

Statistical Advisor

Bernoulli – describes all situations where a "trial" is made resulting in either "success" or "failure", such as when tossing a coin

Beta – arises from a transformation of the F distribution and is typically used to model the distribution of order statistics

Binomial – is useful for describing distributions of binomial events, such as the number of defective components in samples of 20 units taken from production process

Cauchy – is often used in statistics as canonical example of a pathological distribution since both its mean and variances are undefined

Chi-square – the sum of n independent squared random variables, each distributed following the standard normal distribution, is distributed as Chi-square with n degrees of freedom

Exponential – is frequently used to model the time interval between successive random events, example would be the gap length between cars crossing an intersection

Extreme Value – to model extreme events such as size of floods, gust velocities encountered by airplanes, maxima of stock indices over a given year, etc

F – is mostly used in tests of variance (e.g., ANOVA)

Gamma – when modeling the distribution of the life-times of a product such as an electric light bulb, or the serving time taken at a ticket booth at a baseball game

Geometric – if the independent Bernoulli trials are made until a "success" occurs, then the total number of trials required is a geometric random variable

Gompertz – is theoretical distribution of survival times

Logistic – is used to model binary responses (e.g., gender) and is commonly used in logistic regression

Log normal – is often used in simulations of variables such as personal incomes, age at first marriage, or tolerance to poison in animals, etc

Normal – is a bell shaped curve which is symmetrical about mean, is a theoretical function commonly used in inferential statistics as an approximation to sampling distributions, it's a good model for random variable

Pareto – is commonly used in monitoring production processes, example- a machine which produces copper wire will occasionally generate a flaw at some point along the wire

Poisson – distribution of rare events, example - number of accidents per person, number of sweepstakes won per person, etc

Rayleigh – distance of darts from the target in a dart-throwing game

Rectangular – useful for describing random variables with a constant probability density over the defined range

Student's t – is symmetric about 0, its shape is similar to that of standard normal distribution. It is most commonly used in testing hypothesis about the mean of a particular population

Weibull – used when the failure probability varies over time, often used in reliability testing (e.g., ball bearings, etc)

Statistical significance test – Compare the observed distribution of variables against several theoretical distributions and test the discrepancy of the observed data from the respective theoretical distributions

p-value – statistical significance of a result is an estimated measure of the degree to which it is "true"

Z-value – standardized value - value is expressed in terms of its difference from the mean, divided by sd

Confidence intervals – give us a range of values around the mean where we expect the "true" (population) mean is located

t-Test – to evaluate the differences in means between groups. If the resultant t-value is statistically significant then one can conclude that the means in the two variables are different

F-test – for comparison of the variances in two groups, if statistically significant, one can conclude that the variances (variability) in the two groups are different

General Linear Models (GLM) – assumes that the variables in the comparison are normally distributed within the groups

