DegreeLive

BBA | Semester 6 | UK6DSCMGT303

Statistics for Business Decisions

Course Code: UK6DSCMGT303 • Module 1 Notes

degreelive.in

Module 1: Statistics and Measures of Central Tendency

Foundations, Scope, Distrust, Ideal Averages, Mean, Median, Mode, GM, and HM

Chapter 1: The Epistemological Architecture and Business Scope of Statistics

1.1 Definitional Foundations and Historical Evolution

In modern industrial and commercial administration, business decisions cannot be formulated prudently upon unverified intuition, subjective guesswork, or anecdotal folklore. Every operational choice—ranging from supply chain inventory procurement and capital investment budgeting to digital consumer segmentation—demands objective, verifiable quantitative intelligence. The scientific discipline that extracts actionable knowledge from numerical observations is **Statistics**.

Etymologically, the term statistics is derived from the Latin word *status*, the Italian word *statista*, the German word *Statistik*, and the French word *statistique*, each historically denoting a state, political governance, or political arithmetic. In classical antiquity, sovereign monarchs deployed rudimentary statistics primarily for fiscal taxation assessments and military conscription censuses. In modern enterprise administration, statistics has evolved into the universal mathematical language of empirical science, decision theory, and corporate risk management.

In formal academic literature, statistics is defined under two distinct perspectives:

1. **Plural Sense (Statistical Data):** In the plural sense, statistics signifies aggregate numerical facts systematically collected in a specific field of inquiry. Horace Secrist provides the definitive plural formulation: "By statistics we mean aggregates of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other."
2. **Singular Sense (Statistical Methods):** In the singular sense, statistics denotes the entire body of scientific methods deployed for the collection, organization, presentation, analysis, and interpretation of quantitative data. Croxton and Cowden define statistics concisely in this operational sense as "the collection, presentation, analysis and interpretation of numerical data."

1.2 Essential Characteristics of Statistical Data

To qualify as statistical data under Secrist's rigorous definition, numerical observations must satisfy six fundamental criteria:

- **Aggregates of Facts:** An isolated, single numerical figure (e.g., "employee Ramesh earns Rs. 45,000 per month") does not constitute statistics because it offers no basis for comparison. Conversely, a comprehensive dataset showing the monthly wage distribution of 5,000 factory employees constitutes statistical data.

Degreelive

- **Affected by Multiplicity of Causes:** Business phenomena are rarely governed by a single isolated variable. For example, monthly corporate sales revenue is simultaneously influenced by product pricing, advertising expenditure, competitor campaigns, consumer disposable income, seasonal weather patterns, and macroeconomic monetary policies.
- **Numerically Expressed:** Qualitative descriptions such as "high customer satisfaction", "loyal employees", or "inferior product quality" do not become statistical data until they are operationalized through standardized psychometric scales or quantitative metric units.
- **Enumerated or Estimated according to Reasonable Standards of Accuracy:** When examining finite, accessible populations (e.g., number of branches of a bank), complete physical enumeration is feasible. When dealing with vast or infinite universes (e.g., television viewership or consumer preferences), scientific estimation via probability sampling is executed, maintaining disciplined confidence intervals rather than illusory mathematical perfection.
- **Collected in a Systematic Manner for a Predetermined Purpose:** Haphazard, unstructured data accumulation produces noise rather than actionable intelligence. The inquiry objectives, operational variables, sampling frame, and measurement protocols must be formally documented prior to field execution.
- **Placed in Relation to Each Other (Comparability):** Data must be logically homogeneous across time, geography, or administrative units to enable meaningful comparative benchmarking. Comparing the revenue of an Indian steel mill in 2024 against the rainfall of London in 1890 violates statistical comparability.

