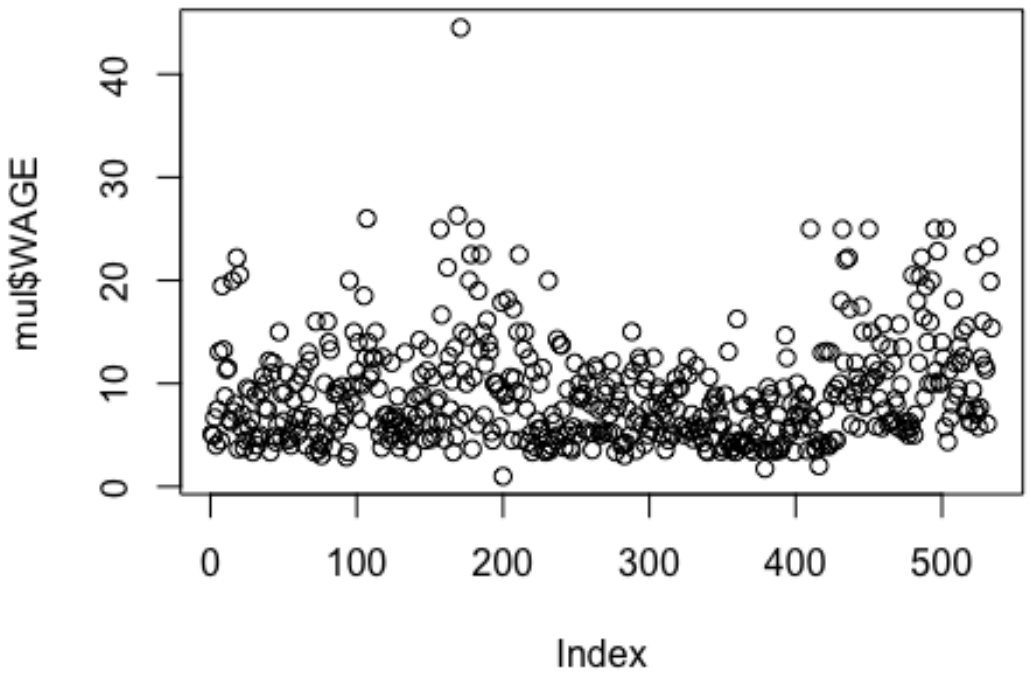
Multicollinearity _Regression

Indika Debnath

07/12/2022

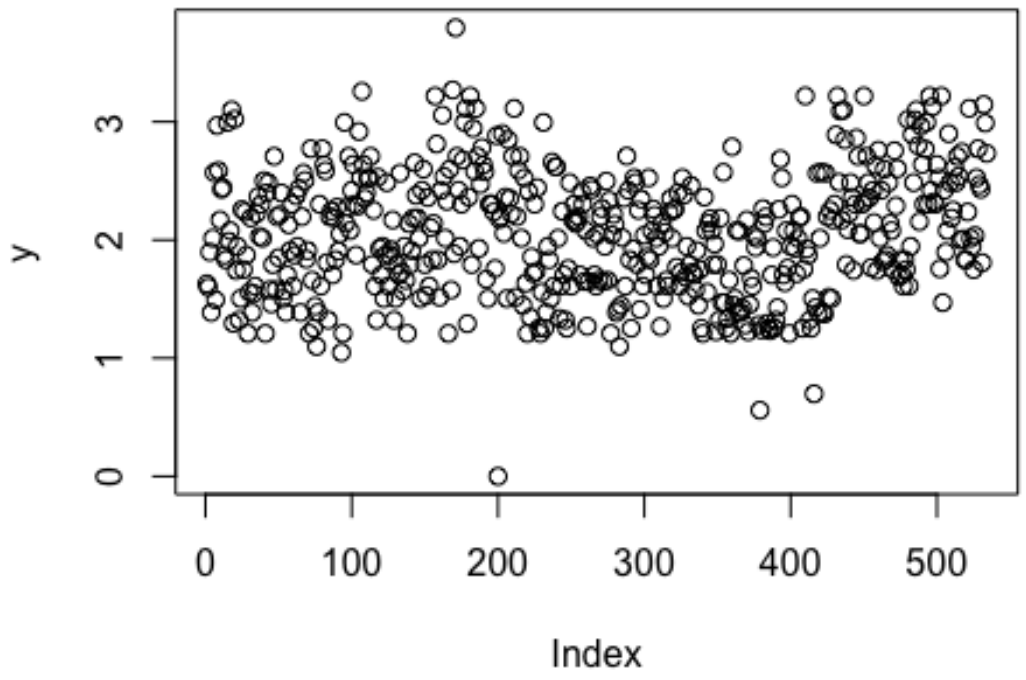
```
#The Current Population Survey (CPS) is used to supplement census information
between census years. These data consist of a random sample of 534 persons
from the CPS, with information on wages and other characteristics of the
workers, including sex, number of years of education, years of work
experience, occupational status, region of residence and union membership. We
wish to determine whether wages are related to these characteristics.

library(readxl)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

mul <-
read_excel("/Users/indikadebnath/Desktop/Stat_Project/Regression_Project/mult
icol.xlsx")
#View(mul)

fit <-
lm(WAGE~OCCUPATION+SECTOR+UNION++EDUCATION+EXPERIENCE+AGE+SEX+MARR+RACE+SOUTH
,data=mul)
summary(fit)

##
## Call:
## lm(formula = WAGE ~ OCCUPATION + SECTOR + UNION + +EDUCATION +
##     EXPERIENCE + AGE + SEX + MARR + RACE + SOUTH, data = mul)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.190 -2.616 -0.684  1.873 37.301
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.0402     6.8790  -0.297  0.76690
```

```
## OCCUPATION   -0.1527     0.1312  -1.165  0.24475
## SECTOR        0.7190     0.3876   1.855  0.06414 .
## UNION         1.5168     0.5250   2.889  0.00403 **
## EDUCATION     1.3262     1.1082   1.197  0.23194
## EXPERIENCE    0.5246     1.1086   0.473  0.63625
## AGE          -0.4282     1.1079  -0.386  0.69931
## SEX          -2.1443     0.3993  -5.370 1.19e-07 ***
## MARR          0.4252     0.4195   1.013  0.31131
## RACE          0.4786     0.2855   1.676  0.09426 .
## SOUTH        -0.6975     0.4285  -1.628  0.10414
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.401 on 523 degrees of freedom
## Multiple R-squared:  0.2805, Adjusted R-squared:  0.2667
## F-statistic: 20.39 on 10 and 523 DF,  p-value: < 2.2e-16

# as we can standard error is high 4.401 and R square is 0.2805

#checking the distribution of dependent variable wage
