

Комп'ютерна статистика

Шкляр Ірина, ксад 4 курс

Робота 2, варіант 4

## 1. Читаємо дані

```
data <- read.csv("~/Downloads/compsta3/distr4.txt", header = TRUE)
```

### Знайдемо основні показники вибірки:

```
> summary(data)
```

X

Min.: 0.01005

1st Qu.: 0.26722

Median: 0.69418

Mean: 0.95424

3rd Qu.: 1.30024

Max.: 4.25546

```
> sd(data$X)      [1] 0.8817646
```

```
> mean(data$X)    [1] 0.9542384
```

```
> min(data)       [1] 0.01004783
```

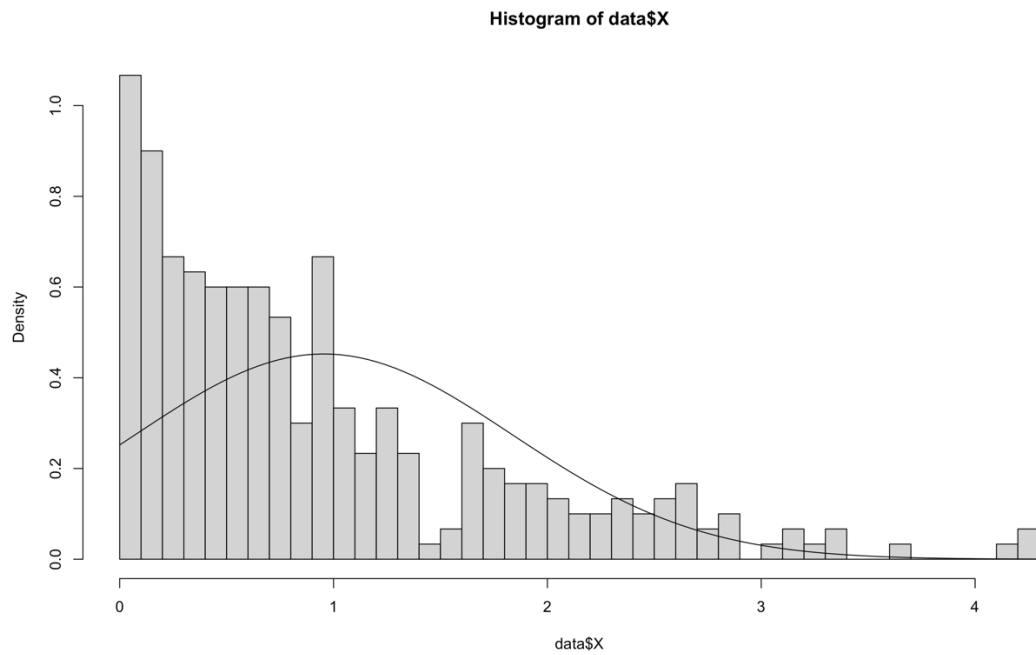
```
> max(data)       [1] 4.255455
```

## 2. Гістограма даних зі щільністю розподілу

а) *нормальний розподіл*

```
hist(data$X, probability = TRUE, right = FALSE, breaks = 50)
```

```
curve(dnorm(x, mean = 0.9542384, sd = 0.8817646), add = TRUE)  --   графік нормальної щільності
```



## PP-діаграма

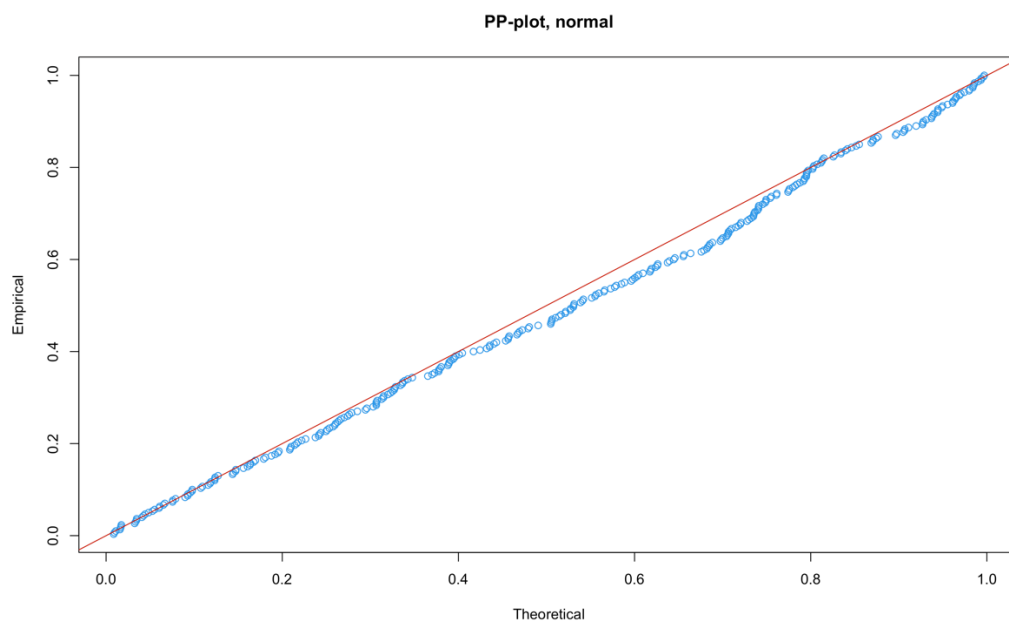
```
set.seed(12345)
```

```
n <- 300
```

```
x <- rnorm(n, mean = 0.9542384, sd = 0.8817646)
```

```
plot(pnorm(sort(x), mean = 0.9542384, sd = 0.8817646), (1:length(x))/length(x), xlab = "Theoretical", ylab = "Empirical", main = "PP-plot", col = 4)
```

```
abline(0,1, col = "red3")
```



