

CR Portfolio 5

Seunghun Lee

2020-11-22 15:57:49

Contents

| | |
|---|---|
| Research Question | 1 |
| Hypotheses | 1 |
| Critical Test Statistics | 2 |
| Sample test statistic results | 2 |
| Conclusion | 3 |
| Visualization | 4 |

Research Question

- A researcher was interested in investigating whether the relationship between test scores of Portuguese class (G1) and weekly study time (numeric: 1 - less than 2 hours; 2 - 2 to 5 hours; 3 - 5 to 10 hours; or 4 - more than 10 hours) was moderated by eagerness to take higher education (binary: yes or no). The researcher tested a hypothesis using a random sample of 649 students in secondary education of Portuguese schools (see student-por.csv).

```
# https://archive.ics.uci.edu/ml/datasets/Student+Performance
perf <- read.csv("./data/student-por.csv", sep = ";") %>%
  select(G1, studytime, higher) %>%
  mutate(across(where(is.character), as.factor))
str(perf)
```

```
## 'data.frame': 649 obs. of 3 variables:
## $ G1 : int 0 9 12 14 11 12 13 10 15 12 ...
## $ studytime: int 2 2 2 3 2 2 2 2 2 2 ...
## $ higher : Factor w/ 2 levels "no","yes": 2 2 2 2 2 2 2 2 2 2 ...
```

Hypotheses

- H_1 : Eagerness to take higher education moderates the relationship between test scores of Portuguese class and weekly study time.
- H_0 : Eagerness to take higher education does not moderate the relationship between test scores of Portuguese class and weekly study time.

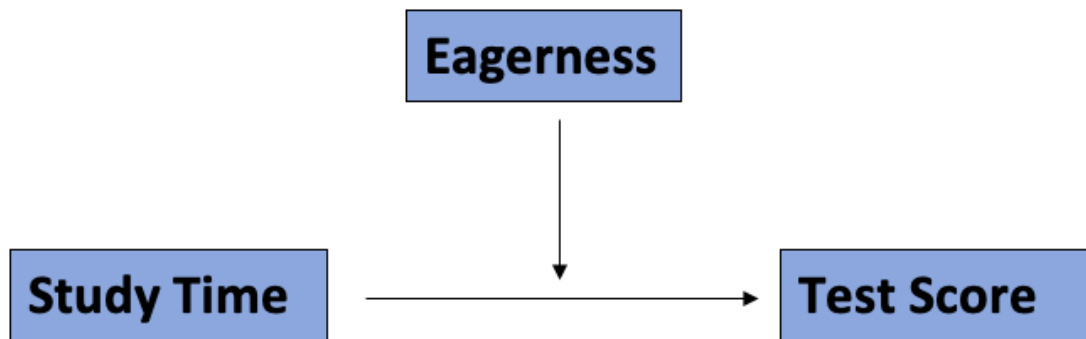


Figure 1: moderation

Critical Test Statistics

- $\alpha = 0.05$, two-tailed, $df = 649 - 3 - 1 = 645$

```
qt(p = 0.975, df = nrow(perf) - 4)
```

```
## [1] 1.963649
```

- Critical $t = \pm 1.964$

Sample test statistic results

- Mean-centered the exogenous variable (studytime) - creating variable **studytime.ctr**

```
perf <- perf %>%
  mutate(studytime.ctr = scale(studytime, center = TRUE, scale = FALSE))
head(perf)
```

```
##   G1 studytime higher studytime.ctr
## 1  0         2   yes    0.06933744
## 2  9         2   yes    0.06933744
## 3 12         2   yes    0.06933744
## 4 14         3   yes    1.06933744
## 5 11         2   yes    0.06933744
## 6 12         2   yes    0.06933744
```

- Dummy coded the moderator variable (higher) - 0 - No; 1 - Yes.

```
summary(lm(G1 ~ studytime.ctr * higher, data = perf))
```

```
##
## Call:
## lm(formula = G1 ~ studytime.ctr * higher, data = perf)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.7409  -1.7409   0.0063   1.5118   7.0063
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      8.5555     0.3551  24.092  <2e-16 ***
## studytime.ctr     -0.1495     0.4117  -0.363   0.7166
## higheryes         3.1336     0.3702   8.465  <2e-16 ***
## studytime.ctr:higheryes  0.8968     0.4307   2.082   0.0377 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.512 on 645 degrees of freedom
## Multiple R-squared:  0.1669, Adjusted R-squared:  0.163
## F-statistic: 43.08 on 3 and 645 DF,  p-value: < 2.2e-16
```

