

815회차 로또 번호 예측

기존의 데이터를 이용한 예측

김규하
송우정
장성문
최주예

INDEX

2009-2018
데이터분석

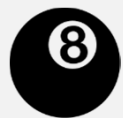
2009-2018
데이터재분석

더 정확한
결과를 위한방안

결론

pre

Lotto 1등의 확률은?



$$\frac{6}{45}$$



$$\frac{1}{66}$$



$$\frac{2}{1419}$$



$$\frac{1}{9933}$$



$$\frac{2}{407253}$$



$$\frac{1}{8145060}$$



이전 데이터를 **분석**하면 당첨 확률을 높일 수 있지 않을까?

2009 – 2018
데이터 분석

1

왜 2009년 이후의 데이터?

최신의 데이터를 중심으로
좀 더 정확한 분석을 하기
위함

2009년 이후로 로또의
추첨에 관련된 물리적인
변화가 있었음

2009 ~ 2018 년도의 데이터를 효율적으로 사용하자!

1

데이터 분석하기

```
1 setwd('C:/Rwork')
2 lotto = read.csv('lotto.csv')
3 head(lotto)
4
5 dat.2018 <- lotto[1:12,14:19]
6 dat.2017 <- lotto[13:64,14:19]
7 dat.2016 <- lotto[65:117,14:19]
8 dat.2015 <- lotto[118:169,14:19]
9 dat.2014 <- lotto[170:221,14:19]
10 dat.2013 <- lotto[222:273,14:19]
11 dat.2012 <- lotto[274:325,14:19]
12 dat.2011 <- lotto[326:378,14:19]
13 dat.2010 <- lotto[379:430,14:19]
14 dat.2009 <- lotto[431:482,14:19]
15
```

분석에 이용되는 기본 데이터를 불러 오는 코드이다. 불러온 데이터를 lotto 변수에 저장한다

Lotto에 존재하는 데이터를 각각의 년도에 맞게 행과 열로 분리하여 년도에 해당하는 변수에 저장한다. 이때 데이터의 형식은 data.frame으로 지정된다.

1

데이터 분석하기

각 년도에 해당하는 변수들을 matrix
형식으로 변화시켜 table을
이용하여 빈도수를 구한다

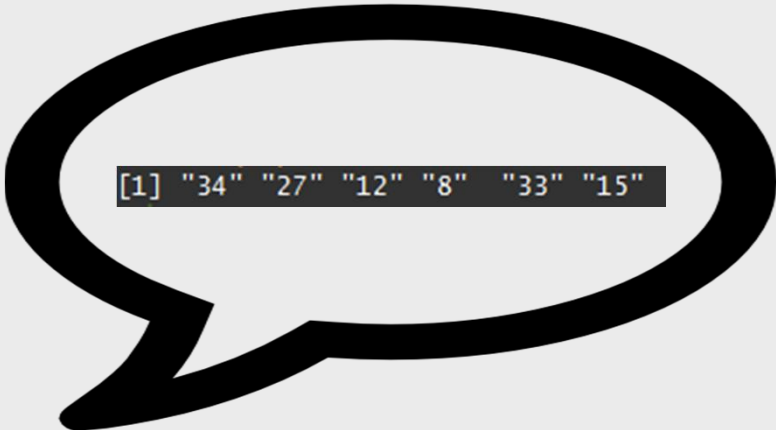
이 때의 변수들은 table
형식이기에 '+' 를 사용하여
하나의 matrix로 결합한다.

```
16 # 기존 데이터
17 num_2009 <- as.matrix(t(table(as.matrix(dat.2009))))
18 num_2010 <- as.matrix(t(table(as.matrix(dat.2010))))
19 num_2011 <- as.matrix(t(table(as.matrix(dat.2011))))
20 num_2012 <- as.matrix(t(table(as.matrix(dat.2012))))
21 num_2013 <- as.matrix(t(table(as.matrix(dat.2013))))
22 num_2014 <- as.matrix(t(table(as.matrix(dat.2014))))
23 num_2015 <- as.matrix(t(table(as.matrix(dat.2015))))
24 num_2016 <- as.matrix(t(table(as.matrix(dat.2016))))
25 num_2017 <- as.matrix(t(table(as.matrix(dat.2017))))
26 num_2018 <- as.matrix(t(table(as.matrix(dat.2018))))
27
28 a <- num_2009+num_2010+num_2011+num_2012+
29   num_2013+num_2014+num_2015+num_2016+num_2017
--
```


