

Online Encyclopedia of Statistical Science

By Stephanie Glen, July 2019 -- www.DataScienceCentral.com

This online book is intended for beginners, college students and professionals confronted with statistical analyses. It is also a refresher for professional statisticians. The book covers over 600 concepts, chosen out of more than 1,500 for their popularity. Entries are listed in alphabetical order, and broken down into 18 parts. Click on the label at the beginning of each part, to access the material covered in the part in question. In addition to numerous illustrations, we have added 100 topics not covered in our online series *Statistical Concepts Explained in Simple English*.

Below are direct links to each part:

[Part 1](#) | [Part 2](#) | [Part 3](#) | [Part 4](#) | [Part 5](#) | [Part 6](#) | [Part 7](#) | [Part 8](#)
[Part 9](#) | [Part 10](#) | [Part 11](#) | [Part 12](#) | [Part 13](#) | [Part 14](#) | [Part 15](#)
[Part 16](#) | [Part 17](#) | [Part 18](#)

We also included a number of visualizations from our series *Statistical Concepts Explained in One Picture*. These visualizations are found after the list of 600 concepts. The most recent version of this book is available [from this link](#), accessible to DSC members only. Topics not yet covered in this version can be found using our data science search engine: [click here](#) for the Data Science Central version and [here](#) for the StatisticsHowTo version.

About the author

Stephanie Glen is the founder of StatisticsHowTo.com, one of the most popular websites related to statistical education for non-statisticians, and also Editor at Data Science Central. Much of the material in this book points to articles on her website.

Other DSC Books

Data Science Central offers several books for free, available exclusively to members of our community. The following books are currently available:

- [Statistics: New Foundations, Toolbox, and Machine Learning Recipes](#)
- [Classification and Regression In a Weekend](#)
- [Azure Machine Learning in a Weekend](#)
- [Enterprise AI - An Application Perspective](#)
- [Applied Stochastic Processes](#)

Statistical Encyclopedia

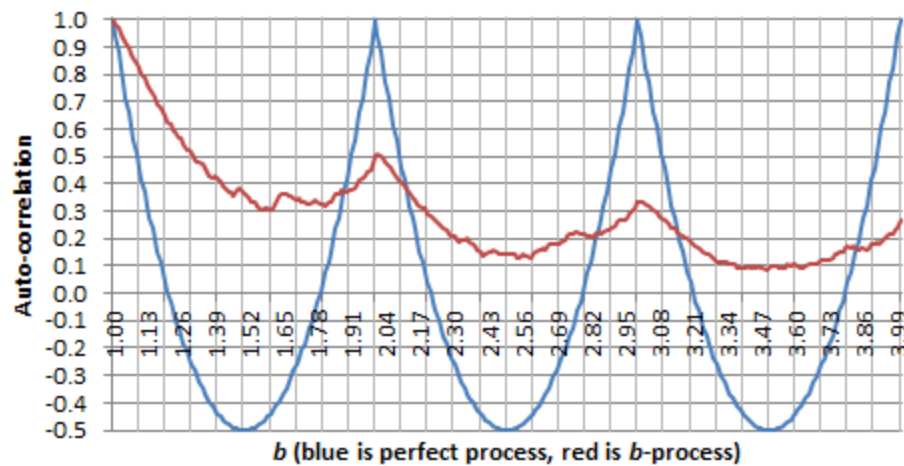
Here we provide online references to more than 600 statistical topics listed in alphabetical order. This listing is followed by several illustrations featuring statistical concepts explained in one picture, in the second half of this book. These “one picture” summaries (starting at page 27) include

- Bayes Theorem
- Naïve Bayes
- Confidence Intervals
- Three Types of Regression
- Support Vector Machines
- Hypothesis Tests
- Cross-validation
- Assumptions of Linear Regression
- R-squared
- Ensemble Methods
- Logistic Regression
- A/B Testing
- ROC Curve
- Sample Size Determination
- Determining the Number of Clusters
- EM Algorithm
- Azure Data Studio
- K Nearest Neighbors
- Prediction Algorithms
- Model Evaluation Techniques
- Regression Analysis

Part 1: [Ab - Be](#)

- 10% Condition in Statistics: What is it?
- 68 95 99.7 Rule in Statistics
- Absolute Error & Mean Absolute Error (MAE)
- Accuracy and Precision: Definition, Examples
- ADF -- Augmented Dickey Fuller Test

Lag-1 auto-correlation, *b*-process vs perfect process



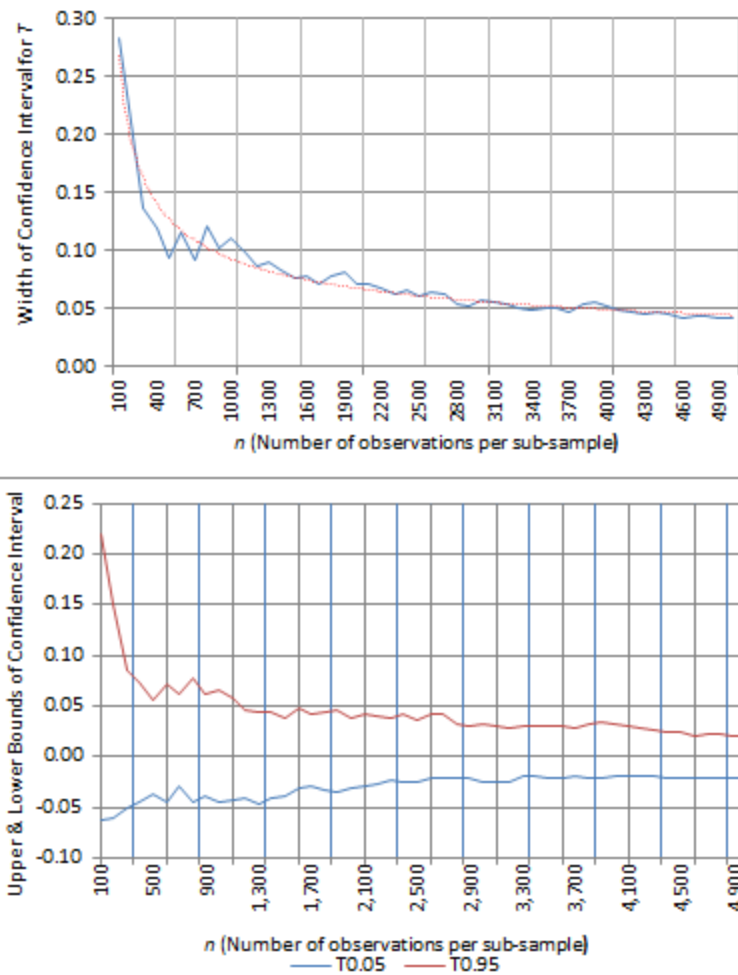
Some **autocorrelations** (source: [here](#))

- Adjusted R2 / Adjusted R-Squared: What is it used for?
- Akaike's Information Criterion: Definition, Formulas
- Alpha Level (Significance Level): What is it?
- Alternate Hypothesis in Statistics: What is it?
- ANCOVA: Analysis of Covariance
- ANOVA: Analysis of Variance
- ANOVA Excel 2013 (One-Way ANOVA) Easy Steps and Video
- Two Way ANOVA in Excel With Replication / Without Replication
- ANOVA Test: Definition, Types, Examples
- Area Between Two Z Values on Opposite Sides of Mean
- Area Principle in Statistics
- Area to the Right of a z score (How to Find it)
- Arithmetic Mean: What it is and How to Find it
- ARMA Model
- Assumption of Independence
- Assumption of Normality / Normality Test
- Assumptions and Conditions for Regression
- Attributable Risk / Attributable Proportion: Definition
- Attribute Variable / Passive Variable: Definition, Examples
- Autoregressive Model: Definition & The AR Process
- Average - Definition - How to Calculate Average
- Average Deviation (Average Absolute Deviation)
- Average Inter-Item Correlation: Definition, Example
- Balanced and Unbalanced Designs: Definition, Examples
- Segmented Bar Chart: Definition & Steps in Excel.
- Bar Chart / Bar Graph: Examples, Excel Steps & Stacked Graphs

- Bartlett's Test: Definition and Examples
- Bayesian Information Criterion (BIC) / Schwarz Criterion
- Bayes' Theorem Problems, Definition and Examples
- Bell Curve (Normal Curve): Definition
- Benjamini-Hochberg Procedure
- Bernoulli Distribution: Definition and Examples
- Bessel's Correction: Why Use N-1 For Variance/Standard Deviation?

Part 2: [Be - Ca](#)

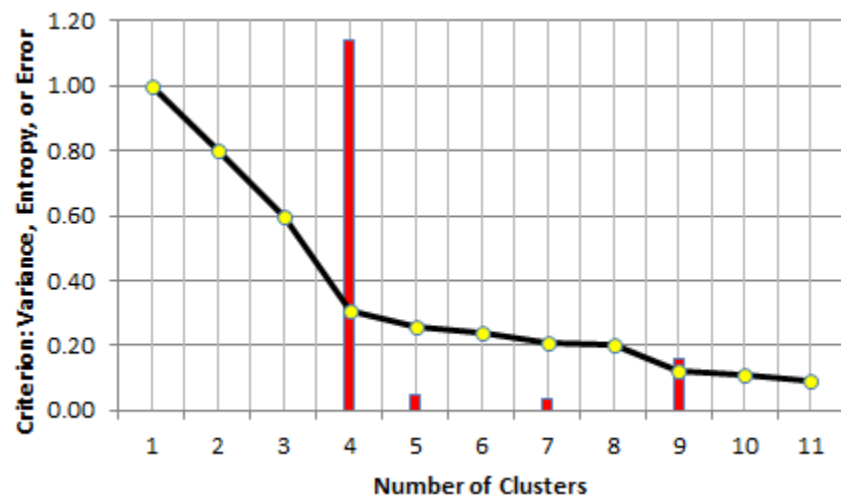
- Beta Weight: Definition, Uses
- Bias in Statistics: Definition, Selection Bias & Survivorship Bias
- Binary Variable: Definition, Examples
- Binomial Coefficient
- Binomial Confidence Interval
- Binomial Probability TI 89: Easy Calculation Steps
- Binomial test / Exact Binomial Test
- Bimodal Distribution: What is it?
- Binomial Theorem: Simple Definition, Formula, Step by Step Videos
- Binomial Distribution: Formula, What it is, and how to use it
- Binomial Experiment: Rules, Examples, and Simple Steps
- Find the Mean of the Probability Distribution / Binomial
- Normal Approximation to the Binomial
- Binomial Distribution Table
- Bivariate Analysis Definition & Example
- Bivariate Normal Distribution / Multivariate Normal (Overview)
- Blinding vs. Double Blind in Statistics: Simple Definition, Examples
- Bootstrap Sample: Definition, Example
- Bowley Skewness: Definition, Formula, Alternate Formula
- Box Cox Transformation
- Box's M Test: Definition



Bootstrap-computed confidence intervals (see [here](#))

- Box Plot (Box and Whiskers)
- Bray Curtis Dissimilarity
- Breusch-Pagan-Godfrey Test: Definition
- Brier Score: Definition, Examples
- Business Statistics: Definition, Step by Step Articles, Videos
- Chi Square P Value Excel: Easy Steps, Video
- Relative Risk and Absolute Risk: Definition and Examples
- Binomial Distribution Calculator
- Expected Value Calculator
- Interquartile Range Calculator
- Linear Regression Calculator
- Permutation Calculator / Combination Calculator
- Variance and Standard Deviation Calculator

Part 3: [Ca - Co](#)



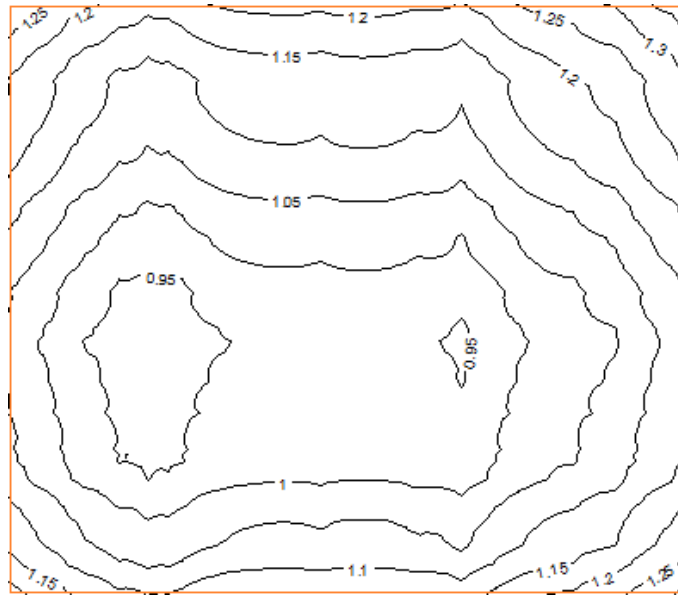
*Elbow rule to determine the number of **clusters** (source: [here](#))*

