

Data Analysis

가설

종교 유무에 따라
이혼율은 다를 것이다

데이터 분석 설계

| 목차

- 분석 주제 선정
- 가설에 따른 분석 가능 변수 구성
- 분석 항목 결정

데이터 분석 설계

분석 주제 선정

종교 유무에 따라 이혼율은 다를 것이다.

가설에 따른 분석 가능 변수 구성

welfare → 가구원과 관련된 여러가지 데이터

rel_marriage → 종교와 결혼 변수의 데이터 값

religion → 종교유무

marriage → 혼인상태

rel_div → 종교에 유무에 따라 이혼한 사람 수

rel_mar → 종교에 유무에 따라 이혼하지 않은 사람 수

rel_mar_div → 종교에 유무에 따라 이혼한 사람 수와 이혼하지 않은 사람 수

rate → 종교에 유무에 따라 결혼한 사람 중 이혼한 사람 비율

분석 항목 결정

religion과 rate와의 관계 분석

데이터 준비

| 목차

- 데이터 불러오기
- 데이터 형태 파악하기

데이터 준비

데이터 불러오기

- 방법
공공 데이터 포털에서 'Koweps_hpc10_2018_beta1.sav'와 'Koweps_Codebook.xlsx' 파일을 다운받는다.
R script에서 다운받은 두 개의 파일을 불러온다.
- 개념설명
'Koweps_hpc10_2018_beta1.sav'에는 가구원과 관련된 여러가지 data들이 저장되어 있다.
'Koweps_Codebook.xlsx'에는 'Koweps_hpc10_2018_beta1.sav'에서 사용된 변수들의 내용을 설명해주는 표이다.
- 명령코드

```
welfare <- read.spss(file = "Koweps_hpc10_2018_beta1.sav", to.data.frame = T)
codebook <- read_excel("c:/Rstudy/Koweps_Codebook.xlsx")
```
- 명령코드 설명
"Koweps_hpc10_2018_beta1.sav" 파일을 data frame 형태로 불러와 welfare 변수에 저장한다.
"Koweps_Codebook.xlsx" 파일을 불러와 codebook 변수에 저장한다.

데이터 준비

데이터 형태 파악하기

- 개념 설명

welfare에 어떤 변수가 어떠한 형태로 있는지 확인한다.

codebook 을 열어 종교, 혼인상태에 해당하는 변수를 파악한다.

- 명령 코드

```
View(welfare)
str(welfare)
View(codebook)
```

- 명령 코드 설명

View(welfare)를 사용하여 welfare의 전체적인 형태를 확인한다.

str(welfare)를 사용하여 어떤 변수가 어떠한 형태로 있는지 간단하게 확인한다.

View(codebook)을 사용하여 열어 종교와 혼인상태에 해당하는 변수와 그에 해당하는 데이터의 의미를 파악한다.

데이터 준비

데이터 형태 파악하기

- 명령 코드 결과
codebook을 열어보면 다음과 같은 내용을 확인 가능하다.

	변수명	설명	내용
1	h10_g3	성별	1.남 2.여
2	h10_g4	태어난 연도	년
3	h10_g10	혼인상태	0.비해당(18세 미만) 1.유배우 2.사별 3.이혼 4.별거 5.미혼(18세이상, 미혼모 포함) 6.기타(사망 등)
4	h10_g11	종교	1.있음 2.없음
5	h10_eco9	직종	직종 코드표 참조
6	p1002_8aq1	일한달의 월 평균 임금	만원
7	h10_reg7	7개 권역별 지역구분	1. 서울 2. 수도권(인천/경기) 3. 부산/경남/울산 4.대구/경북 5. 대전/충남 6. 강원/충북 7.광주/전남/...

종교는 h10_g11 변수를 사용하고 있으며, 1이면 종교 있음, 2이면 종교 없음을 나타낸다.
혼인상태는 h10_g10 변수를 사용하고 있으며, 3일 때 이혼을 나타낸다.
종교와 혼인상태 모두 9일 때는 모름/무응답을 나타낸다.

