G64190069_latihan praktikum 12

Rizal Mujahiddan

5/17/2022

```
library(readxl)
library(EnvStats)
```

```
##
## Attaching package: 'EnvStats'
```

```
## The following objects are masked from 'package:stats':
##
##     predict, predict.lm
```

```
## The following object is masked from 'package:base':
##
##     print.default
```

```
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:EnvStats':
##
##     boxcox
```

```
library(fitdistrplus)
```

```
## Loading required package: survival
```

```
library(ggplot2)
library(psych)
```

```
##
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:psych':
##
##     logit
```

```
## The following object is masked from 'package:EnvStats':
##
##     qqPlot
```

Nomor 1

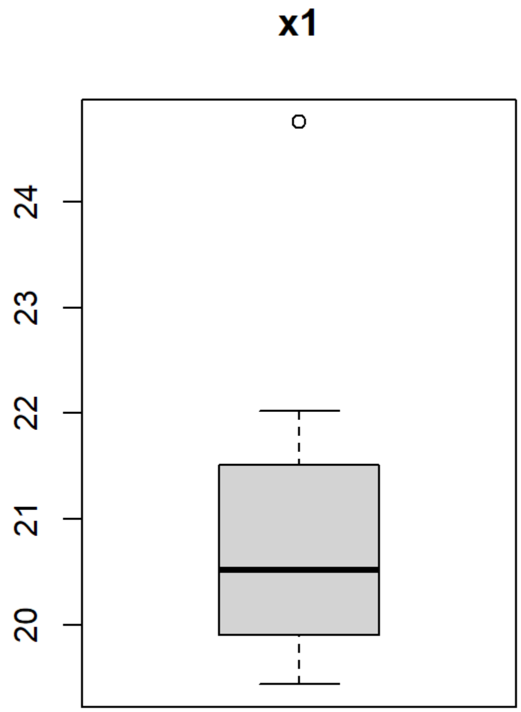
Untuk Nomor 1, Dipergunakan methode mendeteksi pencilan dengan eksplorasi data dengan menggunakan boxplot dengan IQR = 1.5

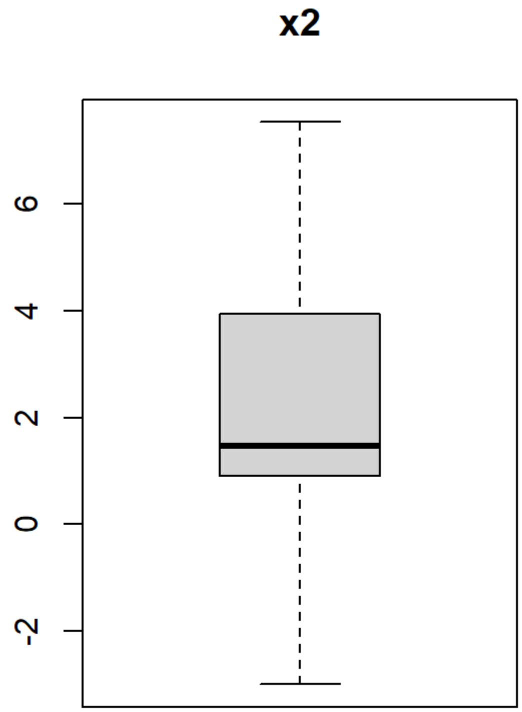
```
data_sheet1 <- read_excel("Data praktikum 12.xlsx",sheet="data 100")
data_sheet2 <- read_excel("Data praktikum 12.xlsx",sheet="data 10")

x1 <- data_sheet2$x1
x2 <- data_sheet2$x2
y1 <- data_sheet1$y1
y2 <- data_sheet1$y2



par(mfrow=c(1,2))
boxplot(x1,main="x1")
boxplot(x2,main="x2")
```





X1 ada satu

outlier yang terlihat, kemudian jika diperhatikan kembali, maka hasilnya cenderung menjulur ke kanan

rosnerTest

```
ros_x1 <- rosnerTest(x1,k=4)
```

```
## Warning in rosnerTest(x1, k = 4): The true Type I error may be larger than
## assumed. See the help file for 'rosnerTest' for a table with information on the
## estimated Type I error level.
```

```
allstat_ros <- ros_x1$all.stats
vec_x1_ros <- x1[allstat_ros[allstat_ros['Outlier']==TRUE,'Obs.Num']]
vec_x1_ros
```

```
## [1] 24.75342
```

X2

Tidak outlier yang terlihat, Kemudian Disini maka cenderung menyebar normal tetapi mean agak bergeser ke kiri

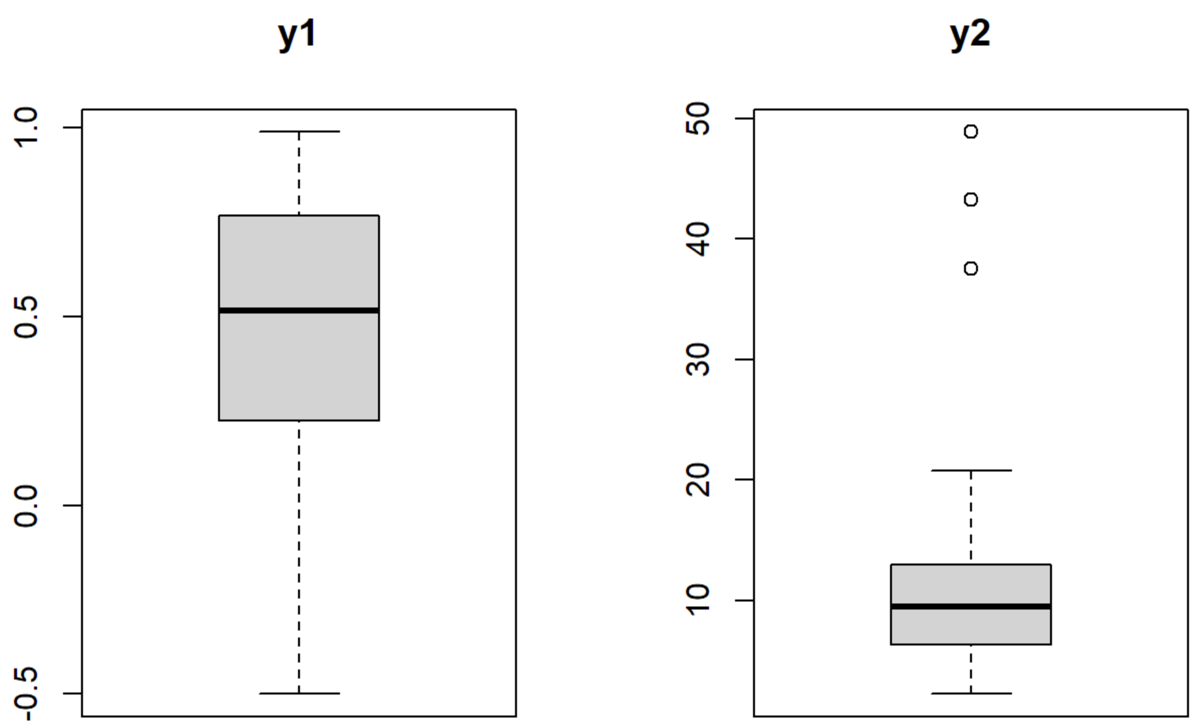
```
ros_x2 <- rosnerTest(x2,k=5)
```

```
## Warning in rosnerTest(x2, k = 5): The true Type I error may be larger than
## assumed. See the help file for 'rosnerTest' for a table with information on the
## estimated Type I error level.
```

```
allstat_ros <- ros_x2$all.stats
vec_x2_ros <- x2[allstat_ros[allstat_ros['Outlier']==TRUE,'Obs.Num']]
vec_x2_ros
```

```
## [1]  7.526452 -2.991378  4.533626  3.939124
```

```
par(mfrow=c(1,2))
boxplot(y1,main="y1")
boxplot(y2,main="y2")
```



Y1

ada satu outlier yang terlihat, kemudian jika diperhatikan kembali, median ada di sebelah kenan

```
ros_y1 <- rosnerTest(y1,k=5)
allstat_ros <- ros_y1$all.stats
vec_y1_ros <- y1[allstat_ros[allstat_ros['Outlier']==TRUE,'Obs.Num']]
vec_y1_ros
```

```
## numeric(0)
```

Y2

Tidak outlier yang terlihat, Kemudian Disini maka cenderung menjulur ke kanan

```
ros_y2 <- rosnerTest(y1,k=5)
allstat_ros <- ros_y1$all.stats
vec_y2_ros <- y1[allstat_ros[allstat_ros['Outlier']==TRUE,'Obs.Num']]
vec_y2_ros
```

```
## numeric(0)
```

