

2021 학년도 2학기 중간과제물(온라인제출용)

교과목명 : 국가통계의 이해

학 번 : 202135-367895

성 명 : 김태정

연 락 처 : 010-4172-4516

o 과제유형 : (공통) 형

o 과 제 명 : 중간과제물 문제

- 이하 과제 작성

1번문제

합계 출산율(TFR)은 한 명의 가임기(15세 부터 49세 까지)의 여성이 평생 평균 몇 명의 자녀를 낳는 지를 나타내는 지표로서 연령별 출산율(ASFR)을 합하여 구한다. 따라서 연령별 출산율을 이용하여 구할 수 있다.

먼저 연령별 출산율(ASFR)을 구하면 아래와 같다. 연령별 출산율은 당해 가임기 여성(15~49)이 낳은 출생아 수를 해당 연령별 여자 연앙인구(7월 1일자 인구)로 나눈후 1000을 곱한 값이다.

$$ASFR = \frac{\text{모(어머니)의 연령별 출생아수}}{\text{해당 연령별 여자인구}} \times 1000$$

그 후 구해진 ASFR을 이용하여 합계 출산율을 구한다.

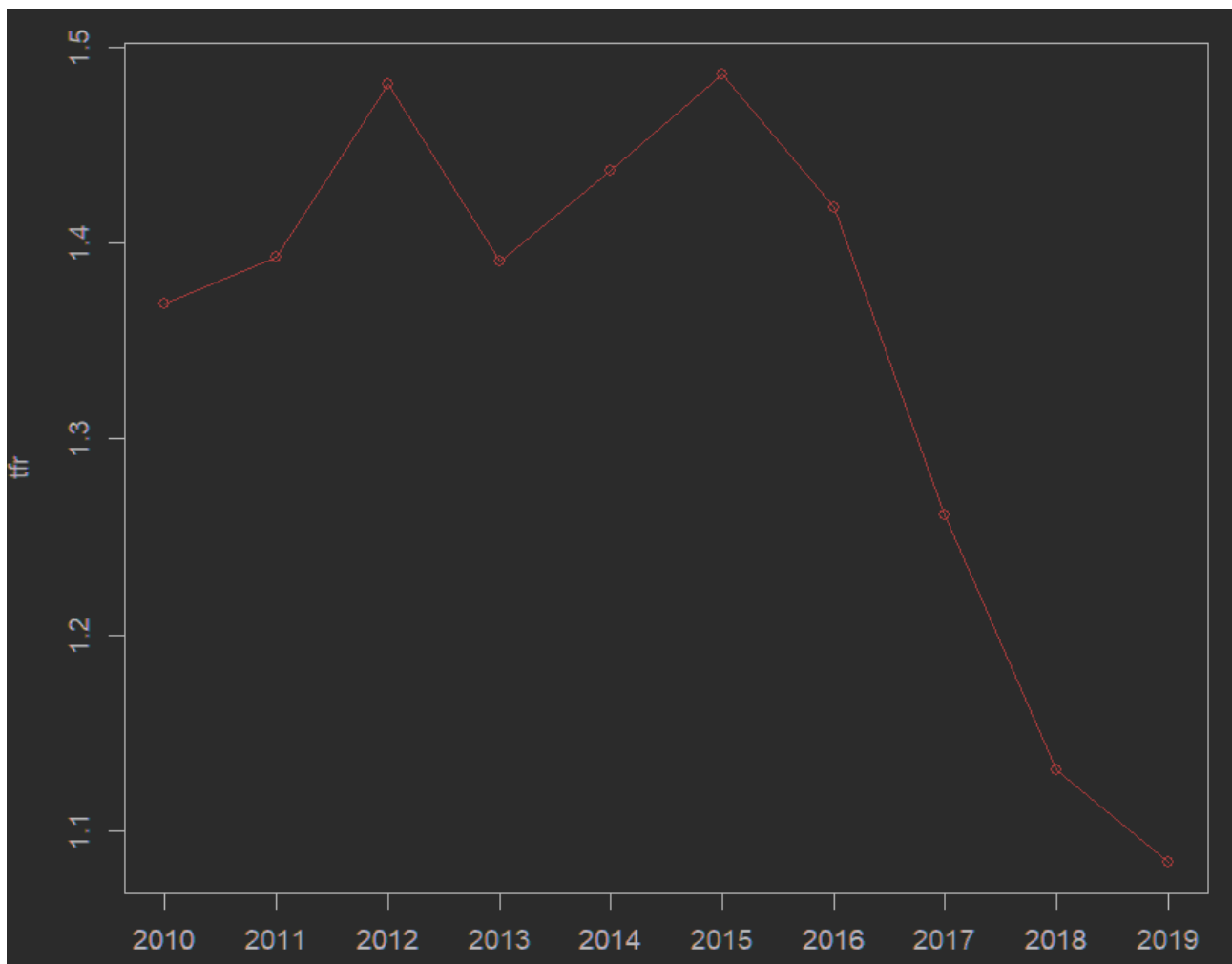
$$TFR = \sum_{a=15}^{49} (ASFR)_a / 1000$$

