Ch4 - Ex14

Thalles Quinaglia Liduares

31/03/2023

Use the data in ATTEND.RAW from wooldridge r package to answer this question. a. To determine the effects of attending lecture on final exam performance, estimate a model relating stndfnl (the standardized final exam score) to atndrte (the percent of lectures attended). Include the binary variables frosh and soph as explanatory variables. Interpret the coeffcient on atndrte, and discuss its significance.

Upload package

```
library(wooldridge)
```

Upload data

```
data<-wooldridge::attend
```

Estimate model by OLS

```
model <- lm(stndfnl ~ atndrte + frosh + soph, data)
summary(model)
```

```
##
## Call:
## lm(formula = stndfnl ~ atndrte + frosh + soph, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2825 -0.6719 -0.0295  0.6791  2.5455
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.501731   0.196314  -2.556 0.010813 *
## atndrte      0.008163   0.002203   3.705 0.000228 ***
## frosh       -0.289894   0.115724  -2.505 0.012478 *
## soph        -0.118446   0.099027  -1.196 0.232078
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9772 on 676 degrees of freedom
## Multiple R-squared:  0.02904,    Adjusted R-squared:  0.02473
## F-statistic: 6.739 on 3 and 676 DF,  p-value: 0.0001752
```

The coefficient on atndrte is 0.008163, which means that, on average, attending one additional lecture is associated with a 0.008163 standard deviation increase in final exam scores, holding constant the effects of being a freshman or sophomore. This coefficient is statistically significant. Therefore, we can conclude that attending lectures has a positive and significant effect on final exam performance.

b. How confident are you that the OLS estimates from part a are estimating the causal e¤ect of attendence? Explain.

Overall, while the OLS estimates in part a provide some evidence of a positive correlation between attendance and final exam scores, we cannot be fully confident that they are estimating the causal effect of attendance without addressing potential sources of bias and endogeneity.

c. As proxy variables for student ability, add to the regression priGPA (prior cumulative GPA) and ACT (achievement test score). Now what is the e¤ect of atndrte?

Discuss how the effect differs from that in part a

```
model2 <- lm(stndfnl ~ atndrte + frosh + soph + priGPA + ACT, data)
summary(model2)
```

```
##
## Call:
## lm(formula = stndfnl ~ atndrte + frosh + soph + priGPA + ACT,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1904 -0.5668 -0.0252  0.5887  2.2915
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.297342   0.308831 -10.677  < 2e-16 ***
## atndrte      0.005225   0.002384   2.191   0.0288 *
## frosh       -0.049469   0.107890  -0.459   0.6467
## soph        -0.159648   0.089772  -1.778   0.0758 .
## priGPA       0.426585   0.081920   5.207 2.55e-07 ***
## ACT          0.084412   0.011168   7.559 1.33e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8851 on 674 degrees of freedom
## Multiple R-squared:  0.2058, Adjusted R-squared:  0.1999
## F-statistic: 34.93 on 5 and 674 DF,  p-value: < 2.2e-16
```

The coefficient on atndrte is now 0.005225, which is smaller than the coefficient in part a. However, it is still statistically significant at the 5% level. This suggests that attending lectures has a positive and significant effect on final exam performance, even after controlling for prior cumulative GPA and achievement test scores.

Overall, the effect of attending lectures on final exam performance in part c is smaller than in part a, which suggests that some of the effect of attendance is being captured by prior cumulative GPA and achievement test scores. However, attending lectures still has a significant positive effect on final exam performance, even after controlling for these variables.

d. What happens to the significance of the dummy variables in part c as compared with part a? Explain.

In part c, we added priGPA and ACT as explanatory variables, which are both highly correlated with final exam scores and likely to be correlated with the attendance variable as well. As a result, some of the variation in final exam scores that was previously being captured by the dummy variables frosh and soph in part a is now being explained by the new explanatory variables in part c.

