BUDA530Module3

## Problem 1 :

The discoveries dataset in the faraway package lists the numbers of “great” inventions and scientific discoveries in each year from 1860-1959. Investigate the number of discoveries over time. Use the appropriate Poisson model to fit the number of discoveries against the year. Hint you may need to create this. Comment on any findings of the model you have.

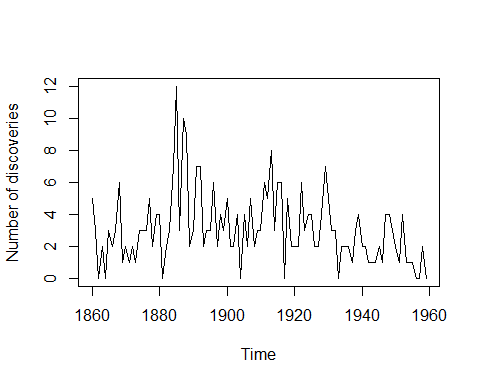
library(faraway)

## Warning: package 'faraway' was built under R version 3.6.2

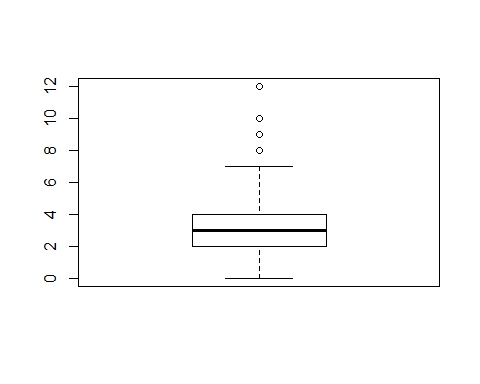
data("discoveries")  
  
summary(discoveries)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0 2.0 3.0 3.1 4.0 12.0

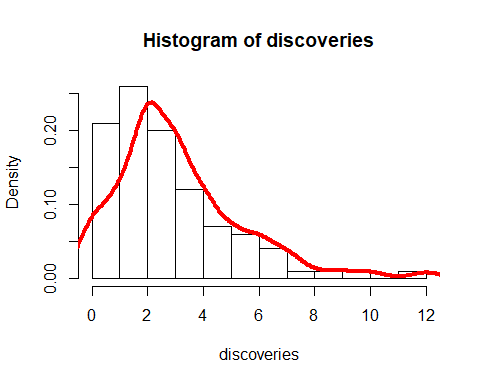
plot(discoveries, ylab="Number of discoveries")



boxplot(discoveries)



hist(discoveries,breaks=10,prob=TRUE)  
lines(density(discoveries, na.rm=T), col="red",lwd=4)



years <- 1860:1959  
  
model <- glm(discoveries~years, family=poisson)  
summary(model)

##   
## Call:  
## glm(formula = discoveries ~ years, family = poisson)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.8112 -0.9482 -0.3533 0.6637 3.5504   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 11.354807 3.775677 3.007 0.00264 \*\*  
## years -0.005360 0.001982 -2.705 0.00683 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 164.68 on 99 degrees of freedom  
## Residual deviance: 157.32 on 98 degrees of freedom  
## AIC: 430.32  
##   
## Number of Fisher Scoring iterations: 5

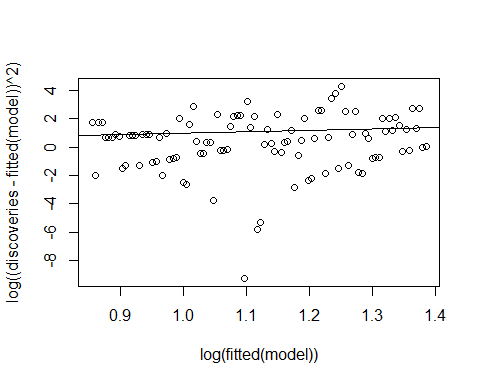
mean(discoveries)

## [1] 3.1

sd(discoveries)

## [1] 2.254065

plot(log(fitted(model)),log((discoveries-fitted(model))^2))  
abline(0,1)



### Interpretation :

As per the graph, more discoveries happens between 1880 to 1900 and large fluctations between these years. Based on the summary of model, years is significant at p-value < 0.05 level. Moreover, histogram shows that data is right-skewed.

Based on statistics of model, one unit increase of years parameter will decrease the discoveries by 0.005360 unit. The plot shows that points are normally distributed to the line which means variance and means are similar. Hence poisson model is fit for the analysis.

## Problem 2 :

The hsb data in the faraway package is a subset of the High School and Beyond study. The variables are school type, socioeconomic status, school type, chosen high school program type, scores in certain classes. Present any tables or plots you may find relevant to this analysis. The goal is to fit a model that explains a persons program choice type based on the observed variables. Investigate the coefficients and comment on any interesting findings. This is a multinomial model with program type as a response (3 levels).

