

평일, 주말의 발병률차이 분산분석석

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
library(dplyr)
```

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

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
## filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
## intersect, setdiff, setequal, union
```

```
library(tidyr)  
library(gvlma)  
library(FinCal)
```

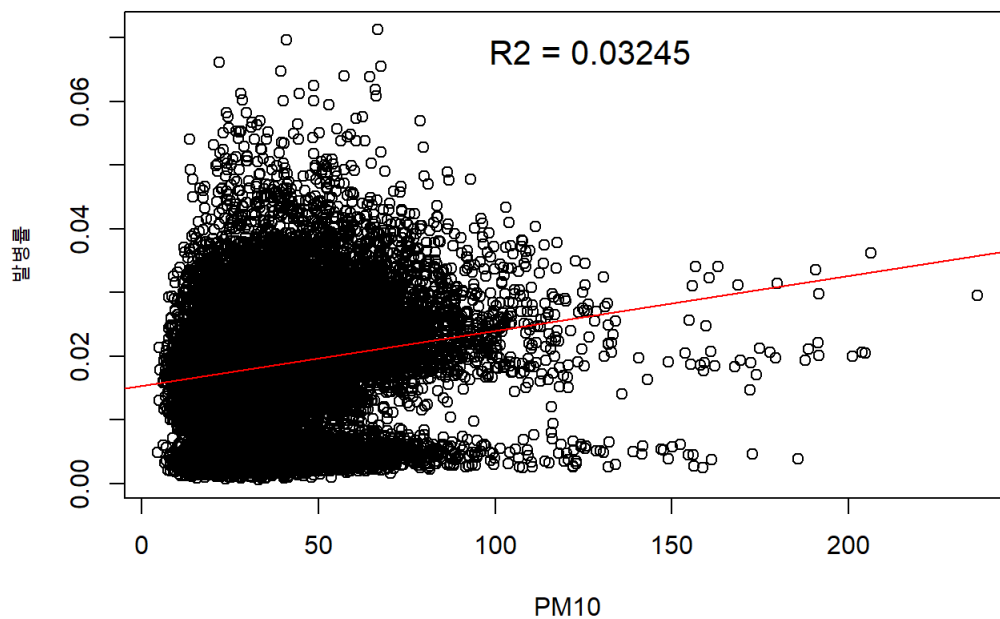
```
## Warning: package 'FinCal' was built under R version 3.6.3
```

```
load('../..../refinedata/analysis/analysis_total_Fixed.rda')
```

```
fit <- lm(발병률 ~ PM10, analysis_total_Fixed)  
summary(fit)
```

```
##  
## Call:  
## lm(formula = 발병률 ~ PM10, data = analysis_total_Fixed)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -0.027529 -0.005917  0.000352  0.005623  0.050752   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) 1.540e-02  1.548e-04   99.49  <2e-16 ***  
## PM10         8.629e-05  3.449e-06   25.02  <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.009506 on 18630 degrees of freedom  
## Multiple R-squared:  0.0325, Adjusted R-squared:  0.03245   
## F-statistic: 625.9 on 1 and 18630 DF,  p-value: < 2.2e-16
```

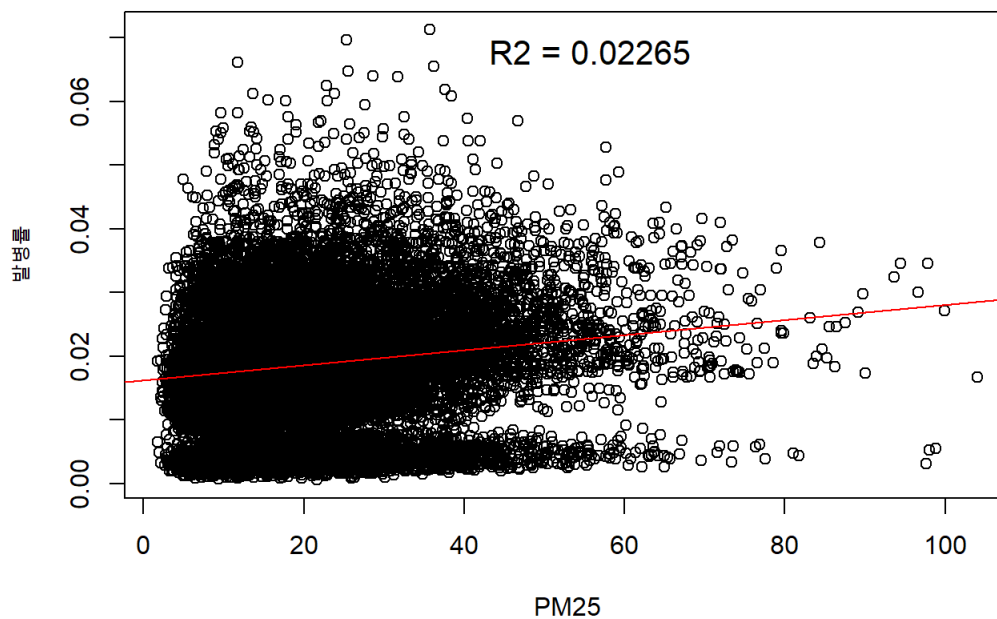
```
plot(발병률 ~ PM10, analysis_total_Fixed)  
abline(fit, col = 'red')  
legend("top", bty="n", cex = 1.3,  
       legend=paste("R2 =",  
                     format(summary(fit)$adj.r.squared, digits=4)))
```



```
fit <- lm(발병률 ~ PM25, analysis_total_Fixed)
summary(fit)
```

```
##
## Call:
## lm(formula = 발병률 ~ PM25, data = analysis_total_Fixed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.024703 -0.006169  0.000352  0.005731  0.050741
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.626e-02  1.433e-04   113.5  <2e-16 ***
