FINAL EXAM

ISyE6420

Released XXXXXX, 12:00pm – due XXXXXX, 11:55pm. This exam is not proctored and 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.

Name _____

Problem	Time to Second Birth	Tasmanian Clouds	Miller Lumber Co	Total
Score	/33	/33	/34	/100

1. Time to Second Birth. The Medical Birth Registry of Norway was established in 1967 and contains information on all births in Norway since that time.[1]

The data set (given in a starter odc file `secondbirth0.odc`) provides the time between first and second births for a selection of 16,341 Norwegian women. This time could be possibly influenced by the age of mother and the fact that the firstborn child died within one year of its birth.

The data set contains the following variables (as columns)

`mage`	age of mother at first birth (in years)
`death`	first child died within one year ($0 =$ no, $1 =$ yes)
`time`	time from first birth to second birth (in days)

(a) Using WinBUGS establish a regression model with variables `mage` and `death` as covariates. What is the 95% CS for the slope β_2 corresponding to variable `death`? Is this variable significant?

(b) By analyzing β_1, the coefficient of covariate `mage`, argue that age of mother is not significant factor in influencing the response `time`.

(c) Helga is a mother with two children. The children are healthy and growing. Helga was 24 when her first child was born. What is the predicted time between the births according to your model.

(d) Emma lost her firstborn child at its birth when she was 28. She gave a birth to the second child. What is the 95% CS for the predicted time between the births according to your model.

Hint: Be careful: mean time and predicted time are not identical. Since sample size is large, limit your MCMC to 10,000 max after the burn in.

Tasmanian Clouds. The data `clouds.csv|dat|xlsx` provided by OzDASL were collected in a cloud-seeding experiment in Tasmania between mid-1964 and January 1971. Analysis of these data is discussed in Miller et al. (1979).

Figure 1: T-Rex Cloud Stomping over Tasmania

The rainfalls for target and control areas are given in inches. Variables `TE` and `TW` are the east and west target areas, respectively, while `CN`, `CS`, and `CNW` are the corresponding rainfalls in the north, south, and northwest control areas, respectively. `S` stands for seeded and `U` for unseeded. Variables `C` and `T` are averages of control and target rainfalls. Variable `DIFF` is the difference `T - C`.

(a) Provide a comprehensive Bayesian additive two-way ANOVA analysis on the response `DIFF` to estimate and test the effects of factors `Season` and `Seeded`.

(b) Repeat the analysis from (a) after adding the interaction term.

Hint: Consult the `Simvastatin` example. You will need only three variables form the data: Season, Seeded, and DIFF. Recode the factor levels in Season and Seeded as numbers.

Miller Lumber Company Customer Survey. Kutner et al. (2005)[2] analyze a data set from a survey of customers of the Miller Lumber Company. The response is the total number of customers (in a representative 2-week period) coming from a census tract of a metropolitan area within 10 miles from the store.

Figure 2: Miller Lumber Company

The covariates include five variables concerning the census tracts: number of housing units, average income in dollars, average housing unit age in years, distance to nearest competitor in miles, and distance to store in miles. Fit and assess a Poisson regression model for the number of customers as predicted by the covariates. The data are in odc starter file `lumber0.odc` and the variable are `customers`, `hunits`, `aveinc`, `aveage`, `distcomp`, and `diststore`.

(a) Propose Poisson model with `customers` as response variable and `hunits`, `aveinc`, `aveage`, `distcomp`, and `diststore`, as covariates. Use noninformative priors on regression coefficients.

(b) If you are to propose a Poisson model with only two covariates which two you will chose? Justify your choice.

(c) Miller Lumber Company is opening a new store in an area for which the covariates are `hunits=720`, `aveinc=70000`, `aveage=6`, `distcomp=4.1`, and `diststore=8`. Find mean response and prediction with 95% for number of customers in a representative 2-week period

---

[2]Kutber, M. H., Nachtsheim, C. J., and Neter, J. (2005). Applied Linear Regression Models- 5th Edition, McGraw Hill/Irwin Series: Operations and Decision Sciences. Miller Lumber Company Example, p. 621.