

Nonlinear Regression

Jing Bu

11/6/2018

目录

1 Nonlinear Regression Functions:	1
1.1 Introduction	1
1.2 Heteroscedasticity	2
1.3 Robust Regression	6
1.4 Logistic and Probit	7
2 homework 1: RAND Experiment	9
2.1 Question1	10
2.2 Question2	11
2.3 Question3	13
2.4 Question4	14

1 Nonlinear Regression Functions:

1.1 Introduction

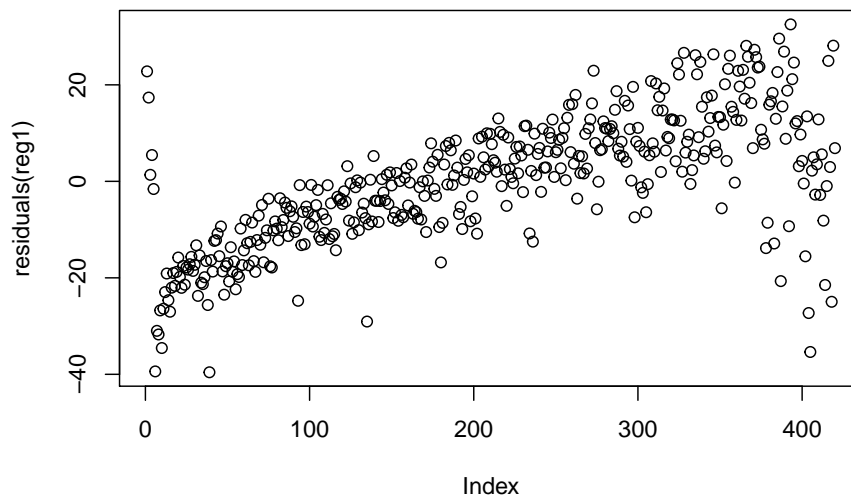
- We extend nonlinear into two cases
 1. nonlinear in Xs
 - Logarithms and Interactions
 2. nonlinear in function
 - Discrete Dependent Variables or Limited Dependent Variables
 - Linear function is not a good prediction function.