plot(mul$WAGE, type='p')
```




```
# it is scattered and not linear in nature
```

```r
# to make wages as linear and improve the R sqaure and standard error we can
take log of wages
 y <-log(mul$WAGE)
plot(y, type='p')
```




```r
# as we can see plot of log(wages) are more linear in nature

# we can see after taking log of dependent variable The R square is improved
0.3185 and standard error got reduced significantly 0.4398
library(readxl)
library(dplyr)
mul <-
read_excel("/Users/indikadebnath/Desktop/Stat_Project/Regression_Project/mult
icol.xlsx")
#View(mul)

fit1 <-
lm(log(WAGE)~OCCUPATION+SECTOR+UNION++EDUCATION+EXPERIENCE+AGE+SEX+MARR+RACE+
SOUTH,data=mul)
summary(fit1)

##
## Call:
## lm(formula = log(WAGE) ~ OCCUPATION + SECTOR + UNION + +EDUCATION +
```

```
##     EXPERIENCE + AGE + SEX + MARR + RACE + SOUTH, data = mul)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.16246 -0.29163 -0.00469  0.29981  1.98248
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.078596   0.687514   1.569 0.117291
## OCCUPATION  -0.007417   0.013109  -0.566 0.571761
## SECTOR       0.091458   0.038736   2.361 0.018589 *
## UNION        0.200483   0.052475   3.821 0.000149 ***
## EDUCATION    0.179366   0.110756   1.619 0.105949
## EXPERIENCE   0.095822   0.110799   0.865 0.387531
## AGE         -0.085444   0.110730  -0.772 0.440671
## SEX         -0.221997   0.039907  -5.563 4.24e-08 ***
## MARR         0.076611   0.041931   1.827 0.068259 .
## RACE         0.050406   0.028531   1.767 0.077865 .
## SOUTH       -0.102360   0.042823  -2.390 0.017187 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4398 on 523 degrees of freedom
## Multiple R-squared:  0.3185, Adjusted R-squared:  0.3054
## F-statistic: 24.44 on 10 and 523 DF,  p-value: < 2.2e-16
```

 # R-squared value of 0.31 is not bad for a cross sectional data of 534 observations. The F-value is highly significant implying that all the explanatory variables together significantly explain the log of wages.

