PROBLEM_SET2

Kafayat Liadi

2025-10-21

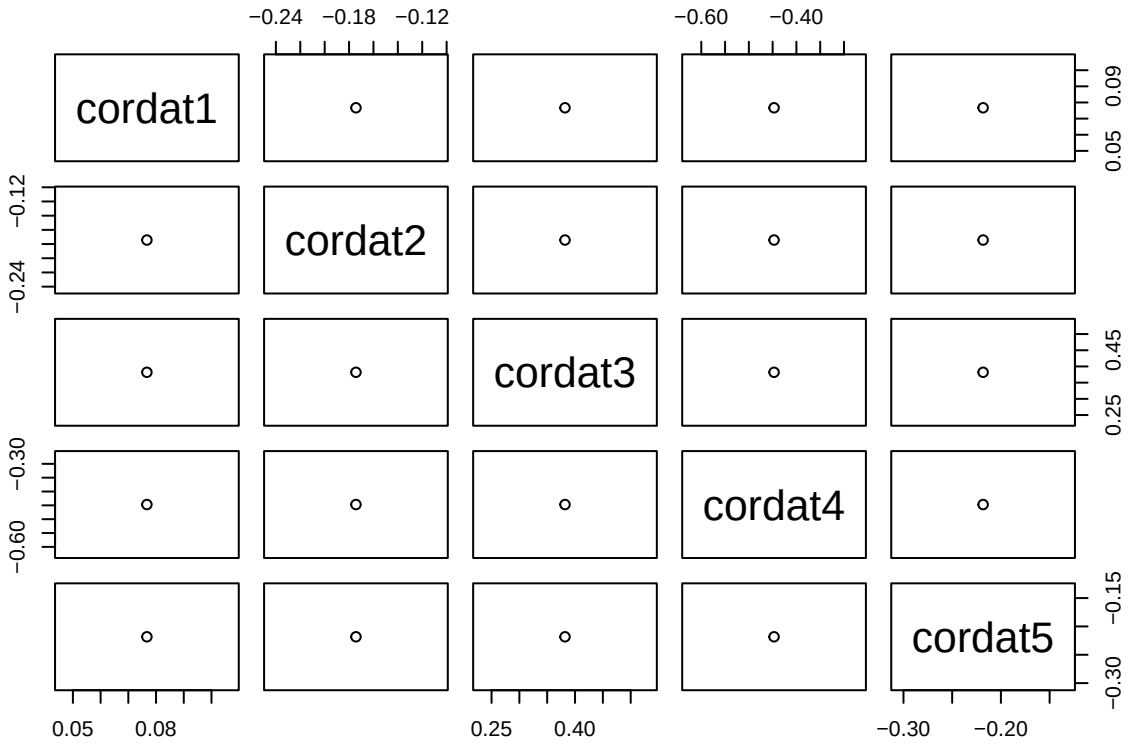
Question 1

```
## Create 2 random variables with 20 observation with 5 repetitions

##  Repetition 1
X1 <- rnorm(20, 4.6, 5)
X2 <- rnorm(20, 0.5, 1.2)


data1 <- data.frame(X1,X2)

# Correlation coefficient 1
cordat1 <- cor(data1$X1,data1$X2)

# Repetition 2
X3<- rnorm(20, 8, 7)
X4 <- rnorm(20, 6.5, 2)
data2 <- data.frame(X3,X4)

cordat2 <- cor(data2$X3,data2$X4)

# Repetition 3
X5<- rnorm(20, 9.5, 8.4)
X6 <- rnorm(20, 12.5, 7.2)
data3 <- data.frame(X5,X6)

cordat3 <- cor(data3$X5,data3$X6)

# Repetition 4
X7<- rnorm(20, 13.5, 0.4)
X8 <- rnorm(20, 2.5, 0.2)
data4 <- data.frame(X7,X8)

cordat4 <- cor(data4$X7,data4$X8)

# Repetition 5
X9<- rnorm(20, 3.5, 2.4)
X10 <- rnorm(20, 9.5, 6.2)
data5 <- data.frame(X9,X10)

cordat5 <- cor(data5$X9,data5$X10)
```

```r
cor_res <- data.frame(cordat1,cordat2,cordat3, cordat4, cordat5)

cor_res
```

```
##      cordat1    cordat2   cordat3    cordat4    cordat5
## 1 0.07670192 -0.1743477 0.3819788 -0.4470467 -0.2182665
```

```r
plot(cor_res)
```




```r
stdev <- sd(cor_res)

cat("The standard deviation of the correlation coefficient:", stdev)
```

```
## The standard deviation of the correlation coefficient: 0.3164846
```

```r
avg <- apply(cor_res,1, FUN = mean)

cat("The average of the correlation coefficient:", avg)
```

