BS2280 – Econometrics I
Homework 4: Hypothesis testing - Solution

1

A researcher hypothesizes that years of schooling, S, may be related to the number of siblings (brothers and sisters), SIBLINGS, according to the relationship

$$S = \beta_1 + \beta_2 SIBLINGS + u$$

She is prepared to test the null hypothesis $H0 : \beta_2 = 0$ against the alternative hypothesis $H1 : \beta_2 \neq 0$ at the 5 % and 1 % levels. She has a sample of 60 observations. The crit,5%ical t values at the 5 % and 1 % significance level are 2.00 and 2.66 respectively. Undertake hypothesis tests for the following scenarios:

1. $\hat{\beta}_2 = -0.20$, s.e.$(\hat{\beta}_2) = 0.07$
2. $\hat{\beta}_2 = -0.12$, s.e.$(\hat{\beta}_2) = 0.07$
3. $\hat{\beta}_2 = 0.06$, s.e.$(\hat{\beta}_2) = 0.07$
4. $\hat{\beta}_2 = 0.20$, s.e.$(\hat{\beta}_2) = 0.07$

To test whether any of the coefficients provided above are statistically significant, we need to follow these steps

1. State the null and alternative hypotheses
2. Select the significance level
3. Select and calculate the test statistics
4. Set the decision rule
5. Make statistical decisions

Now we look at this question

1. Step 1: we know the null hypothesis is $H_0 : \beta_2 = 0$ and alternative hypothesis is $H_1 : \beta_2 \neq 0$.
2. Step 2: the significance levels are 5 % and 1 %.

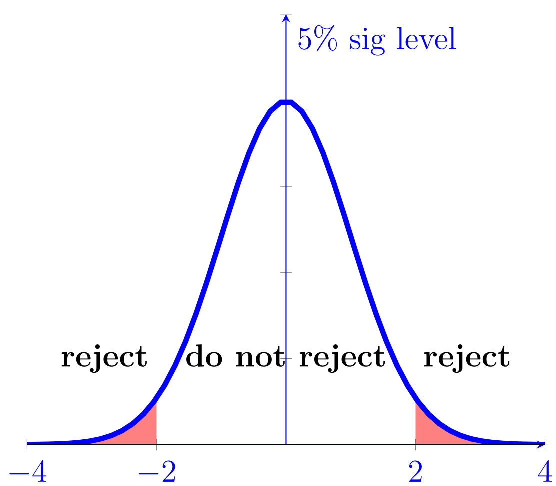
3. Step 3: calculate the t-values, which measure how many standard errors is our coefficient away from zero.

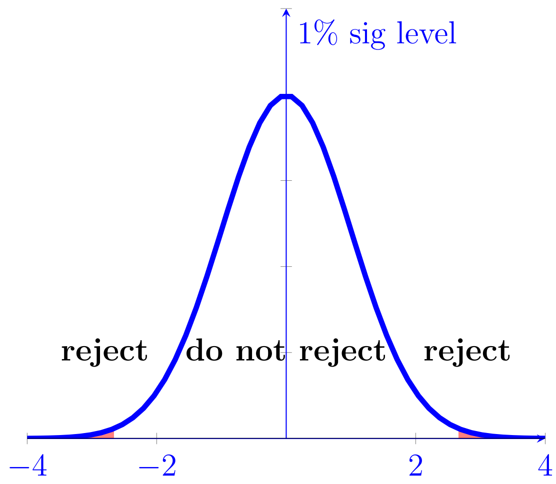
$$t = \frac{\hat{\beta}_2 - \beta_2^0}{s.e.(\hat{\beta}_2)}$$

4. Step 4: The decision rule is: If the t-value is greater than the crit,5%ical t value then we can reject the null Hypothesis and our coefficient is statistically significant.

