1) There are 20 bricks of weight 3.4 Kg each and 30 bricks of weight 3.6 Kg each. Find the Mean, and Standard Deviation of weight for the whole bunch of 50 bricks?

Solution:

$$\text{Mean } (\mu) = \frac{\sum_{i=1}^{i=n} xi}{n} = \frac{20*3.4+30*3.6}{(20+30)} = \frac{68+108}{50} = \frac{176}{50} = \mathbf{3.52}$$

Variance of a population is, $\sigma^2 = \frac{\sum(x-\mu)^2}{n} = \frac{20(3.4-3.52)^2+30(3.6-3.52)^2}{50}$

$\sigma^2 = \frac{20*0.0144+30*0.0064}{50} = \frac{0.288+0.192}{50} = \frac{0.48}{50} = 0.0096$

Standard Deviation, $\boldsymbol{\sigma = 0.09797}$

2) Find the Mean, Standard Deviation of height for the following dataset.

Height (cm)	No of People
150-156	2
157-163	14
164-170	15
171-177	20
178-184	7
185-191	10

Solution:

	Height (cm)	No of People (f)	Mid-point of Height (x)	f*x	x-m	(x-m)²
	150-156	2	153	306	-17.5	306.25
	157-163	14	160	2240	-10.5	110.25
	164-170	15	167	2505	-3.5	12.25
	171-177	20	174	3480	3.5	12.25
	178-184	7	181	1267	10.5	110.25
	185-191	10	188	1880	17.5	306.25
SUM		**68**	**1023**	**11678**		**857.5**

Mean, μ = $\frac{\sum f * x}{\sum f} = \frac{11678}{68} = \mathbf{171.7352}$

In case we use other formula to calculate mean,

$$\text{Mean } (\mu) = \frac{\sum_{i=1}^{i=n} xi}{n} = \frac{1023}{6} = \mathbf{170.5}$$

Variance of a population is, $\sigma^2 = \frac{\sum(x-\mu)^2}{n} = \frac{857.5}{68} = \mathbf{12.6102}$

Standard Deviation, $\boldsymbol{\sigma} = \mathbf{3.551}$

3) Find the correlation coefficient for the following set of observations.

X	7	14	24	30	45	57
Y	24	34	45	50	61	69

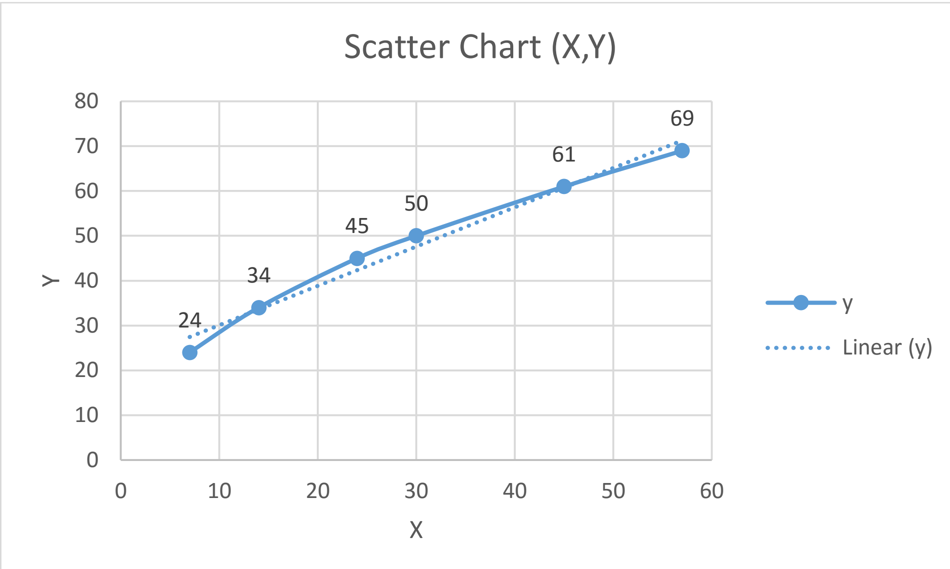
Solution:

$$\text{Correlation Coefficient, } r = r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\ \sqrt{n\sum y_i^2 - (\sum y_i)^2}}.$$

	x	y	x*y	x²	y²
	7	24	168	49	576
	14	34	476	196	1156
	24	45	1080	576	2025
	30	50	1500	900	2500
	45	61	2745	2025	3721
	57	69	3933	3249	4761
SUM	**177**	**283**	**9902**	**6995**	**14739**

$$r = \frac{6*9902-177*283}{\sqrt{6*6995-177*177}\sqrt{6*14739-283*283}} = \frac{59412-50091}{\sqrt{41970-31329}\sqrt{88434-80089}}$$

$$= \frac{9321}{\sqrt{10641}\sqrt{8345}} = \frac{9321}{103.15*91.35} = \frac{9321}{9423.2296} = \mathbf{0.9891}$$



Since 'r' is close to 1 (r = 0.9891) and the curve is almost linear, they are strongly related and positive.

