

2021_R_Basic_FinalExam

Jinwon Lee

2022-09-08

// 1. “제주특별자치도개발공사_기상 정보_20191231.csv” 파일을 다운받아 열어보세요. (Excel로) (2점)

// 2. CSV파일을 R로 불러와서 상위 6행 관측치를 미리보기 하세요. (불러올 때 변수명은 data.raw) (3점)

```
getwd()
## [1] "G:/내 드라이브/202202/R_Basic"

data.raw <- read.csv("./2021/제주특별자치도개발공사_기상 정보_20191231.csv",
                     fileEncoding = "euc-kr")

data.raw %>% head()
##           일자 평균기온... 최대기온... 최소기온... 평균.지표온도...
## 1 2019-01-13      6.62      15.65      5.12      4.87
## 2 2019-01-14      5.44      8.29      3.27      5.08
## 3 2019-01-15      4.93      9.02      2.09      4.85
## 4 2019-01-16      4.97      8.76      1.00      4.75
## 5 2019-01-17      0.59      2.42     -0.55      4.53
## 6 2019-01-18      2.88      5.73      1.07      4.07
## 최대.지표온도... 최소.지표온도... 습도... 일사량.W... 풍속...
## 1           5.19           4.61  90.10      152.5      2.60
## 2           5.19           4.98  84.60      146.3      3.41
## 3           5.11           4.67  84.20      167.7      2.96
## 4           4.87           4.63  77.52      156.2      4.44
## 5           4.86           4.32  81.30      137.2      4.83
## 6           4.33           3.96  74.01      136.1      4.42
```

// 3. 데이터 셋은 몇행 몇열인가요? (반드시, str()함수를 사용하여 확인하세요.) (2점)

```
data.raw %>% str() #> 333행 10열
## 'data.frame':   333 obs. of  10 variables:
## $ 일자          : chr  "2019-01-13" "2019-01-14" "2019-01-15" "2019-01-16"
## ...
## $ 평균기온...    : num  6.62 5.44 4.93 4.97 0.59 2.88 4 7.2 4.7 3.49 ...
## $ 최대기온...    : num  15.65 8.29 9.02 8.76 2.42 ...
## $ 최소기온...    : num  5.12 3.27 2.09 1 -0.55 1.07 -2.03 -1.97 1.21 1.07 ...
## $ 평균.지표온도...: num  4.87 5.08 4.85 4.75 4.53 4.07 3.84 3.7 4.15 4.08 ...
## $ 최대.지표온도...: num  5.19 5.19 5.11 4.87 4.86 4.33 4.03 4 4.32 4.27 ...
## $ 최소.지표온도...: num  4.61 4.98 4.67 4.63 4.32 3.96 3.68 3.51 3.83 3.93 ...
## $ 습도...        : num  90.1 84.6 84.2 77.5 81.3 ...
## $ 일사량.W...     : num  152 146 168 156 137 ...
```

```
## $ 풍속... : num 2.6 3.41 2.96 4.44 4.83 4.42 2.38 2.08 6.15 4.91 ...
```

// 4. 열 이름을 아래와 같이 바꾸세요. (hint: names()) 함수 사용, 다른 함수 사용 가능) (3점)

- “일자”, “평균기온”, “최대기온”, “최소기온”, “평균지표온도”, “최대지표온도”, “최소지표온도”, “습도”, “일사량”, “풍속”

```
names(data.raw) <- c("일자", "평균기온", "최대기온", "최소기온",  
                     "평균지표온도", "최대지표온도", "최소지표온도",  
                     "습도", "일사량", "풍속")  
data.raw %>% names()  
## [1] "일자"          "평균기온"      "최대기온"      "최소기온"      "평균지표온도"  
## [6] "최대지표온도" "최소지표온도" "습도"          "일사량"        "풍속"
```

// 5. 일교차를 계산하여 새로운 열(열이름: 일교차)을 추가하세요. (10점) (hint. 일교차 = 최대기온 - 최소기온)

```
data.raw['일교차'] = data.raw['최대기온'] - data.raw['최소기온']  
data.raw['일교차'] %>% head()  
## 일교차  
## 1 10.53  
## 2 5.02  
## 3 6.93  
## 4 7.76  
## 5 2.97  
## 6 4.66
```

