DegreeLive

BBA | Semester 6 | UK6DSCMGT303

Statistics for Business Decisions

Course Code: UK6DSCMGT303 • Module 2 Notes

degreelive.in

Module 2: Measures of Dispersion and Skewness

Range, Quartile Deviation, Mean Deviation, Standard Deviation, Variance, CV, and Skewness Coefficients

Chapter 1: The Epistemological Significance and Concepts of Dispersion

1.1 The Need for Measuring Dispersion in Business Intelligence

In business analytics, a measure of central tendency provides a single summary value that represents an entire statistical distribution. However, an average alone conveys an incomplete, and often dangerously deceptive, description of a population. Two entirely different commercial datasets may exhibit identical arithmetic means while possessing radically divergent internal structures.

Consider three competing sales branches—Branch X, Branch Y, and Branch Z—each employing five sales executives with monthly commissions recorded in thousands of rupees:

- Branch X: 50, 50, 50, 50, 50 (Mean = 50)
- Branch Y: 45, 48, 50, 52, 55 (Mean = 50)
- Branch Z: 10, 20, 50, 80, 90 (Mean = 50)

While all three branches report an identical average monthly commission of Rs. 50,000, Branch X exhibits complete uniformity with zero variation; Branch Y displays modest, disciplined variation around the mean; and Branch Z suffers from severe operational instability and extreme inequality. To evaluate risk, reliability, and process stability, management requires quantitative tools that measure the extent to which individual observations scatter, spread, or deviate away from the central average. This statistical spread is termed **Dispersion** (or Variation).

According to A. L. Bowley, "Dispersion is the measure of the variation of the items." Brooks and Dick define dispersion as "the degree of scatter or variation of the variables about a central value."

1.2 Managerial Objectives of Measuring Dispersion

The computation of dispersion metrics serves four critical executive purposes:

1. **Assessing the Reliability of an Average:** If the dispersion of a distribution is minimal, individual observations cluster tightly around the average, rendering the mean highly representative of the population. Conversely, if dispersion is vast, the average is unrepresentative and unreliable for operational benchmarking.
2. **Controlling Process Variation:** In total quality management (TQM) and Six Sigma manufacturing, variation is the primary driver of product defects. Measuring dispersion enables industrial engineers to pinpoint machine calibration drift and maintain tight manufacturing tolerances.
3. **Facilitating Comparative Analysis of Stability and Risk:** In investment portfolio management, financial analysts compare the dispersion of asset returns. High dispersion signifies

DegreeLive

high financial volatility and investment risk, whereas low dispersion reflects capital preservation and earnings stability.

4. **Serving as Foundational Inputs for Higher Statistical Theory:** Dispersion metrics—most notably the variance and standard deviation—are essential prerequisites for correlation, regression analysis, statistical quality control, and inferential hypothesis testing.

1.3 Absolute versus Relative Measures of Dispersion

Statistical theory bifurcates measures of dispersion into two distinct operational classes:

- **Absolute Measures of Dispersion:** Expressed in the identical physical units of measurement as the original raw data (e.g., rupees, kilograms, centimeters, hours). While absolute measures quantify the physical magnitude of spread within a single distribution, they cannot be deployed to compare two distributions possessing different units of measurement (e.g., comparing the dispersion of worker wages in rupees against output in kilograms) or differing scales of magnitude (e.g., comparing a startup's sales against a Fortune 500 conglomerate).

  Examples include Range, Quartile Deviation, Mean Deviation, and Standard Deviation.

- **Relative Measures of Dispersion (Coefficients of Dispersion):** Pure, unit-free dimensionless ratios or percentages derived by dividing an absolute measure of dispersion by an appropriate measure of central tendency. Relative measures facilitate direct, unbiased comparisons across disparate series regardless of measurement units or baseline scales.

  Examples include Coefficient of Range, Coefficient of Quartile Deviation, Coefficient of Mean Deviation, and the Coefficient of Variation (CV).

