8.2: 8.2 Deriving OLS Estimators

Now that we have developed some of the rules for differential calculus, we can see how OLS finds values of $\hat{\alpha}$ and $\hat{\beta}$ that minimize the sum of the squared error. In formal terms, let's define the set, $S(\hat{\alpha},\hat{\beta})$, as a pair of regression estimators that jointly determine the residual sum of squares given that: $Y_i=\hat{Y}_i+\epsilon_i=\hat{\alpha}+\hat{\beta}X_i+\epsilon_i$. This function can be expressed:

$$S(\hat{\alpha},\hat{\beta})=\sum_{i=1}^{n}\epsilon_i^2=\sum(Y_i-\hat{Y}_i)^2=\sum(Y_i-\hat{\alpha}-\hat{\beta}X_i)^2$$

First, we will derive $\hat{\alpha}$.

8.2.1 OLS Derivation of $\hat{\alpha}$

Take the partial derivatives of $S(\hat{\alpha},\hat{\beta})$ with-respect-to (w.r.t) $\hat{\alpha}$ in order to determine the formulation of $\hat{\alpha}$ that minimizes $S(\hat{\alpha},\hat{\beta})$. Using the chain rule,

$$\begin{aligned}\frac{\partial S(\hat{\alpha},\hat{\beta})}{\partial\hat{\alpha}}&=\sum 2(Y_i-\hat{\alpha}-\hat{\beta}X_i)^{2-1}*(Y_i-\hat{\alpha}-\hat{\beta}X_i)'\\&=\sum 2(Y_i-\hat{\alpha}-\hat{\beta}X_i)^1*(-1)\\&=-2\sum(Y_i-\hat{\alpha}-\hat{\beta}X_i)\\&=-2\sum Y_i+2n\hat{\alpha}+2\hat{\beta}\sum X_i\end{aligned}$$

Next, set the derivative equal to 0.

$$\frac{\partial S(\hat{\alpha},\hat{\beta})}{\partial\hat{\alpha}}=-2\sum Y_i+2n\hat{\alpha}+2\hat{\beta}\sum X_i=0$$

Then, shift non-$\hat{\alpha}$ terms to the other side of the equal sign:

$$2n\hat{\alpha}=2\sum Y_i-2\hat{\beta}\sum X_i$$

Finally, divide through by $2n$:

$$\begin{aligned}\frac{2n\hat{\alpha}}{2n}&=\frac{2\sum Y_i-2\hat{\beta}\sum X_i}{2n}\\\hat{\alpha}&=\frac{\sum Y_i}{n}-\hat{\beta}*\frac{\sum X_i}{n}\\&=\bar{Y}-\hat{\beta}\bar{X}\\\therefore\hat{\alpha}&=\bar{Y}-\hat{\beta}\bar{X}\end{aligned}\tag{8.1}$$

8.2.2 OLS Derivation of $\hat{\beta}$

Having found $\hat{\alpha}$, the next step is to derive $\hat{\beta}$. This time we will take the partial derivative w.r.t $\hat{\beta}$. As you will see, the steps are a little more involved for $\hat{\beta}$ than they were for $\hat{\alpha}$.

$$\begin{aligned}\frac{\partial S(\hat{\alpha},\hat{\beta})}{\partial\hat{\beta}}&=\sum 2(Y_i-\hat{\alpha}-\hat{\beta}X_i)^{2-1}*(Y_i-\hat{\alpha}-\hat{\beta}X_i)'\\&=\sum 2(Y_i-\hat{\alpha}-\hat{\beta}X_i)^1*(-X_i)\\&=2\sum(-X_iY_i+\hat{\alpha}X_i+\hat{\beta}X_i^2)\\&=-2\sum X_iY_i+2\hat{\alpha}\sum X_i+2\hat{\beta}\sum X_i^2\end{aligned}$$

Since we know that $\hat{\alpha}=\bar{Y}-\hat{\beta}\bar{X}$, we can substitute $\bar{Y}-\hat{\beta}\bar{X}$ for $\hat{\alpha}$.

$$\begin{aligned}\frac{\partial S(\hat{\alpha},\hat{\beta})}{\partial\hat{\beta}}&=-2\sum X_iY_i+2(\bar{Y}-\hat{\beta}\bar{X})\sum X_i+2\hat{\beta}\sum X_i^2\\&=-2\sum X_iY_i+2\bar{Y}\sum X_i-2\hat{\beta}\bar{X}\sum X_i+2\hat{\beta}\sum X_i^2\end{aligned}$$

