**Discriminant Analysis for 2 Groups**

**A Note on Using Misclassification Costs**

R.L. Andrews

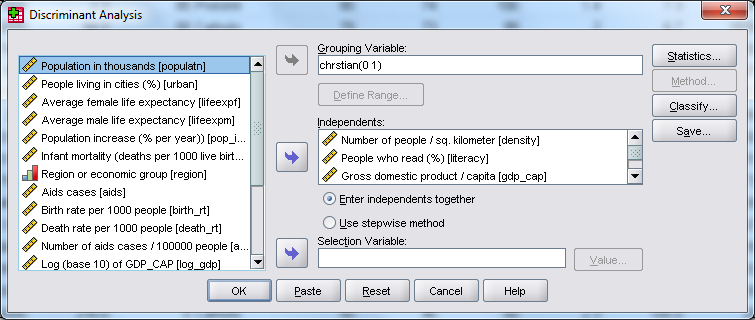
Consider a variable with two groups labeled 1 and 2. p1 denotes the proportion of the total in group 1 and p2 denotes the proportion of the total in group 2. p1 and p2 are the prior probabilities of an item being in the respective groups. C(2|1) is the cost of misclassification for saying that an item from group 1 is in group 2. C(1|2) is the cost of misclassification for saying that an item from group 2 is in group 1

Define a constant K such that .

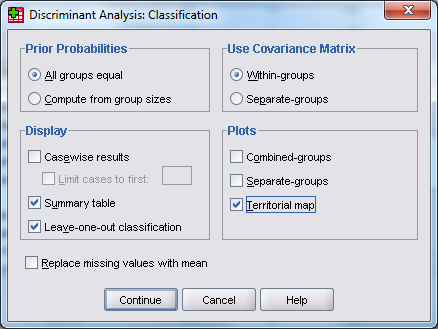
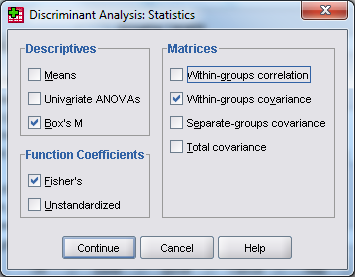
SPSS uses equal misclassification costs for the two groups but one can use K to adjust the prior probabilities to account for unequal misclassification costs.

Using the adjustment for misclassification costs, the input prior for group 1 becomes  and the input prior for group 2 becomes . However, if one prefers to use rather than K, then the adjusted input prior for group 1 is  and the adjusted input prior for group 2 is .

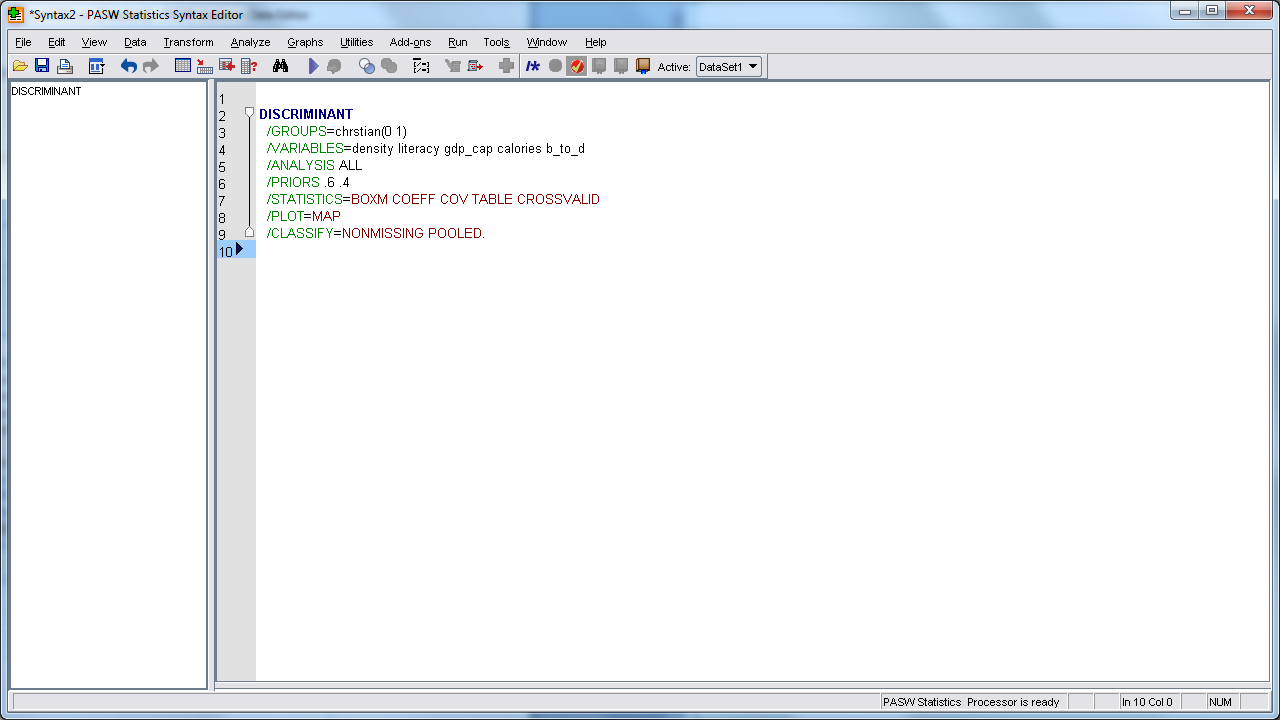
To enter these values in SPSS, set up the analysis using the menus (see the next page) for Statistics and Classify.



The menu for Classify button provides only two options for specifying prior probabilities as is shown in the image on the left.



Clicking on **Summary Table** gives a summary of classification results (hit rate results) for the data in this set. Clicking on **Leave-one-out classification** gives a summary of ***jackknife*** (see page 308 of Stevens 4th or 263 of Stevens 5th) or ***cross-validation*** (page 311 of Hair 6th or page 273 of Hair 7th) classification results for the data in this set. Clicking on **Fisher’s** gives coefficients that can be used to classify a data point into one of the categories. (I can’t find this option in JMP.)

Once these are set up, click on the **Paste** button in the menu on the previous page. For this situation I want to use p1=.6 and p2=.4. I have typed these values in after PRIORS replacing either EQUAL or SIZE. Now click on Run and Current to obtain the analysis output for this Discriminant Analysis.

JMP has fewer options. I can’t find a way to do the **Leave-one-out classification** option forevaluating a model or for calculating the linear discriminant functions. However JMP will provide the squared Mahalanobis distance of each point to the centroid for each group and a point can be classified into the closest group.



The prior probabilities can be stipulated in JMP by selecting Equal Probabilities, Proportional to Occurrence or Other, as shown above. If you select Other then you will get a menu like the one below that will allow you to enter probabilities for each group. If the entered probabilities do not total to 1, JMP adjusts them so that they are proper probabilities totaling to 1.

