

# Modul Eksplorasi Data

oleh

Jimmy Ludin

Diklat Fungsional Statistisi Tingkat Ahli  
Angkatan 20 Tahun 2019

1

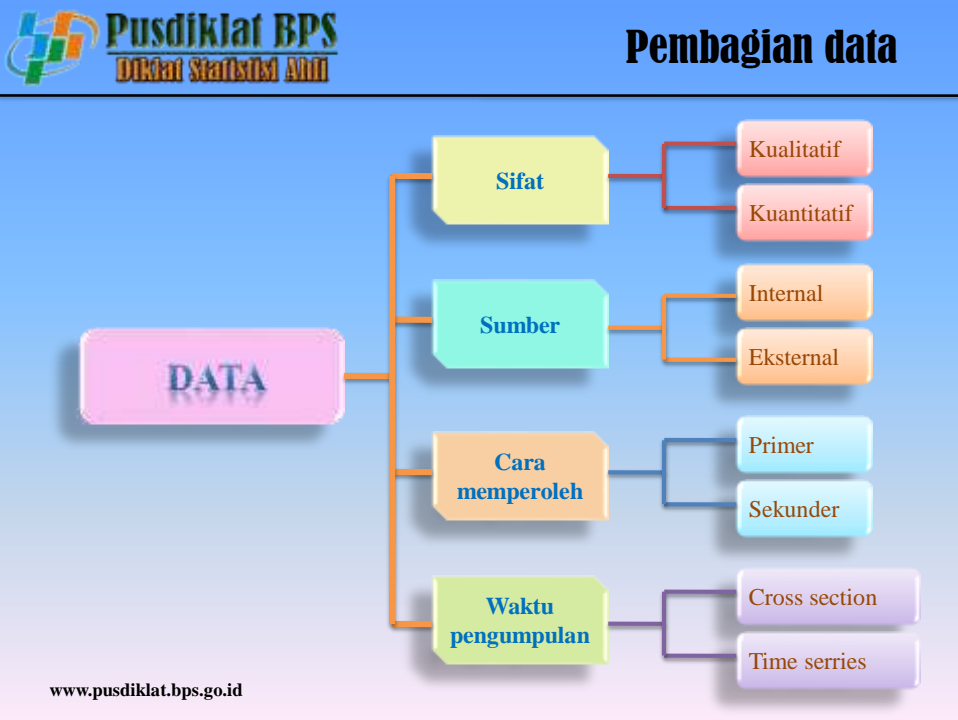
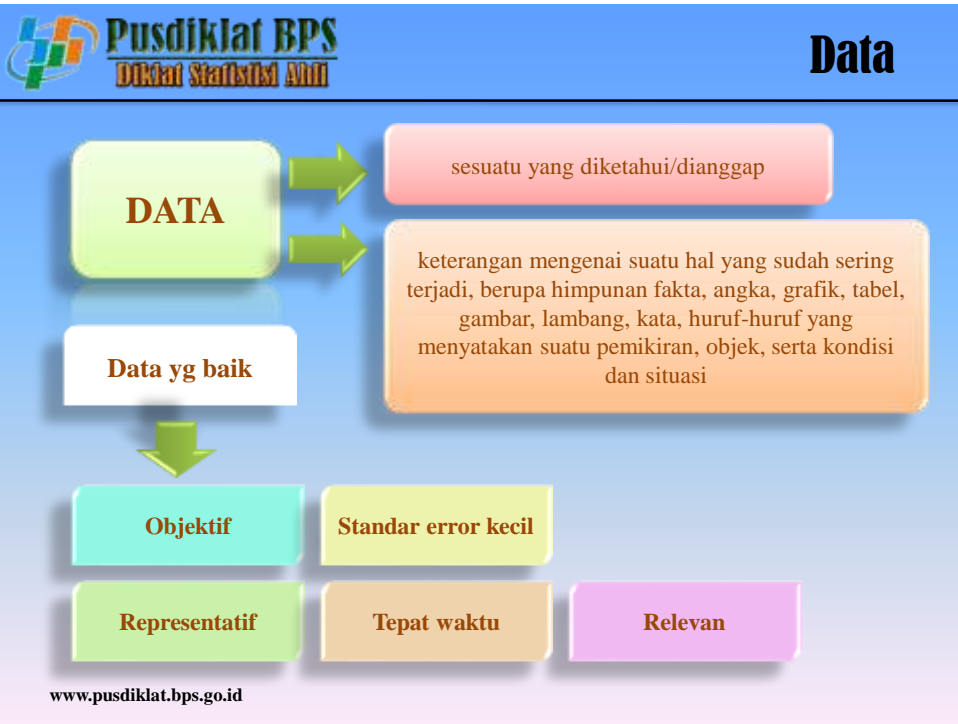
## IT'S About Me....

