MATH3012 – Statistical Methods II
S-Plus Worksheet 6 – More Logistic regression

1. **Obtain the source**. Fire up the internet explorer. Go to the course webpage `http://www.maths.soton.ac.uk/staff/Sahu/teach/math3012` or otherwise. Click on the source file for worksheet 5.

2. **Import the data**

   (a) To do this you just install the MATH3012 files and click on the `shuttle` icon with the `S` symbol on my computer window.

   (b) This dataset concerns the 23 space shuttle flights before the Challenger disaster. The disaster is thought to have been caused by the failure of a number of O-rings, of which there are six in total. The data consist of four columns, the number y of damaged O-rings for each pre-Challenger flight, together with the launch temperature *temp* in degrees Fahrenheit, the pressure *pressure* at which the pre-launch test of O-ring leakage was carried out and the name of the orbiter (*name*; coded 1 = "Atlantis", 2 = "Challenger", 3 = "Columbia", 4 = "Discovery").

   (c) The Challenger launch temperature on 20th January 1986 was 31^0F. By fitting generalised linear models to this data, predict the probability of O-ring damage at the Challenger launch. Is the prediction sensitive to the choice of link function? Comment on the validity of this prediction.

3. We need the column n which should be 6 for all rows. We get this by issuing the command `shuttle$n <- rep(6, 23)`.

4. We first find the best model. We issue the following commands.

   (a) `sh.glm1 <- glm(y/n ~ temp + pressure + name, data=shuttle, family=binomial, weights=n)`

   (b) `summary(sh.glm1)`

   (c) `anova(sh.glm1, test="Chisq")`

   (d) `sh.glm2 <- update(sh.glm1, . ~ . - name -pressure)`

   (e) `anova(sh.glm2, test="Chisq")`

5. **Notes:**

   (a) `name` is a factor, `temp` and `pressure` are continuous covariates.

   (b) The first anova command reveals that `name` and `pressure` are not significant predictors.

   (c) We drop those using the update command.

   (d) The last anova command shows that temperature is significant.

6. To see that the model with only temperature is not a bad fit we work with the residual deviance. Using the following commands we see where the observed value of the deviance is in the theoretical χ^2 distribution.

(a) `x <- seq(10, 50, length=200)`

(b) `plot(x, dchisq(x, 21), type="l")`

(c) `abline(v=18.09)`

7. Now we come to do the prediction at temperature $=31^0$F.

   (a) We first get a new data set. We copy the shuttle data and add another row.

   (b) Type `shnew <- shuttle`. Bring up shnew by double clicking from the object browser. On the 24th row type `NA`, 31, `NA`, `NA` and `NA`. This row gives the value 31 to temperature and 'Not Available' to all other variables. We will use this row to predict the probability at temperature$=31^0$F.

   (c) The command `u <- predict(sh.glm2, shnew, se.fit=T, ci.fit=T, type="response")` predicts the probabilities for all 24 rows of data in `shnew`. Type `u` to see. The values for the first 23 rows are the fitted values corresponding to the observed data. The last row gives the predicted value for temperature$=31^0$F.

8. Some graphs:

   (a) In order to see a better picture we compile the function `niceplot`. Just highlight the statements from niceplot to the closing brace. Type `niceplot()` on the commands window. You will see a graph of what we have been doing!

   (b) The programming done to create the plot is, strictly speaking, not necessary for this unit! However, you are encouraged to do this (read, interpret, see, study) as this will increase your level of statistical and computing skills. It will be great if you can use some of these in this unit and later.

9. Note that we have used the logit link so far. Do all these for the probit and complementary log link. This is left as exercise for you.

10. **The bottom line:** *All of these predictions should be treated wih extreme caution, as we are extrapolating far outside the range of the observed data. However, in the presence of such high uncertainty, it may have been unwise to launch at such a low temperature.*