1

데이터 도출하기

```
34 a <- t(as.matrix(a))  
35 a <- as.numeric(a)  
36 b <- 1:45  
37 names(a) <- b  
38 barplot(a,ylim = c(0,110))  
39 x2 <- head(sort(a, decreasing = T),6)  
40  
41 x2
```



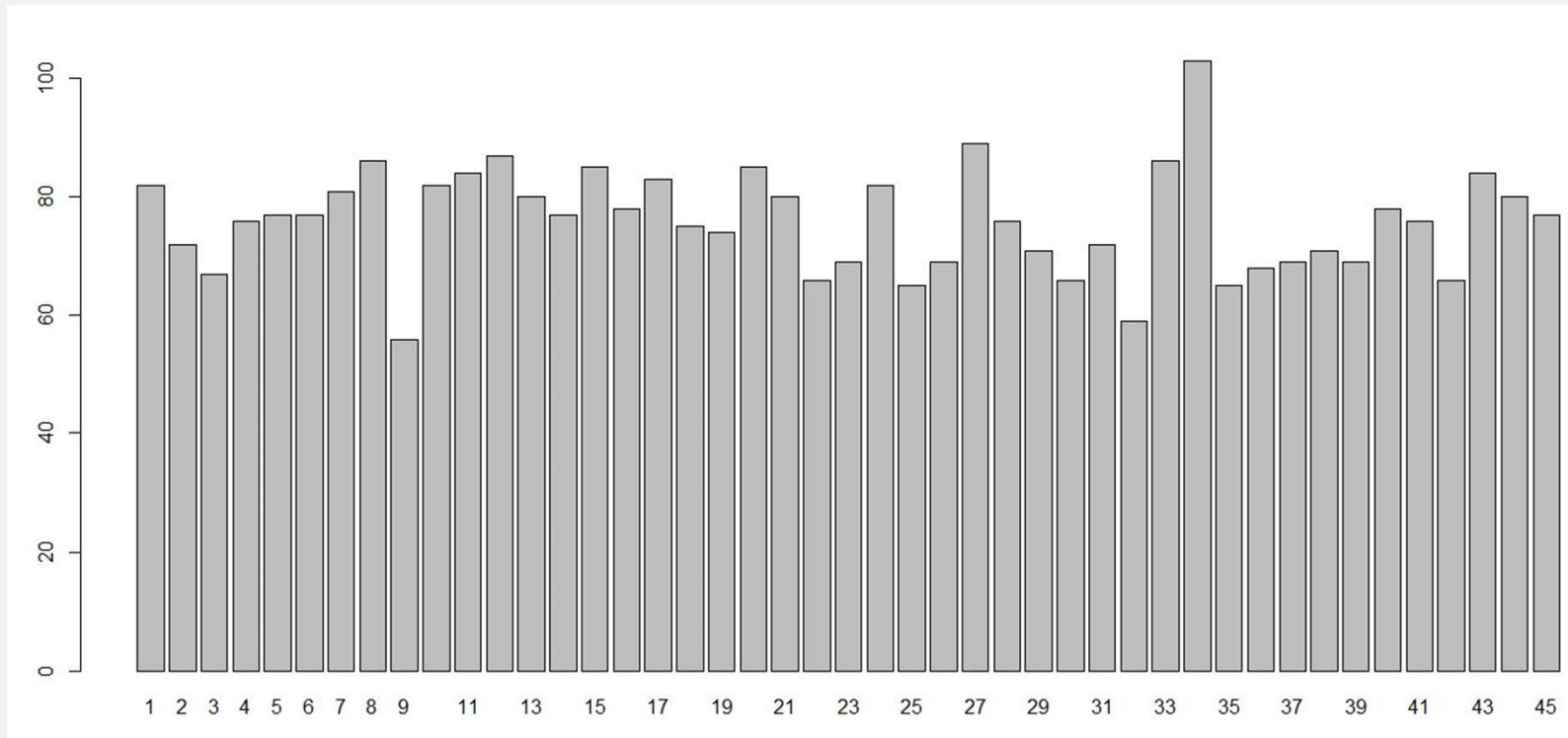
```
[1] "34" "27" "12" "8"  "33" "15"
```



