P. V. THATKAR

List of Formulae (Statistical Methods I)

Statistical Measure	Formula For: Individual Observations	Formula For: Discrete Frequency Distribution	Formula For: Continuous Frequency Distribution	Remark
Arithmetic Mean	$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}$ $= \frac{\sum_{i=1}^{n} X_i}{n}$	$\overline{X} = \frac{X_1 f_1 + X_2 f_2 + ... + X_n f_n}{f_1 + f_2 + ... + f_n}$ $= \frac{\sum_{i=1}^{n} X_i f_i}{\sum_{i=1}^{n} f_i}$	$\overline{X} = \frac{X_1 f_1 + X_2 f_2 + ... + X_n f_n}{f_1 + f_2 + ... + f_n}$ $= \frac{\sum_{i=1}^{n} X_i f_i}{\sum_{i=1}^{n} f_i}$, Where, X_i is the lass mark/mid point of i'th class.	**Def.: Sum of observations divided by total number of observations.**
Combined Mean	$\overline{X}_c = \frac{n_1 \overline{X}_1 + n_2 \overline{X}_2 + ... + n_k \overline{X}_k}{n_1 + n_2 + ... + n_k}$, Where $\overline{X}_c$ is Combined mean of k groups of size $n_1, n_2, ..., n_k$ with means $\overline{X}_1, \overline{X}_2, ..., \overline{X}_k$ respectively.			
Weighted Arithmetic Mean	$\overline{X}_w = \frac{w_1 \overline{X}_1 + w_2 \overline{X}_2 + ... + w_k \overline{X}_k}{w_1 + w_2 + ... + w_k}$ Where $\overline{X}_w$ is Arithmetic mean of k groups with weights $w_1, w_2, ..., w_k$ with means $\overline{X}_1, \overline{X}_2, ..., \overline{X}_k$ respectively.			

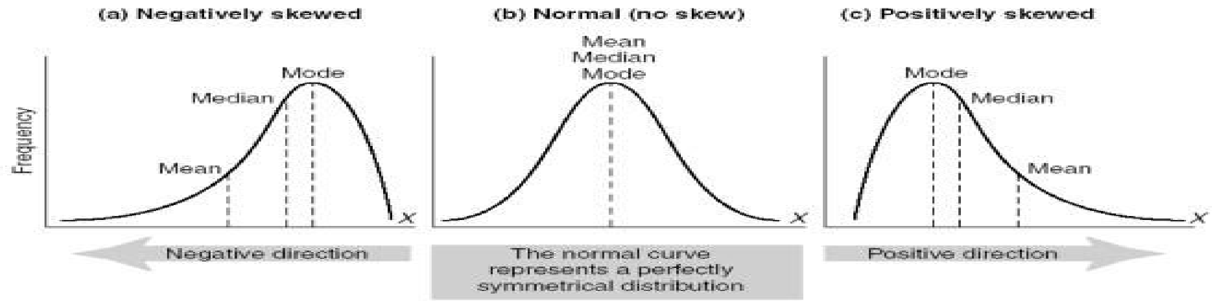
Statistical Measure	Formula For			Remark
	Individual Observations	**Discrete Frequency Distribution**	**Continuous Frequency Distribution**	
Mode	In this case we can find value of mode just by inspection. The observation which is repeated for maximum number of times is mode.	In this case we can find value of mode just by inspection. The observation maximum frequency is mode.	Mode = $l+\left[\frac{f_m - f_{m-1}}{2f_m - f_{m-1} - f_{m+1}}\right]\times h$ Where, l, lower limit of modal class, f_m, frequency of modal class, f_{m-1}, frequency of pre-modal class, f_{m+1}, frequency of post modal class, h class width.	**Def.: The most frequent observation in given set of data**
Median (Q2)	Step1: Arrange obs. as ordered statistic. Step2: count no. of obs. say 'n' If i) 'n' is odd, Median = $\{\frac{(n+1)}{2}\}'th$ Obs. ii) 'n' is even, Median = $\left[\frac{(\frac{n}{2})'th\,Obs. + (\frac{n}{2}+1)'th\,Obs.}{2}\right]$	**Step1: Find the l.c.f. of given distribution.** **Step2: Find N/2.** **Step3: Compare value of N/2 with l.c.f.** **Step 4: Find the l.c.f. which is just greater than or equal to N/2 for the first time, the corresponding value of variable X is Median.**	$Median = l+\left[\frac{\frac{N}{2} - c.f.}{f}\right]\times h$, Where, l, the lower limit of median class, N , is the total frequency, $c.f.$, l.c.f. of premedian class, f, frequency of median class, h, Class width. Median Class is the class for which l.c.f. is just greater than or equal to N/2 for the first time.	**Def.: It is the value which divides the whole distribution in two equal parts.**
Empirical Relation	$(Mean - Mode) = 3\times(Mean - Median)$ ***We can use*** **Emperical relation in following cases.** 1. **When mode is ill defined.** 2. **When median can not be calculated directly .**			

