Module 4
Fundamental Big Data Analysis & Science

Exploratory Data Analysis

- numerical summaries
  - measures of central tendency
  - measures of variation or dispersion
  - measures of association
- rules
  - Chebyshev's inequality rule
  - empirical rule
- analysis types
  - univariate analysis
  - bivariate analysis
  - multivariate analysis

Statistics Mathematics

- interquartile range (IQR)
- mean
- median
- mode
- robustness
- range
- quantile
- quintile
- quartile
- percentile
- population
- bias
- variance
- standard deviation
- z-score
- distributions
  - frequency
  - probability
  - discrete
  - continuous
  - sampling
    - standard error
    - statistical estimator
      - point estimator
      - interval estimator
    - confidence interval
    - skewness
      - positively skewed
      - negatively skewed
  - binomial
  - geometric
  - normal
  - uniform

Statistical Models

- linear regression
- mean squared error
- error term
- residual
- coefficient of determination R^2
- standard error of estimate

Visualization

- bar chart
- line graph
- histogram
- frequency polygons
- scatter plot
- stem and leaf plot
- cross-tabulation
- box and whisker plot
- quantile-quantile (q-q)plot
- lattice plot

Statistics Variable Categories

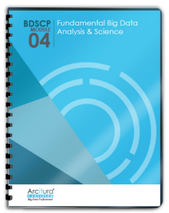
- nominal
- ordinal
- binary
- quantitative
- independent
- random
- qualitative

Big Data Dataset Categories

- high-volume
- high-velocity
- high-variety
- high-veracity
- high-value

Statistics Analysis

- descriptive statistics
- inferential statistics
- correlation
- covariance
- hypothesis testing
  - null hypothesis
  - alternative hypothesis
  - statistical significance
  - p-value
  - type I error
  - type II error
  - critical region



Module 4: Fundamental Big Data Analysis & Science
Official Mind Map Supplement

Arcitura CERTIFIED Big Data Scientist

Arcitura Big Data School - Big Data Scientist Certification Program
Copyright © Arcitura Education Inc. www.arcitura.com

Arcitura TRAINING Big Data Science

Arcitura CERTIFIED Big Data Scientist