Md. Abdul Alim

Chapter: Measures of Dispersion –4

বিচ্ছুরণ

Dispersion: Literal meaning of dispersion is scatteredness. Dispersion is the degree of the scatter or variation of the variables about a central value. A measure of variation is designed to state the extent to which the individual measures differ on an average from the mean.

Example: Let the scores of two batches each of size 4, be as follows:

Batch I	49	50	50	51
Batch II	0	0	100	100

The average score for each batch is 50 the average respondents the first set of scores very well but it is hardly typical of the second set. This is because the values in the first set bunch around the average but the values in the second set are widely scattered. Thus although the two sets have the same mean, they are markedly different in their variability or dispersion. This means that although the two sets of data are quite different in nature, the measure of location has failed to bring out this difference. The need for a measure of dispersion in addition to a measure of location is thus obvious.

Different measures of Dispersion: There are several methods of measuring dispersion. These measures can be divided into two groups:

- Absolute measure
- Relative measure

Absolute measure: Absolute measures of variation are expressed in the same statistical unit in which the original data are given such as rupees, kilograms, tones etc. These values may be used to compare the variation in two or more than two distributions provided the variables are expressed in the same units and have almost the same average value.
Following are the absolute measures of variation or dispersion

- Range
- Quartile Deviation
- Mean Deviation
- Standard Deviation

Relative Measures of Variation: A measure of relative variation is the ratio of a measure of absolute variation to an average. Relative measures may also be used to compare the relative accuracy of data.

Following are the relative measures of variation:

- Co-efficient of range
- Co-efficient quartile deviation
- Co-efficient of mean deviation
- Co-efficient of standard deviation
- Co-efficient of variation

Range: Range is the simplest method of studying variation. It is defined as the difference between the value of the smallest observation and the value of the largest observation included in the distribution. Symbolically, $R = L - S$

where, L = Largest value and S = Smallest value

The relative measures corresponding to range, called the co-efficient of range, is obtained by applying the following formula

$$\text{Co-efficient of Range} = \frac{L-S}{L+S}$$

In a frequency distribution, range is calculated by taking the difference between the lower limit of the lower class and the upper limit of the highest class.

Example # 1: The following are the points of shares of a company from Monday to Saturday:

Day	Prices (Tk.)
Monday	200
Tuesday	210
Wednesday	208
Thursday	160
Friday	220
Saturday	250

Calculate range and co-efficient of range.

Solution: We know that, Range $R = L - S$

Here, $L = 250$ and $S = 160$

$\therefore$ Range = 250-160 = Tk. 90

$$\text{Co-efficient of Range} = \frac{L-S}{L+S} = \frac{250-160}{250+160} = 0.219$$

Example # 2: Calculate coefficient of range and range from the following data:

Profits (Tk. lakhs)	No. of Companies.
10-20	8
20-30	10
30-40	12
40-50	8
50-60	4

Solution: In a frequency distribution, range is calculated by taking the difference between the lower limit of the lower class and the upper limit of the highest class.

Range $= L - S = 60 - 10 = 50$

$$\text{Co-efficient of Range} = \frac{L-S}{L+S} = \frac{60-10}{60+10} = \frac{50}{70} = 0.714$$

- It is the simplest measure of dispersion.
- It is easy to calculate and easy to understand.
- It gives us a quick idea of the variability of the observations involving least amount of time and calculations.

Limitations

- Range is not based on each and every observation of the distribution.
- Range cannot be computed in case of open-end distributions.
- Range cannot tell us extreme observations. For example, observe the following series:

Series A	6	46	46	46	46	46	46	46
Series B	6	6	6	6	46	46	46	46
Series C	6	10	15	25	30	32	40	46

In all the series range is the same (i.e. 46-6 = 40) but it doesn't mean that the distributions are alike. Range is, therefore, more unreliable as a guide to the variation of the values within a distribution.

