

228.371 Computer Lab: Generalized Linear Models

Semester One 2015 - Week 9

Instructions: Read each section and try the commands. Then try the Stream worksheet questions suggested to test your knowledge. The worksheet is “adaptive” which means if you get an answer wrong, you can try again. This quiz is to help you monitor your progress, it does not count toward your mark.

Note that because of fonts, especially for symbols like quotation marks, cutting and pasting commands from this document occasionally will not work - you may have to retype.

1 Link Functions

Load the surface-to-air missile data from last week (SurfaceToAir.txt). Fit the logistic regression model, and then use the binomial model with the probit, cauchit, and complementary log-log functions via the following code:

```
surface <- read.table ("Data/SurfaceToAir.txt", header=TRUE)
m1 <- glm(y ~ TargetSpeed, data=surface, family=binomial(link=logit))
m2 <- glm(y ~ TargetSpeed, data=surface, family=binomial(link=probit))
m3 <- glm(y ~ TargetSpeed, data=surface, family=binomial(link=cauchit))
m4 <- glm(y ~ TargetSpeed, data=surface, family=binomial(link=cloglog))
summary(m1)                ## can be used for the other models as well
anova(m1, test="Chisq")    ## can be used for the other models as well
```

You can picture the differences between the fitted models with the following plot:

```
attach(surface)
ord <- order(TargetSpeed) #this will put the points in order
plot(TargetSpeed[ord], m1$fitted[ord], type='l')
lines(TargetSpeed[ord], m2$fitted[ord], col=2)
lines(TargetSpeed[ord], m3$fitted[ord], col=3)
lines(TargetSpeed[ord], m4$fitted[ord], col=4)
legend("topright", lty=rep(1,4), col=c(1:4),
legend=c("logit", "probit", "cauchit", "cloglog"))
```

Now answer questions 1-3 on the worksheet.

2 Poisson Regression

Load the popcorn data. The response is the number of inedible kernels; the predictors are the temperature, the amount of oil used, and the cooking time. Fit the additive model fitting the main effects as factors:

```
pop <- read.table ("Data/popcorn.txt", header=TRUE)
m1 <- glm(Inedible ~ factor(Temperature)+factor(Oil)+factor(Time), data=pop, family=poisson)
summary(m1)
anova(m1, test="Chisq")
```

Based on this output (and other commands as you see fit) find the best single factor model (call it `m2`). Make the diagnostic plots. Consider whether this model fits well - if it is, the following test should be non-significant.

```
pchisq(m2$deviance, m2$df.residual, lower=FALSE)
```

Now answer questions 4-7 on the worksheet.