UNIVERSITY OF NEW SOUTH WALES
DEPARTMENT OF STATISTICS

MATH5856 Introduction to Statistics and Statistical Computations
Tutorial Problems week 13

1. An article in the *Journal of General Psychology* (1973), pp. 91–95, discusses the relationship between an individual's flicker frequency (the highest frequency for a flickering light source at which the flickering can be detected) and eye colour. For 19 individuals, eye colour and flicker frequency were recorded: Data can be downloaded from Blackboard,

Eye Colour	Flicker Frequency	Eye Colour	Flicker Frequency
Brown	26.80	Blue	25.70
Brown	27.90	Blue	27.20
Brown	23.70	Blue	29.90
Brown	25.00	Blue	28.50
Brown	26.30	Blue	29.40
Brown	24.80	Blue	28.30
Brown	25.70		
Brown	24.50		
Green	26.40		
Green	24.20		
Green	28.00		
Green	26.90		
Green	29.10		

   under `course materials > data sets > flicker.txt`. Save the file (a text file) to your home directory. You can then open the file using the `read.table` command.

   Analyse the data as a one factor experiment. In particular, produce

   - A graph showing the overall mean for the data and the means for the different eye colour groups
   - A graph showing the overall median for the data and the medians for the different eye colour groups

- Boxplots for the flicker frequencies for the eye colour groups
- A summary for a suitable fitted linear model to detect differences between eye colour groups.

Comment on the analysis of variance table for your fitted linear model, and give estimates of the effects for the eye colour groups α_i subject to the constraint $\sum_i \alpha_i = 0$.

2. Three speed reading courses were being evaluated (X, Y and Z). Fifteen male subjects and fifteen female subjects were randomly assigned to each of the three reading courses (five males and five females to each course). The data obtained from this experiment are shown in the table. Enter the data in R, and summarize the data with appropriate

	Group X	Group Y	Group Z
Male	700	480	500
	850	460	550
	820	500	480
	640	570	600
	920	580	610
Female	900	590	520
	880	540	660
	899	560	525
	780	570	610
	899	555	645

plots. Fit a linear model to determine if the three courses are equally effective. Is there an interaction between the reading course and gender? (Hint: first generate an interaction plot, then fit a two factor model with interaction and use the `summary` command to generate the analysis of variance table).

3. The `ethanol` data frame contains 88 measurements generated in an experiment in which ethanol was burned in a single cylinder automobile test engine. The variables are `NOx` (a measure of the nitrogen oxides in the exhaust), `C` (compression ratio of an engine) and `E` (a measure of the richness of the air/ethanol mix used).

Fit a linear model with `NOx` as the response and `E` and `C` as predictors. Examine the F-statistic for testing whether the coefficients for all predictors are zero and the partial t-statistics. Does it seem as though a linear model is helpful for explaining variation in the response? Also, plot appropriate diagnostic quantities for the linear model.

4. In deriving the conditional posterior distribution of β given σ^2 in the Bayesian linear model we used the following result.

   Lemma:

   Let Z be a q-dimensional random vector with density $p(z)$ with

   $$p(z) \propto \exp\left(-\frac{1}{2}\left(z^T A z - 2b^T z\right)\right)$$

   where A is a fixed symmetric positive definite $q \times q$ matrix and b is a fixed $q \times 1$ vector. Then $Z \sim N(A^{-1}b, A^{-1})$.

   Prove this result.

5. In deriving the marginal posterior distribution for σ^2, $p(\sigma^2|y)$ in the Bayesian linear model we used the following result.

   Lemma:

   Let z, b be $q \times 1$ vectors, and A be a symmetric positive definite $q \times q$ matrix. Then

   $$\int \exp\left(-\frac{1}{2}\left(z^T A z - 2b^T z\right)\right) dz = (2\pi)^{q/2}|A|^{1/2} \exp\left(\frac{1}{2} b^T A^{-1} b\right)$$

   where $|A|$ denotes the determinant of A.

   Prove this result

6. Consider one of the standard R data sets, the USArrests data (type `data(USArrests)` to load the data and then `help(USArrests)` to read

about the variables in the USArrests data frame). For this data set, fit a Bayesian linear regression model with the response variable `Assault` and the predictor variable UrbanPop. Use the standard Uniform prior on the parameters β and σ^2, briefly describe how simulation can be carried out in R. In your description, refer to the relevant results given in the lectures.

Using the QR decomposition to help speed up the computation, fit the above model in R. Does it seem as though the predictor is related to the response? Refer to an appropriate 95% credible interval to support your answer.