

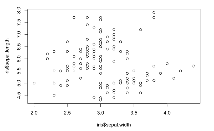
D3: Modelling and Exploration of Multivariate Data: : CHEAT SHEET



# CW 41-44 Introduction D3

## Q Part 1: parametric methods

* a) What is the structure of the dataset? visualising by doing a scatterplot in 2D/3D.



**library**(datasets)  
**data**(iris)

**plot**(iris$Sepal.Width,iris$Sepal.Length)

## Q Part 2: nonparametric methods

* a) What is the relation between Petal.Length and Petal.Width?
* b) What is the expected value for Petal.Lenght for Petal. Width = 1.55. Scatter Plot

idea of performing linear regression

fit1=**lm**(Sepal.Length~Sepal.Width,data=iris)

## Definiton

Stastistical Learning =

* characteristics of probability distributions from data samples
* interference, prediction

## Stastistical Inference

Infer properties of a population by testing hypotheses and deriving estimates

## Machine Learning

Very similar to statistical learning, perhaps more emphasis on algorithms and programming

## Data Mining

Find Pattern in data  
uses mainly Machine Learning methods  
includes also questions of databases/data storage

## Data Visualisation

Create, study and communicate visual representation of data   
find patterns in data

# Prerequisites

## 1. Variance

Covariance

Correlation

## 2. Generating Data Samples

## 3. Feature Scaling

## 4. Regression model