A Course In Business Statistics

4th Edition

Chapter 3

Describing Data Using Numerical Measures

Summary Measures

- **Describing Data Numerically**
  - **Center and Location**
    - Mean
    - Median
    - Mode
    - Weighted Mean
  - **Other Measures of Location**
    - Percentiles
    - Quartiles
  - **Variation**
    - Range
    - Interquartile Range
    - Variance
    - Standard Deviation
    - Coefficient of Variation

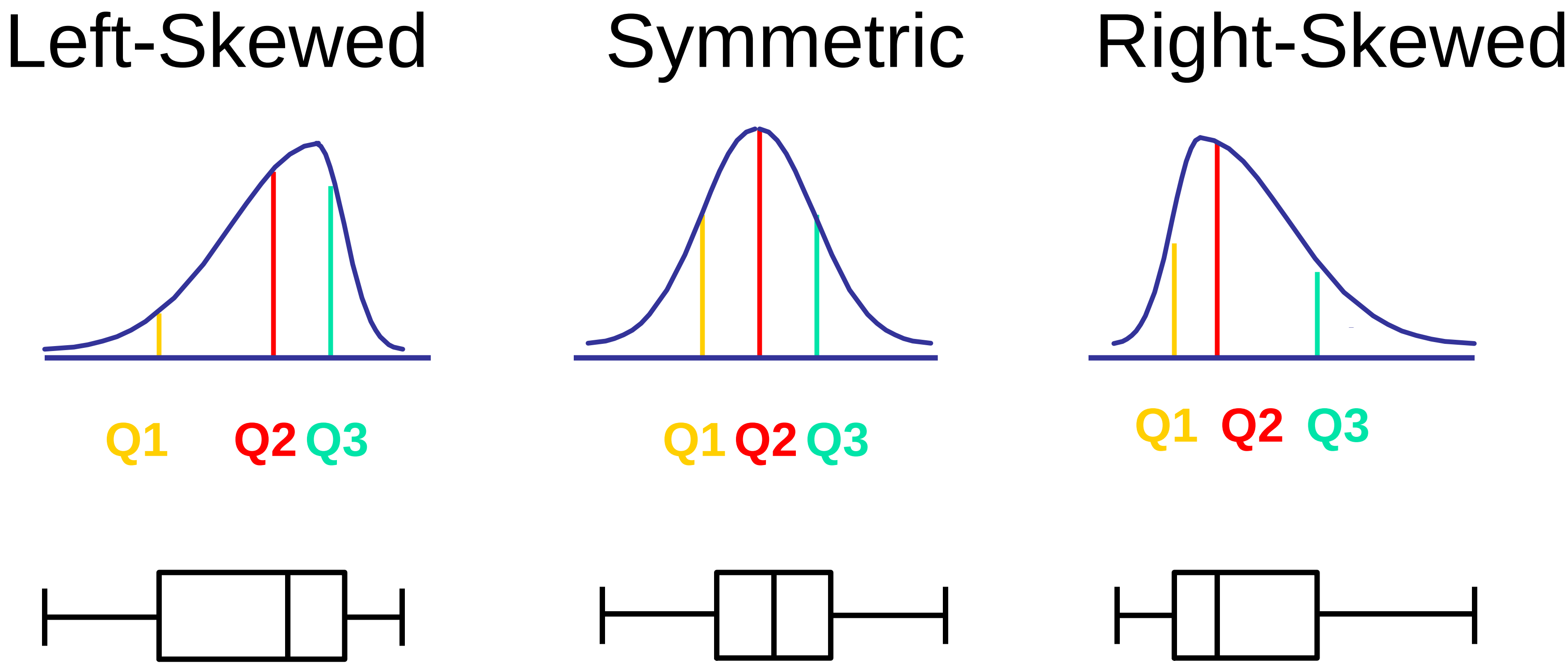
Measures of Shape

- 1) Measures of Skewness
- 2) Measures of Kurtosis

1) Measures of Skewness

- Measures of skewness describe the degree to which the data deviates from symmetry.
- If a distribution is not symmetrical, then it is called as asymmetrical or skewed.
- There are 2 types of skewness:
  - Positively skewed – skewed to the right
  - Negatively skewed – skewed to the left

Distribution Shape and Box and Whisker Plot



Relationship of the Three Measures of Central Tendency for Unimodal Distributions

