EMBA Program
MB-511

Data Science for Managerial Decisions (MB 511)

Instructor
Anant Prakash Awasthi

Data Science for Managerial Decisions (MB 511)

EMBA Program
MB-511

References/Literature

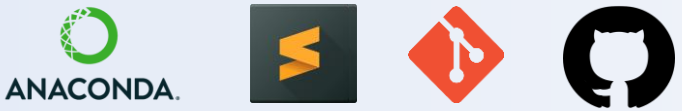
- Jojo Moolayil, “Smarter Decisions The Intersection of IoT and Data Science”, PACKT, 2016.
- Cathy O’Neil and Rachel Schutt , “Doing Data Science”, O'Reilly, 2015.
- David Dietrich, Barry Heller, Beibei Yang, “Data Science and Big data Analytics”, EMC 2013
- Raj, Pethuru, “Handbook of Research on Cloud Infrastructures for Big Data Analytics”, IGI Global
- Management Information System, W.S Jawadekar, Tata Mc Graw Hill Publication.
- Management Information System, David Kroenke, Tata Mc Graw Hill Publication.
- MIS Management Perspective, D.P. Goyal, Macmillan Business Books.

Data Science for Managerial Decisions (MB 511)

EMBA Program
MB-511

Online Resources

Software Resources



Data Science for Managerial Decisions (MB 511)

EMBA Program
MB-511

Program Overview

- Introduction to Data Science
- Information Technology An Overview
- Applications of Data Science in various fields
- MIS and Control Systems
- Data Collection and Data Pre-Processing
- Building Information Systems
- Support Systems for Management Decisions

MIS and Control Systems

EMBA Program
MB-511

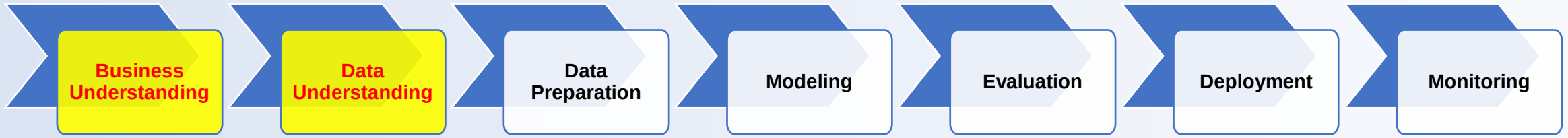
- Introduction to MIS and Control Systems
- Design and Implementation of MIS
- Control Systems in Action
- Challenges and Future Trends

Building Data Science Solutions

CRISP DM Approach

EMBA Program
MB-511

Data is not just a piece of information, but a powerful tool that can transform businesses and drive innovation." - Unknown



Building Data Science Solutions

Levels of Measurements

EMBA Program
MB-511

In statistics, measurement levels, also known as scales of measurement or levels of measurement, refer to the types or levels of data that can be collected or observed. These levels determine the types of statistical analyses that can be performed on the data. There are four commonly recognized measurement levels:

Nominal Level

At this level, data are categorized or labeled without any order or ranking. Nominal data are qualitative and can only be classified into distinct categories. Examples include gender (male, female), eye color (blue, brown, green), or types of cars (sedan, SUV, truck). Nominal data can be summarized using frequencies and mode.

Ordinal Level

In this level, data are ranked or ordered, but the differences between the ranks are not necessarily meaningful or consistent. Ordinal data can be categorized and ranked but do not have consistent intervals between the ranks. Examples include ranking preferences (1st choice, 2nd choice, 3rd choice), education level (high school diploma, bachelor's degree, master's degree), or survey responses using Likert scales (strongly agree, agree, neutral, disagree, strongly disagree). Ordinal data can be summarized using median, mode, and percentiles.

Interval Level

At this level, data are measured on a scale where the intervals between consecutive points are equal and meaningful. However, there is no true zero point. Interval data can be ordered, and the difference between two values is meaningful and consistent, but there is no absolute zero. Examples include temperature measured in Celsius or Fahrenheit, where 0°C or 0°F does not indicate the complete absence of temperature. Interval data can be summarized using mean, median, mode, standard deviation, and percentiles.

