Question 1:

Code in R:

library(foreign)

ds1<-read.dta("hsbdemo.dta")

ds1$index<-c(1:200)

ds1<-subset(ds1, ds1$index<=100)

csp<-xtabs(~ds1$ses+ds1$prog)

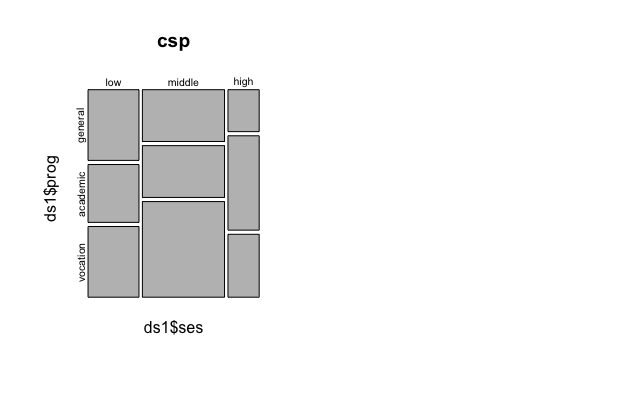
csp

1. The program was chosen by the largest fraction of students with high socio-economic status is: academic
2. 11% of students with low socio-economic status selected the general program
3. In the academic program, there are LESS students with middle socioeconomic status than students with high socio-economic status
4. The least-frequent combination of the two variables is: student with high social-economic with general program.

Question 2:

1. Code in R:

mosaicplot(csp)



1. The students with low ses are LESS likely (as measured in odds) to choose the academic program than students with higher socio-economic status

Code in R:

mosaicplot(csp)

ds1$binac<-NA

ds1$binac[ds1$prog=="academic"]<-"academic"

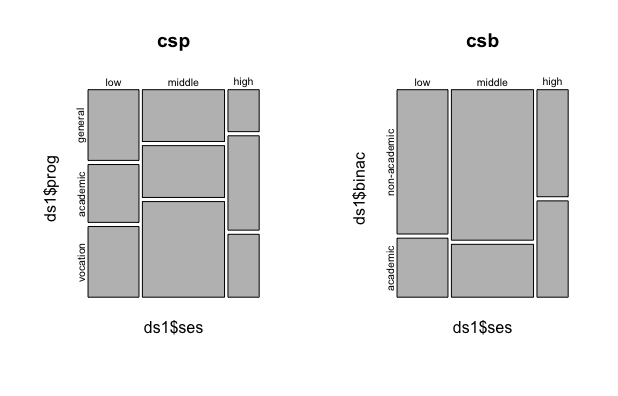
ds1$binac[ds1$prog=="general" | ds1$prog=="vocation"]<-"non-academic"

ds1$binac<-ordered(ds1$binac, levels=c("non-academic","academic"))

csb<-xtabs(~ds1$ses+ds1$binac)

csb

mosaicplot(csb)



loddsratio(csb)

log odds ratios for ds1$ses and ds1$binac

low:middle middle:high

-0.1521507 0.9406080

Question 3:

1. Code in R:

chisq.test(csp)

Result:

Pearson's Chi-squared test

data: csp

X-squared = 4.342, df = 4, p-value = 0.3617

We can see that the p-value = 0.3617 > 0.05 ⇒ the relationship is statistically insignificant.

1. Code in R:

chisq.test(csp)$expected

csp-chisq.test(csp)$expected

ds1$prog

ds1$ses general academic vocation

low 8.68 9.61 12.71

middle 14.00 15.50 20.50

high 5.32 5.89 7.79

ds1$prog

ds1$ses general academic vocation

low 2.32 -0.61 -1.71

middle -1.00 -2.50 3.50

high -1.32 3.11 -1.79

The cells that expected frequencies higher than the observed ones are:

Low ses with general prog

Middle ses with vocation prog

High ses with academic prog

Question 4:

1. Code in R:

fem<-subset(ds1, female=="female")

femsp <- xtabs(~fem$ses+fem$prog)

male<-subset(ds1, female=="male")

malesp <- xtabs(~male$ses+male$prog)

chisq.test(femsp)

chisq.test(malesp)

Result for female:

Pearson's Chi-squared test

data: femsp

X-squared = 1.154, df = 4, p-value = 0.8856

Result for male:

