

DiZhen_1717719

dizhen

2020/4/18

Data

```
KBI <- c(28,68,59,91,70,38,46,57,89,48,74,78,43,76,72,61,63,77,85,31,79,92,76,91,78,103,99,73,88,64,52,7
ADL <- c(39,52,89,57,28,34,42,52,88,90,38,83,30,45,47,90,63,34,76,26,68,85,22,82,80,80,81,30,27,72,46,6
MEM <- c(4,33,17,31,35,3,16,6,41,24,22,41,9,33,36,17,14,35,33,13,34,28,12,57,51,20,20,7,27,9,15,52,26,5
COG <- c(18,9,3,7,19,25,17,26,13,3,13,11,24,14,18,0,16,22,23,18,26,10,16,3,3,18,1,17,27,0,22,13,18,0,19

mydata <- data.frame(KBI,ADL,MEM,COG)
str(mydata)
```

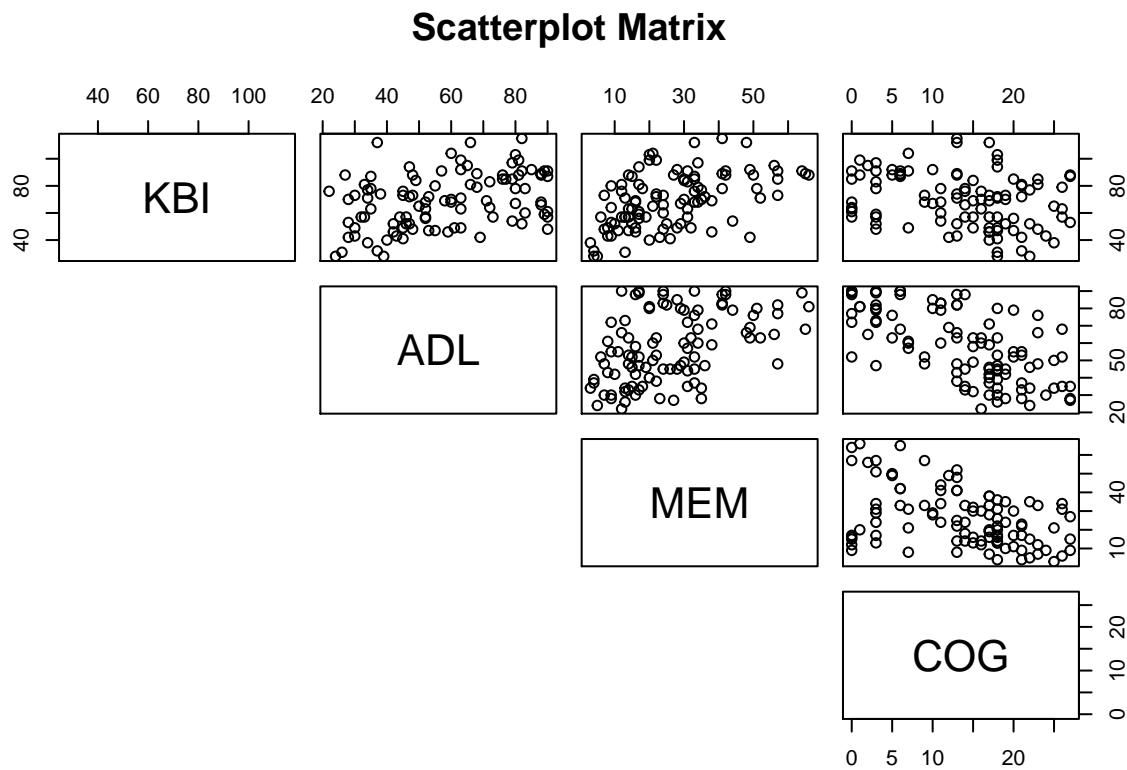
```
## 'data.frame':    100 obs. of  4 variables:
## $ KBI: num  28 68 59 91 70 38 46 57 89 48 ...
## $ ADL: num  39 52 89 57 28 34 42 52 88 90 ...
## $ MEM: num   4 33 17 31 35  3 16  6 41 24 ...
## $ COG: num  18  9  3  7 19 25 17 26 13  3 ...
```

```
head(mydata)
```

```
##   KBI ADL MEM COG
## 1  28  39   4  18
## 2  68  52  33   9
## 3  59  89  17   3
## 4  91  57  31   7
## 5  70  28  35  19
## 6  38  34   3  25
```