   Please note:

   - Our test is a two tailed test, therefore the left border of the non-rejection area will be -2.00 or -2.66 and the right border of the non-rejection area will be +2.00 or +2.66.
   - Looking at the crit,5%ical values shows that to reject the null Hypothesis at a lower significance level, the t-value has to be greater than for a higher significance level! Therefore the nonrejection area at the 1% significance level is wider than at the 5% significance level.
   - If you can reject the H_0 at the 1% significance level, you can always reject it also at the 5% significance level, but if you can reject it at the 5% level does not mean that you can also reject it at the 1% significance level!
   - To get full points in the coursework, ensure that you complete the 5 steps of Hypothesis testing





5. Step 5: Make decisions

1. $\hat{\beta}_2 = -0.20$, s.e.$(\hat{\beta}_2) = 0.07$

$$t = \frac{\hat{\beta}_2 - \beta_2^0}{s.e.(\hat{\beta}_2)} = \frac{-0.20 - 0}{0.07} \approx -2.86$$

We reject H_0 at the 1 % level, as -2.86 falls into the left reject area. Because we can reject it at the 1 % level, we already know that we can also reject it at the 5% level.

2. $\hat{\beta}_2 = -0.12$, s.e.$(\hat{\beta}_2) = 0.07$

$$t = \frac{\hat{\beta}_2 - \beta_2^0}{s.e.(\hat{\beta}_2)} = \frac{-0.12 - 0}{0.07} \approx -1.71$$

Fail to reject H_0 at the 5 % level, as -1.71 is within the non-rejection area (the t-value would have to be smaller than -2.00 or greater than +2.00 so that we could reject it at the 5% level!). As we cannot reject H_0 at the 5% level, we already know that we also cannot reject it at the 1% level!

3. $\hat{\beta}_2 = 0.06$, s.e.$(\hat{\beta}_2) = 0.07$

$$t = \frac{\hat{\beta}_2 - \beta_2^0}{s.e.(\hat{\beta}_2)} = \frac{0.06 - 0}{0.07} \approx 0.86$$

Fail to reject H_0 at the 5 % level, as 0.86 falls into the middle non-rejection area. We also cannot reject it at the 1% level!

4. $\hat{\beta}_2 = 0.20$, s.e.$(\hat{\beta}_2) = 0.07$

$$t = \frac{\hat{\beta}_2 - \beta_2^0}{s.e.(\hat{\beta}_2)} = \frac{0.20 - 0}{0.07} \approx 2.86$$

Reject H_0 at the 1 % level, as 2.86 falls into the middle right reject area. We also reject it at the 5% level!

2

The number of cigarettes smoked per day is regressed on the price of cigarettes per pack in USD. Sample size $n = 807$.
The results are presented in the R output below.

```
Call:
lm(formula = cigs ~ cigpric, data = smoke)

Residuals:
   Min     1Q Median     3Q    Max
-9.224 -8.678 -8.575 11.082 71.332

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.67457    6.17296   1.729   0.0841 .
cigpric     -0.03297    0.10206  -0.323   0.7467
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.73 on 805 degrees of freedom
Multiple R-squared:  0.0001296, Adjusted R-squared:  -0.001112
F-statistic: 0.1044 on 1 and 805 DF,  p-value: 0.7467
```

1. Interpret the intercept and the coefficient of the independent variable.
   If the cigarette price per pack was 0, then individuals would smoke on average 10.67 cigarettes. A 1 USD price increase would decrease the number of cigarettes smoked by 0.032.

2. Write down the test hypotheses for testing the significance of the intercept and coefficient.
   $H_0 : \beta_1 = 0, H_1 : \beta_1 \neq 0$
   $H_0 : \beta_2 = 0, H_1 : \beta_2 \neq 0$

3. Calculate t-statistics for the intercept and the coefficient of cigarette prices. The crit,5%ical t-value at the 5% significance level is 1.96.

   Intercept:
   $t - value = \frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)} = \frac{10.67457-0}{6.17296} \approx 1.729 < t_{crit,5\%} = 1.96$
   We cannot reject H_0. The intercept is statistically insignificant.
   Cigprice:
   $t - value = \frac{\hat{\beta}_2 - \beta_2^0}{s.e.(\hat{\beta}_2)} \frac{-0.03297-0}{0.10206} \approx -0.323 > t_{crit,5\%} = -1.96$
   We cannot reject H_0. The coefficient of cigarette price is statistically insignificant

3

Calculate the 95% confidence interval for the intercept as well as the coefficient of cigarette prices using the R output in Question 2.

