



Research Paper

Bridging Description and Prediction: Numerical Techniques from Arithmetic Mean and Standard Deviation to Correlation and Regression

Chetna Rani Gupta

Assistant Prof. in Mathematics
Multani Mal Modi College, Patiala, Punjab

ABSTRACT: Modern civilization rests on the unseen scaffolding of statistics. Statistics is applied across virtually every field, including Medicine, Business, Technology, Government and Science. Although statistics is a fundamental component of the educational curriculum, many students struggle to recall and apply the correct formulas for numerical problem-solving. Assuming the student possesses the foundational theoretical understanding, this paper focuses on simplifying the calculation techniques for core statistical topics, effectively bridging description and prediction through numerical techniques from the arithmetic mean and standard deviation to correlation and regression. The primary aim is to equip students with straightforward, intuitive methods to easily understand, retain, and apply various statistical formulas.

KEYWORDS: Mean , Standard Deviation, Correlation Coefficient, Regression Coefficient

Received 24 May., 2026; Revised 28 May., 2026; Accepted 02 June., 2026 © The author(s) 2026.
Published with open access at www.questjournals.org

I. INTRODUCTION

Every major pillar of modern society—from life-saving clinical trials in medicine to forecasting models in business and artificial intelligence in technology—relies on statistics. Yet, despite its undeniable importance, a frustrating gap often exists in the classroom. Students frequently grasp the "big picture" theories behind data, but freeze when it comes to the actual math. The math itself becomes a wall, blocking students from applying the very theories they have learned. The formulas feel overwhelming, turning problem solving into a test of memorization rather than a test of understanding. This paper focuses on tearing down that wall. Titled "Bridging Description and Prediction: Numerical Techniques from Arithmetic Mean and Standard Deviation to Correlation and Regression," this study provides straightforward, intuitive methods for statistical calculations. By simplifying the math, this paper empowers students to stop memorizing and start mastering, giving them the practical tools to confidently move from describing data to predicting outcomes.

II. ARITHMETIC MEAN

Notation and terminology used

X represents the series itself and x represents the terms of the series

\bar{X} is used for the mean of the series X

n is the number of terms of the series

A is the assumed mean of the series X , d is the deviation from assumed mean

$$d = X - A$$

c is the common factor of the series d and $d' = \frac{x-A}{c}$

$N = \sum f$ where f is the frequency

m is the mid-values of the class interval of a continuous series

In case of a continuous series, deviation of the mid values are taken from the assumed mean

$$d = m - A, d' = \frac{m-A}{c}$$

\bar{X}_w represents Weighted Arithmetic Mean, $d = X - A_w$ Where A_w is the weighted assumed mean, $d' = \frac{X-A_w}{c}$, c is a common factor of series $d = m - A$

| Arithmetic Mean in case of an Individual Series | | |
|---|--|---|
| Direct Method $\bar{X} = \frac{\sum x}{n}$ | Short-Cut Method $\bar{X} = A + \frac{\sum d}{n}$ | Step Deviation Method $\bar{X} = A + \frac{\sum d'}{n} \times c$ |
| Arithmetic Mean in case of Discrete Series | | |
| Direct Method $\bar{X} = \frac{\sum fx}{N}$ | Short-Cut Method $\bar{X} = A + \frac{\sum fd}{N}$ | Step Deviation Method $\bar{X} = A + \frac{\sum fd'}{N} \times c$ |
| Arithmetic Mean in case of Continuous Series | | |
| Direct Method $\bar{X} = \frac{\sum fm}{N}$ | Short-Cut Method $\bar{X} = A + \frac{\sum fd}{N}$ | Step Deviation Method $\bar{X} = A + \frac{\sum fd'}{N} \times c$ |

Note : Observe that the same fundamental methods are applied to calculate arithmetic mean across Individual Series, Discrete Series and Continuous Series, The only variation lies in determining the values of n, N, d, d' specific to each series. Additionally, Individual series calculations do not involve frequencies, whereas both Discrete and Continuous series require multiplication by the frequency. Furthermore .in the direct method of continuous series, $\sum fm$ is taken instead of $\sum fx$ which is used in discrete series

| Weighted Arithmetic Mean | | |
|--|--|---|
| Direct Method $\bar{X}_w = \frac{\sum wx}{\sum w}$ | Short-Cut Method $\bar{X}_w = A_w + \frac{\sum wd}{\sum w}$ | Step Deviation Method $\bar{X} = A_w + \frac{\sum wd'}{\sum w} \times c$ |
| Combined Arithmetic Mean | | |
| In case of two series $X_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$ | In case of three series $X_{123} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2 + N_3\bar{X}_3}{N_1 + N_2 + N_3}$ | In case of n series $X_{123 \dots n} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2 + N_3\bar{X}_3 \dots + N_n\bar{X}_n}{N_1 + N_2 + N_3 + \dots + N_n}$ |
| Corrected Arithmetic Mean | | |
| Corrected $\sum x =$ incorrect $\sum x$ - incorrect values x + corrected values x | | |
| Apply the formula $\bar{X} = \frac{\text{corrected } \sum x}{n}$ | | |

