Qualitative and Quantitative Analysis

By Dr. David P. Marco, Fellow IIM, CDP, CBIP, IIBA Analytics Expert

President

EWSolutions

EWSolutions' Background

EWSolutions is a Chicago-headquartered strategic partner and full life-cycle systems integrator providing both **award winning** strategic consulting and **full-service implementation services**. This combination affords our client partners a full range of services for any size enterprise information management, metadata management, data governance and data warehouse/business intelligence initiative. Our notable client partner projects have been featured in the Chicago Tribune, Federal Computer Weekly, Journal of the American Medical Informatics Association (JAMIA), Crain's Chicago Business, FBI, The Doings and won the 2004 Intelligent Enterprise's RealWare award, 2007 Excellence in Information Integrity Award nomination, DM Review's 2005 World Class Solutions award, 2016 CIO Review 20 Most Promising Enterprise Architecture providers and 2018 Innovation award.

CHIEF DATA OFFICER MAGAZINE

FBI
Meritorious Service Award

CRAIN'S CHICAGO BUSINESS

Information Integrity Coalition
2007 Excellence in Information Integrity Award Nomination

JAMIA

INSTITUTE OF INFORMATION MANAGEMENT

DMF DATA MANAGEMENT FORUM
Excellence and Thought Leadership

CorporateLiveWire INNOVATION & EXCELLENCE AWARDS 2018 WINNER
2018: Most Innovative in Data Management Solutions

DMReview 2005
World Class Solutions Award Data Management

intelligent enterprise REAL Awards TRANSFORM 2004 WINNER
Best Business Intelligence Application Information Integration
Client: Department of Defense

Chicago Tribune

20 MOST PROMISING ENTERPRISE ARCHITECTURE TECHNOLOGY PROVIDERS - 2016

FCW THE BUSINESS OF FEDERAL TECHNOLOGY

For more information on our Strategic Consulting Services, Implementation Services, or World-Class Training, email us at info@EWSolutions.com

Over 155 Client Partners Since 1997

AFLAC
AgVantis
AON (Hewitt Associates)
Arizona Supreme Court
Bank of America (Countrywide)
Bank of Montreal
BankUnited
Basic American Foods
Becton, Dickinson and Company
Blue Cross Blue Shield Association
Blue Cross Blue Shield companies
Booz Allen Hamilton
Branch Banking & Trust (BB&T)
British Petroleum (BP)
California DMV
California State Fund
Canadian Institute for Health Information (CIHI)
Canadian National Railway
Capella University
Centers for Medicare & Medicaid Services (CMS)
Cigna
College Board
Comcast
Corning Cable Systems
Contech Systems, Inc.
CostCo
D.A. Davidson
Defense Logistics Agency (DLA)
Delta Dental
Department of Defense (DoD)
Department of Homeland Security (DHS)
Department of the Interior
Discover
Driehaus Capital Management
Eli Lilly and Company
Environment Protection Agency (EPA)
Erwin
Fannie Mae
Farmers Insurance Group
Federal Aviation Administration
Federal Bureau of Investigation (FBI)
FedEx
FEMA
Fidelity Information Services
FICO (Fair Issac)
Ford Motor Company
GlaxoSmithKline
Harbor Funds
Harris Bank (BMO)
The Hartford
Harvard Pilgrim HealthCare
Health Care Services Corporation
Henny Penny
HP (Hewlett-Packard)
IDMA (Insurance Data Mgt. Institute)
Information Resources Inc.
International Paper
IQVIA (Quintiles)
Janus Mutual Funds
Johnson Controls
Key Bank
LiquidNet
Loyola Medical Center
Manulife Financial (John Hancock Fin.)
Mayo Clinic
McDonalds
Micron
Microsoft
MoneyGram
NASA
National City Bank (PNC Financial)
Nationwide
Neighborhood Health Plan
NORC
Physicians Mutual Insurance
Pillsbury
Quartz Benefits
Rowan College of South Jersey (RCSJ)
Sallie Mae
Saudi Arabia Minister of the Interior
Saudi Telecom Company (STC)
Schneider National
Secretary of Defense/Logistics
Singapore Defense Science & Technology Agency (DSTA)
Social Security Administration
South Orange County Community College
Spherion
Standard Bank of South Africa
SunBelt Rentals
SunTrust Bank
Surescripts
Target Corporation
Texas CPA
The Regence Group
Thomson Multimedia (RCA)
Thrivent Financial
United Healthcare (UHC)
United Health Group
United Nations (ICAO)
United States Air Force
United States Federal Courts
United States Department of State
United States Fish and Wildlife Service
United States Navy
United States Navy (BUPERS)
United States Transportation Command
University of Michigan
University of Wisconsin Health
USAA
US Cellular
Vera Bradley
Waste Management
Wells Fargo
Wisconsin Department of Transportation
Workers' Compensation Insurance Rating Bureau of California (WCIRB)
Zurich Cantonal Bank

