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:**

= = = **3.52**

Variance of a population is, σ² = =

σ² = = = = 0.0097

Standard Deviation, **σ = 0.09897**

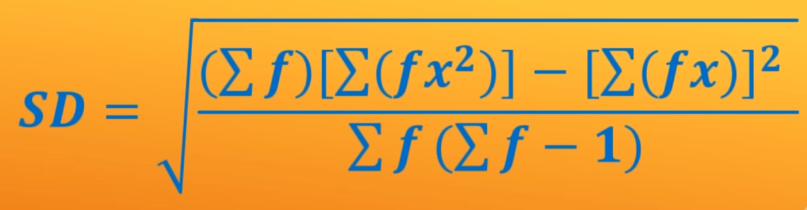
1. **Find the Mean, Standard Deviation of height for the following dataset.**

|  |  |
| --- | --- |
| 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 | f\*x² | |
|  | 150-156 | 2 | 153 | 306 | 46818 | |
|  | 157-163 | 14 | 160 | 2240 | 358400 | |
|  | 164-170 | 15 | 167 | 2505 | 418335 | |
|  | 171-177 | 20 | 174 | 3480 | 605520 | |
|  | 178-184 | 7 | 181 | 1267 | 229327 | |
|  | 185-191 | 10 | 188 | 1880 | 353440 | |
| **SUM** |  | **68** | **1023** | **11678** | **2011840** | |
|  |  |  |  |  = 171.735294 | |

Mean, μ = = = **171.7352**

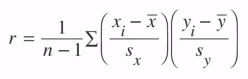


Standard Deviation, σ = = = = 9.7086

1. **Find the correlation coefficient for the following set of observations.**

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

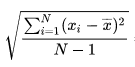
**Solution:**

Correlation coefficient: 

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 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** |

= = = 29.5

= = = 47.16

Sample Standard Deviation for x, Sx = 

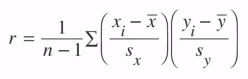
Sx =

Sx = = = 18.83

Sample Standard Deviation for y, Sy =

Sy = = = 16.67

**Z-Score:** How many Standard Deviations (σ) away from the Mean(μ)



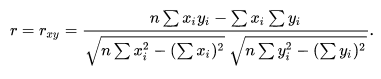
r =

r =

r = =

**r = 0.98**

or, we can calculate the correlation coefficient as below



r = =

= = = = **0.9891**

Since ‘r’ is close to 1 (r = 0.98) and the curve is almost linear, they are strongly related and positive.

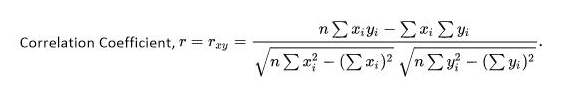
1. **Find the correlation coefficient for the following data set and interpret the results.**

|  |  |  |
| --- | --- | --- |
| 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)

****

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Vehicle model | Mileage (x) | Price (y) | xy |  |  |
| 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 =

=

= = = = **-0.712**

Threshold = = = = 0.475

**Observations: -**

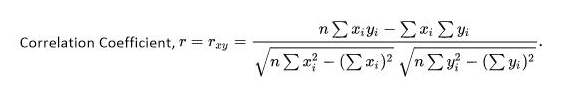
1. r is negative, hence its Inverse correlation or negative correlation.
2. Magnitude of ‘r’ is 0.712, hence it’s a medium correlation coefficient
3. **Local ice cream shop keeps track of how much ice cream sell versus the noon temperature on that day. Here are their figures for the last 12 days.**

**a. Identify if there is a linear or otherwise relationship between Ice Cream Sales and Temperature at noon of the day.**

