SIO 223
Problem Set 4
due 3/18/2020

1. Two objects of unknown weights w_1 and w_2 are weighed in an error-prone pan balance in the following way: (1) object 1 is weighed by itself and the measurement is 3g; (2) object 2 is weighed by itself and the result is 3g; (3) the difference of the weights (weight of object 1 minus the weight of object 2) is measured by placing the objects in different pans, and the result is 1g; (4)the sum of the weights is measured as 7g. The problem is to estimate the true weights of the objects from these measurements.

(a) Set up a linear model $Y = X\beta + e$
(b) Find the LS estimates of w_1 and w_2
(c) Find the estimate of σ^2
(d) Find the estimated standard errors of the LS estimates of part (b)
(e) Estimate $w_1 - w_2$ and its standard error.
(f) Test the null hypothesis $H_0 : \ w_1 \ = \ w_2$

2. The Arrhenius relationship for thermally activated semiconduction in minerals is $\sigma(t) = \sigma_0 e^{-A/kt}$ where $\sigma(t)$ is the electrical conductivity at temperature t, k is Boltzmann's constant and A is the activation energy. This has been used to model the electrical conductivity data for the mineral olivine as a function of temperature. Olivine is a major constituent of Earth's mantle.

•Download and plot the data for the Jackson County dunite in jcd.dat. The file contains a header line followed by 68 data in 2 columns, temperature ($^\circ C$) and conductivity (S/m).

•Explain how to perform an appropriately parametrized linear least squares fit to find the activation energy A and the constant σ_0.

•Using Matlab (or otherwise if you prefer), and omitting the first 14 data in the file, find estimates for σ and A.

•If the errors in the conductivity measurements are 5%, what are the uncertainties in your derived parameters? Use an appropriate analysis of residuals to make a suitable argument as to whether 5% seems like a reasonable estimate of the data uncertainty.

•Redo the analysis with the entire data set. Comment on the differences. Could you modify your modelling approach to accommodate the additional data? If so how?

3. In https://igppweb.ucsd.edu/∼cathy/Classes/SIO223A/faithful.txt you will find data on 272 eruptions events of Yellowstone National Park's Old Faithful geyser. Columns in the file are eruption number, eruption length (in minutes), waiting time to the next eruption (in minutes). Using the techniques we have discussed in this class conduct an exploratory analysis of these data. I offer a couple of preliminary suggestions below, but these are far from comprehensive and I look forward to seeing your more detailed ideas.

•Using the methods described in Chapter 9 of the notes and a kernel of your choice find a kernel density estimate for the distribution of inter-event times. Repeat using a different kernel function and /or bandwidth to show that your results are not strongly dependent on the specific form of kernel used.

•Is the time between the N-th eruption and the previous one correlated with the time between the

previous one and the one before that? Test this formally using the usual methods for testing correlation, and comment on how valid this might be after making a scatterplot.

•What, if any, rule for predicting eruptions is suggested by these data and your analyses? Does this provide any potential physical insight into how eruptions occur?