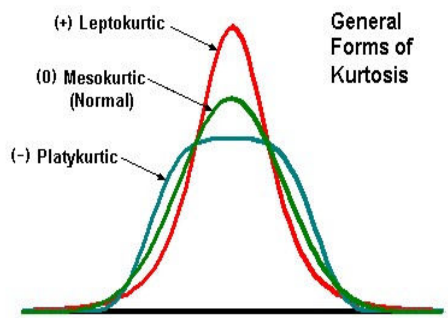
Statistical Measure	Formula For continuous frequency distribution	Remark
i'th Quartile (Qi)	$Q_i = l + \left[\dfrac{\frac{iN}{4} - c.f.}{f}\right] \times h \quad , i = 1, 2, 3.$ Where, l, the lower limit of i'th Quartile class, N, is the total frequency, $c.f.$, l.c.f. of pre-quartile class, f, frequency of quartile class, h, Class width.i'th Quartile Class is the class for which l.c.f. is just greater than or equal to iN/4 for the first time.	**Quartiles are the three values which divides the whole distribution into four equal parts.**
i'th Decile	$D_i = \left[\dfrac{\frac{iN}{10} - c.f.}{f}\right] \times h$; i=1,2,…,9 Where, l, the lower limit of I'th Decile class, N, is the total frequency, $c.f.$, l.c.f. of pre-decile class, f, frequency of i'th decile class, h, Class width. i'th Quartile Class is the class for which l.c.f. is just greater than or equal to iN/10 for the first time.	**Deciles are nine values which divides the whole distribution in to ten equal parts.**
i'th Percentile	$P_i = \left[\dfrac{\frac{iN}{100} - c.f.}{f}\right] \times h$;i=1,2,…,99 Where, l, the lower limit of i'th Percentile class, N, is the total frequency, $c.f.$, l.c.f. of pre-percentile class, f, frequency of i'th percentile class, h, Class width. i'th Percentile Class is the class for which l.c.f. is just greater than or equal to iN/100 for the first time.	**Percentiles are the 99 values which divides whole distribution into hundred equal parts.**
Mode by Graphical Method	Step1: Draw a histogram of given frequency distribution. Step2. Identify the modal bar, pre-modal bar and post-modal bar. Modal bar is the bar having maximum height. Consider the points: lower limit of pre-modal bar, lower limit of modal bar and upper limit of modal bar and upper limit of post-modal bar. Join lower limit of post modal bar to lower limit of modal bar and upper limit of pre-modal bar to upper limit of modal bar. Step3: from the point of intersection of lines drawn in step 2, draw a perpendicular on X axis. The value at the foot of perpendicular is mode.	
Quartiles by Graphical Method	Step1: Draw l.c.f. curve of given frequency distribution. We have to find i'th quartile. Step2: locate value $\frac{iN}{4}$ on Y-axis. From this point draw a parallel line to X axis to meet the curve. Step3: from the point of intersection of line drawn above and l.c.f. curve, draw a perpendicular on X-axis. Value at the foot of perpendicular gives the value of i'th quartile.	