실제로 ASFR은 계급 단위로 되어 있는 경우가 많으므로 아래와 같은 식으로도 생각할 수 있다.

$$TFR = \sum_{a=15 \sim 19}^{45 \sim 49} (ASFR)_a \times \frac{5}{1000}$$

계급은 일반적으로 5인 경우가 많은데 만약 계급 값이 달라진다면 5가 아닌 해당 계급 크기를 곱해 주어야한다.

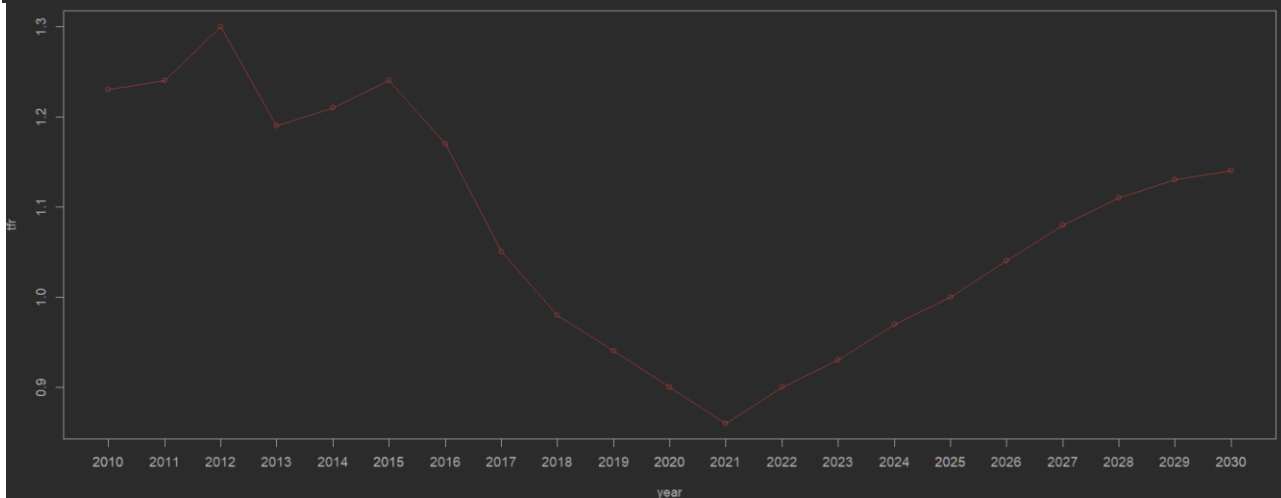
최근 10년간의 합계출산율은 아래와 같다. 차트는 R로 그렸다.



```
df <- read.table('./assignment/question1/data/TFR.csv', sep = ',')
year <- 2010:2019
tfr <- df[1,]
plot(year, tfr, type = 'o', col = 'red')
axis(1, at = 2010:2019)
```

합계출산율을 10년동안 확인해본 결과를 보면 지속적으로 하락세에 있음을 알 수 있다.

```
df <- read.table(
  file =
    './assignment/question1/data/FUTURE_TOTAL_FERTILITY_RATE(2010~2030).csv',
  sep = ',',
  header = T,
  fileEncoding = "utf-8",
  na.strings = '-',
)
year <- 2010:2030
tfr <- df[1,-1]
plot(year, tfr, type = 'o', col = 'red')
axis(1, at = year)
```



