Quantitative methods for economics, finance and management

2019/2020

LIUC
Università Cattaneo

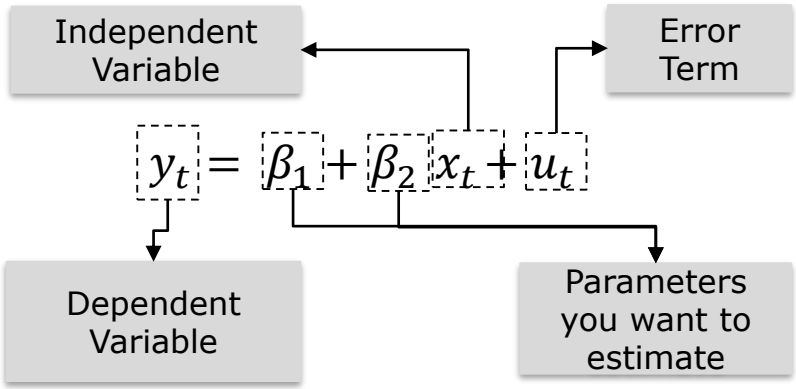
LAB 4

14.10.2019

LIUC
Università Cattaneo

Lab-4: Simple Linear Regression Model (SLRM)

- Given a the simple Linear Regression Model:



- The unknown parameters can be estimated by using OLS estimation as follows:

$$\widehat{\beta_1} = \bar{y} - \widehat{\beta_2}\,\bar{x}$$

$$\widehat{\beta_2} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i}^{N}(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}} = \frac{Cov(x,y)}{Var\,(x)}$$

Lab-4: Coefficient of determination R^2

$$R^2 = \frac{ESS}{TSS}$$

R^2 REPRESENTS THE % VARIATION OF THE DEPENDENT VARIABLE (y) EXPLAINED BY THE VARIATION OF THE FITTED VALUE $\hat{y}$

$$TSS = ESS + RSS$$

THIS EQUALITY HOLDS IF THE LRM INCLUDES THE INTERCEPT

LIUC Università Cattaneo

Lab-4: Coefficient of determination R^2

$$TSS = \sum_{t=1}^{T} (y_t - \bar{y})^2$$

$$ESS = \sum_{t=1}^{T} (\hat{y}_t - \bar{y})^2$$

$$RSS = \sum_{t=1}^{T} \hat{u}_t^2$$

WHERE $\bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t$; GIVEN: $y_t = \beta_1 + \beta_2 X_t + u_t \quad t = 1, \ldots T$

AND $\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 X_t$

$$\hat{u}_t = y_t - \hat{y}_t = y_t - (\hat{\beta}_1 + \hat{\beta}_2 X_t)$$

LIUC Università Cattaneo

Lab-4: Simple Linear Regression Model (SLRM)

CONSIDERING THE FOLLOWING SLRM

$$g = \beta_1 + \beta_2 p + u$$

WHERE:

PRICE IS A (N,1) VECTOR OF PRICE OF GASOLINE

g IS A (N,1) " " " GASOLINE PRODUCTION

u " " " NOISE OR ERROR TERMS

- ESTIMATE β_1 AND β_2 → LS (COMMAND) / → FORMULA

LIUC Università Cattaneo

Lab-4: Simple Linear Regression Model (SLRM)

Dependent Variable: G → PRODUCTION OF GASOLINE MILLION BARRELS PER DAY (MM bbl)
Method: Least Squares
Date: 10/14/19 Time: 18:17
Sample: 1960 1995
Included observations: 36

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	154.0304	11.61636	13.25978	0.0000
PG (PRICE)	31.10752	4.425480	7.029185	0.0000

R-squared	0.592372	Mean dependent var	226.0944
Adjusted R-squared	0.580383	S.D. dependent var	50.59182
S.E. of regression	32.77228	Akaike info criterion	9.870996
Sum squared resid	36516.77	Schwarz criterion	9.958969
Log likelihood	-175.6779	Hannan-Quinn criter.	9.901701
F-statistic	49.40944	Durbin-Watson stat	0.255613
Prob(F-statistic)	0.000000		

IF PRICE OF GASOLINE RISES BY 1 $
THEN PRODUCTION OF GASOLINE RISES BY 31 MM bbl

LIUC Università Cattaneo

Lab-4: Simple Linear Regression Model (SLRM) (1/2)

Program: QMEFM_LAB4_141019 - (c:\users\asseg1\desktop\didattica...

