

Quetelet's Body Mass Index

coop711

2015년 3월 9일

rn96.txt 자료 읽어들이고, 기초통계 요약하기

- rn96.txt 자료 읽어들이기, 원 자료를 보고 header 매개변수 추가

```
rn96<-read.table("rn96.txt",header=TRUE)
```

- rn96 의 자료구조 살피기.

```
str(rn96)
```

```
## 'data.frame':    41 obs. of  2 variables:
## $ height: int   161 155 158 170 160 156 162 158 158 167 ...
## $ weight: int    50 49 42 65 60 52 58 46 45 51 ...
```

- height 와 weight 의 기초통계 살피기

```
summary(rn96)
```

```
##      height      weight
## Min.   :150.0   Min.    :42.00
## 1st Qu.:156.0   1st Qu.:48.00
## Median :159.0   Median :52.00
## Mean   :159.3   Mean    :52.02
## 3rd Qu.:162.0   3rd Qu.:55.00
## Max.   :170.0   Max.    :65.00
```

- 평균과 표준편차만 살펴보려면

```
apply(rn96,2,mean)
```

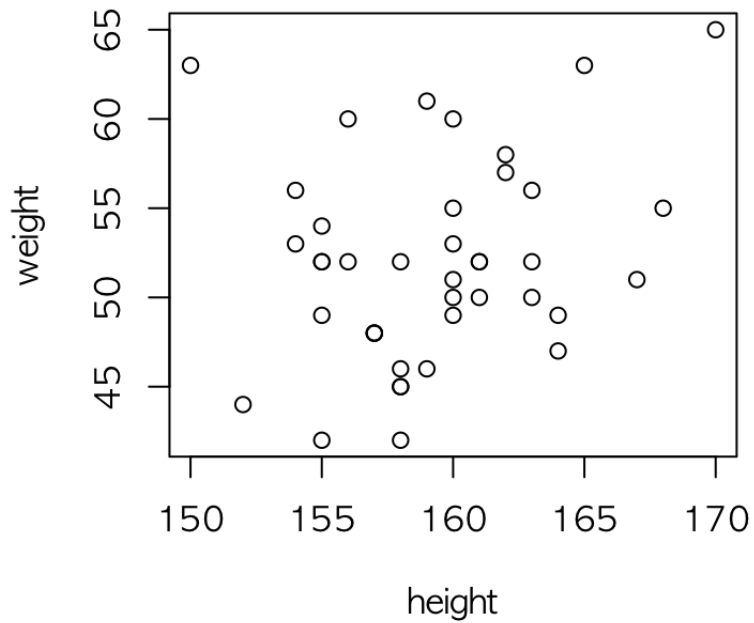
```
##      height      weight
## 159.26829   52.02439
```

```
apply(rn96,2,sd)
```

```
##      height      weight
## 4.266289   5.667838
```

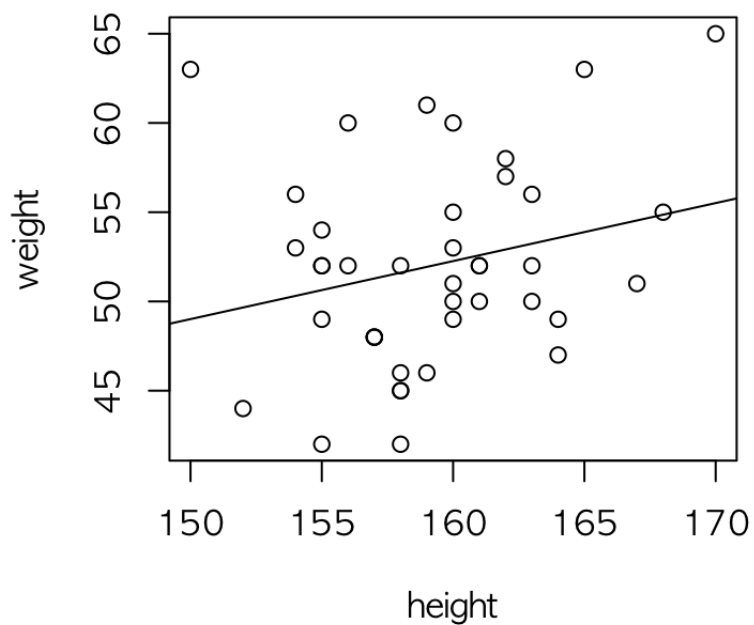
- height 와 weight 의 산점도 그리기

```
plot(weight~height, data=rn96)
```