III. STANDARD DEVIATION

Notation and terminology used

Standard Deviation is represented by σ . Rest are same as used in case of Arithmetic Mean

| Standard Deviation in case of Individual series | | |
|--|---|---|
| Direct Method $\sigma = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2}$ | Short-Cut Method $\sigma = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$ | Step Deviation Method $\sigma = \sqrt{\frac{\sum d'^2}{n} - \left(\frac{\sum d'}{n}\right)^2} \times c$ |
| Standard Deviation in case of Discrete series | | |
| Direct Method $\sigma = \sqrt{\frac{\sum fX^2}{N} - \left(\frac{\sum fX}{N}\right)^2}$ | Short-Cut Method $\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$ | Step Deviation Method $\sigma = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times c$ |
| Standard Deviation in case of Continuous series | | |

| Direct Method | Short-Cut Method | Step Deviation Method |
|--|--|---|
| $\sigma = \sqrt{\frac{\sum fm^2}{N} - \left(\frac{\sum fm}{N}\right)^2}$ | $\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$ | $\sigma = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times c$ |
| Coefficient of Standard Deviation and Coefficient of Variation | | |
| Coefficient of Standard Deviation = $\frac{\sigma}{\bar{x}}$ | | Coefficient of Variation = $\frac{\sigma}{\bar{x}} \times 100$ |

Note : 1. Similar to the arithmetic mean , the foundational methodology for calculating standard deviation remains same across Individual, Discrete, and Continuous series, The only difference lies in computing the n, N, d, d' in each type of series. Further since there are no frequencies in individual series so frequency multiplication is entirely omitted across all the three method. However, for discrete and continuous Series, the multiplication by f is mandatory.

2. Compare the methods to calculate Coefficient of Standard Deviation and Coefficient of Variation note that the sole difference is the multiplication by 100. The Coefficient of Variation is simply the Coefficient of Standard Deviation expressed as a percentage.

IV. KARL PEARSON'S COEFFICIENT OF CORRELATION

Notation and terminology used

Karl Pearson's Coefficient of Correlation is represented by r_{XY}

$dx = x - \bar{X}$ and $dy = y - \bar{Y}$, deviation are taken from the actual mean respectively

A_x and A_y are assumed mean of the series X and Y , respectively

$dx = x - A_x$ and $dy = y - A_y$, deviations are taken from the assumed mean

c_x and c_y are the common factors of series dX and dY

$$d'x = \frac{x-A_x}{c_x} \text{ and } d'y = \frac{y-A_y}{c_y}$$

In case of continuous series, deviation of mid values are taken from the assumed mean

$dx = m_x - A_x$ and $dy = m_y - A_y$, m_x, m_y

are the mid values of the class interval of two continuous series

$$d'x = \frac{m_x-A_x}{c_x} , d'y = \frac{m_y-A_y}{c_y}$$

| Karl Pearson's Coefficient of Correlation in case of individual series/ungrouped series | |
|--|--|
| When deviation are taken from the actual mean $r_{XY} = \frac{\sum(x - \bar{X})(y - \bar{Y})}{\sqrt{\sum(x - \bar{X})^2} \sqrt{\sum(y - \bar{Y})^2}}$ Or $r_{XY} = \frac{\sum dxdy}{\sqrt{\sum(dx)^2} \sqrt{\sum(dy)^2}}$ where $dx = x - \bar{X}$ and $dy = y - \bar{Y}$ and \bar{X} and \bar{Y} are actual mean of series X and Y , respectively | Direct Method When no deviations are taken at all $r_{XY} = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}$ |
| Short Cut Method When deviations are taken from assumed mean $r_{XY} = \frac{n \sum dxdy - (\sum dx)(\sum dy)}{\sqrt{n \sum dx^2 - (\sum dx)^2} \sqrt{n \sum dy^2 - (\sum dy)^2}}$ | Step Deviation Method When deviations are taken from assumed mean $r_{XY} = \frac{n \sum d'xd'y - (\sum d'x)(\sum d'y)}{\sqrt{n \sum d'x^2 - (\sum d'x)^2} \sqrt{n \sum d'y^2 - (\sum d'y)^2}}$ Here multiplication by common factor c_x and c_y from the series X and Y is not done, as coefficient of correlation is independent of scale |
| Karl Pearson's Coefficient of Correlation in case of discrete and continuous series/grouped series | |
| When deviation are taken from actual mean $r_{XY} = \frac{\sum f(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum f(X - \bar{X})^2} \sqrt{\sum f(Y - \bar{Y})^2}}$ OR | Direct Method When no deviations are taken at all $r_{XY} = \frac{N \sum fxy - (\sum fx)(\sum fy)}{\sqrt{N \sum fx^2 - (\sum fx)^2} \sqrt{N \sum fy^2 - (\sum fy)^2}}$ |

