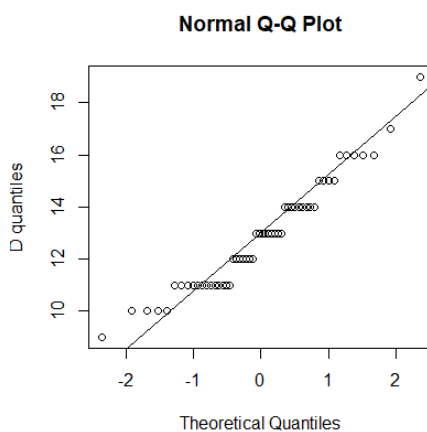


1.

```
D <- c( 14, 11, 13, 13, 13, 15, 11, 16, 10,
       13, 14, 11, 13, 12, 10, 14, 10, 14,
       16, 14, 14, 11, 11, 11, 13, 12, 13,
       11, 11, 15, 14, 16, 12, 17, 9, 16,
       11, 19, 14, 12, 12, 10, 11, 12, 13,
       13, 14, 11, 11, 15, 12, 16, 15, 11)
```

```
x <- fivenum(D)
(pseudosigma <- (x[4]-x[2])/1.34)
#[1] 2.238806
sd(D)
#[1] 2.092631
qqnorm(D, ylab="D quantiles")
abline(x[3],pseudosigma)
```



**정규분포임을 판단:** pseudosigma와 실제 표준편차의 값이 비슷하므로 정규분포에 가깝다고 볼 수 있다.

또한 Normal Q-Q plot에서 자료의 분포가 직선과 잘 대응되므로 정규분포에 가깝다고 볼 수 있다.

**직선의 기울기와 절편:** 절편은 x[3]에 해당하는 13, 기울기는 pseudosigm에 해당하는 2.238806이다.

**직접 계산한 평균,분산 추정치와 확률도에서 계산한 평균,분산 추정치:**

```
matrix(c(mean(D),sd(D),x[3],pseudosigma), nrow=2,
       dimnames = list(c("평균","분산"),c("직접계산","확률도")))
#      직접계산   확률도
평균 12.870370 13.000000
분산  2.092631  2.238806
```

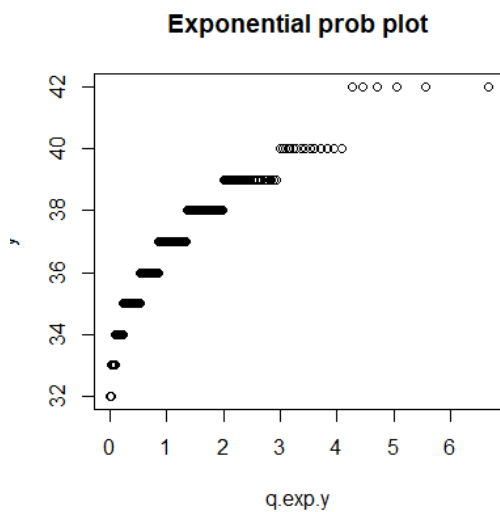
2.

```
x=c(32,33,34,35,36,37,38,39,40,42)
frequency=c(10,33,81,161,224,289,336,369,383,389)
y <- c()
```

```

fr=c()
for(i in 1:length(x)){
  fr[i] = frequency[i]-frequency[i-1]
  fr[1] = 10
}
for(i in 1:length(x)){
  y1 <- rep(x[i],fr[i])
  y <- c(y1,y)
}
q.exp.y <- -log(1-(i-0.5)/n)
plot(q.exp.y, y, main="Exponential prob plot")

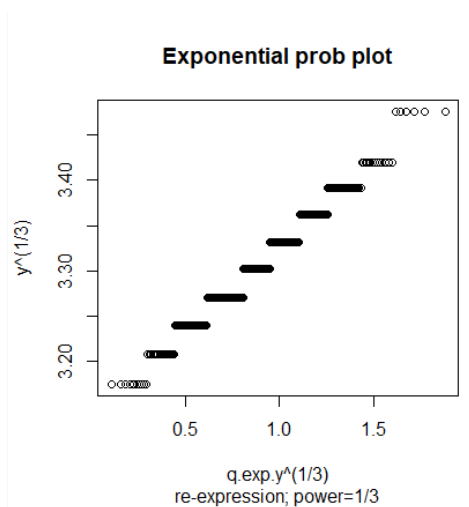
```



```

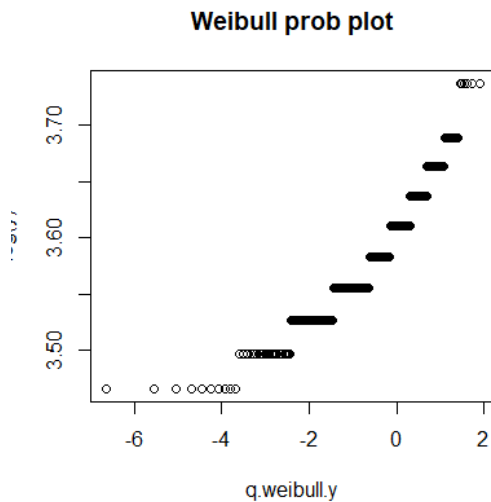
plot(q.exp.y^(1/3), y^(1/3),main="Exponential prob plot",
     sub="re-expression; power=1/3")