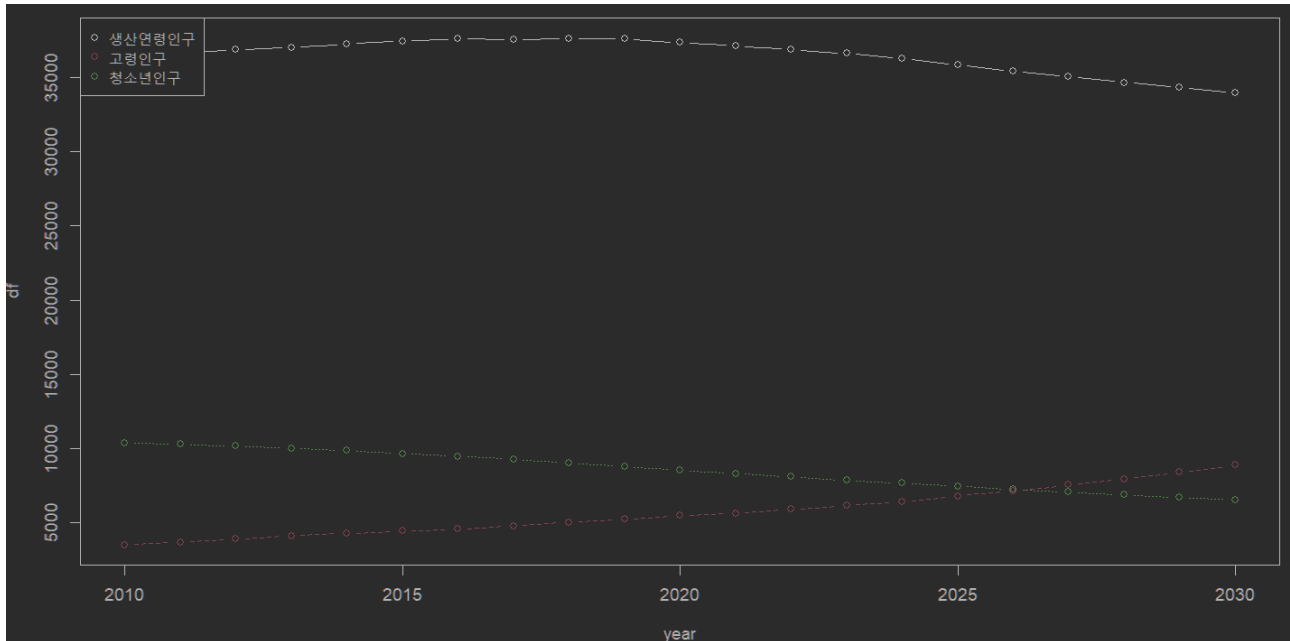
장래 합계출산율은 2017년 까지는 실측치이고 2018년부터 이후는 추정치이다.

장래 합계출산율은 그래프를 보면 2021년에 최저치를 찍고 이후에는 반등할 것이라 기대했었다. 실제로 2021년이 최저치이다. 2019년에 예측한 자료이므로 2021년의 실측치가 0.88이고 추정치가 0.86이므로 유사하다는 걸 알 수 있다.

추후에는 그래프가 반등을 하는데 거기에 대해서 현재 인구수가 많은 1990년대생이 가임기에 진입하면서 지속적으로 합계출산율 감소량이 감소하는 모양새에있기에 다시 반등해서 그 이후 1.1을 유지하지 않겠느냐는 해석이었는데 이 해석이 코로나19로 인하여 틀릴 가능성이 있다고 재기되고 있다.

```
df <- read.table(
  file = './assignment/question1/data/POPULATION_ESTIMATE(2010~2030).csv',
  sep = ',',
  header = T,
  na.strings = '-',
)
df <- t(df[, c(-1, -2)])
colname <- c('생산연령인구', '고령인구', '청소년인구')
colnames(df) <- colname
year <- 2010:2030
```

```
matplot(df, x = year, type = "b", pch = 1, col = 1:3)
legend('topleft', legend = colname, pch = 1, col = 1:3)
```



생산연령별 인구는 계속해서 감소추세에 있고 청소년 인구도 계속해서 감소추세에 있다. 고령인구는 증가 추세인데 이로 인해서 우리나라 인구구조가 고령화 사회가 되어가고 있음을 확인할 수 있다.

