POLS 509: The Linear Model, *Lecture # 4*

Eric Reinhardt

Feb 7, 2005

The Regression Model. Estimation and Inference.

Feb 8 & 10: Week 4. The regression model. Estimation and inference.

- Wooldridge, 21-89, 95-187, 197-198.
- Hagle, 54-56.

Outline

1) Possible approaches to estimating a univariate regression model
2) A simple, 2-point example
3) The general case, with n observations
   Deriving the "normal" equations
   Some useful facts about the OLS estimators
   Interpretation
4) Real-world example: the civil rights voting records of U.S. Supreme Court justices
5) Variability in the estimators
   Deriving the $\text{SE}_{\widehat{\beta}}$s
   Some properties
6) The OLS estimators are BLUE
   Proof ***
7) Inference
   Distributional assumption about errors
   Standardizing $\widehat{\beta}$, with an estimate of the population error variance
   The regression t-statistic
   Interval estimates and hypothesis testing
8) Regression predictions
   The point prediction

SE of the point prediction
CIs of the prediction
Stata PREDICT options... ***

9) The Supreme Court civil rights voting record example revisited...

1 Derivation of OLS Estimators

1.0.1 Step 1: Statement of the problem.

The challenge: Find the $\widehat{\beta}_0$ and $\widehat{\beta}_1$ that minimize the sum of squared residuals,

$$S = \sum_{i=1}^{n} u_i^2.$$

>From the original regression equation, $y_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i + u_i$, we can rewrite this as

$$S = \sum_{i=1}^{n} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i \right)^2,$$

which, expanded, yields

$$\begin{aligned} S &= \sum_{i=1}^{n} \left(y_i^2 - y_i\widehat{\beta}_0 - y_i\widehat{\beta}_1 x_i - \widehat{\beta}_0 y_i + \widehat{\beta}_0^2 + \widehat{\beta}_0\widehat{\beta}_1 x_i - \widehat{\beta}_1 x_i y_i + \widehat{\beta}_1 x_i \widehat{\beta}_0 + \widehat{\beta}_1^2 x_i^2 \right) \\ &= \sum_{i=1}^{n} \left(y_i^2 - 2y_i\widehat{\beta}_0 - 2y_i\widehat{\beta}_1 x_i + 2\widehat{\beta}_0\widehat{\beta}_1 x_i + \widehat{\beta}_0^2 + \widehat{\beta}_1^2 x_i^2 \right) \end{aligned}$$

1.0.2 Step 2: First-order conditions for minimization.

FOC 1:

$$\frac{\partial S}{\partial \widehat{\beta}_0} = \sum_{i=1}^{n} \left(-2y_i + 2\widehat{\beta}_1 x_i + 2\widehat{\beta}_0 \right) = -2\sum_{i=1}^{n} \left(y_i - \widehat{\beta}_1 x_i - \widehat{\beta}_0 \right) = 0.$$

Note that this happens to be the same as $-2\sum_{i=1}^{n} u_i = 0$.

FOC 2:

$$\frac{\partial S}{\partial \widehat{\beta}_1} = \sum_{i=1}^{n}\left(-2y_i x_i + 2\widehat{\beta}_0 x_i + 2\widehat{\beta}_1 x_i^2\right) = -2\sum_{i=1}^{n} x_i\left(y_i - \widehat{\beta}_1 x_i - \widehat{\beta}_0\right) = 0.$$

1.0.3 Step 3: Re-arrange, to find the "normal equations."

>From FOC 1...

$$\begin{aligned}
-2\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_1 x_i - \widehat{\beta}_0\right) &= 0, \\
\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_1 x_i - \widehat{\beta}_0\right) &= 0, \\
\sum_{i=1}^{n} y_i - \widehat{\beta}_1\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\widehat{\beta}_0 &= 0,
\end{aligned}$$

so NE 1:

$$\begin{aligned}
\sum_{i=1}^{n} y_i &= n\widehat{\beta}_0 + \widehat{\beta}_1\sum_{i=1}^{n} x_i, \text{ or if divided by } n\text{...} \\
\overline{y} &= \widehat{\beta}_0 + \widehat{\beta}_1\overline{x} \text{ (see Wooldridge p. 28, eqn. 2-16)}
\end{aligned}$$

