gunyoung

gyc

2022-08-12

df<-airquality  
summary(df)

## Ozone Solar.R Wind Temp   
## Min. : 1.00 Min. : 7.0 Min. : 1.700 Min. :56.00   
## 1st Qu.: 18.00 1st Qu.:115.8 1st Qu.: 7.400 1st Qu.:72.00   
## Median : 31.50 Median :205.0 Median : 9.700 Median :79.00   
## Mean : 42.13 Mean :185.9 Mean : 9.958 Mean :77.88   
## 3rd Qu.: 63.25 3rd Qu.:258.8 3rd Qu.:11.500 3rd Qu.:85.00   
## Max. :168.00 Max. :334.0 Max. :20.700 Max. :97.00   
## NA's :37 NA's :7   
## Month Day   
## Min. :5.000 Min. : 1.0   
## 1st Qu.:6.000 1st Qu.: 8.0   
## Median :7.000 Median :16.0   
## Mean :6.993 Mean :15.8   
## 3rd Qu.:8.000 3rd Qu.:23.0   
## Max. :9.000 Max. :31.0   
##

* 각 열의 요약 통계량과 NA의 수를 확인한 결과
  + Ozone은 37개, Solar.R은 7개 있으므로 NA값을 모두 삭제

str(df) # 모두 수치형데이터(연속형)

## 'data.frame': 153 obs. of 6 variables:  
## $ Ozone : int 41 36 12 18 NA 28 23 19 8 NA ...  
## $ Solar.R: int 190 118 149 313 NA NA 299 99 19 194 ...  
## $ Wind : num 7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ...  
## $ Temp : int 67 72 74 62 56 66 65 59 61 69 ...  
## $ Month : int 5 5 5 5 5 5 5 5 5 5 ...  
## $ Day : int 1 2 3 4 5 6 7 8 9 10 ...

class(df) # 데프

## [1] "data.frame"

summary(df) # NA값 사라진지 확인

## Ozone Solar.R Wind Temp   
## Min. : 1.00 Min. : 7.0 Min. : 1.700 Min. :56.00   
## 1st Qu.: 18.00 1st Qu.:115.8 1st Qu.: 7.400 1st Qu.:72.00   
## Median : 31.50 Median :205.0 Median : 9.700 Median :79.00   
## Mean : 42.13 Mean :185.9 Mean : 9.958 Mean :77.88   
## 3rd Qu.: 63.25 3rd Qu.:258.8 3rd Qu.:11.500 3rd Qu.:85.00   
## Max. :168.00 Max. :334.0 Max. :20.700 Max. :97.00   
## NA's :37 NA's :7   
## Month Day   
## Min. :5.000 Min. : 1.0   
## 1st Qu.:6.000 1st Qu.: 8.0   
## Median :7.000 Median :16.0   
## Mean :6.993 Mean :15.8   
## 3rd Qu.:8.000 3rd Qu.:23.0   
## Max. :9.000 Max. :31.0   
##

* str()로 모든 열들이 수치형(연속형)인것을 확인
* class()로 데이터 프레임
* summary()로 NA 행이 삭제되었는지 확인

library(psych)  
corr.test(df)

## Call:corr.test(x = df)  
## Correlation matrix   
## Ozone Solar.R Wind Temp Month Day  
## Ozone 1.00 0.35 -0.60 0.70 0.16 -0.01  
## Solar.R 0.35 1.00 -0.06 0.28 -0.08 -0.15  
## Wind -0.60 -0.06 1.00 -0.46 -0.18 0.03  
## Temp 0.70 0.28 -0.46 1.00 0.42 -0.13  
## Month 0.16 -0.08 -0.18 0.42 1.00 -0.01  
## Day -0.01 -0.15 0.03 -0.13 -0.01 1.00  
## Sample Size   
## Ozone Solar.R Wind Temp Month Day  
## Ozone 116 111 116 116 116 116  
## Solar.R 111 146 146 146 146 146  
## Wind 116 146 153 153 153 153  
## Temp 116 146 153 153 153 153  
## Month 116 146 153 153 153 153  
## Day 116 146 153 153 153 153  
## Probability values (Entries above the diagonal are adjusted for multiple tests.)   
## Ozone Solar.R Wind Temp Month Day  
## Ozone 0.00 0.00 0.00 0.00 0.56 1.00  
## Solar.R 0.00 0.00 1.00 0.01 1.00 0.56  
## Wind 0.00 0.50 0.00 0.00 0.25 1.00  
## Temp 0.00 0.00 0.00 0.00 0.00 0.65  
## Month 0.08 0.37 0.03 0.00 0.00 1.00  
## Day 0.89 0.07 0.74 0.11 0.92 0.00  
##   
## To see confidence intervals of the correlations, print with the short=FALSE option

* Ozone열과 Wind열과 상관계수가 -0.6
* Ozone열과 Temp열과 상관계수가 0.7
* Probability values는 유의 수준처럼 0.05보다 작으면 *통계적*으로 유의하다고 할 수 있다.