Nomor 2

X1

Trimmed Mean

```
mean(x1,trim = 0.05)
```

```
## [1] 21.02891
```

Winsorzed Mean

```
winsor.mean(x1, trim=0.05)
```

```
## [1] 20.92647
```

M-Estimators

huber

```
fun_hub_x1 <- psi.huber(x1, k = 1.345, deriv = 0)
df_hub_x1 <- data.frame(xx=x1,psi_hub=fun_hub_x1)
ggplot(df_hub_x1,aes(xx,psi_hub)) +
  geom_line() + ggtitle("Function huber in X1")
```


```
M_es_hub <- huber(x1, k = 1.345)
M_es_hub$mu
```

```
## [1] 20.77871
```

X2

Trimmed Mean

```
mean(x2,trim = 0.05)
```

```
## [1] 2.020437
```

Winsorzed Mean

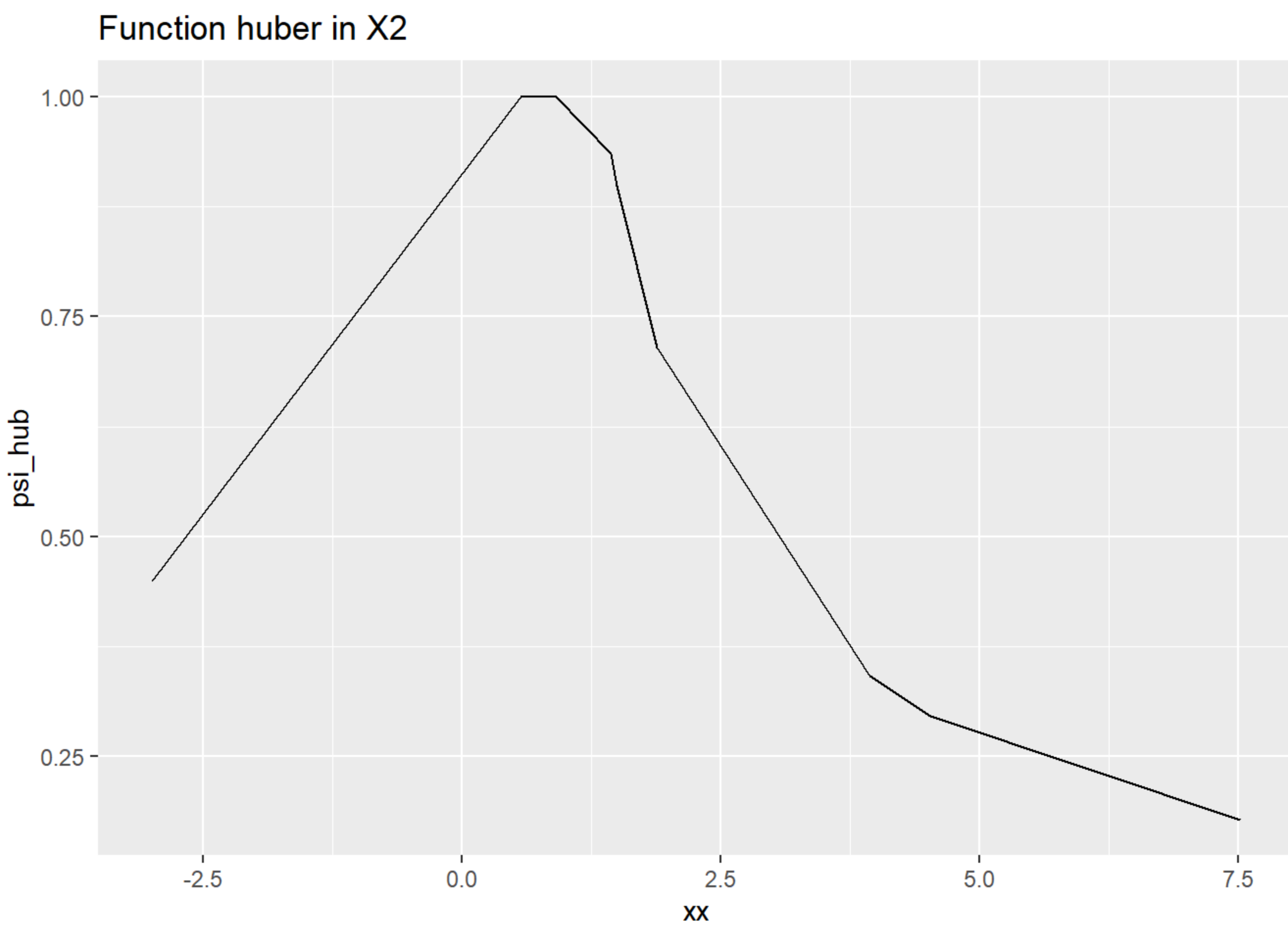
```
winsor.mean(x2, trim=0.05)
```

```
## [1] 2.045956
```

M-Estimators

huber

```
fun_hub_x2 <- psi.huber(x2, k = 1.345, deriv = 0)
df_hub_x2 <- data.frame(xx=x2,psi_hub=fun_hub_x2)
ggplot(df_hub_x2,aes(xx,psi_hub)) +
  geom_line() + ggtitle("Function huber in X2")
```




```
M_es_hub <- huber(x2, k = 1.345)
M_es_hub$mu
```

```
## [1] 1.687373
```

Y1

Trimmed Mean

```
mean(y1,trim = 0.05)
```

```
## [1] 0.5074825
```

Winsorzed Mean

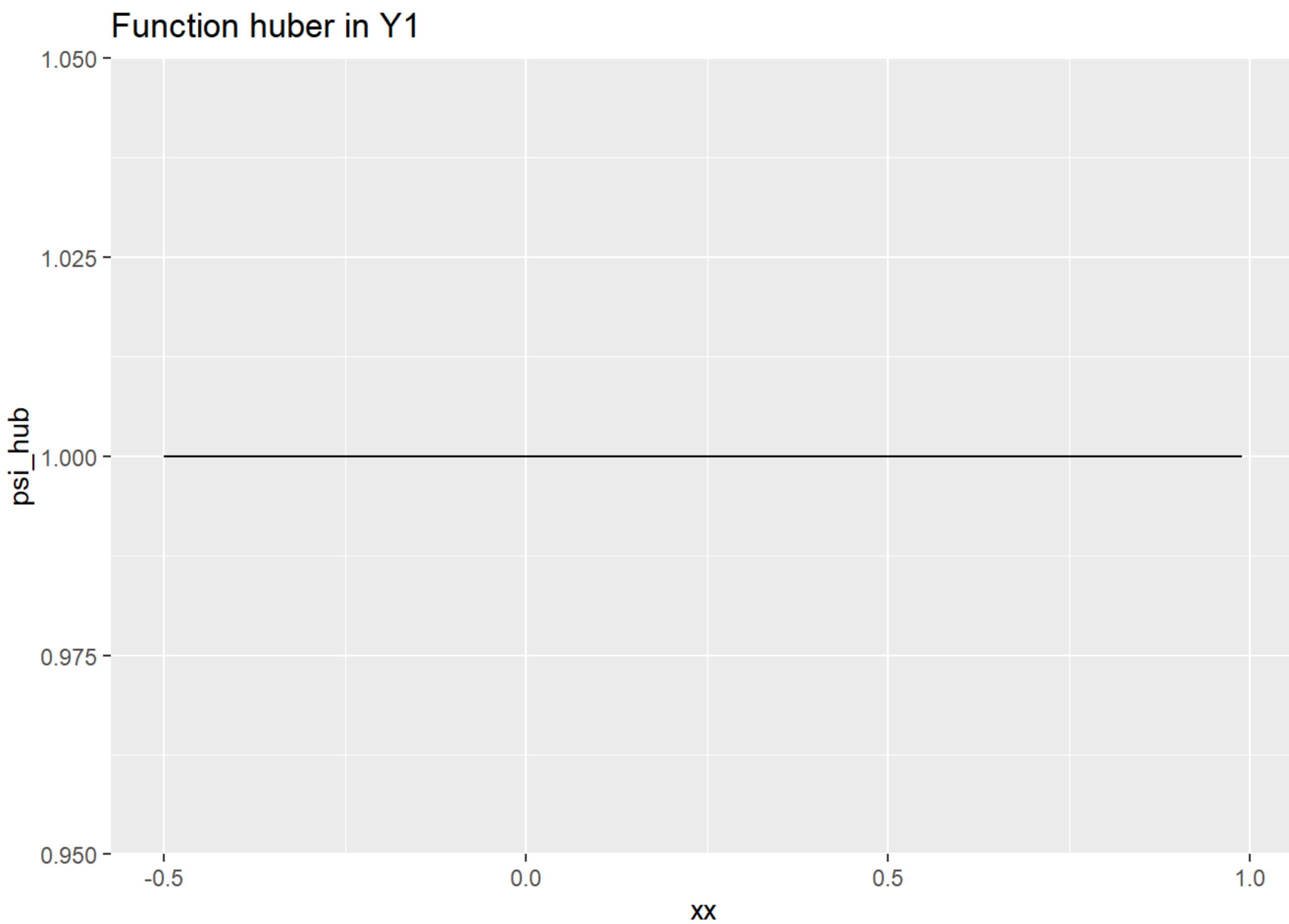
```
winsor.mean(y1, trim=0.05)
```

```
## [1] 0.504838
```

