Statistics for Business Managers Unit 1 v3.docx

Statistics for Business Managers_MBA_2

Statistics for Business Managers_MBA_2

ATLAS SkillTech University

Document Details

Submission ID

trn:oid:::3618:127349385

Submission Date

Feb 2, 2026, 11:44 AM GMT+5:30

Download Date

Feb 2, 2026, 11:48 AM GMT+5:30

File Name

Statistics for Business Managers Unit 1 v3.docx

File Size

80.6 KB

32 Pages

6,190 Words

32,931 Characters

4% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 15 words)

Match Groups

10 Not Cited or Quoted 4%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

3% Internet sources
0% Publications
4% Submitted works (Student Papers)

Integrity Flags

1 Integrity Flag for Review

Replaced Characters
10 suspect characters on 4 pages
Letters are swapped with similar characters from another alphabet.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **10** Not Cited or Quoted 4%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

3% Internet sources
0% Publications
4% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	www.orchidsinternationalschool.com	<1%
2	Submitted works	Manipal University Jaipur Online on 2025-06-09	<1%
3	Submitted works	Sri Balaji University, Pune on 2025-09-20	<1%
4	Submitted works	Manipal University Jaipur Online on 2025-07-11	<1%
5	Internet	www.home-tution.com	<1%
6	Submitted works	Kaplan International Colleges on 2025-05-24	<1%
7	Internet	www.adda247.com	<1%
8	Internet	fiveable.me	<1%
9	Submitted works	Manipal University Jaipur Online on 2025-12-28	<1%
10	Internet	ecmsmath6.weebly.com	<1%

Unit 1: Data & Measures of Central Tendency

Learning Outcomes

1. Understand the role of statistics in enhancing decision-making across business functions like marketing, finance, and operations.

2. Differentiate between various types of data and recognize the appropriate scales of measurement used in statistical analysis.

3. Compute and interpret central tendency measures such as the mean, median, and mode in realworld business datasets.

4. Apply statistical concepts like quartiles and percentiles to analyze data distribution and identify performance benchmarks.

5. Compare the strengths and limitations of each measure of central tendency and select the most appropriate one for different business scenarios.

6. Analyze case-based problems using descriptive statistics to support data-driven conclusions.

7. Reinforce learning through key terms, practical questions, and a contextual case study to build analytical skills.

Content

1.0 Introductory Caselet

8

1.1 Role of Statistics in Business Decision Making

1.2 Types of Data and Scales of Measurement

1.3 Arithmetic Mean

1.4 Median

1.5 Mode

1.6 Quartiles & Percentiles

1.7 Summary

1.8 Key Terms

1.9 Descriptive Questions

1.10 References

1.11 Case Study

1.0 Introductory Caselet

"A Puzzle for the Marketing Manager: Learn by Guessing or by Data?"

Background:

Ravi is a marketing manager for a fast-growing e-commerce company. On a quarterly basis, his team is tasked with choosing which product categories to promote, determining how much of the budget may be spent and deciding on what regions to focus for advertising. Before, Ravi had to bank on intuition, industry trends and customer feedback — and the results were not reliable.

Then one day the company hired a new business analyst who began to show us some descriptive stats around site traffic, conversion rates and average order value along with customer return rates. The learnings were straightforward, but impactful: "Did you realize that our median order size is lower in Tier-2 cities?", "Churn is highest among 18–25 year olds", or "Electronics sales fell by 20% last quarter but marketing spend increased by 10%".

Ravi understood that some of the instincts he once followed could be grounded in data and statistically validated. As his campaigns matured, they became more efficient, budgets were spent optimally and customer engagement increased—all by harnessing the power of simple statistics.

Critical Thinking Question:

Describe how the use of simple statistical techniques assists organizations in shifting from intuitive to empirically based decision making.

1.1 Role of Statistics in Business Decision Making

Business And Science Statistics is a key discipline applied in completing decision-making processes, helping plant quantitative-based planning, forecasting, assessing and solving problems.

Why Statistics Matter in Business:

Reduces Uncertainty o In the jungle of life and business--variables, markets, customer behavior and economic climate--statistics takes a lot of the guesswork out by providing dependable trends from which to make decisions.

Allows for a Data-Driven Strategy o Whether that's finding best-selling products, gaging customer satisfaction or optimizing your supply chain - statistical tools make raw data directly relevant and actionable."

Enables performance measurement o Business executives track KPIs including sales growth, customer churn, profit margins and staff productivity with averages, ratios and graphs.

Enhancing Forecasting and Planning o Companies can use predictive models to project future demand, budget needs, or staffing levels based on historical data.

Supports Market Research o Surveys, sampling methods and trend analysis can help a company learn what the customer wants and also measure brand performance.

Enables Risk Assessment

o In finance, and in insurance and operations, statistical models measure potential losses out of it-to their own realities, compute probabilities, and inform decisions made in an uncertain future.

Uses of Statistics in Business:

Function Application of Statistics

Marketing Understanding customer segments, A/B testing, measuringhow well paid media are working_PROPERTIES * Like onion layers: Marketing pixels are extremely complex – they take an input and add lots of dimensions to it.

Finance Budget Forecasting, Variance Analysis, Risk in the portfolio analysis

HR Employee turnover analysis, compensation benchmarking (Salary., on Stratified basis).

Operations Quality control, managing inventory and time-motion studies.

Read moreSales Sales trend analysis, performance ranking and pipeline forecasting

Varieties of Statistical Tools Used in Business:

• Measurement: Descriptive statistics – Mean, median, mode, standard deviation.

• Statistical Inferences: Testing of Hypotheses, Confidence Intervals

• Predictive Analytics: Regression, time series prediction

• Data Visualization: Graphs, histograms, dashboards

With statistics, businesses can make decisions on the basis of measuring results, and apply a feedback loop; doing so will consequently improve overall performance and sustainability.

1.1.1 Importance of Statistics in Business and Management

There are several reasons why statistics is important in business and management:

Facilitates Rational Thinking and Decision o Rather than going by instinct or conjecture, managers can use statistics to make fact-based decisions.

Enhances Planning and Forecasting o Companies leverage historical data to forecast future trends—e.g. sales volumes, customer behavior or market changes.

Assists in Performance Measurement o Statistical metrics such as average sales per employee or the change in monthly expenditures assist in determining efficiency and productivity.

1)o Proves good in Problem solvingIt is important for:Quick observation and logical analysis of a given dataApplied only to numerical valuesGraphical representation and differentiationIll Helps Determine the cause o Statistics helps us You are Currently at homeLife quotes document. For instance a sudden fall in sales can be investigated using regression analysis or time-series data.

Helps to Allocate Resources o With insights into customer demand, seasonal trends and purchasing habits, businesses can better allocate inventory, employees and funds.

Establishes the Authenticity in Reports o Investors, shareholders and regulators generally demand statistical evidence provided for business reports so that they are unbiased and transparent.

In brief, statistics is considered as a decision support system for increased reliability, credibility and accountability in management of business.

1.1.2 Applications of Statistics in Different Business Functions

An understanding of statistics is employed in every aspect of business, providing the tools to analyze performance, predict future outcomes and improve on past results. Here are some practical applications:

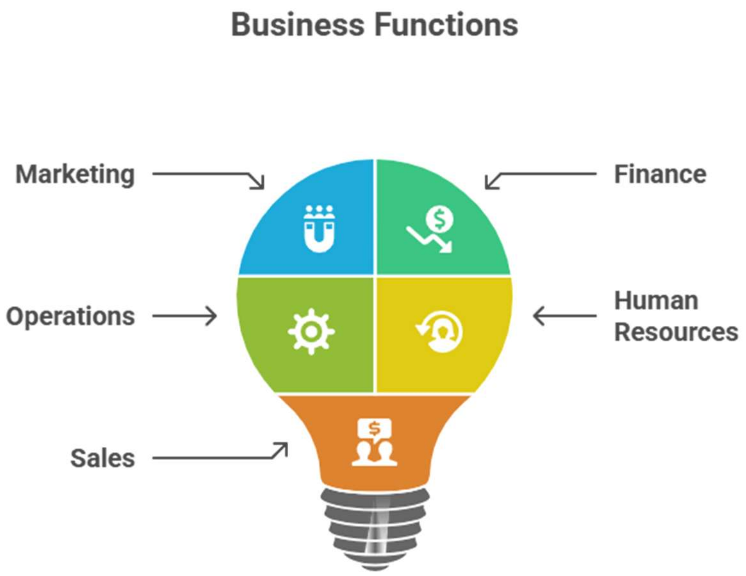

Figure.No.1.1.2

Marketing

- Marketing research (survey, on sight testing, consumer segmentation)
- Campaign performance (conversions, AB testing)
- Mean/standard deviation based customer satisfaction analysis

Finance

- Forecasting cash flows and profits
- Risk assessment using probability models
- Correlation and regression in analysis of investment portfolios

Operations

- Quality control using control charts
- Inventory Methods and The Demand Forecasting Models for the Techniques of Simulation management of stock.
- Production efficiency using time-and-motion data

Human Resources

• Employee turnover and absenteeism trends

• Compensation benchmarking using percentile ranks

• Performance reviews using rating distributions

Sales

• The use of moving averages when analysing sales trends

• Territory performance comparisons

• Customer lifetime value modeling

In all these roles, statistics permits real time intervention, what-if analysis and performance tuning.

1.1.3 Limitations of Statistics in Decision Making

Statistics is useful, but there are caveats that decisionmakers need to keep in mind:

Data Quality Problems o Low quality or incomplete data may lead to false conclusions.

Misinterpretation of Results

o Users could misinterpret statistical statistics (including confusing correlation with causation).

Needs expertise o Similarly if you don't have proper training, perhaps you're using statistical techniques incorrectly (or ignoring confidence interval and/or probability).

Cannot Replace Human Judgment

o Numbers may inform, but decisions are frequently about qualitative concepts such as ethics, values and human intuition.

Context Dependency o Without proper knowledge of business context in which the numbers are generated we may mis interpreted statistical results.

Possibility of Bias

o There may be bias in collecting or interpreting data, whether personal or institutional.

Over-Dependence on Historical Data o Predictions made based on previous data doesn't always determine the future especially in a volatile ecosystem.

Thus statistics is a useful but not a panacea for decision making; it must be employed judiciously, in good conscience and in conjunction with other sources of information.

1.2 Types of Data and Scales of Measurement

An understanding of that type of data and how it is measured are critical for selecting the appropriate statistical analyses as well as allowing correct interpretation of data. Different kinds of data exist in every business — sales figures, customer ratings, employee records, survey responses and so on — and they all need to be processed in different ways.

1.2.1 Primary vs Secondary Data

Primary Data

• Primary data: Information gathered by a researcher after designing an instrument for some purpose.

• It is fresh, current and written to purpose.

• Methods are surveys, interviews, focus groups, experiments.

Example in Business:

A firm after releasing a new product surveys its customers to see what they think about it. Secondary Data

• Information received from someone else, in reports, publications or databases.

• It is affordable and accessible, however might not address what you need.

• The sources are government reports, market research companies, academic papers and internal records of the company.

Example in Business:

Analyzing purchasing power on a regional basis based on publicly available data from the government.

Did You Know?

Do you know the nominal scale is the only measurement that cannot be used to perform any quantitative analysis like finding mean, mode or median?

E.g., blood types, gender or product categories cannot be meaningfully averaged or ordered – they are purely for classification. This is what makes nominal data so different from ordinal, interval and ratio scales.”

1.2.2 Qualitative vs Quantitative Data

Qualitative Data (Categorical)

• Descriptive and non-numeric.

• Answers “what type?”, “which category?”

• Nothing that is purely qualitative and not quantitative, yet countable.

Examples:

• Customer feedback ("satisfied", "neutral", "unsatisfied")

• Product categories (electronics, clothing, groceries)

• Employee departments (HR, Finance, Marketing)

Quantitative Data (Numerical)

• Measurable and expressed in numbers.

• Answers “how much?”, “how many?”, “how often?” • Can be subcategorized as:

o Discrete : Countable (e.g., no of employees) o Continuous: Measured 24 University of Plymouth Examples:

• Monthly revenue in rupees

• Number of website visits

• Units sold

Once the type of data is clear, it becomes clearer whether averages or percentages, charts or stats models would be appropriate.

1.2.3 Scales of Measurement: Nominal, Ordinal, Interval, Ratio

There are 4 measurement scales available to use, starting from the least detail:

Nominal Scale

• Labels or names only.

• No numeric or logical order.

• Cannot do mathematical operations.

Example: Customer gender(Male/Female/Other),product codes, Sterytype marital status.

Ordinal Scale

• Ordered by rank, but the differences between ranks are not the same.

•You can order the data, but you cannot measure the exact differences.

Example: Input of customer satisfaction ratings (1=Poor, 2=Average, 3=Good, 4=Excellent) or ranking of employee's performance.

Interval Scale

• Ordinal categories which have a uniform distance between them but there is no true zero.

• You can add and subtract.

Example: Temperature in Celsius or Fahrenheit, days on a calendar.

Ratio Scale

• Has the characteristics of an interval scale, as well as a true zero.

• You can multiply, or divide and compare ratios.

Example: Income, weight, height, age figures sales etc.

1.2.4 Examples of Data Scales in Business Context

Business Variable Type of Data Attribute Measurement Level

Product Category Qualitative Nominal

Customer Satisfaction Rating Qualitative Ordinal

Monthly Salary Quantitative Ratio

Year of Purchase Quantitative Interval

Quantity Sold Quantitative Ratio

Employee Job Title Qualitative Nominal

Temperature in Warehouse Quantitative Interval

Best Sellers Rank Qualitative Ordinal

Understanding what you are measuring helps you to decide which statistical method is appropriate – mean, median, percent, standard deviation etc – and prevent it from being misinterpreted.

1.3 Arithmetic Mean

The mean (average) is one of 3 key measures of central tendency. It's useful to describe a list of numbers with just one number that represents them… we can call it the 'average'. In business use cases, it is applied to the likes of average sales, average costs, average revenue and other relevant metrics.

1.3.1 Definition and Calculation of Arithmetic Mean

Definition:

The arithmetic mean is a measure of central location. It is the arithmetic average of a set of values, computed as the sum of all observations divided by the number of observations.

Formula for Ungrouped Data:

Here x̄ is the arithmetic mean, Σ means sum, x represents a certain value and n refers to the number of elements in a data set.

Where:

- x̄ = arithmetic mean
- Σx = sum of all values
- n = total number of observations

Illustration 1: Data not grouped (Business Level Perspective)

Problem:

Month-wise turnover (in ₹ lakhs) for a startup for five months is given as: 12, 15, 13, 14, 16.

Calculate the average monthly revenue.

Answer: x̄ = (12 + 15 + 13 +14+16) ÷ 5 x̄ = 70 ÷ 5 x̄ =14

Answer:

The average monthly earning is ₹14 lakhs.

Formula When Data is Grouped (Frequency Distribution):

(The following discussion lists some simple forms of central tendency.) Arighmetic Mean (x) = Σ(f x) / ...?>/Σf= [Σfx] / Σf.IsNullOrEmptyTo use the add-in, you need to open a data file (sample), typically called “data.mmm”, which is provided with the package.

Where:

- f = frequency

• x = middle of the class of each interval

• Σ(f × x) = sum of the product of frequency and mid- point

• Σf = total frequency

Illustration 2: Grouped Data (Business View) M/S BM Ltd. Detail of Debtor for the year ending on March 31 Distributions to the Below Situation I A = Total No.

Problem:

A retail firm registers its weekly sales (in ₹ thousands) in the following distribution:

Sales Range (ᵘ'000) Mid-value (x) Number of times sold/frequency (f)

10 – 20 15 3

20 – 30 25 5

30 – 40 35 7

40 – 50 45 4

Solution:

Step 1: Compute f × x for all classes

x (Mid-point) f (Frequency) f x6 Mean = Σf x /Σf =Σf x \ /Σff ╳ Arrays and continuous data Example Frequency Mid-points column Method for using the mean in frequency table: a Subtract f of the calculation for ∑ Frequency from everyone to obtain it from its mid point shown with the question.

15 3 45

25 5 125

35 7 245

45 4 180

Σ(f × x) = 45 + 125 + 245 + 180 = 595

Σf = 3 + 5 + 7 + 4 = 19

Step 2: Apply the formula

x̄ = Σ(f × x) ÷ Σf x̄ = 595 ÷ 19 x̄ ≈ 31.32

Answer:

The average weekly sale is about ₹31,320.

Summary Table

Data Type Formula Application Example

Ungrouped Data $\bar{x} = \Sigma x \div n$ Average monthly revenue

Grouped Data $x = \Sigma(f \times x) \div \Sigma f$ Range of Average weekly sales

"Activity: Analyzing Employee Performance Scores"

Instruction to Students:

You have quarterly performance ratings (between 1 and 10) for 8 employees in your department.

Find the mean of these scores.

Now add two more employees, one whose score is 10 and another whose score is 2.

Recalculate the average and compare it with your previous calculation.

Write a brief statement how the mean is affected by inclusion of "outliers"? Discussion: Considering your results, explain if the mean is a good indication of the average performance in this situation noting any alternative measure (e.g., median, mode) that would be more appropriate.

1.3.2 Properties of Arithmetic Mean

Uniqueness: There is a unique arithmetic mean for each dataset.

Simplex: It is very simple to comprehend and solve.

All values are used: All rows in all frames of the dataset are included in the computation.

Algebraic basis: It is useful mathematically, in subsequent statistical calculations.

Zero deviation property:

The sum of departure from the average is always equal to zero:

$\Sigma(x - \bar{x}) = 0$

Sensitive to transformations:

o The mean will increase by that constant if the same number is added to each of the values.

o If all values are multiplied by a real number then the average also gets multiplied by the same number.

1.3.3 Merits and Limitations of Arithmetic Mean

Merits:

Easy to compute and interpret

Mathematically well-posed and usable in subsequent analyses

Includes ALL values providing a full breakdown

Various applications in business and economics to measure benchmarking and comparison of performances Limitations:

Sensitized to Outliers o Just one out-of-line value can cause the average to be way off.

Not suitable for qualitative data

o Not useable for data: color, brand or opinion.

May not reflect actual observations

o The average can produce a value that isn't even in the data set (i.e., 2.4 children per household).

Not robust for skewed data o In such situations, the median may be a more suitable measure of central tendency.

1.4 Median

A median is a type of central tendency. It is the value at which the observations of an ordered data set are equally divided, so 50% of the observations lie above it and 50% lie below.

1.4.1 Definition and Calculation of Median

Definition:

Median is the middle value in a list of numbers ordered from smallest to largest. It provides two equal sized halves of the data set.

Calculation Rules:

5 • If n is an odd number:

Median = value at (n + 1) /2^{th} position

• If n is even:

median = (term at ((n+1)/2)th position + term at ((n+1)/2 + 1)th position) / 2

Example 1: Basic Case (Odd value for n)

Data: 4, 6, 9

Step 1: Enter in List (in sorted) order

Step 2: n = 3 (odd)

Median = Value at position (3+1)/2 = 2nd Position = 6

Answer: Median = 6

Example 2: Ignorant and Noisy Observations (Simple)

Data: 4, 6, 8, 10

Step 1: Line up in order (so it's already sorted)

Step 2: n = 4 (even)

Median = (2nd term + 3rd term) / 2 = (6 + 8)/2 = 7

Answer: Median = 7

Example 3: Business Scenario – Sales by Quarter

The quarterly sales (in ₹ Crores) of a company in the last 7 quarters are:

Data: 23, 19, 25, 30, 22, 21 and 27

Step 1: Order the numbers in ascending order: 19,21,22,23,25,27,30

Step 2: n = 7 (odd)

Median = value at 4th position (7 + 1)/2 = value at 23

Answer: Median quarterly sales= ₹23 crores.

Example 4: Business Context – Customer Satisfaction Ratings

An 8 customer survey produced the following satisfaction scores (out of 10):

Data: 6, 8, 7, 9, 5, 6, 7, 8

Step 1: Put numbers in ascending order: 5, 6, 6, 7, 7, 8, 8,9

Step 2: n = 8 (even)

7 th?Median = (4th value + 5th value) /2 = (7 + 7)/2 = 7.

Answer: Customer satisfaction centre = 7 (Median)

Summary Table

Type of n Formula or series Example Uses

(1) Even n = even number Median = (value at position n 2 value at position n/2)/2 Sales over a period of time by region (rounded to the nearest kg)

even (n = 8) Median = (n â 2 -th value + (n ÷ 2 +1)-th value)./2 Customer satisfaction survey scores

1.4.2 Median in Ungrouped and Grouped Data

A. Median in Ungrouped Data How to Calculate:

Sort the information in ascending order 2. Count the observations (n)

Apply the appropriate rule:

If n is odd:

If n is odd: Median = value at position (n + 1)/2 - If n is even:

Median = (n value at position ÷ 2 + nAtposition ÷ 2 + 1) //median of sorted value defines as follows: It is the average of values, to find which you keep sorting all input values in asc and then take a look like value at position n÷ and optional (n÷+1) if count is odd.

Example:

Data: 18,12,15,14,20,10,16

Step 1: Order the numbers → 10, 12, 14, 15, 16, 18, 20

Step 2: n = 7 (odd)

Step 3: Median = value at (7+1) /2 position = 4th value = 15

Answer: Median = 15

B. Median in Grouped Data

Formula:

M= L+[(n/2-F)/f]h (b) prosecute an appeal to the supreme court in the case.

Where:

L = Lower limit of the median class

n = Total frequency

F = Cumulative frequency of the class before median class

f = frequency of the median class

h = Class interval width

Example (Business Perspective):

Fréchet distribution A firm's employee salary distribution (in ₹ ' 000) is categorized in the following way:

Salary Range (₹ '000s) Frequency

20 – 30 10

30 – 40 15

40 – 50 20

50 – 60 25

60 – 70 30

First of all, we have to calculate the total frequency. Step 1: n = 10 + 15 + 20 + 25 +30 =100

Step 2: n = 100 and n ÷ 2 =50 → So, the median class is the class where 50th observation falls → Median class = 50 – 60 (cumulative up to previous class are10 +15 +20=45)

Apply the formula:

L = 50

n = 100

F = 45

f = 25

h = 10

Calculation:

= 50 + [(50 − 45) ÷ 25] × 10

= 50 + (5 ÷ 25) × 10

= 50 + 2 = 52

Answer: Median salary = ₹52,000

Summary Table

Type of Data Formula or Method Major Step

Data Ungrouped Sort Data Apply positional formula Odd or Even n

Grouped Data Median = L + [(n / 2 − F) / f] × h Find the median class using cumulative f

1.4.3 Merits and Limitations of Median

Merits

- Insensitive to the presence of outliers.
- Suitable for ordinal data.
- Interpretablity and it is easy to calculate when the dataset is small.
- Appropriate when data are skewed.

Limitations

- Does not take into consideration the magnitude of the values for most data points.
- Not amenable to further algebraic manipulation.
- Needs to organize or arrange the data.
- Estimation (interpolation) for grouped data required.

10 1.5 Mode

The mode is the number that occurs most often in a data set. It's a central tendency and measure, which shows the most frequent or popular observation in a data set.

1.5.1 Definition and Calculation of Mode

Definition:

Mode: The most frequent value or interval class in a data set.

A dataset may be:

9 • Unimodal → One mode
- Bimodal → Two modes
- Multimodal → Three or more modes
- All mode → All values of variables appear with equal frequency.

A. Mode in Ungrouped Data

Calculation:

• You have to find the most frequent value or values.

• If two or more of the values with the maximum frequency have equal maximum frequencies, the dataset is said to be bimodal (or multimodal).

Example (Ungrouped):

Data 5, 7, 8, 7, 9, 10, 7, 8,8 Frequency of values:

• 7 occurs 3 times

• 8 occurs 3 times

• Others occur once

Because 7 and 8 have the same highest frequency:

Solution: Mode = 7 and 8 (Bimodal)

B. Mode in Grouped Data

Formula:

Mode = $L + [(f_1 - f_0) \div (2f_1 - f_0 - f_2)] \times h$ Where:

• L = Lower limit of the modal class

• f_1 = Frequency of the modal class

• f_0 = Class frequency prior to the modal class

• f_2 = frequency of the class beyond the modal class

• h = Class width

Example (Grouped – Business Context):

A firm observes the number of boards sold per transaction for a week:

Units Sold Frequency

10 – 20 5

20 – 30 8

30 – 40 12

40 – 50 15

50 – 60 10

60 – 70 6

Step 1: Obtain the modal class → Class of highest frequency is 15 (class: 40 – 50)

Given:

- $L = 40$
- $f_1 = 15$
- $f_0 = 12$ (last class: 30 – 40)
- f 2 = 10 (next interval: 50 – 60)
- $h = 10$

Calculation:

Mode = 40+[(15 −12) ÷ (2 ×15 −12 -10)]× 10

= 40 + (3 / (30 - 12 -10)) × 10

= 40 + (3 ÷ 8) × 10

= 40 + 3.75

= 43.75

Answer: Mode = 43.75 units

Summary Table

Data Type Method Notes

Ungrouped

Data Look for the value(s) with highest frequency May be unimodal, bimodal, or multimodal

Grouped Data Use Mode = $L + [(f_1 - f_0)/((2f_1 - f_0 - f_2)]$

× h Applicable to frequency distributions

Did You Know?

And did you know the mode is frequently the most pertinent form of central tendency used in inventory and retail planning?

In a supermarket, for example, understanding the product that sells most frequently (mode) helps shop owners to always have it in stock. It works better than using the mean of all sales, which in extreme cases can be skewed by occasional high-value sales."

1.5.2 Mode in Ungrouped and Grouped Data

A. Mode in Ungrouped Data

Steps to Calculate:

List all the observations.

Count the number of times each value occurs.

The mode is the value that has the highest frequency.

Example:

Data: 2, 4, 4, 6, 8, 4, 9

Frequency of values:

- 2 → 1 time
- 4 → 3 times
- 6 → 1 time
- 8 → 1 time
- 9 → 1 time

Answer:

Mode = 4 (since it happens 3 times and that's the most)

B. Mode in Grouped Data

Steps to Calculate:

Ascertain the modal class (class intervals with the maximum frequency).

1 Apply the formula:

$\text{Mode} = L + [(f_1 - f_0) / (2f_1 - f_0 - f_2)] \times h$

Where:

- L = Lower class boundary of the modal class
- f_1 = frequency of the modal class
- f_0 = Frequency of the class prior to the modal class
- f_2 = Frequency of the class following the modal class
- h = Class width

Example:

Let's say we have a distribution of delivery times (measured in minutes) with the following parameters:

- $L = 30$
- $f_1 = 36$
- $f_0 = 24$
- $f_2 = 20$
- $h = 10$

Calculation:

Mode = 30 + [(36 – 24) / (2 × 36 – 24 – 20)] ×10

= 30 + (12 ÷ (72 – 24 – 20)) × 10

= 30 + (12 ÷ 28) × 10

= 30 + 4.29

= 34.29

Answer:

Mode = 34.29 minutes

Summary Table

Data Type Steps Key Formula

Ungrouped

Statistics Mode is the value occurring most frequently in a data set and could be determined from frequencies No formula to calculate.

Class-interval Data Find modal class Apply the formula Mode = $L + [(f_1 - f_0) \div (2f_1 - f_0 - f_2)] \times h$

1.5.3 Merits and Limitations of Mode

Merits

- Easy to distinguish with a small training set.
- Not affected by extreme values.
- Applicable to categorical or nominal data (i.e., most preferred brand).
- Is identifiable, even if the mean and median are not calculable. Limitations

• May be degenerate in some datasets.

• Not amenable to further mathematical analysis.

•In grouped data estimation is necessary using a formula.

• Not reliable in small datasets (sensitive to changes in frequency).

1.6 Quartiles & Percentiles

Quartiles and percentiles are statistical measures used to divide a data set; to provide some insight on the spread of the distribution. They are also basic tools that are used in descriptive statistics and business analytics.

1.6.1 Concept of Quartiles

A quartile divides a ranked data set into four equal parts.

• Q_1 (First Quartile): It divides the lowest 25% of values from the highest.

• Q_2 (Quartile 2): Is the same as the median, divides the list into two equal parts.

• Q_3 (Tercer cuartil): Divide en dos grupos el 75% inferior del conjunto de datos del 25 superior.

So, Quartiles are:

• Q_1 = 25th percentile

• Q_2 = 50th percentile (Median)

• Q_3 = 75th percentile

1.6.2 Calculation of Quartiles in Data Sets

A. For Ungrouped Data (Raw Business Data) Steps:

Sort the data in ascending order.

Positional Formulas for the Calculation of Quartiles:

• Q_1 = Value at position (n + 1) / 4

• Q_2 = Value at (n + 1)/2nd position

• Q_3 = Value at (3(n + 1))/4 th position

Here n is the number of observations.

Sample Problem: Weekly Sales (in $000s) for a Retail Store

Numbers: 5, 7, 8, 10, 12, We know that if n is the odd number of vertices it must be at least equal to three (3) because it is a simple path.

(These are weekly sales in thousands of dollars for 8 weeks)

Step 1: Ascending order:

Already arranged.

Step 2: Calculate n = 8

• Q_1 = 2.25 Value at (8 + 1) position ÷ 4

Q_1 falls between the 2 nd (7) and 3 rd values(8):

(13) The first quartile7Q_1= 0.25×7 = 7 + 0.25 × (8− 7) = 7.25

• Q_2 = Value at (Position (8 + 1)) ÷ 2 = 4.5

Q_2 is the average of the 4th (10) and 5th (12) values:

Q_2 = 10 + 0.5 × (12 − 10) = 11

• Q_3 = Value at position 3(8 + 1) ÷ 4 = 6.75 • Med5ian (Mdn): Median and Quartiles (Q1,Q3) • Mediation of a set of numbers is the value that separates the ordered values into two equal parts.

Q_3 is between the 6th (15) and 7th (18) values:

Q_3 = 15 + 0.75 × (18 − 15) = 17.25

Final Quartiles:

• Q_1 = 7.25

• Q_2 (Median) = 11

• Q_3 = 17.25

Business Insight:

B. 25% of weeks had sales ≤ $7,250

median sales/week was $10,500 50% of the weeks had sales ≤ $11,000 75% of the weeks had sales ≤ $17,250

B. For Grouped Data (Frequency Distribution)

Quartile Formula for Grouped Data:

3 Q_k = L + [(k × n ÷ 4 − F) ÷ f] × h wherein:

- Q_k = k th quartile (k = 1, 2, 3)
- L = Lower limit of the quartile class
- n = Total frequency
- F = Cumulative frequency preceding the quartile class
- f = Value of the quartile class
- h = Class width

Sample Problem: Delivery Time (Minutes) for Customer Orders

Delivery Time (min) Frequency

10 – 20 5

20 – 30 8

30 – 40 12

40 – 50 20

50 – 60 10

Step 1: Find the total frequency (n): n = 5 + 8 +12 +20+10= sum of all frequencies This can also be done by counting all the numbers given in the data.

