GLM/GAM Online Course Exercises: Day 1

Linear Modeling

1) The **R** data frame **`warpbreaks`** gives the number of **`breaks`** per fixed length of wool during weaving, for two different **`wool`** types, and 3 different weaving **`tensions`**. Using a linear model, establish whether there is evidence that the effect of tension on break rate is dependent on the type of wool. If there is, use **`interaction.plot()`** function to examine the nature of the dependence.

2) The **R** data frame **`cars`** contains data about the stopping distance and speed of cars when the driver was signaled to stop. It takes a fixed reaction time for drivers to apply their brakes, so the car will travel a distance directly proportional to its speed before beginning to slow. However, an automobile's kinetic energy is proportional to the square of its speed, but the brakes can only dissipate that energy, and slow the car, at a constant rate per unit distance traveled.

   Fit three different linear models to this data. Report the results.:
   a) dist ~ $\beta_0 + \beta_1(\text{speed}) + \beta_2(\text{speed}^2) + e$
   b) dist ~ $\beta_1(\text{speed}) + \beta_2(\text{speed}^2) + e$
   c) dist ~ $\beta_1(\text{speed}) + e$

   Which model seems to fit better? Why?

   Using the second model in the above list, estimate the average time that it takes a driver to apply the brakes (there are 5280 feet in a mile).

Generalized Linear Modeling

3) The following table shows numbers of occasions when inhibition (i.e., no flow of current across a membrane) occurred within 120 s, for different concentrations of the protein peptide-C. The outcome *yes* implies that inhibition has occurred. Use logistic regression to model the probability of inhibition as a function of protein concentration. Report and plot your results fully. Interpret your results.

conc	0.1	0.5	1	10	20	30	50	70	80	100	150
no	7	1	10	9	2	9	13	1	1	4	3
yes	0	0	3	4	0	6	7	0	0	1	7

4) The **R** data frame **`ACF1`** in the package **`DAAG`** consists of two columns: `count` and `endtime`. The first column contains the counts of simple aberrant foci (ACFs). These are aberrant aggregations of tube-like structures in the rectal end of 22 rat colons after administration of a dose of the carcinogen azoxymethane. Each rat was sacrificed after 6, 12 or 18 weeks. Create a scatterplot of `count` by (~) `endtime`.

Run two **`glm`** models. The first specifies `count` (as the response variable) as predicted by `endtime` (the explanatory variable) and uses a `poisson` family for the distribution. Plot the results. Interpret the results. Then run a second model adding an `endtime^2` term to the right hand side to accommodate a possible quadratic effect. Plot the results. Interpret the results. Compare the two models with an `anova` table. Which model 'fits' better? Why?