

데이터는 기상청의 청주시의 17년 5월 15일부터 30일까지의 지상기상관측데이터이다.

Temp : 온도 wind_speed : 풍속, see_level : 해면기압
humidity : 습도를 나타낸다.

종속변수 temp

독립변수 wind_speed, see_level, humidity

```
install.packages("perturb")
install.packages("car")
install.packages("lmtest")
library(perturb)
library(car)
library(lmtest)
```

데이터 읽기

```
obs <- read.csv("D:/회귀분석/observation.csv", header=T)
```

상관계수 낮은 (see_level, humidity) 회귀 분석

```
see_level <- lm(temp~see_level, obs) # 다른 독립 변수들보다 덜 유의함.
```

```
summary(see_level)
```

```
humidity <- lm(temp~humidity, obs)
```

```
summary(humidity)
```

```
sh <- lm(temp~see_level+humidity, obs)
```

```
summary(sh)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	208.10783	73.09042	2.847	0.00465 **
see_level	-0.18541	0.07224	-2.566	0.01066 *

다른 변수들보다
덜 유의함

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	30.016613	0.469873	63.88	<2e-16 ***
humidity	-0.204655	0.009521	-21.49	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	198.824014	48.716823	4.081	5.47e-05 ***
see_level	-0.166881	0.048159	-3.465	0.000591 ***
humidity	-0.204080	0.009387	-21.741	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
##상관계수 유의한 (wind_speed, humidity) 회귀 분석
wind_speed <- lm(temp~wind_speed, obs)
summary(wind_speed)
wh <- lm(temp~wind_speed+humidity, obs)
summary(wh)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	15.5666	0.4174	37.29	<2e-16 ***
wind_speed	2.8749	0.2131	13.49	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	26.88612	0.83471	32.210	< 2e-16 ***
wind_speed	0.95830	0.21354	4.488	9.59e-06 ***
humidity	-0.17281	0.01169	-14.784	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##다중공선성

```
model <- lm(temp~., obs)
summary(model) #see_level 유의하지 않으므로 제거
#vif
vif(model)
#condition index
colldiag(model)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	109.90238	55.65614	1.975	0.04904 *
wind_speed	0.77899	0.24475	3.183	0.00158 **
see_level	-0.08149	0.05463	-1.492	0.13660
humidity	-0.17849	0.01228	-14.540	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> vif(model)
```

	see_level	humidity
wind_speed	2.087126	1.752145

```
> colldiag(model)
```

Condition

Index	Variance Decomposition Proportions
	intercept wind_speed see_level humidity
1	1.000 0.000 0.006 0.000 0.004
2	3.652 0.000 0.241 0.000 0.089
3	10.623 0.000 0.506 0.000 0.806
4	960.498 1.000 0.247 1.000 0.100

```
##최종 모형
final<-lm(temp~wind_speed+humidity, obs)
summary(final)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	26.88612	0.83471	32.210	< 2e-16 ***
wind_speed	0.95830	0.21354	4.488	9.59e-06 ***
humidity	-0.17281	0.01169	-14.784	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
##잔차 분석
#정규성
shapiro.test(resid(final))
```

```
#잔차의 등분산성
plot(resid(final))
```

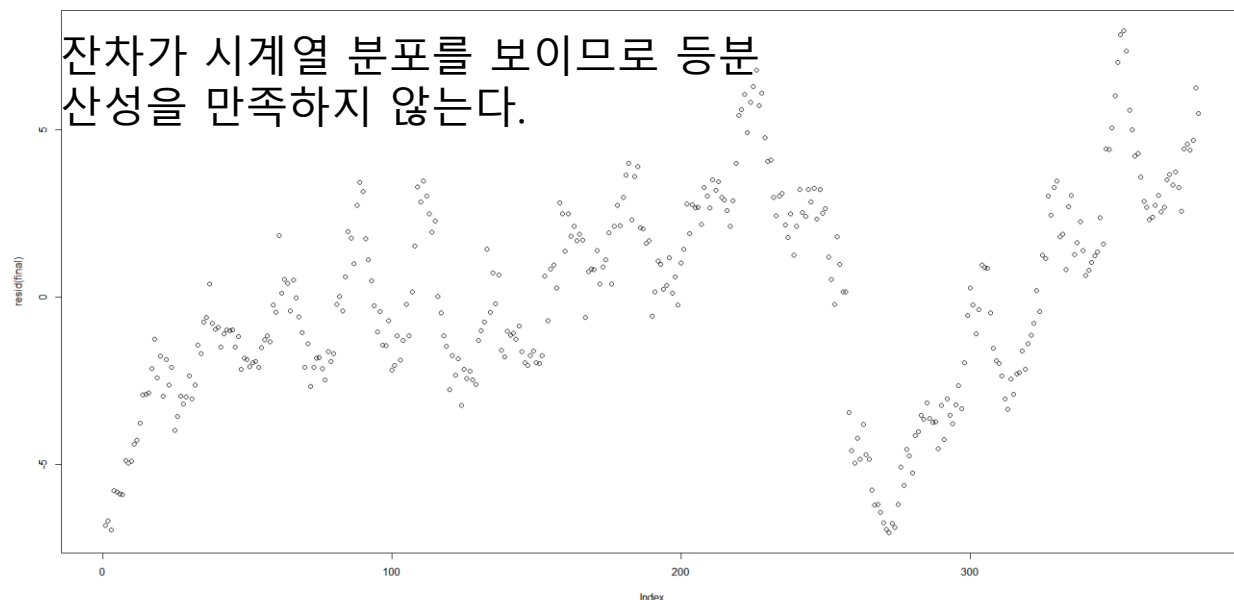
```
#잔차의 독립성
dwtest(final)
```