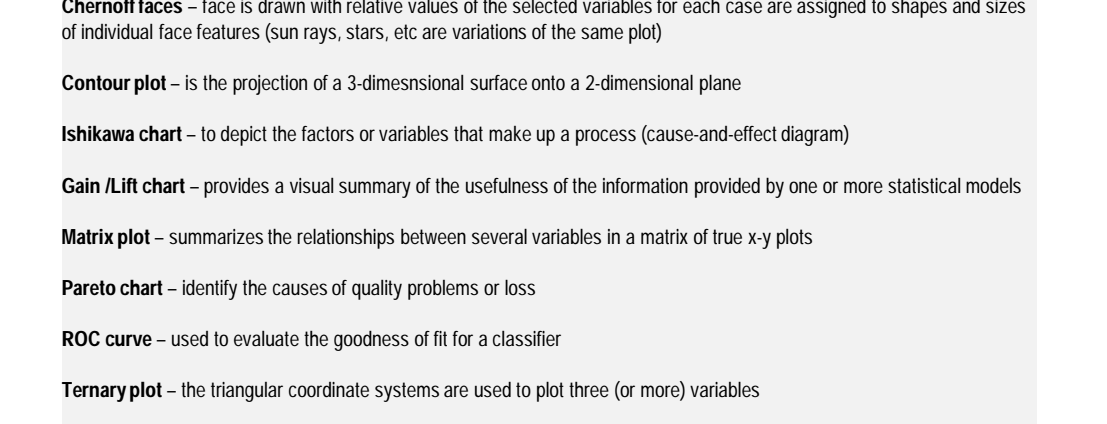
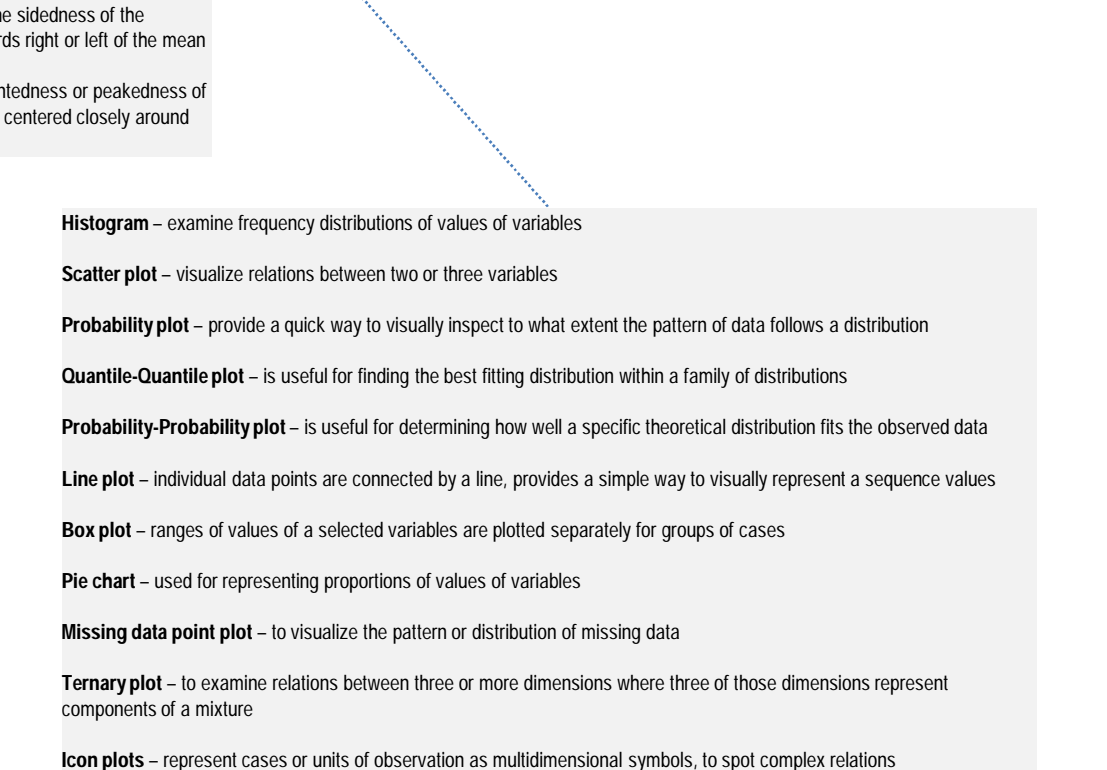
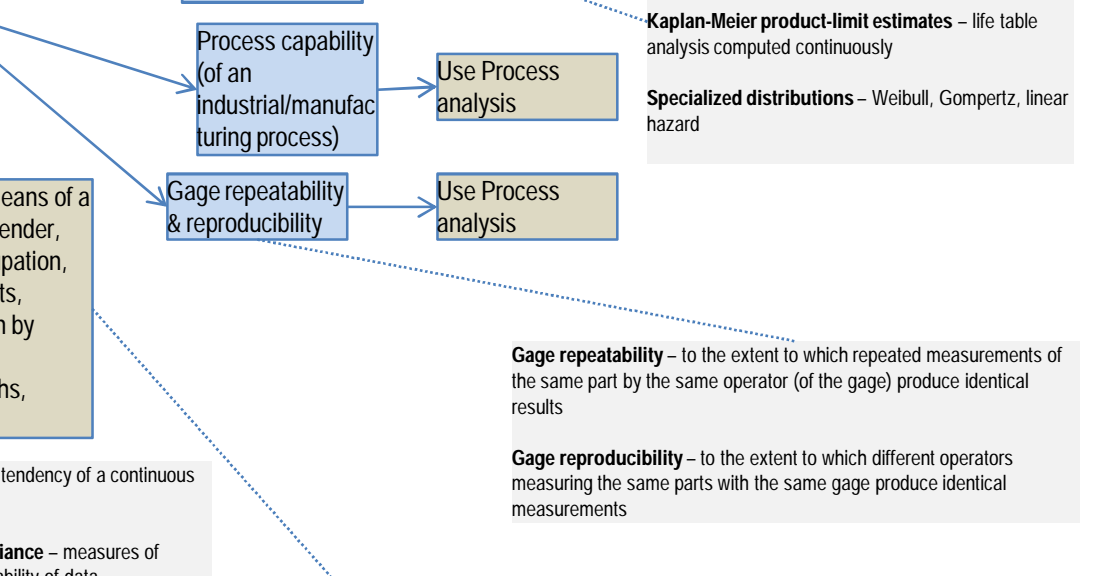
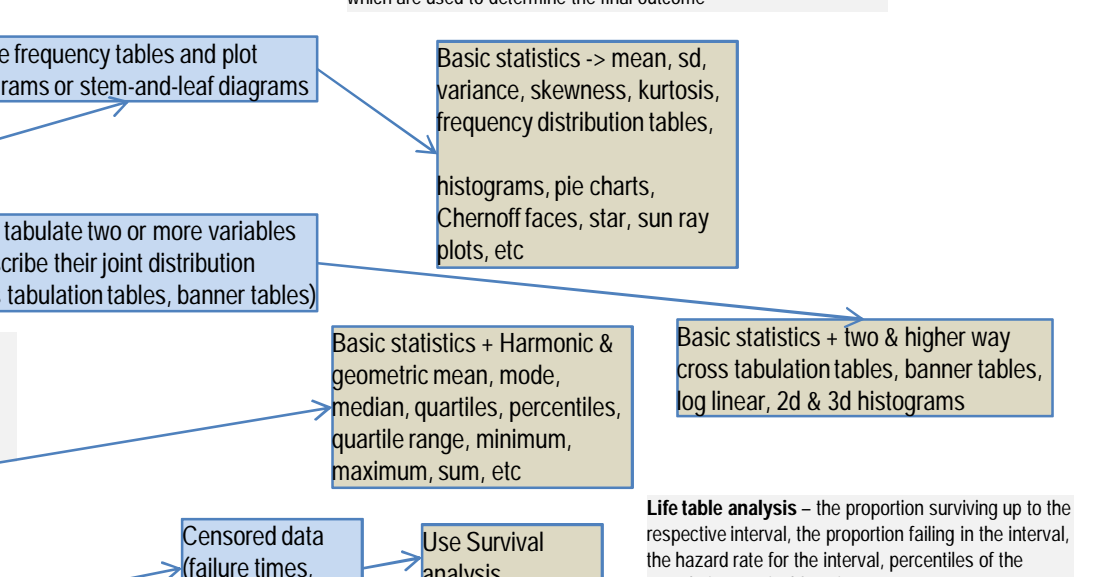
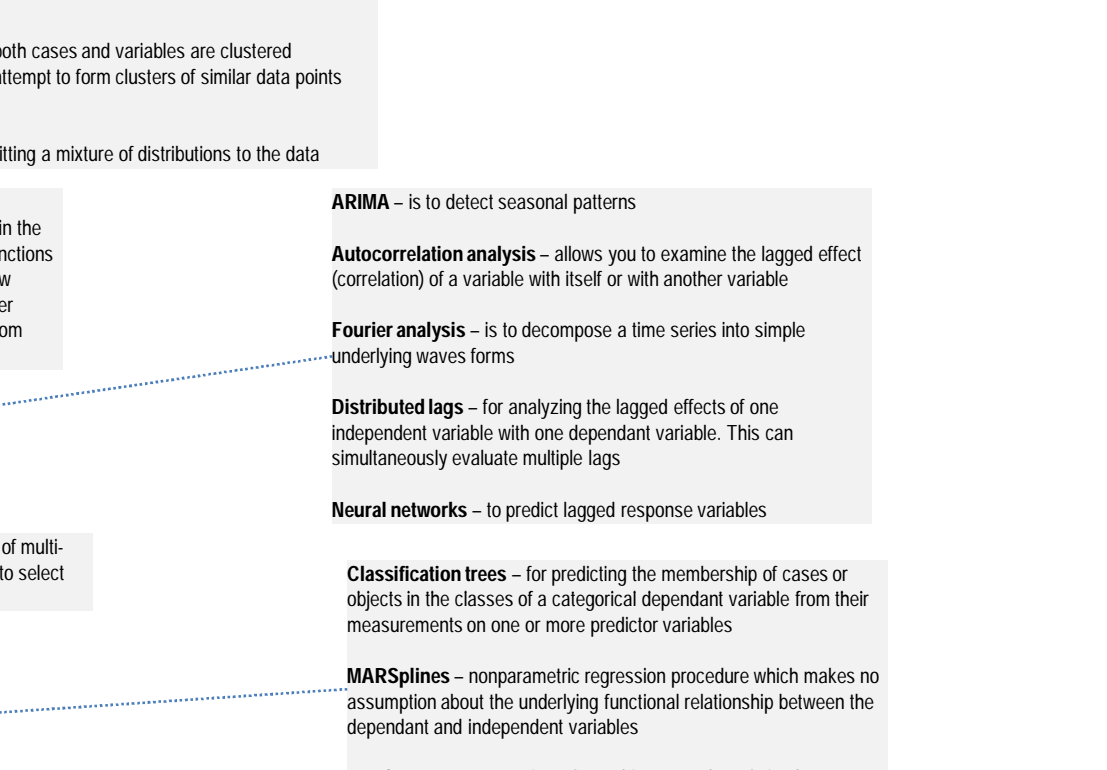
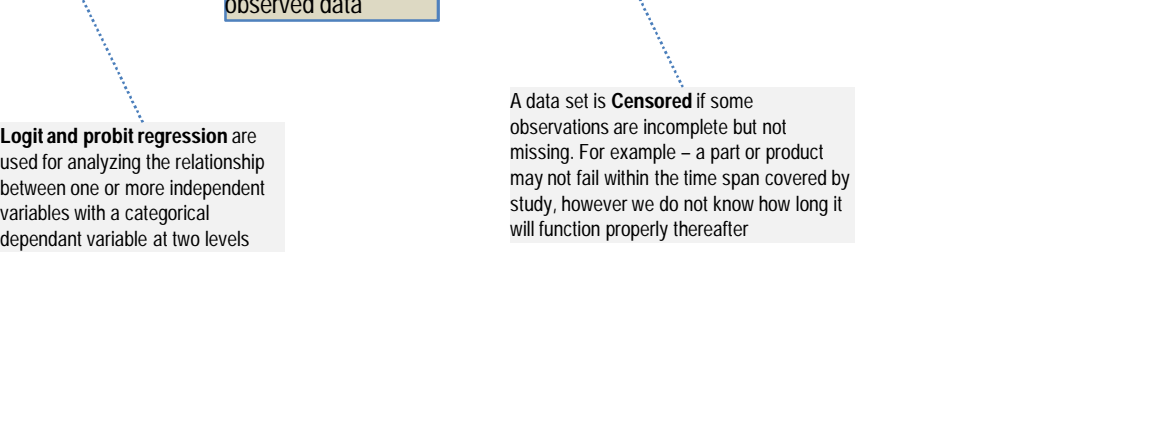