Uses of Range: Despite serious limitations range is useful in the following cases:

✓ **Quality control**

The object of quality control is to keep a check on the quality of the product without 100% inspection.

✓ **Fluctuations in the share prices**

Range is useful in studying the variations in the price of stocks and shares and other commodities etc. They are very sensitive to price changes from one period to another. For example, by computing range we can get an idea about the range of variation of, say, gold prices. If the minimum price for 10 gm during 1989-90 was Tk. 3010 and the maximum price Tk. 3350 this at once tell us about the range of variation

i.e. 3350-3010 = 340.

✓ **Weather forecasts**

The meteorological department does make use of the range in determining the difference between the minimum temperature and maximum temperature. This information is of great concern to the general public because they know as to within what limits the temperature is likely to vary on a particular day.

Md. Moyazzem Hossain
Lecturer, Dept. of Statistics, JU

Mean deviation: Mean deviation is obtained by calculating the absolute deviations of each observation from mean (or median or mode) and then averaging these deviations by taking their arithmetic mean.

Let $X_1, X_2, \ldots, X_n$ be n observations of a variable with mean ($\bar{x}$), median (M_e) and Mode (M_o) then mean deviation is defined by:

For Ungrouped Data

$$M.D_{(\bar{x})} = \frac{1}{n}\sum|x-\bar{x}|$$

$$M.D_{(M_e)} = \frac{1}{n}\sum|x-M_e|$$

$$M.D_{(M_o)} = \frac{1}{n}\sum|x-M_o|$$

$$M_e = L + \frac{\frac{n}{2} - p.c.f}{f} \times h$$

$$M_o = L + \frac{\Delta_1}{\Delta_1+\Delta_2} \times h.$$

For grouped Data

$$M.D_{(\bar{x})} = \frac{1}{n}\sum f|x-\bar{x}|$$

$$M.D_{(M_e)} = \frac{1}{n}\sum f|x-M_e|$$

$$M.D_{(M_o)} = \frac{1}{n}\sum f|x-M_o|$$

Where, $n = \sum f$

Co-efficient of Mean deviation (C.M.D): Co-efficient of mean deviation is the ratio of the mean deviation measured from certain measure of central location to the corresponding measure of central location and is defined as follows:

$$C.M.D_{(\bar{x})} = \frac{M.D_{\bar{x}}}{\bar{x}}$$

$$C.M.D_{(M_e)} = \frac{M.D_{(M_e)}}{M_e}$$

$$C.M.D_{(M_o)} = \frac{M.D_{(M_o)}}{M_o}$$

Problem: Calculate the mean deviation from

(a) Arithmetic mean
(b) Mode
(c) Median

In respect of the marks obtained by nine students given below and show that the mean deviation from median is minimum.

Marks (Out of 25)	7	4	10	9	15	12	7	9	7

Md. Moyazzem Hossain
Lecturer, Dept. of Statistics, [illegible]

Solution: Calculation of mean deviation from mean, median and mode:

We know that,

$$M.D_{(\bar{x})} = \frac{1}{n}\sum|x-\bar{x}|$$

$$M.D_{(M_e)} = \frac{1}{n}\sum|x-M_e|$$

$$M.D_{(M_o)} = \frac{1}{n}\sum|x-M_o|$$

Now, Mean $\bar{x} = \frac{\sum x_i}{n} = \frac{80}{9} = 8.89$

For calculating median the items have to be arranged

Marks (Out of 25)	4	7	7	7	9	10	10	12	15

$\therefore$ Median = size of $\frac{n+1}{2}$ th item = $\frac{9+1}{2}$ = 5th item

Here, Size of 5th item = 9.

Hence, median = 9.

