

Birthday Problem

데이터테크전공 20173204 곽명빈

2020-12-01

Data

	A	B	C	D	E
1					
2	학(적)번	Q1	Q2	팀	
3	20196201	나	1,7	Black	
4	20196202	나	11,15	Black	
5	20195105	나	2,23	Black	
6	20196203	가	10,31	Black	
7	20165207	나	7,15	Black	
8	20196204	가	11,14	Red	
9	20196205	나	7,11	Red	
10	20196206	나	4,16	Red	
11	20196207	가	5,6	Black	
12	20196208	나	4,30	Red	
13	20196209			Black	
14	20196210	가	9,1	Red	
15	20196211	가	9,7	Red	
16	20196212	가	9,1	Black	
17	20196214	나	9,29	Red	
18	20196215	나	3,31	Black	
19	20196216	나	11,5	Red	
20	20196217	나	2,3	Red	
21	20196218	나	10,24	Red	
22	20133303			Red	
23	20196219	나	11,18	Black	
24	20196220	나	7,11	Black	

생일 문제의 질문

() 학과 학번 () 이름 ()

1. 통계학입문 수업 듣는 학생 중 나와 생일이 같은 사람은?

- 가. 있을 것이다.
- 나. 없을 것이다.

2. 나의 생일은 ()월 ()일이다.

Q2. Birthday Problem

생일이 같은 사람의 수효

학번	생일	그룹
20196278	01월22일	Black
20196283	01월22일	Red
20196205	07월11일	Red

학번	생일	그룹
20196220	07월11일	Black
20183940	07월12일	Black
20196277	07월12일	Red
20196210	09월01일	Red
20196212	09월01일	Black
20196225	09월01일	Black
20196203	10월31일	Black
20196285	10월31일	Black

어느 날에 몇 명씩 생일이 같은가?

01월22일	07월11일	07월12일	09월01일	10월31일	계
2	2	2	3	2	11

생일이 같은 사람은 몇 명 정도 기대되는가?

N 을 전체 인원이라 할 때, 기대 인원은 $N \times \{1 - (\frac{364}{365})^{N-1}\}$, 분산은 $N \times \{1 - (\frac{364}{365})^{N-1}\} + N \times (N - 1) \times \{1 - (\frac{363}{365})^{N-2}\}$ 로 계산된다.

무응답이거나 결석한 학생을 제외한 응답 인원 62명에 대하여 기대인원을 계산하면 9.6명, 표준오차는 3.1명으로 계산되어 관찰된 값이 그 범위 근처에 있음을 알 수 있다.

기대값의 계산

```
## [1] 9.6
```

표준오차의 계산

```
## [1] 3.1
```

태어난 달의 분포는?

	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월	계
Red	3	5	1	4	2	4	1	5	2	3	1	1	32
Black	3	4	2	2	6	2	2	2	2	0	5	0	30
계	6	9	3	6	8	6	3	7	4	3	6	1	62

랜덤화 효과 검증

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates): .

Test statistic	df	P value
12.01	NA	0.3913

월별로 고르게 출생하였는가?

Chi-squared test for given probabilities with
simulated p-value (based on 2000 replicates): .

Test statistic	df	P value
11.94	NA	0.3833

Q1. 나와 생일 같은 사람이 있을까?

나와 생일이 같은 사람이 있겠느냐는 질문에 3/4은 없을 것이라고 답을 했습니다. 생일이 같은 사람이 11명이나 되는 데 대해서 어떻게 생각할까요?

	있을 것이다	없을 것이다	결석	계
Red	8	24	7	39
Black	8	22	7	37
계	16	46	14	76

Pearson's Chi-squared test with Yates'
continuity correction: .

Test statistic	df	P value
0	1	1

% 비교.

	있을 것이다	없을 것이다	계
Red	25.0	75.0	100.0
Black	26.7	73.3	100.0

% 합계

	있을 것이다	없을 것이다	계
계	25.8	74.2	100.0

Comments

이번 시간에는 나와 생일이 같은 사람이 몇 명이나 있을까? 라는 주제에 대해 공부할 수 있었습니다. 처음 수업을 듣기 전에는 나와 생일이 같은 사람은 없다고 생각을 하였습니다. 하지만 생일이 같은 사람의 수효를 살펴본 결과 생일이 같은 사람이 11명이나 된다는 사실에 놀라웠습니다. 기댓값과 표준오차를 계산해본 결과 11에 근접한다는 사실을 알 수 있었습니다. piping과 piping 아닌 것 두 개 다 올려주셔서 아직 piping이 완전히 익숙하지 않지만, 더 쉽게 이해할 수 있었습니다. 데이터를 월 일로 분리 후에 정리한 다음 날짜순으로 정리하는 방법을 배울 수 있었습니다. 그 덕분에 데이터를 쉽게 이해할 수 있게 되었습니다. 중복된 생일을 찾고 집계하는 방법을 알게 되었습니다. 많은 사람이 나와 같은 생일을 가진 사람은 없을 것이라는 답변을 한 걸 보니 다 비슷한 생각을 가진다는 것을 알 수 있었습니다.