

Statistika dan Probabilitas

Dhita Diana Dewi, M.Stat.

Kontrak Perkuliahan

Tata Tertib Kelas

- Toleransi Keterlambatan 15 menit
- Tidak ada toleransi untuk TITIP ABSEN
- Baju rapi, tidak ketat, dan memakai sepatu
- Presensi 80%
- Wajib menjaga kebersihan dan ketenangan kelas
- Dilarang makan, minum, merokok, menggunakan headset, dll. yang bersifat mengganggu ketenangan kelas
- Tugas dikumpulkan harus tepat waktu
- Tugas hasil plagiarisme, maka dinilai nol
- QUIZ tanpa pemberitahuan terlebih dahulu

Komponen Penilaian

Item	Bobot
Presensi	10%
Quiz	15%
Tugas	15%
UTS	30%
UAS	30%
TOTAL	100%

Statistika dan Probabilitas (3 SKS)

UTS : DASAR-DASAR ILMU STATISTIKA, NILAI-2 STATISTIK,
PENGANTAR PELUANG

UAS : SAMPLING, PENAKSIRAN PARAMETER, PENGUJIAN
HIPOTESIS dan ANALISIS VARIANS.

Refferensi :

1. Walpole, G.R&Myers, R.H. 1986

Ilmu peluang dan statistika untuk insinyur dan ilmuwan,

2. Sudjana, 2002, Metoda Statistika

3. Spiegel,M.R 1984 THEORY AND PROBLEMS OF STATISTICS

Chapter 1

Pengertian Dasar & Statistika Deskriptif

Apa itu Statistik?

Statistik dan Statistika

- STATISTIKA ?

Ilmu yang membahas cara-cara pengumpulan data, pengolahan data (analisis), dan penarikan kesimpulan berdasarkan kumpulan data dan analisis yang dilakukan. (Sudjana, 2002).

- STATISTIK?

Catatan angka-angka (bilangan) yang dikumpulkan, ditabulasi, digolongkan, sehingga dapat memberikan informasi yang berarti mengenai masalah atau gejala (KBBI)

1

Descriptive Statistics

- Penyajian Data
- Pengelompokan Data dan Histogram
- Measures of Central Tendency (Ukuran Gejala Pusat)
- Measures of Variability
- Skewness and Kurtosis
- Hubungan antara Mean dan Standard Deviation

1

LEARNING OBJECTIVES

Setelah mempelajari bagian ini diharapkan dapat :

- Membedakan *qualitative data* dan *quantitative data*.
- Menjelaskan Skala Pengukuran: nominal, ordinal, interval, dan ratio .
- Menjelaskan perbedaan antara : populasi dan sampel.
- Menghitung dan menginterpretasikan kuantil.
- Menghitung dan menjelaskan Ukuran Gejala Pusat (*measures of central tendency*)
- Menyajikan data melalui tabel dan grafik.
- Menggunakan Excel templates untuk menghitung berbagai ukuran statistik dan membuat grafik.

WHAT IS STATISTICS?

- **Statistics** adalah suatu ilmu yg membantu kita untuk membuat keputusan yg terbaik.
- **Statistics** mengajari kita bagaimana untuk meringkas, menganalisis, dan membuat kesimpulan berdasarkan data sehingga kita dapat membuat keputusan .

Pembagian Statistika

- **Descriptive Statistics**

- ✓ Mengumpulkan
- ✓ Mengorganisir
- ✓ Meringkas
- ✓ Menyajikan
- ✓ Menganalisis

- **Inferential Statistics**

- ✓ Memprediksi dan meramalkan nilai parameter populasi
- ✓ Pengujian Hipotesis untuk nilai parameter populasi
- ✓ Membuat kesimpulan

Types of Data

- **Qualitative -
Categorical atau
Nominal:**

Contoh:

- ✓ **Warna**
- ✓ **Gender**
- ✓ **Suku bangsa**

- **Quantitative -
Measurable atau
Countable
(numerical):**

Contoh:

- ✓ **Temperatur**
- ✓ **Besar Gaji**
- ✓ **Skor Ujian**

Skala Pengukuran

- **Skala Nominal**
- **Skala Ordinal**
- **Skala Interval**
- **Skala Ratio**

Sampel dan Populasi

- **Populasi** berisi suatu set seluruh pengukuran dari karakteristik yang diteliti
- **Sample** adalah bagian dari pengukuran yang dipilih dari populasinya
- **Sensus** adalah pengamatan terhadap keseluruhan obyek penelitian dalam populasinya tanpa ada yang terlewat

Sampel dan Populasi



Populasi (N)



Sampel (n)

Tabel Distribusi Frekuensi

- Penyajian Data secara dikelompokkan dalam kelas interval
 - Jumlah observasi
 - N untuk Populasi
 - n untuk sampel
- Kelas **midpoint** atau Titik Tengah merupakan nilai tengah dari tiap kelas interval
- **Relative frequency** adalah persentase dari tiap frekuensi dari tiap kelas interval
 - ✓ Jumlah dari relative frequencies = 1

