

# 회귀분석론\_hw3

212STG18 예지혜

2021년 3월 9일

## 1.1

```
library(alr4)
```

```
## Warning: package 'alr4' was built under R version 3.6.3
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
## Loading required package: effects
```

```
## Warning: package 'effects' was built under R version 3.6.3
```

```
## Registered S3 methods overwritten by 'lme4':
##   method                      from
##   cooks.distance.influence.merMod car
##   influence.merMod             car
##   dfbeta.influence.merMod      car
##   dfbetas.influence.merMod     car
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

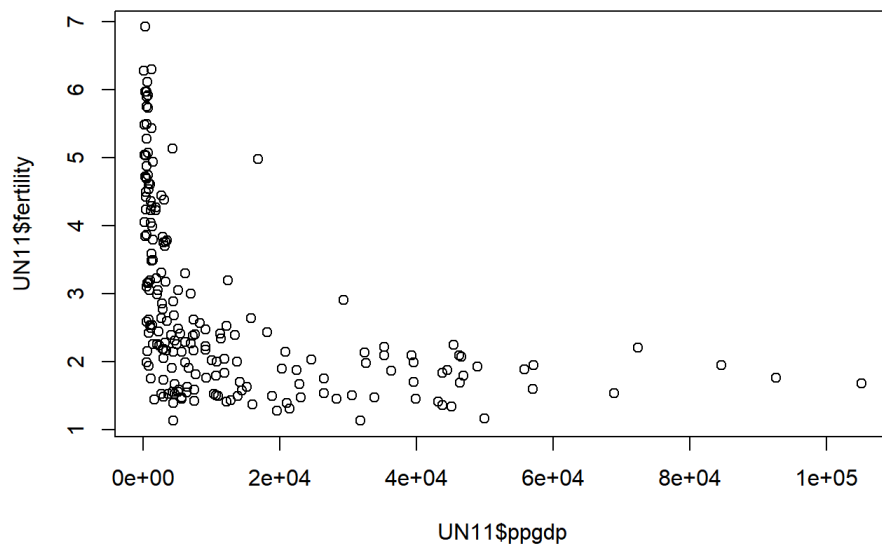
```
head(UN11)
```

```
##           region group fertility  ppgdp lifeExpF pctUrban
## Afghanistan   Asia  other    5.968   499.0   49.49     23
## Albania        Europe other    1.525  3677.2   80.40     53
## Algeria         Africa africa    2.142  4473.0   75.00     67
## Angola          Africa africa    5.135  4321.9   53.17     59
## Anguilla        Caribbean other    2.000 13750.1   81.10    100
## Argentina      Latin Amer other    2.172  9162.1   79.89     93
```

1.1.1 predictor : ppgdp, response : fertility

### 1.1.2

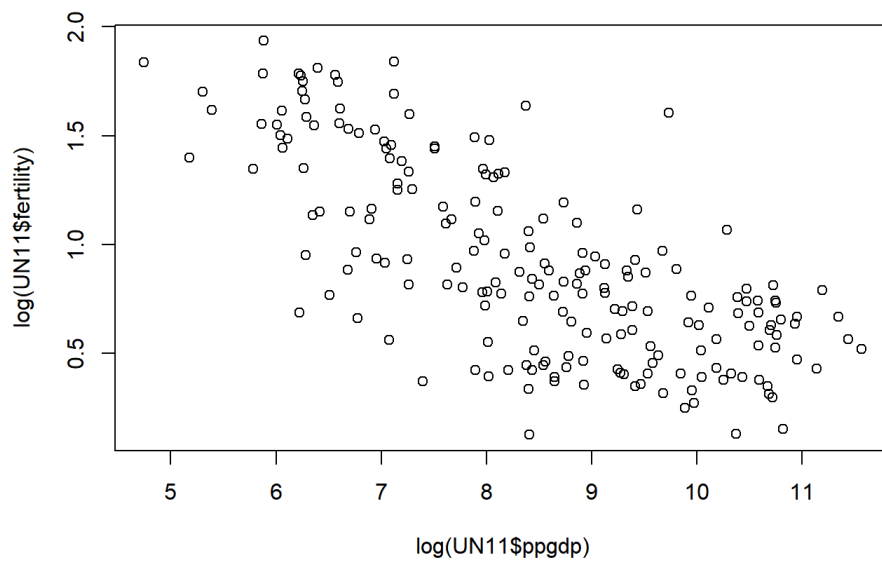
```
plot(UN11$ppgdp, UN11$fertility)
```



1차 직선 함수로의 피팅은 좋지 않아 보인다. 선형 관계가 보이지 않을 뿐 아니라, ppgdp가 낮은 값에 매우 몰려있어 적합이 어렵다.

### 1.1.3

