Ordinary least squares (OLS) is a mathematical technique used to estimate a relationship between different variables. The most simple version of this relationship is:

\[ Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \]

where \( Y_i \) is the value of the dependent variable, \( X_i \) is the value of the explanatory variable and \( \varepsilon_i \) is the value of the stochastic error term for the \( i^{th} \) observation.

The result of this estimation procedure is estimates of the coefficients, \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \), which are called \( \beta_0 \)-hat and \( \beta_1 \)-hat. These coefficients are used to generate estimates of or predicted values for the dependent variable which are called \( \hat{Y}_i \)-hat and we can say that

\[ \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \]

The difference between the actual value of \( Y_i \) and its estimated value, \( \hat{Y}_i \), is equal to \( \varepsilon_i \), the error term. This can be written as:

\[ Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \varepsilon_i \]

\[ Y_i - \hat{Y}_i = \varepsilon_i \]

So, what do these things look like on a graph?
What do $e_i$ look like?

The goal of OLS is to minimize the sum of $e_i^2$, $e_2^2$, and $e_3^2$.

What do Yi-hat look like?
The line relating X and Y that is calculated by OLS is good because it minimizes the sum of the squared errors. That is it minimizes:

$$\sum e_i^2$$

It is equivalent to say that it minimizes:

$$\sum (y_i - \hat{y}_i)^2$$

This has three good characteristics:
1. The regression line goes through the point $\left( \bar{X}, \bar{Y} \right)$ which is the mean of the data
2. The sum of the errors or residuals is zero
3. OLS gives the "best" estimation, depending on some conditions and definitions

**Definitions**

Standard Error of the Estimate (SEE)

$$\text{SEE} = \sqrt{\frac{\sum e_i^2}{n - 2}}$$

Total Sum of Squares (TSS)

$$\text{TSS} = \sum (y_i - \bar{Y})^2$$
Explained Sum of Squares (ESS)

\[
ESS = \sum (\hat{Y}_i - \bar{Y})^2
\]
Residual Sum of Squares (RSS)

\[ RSS = \sum (e_i^2) = \sum (\hat{Y}_i - Y_i)^2 \]

To put this all together: TSS = ESS + RSS

\[ R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \]

A higher \( R^2 \) means that the model being estimated explains a higher level of variation in the dependent variable.

If the \( R^2 \) is zero, then the model offers no information about the dependent variable and the best prediction you can make about the value of the dependent variable is its mean. The "explanatory" variables really offer no explanatory power whatsoever.

What does this look like?
Multivariate Regression
There aren't a lot of interesting questions to be answered using one explanatory variable. It's more fun to look at a number of explanatory variables through multivariate regressions, or regression on multiple variables. For a model with $K$ explanatory variables this would be:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki} + \epsilon_i$$

Each of the coefficients, $\beta_k$, is the partial derivative of the dependent variable, $Y_i$ with respect to the explanatory variable $X_{ki}$. That is, the coefficient is the expected change in $Y$ resulting from a one unit change in $X$ holding the other variables constant.

A Beef Example (Studenmund, p. 44)
Consider the following estimated model:

$$\hat{B}_t = 37.53 - 0.88P_t + 11.9Yd_t$$

where
- $B_t$ = the per capita consumption of beef in year $t$ in pounds per person
- $P_t$ = the price of beef in year $t$ (in cents per pound)
- $Yd_t$ = the per capita disposable income in year $t$ (in thousands of dollars)

Questions
1. What is the interpretation of the coefficient on $P_t$?
2. What is the interpretation of the coefficient on $Yd_t$?
3. Is beef a normal good?
4. Do these estimated coefficients conform to the law of demand?
5. According to the model, what would happen to per capita consumption of beef if the price rose by $0.02/pound?
6. According to the model, what would happen to per capita consumption of beef if per capita disposable income rose by $2,000?
7. According to the model, what would happen to per capita consumption of beef if the price of beef doubled?
8. According to the model, what would happen to per capita consumption of beef if the price rose by $0.50? Do you believe this result? Explain the problem and offer a solution.
9. How would the coefficient estimates change if beef consumption was expressed in kilograms per person? If the price was expressed in dollars per pound?

Adjusted $R^2$

Now, an ideal model will have a lot of explanatory power. This means that $\text{ESS/TSS} = R^2$ should be as close as possible to 1. However, adding more and more variables to the equation to be estimated will never decrease the $R^2$ and will often increase it. The result is that a model with many unrelated explanatory variables will appear to have very high explanatory power, but this high $R^2$ will merely be a result of spurious correlation.

To correct for this, we calculate the adjusted $R^2$.

$$\text{Adj. } R^2 = R^2 - \left(1 - R^2\right) \frac{K}{n - K - 1}$$

where

- $n$ = the number of observations in the data set
- $K$ = the number of slope coefficients estimated

For example, in the beef regression above, $n$ would be the number of years in which data were gathered and $K$ would be two because one coefficient was estimated for price and one was estimated for income.

Adding a variable to a regression, even if it has nothing to do with the dependent variable, is likely to increase $R^2$ but will also increase $K$ and may decrease the adjusted $R^2$.

