

1. (1)

```

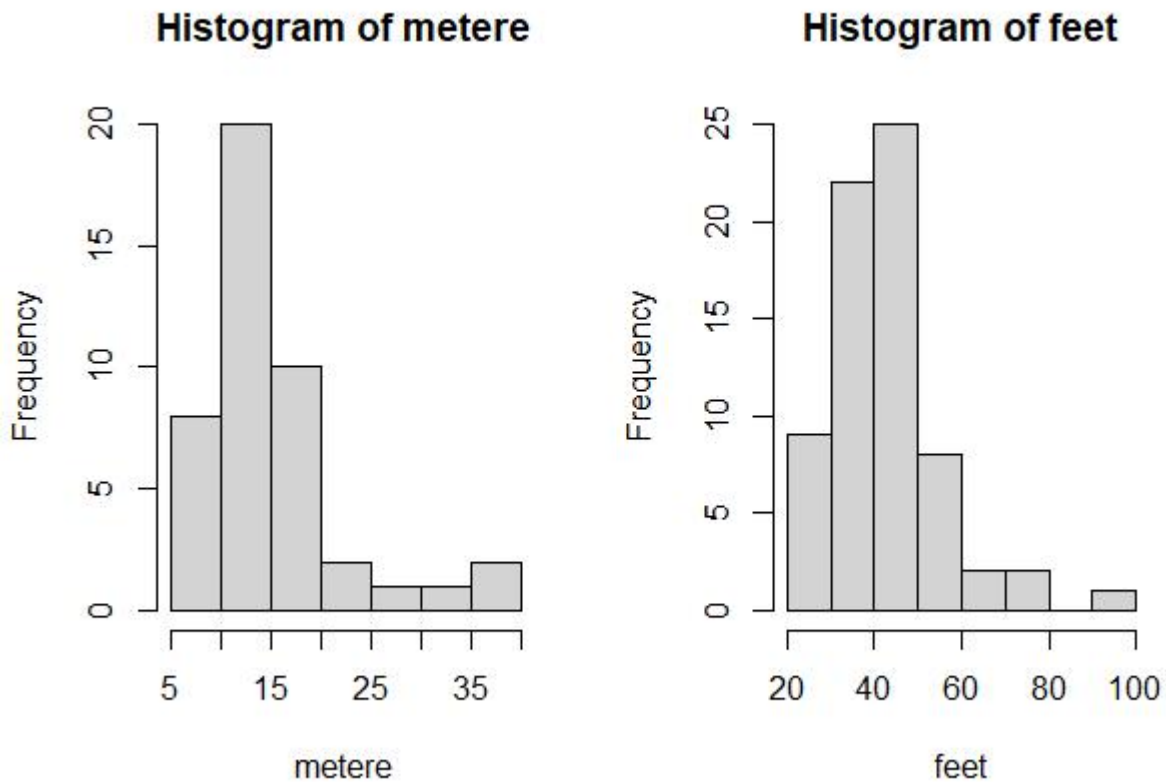
metere <- c(8, 9, 10, 10, 10, 10, 10, 10, 10, 11, 11, 11, 11, 12, 12, 13, 13, 13,
            14, 14, 14, 15, 15, 15, 15, 15, 15, 15, 15, 16, 16, 16, 17, 17, 17, 17,
            18, 18, 20, 22, 25, 27, 35, 38, 40)
feet <- c(24, 25, 27, 30, 30, 30, 30, 30, 30, 32, 32, 33, 34, 34, 34, 35, 35, 36, 36, 36, 37, 37,
          40, 40, 40, 40, 40, 40, 40, 40, 40, 41, 41, 42, 42, 42, 42, 43, 43, 44, 44, 44, 45, 45, 45,
          45, 45, 45, 46, 46, 47, 48, 48, 50, 50, 50, 51, 54, 54, 54, 55, 55, 60, 60, 63, 70, 75, 80,
          94)

```

```

par(mfrow=c(1,2))
hist(metere)
hist(feet)