GSA Schedule

For more information on our Strategic Consulting Services, Implementation Services, or World-Class Training email us at **Info@EWSolutions.com**

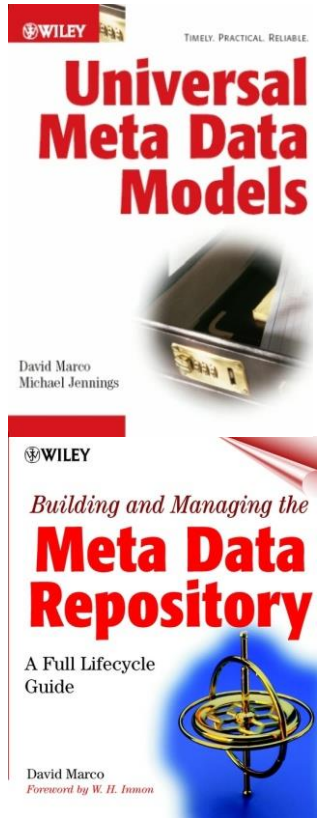
Over 155 Successful Client Partner Implementations
Contact us at INFO@EWSolutions.com to become our next success

Dr. David P. Marco – Professional Profile

Best known as the world's foremost authority on metadata management and the father of the Managed Metadata Environment, he is an internationally recognized expert in the fields of data governance, digital transformation, data literacy, big data, data warehousing, master data management and data management. In 2004 Dr. David P. Marco was named the **"Melvil Dewey of Metadata"** by **Crain's Chicago Business** as he was selected to their very prestigious **"Top 40 Under 40"** list. Dr. Marco has authored several books including the widely acclaimed **"Universal Metadata Models"** (Wiley, 2004) and the classic **"Building and Managing the Metadata Repository: A Full Life-Cycle Guide"** (Wiley, 2000).

- 2022 **CDO Magazine Top Data Consultants in North America**
- 2021, he was named an **IIBA Analytics Expert**
- 2020 received a lifetime professional **Fellowship from the Institute of Information Management** (their highest honor)
- President of Data Management University (**DataManagementU.com**)
- Author of several best-selling information technology books, including the top 2 sellers in metadata management history
- **2016 Data Management Channel Expert** for Business Analytics Collaborative
- **2008 Data Management Hall of Fame (**Professional Achievement Award**)**
- **2007 DePaul University named** him one of their **"Top 14 Alumni Under 40"**
- Selected to the prestigious **2004 Crain's Chicago Business "Top 40 Under 40"**
- Worked with over 90 client partners to successfully build their data management & governance programs
- Presented hundreds of keynotes/seminars across four continents
- Published hundreds of IT articles, some of which were translated into Mandarin, Russian, Italian, Portuguese, and others
- Taught at the **University of Chicago** and **DePaul University**
- Holds Ph.D. (Hon), MBA, Fellow IIM, CDP, CCP, and CBIP certifications

