CS7DS3 Assignment 1

March 28, 2025

To be submitted by **1 p.m.** on **Monday 21st April, 2025**. Submit solutions through blackboard. Legible scans are fine for some questions but please be aware that the Turnitin submission requires a minimum amount of printed text.

Please remember to print your **name** and **student number** on your submission.

If you have any questions about the assignment, email me: arwhite@tcd.ie.

Problem set - 15 marks total

The time T until failure of electrical components can be modelled using an exponential distribution, with rate parameter $\theta > 0$, so that $T \sim \mathcal{E}(\lambda)$. This distribution has a pdf of the form

$$p(t \mid \theta) = \theta \exp(-\theta t),$$

with $\mathbb{E}[T] = 1/\theta$.

1) Discuss the similarities between the form of $p(t \mid \theta)$ and the standard exponential family form, which we defined in class to be:

$$p(t \mid \theta) = h(t)g(\theta) \exp\{\phi(\theta)s(t)\}.$$

Clearly identify how the natural parameter $\phi(\theta)$, sufficient statistic $s(t)$, normalising constant $h(t)$ and auxillary function $g(\theta)$ correspond to different elements of the exponential distribution in this case.

[2 marks]

2) Suppose that covariates x_1 and x_2 are also available for analysis. A log-linear model can be used to model time to failure in this case. Specifically, we assume that the rate parameter θ can be expressed in the form $\theta = \exp\left(-\beta_0 - \beta_1 x_1 - \beta_2 x_2\right)$.

i) Is choosing to model θ in a log-linear setting (i.e., as an exponential function of the covariates) a surprising choice in this case? Explain your answer. **[2 marks]**

ii) Show that under this model, $\mathbb{E}[T] = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$. **[1 marks]**

3) A study was performed in which $n = 100$ electrical components were observed to fail at times $t_1, \ldots, t_n$ respectively. The type of component, x_1, and the grade of material x_2 used to construct the component, x_2, were also recorded. Two types of component (A or B) were available for the study, with $x_1 = 1$ if the component was Type B, otherwise $x_1 = 0$. Material was graded on a scale from 1 to 5, 1 indicating highest quality material and 5 indicating lowest quality. For the purposes of the analysis, it was agreed that x_2 could be treated as being continuous.

An analysis of this data set was performed using the `brms` package, with selected output shown on pp. 3–4. Use this output to address the questions below. Justify your answer in each case.

i) Was MCMC performance satisfactory in this case? **[2 marks]**

ii) How should the intercept parameter β_0 be interpreted for this model? **[2 marks]**

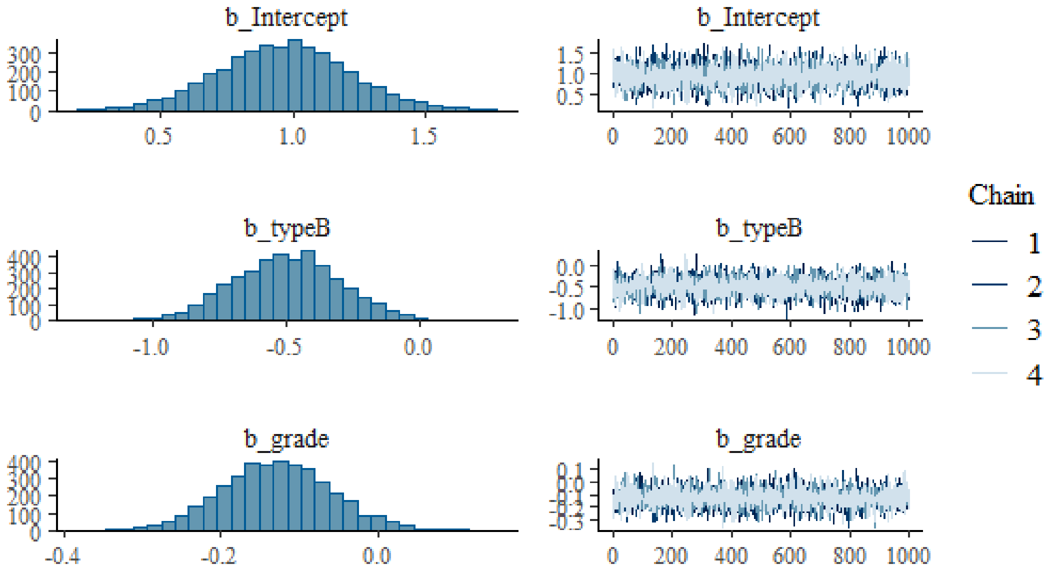
iii) Is there any evidence that the Type A and Type B components are different, and if so, which type of component is better, i.e., takes a longer time to fail on average? **[3 marks]**

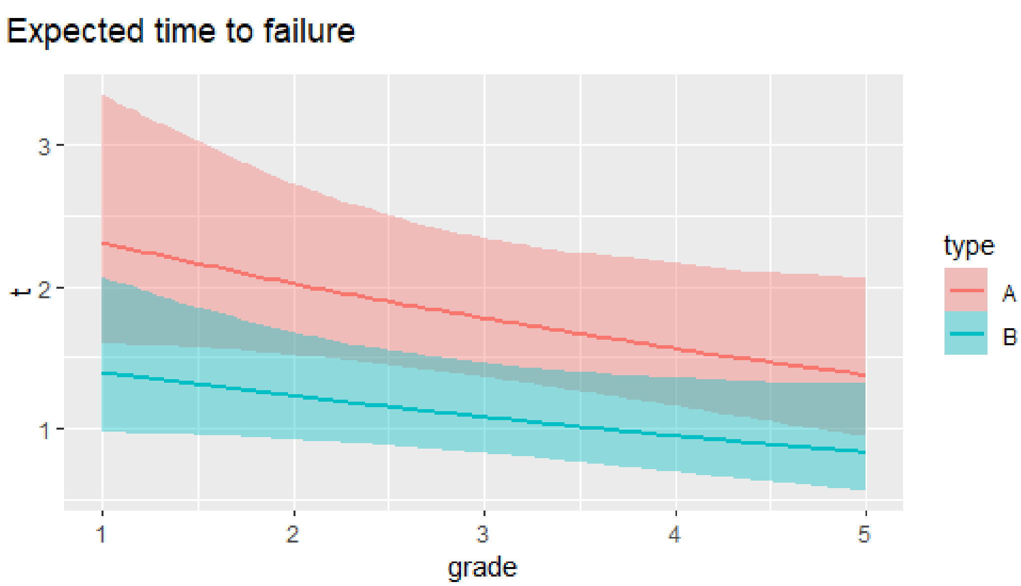
iv) In your opinion, what is more important to take into account when assessing time to failure for a component, the type or grade of component? **[3 marks]**

```
> summary(fit)
 Family: exponential
  Links: mu = log
Formula: y ~ type + grade
   Data: dat (Number of observations: 100)
  Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup draws = 4000

Regression Coefficients:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     0.96      0.24     0.49     1.46 1.00     4172     2637
typeB        -0.50      0.20    -0.88    -0.11 1.00     3728     3197
grade        -0.13      0.07    -0.27     0.01 1.00     4255     2966

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```


b_Intercept
b_Intercept
b_typeB
b_typeB
b_grade
b_grade
Chain
1
2
3
4


Expected time to failure
t
grade
type
A
B