

## 13.11: Example of How to Test a Hypothesis Using Regression

Let's walk through the steps for testing the hypothesis that sleep hours will be useful in predicting quiz scores using Data Set 12.1. This will include summarizing some of the computations we learned in prior sections of this chapter.

### Steps in Hypothesis Testing

In order to test a hypothesis, we must follow these steps:

#### 1. State the hypothesis.

A summary of the research hypothesis and corresponding null hypothesis in sentence and symbol format are shown below. However, researchers often only state the research hypothesis using a format like this: *It is hypothesized that hours of sleep will be useful for predicting quiz scores.*

Non-Directional Hypothesis for a Regression

Research hypothesis	Hours of sleep will predict quiz scores.	$H_A : \beta_1 \neq 0$
Null hypothesis	Hours of sleep will not predict quiz scores.	$H_0 : \beta_1 = 0$

#### 2. Choose the inferential test (formula) that best fits the hypothesis.

The usefulness of one quantitative variable for predicting another quantitative variable is being tested so the appropriate test is simple regression.

#### 3. Determine the critical value.

In order to determine the critical value for a regression, three things must be identified:

1. the alpha level,
2. the degrees of freedom for the model ( $df_M$ ), and
3. the degrees of freedom for the error ( $df_E$ )

The alpha level is often set at .05 unless there is reason to adjust it such as when multiple hypotheses are being tested in one study or when a Type I Error could be particularly problematic. The default alpha level can be used for this example because only one hypothesis is being tested and there is no clear indication that a Type I Error would be especially problematic. Thus, alpha can be set to 5%, which can be summarized as  $\alpha = .05$ .

The  $df_M$  and the  $df_E$  must also be calculated. The  $df_M$  is equal to the number of predictor variables. In simple bivariate regression only one predictor variable is used (the  $X$ -variable) so the  $df_M = 1$ . In simple bivariate regression the  $df_M = n - 2$ . Thus, the two forms of  $df$  for Data Set 12.1 are as follows:

$$df_M = 1$$

$$df_E = 8$$

The alpha level and  $df$ s are used to determine the critical value for the test. The full tables of the critical values for  $F$  are located in Appendix E. Under the conditions of an alpha level of .05,  $df_M = 1$ , and  $df_E = 8$ , the critical value (CV) is 5.318 (see Appendix E).

$$CV = 5.318$$

The critical value represents the value which must be exceeded in order to declare a result significant. It represents the threshold of evidence needed to be confident a hypothesis is true. Regression uses ANOVA for which the result can only be a positive value. Thus, the obtained  $F$ -value must be greater than 5.318 to be declared significant when using Data Set 12.1.

#### 4. Calculate the test statistic.

An ANOVA  $F$ -value is used to see if an  $X$ -variable is useful for predicting a  $Y$ -variable. The formula for this is:

$$F = \frac{RSS \div df_M}{SSE \div df_E}$$

See the section titled *Testing a Regression Model with ANOVA* earlier in this chapter for details on how SSR and SSE are calculated. For this section, we will show the abbreviated computations once those values and the  $dfs$  are known. For Data Set 12.1, the results are computed as follows:

$$F = \frac{1,728.0234 \div 1}{195.9766 \div 8} = \frac{1,728.0234}{24,4971} = 70.5399$$

##### 5. Apply a decision rule and determine whether the result is significant.

Assess whether the obtained value for  $F$  exceeds the critical value as follows:

The critical value is 5.318.

The obtained  $F$ -value, rounded to the hundredths place, is 70.54.

The obtained  $F$ -value exceeds (i.e. is greater than) the critical value, thus, the result is significant.

##### 6. Calculate the effect sizes and relevant secondary analyses.

When it is determined that the result is significant, effect sizes should typically be computed. However, in correlation and regression, a secondary analysis is typically given rather than an effect size. Thus, the foci of step 6 for regression is to calculate and interpret the coefficient of determination and the slope of the regression line.

The **coefficient of determination** is the percent of variation in the  $Y$ -variable that is accounted for by variance in the  $X$ -variable. It can also be described as a calculation of how well a model using one variable ( $X$ ) can be used to estimate the other ( $Y$ ). These refer to two ways of interpreting and describing the same thing with the former more applicable to correlation and the latter more applicable to regression. The coefficient of determination results and interpretation will be summarized here as they were already covered for Data Set 12.1 in Chapter 12. For a full review of how to compute and interpret this value, see Chapter 12.

The symbol and formula for the coefficient of determination are the same and are written as follows:

$$r^2$$

The coefficient of determination for Data Set 12.1 would be computed as follows:

$$r^2 = (0.9477 \dots)^2$$

$$r^2 = 0.8981 \dots$$

This result is quite large and can be reported as a percent and interpreted as follows for Data Set 12.1:

Approximately 89.8% of the variance in quiz scores was accounted for by variance in hours of sleep.

In addition, when a regression ANOVA is significant, the slope should be interpreted and supported with an evidence string for the  $t$ -test. These are generally computed using SPSS rather than by hand and are as follows:

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	38.5529	5.177		7.447	<.001
	Hours of Sleep	6.3765	.759	.948	8.399	<.001

a. Dependent Variable: Quiz Scores

The slope, when interpreted and supported with the  $t$ -test results, would be reported as follows:

*For every one hour increase in sleep, there is a 6.38 unit increase predicted in quiz scores,  $t(8) = 8.40, p < .05$ .*

Note that the  $df_E$  is used for the  $t$ -test in a regression.

##### 7. Report the results in American Psychological Associate (APA) format.

Results for inferential tests are often best summarized using a paragraph that states the following:

- the hypothesis and specific inferential test used,
- the main results of the test and whether they were significant,
- any additional results that clarify or add details about the results,
- whether the results support or refute the hypothesis.

Following this, the results for our hypothesis with Data Set 12.1 can be written as shown in the summary example below.

#### APA Formatted Summary Example

A simple regression was used to test the hypothesis that hours of sleep would predict quiz scores. Consistent with the hypothesis, hours of sleep was a significant predictor of quiz scores,  $F(1, 8) = 70.54$ ,  $p < .05$ . Approximately 89.8% of the variance in quiz scores was accounted for by variance in hours of sleep. For every one hour increase in sleep, there is a 6.38 unit increase predicted in quiz scores,  $t(8) = 8.40$ ,  $p < .05$ .

As always, the APA-formatted summary provides a lot of detail in a particular order. To understand how to read and create a summary like this, review the detailed walk-through in Chapter 7. For a brief review of the structure for the APA-formatted summary of the omnibus test results, see the summary below.

The following breaks down what each part represents in the evidence string for the ANOVA and  $t$ -test results in the APA-formatted paragraph above:

Symbol for the test	Degrees of Freedom	Obtained Value	$p$ -Value
$F$	(1, 8)	= 70.54	$p < .05$ .
$t$	(8)	= 8.40	$p < .05$ .

- How is the ANOVA part of regression interpreted and reported in APA-format?
- How is the  $t$ -test part of regression interpreted and reported in APA-format?
- When a regression is significant, how is the coefficient of determination interpreted and reported?
- When a regression is significant, how is the slope interpreted and reported?

This page titled [13.11: Example of How to Test a Hypothesis Using Regression](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by .