Multivariate Location and Scatter
Lecture 9
Regression Analysis

Pauliina Ilmonen

Outline

Multivariate Regression Analysis

Multivariate Regression Analysis

Consider the linear regression model

$$Y = X\beta + \varepsilon,$$

where $Y \in \mathbb{R}^{n \times p}$ is a data matrix that contains the observed values of p response variables, $X \in \mathbb{R}^{n \times q}$ is a data matrix that contains the observed values of q explanatory variables and $\varepsilon \in \mathbb{R}^{n \times p}$ is the matrix of the residuals.

We assume that the residuals are i.i.d. and centered with respect to some location score function T (i.e. $E[T(\varepsilon_i)] = 0$).

Multivariate Regression Analysis

We make the following technical assumptions:

$$\frac{1}{n} XX^T \rightarrow_{wp1} D, \quad \text{as} \quad n \rightarrow \infty$$

and

$$\max \frac{x_i^T C^T C x_i}{\sum_{i=1}^{n} x_i C^T C x_i} \rightarrow_{wp1} 0, \quad \text{as} \quad n \rightarrow \infty$$

for some positive definite $D \in \mathbb{R}^{q \times q}$ and for all $C \in \mathbb{R}^{p \times q}$ with rank > 0.

Significance Testing

Significance Testing

Consider the null hypothesis

$$H_0 : \beta = 0.$$

Note that under the null hypothesis, we have that $E[T(y_i)] = 0$.

Significance Testing

Let $B = E[T(\varepsilon_i)T(\varepsilon_i)^T]$ be finite and let $T(Y) = (T(y_1), T(y_2), ..., T(y_n))^T$.

Now, under the null, we have that

$$\frac{1}{\sqrt{n}}\text{vec}(T(Y)^T X) \to_d N_{pq}(0, D \otimes B)$$

and the test statistic

$$Q(X, Y) = n{\cdot}\text{trace}(T(Y)^T X(X^T X)^{-1} X^T T(Y)(T(Y)^T T(Y))^{-1}) \to_d \chi^2(pq).$$

Significance Testing

Alternative version of the test statistic above can be obtained by inner standardization.

Estimation

Estimation

Estimation is usually based on applying a score function such that the estimate $\hat{\beta}$ solves the equation

$$T(\hat{\varepsilon})^T X = 0.$$

We search for an estimate $\hat{\beta}$ such that the standardized estimated residuals do not correlate with the explanatory variables.

Examples

Examples

L_2 regression...

Spatial signs and spatial ranks...

Spatial Sign

The spatial sign $U(x_i)$ is defined as

$$U(x_i) = \begin{cases} x_i(||x_i||)^{-1} & x_i \neq 0, \\ 0 & x_i = 0. \end{cases}$$

Spatial Rank

The spatial rank $R(x_i)$ is defined as

$$R(x_i) = \frac{1}{n}\sum_{j=1}^{n} U(x_i - x_j).$$

References

H. Oja, Multivariate Nonparametric Methods With R, Springer-Verlag, New York (2010).