

:: Praktikum Statistika menggunakan R ::

02. Statistika Deskriptif

Sari Numerik

MA2181 Analisis Data / MA2081
Statistika Dasar / MA2082 Biostatistika

Kelompok Keilmuan Statistika

Laboratorium Statistika dan Komputasi Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam



TUJUAN PRAKTIKUM

1

Mengambil informasi dari data (*data summary*) sebagai gambaran awal bagi peneliti/pengamat, khususnya dalam mengeksplorasi data mentah yang ada.

2

Hasil yang diharapkan dari informasi data adalah menentukan distribusinya



Ukuran Pemusatan Data

Rata-rata Aritmatika

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Rata-rata Geometri

$$GM = \sqrt[n]{x_1 x_2 \dots x_n}$$

Modus

nilai yang paling sering muncul"

Kuartil bawah

$$q_1 = x_{\left(\frac{n+1}{4}\right)}$$

Kuartil atas

$$q_3 = x_{\left(\frac{3(n+1)}{4}\right)}$$

Median

$$q_2 = x_{\left(\frac{n+1}{2}\right)}$$

Trirata

$$TRI = \frac{q_3 + q_1 + 2q_2}{4}$$



Ukuran Penyebaran Data

Variansi

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Simpangan Baku

$$s = \sqrt{s^2}$$

Jangkauan (*Range*)

$$range = x_{maks} - x_{min}$$

Standar Error

$$SE = \frac{s}{\sqrt{n}}$$

Simpangan Kuartil

$$d_q = q_3 - q_1$$

Ukuran Kemencengan (*skewness*) dan Kelancipan (*kurtosis*)

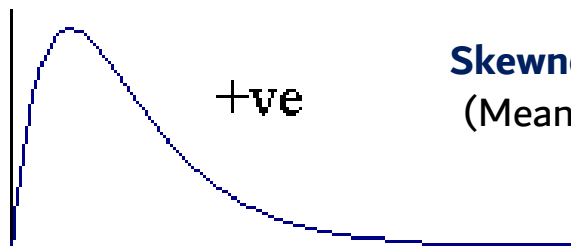
Skewness

$$\hat{\gamma} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{3}{2}}}$$

Kurtosis

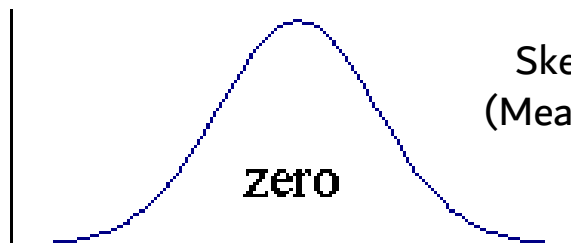
$$\hat{K}(x) = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2}$$

Skewness



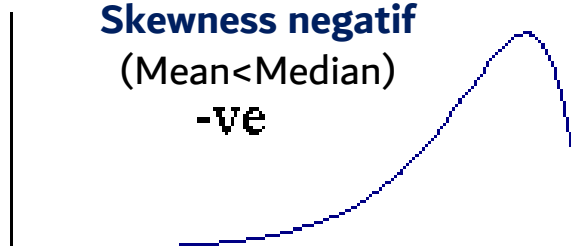
+ve

Skewness positif
(Mean > Median)



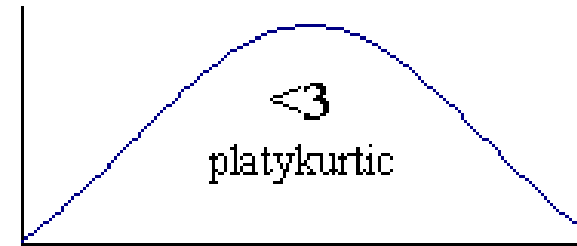
zero

Skewness=0
(Mean=Median)