This means that the estimated coefficients for frosh and soph may be smaller in magnitude and less significant in part c compared to part a, since their explanatory power is now being shared with the other variables in the model.

e. Add the squares of priGPA and ACT to the equation. What happens to the coefficient on atndrte? Are the quadratics jointly significant?

```
model3 <- lm(stndfnl ~ atndrte + frosh + soph + priGPA + ACT+I(priGPA^2)+I(ACT^2), data)
summary(model3)
```

```
##
## Call:
## lm(formula = stndfnl ~ atndrte + frosh + soph + priGPA + ACT +
##     I(priGPA^2) + I(ACT^2), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.14241 -0.54902 -0.02163  0.56155  2.36547
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.384812   1.239361   1.117  0.26424
## atndrte      0.006232   0.002358   2.642  0.00842 **
## frosh       -0.105337   0.106975  -0.985  0.32513
## soph        -0.180729   0.088635  -2.039  0.04184 *
## priGPA      -1.526140   0.473972  -3.220  0.00134 **
## ACT         -0.112433   0.098172  -1.145  0.25251
## I(priGPA^2)  0.368218   0.088985   4.138 3.95e-05 ***
## I(ACT^2)     0.004182   0.002169   1.928  0.05425 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8718 on 672 degrees of freedom
## Multiple R-squared:  0.2316, Adjusted R-squared:  0.2236
## F-statistic: 28.94 on 7 and 672 DF,  p-value: < 2.2e-16
```

```
library(car)
```

```
## Carregando pacotes exigidos: carData
```

```
linearHypothesis(model3, c("I(priGPA^2)=0", "I(ACT^2)=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## I(priGPA^2) = 0
## I(ACT^2) = 0
##
## Model 1: restricted model
## Model 2: stndfnl ~ atndrte + frosh + soph + priGPA + ACT + I(priGPA^2) +
##     I(ACT^2)
##
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    674 527.96
## 2    672 510.79  2    17.172 11.296 1.496e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that the coefficient on atndrte has increased to 0.006, indicating a larger positive effect of attendance on final exam scores compared to part c. The squared terms for priGPA and ACT are both statistically significant, indicating nonlinear relationships between these variables and final exam scores. Also, based on F-Test, we observe that quadratic terms are jointly significant in explain the endogenous variable.

f. To test for a nonlinear effect of atndrte, add its square to the equation from part e. What do you conclude?

```
model4 <- lm(stndfnl ~ atndrte + frosh + soph + priGPA + ACT+I(priGPA^2)+I(ACT^2)+I(atndrte^
2), data)
summary(model4)
```

```
##
## Call:
## lm(formula = stndfnl ~ atndrte + frosh + soph + priGPA + ACT +
##     I(priGPA^2) + I(ACT^2) + I(atndrte^2), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.14165 -0.54816 -0.02205  0.55940  2.36637
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.394e+00  1.267e+00   1.100  0.27159
## atndrte       5.843e-03  1.092e-02   0.535  0.59282
## frosh        -1.054e-01  1.071e-01  -0.984  0.32537
## soph         -1.808e-01  8.875e-02  -2.038  0.04199 *
## priGPA       -1.525e+00  4.757e-01  -3.205  0.00141 **
## ACT          -1.123e-01  9.828e-02  -1.143  0.25339
## I(priGPA^2)   3.679e-01  8.944e-02   4.113 4.38e-05 ***
## I(ACT^2)      4.180e-03  2.171e-03   1.925  0.05461 .
## I(atndrte^2)  2.873e-06  7.871e-05   0.036  0.97089
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8725 on 671 degrees of freedom
## Multiple R-squared:  0.2316, Adjusted R-squared:  0.2225
## F-statistic: 25.28 on 8 and 671 DF,  p-value: < 2.2e-16
```

There is no significant evidence to suggest that the quadratic effect of attendance rate I(atndrte^2) has an impact on the final exam score, as the p-value is greater than 0.05. Therefore, we can conclude that the variable I(atndrte^2) is not statistically significant in explaining the variation in the final exam score, and it can be removed from the model without affecting the overall model fit.