Visualization

```
pairs(~KBI + ADL + MEM + COG, data = mydata,
      lower.panel = NULL,
      main = "Scatterplot Matrix")
```



Question 1

```
mydata.LM1 <- lm(KBI ~ ADL + MEM + COG, data = mydata)
summary(mydata.LM1)
```

```
##
## Call:
## lm(formula = KBI ~ ADL + MEM + COG, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -42.037 -10.535  -1.503   9.213  43.151
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  40.4908    10.1030   4.008 0.000121 ***
## ADL           0.2162     0.1168   1.851 0.067273 .
## MEM           0.5547     0.1300   4.267 4.65e-05 ***
## COG           0.1210     0.3003   0.403 0.687978
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.26 on 96 degrees of freedom
```

```
## Multiple R-squared:  0.282, Adjusted R-squared:  0.2596
## F-statistic: 12.57 on 3 and 96 DF,  p-value: 5.315e-07
```

The multiple regression equation is:

$$KBI = 40.5 + 0.2 \times ADL + 0.6 \times MEM + 0.1 \times COG$$

The mean KBI when ADL, MEM and COG are zero is 40.5. For every extra unit of ADL, the expected KBI increases by 0.2 points, holding other variables constant. For every extra unit of MEM, the mean KBI increases by 0.6 points, holding other variables constant. For every extra unit of COG, the mean KBI increases by 0.1 points, holding other variables constant.

Question 2

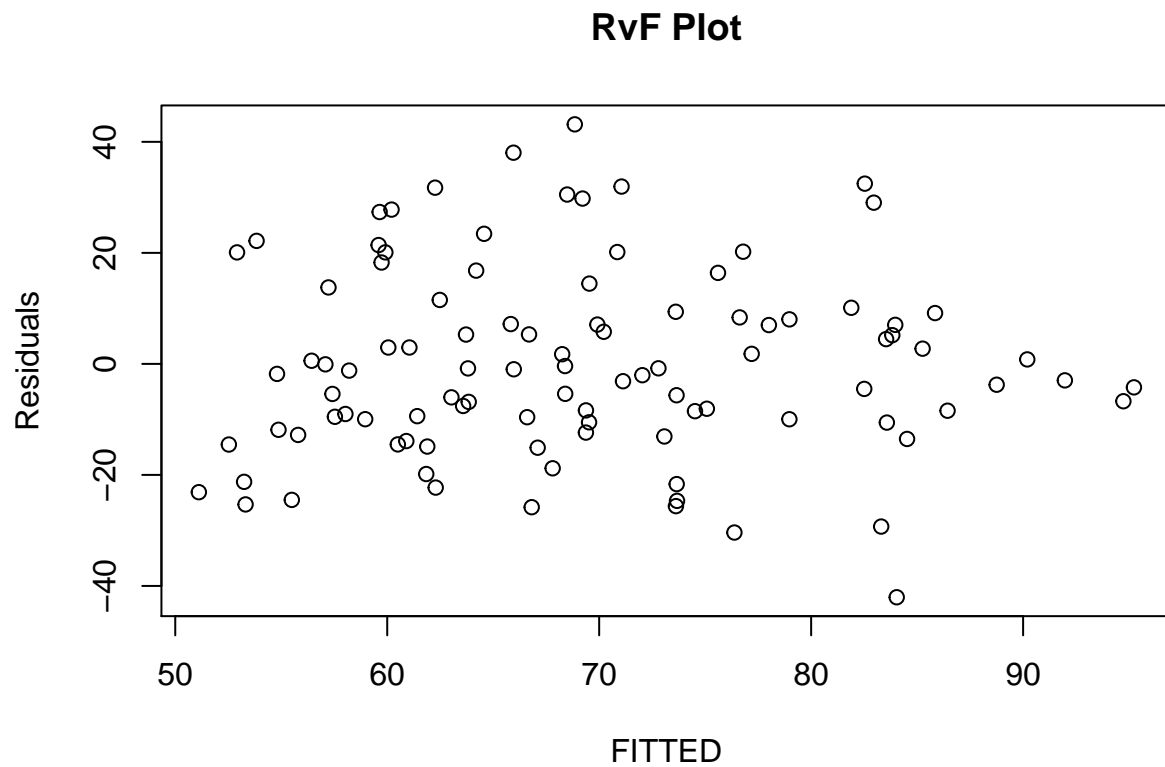
```
mydata$FITTED <- predict(mydata.LM1, type = "response")
mydata$RESID <- resid(mydata.LM1)
head(mydata)
```

