Esame di Statistical Modeling

Mario Pedol

4/20/2021

Set Up Environment

Prima di iniziare a svolgere l'esercizio è necessario fare il caricamento di tutte le librerie necessarie ai fini dell'analisi. Si defenisce inoltre, la funzione per il test di white necessaria in seguito e si caricano i dati.

```r
library(skedastic)
library(car)
library(describedata)
library(psych)
library(klaR)
library(olsrr)
library(sandwich)
library(systemfit)
library(DataCombine)
library(lmtest)
white.test<-function(lmod){

  u2<-lmod$residuals^2

  y<-lmod$fitted

  R2u<-summary(lm(u2~y+I(y^2)))$r.squared

  LM<-length(y)*R2u

  p.val<-1-pchisq(LM,2)

  data.frame("Test Statistic"=LM, "P"=p.val)
}
data<-read.csv("~/Desktop/sm_esame290421.csv", header=TRUE)
head(data)
```

```
##   time        x1          x2        x3        y
## 1  192  6.131982 -0.70076947  50.93935 20.67729
## 2  122 14.256518  1.25404892 330.99816 22.84596
## 3   28  6.296125  0.07431125  79.15460 23.23476
## 4   94 18.650870  2.41529786 724.94578 28.12963
## 5   57  7.283519 -0.68889414  63.88759 23.69631
## 6  185 11.781854 -1.92217621 293.91335 21.15747
```

Statistiche Descrittive

Prima di iniziare ad esplorare i dati con opportune statistiche descrittive effettuo un controllo per verificare che il datastet sia stato letto correttamente, se così non fosse, occorre modificare i parametri di lettura o il type delle variabili.

```r
str(data)
```

```
## 'data.frame':    200 obs. of  5 variables:
##  $ time: int  192 122 28 94 57 185 104 12 6 15 ...
##  $ x1  : num  6.13 14.26 6.3 18.65 7.28 ...
##  $ x2  : num  -0.7008 1.254 0.0743 2.4153 -0.6889 ...
##  $ x3  : num  50.9 331 79.2 724.9 63.9 ...
##  $ y   : num  20.7 22.8 23.2 28.1 23.7 ...
```

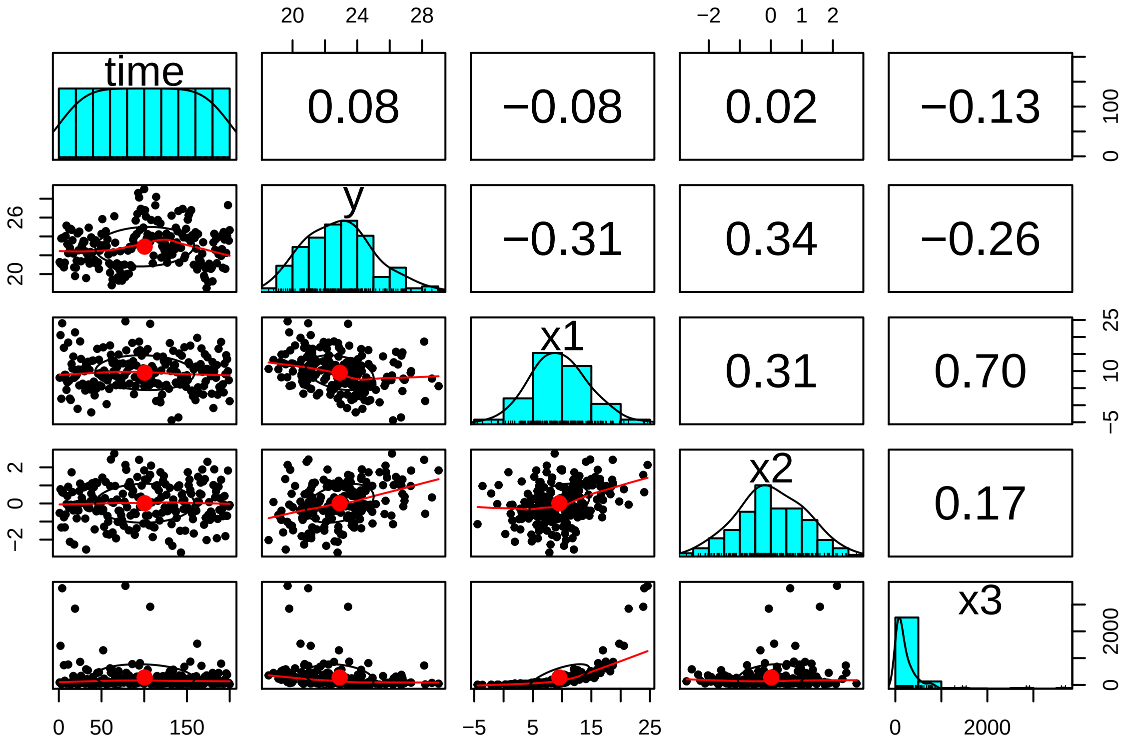
I dati sembrano essere stati letti senza errori.
Ora è possibile svolgere delle analisi descrittive che ci permettano di prendere confidenza con i dati.

```r
summary(data)
```

```
##       time              x1               x2                  x3          
##  Min.   :  1.00   Min.   :-4.440   Min.   :-2.71699   Min.   :   3.281  
##  1st Qu.: 50.75   1st Qu.: 6.059   1st Qu.:-0.66916   1st Qu.:  61.562  
##  Median :100.50   Median : 9.366   Median :-0.05528   Median : 131.239  
##  Mean   :100.50   Mean   : 9.533   Mean   : 0.01307   Mean   : 276.579  
##  3rd Qu.:150.25   3rd Qu.:12.549   3rd Qu.: 0.89788   3rd Qu.: 297.468  
##  Max.   :200.00   Max.   :24.603   Max.   : 2.76130   Max.   :3699.064  
##        y        
##  Min.   :18.52  
##  1st Qu.:21.20  
##  Median :22.84  
##  Mean   :22.91  
##  3rd Qu.:24.18  
##  Max.   :29.02
```