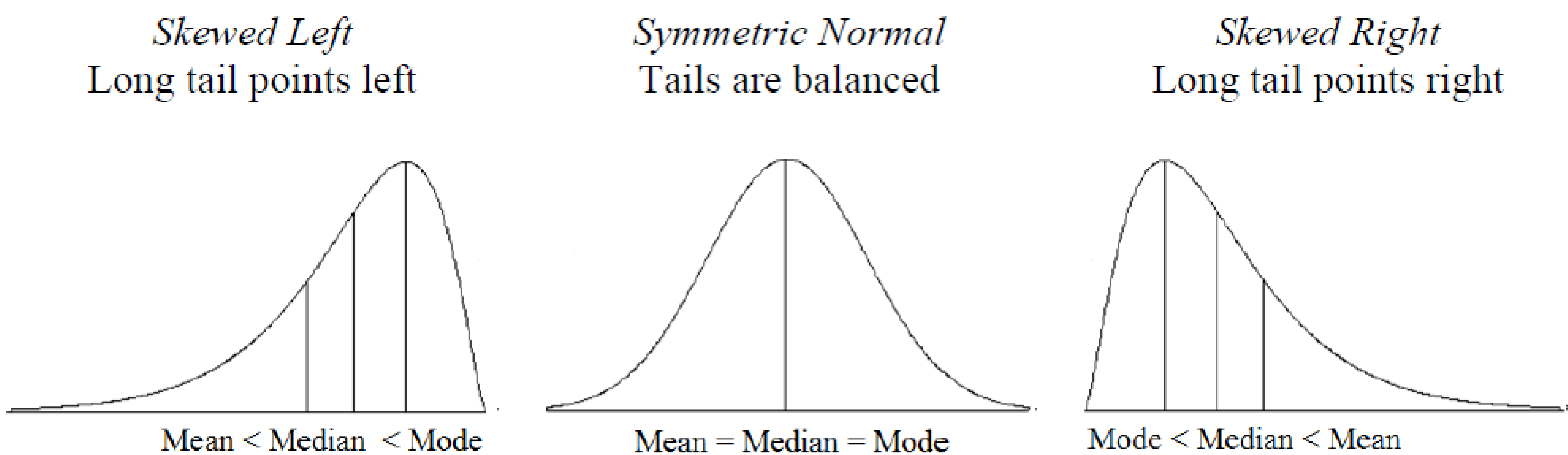


Figure 1. Sketches showing general position of mean, median, and mode in a population.

- Source: Doane, D.P., Seward, L. E. (2011). Measuring Skewness: A Forgotten Statistic?, Journal of Statistics Education, Vol.19, No.2.

- If a distribution is symmetrical, it does not have to be bell-shaped.

- If a distribution is skewed, it does not have to be uni-modal (having only one mode).

- Source: Doane, D.P., Seward, L. E. (2011). Measuring Skewness: A Forgotten Statistic?, Journal of Statistics Education, Vol.19, No.2.

