Di Zhen (1717719)

dizhen

2020.3.28

Set working directory:

```
setwd("D:/git/DPH112-xjtlu/week05")
```

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,

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
```

Question 1

```
# pairs(~KBI + ADL + MEM + COG, data = mydata,
#       lower.panel = NULL,
```

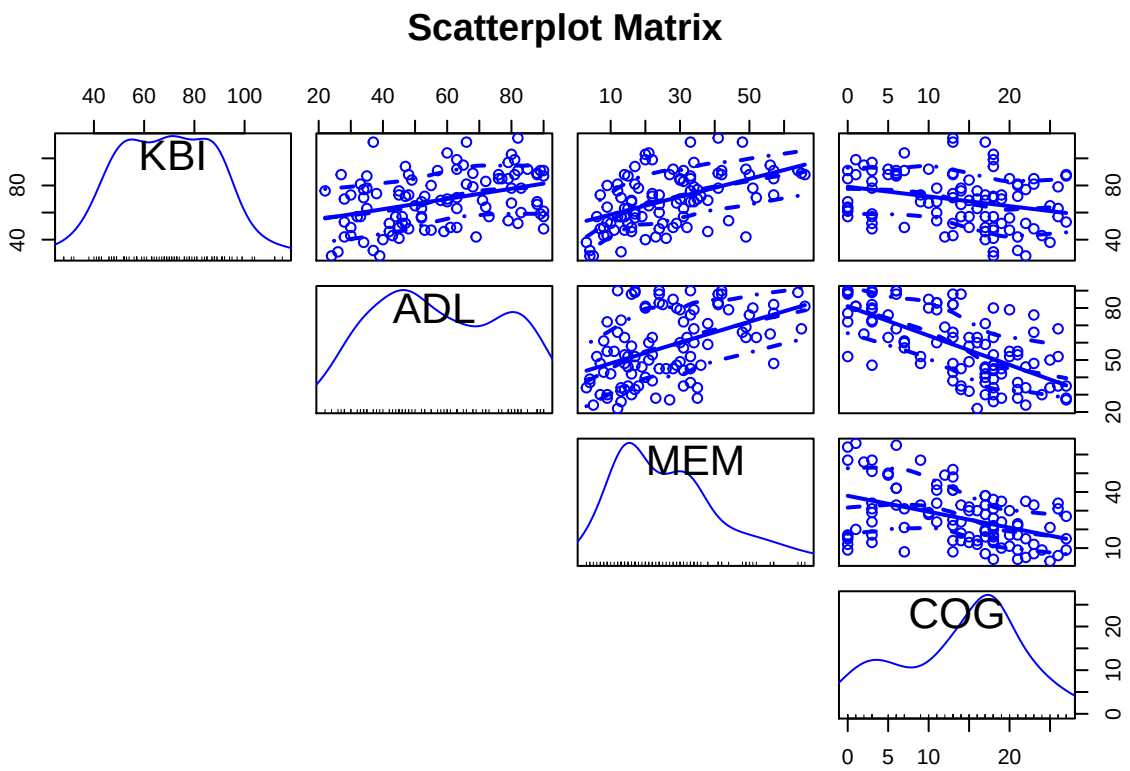
```
#       main = "Scatterplot Matrix")

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

```
## Loading required package: car
```

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

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



Question 2

KBI is positively correlated with ADL and MEM, slightly negatively correlated with COG. ADL is positively correlated with MEM and negatively correlated with COG. MEM is negatively correlated with COG.