1.2 Heteroscedasticity

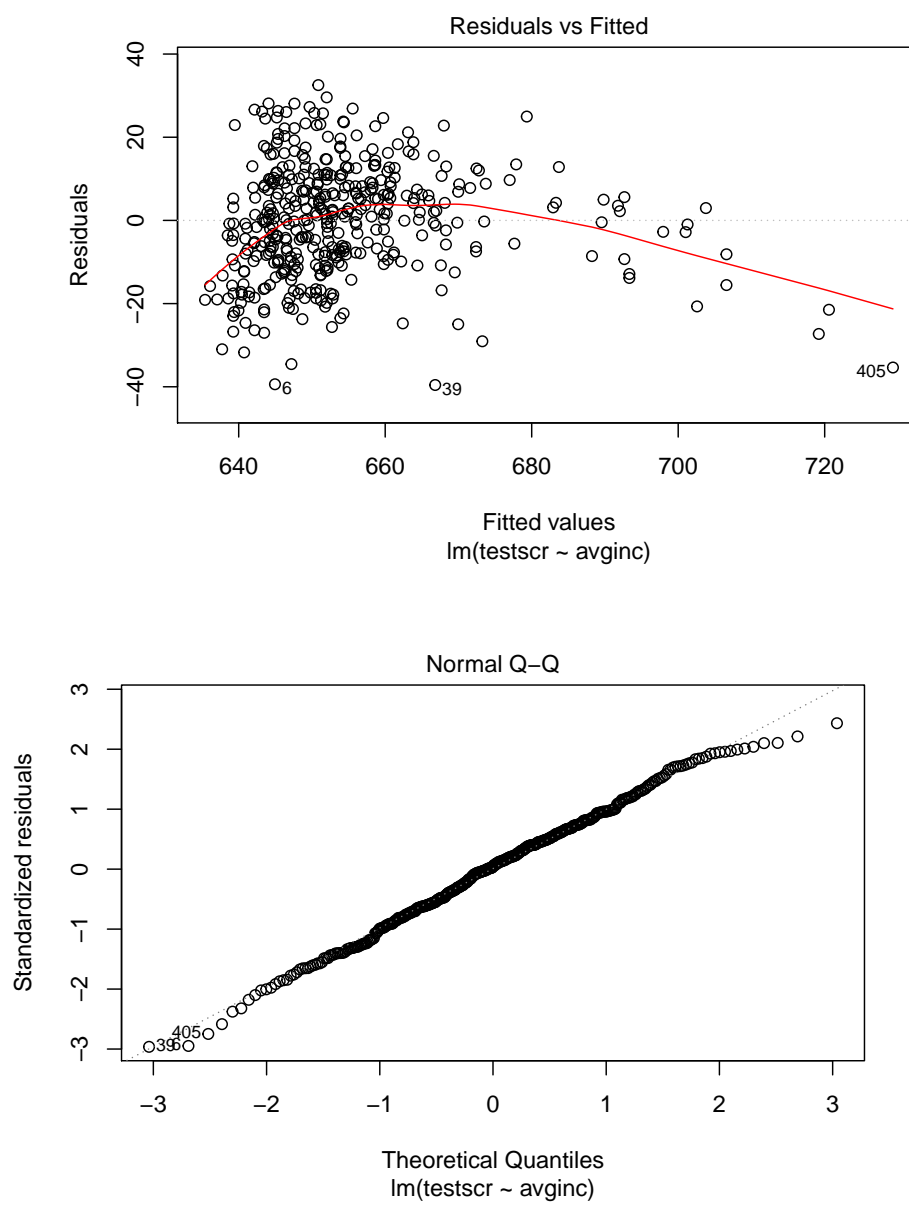
1.2.1 Plot

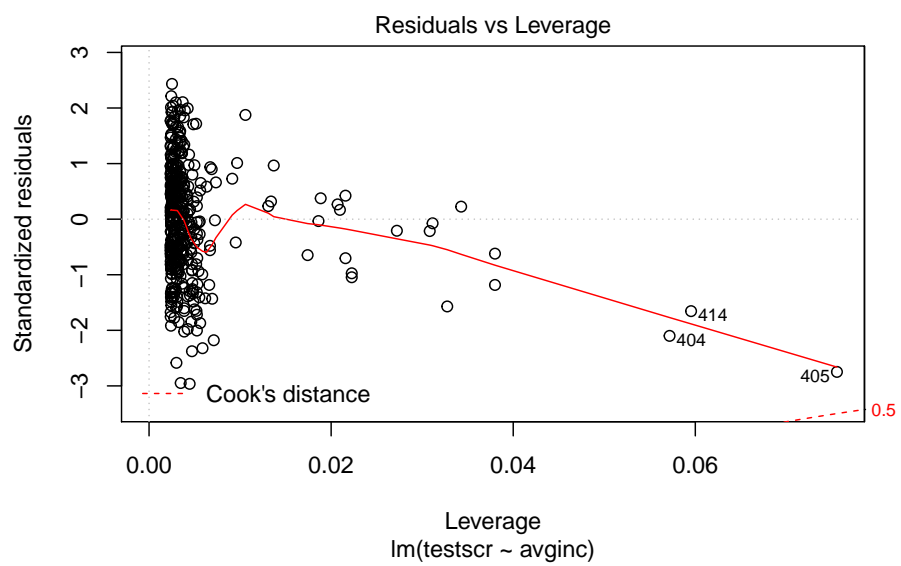
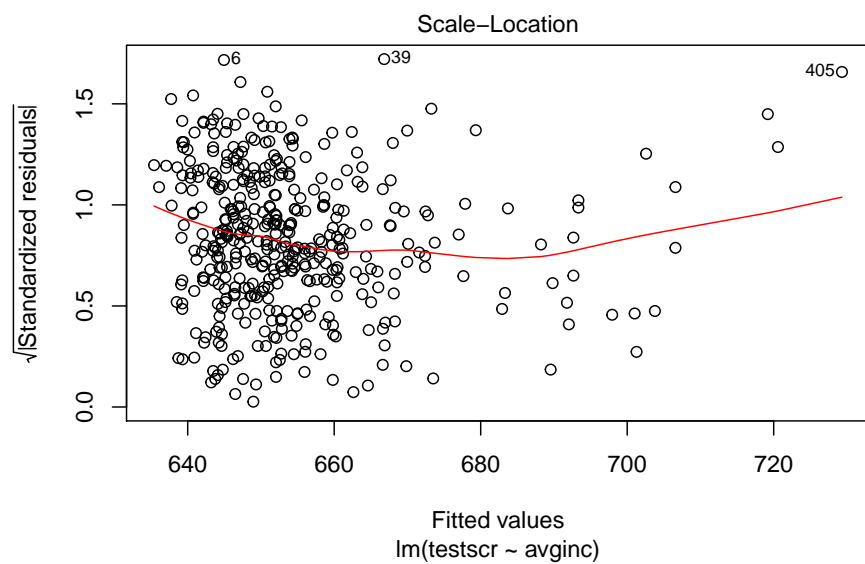
```
library("foreign")  
caschool<-read.dta("/Users/admin/Desktop/teaching assiatant/Econometrics/teaching assis
```

```
reg1<-lm(testscr ~ avginc,data=caschool)  
plot(residuals(reg1))
```



```
plot(reg1)
```