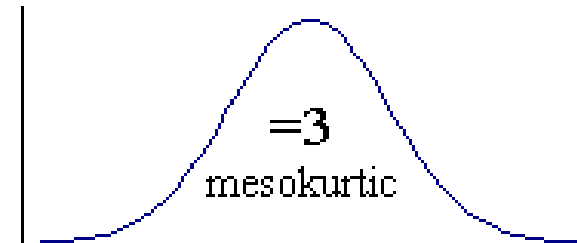


Skewness negatif
(Mean < Median)
-ve

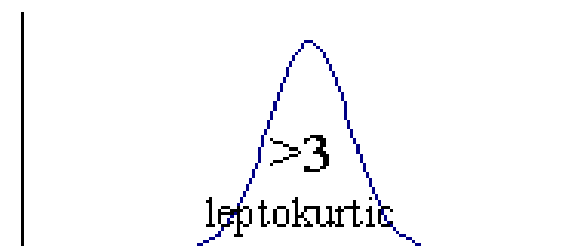
Kurtosis



<3
platykurtic



=3
mesokurtic



>3
leptokurtic

Contoh 1

Data berikut menyatakan laju pencacahan unsur Cesium Cs-137 selama 25 kali per 10 detik:

30	35	32	27	18	36
19	43	44	39	24	33
44	39	55	34	35	40
46	30	34	35	43	21
26	35	37	29	23	24

Import Data

Ketikkan data Cesium pada Contoh 1 di media input yang dipilih seperti Ms. Excel, simpan file dengan format csv atau xlsx. Misalkan simpan file dengan nama “Cesium” pada direktori “D:/Praktikum stat”.

Pada R Editor, tuliskan :

Untuk file csv

```
Cesium <- read.table("D:/Praktikum stat/Cesium.csv")
```

Untuk file xlsx

```
library(xlsx)  
Cesium <- read.xlsx("D:/Praktikum stat/Cesium.xlsx",  
sheetName = "Sheet1")
```

Syntax Statistika Deskriptif

```
# STATISTIKA DESKRIPTIF
```

```
summary(Cesium)
```

```
V1
```

```
Min.    :18.00
```

```
1st Qu.:27.50
```

```
Median :34.50
```

```
Mean    :33.67
```

```
3rd Qu.:39.00
```

```
Max.    :55.00
```

```
(jumlah = sum(Cesium))
```

```
1010
```

```
(modus = names(sort(-table(Cesium)))[1])
```

```
"35"
```

```
library(psych)
```

```
describe(Cesium)
```

vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se	
X1	1	30	33.67	8.7	34.5	33.62	8.15	18	55	37	0.15	-0.43	1.59

Syntax Statistika Deskriptif

```
library(Hmisc)
describe(Cesium)
```

Cesium

1 Variables 30 Observations

V1

	n	missing	distinct	Info	Mean
	30	0	21	0.996	33.67
Gmd		.05	.10	.25	.50
	9.963	19.9	22.8	27.5	34.5
	.75	.90	.95		
	39.0	44.0	45.1		

lowest : 18 19 21 23 24, highest: 40 43 44 46 55

```
library(pastecs)
stat.desc(Cesium)
```

V1

nbr.val	30.0000000
nbr.null	0.0000000
nbr.na	0.0000000
min	18.0000000
max	55.0000000
range	37.0000000
sum	1010.0000000
median	34.5000000
mean	33.6666667
SE.mean	1.5875473
CI.mean.0.95	3.2468989
var	75.6091954
std.dev	8.6953548
coef.var	0.2582779



Keluaran Statistika Deskriptif

Pemusatan	
Banyaknya	50
Jumlah	1016
Rata-rata	20,32
Minimum	7
Kuartil Bawah	14,25
Median	19,5
Kuartil Atas	24
Maksimum	45
Modus	24

Rataan >
median

Persebaran	
Variansi	70,01
Simpangan Baku	8,37
Simpangan Kuartil	9,75
Jangkauan	73
Standar Error	1,59
Kemencengan ($\hat{\gamma}=0$)	0,15
Keruncingan ($\widehat{K(x)}=3$)	-0,43

Kemencengan positif : data menumpuk di nilai yang lebih kecil dari rata-rata. **Keruncingan** < 3 : distribusi lebih landai dari distribusi normal.