- Jimmy Ludin, SST., M.Si
  - Lahir di Pekanbaru
  - Statistika Ekonomi STIS (1996 – 2001)
- BPS Provinsi Papua (Feb 2002 – Mar 2008)
  - Pusdiklat BPS (Apr 2008)
- Magister Statistika Komputasi ITS (2009 – 2011)
  - Widyaiswara Muda (Mei 2012)
  - Email : jimmy.ludin@bps.go.id



## Pendahuluan







Tabel yg memuat keterangan mengenai satu karakteristik

Tabel 1. Jumlah peserta diklat fungsional statistisi ahli menurut umur tahun 2014

Umur	Jumlah peserta diklat
(2)	(1)
22	25
23	18
24	21
25	11
Jumlah	75

Tabel yg menghubungkan  
dua hal atau dua  
karakteristik

Tabel 2. Jumlah peserta orientasi mahasiswa baru Universitas Indonesia tahun 2015

Fakultas	Jawa	Luar jawa
(2)	(1)	(3)
Ekonomi	30	3
Ilmu Pendidikan dan keguruan	36	12
Ilmu Alam	24	4
Teknik	32	5
Hukum	35	4
Kedokteran	12	7
Jumlah	169	35

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

Tabel 3. Pertumbuhan volume dan nilai ekspor dan impor komoditas hortikultura tahun 1996-2002

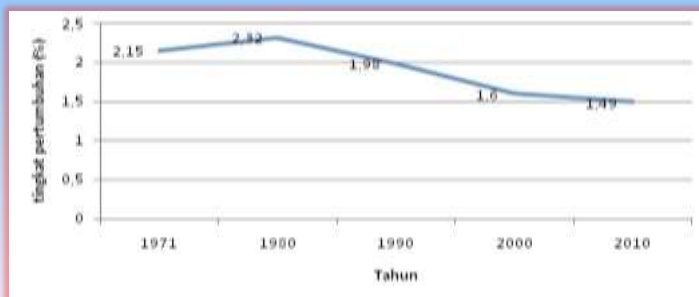
Komoditas	1996-1997		1997-1999		2000-2002	
	Volume	Nilai	Volume	Nilai	Volume	Nilai
Ekspor:						
Sayuran segar/dingin	-36,89	-44,27	-0,82	-8,16	-13,24	25,28
Impor:						
Sayuran segar/dingin	11,25	19,77	15,41	-9,01	8,45	16,18

Sumber : FAO, berbagai tahun (diolah)

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

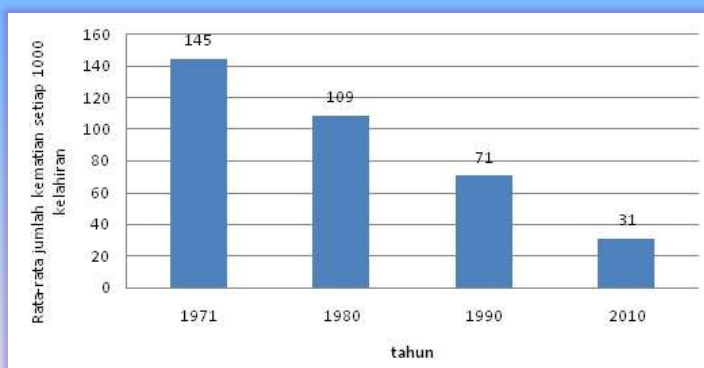
Untuk mengetahui perkembangan suatu hal atau kegiatan yang disajikan dalam bentuk garis

