

Econometrics Analysis

Setup Code

```
# Tải các packages
install.packages('readxl')
library('readxl')
install.packages('olsrr')
library('olsrr')
# Import data
data <- read_excel(file.choose())
# Chuyển về Numeric
full <- mutate_all(data, function(x)
as.numeric(as.character(x)))
```

```
> # Tải các packages
> library('readxl')
Warning message:
package 'readxl' was built under R version 4.2.3
> # Import data
> data <- read_excel(file.choose())
> # Chuyển về Numeric
> library('olsrr')
```

Attaching package: 'olsrr'

The following object is masked from 'package:datasets':

 rivers

```
Warning message:
package 'olsrr' was built under R version 4.2.3
>
> full <- mutate_all(data, function(x)
as.numeric(as.character(x)))
Error in mutate_all(data, function(x)
as.numeric(as.character(x))) :
  could not find function "mutate_all"
```

Câu 1

1) Hồi quy Số lượng quần áo theo Phụ cấp

```
> # 1. Hồi quy Số lượng quần áo theo Phụ cấp
> cau1 <- lm(slquanao~phucap,data=full)
> summary(cau1)
Call:
lm(formula = slquanao ~ phucap, data = full)

Residuals:
    Min       1Q   Median       3Q      Max
-13.697   -9.002   -5.002    4.303   86.040

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.396e+01  2.330e+00   5.991  1.13e-08 ***
phucap        3.474e-04  5.740e-04   0.605    0.546
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.54 on 178 degrees of freedom
Multiple R-squared:  0.002054,    Adjusted R-squared:  -
0.003553
F-statistic: 0.3663 on 1 and 178 DF,  p-value: 0.5458
```

Nếu phụ cấp tăng 1 (nghìn đồng) thì số lượng quần áo mua (cái/năm) tăng $3.474e-04$.

2) Hồi quy Số lượng quần áo theo Thu nhập

```
> cau2 <- lm(slquanao~thunhap,data=full)
> summary(cau2)
```

Call:

```
lm(formula = slquanao ~ thunhap, data = full)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-33.379	-8.395	-4.096	4.627	85.053

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	1.410e+01	1.290e+00	10.929
thunhap	8.513e-04	4.593e-04	1.854

Pr(>|t|)

(Intercept)	<2e-16 ***
thunhap	0.0654 .

Signif. codes:

0	'****'	0.001	'***'	0.01	'**'	0.05	'.'
0.1	' '	1					

Residual standard error: 15.41 on 178 degrees of freedom

Multiple R-squared: 0.01894, Adjusted R-squared: 0.01343

F-statistic: 3.436 on 1 and 178 DF, p-value: 0.06544

3) Hồi quy Số lượng quần áo theo Đi chơi

```
> cau3 <- lm(slquanao~dichoi,data=full)
> summary(cau3)
```

Call:

```
lm(formula = slquanao ~ dichoi, data = full)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-37.067	-6.786	-2.797	2.149	90.236

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	7.7427	1.5546	4.980
dichoi	1.0108	0.1565	6.459

Pr(>|t|)

(Intercept)	1.49e-06	***
dichoi	9.75e-10	***

Signif. codes:

0	'***'	0.001	'**'	0.01	'*'	0.05	'.'
0.1	' '	1					

Residual standard error: 14 on 178 degrees of freedom

Multiple R-squared: 0.1899, Adjusted R-squared: 0.1853

F-statistic: 41.71 on 1 and 178 DF, p-value: 9.751e-10

4) Hồi quy Số lượng quần áo theo Đi học

```
> cau4 <- lm(slquanao~dihoc,data=full)
> summary(cau4)
```

Call:

```
lm(formula = slquanao ~ dihoc, data = full)
```

Residuals:

Min	1Q	Median	3Q	Max
-14.740	-8.490	-5.708	4.268	86.970

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	17.7728	3.2049	5.545
dihoc	-0.6775	0.7819	-0.866

Pr(>|t|)

(Intercept) 1.05e-07 ***

dihoc 0.387

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1

Residual standard error: 15.53 on 178 degrees of freedom

Multiple R-squared: 0.0042, Adjusted R-squared: -0.001395

F-statistic: 0.7507 on 1 and 178 DF, p-value: 0.3874

5) Hồi quy Số lượng quần áo theo Săn lòng

```
> cau5 <- lm(slquanao~sanlong,data=full)
> summary(cau5)
```

Call:

lm(formula = slquanao ~ sanlong, data = full)

Residuals:

Min	1Q	Median	3Q	Max
-29.809	-8.638	-4.594	5.005	85.362

```

Coefficients:
              Estimate Std. Error t value
(Intercept) 1.410e+01  1.206e+00  11.693
sanlong      1.785e-03  6.649e-04   2.685
              Pr(>|t|)
(Intercept) < 2e-16 ***
sanlong      0.00794 **
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
  0.1 ' ' 1

Residual standard error: 15.25 on 178 degrees of freedom
Multiple R-squared:  0.03892, Adjusted R-squared:  0.03352
F-statistic: 7.209 on 1 and 178 DF,  p-value: 0.007941