Sebanyak selang kepercayaan 1- σ atau **68% data** berada pada interval nilai **11,95 sampai 28,69** (rata-rata \pm simpangan baku).



:: Praktikum Statistika menggunakan R ::

02. Statistika Deskriptif

Grafik

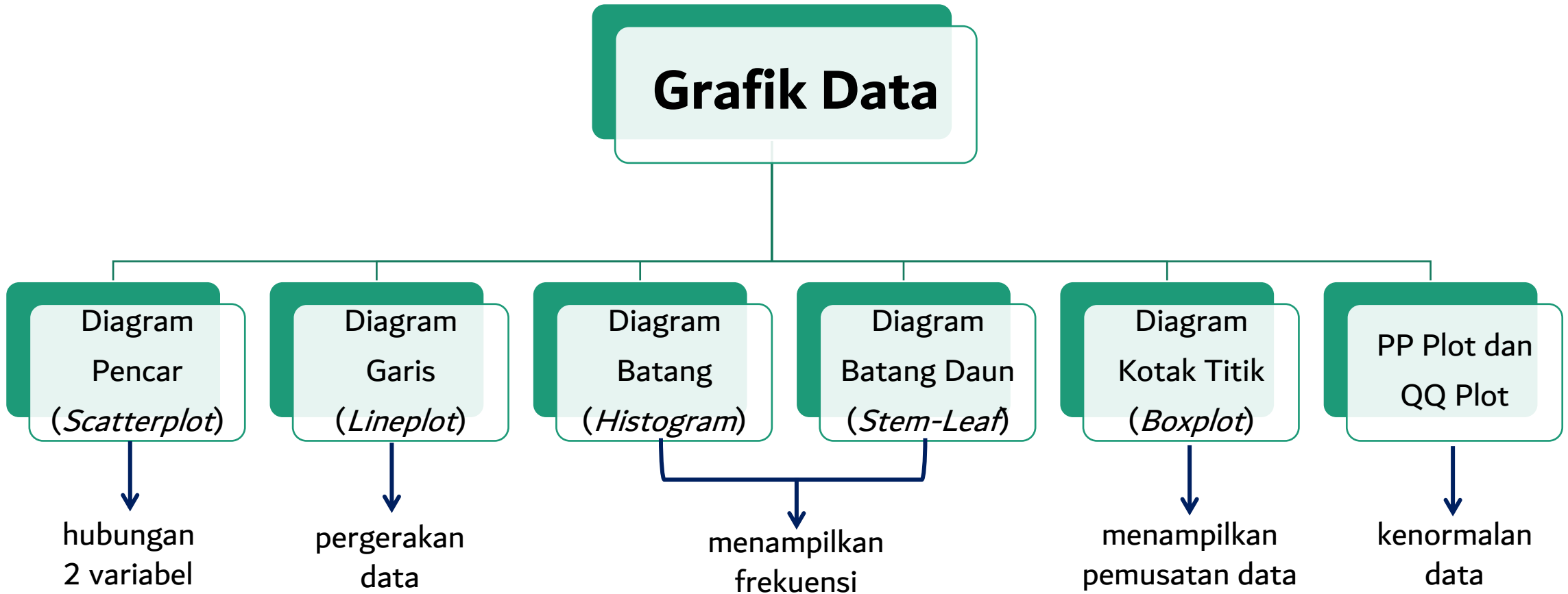
MA2181 Analisis Data / MA2081
Statistika Dasar / MA2082 Biostatistika

Kelompok Keilmuan Statistika

Laboratorium Statistika dan Komputasi Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam



Grafik Data



Contoh 2

Data *Estates Production by Crops*, Indonesia, 1990-2001 (Ton)