M-Estimators

huber

```
fun_hub_y1 <- psi.huber(y1, k = 1.345, deriv = 0)
df_hub_y1 <- data.frame(xx=y1,psi_hub=fun_hub_y1)
ggplot(df_hub_y1,aes(xx,psi_hub)) +
  geom_line() + ggtitle("Function huber in Y1")
```




```
M_es_hub <- huber(y1, k = 1.345)
M_es_hub$mu
```

```
## [1] 0.5045575
```

Y2

Trimmed Mean

```
mean(y2,trim = 0.05)
```

```
## [1] 9.939848
```

Winsorzed Mean

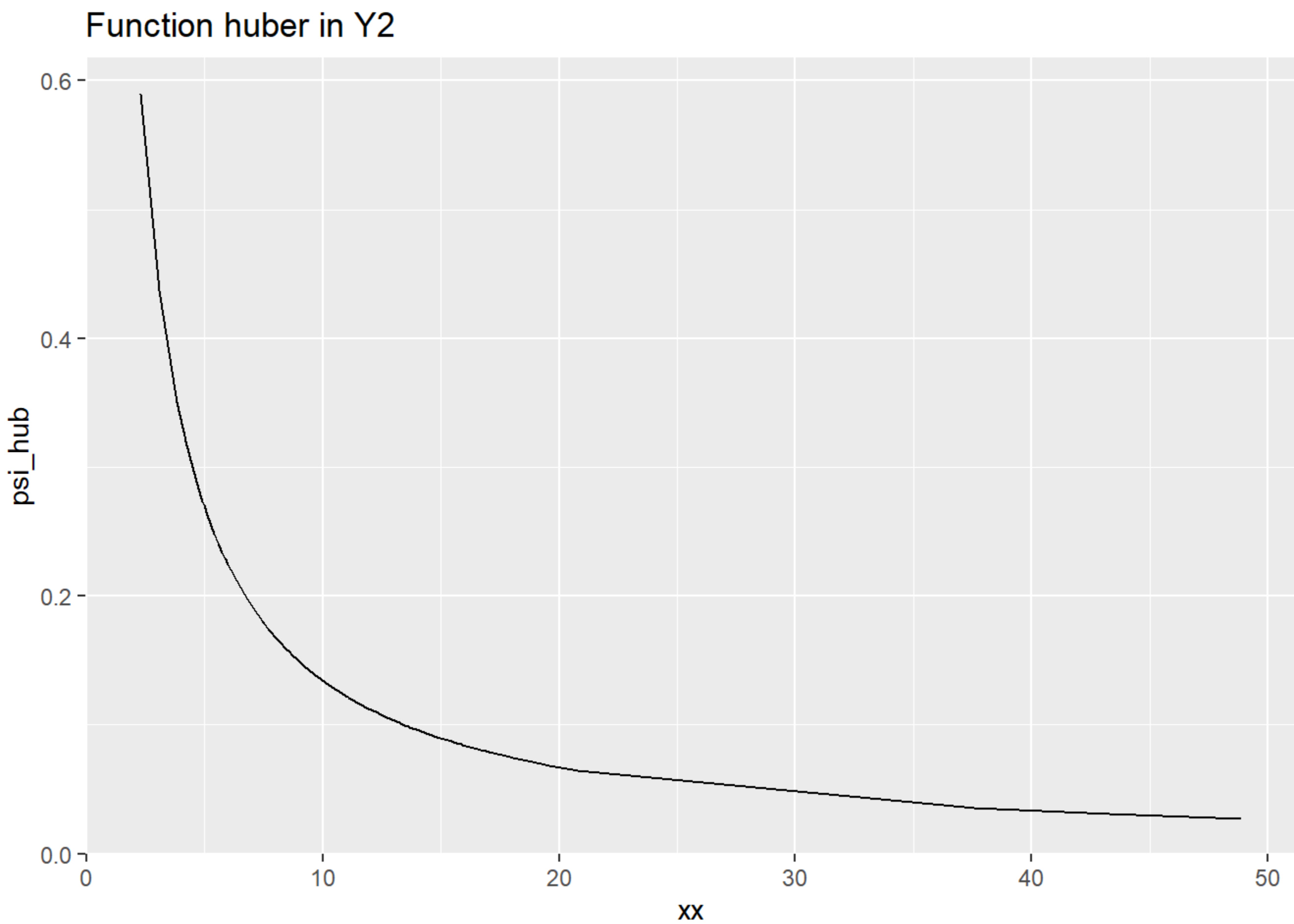
```
winsor.mean(y2, trim=0.05)
```

```
## [1] 10.11176
```

M-Estimators

huber

```
fun_hub_y2 <- psi.huber(y2, k = 1.345, deriv = 0)
df_hub_y2 <- data.frame(xx=y2,psi_hub=fun_hub_y2)
ggplot(df_hub_y2,aes(xx,psi_hub)) +
  geom_line() + ggtitle("Function huber in Y2")
```




```
M_es_hub <- huber(y2, k = 1.345)
M_es_hub$mu
```

```
## [1] 9.890736
```

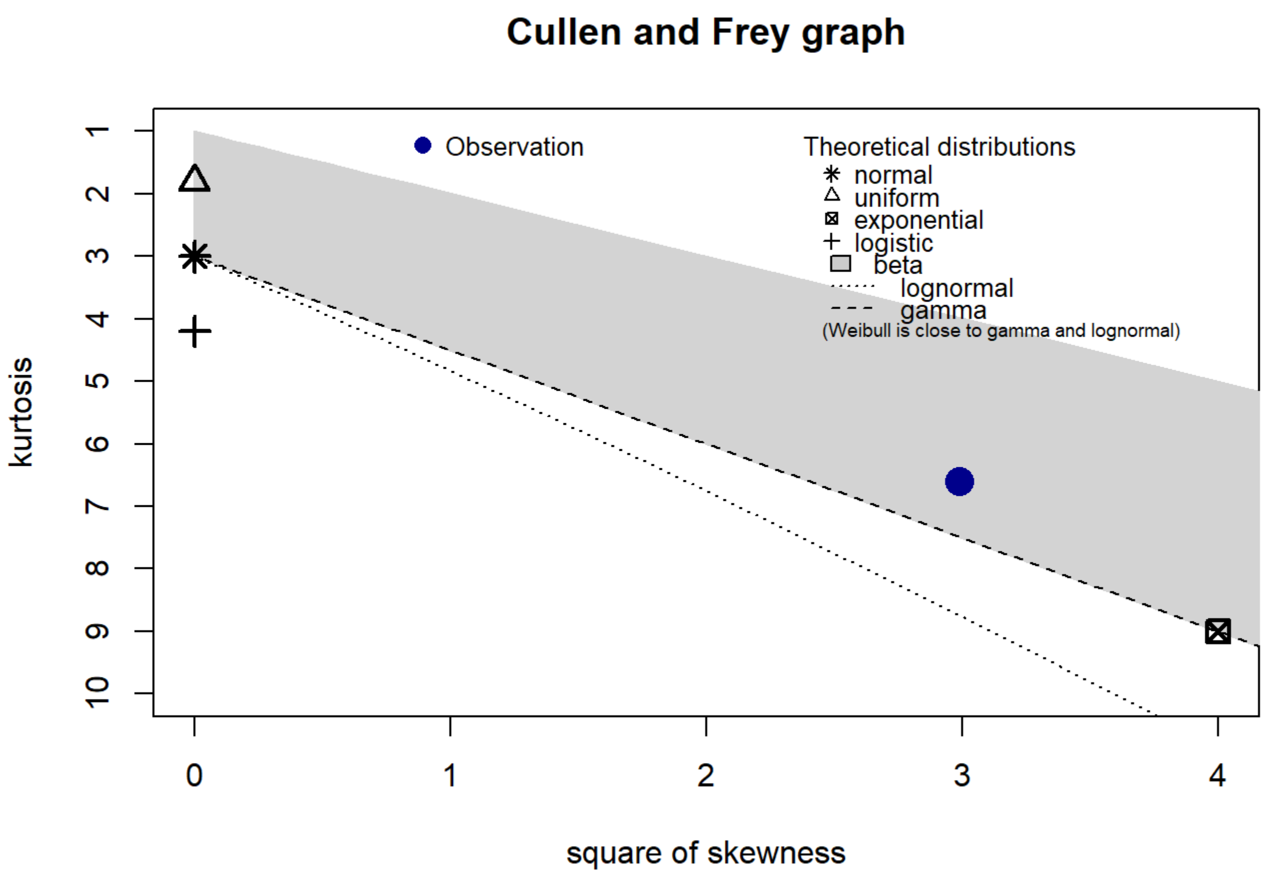
Nomor 3

IQR

Berdasarkan dari link ini (https://web.ipac.caltech.edu/staff/fmasci/home/astro_refs/RobustEstimators.pdf), Maka Sebelum kita menentukan nilai IQR dari suatu data, maka kita lihat distribusi apa yang cocok dengan data tersebut