>From FOC 2...

$$\begin{aligned}
-2\sum_{i=1}^{n} x_i\left(y_i - \widehat{\beta}_1 x_i - \widehat{\beta}_0\right) &= 0, \\
\sum_{i=1}^{n} x_i y_i - \widehat{\beta}_1\sum_{i=1}^{n} x_i^2 - \widehat{\beta}_0\sum_{i=1}^{n} x_i &= 0,
\end{aligned}$$

hence NE 2:

$$\sum_{i=1}^{n} x_i y_i = \widehat{\beta}_0\sum_{i=1}^{n} x_i + \widehat{\beta}_1\sum_{i=1}^{n} x_i^2.$$

1.0.4 Step 4: Solve for $\widehat{\beta}_0$ in NE 1.

Multiply both sides by $\frac{1}{n}$, yielding

$$\frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\left(n\widehat{\beta}_0\right) + \frac{1}{n}\left(\widehat{\beta}_1\sum_{i=1}^{n} x_i\right),$$

which simplifies to

$$\overline{y} = \widehat{\beta}_0 + \widehat{\beta}_1\overline{x}.$$

The first unknown, $\widehat{\beta}_0$, thus is

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1\overline{x}.$$

1.0.5 Step 5: Solve for $\widehat{\beta}_1$ in NE 2, by substitution.

From NE 2, substituting in the expression for $\widehat{\beta}_0$ from above, we have

$$\sum_{i=1}^{n} x_i y_i = \left(\overline{y} - \widehat{\beta}_1\overline{x}\right)\sum_{i=1}^{n} x_i + \widehat{\beta}_1\sum_{i=1}^{n} x_i^2,$$

or

$$\begin{aligned}
\sum_{i=1}^{n} x_i y_i &= \left(\overline{y} - \widehat{\beta}_1\overline{x}\right) n\left(\frac{1}{n}\right)\sum_{i=1}^{n} x_i + \widehat{\beta}_1\sum_{i=1}^{n} x_i^2, \\
\sum_{i=1}^{n} x_i y_i &= \left(\overline{y} - \widehat{\beta}_1\overline{x}\right) n\overline{x} + \widehat{\beta}_1\sum_{i=1}^{n} x_i^2, \\
\sum_{i=1}^{n} x_i y_i &= n\overline{x}\overline{y} - n\widehat{\beta}_1\overline{x}^2 + \widehat{\beta}_1\sum_{i=1}^{n} x_i^2.
\end{aligned}$$

Pause here, and note the following two equalities:
(a)

$$\sum_{i=1}^{n} x_i^2 - n\overline{x}^2 = \sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2,$$

because

$$
\begin{aligned}
\sum_{i=1}^{n}(x_i-\overline{x})^2 &= \sum_{i=1}^{n}\left(x_i^2-2x_i\overline{x}+\overline{x}^2\right) \\
&= \sum_{i=1}^{n}x_i^2-2\overline{x}\sum_{i=1}^{n}x_i+\sum_{i=1}^{n}\overline{x}^2 \\
&= \sum_{i=1}^{n}x_i^2-2\overline{x}\left(n\frac{1}{n}\right)\sum_{i=1}^{n}x_i+n\overline{x}^2 \\
&= \sum_{i=1}^{n}x_i^2-2n\overline{x}^2+n\overline{x}^2 \\
&= \sum_{i=1}^{n}x_i^2-n\overline{x}^2.
\end{aligned}
$$

Through a similar trick,
(b)

$$\sum_{i=1}^{n}x_iy_i-n\overline{xy}=\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y}),$$

(It's only different by virtue of changing one of the x_i's to a y_i in (a) above, more or less.)