Mode = 7 (since 7 is repeated the maximum number of items i.e. 3)

Calculation of Mean Deviation

Marks x	**Deviations from Mean** = $\lvert x_i-\bar{x}\rvert$	**Deviations from Median** = $\lvert x_i-Me\rvert$	**Deviations from Mode** = $\lvert x_i-Mo\rvert$
7	1.89	2	0
4	4.89	5	3
10	1.11	1	3
9	0.11	0	2
15	6.11	6	8
12	3.11	3	5
7	1.89	2	0
9	0.11	0	2
7	1.89	2	0
Total	$\sum\lvert x_i-\bar{x}\rvert = 21.11$	$\sum\lvert x_i-M_e\rvert = 21$	$\sum\lvert x_i-M_o\rvert = 23$

Now we have,

$$M.D_{(\bar{x})} = \frac{1}{n}\sum|x-\bar{x}| = \frac{1}{9}\times 21.11 = 2.34$$

$$M.D_{(M_e)} = \frac{1}{n}\sum|x-M_e| = \frac{1}{9}\times 21 = 2.33$$

$$M.D_{(M_o)} = \frac{1}{n}\sum|x-M_o| = \frac{1}{9}\times 23 = 2.56$$

Note: From these calculations it is clear that the mean deviation is least from median.

Md. Moyazzem Hossain

For grouped Data

$$M.D_{(\bar{x})} = \frac{1}{n}\sum f|x-\bar{x}|$$

$$M.D_{(M_e)} = \frac{1}{n}\sum f|x-M_e|$$

$$M.D_{(M_o)} = \frac{1}{n}\sum f|x-M_o|$$

where, $n=\sum f$

Problem: Calculate mean deviation (taking deviations from mean) from the following data:

x	2	4	6	8	10
f	1	4	6	4	1

Solution: We know that, $M.D_{(\bar{x})} = \frac{1}{n}\sum f|x-\bar{x}|$

Now, $\bar{x} = \frac{\sum fx}{n}$

By using calculator we get, $\sum fx = 96$ and $n = \sum f = 16$.

$$\therefore \bar{x} = \frac{\sum fx}{n} = \frac{96}{16} = 6$$

Calculation of mean deviation

x	f	**Deviations from Mean** $=\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$
2	1	4	4
4	4	2	8
6	6	0	0
8	4	2	8
10	1	4	4
Total			$\sum f\lvert x-\bar{x}\rvert = 24$

$$\therefore M.D_{(\bar{x})} = \frac{1}{n}\sum f|x-\bar{x}| = \frac{1}{16}\times 24 = 1.5$$

Problem: Age distribution of hundred life insurance policy holders is as follows:

Age as on nearest birthday	Number
17-20	9
20-26	16
26-36	12
36-41	26
41-51	14
51-56	12
56-61	6
61-71	5

Calculate mean deviation from median age.

Md. Moyazzem Hossain
Lecturer, Dept. of Statistics, JU

Solution: We know that, $M.D_{(M_e)} = \frac{1}{n}\sum f|x - M_e|$

Calculation of mean deviation

Age as on nearest birthday	Number f	$c.f$	Mid point x	Deviations from Median $= \|x - M_e\|$	$f\|x - M_e\|$
17-20	9	9	18.5	19.75	177.75
20-26	16	25	23	15.25	244.00
26-36	12	37	31	7.25	87.00
36-41	26	63	38.5	0.25	6.50
41-51	14	77	46	7.75	108.50
51-56	12	89	53.5	15.25	183.00
56-61	6	95	58.5	20.25	121.50
61-71	5	100	66	27.75	138.75
Total	$\sum f_i = 100$				$\sum f\|x - M_e\| = 1067$

We know that, Median $= L + \frac{\frac{N}{2} - p.c.f}{f} \times i$

Here, $\frac{N}{2} = \frac{100}{2} = 50$

Here median class is 36-41. $\therefore L = 36$, $p.c.f = 37$, $f = 26$, $i = 5$

So we have, Median $= 36 + \frac{50 - 37}{26} \times 5 = 38.25$

$\therefore M.D_{(M_e)} = \frac{1}{n}\sum f|x - M_e| = \frac{1}{100} \times 1067 = 10.67$

Merits

- It is easy to understand.
- It is relatively easy to calculate.
- It takes all the observations into account.
- It is less affected by the extreme values.

