

# Statistical Methods and Data Analysis in Developmental Psychology

## Exam Simulation

Prof Antonio Calcagni, Dr. Filippo Gambarota

The dataset `anxiety.rda` contains data about adolescents with or without anxiety disorders. The dataset contains the following variables:

- **anx**: having (1) or not (0) the anxiety disorder
- **selfesteem**: the self-esteem score from 0 (low self-esteem) to 10 (high self-esteem)
- **socialnetwork**: the estimated size of the social network from 0 (very small) to 100 (big)
- **age**: the age in years
- **family**: If the family had (“yes”) or not (“no”) history of anxiety disorders

The dataset can be loaded using `load()`. Make sure to check the dataset structure and whether categorical variables are interpreted as factors from R.

### 1 Problem: Identify the number of statistical units **n** and the type of variables

**a)** 190 observations, 2 categorical variables and 2 numeric variables **b)** 200 observations, 1 categorical variables and 4 numeric variables **c)** 205 observations, 1 categorical variables and 4 numeric variables **d)** 200 observations, 5 categorical variables and 0 numeric variables

- 2 Problem: Make an appropriate plot of univariate distributions of predictors response variable
- 3 Problem: Calculate from observed data the odds ratio of having anxiety as a function of family anxiety history
- 4 Problem: Define and fit an appropriate additive model to predict the probability of having anxiety as a function of selfesteem and family. Intepret the results.
- 5 Problem: From the fitted model, find the probability and the 95% confidence interval that a subject without anxiety familiarity and self esteem = 3 has axiety disorders.
- 6 Problem: fit a model including also the social network effect and intepret the parameters of numerical predictors using the divide by 4 rule.
- 7 Problem: perform a statistical test to compare the residual deviance of the null model, model with and model without the socialnetwork predictor and intepret the result.
- 8 Problem: extract the DFBETAs value from the model created in the previous step. Identify (if any) problematic observations (i.e., marked as outlier in all coefficients but the Intercept) using a cut-off of  $2/\sqrt{n}$  ( $n$  is the sample size) and intepret the results. In case of problematic observations re-fit the model without that observations and comment the results.
- 9 Problem: The classification accuracy of the model created at Problem 6 (`anx ~ family + selfesteem + socialnetwork`) is:

- a) 0
- b) 0.65
- c) 33
- d) 0.85

- 10 Problem: Fit the same model considered in the previous problem but using the median-centered self-esteem. How parameters interpretation change?