#However, coming to the individual regression coefficients, it is seen that as many as four variables (occupation, education, experience, age) are not statistically significant and two (marital status and south) are significant only at 10 % level of significance.

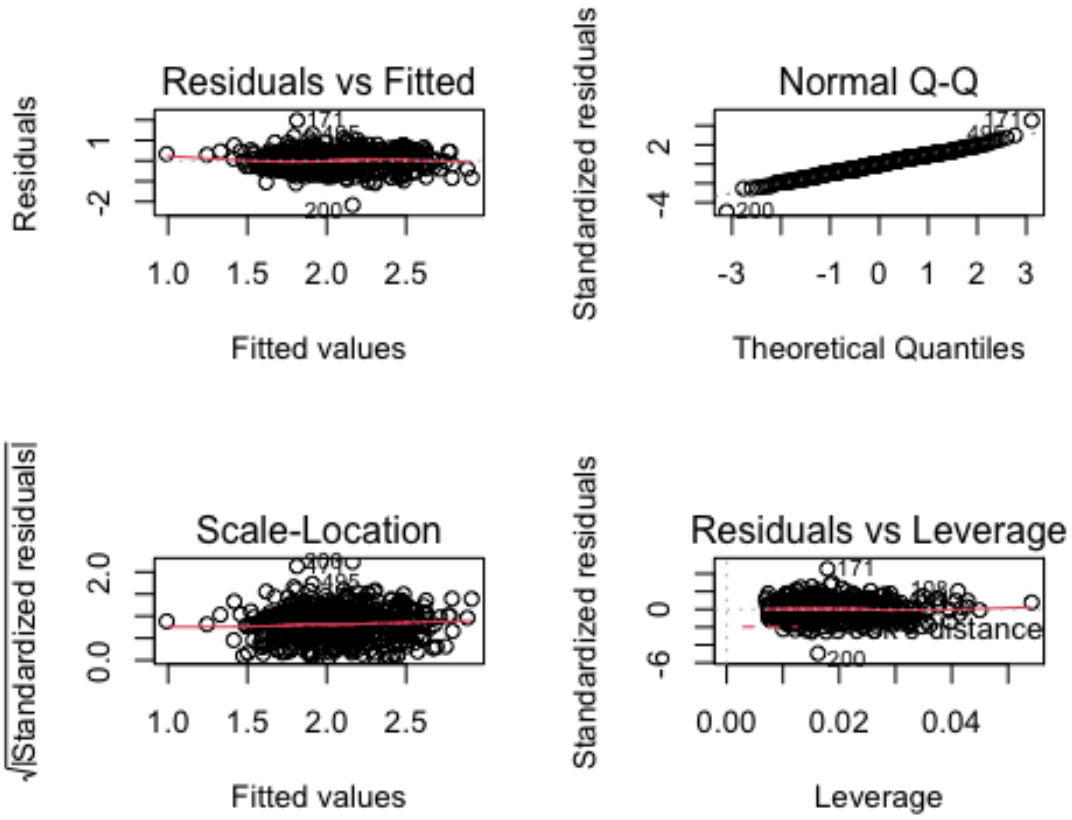
#here H0:b1=b2=b3=b4=b5=b6=b7=b8=b9=b10 #independent variables are not influencing dependent variable

#HA:b1!= or b2!= or b3!= or b4!= or b5!= or b6!= or b7!=or b8!= or b9!= or b10!
#independent variables are not influencing dependent variable

#we can see P value is less than 0.05 (2.2e-16) that means at least 1 of the independent variables are influencing dependent variable, we can reject our null hypothesis and be sure that the influence of independent variable on dependent variable not just happens by random chance