4) Find the correlation coefficient for the following data set and interpret the results.

Vehicle model	Mileage (m/g)	Price $'000
1	19	14.94
2	19	14.8
3	20	24.76
4	20	14.93
5	20	13.95
6	21	17.88
7	21	11.65
8	22	17.9
9	23	21.5
10	24	13.25
11	25	9.6
12	17	13.95
13	28	13.07
14	32	6.6

15	33	9.41
16	34	5.87
17	35	6.49

Solution:

Let’s consider that based on mileage and other factors, price of the vehicle is decided.

So, let’s assume that, Mileage = x (independent) and Price = y (dependent)

$$\text{Correlation Coefficient, } r = r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\ \sqrt{n\sum y_i^2 - (\sum y_i)^2}}.$$

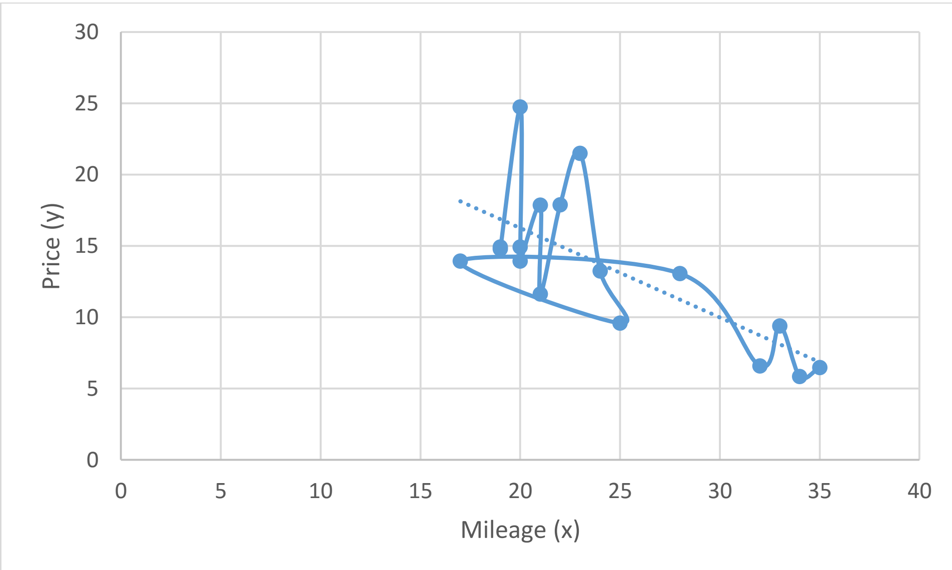
Vehicle model	Mileage (x)	Price (y)	xy	x^2	y^2
1	19	14.94	283.86	361	223.2036
2	19	14.8	281.2	361	219.04
3	20	24.76	495.2	400	613.0576
4	20	14.93	298.6	400	222.9049
5	20	13.95	279	400	194.6025
6	21	17.88	375.48	441	319.6944
7	21	11.65	244.65	441	135.7225
8	22	17.9	393.8	484	320.41
9	23	21.5	494.5	529	462.25
10	24	13.25	318	576	175.5625
11	25	9.6	240	625	92.16
12	17	13.95	237.15	289	194.6025
13	28	13.07	365.96	784	170.8249
14	32	6.6	211.2	1024	43.56
15	33	9.41	310.53	1089	88.5481
16	34	5.87	199.58	1156	34.4569
17	35	6.49	227.15	1225	42.1201
SUM	**413**	**230.55**	**5255.86**	**10585**	**3552.721**

$$r = \frac{17*5255.86-413*230.55}{\sqrt{17*10585-413*413}\sqrt{17*3552.721-230.55*230.55}}$$

$$= \frac{89349.62 - 95217.15}{\sqrt{179945 - 170569}\sqrt{60396.257 - 53153.3025}}$$

$$= \frac{-5867.53}{\sqrt{9376}\sqrt{7242.9545}} = \frac{-5867.53}{96.82 * 85.1} = \frac{-5867.53}{8240.2115} = \mathbf{-0.712}$$

$$\text{Threshold} = \frac{1.96}{\sqrt{n}} = \frac{1.96}{\sqrt{17}} = \frac{1.96}{4.123} = 0.475$$



