Business Analytics & Machine Learning

Course Overview

Prof. Dr. Martin Bichler

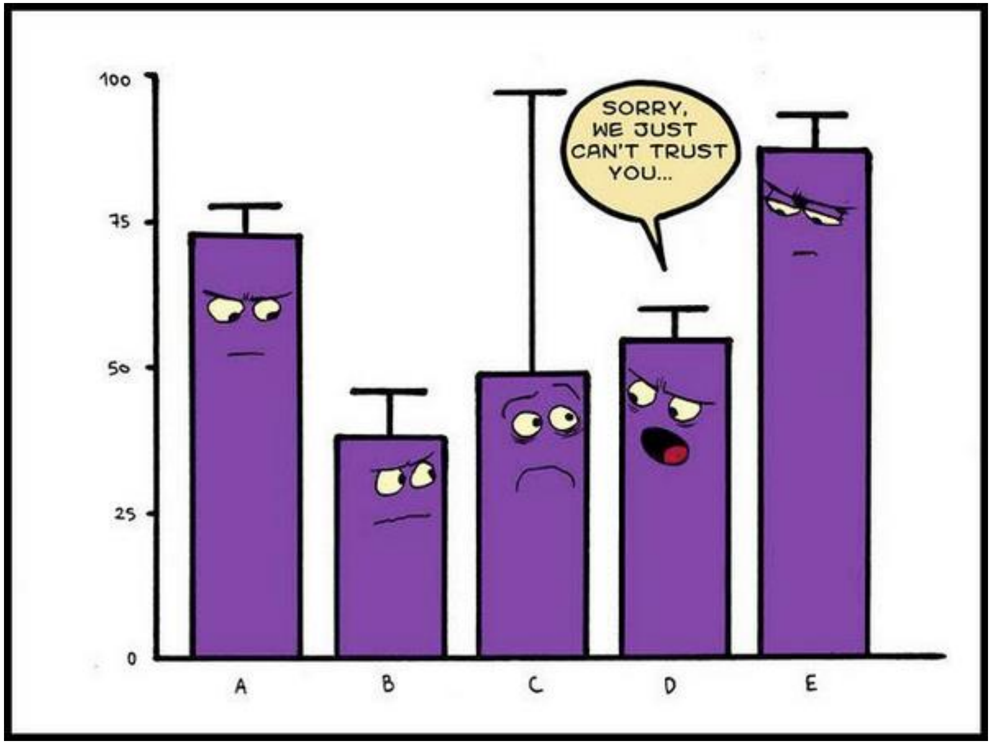
Decision Sciences & Systems

Department of Informatics

Technical University of Munich

Agenda for Today

1. Understand what this course is all about
2. Learn about organization, grading, and tutor groups
3. Refresh basic statistical concepts



Chair for Decision Sciences & Systems

(Prof. Bichler and Prof. Brandt)

Focus in research:

Game theory, mathematical optimization, market design, and social choice.

Core courses:

- *Winter term*
  - Auction Theory and Market Design
  - Business Analytics and Machine Learning
  - Computational Social Choice
  - Seminar on Markets, Algorithms, Incentives, and Networks

- *Summer term*
  - Operations Research
  - Algorithmic Game Theory
  - Seminar on Economics and Computation
  - Seminar on Data Mining

Introductory Remarks

- This course provides an introduction to data analysis with a focus on problems in business and economics.
- There are more than 1200 students enrolled having very (!) diverse backgrounds. As a result, we aim to make minimal assumptions on prerequisites and only require an introduction to statistics.
- Later classes (e.g., dimensionality reduction, neural networks) require some linear algebra and basic concepts in convex optimization. Given the diversity of students, we limit the mathematical background to the essentials necessary to understand different techniques.
- After this course you should be able to analyze mid-sized data sets, perform regression and classification tasks with wide-spread methods, and be able to interpret the results.
- We aimed for an in-person course in 2021, but decided to provide short video lectures again due to the large number of students entrolled.

New Name for this Course

Starting in the winter term 2021/22, the former module "*IN2028 Business Analytics*" is being replaced by the new module "*IN2028 Business Analytics and Machine Learning*".

It is one of the modules intended for the Bachelor programs that serves as prerequisite for machine learning classes in the Master. There shall be similar modules offered in the BSc Informatics and Bioinformatics.

Since the name change has only been announced recently, you will still see "*IN2028 Business Analytics*“ in some types of study and some FPSOs.

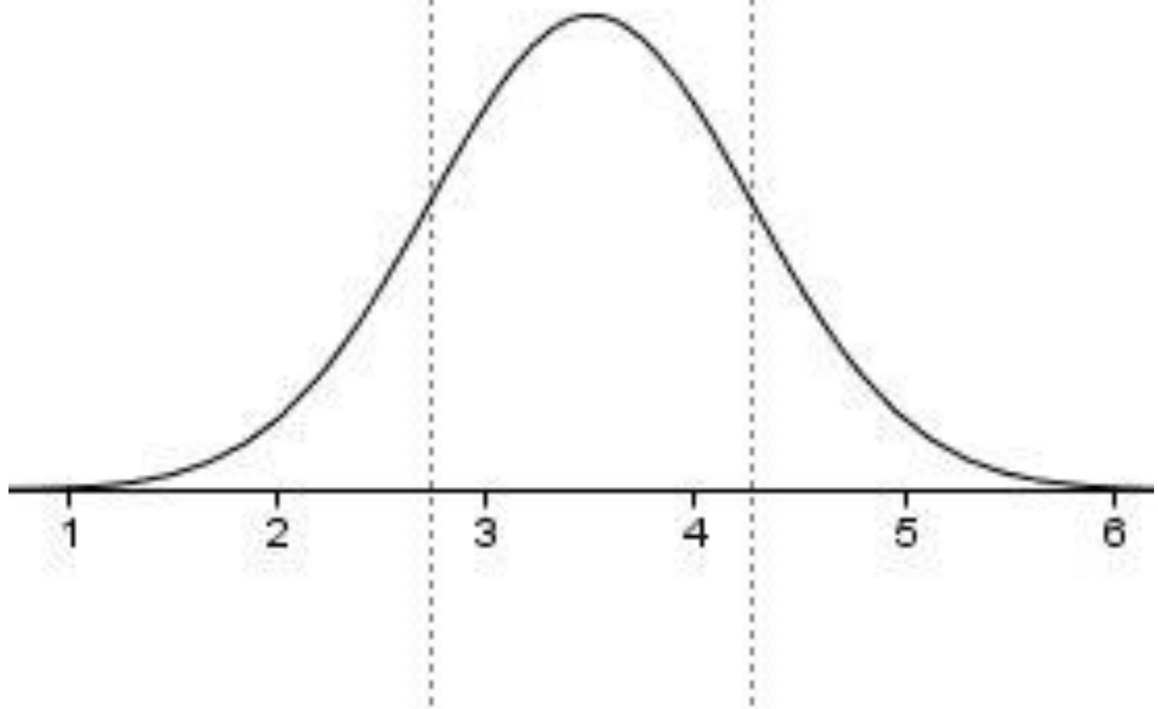
To keep the confusion to a minimum, we did not yet adapt the slides to the new name.

This Course is for …

I know very little about statistics and programming

I have heared statistics and algorithms in my bachelor program

I have already attended a machine learning or deep learning course



Business Analytics

Business analytics makes extensive use of statistical analysis, including explanatory and predictive modeling, and fact-based management to drive decision making. It is therefore closely related to management science. Analytics may be used as input for human decisions or may drive fully automated decisions.

