

loglinear

Dong Luo

22 October 2018

Data Cleaning

We exclude the observations with age less than 18 since our research questions only focus on adult.

We cut each of the continuous variables `CHOPER1`, `FATPER1`, and `PROPER1`, which stand for the percentage of energy comes from carbohydrate, fat, and protein, into three distinct levels, from low, medium, to high.

| | Carbohydrate | Fat | Protein |
|-----------|--------------|----------|----------|
| low(%) | [0,45] | [0,20] | [0,15] |
| medium(%) | (45,65] | (20,35] | (15,25] |
| high(%) | (65,100] | (35,100] | (25,100] |

By dividing we are interested in the mean proportion of each diet types, so as to identify what are the most popular diet types.

```
tables.3way<-table(carb.cat,fat.cat,protein.cat)
table.y<-NULL
for(i in 1:3){
  for(j in 1:3){
    table.y<-c(table.y,tables.3way[j,,i])
  }
}
table.fat<-factor(rep(c('low','medium','high'),9))
table.carb<-factor(rep(c('low','medium','high'),each=3,times=3))
table.protein<-factor(rep(c('low','medium','high'),each=9))
loglin.dat<-data.frame(y=table.y,
                       fat=table.fat,
                       carb=table.carb,
                       protein=table.protein)
diet.prop<-cbind(loglin.dat[,1]/nrow(adultData),
                 loglin.dat[,2:4])
colnames(diet.prop)[1]<-'proportion'

# sort the df by proportion
diet.prop<-diet.prop[order(
  diet.prop$proportion,decreasing=TRUE),]

# see most common diet types in our sample
kable_styling(kable(
  head(diet.prop),
  row.names = F,
  format='latex',
  booktabs=TRUE),position = "center")
```

| proportion | fat | carb | protein |
|------------|--------|--------|---------|
| 0.1902491 | medium | medium | medium |
| 0.1785904 | high | low | medium |
| 0.1383148 | medium | low | medium |
| 0.1267621 | medium | medium | low |
| 0.0714361 | medium | low | high |
| 0.0557499 | high | low | low |