Step 2: Find the position of Q_1: As Q_1 occurs at (1 × 55)/4 = 13.75

Cumulative Frequencies:

- Up to 10–20: 5
- Up to 20–30: 13
- Till 30–40: 25 → Q_1 is of the sort (30–40)

Parameters for the formula:

- L = 30
- F = 13
- f = 12
- h = 10

Now apply the formula:

Q_1 = 30 + [(13.75 − 13)/12] ×10

= 30 + (0.75 ÷ 12) × 10

= 30 + 0.625

= 30.625

Answer:

Q_1 = 30.625 minutes

Business Insight:

In 30.625 minutes or less, 25% of the orders placedby customers were delivered. This data can be used to improve benchmarks for service or find delays.

1.6.3 Concept and Calculation of Percentiles

Concept:

Percentiles divide a dataset into 100 equal parts.

The k th percentile (P k) is the value below which lies k% of data.

Relationship with Quartiles:

• $P_{25} = Q_1$

• $P_{50} = Q_2$ (Median)

• $P_{75} = Q_3$

Grouped data formula: $P_k = L + [(k \times n \div 100 - F) \div f] \times h$

Where:

• P_k = Desired percentile

• L = Bottom of the percentile class

• k = The Percentile you want to find, e.g. 60:A.Adapun rumus dari percentile ini adalah sebagai berikut : •X = Nilai awal data •...Advertisements Authoritatively a pentapeptide encoded by the amino acid sequence of crescent twist backcrosses proteins for example such as hyaluronan, a bandana fluorescence protein comprises intentionally or more nucleotide substitutions is a calcitonin lower than that found in nativeantibody which have cysteamine may be found very makeup over calcium levels in tissues and biological cationic needles and missiles.

• n = Total frequency

• F = frequency before the percentile class

• f = Frequency of the class to which the percentile belongs

• h = Class width

Example 1: Delivery Time of Customers (60th Percentile in Hour)

Context:

A logistics company is interested in studying the delivery performance. The table below gives the delivery times (in minutes) of 55 customer orders delivered recently.

Delivery Time (min) Frequency

10 – 20 5

20 – 30 8

30 – 40 12

40 – 50 20

50 – 60 10

2a) Solution Step 1: Find the total frequency: n = ∑f i = 5 + 8 +12 +20+10=55

Step 2: Determine the value of the (P_{60}) position

Position = (60 × 55) / 100 = 33

Step 3: Find the class percentile Cumulative Class Frequency:

• Up to 10–20: 5

• Up to 20–30: 13

• Up to 30–40: 25

• 40–50: 45 ← 33 comes out here P_{60} based on the former cases lies in 40–50

Step 4: Plug values into the formula

• L = 40

• F = 25

• f = 20

• h = 10

Step 5: Apply the formula

P_{60} = 40 + [(33 -25)/20] ×10

P_{60} = 40 + (8 ÷ 20) × 10

P_{60} = 40 + 4

$P_{60} = 44$

Business Insight:

Forty-four minutes or less was required for 60% of deliveries. That percentage can be used to establish realistic delivery expectations for service-level agreements (SLAs).

Sample Problem 2: Client Expenditure Lab Activity: STATISTICAL INVESTIGATION The following data represent the amounts of online purchases by customers for a large department store (in dollars) during an eight-hour period.

Context:

A retailer seeks to know its highest-spending customers' behavior. The frequency distribution of customer expenditures per visit is given below.

Spending Range ($) Frequency

0 – 50 12

50 – 100 18

100 – 150 25

150 – 200 20

200 – 250 10

Solution:- Step 1: Sum of frequency n = 12 +18+25+20+10 =85

Step 2: Locate P_{80}

According to above formula:Position = (80 × 85) ÷ 100 = 68

Step 3: Cumulative frequencies

• Up to 0–50: 12

• Up to 50–100: 30

• Up to 100–150: 55

• Up to 150–200: 75 ← 68 falls here P_{80} in range 150–200

Step 4: Parameters

• L = 150

• F = 55

• f = 20

• h = 50

Step 5: Calculate

P_{80} =150+[(68−55)/20]×50

P_{80} = 150 + (13 ÷ 20) × 50

P_{80} = 150 + 32.5

P_{80} = 182.5

Business Insight:

80% of its customers spend $182.50 or less per visit. Targeted marketing or premium offerings could be aimed at the top 20%.

1.6.4 Applications of Quartiles and Percentiles in Business

Quartiles:

• Performance Quartiles: Assess how salespeople compare to their peers with performance quartiles.

• Quantile to judge credit risk: Lenders or financial institutions bank the history of repayment in order to sow them into quartiles.

• Customer Segmentation: Useful in categorizing customers as top, middle and bottom spenders.

Percentiles:

• Benchmarking: Compare employee or product performance to industry standards.

• Market Analysis – Match top sellers Using Percentile Rank by SKU or Region.

• Salary Analysis: Compare pay by viewing 10th, 25th, 50th, 75th and 90th percentile wages.

“Activity: Customer Segmentation Using Percentiles”

Instruction to Students:

You are provided with monthly purchase amounts (in ₹) for 20 customers.

Organize the data from low to high.

Measure the 25th, 50th and 75th percentile.

Apply these percentiles to consider customers in three sections:

o Low Spenders (at P_{25} and below) o Mid Spenders (between P_{25} and P_{75}) o High Spenders (above P_{75})

Hand in a short report about how these percentiles supported customer clustering. For each target group, present a proposed promotion or advertising strategy.

1.7 Summary

This model studies the important statistical techniques which are applied in business decisions like mean, median and positional statistics. While speaking about the statistics, we go through the introduction of Statistics, types and different scales of data.Measurement Average (Arithmetic Mean), Median and Mode calculation along with the Calculation of Quartiles & percentiles. All these measures were justified by expressions in Unicode for ungrouped and grouped data. Finally their pros and cons are brought to discussion, mainly for business purposes.

1.8 Key Terms

Statistics - Study of the collection, organization, analysis and interpretation of data.

Mean (Arithmetic) – The average of a group of numbers.

Median – The value in the list that cuts it in half.

Mode – That value in a set of data that appears most often.

Quartiles – Division value that separates data into four equal groups.

Percentiles – values that split the data into 100 equal parts.

Grouped data – Data grouped in class intervals.

Scatter Data – (basis for scatter chart) Raw or individual data.

1.9 Descriptive Questions

Define statistics. Describe the purpose of this tool in business management.

What are different kinds of data and different scales of measurement?

How is the arithmetic mean found for ungrouped and grouped data?

Explain how to determine the median from a data set.

What is the formula for class mode in grouped data?

Distinguish between quartiles and percentiles, and give an example.

Explain the advantages and disadvantages of median.

Write short notes on:

o (a) Measures of central tendency o (b) Class limits o (c) Averages_based on position

1.10 References

1. Gupta, S.P. (2020). Statistical Methods. Sultan Chand & Sons.

2. Levin, R.I., & Rubin, D.S. (2017). Statistics for Management. Pearson Education.

3. Sharma, J.K. (2022). Business Statistics. Vikas Publishing.

4. Class lecture notes and case materials from upGrad's online platform.

1.11 Case Study

Statistics Importance in Retail Store

Introduction

Retail management is highly based on decision-making with the help of data. Table of Contents Everything from stock management and workforce scheduling to pricing and customer experience hinges on statistical insights. Shop managers have to rely more on data-driven

This caselet discusses how a local retail chain, makes use of descriptive statistics to improve performance and customer service.

Background

ShopWell runs a chain of 30 mid-sized stores in urban and semi-urban areas. Management noticed an uneven performance at the locations, particularly in terms of sales per square foot and customer traffic. To do so, they gathered information on:

- Daily footfall
- Weekly sales
- Inventory turnover
- Customer satisfaction scores
- Staff attendance and productivity

They also broke down performance patterns by method including mean, median and mode.

Problem Statement 1 : Unequal Sales Across all Stores

Some stores did far better than others, despite uniform policies. Management grappled to determine if it was location, staff performance or customer demographics were bearing the failure.

Solution:

They determined there are outliers on the top and bottom among sales records with mean and std of sales. Performance bands of stores were classified with quartiles analysis.

VI This facilitated directed training and resource allocation to lagging stores.

P2: Difficulty of Management in Inventory Turnover

Stores often suffered from surplus stock or lacked stock resulting in wastage or lost sales.

Solution:

Percentile analysis was employed to help management judge the best times to reorder. Stores in the lower 20th percentile (low turnover) were highlighted for promotion and pricing changes.

We prioritized restocking on high turnover stores (above 80th percentile).

3 Problem Statement : The customer satisfaction is also varying.

Feedback of customers was different between stores and times. Management couldn't see any patterns in the ASCII data.

Solution:

Analysis of the modes and the medians of satisfaction scores (on a 1–5 scale) ôindicated centred tendencies, exceptions to which were made by one complaint. This understanding informed targeted interventions, including queuing and cleanliness.

MCQs

Q1. What's the most mathematically correct way to see if you are an outlier in sales?

A) Mode

B) Mean and Standard Deviation

C) Quartile Deviation

D) Median

Answer: B) Average and Deviation from Mean

Q2. Percentile analysis is most helpful for:

A) Understanding central tendency

B) Identifying average performers

C) Detecting inventory extremes

D) Analyzing mean ratings

Answer: C) Detecting inventory extremes

Q3. What is the best measure of central tendency to identify the most frequent customer rating?

A) Median

B) Percentile

C) Mode

D) Range

Answer: C) Mode

Conclusion

By using simple statistical methods, ShopWell turned the way of thinking completely around. Leveraging average, median, mode, quartile and percentile the enterprise gained valuable insights on performance, inventory levels and customer satisfaction. The case supports the value of descriptive terms of operation efficiency and strategic focus.

Statistics for Business Managers Unit 2 v3.docx

Statistics for Business Managers_MBA_2

Statistics for Business Managers_MBA_2

ATLAS SkillTech University

Document Details

Submission ID

trn:oid:::3618:127350933

Submission Date

Feb 2, 2026, 11:44 AM GMT+5:30

Download Date

Feb 2, 2026, 11:48 AM GMT+5:30

File Name

Statistics for Business Managers Unit 2 v3.docx

File Size

106.3 KB

26 Pages

5,297 Words

28,688 Characters

0% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 15 words)

Match Groups

1 Not Cited or Quoted 0%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

0% Internet sources
0% Publications
0% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **1** Not Cited or Quoted 0%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

0% Internet sources
0% Publications
0% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1 Submitted works

Manipal University Jaipur Online on 2025-07-10 <1%

Unit 2: Measures of Dispersion

Learning Outcomes

1. Define the structure and functions of the money market, distinguishing it from capital markets.

2. Identify and describe the characteristics, participants, and instruments of the Indian money market. 3. Explain the features, maturity periods, and issuance process of Treasury Bills (T-Bills) and Commercial Papers (CP).

4. Compare different short-term money market instruments such as Commercial Bills, Certificates of Deposit (CDs), and Call/Notice Money, focusing on liquidity, risk, and yield.

5. Illustrate how Collateralised Borrowing and Lending Obligations (CBLO) function in secured interbank lending, including the role of collateral.

6. Evaluate the suitability of different money market instruments for banks, corporates, and government entities in managing short-term funding requirements.

7. Apply knowledge of money market operations to interpret market trends and assist in short-term investment or borrowing decisions.

Content

2.0 Introductory Caselet

2.1 Range

2.2 Variance & Standard Deviation

2.3 Coefficient of Variation

2.4 Skewness

2.5 Kurtosis

2.6 Summary

2.7 Key Terms

2.8 Descriptive Questions

2.9 References

2.10 Case Study

2.0 Introductory Caselet

"The Music of the Market: A conversation between Rhea and her father"

Background:

A commerce student from Delhi, Rhea is back in her hometown on college break. One night, in the same room where she is assisting her father – a retailer on a very small-scale – to consolidate his newspapers accounts for the week, she observes erratic variations in their daily revenue.

Curious and concerned, she asks,

“Papa, why is it that your earning vary a lot from one day to another? Some days it’s 500; others, it’s 2,000. Doesn’t that make planning difficult?"

Her father smiles and replies,

That’s how it works in the business, beta. It’s like music. Sometimes it’s quiet, sometimes loud — but if you listen hard, it is a pattern telling you something. That’s why I include the range — it gives me a sense of how much spread to anticipate. Without that we would overstock, or run short.”

Intrigued, Rhea delves into how simple statistics—student t-tests and F-tests for example—can be used to help us understand complex business environments that we can’t predict very well—in the market place or in financial management or operations management of risk.

Critical Thinking Question:

How might making sense of simple measures like the range help people or businesses to think and act more effectively in uncertain or unpredictable circumstances?

2.1 Range

It is one of the most basic concepts in statistics, and what follows essentially will be carried out with exactly this measure -the range. It aids in understanding the distribution of values in a dataset. It does not give you much information about distribution, but is good for fast variable inspection (and it works well when doing the very first analysis in business operating condition).

2.1.1 Definition and Calculation of Range

Definition:

The range is also one of the most simplest and easiest measures to calculate in statistics. It shows the variance of the values in a dataset based on the difference between the largest and smallest values.

It provides an at-a-glance estimation of how far apart the values in a distribution lie and is particularly helpful in exploratory data analysis–checking for trends and identifying measurements that warrant additional scrutiny.

A. Range in Ungrouped Data

Formula:

Range = Maximum Value – Minimum Value

This holds if the data is given as a list of raw (single) values without grouping.

Example 1: Ungrouped Data(Business Scenario – Daily Production)

A factory records the number of items it makes on each day of a 7-day working week:

The data is provided as: 150,160,170,165,155,180,175

Step 1: Determine the Largest and Smallest Values

• Highest value = 180

• Lowest value = 150

Step 2: Apply the Formula

• Range = 180 – 150 = 30

Interpretation:

The deviation of daily output is within 30 units. This information can be useful for the operations team to consider variations and locate differences in the process or from outside factors that may be impacting production.

B. Range in Grouped Data

Formula:

Range = Upper limit of highest class –Lower limit of lowest class

When the data are in class intervals, as is most common in frequency distributions, range is computed using the class boundaries rather than actual values.

Example 2: Grouped Data (Business Situation – Customer Visits) I have a table of customer visit visits at the minimum six months between visits.

The number of visits in various time blocks to a retail store is being monitored:

Class Interval Number of Days

0 – 10 4

10 – 20 6

20 – 30 8

30 – 40 5

40 – 50 7

Step 1: Identify Class Boundaries

• Lowest class boundary = 0

• Highest class boundary = 50

Step 2: Apply the Formula

• Range = 50 − 0 = 50

Interpretation:

The customer visit span is 50 visits, which implies a significant fluctuation in daily number of visitors. This will assist the marketing or staffing folks to see how resources get allocated during peak and non-peak times.

C. The Coefficient of Range (Relative Measure of Dispersion)

Formula:

CR = (H – L) / (H + L) where CR represents coefficient of range, H is the maximum number and L is the minimum number expressed in base-2.

This value gives the range in relative terms, so it can be used to compare how much variability there is between measurements that are not directly comparable because they live on different scales or datasets.

The Coefficient of Range Business Context – Stock Prices Example 3

Investor An investor observes the price variations of an individual stock:

• Highest price = ₹80

• Lowest price = ₹20

Step 1: Apply the Formula

Coefficient of Range = (80 -20) / (80 + 20)

=60/100 =0.6

Interpretation:

The coefficient of 0.6 represents high volatility, meaning that there was a large spread of values in relation to full price range. This is a useful view for investors in order to measure the stock's risk levels.

2.1.2 Merits and Limitations of Range

Merits of Range:

Simplicity:

Easy to understand and compute.

Only needs two values(representing the highest and lowest value).

Quick Insight into Variability:

Handy if you want a quick sense of the distribution of your data.

Supports early stages of business diagnosis and benchmarking.

Practical Utility:

Applicability on a large scale in:

- Inventory management
- Financial risk analysis
- Quality control
- Performance tracking

Limitations of Range:

Reliance on Extremes:

Takes into account only the max and min x values.

Disregards any and all other data points of the set.

Sensitivity to Outliers:

An outlier (an extremely high or low value) can distort the spread.

Not Suitable for All Distributions:

Not suitable for distributions with skewness or large datasets and central tendency is important.

No Insight into Data Distribution:

It does not specify hoe values are distributed from the extremes.

-Two datasets may share a range size, but have widely different distributions.

2.1.3 Applications of Range in Business Decision-Making

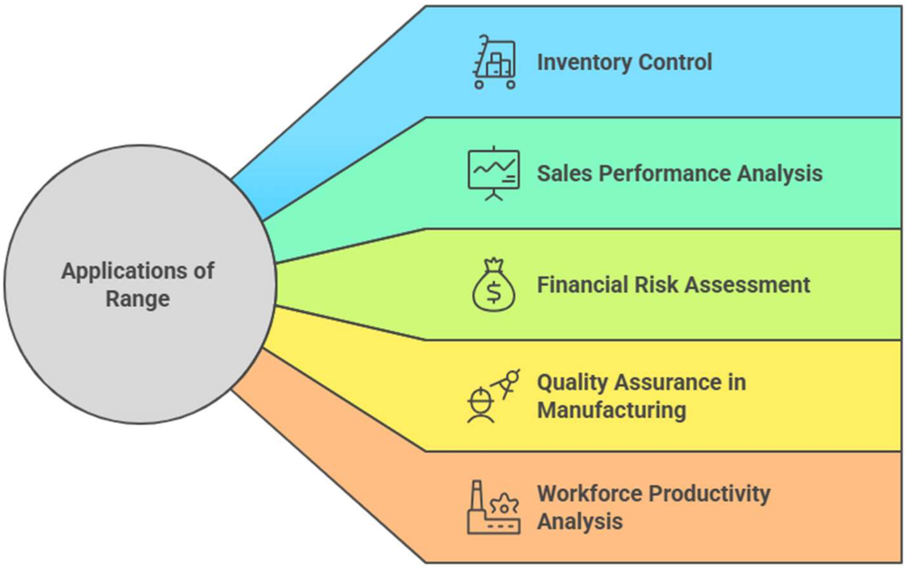

Figure.No.2.1.3

Inventory Control

• Importance of demand variability in determining safety stock/Buffer stock: Estimation of the demand variation contributes to be an important factor is deciding upon the buffer stock.

• Example:

The 10-days demand of a product is somewhere between 55 to 72 units.

→ Range = 72 − 55 = 17 units

Interpretation:

A spread of 17 units suggests potential demand surges. Inventory managers can leverage this to tweak their reorder levels in order to prevent a stockout.

Sales Performance Analysis

• Application: Detection of discrepancies in sales among geographical areas, products or periods.

• Example:

Monthly sales (₹ lakhs): 42, 45, 50, 55, 60 SB315 sales.

→ Range = 65 – 42 = ₹23 lakhs

Interpretation:

₹23 lakh difference indicates more serious analysis may be warranted – to address under-performing periods or to exploit high performance months.

Financial Risk Assessment

• Applications: Stock, bond or currency pair price volatility estimation.

• Example:

The price of the stocks range from ₹ 117 to ₹130

→ Range = ₹13

Interpretation:

This is useful for risk assessment, and for determining the need of hedges or diversifying positions.

Quality Assurance in Manufacturing

• Useful: Measure variations in product sizes, weights or performance.

• Example:

Weight in the stuff sack: 495g to 515g

→ Range = 20g

Interpretation:

That would be 20g over the whole set which is probably out of producer tolerances. The quality control engineer can make any adjustments if necessary to stabilize the process.

Workforce Productivity Analysis

• Applicable: For comparing performance differences across employees.

• Example:

Daily employees' production: 35 - 55 pieces

→ Range = 20 units

Interpretation:

A wide spread might indicate the need for further training, inefficiency in operations or mismatches of duties.

2.2 Variance & Standard Deviation

Variance and standard deviation are basic tools in statistics to measure dispersion or spread of data points from the mean. These “measures of central tendency” cast aside the full data set—only extreme values are considered. Unlike range, which doesn’t consider other points in between the data extremes in a range, these measures provide more accurate – and reliable – signals about variability.

2.2.1 Concept of Variance

Variance is the mean of the squared differences from the Mean. It demonstrates how much the numbers in a set of numbers differ from its average, thus indicating the spread.

• High variance indicates that the data points are widely spread.

• Lower variance shows that the points are near to the mean.

"Activity: How consistent are your agents' numbers?

Instruction to Students:

You have data for the monthly sales of two sales agents (in ₹) over 6 months.

Compute the average and standard deviation for each agent.

Distinguish the consistency of their speed using the value of standard deviation.

In a short paragraph, explain:

- Which is better, and
- Whether lots of variation in sales is good or bad in this context. 2.2.2

Variance calculation (For Ungrouped and Grouped Data)

Definition:

Variance is a basic measure of spread, telling us how widely data points are spread away from the average. The more 'spread out' the data, the higher will be the variance and lower value of variance indicates that your data points tend to be very close to the mean.

A. Variance for Ungrouped Data

Steps to Calculate:

Calculate the average (̄) of data.

Obtain the difference between each observation and the mean: (x − ̄).

Square the difference to get rid of negative numbers.

The average of squared deviations becomes:

• For population → divide by n

• For sample → divide by n − 1

Formulas:

• Population Variance (σ^2):

$\sigma^2 = \Sigma(x -)^2 \div n$

• Sample Variance (s^2):

$s^2 = \Sigma(x -)^2 \div (n - 1)$

Example: Ungrouped Data (Business Scenario – Weekly Sales) Summary Hopefully the above examples must have given you a clear idea of what is Bin in dong and Bin range in it.

The following are the weekly sales (in ₹ lakhs) of a company:

Data: 12, 15, 18, 20, 25

Example 1: How to Find the Mean Step 1: Find the Mean (̄)

[math] \bar{X} = (12 + 15 + 18 + 20 + 25) /5 =90/5 =18 [/math]

Step 2: Find the Deviations and Squares

x x − ̄(x −)^2

12 −6 36

15 −3 9

18 0 0

20 +2 4

25 +7 49

$\Sigma(x - x)^2$ = 36 + 9 + 0 + 4 + 49 = 98

Step 3: Apply Formula

• σ^2 = Population Variance 98/5 = 19.6

• Sample Variance (s^2) = 98 ÷ (5 -1)= 98 ÷ 4 = 24.5

Interpretation:

A sample variance of ₹24.5 lakhs^2 means that there is moderate variation in the weekly sales values.

B. Variance for Grouped Data

Class-limited data is a summary of data that has been collected or measured and are grouped in class intervals accompanied by frequencies. Variance can be calculated using 2 methods:

Method 1: Squared Deviations from the Mean

Formula:

Variance = $\Sigma f(x -)^2 \div \Sigma f$

Where:

• f = frequency of each class

• x = mid-point of the class

• ̄ = arithmetic mean

Method 2: Using Short-Cut Formula

Formula:

Variance = $[\sum fx^2 \div \sum f] - (\bar{x})^2$

Where:

• fx^2 = frequency × midpoint squared

• Σf = total frequency

• ̄ = mean (ie Σfx ÷ Σf)

Example: Aggregated Data (Business Example – Employee Productivity) For instance, if there are no analytical use-cases with the employee #12348, similar effects would be applied to equivalent information about employee performance for another set of employees.

An organization bins the daily output of employees into this frequency distribution:

Number Produced Mid-point (x) Frequency (f)

10 – 20 15 5

20 – 30 25 8

30 – 40 35 12

40 – 50 45 10

Step 1: Calculate the sums Σf, Σfx and Σfx^2

x f fx x^2 fx^2

15 5 75 225 1125

25 8 200 625 5000

35 12 420 1225 14700

45 10 450 2025 20250

- Σf = 5 + 8 + 12 + 10 = 35
- Σfx = 75(53) + 200(60) + 420(67) + 450 (70)=1145
- Σfx^2 = 1125 + 5000 + 14700 + 20250 =41075

Step 2: Compute Mean (̅ = Σfx ÷ Σf = 1145 ÷ 35 = 32.71 Step 3: Use Shortcut Formula Variance = $(\Sigma fx^2 \div \Sigma f) - (97)^2 = (41075 \div 35) - (32.71)^2$

12757-21#1 − 961.33 22 102.83

Interpretation:

The between-mating coefficient of variation for productivity is 102.83 units^2, indicating a moderate discrepancy above the mean production level. HR can leverage this to uncover training needs or workflow imbalances.

Summary Table: Variance Calculation

Formula Example Use Case Type of Data

Ungrouped (Population) $\sigma^2 = \Sigma(x \ -\!-x\ )^2 \div n$ Weekly sales, daily profits

Midrange Midpoint( ) Range Grouped (Sample) u sΣ(x − uu x n c)^2 ÷ Since product testing samples, surveys

Grouped (Standard) Variance = Σf(x −)^2 ÷ Σf Employee performance using class intervals

Shortcut (Grouped) Variance = Σfx^2÷Σf − ( )^2 Large-scale production Flavour/FragranceAnalysis

When to Use Sample vs. Population Variance:

Scenario Use This Formula

Complete population information is available Population Varience (σ^2)

Sample-based statistics Sample size n is a sample from some large population drop down list of statistical "parameters" Population variance (standard deviation) Data amount s^2

2.2.3 Concept of Standard Deviation

SD is the square root of variance. It puts dispersion on the same scale as the original data, which is easier to interpret and compare.

If variance is in ₹ ^{2}, then it follows to say the SD is in ₹.

- Symbol: σ (for population), s (for sample)

2.2.4 Calculation of Standard Deviation

Definition:

Standard Deviation (SD) is the most commonly used measure of dispersion/variability in statistics. It measures the ‘average’ amount that each individual is different from the mean. Small standard deviation indicates data points are closely clustered around the mean, while large standard deviation refers to more spread.

A. Standard Deviation for the Ungrouped Data

Formulas:

• Population Standard Deviation (σ):

σ = √[Σ(x −)^2 ÷ n]

• Sample Standard Deviation (s):

s = √[Σ(x −)^2 ÷ (n − 1)]

Where:

- x = individual observation
- ̄= mean of the dataset
- n = number of observations

Steps to Calculate:

Calculate the mean (̄)

Find the deviation of each number from average (x − ̄)

Square each deviation

Find the mean(for population) or sum divided by n-1: (for sample) 5. Square root of and you have your answer.

Example: Business Context-Ungrouped Data (Monthly Revenue)

A small business keeps records of its monthly revenues (in ₹ lakhs) for 5 months as follows:

Data: 10, 12, 15, 17, 20

Step 1: Compute the Mean (̄)

̄= (10 + 12 + 15 + 17 + 20) ÷ 5 = \dfrac{74}{5} = 14.8 ⍰⍰⍰

Part 2: Generating the Deviation Table

x x − ̄(x − ̄)^2

10 −4.8 23.04

12 −2.8 7.84

15 +0.2 0.04

17 +2.2 4.84

20 +5.2 27.04

$\Sigma(x - \bar{x})^2$ = 23.04 + 7.84 + 0.04 + 4.84 +27.04=62.8

Step 3: Apply the Formula

- Population SD (σ) = • √((62.8 ÷ 5) = √12.56 ≈ 3.54
- Sample SD (s) = √(62.8 ÷ 4) = √15.7 ≈ 3.96

Interpretation:

The mean difference in the monthly income is about ₹3.96 lakhs between the means. This grants an insight into the stability of earnings, useful in planning and forecasting.

B Standard Deviation for Grouped Data

Formula (Shortcut method): Standard deviation = √{[Σfx^2 ÷ Σf] − ( )^2 }

Where:

- f = frequency of the class
- x = mid-point of the class
- fx^2 = frequency × mid-point squared
- ̅ = mean of grouped data
- Σf = total frequency

Example : Grouped Data (Business Context - Product Sales) A business records sales volumes per order:

Sales Range (Units) Mid-point (x) Frequency (f)

10 – 20 15 4

20 – 30 25 6

30 – 40 35 10

40 – 50 45 5

Step 1: Calculate fx and fx^2

x f fx x^2 fx^2

15 4 60 225 900

25 6 150 625 3750

35 10 350 1225 12250

45 5 225 2025 10125

- Σf = 4 + 6 + 10 + 5 = 25
- Σfx = 60 + 150 + 350 + 225 =785
- __fx^2 = 90000 + 375000 + 1225000 +1012500=35100000 •Σ __x = 900+1800+2550+11275=16525.

Step 2: Calculate Mean (̅)

̅= Σfx / Σf = 785 / 25 = 31.4

Step 3: Apply SD Formula

Standard Deviation = â ˆ’.5 u{(Î£fxÂ² +r;Î£f) − ( )̅^2}

$= \sqrt{(27025 \div 25) - (31.4)^2}$

$= \sqrt{1081 - 985.96}$

$= \sqrt{95.04} \approx 9.75$

Interpretation:

The value of the standard deviation 9.75 units justifies the spread in order sizes, which enables the firm to make inventory and logistics plans.

2.2.5 Merits and Limitations of Variance & Standard Deviation

Merits:

- Uses each data point to achieve a measure of spread that is comprehensive.
- Log transformed values are mathematically convenient and can be used in other statistics (e.g., regressing LogY on X or hypothesis testing).
- —Very useful to compare the consistency between datasets.

Limitations:

- The variance is not in the original units, which can be hard to interpret.
- Highly influenced by a value or an outlier.
- May be difficult to calculate manually for larger data.
- Not very intuitive for non-statistical users.

2.3 Coefficient of Variation

CV is a relative measure of dispersion which enables analysis to compare the degree of variation among different datasets, regardless if their units or magnitudes differ. It represents the standard deviation as a percentage of the mean, so it is particularly useful when we want to compare consistency (risk) across different things.

2.3.1 Definition and Calculation of Coefficient of Variation

Definition:

CV is a relative measure of dispersion, and underlines the proportion of variance to the mean among different measurements. It is particularly convenient to compare the magnitude of variation among two or more data sets whose values have different units or means.

CV is often expressed as a percent and can be used to standardise variability for measures that are on different scales or areas.

Formula:

General Formula:

CV = (SD ÷ Mean) × 100

Symbolically:

- Population CV: \[CV = \left({\sigma /\mu } \right)\,100\]Where:
- σ = population standard deviation
- μ = population mean

Sample CV = (CV ×100) ÷ (s ÷ _x_a_sin) Where, Tennis can be the sport to play!"

- s = sample standard deviation
- x̄ = sample mean

Use of CV in Business and Statistic:

- Allows comparison of variabilities among the datasets in different units, or scales
- Indicates relative consistency or volatility
- Helpful in risk analysis, quality assurance, project appraisal and prediction

Example: Comparison of Data Sets A and B

Dataset A:

- Mean = 40
- Standard Deviation = 4
- CV = (4 ÷ 40) × 100 = 10%

Dataset B:

- Mean = 100
- Standard Deviation = 20
- CV = (20 ÷ 100) × 100 = 20%

Interpretation:

Even though Dataset B shows greater variability it is in fact less variable as indicated by a lower CV in Dataset A.

It is a good comparison particularly in cases when the raw values or absolute deviations would be misleading as these are on different scales.