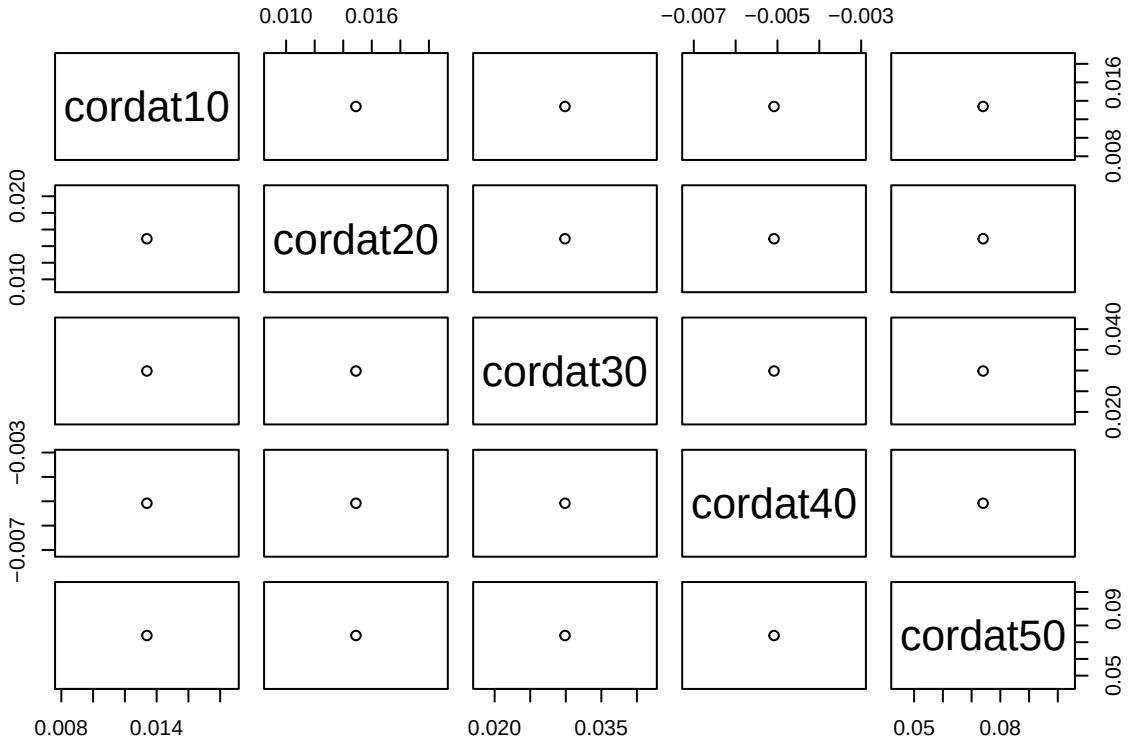
```
## The average of the correlation coefficient: -0.07619603
```

