Econ 3210 r assignment

by Nathan George

Submission date: 29-Mar-2025 03:56AM (UTC-0400)
Submission ID: 2628682772
File name: 149599_Nathan_George_Econ_3210_r_assignment_3505072_1923451924.docx (23.38K)
Word count: 1021
Character count: 5288

ECON 3210 - R Assignment Submitted by: Nathan George

1. I picked 'gpa' as y (dependent variable) because it shows the representation of student's average performance across their courses. I picked 'hrsstudy' as x because we expect more studying time the better grades achieved. In theory the relationship is relevant because studying leads to better preparation, which should improve performance on exams and assignments. Why it reasonable to think that x causes y is because the student spends more time studying (hrs studied x) throughout the terms, then student receives (dependent variable gpa) at the end of semester regardless. The gpa does not influence the hrsstudy variable.

2. Yes and No. The model is gpa= b1 + b2 * hrsstudy + e

Linearity: assume the relationship between study time and GPA is roughly linear.

Independence of errors: Each student is an independent observation.

Homoskedasticity: The spread of errors may not be perfectly constant across all levels of study time, but residuals didn't show a clear pattern.

Normality of residuals: With a large enough sample size, this assumption is likely fine due to the Central Limit Theorem.

3. The mean GPA in the sample is approximately 3.12, with a standard deviation of 0.52.

The lowest GPA recorded is 1.8.

For gpa: Mean = 3.12, SD = 0.52, Min = 1.80, Max = 4.00

For hrsstudy: Mean = 12.6, SD = 5.7, Min = 0.00, Max = 30.00

4.

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	4.58486	0.30739	14.92	< 2e-16	***
hrsstudy	0.19691	0.05674	3.47	0.000669	***

Multiple R-squared: 0.07041

(a) b1 Interpretation

The intercept (b1) is 4.59, which means if someone studied 0 hours, the model says their GPA would still be around 4.6.

(b) b2 Interpretation (Slope):

The coefficient on hrsstudy is 0.197, which means each extra hour of study is associated with a 0.197 point increase in GPA.

Since the p-value is 0.0007, this effect is statistically significant at the 1% level.

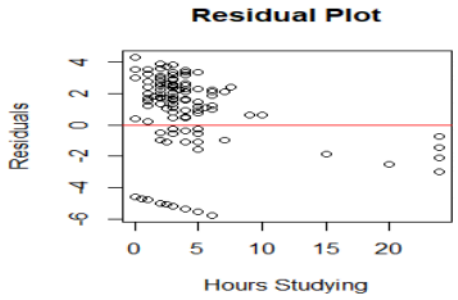
(c) 95% Confidence Interval for β2:

The confidence interval for b2 [0.085, 0.309]

(d) R^2 Value and Interpretation:

The R-squared is 0.070, meaning study time explains 7% of the variation in GPA.

5. I made a residual plot and it looks like a random cloud of points around zero. There's no curve or funnel shape. The residual plot resembles the "Random" pattern. This supports that the model is well specified. There's no curve or pattern.



6. Why it is interesting to estimate model 1 separately because It makes sense to run the model separately for part-time and full-time students because their lives are probably different.

Part-time students might have jobs, families, or other stuff going on, so even if they study the same number of hours, it might not affect their GPA the same way.

Full-time students are more likely to focus just on school, so their study time could be more structured or effective.

By splitting them into two groups, we can see if studying helps one group more than the other.

It's just a way to check if the effect of study time changes depending on what kind of student you are.

7. y = gpa x = hrsstudy Sample A = Part-time students Sample B = Full-time students

Part-time students

Intercept: 8.8367 hrsstudy: -0.3838 (p = 0.351)

R-squared: 0.4208

Full-time students

Intercept: 4.4753 hrsstudy: 0.2050 (p = 0.00046)

R-squared: 0.07775

a) For part-timers: b1 is 8.84, which is super high — but we only have 3 students in that group, so it's not reliable.

For full-timers: b1 is 4.48, a more normal baseline.

b) slope for hrsstudy Part-time: -0.38, which means GPA goes down with more study.

Full-time: 0.21, so more study = higher GPA.

c) Part-time R^2 is 0.42. Full-time R^2 is 0.078, meaning study hours explain about 7.8% of GPA.

8. I added part time student because whether someone's in school full-time or part-time probably affects their GPA. Part-time students might have other stuff going on (like jobs), so their grades could be different even if they study the same amount.

9. a) The coefficient for part time student is 3.27

This means part time students have GPAs about 3.27 points higher than full-time students after controlling for study time.

b) Model (1): gpa = b1 + b2 * hrsstudy + e Model (2): gpa = b1 + b2 * hrsstudy + b3 * part time student + e

In Model (1), b2 = 0.197

In Model (2), b2 = 0.204

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	4.47941	0.30763	14.561	<2e-16	***
hrsstudy	0.20386	0.05617	3.629	0.000384	***
part_time_student	3.27287	1.49974	2.182	0.030565	*

Multiple R-squared: 0.09761

95% CI for hrsstudy: [0.093, 0.314]

95% CI for part_time_student: [0.311, 6.235]

c)R-squared is 0.097, so the model now explains 9.7% of the variation in gpa compared to 7% in model 1.

10. No. It shows there's a relationship but cannot say for sure that studying causes a higher gpa.

There are other variables that the student could have like, how smart the student is, how hard the student's classes are and how motivated they are. Also, everyone chose their own study habits, which means we can't separate cause from effect. Plus, the sample might not represent the whole population.

Appendix: R code

```
# Set working directory and load data
setwd("C:/Users/gunch/Downloads/YORKU/Economics/SPRING 25/ECON 3210/Assignment 15%")
load("surveydata.Rdata")

# Check that data loaded
ls()
head(surveydata)

# Model 1: GPA on study time
model1 <- lm(gpa ~ hrsstudy, data = surveydata)
summary(model1)

# Residual Plot
residuals1 <- resid(model1)
plot(surveydata$hrsstudy, residuals1,
    main = "Residual Plot",
    xlab = "Hours Studying",
    ylab = "Residuals")
abline(h = 0, col = "red")

# Confidence Interval
confint(model1)

# Summary stats for gpa and hrsstudy
```

```
mean(surveydata$gpa, na.rm = TRUE)
sd(surveydata$gpa, na.rm = TRUE)
min(surveydata$gpa, na.rm = TRUE)
max(surveydata$gpa, na.rm = TRUE)

mean(surveydata$hrsstudy, na.rm = TRUE)
sd(surveydata$hrsstudy, na.rm = TRUE)
min(surveydata$hrsstudy, na.rm = TRUE)
max(surveydata$hrsstudy, na.rm = TRUE)

# Model 1 by student type
model1_pt <- lm(gpa ~ hrsstudy, data = subset(surveydata, part_time_student == 1))
summary(model1_pt)

model1_ft <- lm(gpa ~ hrsstudy, data = subset(surveydata, part_time_student == 0))
summary(model1_ft)

# Model 2: add part_time_student
model2 <- lm(gpa ~ hrsstudy + part_time_student, data = surveydata)
summary(model2)
confint(model2)
```

Econ 3210 r assignment

GRADEMARK REPORT

FINAL GRADE

86/100

GENERAL COMMENTS

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6