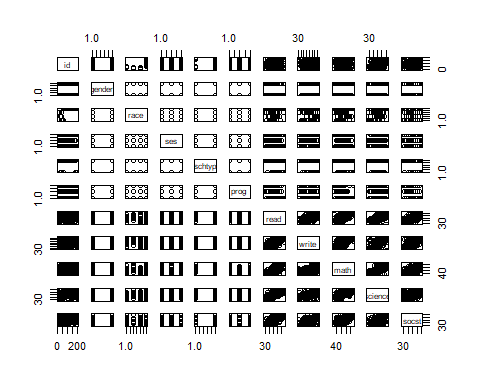
library(faraway)  
library(nnet)  
data(hsb)  
head(hsb)

## id gender race ses schtyp prog read write math science socst  
## 1 70 male white low public general 57 52 41 47 57  
## 2 121 female white middle public vocation 68 59 53 63 61  
## 3 86 male white high public general 44 33 54 58 31  
## 4 141 male white high public vocation 63 44 47 53 56  
## 5 172 male white middle public academic 47 52 57 53 61  
## 6 113 male white middle public academic 44 52 51 63 61

str(hsb)

## 'data.frame': 200 obs. of 11 variables:  
## $ id : int 70 121 86 141 172 113 50 11 84 48 ...  
## $ gender : Factor w/ 2 levels "female","male": 2 1 2 2 2 2 2 2 2 2 ...  
## $ race : Factor w/ 4 levels "african-amer",..: 4 4 4 4 4 4 1 3 4 1 ...  
## $ ses : Factor w/ 3 levels "high","low","middle": 2 3 1 1 3 3 3 3 3 3 ...  
## $ schtyp : Factor w/ 2 levels "private","public": 2 2 2 2 2 2 2 2 2 2 ...  
## $ prog : Factor w/ 3 levels "academic","general",..: 2 3 2 3 1 1 2 1 2 1 ...  
## $ read : int 57 68 44 63 47 44 50 34 63 57 ...  
## $ write : int 52 59 33 44 52 52 59 46 57 55 ...  
## $ math : int 41 53 54 47 57 51 42 45 54 52 ...  
## $ science: int 47 63 58 53 53 63 53 39 58 50 ...  
## $ socst : int 57 61 31 56 61 61 61 36 51 51 ...

plot(hsb)



summary(hsb)

## id gender race ses   
## Min. : 1.00 female:109 african-amer: 20 high :58   
## 1st Qu.: 50.75 male : 91 asian : 11 low :47   
## Median :100.50 hispanic : 24 middle:95   
## Mean :100.50 white :145   
## 3rd Qu.:150.25   
## Max. :200.00   
## schtyp prog read write   
## private: 32 academic:105 Min. :28.00 Min. :31.00   
## public :168 general : 45 1st Qu.:44.00 1st Qu.:45.75   
## vocation: 50 Median :50.00 Median :54.00   
## Mean :52.23 Mean :52.77   
## 3rd Qu.:60.00 3rd Qu.:60.00   
## Max. :76.00 Max. :67.00   
## math science socst   
## Min. :33.00 Min. :26.00 Min. :26.00   
## 1st Qu.:45.00 1st Qu.:44.00 1st Qu.:46.00   
## Median :52.00 Median :53.00 Median :52.00   
## Mean :52.65 Mean :51.85 Mean :52.41   
## 3rd Qu.:59.00 3rd Qu.:58.00 3rd Qu.:61.00   
## Max. :75.00 Max. :74.00 Max. :71.00

table( hsb$gender, hsb$prog)

##   
## academic general vocation  
## female 58 24 27  
## male 47 21 23

table(hsb$ses, hsb$prog)

##   
## academic general vocation  
## high 42 9 7  
## low 19 16 12  
## middle 44 20 31

table(hsb$schtyp, hsb$prog)

##   
## academic general vocation  
## private 24 6 2  
## public 81 39 48

model1 <- multinom(prog~., data=hsb)

## # weights: 45 (28 variable)  
## initial value 219.722458   
## iter 10 value 181.098338  
## iter 20 value 154.577078  
## iter 30 value 152.478856  
## final value 152.478368   
## converged

summary(model1)

## Call:  
## multinom(formula = prog ~ ., data = hsb)  
##   
## Coefficients:  
## (Intercept) id gendermale raceasian racehispanic  
## general 4.263658 -0.007332836 -0.04666403 1.2170225 -0.8702109  
## vocation 7.845921 -0.003680462 -0.29724832 -0.7863428 -0.3236628  
## racewhite seslow sesmiddle schtyppublic read  
## general 0.8609754 1.1547399 0.7430976 0.1384853 -0.05445264  
## vocation 0.6223190 0.0728241 1.1897765 1.8285649 -0.04078359  
## write math science socst  
## general -0.03716360 -0.1037470 0.1065258 -0.01786542  
## vocation -0.03220268 -0.1099712 0.0537472 -0.07959798  
##   
## Std. Errors:  
## (Intercept) id gendermale raceasian racehispanic  
## general 1.960941 0.007678009 0.4587870 1.064969 0.9286986  
## vocation 2.288984 0.008408855 0.5048241 1.476435 0.8924359  
## racewhite seslow sesmiddle schtyppublic read write  
## general 0.9438010 0.6134530 0.5096129 0.7338284 0.03300204 0.03398842  
## vocation 0.9519097 0.7067682 0.5739217 0.9981540 0.03583547 0.03597627  
## math science socst  
## general 0.03556357 0.03331314 0.02737227  
## vocation 0.03885464 0.03445137 0.02963317  
##   
## Residual Deviance: 304.9567   
## AIC: 360.9567

step(model1)

