RESEARCH SCHOOL OF
FINANCE, ACTUARIAL STUDIES AND STATISTICS
College of Business & Economics, The Australian National University

GENERALISED LINEAR MODELS

(STAT3015/STAT7030)

Assignment 1 for 2015

Instructions

- This assignment is worth 15% of your overall marks for this course (for all students, enrolled in either STAT3015 or STAT7030). If you wish, you may work together with another student in doing the analyses and present a single (joint) report. If you choose to do this then both of you will be awarded the same total mark. A STAT3015 student may work with a STAT7030 student. You may NOT work in groups of more than two students and the usual ANU examination rules on plagiarism still apply with respect to people not in your group.
- Research School of Finance, Actuarial Studies and Statistics (RSFAS) assignment cover sheets are available on Wattle. Please complete and attach a copy of the cover sheet to the front of your assignment. Remember to keep a copy of your assignment.
- Assignments should be written, typed or printed on sheets of A4 paper stapled together at the top left-hand corner (do not submit the assignment in plastic covers or envelopes). Your assignment may include some carefully edited computer output (e.g. graphs) showing the results of your data analysis and a discussion of those results. Please be selective about what you present – only include as many pages and as much computer output as necessary to justify your solution and be concise in your discussion of the results. Clearly label each part of your report with the question number and the part of the question that it refers to. This semester, we will be trialling online submission – if you wish to use that option, follow the instructions on Wattle.
- Unless otherwise advised, use a significance level of 5%.
- Marks may be deducted if these instructions are not strictly adhered to, and marks will certainly be deducted if the total report is of an unreasonable length, i.e. more than 10 pages including graphs. You may include, as an appendix, any *R* commands you used to produce your computer output. This appendix and the cover sheet are in addition to the above page limits; but the appendix will generally not be marked, only checked if there is some question about what you have actually done.
- This assignment will be marked by the course tutor, Xu Shi. Assignments should be submitted either online via Wattle or in the assignment box for this course located next to the RSFAS office by **4 pm Friday 4 September 2015**. You may ask your tutor or me (the lecturer, Ian McDermid) questions about this assignment, in person, up to the deadline (4 pm on Friday 4 September 2015), after which we will NOT answer any further questions about this assignment, until after the marked assignments have been returned to students. Answers to questions in writing posted on Wattle or sent to me via e-mail will be posted on Wattle, but must be received no later than 4 pm Wednesday 2 September 2015.
- Late assignments will NOT be accepted after the deadline without an extension. Extensions will usually be granted on medical or compassionate grounds on production of appropriate evidence, but must have the permission of both your tutor and me by no later than 4 pm Wednesday 2 September 2015. Even with an extension, all assignments must be submitted reasonably close to the above deadline to allow time for your tutor to mark all assignments in time for the tutorials in week 8 (first of which is 9 am on Wednesday 23 September 2015), when the assignment solutions will be released and discussed.

Question 1 **(30 marks)**

The data in the file **ClothDye.txt** (available on Wattle) are taken from problem 6.17 on pages 286-287 of the text by Montgomery, Douglas C. (1997) *Design and Analysis of Experiments*, 4th Edn, Wiley, New York. Montgomery describes the experiment that generated these data as follows: "The quality control department of a fabric finishing plant is studying the effect of several factors on the dyeing of cotton-synthetic cloth used to manufacture men's shirts. Three operators, three cycle times and two temperatures were selected, and three small specimens of cloth were dyed under each set of conditions. The finished cloth was compared to a standard, and a numerical score was assigned."

(a) The response variable for this experiment is the score assigned to each specimen of dyed cloth that measures the quality of the result (presumably better scores indicate better quality). Describe the possible explanatory factors in this experiment and how many levels there are for each factor. How many experimental treatments are there and how many replications of each treatment? **(3 marks)**

(b) Assuming that the experimenters are primarily interested in how the scores vary in response to the combinations of cycle times (presumably measured in minutes) and temperature (which is probably in degrees Fahrenheit); read the data into *R* and present plots that illustrate the mean scores for the combinations of these two factors. **(3 marks)**

(c) Assuming that the experimenters do want to control for differences between the three operators, but are not interested in any interactions other than a possible interaction between cycle time and temperature, choose an appropriate linear model. In fitting this model, treat the experimental factors as nominal rather than continuous variables and read part (d) below, before deciding how to parameterise the model. Give a complete algebraic description of the model, the constraints and assumptions used. Then use *R* to fit the model and present residual and other diagnostic plots (no more than 3 plots) and discuss whether or not these plots suggest any problems with the underlying assumptions. **(6 marks)**

(d) Select and present some summary output for your model in part (c). Ignoring for the moment any problems with the fit of your model, what does the model suggest about the relationship between the scores and the combinations of cycle time and temperature? Which combinations produce results that are significantly different from the overall mean score? **(7 marks)**

(e) Fit other models to investigate possible interactions between the operators and the other experimental factors. Select one of these models and present and discuss the ANOVA table. Do not present the residual plots for your chosen model, but discuss any differences between these plots and the ones you produced in part (c). **(4 marks)**

(f) In the model described in part (c), one of the experimental factors was effectively used as a blocking variable, but treated as a fixed effect. Refit the model in part (c) as a mixed effects model, treating this blocking factor as a random effect. Describe the changes this makes to the algebraic description of the model. **(3 marks)**

(g) Examine the summary output for the model in part (f). Aside from the treatment of the blocking variable; have there been any changes to the conclusions about the fixed effects in the model? **(2 marks)**

(h) Has blocking been effective in this instance? What proportion of the overall variability is due to variation between blocks? **(2 marks)**