Back to the solution from NE 2, above, we had

$$\sum_{i=1}^{n}x_iy_i=n\overline{xy}-n\widehat{\beta}_1\overline{x}^2+\widehat{\beta}_1\sum_{i=1}^{n}x_i^2.$$

Factoring out $\widehat{\beta}_1$, we have

$$\sum_{i=1}^{n}x_iy_i=\widehat{\beta}_1\left(\sum_{i=1}^{n}x_i^2-n\overline{x}^2\right)+n\overline{xy},$$

and if we substitute in equality (a) above, we get

$$\sum_{i=1}^{n}x_iy_i=\widehat{\beta}_1\sum_{i=1}^{n}(x_i-\overline{x})^2+n\overline{xy}.$$

Solving for $\widehat{\beta}_1$, we step to

$$\widehat{\beta}_1 \sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i y_i - n\overline{xy},$$

which, after subbing the expression from equality (b) in for the RHS of this, becomes

$$\widehat{\beta}_1 \sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}).$$

Now we divide, and, voila!,

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}.$$

As Darth Vader would say, "Impressive, my student, but you are not a Jedi yet."

Just to restate from the first part of the solution,

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x},$$

which we write with $\widehat{\beta}_1$ still in, for convenience. These are the OLS estimators for β_0 and β_1, in a model with a constant and just one explanatory variable (i.e., univariate regression).

1.0.6 Step 6: How do we know this is a minimum and not a maximum?

For this to be a minimum, we must take the second-order partials and use the discriminant. Namely, it must be true that

$$\frac{\partial^2 S}{\partial \widehat{\beta}_0^2} \times \frac{\partial^2 S}{\partial \widehat{\beta}_1^2} - \left(\frac{\partial^2 S}{\partial \widehat{\beta}_0 \widehat{\beta}_1} \right)^2 > 0$$

and

$$\frac{\partial^2 S}{\partial \widehat{\beta}_1^2} \text{ or } \frac{\partial^2 S}{\partial \widehat{\beta}_0^2} > 0.$$

Using the first partial derivatives in the normal equations above, we see that

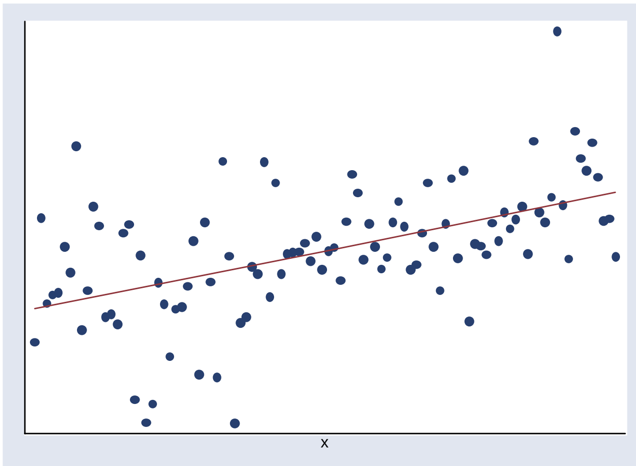
$$\begin{aligned}
\frac{\partial^2 S}{\partial \widehat{\beta}_0^2} &= \frac{\partial \left[-2\sum_{i=1}^{n} \left(y_i - \widehat{\beta}_1 x_i - \widehat{\beta}_0 \right) \right]}{\partial \widehat{\beta}_0} = 2n \\
\frac{\partial^2 S}{\partial \widehat{\beta}_1^2} &= \frac{\partial \left[-2\sum_{i=1}^{n} x_i \left(y_i - \widehat{\beta}_1 x_i - \widehat{\beta}_0 \right) \right]}{\partial \widehat{\beta}_1} = 2\sum_{i=1}^{n} x_i^2 \\
\text{and } \frac{\partial^2 S}{\partial \widehat{\beta}_0 \widehat{\beta}_1} &= 2\sum_{i=1}^{n} x_i.
\end{aligned}$$

The discriminant is

$$\begin{aligned}
& \frac{\partial^2 S}{\partial \widehat{\beta}_0^2} \times \frac{\partial^2 S}{\partial \widehat{\beta}_1^2} - \left(\frac{\partial^2 S}{\partial \widehat{\beta}_0 \widehat{\beta}_1} \right)^2 \\
&= \left[2n \left(2\sum_{i=1}^{n} x_i^2 \right) \right] - \left[2\sum_{i=1}^{n} x_i \right]^2 \\
&= 4n\sum_{i=1}^{n} x_i^2 - 4\left(\sum_{i=1}^{n} x_i \right)^2.
\end{aligned}$$

If this is strictly positive, then, since $\frac{\partial^2 S}{\partial \widehat{\beta}_0^2} > 0$, we have a minimum.

