

TEST2

2023-04-05

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##   filter, lag

## The following objects are masked from 'package:base':
##   intersect, setdiff, setequal, union

library(ggplot2)
TestExer2 <- read.delim("~/Desktop/Rdata/TestExer2.txt")
names(TestExer2)

## [1] "Observ." "FGPA"      "SATM"      "SATV"      "FEM"

View(TestExer2)
```

(a).

```
Regression <- lm(FGPA~SATV, data = TestExer2)
summary(Regression)

##
## Call:
## lm(formula = FGPA ~ SATV, data = TestExer2)
##
## 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
```

(i)

the coefficient of SATV = 0.063

its standard error = 0.027

p-value = 0.022

(ii)

Before SATV increased,

```
n <- 609
mu <- mean(TestExer2$FGPA)
sigma <- sqrt(0.15506^2/n)
interval <- mu + c(-1.96, 1.96) * sigma

interval

## [1] 2.780481 2.805112
```

After SATV increased,

```
TestExer2 <- TestExer2 %>%
  mutate(SATV2 = SATV + 1)
Regression <- lm(FGPA~SATV2, data = TestExer2)
summary(Regression)
```

```
##
## Call:
## lm(formula = FGPA ~ SATV2, data = TestExer2)
##
## 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)   2.37865    0.18256   13.03  <2e-16 ***
## SATV2         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
```

```
n <- 609
mu <- 2.792796
sigma <- sqrt(0.18256^2/n)
interval <- mu + c(-1.96, 1.96) * sigma

interval

## [1] 2.778296 2.807296
```

in SATV increased, the confidence interval is widen.

(b).

on SATM,

```
Regression <- lm(FGPA~SATM, data = TestExer2)
summary(Regression)
```

```
##
## Call:
## lm(formula = FGPA ~ SATM, data = TestExer2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.36083 -0.31550 -0.02403  0.32903  1.16111
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.85133    0.19303   9.591  < 2e-16 ***
## SATM          0.15067    0.03075   4.899 1.23e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4518 on 607 degrees of freedom
## Multiple R-squared:  0.03804,    Adjusted R-squared:  0.03646
## F-statistic: 24 on 1 and 607 DF,  p-value: 1.235e-06
```

```
n <- 609
mu <- 2.792796
sigma <- sqrt(0.19303^2/n)
interval <- mu + c(-1.96, 1.96) * sigma

interval

## [1] 2.777465 2.808127
```

```
TestExer2 <- TestExer2 %>%
  mutate(SATM2 = SATM + 1)
```

```
Regression <- lm(FGPA~SATM2, data = TestExer2)
summary(Regression)
```

```
##
## Call:
## lm(formula = FGPA ~ SATM2, data = TestExer2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.36083 -0.31550 -0.02403  0.32903  1.16111
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.70066    0.22366   7.604 1.10e-13 ***
## SATM2         0.15067    0.03075   4.899 1.23e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4518 on 607 degrees of freedom
## Multiple R-squared:  0.03804,    Adjusted R-squared:  0.03646
## F-statistic: 24 on 1 and 607 DF,  p-value: 1.235e-06
```

```
n <- 609
mu <- 2.792796
sigma <- sqrt(0.22366^2/n)
interval <- mu + c(-1.96, 1.96) * sigma

interval

## [1] 2.775032 2.810560
```

on FEM,

```
Regression <- lm(FGPA~FEM, data = TestExer2)
summary(Regression)
```

```
##
## Call:
## lm(formula = FGPA ~ FEM, data = TestExer2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22824 -0.30524 -0.02524  0.29176  1.21976
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.72824    0.02348 116.217  < 2e-16 ***
## FEM           0.16659    0.03771   4.418 1.18e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4534 on 607 degrees of freedom
## Multiple R-squared:  0.03115,    Adjusted R-squared:  0.02955
## F-statistic: 19.52 on 1 and 607 DF,  p-value: 1.182e-05
```

```
n <- 609
mu <- 2.792796
sigma <- sqrt(0.02348^2/n)
interval <- mu + c(-1.96, 1.96) * sigma

interval

## [1] 2.790931 2.794661
```

```
TestExer2 <- TestExer2 %>%
  mutate(FEM2 = FEM + 1)
```

```
Regression <- lm(FGPA~FEM2, data = TestExer2)
summary(Regression)
```

```
##
## Call:
## lm(formula = FGPA ~ FEM2, data = TestExer2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22824 -0.30524 -0.02524  0.29176  1.21976
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.56165    0.05546  46.192  < 2e-16 ***
## FEM2          0.16659    0.03771   4.418 1.18e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4534 on 607 degrees of freedom
## Multiple R-squared:  0.03115,    Adjusted R-squared:  0.02955
## F-statistic: 19.52 on 1 and 607 DF,  p-value: 1.182e-05
```

```
n <- 609
mu <- 2.792796
sigma <- sqrt(0.05546^2/n)
interval <- mu + c(-1.96, 1.96) * sigma

interval

## [1] 2.788391 2.797201
```

(c).

```
CM <- cor(TestExer2[2:5])
round(CM,3)
```

```
##          FGPA      SATM      SATV      FEM
## FGPA 1.000  0.195  0.092  0.176
## SATM 0.195  1.000  0.288  -0.163
## SATV 0.092  0.288  1.000  0.034
## FEM  0.176 -0.163  0.034  1.000
```

The reason for the differences between the outcomes in parts (a) and (b) is that the correlations are different.

Between FGPA and SATV (+0.092), There is only a small correlation. On the other hand, the correlations of both FGPA and SATM (+0.195), and the correlation of FGPA and the dummy variable FEM (+0.176) is higher.

(d).

```
Regression <- lm(SATV~FGPA, data = TestExer2)
summary(Regression)
```

```
##
## Call:
## lm(formula = SATV ~ FGPA, data = TestExer2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.40200 -0.48100 -0.02508  0.45275  2.01087
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.18880    0.16713   31.05  <2e-16 ***
## FGPA          0.13465    0.05905    2.28  0.0229 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6701 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
```

F-statistic = 5.201

t-value = 2.28

(t-value) 2 = 5.1984