
WHAT IS A REGRESSION?

1.0	What We Need to Know When We Finish This Chapter	2
1.1	Why Are We Doing This?	3
1.2	Education and Earnings	5
1.3	What Does a Regression Look Like?	6
1.4	Where Do We Begin?	6
1.5	Where's the Explanation?	7
1.6	What Do We Look for in This Explanation?	9
1.7	How Do We Interpret the Explanation?	12
1.8	How Do We Evaluate the Explanation?	17
1.9	R^2 and the F -statistic	19
1.10	Have We Put This Together in a Responsible Way?	20
1.11	Do Regressions Always Look Like This?	25
1.12	How to Read This Book	28
1.13	Conclusion	28
	Exercises	29

1.0 What We Need to Know When We Finish This Chapter

This chapter explains what a regression is and how to interpret it. Here are the essentials.

1. **Section 1.4:** The *dependent* or *endogenous* variable measures the behavior that we want to explain with regression analysis.
2. **Section 1.5:** The *explanatory*, *independent*, or *exogenous* variables measure things that we think might determine the behavior that we want to explain. We usually think of them as *predetermined*.
3. **Section 1.5:** The *slope* estimates the effect of a change in the explanatory variable on the value of the dependent variable.
4. **Section 1.5:** The *t*-statistic indicates whether the explanatory variable has a discernible association with the dependent variable. The association is discernible if the *p*-value associated with the *t*-statistic is .05 or less. In this case, we say that the slope is *statistically significant*. This generally corresponds to an absolute value of approximately two or greater for the *t*-statistic itself. If the *t*-statistic has a *p*-value that is greater than .05, the associated slope coefficient is *insignificant*. This means that the explanatory variable has no discernible effect.
5. **Section 1.6:** The *intercept* is usually uninteresting. It represents what everyone has in common, rather than characteristics that might cause individuals to be different.
6. **Section 1.6:** We usually interpret only the slopes that are statistically significant. We usually think of them as indicating the effect of their associated explanatory variables on the dependent variable *ceteris paribus*, or *holding constant all other characteristics that are included in the regression*.
7. **Section 1.6:** *Continuous variables* take on a wide range of values. Their slopes indicate the change that would be expected in the dependent variable if the value of the associated explanatory variable increased by one unit.
8. **Section 1.6:** *Discrete variables*, sometimes called *categorical variables*, indicate the presence or absence of a particular characteristic. Their slopes indicate the change that would occur in the dependent variable if an individual who did not have that characteristic were given it.
9. **Section 1.7:** Regression interpretation requires three steps. The first is to identify the discernible effects. The second is to understand their magnitudes. The third is to use this understanding to verify or modify

the behavioral understanding that motivated the regression in the first place.

10. **Section 1.7:** Statistical significance is *necessary* in order to have interesting results, but not *sufficient*. Important slopes are those that are both statistically significant and substantively large. Slopes that are statistically significant but substantively small indicate that the effects of the associated explanatory variable can be reliably understood as unimportant.
11. **Section 1.7:** A *proxy* is a variable that is related to, but not exactly the variable we really want. We use proxies when the variables we really want aren't available. Sometimes this makes interpretation difficult.
12. **Section 1.8:** If the *p-value* associated with the *F-statistic* is .05 or less, the collective effect of the ensemble of explanatory variables on the dependent variable is statistically significant.
13. **Section 1.8:** *Observations* are the individual examples of the behavior under examination. All of the observations together constitute the *sample* on which the regression is based.
14. **Section 1.8:** The R^2 , or *coefficient of determination*, represents the proportion of the variation in the dependent variable that is explained by the explanatory variables. The adjusted R^2 modifies the R^2 in order to take account of the numbers of explanatory variables and observations. However, neither measures statistical significance directly.
15. **Section 1.9:** *F-statistics* can be used to evaluate the contribution of a subset of explanatory variables, as well as the collective statistical significance of all explanatory variables. In both cases, the *F-statistic* is a transformation of R^2 values.
16. **Section 1.10:** Regression results are useful only to the extent that the choices of variables in the regression, variable construction, and sample design are appropriate.
17. **Section 1.11:** Regression results may be presented in one of several different formats. However, they all have to contain the same substantive information.