OLS Regression

Jing Bu

10/23/2018

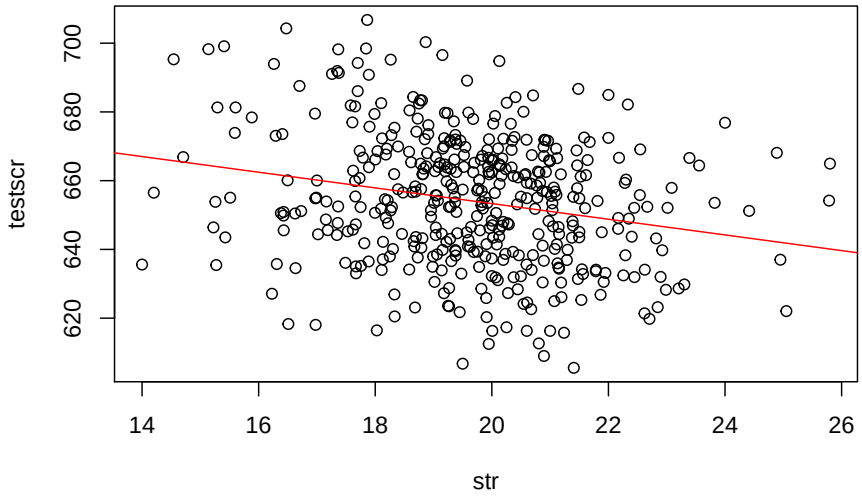
目录

1 Plot

1.1 Scatter Plot

- Draw a scatter plot of the variable testscr against str:

```
attach(cars_data_small)
plot(str, testscr)
abline(lm(testscr ~ str , data = cars_data_small),col = "red")
```

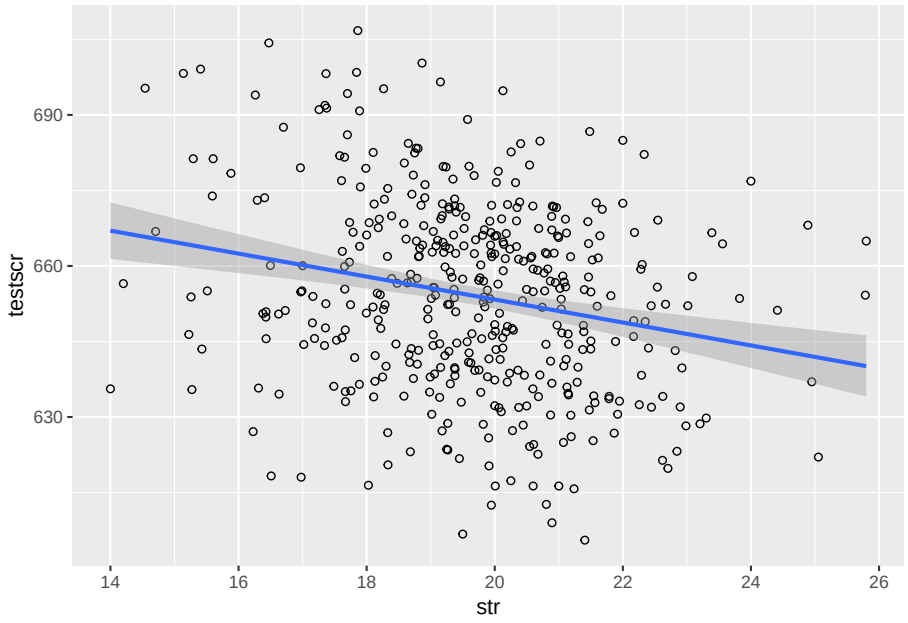



```
lm(formula, data, subset, weights, na.action,
   method = "qr", model = TRUE, x = FALSE, y = FALSE, qr = TRUE,
   singular.ok = TRUE, contrasts = NULL, offset, ···)
```