Dalla summary si vedono range non troppo elevati per x_1, x_2 che presentano anche un valore minimo negativo, mentre y presenta un range contenuto.
Range elvati si trovano invece per la varibile *time* e sopratutto x_3.
Proseguendo con le analisi si decide di utilizzare la funzione pairs.panels sulle variabili numeriche che ci consente di valutre le dstribuzioni e le correlazione delle variabili selezionate contemporanemente.

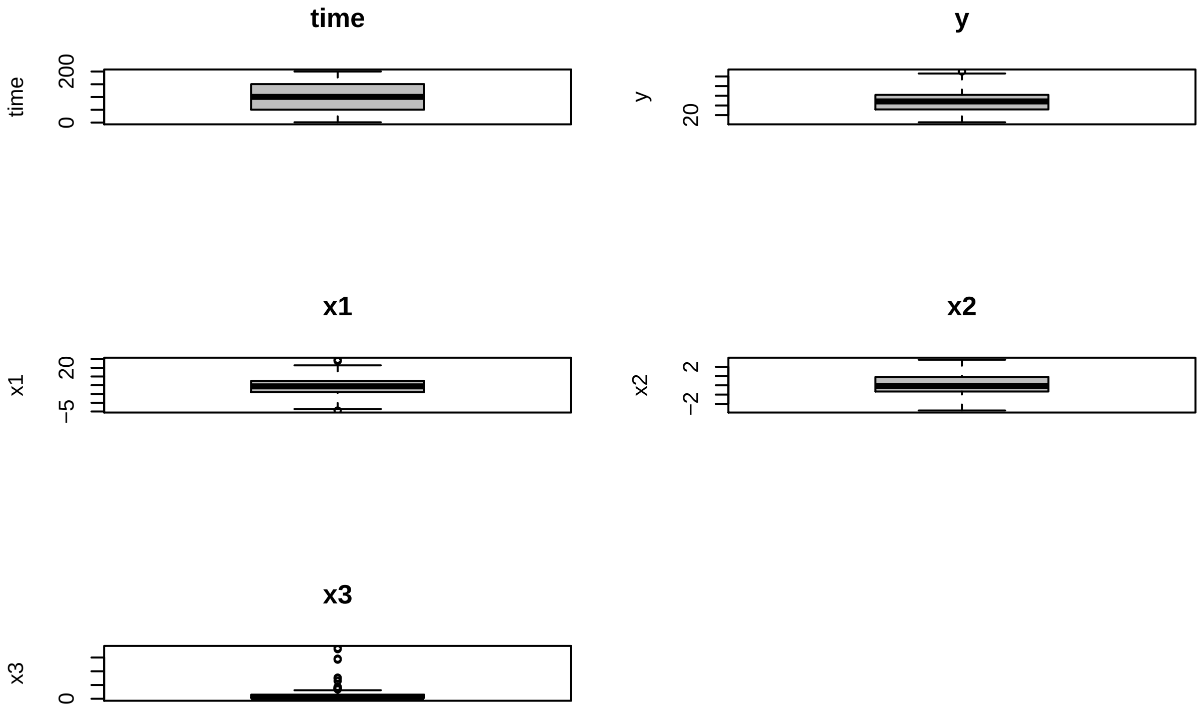
```r
var_num<-c('time','y','x1','x2','x3')
pairs.panels(data[,var_num])
```



Dal grafico emergono distrbuzioni piuttosto simmetriche tranne per x_3 che risulta fortemnte schiacciata a destra.

Si notano alcune correllazioni negative, anche se non particolarmente elevate come x_3 con y, mentre invece la correlazione più alta al 70% risulta essere quella tra x_3 e x_1.

```
par(mfrow=c(3,2))
for(i in var_num){
  boxplot(data[,i],main=i,col="grey",ylab=i)
}
```



Dai Box Plot risulta che x_3 è particolarmente schiaccita verso lo 0 con alcuni valori anomali, risulta anche qualche osservazione anomala per x_1 e y.

Modelli

Si passa ora alla specifica del modello e alla stima dei paramentri rischiesta dal punto 2 dell'esercizio.

```r
mod<-lm(y~x1+x2+I(log(x3)),data)
summary(mod)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + I(log(x3)), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6859 -1.1202 -0.0515  0.8360  4.9887
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 27.06768    1.51116  17.912  < 2e-16 ***
## x1           0.02248    0.13660   0.165  0.86943
## x2           0.70371    0.19458   3.617  0.00038 ***
## I(log(x3))  -0.89438    0.56447  -1.584  0.11470
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.748 on 196 degrees of freedom
## Multiple R-squared:  0.3183, Adjusted R-squared:  0.3078
```

```
## F-statistic:  30.5 on 3 and 196 DF,  p-value: 3.162e-16
```

Nel modello risultano molto significativi i parametri dell'intercetta e x_2, mentre non sono significativi x_1 e $log(x_3)$.
I parametri possono essere interpretati come segue:

1. x1: Un incremento unitario di x_1 porta ad un aumento dello 0.0225 del volore di y, al netto di tutte le altre variabili.
2. x2: Un incremento unitario di x_2 porta ad un aumento dello 0.7037 del volore di y, al netto di tutte le altre variabili.
3. log(x3): Per l'incremento dell'1% del tasso di x_3 y diminuisce dello 0.0089% (unità) al netto di tutte le altre varibaili.

