Predicting Freshman grade using entrance exam scores

Sanjaya J Shetty

29/05/2021

Question and Answers

- **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).
- **Answers:**

Table 1	Estimate	std error	p-value
(Intercept)	2.44173	0.15506	<2e-16
SATV	0.06309	0.02766	0.0229

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

- **Answers:**

Confidence Interval (95%) = intercept +/- (2*(SE))

Lower Limit = 0.0630 - 2*0.02766 = 0.0076

Upper Limit = 0.0630 + 2*0.02766 = 0.1183

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

- **Answers:**

Table 2	Estimate	std error	p-value
(Intercept)	1.55705	0.21610	1.73e-12
SATV	0.01416	0.02793	0.612
SATM	0.17274	0.03193	9.07e-08
FEM	0.20027	0.03738	1.20e-07

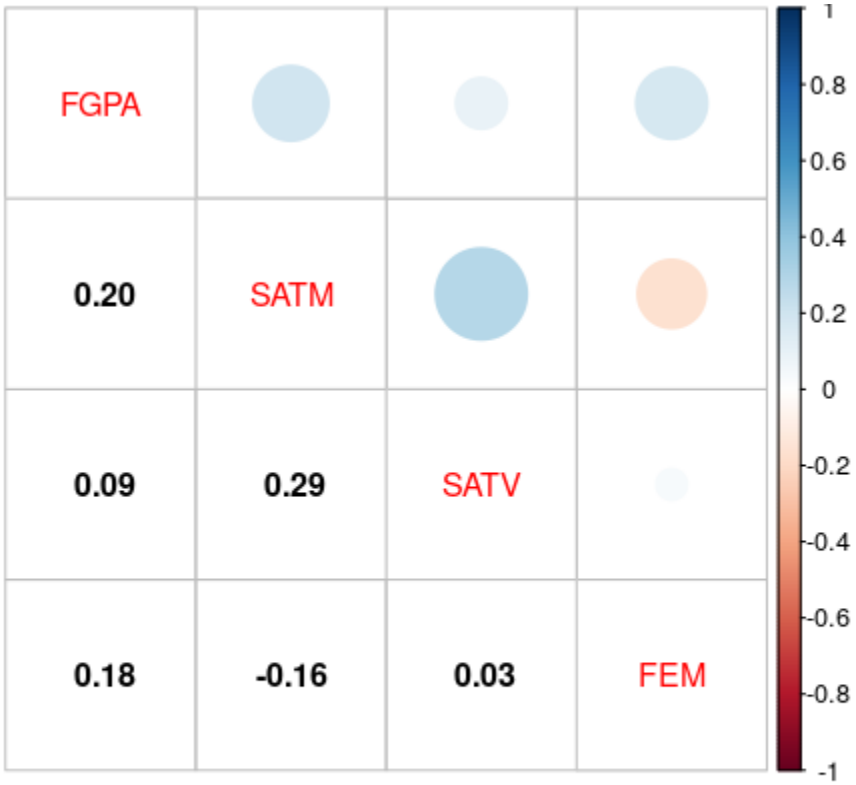
Confidence Interval (95%) = intercept +/- (2*(SE))

Lower Limit = 0.01416 - 2*0.02766 = -0.04116

Upper Limit = 0.01416 + 2*0.02766 = 0.06948

- **(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).

- **Answers:**



when u check the p value in table 2, we could see for SATV the p-value will be 0.612, hence it is statistically insignificant.

- **(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.

- **Answers:**

F statistic = 0.2572

(*Check Anova Table*)

or can it can also be found using R^2 values from the model2 and model3

- **(ii)** Check numerically that F = t^2.

- **Answers:**

t value from the model2 is 0.507

hence $t^2 = (0.507)^2 = 0.257 = F$

Method

Load the library

```
library(readxl)
```

Load the Data

```
setwd("~/Personal not your shit/H./Online Classes/R programming/Econometrics/Week2")

data <- read_xls(paste0(getwd(), '/data/TestExer2-GPA-round2.xls'))
```

Fitting a linear line

```
model1 <- lm(FGPA~SATV, data)
summary(model1)
```

```
##
## Call:
## lm(formula = FGPA ~ SATV, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.38333 -0.30694 -0.02763  0.32359  1.14037
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.44173    0.15506   15.75   <2e-16 ***
## SATV         0.06309    0.02766    2.28   0.0229 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4587 on 607 degrees of freedom
## Multiple R-squared:  0.008495,   Adjusted R-squared:  0.006861
## F-statistic: 5.201 on 1 and 607 DF,  p-value: 0.02293
```

Confidence Interval (95%) = intercept +/- (2*(SE))