Question 2

```r
## Create random 2 random variables with sample size of 1000 in 5 repetitions

# Repetition 1
X10 <- rnorm(1000, 4.6, 5)
X20 <- rnorm(1000, 0.5, 1.2)


data10 <- data.frame(X10,X20)

# Correlation coefficient 1
cordat10 <- cor(data10$X10,data10$X20)

# Repetition 2
X30<- rnorm(1000, 8, 7)
X40 <- rnorm(1000, 6.5, 2)
data20 <- data.frame(X30,X40)

cordat20 <- cor(data20$X30,data20$X40)

# Repetition 3
X50<- rnorm(1000, 9.5, 8.4)
X60 <- rnorm(1000, 12.5, 7.2)
data30 <- data.frame(X50,X60)

cordat30 <- cor(data30$X50,data30$X60)

# Repetition 4
X70<- rnorm(1000, 13.5, 0.4)
X80 <- rnorm(1000, 2.5, 0.2)
data40 <- data.frame(X70,X80)

cordat40 <- cor(data40$X70,data40$X80)

# Repetition 5
X90<- rnorm(1000, 3.5, 2.4)
X100 <- rnorm(1000, 9.5, 6.2)
data50 <- data.frame(X90,X100)

cordat50 <- cor(data50$X90,data50$X100)



cor_res2 <- data.frame(cordat10,cordat20,cordat30, cordat40, cordat50)
cor_res2
```

```
##     cordat10   cordat20   cordat30     cordat40   cordat50
## 1 0.01338677 0.01488989 0.02989007 -0.005081923 0.07400493
```

```r
plot(cor_res2)
```



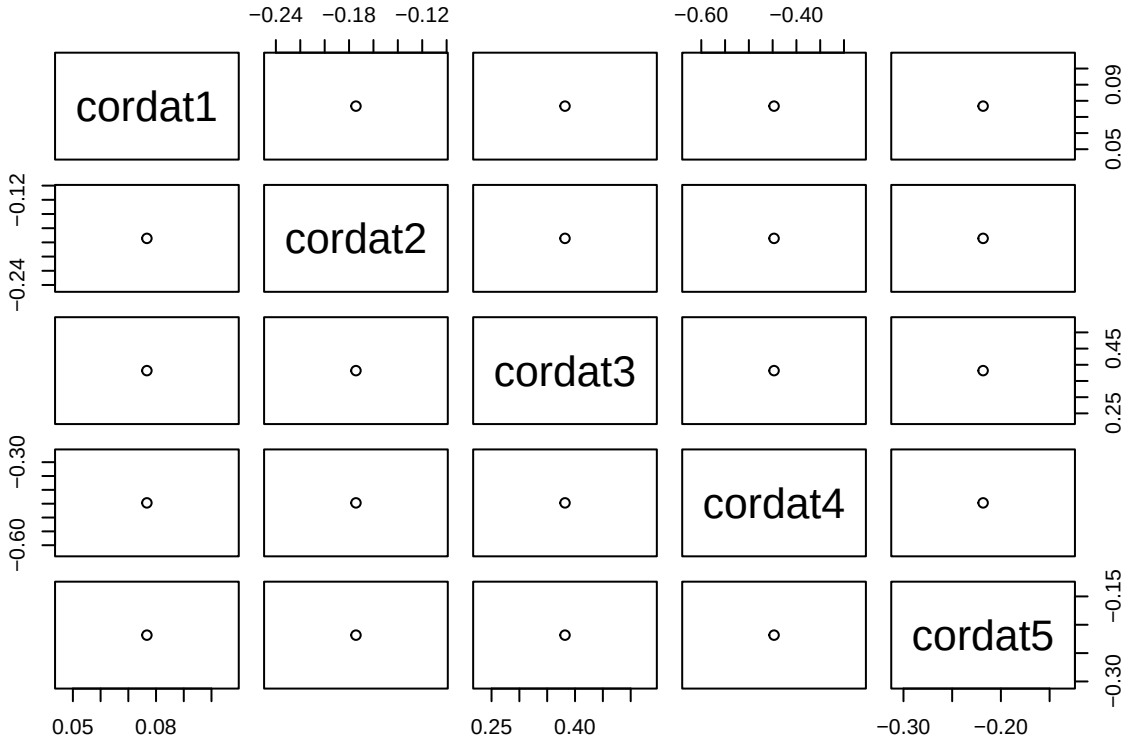

```r
stdev2 <- sd(cor_res2)

cat("The standard deviation of the correlation coefficient:", stdev2)
```