Email: DMarco@EWSolutions.com



Agenda

- Data Literacy Background & Fundamentals
- Qualitative and Quantitative Types of Data
- Hypothesis Testing
- Correlation

Data Literacy Defined

- ***Data Literacy:*** ***is the ability to create, manage, read, work with, and analyze data to ensure and maximize the data's accuracy, trust, and value to the organization***

Qualitative and Quantitative Types of Data

Qualitative and Quantitative

- **Qualitative:** categories and characteristics that cannot be measured numerically
  - Cannot be aggregated
- **Quantitative:** numerically measurable categories and characteristics
  - Can be aggregated

Name	Age	Height	Gender	Favorite Food
Alyssa	12	5' 1"	F	Italian Bread
David	42	6' 0"	M	Steak
Elizabeth	32	5' 11"	F	Chocolate Cake
Matthew	5	3' 10"	M	Chicken Wings

Name, Gender and **Favorite Food** are qualitative

Age and **Height** are quantitative

Qualitative

- **Qualitative:** categories and characteristics that cannot be measured numerically
- Qualitative has 2 types of variables that both represent ranking:
  - Nominal Qualitative Variables
  - Ordinal Qualitative Variables

Nominal Qualitative Variables

- **Nominal Qualitative Variables** represent categories that cannot be ranked.
- **Example:** possibly we are doing an analysis on vegetables like broccoli, spinach, and carrots. These are nominal qualitative variables because they are not ranked. A carrot is not of a higher rank than spinach
- **Learning Point:** You can remember this type of a variable by **Nominal = Named**

Ordinal Qualitative Variables

- **Ordinal Qualitative Variables: these** are variables that can be ranked. They are not numerically measurable, but there is a logical rank-order
- Example: Ordinal variables are often used with qualitative surveys:
  - Extremely dissatisfied, Dissatisfied, Neither satisfied nor dissatisfied, Satisfied, Extremely satisfied
  - Never, Sometimes, Oftentimes, Always
- Ordinal Qualitative Variables can often be converted into Quantitative variables by giving them numeric equivalents (1 = Never, 2 = Sometimes, 3 = Oftentimes, 4 = Always)
- **Learning Point:** You can remember this type of a variable by **Ordinal = Order**

Quantitative

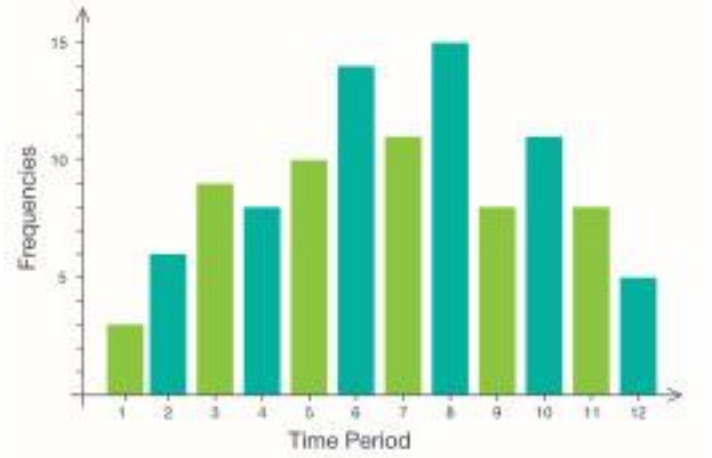
- **Quantitative:** numerically measurable categories and characteristics
  - Can be aggregated
- Quantitative has 2 types of variables:
  - Discrete Variables (Data)
  - Continuous Variables (Data)

Discrete Variables

- **Discrete (or Attribute) Variables (Data)** consists of numerical data that can only be counted and cannot be measured or divided
- Characteristics of discrete data
  - Data can be counted. It is usually an integer (a number without a fractional component)
  - Cannot be measured (e.g., your weight is not discrete since it can be measured)
  - Data cannot be divided into smaller pieces (e.g., "Number of Children in a Family" cannot be divided as there is no such thing as a half child)
  - Has a limited number of possible values (e.g., months in a year)
- **Example:** number of children in a family, number of students in a class, number of default loans, number of touchdowns in a football game, etc.
- **Learning Point:** Histograms and pie charts are great for displaying discrete variables