Year	Hevea Rubber	Palm Oil	Palm Kernel	Cocoa	Coffee	Tea	Quinine	Tobacco
1990	315,3	2.096.900	445,8	41,5	25,5	129,1	1,9	3,5
1991	330,1	1.843.600	406,2	30,6	26,4	125	2,1	4,9
1992	335	2.186.000	483,1	39,5	23,9	113	2,7	7,5
1993	335	2.288.300	524,6	42,7	20,9	100	6,00	3,1
1994	326,4	1.930.300	472,1	43,7	19,7	98	3,00	5,1
1995	341	2.476.400	605,3	46,4	20,8	111,1	3,00	9,9
1996	334,6	2.569.500	626,6	48,8	28,5	132	4,00	7,1
1997	330,5	4.081.100	927,5	65,9	30,6	121	5,00	7,8
1998	332,6	4.013.100	912,1	60,9	28,5	132,7	4,00	7,7
1999	303,6	4.024.821	914,7	58,9	27,5	130,5	9,17	5,8
2000	336,2	4.094.073	930,6	60,6	29,5	127,9	9,34	6,3
2001	328,3	4.152.596	946,9	65,3	28,7	129,3	9,20	5,1

Bagaimana cara menyarikan informasi - informasi yang terkandung dalam data di atas dalam bentuk grafik?



Diagram Pencar (*Scatterplot*)

Plot yang menandai setiap observasi dengan titik dalam bidang koordinat segiempat XY. Diagram pencar dapat dimanfaatkan sebagai acuan awal untuk melihat hubungan yang mungkin dari dua variabel yang diamati.

```
# 1. scatterplot  
# plot(x, y, main="judul", xlab="judul sumbu x", ylab="judul sumbu y")  
plot(estates$Coffee, estates$Tea, main="Scatterplot Produksi Kopi Terhadap  
Teh (ton)", xlab="Kopi", ylab="Teh")
```

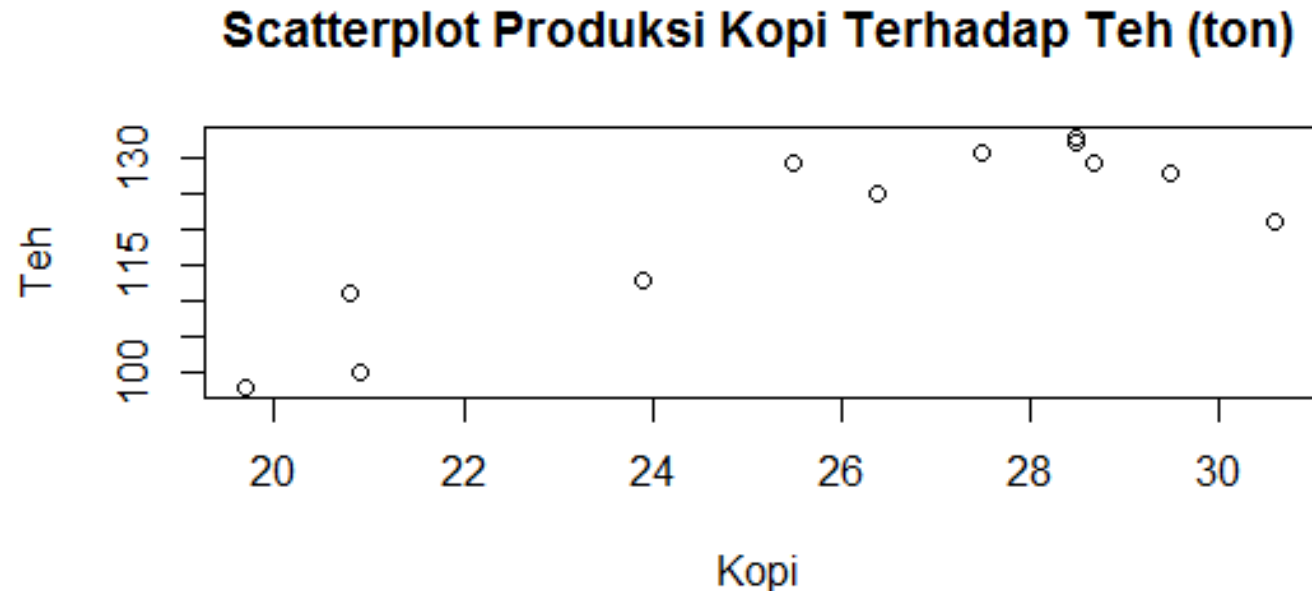


Diagram Garis (*Lineplot*)