In generale, il modello ha un R^2 piccolo circa 32%, tuttavia si rifiuta l'ipotesi nulla $H_0 : R^2 = 0$ alle usuali soglie di significatività, quindi si accetta il modello.

Si procede ora a testatre le ipotesi del modello classico.
Multicollinearità

```r
ols_vif_tol(mod)
```

```
##    Variables  Tolerance       VIF
## 1         x1 0.03149537 31.750696
## 2         x2 0.34794905  2.873984
## 3 I(log(x3)) 0.03470776 28.812002
```

```r
ols_eigen_cindex(mod)
```

```
##    Eigenvalue Condition Index    intercept           x1           x2
## 1 2.885319483        1.000000 0.0007539820 7.926600e-04 0.0004437957
## 2 1.007306244        1.692451 0.0000617798 3.105362e-05 0.3387415037
## 3 0.106144918        5.213716 0.0307002236 3.294456e-02 0.0437347227
## 4 0.001229354       48.446060 0.9684840146 9.662317e-01 0.6170799779
##     I(log(x3))
## 1 2.256514e-04
## 2 1.019403e-05
## 3 9.465887e-08
## 4 9.997641e-01
```

Si nota subito che i valori della varianza fattoriale sono elvati per x_1 e x_3, anche il condition index è elevato per tali valori. Per risolvere la presenza di multicollinearità allora si decide di togliere dal modello l'esplicativia con l'autovalore più elvato ossia x_1.

```r
mod<-lm(y~x2+I(log(x3)),data)
summary(mod)
```

```
##
## Call:
## lm(formula = y ~ x2 + I(log(x3)), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6743 -1.1198 -0.0646  0.8120  4.9948
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  26.8348     0.5295  50.682  < 2e-16 ***
## x2            0.7295     0.1148   6.354 1.42e-09 ***
## I(log(x3))   -0.8031     0.1052  -7.634 9.57e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.743 on 197 degrees of freedom
## Multiple R-squared:  0.3182, Adjusted R-squared:  0.3112
## F-statistic: 45.96 on 2 and 197 DF,  p-value: < 2.2e-16
```

Come ci si poteva aspettare anche se l'R^2 non è variato, ora tutti i parametri del modello sono significativi.

```r
ols_vif_tol(mod)
```

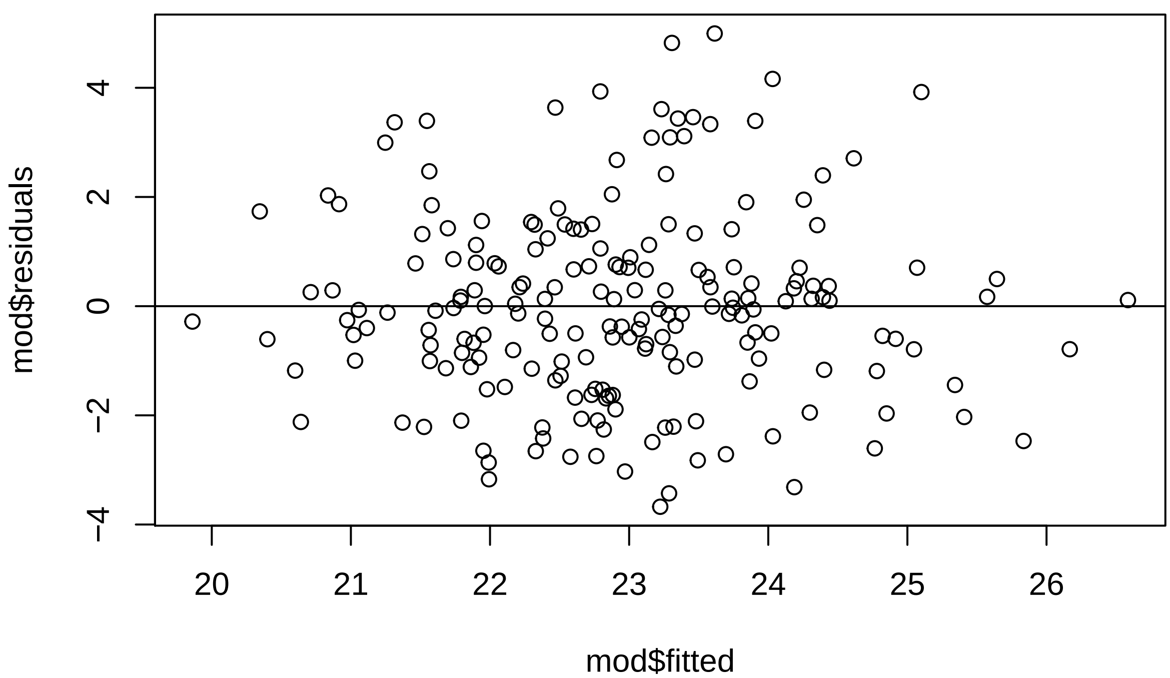
```
##    Variables Tolerance      VIF
## 1         x2 0.9944296 1.005602
## 2 I(log(x3)) 0.9944296 1.005602
```

Non sembrano esserci più problemi di multicollinerità, si procede dunque a testare le altre ipotesi.

