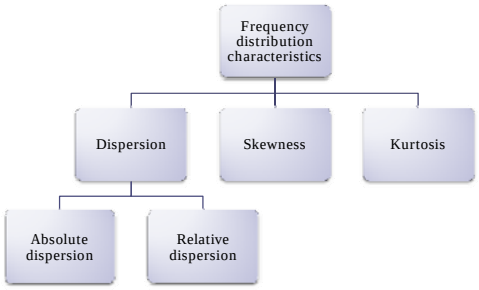
Dispersion, Skewness and Kurtosis

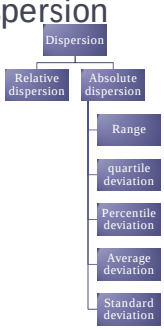
Presented by:
Mahendra AN

Sources:
http://business.clayton.edu/arjomand/business/stat%20presentations/busa3101.html
Djarwanto, **Statistik Sosial Ekonomi** Bagian pertama edisi 3, BPFE, 2001

Framework



Dispersion



- Dispersion is separate measures of values among its central tendency.

Absolute Dispersion

Range
- Simplest dispersion measure
- Effected by extreme values

$$R = X_{largest} - X_{smallest}$$

Quartile Deviation
- Dispersion of inter quartile range

$$QD = \frac{Q_3 - Q_1}{2}$$

Absolute Dispersion (cont.)

Percentile Deviation
- Dispersion of inter percentile range (P_{10} and P_{90})

$$PD = \frac{P_{90} - Q_{10}}{2}$$

Average deviation (mean deviation)
- Ungrouped data

$$AD = \frac{\sum|X - \bar{X}|}{n}$$

- Grouped data

$$AD = \frac{\sum f|m - X|}{n}$$

Absolute Dispersion (cont.)

Standard deviation $s = \sqrt{variance}$
- Ungrouped data

small sample $n < 30$

$$s = \sqrt{\frac{\sum(X - \bar{X})^2}{n-1}}$$

big sample $n \geq 30$

$$s = \sqrt{\frac{\sum(X - \bar{X})^2}{n}}$$

σ = population
s = sample

- Grouped data

small sample $n < 30$

$$s = \sqrt{\frac{\sum f(m - \bar{X})^2}{n-1}}$$

big sample $n \geq 30$

$$s = \sqrt{\frac{\sum f(m - \bar{X})^2}{n}}$$

Standard Units

- To compare two or more distributions
- Standard unit show deviation of a variable value (X) on mean ($\bar{X}$) in standard deviation unit (s)
- Commonly base on zero value (Z=0)
- Example: Z=1.2 is better than z=1.0

$$Z = \frac{X - \bar{X}}{S}$$

Relative Dispersion

- To know smallest variation in a distribution
- To compare two or more frequency distribution
- All of standard dispersion measurement can be used.
- Stated in coefficient of variation

$$V = \frac{s}{\bar{X}} \cdot 100; or\ V = \frac{Range}{\bar{X}} \cdot 100; or\ V = \frac{AD}{\bar{X}} \cdot 100; or\ V = \frac{QD}{Md} \cdot 100; or\ V = \frac{(Q_3 - Q_1)}{(Q_3 + Q_1)} \cdot 100$$

Skewness

- An important measure of the shape of a distribution is called skewness
- The formula for computing skewness for a data set is somewhat complex

Skewness (cont.)

Karl Pearson method

- Base on mean and median values

$$Sk = \frac{X - M_o}{S}; or\ Sk = \frac{3(X - M_d)}{S}$$

Bowley method

- Base on quartile values

$$Sk = \frac{(Q_3 + Q_2) - (Q_2 - Q_1)}{(Q_3 + Q_2) + (Q_2 - Q_1)}$$

Skewness (cont.)

10 – 90 percentile's method

- Base on percentile $Sk = \frac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{(P_{90} - P_{10})}$
- The better measurement for skewness base on the third moment.

$$M_3 = \frac{\Sigma(X - \bar{X})^3}{n}\ for\ big\ sample \qquad \beta_1 = \frac{M_3^2}{M_2^3}$$

$$M_2 = \frac{\Sigma(X - \bar{X})^2}{n}\ for\ big\ sample \qquad \beta_1 = relative\ skewness$$

Skewness (cont.)

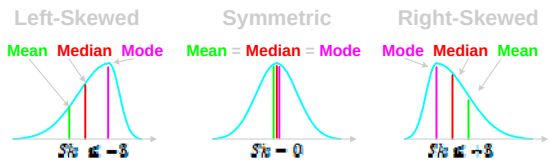
- Grouped data skewness
- Third moment method is used

$$M_2 = \frac{\Sigma fd'^2}{n} - \left(\frac{\Sigma fd'}{n}\right)^2 \qquad \beta_1 = \frac{M_3^2}{M_2^3}$$

$$M_3 = \frac{\Sigma fd'^3}{n} - 3\frac{\Sigma fd'}{n} \cdot \frac{\Sigma fd'^2}{n} + 2\left(\frac{\Sigma fd'}{n}\right)^3$$

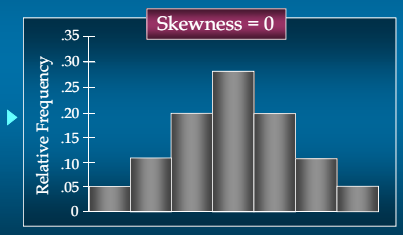
Deciding skewness

$$-3 \leq Sk \leq +3$$



Distribution Shape: Skewness

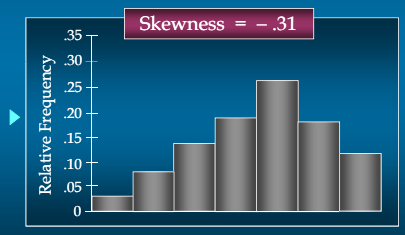
- Symmetric (not skewed, SK = 0)
  - If skewness is zero, then
  - Mean and median are equal.



Slide 14

Distribution Shape: Skewness

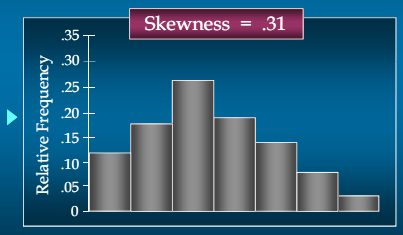
- Moderately Skewed Left
  - Is skewness is negative (left skewed SK >= -3), then
  - Mean will usually be less than the median.



Slide 15

Distribution Shape: Skewness

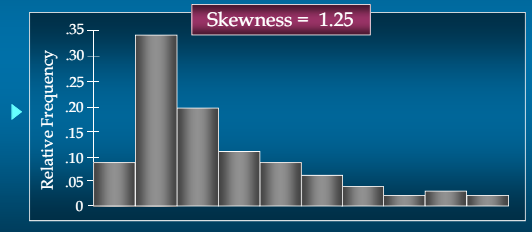
- Moderately Skewed Right
  - If skewness is positive (right skewed, SK >= +3), then
  - Mean will usually be more than the median.



Slide 16

Distribution Shape: Skewness

- Highly Skewed Right
  - Skewness is positive (often above 1.0).
  - Mean will usually be more than the median.



Slide 17

Kurtosis

- Measurement of peakedness
- Difficult to interpret
- Only in theoretic not in practice