1.3 Functions and Practical Utility of Statistics in Business Decisions

Statistics fulfills six indispensable functions across the corporate enterprise hierarchy:

- **Condensation of Massive Datasets:** Raw corporate operational data—such as millions of point-of-sale retail transactions—is cognitively unmanageable. Statistical summaries (frequency distributions, measures of central tendency, dispersion metrics) condense voluminous data into lucid executive summaries without sacrificing vital structural patterns.
- **Facilitating Comparative Analysis:** Statistics provides standardized relative metrics (ratios, percentages, index numbers, z-scores) that enable management to evaluate performance across disparate operating divisions, competing brands, and historical fiscal quarters.
- **Formulating and Testing Business Hypotheses:** When marketing leadership considers whether a redesigned e-commerce interface boosts conversion rates, inferential statistics provides empirical hypothesis testing tools (t-tests, ANOVA, Chi-Square) to determine whether observed sales surges represent genuine consumer behavior shifts or random chance.
- **Forecasting and Prospective Planning:** Modern supply chains rely upon statistical regression, econometric forecasting, and time-series decomposition to project future product demand, raw material requirements, and cash flow obligations.
- **Quality Control and Process Optimization:** In manufacturing operations, Statistical Process Control (SPC) charts monitor production tolerances, distinguishing between routine common-cause process variation and disruptive special-cause defects.

Degreelive

- **Measuring Risk and Financial Volatility:** Portfolio managers use standard deviations, variance-covariance matrices, and beta coefficients to quantify investment risk, optimize capital asset allocation, and hedge foreign exchange exposure.

1.4 Inherent Limitations of the Statistical Method

Despite its analytical power, statistics possesses severe structural limitations that business decision-makers must never overlook:

- *Statistics Deals Only with Quantitative Aggregates:* Statistics cannot analyze non-quantifiable qualitative phenomena directly (e.g., organizational leadership integrity, worker morale, brand prestige) unless converted into proxy scale metrics.
- *Statistics Deals with Groups, Not with Specific Individuals:* Statistical conclusions represent aggregate probabilistic tendencies across populations. A mortality table may establish that life expectancy is 72 years, but it cannot predict the specific lifespan of an individual policyholder.
- *Statistical Laws are Probabilistic, Not Exact:* Unlike the absolute universal laws of classical physics or mathematics ($2 + 2 = 4$), statistical laws hold true only on average over large numbers of observations and across extended time horizons.
- *Data Can Be Readily Misinterpreted or Manipulated:* Devoid of rigorous contextual understanding, statistical figures can be distorted deliberately or innocently to support fallacious assertions.
- *Statistics Is Only a Decision Tool, Not a Replacement for Judgment:* Quantitative models supply empirical evidence, but final executive choices must integrate qualitative legal, ethical, macroeconomic, and geopolitical considerations.

Chapter 2: The Phenomenology of Distrust in Statistics and Ethical Safeguards

2.1 The Roots and Manifestations of Public Skepticism

Public discourse frequently harbors profound skepticism toward statistical claims, crystallized in famous satirical aphorisms such as "Statistics can prove anything" and Benjamin Disraeli's cynical ranking: "There are three kinds of lies: lies, damned lies, and statistics." This pervasive distrust does not stem from flaws in statistical mathematical theory itself; rather, it arises from the pervasive misuse, deliberate manipulation, and incompetent application of statistical tools by commercial advertisers, political campaigners, and biased advocates.

2.2 Common Methods of Statistical Distortion and Deception

Decision-makers must identify and dismantle seven classic statistical fallacies:

1. **Selection Bias and Unrepresentative Samples:** Drawing conclusions about a national consumer population based on a self-selected online poll of affluent metropolitan smartphone users.

2. **Deliberate Omission of Sample Size:** Commercial claims asserting that "75% of doctors recommend our brand" without disclosing that the sample comprised only four surveyed physicians.
3. **Spurious Correlation and Confounding Variables:** Confusing mathematical correlation with causal mechanisms (e.g., noting that ice cream sales and drowning incidents correlate positively during summer months, ignoring the confounding role of ambient temperature).
4. **Misleading Graphical Truncation:** Truncating the vertical Y-axis of a bar chart to make a minuscule 0.5% sales increase visually resemble a dramatic tenfold commercial expansion.
5. **The Fallacy of the Wrong Average:** Selectively reporting the Arithmetic Mean instead of the Median when describing severely skewed distributions (e.g., using mean household wealth in a town where a billionaire resides to present an illusion of widespread municipal prosperity).
6. **Ignoring Operational Definitions and Context:** Presenting unemployment statistics that define an employed individual as anyone working a single hour per week, masking acute underemployment.
7. **Post-Hoc Data Dredging (P-Hacking):** Testing dozens of unrelated variables against an outcome until a random, statistically significant relationship ($p < 0.05$) emerges by sheer chance, reporting it as a groundbreaking commercial insight.