## Start: AIC=360.96  
## prog ~ id + gender + race + ses + schtyp + read + write + math +   
## science + socst  
##   
## trying - id   
## # weights: 42 (26 variable)  
## initial value 219.722458   
## iter 10 value 171.814970  
## iter 20 value 153.793692  
## iter 30 value 152.935260  
## final value 152.935256   
## converged  
## trying - gender   
## # weights: 42 (26 variable)  
## initial value 219.722458   
## iter 10 value 180.058431  
## iter 20 value 154.812587  
## iter 30 value 152.663811  
## final value 152.663786   
## converged  
## trying - race   
## # weights: 36 (22 variable)  
## initial value 219.722458   
## iter 10 value 180.513235  
## iter 20 value 156.699696  
## final value 155.608810   
## converged  
## trying - ses   
## # weights: 39 (24 variable)  
## initial value 219.722458   
## iter 10 value 178.070318  
## iter 20 value 159.672299  
## final value 158.649746   
## converged  
## trying - schtyp   
## # weights: 42 (26 variable)  
## initial value 219.722458   
## iter 10 value 182.731673  
## iter 20 value 156.661868  
## iter 30 value 154.540548  
## final value 154.540545   
## converged  
## trying - read   
## # weights: 42 (26 variable)  
## initial value 219.722458   
## iter 10 value 173.465840  
## iter 20 value 154.974752  
## iter 30 value 153.967893  
## final value 153.967881   
## converged  
## trying - write   
## # weights: 42 (26 variable)  
## initial value 219.722458   
## iter 10 value 183.360638  
## iter 20 value 154.719974  
## iter 30 value 153.192008  
## final value 153.191984   
## converged  
## trying - math   
## # weights: 42 (26 variable)  
## initial value 219.722458   
## iter 10 value 182.410292  
## iter 20 value 159.634346  
## iter 30 value 158.823267  
## final value 158.823263   
## converged  
## trying - science   
## # weights: 42 (26 variable)  
## initial value 219.722458   
## iter 10 value 181.365160  
## iter 20 value 158.514623  
## iter 30 value 158.121328  
## iter 30 value 158.121327  
## iter 30 value 158.121327  
## final value 158.121327   
## converged  
## trying - socst   
## # weights: 42 (26 variable)  
## initial value 219.722458   
## iter 10 value 188.794964  
## iter 20 value 157.284268  
## iter 30 value 156.585430  
## final value 156.585428   
## converged  
## Df AIC  
## - race 22 355.2176  
## - gender 26 357.3276  
## - id 26 357.8705  
## - write 26 358.3840  
## - read 26 359.9358  
## <none> 28 360.9567  
## - schtyp 26 361.0811  
## - socst 26 365.1709  
## - ses 24 365.2995  
## - science 26 368.2427  
## - math 26 369.6465  
## # weights: 36 (22 variable)  
## initial value 219.722458   
## iter 10 value 180.513235  
## iter 20 value 156.699696  
## final value 155.608810   
## converged  
##   
## Step: AIC=355.22  
## prog ~ id + gender + ses + schtyp + read + write + math + science +   
## socst  
##   
## trying - id   
## # weights: 33 (20 variable)  
## initial value 219.722458   
## iter 10 value 172.925326  
## iter 20 value 156.065379  
## final value 155.776076   
## converged  
## trying - gender   
## # weights: 33 (20 variable)  
## initial value 219.722458   
## iter 10 value 179.707679  
## iter 20 value 156.318319  
## final value 155.851235   
## converged  
## trying - ses   
## # weights: 30 (18 variable)  
## initial value 219.722458   
## iter 10 value 181.476339  
## iter 20 value 161.172288  
## final value 161.071677   
## converged  
## trying - schtyp   
## # weights: 33 (20 variable)  
## initial value 219.722458   
## iter 10 value 184.444565  
## iter 20 value 160.437562  
## final value 159.084037   
## converged  
## trying - read   
## # weights: 33 (20 variable)  
## initial value 219.722458   
## iter 10 value 175.690487  
## iter 20 value 157.298264  
## final value 156.720325   
## converged  
## trying - write   
## # weights: 33 (20 variable)  
## initial value 219.722458   
## iter 10 value 184.945637  
## iter 20 value 156.498714  
## final value 156.136288   
## converged  
## trying - math   
## # weights: 33 (20 variable)  
## initial value 219.722458   
## iter 10 value 193.026915  
## iter 20 value 162.759806  
## final value 162.292860   
## converged  
## trying - science   
## # weights: 33 (20 variable)  
## initial value 219.722458   
## iter 10 value 181.977682  
## iter 20 value 161.339132  
## final value 161.233380   
## converged  
## trying - socst   
## # weights: 33 (20 variable)  
## initial value 219.722458   
## iter 10 value 188.741082  
## iter 20 value 159.946523  
## final value 159.465122   
## converged  
## Df AIC  
## - id 20 351.5522  
## - gender 20 351.7025  
## - write 20 352.2726  
## - read 20 353.4407  
## <none> 22 355.2176  
## - ses 18 358.1434  
## - schtyp 20 358.1681  
## - socst 20 358.9302  
## - science 20 362.4668  
## - math 20 364.5857  
## # weights: 33 (20 variable)  
## initial value 219.722458   
## iter 10 value 172.925326  
## iter 20 value 156.065379  
## final value 155.776076   
## converged  
##   
## Step: AIC=351.55  
## prog ~ gender + ses + schtyp + read + write + math + science +   
## socst  
##   
## trying - gender   
## # weights: 30 (18 variable)  
## initial value 219.722458   
## iter 10 value 172.662548  
## iter 20 value 156.063823  
## final value 156.032828   
## converged  
## trying - ses   
## # weights: 27 (16 variable)  
## initial value 219.722458   
## iter 10 value 174.614066  
## iter 20 value 161.475590  
## final value 161.472216   
## converged  
## trying - schtyp   
## # weights: 30 (18 variable)  
## initial value 219.722458   
## iter 10 value 180.749264  
## iter 20 value 159.825179  
## final value 159.649518   
## converged  
## trying - read   
## # weights: 30 (18 variable)  
## initial value 219.722458   
## iter 10 value 