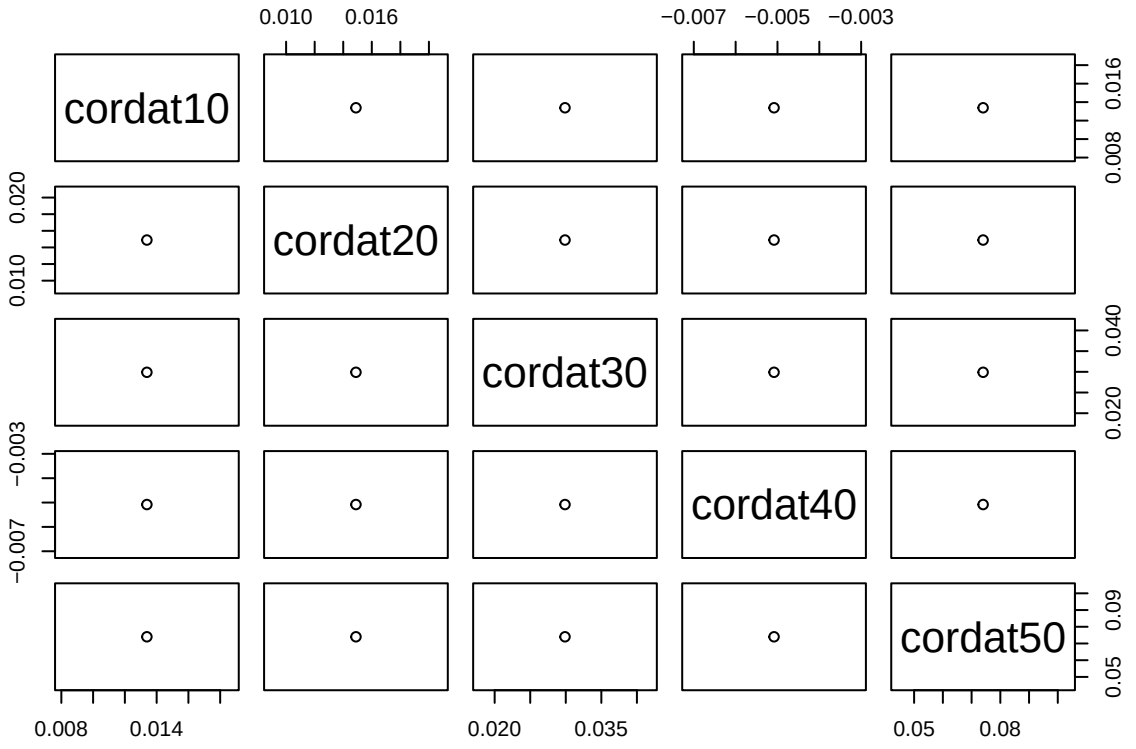
```
## The standard deviation of the correlation coefficient: 0.0298602
```

```r
avg2 <- apply(cor_res2,1, FUN = mean)

cat("The average of the correlation coefficient:", avg2)
```

```
## The average of the correlation coefficient: 0.02541795
```

Compare results of the two sample sizes (20 & 1000) tested





```r
cat("The standard deviation of the correlation coefficients of 20 sample size:", stdev)
```

```
## The standard deviation of the correlation coefficients of 20 sample size: 0.3164846
```

```r
cat("The standard deviation of the correlation coefficient of 1000 sample size:", stdev2)
```

```
## The standard deviation of the correlation coefficient of 1000 sample size: 0.0298602
```

```r
cat("The average of the correlation coefficient of 20 sample size:", avg)
```

```
## The average of the correlation coefficient of 20 sample size: -0.07619603
```

```r
cat("The average of the correlation coefficient of 1000 sample size:", avg2)
```

```
## The average of the correlation coefficient of 1000 sample size: 0.02541795
```

