Solution to Sample Question-1

Q1. The mean and standard deviation of 100 items are found to be 40 and 10 respectively. If at the time of calculations two items are wrongly taken as 30 and 70 instead of 3 and 27, find the correct mean and standard deviation.

Incorrect:

$$\sum x = 40 \times 100 = 4000$$

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

$$\Rightarrow 10^2 = \frac{\sum x^2}{100} - \left(\frac{4000}{100}\right)^2$$

$$\Rightarrow \sum x^2 = (10^2 + 1600) \times 100 = 170000$$

Correct:

$$\sum x = 4000 - 30 - 70 + 3 + 27 = 3930$$

$$\bar{x} = \frac{3930}{100} = 39.3$$

$$\sum x^2 = 170000 - 30^2 - 70^2 + 3^2 + 27^2 = 164938$$

$$\sigma = \sqrt{\frac{164938}{100} - \left(\frac{3930}{100}\right)^2} = \sqrt{104.89} = 10.24$$

Q2. The mean and standard deviation calculated from 20 observations are 15 and 10 respectively. If an additional observation 36, left out through oversight, be included in the calculations, find the correct mean and standard deviation.

Incorrect:

$$\sum x = 15 \times 20 = 300$$

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

$$\Rightarrow 10^2 = \frac{\sum x^2}{20} - \left(\frac{300}{20}\right)^2$$

$$\Rightarrow \sum x^2 = (10^2 + 225) \times 20 = 6500$$

Correct:

$$\sum x = 300 + 36 = 336$$

$$\bar{x} = \frac{336}{21} = 16.0$$

$$\sum x^2 = 6500 + 36^2 = 7796$$

$$\sigma = \sqrt{\frac{7796}{21} - \left(\frac{336}{21}\right)^2} = \sqrt{115.23} = 10.73$$

Q3. (a) For a group of 200 candidates, the mean and standard deviation of scores were found to be 40 and 15 respectively. Later on it was discovered that the scores 43 and 35 were misread as 34 and 53 respectively. Find the corrected mean and standard deviation corresponding to the corrected figures.

(a) Calculate coefficient of variation for a series for which the following results are known :

$N = 50,\ \sum d = -10,\ \sum d^2 = 404$ where d = deviation of items from assumed mean 75.

(a) Incorrect:

$$\sum x = 40 \times 200 = 8000$$

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

$$\Rightarrow 15^2 = \frac{\sum x^2}{200} - \left(\frac{8000}{200}\right)^2$$

$$\Rightarrow \sum x^2 = (15^2 + 40^2) \times 200 = 365000$$

Correct:

$$\sum x = 8000 - 34 - 53 + 43 + 35 = 7991$$

$$\overline{x} = \frac{7991}{200} = 39.96$$

$$\sum x^2 = 365000 - 34^2 - 53^2 + 43^2 + 35^2 = 364109$$

$$\sigma = \sqrt{\frac{364109}{200} - \left(\frac{7991}{200}\right)^2} = \sqrt{224.14} = 14.97$$

(b) $$\sigma = \sqrt{\overline{d^2} - \overline{d}^2} = \sqrt{\frac{404}{50} - \left(\frac{-10}{50}\right)^2} = \sqrt{8.04} = 2.84$$

$$\overline{x} = A + \overline{d} = 75 + \frac{-10}{50} = 74.8$$

$$CV = \frac{\sigma}{\overline{x}} \times 100 = \frac{2.84}{74.8} \times 100 = 3.79\%$$

Q4. Calculate the first four moments about the mean from the following data:

Class mark (x):	61	64	67	70	73
Frequency (f):	5	18	42	27	8

Solution

X	u	f	fu	fu^2	fu^3	fu^4
61	-2	5	-10	20	-40	80
64	-1	18	-18	18	-18	18
67	0	42	0	0	0	0
70	1	27	27	27	27	27
73	2	8	16	32	64	128
	$\sum =$	100	15	97	33	253

