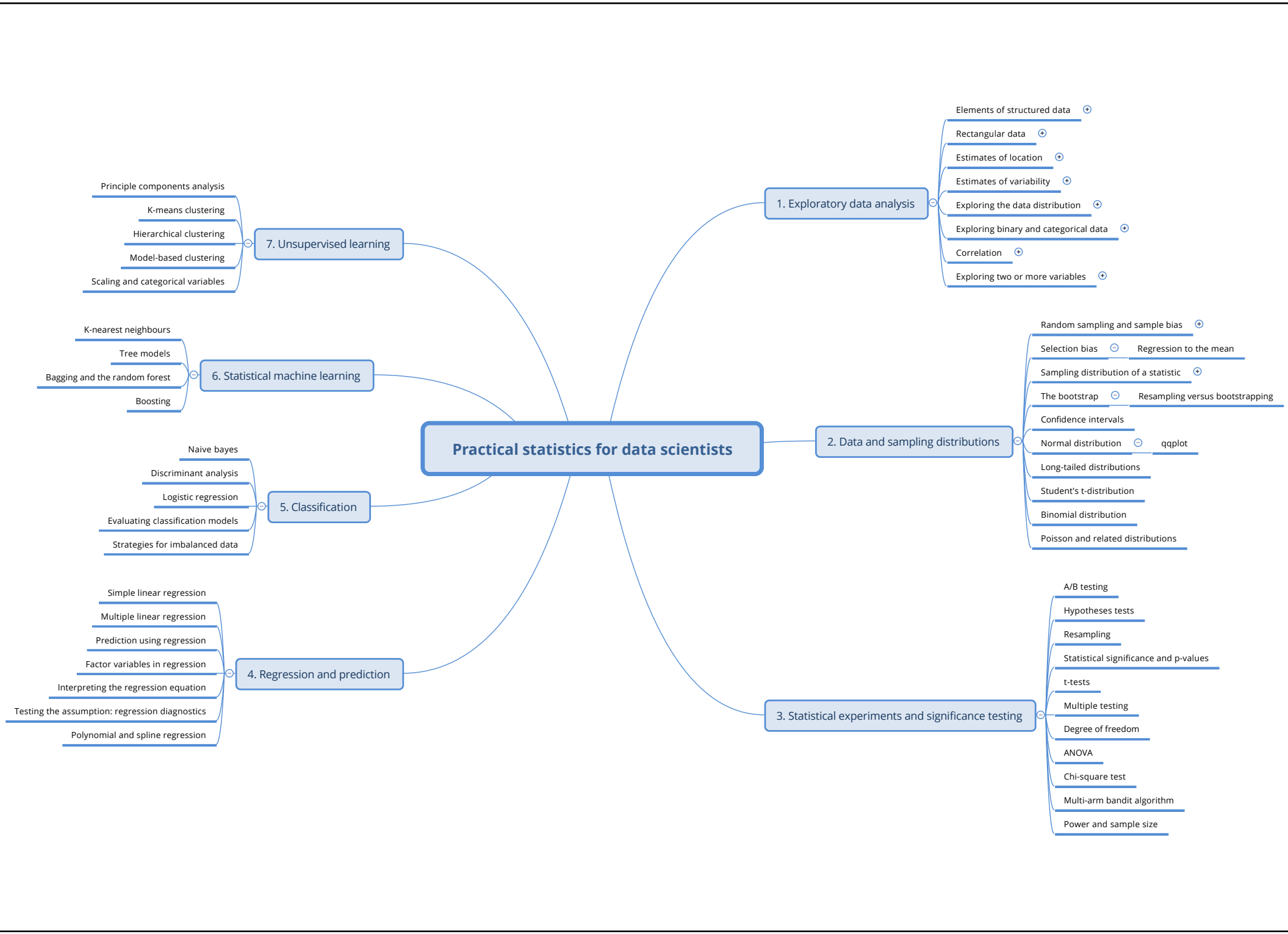
Practical statistics for data scientists
1. Exploratory data analysis
Elements of structured data
Rectangular data
Estimates of location
Estimates of variability
Exploring the data distribution
Exploring binary and categorical data
Correlation
Exploring two or more variables
2. Data and sampling distributions
Random sampling and sample bias
Selection bias
Regression to the mean
Sampling distribution of a statistic
The bootstrap
Resampling versus bootstrapping
Confidence intervals
Normal distribution
qqplot
Long-tailed distributions
Student's t-distribution
Binomial distribution
Poisson and related distributions
3. Statistical experiments and significance testing
A/B testing
Hypotheses tests
Resampling
Statistical significance and p-values
t-tests
Multiple testing
Degree of freedom
ANOVA
Chi-square test
Multi-arm bandit algorithm
Power and sample size
4. Regression and prediction
Simple linear regression
Multiple linear regression
Prediction using regression
Factor variables in regression
Interpreting the regression equation
Testing the assumption: regression diagnostics
Polynomial and spline regression
5. Classification
Naive bayes
Discriminant analysis
Logistic regression
Evaluating classification models
Strategies for imbalanced data
6. Statistical machine learning
K-nearest neighbours
Tree models
Bagging and the random forest
Boosting
7. Unsupervised learning
Principle components analysis
K-means clustering
Hierarchical clustering
Model-based clustering
Scaling and categorical variables