Omoschedasticità

Per testare questa ipotesi se deciso ti usare il seguente grafico e il test di White.

```r
par(mfrow=c(1,1))
plot(mod$fitted, mod$residuals)
abline(h=0)
```




```r
white.test(mod)
```

```
##   Test.Statistic         P
```

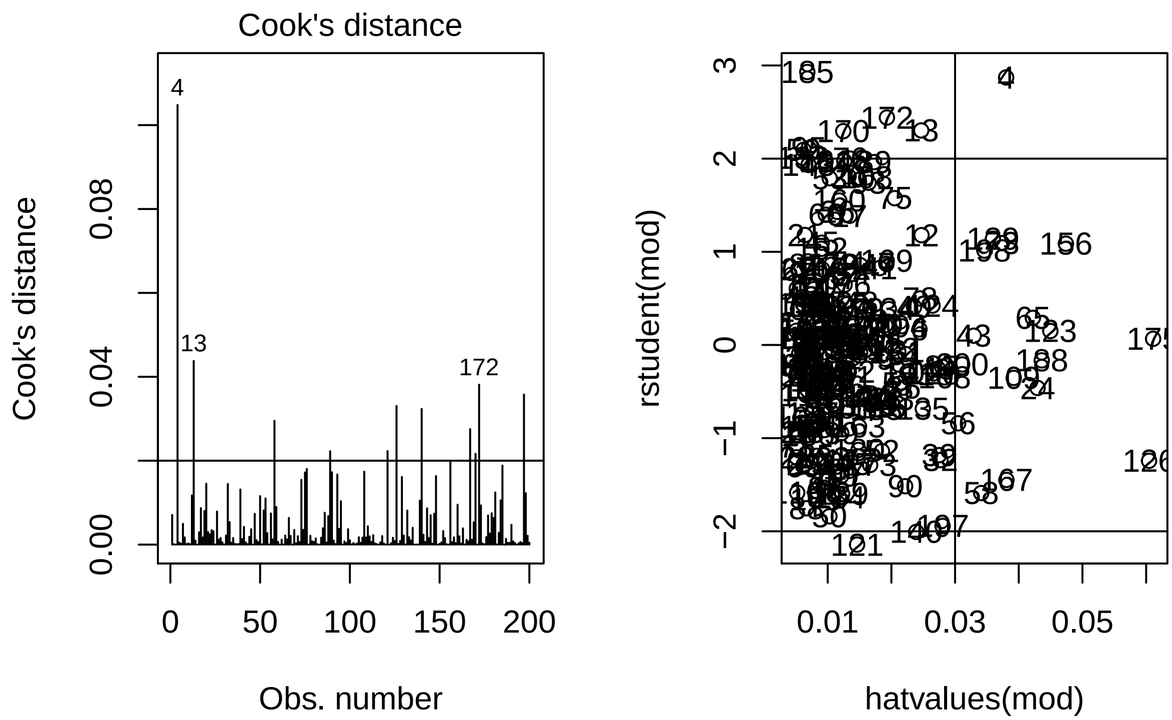
```
## 1        3.792393 0.1501386
```

Dal grafico è difficile capire se i residui presentino un andamento a ventaglio che indichi la presenza di eteroschedasticità. Affidandoci al test di white, non si rifiuta l'ipotesi nulla che i residui siano normali. si precede dunque con la prossima ipotesi.

Outlier

Per verificare la presenza di outlier si analizzano i grafici della distanza di cook e dei residui studentizzati con le loro reltive soglie.

```
par(mfrow=c(1,2))
k=length(coef(mod))
n=nrow(data)
plot(mod, which = 4)
abline(h=4/n)
plot(hatvalues(mod), rstudent(mod))
text(hatvalues(mod), rstudent(mod))
abline(h=2)
abline(h=-2)
abline(v=2*k/n)
abline(v=-2*k/n)
```



Si nota la presenza di diverisi valori anomali e influenti, in particolar modo si identificano le osservazioni 4, 13, 172, 175, 126, ecc. . . , si passa alla rimozione di tali osservazioni secondo le soglie qui sopra indicate.

```
nout<-data[hatvalues(mod)<=2*k/n& abs(rstudent(mod))<2 & cooks.distance(mod)<4/n,]
mod<-lm(y~x2+I(log(x3)),nout)
summary(mod)
```