183.217967  
## iter 20 value 156.956223  
## final value 156.905034   
## converged  
## trying - write   
## # weights: 30 (18 variable)  
## initial value 219.722458   
## iter 10 value 176.860078  
## iter 20 value 156.634024  
## final value 156.325078   
## converged  
## trying - math   
## # weights: 30 (18 variable)  
## initial value 219.722458   
## iter 10 value 183.819884  
## iter 20 value 162.568961  
## final value 162.533639   
## converged  
## trying - science   
## # weights: 30 (18 variable)  
## initial value 219.722458   
## iter 10 value 185.688554  
## iter 20 value 161.852139  
## final value 161.818793   
## converged  
## trying - socst   
## # weights: 30 (18 variable)  
## initial value 219.722458   
## iter 10 value 180.870401  
## iter 20 value 159.648982  
## final value 159.589251   
## converged  
## Df AIC  
## - gender 18 348.0657  
## - write 18 348.6502  
## - read 18 349.8101  
## <none> 20 351.5522  
## - ses 16 354.9444  
## - socst 18 355.1785  
## - schtyp 18 355.2990  
## - science 18 359.6376  
## - math 18 361.0673  
## # weights: 30 (18 variable)  
## initial value 219.722458   
## iter 10 value 172.662548  
## iter 20 value 156.063823  
## final value 156.032828   
## converged  
##   
## Step: AIC=348.07  
## prog ~ ses + schtyp + read + write + math + science + socst  
##   
## trying - ses   
## # weights: 24 (14 variable)  
## initial value 219.722458   
## iter 10 value 174.143433  
## iter 20 value 161.751125  
## iter 20 value 161.751124  
## iter 20 value 161.751124  
## final value 161.751124   
## converged  
## trying - schtyp   
## # weights: 27 (16 variable)  
## initial value 219.722458   
## iter 10 value 180.674669  
## iter 20 value 159.902074  
## final value 159.901300   
## converged  
## trying - read   
## # weights: 27 (16 variable)  
## initial value 219.722458   
## iter 10 value 182.139891  
## iter 20 value 157.256553  
## final value 157.255365   
## converged  
## trying - write   
## # weights: 27 (16 variable)  
## initial value 219.722458   
## iter 10 value 176.827677  
## iter 20 value 156.410686  
## final value 156.406678   
## converged  
## trying - math   
## # weights: 27 (16 variable)  
## initial value 219.722458   
## iter 10 value 183.645245  
## iter 20 value 162.999887  
## final value 162.998232   
## converged  
## trying - science   
## # weights: 27 (16 variable)  
## initial value 219.722458   
## iter 10 value 185.984215  
## iter 20 value 162.121250  
## final value 162.117077   
## converged  
## trying - socst   
## # weights: 27 (16 variable)  
## initial value 219.722458   
## iter 10 value 180.818124  
## iter 20 value 159.843366  
## final value 159.841915   
## converged  
## Df AIC  
## - write 16 344.8134  
## - read 16 346.5107  
## <none> 18 348.0657  
## - ses 14 351.5022  
## - socst 16 351.6838  
## - schtyp 16 351.8026  
## - science 16 356.2342  
## - math 16 357.9965  
## # weights: 27 (16 variable)  
## initial value 219.722458   
## iter 10 value 176.827677  
## iter 20 value 156.410686  
## final value 156.406678   
## converged  
##   
## Step: AIC=344.81  
## prog ~ ses + schtyp + read + math + science + socst  
##   
## trying - ses   
## # weights: 21 (12 variable)  
## initial value 219.722458   
## iter 10 value 175.697433  
## final value 162.312774   
## converged  
## trying - schtyp   
## # weights: 24 (14 variable)  
## initial value 219.722458   
## iter 10 value 183.996922  
## iter 20 value 160.624557  
## final value 160.624514   
## converged  
## trying - read   
## # weights: 24 (14 variable)  
## initial value 219.722458   
## iter 10 value 171.169761  
## iter 20 value 157.775586  
## final value 157.775540   
## converged  
## trying - math   
## # weights: 24 (14 variable)  
## initial value 219.722458   
## iter 10 value 175.971820  
## iter 20 value 164.774187  
## final value 164.774168   
## converged  
## trying - science   
## # weights: 24 (14 variable)  
## initial value 219.722458   
## iter 10 value 175.048445  
## iter 20 value 162.188199  
## final value 162.188190   
## converged  
## trying - socst   
## # weights: 24 (14 variable)  
## initial value 219.722458   
## iter 10 value 170.620432  
## iter 20 value 161.495973  
## final value 161.495961   
## converged  
## Df AIC  
## - read 14 343.5511  
## <none> 16 344.8134  
## - ses 12 348.6255  
## - schtyp 14 349.2490  
## - socst 14 350.9919  
## - science 14 352.3764  
## - math 14 357.5483  
## # weights: 24 (14 variable)  
## initial value 219.722458   
## iter 10 value 171.169761  
## iter 20 value 157.775586  
## final value 157.775540   
## converged  
##   
## Step: AIC=343.55  
## prog ~ ses + schtyp + math + science + socst  
##   
## trying - ses   
## # weights: 18 (10 variable)  
## initial value 219.722458   
## iter 10 value 166.035430  
## final value 163.818866   
## converged  
## trying - schtyp   
## # weights: 21 (12 variable)  
## initial value 219.722458   
## iter 10 value 171.145160  
## iter 20 value 162.019070  
## iter 20 value 162.019070  
## iter 20 value 162.019070  
## final value 162.019070   
## converged  
## trying - math   
## # weights: 21 (12 variable)  
## initial value 219.722458   
## iter 10 value 171.943351  
## final value 169.805302   
## converged  
## trying - science   
## # weights: 21 (12 variable)  
## initial value 219.722458   
## iter 10 value 165.438528  
## final value 162.495650   
## converged  
## trying - socst   
## # weights: 21 (12 variable)  
## initial value 219.722458   
## iter 10 value 167.071471  
## final value 165.099779   
## converged  
## Df AIC  
## <none> 14 343.5511  
## - ses 10 347.6377  
## - schtyp 12 348.0381  
## - science 12 348.9913  
## - socst 12 354.1996  
## - math 12 363.6106

