The Simple Linear Regression Model

David A. Hughes, Ph.D.

Auburn University at Montgomery

david.hughes@aum.edu

September 23, 2021

Introduction

- Previously, we examined bivariate relationships by assessing directionality, strength of association, and statistical significance.
- Our approaches to these tasks, however, told us little about how discrete changes in X might affect Y, and we were unable to control for alternative explanations for outcomes in the dependent variable.
- Moving forward, we will use statistical regression analysis to address these concerns.

The bivariate linear model

- Recall from unit 1 that our econometric model consists of a DV (Y_i), an IV (X_i), a slope coefficient (β_1), an intercept term (β_0), and an error term (u_i).

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- The beta coefficients determine our best guess for the DV for any given value of X_i, and u_i accounts for any error in that guess.
- When we make inferences about population parameters of interest using sample data, we denote effect parameters using a "hat":

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i,$$

which is the notation we'll stick with from here on.

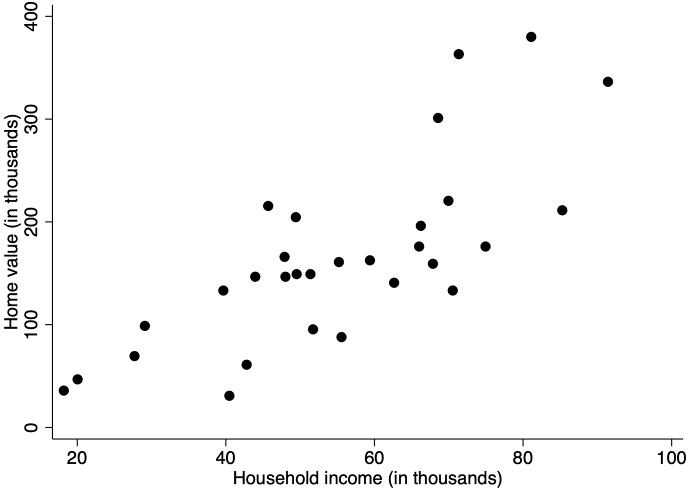
The bivariate linear model (cont'd.)

- Let $\hat{Y}_i$ denote our best guess for the value of the dependent variable given some level of input, X_i:

$$\begin{aligned}\hat{Y}_i &= \hat{\beta}_0 + \hat{\beta}_1 X_i \\ \hat{Y}_i &= Y_i - \hat{u}_i.\end{aligned}$$

- We'll refer to this expression as the "linear predictor."
- Our task is to come up with some values of $\hat{\beta}_0$ and $\hat{\beta}_1$ that best describe the linear relation between X and Y.

Example of a bivariate, linear relationship



The line of best fits

- How do we calculate the line of best fits when we think about the relationship between X and Y?
- What kinds of properties should it have in an ideal world?
- To estimate the line of best fits, we'll make use of a technology known as ordinary least squares (OLS).

Calculating the bivariate OLS Line

- We need some way to summarize the amount of error in our model and then to minimize it.

$$
\begin{aligned}
Y_i &= \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i, && (1)\\
\hat{u}_i &= Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i), && (2)\\
\hat{u}_i^2 &= \left[Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)\right]^2 && (3)\\
\sum_{i=1}^{n} \hat{u}_i^2 &= \sum_{i=1}^{n} \left[Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)\right]^2 && (4)\\
\min \sum_{i=1}^{n} \hat{u}_i^2 &= \min \sum_{i=1}^{n} \left[Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)\right]^2 && (5)
\end{aligned}
$$

The OLS line (Continued)

- We find the values of $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize the sum of squared errors in Equation (5).
- It turns out that, after some calculus and algebraic reorganization, we get:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}, \tag{6}$$