- 선형회귀선 추가하기

```
plot(weight~height, data=rn96)  
abline(lm(weight~height, data=rn96)$coefficient)
```



- 선형모형으로 분석하기

```
rn96.lm<-lm(weight~height, data=rn96)
```

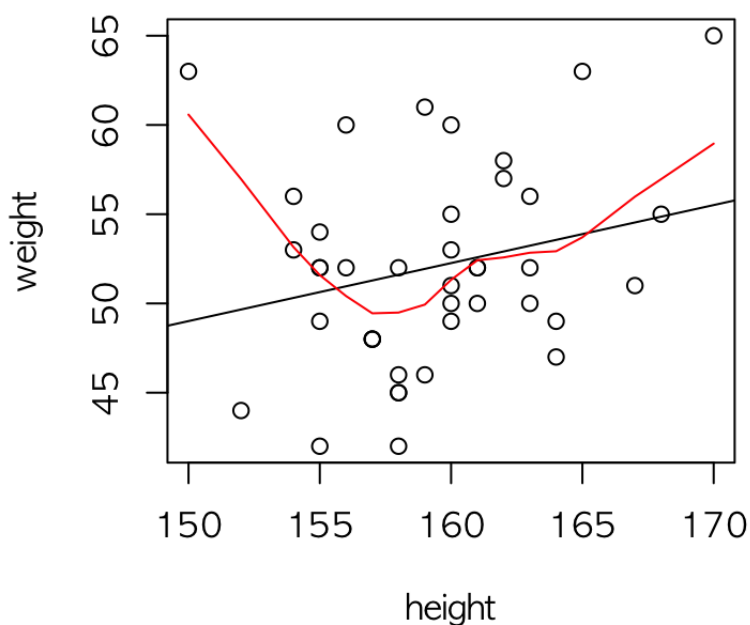
- 회귀계수와 관련 통계량 살펴보기

```
summary(rn96.lm)
```

```
##
## Call:
## lm(formula = weight ~ height, data = rn96)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.6120 -3.2868 -0.5875  2.7622 13.9893
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.2369     32.8626   0.007   0.994
## height        0.3252      0.2063   1.576   0.123
##
## Residual standard error: 5.565 on 39 degrees of freedom
## Multiple R-squared:  0.0599, Adjusted R-squared:  0.0358
## F-statistic: 2.485 on 1 and 39 DF, p-value: 0.123
```

- lowess local smoother 추가하기

```
plot(weight~height, data=rn96)
abline(lm(weight~height, data=rn96)$coefficient)
lines(lowess(rn96$height,rn96$weight),col="red")
```



BMI 계산하고 줄기-잎 그리기

- 조금 편하게 작업하기 위해서

```
attach(rn96)
search()
```

```
## [1] ".GlobalEnv"      "rn96"              "package:stats"
## [4] "package:graphics" "package:grDevices" "package:utils"
## [7] "package:datasets" "KoreaEnv"          "package:methods"
## [10] "Autoloads"        "package:base"
```

- BMI 공식은 $\frac{\text{몸무게}(kg)}{\text{키}(m)^2}$

```
BMI<-round(weight/(height/100)^2,digits=1)
head(cbind(rn96, BMI))
```

```
## height weight BMI
## 1 161 50 19.3
## 2 155 49 20.4
## 3 158 42 16.8
## 4 170 65 22.5
## 5 160 60 23.4
## 6 156 52 21.4
```

- BMI 값들의 줄기-잎 그림 그리기

```
stem(BMI)
```

```
##
## The decimal point is at the |
##
## 16 | 855
## 18 | 00223480135555569
## 20 | 11478145667
## 22 | 1355146
## 24 | 17
## 26 |
## 28 | 0
```

