Moments

Mean and median

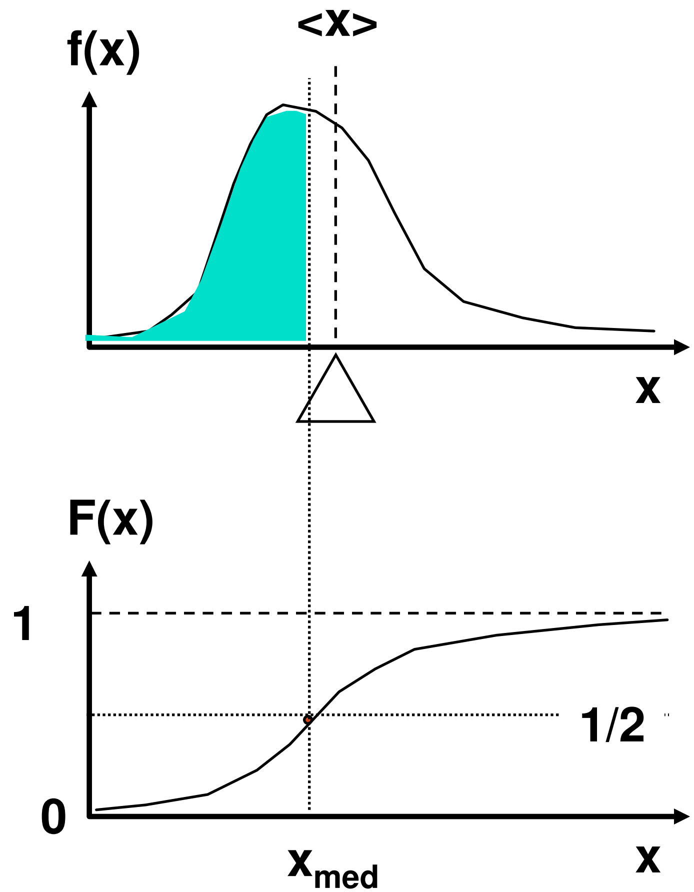
- Mean value (centre of gravity)

$$\langle x \rangle \equiv \int f(x)\, x\, dx$$

- Median value (50th percentile)

$$F(x_{\text{med}}) \equiv \frac{1}{2}$$

$$P(x < x_{\text{med}}) = P(x > x_{\text{med}})$$

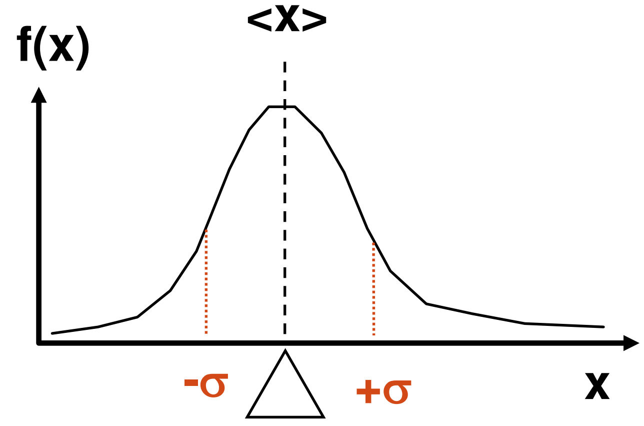


Variance and standard deviation

- Standard deviation σ measures width of distribution.
- Variance σ^2 (moment of inertia)

$$\sigma^2 = \mathrm{Var}(x) \equiv \left\langle [x - \langle x \rangle]^2 \right\rangle$$

$$= \int f(x)[x - \langle x \rangle]^2 \, dx$$



Moments

A moment is a quantitative measure of the shape of a set of points.

The first moment is called the mean which describes the center of the distribution.

The second moment is the variance which describes the spread of the observations around the center.

Other moments describe other aspects of a distribution such as how the distribution is skewed from its mean or peaked.

A moment designates the power to which deviations are raised before averaging them.

Central (or Mean) Moments

In mean moments, the deviations are taken from the mean.