X1

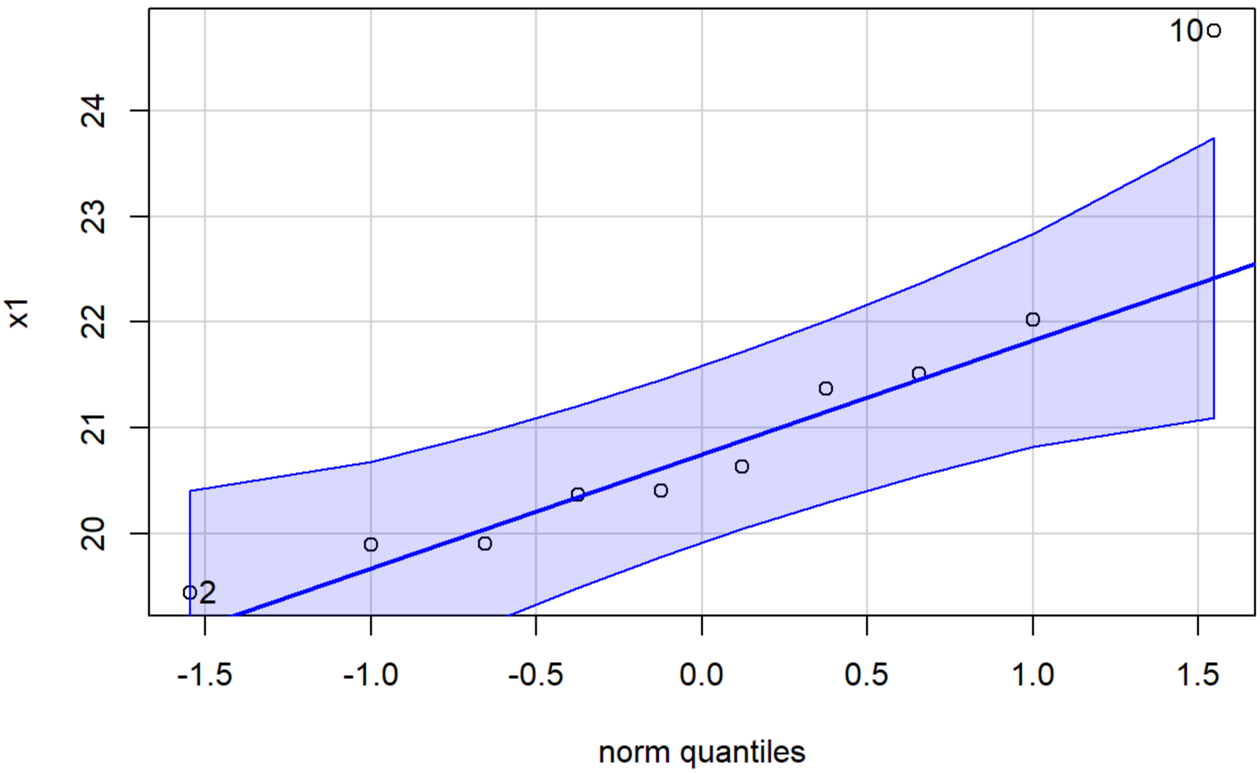
```
descdist(x1)
```




```
## summary statistics
## ------
## min:  19.4353   max:  24.75342
## median:  20.51857
## mean:  21.02891
## estimated sd:  1.537735
## estimated skewness:  1.728735
## estimated kurtosis:  6.605395
```

Dengan Menggunakan Graph Cullen and Frey graph, Maka kemungkinan distribusi x1 antara beta dan gamma. Tetapi dikarenakan error menggunakan Beta, maka bisa kita cek gamma. Tetapi sebelum itu, kita gunakan qqplot dahulu yaaa dengan distibusi normal terlebih dahulu

```
# qqnorm(x1, pch = 1, frame = FALSE)
# qqline(x1, col = "steelblue", lwd = 2)
qqPlot(x1)
```



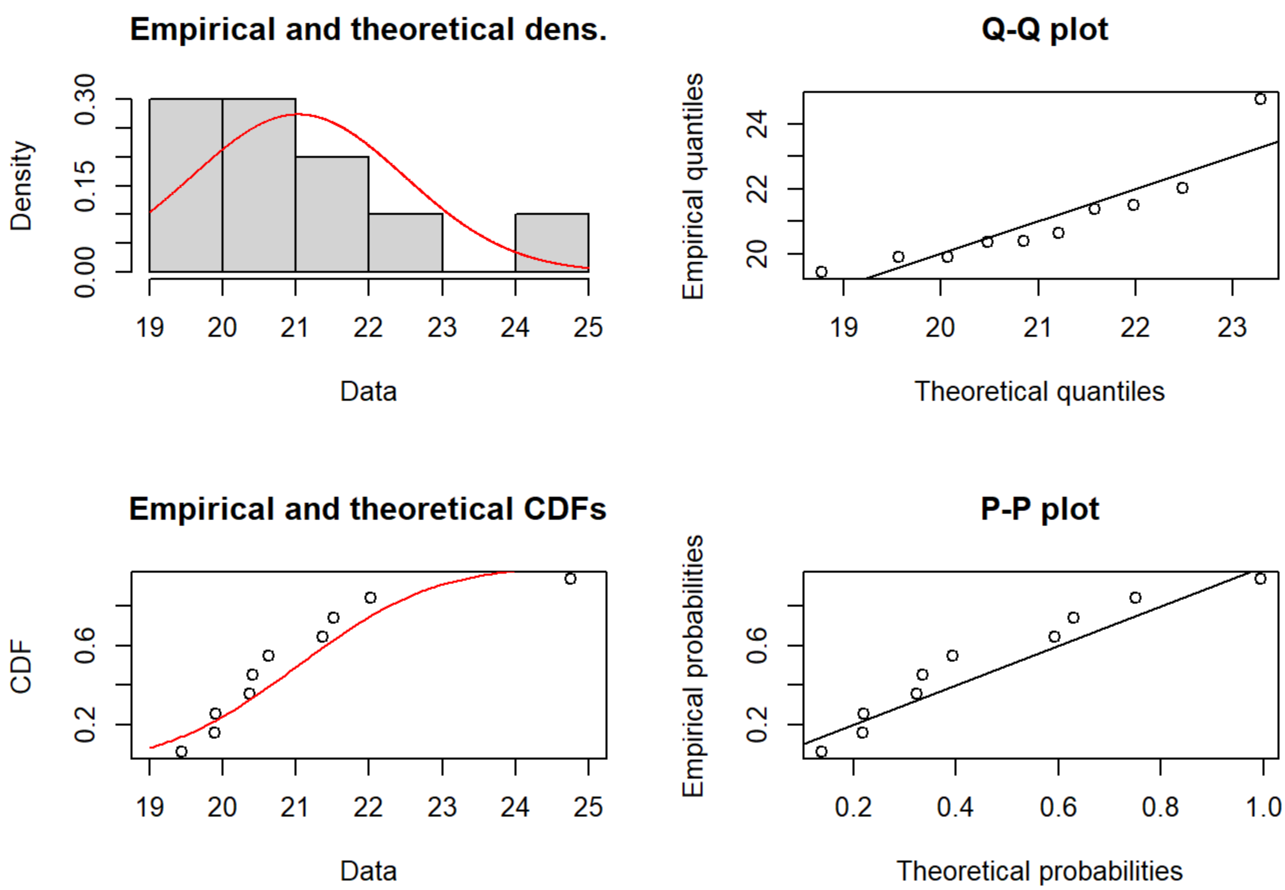

```
## [1] 10  2
```

Setelah diperhatikan, Sebenarnya distribusi ini mendekati normal, tetapi dikarenakan adanya outlier, maka jika kita test dengan ks.test

```
fit_x1 <- fitdist(x1,"norm")
summary(fit_x1)
```

```
## Fitting of the distribution ' norm ' by maximum likelihood
## Parameters :
##       estimate Std. Error
## mean 21.028911  0.4613204
## sd    1.458823  0.3262021
## Loglikelihood:  -17.96569   AIC:  39.93137   BIC:  40.53654
## Correlation matrix:
##      mean sd
## mean    1  0
## sd      0  1
```

```
plot(fit_x1)
```




```
ks.test(x1,"pnorm",mean=21.028911,sd=1.458823)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  x1
## D = 0.20699, p-value = 0.7123
## alternative hypothesis: two-sided
```

```
shapiro.test(x1)
```