Gambar 1. Tingkat pertumbuhan penduduk Indonesia tahun 1971-2010



[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

Gambar 2. Jumlah kematian bayi tahun 1971-2010



[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

Grafik berupa lingkaran, dimana luas lingkaran merupakan komponen dari beberapa nilai

Gambar 3. Umur remaja saat pertama kali merokok



[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

Merupakan grafik berupa peta

Gambar 4. Data kunjungan pasien Puskesmas Sukajaya bulan Januari-Juni tahun 2010



[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

Gambar 5. Penjualan motor Yamaha Jupiter di wilayah DKI Jakarta tahun 2010



[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)



[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)





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

$\bar{x}$  = Rata-rata

$\sum_{i=1}^n x_i$  = Jumlah data

$n$  = Banyaknya data

$i$  = 1, 2, ..., n

### Digunakan Jika:

Jenis datanya numerik (interval/rasio)

Sebaran datanya simetrik

Tidak ada data pencilan (outlier)

Pengeluaran rata-rata perbulan dari 6 rumah tangga di suatu daerah adalah sebagai berikut

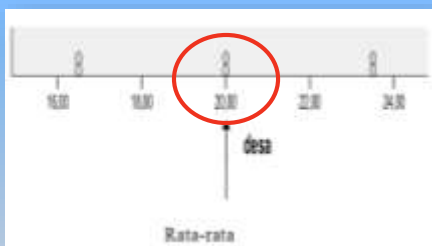
Daerah	1	2	3	4	5	6
Perdesaan A	2.000.000	2.300.000	1.600.000	2.000.000	2.400.000	1.700.000
Perkotaan B	800.000	5.000.000	700000	800.000	1.200.000	3.500.000

$$\bar{x}_A = \frac{\sum x}{n} = \frac{2+2,3+1,6+2,0+2,4+1,7}{6} = 2(\text{juta})$$

$$\bar{x}_B = \frac{\sum x}{n} = \frac{0,8+5+0,7+0,8+1,2+3,5}{6} = 2(\text{juta})$$

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

Gambaran data diatas adalah:



Gambar 1. Ilustrasi letak nilai rata-rata pada Perdesaan A



Gambar 2. Ilustrasi letak nilai rata-rata pada Perkotaan B

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

$$Me = Q_2 = \begin{cases} x_{\frac{n+1}{2}}, & \text{jika } n \text{ ganjil} \\ \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}, & \text{jika } n \text{ genap} \end{cases}$$

Me = Median

$$x_{\frac{n+1}{2}} = \text{Data ke } \frac{n+1}{2}$$

$$x_{\frac{n}{2}} = \text{Data ke } \frac{n}{2}$$

$$x_{\frac{n}{2}+1} = \text{Data ke } \frac{n}{2} + 1$$

www.pusdiklat.bps.go.id

## Digunakan Jika:

Rata - rata tak memenuhi syarat pada data berjenis interval/rasio

Data tidak simetrik

inferensia jika sebaran data tidak normal dengan selang kepercayaan

Tidak ada data pencilan (outlier)

## Contoh 2

Hitunglah median dari nilai ujian matematika kelas 3 SMU berikut ini:  
8; 4; 5; 6; 7; 6; 7; 7; 2; 9

**Mengurutkan data:** 2; 4; 5; 6; 6; 7; 7; 7; 8; 9 → **n = 10**

**posisi Me =  $\frac{1}{2}(10+1) = 5,5$  (antara data ke-5 dan ke-6)**

jadi **Median =  $\frac{1}{2} (6+7)$**

**= 6,5**

**(rata-rata dari 2 data yang terletak pada urutan ke-5 dan ke-6)**

www.pusdiklat.bps.go.id

**Digunakan:**

untuk tujuan deskriptif  
karena tidak  
mempertimbangkan  
sebaran data

Jika nilai pengamatan  
sangat bervariasi dari  
nilai pusatnya, modus  
tidak cocok untuk  
mengambarkan ukuran  
pemusatan