- Case-Control Study: Definition, Real Life Examples
- Case Studies: Case Study Definition and Steps
- Cauchy-Schwarz Inequality: Simple Definition, Example & Proof
- Causation Definition & Causation vs. Correlation
- Categorical Variable in Statistics: What is it?
- Center of a Distribution: Definition, How to Find it
- Central Limit Theorem: Definition and Examples in Easy Steps
- Central Tendency (Measures of Location): Definition and Examples
- Chauvenet's Criterion
- Chebyshev's Theorem / Inequality: Calculate it by Hand / Excel
- Chi Squared Table (Right Tail)
- Chi-Square Statistic: How to Calculate It / Distribution
- Chi-square Test for Normality
- Choose Bin Sizes for Histograms in Easy Steps + Sturge's Rule
- Chow Test: Definition & Examples
- Classical Probability: Definition and Examples
- Clustered Standard Errors: Definition
- Cluster Sampling in Statistics: Definition, Types
- Coefficient Definition: Different Types in Statistics and Math
- Coefficient of Association: Definition, Types, Examples
- Coefficient of Dispersion
- Coefficient of Determination (R Squared): Definition, Calculation
- How to Find a Coefficient of Variation
- Cohen's D: Definition, Examples, Formulas

- Cohen's Kappa Statistic
- Cohort Study (Retrospective, Prospective): Definition, Examples
- Cointegration: Definition, Examples, Tests
- Combined Mean: Definition, Examples
- Comparison of Means
- Concordant Pairs and Discordant Pairs
- Concurrent Validity Definition and Examples
- Conditional Probability: Definition & Examples
- Conditional Distribution: Definition and Examples
- Conditional Expectation: Definition & Step by Step Example
- Confidence Interval: How to Find a Confidence Interval: The Easy Way!
- Confidence Level: What is it?
- Conservative in Statistics
- Consistent Estimator: Consistency Definition & Examples
- Construct Validity: Simple Definition, Statistics Used

Part 4: [Co - Cu](#)

- Content Validity (Logical or Rational Validity)
- Contingency Coefficient: Definition
- Contingency Table: What is it used for?
- Continuity Correction Factor: What is it?
- Continuous Probability Distribution
- Continuous Variable Definition (Continuous Data)
- Contour Plots: Definition, Examples
- Control Group: Definition, Examples and Types
- Control Variable: Simple Definition
- Convenience Sampling (Accidental Sampling): Definition, Examples
- Convergent Validity and Discriminant Validity: Definition, Examples
- Cook's Distance / Cook's D: Definition, Interpretation
- Correlation Coefficient: Simple Definition, Formula, Easy Calculati...
- Correlation in Statistics: Correlation Analysis Explained
- Correlation Matrix: Definition
- Counterbalancing in Research
- Covariance in Statistics: What is it? Example
- Covariate Definition in Statistics
- Cramer-Rao Lower Bound
- Criterion Validity: Definition, Types of Validity
- Critical Values: Find a Critical Value in Any Tail
- Criterion Variable: Definition, Use and Examples
- Critical Z Value TI 83: Easy Steps for the InvNorm Function

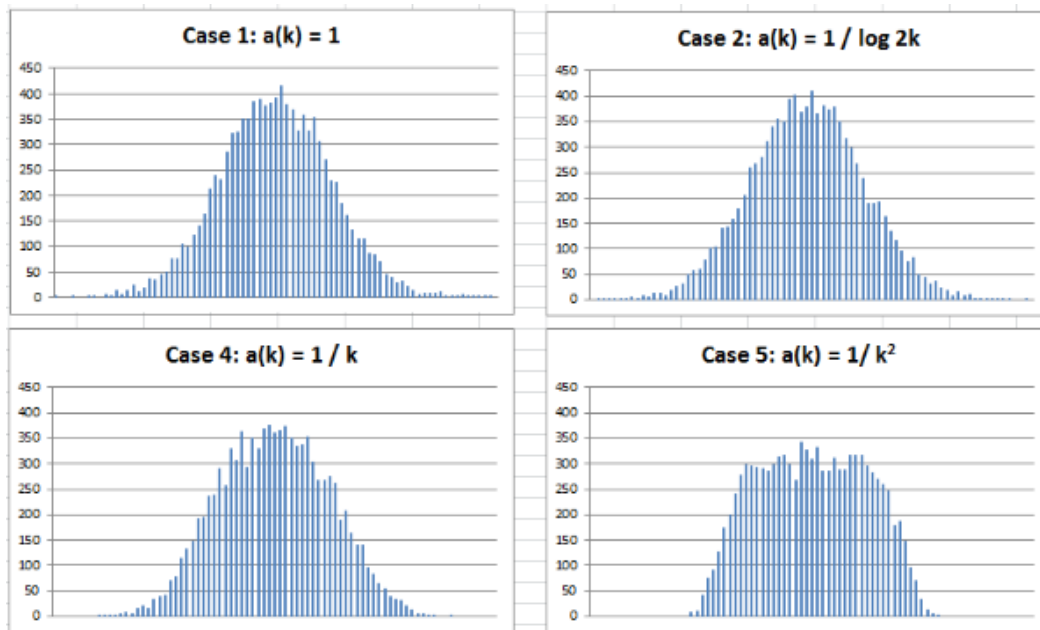


Contour plot (source: [here](#))

- Cronbach's Alpha: Simple Definition, Use and Interpretation
- C-Statistic: Definition, Examples, Weighting and Significance
- Cumulative Distribution Function CDF

Part 5: [Cu - Du](#)

- Cumulative Frequency Distribution: Simple Definition, Easy Steps
- Cumulative Frequency Table in Excel: Easy Steps
- Data Analysis & Exploratory Data Analysis (EDA)
- Data Collection Methods
- What is a Decile?
- Decision Tree: Definition and Examples
- Degrees of Freedom: What are they?
- Density Curve Examples
- Dependent Events and Independent Events
- Dependent Variable: Definition and Examples
- Descriptive Statistics: Definition & Charts and Graphs
- Design Effect: Definition, Examples
- Deterministic: Definition and Examples
- Detrend Data
- Dichotomous Variable: Definition
- Dice Roll Probability: 6 Sided Dice
- Difference Between BinomPDF and BinomCDF
- Curse of Dimensionality & High Dimensional Data



Density curve and central limit theorem (source: [here](#))

- Acyclic Graph & Directed Acyclic Graph: Definition, Examples
- Direction of Association in Statistics: What is it?
- Discrete Probability Distribution: Definition & Examples
- What is a Discrete Variable in Statistics?
- Discrete vs Continuous variables: How to Tell the Difference
- Disjoint Events: Definition, Examples
- Dispersion / Measures of Dispersion: Definition
- Dixon's Q Test: Definition, Step by Step Examples + Q Critical Values Tables
- Dot Plot in Statistics: What it is and How to read one
- Double Sampling: Simple Definition, Types
- Dummy Variables / Indicator Variable: Simple Definition, Examples
- Duncan's Multiple Range Test (MRT)
- Dunnett's Test / Dunnett's Method: Definition
- Dunn's test: Definition

Part 6: [Du - Ex](#)

- Durbin Watson Test & Test Statistic
- Ecological Validity: Definition and Examples
- EM Algorithm (Expectation-maximization): Simple Definition
- Empirical Distribution Function / Empirical CDF
- Empirical Rule: What is it?

$$E[R_n] = \frac{1}{\lambda} \cdot \sum_{k=1}^{n-1} \frac{1}{k},$$

$$\text{Var}[R_n] = \frac{1}{\lambda^2} \cdot \sum_{k=1}^{n-1} \frac{1}{k^2}.$$

Range of **exponential distribution** (source: [here](#))

- Endogenous Variable and Exogenous Variable: Definition and Classifying
- Erlang Distribution: Definition, Examples
- Error Term: Definition and Examples
- Estimator: Simple Definition and Examples
- Eta Squared / Partial Eta Squared
- Excel Data Analysis ToolPak: Easy Steps and Video 2016-2007
- Excel Multiple Regression (Polynomial Regression)
- Excel PERCENTRANK Function, PERCENTILE & RANK
- Excel Regression Analysis Output Explained
- Expected Frequency: Definition, Formula, Calculation
- Expected Monetary Value EMV: Definition & Example
- Expected Value in Statistics: Definition and Calculations
- Experimental Design
- Extrapolation & Interpolation: What are they?
- Confounding Variable: Simple Definition and Example
- Fixed Effects / Random Effects / Mixed Models and Omitted Variable ...
- Experimental Group (Treatment Group): Definition, Examples
- Expert Sampling / Judgment Sampling
- Explanatory Variable & Response Variable: Simple Definition and...
- Exponential Distribution / Negative Exponential: Definition, Examples
- Exponential Smoothing: Definition of Simple, Double and Triple
- External Validity Definition & Examples
- Extraneous Variable Simple Definition

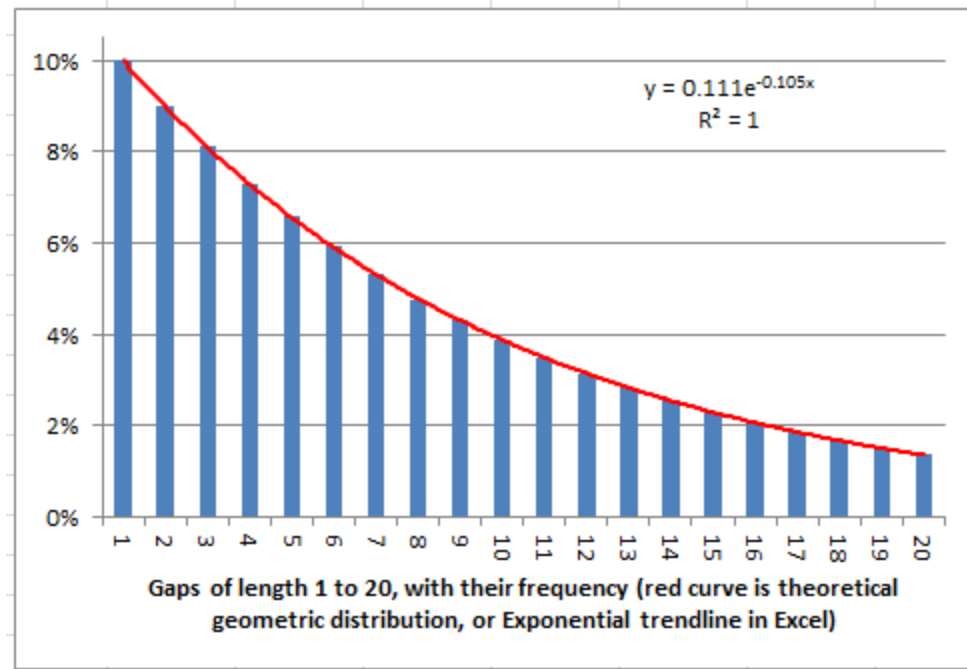
Part 7: [Fa - Fr](#)

- F Statistic / F Value: Definition and How to Run an F-Test
- F Table for alpha levels from .01 to .10
- F Test: Simple Definition, Step by Step Examples
- Face Validity: Definition and Examples
- Factor analysis: Easy Definition

- Factorial of a Number (!): Solve Factorials in Easy Steps
- False Discovery Rate: Simple Definition, Adjusting for FDR
- False Positive and False Negative: Definition and Examples
- Familywise Error Rate (Alpha Inflation): Definition
- Fat Tail Distribution: Definition, Examples
- 5 Number Summary in Excel: Easy Steps with Video
- Outliers: Finding Them in Data, Formula, Examples. Easy Steps
- How to Find Pearson's Coefficient of Skewness in Excel
- Pooled Sample Standard Error: How to Calculate it
- Regression Slope Intercept: How to Find it in Easy Steps
- Standard Error Excel 2013 in Easy Steps
- Standard Error of Regression Slope
- How to Find t Critical Value on TI 83
- Variance in Minitab: How to Find it
- Finite Population Correction Factor FPC: Formula, Examples
- Fisher Z-Transformation
- Fleiss' Kappa
- Fmax / Hartley's Test: Definition, Step by Step Example, Table
- Fractile Definition Usage and How to Calculate
- Frequency Distribution Table: Examples, How to Make One
- Frequency Distribution Table in Excel -- Easy Steps!
- Frequency Polygon: Definition and How to Make One
- Friedman's Test / Two Way Analysis of Variance by Ranks

Part 8: [Fu - Ho](#)

- Fundamental Counting Principle (The Multiplication Counting Rule)
- Funnel Plot: Definition, Examples
- Gamma Coefficient (Goodman and Kruskal's Gamma) & Yule's Q
- Gamma Distribution: Definition, PDF, Finding in Excel
- Gamma Function: Definition, Properties
- Gauss Markov Theorem & Assumptions
- General Linear Model (GLM): Simple Definition / Overview
- Geometric Distribution: Definition & Example
- Geometric Mean: Definition, Examples, Formula, Uses
- Goodness of Fit Test: What is it?
- Granger Causality: Definition, Running the Test
- Greatest Possible Error: Easy Definition, Step by Step Examples
- Grounded Theory: Simple Definition and Examples
- Grouped Data / Ungrouped Data: Definition, Examples
- Grubbs' Test for Outliers (Maximum Normed Residual Test)



