

# Introduction to R

## *Session 8 exercises*

Statistical Consulting Centre

2 March, 2017

## 1 Linear regression

- (i) Perform a linear regression between age (explanatory variable) and nerdy score (dependent variable).
- (ii) Are the estimated intercept and slope significantly different from zero?
- (iii) Examine the residuals of the fitted linear model.
- (iv) Add the fitted line to the scatterplot of nerdy score against age.
- (v) What conclusions can you draw? Do you think age and nerdy score are linearly correlated?

## 2 Logistic Regression

### 2.1 Continuous explanatory variable

- (i) Suppose we want to model the probability of being male, i.e., `gender = Male`. First, ensure that `gender` is a variable with a correct type.
- (ii) Fit a logistic model with `gender` as the response variable and `nerdy.sc` as the explanatory variable.
- (iii) Perform an analysis of deviance to determine the overall significance of `nerdy.sc`.
- (iv) Calculate the estimated slope of the logistic regression. What can you conclude about the slope?

### 2.2 Categorical explanatory variable

- (i) We now want to model the probability of living with a partner given age group. `partner` is already of type `factor`. Now, generate a one-way table of `partner` to examine its contents.
- (ii) Set `partner = Yes` as the reference level.
- (iii) Once again generate the one-way frequency table of `partner`.
- (iv) Fit a logistic model with `partner` as the response variable and `age.group` as the explanatory variable.

- (v) Is **age.group** a significant predictor of whether or not an individual in particular age group has a partner?
- (vi) Generate a two-way frequency table of **partner** against **age.group**.
- (vii) Convert these frequencies to percentages of age group total. Does this table agree with your earlier conclusion?