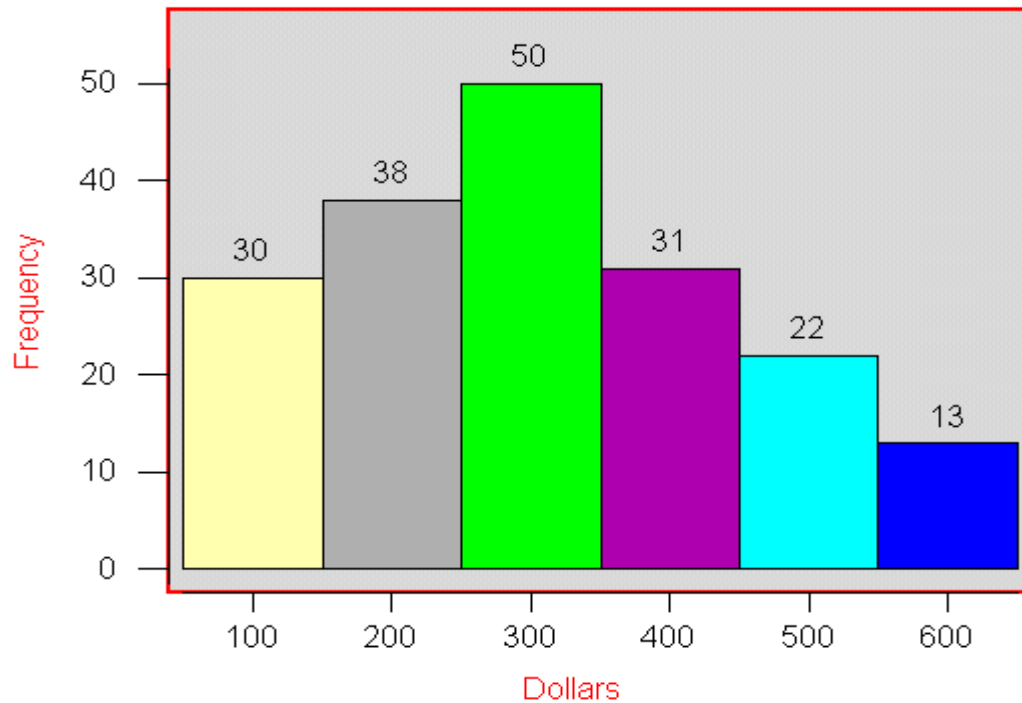
Contoh: Tabel Distribusi Frekuensi

x	f(x)	f(x)/n
Spending Class (\$)	Frequency (number of customers)	Relative Frequency
0 to less than 100	30	0.163
100 to less than 200	38	0.207
200 to less than 300	50	0.272
300 to less than 400	31	0.168
400 to less than 500	22	0.120
500 to less than 600	13	0.070
	<hr/>	<hr/>
	184	1.000

- Example of relative frequency: $30/184 = 0.163$
- Sum of relative frequencies = 1

Contoh Histogram

Frequency Histogram



Contoh Histogram

Relative Frequency Histogram



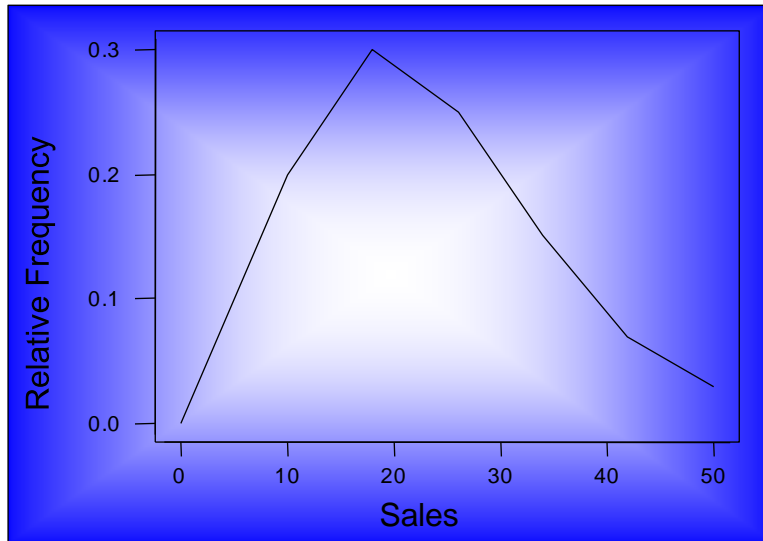
Cumulative Frequency Distribution

x	F(x)	F(x)/n
Spending Class (\$)	Cumulative Frequency	Cumulative Relative Frequency
Kurang dari 100	30	0.163
Kurang dari 200	68	0.370
Kurang dari 300	118	0.641
Kurang dari 400	149	0.810
Kurang dari 500	171	0.929
Kurang dari 600	184	1.000

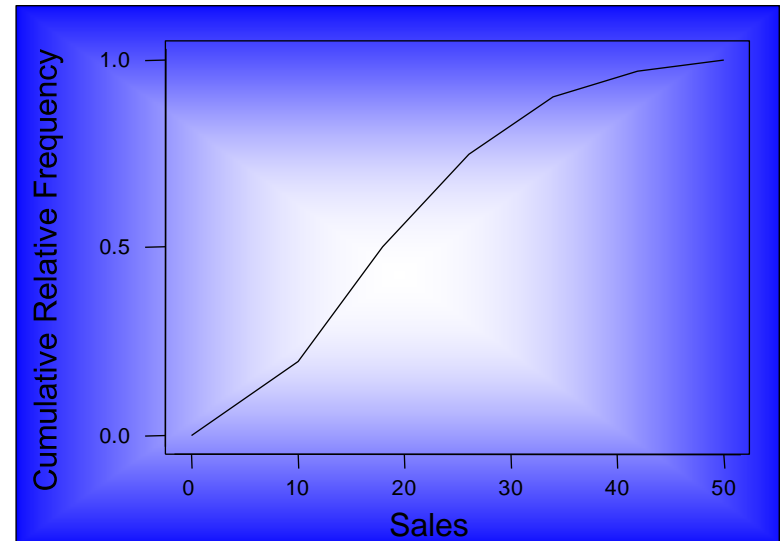
The **cumulative frequency** of each group is the sum of the frequencies of that and all preceding groups.