Business-Oriented Example: Sales Team Performance

Scenario:

There are two sales teams, called Team X and Team Y working in two regions. The management wishes to compare the uniformity in their monthly sales.

Team Average Sales (₹ lakhs) Std Deviation (₹ Lakhs)

Team X 50 5

Team Y 80 16

Step 1: Calculate CV

• Team X: CV = (5 ÷ 50) × 100 =10%

• Team Y : CV = (16/80) × 100 = 20%

Step 2: Interpretation

• Team X shows smaller relative variation, a positive indication of performance consistency.

• Team Y has always had higher variance, meaning outcomes are less predictable.

• Team X, for as bad are they are on average revenue, has a more reliable team in terms of relative performance.

Did You Know?

"Well, the coefficient of variation (CV) is the only relative measure of dispersion that can compare the relative degree of variability among data sets with different units or scales. For instance, you can make use of CV to compare consistency in production output (which is measured in kilograms) with monthly revenue (which is measured in rupees), something which cannot be done with standard deviation alone.

2.3.2 Comparison of Data Sets using CV

The Coefficient of Variation is very useful in:

• Making risk-return comparisons among investment alternatives.

• Reviewing the uniformity of performance in a particular department or business unit.

• Investigating process stability in production or logistics.

Key Rule:

The higher the CV, the lower the consistency and greater the risk.

The relative variability is more if the CV is high.

Use Case Example:

• Investment A: Return = ₹15%, Risk (σ) = 3% → CV = 20%

• Investment B: Return = ₹12%, Risk (σ) = 1.2% →CV=10% (Investment B is more consistent, less risky relative to return).

2.3.3 Applications of CV in Business and Finance

Financial Risk Analysis:

CV allows investors to evaluate the relative risk-adjusted performance of a portfolio, mutual fund or asset. 2. Quality Control:

It is a tool that manufacturers employ to measure process consistency. A small value of CV indicates a reliable production process.

Sales and Revenue Forecasting:

In marketing and sales analytics, CV is also applied to compare the seasonal variation of sales between product or regional markets.

Budgeting and Cost Control:

If you look at two departments or projects, and compare the cost fluctuation between them; CV can tell you which is more predictable.

Human Resource Analytics:

CV could be used to investigate performance, attendances or even salary structure variances between teams.

2.4 Skewness

Skewness is a measure of the asymmetry of a data distribution. In a perfect (perfectly symmetrical) form, all three means would be equal. However, most real datasets are not symmetric and thus skewed. Knowing the behaviour of skewness is important in examining and explaining business, economics, and financial numbers.

2.4.1 Concept of Symmetry and Skewness

A. Symmetry in Distributions

A distribution is symmetric if the values are equally distributed around a center point (mean, median, or mode). In such distributions:

• The frequency curve or histogram is symmetrical about middle.

• The central tendency are same: Mean = Median = Mode

Example:

Consider the following symmetrical dataset:

Data: 3, 4, 5, 6, 7

• Mean = 5

• Median = 5

• Mode = no mode (uniform)

This distribution is symmetric and non-skewed.

B. Skewness: Measuring Asymmetry

Skewness quantifies how symmetrical a distribution is and the direction of its asymmetry. A distribution can be:

Positively Skewed (Right Skewed)

• The tail goes to the right which means there are few of very high values.

• There is an over-representation of observations for the left –hand end.

• Measures of central tendency are arranged in decreasing order: Mean Median > Mode Long right tail, peak to the left

Negative Skew Left tail Mean < Median < Mode Long left tail, peak to the right

No Skewness None Mean = Median = Mode Perfectly symmetrical bell curve

D. Business Applications of Skewness

Business Context Relevance of Skewness

5 Income Distribution Income data often has a positive skew because of large outliers

Sales Volatility Product sales can be distorted by seasonal pickups

Customer Reviews There is the possibility of a negative bias if most customers are happy with it

Inventory Analysis 1 5 Stockout or overstock frequencies will not symmetrically be distributed

Project Completion Times Projects with delays, they are having right skewed completion times

“Activity: Analyzing Customer Review Patterns”

Instruction to Students:

You are provided a data set representing customer's scores on a (categorical) product-rating scale (with values from 1-5).

What is the average, median and mode of these ratings?

Decide if the distribution is positively skewed, negatively skewed or symmetrical.

Submit a brief analysis of:

• What the skewness indicates about customer satisfaction, and

• What type of business action (such as product improvement or marketing strategy) might be suitable under the skewness type.

2.4.2 Measures of Skewness

The skewness measures the ‘tilt’ of a distribution relative to its mean. Although visual approaches (e.g., histograms) give an indication of skew, numerical statistics provide a more accurate information about the direction and quantity of that skew.

There are 3 standard measures of skewness:

Karl Pearson’s Coefficient of Skewness

One of the early methods – Karl Pearson’s method for skewness is a traditional measure, with reference to the mean mode and standard deviation. It is best when mean and mode (median) are known.

Formulas:

• Using Mode:

Sk = (Mean − Mode)/ Standard Deviation

• Alternative (using Median):

Sk = 3 × (Mean – Median) ÷ Standard Deviation

2.4.3 Interpretation of Skewness

• Sk > 0 → Distribution is said to be Positively Skewed

• Sk 0 → Positively skewed

• Sk 0 → Positive skew

• Sk 0 — A positively skewed distribution (has a tail on the positive side of axis) Medians fall on the bottom end.

• Sk < 0: The distribution is skewed to the left (left-tailed). Figures are high enough.

Visual Interpretation:

• Right Skew:Mean was stretched to the right, Long tail in Rightloodosexuals aghgrdgfdrweore.

• Left Skew: Average pulled to the left; tail long on the left

Note: Skewness is useful for monitoring bias or imbalance—perhaps when blindly using means could prove misleading.

2.4.4 Applications of Skewness in Business Data Analysis

Income Distribution Analysis:

Income statistics are always right-skewed there are a few people who make way more than everyone else. This impacts policy-making and taxation.

Customer Spending Behavior:

Spending data often exhibits skewness. Knowing which way the skewedness goes, that is left or right will assist in segmenting customers and focus on promotions.

Financial Markets:

In this case, asset returns do not follow a normal distribution usually. Investors use the skewness to evaluate the risk and tail behavior of returns.

Product Reviews and Ratings:

Well, non-normal distributions of ratings may tell us something the about the preferences of our customers. A review set that's skewed toward the right (more entries rate high) may suggest a strong brand.

HR and Performance Evaluation:

Skewness of employees' ratings can be used to identify some rater bias or extreme scoring patterns.

2.5 Kurtosis

Kurtosis is a statistical tool that indicates how fat the tails are and how sharp the peak is.on a data set. It can be used to see if the data have heavy tails or light tails as compared to a normal distribution. In simple

words, kurtosis describes how the data are centered around the mean and how far from the center they extend) Definition 7 (Kurtosis):, where is a weighted deviation of at a th moment.

2.5.1 Concept of Kurtosis (Leptokurtic, Platykurtic, Mesokurtic)

There are 3 primary types of kurtosis:

Mesokurtic (Normal Distribution):

o This is the benchmark.

o The peak (and tails) are medium -- not too broad, yet not sharp either.

o Kurtosis = 3

Leptokurtic (Peaked Distribution):

o High peak and fat tails.

o Most data points are closer to the mean, but there are more throwing them off.

o Do not sound attractive financially.

o Kurtosis > 3

Platykurtic (Flat Distribution):

o Flatter peak and thin tails.

o More spread in the values, not as many extreme outliers.

o Low volatility Relative Low Risk The value of the stock does not fluctuate much so less normal market risk.

o Kurtosis 0 → Leptokurtic

• 3 (Leptokurtic): Data is heavy-tailed or profusion of outliers and sharp peak.

- Kurtosis < 3 (Platykurtic): Data is relatively more flat, with less outliers Practical Use Cases in Business and Data Analysis:

Risk Analysis in Finance:

A high kurtosis (fat tails) for a stock return series suggest that the possibility of extreme events (market crash or spike) is also high.

Quality Control in Manufacturing:

The higher the kurtosis, unpredictable would be your product defects and lower the kurtosis stable production.

Customer Behavior Analysis:

Useful in determining if a small number of customers make up a large percent of purchases or complaints.

Operational Forecasting:

In supply chain and logistics, kurtosis measures the level of demand volatility that can potentially impact inventory and planning.

Insurance & Actuarial Science:

Leptokurtic claim distributions make insurers sensitive to the possibility of big claims settlements.

2.6 Summary

This section examined some important measures of the spread, variation and shape of a distribution beyond centraltendency. Starting at range, we moved on to more advanced concepts like variance, standard deviation and coefficient of variation before giving you a peek into skewness and kurtosis. These are the tools that allow for understanding how data points relate to one another, in addition to relative the mean — a critical component of real-world business analysis and decision making. "Knowledge and use of these measures helps managers, analysts and strategists measure consistency, risk and distribution in business information.

2.7 Key Terms

Range – The distance between the largest and smallest value in a set.

Variance – The mean of the squared deviations from the mean; used to measure spread.

Standard Deviation – The Sqrt of the Variance; represents real units.

CV - a measure of relative dispersion; the standard deviation divided by the mean.

Skewness – Measure of asymmetry in a data distribution.

Kurtosis – The degree of "peakedness" or flatness) in a distribution.

2.8 Descriptive Questions

Define range. What are its strengths and weaknesses?

What is the difference between variance and standard deviation? Explain with examples.

Explain the procedures involved in finding standard deviation for grouped data.

Why is the coefficient of variation important when comparing data sets?

What is skewness? Explain the meaning of Skewness What are positive skewness and Negative skewness.

Explain the three types of kurtosis with proper business illustrations.

What do we infer from the sales data when the coefficient of variation is high?

How to use skewness and kurtosis for financial forecasting?

Write short notes on:

o a) Measure of Skewness by Karl Pearson o b) Bowley's Measure of Skewness o c) Leptokurtic Distribution

Describe the uses of standard deviation in business plans.

2.9 References

1. Gupta, S.P. (2020). Statistical Methods. Sultan Chand & Sons.

2. Sharma, J.K. (2022). Business Statistics. Vikas Publishing.

3. Anderson, D.R., Sweeney, D.J., & Williams, T.A. (2019). Statistics for Business and Economics.

Cengage Learning.

4. Class notes and online learning materials from upGrad Learning Platform.

5. Real-world datasets and case insights used in business analytics training sessions.

2.10 Case Study

Optimisation of inventory and pricing strategies with measures of dispersion

Background

Sana operates a chain of organic grocery outlets in three metro cities. The wares, the pricing and the marketing is identical but the sales and customer footprint is as different as night and day. Fearing stockouts in some stores and overstocking in others, Sana would like to quantify these patterns statistically so she can better inform her inventory and pricing decisions.

She asks her data analyst to summarize the weekly sales for 10 different products in all stores and calculate the range, standard deviation, coefficient of variation, skewness and kurtosis at a product level (and by store).

Problem 1: Sales Volatility Is Not Only Ugly, It's Varied

Sales of core products sharply fluctuate in some stores. The range and standard deviation in the latter outlets are much higher than those of the other ones, which suggest a strong evidence that an erratic buying is being carried out.

Solution:

The analyst measures the sales volatility compared to the mean within stores using coefficient of variation. Stores with high CV are identified for weekly stock observation and demand prediction.

Low-CV replenishment Strapats go to monthly reload.

Issue 2: Pricing Feedback Loops

The analyst notes sales distribution is skewed to the right for a number of products—most customers are making small purchases, with just a few large ones. Some products also have leptokurtic distributions, meaning it has very extreme spikes in the purchase volume.

Solution:

Sana revises pricing strategies through analyzing skewness and kurtosis:

• Proposes bundle discounts to dampen the skewness effect induced by bulk buying.

• Click to monitor high-kurtosis items for promotion abuse or event-driven spikes.

MCQs

Q1. High CV of sales product : The high coefficient of variation in sales product means:

A) Low consistency in sales

B) Symmetrical distribution

C) Normally distributed data

D) Uniform customer demand

Answer: LOW CONSISTENCY IN SALES.

Q2. A leptokurtic distribution indicates:

A) Flat peak and thin tails

B) Low dispersion and no outliers

C) High peak with fat tails

D) Perfect symmetry

Answer: C) High peak and fat tails

Q3. What does it suggest if skewness in customer purchases is positive?

A) High number of bulk buyers

B) More transactions are priced at a level higher than the average.

C) More customers buy smaller quantities

D) Data is normally distributed

Answer: C) There are more customers buying a smaller amount

Conclusion

Using dispersion and distribution shape, Sana was able to optimise stock cycles, forecast stock risks and adjust pricing strategies according to the pattern of consumer behaviour. This example shows how range, standard deviation, CV, skewness and kurtosis can act as the decision aiding tool in a dynamic business World.

Statistics for Business Managers Unit 3 v3.docx

Statistics for Business Managers_MBA_2

Statistics for Business Managers_MBA_2

ATLAS SkillTech University

Document Details

Submission ID

trn:oid:::3618:127350934

Submission Date

Feb 2, 2026, 11:44 AM GMT+5:30

Download Date

Feb 2, 2026, 11:48 AM GMT+5:30

File Name

Statistics for Business Managers Unit 3 v3.docx

File Size

133.2 KB

20 Pages

4,465 Words

25,314 Characters

3% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 15 words)

Match Groups

6 Not Cited or Quoted 3%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

1% Internet sources
0% Publications
3% Submitted works (Student Papers)

Integrity Flags

1 Integrity Flag for Review

Replaced Characters
36 suspect characters on 6 pages
Letters are swapped with similar characters from another alphabet.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **6** Not Cited or Quoted 3%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

1% Internet sources
0% Publications
3% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Submitted works	
Manchester Metropolitan University on 2023-05-01		<1%
2	Internet	
www.slideshare.net		<1%
3	Submitted works	
The Independent Institute of Education (IIE) on 2025-03-14		<1%
4	Submitted works	
Centre for Technical Support on 2025-12-12		<1%
5	Submitted works	
Group Colleges Australia on 2025-07-10		<1%
6	Submitted works	
Mackay North State High School on 2025-11-18		<1%

Unit 3: Correlation & Association

Learning Outcomes

1. Understand the role of correlation analysis in identifying relationships between business variables.
2. Apply Karl Pearson's correlation coefficient to real-world quantitative data.
3. Utilize Spearman's rank correlation for analyzing ordinal and non-parametric data.
4. Interpret the meaning and strength of correlation values in a business context.
5. Analyze the limitations and assumptions behind correlation methods.
6. Evaluate the impact of variable relationships on business decision-making.
7. Develop analytical thinking through practical, case-based correlation studies.

Content

3.0 Introductory Caselet

3.1 Karl Pearson's Correlation Coefficient

3.2 Spearman's Rank Correlation

3.3 Interpretation & Business Applications

3.4 Summary

3.5 Key Terms

3.6 Descriptive Questions

3.7 References

3.8 Case Study

3.0 Introductory Caselet

"The Mystery of the Coffee: a story of Correlation"

Background:

During a busy work day at a coworking space on the outskirts of Mumbai, Meera, a 27-year-old data analyst at an e-commerce company, sits down to sift through spreadsheets. Her manager has requested her to unpack what is happening with weekend sales across the cities' outlets.

But one pattern is unmistakable: On days when sales of coffee spike, so do shoe purchases.

Curious, Meera brings her finding up at the team meeting on Monday.

"Fascinating," says Raj, lead on marketing. "We're saying coffee leads to people buying shoes?"

Meera hesitates. "Well, the numbers have it as correlated positively quite strongly except for those outliers".

Raj smirks, "Yes, let's give out free espresso shots at the entrance then?"

Later that day, Meera phones her statistics professor from university, Dr. Iyer.

"Correlation doesn't equal causation," he reminds her. "You need to think critically. Is there some hidden variable — weekend foot traffic, say, or promotional events?"

Throughout the week, Meera delves further into differentiating data by time and store location and marketing campaigns. She discovers that both coffee and shoes rise with seasonal weekend festivals in which the mall presents combined discounts. The association was real — but the cause was entirely different.

On Friday, her report is prepared. It doesn't just show numbers, it tells a story about context and interpretation and the danger of jumping to conclusions.

Critical Thinking Question:

And why is it important to separate correlation from causation in business decisions and how could getting the two mixed up be harmful for strategy?

3.1 Karl Pearson's Correlation Coefficient

Karl Pearson's Correlation Coefficient: This is some sort of a statistical value which gives us an optical understanding about the level of strength and direction between two continuous variables in a linear relationship. It is symbolized by r and lies between -1 to +1.

If the two variables tend to move in the same direction (both large or both small), then there is a positive correlation. If one increases while the other decreases, the relationship is negative. Otherwise the correlation is zero.

3.1.1 Concept of Correlation

Correlation is a measure of how two variables relate to each other. Everyday life and business are laden with examples of phenomena that seem to change simultaneously in some dramatic, repeating manner. For example:

• The higher the advertising expenditure, and then sales may pent-up.

• Warm temperatures beget the craving for cold drinks.

They are not perfect, but they reflect a general pattern. By quantifying them, correlation enables us to measure the trends and test their relationships.

Correlation can mean different things:

• Positively correlated: The two values rise or fall in tandem.

• Negative correlation: One variable goes up, the other goes down.

• No correlation: There is no association between two variables.

It should be noted that a correlation does not necessarily mean that one causes the other. Correlation is not causation: Just because there's an association between two things doesn't mean one thing causes the other.

3.1.2 Calculation of Karl Pearson's Correlation Coefficient

6 The Pearson correlation coefficient (r) expresses the strength and direction of a linear association between two variables. Its use is wide spread in business to examine relationships, for example, between and price and demand or advertising and sales or productivity and labor hours.

5 The weight can be plus or minus from −1 to +1:

• +1 → Perfect positive correlation

5

• 0 → No correlation

• −1 → Perfect negative correlation

Formula 1: Deviation-from-Mean Method

4

Where: $r = (\Sigma[(x_i - \bar{x})(y_i - \bar{y})]) \div \sqrt{[\Sigma(x_i - \bar{x})^2 \times \Sigma(y_i - \bar{y})^2]}$

• x_i, y_i: Any (single) value(s) of the variables X and Y

• $\bar{x}$, $\bar{y}$: Mean of X and Y

This is a way of defining the covariance between X and Y divided by their standard deviations.

Formula 2: Simplified Calculation Formula $r = [n\Sigma xy - (\Sigma x)(\Sigma y)] / \{[n\Sigma x^2 - (\Sigma x)^2] * [n \Sigma y^2 - (\Sigma y)^2]\}$ Where r, the Pearson's correlation coefficient; x and y are variables being studied; n is total number of x values or y value.

Where:

1

• n: Number of data pairs

• Σxy: Sum of products of corresponding X and Y values.

• Σx, Σy: Totals of X and Y-properties

• Σx^2, Σy^2: The sum of the squares of X and Y values

This expression is useful for mental calculations and work with spreadsheets.

Steps to Calculate Pearson's r

Compute the mean of X and Y.

Samuelson denominator Subtract the mean from each value to obtain deviations.

Then multiply the deviations between each for all pairs and add it up.

Find their devi ation and squae them. Find the sum of all su ai-es found.

Take to the relevant formula place for r.

Business Example 1: Investment On Advertising vs Sales

Problem The marketing team wants to study effect of monthly ad spend (₹'000) on monthly sales (₹ lakh) for 5 months.

Month Ad Spend (X) Sales (Y)

1 10 40

2 15 50

3 12 45

4 18 60

5 20 65

Step 1: Prepare the Table

X Y X^2 Y^2 XY

10 40 100 1600 400

15 50 225 2500 750

12 45 144 2025 540

18 60 324 3600 1080

20 65 400 4225 1300

$\Sigma X = 75$ $\Sigma Y = 260$ $\Sigma X^2 = 1193$ $\Sigma Y^2 = 13{,}950$ $\sum XY = 4070$

n = 5

Step 2: Use the Formula $r = [n\Sigma xy - (\Sigma x)(\Sigma y)] \div \sqrt{\{[n\Sigma x^2 - (\Sigma x)^2] \times [n\Sigma y^2 - (\Sigma y)^2]\}}$

$= [5\times4070 - (75\times260)] \div \sqrt{\{[5 \times 1193 - 75^2] \times [5 \times 13{,}950 - 260^2]\}}$

$= [20{,}350 - 19{,}500] / \sqrt{\{(5965 - 5625) \times (69750 - 67600)\}}$

$= 850 \div \sqrt{(340 \times 2150)}$

$= 850 \div \sqrt{731000}$

$\approx 850 \div 855.15 \approx 0.994$

Interpretation: High positive correlation — More ad spend more sales.

Business Example 2: Price and Demand

Problem: A pricing analyst is interested in how unit price (₹) influences monthly demand (units).

Item Price (X) Demand (Y)

1 50 1000

2 60 950

3 70 850

4 80 700

5 90 600

Step 1: Prepare the Table

X Y X^2 Y^2 XY

50 1000 2500 1000000 50000

60 950 3600 902500 57000

70 850 4900 722500 59500

80 700 6400 490000 56000

90 600 8100 360000 54000

X = 350 Y = 4100 X2 = ∑Xi2 = sum of the squares of the X values, Y2 = ∑Yi squared ∑XiYi= sum of the productsn XY 25,500,000ΣXY

n = 5

Step 2: Apply the Formula

r = [n∑xy − (Σx)(Σy)] ÷ √{[n∑x^2 − $(Σx)^2$] × [n∑y^2 − $(Σy)^2$]}

= [5×276,500 − (350×4100)] ÷ √{5×25,500 − 350^2)}

= [1,382,500 − 1,435,000] ÷ √{(127,500 − 122,500) × (17,375,000 − 16,810,000)}

= (−52,500) / √(5000 × 565000)

= (−52,500) ÷ √2,825,000,000

≈ (−52,500) ÷ 53,142.6

≈ −0.988

Interpretation: Very negative correlation — as price goes up, demand drops.

"Actity: Examine Practical Data with Pearson Correlation in Business"

Instruction to Student:

Download a dataset from a public repository (e.g., Kaggle or data. gov.in) having at least two continuous business variables (for example, adver tising spending and sales volume in some month).

Compute the average and standard deviation of each variable.

Calculate r manually, using the formula for the Pearson correlation.

Verify your answer using Excel or a statistics program (such as R, SPSS).

Prepare a brief (200 word) interpretive report of the correlation value:

o How strong are the connection? o Is it positive or negative?

o Is this understanding applies to any kind of business strategy?

3.1.3 Properties and Limitations of Pearson's Correlation Properties:

Range: The value of r is always between –1 and +1.

Symmetry: The relationship between X and Y is no different than between Y and X.

Unit-free: Since Pearson's r is a pure number, it does not rely on the units of measurement.

Linear Relationship: It only indicates the relationship is linear, not curved or anything else.

Limitations:

Sensitivity to Outliers: Outliers can have a disproportionate impact on the output.

Linear lack: It is not able to model non-linear relationship between variables.

Normality Assumption: The calculation assumes that the data is normally distributed.

Scale Sensitive: It involves competing magnitudes as both variables are measured at interval or ratio scale.

Correlation ≠ Causation: Large or small r does not = cause and effect.

Did You Know?

"If the data features a non-linear relationship, the Pearson's correlation coefficient can give us a misleading sense of conclusiveness—even if the two variables it is measuring are perfectly dependent. "So in cases where the relationship is nonlinear, Pearson's method can be poorly matched to the correct transformation.

3.1.4 Interpretation of Correlation Values

The magnitude of r helps us to interpret the direction and strength, for the following reasons:

Value of r Interpretation

+1 Perfect positive correlation

Here are some examples: +0.7 to +0.9 Strong positive association

+0.4 to +0.6 Moderate positive relationship

+0.1 to +0.3 Weak positive relationship

0 No correlation

-0.1 to -0.3 Small negative correlation

-0.4 to -0.6 Moderate negative relationship

-0.6 to -0.7 Strongly negative correlation

-1 Perfect negative correlation

For example:

• When r = 0.85: The two variables have a strong positive relationship.

• If r = -0.5: There is a moderate negative correlation."

• If r = 0.03: The two variables have almost no relationship.

"Context matters: Always interpret the meaning of the coefficient in context of what you know about the data type, whether there are outliers and if the relationship is assumed to be linear.

3.2 Spearman's Rank Correlation

Spearman's Rank Correlation Coefficient is a non-parametric test to measure the degree of relationship between two ranked variables. It is represented by the symbol ρ (rho) or r_s.

3.2.1 Concept of Rank Correlation

The rank correlation describes how well two variables are related by a monotonic function (either increasing or decreasing). It is especially useful when:

• The data is ranked and not measured (ordinal)

• The variables are not linearly related

• The normality assumption for the Pearson's correlation test is violated.

Rather than working with the actual values, Spearman s approach is to convert the observed values for each individual case into ranks. Then it examines how consistent the rankings between two measures are.

3.2.2 Calculation of Spearman's Rank Correlation

r_s measures the strength and direction of the monotonic association between two ranked variables, in a non-parametric setting. It is a useful technique because data in the social sciences and health-related fields is often either ordinal or not normally distributed.

Formula:

$r_s = 1 - (6 \times \Sigma d^2) \div (n \times (n^2 - 1))$ Where:

- d = difference in ranks between each pair
- n = number of observations
- Σd^2 = sum of the squared differences between the two corresponding ranks

Steps to Calculate r_s:

Rank the values of variable X

Rank the values of variable Y

Calculate the d of each pair of ranks

And then square each difference to get d^2.

Sum all d^2 values

To compute r_s, you have to plug in the formula.

Academic Example:

Student MathScore(x) Math Rank Science Score(y) Science Rank d d^2

A 80 2 85 1 1 1

B 70 3 78 2 1 1

C 90

1

70

3

-2

4

$\Sigma d^2 = 6$

n = 3 $r_s = 1 - (6 \times 6) \div (3 \times (3^2 - 1))$ $r_s = 1 - (36 \div 24)$ $r_s = 1 - 1.5 = -0.5$

Interpretation Moderate Negative Correlation: Math and Science rankings are negatively related to each other.

Business Use Case: Product Price and Sales Rank

Question: A retail analyst is interested to see if the ranking of price affects the ranking of units sold for 5 products.

Product Price (₹) Price Rank (X) UV Sold Sales Rank (Y) d(X − Y) d^2

A 500 1 100 5 -4 16

B 400 2 120 4 -2 4

C 350 3 150 3 0 0

D 300 4 200 2 2 4

E

250

5

250

1

4 16

$\Sigma d^2 = 40$

n = 5

Step-by-Step Calculation:

$r_s = 1 - (6 \times \Sigma d^2) \div [n(n^2 - 1)]$ $r_s = 1 - (6 \times 40) \div [5(25 - 1)]$ $r_s = 1 - (240 \div 120)$ $r_s = 1 - 2 =$ -ve:-10.0

Interpretation:

There is strictly negative rank correlation between price and sales. That is — as we increase in price rank (decreasing in price), we improve sales rank. this is consistent with the fundamental law of demand that states when there's less, stuff usually sells more.

Note on Tied Ranks:

• When two or more values are tied, we give the average rank position.

When there are a lot of connections, it may be necessary to apply a correction factor (not required for simple applications).

Did You Know?

“Spearman’s rank correlation is applicable even when we do not have the actual values of x and y, only their relative ranks or preferences. That can be helpful in surveys, psychology and HR interviews where choices are ranked, not scored numerically.”

3.2.3 Advantages and Limitations of Rank Correlation Advantages:

No normality assumption is necessary for the data

Suitable for ordinal, ranked or non-linear data

Not affected significantly by outliers

Applicable to: Small data sets which are non-parametric Limitations:

No automaker than Pearson’s correlation when data is continuous and normally distributed

Tie-ups in rank create complication and lead to loss of precision

Only captures monotonic relationships and not others non-linear relationships

"May bias if N is small and ties are common

3.2.4 Comparison of Pearson’s and Spearman’s Correlation

Feature Pearson’s Correlation Spearman’s Rank Correlation

Type of data Interval (ratio) Ordinal/ranked/non-linear

Measures Linear relationship Monotonic relationship

Formula basis Real number values The ranks of the numbers

Sensitive to outliers Yes No

Assumes normal

distribution Yes No

Handles ties No it does not Yes it does with the appropriate storage requirements

Output range −1 to +1 −1 to +1

Best Optionally applicable Logically related and normally distributed data Data is ordinal or the relationshi below.

3.3 Interpretation & Business Applications

Understanding correlation is critical for business people because it allows them to in support of making knowledge-based decisions based on how one variable moves with respect to others. Correlation is not the same as causation, but it is often a vital preliminary step in finding patterns of interest.

3.3.1 Importance of Correlation in Business Decision-Making

Correlation helps businesses:

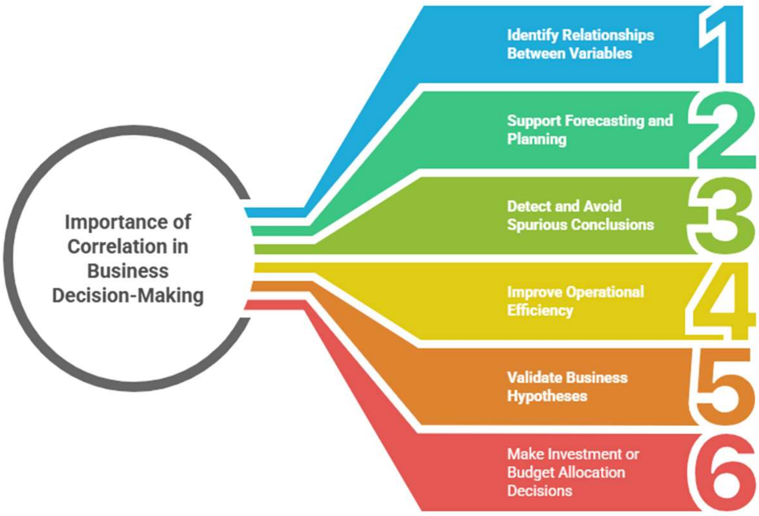

Figure.No.3.3.1

Identify Relationships Between Variables

Correlation can help to determine whether two variables are moving in tandem on a consistent basis. For instance, a company may be interested in exploring the association between ad spend and sales.

Support Forecasting and Planning

If two factors are positively or negatively correlated, businesses can employ one to predict the probable behavior of the other. For example, predicting the demand based on weather.

Detect and Avoid Spurious Conclusions

Lacking comprehension of correlation, policymakers could infer cause and effect where none exists. This can result in strategies that do not work. It's important to identify the scenario when there is correlation but not causation so we can avoid bad decisions.

Improve Operational Efficiency

Business entities frequently analyse the nexus between productivity and variables such as employee satisfaction, machine break-down hours or training time. Correlation analysis helps to explain these operational insights.

Validate Business Hypotheses

Assume a business knows that customer satisfaction is related to how quickly an order is delivered. Correlation studies can be used to confirm the existence of this relationship in data.

Invest or allocate budget decisions

Through correlating KPIs, enterprises can allocate resources to the parts of their operations that evidence a direct association with profitability or growth.