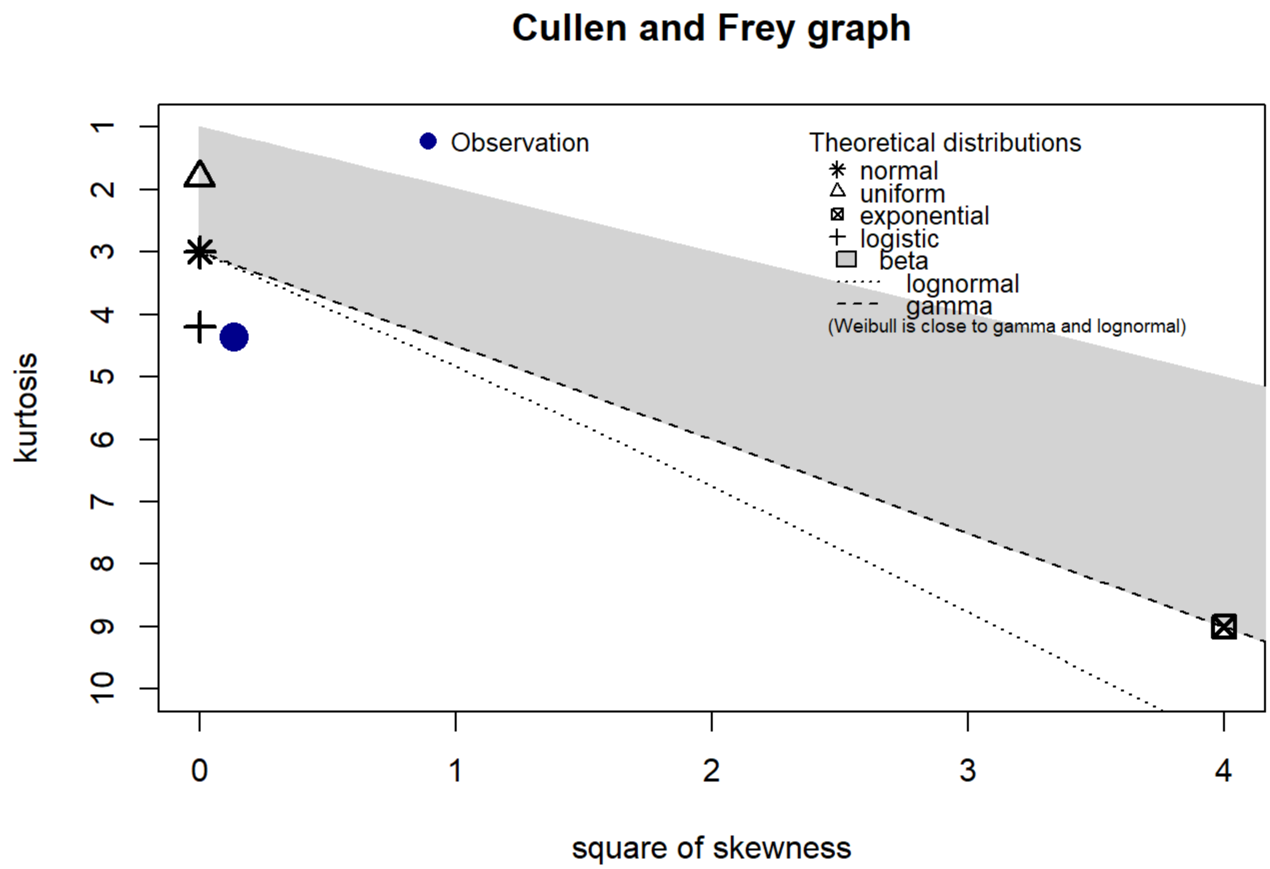
```
##
##  Shapiro-Wilk normality test
##
## data:  x1
## W = 0.83965, p-value = 0.0437
```

jika diperhatikan, jika ks.test normal, sedangkan shapiro-wilk test ini tidak normal, untuk p-value shapiro-wilk ini, mendekati normal, meskipun mendekati p-value.

X1 itu adalah Distribusi Normal (mean=21.028911,sd=1.458823)

X2

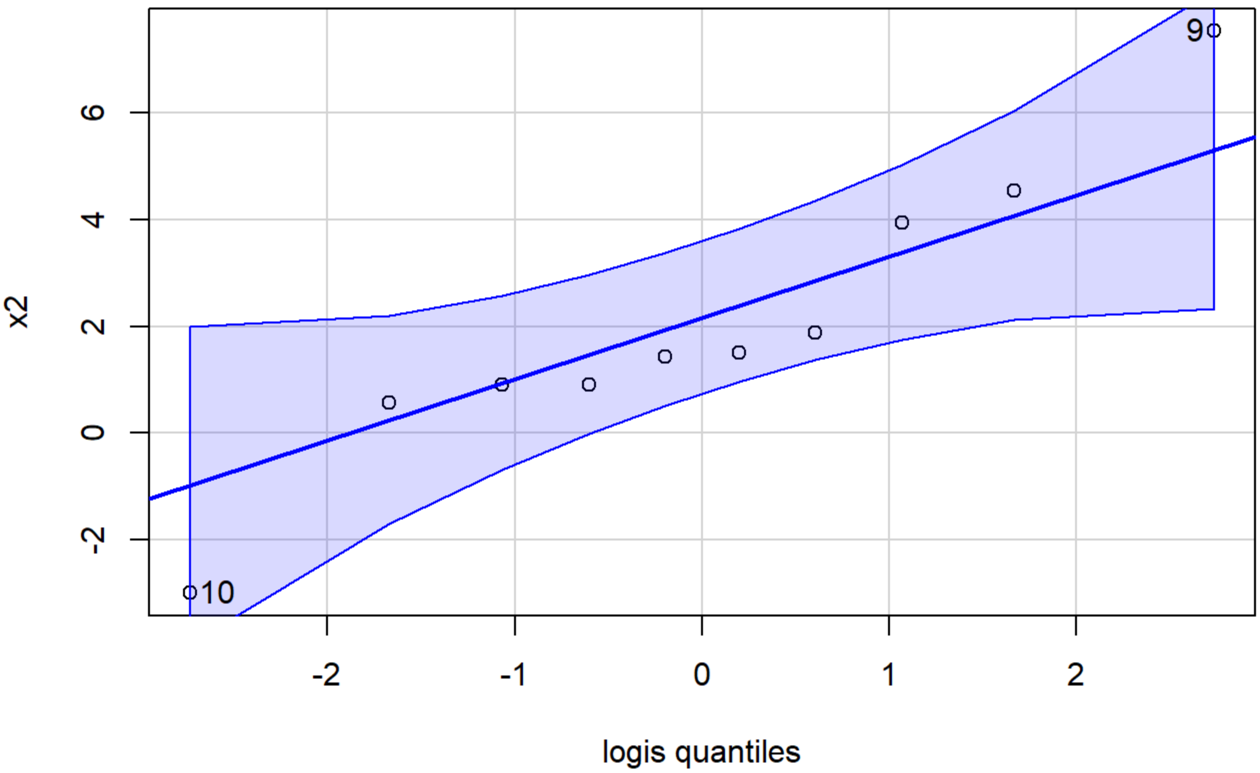
```
descdist(x2)
```




```
## summary statistics
## ------
## min:  -2.991378   max:  7.526452
## median:  1.468013
## mean:  2.020437
## estimated sd:  2.802018
## estimated skewness:  0.3648262
## estimated kurtosis:  4.365285
```

Disini, banyak sekali kemungkinan yaaa. jika diperhatikan, distribusinya ada logistic , atau mungkin lognormal. Kita cek qqplot yang logistic yaaa

```
qqPlot(x2,"logis")
```