Chapter 2: Positional and Distance Measures: Range and Quartile Deviation

2.1 The Range and Its Coefficient

The **Range** (R\$) is the simplest measure of dispersion. It represents the absolute mathematical difference between the largest (maximum) observation and the smallest (minimum) observation in a series:

> $$\text{Absolute Range } (R) = L - S$$
>
> $$\text{Coefficient of Range} = \frac{L-S}{L+S}$$
>
> **Where:** L is the Largest Value (or Upper Limit of highest class interval), and S is the Smallest Value (or Lower Limit of lowest class interval).

Managerial Merits and Demerits: The range is instantly calculated and intuitively understood, finding wide application in daily weather reports, stock market daily high-low price spreads, and industrial quality control charts (X-bar and R-charts). However, its severe demerits include: (a) it relies exclusively on the two extreme outlier values, completely ignoring the internal distribution of the

remaining observations; (b) it is hyper-sensitive to anomalous outlier spikes; and (c) it cannot be computed for open-ended frequency distributions.

2.2 The Quartile Deviation (Semi-Interquartile Range)

To overcome the outlier vulnerability of the range, statisticians utilize the **Interquartile Range (IQR)**, which measures the spread of the middle 50% of the distribution by computing the difference between the Third (Upper) Quartile (Q_3) and the First (Lower) Quartile (Q_1).

The **Quartile Deviation (QD)** (or **Semi-Interquartile Range**) is defined as half of the interquartile range:

$$\text{Interquartile Range (IQR)} = Q_3 - Q_1$$

$$\text{Quartile Deviation } (QD) = \frac{Q_3 - Q_1}{2}$$

$$\text{Coefficient of Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Managerial Suitability: Quartile deviation is superior to the range because it is unaffected by extreme tails, making it the preferred measure of dispersion for open-ended distributions (e.g., "sales under Rs. 10,000" or "sales above Rs. 1,00,000") and severely skewed socioeconomic income datasets. However, it is not based on all observations and is not amenable to advanced algebraic treatments.

Chapter 3: Mean Deviation (Average Deviation)

3.1 Concept and Methodological Logic

The **Mean Deviation (MD)** is the arithmetic mean of the absolute numerical deviations of all observations taken from a measure of central tendency (typically the Mean or the Median). Unlike range and quartile deviation, mean deviation is based on every single observation in the dataset.

In standard algebra, the sum of deviations taken from the actual mean is identically zero ($\sum(X - \bar{X}) = 0$). To prevent positive and negative deviations from canceling each other out, the mean deviation takes the **Absolute Values** of deviations, ignoring algebraic signs ($|d| = |X - A|$).

3.2 Computational Formulas across Data Series

1. Individual Series

$$MD_{\bar{X}} = \frac{\sum |X - \bar{X}|}{N} \quad \text{and} \quad MD_M = \frac{\sum |X - M|}{N}$$

2. Discrete and Continuous Frequency Distributions

$$MD = \frac{\sum f|d|}{N} = \frac{\sum f|X-A|}{\sum f}$$

Where: X represents individual values in discrete series or class mid-points (m) in continuous series, and A represents either $\bar{X}$ or M.

3. Coefficients of Mean Deviation

$$\text{Coefficient of } MD_{\bar{X}} = \frac{MD_{\bar{X}}}{\bar{X}} \quad \text{and} \quad \text{Coefficient of } MD_M = \frac{MD_M}{M}$$

3.3 Mathematical Invariant: Minimal Sum of Absolute Deviations

A profound mathematical property of the median establishes that the sum of the absolute deviations of observations is at a minimum when taken from the **Median** rather than any other point:

$$\sum |X - M| < \sum |X - A| \quad \text{for any } A \neq M$$

Theoretical Principle: Consequently, the Mean Deviation calculated from the Median is mathematically smaller than the Mean Deviation calculated from the Mean ($MD_M \leq MD_{\bar{X}}$). However, in commercial practice, mean deviation from the mean is also widely reported.