## Call:  
## multinom(formula = prog ~ ses + schtyp + math + science + socst,   
## data = hsb)  
##   
## Coefficients:  
## (Intercept) seslow sesmiddle schtyppublic math  
## general 2.587029 0.87607389 0.6978995 0.6468812 -0.1212242  
## vocation 6.687272 -0.01569301 1.2065000 1.9955504 -0.1369641  
## science socst  
## general 0.08209791 -0.04441228  
## vocation 0.03941237 -0.09363417  
##   
## Residual Deviance: 315.5511   
## AIC: 343.5511

modS <- step(model1)

## Start: AIC=360.96  
## prog ~ id + gender + race + ses + schtyp + read + write + math +   
## science + socst  
##   
## trying - id   
## # weights: 42 (26 variable)  
## initial value 219.722458   
## iter 10 value 171.814970  
## iter 20 value 153.793692  
## iter 30 value 152.935260  
## final value 152.935256   
## converged  
## trying - gender   
## # weights: 42 (26 variable)  
## initial value 219.722458   
## iter 10 value 180.058431  
## iter 20 value 154.812587  
## iter 30 value 152.663811  
## final value 152.663786   
## converged  
## trying - race   
## # weights: 36 (22 variable)  
## initial value 219.722458   
## iter 10 value 180.513235  
## iter 20 value 156.699696  
## final value 155.608810   
## converged  
## trying - ses   
## # weights: 39 (24 variable)  
## initial value 219.722458   
## iter 10 value 178.070318  
## iter 20 value 159.672299  
## final value 158.649746   
## converged  
## trying - schtyp   
## # weights: 42 (26 variable)  
## initial value 219.722458   
## iter 10 value 182.731673  
## iter 20 value 156.661868  
## iter 30 value 154.540548  
## final value 154.540545   
## converged  
## trying - read   
## # weights: 42 (26 variable)  
## initial value 219.722458   
## iter 10 value 173.465840  
## iter 20 value 154.974752  
## iter 30 value 153.967893  
## final value 153.967881   
## converged  
## trying - write   
## # weights: 42 (26 variable)  
## initial value 219.722458   
## iter 10 value 183.360638  
## iter 20 value 154.719974  
## iter 30 value 153.192008  
## final value 153.191984   
## converged  
## trying - math   
## # weights: 42 (26 variable)  
## initial value 219.722458   
## iter 10 value 182.410292  
## iter 20 value 159.634346  
## iter 30 value 158.823267  
## final value 158.823263   
## converged  
## trying - science   
## # weights: 42 (26 variable)  
## initial value 219.722458   
## iter 10 value 181.365160  
## iter 20 value 158.514623  
## iter 30 value 158.121328  
## iter 30 value 158.121327  
## iter 30 value 158.121327  
## final value 158.121327   
## converged  
## trying - socst   
## # weights: 42 (26 variable)  
## initial value 219.722458   
## iter 10 value 188.794964  
## iter 20 value 157.284268  
## iter 30 value 156.585430  
## final value 156.585428   
## converged  
## Df AIC  
## - race 22 355.2176  
## - gender 26 357.3276  
## - id 26 357.8705  
## - write 26 358.3840  
## - read 26 359.9358  
## <none> 28 360.9567  
## - schtyp 26 361.0811  
## - socst 26 365.1709  
## - ses 24 365.2995  
## - science 26 368.2427  
## - math 26 369.6465  
## # weights: 36 (22 variable)  
## initial value 219.722458   
## iter 10 value 180.513235  
## iter 20 value 156.699696  
## final value 155.608810   
## converged  
##   
## Step: AIC=355.22  
## prog ~ id + gender + ses + schtyp + read + write + math + science +   
## socst  
##   
## trying - id   
## # weights: 33 (20 variable)  
## initial value 219.722458   
## iter 10 value 172.925326  
## iter 20 value 156.065379  
## final value 155.776076   
## converged  
## trying - gender   
## # weights: 33 (20 variable)  
## initial value 219.722458   
## iter 10 value 179.707679  
## iter 20 value 156.318319  
## final value 155.851235   
## converged  
## trying - ses   
## # weights: 30 (18 variable)  
## initial value 219.722458   
## iter 10 value 181.476339  
## iter 20 value 161.172288  
## final value 161.071677   
## converged  
## trying - schtyp   
## # weights: 33 (20 variable)  
## initial value 219.722458   
## iter 10 value 184.444565  
## iter 20 value 160.437562  
## final value 159.084037   
## converged  
## trying - read   
## # weights: 33 (20 variable)  
## initial value 219.722458   
## iter 10 value 175.690487  
## iter 20 value 157.298264  
## final value 156.720325   
## converged  
## trying - write   
## # weights: 33 (20 variable)  
## initial value 219.722458   
## iter 10 value 184.945637  
## iter 20 value 156.498714  
## final value 156.136288   
## converged  
## trying - math   
## # weights: 33 (20 variable)  
## initial value 219.722458   
## iter 10 value 193.026915  
## iter 20 value 162.759806  
## final value 162.292860   
## converged  
## trying - science   
## # weights: 33 (20 variable)  
## initial value 219.722458   
## iter 10 value 181.977682  
## iter 20 value 161.339132  
## final value 161.233380   
## converged  
## trying - socst   
## # weights: 33 (20 variable)  
## initial value 219.722458   
## iter 10 value 188.741082  
## iter 20 value 159.946523  
## final value 159.465122   
## converged  
## Df AIC  
## - id 20 351.5522  
## - gender 20 351.7025  
## - write 20 352.2726  
## - read 20 353.4407  
## <none> 22 355.2176  
## - ses 18 358.1434  
## - schtyp 20 358.1681  
## - socst 20 358.9302  
## - science 20 362.4668  
## - math 20 364.5857  
## # weights: 33 (20 variable)  
## initial value 219.722458   
## iter 10 value 172.925326  
## iter 20 value 156.065379  
## final value 155.776076   
## converged  
##   
## Step: AIC=351.55  
## prog ~ gender + ses + schtyp + read + write + math + science +   
## socst  
##   
## trying - gender   
## # weights: 30 (18 variable)  
## initial value 219.722458   
## iter 10 value 172.662548  
## iter 20 value 156.063823  
## final value 156.032828   
## converged  
## trying - ses   
## # weights: 27 (16 variable)  
## initial value 219.722458   
## iter 10 value 174.614066  
## iter 20 value 161.475590  
## final value 161.472216   
## converged  
## trying - schtyp   
## # weights: 30 (18 variable)  
