

R 활용 통계심화

종합도전문제

📋 도전문제

- ❖ Data폴더에 있는 모든 엑셀 파일의 자료를 읽어 하나의 데이터프레임으로 만듦
 - ✓ 엑셀파일은 월별로 측정된 자료로 파일명은 "data**.xlsx"
 - **은 월을 나타냄. 예 2월 ⇒ 02
 - ✓ 각 엑셀 파일의 변수명(colname)은 Y, X1, X2, Group으로 되어 있음
- ❖ 각 엑셀 파일의 자료를 이용하여 (X1,Y), (X2, Y)의 산점도 2개를 한 화면에 그리고 엑셀 파일명과 동일한 jpg 파일로 저장함
 - ✓ 각 산점도에서 점은 Group (A, B)별로 다른 색으로 표시함
- ❖ 각 엑셀 파일의 자료를 이용하여 Group별로 회귀분석($Y \sim X1 + X2$)을 실시하고 회귀계수의 월별 추이를 Group로 나누어 그림을 그리기

📋 도전문제

- 🔗 Data폴더에 있는 모든 엑셀 파일의 자료를 읽어 하나의 데이터프레임으로 만듦
- ✓ Data폴더에 있는 엑셀 파일 ^{→ 태그 필요} 확인 (프로그램 ⇔ 수작업)
- ✓ 파일명 가져오기(프로그램 ⇔ 수작업)
- ✓ 엑셀 데이터 가져오기(어떤 Package 사용)
- ✓ 반복 과정을 통해 데이터 프레임에 새로운 데이터 프레임 추가

📝 도전문제

✧ 각 엑셀 파일의 자료를 이용하여 (X1,Y), (X2, Y)의 산점도 2개를 한 화면에 그리고 엑셀 파일명과 동일한 jpg 파일로 저장함

parC
mfrow
=c(2,1,1)
✓ 엑셀파일명과 동일한 파일명의 jpg 파일명 만들기(프로그램 ⇔ 수작업)

✓ 한 화면에 두 산점도(Y~X1), (Y~X2) 그리기

○ 그룹별로 다른 색으로 표시(A 검은색, B 빨간색)

✓ 산점도를 jpg 파일로 저장하기(프로그램 ⇔ 수작업)

✓ 반복과정을 통해 모든 엑셀파일의 자료를 그리기 //



📄 도전문제

- ❖ 각 엑셀 파일의 자료를 이용하여 Group별로 회귀분석($Y \sim X_1 + X_2$)을 실시하고 회귀계수의 월별 추이를 Group로 나누어 그림을 그리기
 - ✓ 회귀분석하고 회귀계수 계산
 - ✓ 그룹별로 회귀분석 (프로그램 ↔ 수작업)
 - ✓ 반복작업을 통해 모든 엑셀파일의 자료를 회귀분석하고 회귀계수를 차례대로 추가
 - ✓ 회귀계수(X_1 과 X_2)를 한 그림에 표시

Data폴더 위치로 변경 → 결과를 저장할 때 이곳에 저장해서 편리함.

setwd("E:/통계프로그래밍/data")

setwd("E:\\\\통계프로그래밍\\\\data") *여들래시 사용시 2개 (escape char과 구분)*

getwd()

*(read_xlsx : xlsx만 가능 (2013 이후)
read_xls : (2013 이전)
read_excel : 자동인식)*

파일 목록 *(현재 내 작업공간에 있는 것)*

파일이름 <- dir()

Excel 파일 확인(xlsx만 사용)

파일개수 <- length(파일이름)

이름분해 <- strsplit(파일이름,"[[:punct:]]") *.을 기준으로 자르기*

- 이름분해: list *파일이름 중간에 .이 있는 경우 3개로 쪼개짐. 이때 마지막 2개만 확인*

확장자 <- sapply(이름분해,function(x) x[length(x)])

엑셀파일 <- 파일이름[확장자 == "xlsx"]

엑셀파일 <- sort(엑셀파일)

확장자를 대문자로 작성한 경우도 있을 수 있음.

tolower(확장자) == "xlsx" 가 좋은 방법

리드값 각각을 집어넣음.