2.3 Ethical Safeguards and Professional Best Practices

To eliminate distrust, modern management demands strict adherence to empirical integrity protocols: (a) transparent public disclosure of sample sizes, response rates, and sampling frames; (b) publishing full methodology blueprints and raw data distributions alongside summary metrics; (c) reporting measures of dispersion (standard deviation, interquartile range) whenever central tendency averages are presented; and (d) maintaining value-neutral analytical objectivity.

Chapter 3: Theoretical Foundations of Central Tendency and the Ideal Average Criteria

3.1 Concept and Managerial Significance of Central Tendency

When an enterprise compiles thousands of numerical records—such as monthly sales receipts across nationwide branches—the individual observations cluster naturally around a central focal point. **Central Tendency** refers to this statistical tendency of quantitative data to concentrate around a typical middle value. A **Measure of Central Tendency** (or **Average**) is a single representative value that synthesizes and characterizes the entire mass of data, acting as a cognitive shorthand for the population.

The primary objectives of computing averages in business administration are: (1) to condense massive, complex datasets into a single graspable metric; (2) to facilitate meaningful comparative benchmarking between diverse business units or operating periods; (3) to provide baseline inputs for higher-order statistical modeling (dispersion, skewness, regression, hypothesis testing); and (4) to establish operational norms and performance targets.

3.2 Yule's Criteria for an Ideal Average

In classical statistical theory, G. Udny Yule formulated six rigorous mathematical and practical criteria that an ideal measure of central tendency must fulfill:

1. **Rigidly Defined:** The average must be governed by an unambiguous mathematical formula so that different investigators analyzing the identical dataset arrive at the exact same numerical result, leaving no room for subjective interpretation.
2. **Based on All Observations:** The average should incorporate every single data item in the series. If an average ignores substantial subsets of data, it ceases to be genuinely representative of the total distribution.
3. **Easily Understandable and Simple to Compute:** The metric must possess intuitive conceptual clarity so that non-technical managers and business stakeholders can readily interpret its operational meaning.
4. **Amenable to Further Algebraic Treatment:** The mathematical form of the average should facilitate advanced algebraic manipulation—such as computing combined averages of merged groups or calculating mathematical deviations—without returning to raw micro-data.
5. **Sampling Stability:** If multiple random samples are drawn independently from the same underlying population, the calculated averages across samples should exhibit minimal variation. High sampling stability is vital for inferential testing.
6. **Resilient against Extreme Values:** The average should not be unduly distorted or pulled off-center by a few abnormally large or small outlier observations.

3.3 The Taxonomy of Averages

Statistical averages are categorized into two fundamental groups:

- **Mathematical Averages:** Derived through comprehensive algebraic computation involving every data value. Comprises the **Arithmetic Mean**, **Geometric Mean**, and **Harmonic Mean**.
- **Positional Averages:** Determined primarily by the relative physical location or position of values within an ordered array. Comprises the **Median** and **Mode** (along with partition values: Quartiles, Deciles, and Percentiles).

Chapter 4: Mathematical Theory and Computation of the Arithmetic Mean

4.1 Definitional Framework and Computation across Data Series

The **Arithmetic Mean** (commonly termed the mean or simple average) is the most widely utilized measure of central tendency. It is defined as the quotient obtained by dividing the sum total of all values in a statistical series by the total number of observations.

1. Individual Series (Raw Ungrouped Data)

Where a series consists of N individual observations $X_1, X_2, \ldots, X_N$\$:

> Direct Method: $\bar{X} = \frac{\sum X}{N} = \frac{X_1 + X_2 + \cdots + X_N}{N}$
>
> Short-Cut (Assumed Mean) Method: $\bar{X} = A + \frac{\sum d}{N}$
>
> **Where:** A is an arbitrarily assumed mean, and $d = X - A$ represents the deviation of each individual observation from the assumed mean.