```
plot(log(UN11$ppgdp), log(UN11$fertility))
```



기울기가 음수인 1차 직선을 피팅시키기에 적합해 보인다. 또한,  $\log(\text{ppgdp})$ 는 전체 범위에 걸쳐 고르게 분포해있다.

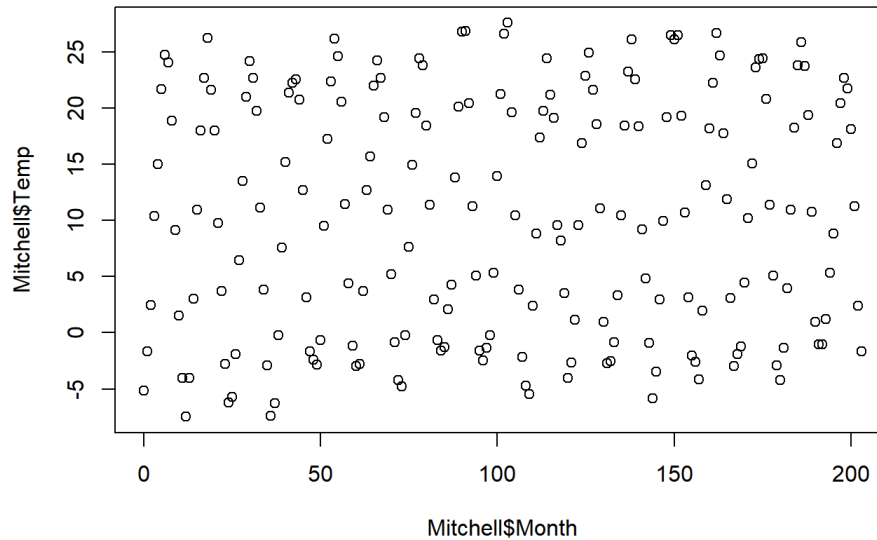
### 1.3

```
head(Mitchell)
```

```
##      Month      Temp
## 1      0 -5.18333
## 2      1 -1.65000
## 3      2  2.49444
## 4      3 10.40000
## 5      4 14.99440
## 6      5 21.71670
```

## 1.3.1

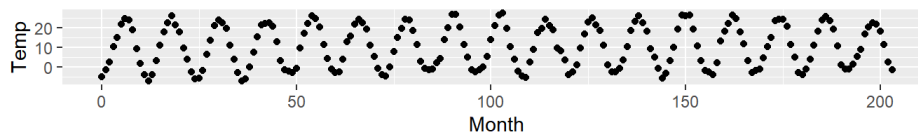
```
plot(Mitchell$Month, Mitchell$Temp)
```



Month에 상관없이 Temp가 고르게 분포해있다. Month와 Temp의 어떠한 상관성을 발견하기 어렵다.

## 1.3.2

```
library(ggplot2)
ggplot(Mitchell, aes(x=Month, y=Temp)) + geom_point() + coord_fixed(ratio = 0.5)
```



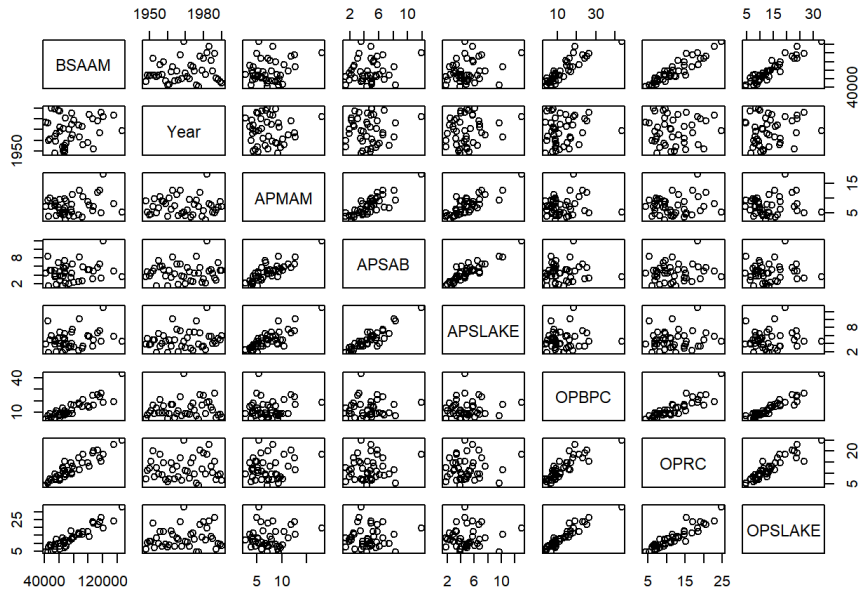
x축의 길이를 길게 늘렸더니, 월(시간)이 지남에 따라 주기적으로 온도가 상승, 하강을 반복하는 것을 발견할 수 있다.

## 1.5

```
head(water)
```

```
##      Year  APMAM  APSAB  APSLAKE  OPBPC  OPRC  OPSLAKE  BSAAM
## 1 1948   9.13   3.58   3.91   4.10   7.43   6.47  54235
## 2 1949   5.28   4.82   5.20   7.55  11.11  10.26  67567
## 3 1950   4.20   3.77   3.67   9.52  12.20  11.35  66161
## 4 1951   4.60   4.46   3.93  11.14  15.15  11.13  68094
## 5 1952   7.15   4.99   4.88  16.34  20.05  22.81 107080
## 6 1953   9.70   5.65   4.91   8.88   8.15   7.41  67594
```

```
pairs(BSAAM ~ ., data=water)
```



각 지역별 BSAAM(stream runoff)의 시계열 자료를 살펴보면, OPBPC와 OPRC, OPSLAKE 지역에서 꾸준히 증가하고 있는 것을 확인할 수 있다. 나머지 지역에서는 증가한다고 단정짓기는 애매하나, 분포 자체는 서로 유사하다.

또한 지역 간의 plot을 통해 서로 연관성 있는 지역들을 찾을 수 있다. APNAM과 APSAB, APSLAKE 지역끼리, OPBPC와 OPRC, OPSLAKE 지역끼리 양의 상관성이 높아보이는데, 이는 이 지역들의 BSAAM 경향이 비슷하다고 해석할 수 있다.

년도 자체는 그 어떤 변수와도 상관성이 높지 않아보이며, 이는 BSAAM이 지역별 특색과는 어떠한 관련성이 있지만, 시간에는 의존하지 않는다고 해석할 수 있다.