MAST30027: Modern Applied Statistics

Assignment 1, 2019.

Due: 5pm Friday August 23rd

This assignment is worth 12% of your total mark. To get full marks, show your working including the R code you use. Sign a plagiarism declaration form using a link on LMS.

1. Fit a binomial regression model to the O-rings data from the Challenger disaster, using a *probit* link. You must use R (but without using the `glm` function); I want you to work from first principles.

   Your report should include the following:

   (a) parameter estimates;

   (b) 95% CIs for the parameter estimates;

   (c) a likelihood ratio test for the significance of the temperature coefficient;

   (d) an estimate of the probability of damage when the temperature equals 31 Fahrenheit (your estimate should come with a 95% CI, as all good estimates do);

   (e) a plot comparing the fitted probit model to the fitted logit model. To obtain the fitted logit model, you are allowed to use the `glm` function.

2. The R object `assign1.Robj` contains a subset of the `pima` data set. For details of the `pima` data set, please see the practical problem 2 for the week 2.

```
> load("assign1.Robj")
> str(pima_subset)
'data.frame': 532 obs. of  2 variables:
 $ bmi : num  33.6 26.6 28.1 43.1 31 30.5 30.1 25.8 45.8 43.3 ...
 $ test: int  1 0 0 1 1 1 1 1 1 0 ...
```

   We will fit a model with `test` as a response and `bmi` as a predictor to see the relationship between the odds of a patient showing signs of diabetes and his/her bmi. The odds o and probability p are related by

$$o = \frac{p}{1-p} \quad p = \frac{o}{1+o}.$$

   (a) Please estimate the amount of increase in the log(odds) when the bmi increases by 10.

   (b) Give a 95% CI for the estimate.

   You are allowed to use the `glm` function.

3. The gamma distribution with shape $\nu > 0$ and rate $\lambda > 0$ has p.d.f.

$$f(x; \nu, \lambda) = \frac{\lambda^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\lambda x}$$

   for $x > 0$.

   (a) Show that the gamma distribution is an exponential family.

   (b) Obtain the canonical link and the variance function.