데이터 가공 및 통합

| 목차

- 필요한 변수의 이름 변경한 후 추출
- 필요한 데이터 이름 변경한 후 불필요한 데이터 제거
- 데이터 값에 따라 그룹화하고 카운트 한 후 가로 결합
- 결혼한 사람 중 이혼한 사람의 비율을 구하기

데이터 가공 및 통합

필요한 변수의 이름 변경한 후 추출

- 방법

marriage : h10_g10

religion : h10_g11

- 명령 코드

```
welfare <- rename(welfare, marriage = h10_g10, religion = h10_g11)  
rel_marriage <- welfare %>% select(religion, marriage)
```

데이터 가공 및 통합

필요한 변수의 이름 변경한 후 추출

- 명령 코드 설명
welfare 변수 중에 h10_g11을 religion 으로, h10_g10을 marriage 로 이름을 변경한다.
welfare에서 religion과 marriage 변수 데이터들만 뽑아서 rel_marriage에 저장한다.
- 명령 코드 결과 (rel_marriage)

	religion	marriage
1	2	2
2	2	2
3	2	2
4	1	3
5	1	2
6	1	1
7	1	1
8	1	0
9	1	1
10	1	1

데이터 가공 및 통합

필요한 데이터 이름 변경한 후 불필요한 데이터 제거

- 방법

- 1. Religion 변수 데이터 이름 변경하기

- have_religion : 1(있음)

- no_religion : 2(없음)

- 2. Marriage 변수 데이터 이름 변경하기

- no_divorce : 1(유배우), 2(사별), 4(별거)

- divorce : 3(이혼)

- NA : 0(비해당, 18세 미만), 5(미혼, 18세 이상), 6(기타)

- 결혼 했던 사람 중에 이혼한 사람의 비율을 구할 것 이기 때문에 0, 5, 6 데이터를 제거

- 명령 코드

```
rel_marriage$religion <- ifelse(rel_marriage$religion == 1, 'have_religion', ifelse(rel_marriage$religion == 2, 'no_religion', NA))
rel_marriage$marriage <- ifelse(rel_marriage$marriage == 3, 'divorce', ifelse(rel_marriage$marriage >= 1 &
rel_marriage$marriage <= 4, 'no_divorce', NA))
rel_marriage <- rel_marriage %>% filter(!is.na(marriage) & !is.na(religion))
```

데이터 가공 및 통합

필요한 데이터 이름 변경한 후 불필요한 데이터 제거

- 명령 코드 설명
religion 변수 데이터가 1이면 have_religion로, 2이면 no_religion로, 그 외는 NA로 변경
marriage 변수 데이터가 3이면 divorce로, 1, 2, 4이면 no_divorce로, 그 외는 NA로 변경
religion과 marriage의 결측치(NA) 제거
- 명령 코드 결과 (rel_marriage)

	religion	marriage
1	no_religion	no_divorce
2	no_religion	no_divorce
3	no_religion	no_divorce
4	have_religion	divorce
5	have_religion	no_divorce
6	have_religion	no_divorce
7	have_religion	no_divorce
8	have_religion	no_divorce
9	have_religion	no_divorce
10	no_religion	no_divorce

데이터 가공 및 통합

데이터 값에 따라 그룹화하고 카운트 한 후 가로결합

- 방법
이혼한 사람들을 종교유무에 따라 그룹화 한 후 그 수를 카운트
배우가 있는 사람들을 종교유무에 따라 그룹화한 후 그 수를 카운트
종교를 기준으로 가로 결합한다.
- 명령 코드

```
rel_div <- rel_marriage %>% filter(marriage == 'divorce') %>% group_by(religion) %>% summarise(divorce=n())  
rel_noDiv <- rel_marriage %>% filter(marriage == 'no_divorce') %>% group_by(religion) %>% summarise(no_divorce=n())  
rel_mar <- full_join(rel_div, rel_noDiv, by='religion')
```

데이터 가공 및 통합

데이터 값에 따라 그룹화하고 카운트 한 후 가로결합

- 명령 코드 설명
 - rel_marriage에서 marriage 데이터 값이 divorce인 것을 religion에 따라 그룹화한 후 그 수를 카운트해서 rel_div에 저장
 - rel_marriage에서 marriage 데이터 값이 marriage인 것을 religion에 따라 그룹화한 후 그 수를 카운트해서 rel_noDiv에 저장
 - rel_div와 rel_noDiv를 religion을 기준으로 가로결합한 값을 rel_mar에 저장
- 명령 코드 결과 (rel_mar)

	religion	divorce	no_divorce
	<chr>	<int>	<int>
1	have_religion	328	5536
2	no_religion	384	5096