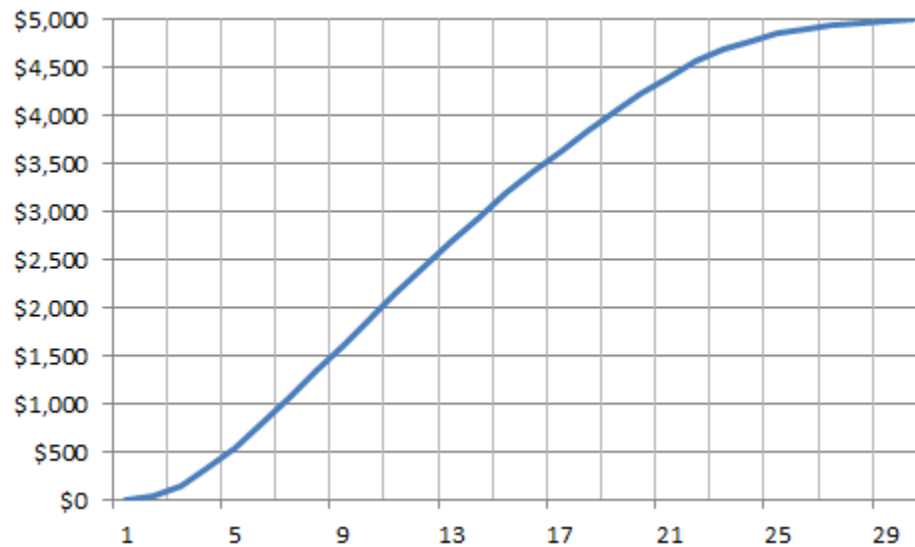
Geometric distribution and gap test (source: [here](#))

- Guttman Scale (Cumulative Scale): Definition & Examples
- Hamiltonian Cycle: Simple Definition and Example
- Haphazard Sampling: Definition, Examples, Advantages/Disadvantages
- Harmonic Mean: Definition, Formula, Examples
- Hausman Test for Endogeneity (Hausman Specification Test)
- Hazard Ratio: Definition, Examples & Log of the Hazard
- Heavy Tailed Distribution & Light Tailed Distribution: Definiti...
- Hedges' g: Definition, Formula
- Heterogeneity and Heterogeneous Data in Statistics
- Heteroscedasticity: Simple Definition and Examples
- Hidden Markov Model: Simple Definition & Overview
- Hierarchical Clustering / Dendrogram: Simple Definition, Examples
- Histogram in Excel: Easy Steps 2016, 2013, 2010-2007
- Histogram: Make a Chart in Easy Steps
- Holm-Bonferroni Method: Step by Step
- Homogeneity, Homogeneous Data & Homogeneous Sampling
- Homoscedasticity / Homogeneity of Variance/ Assumption of Equal Var...
- Hosmer-Lemeshow Test: Definition

Part 9: [Ho - Kr](#)

- Hotelling's T-Squared: Simple Definition

- Normal Distribution Probability in Excel: All Versions up to 2016
- Hypergeometric Distribution: Examples and Formula



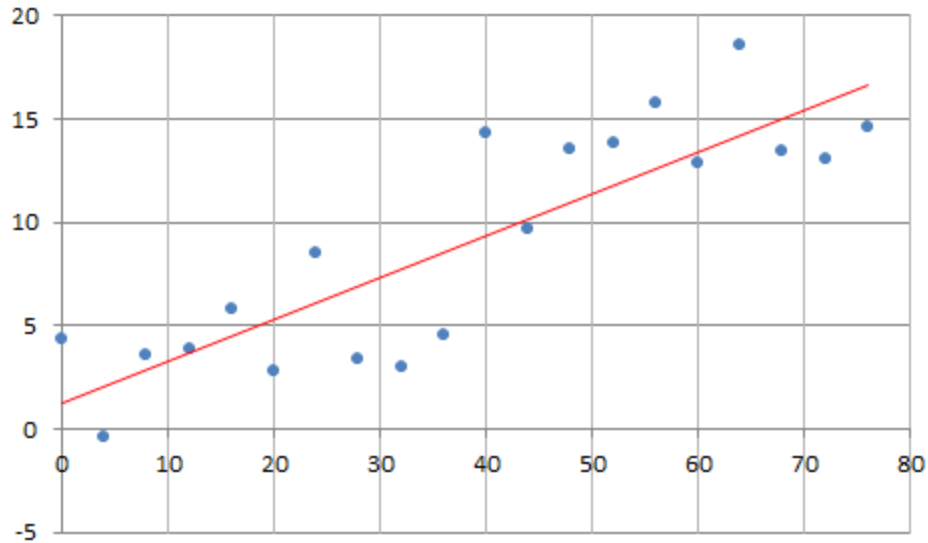
Growth model with S-curve (see [here](#))

- Hypothesis Testing
- Hypothesis Test in Excel for the Population Mean (Large Sample)
- Idempotent Matrix: Definition, Examples
- IID Statistics: Independent and Identically Distributed Definition ...
- Independent Random Variables: Definition, Examples
- Independent Samples T Test (Unpaired Samples): Definition, Running
- Independent Variable (Treatment Variable) Definition and Uses
- Index Number: Simple Definition
- Inferential Statistics: Definition, Uses
- Information Bias (Observation Bias): Definition, Examples
- Instrumental Variable: Definition & Overview
- Interaction Effect, Statistical Interactions & Interacting Vari...
- Internal Consistency Reliability: Definition, Examples
- Internal Validity: Definition and Examples
- Inter-rater Reliability IRR: Definition, Calculation
- Interquartile Range (IQR): What it is and How to Find it
- Interval Estimate: Definition, Examples
- Interval Scale & Interval Variable: Simple Definition & Exa...
- Intervening variable: Simple Definition & Examples
- Intraclass Correlation
- Inverse Gamma Distribution: Definition, Mean, Variance, PDF

- Inverse Normal Distribution
- Jaccard Index / Similarity Coefficient
- Jarque-Bera Test
- Joint Probability and Joint Distributions: Definition, Examples
- Kaiser-Meyer-Olkin (KMO) Test for Sampling Adequacy
- Kendall's Tau (Kendall Rank Correlation Coefficient)
- Kolmogorov-Smirnov Goodness of Fit Test
- KPSS Test: Definition and Interpretation
- Krippendorff's Alpha Reliability Estimate: Simple Definition

Part 10: [Kr - Me](#)

- Kruskal Wallis H Test: Definition, Examples & Assumptions
- Kuder-Richardson 20 (KR-20) & 21 (KR-21)
- Kurtosis: Definition, Leptokurtic, Platykurtic
- Lasso Regression: Simple Definition
- Law of Large Numbers / Law of Averages
- Least Squares Regression Line: Ordinary and Partial
- Levels in Statistics
- Levene Test for Equality of Variances
- Likelihood Function: Overview / Simple Definition
- Likelihood Ratio (Medicine): Basic Definition, Interpretation
- Likelihood-Ratio Tests (Probability and Mathematical Statistics)
- Likert Scale Definition and Examples
- Limiting Distribution (Asymptotic Distribution): Definition and Exa...
- Linear Relationship: Definition, Examples
- Linear Regression: Simple Steps and Video - Find the Equation, Coef...
- Vector Transformation & Linear Transformation
- Line Graph - Definition and Easy Steps to Make One
- Line of Best Fit: What it is, How to Find it
- Lognormal Distribution: Definition, Examples
- Log Odds: Definition and Worked Statistics Problems
- Lowess Smoothing in Statistics: What is it?
- Mahalanobis Distance: Simple Definition, Examples
- Main Effect: Definition and Examples
- Manipulated Variable: Definition
- Mann Kendall Trend Test: Definition, Running the Test
- Mann Whitney U Test
- Margin of Error: Definition, How to Calculate in Easy Steps
- Marginal Distribution



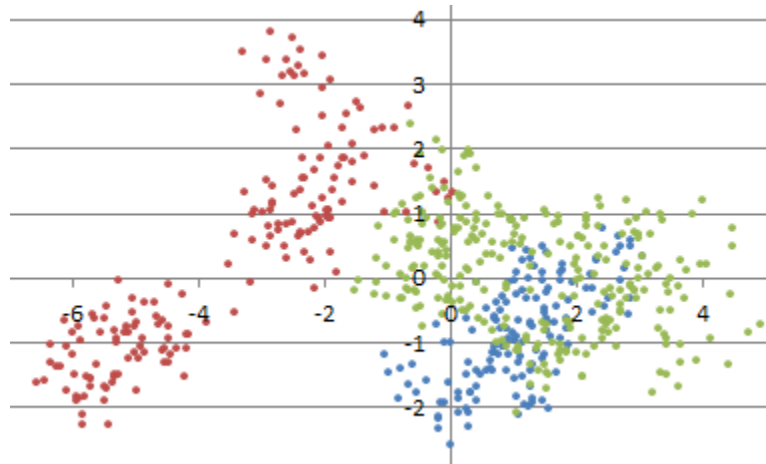
How to **normalize** R-squared (source: [here](#))

- Marginal Effects: Definition
- Matched Samples: Definition, Examples
- Matrices and Matrix Algebra
- Maximum Variation Sampling
- McNemar Test Definition, Examples, Calculation
- Mean, Median, Mode: What They Are, How to Find Them
- Mean Difference / Difference in Means (MD)
- Mean Error: Definition
- Mean Square Between: Definition & Examples

Part 11: [Me - Na](#)

- Mean Squared Error: Definition and Example
- Median: Definition, How to Find it, Formula
- Measurement Error (Observational Error)
- Measurement Variable: Simple Definition & Examples
- Measures of Spread: Definitions, Examples
- Measures of Variation: Definition, Types and Examples
- Median Absolute Deviation
- Memoryless Property: Definition and Examples in Statistics
- Method of Moments Definition and Example
- Metropolis-Hastings Algorithm / Metropolis Algorithm
- Middle Fifty in Statistics: What is it?
- Midpoint / Class Mark (in Statistics)
- Midrange: Definition and How to Find the Midrange
- Minimum Spanning Tree: Definition, Examples, Prim's Algorithm

- Misleading Graphs: Real Life Examples
- Misleading Statistics Examples in Advertising and The News
- Mixture Distribution: Definition and Examples
- How to Find the Mode or Modal Value



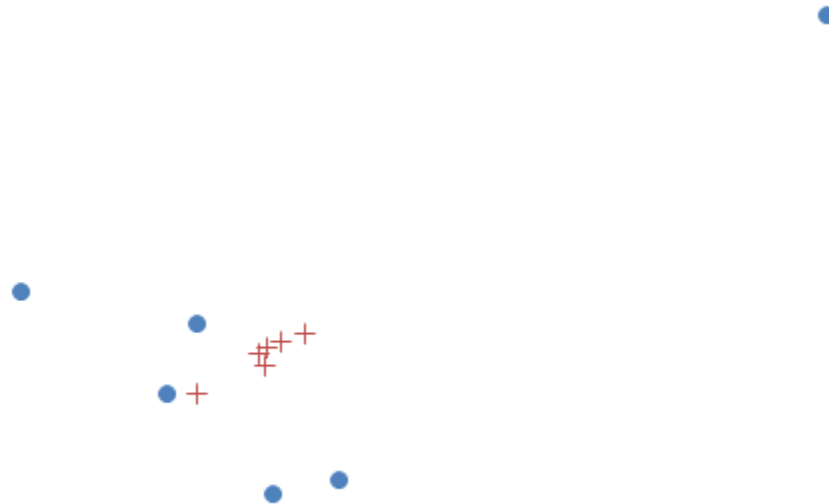
2-D mixture model (source: [here](#))

- Model Misspecification
- Moderating Variable (or Moderator)
- Moment in Statistics: Definition, Examples
- Moment Generating Function MGF: Definition, Examples
- Monotonic Relationship: Definition
- Monty Hall Problem: Solution Explained Simply
- Moran's I: Definition, Examples
- Moving Average: What it is and How to Calculate it
- Multicollinearity: Definition, Causes, Examples
- Multidimensional Scaling: Definition, Overview, Examples
- Multimodal Distribution Definition and Examples
- Multinomial Distribution: Definition, Examples
- Multiplication Rule Probability: Definition, Examples
- Multistage Sampling: Definition, Real Life Examples
- Multivariate Analysis
- Mutually Exclusive Event: Definition, Examples, Unions
- Mutually Inclusive Events: Definition, Examples
- Nash Equilibrium: Simple Definition and Examples

Part 12: [Ne - Pa](#)

