Empirical Exercise - E4.1

Chi-Yuan Fang

2021-03-21

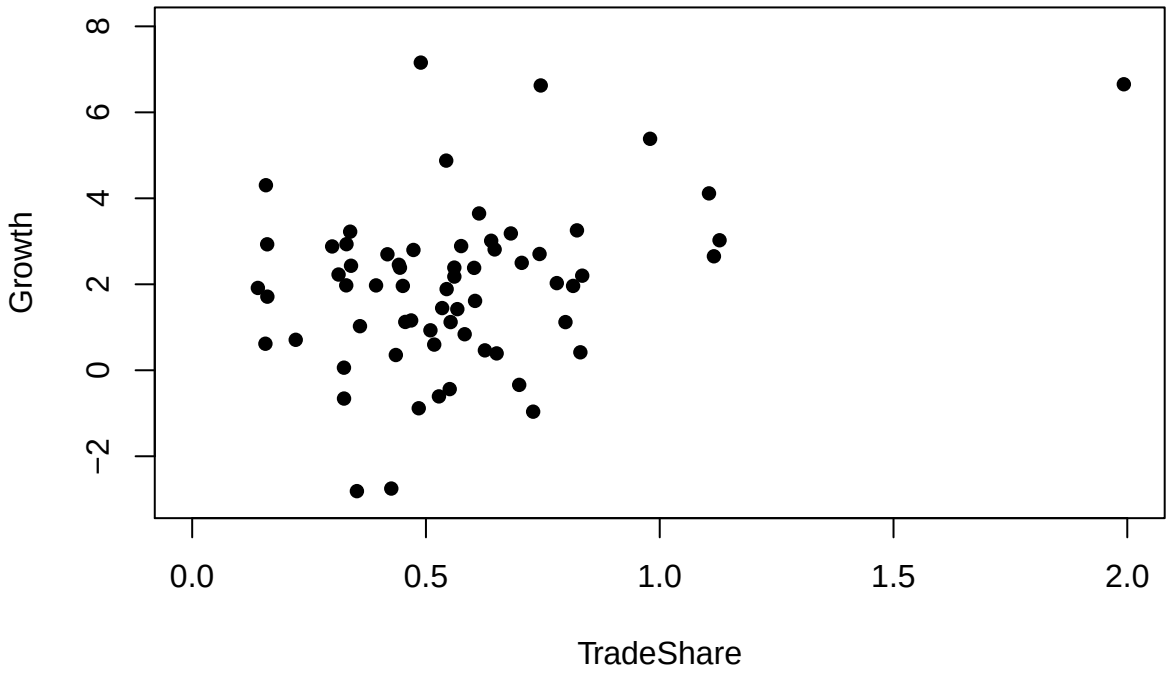
On the text website, http://www.pearsonglobaleditions.com, you will find the data file **Growth**, which contains data on average growth rates from 1960 through 1995 for 65 countries, along with variables that are potentially related to growth. A detailed description is given in **Growth_Description**, also available on the website. In this exercise, you will investigate the relationship between growth and trade.

a. Construct a scatterplot of average annual growth rate (*Growth*) on the average trade share (*TradeShare*). Does there appear to be a relationship between the variables?

Solution

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

plot(x = Growth$tradeshare,
     y = Growth$growth,
     pch = 16, # filled circle
     col = "black",
     xlim = c(0, 2),
     ylim = c(-3, 8),
     xlab = "TradeShare",
     ylab = "Growth",
     main = "E4.1 (a)")
```



b. One country, Malta, has a trade share much larger than the other countries. Find Malta on the scatterplot. Does Malta look like an outlier?

Solution

```
Growth[Growth$country_name =="Malta",]
```

```
## # A tibble: 1 x 8
##   country_name growth   oil rgdp60 tradeshare yearsschool rev_coups
##   <chr>         <dbl> <dbl>  <dbl>      <dbl>       <dbl>     <dbl>
## 1 Malta          6.65     0   1374       1.99        5.64         0
## # ... with 1 more variable: assasinations <dbl>
```

Malta is the "outlying" observation with a trade share of 1.99.

c. Using all observations, run a regression of *Growth* on *TradeShare*. What is the estimated slope? What is the estimated intercept? Use the regression to predict the growth rate for a country with a trade share of 0.5 and for another with a trade share equal to 1.0.

Solution

```
# regression
E41c <- lm(formula = growth ~ tradeshare, data = Growth)

# estimated intercept, estimated slope
summary(E41c)
```

```
##
## Call:
## lm(formula = growth ~ tradeshare, data = Growth)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -4.3739 -0.8864  0.2329  0.9248  5.3889
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.6403     0.4900   1.307  0.19606
## tradeshare    2.3064     0.7735   2.982  0.00407 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.79 on 63 degrees of freedom
## Multiple R-squared:  0.1237, Adjusted R-squared:  0.1098
## F-statistic: 8.892 on 1 and 63 DF,  p-value: 0.00407
```

```r
# predict value
E41c_predict <- function(x){
  E41c$coefficients %*% matrix(c(1, x), ncol = 1)

}

# predict value: tradeshare = 0.5
E41c_predict(0.5)
```

```
##          [,1]
## [1,] 1.793482
```

```r
# predict value: tradeshare = 1.0
E41c_predict(1)
```

```
##          [,1]
## [1,] 2.946699
```

d. Estimate the same regression, excluding the data from Malta. Answer the same questions in (c).

Solution