// 6. 일교차의 평균, 표준편차, 최대값, 최소값을 구하세요. (10점)

```
data.raw %>%  
  summarise(평균 = mean(일교차),  
            표준편차 = sd(일교차),  
            최대값 = max(일교차),  
            최소값 = min(일교차))  
## 평균 표준편차 최대값 최소값  
## 1 9.683754 4.471683 21.23 1.14
```

// 7. 일교차가 15°C 이상인 날의 행만을 추출하세요. (10점)

```
data.raw %>% filter(일교차 > 15) %>% head()  
## 일자 평균기온 최대기온 최소기온 평균지표온도 최대지표온도 최소지표온도  
## 1 2019-02-06 5.79 13.37 -3.45 5.23 5.83 4.87  
## 2 2019-03-09 7.23 16.37 -1.97 7.07 7.50 6.62  
## 3 2019-03-16 7.45 16.04 -1.83 7.20 7.74 6.69  
## 4 2019-03-18 10.13 22.34 2.15 8.14 8.84 7.76  
## 5 2019-03-27 12.35 18.53 3.16 8.67 9.58 8.21  
## 6 2019-03-28 13.20 22.50 1.27 9.29 9.89 8.82  
## 습도 일사량 풍속 일교차  
## 1 76.17 191.8 1.80 16.82
```

```
## 2 59.19 227.1 1.65 18.34
## 3 64.47 226.2 4.05 17.87
## 4 56.02 233.5 3.23 20.19
## 5 68.79 231.5 3.46 15.37
## 6 51.99 236.1 2.36 21.23
```

// 8. 월별 일교차를 계산하세요. (15점)

- (hint) substr() : 문자열 일부 추출 ex) substr("Statistics", 1, 4) → "Stat"
- step1) 일자 열에서 "월"만 추출하여 새로운 열(새로운 열 이름 : 월)을 추가하세요.
 - ex) substr("2019-01-13",6,7) → "01" : 6번째에서 7번째까지 문자 출력
- step2) group_by(), summarise() 또는 aggregate() 함수를 적용하여 월별 평균값을 구하세요.

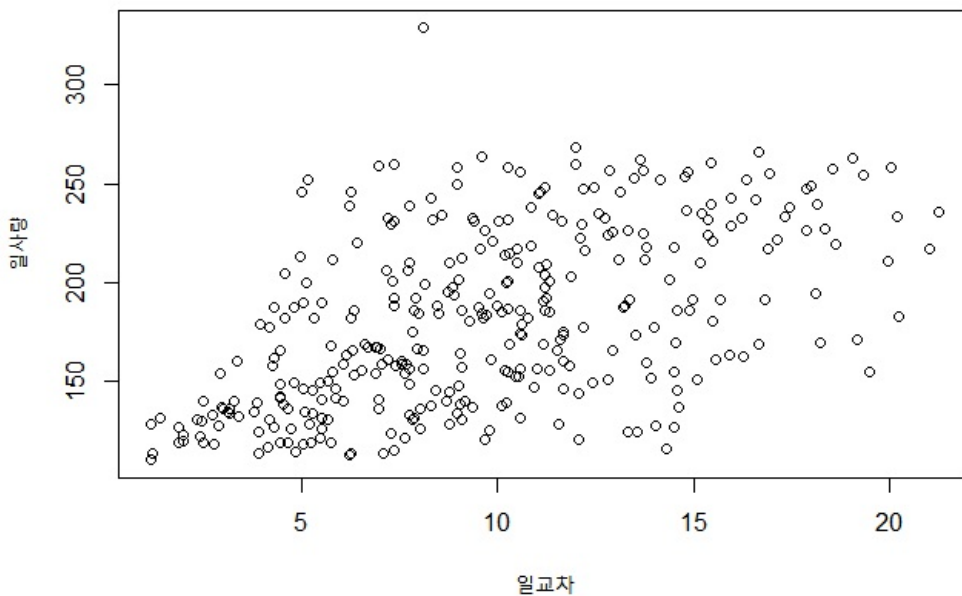
```
data.raw['월'] = substr(data.raw$일자, 6, 7)

data.raw %>%
  group_by(월) %>%
  summarise(월평균_일교차 = mean(일교차))
## # A tibble: 12 × 2
##   월   월평균_일교차
##   <chr>         <dbl>
## 1 01             8.33
## 2 02             8.55
## 3 03            11.7
## 4 04            11.8
## 5 05            11.8
## 6 06            10.0
## 7 07             6.95
## 8 08             7.52
## 9 09             8.41
## 10 10            8.88
## 11 11            11.4
## 12 12            9.45
```

// 9.