## PM25        1.181e-04  5.675e-06    20.8  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009554 on 18630 degrees of freedom
## Multiple R-squared:  0.02271,    Adjusted R-squared:  0.02265
## F-statistic: 432.8 on 1 and 18630 DF,  p-value: < 2.2e-16
```

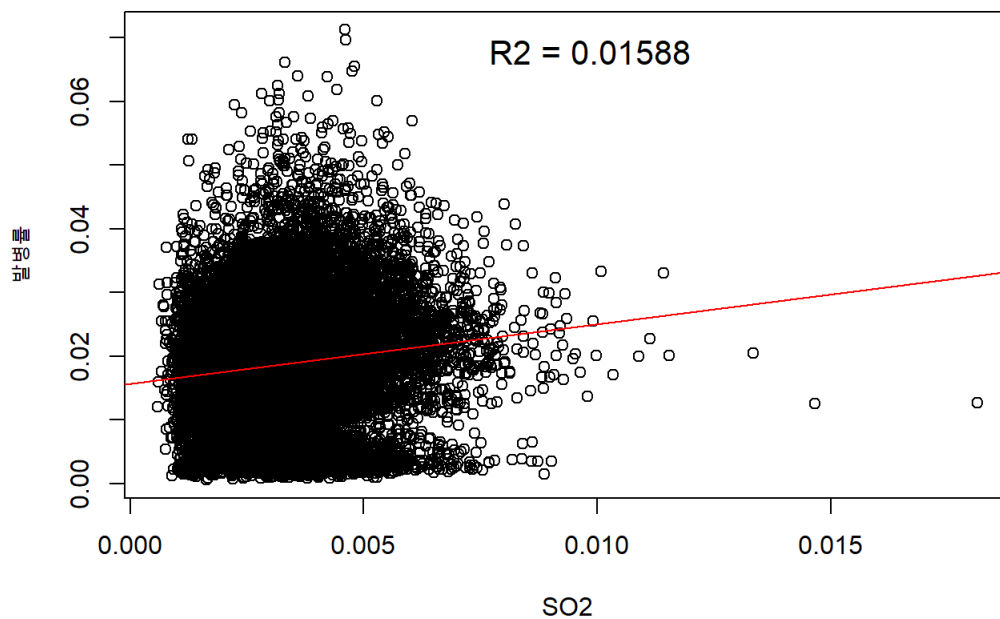
```
plot(발병률 ~ PM25, analysis_total_Fixed)
abline(fit, col = 'red')
legend("top", bty="n", cex = 1.3,
      legend=paste("R2 =",
                    format(summary(fit)$adj.r.squared, digits=4)))
```



```
fit <- lm(발병률 ~ SO2, analysis_total_Fixed)
summary(fit)
```

```
##
## Call:
## lm(formula = 발병률 ~ SO2, data = analysis_total_Fixed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.022476 -0.006385  0.000399  0.005758  0.051214
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0157114  0.0001943   80.85  <2e-16 ***
## SO2          0.9328686  0.0537204   17.36  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009588 on 18630 degrees of freedom
## Multiple R-squared:  0.01593,    Adjusted R-squared:  0.01588
## F-statistic: 301.6 on 1 and 18630 DF,  p-value: < 2.2e-16
```

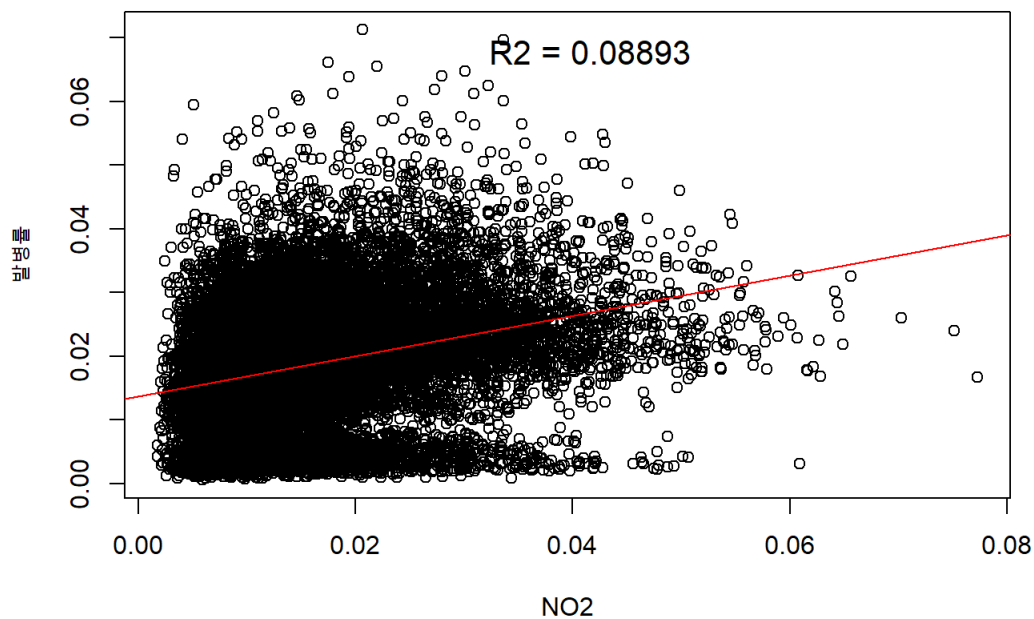
```
plot(발병률 ~ SO2, analysis_total_Fixed)
abline(fit <- lm(발병률 ~ SO2, analysis_total_Fixed), col = 'red')
legend("top", bty="n", cex = 1.3,
      legend=paste("R2 =",
                    format(summary(fit)$adj.r.squared, digits=4)))
```



```
fit <- lm(발병률 ~ NO2, analysis_total_Fixed)
summary(fit)
```

```
##
## Call:
## lm(formula = 발병률 ~ NO2, data = analysis_total_Fixed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.029929 -0.005693  0.000156  0.005468  0.050973
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0136711  0.0001391   98.28  <2e-16 ***