For an alternative approach, consider the hypothetical regression graph in the scatterplot below:



The regression line provides a "best fit" to the data, and it uses the OLS estimators for the y-intercept ($\widehat{\beta}_0$) and slope ($\widehat{\beta}_1$). What happens to S, the sum of squared residuals, as we shift the line away from this intercept and slope? What line position, in particular, would *maximize* S? Could we increase S by shifting the slope far, far up the y-axis? Certainly, and we could do so infinitely; likewise with the line slope. There is no critical point yielding a maximum for S, in other words. Hence our critical point must be yielding a minimum.

Now, my student, you have become a Jedi.

2 Some Useful and Interesting Facts

1. From the first 'normal equation', we can show that the sum of the error terms, and thus likewise the mean error, is equal to zero.

2. Also from the first, as rearranged above, we had

$$\overline{y} = \widehat{\beta}_0 + \widehat{\beta}_1 \overline{x}.$$

This implies that all OLS-derived lines must pass through the point of means, $(\overline{x}, \overline{y})$.

3. From the second NE, we can show that $\sum_{i=1}^{n} x_i u_i = 0$, which implies that the sample covariance between the regressor (x) and the error term (u) is zero.

4. You have seen the components of $\widehat{\beta}_1$ before...

Since $Cov(x,y) = E\left[(x-\mu_x)(y-\mu_y)\right] = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})$, and since $Var(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$, then because $\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}$,

$$\widehat{\beta}_1 = \frac{Cov(x,y)}{Var(x)}.$$

(See Wooldridge p. 29.)

3 Estimates of Variability

We would like to use our estimators of the βs to infer something about the *population* of interest, not merely to characterize the line that best fits the *sample's* data. (This is why we call them "estimators.") To start with, let's re-write the formula for our $\widehat{\beta}_1$ estimator in the simple regression, $y_i = \beta_0 + \beta_1 x_i + u_i$:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}.$$

Note that the numerator of this expression,

$$
\begin{aligned}
& \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) \\
= \; & \sum_{i=1}^{n} (x_i y_i - x_i \overline{y} - \overline{x} y_i + \overline{xy}) \\
= \; & \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \overline{y} - \sum_{i=1}^{n} \overline{x} y_i + \sum_{i=1}^{n} \overline{xy} \\
= \; & \sum_{i=1}^{n} x_i y_i - n\overline{xy} - \sum_{i=1}^{n} \overline{x} y_i + n\overline{xy} \\
= \; & \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} \overline{x} y_i = \sum_{i=1}^{n} (x_i y_i - \overline{x} y_i) \\
= \; & \sum_{i=1}^{n} (x_i - \overline{x}) y_i .
\end{aligned}
$$

Whoa! So

$$
\widehat{\beta}_1 = \frac{\sum\limits_{i=1}^{n} (x_i - \overline{x}) y_i}{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2}.
$$

Let's plug the regression equation back in for y_i, transforming this into

$$
\widehat{\beta}_1 = \frac{\sum\limits_{i=1}^{n} (x_i - \overline{x}) (\beta_0 + \beta_1 x_i + u_i)}{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2},
$$

noting that, on the top part,

$$
\begin{aligned}
& \sum_{i=1}^{n} (x_i - \overline{x}) (\beta_0 + \beta_1 x_i + u_i) \\
= & \sum_{i=1}^{n} [(x_i - \overline{x}) \beta_0 + (x_i - \overline{x}) \beta_1 x_i + (x_i - \overline{x}) u_i] \\
= & \beta_0 \sum_{i=1}^{n} (x_i - \overline{x}) + \beta_1 \sum_{i=1}^{n} (x_i - \overline{x}) x_i + \sum_{i=1}^{n} (x_i - \overline{x}) u_i.
\end{aligned}
$$