For Ungrouped Data:

$$\text{First } \textit{Population Moment about Mean}=\mu_1 = \frac{\sum(x_i - \mu)}{N}$$

$$\textit{Second Population Momentabout Mean}=\mu_2 = \frac{\sum(x_i - \mu)^2}{N}$$

$$\textit{First Sample Momentabout Mean}=m_1 = \frac{\sum(x_i - \bar{x})}{n}$$

$$\textit{Second Sample Momentabout Mean}=m_2 = \frac{\sum(x_i - \bar{x})^2}{n}$$

In General,

$$\text{r}^{th} \text{ Population Moment about Mean}=\mu_r = \frac{\sum(x_i - \mu)^r}{N}$$

$$\text{r}^{th} \text{ Sample Moment about Mean}=m_r = \frac{\sum\left(x_i - \overline{x}\right)^r}{n}$$

Central (or Mean) Moments

Formula for Grouped Data:

$$\text{r}^{th}\ \text{Population Moment about Mean} = \mu_r = \frac{\sum f\left(x_i - \mu\right)^r}{\sum f}$$

$$\text{r}^{th}\ \text{Sample Moment about Mean} = m_r = \frac{\sum f\left(x_i - \overline{x}\right)^r}{\sum f}$$

Moments about (arbitrary) Origin

- If the deviations are taken from some arbitrary number ('a' called origin), then moments are called moments about arbitrary origin 'a'.

For Ungrouped Data:

$$r^{th} \text{ Population Moment about Origin } 'a' = \mu_r' = \frac{\sum (x_i - a)^r}{N}$$

$$r^{th} \text{ Sample Moment about Origin } 'a' = m_r' = \frac{\sum (x_i - a)^r}{n}$$

For Grouped Data:

$$r^{th} \text{ Population Moment about Origin } 'a' = \mu_r' = \frac{\sum f (x_i - a)^r}{\sum f}$$

$$r^{th} \text{ Sample Moment about Origin } 'a' = m_r' = \frac{\sum f (x_i - a)^r}{\sum f}$$

Moments about zero

- If origin is taken as zero. i.e. a=0, moments are called moments about zero.

For Ungrouped Data:

$$\text{r}^{th}\ \textit{Population Moment about Zero}=\mu_r' = \frac{\sum(x_i - 0)^r}{N} = \frac{\sum(x_i)^r}{N}$$

$$r^{th}\ \textit{Sample Moment about Zero}=m_r' = \frac{\sum(x_i - 0)^r}{n} = \frac{\sum(x_i)^r}{n}$$

For Grouped Data:

$$\text{r}^{th}\ \textit{Population Moment about Zero}=\mu_r' = \frac{\sum f(x_i-0)^r}{\sum f} = \frac{\sum f(x_i)^r}{\sum f}$$

$$r^{th}\ \textit{Sample Moment about Zero}=m_r' = \frac{\sum f(x_i-0)^r}{\sum f} = \frac{\sum f(x_i)^r}{\sum f}$$

Conversion from Moments about Mean to Moments about Origion

- **Sample Moments about Mean in terms of Moments about Origion.**

$$m_1 = m_1' - m_1' = 0$$
$$m_2 = m_2' - (m_1')^2$$
$$m_3 = m_3' - 3m_2'm_1' + 2(m_1')^3$$
$$m_4 = m_4' - 4m_3'm_1' + 6m_2'(m_1')^2 - 3(m_1')^2$$

Population Moments about Mean in terms of Moments about Origion.

$$\mu_1 = \mu_1' - \mu_1' = 0$$
$$\mu_2 = \mu_2' - (\mu_1')^2$$
$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$$
$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^2$$

Moment Ratios

Ratios involving moments are called moment-ratios.

Most common moment ratios are defined as:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}, \beta_2 = \frac{\mu_4}{\mu_2^2}$$

Since these are ratios and hence have no unit.

For symmetric distributions, β_1 is equal to zero. So it is used as a measure of skewness.

β_2 is used to explain the shape of the curve and it is a measure of peakedness.

For normal distribution (Bell-Shaped Curve), $\beta_2 = 3$.

