Guide to Interpreting OLS Regression Coefficients

GV481

October 4, 2023

1 Continuous-Binary DVs and IVs

This guide will introduce you to the ways we interpret regression coefficients when dependent (Y) and independent (X) variables are measured differently. Let's assume, we want to understand the relationship between education and income: $Income = \beta_0 + \beta_1 Education + \epsilon$

Coefficient	DV	IV	Example	Interpretation
β_0	-	-		β_0 gives the average value of $Income$ when education = 0.
β_1	Continuous	Continuous	Association between years of education and pounds earned per year.	For every additional 1 year of education, there is β_1 change in pounds earned per year.
	Binary	Continuous	Association between years of education and whether the person is a high-earner (over 80K pounds earned per year).	For every additional 1 year of education, there is a β_1 change in the probability that an individual is a high earner (or a $\beta_1 \times 100$ percentage point change).
	Continuous	Binary	Association between having a post-graduate degree and pounds earned per year	The difference in pounds earned per year between those who do not have and those who have a post-graduate degree is β_1.
	Binary	Binary	Association between being a high earner and having a post-graduate degree.	The difference in the probability of being a high earner between those who have and do not have a post-graduate degree is β_1.

2 Logging DVs and IVs

Often we want to change how we measure the dependent or independent variables to get a better distribution in the data. For example, when our variables are left or right-skewed, it is often advised to log the variable to get a more normal distribution. This changes the interpretation of the β_1 coefficient.

Coefficient	DV	IV	Example	Interpretation
β_1	log(Continuous)	Continuous	Association between logged pounds earned per year and years of education	For every additional year of education, there is a $100\times\beta_1$ percent change in earnings (NOT log earnings).
	log(Continuous)	Binary	Association between logged pounds earnings and whether the individual has post-graduate education or not.	The % difference in earnings between those who have an education vs. those who do not is $100 \times_b eta_1$.
	log(Continuous)	log(Continuous)	Association between logged pounds earnings and logged years of education.	For every 1% increase in education, there is a β_1 percent increase in pounds earned per year.
	Continuous	log(Continuous)	Association between earnings in pounds and logged education	For every 1% increase in years of education, there is a $(\beta_1/100)$ pounds change in yearly earnings.
	Binary	log(Continuous)	Association between being a high earner and logged years of education.	For every 1% change in years of education, there is a $(\beta_1/100)$ change in the probability of being a high earner.

3 Percent Vs. Percentage Point

While percent and percentage points are closely related, we have to be careful which one we use when we interpret coefficients. For example, assume that we want to estimate the relationship between the number of hours a student has studied and their test score. Test score is measured in percentage terms and ranges from 0 to 100. We estimate the following model:

$$TestScore_i = \beta_0 + \beta_1 StudyHours_i + \epsilon_i$$

Let's assume that we get the following results:

$$Test\hat{S}core_i = 20 + 10.5StudyHours_i$$

. We would interpret this as for every additional hour studied, on average, students score 10.5 percentage points higher on the test. However, I can also express this change in percent terms. Let's say the average student studied got a test score of 55. For the average student, an additional hour of study would result in a score of $55 + 10.5 = 65.5$. This corresponds to an about 19 percent increase in test scores, because $65.5/55 = 0.19$.

Things get a bit more complex when we talk about probabilities. The probability of some event taking place is between 0 and 1. When we have a binary variable as the dependent variable, we often interpret β_1 as a change in the probability of Y taking place. For example, we may be interested in a student passing an exam, instead of test scores. If we want to know the relationship between passing an exam and number of hours studied, we can estimate:

$$PassExam_i = \beta_0 + \beta_1 StudyHours_i + \epsilon_i.$$

Suppose, we get the following result:

$$Pass\hat{E}xam_i = 0.4 + 0.05 StudyHours_i + \epsilon_i.$$

. In this case, we would say that an additional hour of study is associated with a 0.05 increase in the probability that a student passes OR that there is a 5 percentage point increase in the likelihood of the student passing. This is because if we rescale the probability of passing the exam from 0-1 to 0-100, we can easily multiply the β_1 by a 100 as well. Now, let's say that the average student's probability of passing the exam is 0.6. Then an extra hour of study would result in a 0.65 probability of passing the exam, an 8% increase.