Assingment 8

Chitresh Kumar

if(!require("pacman")) install.packages("pacman")  
pacman::p\_load(tidyverse, reshape, gplots, ggmap, RStata,haven,  
 data.table,margins,pastecs,MASS,lmtest,broom,car,stargazer,sandwich,knitr,dplyr)  
search()  
theme\_set(theme\_classic())

df<-read\_dta('ivreg2.dta')  
head(df)

## # A tibble: 6 x 4  
## x y z1 z2  
## <dbl> <dbl> <dbl> <dbl>  
## 1 -0.965 1.16 0.438 -1.17   
## 2 -2.33 1.53 -2.51 -1.43   
## 3 0.472 4.78 -0.449 -0.0394  
## 4 -3.43 -3.58 -0.848 0.530   
## 5 0.138 2.14 0.729 0.0836  
## 6 -1.53 1.03 -0.638 -0.603

**PART b**

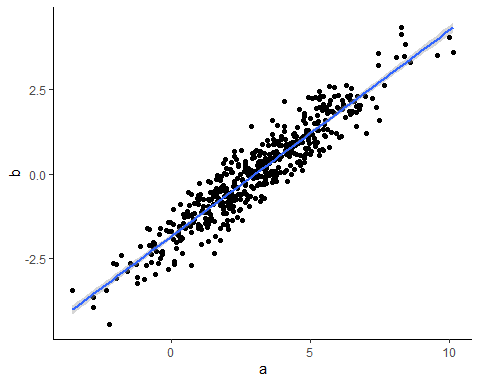
a=df$y  
b=df$x  
   
e=a-3-b

cor(b,e)

## [1] 0.65136

**PART c**

e\_y=3+b  
ggplot(df,aes(x=a,y=b))+  
 geom\_point()+  
 geom\_smooth(method="lm")



**PART D**

lmdata\_1<- df %>% slice(1:10)  
lm\_1<-lm(y~x,data=lmdata\_1)  
summary(lm\_1)

##   
## Call:  
## lm(formula = y ~ x, data = lmdata\_1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.6450 -0.6888 -0.2390 0.4484 1.9556   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.7775 0.3608 7.698 5.76e-05 \*\*\*  
## x 1.3722 0.1727 7.945 4.59e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.136 on 8 degrees of freedom  
## Multiple R-squared: 0.8875, Adjusted R-squared: 0.8735   
## F-statistic: 63.12 on 1 and 8 DF, p-value: 4.589e-05

lmdata\_2<- df %>% slice(1:20)  
lm\_2<-lm(y~x,data=lmdata\_2)  
summary(lm\_2)

##   
## Call:  
## lm(formula = y ~ x, data = lmdata\_2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.83171 -0.52577 0.08304 0.45379 1.75205   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.0169 0.2036 14.81 1.59e-11 \*\*\*  
## x 1.3876 0.1211 11.46 1.05e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.9056 on 18 degrees of freedom  
## Multiple R-squared: 0.8795, Adjusted R-squared: 0.8728   
## F-statistic: 131.4 on 1 and 18 DF, p-value: 1.053e-09

lmdata\_3<- df %>% slice(1:100)  
lm\_3<-lm(y~x,data=lmdata\_3)  
summary(lm\_3)

##   
## Call:  
## lm(formula = y ~ x, data = lmdata\_3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.1199 -0.5289 0.0271 0.5255 1.7940   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.00783 0.07872 38.21 <2e-16 \*\*\*  
## x 1.40164 0.05330 26.30 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7864 on 98 degrees of freedom  
## Multiple R-squared: 0.8759, Adjusted R-squared: 0.8746   
## F-statistic: 691.5 on 1 and 98 DF, p-value: < 2.2e-16

lmdata\_4<- df %>% slice(1:500)  
lm\_4<-lm(y~x,data=lmdata\_4)  
summary(lm\_4)

##   
## Call:  
## lm(formula = y ~ x, data = lmdata\_4)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.20345 -0.51588 -0.01086 0.52412 2.26606   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.01825 0.03410 88.5 <2e-16 \*\*\*  
## x 1.45352 0.02367 61.4 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7624 on 498 degrees of freedom  
## Multiple R-squared: 0.8833, Adjusted R-squared: 0.8831   
## F-statistic: 3770 on 1 and 498 DF, p-value: < 2.2e-16

**PART e**

df$e<-e  
cor(df)

## x y z1 z2 e  
## x 1.0000000 0.9398447 0.620821104 0.28948601 0.651359982  
## y 0.9398447 1.0000000 0.399870154 0.19965601 0.871374239  
## z1 0.6208211 0.3998702 1.000000000 -0.01530765 -0.003447192  
## z2 0.2894860 0.1996560 -0.015307651 1.00000000 0.027708992  
## e 0.6513600 0.8713742 -0.003447192 0.02770899 1.000000000