- Negative Binomial Experiment / Distribution: Definition, Examples

- Nested Model, ANOVA and Factors: Simple Definitions and Examples
- Nominal Ordinal Interval Ratio: Examples
- Nominal Variable: Definition and Examples
- Differential & Non-Differential Misclassification



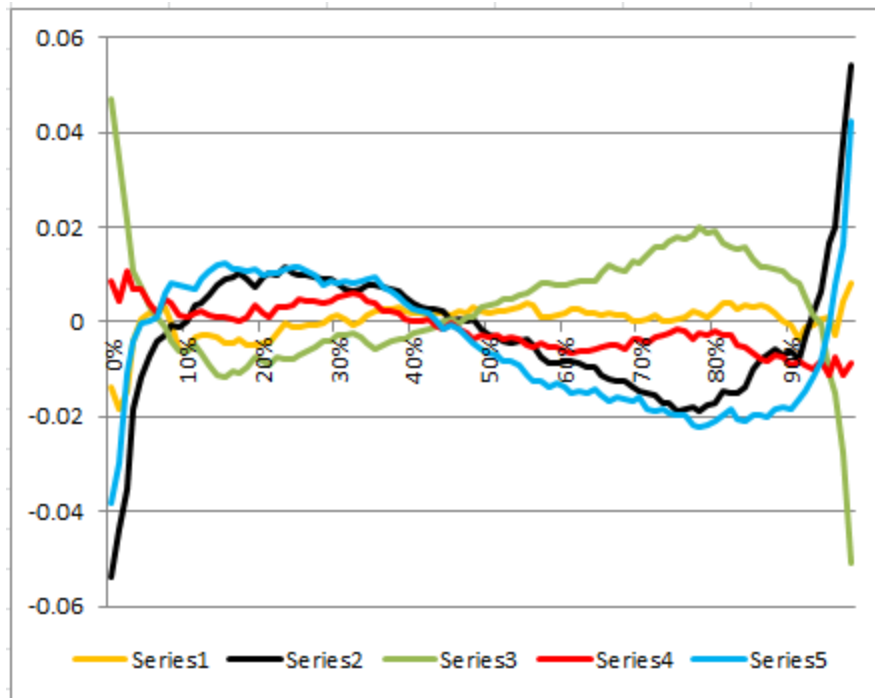
*Impact of **outliers** on centroid (source: [here](#))*

- Nonlinear Regression: Simple Definition & Examples
- Non-Probability Sampling: Definition, Types
- Non Response Bias: Definition, Examples
- Normalized Data / Normalization
- Area Under a Normal Curve: Find it in Easy Steps with Video
- Non Normal Distribution
- Normal Distributions: Definition, Word Problems
- Normal Probability Plot: Definition, Examples
- Normal Probability Practice Problems and Answers
- Nuisance Variable & Nuisance Parameter: Definition, Examples
- Null Hypothesis Definition and Examples, How to State
- Support or Reject Null Hypothesis in Easy Steps
- Number Needed to Harm NNH: Definition
- Observation in Statistics: Simple Definition & Examples
- Observer Bias / Research or Experimenter Bias: Definition, Examples...
- Odds Ratio Calculation and Interpretation
- Ogive Graph / Cumulative Frequency Polygon in Easy Steps
- Omega Squared: Definition
- One Sample T Test: How to Run It, Step by Step
- One Sample Z Test: How to Run One

- Open Ended Distribution: Definition and Examples
- Order Effects: Definition, Examples and Solutions
- Order of Integration (Time Series): Simple Definition / Overview
- Order Statistics: Simple Definition, Examples
- Ordinal Numbers, Variables and Data: Definition and Examples
- Pairwise Independent, Mutually Independent: Definition, Example
- Parallel Design / Parallel Group Study
- Parallel Forms Reliability (Equivalent Forms)
- What is a Parameter in Statistics?
- Non Parametric Data and Tests (Distribution Free Tests)

Part 13: [Pa - Pr](#)

- Parametric Statistics, Tests and Data
- Pareto Distribution Definition
- Parsimonious Model: Definition, Ways to Compare Models
- Partial Correlation & Semi-Partial: Definition & Example
- Pearson Mode Skewness: Definition and Formulas
- Pearson's Coefficient of Skewness
- Percent Error & Percent Difference: Definition & Examples
- Percentiles, Percentile Rank & Percentile Range
- Z score to Percentile Calculator and Manual Methods
- Performance Bias: Definition and Examples
- Permutation, Combination and Derangement: Formula, Examples
- Permuted Block Randomization
- PERT Distribution / Beta-PERT: Definition, Examples
- Phi Coefficient (Mean Square Contingency Coefficient)
- Pie Chart: Definition, Examples, Make one in Excel/SPSS
- Pillai's Trace
- Point-Biserial Correlation & Biserial Correlation: Definition
- Point Estimate: Definition
- Poisson Distribution / Poisson Curve: Simple Definition
- Pooled Standard Deviation
- What is a Population in Statistics?
- Population Density Definition
- Population Mean Definition
- Population Proportion
- Population Variance: Definition and Examples
- Posterior Probability & the Posterior Distribution
- Post-Hoc Definition and Types of Post Hoc Tests
- Power Law and Power Law Distribution

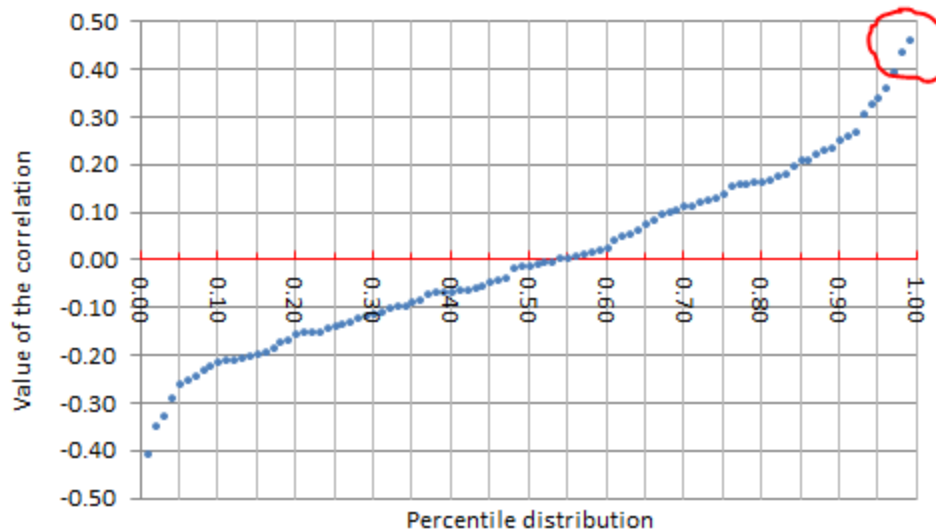


Model fitting using **percentile** tests (see [here](#))

- Practice Effect & Carry Over Effect Definition & Examples
- Prediction Interval: Simple Definition, Examples
- Predictive Validity
- Primary Data & Secondary Data: Definition & Example
- Probabilistic: Definition, Models and Theory Explained
- Probability Frequency Distribution: How to Solve Problems in Easy S...
- Probability Introduction: Articles and Videos with Solutions!
- Probability of an Event: Simple Steps in Plain English

Part 14: [Pr - Re](#)

- Probability Density Function/Probability Distribution Function
- Probability Distribution: List of Statistical Distributions
- Probability Mass Function (PMF) / Frequency Function: Definition
- Probability of A and B / A or B
- Probability of Picking From a Deck of Cards
- Probability of Selecting a Person from a Group or Committee
- Probability Problems: Solve them the easy way!
- Probability Sampling: Definition, Types, Advantages and Disadvantages
- Propensity Score Matching: Definition & Overview
- Prospective Study: Definition, Examples
- Purposive Sampling (Deliberate Sampling)



*How to lie with **p-values**? (source: [here](#))*

- P-Value in Statistical Hypothesis Tests: What is it?
- Pygmalion Effect / Rosenthal Effect: Definition, Examples
- Q Q Plots: Simple Definition & Example
- Quadratic Mean / Root Mean Square
- Quadratic Regression: Simple Definition, TI-Calculator Instructions
- Qualitative Variable: Definition and Examples
- Quantile: Definition and How to Find Them in Easy Steps
- Quantitative Variables (Numeric Variables) in Statistics
- What are Quartiles?
- Quota Sampling: Definition and Examples
- Q-Value: Definition and Examples
- Random Event: Definition, How to Find Probability
- Randomization in Statistics and Experimental Design
- Random Variable: What is it in Statistics?
- Range of a Set of Data in Math and Statistics
- Range Rule of Thumb
- Ratio Scale: Definitions, Examples of Ratio Variables
- Rayleigh Distribution: Definition, Uses, Mean, Variance
- Recall Bias: Definition, Examples, Strategies to Avoid it
- Receiver Operating Characteristic (ROC) Curve: Definition, Example
- Regression Analysis: Step by Step Articles, Videos, Simple Definitions
- Regression Equation: What it is and How to use it
- Row Echelon Form & Reduced Row Echelon Form
- Regression to the Mean: Definition, Examples
- Regularization: Simple Definition, L1 & L2 Penalties

- Rejection Region (Critical Region) for Statistical Tests

	Type of Distribution	Expectation	Standard deviation
Uniform	Short tail	1	$\frac{1}{n}$
Gaussian	Medium tail	$\sqrt{\log n}$	$\frac{1}{\sqrt{n}}$
Exponential	Fat tail	$\log n$	1

*Distribution of the **range** (source: [here](#))*

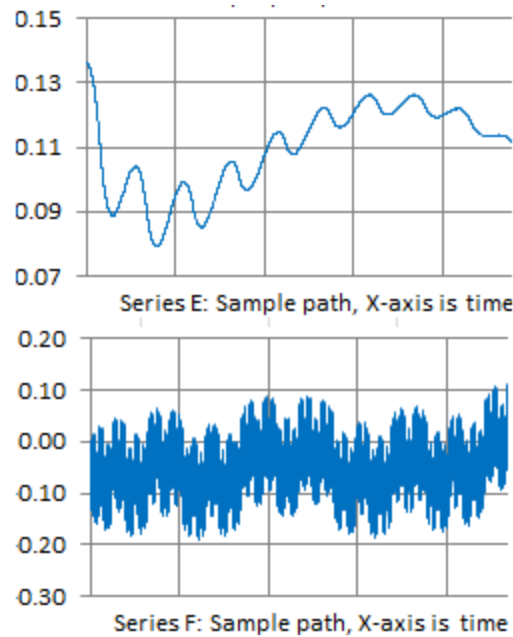
Part 15: [Re - Sc](#)

- Relative Error: Definition, Formula, Examples
- Relative Frequency Distribution: Definition and Examples
- Relative Frequency Histogram: Definition and How to Make One
- Relative Standard Deviation: Definition & Formula
- Reliability and Validity in Research: Definitions, Examples
- Reporting Statistics APA Style
- What is Rescaling Data ? Statistics Definitions and Examples
- Research Methods: Qualitative Research and Quantitative Research
- Residual Values (Residuals) in Regression Analysis
- Residual Plot: Definition and Examples
- Sum of Squares: Residual Sum, Total Sum, Explained Sum, Within
- Resistance & Resistant Measures in Statistics
- Responding Variable
- Response Bias: Definition and Examples
- Reverse Causality: Definition, Examples
- Ridge Regression: Simple Definition
- RMSE: Root Mean Square Error
- Same Birthday Odds: Higher Than You Think!
- Sample in Statistics: What it is, How to find it
- Sample Mean: Symbol (X Bar), Definition, and Standard Error
- Sample Space Examples and The Counting Principle
- Sample Size in Statistics (How to Find it): Excel, Cochran's Formula
- Sample Variance: Simple Definition, How to Find it in Easy Steps
- Sampling in Statistics: Different Sampling Methods, Types & Error
- Sampling Distribution: Definition, Types, Examples
- Sampling Distribution of the Sample Proportion
- Sampling Frame / Sample Frame Definition

- Sampling Variability: Definition
- Sampling With Replacement / Sampling Without Replacement
- Scales of Measurement / Level of Measurement
- Scale Variable: Definition
- Scatter Plot / Scatter Chart: Definition, Examples, Excel
- Scheffe Test: Definition, Examples, Calculating (Step by Step)

Part 16: [Se - St](#)