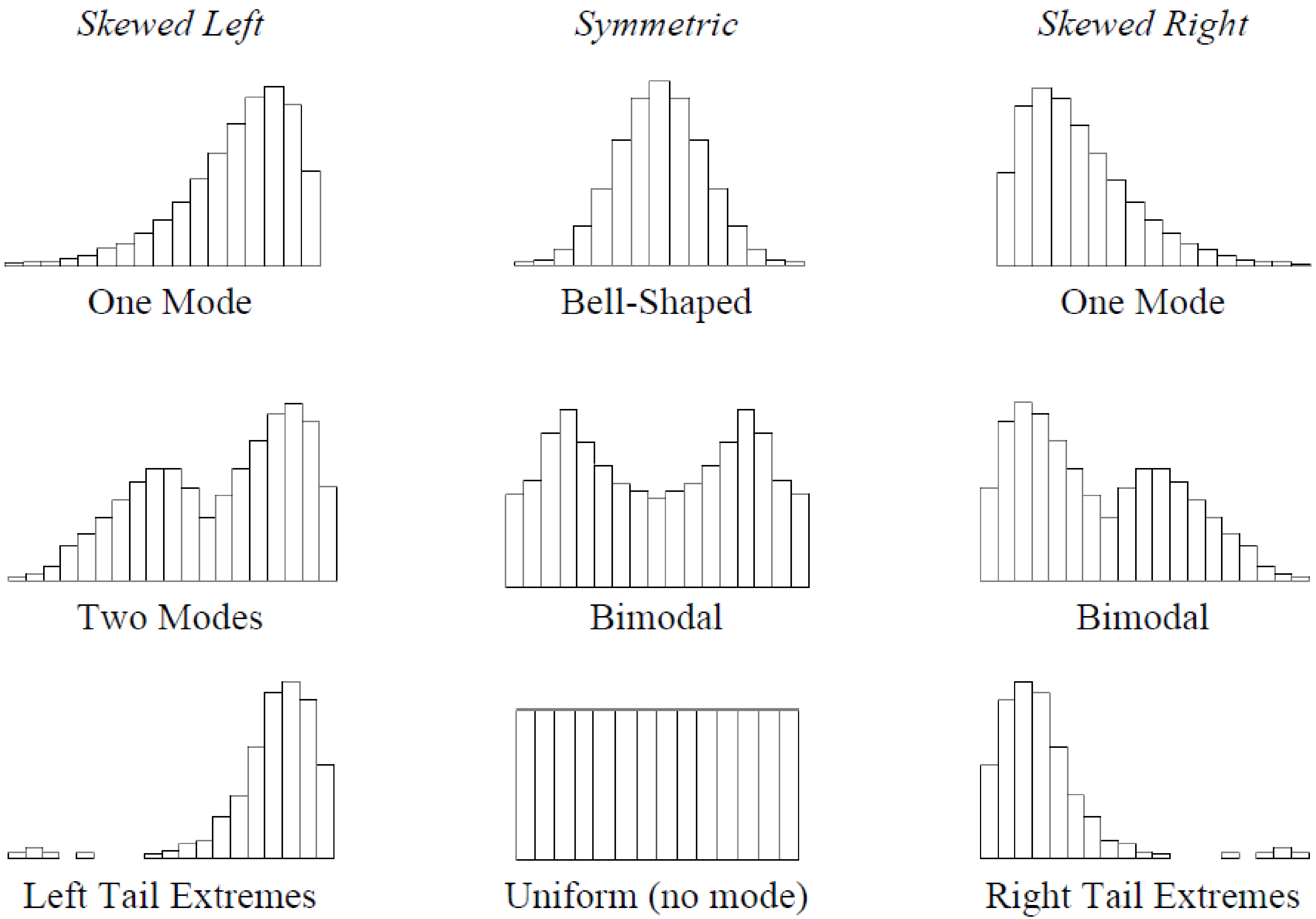


Figure 2. Illustrative prototype histograms.

How to detect Skewness? Charts vs Statistical Measures?

- The simplest way to see if a distribution is skewed or symmetrical is to construct a histogram or a box-plot.
- A **measure of skewness** is a single value that indicates the degree and direction of asymmetry.
- Signs of skewness measures:
  - Skewness = 0 Symmetrical distribution
  - Skewness >0 Skewed to the right (positively skewed)
  - Skewness <0 Skewed to the lef (negatively skewed)
- The larger the measure of skewness, the more skewed the distribution is.

Measures of Skewness

- **There are a few measures for skewness:**
  - Pearson's 1st coefficient of skewness: Based on the distance between the mean and the mode

$$Skewness_1 = \frac{\bar{X} - Mode}{\sigma}$$

  - Pearson's 2nd coefficient of skewness: Based on the distance between the mean and the median

$$Skewness_2 = \frac{3(\bar{X} - Median)}{\sigma}$$

  - Skewness based on the quartiles (boxplot)

$$Skewness_4 = \frac{(Q_3 - Median) - (Median - Q_1)}{Q_3 - Q_1}$$

Measures of Skewness

- **Most traditional measure for skewness:**

  - Fisher-Pearson coefficient of skewness: Based on the 2nd and 3rd moments around the mean.

$$Skewness_3 = \alpha_3 = \frac{\mu_3}{\sigma^3}$$

rth Moment Around the Mean

- The rth moment around the mean is denoted by

$$\mu_r$$

- and it is calculated as follows:

$$\mu_r = \frac{\sum (X_i - \bar{X})^r}{N} \qquad \mu_r = \frac{\sum f_i (X_i - \bar{X})^r}{N} \qquad \mu_r = \frac{\sum X_i (m_i - \bar{X})^r}{N}$$

- Note that:

$$\mu_0 = 1 \qquad \mu_1 = 0 \qquad \mu_2 = \text{Variance}$$

Fisher-Pearson coefficient of skewness

$$\alpha_3 = \frac{\mu_3}{\sigma^3}$$

Interpretation of Fisher-Pearson's Skewness Measure

- Bulmer, M. G., *Principles of Statistics* (Dover, 1979) — a classic — suggests this rule of thumb:
  - If skewness is less than −1 or greater than +1, the distribution is **highly skewed**.
  - If skewness is between −1 and −½ or between +½ and +1, the distribution is **moderately skewed**.
  - If skewness is between −½ and +½, the distribution is **approximately symmetric**.

