MEASURES OF SHAPE

Measures of shape are values quantifying several shape characteristics of a bar chart or histogram. There are generally two categories of shape measurements: skewness measurements and kurtosis measurements. Shape measurements only make sense when studying quantitative variables measured on an interval or ratio scale.

MOMENTS

MOMENTS FOR UNGROUPED DATA

If $x_1, x_2, x_3, \ldots, x_n$, are the N values assumed by the variable X, we define the quantity

$$\boldsymbol{m_r} = \frac{\sum x_i^{\,r}}{n}$$

called the r^{th} moment. (simple moments)
The first moment with r = 1 is the arithmetic mean $\boldsymbol{\overline{X}}$.

The r^{th} moment about the mean $\boldsymbol{\overline{X}}$ (centered moments) is defined as

$$\boldsymbol{M_r} = \frac{\sum(\boldsymbol{X_i} - \boldsymbol{\overline{X}})^r}{\boldsymbol{n}}$$

If r = 1, then $M_1 = 0$. If r = 2, then $M_2 = s$ (the variance).

Example: Calculate the simple moments m_1, m_2, m_3, and centered moments M_1, M_2, M_3 and M_4 for the following set of numbers

3, 6, 11, 18, 7

Solution

- simple moments

$$m_1 = \frac{3+6+11+18+7}{5} = 9$$

$$m_2 = \frac{3^2+6^2+11^2+18^2+7^2}{5} = 107.8\ 9$$

$$m_3 = \frac{3^3 + 6^3 + 11^3 + 18^3 + 7^3}{5} = 1549.8$$

$$M_1 = \frac{\sum(3-9) + (6-9) + (11-9) + (18-9) + (7-9)}{5} = 0$$

$$M_2 = \frac{\sum(3-9)^2 + (6-9)^2 + (11-9)^2 + (18-9)^2 + (7-9)^2}{5} = 26.8$$

$$M_3 = \frac{\sum(3-9)^3 + (6-9)^3 + (11-9)^3 + (18-9)^3 + (7-9)^3}{5} = 97.2$$

$$M_4 = \frac{\sum(3-9)^4 + (6-9)^4 + (11-9)^4 + (18-9)^4 + (7-9)^4}{5} = 1594$$

MOMENTS FOR GROUPED DATA

If $x_1, x_2, x_3, \ldots, x_k$, are data points and $n_1, n_2, x_3, \ldots, n_k$, represent their respective frequencies, then, the above moments are given by

$$\boldsymbol{m_r} = \frac{\sum n_i x_i^{\,r}}{\sum \boldsymbol{n_i}}$$

$$M_r = \frac{\sum n_i (X_i - \overline{X})^r}{\sum n_i}$$

Example: Find the simple moments m_2, and m_3, and centered moments M_3 and M_4 for the following frequency distribution

x_i	[20 25[	[25 30[	[30 35[	[35 40[	[40 45[	[45 50[	[50 55[	[55 60[
n_i	4	10	24	34	14	8	4	2

Solution

xi	n_i	c_i	$n_i c_i$	$n_i c_i^2$	$n_i c_i^3$	$X_i - \overline{X}$	$n_i(X_i - \overline{X})^3$	$n_i(X_i - \overline{X})^4$
[20 25[	4	22.5	90	2025	45562,5	-14,7	-12706,092	186779,552
[25 30[	10	27.5	275	7562,5	207968,75	-9,7	-9126,73	88529,281
[30 35[	24	32.5	780	25350	823875	-4,7	-2491,752	11711,2344
[35 40[	34	37.5	1275	47812,5	1792968,75	0,3	0,918	0,2754
[40 45[	14	42.5	595	25287,5	1074718,75	5,3	2084,278	11046,6734
[45 50[	8	47.5	380	18050	857375	10,3	8741,816	90040,7048
[50 55[	4	52.5	210	11025	578812,5	15,3	14326,308	219192,512
[55 60[	2	57.5	115	6612,5	380218,75	20,3	16730,854	339636,336
SUM	**100**	/	**3720**	**143725**	**5761500**	/	**17559,6**	**946936,57**