```

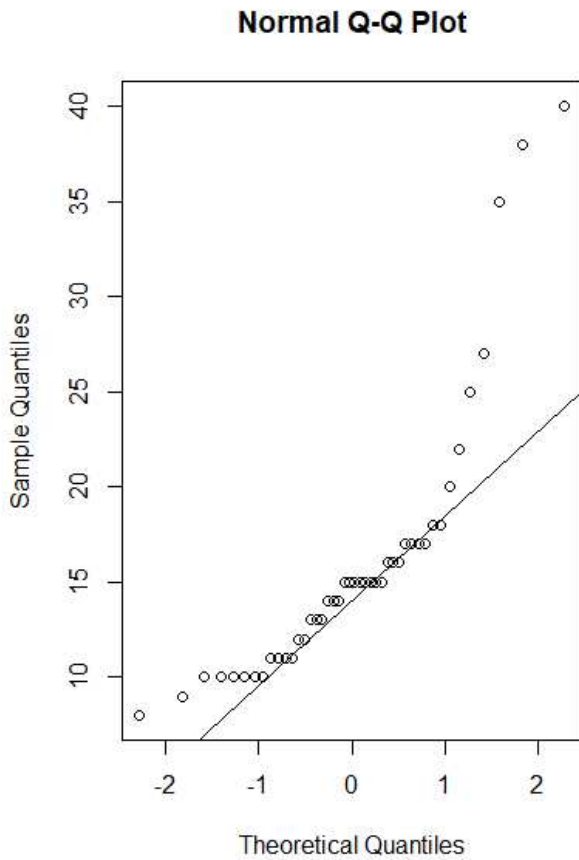


metere와 feet 모두 right skewness 경향이 강하기 때문에 정규분포가 아닐 것으로 생각된다. 정규분포에 대응하는 qqplot을 그리기 위해 qqnorm 함수로 분석하면 다음과 같다.

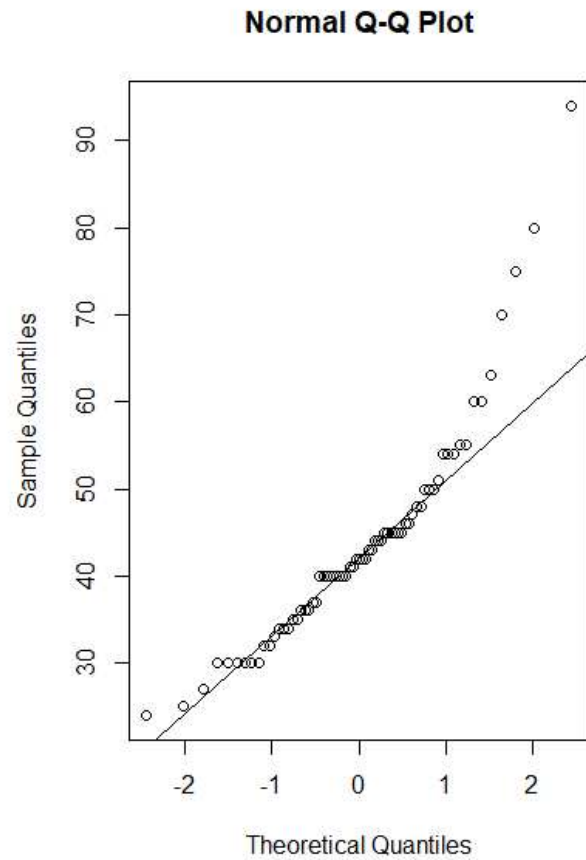
```

qqnorm(metere)
qqline(metere)
qqnorm(feet)
qqline(feet)

```



metere



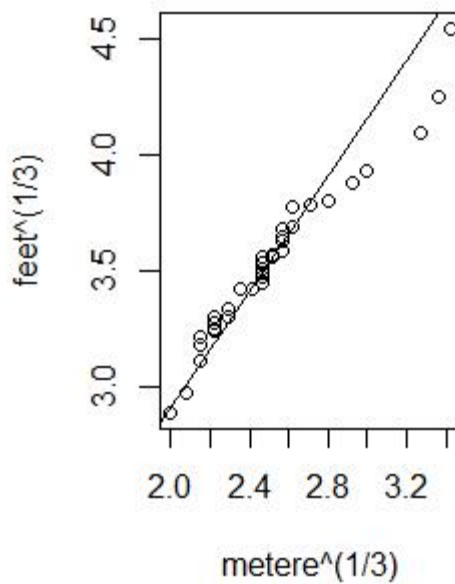
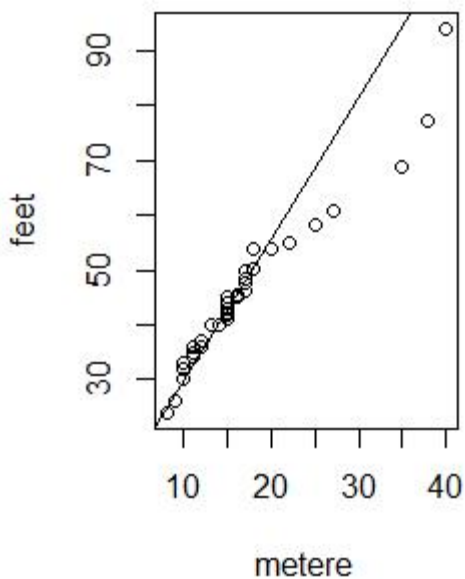
feet