```

Chọn ra mô hình có R2 lớn nhất

Tạo bảng gồm các giá trị của các mô hình trên

```

> # 6. Tính và chọn mô hình có R^2 lớn nhất
> # Gán biến cho R^2
> r2_cau1 = summary(cau1)$r.squared
> r2_cau2 = summary(cau2)$r.squared
> r2_cau3 = summary(cau3)$r.squared
> r2_cau4 = summary(cau4)$r.squared
> r2_cau5 = summary(cau5)$r.squared
> # Tạo bảng
> table6 <-
matrix(c(r2_cau1,r2_cau2,r2_cau3,r2_cau4,r2_cau5),ncol=1,byrow=
TRUE)
> colnames(table6) <- c("R Squared")
> rownames(table6) <- c("Câu 1","Câu 2","Câu 3","Câu 4","Câu
5")
> table6
              R Squared
Câu 1 0.002053898
Câu 2 0.018938372
Câu 3 0.189858770
Câu 4 0.004199629
Câu 5 0.038922096

```

Chọn ra mô hình có R2 lớn nhất

```
> max(table6)
[1] 0.1898588
```

Mô hình có R2 lớn nhất là mô hình Hồi quy Số lượng quần áo theo Đi chơi

6) Lấy các biến ở câu 1), chạy các dạng:

Hàm hồi quy Full Chung

```
> # Mô hình Full Chung
> full_chung <- lm(slquanao ~ ., data = full)
> summary(full_chung)
```

Call:

```
lm(formula = slquanao ~ ., data = full)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-27.846	-6.137	-2.280	3.396	84.963

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	8.5942328	8.1999506	1.048	0.2961	
nam	-5.3044700	2.3060494	-2.300	0.0227	*
phucap	-0.0007327	0.0005689	-1.288	0.1996	
thunhap	0.0003083	0.0004583	0.673	0.5022	
dihoc	-0.0343286	0.7220472	-0.048	0.9621	
dichoi	0.9723914	0.1683232	5.777	3.73e-08	***
sanlong	0.0007741	0.0006656	1.163	0.2465	
kaki	2.1785735	4.0642855	0.536	0.5927	
jeans	7.3782358	3.7699965	1.957	0.0520	.
cotton	6.0427203	3.3266293	1.816	0.0711	.

thuongmai	0.8273485	4.4862734	0.184	0.8539
mxh	-3.8618894	3.3089262	-1.167	0.2449
ngoaiduong	-1.4594681	7.0999222	-0.206	0.8374
nguoithan	-4.5620335	5.0988137	-0.895	0.3722
tot	3.8346069	6.0168113	0.637	0.5248
binhthuong	-0.4384577	5.7344886	-0.076	0.9391

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.66 on 164 degrees of freedom

Multiple R-squared: 0.29, Adjusted R-squared: 0.225

F-statistic: 4.465 on 15 and 164 DF, p-value: 4.92e-07

Tuyến tính | $Y = \beta_0 + \beta X$

```
> tuyentinh4 <- lm(slquanao~dichoi, data = full)
> summary(tuyentinh4)
```

Call:

lm(formula = slquanao ~ dichoi, data = full)

Residuals:

Min	1Q	Median	3Q	Max
-37.067	-6.786	-2.797	2.149	90.236

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.7427	1.5546	4.980	1.49e-06 ***
dichoi	1.0108	0.1565	6.459	9.75e-10 ***

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14 on 178 degrees of freedom

Multiple R-squared: 0.1899, Adjusted R-squared: 0.1853

F-statistic: 41.71 on 1 and 178 DF, p-value: 9.751e-10

a. Ý nghĩa hệ số góc

Tuyến tính - Logarit | $Y = \beta_0 + \beta_1 \ln(X)$

```
> linlog4 <- lm(slquanao ~
log(dichoi), data=full, dichoi>0)
> summary(linlog4)
```

Call:
lm(formula = slquanao ~ log(dichoi), data = full, subset =
dichoi >
0)

Residuals:

	Min	1Q	Median	3Q	Max
	-25.509	-7.952	-3.370	3.828	91.150

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.330	2.440	1.774	0.0778 .
log(dichoi)	6.521	1.296	5.031	1.21e-06 ***

Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.54 on 173 degrees of freedom
Multiple R-squared: 0.1276, Adjusted R-squared: 0.1226
F-statistic: 25.31 on 1 and 173 DF, p-value: 1.215e-06

a. Ý nghĩa hệ số góc

Hàm hồi quy qua gốc tọa độ

Phạm Trí Cao - Khoa Toán - Thống kê đang trình bày

MÔ HÌNH HỒI QUY ĐƠN

Không qua gốc tọa độ