An Example

i	X	X-Xmean	(X-Xmean)^2	(X-Xmean)^3	(X-Xmean)^4
1	53	5	25	125	625
2	3	-45	2025	-91125	4100625
3	47	-1	1	-1	1
4	30	-18	324	-5832	104976
5	58	10	100	1000	10000
6	39	-9	81	-729	6561
7	100	52	2704	140608	7311616
8	41	-7	49	-343	2401
9	96	48	2304	110592	5308416
10	13	-35	1225	-42875	1500625
Sum	**480**	**0**	**8838**	**111420**	**18345846**
	48				

$$\bar{X} = \frac{480}{10} = 48$$

$$\sigma = \sqrt{\frac{\sum(X-\bar{X})^2}{N}} = \sqrt{\frac{8838}{10}} = 29.73$$

$$\mu_3 = \frac{\sum(X-\bar{X})^3}{N} = \frac{111420}{10} = 11142$$

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{11142}{29.73^3} = 0.42$$

$$\mu_4 = \frac{\sum(X-\bar{X})^4}{N} = \frac{18345846}{10} = 1834584.6$$

$$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{1834584.6}{29.73^4} = 2.35$$

Another example:

Female Life Expectancy at birth (years) High Income OECD (2009)				
Australia	84		Japan	86
Austria	83		Korea, Rep.	84
Belgium	82		Luxembourg	83
Canada	83		Netherlands	83
Czech Republic	80		New Zealand	82
Denmark	81		Norway	83
Estonia	80		Poland	80
Finland	83		Portugal	82
France	85		Slovak Republic	79
Germany	83		Slovenia	82
Greece	83		Spain	85
Hungary	78		Sweden	83
Iceland	83		Switzerland	84
Ireland	82		United Kingdom	82
Israel	84		United States	81
Italy	84			

Source: World Bank

Female Life Expectancy at birth (years) Sub-Saharan Africa (2009)

Angola	52	Gabon	63	Niger	54
Benin	57	Gambia, The	59	Nigeria	52
Botswana	52	Ghana	64	Rwanda	56
Burkina Faso	55	Guinea	55	Sao Tome and Principe	66
Burundi	51	Guinea-Bissau	49	Senegal	60
Cameroon	52	Kenya	57	Seychelles	78
Cape Verde	78	Lesotho	46	Sierra Leone	48
Central African Republic	48	Liberia	56	Somalia	52
Chad	50	Madagascar	68	South Africa	52
Comoros	62	Malawi	53	Sudan	63
Congo, Dem. Rep.	49	Mali	52	Swaziland	48
Congo, Rep.	58	Mauritania	60	Tanzania	57
Cote d'Ivoire	55	Mauritius	77	Togo	58
Eritrea	63	Mozambique	50	Uganda	54
Ethiopia	60	Namibia	62	Zambia	48
				Zimbabwe	47

	High Income OECD	Sub-Saharan Africa
N	31	46
$\sum X$	2557	2606
$\bar{X}$	2557/31 = 82 years	2606/46 = 57 years
$\sum X^2$	211007	150384
$\sigma^2 = \frac{\sum X^2}{N} - \bar{X}^2$	$= \frac{211007}{31} - 82^2 = 3.088$	$= \frac{150384}{46} - 57^2 = 59.748$
σ	$\sqrt{3.088} = 1.757$	$\sqrt{59.748} = 7.729$
$\mu_3 = \frac{\sum(X - \bar{X})^3}{N}$	$= \frac{-91.49}{31} = -2.95$	$= \frac{24591.87}{46} = 534.61$
$\alpha_3 = \frac{\mu_3}{\sigma^3}$	$= \frac{-2.95}{1.757^3} = \mathbf{-0.544}$	$= \frac{534.61}{7.729^3} = \mathbf{1.158}$