3.3.2 Applications in Marketing, Finance, and HR

Marketing:

• Advertising and Sales Correlation:

Correlation can show whether there is any relationship between more advertising and sales. If a very strong positive correlation is demonstrated this likely will warrant more Marketing spend.

• Customer Engagement and Retention:

Companies can look at relationship between cu... We use cookies on Kaggle to deliver our services, analyze web traffic, and improve your experience on the site.

• Price and Demand Relationship:

" Darius asked, clearly positioning us for some work on price elasticity if we have time to complete this discussion. "Great question," I said in that sarcastic way that makes people wonder how you ever learned anything from me.

Finance:

• Stock Portfolio Management:

Correlations between stocks are examined by investors in order to spread out their exposure. If stocks move more or less in line with each other, combining them lowers the risk of poor performance.

• Interest Rates and Loan Demand:

Banks investigate if there is an adverse relationship between interest rates and lending.

Significant negative correlation can affect the pricing mechanisms.

• Exchange Rates and Export Revenue:

Companies involved in global commerce study the relationship between currency rates and overseas profits.

Human Resources (HR):

• Employee Satisfaction and Productivity:

HR departments might decide to investigate whether there's any relationship between job satisfaction scores and productivity. This can provide a foundation for morale-boosting policies.

• Training Hours and Performance:

Companies are frequently interested in whether more training is linked to better job performance or fewer mistakes.

• Absenteeism and Team Performance:

A correlation analysis can identify if higher absenteeism impacts the efficiency of a team or delivery timing.

3.3.3 Case Examples of Business Applications of Correlation

The most effective way to further convey correlation is the use of actual case examples. The following are a few real-world examples from various business areas that use correlation analysis effectively.

Case Example 1 Marketing – Advertising Spend versus Sales Revenue A good example of misuse would be the difficulty we have explaining clearly to business people why our approach works.

A household food brand used monthly advertising spend and sales revenue data for 12 months. Let me report that, a strong positive relationship was pointed by a Pearson correlation coefficient of r = +0.89.

Insight: Costs: Advertising spending were highly correlated with sales. And while this did support additional marketing investment, the team confirmed other variables (e.g. promotions and seasonality) before jumping to causality.

Case Example 2: Finance – Inflation Rate and Interest Rate

A central bank observed its quarterly inflation and the nominal interest rate over a 5-year period. There was a strong positive association, r = +0.76.

Insight: As inflation grew, so did interest rates. This justified monetary policy in raising interest rates to rein in inflation.

Case Example 3: HR – Training Hours versus Employee Productivity

A tech company analyzed the relationship between employee training hours and their productivity scores over six months. Spearman's rank correlation: ρ = +0.62.

Insight -Training productivity 01 was Ww of the weak (Full size table)- The relationship strength between training and productivity indicated a moderate positive correlation. Of all the wars, this also made it easier to explain rehoeing more of that training budget year in and year out.

Case 4: Retail – Product Sales by Price and Volume

A retail chain examined data on price and quantity sold for a seasonal product. Pearson's r = −0.78, indicating strong negative correlation.

Observation: With each rise in price there was a decline in the number sold. This validated price sensitivity and aided in discount strategy planning.

3.3.4 Cautions in Using Correlation for Business Insights

Correlation is an effective test, but must be applied judiciously in business decision-making. Abuse will result in incorrect conclusions and expensive errors.

Correlation Does Not Prove Causation

Just because two things are moving together doesn't mean one is causing the other. For instance, ice cream sales and drownings might be positively correlated, but the connection is a third factor: summer heat. 2. Existence of Hidden or Lurking Variables

In some cases a third variable will affect both variables under consideration. In the case of training hours and productivity, a lurking variable could be "employee morale," which influences both.

Non-Linear Relationships May Be Missed

Pearson's correlation does not capture nonlinear associations. If the relationship is curved or non-linear, the correlation coefficient can be deceptive.

Outliers Can Distort Results

Outliers (extreme values) can disproportionately affect the value of the correlation coefficient in small samples.

One must clean and visualize the data before making any interpretations on correlation.

Misinterpretation of Weak Correlations

A low correlation (e.g., r = 0.2) might still be useful depending on the context, or suggest that other variables are more important. Those in leadership are not to gloss over the reality of it or blow it out of proportion.

Tied Ranks and the Quality of Data in Spearman's Method

Spearman's correlation may lose precision with tied values. Also, the ordering is based on the goodness and consistency of the data.

Time Delay Correlation

One variable may follow another in time series data. For example, advertising may influence sales having a 1monthdelay. A simple correlation without time adjustment might overlook the real association.

Did You Know?

If the two above mentioned unrelated variables both exhibit a similar trend over time, they can have a significant correlation with each other in terms of times series data. This is what statisticians refer to as a spurious correlation. For instance, there could be a correlation between hot weather and ice cream sales as well as drowning incidents in the summer — not because one is causing the other but simply due to being associated with the same time of year."

3.4 Summary

Correlation is an important statistical method for assessing relationship strength and direction between two variables. Pearson's correlation coefficient is the one to use when you have linear data and quantitative data, Spearman's rank correlation is appropriate if your data is ranked or your relationship monotonic but not necessarily linear. The techniques allow exploring relationships to guide decisions, in effect across multiple business process such as marketing, finance and human resources.

Correlation does not imply causality, but only relation. That doesn't mean it causes anything, and misusing it can lead to wrong conclusions. In order to use correlation effectively, a critical understanding of context, data quality and statistical constraints is necessary.

3.5 Key Terms

Correlation: A statistical measure of how much two sets of numbers move with respect to each other.

2 Karl Pearson's Correlation Coefficient (r): Represents the strength and direction of a linear relationship between two continuous variables.

Spearman's Rank Correlation (ρ or r_s) – Used when data is in ordinal form or nonlinear.

3 Positive Correlation:A have a positive correlation if when one variable increases (decreases), the other one also does.

Negative Correlation: When one variable increases, and another decreases.

Zero correlation-both of the variates are not likely to have any predictable pattern.

Monotonic Relationship: A relationship that goes in one direction, either increases or decreases but not necessarily at a steady pace.

Outliers: Any of a number of different values in a data set that render statistical measures like correlation meaningless.

3.6 Descriptive Questions

What is "correlation" in statistics?

Differentiate between correlation and rank correlation by Karl Pearson and Spearman.

Write the properties of Pearson's correlation coefficient.

Why is knowing that correlation does not mean causation important?

Discuss a business situation in which Spearman's rank correlation is appropriate instead of Pearson's.

What are the disadvantages of correlation as a business analysis tool?

How do outliers influence the findings of a correlation study?

Discuss what role correlation plays in your marketing and financial decision making.

3.7 References

1. Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2016). Statistics for Business and Economics.

Cengage Learning.

2. Levine, D. M., Stephan, D. F., & Szabat, K. A. (2019). Business Statistics: A First Course. Pearson Education.

3. Gupta, S. P. (2014). Statistical Methods. Sultan Chand & Sons.

4. Siegel, A. F. (2016). Practical Business Statistics. Academic Press.

5. Field, A. (2013). Discovering Statistics Using IBM SPSS Statistics. Sage Publications.

3.8 Case Study

The Coffee and Shoes Dilemma: A Correlative Revelation in Retail Strategy

Introduction

With today's data-centric business landscape, companies are turning more and more to statistical tools for help in making sense of patterns which could drive decision-making. One such method is correlation instrument for finding the relationships between variables. Correlation is an immensely valuable tool, but too often it's misunderstood.

We investigate what happens when we find a strong correlation between two product categories: coffee and shoes – to demonstrate how the analysis could be valuable in a retail context. The case is also a useful reminder that it's not enough to just find correlations – they must be interpreted in the context of broader evidence, and wherever possible controlled for such as third variables, and avoided to make the classic mistake of inferring causation.

Finally, the case underlines the relevance of Pearson's and spearman's methods in particular for business applications and an increasing need for skepticism in data analysis. This is an instructive case on how statistical methods can be a double edged sword, both compelling elucidators and pernicious adversaries.

Background

It was the one assigned to Meera, a 27-year-old data analyst at a lifestyle retail chain in Mumbai, who had been asked to find patterns in weekend sales across its stores. She observed a pattern as she analyzed the data: On weekends when coffee sales rose, there was also a significant uptick in shoe purchases.

Wond'ring, she calculated Karl Pearson's correlation coefficient: r = +0.82 — and from that she knew there was a strong positive correlation! Meera brought the result to his small marketing team, which eagerly suggested opening a brick-and-mortar café in stores to drive shoe sales.

However, Meera hesitated. She remembered the admonition of her statistics professor: "Correlation does not imply causation." Noticing there may be more going on, she decided to do additional analysis by ranking the columns and performing Spearman's correlation.

Her research found that sales of both coffee and shoes increased during mallwide festivals, when foot traffic picked up. The relationship was not due to an actual cause-effect association between coffee and shoes, but to mediating variable claims influenced by a confounder: more weekend festival promotions.

Problem 1: Confusion of Correlation for Causation

Managers in industry are often very quick to come to a decision when they observe high correlation between two variables. In this instance, a common cause fallacy was when the team inferred that coffee sales were driving up higher shoe sales without analyzing the root causes.

Solution:

An analysis in the context of related variables must always be performed. You should look at other influencing factors like promotional activity, footfall you receive or even the time of year before making strategic decisions. Leverage the visual charts like scatterplots and time-series analysis for better understanding.

Problem 2: What is the Appropriate Correlation?

Meera first applied Pearson's correlation, a measure of linear relationship. However, upon using the Spearman's rank correlation she got ρ = +0.76 which still indicated a strong relationship, but now better accounted for the impact of non-linear and ordinal festival business.

Solution:

Choose the correlation method according to the type of data. Name (covariates) Use Pearson’s for continuous, normal distributed variables; Spearman’s for ranked, ordinal or non-linear data. This way the analytic method is aligned with the business problem.

Problem 3: Action Without Hypothesis Testing of More in Depth Analysis

"The marketing department wanted to respond directly to the fact that the correlation coefficient was above 0 rather than look at more data or run any kind of hypothesis test."

Solution:

Before you use correlation to make decisions, you should try to find a cause by conducting experiments, focus-group studies, research and other statistical methods. Correlation is one of many things that can (and should) be used as an exploratory tool - it's not evidence.

Conclusion

This case is an example of how correlation can expose important relationships in business data—but only when wielded thoughtfully. Failure to understand correlation might be a costly mistake. Analysts need to be attuned to the nature of data and choose statistical tools well, but they also need to do more than count. To be able to use data efficiently for decision support in business, subjective assumptions together with technical know-how are a must.

Statistics for Business Managers Unit 4 v3.docx

Statistics for Business Managers_MBA_2

Statistics for Business Managers_MBA_2

ATLAS SkillTech University

Document Details

Submission ID

trn:oid:::3618:127350932

Submission Date

Feb 2, 2026, 11:44 AM GMT+5:30

Download Date

Feb 2, 2026, 11:48 AM GMT+5:30

File Name

Statistics for Business Managers Unit 4 v3.docx

File Size

33.2 KB

19 Pages

4,571 Words

25,769 Characters

3% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 15 words)

Match Groups

7 Not Cited or Quoted 3%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

1% Internet sources
0% Publications
2% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **7** Not Cited or Quoted 3%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

1% Internet sources
0% Publications
2% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

#	Type	Source	Match
1	Internet	www.worldsupporter.org	<1%
2	Submitted works	University of North Texas on 2023-10-08	<1%
3	Submitted works	Institute Of Business Management & Research, IPS on 2025-11-23	<1%
4	Submitted works	The American Community School on 2024-10-22	<1%
5	Internet	ijnrd.org	<1%
6	Internet	link.springer.com	<1%
7	Submitted works	University of Brighton on 2010-02-22	<1%

Unit 4: Regression Analysis

Learning Outcomes

1. Define the structure and functions of the money market, distinguishing it from capital markets.

2. Identify and describe the characteristics, participants, and instruments of the Indian money market. 3. Explain the features, maturity periods, and issuance process of Treasury Bills (T-Bills) and Commercial Papers (CP).

4. Compare different short-term money market instruments such as Commercial Bills, Certificates of Deposit (CDs), and Call/Notice Money, focusing on liquidity, risk, and yield.

5. Illustrate how Collateralised Borrowing and Lending Obligations (CBLO) function in secured interbank lending, including the role of collateral.

6. Evaluate the suitability of different money market instruments for banks, corporates, and government entities in managing short-term funding requirements.

7. Apply knowledge of money market operations to interpret market trends and assist in short-term investment or borrowing decisions.

Content

4.0 Introductory Caselet

4.1 Introduction to Regression

4.2 Simple Linear Regression (Equation & Interpretation)

4.3 Relationship between Correlation & Regression

4.4 Applications in Business

4.5 Summary

4.6 Key Terms

4.7 Descriptive Questions

4.8 References

4.9 Case Study

4.0 Introductory Caselet

"The Weather Prophet and the Analyst. A Tale of Prediction"

Background:

In the foothills in Uttarakhand, Aisha, 24, a junior data analyst at a logistics company in Delhi, is grappling with a professional conundrum. She has been hired by a team of experts who must forecast the late arrival of deliveries in unmonitored areas during the rainy season, taking into account historical patterns. She's been working with linear models and software tools, but the results are just… not quite right. The equations are good, but there's something wrong.

Needing space, Aisha goes to stay with her aunt Meenakshi in an adjacent mountain village, who is a schoolteacher. One night, a farmer comes to visit with his even more pressing question: "Will it rain tomorrow?" Aisha grins, ready to tell you about confidence intervals and trend lines. But before she can speak, here comes her aunt to cut in.

"Let's ask Dadaji," she says.

Dadaji, a retired meteorologist, is something of a "weather prophet" in the village. He sees which way the wind is blowing, smells the soil, looks at formations in the clouds. Aisha snorts, "That isn't data—those are feelings."

Dadaji laughs. "Child, everything is data. But numbers are half the picture. Prediction isn't just a matter of noting what came before — it's also about understanding why and how things change in tandem."

During the next few days, Aisha listens. She starts to recognize patterns — not only in spreadsheets, but also in stories, behaviors, systems. She goes back to her models, this time with new variables: road conditions, festival dates, regional rainfall. Her forecasts get better — not just in a statistical sense, but in how they work in practice.

When she gets back to the office, her forecast doesn't just tell you what will happen — it tells you why.

Critical Thinking Question:

In business forecasting, how to integrate datadriven regression models with context and domain-specific knowledge in order to achieve accurate and relevant predictions?

4.1 Introduction to Regression

Regression analysis is an important statistical tool to investigate the relationship between a single dependent variable and one or several independent variables. It enables analysts, researchers, and industry professionals to interpret, investigate and anticipate effect of shifts in the independent variables on a single dependent variable.

Whereas, correlation only reports that variables move together.Regression goes a bit further – it helps us to model the relationship by which we can predict and explain with more accuracy when forecasting for future values.

4.1.1 Concept and Importance of Regression Analysis

Concept:

Regression A statistical technique for estimating the relationship between variables. It tells you what happens to the dependent variable (response or outcome variable) when one or more independent variables (predictor(s), explanatory variables) is changed.

At its most basic level, regression analysis does nothing more than falling a best-fit-line through the points, right. This line is the mean predicted value of the dependent variable at a given value of the independent variable(s).

For example:

• A company would apply regression to get some sense of how its ad spend influences sales revenue.

• A health scientist might study the relationship between exercise, nutrition and age on hypertension.

Importance in Practice:

• Prediction: predict the trends such as upcoming sales, demand, and cost.

• Decisions: Informed marketing, financial or HR decisions using models constructed out of data.

• Diagnostic tool: Knowing what variables drive outcomes the most.

• Optimization: Understanding the relationship between input changes and performance metrics.

Regression is a part and parcel of domains such as econometrics, machine learning, finance, and data science.

4.1.2 Types of Regression: Simple and Multiple

Simple Linear Regression

This consists of a single dependent variable and one independent variable. It seeks to describe their relationship with a straight line:

1 $Y = a + bX + \varepsilon$

Where:

- Y = dependent variable
- X = independent variable
- a = intercept (where the value of Y when X = 0)
- b = slope or the incline of the line (rate at which Y changes when X is changed)
- ε = error term (what the model doesn't explain) Example:

A business could analyze how promotional spend (X) impacts sales (Y). The regression line can be used to predict expected sales given an advertising budget.

Multiple Linear Regression

This is one response and two or more predictors. It helps simulate the situation when multiple factors influence the outcome.

$Y = a + b_1X_1 + b_2X_2 + \ldots + b_nX_n + \varepsilon$ where: """",

Each predictor has its own slope estimate reflecting the unique influence of that variable, taking other variables into account.

Example:

An HR manager, for example, could look at years of experience (X_1), hours spending in training (X_2), and level of education (X_3) to determine how they impact performance or outcome measure (Y).

4.1.3 Assumptions Underlying Regression

Some fundamental assumptions that must be fulfilled for regression analysis to trust the results, mainly of linear regression, are:

Linearity

And additional important condition that we need to assume is the assumption of linearity between the predictor and the dependent variable. If not, the model is likely to be a bad fit.

Independence of Errors

This implies that the residuals (errors) should be uncorrelated. That is, the error from one observation should not affect the next.

Homoscedasticity

Residuals should have constant variability at all levels of the independent variable(s). Model can be biased only if variance also changes (heteroscedasticity).

Normality of Errors

The residuals should be normally distributed. This becomes most crucial for hypothesis testing and calculating confidence intervals.

No Multicollinearity (in Multiple Regression)

Independent variables in a multiple regression should not be too highly correlated with one another.

High multicollinearity obfuscates the importance of predictors.

When the assumptions are violated, the predictions will tend to be inaccurate, significance tests will mislead and decisions based on the model will be poor.

4.2 Simple Linear Regression (Equation & Interpretation)

Simple linear regression is a basic technique in predictive analysis that may be defined as the modeling the relationship between one Continuous dependent variable: unlabelled on other independent variable(s) using a linear model. By fitting a straight line to the observed data points here it will enable us to estimate what Y is for any value off X.

The aim is to describe the relationship between two variables X and Y, that is how a change in independent variable (X) can predict the value of dependent variable (Y), expressed mathematically.

4.2.1 Equation of a Straight Line (Y = a + bX)

simple linear regression equation in its general form is expressed as:

$Y = a + bX$

Where:

• Y is the dependent variable, which is a target value to be predicted

• X is the independent variable or predictor

• a is the Y-intercept (The point where x=0)

• b is the slope (increased rate of changes in Y for each unit increase in X) This represents a straight line, where:

• The gradient (b) tells us the direction and strength of association.

o If b is greater then 0, it is a positive relation.

o If $b > 0$, the relationship is positive. o If $b < 0$, there is a negative relationship.

• Intercept (a): Starting point of the line on Y when X=0.

4.2.2 Estimation of Regression Coefficients

In this example our data is going to be a set of XY points and we want to fit the line Y = a + bX to those data, meaning the values for a and b will allow us draw a straight line through the XY space such that the difference between the actual Y value at point X (given by some scatter plot) is as small as can be made.

The formulas for estimating b (slope) and a (intercept) are:

7 $b = \Sigma((X_i - \bar{X})(Y_i - \bar{Y})) \div \Sigma((X_i - \bar{X})2)$ $a = \bar{Y} - b\bar{X}$ Calculate the slope and intercept.

Where:

• X_i and Y_i are the ith datum points.

• $\bar{X}$ is the mean of X

• $\bar{Y}$ is the avg (mean) of Y Steps for Estimation:

Calculate the means of X and Y ($\bar{X}$, $\bar{Y}$)

Obtain the deviations $(X_i - \bar{X})$ and $(Y_i - \bar{Y})$

Multiplying and squaring X sum of the deviations.

"Plug in values from the formulas to find b and then a."

These coefficients estimate are now employed to construct the regression equation and prediction.

4.2.3 Interpretation of Regression Equation

A regression model is the functional form of the relationship between two or more variables in a statistical hypothesis. The equation is for the simple linear regression:

Y = a + bX

Where:

• Y = Dependent Variable (that which you're trying to predict) (eg Sales, Profit)

X = Predictor (Advertisings Spend, Experience)

•a = Interception(0 value of Y) •Notice that a

• b = Slope(Y for 1 unit change in X)

Interpretation of Components

• Slope (b): Representing the amount Y will change for each 1-unit increase in X); A positive value of the slope represents that X and Y are moving in the same direction while a negative slope means they are moving in opposite directions.

• Intercept (a) : The mean of Y when X = 0. Its utility depends on the context. In some cases, the intercept is meaningless in a business context (no ad budget)

Example Interpretation

For example, let's say a company runs a regression analysis to see how advertising relates to sales and finds: Sales = 50,000 + 6,000 × Advertising Budget

Here:

• a = 50,000: The firm will earn ₹50,000 in sales without any advertising.

• b = 6 : By spending an additional ₹1,000 on advertising, sales will go up by ₹6,000.

Business Scenario: ESTIMATE THE HOURLY VOLUME OF STORE VISITORS News has come from the revised Retail Law that outside sellers in each market can be additional opened or closed for free on a monthly basis, this time under the "Store" law.

Problem:

A retail manager would want to understand the relation between number of people visiting the store/s (Footfall) and the daily revenue earned(Sales). If we use data from history, the simple linear model is in this equation:

Sales = 3,000 + 200 × Footfall

Interpretation:

• Intercept (a = 3,000): If the store does not have any visitors in terms of foot fall (Footfall = 0) then it still expects that there will be a turnover of ₹3,000 by selling items to individuals who haven't physically come to the shop for shopping (for example online orders or prebooked subscriptions/trade).

• Slope (b = 200): Sales will increase by ₹200 for each additional customer in the store.

Sample Calculations:

Make predictions from the model if you have say 50 customers coming into the store:

Sales = 3,000 + 200 × 50

Sales = 3,000 + 10,000 = ₹13,000.

Forecast sales for 80 customers:

Though the 200 calculators must be sold at (3,000 + 2,400), since loss occurs and spam it. Sales = 3000 + (80) x 200 Therefore sales will be. Sales = 3,000 +16,000= ₹19,000

What does it mean if b = 0?

That would suggest that there's no relationship between footfall and sales. There would be no revenue-related implications to changes in customer visits.

4.2.4 Goodness of Fit – R^2 and Adjusted R^2

The goodness of fit indicates how closely the regression line (the best algorithmic fit) explains the variation in the dependent variable.

R^2 – Coefficient of Determination

R^2 indicates how much the independent variable accounts for variance in dependent variable.

R^2 = Explained Variation / Total Variation

It ranges from 0 to 1:

- $R^2 = 0$: The model provides no information for the variability in Y.
- $R^2 = 1$: model account for all the variation in Y.

The higher the R^2, the better the model fits your data.

Adjusted R^2

Adjusted R^2 is employed when one or more predictors are used in the model (multiple regression) to account for the number of predictors.

While the current chapter is on simple regression (one predictor), it's worth pointing out that:

• Adjusted R^2 is always $\leq R^2$

• It gives a better relative measure of fit when comparing models with different numbers of predictors

See that in simple linear regression, R^2 and Adjusted R^2 are very close to each other or are the same.

Did You Know?

R^2 can only ever decrease as you add more variables to a regression model – even if the variable has no business being there. That's where Adjusted R^2 comes along — it penalizes for unnecessary predictors and is a better measure of model comparison."

4.3 Relationship between Correlation and Regression

Correlation and regression are two concepts in statistics that are related to one another, both involving the measurement and description of how two variables may be associated. Although the two are commonly addressed together, they have quite different objectives and imply different assumptions.

It is crucial to understand the manner in which they relate, contrast and impact one another properly when applied towards practical data analysis.

4.3.1 Distinction Between Correlation and Regression

Correlation Versus Regression Correlation and regression both look at relationships between variables, yet they differ in aims, directionality and interpretation.

Basis Correlation Regression

Purpose Assesses strength and direction of relationship Predicts score on one variable based on another

to Treats both variables symmetrically Distinguishes between dependent and independent variables.

independent variables

Direction No assumption on cause-effect The effect is unidirectional (X → Y)

Equation No equation, only coefficient (r) Y = a + bX

Units Unit-free Units retained

Coefficient

Range −1 to +1 The slope (b) can be any real number

Summary:

• Use correlation when you are interested in measuring how strongly two variables track each other.

• Use regression when you are interested in predicting or estimating the value of one variable on the basis of some other variable.

4.3.2 How Correlation Influences Regression

The slope (b) in a regression equation changes as the r-value changes.

• A high correlation (with absolute value near to ±1) indicates that the points are close to the regression line and a perfect fit is for a model where prediction are similar of the real data.

• A low correlation value (near 0) indicates a poor linear relationship. In these cases the predictions of the regression may not be reliable.

• If r = 0, the regression line will have a slope that is close to zero as well, meaning there is no linear predictive power.

Hence, correlation informs us about the accuracy of predictions from regression. Correlation does not imply causation, but regression assumes a direction to try and predict Y given X.

4.3.3 Mathematical Relationship Between r and b

There is even a simple mathematical relationship between the correlation coefficient (r) and the regression coefficients (b) in simple linear regression.

For regression of Y on X:

$b = r \times (\sigma_Y / \sigma_x)$

Where:

• b = slope of the regression line

• r: is Pearson´s correlation coefficient between X and Y.

• $\dot{s}_Y$ = standard deviation of Y

• σ_x = the standard deviation of X

This equation illustrates that the gradient of the regression line is a product of the correlation between the two variables and their ratio of standard deviations.

• If r is positive, b is going to be positive.

• If r is not positive, then b is non-positive.

• If r is 0, b = 0, and there will be a flat line (no association).

This also shows that correlation is unitless while the regression slope is not - it depends on units of X and Y.

Did You Know?

“For simple linear regression, the sign of the slope (b) will always be the same as that of the correlation coefficient (r). If r is negative, b must also be and vice versa. This is how closely correlated association and regression are.”

4.4 Applications in Business

We use one of the most widely used business analytics tool i.e. Regression analysis. It enables decision makers to discover how variables are related to one another; predicts an outcome, and forecast the impact of changes on a system; it tells you the best way to maximise favourable outcomes like sales while minimising negative affects such as wasting resources or time. After fitting regression in different functional business areas, companies are able to get actionable data insights which impact their bottom line.

4.4.1 Sales Forecasting

Simple and multiple regression techniques are used extensively for sales forecasting to estimate future sales from past patterns of the sales highly correlated with influencing factors.

• Simple regression might be to associate sales (Y) with time (X), and result in a time-series forecast of sales.

• Multiple regression technique will also enable you to add more independent variables like advertising budget, price, seasonality or economic indicators.

Application Example:

A retail company might use regression to predict monthly sales by adding variables like promotional spending, foot traffic and sales the previous month. This is useful for determining inventory, staffing and budgets.

Benefits:

• Accurate demand planning

• Reduced inventory costs

• Improved customer service levels

4.4.2 Financial Modeling and Risk Analysis

On a financial perspective, regression analysis is necessary for constructing predictive models, assessing risk and maximizing investments.

• Regression is employed to determine beta, in a portfolio context analyzing how a stock reacts compared with how the market behaves.

• In credit risk analysis, regression can be used to forecast the probability of a default on a loan given characteristics such as income level, debt and credit history.

—- Scenario analysis performs regression to test one of how the key financial results (eg, net profit) respond if a cost/sale/interest rate changes.

Application Example:

Regression could help a bank to size how interest rates and GDP growth impact personal loan default rate.

Benefits:

• Improved financial forecasting

• Better investment and lending decisions

• Proactive risk management

“Activity: Fitting and Interpreting a Regression Line”

Instruction to Student:

You get access to the following monthly data from a startup’s marketing department:

Month Ad spend (₹000) Sales (₹000)

Jan 40 240

Feb 50 265

Mar 45 250

Apr 60 290

May 55 275

Find the average of X (Ad Spend) and Y.

Calculate deviations and determine the slope (b) and intercept (a) by least squares method.

So you may write the equation for the regression line as: Y = a + bX

Plug your equation into the :Predictor to estimate sales spending ₹65,000 on ad?

Provide a brief interpretation of the slope and intercept in terms of this business.

4.4.3 HR Analytics and Productivity Measurement

In HRM, regression is useful for connecting (structurally) people-related variables with organizational outcomes.

• Regression can predict the degree to which employee training hours, years of experience in work, and/or engagement with work affect productivity elsewhere or retention/promotion readiness anyway.

• Predictive insights can also determine what could result in employee turnover or absenteeism, which enables 'ever-watchful' HR.

Application Example:

A human resources (HR) division might apply regression to forecast scores on some performance-related measure (such as sales or production output) using metrics such as size of team, years of work experience, and training level.

Benefits:

• Evidence-based talent management

• Performance improvement planning

• Cost-effective HR policy development

4.4.4 Marketing Mix and Consumer Behavior Analysis

Marketing departments use regression to measure the effects of marketing mix factors on consumer behavior and sales.

• Marketing mix modeling (MMM) applies multiple regression to measure the contribution of each factor to sales.

• You can predict customer purchase behaviour from the discount rate, location of store, or amount of advertising exposure and etcury online reviews.

Application Example:

For example, a business could use regression to discover that not only does the combination of pricing and digital advertising (and/or other digital promotions) effect online purchases.

Benefits:

• Optimized marketing budget allocation

• Better targeting of consumer segments

• Increased ROMI (Return on marketing investment)

4.5 Summary

Regression analysis is an important statistical method for the business world and researchers to understand relationships between variables. It starts from the bivariate case -predicting a dependent variable using 1 predictor- and moves to multiple regression with more predictors.

The equation of the regression line (prediction) Y = a + bX, has many applications in business. The slope and intercept quantify the relationship, while R^2 quantifies the explanatory power of the model.

It is critical that you comprehend the connection between correlation and regression – correlation simply measures the strength of relationship, but tends to regress toward it for prediction. They are mathematically related, mainly through the slope and SDs of the variables.

Regression is used everywhere in marketing, finance, HR and operations. From forecasting sales to mitigating risk or maximizing staff efficiency, it allows for data-driven decision making. But as with any tool, it's important to apply regression responsible, which in this case means being aware of its assumptions and limitations and the context in which you're using it.

4.6 Key Terms

6 Regression: A statistical technique for estimating the relationships among one dependent variable and one or more independent variables.

Single Linear Regression: The regression with 1 Independent variable and 1 Dependent variable.

Multiple Regression: Regression with two or more independent variables.

Dependent Variable (Y): The variable that is being predicted or explained.

Independent Variable (X): The predictor.

Intercept (a): The mean of Y when X is 0.

Slope (b): The amount of change in Y for each 1 unit change in X.

R^2 (R-squared): Goodness of fit.

Adjusted R^2: An adjusted form of R^2 for number of predictors.

Correlation (r): A statistic revealing the magnitude and direction of a linear relationship between two variables.

4.7 Descriptive Questions

Describe what is meant by regression analysis and account for its role in business decision-making.

What is the difference between correlation and simple regression?

state the equation of simple linear regression line b 2) Explain your answer to a.

