

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

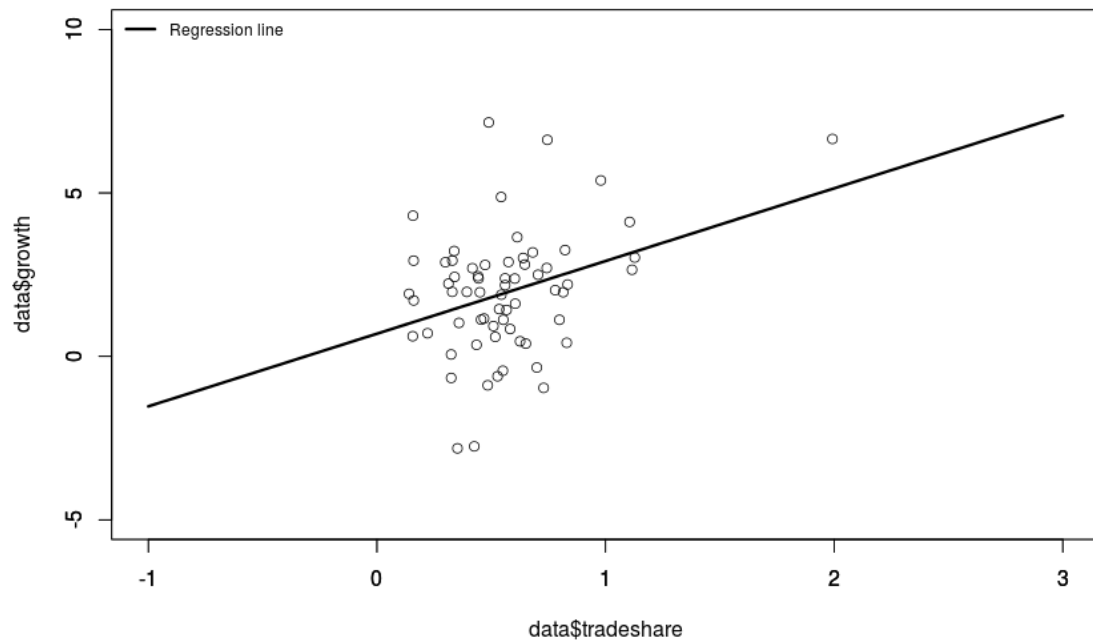
Estimate a regression of growth on tradeshr, using the “robust” option.

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

$$\beta_{1.0} = 0$$

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

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

$$t = 3.2723 \quad t > 1.96, \text{ so we have to reject } H_0$$

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) $(corr(growth, tradeshare))^2 = 0.1236802$

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.

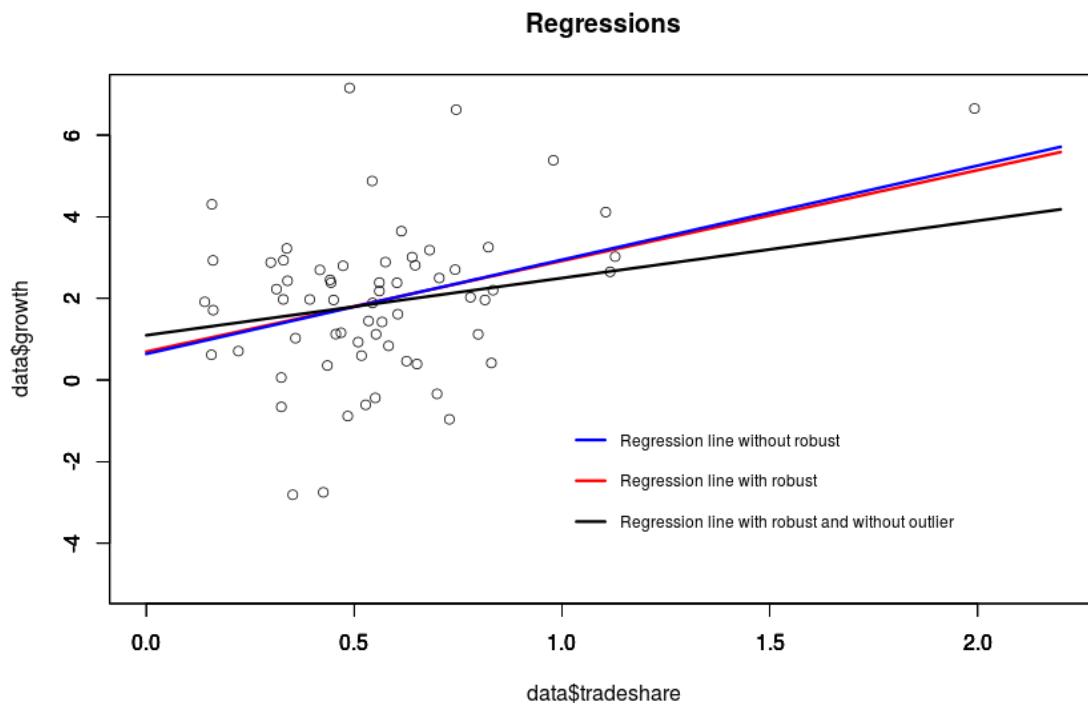
g) $RMSE = \sqrt{\frac{\sum_{i=1}^n \hat{u}_i^2}{n}} = 1.7624$

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 3

Construct a new variable, `lorgdp60`, which equals one if the country's GDP is in the bottom quartile of GDP for 1960 and equals zero otherwise. Estimate a regression of growth on `lorgdp60`, using the “robust” option.