Chapter 4: Standard Deviation, Variance, and Advanced Algebraic Properties

4.1 Definitional Foundations

Introduced by Karl Pearson in 1893, the **Standard Deviation** (conventionally denoted by the Greek lower-case letter sigma, σ) is universally regarded as the most rigorous, reliable, and scientifically preeminent measure of dispersion. It is defined as the positive square root of the arithmetic mean of the squared deviations of values taken from their arithmetic mean.

The square of the standard deviation is termed **Variance** (σ^2):

$$\text{Variance } (\sigma^2) = \frac{\sum (X-\bar{X})^2}{N} \implies \text{Standard Deviation } (\sigma) = \sqrt{\frac{\sum (X-\bar{X})^2}{N}}$$

Mathematical Superiority: By squaring the deviations ($(X - \bar{X})^2$), the standard deviation eliminates negative signs through rigorous algebraic squaring rather than the mathematically arbitrary sign-dropping of mean deviation.

4.2 Computational Methodologies across Data Distributions

1. Individual Series

Direct Method: $\sigma = \sqrt{\frac{\sum x^2}{N}}$ where $x = X - \bar{X}$

Short-Cut (Assumed Mean) Method: $\sigma = \sqrt{\frac{\sum d^2}{N} - \left(\frac{\sum d}{N}\right)^2}$ where $d = X - A$

2. Discrete and Continuous Frequency Distributions

In continuous series, where m is the class mid-point, A is the assumed mean, and c is the uniform class interval width, the **Step-Deviation Method** provides computational elegance:

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

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

Where: $N = \sum f$. The class width c is multiplied strictly *outside* the square root radical.

4.3 Mathematical Properties of the Standard Deviation

The standard deviation possesses six fundamental algebraic theorems that underpin modern econometrics and statistical inference:

1. **Minimal Root Mean Square Deviation:** The standard deviation is strictly smaller than the root mean square deviation taken from any point other than the arithmetic mean ($\sigma < sA\$$).
2. **Change of Origin Invariance:** The standard deviation is completely independent of a change of origin. If a constant a is added to or subtracted from every value in a dataset ($Y = X \pm a\$$), the standard deviation remains entirely unchanged: $\sigma Y = \sigma X\$$.
3. **Change of Scale Responsiveness:** The standard deviation is directly responsive to a change of scale. If every value is multiplied or divided by a constant b ($Y = bX\$$), the standard deviation scales by the absolute magnitude of that factor: $\sigma Y = |b|\sigma X\$$.
4. **Combined Standard Deviation:** If two independent sub-groups have known sample sizes ($N1, N2\$$), arithmetic means ($\bar{X}1, \bar{X}2\$$), and standard deviations ($\sigma 1, \sigma 2\$$), the combined pooled standard deviation ($\sigma 12\$$) of the consolidated group is formulated as:

   $$\sigma 12 = \sqrt{\frac{N1(\sigma 1^2 + d1^2) + N2(\sigma 2^2 + d2^2)}{N1 + N2}}$$

   Where: $d_1 = \bar{X}_1 - \bar{X}_{12}$ and $d_2 = \bar{X}_2 - \bar{X}_{12}$, and $\bar{X}_{12}$ represents the combined arithmetic mean.

5. **The Empirical Rule for Normal Distributions:** In a symmetrical, bell-shaped Gaussian distribution:
   - $\bar{X} \pm 1\sigma$ encompasses approximately **68.27%** of all observations.

- $\bar{X} \pm 2\sigma$ encompasses approximately **95.45%** of all observations.
- $\bar{X} \pm 3\sigma$ encompasses approximately **99.73%** of all observations.