엑셀개수 <- length(엑셀파일)

전체자료 <- NULL

for (i in 1:엑셀개수)

응용코드 파일에 직접 작성

① read_xl 사용 가능 경우

② 파일이름 그대로 넣는 경우

{

dfrm <- xlsx::read.xlsx(엑셀파일[i],sheetIndex=1)

dfrm\$month <- i

전체자료 <- rbind(전체자료,dfrm)

}

빈공간이 있다가 데이터프레임이 들어오면 데이터프레임이 됨

head(전체자료)

(NULL로 만들면 상황에 따라 바뀜)

tail(전체자료)

#####

ggplot2 저장 → ggsave

1부터
그림파일 <- substr(엑셀파일[1],*이름의 전체 길이* *1부터* *.xlsx* *↓* *nchar(엑셀파일)-5*)

그림파일 <- paste(그림파일,".jpg",sep="") *→ 파일명과 동일한 이름의 그림파일 만들기*

for (i in 1:엑셀개수)

{

dfrm <- xlsx::read.xlsx(엑셀파일[i],sheetIndex=1)

jpeg(그림파일[i]) *고급크기 지정도 가능 (픽셀 단위)*

par(mfrow=c(1,2)) *→ 화면 밖으로 나게 됨. 단, 중간에 뭔가 잘못했을 경우 하나씩 어긋나게 됨*

dfrm\$Group <- as.factor(dfrm\$Group)

char로 읽어들이

plot(Y~X1,col=Group,data=dfrm)

다시 legend("topleft",legend=c("A","B"),pch=1,col=c(1,2))

plot(Y~X2,col=Group, dfrm)

모양 색

legend("topright",legend=c("A","B"),pch=1,col=c(1,2))

특정 위치 지정

dev.off() device off
완성했음을 뜻함



by 알아서 2개씩
by(— , 전체 \$ with)
→ 원본 보전

#####

시점 x $X1 * X2 \rightarrow X1, X2, X1 \text{과 } X2 \text{의 상호작용}$

result <- lm(Y~X1+X2,data=dfrm)

anova(result) X1, X2 (주변자)

```
> result <- lm(Y~X1+X2,data=dfrm)
> anova(result)
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)    
X1          1  189308   189308    71.965 1.406e-15 ***
X2          1  5772446   5772446  2194.364 < 2.2e-16 ***
Residuals 273   718148     2631                      
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

<모두 같은 표현>

```
> result[[1]]
(Intercept)      X1      X2
-53.684536   10.361559   5.023466
> coefficients(result)
(Intercept)      X1      X2
-53.684536   10.361559   5.023466
> result$coefficients
(Intercept)      X1      X2
-53.684536   10.361559   5.023466
> result$coeff
(Intercept)      X1      X2
-53.684536   10.361559   5.023466
```

↑
↓

2월 데이터만 있음.

getCoeff <- function(df) {

```
  result <- lm(Y~X1+X2,df)
  beta <- coefficients(result)
  return(beta)
}
```

```
> predict(result)[1:10]
      1      2      3      4      5      6      7      8      9     10
2276.560 2646.476 2655.330 2490.900 2660.514 2383.671 2412.543 2589.272 2404.370 2391.445
```

A그룹 <- NULL

B그룹 <- NULL

for(i in 1:엑셀개수)

```
{
  dfrm <- xlsx::read.xlsx(엑셀파일[i],sheetIndex=1)
  A그룹, B그룹 따지 (총 6개 data)
  dfrm$Group <- as.factor(dfrm$Group)
  result <- by(dfrm, dfrm$Group, getCoeff)
  계수리스트 <- lapply(result,function(x) x)
  (A그룹 <- rbind(A그룹,계수리스트$A))
  (B그룹 <- rbind(B그룹,계수리스트$B))
}
```

