distribution of Y conditional on X.



Note: The assumptions of linearity, constant variance, and normality in a simple regression. The graph shows conditional population distributions $p(Y|x)$ of Y for several values of explanatory variable X. The conditional means of Y and x are denoted $\mu_1, ...\mu_5$.

Gauss-Markov Theorem

We used the method of ordinary least squares (OLS) to fit a model to our data, solving for α and β. Is this a good model for our data?

The OLS estimator is BLUE when the Gauss-Markov assumptions are satisfied.

- B best (efficient)
- L linear
- U unbiased
- E estimator

In other words, if the Gauss-Markov assumptions are satisfied, then the OLS estimator is the one with minimum variance of sampling distribution that avoids bias.

Gauss-Markov Assumptions

- There is a weak set and a strong set of Gauss-Markov assumptions. The weak set is sufficient to satisfy the Gauss-Markov Theorem.
- To form the strong set, we add a final assumption: that the disturbances are normally distributed.
- This allows us to draw inferences about our estimates. For example, to draw inferences, we would need to build confidence intervals and conduct hypothesis tests.

Gauss-Markov Assumptions

The weak set of assumptions (enough to prove OLS is BLUE):

1. **Linearity:** The expected value of the disturbance term is 0.

$$\begin{aligned} \mu_i \equiv \mathbb{E}(Y_i) \quad \equiv \mathbb{E}(Y|x_i) \quad &= \mathbb{E}(\alpha + \beta_{xi} + \varepsilon_i) \\ &= \alpha + \beta x_i + \mathbb{E}(\varepsilon_i) \\ &= \alpha + \beta x_i + 0 \end{aligned}$$

Gauss-Markov Assumptions

The weak set of assumptions (enough to prove OLS is BLUE):

2. **Nonstochastic regressors:** The X values are independent of the error term (or X is exogenous).
   - $cov(X_i, \varepsilon_i) = 0$
   - Violations can arise because of measurement error on X, omitted confounder(s), or simultaneous causation.

Gauss-Markov Assumptions

The weak set of assumptions (enough to prove OLS is BLUE):

3. **Homoskedasticity:** constant error variance across values of X_i. Because the distribution of errors is the same as the distribution of the response variable around the population regression line, constant error variance implies constant conditional variance of Y given X.

$$
\begin{aligned}
\mathbb{V}(Y|x_i) &= \mathbb{E}[(Y_i - \mu_i)^2] \\
&= \mathbb{E}[(Y_i - \alpha - \beta x_i)^2] \\
&= \mathbb{E}(\varepsilon_i^2) \\
&= \sigma_\varepsilon^2
\end{aligned}
$$

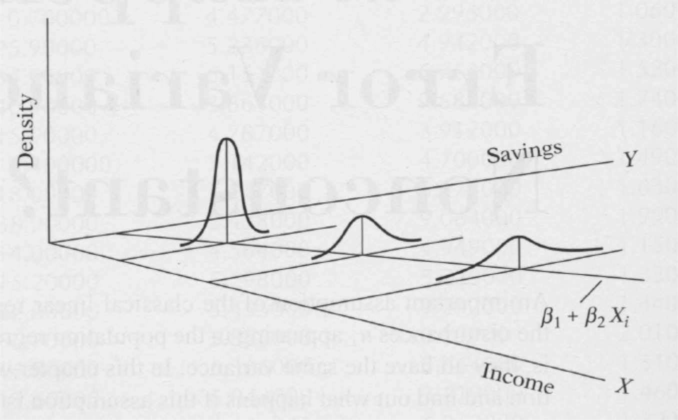
Note that because the mean of ε_i is 0, its variance is simply $\mathbb{E}(\varepsilon_i^2)$.

Gauss Markov Assumptions

Mathematical reminder: Let X be a discrete random variable with observations $x_1, ...x_n$. Then the variance of X is defined as:

$$\mathbb{V}(X) = \sigma_X^2 = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

Visualize Heteroskedasticity



Gauss-Markov Assumptions

The weak set of assumptions (enough to prove OLS is BLUE):

4. **Independence:** the observations are sampled independently. Any pair of errors ε_i and ε_j are independent for $i \neq j$
   - Violation: if there is autocorrelation between disturbances, such as in time series data.
   - Remedy with corrections (e.g. AR(1)) when you know something about the data generating process.

