

Department of Statistics

COURSE STATS 330

Assignment 4, 2008

Instructions: Hand in your completed assignment to the Student Resource Centre by 4pm on **Monday October 6.**

The data for this assignment arose from an agricultural experiment to investigate the factors that affect the level of Vitamin C in cabbages. The data are in a data frame `cabbages` that is included in R, in the library MASS. To access the data frame, type

```
library(MASS)
data(cabbages)
```

This data frame contains the following columns:

<code>Cult</code>	Factor giving the cultivar of the cabbage, two levels: <code>c39</code> and <code>c52</code> .
<code>Date</code>	Factor specifying one of three planting dates: <code>d16</code> , <code>d20</code> or <code>d21</code> .
<code>HeadWt</code>	Weight of the cabbage head, presumably in kg.
<code>VitC</code>	Vitamin C content, in undefined units.

Your task in this assignment is to assess the effect of `Cult`, `Date` and `HeadWt` on the vitamin content. Specifically,

1. Draw a trellis plot and use it to assess graphically which of the factors are related to the response. Report your conclusions.[10 marks]
2. Fit a suitable model to the data. Are there significant interactions between the explanatory variables? [10 marks]
3. Using the model you fitted in (2), and interpreting the coefficients in the regression summary, discuss how the explanatory variables affect the response. [10 marks]
4. For the model you fitted in 2, calculate a confidence interval for the mean vitamin C content for each combination of `Date` and `Cult`, fixing the value of `HeadWt` at the mean head weight for the corresponding cell. [10 marks] Hint: use the `predict` function. [10 marks]

PTO

Question for STATS 763 only.

Write an R program to evaluate the log-likelihood for the CHD data in lecture 20, over a grid of α and β values. Use 50 α -values equally spaced from -5.4, to -5.2, and 50 β -values equally spaced from 0.105 to 0.115. Store the result as a matrix with the rows corresponding to the α values and the columns corresponding to the β values, and the matrix entries corresponding to the values of the log-likelihood.

Then plot the log-likelihood as a contour plot. To get a good plot, you may need to manipulate the contour levels (use the levels argument).

Mark the maximum value by eye on the plot, and then verify that the estimates obtained using glm correspond to these maximum values.

R hint: You can evaluate the log-likelihood using the expression

```
sum( (a+b*chd.df$age) * chd.df$chd - log(1+exp(a+b*chd.df$age)) )
```

The formula for the log-likelihood is given in lecture 21.