metere와 feet 모두 곡선 경향이 나타나므로 정규분포로 가정하기 어렵다고 생각된다.

(2)

```
qqplot(metere, feet)
abline(line(qqplot(metere, feet)))
```

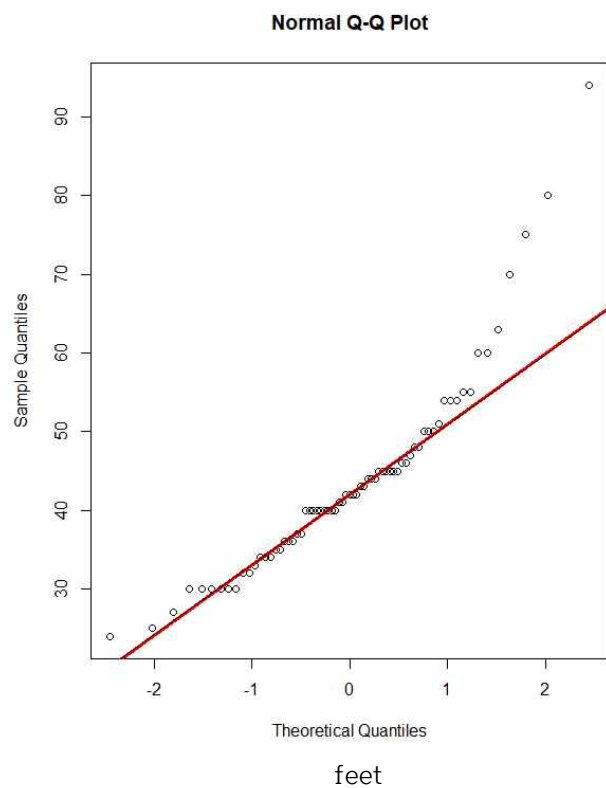
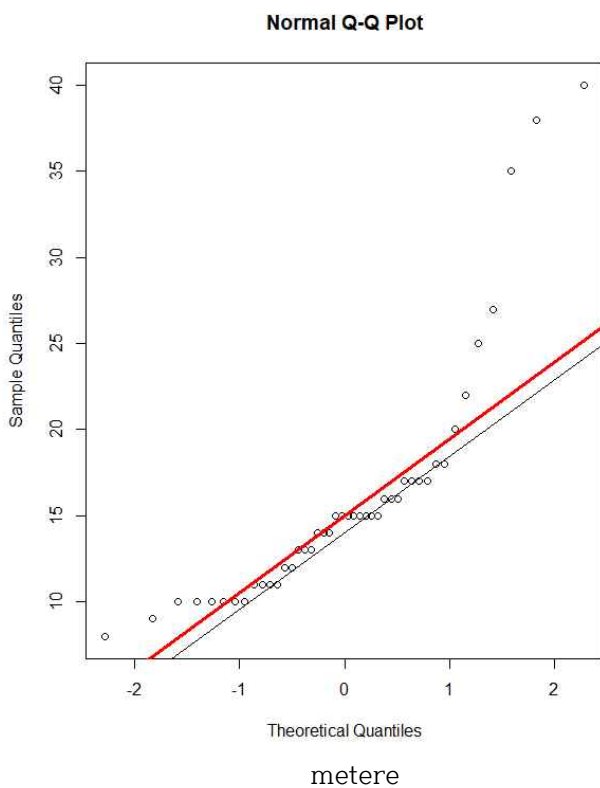
```
qqplot(metere^(1/3), feet^(1/3))
abline(line(qqplot(metere^(1/3), feet^(1/3))))
```