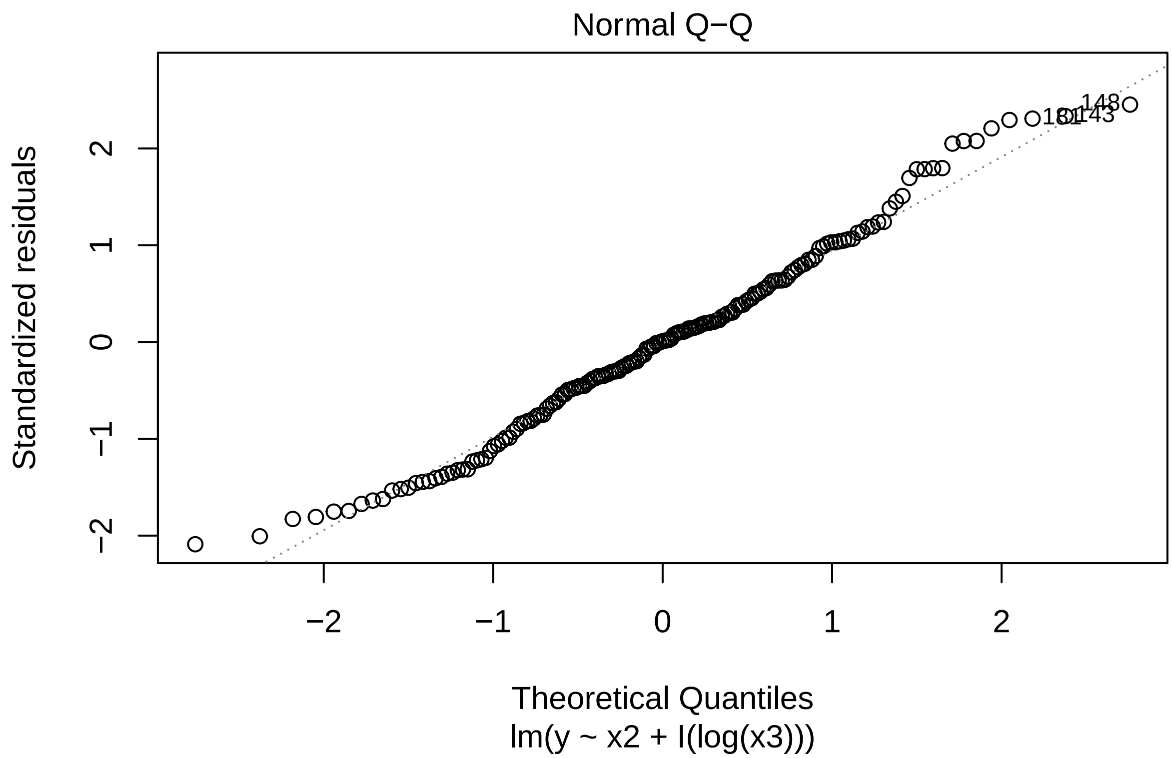
```
##
## Call:
## lm(formula = y ~ x2 + I(log(x3)), data = nout)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0923 -0.9852  0.0121  0.9390  3.6285
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  26.5545     0.5778  45.955  < 2e-16 ***
## x2            0.8466     0.1201   7.048 4.40e-11 ***
## I(log(x3))   -0.7716     0.1156  -6.672 3.45e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.491 on 169 degrees of freedom
## Multiple R-squared:  0.3429, Adjusted R-squared:  0.3351
## F-statistic: 44.09 on 2 and 169 DF,  p-value: 3.904e-16
```

Il modello sembra avere aumentato l'R^2 al 34% grazie alla rimozione dei valori Outlier.

Normalità

Si testa ora come ultima ipotesi la normalità, a tale scopo verranno utilizzati i test di shapiro wilk e Kolmogorov-Smirnov e il seguente grafico QQplot

```
plot(mod, which = 2)
```




```
ols_test_normality(mod$residuals)
```

```
## -----------------------------------------------
##        Test             Statistic       pvalue
## -----------------------------------------------
## Shapiro-Wilk              0.986          0.0843
## Kolmogorov-Smirnov        0.0444         0.8861
## Cramer-von Mises         11.6615         0.0000
## Anderson-Darling          0.4232         0.3164
## -----------------------------------------------
```

Dal grafico i residui risultano approssivativamente normali, anche se i valori sulle code si discotano un po' dai quantili teorici.
Infatti, i test non rifiutano l'ipotesi nulla di normalità.

Autocorrelzione

Come richiesto dall'esercizio si ristima il modello partendo dall'utlimo trovato precedentemente, il modello utilizzerà tutte le osservazioni quindi non vengono omessi dati outlier.
Prima di procedere però si preferisce, poiché si ha a disposizione di ordinare i dati secondo la varibile *time* non inclusa nel modello.

```
data<-data[order(data$time),]
mod<-lm(y~x2+I(log(x3)),data)
summary(mod)
```

```
##
## Call:
## lm(formula = y ~ x2 + I(log(x3)), data = data)
```