```
> result
dfmr$Group: A
(Intercept)      X1      X2
  4.926500   10.01790   4.80853

dfmr$Group: B
(Intercept)      X1      X2
-16.89602    10.19171   5.05132
> 계수리스트
$A
(Intercept)      X1      X2
  4.926500   10.01790   4.80853
$B
(Intercept)      X1      X2
-16.89602    10.19171   5.05132
```

```
> A그룹
(Intercept)      X1      X2
[1,]  20.324174  10.059385  3.089959
[2,]   6.195838  10.224269  3.110504
[3,]   7.318923   9.999566  3.106793
[4,]   4.439471  10.109453  3.103336
[5,]   2.661936  10.347956  5.112366
[6,]  13.224172  10.161717  5.101991
[7,]   2.880790  10.180130  5.104792
[8,]  29.438962   9.954265  5.063199
[9,]  11.095531  10.021163  4.790322
[10,]  4.026789   9.992731  4.815967
[11,]  42.940851   9.946072  4.741105
[12,]  4.926500  10.017901  4.808530
```



회귀계수 = data.frame(A그룹,B그룹)

names(회귀계수) = c("절편A","beta1A","beta2A","절편B","beta1B","beta2B")

par(mfrow=c(2,1),mar=c(5,5,2,5))

plot(A그룹[2],type="l",xlab="월",ylab="X1 회귀계수",main="A그룹 계수추이")

par(new=T) line: 선으로 연결 (둘 다 그래프 경우 both 이용)

plot(A그룹[3],type="l",col="blue", lty=2, axes=F,xlab=NA,ylab=NA)

axis(side = 4)

mtext(side = 4, line = 3, "X2 회귀계수",col="blue")

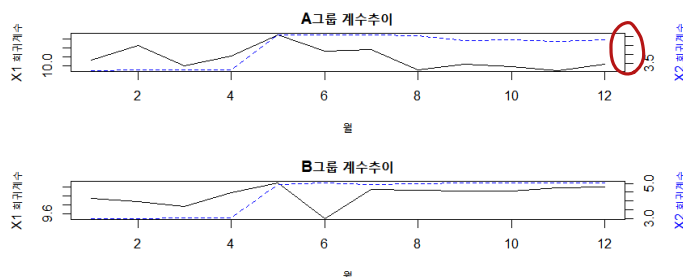
plot(B그룹[2],type="l",xlab="월",ylab="X1 회귀계수",main="B그룹 계수추이")

par(new=T)

plot(B그룹[3],type="l",col="blue", lty=2, axes=F,xlab=NA,ylab=NA)

axis(side = 4)

mtext(side = 4, line = 3, "X2 회귀계수",col="blue")



```
#####
```

```
전체자료 <- NULL
```

```
A그룹 <- NULL
```

```
B그룹 <- NULL
```

```
for (i in 1:엑셀개수)
```

```
{
```

```
  dfrm <- xlsx::read.xlsx(엑셀파일[i],sheetIndex=1)
```

```
  dfrm$month <- i
```

```
  ### 자료 추가
```

```
  전체자료 <- rbind(전체자료,dfrm)
```

```
  ## 그리 그리기 & 저장
```

```
  jpeg(그림파일[i])
```

```
  par(mfrow=c(1,2))
```

```
  plot(Y~X1,col=Group, dfrm)
```

```
  legend("topleft",legend=c("A","B"),pch=1,col=c(1,2))
```

```
  plot(Y~X2,col=Group, dfrm)
```

```
  legend("topright",legend=c("A","B"),pch=1,col=c(1,2))
```

```
  dev.off()
```

```
  # 회귀계수 계산 & 추가
```

```
  result <- by(dfrm, dfrm$Group, getCoff)
```

```
  계수리스트 <- lapply(result,function(x) x)
```

```
  A그룹 <- rbind(A그룹,계수리스트$A)
```

```
  B그룹 <- rbind(B그룹,계수리스트$B)
```

```
}
```