Source: http://en.wikipedia.org/

Analytics Landscape

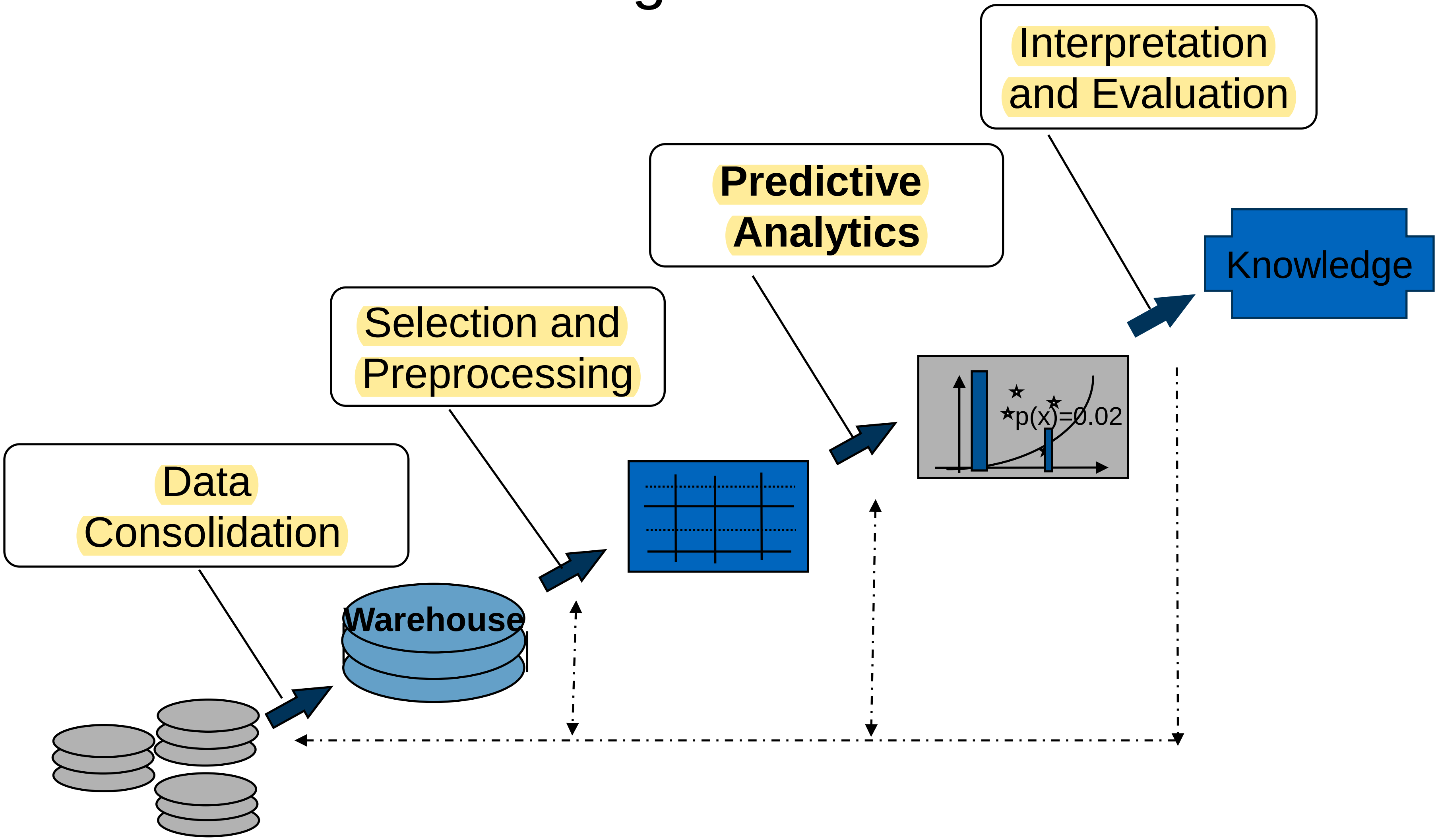
			Academic Fields
Stochastic Optimization	How can we achieve the best outcome including the effects of variability?	Prescriptive	Operations Research
Optimization	How can we achieve the best outcome?		
Predictive modeling	What will happen next if ?	Predictive	Econometrics & Machine Learning
Forecasting	What if these trends continue?		
Simulation	What could happen.... ?		
Alerts	What actions are needed?		
Query/drill down	What exactly is the problem?	Descriptive	Data Management
Ad hoc reporting	How many, how often, where?		
Standard Reporting	What happened?		

Competitive Advantage

Degree of Complexity

Based on: Competing on Analytics, Davenport and Harris, 2007

From Data to Knowledge



Predictive Analytics

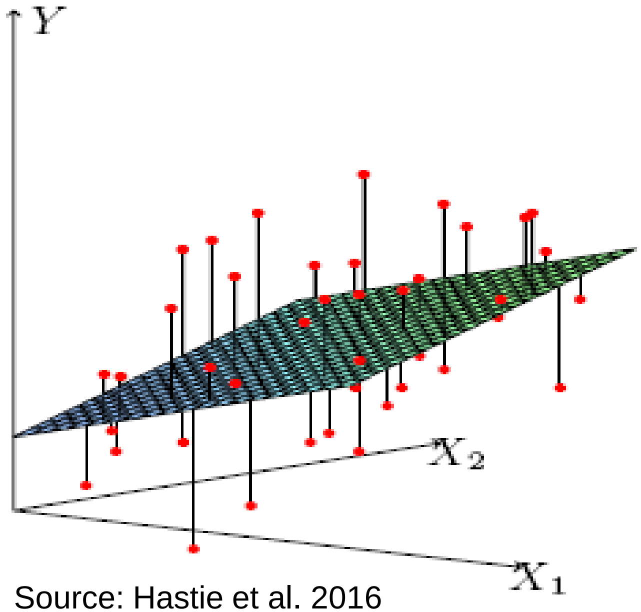
Predictive Analytics draws on methods from different fields, in particular from econometrics and machine learning. The following definitions are taken from en.wikipedia.org.

Econometrics is the application of statistical methods to economic data in order to give empirical content to economic relationships. More precisely, it is "the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference".

Machine learning (ML) is the study of computer algorithms that can improve automatically through experience and by the use of data. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so.

Numerical Prediction

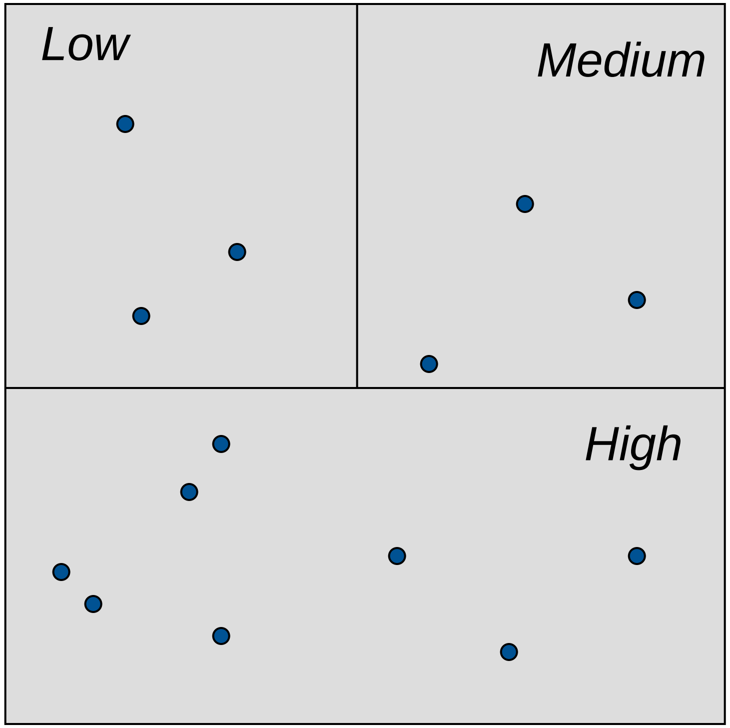
- Given a collection of data with known numeric outputs, create a function that outputs a predicted value from a new set of inputs
- For example, given age and income of a person, predict monthly expenses dining.



Source: Hastie et al. 2016

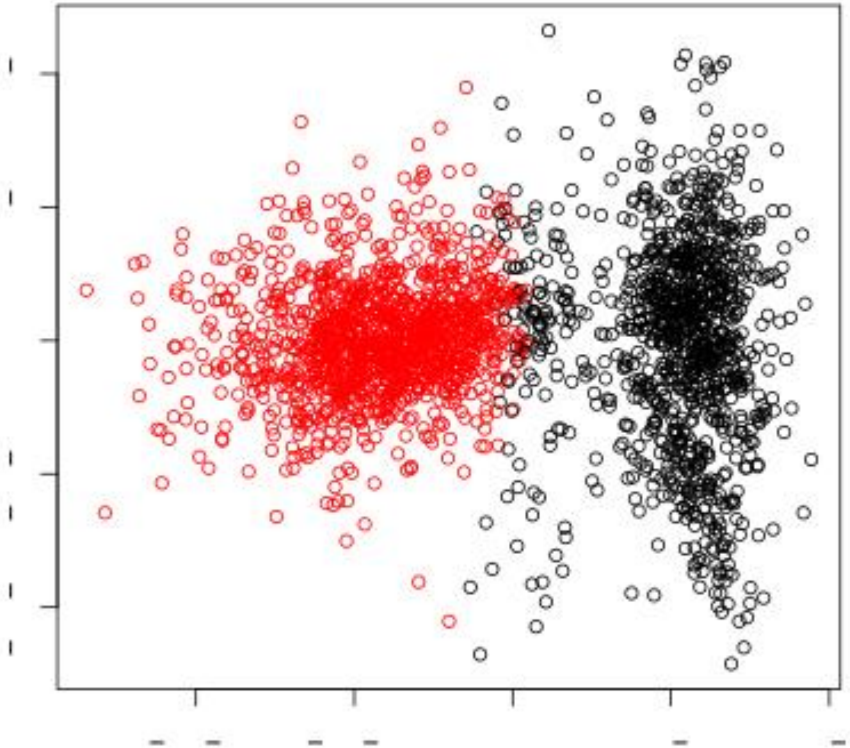
Classification

- From data with known labels, create a *classifier* that determines which label to apply to a new observation
- E.g. Identify new loan applicants as low, medium, or high risk based on existing applicant behavior



Clustering

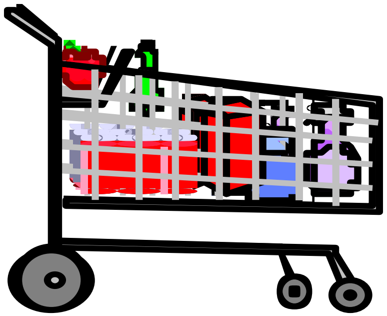
- Identify “natural” groupings in data
- Unsupervised learning, no predefined groups
- E.g. Identify clusters of “similar” customers