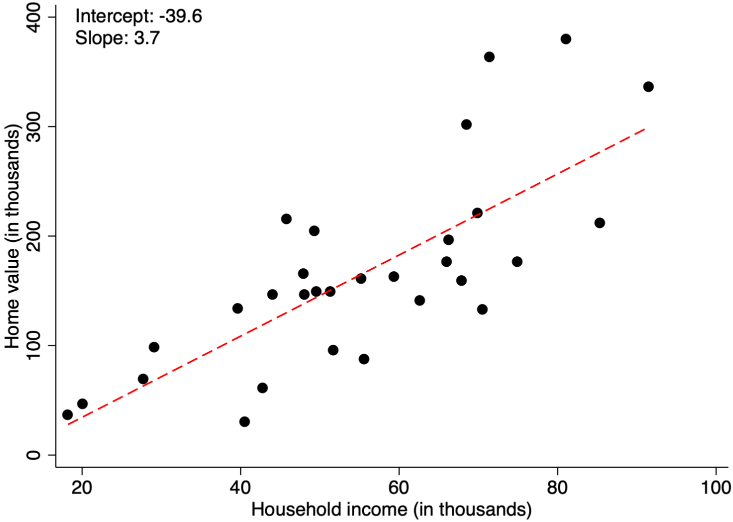
$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}. \tag{7}$$

- Note that the numerator in Equation (6) looks an awful like the formula for covariance while the denominator looks a lot like the formula for variance.
- Indeed, $\hat{\beta}_1 = r_{x,y}\left(\frac{\sigma_y}{\sigma_x}\right)$.

Interpreting the OLS coefficients

- We interpret $\hat{\beta}_0$ and $\hat{\beta}_1$ just as we would for any other type of line.
- The intercept ($\hat{\beta}_0$) tells us the predicted value of $\hat{Y}_i$ given $X_i = 0$.
- The slope coefficient ($\hat{\beta}_1$) tells us the predicted change in $\hat{Y}_i$ for every one-unit change in X_i.
- Remember the following: "For every one-unit increase in X, there is a predicted $\hat{\beta}_1$ change in Y."

The OLS line in practice



An Example: Height and Weight

- Suppose we collect the height and weight of 4 individuals.
- We get height (in inches) among our 4 individuals: $H = \{60, 72, 65, 75\}$.
- And suppose we get weight (in pounds) among our 4 individuals: $W = \{140, 210, 175, 195\}$.
- Suppose: $\text{Weight}_i = \beta_0 + \beta_1 \text{Height}_i + u_i$.
- Let's plot the scatterplot and calculate $\hat{\beta}_0$ and $\hat{\beta}_1$ by hand.

An example (cont'd.)

i	H	W	$\bar{H}$	$\bar{W}$	$H-\bar{H}$	$W-\bar{W}$	$(H-\bar{H})(W-\bar{W})$	$(H-\bar{H})^2$
1	60	140	68	180	-8	-40	320	64
2	72	210	68	180	4	30	120	16
3	65	175	68	180	-3	-5	15	9
4	75	195	68	180	7	15	105	49

Table: Setting up the OLS equations

An example (cont'd.)

- We begin by calculating the slope coefficient, $\hat{\beta}_1$:

$$\begin{aligned}\hat{\beta}_1 &= \frac{320+120+15+105}{64+16+9+49} \\ &= \frac{560}{138} \\ &= 4.1,\end{aligned}$$

which tells us that for every additional inch tall a person is, they are predicted to gain 4.1 pounds.

- Next, the intercept:

$$\begin{aligned}\hat{\beta}_0 &= 180-(4.1)68 \\ &= -96.1,\end{aligned}$$

which tells us that a person who is 0 inches tall is predicted to weigh -96.1 pounds (huh?).

Properties of OLS

- The sum and sample average of residuals will equal zero, $\sum_{i=1}^{n} \hat{u}_i = 0$.
- The sample covariance between the independent variable and residuals is also zero: $\sum_{i=1}^{n} x_i \hat{u}_i = 0$.
- The point, $(\bar{x}, \bar{y})$, is always located on the regression line.

Goodness of fit in OLS: R^2

- First, we have an $R^2 \in [0,1]$, which measures the proportion of variance in the DV that the IV is explaining.
- More specifically, R^2 measures the ratio of "explained" to "total" model variance:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n}\hat{u}^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}.$$

- From our example above looking at height and weight, we get $R^2 = 0.83$. How do we interpret this figure?

OLS and transformed variables

- Suppose we had an independent variable measured as a proportion and had a $\hat{\beta} = 0.50$.
- How would this effect change were we to transform the independent variable into a percentage?