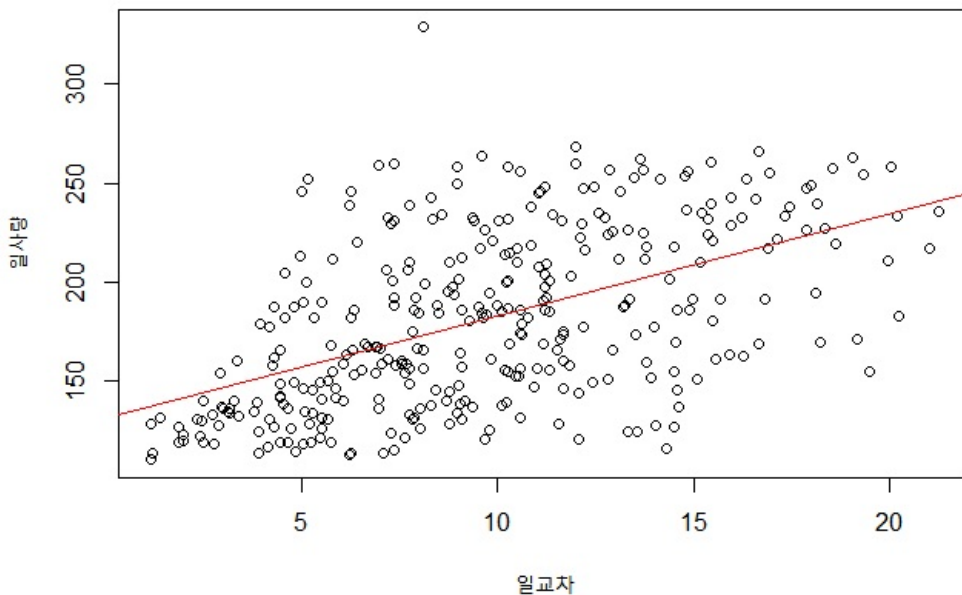
- 1. 일교차와 일사량에 대한 산점도를 그리세요. (plot()함수 이용) (5점)

```
plot(formula = 일사량~일교차, data = data.raw)
```



- 2. 회귀선을 추가하세요. (lm(), abline() 함수 이용) (5점)

```
plot(formula = 일사량~일교차, data = data.raw)
abline(lm(data.raw$일사량~data.raw$일교차), col = 'red')
```

