Homework 4

Shun-Lung Chang, Dilip Hiremath

```r
library(foreign)
hsb <- read.dta('data/hsbdemo.dta')
```

1. **Cross-tabulate the variables ses and prog.**

(a) (half a point) Which program was chosen by the largest fraction of students with high socio-economic status?

(b) (half a point) How many percent of students with low socio-economic status selected the general program?

(c) (half a point) In the academic program are there more students with middle socioeconomic status than students with high socio-economic status?

(d) (half a point) What is the least-frequent combination of the two variables?

2. **You continue with your analysis of the relationship between ses and prog.**

(a) (half a point) Draw a mosaicplot visualising the contingency table of program choice and socio-economic status.

(b) (1.5 points) Are students with low ses less likely (as measured in odds) to choose the academic program than students with higher socio-economic status? Calculate the odds ratios for choosing the academic program comparing students with low ses to students with middle ses and to students with high ses. [hint: use the command loddsratio in the package vcd. First, aggregate the variable prog into a binary variable indicating whether the student has chosen an academic program yes or no.]

3. **Now, you assess the relationship between prog and ses using the χ^2-statistic.**

(a) (1 point) Calculate the χ^2-test to assess the relationship between ses and prog. Is the relationship statistically significant?

(b) (1 point) Calculate the expected frequencies under the assumption that socio-economic status has no effect on program choice. For which cells are expected frequencies higher than the observed ones?

4. **In the following, perform the last analysis separately for female and male students.**

(a) (half a point) Calculate the χ^2-test to assess the relationship between ses and prog.

(b) (half a point) Calculate the expected frequencies under the assumption that socio- econmic status has no effect on program choice. For which cells are expected frequencies higher than the observed ones?

(c) (half a point) Do the results differ for the two sexes?

(d) (half a point) Visualise the relationships using mosaicplots. Get any differences between females and males in relation to socio-economic status and program choice visible in the plots?

5. **Create a multinomial logistic regression model using prog as dependent variable and the fol- lowing predictors: female, ses, schtype, read, write, math, science, honors, awards. [hint: use the function multinom in hte package nnet.]**

(a) (half a point) How large is the AIC score for this model?

(b) (1.5 points) The default output does not include p-values. Compute p-values based on the Wald-test statistics and determine the coefficients that are statistically significantly different from zero!