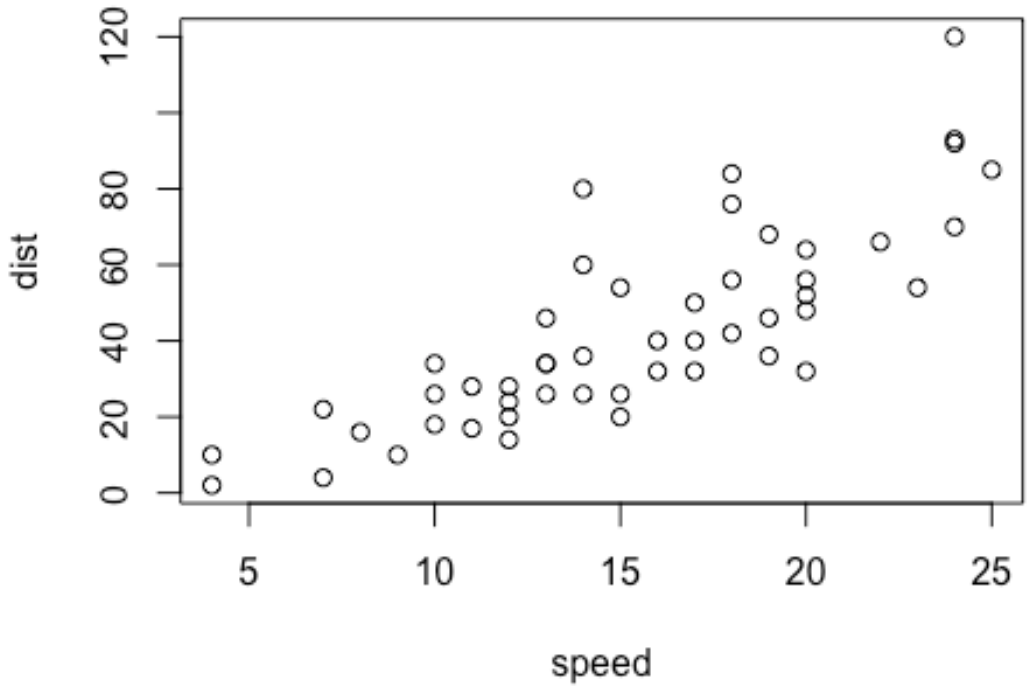
R Notebook

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

```
plot(cars)
```



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.

pdf("output.pdf")

install.packages('bayesm') library(bayesm) data(margarine) head(margarine) library(mlogit) library(tinytex)

#Q1

data = margarine$choicePrice #Average apply(as.matrix(margarine$choicePrice[,c(3:12)]),2,mean) #dispersion apply(as.matrix(margarine$choicePrice[,c(3:12)]),2,sd)apply(as.matrix(margarinechoicePrice[,c(3:12)]),2,range) apply(as.matrix(margarine$choicePrice[,c(3:12)]),2,var)

#Market share choice frequency nrow(data)

tb_share <-table(data$choice) names(tb_share) <- colnames(data)[3:12][as.numeric(names(tb_share))] tb_share

#market share by product characteristics(choice frequency by price bins:below average, over average) df <-data.frame() for(i in c(1:10)){ df_new <- data[which(data$choice==i),c('hhid','choice',colnames(data)[i+2])] colnames(df_new)[3] <- 'price' df <-rbind(df,df_new) }

avg_price <-tapply(df$price, df$choice,mean) avg_price <- data.frame(choice=names(avg_price),avg_prc=avg_price)

df <-merge(df,avg_price,by='choice')

over_average <-tapply(ifelse(df$price >= df$avg_prc,1,0),df$choice, sum)names(over_average) < -colnames(data)[3:12][as.numeric(names(over_average))]below_average < -tapply(ifelse(df$price<df$avg_prc, 1,0), df$choice,sum) names(below_average) <- colnames(data)[3:12][as.numeric(names(below_average))]

fre_by_price_bins<-rbind(tb_share,over_average,below_average)

#Illustrate the mapping between observed attributes and choices The combination of Parkay brand and stick type has the largest market share. The lowest market share is the combination of House Brand and tub Type. From the aspect of product type, the market share of stick is generally larger than the market share of tub. From the aspect of Brands, Parkay is the most popular brand.

#Q2 First Model

#choice matrix #choice matrix #choice matrix

clogit=function(param){ data = margarine$choicePrice ni=nrow(data) nj=ncol(data[,3:12])

df <-data.frame() for(i in c(1:10)){ df_new <- data[which(data$choice==i),c('hhid','choice',colnames(data)[i+2])] colnames(df_new)[3] <- 'price' df <-rbind(df,df_new) }

df$choice < -as.factor(df$choice) df <-cbind(df,as.data.frame(model.matrix(~choice-1,df))) Y=df[,4:13]

intercept=matrix(rep(param[1:nj-1],each=ni),nrow=ni,ncol=nj-1) XB=cbind(0,intercept)+data[,3:12]*param[nj] eXB=exp(XB) prob=eXB/rowSums(eXB) llike=sum(Y*log(prob)) return(-llike) } set.seed(1000) param=runif(10,-1,1) model1=optim(param,clogit,method="BFGS") model1$par

#interpretation of coefficient increasing in price will decrease the probability of choosing choice 1-9 but increase the probability of choosing choice 10.

#Q3 Second Model

data = margarine*choicePricedata2* = *margarine*demos data3 = merge(data,data2[,c(1:2)],by='hhid')

mlogit

mlogit = function(param) { data = data3 Income = data*Incomechoice* = *data*choice

ni = nrow(data) nj = length(unique(choice)) ut = mat.or.vec(ni,nj) # multinomial logit pn1 = param[1:nj]

multinomial logit for (j in 1:nj) {
ut[,j] = Income*pn1[j] } #exp(XB) prob = exp(ut)
#sprob = rowsums(prob) # sum_j exp(XB) denominator prob = sweep(prob,MARGIN=1,FUN="/",STATS=rowSums(prob)) # an example of how to construct

avg_param <-c() for(j in 1:nj){ avg_param<-c(avg_param,mean(prob[,j]*param[j])) } # match prob to actual choices probc = NULL me=mat.or.vec(ni,nj)

for (i in 1:ni) { probc[i] = prob[i,choice[i]] for(j in 1:nj){ me[i,j] = probc[i]*(param[j]-avg_param[j]) }

} probc[probc>0.999999] = 0.999999 probc[probc<0.000001] = 0.000001 like = sum(log(probc)) result = list(like=-like,me=me) return(result) }

set.seed(1000) param = rep(0,10)

like_fun = function(param){ result = mlogit(param) like=result$like return(like) } model <-optim(param,like_fun,method='BFGS') model

#Interprete coefficient on family Increasing in family income could reduce the probability of choosing choice 1.

#Q4

#marginal effect of first model

#marginal effect of Second model me=mlogit(model*par*)me summary(me) Increasing one unit of income could change the probability of choosing the choice.

#Q5 #combining dataset of price and income as data4 data = margarine$choicePricedata2 = margarine$demos data4 = merge(df,data2[,c(1:2)],by='hhid')