Data were collected from a group of citizens, and the government wants to understand the difference between full time and part time workers. They want to be able to classify what type of worker someone is likely to be depending on their demographic variables. There are three classifications – working full time, working part time or not working.

IV/Classification:

Wrkstat – Work status (full-time, part-time, not working)

DV/Predictors:

Happy – how happy in general a respondent is (very, pretty).

Childs – Number of children in the home.

Age – age of respondent.

Educ – highest year of school completed (continuous, <12 high school, 12-16 college, 17+ graduate)

Sex – (1 male, 2 female).

Race – (1 white, 2 black, 3 other).

Income – total family income.

Rincome – only respondent’s income.

**Research question:** Can we accurately predict the difference between the different work statuses? What variables predict the different workers?

Write-ups should include:

* Assumptions / Screening
* Significant function information (Chi-square/Wilks)
  + Include R2 – square the canonical correlation
* Important predictors for that function and what they mean
  + You only need to interpret the ones that have structural weights over absolute value | .3 |.
* Classification – did it classify correctly or not?
  + Where are you getting them right? Which group are you least correct? Include the table.

Questions:

How does someone get classified into one group over another in a discriminate function analysis?

Why is it problematic to have more than 3 classifications?

How is discriminate analysis similar to MANOVA?