STAT2008/STAT2014/STAT6014

Tutorial 3

Question 1.

The data file **Lubricant.csv** (available on Wattle) contains 53 measurements of the viscosity of a particular lubricating agent at various temperatures and pressures. The names of the three variables in the data are **viscos, pressure** and **tempC**. In tutorial 2, we have already used this data to fit a simple linear regression with viscosity as the response and pressure as the predictor variable. Based on this fitted model, please answer the following questions.

(a) Use **anova()** to produce the Analysis of Variance table for the regression. What is the MSE for this regression? How many degrees of freedom are associated with this MSE? Have a look at the F statistic and p-values in the ANOVA table. What is your conclusion? Please also manually compute the anova table and also look at the related results produced by the **summary()** function.

(b) What are the standard errors of the slope and intercept? Manually check the standard error of the slope coefficient by calculating SSE and S_{xx}.

(c) We are interested in testing the significance of the slope coefficient. Give full details of a t-test on the estimated slope coefficient. Find the p-value for the test statistic. Are they consistent with the F-test that you obtained in tutorial 2?

(d) We are interested in testing if the slope coefficient is 0.001. Find the p-value associated with the test.

(e) Now for a more advanced question which will take a lot of R coding. Again plot viscosity against pressure, but this time use a different plotting symbol to indicate which value of temperature is associated with each data point. [Hint: use type="n" to start with a blank plot and then use the points() function and the pch option to add the points for the different levels of tempC ? you will probably need to do some searching through the help files associated with the graphical parameters help(par).] What do you notice from your plot? Fit separate simple linear regression lines for each level of temperature and add these to your plot (and possibly include a suitable legend). Do the slopes of these separate models appear to be the same or different?

Question 2. Ian has been having problems with hearing for over 20 years, but has only been wearing hearing aids for the last decade. A couple of years ago, he purchased some brand new hearing aids and decided to also complete an on-line program designed to improve his listening skills (at the suggestion of his audiologist, who arranged free access to the program). He has only recently managed to finally complete the 11 daily sessions of the full program. Among the 11 sessions, he did the first few days of the course at the rate of about once a week, but then had almost a 6 month gap before he completed the later days at the rate of almost one session a day. Here are his daily cumulative scores:

Day	Cumulative Score
0	0
1	195
2	351
3	503
4	683
5	847
6	1011
7	1193
8	1378
9	1561
10	1743
11	1925

Information provided about interpreting these scores is presented in the file **LACE Results.pdf**, which is available on Wattle.

Can he reasonably conclude that he has done significantly better in the program than someone in the typical range ("Many people training with LACE get between 60 and 120 points per day")? Analyse the above data in R and conduct a test of some appropriately chosen hypotheses.