1.2.2 Test Heteroscedasticity

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

```
## Loading required package: zoo
```

```
##
```

```
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      as.Date, as.Date.numeric
```

```
reg2<-lm(testscr ~ avginc,data=caschool)
bptest(reg2)
```

```
##
```

```
## studentized Breusch-Pagan test
```

```
##
```

```
## data: reg2
```

```
## BP = 0.078688, df = 1, p-value = 0.7791
```

```
summary(reg2)
```

```
##
```

```
## Call:
```

```
## lm(formula = testscr ~ avginc, data = caschool)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -39.574  -8.803   0.603   9.032  32.530
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 625.3836      1.5324  408.11   <2e-16 ***
## avginc      1.8785      0.0905   20.76   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.39 on 418 degrees of freedom
## Multiple R-squared:  0.5076, Adjusted R-squared:  0.5064
## F-statistic: 430.8 on 1 and 418 DF,  p-value: < 2.2e-16
```

```
avginc_log<-exp(caschool$avginc)
reg3<-lm(testscr ~ avginc_log,data=caschool)
bptest(reg3)
```

```
##
## studentized Breusch-Pagan test
##
## data:  reg3
## BP = 0.50589, df = 1, p-value = 0.4769
```

1.3 Robust Regression

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

```
## Loading required package: Matrix
##
## Attaching package: 'lfe'
##
## The following object is masked from 'package:lmtest':
##
##      waldtest
```

```
reg4<-felm(testscr~avginc,data=caschool)
summary(reg4)

##
## Call:
##   felm(formula = testscr ~ avginc, data = caschool)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -39.574  -8.803   0.603   9.032  32.530
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  625.3836     1.5324  408.11  <2e-16 ***
## avginc        1.8785      0.0905   20.76  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.39 on 418 degrees of freedom
## Multiple R-squared(full model): 0.5076   Adjusted R-squared: 0.5064
## Multiple R-squared(proj model): 0.5076   Adjusted R-squared: 0.5064
## F-statistic(full model):430.8 on 1 and 418 DF, p-value: < 2.2e-16
## F-statistic(proj model): 430.8 on 1 and 418 DF, p-value: < 2.2e-16
```

1.4 Logistic and Probit

```
library("foreign")
womenwk<-read.dta("/Users/admin/Desktop/teaching assiatant/Econometrics/teaching assist

logit<-glm(work~age+married+children+education,data=womenwk,family = binomial(link=logi
summary(logit)

##
```

```
## Call:
## glm(formula = work ~ age + married + children + education, family = binomial(link =
##      data = womenwk)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -2.6212  -0.9292   0.4614   0.8340   2.0455
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.159247   0.332040 -12.526 < 2e-16 ***
## age          0.057930   0.007221   8.022 1.04e-15 ***
## married      0.741777   0.126471   5.865 4.49e-09 ***
## children     0.764488   0.051529  14.836 < 2e-16 ***
## education    0.098251   0.018652   5.268 1.38e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2532.4  on 1999  degrees of freedom
## Residual deviance: 2055.8  on 1995  degrees of freedom
## AIC: 2065.8
##
## Number of Fisher Scoring iterations: 5
```

```
probit<-glm(work~age+married+children+education,data=womenwk,family = binomial(link= pr
summary(probit)
```

```
##
## Call:
## glm(formula = work ~ age + married + children + education, family = binomial(link =
##      data = womenwk)
##
```



```
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -2.7594  -0.9414   0.4552   0.8459   2.0427
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.467365   0.192291 -12.831 < 2e-16 ***
## age          0.034721   0.004232   8.204 2.33e-16 ***
## married      0.430857   0.074310   5.798 6.71e-09 ***
## children     0.447325   0.028642  15.618 < 2e-16 ***
## education    0.058365   0.011018   5.297 1.18e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2532.4  on 1999  degrees of freedom
## Residual deviance: 2054.1  on 1995  degrees of freedom
## AIC: 2064.1
##
## Number of Fisher Scoring iterations: 5
```

2 homework 1: RAND Experiment

- The RAND Health Insurance Experiment (HIE), which ran from 1974 to 1982, was one of the most influential social experiments in research history. The HIE enrolled 3,958 people aged 14 to 61 from six areas of USA. The HIE sample excluded Medicare participants and most Medicaid and military health insurance subscribers. HIE participants were randomly assigned to one of 14 insurance plans. Participants did not have to pay insurance premiums, but the plans had a variety of provisions related to cost sharing, leading to large differences in the amount of insurance they offered.