Association Rule Analysis

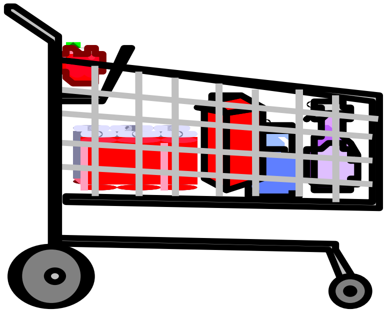
- Identify relationships in data from co-occuring terms or items
- E.g., analyze grocery store purchases to identify items most commonly purchased together

Milk, eggs, sugar, bread



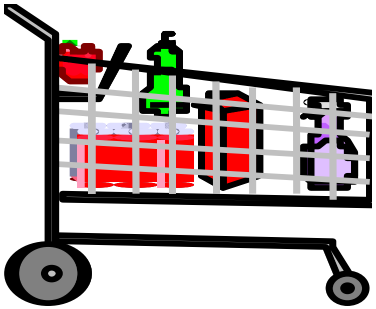
Customer1

Milk, eggs, cereal, bread



Customer2

Eggs, sugar



Customer3

Machine Learning Terminology

In machine learning these tasks are categorized as supervised and usupervised learning:

Supervised learning:

- Supervised methods are thought to attempt the discovery of the relationships between input attributes and a target attribute.
- A training set is given and the objective is to form a description that can be used to predict unseen examples.
- Examples: Classification, numerical prediction

Unsupervised learning:

- There is no supervisor and only input data is available.
- The aim is now to find regularities, irregularities, relationships, similarities and associations in the input.
- Examples: Clustering, association rules

What is a Model?

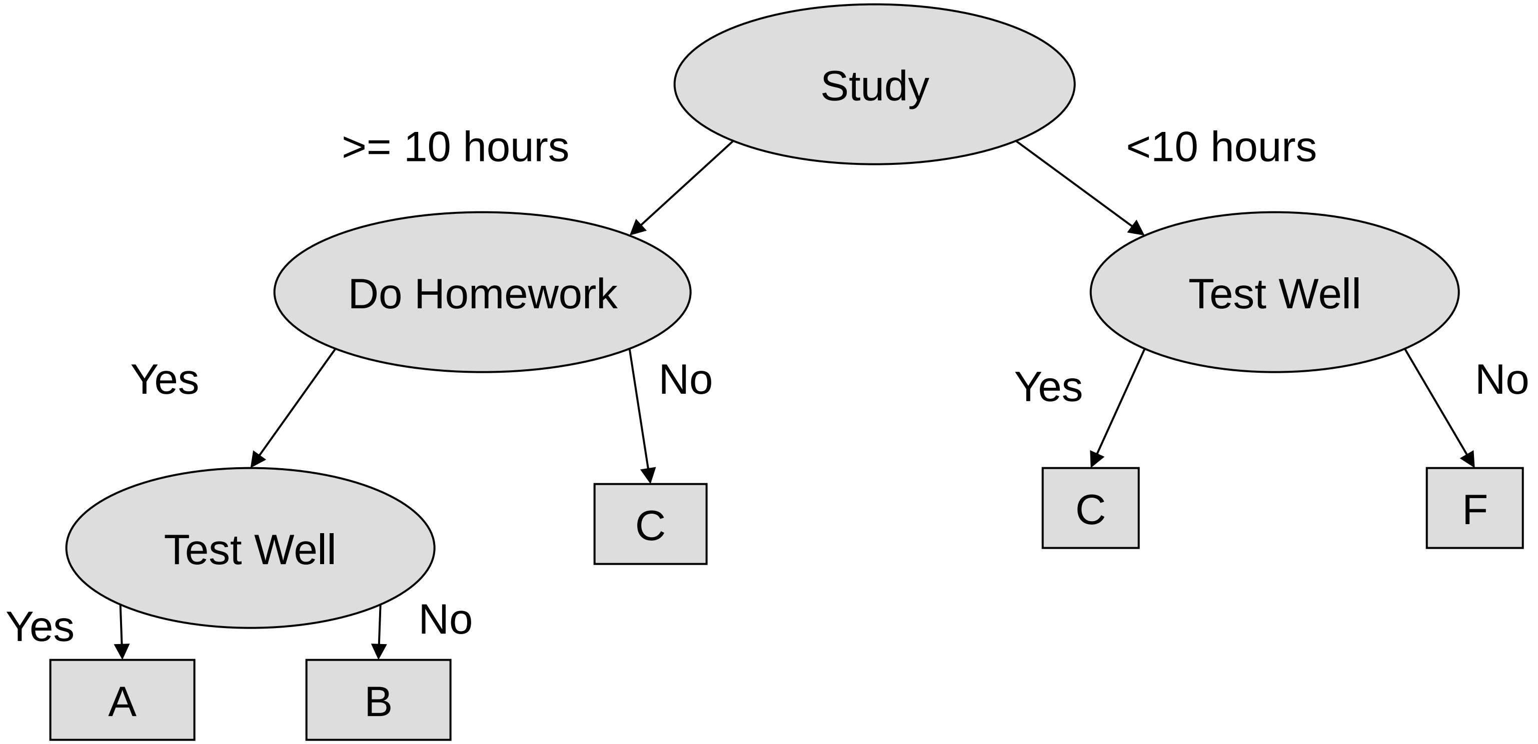
- A representation of a system that allows for investigation of the properties of the system and, in some cases, prediction of future outcomes.

- Linear functions as a well-known example
  – Mathematical combination of attribute values
  – Expenses based on age and income
  – Could be a non-linear or linear model

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e_i$$

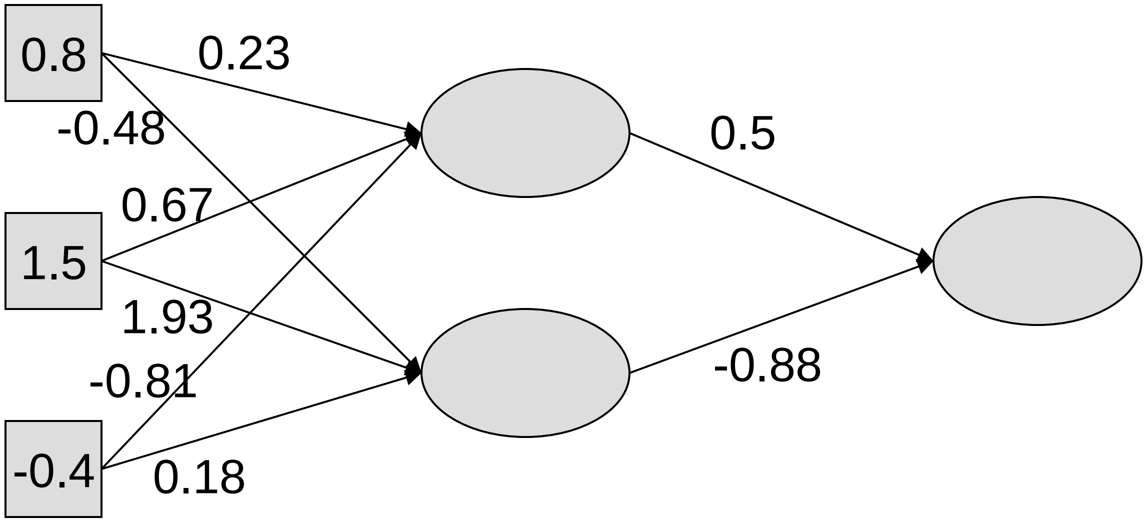
Models

Decision Trees



Models

Neural Networks



What do Data Scientists Work on?