1.2 ggplot2

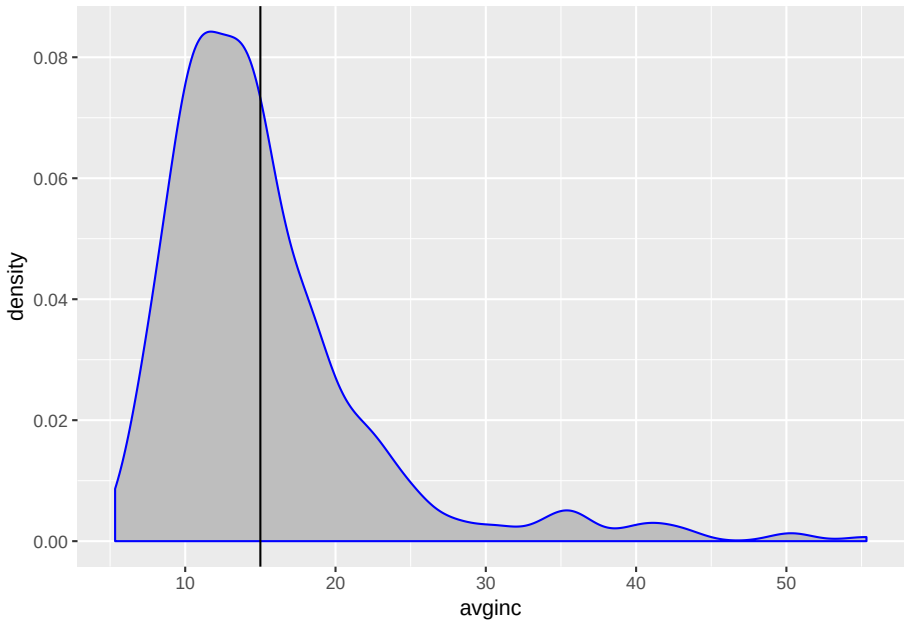
```
library("ggplot2")
ggplot(data =cars_data_small,aes(x=str, y=testscr)) +
```

```
geom_point(shape=1) + # Use hollow circles
geom_smooth(method=lm) # Add linear regression line
```

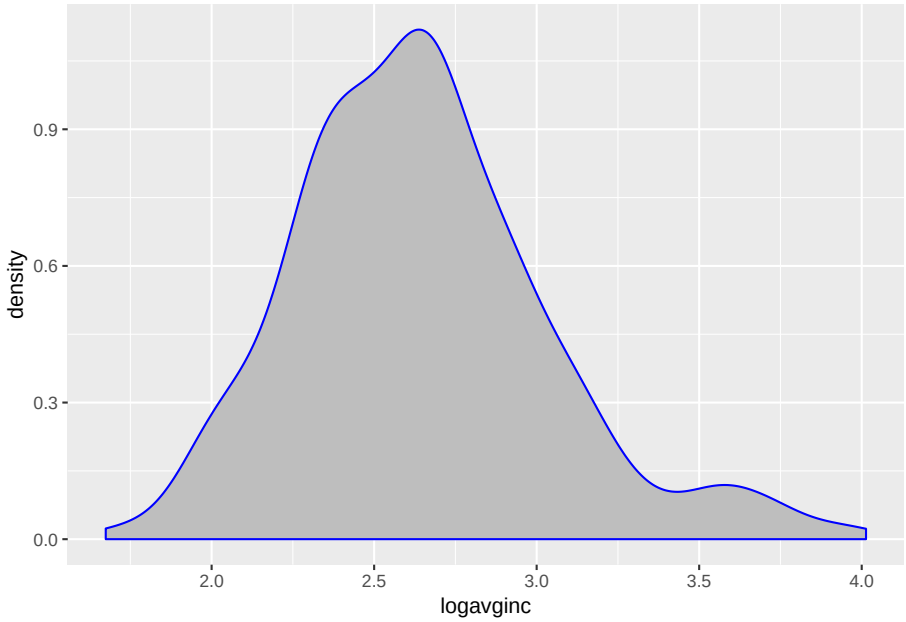


1.3 A kdensity distribution of income

```
cars_data$inc <- with(cars_data,avginc >=15)
ggplot(cars_data,aes(x=avginc))+
geom_density(fill="grey",color ="blue")+
geom_vline(xintercept = 15)
```