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

**Contoh 3**

Contohnya, seorang intelijen Negara member informasi pada kepolisian bahwa buronan yang dicari sering muncul bersama-sama antara tanggal 5-10 tiap bulan. Dalam satu bulan, buronan hanya muncul 1 kali bersama-sama untuk konsolidasi. Pihak kepolisian harus memutuskan sebuah tanggal untuk melakukan penggrebekan.

**Data tanggal setiap bulan mengenai kemunculan para buronan yang direkam selama 2 tahun adalah sebagai berikut:**

**6 5 5 5 6 6 9 5 5 7 8 5 7 5 7 5 5 7 7 5 6 5 10 5**

**Kemunculan nilai terbanyak : 5 (12 kali)**

**Dengan demikian, pada bulan ini pihak kepolisian akan melakukan penggerebekan terhadap para buronan tepat pada tanggal 5**

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

### Perbandingan rata-rata, median, dan modus

Ukuran Pemusatan	Kelebihan	Kekurangan
Rata-rata	<ol style="list-style-type: none"> <li>1. Mempertimbangkan semua nilai</li> <li>2. Menggambarkan mean populasi</li> <li>3. Variasinya stabil</li> <li>4. Data homogen</li> </ol>	<ol style="list-style-type: none"> <li>1. Mudah terpengaruh oleh nilai ekstrim</li> <li>2. Kurang baik untuk data heterogen</li> </ol>
Median	<ol style="list-style-type: none"> <li>1. Tidak terpengaruh oleh nilai ekstrim</li> <li>2. Cocok untuk data heterogen</li> </ol>	<ol style="list-style-type: none"> <li>1. Tidak mempertimbangkan semua nilai</li> <li>2. Kurang menggambarkan mean populasi</li> </ol>
Modus	<ol style="list-style-type: none"> <li>1. Tidak terpengaruh oleh nilai ekstrim</li> <li>2. Cocok untuk data homogen maupun heterogen</li> </ol>	<ol style="list-style-type: none"> <li>1. Kurang menggambarkan mean populasi</li> <li>2. Modus bias lebih dari satu</li> </ol>

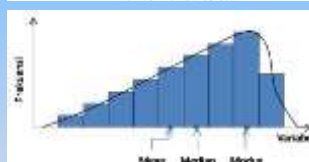
[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

