

Red and Black 170303 : id Masked

coop711

2018-03-20

Data

```
class_roll <- read.table("../data/class_roll_masked.txt",
                        header = TRUE,
                        stringsAsFactors = FALSE,
                        encoding = "UTF-8")

str(class_roll)

## 'data.frame':   160 obs. of  6 variables:
##  $ dept   : chr   "○○학과" "○○학과" "○○학과" "○○학과" ...
##  $ id     : int   20119999 20119999 20179999 20149999 20169999 20129999 20149999 20169999 20179999 2
0129999 ...
##  $ name   : chr   "강○○" "강○○" "강○○" "강○○" ...
##  $ year   : int    4 4 1 4 2 3 4 2 1 3 ...
##  $ email  : chr   "user_name@naver.com" "user_name@hanmail.net" "user_name@naver.com" "user_name@han
mail.net" ...
##  $ cell_no: chr   "010-9164-xxxx" "010-8574-xxxx" "010-6435-xxxx" "010-2066-xxxx" ...
```

Randomization

```
# set.seed(107)
N <- nrow(class_roll)
class_roll$group <- sample(1:N) %% 2 + 1
class_roll$group <- factor(class_roll$group,
                          labels = c("Red", "Black"))

red_id <- which(class_roll$group == "Red")
black_id <- which(class_roll$group == "Black")
```

학번

```
ID_16 <- factor(ifelse(substr(class_roll$id, 1, 4) >= 2016,
                          "younger_16", "older_16"),
                levels = c("younger_16", "older_16"))
kable(table("그룹" = class_roll$group,
            "16학번 기준" = ID_16))
```

	younger_16	older_16
Red	48	32
Black	39	41

```
ID_15 <- factor(ifelse(substr(class_roll$id, 1, 4) >= 2015,
                          "younger_15", "older_15"),
                levels = c("younger_15", "older_15"))
kable(table("그룹" = class_roll$group,
            "15학번 기준" = ID_15))
```

	younger_15	older_15
Red	53	27
Black	47	33

```
ID_14 <- factor(ifelse(substr(class_roll$id, 1, 4) >= 2014,
                          "younger_14", "older_14"),
                levels = c("younger_14", "older_14"))
kable(table("그룹" = class_roll$group,
            "14학번 기준" = ID_14))
```

	younger_14	older_14
Red	62	18
Black	58	22

```
ID_13 <- factor(ifelse(substr(class_roll$id, 1, 4) >= 2013,
                          "younger_13", "older_13"),
                levels = c("younger_13", "older_13"))
kable(table("그룹" = class_roll$group,
            "13학번 기준" = ID_13))
```

	younger_13	older_13
Red	75	5
Black	71	9

e-mail 서비스업체

```
email_list <- strsplit(class_roll$email, "@", fixed = TRUE)
mail_com <- sapply(email_list, `[`, 2)
kable(table("그룹" = class_roll$group,
            "e-mail" = mail_com))
```

	daum.net	gmail.com	hanmail.net	nate.com	naver.com
Red	2	3	5	4	66
Black	0	3	4	3	69

성씨 분포

```
f_name <- substring(class_roll$name,
                    first = 1, last = 1)
kable(table("Group" = class_roll$group,
            "Family Name" = f_name))
```

	강	고	구	권	김	나	명	문	박	반	방	배	서	성	손	송	신	심	안	양	우	유	윤	이	임	장	전	정	조	차	최	하	한
Red	3	1	1	1	22	1	1	0	6	1	0	1	3	1	2	2	0	1	3	1	1	1	0	10	3	2	0	3	4	2	2	1	0
Black	3	0	0	3	14	0	0	1	7	0	1	1	1	0	0	1	3	0	2	0	0	4	5	11	0	1	2	5	3	0	7	0	2

많이 나오는 성씨

```
f_name_f <- factor(ifelse(f_name %in% c("김", "이", "박"),
                          f_name, "기타"),
                  levels = c("김", "이", "박", "기타"))
kable(table("Group" = class_roll$group,
            "Family Name" = f_name_f))
```

	김	이	박	기타
Red	22	10	6	42
Black	14	11	7	48

출석부에서 20명 비복원 랜덤 표집

```
# set.seed(1)
kable(class_roll[sample(1:nrow(class_roll), size = 20), ])
```

	dept	id	name	year	email	cell_no	group
5	○○학과	20169999	강○○	2	user_name@naver.com (mailto:user_name@naver.com)	010-8820-xxxx	Red
140	○○학과	20169999	조○○	2	user_name@naver.com (mailto:user_name@naver.com)	010-2425-xxxx	Black
74	○○학과	20179999	손○○	1	user_name@hanmail.net (mailto:user_name@hanmail.net)	010-9450-xxxx	Red
76	○○학과	20179999	송○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-8914-xxxx	Black
58	○○학과	20149999	박○○	4	user_name@naver.com (mailto:user_name@naver.com)	010-7515-xxxx	Black
98	○○학과	20169999	윤○○	2	user_name@naver.com (mailto:user_name@naver.com)	010-5150-xxxx	Black
117	○○학과	20169999	이○○	2	user_name@naver.com (mailto:user_name@naver.com)	010-4650-xxxx	Black
49	○○학과	20179999	나○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-2912-xxxx	Red
27	○○학과	20129999	김○○	4	user_name@hanmail.net (mailto:user_name@hanmail.net)	010-8505-xxxx	Black
109	○○학과	20179999	이○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-7663-xxxx	Red
138	○○학과	20169999	조○○	2	user_name@naver.com (mailto:user_name@naver.com)	010-8611-xxxx	Red
113	○○학과	20179999	이○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-5113-xxxx	Black
15	○○학과	20169999	김○○	2	user_name@naver.com (mailto:user_name@naver.com)	010-9264-xxxx	Red
141	○○학과	20179999	조○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-3467-xxxx	Black
81	○○학과	20139999	신○○	4	user_name@naver.com (mailto:user_name@naver.com)	010-9990-xxxx	Black
133	○○학과	20139999	정○○	3	user_name@naver.com (mailto:user_name@naver.com)	010-6343-xxxx	Black
90	○○학과	20139999	유○○	3	user_name@naver.com (mailto:user_name@naver.com)	010-9984-xxxx	Black
4	○○학과	20149999	강○○	4	user_name@hanmail.net (mailto:user_name@hanmail.net)	010-2066-xxxx	Red
158	○○학과	20159999	황○○	3	user_name@naver.com (mailto:user_name@naver.com)	010-5656-xxxx	Black
13	○○학과	20179999	김○○	1	user_name@naver.com (mailto:user_name@naver.com)	010-2386-xxxx	Red

set.seed() 의 용법

set.seed() 를 이용하면 랜덤넘버에 의존하는 실험을 재현할 수 있다. 다음 코드를 반복 수행하거나 다른 사람들의 수행결과와 비교해 보라.

세 결과가 모두 다른 경우

```
sample(1:6, size = 2)

## [1] 5 1

sample(1:6, size = 2)

## [1] 4 5

sample(1:6, size = 2)

## [1] 1 5
```

세 번의 수행 결과가 똑같이 반복되는 경우

```
set.seed(1)
sample(1:6, size = 2)

## [1] 2 6

sample(1:6, size = 2)

## [1] 4 5

sample(1:6, size = 2)

## [1] 2 5
```

동일한 결과를 반복적으로 얻는 경우

```
set.seed(1)
sample(1:6, size = 2)

## [1] 2 6

set.seed(1)
sample(1:6, size = 2)

## [1] 2 6

set.seed(1)
sample(1:6, size = 2)

## [1] 2 6
```