feet과 metere를 각각 대응시켜 그린 QQplot이 대부분 직선 경향으로 분포한다. 자료가 하단에 밀집해 분포하는 경향이 있으므로 power=1/3을 적용한 QQplot도 함께 그려 확인한다. 자료의 대부분이 직선 위에 분포하고 있으므로 두 분포는 동일하다고 할 수 있다.

```
qqnorm(metere)
qqline(metere)
x = fivenum(metere)
pseudosigma = (x[4]-x[2])/1.34
abline(x[3],pseudosigma, col="red", lwd=3)
```

```
qqnorm(feet)
qqline(feet)
y = fivenum(feet)
pseudosigmaf = (y[4]-y[2])/1.34
abline(y[3],pseudosigmaf, col="red", lwd=3)
```



metere, feet으로 그린 Normal Q-Qplot은 다음과 같다. Q-Qplot에서 데이터와 잘 대응되는 직선을 qqline()과 abline(중앙값, pseudosigma)로 그려 각각 검은색, 빨간색으로 표시한다. metere의 경우 빨간 직선이 더 데이터를 잘 나타내고, feet의 경우 둘이 동일하므로 Q-Qplot에서 구하는 평균과 분산을 각각 중앙값, pseudosigma로 구한다.

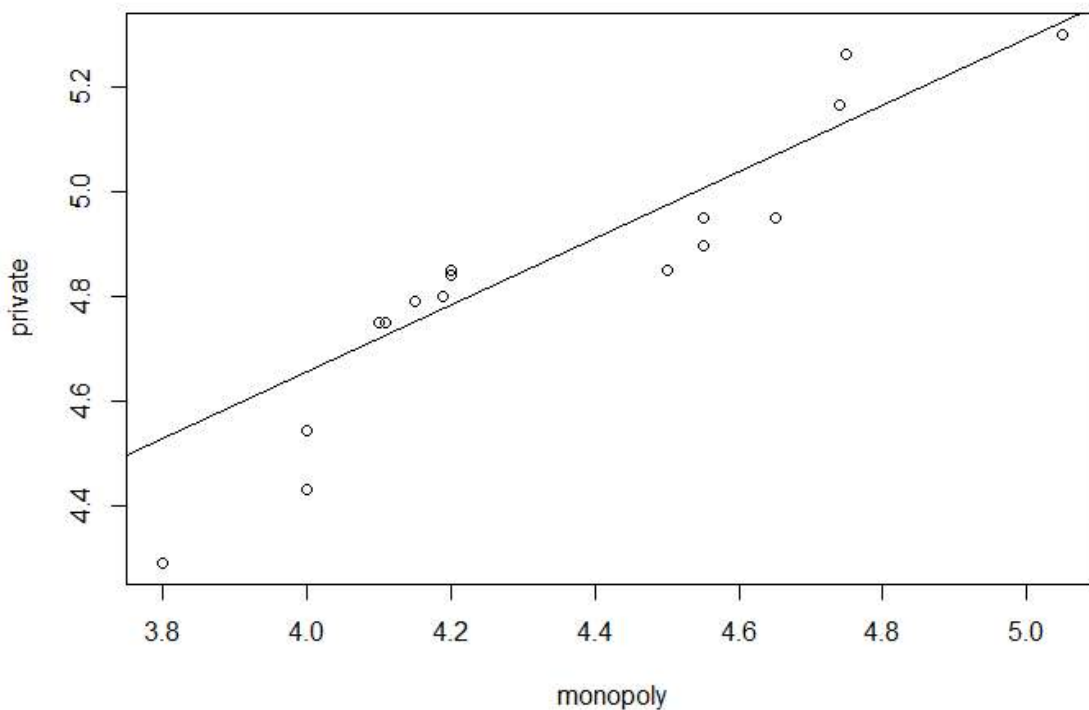
```
matrix(c(mean(metere),var(metere),x[3],pseudosigma,mean(feet),var(feet),y[3],pseudosigmaf),ncol=2
      ,dimnames = list(c("mean from sample estimates","var from sample estimates","mean
from QQplot","var from QQplot"),c("metere","feet")))
#               metere      feet
mean from sample estimates 16.022727  43.695652
var from sample estimates  51.045983 156.185422
mean from QQplot          15.000000  42.000000
var from QQplot           4.477612   8.955224
```

