

NIMBLE Project Report

Tristan Sones-Dykes

Date: November 06, 2023

Data

The input data format is as follows:

| yi | xi | zi |
|-----|-----|-----|
| ... | ... | ... |

The variables represent the total length of time spent telephoning (yi), the number of telephone calls made (xi), and a binary feature for weekdays (zi).

Model Specification

Likelihood

The model assumes a linear relationship between the response variable (yi) and the explanatory variables (xi and zi) as follows:

$$yi = \alpha + \beta xi + \gamma zi + \epsilon_i$$

Where $\epsilon_i \sim N(0, \sigma^2)$ represents the error term.

Priors

The prior distributions of the variables are as follows:

- $\alpha \sim N(\mathfrak{t}_\alpha, \sigma_\alpha^2)$
- $\beta \sim N(\mathfrak{t}_\beta, \sigma_\beta^2)$
- $\gamma \sim N(\mathfrak{t}_\gamma, \sigma_\gamma^2)$
- $\sigma^2 \sim \Gamma^{-1}(a, b)$

(1) Vague priors for these are:

- $\alpha \sim N(0, 10^5)$
- $\beta \sim N(0, 10^5)$
- $\gamma \sim N(0, 10^5)$
- $\sigma^2 \sim \Gamma^{-1}(0.0001, 0.0001)$

This ensures both the normals and inverse gamma is very flat, and therefore vague.

Data Analysis

Exploration

(2a) The data is plotted below:

```
ggplot(df, aes(x = xi, y = yi, colour = factor(zi))) +  
  geom_point() +  
  labs(x = "Number of calls",  
       y = "Time spent telephoning (mins)",  
       colour = "Weekday",  
       title = "Time spent telephoning vs. number of calls")
```