| | |
|--|--|
| $r_{XY} = \frac{\sum f dx dy}{\sqrt{\sum f(dx)^2} \sqrt{\sum f(dy)^2}}$ <p>where $dx = x - \bar{X}$ and $dy = y - \bar{Y}$ and \bar{X} and \bar{Y} are actual mean of series X and Y, respectively</p> | |
| <p>Short Cut Method When deviations are taken from assumed mean</p> $r_{XY} = \frac{N \sum f dx dy - (\sum f dx)(\sum f dy)}{\sqrt{N \sum f dx^2 - (\sum f dx)^2} \sqrt{N \sum f dy^2 - (\sum f dy)^2}}$ | <p>Step Deviation Method When deviations are taken from assumed mean</p> $r_{XY} = \frac{N \sum f d' x d' y - (\sum f d' x)(\sum f d' y)}{\sqrt{N \sum f d' x^2 - (\sum f d' x)^2} \sqrt{N \sum f d' y^2 - (\sum f d' y)^2}}$ <p>Note: Here multiplication by common factors c_x and c_y from the series X and Y is not done as coefficient of correlation is independent of scale</p> |

- Note :** 1. In all the methods except the method when deviation are taken from the actual mean, observe that a particular type of same pattern in the formula in direct ,shortcut and step deviation method ,only difference is in the use $fx, fy, fxy, fdx, fdy, f dx dy$ and $fdx', fd'y, fd'xd'y$ according to the type of method
2. On comparing all the formulas with that of individual series, note that f is mandatory in all the formulas for discrete and continuous series
3. Observe that the same formula for each method for continuous series as that of in case of discrete series, only difference is in the calculation of $dx, dy, d'x$ and $d'y$. Here deviation is taken from the midpoints because of the presence of class intervals

| |
|--|
| Corrected Correlation Coefficient |
| Corrected $\sum x =$ incorrect $\sum x$ - incorrect values x + corrected values x |
| Corrected $\sum y =$ incorrect $\sum y$ - incorrect values y + corrected values y |
| Corrected $\sum xy =$ incorrect $\sum xy$ - incorrect values xy + corrected values xy |
| Corrected $\sum x^2 =$ incorrect $\sum x^2$ - incorrect values x^2 + corrected values x^2 |
| Corrected $\sum y^2 =$ incorrect $\sum y^2$ - incorrect values y^2 + corrected values y^2 |
| Apply the formula $r_{XY} = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}$ Take the corrected values of all the summations |

V. REGRESSION COEFFICIENT

Notation and terminology used

Regression Coefficient of X on Y is represented by b_{XY} and Regression Coefficient of Y on X is represented by b_{YX} , Others are the same already used in the paper only

| | |
|---|---|
| Regression Coefficients in case of individual series /ungrouped data: | |
| Regression Coefficient of X on Y b_{XY} | Regression Coefficient of Y on X b_{YX} |
| <p>Direct Method When no deviations are taken at all</p> $b_{XY} = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum Y^2 - (\sum y)^2}$ | <p>Direct Method When no deviations are taken at all</p> $b_{YX} = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2}$ |
| <p>Short Cut Method When deviations are taken from the assumed mean</p> $b_{XY} = \frac{n \sum dx dy - (\sum dx)(\sum dy)}{n \sum dy^2 - (\sum dy)^2}$ | <p>Short Cut Method When deviations are taken from the assumed mean</p> $b_{YX} = \frac{n \sum dx dy - (\sum dx)(\sum dy)}{n \sum dx^2 - (\sum dx)^2}$ |

| | |
|---|---|
| <p>Step Deviation Method When deviations are taken from the assumed mean</p> $b_{XY} = \frac{n \sum d'x d'y - (\sum d'x)(\sum d'y)}{n \sum d'y^2 - (\sum d'y)^2} \times \frac{c_x}{c_y}$ | <p>Step Deviation Method When deviations are taken from the assumed mean</p> $b_{YX} = \frac{n \sum d'x d'y - (\sum d'x)(\sum d'y)}{n \sum d'x^2 - (\sum d'x)^2} \times \frac{c_y}{c_x}$ |
| Regression Coefficients in case of grouped data covering both discrete and continuous series | |
| Regression Coefficient of X on Y b_{XY} | Regression Coefficient of Y on X b_{YX} |
| <p>Direct Method When no deviations are taken at all</p> $b_{XY} = \frac{N \sum fxy - (\sum fx)(\sum fy)}{N \sum fy^2 - (\sum fy)^2}$ | <p>Direct Method When no deviations are taken at all</p> $b_{YX} = \frac{N \sum fxy - (\sum fx)(\sum fy)}{N \sum fx^2 - (\sum fx)^2}$ |
| <p>Short Cut Method When deviations are taken from assumed mean</p> $b_{XY} = \frac{N \sum fdxdy - (\sum fdx)(\sum fdy)}{N \sum fdy^2 - (\sum fdy)^2}$ | <p>Short Cut Method When deviations are taken from the assumed mean</p> $b_{YX} = \frac{N \sum fdxdy - (\sum fdx)(\sum fdy)}{N \sum fdx^2 - (\sum fdx)^2}$ |
| <p>Step Deviation Method When deviations are taken from the assumed mean</p> $b_{XY} = \frac{N \sum fd'xd'y - (\sum fd'x)(\sum fd'y)}{N \sum fd'y^2 - (\sum fd'y)^2} \times \frac{c_x}{c_y}$ | <p>Step Deviation Method When deviations are taken from the assumed mean</p> $b_{YX} = \frac{N \sum fd'xd'y - (\sum fd'x)(\sum fd'y)}{N \sum fd'x^2 - (\sum fd'x)^2} \times \frac{c_y}{c_x}$ |