- 조금더 알아보기 쉽게 매개변수를 바꾸기

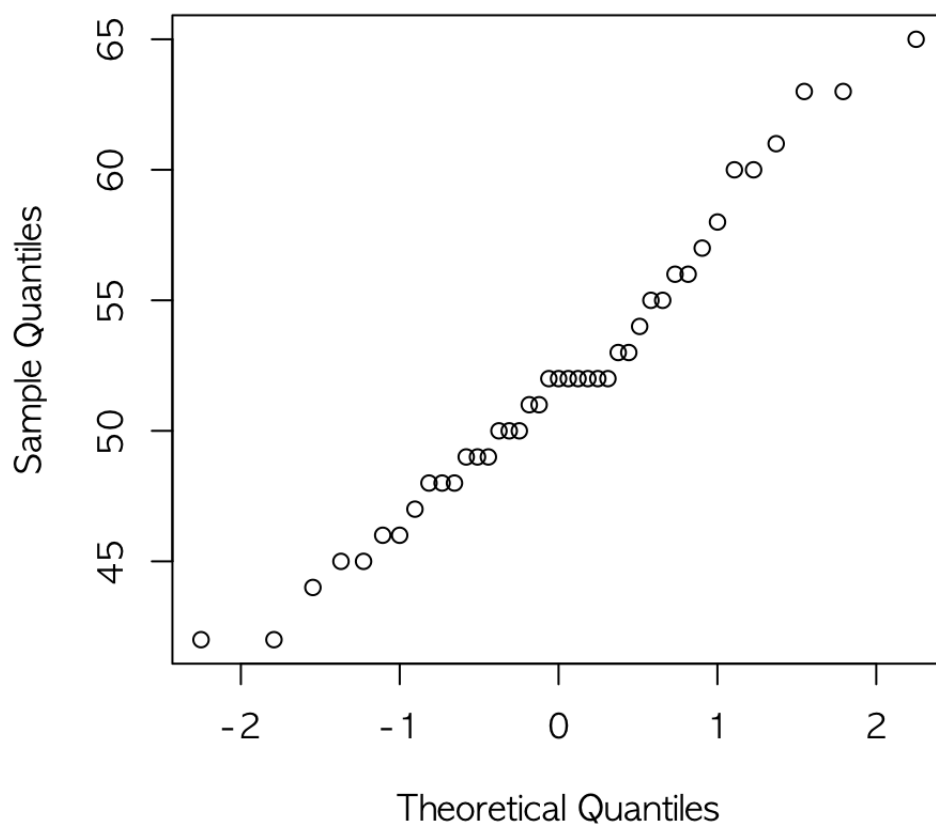
```
stem(BMI,scale=2)
```

```
##
## The decimal point is at the |
##
## 16 | 8
## 17 | 55
## 18 | 0022348
## 19 | 0135555569
## 20 | 11478
## 21 | 145667
## 22 | 1355
## 23 | 146
## 24 | 17
## 25 |
## 26 |
## 27 |
## 28 | 0
```

- 각 변수는정규분포에 가까울까?

