

## 6.3: Regression Coefficients

The whole goal of the regression analysis was to test the hypothesis that the dependent variable,  $Y$ , was in fact dependent upon the values of the independent variables as asserted by some theory, such as the consumption and income example. Looking at the estimated equation under Figure 13.8, we see that this amounts to determining the values of  $b_0$  and  $b_1$ .

The regression analysis output provided by the computer software will produce an estimate of  $b_0$  and  $b_1$ , and any other  $b$ 's for other independent variables that were included in the estimated equation. The issue is how good are these estimates? In order to test a hypothesis concerning any estimate, we have found that we need to know the underlying sampling distribution. It should come as no surprise by now in the course that the answer is going to be the normal distribution. This can be seen by remembering the assumption that the error term in the population is normally distributed. If the error term is normally distributed and the variance of the estimates of the equation parameters,  $b_0$  and  $b_1$ , are determined by the variance of the error term, it follows that the variances of the parameter estimates are also normally distributed. And indeed this is just the case.

To test whether or not  $Y$  does indeed depend upon  $X$ , or in our example, that whether consumption depends upon income, we need only test the hypothesis that  $b_1$  equals zero. This hypothesis would be stated formally as:

$$H_0 : b_1 = 0$$

$$H_a : b_1 \neq 0$$

If we cannot reject the null hypothesis, we must conclude that our theory has no validity. If we cannot reject the null hypothesis that  $b_1 = 0$  then  $b_1$ , the coefficient of Income, is zero and zero times anything is zero. Therefore the effect of Income on Consumption is zero. There is no relationship as our theory had suggested.

Notice that as before, we have set up the presumption, the null hypothesis, as "no relationship". This puts the burden of proof on the alternative hypothesis. In other words, if we are to validate our claim of finding a relationship, we must do so with a level of significance greater than 95 percent typically. No relationship exists, and to be able to make the claim that we have actually added to our body of knowledge we must do so with significant probability of being correct.

The test statistic for this test comes directly from our old friend, the t-test formula:

$$t_c = \frac{b_1 - \beta_1}{S_{b_1}}$$

where  $b_1$  is the estimated value of the slope of the regression line,  $\beta_1$  is the hypothesized value of slope of the regression line, which is always zero, and  $S_{b_1}$  is the standard deviation of the estimate of  $b_1$ . In this case we are asking how many standard deviations is the estimated slope away from the hypothesized slope. This is exactly the same question we asked before with respect to a hypothesis about a mean: how many standard deviations is the estimated mean, the sample mean, from the hypothesized population mean?

The decision rule for acceptance or rejection of the null hypothesis follows exactly the same form as in all our previous test of hypothesis. Namely, if the calculated value of  $t$  (or  $Z$ ) falls into the tails of the distribution, where the tails are defined by  $\alpha$ , the required significance level in the test, we have enough evidence to reject the null hypothesis. If on the other hand, the calculated value of the test statistic is within the critical region, we fail to reject the null hypothesis.

If we conclude that we reject the null hypothesis, we are able to state with  $(1 - \alpha)$  level of confidence that the slope of the line is given by  $b_1$ . This is an extremely important conclusion. Regression analysis not only allows us to test if a cause and effect relationship exists, we can also determine the magnitude of that relationship, if one is found to exist. It is this feature of regression analysis that makes it so valuable. If models can be developed that have statistical validity, we are then able to simulate the effects of changes in variables that may be under our control with some degree of probability, of course. For example, if intentional advising is demonstrated to affect student retention, we can determine the effects of changing to intentional advising and decide if the increased retention are worth the added expense.

This page titled [6.3: Regression Coefficients](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Yang Lydia Yang](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- [13.4: The Regression Equation](#) by OpenStax is licensed [CC BY 4.0](#). Original source: <https://openstax.org/details/books/introductory-business-statistics>.