Variable Names	Discriptions
any_ins	= 1 if has any health insurance assigned; = 0 otherwise
female	= 1 if female; = 0 otherwise
blackhisp	= 1 if nonwhite; = 0 otherwise
educper	years of education
hosp	hospitalized last year
ghindx	pre-treatment outcome : general health index
cholest	pre-treatment outcome: cholesterol level (mg/dl)
ghindx	post-treatment outcome: general health index
cholestx	post-treatment outcome: cholesterol level (mg/dl)

2.1 Question1

- Generate basic summary statistics such as mean, standard deviation, and total number of observations for variables female, blackhisp,educper, ghindx, and cholest, separated by the any_ins indicator. Your table for each variables should look similar as the table in Question 1. You can also combine all the information in just one table.

```
library("foreign")
library("plyr")
rand<-read.dta("/Users/admin/Desktop/teaching assiatant/Econometrics/teaching assistant
ddply(rand,"any_ins",summarize,mean_female=mean(female,na.rm=TRUE),sd_female=sd(female,
```

```
##  any_ins mean_female sd_female mean_blackhisp sd_blackhisp mean_educper
## 1      0  0.5599473 0.4967206      0.1716667  0.3774051    12.10483
## 2      1  0.5303315 0.4991572      0.1465824  0.3537590    11.93619
##  sd_educper mean_ghindx sd_ghindx mean_cholest sd_cholest  obs
## 1  2.881461    70.95892 14.65553    207.0904  42.44610 2689
## 2  3.005082    69.93396 15.02703    204.1117  43.54407 3198
```

Variable	any_ins=0	any_ins=1
female	0.560 [0.497]	0.530 [0.499]
blackhisp	0.1717 [0.377]	0.1466 [0.354]
educper	12.105 [2.882]	11.936 [3.005]
ghindx	70.959 [14.656]	69.934 [15.027]
cholest	207.090 [42.446]	204.112 [43.544]
obs	2689	3198

2.2 Question2

- Three personal characteristics variables in question 1 (thus female, blackhisp and educper) and two pre-treatment outcome variables (thus ghindx, and cholest) can be considered as baseline outcome variables. For these variables, test the null hypothesis that there is no mean difference ($H_0: = 0$). (You should write down the equation of the statistic which you are going to use.) And What the baseline outcomes mean to the experiment?

```
t.test(female~any_ins,data=rand)
```

```
##
## Welch Two Sample t-test
##
## data: female by any_ins
## t = 1.4753, df = 1149.2, p-value = 0.1404
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.009770902 0.069002586
```

```
## sample estimates:
## mean in group 0 mean in group 1
##      0.5599473      0.5303315
```

```
t.test(blackhisp~any_ins,data=rand)
```

```
##
## Welch Two Sample t-test
##
## data:  blackhisp by any_ins
## t = 1.4811, df = 865.62, p-value = 0.1389
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.008156602  0.058325178
## sample estimates:
## mean in group 0 mean in group 1
##      0.1716667      0.1465824
```

```
t.test(educper~any_ins,data=rand)
```

```
##
## Welch Two Sample t-test
##
## data:  educper by any_ins
## t = 1.3441, df = 1030.1, p-value = 0.1792
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.07756339  0.41484572
## sample estimates:
## mean in group 0 mean in group 1
##      12.10483      11.93619
```

```
t.test(ghindx~any_ins,data=rand)

##
## Welch Two Sample t-test
##
## data: ghindx by any_ins
## t = 2.2577, df = 4037.9, p-value = 0.02402
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1349045 1.9150177
## sample estimates:
## mean in group 0 mean in group 1
##      70.95892      69.93396
```

```
t.test(cholest~any_ins,data=rand)

##
## Welch Two Sample t-test
##
## data: cholest by any_ins
## t = 1.7837, df = 2177.2, p-value = 0.07461
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.2961282 6.2535693
## sample estimates:
## mean in group 0 mean in group 1
##      207.0904      204.1117
```

2.3 Question3

- For the two post-treatment variables, (thus ghindx, and cholest) repeat the exercises you just did (compute group mean, standard deviation, and run a t-test)

```
t.test(ghindx~any_ins,data=rand)

##
## Welch Two Sample t-test
##
## data: ghindx by any_ins
## t = 4.7132, df = 4745.3, p-value = 2.508e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.170210 2.837027
## sample estimates:
## mean in group 0 mean in group 1
##      70.11108      68.10746
```

```
t.test(cholestd~any_ins,data=rand)

##
## Welch Two Sample t-test
##
## data: cholestx by any_ins
## t = -0.68553, df = 3504.8, p-value = 0.4931
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -3.734833 1.799709
## sample estimates:
## mean in group 0 mean in group 1
##      200.8954      201.8630
```

2.4 Question4

- Interpret the results you obtain (the mean comparison results for characteristic variables, pre-treatment outcome variables, and post-treatment outcome variables).