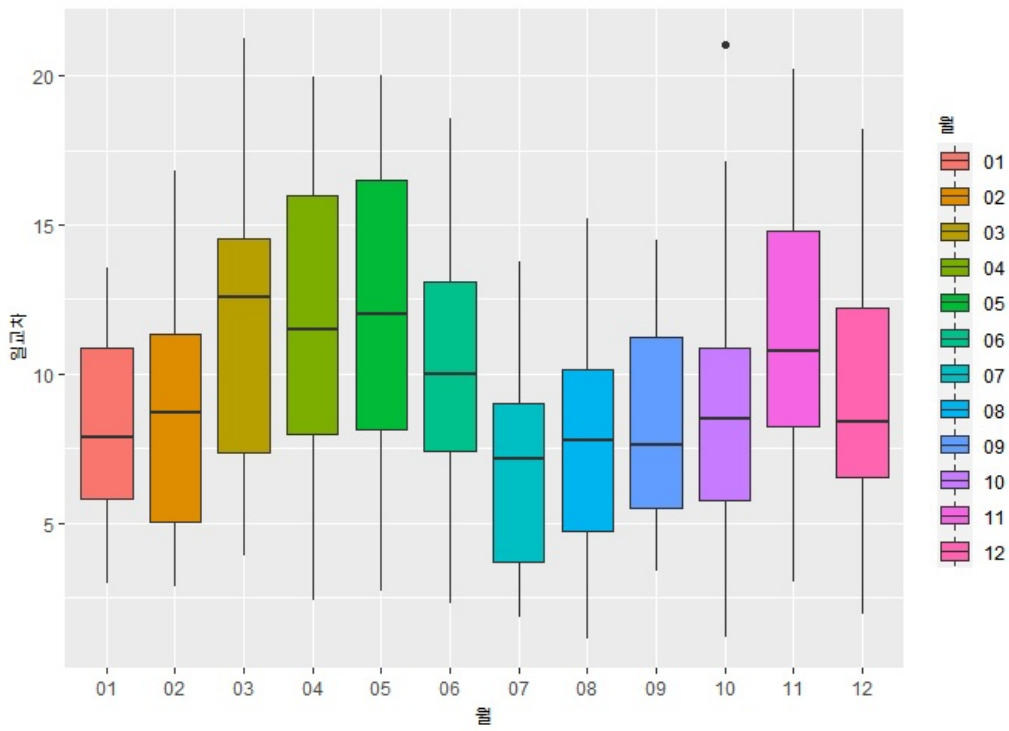


- 3. 상관계수를 산출하세요. (cor() 함수 이용) (10점)

```
cor(x = data.raw['일교차'], y = data.raw['일사량'])
## 일사량
## 일교차 0.5229626
```

// 10. 아래와 같이 월별 일교차의 분포를 ggplot2 패키지의 그룹별 상자그림을 활용하여 그려보세요. (15점)

```
data.raw %>%
  ggplot(aes(월, 일교차, fill = 월)) +
  geom_boxplot()
```



// 11.

- 1. 아래와 같이 습도 열 1번째, 3번째, 5번째, 10번째에 결측값 NA를 입력하세요. (5점)

```
data.raw[c(1, 3, 5, 7), '습도'] <- NA
```

```
data.raw %>% head(10)
```

```
##          일자 평균기온 최대기온 최소기온 평균지표온도 최대지표온도 최소지표온도
## 1 2019-01-13      6.62    15.65      5.12          4.87          5.19          4.61
## 2 2019-01-14      5.44      8.29      3.27          5.08          5.19          4.98
## 3 2019-01-15      4.93      9.02      2.09          4.85          5.11          4.67
## 4 2019-01-16      4.97      8.76      1.00          4.75          4.87          4.63
## 5 2019-01-17      0.59      2.42     -0.55          4.53          4.86          4.32
## 6 2019-01-18      2.88      5.73      1.07          4.07          4.33          3.96
## 7 2019-01-19      4.00     10.13     -2.03          3.84          4.03          3.68
## 8 2019-01-20      7.20     11.59     -1.97          3.70          4.00          3.51
## 9 2019-01-21      4.70     10.21      1.21          4.15          4.32          3.83
## 10 2019-01-22      3.49      6.39      1.07          4.08          4.27          3.93
##          습도 일사량 풍속 일교차 월
## 1      NA    152.5 2.60   10.53 01
## 2  84.60    146.3 3.41    5.02 01
## 3      NA    167.7 2.96    6.93 01
## 4  77.52    156.2 4.44    7.76 01
## 5      NA    137.2 4.83    2.97 01
## 6  74.01    136.1 4.42    4.66 01
## 7      NA    177.5 2.38   12.16 01
## 8  89.40    124.5 2.08   13.56 01
## 9  85.10    147.6 6.15    9.00 01
## 10 76.41    182.4 4.91    5.32 01
```

- 2. 결측값(NA)가 포함된 행들을 출력하세요. (10점)

```
data.raw[is.na(data.raw$습도), ]
```

```
##          일자 평균기온 최대기온 최소기온 평균지표온도 최대지표온도 최소지표온도
```

```
## 1 2019-01-13      6.62    15.65      5.12          4.87          5.19          4.61
```

```
## 3 2019-01-15      4.93      9.02      2.09          4.85          5.11          4.67
```

```
## 5 2019-01-17      0.59      2.42     -0.55          4.53          4.86          4.32
```

```
## 7 2019-01-19      4.00     10.13     -2.03          3.84          4.03          3.68
```

```
## 습도 일사량 풍속 일교차 월
```

```
## 1  NA  152.5 2.60  10.53 01
```

```
## 3  NA  167.7 2.96   6.93 01
```

```
## 5  NA  137.2 4.83   2.97 01
```

```
## 7  NA  177.5 2.38  12.16 01
```