따라서 sample estimates와 QQplot에서 구한 mean과 variance는 위와 같다,

2.

```
monopoly = c(4.65, 4.55, 4.11, 4.15, 4.20, 4.55, 3.80, 4.00, 4.19, 4.75,
             4.74, 4.50, 4.10, 4.00, 5.05, 4.20)
private = c(4.82, 5.29, 4.89, 4.95, 4.55, 4.90, 5.25, 5.30, 4.29, 4.85,
            4.54, 4.75, 4.85, 4.85, 4.50, 4.75, 4.79, 4.85, 4.79, 4.95,
            4.95, 4.75, 5.20, 5.10, 4.80, 4.29)
```

```
abline(line(qqplot(monopoly, private)))
```

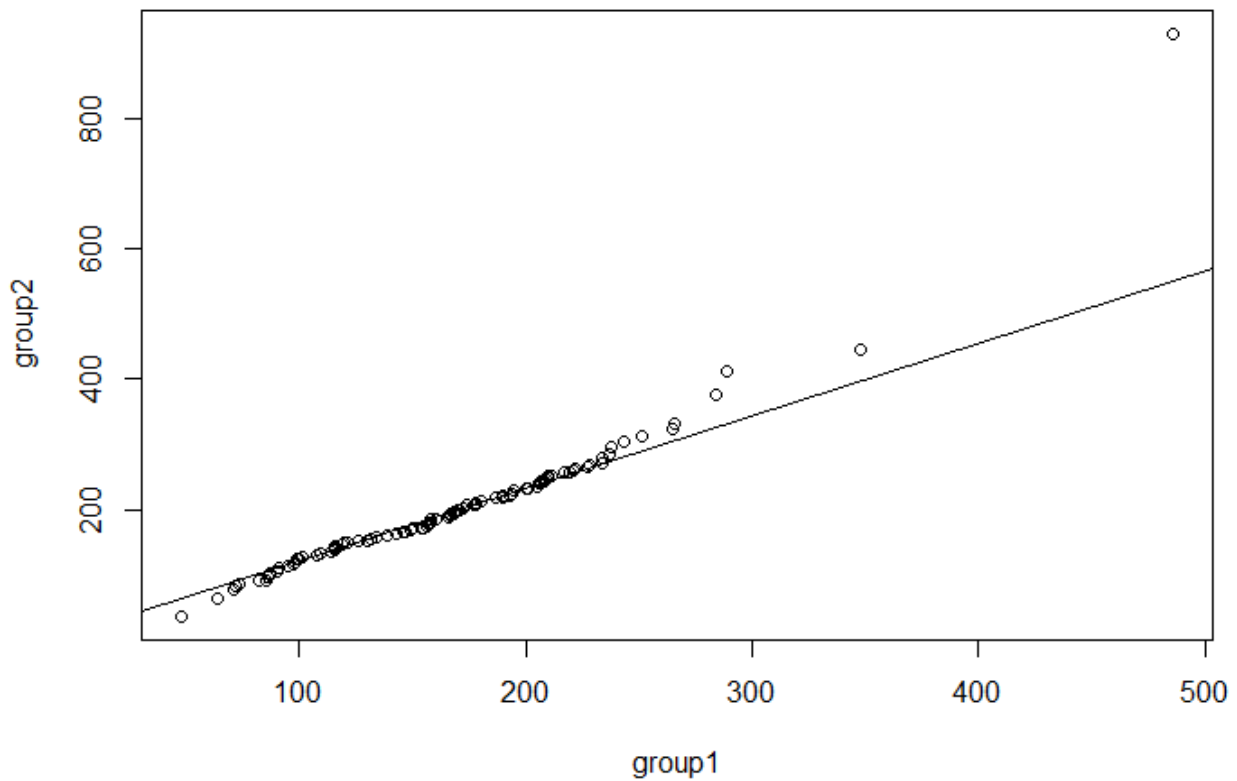


qqplot과 그에 대응하는 직선을 그려보면 점들의 위치가 전반적으로 직선을 이루고 있다는 것을 알 수 있다. 따라서 두 그룹의 분포는 같은 분포라고 해석할 수 있다.

3.

```
group1 <- BLOODFAT.DAT의 1그룹 부분  
group2 <- BLOODFAT.DAT의 2그룹 부분
```

```
qqplot(group1,group2)  
abline(line(qqplot(group1, group2)))
```



group1과 group2간의 QQplot이 뚜렷한 직선의 경향을 보이고 있으므로 둘의 분포는 동일하다. 따라서 그룹1과 그룹2의 콜레스테롤 분포는 동일하다.