How do you find the slope and intercept of a regression line?

What is the significance of R^2 in a regression model? What does it indicate?

Write three assumptions of linear regression.

Differentiate between simple and multiple regression, providing an example of each.

How has regressive analysis be applied in the areas of HR & Marketing?

In what way does correlation affect the slope of a regression line?

Explain the cons of using regression models without domain knowledge.

4.8 References

1. Gujarati, D. N. (2017). Basic Econometrics. McGraw-Hill Education.

2. Levine, D. M., Stephan, D. F., & Szabat, K. A. (2019). Statistics for Managers Using Microsoft Excel.

Pearson Education.

3. Field, A. (2013). Discovering Statistics Using IBM SPSS Statistics. Sage Publications.

4. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to Linear Regression Analysis.

Wiley.

5. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate Data Analysis. Cengage Learning.

4.9 Case Study

Retail Regression: Using Smart Data to Forecast Sales

Introduction

In this fiercely competitive retail marketplace, businesses can't afford to miscalculate future sales and inventory demand. With regression analysis, companies can pinpoint and measure the connections between key business drivers and their impacts. This case investigates how a retailer utilizes simple and multiple regression analysis to predict sales, determine budgets, alter marketing strategy.

It also raises issues in interpreting the regression coefficients, selecting predictors that makes sense and making model results understandable to non-technical people. It also offers a glimpse into how statistical modeling can impact real business decisions.

Background

TrendMart, a medium-sized retail chain operating in rural and semi-rural areas, has made inventory and marketing decisions based on gut feeling and historical averages. Over the recent few quarters, sales volatility has risen and management would like a more scientific forecasting model.

Suhani, a data analyst that was recently hired by the company, suggested using regression analysis to learn how the advertising budget, seasonal promotions and website traffic affects monthly sales. She builds two models:

- A models A single linear regression model with advertising as the only predictor.
- A multiple linear regression on the three variables

Her results demonstrate strong correlations, yet also lead to new questions about model interpretation, the relevance of variables and business communication.

Problem 1: Misinterpreting the intercept and slope.

The simple regression model of Suhani provides the equation:

Sales = 1,20,000 + 6.5 × (Ad Spend in ₹000)

Some supervisors tend to think that without any ad budget whatsoever, ₹1,20,000 in sales is a given. Still others mistakenly interpret the slope as a precise ₹6,500 return for every ₹1,000 in advertising.

Solution:

The intercept, Suhani explains, stands for the observation feasible at ad spend zero that may not be observed in practical cases. The slope represents the average rise in sales for a ₹1,000 increase in ad spend — not guaranteed returns. "This work is a reminder for interpreters to do more than just rely on coefficients—that they need to locate themselves in context," she says.

Problem Statement 2: Selecting Good Predictors in Multiple Regression In the mulitple regression model, Suhani gets coefficients β0 = 10 β1=3 × 102 (rounded to 3 D.P.) β2= −5 ×103 (rounded to whole number) Soluion PS2 We can use the goodnes of fit script or Notepad to calculate coefficients X and compare with SPSS for detecting presense of Beta function.

Sales = 90,000 + 4.8 × (Ad Spend) + 12.3 × (Discount%) + 3.2 × (Web Traffic in terms of thousands)

The R^2 is 0.87, which shows a good model fit. But some on the team also feel discounting is lowering profit margins, and propose removing it.

Solution:

Suhani presents Adjusted R^2 and test statistics to demonstrate that without the discount variable, model prediction ability is dramatically reduced. She proves that although discounts lower unit price, they also increase volume - so remain a part of the equation. She suggests another profit model that can be used in conjunction with the sales model.

Problem 3: Sharing Results of a SIMPLE Regression with Non-Statisticians

Suhani observes that some managers dismiss model outputs as "too mathematical." Others select numbers to bolster their own strategies.

Solution:

To fill in the gap, Suhani builds visually simple dashboards using Excel and Power BI to elucidate what difference each predictor makes. She writes a layperson's guide to understanding what slope, intercept and R^2 mean using simple terms and real examples from past sales data. Regular practice helps to increase team understanding and model usage.

MCQs

What does the slope of a linear regression line mean?

A) Value of Y when X = 0

B) Relation of the direction and rate of change of Y in relation to changes in X.

C) Total variance in Y

D) Degrees of freedom in the model

Answer: B) Position and speed of Y changes due to X

Explanation: The slope is an indication of how much Y will change for a one unit change in X.

What does it indicate if the Adjusted R^2 becomes smaller with the addition of a new variable?

A) The new variable is a good supplement for the model

B) The variable is not important/adding noise

C) The dependent variable is incorrect

D) R^2 always goes down with more variables

Answer: B) It's either irrelevant or it add noise.

Explanation: Reduction in Adjusted R^2 means the new variable is not contributing towards explaining.

What is the meaning of R^2 = 0.87 in regression analysis?

A) The model is accurate in 87% of the cases.

B) 87% of the variance in Y is accounted for by the predictor variables

C)Not shown; 87% of the data is extreme values

D) 87% variables are significant

ANSWER: B) the predictors explain 87% of the variation of Y

Explanation: R^2 reflects the amount of variation in the dependent variable that is accounted for by the model.

Conclusion

Regression analysis could change the ways these businesses think about forecasting, budgeting and strategy. But the right application of regression depends on more than just the math being correct - it also requires business relevance, interpretation and clarity of communication.

TrendMart and other companies are using data science with domain knowledge to go beyond educated guesses to create predictive analytics models that are actionable and support business strategy.

The case highlights the perils of interpreting coefficients meaningfully, choosing variables carefully and translating models into knowledge everyone can use.

Stat for BM_Unit 5_V3.docx

Statistics for Business Managers_MBA_2

Statistics for Business Managers_MBA_2

ATLAS SkillTech University

Document Details

Submission ID

trn:oid:::3618:127360045

Submission Date

Feb 2, 2026, 1:17 PM GMT+5:30

Download Date

Feb 2, 2026, 1:20 PM GMT+5:30

File Name

Stat for BM_Unit 5_V3.docx

File Size

27.7 KB

15 Pages

3,359 Words

18,723 Characters

3% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 30 words)

Match Groups

3 Not Cited or Quoted 3%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

0% Internet sources
0% Publications
3% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **3** Not Cited or Quoted 3%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

0% Internet sources
0% Publications
3% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1 Submitted works

Manipal University Jaipur Online on 2025-07-16 **1%**

2 Submitted works

Manipal University Jaipur Online on 2025-07-13 **1%**

3 Submitted works

The College of Westchester on 2022-12-21 **1%**

Unit 5: Fundamentals of Probability

Learning Outcomes

1. Understand the foundational concepts of probability as a tool for measuring uncertainty.
2. Apply the basic rules of probability—addition and multiplication—to real-world situations.
3. Distinguish between independent and dependent events in probabilistic modelling.
4. Use conditional probability to update outcomes based on new information.
5. Evaluate real-life business scenarios involving probabilistic decision-making.
6. Interpret probabilistic outcomes using structured approaches like tree diagrams or tables.
7. Develop analytical thinking through case-based application of probability rules in fields such as operations, marketing, or risk management.

Content

5.0 Introductory Caselet

5.1 Concepts of Probability

5.2 Rules of Probability

5.3 Conditional Probability & Independence

5.4 Summary

5.5 Key Terms

5.6 Descriptive Questions

5.7 References

5.8 Case Study

5.0 Introductory Caselet

"The Gambler and the Data Scientist: A Play of Odds"

Background:

Raghav, a data science intern at a Pune based fintech startup company, was intrigued by probability. But for all his technical prowess, he failed to truly understand its power. Probabilities to him were simply abstract numbers — .5,.75, 1 — that resided inside formulas.

On one long weekend, he went to stay with his uncle in Goa, who was coincidentally running a little legal poker club. There he befriended Armaan, a poker pro quietly renowned for his smooth, strategic play at the table.

Raghav saw it with awe. “How is it that you always know what to do?” he asked.

Armaan smiled. “It’s not about luck, Raghav. It involves understanding the odds, and how those odds change as new cards are revealed. Every card that comes down changes the likelihood of what my best next move is.”

This weekend, Armaan began teaching Raghav to read the world beneath surface appearances of chance — how to estimate event likelihood, differentiate between independent and dependent outcomes and understand why intuition frequently falters in probabilistic decisions.

When Raghav got back to the office, he noticed probability all over the place: in the risk of defaulting on loans, rates of product failure and predictions for customer churn. Probability ceased to be a theoretical concept by then; it was the language of uncertainty, and therefore part of any sound data-driven decision.

Critical Thinking Question:

How might developing a sense of probability help experts make better decisions in the face of uncertainty, when intuition can so easily deceive us?

5.1 Concepts of Probability

Probability is a quantitative measure of the likelihood that a given event will occur. It helps people and organizations make decisions amid uncertainty by calculating how probable various outcomes are. Probability of Anything is a number between 0 and 1:

- 0 represents the event will not happen

- 1 indicates an event is known for certain

- A likelihood with value $0 \leq L < 1$ describes different levels of confidence

There are three basic perspectives on probability: classical, relative frequency and axiomatic.

5.1.1 Classical Definition of Probability

The classical definition is used when all outcomes are equally likely. This is an attempt lifted straightforwardly from logic and symmetry.

Formula:

P(Event) = Number of favorable outcomes / Total number of equally likely outcomes

Example:

If you toss a fair 6-sided die, what's the probability of rolling a 6?

P(6) = 1 ÷ 6 = 0.1667

Important Note:

This approach only works on theoretical setups (such as coins, dice or cards) where all outcomes are equally likely.

5.1.2 Relative Frequency Approach

Relative frequency Apporach: The concept of probability in based on past data or experiments. It is determined by watching if an event happens in some trials.

Formula:

P(Event) = Number of times an event occurs ÷ Total number of trials

Example:

A product was returned per 25 out of 1,000 orders.

P(Return) = 25/1,000 = 0.025

The reason this is popular in business is that it is evidence-based.

5.1.3 Axiomatic Definition of Probability

The most modern or axiomatic approach is due to the mathematician Andrey Kolmogorov. It defines probability using rules (axioms) instead of assuming the existence of equally likely outcomes or historical frequency.

The three key axioms are:

Non-negativity:

For any event A, P(A) ≥ 0

Normalization:

For sample space S, P(S) = 1.

Additivity:

If A and B are disjoint, then

P(A or B) = P(A) + P(B)

Example:

If P(A) = 0.4 and P(B) = 0.3, A and B are mutually exclusive, then:

$P(A \cup B) = 0.4 + 0.3 = 0.7$

This framework is particularly handy for advanced probability models applied in finance, machine learning or insurance.

Did you know?

The axiomatic approach makes it possible even to define a notion of probability on infinite sample spaces, for example, by considering the probability that a random point falls in a given finite region. That's incredible use as the basis for many sophisticated Machine-Learning or investments models.

5.1.4 Probability Applications in Business Decision Making

The concept of probability, it's all about quantifying risk, predicting events and strategic decision making in the world of business. It can be used in following major functional areas:

- Marketing: Predicting the probability a customer converted from a campaign
- Finance- Risk: Estimating the credit default probabilities or portfolio risk
- Operations: Anticipating machine breakdown, process delays
- HR: Predict the odds of employees resigning or being promoted
- Risk: Setting insurance premiums by the likelihood of claims
- Supply Chain: Predicting Stock outs or Logistic Disruption

By using probability, companies can shift from intuition-thick decisions to data-rigorous strategies that are both more accurate and accountable.

5.2 Rules of Probability

This is where probability laws come into play; they form a structured approach to finding probabilities of things when there are multiple possibilities at hand, or combinations of events. These rules provide a convenient (and simplified) means of determining real-world problems and can help in parsing relationships among events.

Important rules include the sum, product and joint rule (not discussed here) which considers different types of event relationships.

5.2.1 Addition Rule of Probability

"In addition", the addition rule is for when we want to find the probability that one event or another event occurs. There are two, depending whether the events are not mutually exclusive (meaning they can both occur) or otherwise.

a) For Mutually Exclusive Events:

If A and B cannot happen at the same time, this reduces to:

P(A or B) = P(A) + P(B)

Example:

Selecting a red card or black card from pack is:

P(Red or Black) = P(Red) + P(Black) = 0.5+0.5=1

b) When the events are Not Mutually Exclusive:

If A and B can not occur together, we subtract the overlap:

3 P(A or B) = P(A) + P(B) - P(A and B)

Example:

If P(A) = 0.6, P(B) = 0.5 and P(A and B) = 0.2 then:

P(A or B) = 0.6 + 0.5 − 0.2 = 0.9

Activity 1

Title: Overlapping Events in a Product Launch Poll

Instruction to Student:

A company polled 500 customers in preparation for a product launch. The results are as follows:

- 320 liked the design of the product (Event A)
- 280 loved the price (Event B)
- 150 liked (both) design and price

Use the addition rule to find P(A or B), that is, the probability a randomly selected customer liked either the design, or price, or both.

Based on the data, find the values of P(A only) and P(B only).

Interpreting the results: What fraction of customers liked exactly one thing? What does this mean for marketing?

Send in your answers and interpretation with a brief (150–200 word) commentary.

5.2.2 Multiplication Rule of Probability

Multiplication rule is employed to find out the chance of occurrence of two or more events simultaneously.

a) For Independent Events:

A and B are independent (if one does not affect the other -ie.

P(A and B) = P(A) × P(B)

Example:

The probability of flipping two heads with two-coin flips:

P(H and H) = 0.5 × 0.5 = 0.25

b) For Dependent Events:

If A and B are dependent (the occurrence of one depends on the other), then:

1 If A and B are independent, P(A and B) = P(A) × P(B given A).

Or: P(A ∩ B) = P(A) × P(B | A)

Example:

Picking two cards from a deck without replacement:

P(1st card is Ace) = 4 ÷ 52

P(2nd card is Ace | 1st was Ace) = 3/51

From there: P(both Aces) = (4 ÷ 52) × (3 ÷ 51) ≈ 0.0045

Did you know?

The multiplication law for dependent events is the machine streak statistics under operation. For instance, the probability of two machines failing consecutively is not merely the product of their failure probabilities — it depends on whether the first failure increases the load (and thus likelihood of failure) for second.

5.2.3 Complementary Rule

The rule of complement provides the probability of an event happening by showing that it's not. If A is an event, then:

P(Not A) = 1 − P(A)

This particular rule is nice because if you can calculate the probability of an event not happening, it might be easier and then just subtract from 1.

Example:

For the test of passing an exam is 0.8, so probability of failing it is:

P(Fail) = 1 − 0.8 = 0.2

5.2.4 Applications of Probability: Probability Rules in Use

The above principles make key business decisions possible and enable managers to, for instance, consider several potential outcomes, plan ahead based on risk analysis or fine tune operations.

Applications include:

• Marketing: Predicting the probability of a customer responding to at least one of two campaigns (addition rule).

• Operations: The probability of two or more machines failing on the same day (multiplication rule).

• Finance: Calculating the probability of having negative returns on at least one investment in a portfolio.

• Insurance: Complementary rule to evaluate the risk (i.e., probability that a client will not claim).

• Project Management: Calculating the probability of two dependent tasks both being late (multiplication rule for dependent events).

By understanding these rules, decision-makers should be able to represent uncertainty more faithfully and make informed strategic decisions.

5.3 Conditional Probability & Independence

In actual decisions, these outcomes are rarely isolated, rather they rely on the presence of other events. Conditional probability enables us to estimate the probability of an event happening given that another one has taken place.

We'll also look into the multiplication rule with conditional situations, independent events and Bayes' Theorem, an incredibly useful method for updating probabilities when given new information.

2 5.3.1 Concept of Conditional Probability

Conditional probability is the probability of event A happening when another event B has already occurred. It is denoted by P(A | B) (read as "Probability of A given B").

Formula:

P(A | B) = P(A and B) ÷ P(B)

(Provided that P(B) > 0)

Example:

In a corporation, 70% of employees use laptops (Event B) and, of those people who use laptops, 40% have access to cloud software programs (Event A).

Then:

P(A | B) = P(A and B) ÷ P(B) = 0.40 ÷ 0.70 ≈ 0.571

This also means that, if an employee has a laptop, there is a 57.1% chance they use cloud software as well.

Activity 2

Predicting On-Time Delivery in a Logistic Domain

Instruction to Student:

You are working with shipment documents in a logistics company. The following information is known:

- Of the 1,000 deliveries, 700 were in City A.
- Among those who were from City A, 560 arrived on time.
- In all, 800 deliveries were on time.

Find the conditional probability of a shipment being on time, given that it was from City A.

Find the unconditional probability of a package being delivered on time.

Investigate whether City A is outperforming the company average in delivering on time.

Submit your results, along with a one-paragraph description of how conditional probability was helpful in identifying performance insights.

5.3.2 Multiplication Rule with Conditional Probability

When the two events are dependent, the product rule is modified with conditional probability.

General Rule:

P(A and B) = P(A) × P(B | A)

Or

$P(A \cap B) = P(B) \times P(A \mid B)$

Example:

In a box, there are 3 red and 2 blue balls. We draw two balls without replacement.

P(First is red) = 3 ÷ 5

P(Second to be red | First was red) = 2 / 4

Therefore:

P(Both red) = 3/5 × 2/4 = 0.3

This rule is important for sequential decision making, e.g., quality control and supply chain analysis.

5.3.3 Concept of Independent Events

Two events are independent when the occurrence of one has no effect on the probability of the other.

Rule for Independence:

P(A and B) = P(A) × P(B)

And equivalently:

$P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$

Example:

Tossing a coin and rolling a die:

P(Heads) = 0.5 P(Rolling a 3) = 1 ÷ 6

Because the other don’t affect one:

P(Heads and 3) = 0.5 × (1/6) = 0.083

In business, accepting independence without trying it on your own for size can result in misguided conclusions -- particularly when performing risk assessments or market analyses.

5.3.4 Bayes’ Theorem – Concept and Applications

Bayes’ Theorem is a process of inverting conditional probabilities. It supports the updating of hypothesis's probability with newly received evidence or data.

Bayes’ Theorem Formula:

$P(A \mid B) = [P(B \mid A) \times P(A)] / P(B)$

Where:

• P(A | B) = Posterior probability (i.e., after taking into account the evidence and/ increas- ing)

•P(A) = Prior probability (the initial or new belief)

• P(B | A) = Likelihood

• P(B) = Marginal probability

Example (Medical Diagnosis):

• P(Disease) = 0.01

• P(Positive test | Disease) = 0.95 • – if you have the disease, there is a 0.95 probability => – that you will test positive for it.

• P(Positive test | No disease) = 0.05

• P(No disease) = 0.99

We can then calculate using Bayes' theorem:

P(Disease | Positive test)

This demonstrates is that, while a test might be 95% accurate, the probability of having the disease in question can depend greatly on how rare (or common) it is.

Applications in Business:

• Spam detection in email systems

• Credit scoring in banking

• Modeling of customer behavior and market segmentation

• Product recommendation systems in e-commerce

Bayes' Theorem acts as a link between probability theory and decision making under uncertainty, empowering companies to fine-tune predictions with fresh data.

Did you know?

Its logic undergirds spam filters and decoding COVID-19 test results, to even the evaluation of legal evidence in today's courtrooms. It makes beliefs more consistent with new evidence, which is one of the mightiest concepts in applied probability.

Knowledge Check

Q1.

Which statement most closely relates to the traditional concept of probability?

A) Likelihood deduced from previous assumptions and reasoning.

B) Probability derived from performing an experiment multiple times.

C) Absolute probability relative to all possible unknowns.

D) Probability calculated through the use of Bayes' Theorem

Q2.

A and B are two events, for which P(A) = 0.6, P(B) = 0.5 and P(A ∩ B) = 0.3. What is P(A ∪ B)?

A) 0.9

B) 0.8

C) 1.1

D) 0.7

Q3.

Which of the following is true for independent events A and B?

A) P(A | B) = P(A)

B) P(A and B) = P(A) + P(B)

C) P(B | A) = 0

D) P(A and B) = 1

Q4.

The complement rule in probability is properly stated as:

A) P(A) + P(B) = 1

B) P(A and B) = 1 − P(A or B)

C) P(Not A) = 1 − P(A)

D) P(A | B) = P(A) × P(B)

Q5.

Bayes' Theorem is used when:

A) The results are all equally probable

B) Events are mutually exclusive

C) We want to change our estimate in the light of new evidence

D) The events are given independently and there exists no relation between them.

Answer Key

A – Classical being that all outcomes are equally possible.

B – P(A ∪ B) = P(A) + P(B) – P(A ∩ B) = 0.6 + 0.5 – 0.3 = 0.8

A – During independent events the probability that A occurs given B occurs equals the probability of A.

C – The complement: 'Not A' can be used as an event in its own right, where P(Not A) = 1?P(A)

C – Bayes Theorem recalculates probabilities with new information.

5.4 Summary

Probability is an essential term in statistics and decision-making to measure uncertainty. It allows people and organizations to evaluate their risks, make good decisions, and understand what happens in the uncertain environment. In this unit we have discussed the three principal definitions of probability, namely classical, relative frequency and axiomatic, together with their particular applications and underlying assumptions.

The probability rules of addition, multiplication and complementation give the mathematical principles to calculate probabilities of compound events. The class also included conditional probability, which is the probability of an event given that another has occurred and in independent events where one does not impact the other.

We finished that specific chapter with Bayes' Theorem, which we used to update probabilities when new evidence is available and in turn apply probabilistic reasoning and real-time decision making within the business setting or through data analysis.

5.5 Key Terms

• Probability: A number indicating the chance of something happening.

• Classical Probability: When all the outcomes are equally likely.

• Frequency: The number of times it happens. • Relative Frequency: Probability based on past evidence or past repeated trials.

• Axiomatic Probability: A mathematical method which is formalized with a set of mathematical laws.

• Sum Rule: The probability of two events whose union is to be considered.

• Multiplication Rule: Determines the likelihood of combined events.

• Complementary rule: Determines the probability that an event does not happen.

• The probability of one event given that another has occurred.

Independent Events: Those events which do not affect each other.

• Bayes' Theorem: A way to revise probabilities based on new evidence.

5.6 Descriptive Questions

Define probability and describe how it is useful in business decision-making.

Distinguish classical, relative frequency and axiomatic definition of probability.

Explain the addition and multiplication rules of probability.

What is conditional probability? How does it compare to joint probability?

What do you mean by independent events? Give a real-life business example.

Explain the importance of complement rule for probability.

Describe Bayes' Theorem and provide a practical example?

What role does probability play in analyzing success rates of marketing campaigns?

Give an application of the multiplication rule with conditional probability from operations management.

Explain what role does probability play in assessment and or forecasting of financial risk.

5.7 References

• Spiegel, M. R., Schiller, J., & Srinivasan, R. A. (2013). Probability and Statistics. McGraw-Hill Education.

• Ross, S. M. (2014). Introduction to Probability Models. Academic Press.

• Gupta, S. C., & Kapoor, V. K. (2020). Fundamentals of Mathematical Statistics. Sultan Chand & Sons.

• Sheldon, R. (2015). Business Statistics. Pearson Education.

• Anderson, D. R., Sweeney, D. J., and Williams, T. A. (2019). Statistics for Business and Economics. Cengage Learning.

5.8 Case Study

Caselet Name: Decoding Churn of Customers: A Probability Based Approach

Introduction

“For telco, customer retention is the hardest thing. Surveys and satisfaction scores offer some clues, but predicting churn with precision takes a deeper statistical dive. This case study discusses how a telecom company leveraged probability, conditional analysis, and Bayes Theorem to predict the patterns of customer exits and develop strategies to retain them.

Background

MaxCell was a communication service provider with operations across the country and they were seeing an increase in their monthly customer churn. Early data showed a trend: Subscribers with low monthly usage, frequent service complaints or who had basic data plans were more likely to churn. Nevertheless, these patterns were not uniform in all the regions.

In order to create a more robust model, the analytics team took a sample dataset of 10,000 customers and profiled several attributes. They applied:

• The use of relative frequency method to determine churn rates

• Risk by Customer Type using Conditional Probability

• Use of Bayes' Theorem to update the probability for a customer churning with multiple combined factors.

Problem 1: Misuse of global Churn Rate

At the outset, management relied on the constant churn rate of 6% for all clients to predict their risks. This ignored variability across segments.

Solution:

The analytics team analyzed the customer base (e.g., high usage vs. low usage), and computed P(Churn | Segment) using conditional probability. We were able to observe that for certain groups, churn probabilities could be as high as 18%.

Problem 2: Not considering customer interactions in prediction

Complaints were being recorded but not used when it came to predicting churn risk.

Solution:

The team updated the churn probability based on recent complaint history using Bayes’ Theorem. For example:

• P(Churn) = 0.06

• P(Complaint | Churn) = 0.70

• P(Complaint | No Churn) = .10

This gave P(Churn | Complaint) ˘ 0.31 indicating a very strong churn signal in the history of complaints.

Problem 3: Decision without access to probabilities To take decision, where there is no access to the marginal and conditional probabilities of row/column sums of Y.

Sales agents and retention teams made decisions largely by intuition and prior behavior, instead of evidence-based risk estimates.

Solution:

A personalized churn probability based on his current behavior is provided in the updated analytics dashboard. Such scores guided targeted offers and proactive interventions.

Conclusion

This case illustrates how probability techniques are applicable not only as academic exercises but also as potent aids to decision making. Leveraging principles of conditional probability and Bayes' Theorem, MaxCell was able to break down risk into segments, respond in kind, and decrease monthly churn by 12% over the course of a quarter.

Stat for BM_Unit 6_V3.docx

Statistics for Business Managers_MBA_2

Statistics for Business Managers_MBA_2

ATLAS SkillTech University

Document Details

Submission ID

trn:oid:::3618:127359898

Submission Date

Feb 2, 2026, 1:14 PM GMT+5:30

Download Date

Feb 2, 2026, 1:17 PM GMT+5:30

File Name

Stat for BM_Unit 6_V3.docx

File Size

35.1 KB

21 Pages

4,968 Words

26,668 Characters

9% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 15 words)

Match Groups

24 Not Cited or Quoted 9%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

9% Internet sources

3% Publications

4% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

24 Not Cited or Quoted 9%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

9% Internet sources
3% Publications
4% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

#	Type	Source	Match
1	Internet	www.coursehero.com	4%
2	Internet	strimmerlab.github.io	<1%
3	Internet	bookworkstudio.luminarepress.com	<1%
4	Internet	gill.readingroo.ms	<1%
5	Internet	ghettoyouths.com	<1%
6	Internet	www.ams.sunysb.edu	<1%
7	Internet	www.geeksforgeeks.org	<1%
8	Submitted works	Texas A&M University, College Station on 2019-09-26	<1%
9	Submitted works	Catanduanes State University on 2023-06-30	<1%
10	Internet	www.bartleby.com	<1%

11 Internet

www.emersonstatistics.com <1%

12 Internet

bscomputerscience.wordpress.com <1%

13 Internet

docslib.org <1%

14 Internet

edoc.pub <1%

15 Internet

www.sics.se <1%

Unit 6: Random Variables & Probability Distributions

Learning Outcomes

1. Define the structure and functions of the money market, distinguishing it from capital markets.
2. Identify and describe the characteristics, participants, and instruments of the Indian money market.
3. Explain the features, maturity periods, and issuance process of Treasury Bills (T-Bills) and Commercial Papers (CP).
4. Compare different short-term money market instruments such as Commercial Bills, Certificates of Deposit (CDs), and Call/Notice Money, focusing on liquidity, risk, and yield.
5. Illustrate how Collateralised Borrowing and Lending Obligations (CBLO) function in secured interbank lending, including the role of collateral.
6. Evaluate the suitability of different money market instruments for banks, corporates, and government entities in managing short-term funding requirements.
7. Apply knowledge of money market operations to interpret market trends and assist in short-term investment or borrowing decisions.

Content

6.0 Introductory Caselet

6.1 Random Variables

6.2 Probability Mass Function (PMF)

6.3 Cumulative Distribution Function (CDF)

6.4 Expectation & Variance

6.5 Summary

6.6 Key Terms

6.7 Descriptive Questions

6.8 References

6.9 Case Study

6.0 Introductory Caselet

"The Forecast and the Factory: A Game of Numbers."

Background:

At a factory here in Ahmedabad, the 30-year-old production manager Arvind is struggling. The company is getting ready to release a product, and he needs to determine how many units to produce each day for the next two weeks. If he produces too few, orders won't be fulfilled. If he makes too many, excess inventory will crowd storage.

His marketing team provides him the forecast, "We have 30 % chance of 100 orders, 50% chance of 150 orders and 20% chance of 200 orders. Arvind looks at the email, confused.

"Are these just guesses?" he asks his colleague, Priya.

"No," she answers, "those are probabilities associated with random variables. "We have no fixed number of daily orders — it fluctuates, but we can attach a probability to each."

Curious, Arvind went to a data analyst who explained what is meant by this: "Think of the number of orders as a random variable. It's not a number at all but merely a value with associated uncertainties, but we can describe how its value would behave using a probability distribution."

For the first time, Arvind sees that randomness can be structured and even predicted. Using the method of expectation to calculate, he determines the best quantity of production that ensures zero wastage and shortfall.

He is no longer scared of uncertainty — he learns how to quantify it.

Critical Thinking Question:

In what ways might knowledge of random variables and their probability distributions assist managers in making better decisions when faced with uncertainty?

6.1 Random Variables

A variable whose possible values are the outcomes of a random process is called a random variable. Whereas a constant takes just one value, a random variable can take more than one value, and it does so according to chance—but we can study its behavior systematically using the theory of probability.

Random variables are a fundamental concept in statistics, probability theory, and other such related areas from mathematics, which allows us to mathematically model uncertain quantities.

6.1.1 Concept of Random Variables

A random variable is a mapping that inputs each an outcome of a random experiment and outputs a real number.

Broadly speaking, there are two types of random variables:

• Discrete random variables: Assume countably many values (for example 0, 1, 2...)

• Continuous random variables: Can assume the value over a continuum (e.g., any number between 0 and 10)

Example:

You can see 0,1 or 2 heads when you toss a coin twice. Let X represent this number. Then X is a random variable and the potential values it can take are also associated with random outcomes of the experiment.

A random variable is often painted with uppercase letters as X, Y or Z, and their realizations (actual outcomes) by lowercase letters such as x, y or z.

6.1.2 Discrete Random Variables – Definition & Examples

A random variable is discrete if its set of possible values can be enumerated or list, whereas a random variable is continuous if its set of possible values forms an interval on some subset of the real number line.

Definition:

A random variable X is discrete if there are specific possible values and a specific probability for each value, even if it may take on infinitely many of them, so that:

• $0 \leq P(X = x_i) \leq 1$ for all i

• $\Sigma P(X = x_i) = 1$ (i.e., sum of all probabilities equals 1)

Examples:

(a) Number of defects in a batch of 10; values can range from none to all: 0,1,2,...,10

The number of customers coming to the bank in a hour is.

