Erasmus Econometrics - Test Assignment 2

Harro Cyranka

7/21/2017

Question (a):

(i) Regress FGPA on a constant and SATV. Report the coefficient of SATV and its standard error and p-value (give your answers with 3 decimals).

```
## [1] "The coefficient for SAT Verbal Test is: 0.063"
## [1] "The Standard Error of the coefficient for SAT Verbal Test is: 0.028"
## [1] "T-statistic of the coefficient for SAT Verbal Test is: 2.28"
## [1] "P-Value Associated with the t-statistic for the coefficient for SAT Verbal Test is: 0.023"
```

(ii) Determine a 95% confidence interval (with 3 decimals) for the effect on FGPA of an increase by 1 point in SATV.

```
#(ii) Calculating a 95% Confidence Interval for SATV
round(confint(fgpa_satv, 'SATV',level = 0.95),3)
```

```
##      2.5 % 97.5 %
## SATV 0.009  0.117
```

Question (b):

- Answer questions (a-i) and (a-ii) also for the regression of FGPA on a constant, SATV, SATM, and FEM.

(i) Regression coefficient, standard error, and corresponding p-value for SATV in the new multiple regression setting:

```
## [1] "The coefficient for SAT Verbal Test is: 0.014"
## [1] "The Standard Error of the coefficient for SAT Verbal Test is: 0.028"
## [1] "T-statistic of the coefficient for SAT Verbal Test is: 0.507"
## [1] "P-Value Associated with the t-statistic for the coefficient for SAT Verbal Test is: 0.612"
```

(ii) 95% confidence interval (with 3 decimals) for the effect on FGPA of an increase by 1 point in SATV in the multiple regression setting.

```
#Calculating a 95% Confidence Interval for SATV
round(confint(fgpa_all, 'SATV', level = 0.95), 3)
```

```
##       2.5 % 97.5 %
## SATV -0.041  0.069
```

Question (c):

- Determine the (4×4) correlation matrix of FGPA, SATV, SATM, and FEM. Use these correlations to explain the differences between the outcomes in parts (a) and (b).

Here is the correlation matrix (Since the matrix is symmetric, I made the elements above the diagonal blank):

```
##      FGPA    SATM     SATV   FEM
## FGPA "1"     ""       ""     ""
## SATM "0.2"   "1"      ""     ""
## SATV "0.09"  "0.29"   "1"    ""
## FEM  "0.18"  "-0.16"  "0.03" "1"
```

Commentary on the results:

By checking the correlation matrix we observe that the SAT verbal score is not the only variable that has a linear relationship with the outcome. A student's SAT math score and the student's gender are also correlated with freshman GPA. The results in the first part come from a simple regression model, where SATV is the only independent variable. Therefore, the variation in the outcome is explained solely by the variation in SATV and by the variation in the error term. On average, we assume that the error term has, on average, no effect on the expected value of y.

However as the correlation matrix shows, there are other variables that could potentially be useful in our model. Once we include them in the model and control for them (as we did in the multiple regressions setting), the initial relationship between SATV and the FGPA changes. The relationship between SATV and the variation in the y that remains after controlling for the other variables is not statistically significant anymore.

Question (d):

(i) Perform an F-test on the significance (at the 5% level) of the effect of SATV on FGPA, based on the regression in part (b) and another regression.

```
##Performing an F-Test with the linear restriction that coef for SATV is equal to zero
fgpa_minus_satv <- lm(FGPA ~ FEM + SATM, data = x)

##Getting the restricted and unrestricted R^2
r2_unrestricted <- summary(fgpa_all)$r.squared
r2_restricted <- summary(fgpa_minus_satv)$r.squared

f_numerator <- (r2_unrestricted - r2_restricted)/1
f_denominator <-(1 - r2_unrestricted)/(nrow(x) - 4)

f_value <- round(f_numerator/f_denominator,6) ##F-Value

##F-Value
print(f_value)
```

```
## [1] 0.257155
```

```
##P-Value Associated with the F-Value
print(1 - pf(f_value, 1,605))
```

```
## [1] 0.6122663
```

(ii) Check numerically that F = t^2

The code below performs that operation automatically:

```
t_value_for_SATV = round(coef(summary(fgpa_all))[,"t value"][4],6)
round(t_value_for_SATV**2,6) == round(f_value,6)
```

```
## SATV
## TRUE
```