	2011	2013	2015
Improving understanding of customers	33%	45%	46%
Retaining customers	30%	36%	37%
Improving customer experiences	22%	36%	36%
Selling products / services to existing customers	23%	33%	35%
Market research / survey analysis	29%	36%	34%
Acquiring customers	23%	32%	32%
Improving direct marketing programs	22%	27%	30%
Sales forecasting	19%	27%	27%
Fraud detection or prevention	21%	23%	26%
Risk management / credit scoring	22%	26%	25%
Price optimization	14%	22%	23%
Manufacturing improvement	10%	15%	17%
Medical advancement / drug discovery / biotech / genomics	12%	17%	17%
Supply chain optimization	7%	11%	15%
Investment planning / optimization	11%	13%	14%
Software optimization	7%	9%	11%
Website or search optimization	8%	12%	10%
Human resource applications	4%	8%	9%
Collections	6%	7%	8%
Language understanding	4%	7%	8%
Criminal or terrorist detection	4%	4%	7%

Source: Rexter Analytics Data Science Survey, 2016

Example Application: Churn Prediction

Churn: the proportion of contractual customers or subscribers who leave a supplier or service provider during a given time period

Example

- Churn rates of a telecom at 2.1% per month
- Causes: increased competition, lack of differentiation, market saturation
- Cost: $300 to $700 cost of replacement of a lost customer in terms of sales support, marketing, advertising, etc.
- Response: Targeted retention strategies

Churn Prediction as Classification Task

Churn as a Classification problem:

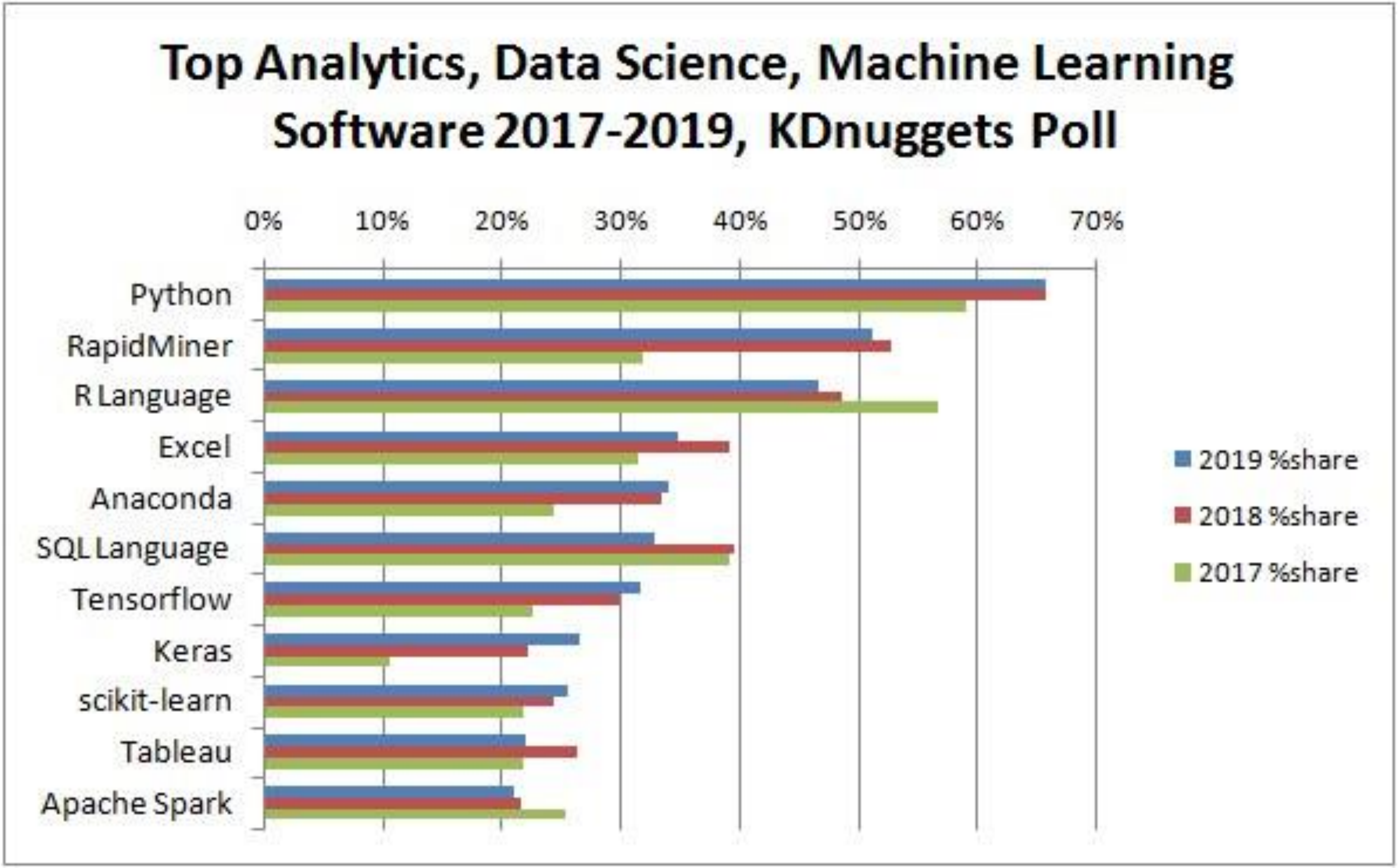
Classify a customer *i* characterized by *p* variables

$x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ as

- Churner $y_i = +1$
- Non-churner $y_i = -1$

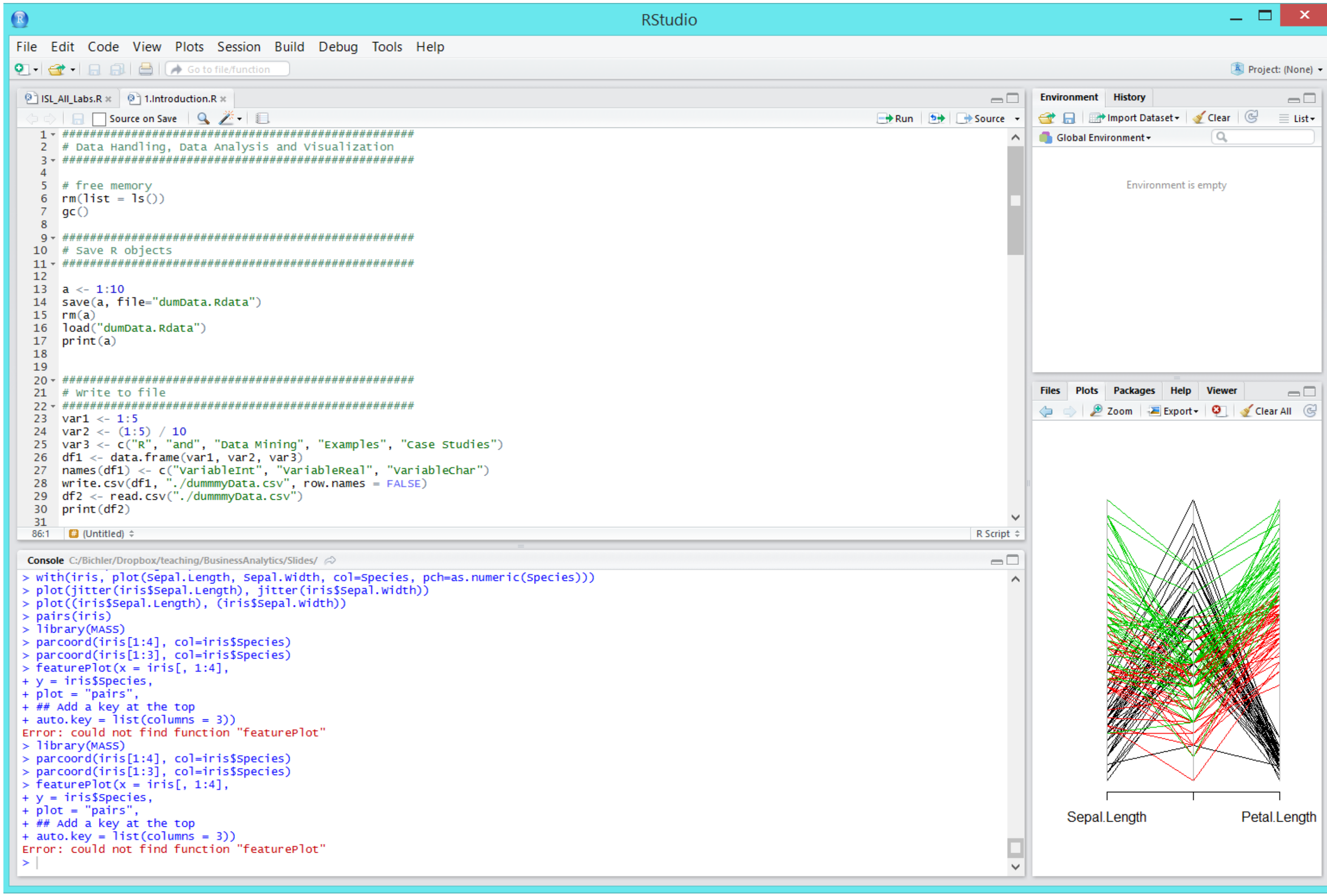
Churn is the response binary variable to predict: $y_i = f(x_i)$

Choice of the binary choice model $f(\,.\,)$?



