

Course EPIB-621 - Data Analysis for the Health Sciences

Assignment 3

1. Consider the data set called `assign3num1.txt` on the course web page.

(a) Run a linear regression of y on x_1 and x_2 . Report the usual summary from R, including the R^2 value.

(b) Plot a histogram of residuals, and report any outliers. Identify the outlier by plotting x_1 versus the residuals.

(c) Delete that outlier from the data set, and rerun the linear regression. By how much do the parameter estimates and their confidence intervals change? How much does the R^2 value change by?

2. The FIM is a measure of functional independence, often used in the emergency room following an accident or trauma. Higher values of the FIM indicate better functioning, with 126 being considered as “normal”. In this question we will consider the effect of alcohol on the FIM. This is an interesting question, since it is not clear whether alcohol will have a positive (e.g., more relaxed during an accident may lead to less injuries) or negative (e.g., worse accidents if under the influence) effect on the FIM, on average.

(a) The course web site has a data set called `fimdata.txt`. Download this file somewhere to your hard disk. Open R and read in the data set (using a command such as

```
fim.dat <-read.table(file, header=T)
```

where `file` points to where you stored the data set, such as “`c://temp//fimdata.txt`”. The variables include age in years, sex (1=male, 0=female), fim, and alcohol level on arrival at the emergency room. Print out some descriptive statistics for these variables, for example using the R command `summary(fim.dat)`. Print out plots of age versus fim and alcohol versus fim. Do you see any trends?

(b) To use the data set in WinBUGS, you will need to add `[]`'s after each variable name. Open WinBUGS, and cut and paste in the data set you saved. Change the first line to

```
sex[] age[] fim[] alcohol[]
```

Run a simple linear regression in WinBUGS, with `fim` as the dependent variable, and `age`, `sex`, and `alcohol` level. Report your results, and provide an interpretation for each beta coefficient.

(c) It is easy to make predictions for any variable combination using WinBUGS, simply by adding a single line to the program for each prediction you want to make. For example, if you want to make a `fim` prediction for a male aged 50 with an alcohol level of 80, add a line like:

```
pred.fim.male.50.80 <- alpha + beta.sex*1 + beta.age*50 +
beta.alc*80
```

and then monitor the new parameter you created, `pred.fim.male.50.80`. Create predictions for males aged 30 and alcohol level of 85 and for females aged 50 with alcohol level of zero. Create another variable that monitors the difference between these two groups, with a command such as

```
diff <- pred.fim.male.30.85 - pred.fim.female.50.0
```

Run a program with these additional lines, report the outcomes, and comment of the difference between the two sets of predictions you made.

3. In this question we will use the `bicreg` program to examine various models relating to predicting IQ.

(a) Obtain the data file “`brain.data.txt`” from the course web page. Variables used are:

CCMIDSA:	Corpus Collasum Surface Area (cm ²)
IQ:	Intelligence Quotient
HC:	Head Circumference (cm)
ORDER:	Birth Order
SEX:	Sex (1=Male 0=Female)
TOTSA:	Total Surface Area (cm ²)
TOTVOL:	Total Brain Volume (cm ³)
WEIGHT:	Body Weight (kg)

Read the data set into R (for example, using a “`read.table`” command, note that a header is present, so use “`header=T`”), and use the “`pairs`” command to produce a

pairwise plot. Comment on which variables seem to be related or not, just looking at the graphs alone.

(b) Use the bicreg program to perform model selection on this data set. Provide the output in your answer.

(c) Look at the best model. How much better is it than the second best model?

4. Average blood mercury levels vary from region to region, because of different degrees of fish consumption, different origins for available fish, and other factors. The data set called fish.txt contains the mercury concentrations (in μ grams per liter) for 20 randomly selected persons from each of five different regions. Assume that within each region, mercury levels vary randomly around the region specific mean, and that the region specific means also vary randomly around an overall mean value. Run a random effects model in R, and report the parameter estimates for both region specific and overall mercury levels.

5. Repeat question 4, but use a Bayesian hierarchical model. Provide your program in WinBUGS, and monitor and report on all unknown parameters from the model.