Note: Compare the similarities and dissimilarities in the methods mentioned above. Note the difference in the denominators of Regression Coefficient of Y on X and that of X on Y and in use of $\frac{c_y}{c_x}$ or $\frac{c_x}{c_y}$

VI. COVARIANCE OF SERIES X AND Y

Notation and terminology used

Covariance of Series X any Y is represented by $Cov(X, Y)$. Others are same already used in the paper only

Covariance of Series X any Y in case of individual series /ungrouped Series

| | |
|---|---|
| <p>When deviation are taken from actual mean</p> $Cov(X, Y) = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{n}$ | <p>Short Cut Method When deviations are taken from assumed mean</p> $Cov(X, Y) = \frac{n \sum dxdy - (\sum dx)(\sum dy)}{n^2}$ |
| <p>Direct Method When no deviations are taken at all</p> $Cov(X, Y) = \frac{n \sum xy - (\sum x)(\sum y)}{n^2}$ | <p>Step Deviation Method When deviations are taken from assumed mean</p> $Cov(X, Y) = \frac{n \sum d'xd'y - (\sum d'x)(\sum d'y)}{n^2} \times c_x \times c_y$ |
| Covariance of Series X any Y in case of grouped data covering both discrete and continuous series | |
| <p>When deviation are taken from actual mean</p> $Cov(X, Y) = \frac{\sum f(X - \bar{X})(Y - \bar{Y})}{N}$ | <p>Short Cut Method When deviations are taken from assumed mean</p> $Cov(X, Y) = \frac{N \sum fdxdy - (\sum fdx)(\sum fdy)}{N^2}$ |
| <p>Direct Method When no deviations are taken at all</p> $Cov(X, Y) = \frac{N \sum fxy - (\sum fx)(\sum fy)}{N^2}$ | <p>Step Deviation Method When deviations are taken from assumed mean</p> $Cov(X, Y) = \frac{N \sum fd'xd'y - (\sum fd'x)(\sum fd'y)}{N^2} \times c_x \times c_y$ |

| | |
|--|--|
| | $= \frac{N \sum f d' x d' y - (\sum f d' x) (\sum f d' y)}{N^2} \times c_x \times c_y$ |
|--|--|

Note : Observe the presence of c_x, c_y in the step deviation method

VII. OTHER IMPORTANT FORMULAS OF CORELATION AND REGRESSION

| | |
|---|---|
| Some more formula regarding correlation and regression | |
| Regression line of X on Y $X - \bar{X} = b_{XY}(Y - \bar{Y})$ | Regression line of Y on X $Y - \bar{Y} = b_{YX}(X - \bar{X})$ |
| Standard Error = $\frac{1-(r_{XY})^2}{\sqrt{n}}$ | Probable Error = $0.6745 \times \frac{1-(r_{XY})^2}{\sqrt{n}}$ = 0.6745 × Standard Error |
| Standard Error of an Estimate of X on Y $S_{XY} = \sigma_X \sqrt{1 - (r_{XY})^2}$ | Standard Error of an Estimate of Y on X $S_{YX} = \sigma_Y \sqrt{1 - (r_{XY})^2}$ |
| Standard Error of an Estimate of X on Y $S_{XY} = \sqrt{\frac{\sum (y - \hat{y})^2}{n}}, Y - \hat{Y} = b_{YX}(X - \bar{X})$ | Standard Error of an Estimate of X on Y $S_{YX} = \sqrt{\frac{\sum (x - \hat{x})^2}{n}}, X - \hat{X} = b_{XY}(Y - \bar{Y})$ |

VIII. PROPERTIES OF CORRELATION AND REGRESSION

- Coefficient of Correlation is independent of change of origin
- Coefficient of Correlation is independent of change of scale
- Range of coefficient of correlation is from -1 to 1
- Correlation coefficient is a pure number
- The sign of correlation indicates the direction of the relationship
- If X and Y are independent variables then Correlation coefficient is zero but the converse is not true
- The square of coefficient r^2 of Correlation is referred to as coefficient of determination
- Correlation coefficient possesses the property of symmetry

$$r_{xy} = r_{yx}$$
- Correlation coefficient is the geometric mean of two regression coefficients $r = \pm \sqrt{b_{yx} b_{xy}}$ Or $r^2 = b_{yx} b_{xy}$
- Both of the regression coefficient have same sign
- The product of two regression coefficient is always less than or equal to one
- The arithmetic mean of two regression coefficient is always more than equal to geometric mean of two regression coefficients $\frac{b_{yx} + b_{xy}}{2} \geq \sqrt{b_{yx} b_{xy}}$
- Regression coefficient can also be calculated by using Coefficient of Correlation and standard deviations $b_{xy} = r \frac{\sigma_x}{\sigma_y}$ or $b_{yx} = r \frac{\sigma_y}{\sigma_x}$