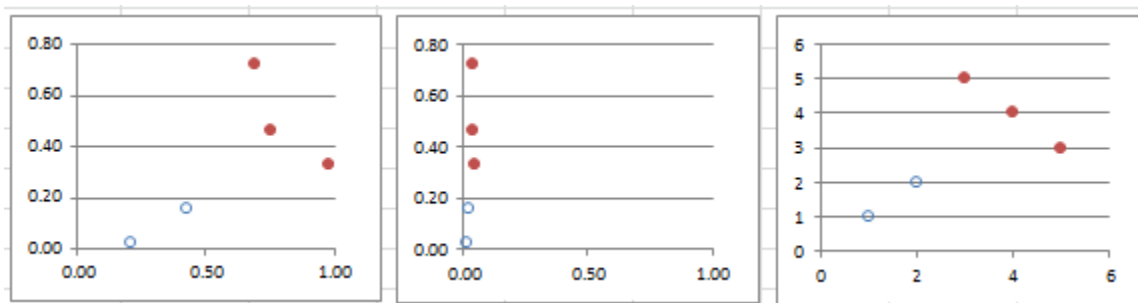
- Self-Selection Bias
- Semantic Differential Scale: Definition, Examples
- Semi Interquartile Range / Quartile Deviation
- Sensitivity vs Specificity and Predictive Value
- Sequential Sampling: Definition, Advantages/Disadvantages
- Serial Correlation / Autocorrelation: Definition, Tests
- Shapes of Distributions: Definitions, Examples
- Shapiro-Wilk Test: What it is and How to Run it
- Sig(2-Tailed): Interpreting Results
- Sigma / sqrt (n) -- why is it used?
- Significant Digits / Figures and Rounding in Statistics
- Sign Test: Step by Step Calculation
- Simple Random Sample: Definition and Examples
- What is Simpson's Paradox?
- Simpson's Diversity Index: Definition, Formula, Calculation
- Simultaneity Bias: Simple Definition
- Skewed Distribution: Definition, Examples
- Skewness: Equations for common graphs and distributions
- Snowball Sampling: Definition, Advantages and Disadvantages
- Somers' D: Simple Definition
- Spearman-Brown Formula
- Spearman Rank Correlation (Spearman's Rho)
- Split-Half Reliability: Definition, Steps
- SPSS Tutorial (for Beginners): Learn Online in Simple Steps
- Spurious Correlation: Examples from Real Life and the News
- Standard Deviation: Simple Definition, Step by Step Video
- What is the Standard Error of a Sample ?
- Standardized Residuals in Statistics: What are They?
- Standard Error of Measurement (SEm): Definition, Meaning
- Standardized Beta Coefficient: Definition & Example



- Standardized Test Statistic: What is it?
- Standardized Values: Example
- Standardized Variables: Definition, Examples
- Stanine Score: Definition, Examples, How to Convert
- Stationarity: Definition, Examples, Types

- Statistical Analysis: Definition, Examples
- Statistical Conclusion Validity
- Statistical Power: What it is, How to Calculate it
- Statistical Relationship: Definition, Examples
- Statistical Treatment
- Statistics Basics
- Statistics Symbols in Alphabetical Order
- What is Statistical Significance?
- Statistics Definitions in Plain English with Examples
- Stemplot in Statistics: What is it? How to Make One
- STEN Score
- Stepwise Regression
- Stochastic Model / Process: Definition and Examples
- Stratified Random Sample: Definition, Examples
- Studentized Range Distribution
- Subjective Probability: Definition & Examples

- Success/Failure Condition: Definition, Examples
- Summary Statistics: Definition and Examples
- Summation Notation (Sigma Notation): Definition and Use
- Symmetric Distribution in Statistics
- Systematic Error / Random Error: Definition and Examples
- Systematic Sampling: Definition & Examples + Repeated Samples
- T Statistic: Definition, Types and Comparison to Z Score
- T Critical Value: Easy Definition, Calculating
- T-Score vs. Z-Score: What's the Difference?
- T Score Formula: Calculate in Easy Steps
- T Test (Student's T-Test): Definition and Examples
- T-Distribution / Student's T: Definition, Step by Step Articles, Video
- T-Distribution Table (One Tail and Two-Tails)



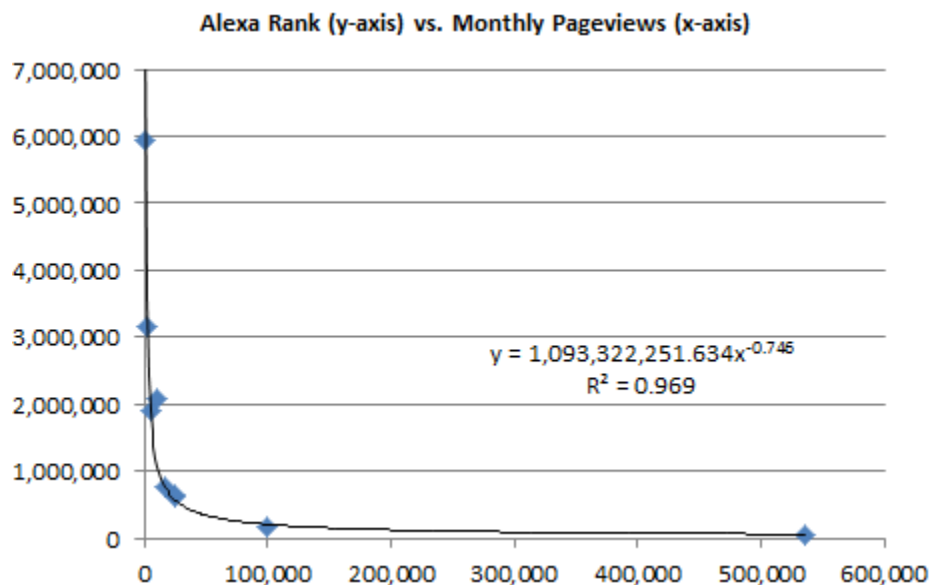
Scale-invariant clustering (source: [here](#))

- One Tailed Test or Two in Hypothesis Testing: How to Decide
- Test-Retest Reliability / Repeatability
- Test Statistic: What is it? Types of Test Statistic
- Tetrachoric Correlation: Definition, Examples, Formula
- Theoretical Probability Definition and Examples
- Three-Way ANOVA: Definition
- Thurstone Scale: Definition, Examples
- Timeplot / Time Series: Definition, Examples & Analysis
- Total Probability Rule / Law of Total Probability Theorem
- Trend Analysis: Simple Definition, Examples
- Trimmed Mean / Truncated Mean: Definition, Examples
- Tukey Test / Tukey Procedure / Honest Significant Difference
- Two-Sample T-Test: When to Use it
- Two Sample z Test Excel 2013: Easy Steps
- Type I & Type II Errors (Decision Errors): Easy Definition, Exa...
- Types of graphs used in Math and Statistics

- Types of Variables in Statistics and Research
- Unbiased in Statistics: Definition and Examples
- Unequal Sample Sizes

Part 18: Un - Z

- Unidimensionality: Definition, Examples
- Uniform Distribution / Rectangular Distribution: What is it?
- Unimodal Distribution in Statistics



Zipf law (source: [here](#))

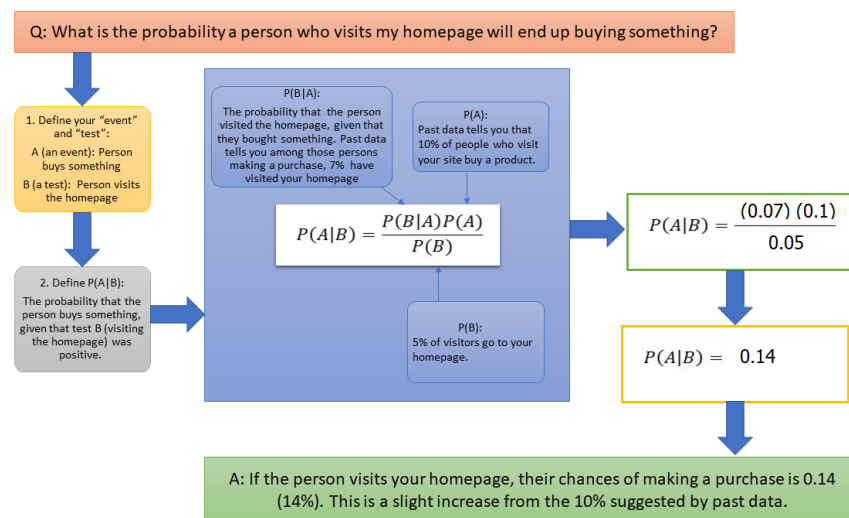
- Unit Root: Simple Definition, Unit Root Tests
- Univariate Analysis: Definition, Examples
- Upper and Lower Fences
- Upper Hinge and Lower Hinge
- Validity Coefficient: Definition and How to Find it
- Variability in Statistics: Definition, Examples
- Variance: Simple Definition, Step by Step Examples
- Variance Inflation Factor
- Voluntary Response Sample in Statistics: Definition
- Wald Test: Definition, Examples, Running the Test
- Weibull Distribution and Weibull Analysis
- Weighted Least Squares: Simple Definition, Pluses and Minuses
- Weighted Mean: Formula: How to Find Weighted Mean
- Weighting Factor, Statistical Weight and Weight Functions
- Welch's ANOVA: Definition, Assumptions

- Welch's Test for Unequal Variances
- Yates Correction: What is it used for in Statistics?
- White Test: Definition, Examples
- Wilcoxon Signed Rank Test: Definition, How to Run
- Wilks' Lambda: Simple Definition
- Winsorize: Definition, Examples in Easy Steps
- Within-Group Variation: Definition and Examples
- Y Hat: Definition
- Z Alpha/2 ($z_{\alpha/2}$): What is it, How to Find it
- Zero-Order Correlation: Definition, Examples
- Z Test: Definition & Two Proportion Z-Test
- Z-Score: Definition, Formula and Calculation
- Z-table (Right of Curve or Left)

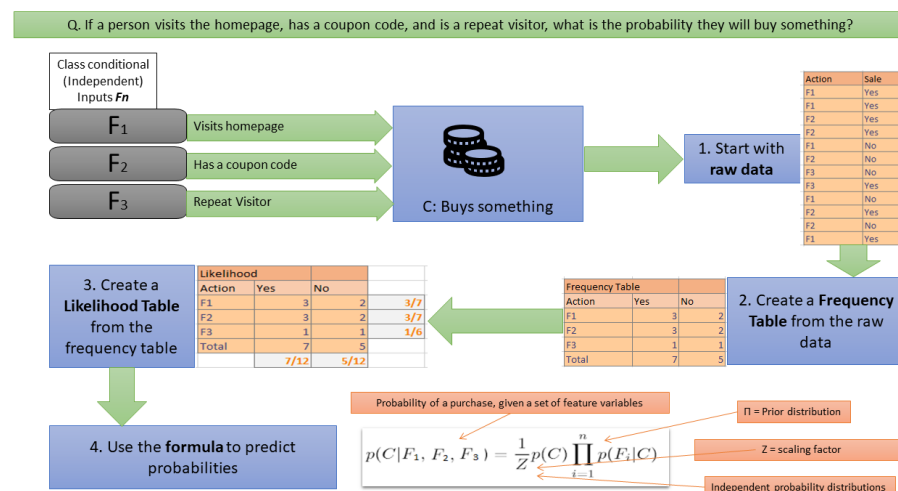
Statistical Topics Explained in One Picture

Below is a selection of topics covering various aspects of statistical science. You can **zoom in (using the browser or Acrobat zoom button) to enlarge the pictures**: they are in high resolution. A link to where the picture was first posted is provided in each case, allowing you to find more information about the topic in question. More “one picture” summaries can be found [here](#).