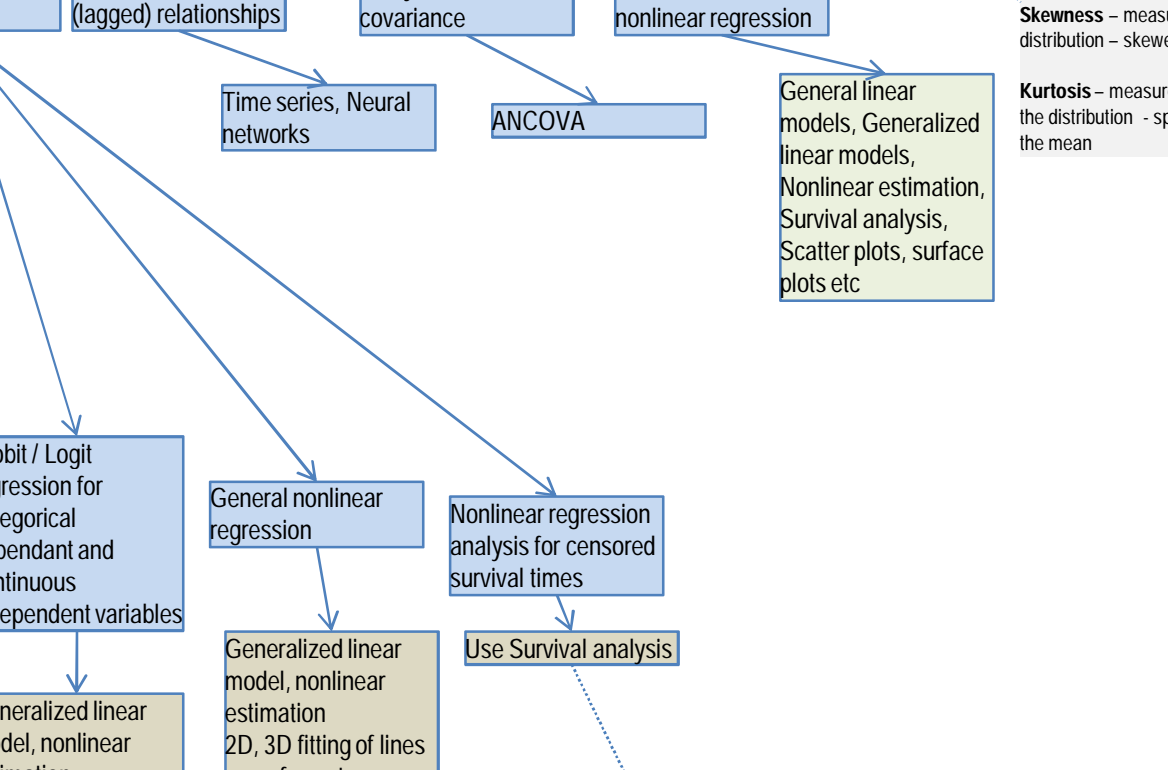
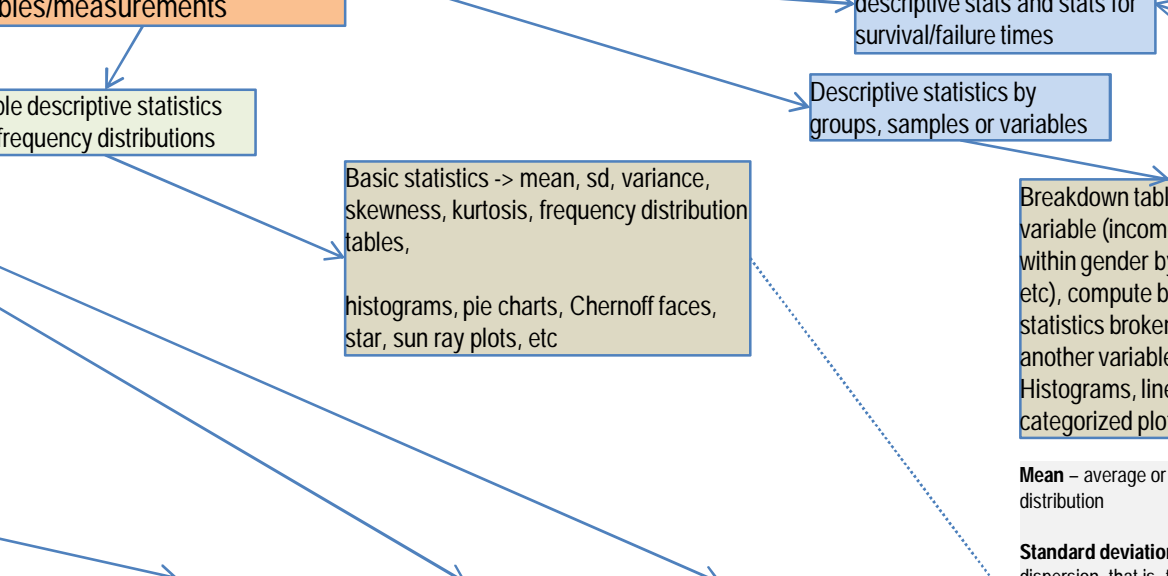
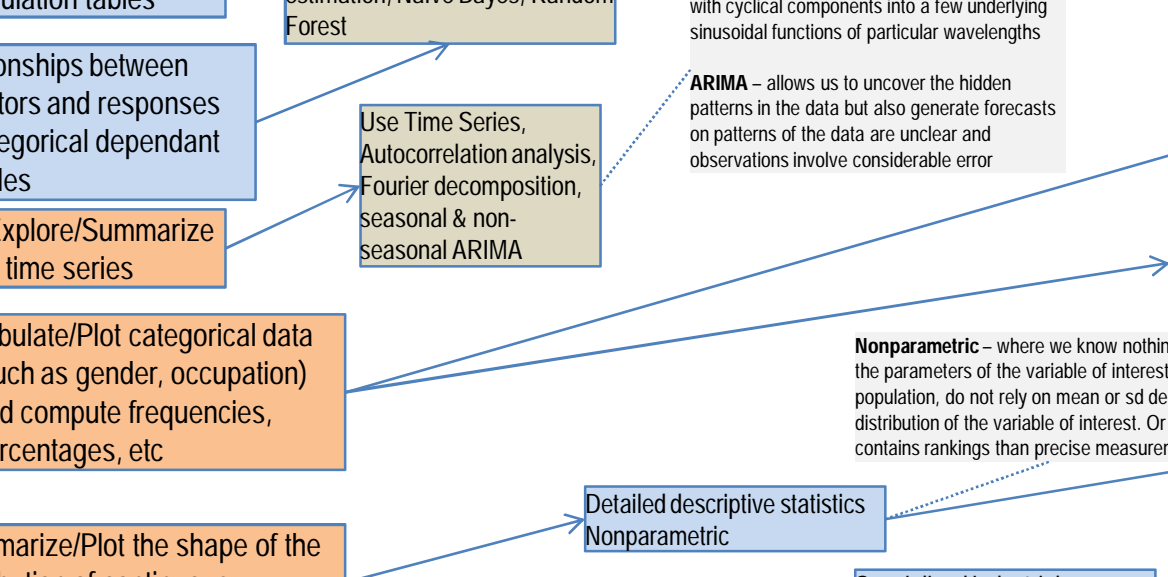
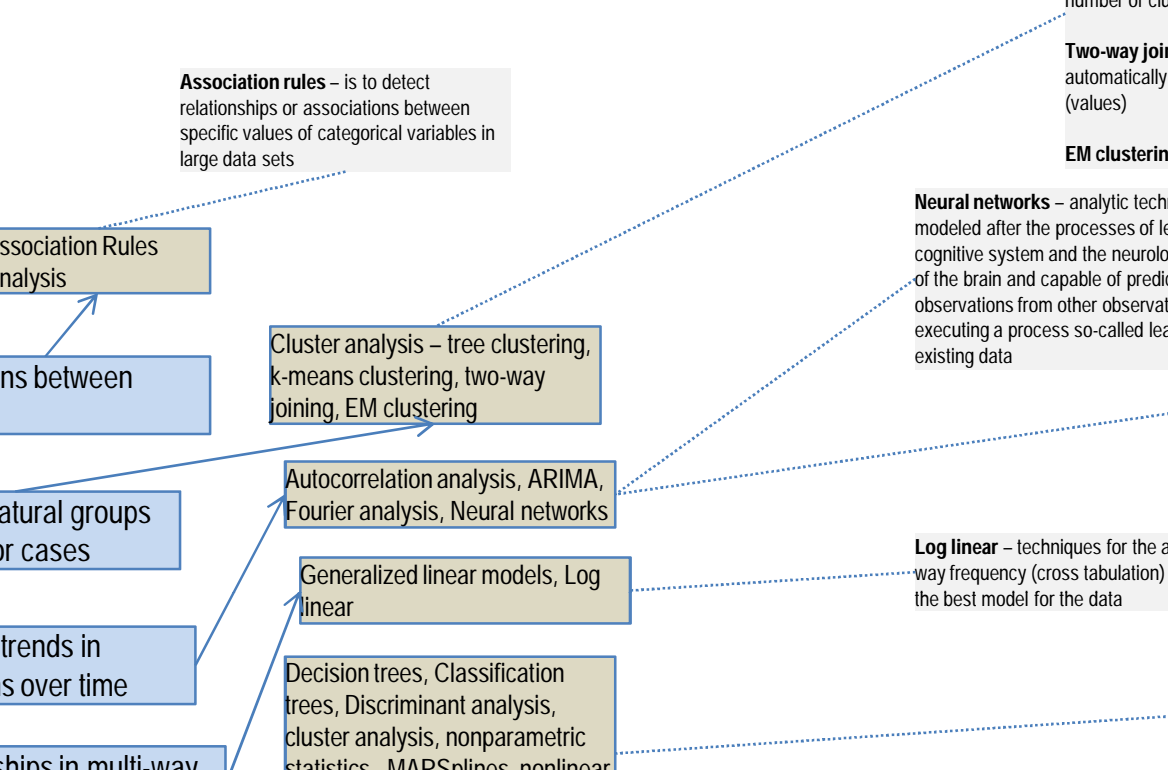
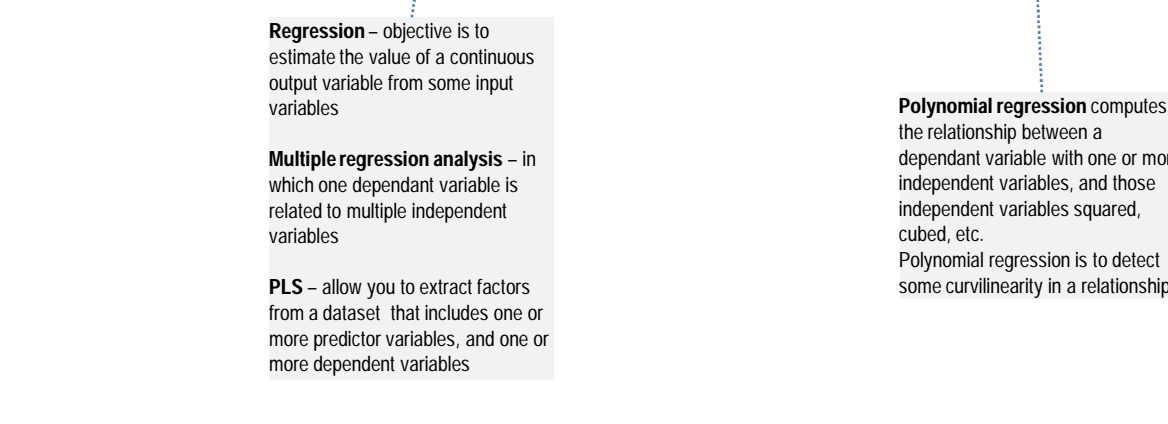
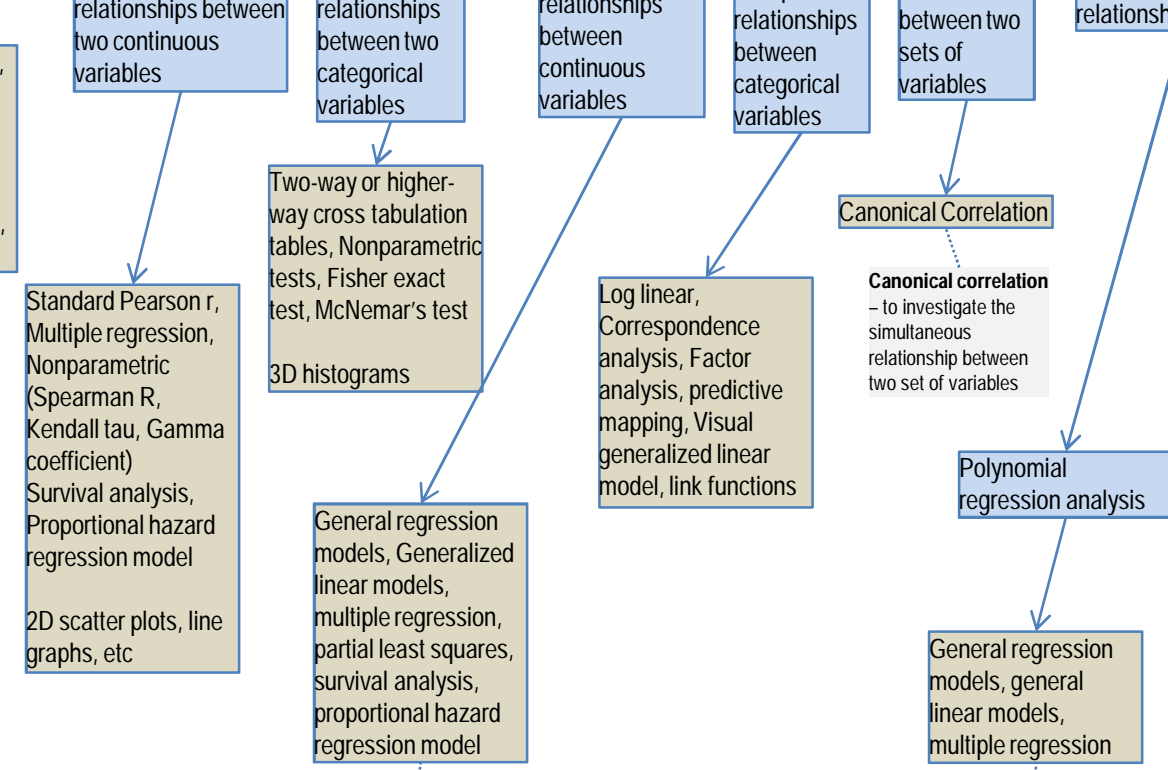
