

# Randomization

coop711

2015년 4월 6일

## 출석부 자료로부터 랜덤화하는 과정

e-campus의 온라인 접속횟수 기록을 다운로드 받은 엑셀 파일에서 필요한 부분을 발췌한 자료 읽어들이기.

```
class.roll<-read.table("class_roll.txt", header=TRUE, sep="")
```

red 군에 속할 ID를 랜덤추출하는 과정. 우선, red 에 속할 인덱스를 추출하고, ifelse()를 이용하여 red 군과 black 군 구분. %in% 의 용법에 유의. factor()로 설정하는 이유를 생각해 볼 것.

```
red.id<-sample(1:71, size=35)  
red.id
```

```
## [1]  2 28 44 46 37 35 55 64 12 29 11 47  4 45 18 25  8 61 70 19 26 20  5  
## [24] 51 32 24 54 23 68 49 63 48 15 57 52
```

```
class.roll$group<-factor(ifelse((1:71) %in% red.id, "red", "black"), levels=c("red", "black"))
```

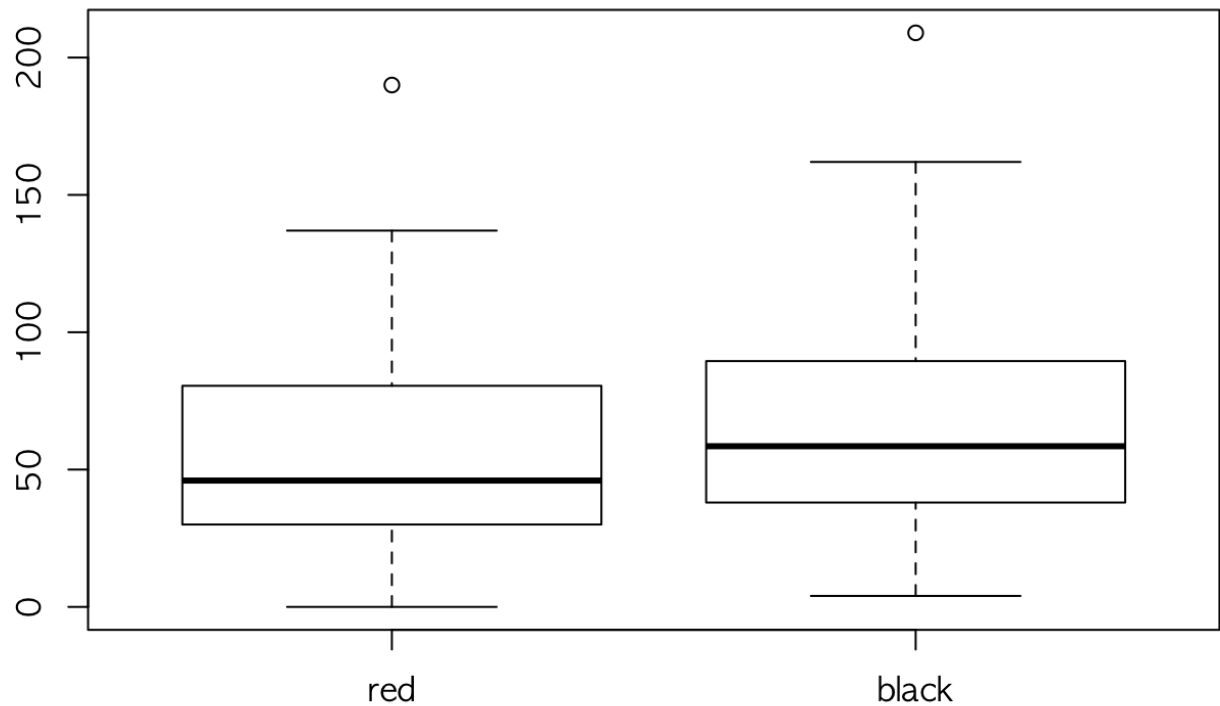
randomization이 적절히 이루어졌는지 기초통계값으로 확인. 만족스럽지 않으면 앞의 과정 반복.

```
tapply(class.roll$n.accesses,class.roll$group,summary)
```

```
## $red  
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##      0.00  30.00   46.00   57.83  80.50  190.00  
##  
## $black  
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##      4.00  38.50   58.50   68.33  87.75  209.00
```

Boxplot 으로 확인.