Gauss-Markov Assumptions

Additional (strong) assumption needed to draw statistical inferences:

5. **Normality:** the errors are normally distributed.
   - $\varepsilon_i \sim N(0, \sigma^2_\varepsilon)$
   - Equivalently, the conditional distribution of the response variable is normal: $Y_i \sim N(\alpha + \beta x_i, \sigma^2_\varepsilon)$

It means that the OLS estimator is the most efficient among *all* unbiased estimators, not just the linear unbiased estimators.

Properties of Least Squares Estimator

- The least squares intercept and slope are linear estimators
- The sample least squares coefficients are unbiased estimators of the population regression coefficients:

$$
\begin{aligned}
\mathbb{E}(A) &= \alpha \\
\mathbb{E}(B) &= \beta
\end{aligned}
$$

- Both A and B have simple sampling variances:

$$
\begin{aligned}
\mathbb{V}(A) &= \frac{\sigma_\varepsilon^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2} \\
\mathbb{V}(B) &= \frac{\sigma_\varepsilon^2}{\sum (x_i - \bar{x})^2} = \frac{\sigma_\varepsilon^2}{(n-1) S_X^2}
\end{aligned}
$$

- Under the assumption of normality, the least squares coefficients *A* and *B* are themselves normally distributed.
  - Even if the errors are not normally distributed, as the sample size grows large, *A* and *B* are approximately normal under a broad range of conditions due to CLT.

Sampling variances

We like the normal distribution for several reasons:

- By the Central Limit Theorem, the sum of a large number of independent and identically distributed random variables converges to a normal distribution.
- Ease of inference. The sum of normally distributed variables has a normal distribution, hence $\hat{\alpha}$ and $\hat{\beta}$ have normal sampling distributions.
- With a small sample size, the normality assumption allows us to use *t*, *F* and χ^2 tests in regression models.
- Many natural phenomena follow the normal distribution, it is well known, and is relatively simple with only two parameters.

Calculating standard errors

The variance of the residuals provides an unbiased estimator of σ_ε^2:

$$\hat{\sigma} = SE(E_i) = \sqrt{\frac{\sum E_i^2}{n-2}}$$

where 2 represents the number of coefficients estimated by the simple regression model.

Calculating standard errors

From this, we compute the standard error of each sample coefficient:

$$SE(A) = \hat{\sigma}\sqrt{\frac{\sum X_i^2}{n\sum(X_i - \bar{X})^2}}$$

$$SE(B) = \frac{\hat{\sigma}}{\sqrt{\sum(X_i - \bar{X})^2}}$$

With the normality assumption, we can use these standard errors to construct confidence intervals and perform hypothesis testing.

Confidence Interval & Hypothesis Test

To construct a $100(1-a)\%$ confidence interval for our slope, we take:

$$\beta = B \pm t_{a/2} SE(B)$$

where $t_{a/2}$ is the critical value of t with $n-2$ degrees of freedom and a probability of $a/2$ to the right and $SE(B)$ is the standard error of B.

To test the hypothesis $H_0 : \beta = \beta_0$, calculate the test statistic:

$$t_0 = \frac{B - \beta_0}{SE(B)}$$

Hypothesis Test

- By default, our usual software conducts the hypothesis test:

$$
\begin{aligned}
H_0 &: \quad \beta = 0 \\
H_1 &: \quad \beta \neq 0
\end{aligned}
$$

  where H_0 is the null hypothesis and H_1 is the alternative hypothesis. This is the two-tailed alternative hypothesis

- A one tailed alternative hypothesis test would be:

$$
\begin{aligned}
H_0 &: \quad \beta = 0 \\
H_1 &: \quad \beta < 0
\end{aligned}
$$

Table of Contents

Multivariate linear regression

- The main difference between bivariate and multiple linear regression: the slope coefficients for the explanatory variables in multiple regression are **partial** coefficients.
- The slope represents the effect on the response variable of a one unit increment in the corresponding explanatory variable, holding other variables constant.
- The simple (bivariate) regression slope simply ignores these other explanatory variables.

So for two explanatory variables, we now have a model that describes a plane in a three dimensional space defined by $\{X_1, X_2, Y\}$ space.

$$
\begin{aligned}
\hat{Y} &= A + B_1 X_1 + B_2 X_2 \\
E_i &= Y_i - \hat{Y}_i = Y_i - (A + B_1 X_{i1} + B_2 X_{i2})
\end{aligned}
$$

To make the plane come as close to the points as possible, we want values of A, B_1 and B_2 that minimize the sum of squared residuals.