## QQ-діаграма

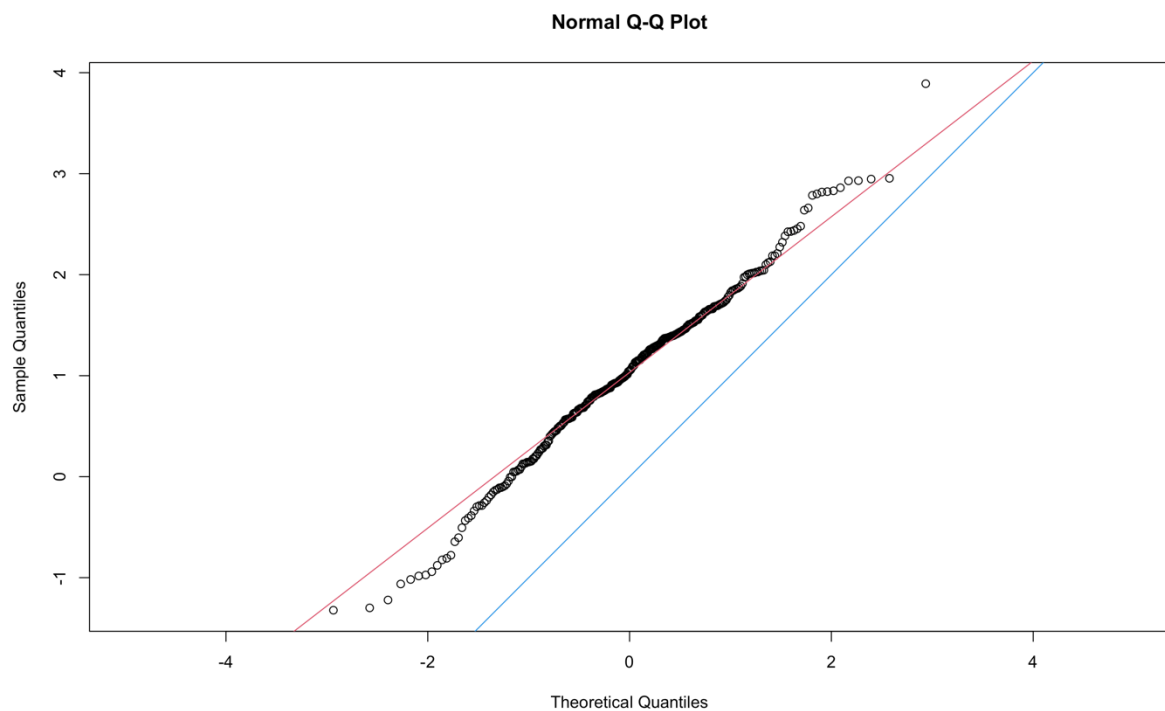
```
n <- 300
```

```
x <- rnorm(n, mean = 0.9542384, sd = 0.8817646)
```

`qqnorm(x, asp = 1)` # qq-діаграма для норм. розподілу, використовує квантилі стандартного нормального по осі x, тому лінія, вздовж якої розташовані точки діаграми не є прямою  $y=x$ .

