UNIVERSITY OF EDINBURGH
Business School

Predictive Analytics and Modelling of Data

CMSE11428 (2020-2021)

Dr Xuefei Lu
The University of Edinburgh Business School

standardized (regression) coefficients

If you want to use multiple regression coefficients to determine variable importance, use 'standardized (regression) coefficients'

$$\beta_i^* = \frac{s_i}{s_y}\beta_i,$$

where s_i and s_y are the estimated standard deviations of X_i and Y

Logistic Regression

$$\ln\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

- **Appropriate outcome structure**
  Dependent variable is binary
- **Linearity of independent variables and log odds**
  logistic regression does not require a linear relationship between the dependent and independent variables, but linear relation between independent and log of odds
- **The absence of multicollinearity**
  The logistic regression assumes that there is minimal or no multicollinearity among the independent variables.
- **Observation independence**
  The Logistic regression assumes the observations to be independent of each other.
- **Large sample size**
- Minimum of 10 obs with the leaset frequent outcome of each independent variable:
- 4 predictors, expected prob of least frequent outcome 0.1 ➔ n = 10*4/0.1 = 400

UNIVERSITY OF EDINBURGH
Business School