Multivariate linear regression

As in simple regression, we find the parameters to minimize:

$$S(A, B_1, B_2) = \sum E_i^2 = \sum (Y_i - A + B_1 X_{i1} + B_2 X_{i2})^2$$

by solving a system of partial differential equations. We take the partial derivatives, set them equal to zero and obtain three linear equations with three unknowns. This (usually) allows a unique solution for the least squares coefficients (unless the variables X_1 and $X2$ are perfectly correlated.)

$$\begin{aligned}
\frac{\partial S(A, B_1, B_2)}{\partial A} &= \sum (-1)(2)(Y_i - A - B_1 X_{i1} - B_2 X_{i2}) \\
\frac{\partial S(A, B_1, B_2)}{\partial B_1} &= \sum (-X_{i1})(2)(Y_i - A - B_1 X_{i1} - B_2 X_{i2}) \\
\frac{\partial S(A, B_1, B_2)}{\partial B_2} &= \sum (-X_{i2})(2)(Y_i - A - B_1 X_{i1} - B_2 X_{i2})
\end{aligned}$$

We refer to Fox Chapter 5.2 for an example.

Figure 5.6 The multiple-regression plane, showing the partial slopes B_1 and B_2 and the residual E_i for the ith observation. The white dot in the regression plane represents the fitted value. Compare this graph with Figure 5.2 for simple regression.

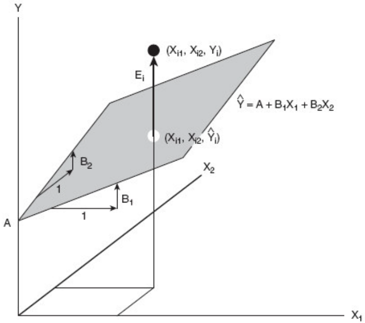
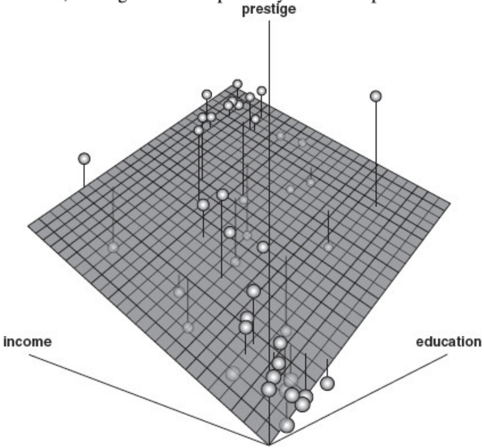


Figure 5.7 The multiple-regression plane in Duncan's regression of prestige on education and income. The two sets of parallel lines on the regression plane represent the partial relationship of prestige to each explanatory variable, holding the other explanatory variable at particular values.



Multiple Correlation

As in simple regression, the residual standard error in multiple regression measures the average size of the residuals. As before, we divide bt the degrees of freedom $n - k - 1$, so the standard error of the regression is:

$$S_E = \sqrt{\frac{\sum E_i^2}{n-k-1}}$$

The sum of squares in the multiple regression are defined the same as in the simple case:

$$\begin{aligned} TSS &= \sum(Y_i - \bar{Y})^2 \\ RegSS &= \sum(\hat{Y}_i - \bar{Y})^2 \\ RSS &= \sum(Y_i - \hat{Y})^2 = \sum E_i^2 \end{aligned}$$

Multiple Correlation

The R^2 multiple correlation represents the proportion of variation in the response variable captured by the regression:

$$R^2 \equiv \frac{RegSS}{TSS}$$

And the Adjusted R^2 is given by:

$$\tilde{R}^2 = 1 - \frac{\frac{RSS}{n-k-1}}{\frac{TSS}{n-1}}$$

Example with Prestige Data

For an example, let's turn to the `Prestige.txt` data and a linear regression.