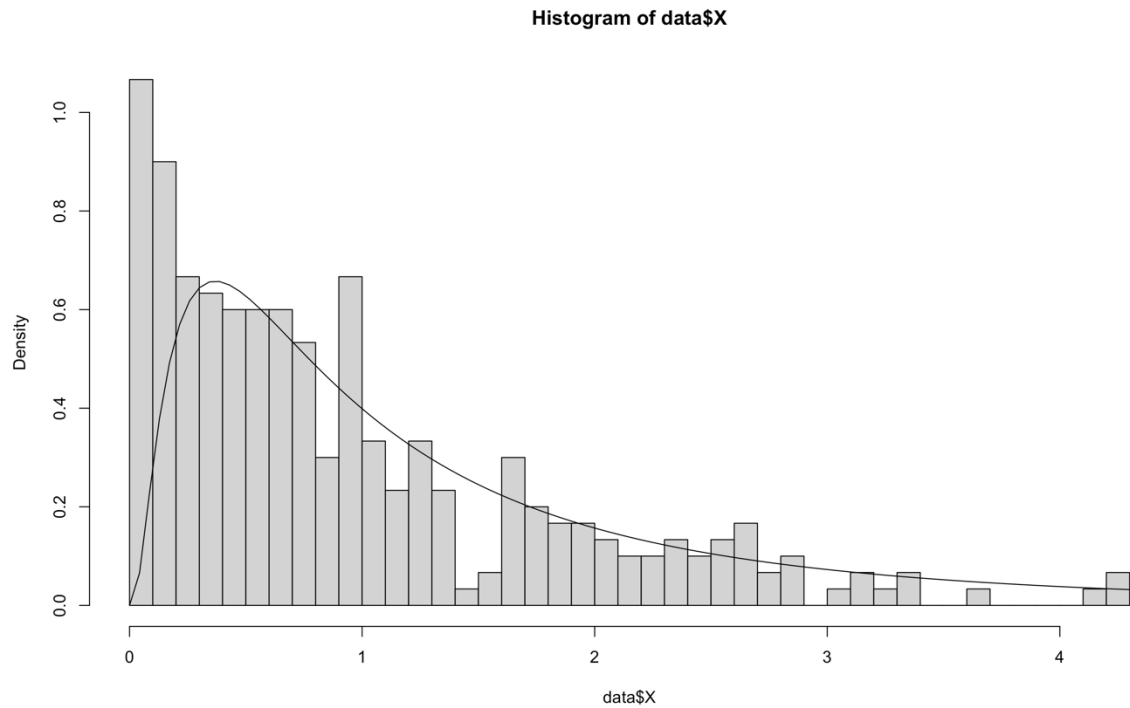
`qqline(x, col = 2)` # оцінює за даними вибірки мат. спод. b та станд. відх-ня a і малює пряму  $y = ax + b$ .

```
abline(0,1, col = 4)
```

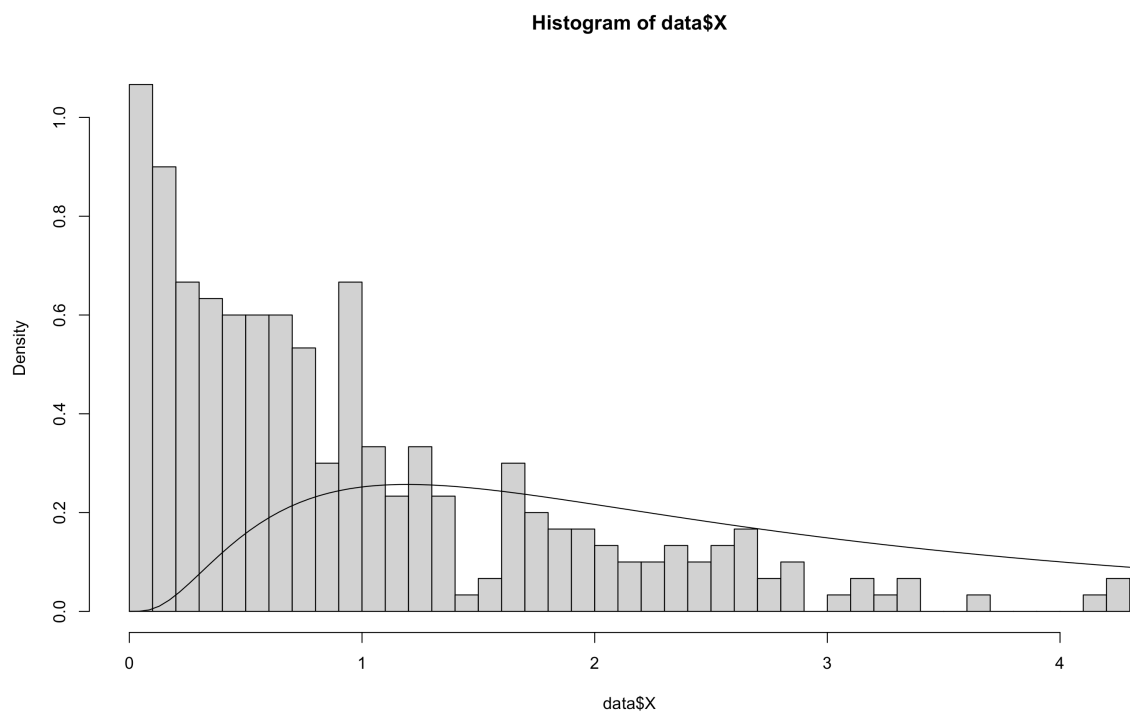


## б) логнормальний розподіл

```
hist(data$X, probability = TRUE, right = FALSE, breaks = 50)
curve(dlnorm(x, meanlog = 0, sdlog = 1), add = TRUE)
```