## initial value 219.722458   
## iter 10 value 180.749264  
## iter 20 value 159.825179  
## final value 159.649518   
## converged  
## trying - read   
## # weights: 30 (18 variable)  
## initial value 219.722458   
## iter 10 value 183.217967  
## iter 20 value 156.956223  
## final value 156.905034   
## converged  
## trying - write   
## # weights: 30 (18 variable)  
## initial value 219.722458   
## iter 10 value 176.860078  
## iter 20 value 156.634024  
## final value 156.325078   
## converged  
## trying - math   
## # weights: 30 (18 variable)  
## initial value 219.722458   
## iter 10 value 183.819884  
## iter 20 value 162.568961  
## final value 162.533639   
## converged  
## trying - science   
## # weights: 30 (18 variable)  
## initial value 219.722458   
## iter 10 value 185.688554  
## iter 20 value 161.852139  
## final value 161.818793   
## converged  
## trying - socst   
## # weights: 30 (18 variable)  
## initial value 219.722458   
## iter 10 value 180.870401  
## iter 20 value 159.648982  
## final value 159.589251   
## converged  
## Df AIC  
## - gender 18 348.0657  
## - write 18 348.6502  
## - read 18 349.8101  
## <none> 20 351.5522  
## - ses 16 354.9444  
## - socst 18 355.1785  
## - schtyp 18 355.2990  
## - science 18 359.6376  
## - math 18 361.0673  
## # weights: 30 (18 variable)  
## initial value 219.722458   
## iter 10 value 172.662548  
## iter 20 value 156.063823  
## final value 156.032828   
## converged  
##   
## Step: AIC=348.07  
## prog ~ ses + schtyp + read + write + math + science + socst  
##   
## trying - ses   
## # weights: 24 (14 variable)  
## initial value 219.722458   
## iter 10 value 174.143433  
## iter 20 value 161.751125  
## iter 20 value 161.751124  
## iter 20 value 161.751124  
## final value 161.751124   
## converged  
## trying - schtyp   
## # weights: 27 (16 variable)  
## initial value 219.722458   
## iter 10 value 180.674669  
## iter 20 value 159.902074  
## final value 159.901300   
## converged  
## trying - read   
## # weights: 27 (16 variable)  
## initial value 219.722458   
## iter 10 value 182.139891  
## iter 20 value 157.256553  
## final value 157.255365   
## converged  
## trying - write   
## # weights: 27 (16 variable)  
## initial value 219.722458   
## iter 10 value 176.827677  
## iter 20 value 156.410686  
## final value 156.406678   
## converged  
## trying - math   
## # weights: 27 (16 variable)  
## initial value 219.722458   
## iter 10 value 183.645245  
## iter 20 value 162.999887  
## final value 162.998232   
## converged  
## trying - science   
## # weights: 27 (16 variable)  
## initial value 219.722458   
## iter 10 value 185.984215  
## iter 20 value 162.121250  
## final value 162.117077   
## converged  
## trying - socst   
## # weights: 27 (16 variable)  
## initial value 219.722458   
## iter 10 value 180.818124  
## iter 20 value 159.843366  
## final value 159.841915   
## converged  
## Df AIC  
## - write 16 344.8134  
## - read 16 346.5107  
## <none> 18 348.0657  
## - ses 14 351.5022  
## - socst 16 351.6838  
## - schtyp 16 351.8026  
## - science 16 356.2342  
## - math 16 357.9965  
## # weights: 27 (16 variable)  
## initial value 219.722458   
## iter 10 value 176.827677  
## iter 20 value 156.410686  
## final value 156.406678   
## converged  
##   
## Step: AIC=344.81  
## prog ~ ses + schtyp + read + math + science + socst  
##   
## trying - ses   
## # weights: 21 (12 variable)  
## initial value 219.722458   
## iter 10 value 175.697433  
## final value 162.312774   
## converged  
## trying - schtyp   
## # weights: 24 (14 variable)  
## initial value 219.722458   
## iter 10 value 183.996922  
## iter 20 value 160.624557  
## final value 160.624514   
## converged  
## trying - read   
## # weights: 24 (14 variable)  
## initial value 219.722458   
## iter 10 value 171.169761  
## iter 20 value 157.775586  
## final value 157.775540   
## converged  
## trying - math   
## # weights: 24 (14 variable)  
## initial value 219.722458   
## iter 10 value 175.971820  
## iter 20 value 164.774187  
## final value 164.774168   
## converged  
## trying - science   
## # weights: 24 (14 variable)  
## initial value 219.722458   
## iter 10 value 175.048445  
## iter 20 value 162.188199  
## final value 162.188190   
## converged  
## trying - socst   
## # weights: 24 (14 variable)  
## initial value 219.722458   
## iter 10 value 170.620432  
## iter 20 value 161.495973  
## final value 161.495961   
## converged  
## Df AIC  
## - read 14 343.5511  
## <none> 16 344.8134  
## - ses 12 348.6255  
## - schtyp 14 349.2490  
## - socst 14 350.9919  
## - science 14 352.3764  
## - math 14 357.5483  
## # weights: 24 (14 variable)  
## initial value 219.722458   
## iter 10 value 171.169761  
## iter 20 value 157.775586  
## final value 157.775540   
## converged  
##   
## Step: AIC=343.55  
## prog ~ ses + schtyp + math + science + socst  
##   
## trying - ses   
## # weights: 18 (10 variable)  
## initial value 219.722458   
## iter 10 value 166.035430  
## final value 163.818866   
## converged  
## trying - schtyp   
## # weights: 21 (12 variable)  
## initial value 219.722458   
## iter 10 value 171.145160  
## iter 20 value 162.019070  
## iter 20 value 162.019070  
## iter 20 value 162.019070  
## final value 162.019070   
## converged  
## trying - math   
## # weights: 21 (12 variable)  
## initial value 219.722458   
## iter 10 value 171.943351  
## final value 169.805302   
## converged  
## trying - science   
## # weights: 21 (12 variable)  
## initial value 219.722458   
## iter 10 value 165.438528  
## final value 162.495650   
## converged  
## trying - socst   
## # weights: 21 (12 variable)  
## initial value 219.722458   
## iter 10 value 167.071471  
## final value 165.099779   
## converged  
## Df AIC  
## <none> 14 343.5511  
## - ses 10 347.6377  
## - schtyp 12 348.0381  
## - science 12 348.9913  
## - socst 12 354.1996  
## - math 12 363.6106

