



Hyrje në shkencën e të dhënave

Pjesa 3 – Teknikat

Prof. Asoc. Dr. Ermir Rogova

Data analysis vs. data analytics



Data analysis = hands-on data exploration and evaluation



Data analytics = broader term that includes data analysis as a necessary part



Analysis – typically looks at what happened in order to explain



Analytics – models the future or predicts a result

Types of analysis/analytics (that we care about here)

Descriptive

Diagnostic

Predictive

Prescriptive

Exploratory

Mechanistic

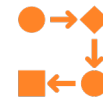
1. Descriptive analysis



Typically it is the first kind of data analysis performed on a data set.



Usually it is applied to large volumes of data, such as census data.



Description and interpretation processes are different steps.



Core concepts/techniques:

Frequency distribution:
histogram

Measures of centrality: mean, median, mode

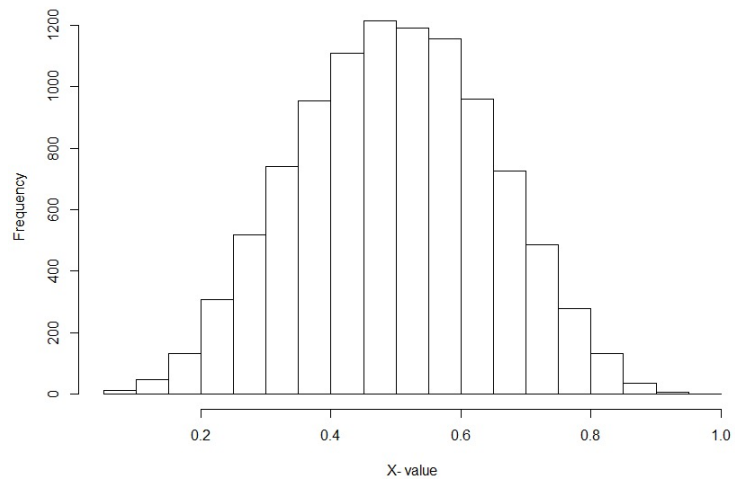
Dispersion of distribution: range, interquartile range, variance, standard deviation

Variables

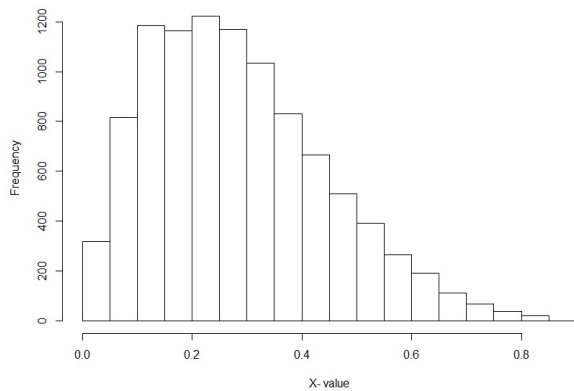
- Categorical
 - Independent or predictor
 - Dependent or outcome
- Nominal
- Ordinal
- Interval
- Ratio

Data distributions

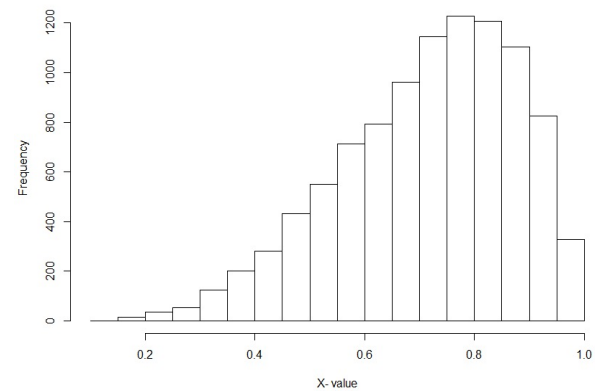
Normal distribution



Skewed right



Skewed left



Measures of centrality

- Mean
- Median
- Mode

Dispersion of a distribution

- Range
- Interquartile range
- Variance
- Standard deviation

2. Diagnostic analytics



Used for discovery, or to determine why something happened



Most commonly used technique: correlation

3. Predictive analytics



Obtain *insight* from *hindsight* as we identify patterns from existing data



Then use such *insights* to predict the future: *foresight*

4. Prescriptive analytics

An area of business analytics dedicated to finding the best course of action for a given situation



Typical steps:

Start by first analyzing the situation
(using descriptive analysis)

Move toward finding connections
among various parameters/variables,
and their relation to each other

Use this to address a specific problem,
more likely that of prediction

5. Exploratory analysis



An approach to analyzing data sets to find previously unknown relationships



Often involves visualization techniques



Useful when we lack a clear question or a hypothesis

6. Mechanistic analysis



Involves understanding the exact changes in variables that lead to changes in other variables for individual objects



Most common (and powerful) technique: regression



Regression analysis is a process for estimating relationships among variables: typically from predictors to outcome

Hands-on with regression

#	Attitude	Score
1	65	129
2	67	126
3	68	143
4	70	156
5	71	161
6	72	158
7	72	168
8	73	166
9	73	182
10	75	201

Pearson's correlation coefficient $r = 0.94$

S.D. of Attitude = 3.10, Score = 22.80

$$\beta_1 = 0.94 \times \frac{22.80}{3.10} = 6.91$$

$$\beta_0 = 159 - (6.91 \times 70.6) = -324.7$$

$$y = -324.7 + (6.91 \times 78) = 214$$

Summary

- Data analysis: hands-on data exploration and evaluation. Analysis looks backwards, providing marketers with a historical view of what has happened.
- Data analytics: defines the science behind the analysis. The science means understanding the cognitive processes an analyst uses to understand problems and explore data in meaningful ways. It is used to model the future or predict a result.
- Analysis/analytics types
 - Descriptive analysis
 - Diagnostic analytics
 - Predictive analytics
 - Prescriptive analytics
 - Exploratory analysis
 - Mechanistic analysis

Temat e javës së katërt

- Unix/Linux
- Connecting to a Unix/Linux Server
- Basic Commands
- Editing on Unix/Linux
- Redirections and Piping
- Solving Small Problems with Unix/Linux



Pyetje???