

Analysis Plan

Emma & Rob

2021/02/24

Hypothesis

Based on previous exploratory data analysis (found in `old_faithful_eda.pdf`) we propose the following hypothesis:

There are 2 distinct types of eruption from Old Faithful: 1) short frequent eruptions, 2) long infrequent eruptions.

We will use the dataset `oldfaithful` available from the R package `datasets`. See [here](#) for the documentation.

Analysis plan

We will fit two Gaussian mixture models with 1 and 2 components. The number of components will be determined by the model with the highest Integrated Complete Data Likelihood Criterion (ICL). This criterion is similar to the Bayesian information criterion but is more suited to classification problems.

The true number of components will be given by the model with the highest Integrated Complete Data Likelihood (ICL). The model will be fit in the R package `RMixmod`.