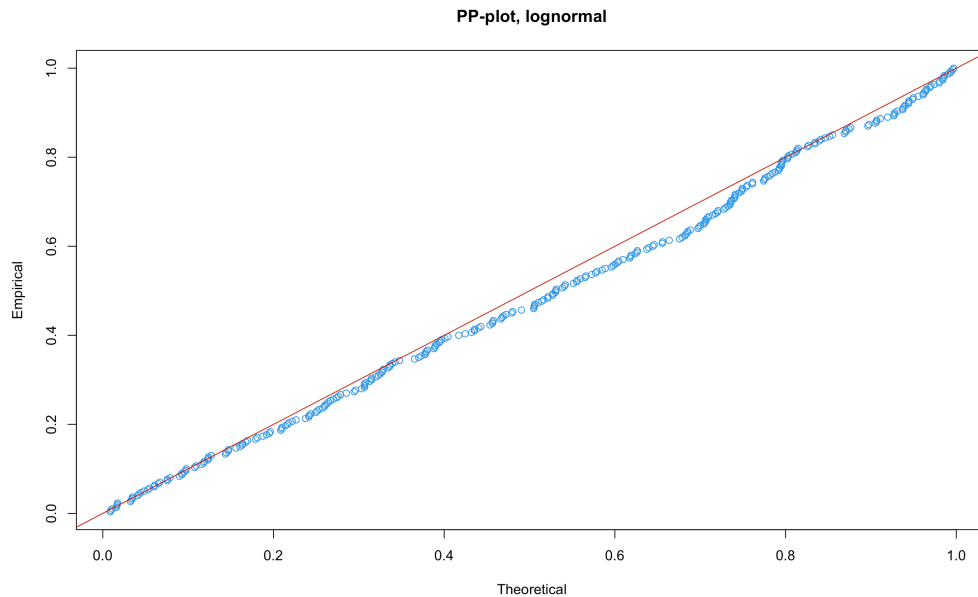


```
hist(data$X, probability = TRUE, right = FALSE, breaks = 50)
curve(dlnorm(x, meanlog = 0.9542384, sdlog = 0.8817646), add = TRUE)
```



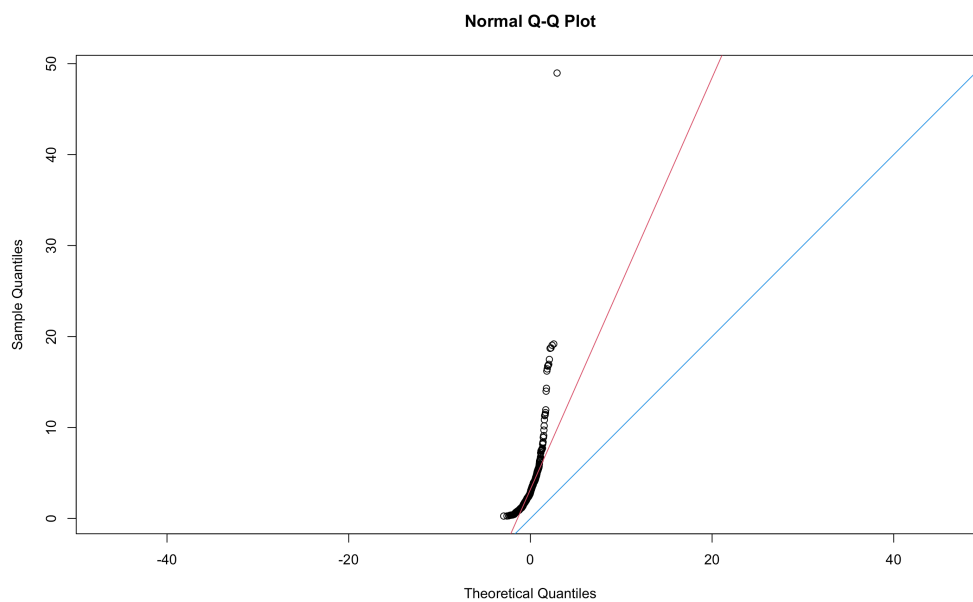
## PP-діаграма

```
set.seed(12345)
n <- 300
x0 <- rlnorm(n, meanlog = 0.9542384, sdlog = 0.8817646)
plot(plnorm(sort(x0), meanlog = 0.9542384, sdlog = 0.8817646), (1:length(x0))/length(x0), xlab = "Theoretical",
ylab = "Empirical", main = "PP-plot, lognormal", col = 4)
abline(0,1, col = "red3")
```



