

Exercises Day 4

PSY8003

Matthias Mittner

spring 2022

Exercise 1

The main variable here is `yesno` which indicates if the email is spam or not. The other variables are explanatory variables. They are:

- `crl.tot` - total length of words that are in capitals
- `dollar` - frequency of the \$ symbol, as percentage of all characters
- `bang` - frequency of the ! symbol, as a percentage of all characters,
- `money` - frequency of the word money, as a percentage of all words,
- `n000` - frequency of the text string 000, as percentage of all words,
- `make` - frequency of the word make, as a percentage of all words.

The goal is mainly to predict whether a future email is spam or not based on these explanatory variables.