

**Problem set 2**  
**DUE: Mon. September 8, 2014**

**Exercise 1**

Compute the sample mean and standard deviation of growth and tradeshr.

**Solution**

$$E(\text{growth}) = 1.9427 \quad E(\text{tradeshr}) = 0.5647$$

$$\sigma(\text{growth}) = 1.37736 \quad \sigma(\text{tradeshr}) = 0.5378386$$

**Exercise 2**

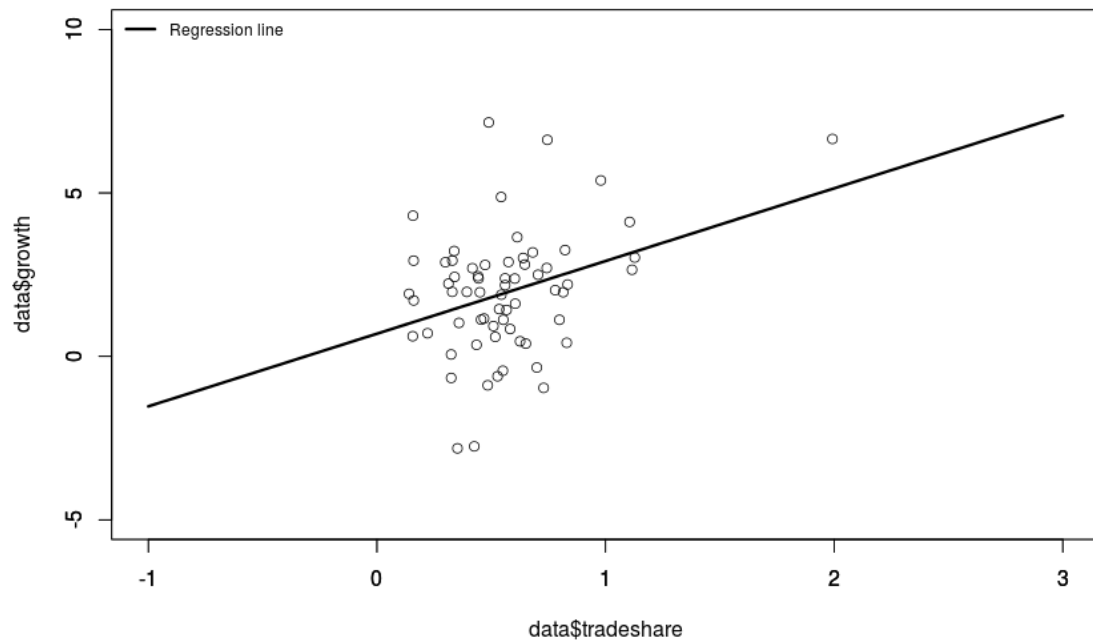
Estimate a regression of growth on tradeshr, using the “robust” option.

- a) What is the coefficient on tradeshr? Explain in words what it means. Is the numerical value of your estimate large or small in an economic (real-world) sense?
- b) Graph the data points and the estimated regression line.
- c) Is the slope coefficient statistically significantly different from zero at the 5% significance level? Show how you reach this conclusion.
- d) Report the 95% confidence interval for  $\beta_1$ , the slope of the population regression line.
- e) What is the  $R^2$  of this regression? What does this mean?
- f) Compute the correlation coefficient between growth and tradeshr, and compare its square to the  $R^2$ . How are the correlation coefficient and the  $R^2$  related?
- g) What is the value of the root mean squared error of the regression? What does this mean?
- h) Based on your graph from (b), does the regression error appear to be homoskedastic or heteroskedastic?
- i) Run the regression again without the “robust” option. Compare the results to what you obtained with the “robust” option. What is different?
- j) You should see an outlier in the data set. Rerun the regression (with the “robust” option), dropping the outlier. Does dropping the outlier make a qualitative difference to your results? Explain.

- k) What is the outlying observation? Considering the economics of the relation you are investigating, in your judgment should that outlier be omitted from the regression? (You might need to do a bit of research about that outlier to answer this question properly.)

### Solution

- a) The slope coefficient on tradeshr is 2.2233374. It means that if tradeshr will be changed by one, growth will be changed in 2.2233374. I personally think that such value of slope parameter is large in a real-world sense.



- b)

c) 
$$t = \frac{\hat{\beta} - \beta_{1.0}}{SE(\hat{\beta})}$$

$$\beta_{1.0} = 0$$

$$SE(\hat{\beta}) = \sqrt{\hat{\sigma}_{\hat{\beta}}^2}$$

$$\hat{\sigma}_{\hat{\beta}}^2 = \frac{1}{n} * \frac{\sum_{i=1}^n (X_i - \bar{X})^2 \hat{u}^2}{\left[ \frac{1}{n} * \sum_{i=1}^n (X_i - \bar{X})^2 \right]^2}$$

$$t = 3.221527 \quad t > 1.96, \text{ so we have to reject } H_0$$

d)  $[\hat{\beta}_1 - 1.96 * SE(\hat{\beta}_1), \hat{\beta}_1 + 1.96 * SE(\hat{\beta}_1)] = [0.870643, 3.576032]$

e)  $R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = 0.1149446$

It means that the regression on tradeshare predicts growth bad.

f)  $(\text{corr}(\text{growth}, \text{tradeshare}))^2 = 0.1236802$

As  $R^2$  it shows that the value of tradeshare is not very much correlated with growth, however we cannot say that this values are not correlated.