Next, we can substitute $\frac{\sum Y_i}{n}$ for $\bar{Y}$ and $\frac{\sum X_i}{n}$ for $\bar{X}$ and set it equal to 0.

$$\frac{\partial S(\hat{\alpha},\hat{\beta})}{\partial\hat{\beta}}=-2\sum X_iY_i+\frac{2\sum Y_i\sum X_i}{n}-\frac{2\hat{\beta}\sum X_i\sum X_i}{n}+2\hat{\beta}\sum X_i^2=0$$

Then, multiply through by $\frac{n}{2}$ and put all the $\hat{\beta}$ terms on the same side.

$$\begin{aligned}n\hat{\beta}\sum X_i^2-\hat{\beta}(\sum X_i)^2&=n\sum X_iY_i-\sum X_i\sum Y_i\\\hat{\beta}(n\sum X_i^2-(\sum X_i)^2)&=n\sum X_iY_i-\sum X_i\sum Y_i\\\therefore\hat{\beta}&=\frac{n\sum X_iY_i-\sum X_i\sum Y_i}{n\sum X_i^2-(\sum X_i)^2}\end{aligned}$$

The $\hat{\beta}$ term can be rearranged such that:

$$\hat{\beta}=\frac{\Sigma(X_i-\bar{X})(Y_i-\bar{Y})}{\Sigma(X_i-\bar{X})^2}\tag{8.2}$$

Now remember what we are doing here: we used the partial derivatives for $\sum\epsilon^2$ with respect to $\hat{\alpha}$ and $\hat{\beta}$ to find the values for $\hat{\alpha}$ and $\hat{\beta}$ that will give us the smallest value for $\sum\epsilon^2$. Put differently, the formulas for $\hat{\beta}$ and $\hat{\alpha}$ allow the calculation of the error-minimizing slope (change in Y given a one-unit change in X) and intercept (value for Y when X is zero) for any data set representing a bivariate, linear relationship. No other formulas will give us a line, using the same data, that will result in as small a squared-error. Therefore, OLS is referred to as the Best Linear Unbiased Estimator (BLUE).

8.2.3 Interpreting $\hat{\beta}$ and $\hat{\alpha}$

In a regression equation, $Y=\hat{\alpha}+\hat{\beta}X$, where $\hat{\alpha}$ is shown in Equation (8.1) and $\hat{\beta}$ is shown in Equation (8.2). Equation (8.2) shows that for each 1-unit increase in X you get $\hat{\beta}$ units to change in Y. Equation (8.1) shows that when X is 0, Y is equal to $\hat{\alpha}$. Note that in a regression model with no independent variables, $\hat{\alpha}$ is simply the expected value (i.e., mean) of Y.

The intuition behind these formulas can be shown by using `R` to calculate "by hand" the slope ($\hat{\beta}$) and intercept ($\hat{\alpha}$) coefficients. A theoretical simple regression model is structured as follows:

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

- α and β are constant terms
- α is the intercept
- β is the slope
- X_i is a predictor of Y_i
- ϵ is the error term

The model to be estimated is expressed as $Y=\hat{\beta}+\hat{\beta}X+/epsilon$.

As noted, the goal is to calculate the intercept coefficient:

$\hat{\alpha}=\bar{Y}-\hat{\beta}\bar{X}$ and the slope coefficient: $\hat{\beta}=\frac{\Sigma(X_i-\bar{X})(Y_i-\bar{Y})}{\Sigma(X_i-\bar{X})^2}$

Using `R`, this can be accomplished in a few steps. First, create a vector of values for `x` and `y` (note that we chose these values arbitrarily for the purpose of this example).

```
x <- c(4,2,4,3,5,7,4,9)
x
```

```
## [1] 4 2 4 3 5 7 4 9
```

```
y <- c(2,1,5,3,6,4,2,7)
y
```

```
## [1] 2 1 5 3 6 4 2 7
```

Then, create objects for $\bar{X}$ and $\bar{Y}$:

```
xbar <- mean(x)
xbar
```

```
## [1] 4.75
```

```
ybar <- mean(y)
ybar
```

```
## [1] 3.75
```

Next, create objects for $(X-\bar{X})$ and $(Y-\bar{Y})$, the deviations of X and Y around their means:

```
x.m.xbar <- x-xbar
x.m.xbar
```

```
## [1] -0.75 -2.75 -0.75 -1.75  0.25  2.25 -0.75  4.25
```

```
y.m.ybar <- y-ybar
y.m.ybar
```

```
## [1] -1.75 -2.75  1.25 -0.75  2.25  0.25 -1.75  3.25
```