IX. NUMERICALS

9.1 Continuous series covering mean and standard deviation only by applying all the three methods Direct, Short Cut and Step Deviation methods

| Class Intervals | Frequency (f) | Mid Value of Class Interval (m) | mf | A = 50 d = m - A | fd | c = 10' $\frac{d}{m - A} = \frac{d}{c}$ | fd' |
|-----------------|---------------|---------------------------------|------------------|---------------------|-----------------|--|-----------------|
| 0-20 | 4 | 10 | 40 | -40 | -160 | -4 | -16 |
| 20-40 | 10 | 30 | 300 | -20 | -200 | -2 | -20 |
| 40-60 | 15 | 50 | 750 | 0 | 0 | 0 | 0 |
| 60-80 | 20 | 70 | 1400 | 20 | 400 | 2 | 40 |
| 80-100 | 11 | 90 | 990 | 40 | 440 | 4 | 44 |
| | $\sum f = 60$ | | $\sum fm = 3480$ | | $\sum fd = 480$ | | $\sum fd' = 48$ |

| | | |
|---|--|---|
| Arithmetic Mean by Direct Method $\bar{X} = \frac{\sum fm}{N} = \frac{3480}{60} = 58$ | Arithmetic Mean by Short-Cut Method $\bar{X} = A + \frac{\sum fd}{N} = 50 + \frac{480}{60} = 58$ | Arithmetic Mean by Step Deviation Method $\bar{X} = A + \frac{\sum fd'}{N} \times c$ $= 50 + \frac{48}{60} \times 10 = 58$ |
|---|--|---|

Standard Deviation by Direct method

| Class Intervals | Frequency (f) | Mid Value of Class Interval (m) | fm | m ² | fm ² |
|-----------------|---------------|---------------------------------|----------------|----------------|--------------------|
| 0-20 | 4 | 10 | 40 | 100 | 400 |
| 20-40 | 10 | 30 | 300 | 900 | 9000 |
| 40-60 | 15 | 50 | 750 | 2500 | 37500 |
| 60-80 | 20 | 70 | 1400 | 4900 | 98000 |
| 80-100 | 11 | 90 | 990 | 8100 | 89100 |
| | $\sum f=60$ | | $\sum fm=3480$ | | $\sum fm^2=234000$ |

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

$$= \sqrt{\frac{234000}{60} - \left(\frac{3480}{60}\right)^2}$$

$$= \sqrt{3900 - 3364} = \sqrt{536}$$

$$= 23.152$$

Standard Deviation by Short Cut Method and Step deviation Method

| Class Intervals | Frequency (f) | Mid Value of Class Interval (m) | d = m - A A = 50 | fd | d ² | fd ² | d' = $\frac{m-A}{c}$ c = 10 | fd' | d' ² | fd' ² |
|-----------------|---------------|---------------------------------|---------------------|---------------|----------------|-------------------|--------------------------------|---------------|-----------------|------------------|
| 0-20 | 4 | 10 | -40 | -160 | 1600 | 6400 | -4 | -16 | 16 | 64 |
| 20-40 | 10 | 30 | -20 | -200 | 400 | 4000 | -2 | -20 | 4 | 40 |
| 40-60 | 15 | 50 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 60-80 | 20 | 70 | 20 | 400 | 400 | 8000 | 2 | 40 | 4 | 80 |
| 80-100 | 11 | 90 | 40 | 440 | 1600 | 17600 | 4 | 44 | 16 | 176 |
| | $\sum f=60$ | | | $\sum fd=480$ | | $\sum fd^2=36000$ | | $\sum fd'=48$ | | $\sum fd'^2=360$ |

Short Cut Method

$$\sigma = \sqrt{\frac{\sum fd}{N} - \left(\frac{\sum fd}{N}\right)^2} = \sqrt{\frac{36000}{60} - \left(\frac{480}{60}\right)^2} = \sqrt{600 - 64} = \sqrt{536} = 23.152$$

Step Deviation method

$$\sigma = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times c$$

$$= \sqrt{\frac{360}{60} - \left(\frac{48}{60}\right)^2} \times 10 = \sqrt{\frac{360 \times 60 - 2304}{3600}} \times 10 = \sqrt{\frac{21600 - 2304}{3600}} \times 10 = \sqrt{\frac{9296}{3600}} \times 10$$

$$= \sqrt{5.36} \times 10 = \sqrt{536} = 23.152$$

Coefficient of Standard Deviation = $\frac{\sigma}{\bar{X}} = \frac{\sqrt{536}}{58} = \frac{23.151673805580}{58} = 0.399$

Coefficient of Variation = $\frac{\sigma}{\bar{X}} \times 100 = \frac{\sqrt{536}}{58} = \frac{23.151673805580}{58} \times 100$
 $= 0.3991667897513 \times 100 = 39.917$

Note: If in the above example class intervals are to be ignored and mid-point value *m* is replaced by *x* then the new series so obtained will represent discrete series and same results will hold true for the discrete series obtained so