2번문제

(1)번문제

실업률과 고용률은 각각 아래의 산식에 의해서 정의된다.

$$\text{실업률}(\%) = \frac{\text{실업자수}}{\text{경제활동인구}} \times 100$$

$$\text{고용률}(\%) = \frac{\text{취업자수}}{\text{15세 이상 인구}} \times 100$$

실업률을 보면 알겠지만 경제활동인구라는 개념을 알아야 구할 수 있다.

$$\text{경제활동인구} = \text{취업자} + \text{실업자}$$

실업률과 고용률은 비슷해 보이지만 다른점이 있는데 고용률은 15세 인구 전체를

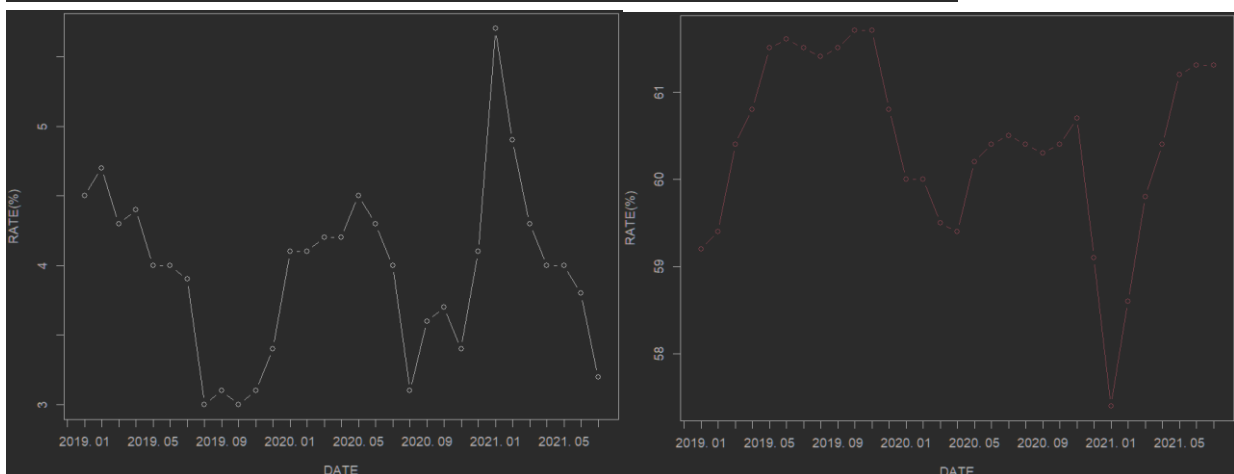
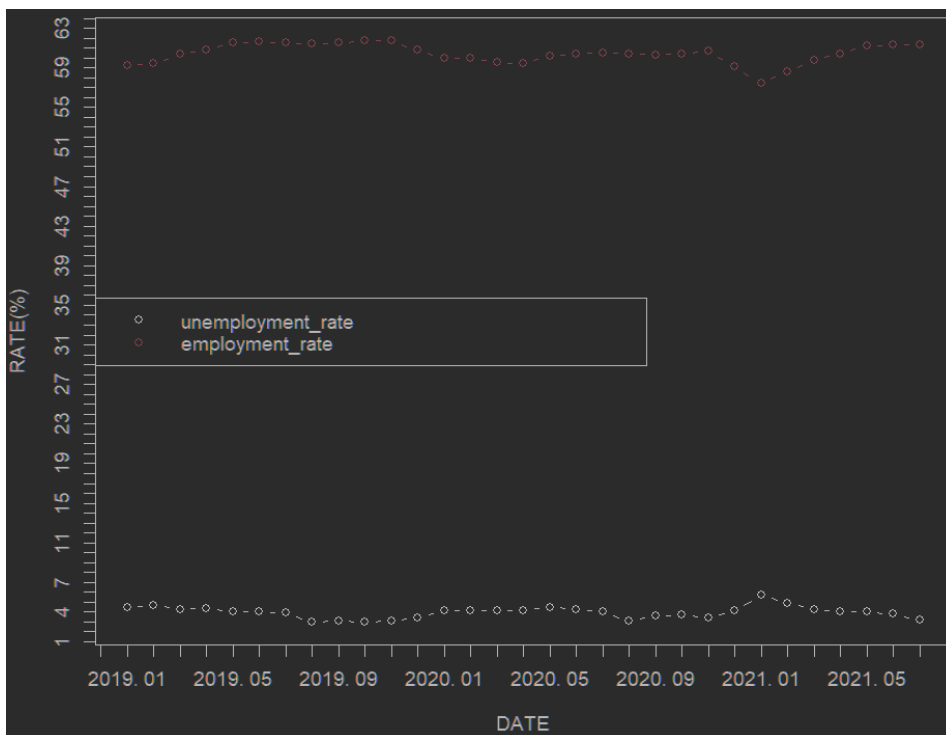
나타낸다는 점이고 실업률은 경제활동인구만 본다는 것이다. 그러면 실업자에 대한 기준을 봐야 알 수 있다. 실업자의 기준은 통계청 사이트¹에서 이를 확인할 수 있다.