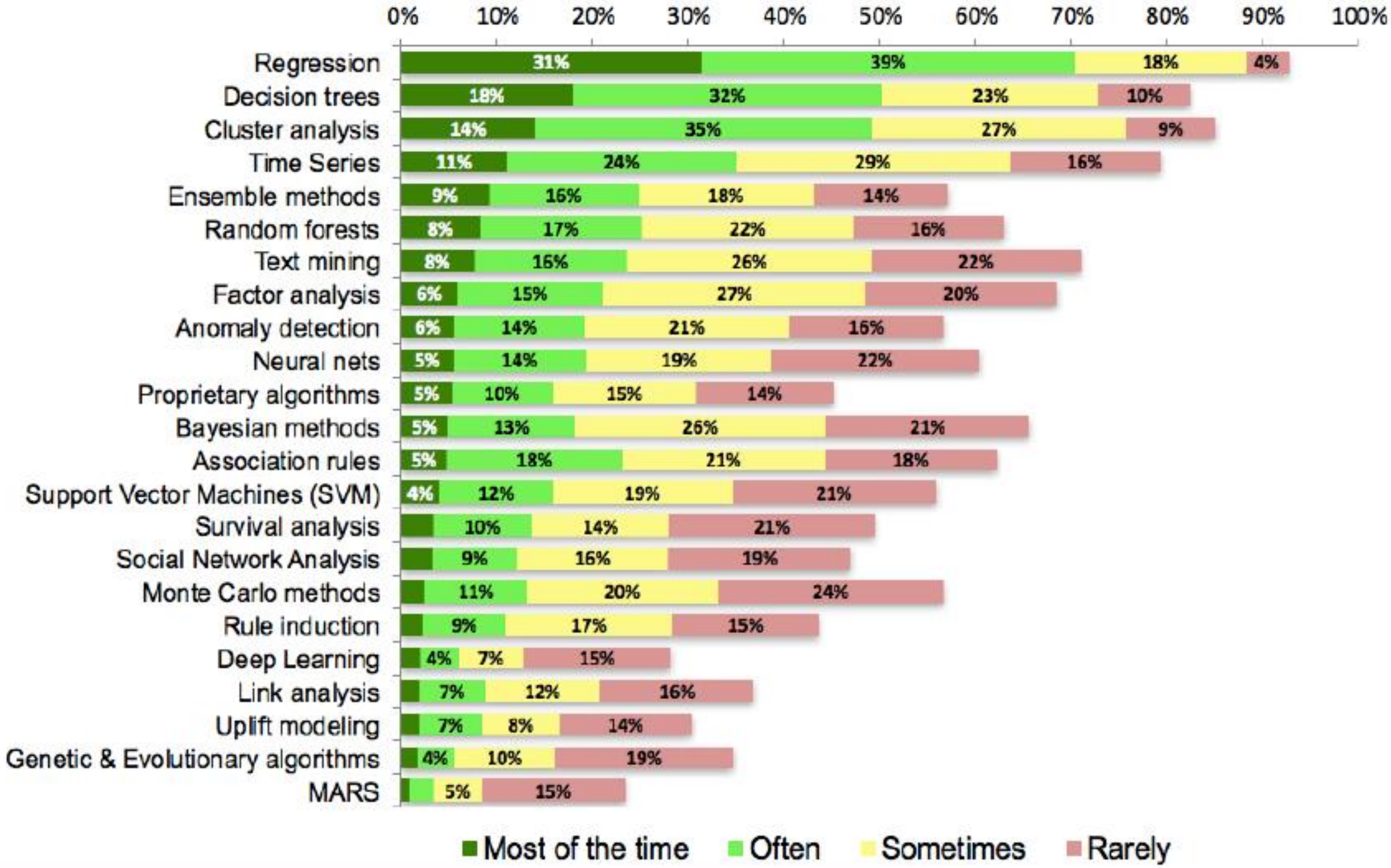
Source: KDnuggets Analytics/Data Science 2019 Software Poll

Tools Used in this Course: R Studio



Our Core Algorithms Remain the Same

- Regression, decision trees, and cluster analysis continue to form a triad of core algorithms for most data miners. This has been consistent since the first Data Miner Survey in 2007.

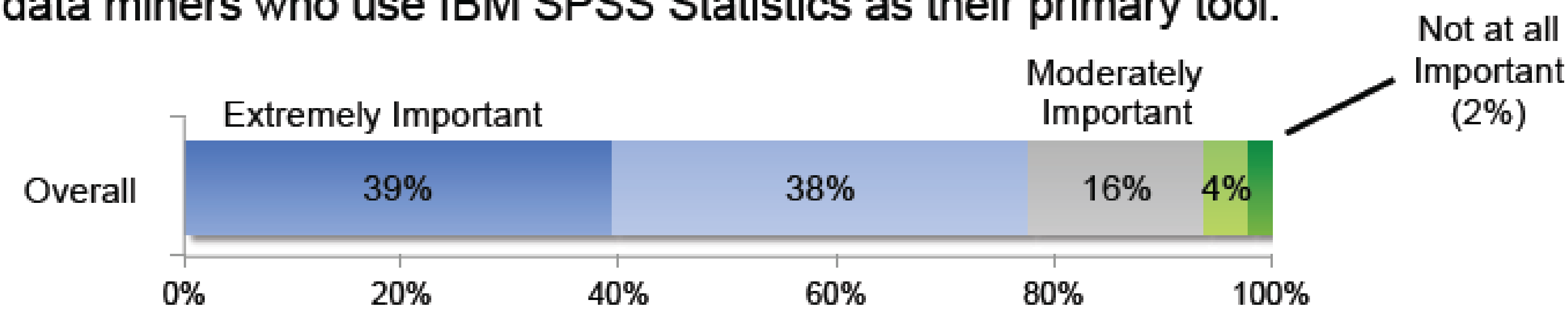


Question: What algorithms / analytic methods do you TYPICALLY use? (Select all that apply)

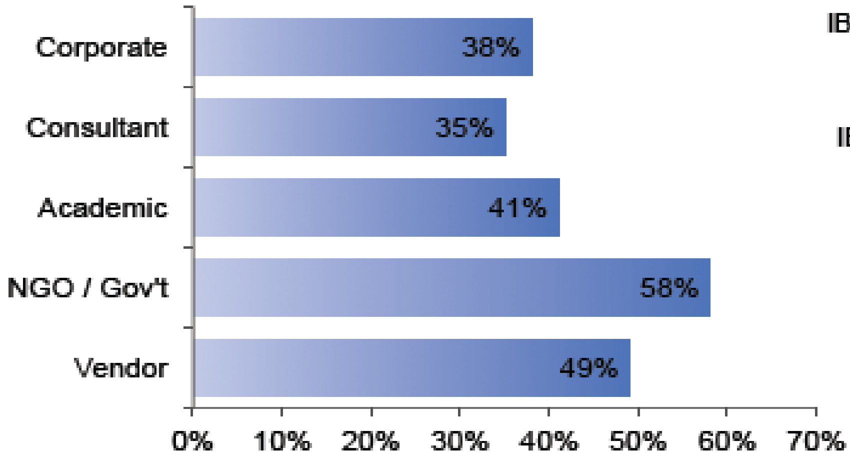
Source: Rexter Analytics

Importance of Model Explainability

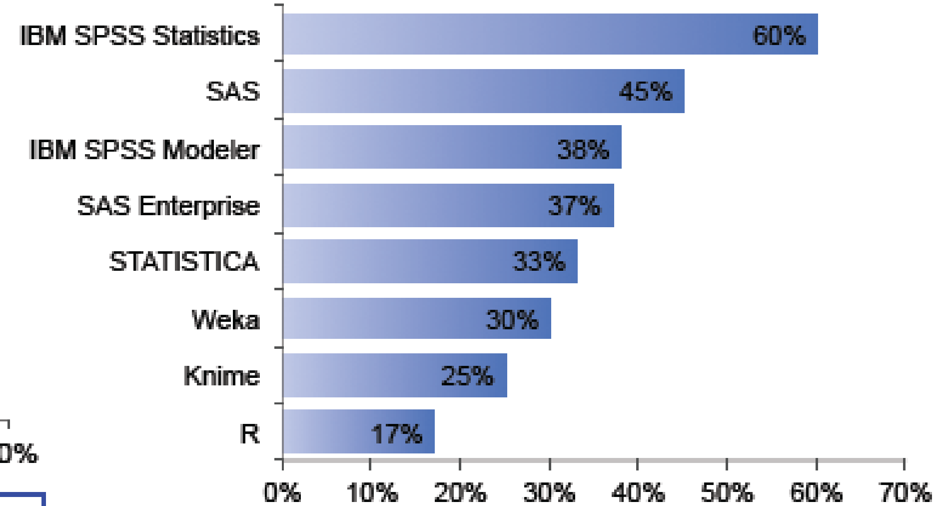
- Model explainability / transparency is important to most data miners.
- It is particularly important to data miners working in NGO / Gov't settings and to data miners who use IBM SPSS Statistics as their primary tool.



Percent indicating Model Explainability is "Extremely Important" by Employment Type



Percent indicating Model Explainability is "Extremely Important" by Primary Tool Used



Question: How important is model explainability / transparency to you?