## NO2          0.3172497  0.0074373   42.66  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009225 on 18630 degrees of freedom
## Multiple R-squared:  0.08898,    Adjusted R-squared:  0.08893
## F-statistic: 1820 on 1 and 18630 DF,  p-value: < 2.2e-16
```

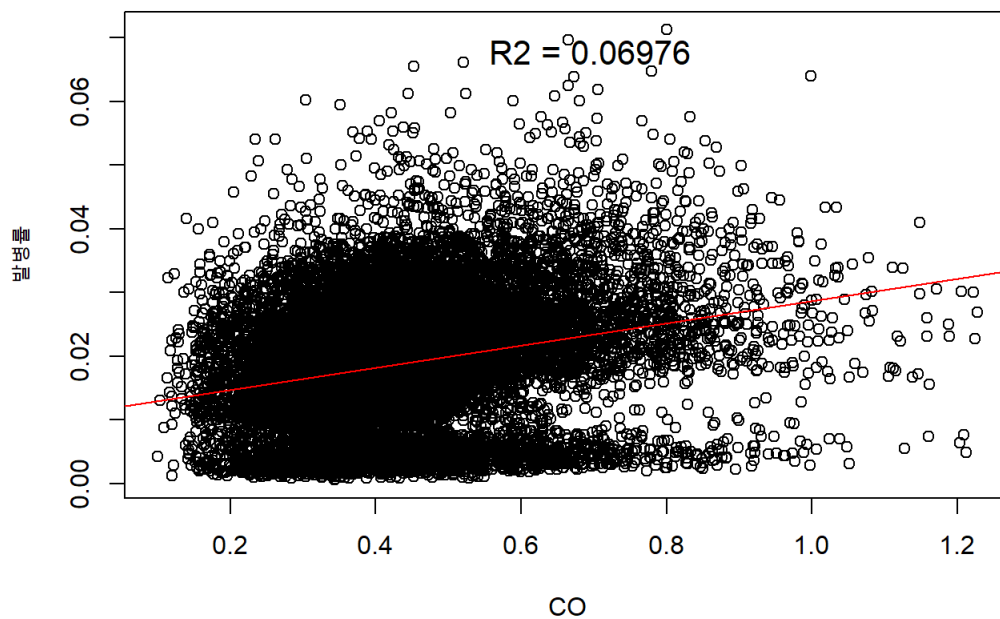
```
plot(발병률 ~ NO2, analysis_total_Fixed)
abline(fit <- lm(발병률 ~ NO2, analysis_total_Fixed), col = 'red')
legend("top", bty="n", cex = 1.3,
      legend=paste("R2 =",
                    format(summary(fit)$adj.r.squared, digits=4)))
```



```
fit <- lm(발병률 ~ CO, analysis_total_Fixed)
summary(fit)
```

```
##
## Call:
## lm(formula = 발병률 ~ CO, data = analysis_total_Fixed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.027487 -0.005492  0.000348  0.005608  0.046812
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0112570  0.0002144   52.50  <2e-16 ***
## CO           0.0174698  0.0004672   37.39  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009321 on 18630 degrees of freedom
## Multiple R-squared:  0.06981,    Adjusted R-squared:  0.06976
## F-statistic: 1398 on 1 and 18630 DF,  p-value: < 2.2e-16
```

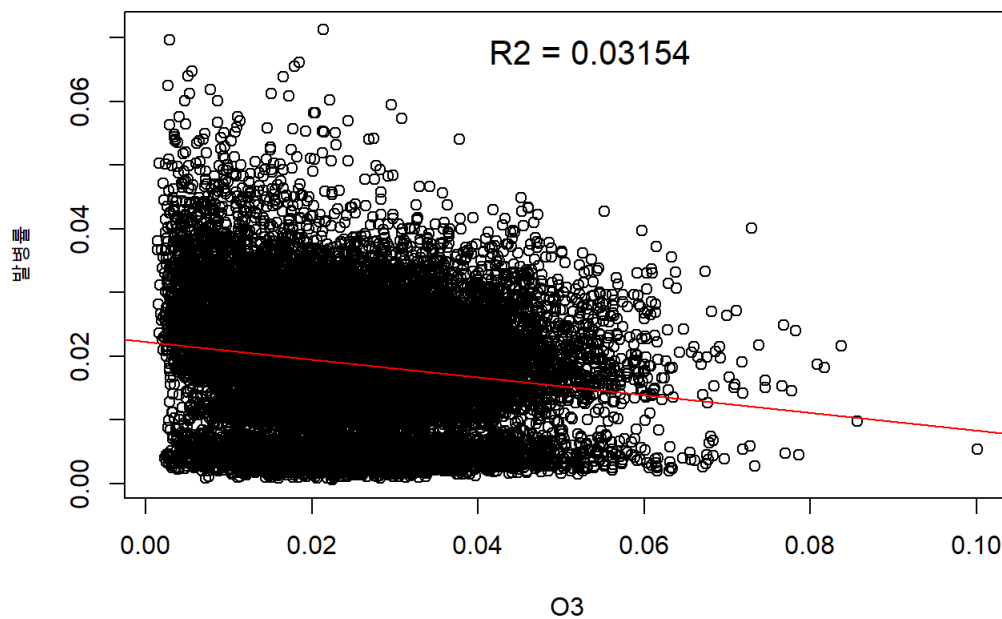
```
plot(발병률 ~ CO, analysis_total_Fixed)
abline(fit <- lm(발병률 ~ CO, analysis_total_Fixed), col = 'red')
legend("top", bty="n", cex = 1.3,
      legend=paste("R2 =",
                    format(summary(fit)$adj.r.squared, digits=4)))
```



```
fit <- lm(발병률 ~ O3, analysis_total_Fixed)
summary(fit)
```

```
##
## Call:
## lm(formula = 발병률 ~ O3, data = analysis_total_Fixed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.020415 -0.006575  0.000448  0.005728  0.051936
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.022243   0.000154  144.44  <2e-16 ***