6. **Empirical Relationships among Dispersion Metrics:** In moderately symmetrical distributions:

> $$QD \approx \frac{2}{3}\sigma = 0.6745\sigma \quad \text{and} \quad MD \approx \frac{4}{5}\sigma = 0.7979\sigma$$
>
> $$6QD \approx 5MD \approx 4\sigma$$

Chapter 5: The Coefficient of Variation (CV) and Business Risk Diagnostics

5.1 Definitional Architecture

Formulated by Karl Pearson, the **Coefficient of Variation (CV)** is the percentage ratio of the standard deviation to the arithmetic mean:

> $$CV = \frac{\sigma}{\bar{X}} \times 100$$

5.2 Executive Interpretation and Decision Rules

The Coefficient of Variation is the definitive relative measure for comparing consistency, stability, uniformity, and risk across competing commercial series:

Analytical Scenario	Higher Coefficient of Variation (Higher CV)	Lower Coefficient of Variation (Lower CV)
Factory Worker Production	Less Consistent, Erratic, High Fluctuation	**More Consistent, Uniform, Stable Production**
Stock Market Equity Returns	**Higher Financial Risk, Elevated Volatility**	Lower Financial Risk, Capital Preservation
Product Quality Tolerances	High Defect Rates, Inconsistent Quality	**Disciplined Process Control, High Reliability**
Regional Sales Performance	Vulnerable, Highly Erratic Revenue Cycles	**Resilient, Predictable Revenue Streams**

> **Universal Managerial Decision Rule:** When comparing consistency, stability, or homogeneity, the series possessing the **LOWER Coefficient of Variation** is judged to be superior and more reliable. When

> evaluating financial risk, the series with the **HIGHER CV** carries greater risk per unit of expected return.

Chapter 6: The Mathematical Foundations and Measures of Skewness

6.1 The Concept of Skewness (Asymmetry)

While central tendency identifies the central location of data and dispersion measures the spread around that center, neither reveals whether the distribution is symmetrical or distorted toward one tail. **Skewness** signifies the lack of symmetry in a statistical frequency distribution. A distribution is skewed when observations are clustered heavily on one side of the average, trailing off into an elongated tail on the opposite side.

6.2 Symmetrical versus Asymmetrical Distributions

- **Symmetrical (Bell-Shaped) Distribution:** A distribution where values at equal distances on either side of the central peak occur with identical frequencies. The curve is perfectly balanced. In a symmetrical distribution:

$$\text{Mean} = \text{Median} = \text{Mode} \implies \bar{X} = M = Z$$
$$Q_3 - M = M - Q_1 \quad \text{and} \quad \text{Skewness} = 0$$

- **Positively Skewed Distribution (Right-Skewed):** The distribution possesses an elongated tail extending toward the right (positive) side of the scale, pulled outward by exceptionally large values. Typical of corporate executive salaries and household wealth distributions:

$$\text{Mean} > \text{Median} > \text{Mode} \implies \bar{X} > M > Z$$
$$Q_3 - M > M - Q_1$$

- **Negatively Skewed Distribution (Left-Skewed):** The distribution possesses an elongated tail stretching toward the left (negative) side, pulled down by exceptionally low values (e.g., student age distributions in adult evening MBA classes):

$$\text{Mode} > \text{Median} > \text{Mean} \implies Z > M > \bar{X}$$
$$M - Q_1 > Q_3 - M$$

6.3 Measures of Skewness

1. Karl Pearson's Coefficients of Skewness

Karl Pearson formulated two standardized mathematical coefficients based on the distance between central averages relative to the standard deviation:

First Coefficient (Mode-Based): $Sk_p = \frac{X - Z}{\sigma}$

Second Coefficient (Median-Based): $Sk_p = \frac{3(X - M)}{\sigma}$

Theoretical Range: Karl Pearson's coefficient typically fluctuates within the boundaries of **-3 to +3**. A positive coefficient indicates positive skewness; zero indicates perfect symmetry; a negative coefficient indicates negative skewness.