이를 통해 기존의 10년치 데이터의 빈도수에 의한
가장 당첨확률이 높은 숫자 6개를 도출 할 수 있다

1

그래프를 이용한 결과



〈 앞의 코드를 이용해 도출된 번호 별 빈도수〉

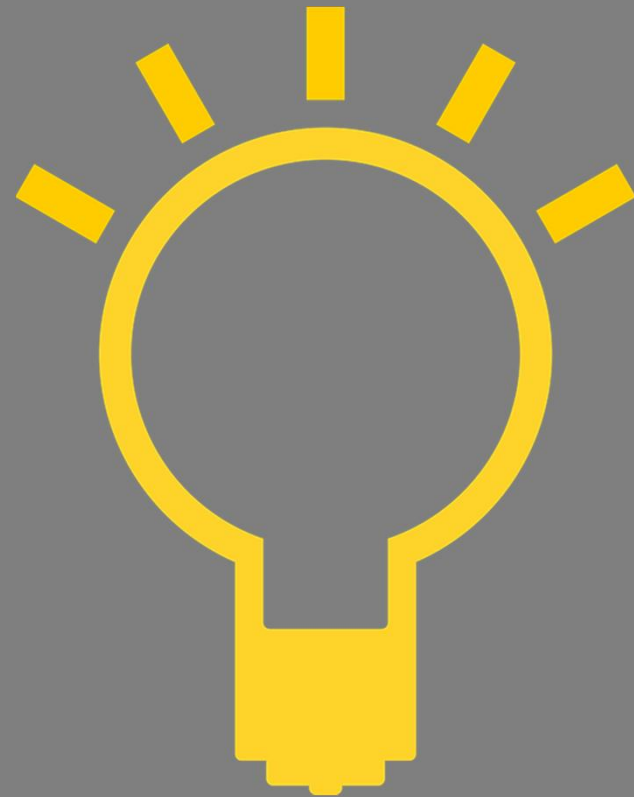


하지만 모든 데이터가 동일한 가치를
가지지 않을 텐데 그럼 각자의 데이터에
차이를 두어야 하지 않을까?

각 데이터마다 **가중치**를 다르게 주면 더
정확한 결과를 얻을 수 있겠다!

2018년에 가까운 데이터들에게는 큰 가중치
를, 멀어질수록 가중치를 줄여가면 되겠다!

가중치를 데이터마다 다르게 부여하게
된다면 가중치가 크게 할당 받은 항목에
좀 더 큰 가치를 부여하게 됨으로써
우리가 원하는 데이터에 대해 기존보다 더
적합한 결과를 얻을 수 있을 거야