1. 조사대상기간에 수입이 있는 일을 하지 않음
2. 지난 4주간 적극적으로 구직활동을 하였음
3. 조사대상기간에 일이 주어지면 즉시 취업이 가능한 사람

이 대상에 들어가야만 실업자가 되기에 주부나 학생들은 통계에서 제외된다. 즉 고용률의 취지는 15세 이상의 사람들 중에서 얼마만큼이 취직이 되었느냐 라는 것을 보는 것이고 실업률은 취직하고 싶은 사람 중에서 얼마만큼 취직이 되었느냐를 보는 것이다. 취지 자체가 다르기에 쓰임새도 다르다. 고용률은 현재 사회에서 얼마만큼의 사람이 일하고있는 지를 기계적으로 나타낸 것이고 실업률은 취업의 의지가 있는 사람이 취직을 못하고 있느냐를 나타낸다. 만약 의지와는 별개로 한국 사회에서 일하는 사람이 얼마나 되느냐를 알고 싶다면 고용률이 올바른 통계일 것이다. 가령 여성의 사회진출이 낮은 사회라면 고용률이 낮게 나올 것이다. 하지만 실업률은 이 상황을 반영하지 못할 것이다. 그러므로 통계에서 사용자의 선택에 따라서 잘 취사선택해야한다.

(2)번문제

```
df <-  
read.table('./assignment/question2/data/ECONOMICALLY_ACTIVE_POPULATION.csv',  
           sep = ',',  
           fileEncoding = 'utf-8')[, -1]  
unemployment_rate <- t(df[3, c(T, F)])  
employment_rate <- t(df[3, c(F, T)])  
value_range <- t(df[1, c(F, T)])  
data <- data.frame(cbind(unemployment_rate, employment_rate))  
colnames(data) <- c('unemployment_rate', 'employment_rate')  
rownames(data) <- value_range  
matplot(data, type = "b", pch = 1, col = 1:3, ylab = "RATE(%)", xlab = "DATE",  
        labels = F)  
axis(side = 2, at = 1:100)  
axis(side = 1, at = seq_along(value_range), value_range)  
legend('left', legend = c('unemployment_rate', 'employment_rate'), pch = 1, col  
      = 1:3)
```



