

9.1: Hypothesis Tests for Regression Coefficients

Hypothesis testing is the key to theory building. This chapter is focused on empirical hypothesis testing using OLS regression, with examples drawn from the accompanying class dataset. Here we will use the responses to the political ideology question (ranging from 1=strong liberal to 7=strong conservative), as well as responses to a question concerning the survey respondents' level of risk that global warming poses for people and the environment.¹⁵

Using the data from these questions, we posit the following hypothesis:

H1H1: On average, as respondents become more politically conservative, they will be less likely to express increased risk associated with global warming.

The null hypothesis, H_0H_0 , is $\beta=0\beta=0$, posits that a respondent's ideology has no relationship with their views about the risks of global warming for people and the environment. Our working hypothesis, H_1H_1 , is $\beta<0\beta<0$. We expect $\beta\beta$ to be less than zero because we expect a *negative* slope between our measures of ideology and levels of risk associated with global warming, given that a larger numeric value for ideology indicates a more conservative respondent. Note that this is a *directional* hypothesis since we are posting a negative relationship. Typically, a directional hypothesis implies a one-tailed test where the critical value is 0.05 on one side of the distribution. A *non-directional* hypothesis, $\beta\neq 0\beta\neq 0$ does not imply a particular direction, it only implies that there is a relationship. This requires a two-tailed test where the critical value is 0.025 on both sides of the distribution.

To test this hypothesis, we run the following code in R .

Before we begin, for this chapter we will need to make a special data set that just contains the variables `glbcc_risk` and `ideol` with their missing values removed.

```
#Filtering a data set with only variables glbcc_risk and ideol
ds.omit <- filter(ds) %>%
  dplyr::select(glbcc_risk,ideol) %>%
  na.omit()
#Run the na.omit function to remove the missing values
```

```
ols1 <- lm(glbcc_risk ~ ideol, data = ds.omit)
summary(ols1)
```

```
##
## Call:
## lm(formula = glbcc_risk ~ ideol, data = ds.omit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.726 -1.633  0.274  1.459  6.506
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)  10.81866    0.14189   76.25 <0.0000000000000002 ***
## ideol        -1.04635    0.02856  -36.63 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.479 on 2511 degrees of freedom
## Multiple R-squared:  0.3483, Adjusted R-squared:  0.348
## F-statistic: 1342 on 1 and 2511 DF, p-value: < 0.00000000000000022
```

To know whether to accept the rejecting of the null hypothesis, we need to first understand the standard error associated with the model and our coefficients. We start, therefore, with consideration of the residual standard error of the regression model.

9.1.1 Residual Standard Error

The residual standard error (or standard error of the regression) measures the spread of our observations around the regression line. As will be discussed below, the residual standard error is used to calculate the standard errors of the regression coefficients, AA and BB.

The formula for the residual standard error is as follows:

$$SE = \sqrt{\sum e_i^2 / (n - 2)} \quad (9.1) \quad SE = \sqrt{\sum e_i^2 / (n - 2)}$$

To calculate this in R, based on the model we just ran, we create an object called `Se` and use the `sqr` and `sum` commands.

```
Se <- sqrt(sum(ols1$residuals^2)/(length(ds.omit$glbcc_risk)-2))
Se
```

```
## [1] 2.479022
```

Note that this result matches the result provided by the `summary` function in R, as shown above.

For our model, the results indicate that: $\hat{Y}_i = 10.8186624 - 1.0463463X_i + e_i$. Another sample of 2513 observations would almost certainly lead to different estimates for AA and BB. If we drew many such samples, we'd get the sample distribution of the estimates. Because we typically cannot draw many samples, we need to estimate the sample distribution, based on our sample size and variance. To do that, we calculate the standard error of the slope and intercept coefficients, $SE(B)$ and $SE(A)$. These standard errors are our estimates of how much variation we would expect in the estimates of BB and AA across different samples. We use them to evaluate whether BB and AA are larger than would be expected to occur by chance if the real values of BB and/or AA are zero (the null hypotheses).