model2 <- multinom(prog~science+math+socst+ses+schtyp, data=hsb)

## # weights: 24 (14 variable)  
## initial value 219.722458   
## iter 10 value 171.169761  
## iter 20 value 157.775586  
## final value 157.775540   
## converged

summary(model2)

## Call:  
## multinom(formula = prog ~ science + math + socst + ses + schtyp,   
## data = hsb)  
##   
## Coefficients:  
## (Intercept) science math socst seslow  
## general 2.587029 0.08209791 -0.1212242 -0.04441228 0.87607389  
## vocation 6.687272 0.03941237 -0.1369641 -0.09363417 -0.01569301  
## sesmiddle schtyppublic  
## general 0.6978995 0.6468812  
## vocation 1.2065000 1.9955504  
##   
## Std. Errors:  
## (Intercept) science math socst seslow sesmiddle  
## general 1.686492 0.02787694 0.03213345 0.02344856 0.5758781 0.4930330  
## vocation 1.945363 0.02864929 0.03591701 0.02586717 0.6690861 0.5571202  
## schtyppublic  
## general 0.545598  
## vocation 0.812881  
##   
## Residual Deviance: 315.5511   
## AIC: 343.5511

output <- summary(model2)  
  
predict(model2, newdata = data.frame(science=35, math=70, socst= 35, ses = c("high","low","middle"),schtyp=c( "private")), type="prob")

## academic general vocation  
## 1 0.9818315 0.01007192 0.008096565  
## 2 0.9682859 0.02385355 0.007860535  
## 3 0.9540420 0.01966694 0.026291033

predict(model2, newdata = data.frame(science=30, math=40, socst= 35, ses=c("low","middle","high"),schtyp="public"), type="prob")

## academic general vocation  
## 1 0.1934177 0.22914975 0.5774325  
## 2 0.0824691 0.08175875 0.8357722  
## 3 0.2209466 0.10900251 0.6700509

predict(model2, newdata = data.frame(science=30, math=40, socst= 35, ses="low",schtyp="public"), type="prob")

## academic general vocation   
## 0.1934177 0.2291498 0.5774325

predict(model2, newdata = data.frame(science=52, math=40, socst= 40, ses="middle",schtyp="public"), type="prob")