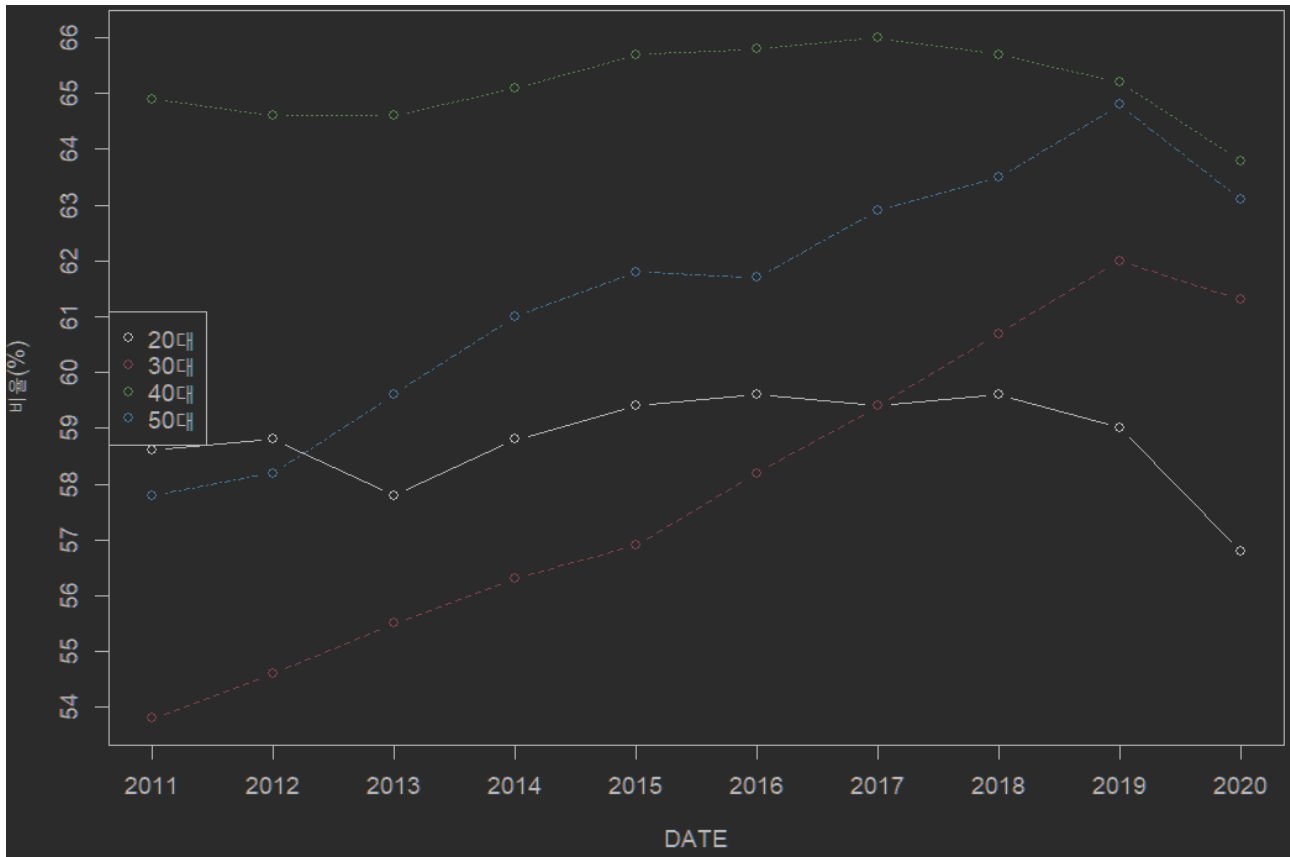
R로 그린 그래프에서 흰색은 실업률이고 빨강은 고용률을 나타낸다. 전체적인 경향을 보면 실업률이 증가하면 고용률이 감소하고 실업률이 감소하면 고용률이 증가하는 반비례 관계인 것을 알 수 있다.

(3)번문제

```
df <-
read.table('./assignment/question2/data/EMPLOYED_BY_GENDER_AND_AGE_RATE.csv',
sep = ',')[-1, c(-1, -2)]
#rate <- colSums(df)
df <- t(df)

value_range <- 2011:2020
```

```
matplot(df, type = "b", pch = 1, col = 1:4, ylab = "비율(%)", xlab = "DATE",
labels = F)
axis(side = 2, at = 1:10)
axis(side = 1, at = seq_along(value_range), value_range)
legend('left', legend = c('20대', '30대', '40대', '50대'), pch = 1, col = 1:4)
```