For sample data, moment ratios can be similarly defined as:

$$b_1 = \frac{m_3^2}{m_2^3}, b_2 = \frac{m_4}{m_2^2}$$

Skewness

A distribution in which the values equidistant from the mean have equal frequencies and is called **Symmetric Distribution**.

Any departure from symmetry is called **skewness**.

In a perfectly **symmetric distribution, Mean=Median=Mode** and the two tails of the distribution are equal in length from the mean. These values are pulled apart when the distribution departs from symmetry and consequently one tail become longer than the other.

If right tail is longer than the left tail then the distribution is said to have **positive skewness**. In this case, **Mean>Median>Mode**

If left tail is longer than the right tail then the distribution is said to have **negative skewness**. In this case, **Mean<Median<Mode**

Skewness

When the distribution is symmetric, the value of skewness should be zero.

Karl Pearson defined coefficient of Skewness as:

$$Sk = \frac{Mean - Mode}{SD}$$

Since in some cases, Mode doesn't exist, so using empirical relation,

$$Mode = 3Median - 2Mean$$

We can write,

$$Sk = \frac{3\left(Median - Mean\right)}{SD}$$

(it ranges b/w -3 to +3)

Skewness

- According to Bowley (a British Statistician):

Bowley's coefficient of skewness (also called Quartile skewness coefficient)

$$sk = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{Q_1 - 2Q_2 + Q_3}{Q_3 - Q_1} = \frac{Q_1 - 2Median + Q_3}{Q_3 - Q_1}$$

Another measure of skewness mostly used is by using moment ratio (denoted by $\sqrt{\beta_1}$):

$$sk = \frac{\mu_3}{\sigma^3}, \qquad \text{for population data}$$

$$sk = \frac{m_3}{s^3}, \qquad \text{for sample data}$$

For symmetric distributions, it is zero and has positive value for positively skewed distribution and take negative value for negatively skewed distributions.

Kurtosis

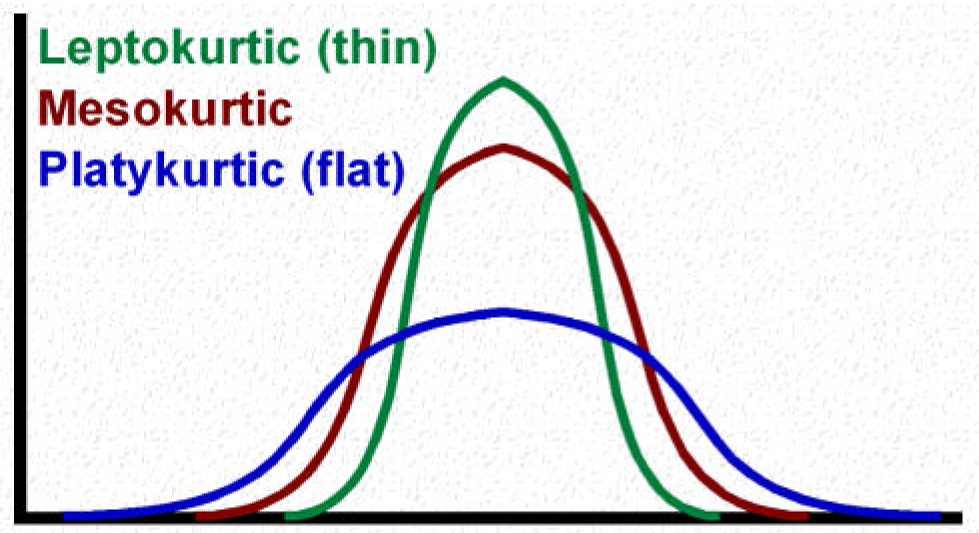
Karl Pearson introduced the term Kurtosis (literally the amount of hump) for the degree of peakedness or flatness of a unimodal frequency curve.

When the peak of a curve becomes relatively high then that curve is called Leptokurtic.

When the curve is flat-topped, then it is called Platykurtic.

Since normal curve is neither very peaked nor very flat topped, so it is taken as a basis for comparison.