2. Bowley's Quartile Coefficient of Skewness

A. L. Bowley constructed a positional measure of skewness based on the relative distances of the Upper and Lower Quartiles from the Median:

$$Sk_b = \frac{(Q3 - M) - (M - Q1)}{(Q3 - M) + (M - Q1)} = \frac{Q3 + Q1 - 2M}{Q3 - Q1}$$

Theoretical Range: Bowley's coefficient is strictly bounded between **-1 and +1**. It is particularly valuable for open-ended frequency distributions where the mean and standard deviation cannot be computed.

Chapter 7: Comprehensive Worked Numerical Problems in Dispersion and Skewness

7.1 Worked Problem 1: Continuous Series Standard Deviation, Variance, and CV

The daily production output (in metric units) of 100 manufacturing shifts in an automated factory is distributed across continuous class intervals as follows:

Output Class	Shifts (f)	Mid-Point (m)	d' = (m - 45) / 10	f · d'	(d')^2	f · (d')^2
20–30	10	25	-2	-20	4	40
30–40	20	35	-1	-20	1	20
40–50	40	45 (A)	0	0	0	0
50–60	20	55	+1	+20	1	20
60–70	10	65	+2	+20	4	40
Total	**N = 100**			**Sum fd' = 0**		**Sum f(d')^2 = 120**

Step-by-Step Mathematical Resolution:

1. Calculation of Arithmetic Mean:

$$\bar{X} = A + \left(\frac{\sum fd'}{N} \times c\right) = 45 + \left(\frac{0}{100} \times 10\right) = 45.00 \text{ units}$$

2. Calculation of Standard Deviation and Variance:

$$\sigma = \sqrt{\frac{\sum f(d')^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times c = \sqrt{\frac{120}{100} - (0)^2} \times 10$$

$$\sigma = \sqrt{1.20} \times 10 = 1.095445 \times 10 = 10.954 \approx 10.95 \text{ units}$$

$$\text{Variance } (\sigma^2) = (10.954)^2 = 120.00 \text{ units}^2$$

3. Calculation of Coefficient of Variation:

$$CV = \frac{\sigma}{\bar{X}} \times 100 = \frac{10.954}{45.00} \times 100 = 24.34\%$$

7.2 Worked Problem 2: Comparative Consistency Analysis of Two Factories

An enterprise operates two assembly factories, Factory Alpha and Factory Beta, manufacturing identical consumer electronic units. Weekly output records over a one-year operating horizon provide the following operational metrics:

- **Factory Alpha:** Mean Weekly Output ($\bar{X}_A$): 800 units; Standard Deviation (σ_A): 40 units.
- **Factory Beta:** Mean Weekly Output ($\bar{X}_B$): 1,200 units; Standard Deviation (σ_B): 48 units.

Required:

1. Which factory exhibits greater aggregate production scale?
2. Which factory demonstrates greater consistency and stability in manufacturing operations?

Step-by-Step Mathematical Resolution:

1. Scale Evaluation:

> Factory Beta achieves higher average weekly output (1,200 units vs. 800 units for Factory Alpha), demonstrating 50% greater production scale.

2. Consistency Evaluation via Coefficient of Variation:

$$CV_A = \frac{\sigma_A}{\bar{X}_A} \times 100 = \frac{40}{800} \times 100 = 5.00\%$$

$$CV_B = \frac{\sigma_B}{\bar{X}_B} \times 100 = \frac{48}{1{,}200} \times 100 = 4.00\%$$

> **Managerial Decision:** Although Factory Beta possesses a slightly higher absolute standard deviation (48 units vs. 40 units), its Coefficient of Variation is significantly lower (**4.00% vs. 5.00%**). Therefore, **Factory Beta is more consistent, stable, and disciplined** in its production throughput relative to its operational scale.

7.3 Worked Problem 3: Combined Standard Deviation Computation

Two industrial shifts produce metallic rods. Shift 1 has $N_1 = 50$, mean diameter $\bar{X}_1 = 20.0$ mm, and standard deviation $\sigma_1 = 2.0$ mm. Shift 2 has $N_2 = 100$, mean diameter $\bar{X}_2 = 23.0$ mm, and standard deviation $\sigma_2 = 3.0$ mm.