### Perbandingan rata-rata, median, dan modus

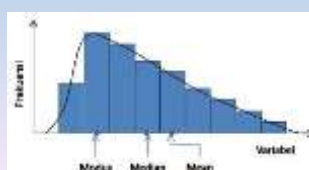
**Rata-rata = Median = Modus**



**Rata-rata < Median < Modus**



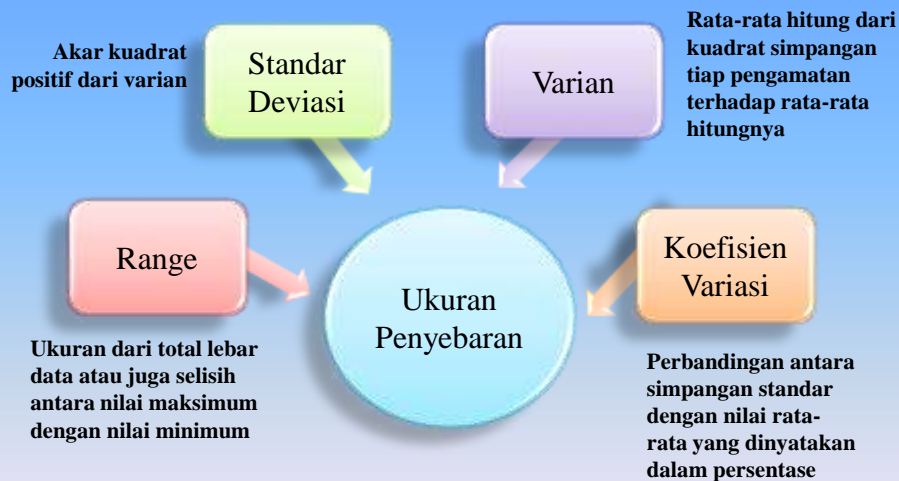
**Rata-rata > Median > Modus**



**Hubungan empiris antara rata-rata, median, dan modus:**

$$\bar{X} - \text{Mod} = 3(\bar{X} - \text{Med})$$

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)



## Range

$$\text{Range} = \text{nilai maksimum} - \text{nilai minimum}$$

Range tidak cocok digunakan bila ada nilai ekstrem

### Contoh 4

Carilah range data berikut!  
Pengeluaran rata-rata perbulan (dalam ratusan ribu) dari 6 rumah tangga di suatu daerah

Daerah	1	2	3	4	5	6
Pedesaan A	20	23	16	20	24	17
Perkotaan B	8	50	7	8	12	35

Dari contoh 1, didapatkan range pedesaan A = 8 dan range perkotaan B = 43.

## Populasi

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} \text{ atau } \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{N}}{N}}$$

## Sampel

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \text{ atau } \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}}$$

### Contoh 5

Diberikan data mengenai hasil perolehan nilai pada 2 Quiz yg berbeda, sebagai berikut ini :

	1	2	3	4	5	6	7	8	9	10	11
Quiz 1:	1	20	20	20	20	20	20	20	20	20	20
Quiz 2:	2	3	4	5	6	14	15	16	17	18	19

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

Quiz I: rata-rata = 18.27

Quiz 2: rata-rata = 10.82

No	Quiz 1 ( $x_i$ )	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	Quiz 2 ( $x_i$ )	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	1	-17.27	298.35	2	-8.82	77.76
2	20	1.73	2.98	3	-7.82	61.12
3	20	1.73	2.98	4	-6.82	46.49
4	20	1.73	2.98	5	-5.82	33.85
5	20	1.73	2.98	6	-4.82	23.21
6	20	1.73	2.98	14	3.18	10.12
7	20	1.73	2.98	15	4.18	17.49
8	20	1.73	2.98	16	5.18	26.85
9	20	1.73	2.98	17	6.18	38.21
10	20	1.73	2.98	18	7.18	51.58
11	20	1.73	2.98	19	8.18	66.94
	<b>Jumlah</b>		328.1818			453.6364

Quiz 1:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{328.18}{11-1}} = 5.73$$

$$\text{varian} = s^2 = 5.73^2 = 32.82$$

Quiz 2:

$$s = \sqrt{\frac{453.64}{11-1}} = 6.74$$

$$\text{varian} = s^2 = 6.74^2 = 45.36$$

Quiz ke-2 lebih bervariasi dibandingkan dengan Quiz ke-1.

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

$$C.V = \frac{s}{\bar{x}} \times 100\%$$

### Contoh 6

Perhatikan gugus data untuk Kelompok A dan Kelompok B

A	2	4	5	6	6	7	7	7	8	9
B	3	6	7	9	9	10	10	10	11	12

Jawab:

Kelompok A: Rata-rata = 6.1; s = 2.0

Kelompok B: Rata-rata = 8.7; s = 2.7

$$CV = \frac{s}{\bar{x}} \times 100\% = \frac{2.0}{6.1} \times 100\% = 33.2\%$$

$$CV = \frac{s}{\bar{x}} \times 100\% = \frac{2.7}{8.7} \times 100\% = 30.7\%$$

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

untuk melihat  
ketidaksimetrisan  
data kuantitatif  
yang berskala besar

Kuantil

Box Plot

representasi grafik  
dari sekelompok  
data yang memuat 5  
ringkasan data:  
median, Q1, Q3,  
minimum dan  
maksimum

