What is this package about?
Statistical analysis of data is essential to make valid statements on the association of observed variables. For instance,

- Is diabetes related to heart failure?
- What is growth curve of children?
- What is the shelf life of light bulbs?
- Is there a trend in the stock exchange?
- Is the heart rate of females more stable than of males?

To answer these types of questions it is important to model data that would explicitly quantify possible relationships between variables. The coherent package on statistics provides three related areas of statistical analysis that uses statistical models: **generalized linear models**, **survival and reliability models**, and **time series models**. Each of these topics is provided in a 5 ECT course. The courses provide theoretical background but also provide a lot of practice. You will work with the software packages SAS and [R].

### Course descriptions

#### Survival and reliability analysis JBM210
For certain types of studies, the time to an event is being captured. For instance, how long does it take before a light bulb fails or how long does it take for someone to develop cancer? When multiple units (light bulbs or people) are being measured we may model times to event with a certain probability distribution. This distribution may be affected by certain explanatory variables. For instance, would light emitting diodes fail faster than compact fluorescent lamps or do man develop cancer faster than woman? In many studies though, not all times to events are being observed, just because the events will take longer than the study length, which introduces censoring of data. Only a maximum time without having
observed an event is then being captured. This is an important and specific aspect of survival and reliability analysis. Additionally, in reliability studies it is sometimes possible to perform an accelerated study where products are intentionally put under additional stress (higher temperatures) to impact or shorten the survival times. This course discusses several forms of probability distributions for survival times as well as semi-parametric methods like Cox-proportional hazard models. It also explains how we can visualize survival times and how we can model survival times as function of one or more explanatory variables. A specific aspect is the estimation procedures. The different models will be applied to real data sets to practice the theory.

**Generalized linear statistical models JBM200**

This course discusses the most common and relevant statistical models that would connect a certain outcome to one or more different explanatory variables. This type of modeling plays a dominant role in many different sciences and disciplines and is referred to as generalized linear models. It contains models like linear regression and logistic regression. The models that will be discussed during the course would make it possible to answer questions like:

- How does the body mass index affect the risk of a heart attack and is this different for man and woman?
- How does advertisement improve the profit or revenue of a product and is this different for types of advertisements?
- How does processing time affect the number of product failures and is this different for different products?

The theory of the statistical models is based on three components. The first component focuses on parametric probability distributions that may fit with or describe the outcome of interest (heart attack, profit, number of failures). The second component would formulate specific linear models of the explanatory variables (body mass index and gender; amount and type of advertisement; processing time and type of product) that may affect the outcome. The third component would describe the relation between the parameters of the probability distribution and the linear model through what is called link function. This connects the outcome to the explanatory variables. The general theory is discussed together with appropriate methods of estimating the model parameters and several data sets will be analyzed to practice with the models on real data sets.

**Multivariate Data Analysis JBM220**

*Detecting patterns and relationships in multiple variable datasets*

Experimental studies and research may result in complex data sets containing several related responses. For their analysis traditional univariate, one-variable at a time, statistical methods may not be adequate as the analysis may gain from a simultaneous examination of all relevant variables. With Multivariate Data Analysis correlation or association between
variables can be taken into account, while modeling or analyzing several variables simultaneously. They provide the means for both describing and exploring data, aiming to extract the underlying patterns and structure.

In this course 'Multivariate Data Analysis' a range of multivariate techniques for analyzing correlated or associated variables are discussed and underlying principles are explained. Topics treated are: Exploratory Data Analysis for Multivariate Data, Multivariate Analysis of Variance: MANOVA. Data Reduction: Principal Component Analysis, Factor Analysis and Partial Least Squares Methods, Data Grouping: Discriminant Function Analysis and Cluster Analysis. They make it possible to:

- Link process settings to a sensory panel assessment,
- Calibrate a measurement procedure,
- Model a production process to increase quality and / or to reduce production losses.

During the course the use of representative statistical software is integrated and participants get the opportunity for hands-on experience in applying multivariate data analysis. After successful completion of the course participants should be able to perform multivariate data analyses independently.