

**DUE: February 11, 2016**

1. The yield ( $Y$ ) of a chemical process depends on the temperature ( $X_1$ ) and pressure ( $X_2$ ). The following nonlinear regression model is expected to be applicable:

$$Y_i = \theta_0(X_{i1})^{\theta_1}(X_{i2})^{\theta_2} + \epsilon_i$$

Prior to beginning full-scale production, 18 tests were undertaken to study the process yield for various temperature and pressure combinations. The results (data) are given on the course Blackboard webiste in the file:

**process2.dat**

where the first column is  $Y$ , 2nd column is  $X_1$  and third is  $X_2$ .

- (a) Perform a grid search of possible starting values. The following ranges have been postulated:  $1 \leq \theta_0 \leq 21$ ,  $0.2 \leq \theta_1 \leq 0.8$ ,  $0.1 \leq \theta_2 \leq 0.7$ . For  $\theta_0$  consider 1,6,11,16,21, for  $\theta_1$  consider .2, .3, ..., .8, for  $\theta_2$  consider .1, .2, ..., .7. Evaluate the least squares criterion for each and identify the potential starting values.
- (b) Obtain starting values by ignoring the random error and taking a logarithmic transformation. Use a multiple linear regression to get starting values. Show your work and report your starting values.
- (c) Using the starting values in part (b), obtain least squares estimates of the parameters using each of the following methods: steepest decent, Newton-Raphson, and Gauss-Newton. Are there differences between the estimates and/or the number of iterations to convergence?
- (d) Consider the results from the Gauss-Newton method; report if the parameter estimates are "close-to-linear".
- (e) Using the Gauss-Newton results, plot the residuals against the predicted values,  $X_1$  and  $X_2$ . Also consider a normal probability plot. (Attach all plots; label them and reference them so they are easily found!!). Discuss these plots.
  - i. Test the hypothesis  $H_0 : \theta_1 = \theta_2$  vs.  $H_a : \theta_1 \neq \theta_2$  using  $\alpha = .05$ .
  - ii. Obtain approximate joint confidence intervals for the parameters  $\theta_1$  and  $\theta_2$  using the Bonferroni procedure and a 95% family confidence interval.
  - iii. What do you conclude about the parameters  $\theta_1$  and  $\theta_2$  based on the results in part (i) and (ii)?

2. The following data are from a nutrient requirement experiment on swine. In particular, a dietary supplement (Isoleucine) was considered at 6 different percentage levels. There are four replicates at each level and the response in terms of weight gain (in grams) per day was measured. For this experiment it is expected that a plateau model is appropriate, but it is not clear *a priori* whether the initial weight gain (before the plateau) is linear or quadratic (although, the scientists suspect that the quadratic model is most appropriate). Fit both a linear and quadratic plateau model to these data and suggest which model is most appropriate. Include all of the data but ignore the replication effect in this analysis. Write your answer in a paragraph form, making sure to include information about how you fit the model (e.g., starting values, nonlinear estimation method, etc.). Include a snippet of your SAS code to show how the model was fit.

x (Dietary Isoleucine %)	y (Gain, g/d)
0.38	420
0.38	430
0.38	450
0.38	530
0.42	525
0.42	545
0.42	645
0.42	685
0.46	560
0.46	660
0.46	710
0.46	820
0.50	560
0.50	720
0.50	750
0.50	760
0.54	650
0.54	690
0.54	760
0.54	810
0.58	650
0.58	660
0.58	720
0.58	865

3. Does/response modeling is an important part of many animal health studies. In this setting, the animal's response to various dosages of some compound are measured and a does/response curve is fit. Typically, these curves are somewhat sigmoidal in shape, often modeled by some form of logistic function. The following data are from a study in which interest is in the potency for a current production lot (test) of a veterinary vaccine relative to a reference lot (ref), which has previously been show to demonstrate efficacy. In particular, the following table shows the dilution (dosage), and the response for two replicates. In addition, there are indicators as to whether the sample is the reference (Ref=1 for reference, Ref=0 if not) or the test vaccine (Test = 1 if it is the test vaccine, Test=0 if it is not). We seek to fit the following 3-parameter logistic model:

$$Y_i = \frac{a}{\left(1 + \exp\left(-\frac{(x_i - x_0)}{b}\right)\right)},$$

where  $x_i = \log_2(\text{dilution})$ ,  $a$  - upper limit,  $b$  - rate of change, and  $x_0$  -  $\log_2(\text{midpoint})$ . Note, our chief interest is whether the response is different for the test vaccine versus the reference vaccine, and if so, in what way it is different. Ignore the replication effect in your model (but use all of the data). Again, provide a written answer and make sure you provide detail to justify how you fit the model (e.g., assumptions, etc.). Include a snippet of SAS code to show how your model was fit.

$\log_2(\text{Dilution})$	Ref	Test	Response	
			Rep 1	Rep 2
1	0	1	0.072	0.046
2	0	1	0.092	0.086
3	0	1	0.138	0.122
4	0	1	0.199	0.184
5	0	1	0.267	0.257
6	0	1	0.345	0.344
7	0	1	0.436	0.428
8	0	1	0.518	0.516
1	1	0	0.140	0.071
2	1	0	0.113	0.090
3	1	0	0.141	0.132
4	1	0	0.203	0.194
5	1	0	0.284	0.259
6	1	0	0.364	0.361
7	1	0	0.451	0.450
8	1	0	0.529	0.504

4. The North American Breeding Bird Survey (BBS) is conducted each breeding season by volunteer observers. The observers count the number of various species of birds along specified routes. The collected data are used for several purposes, including the study of the range of bird species, and the variation of the range and abundance over time. For this problem we are interested in the abundance of a particular bird species over time. In particular, we are concerned with the

relative abundance of the House Finch (*Carpodacus mexicanus*). The species is found throughout the U.S. However, it is native to the Western U.S. and Mexico. The Eastern population is a result of a 1940 release of caged birds in New York. The birds were being sold illegally in New York City as “Hollywood Finches” and were released by rare bird merchants in an attempt to avoid prosecution. Within three years there were reports of the birds breeding in the New York area. Because the birds are prolific breeders and their juveniles disperse over long distances, the House Finch quickly expanded to the west. Simultaneously, as the human population on the west coast expanded eastward (and correspondingly, changed the environment) the House Finch expanded eastward as well. By the late 1990’s, the two populations met in the Central Plains of North America.

The file `finch6695.dat` contains two columns of data: (1) the year, and (2) the abundance (total count) of House Finches in the eastern half of the U.S. from 1966 to 1995.

Your task is to find a growth model that fits these data well. Of course, you must document and justify your procedure.