Goals of this Course

- Learn data analysis methods with a focus on
  - problems in business and economics, and
  - causal inference, which is particularly challenging when analyzing human decision behavior

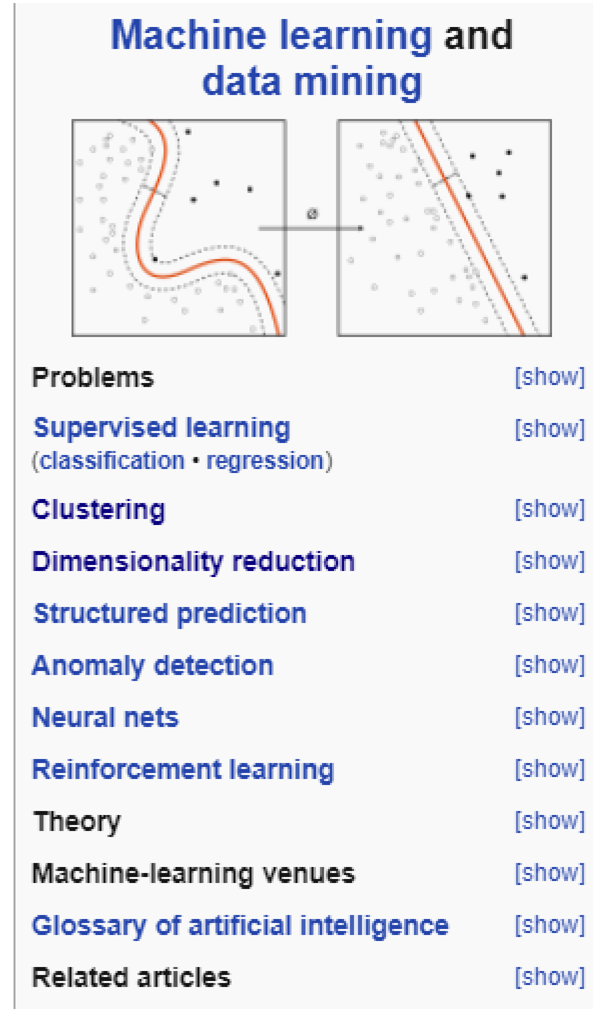
- Understand and interpret techniques for
  - *numerical prediction*
  - *classification*
  - *clustering* and *dimensionality reduction*

- Learn to analyze data with the *R programming language*
  - during the Analytics Cup you analyze data sets as part of the tutorials in small groups

Please note:
- This is an introductory course on data analysis.
- We expect that you had an introductory course in statistics and algorithms.
- Students in this class have very different prerequisites.

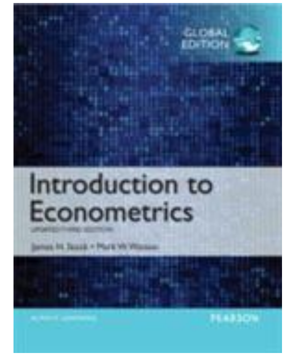
Course Content

- Introduction
- Regression Analysis
- Regression Diagnostics
- Logistic and Poisson Regression
- Naive Bayes and Bayesian Networks
- Decision Tree Classifiers
- Data Preparation and Causal Inference
- Model Selection and Learning Theory
- Ensemble Methods and Clustering
- High-Dimensional Problems
- Association Rules and Recommenders
- Neural Networks

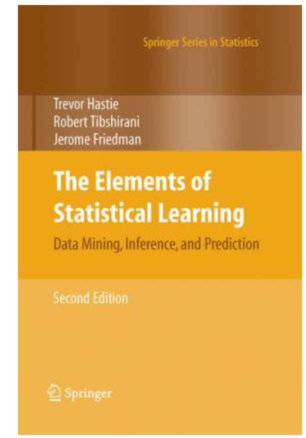


Check out related entries on wikipedia.org!

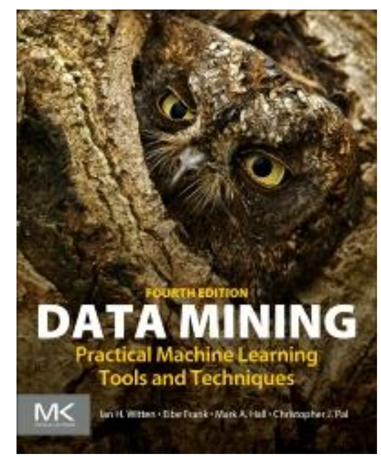
Primary Literature



- **Introduction to Econometrics**
  – Stock, James H., and Mark W. Watson



- **The Elements of Statistical Learning**
  – Trevor Hastie, Robert Tibshirani, Jerome Friedman, 2014
  – https://web.stanford.edu/~hastie/ElemStatLearn/



- **Data Mining: Practical Machine Learning Tools and Techniques**
  – Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher Pal, 2016
  – http://www.cs.waikato.ac.nz/ml/weka/book.html

- **An Introduction to Statistical Learning: With Applications in R**
  – Gareth James, Trevor Hastie, Robert Tibshirani, 2014
  – http://www-bcf.usc.edu/~gareth/ISL/

Parts of the slides have kindly been provided by
Prof. Dr. Gregory Piatetsky-Shapiro and Prof. Gary Parker
(Univ. of Connecticut).

This Course is Available to Students from …

- MSc Information Systems => **BSc** Information Systems
- MSc Informatics, MSc Games Engineering, MSc Data Engineering & Analytics
- MSc Management and Technology, MSc Consumer Affairs
- MSc Mathematics, MSc Mathematics in Operations Research

Students from IN, GE, and DE&A can choose one class in Analytics and one class in Machine Learning:

Analytics

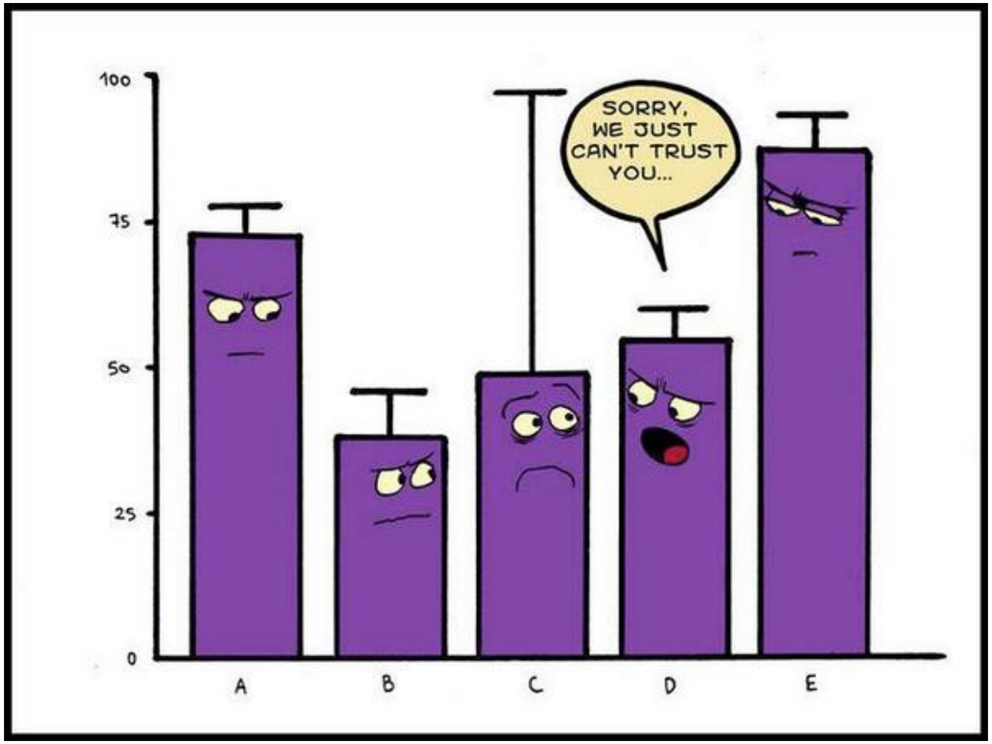
* Data Mining, IN2030, 2V, WS, Prof. Runkler
* Business Analytics, IN2028, 2V+2Ü, WS, Prof. Bichler
* Data Analysis and Visualization in R, IN2339, 2V+4Ü, WS, Prof. Gagneur

Machine Learning

* Statistical Modeling & Machine Learning, IN2332, 4V+4Ü, SS, Prof. Gagneur
* Machine Learning, IN2064, 4V+2Ü, WS, Prof. Günnemann

Agenda for Today

1. Understand what this course is all about
2. **Learn about organization, grading, and tutor groups**
3. Refresh basic statistical concepts



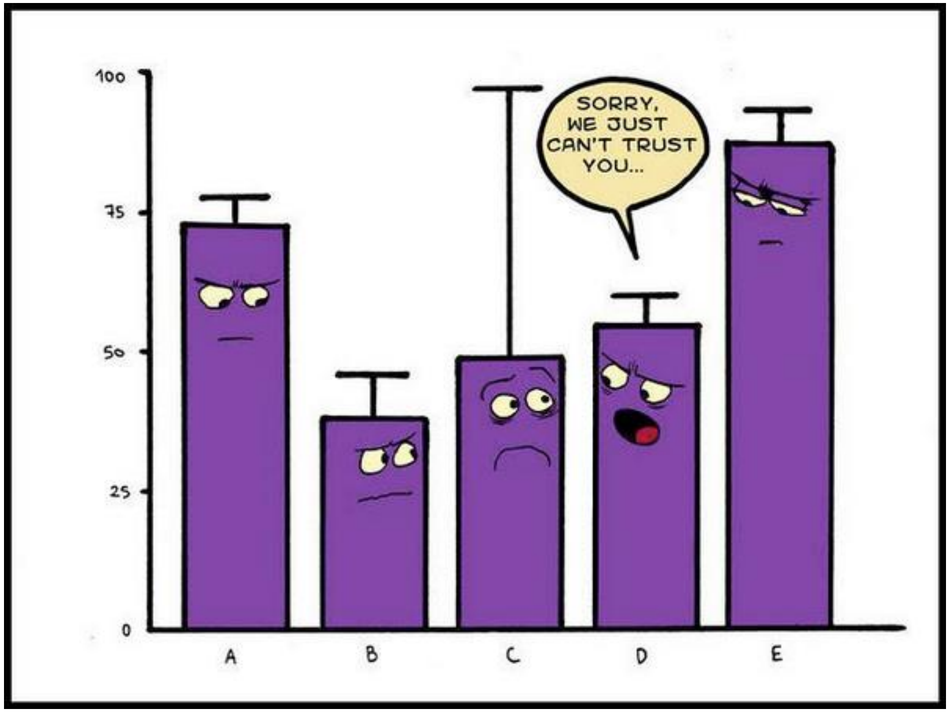
Agenda for Today

1. Understand what this course is all about
2. Learn about organization, grading, and tutor groups
3. Homework: **refresh basic statistical concepts**

In the first tutorials, we will recap important concepts from inferential statistics and introduce the R programming language, required for the rest of the course.