## O3          -0.138942   0.005636  -24.65  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009511 on 18630 degrees of freedom
## Multiple R-squared:  0.03159,    Adjusted R-squared:  0.03154
## F-statistic: 607.7 on 1 and 18630 DF,  p-value: < 2.2e-16
```

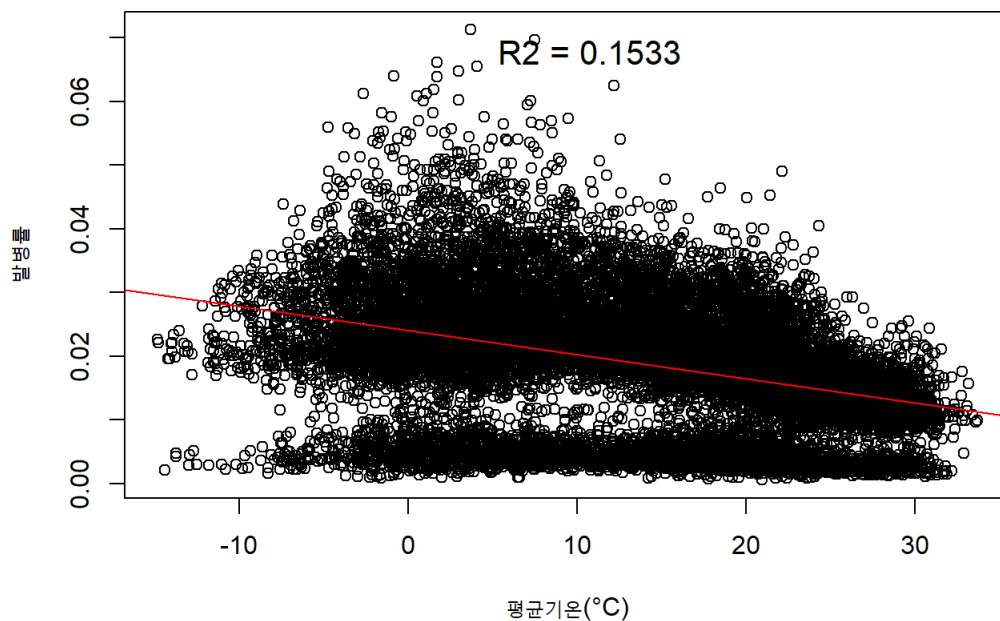
```
plot(발병률 ~ O3, analysis_total_Fixed)
abline(fit <- lm(발병률 ~ O3, analysis_total_Fixed), col = 'red')
legend("top", bty="n", cex = 1.3,
      legend=paste("R2 =",
                    format(summary(fit)$adj.r.squared, digits=4)))
```



```
fit <- lm(발병률 ~ `평균기온(°C)`, analysis_total_Fixed)
summary(fit)
```

```
##
## Call:
## lm(formula = 발병률 ~ `평균기온(°C)`, data = analysis_total_Fixed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.027358 -0.003804  0.000516  0.004799  0.048605
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.400e-02  1.100e-04   218.28  <2e-16 ***
## `평균기온(°C)` -3.781e-04  6.508e-06   -58.09  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.008893 on 18630 degrees of freedom
## Multiple R-squared:  0.1534, Adjusted R-squared:  0.1533
## F-statistic: 3374 on 1 and 18630 DF, p-value: < 2.2e-16
```

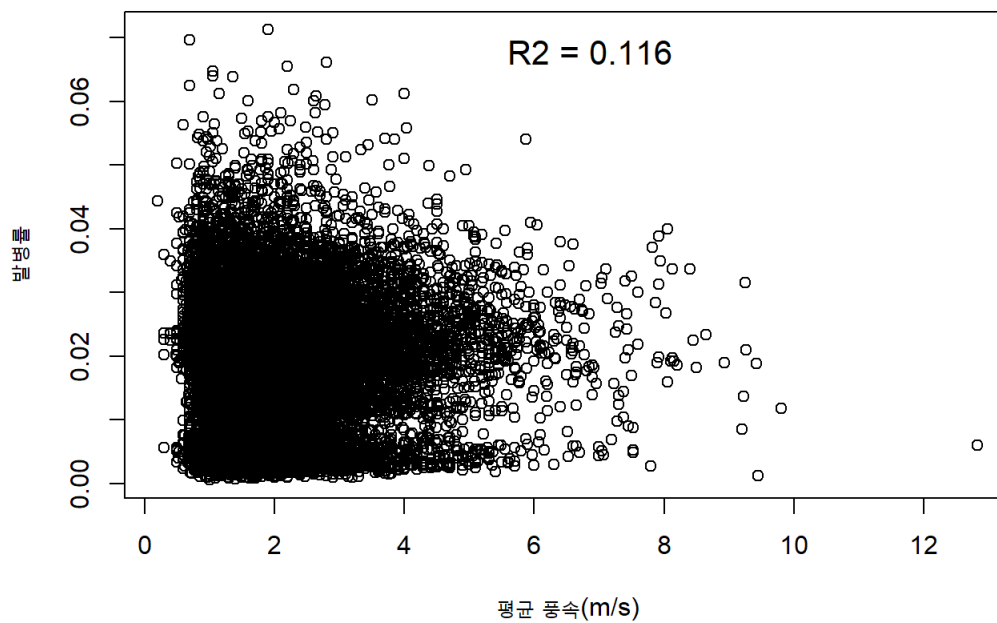
```
plot(발병률 ~ `평균기온(°C)`, analysis_total_Fixed)
abline(fit, col = 'red')
legend("top", bty="n", cex = 1.3,
      legend=paste("R2 =",
                    format(summary(fit)$adj.r.squared, digits=4)))
```



```
fit <- lm(발병률 ~ `평균 현지기압(hPa)`, analysis_total_Fixed)
summary(fit)
```

```
##
## Call:
## lm(formula = 발병률 ~ `평균 현지기압(hPa)`, data = analysis_total_Fixed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.023007 -0.004679  0.000427  0.005189  0.048897
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -3.361e-01  7.179e-03  -46.81  <2e-16 ***
