Multivariate statistics
Lecture 06:
Univariate Multiple Linear Regression, UMLR

Literature and Problems: See course schedule

Agenda (preliminary):

SLIDE: Overview of linear regression models and terminology

Univariate Multiple Linear Regression

- model: response variable Y, predictor variables $z_1,\ldots,z_r$, error term ε and unknown regression coefficients $\beta_0,\ldots,\beta_r$
- n observations
- assumptions
- model and observations on matrix form, $Y = Z\beta+\varepsilon$
- Least Squares estimation of β --> estimates Y_{hat} and ε_{hat}
- eigen decomposition of Z^TZ with full rank --> $(Z^TZ)^{-1}$
- Z^TZ with rank < r+1 --> pseudo-inverse $(Z^TZ)^{\%}$
- mean and covariance matrix of β_{hat} and ε_{hat}
- estimation of $Var[\varepsilon] = \sigma^2$
- Sum of Squares decomposition $SS_T = SS_R + SS_E$
- coefficient of determination, R^2 and R_{adj}^2

Inference on β

- MVN assumption on ε --> MVN Y, MVN β_{hat}, MVN ε_{hat}, chi-square $s^2 = \sigma_{hat}^2$
- confidence region for β
- simultaneous and Bonferroni confidence intervals for β_i
- test of hypothesis H_0: $\beta = \beta_0$
- test of hypothesis H_0: $\beta_{(2)} = 0$ (where $\beta = [\ \beta_{(1)}\ \beta_{(2)}\ ]$) using the extra-sum-of-squares concept, $ESS(\beta_{(1)}\text{-->}\beta) = SS_R(\beta)-SS_R(\beta_{(1)}) = SS_E(\beta_{(1)})-SS_E(\beta)$
- test of significant regression model, H_0: $\beta_1 = \ldots = \beta_r = 0$

Prediction (inference concerning new observations)

- estimation of mean of new observations at $z = z_0$, $E[Y(z_0)]$
- estimation of single new observations at $z = z_0$, $Y(z_0)$

Model check (analysis of residuals ε_{hat})

- independence ?
- constant variance ?
- normal distributed ?

ANOVA as a special case of UMLR

- UMLR: z_i continuous predictor variables --> ANOVA: z_i indicator variables
- SLIDE: example from textbook
- ESS concept can be used to test for interactions and factor effects