#When heteroscedasticity is present in a regression analysis, the results of the analysis become hard to trust. Specifically, heteroscedasticity increases the variance of the regression coefficient estimates, but the regression

```
model doesn't pick up on this.

#This makes it much more likely for a regression model to declare that a term
in the model is statistically significant, when in fact it is not

#Further we can plot the model diagnostic checking for other problems such as
normality of error term, heteroscedasticity etc.

#The par() function allows to set parameters to the plot. The mfrow()
parameter allows to split the screen in several panels. Subsequent charts
will be drawn in panels. We have to provide a vector of length 2 to mfrow() :
number of rows and number of columns.

# mfcol() does the same job but draws figure by columns instead of by row.

par(mfrow=c(2,2))
plot(fit1)#no heterroskedastcty

## Warning: not plotting observations with leverage one:
##   444
```



```
#Thus, the diagnostic plot is also look fair.
#So, possibly the multicollinearity problem is the reason for not getting
many insignificant regression coefficients.

# we can check heterroskedastcty with Breusch-Pagan test also
library(lmtest)

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

lmtest::bptest(fit1)

##
##  studentized Breusch-Pagan test
##
## data:  fit1
## BP = 15.813, df = 10, p-value = 0.1051

#H0: Homoscedasticity is present (the residuals are distributed with equal
variance)
#HA:Heteroscedasticity is present (the residuals are not distributed with
equal variance)

# as the pvalue >0.05 we can't reject null hypo thesis so it concludes
Heteroscedasticity is not present in the data

#here we are trying to find which of the independent variables are actually
statistically significant

#A non significant varibale become significant when there is
heterscedasticity or multicollienartity as we are now sure there is no
heterscedasticity So, possibly the multicollinearity problem is the reason
for not getting many insignificant regression coefficients


#checking multicollinearity#
library(car)

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode
```

```
vif(fit1)
```