## `평균 현지기압(hPa)`  3.527e-04  7.134e-06   49.44  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009087 on 18630 degrees of freedom
## Multiple R-squared:  0.116, Adjusted R-squared:  0.116
## F-statistic: 2445 on 1 and 18630 DF, p-value: < 2.2e-16
```

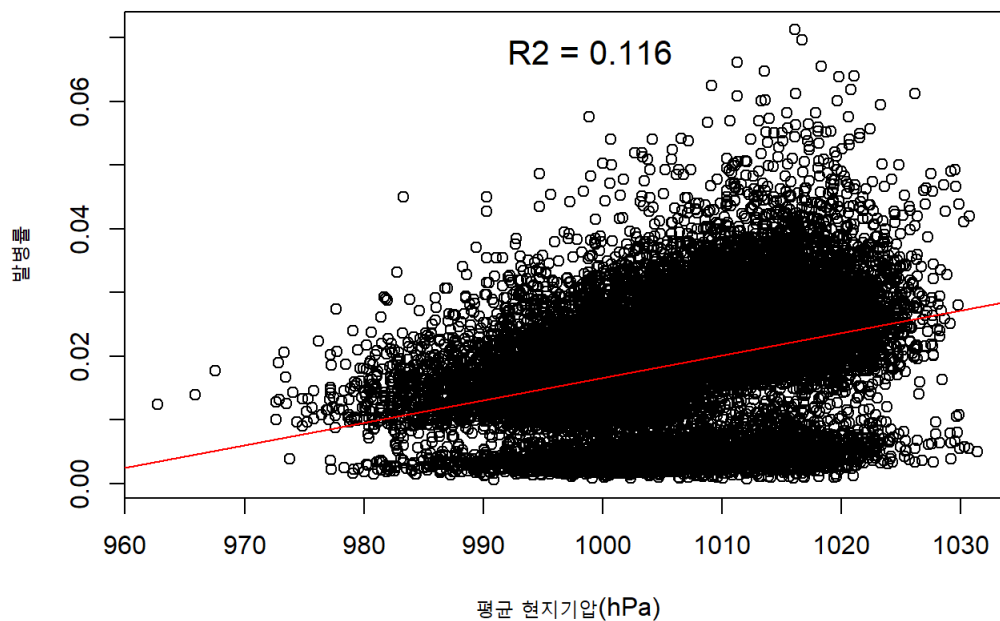
```
plot(발병률 ~ `평균 풍속(m/s)`, analysis_total_Fixed)
abline(fit, col = 'red')
legend("top", bty="n", cex = 1.3,
      legend=paste("R2 =",
                    format(summary(fit)$adj.r.squared, digits=4)))
```

```
fit <- lm(발병률 ~ `평균 현지기압(hPa)` ,analysis_total_Fixed)
summary(fit)
```

```
##
## Call:
## lm(formula = 발병률 ~ `평균 현지기압(hPa)` , data = analysis_total_Fixed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.023007 -0.004679  0.000427  0.005189  0.048897
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -3.361e-01  7.179e-03  -46.81  <2e-16 ***
## `평균 현지기압(hPa)`  3.527e-04  7.134e-06   49.44  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009087 on 18630 degrees of freedom
## Multiple R-squared:  0.116, Adjusted R-squared:  0.116
## F-statistic: 2445 on 1 and 18630 DF, p-value: < 2.2e-16
```

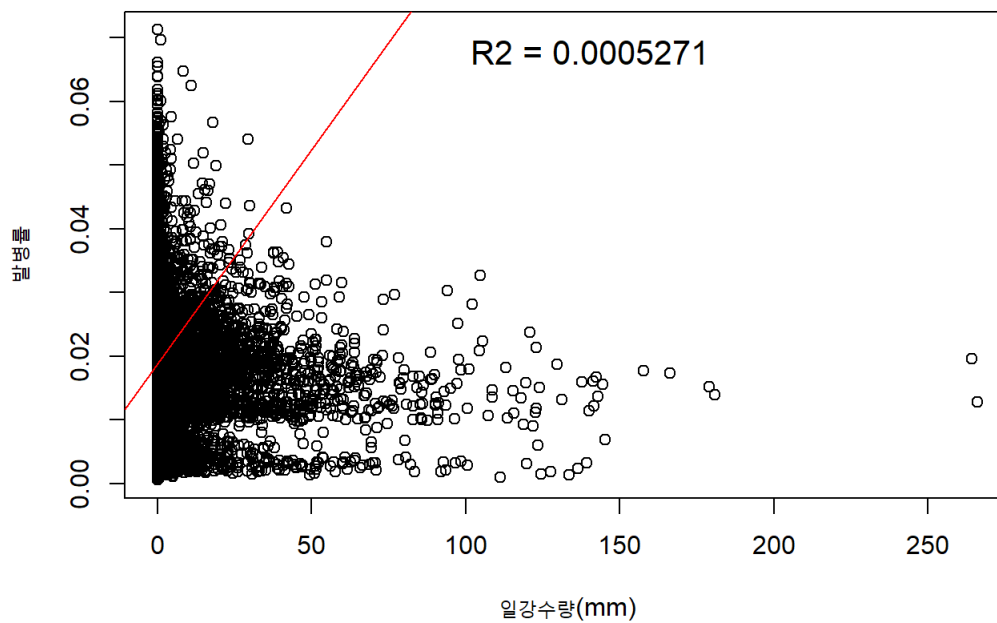
```
plot(발병률 ~ `평균 현지기압(hPa)` ,analysis_total_Fixed)
abline(fit, col = 'red')
legend("top", bty="n", cex = 1.3,
      legend=paste("R2 =",
                    format(summary(fit)$adj.r.squared, digits=4)))
```



```
fit <- lm(발병률 ~ `일 최심신적설(cm)`, analysis_total_Fixed)
summary(fit)
```

```
##
## Call:
## lm(formula = 발병률 ~ `일 최심신적설(cm)`, data = analysis_total_Fixed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.025459 -0.006641  0.000438  0.005861  0.052375
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.884e-02  7.111e-05  264.86  <2e-16 ***