##	KBI	ADL	MEM	COG	FITTED	RESID
## 1	28	39	4	18	53.31764	-25.317639
## 2	68	52	33	9	71.12423	-3.124227
## 3	59	89	17	3	69.52256	-10.522556
## 4	91	57	31	7	70.85388	20.146118
## 5	70	28	35	19	68.25490	1.745095
## 6	38	34	3	25	52.52884	-14.528836

Question 3

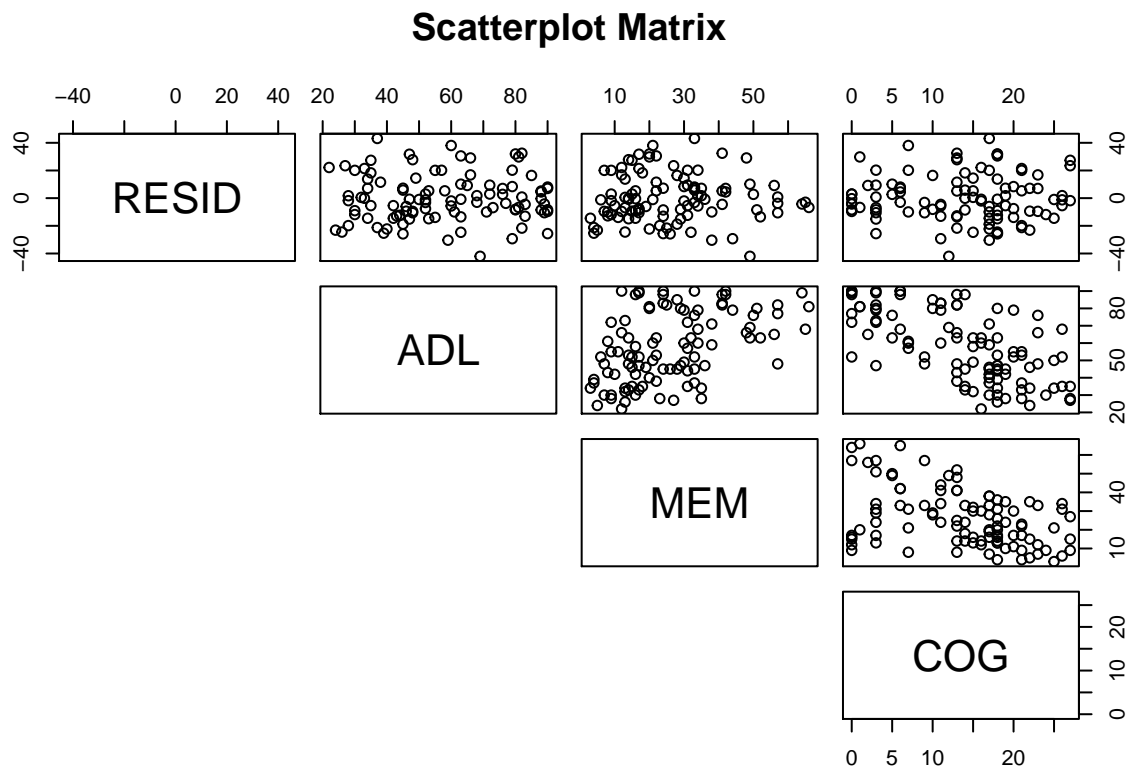
1. Plot the residuals against the fitted values (RvF plot)

```
plot(RESID-FITTED, data = mydata,
     main = "RvF Plot",
     ylab = "Residuals")
```



2. Plot the residuals against each of the predictors

```
pairs(~RESID + ADL + MEM + COG, data=mydata,  
      lower.panel=NULL,  
      main = "Scatterplot Matrix")
```



In the RvF plot, we do not notice any systematic non-linear deviation. In the scatterplot matrix, we are only interested in the first row. Neither shows any systematic non-linear deviation.