Steam and  
Leaf

teknik untuk  
menggambar pola  
sebaran bagi data yang  
berukuran kecil

Menggunakan  
Software

Eksplorasi  
Data

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)



Untuk menggambarkan pola sebaran data berukuran kecil

### Contoh 7

Data pengeluaran rumah tangga di suatu daerah untuk 44 rumah tangga (dalam ratusan ribuan) adalah sebagai berikut:

47, 11, 46, 33, 19, 42, 27, 22, 62, 10, 44, 2, 15, 21, 67, 20, 26, 25, 6, 53, 18, 3, 30, 7, 21, 25, 20, 40, 16, 8, 4, 10, 46, 31, 14, 15, 8, 10, 19, 17, 12, 16, 42, 16

Batang ← 45 → Daun

Batang	Daun
0	2 6 3 7 8 4 8
1	1 9 0 5 8 6 0 4 5 0 9 7 2 6 6
2	7 2 1 0 6 5 1 5 0
3	3 0 1
4	7 6 2 4 0 6 2
5	3
6	2 7

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

10, 15, 16, 17, 18, 20, 21, 21, 22, 22, 23, 25, 26, 27, 30, 32, 44, 53, 67

$Q_1$

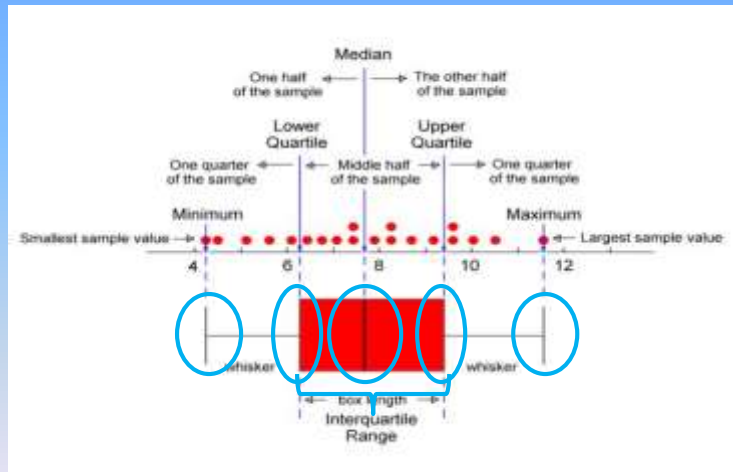
10, 15, 16, 17, 18, 20, 21, 21, 22, 22, 23, 25, 26, 27, 30, 32, 44, 53, 67

$Q_2$

10, 15, 16, 17, 18, 20, 21, 21, 22, 22, 23, 25, 26, 27, 30, 32, 44, 53, 67

$Q_3$

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

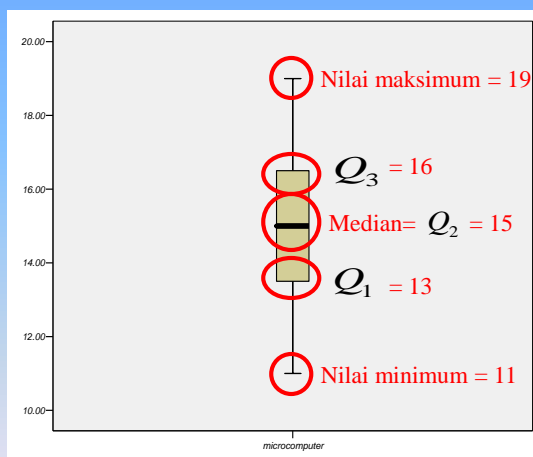


[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

## Contoh 8

Berikut 20 data penelitian tentang penggunaan microcomputer selama seminggu (dalam jam) oleh mahasiswa pada jurusan matematika di suatu perguruan tinggi:

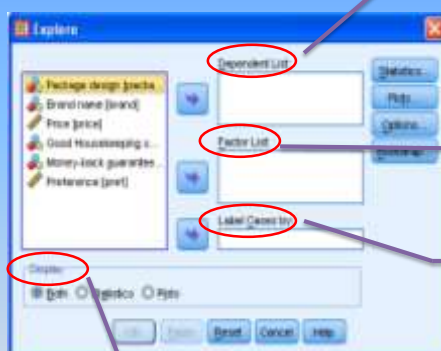
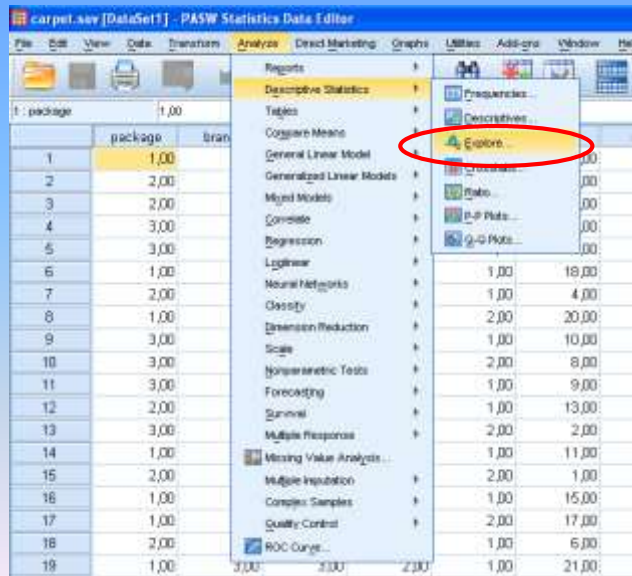
12, 16, 12, 13, 16, 14, 15,  
15, 16, 17, 18, 14, 18, 19, 11,  
15, 13, 15, 17, 14



[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

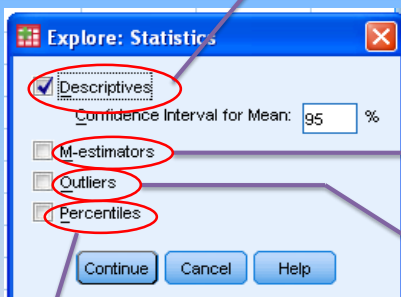
- 1 **Aktifkan data**
- 2 **Analyze**
- Descriptive statistics**
- Eksplora**

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)



[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

## Statistics



menampilkan output mean, median, modus, 5% trimmed mean, standar error, varians, standar deviasi, minimum, maksimum, range, interquartile range, skewness, standar error skewness, kurtosis dan standar error kurtosis

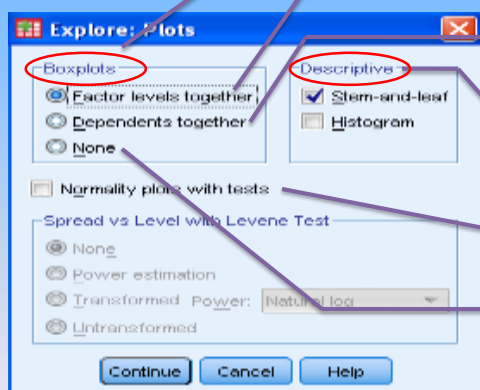
menampilkan output robust maximum-likelihood estimators of central tendency

menampilkan output lima data terkecil dan lima data terbesar. Pada outputnya akan ditampilkan nilai extreme

akan menampilkan output persentil: 5, 10, 25, 50, 75, 90, 95 dan Tukey's hinges

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

## Plots



organisasi output boxplot

boxplot dikelompokkan berdasarkan faktor

boxplot dikelompokkan berdasarkan dependent variabel untuk faktor yang sama

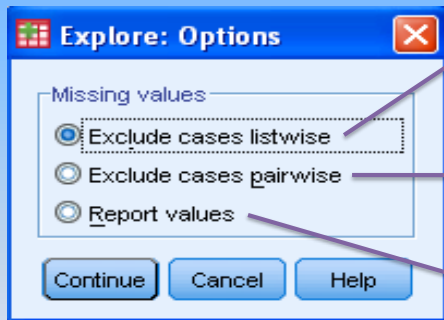
menampilkan plot descriptive

menampilkan output uji kenormalan

tidak menampilkan boxplot

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

**Option**



missing value tidak  
diikutkan dalam analisis

output menyertakan hasil  
analisis dengan missing value  
dan tidak dengan missing  
value

mendefinisikan missing  
value sebagai data  
tersendiri

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

**Contoh 8**

Perhatikan contoh data berikut:

47, 11, 46, 33, 19, 42, 27, 22, 62, 10, 44, 2, 15, 21, 67, 20, 26, 25, 6, 53, 18, 3, 30, 7,  
21, 25, 20, 40, 16, 8, 4, 10, 46, 31, 14, 15, 8, 10, 19, 17, 12, 16, 42, 16

**Explore**

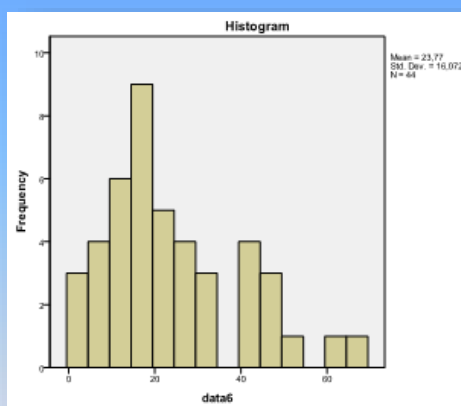
Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
data6	44	100,0%	0	,0%	44	100,0%

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

Descriptives		
	Statistic	Std. Error
data6 Mean	23,77	2,423
95% Confidence Interval for Mean	18,89	
Lower Bound	28,66	
Upper Bound	22,74	
5% Trimmed Mean	19,50	
Median	258,319	
Variance	16,072	
Std. Deviation	2	
Minimum	67	
Maximum	65	
Range	21	
Interquartile Range	,944	,357
Skewness	,250	,702
Kurtosis		

Percentiles								
		Percentiles						
		5	10	25	50	75	90	95
Weighted Average(Definition 1)	data6	3,25	6,50	11,25	19,50	32,50	46,50	59,75
Tukey's Hinges	data6			11,50	19,50	32,00		

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)



#### Stem-and-Leaf Plot

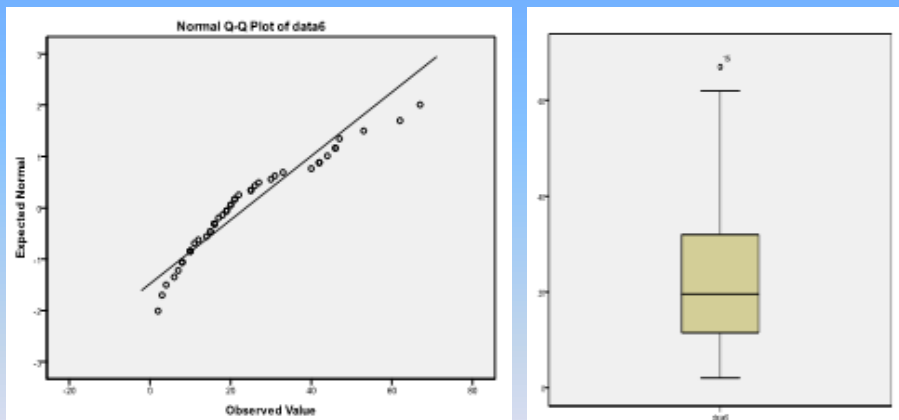
```

Frequency  Stem & Leaf
 7,00      0 . 2346788
15,00      1 . 000124556667899
 9,00      2 . 001125567
 3,00      3 . 013
 7,00      4 . 0224667
 1,00      5 . 3
 1,00      6 . 2
 1,00      Extremes  (>=67)

```

Stem width: 10  
Each leaf: 1 case(s)

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)



[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)

**Terima kasih**

[www.pusdiklat.bps.go.id](http://www.pusdiklat.bps.go.id)