## QQ-діаграма

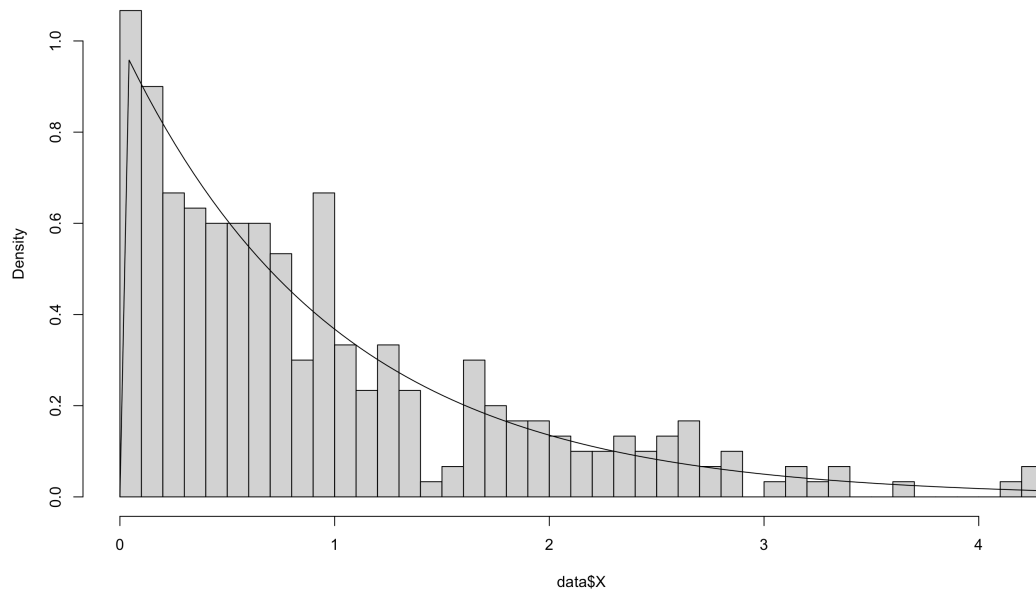
```
n <- 300
x0 <- rlnorm(n, meanlog = 0.9542384, sdlog = 0.8817646)
qqnorm(x0, asp = 1)
qqline(x0, col = 2)
abline(0,1, col = 4)
```



в) експоненційний розподіл

```
hist(data$X, probability = TRUE, right = FALSE, breaks = 50)  
curve(dexp(x, rate=1), add = TRUE)
```