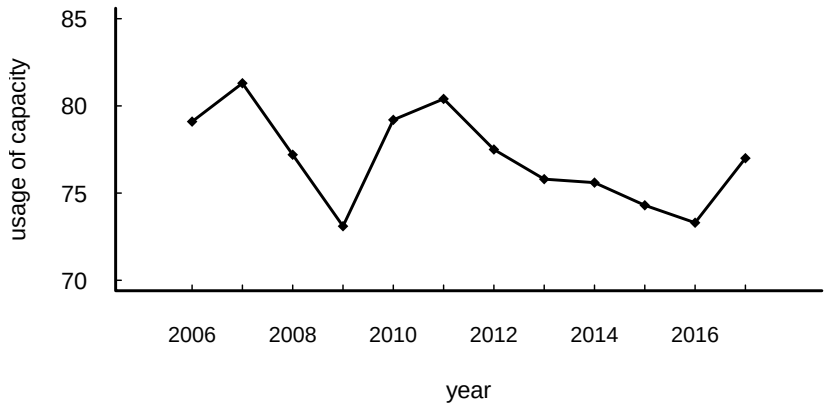



```
cars_data$logavginc <- log(cars_data$avginc)
ggplot(cars_data,aes(x=logavginc))+
geom_density(fill="grey",color ="blue")
```

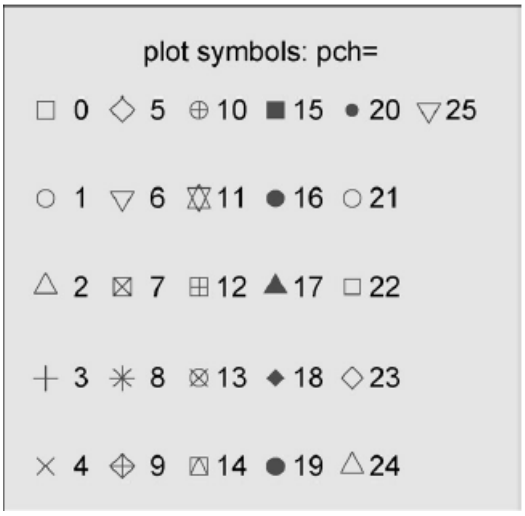


1.4 extra image

```
library(readxl)
image <- read_excel("/Users/admin/Desktop/teaching assiatant/Econometrics/teaching assi
attach(image)
opar<-par(no.readonly = TRUE)
par(pch=18,lwd=2)
par(cex=1,cex.axis=1,cex.lab=1)
par(font=1,font.axis=1,font.lab=1)
par(pin=c(5,2))
plot(year,rate,type="o",bty="l",ann=FALSE,xaxt="n",yaxt="n",xlim=c(2005,2018),ylim=c(70
title(xlab="year",ylab="usage of capacity")
axis(1,at=year,tck=0.02,cex.axis=0.95,las=0)
axis(2,tck=0.02,las=2,cex.axis=1)
```



```
par(opar)
```



2 T-test in R

2.1 single sample

- t-test for scores

```
summary(cars_data_small$testscr)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   605.5   640.0   654.5   654.2   666.7   706.8
```

```
t.test(cars_data_small$testscr,alternative = "two.sided",mu = 650)
```

```
##
##  One Sample t-test
##
## data:  cars_data_small$testscr
## t = 4.4708, df = 419, p-value = 1.005e-05
## alternative hypothesis: true mean is not equal to 650
## 95 percent confidence interval:
```

```
##  652.3291 655.9840
## sample estimates:
## mean of x
##  654.1565
```

3 OLS Regression

- https://cran.r-project.org/doc/contrib/Ricci-refcard-regression.pdf

```
fm1 <- lm(testscr ~ str,data = cars_data_small)
summary(fm1)
```

```
##
## Call:
## lm(formula = testscr ~ str, data = cars_data_small)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -47.727 -14.251   0.483  12.822  48.540
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 698.9330     9.4675  73.825  < 2e-16 ***
## str          -2.2798     0.4798  -4.751 2.78e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.58 on 418 degrees of freedom
## Multiple R-squared:  0.05124,    Adjusted R-squared:  0.04897
## F-statistic: 22.58 on 1 and 418 DF,  p-value: 2.783e-06
```

3.1 OLS Regression 2

```
fm2 <- lm(testscr ~ str,data = cars_data)

summary(fm2)
```

```
##
## Call:
## lm(formula = testscr ~ str, data = cars_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -47.727 -14.251   0.483  12.822  48.540
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 698.9330     9.4675  73.825  < 2e-16 ***
## str          -2.2798     0.4798  -4.751 2.78e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.58 on 418 degrees of freedom
## Multiple R-squared:  0.05124,    Adjusted R-squared:  0.04897
## F-statistic: 22.58 on 1 and 418 DF,  p-value: 2.783e-06
```