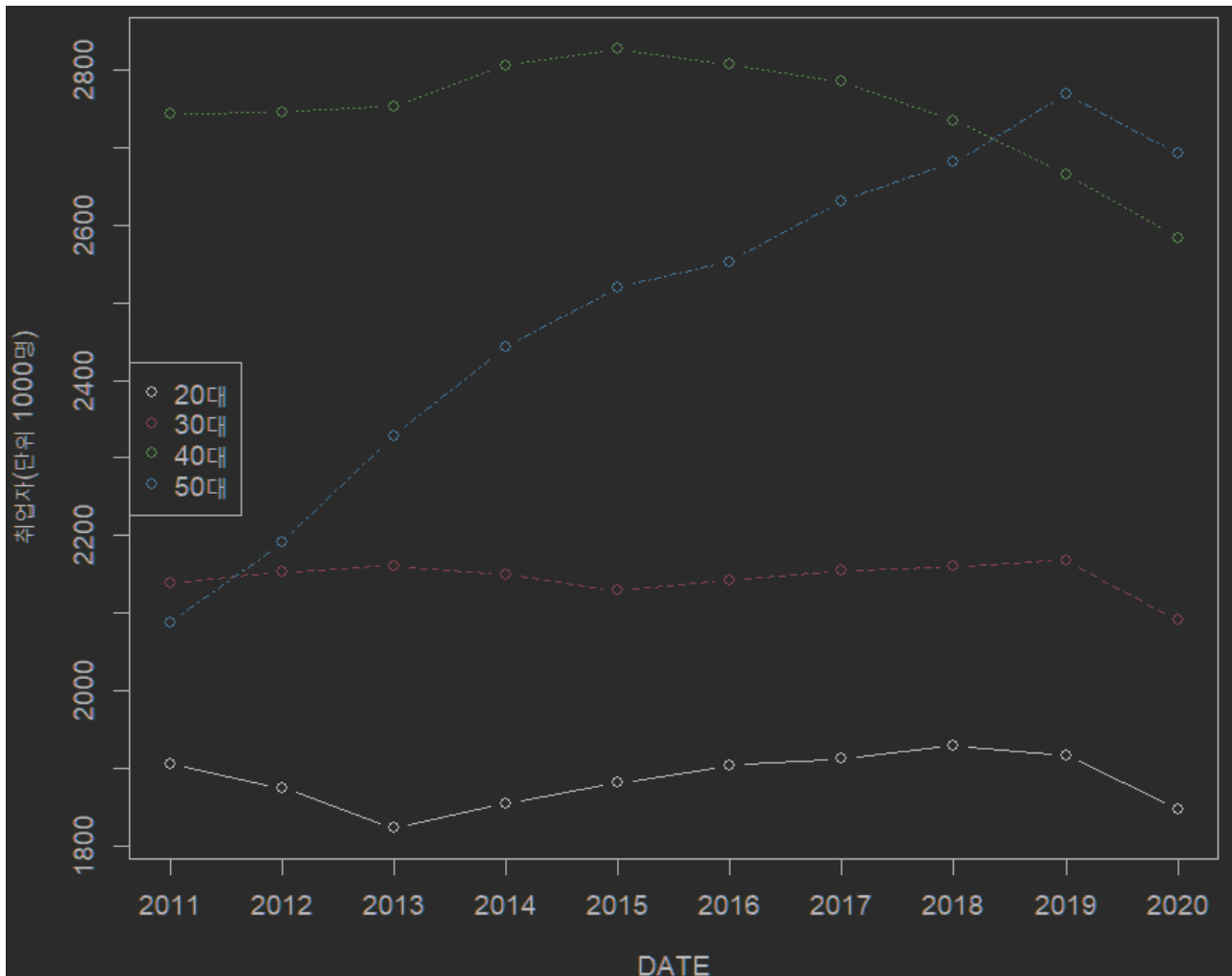


해당데이터는 근 10년간 여성의 고용률을 알아보자. 실업률을 알아보지 않는 이유는 해당 문제가 여성 경제활동 규모를 알고 싶는데 실업률로 조사할 경우 취직의 의사가 없는 사람은 통계에서 제외되기 때문에 규모를 파악할 수 없기 때문이다. 해당 데이터를 통해서 20대와 40대의 고용률의 경우 꾸준히 일정 수준을 유지하고 있지만 50대와 30대의 고용률이 폭발적이게 늘어난 것을 확인할 수 있다. 반면에 코로나의 여파로 2019년부터 고용률에 감소세가 있음을 확인할 수 있다.

```
df <- read.table('./assignment/question2/data/EMPLOYED_BY_GENDER_AND_AGE.csv',
sep = ',')[-1, c(-1, -2)]
#rate <- colSums(df)
df <- t(df)

value_range <- 2011:2020
matplot(df, type = "b", pch = 1, col = 1:4, ylab = "취업자(단위 1000 명)", xlab =
```

```
"DATE", labels = F)
axis(side = 2, at = seq(from = 0, to = 7000, by = 100))
axis(side = 1, at = seq_along(value_range), value_range)
legend('left', legend = c('20 대', '30 대', '40 대', '50 대'), pch = 1, col = 1:4)
```



취업자의 수로 봤을 때는 20대와 30대는 일정한 수준을 유지중이고 40대는 꾸준히 감소세에 있으나 50대의 경우 근 시일내에 폭발적으로 증가한 것을 확인할 수 있다.