**PART f**

lmdata\_1<- df %>% slice(1:10)  
lm\_i1<-lm(x~z1,data=lmdata\_1)  
x\_fit<-fitted(lm\_i1)  
sls\_1<-lm(y~x\_fit,data=lmdata\_1)  
summary(sls\_1)

##   
## Call:  
## lm(formula = y ~ x\_fit, data = lmdata\_1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.8950 -1.4894 -0.5804 2.4329 3.0017   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.7144 0.8705 3.118 0.0143 \*  
## x\_fit 1.0640 0.5142 2.069 0.0723 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.733 on 8 degrees of freedom  
## Multiple R-squared: 0.3486, Adjusted R-squared: 0.2672   
## F-statistic: 4.282 on 1 and 8 DF, p-value: 0.07231

lmdata\_2<- df %>% slice(1:20)  
lm\_i2<-lm(x~z1,data=lmdata\_2)  
x\_fit2<-fitted(lm\_i2)  
sls\_2<-lm(y~x\_fit2,data=lmdata\_2)  
summary(sls\_2)

##   
## Call:  
## lm(formula = y ~ x\_fit2, data = lmdata\_2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.7627 -1.2284 0.2782 1.3325 3.6378   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.0810 0.5024 6.132 8.61e-06 \*\*\*  
## x\_fit2 1.0263 0.3952 2.597 0.0182 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.225 on 18 degrees of freedom  
## Multiple R-squared: 0.2726, Adjusted R-squared: 0.2322   
## F-statistic: 6.745 on 1 and 18 DF, p-value: 0.01821

lmdata\_3<- df %>% slice(1:100)  
lm\_i3<-lm(x~z1,data=lmdata\_3)  
x\_fit3<-fitted(lm\_i3)  
sls\_3<-lm(y~x\_fit3,data=lmdata\_3)  
summary(sls\_3)

##   
## Call:  
## lm(formula = y ~ x\_fit3, data = lmdata\_3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.8057 -1.4028 0.2217 1.5811 4.0127   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.9771 0.2058 14.463 < 2e-16 \*\*\*  
## x\_fit3 0.9363 0.2217 4.223 5.41e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.053 on 98 degrees of freedom  
## Multiple R-squared: 0.1539, Adjusted R-squared: 0.1453   
## F-statistic: 17.83 on 1 and 98 DF, p-value: 5.408e-05

lmdata\_4<- df %>% slice(1:500)  
lm\_i4<-lm(x~z1,data=lmdata\_4)  
x\_fit4<-fitted(lm\_i4)  
sls\_4<-lm(y~x\_fit4,data=lmdata\_4)  
summary(sls\_4)

##   
## Call:  
## lm(formula = y ~ x\_fit4, data = lmdata\_4)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.8886 -1.4938 0.0333 1.3726 7.6623   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.03150 0.09153 33.119 <2e-16 \*\*\*  
## x\_fit4 0.99613 0.10232 9.736 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.046 on 498 degrees of freedom  
## Multiple R-squared: 0.1599, Adjusted R-squared: 0.1582   
## F-statistic: 94.78 on 1 and 498 DF, p-value: < 2.2e-16

**PART g**

lmdata\_g1<- df %>% slice(1:10)  
lm\_g1<-lm(x~z2,data=lmdata\_g1)  
x\_fitg1<-fitted(lm\_g1)  
sls\_g1<-lm(y~x\_fitg1,data=lmdata\_g1)  
summary(sls\_g1)

##   
## Call:  
## lm(formula = y ~ x\_fitg1, data = lmdata\_g1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.9100 -1.5470 -0.7599 2.0799 5.8201   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 1.892 3.697 0.512 0.623  
## x\_fitg1 -2.950 17.282 -0.171 0.869  
##   
## Residual standard error: 3.38 on 8 degrees of freedom  
## Multiple R-squared: 0.00363, Adjusted R-squared: -0.1209   
## F-statistic: 0.02915 on 1 and 8 DF, p-value: 0.8687

lmdata\_g2<- df %>% slice(1:20)  
lm\_g2<-lm(x~z2,data=lmdata\_g2)  
x\_fitg2<-fitted(lm\_g2)  
sls\_g2<-lm(y~x\_fitg2,data=lmdata\_g2)  
summary(sls\_g2)

##   
## Call:  
## lm(formula = y ~ x\_fitg2, data = lmdata\_g2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.8580 -1.4576 0.2981 1.5408 5.0120   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.2433 0.7499 4.325 0.000408 \*\*\*  
## x\_fitg2 0.1110 2.6571 0.042 0.967140   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.609 on 18 degrees of freedom  
## Multiple R-squared: 9.693e-05, Adjusted R-squared: -0.05545   
## F-statistic: 0.001745 on 1 and 18 DF, p-value: 0.9671

lmdata\_g3<- df %>% slice(1:100)  
lm\_g3<-lm(x~z2,data=lmdata\_g3)  
x\_fitg3<-fitted(lm\_g3)  
sls\_g3<-lm(y~x\_fitg3,data=lmdata\_g3)  
summary(sls\_g3)