Demerits

- It is not amenable to further algebraic treatment.
- It cannot be calculated if the extreme classes of the frequency distribution are open.
- It is less stable than standard deviation.

Uses: Because of its simplicity in meaning and computation it is especially effective in reports presented to the general public or to groups not familiar with statistical methods. This measure is useful for small samples with no elaborately analysis required. Incidentally it may be mentioned that the National Bureau of Economic Research has found in its work on forecasting business cycles, that the mean deviation in the most practical measure of variation to use for this purpose.

Md. Moyazzem Hossain
Lecturer, Dept. of Statistics, JU

Standard deviation: Standard deviation may be defined as the positive square root of the arithmetic mean of the squares of deviations of given observations from their arithmetic mean.

For ungrouped data: Let $x_1, x_2, \ldots, x_n$ denote n values of a variable x. The standard deviation denoted by S_x is defined as

$$S_x = \sqrt{\frac{\sum (x-\bar{x})^2}{n}} = \sqrt{\left\{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2\right\}} = \sqrt{\left\{\frac{\sum x^2}{n} - \left(\bar{x}\right)^2\right\}} \text{ where, } \bar{x} = \frac{\sum x}{n}.$$

Problem: Find the standard deviation from the weekly wages of ten workers working in a factory:

Workers	wages	workers	wages
A	320	F	340
B	310	G	325
C	315	H	321
D	322	I	320
E	326	J	331

Solution: We know that,

$$S_x = \sqrt{\left\{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2\right\}}$$

By using calculator we get, $\sum x^2 = 1043912$, $\sum x = 3230$ and $n = 10$.

So that, $S_x = \sqrt{\left\{\frac{1043912}{10} - \left(\frac{3230}{10}\right)^2\right\}} = 7.886 = 7.89$

For grouped data: If $x_1, x_2, \ldots, x_n$ occurs with frequency $f_1, f_2, \ldots, f_n$ respectively, the standard deviation is defined as

$$S_x = \sqrt{\frac{\sum f\left(x-\bar{x}\right)^2}{N}} = \sqrt{\left\{\frac{\sum fx^2}{N} - \left(\frac{\sum fx}{N}\right)^2\right\}} = \sqrt{\left\{\frac{\sum fx^2}{N} - \left(\bar{x}\right)^2\right\}}$$

where $N = \sum f$ and $\bar{x}$ is the arithmetic mean of the distribution.

$$* \; S_x = \sqrt{\frac{\sum(x-\bar{x})^2}{n}} = \sqrt{\frac{1}{n}\left(\sum x^2 - 2\cdot\bar{x}\sum x + \bar{x}^2\cdot\sum 1\right)}$$

$$= \sqrt{\frac{\sum x^2}{n} - \frac{1}{n}\left\{2(\sum x)^2 - n\cdot\left(\frac{\sum x}{n}\right)^2\right\}}$$

$$= \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

Problem: An analysis of production rejects resulted in the following figures:

No. of rejects per operator	No. of operators
21-25	5
26-30	15
31-35	28
36-40	42
41-45	15
46-50	12
51-55	3

Solution: We know that $S_x = \sqrt{\left\{\frac{\sum fx^2}{N} - \left(\frac{\sum fx}{N}\right)^2\right\}}$ where, $N = \sum f$.

Calculation of Standard Deviation

No. of rejects per operator	No. of operators f	Mid point x	x^2	fx	fx^2
21-25	5	23	529	115	2645
26-30	15	28	784	420	11760
31-35	28	33	1089	924	30492
36-40	42	38	1444	1596	60648
41-45	15	43	1849	645	27735
46-50	12	48	2304	576	27648
51-55	3	53	2809	159	8427
Total	120		10808	4435	169355

From the data by using calculator we get,

$\sum fx = 4435$, $N = \sum f = 120$ and $\sum fx^2 = 169355$

$$\therefore S_x = \sqrt{\left\{\frac{169355}{120} - \left(\frac{4435}{120}\right)^2\right\}} = 6.7359 = 6.74$$

Merits

- The standard deviation is the best measure of variation because of its mathematical characteristics. It is based on every item the distribution.