We have the following variables for the Canadian occupation data:

- Prestige = some probably problematic measure of how socially esteemed a job is (arbitrary scale)
- Education = average education for that occupation
- Income = average income for that occupation ($)
- Women = percentage of women in that occupation

OLS_example.r

Multivariate linear regression - General Case

Obviously, the more variables we include in our regression, the more difficult our calculation becomes and it becomes impossible to visualize. For the case of k explanatory variables, the task is to find the A and Bs that minimize the sum of squared residuals:

$$S(A, B_1, B_2, \ldots, B_k) = \sum_{i=1}^{n} [Y_i - (A + B_1 X_{i1} + B_2 X_{i2} + \cdots + B_k X_{ik})]^2$$

The minimization of the sum-of-squares function produces the normal k equations with k unknowns. If none of our variables are perfectly collinear (i.e. X_j cannot be recreated by a combination of other $X_{i \neq j}$), then our system of normal equations has a unique solution.

$$\begin{aligned}
An + B_1 \sum X_{i1} + B_2 \sum X_{i2} + \ldots + B_k \sum X_{ik} &= \sum Y_i \\
A \sum X_{i1} + B_1 \sum X_{i1}^2 + B_2 \sum X_{i1} X_{i2} + \ldots + B_k \sum X_{i1} X_{ik} &= \sum X_{i1} Y_i \\
A \sum X_{i2} + B_1 \sum X_{i1} X_{i2} + B_2 \sum X_{i2}^2 + \ldots + B_k \sum X_{i2} X_{ik} &= \sum X_{i2} Y_i \\
&\vdots \\
A \sum X_{ik} + B_1 \sum X_{i1} X_{ik} + B_2 \sum X_{i2} X_{ik} + \ldots + B_k \sum X_{ik}^2 &= \sum X_{ik} Y_i
\end{aligned}$$

Sidebar on multicollinearity!

Multivariate Linear Regression

We cannot write out a general solution to the normal equations without specifying the number of explanatory variables k and even for k as small as 3, the solution would be very complicated.

But because the normal equations are linear and because there are as many equations as unknown regression coefficients $(k+1)$, there is usually a unique solution for the coefficients $A, B_1, B_2, ..., B_k$. Dividing the first normal equation through by n reveals that the least-squares surface passes through the point of means $(\bar{X}_1, \bar{X}_2, ..., \bar{X}_k)$.

Clearly we need a better way to deal with this!

Table of Contents

Linear Model with Matrix Algebra

We showed how adding just a few X variables to our linear regression model makes a lot of cumbersome mathematical expressions.

Matrix algebra helps us consolidate our notation and employ a number of tricks to produce the OLS model and work with complicated dataframes. It also gives us the perspective of a geometrical representation of the least squares fit.

Visualize LS Fit as Projection

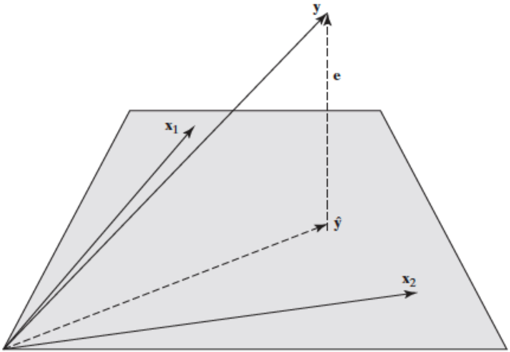


FIGURE 3.2 Projection of **y** into the Column Space of **X**.

Source: Greene 2012, p.72.

Linear Model in Matrix Form

Let's tweak our notation for the general linear model:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \varepsilon_i$$

We can collect our regressors into a row vector, append a 1 for the constant, and place the parameters into a column vector to rewrite this as:

$$Y_i = [1, x_{i1}, ...x_{ik}] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \varepsilon_i$$

$$Y_i = \underset{(1 \times k+1)}{\boldsymbol{x}_i'} \quad \underset{(k+1 \times 1)}{\beta} \quad + \varepsilon_i$$

This was for a single observation i. What about for a sample of n observations?

Linear Model in Matrix Form

For n observations, we have n such equations that can be stacked up and organized. Our system of linear equations can be rewritten in matrix form:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\underset{(n \times 1)}{\boldsymbol{y}} = \underset{(n \times k+1)}{\boldsymbol{X}} \; \underset{(k+1 \times 1)}{\beta} + \underset{(n \times 1)}{\varepsilon}$$

Linear Model in Matrix Form

Let's restate the assumptions of the linear model in matrix form.

Note that the error term ε is a vector random variable. Errors are assumed to be independent and normally distributed with zero expectation and common variance.

Thus ε follows a multivariate normal distribution with expectation $\mathbb{E}(\varepsilon) = \mathbf{0}$ and covariance matrix $\mathbb{V}(\mathbb{E}(\varepsilon\varepsilon')) = \sigma_\varepsilon^2 \mathbf{I}_n$, where $\mathbf{I}$ is the n by n identity matrix.

