Homework 6

Due: Friday 4/28 at 5pm
See general homework tips and submit your files via the course website.

For all exercises, use the SIDS data set defined in HW6Data.sas file. The data in sids.dat were collected by Spicer et al. (1987) in an investigation of sudden infant death syndrome (SIDS). The two groups here consist of 16 SIDS victims and 49 controls. All the infants have a gestational age of 37 weeks or more and were regarded as full term. The variables are as follows:

- **Group:** 1=control, 2=SIDS
- **HR:** Heart rate, bpm
- **BW:** birthweight, grams
- **Factor 68:** A value arising from spectral analysis of 24 h recordings of electrocardiograms and respiratory movements made on each child
- **Gestational age:** weeks

Exercise 1

a) Perform a discriminant analysis for **group** as a function of all the continuous variables (all variables except group and id). Test whether LDA or QDA is more appropriate. Comment on what the MANOVA tests tell us about possibility to discriminate between infants groups based on these variables.
b) Comment on the cross-validation error results and how well the discrimination matches the groups.

Exercise 2

a) Repeat Exercise 1 using stepwise discrimination. Comment on which predictors are chosen based on the stepwise discrimination procedure.
b) Comment on the cross-validation error results and how well the discrimination matches the groups. Compare the separation performance between this discriminant method versus the model with all predictors obtained in Exercise 1.

Exercise 3

Training and test sets give a more realistic estimate of misclassification. Now we split the infant sids data into a training set (N=45) and a test set (N=20). Observations from the test data are randomly chosen using **proc surveyselect**. Perform a discriminant analysis for **group** on the training set using the model chosen from Exercise 2. How does the discriminant analysis perform on the test data?