Suatu jenis diagram untuk pengamatan data dari waktu ke waktu secara berurutan. Sumbu X menunjukkan waktu pengamatan, Y nilai data pengamatan.

```
# 2. dotplot  
# plot(x, type="o", main="judul",xlab="judul sumbu x",ylab="judul sumbu y")  
plot(estates$Palm.Oil,type="o",main="Diagram Garis Produksi Minyak Kelapa",  
      xlab="Tahun ke-", ylab="minyak kelapa (ton)")
```

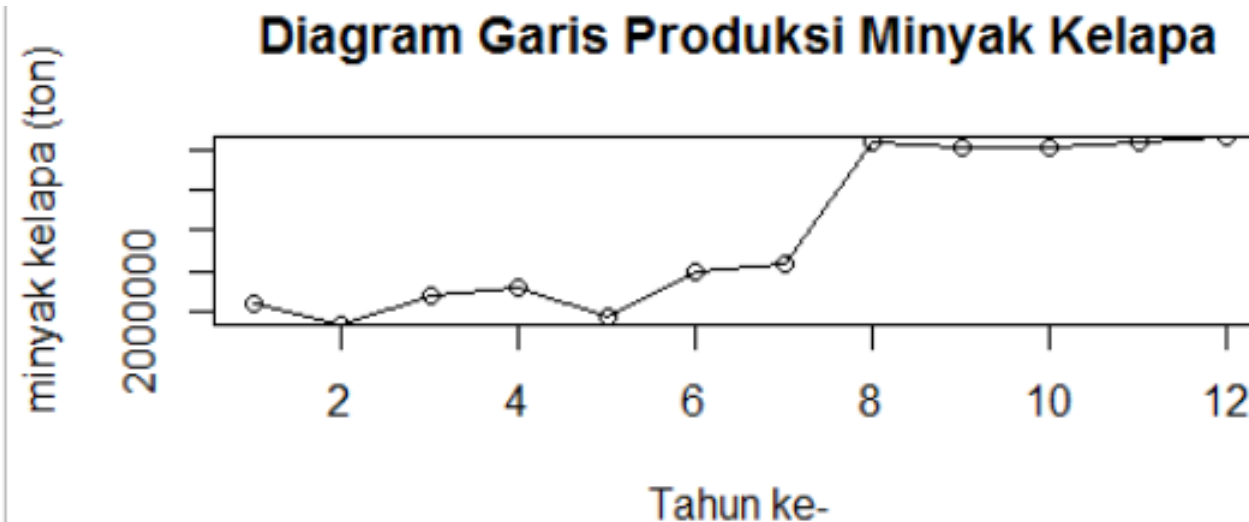


Diagram Batang (*Histogram*)

Diagram menggambarkan diagram batang antara data terhadap frekuensi, frekuensi kumulatif, persentase atau persentase kumulatif.

```
# 3. histogram  
# hist(data, freq, main="judul",xlab="judul sumbu x",ylab="judul sumbu y")  
# freq = TRUE (frekuensi), FALSE (frekuensi relatif)  
hist(Cesium$V1,freq=T, main="Laju Pencacahan Cs-137", xlab="Laju (per/ 10  
      detik)", ylab="Frekuensi")
```