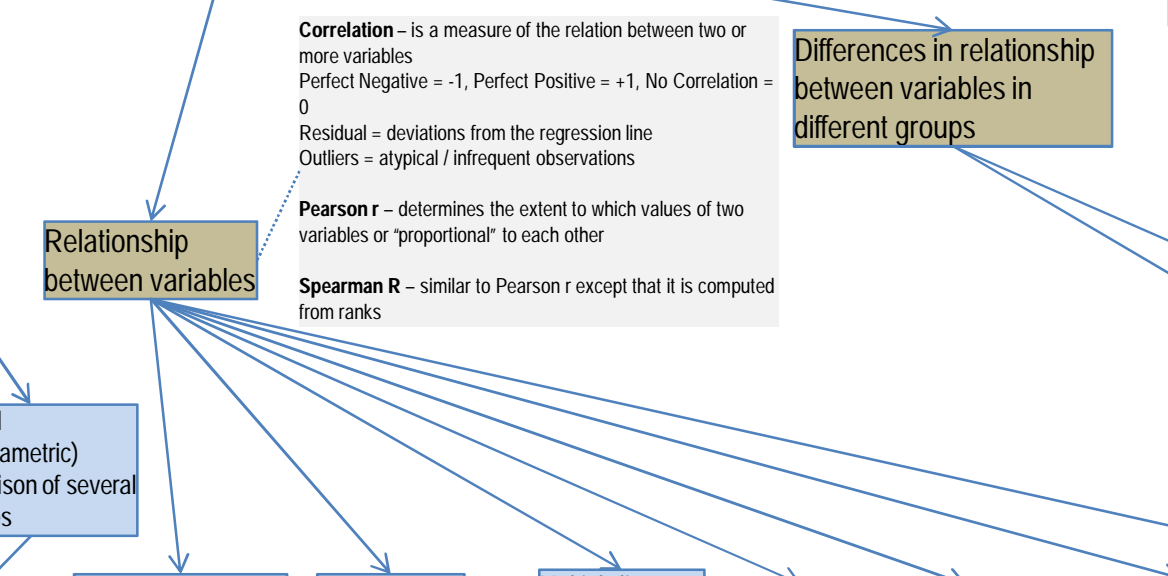
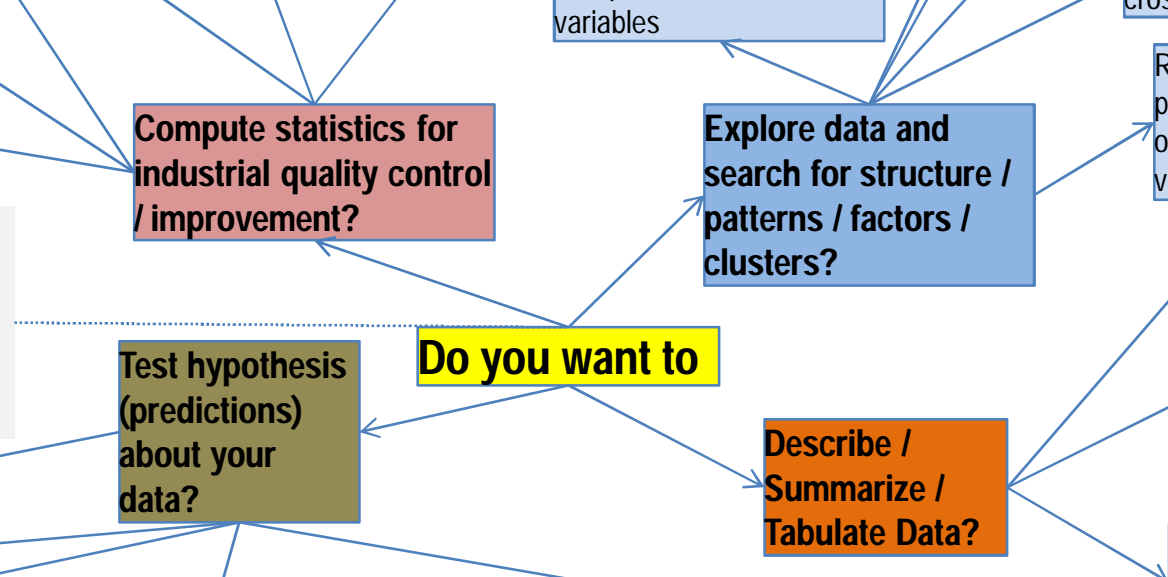
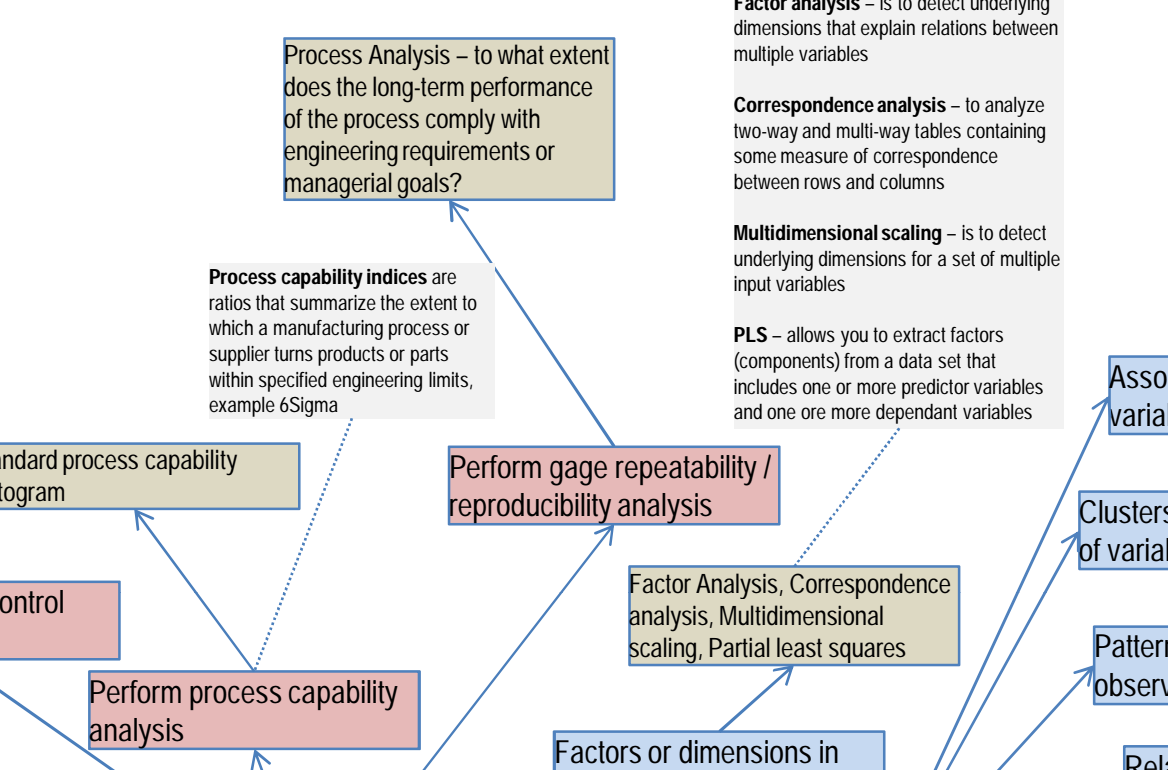
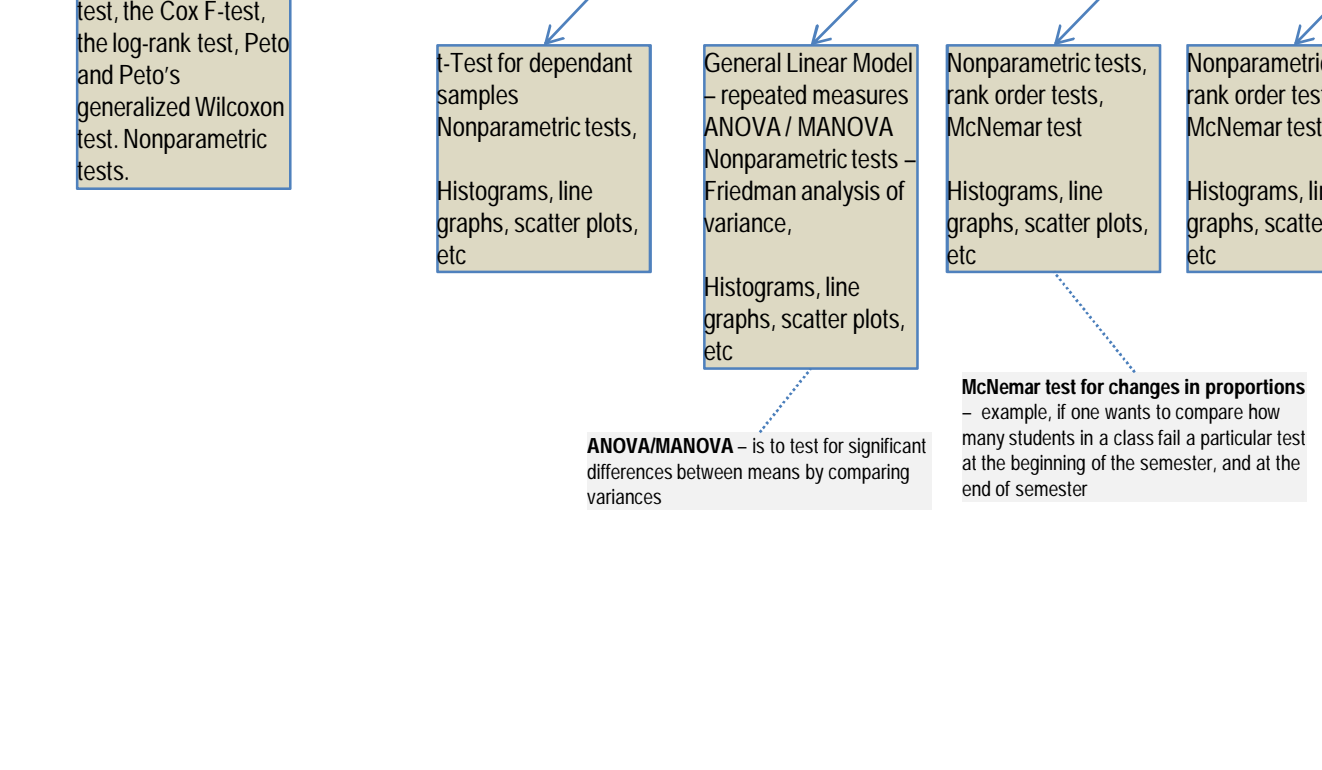
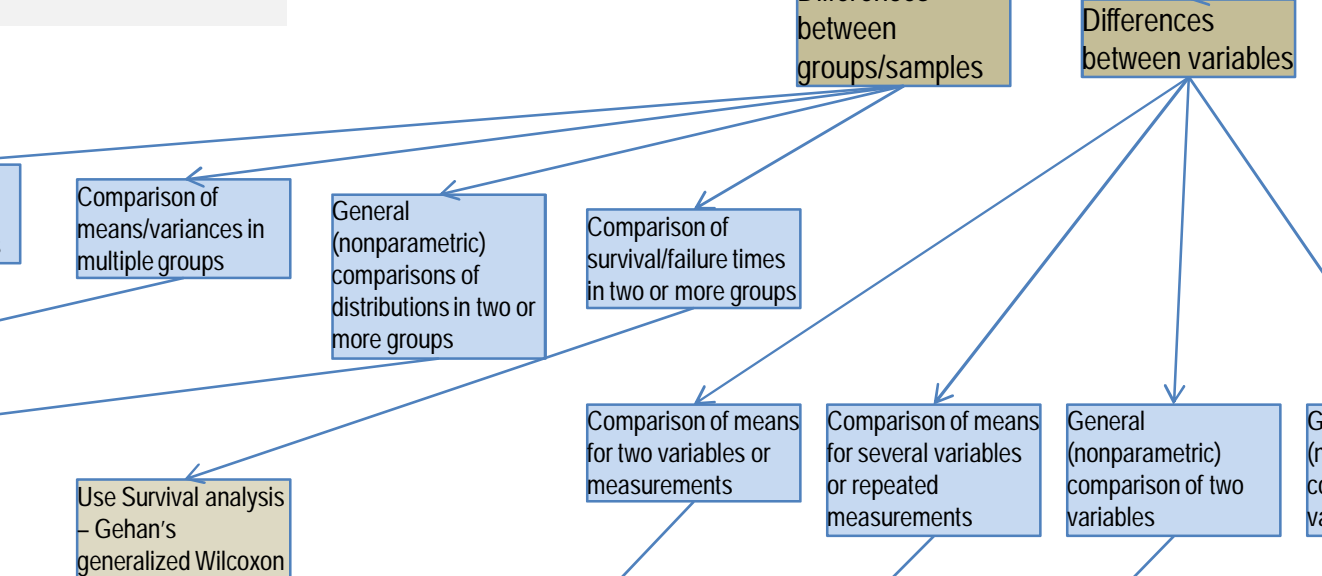
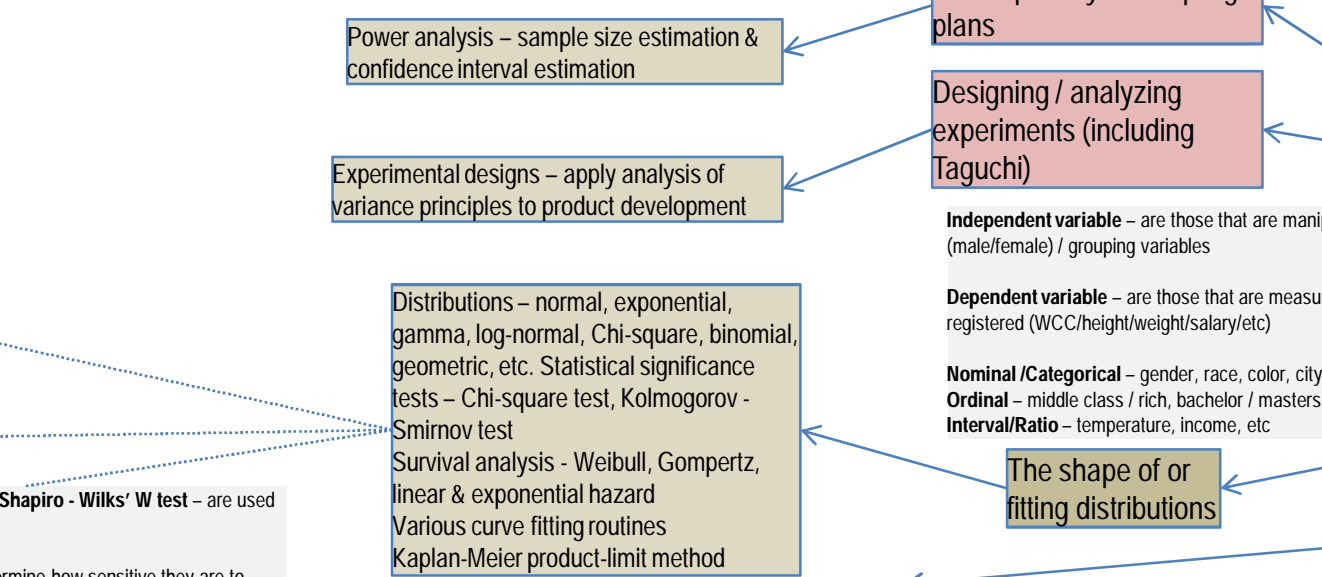