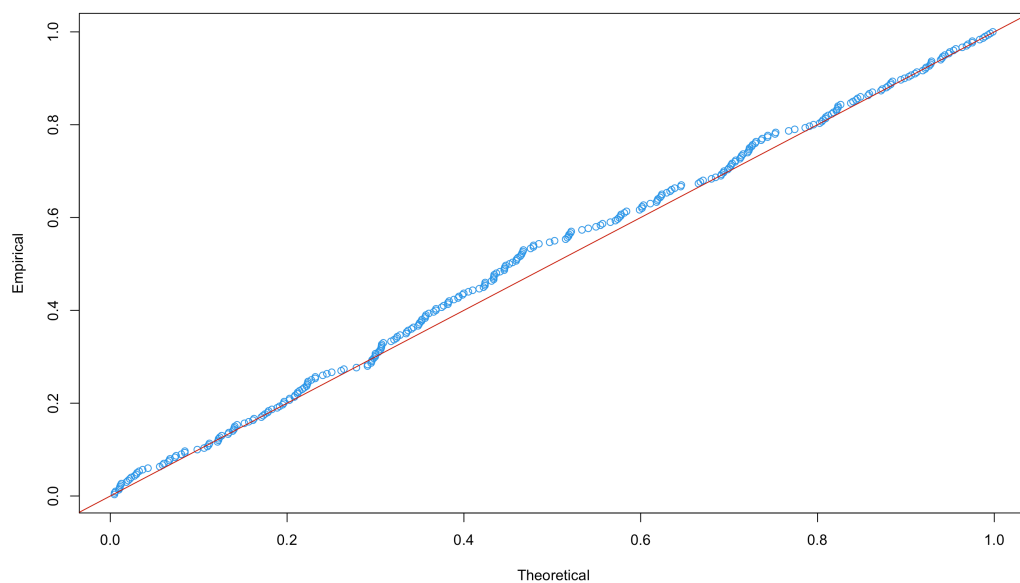
Histogram of data\$X



## PP-діаграма

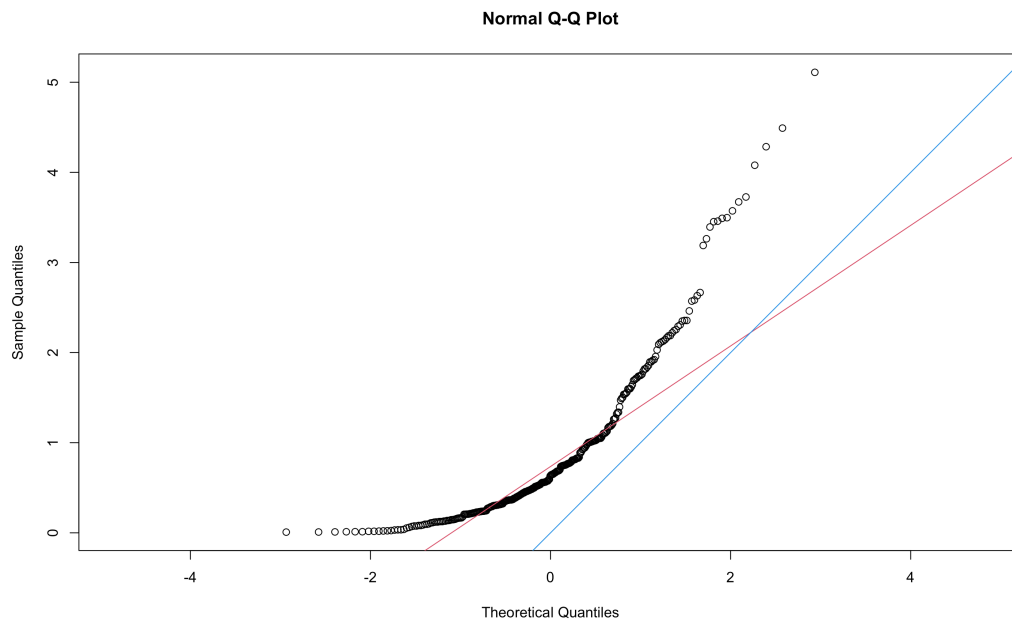
```
set.seed(12345)  
n <- 300  
x1 <- rexp(n, rate=1)  
plot(pexp(sort(x1), rate=1), (1:length(x1))/length(x1), xlab = "Theoretical", ylab = "Empirical", main =  
      "PP-plot, exponential", col = 4)  
abline(0,1, col = "red3")
```

PP-plot, exponential



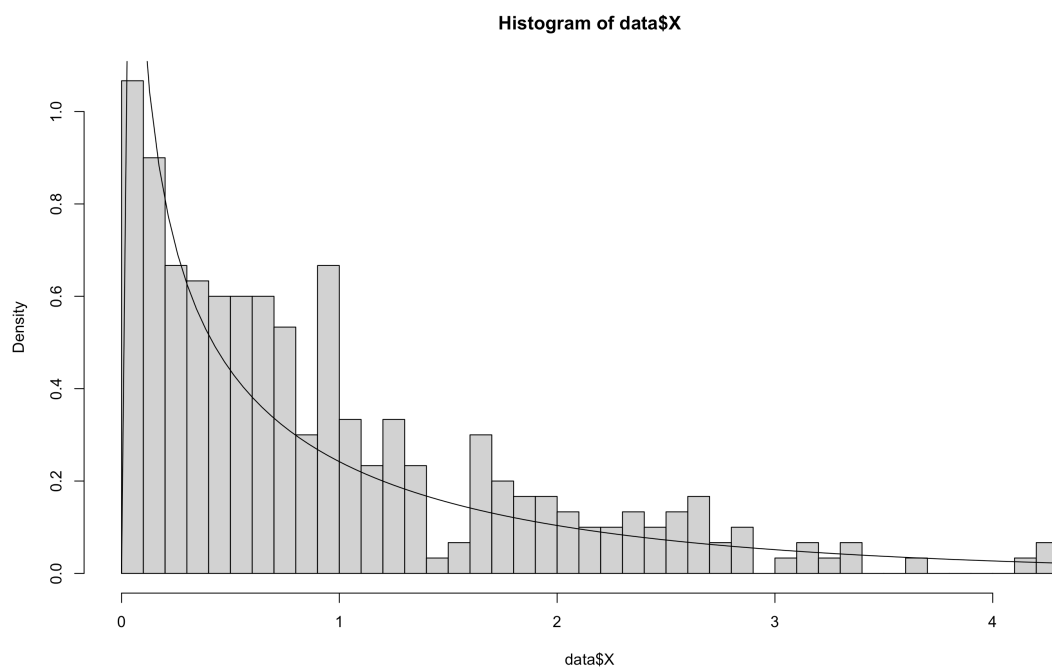
## QQ-діаграма

```
n <- 300  
x1 <- rexp(n, rate=1)  
qqnorm(x1, asp = 1)  
qqline(x1, col = 2)  
abline(0,1, col = 4)
```



