MATH 523: EXAMPLES - POISSON DATA SETS

EXAMPLE: Scottish Lip Cancer data

```
1  ScottishLip<-read.table("ScottishLipCancer.txt")
2  names(ScottishLip)<-c("ID","Y","E","AFF","LAT","LONG")
3
4  glm0<-glm(Y~1,family=poisson,data=ScottishLip)
5  glm0.off<-glm(Y~1+offset(log(E)),family=poisson,data=ScottishLip)
6  glm1.off<-glm(Y~AFF+offset(log(E)),family=poisson,data=ScottishLip)
7  glm2.off<-glm(Y~AFF+LAT+offset(log(E)),family=poisson,data=ScottishLip)
8  glm3.off<-glm(Y~AFF+LAT+LONG+offset(log(E)),family=poisson,data=ScottishLip)
9  glm4.off<-glm(Y~AFF+LAT+AFF:LAT+offset(log(E)),family=poisson,data=ScottishLip)
10
11 anova(glm0.off,glm1.off,glm2.off,glm3.off)
12
13 summary(glm2.off)
14
15 anova(glm0.off,glm1.off,glm2.off,glm4.off)
16 summary(glm4.off)
17
18 glm5.off<-glm(Y~AFF+LAT+AFF:LAT+offset(log(E)),family=quasipoisson,data=ScottishLip)
19 summary(glm5.off)
```

Notes:

- An offset model is used, with offset equal to the Expected value.
- The `anova` function can be used to tabulate the Analysis of Deviance tests.
- The colon symbol in lines 9 and 18 indicates an interaction
- The `family=quasipoisson` command in line 18 uses an estimate of dispersion ϕ, instead of setting $\phi = 1$ to compute the standard errors.

EXAMPLE: Political Affiliations of College Students:

```
1  Y.aff<-c(34,61,16,31,19,17,19,23,16,23,39,12)
2  C.aff<-gl(4,3)
3  A.aff<-gl(3,1,12)
4
5  Affil<-data.frame(Y.aff,C.aff,A.aff)
6  names(Affil)<-c("Y","College","Affiliation")
7
8  glm0.aff<-glm(Y~1,family=poisson,data=Affil)
9  glm1.aff<-glm(Y~C,family=poisson,data=Affil)
10 glm2.aff<-glm(Y~A,family=poisson,data=Affil)
11 glm3.aff<-glm(Y~A+C,family=poisson,data=Affil)
12 glm4.aff<-glm(Y~A*C,family=poisson,data=Affil)
13
14 anova(glm3.aff,glm4.aff)
15 qchisq(0.95,6)
16
17 anova(glm2.aff,glm3.aff)
18 anova(glm1.aff,glm3.aff)
19
20 summary(glm4.aff)
21
22 fitted(glm4.aff)
```

Notes:

- The $*$ in line 12 fits the *full factorial* model (that is, main effects and all interactions), which for the log-linear case is the full or *saturated* model.

EXAMPLE: Classroom behaviour data:

```
1  Y.behav<-c(16,7,15,34,5,3,1,1,3,8,1,3)
2  Y.Dev<-factor(rep(c("ND","D"),each=6))
3  Y.Risk<-factor(rep(c("N","Y"),6))
4  Y.School<-rep(factor(rep(c("Low","Med","High"),each=2)),2)
5
6  Behaviour<-data.frame(Y.behav,Y.Dev,Y.Risk,Y.School)
7  names(Behaviour)<-c("Y","Deviant","AtRisk","Adversity")
8
9  glm7.Behav<-glm(Y~Deviant*AtRisk*Adversity,family=poisson,data=Behaviour)
10
11 anova(glm7.Behav)
12
13 glm6.Behav<-update(glm7.Behav,~.-Deviant:AtRisk:Adversity)
14 summary(glm6.Behav)
15
16 glm5.Behav<-update(glm6.Behav,~.-AtRisk:Adversity)
17 summary(glm5.Behav)
18
19 anova(glm5.Behav,glm6.Behav,glm7.Behav)
```

Notes:

- The `anova` function in line 11 produces a sequence of fits of all the models nested inside the saturated model `glm7.Behav`
- The `update` function allows previously fitted models to be updated by the omission or inclusion of one or more terms.