```



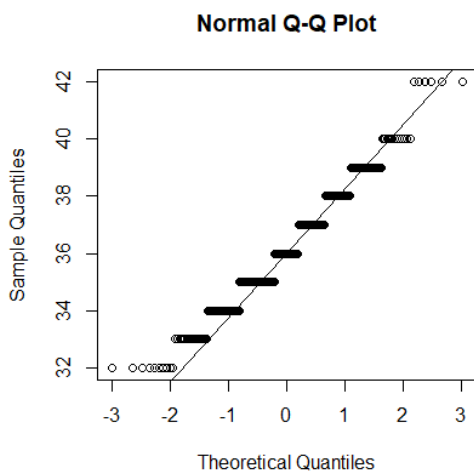
**지수분포 판단:** Exponential probability plot을 그려봤으나 분포는 곡선의 형태여서 직선에 잘 대응되지 않는다. 다만 1/3의 power를 적용한 경우 직선에 잘 대응된다는 것을 확인할 수 있다. 따라서 1/3의 power를 적용한 지수분포로 볼 수 있다고 생각된다.

```
q.weibull.y <- log(q.exp.y)
plot(q.weibull.y, log(y), main="Weibull prob plot")
```



**와이블 분포 판단:** Weibull probability plot을 그려봤으나 분포는 곡선의 형태여서 직선에 잘 대응되지 않기 때문에 와이블분포를 이룬다고 보기 어렵다.

```
qqnorm(y)
x <- fivenum(y)
pseudosigma <- (x[4]-x[2])/1.34
abline(x[3],pseudosigma)
```

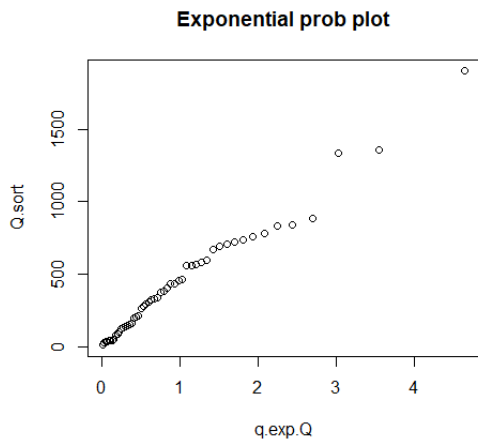


```
> pseudosigma
[1] 2.238806
> sd(y)
[1] 2.042171
```