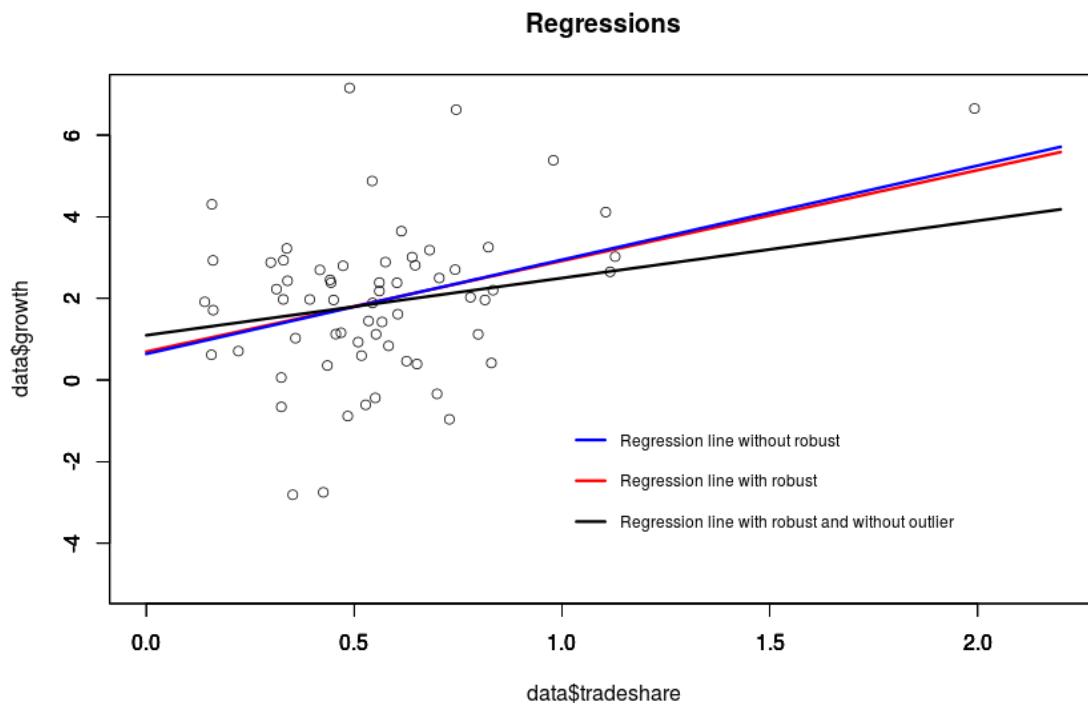
g)  $\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n \hat{u}_i^2}{n}}$

h) It seems to be heteroskedastic.

i) According to the graphic of regressions, regression without robustness has a little bit different coefficients.

j) According to the graphic of regressions, regression without the outlier has very different coefficients.

k) Outlying observation is Malta. We have to omit the outlier because linear regression is sensitive for it.



**Exercise 4**

Table 2 presents the results of three regressions, one in each column. Estimate the indicated regressions and fill in the values (you may either handwrite or type the entries in, at your convenience; if you choose to type up the table, an electronic copy of Table 2 in .doc format is available on the course Web site). For example, to fill in column (1), estimate the regression with growth as the dependent variable and tradeshr and school60 as the independent variables, using the “robust” option, and fill in the estimated coefficients and standard errors; also compute and fill in the value of the F-statistic and p-value testing the hypothesis that the coefficients on tradeshr and school60 are both zero. The adjusted  $R^2$  can be computed by rerunning this regression without the “robust” option. Note: the regressions in Table 2 all exclude the observation on Malta.

**Solution**

<b>Regressor</b>	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>
tradeshare	(1.5412032)	(1.3686513435)	(0.8505190649)
school60	(0.2105808)	(0.4730324670)	(0.4478457993)
capstock60	-	(-0.0003034976)	(-0.0003690296)
revc	-	-	(-2.7211122223)
civil	-	-	( 0.3973022112)
Intercept	0.0972978	0.1509167230	1.0709842944
<b>F-statistics testing the hypothesis that the population coefficients on the indicated regressors are all zero:</b>			
tradeshare, school60	(4.152)	(5.5003)	(4.6298)
tradeshare, school60, capstock60	-	(3.9728)	( 3.0852)
revc, civil	-	-	(5.1212)
<b>Regression summary statistics</b>			
$\overline{R^2}$	0.1331	0.2078	0.23
$R^2$	0.1606	0.2455	0.2911
Regression RMSE	1.65775	1.572138	1.534965
n	64	64	64