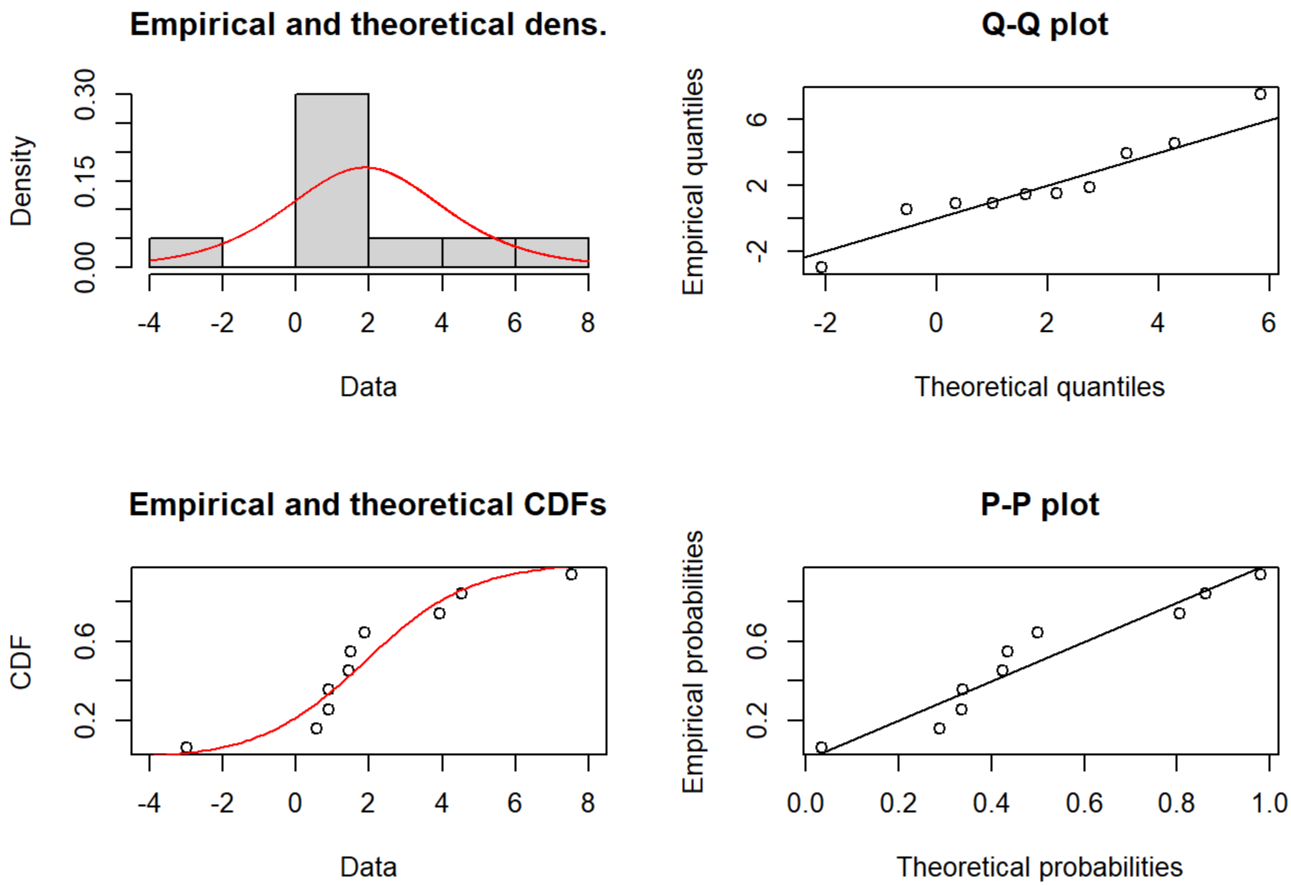

```
## [1]  9 10
```

Kemudian, kita lihat, bahwa x2 distribusi lognormal , jika diperhatikan, bahwa ada outlier yaaa. ada 2 outlier yaa, yang lain bisa dikatakan distribusi logistik

```
fit_x2 <- fitdist(x2,"logis")
fit_x2
```

```
## Fitting of the distribution ' logis ' by maximum likelihood
## Parameters:
##          estimate Std. Error
## location 1.882585  0.7817310
## scale    1.446234  0.3932554
```

```
plot(fit_x2)
```



Bisa dilihat

bahwa sebenarnya data ini merupakan distribusi logistik

```
ks.test(x2,"plogis",location=1.882585,scale=1.446234)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  x2
## D = 0.19993, p-value = 0.7491
## alternative hypothesis: two-sided
```

```
shapiro.test(x2)
```

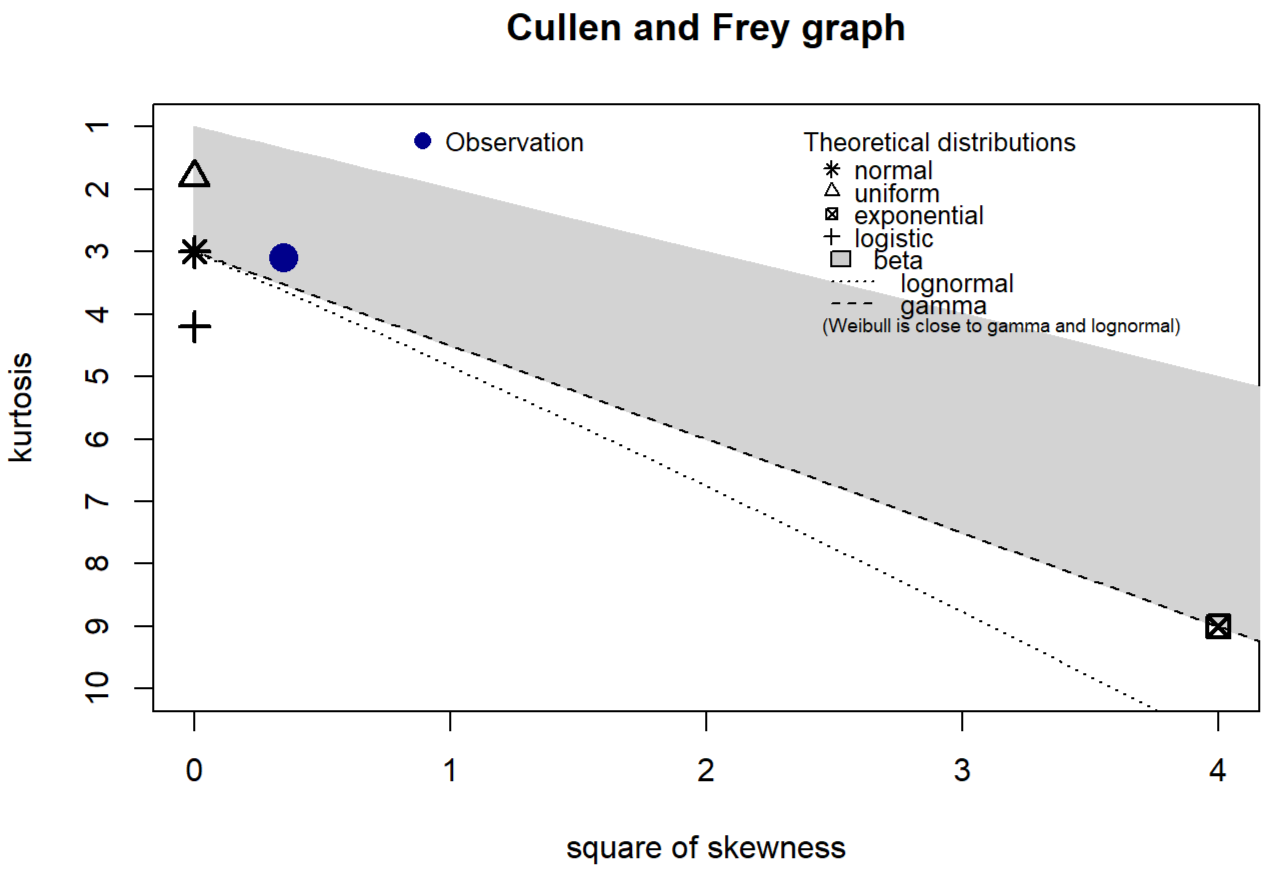
```
##
##  Shapiro-Wilk normality test
##
## data:  x2
## W = 0.93129, p-value = 0.4607
```

jika diperhatikan, jika ks.test logis, sedangkan shapiro-wilk test ini tidak normal, berarti tepat bahwa data ini distribusi logistik dan juga bukan distribusi normal

X2 itu adalah Distribusi Logistik (location=1.882585,scale=1.446234)

