In-Class Lab 7

ECON 425 (Justin Heflin, West Virginia University)

March 3, 2023

The purpose of this lab is to practice using R to conduct hypothesis tests. The lab may be completed as a group. To receive credit, upload your .R script to the appropriate place on eCampus ("In-Class Labs'' folder).

For starters

Open a new R script (named `ICL7_XYZ.R`, where `XYZ` are your initials)

Clean out/"Sweep'' R Studio

Click the broom in the Environment panel (top-right), it is directly below the Tutorial button. Also, in the bottom-right panel, click the Plots button and then click the broom in that panel. This should help with loading things into R.

Hypothesis Testing

$$colGPA = \beta_0 + \beta_1 hsGPA + \beta_2 ACT + \beta_3 skipped + \epsilon$$

- colGPA: college GPA
- hsGPA: high school GPA
- ACT: ACT score
- skipped: average lectures missed per week

Let's develop hypotheses for each slope coefficient:

$H_0 : \beta_{hsGPA} = 0; \quad H_0 : \beta_{ACT} = 0; \quad H_0 : \beta_{skipped} = 0$

$H_A : \beta_{hsGPA} \neq 0; \quad H_A : \beta_{ACT} \neq 0; \quad H_A : \beta_{skipped} \neq 0$

Each hypothesis test for slope coefficients above are two-tailed.

$$\widehat{colGPA} = 1.39 + \underset{(.094)}{.412} hsGPA + \underset{(.011)}{.015} ACT - \underset{(.026)}{.083} skipped$$

Calculate t-statistic for each slope coefficient

$$t_{hsGPA} = \frac{.412}{.094} \rightarrow 4.383 \qquad t_{ACT} = \frac{.015}{.011} \rightarrow 1.364 \qquad t_{skipped} = \frac{-.083}{.026} \rightarrow -3.192$$

Degrees of Freedom (DF) = 137 (N - K - 1 $\implies$ 141 - 3 - 1)

Select level of significance: $\alpha = 5\%$ (0.05)

- Critical t-value based on the DF, chosen level of significance, and two-tail test is about 1.96 from t-table

Compare $t_{skipped}$ = -3.192 to critical t-value $t_c = 1.96$ using the decision rule:

(Reject H_0 if $|t_k| > t_c$ AND if t_k also has the sign implied by H_A)

$\implies |-3.192| > 1.96$; $\beta_{skipped} \neq 0$ which matches our H_A

This means we can reject the null hypothesis can conclude skipping lectures does indeed tend to have a negative relationship with college GPA (holding the other variables in the equation constant)

Hypothesis Testing in R

```
library(wooldridge)
GPA_data <- as.data.frame(gpa1)
```

```
regression1 <- lm(colGPA ~ hsGPA + ACT + skipped, data = GPA_data)
summary(regression1)
```

```
##
## Call:
## lm(formula = colGPA ~ hsGPA + ACT + skipped, data = GPA_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.85698 -0.23200 -0.03935  0.24816  0.81657
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.38955    0.33155   4.191 4.95e-05 ***
## hsGPA        0.41182    0.09367   4.396 2.19e-05 ***
## ACT          0.01472    0.01056   1.393  0.16578
## skipped     -0.08311    0.02600  -3.197  0.00173 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3295 on 137 degrees of freedom
## Multiple R-squared:  0.2336, Adjusted R-squared:  0.2168
## F-statistic: 13.92 on 3 and 137 DF,  p-value: 5.653e-08
```

Now let's take a look at the confidence interval for each regression coefficient

```
confint(regression1, level = 0.95)
```

```
##                    2.5 %      97.5 %
## (Intercept)  0.733929519  2.04517814
## hsGPA        0.226581851  0.59705049
## ACT         -0.006171074  0.03561154
## skipped     -0.134523444 -0.03170283
```

```
install.packages("GGally")
```

```
library(ggplot2)
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
ggcoef(regression1, exclude_intercept = TRUE, vline_linetype = "solid",
       vline_color = "blue", errorbar_color = "darkorange",
       errorbar_height = 0.15)
```