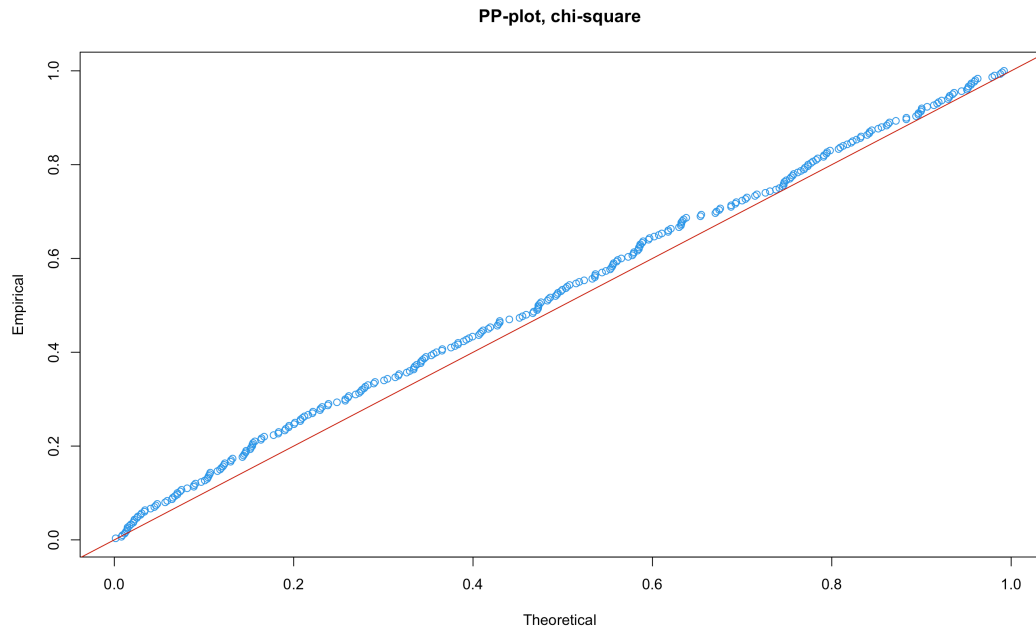
## г) *хі-квадрат розподіл*

```
hist(data$X, probability = TRUE, right = FALSE, breaks = 50)  
curve(dchisq(x, df=1), add = TRUE)
```



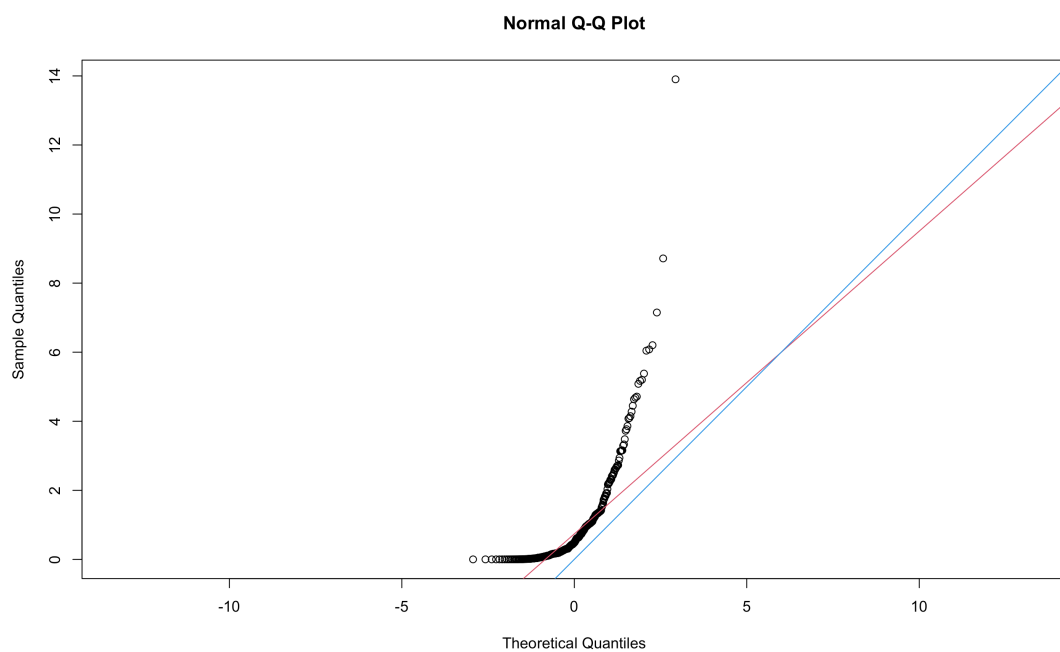
## PP-діаграма

```
set.seed(12345)
n <- 300
x2 <- rchisq(n, df=1)
plot(pchisq(sort(x2), df=1), (1:length(x2))/length(x2), xlab = "Theoretical", ylab = "Empirical", main =
      "PP-plot, chi-square", col = 4)
abline(0,1, col = "red3")
```