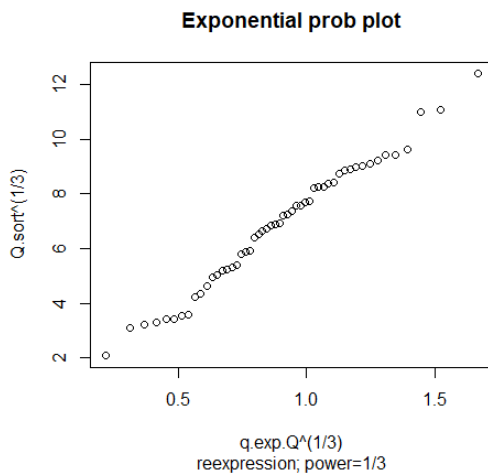
**정규 분포 판단:** Normal Q-Q plot을 그려보니 분포가 직선에 잘 대응된다. 실제로 절편을 중앙값, 기울기를 pseudosigma로 설정한 직선을 그려보면 대부분의 데이터가 직선위에 분포한다는 것을 확인할 수 있다. 또한 앞서 구한 pseudosigma와 실제 sd가 비슷한 값이므로 정규분포를 가정하는것이 적합하다고 생각된다. 이를 가정하고 추정한 parameter는  $\mu=36(\text{median})$ ,  $\sigma=2.238806(\text{pseudosigma})$ 이다. 실제 자료에서 구한 parameter는  $\mu=36.1671$ ,  $\sigma=2.042171$ 이므로, 근사하다는 것이 확인된다.

3.

```
Q <- QUAKES.DAT
Q.sort <- sort(Q)
n.Q <- length(Q)
i <- 1:n.Q
q.exp.Q <- -log(1-(i-0.5)/n.Q)
plot(q.exp.Q, Q.sort, main="Exponential prob plot")
```

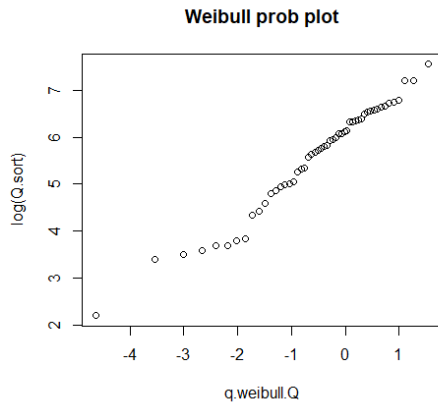


```
plot(q.exp.Q^(1/3), Q.sort^(1/3), main="Exponential prob plot",
     sub="reexpression: power=1/3")
```



**지수분포 판단:** Exponential probability plot을 그려보면 좌측 아래에 데이터가 직선 형태로 모여있는 것으로 보인다. 그러나 power=1/3으로 재표현했을때 다소 concave한 곡선성이 나타나는것을 확인할 수 있다.

```
q.weibull.Q <- log(q.exp.Q)
plot(q.weibull.Q, log(Q.sort), main="Weibull prob plot")
```



**와이블분포 판단:** 데이터가 밀집해있는 우측 상단을 중심으로 확인해보면 다소 concave한 곡선성이 나타나는것을 확인할 수 있다.

지수분포, 와이블분포를 가정했을때 모두 곡선성이 나타나므로 감마분포의 parameter를 찾아 감마분포의 적합성을 판단해 본다.

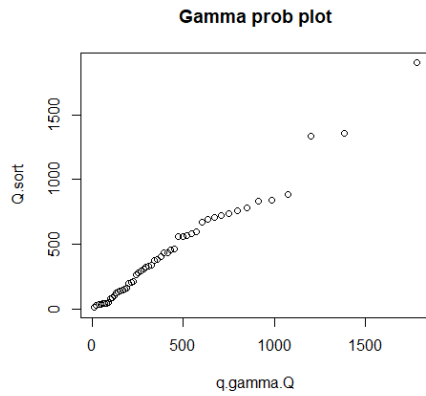
```
(mean.Q <- mean(Q))
# 425.7115
(var.Q <- var(Q))
# 146769
(shape.Q <- mean.Q^2/var.Q)
# 1.2348
(scale.Q <- mean.Q/var.Q)
# 0.002900555
```

표본에서 직접 구한 감마분포의 parameter는  $\alpha = 1.2348$ ,  $\beta = 0.002900555$  이다.

```
library(MASS)
fitdistr(Q, "gamma")
#      shape      rate
# 1.1407136376 0.0026795460
# (0.1840401630) (0.0004997581)
```