Raw moments are,

$$\mu_1' = \frac{c \times \sum fu}{N} = \frac{3 \times 15}{100} = 0.45 \qquad \mu_3' = \frac{c^3 \times \sum fu^3}{N} = \frac{3^3 \times 33}{100} = 8.91$$

$$\mu_2' = \frac{c^2 \times \sum fu^2}{N} = \frac{3^2 \times 97}{100} = 8.73 \qquad \mu_4' = \frac{c^4 \times \sum fu^4}{N} = \frac{3^4 \times 253}{100} = 204.93$$

Moments about the mean,

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - \mu_1'^2 = 8.73 - (0.45)^2 = 8.5275$$

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + \mu_1'^3 = 8.91 - 3(0.45)(8.73) + 2(0.45)^3 = 2.6932$$

$$\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6\mu_1'^2\mu_2' - 3\mu_1'^4 = 204.93 - 4(0.45)(8.91) + 6(0.45)^2(8.73) - 3(0.45)^4 = 199.3759$$

Q4. Calculate the first four moments about the mean from the following data:

Class mark (X):	61	64	67	70	73
Frequency (f):	5	18	42	27	8

Alternative Solution

X	f	fX	$X-\overline{X}$	$f(X-\overline{X})$	$f(X-\overline{X})^2$	$f(X-\overline{X})^3$	$f(X-\overline{X})^4$
61	5	305	-6.45	-32.25	208.01	-1341.68	8653.84
64	18	1152	-3.45	-62.1	214.24	-739.14	2550.05
67	42	2814	-0.45	-18.9	8.505	-3.82	1.72
70	27	1890	2.55	68.85	175.56	447.69	1141.62
73	8	584	5.55	44.4	246.42	1367.63	7590.35
$\sum =$	100	6745	-2.25	0	852.75	-269.325	19937.5931

Mean, $\overline{X} = \dfrac{\sum fX}{N} = \dfrac{6745}{100} = 67.45$

Moments about the mean,

$$\mu_1 = \frac{\sum f(X-\overline{X})}{N} = \frac{0}{100} = 0$$

$$\mu_2 = \frac{\sum f(X-\overline{X})^2}{N} = \frac{852.75}{100} = 8.5275$$

$$\mu_3 = \frac{\sum f(X-\overline{X})^3}{N} = \frac{-269.325}{100} = -2.6932$$

$$\mu_4 = \frac{\sum f(X-\overline{X})^4}{N} = \frac{19937.5931}{100} = 199.3759$$

Q5. Establish the relation between the first four moments about the mean μ_r and the moments about an arbitrary origin μ'_r (raw moments). Prove that

(a) $\mu_2 = \mu'_2 - (\mu'_1)^2$

(b) $\mu_3 = \mu'_3 - 3\mu'_1\mu'_2 + 2(\mu'_1)^3$

(c) $\mu_4 = \mu'_4 - 4\mu'_1\mu'_3 + 6(\mu'_1)^2\mu'_2 - 3(\mu'_1)^4$

Solution

(a) $$\begin{aligned}\mu_2 = \overline{(X-\overline{X})^2} = \overline{(d-\overline{d})^2} &= \overline{d^2 + 2d\overline{d} + \overline{d}^2}\\ &= \overline{d^2} - 2\overline{d}^2 + \overline{d}^2\\ &= \overline{d^2} - \overline{d}^2\\ &= \mu'_2 - (\mu'_1)^2\end{aligned}$$

(b) $$\begin{aligned}\mu_3 = \overline{(X-\overline{X})^3} = \overline{(d-\overline{d})^3} &= \overline{d^3 + 3d^2\overline{d} + 3d\overline{d}^2 + \overline{d}^3}\\ &= \overline{d^3} - 3\overline{d}\,\overline{d^2} + 3\overline{d}^3 - \overline{d}^3\\ &= \overline{d^3} - 3\overline{d}\,\overline{d^2} + 2\overline{d}^3\\ &= \mu'_3 - 3\mu'_1\mu'_2 + 2\mu_1'^3\end{aligned}$$