## QQ-діаграма

```
n <- 300
x2 <- rchisq(n, df=1)
qqnorm(x2, asp = 1)
qqline(x2, col = 2)
abline(0,1, col = 4)
```

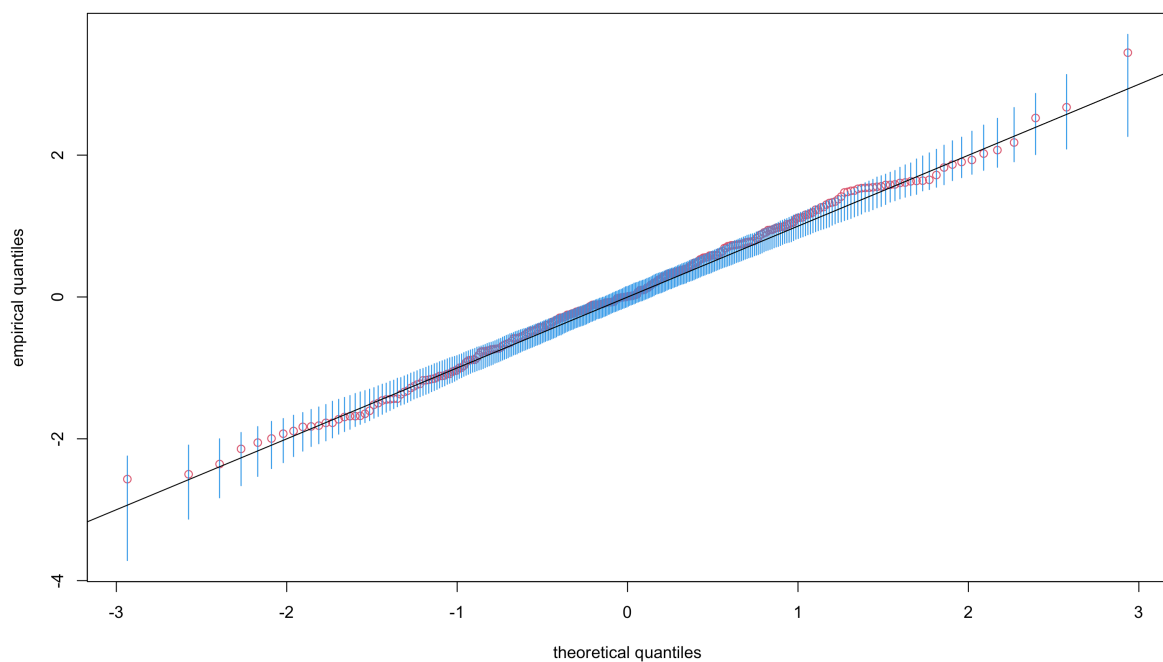




З графіків видно, що найкраще описує розподіл даних *нормальний* розподіл.

### 3. QQ-діаграма з прогнозними інтервалами для нормального розподілу

```
QQplot<-function(x,K=1000,alpha=0.05){  
  n<-length(x)  
  normQ<-qnorm((1:n-0.5)/n)  
  sx<-sort(x)  
  W<-matrix(rnorm(K*n),nrow=n,ncol=K)  
  W<-apply(W,2,sort)  
  tops<-apply(W,1,quantile,probs=1-alpha/2)  
  bots<-apply(W,1,quantile,probs=alpha/2)  
  plot(c(normQ,normQ,normQ),c(tops,bots,sx),type="n", xlab="theoretical quantiles", ylab="empirical  
quantiles")  
  points(normQ,sx,col=2)  
  segments(normQ,bots,normQ,tops,col=4)  
  abline(0,1,col=1)  
}  
x<-rnorm(300)  
QQplot(x)
```



Остаточно, бачимо, що розподіл даних відповідає теоретичному.