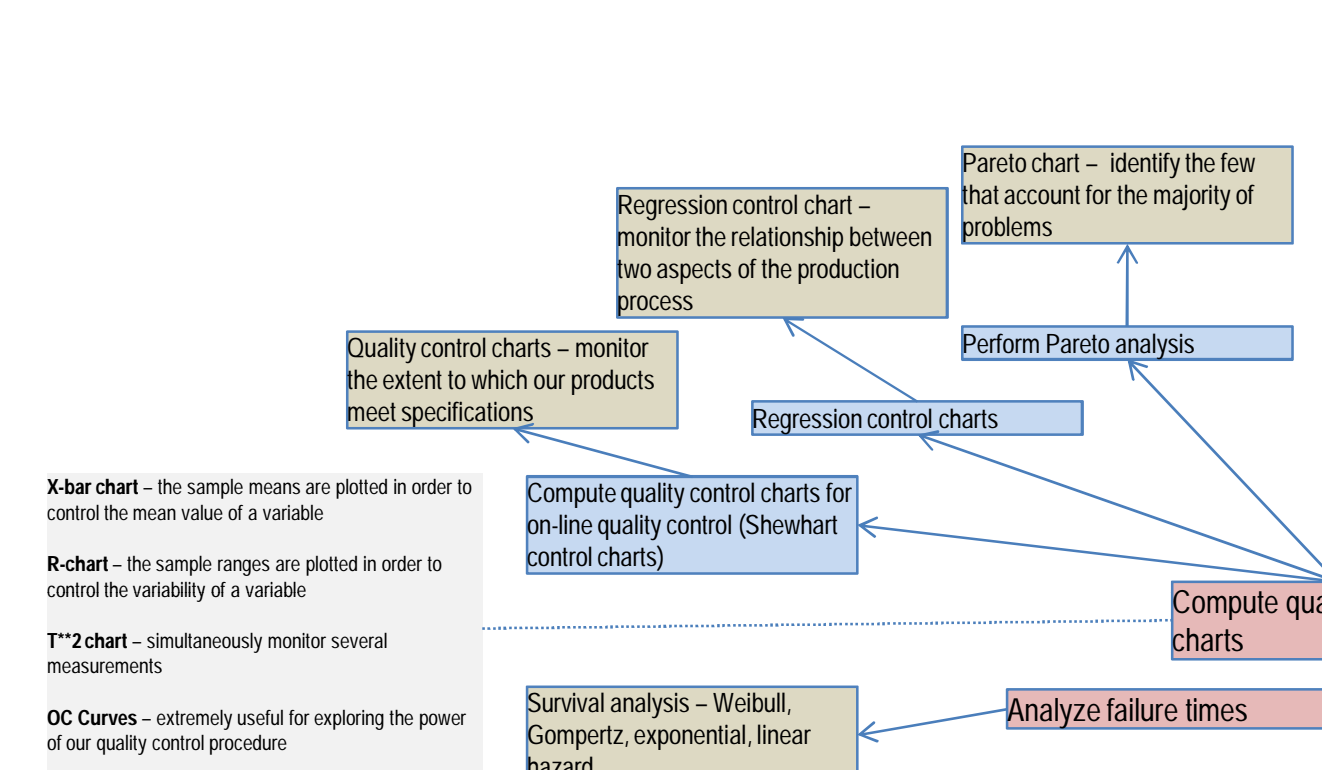
Generalized Linear Models (GLZ) – doesn't assume that the variables in the analysis follow normal distribution

Variance components & Mixed model ANOVA/MANOVA – techniques for analyzing research designs with random effects, including the estimation of variance components for such effects. It is also well suited for analyzing large main effect designs, and designs with many factors where the higher order interactions are not of interest, and analysis involving case weights

Discriminant function analysis – how to identify the specific variables that show different means in different groups

ANOVA/MANOVA – is to test for significant differences between means by comparing variances

McNemar test for changes in proportions – example, if one wants to compare how many students in a class fail a particular test at the beginning of the semester, and at the end of semester



Reference: StatSoft Online Textbook. Prepared By: Rajkumar BN (Draft version)