## `일 최심신적설(cm)` 6.732e-04  2.046e-04   3.29   0.001 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009662 on 18630 degrees of freedom
## Multiple R-squared:  0.0005807, Adjusted R-squared:  0.0005271
## F-statistic: 10.83 on 1 and 18630 DF, p-value: 0.001003
```

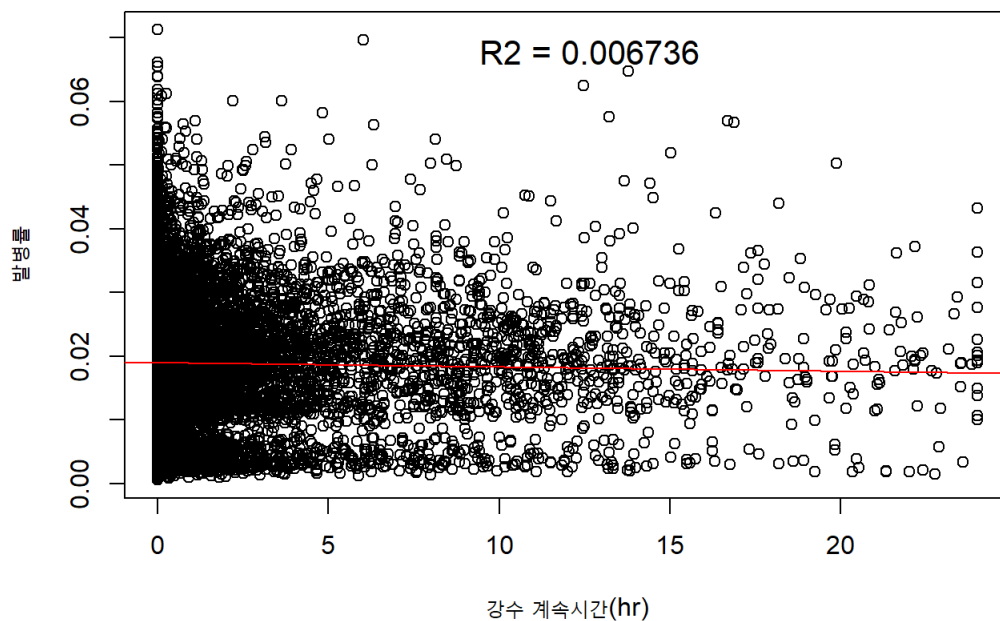
```
plot(발병률 ~ `일강수량(mm)`, analysis_total_Fixed)
abline(fit, col = 'red')
legend("top", bty="n", cex = 1.3,
      legend=paste("R2 =",
                    format(summary(fit)$adj.r.squared, digits=4)))
```



```
fit <- lm(발병률 ~ `일강수량(mm)`, analysis_total_Fixed)
summary(fit)
```

```
##
## Call:
## lm(formula = 발병률 ~ `일강수량(mm)`, data = analysis_total_Fixed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.018497 -0.006516  0.000494  0.005803  0.052122
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.909e-02  7.347e-05  259.81  <2e-16 ***
## `일강수량(mm)` -6.800e-05  6.026e-06 -11.29  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009632 on 18630 degrees of freedom
## Multiple R-squared:  0.006789,    Adjusted R-squared:  0.006736
## F-statistic: 127.3 on 1 and 18630 DF,  p-value: < 2.2e-16
```

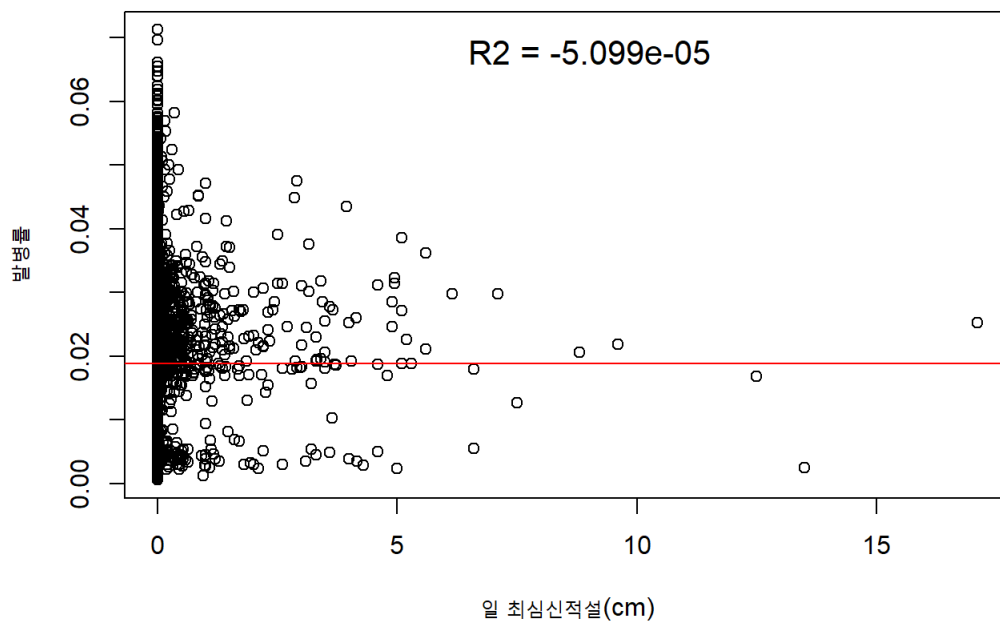
```
plot(발병률 ~ `강수 계속시간(hr)`, analysis_total_Fixed)
abline(fit, col = 'red')
legend("top", bty="n", cex = 1.3,
      legend=paste("R2 =",
                    format(summary(fit)$adj.r.squared, digits=4)))
```



```
fit <- lm(발병률 ~ `강수 계속시간(hr)`, analysis_total_Fixed)
summary(fit)
```

```
##
## Call:
## lm(formula = 발병률 ~ `강수 계속시간(hr)`, data = analysis_total_Fixed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.018260 -0.006652  0.000448  0.005856  0.052359
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.885e-02  7.612e-05  247.651  <2e-16 ***