- It is possible to calculate the combined standard deviation of two or more groups. This is not possible with any other measure.
- For comparing the variability of two or more distributions coefficient of variation is considered to be most appropriate and this measure is based on mean and standard deviation.
- Standard deviation is most prominently used in further statistical work. For example, in comparing skewness, correlation etc., and use in made of standard deviation.

Limitation

- As compared to other measures it is difficult to compute.
- It gives more weight to extreme values and less to those which are near the mean.

Theorem 1: Standard deviation is independent on change of origin but not of scale.

Proof: Let $x_1, x_2, \ldots, x_k$ be the mid values of the classes of a frequency distribution and let $f_1, f_2, \ldots, f_k$ be their corresponding frequencies. Also let $u_i = \frac{x_i - A}{h}$ where, u_i, A and h are changed variate, origin and scale respectively.

Now, $x_i = hu_i + A$ or $\bar{x} = h\bar{u} + A$.

We know that, $S_x = \sqrt{\frac{1}{n}\sum f_i (x_i - \bar{x})^2}$. Putting the value of x_i and $\bar{x}$ we get,

$$S_x = \sqrt{\frac{1}{n}\sum f_i (x_i - \bar{x})^2}$$

$$= \sqrt{\frac{1}{n}\sum f_i (hu + A_i - h\bar{u} - A)^2} = \sqrt{\frac{h^2}{n}\sum f_i (u_i - \bar{u})^2} = \sqrt{h^2 S_u^2}$$

where, S_u^2 is the variance of u variate.

$\therefore S_x = |hS_u|$ showing that standard deviation is independent on change of origin but not of scale.

Theorem 2: Standard deviation is the least possible root mean square deviation i. e. root mean square deviation is the least when the deviations are taken from the arithmetic mean.

Proof: Let $x_1, x_2, \ldots, x_k$ be the mid values of the classes of a frequency distribution and let $f_1, f_2, \ldots, f_k$ be their corresponding frequencies. Also let $\bar{x}$ be the mean of the observations and A be any arbitrary value.

Now we have, $S'^2 = \frac{1}{n}\sum f_i (x_i - A)^2 = \frac{1}{n}\sum f_i \left[(x_i - \bar{x}) + (\bar{x} - A)\right]^2$

$$S'^2 = \frac{1}{n}\left[\sum f_i (x_i - \bar{x})^2 + 2(\bar{x} - A)\sum f_i (x_i - \bar{x}) + n(\bar{x} - A)^2\right]$$

Md. Moyazzem Hossain
Lecturer, Dept. of Statistics, IU

$\sqrt{\frac{1}{n}\sum f_i (x_i - \cancel{\bar{x}})^2} \leq \sqrt{\frac{1}{n}\sum f (x_i - A)^2}$

$$S'^2 = \frac{1}{n}\sum f_i(x_i - \bar{x})^2 + 0 + \text{ a positive value}$$

$$S'^2 = S^2 + \text{ a positive value}$$

Therefore, $S'^2 > S^2$ i. e. $S' > S$. Hence the theorem is proved.

Theorem 3: For two observations, standard deviation is the half of the range.

Proof: Let X_1 and X_2 be two observations. Then $\bar{X} = \frac{(X_1 + X_2)}{2}$ where $\bar{X}$ is the arithmetic mean. Let S denote the standard deviation.