```

Call:
lm(formula = wage ~ brthord)

Coefficients:
(Intercept) 1054.736      23.876  44.176 < 2e-16 ***
brthord      -36.822      8.589  -4.287 2.02e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 399.8 on 850 degrees of freedom
(83 observations deleted due to missingness)
Multiple R-squared:  0.02117    Adjusted R-squared:  0.02001

```

Qua gốc tọa độ

```

Call:
lm(formula = wage ~ 0 + brthord)

Coefficients:
brthord 273.960      8.559  30.65 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 745.4 on 851 degrees of freedom
(83 observations deleted due to missingness)
Multiple R-squared:  0.5246    Adjusted R-squared:  0.5241
( R^2 = -2.2262 và R^2 hiệu chỉnh = -2.2262 )

```

52

08:24 | sxq-usgq-nyo

```

> origin4 <- lm(slquanao ~ 0 + dichoi,data=full)
> summary(origin4)

```

Call:

```
lm(formula = slquanao ~ 0 + dichoi, data = full)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-46.653	-2.045	2.029	6.882	96.823

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
dichoi	1.5884	0.1119	14.2	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.91 on 179 degrees of freedom
Multiple R-squared: 0.5298, Adjusted R-squared: 0.5271
F-statistic: 201.7 on 1 and 179 DF, p-value: < 2.2e-16

a. Ý nghĩa hệ số góc

Hàm hồi quy dạng hàm bậc 2

```

> power4 <- lm(slquanao ~ dichoi + I(dichoi^2),data=full)
> summary(power4)

Call:
lm(formula = slquanao ~ dichoi + I(dichoi^2), data = full)

Residuals:
    Min       1Q   Median       3Q      Max
-41.994  -6.376  -2.815   3.255  89.192

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.90365     2.26475   4.373 2.09e-05 ***
dichoi         0.40564     0.48771   0.832   0.407
I(dichoi^2)    0.02325     0.01775   1.310   0.192
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.98 on 177 degrees of freedom
Multiple R-squared:  0.1976, Adjusted R-squared:  0.1886
F-statistic: 21.8 on 2 and 177 DF, p-value: 3.442e-09

```

a. Ý nghĩa hệ số góc

b. Chọn mô hình

Lập bảng so sánh các chỉ số của các mô hình.

```

> # Tuyến tính
> r2_tuyentinh4 = summary(tuyentinh4)$r.squared
> aic_tuyentinh4 = AIC(tuyentinh4)
> bic_tuyentinh4 = BIC(tuyentinh4)
> tuyentinh4_cp=ols_mallows_cp(tuyentinh4,
full_chung)
>
> # Lin - Log
> r2_linlog4 = summary(linlog4)$r.squared
> aic_linlog4 = AIC(linlog4)
> bic_linlog4 = BIC(linlog4)
> linlog4_cp=ols_mallows_cp(linlog4, full_chung)

```

```

>
> # Qua Gốc Tọa Độ
> r2_origin4 = summary(origin4)$r.squared
> aic_origin4 = AIC(origin4)
> bic_origin4 = BIC(origin4)
> origin4_cp=ols_mallows_cp(origin4, full_chung)
>
> # Bậc Hai
> r2_power4 = summary(power4)$r.squared
> aic_power4 = AIC(power4)
> bic_power4 = BIC(power4)
> power4_cp=ols_mallows_cp(power4, full_chung)
>
> table7 <-
matrix(c(aic_tuyentinh4,bic_tuyentinh4,tuyentinh4_cp,r2_tu
yentinh4,
+
aic_linlog4,bic_linlog4,linlog4_cp,r2_linlog4,
+
aic_origin4,bic_origin4,origin4_cp,r2_origin4,
+
aic_power4,bic_power4,power4_cp,r2_power4),ncol=4,byrow=TR
UE)
> colnames(table7) <- c("AIC","BIC","CP","R Squared")
> rownames(table7) <- c("Tuyến tính","Lin Log","Gốc tọa
độ","Bậc 2")
>
> table7

```

	AIC	BIC	CP	R Squared
Tuyến tính	1464.988	1474.567	11.12393	0.1898588
Lin Log	1437.630	1447.124	25.14523	0.1276260
Gốc tọa độ	1486.471	1492.857	37.20021	0.5297689
Bậc 2	1465.252	1478.023	11.32758	0.1976360