A good general rule to use in choosing between models is to choose the model with a higher adjusted $R^2$. If you are considering adding a new variable to a regression, see if it increases the adjusted $R^2$. If it does, then it should perhaps be added.

This, however, is a rule which should be applied only carefully, as the example in Studenmund 2.5 explains.

The best way to choose variables for a regression is to research the dependent variable and, on the basis of your understanding of the variable, decide which variables should be included prior to doing any regressions. Your model should be theoretically sound and you should be able to offer a good explanation for your inclusion of each explanatory variable.

Degrees of Freedom

The number of degrees of freedom in a regression is equal to the difference between the number of observations in the data set ($n$) minus the number of coefficients to be estimated ($K+1$).

It must be the case that $n-K-1 \geq 0$.

To see why, consider the case of a univariate regression with one data point.

$$Y_i = b_0 + b_1 X_i$$

Because there are two coefficients to be estimated ($b_0$ and $b_1$) we have $K+1=1+1=2$. So, we must have at least two points to estimate a line here. This can be seen quite easily graphically.

1. What if you try to estimate an OLS relationship with one observation?
2. How can you estimate an OLS relationship with one observation? What do you need to assume?
Tough Questions You Can Ask

Studenmund (p. 49) offers several questions you can and should ask when reading a report involving an OLS regressions.

1. Is the equation supported by sound theory?
   At one level, ask yourself if the included explanatory variables make sense and if there are other variables you believe should be included. It may be that there is a very good reason for exclusion of some variables, but you should ask.

2. How well does the regression as a whole fit the data?
   This relates to the $R^2$. A low $R^2$ doesn't necessarily mean you should condemn the model, but it should raise some flags about what the results should be used for. It should be understood that if the $R^2$ is low, the predictive power for any one observation may be very low, although it might be good for a large number of observations. Similarly, a very high $R^2$ might suggest something suspicious.

3. Is the data set reasonably large and accurate?
   The number of observations is important, but even more important is the number of degrees of freedom. Further, you should ask yourself if all of the variables seem quantifiable and measurable and how accurately they may have been measured.

4. Is OLS the best estimator to be used for this equation?
   There are some other options we will discuss, although they are all basically variations on the OLS theme.

5. How well do the estimated coefficients correspond to the expectations developed by the researcher before the data were collected?
   Look for weird signs on the estimated coefficients. If quantity demanded is positively related to price, for example, there is something suspicious going on and you should ask questions.

6. Are all the obviously important variables included in the equation?

7. Has the most theoretically logical functional form been used?
   Consider the beef example. Does the prediction about the effect of a $0.50/pound increase in the price of beef make sense? A different model may be appropriate here. Explanatory variables may need to be raised to some power to more accurately describe how they affect the dependent variable.

8. Does the regression appear to be free of major econometric problems?
   We'll discuss some more of these. The one we have talked about that you should remember is endogeneity. That is, can you imagine a model in which one of the explanatory variables is explained by other explanatory variables?
**Answers to Questions on the Beef Model**

1. What is the interpretation of the coefficient on $P_t$?
   This is the change in annual, per capita beef consumption in pounds resulting from a $0.01$ increase in the price of beef.

2. What is the interpretation of the coefficient on $Y_{dt}$?
   This is the change in annual, per capita beef consumption in pounds resulting from a $1,000$ increase in per capita income.

3. Is beef a normal good?
   Yes, the data seem to suggest that it is because the coefficient on income is positive, so as income rises (holding other things constant), so will beef consumption.

4. Do these estimated coefficients conform to the law of demand?
   Yes they do. The estimated coefficient on price is negative, suggesting that as the price of beef rises (holding other things constant) the quantity demanded will fall.

5. According to the model, what would happen to per capita consumption of beef if the price rose by $0.02$/pound?
   If the price of beef rises by $0.02$/pound, the model predicts that annual per capita consumption will fall by $2 \times 0.88 = 1.76$ pounds.

6. According to the model, what would happen to per capita consumption of beef if per capita disposable income rose by $2,000$?
   If per capita disposable income rose by $2,000$, the model predicts that annual per capita beef consumption would rise by $11.9 \times 2 = 23.8$ pounds.

7. According to the model, what would happen to per capita consumption of beef if the price of beef doubled?
   We can't tell without knowing the current price of beef because we don't know the amount of the increase in cents per pound.

8. According to the model, what would happen to per capita consumption of beef if the price rose by $0.50$?
   Do you believe this result? Explain the problem and offer a solution.
   If the price of beef rose by $0.50$/pound, annual per capita consumption would fall by $50 \times 0.88 = 44$ pounds.
   This seems a bit drastic and I would suggest a model in which beef consumption was maybe related to the square root of the price.

9. How would the coefficient estimates change if beef consumption was expressed in kilograms per person?
   Because there are $2.2$ pounds per kilogram, each coefficient would be divided by $2.2$. For example, to get a one kilogram increase in per capita consumption, the necessary price decrease would have to be $2.2$ times what is needed to get a one pound increase.

If the price was expressed in dollars per pound?
   If the price was expressed in dollars per pound the coefficient on $P_t$ would be multiplied by $100$. In any case, a one dollar increase in the price of beef will decrease annual per capita consumption by $88$ pounds.