Frequency Polygon and Ogive

Relative Frequency Polygon



Ogive



(Cumulative frequency or
relative frequency graph)

Langkah-2 Menyajikan Data dalam Dist.Frekuensi:

1. Tentukan Range
2. Tentukan banyak kelas Interval k
3. Tentukan Panjang Kelas Interval = $p = i$
Tentukan Kelas interval pertama

Contoh :

Data di bawah ini menunjukkan masa hidup dari 30 lampu pijar yang dipilih secara acak dari pabrik lampu pijar “G”

43	54	51	12	10	36	45	51	43
23	24	36	41	45	48	50	61	76
78	54	19	21	32	38	40	38	50
28	25	36	(dalam puluhan jam)					

Sajikan data di atas dalam : Tabel Distribusi Frekuensi ~ Tabel Distribusi Frekuensi Relatif- Tabel Distribusi Frekuensi Kumulatif – Tabel Distribusi Frekuensi Kumulatif Relatif

Sajikan Pula Dalam Grafik : Histogram – Polygon ~ Ogive

Apa yang terlihat pada Diagram Batang Daun ?

Berikan interpretasi untuk semua hasil di atas.

Catatan:

Dari polygon dapat terlihat apakah model sampel/populasinya miring atau simetris

Summary Measures: Population Parameters Sample Statistics

- Measures of Central Tendency

- ✓ Median
- ✓ Modus
- ✓ Mean

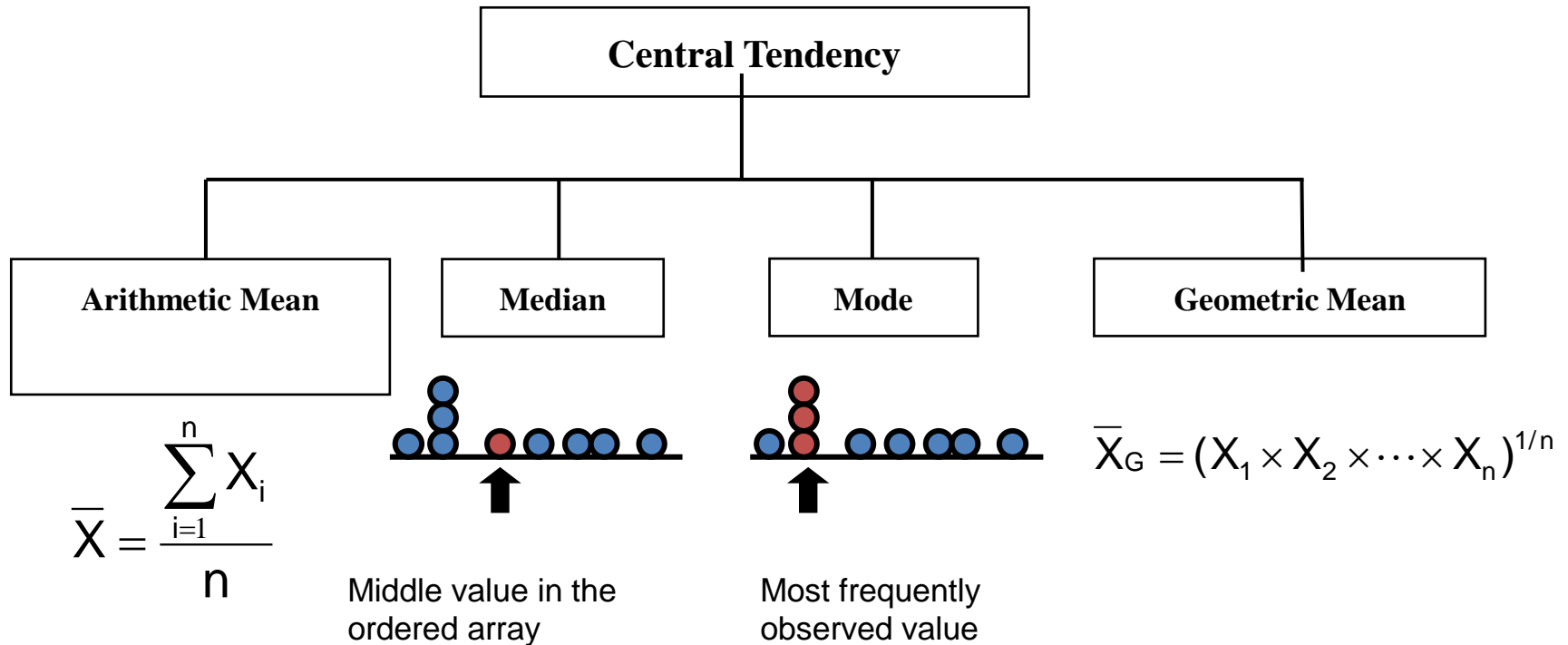
- Measures of Variability

- ✓ Range
- ✓ Interquartile range
- ✓ Variance
- ✓ Standard Deviation

- Other summary measures:

- ✓ Skewness
- ✓ Kurtosis

Measures of Central Tendency Summary



Contoh:

Misalkan diperoleh data waktu penyelesaian sejenis pekerjaan oleh 13 orang pegawai divisi A dan 12 pegawai divisi B yang dipilih secara acak (dalam menit):

Divisi A

5,5,5, 6,7,8,9,6, 4,3,4,1,2

Divisi B

5,7,8,6, 5,10,9,4,3,2,1, 5

1. Buatlah diagram titiknya
2. Bandingkan berdasarkan nilai : mean, median, modus, RAK, simpangan baku, kemiringan kurva
3. Jika ada seorang pegawai mengerjakan pekerjaan tersebut dalam tempo 7.5 menit bagaimana posisi dia jika masuk di Divisi A atau B?
4. CV

Arithmetic Mean or Average

Mean dari satu set observasi adalah nilai dari jumlah observasi dibagi dengan banyaknya observasi.

Population Mean

$$\mu = \frac{\sum_{i=1}^N x}{N}$$

Sample Mean

$$\bar{x} = \frac{\sum x}{n}$$

Ukuran Variabilitas (Dispersion)

- **Range**
 - ✓ Perbedaan antara nilai terbesar dengan nilai terkecil
- **Interquartile Range**
 - ✓ Perbedaan antara Quartil atas dengan quartil bawah ($Q_3 - Q_1$)
- **Variance**
 - ✓ Rata – rata kuadrat jarak dari data dengan mean
- **Standard Deviation**
 - ✓ Akar dari variance

Variance and Standard Deviation

Population Variance

$$\begin{aligned}\sigma^2 &= \frac{\sum_{i=1}^N (x - \mu)^2}{N} \\ &= \frac{\sum_{i=1}^N x^2 - \frac{\left(\sum_{i=1}^N x\right)^2}{N}}{N} \\ \sigma &= \sqrt{\sigma^2}\end{aligned}$$

Sample Variance

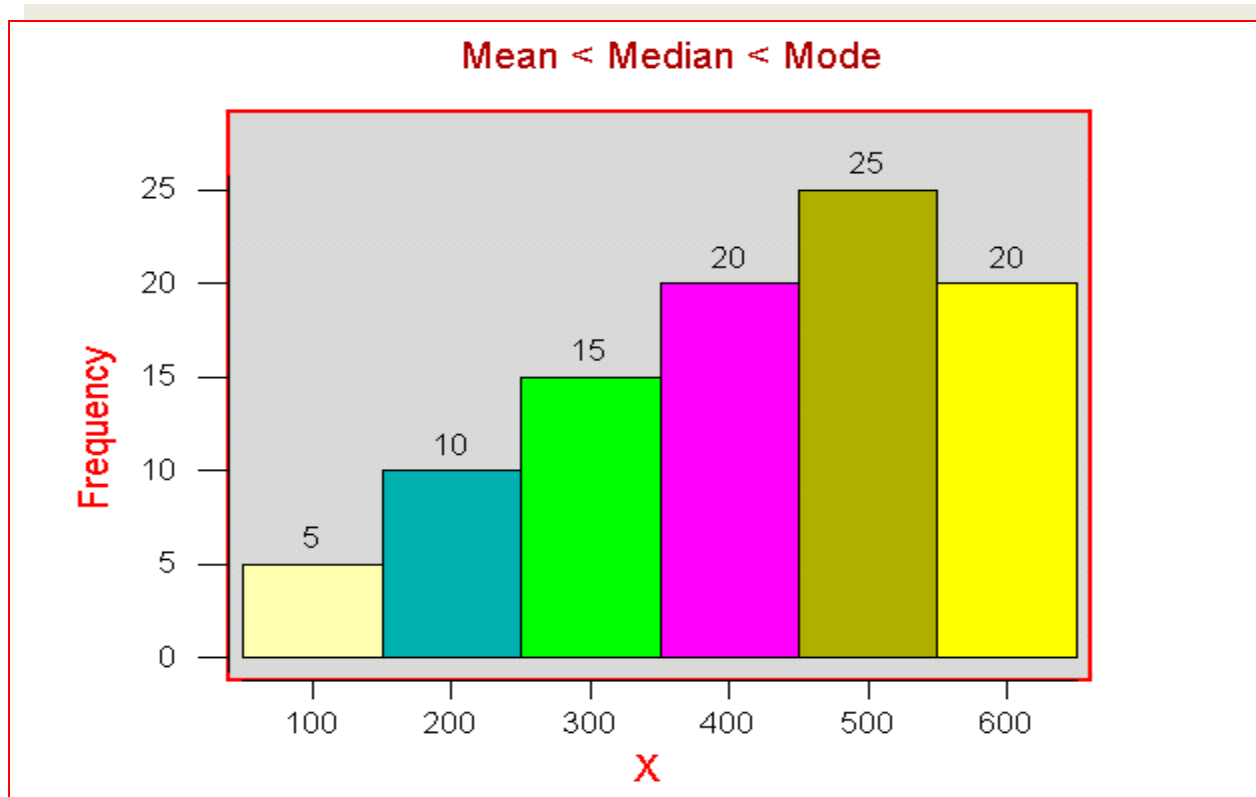
$$\begin{aligned}s^2 &= \frac{\sum_{i=1}^n (x - \bar{x})^2}{(n - 1)} \\ &= \frac{\sum_{i=1}^n x^2 - \frac{\left(\sum_{i=1}^n x\right)^2}{n}}{(n - 1)} \\ s &= \sqrt{s^2}\end{aligned}$$