Lower Limit = 0.0630 - 2*0.02766 = 0.0076

Upper Limit = 0.0630 + 2*0.02766 = 0.1183

```r
model2 <- lm(FGPA~SATV+SATM+FEM, data)

summary(model2)
```

```
##
## Call:
## lm(formula = FGPA ~ SATV + SATM + FEM, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.31351 -0.29883 -0.02146  0.29419  1.09966
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.55705    0.21610   7.205 1.73e-12 ***
## SATV         0.01416    0.02793   0.507    0.612
## SATM         0.17274    0.03193   5.410 9.07e-08 ***
## FEM          0.20027    0.03738   5.358 1.20e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4418 on 605 degrees of freedom
## Multiple R-squared:  0.08296,    Adjusted R-squared:  0.07842
## F-statistic: 18.24 on 3 and 605 DF,  p-value: 2.411e-11
```

Confidence Interval (95%) = intercept +/- (2*(SE))

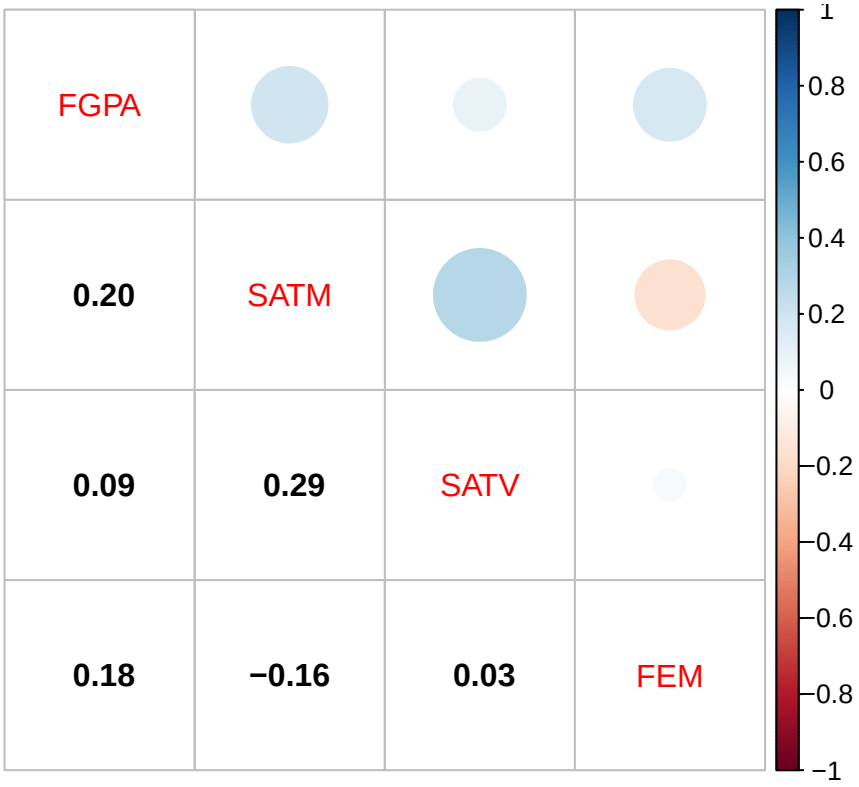
Lower Limit = 0.01416 - 2*0.02766 = -0.04116

Upper Limit = 0.01416 + 2*0.02766 = 0.06948

```r
dataCorr <- subset(data, select = c("FGPA","SATM","SATV","FEM"))

cor <-cor(dataCorr)

cor
```

```
##            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
```

```r
library(corrplot)

corrplot.mixed(cor, lower.col = 'black')
```




```
model3 <- lm(FGPA~SATM+FEM, data)

summary(model3)
```

```
##
## Call:
## lm(formula = FGPA ~ SATM + FEM, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29474 -0.29525 -0.01836  0.29468  1.10000
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.60515    0.19405   8.272 8.39e-16 ***
## SATM         0.17755    0.03046   5.828 9.10e-09 ***
## FEM          0.20188    0.03722   5.424 8.44e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4416 on 606 degrees of freedom
## Multiple R-squared:  0.08257,    Adjusted R-squared:  0.07955
## F-statistic: 27.27 on 2 and 606 DF,  p-value: 4.559e-12
```

```r
ssr <- sum((model3$residuals)^2)
ssr
```

```
## [1] 118.1512
```

```r
ssr <-  sum((model2$residuals)^2)

ssr
```

```
## [1] 118.101
```

```r
anova(model2, model3)
```

```
## Analysis of Variance Table
##
## Model 1: FGPA ~ SATV + SATM + FEM
## Model 2: FGPA ~ SATM + FEM
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    605 118.10
## 2    606 118.15 -1 -0.050199 0.2572 0.6123
```