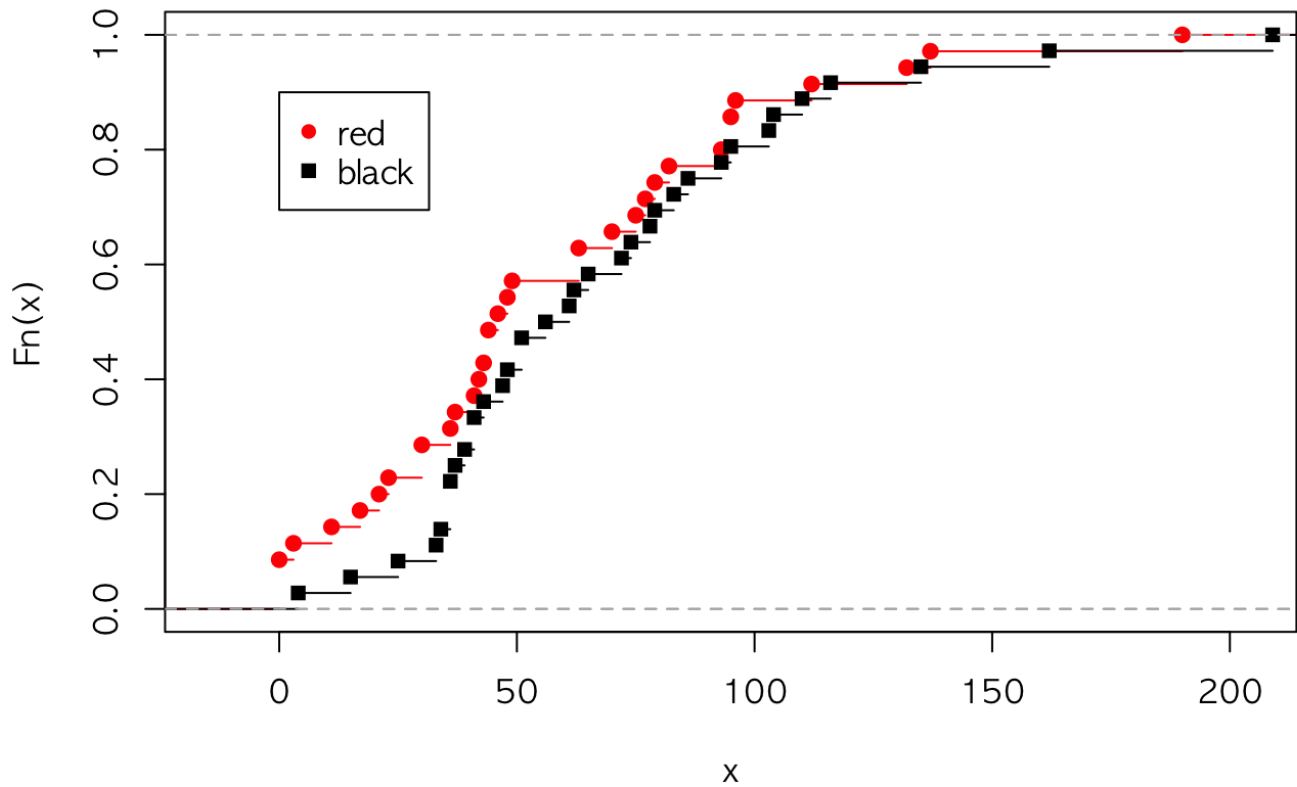
```
boxplot(n.accesses~group,data=class.roll)
```



Empirical CDF 로 확인. 먼저 black.id 를 red.id 아닌 인덱스 로 정의

```
black.id<-(1:71)[-red.id]
plot(ecdf(class.roll$n.accesses[red.id]), col="red",main="표본누적분포함수")
lines(ecdf(class.roll$n.accesses[black.id]), pch=15)
legend(x=0,y=0.9,legend=c("red","black"),pch=16:15,col=c("red","black"))
```

## 표본누적분포함수



13학번 이후와 12학번 이전으로 ID 구분. n.accesses 를 100 회 기준, 혹은 50회 기준으로 나누어 randomization 의 효과 확인. table()의 value를 잘 파악하기 위하여 어떤 조치들을 취하였는지 잘 살필 것. factor()로 굳이 설정한 것, dnn=NULL 이 있고 없과의 차이들을 잘 파악할 것.

```
ID.13<-factor(ifelse(substr(class.roll$ID, 1, 4) >= 2013, "younger.13", "older.13"), levels=c("younger.13", "older.13"))
table(ID.13)
```

```
## ID.13
## younger.13  older.13
##           36         35
```

```
table(ID.13[red.id])
```

```
##
## younger.13  older.13
##           19         16
```

```
table(ID.13[black.id])
```

```
##
## younger.13  older.13
##           17         19
```

```
ID.12<-factor(ifelse(substr(class.roll$ID, 1, 4) >= 2012, "younger.12", "older.12"), levels=c("younger.12", "older.12"))
table(ID.12)
```

```
## ID.12
## younger.12 older.12
##          46        25
```

```
table(ID.12[red.id])
```

```
##
## younger.12 older.12
##          21        14
```

```
table(ID.12[black.id])
```

```
##
## younger.12 older.12
##          25        11
```

```
n.accesses.100<-factor(ifelse(class.roll$n.accesses >= 100, "above 100", "less than 100"), levels=c("above 100", "less than 100"))
table(n.accesses.100)
```

```
## n.accesses.100
##   above 100 less than 100
##           11          60
```

```
table(n.accesses.100[red.id])
```

```
##
##   above 100 less than 100
##           4           31
```

```
table(n.accesses.100[black.id])
```

```
##
##   above 100 less than 100
##           7           29
```

```
n.accesses.50<-factor(ifelse(class.roll$n.accesses >= 50, "above 50", "less than 50"), levels=c("above 50", "less than 50"))
table(n.accesses.50)
```

```
## n.accesses.50
##      above 50 less than 50
##           36           35
```

```
table(n.accesses.50[red.id])
```

```
##
##      above 50 less than 50
##           15           20
```

```
table(n.accesses.50[black.id])
```

```
##
##      above 50 less than 50
##           21           15
```