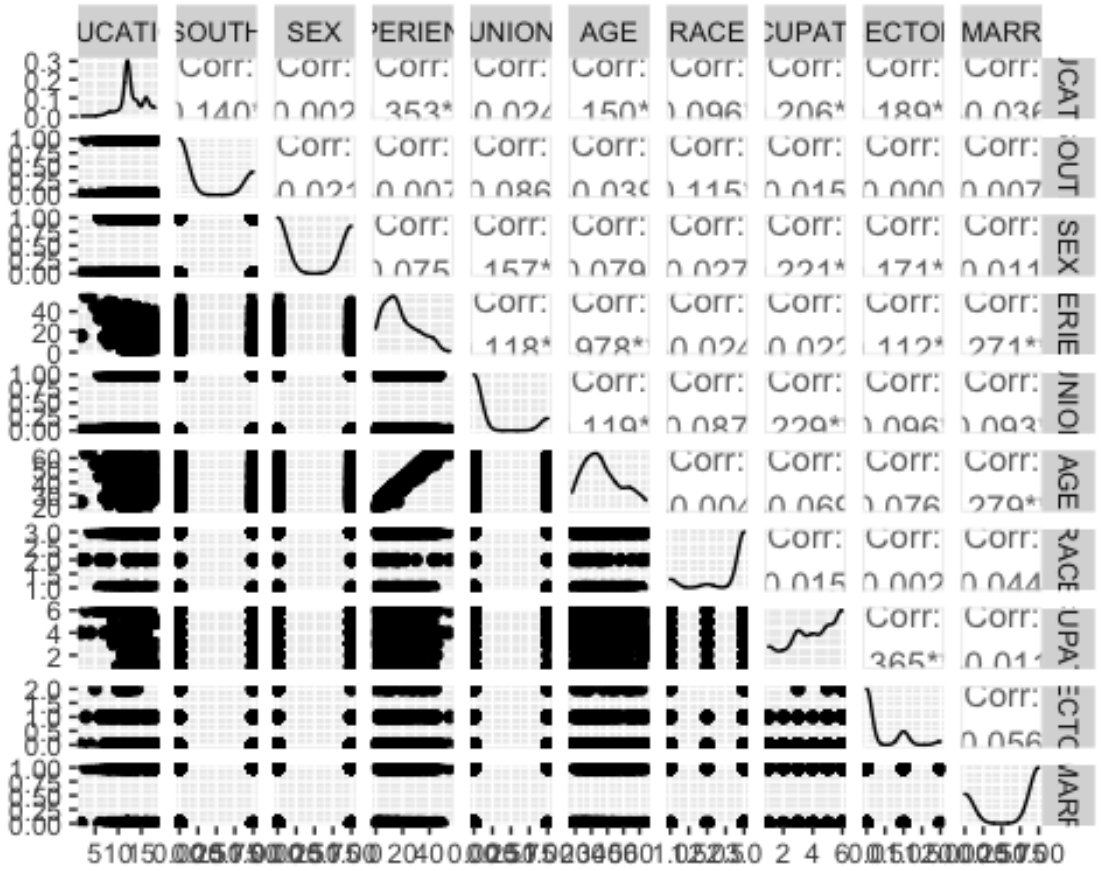
```
##  OCCUPATION      SECTOR       UNION   EDUCATION  EXPERIENCE         AGE
##    1.298232    1.198670    1.120861  231.195580 5184.093895 4645.664977
##         SEX        MARR        RACE       SOUTH
##    1.091634    1.096130    1.037138    1.046828
```

```
# we can see from the VIF result EDUCATION,EXPERIENCE,AGE have VIF>5 which
indicates there is high multicollinearity between these independent variables

#we can use the corr command to create a correlation matrix to view the
correlation coefficients between each of the variables in the model, which
can help us identify which variables might be highly correlated with each
other and could be causing the problem of multicollinearity:

#further diagnosis of the problem, let us first look at the pair-wise
correlation among the explanatory variables
library(GGally)
```

```
## Loading required package: ggplot2
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
library(ggplot2)
x <- mul[,2:11]
ggpairs(x)
```




```
#The correlation matrix shows that the pair-wise correlation among all the
explanatory variables are not very high, except for the pair age -
experience.

#The high correlation between age and experience might be the root cause of
multicollinearity.

#Again by looking at the partial correlation coefficient matrix among the
variables, it is also clear that the partial correlation between experience -
education, age - education and age - experience are quite high.

#install.packages("corpcor")
library(corpcor)
cor2pcor((cov(x)))
```

```
##               [,1]         [,2]         [,3]        [,4]         [,5]
##  [1,]  1.000000000 -0.031750193  0.051510483 -0.99756187 -0.007479144
##  [2,] -0.031750193  1.000000000 -0.030152499 -0.02231360 -0.097548621
##  [3,]  0.051510483 -0.030152499  1.000000000  0.05497703 -0.120087577
##  [4,] -0.997561873 -0.022313605  0.054977034  1.00000000 -0.010244447
##  [5,] -0.007479144 -0.097548621 -0.120087577 -0.01024445  1.000000000
##  [6,]  0.997261601  0.021525073 -0.053697851  0.99987574  0.012238897
##  [7,]  0.017230877 -0.111197596  0.020017315  0.01088849 -0.107706183
```

```
##  [8,]  0.029436911  0.008430595 -0.142750864  0.04205856  0.212996388
##  [9,] -0.021253493 -0.021518760 -0.112146760 -0.01326166 -0.013531482
## [10,] -0.040302967  0.030418218  0.004163264 -0.04097664  0.068918496
##              [,6]         [,7]         [,8]         [,9]        [,10]
##  [1,]  0.99726160  0.017230877  0.029436911 -0.021253493 -0.040302967
##  [2,]  0.02152507 -0.111197596  0.008430595 -0.021518760  0.030418218
##  [3,] -0.05369785  0.020017315 -0.142750864 -0.112146760  0.004163264
##  [4,]  0.99987574  0.010888486  0.042058560 -0.013261665 -0.040976643
##  [5,]  0.01223890 -0.107706183  0.212996388 -0.013531482  0.068918496
##  [6,]  1.00000000 -0.010803310 -0.044140293  0.014565751  0.045090327
##  [7,] -0.01080331  1.000000000  0.057539374  0.006412099  0.055645964
##  [8,] -0.04414029  0.057539374  1.000000000  0.314746868 -0.018580965
##  [9,]  0.01456575  0.006412099  0.314746868  1.000000000  0.036495494
## [10,]  0.04509033  0.055645964 -0.018580965  0.036495494  1.000000000
```

```
#The 'mctest' package in R provides the Farrar-Glauber test and other
relevant tests for multicollinearity.

library(mctest)
fit <-
lm(log(WAGE)~OCCUPATION+SECTOR+UNION++EDUCATION+EXPERIENCE+AGE+SEX+MARR+RACE+
SOUTH,data=mul)

omcdiag(fit)

##
## Call:
## omcdiag(mod = fit)
##
##
## Overall Multicollinearity Diagnostics
##
##                        MC Results detection
## Determinant |X'X|:         0.0001         1
## Farrar Chi-Square:      4818.3895         1
## Red Indicator:             0.1983         0
## Sum of Lambda Inverse: 10068.8439         1
## Theil's Method:            0.8845         1
## Condition Number:        739.7337         1
##
## 1 --> COLLINEARITY is detected by the test
## 0 --> COLLINEARITY is not detected by the test

 #The value of the standardized determinant is found to be 0.0001 which is
very small. The calculated value of the Chi-square test statistic is found to