Given that $\sum_{i=1}^{n} (x_i - \overline{x}) = 0$ and $\sum_{i=1}^{n} (x_i - \overline{x}) x_i = \sum_{i=1}^{n} (x_i - \overline{x})^2$, our estimator then becomes

$$
\begin{aligned}
\widehat{\beta}_1 &= \frac{\beta_1 \sum_{i=1}^{n} (x_i - \overline{x})^2 + \sum_{i=1}^{n} (x_i - \overline{x}) u_i}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \\
&= \beta_1 + \frac{\sum_{i=1}^{n} (x_i - \overline{x}) u_i}{\sum_{i=1}^{n} (x_i - \overline{x})^2}.
\end{aligned}
$$

This makes a very clear point: our estimate is the sum of two components, one being the true population parameter β_1, and the other being a chance or random element based on u_i. After all, each time we draw a different sample, we get a somewhat different estimate of the same population parameter, since the next sample itself is somewhat different than the original. The estimator is thus a random variable, governed by its own sampling distribution. Fine, but what **is** that distribution?

(See Wooldridge p. 56.) Let's call the variance of $\widehat{\beta}_1$'s sampling distribution $Var(\widehat{\beta}_1)$. From our work with expected values and variances, we know that the variance of a constant is zero, so we can zero out all the nonrandom components when taking the variance of $\widehat{\beta}_1$. That leaves

$$
\begin{aligned}
Var(\widehat{\beta}_1) &= \frac{Var\left(\sum_{i=1}^{n}(x_i-\overline{x})\,u_i\right)}{\left[\sum_{i=1}^{n}(x_i-\overline{x})^2\right]^2} \\
&= \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2\,Var(u_i)}{\left[\sum_{i=1}^{n}(x_i-\overline{x})^2\right]^2}.
\end{aligned}
$$

We assume in ordinary regression that $Var(u_i)=\sigma^2$, a constant. Hence

$$
\begin{aligned}
Var(\widehat{\beta}_1) &= \frac{\sigma^2\sum_{i=1}^{n}(x_i-\overline{x})^2}{\left[\sum_{i=1}^{n}(x_i-\overline{x})^2\right]^2} \\
&= \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}.
\end{aligned}
$$

That is the variance of $\widehat{\beta}_1$'s sampling distribution. Analogously,

$$
Var(\widehat{\beta}_0)=\frac{\sigma^2\sum_{i=1}^{n}x_i^2}{n\sum_{i=1}^{n}(x_i-\overline{x})^2}.
$$

To summarize, the sampling distributions for the regression estimates are:

$$
\begin{aligned}
&\widehat{\beta}_0\,\tilde{}\,N\left[\beta_0,\frac{\sigma^2\sum x_i^2}{n\sum(x_i-\overline{x})^2}\right] \\
&\widehat{\beta}_1\,\tilde{}\,N\left[\beta_1,\frac{\sigma^2}{\sum(x_i-\overline{x})^2}\right].
\end{aligned}
$$

This looks good but doesn't help us in practice, since we do not actually know the true **population** variance of the **error** term u_i. We know the variance of the **residuals** in the **sample**, but it is important to note the distinction.

Wooldridge pps. 57-58 shows that an unbiased estimator of σ^2, $\widehat{\sigma}^2$, is $\frac{1}{n-2}\sum_{i=1}^{n}\widehat{u}_i^2$, which is nearly the sample variance of the (sample) residuals. Plug this back in and we get

$$Var(\widehat{\beta}_1) = \frac{\frac{1}{n-2}\sum_{i=1}^{n}\widehat{u}_i^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}.$$

The square root of this yields the **standard error of** $\widehat{\beta}_1$:

$$\begin{aligned} \text{se}\left(\widehat{\beta}_1\right) &= \sqrt{\frac{\frac{1}{n-2}\sum_{i=1}^{n}\widehat{u}_i^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}}, \text{ or just} \\ &= \frac{\widehat{\sigma}}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}. \end{aligned}$$