The assumptions of the linear model can be written compactly as:

$$\varepsilon \overset{\mathrm{d}}{\sim} N_n(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_n)$$

Linear Model in Matrix Form

It follows that the dependent variable **y** is also normally distributed. The expectation of $\boldsymbol{y}$ is:

$$\begin{aligned} \mathbb{E}(\boldsymbol{y}) &= \mathbb{E}(\boldsymbol{X}\beta + \varepsilon) \\ &= \boldsymbol{X}\beta + \mathbb{E}(\varepsilon) \\ &= \boldsymbol{X}\beta \end{aligned}$$

The variance of $\boldsymbol{y}$ is:

$$\begin{aligned} \mathbb{V}(\boldsymbol{y}) &= \mathbb{E}[(\boldsymbol{y} - \mathbb{E}(\boldsymbol{y}))(\boldsymbol{y} - \mathbb{E}(\boldsymbol{y}))'] \\ &= \mathbb{E}[(\boldsymbol{y} - \boldsymbol{X}\beta)(\boldsymbol{y} - \boldsymbol{X}\beta)'] \\ &= \mathbb{E}(\varepsilon\varepsilon') \\ &= \sigma_\varepsilon^2 \boldsymbol{I} \end{aligned}$$

So the distribution of $\boldsymbol{y}$ is:

$$\Rightarrow \boldsymbol{y} \overset{\mathrm{d}}{\sim} N_n(\mathbf{X}\beta, \sigma_\varepsilon^2 \mathbf{I}_n)$$

Matrix Algebra...

We will use this notation and some other concepts from linear algebra to derive the OLS estimator. Matrix notation gives us the ability to express complicated models and translate them directly into data manipulations performed in **R**. What concepts from linear algebra do we need to use?

- Vector, scalar, matrix – as objects comprised of data
- Matrix addition, multiplication
- Transpose, inverse
- Conformability
- Identity matrix $\boldsymbol{I}_{n \times n}$
- Rank of matrix; non-singular matrix; positive-definite matrix
- Projection (next time)

Reminder, methods track students should fully understand all the moving parts here. Everyone else should work toward developing a general intuition but need not be able to do the math themselves.

Matrix Algebra...

Useful information on matrix algebra can be found here:

Appendices to Fox Textbook

Why should I know this?

- Open the black box of regression, **THE** core tool you will use in empirical political science.
- Some common and avoidable research mistakes arise because people do not understand how and why the models work, leading to poor analysis choices and erroneous findings.
- If you understand how and why, you don't have to memorize long (forgettable) checklists of best practices.
- Methods track students: Mastery of the regression model is one important requirement for passing the methods comprehensive exams.

Least-Squares Fit with Matrix Algebra

To find the least squares coefficients, we write the fitted linear model as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e}$$

where $\boldsymbol{b} = [B_0, B_1, ..., B_k]'$ is the vector of fitted coefficients and $\boldsymbol{e} = [E_0, E_1, ..., E_n]'$ is the vector of residuals. We seek the coefficient vector $\boldsymbol{b}$ that minimizes the RSS.

$$\begin{aligned} S(\boldsymbol{b}) &= \sum E_i^2 = \boldsymbol{e}\boldsymbol{e}' = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) \\ &= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\boldsymbol{b} - \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} \\ &= \boldsymbol{y}'\boldsymbol{y} - (2\boldsymbol{y}'\boldsymbol{X})\boldsymbol{b} + \boldsymbol{b}'(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{b} \end{aligned}$$

(Note on commutative property.)

Least-Squares Fit with Matrix Algebra

Take derivative with respect to the coefficient vector $\boldsymbol{b}$ and set it equal to zero. Note that we can do this in one compact equation rather than solving partial differentials for each α and β as before.

$$\begin{aligned}\frac{\partial S(\boldsymbol{b})}{\partial \boldsymbol{b}} &= \frac{\partial}{\partial \boldsymbol{b}}(\boldsymbol{y}'\boldsymbol{y} - (2\boldsymbol{y}'\boldsymbol{X})\boldsymbol{b} + \boldsymbol{b}'(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{b}) \\ \mathbf{0} &= \mathbf{0} - 2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b}\end{aligned}$$

This means there are $k+1$ normal equations in the same number of unknown coefficients. Now solve for $\boldsymbol{b}$ (divide by scalar 2, premultiply by $(\boldsymbol{X}'\boldsymbol{X})^{-1}$):

$$\begin{aligned}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} &= \boldsymbol{X}'\boldsymbol{y} \\ \boldsymbol{b} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}\end{aligned}$$

Least-Squares Fit with Matrix Algebra

We require a couple things to uniquely solve for the $\boldsymbol{b}$ least squares coefficients.

