STSB6816 Test 3 of 2023

Mathematical Statistics and Actuarial Science; University of the Free State

2023/06/13

## Time: 180 minutes; Marks: 50

# Instructions

* Answer all questions in a single R Markdown document. Please knit to PDF or Word at the end and submit both the PDF/Word document and the “.Rmd” file for assessment, in that order.
* Label questions clearly, as it is done on this question paper.
* All results accurate to about 3 decimal places.
* Show all derivations, formulas, code, sources, and reasoning.
* Intervals should cover 95% probability unless stated otherwise.
* No communication software, devices, or communication capable websites may be accessed prior to submission. You may not (nor even appear to) attempt to communicate or pass information to another student.

# Question 1

The data is provided at <https://ufs.blackboard.com>. It contains data from an experiment on the “Pharmacokinetics of Theophylline”. 12 subjects (Subject) were each weighed (Wt) and given a slightly different dose (Dose) of this substance at time 0. Their blood concentration (conc) was measured over time (Time). Your goal is to predict the log blood concentration curve of a random future subject.

We will assume that the log concentration curves follow the formula $η+λt$. $η$ (eta) measures where the curve would start if absorption was instantaneous, and $λ$ (lambda) measures how the concentration drops over time $\left(t\right)$.

We can then construct regression Model 1 by assuming an error distribution around the curve:

$$\begin{matrix}y\_{i}&∼t\left(ν,μ\_{i},σ\right), i=1…n\\μ\_{i}&=λ\_{s\_{i}}t\_{i}+η\_{s\_{i}}\\λ\_{j}&∼N\left(λ\_{0},τ\_{1}^{2}\right), j=1…n\_{s}\\η\_{j}&∼N\left(η\_{0},τ\_{2}^{2}\right)\\lnπ\left(ν,σ,τ\_{1},τ\_{2},λ\_{0},η\_{0}\right)&=-2logσ+logν-3log\left(ν+0.75\right)-2logτ\_{1}-2logτ\_{2}+k\\where s\_{i}& denotes the subject number of observation i\\n\_{s}& denotes the number of subjects\\n& denotes the number of observations in total\end{matrix}$$

Note that Model 1 does not consider any explanatory variables other than the random effects induced by the assumption that each subject has their own curve. We are interested in the average curve, that will hopefully be indicative of a random future subject. Usually, one might model the correlation between the random intercept and random slope parameters explicitly, but the implied correlation will suffice today.

**1.1)** What does modelling the data on the log scale as in Model 1 imply with regard to the variation (in terms of standard deviation) around the curve on the two scales? **[3]**

**1.2)** Import the data set into R and explore it visually. You could draw line plots with a line for each subject, perhaps coloured by an explanatory variable; or a table of averages per subject next to their dose and weight. Discuss what you see. **[5]**

**1.3)** Fit Model 1 on this data and discuss your estimates of $η\_{0}$ and $λ\_{0}$, along with their 95% intervals, in both statistical terms and practical terms. **[14]**

The bio-availability of the substance is related to the Area Under the Curve (AUC). We are most interested in the area under the blood concentration curve (not log) between hours 2 and 14 specifically. Assuming the model fits, the

$$AUC≈0.1\sum\_{i=1}^{121}exp\left(y\_{t\_{i}}^{new}|y\right), t\_{i}=2,2.1,2.2,…,13.9,14$$

**1.4)** Illustrate the posterior density of AUC for a random future subject. Please include only the lower 98% of predictions in any graphical illustration. [Partial credit will be given for a rough estimate of AUC.] **[8]**

Now consider the explanatory variables *weight* and *dose*. Including them as part of the intercept produces Model 2, which has the following changes:

$$\begin{matrix}μ\_{i}&=η\_{s\_{i}}+λ\_{s\_{i}}t\_{i}+β\_{1}\*w\_{j}+β\_{2}\*d\_{j}), j=1…n\_{s}\\w\_{j}& denotes the standardised weight of subject j\\d\_{j}& denotes the standardised dose of subject j\end{matrix}$$

**1.5)** Standardise the explanatory variables using the mean-standard deviation approach, then fit the model with standardised explanatory variables and give estimates of those coefficients (betas). **[7]**

**1.6)** Compare the fit of the two models, and then explain what your model comparison implies regarding the significance of the explanatory variables as a set. **[6]**

**1.7)** Plot the data of any one observed subject from the experiment. On the same plot show the fitted curve of that subject and 95% prediction intervals around the curve. You may use either the log or original scale. **[7]**