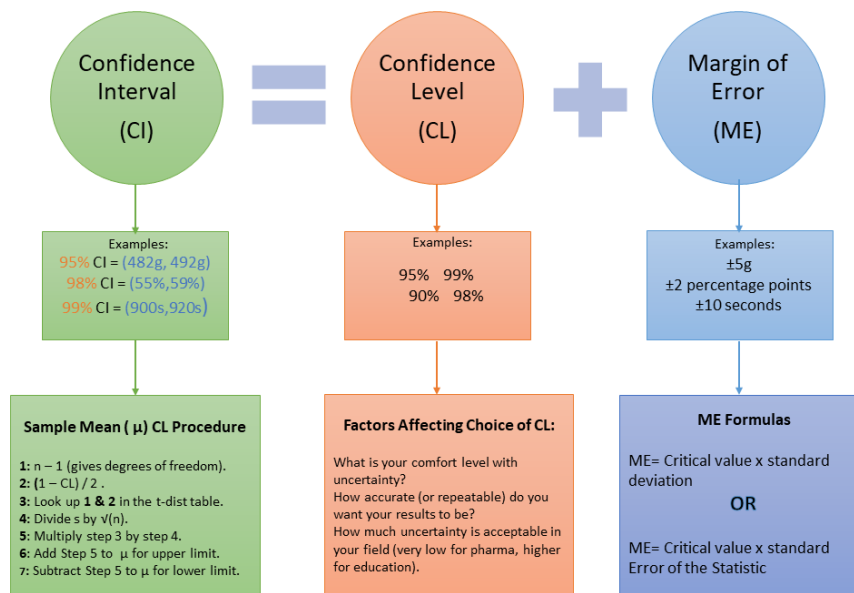
1. Bayes Theorem ([source](#))



2. Naïve Bayes ([source](#))



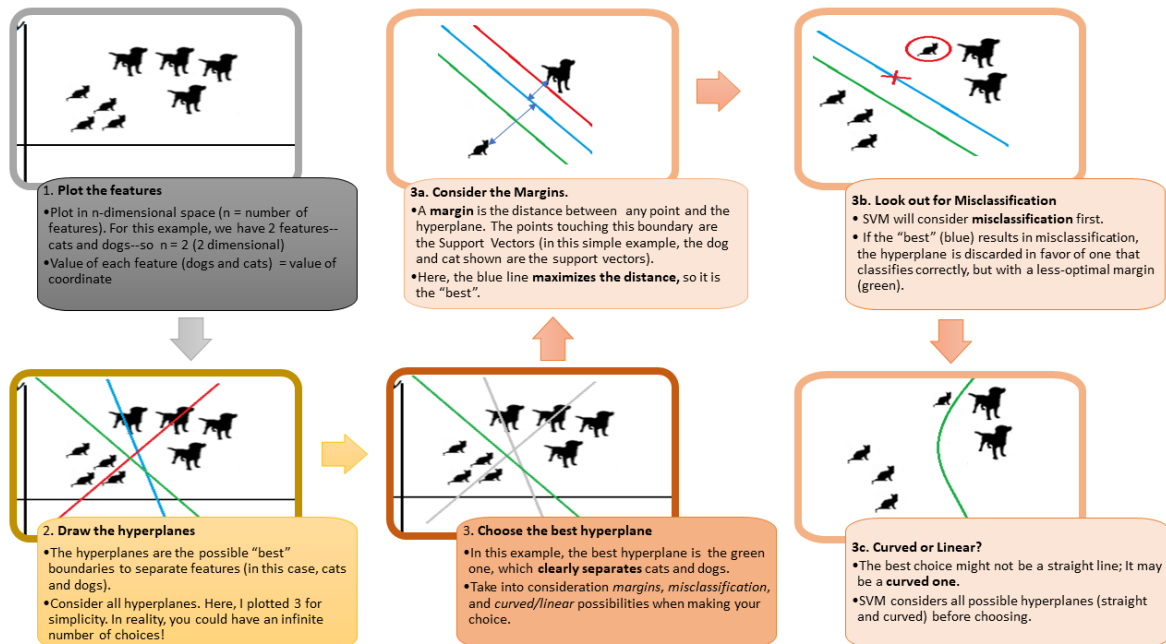
3. Confidence Intervals (source)



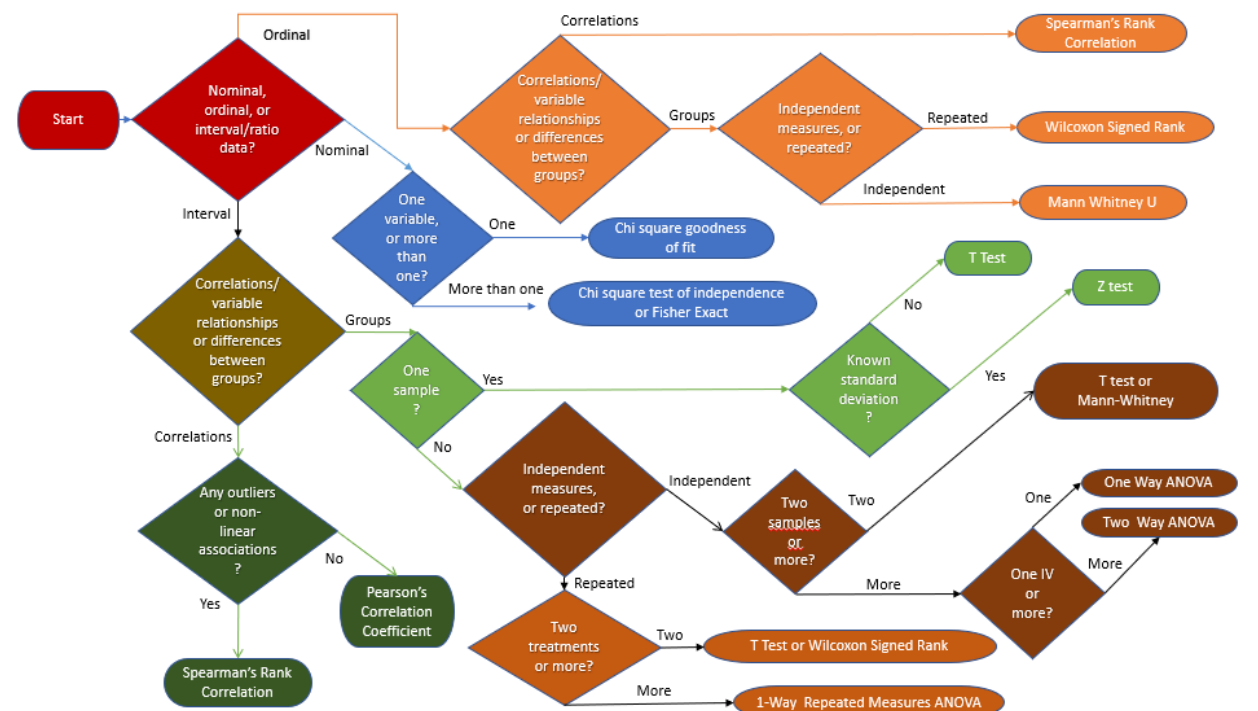
4. Three Types of Regression (source)

LINEAR REGRESSION	LOGISTIC REGRESSION	POISSON REGRESSION
<ol style="list-style-type: none"> 1 Econometric modelling 2 Marketing Mix Model 3 Customer Lifetime Value 	<ol style="list-style-type: none"> 1 Customer Choice Model 2 Click-through Rate 3 Conversion Rate 4 Credit Scoring 	<ol style="list-style-type: none"> 1 Number of orders in lifetime 2 Number of visits per user
Continuous \Rightarrow Continuous	Continuous \Rightarrow True/False	Continuous \Rightarrow 0,1,2,...
$y = \alpha_0 + \sum_{i=1}^N \alpha_i x_i$	$y = \frac{1}{1 + e^{-z}}$	$y \sim \text{Poisson}(\lambda)$
$\ln \lambda = \alpha_0 + \sum_{i=1}^N \alpha_i x_i$	$z = \alpha_0 + \sum_{i=1}^N \alpha_i x_i$	$\ln \lambda = \alpha_0 + \sum_{i=1}^N \alpha_i x_i$
<code>lm(y ~ x1 + x2, data)</code>	<code>glm(y ~ x1 + x2, data, family=binomial())</code>	<code>glm(y ~ x1 + x2, data, family=poisson())</code>
1 unit increase in x increases y by α	1 unit increase in x increases log odds by α	1 unit increase in x multiplies y by e^α

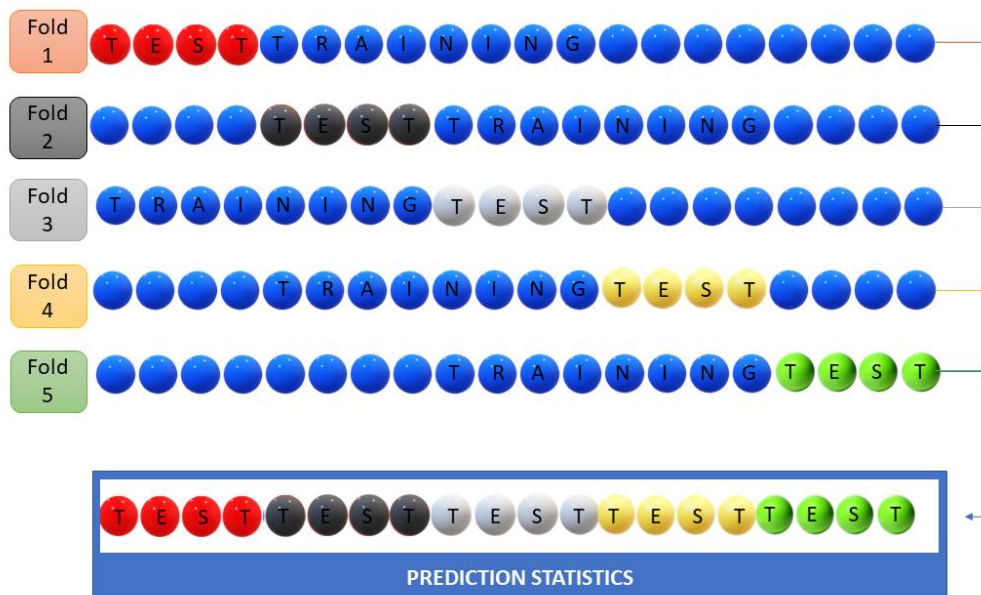
4. Support Vector Machines (source)



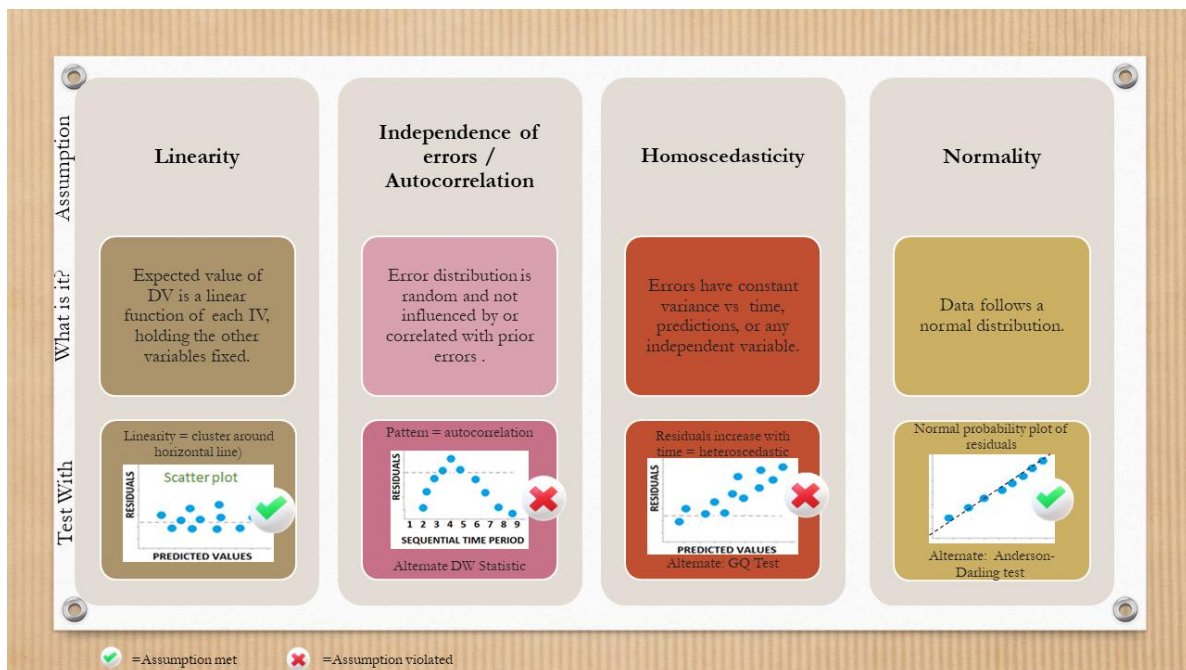
6. Hypothesis Tests (source)



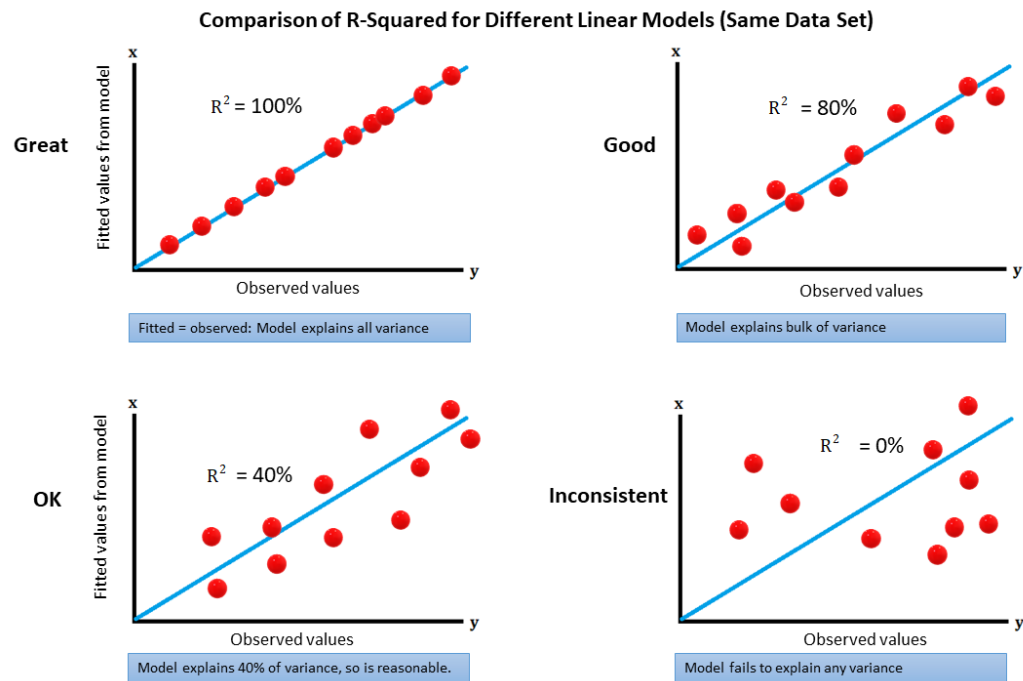
7. Cross-validation (source)