2. Discrete Series (Ungrouped Frequency Distribution)

Where distinct values $X_1, X_2, \ldots, X_n$ occur with respective frequencies $f_1, f_2, \ldots, f_n$\$, with total frequency $N = \sum f$\$:

> Direct Method: $\bar{X} = \frac{\sum (fX)}{N}$
>
> Short-Cut Method: $\bar{X} = A + \frac{\sum (fd)}{N}$
>
> **Where:** $d = X - A$, and $N = \sum f$.

3. Continuous Series (Grouped Frequency Distribution with Class Intervals)

Where data is organized into class intervals (e.g., 10–20, 20–30), the mid-value (m\$) of each class interval represents that class: $m = \frac{\text{Lower Limit} + \text{Upper Limit}}{2}$\$.

> Direct Method: $\bar{X} = \frac{\sum (fm)}{N}$
>
> Step-Deviation Method: $\bar{X} = A + \left(\frac{\sum (fd')}{N} \times c \right)$
>
> **Where:** $d' = \frac{m - A}{c}$, A is the assumed mean, c represents the uniform class interval magnitude, and $N = \sum f$.

4.2 Fundamental Mathematical Properties of the Arithmetic Mean

The arithmetic mean possesses five unique algebraic properties that make it the mathematical cornerstone of statistical theory:

1. **Zero Sum of Algebraic Deviations:** The algebraic sum of the deviations of individual observations from their actual arithmetic mean is always equal to zero:

   > $$\sum (X - \bar{X}) = 0 \quad \text{and} \quad \sum f(X - \bar{X}) = 0$$
   >
   > **Significance:** The mean acts as the exact physical center of gravity of the distribution; positive deviations above the mean perfectly balance negative deviations below it.

2. **Minimum Sum of Squared Deviations:** The sum of the squared deviations of values from their arithmetic mean is strictly smaller than the sum of squared deviations taken from any other

arbitrary point (A):

$$\sum(X - \bar{X})^2 < \sum(X - A)^2 \quad \text{for any } A \neq \bar{X}$$

Significance: This property provides the foundational justification for the **Method of Least Squares** in regression analysis.

3. **Aggregate Value Reproduction:** The product of the arithmetic mean and the total number of observations equals the grand aggregate sum of all values in the population:

$$N\bar{X} = \sum X$$

4. **Combined Arithmetic Mean:** If two or more independent sub-groups have known sample sizes ($N_1, N_2, \ldots, N_k$\$) and known arithmetic means ($\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k$\$), the combined pooled mean of the consolidated group can be derived directly without accessing raw records:

$$\bar{X}_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$$
$$\bar{X}_{12\ldots k} = \frac{\sum(N_i\bar{X}_i)}{\sum N_i}$$

5. **Linear Shift Invariance:** If every value in a distribution is increased by an additive constant a\$ and multiplied by a scale factor b\$ ($Y = a + bX$\$), the new arithmetic mean changes accordingly: $\bar{Y} = a + b\bar{X}$\$.

4.3 Weighted Arithmetic Mean

In many business applications, different items do not carry equal economic importance. For example, in computing a Cost of Living Index, household spending on staple food grains carries vastly higher importance than spending on entertainment. The **Weighted Arithmetic Mean** assigns explicit numerical weights ($W_1, W_2, \ldots, W_n$\$) proportional to the relative operational significance of each item:

$$\bar{X}_w = \frac{\sum(W \cdot X)}{\sum W}$$

Chapter 5: Positional Averages: Median and Partition Values

5.1 The Concept and Computational Logic of the Median

The **Median** (M\$) is the positional middle value of a statistical series when all individual observations are arranged in order of magnitude (ascending or descending). It partitions the total

frequency distribution into exactly two equal halves: 50% of observations lie below the median, and 50% lie above it.

1. Median in Individual Series

Data must first be arranged in ascending order:

> $$\text{Position of Median} = \left(\frac{N+1}{2}\right)^{\text{th}} \text{ item}$$
>
> If N is odd, the median is the exact middle observation. If N is even, the median is the arithmetic mean of the two middle observations: $\frac{N}{2}$ and $\frac{N}{2}+1$.