Skewness and Kurtosis

- **Skewness** (Measure of asymmetry of a frequency distribution)
 - Skewed to left
 - Symmetric or unskewed
 - Skewed to right
- **Kurtosis** (Measure of flatness or peakedness of a frequency distribution)
 - **Platykurtic** (relatively flat)
 - **Mesokurtic** (normal)
 - **Leptokurtic** (relatively peaked)

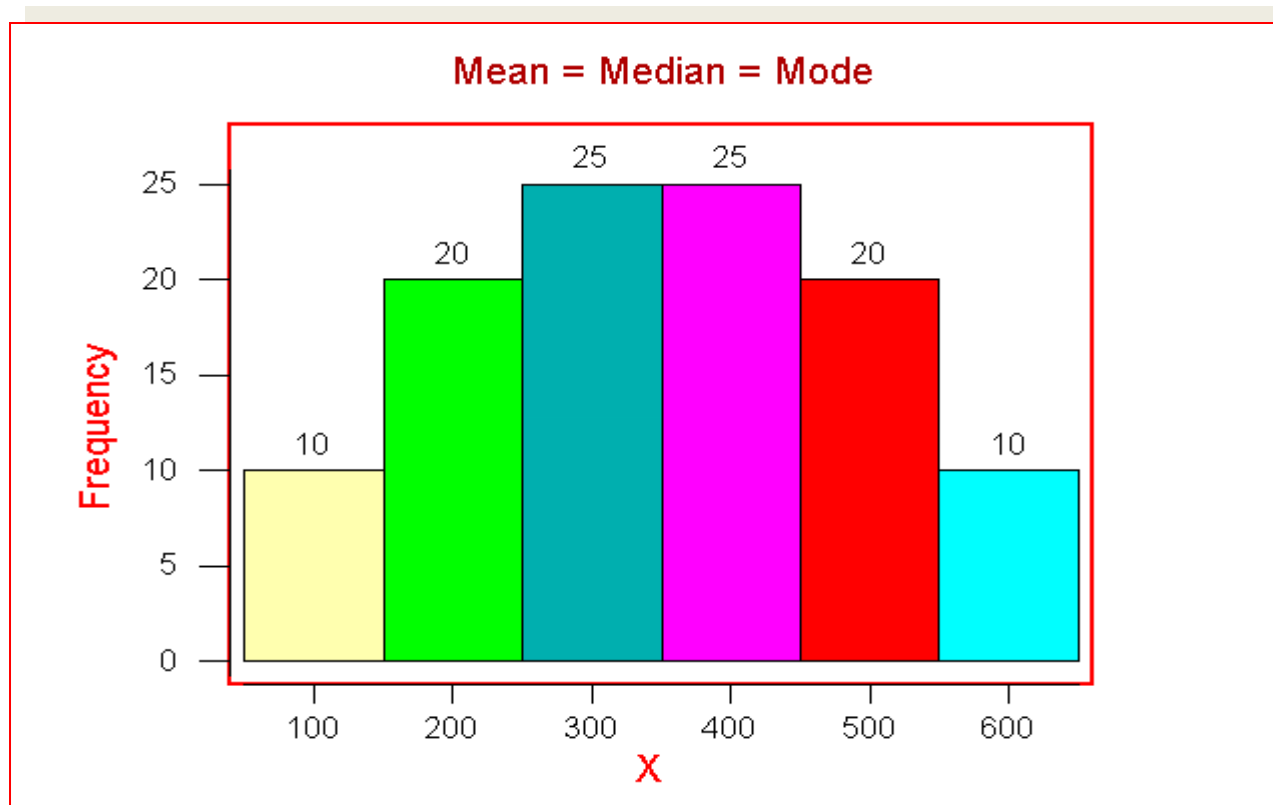
Skewness

Skewed to left



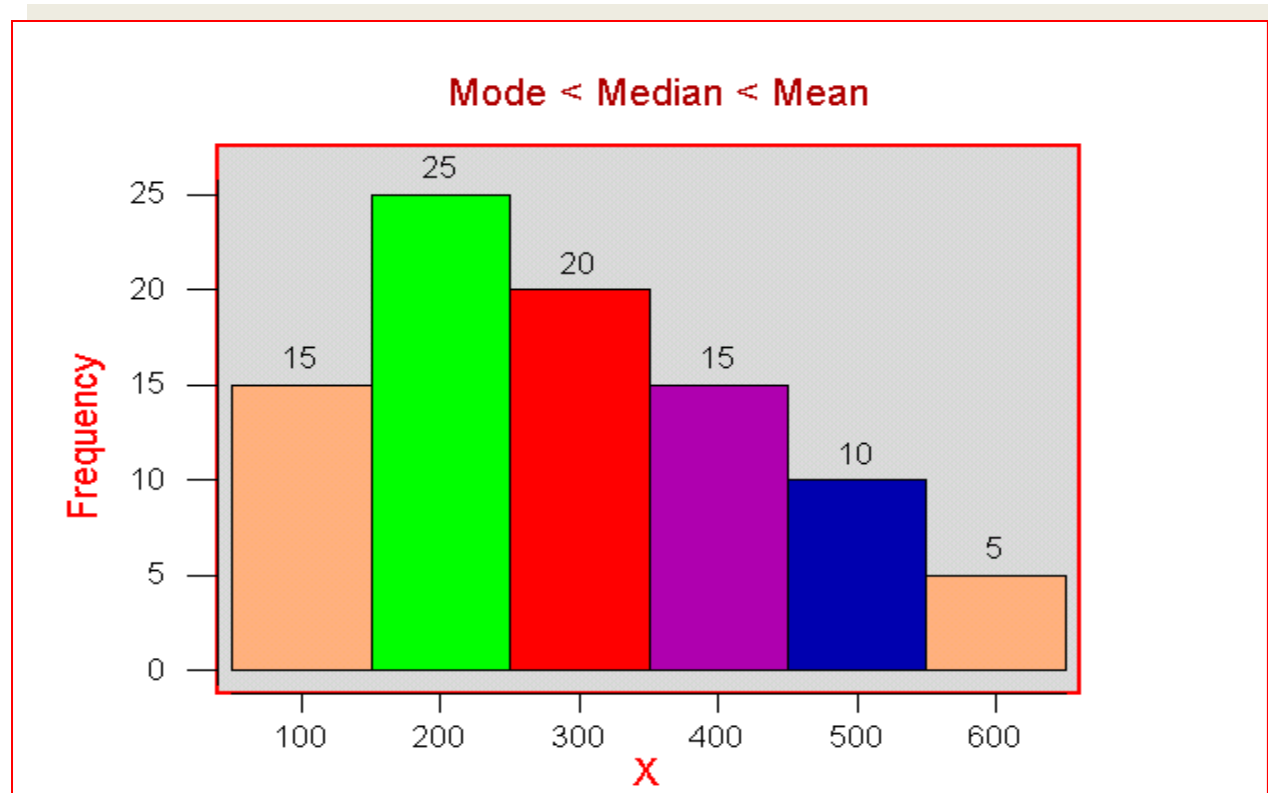
Skewness

Symmetric



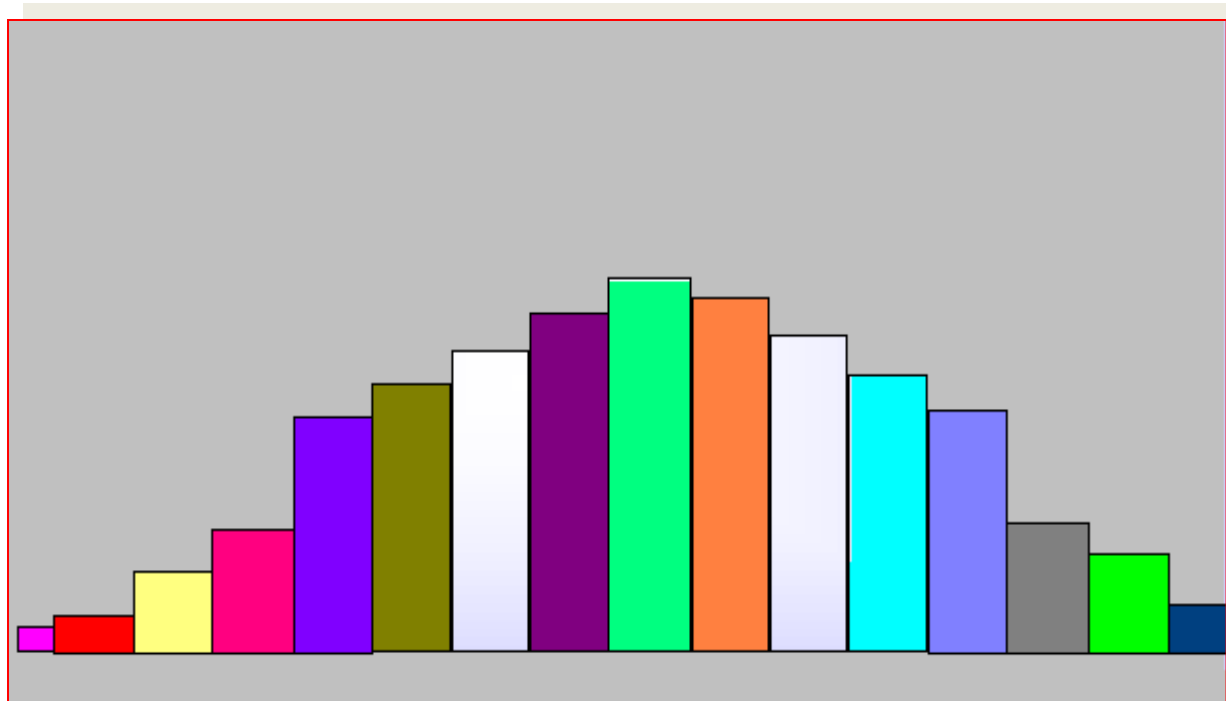
Skewness

Skewed to right



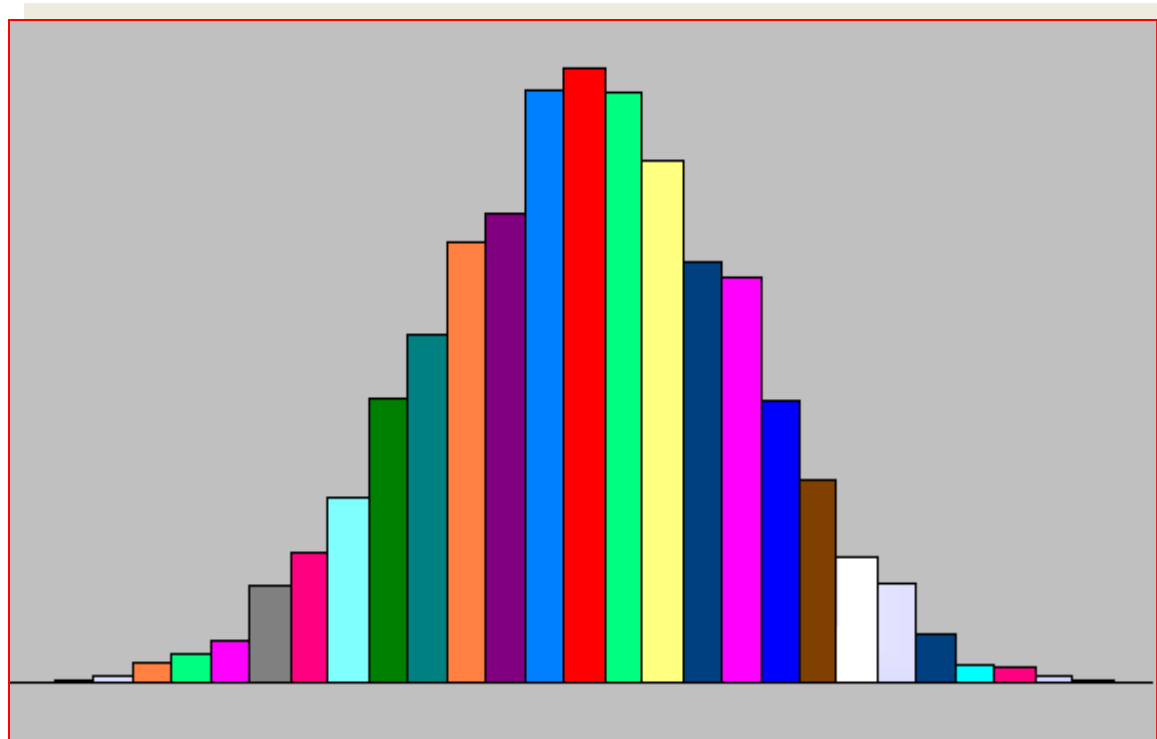
Kurtosis

Platykurtic - flat distribution



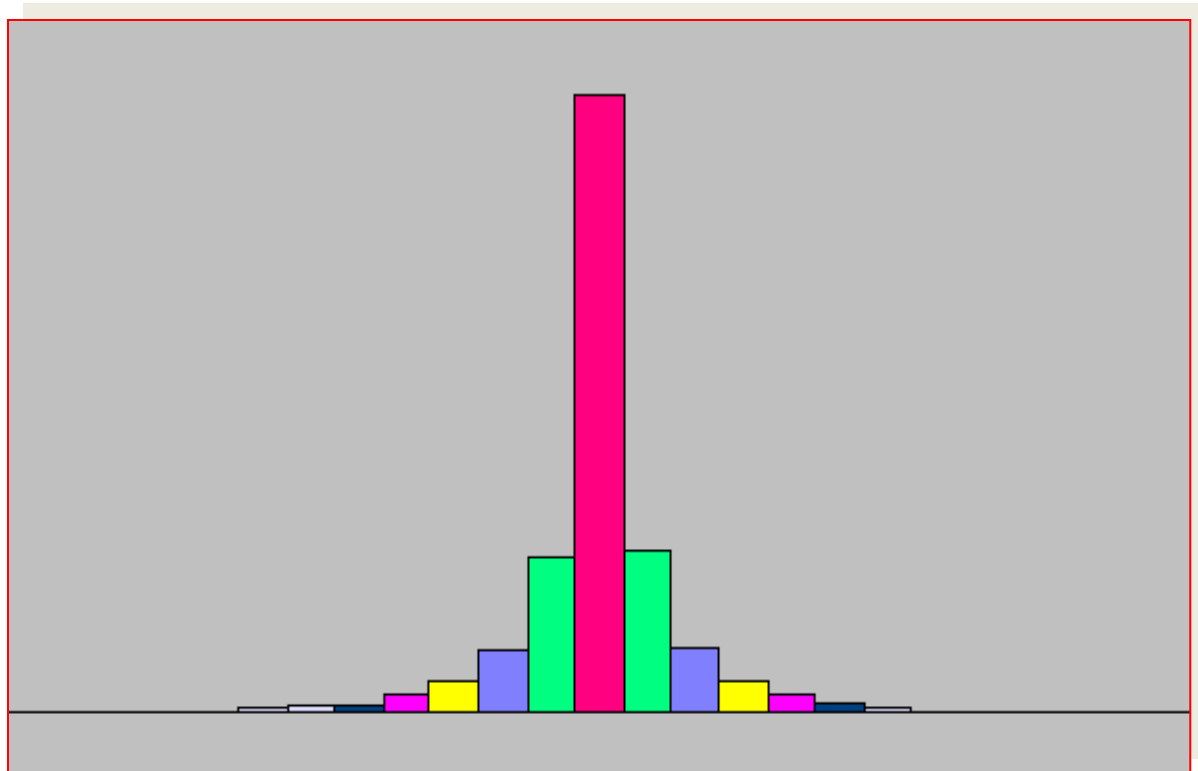
Kurtosis

Mesokurtic - not too flat and not too peaked



Kurtosis

Leptokurtic - peaked distribution



Hubungan antara Mean dan Standard Deviation

- **Chebyshev's Theorem**

- ✓ Applies to **any** distribution, regardless of shape
- ✓ Places lower limits on the percentages of observations within a given number of standard deviations from the mean