2009 – 2018
데이터 재분석

3

데이터 재분석하기

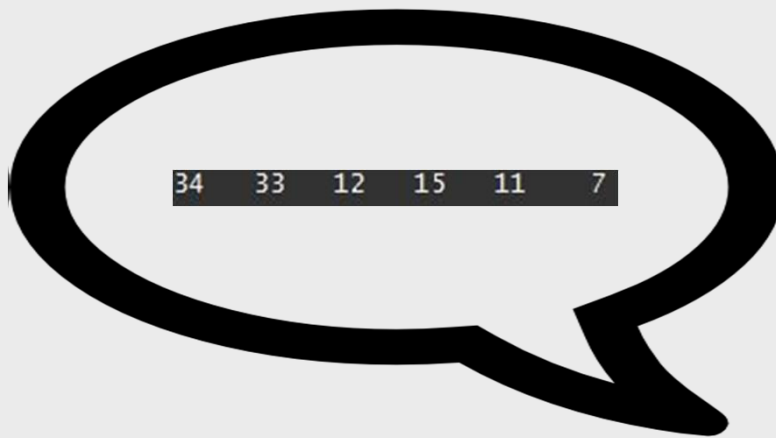
가중치 데이터

```
num_2009 <- as.matrix(t(table(as.matrix(dat.2009))))*0.1  
num_2010 <- as.matrix(t(table(as.matrix(dat.2010))))*0.2  
num_2011 <- as.matrix(t(table(as.matrix(dat.2011))))*0.3  
num_2012 <- as.matrix(t(table(as.matrix(dat.2012))))*0.4  
num_2013 <- as.matrix(t(table(as.matrix(dat.2013))))*0.5  
num_2014 <- as.matrix(t(table(as.matrix(dat.2014))))*0.6  
num_2015 <- as.matrix(t(table(as.matrix(dat.2015))))*0.7  
num_2016 <- as.matrix(t(table(as.matrix(dat.2016))))*0.8  
num_2017 <- as.matrix(t(table(as.matrix(dat.2017))))*0.9  
num_2018 <- as.matrix(t(table(as.matrix(dat.2018))))*1
```

다른 부분은 기존의 코드와 유사하지만
이번에는 각각의 빈도수에 현재부터
과거 순으로 가중치를 줄여가며
빈도에 대한 변화를 준다

3

데이터 다시 도출하기



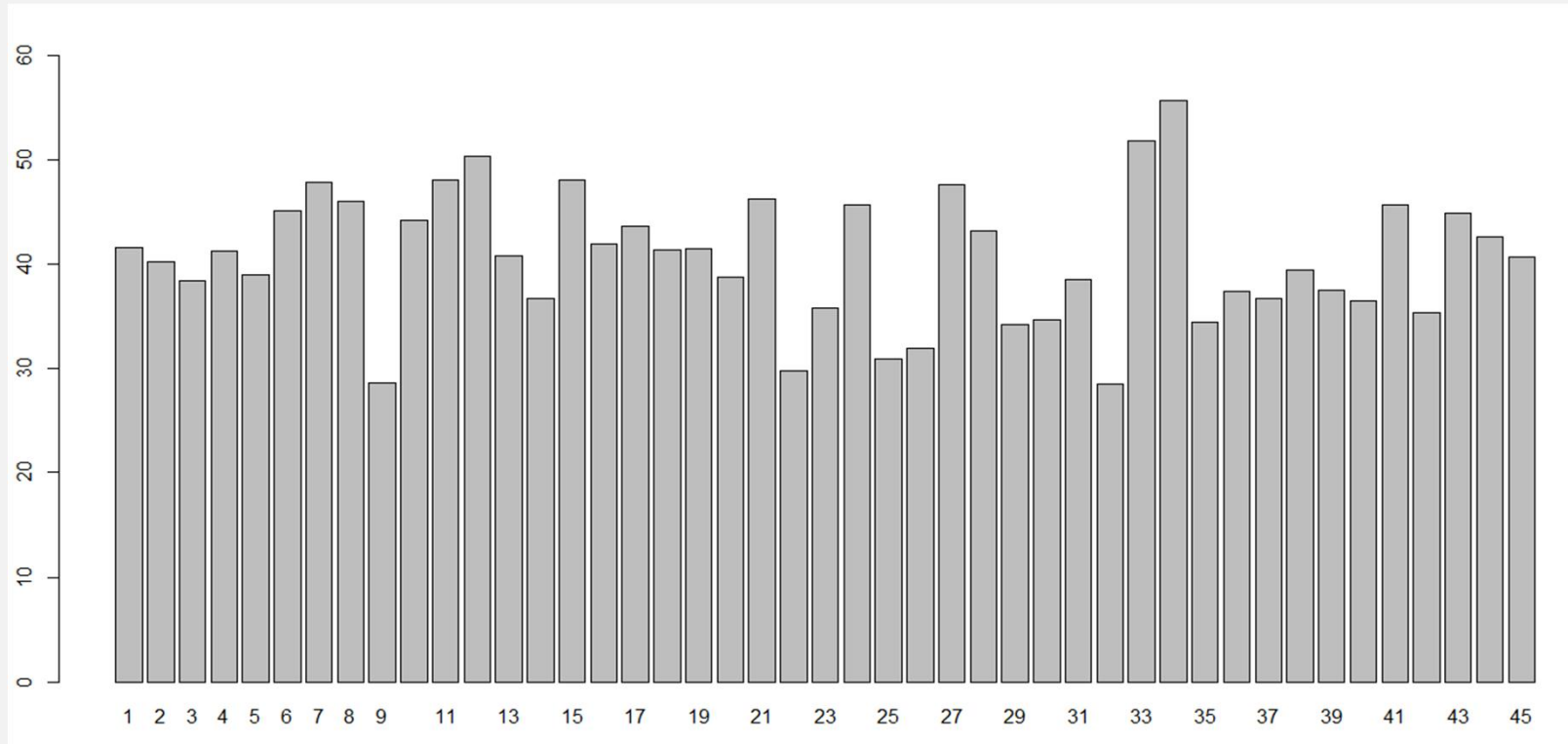
| | origin | weight |
|------|--------|--------|
| [1,] | "34" | "34" |
| [2,] | "27" | "33" |
| [3,] | "12" | "12" |
| [4,] | "8" | "15" |
| [5,] | "33" | "11" |
| [6,] | "15" | "7" |
| . | ... | ... |