The standard error for BB, $SE(B)$ is:

$$SE(B) = SE / \sqrt{TSS_X} \quad (9.2) \quad SE(B) = SE / \sqrt{TSS_X}$$

where SE is the residual standard error of the regression, (as shown earlier in equation 9.1). TSS_X is the total sum of squares for XX , that is the total sum of the squared deviations (residuals) of XX from its mean \bar{X} ; $\sum (X_i - \bar{X})^2$. Note that the greater the deviation of XX around its mean as a proportion of the standard error of the model, the smaller the $SE(B)$. The smaller $SE(B)$ is, the less variation we would expect in repeated estimates of BB across multiple samples.

The standard error for AA, $SE(A)$, is defined as:

$$SE(A) = SE * \sqrt{1/n + \bar{X}^2 / TSS_X} \quad (9.3) \quad SE(A) = SE * \sqrt{1/n + \bar{X}^2 / TSS_X}$$

Again, the SE is the residual standard error, as shown in equation 9.1.

For AA, the larger the data set, and the larger the deviation of XX around its mean, the more precise our estimate of AA (i.e., the smaller $SE(A)$ will be).

We can calculate the SE of AA and BB in R in a few steps. First, we create an object `TSSx` that is the total sum of squares for the XX variable.

```
TSSx <- sum((ds.omit$ideol - mean(ds.omit$ideol, na.rm = TRUE))^2)
TSSx
```

```
## [1] 7532.946
```

Then, we create an object called `SEa`.

```
SEa <- Se*sqrt((1/length(ds.omit$glbcc_risk))+(mean(ds.omit$ideol,na.rm=T)^2/TSSx))
SEa
```

```
## [1] 0.1418895
```

Finally, we create SEb .

```
SEb <- Se/(sqrt(TSSx))
SEb
```

```
## [1] 0.02856262
```

Using the standard errors, we can determine how likely it is that our estimate of $\beta\beta$ differs from 00; that is how many standard errors our estimate is away from 00. To determine this we use the tt value. The tt score is derived by dividing the regression coefficient by its standard error. For our model, the tt value for $\beta\beta$ is as follows:

```
t <- ols1$coef[2]/SEb
t
```

```
##      ideol
## -36.63342
```

The tt value for our BB is 36.6334214, meaning that BB is 36.6334214 standard errors away from zero. We can then ask: What is the probability, *pp value*, of obtaining this result if $\beta=0\beta=0$? According to the results shown earlier, $p=2e-16p=2e-16$. That is remarkably close to zero. This result indicates that we can reject the null hypothesis that $\beta=0\beta=0$.

In addition, we can calculate the confidence interval (CI) for our estimate of BB. This means that in 95 out of 100 repeated applications, the confidence interval will contain $\beta\beta$.

In the following example, we calculate a 95%95% CI. The CI is calculated as follows:

$B \pm 1.96(SE(B))(9.4)B \pm 1.96(SE(B))$

We can easily calculate this in R . First, we calculate the upper limit then the lower limit and then we use the `confint` function to check.

```
Bhi <- ols1$coef[2]-1.96*SEb
Bhi
```

```
##      ideol
## -1.102329
```

```
Blow <- ols1$coef[2]+1.96*SEb
Blow
```

```
##      ideol
## -0.9903636
```

```
confint(ols1)
```

```
##                2.5 %      97.5 %  
## (Intercept) 10.540430 11.0968947  
## ideol       -1.102355 -0.9903377
```

As shown, the upper limit of our estimated BB is -0.9903636, which is far below 00, providing further support for rejecting H_0 .

So, using our example data, we tested the working hypothesis that political ideology is negatively related to the perceived risk of global warming to people and the environment. Using simple OLS regression, we find support for this working hypothesis and can reject the null.

This page titled [9.1: Hypothesis Tests for Regression Coefficients](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Jenkins-Smith et al. \(University of Oklahoma Libraries\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.