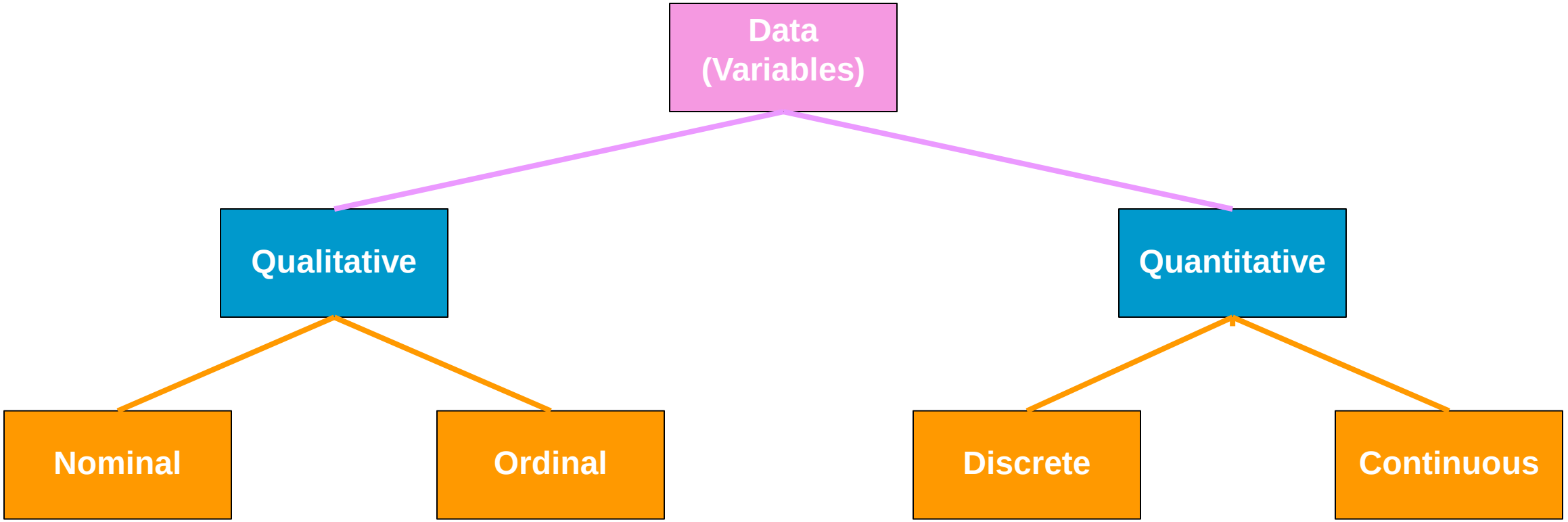
Continuous Variables

- **Continuous Variables** can be meaningfully divided into finer levels. It can be measured on a scale or continuum and can have almost any numeric value. For example, you can measure a person's height at very precise scales — feet, inches, meters, centimeters, millimeters, etc.
- Can take (almost) any value between two numbers (e.g., between 100 pounds and 150 pounds are literally millions of possible heights: 122.5 lbs., 155.701957 lbs., etc.
- **Learning Point:** Histograms and line graphs are often used to show trends in continuous variables



Histogram

Qualitative & Quantitative Variables



- Carrots, peas, etc.
- Product categories, etc.

- Never, Sometimes, etc.
- Dissatisfied, Neutral, etc.