Pearson's Chi-squared test

data: malesp

X-squared = 7.5787, df = 4, p-value = 0.1083

1. Code in R:

chisq.test(femsp)$expected

femsp-chisq.test(femsp)$expected

chisq.test(malesp)$expected

malesp-chisq.test(malesp)$expected

For female:

fem$prog

fem$ses general academic vocation

low 0.1818182 0.4545455 -0.6363636

middle -0.8545455 -0.8363636 1.6909091

high 0.6727273 0.3818182 -1.0545455

The cells that expected frequencies higher than the observed ones are:

Low ses with general prog

Low ses with academic prog

Middle ses with vocation prog

High ses with general prog

High ses with academic prog

For male:

male$prog

male$ses general academic vocation

low 3.0000000 -2.2666667 -0.7333333

middle -2.3333333 0.1111111 2.2222222

high -0.6666667 2.1555556 -1.4888889

The cells that expected frequencies higher than the observed ones are:

Low ses with general prog

Middle ses with academic prog

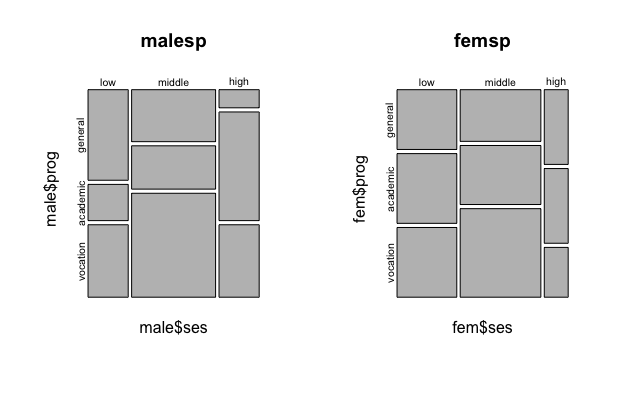
Middle ses with vocation prog

High ses with academic prog

1. The results has some different for the two sexes for female low ses with academic prog has positive value. Male attends academic prog with middle ses more than female. High ses female joins general prog more than model expected.
2. Code in R:

mosaicplot(malesp)

mosaicplot(femsp)



We can see that with low ses the proportion of female goes to academic prog more than male. Meanwhile with high ses, male tends to attend academic prog more than female. The middle ses female also has bigger portion of attending academic prog than the proportion of male with the same ses.

Question 5:

1. Code in R:

temp1<- multinom(prog~female+ses+schtyp+read+write+math+science+honors+awards, data=ds1)

summary(temp1)

The AIC score for this model is: 226.8863

1. Code in R:

wval<-summary(temp1, cor=FALSE, Wald=TRUE)$Wald.ratios

pval<-(1-pnorm(abs(wval),0,1))\*2

pval<0.05

The significant coefficients different from zero are:

Science and academic

Vocation and sesmiddle

Academic and honorenrolled

Vocation and honorenrolled

Question 6:

1. Code in R:

temp2<-step(temp1, direction="backward")

summary(temp2)

The predictors include in the models are: schtypprivate and math.

1. Code in R:

BIC(temp2)

The BIC score of the model is: 229.2506

1. Code in R:

logLik(temp2)

The log-likelihood score is: -100.8098 (df=6)

Question 7:

Code in R:

dftemp2<-data.frame(schtyp=rep(c("private", "public"), each=3),

math=rep(mean(ds1$math),6))

predict(temp2, dftemp2, "probs")

general academic vocation

1 0.3101117 0.6081881 0.08170023

2 0.3101117 0.6081881 0.08170023

3 0.3101117 0.6081881 0.08170023

4 0.2823303 0.2668204 0.45084927

5 0.2823303 0.2668204 0.45084927

6 0.2823303 0.2668204 0.45084927

The probability of students who are going for general program are almost the same form public and private school.

The probability of students who are going for academic program from private school is much higher than the ones who are from public school.

The probability of students who are going for vocation program from private school is much lower than the ones who are from public school.

Question 8:

Code in R:

dftemp31<-data.frame(schtyp=rep(c("private", "public"), each=153),

math=rep(c(30:80),6))

dftemp32<-predict(temp2, dftemp31, "probs")

dftemp34<-cbind(dftemp31, dftemp32)

by(dftemp34[,3:5], dftemp34$schtyp, colMeans)

dftemp34$schtyp: private

general academic vocation

0.28893425 0.63835616 0.07270959

-----------------------------------------------------------------------------

dftemp34$schtyp: public

general academic vocation

0.3177313 0.3575797 0.3246889

The math grade show the same results as in question 7. The probabilities are almost the same.

The probability of students who are going for general program are almost the same form public and private school.

The probability of students who are going for academic program from private school is much higher than the ones who are from public school.

The probability of students who are going for vocation program from private school is much lower than the ones who are from public school.