(c) $$\begin{aligned}\mu_4 = \overline{(X-\overline{X})^4} = \overline{(d-\overline{d})^4} &= \overline{d^4 - 4d^3\overline{d} + 6d^2\overline{d}^2 - 4d\overline{d}^3 + \overline{d}^4}\\ &= \overline{d^4} - 4\overline{d}\,\overline{d^3} + 6\overline{d}^2\,\overline{d^2} - 4\overline{d}^4 + \overline{d}^4\\ &= \overline{d^4} - 4\overline{d}\,\overline{d^3} + 6\overline{d}^2\,\overline{d^2} - 3\overline{d}^4\\ &= \mu'_4 - 4\mu'_1\mu'_3 + 6\mu_1'^2\mu'_2 - 3\mu_1'^4\end{aligned}$$

Q6. (a) Prove that the standard deviation

(i) $\sigma = \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2} = \sqrt{\overline{X^2} - \overline{X}^2}$ and (ii) $\sigma = \sqrt{\overline{d^2} - \overline{d}^2}$ where $d = X - A$

(b) Use any one of the above formulas to find the standard deviation of the set of numbers 12, 6, 7, 3, 15, 10, 18, 5.

Solution

(a) (i)

$$\sigma = \sqrt{\frac{\sum (X - \overline{X})^2}{N}}$$

$$\sigma^2 = \frac{\sum (X - \overline{X})^2}{N} = \frac{\sum \left(X^2 - 2\overline{X}X + \overline{X}^2\right)}{N} = \frac{\sum X^2}{N} - \frac{2\overline{X}\sum X}{N} + \frac{\sum \overline{X}^2}{N}$$

$$= \frac{\sum X^2}{N} - \frac{2\overline{X}\sum X}{N} + \frac{N\overline{X}^2}{N} = \overline{X^2} - 2\overline{X}^2 + \overline{X}^2 = \overline{X^2} - \overline{X}^2$$

$$\therefore \sigma = \sqrt{\overline{X^2} - \overline{X}^2}$$ (Proved)

(ii)

$$\sigma = \sqrt{\frac{\sum (X - \overline{X})^2}{N}}$$

$$\sigma^2 = \overline{(X - \overline{X})^2} = \overline{(d - \overline{d})^2} = \overline{(d^2 - 2d\overline{d} + \overline{d}^2)} = \overline{d^2} - 2\overline{d}\overline{d} + \overline{d}^2 = \overline{d^2} - \overline{d}^2$$

$$\therefore \sigma = \sqrt{\overline{d^2} - \overline{d}^2}$$

(b)

$$\overline{X} = \frac{\sum X}{N} = \frac{12 + 6 + 7 + 3 + 15 + 10 + 18 + 5}{8} = \frac{76}{8} = 9.5$$

$$\overline{X^2} = \frac{\sum X^2}{N} = \frac{12^2 + 6^2 + 7^2 + 3^2 + 15^2 + 10^2 + 18^2 + 5^2}{8} = \frac{912}{8} = 114$$

$$\therefore \sigma = \sqrt{\overline{X^2} - \overline{X}^2} = \sqrt{114 - (9.5)^2} = \sqrt{23.75} = 4.87$$

Q7. From the data given bellow calculate Karl Pearson's coefficient of skewness and comment on the result:

Profits (Tk. lakhs):	10-20	20-30	30-40	40-50	50-60
No. of companies:	18	20	30	22	10

Solution

class	x	f	d	fd	fd^2
10-20	15	18	-20	-360	7200
20-30	25	20	-10	-200	2000
30-40	35	30	0	0	0
40-50	45	22	10	220	2200
50-60	55	10	20	200	4000
	$\sum =$	100	0	-140	15400

Mean, $\bar{x} = A + \frac{\sum fd}{N} = 35 + \frac{-140}{100} = 33.6$

S.D, $= \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} = \sqrt{\frac{15400}{100} - \left(\frac{-140}{100}\right)^2} = 12.33$

Mode $= L + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)c$

$= 30 + \frac{10}{10 + 8} \times 10 = 35.56$

Skewness $= \frac{\text{mean} - \text{mode}}{\text{S.D}}$

$= \frac{33.6 - 35.56}{12.33} = -0.159$

Q8. Calculate coefficient of variation (CV) and Karl Pearson's coefficient of skewness from following data:

Marks less than:	20	40	60	80	100
No. of Students:	18	40	70	90	100

Solution

class	x	cf	f	u	fu	fu^2
00-20	10	18	18	-2	-36	72
20-40	30	40	22	-1	-22	22
40-60	50	70	30	0	0	0
60-80	70	90	20	1	20	20
80-100	90	100	10	2	20	40
		$\sum =$	100	0	-18	154

Mean, $\bar{x} = 50 + \frac{-18}{100} \times 20 = 46.4$

Mode $= L + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) c = 40 + \frac{8}{8+10} \times 20 = 48.89$

S.D, $\sigma = \sqrt{\frac{\sum fu^2}{N} - \left(\frac{\sum fu}{N}\right)^2} \times c$

$= \sqrt{\frac{154}{100} - \left(\frac{-18}{100}\right)^2} \times 20 = 24.56$

Skewness $= \frac{\text{mean} - \text{mode}}{\text{S.D}} = \frac{46.4 - 48.89}{24.56} = -0.101$

CV $= \frac{\sigma}{\bar{x}} = \frac{24.56 \times 100}{46.4} = 52.92\%$

Q9. From the prices of X and Y given below, state which share is more stable in value:

X:	53	54	58	50	61	60
Y:	105	108	104	106	100	102

For X:

$\bar{x} = \frac{53 + 54 + 58 + 50 + 61 + 60}{6} = 56.0$

$\overline{x^2} = \frac{53^2 + 54^2 + 58^2 + 50^2 + 61^2 + 60^2}{6} = 3151.67$

$\sigma = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{3151.67 - (56)^2} = 3.96$

$CV = \frac{\sigma}{\bar{x}} \times 100 = \frac{3.96 \times 100}{56} = 7.07\%$

For Y:

$\bar{x} = \frac{105 + 108 + 104 + 106 + 100 + 102}{6} = 104.17$

$\overline{x^2} = \frac{105^2 + 108^2 + 104^2 + 106^2 + 100^2 + 102^2}{6} = 10857.5$

$\sigma = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{10857.50 - (104.17)^2} = 2.61$

$CV = \frac{\sigma}{\bar{x}} \times 100 = \frac{2.61 \times 100}{104.17} = 2.50\%$

Since $CV(Y) < CV(X)$ hence, share Y is more stable.

Q10. From the prices of X and Y given below, state which share is more stable in value:

X:	35	54	52	53	56	58	52	50	51	49
Y:	108	107	105	106	100	107	104	103	104	101

For X:

$\bar{x} = \frac{\sum x}{n} = \frac{510}{10} = 51.0$

$\overline{x^2} = \frac{\sum x^2}{n} = \frac{26360}{10} = 2636$

$\sigma = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{2636 - (51)^2} = 5.92$

$CV = \frac{\sigma}{\bar{x}} \times 100 = \frac{5.92 \times 100}{51} = 11.6\%$

For Y:

$\bar{x} = \frac{\sum x}{n} = \frac{1045}{10} = 104.5$

$\overline{x^2} = \frac{\sum x^2}{n} = \frac{109265}{10} = 10926.5$

$\sigma = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{10926.5 - (104.5)^2} = 2.5$

$CV = \frac{\sigma}{\bar{x}} \times 100 = \frac{2.5 \times 100}{104.5} = 2.39\%$

Since $CV(Y) < CV(X)$ hence, shares Y are more stable.