P. V. THATKAR

Statistical Measure	Formula For: Individual Observations	Formula For: Discrete Frequency Distribution	Formula For: Continuous Frequency Distribution	Remark
Range	Range = L – S , Where, L= largest Observation, S= Smallest Observation, It is absolute measure of dispersion.			
Coefficient of range	Coefficient of Range = $\frac{L-S}{L+S}$ Where, L= largest Observation, S= Smallest Observation, It is Relative measure of dispersion.			
Quartile Deviation	$Q.D.=\frac{Q_3-Q_1}{2}$, It is absolute measure of dispersion.			
Coefficient of Q. D.	$Q.D.=\frac{Q_3-Q_1}{Q_3+Q_1}$, It is Relative measure of dispersion.			
Variance	$\sigma_X^2=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}$ **OR** $\sigma_X^2=\frac{\sum_{i=1}^{n}X_i^2}{n}-(\bar{X})^2$	$\sigma_X^2=\frac{\sum_{i=1}^{n}f_i(X_i-\bar{X})^2}{\sum_{i=1}^{n}f_i}$ **OR** $\sigma_X^2=\frac{\sum_{i=1}^{n}f_iX_i^2}{\sum_{i=1}^{n}f_i}-(\bar{X})^2$	$\sigma_X^2=\frac{\sum_{i=1}^{n}f_i(X_i-\bar{X})^2}{\sum_{i=1}^{n}f_i}$, **OR** $\sigma_X^2=\frac{\sum_{i=1}^{n}f_iX_i^2}{\sum_{i=1}^{n}f_i}-(\bar{X})^2$ X_i is the mid point of	
Standard Deviation	$\sigma_X=\sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}}$ or $\sigma_X=\sqrt{\frac{\sum_{i=1}^{n}X_i^2}{n}-(\bar{X})^2}$	$\sigma_X=\sqrt{\frac{\sum_{i=1}^{n}f_i(X_i-\bar{X})^2}{f_i}}$ Or $\sigma_X=\sqrt{\frac{\sum_{i=1}^{n}f_iX_i^2}{f_i}-(\bar{X})^2}$		
Coefficient of Variation	$C.V.=\frac{\sigma_X}{\bar{X}}\times 100\ \%$, It is Relative measure of dispersion. It is used to check consistency of data under study or for Comparison purpose. Less C. V. More consistency, more C. V. less consistency.			
Combined Variance	$\sigma_c^2=\frac{n_1(\sigma_1^2+d_1^2)+n_2(\sigma_2^2+d_2^2)+....+n_k(\sigma_k^2+d_k^2)}{n_1+n_2+.....+n_k}$, where $\bar{X}=\frac{\sum_{i=1}^{n}X_i}{n}$, $d_1^2=(\bar{X}_1-\bar{X}_c)^2$, $d_2^2=(\bar{X}_2-\bar{X}_c)^2$			