2. Median in Discrete Series

Cumulative frequencies (cf\$) are compiled. The median position is identified using $\frac{N+1}{2}$\$, and the value corresponding to the cumulative frequency equal to or immediately exceeding this rank is the median.

3. Median in Continuous Series

In continuous grouped data, the median class is identified using the position $\frac{N}{2}$\$. The exact interpolated median value within the class is computed using the linear interpolation formula:

> $$M = L + \left(\frac{\frac{N}{2} - cf}{f}\right) \times c$$
>
> **Where:** L = Lower boundary limit of the median class;
> N = Total frequency ($\sum f$);
> cf = Cumulative frequency of the pre-median class (immediately preceding the median class);
> f = Simple frequency of the median class itself;
> c = Class width (magnitude) of the median class.

5.2 Partition Values: Quartiles, Deciles, and Percentiles

Partition values divide an ordered distribution into equal fractional divisions:

- **Quartiles:** Divide data into four equal parts using three partition points (Q_1, Q_2, Q_3\$). Q_1 (Lower Quartile) demarcates the bottom 25%; Q_2 is the Median (50%); and Q_3 (Upper Quartile) demarcates the top 25%:

  > $$Q1 = L + \left(\frac{\frac{N}{4} - cf}{f}\right) \times c \quad \text{and} \quad Q3 = L + \left(\frac{\frac{3N}{4} - cf}{f}\right) \times c$$

- **Deciles:** Divide data into ten equal parts using nine points ($D_1, D_2, \ldots, D_9$\$), where D_k represents the $\frac{kN}{10}$ positional value.

- **Percentiles:** Divide data into one hundred equal parts using ninety-nine points ($P_1, P_2, \ldots, P_{99}$), where P_k represents the $\frac{kN}{100}$ positional value. Widely deployed in standardized testing, customer satisfaction benchmarks, and income distribution percentiles.

5.3 Merits, Demerits, and Managerial Suitability of the Median

Primary Merits: (1) immune to extreme outlier distortion; (2) can be calculated cleanly for open-ended frequency distributions (e.g., "above Rs. 1,00,000"); (3) possesses clear physical meaning; and (4) can be determined graphically from cumulative frequency ogives.

Demerits: (a) not based on all observations; (b) not amenable to combined group algebraic manipulation; and (c) exhibits lower sampling stability than the mean.

Chapter 6: The Mode and Empirical Frequency Distribution Relationships

6.1 Concept and Determination of the Mode

The **Mode** (Z) is that value in a statistical distribution which occurs with the greatest frequency. It represents the point of maximum concentration, peak popularity, or dominant fashion in a series. In commercial management, the mode identifies the most popular ready-made shoe size, the highest-selling automobile color variant, or the most frequent daily order size.

1. Mode in Individual and Discrete Series

Determined by physical inspection identifying the value with the highest frequency. When two values share the peak frequency, the distribution is **Bimodal**; when multiple peaks exist, it is **Multimodal**.

2. Mode in Continuous Series

The modal class is identified by inspection (or through grouping tables when frequencies fluctuate irregularly). The exact mode is interpolated using:

$$Z = L + \left(\frac{f1 - f0}{2f1 - f0 - f2} \right) \times c$$

Where: L = Lower boundary limit of the modal class;
f_1 = Frequency of the modal class;
f_0 = Frequency of the pre-modal class (immediately preceding);
f_2 = Frequency of the post-modal class (immediately succeeding);
c = Class width of the modal class.

6.2 The Empirical Relationship among Mean, Median, and Mode

In a perfectly symmetrical, bell-shaped distribution, the mean, median, and mode coincide at identical values: $\bar{X} = M = Z$. However, in moderately asymmetrical (skewed) frequency distributions, Karl Pearson established the robust **Empirical Relationship**:

$$\text{Mode} = 3 \cdot \text{Median} - 2 \cdot \text{Mean} \implies Z = 3M - 2\bar{X}$$

$$\bar{X} - Z = 3(\bar{X} - M)$$

Managerial Utility: When a distribution has ill-defined modal boundaries, open-ended tails, or bimodal tendencies, the mode is reliably estimated using the empirical relationship from the calculated mean and median.