- simple moments

$$m_2 = \frac{143725}{100} = 143.725$$

$$m_3 = \frac{5761500}{100} = 57615$$

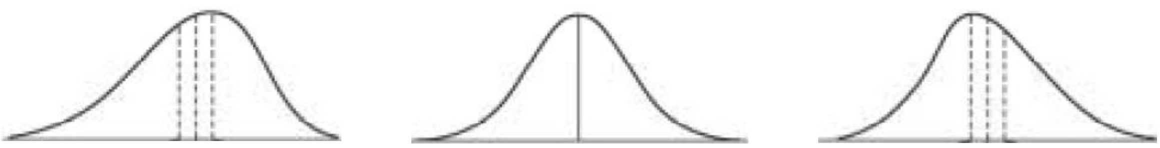
- centered moments

$$M_3 = \frac{17559.6}{100} = 175.596$$

$$M_4 = \frac{946936.57}{100} = 9469.36$$

SKEWNESS

Skewness is the degree of asymmetry, or departure from symmetry, of a distribution. If the frequency curve of a distribution has a longer tail to the right of the central maximum than to the left, the distribution is said to be skewed to the right, or to have positive skewness. If the reverse is true, it is said to be skewed to the left, or to have negative skewness.



skewed to the left | symmetrical | skewed to the right

SKEWNESS MEASUREMENTS

Skewness can be measured by the following coefficients

- The Pearson coefficients of skewness

<u>- Pearson's first coefficient of skewness</u>

$$P_1 = \frac{\bar{X} - M_O}{S}$$

$P_1 = 0$ the distribution is symmetric
$P_1 > 0$ the distribution is said to be skewed to the right
$P_1 < 0$ the distribution is said to be skewed to the left

- Pearson's second coefficient of skewness

$$P_2 = \frac{3(\bar{X} - M_e)}{S}$$

$P_2 = 0$ the distribution is symmetric
$P_2 > 0$ the distribution is said to be skewed to the right
$P_2 < 0$ the distribution is said to be skewed to the left

- Pearson's coefficient of skewness β_1

$$\beta_1 = \frac{M_3^2}{S^3}$$

For a symmetric distribution $\beta_1 = 0$
For a skewed distribution $\beta_1 \neq 0$

Example: Find Pearson's coefficients of skewness for the following frequency distribution

x_i	[20 25[	[25 30[	[30 35[	[35 40[	[40 45[	[45 50[	[50 55[	[55 60[
n_i	4	10	24	34	14	8	4	2

Solution

xi	n_i	c_i	$n_i c_i$	$X_i - \bar{X}$	$n_i(X_i - \bar{X})^2$	$n_i(X_i - \bar{X})^3$
[20 25[	4	22.5	90	-14,7	864,36	-12706,092
[25 30[	10	27.5	275	-9,7	940,9	-9126,73
[30 35[	24	32.5	780	-4,7	530,16	-2491,752
[35 40[	34	37.5	1275	0,3	3,06	0,918
[40 45[	14	42.5	595	5,3	393,26	2084,278
[45 50[	8	47.5	380	10,3	848,72	8741,816
[50 55[	4	52.5	210	15,3	936,36	14326,308
[55 60[	2	57.5	115	20,3	824,18	16730,854
SUM	**100**	/	**3720**	/	5341	**17559,6**

We have calculated before $\bar{X} = 37.2$ $M_e = 36.76$ $S = 7.34$

$$M_O = 35 + \frac{34 - 24}{34 - 24 + 34 - 14} * 5 = 36.66$$

- Pearson's first coefficient of skewness P_1

$$P_1 = \frac{37.2 - 36.66}{7.34} \ 0.073$$

Since $P_1 > 0$ the distribution is slightly skewed to the right

- Pearson's second coefficient of skewness P_2

$$P_2 = \frac{3(37.2 - 36.76)}{7.34} = 0.18$$

Since $P_2 > 0$ the distribution is slightly skewed to the right

- Pearson's coefficient of skewness β_1

$$\beta_1 = \frac{(175.596)^2}{(7.34)^3} = 77.972$$

Since $\beta_1 > 0$ the distribution is skewed

- Fisher's coefficient of skewness F_1

$$F_1 = \frac{M_3}{S^3}$$

$F_1 = 0$ the distribution is symmetric
$F_1 > 0$ the distribution is said to be skewed to the right
$F_1 < 0$ the distribution is said to be skewed to the left

- Yule's coefficient of skewness C_y

$$C_y = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{Q_3 - 2Q_2 - Q_1}{Q_3 - Q_1}$$