- a) What is the coefficient on `lorgdp60`? Explain in words what this means. Is the numerical value of your estimate large or small in an economic (real-world) sense?
- b) Test the hypothesis that the mean growth rate from 1960-1995 is the same for countries with `lorgdp60` = 1 as it is for countries with `lorgdp60` = 0, against the alternative that they differ, at the 5% significance level.
- c) Using the “summarize” command, compute the sample average of growth for countries with `lorgdp60` = 1 and then again for countries with `lorgdp60` = 0; from this compute the difference in mean GDP growth rates for the two groups and construct the differences-of-means t-statistic testing the hypothesis that the mean growth rates are the same.
- d) Reestimate the regression of growth on `lorgdp60`, without the “robust” option. How does the t-statistic computed in (c) compare to the t-statistic on the slope coefficient in the regression of growth on `lorgdp60` obtained with and without the “robust” option? Explain.

Solution

- a) $\hat{\beta}_1 = -1.640866$

It means that when `lorgdp60` = 1 growth decreases by 1.64. In economic sense the value of the coefficient can be large.

- b) $t = \frac{\bar{Y}_m - \bar{Y}_n - d_0}{SE(\bar{Y}_m - \bar{Y}_n)} = 2.284999$

So we have to reject the idea that the means are equal.

- c) Due to the fact that we didn't use stata this task is equal to previous one.
- d) The differences are in counting SE. When regression is robust we have to estimate the error u_i

$$SE(\hat{\beta}_1) = \sqrt{\hat{\sigma}_{\hat{\beta}_1}^2}$$

in robust case:

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

in not robust case:

$$\tilde{\sigma}_{\hat{\beta}}^2 = \frac{s_{\hat{u}}^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

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 you 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

Regressor	(1)	(2)	(3)
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
F-statistics testing the hypothesis that the population coefficients on the indicated regressors are all zero:			
tradeshare, school60	(4.152)	(5.5003)	(4.6298)
tradeshare, school60, capstock60	-	(3.9728)	(3.0852)
revc, civil	-	-	(5.1212)
Regression summary statistics			
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

Exercise 5

Use Table 2 to answer the following questions.

- a) Write the regression in column (1) in “equation form,” with the standard error below the respective regression coefficient.
- b) Explain in words what the coefficient on *school60* means in regression (1).
- c) Using regression (1), test the hypothesis that the coefficient on *tradeshare* is zero, against the alternative that it is nonzero, at the 5% significance level. In everyday words (not statistical terms), what precisely is the hypothesis that you are testing?
- d) Does the coefficient on *tradeshare* differ in regressions (1), (2), and (3) in a substantively important way, that is, is the difference between the three estimates large in a real-world sense?
- e) Economic theory predicts that *tradeshare*, *school60*, and *capstock60* all are determinants of economic growth. Use regression (2) to test the hypothesis (at the 5% significance level) that the coefficients on these three economic variables are all zero, against the alternative that at least one coefficient is nonzero.
- f) In regression (3), is the coefficient on *revc* statistically significant at the 5% significance level? Is the coefficient on *civil* statistically significant at the 5% significance level?
- g) Use the heteroskedasticity-robust F-statistic to test the hypothesis (at the 1% significance level) that the coefficients on the political variables (*revc* and *civil*) in regression (3) are both zero, against the alternative that one or the other or both are nonzero. Discuss in light of your answer to part (f).
- h) The neoclassical theory of human capital in economics suggests that countries with more human capital – that is, a better educated work force – will have a higher rate of productivity and therefore have a higher growth rate. Is this prediction borne out in the regression results? Explain.

Solution

- a)

0.0972978	+	1.5412032	<i>tradeshare</i>	+	1.5412032	<i>school60</i>
(0.71071231)			(0.91985029)			(0.07660823)
- b) The *school60* means that it has increased/decreased by one the growth increased by 0.07660823.

c) $t = \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} = \frac{1.5412032}{0.91985029} = 1.667301$

$|t| < 1.96$ -the hypothesis is accepted in 5% significance level.

d) decrease importance of regressors during increasing of number of regressors

e) Lets calculate F-test for this hypothesis from R

`regr <- lmRob(newData$growth newData$tradeshare+newData$yearsschool+newData$rgdp60) anova(regr, lmRob(newData$growth 1))`

The F-test is:

$$F = 3.9728$$

$F > 2.6$ - hence, hypothesis is rejected in 5% significance level.

f) Lets calculate the standards error for equation

$$\begin{array}{cccccc} 1.071 & + & 0.851 & + & 0.448 & - & 0.0004 & - & 2.721 & + & 0.397 \\ (0.6664) & & (0.8171) & & (0.1218) & & (0.0001) & & (0.952) & & (0.4153) \end{array}$$

And calculate t-test

For coefficient on revc

$$t = \frac{\hat{\beta}_{revc} - \beta_{revc}}{SE(\hat{\beta}_{revc})} = \frac{2.721}{0.952} = 2.858193$$

$|t| > 1.96$ -the hypothesis is rejected in 5% significance level.

$$\text{For coefficient on assassinations } t = \frac{\hat{\beta}_{as} - \beta_{as}}{SE(\hat{\beta}_{as})} = \frac{0.397}{0.4153} = 0.9559355$$

$|t| < 1.96$ -the hypothesis is accepted in 5% significance level.

g) Lets calculate robust F-test in R

`anova(lmRob(newData$growth newData$tradeshare+newData$yearsschool+newData$rgdp60+newData$revcoups+newData$assassinations), lmRob(newData$growth newData$yearsschool + newData$rgdp60))`

$$F = 5.1212$$

$|F| > 4.61$ - the hypothesis is rejected in 1% significance level.

h) According to the coefficients, yearsschool doesn't have the biggest influence on growth. So the theory fails.