Y1

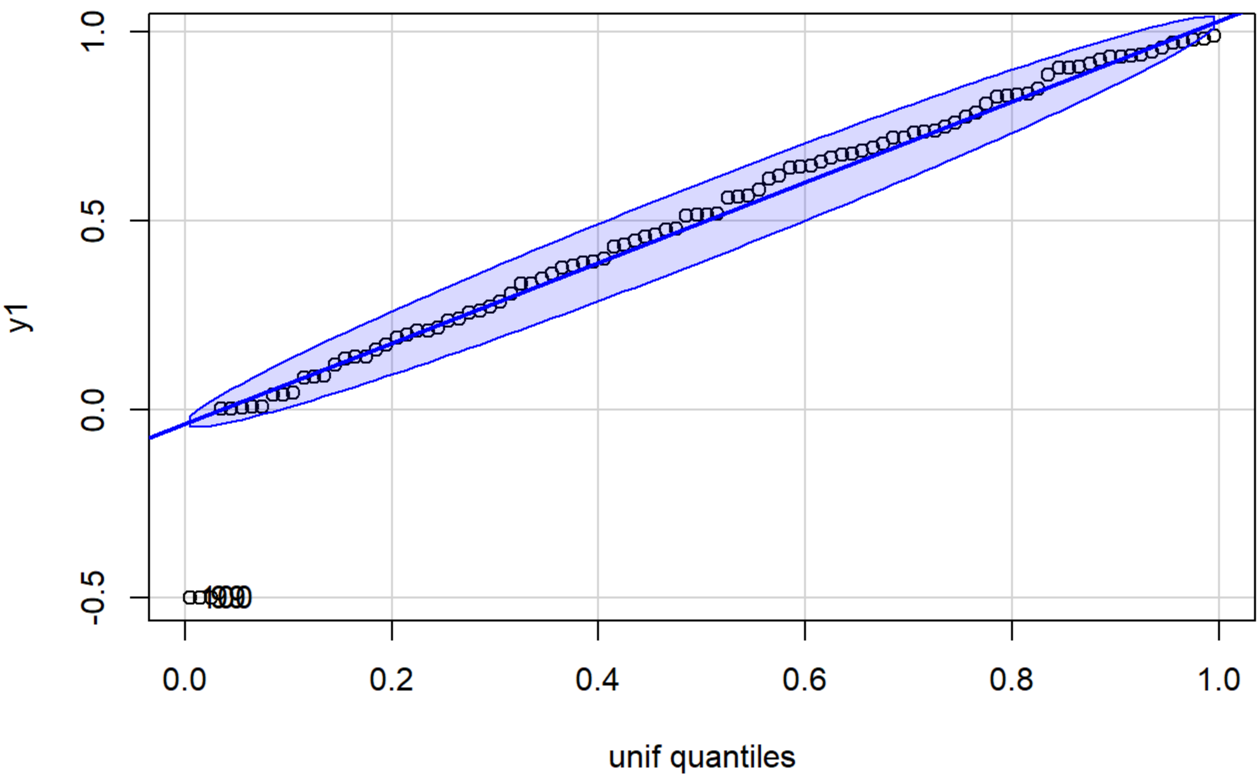
```
descdist(y1)
```




```
## summary statistics
## ------
## min:  -0.499999   max:  0.9888917
## median:  0.5144103
## mean:  0.4906979
## estimated sd:  0.3482005
## estimated skewness:  -0.5900393
## estimated kurtosis:  3.094567
```

Bisa diperhatikan, bahwa Asumsi distribusi uniform, normal , gamma, dan logistik. berarti kita cek yahhh 4 distribusi, berarti kita cek yaaa dahulu uniform

```
qqPlot(y1,distribution="unif")
```



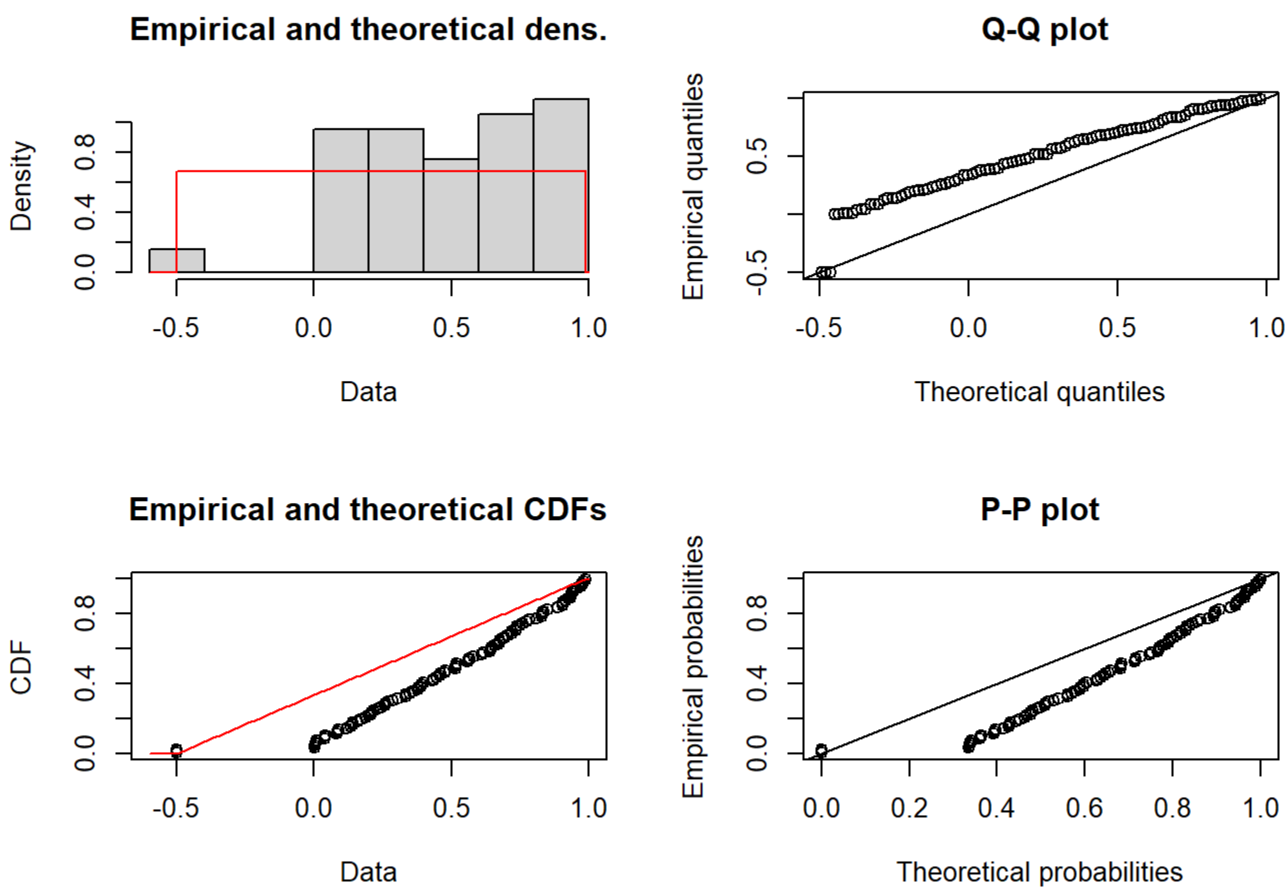

```
## [1] 100  99
```

Jika diperhatikan, ada beberapa partikel yang melewati batas partikel yaaa. nah itu outlier yaaa, ada sedikit kesalahan disitu, kemudian hasilnya pada qqplot sudah pasti adalah distribusi uniform dikarenakan sudah tepat

```
fit_y1 <- fitdist(y1,"unif")
fit_y1
```

```
## Fitting of the distribution ' unif ' by maximum likelihood
## Parameters:
##       estimate Std. Error
## min -0.4999990         NA
## max  0.9888917         NA
```

```
plot(fit_y1)
```




```
ks.test(y1,"punif")
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  y1
## D = 0.072115, p-value = 0.6758
## alternative hypothesis: two-sided
```

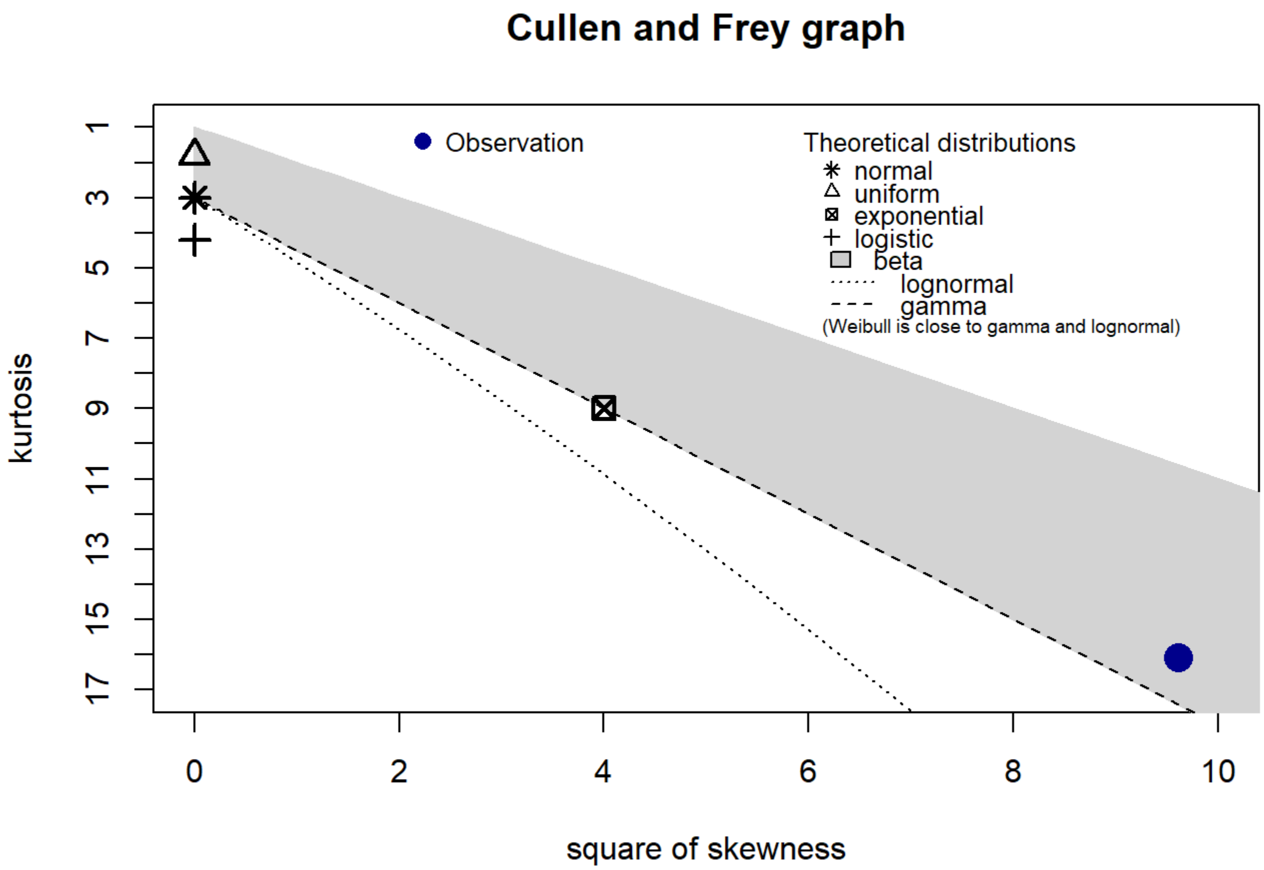
```
shapiro.test(y1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  y1
## W = 0.94574, p-value = 0.0004397
```