Assumption 1: Linearity

- You might be surprised to learn that linear regression coefficients are required to be. . . linear.
- There's nothing wrong with having non-linear variables.
- But OLS requires linearity in the parameters (i.e., the β coefficients).

Assumption 2: Random Sampling

- OLS assumes that we gathered our data as a random sample of size, n, $\{(x_i, y_i) : i = 1, 2, \ldots, n\}$.
- Often in the social sciences, we are unable to meet this expectation (e.g., time-series designs).
- We'll have to develop technology at a later time to deal with this sorts of issues, but for the most part, cross-sectional designs can be treated like random samples.

Assumption 3: X Must Vary

- Not the most interesting assumption, granted, but it's necessary.
- If the standard deviation of X is zero, either we were very unlucky in our sample, or the phenomenon we're examining isn't very interesting.

Assumption 4: Zero Conditional Mean

- For any given value of x, the error has an expected value of zero, $E(u_i \mid x_i) = 0$, for all i.
- Remember that when we discussed the concept of endogeneity, we framed it as a correlation between the independent variable and the error term.
- Similarly, here we say that the average value of u does not depend upon the value of x. They are independent.

Theorem 1: Unbiasedness of OLS

- Using Assumptions 1 through 4 above, we can conclude that the $\hat{\beta}$s derived via OLS are unbiased.
- That is, $E(\hat{\beta}_0) = \beta_0$ and $E(\hat{\beta}_1) = \beta_1$.
- But it is important to remember that unbiasedness is a feature of the sampling distributions of the $\hat{\beta}$s and does not guarantee that $\hat{\beta} \approx \beta$. On average, though, it will.

OLS and Unbiasedness

- So let's say we have our $\hat{\beta}$s in hand, and now we'd like to ask a simple question: "Just how close to the real βs are these things?"
- Under the central limit theorem, we know that the distribution of $\hat{\beta}$ is normal and therefore unbiased because the mean of all $\hat{\beta}$s will converge upon β.
- This leaves us with the error term.
- If our error term and our independent variable are unrelated to one another, then they are exogenous, and we have unbiased coefficients.
- As X_i and u_i become more correlated, our $\hat{\beta}$ becomes more biased.

Assumption 5: Homoskedasticity

- The variance in the error term, conditional on x, is constant, $\text{Var}(u \mid x) = \sigma^2$.
- This assumption simply says that the variance around x for some given value can't be bigger or smaller compared to other values.
- Interestingly, even in the presence of heteroskedasticity, $\hat{\beta}$ coefficients continue to be unbiased.

Theorem 2: Sampling Variances

- Due to Assumptions 1 through 5, we can show that:

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

- And,

$$\text{Var}(\hat{\beta}_0) = \frac{\sigma^2\left(\sum_{i=1}^{n} x_i^2/n\right)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

- We see that the larger the error variance, the larger $\text{Var}(\hat{\beta}_1)$. Interestingly, though, the greater the variance in x, the greater precision we get in $\hat{\beta}_1$.

Theorem 3: Unbiased Error Variance

- Due to Assumptions 1 through 5, we can show that the expected value of the error variance of the sample equals that of the population, $E(\hat{\sigma}^2) = \sigma^2$.
- This means that when we use $\hat{\sigma}^2$ to estimate the variance of our $\hat{\beta}$s, we're getting unbiased results there too.

More on error variance

- If we take the square root of the error variance term, we get something called the standard error of the regression, $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$.
- This is essentially a measure of the standard deviation in y once the effect of x has been taken out.
- Of special interest to us, though, is how $\hat{\sigma}$ helps us to measure the standard error in $\hat{\beta}$:

$$se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

- We'll use standard errors a lot moving forward as we learn to hypothesis-test with regression.

Conclusion

- In this unit, we introduced the simple, bivariate linear regression model along with some of its basic properties and assumptions.
- Moving forward, we will learn how to build out our model to include more than just one predictor variable.
- We'll also consider the efficiency of our OLS estimator compared to other feasible estimators.
- And from there we will pick up the task of using OLS to test hypotheses.