Empirical Exercise - E6.2

Chi-Yuan Fang

2021-03-21

Using the data set **Growth** described in Empirical Exercise E4.1, but excluding the data for Malta, carry out the following exercises.

a. Construct a table that shows the sample mean, standard deviation, and minimum and maximum values for the series $Growth, TradeShare, YearsSchool, Oil, Rev\_Coups, Assassinations$, and $RGDP60$. Include the appropriate units for all entries.

Solution

```
library(ggplot2); library(dplyr);
library(jtools); library(ggstance);
```

```
##
## Attaching package: 'ggstance'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     geom_errorbarh, GeomErrorbarh
```

```
library(broom.mixed); library(huxtable)
```

```
## Warning in checkMatrixPackageVersion(): Package version inconsistency detected.
## TMB was built with Matrix version 1.3.2
## Current Matrix version is 1.2.18
## Please re-install 'TMB' from source using install.packages('TMB', type = 'source') or ask CRAN for a
```

```
## Registered S3 method overwritten by 'broom.mixed':
##   method      from
##   tidy.gamlss broom
```

```
##
## Attaching package: 'huxtable'
```

```
## The following object is masked from 'package:dplyr':
##
##     add_rownames
```

```
## The following object is masked from 'package:ggplot2':
##
##     theme_grey
```

```
# import data
library(readxl)
Growth <- read_xlsx("Growth/Growth.xlsx")

E62a <- function(x){
  # meam
  mu <- mean(x)
```

```
  # standard deviation
  SD <- sd(x)

  # minimum
  MIN <- min(x)

  # maximum
  MAX <- max(x)

  Table <- data.frame(mu, SD, MIN, MAX)

  colnames(Table) <- c("Mean", "Standard Deviation", "Minimum", "Maximum")

  Table

}

E62a_output <- as.list(matrix(ncol = 4))



for (i in 2:ncol(Growth)){
  E62a_output[[i]] <- E62a(Growth[[i]])

}

names(E62a_output) <- variable.names(Growth)

E62a_output
```

```
## $country_name
## [1] NA
##
## $growth
##       Mean Standard Deviation   Minimum  Maximum
## 1 1.942715            1.89712 -2.811944 7.156855
##
## $oil
##   Mean Standard Deviation Minimum Maximum
## 1    0                  0       0       0
##
## $rgdp60
##       Mean Standard Deviation  Minimum  Maximum
## 1 3103.785           2512.657 366.9999 9895.004
##
## $tradeshare
##       Mean Standard Deviation  Minimum  Maximum
## 1 0.564703          0.2892703 0.140502 1.992616
##
## $yearsschool
##       Mean Standard Deviation Minimum Maximum
## 1 3.985077              2.542     0.2   10.07
##
```

```
## $rev_coups
##        Mean Standard Deviation Minimum   Maximum
## 1 0.1674501          0.2246798       0 0.9703704
##
## $assasinations
##        Mean Standard Deviation Minimum  Maximum
## 1 0.2775641          0.4915284       0 2.466667
```

b. Run a regression of $Growth$ on $TradeShare, YearsSchool, Rev_Coups, Assassinations$, and $RGDP60$. What is the value of the coefficient on Rev_Coups? Interpret the value of this coefficient. Is it large or small in a real-world sense?

Solution

```
E62b <- lm(growth ~ tradeshare + yearsschool + rev_coups + assasinations + rgdp60, data = Growth)

summ(E62b)
```

```
## MODEL INFO:
## Observations: 65
## Dependent Variable: growth
## Type: OLS linear regression
##
## MODEL FIT:
## F(5,59) = 6.61, p = 0.00
## R² = 0.36
## Adj. R² = 0.30
##
## Standard errors: OLS
## ------------------------------------------------
##                        Est.   S.E.   t val.      p
## ------------------- ------- ------ -------- ------
## (Intercept)            0.49   0.69     0.71   0.48
## tradeshare             1.56   0.76     2.06   0.04
## yearsschool            0.57   0.14     4.13   0.00
## rev_coups             -2.16   1.11    -1.94   0.06
## assasinations          0.35   0.48     0.74   0.46
## rgdp60                -0.00   0.00    -3.17   0.00
## ------------------------------------------------
```

c. Use the regression to predict the average annual growth rate for a country that has average values for all regressors.

Solution

```
# Sample means cross the sample linear regression.
mean(Growth$growth)
```

```
## [1] 1.942715
```

```
# predict value
E62c <- function(x){
  E62b$coefficients %*% matrix(c(1, x), ncol = 1)

}

E62c_X <- c(mean(Growth$tradeshare),
            mean(Growth$yearsschool),
```

```
            mean(Growth$rev_coups),
            mean(Growth$assasinations),
            mean(Growth$rgdp60))

E62c(E62c_X)
```

```
##           [,1]
## [1,] 1.942715
```

d. Repeat (c), but now assume that the country's value for *TradeShare* is one standard deviation above the mean.

Solution

```
E62d_X <- c(mean(Growth$tradeshare) + sd(Growth$tradeshare),
            mean(Growth$yearsschool),
            mean(Growth$rev_coups),
            mean(Growth$assasinations),
            mean(Growth$rgdp60))

E62c(E62d_X)
```

```
##           [,1]
## [1,] 2.394468
```

e. Why is *Oil* omitted from the regression? What would happen if it were included?

Solution

The variable "oil" takes on the value of 0 for all 64 countries in the sample. This would generate perfect multicollinearity.

```
sum(Growth$oil == 1)
```

```
## [1] 0
```