Homework 5

ISyE 6420
Fall 2019

Due November 10, 2019, 11:55pm. HW5 is not time limited except the due date. Late submissions will not be accepted.

Use of all available electronic and printed resources is allowed except direct communication that violates Georgia Tech Academic Integrity Rules.

1. Cross-validating a Bayesian Regression. In this exercise covariates x_1 and x_2 are simulated as[1]

```
x1 = rand(1, 40);  x2 = floor(10 * rand(1,40)) + 1;
```

and the response variable y is obtained as

```
y = 2 + 6 * x1 - 0.5 * x2 + 0.8*randn(size(x1));
```

Write a WinBUGS program that takes 20 triples (x_1, x_2, y) to train the linear regression model $\hat{y} = b_0 + b_1 x_1 + b_2 x_2$ and then uses the remaining 20 triples to evaluate the model by comparing the original responses $y_i,\ i = 21, \ldots, 40$, with regression-predicted values $\hat{y}_i,\ i = 21, \ldots, 40$. The comparison would involve calculating the MSE, the mean of $(y_i - \hat{y}_i)^2,\ i = 21, \ldots, 40$.

This is an example of how a cross-validation methodology is often employed to assess statistical models.

How do the Bayesian estimators of β_0, β_1, β_2, and σ compare to the "true" values 2, 6, -0.5, and 0.8?

2. Body Fat from Linear Regression. Excess adiposity is a risk factor for a range of diseases, leading to increased morbidity and mortality. Body fat (BF) can be measured

[1] Or in Python or R:

```
import numpy as np
x1 = np.random.uniform(0,1, 40)
x2 = np.floor(10 * np.random.uniform(0,1,40)) + 1
y = 2 + 6 * x1 - 0.5 * x2 + 0.8*np.random.normal(0,1,len(x1))
====
x1 <- runif(40)
x2 <- floor(10 * runif(40)) + 1
y <- 2 + 6 * x1 - 0.5 * x2 + 0.8*rnorm(length(x1))
```

by several techniques such as skin-fold measurements bioelectrical impedance analysis (BIA) and dual-energy X-ray absorptiometry (DEXA). Most of these techniques are not used in the clinical practice or they are not adequate when large populations are considered.

Fuster-Para et al. (2015)[2] compare several linear models for predicting the body fat (BF) from Age, Body Mass Index (BMI), Body Adiposity Index (BAI) and Gender.

Data set `RegBF.csv|xlsx` provides data on Age (in years), Body Adiposity Index (BAI), Body Mass Index (BMI), Body Fat (BF), and Gender (0 for males and 1 for females), of 3,200 adults from Mallorca (Spain). To save you some time a starter file `BFReg.odc` is provided.

Percentage of body fat mass was obtained by Tetrapolar Bioelectrical Impedance Analysis (BIA) system (BF-350, Tanita Corp, Tokyo, Japan). The BAI is defined as

$$\frac{\text{hip circumference in cm}}{(\text{height in m})^{1.5}} - 18.$$

We are interested in predicting BF from Age, BAI, BMI, Gender, and BB. BB is a new variable defined as `BB = BAI * BMI`, and as such, describes the interaction between BAI and BMI.

(a) Suggest two models: first with all predictors, and the second with single best predictor. Explain how did you choose the best predictor.

(b) A new person is to be evaluated using the two models from (a). The covariates are: Age = 35, BAI=26, BMI=20, Gender = 0, BB=520. What are the predicted BF's from the two models.

3. Shocks. An experiment was conducted (Peter Lee, 2009; Dalziel et al., 1941) to assess the effect of small electrical currents on farm animals, with the eventual goal of understanding the effects of high-voltage power lines on livestock. The experiment was carried out with seven cows using six shock intensities, 0, 1, 2, 3, 4, and 5 milliamps (shocks on the order of 15 milliamps are painful for many humans). Each cow was given 30 shocks, 5 at each intensity, in random order. The entire experiment was then repeated, so each cow received a total of 60 shocks. For each shock the response, mouth movement, was either present or absent. The data as quoted give the total number of responses, out of 70 trials, at each shock level. We ignore cow differences and differences between blocks (experiments).

[2]Fuster-Parra, P., Bennasar-Veny, M., Tauler, P., Yañez, A., López-González, A. A., and Antoni Aguiló, A. (2015). A comparison between multiple regression models and CUN-BAE equation to predict body fat in adults. PLOS One, `DOI:10.1371/journal.pone.0122291.`

Current (milliamps) x	Number of responses y	Number of trials n	Proportion of responses p
0	0	70	0.000
1	9	70	0.129
2	21	70	0.300
3	47	70	0.671
4	60	70	0.857
5	63	70	0.900

Using logistic regression and noninformative priors on its parameters, estimate the proportion of responses after a shock of 2.5 milliamps. Find 95% credible set for the population proportion.