Ratio Level

This is the highest level of measurement, where data have all the properties of interval data, but with a true zero point. In ratio data, zero indicates the complete absence of the attribute being measured. Examples include height, weight, time, and money. Ratio data can be ordered, and the differences between values are meaningful and consistent, and there is a true zero point. Ratio data can be summarized using mean, median, mode, standard deviation, percentiles, and can also involve ratios and proportions.

Building Data Science Solutions

Levels of Measurements – Nominal Data

EMBA Program
MB-511

Nominal data represent categories or labels without any inherent order or ranking. These categories are distinct and mutually exclusive, meaning each observation can only belong to one category. Nominal data are qualitative in nature and cannot be quantified.
Here are examples of nominal data:

- **Gender**: Categories include male, female, and sometimes other genders like non-binary or transgender.
- **Marital Status**: Categories can include married, single, divorced, widowed, etc.
- **Eye Color**: Categories include blue, brown, green, hazel, etc.
- **Nationality**: Categories can include American, British, Chinese, Indian, etc.
- **Ethnicity/Race**: Categories may include Caucasian, African American, Hispanic, Asian, etc.
- **Political Party Affiliation**: Categories can include Democrat, Republican, Independent, Green Party, etc.
- **Type of Pet**: Categories can include dog, cat, fish, bird, etc.
- **Blood Type**: Categories include A, B, AB, O.
- **Favorite Color**: Categories can include red, blue, green, yellow, etc.
- **Hair Color**: Categories include blonde, brunette, black, red, etc.

Building Data Science Solutions

Levels of Measurements – Nominal Data

EMBA Program
MB-511

While nominal data serve essential roles in various fields, they also come with limitations that researchers and analysts need to consider:

- Lack of Order: Nominal data do not convey any inherent order or ranking among categories. For instance, in a dataset representing different countries, the categories (countries) are distinct, but there's no inherent order or hierarchy among them.
- Limited Statistical Analysis: Due to the absence of numerical values, statistical operations like arithmetic calculations (e.g., mean, median) are not meaningful for nominal data. Instead, nominal data are typically analyzed using frequency distributions, mode, and measures of association like chi-square tests.
- Loss of Information: Nominal data may lose information about the magnitude of differences between categories. For instance, if a dataset records different types of fruits (e.g., apple, banana, orange), nominal coding doesn't capture the potential differences in taste, size, or nutritional content between these categories.

Building Data Science Solutions

Levels of Measurements – Nominal Data

EMBA Program
MB-511

- Potential Misinterpretation: Care must be taken when interpreting results based on nominal data, as assigning numerical codes to categories can inadvertently imply order or magnitude where none exists. For example, assigning numerical values (1, 2, 3) to categories such as "low," "medium," and "high" can imply a quantitative relationship that may not be valid.
- Limited Relationship Analysis: Nominal data are not conducive to assessing relationships or correlations between variables. While measures of association like chi-square tests can identify associations between nominal variables, they do not provide information about the strength or direction of these relationships.
- Difficulty in Handling Missing Data: Handling missing data in nominal variables can be challenging, especially if the missingness is non-random. Techniques for dealing with missing data, such as imputation, may not be straightforward for nominal data due to their categorical nature.

Building Data Science Solutions

Levels of Measurements – Ordinal Data

EMBA Program
MB-511

Ordinal data represent categories that have a natural order or ranking. Unlike nominal data, ordinal data not only allow for classification into distinct categories but also provide information about the relative position or rank of the categories. Here are key points to discuss regarding ordinal data:

- Ordered Categories: Ordinal data consist of categories that can be ordered or ranked according to some criterion, but the differences between the categories may not be uniform or precisely quantifiable. For example, in a survey asking respondents to rate their satisfaction on a scale from "very dissatisfied" to "very satisfied," the categories have a clear order.
- Limited Quantifiability: While ordinal data have a sense of order, the intervals between categories may not be uniform or measurable. For instance, the difference in satisfaction between "very dissatisfied" and "somewhat dissatisfied" may not be the same as the difference between "somewhat satisfied" and "very satisfied."
- No Absolute Zero: Ordinal data lack a true zero point, meaning there is no category that represents the absence of the attribute being measured. For example, in a ranking of preferences (1st, 2nd, 3rd choice), a rank of "1st choice" doesn't imply a zero preference.