9.2 Individual series done by short cut method covering all the topics understudy

| | | | | | | |
|---|---|-------------------------|-------------------------|-------|-----------------|-----------------|
| X | Y | dx = x - A _x | dy = y - A _y | dx dy | dx ² | dy ² |
|---|---|-------------------------|-------------------------|-------|-----------------|-----------------|

Bridging Description and Prediction: Numerical Techniques from Arithmetic Mean and..

| | | | | | | |
|------------|------------|---------------|---------------|-------------------|-------------------|-------------------|
| 45 | 56 = A_y | -11 | 0 | 0 | 121 | 0 |
| 55 | 50 | -1 | -6 | 6 | 1 | 36 |
| 56 = A_x | 48 | 0 | -8 | 0 | 0 | 64 |
| 58 | 60 | 2 | 4 | 8 | 4 | 16 |
| 60 | 62 | 4 | 6 | 24 | 16 | 36 |
| 65 | 64 | 9 | 8 | 72 | 81 | 64 |
| | | $\Sigma dx=3$ | $\Sigma dy=4$ | $\Sigma dxdy=110$ | $\Sigma dx^2=223$ | $\Sigma dy^2=216$ |

Arithmetic Mean

| | |
|--|---|
| <p>Arithmetic Mean of series X</p> $\bar{X} = A_x + \frac{\sum dx}{n} = 56 + \frac{3}{6} = 56.05$ | <p>Arithmetic Mean of series y</p> $\bar{Y} = A_y + \frac{\sum d}{n} = 56 + \frac{4}{6} = 56.66$ |
|--|---|

Standard Deviation of X

$$\sigma_x = \sqrt{\frac{\sum dx^2}{n} - \left(\frac{\sum dx}{n}\right)^2} = \sqrt{\frac{223}{6} - \left(\frac{3}{6}\right)^2} = \sqrt{\frac{6 \times 223 - 9}{36}} = \sqrt{\frac{1338 - 9}{36}} = \sqrt{\frac{1329}{36}}$$

Standard Deviation of Y

$$\sigma_y = \sqrt{\frac{\sum dy^2}{n} - \left(\frac{\sum dy}{n}\right)^2} = \sqrt{\frac{216}{6} - \left(\frac{4}{6}\right)^2} = \sqrt{\frac{6 \times 216 - 16}{36}} = \sqrt{\frac{1296 - 16}{36}} = \sqrt{\frac{1280}{36}}$$

Coefficient of Correlation

$$r_{XY} = \frac{n \sum dxdy - (\sum dx)(\sum dy)}{\sqrt{n \sum dx^2 - (\sum dx)^2} \sqrt{n \sum dy^2 - (\sum dy)^2}}$$

$$r_{XY} = \frac{6 \times 110 - 3 \times 4}{\sqrt{6 \times 223 - 3^2} \sqrt{6 \times 216 - 4^2}} = \frac{660 - 12}{\sqrt{1338 - 9} \sqrt{1296 - 16}} = \frac{648}{\sqrt{1329} \sqrt{1280}}$$

Covariance of Series X any Y

$$Cov(X, Y) = \frac{n \sum dxdy - (\sum dx)(\sum dy)}{n^2} = \frac{6 \times 110 - 3 \times 4}{6^2} = \frac{660 - 12}{6^2} = \frac{648}{36}$$

Regression Coefficient of Y on X

$$b_{YX} = \frac{n \sum dxdy - (\sum dx)(\sum dy)}{n \sum dx^2 - (\sum dx)^2} = \frac{6 \times 110 - 3 \times 4}{6 \times 223 - 3^2} = \frac{648}{1329}$$

Regression Coefficient of X on Y

$$b_{XY} = \frac{n \sum dxdy - (\sum dx)(\sum dy)}{n \sum dy^2 - (\sum dy)^2} = \frac{6 \times 110 - 3 \times 4}{6 \times 216 - 4^2} = \frac{648}{1280}$$

| Some Observations | | |
|---|---|---|
| $r_{XY} = \frac{Cov(X, Y)}{\sigma_x \sigma_y}$ $\frac{Cov(X, Y)}{\sigma_x \sigma_y} = \frac{\frac{648}{36}}{\sqrt{\frac{1329}{36}} \times \sqrt{\frac{1280}{36}}}$ $= \frac{648}{\sqrt{1329} \sqrt{1280}}$ $= r_{XY}$ | $b_{XY} = \frac{Cov(X, Y)}{\sigma_y^2}$ $\frac{Cov(X, Y)}{\sigma_y^2} = \frac{\frac{648}{36}}{\frac{1280}{36}} = \frac{648}{1280} = b_{XY}$ | $b_{YX} = \frac{Cov(X, Y)}{\sigma_x^2}$ $\frac{Cov(X, Y)}{\sigma_x^2} = \frac{\frac{648}{36}}{\frac{1329}{36}} = \frac{648}{1329} = b_{YX}$ |
| $r_{XY} = \sqrt{b_{YX} \times b_{XY}}$ $\sqrt{b_{YX} \times b_{XY}} = \sqrt{\frac{648}{1329} \times \frac{648}{1280}}$ $= \frac{648}{\sqrt{1329} \sqrt{1280}}$ $= r_{XY}$ | $b_{XY} = r_{XY} \frac{\sigma_x}{\sigma_y}$ $r_{XY} \frac{\sigma_x}{\sigma_y} = \frac{648}{\sqrt{1329} \sqrt{1280}} \times \frac{\sqrt{\frac{1329}{36}}}{\sqrt{\frac{1280}{36}}}$ $= \frac{648}{1280} = b_{XY}$ | $b_{YX} = r_{XY} \frac{\sigma_y}{\sigma_x}$ $r_{XY} \frac{\sigma_y}{\sigma_x} = \frac{648}{\sqrt{1329} \sqrt{1280}} \times \frac{\sqrt{\frac{1280}{36}}}{\sqrt{\frac{1329}{36}}}$ $= \frac{648}{1329} = b_{YX}$ |