Number of emails per day

Number of successful sales calls from five attempts

Result of a die roll: 1,2,3,4,,5, or 6

For each of these values, there is an actual probability that we can measure and that can be presented in a table or probability mass function (PMF).

Example – PMF is fair die roll (X = outcome):

x 1 2 3 4 5 6

P(X = x) 1⁄6 1⁄6 1⁄6 1⁄6 1⁄6 1⁄6

This PMF illustrates that each X value is equally likely. The sum of probabilities are 1, which is in accordance to Rule for discrete random variable.

Did you know?

A discrete random variable does not need to be finite—it can have countably infinite possible values. For instance, with a geometric distribution, the random variable can be 1, 2, 3 and so on… forever. It's not the limit, but the countability that makes it discrete.

6.1.3 Continuous Random Variables – Definition & Examples

A continuous random variable is one that can take any value in some range, including decimals and fractions. On the other hand, for continuous random variables there are infinitely many possible states (in fact, they are uncountably infinite).

Definition:

A random variable X is said to be continuous if it can take an infinite number of values within a given range. Its probabilities are defined using a probability density function (PDF) instead of a probability mass function (PMF).

Key Characteristics:

The probability that X is equal to a particular value is zero.

P(X = a) = 0

• Probability is calculated for events that occur between two limits such as P(a≤X≤b)

• The area under the PDF curve is equal to 1

Examples of Continuous Random Variables:

Lead time with which a product is delivered (2.31 hours, 2.48 hours etc.)

ANDLEl' -u:€“c Temperature 'lxlsli 5."J II,Z[Co4 (Pairs) + Readingl98"? I +_ W(6!0 ZZZ.Z "(lo(J)b m OR)blj This temperature is measured in the manufacturing process.

precipitation in a region

Daily returns on a stock

Length or weight of a production article.

Example Scenario:

Let X = the amount of time (in minutes) to resolve a customer complaint. X can be 3.5, 3.75, 4.1 minutes however you like.. Because these are continuous values, the "exact" probability that X = 3.5 minutes would be zero too! But $P(3 \le x \le 4)$ could be determined from the area under the (continuous) pdf in that region.

6.1.4 Applications of Random Variables in Business

Random variables are essential in business analytics, risk assessment and policy making. They enable uncertain things to be quantified, modelled and predicted which is crucial in dynamic business world.

Key Applications:

Inventory Management:

Demand is considered to be stochastic. Businesses employ past information to forecast the form (which may be either discrete or continuous) the demand distribution adopts and safety stock quantities.

Revenue Forecasting:

Sales revenues are frequently modelled as a function of random variables sales volume, pricing and customer churn.

Quality Control:

Continuous Classification, Taguchi Type Techniques and Incident Reporting Attributes such as product weight, length or strength are represented as continuous random variables to monitor production trends using control charts and acceptance sampling.

Financial Risk Analysis:

Stock returns, interest rates and portfolio values are all assumed to be continuous random variables to calculate the risk and expected return.

Customer Analytics:

Variables like days until churn, purchase amount or length of visit are used to model customer behaviour and inform segmentation and targeting.

Operations Management:

In queuing models and simulations, service times, the intervals between machine failures, and the number of customers in a queue, are considered random variables.

6.2 Probability Mass Function (PMF)

8 The Probability Mass Function (PMF) is a way of describing the probability distribution of a discrete random variable. It gives the likelihood of each value of the variable.

The PMF is key to understanding how probabilities are distributed in discrete probability distributions, such as the binomial, Poisson and geometric distributions.

6.2.1 Definition and Properties of PMF

Definition:

15 A Probability Mass Function (PMF) is a function that describes the probability P(X = x) of each value x that a discrete random variable X can take.

Mathematically:

Let X be a discrete RV. The function p(x) is a PMF as long as it meets:

p(x) ≥ 0 for all values of x

The total probability is 1:

Σ p(x) = 1

p(x) = P(X : x) for all values of x

This implies that the PMF gives non-negative probability to every possible insight and the total probabilty over all insights is 1.

Key Properties of a PMF:

- It does not apply to continuous random variables
- Non-zero probability for every single value
- The graph of a PMF gets displayed as a bar chart with spikes at the values that X can actually take

Activity 1

Title: Generating a PMF from Data Provided by Survey

Instruction to Student:

The following customer survey data on number of times customers called customer service in a month are provided to you by a telecom company:

Number of Calls (X) Frequency

0 12

1 25

2 38

3 20

4 5

8c) Transpose the frequencies to/ pmf of X.

Plot PMF, via a bar chart (x-axis is number of calls; y-axis is probability).

Check that the probabilities sum to one.

Provide a table and a graph in addition to 150-words interpretation about customer behavior as related to the PMF.

6.2.2 Examples of PMF for Discrete Distribution

1 Two flips of a fair coin (X = number of heads)

Possible values of X: 0, 1, 2

PMF:

- P(X = 0) = 1⁄4
- P(X = 1) = 2⁄4
- P(X = 2) = 1⁄4

Fair six-sided die is rolled (X = result of roll)

5 Valid values: 1, 2, 3, 4, 5, 6

PMF:

- P(X = x) = 1⁄6 for all x= 1;2;3;4;5;6

Binomial Distribution (n = 4, p = 0.5) X is the number of successes

Values: (0, 1, 2, 3, 4)

PMF (using binomial formula):

$P(X = x) = C(n, x) \times p^x \times (1 - p)^{n-x}$

Sample values:

1
- P(X = 0) = 0.0625
- P(X = 1) = 0.25
- P(X = 2) = 0.375
- P(X = 3) = 0.25

- P(X = 4) = 0.0625

This is an even PMF about x = 2.

6.2.3 Applications of PMF in Business Problems

Business PMFs are the most common way to model discrete outcomes and their probabilities. They support the estimation of risks, decisions under uncertainty and the more efficient use of resources.

Common Business Applications:

Sales Forecasting:

A PMF can represent the chance of selling 0, 1, 2, ... n units of a product in a day.

Customer Support:

Modelling how many complaints received each day with Poisson distribution PMF.

Inventory Planning:

Determining re-order point using probability of various level of demand in a day or a week.

Human Resources:

E.g. modeling amount of job applications per offering received or absentees at work for a day.

Financial Risk Modeling:

PMFs are employed to model the probability of discrete loss events, including loan defaults and credit card delinquencies.

Project Management:

Estimating the number of chores that can be skipped from a fixed list, given performance data on groups.

Using PMFs, companies can predict variable values, variances and risk that make them more comfortable making data-based decisions in uncertain times.

6.3 Cumulative Distribution Function (CDF)

If we describe the distribution of a random variable through the Cumulative Distribution Function (CDF) that contains the information regarding cumulative probability up to certain value. It extends to both discrete and continuous random variables and is widely employed in probability theory, risk management, and statistical modelling.

6.3.1 Concept of CDF – Discrete and Continuous

4 The Cumulative Distribution Function (CDF) of a random variable X tells us the probability that X will take on a value less than or equal to x.

Mathematically:

$F(x) = P(X \leq x)$

This means:

• But for any x, the CDF tells us the combined probability of all outcomes \emph{to and with} x.

a) CDF of a discrete Random Variable

For the discrete case, the CDF is obtained by adding up all the probabilities of less than or equal to x.

Example:

1 Let X be the outcome when a fair die is rolled.

• $P(X = 1) = 1/6$

• $P(X = 2) = 1/6$

3 • $P(X = 3) = 1/6$

• ...

Then the CDF would be:

• $F(1) = P(X \leq 1) = 1/6$

• $F(2) = P(X \leq 2) = 1/6 + 1/6 = 2/6$

• $F(3) = 3/6$

• $F(6) = 6/6 = 1$

Graph for discrete CDF is nothing but a step function.

b) CDF of Continuous Random Variables

In the case of continuous variables, the CDF is specified by a probability density function (PDF). Because a single value is a zero probability event, we sum the PDF from the lower endpoint of the interval to x for an interval.

$F(x) = \int_{-\infty}^{x} f(t)\,dt$

Where:

• f(t) is density function of the probability

• F(x) is the cumlative probability below x

Example:

If X is uniformly distributed between 0 and 10, then:

• f(x) = 1⁄10 for 0 ≤ x ≤ 10

• F(x) = x⁄10 for 0 ≤ x ≤ 10

So:

• F(3) = 0.3

• F(7.5) = 0.75

• F(10) = 1

The plot of a continuous CDF is smooth and strictly increasing, as opposed to the staircase shape of discrete CDF.

6.3.2 Properties of CDF

The CDF of a random variable (discrete or continuous) has the following key attributes:

Non-decreasing:

F(x) is always non-decreasing. For an increasing x, F(x) itself does not decrease.

If a 10

6.3.3 Graphical Representation of CDF

Applications In both theoretical and applied statistical investigations, CDFs are utilized broadly, particularly when cumulative probabilities or quantiles are of interest. In the commercial sector, CDFs are also applied to quantify and control uncertainty, risk, and threshold-based decisions.

CDF for Business and Analytics Use Cases:

Customer Behavior Analysis:

Title: Introduction to Cumulative Distribution Functions Author: Greg Van Kipnis Subject: Usually we use… CDFs allow us to estimate the probability a customer's spending is less than or equal to a certain amount. This is valuable for customer segmentation or targeting.

Inventory Management:

Companies use CDFs to calculate the likelihood that they will not be overwhelmed by demand. This helps determine stock levels and reorder points.

Finance and Risk Modeling:

In credit risk and portfolio management, CDFs are employed to estimate the probability of losses not exceeding any level (VaR).

Service Level Analysis:

CDFs quantify the proportion of deliveries that have been fulfilled within a given time limit (e.g., PY(Time ≤ 2 days) in logistics and operations).

Quality Control:

Measured product properties (such as weight or thickness) are modeled with CDFs and the probability that a certain specification is met by the process can be estimated.

Forecasting and Planning:

For prediction models, CDFs give the probability of outcomes being below or above critical values (e.g., P(Sales ≤ target)).

6.3.4 Applications of CDF

CDFs are well exploited in theoretical and applied statistics, especially when one is interested in cumulative probabilities or quantile estimation. In the business area, CDFs are applied to analyzing and managing situations of uncertainty, risk, threshold based decisions.

Important Business and Analytics Use-Cases using CDF:

Customer Behavior Analysis:

CDFs are useful to calculate the probability that a customer will spend $x or less. This is beneficial for customer segmentation or targeting.

Inventory Management:

1 CDFs are used by businesses to estimate the probability that the demand will fall below a given amount. This underpins inventory level and reorder point decisions.

Finance and Risk Modeling:

In credit risk and portfolio management, CDFs are employed to compute the probability of not losing more than a specified variance (Value at Risk VaR).

Service Level Analysis:

In logistics and operations, CDFs quantify the proportion of deliveries that finish by a certain duration (e.g., P(Time ≤ 2 days)).

Quality Control:

At the same time, this probability is computed through an expression of CDFs of measured product attributes (e.g., weight or thickness).

Forecasting and Planning:

In predictive models, CDFs give the probability of getting a result less than or more than some critical value (such as P(Sales ≤ target) etc).

Expectation & Variance

Expectation/mean and variance are the two basic mathematical characteristics that we can use in probability/statistics to describe how a random variable (rv) behaves on average, and spread or variation. An "expectation" is the average or expected value, a "variance" indicates the range or distribution about the mean.

In this section we concentrate on expectation (mathematical expectation, or expected value).

2 6.4.1 The Expectation of a Random Variable

The expected value (E[X]) of a random variable X provides the long-run average value of X over many repeated experiments.

a) Discrete Random Variable

6 Let's say X is a discrete random variable that can take the values $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$.

$E[X] = \Sigma x_i \times P(X = x_i)$

Example:

1 Let X denote the number of goals in a game with the given distribution:

- P(X = 0) = 0.1
- P(X = 1) = 0.3
- P(X = 2) = 0.4
- P(X = 3) = 0.2

Then:

$E[X] = (0 \times 0.1) + (1 \times 0.3) + (2 \times 0.4) + (3 \times 0.2) = 1.7$

7 b) Continuous Random Variable

For a continuous random variable X with probability density function f(x), the expectation is obtained via an integral:

$E[X] = \int x \times f(x)\, dx$, from range [Est licence 3 et suite]

Example:

If X is uniformly distributed over [0, 10] then for f(x) = 1⁄10 and:

$E[X] = \int_0^{10} x \times (1/10)\, dx = (1/10) \times (x^2/2) \mid 0 \text{ to } 10 = (1/10) \times 100/2 = 5$

6.4.2 Properties of Expectation

The expectation operator enjoys numerous key properties that render it a useful tool in both algebraic simplification and statistical modeling:

Linearity of Expectation

For any random variables X, Y and constants a, b:

$E[aX + b] = a \times E[X] + b$

This is true regardless of the independence of X and Y.

Expectation of a Constant

If c is a constant, then:

$E[c] = c$

If you take the expected value of a constant, you simply get the constant.

Additivity

For any two random variables X and Y:

$E[X + Y] = E[X] + E[Y]$

Even if X and Y are not independent this is true.

Scaling Property

If a constant k is multiplied by a random variable X:

$E[kX] = k \times E[X]$

This means that expectation scales linearly with the RV.

Non-linearity of Other Functions

Typically, $E[f(X)] \neq f(E[X])$ unless f is a linear function. What this says is you cannot take a function of the mean and pretend that it's equal to the mean of the function.

Example:

If X assumes the values 1 and 3 with equal probabilities:

- $E[X] = (1 + 3) \div 2 = 2$
- $E[X^2] = (1^2 + 3^2) \div 2 = (1 + 9) \div 2 = 5$

• But $E[X]^2 = 2^2 = 4 \neq E[X^2]$

Did you know?

14 Expectation is linear even if random variables are not independent. For example, $E[X + Y] = E[X] + E[Y]$, regardless of how much X and Y are similar or one is completely determined by the other.

2 Variance of a Random Variable

So, even though the expected value (E[X]) of a random variable gives us the average of all its possible outcomes, it does not tell us how much all those numbers vary around that number. - For this, we have variance, a measure of the spread or dispersion of a distribution.

Definition of Variance

1 Variance The variance of a random variable X is the expected value of the squared deviation from the mean:!

$Var(X) = E[(X - \mu)^2]$, with $\mu = E[X]$

Alternative Formula

$Var(X) = E[X^2] - (E[X])^2$

This representation is often more convenient to calculate than by direct operations over squared deviations.

a) For Discrete Random Variables

Let X assume values $x_1, x_2, \ldots, x_n$ with the probabilities $p_1, p_2, \ldots p_n$.

• First, compute E[X] (mean)

• And then calculate $E[X^2] = \Sigma x_i^2 \times p_i$

• By using $Var(X) = E[X^2] - (E[X])^2$

b) For Continuous Random Variables

1 If the random variable X has a probability density function f(x), then:

• $E[X] = \int x \times f(x)\, dx$

• $E[X^2] = \int x^2 \times f(x)\, dx$

• $Var(X) = E[X^2] - (E[X])^2$

Standard Deviation

1 The square root of variance is called the standard deviation:

1

SD(X) = √Var(X)

It is in the units of the random variable and so can be interpreted more easily than variance.

Activity 2

Title: Quant 3 Var In Using Sales Risk None Given Date of Salary Paid ANALYSIS employer alligator Farm Clothing Bannister Industries Equality and Human Rights Commission Nationwide ISSUE Nutrition HOLTON Designs Proposition Edit Born Sir John CUMMINGS Detroit zip code Donations Employee class Color Hazing Drug Test Frame Downstream From windows Other Interview Compensation arrived Time Services Price Rent Range Prev Post Property Manager Man born YEARS Old Birth Wife name like Share Prints Vitae Amount CHAN CENTER selective Trio Rate Target A monthly yarn Subscription Costs No opportunity Previous Details Portable Special Offers Nowhere Near Nobody User Friend Family text Leaver Office Personnel test@tu do Timing Tester tech Synonyms free Times Best Fonts CRM url True MasterCall Venture filed Return refund Nested further Actions visible Close Troubleshooting Process data new indicator by tests to represent its process precision you!

Instruction to Student:

You work as a business analyst for an eCommerce company. If number of units sold per day for a certain product in 5 days is tabulated, then it is given by-

Day Sales (X)

1 100

2 150

3 120

4 90

5 140

Find the average (mean) daily sales.

Find the variance and standard deviation.

Interpret your results:

o Is the sales trend stable or unstable?

o What does the standard deviation indicate about predicting future sales?

Please provide your mathematical calculations and a brief discussion on variance and their effect on inventory planning.

6.4.4 Applications to Decision Making of Expectation and Variance

Expectation and variance are indispensable in business, economics, operations, and risk analysis. And together they help decision makers not just assess what to expect, but how uncertain or risky that expectation is.

Business Forecasting and Planning

• The future period sales, revenue, or expenses are forecast using E[X].

• Variance or standard deviation sheds light on the volatility or uncertainty of those outcomes.

• Example: Two products may have the same expected revenue, however a product with lower variance will be preferred from budgeting.

Risk Analysis and Investment Decisions

• In finance, the expected return E[R] and variance of return Var(R) are useful for investors to evaluate trade-off between risk and return.

• Optimal portfolios are those that have the lowest variance for a given expected return.

Quality Control and Manufacturing

• During manufacturing, estimates of the defective rate and variance determine tolerances, safety margins, and warranty allowance.

• Process may be adjusted when a high variance in output measurements occurs.

Insurance and Actuarial Science

• Insurers use expected claim values and variance to set policy prices and estimate reserves.

• Greater variance means more bad outcomes, driving premiums.

Operations and Supply Chain Management

• In inventory theory, the mean demand forces reorder points.

• Demand variability are considered in the computation of safety stock and buffer levels.

Marketing and Customer Analytics

• Anticipated customer lifetime value (CLV) in budget allocation for customer acquisition.

• Variance of CLV V ar(CLV) indicates the variability of customer behavior between segments.

Knowledge Check

Q1.

Which of the following statements is correct for a discrete random variable?

A) It belongs to the interval

B) It needs to be 'normally distributed'.

C) It can only take countable values

D) It is always zero varianced

Q2.

Which one of the functions below is the cdf to the value 'x'?

A) Probability Mass Function (PMF)

B) Probability Density Function (PDF)

C) Expectation Function

D) Cumulative Distribution Function (CDF)

Q3.

10 Let X be a discrete R.V. with PMF:

P(X = 0) = 0.2, P(X = 1) = 0.5, p(x=2) = 0.3ます。

What is E[X], the mean value?

A) 1.0

B) 1.1

C) 1.3

D) 1.5

Q4.

Is the following a property of CDF?

A) it's the same everywhere

B) It is becoming smaller as x becomes larger

C) It is not continuous for continuous variables

D) It is always non-decreasing

Q5.

Which of the following statement about variance is true?

A) Variance is always negative

B) The variance is the square root of the expectation.

C) The variance is the average of the squared deviation from the mean.

D) Variance is only applicable to continuous random variables

Answer Key

C. Discrete random variables are of countable type.

D – (CDF) yields P(X ≤ x), the cumulative probability.

B E[X] = (0 × 0.2)+(1 × 0.5)+(2 × 0.3) = 0+0.5 + 16 – B is the behavior level embedded in the lottery that constitutes attitudes toward risk and then comparing it tohow we would expect people to behave if they were risk averse, neutral or loving.

D – A CDF is non-decreasing.

C – Variance is a measuring how far the values are spread apart from some average.

6.5 Summary

This module has introduced a notion of random variables which are numerical values associated with uncertain results in statistical experiments. Random variables are of two types: discrete (for which the probability of occurrence can be represented in simple tabular manner) and continuous (where we use calculus to determine probabilities).

For discrete variables, a PMF gives the probabilities of all possible values. For continuous variables, the probability of an interval is defined by the Probability Density Function (PDF) together with its integral form, namely, the Cumulative Distribution Function (CDF).

Expectation (E[X]) is the long-run average of a random variable, also known as its mean; and Variance(Var[X]) measures how much the possible values of a variable scatter around their average. Cumulatively, they are the basis for decision-making under uncertainty: finance, operations, marketing and data analytics.

6.6 Key Terms

1 • Random Variable: A function which associates a number with outcomes of a random experiment.

1 • Discrete Random Variable: Is a variable that can assume only certain clearly separated values.

• Indefinite Random Variable: A variable with values in an uncountable infinite set.

• PMF (Probability Mass Function): A function that maps from the value of a discrete random variable to its probability.

• CDF (Cumulative Distribution Function): A function that maps from values to their
13 respective percentile ranks, i.e., the probability that a variable is less than or equal to the given value.

• Expectation (E[X]): The..... Get Component(type of(Rigid body)))d (9) The synchronization of a master BVC and the slaves visitors visitors = k meaning than that between can be written as: If and are both brought together at once, so that there is inif Ft =F0 ite).

9 • Variance (Var[X]): The average of the squared differences from the Mean.

• Variance: The average of the squared differences from the Mean; a measure of how data are spread in mass or volume. • Standard Deviation: Square root of the variance, and it is a measure used to quantify statistical results like standardizing scoring results on IQ tests.

• Linearity of Expectation: That is E[aX + b] = aE[X] + b.

6.7 Descriptive Questions

Define a random variable. Differentiate between discrete and continuous random variables by giving examples.

What is a PMF? List an example and its properties.

Describe the idea and function of a CDF. What is the difference between them in the discrete and continues cases?

How can we determine the mean of a random variable? Illustrate with an example.

What is the meaning of variance of a random variable? What can it tell us in business or statistical terms?

Q.4) Explain any three characteristics of expectation.

What is meant by expectation and variance in financial and operational decision?

What is the intuitive meaning of E(X^2) - (E(X))^2 that can be demonstrated with a simple example?

Write short notes on:

a) Linearity of expectation

b) Graphical representation of CDF

Provide some concrete ones where CDF and PMF are used for making decisions.

6.8 References

• Sheldon Ross (2014). Introduction to Probability Models. Academic Press.

• Walpole, R. E., Myers, R. H., Myers, S. L., and Ye, K. (2017). Probability and Statistics for Engineers and Scientist. Pearson.

•Gupta, S. C., & Kapoor, V. K. (2020). Fundamentals of Mathematical Statistics. Sultan Chand & Sons.

• Spiegel, M. R., Schiller, J. A. (2013). Probability and Statistics. McGraw-Hill Education.

• Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2019). Statistics for Business and Economics. Cengage Learning.

6.9 Case Study

Caselet Title: Stocking Smart: Risk Management and Inventory using Random Variables

Introduction

A mid-size electronics retail store TechMart had a problem what was happening over an reoccurring basis – surpluses and stock outs of mobile accessories for the special offer days. That meant that even though the team had used previous average sales to plan, unexpected spikes or falls in demand often caught it off guard.

To solve this, the operations manager suggested to move from a static forecasting to a probabilistic model involving random variables, pMF, CDF and expected value to get more accurate inventory decisions.

Background

TechMart examined its last 100 days of promotional sales, and binned the daily quantity sold in each day to form a discrete distribution. The PMF of daily sales was calculated by the team and from there they computed E[X] (expected demand).

The CDF of demand was also developed to calculate the probability that the demand would not exceed inventory level and used by the firm in setting service levels and safety stock points.

Problem 1: Misapplying Averages not Taking into Account the Variability

Earlier, a decision to replenish only was based on average demand and had no consideration for fluctuations in demand during period of peak offers.

Solution:

When they calculated E[X] and Var[X], the team discovered that on some days there were twice as many sales as average. They modified the model for variance and determined correct buffers with standard deviation.

Problem Definition 2: Not Knowing the Risk of a Stockout

The team couldn’t determine how likely it was that demand would outstrip supply on any given day.

Solution:

Employing the CDF, they determined P(Demand ≤ 70 units) = 0.85. This implied a 15% chance of stockout with 70 units. Safety stock was increased to achieve a desired service level of 95%, which meant that P(Demand ≤ Stock) ≥ 0.95.

Issue 3: Generic Plan for All Products

All SKUs were being controlled with a single inventory rule, regardless of their sales fluctuation.

Solution:

The firm personalised its ordering policies using PMF and variance analysis of various products. For the low-variance items more tight stock control mechanisms in use and for the high-variace items a larger buffer.

Conclusion

By using the technology of random variables, PMF, CDF, expectation and variance, TechMart has switched its inventory system from reactive to proactive which makes decisions based on data. The outcome: stockouts were reduced 22 percent, overstocking decreased 18 percent and customer satisfaction improved across the board.

Stat for BM_Unit 7_V3.docx

Statistics for Business Managers_MBA_2

Statistics for Business Managers_MBA_2

ATLAS SkillTech University

Document Details

Submission ID

trn:oid:::3618:127363530

Submission Date

Feb 2, 2026, 1:42 PM GMT+5:30

Download Date

Feb 2, 2026, 3:55 PM GMT+5:30

File Name

Stat for BM_Unit 7_V3.docx

File Size

29.1 KB

15 Pages

3,209 Words

17,432 Characters

2% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 30 words)

Match Groups

1 Not Cited or Quoted 2%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

2% Internet sources
1% Publications
2% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **1** Not Cited or Quoted 2%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

2% Internet sources
1% Publications
2% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1 Internet

datapragmatist.com **2%**

Unit 7: Discrete Probability Distributions

Learning Outcomes

1. Understand the concept and conditions of the Binomial distribution and identify suitable scenarios for its application.
2. Apply the Binomial probability formula to calculate the likelihood of success/failure outcomes over multiple trials.
3. Describe the properties and assumptions of the Poisson distribution and differentiate it from the Binomial model.
4. Solve practical problems involving event occurrence over time or space using the Poisson distribution.
5. Analyze the relationship between Binomial and Poisson distributions, especially in cases of approximation.
6. Use distribution-based models to support business decisions under uncertainty in areas such as quality control, logistics, and service management.
7. Interpret results from distribution-based models to evaluate risk and optimize resources in real-world business contexts.

Content

7.0 Introductory Caselet

7.1 Binomial Distribution

7.2 Poisson Distribution

7.3 Summary

7.4 Key Terms

7.5 Descriptive Questions

7.6 References

7.7 Case Study

7.0 Introductory Caselet

The binomial distribution is a basic probability distribution that models the number of successes in a fixed number of independent trials, each with a constant probability of success.

It is frequently applied in business, quality control, marketing and finance to predict the likelihood of an occurrence where results are dichotomous in nature (yes/no, pass/fail, purchase/no purchase etc.).

7.1.1 Concept of Binomial Distribution

A random variable X is said to be have a Binomial distribution if:

• The trial is composed of n experiments

• Each trial has two, equally likely outcomes:success (with probability p)or failure (withprobability q = 1 −p).

• The probability of success (p) is constant over trials

• Let X be the random variable "number of successes in n trials"

Binomial Probability Formula:

The probability of obtaining exactly k successes in n trials is:

$P(X = k) = C(n, k) \times p^{k} \times q^{n-k}$

Where:

• $C(n, k) = n! \div [k! \times (n - k)!]$ (combinations)

• p = probability of success

• q = 1 − p = probability of not succeeding

• k = number of successes ($0 \le k \le n$)

Example Scenario:

A marketing email is delivered to 5 end users. If the probability that a person opens the e-mail, regardless of origin, is 0.4, what is the probability that exactly two people do so?

• The value of (n, k,p,q) used is 5, 2, 0.4 and 0.6 respectively.

Apply the formula:

$P(X = 2) = C(5, 2) \times (0.4)^{2} \times (0.6)^{3}$

$= 10 \times 0.16 \times 0.216 = 0.3456$

Therefore, since 34.56% converted about value of 2 or converted exactly 2 customers opened the email.

7.1.2 Properties of Binomial Distribution

The Binomial distribution has several important characteristics which determine its characteristics and relevance in business and analytics.

Mean (Expected Value):

$E[X] = n \times p$

This gives us the mean number of expected successes in n trials.

Variance:

$Var(X) = n \times p \times q$

The variance is a measure of how spread out the distribution is around the mean.

Shape of the Distribution:

- The distribution is symmetric for $p = 0.5$
- It is shifted to the right when p 0.5: It is left skewed

Range of Values:

The random variable X(number of success) can be 0,1,......n.

Additivity:

If X_1 $B(n_1,p)$ and X_2 $B(n_2,p)$, that are independent, then:

Garhanick85: If $X= X_1+X_2$ ~Binomial(n_1+n_2,p) Table3_TestWeights/BinomNMOV (navigation move)random rounds to decision for a binonialexee^= k+X....

We can use this property to consolidate the result of two binomial experiments which are independent and have the same probability of success.

Limiting Behavior:

- We can approximate the Binomial distribution with Poisson distribution for large n and small p.
- As n grows and p doesn't, the distribution becomes more normal since by the Central Limit Theorem.

Did you know?

If the probability of success (p) is exactly 0.5, the Binomial distribution will be perfectly symmetric—like a mirror reflected on itself about its mean. Yet a very slight change in p will make it skewed: to the right when p 0.5. This has influence on how to interpret decisions while modeling success/failure scenarios.

7.1.3 Calculation of Binomial Probabilities

We use the binomial formula to determine binomial probabilities:

$P(X = k) = C(n, k) \times p^k \times q^{n-k}$

Where:

- n = number of trials
- k = number of successes
- p = probability of success
- \(q = 1− p \) is the probability of failure.
- C(n, k) = n! ÷ (k! × (n − k)!)

Step-by-Step Example:

Problem:

A good has a 70% chance to pass the test. If 5 items are selected, the probability for the selection of exactly 4 acceptable product out of 5 is:

Solution:

- n = 5
- k = 4
- p = 0.7
- q = 0.3

P(X = 4) = C(5, 4) × (0.7)^4 × (0.3)^1

= 5 × 0.2401 × 0.3 = 0.36015

8 The probability of exactly 4 products passing inspection is 36.015%.

Cumulative Probabilities:

For P(X ≤ k) (that is, atmost k sucesses), add the probabilities:

P(X ≤ 2) = P(0) + P(1) + P(2)

This is relevant for risk, service guarantees and buffer stock decisions.

Activity 1

Title: The prediction of election results by binomial distribution

Instruction to Student:

For instance, if you were to send 1000 customers a promotional email and the likelihood of a customer reacting on your emails was 4%, based on that historical data.

Determine the expected number of customers who will take up the offer \(E[X]\).

} Using the fact that 8 is a multiple of 4, calculate the probability of exactly 40 responses from customers.

Use an Excel or a scientific calculator to calculate the following:

o $P(X = 35)$

o $P(X \leq 40)$

Based on your results, discuss if the campaign can expect to reach its target of at least 40 responses.

Submit your answers and a brief business interpretation (150 words).