be 4818.3895 and it is highly significant thereby implying the presence of
multicollinearity in the model specification.
```

```
#now individual level  multicolleinearity check

imcdiag(fit)

##
## Call:
## imcdiag(mod = fit)
##
##
## All Individual Multicollinearity Diagnostics Result
##
##                  VIF    TOL          Wi          Fi Leamer      CVIF Klein
## OCCUPATION    1.2982 0.7703     17.3637     19.5715 0.8777    1.3620     0
## SECTOR        1.1987 0.8343     11.5670     13.0378 0.9134    1.2576     0
## UNION         1.1209 0.8922      7.0368      7.9315 0.9445    1.1759     0
## EDUCATION   231.1956 0.0043  13402.4982  15106.5849 0.0658  242.5527     1
## EXPERIENCE 5184.0939 0.0002 301771.2445 340140.5368 0.0139 5438.7545     1
## AGE        4645.6650 0.0002 270422.7164 304806.1391 0.0147 4873.8761     1
## SEX           1.0916 0.9161      5.3351      6.0135 0.9571    1.1453     0
## MARR          1.0961 0.9123      5.5969      6.3085 0.9551    1.1500     0
## RACE          1.0371 0.9642      2.1622      2.4372 0.9819    1.0881     0
## SOUTH         1.0468 0.9553      2.7264      3.0731 0.9774    1.0983     0
##              IND1   IND2
## OCCUPATION 0.0132 0.6125
## SECTOR     0.0143 0.4419
## UNION      0.0153 0.2875
## EDUCATION  0.0001 2.6546
## EXPERIENCE 0.0000 2.6656
## AGE        0.0000 2.6656
## SEX        0.0157 0.2238
## MARR       0.0157 0.2338
## RACE       0.0166 0.0955
## SOUTH      0.0164 0.1193
##
## 1 --> COLLINEARITY is detected by the test
## 0 --> COLLINEARITY is not detected by the test
##
## OCCUPATION , EDUCATION , EXPERIENCE , AGE , MARR , RACE , coefficient(s)
are non-significant may be due to multicollinearity
##
## R-square of y on all x: 0.3185
##
## * use method argument to check which regressors may be the reason of
collinearity
## ==================================

# we can see education,experience,age multicollinearity is detected.

#The VIF, TOL and Wi columns provide the diagnostic output for variance
inflation factor, tolerance and Farrar-Glauber F-test respectively.
```

```
#The F-statistic for the variable 'experience' is quite high (5184.0939) followed by the variable 'age' (F -value of 4645.6650) and 'education' (F-value of 231.1956). The degrees of freedom is(k-1,n-k)or (10-1, 534-10)=(9,524)

#For this degrees of freedom at 5% level of significance, the theoretical value of F is 1.89774. Thus, the F test shows that either the variable 'experience' or 'age' or 'education' will be the root cause of multicollinearity.

#Though the F -value for 'education' is also significant, it may happen due to inclusion of highly collinear variables such as 'age' and 'experience'.

#Experience has the higest 5184 VIF

#Finally, for examining the pattern of multicollinearity, it is required to conduct t-test for correlation coefficient.

#In R, there are several packages for getting the partial correlation coefficients along with the t- test for checking their significance level.

#We'll the 'ppcor' package to compute the partial correlation coefficients along with the t-statistic and corresponding p-values.

 #install.packages('ppcor')
 library(ppcor)
```

```
## Loading required package: MASS
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
 pcor(x,method = "pearson")
```

```
## $estimate
##               EDUCATION        SOUTH          SEX  EXPERIENCE        UNION
## EDUCATION   1.000000000 -0.031750193  0.051510483 -0.99756187 -0.007479144
## SOUTH      -0.031750193  1.000000000 -0.030152499 -0.02231360 -0.097548621
## SEX         0.051510483 -0.030152499  1.000000000  0.05497703 -0.120087577
## EXPERIENCE -0.997561873 -0.022313605  0.054977034  1.00000000 -0.010244447
## UNION      -0.007479144 -0.097548621 -0.120087577 -0.01024445  1.000000000
## AGE         0.997261601  0.021525073 -0.053697851  0.99987574  0.012238897
## RACE        0.017230877 -0.111197596  0.020017315  0.01088849 -0.107706183
## OCCUPATION  0.029436911  0.008430595 -0.142750864  0.04205856  0.212996388
## SECTOR     -0.021253493 -0.021518760 -0.112146760 -0.01326166 -0.013531482
## MARR       -0.040302967  0.030418218  0.004163264 -0.04097664  0.068918496
```

```