9.3 Continuous series done by Step Deviation method covering all the topics understudy

Table1

| Income in Rs. | Savings in Rs. (X) | | 25-75 | 75-125 | 125-175 | 175-225 | Total ((f) |
|---------------|--------------------|-------------------|-------|--------|---------|---------|------------|
| | $m_y \downarrow$ | $m_x \rightarrow$ | 50 | 100 | 150 | 200 | |
| 300-500 | 400 | | 10 | 4 | Nil | Nil | 14 |
| 500-700 | 600 | | 8 | 12 | 24 | 6 | 50 |
| 700-900 | 800 | | Nil | 9 | 7 | 2 | 18 |
| 900-1100 | 1000 | | Nil | Nil | 10 | 5 | 15 |
| 1100-1300 | 1200 | | Nil | Nil | 9 | 4 | 13 |
| | Total (f) | | 18 | 25 | 50 | 17 | 110 |

Table 2

| Saving in Rs | 25-75 | 75-125 | 125-175 | 175-225 | |
|-----------------------|-------|--------|---------|---------|---------------------|
| $m_x \rightarrow$ | 50 | 100 | 150 | 200 | |
| f | 18 | 25 | 50 | 17 | $\Sigma f=110$ |
| $dx = x - 100$ | -50 | 0 | 50 | 100 | |
| $d'x = \frac{dx}{50}$ | -1 | 0 | 1 | 2 | |
| $d'x^2$ | 1 | 0 | 1 | 4 | |
| $fd'x$ | -18 | 0 | 50 | 34 | $\Sigma fd'x=66$ |
| $fd'x^2$ | 18 | 0 | 50 | 68 | $\Sigma fd'x^2=136$ |

Table 3

| Income in Rs. | Mid values $m_y \downarrow$ | f | $\frac{dy}{= y - 800}$ | $d'y = \frac{dy}{200}$ | $d'y^2$ | $fd'y$ | $fd'y^2$ |
|---------------|-----------------------------|-----|------------------------|------------------------|---------|--------|----------|
| 300-500 | 400 | 14 | -400 | -2 | 4 | -28 | 56 |
| 500-700 | 600 | 50 | -200 | -1 | 1 | -50 | 50 |
| 700-900 | 800 | 18 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | |
|-----------|------|------------------|-----|---|---|---------------------|-----------------------|
| 900-1100 | 1000 | 15 | 200 | 1 | 1 | 15 | 15 |
| 1100-1300 | 1200 | 13 | 400 | 2 | 4 | 26 | 52 |
| | | $\Sigma f = 110$ | | | | $\Sigma fd'y = -37$ | $\Sigma fd'y^2 = 173$ |

Table 4

| $d'y$ ↓ | $d'x$ → | -1 | 0 | 1 | 2 | $fd'xd'y$ | | | |
|------------|------------|----|-----|---|-----|-----------|-----|-----|-----------------------|
| -2 | 10 | 20 | 4 | 0 | Nil | 0 | Nil | 0 | 20 |
| -1 | 8 | 8 | 12 | 0 | 24 | -24 | 6 | -12 | -28 |
| 0 | Nil | 0 | 9 | 0 | 7 | 0 | 2 | 0 | 0 |
| 1 | Nil | 0 | Nil | 0 | 10 | 10 | 5 | 10 | 20 |
| 2 | Nil | 0 | Nil | 0 | 9 | 18 | 4 | 16 | 34 |
| $fd'xd'y$ | | 28 | 0 | 4 | 14 | | | | $\Sigma fd'xd'y = 46$ |

Obtained after calculation

$N = \Sigma f = 110$, $c_x = 50$, $c_y = 200$, $\Sigma fd'x = 66$, $\Sigma fd'x^2 = 136$, $\Sigma fd'y = -37$, $\Sigma fd'y^2 = 173$, $\Sigma fd'xd'y = 46$

| | |
|--|--|
| <p>Arithmetic Mean of series X</p> $\bar{X} = A_x + \frac{\Sigma fd'x}{N} \times c_x$ $= 100 + \frac{66}{110} \times 50$ $= 100 + 30 = 130$ | <p>Arithmetic Mean of series Y</p> $\bar{Y} = A_y + \frac{\Sigma fd'y}{N} \times c_y$ $= 800 + \frac{-37}{110} \times 200$ $= 800 - 67.272 = 732.728$ |
|--|--|