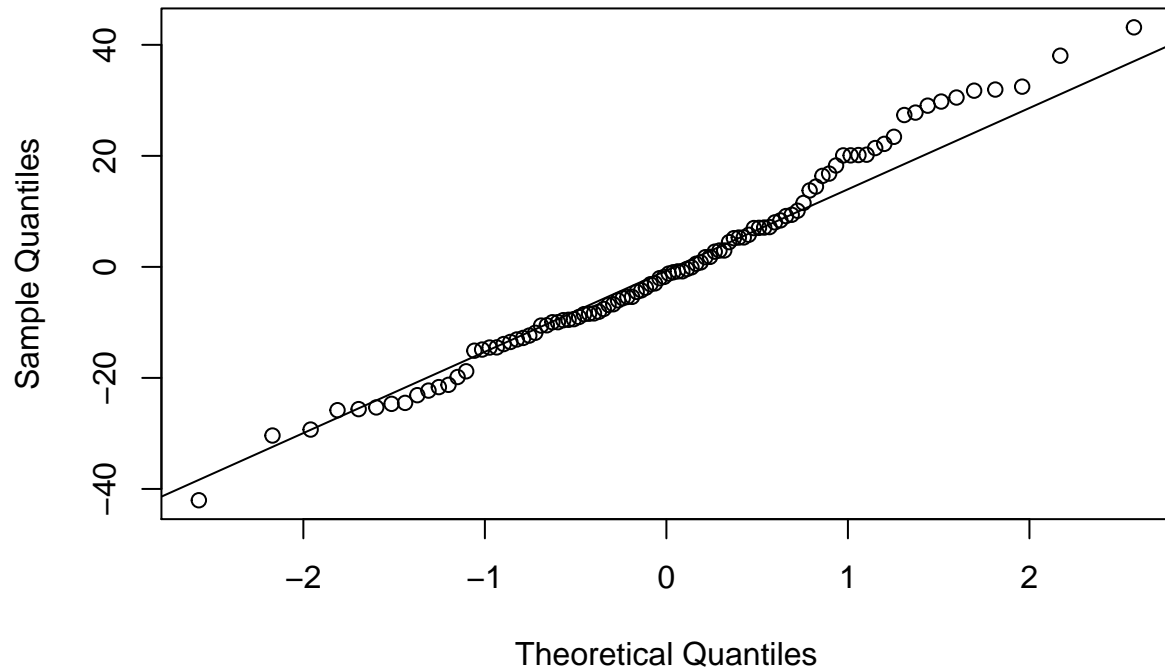
Conclusion: overall, the assumption of normality is not violated.

Question 4

Produce a quantile-quantile plot of the residuals.

```
qqnorm(mydata$RESID,
       main = "Normal Quantile-Quantile Plot of Ordinary Residuals")
qqline(mydata$RESID)
```

Normal Quantile–Quantile Plot of Ordinary Residuals



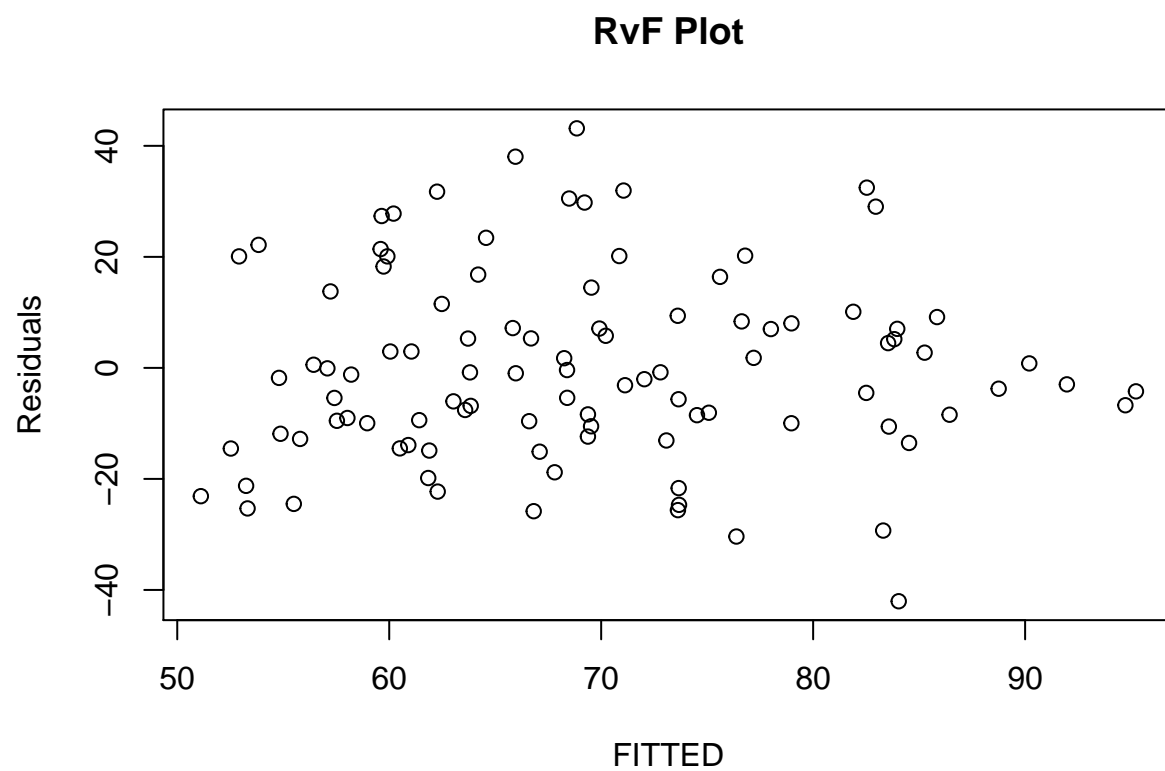
The graph shows some deviation in the right tail. Although the points hug the qq line quite closely, this deviation affects more than 10 observations.