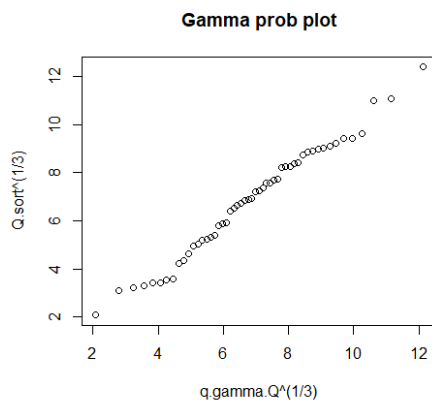
fitdistr 함수로 추정한 감마분포의 parameter는  $\alpha = 1.1407136376$  ,  $\beta = 0.0026795460$  이다.

```
q.gamma.Q <- qgamma((i-0.5)/n.Q, shape.Q, scale.Q)
plot(q.gamma.Q, Q.sort, main="Gamma prob plot")
```



직선성을 더 자세히 확인하기 위해  $power=1/3$ 을 적용하면 다음과 같다.

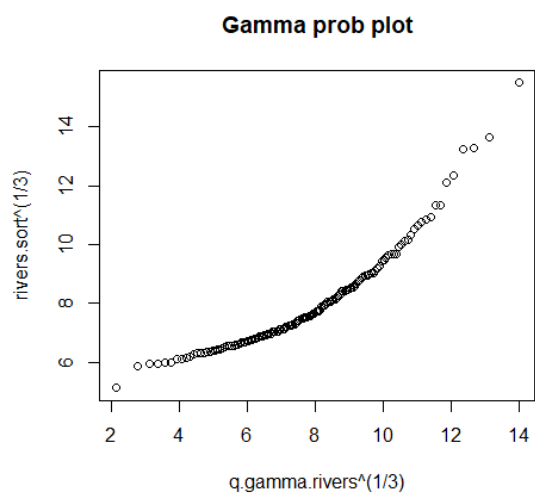
```
plot(q.gamma.Q^(1/3), Q.sort^(1/3), main="Gamma prob plot")
```



마찬가지로 다소 concave한 경향이 나타나는것을 확인할 수 있다.

$power$ 이 0에 가까운 효과를 가지도록  $\log$ 를 취해 확인하면 다음과 같다.

```
plot(log(q.gamma.Q), log(Q.sort), main="Gamma prob plot")
```



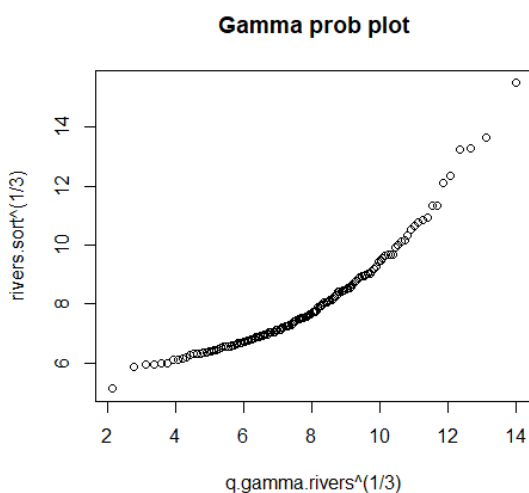
자료가 밀집된 부분에서 concave한 경향이 계속 나타나고 있음을 확인할 수 있다.

지수분포, 와이블분포, 패러미터를 추정한 감마분포를 가정한 확률도에서 모두 곡선의 경향을 보인다. 직선에 잘 대응되는 분포가 드러나야 표본의 분포에 적합하다고 판단할 수 있다. 따라서 거의 동일한 수준으로 가장 직선에 가까운 분포인 지수분포와 감마분포가 적합하다고 판단된다.

[추가속제]

수업에서 다룬 river데이터의 감마분포 확률도를  $\text{power}=1/3$ 을 취해 분석하면 다음과 같다.