```r
# excluding the data from Malta
Growth_n <- Growth[Growth$country_name !="Malta",]

# regression
E41d <- lm(formula = growth ~ tradeshare, data = Growth_n)

# estimated intercept, estimated slope
summary(E41d)
```

```
##
## Call:
## lm(formula = growth ~ tradeshare, data = Growth_n)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4247 -0.9383  0.2091  0.9265  5.3776
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.9574     0.5804   1.650   0.1041
## tradeshare    1.6809     0.9874   1.702   0.0937 .
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.789 on 62 degrees of freedom
## Multiple R-squared:  0.04466,    Adjusted R-squared:  0.02925
## F-statistic: 2.898 on 1 and 62 DF,  p-value: 0.09369
```

```r
# predict value
E41d_predict <- function(x){
  E41d$coefficients %*% matrix(c(1, x), ncol = 1)

}

# predict value: tradeshare = 0.5
E41d_predict(0.5)
```

```
##          [,1]
## [1,] 1.797863
```

```r
# predict value: tradeshare = 1.0
E41d_predict(1)
```

```
##          [,1]
## [1,] 2.638315
```

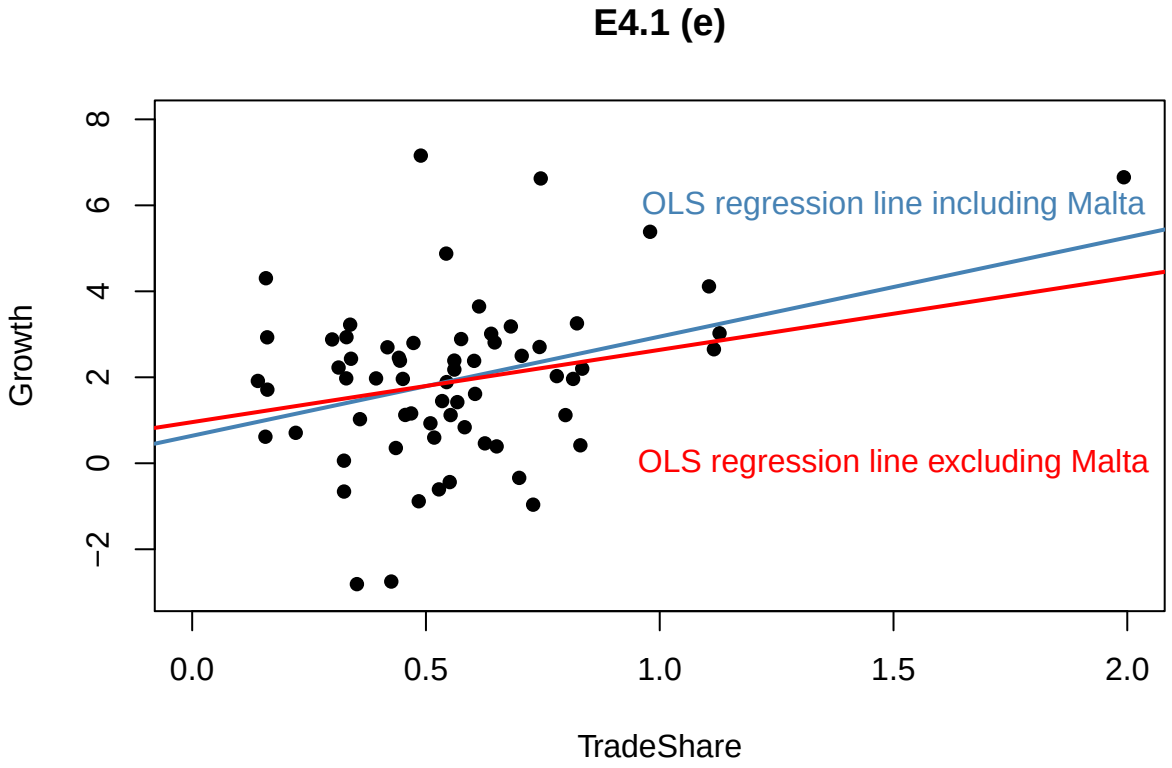
e. Plot the estimated regression functions from (c) and (d). Using the scatterplot in (a), explain why the regression function that includes Malta is steeper than the regression function that excludes Malta.

Solution

```r
plot(x = Growth$tradeshare,
     y = Growth$growth,
     pch = 16, # filled circle
     col = "black",
     xlim = c(0, 2),
     ylim = c(-3, 8),
     xlab = "TradeShare",
     ylab = "Growth",
     main = "E4.1 (e)")

# with Malta
abline(E41c, lwd = 2, col = "steelblue")
text(1.5, 6, "OLS regression line including Malta", col = "steelblue")

# without Malta
abline(E41d, lwd = 2, col = "red")
text(1.5, 0, "OLS regression line excluding Malta", col = "red")
```



f. Where is Malta? Why is the Malta trade share so large? Should Malta be included or excluded from the analysis?

Solution

Malta is an island nation in the Mediterranean Sea, south of Sicily.

Malta is a freight transport site, which explains its large "trade share." Many goods coming into Malta (imports into Malta) and are immediately transported to other countries (as exports from Malta).

Thus, Malta's imports and exports are unlike the imports and exports of most other countries. Malta should not be included in the analysis.