$C_y = 0$ the distribution is symmetric
$C_y > 0$ the distribution is said to be skewed to the right
$C_y < 0$ the distribution is said to be skewed to the left

Example: Find the coefficients of skewness of Fisher and Yule for the previous frequency distribution
Solution

- Fisher's coefficient of skewness F_1

$$F_1 = \frac{175.596}{(7.34)^3} = 0.444$$

Since $F_1 > 0$ the distribution is skewed to the right

- Yule's coefficient of skewness C_y

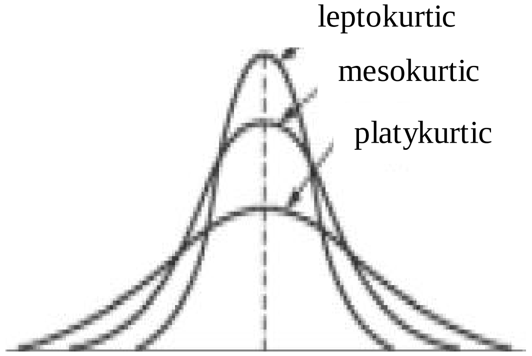
We have calculated before $Q_1 = 32.29$ $Q_2 = 36.76$ $Q_3 = 41.07$, so

$$C_y = \frac{(41.07 - 36.76) - (36.76 - 32.29)}{41.07 - 32.29} = -0.037$$

Unlike other coefficients of skewness Yule's coefficient ($C_y < 0$) indicates that the distribution is skewed to the left.

KURTOSIS

Kurtosis is the degree of peakedness of a distribution, usually taken relative to a normal distribution.
A peaked distribution is called leptokurtic, while one which is flat-topped is called platykurtic. A normal distribution, which is not very peaked or very flat-topped, is called mesokurtic.



- Pearson's coefficient of kurtosis β_2

$$\boldsymbol{\beta_2} = \frac{M_4}{S^4}$$

$\boldsymbol{\beta_2} = 3$ the distribution is mesokurtic
$\boldsymbol{\beta_2} > 3$ the distribution is leptokurtic
$\boldsymbol{\beta_2} < 3$ the distribution is platykurtic

- Fisher's coefficient of kurtosis $\boldsymbol{F_2}$

$$\boldsymbol{F_2} = \frac{M_4}{S^4} - \mathbf{3} = \boldsymbol{\beta_2} - \mathbf{3}$$

$F_2 = 0$ the distribution is mesokurtic
$F_2 > 0$ the distribution is leptokurtic
$F_2 < 0$ the distribution is platykurtic

- percentile coefficient of kurtosis: this coefficient is based on both quartiles and percentiles and is given by

$$\kappa = \frac{Q_3 - Q_1}{P_{90} - P_{10}}$$

$\kappa = 0.263$ the distribution is mesokurtic
$\kappa > 0.263$ the distribution is leptokurtic
$\kappa < 0.263$ the distribution is platykurtic

Example: Find the coefficients of skewness of Fisher and Yule for the previous frequency distribution

Solution

xi	n_i	$X_i - \overline{X}$	$n_i(X_i - \overline{X})^2$	$n_i(X_i - \overline{X})^4$
[20 25[	4	-14,7	864,36	186779,552
[25 30[	10	-9,7	940,9	88529,281
[30 35[	24	-4,7	530,16	11711,2344
[35 40[	34	0,3	3,06	0,2754
[40 45[	14	5,3	393,26	11046,6734
[45 50[	8	10,3	848,72	90040,7048
[50 55[	4	15,3	936,36	219192,512
[55 60[	2	20,3	824,18	339636,336
SUM	**100**	/	5341	**946936,57**

- Pearson's coefficient of kurtosis $\boldsymbol{\beta_2}$

$$\boldsymbol{\beta_2} = \frac{9469.36}{(7.34)^4} = \mathbf{3.262}$$

Since $\beta_2 > 3$ the distribution leptokurtic

- Fisher's coefficient of kurtosis $\boldsymbol{F_2}$

$$\boldsymbol{F_2} = \frac{9469.36}{(7.34)^4} - 3 = 3.262 - 3 = 0.262$$

Since $F_2 > 0$ the distribution is leptokurtic

- percentile coefficient of kurtosis:

$$\kappa = \frac{Q_3 - Q_1}{P_{90} - P_{10}}$$