Y1 itu Distribusi Uniform (min=-0,4999990,max=0.3464552)

Y2

```
descdist(y2)
```



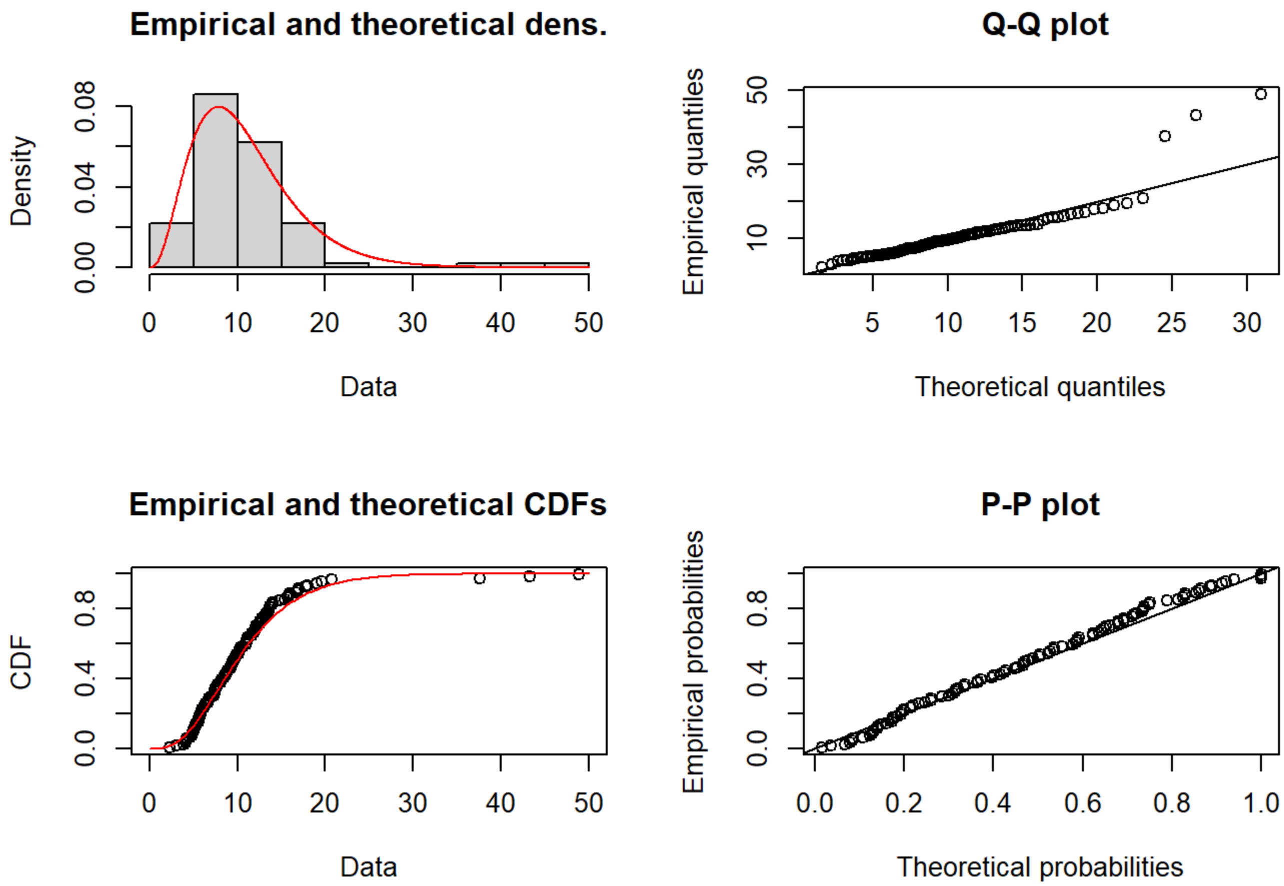

```
## summary statistics
## ------
## min:  2.279965   max:  48.86305
## median:  9.499855
## mean:  10.82127
## estimated sd:  7.081105
## estimated skewness:  3.10021
## estimated kurtosis:  16.07243
```

bisa diperhatikan bahwa observasi ini, bisa gamma atau beta distribusinya, tapi kita cek dulu normal deh

```
fitdist(y2,"gamma")
```

```
## Fitting of the distribution ' gamma ' by maximum likelihood
## Parameters:
##        estimate Std. Error
## shape 3.6125932 0.48912413
## rate  0.3338428 0.04849241
```

```
plot(fitdist(y2,"gamma"))
```




```
## hubungan anatara gamma denganchisquare adalah sebagai berikut https://programmathically.com/chi-square-distribution-and-degrees-of-freedom/
ks.test(y2,"pgamma",shape=5,rate=0.5)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  y2
## D = 0.04966, p-value = 0.9661
## alternative hypothesis: two-sided
```

dikarenakan shape = 5 dan juga rate = 0.5 maka bisa dikatakan bahwa shape =5 dengan syarat rate=0.5 –> df = 10

```
ks.test(y2,"pchisq",df=10)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  y2
## D = 0.04966, p-value = 0.9661
## alternative hypothesis: two-sided
```

Setelah diujicoba dan diperhatikan dengan seksama , maka hasilnya adalah

Y2 in distribution Chi Square (df=10)

Perhitungan IQR

Rujukan adalah link ini ya (https://dc.etsu.edu/cgi/viewcontent.cgi?article=2187&context=etd)

1. X1 Distribusi Normal (mean=21.028911,sd=1.458823)
2. X2 Distribusi Logistik (location=1.882585,scale=1.446234)
3. Y1 Distribusi Uniform (min=-0,4999990,max=0.3464552)
4. Y2 Distribusi Chisquare (df=10)

X1

Nilai IQR dari X1 adalah

```
## Dikarenakan Normal maka Formulanya
## IQRnorm ≈ 1.34900σ
IQR_X1 <- 1.349*sd(x1)
print(paste('IQR_X1 =',IQR_X1,sep=' '))
```

```
## [1] "IQR_X1 = 2.0744039770859"
```

X2

Nilai IQR dari X2 adalah

```
## Dikarenakan Logistic maka Formulanya
IQR_X2 <- IQR(x2)
print(paste('IQR_X2 =',IQR_X2,sep=' '))
```

```
## [1] "IQR_X2 = 2.52318627238694"
```

Y1

Nilai IQR dari Y1 adalah

```
## Dikarenakan Logistic maka Formulanya
## IQRunif ≈ 0.5(θ2 - θ1)
IQR_Y1 <- 0.5*(0.3464552-(-0.4999990))
print(paste('IQR_Y1 =',IQR_Y1,sep=' '))
```

```
## [1] "IQR_Y1 = 0.4232271"
```

Y2

Nilai IQR dari Y2 adalah

```
## Dikarenakan Logistic maka Formulanya
## IQRunif ≈ 2(e^μ)sinh(0.6745σ)
IQR_Y2 <- IQR(y2)
print(paste('IQR_Y2 =',IQR_Y2,sep=' '))
```

```
## [1] "IQR_Y2 = 6.51586843556726"
```

Jika diperhatikan bahwa hasil IQR itu sendiri adalah tidak berpengaruh terhadap pencilan dikarenakan data tersebut adalah pada quartil 1 dan 3, sehingga pencilan tidak dianggap.