

1. In file “data3.txt” (zipped with the assignment) you will find the recorded variables Y, X1, X2, X3 (continuous) and W (categorical with three levels) on 150 cases. Using these data answer the following questions:

(a) Run the parametric one-way ANOVA of each of the continuous variables (Y, X1, X2, X3) on the categorical variable (W). Specifically,

- (i) provide a graphical representation of each of the continuous versus the categorical variable
- (ii) provide the ANOVA output
- (iii) check the assumptions and provide alternatives if any of them fails.
- (b) Provide a scatter-plot matrix of Y, X1, X2, X3, annotating the different levels of W in each plot using a different color.
- (c) Run the regression model of Y on X1
- (d) Run the regression model of Y on all the remaining variables (X1, X2, X3, W), including the non-additive terms (i.e. interactions of the continuous predictors with the categorical).
- (e) Examine the regression assumptions and provide alternatives if any of them fails.
- (f) Use the “stepwise regression” approach to examine whether you can reduce the dimension of the model. Provide the interpretation of the coefficients of the final model.
- (g) Using the `cut()` function, create a categorical variable (named Z) with 4 levels based on the quantiles of X3. Provide the contingency table of X3 and W.
- (h) Run the parametric two-way ANOVA of Y on the categorical variables W and Z (including only the main effects). Provide the fit, examine the assumptions and comment on the significance of the terms.