Statistical Measure	Formula For		Remark
	Individual Observations	**Discrete Frequency Distribution / Continuous Frequency Distribution**	
Moments About Mean (Central Moments)	$\mu_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})}{n}$, $\mu_2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$ $\mu_3 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^3}{n}$, $\mu_4 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^4}{n}$	$\mu_1 = \frac{\sum_{i=1}^{n} f_i(X_i - \bar{X})}{\sum_{i=1}^{n} f_i}$, $\mu_2 = \frac{\sum_{i=1}^{n} f_i(X_i - \bar{X})^2}{\sum_{i=1}^{n} f_i}$, $\mu_3 = \frac{\sum_{i=1}^{n} f_i(X_i - \bar{X})^3}{\sum_{i=1}^{n} f_i}$, $\mu_4 = \frac{\sum_{i=1}^{n} f_i(X_i - \bar{X})^4}{\sum_{i=1}^{n} f_i}$	
Moments About Origin (Raw Moments)	$\mu_1' = \frac{\sum_{i=1}^{n} X_i}{n}$, $\mu_1' = \frac{\sum_{i=1}^{n} X^2{}_i}{n}$, $\mu_1' = \frac{\sum_{i=1}^{n} X_i^3}{n}$, $\mu_1' = \frac{\sum_{i=1}^{n} X_i^4}{n}$	$\mu_1' = \frac{\sum_{i=1}^{n} f_i X_i}{\sum_{i=1}^{n} f_i}$, $\mu_2' = \frac{\sum_{i=1}^{n} f_i X^2{}_i}{\sum_{i=1}^{n} f_i}$, $\mu_3' = \frac{\sum_{i=1}^{n} f_i X^3{}_i}{\sum_{i=1}^{n} f_i}$, $\mu_4' = \frac{\sum_{i=1}^{n} f_i X_i^4}{\sum_{i=1}^{n} f_i}$	
Moments About Arbitrary point 'a'	$\mu_1'(a) = \frac{\sum_{i=1}^{n}(X_i - a)}{n}$, $\mu_2'(a) = \frac{\sum_{i=1}^{n}(X_i - a)^2}{n}$ $\mu_3'(a) = \frac{\sum_{i=1}^{n}(X_i - a)^3}{n}$, $\mu_4'(a) = \frac{\sum_{i=1}^{n}(X_i - a)^4}{n}$	$\mu_1'(a) = \frac{\sum_{i=1}^{n} f_i(X_i - a)}{\sum_{i=1}^{n} f_i}$, $\mu_2'(a) = \frac{\sum_{i=1}^{n} f_i(X_i - a)^2}{\sum_{i=1}^{n} f_i}$, $\mu_3'(a) = \frac{\sum_{i=1}^{n} f_i(X_i - a)^3}{\sum_{i=1}^{n} f_i}$, $\mu_4'(a) = \frac{\sum_{i=1}^{n} f_i(X_i - a)^4}{\sum_{i=1}^{n} f_i}$	
Relation between Raw moments and Central moments	$\mu_1 = 0$ $\quad$ $\mu_2 = \mu_2' - \mu_1'^2$ $\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$ $\quad$ $\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$		

Sr. No.	Measures of Skewness and Kurtosis	Remark
Graphical Method	 FIGURE 15.6 Examples of normal and skewed distributions	
Karl Pearson's Measure of Skewness	$Skewness = Mean - Mode$ **or** $Skewness = 3(Mean - Median)$	
Karl Pearson's coefficient of Skewness	$S_{kp} = \frac{(Mean - Mode)}{Std.Dev.}$ **Or** $S_{kp} = \frac{3(Mean - Median)}{Std.Dev.}$ **If** S_{kp} **= 0, the given distribution is symmetric,** S_{kp} **> 0, the given distribution is positively skewed.** S_{kp} **< 0. the given distribution is negatively skewed.**	
Bowley's Measure of Skewness	**Skewness =** $Skewness = (Q_3 - Q_2) - (Q_2 - Q_1)$	
Bowley's Coefficient of Skewness	$S_B = \frac{(Q_3 + Q_1 - 2Q_2)}{Q_3 - Q_1}$, **If** S_B **= 0, the given distribution is symmetric.** S_B **> 0, the given distribution is positively skewed.** S_B **< 0, the given distribution is Negatively skewed.**	

Sr. No.	Measure of Skewness and Kurtosis	Remark
Karl Pearson's Moment coefficient of skewness	$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$, $\gamma_1 = \pm\sqrt{\beta_1}$ If, $\gamma_1 = 0$, the given distribution is symmetric. $\gamma_1 > 0$, the given distribution is positively skewed. $\gamma_1 < 0$, the given distribution is negatively skewed.	
Graphical method		
Coefficient of kurtosis	$\beta_2 = \frac{\mu_4}{\mu_2^2}$, $\gamma_2 = \beta_2 - 3$ If, $\gamma_2 = 0$, the given distribution is Mesokurtic $\gamma_2 > 0$, the given distribution is leptokurtic $\gamma_2 < 0$, the given distribution is platykurtic	