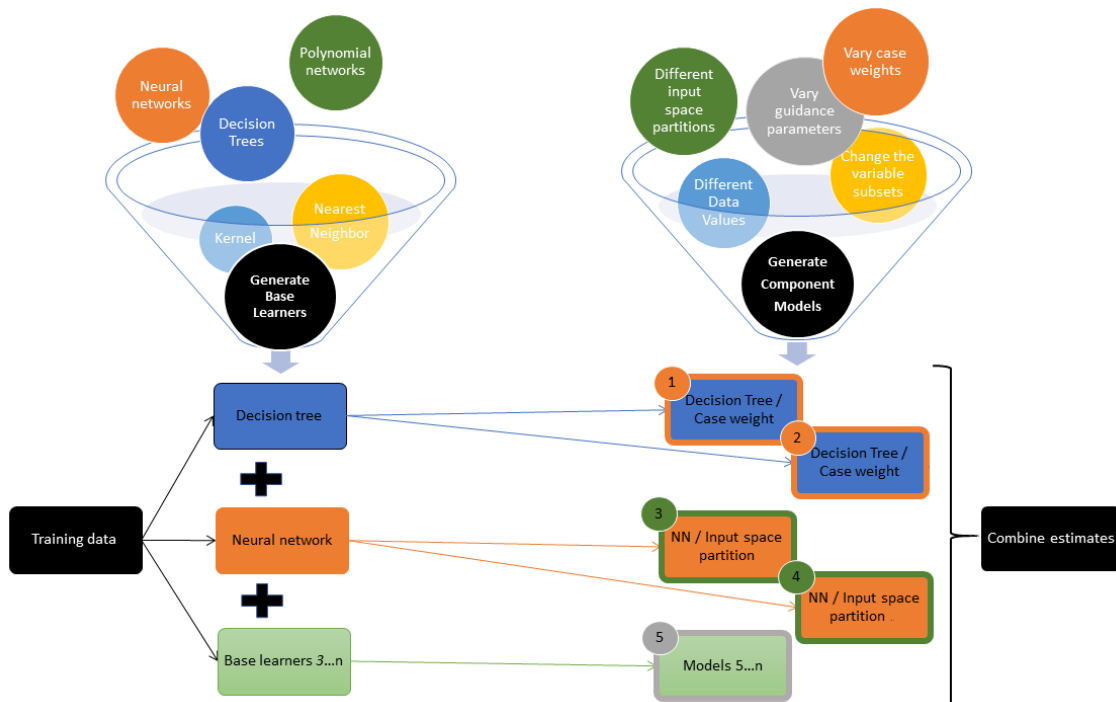
8. Assumptions of Linear Regression (source)



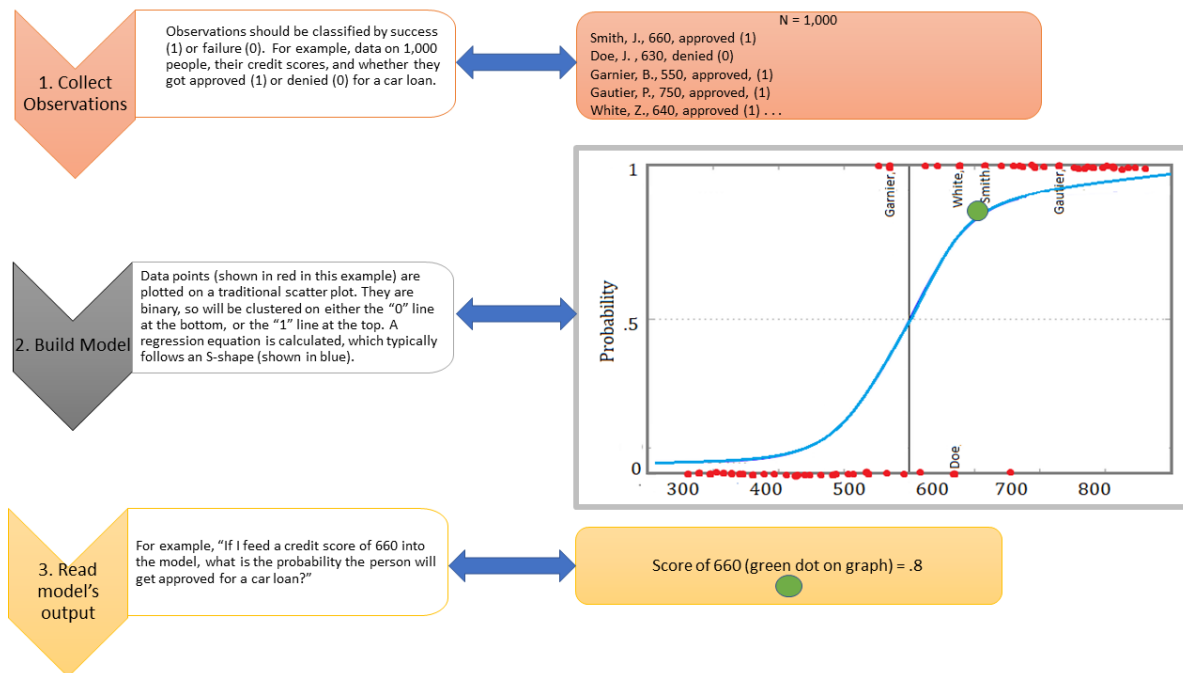
9. R-squared ([source](#))



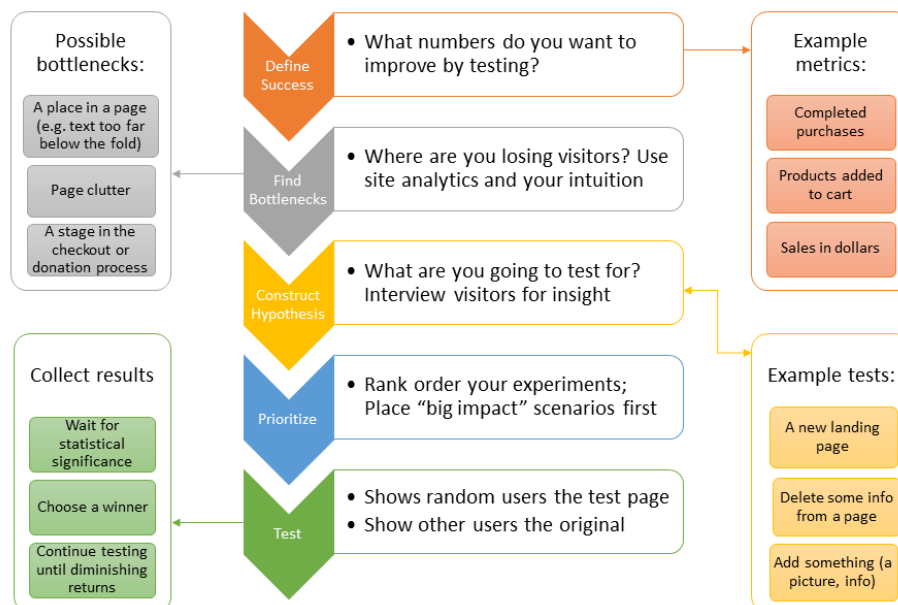
10. Ensemble Methods ([source](#))



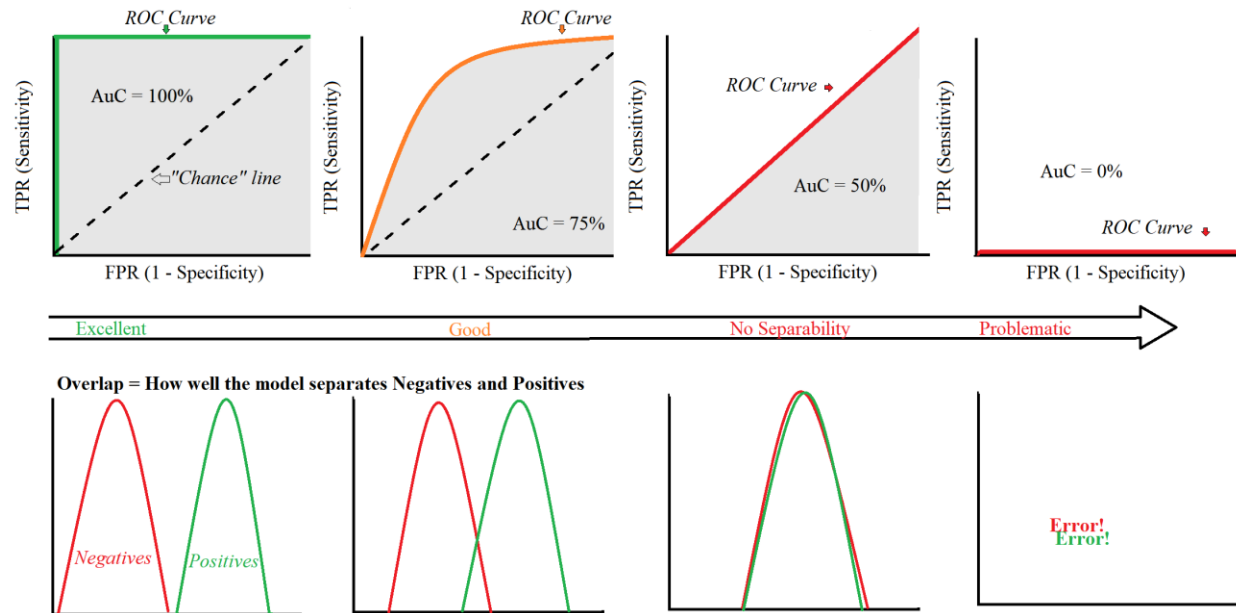
11. Logistic Regression (source)



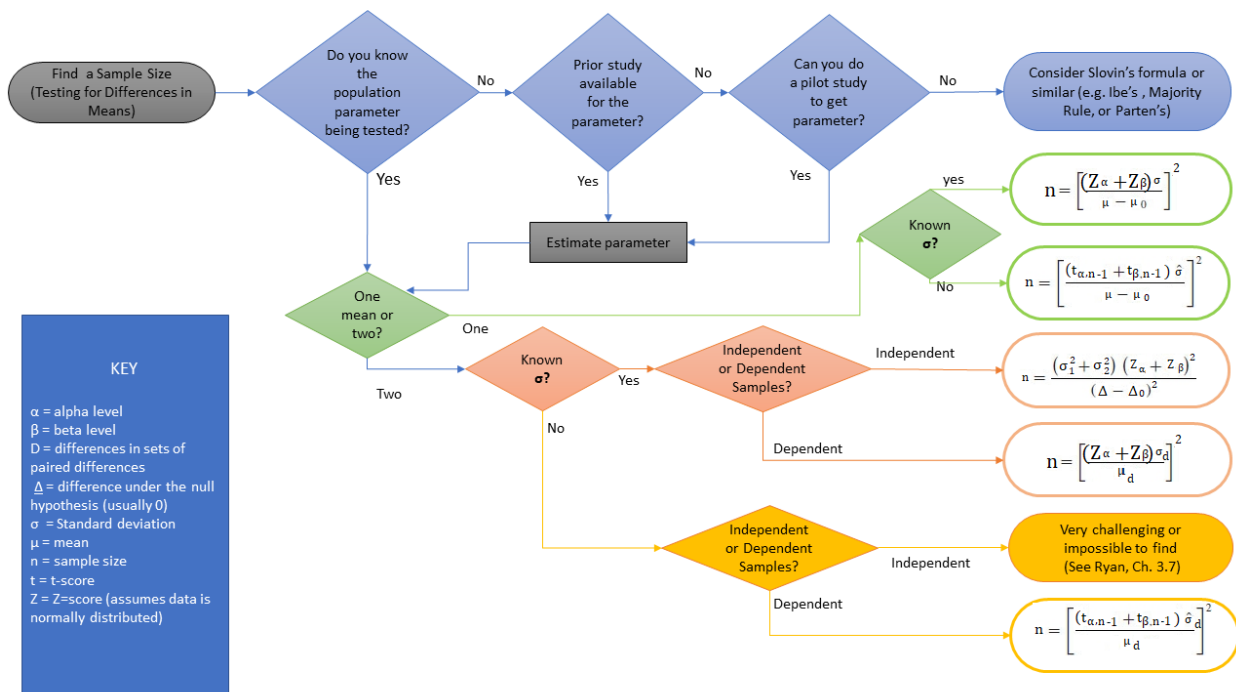
12. A/B Testing (source)