Conclusion: The assumption of normality is violated.

Question 5

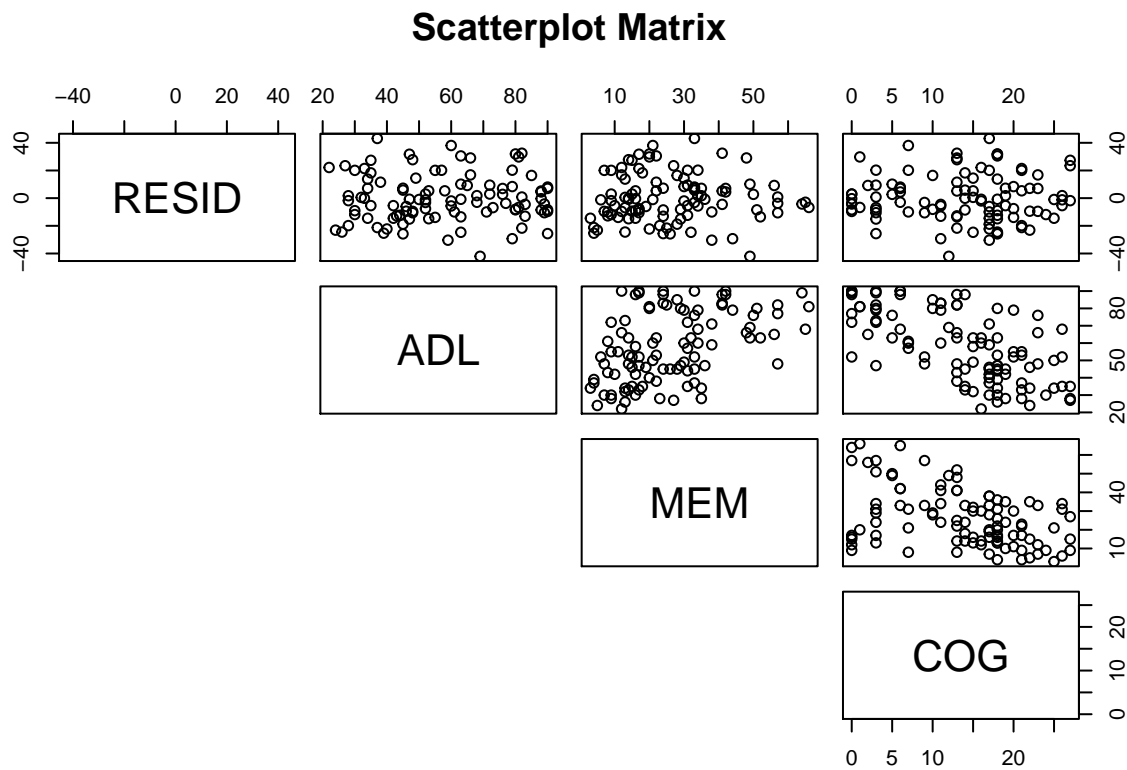
1. Plot the residuals against the fitted values (RvF plot)

```
plot(RESID~FITTED, data = mydata,  
     main = "RvF Plot",  
     ylab = "Residuals")
```



2. Plot the residuals against each of the predictors

```
pairs(~RESID + ADL + MEM + COG, data=mydata,  
      lower.panel=NULL,  
      main = "Scatterplot Matrix")
```



There is no wedge-shaped patterns or trumpeting.