3.2 OLS Regression 3

```
fm3 <- lm(testscr ~ str+meal_pct,data = cars_data)

summary(fm3)
```

```
##
## Call:
## lm(formula = testscr ~ str + meal_pct, data = cars_data)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.657  -5.518  -0.168   5.645  34.494
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 702.91131    4.70025 149.548  < 2e-16 ***
## str          -1.11723    0.24036  -4.648  4.5e-06 ***
## meal_pct     -0.59975    0.01676 -35.775  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.222 on 417 degrees of freedom
## Multiple R-squared:  0.7668, Adjusted R-squared:  0.7657
## F-statistic: 685.8 on 2 and 417 DF,  p-value: < 2.2e-16
```

3.3 OLS Regression 4

```
fm4 <- lm(testscr ~ str+meal_pct+computer,data = cars_data)

summary(fm4)
```

```
##
## Call:
## lm(formula = testscr ~ str + meal_pct + computer, data = cars_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.638  -5.620  -0.144   5.690  34.544
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.031e+02  4.775e+00 147.239  < 2e-16 ***
```

```
## str         -1.129e+00  2.471e-01  -4.566 6.55e-06 ***
## meal_pct    -5.999e-01  1.679e-02 -35.723  < 2e-16 ***
## computer     2.108e-04  1.052e-03   0.200    0.841
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.233 on 416 degrees of freedom
## Multiple R-squared:  0.7669, Adjusted R-squared:  0.7652
## F-statistic: 456.1 on 3 and 416 DF,  p-value: < 2.2e-16
```

3.4 OLS Regression 5

```
fm5 <- lm(testscr ~ str+I(str^2),data = cars_data)

summary(fm5)
```

```
##
## Call:
## lm(formula = testscr ~ str + I(str^2), data = cars_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -47.349 -14.226   0.556  12.981  48.593
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 739.4456    60.7179  12.178   <2e-16 ***
## str          -6.4490     6.1906  -1.042    0.298
## I(str^2)      0.1063     0.1573   0.676    0.500
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.59 on 417 degrees of freedom
## Multiple R-squared:  0.05228,    Adjusted R-squared:  0.04773
```

```
## F-statistic:  11.5 on 2 and 417 DF,  p-value: 1.374e-05
```

3.5 OLS Regression 6

```
fm6 <- lm(testscr ~ str+str:el_pct,data = cars_data)

summary(fm6)
```

```
##
## Call:
## lm(formula = testscr ~ str + str:el_pct, data = cars_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -49.053 -10.119  -0.704   9.873  43.898
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 677.610055   7.504411  90.295   <2e-16 ***
## str          -0.673273   0.387181  -1.739   0.0828 .
## str:el_pct   -0.032357   0.001973 -16.398   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.51 on 417 degrees of freedom
## Multiple R-squared:  0.4232, Adjusted R-squared:  0.4204
## F-statistic:   153 on 2 and 417 DF,  p-value: < 2.2e-16
```

符　号	用　途
~	分隔符号，左边为响应变量，右边为解释变量。例如，要通过x、z和w预测y，代码为y ~ x + z + w
+	分隔预测变量
:	表示预测变量的交互项。例如，要通过x、z及x与z的交互项预测y，代码为y ~ x + z + x:z
*	表示所有可能交互项的简洁方式。代码y~ x * z * w可展开为y ~ x + z + w + x:z + x:w + z:w + x:z:w
^	表示交互项达到某个次数。代码y ~ (x + z + w)^2可展开为y ~ x + z + w + x:z + x:w + z:w
.	表示包含除因变量外的所有变量。例如，若一个数据框包含变量x、y、z和w，代码y~.可展开为y ~ x + z + w
-	减号，表示从等式中移除某个变量。例如，y ~ (x + z + w)^2 - x:w可展开为y ~ x + z + w + x:z + z:w
-1	删除截距项。例如，表达式y ~ x - 1拟合y在x上的回归，并强制直线通过原点
I()	从算术的角度来解释括号中的元素。例如，y~x+(z+w)^2将展开为y~x+z+w+z:w。相反，代码y ~ x + I((z + w)^2)将展开为y ~ x + h，h是一个由z和w的平方和创建的新变量
function	可以在表达式中用的数学函数。例如，log(y) ~ x + z + w表示通过x、z和w来预测log(y)