Required: Compute the combined standard deviation of the consolidated 150-rod production batch.

Step-by-Step Mathematical Resolution:

1. Combined Mean:

$$\bar{X}_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2} = \frac{50(20) + 100(23)}{50 + 100} = \frac{1{,}000 + 2{,}300}{150} = \frac{3{,}300}{150} = 22.0 \text{ mm}$$

2. Deviation of Individual Means from Combined Mean:

$$d_1 = \bar{X}_1 - \bar{X}_{12} = 20.0 - 22.0 = -2.0 \implies d_1^2 = 4.0$$
$$d_2 = \bar{X}_2 - \bar{X}_{12} = 23.0 - 22.0 = +1.0 \implies d_2^2 = 1.0$$

3. Combined Standard Deviation Formula:

$$\sigma_{12} = \sqrt{\frac{N_1(\sigma_1^2 + d_1^2) + N_2(\sigma_2^2 + d_2^2)}{N_1 + N_2}}$$
$$\sigma_1^2 + d_1^2 = 2.0^2 + 4.0 = 4.0 + 4.0 = 8.0$$
$$\sigma_2^2 + d_2^2 = 3.0^2 + 1.0 = 9.0 + 1.0 = 10.0$$
$$\sigma_{12} = \sqrt{\frac{50(8.0) + 100(10.0)}{150}} = \sqrt{\frac{400 + 1{,}000}{150}} = \sqrt{\frac{1{,}400}{150}} = \sqrt{9.333} \approx 3.055 \text{ mm}$$

7.4 Worked Problem 4: Karl Pearson and Bowley Coefficients of Skewness

A regional distribution network recorded monthly fuel efficiency metrics across 200 fleet vehicles. Statistical processing provided the following computed aggregates:

- Arithmetic Mean (\bar{X}): 18.50 km/litre
- Median (M): 19.00 km/litre
- Mode (Z): 20.00 km/litre

- Standard Deviation (σ): 2.50 km/litre
- First Quartile (Q_1): 16.80 km/litre
- Third Quartile (Q_3): 20.60 km/litre

Required: Compute (a) Karl Pearson's Coefficient of Skewness (both Mode and Median methods); and (b) Bowley's Quartile Coefficient of Skewness. Interpret the directional asymmetry.

Step-by-Step Mathematical Resolution:

1. Karl Pearson's First Coefficient (Mode-Based):

> $$Sk_p = \frac{\bar{X} - Z}{\sigma} = \frac{18.50 - 20.00}{2.50} = \frac{-1.50}{2.50} = -0.60$$

2. Karl Pearson's Second Coefficient (Median-Based):

> $$Sk_p = \frac{3(\bar{X} - M)}{\sigma} = \frac{3(18.50 - 19.00)}{2.50} = \frac{3(-0.50)}{2.50} = \frac{-1.50}{2.50} = -0.60$$
>
> Both Pearson formulas yield perfectly congruent coefficients of **-0.60**.

3. Bowley's Quartile Coefficient of Skewness:

> $$Sk_b = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} = \frac{20.60 + 16.80 - 2(19.00)}{20.60 - 16.80}$$
>
> $$Q_3 + Q_1 = 37.40, \quad 2M = 38.00 \implies \text{Numerator} = 37.40 - 38.00 = -0.60$$
>
> $$\text{Denominator} = 20.60 - 16.80 = 3.80$$
>
> $$Sk_b = \frac{-0.60}{3.80} \approx -0.158$$
>
> **Analytical Synthesis:** Both Pearson's coefficient (-0.60) and Bowley's coefficient (-0.158) are strictly negative ($Sk < 0$), and the mean is smaller than the median and mode ($\bar{X} < M < Z$). This confirms that the fleet fuel efficiency distribution is **Negatively Skewed (Left-Skewed)**, possessing an elongated tail of vehicles exhibiting abnormally depressed fuel performance.