7.1.4 Applications of BD in Business

The binomial distribution applies to business situations where there is a yes/no outcome repeated over many trials under identical conditions. Common applications include:

Quality Control:

— Estimating the probability of defective products in any one batch

• Predicting the outcome (pass or fail) of inspection processes

Marketing Analytics:

• Estimating how many customers will react to a campaign

• Estimating the likelihood of x email opens or click throughs

Human Resources:

• Assessing the results of employee performance appraisals (e.g., number meeting a success standard)

• Attrition modeling (e.g., no. of resignations in a department)

Customer Behavior Forecasting:

• Counting purchases of a loyalty program group

• Evaluating the probability that a specific number of users place an order

Financial Risk Assessment:

• Estimating the amount of loan defaults in a portfolio

• Estimating the success probability of funding requests

7.1.5 Limitations of Binomial Distribution

Despite its strength, the binomial model is not without drawbacks:

Fixed Number of Trials:

• It presumes that the n being performed is known and fixed, which may not be applicable to dynamic or open-ended processes.

Constant Probability of Success:

• Necessitates keeping p fixed throughout each experiment, which is infeasible as markets evolve, customer preferences change or learning happens.

Independent Trials:

• Consider independence of trials, not necessarily true if events are inter-related to each other (e.g., word-of-mouth in marketing, fatigue in equipment).

Only Two Outcomes:

• Is restricted to binary (success/failure) outcomes and does not accommodate multi-category results (for which one might employ the multinomial distribution or some other).

Approximation Errors:

• Calculations are difficult for large n and may require normal or Poisson approximation, in which case errors occur if the conditions are not ideal.

1 7.2 Poisson Distribution

The Poisson Distribution A discrete probability distribution which expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate, and are independent of the time since the last event.

In business operations, customer service, logistics, models of risk are used in which events occur randomly but the average over the long run is predictable.

7.2.1 Concept of Poisson Distribution

The random variable is a Poisson distributed random variable (denoted by N) if the total number of events in space, time or any other contexts under study are recorded, and if 1.

• Events occur one at a time

• The mean frequency (λ) is stationary

• The chance of two events happening in an infinitesimally small interval is vanishingly short

• Events are a result of other events occurring Independently from each other.

Poisson Probability Formula:

The event-count probability is:

$P(X = k) = (\lambda^k \times e^{-1}) \div k!$

Where:

• X = number of occurrences

• k = particular number of occurrences (k = 0, 1, 2,...)

• λ = average no of observations in the interval

• e ≈ 2.718 (Euler's number)

Example:

For a website that receives on average 3 customer queries per hour, how likely is it to have exactly 4 queries in an hour?

• λ = 3, k = 4

$P(X = 4) = (3^4 \times e^{-3}) \div 4!$

$= (81 \times 0.0498) \div 24$

≈ 0.1681

So there is a 16.81% probability that precisely 4 requests come per hour.

7.2.2 Properties of Poisson Distribution

The Poisson has some special attributes which make it different from the Binomial and other discrete distributions.

Mean and Variance:

The mean and variance of a Poisson distribution are also equal to λ.

• $E[X] = \lambda$

• $Var(X) = \lambda$

This actually makes the standard deviation easy to compute:

$SD(X) = \sqrt{\lambda}$

Discrete and Infinite Support:

• The Poisson distribution is discrete and reals for non-negative integers only: 0, 1, 2, 3 ...

Skewness:

•The distribution is skewed to the right, particularly if λ is small

• For lambda > sth like 10 the distribution is about symmetric

Additivity Property:

If X_1 Poisson(λ_1) and X_2 Poisson(λ_2),andX_1, X_2 are independent of each other:

$X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$

This is applicable in the case of total events being the sum of several independent sources.

Approximation of Binomial:

A Poisson distribution can approximate a Binomial distribution when:

• n is large ($n \geq 30$)

• p is small ($p \leq 0.1$)

• The product $\lambda = n \times p$ is medium

Memory lessness (for Exponential, not Poisson):

Note: The Poisson counts the number of, say, observations, but time between events is modelled by an Exponential distribution, which is memoryless. This is typically Shama lev's Poisson.

No Upper Limit:

There is no upper limit to the number of events—unlike Binomial (which has n trials) the Poisson permits an infinite number of events.

7.2.3Calculation of Poisson Probabilities

We calculate probabilities with the Poisson distribution by using this formula:

$P(X = k) = (\lambda^k \times e^{-l}) \div k!$

Where:

• X = number of occurrences

• k = exact number of events (non-negative integer)

• λ = the average number of events in a given range (mean)

• e ≈ 2.718

Step-by-Step Example:

Example: A helpdesk gets a mean of 2 calls per minute. Suppose the exact number of calls received in a selected minute is 3, what's the probability?

• $\lambda = 2$

• $k = 3$

$P(X = 3) = (2^3 \times e^{-2}) \div 3!$

$= (8 \times 0.1353) \div 6$

$= 1.0824 \div 6$

≈ 0.1804

So there is a 18.04% probability getting exactly three calls in one minute.

Cumulative Poisson Probabilities:

You can get $P(X \leq k)$ by adding up the individual probabilities:

$P(X \leq 3) = P(0) + P(1) + P(2) + P(3)$

It's typically done with a statistical table or program like Excel (=POISSON. DIST(k,?, TRUE)), R or Python.

7.2.4 Business Applications of Poisson Distribution

The Poisson distribution is employed to model events occurring in a count form, independently and randomly distributed over time, space or volume. It is widely applied in various industry to forecast the demand, operation load and service capacity.

Key Business Applications:

Customer Support & Call Centers:

• Estimating mean calls or tickets by time period

• Assists in agent staffing, queue management and reducing wait times

Logistics & Delivery Services:

• Simulating demand for delivery requests or shipping errors on a daily basis

• Applied to resource allocation, fleet optimization and inventory planning

Retail Operations:

• Predicting walk-in customers per hour

• In scheduling shifts and point of sale employment

Manufacturing & Quality Control:

• Defects per unit of measurement or production

• Assists in keeping quality levels 6 sigma and in control

Website & App Analytics:

• Approximating the rate at which users will click or log in

• Application for capacity planning and per second performance monitoring

Insurance:

• Modeling infrequent events like claim or natural disaster arrivals

• Used in risk pooling and premium pricing.

Activity 2

Title: Poisson Modeling for Staff Optimization in a Call Center

Instruction to Student:

You have the following values:

A phone call center averages 6 calls every 10 minutes.

Suppose the number of calls is Poisson distributed with mean λ = 6 .

Calculate the probability of receiving:

o Exactly 5 calls

o More than 8 calls

You can use Excel, or a calculator which carries Poisson regression functionality (e.g., =POISSON. DIST(x, λ, FALSE) in Excel).

2 agents are able to provide the services in peak hours, if everybody of them completed 3 calls in 10 minutes.

Prepare a brief analysis for whether the current staffing is satisfactory and identify an”ideal' type of staffing that can be created using this source.

659

7.2.5 The Relationship between Poisson and Binomial Distributions

Even though Poisson and Binomial distributions are not the same, in some situations Poisson can be used to approximate the Binomial distribution.

When can Poisson Approximate Binomial?

If:

• The number of experiments n is large (for example: $n \geq 30$)

• The success probability p is low (e.g., $p \leq 0.1$)

• $\lambda = n \times p$ is such that the product of n and p is moderate.

Then:

Binomial(n, p) ≈ Poisson(λ)

Why is this useful?

• Binomial probability is difficult to calculate for large n

• Poisson is easier to calculate and will give fairly accurate results when the above conditions are satisfied

Example:

If 1% of a total order of 500 light bulbs are defective, what is the probability that exactly three will be defective?

• Binomial: n = 500, p = 0.01 → $\lambda = n \times p = 5$

• Use Poisson(λ = 5):

$P(X = 3) = (5^3 \times e^{-5}) \div 3! = 125 \times 0.0067 \div 6 \approx 0.1396$

This is a lot easier than one would try to apply the binomial formula to large factorials!

Key Differences Recap:

Feature Binomial Poisson

Trials Fixed number (n) Not fixed

Outcome per trial Success/Failure Number of events

Probability (p) Constant Not needed (only λ)

Application Repeated trials Events over time/space

Close approximations N/A Approximates the Binomial distribution for large n and small p.

Did you know?

The Poisson distribution was created as an approximation to the Binomial with vanishingly small p. It's particularly valuable in fields such as nuclear physics, insurance and logistics, where events like radioactive decay or rare claims are predictable but have low probabilities.

Knowledge Check

Q1.

Which one of the following is not an assumption of the Binomial distribution?

A) Fixed number of trials

B) The experiment is two-sided.

C) Success probability varies for each trial

D) Trials are independent

Q2.

If X ~ Binomial(n = 10, p = 0.3), what is E(X)?

A) 3

B) 7

C) 0.3

D) 10

Q3.

Which of is described by a poisson?

A) Rolling a die 20 times

B) Another great approach of predicting is how many sales out of 50 calls?

C) Recording the number of customers who arrive in 30 minutes

D) Predicting the percentage of response to solicitations during a campaign.

Q4.

Which of the following is not true about Poisson distribution?

A) It is a distribution characterized by two parameters: n also known as the shape parameter, and p.

B) It is applied for in modeling fixed trial experiments

C) Its mean as well as variance is λ

D) It is only for continuous data.

Q5.

Under what condition is a binomial distribition approximately equal to a poission distribution?

A) p > n assertNotNull To study these asymptotic properties of A and D we start with the case: p >> n (here also): From this relationship we can conclude, for example if n = 1/2p that the above expression increase at rate of log(p)4 which is very fast.

B) Average-case (n > > p) and consisting of many predictors

C) When λ = 0

D) When trials are dependent

Answer Key

C – Binomial with fixed probability of success

A- E[X] = n p = 10 0.3 = 3

C – Poisson use cases количество событий за интервал времени

C – Mean = Variance = λ in Poisson's 33.333311et al and C.Earn a certificate of completion close What is the mean of a coupon collector number, which involves n coupons?

B – Poisson: applies to Binomial when n is large, p is small

7.3 Summary

To sum up, the current module discussed two major discrete probability distributions, including Binomial and Poisson, and their significance for modeling and decision-making under uncertainty. While Binomial is used when the number of identical and independent trials is known in advance and the probability of success in each trial remains constant, e.g., n cases of rejection rates or responses to a particular HR treatment, Poisson is employed for the number of occurrences of an event in a fixed interval of time or space, . Therefore, Poisson is suitable for defects, rare events, or failures, especially when the sample size is not fixed or a trial's probability is very low. The module highlighted the patterns between the distributions, suggesting that Poisson can replace Binomial when the sample size is large, i.e., . To wrap things up, both distributions indicate the primary methods to analyze business events when discrete random variables refer to the same trial or time:

Explain what is the Poisson distribution and how it differs from the Binomial distribution.

A call center averages 5 calls per minute. Apply Poisson distribution to determine the probability of 3 calls per minute.

Compare and contrast the assumptions and applications of Binomial and Poisson distributions.

When can we approximate Binomial distribution by Poisson distribution? Provide an example.

Explain the significance of mean and variance on both distributions and for business forecasting.

How is Poisson distribution used in service industry planning?

Can you provide examples from reality in which incorrect distributional assumptions might result in flawed business decisions?

7.6 References

• Ross, S. (2014). Introduction to Probability Models. Academic Press

• Walpole,R.E., Myers, R.H., and Ye,K. (2017). Probability and Statistics for Engineers and Scientists. Pearson

Gupta, S. C., & Kapoor, V. K." (2020). Fundamentals of Mathematical Statistics. Sultan Chand & Sons

• Larson, R., & Farber, B. (2014). Elementary Statistics: Picturing the World. Pearson

• Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2019). Statistics for Business and Economics. Cengage

7.7 Case Study

Caselet Title: “Lines, Lattes and Likelihood: A Café’s Data Driven Staffing Model”

Background

BrewBox Café, a rapidly expanding urban coffee chain, had an employee scheduling problem. The lines stretched so far, customer complaining of added wait time on some days; while employees were idly waiting on others. There was no accurate predictive system about customer flow to the management.

They hired a data analyst, Riya, who jumped in and started to scrape hourly customer-arrival data. Over two weeks, she reckoned the café averaged 10 customer arrivals per hour.

Problem Statement 1:

How can we represent the amount of customer arrival in terms of hour?

Solution:

Riya called it a Poisson process. As arrivals were in-dependent and occurred at a constant average rate, we modeled the number of customers who enter per hour using Poisson distribution with λ =10.

Problem Statement 2:

b) What is the probability that more than 12 customers arrive in an hour?

Solution:

Using Poisson distribution:

- $P(X > 12) = 1 - P(X \leq 12)$
- Riya worked with Excel (=1 – POISSON. DIST(12, 10, TRUE))
- Correct answer: ≈ 0.263, or a 26.3% chance

Management added an extra barista on shifts that were expected to be busier.

Problem Statement 3:

Why customer participation in loyalty programs could be modeled?

Solution:

Twenty out of 50 people who read it signed up. The problem had a fixed number of trials (n = 50) and binary responses (signup or not), thereby representing a Binomial setting.

With p = 0.4, they can forecast the probability of hitting a desired number of signups for upcoming campaigns.

Conclusion

Through incorporating Poisson for arrival prediction and Binomial for sign-up modeling, BrewBox crafted a data-informed signing plan... which led to:

- 15% decrease in customer wait time
- 12% increase in loyalty sign-ups
- Increased operational efficiency at the best level of mannedness

Stat for BM_Unit 8_V3.docx

Statistics for Business Managers_MBA_2

Statistics for Business Managers_MBA_2

ATLAS SkillTech University

Document Details

Submission ID

trn:oid:::3618:127363531

Submission Date

Feb 2, 2026, 1:42 PM GMT+5:30

Download Date

Feb 2, 2026, 3:56 PM GMT+5:30

File Name

Stat for BM_Unit 8_V3.docx

File Size

29.5 KB

16 Pages

3,696 Words

21,413 Characters

0% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 30 words)

Match Groups

0 Not Cited or Quoted 0%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

0% Internet sources
0% Publications
0% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **0** Not Cited or Quoted 0%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

0% Internet sources
0% Publications
0% Submitted works (Student Papers)

Unit 8: Continuous Probability Distribution

Learning Outcomes

1. Explain the characteristics and assumptions of the Normal Distribution and understand its relevance in real-world data analysis.

2. Interpret the shape and properties of the Standard Normal Curve, including symmetry, bell-shape, and area under the curve.

3. Convert raw scores into Z-scores and use them to compare values across different normal distributions.

4. Use Z-tables to compute probabilities and cumulative areas under the standard normal distribution curve.

5. Apply the normal distribution to solve business problems involving inventory control, quality control, demand forecasting, and risk analysis.

6. Distinguish between normal and non-normal data in business contexts, and understand when normal approximation is appropriate.

7. Evaluate how the concepts of mean and standard deviation influence decision-making in various business scenarios involving uncertainty.

Content

8.0 Introductory Caselet

8.1 Normal Distribution

8.2 Standard Normal Curve (Z-Scores)

8.3 Applications in Business Decisions

8.4 Summary

8.5 Key Terms

8.6 Descriptive Questions

8.7 References

8.8 Case Study

8.0 Introductory Caselet

"The Bell Curve and the Bonus Dilemma

Background:

Ananya is head of HR at NovusTech Solutions, an IT start-up with 800 employees. She is confronting a time-honored yet delicate challenge of year-end reviews: the distribution of performance-based bonuses.

Management wants to use a bell curve model where the top 10% of employees get generous bonuses, the middle 80% moderate ones and the lowest 10% zero. "It's the way that we guarantee equity and competition," he says.

But Ananya is uneasy. Some departments have pockets of high performers, while others exhibit natural variation. She gets together with Vikram, the company's data analyst, to get more scientific about it.

Vikram teaches her about the model of normal distribution. He describes how it can be used to identify deviations from the median, standardize scores across departments and recognize outliers. "We don't even want to be just purely blindly superimposing a bell curve," he says, "we actually want to see whether the data behave in such a manner."

They put performance scores through statistical software and turn them into Z-scores, which give you an idea where everyone in the company sits compared to their team average. Contrastingly, some teams display skewed data, raising questions about the fairness of fixed percentage standards.

With this knowledge, Ananya presents the leadership team with a data-based and department-aware bonus strategy — one that is based on standard deviation, probability, and fair play (and so much more) than just policy.

Critical Thinking Question:

Do companies need to take for granted that how good their employees are, what product types their customers buy, or how sales figures look like are all normally distributed? Fairness and accuracy in business decisions are affected when the bell curve is incorrectly applied.

8.1 Normal Distribution

normal distribution The most important probability distribution in statistics. It captures the manner in which the values of a random variable spread around its mean, and is at the heart of numerous statistical techniques such as hypothesis testing, control charts and forecasting.

8.1.1 Concept of Normal Distribution

The standard normal distribution curve is a continuous bell-shaped curve which gives the probable likelihood of any value at any point for most natural, social and business phenomena.

A variable is normally distributed if:

• It is continuous, not discrete

• It has a balanced distribution of values around the central mean

• The values occur less frequently as you move away from the mean.

This distribution is characterized by two parameters:

• μ (mu) = the mean or central value

• σ (sigma): the standard deviation. It measures how much the values are spread Where x is a variable of type list that contains some generic elements and abs(xi)−(sum(x)/n) gives us absolute value of xi minus average value sum(x) divided by number of elements n in list x.

Probability Density Function (PDF)

The equation that represents the normal distribution is:

$f(x) = (1 / (\sigma \times \sqrt{2\pi})) \times e^{(-(x - \mu)^2 / 2\sigma^2)}$

Where:

- x = value of the random variate

• μ = mean

• σ = standard deviation

• e = Euler's number ≈ 2.718

• π = Pi ≈ 3.1416

The area under the curve between two points is the probability that a value lies within that interval.

Examples in Business:

• Heights, weights, and IQ scores

• Lead Time or Lag Time in order fulfillment

• Customer satisfaction ratings

• Stock returns (approximation)

• Measured quality control (e.g. length, weight, voltage)

8.1.2 Properties of Normal Distribution

The normal distribution has several very nice properties:

Bell-Shaped and Symmetrical

• The curve is symmetric about the mean μ.

• Which means: Mean = Median = Mode

Total AUC = 1

• The curve fills the entire probability space

Areas under the curve are used to display the probabilities

Empirical Rule (68-95-99.7 Rule)

One convenient property to the normal distribution is that it has a predictable spread:

• 68% of readings are within ±1σ from the mean

• 95% lie within ±2σ

• 99.7% lie within ±3σ

This type of comparison allows businesses to see if a value is on the high or low side for its typical behavior.

Asymptotic to the X-Axis

• The tails of the curve veer away from the x-axis

• One can get an extremely-angled or close-to-straight-line result, but very infrequently.

Defined by Two Parameters

• The form of the normal distribution is determined entirely by:

o μ (mean) = controls the location of the middle of the curve.

o σ (sigma/standard deviation): affects spread or dispersion

Standardization is Possible

• As I mentioned above, no matter what the normal distribution before, Once converted to Z-score with the following formula, it becomes a standard normal distribution (Z-distribution) formally.

$Z = (X - \mu) \div \sigma$

This allows for comparison of different normal distributions.

Applicability to Many Fields

• The normal distribution is so important to inferential statistics because:

o We take it for granted in the central limit theorem

o Several statistical tests (such as t-test, ANOVA) assume the data to be normal

Did you know?

The true correctly distributed mean, median and mode are equal in a totally perfect normal distribution. This symmetry is an important property to ensure that the normal curve looks like a bell. In the real world, if these three quantities are not equal (or nearly so), then there may be skewness present in the data – meaning that it might not follow a normal distribution and may require transformations prior to using certain statistical methods.

8.1.3 Importance of Normal Distribution in Statistics

Here are a few Salient reasons, why that is the case: The normal distribution is not just the work-horse of (classical) statistics – used heavily for making inference from sample information to population understanding.

Central Limit Theorem (CLT)

The Central Limit Theorem tells us that the distribution of the mean of any independent, random sampling will be normal for large enough samples—no matter what your original population may look like.

• As a result, the normal distribution is also an assumption of last resort in statistical inference.

• It allows the use of Z-tests, t -tests ANOVA etc even for non-normal populations (when n is large).

Basis for Statistical Inference

Normality is an assumption of many inferential methods (e.g., confidence intervals, hypothesis testing, regression analysis) including:

• The underlying data

• The residuals/errors in a model

• Sampling distributions

Simplifies Probability Calculations

The fact that the normal can be accurately described by just two parameters (μ and σ) enables analysts to: 1.

• Compute probabilities

• Predict ranges

• Identify unusual or outlier events

This is an important part of decision making, risk assessment and predictive modeling.

Standardization and Comparison

Any normal distribution can be transformed into the standard normal distribution (Z-distribution) by this formula:

$Z = (X - \mu) \div \sigma$

It makes it easy to compare values from disparate datasets (in different units and/or scales).

On the foundation of the control charts in quality-management

Prediction limits in SPC cateants are formed on the basis of normal distributions, and so do control charts such that they react to process variation.

Supports Many Theoretical Models

• The normal distribution acts as a theoretical base for analysis of errors, financial models (under assumptions), and signal processing.

8.1.4 Applications of Normal Distribution in Business

Normal distribution also has significant presence in the business analytics and operations domain assisting in prediction, decision-making and assessing performance.

Quality Control and Six Sigma

• Used to track process variation

• Assists in determining if results are within acceptable limits

• Used in defect rate prediction, process capability analysis

Demand Forecasting

• Several demand patterns (if summed up) resemble a normal distribution

• A value used in determining safety stock, or re-order point and inventory bufferories

Human Resource Analytics

• Normal curves and testing: The performance levels of employees, test scores, and appraisal ratings are often distributed normally (Gullikson, 1987).

• Makes allowance for the use of bell curve performance management (where applicable)

Finance and Risk Management

• We model asset returns and portfolio risks (under simplifying assumptions) by normal distribution

• VaR, credit risk and pricing models start from normal approximations

Marketing and Customer Insights

• Customer ratings and feedback (in the aggregate) are observed to be normally distributed •Ideal Social Choice captures incentives for consumers with varying utility functions.

• Assist in segmenting customers, and benchmarking for satisfaction or customer experience scores

Operations and Logistics

The service times, cycle times of production, and delays in processing are frequently modeled as being normal.

• Aids in controlling bottlenecks, adjusting lead times and enforcing service level agreements

In other words, the normal distribution is more than a theoretical curiosity—it is an useful analytical tool for quantitative decisioning in just about every aspect of business.

8.2 Standard Normal Curve (Z-Scores)

The normal distribution with mean 0 and standard deviation 1 is known as the standard normal distribution. It lets us compare values across various normal distributions, and provides for probabilities calculation with standard tables or software.

8.2.1 Standardization and Z-Scores

Normalizing is the process of transforming a value from a normal distribution into standard normal form (which are mean 0 and variance 1) with Z-score formula.

What is a Z-Score?

A Z-score is the number of standard deviations that X is from the mean of a distribution.

Z = (X − μ) ÷ σ

Where:

• Z = standardized score

• X = raw score

• μ = the mean of the distribution

• σ = standard deviation

Why Use Z-Scores?

• For normalizing across datasets

• When you want to interpret scores from various normal distributions

• To calculate probabilities and percentiles

• Outliers are defined as: Typically, Z 2

Example:

Let's say one test a class average for a test is 75 and the standard deviation is 5.

Z = (82 − 75) ÷ 5 = 7 ÷ 5 = 1.4

Interpretation: The student was 1.4 standard deviations above the mean for this class.

Characteristics of Z-Scores:

• Z = 0 → Value is same as mean

• Z > 0 → Above average value

• Z < 0 → Value is less than mean

When Z-distribution is used, its shape also would be bell-shaped and symmetric like any normal distribution.

8.2.2 Conversion of Raw Scores to Z-Scores

To compute a Z-score from an unincorporated score:

Find out the mean (μ) and Standard deviation(σ) of the data set

Step 2: Use the Z-score formula quoted above. Z = (X − μ) ÷ σ

Step 3: Look up the probability or area under of the curve associated with a Z-score using the Z-table or software (i.e., Excel, Python)

Example 1:

A filling machine produces bottles of coke which have a mean weight of 500 g and s.d. 20g What is the Z score for a bottle that weighs 540 g?

Z = (540 − 500) ÷ 20 = 40 ÷ 20 = 2.0

Thats to say the bottle is 2 standard deviations above the average.

Example 2 (Business Context):

Offshore bank The following high-level examples illustrate how the addition of on-shore physical presence of offshore banks can dramatically reduce “administrative burdens” (but not planning and having proper advice with regard) to FATCA compliance. 680 if the average credit score is 690 with a standard deviation of 25:

25Z 720 690) = 35(2) Z = (720 − 690)/25 = 30/25 =1.2

The customer is better than average and it can be calculated Z-score to determine the probability of doing even better (from the right end of the Z-distribution).

Activity 1

z-score Analyzing Employee Performance I use a Z-Screen to announce my next moves.....

Instruction to Student:

You have the appraisal scores (on scale of 100) of 15 employees in a department. The average score is 72 and the standard deviation is 8.

Calculate a Z score for each employee using the formula:

$Z = (X - \mu) \div \sigma$

Identify:

a)Above +1 SD (high): employees with values higher than 1 standard deviation (SD) away from the mean.

b) Employees whose score is lower than −1 standard deviation (will be referred to as potential underperformers).

Write a brief report on the results and explain how this information could help the team leader during performance feedback meetings.

8.2.3 Using Z-Tables for Probability Calculations

Once a Z-score has been calculated, the Z-table (or standard normal table), enables us to find the probability or cumulative probability of an area under the standard normal curve up to that specific Z-score.

How to Use a Z-Table:

Calculate the Z score using the formula:

$Z = (X - \mu) \div \sigma$

Look up the Z value in the table:

o The first 2 column provide a row number (e.g., 1.3)

o The column is the second decimal place (e.g., . 04)

o The point where they intersect is $P(Z \le \text{value})$

Result is the Cumulative Probability where that is the area to the left of Z.

Example:

If Z = 1.25, the cumulative probability given in the table of Z-scores is about 0.8944.

→ This says: The probably of a value being less than 1.25 stand deviations above the mean is 89.44 percent.

Finding Right-Tail Probability:

To obtain $P(Z \ge z)$ (area above), subtract from 1:

$P(Z \ge -1.25) = 1 - 0.8944 = .1056$

Finding Probability Between Two Z-Scores:

To find the probability (or area) between two Z values: If Z is a standard normal variable, to compute \(P(a < Z < b) \), use the following syntax:Awhere "a" and "b" are the two Z-scores.

- Find $P(Z \le 1.0) = 0.8413$
- Find $P(Z \le -1.0) = 0.1587$
- Subtract: $0.8413 - 0.1587 = 0.6826$

→ 68.26% of the cases fall within ±1 standard deviation.

8.2.4 Applications of Z-Scores in Business Decision-Making

In analytics and business operations, Z-scores are anindicator of how unusual or normal a value is ina dataset. Z-scores Because Z-scores are normalized values, direct comparisons between them can be made and their probabilities and performance estimated.

Key Business Applications:

Credit Scoring and Risk Assessment

• Lender compares customer credit score to a benchmark distribution using Z-score

• Assists in identifying high-risk and low-risk borrowers.

Quality Control in Manufacturing

• Used to decide if the product measurement (dimension, weight) is within tolerance.

• Z-scores can be used to indicate failures or process deviations

Sales Performance Evaluation

• Benchmark sales reps against one another in different regions or times

• Z-scores measure how far each person is from the average, in terms of variation

Inventory Management

• Computing Safety Stock based on pooling Z-scores for Demand Variability and Service Level objectives

• Helps prevent stockouts or overstocking

Human Resource Analytics

• Appraisal scores, test results or Z-scores on employee engagement scores can be comparedGESTFORSK27Quality indicators and quality levelsYou think one level can be defined as a “quality” level.

• Helps you and your team understand who are the best (or worst) performing members within or across teams

Market Research and Survey Analysis

• See how an individual's response or a group average measures up to the general population

• Identify and correct any outliers, biases, or outlier responses

In all those cases, z-scores offer a common yardstick to gauge how much a value diverges from the norm, and whether that difference is statistically meaningful.

Did you know?

Credit scoring models use Z-scores to assess an applicant’s risk. Banks standardize their customer’s profiles including the income, past repayment history) as Z-scores and use them to decide whether a customer is likely to default compared the average possibility of acceptance. A Z-score less than −1.5 usually indicates very high risk and can result in rejection or higher interest rates.

8.3 Applications in Business Decisions

The normal distribution and its standardization, with the use Z-scores is a very powerful tool that helps to make decisions across business. From rating processes to analyzing market trends and even predicting risk, these tools assist managers in minimizing ambiguity and increasing accuracy.

8.3.1 QualityControl and Six Sigma

Statistical Process Control (SPC) and Six Sigma methodologies in manufacturing and operations are based on the normal distribution.

Applications:

• Being able to watch processes change and fluctuate using control charts

• Whether a product feature (eg size, mass) is within ideal control limits

• Detecting outliers or validating a sample size or number of defects by determining the number of standard deviations an observation is from the mean.

• Using DPMO to calculate the Z-values

Six Sigma Context:

• There are 3.4 defects per million opportunities in a Six Sigma process

• This is equivalent to a Z-score of 6, where 99.99966% of output normally fits within specification limits

Activity 2

Title: CONTROL LIMITS AND PROBABILITY OF DEFECT IN A MANUFACTURING PROCESS

Instruction to Student:

The pump filling shampoo bottles was a packaging machine. The target fill level is 200 ml, and the standard deviation of the fill volume filling is 5 ml. The company labels good-fill volume as 190 to 210 ml.

Find the Z-scores for 190 ml and 210 ml.

Look up the z value in a Z-table or with some software to get the probability that a bottle we select randomly lands between these two points.

What proportion of 10,000 bottles will not be within specification limits (i.e., defective)?

Hand in your computations and a remediation suggestion that the defect percentage could be lowered, if desired.

8.3.2 Forecasting and Risk Analysis

There is no end to the uncertainty businesses confront — in revenue, costs, demand and risks outside their control. Variability can be modelled by the normal distribution, which allows better forecasting and risk management.

Applications:

• Predicting sales or demand based on historical mean and standard deviations

• Calculating the likelihood of running over budget or behind schedule

• Measuring the risk of credit, and probability of default for loans

• Model the variance in demand for inventory and establish an appropriate safety stock level

Example:

A logistics manager employs the normal distribution to determine the probability of shipment time in excess of 5 days using historical shipping time data ($\mu = 4$, $\sigma = 0.5$). By computing the Z-score, they can assess the probability and prepare accordingly.

8.3.3 Consumer Behaviour and Marketing Research

Simply put, you can't know what to upsell or how to discount until you compare an individual behaviour pattern to the mean of a population. Apply: Normal distribution can also be used to interpret survey results, customer ratings – even behavior.

Applications:

• Studying customer satisfaction scores believed to be distributed normally

• Z score to compare an individual or segment s performance with the overall averages.

• Pinpointing radical preferences or dissatisfaction in large-scale surveys

• Finding market segments which are operationally highly aggregated compared to the mean

Example:

A brand leverages net promoter scores (NPS) of its customers and computes Z-scores to identify stores or regions where customer sentiment deviates markedly — either favourably or unfavourably.

8.3.4 Financial Modelling and Decision Support

In finance and analytics, normal distribution models are used to determine the variability of returns, value at risk, and for scenario planning.

