

GLMs

Fewer hints this time remember to set your working directory and read in the data we're going to look at a dataset called roadkills, this is number of frogs run over on roads. It is a .txt file so you will have to use

```
read.table('RoadKills.txt', header = T)
```

so lets take a look at the data to see how it looks

```
hist(roadkill$BufoCalamita)
```

```
hist(roadkill$OPEN.L)
```

so say we are interested in the effect of open land on the number of frogs run over

shall we try plotting it first to see if there is a trend we might be interested in

```
plot(roadkill$BufoCalamita ~ roadkill$OPEN.L)
```

excluding some funky very high dead frogs it looks like there might be a slight decrease

lets try our first GLM, our dead frogs was poisson distributed so lets use a poisson glm

```
glm1 <- glm(BufoCalamita ~ OPEN.L, data = roadkill, family = poisson())
```

```
summary(glm1)
```

and we can get diagnostic plots from it

```
plot(glm1)
```

this model isn't great but not totally terrible either, this is kind of realistic in what you will get

look at that we have a significant effect, but there are more things that might be effecting dead frogs, lets make a larger model.

here we are looking at the effect of open land, and shrub density and road density.

```
glm2 <- glm(BufoCalamita ~ OPEN.L * SHRUB * L.D.ROAD, roadkill, family = poisson())  
summary(glm2)
```

ok so we have some significant effects here but not all, so lets try removing non significant ones from the model

well open L is significant, ut so is the interaction between that road density so we need those things in

```
glm3 <- glm(BufoCalamita ~ OPEN.L * L.D.ROAD, roadkill, family = poisson())  
summary(glm3)
```

well it looks like it might have been improved but is the model actually better?

lets compare

```
anova(glm2, glm3, test = 'Chisq')
```

Well glm3 is significantly different than glm2 but it has a higher residual deviance meaning it is actually a worse model!

ok so we have our statistics done shall we plot our effect in a proper way?

```
plot(roadkill$BufoCalamita ~ roadkill$OPEN.L)
```

so how do we get a line one there?

this is quite tricky

here we are making a new sequence of possible OPEN L values

```
pred <- expand.grid(OPEN.L = seq(0,100, by = 1))
```

we are then using a model to get predicted values for them

```
pred$fit <- predict(glm1, newdata= pred, type='response')
```

and using those to plot a line

```
lines(fit ~ OPEN.L, data = pred)
```

ok lets try doing one with a binomial distribution. read in the gala.txt data

take a look at the data first

so say I want us to investigate what effects the proportion of endemic species?

we don't have that data but we do have number of endemics and number of species

we can make a new column by referencing a column that doesn't exist

```
gala$PropEnd <- gala$Endemics/gala$Species  
hist(gala$PropEnd)
```

as a proportion is bounded between 0 and 1 and can take a lot of weird distributions within that it is definitely not normal

fortunately, we have a glm family for this!

lets investigate whether area or elevation are having an effect on proportion of endemics

```
glm4<- glm(PropEnd ~ Area * Elevation, data = gala, weights = Species , family = binomial())
```

we are not using 0 or 1 success trials so we have to specify a weights, essentially the total number of trials in this case the number of species.

```
summary(glm4)
```

Ok so we have significant things here and some not

lets try using the drop1 command

```
drop1(glm4, test = 'Chisq')
```

this checks the significance of your interaction, it says this is significant so in reality we can't drop any terms. But we could do it on our previous glm on the roadkills.

```
drop1(glm2, test = 'Chisq')
```

yep its telling you to drop that interaction which we did ourselves. drop1 is useful to know which way it is best to simplify the model.

lets go back to our mammal data set

do a lm to see if order has an effect on adult body mass

```
MammalLM <- lm(AdultBodyMass_g ~ Order, MammalData)
summary(MammalLM)
```

ok so we have an lm with a factor with multiple levels in there, and we want to see what the difference between the levels are.

for this we are going to need a new package (there is now a tukey function built into base R but we need to practice new packages)

this installs the package, you only need to do this once after that you have it forever.

```
install.packages('multcomp')
```

this loads the package in

```
library(multcomp)
```

ok now we have a package to do what we want, in R there are packages for pretty much everything you can imagine.

```
tukey <- glht(MammalLM, linfct = mcp(Order = 'Tukey'))
```

this command does a tukey test now we just look at it to see what is significant

```
summary(tukey)
```

this spits out a lot of things and takes a long time but is what we want.

the confint and plot commands are incredibly useful for visualising where there is significance.

```
confint(tukey)
```

```
plot(tukey)
```

however in this case we have so many levels it's hard to see what is what

Ok now load in the galliformesData.

Investigate what is having an impact on Clutch size in galliformes.