Chapter 7: Specialized Mathematical Averages: Geometric Mean and Harmonic Mean

7.1 The Geometric Mean (GM)

The **Geometric Mean** of a series of N positive observations is defined as the N^{th} root of the product of all values:

$$GM = \sqrt[N]{X_1 \times X_2 \times \cdots \times X_N} = (X_1 \cdot X_2 \dots X_N)^{\frac{1}{N}}$$

$$\log(GM) = \frac{\sum \log x}{N} \implies GM = \text{Antilog}\left(\frac{\sum \log x}{N}\right)$$

$$\text{Continuous / Frequency Series: } GM = \text{Antilog}\left(\frac{\sum (f \log m)}{N}\right)$$

Appropriate Business Applications of GM:

- Calculating compound average growth rates (CAGR) of sales, revenues, and population over chronological time.
- Constructing Index Numbers (Fisher's Ideal Index relies upon GM).
- Averaging ratios, percentage expansions, and multiplicative factors.
- *Limitation:* Cannot be calculated if any observation is zero or negative.

7.2 The Harmonic Mean (HM)

The **Harmonic Mean** is defined as the reciprocal of the arithmetic mean of the reciprocals of the individual values:

$$HM = \frac{N}{\sum\left(\frac{1}{X}\right)} = \frac{1}{\frac{1}{X_1} + \frac{1}{X_2} + \cdots + \frac{1}{X_N}}$$

$$\text{Continuous / Frequency Series: } HM = \frac{N}{\sum\left(\frac{f}{m}\right)}$$

Appropriate Business Applications of HM:

- Averaging rates, speeds, and price ratios where the variable under study is expressed per constant unit of another quantity (e.g., speed in kilometers per hour when travel distances are equal; purchasing power of the rupee when spending amounts are fixed).

- *Limitation:* Gives disproportionate weight to small values; cannot be computed if any observation is zero.

7.3 Relative Magnitudes of Averages

For any series of positive, unequal numerical observations, the three mathematical averages maintain the immutable inequality relationship:

> $$\text{Arithmetic Mean } (AM) \geq \text{Geometric Mean } (GM) \geq \text{Harmonic Mean } (HM)$$
>
> $$AM = GM = HM \quad \text{if and only if all individual observations are identical}$$
>
> $$\text{For two positive numbers } a \text{ and } b: \quad GM = \sqrt{AM \times HM}$$

Chapter 8: Comprehensive Worked Practical Problems with Step-by-Step Solutions

8.1 Worked Problem 1: Continuous Grouped Series Mean, Median, and Mode

The monthly overtime hours logged by 100 technicians in an automated manufacturing facility are distributed across class intervals as follows:

Overtime Hours Class	Number of Technicians (f)	Mid-Value (m)	Step Deviation d' = (m - 35) / 10	f · d'	Cumulative Frequency (cf)
10–20	12	15	-2	-24	12
20–30	18	25	-1	-18	30
30–40	35	35 (A)	0	0	65
40–50	20	45	+1	+20	85
50–60	15	55	+2	+30	100
Total	**N = 100**			**Sum fd' = +8**	

Step-by-Step Mathematical Computation:

1. Calculation of Arithmetic Mean:

> $$\bar{X} = A + \left(\frac{\sum (fd')}{N} \times c \right) = 35 + \left(\frac{8}{100} \times 10 \right) = 35 + 0.8 = 35.80 \text{ hours}$$

2. Calculation of Median:

$$\text{Position} = \frac{N}{2} = \frac{100}{2} = 50^{\text{th}} \text{ item}$$

The 50th item falls in the **30–40** class interval ($cf = 65$).

$$L = 30, \quad cf = 30, \quad f = 35, \quad c = 10$$

$$M = L + \left(\frac{\frac{N}{2} - cf}{f}\right) \times c = 30 + \left(\frac{50-30}{35}\right) \times 10 = 30 + \left(\frac{20}{35} \times 10\right) = 30 + 5.71 = 35.71 \text{ hours}$$

3. Calculation of Mode:

The modal class is **30–40** with peak frequency $f_1 = 35$.