Check out http://onlinestatbook.com as an online source.

For this week please revisit the concepts on the following slides. Slides are only meant as a refresher.



Statistics

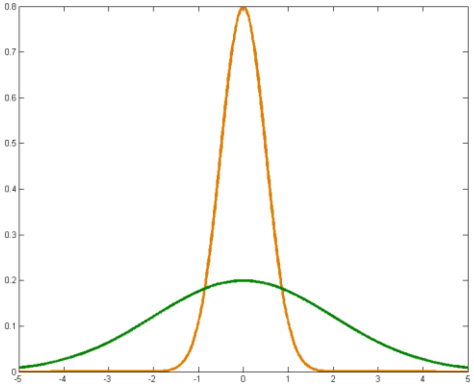
- **Descriptive statistics** can be used to summarize the data, either numerically or graphically, to describe the sample (e.g., mean and standard deviation)

- **Inferential statistics** is used to model patterns in the data, accounting for randomness and drawing inferences about the larger population. These inferences may take the form of
  - estimates of numerical characteristics (estimation),
  - answers to yes/no questions (hypothesis testing),
  - forecasting of future observations (forecasting),
  - descriptions of association (correlation), or
  - modeling of relationships (regression).

Random Variables

- X is a random variable if it represents a random draw from some population and is associated with a probability distribution
  - a discrete random variable can take on only selected values (e.g., Binomial or Poisson distributed)
  - a continuous random variable can take on any value in a real interval (e. g., uniform, Normal or Chi-Square distributions)
- For example, a Normal distribution, with mean μ and variance σ^2 is written as $N(\mu, \sigma^2)$ has a probability density function (pdf) of:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



The Standard Normal

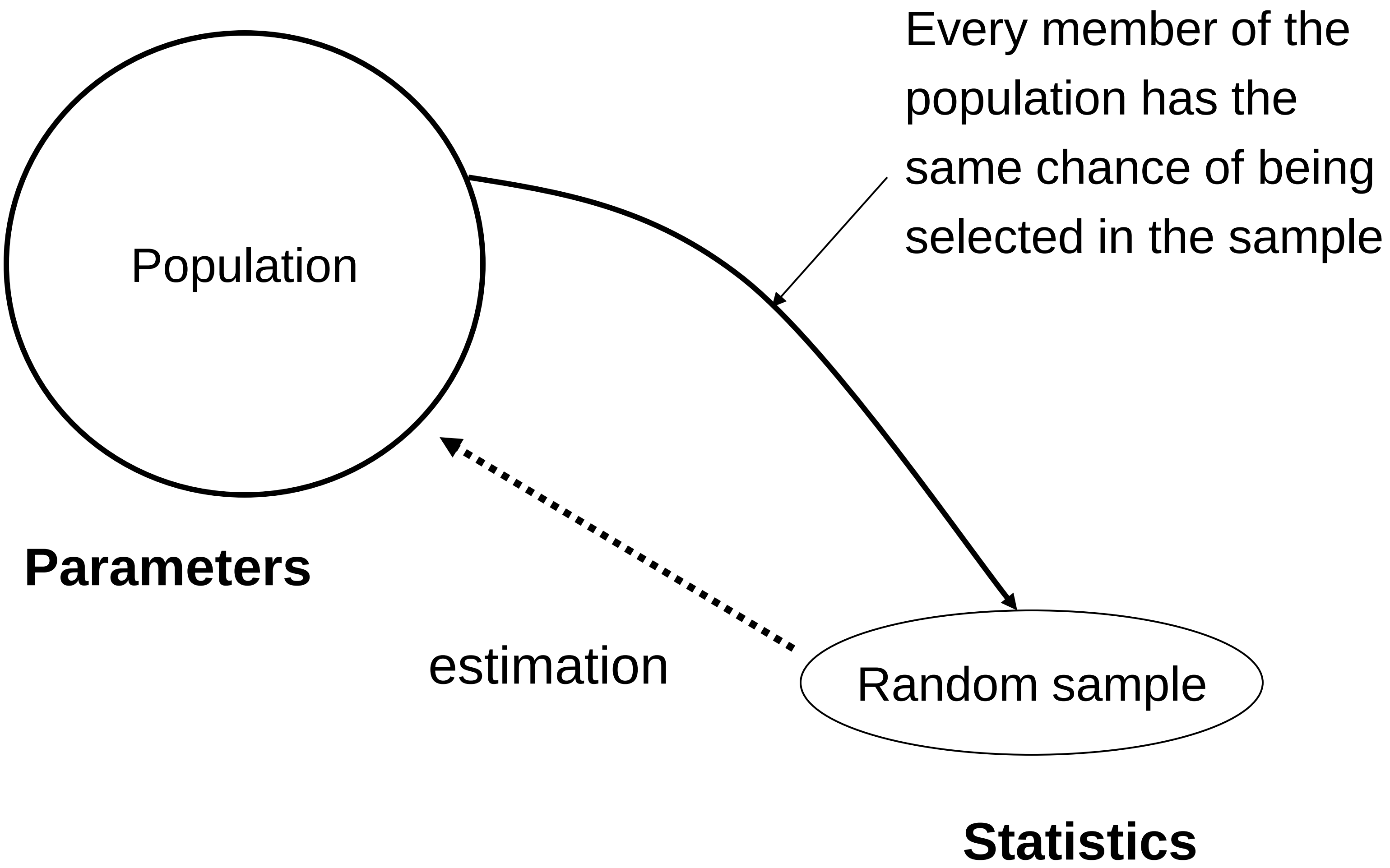
- Any random variable can be "standardized" by subtracting the mean, μ, and dividing by the standard deviation, σ , so $E(Z) = 0$, $Var(Z) = 1$
- Thus, the standard normal, $N(0,1)$, has the probability density function (pdf)

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

- For a pdf, $f(x)$, where $f(x)$ is $P(X = x)$, the cumulative distribution function (cdf), $F(x)$, is $P(X \leq x)$; $P(X > x) = 1 - F(x) = P(X < -x)$
- For the standard normal, $\varphi(z)$, the cdf is

$$\Phi(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-t^2}{2}} dt$$

Statistical Estimation



Expected Value of X: Population Mean $E(X)$

- The expected value of a probability weighted average of $X,\ E(X),$ is the mean or expected value of the distribution of X, denoted by μ_x
- Let $f(x_i)$ be the (discrete) probability that $X = x_i$, then

$$\mu_x = E(X) = \sum_{i=1}^{n} x_i f(x_i)\ or \int_{-\infty}^{\infty} x f(x) dx$$

Example: Expected Value

Students were surveyed and told to pick the number of hours that they play online games each day. The probability distribution is given below.

# of Hours *x*	Probability *P*(*x*)
0	.3
1	.4
2	.2
3	.1

Compute a “weighted average” by multiplying each possible time value by its probability and then adding the products

$$E(X) = 0(.3) + 1(.4) + 2(.2) + 3(.1) = 1.1$$

Random Samples and Sampling

- For a random variable X, repeated draws from the same population can be labeled as $X_1, X_2, \ldots, X_n$
- If every combination of n sample points has an equal chance of being selected, this is a random sample
- A random sample is a set of independent, identically distributed (i.i.d) random variables

Examples of Estimators

- Suppose we want to estimate the **population mean**
- Suppose we use the formula for $E(X)$, but substitute $1/n$ for $f(x_i)$ as the probability weight since each point has an equal chance of being included in the sample, then we can calculate the sample mean:

$$\bar{X} = \frac{1}{n} \sum_{1}^{n} X_i$$

- $\bar{X}$ describes the random variable for the **arithmetic mean of the sample**, while $\bar{x}$ is the mean of a particular realization of a sample