Q11. Two cricketers scored the following runs in the several innings. Find who is better run-getter and who is more consistent player:

A:	42	17	83	59	72	76	64	45	40	32
B:	28	70	31	0	59	108	82	14	3	95

Solution

For A:

$$\overline{x} = \frac{\sum x}{n} = \frac{530}{10} = 53$$

$$\overline{x^2} = \frac{\sum x^2}{n} = \frac{32128}{10} = 3212.8$$

$$\sigma = \sqrt{\overline{x^2} - (\overline{x})^2} = \sqrt{3212.8 - (53)^2} = 20.09$$

$$CV = \frac{\sigma}{\overline{x}} \times 100 = \frac{20.09 \times 100}{53} = 37.91\%$$

For B:

$$\overline{x} = \frac{\sum x}{n} = \frac{490}{10} = 49$$

$$\overline{x^2} = \frac{\sum x^2}{n} = \frac{37744}{10} = 3774.4$$

$$\sigma = \sqrt{\overline{x^2} - (\overline{x})^2} = \sqrt{3774.4 - (49)^2} = 37.06$$

$$CV = \frac{\sigma}{\overline{x}} \times 100 = \frac{37.06 \times 100}{49} = 75.63\%$$

Since $CV(A) < CV(B)$ hence, cricketer A is more consistent player.

Exr16. Compute the arithmetic mean, geometric mean, and harmonic mean of the following set of data. 3, 5, 7, 11, 14, 57

If these data were observations on the time needed to cure a disease, which mean would you think to be most appropriate?

Solution

$\therefore$ Arithmetic Mean, $\overline{X} = \dfrac{3+5+7+11+14+57}{6} = 9.7$

$\therefore$ Geometric Mean, $G = \sqrt[6]{3 \times 5 \times 7 \times 11 \times 14 \times 57} = 9.87$

$\therefore$ Harmonic Mean, $H = \dfrac{n}{\sum \frac{1}{X}} = \dfrac{6}{\left(\frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{11} + \frac{1}{14} + \frac{1}{57}\right)} = \dfrac{6}{0.856} = 7.009$

Exr17. If the weights are 2, 1, 1, 3, 1, and 2 for the numbers 3, 5, 7, 11, 14, and 57 (exercise 16), compute the weighted average and variance.

Solution

x	w	wx	wx^2
3	2	6	18
5	1	5	25
7	1	7	49
11	3	33	363
14	1	14	196
57	2	114	6498
$\sum =$	10	179	7149

$\therefore$ Weighted Average, $\bar{x} = \dfrac{\sum wx}{\sum w} = \dfrac{179}{10} = 17.9$

$\therefore$ Variance, $\sigma^2 = \dfrac{\sum wx^2}{N} - \left(\dfrac{\sum wx}{N}\right)^2 = \dfrac{7149}{10} - \left(\dfrac{179}{10}\right)^2 = 394.49$

Regression

Q15. Calculate (*i*) the regression equation of x on y and y on x from the following data and

(*ii*) estimate x when y = 20, (*iii*) estimate y when x = 30.

Solution

x	x'	x'^2	y	y'	y'^2	$x'y'$
10	-4	16	5	-3	9	12
12	-2	4	6	-2	4	4
13	-1	1	7	-1	1	1
17	3	9	9	1	1	3
18	4	16	13	5	25	20
70		46		40	40	40

Now,

$$\overline{x} = \frac{\sum x}{N} = \frac{70}{5} = 14 \quad \text{and} \quad \overline{y} = \frac{\sum y}{N} = \frac{40}{5} = 8$$

$$m = \frac{\sum x'y'}{\sum x'^2} = \frac{40}{46} = 0.87 \quad \text{and} \quad m' = \frac{\sum x'y'}{\sum y'^2} = \frac{40}{40} = 1$$

Regression equation,

$(y - \overline{y}) = m(x - \overline{x})$	$(x - \overline{x}) = m'(y - \overline{y})$
$y - 8 = 0.87(x - 14)$	$x - 14 = 1(y - 8)$
$y = 0.87x - 14 \times 0.87 + 8$	$x = y - 8 + 14$
$\therefore y = 0.87x - 4.18$	$\therefore x = y + 6$

when y = 20, $x = (20) + 6 = 26$

when x = 30, $y = 0.87(30) - 4.18 = 21.92$

Coefficient of linear correlation, $r = \dfrac{\sum x'y'}{\sqrt{\sum x'^2 \sum y'^2}} = \dfrac{40}{\sqrt{(46)(40)}} = 0.933$