Dùng hàm chọn các mô hình với chỉ số tốt nhất

	Tuyến tính	Lin log	Hồi quy qua gốc tọa độ	Dạng hàm bậc 2	Chọn mô hình
AIC	0,5	0,51	0,63	0,72	Tuyến tính
BIC	0,47	0,46	0,56	0,78	Lin log
Cp	2,1	1,9	1,7	1,2	Dạng hàm bậc 2
R ² hiệu chỉnh	0,73	0,68	0,72	0,85	Dạng hàm bậc 2

```

> best_aic=min(table7[,1], na.rm=FALSE)
> best_bic=min(table7[,2], na.rm=FALSE)
> best_cp=min(table7[,3], na.rm=FALSE)
> best_r2=max(table7[,4], na.rm=FALSE)
> best <-
matrix(c(best_aic,best_bic,best_cp,best_r2),ncol=1,byrow=TRUE)
> colnames(best) <- c("Best")
> rownames(best) <- c("AIC","BIC","CP","R Squared")
> best

              Best
AIC      1464.9878790
BIC      1474.5667495
CP         11.1239290
R Squared   0.5297689

```

Minh họa

--	--	--	--	--

8) Lấy các biến ở câu 4), chạy các dạng:

Hàm hồi quy log-log

```
> # Tuyến tính log
> ttlog4 <-
lm(log(slquanao)~log(dichoi),data=full,slquanao>0&dichoi>0
)
> summary(ttlog4)
```

Call:

```
lm(formula = log(slquanao) ~ log(dichoi), data = full,
    subset = slquanao >
      0 & dichoi > 0)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-3.07626	-0.41785	-0.06103	0.43918	2.60599

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.72349	0.12309	14.002	< 2e-16 ***
log(dichoi)	0.39774	0.06538	6.084	7.33e-09 ***

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7336 on 173 degrees of freedom

Multiple R-squared: 0.1762, Adjusted R-squared: 0.1715

F-statistic: 37.01 on 1 and 173 DF, p-value: 7.334e-09

a. Ý nghĩa hệ số góc

Hàm hồi quy log lin (log-tuyến tính)

```
> # Log Lin
```

```

> loglin4 <-
lm(log(slquanao)~dichoi,data=full,slquanao>0)
> summary(loglin4)

Call:
lm(formula = log(slquanao) ~ dichoi, data = full, subset =
  slquanao >
    0)

Residuals:
    Min       1Q   Median       3Q      Max
-3.6109 -0.4538  0.0615  0.4704  2.5285

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.967093    0.083285   23.619  < 2e-16 ***
dichoi        0.054794    0.008384    6.535 6.46e-10 ***
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7503 on 178 degrees of freedom
Multiple R-squared:  0.1935, Adjusted R-squared:  0.189
F-statistic: 42.71 on 1 and 178 DF, p-value: 6.463e-10

```

a. Ý nghĩa hệ số góc

b. Chọn mô hình

Ta tính toán các tiêu chuẩn trong phần mềm R của các mô hình và lập bảng so sánh như sau:

```

> # 8b. Chọn mô hình
> # Log Log
> r2_ttlog4 = summary(ttlog4)$r.squared
> aic_ttlog4 = AIC(ttlog4)
> bic_ttlog4 = BIC(ttlog4)
> ttlog4_cp=ols_mallows_cp(ttlog4, full_chung)
>
> # Log Lin
> r2_loglin4 = summary(loglin4)$r.squared
> aic_loglin4 = AIC(loglin4)

```

```

> bic_loglin4 = BIC(loglin4)
> loglin4_cp=ols_mallows_cp(loglin4, full_chung)
>
> table8 <-
matrix(c(aic_ttlog4,bic_ttlog4,ttlog4_cp,r2_ttlog4,
+
aic_loglin4,bic_loglin4,loglin4_cp,r2_loglin4),ncol=4,byrow=TRUE)
> colnames(table8) <- c("AIC","BIC","CP","R Squared")
> rownames(table8) <- c("Tuyến tính Log","Log Lin")
> table8

```

	AIC	BIC	CP	R Squared
Tuyến tính Log	392.1792	401.6736	-170.5010	0.1762339
Log Lin	411.3724	420.9512	-175.4629	0.1935117

Ta tìm các giá trị tốt nhất của chỉ số đánh giá bằng hàm max()

```

> # Đánh giá
> best_aic=min(table7[,1], na.rm=FALSE)
> best_bic=min(table7[,2], na.rm=FALSE)
> best_cp=min(table7[,3], na.rm=FALSE)
> best_r2=max(table7[,4], na.rm=FALSE)
> best <-
matrix(c(best_aic,best_bic,best_cp,best_r2),ncol=1,byrow=TRUE)
> colnames(best) <- c("Best")
> rownames(best) <- c("AIC","BIC","CP","R Squared")
> best

```

	Best
AIC	1437.6300793
BIC	1447.1244372
CP	11.1239290
R Squared	0.5297689

Có thể thấy giá trị các chỉ số đánh giá của hai mô hình xấp xỉ nhau. Tuy nhiên, mô hình Tuyến tính - Log có các giá trị tốt nhất.