We have,

$$S^2 = \frac{1}{2}\left[\left\{X_1 - \frac{(X_1+X_2)}{2}\right\}^2 + \left\{X_2 - \frac{(X_1+X_2)}{2}\right\}^2\right]$$

$$= \frac{1}{2}\left[\left(\frac{X_1 - X_2}{2}\right)^2 + \left(\frac{X_2 - X_1}{2}\right)^2\right] = \left(\frac{X_1 - X_2}{2}\right)^2$$

Therefore, $S = \left|\frac{X_1 - X_2}{2}\right|$. Hence proved.

Theorem 4: The standard deviation of first n natural number is $\sqrt{\frac{(n^2-1)}{12}}$.

Proof: The n natural numbers are 1, 2, 3, ..., n.

Their mean,

$$\bar{X} = \frac{1+2+3+\ldots+n}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}$$

Now, $\sum X^2 = 1^2 + 2^2 + 3^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}$.

We know,

$$S^2 = \frac{\sum X^2}{n} - \bar{X}^2 = \frac{n(n+1)(2n+1)}{6n} - \left\{\frac{(n+1)}{2}\right\}^2$$

$$= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{(n+1)}{2}\left\{\frac{2n+1}{3} - \frac{n+1}{2}\right\} = \frac{(n+1)(n-1)}{12} = \frac{n^2-1}{12}$$

therefore, $S = \sqrt{\frac{(n^2-1)}{12}}$. Hence proved.

Theorem 5: If $\bar{X}$ and S be the arithmetic mean and standard deviation respectively of n non-negative observations, then $\bar{X}\sqrt{(n-1)} \geq S$.

Proof: Let us consider $X_1, X_2, \ldots, X_n$ be n non-negative observations.

We know $\sum_{i=1}^{n} X_i = n\bar{X}$ and $nS^2 = \sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}$.

Now $\left(\sum_{i=1}^{n} X_i\right)^2 = \sum_{i=1}^{n} X_i^2 + \sum_{i \neq j} X_i X_j$. Since X_i's are non-negative, $\sum_{i \neq j}^{n} X_i X_j \geq 0$.

2

Hence, $\left(\sum_{i=1}^{n} X_i\right)^2 \geq \sum_{i=1}^{n} X_i^2$. Subtracting $\frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}$ from both sides, we have,

$$\left(\sum_{i=1}^{n} X_i\right)^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n} \geq \sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}$$

or $\frac{(n-1)}{n}.(n\bar{X})^2 \geq nS^2$

or $(n-1).\bar{X}^2 \geq S^2$

or $\sqrt{(n-1)}\bar{X} \geq S$. Hence proved.

Co-efficient of Standard Deviation: Co-efficient of standard deviation is defined by

$$C.S.D = \frac{S.D}{Mean} = \frac{\sigma}{\bar{x}} = \frac{s}{\bar{x}}$$

Co-efficient of Variation: The relative measure of dispersion based upon standard deviation is called co-efficient of standard deviation. The co-efficient of standard deviation multiplied by 100 gives the co-efficient of variation.

Thus, Co-efficient of variation (C.V)

$$= \frac{\text{Standard deviation}}{\text{Mean}} \times 100$$

$$= \frac{\sigma}{\bar{x}} \times 100 \text{ where, } \sigma \text{ and } \bar{x} \text{ are both measured in the same units.}$$

Note: Co-efficient variation is more useful when the two distributions are entirely different and the units of measurement are also different. Co-efficient of variation being a pure number is independent of the units of measurement and thus is suitable for comparing the variability, homogeneity and uniformly of two or more distributions. The series having greater C.V is said to be more variable than the other and the series having lesser C.V is said to be more consistent (or homogenous) than the other. Co-efficient of variation is, however unreliable if $\bar{x}$ is near to zero.

