**Question 1**

In this context, the response is the number of car crashes on sections of highways in California (US). The number of car crashes is discrete defined on which suggests a Poisson distribution with log link, **.** The probability mass function of is given by

An appropriate statistical formula is

**Question 2**

When doing a variable selection search, the full model is regarded as a certain model to start. Then all subsets of the full model will be included in a variable selection search, such as the models only with main effects and the models having main effects plus all pairwise interactions. I will use the function to find best models by AIC and BIC respectively. After ranking AIC and BIC, I will get the best models.

**Question 3**

**Model 1 By AIC**

**Model 2 By AIC**

**Model 3 By AIC**

**Question 4**

**Model 4 By BIC (same as Model 2)**

**Model 5 By BIC**

**Model 6 By BIC**

**Question 5**

The firs focus concerns the mean of response and I wish to select a model that best estimates the expected number of car crashes. The second focus is the probability that the number of car crashes is greater than 8, that is .

**Question 6**

When doing the search with FIC, I will use models having main effects plus the pairwise interactions. Only the intercept is fixed. Because there are the problems with singularities of matrices in the random sample, I have to leave out some interactions from all of the models, such as and

All subsets of this model will be searched.

**Question 7**

**Model 7 By Focus1**

**Model 8 By Focus1**

**Model 9 By Focus1**

**Model 10 By Focus2**

**Model 11 By Focus2**

**Model 12 By Focus2**

**Question 8**

By AIC and BIC, 5 models are selected totally because Model 4 by BIC is same as Model 2 by AIC. These models all have main effects plus some interactions. Model 1 is the most complex model by AIC. As for models selected by FIC, these models are relatively simple and different from models by AIC or BIC. They give similar estimated focuses according to focus 1 and focus 2, respectively. Model 12 is the most complex model selected by FIC.

**Question 9**

studentnumber = 819511

fulldata = read.table("dataHW.txt",header=T)

set.seed(studentnumber)

rownumbers = sample(1:nrow(fulldata),600,replace=F)

mydata = fulldata[rownumbers,]

attach(mydata)

hist(mydata$Y)

library(glmulti)

#AIC

fullmodel=glm(Y~X1+X2+X3+X4+X5,data=mydata,family = poisson(link="log"))

allfits1 <- glmulti(fullmodel,level = 2,crit = "aic")

weightable(allfits1)[1:3,]

#BIC

allfits2 <- glmulti(fullmodel,level = 2,crit = "bic")

weightable(allfits2)[1:3,]

library(fic)

wide.mydata <- glm(Y~.^2-X2:X3-X3:X5,data=mydata,family = poisson(link="log"))

#focus 1

focus1 <- function(par,X) exp((X %\*% par))

inds0 <- c(1,rep(0,13))

combs <- all\_inds(wide.mydata,inds0)

X.all <- model.matrix(wide.mydata)

X <- colMeans(X.all)

fic1 <- fic(wide=wide.mydata,inds = combs,inds0 = inds0,focus = focus1,X=X)

best1 <- order(fic1$rmse.adj)[1:3]

fic1[best1,]

#focus 2

focus2 = function(par, X) 1-ppois(8,lambda=exp(X %\*% par))

fic2 = fic(wide=wide.mydata,inds=combs,inds0=inds0,focus=focus2,X=X.all)

best2 <- order(fic2$rmse.adj)[1:3]

fic2[best2,]