Run | Print | Save | SaveAs | Cut | Copy | Paste | InsertTxt | Find | Replace | Wrap+/- | LineNum+/-

```
'gasoline dataset'
'g is the production of gasoline'
'pg is the price of gasoline'
'Estimate gasoline supply equation by using OLS approach with command ls'
ls g c pg
'Estimate gasoline supply equation by using OLS approach with formula'
'generate a  new vector z1 equals 1 (constant)'
genr z1 = 1
'generate a new vector z2 equals price of gasoline (explanatory variable)'
genr z2 = pg
'generate a new vector y equals consumption (dependent variable)'
genr yy = g
'compute the sample average of production of gasoline'
scalar avgyy = @mean(yy)
'compute the sample average of  price of gasoline'
scalar avgz2 = @mean(z2)
'compute production of gasoline in deviation from its mean '
genr yydevavgyy = yy - avgyy'
'compute price of gasoline in deviation from its mean '
genr z2devavgz2 = z2 - avgz2
'compute the square of production of gasoline in deviation from the mean '
genr z2devavgz2sq = z2devavgz2^2
'define new variable equals the number of observations'
scalar nobs = @obs(yy)
'define k=2=number of explanatory variables (constant+ income)'
scalar k = 2
genr z2devyydev= @sum(z2devavgz2*yydevavgyy)
scalar z2dev_yydev = z2devyydev(1)
```


LIUC Università Cattaneo

Lab-4: Simple Linear Regression Model (SLRM) (2/2)

Program: QMEFM_LAB4_141019 - (c:\users\asseg1\desktop\didattica...

Run | Print | Save | SaveAs | Cut | Copy | Paste | InsertTxt | Find | Replace | Wrap+/- | LineNum+/-

```
'define the sample covariance between x and y'
scalar covz2yy =  z2dev_yydev/(nobs-1)
genr z2devAvgz_sq= @sum(z2devavgz2^2)
'define the sample variance of x '
scalar varz2 =  z2devAvgz_sq(1) /(nobs-1)
'compute the estimated coefficients'
scalar beta2hat = covz2yy/varz2
scalar beta1hat = avgyy- beta2hat*avgz2
show yy
'compute the fitted values'
genr yyhat = beta1hat*z1 + beta2hat*z2
'compute the residuals'
genr res = yy-yyhat
show res
'actual values  = fitted values +residuals'
genr y1 = yyhat+res
'fitted values  = actual values +residuals'
genr yyhat2 = yy+res
```


LIUC Università Cattaneo

Lab-4: Simple Linear Regression Model (SLRM)

- Considering the following simple regression model:

$$W = \beta_1 + \beta_2 Ed + u$$

where:

- W is a vector of **hourly wages**
- Ed is a vector of **year of schooling**.
- u is a vector of noise ("classical" error terms)
- **You are requested to carry out the following duties:**

1. Perform the regression equation to get the estimates of β_1 and β_2
2. What is the interpretation of β_2? Explain.
3. Find the vector of the fitted values $\widehat{price}$
4. Find the vector of residual $\hat{u}$
5. Compute:
   a) Total Sum of Square (TSS)
   b) Explained sum of square (ESS)
   c) Coefficient of determination R^2

LIUC Università Cattaneo

Lab-4: Education dataset - SLRM

Equation: EQ01 Workfile: QMEFM_LAB4_141019::education\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: WAGE
Method: Least Squares
Date: 10/14/19 Time: 18:30
Sample: 1 4733
Included observations: 4733