Conclusion: The assumption of equal variance of residuals is not violated.

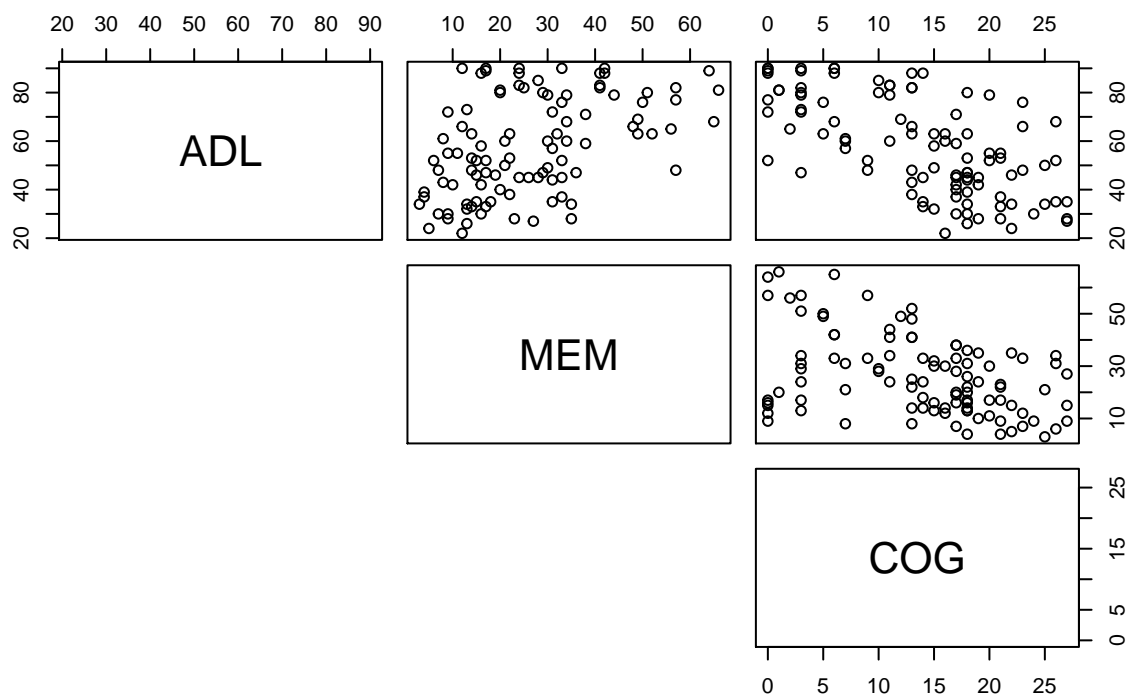
Question 6

Evidence about independence is available from the design of data.

Question 7

```
pairs(~ADL + MEM + COG, data=mydata,
      lower.panel=NULL,
      main = "Collinearity between Predictors")
```


Collinearity between Predictors



From the plot, we observe little linear relationship between predictors. There is negative relationships between ADL and COG, and between MEM and COG. There is positive relationship between MEM and ADL.

```
cor(mydata$ADL,mydata$MEM)
```

```
## [1] 0.4557426
```

```
cor(mydata$ADL,mydata$COG)
```

```
## [1] -0.6470085
```

```
cor(mydata$MEM,mydata$COG)
```

```
## [1] -0.4297546
```

The correlation between ADL and MEM is 45.57%. The correlation between ADL and COG is -64.70%. The correlation between MEM and COG is -42.98%.

```
# if(!require(car)){install.packages("car")}
library(car)
```

```
## Loading required package: carData
```

```
vif(mydata.LM1)
```

```
##          ADL          MEM          COG  
## 1.842791 1.314053 1.790778
```

The VIF(ADL), VIF(MEM), VIF(COG) are all < 10 , indicating the absence of collinearity.

Conclusion: there is no multicollinearity between the predictors.

THE END