##                    AGE          RACE   OCCUPATION       SECTOR         MARR
## EDUCATION   0.99726160  0.017230877  0.029436911 -0.021253493 -0.040302967
## SOUTH       0.02152507 -0.111197596  0.008430595 -0.021518760  0.030418218
## SEX        -0.05369785  0.020017315 -0.142750864 -0.112146760  0.004163264
## EXPERIENCE  0.99987574  0.010888486  0.042058560 -0.013261665 -0.040976643
## UNION       0.01223890 -0.107706183  0.212996388 -0.013531482  0.068918496
## AGE         1.00000000 -0.010803310 -0.044140293  0.014565751  0.045090327
## RACE       -0.01080331  1.000000000  0.057539374  0.006412099  0.055645964
## OCCUPATION -0.04414029  0.057539374  1.000000000  0.314746868 -0.018580965
## SECTOR      0.01456575  0.006412099  0.314746868  1.000000000  0.036495494
## MARR        0.04509033  0.055645964 -0.018580965  0.036495494  1.000000000
##
## $p.value
##            EDUCATION      SOUTH         SEX EXPERIENCE        UNION
AGE
## EDUCATION  0.0000000 0.46745162 0.238259049  0.0000000 8.641246e-01
0.0000000
## SOUTH      0.4674516 0.00000000 0.490162786  0.6096300 2.526916e-02
0.6223281
## SEX        0.2382590 0.49016279 0.000000000  0.2080904 5.822656e-03
0.2188841
## EXPERIENCE 0.0000000 0.60962999 0.208090393  0.0000000 8.146741e-01
0.0000000
## UNION      0.8641246 0.02526916 0.005822656  0.8146741 0.000000e+00
0.7794483
## AGE        0.0000000 0.62232811 0.218884070  0.0000000 7.794483e-01
0.0000000
## RACE       0.6933788 0.01070652 0.646920379  0.8032546 1.345383e-02
0.8047625
## OCCUPATION 0.5005235 0.84704000 0.001027137  0.3356824 8.220095e-07
0.3122902
## SECTOR     0.6267278 0.62243025 0.010051378  0.7615531 7.568528e-01
0.7389200
## MARR       0.3562616 0.48634504 0.924111163  0.3482728 1.143954e-01
0.3019796
##                  RACE   OCCUPATION       SECTOR      MARR
## EDUCATION  0.69337880 5.005235e-01 6.267278e-01 0.3562616
## SOUTH      0.01070652 8.470400e-01 6.224302e-01 0.4863450
## SEX        0.64692038 1.027137e-03 1.005138e-02 0.9241112
## EXPERIENCE 0.80325456 3.356824e-01 7.615531e-01 0.3482728
## UNION      0.01345383 8.220095e-07 7.568528e-01 0.1143954
## AGE        0.80476248 3.122902e-01 7.389200e-01 0.3019796
## RACE       0.00000000 1.876376e-01 8.833600e-01 0.2026017
## OCCUPATION 0.18763758 0.000000e+00 1.467261e-13 0.6707116
## SECTOR     0.88336002 1.467261e-13 0.000000e+00 0.4035489
## MARR       0.20260170 6.707116e-01 4.035489e-01 0.0000000
##
## $statistic
##               EDUCATION      SOUTH         SEX   EXPERIENCE      UNION
## EDUCATION     0.0000000 -0.7271618  1.18069629 -327.2105031 -0.1712102
```

```
## SOUTH        -0.7271618  0.0000000 -0.69053623   -0.5109090 -2.2436907
## SEX           1.1806963 -0.6905362  0.00000000    1.2603880 -2.7689685
## EXPERIENCE -327.2105031 -0.5109090  1.26038801    0.0000000 -0.2345184
## UNION        -0.1712102 -2.2436907 -2.76896848   -0.2345184  0.0000000
## AGE         308.6803174  0.4928456 -1.23097601 1451.9092015  0.2801822
## RACE          0.3944914 -2.5613138  0.45830912    0.2492636 -2.4799336
## OCCUPATION    0.6741338  0.1929920 -3.30152873    0.9636171  4.9902208
## SECTOR       -0.4866246 -0.4927010 -2.58345399   -0.3036001 -0.3097781
## MARR         -0.9233273  0.6966272  0.09530228   -0.9387867  1.5813765
##                     AGE       RACE OCCUPATION     SECTOR        MARR
## EDUCATION   308.6803174  0.3944914  0.6741338 -0.4866246 -0.92332727
## SOUTH         0.4928456 -2.5613138  0.1929920 -0.4927010  0.69662719
## SEX          -1.2309760  0.4583091 -3.3015287 -2.5834540  0.09530228
## EXPERIENCE 1451.9092015  0.2492636  0.9636171 -0.3036001 -0.93878671
## UNION         0.2801822 -2.4799336  4.9902208 -0.3097781  1.58137652
## AGE           0.0000000 -0.2473135 -1.0114033  0.3334607  1.03321563
## RACE         -0.2473135  0.0000000  1.3193223  0.1467827  1.27577106
## OCCUPATION   -1.0114033  1.3193223  0.0000000  7.5906763 -0.42541117
## SECTOR        0.3334607  0.1467827  7.5906763  0.0000000  0.83597695
## MARR          1.0332156  1.2757711 -0.4254112  0.8359769  0.00000000
##
## $n
## [1] 534
##
## $gp
## [1] 8
##
## $method
## [1] "pearson"
```