Standard Deviation

| | |
|---|--|
| <p>Standard Deviation of series X</p> $\sigma_x = \sqrt{\frac{\Sigma fd'x^2}{N} - \left(\frac{\Sigma fd'x}{N}\right)^2} \times c_x$ $= \sqrt{\frac{136}{110} - \left(\frac{66}{110}\right)^2} \times 50$ $= \sqrt{\frac{136 \times 110 - 4356}{12100}} \times 50$ $= \sqrt{\frac{4960 - 4356}{12100}} \times 50$ $= \sqrt{\frac{10604}{12100}} \times 50 = 46.807$ | <p>Standard Deviation of series Y</p> $\sigma_y = \sqrt{\frac{\Sigma fd'y^2}{N} - \left(\frac{\Sigma fd'y}{N}\right)^2} \times c_y$ $= \sqrt{\frac{173}{110} - \left(\frac{-37}{110}\right)^2} \times 200$ $= \sqrt{\frac{173 \times 110 - 1369}{12100}} \times 200$ $= \sqrt{\frac{19030 - 1369}{12100}} \times 200$ $= \sqrt{\frac{17661}{12100}} \times 200 = 241.627$ |
|---|--|

Coefficient of correlation and Covariance of Series X any Y

| | |
|--|---|
| $r_{XY} = \frac{N \sum fd'xd'y - (\sum fd'x)(\sum fd'y)}{\sqrt{N \sum fd'x^2 - (\sum fd'x)^2} \sqrt{N \sum fd'y^2 - (\sum fd'y)^2}}$ $= \frac{110 \times 46 - 66(-37)}{\sqrt{110 \times 136 - (66)^2} \sqrt{110 \times 137 - (-37)^2}}$ $= \frac{5060 + 2442}{7502} = \frac{\sqrt{14960 - 4356} \sqrt{19030 - 1369}}{\sqrt{187277244}} = 0.5482$ | $Cov(X, Y) = \frac{N \sum fd'xd'y - (\sum fd'x)(\sum fd'y)}{N^2} \times c_x \times c_y$ $= \frac{110 \times 46 - 66(-37)}{(110)^2} \times 50 \times 200$ $= \frac{5060 + 2442}{12100} = \frac{7502 \times 1000}{12100} = 620$ |
|--|---|

Regression Coefficient

| | |
|--|---|
| <p>Regression Coefficient of X on Y</p> $b_{XY} = \frac{N \sum fd'xd'y - (\sum fd'x)(\sum fd'y)}{N \sum fd'y^2 - (\sum fd'y)^2} \times \frac{c_x}{c_y}$ $= \frac{110 \times 46 - 66(-37)}{110 \times 137 - (-37)^2} \times \frac{50}{200}$ $= \frac{5060 + 2442}{19030 - 1369} \times \frac{1}{4} = \frac{7502}{17661} \times \frac{1}{4}$ $= \frac{7502}{70644} = 0.106$ | <p>Regression Coefficient of Y on X</p> $b_{YX} = \frac{N \sum fd'xd'y - (\sum fd'x)(\sum fd'y)}{N \sum fd'x^2 - (\sum fd'x)^2} \times \frac{c_y}{c_x}$ $= \frac{110 \times 46 - 66(-37)}{110 \times 136 - (66)^2} \times \frac{200}{50}$ $= \frac{110 \times 46 - 66(-37)}{114960 - 4356} \times 4 = \frac{7502}{17661} \times 4$ $= \frac{7502}{10604} \times 4 = \frac{30008}{10604} = 2.829$ |
|--|---|

X. CONCLUSION

Statistics is a vital tool but students/learners often are stuck memorizing long, complicated formulas. The gap between understanding statistical theory and executing statistical calculations has long been a barrier in the learning process. This paper shows that the math does not have to be so hard. Once students/learners realize that the core method does not change, the topic becomes much less scary. By stripping away the unnecessary complexity of the math, students/learners can finally stop worrying about the formulas and start confidently using statistics to understand and predict the world around them. By making the math accessible, we ensure that the next generation of professionals can seamlessly bridge description and prediction, applying the power of statistics to solve the complex problems of tomorrow.

ACKNOWLEDGEMENT

I would like to express my gratitude to GLM5 for assisting me in drafting this English text. Any book on statistical methods consists of standard content; however, in the paper focus is on a highly logical sequencing of topics.

REFERENCES

- [1]. Gupta, S. C., & Kapoor, V. K., Fundamentals of mathematical statistics, Sultan Chand & Sons, 2020.
- [2]. Gupta, S. P., Statistical methods, Sultan Chand & Sons, 2021.
- [3]. Newbold, P., Carlson, W. L., & Thorne, B., Statistics for business and economics Pearson Education, 2019.
- [4]. Sharma, J. K. , Business statistics (4th ed.). Vikas Publishing House Pvt Limited, 2019.
- [5]. Spiegel, M. R., & Stephens, L. J., Schaum's outline of statistics, McGraw-Hill, 2008.