Building Data Science Solutions

Levels of Measurements – Ordinal Data

EMBA Program
MB-511

- Examples of Ordinal Data: Examples of ordinal data include rankings (e.g., ranks in a competition), Likert scale responses (e.g., strongly agree, agree, neutral, disagree, strongly disagree), educational levels (e.g., high school, bachelor's degree, master's degree), and socioeconomic status categories (e.g., low-income, middle-income, high-income).
- Statistical Analysis: Descriptive statistics such as median and mode are commonly used to summarize ordinal data. While the mean can technically be calculated for ordinal data, it may not always be meaningful due to the lack of equal intervals between categories. Percentiles and quartiles are also useful for understanding the distribution of ordinal data.
- Limited Arithmetic Operations: Arithmetic operations like addition and subtraction are generally not meaningful for ordinal data due to the unequal intervals between categories. However, some statistical analyses, such as non-parametric tests like the Mann-Whitney U test or Kruskal-Wallis test, can be used to analyze ordinal data without making assumptions about the distribution.

Building Data Science Solutions

Levels of Measurements – Interval Data

EMBA Program
MB-511

Interval data represent measurements where the intervals between values are equal and meaningful, but there is no true zero point. Here's a discussion on interval data:

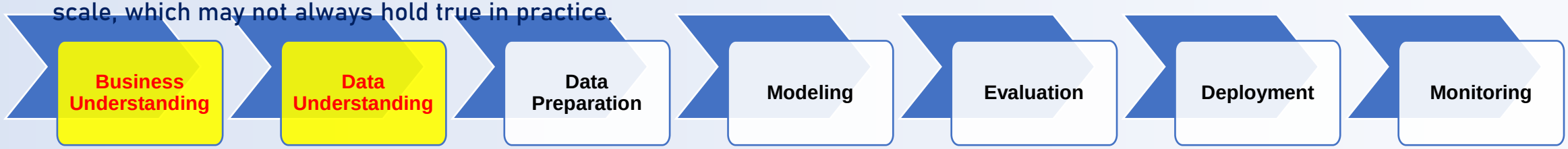
- Equal Intervals: Interval data are characterized by equal intervals between consecutive points on the scale. This means that the difference between any two adjacent values is consistent and represents the same amount of the attribute being measured. For example, the difference between 20°C and 30°C is the same as the difference between 70°F and 80°F.
- Meaningful Differences: Unlike ordinal data, interval data allow for meaningful interpretation of the differences between values. These differences represent quantitative changes in the attribute being measured. For instance, in a temperature scale, a difference of 5°C indicates the same magnitude of change regardless of the starting point.
- No True Zero Point: One crucial characteristic of interval data is the absence of a true zero point. This means that a value of zero does not indicate the complete absence of the attribute being measured but rather represents an arbitrary point on the scale. For example, a temperature of 0°C or 0°F does not imply the absence of temperature but rather a specific point on the temperature scale.

Building Data Science Solutions

Levels of Measurements – Interval Data

EMBA Program
MB-511

- Examples of Interval Data: Common examples of interval data include temperature measured in Celsius or Fahrenheit, calendar dates (e.g., years, months, days), and standardized test scores (e.g., IQ scores, SAT scores). In these examples, the intervals between values are consistent and meaningful, but there is no true zero point.
- Statistical Analysis: Interval data allow for a wide range of statistical analyses, including measures of central tendency (mean, median, mode), measures of dispersion (standard deviation, variance), and parametric tests such as t-tests and ANOVA. However, it's essential to be mindful of the scale's arbitrary zero point and interpret results accordingly.
- Transformation: Interval data can be transformed without losing the inherent structure of the data. For example, converting temperature measurements from Celsius to Fahrenheit or vice versa preserves the equal intervals between values, although it changes the scale's reference point.
- Limitations: Despite their utility, interval data have limitations. The absence of a true zero point means that ratios and proportions are not meaningful. For instance, a temperature difference of 10°C is not equivalent to a temperature difference of 10°F. Additionally, interval scales assume that the intervals between values are uniform across the entire scale, which may not always hold true in practice.