1. **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:**

****

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Temperature (centigrade) (x) | Ice Cream Sales ($) (y) | x\*y | x² | y² |
|  | 14.2 | 215 | 3053 | 201.64 | 46225 |
|  | 16.4 | 325 | 5330 | 268.96 | 105625 |
|  | 11.9 | 185 | 2201.5 | 141.61 | 34225 |
|  | 15.2 | 332 | 5046.4 | 231.04 | 110224 |
|  | 18.5 | 406 | 7511 | 342.25 | 164836 |
|  | 22.1 | 522 | 11536.2 | 488.41 | 272484 |
|  | 19.4 | 412 | 7992.8 | 376.36 | 169744 |
|  | 25.1 | 614 | 15411.4 | 630.01 | 376996 |
|  | 23.4 | 544 | 12729.6 | 547.56 | 295936 |
|  | 18.1 | 421 | 7620.1 | 327.61 | 177241 |
|  | 22.6 | 445 | 10057 | 510.76 | 198025 |
|  | 17.2 | 408 | 7017.6 | 295.84 | 166464 |
| **SUM** | **224.1** | **4829** | **95506.6** | **4362.05** | **2118025** |
| **Average (Mean)** | **18.675** | **402.4167** | **7958.883** | **363.5042** | **176502.1** |

r =

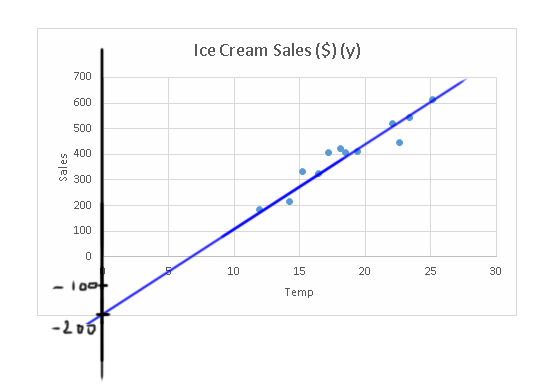
=

= = = =

**r = 0.9576**

Threshold = = = 0.566

Correlation coefficient is greater than threshold and r is close to 1, hence it’s a very strong correlation.

Scatter plot for the given values looks like the below graph

By looking at the scatter-plot, we can conclude that there is a linear relationship between temperature and ice-cream sales and they are directly proportional to each other.

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 = = = = **33.21**

y-intercept (from the graph), **c = -200**

so, the equation of the curve is,

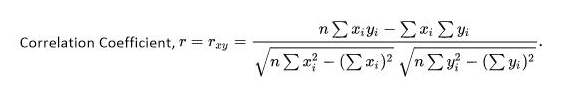
y = 33.21x – 200

When x = 26.5°c, y (Ice Cream Sales) = 33.21\*26.5-200 = 880.065 – 200 = **680.065**

1. **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) | x\*y | x² | y² |
|  | 100 | 15 | 1500 | 10000 | 225 |
|  | 75 | 11 | 825 | 5625 | 121 |
|  | 80 | 15 | 1200 | 6400 | 225 |
|  | 90 | 14 | 1260 | 8100 | 196 |
|  | 60 | 8 | 480 | 3600 | 64 |
|  | 50 | 9 | 450 | 2500 | 81 |
|  | 25 | 2 | 50 | 625 | 4 |
|  | 40 | 5 | 200 | 1600 | 25 |
| **SUM** | **520** | **79** | **5965** | **38450** | **941** |
| **Average (Mean)** | **65** | **9.875** | **745.625** | **4806.25** | **117.625** |

r =

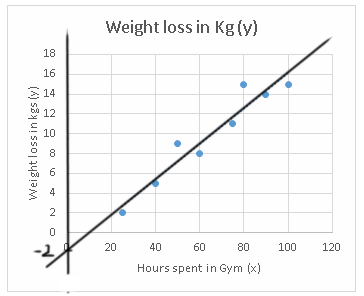
=

= = =

**r = 0.9597**

Threshold = = = 0.695

Correlation coefficient is greater than threshold and r is close to 1, hence it’s a very strong correlation.



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 = = = = **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**