UNIVERSITY OF NEW SOUTH WALES
DEPARTMENT OF STATISTICS

MATH5856 Introduction to Statistics and Statistical Computations
Tutorial Problems week 12 (Bayesian inference and Linear models)

1. In each of the following cases, derive the posterior distribution:

   (a) $x_1, \ldots, x_n$ are a random sample from the distribution with probability function

   $$p(x|\theta) = \theta^{x-1}(1-\theta); \quad 1, 2, \ldots$$

   with the Beta(p, q) prior distribution

   $$p(\theta) = \frac{\theta^{p-1}(1-\theta)^{q-1}}{Be(p, q)}, \quad 0 < \theta < 1$$

   (b) $x_1, \ldots, x_n$ are a random sample from the distribution with probability density function

   $$p(x|\theta) = \frac{e^{-\theta}\theta^x}{x!}, \quad x = 0, 1, \ldots$$

   with the prior distribution

   $$p(\theta) = e^{-\theta}, \quad 0 < \theta$$

2. The proportion, θ, of defective items in a large shipment is known, but expert assessment assigns θ the Beta(2,200) prior distribution. If 100 items are selected at random from the shipment, and 3 are found to be defective, what is the posterior distribution of θ?

   If another statistician, having observed the 3 defectives, calculated her posterior distribution as being a beta distribution with mean 4/102 and variance 0.0003658, then what prior distribution had she used?

3. The diameter of a component from a long production run varies according to a $N(\theta, 1)$ distribution. An Engineer specifies that the prior distribution for θ is $N(10, 0.25)$. In one production run 12 components are sampled and found to have a sample mean diameter of 31/3. Use this information to calculate the probability that the mean component diameter is at least 10 units.

4. The table below shows the number of deaths caused by firearms in Australia over a number of years expressed as a rate per $100,000$ of population.

Year	Rate	Year	Rate
1983	4.31	1991	3.67
1984	4.42	1992	3.61
1985	4.52	1993	2.98
1986	4.35	1994	2.95
1987	4.39	1995	2.72
1988	4.21	1996	2.96
1989	3.40	1997	2.30
1990	3.61		

(a) Enter the data set in a data frame with columns `Year` and `Rate`.

(b) Do a scatter plot of `Rate` against `Year`. Do you think a simple linear regression model is appropriate?

(c) Fit a regression model with only the intercept term in.

(d) Fit a regression model with `Rate` as a regressor, but no intercept.

(e) Fit the simple linear regression model with both intercept and `Rate` as a regressor. From the `summary` output for the fitted model, state whether you think there is strong evidence for a linear trend in death rates over time.

(f) By using the `plot` command, produce residual and diagnostic plots for the fitted model. What do the plots tell you?

5. The heights and weights of 7 people are given in the table below. Fit

Height	Weight
68	155
61	99
63	115
70	205
69	170
65	125
72	220

a simple linear regression model to predict height from weight. Give a prediction of height when weight is 100, and give 95% confidence limits for this prediction. Also, compute the residuals and plot them against the fitted values.

6. Obtain the `cheese` data from Blackboard. This dataset contains the variables `taste`, `acetic`, `H2S` and `lactic`.

   (a) The command `pairs` allows you to make pairwise plots of all the variables in a data frame, using this command for the cheese data, do you think it would be reasonable to fit linear regression models with `taste` as response, using some or all of the predictors?

   (b) Fit a model with `taste` and response with all the predictors, do you think any of the predictors should be included in the model? use an appropriate statistical test and write down the corresponding hypotheses.

   (c) Using an appropriate hypothesis test, sequentially select the least important predictors to exclude from the full model, which model is the most parsimonious?

7. Prove the following vector expectations

   (a) If a is a $k \times 1$ vector of constants then $E(a) = a$.

   (b) If a is a $k \times 1$ vector of constants, and Y is a $k \times 1$ random vector with $E(Y) = \mu$, then $E(a^T Y) = a^T \mu$.

(c) If A is an $n \times k$ matrix, and Y is a $k \times 1$ random vector with $E(Y) = \mu$, then $E(AY) = A\mu$.

8. (a) Let Y be a $k \times 1$ random vector with $Var(Y) = V$. If a is a $k \times 1$ vector of real numbers, then

$$Var(a^T Y) = a^T V a.$$

(b) Let Y be a $k \times 1$ random vector with $Var(Y) = V$. Let A be a $k \times k$ matrix. If $Z = AY$, then

$$Var(Z) = AVA^T.$$