6 Data Science (djw1005)



This plot shows a collection of datapoints $(x_i, y_i, group_i)$ from two groups, together with fitted lines for each group. Based on our scientific understanding of the dataset, the lines should be parallel.

(a) Describe a probability model and a fitting procedure that might be behind this plot. Explain how you enforce the 'parallel lines' requirement. Give pseudocode for the fitting procedure. Explain briefly how the plot is produced. [Note: When describing a probability model, you should state your assumptions: what probability distributions are you using, and why have you chosen them?]

[7 marks]

- (b) We would like to report our confidence about the offset between the two groups. Explain the Bayesian approach to computing this confidence, and give pseudocode. [7 marks]
- (c) We have been asked to predict y at some new value x^* , for both groups. Explain how to compute confidence intervals for these two predictions, and give pseudocode. [6 marks]