The formula for calculating the confidence interval for the intercept is

$$\hat{\beta}_1 - s.e.(\hat{\beta}_1) \times t_{crit,5\%} \leq \beta_1 \leq \hat{\beta}_1 + s.e.(\hat{\beta}_1) \times t_{crit,5\%}$$

The regression output shows that the point estimate $\hat{\beta}_1$ is 10.67457 and its standard error is 6.17296:

$$\hat{\beta}_1 - s.e.(\hat{\beta}_1) \times t_{crit,5\%} \leq \beta_1 \leq \hat{\beta}_1 + s.e.(\hat{\beta}_1) \times t_{crit,5\%}$$

$$10.67457 - 6.17296 \times 1.96 \leq \beta_1 \leq 10.67457 + 6.17296 \times 1.96$$

The formula for calculating the confidence interval for the cigarette coefficient is

$$\hat{\beta}_2 - s.e.(\hat{\beta}_2) \times t_{crit,5\%} \leq \beta_2 \leq \hat{\beta}_2 + s.e.(\hat{\beta}_2) \times t_{crit,5\%}$$

The regression output shows that the point estimate $\hat{\beta}_2$ is -0.03297 and its standard error is 0.10206:

$$\hat{\beta}_2 - s.e.(\hat{\beta}_2) \times t_{crit,5\%} \leq \beta_2 \leq \hat{\beta}_2 + s.e.(\hat{\beta}_2) \times t_{crit,5\%}$$

$$-0.03297 - 0.10206 \times 1.96 \leq \beta_2 \leq -0.03297 + 0.10206 \times 1.96$$

4

Calculate the F statistic for the regression undertaken in Question 2 using ESS and RSS presented in the R anova output table above. Check that the F statistic derived from R^2 is the same. Perform the F test, whereby the crit,5%ical F-value at the 5% significance level is approximately 3.8415.

The formula for the F-statistics are

$$F(1, 805) = \frac{ESS/(k-1)}{RSS/(n-k)}$$

or

$$F(1, 805) = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

```
Analysis of Variance Table

Response: cigs
           Df Sum Sq Mean Sq F value Pr(>F)
cigpric     1     20  19.672
Residuals 805 151734 188.489
```

We follow hypothesis testing five steps:

1. Step 1: the null hypothesis is $H_0 : R^2 = 0$ and alternative hypothesis is $H_1 : R^2 \neq 0$.
2. Step 2: the significance level is 5 %.
3. Step 3: calculate the F-values using either formula

$$F(1,805) = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{20/(2-1)}{151734/(807-2)} \approx 0.106$$

$$F(1,805) = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0.0001296/(2-1)}{(1-0.0001296)/(807-2)} \approx 0.104$$

4. Step 4: The decision rule is: If the F-value is greater than the $F_{crit,5\%}$ then we can reject the null Hypothesis, so our model is statistically significant.

$$F(1,805) \approx 0.104 < F_{crit,5\%} = 3.8415$$

5. Step 5: Make decisions
   Therefore we cannot reject the H_0 and our model is statistically insignificant.

5

The number of cigarettes per day is regressed on the age of participants. Use the R output tables below to answer Questions 2 – 4 again.

```
Call:
lm(formula = cigs ~ age, data = smoke)

Residuals:
   Min     1Q Median     3Q    Max
-9.498 -8.929 -7.991 10.669 71.372

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.06698    1.26597
age         -0.03348    0.02838
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.72 on 805 degrees of freedom
Multiple R-squared:  0.001726,  Adjusted R-squared:  0.0004856
F-statistic:      on 1 and 805 DF,  p-value:


Analysis of Variance Table

Response: cigs
           Df Sum Sq Mean Sq F value Pr(>F)
age         1    262  261.88
Residuals 805 151492  188.19
```

See results in table below.

```
Call:
lm(formula = cigs ~ age, data = smoke)

Residuals:
   Min      1Q Median      3Q     Max
-9.498 -8.929 -7.991 10.669 71.372

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.06698    1.26597   7.952  6.2e-15 ***
age         -0.03348    0.02838  -1.180    0.238

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.72 on 805 degrees of freedom
Multiple R-squared:  0.001726,  Adjusted R-squared:  0.0004856
F-statistic: 1.392 on 1 and 805 DF,  p-value: 0.2385


95% Confidence interval:
                 2.5 %      97.5 %

(Intercept)  7.5819831 12.55198464
age         -0.0891801  0.02222761


Analysis of Variance Table


Response: cigs
           Df Sum Sq Mean Sq F value Pr(>F)
age         1    262  261.88  1.3916 0.2385
Residuals 805 151492  188.19
```