- 7 children
- 205 employees
- 1,754 mobile phones

- 4.72 feet
- 192.975 pounds
- $9,397.759272

Hypothesis Testing

Hypothesis Testing

- **Hypothesis:** A tentative explanation for an observation, phenomenon, or scientific problem that can be tested by further investigation. An assumption.
- **Hypothesis Testing:** is the formal procedures used to test whether a hypothesis can be accepted or not.
  - Uses data from a sample to draw conclusions (inferences) about a population
  - Sometimes called “educated guessing”
- **Inference** is the process of drawing conclusions about a population based on a sample of the data.
  - Typically, its not practical to obtain all the measurements in a given population
- **Null Hypothesis:** assumes there is no statistical importance between the two variables in the hypothesis.
- **Alternative Hypothesis:** supports a statistically significant result. The opposite of a null hypothesis. Rejects the null hypothesis.
- **Learning Point:** If an analyst has data for the entire population, they don't need to make inferences about the difference between groups within that population.

Hypothesis Testing – Example

- Hypothesis testing begins with the creation of null and alternative hypothesis statements. Let's look at an example.
  - Null hypothesis states that the compound (medication) will not impact health outcomes. It proposes that those who receive the treatment will not have different outcomes from those who do not
  - Alternative hypothesis states that there will be a difference. It proposes that those receiving the medication will show improved health outcomes as compared to those who do not
  - **Hypothesis tests assume that the null hypothesis is true!**
    - Innocent until proven guilty
  - Hypothesis tests take the number of samples, the size of the difference measured, and the amount of variation observed in each group into account
  - If the group that received the medication has a 2% improved health outcome as compared to a placebo group, we may conclude that 2% is not statistically significant enough to disprove the null hypothesis

Correlation

Correlation

- **Correlation:** is a technique that can show whether and how strongly pairs of quantitative variables are related
- Correlation is limited to being between -1 and 1
  - **Correlation = -1:** If it is -1, then variables are perfectly negatively correlated. That means if one variable is moving in one direction, then another is moving in the opposite direction
  - **Correlation = 0:** Means the variables do not have any correlation
  - **Correlation = 1:** If it is +1, then variables are known as perfectly positively correlated. Both variables are moving in positive directions
- Example, suppose we study the correlation between individuals' average income and the revenues of 2 companies (Chick-Fil-A and Hobby Lobby)
  - Our hypothesis is that the higher a person's income, the greater the revenues of these 2 companies
  - After analyzing the data, we find that Chick-Fil-A has a Correlation Coefficient = 0.07 and Hobby Lobby has a Correlation Coefficient = 0.81
  - We can see that both companies are positively collated to a person's invoice; however, Hobby Lobby's correlation is much higher

Correlation and Causation

- **Learning Point:** Correlation does not always equal causation
  - 100% of adults who have died have eaten an apple at some point in their life. Correlation Coefficient =1; therefore, apples cause death. **WRONG**
- Correlation tells us only how strongly a pair of quantitative variables are linearly related. **It does not explain the how or why**
- For example, sales of air conditioners correlate with sales of sunscreen. Are people buying air conditioners because they bought sunscreen or vice versa? No. The cause of both purchases is clearly something else, in this case, hot weather
- **Learning Point:** Establishing that one thing causes another takes a detailed, methodical and unbiased analysis

In the Final Analysis

In Conclusion

- Qualitative and quantitative analysis techniques differ and their use
- Correlation does not always equal causation
- Remove as much bias as you can from your analytics

Questions & Answers

Data Management Consulting Services

Over 155 Client Partners

100% Data Management Success since 1997

World-Class Training

More than 40 courses, taught by industry experts. Over 150 satisfied client partners

GOVERNANCE METADATA ANALYTICS — EWSolutions METHODOLOGIES EWSOLUTIONS.COM

EWSolutions™

DATAMANAGEMENT EDUCATION.COM

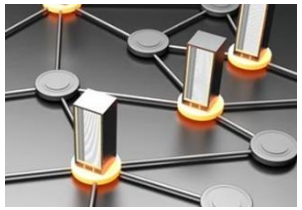
Online, On-Demand Courses

Data Warehouse Data Models

DataManagementU.com

Free knowledge portal

Big Data Meta Model



Over 155 Successful Client Partner Implementations

Contact us at INFO@EWSolutions.com to become our next success

Questions