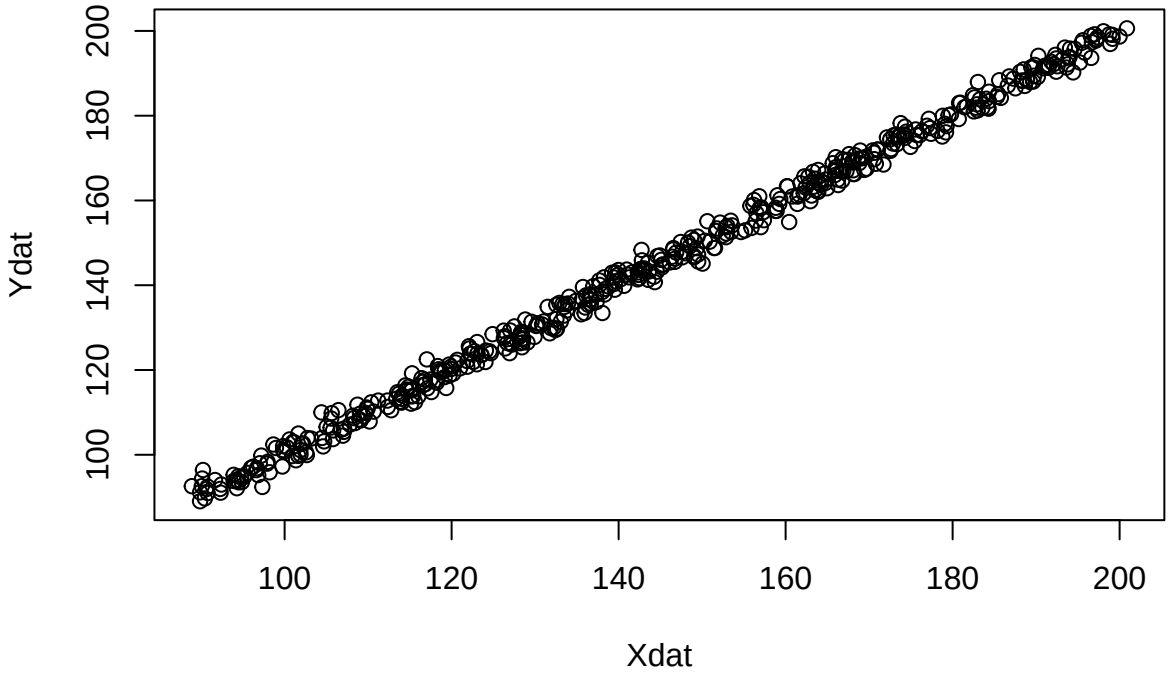
With a reduced standard deviation estimate of the correlation coefficient in the large sample size tested (1000) compared to the smaller sample size (20), result suggests reduced variation in the estimated correlation coefficients when the sample size is increased. Additionally, correlation coefficients estimated is reduced with increased sample size. However, it is important to note our result may change if the number of repetition considered is increased. Probably, 5 repetition used in this exercise may be too small.

Question 3

```r
Z <- runif(500, 90, 200)

Xdat <- Z + rnorm(500,0.4, 1.2)

Ydat <- Z + rnorm(500,0.7, 1.5)

corXY <- cor(Xdat,Ydat)
cat("The correlation between X and Y is :", corXY)
```

```
## The correlation between X and Y is : 0.998052
```

```r
plot(Xdat,Ydat)
```




```
cat("The correlation between Z and X is :", cor(Z,Xdat))
```

```
## The correlation between Z and X is : 0.99922
```

```
cat("The correlation between Z and Y is :", cor(Z,Ydat))
```

```
## The correlation between Z and Y is : 0.9987769
```

There is a linear relationship between X and Y.

Result suggest that correlation does not imply causation. X and Y are derived from Z, which suggest they are independent of one another. However, the correlation between the three variables (X & Y, X & Z and Y&Z) is ~0.99