##   
## Call:  
## lm(formula = y ~ x\_fitg3, data = lmdata\_g3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.8871 -1.3593 -0.0861 1.5005 5.3739   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.9902 0.2137 13.996 < 2e-16 \*\*\*  
## x\_fitg3 1.1349 0.3542 3.204 0.00183 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.124 on 98 degrees of freedom  
## Multiple R-squared: 0.09483, Adjusted R-squared: 0.08559   
## F-statistic: 10.27 on 1 and 98 DF, p-value: 0.001828

lmdata\_g4<- df %>% slice(1:500)  
lm\_g4<-lm(x~z2,data=lmdata\_g4)  
x\_fitg4<-fitted(lm\_g4)  
sls\_g4<-lm(y~x\_fitg4,data=lmdata\_g4)  
summary(sls\_g4)

##   
## Call:  
## lm(formula = y ~ x\_fitg4, data = lmdata\_g4)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.8896 -1.3373 -0.1368 1.4042 7.5446   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.02946 0.09804 30.901 < 2e-16 \*\*\*  
## x\_fitg4 1.06665 0.23458 4.547 6.84e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.187 on 498 degrees of freedom  
## Multiple R-squared: 0.03986, Adjusted R-squared: 0.03793   
## F-statistic: 20.68 on 1 and 498 DF, p-value: 6.839e-06

**PART h**

data\_1<- df %>% slice(1:10)  
comb\_1<-lm(x~z2+z1,data=data\_1)  
fit\_1<-fitted(comb\_1)  
sls\_lm1<-lm(y~fit\_1,data=data\_1)  
summary(sls\_lm1)

##   
## Call:  
## lm(formula = y ~ fit\_1, data = data\_1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.0639 -1.3528 -0.4129 2.4800 2.9570   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.7114 0.8746 3.100 0.0147 \*  
## fit\_1 1.0491 0.5140 2.041 0.0755 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.745 on 8 degrees of freedom  
## Multiple R-squared: 0.3425, Adjusted R-squared: 0.2603   
## F-statistic: 4.166 on 1 and 8 DF, p-value: 0.07553

data\_2<- df %>% slice(1:20)  
comb\_2<-lm(x~z2+z1,data=data\_2)  
fit\_2<-fitted(comb\_2)  
sls\_lm2<-lm(y~fit\_2,data=data\_2)  
summary(sls\_lm2)

##   
## Call:  
## lm(formula = y ~ fit\_2, data = data\_2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.9387 -0.9457 0.3604 1.4182 3.6502   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.0852 0.5045 6.115 8.92e-06 \*\*\*  
## fit\_2 1.0026 0.3924 2.555 0.0199 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.235 on 18 degrees of freedom  
## Multiple R-squared: 0.2661, Adjusted R-squared: 0.2254   
## F-statistic: 6.528 on 1 and 18 DF, p-value: 0.01989

data\_3<- df %>% slice(1:100)  
comb\_3<-lm(x~z2+z1,data=data\_3)  
fit\_3<-fitted(comb\_3)  
sls\_lm3<-lm(y~fit\_3,data=data\_3)  
summary(sls\_lm3)

##   
## Call:  
## lm(formula = y ~ fit\_3, data = data\_3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.1040 -1.4575 0.2193 1.5744 3.9015   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.9808 0.1962 15.191 < 2e-16 \*\*\*  
## fit\_3 0.9921 0.1833 5.413 4.41e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.958 on 98 degrees of freedom  
## Multiple R-squared: 0.2302, Adjusted R-squared: 0.2223   
## F-statistic: 29.3 on 1 and 98 DF, p-value: 4.407e-07

data\_4<- df %>% slice(1:500)  
comb\_4<-lm(x~z2+z1,data=data\_4)  
fit\_4<-fitted(comb\_4)  
sls\_lm4<-lm(y~fit\_4,data=data\_4)  
summary(sls\_lm4)

##   
## Call:  
## lm(formula = y ~ fit\_4, data = data\_4)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.1205 -1.3817 -0.0526 1.2986 7.3349   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.03113 0.08920 33.98 <2e-16 \*\*\*  
## fit\_4 1.00899 0.08984 11.23 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.994 on 498 degrees of freedom  
## Multiple R-squared: 0.2021, Adjusted R-squared: 0.2005   
## F-statistic: 126.1 on 1 and 498 DF, p-value: < 2.2e-16

A close up of text on a whiteboard

Description automatically generated

A close up of text on a whiteboard

Description automatically generated

A close up of text on a white background

Description automatically generated

A close up of text on a whiteboard

Description automatically generated

A close up of text on a whiteboard

Description automatically generated

A close up of text on a whiteboard

Description automatically generated

A close up of text on a whiteboard

Description automatically generated

A close up of text on a white background

Description automatically generated

A close up of text on a white background

Description automatically generated