Standardized Coefficients

```
summary(lm(scale(G1) ~ scale(studytime.ctr)*higher, data = perf))
```

```
##
## Call:
## lm(formula = scale(G1) ~ scale(studytime.ctr) * higher, data = perf)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.2768 -0.6341   0.0023   0.5507   2.5522
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -1.03580     0.12935  -8.007 5.48e-15 ***
## scale(studytime.ctr) -0.04518     0.12439  -0.363   0.7166
## higheryes         1.14144     0.13484   8.465 < 2e-16 ***
## scale(studytime.ctr):higheryes  0.27097     0.13015   2.082   0.0377 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9149 on 645 degrees of freedom
## Multiple R-squared:  0.1669, Adjusted R-squared:  0.163
## F-statistic: 43.08 on 3 and 645 DF,  p-value: < 2.2e-16
```

Conclusion

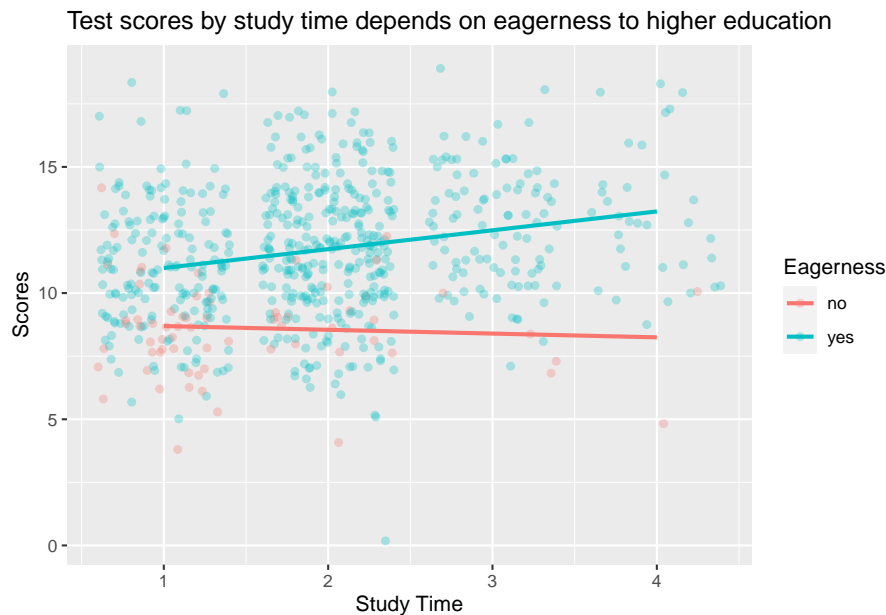
- The results led me to reject H_0 and infer that eagerness to take higher education significantly moderates the relationship between test scores and weekly study time.
- $[B = 0.897, \beta = 0.271, t(645) = 2.082, p < 0.05]$

Visualization

- I used the uncentered values on the predictor for the interaction plot.

```
perf %>%  
  ggplot(aes(x = studytime, y = G1, color = higher)) +  
  geom_jitter(alpha = 0.3) +  
  geom_smooth(se = FALSE, method = "lm") +  
  labs(title = "Test scores by study time depends on eagerness to higher education",  
        x = "Study Time",  
        y = "Scores",  
        color = "Eagerness")
```

```
## 'geom_smooth()' using formula 'y ~ x'
```



- The pattern of results indicates that the prediction of test scores (G1) by weekly study time depends on the eagerness to higher education of the person. The significant interaction effect that was found is reflected in the different pattern of results that depend on the eagerness to higher education. It shows an ordinal interaction - students with eagerness to higher education always had higher scores on the Portuguese test than ones without eagerness to higher education. There is a positive relationship between test scores and weekly study time in students with eagerness to higher education. On the other hand, students without eagerness to higher education received the lower test scores as they increased the study time (negative relationship).