**확률과 통계 기말고사 분석 보고서**

- 모기 개체수와 기온, 모기 개체수와 습도의 상관관계 비교 -

**201835612컴퓨터공학과**

**경은하**

**1. 2018년도 14주~44주 동안의 인천광역시 모기밀도 결과 전처리**

## 'data.frame': 371 obs. of 19 variables:  
## $ 월 : int 4 4 4 4 4 4 4 4 4 4 ...  
## $ 연중주수 : Factor w/ 31 levels "14주","15주",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ 채집장소 : Factor w/ 12 levels "인천광역시 강화 강화읍 (대산리)",..: 8 9 11 10 12 7 6 5 4 3 ...  
## $ 얼룩날개모기류: int 0 0 0 1 0 0 1 0 7 0 ...

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  
## $ 금빛어깨숲모기: int 0 0 0 0 0 0 0 0 0 0 ...

1. 연중주수 변수 전처리

mosquito <- rename(mosquito, week = "연중주수")

mosquito$week <- gsub("주$", "", mosquito$week)  
mosquito$week <- as.numeric(mosquito$week)

2. 채집장소 변수 전처리 : 채집장소의 내용을 아래 표와 같이 코드명으로 변경한다.

|  |  |  |  |
| --- | --- | --- | --- |
| 채집장소 | 코드명 | 채집장소 | 코드명 |
| 인천광역시 계양구 선주지동 | 1 | 인천광역시 강화 송해면 (솔정리) | 6 |
| 인천광역시 부평구 부평동 | 2 | 인천광역시 강화 선원면 (금월리) | 7 |
| 인천광역시 서구 연희동 | 3 | 인천광역시 강화 삼산면 (석모리) | 8 |
| 인천광역시 서구 백석동 | 4 | 인천광역시 강화 교동면 (대룡리) | 9 |
| 인천광역시 중구 운남동 (영종도) | 5 | 인천광역시 강화 강화읍 (대산리) | 10 |
| 인천광역시 강화 송해면 (숭뢰리) | 6 | 인천광역시 강화 강화읍 (월곳리) | 10 |

3. 모기 개체수의 총합을 나타내는 total 변수 생성

**2. 2018년도 14~44주 동안의 인천광역시 계양구 계양1동~강화읍 매시각 기온 데이터 전처리**

## 'data.frame': 5864 obs. of 3 variables:  
## $ format..day : Factor w/ 39 levels " "," 1"," 10",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ hour : int 0 100 200 300 400 500 600 700 800 900 ...  
## $ value.location.56\_127.Start...20180401: num 15.7 17.2 17.3 16.5 16.5 ...

|  |  |
| --- | --- |
| 모기밀도 데이터 채집장소 중 법정동과 행정동이 다른 장소 | 행정동명 |
| 인천광역시 계양구 선주지동 | 인천광역시 계양1동 |
| 인천광역시 부평구 부평동 | 인천광역시 부평1동~부평6동 |
| 인천광역시 서구 백석동 | 인천광역시 서구 검암 경서동 |
| 인천광역시 중구 운남동 (영종도) | 인천광역시 영종동 |
| 인천광역시 강화읍 대산리, 월곳리 | 인천광역시 강화읍 |

1. format: day, hour, value.location.56\_127.Start…20180401 변수 전처리

gyeyang1\_TEMP <- rename(gyeyang1\_TEMP, day = format..day) #변수명 day로 변경

gyeyang1\_TEMP <- gyeyang1\_TEMP %>% filter(!is.na(hour)) #결측지 제거

gyeyang1\_TEMP <- rename(gyeyang1\_TEMP, TEMP = value.location.56\_127.Start...20180401) #변수명 TEMP로 바뀜  
gyeyang1\_TEMP <- gyeyang1\_TEMP %>% #결측지 제거  
 filter(!is.na(TEMP))

4. week 변수 생성

k <- 14  
 for (i in 1:5856) {  
 gyeyang1\_TEMP$week[i] <- k  
 ifelse(i %% 168 == 0, k <- k + 1, k)  
 }

5. gyeyang1\_TEMP\_final : 계양1동의 주차별 평균 온도표 만들기

gyeyang1\_TEMP\_final <- gyeyang1\_TEMP %>%   
 filter(!(week > 44)) %>%   
 group\_by(week) %>%   
 summarise(mean\_TEMP = mean(TEMP)) %>%   
 mutate(채집장소 = 1)