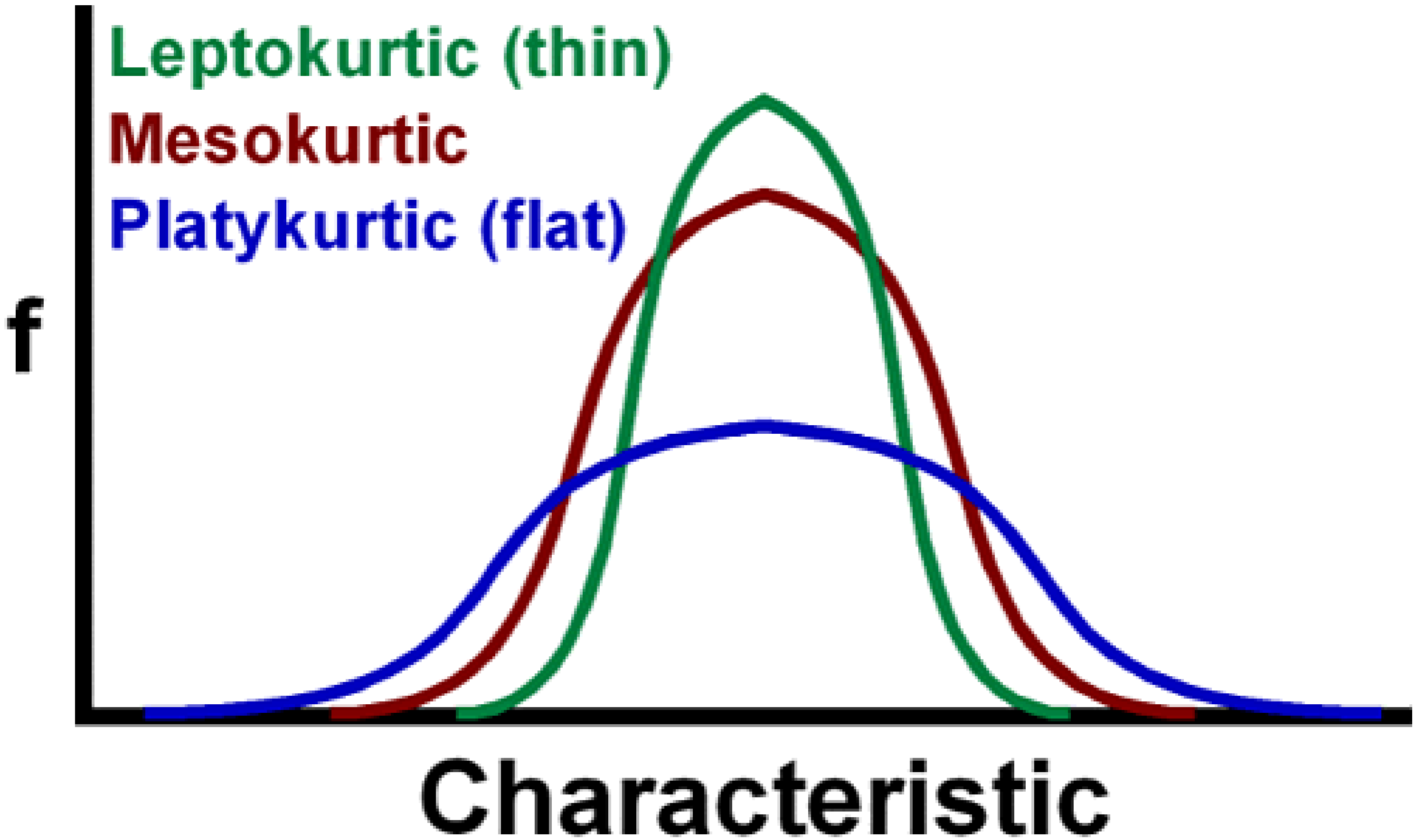
2) Measure of Kurtosis

- Karl Pearson introduced the following terms to classify a unimodal distribution according to the shape of its hump (kambur) as compared to a normal distribution with the same variance:
  - Kurtosis = 3 Symmetrical distribution (mesokurtic)
  - Kurtosis >3 Peaked distribution (leptokurtic)
  - Kurtosis <3 Flat distribution (platykurtic - platus)
- The larger the measure of kurtosis, the more peaked or flattened the distribution is.
  - Kurtosis – 3 shows the «excess kurtosis».

Measures of Kurtosis

- **Most traditional measure for kurtosis:**
  - Coefficient of kurtosis: Based on the 2nd and 4th moments around the mean.

$$Kurtosis = \alpha_4 = \frac{\mu_4}{\sigma^4}$$


Leptokurtic (thin)
Mesokurtic
Platykurtic (flat)
f
Characteristic

	High Income OECD	Sub-Saharan Africa
N	31	46
$\bar{X}$	2557/31 = 82 years	2606/46 = 57 years
σ	$\sqrt{3.088} = 1.757$	$\sqrt{59.748} = 7.729$
$\mu_4 = \frac{\sum(X - \bar{X})^4}{N}$	$= \frac{935.88}{31} = 30.19$	$= \frac{680201}{46} = 14786.98$
$\alpha_4 = \frac{\mu_4}{\sigma^4}$	$= \frac{30.19}{1.757^4} = 3.165$ Mesokurtic Excess kurtosis = 0.165	$= \frac{14786.98}{7.729^4} = 4.14$ Leptokurtic Excess K = 1.14