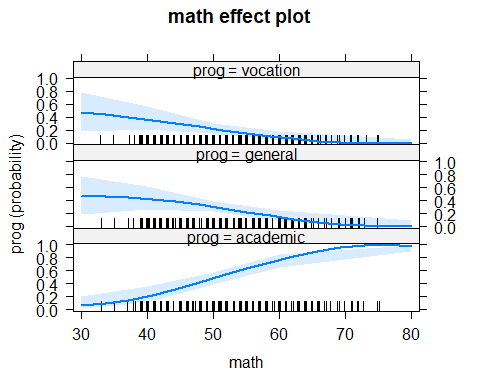
## academic general vocation   
## 0.04776696 0.23085180 0.72138124

library(effects)

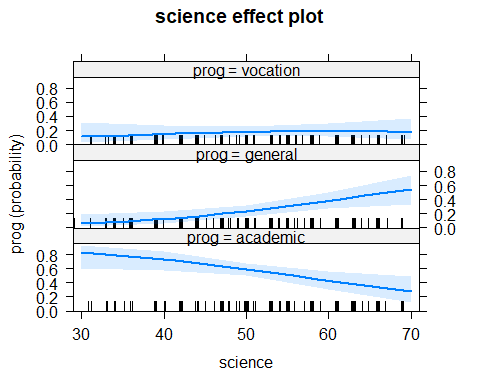
## Loading required package: carData

## lattice theme set by effectsTheme()  
## See ?effectsTheme for details.

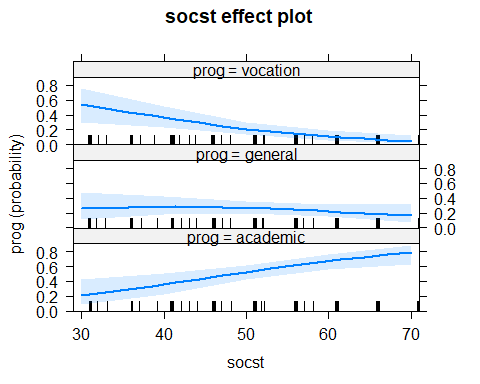
plot(Effect("math",model2))



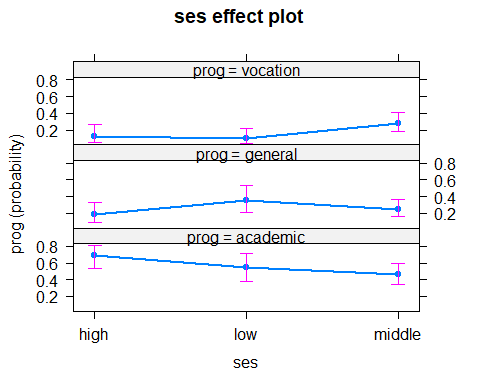
plot(Effect("science",model2))



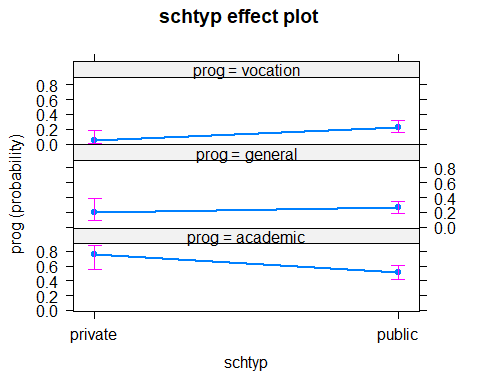
plot(Effect("socst",model2))



plot(Effect("ses",model2))



plot(Effect("schtyp",model2))



z <- output$coefficients/output$standard.errors  
p <- (1 - pnorm(abs(z), 0, 1))\*2  
  
Pclass2 <- rbind(output$coefficients[1,],output$standard.errors[1,],z[1,],p[1,])  
rownames(Pclass2) <- c("Coefficient","Std. Errors","z stat","p value")  
knitr::kable(Pclass2)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | (Intercept) | science | math | socst | seslow | sesmiddle | schtyppublic |
| Coefficient | 2.5870289 | 0.0820979 | -0.1212242 | -0.0444123 | 0.8760739 | 0.6978995 | 0.6468812 |
| Std. Errors | 1.6864918 | 0.0278769 | 0.0321335 | 0.0234486 | 0.5758781 | 0.4930330 | 0.5455980 |
| z stat | 1.5339706 | 2.9450114 | -3.7725237 | -1.8940300 | 1.5212835 | 1.4155229 | 1.1856370 |
| p value | 0.1250369 | 0.0032294 | 0.0001616 | 0.0582210 | 0.1281887 | 0.1569152 | 0.2357657 |

### Interpretation :

Based on the step function, 5 predictors are selected for response program Type with lowest AIC 343.5511. Baseline of the model with 5 variables : science, math, socst, ses and schtyp is academic program types with high socioeconomic class and private school.

Predicting the probablities of variables at different values,all five variables are making difference on the response. Moreover, effects plot shows similar visibility that varaibles science, math and socst are significant for choosing the program type while schtype and ses are significant at some level when choosing acedemic prog type.

Based on p-values, science, math and socst scores are significant at 0.05 level hence we reject the null hypothesis for these variables. However, p-values of ses and schtyp are at 0.1 level and are signigicant factor for chosing program type as per analysis.

Summary of model says that students with low math and socst scores in middle socioeconomic class attended public school are chosing vocation progam type against academic. With high math score studied in private school in high socieconomic class, students tend to chose academic program.

Looking at coefficients of summary, as the tuning parameters (predictors) increases by one unit, the log odds will decrease by that unit. For example, if science score increases by 0.0820 and all others are set to zero than response variable program type will increase by 0.2669 unit ( 2.597 - 0.0820). If predictors are set to zero, log odds are 2.5870 against academic program type.