**3. 기온과 모기개체수 간의 관계표 만들기**

1. TEMP 데이터 생성 : 각 행정구역의 주차별 기온 데이터를 모두 모아놓은 데이터이다.

TEMP <- bind\_rows(gyeyang1\_TEMP\_final, bupyeong\_TEMP\_final, yeonhee\_TEMP\_final, geomam\_TEMP\_final, youngjong\_TEMP\_final, songhae\_TEMP\_final, seonwon\_TEMP\_final, samsan\_TEMP\_final, gyodong\_TEMP\_final, ganghwa\_TEMP\_final)

1. Final\_TEMP 데이터 생성 : week와 채집장소를 기준으로 TEMP와 mosquito 데이터 병합.

Final\_TEMP <- merge(mosquito, TEMP, by = c("week", "채집장소"))

1. 기온과 모기개체수 간의 관계표 생성

TEMP\_Mosquito <- Final\_TEMP %>%  
 group\_by(week) %>%   
 mutate(mosquito = sum(total)) %>%   
 mutate(TEMP = mean(mean\_TEMP)) %>%   
 select(월, week, TEMP, mosquito)  
TEMP\_Mosquito <- unique(TEMP\_Mosquito)

*계양1동과 같이 다른 행정구역도 전처리 과정이 동일하다.*

**4. 2018년도 14~44주까지의 인천광역시 계양구 계양1동~인천광역시 강화읍 매시각 상대습도 데이터 전처리**

## 'data.frame': 5864 obs. of 3 variables:  
## $ format..day : Factor w/ 39 levels " "," 1"," 10",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ hour : int 0 100 200 300 400 500 600 700 800 900 ...  
## $ value.location.56\_127.Start...20180401: num 65 65 67 72 74 75 77 80 83 84 ...

1. format: day, value location:56\_127 Start : 20180401, hour 변수 전처리

# 변수명 day로 변경  
gyeyang1\_Humidity <- rename(gyeyang1\_Humidity, day = format..day)

daygyeyang1\_Humidity <- rename(gyeyang1\_Humidity, Humidity = value.location.56\_127.Start...20180401) # 변수명 Humidity로 변경

gyeyang1\_Humidity <- gyeyang1\_Humidity %>% filter(!is.na(hour)) # 결측치 제거

4. Humidity 변수 전처리

table(gyeyang1\_Humidity$Humidity) # 습도 이상치 확인

gyeyang1\_Humidity$Humidity <- ifelse(gyeyang1\_Humidity$Humidity == -1 | gyeyang1\_Humidity$Humidity == 0, NA, gyeyang1\_Humidity$Humidity)

상대습도로 존재할 수 없는 -1과 0이 굉장히 많이 있는 것을 알 수 있다. 따라서 계양동과 가장 가까운 곳에 있는 관측장소인 인천관측소에서 측정한 습도 데이터로 이상치를 대체한다.

스크린샷이(가) 표시된 사진

자동 생성된 설명

5. 계양1동 week 변수 생성 (기온데이터와 동일하므로 코드 포함 X)

6. 주별 평균 상대습도 표 생성 (기온데이터와 동일한 방식이므로 코드 포함X)

7. 결측치가 발생한 곳에 인천광역시 습도 데이터로 대체시킨다.

gyeyang1\_Humidity\_final$mean\_Humidity[15:31] <- c(83.21429, 78.64286, 75.50000, 64.62857, 77.10000, 58.27143, 73.12857, 82.84286, 67.81429, 62.08571, 74.95714, 53.98571, 67.65714, 52.87143, 60.90000, 62.85714, 61.81429)

검암 경서동도 계양1동과 같이 위의 방식으로 결측치를 대체시켜주고 이외의 지역은 결측치가 적어 결측치를 제외하고 주별 평균 상대습도 표를 만든다.

**5. 습도와 모기개체수 간의 관계표 만들기(기온데이터와 동일한 방식이므로 소스코드는 포함X)**

1. Humidity 데이터 생성 : 행정구역의 주별 평균 습도 데이터를 Humidity 데이터에 합침.
2. Final\_Humidity 데이터 생성 : week와 채집장소를 기준으로 Humidity 데이터와 mosquito 데이터를 합침.
3. Humidity\_Mosquito 데이터 생성 : 습도와 모기개체수 간의 관계표 생성