- Because the rank of $\boldsymbol{X}$ can be no greater than the smaller of n and $k+1$, we require at least as many observations (n) as there are coefficients in the model ($k+1$). *Why?*
- The $k+1$ columns of $\boldsymbol{X}$ must be linearly independent. *Why?*

Check that this is a minimum, not a maximum, by looking at the sign of the second derivative.

$$\frac{\partial^2 S(\boldsymbol{b})}{\partial \boldsymbol{b} \partial \boldsymbol{b}'} = 2\boldsymbol{X}'\boldsymbol{X}$$

Because $\boldsymbol{X}'\boldsymbol{X}$ is positive-definite when $\boldsymbol{X}$ is of full rank, the solution $\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ represents a *minimum* of $S(\boldsymbol{b})$.[2]

[2] A square matrix such as $\boldsymbol{X}'\boldsymbol{X}$ is called positive definite if it is symmetric and all its eigenvalues λ are positive. This is satisfied because, roughly speaking, none of $\boldsymbol{X}$'s columns can be recreated by combinations of other columns and there are more observations than variables, so we can actually fit a single hyperplane through the data.

Translating Notation

Table 9.1 Comparison Between Simple Regression Using Scalars and Multiple Regression Using Matrices

	Simple Regression	*Multiple Regression*
Model	$Y = \alpha + \beta x + \varepsilon$	$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
Least-squares estimator	$B = \frac{\sum x^* Y^*}{\sum x^{*2}}$	$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$
	$= (\sum x^{*2})^{-1} \sum x^* Y^*$	
Sampling variance	$V(B) = \frac{\sigma_\varepsilon^2}{\sum x^{*2}}$	$V(\mathbf{b}) = \sigma_\varepsilon^2(\mathbf{X}'\mathbf{X})^{-1}$
	$= \sigma_\varepsilon^2(\sum x^{*2})^{-1}$	
Distribution	$B \sim N[\beta, \sigma_\varepsilon^2(\sum x^{*2})^{-1}]$	$\mathbf{b} \sim N_{k+1}[\boldsymbol{\beta}, \sigma_\varepsilon^2(\mathbf{X}'\mathbf{X})^{-1}]$

NOTE: Subscripts are suppressed in this table; in particular, $x^* \equiv x_i - \overline{x}$ and $Y^* \equiv Y_i - \overline{Y}$.

SOURCE: Adapted from Wonnacott and Wonnacott (1979, Table 12-1), *Econometrics, Second Edition*.

Summary

What we did today:

1. Derived the least squares estimator and showed in R how OLS chooses slope and intercept to minimize the RSS.

Summary

What we did today:

1. Derived the least squares estimator and showed in R how OLS chooses slope and intercept to minimize the RSS.
2. Explained the Gauss Markov assumptions and how they can enable us to use our least squares model for statistical inference.

Summary

What we did today:

1. Derived the least squares estimator and showed in R how OLS chooses slope and intercept to minimize the RSS.
2. Explained the Gauss Markov assumptions and how they can enable us to use our least squares model for statistical inference.
3. Extended the model to multiple variables and showed how solving a system of linear equations gets really complicated really fast!

Summary

What we did today:

1. Derived the least squares estimator and showed in R how OLS chooses slope and intercept to minimize the RSS.
2. Explained the Gauss Markov assumptions and how they can enable us to use our least squares model for statistical inference.
3. Extended the model to multiple variables and showed how solving a system of linear equations gets really complicated really fast!
4. Expressed the equations compactly using matrix notation.

Summary

What we did today:

1. Derived the least squares estimator and showed in R how OLS chooses slope and intercept to minimize the RSS.
2. Explained the Gauss Markov assumptions and how they can enable us to use our least squares model for statistical inference.
3. Extended the model to multiple variables and showed how solving a system of linear equations gets really complicated really fast!
4. Expressed the equations compactly using matrix notation.
5. Rederived the OLS estimator using matrix notation so that it applies to many X variables.

Next time

- Problem set 1 due April 17.
- More on properties of least squares estimator and statistical inference for multivariate linear models