

Module Title: Multivariate data analysis
Module Code: DA0306
Maximum Number of Students: 60
Total ECTS Credits 2
Notional Learning Hours (a) Contact Time - 14h (b) Private Study - 36h Format of Teaching: Lectures 10 h Laboratories or Practicals 4 h Other 0 h Teaching Strategy: Formal lectures in 60/90 min timetable. Two supervised computer exercises.
Convener: B. Grung
University: University of Bergen
Language of Tuition: English
Module Description - The Purpose or Aims: <ol style="list-style-type: none"> 1. To introduce the concept of latent variables 2. Using latent variables to detect and describe the main variation patterns in a data set 3. Using Principal Component Analysis for exploratory data analysis and classification
Learning Outcomes: At the end of the module the student is expected to be able to: <ol style="list-style-type: none"> 1. Understand the mathematics behind latent variable methods 2. Perform an adequate pretreatment of a data set prior to multivariate analysis 3. Use commercial software to perform PCA on a data set 4. Use SIMCA classification to perform classification with and without a priori information
Summary of Course Content: Data sets from the modern analytical laboratory are usually highly collinear in nature. Modern analytical instruments produce huge amounts of data. To find the information present in such data sets require the use of multivariate modeling techniques. Principal Component Analysis (PCA) is demonstrated to be an extremely powerful tool for detection of sample and variable correlations. Various useful plots that aid the analyst in his quest for information are presented. A classification technique (SIMCA), based on PCA, is presented. Various common types of data pretreatment are presented.
Transferable Skills Taught: The ability to use commercially available software to perform explorative multivariate data analysis and multivariate classification. An understanding of the mathematics behind latent variable modelling.
Assessment Methods: Written exam (70%). Computer exercises (30%)

Assessment Criteria:Threshold

LO1 – Knowledge of variable and object space. The use of distance and angle as similarity measures.

LO2 – Column centring of data.

LO3 – Being able to perform a simple multivariate analysis under guidance

LO4 – Understanding how residuals are used to describing class belonging

Good

LO1 – Graphical definition of scores, loading and residuals.

LO2 – Normalization and standardization

LO3 – Being able to perform simple multivariate analysis without guidance

LO4 – Understanding the concepts of residual standard deviation for a sample and a class. Conceptual understanding of modeling and discriminating power. Being able to classify new samples.

Excellent

LO1 – Understanding the NIPALS algorithm for PCA

LO2 – Knowledge of more advanced pre-treatment techniques, such as differentiation and filtering.

LO3 – Being able to independently perform PCA on a data set; including the choice of proper pre-treatment.

LO4 – Understanding the mathematics behind SIMCA classification and cross validation.

Resource Implications of Proposal and Proposed Solutions:

Lecture notes will be available for students.

Recommended reading: R.G. Brereton (2003): *Chemometrics – Data analysis for the Laboratory and Chemical Plant*, Wiley