Observations: -

i) r is negative, hence its Inverse correlation or negative correlation.
ii) Magnitude of ‘r’ is 0.712, hence it’s a medium correlation coefficient

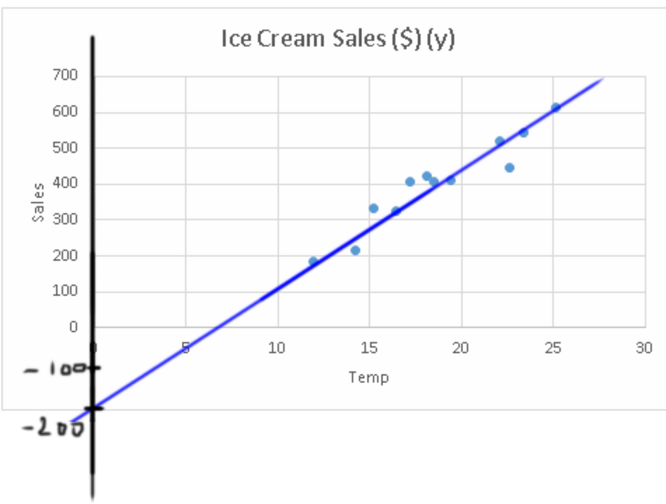
5) Local ice cream shop keeps track of how much ice cream sell versus the noon temperature on that day. Here are their figures for the last 12 days. Identify if there is a linear or otherwise relationship between Ice Cream Sales and Temperature at noon of the day. Predict the Ice Cream sales if the noon temperature is 26.5 degree centigrade.

Temperature (centigrade)	Ice Cream Sales ($)
14.2	215
16.4	325
11.9	185
15.2	332

18.5	406
22.1	522
19.4	412
25.1	614
23.4	544
18.1	421
22.6	445
17.2	408

Solution:

Based on the values, we can get the below graph



We know the general equation of line, y = mx+c,

where m is slope and c is y-intercept

Let's derive the slope of the line, m = $\frac{y2-y1}{x2-x1} = \frac{614-325}{25.1-16.4} = \frac{289}{8.7}$ = 33.21

y-intercept, c = -200

so, the equation of the curve is,

y = 33.21x – 200

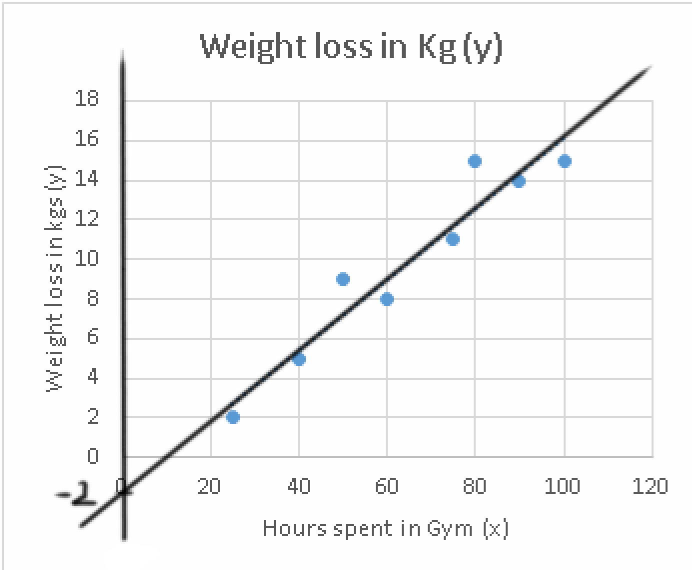
When x = 26.5, y = 33.21*26.5-200 = 880.065 – 200 = **680.065**

6) **Weight loss of a person is assumed to depend on the number of hours of exercise in gym. Observed values of these for 8 people are given in the table below. Validate if the assumption is right. Predict the weight loss for 70 hours of exercise in the gym.**

Hours spent in gym	Weight loss in Kg
100	15
75	11
80	15
90	14
60	8
50	9
25	2
40	5

Solution:

	Hours spent in gym (x)	Weight loss in Kg (y)
	100	15
	75	11
	80	15
	90	14
	60	8
	50	9
	25	2
	40	5
SUM	**520**	**79**
Average (Mean)	**65**	**9.875**



We know the general equation of line, y = mx+c,

where m is slope and c is y-intercept

Let's derive the slope of the line, m = $\frac{y2-y1}{x2-x1} = \frac{14-5}{90-40} = \frac{9}{50}$ = 0.18

y-intercept, c = -2

so, the equation of the curve is,

y = 0.18x – 2

It's evident from the curve that as the number of hours in gym is increasing, weight loss is increasing. They are positively correlated and the correlation is strong.

When x = 70, y =0.18*70 - 2 = 12.6 – 2 = **10.6**