Building Data Science Solutions

Levels of Measurements – Ratio Data

EMBA Program
MB-511

Ratio scale represents the highest level of measurement in statistics, characterized by measurements that have a true zero point, equal intervals between values, and meaningful ratios between values. Here's a detailed discussion on ratio scale:

- True Zero Point: The defining feature of a ratio scale is the presence of a true zero point, which indicates the complete absence of the attribute being measured. Unlike interval scales, where zero is arbitrary, a zero value on a ratio scale represents a true absence of the attribute. For example, a weight of 0 kg indicates the absence of weight.
- Equal Intervals: Similar to interval data, ratio scale data have equal intervals between consecutive points on the scale. This means that the difference between any two adjacent values is consistent and represents the same amount of the attribute being measured.
- Meaningful Ratios: Ratio scale data allow for meaningful interpretation of ratios and proportions between values. A ratio of two values indicates the quantitative relationship between them. For example, if one object weighs twice as much as another object, the ratio of their weights is 2:1.

Building Data Science Solutions

Levels of Measurements – Ratio Data

- Absolute Zero Point: The presence of a true zero point allows for the calculation of meaningful ratios and proportions. For instance, if an object has a length of 0 cm, it means the object has no length at all. This allows for the meaningful interpretation of ratios such as one object being twice as long as another.
- Examples of Ratio Scale: Examples of data measured on a ratio scale include height, weight, length, time, temperature measured in Kelvin, and counts of objects or events. In each case, zero represents the absence of the attribute being measured, and the intervals between values are equal and meaningful.
- Statistical Analysis: Ratio scale data allow for a wide range of statistical analyses, including measures of central tendency (mean, median, mode), measures of dispersion (standard deviation, variance), and parametric tests such as t-tests and ANOVA. Additionally, ratios and proportions calculated from ratio scale data are meaningful and can be used for further analysis.

Building Data Science Solutions

Levels of Measurements – Ratio Data

EMBA Program
MB-511

- Multiplicative Operations: Ratio scale data support multiplicative operations, such as multiplication and division. This allows for the calculation of proportions, percentages, and other relative measures. For example, doubling a weight on a ratio scale results in a meaningful increase in weight.

- Limitations: While ratio scale data offer many advantages, they are not always applicable in every situation. Some variables, such as socioeconomic status or educational level, may not have a true zero point and therefore cannot be measured on a ratio scale. Additionally, ratio scale data may not be appropriate for variables that are inherently bounded or constrained, such as percentages or proportions.

Building Data Science Solutions

Levels of Measurements – Summary – Do and Don'ts

EMBA Program
MB-511

Comparison [edit]

Incremental progress	Measure property	Mathematical operators	Advanced operations	Central tendency	Variability
Nominal	Classification, membership	=, ≠	Grouping	Mode	Qualitative variation
Ordinal	Comparison, level	>, <	Sorting	Median	Range, interquartile range
Interval	Difference, affinity	+, −	Comparison to a standard	Arithmetic mean	Deviation
Ratio	Magnitude, amount	×, /	Ratio	Geometric mean, harmonic mean	Coefficient of variation, studentized range

https://en.wikipedia.org/wiki/Level_of_measurement

Quiz and Assignment

EMBA Program
MB-511

As Moodle is in picture now. I will share the assignments and quizzes with Course Administration team. They will share the details with the group.

EMBA Program
MB-511

Have a question?

Feel Free to Reach out at

- +91-88846-52929 (WhatsApp)
- anant.awasthi@outlook.com (E-Mail)