STSB6816 Test 2 of 2022

Mathematical Statistics and Actuarial Science; University of the Free State

2022/05/26

## Time: 180 minutes; Marks: 50

# Instructions

* Answer all questions in a single R Markdown document. Please knit to Word or PDF 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 at least 1 decimal place, ensure that simulation error is small enough (by doing enough simulations).
* Show all derivations, formulas, code, sources and reasoning.
* Intervals should cover 95% probability unless stated otherwise.
* No communication software, no devices, and no 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

A patient is asked to take their blood sugar readings every morning for four weeks. They present the following to the doctor:

8.6, 5.6, 7.2, Forgot, 6.9, 6.8, 5.4, 6.8, 5.9, Didn’t have time, 4.4, 5.9, 4.8, 5.2, 3.7, 5.3, ? (illegible), 3.4, 4, 4.2, 3.1, BDL, 3.6, BDL, 3.2, BDL, BDL, BDL; where BDL means below the detection limit of 3.0 on the home machine.

The doctor asks you to tell them what the missing and censored values might have been, with uncertainty.

**1.1)** Capture the data. Ensure that you have columns for Day (1 to 28), Values (numeric), Missing (indicator with 3 1s), and BDL (indicator with 5 1s). Check that you have captured the data correctly by giving a summary and showing that the observed values sum to 104. The data is also given in tabular format below if that makes capturing easier for you. **[ 4 ]**

|  | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | Day 6 | Day 7 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Week 1 | 8.6 | 5.6 | 7.2 | Forgot | 6.9 | 6.8 | 5.4 |
| Week 2 | 6.8 | 5.9 | Didn’t have time | 4.4 | 5.9 | 4.8 | 5.2 |
| Week 3 | 3.7 | 5.3 | ? (illegible) | 3.4 | 4 | 4.2 | 3.1 |
| Week 4 | BDL | 3.6 | BDL | 3.2 | BDL | BDL | BDL |
|  |  |  |  |  |  |  |  |

**1.2)** Fit a linear model with Student-t errors through the points on the log scale (equivalent to an exponential slope on the original scale). Use the objective log prior $logπ\left(v,m,s\right)=log\left(v\right)-3log\left(v+0.75\right)-2log\left(s\right)+c\_{1}$. Incorporate the censored observations appropriately. Give a summary of the parameters. **[ 20 ]**

*HINT* First fit the model with a normal distribution on only the complete data and make sure you get sensible results. You can do the rest of the test and get everything nearly right on that alone. Then switch to *t* and add the prior, and lastly incorporate the censoring only when you are sure it is working without it.

**1.3)** For the first missing day (Day 4) obtain a reasonable approximation of the posterior predictive distribution given the model and other data. Give a histogram or density plot of some kind to illustrate it. **[ 7 ]**

*HINT* For all plots limit the range of sugar values to lie from 0 to 15.

**1.4)** For each day (1 to 28), obtain a reasonable approximation of the posterior predictive distribution given the model and data. Based on these distributions calculate **median values and highest posterior density intervals for each day**. Produce a plot on the original scale showing these expected values and intervals as curves, as well as the observed data points. **[ 11 ]**

**1.5)** Give an assessment of your model fit, including at least: whether the observed coverage is roughly in line with the nominal level and whether the fit seems reasonable. **[ 4 ]**

**1.6)** What is the probability that the sugar level will be below detection limit on Day 30? **[ 2 ]**

**1.7)** Explain why it would be wrong to extrapolate your fit to the left (Day -1, -2, etc.) specifically. **[ 2 ]**

**1.8)** For up to 2 bonus marks, illustrate the censored observations on your final plot in a sensible way.