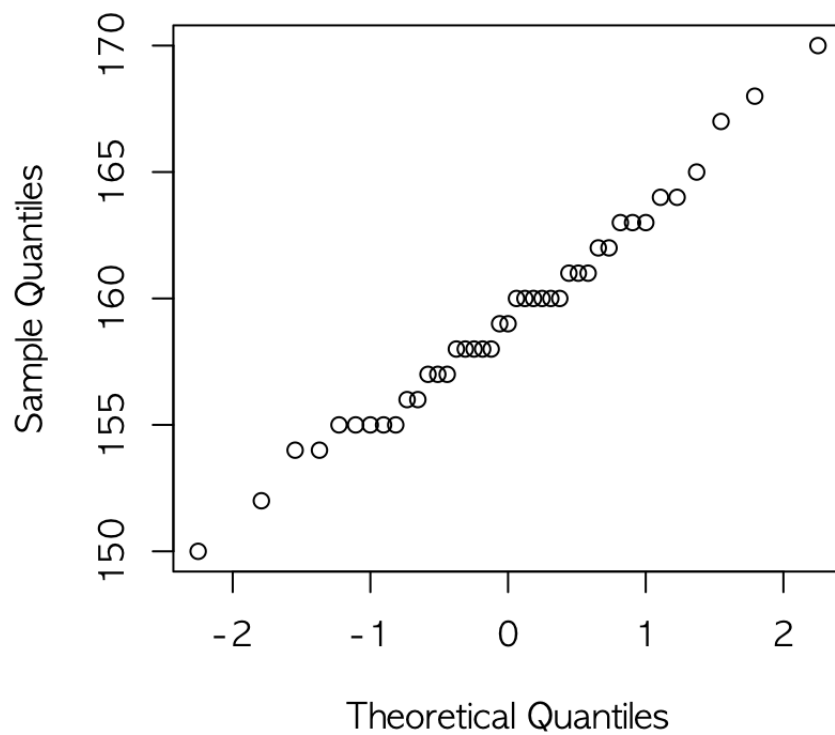
```
qqnorm(weight)
```

Normal Q-Q Plot



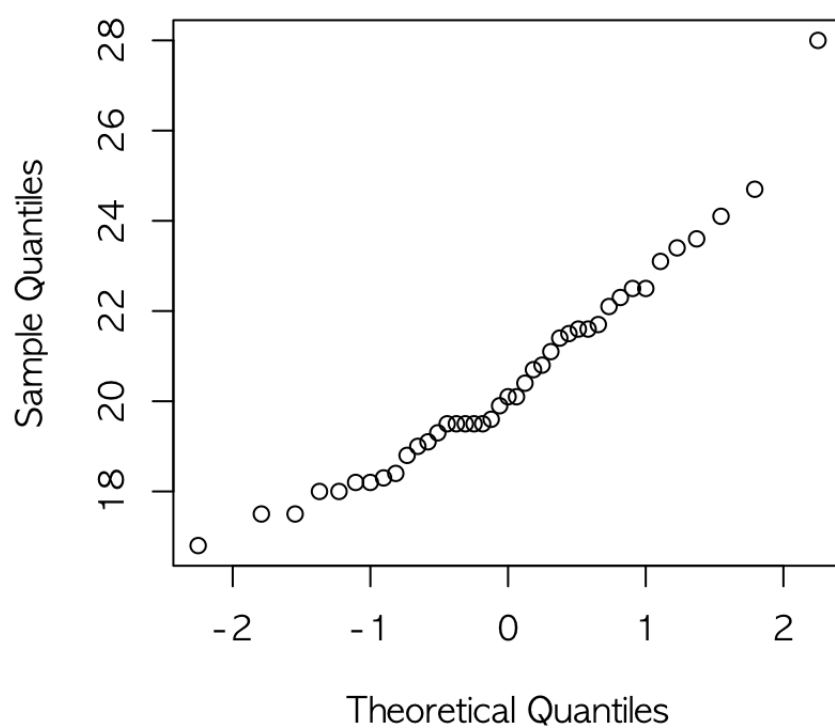
```
qqnorm(height)
```

Normal Q-Q Plot



```
qqnorm(BMI)
```

Normal Q-Q Plot



- weight 와 height 의 줄기-잎 그림

```
stem(weight)
```

```
##
## The decimal point is at the |
##
## 42 | 00
## 44 | 000
## 46 | 000
## 48 | 000000
## 50 | 00000
## 52 | 000000000
## 54 | 000
## 56 | 000
## 58 | 0
## 60 | 000
## 62 | 00
## 64 | 0
```

```
stem(height)
```

```
##
## The decimal point is at the |
##
## 150 | 0
## 152 | 0
## 154 | 0000000
## 156 | 00000
## 158 | 0000000
## 160 | 000000000
## 162 | 00000
## 164 | 000
## 166 | 0
## 168 | 0
## 170 | 0
```

R base package 에서 제공하는 정규성 테스트

- Shapiro-Wilk test, `shapiro.test()`

```
shapiro.test(weight)
```

```
##
## Shapiro-Wilk normality test
##
## data: weight
## W = 0.9694, p-value = 0.3304
```

```
shapiro.test(height)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  height  
## W = 0.9836, p-value = 0.8096
```

```
shapiro.test(BMI)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  BMI  
## W = 0.9466, p-value = 0.05333
```

- Kolmogorov-Smirnov test, `ks.test()`

```
ks.test(weight, "pnorm", mean=mean(weight), sd=sd(weight))
```

```
## Warning in ks.test(weight, "pnorm", mean = mean(weight), sd = sd(weight)):  
## ties should not be present for the Kolmogorov-Smirnov test
```

```
##  
##  One-sample Kolmogorov-Smirnov test  
##  
## data:  weight  
## D = 0.1359, p-value = 0.4355  
## alternative hypothesis: two-sided
```

```
ks.test(height, "pnorm", mean=mean(height), sd=sd(height))
```

```
## Warning in ks.test(height, "pnorm", mean = mean(height), sd = sd(height)):  
## ties should not be present for the Kolmogorov-Smirnov test
```

```
##  
##  One-sample Kolmogorov-Smirnov test  
##  
## data:  height  
## D = 0.0904, p-value = 0.8906  
## alternative hypothesis: two-sided
```

```
ks.test(BMI, "pnorm", mean=mean(BMI), sd=sd(BMI))
```



```
## Warning in ks.test(BMI, "pnorm", mean = mean(BMI), sd = sd(BMI)): ties  
## should not be present for the Kolmogorov-Smirnov test
```

```
##  
## One-sample Kolmogorov-Smirnov test  
##  
## data: BMI  
## D = 0.1195, p-value = 0.6018  
## alternative hypothesis: two-sided
```