Sr,. No.	Properties of statistical measures
1.	**Effect of Change of Scale & Change of origin on statistical measures**
	If, $U = \frac{X-a}{h}, \quad V = \frac{Y-b}{k}$
Mean	$\overline{X} = h \times \overline{U} + a, \quad \overline{Y} = k \times \overline{V} + b$ **i.e. mean is not independent of change of origin as well as change of scale.**
Standard Deviation	$\sigma_X = h \times \sigma_U$, $\sigma_Y = k \times \sigma_V$ **i.e. standard deviation is independent of change of origin but not the scale.**
Variance	$\sigma_X^2 = h^2 \times \sigma_U$, $\sigma_Y^2 = k^2 \times \sigma_V$ **i.e. variance is independent of change of origin but not the scale.**
Moments	$\mu_r \; of \; X = h^r \times \mu_r \; of \; U \quad and \quad \mu_r \; of \; Y = h^r \times \mu_r \; of \; V$ **,** **i.e. central moments are independent of change of origin but not the scale.**
Mean	**Sum of deviations taken from mean is always equal to zero.** **i.e.** $\sum_{i=1}^{n} (X_i - \overline{X}) = 0$
Median	**Sum of squares of deviations taken from median is always minimum.** **i.e.** $\sum_{i=1}^{n} (X_i - median)^2 \; is$ **Minimum.**

Requisites of good statistical measures.	
Central Tendency/ Dispersion	1. It should be rigidly defined. 2. It should be easy to understand and calculate. 3. It should be based on all observations. 4. It should be capable of further mathematical treatment. 5. It should be least affected by extreme observations. 6. It should be least affected by fluctuations of sampling.
Merits and demerits of Statistical measures.	
Mean	Merits: 1. It is rigidly defined. 2. It is easy to calculate and understand. 3. It is based on all observations. 4. It is capable of further mathematical treatment. Demerits: 1. It is too sensitive to extreme observations. 2. It can not be calculated in case of open end classes. 3. It can not be determined by inspection nor can it be located graphically. 4. It can not be used with qualitative type of data.
Median	Merits: 1. It is rigidly defined. 2. It is not affected by extreme observations. 3. It can be calculated in case of open end classes. 4. It can be located graphically. 5. It can be calculated while dealing with qualitative type of data. Demerits: 1. It is not based on all observations. 2. It is not suitable for further mathematical treatment. 3. It is affected more by fluctuations of sampling.
Mode	Merits: 1. Mode is easy to calculate and understand.

<table>
<tr><td></td><td colspan="2">2. It is not affected by extreme observations.
3. It can be obtained in case of open end classes.
Demerits:
1. It is sometimes ill defined.
2. It is not based on all observations.
3. It is not capable of further mathematical treatment.
4. Mode is affected to a greater extent by fluctuations of sampling.</td></tr>
<tr><td>Range</td><td colspan="2">Merits:
1. It is simplest measure of dispersion.
2. It is rigidly defined.
Demerits:
1. It is not based on all observations.
2. It is very much affected by fluctuations of sampling.
3. It is unreliable measure of dispersion.
4. It can not be used while dealing with open end classes.
5. It is not suitable of further mathematical treatment.</td></tr>
<tr><td>Types of Measures of Dispersion</td><td><u>Absolute Measure of Dispersion</u>
1. The measures of dispersion which are expressed in terms of original units of data are termed as absolute measures.
2. Such measures are not suitable for comparing the variability of two distributions which are expressed in different units of measurement.
3. e.g. Range, Quartile deviation, Standard deviation, Variance etc.</td><td><u>Relative Measure of Dispersion</u>
1. Relative measures of dispersion are obtained as ratios or percentages and thus pure numbers independent of units of measurement.
2. For comparing variability of two distributions, relative measure of dispersion can be used.
3. e.g. Co-efficient of range, Co-efficient of quartile deviation, co-efficient of variance etc.</td></tr>
<tr><td>Quartile Deviation</td><td colspan="2"></td></tr>
<tr><td></td><td colspan="2"></td></tr>
</table>

