

## Homework 5

ISyE 6420

Fall 2022

**1. Paddy Soil Adhesion.** Pan and Lu (1998) provide measurements of adhesion on 43 pairs of samples of paddy soil to steel and rubber. From 1974 to 1983, during the rice-growing season, the adhesion of soils to steel and to rubber were measured in situ simultaneously in paddy fields in South China. As steel and rubber have long been the most important materials used for wetland running gears such as wheel and track, it is expected that the adhesion to them would be roughly the same. The adhesion was measured with an adhesometer. Data set `paddy.dat` has two columns: (1) adhesion to steel, and (2) adhesion to rubber. Both measurements are given in kPa.

(a) Fit the linear regression model where the response variable  $y$  is adhesion to rubber. Report the parameter estimates and Bayesian  $R^2$  that are obtained via the summary output.

(b) If the adhesions to steel and rubber are comparable, the regression should have  $\beta_0 = 0$  and  $\beta_1 = 1$  as population parameters. Investigate each claim separately using Bayesian methods and report your findings (hint: step function will be useful in BUGS).

(c) Use the model to predict the adhesion to rubber for which adhesion to steel was 2. Find the 95% credible set for a single predictive response.

**2. Third-degree Burns.** The data for this exercise, discussed in Fan et al. (1995), refer to  $n = 435$  adults who were treated for third-degree burns by the University of Southern California General Hospital Burn Center. The patients were grouped according to the area of third-degree burns on the body. For each midpoint of the groupings  $\log(\text{area} + 1)$ , the number of patients in the corresponding group who survived and the number who died from the burns was recorded:

$\log(\text{area} + 1)$	<i>Survived</i>	<i>Died</i>
1.35	13	0
1.60	19	0
1.75	67	2
1.85	45	5
1.95	71	8
2.05	50	20
2.15	35	31
2.25	7	49
2.35	1	12

(a) Fit the logistic regression on the probability the covariate  $x = \log(\text{area} + 1)$  and report the model output. What is the deviance?

(b) Using your model, find the posterior probability of survival for a person for which  $\log(\text{area} + 1)$  equals 2 .

**3. SO<sub>2</sub>, NO<sub>2</sub>, and Hospital Admissions.** Fan and Chen (1999) discuss a public health data set consisting of daily measurements of pollutants and other environmental factors in Hong Kong between January 1, 1994 and December 31, 1995. The association between levels of pollutants and the number of daily hospital admissions for circulation and respiratory problems is of particular interest.

The data file `hospitaladmissions.dat` consists of six columns:

1. year
2. month
3. day in month
4. concentration of sulfur dioxide SO<sub>2</sub>
5. concentration of pollutant nitrogen NO<sub>2</sub>
6. daily number of hospital admissions

In this problem, only the last three columns are needed. Use noninformative priors on all parameters in your model.

(a) Fit a Bayesian Poisson regression model and report the summary of the model output. Explain how the expected number of hospital admissions varies with the levels of SO<sub>2</sub> and NO<sub>2</sub>.

(b) Suppose that on a particular day the levels of SO<sub>2</sub> and NO<sub>2</sub> were measured as 44 and 100, respectively. Estimate the expected number of hospital admissions and report 95% credible set.