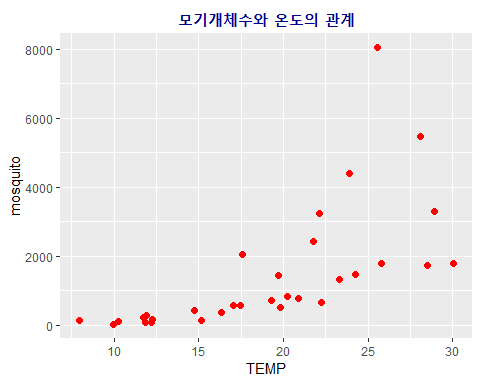
**6. 습도와 모기개체수 간의 관계, 온도와 모기개체수 간의 관계 표를 합친 최종 Final 데이터 생성**

Final <- merge(TEMP\_Mosquito, Humidity\_Mosquito, by = c("월", "week", "mosquito"))  
Final <- Final %>%   
 arrange(week)  
head(Final, 3)

## 월 week mosquito TEMP Humidity  
## 1 4 14 116 10.234127 70.75223  
## 2 4 15 13 9.943204 60.34127  
## 3 4 16 88 11.872619 54.67708

*다른 행정구역도 계양1동의 습도 데이터 전처리과정과 같다.*

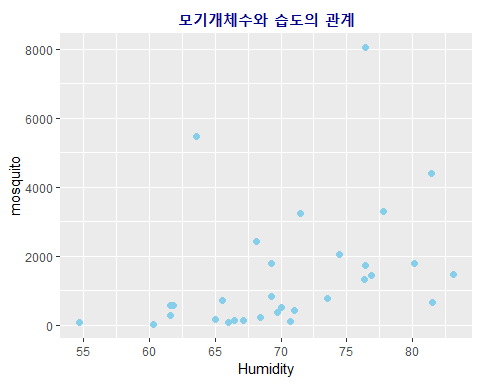
**7. 온도와 모기개체수 간의 상관관계를 보여주는 인터랙티브 그래프 만들기**  
ggplot(data = Final, aes(x = TEMP, y = mosquito)) +   
 geom\_point(size = 2, color = "red") +  
 ggtitle("모기개체수와 온도의 관계") +  
 theme(plot.title = element\_text(face = "bold", hjust = 0.5, size = 15, color = "darkblue"))



*인터랙티브 그래프 :* [*file:///C:/Users/pc/Desktop/R/FinalProject/Interactive\_TEMP.html*](file:///C:/Users/pc/Desktop/R/FinalProject/Interactive_TEMP.html)

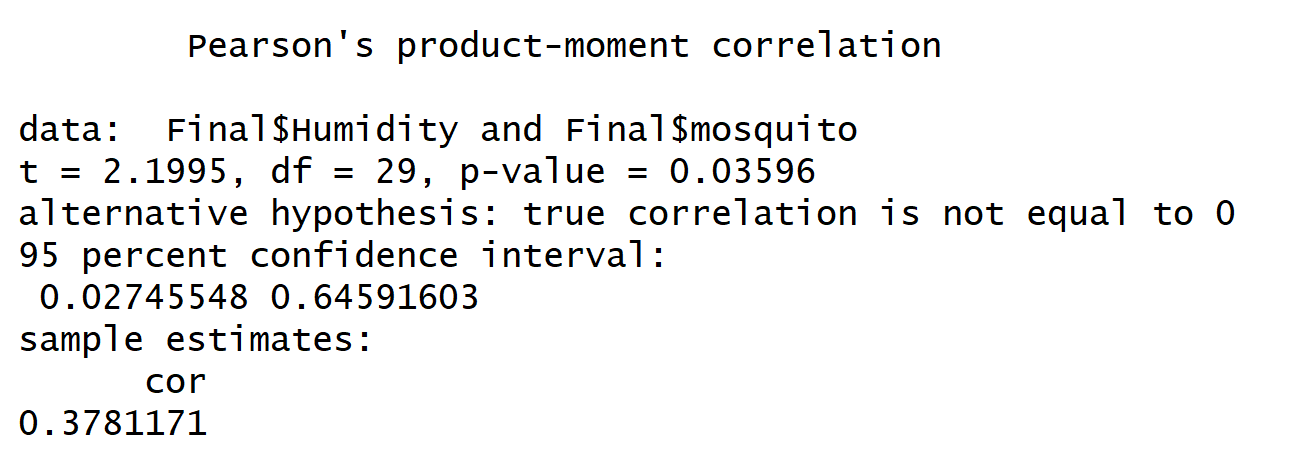
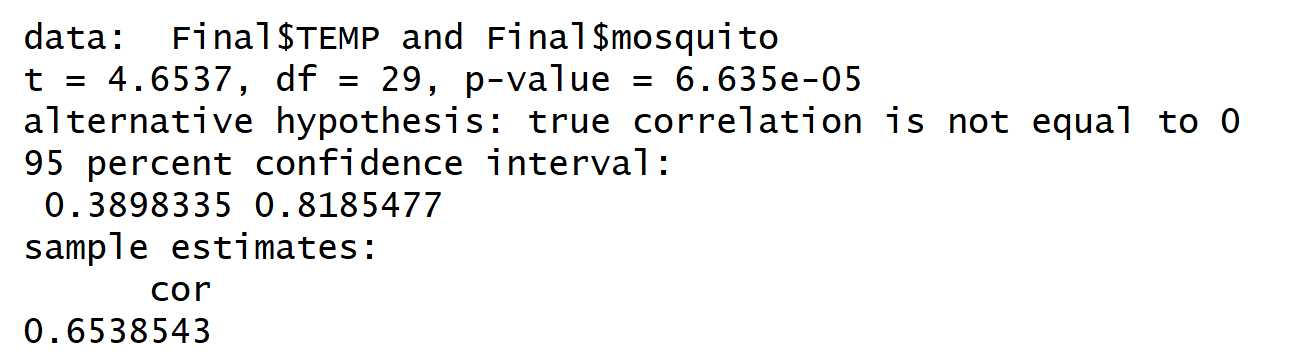
**8. 습도와 모기개체수 간의 상관관계를 보여주는 인터랙티브 그래프 만들기**

