**MKT 6323 Database Marketing**

**Assignment 3**

**Chih-Kai Wang**

**Kuo-Chi Lin**

**Te-Hsuan Lung**

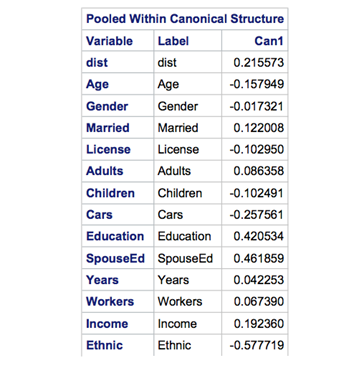
**Wei-Min Huang**

**Yu-Min Wang**

**Nov. 9, 2017**

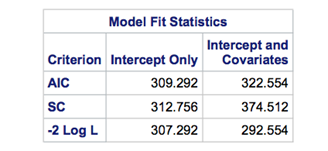
**i) Based on your results in (b) which demographic variables help discriminate potential users of public transportation from non-users. Why?**

To find variables that help discriminate potential users of public transportation from non-users, we have to look at the pooled within canonical structure of each variable. From the table below, we can see that variables Ethnic, Education and SpouseEd has the most significant canonical structure(-0.578, 0.420, 0.461). It can be concluded that these three variables help discriminate potential users of public transportation from non-users.



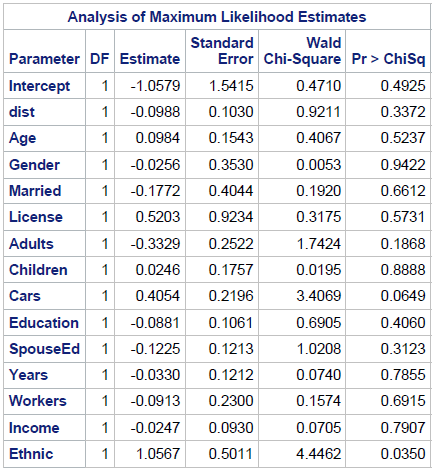
**ii) How good is your logistic regression – what criteria would you use to ascertain this?**

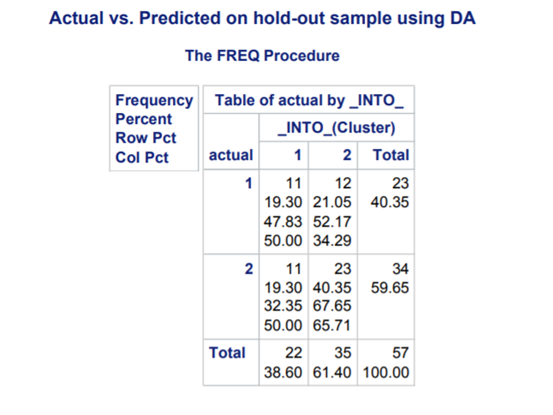
To see how good is the logistic regression, we have to look at the AIC, SC and -2Log L of the regression. We can see AIC and SC become even larger from “intercept only” to ”intercept and covariates“, it means the logistic model is not a good fit. -2LogL only decreases a little from 307.29 to 292.55, it says the variables can not explain lot.



**(iii) Which demographic variables are significant and why?**

To see which variables are significant, we have to look at the Pr > ChiSq which should be < 0.05. Looking at the table above, we can see only the variable Ethnic with Pr = 0.035 < 0.05, which can be concluded that Ethnic is the only significant variable.



**iv) Comment on the cross-tabulation in (c). What does it tell you about the effectiveness of your classification procedure?  
**

The above table indicates that the proportion of correct clustering is (11+23)/57 = 59.65% by adding actual consumers not using mass transit who are classified as cluster 1, 11 samples, and actual consumers using mass transit who are classified as cluster 2, 23 samples, then divided by total samples , 57 samples.