```
> shapiro.test(resid(final))
```

Shapiro-Wilk normality test

data: resid(final)
W = 0.99196, p-value = 0.03822

P-value 가 0.05미만으로 귀무가설을 기각한다. 즉, 잔차의 정규성을 만족하지 않는다.



잔차가 시계열 분포를 보이므로 등분산성을 만족하지 않는다.

```
> dwtest(final)
```

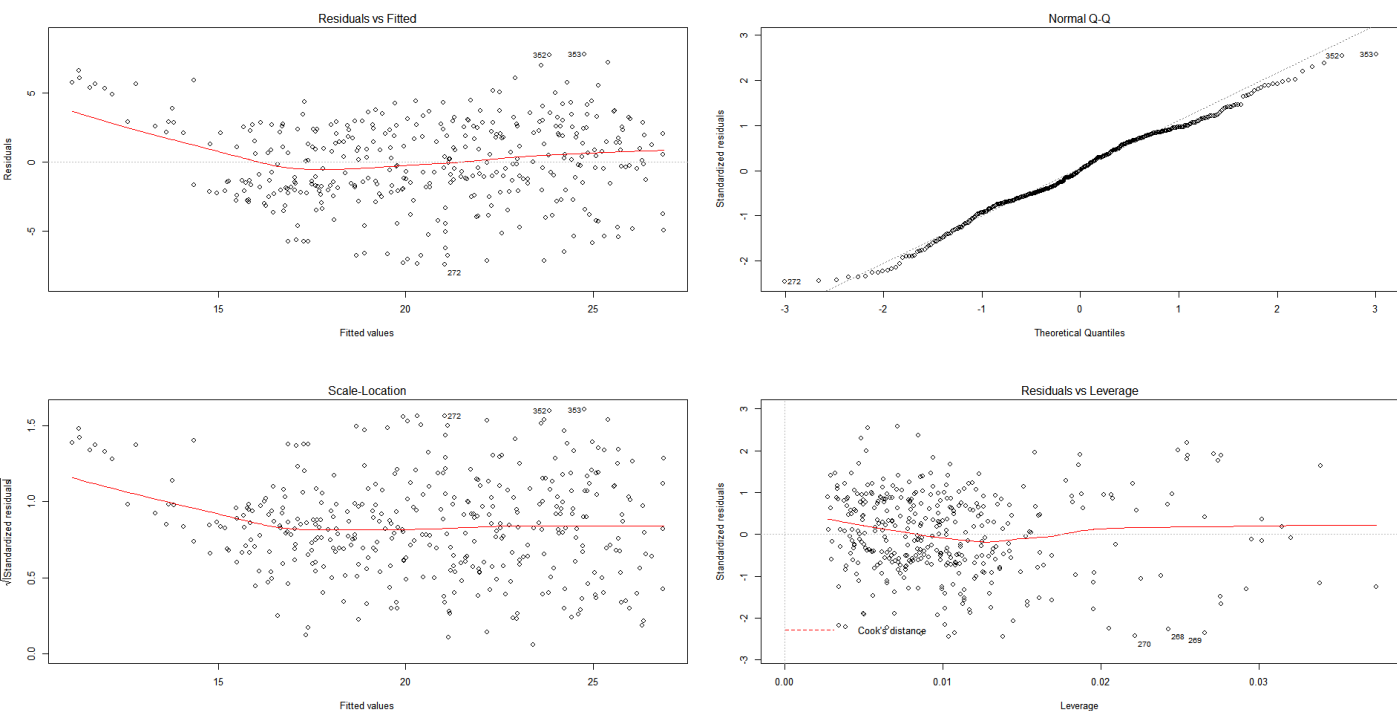
Durbin-Watson test

data: final
DW = 0.081847, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0

Dw-test에서 dw가 0과 가까우므로 독립성을 만족하지 않는다.

위의 잔차분석을 통해 본 결과 정규성, 등분산성, 독립성을 만족하지 않는다.

Plot(model)을 통해서 잔차분석을 다시 해보았다.



Q-Q plot 의 결과, 대각선 선과 확연한 차이가 없으므로 잔차가 정규분포를 따른다고 추측할 수 있다.

또한 잔차의 등분산성과 독립성은 자료 자체가 군데군데 쏠려서 분포되어있어서 만족하지 않음을 알 수 있다.

하지만 선형모형이므로 잔차가 최종모형에 큰 영향을 끼치지 않을것이라고 예측할 수 있다.