Then, calculate $\hat{\beta}$:

$$\hat{\beta}=\frac{\Sigma(X_i-\bar{X})(Y_i-\bar{Y})}{\Sigma(X_i-\bar{X})^2}$$

```
B <- sum((x.m.xbar)*(y.m.ybar))/sum((x.m.xbar)^2)
B
```

```
## [1] 0.7183099
```

Finally, calculate $\hat{\alpha}$

$$\hat{\alpha}=\bar{Y}-\hat{\beta}\bar{X}$$

```
A <- ybar-B*xbar
A
```

```
## [1] 0.3380282
```

To see the relationship, we can produce a scatterplot of `x` and `y` and add our regression line, as shown in Figure 8.2.4. So, for each unit increase in x, y increases by 0.7183099 and when x is 0, y is equal to 0.3380282.

```
plot(x,y)
lines(x,A+B*x)
```

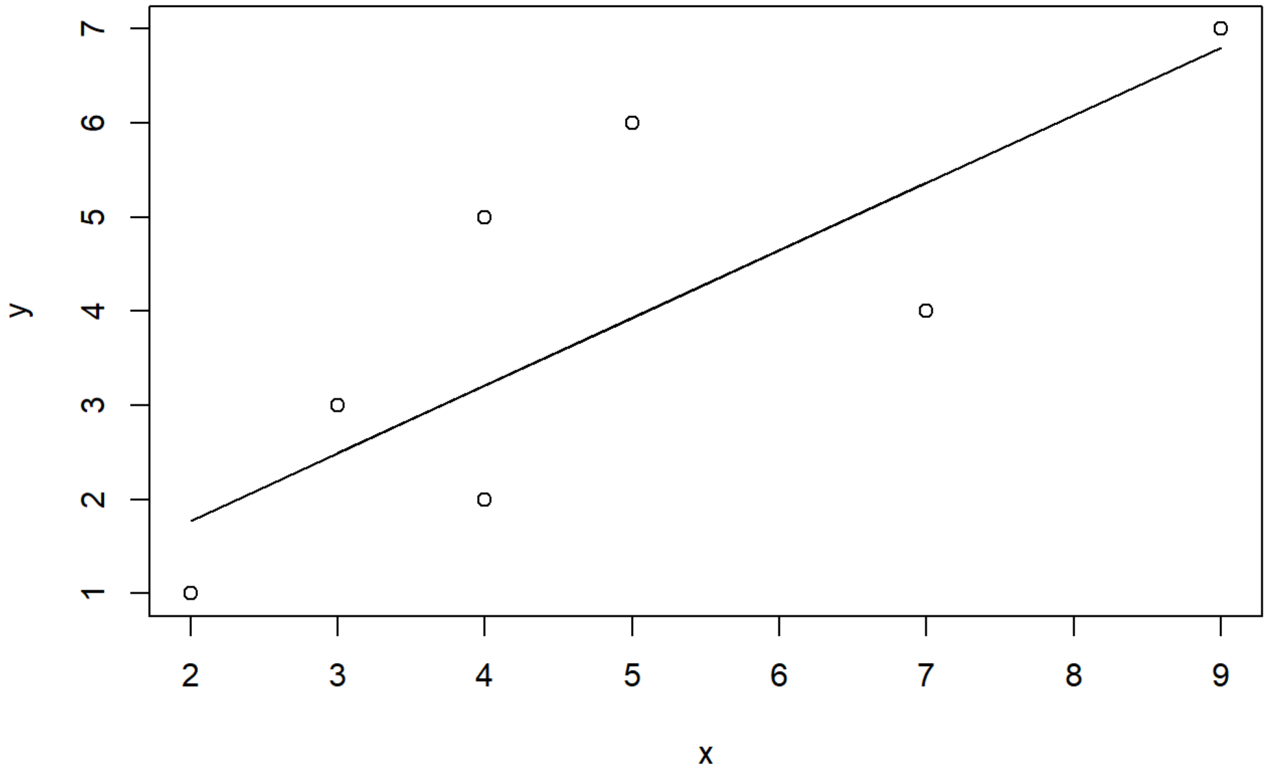

Figure 8.2.4: Simple Regression of xx and yy

```
dev.off()
```

```
## RStudioGD
##          2
```