What makes multivariate analysis different?

What makes multivariate analysis different?

First, Dr Pavel Krivitsky will discuss some general aspects of multivariate analysis in the video below.

Transcript

Usually, when studying complex phenomena, **many** variables are required. Besides, the process of studying is usually an iterative one with many variables often added or deleted from the study. Multivariate analysis deals with developing methods for better understanding the relationships between the many variables included in the analysis of such complex phenomena.

In previous studies, you may have learned about a variety of methods for analysing many variables. For example, you have probably learned about the *multiple regression* linear model:

$$Y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_p x_{i,p} + \epsilon_i,\ i = 1, \ldots, n$$

where Y_i is the ith observation of the response variable, $x_{i,k}$ ith observation of the kth predictor variable, and ϵ_i the ith error. However, in this regression, we designate the p predictors as fixed (conditioned on) and only *one* variable per observation is random. Typically, we assume that the ϵ_is and therefore Y_is are independent (conditional on the xs) or at least uncorrelated. Contrast this with a *multivariate* linear model,

$$\begin{aligned} Y_{i,1} &= \beta_{0,1} + \beta_{1,1} x_{i,1} + \beta_{2,1} x_{i,2} + \cdots + \beta_{p,1} x_{i,p} + \epsilon_{i,1}, \\ Y_{i,2} &= \beta_{0,2} + \beta_{1,2} x_{i,1} + \beta_{2,2} x_{i,2} + \cdots + \beta_{p,2} x_{i,p} + \epsilon_{i,2}, \end{aligned}$$

where $Y_{i,1}$ and $Y_{i,2}$ are the ith observations of two distinct response variables, and $\epsilon_{i,1}$ and $\epsilon_{i,2}$ may be correlated. The multivariate linear model can be used when multiple observations are taken on each individual in the sample, and it can allow us to model the relationships among these measurements.

Difficulties in such a process include:

- There is more data to analyse.
- More involved mathematics are necessary.
- Computer intensive methods are involved in the process.

Objectives of multivariate methods are the following:

- **Data reduction:** presenting the phenomenon as simply as possible **but** without sacrificing valuable information. Typical *representative method*: Principal components analysis. Sometimes, this reduction
  is achieved by introducing a small number of unobservable (latent) variables when trying to explain a large number of observable output variables. Representative methods: *factor analysis* and *covariance*
  structure analysis.
- **Sorting or grouping:** creating groups of "similar" objects or variables that in a sense are more closer to each other than to objects outside the group; and finding reasonable explanation for the existing grouping. *Representative methods*: Factor Analysis, Cluster Analysis, Discriminant Analysis.
- **Investigation of dependence among variables:** finding which sets of variables can be considered as independent and which are "more dependent"; and "measuring" the dependence. *Representative*
  Methods: Correlation Analysis, Partial Correlations, Canonical Correlations.
- **Prediction:** predicting values of one or more variables on the basis of observations of other variables that have been found to influence the former variables: a basic but important goal. *Representative*:
  Multivariate Regression.
- **Hypothesis testing:** either validating assumptions (e.g., normality) on the basis of which certain analysis is being done or to reinforce some prior modelling convictions (e.g., equality of parameters). Hypothesis testing is relevant to the applications of all multivariate methods we will be dealing with.

As a basic **mathematical model** for our analyses in this course the **multivariate normal distribution** will be used. The reasons for this are our limited time and the complexity of other approaches. Although in practice also other distributions are relevant, modelling based on the multivariate normal distribution can still be a very good approximation.

Watch: Copula stock return

Watch the copula stock example below by Dr Pavel Krivitsky.

Transcript