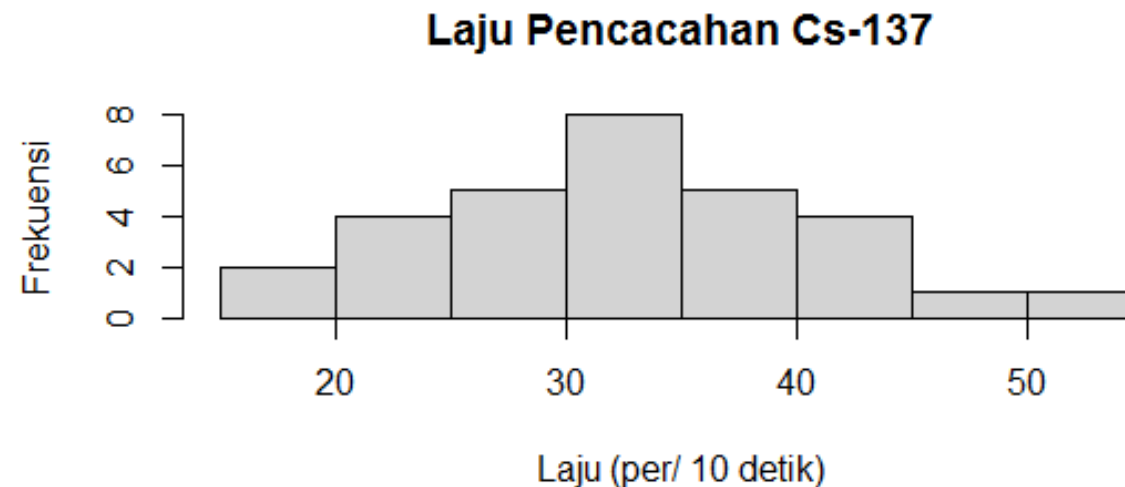


Diagram Batang Daun (*Stem Leaf*)

Diagram dapat memberikan gambaran bentuk distribusi data dan mampu mempertahankan tampilan data asli yang disusun terurut. Lebih jauh, tampilan diagram ini seperti bentuk histogram yang dirotasi 90° searah jarum jam.

```
# 4. steam leaf  
# stem(data)  
stem(Cesium$V1)
```

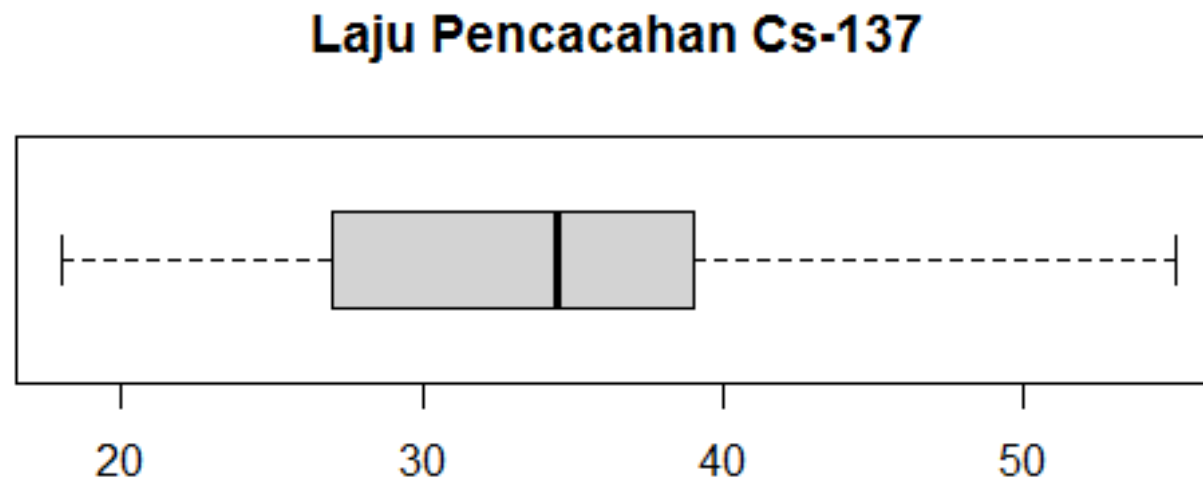
The decimal point is 1 digit(s) to the right of the |

1		89
2		1344
2		679
3		002344
3		55556799
4		03344
4		6
5		
5		5

Diagram Kotak Titik (*Boxplot*)

Merupakan salah satu grafik yang paling sering dipakai. Grafik ini dapat memperlihatkan data yang memiliki nilai ekstrim dan dapat langsung melihat nilai kuartilnya.

```
# 5. boxplot  
# boxplot(data, horizontal=T, main="judul")  
boxplot(Cesium$V1, horizontal=T, main="Laju Pencacahan Cs-137")
```