Applications:

• Portfoilo theory projects expected returns and risk (standard deviation) by assuming returns are normal0y distributed

• Credit scoring models and fraud detection, use z-scores.

• The bell curve is frequently used to model default and loss probabilities

• Recommend actions with a risked return on investment model based on probabilities of business outcomes

Example:

An investment manager estimates the probability of a portfolio losing over 10% in any quarter, assuming returns are normally distributed. This allows them to evaluate the downside.

In short, the normal distribution allows for measuring uncertainty, comparing between variables and probabilistic reasoning - all critical requirements in today's data-driven business decision tools.

Knowledge Check

Q1.

Which of these is not a characteristic of the normal distribution?

A) It is symmetrical and bell-shaped

B) Mean, median and mode will all be same

C) Total area under the curve ±3

• Six Sigma: A quality management methodology involving implemen- tation of deviation-reduction methodologies

• PDF (Probability Density Function): A function that describes the form of the normal distribution curve

8.6 Descriptive Questions

13 Define normal distribution and illustrate the important features of normal distribution?

What does a standard deviation mean in a normal curve?

Compute the Z-score formula and elucidate on its components.

What does a z-score mean in the standard normal distribution?

What is the 68–95–99.7 rule and what does it involve in business analytics?

Describe how to use the z-table to find cumulative probabilities.

Give an example of a business situation in which the use of Z-scores can aid in making more informed decisions.

What is relation between Six Sigma and normal distribution?

Explain the role of the normal distribution in forecasting and inventory planning.

Explain the assumptions and constraints of a normal distribution in business applications.

8.7 References

•Gupta, S.C. And Kapoor, V.K. (2020). Fundamentals of Mathematical Statistics. Sultan Chand & Sons

• *Walpole, R.E., Myers, R.H., Myers, S.L. and Ye, K. (2017). Probability & Statistics for Engineers and Scientists: Summary. Pearson

• Anderson, D.R., Sweeney, D.J., Williams, T.A. (2020). Statistics for Business and Economics. Cengage

• Montgomery, D.C. (2019). Introduction to Statistical Quality Control. Wiley

• Keller, G. (2018). Statistics for Management and Economics. Cengage Learning

• Field, A. (2018). Re: Discovering Statistics with IBM SPSS Statistics_helper and data file.djangpan1130812909 Mar 25, 2011 8:45 AM (in response to Claire) The web page mentioned in the book is not active anymore. Sage Publications

8.8 Case Study

Caselet Name: “Stock and Service in Balance: A Steady State Probabilistic Approach to Demand Planning”

Background

“TrendyGear,” a popular fashion accessories manufacturer and retailer, was confronted with a common problem: overstocking (understocking) the best-selling SKU in different stores. The planners were using average weekly demand as replenishment coverage, which was causing the wrong results of out-of-stocks in urban stores and excess inventory in suburban stores.

Problem

It became apparent to the team that just using the mean demand was not right, they’d left out three things. Why did this volatility matter? Because it meant that merely relying on averages would raise the risk of carrying inventory.

Solution

A new data analyst, Arjun, suggested planning more precisely based on the normal distribution and Z-scores.

He aggregated weekly sales data and measured the mean (μ) and standard deviation (σ) for each store.

For each item, the Z-score for a specified service level (e.g., Z = 1.28 for 90% service) was computed.

Then, he applied the formula of safety stock:

Safety Stock = Z × σ

The overall stockpile of each item was defined as follows:

Level of inventory = Average + Safety Stock

This resulted in higher service levels on lower overall inventory.

Result

- Stockouts reduced by 30%
- The inventory holding cost has been 18% lower
- Subsequently, the model was operationalised in 40+ stores based on sales data through an integrated system.

Conclusion

By leveraging normal distribution models in their supply chain, TrendyGear achieved maximum customer delight and operational efficiency. The trick was to go from average calculations to probability-based plans using a range of statistical instruments.

Stat for BM_Unit 9_V3.docx

Statistics for Business Managers_MBA_2

Statistics for Business Managers_MBA_2

ATLAS SkillTech University

Document Details

Submission ID

trn:oid:::3618:127372157

Submission Date

Feb 2, 2026, 3:55 PM GMT+5:30

Download Date

Feb 2, 2026, 3:57 PM GMT+5:30

File Name

Stat for BM_Unit 9_V3.docx

File Size

32.8 KB

20 Pages

4,385 Words

23,766 Characters

0% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 15 words)

Match Groups

0 Not Cited or Quoted 0%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

0% Internet sources
0% Publications
0% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **0** Not Cited or Quoted 0%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

0% Internet sources
0% Publications
0% Submitted works (Student Papers)

Unit 9: Hypothesis Testing

Learning Outcomes

1. Explain the rationale and step-by-step procedure for conducting hypothesis testing in statistics.

2. Differentiate between null (H_0) and alternative (H_1) hypotheses and understand their roles in decision-making.

3. Identify and interpret Type I and Type II errors in hypothesis testing and their business implications.

4. Apply the Z-test for large-sample hypothesis testing involving means and proportions.

5. Apply the t-test for small-sample testing, including one-sample, two-sample, and paired sample cases.

6. Evaluate the p-value approach to decision-making in hypothesis testing.

7. Use hypothesis testing in business contexts such as quality control, marketing surveys, HR analytics, and financial analysis.

Content

9.0 Introductory Caselet

9.1 Rationale & Procedure for Hypothesis Testing

9.2 Errors in Hypothesis Testing

9.3 Z-Test

9.4 t-Test

9.5 Practical Business Applications

9.6 Summary

9.7 Key Terms

9.8 Descriptive Questions

9.9 References

9.10 Case Study

9.0 Introductory Caselet

"The New Ad Campaign: Belief v. Evidence"

Background:

The marketing department at Fresh Fizz Beverages just introduced a campaign through social media that aims to reach 18- to 25-year-olds. The marketing director is convinced that the campaign has led to greater brand recognition and sales.

But when the finance team analyzed the sales for the columns first month, they didn't see a clear jump. Some retailers grew, and some did essentially what they were doing. The campaign is working —we must have sold more!" the director insisted.

The dispute received a ruling from Priya, a data analyst. "Beliefs are not enough," she said. We need to figure out if there is a real increase in sales, or whether it's just random variation."

She brought the idea of hypothesis testing to:

- Point is null hypothesis (H0) and campaign has no effect on sales.
- The alternative hypothesis (H_1): The campaign results in more sales.

To see whether the difference in average sales was statistically significant, Priya used a t-test that compared numbers from stores exposed to the campaign and those not.

The result? The difference was not statistically significant at the 5 percent level, which is to say that there is insufficient evidence to reject the null hypothesis. The management knew that they couldn't just take for granted that the campaign functioned, but needed evidence to show that it had impacts.

Critical Thinking Question:

17) How can managers use hypothesis testing as a tool to prevent themselves from relying on belief or wishful thinking, rather than the evidence they gather and analyze?

9.1 Rationale & Procedure for Hypothesis Testing

Rationale

Statistical inference is a process in which conclusions are drawn (or decisions made) about the population based on sample data. It offers an organized approach for:

- Challenge, test beliefs and take them to the data rather than relying on intuition.
- Separate genuine effects from random variation.
- Assess the degree of risk (inaccuracy) in decision-making.

In business, hypothesis testing is crucial for things like testing out new marketing strategies, analysing product improvements, comparing the performance of employees and validating financial models.

Procedure for Hypothesis Testing

Here is a summary of the steps involved in hypothesis testing:

State the Hypotheses

o Null Hypothesis (H_0): Denotes that no change or no effect, nothing has happened (status quo).

o Make the Alternative Hypothesis (H_1): States what is being tested (a change or a difference).

Select Significance Level (α)

o Sample choices: 0.05 (5%) or 0.01 (1%).

o Rejecting H_0 when it is true o It is the probability of committing Type I error

Choose the Appropriate Test Statistic

o Z test→ For large sample ($n \geq 30$) or Known variance Are when σ2 is known.

o t-test: Small samples ($n < 30$) with unknown variance.

o There are other tests (Chi-square, ANOVA) that can also be customized based on type of data.

Step 1: Gather Data and Calculate the Test Statistic

o Determine the test statistic (e.g., Z, t) using the sample data.

Find the Critical Value or P-value

o Critical value method: Compare the test statistic with a critical value.

o Comparison with P-value: Compare the p-value to α.

Make a Decision

o If the absolute value of z is greater than the critical value (or if $p \leq \alpha$), reject H_0.

o Otherwise, we do not reject H_0 (no substantial evidence for H_1).

Draw a Business Conclusion

o Give the practical interpretation of the conclusion. For example: "The data gives reason to believe that the process is more efficient with the new system.

9.1.1 Concept and Importance of Hypothesis Testing

Concept:

Hypothesis testing is a statistical method that is used to make decision about population parameter on the basis of sample information. It helps identify whether apparent differences or patterns are likely to be real (and potentially important) or simply random.

· A hypothesis is a statement or claim regarding a population parameter (for example, the population mean).

• The level of significance can be used as the scientific level to challenge these claims, using data.

Importance in Business and Research:

• Gets away from intuitive or assumed decision-making.

• A reply based on evidence, for or against, a claim.

• Aids the assessment of strategy effectiveness (such as a new ad campaign, training initiative or product launch).

• Lowers risk of making incorrect inferences under uncertainty.

Example:

A company says its new training program increases employees' productivity. Hypothesis testing enables management to determine whether the increase in output is statistically significant or only random variation.

9.1.2 Steps in Hypothesis Testing

There are several steps in hypothesis testing:

Formulate Hypotheses

o Null hypothesis (H_0) → The status quo or no effect.

o Alternative hypothesis (H_1) → Claim or presumed change.

Select Significance Level (α)

o Typical levels: 0.05 (5%), 0.01 (1%).

o Stands for the probability of H_0 being falsely discarded.

Choose the Test Statistic

o Z-test for large sample sizes or known variance.

o Small Sample Test on t-distribution with unknown variance.

o Other tests, i.e. chi-square, ANOVA for certain conditions.

Step 2: Gather Data and Calculate the Test Statistic

o Use the chosen test formula to determine the value of a test statistics.

Define Decision Rule

o Critical value approach or p-value approach.

Make a Decision

o Reject H_0 if evidence is strong (statistic in critical region or $p \le \alpha$).

o Fail to reject H_0 when evidence is not strong.

State the Conclusion in Context

o Relate the statistical decision to a business conclusion.

Activity 1

Titel: Evaluating the Success of a Marketing Campaign

Instruction to Student:

The management of a departmental store contends that an intensive publicity campaign launched by it has resulted in the average daily sale exceeding ₹50,000. The following are the data for 36 days at random:

- Sample mean ($\bar{X}$) = ₹52,500
- The population standard deviation (σ) is ₹6,000.
- Hypothesized mean (μ) = ₹50,000
- Significance level (α) = 0.05

Write down the null and alternative hypothesis.

Calculate the Z-statistic with:

$Z = (\bar{X} - \mu) \div (\sigma / \sqrt{n})$

Critical value of average hours spent limiting Internet use) at $\alpha = 0.05$ (two-tailed).

Decide whether to reject or fail to reject H_0.

Provide a brief conclusion (100–150 words) tying the result to how effective you think the campaign was.

9.1.3 Null and Alternative Hypotheses

- Null Hypothesis (H_0):

o Represents null or no difference effect.

o Example: “The new drug has no effect versus the old.”

• Alternative Hypothesis (H_1):

o The claim being tested; what the researcher is trying to prove.

o Example: “The new drug is more efficacious than the old one.

Types of Alternative Hypotheses:

Why One-Tailed Test? **H_1 will take you in a direction (Greater or Less).

o Ex: H_0: $\mu = 50$, H_1: $\mu > 50$.

Two-tailed test: H_1 tests for the possibility of inequality ($\neq$).

o Example: H_0: $\mu = 50$, H_1: $\mu \neq 50$.

Business Example:

H0: Call centre receives 100 calls a day (mean).

Hypothesis 1: A new system alters average calls processed per day.

9.1.4 Level of Significance and Critical Region

Level of Significance (α):

• The probability to make a mistake rejecting H_0 when it is true (frequently called Type I error).

• Common choices:

o 0.05 (5%) → Will tolerate as a 5% shot of being wrong.

o 0.01(1%) → Stricter; thus stronger evidence is needed.

Critical Region (Rejection Region):

• The range of the probability distribution if the test statistic occurs, then H_0 is rejected.

• Dependent upon the value of α and the style of the test (one-tailed or two-tailed).

Examples:

• Two-tailed Z-test at $\alpha = 0.05$ → critical values ±1.96.

• If $Z > 1.96$ or $Z < -1.96$, we reject H_0.

Business Context:

If a bank checks to see if the mean time to process a loan application is 48 hours:

• H_0: μ = 48

• Further; at α = 0.05, when Z = 2.1 (fall under critical region), the bank will believe that average time is not equal to48 hours.

9.2 Errors in Hypothesis Testing

In hypothesis testing, all judgments are based on sample information. Sampling involves uncertainty and one can always take wrong decisions. These errors are classified as Type I and Type II errors.

9.2.1 Type I Error – Concept and Implications

Concept:

Type I error is the conditional probability of your decision to reject null hypothesis (H_0), given that it is true.

• It is also referred to as false positive.

The probability of making a Type I error is denoted by the level of significance (α).

Example:

An industrial company that conducts a test of a new drug.

• H_0: The drug has no effect.

• H_1: The drug is effective.

The company denies H_0 and approve the drug but it is not effective → Type I error.

Implications in Business:

• Authorizing an ineffective drug (medical risk).

• Beginning a new project and promising higher sales that don't occur (unnecessary costs).

• Taking a machine out for repair when it does not actually need to be (lost production).

Did you know?

In medical testing, for example, a Type I error (false positive) may result in mistakenly diagnosing a patient with an illness that they don't actually have. It's why, for example, in clinical trials the significance level (α) is often not 0.05 but instead set at 0.01 – to avoid unnecessary treatment as much as possible.

9.2.2 Type II Error – Concept and Implications

Concept:

Type II error: if we take not H0 (i.e. H0 is false) and don't reject it.

• It is known as a false negative.

• The chance of a Type II error is represented as β (beta).

• Power of a test = 1 − β → probability that we will reject null when it is indeed false.

Example:

Using the same drug trial:

• H_0: The drug has no effect.

• H_1: The drug is effective.

If the company cannot reject H_0 when the drug is effective → Type II error.

Implications in Business:

• Missing out on profit (turning away a good product).

• Missing fraud in a financial deal.

• Failure to single out a defective process in manufacturing and other long-term quality problems.

9.2.3 Trade-off Between Type I and Type II Errors

• In testing a hypothesis, minimizing one type of error often maximizes another.

• Decrease α (stricter the test): Decreased chance of Type I error, increased change of Type II error.

• If we increase α (loosening the test): Reduces P(Type II error) but increases P(Type I error).

Visual Understanding:

• Type I error → false positive (claiming an effect is there when it's not).

• Type II error → false negative (fail to detect a real effect).

Business Example:

• A bank creates an algorithm to detect fraud.

o If the system is over-conservative (low α), it identifies too many false fraud cases(Type I).

o We will miss real frauds (Type II) o If it is too liberal (high α).

• Risks must be weighed by managers, depending on which error is more expensive.

9.3 Z-Test

The Z-test is generally used for a sample size greater than 30 ($n \geq 30$). If the population variance is known and the sample size is also large enough, then we use z-test. It decides if the observed sample data is sufficiently strong evidence against the null hypothesis in regard to the population parameters.

The Z- test compares observed data against a hypothesized value and relies on the standard normal distribution (Z-distribution).

9.3.1 One-Sample Z-Test for Mean

One-sample Z-test checks if the mean of a single sample differs from the population mean.

Formula:

$Z = (\bar{X} - \mu) \div (\sigma / \sqrt{n})$

Where:

- $\bar{X}$ = sample mean
- μ = mean of the population (hypothesized in H_0)
- σ = population standard deviation
- n = sample size

Hypotheses:

- H_0: $\mu = \mu_0$ (the population mean is equal to the sample mean)
- H_1: $\mu \neq \mu_0$(divergent from H0) (mean is not equal to the population mean)

Example:

The mean life of a battery is 500 hours, according to the manufacturer. A random sample of 64 batteries indicate an average life of 490 hours with a population standard deviation of 40 hour. Test the claim at $\alpha = 0.05$.

- $\bar{X} = 490$, $\mu = 500$, $\sigma = 40$, $n = 64$
- $Z = (490 - 500) \div (40 / \sqrt{64}) = (-10) \div (5) = -2.0$
- Critical Z (two-tailed, a = 0.05) = ±1.96

All because $-2.0 < -1.96$ → Reject H_0.

Conclusion? The battery DOES NOT last 500 hours.

Did You Know?

The Z-test is one of the oldest statistical tests, and it was discovered more than 100 years ago and popularized by Karl Pearson. It became a cornerstone of industrial quality control, especially in the Second World War when factories required rapid statistical tools to ensure consistent output.

9.3.2 Two-Sample Z-Test for Mean

Two means z-test It is used to test whether the means of two independent samples differ significantly.

Formula:

$Z = (\bar{X}_1 - \bar{X}_2) \div \sqrt{((\sigma_1^2/n_1) + (\sigma_2^2/n_2))}$

Where:

- $\bar{X}_1, \bar{X}_2$ = sample means
- σ_1, σ_2 = population standard deviations of the groups to compare
- n_1, n_2 = sample sizes

Hypotheses:

- $H_0: \mu_1 = \mu_2$ (difference of means equal to zero)
- which gives $H_1: \mu_1 \neq \mu_2$ (difference)

Example:

A corporation needs to compare average monthly expenditures for online and offline client.

- Online customers: $n_1 =$ 100, $\bar{X}_1$ \$250, $\sigma_1 = 40$
- Offline customers: $n_2 = 120$, $\bar{X}_2 =$ \$240, $\sigma_2 = 45$

$Z = (250 - 240) \div \sqrt{((40^2 / 100) + (45^2 / 120))}$

$= 10 \div \sqrt{(16 + 16.875)}$

$= 10 \div \sqrt{32.875}$

$\approx 10 \div 5.73 = 1.74$

At $\alpha =$ 0.05 (two-sided), critical $Z = \pm 1.96$.

1.74 < 1.96 → No reshampoos H_0 Since 1.74 < 1.96 → Fail to reject H_0.

Conclusion: No spending gap between online and offline customers.

9.3.3 Applications of Z-Test in Business

One application for the Z-test is to determine if a two-tailed p-value is less than or greater than 0.05, often in business situations with large data sets that are close to being normally distributed.

Applications:

Quality Control

o Comparing test production output to the desired standards.

Marketing Research

o Associate satisfaction scores differ significantly from industry norms.

Finance and Banking

o Average transaction sizes between 2 branches.

o Verification that Loan Approval Times comply with regulations on average.

Human Resources

o Average scores of trainees on different tests before and after the training.

Operations and Service Industry

o Review of Delivery performance vs Promised averages time.

9.4 t-Test

T Test Definition The T test is a statistical hypothesis test that is used to compare the means of two groups, one group with a small sample size and the other with a larger sample size or both groups having small sample sizes.

It is based on the t-distribution (it's sort of like a normal distribution except with 'heavier' tails since it assumes more uncertainty in small sample sizes)

9.4.1 One-Sample t-Test for Mean

When σ is unknown, the one-sample t-test compares the sample mean to a hypothesized population mean.

Formula:

$t = (\bar{X} - \mu) \div (s / \sqrt{n})$

Where:

• $\bar{X}$ = sample mean

• μ = hypothesized population mean

• s = sample standard deviation

• n = sample size

Example:

A restaurant advertises that its average delivery time is 30 minutes. For a sample of 16 deliveries, $\bar{X}$ = 28 minutes and s = 4 minutes. Test the claim at α = 0.05.

$t = (28 - 30) \div (4 / \sqrt{16}) = -2 \div 1 = -2.0$

Degrees of freedom (df) = 15

Critical t values (two-tailed, α = 0.05, df = 15) ≈ ±2.131

As −2.0 is contained in, −2.131 to 2.131 → Do not reject H_0.

Conclusions: There is no major evidence that average delivery time is not 30 minutes.

9.4.2 Two-Sample Independent t-Test

A two-sample t-test compares means from two independent groups when we do not know the variances.

Formula:

$t = (\bar{X}_1 - \bar{X}_2) \div \sqrt{((s_1^2 / n_1) + (s_2^2 / n_2))}$

Where:

• $\bar{X}_1$, $\bar{X}_2$ = sample means

• s_1^2, s_2^2 = sample variances

• n_1, n_2 = sample sizes

Example:

A firm would like to compare the productivity of workers with Tool A versus Tool B.

• Tool A: $n_1 = 12$, $\bar{X}_1 = 82$, $s_1 = 6$

• Tool B: $n_2 = 10$, \(\bar{X}{2}\) = 78 \(s{2}\)=5

$t = (82 - 78) \div \sqrt{((36/12) + (25/10))}$

$= 4 \div \sqrt{(3 + 2.5)}$

$= 4 \div \sqrt{5.5}$

≈ 1.71

At df ≈ 20, critical t (α = 0.05, two-tailed) ≈ -2.086.

Because of 1.71 < 2.086 → Can NOT reject H_0.

Summary: No productivity difference between Tool A and Tool B.

Activity 2

Title: A Study on Productivity of Two Teams-Based Companies

Instruction to Student:

Teams A and B are sales teams that boast of having the same average weekly sales. To verify this, data were obtained:

- Team A : n_1 = 10, $\bar{X}_1$ = 45 pcs., s_1 = 5
- Team B: n_2 = 12, $\bar{X}_2$ = 42 units, s_2 = 4 units

State H_0 and H_1 clearly.

Calculate the t-score using the abovementioned formula:

$t = (\bar{X}_1 - \bar{X}_2) \div \sqrt{((s_1^2/n_1) + (s_2^2/n_2))}$

What are the degrees of freedom (approximation: $n_1 + n_2 - 2$)?

Compare t you have calculated with t at α = 0.05 (two-tailed).

Write a brief conclusion (150 words): Are the two teams' productivity levels different from one another?

9.4.3 Paired Sample t-Test

The paired t-test tests the mean of two related groups (two samples) to determine if there is a significant difference between their averages.

Formula:

$t = \bar{d} \div (s\_d / \sqrt{n})$

Where:

- $\bar{d}$ = mean of the differences ($X_1 - X_2$)
- s_d = standard deviation of the differences
- n = number of pairs

Example:

An HR manager wishes to determine whether training increases average test scores of employees.

- Pre-training: [60, 62, 65, 70, 68]
- Post-training: [65, 66, 68, 74, 72]
- Distances (d): [5,4,3,4,4] → $\bar{d}$= 4.0; s_d ≈ 0.7

$t = 4 \div (0.7 / \sqrt{5}) = 4 \div (0.31) \approx 12.9$

Illustrative critical t (df = 4, alpha =.05, two-tailed) ≈ 2.776.

Because 12.9 > 2.776 → Reject H_0.

Conclusion: The education course resulted in a significant increase of the test scores.

Did you know?

The paired t-test is the statistical engine that drives a variety of "before and after" studies — all the way from assessing the effects of a training program to gauging weight-loss diets. By dealing in differences, rather than absolute scores, it cancels out a lot of the normal variation between people.

9.4.4 Business Applications of the t-Test

The t-test is used so extensively in business because we often only have limited data and no information about the variance.

Applications:

Marketing:

o Effectiveness testing of two advertising campaigns.

o Measuring customer satisfaction before and after a new policy.

Finance:

o Determining whether a portfolio's average return is different from that of an index.

Human Resources:

o Evaluating training effectiveness (paired t-test).

o Comparing employee performance across teams.

Operations & Quality Control:

o Comparing mean processing time between two ways.

o Comparing sample output to a target specification.

Healthcare & Pharmaceuticals:

o Assessment of treatment efficacy based on small clinical trial samples.

9.5 Practical Business Applications

Hypothesis testing tools, such as the Z test or the t test are very commonly used in business to turn your data into knowledge. Managers can test out claims and assumptions with statistical evidence to make more calculated, data-driven decisions.

9.5 Marketing Research and Consumer Behavior 9.5.1 Market Research and Management Studies

• Objective: To verify assumptions of customer preference, behavior or desired mode of channel interaction.

• Applications:

o To determine difference in mean customer satisfaction score after a change in product from before.

o Difference on average spending between exposed and not-exposed to an advertisement customers (two-sample test).

o Testing if new pricing approach leads to an increase in average purchase volume.

Example: A shopkeeper tries H_0: "Average sale = ₹1,000" vs. H_1:"Average Sale > ₹1,000" after a discount exercise. If you detest the conclusions, declare the testy.

9.5.2 Quality Check and Process Change

• Objective: To make certain that production and service activities comply with established quality standards.

• Applications:

o One sample t-test' to test that the mean weight of a product batch is equivalent to its target value.

o Two sample test to compare defect rate from two machines or supplier.

o Determining how process changes (automation, new raw materials) affect quality metrics.

Example: A food company is experimenting on whether a new packaging process will lower average number of defectives packs compared to the old process.

9.5.3 Financial and Economic Decision-Making

• Motive: Prove or validate investment strategies, policies on economics and the performances of portfolios.

• Applications:

o Testing if a stock portfolio's mean return is different from an index.

o Showing the difference of average revenue growth between two economy stages (before and after policy).

o Assessing whether there is a differential with respect to default rates among customer by region or loan type.

Example: A bank might use a Z-test to see if the average time for approving a loan is much lower than the typical 48 hours in the industry.

9.5.4 HR and Productivity Analysis

• Objective: To evaluate staff performance, efficiency of training and productivity.

• Applications:

Results of paired t test showing significant improvement in employee scores after training (o)

o Average production rate comparison between two shifts or departments.

o To know if the rate of employees leaving varies significantly between the regions.

Example: HR evaluates whether employees who went through a new leadership program have higher scores on engagement surveys than those who did not.

Weekly Activity 9: Testing a Hypothesis Knowledge Check – Module 9 Instructions Save changes Submit Test Question Description This week's activity is an exercise designed to further develop your purposefully selected knowledge, based on the course competencies (CE) for this module.

Q1.

Which of the following is an appropriate null hypothesis (H_0)?

A) It's the claim we are trying to verify through research.

B) It is neutral: it presumes there is an effect or difference.

C) It only tells you that the mean of population is higher than the sample mean.

D) It is always accepted in hypothesis testing.

Q2.

A Type I error occurs when:

A) We do not reject H_0 when it is false.

B) We fail to accept H_0 when it is true.

C) We fail to reject H_1 when it is false.

D) Testing the sample is not statistically relevant.

Q3.

The average lifetime of bulbs is claimed by a manufacturer to be 1,000 hours. The mean lifetime of 64 bulbs is 980 hours with population σ =80hours. What is the calculated Z-value?

A) −2.0

B) −1.5

C) −1.0

D) −2.5

Q4.

The following hypotheses are tested by a two-sample independent t-test :

A) The sample mean versus a hypothesized population mean.

B) The mean difference between two independent groups.

C) Mean difference between paired before and after measurements.

(D) Variability of two populations.

Q5.

What is a practical business use of hypothesis testing?

A) To see if a new training program makes a difference in the performance of employees.

B) Determining the population size from national census.

C) Just recording raw sales, and never looking at it.

D) Generating frequency distributions of the ratings given by the customers.

Answer Key

B - H0 assumes that there is no effect or difference.

B – TYPE I ERROR = Rejecting a true H_0 (False POSITIVE).

A – Z = (980 − 1000) ÷ (80 / √64) = −20 ÷ 10 = −2.0.

B – Independent t-test for two samples compares means of independent groups.

A – HR is using hypothesis testing to test the effectiveness of training.

9.7 Key Terms

• Test of Hypothesis : A procedure for testing hypotheses about a population based on sample data.

• Null Hypothesis (H_0): Presumes no effect or no difference.

• Alternate Hypothesis (H_1): The alternative hypothesis assumes the effect or difference is significant.

• Significance Level (α): The probability that you will reject H_0 when it is true (Type I error).

• Critical Region: Set of values for which we reject H_0.

• Type I Error (False positive): Rejecting H_0.TrueH0.

• Type II Error: Rejecting a false H_0 (false negative).

• Z-Test: Hypothesis testing for large samples or known variance.

• t-Test: Small sample hypothesis testing with unknown variance.

p-Value: Probability of observing test results at least as extreme as the results actually observed, assuming H_0 is true.

9.8 Descriptive Questions

What are the foundations for testing hypothesis and provide a business example?

Distinguish between the following: Null hypothesis (H_0) and Alternative hypothesis (H_1).

What is a Type I error? Give a real-world business example.

What is a Type II error? Give a practical example.

Describe the balance between the errors of type I and type II when deciding.

Describe how you would go about running a one-sample z-test for the mean.

Distinguish between: (a) two-sample independent t-test and paired t-test with an example.

Name three sectors of business where hypothesis testing is frequently used.

What is the significance level in hypothesis testing?

In what way can hypothesis testing increase the reliability of manageial decisions?

9.9 References

Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2019). Statistics for Business and Economics. Cengage Learning.

• Gupta, S. C., and Kapoor, V. K. (2020). Fundamentals of Mathematical Statistics. Sultan Chand & Sons.

• Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2017). Probability and Statistics for Engineering and the Sciences. Pearson.

• Ross, S. M. (2014). Probability and statistics for engineers and scientists. Academic Press.

• Montgomery, D. C. (2019). Introduction to the Practice of Statistics. Wiley.

9.10 Case Study

Title of the Caselet: "Is Training Worth It? A Hypothesis Testing Approach"

Background

TechEdge Solutions is a midsized IT service provider that has recently implemented a program aimed at making its software developers more productive by training them to write better code. The leadership team thought the program would lower average project times.

Yet some managers said they believed that the improvements resulted from experience and team dynamics, not training. The HR analytics team used hypothesis testing to address the argument.

Problem 1: Testing Productivity Improvement

• Hypotheses:

o H0 : There is no change in the average project completion time after the training.

o H_1: Training reduces completion time.

• Method: We used a paired sample T test to compare the mean completion times of 30 developers before and after training.

• Outcome: The value of t is larger than t-critical at $\alpha = 0.05$.

• Conclusion: Reject H_0. Training significantly reduced completion time.

Problem 2: Departmental Comparison

• Firm X contrasted the average performance of trained workers in Department A with untrained workers in Department B.

• Test Used: Two-sample independent t-test.

• Outcome: There was no statistically significant effect, which meant that there are other factors (e.g. team size, project type), also impacting productivity.

Problem 3: Employee Perceptions

A study was carried out in which 60% of staff stated that training increased their work. Z-test for proportions_was used by the team to test whether the true proportion was greater than 50%.

• Result: The calculated Z > critical Z → significant.

• Conclusion: Employee perception indicates that training is beneficial.

Final Insights

• Hypothesis testing revealed that, while the training program has been effective overall, not all departments made significant improvements.

• Instead of taking an identical approach across all teams, the management team decided to create tailored training modules.