ggplot(data = Final, aes(x = Humidity, y = mosquito)) +   
 geom\_point(size = 2, color = "skyblue") +  
 ggtitle("모기개체수와 습도의 관계") +  
 theme(plot.title = element\_text(face = "bold", hjust = 0.5, size = 15, color = "darkblue"))

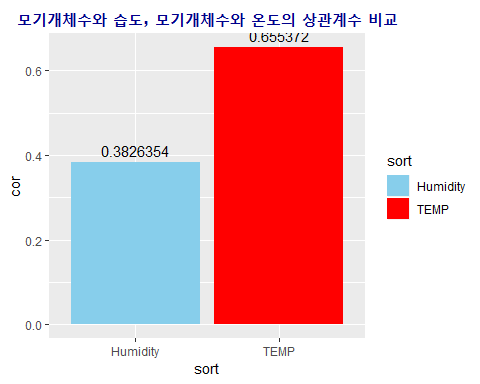


*인터랙티브 그래프 :* [*file:///C:/Users/pc/Desktop/R/FinalProject/Interactive\_Humidity.html*](file:///C:/Users/pc/Desktop/R/FinalProject/Interactive_Humidity.html)

**9. 온도와 모기개체수, 습도와 모기개체수의 상관계수 확인**



cor <- data.frame(sort = c("TEMP", "Humidity"),  
 cor = c(0.655372, 0.3826354))  
ggplot(data = cor, aes(x = sort, y = cor, fill = sort)) +  
 geom\_col() +  
 scale\_fill\_manual(values = c("skyblue", "red")) +  
 ggtitle("모기개체수와 습도, 모기개체수와 온도의 상관계수 비교") +  
 theme(plot.title = element\_text(face = "bold", hjust = 0.5, size = 15, color = "darkblue")) +  
 geom\_text(mapping = aes(label = cor), size = 4, vjust = -0.5)



*인터랙티브 그래프 :* [*file:///C:/Users/pc/Desktop/R/FinalProject/Interactive\_Final.html*](file:///C:/Users/pc/Desktop/R/FinalProject/Interactive_Final.html)

**10. 의사결정 결론** : 온도와 모기개체수의 상관계수는 0.6538543로 온도가 높아질수록 모기 개체수가 증가한다는 상관관계가 있고, 습도와 모기개체수의 상관계수는 0.3826354로 서로 관련이 많지 않다.

**11. 본인만의 차별화 분석 포인트**

1. 산점도 그래프와 막대그래프에서 온도와 습도를 구분하는 방법으로 온도하면 떠오르는 색인 빨간색을, 습도하면 떠오르는 색깔인 파란색을 사용하였다.

2. 인터랙티브 그래프를 사용하여 세부적인 데이터를 알 수 있도록 하였다.

*Github :* [*https://github.com/EUNHAGYEONG/Rdata\_*](https://github.com/EUNHAGYEONG/Rdata_)