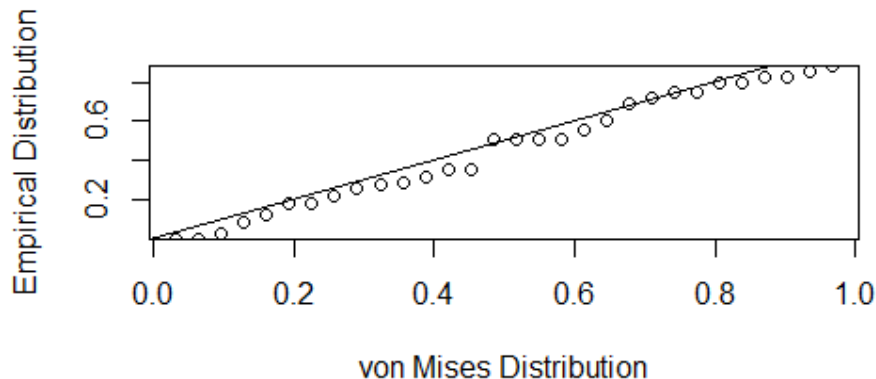


PP Plot dan QQ Plot

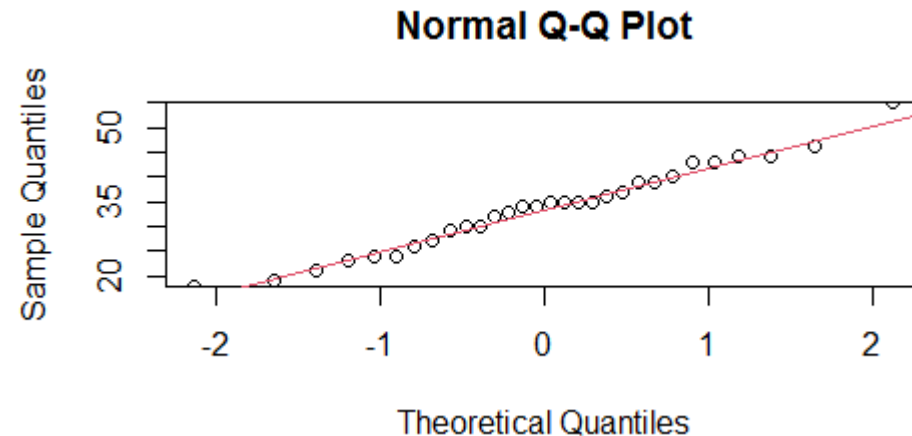
Kedua plot ini digunakan untuk mengidentifikasi kenormalan distribusi data. **QQ Plot** menganalisis plot grafik antara variabel *quantile* dengan *quantile* setiap anggotanya. **PP Plot** menganalisis plot grafik antara variabel proporsi kumulatif dengan proporsi setiap anggotanya.

Install paket "CircStats"

```
# 6. pp plot
library(CircStats)
pp.plot(Cesium$V1, ref.line=TRUE)
mu      kappa
4.726661 0.2101425
```



```
# 6. qq plot
qqnorm(Cesium$V1)
qqline(Cesium$V1, col=2)
```



Tim Penyusun



Dr. Utriweni Mukhaiyar

Dosen KK Statistika

Kepala Laboratorium Statistika dan Komputasi Statistika



Fatia Amalia, S.Si

Asisten KK Statistika

Pengajar Semester I – 2020/2021



Dr. Udjianna S. Pasaribu

Dosen KK Statistika, MA2181 Analisis Data



Dr. Rr. Kurnia Novita Sari

Dosen KK Statistika, MA2181 Analisis Data



Dr. Sandy Vantika

Dosen KK Statistika,

MA2181 Analisis Data / MA2081 Statistika Dasar



Dr. Sapto Wahyu Indratno

Dosen KK Statistika, MA2082 Biostatistika



Yuli Sri Afrianti, S.Si., MT, MBA.

Dosen KK Statistika,

MA2181 Analisis Data / MA2081 Statistika Dasar



Dr. Utriweni Mukhaiyar

Dosen KK Statistika, MA2082 Biostatistika



Referensi

- KK Statistika ITB. 2019. *Modul Statistika Dasar*. Bandung.

Selamat Praktikum!