두 가지 그래프를 종합해서 봤을 때 50대에서의 여성 경제활동 규모가 가장 크게 증가하고 있음을 알 수 있다.

3번문제

(1)번문제

가중치는 품목의 총합을 1000으로 뒀을 때 그에 따른 비율을 의미한다. 예를들어

전체가 300개라면 300을 1000으로 잡게 되며 그중의 특정 항목이 100개일경우 가중치는 333이된다.

```
item <- c('쌀', '닭고기', '배추', '사과')
year2015 <- data.frame(list(count = c(10, 7.5, 7, 8), price = c(5000, 2000, 1000, 1000)))
rownames(year2015) <- item
year2016 <- data.frame(list(count = c(10, 6.5, 8, 7), price = c(5500, 1800, 800, 1300)))
rownames(year2016) <- item
year2017 <- data.frame(list(count = c(9.5, 8, 6, 8), price = c(5700, 2000, 1200, 1200)))
rownames(year2017) <- item
for (i in 2015:2017) {
  print(i)
  cur <- get(paste0('year', i))
  weight <- round(cur$count * 1000 / sum(cur$count), 1)
  print(cbind(cur, weight))
}
```

```
[1] 2015
      count price weight
쌀      10.0  5000   307.7
닭고기   7.5  2000   230.8
배추     7.0  1000   215.4
사과     8.0  1000   246.2
[1] 2016
      count price weight
쌀      10.0  5500   317.5
닭고기   6.5  1800   206.3
배추     8.0   800   254.0
사과     7.0  1300   222.2
[1] 2017
      count price weight
쌀       9.5  5700   301.6
닭고기   8.0  2000   254.0
배추     6.0  1200   190.5
사과     8.0  1200   254.0
```

R로 코드를 작성하였고 각 연도별 가중치(weight)는 위와 같다.

(2)번문제

라이파레스산식은 가중산술평균법이라고도 한다. 해당 산식은 아래와 같다.

$$L_{year1, year2}^{year1=100} = \frac{\sum P_i^{year1} Q_i^{year2}}{P_i^{year1} Q_i^{year1}} \times 100 = \sum W_i^{year1} (P_i^{year2} / P_i^{year1}) \times 100$$

가중 산술평균법은 단순 산술평균과는 달리 가중치를 뒤서 산술평균을 하는 것을 의미한다. 가중치는 1번문제에서 확인할 수 있듯이 전체에서 해당 품목이 차지하는 비율로 확인할 수 있다. 일반 산술평균을 사용하지 않는 것은 척도가 다른 품목들을 단순히 더하게 된다면 그 가치들을 왜곡하기 때문이다.

```
year2015 <- data.frame(list(count = c(10, 7.5, 7, 8), price = c(5000, 2000, 1000, 1000)))
year2016 <- data.frame(list(count = c(10, 6.5, 8, 7), price = c(5500, 1800, 800, 1300)))
year2017 <- data.frame(list(count = c(9.5, 8, 6, 8), price = c(5700, 2000, 1200, 1200)))
for (i in 2015:2017) {
  cur <- get(paste0('year', i))
  weight <- round(cur$count * 1000 / sum(cur$count), 1)
  assign(paste0('year', i), cbind(cur, weight))
}

rate2016 <- sum(year2015$weight * (year2016$price / year2015$price)) / 10
rate2017 <- sum(year2015$weight * (year2017$price / year2015$price)) / 10
for (i in 2016:2017) {
  cur <- get(paste0('rate', i))
  print(i)
  print(cur)
}
```

```
[1] 2016
[1] 103.857
[1] 2017
[1] 113.5498
```

결과를 보면 2016년 물가 상승률이 2017년 보다 더 낮다는 걸 알 수 있다.

ⁱ http://kostat.go.kr/understand/info/info_lge/1/detail_lang.action?bmode=detail_lang&cd=SL4139