$$L = 30, \quad f_1 = 35, \quad f_0 = 18, \quad f_2 = 20, \quad c = 10$$

$$Z = L + \left(\frac{f1 - f0}{2f1 - f0 - f2}\right) \times c = 30 + \left(\frac{35-18}{2(35)-18-20}\right) \times 10$$

$$Z = 30 + \left(\frac{17}{70-38}\right) \times 10 = 30 + \left(\frac{17}{32} \times 10\right) = 30 + 5.31 = 35.31 \text{ hours}$$

Verification via Empirical Formula:
$Z = 3M - 2\bar{X} = 3(35.71) - 2(35.80) = 107.13 - 71.60 = 35.53$ hours. Closely corroborates the direct interpolation value.

8.2 Worked Problem 2: Combined Arithmetic Mean Analysis

A manufacturing conglomerate operates three distinct production plants with operational wage parameters as follows:

- **Plant A:** $N_1 = 300$ factory workers; Mean Monthly Wage $\bar{X}_1 = \text{Rs. } 24,000$.
- **Plant B:** $N_2 = 200$ factory workers; Mean Monthly Wage $\bar{X}_2 = \text{Rs. } 27,000$.
- **Plant C:** $N_3 = 500$ factory workers; Mean Monthly Wage $\bar{X}_3 = \text{Rs. } 21,600$.

Required: Compute the overall combined mean monthly wage for the entire conglomerate workforce.

Step-by-Step Resolution:

$$\bar{X}_{123} = \frac{N_1X_1 + N_2X_2 + N_3X_3}{N_1 + N_2 + N_3}$$

$$\sum N = 300 + 200 + 500 = 1,000 \text{ total workers}$$

$$N_1\bar{X}_1 = 300 \times 24,000 = \text{Rs. } 72,00,000$$

$$N_2\bar{X}_2 = 200 \times 27,000 = \text{Rs. } 54,00,000$$

$$N_3\bar{X}_3 = 500 \times 21,600 = \text{Rs. } 1,08,00,000$$

$$\text{Grand Total Payroll} = 72,00,000 + 54,00,000 + 1,08,00,000 = \text{Rs. } 2,34,00,000$$

$$\bar{X}_{123} = \frac{2,34,00,000}{1,000} = \text{Rs. } 23,400.00$$

Managerial Insight: The combined mean of Rs. 23,400 accurately weights each plant by its workforce scale, pulling the conglomerate mean toward Plant C's dominant 500-worker base.

8.3 Worked Problem 3: Harmonic Mean for Average Speed and Rate

An executive logistics transport van delivers urgent retail supplies between a central warehouse and three retail distribution hubs:

- Leg 1: Travels 60 kilometers at a steady speed of 30 km/h.
- Leg 2: Travels 60 kilometers at a steady speed of 40 km/h.
- Leg 3: Travels 60 kilometers at a steady speed of 60 km/h.

Required: Compute the authentic average speed for the complete 180 km journey.

Resolution: Because the distances traveled across all three segments are identical (constant distance = 60 km), the correct average speed is governed by the **Harmonic Mean**:

$$HM = \frac{1}{\frac{1}{X_1} + \frac{1}{X_2} + \frac{1}{X_3}} = \frac{1}{\frac{1}{30} + \frac{1}{40} + \frac{1}{60}}$$

Common Denominator (120): $\frac{4}{120} + \frac{3}{120} + \frac{2}{120} = \frac{9}{120} = \frac{3}{40}$

$$HM = \frac{3}{\frac{3}{40}} = 3 \times \frac{40}{3} = 40.00 \text{ km/h}$$

Verification via Total Physical Time:
Time Leg 1 = 60 / 30 = 2.0 hours.
Time Leg 2 = 60 / 40 = 1.5 hours.
Time Leg 3 = 60 / 60 = 1.0 hour.
Total Time = 2.0 + 1.5 + 1.0 = 4.5 hours.
Average Speed = Total Distance / Total Time = 180 / 4.5 = **40.00 km/h**.
(Note: The simple arithmetic mean of speeds would yield (30 + 40 + 60) / 3 = 43.33 km/h, which is physically false).