- **Empirical Rule**

- ✓ Applies only to roughly **mound-shaped** and **symmetric** distributions
- ✓ Specifies approximate percentages of observations within a given number of standard deviations from the mean

Chebyshev's Theorem

Paling sedikit $\left(1 - \frac{1}{k^2}\right)$ dari elemen/observasi
suatu distribusi berada dalam **k** standard
deviations dari mean

**At
least**

$1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} = 75\%$
$1 - \frac{1}{3^2} = 1 - \frac{1}{9} = \frac{8}{9} = 89\%$
$1 - \frac{1}{4^2} = 1 - \frac{1}{16} = \frac{15}{16} = 94\%$

**Lie
within**

2
3
4

**Standard
deviations**
of the mean

Empirical Rule

68%	Lie within	1 standard deviation of the mean
95%		2 standard deviations of the mean
All		3 standard deviations of the mean

Soal:

30 ibu rumah tangga yang dipilih secara acak dari sebuah perumahan P ditanya tentang besar pengeluaran bulan januari 2009 , jawaban mereka adalah (dalam ratusan ribu rupiah) :

30 40 35 25 50 55 45 38 35 40
39 45 38 39 42 43 48 45 38 41
20 25 28 32 35 34 35 28 25 28

Jawaban 60 ibu rumah tangga yang tinggal di perumahan Q adalah :

Pengeluaran	Jumlah ibu Rumah Tangga
15 - < 30	8
30 - < 40	12
40 - < 45	20
45 - < 50	11
50 - < 60	9

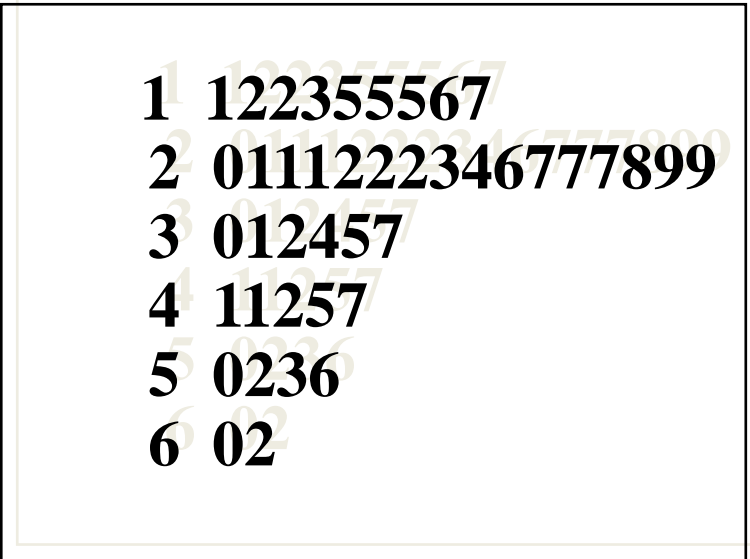
- a. Di perumahan manakah pengeluaran ibu Rumah tangga yang lebih bervariasi ?
- b. Di daerah manakah pengeluaran ibu rumah tangga yang memiliki kurva positif ? Jelaskan artinya !
- c. Tentukan ada berapa persen ibu rumah tangga yang pengeluarannya berada dalam dua simpangan baku dari rata – ratanya .
- d. Jika ada seorang ibu rumah tangga membelanjakan uang nya sebesar Rp. 4.500.000 , maka bagaimanakah posisinya jika dia berasal dari perumahan P atau Q ?
- e. Sajikan diagram batang daun untuk pengeluaran Ibu rumah tangga dari perumahan P.
- f. Gambarkan variasi untuk 50% pengeluaran dari dua kelompok tersebut kemudian jelaskan.

1-9 Exploratory Data Analysis - EDA

Techniques to determine relationships and trends, identify outliers and influential observations, and quickly describe or summarize data sets.

- **Stem-and-Leaf Displays**
 - ✓ Quick-and-dirty listing of all observations
 - ✓ Conveys some of the same information as a histogram
- **Box Plots**
 - ✓ Median
 - ✓ Lower and upper quartiles
 - ✓ Maximum and minimum

Example 1-8: Stem-and-Leaf Display



A stem-and-leaf display showing task performance times. The stems are 1 through 6, and the leaves are the digits following the stem. The display is enclosed in a box with a double border.

Stem	Leaves
1	1 2 2 3 5 5 5 6 7
2	0 1 1 1 2 2 2 3 4 6 7 7 7 8 9 9
3	0 1 2 4 5 7
4	1 1 2 5 7
5	0 2 3 6
6	0 2

Figure 1-17: Task Performance Times

Box Plot

Elements of a Box Plot