Example: Calculate co-efficient of variation from the following data:

Profits (Rs. crores)	10-20	20-30	30-40	40-50	50-60
No. of Companies	8	12	20	6	4

Solution: We know that,

$$C.V = \frac{\sigma}{\bar{x}} \times 100 \text{ and } \bar{x} = \frac{\sum fx}{N}, \sigma = \sqrt{\left\{\frac{\sum fx^2}{N} - \left(\frac{\sum fx}{N}\right)^2\right\}}$$

Calculation of coefficient of variation

Profits (Tk. crores)	Midpoint (x)	No. of cost (f)	x^2	fx	fx^2
10-20	15	8	225	120	1800
20-30	25	12	625	300	7500
30-40	35	20	1225	700	24500
40-50	45	6	2025	270	12150
50-60	55	4	3025	220	12100
Total		50	7125	1610	58050

By using calculator we get $\sum fx = 1610$, $\sum fx^2 = 58050$ and $N = \sum f = 50$.

$$\therefore \bar{x} = \frac{\sum fx}{N} = \frac{1610}{50} = 32.2$$

$$\therefore \sigma = \sqrt{\left\{\frac{\sum fx^2}{N} - \left(\frac{\sum fx}{N}\right)^2\right\}} = \sqrt{\left\{\frac{58050}{50} - \left(\frac{1610}{50}\right)^2\right\}} = 11.14$$

$$\therefore \text{Co-efficient of variation} = \frac{\sigma}{\bar{x}} \times 100 = \frac{11.14}{32.2} \times 100 = 34.596 = 34.6\%$$

Example: A purchasing agent obtained samples of 60 watt bulbs from two companies. He had the samples tested in his own laboratory for length of life with the following result:

Length of life (in hours)	Company A	Company B
1700-1900	10	3
1900-2100	16	40
2100-2300	20	12
2300-2500	8	3
2500-2700	6	2

a) Which company's bulbs do you think are better?
b) If prices of both types are the same, which company's bulbs would you buy?

...tion: Calculation of Mean and co-efficient of variation

...gth of life (in hours)	Midpoint	Company A	Company B
1700-1900	1800	10	3
...0-2100	2000	16	40
...00	2200	20	12
[illegible]	2400	8	3
[illegible]	2600	6	2

Md. Moyazzem Hossain
Lecturer, Dept. of Statistics, JU

For Company A: By using calculator we get, $\sum fx = 128800$, $\sum fx^2 = 279840000$ and $N = \sum f = 60$

$$\therefore \bar{x} = \frac{\sum fx}{N} = \frac{128800}{60} = 2146.67 \text{ and}$$

$$\sigma = \sqrt{\left\{\frac{\sum fx^2}{N} - \left(\frac{\sum fx}{N}\right)^2\right\}}$$

$$= \sqrt{\left\{\frac{279840000}{60} - \left(\frac{128800}{60}\right)^2\right\}} = 236.267 = 236.27$$

$$\therefore \text{Co-efficient of variation} = \frac{\sigma}{\bar{x}} \times 100 = \frac{236.27}{2146.67} \times 100 = 11.00 = 11\%$$

For Company B: By using calculator we get

$$\sum fx = 124200,\ \sum fx^2 = 258600000 \text{ and } N = \sum f = 60$$

$$\therefore \bar{x} = \frac{\sum fx}{N} = \frac{124200}{60} = 2070 \text{ and}$$

$$\sigma = \sqrt{\left\{\frac{\sum fx^2}{N} - \left(\frac{\sum fx}{N}\right)^2\right\}}$$

$$= \sqrt{\left\{\frac{258600000}{60} - \left(\frac{124200}{60}\right)^2\right\}} = 158.429 = 158.43$$

$$\therefore \text{Co-efficient of variation} = \frac{\sigma}{\bar{x}} \times 100 = \frac{158.43}{2070} \times 100 = 7.65\%$$

(a) Since average is higher in case of company A, hence bulbs of company A are better.

(b) Co-efficient of variation is less for company B. Hence if prices are same we will prefer to buy Company's B's bulbs.

Md. Mozammel Hossain
Lecturer, Dept. of Statistics, [illegible]