데이터 가공 및 통합

결혼한 사람 중 이혼한 사람의 비율을 구하기

- 방법

$$\text{결혼한 사람 중 이혼한 사람의 비율} = \frac{\text{이혼한 사람}}{\text{이혼한 사람 수와 이혼하지 않은 사람 수의 합}} \times 100 \%$$

- 명령 코드

```
rel_mar$rate <- ( rel_mar$divorce / (rel_mar$divorce + rel_mar$marriage) ) * 100
```

- 명령 코드 설명

결혼한 사람 중 이혼한 사람의 비율을 구해 rate 파생변수에 저장

- 명령 코드 결과 (rel_mar)

데이터 분석 및 결론

| 목차

- 그래프를 통해 분석 결과 시각화
- 결론

데이터 분석 및 결론

그래프를 통해 분석 결과 시각화

-방법

ggplot을 이용해서 종교유무와 결혼한 사람 중 이혼한 사람의 비율과의 관계를 시각화 해보기

-명령코드

```
ggplot(rel_mar, aes(x=religion, y=rate)) + geom_col() + ylab('divorce rate (%)')
```

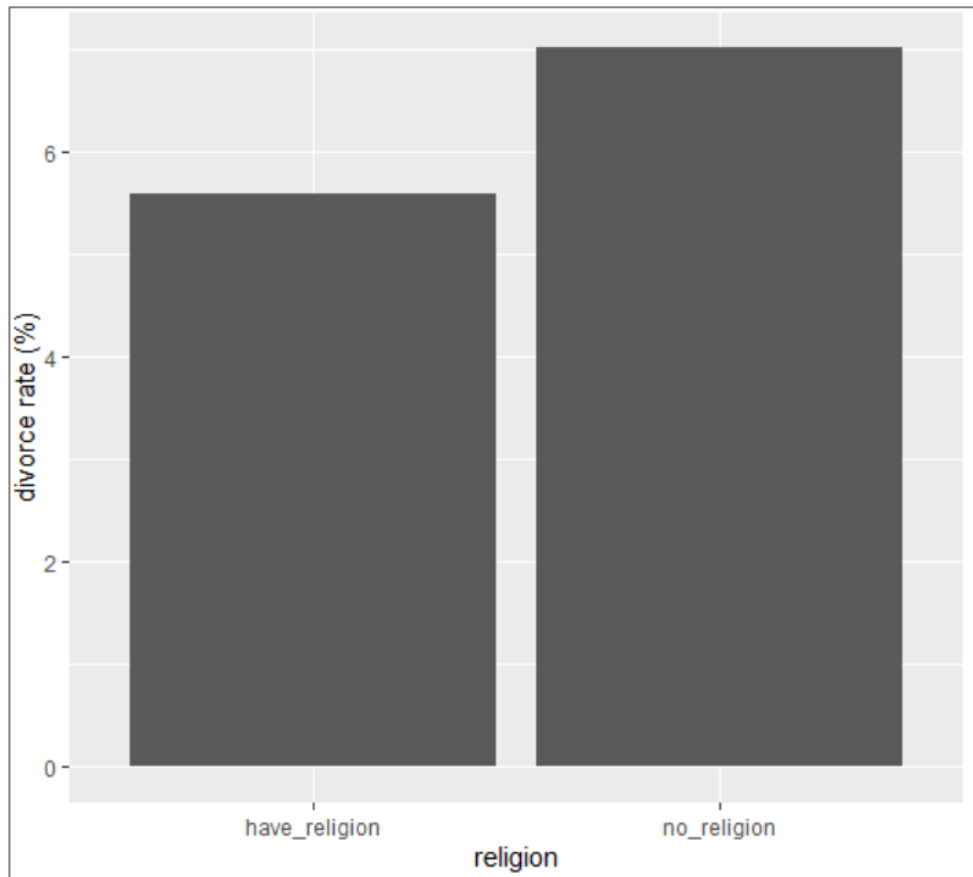
-명령 코드 설명

x좌표에 religion, y좌표에 rate 값을 놓고, 막대그래프를 그려 시각화 한다.

데이터 분석 및 결론

그래프를 통해 분석 결과 시각화

- 명령 코드 결과



데이터 분석 및 결론

결론

종교 유무에 따라 이혼율이 다를 것이다 라는 가설은 참으로 증명되었다.
또한 종교를 가진 사람보다 종교를 가지지 않은 사람이 이혼율이 더 높다는 것을 알 수 있다.