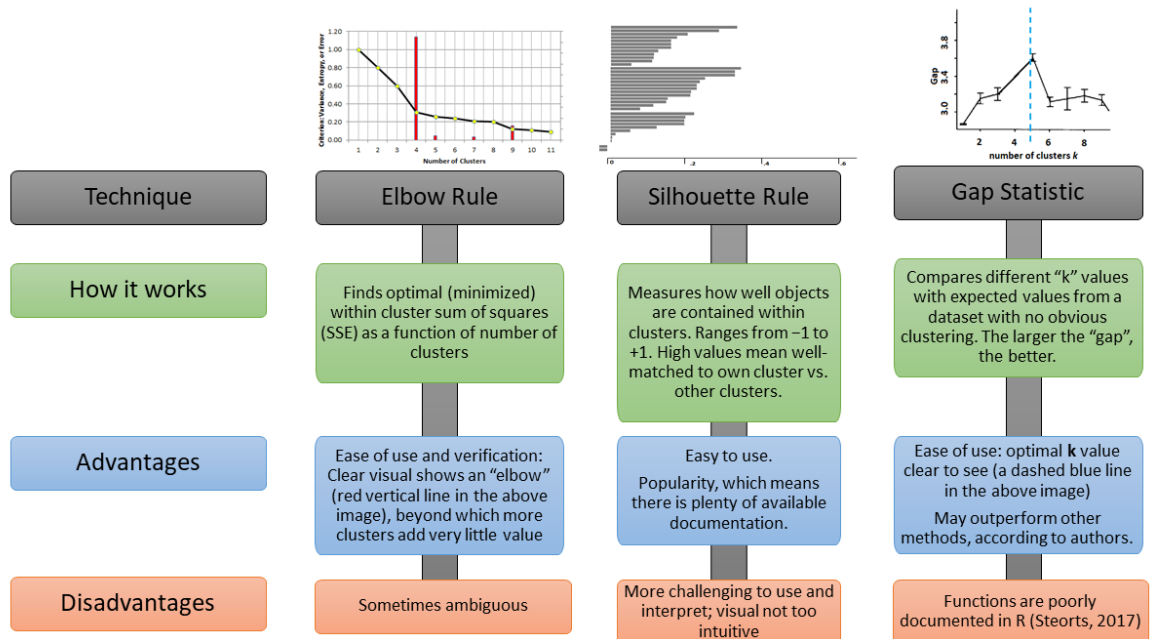
13. ROC Curve (source)



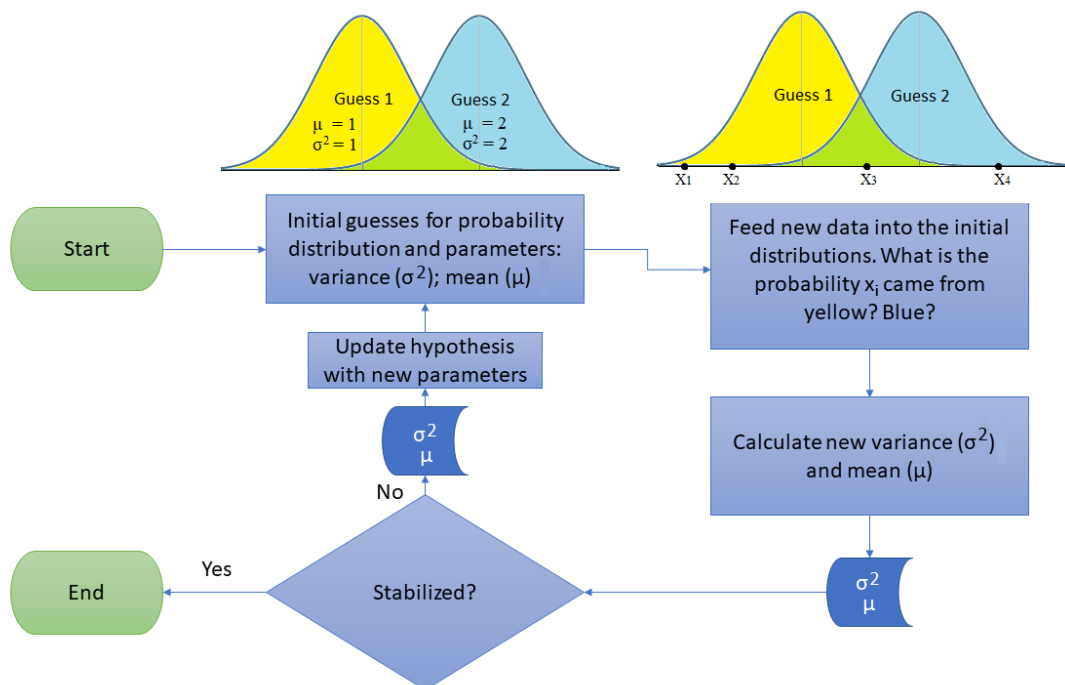
14. Determining Sample Size (source)



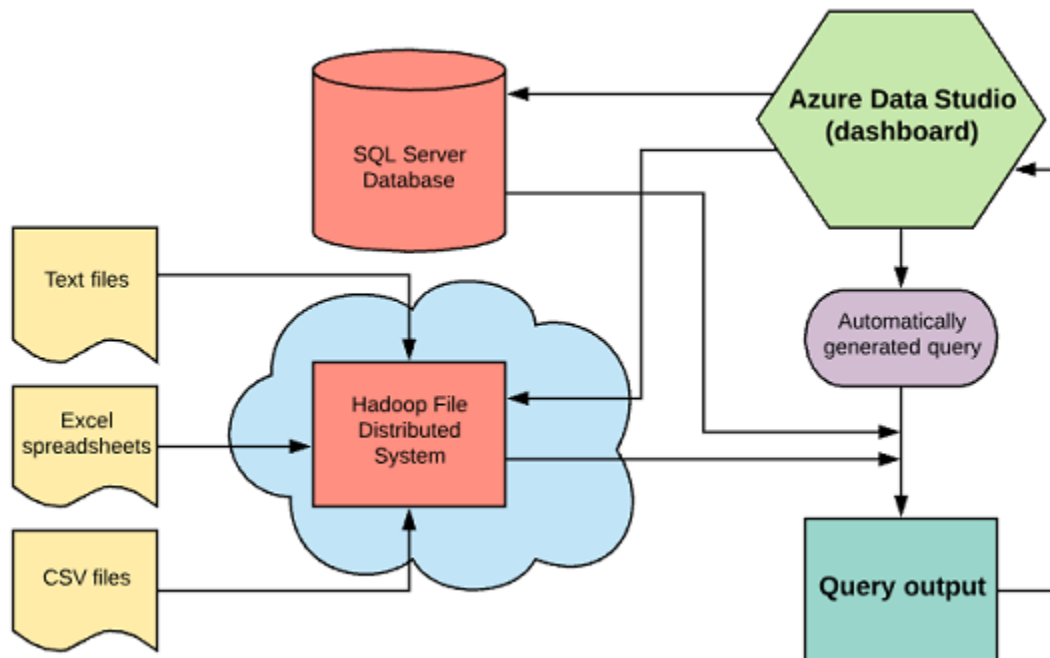
15. Determining the Number of Clusters (source)



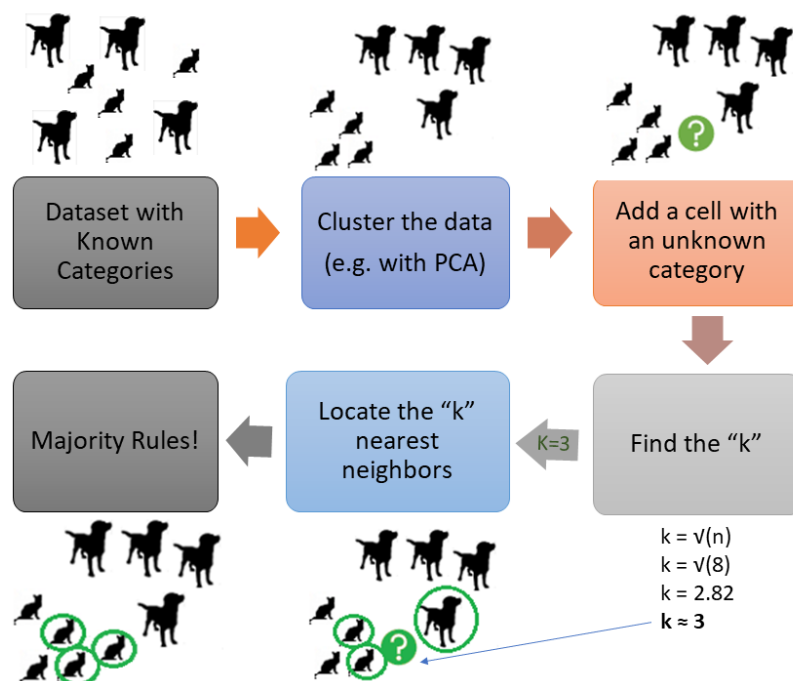
16. EM Algorithm (source)










17. Azure Data Studio (source)



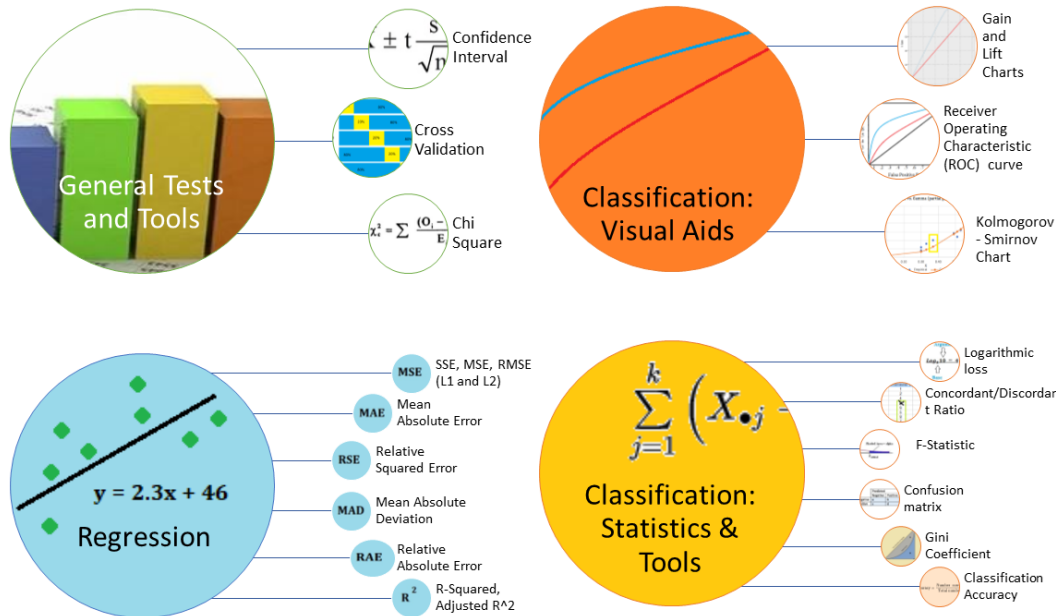
18. K Nearest Neighbors Clustering (source)



19. Prediction Algorithms ([source](#))

<div>  dataiku </div> TOP PREDICTION ALGORITHMS				
TYPE	NAME	DESCRIPTION	ADVANTAGES	DISADVANTAGES
Linear	 Linear regression	The “best fit” line through all data points. Predictions are numerical.	Easy to understand – you clearly see what the biggest drivers of the model are.	X Sometimes too simple to capture complex relationships between variables. X Tendency for the model to “overfit”.
	 Logistic regression	The adaptation of linear regression to problems of classification (e.g., yes/no questions, groups, etc.)	Also easy to understand.	X Sometimes too simple to capture complex relationships between variables. X Tendency for the model to “overfit”.
Tree-based	 Decision tree	A graph that uses a branching method to match all possible outcomes of a decision.	Easy to understand and implement.	X Not often used on its own for prediction because it's also often too simple and not powerful enough for complex data.
	 Random Forest	Takes the average of many decision trees, each of which is made with a sample of the data. Each tree is weaker than a full decision tree, but by combining them we get better overall performance .	A sort of “wisdom of the crowd”. Tends to result in very high quality models. Fast to train.	X Can be slow to output predictions relative to other algorithms. X Not easy to understand predictions.
	 Gradient Boosting	Uses even weaker decision trees, that are increasingly focused on “hard” examples .	High-performing.	X A small change in the feature set or training set can create radical changes in the model. X Not easy to understand predictions.
Neural networks	 Neural networks	Mimics the behavior of the brain. Neural networks are interconnected neurons that pass messages to each other. Deep learning uses several layers of neural networks put one after the other.	Can handle extremely complex tasks - no other algorithm comes close in image recognition.	X Very, very slow to train, because they have so many layers. Require a lot of power. X Almost impossible to understand predictions.

20. Model Evaluation Techniques (source)



21. Regression Analysis (source)