The normal curve is called Mesokurtic.



Kurtosis

- Kurtosis is usually measured by the moment ratio (β_2).

$$Kurt = \beta_2 = \frac{\mu_4}{\mu_2^2}, \qquad \text{for population data}$$

$$Kurt = b_2 = \frac{m_4}{m_2^2}, \qquad \text{for sample data}$$

For a normal distribution, kurtosis is equal to 3.

When is greater than 3, the curve is more sharply peaked and has narrower tails than the normal curve and is said to be leptokurtic.

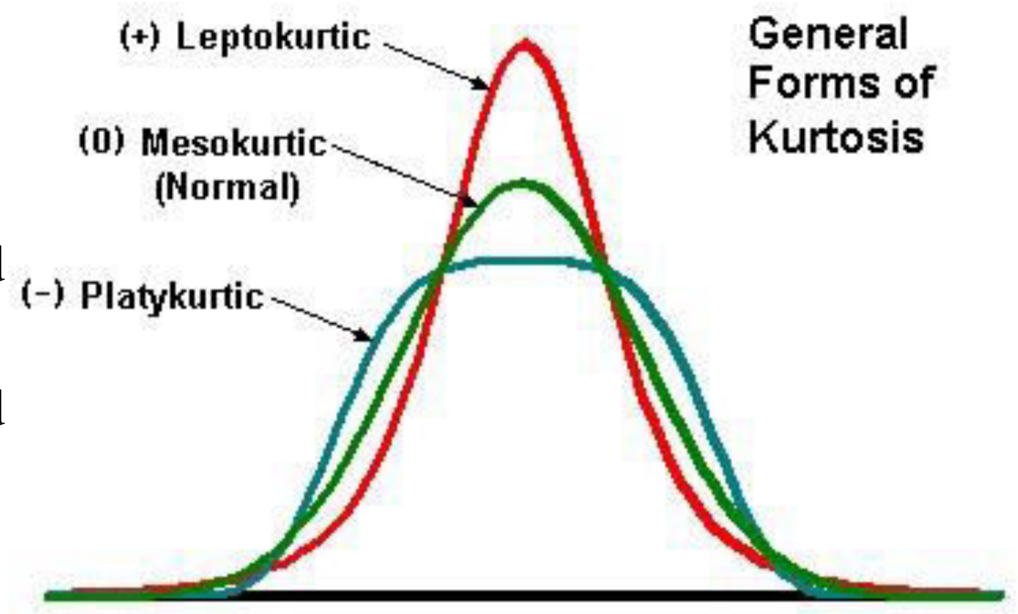
When it is less than 3, the curve has a flatter top and relatively wider tails than the normal curve and is said to be platykurtic.

Kurtosis

Excess Kurtosis (EK): It is defined as:

EK=Kurtosis-3

- For a normal distribution, EK=0.
- When EK>0, then the curve is said to be Leptokurtic.
- When EK<0, then the curve is said to be Platykurtic.



Kurtosis

Another measure of Kurtosis, known as Percentile coefficient of kurtosis is:

$$\text{Kurt}=\frac{Q.D}{P_{90}-P_{10}}$$

Where,

Q.D is semi-interquartile range=Q.D=$(Q_3-Q_1)/2$

P_{90}=90th percentile

P_{10}=10th percentile

Describing a Frequency Distribution

To describe the major characteristics of a frequency distribution, we need to calculate the following five quantities:

- The total number of observations in the data.
- A measure of central tendency (e.g. mean, median etc.) that provides the information about the center or average value.
- A measure of dispersion (e.g. variance, SD etc.) that indicates the spread of the data.
- A measure of skewness that shows lack of symmetry in frequency distribution.
- A measure of kurtosis that gives information about its peakedness.

Describing a Frequency Distribution

It is interesting to note that all these quantities can be derived from the first four moments.

For example,

- The first moment about zero is the arithmetic mean
- The second moment about mean is the variance.
- The third standardized moment is a measure of skewness.
- The fourth standardized moment is used to measure kurtosis.

Thus first four moments play a key role in describing frequency distributions.