

# MATH 4044 – Statistics for Data Science

---

## Practical Week 9

### Question 1

The data for this practical is stored in a SAS data file called `hsb2.sas7bdat` located in `mydata` library on the SAS OnDemand server.

This data file contains 200 observations from a sample of high school students with demographic information about the students, such as their gender (`female`), socio-economic status (`ses`) and ethnic background (`race`). It also contains a number of scores (out of 100) on standardized tests, including tests of reading (`read`), writing (`write`), mathematics (`math`) and social studies (`socst`).

- (a) Perform and analyse a factorial ANOVA model to determine whether there is statistically significant difference in average writing scores by gender and socio-economic status. Include tests for interaction. Interpret the results. Compare to the results from part (a).
- (b) Define dummy variables for the variable `ses`. Relate `write` to `ses` and `female` (which already is a dummy variable) including interaction using multiple regression. Discuss your results and compare to results from part (a).

**Note:** You can write your own code (see topic notes) or use Tasks. If you choose to use the Tasks menu, make sure that PROC GLM is used.