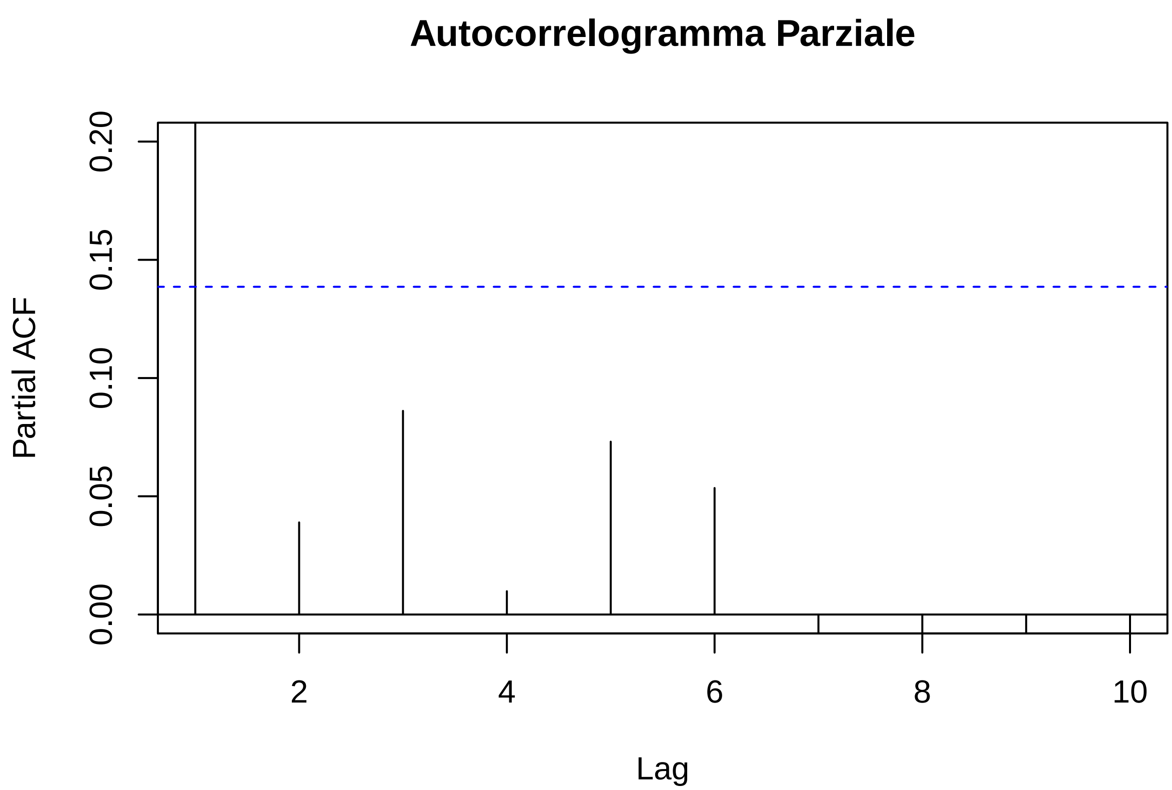
```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6743 -1.1198 -0.0646  0.8120  4.9948
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  26.8348     0.5295  50.682  < 2e-16 ***
## x2            0.7295     0.1148   6.354 1.42e-09 ***
## I(log(x3))   -0.8031     0.1052  -7.634 9.57e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.743 on 197 degrees of freedom
## Multiple R-squared:  0.3182, Adjusted R-squared:  0.3112
## F-statistic: 45.96 on 2 and 197 DF,  p-value: < 2.2e-16
```

Si passa ora verificare l'ipotesi di autocorrelazione tramite un opprtuno grafico (correlogramma Parziale) e un opportuno test, Durbin Watson.

```r
durbinWatsonTest(mod)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1       0.8308835     0.3339711       0
##  Alternative hypothesis: rho != 0
```

```r
autocorr <- pacf(resid(mod),lwd=2, plot = F)
plot(autocorr,xlim = c(1,10), ylim = c(0,0.2), main = "Autocorrelogramma Parziale")
```



Dal grafico si sopsetta un autocorrelazione di primo grado, si rifiuto inoltre l'ipotesi nulla del test di DW che i residui siano incorralti.

Si rislove la situzione mediante il metodo di Chorcane-Orcutt descritto e sviluppato nei passaggi successivi:

```r
data$x5<-log(data$x3)
data$u_hat<-mod$residuals
data<-slide(data, Var='u_hat', TimeVar='time', NewVar='u_hat_lag')

#Calcolo il moello ausliario
aux<-lm(u_hat~u_hat_lag,data)
rho<-aux$coefficients[2]

#Definisco le varibili esplicative ritardate
data<-slide(data, Var='y', TimeVar='time', NewVar='y_lag')
data<-slide(data, Var='x1', TimeVar='time', NewVar='x1_lag')
data<-slide(data, Var='x5', TimeVar='time', NewVar='x5_lag')

#Definisco le varibili trasformate
data$inter<-1-rho
data$y_t<-data$y-rho*data$y_lag
data$x1_t<-data$x1-rho*data$x1_lag
data$x5_t<-data$x5-rho*data$x5_lag

#Modello Trasformato C-O
mod2<-lm(y_t~0+inter+x1_t+x5_t,data)
summary(mod2)
```

```
##
## Call:
## lm(formula = y_t ~ 0 + inter + x1_t + x5_t, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1097 -0.8311  0.0152  0.8684  3.7743
##
## Coefficients:
##       Estimate Std. Error t value Pr(>|t|)
## inter 30.46641    0.73357  41.532  < 2e-16 ***
## x1_t   0.34899    0.04188   8.334 1.35e-14 ***
## x5_t  -2.20610    0.18438 -11.965  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.196 on 196 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.9231, Adjusted R-squared:  0.9219
## F-statistic: 783.8 on 3 and 196 DF,  p-value: < 2.2e-16
```

Il modello trasformato ha aumentato significativamente il suo R^2 al 92%, questo risulta essere sospetto, per mancanza di tempo non si riesce a controllare eventuali errori di implementazione. Rifacendo il test di DW risulta essere ancora rifiutata l'ipotesi di incorrelazione, questo può significare che l'utocorrelazione vada oltre il primo lag.
Quindi bisognerebbe sviluppare il modello precedente oltre il primo ritrado.

In alternativa si potrebbe sviluppare in modello autoregressivo AR con il giusto ordine che tenga appunto presente di tale problema.

PARTE TEORICA

DOMANDA 1

Una volta che sono state identificate tutte le relazioni tra le y_i e le covariate, ci si pongono le domande se esse sono significative dal punto di vista statistico e se il modello si adatta bene ai dati.