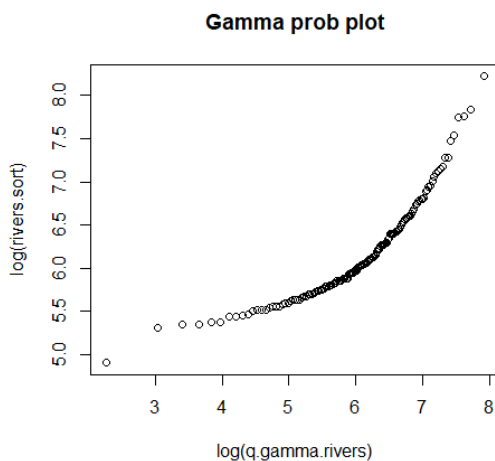
```
plot(q.gamma.rivers^(1/3), rivers.sort^(1/3), main="Gamma prob plot")
```



$\text{power}=1/3$ 을 적용한 상태에서는 뚜렷하게 convex한 경향이 나타나고 있음이 확인된다. 따라서 자료가 직선에 잘 맞지 않으므로 Gamma 분포가 적절하다고 판단하기 힘들다.

또한  $\text{power}$ 가 0에 가까운 효과를 주도록 log를 취하면 다음과 같다.

```
plot(log(q.gamma.rivers), log(rivers.sort), main="Gamma prob plot")
```



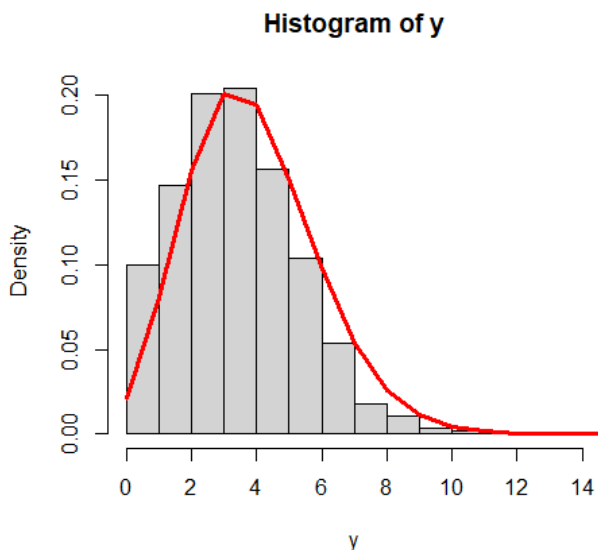
역시 concave한 경향이 뚜렷히 나타나므로 Gamma분포는 적절하지 않음을 알 수 있다.

4.

```
x=c(seq(0,14))
frequency=c(57,203,383,525,532,408,272,139,46,27,10,4,0,1,1)
y <- c()
```

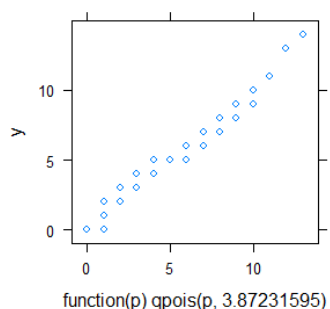
```
for(i in 1:length(x)){
  y1 <- rep(x[i],frequency[i])
  y <- c(y1,y)
}
hist(y,probability = T)
```

```
library(lattice)
fitdistr(y,"poisson")
#lambda=3.87231595
x1<-c(0:40)
yp=dpois(x1,lambda=3.87231595)
lines(x1,yp,type='l',col="red")
```



fitdistr 함수로 추정한 poisson분포의 lambda는 3.87231595이다. 이를 적용해 poisson분포의 확률밀도함수를 붉은색 선으로 자료의 히스토그램 위에 그리면 위와 같다. 자료는 해당 포아송 분포와 비슷하다는 것을 확인할 수 있다.

```
qqmath(y, distribution = function(p) qpois(p,3.87231595 ))
```



자료에 대한 lambda=3.87231595인 포아송분포로부터 유도한 확률 플롯은 이와 같다. 직선의 경향이 나타남을 확인할 수 있다. 따라서 위의 히스토그램과 종합해 봤을 때 자료는 lambda=3.87231595인 포아송분포에 적합하다고 판단할 수 있다.