Estimators Should be Unbiased

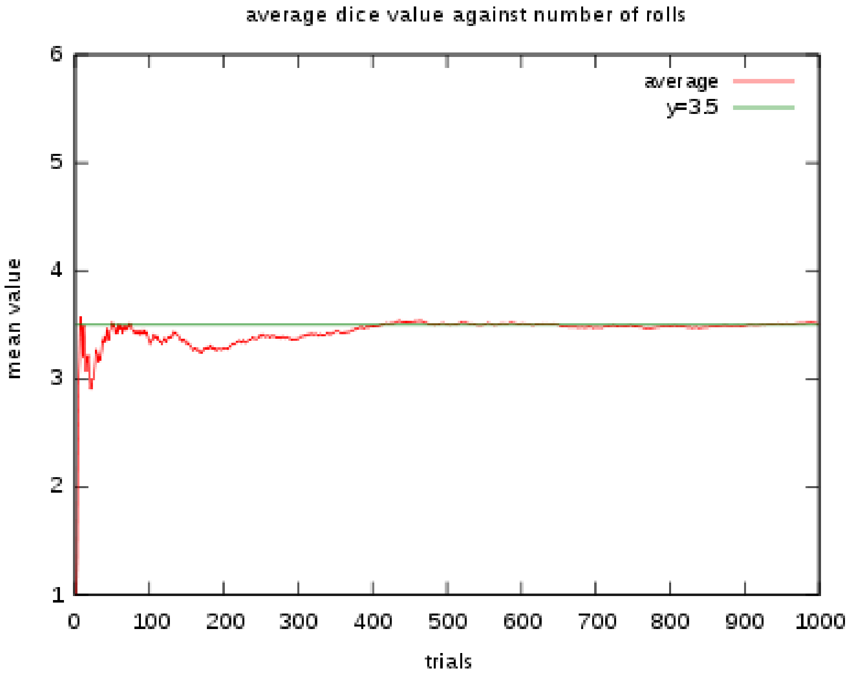
- An estimator (e.g., the arithmetic sample mean) is a statistic (a function of the observable sample data) that is used to estimate an unknown population parameter (e.g., the expected value)
- We want the estimator to be right, on average, i.e. unbiased.
- In our case, the sample mean $\bar{X}$ should be an unbiased estimator for the population mean μ_X:

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mu_X = \frac{1}{n} n\mu_X = \mu_X$$

Rolling a Dice

Wikipedia: Expected value of 3.5 as the
number of die rolls grows.



According to the **law of large numbers**, the sample mean converges to the expected value of the population distribution.

Law of Large Numbers

Proposition (Weak Law of Large Numbers)

$$\lim_{\{n \to \infty\}} \Pr(|\bar{X}_n - \mu| > \varepsilon) = 0$$

In other words: $n \to \infty$, $\bar{X}_n \to \mu$

Proof

Remember Chebyshev‘s inequality: No more than a certain fraction of values can be more than a certain distance from the mean.

$$\Pr(|X - \mu| > \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$$

Replace X by the sample mean of n i.i.d random variables $\bar{X}_n$

$$\left.\Pr(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{Var(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}\right\} \to 0 \text{ with } n \to \infty$$

$$\bar{X}_n = \frac{1}{n}\Sigma_{i=1}^{n} X_i,\ Var(\bar{X}_n) = \frac{1}{n^2}\Sigma_{i=1}^{n} Var(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Standard Error of the Sample Mean

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=0}^{n}(X_i)\right) = \frac{1}{n^2}\sum_{i=1}^{n}Var(X_i)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

$$\sigma_{\bar{X}} = SD(\bar{X}) = \sqrt{Var(\bar{X})} = \frac{\sigma}{\sqrt{n}}$$

Rule: $Var[aX + b] = a^2 Var[X]$

The **standard error of the sample mean** is an estimate of how far the sample mean is likely to be from the population mean. This means, the standard error of the mean tells you how accurate your estimate of the mean is likely to be.
The **standard deviation** of the sample is the degree to which individuals within the sample differ from the sample mean

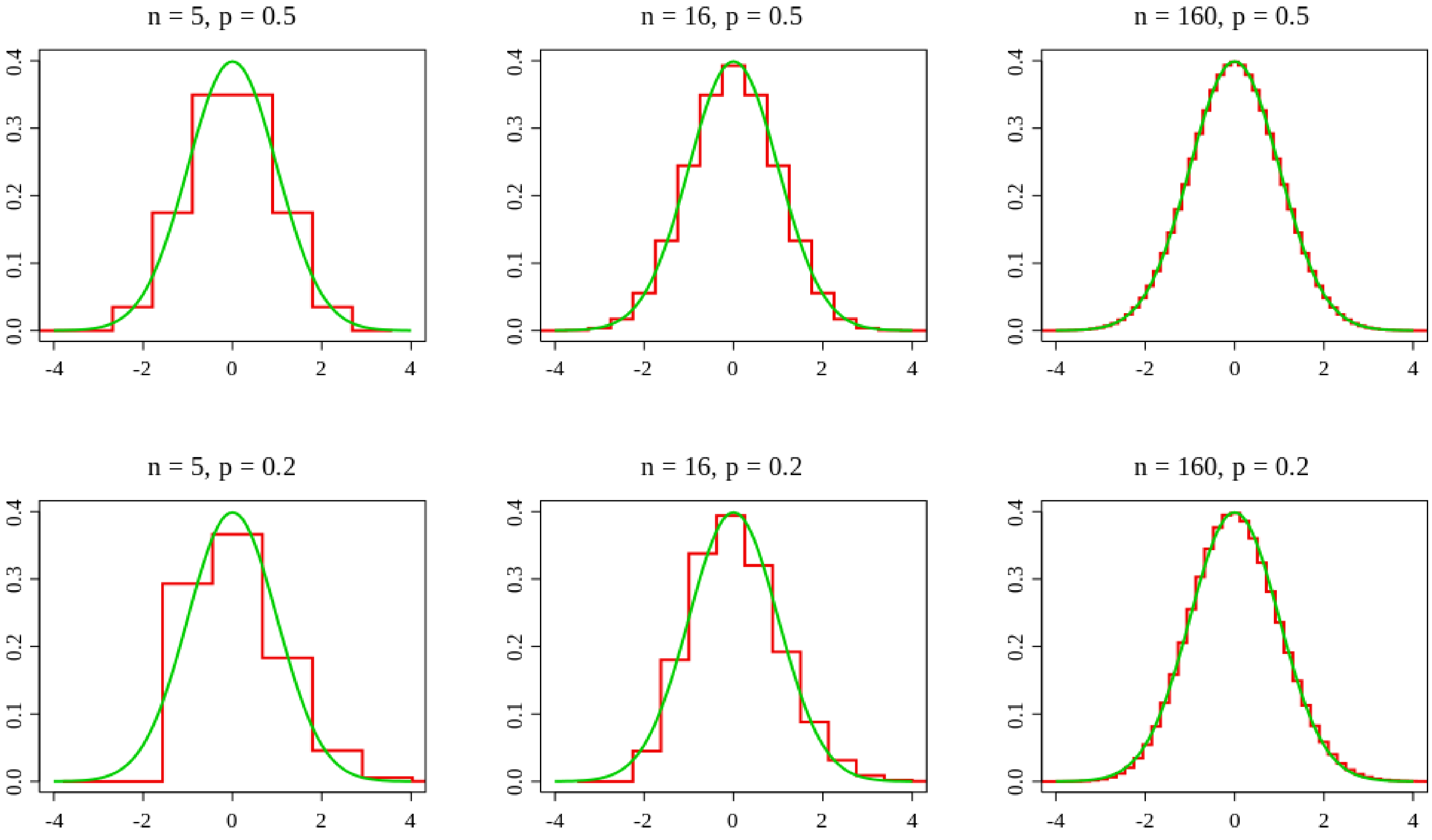
The Central Limit Theorem

- The *central limit theorem* states that the standardized average of any population of i.i.d. random variables X_i with mean μ_X and variance σ^2 is asymptotically $\sim N(0,1)$ as n goes to infinity.

$$Z = \frac{\bar{X} - \mu_X}{\sigma/\sqrt{n}} \sim N(0,1)$$

- This means, when independent random variables are added, their normalized sum tends toward a normal distribution even if the original variables themselves are not normally distributed.

Binomial Distributions and the Normal Distribution

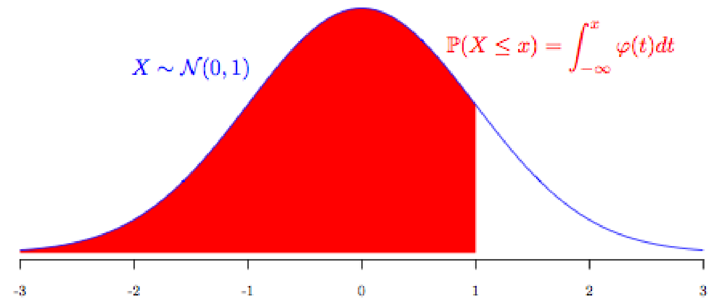


Source: Wikipedia

Summary: Sampling Distribution of the Mean

- We can say something about the distribution of sample statistics (such as the sample mean)
- The sample mean is a random variable, and consequently it has its own distribution and variance
- The distribution of sample means for different samples of a population is centered on the population mean
- The mean of the sample means is equal to the population mean
- If the population is normally distributed or when the sample size is large, sample means are distributed normally (Central Limit Theorem)

Question



What is the probability that a sample of 100 randomly selected elements with a mean of 300 or more gets selected if the true population mean is 288 and the population standard deviation is 60?

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990