TestExercise2

Jason

17/12/2017

Introduction

This test exercise is of an applied nature and uses data that are available in the data file TestExer2. The exercise is based on Exercise 3.14 of 'Econometric Methods with Applications in Business and Economics'. The question of interest is whether the study results of students in Economics can be predicted from the scores on entrance tests taken before they start their studies. More precisely, you are asked to investigate whether verbal and mathematical entrance tests predict freshman grades of students in Economics. Data are available for 609 students on the following variables:

- FGPA: Freshman grade point average (scale 0-4)
- SATV: Score on SAT Verbal test (scale 0-10)
- SATM: Score on SAT Mathematics test (scale 0-10)
- FEM: Gender dummy (1 for females, 0 for males)

Loading in the dataset

Question 1a

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).

```
simpleModel <- lm(resultsData$FGPA~resultsData$SATV)
summary(simpleModel)$coefficients
```

```
##                   Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)      2.44173246 0.15506207 15.746807 4.257353e-47
## resultsData$SATV 0.06308585 0.02766362  2.280462 2.292611e-02
```

- Coefficient of SATV = 0.063
- Standard Error = 0.028
- p-value = 0.023

Question 1b

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

```
confint(simpleModel, 'resultsData$SATV', level=0.95)
```

```
##                         2.5 %    97.5 %
## resultsData$SATV 0.008757813 0.1174139
```

- Confidence Interval = (0.009, 0.117)

Question 2a

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

```
complexModel <- lm(resultsData$FGPA~resultsData$SATV + resultsData$SATM + resultsData$FEM)
summary(complexModel)$coefficients
```

```
##                   Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)      1.5570482 0.21609551 7.2053704 1.729863e-12
## resultsData$SATV 0.0141619 0.02792697 0.5071047 6.122662e-01
## resultsData$SATM 0.1727359 0.03192671 5.4103874 9.071480e-08
## resultsData$FEM  0.2002716 0.03738085 5.3575989 1.200266e-07
```

- Coefficient of SATV = 0.014
- Standard Error = 0.028
- p-value = 0.612

Question 2b

```
confint(complexModel, 'resultsData$SATV', level=0.95)
```

```
##                         2.5 %     97.5 %
## resultsData$SATV -0.04068368 0.06900747
```

- Confidence Interval = (-0.041, 0.069)

Question 3

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).

```
resultsData$Observ. <- NULL
cor(resultsData)
```

```
##             FGPA       SATM       SATV         FEM
## FGPA 1.00000000  0.1950404 0.09216712  0.17649071
## SATM 0.19504042  1.0000000 0.28780108 -0.16268037
## SATV 0.09216712  0.2878011 1.00000000  0.03357664
## FEM  0.17649071 -0.1626804 0.03357664  1.00000000
```

The correlation matrix indicates the strength of the relantionship between the variables. Values closest to -1 or +1 indicate that they have a strong correlation with each other. For e.g. if the value is equal to 1 for two variables x and y, that means when x increases, y also increases.

The coefficient in answer (a) for SATV was much higher than answer (b) because it included for the direct effect caused by the other variables SATM and FEM as they are positiviely correlated with SATV.

Question 4a

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. Note: Use the F-test in terms of SSR or R2 and use 6 decimals in your computations. The relevant critical value is 3.9.

```
unrestricted <- lm(resultsData$FGPA ~resultsData$SATV + resultsData$SATM + resultsData$FEM)
restricted <- lm(resultsData$FGPA ~resultsData$SATM + resultsData$FEM)

unrestrictedR <- summary(unrestricted)$r.squared

restrictedR <- summary(restricted)$r.squared
```

Number of regressors excluded

```
g <- 2
```

Number of observations

```
n <- 609
```

Number of regressors in unrestricted model

```
k <- 3
```

```
fValue <- (((unrestrictedR) - (restrictedR))/g)/((1-(unrestrictedR))/(n-k))
fValue
```

```
## [1] 0.1287901
```

- F-value = 0.257

Question 4b

Check numerically that F = t^2

```
0.257^2
```

```
## [1] 0.066049
```