Il primo test che si può effettuare è un test normale per i singoli parametri a varianza nota che può essere definito come segue:

$$H_0: \beta_j = 0 \qquad H_1: \beta_j \neq 0$$

Dove si ha che:

$$\widehat{B_j} \sim N(\beta_j, \frac{\sigma^2}{n\sigma_{jj}^{-1}})$$

E la statistica test pari a:

$$Z = \frac{\widehat{\beta_j} - \beta_j}{\sqrt{\frac{\sigma^2}{n\sigma_{jj}^{-1}}}}$$

Si rifiuta H_0 se l'intervallo centrale non comprende il valore β_j ricavato dal campione **Figura1**, se rifiuto l'ipotesi nulla quindi, allora ritengo β_j significativamente diverso da 0, cioè statisticamente significativo.

Tuttavia, poiché nella realtà è inverosimile che la varianza sia nota, si rende necessario passare a un test statistico per singoli parametri con varianza ignota che può essere formulato come segue:

$$H_0: \beta_j = 0 \qquad H_1: \beta_j \neq 0$$

Dove si ha che:

$$\widehat{B_j} \sim N(\beta_j, SE(\widehat{B_j}))$$

E la statistica test pari a:

$$T = \frac{\widehat{\beta_J} - \beta_j}{SE(\widehat{B_J})} = \frac{Z}{\sqrt{\frac{\chi^2_{n-k-1}}{n-k-1}}} \sim t_{n-k-1}(sotto\ H_0)$$

Come il precedente test se si rifiuta H_0 **Figura2** allora β_j è statisticamente significativo.

Per valutare la bontà del modello invece, è necessario fare un test sull'R^2 del tipo:

$$H_0: R^2 = 0 \qquad H_1: R^2 \neq 0$$

Si considerano le seguenti distribuzioni sotto H_0:

- $\frac{SSR}{\sigma^2}$ è $determinazione\ di\ \chi^2_{n-k-1}$
- $\frac{MSS}{\sigma^2}$ è $determinazione\ di\ \chi^2_k$
- $\frac{SST}{\sigma^2}$ è $determinazione\ di\ \chi^2_{n-1}$

Da ciò consegue che la statistica test che si prende in considerazione, è sotto H_0, una F di Snedecor:

$$F = \frac{MSS/k}{SSR/(n-k-1)} \sim F_{k,n-k-1}(sotto\ H_0)$$

In questo caso se si rifiuta H_0 **Figura3,** allora si accetta il modello.

Infine, è possibile effettuare un test F su più parametri in questo caso si andrà a valutare l'ipotesi nulla $H_0: \beta_1, \dots, \beta_j, \dots, \beta_q = 0$ (ipotesi alternativa, almeno un parametro diverso da 0) sotto cui si avrà una statistica test del tipo:

$$F = \frac{\frac{MSS - MSS_1}{k-q}}{SSR/(n-p-1)} \sim F_{k-q,n-k-1}(sotto\ H_0)$$

Dove MSS_1 è la devianza spiegata del modello che considera sole le ultime *k-q* variabili con *q<k*, mentre *MSS* è la devianza spiegata del modello completo, tale per cui *MSS>*MSS_1 e SSR_1*>SSR*.

Oltre ai test d'ipotesi per verificare la significatività dei parametri è possibile anche calcolare l'intervallo di confidenza di un parametro così da misurarne il possibile impatto.

Tra test di ipotesi e intervalli di confidenza sussiste una relazione, infatti, nel test d'ipotesi respingiamo H_0 se il p-value è piccolo, dunque, la probabilità che il parametro nel campione ha un valore estremo anche se H_0 è vero, è piccola, mentre in un intervallo di confidenza al 95%, ad esempio, ci fornisce un intervallo di valori plausibili per il coefficiente angolare dove il 95% degli intervalli comprende il valore del coefficiente angolare stesso, in questo caso se il parametro campionario non è compreso nel 95% della distribuzione allora riusciamo a respingere H_0.

FIGURA 1

SOTTO H_0 LA DISTRIBUZIONE DI β_J RISULTA ESSERE:

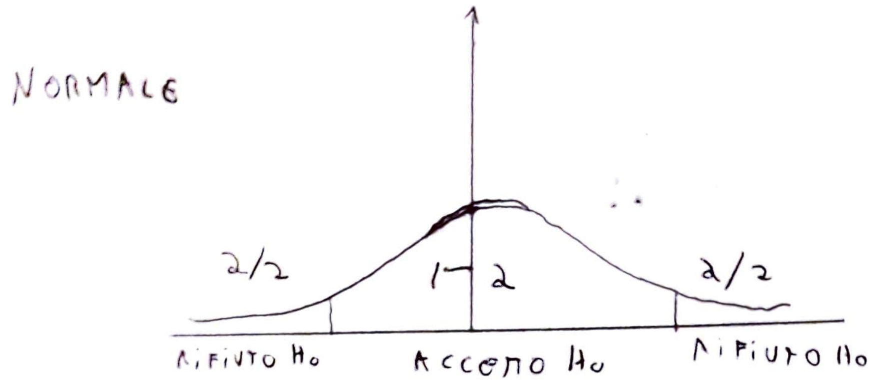


FIGURA 2

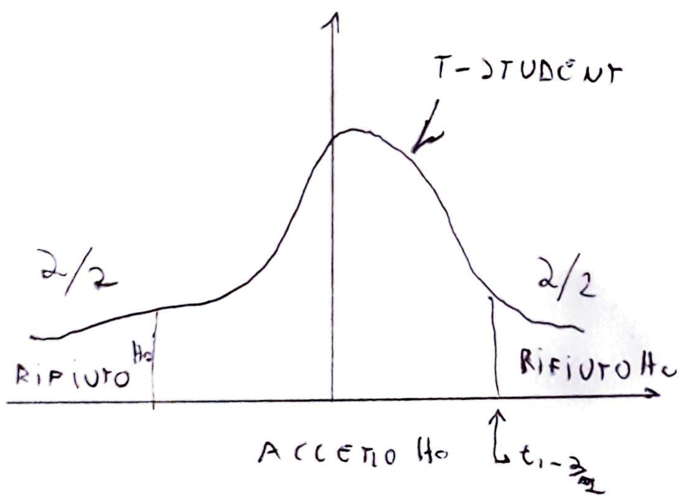
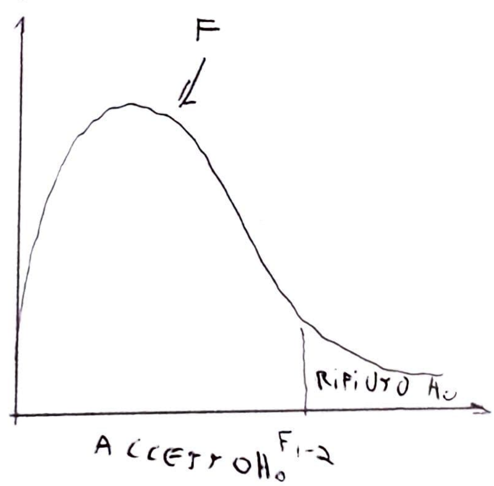


FIGURA 3



DOMANDA 2

Nel Modello SURE si hanno regressori diversi per ogni equazione all'interno dell'insieme complessivo dei regressori per l'insieme delle equazioni del modello. Questo permette di risolvere i problemi di una numerosità diversa delle osservazioni delle diverse equazioni, dove si indica con $\sum_j r_j$ la somma di tutti i regressori nelle diverse equazioni e $n = \sum_j n_j$ la numerosità delle osservazioni nel complesso.

Il modello SURE può essere quindi formulato come:

$$y_1 = \beta_{10} + \beta_{11}z_1 + \cdots + \beta_{1r}z_r + \varepsilon_1$$
$$y_2 = \beta_{20} + \beta_{22}z_2 + \cdots + \beta_{2r}z_r + \varepsilon_2$$
$$\vdots$$
$$y_m = \beta_{m0} + \beta_{m1}z_1 + \cdots + \beta_{mr}z_r + \varepsilon_m$$

Con struttura di matrice di varianze e covarianze pari a:

$$E(\varepsilon\varepsilon') = \textstyle\sum_\varepsilon \ (FIGURA\ 4)$$

Dove:

1. Gli errori sono omoschedastici e incorrelati all'interno delle stesse equazioni.
2. Gli errori sono eteroschedastici tra equazioni diverse.
3. Gli errori sono correlati per lo stesso individuo e incorrelati tra individui diversi fra diverse equazioni.

Per ottenere le soluzioni dato il diverso n e la diversa tipologia delle variabili esplicative in ogni equazione si esprimono Y, B, E in termini di supervettori e Z come matrice diagonale a blocchi, dove m sono le equazioni e r_j le variabili esplicative Z_j dell'equazione j-esima. Quindi, il modello multivariato può essere riscritto nel seguente modo:

$$y^* = \beta^* Z^* + \varepsilon^*$$

Con:

$$\hat{\beta}^* = y^* \Sigma_\varepsilon^{-1} Z^{*\prime} (Z^* \Sigma_\varepsilon^{-1} Z^{*\prime})^{-1}$$

Si nota in particolare come a differenza del GLS:

- Il modello si caratterizza dalla presenza delle variabili esplicative Z_A, Z_B, Z_C che variano da equazione ed equazione.
- Gli errori sono omoschedastici e incorrelati nella stessa equazione.
- Gli errori sono eteroschedastici, correlati per lo stesso individuo.
- Gli errori sono incorrelati fra individui diversi fra diverse equazioni.

Il modello viene infine scelto rispetto al modello multivariato classico e generalizzato come:

a. Scelta di variabili esplicative che sono diverse da equazione ad equazione.
b. Se si omettono delle variabili cambiano tutti i coefficienti di regressione.
c. Cercare il giusto trade-off tra adattamento ai dati e parsimonia equazione per equazione. Se la diminuzione dell'R^2 e di F risulta minima, allora si opta per eliminare quella variabile dalla specifica equazione.
d. Bisogna verificare l'omoschedasticità e la correlazione sia all'interno che fra equazioni. Si possono costruire così modelli multivariati con

covariate diverse in ogni equazione utilizzando diversi tipi di matrici di varianze covarianze degli errori.

e. Scelta tra modello lineare generalizzato e Sure: in questo caso si possono scartare le variabili non significative.

FIGURA 4

$$E(\varepsilon'\varepsilon) = \Sigma_\varepsilon = \begin{bmatrix} \sigma_1^2 I_N & \sigma_{12}^2 I_N & \cdots & \sigma_{1M}^2 I_N \\ \sigma_{12}^2 I_N & \ddots & & \vdots \\ \vdots & & \ddots & \\ \sigma_{1M}^2 I_N & \cdots & \cdots & \sigma_M^2 I_N \end{bmatrix}$$