#As expected the high partial correlation between ‘age’ and ‘experience’ is found to be highly statistically significant.

#Similar is the case for ‘education – experience’ and ‘education – age’ .

#Not only that even some of the low correlation coefficients are also found to be highyl significant. Thus, the Farrar-Glauber test points out that Experience has the higest 5184 VIF(from imcdag test) is the root cause of all multicollinearity problem.

#remedial measures

#There are several remedial measure to deal with the problem of multicollinearity such Prinicipal Component Regression, Ridge Regression, Stepwise Regression etc.

#However, in the present case, we’ll go for the exclusion of the variables for which the VIF values are above 10 and as well as the concerned variable logically seems to be redundant.

```
#Age and experience will certainly be correlated

#why to use both of them?

#If we use 'age' or 'age-squared', it will reflect the experience of the
respondent also.

# Experience has the higest 5184 VIF value

#Thus, we try to build a model by excluding 'experience', estimate the model
and go for further diagnosis for the presence of multicollinearity

# model after excluding  Experience

fit2 <-
lm(log(WAGE)~OCCUPATION+SECTOR+UNION++EDUCATION+AGE+SEX+MARR+RACE+SOUTH,data=
mul)
summary(fit2)

##
## Call:
## lm(formula = log(WAGE) ~ OCCUPATION + SECTOR + UNION + +EDUCATION +
##     AGE + SEX + MARR + RACE + SOUTH, data = mul)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.16018 -0.29085 -0.00513  0.29985  1.97932
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.501358   0.164794   3.042 0.002465 **
## OCCUPATION  -0.006941   0.013095  -0.530 0.596309
## SECTOR       0.091013   0.038723   2.350 0.019125 *
## UNION        0.200018   0.052459   3.813 0.000154 ***
## EDUCATION    0.083815   0.007728  10.846  < 2e-16 ***
## AGE          0.010305   0.001745   5.905 6.34e-09 ***
## SEX         -0.220100   0.039837  -5.525 5.20e-08 ***
## MARR         0.075125   0.041886   1.794 0.073458 .
## RACE         0.050674   0.028523   1.777 0.076210 .
## SOUTH       -0.103186   0.042802  -2.411 0.016261 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4397 on 524 degrees of freedom
## Multiple R-squared:  0.3175, Adjusted R-squared:  0.3058
## F-statistic: 27.09 on 9 and 524 DF,  p-value: < 2.2e-16

#Now by looking at the significance level, it is seen that out of nine of
regression coefficients, eight are statistically significant. The R-square
value is 0.31 and F-value is also very high and significant too.
```

```
library(car)
vif(fit2)

## OCCUPATION     SECTOR      UNION  EDUCATION        AGE        SEX
MARR
##   1.295935   1.198460   1.120743   1.125994   1.154496   1.088334
1.094289
##       RACE      SOUTH
##   1.037015   1.046306

#Even the VIF values for the explanatory variables have reduced to very lower
values.

#Now the model is free from Multicollinearity
```