기존의 6개의 숫자와 달리 가중치를 부여했을 때의 당첨번호가
변경된 것을 볼 수 있으며 기존에 존재하던 번호들도 순위가 변경된 것을 볼 수 있다.

3

가중치를 이용한 결과의 그래프

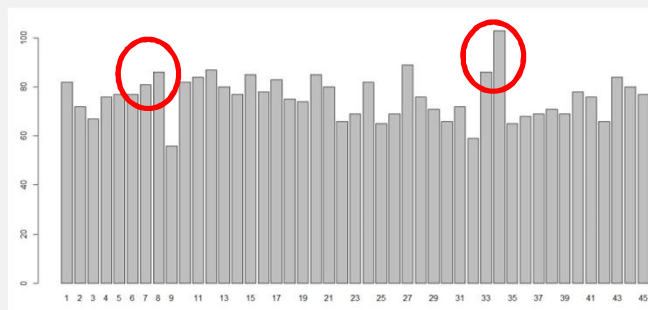


〈 앞의 코드를 이용해 도출된 번호 별 변화된 빈도수 〉



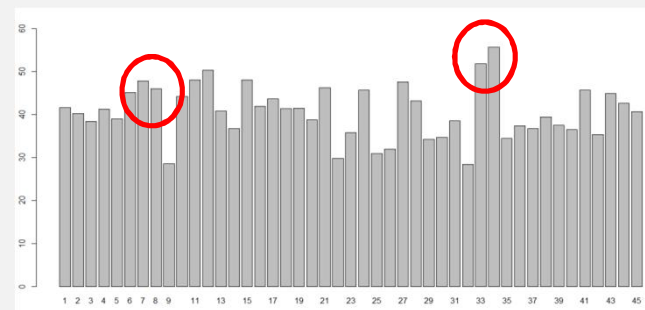
결론

original



$\langle 34, 27, 12, 8, 33, 15 \rangle$

weight



$\langle 34, 33, 12, 15, 11, 7 \rangle$

감사합니다