Question 3

```r
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 equition is:

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

Question 4

1) The intercept is 40.5. If ADL = MEM = COG = 0 and have their meanings (or in the scope), the KBI is equal to 40.5. If ADL = MEM = COG = 0 is out of scope, the intercept has no practical meaning.

2) The coefficient of ADL is 0.2. This means if the score of total activities of daily living(ADL) increases 1 unit, the expected KBI will increase by 0.2, holding other variables constant.

3) The coefficient of MEM is 0.6. This means if the score of memory and behavioral problems (MEM) increases 1 unit, the expected KBI will increase by 0.6, holding other variables constant.

4) The coefficient of COG is 0.1. This means if the score of cognitive impairment (COG) increases 1 unit, the expected KBI will increase by 0.1, holding other variables constant.

Question 5

1) Test the null hypothesis of $H_0 : intercept = 0$ at $\alpha = 0.03$:

The output gives the value of the t statistics as 4.0 and the p-value is 1.21×10^{-4}. The p-value is less than $\alpha = 0.03$ providing us with evidence to reject the null hypothesis.

2) Test the null hypothesis of $H_0 : \beta_{ADL} = 0$ at $\alpha = 0.03$:

The output gives the value of the t statistics as 1.9 and the p-value is 0.07. The p-value is more than $\alpha = 0.03$,thus we fail to reject the null hypothesis.

3) Test the null hypothesis of $H_0 : \beta_{MEM} = 0$ at $\alpha = 0.03$:

The output gives the value of the t statistics as 4.3 and the p-value is 4.65×10^{-5}. The p-value is less than $\alpha = 0.03$ providing us with evidence to reject the null hypothesis.

4) Test the null hypothesis of $H_0 : \beta_{COG} = 0$ at $\alpha = 0.03$:

The output gives the value of the t statistics as 0.4 and the p-value is 0.69. The p-value is more than $\alpha = 0.03$, thus we fail to reject the null hypothesis.

Question 6

```
confint(mydata.LM1, level = 0.97)
```

```
##                    1.5 %     98.5 %
## (Intercept) 18.23563084 62.7459565
## ADL         -0.04111709  0.4734715
## MEM          0.26831265  0.8409940
## COG         -0.54052559  0.7824503
```

1) The 97% confidence interval for intercept is from 18.2 to 62.7. We are 97% confident that the intercept estimate lies between the interval 18.2 and 62.7, because on repeated sampling, 97% of intervals constructed in ths manner will contain the true intercept.

2) The 97% confidence interval for slope of ADL is from 0 to 0.5. We are 97% confident that the slope estimate of ADL lies between the interval 0 and 0.5, because on repeated sampling, 97% of intervals constructed in ths manner will contain the true slope of ADL.

3) The 97% confidence interval for slope of MEM is from 0.3 to 0.8. We are 97% confident that the slope estimate of MEM lies between the interval 0.3 and 0.8, because on repeated sampling, 97% of intervals constructed in ths manner will contain the true slope of MEM.

4) The 97% confidence interval for slope of COG is from -0.5 to 0.8. We are 97% confident that the slope estimate of COG lies between the interval -0.5 and 0.8, because on repeated sampling, 97% of intervals constructed in ths manner will contain the true slope of COG.

Question 7

```
mydata.NEW1 <- data.frame(ADL = 1, MEM = 2, COG = 3)
mydata.NEW2 <- data.frame(ADL = 40, MEM = 50, COG = 60)
predict(mydata.LM1, mydata.NEW1,
        interval = "prediction",
        level = 0.97)
```

```
##        fit        lwr      upr
## 1 42.17916 -0.8947255 85.25305
```

```r
predict(mydata.LM1, mydata.NEW1,
        interval = "confidence",
        level = 0.97)
```

```
##        fit      lwr      upr
## 1 42.17916 21.93147 62.42685
```

```r
predict(mydata.LM1, mydata.NEW2,
        interval = "prediction",
        level = 0.97)
```

```
##        fit     lwr      upr
## 1 84.12829 35.0755 133.1811
```

```r
predict(mydata.LM1, mydata.NEW2,
        interval = "confidence",
        level = 0.97)
```

```
##        fit      lwr      upr
## 1 84.12829 53.13176 115.1248
```

Suppose there are two new caregivers. One got ADL=1, MEM=2, COG=3, the other got ADL = 40, MEM = 50, COG = 60.

1) For the first caregiver:

The 97% confidence interval for mean KBI is (-0.9, 85.2). The 97% prediction interval for KBI is (21.9, 62.4).

2) For the second caregiver:

The 97% confidence interval for mean KBI is (35.1, 133.2). The 97% prediction interval for KBI is (53.1, 115.1).

THE END