## `강수 계속시간(hr)` 5.372e-06  2.400e-05   0.224    0.823
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009665 on 18630 degrees of freedom
## Multiple R-squared:  2.689e-06, Adjusted R-squared: -5.099e-05
## F-statistic: 0.0501 on 1 and 18630 DF, p-value: 0.8229
```

```
plot(발병률 ~ `일 최심신적설(cm)`, analysis_total_Fixed)
abline(fit, col = 'red')
legend("top", bty="n", cex = 1.3,
      legend=paste("R2 =",
                    format(summary(fit)$adj.r.squared, digits=4)))
```



```
analysis_total_Fixed <- as.data.frame(analysis_total_Fixed)

analysis_total_Fixed$day <- weekdays(analysis_total_Fixed$일시)
analysis_total_Fixed$weekend <- ifelse(analysis_total_Fixed$day %in% c('토요일', '일요일'), '주말', '평일')

summary(group_by(analysis_total_Fixed, weekend))
```

```
## 시도코드 일시 발생건수 시도
## 11 : 1096 Min. :2016-01-01 Min. : 4.0 Length:18632
## 26 : 1096 1st Qu.:2016-09-30 1st Qu.: 164.0 Class :character
## 27 : 1096 Median :2017-07-01 Median : 350.0 Mode :character
## 28 : 1096 Mean :2017-07-01 Mean : 564.3
## 29 : 1096 3rd Qu.:2018-04-01 3rd Qu.: 590.0
## 30 : 1096 Max. :2018-12-31 Max. :7197.0
## (Other):12056
## 평균기온 (°C) 최저기온 (°C) 최고기온 (°C) 평균 풍속 (m/s)
## Min. : -14.840 Min. : -19.8000 Min. : -10.93 Min. : 0.200
## 1st Qu.: 5.216 1st Qu.: 0.3321 1st Qu.: 10.62 1st Qu.: 1.327
## Median : 14.696 Median : 9.7375 Median : 20.30 Median : 1.800
## Mean : 13.612 Mean : 9.1073 Mean : 18.83 Mean : 2.080
## 3rd Qu.: 22.000 3rd Qu.: 18.0000 3rd Qu.: 27.07 3rd Qu.: 2.580
## Max. : 33.700 Max. : 30.3000 Max. : 39.60 Max. :12.825
##
## 평균 현지기압 (hPa) 일 최심신적설 (cm) 일강수량 (mm) 강수 계속시간 (hr)
## Min. : 962.7 Min. : 0.00000 Min. : 0.000 Min. : 0.0000
## 1st Qu.:1000.1 1st Qu.: 0.00000 1st Qu.: 0.000 1st Qu.: 0.0000
## Median :1006.5 Median : 0.00000 Median : 0.000 Median : 0.0000
## Mean :1006.3 Mean : 0.03338 Mean : 3.396 Mean : 1.1644
## 3rd Qu.:1013.4 3rd Qu.: 0.00000 3rd Qu.: 0.640 3rd Qu.: 0.7609
## Max. :1031.4 Max. :17.10000 Max. :266.000 Max. :24.0000
##
## 최다풍향 (16방위) SO2 CO O3
## Min. : 1.000 Min. :0.000587 Min. :0.1000 Min. :0.00144
## 1st Qu.: 5.000 1st Qu.:0.002444 1st Qu.:0.3365 1st Qu.:0.01488
## Median :10.000 Median :0.003246 Median :0.4110 Median :0.02324
## Mean : 9.196 Mean :0.003373 Mean :0.4351 Mean :0.02437
## 3rd Qu.:13.000 3rd Qu.:0.004122 3rd Qu.:0.5077 3rd Qu.:0.03238
## Max. :16.000 Max. :0.018143 Max. :1.2274 Max. :0.10008
##
## NO2 PM10 PM25 년도
## Min. :0.001788 Min. : 4.40 Min. : 1.766 Length:18632
## 1st Qu.:0.009792 1st Qu.: 26.02 1st Qu.: 12.935 Class :character
## Median :0.014133 Median : 36.36 Median : 19.711 Mode :character
## Mean :0.016349 Mean : 40.08 Mean : 22.030
## 3rd Qu.:0.020753 3rd Qu.: 50.01 3rd Qu.: 28.461
## Max. :0.077245 Max. :236.50 Max. :104.009
##
## 인구수 발병률 day weekend
## Min. : 242507 Min. :0.0005914 Length:18632 Length:18632
## 1st Qu.: 1511214 1st Qu.:0.0122051 Class :character Class :character
## Median : 1826174 Median :0.0193046 Mode :character Mode :character
## Mean : 3025904 Mean :0.0188578
## 3rd Qu.: 2936117 3rd Qu.:0.0247080
## Max. :13103188 Max. :0.0712102
##
```

```
table(analysis_total_Fixed$weekend)
```

```
##
## 주말 평일
## 5338 13294
```

```
t.test(formula = 발병률 ~ weekend, data = analysis_total_Fixed)
```

```
##
## Welch Two Sample t-test
##
## data: 발병률 by weekend
## t = -74.974, df = 9799.3, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.01057840 -0.01003934
## sample estimates:
## mean in group 주말 mean in group 평일
## 0.01150235 0.02181122
```

```
boxplot(formula = 발병률 ~ weekend, data = analysis_total_Fixed, col=rainbow(4, alpha = 0.5))
```