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5.202605	0.465486	-11.17672	0.0000
EDUC	1.156924	0.034457	33.57628	0.0000

R-squared	0.192437	Mean dependent var	10.18720
Adjusted R-squared	0.192266	S.D. dependent var	6.213761
S.E. of regression	5.584556	Akaike info criterion	6.278309
Sum squared resid	147546.9	Schwarz criterion	6.281040
Log likelihood	-14855.62	Hannan-Quinn criter.	6.279269
F-statistic	1127.367	Durbin-Watson stat	0.358700
Prob(F-statistic)	0.000000		

AN ADDITIONAL YEAR OF EDUCATION CORRESPOND TO AN INCREASE IN HOURLY WAGES OF 1,15 $
(LEVEL - LEVEL SPECIFICATION)

LIUC Università Cattaneo

Lab-4: Education dataset - SLRM (1/3)

Program: QMEFM_LAB4_141019 - (c:\users\asseg1\desktop\didattica...

Run | Print | Save | SaveAs | Cut | Copy | Paste | InsertTxt | Find | Replace | Wrap+/- | LineNum+/-

```
'education dataset'
'Estimate wage-education equation by using OLS approach with command ls'
ls wage c educ
'rename dependent and covariates'
genr x1 = 1
genr x2 = educ
genr y = wage
'Estimate wage-education equation by using OLS approach with command ls'
ls y x1 x2
'store estimated coefficients'
scalar beta1hat = c(1)
scalar beta2hat = c(2)
'compute the fitted values'
genr yhat = beta1hat*x1 + beta2hat*x2
'compute the residuals'
genr res = y-yhat
show res
'show that the sum of fitted values and residuals equals the actual values'
genr y1 = yhat+res
'compute the average of the residuals'
scalar avgres = @mean(res)
'compute the average of the fitted values'
scalar avgyhat = @mean(yhat)
'compute the average of the actual values'
scalar avgy = @mean(y)
'show that the avgy coincedes with avgyhat'
show avgyhat
show avgy
```


LIUC Università Cattaneo

Lab-4: Education dataset - SLRM (2/3)

Program: QMEFM_LAB4_141019 - (c:\users\asseg1\desktop\didattica...

Run | Print | Save | SaveAs | Cut | Copy | Paste | InsertTxt | Find | Replace | Wrap+/- | LineNum+/-

```
'compute the total sum of squares (TSS)'
genr ss = (y - avgy)^2
scalar tss = @sum(ss)
'compute the explained  sum of squares (ESS)'
genr es = (yhat - avgy)^2
scalar ess = @sum(es)
'compute the residual sum of squares (RSS)'
genr rs =  (res- avgres)^2
scalar  rss = @sum(rs)
'compute the TSS by exploiting the relationship between ESS and RSS'
scalar tss1 = ess +rss
'show that tss (derived by using the previous formula) = tss1'
show tss
show tss1
'compute the total sum of squares as follows'
scalar sdy =  @stdev(y)
scalar tobs = @obs(y)
scalar tss2 = ((sdy)^2)*(tobs-1)
'show that tss2 (derived by using the previous formula) = tss = tss1'
show tss1
show tss2
show tss
'compute the coefficient of determination in 2 different ways'
scalar Rsq1 = ess/tss
scalar Rsq2 = 1-rss/tss
'show that tss2 (derived by using the previous formula) = tss = tss1'
show Rsq1
show Rsq2
```


LIUC Università Cattaneo

Lab-4: Education dataset - SLRM (3/3)

```
'show that 1 = ess/tss + rss/tss'
scalar one = ess/tss + rss/tss
'find coeffcient of determination  equals zero'
ls y c
smpl 10 11
'find coeffcient of determination  equals 1 (deterministic model)'
ls y x1 x2
expand 1 4733
```