4 Published Tables: Stargazer

```
#install.packages("stargazer",repos = "http://mirrors.ustc.edu.cn/CRAN/")
library(stargazer)
```

```
##
## Please cite as:

##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics T

##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

4.1 Published Tables: Stargazer1

```
stargazer(fm1,fm3,fm4, title="Regression Results1")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu % Date and time: 周一, 10 月 22, 2018 - 23:04:49

表 1: Regression Results1

	Dependent variable:		
	testscr		
	(1)	(2)	(3)
str	−2.280***	−1.117***	−1.129***
	(0.480)	(0.240)	(0.247)
meal_pct		−0.600***	−0.600***
		(0.017)	(0.017)
computer			0.0002
			(0.001)
Constant	698.933***	702.911***	703.074***
	(9.467)	(4.700)	(4.775)
Observations	420	420	420
R^2	0.051	0.767	0.767
Adjusted R^2	0.049	0.766	0.765
Residual Std. Error	18.581 (df = 418)	9.222 (df = 417)	9.233 (df = 416)
F Statistic	22.575*** (df = 1; 418)	685.756*** (df = 2; 417)	456.132*** (df = 3; 416)

Note: *p<0.1; **p<0.05; ***p<0.01

4.2 Published Tables: Stargazer2

```
stargazer(fm1,fm3,fm4,title="Regression Results2", se = NA)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu % Date and time: 周一, 10 月 22, 2018 - 23:04:50

4.3 Published Tables: Stargazer3

```
stargazer(fm1,fm3,fm4, title="Regression Results3", add.lines = TRUE)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu % Date and time: 周一, 10 月 22, 2018 - 23:04:50

4.4 Published Tables: Stargazer4

```
stargazer(fm1,fm3,fm4, title="Regression Results4", ci = TRUE)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu % Date and time: 周一, 10 月 22, 2018 - 23:04:50

表 2: Regression Results2

	Dependent variable:		
	testscr		
	(1)	(2)	(3)
str	−2.280	−1.117***	−1.129***
		(0.240)	(0.247)
meal_pct		−0.600***	−0.600***
		(0.017)	(0.017)
computer			0.0002
			(0.001)
Constant	698.933	702.911***	703.074***
		(4.700)	(4.775)
Observations	420	420	420
R^2	0.051	0.767	0.767
Adjusted R^2	0.049	0.766	0.765
Residual Std. Error	18.581 (df = 418)	9.222 (df = 417)	9.233 (df = 416)
F Statistic	22.575*** (df = 1; 418)	685.756*** (df = 2; 417)	456.132*** (df = 3; 416)
Note:			*p<0.1; **p<0.05; ***p<0.01

表 3: Regression Results3

	Dependent variable:		
	testscr		
	(1)	(2)	(3)
str	−2.280***	−1.117***	−1.129***
	(0.480)	(0.240)	(0.247)
meal_pct		−0.600***	−0.600***
		(0.017)	(0.017)
computer			0.0002
			(0.001)
Constant	698.933***	702.911***	703.074***
	(9.467)	(4.700)	(4.775)
TRUE			
Observations	420	420	420
R^2	0.051	0.767	0.767
Adjusted R^2	0.049	0.766	0.765
Residual Std. Error	18.581 (df = 418)	9.222 (df = 417)	9.233 (df = 416)
F Statistic	22.575*** (df = 1; 418)	685.756*** (df = 2; 417)	456.132*** (df = 3; 416)

Note: *p<0.1; **p<0.05; ***p<0.01

表 4: Regression Results4

	Dependent variable:		
	testscr		
	(1)	(2)	(3)
str	−2.280***	−1.117***	−1.129***
	(−3.220, −1.339)	(−1.588, −0.646)	(−1.613, −0.644)
meal_pct		−0.600***	−0.600***
		(−0.633, −0.567)	(−0.633, −0.567)
computer			0.0002
			(−0.002, 0.002)
Constant	698.933***	702.911***	703.074***
	(680.377, 717.489)	(693.699, 712.124)	(693.715, 712.433)
Observations	420	420	420
R^2	0.051	0.767	0.767
Adjusted R^2	0.049	0.766	0.765
Residual Std. Error	18.581 (df = 418)	9.222 (df = 417)	9.233 (df = 416)
F Statistic	22.575*** (df = 1; 418)	685.756*** (df = 2; 417)	456.132*** (df = 3; 416)

Note: *p<0.1; **p<0.05; ***p<0.01