# STATS 201 Assignment 4

#### **Model Answers**

## Loading required package: s20x

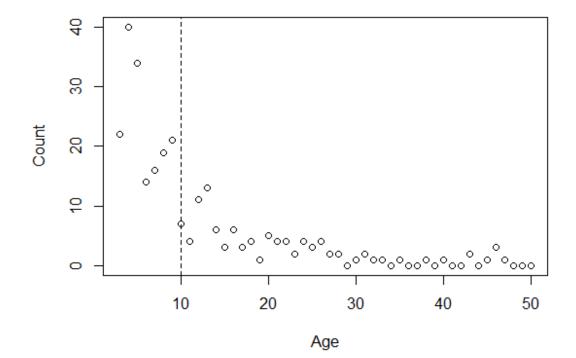
### **Question 1**

### **Questions of Interest**

The questions we want answered are: how does the population of geoducks change with age, did the storm event have any impact on the population, and did the relationship between age and the population count change after the storm?

## Read in and inspect the data:

```
Geoduck.df=read.table("Geoduck.txt",header=T)
plot(Count~Age,data=Geoduck.df)
abline(v=10,lty=2)
```

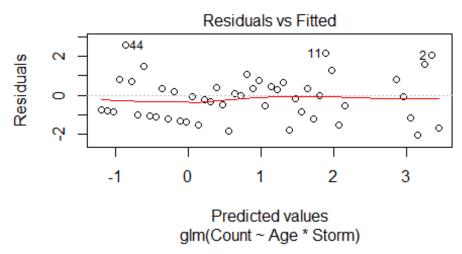


### Comment on plot(s) and/or summary statistics

The scatter plot shows an exponental (i.e., multiplicative) decline in geoduck count with age. There is some indication of a storm effect at age 10, but it is hard to be sure.

#### Fit and check model

```
duckfit1=glm(Count~Age*Storm,family=poisson,data=Geoduck.df)
plot(duckfit1,which=1)
```



```
summary(duckfit1)
##
## Call:
## glm(formula = Count ~ Age * Storm, family = poisson, data = Geoduck.df)
##
## Deviance Residuals:
##
       Min
                 10
                      Median
                                    3Q
                                            Max
           -1.0981
                     -0.1192
## -2.0799
                                0.5128
                                         2.5695
##
## Coefficients:
##
               Estimate Std. Error z value Pr(>|z|)
                                    15.913
                                              <2e-16 ***
                3.72147
                           0.23386
## (Intercept)
## Age
               -0.09560
                           0.03925
                                     -2.436
                                              0.0149 *
               -0.74423
                           0.33266
                                     -2.237
                                              0.0253 *
## Storm
## Age:Storm
                0.01216
                           0.04068
                                      0.299
                                              0.7650
                   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Signif. codes:
## (Dispersion parameter for poisson family taken to be 1)
##
       Null deviance: 458.589
                               on 47
                                       degrees of freedom
##
## Residual deviance: 60.985 on 44 degrees of freedom
```

```
## AIC: 189.51
##
## Number of Fisher Scoring iterations: 5
1-pchisq(60.985,44)
## [1] 0.04567957
duckfit2=glm(Count~Age*Storm,family=quasipoisson,data=Geoduck.df)
summary(duckfit2)
##
## Call:
## glm(formula = Count ~ Age * Storm, family = quasipoisson, data =
Geoduck.df)
##
## Deviance Residuals:
                      Median
                                   3Q
                                           Max
       Min
                 1Q
## -2.0799 -1.0981 -0.1192
                               0.5128
                                        2.5695
##
## Coefficients:
               Estimate Std. Error t value Pr(>|t|)
                                             <2e-16 ***
## (Intercept) 3.72147
                           0.28203 13.195
## Age
               -0.09560
                           0.04734 -2.020
                                             0.0495 *
               -0.74423
                           0.40118 -1.855
                                             0.0703 .
## Storm
## Age:Storm
               0.01216
                           0.04906
                                     0.248
                                             0.8054
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 1.454374)
##
##
       Null deviance: 458.589 on 47 degrees of freedom
## Residual deviance: 60.985 on 44 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
duckfit3=glm(Count~Age+Storm, family=quasipoisson, data=Geoduck.df)
summary(duckfit3)
##
## Call:
## glm(formula = Count ~ Age + Storm, family = quasipoisson, data =
Geoduck.df)
##
## Deviance Residuals:
       Min
                 1Q
                      Median
                                   3Q
                                           Max
##
                               0.5151
## -2.0980 -1.0881
                    -0.1776
                                        2.5915
##
## Coefficients:
               Estimate Std. Error t value Pr(>|t|)
##
```

```
## (Intercept) 3.65761
                          0.11672 31.338 < 2e-16 ***
                           0.01243 -6.782 2.15e-08 ***
## Age
               -0.08428
                          0.23351 -2.841 0.00673 **
## Storm
               -0.66339
## ---
                  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Signif. codes:
##
## (Dispersion parameter for quasipoisson family taken to be 1.438666)
##
##
       Null deviance: 458.589 on 47 degrees of freedom
## Residual deviance: 61.075 on 45
                                     degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
exp(confint(duckfit3))
## Waiting for profiling to be done...
##
                    2.5 %
                              97.5 %
## (Intercept) 30.7695824 48.6534515
               0.8961418 0.9409958
## Age
## Storm
               0.3233277 0.8086895
100*(exp(confint(duckfit3))[2:3,]-1)
## Waiting for profiling to be done...
             2.5 %
                      97.5 %
##
         -10.38582
## Age
                   -5.900423
## Storm -67.66723 -19.131046
```

### **Method and Assumption Checks**

A Poisson log-linear model was fitted to the geoduck counts. The explanatory variables were Age and an indicator variable for the storm event. The initial model included interaction. Residuals looked fine but there was evidence that over–dispersion was present (P-value = 0.0457) so the model was refitted as quasipoisson, The interaction term was not significant and so was removed. After this, both Age and Storm were significant so the model was not simplified any further.

The Age and Storm effects were significant and the final model was

$$\log(\mu_i) = \beta_0 + \beta_1 \times age_i + \beta_2 \times Storm_i$$

where  $Count_i$  is the count of geoduck of age i geoduck and has an overdispersed distribution with mean  $\mu_i$ . (Note: this is not a Poisson distribution in this case as we used a QuasiPoisson model.)  $Storm_i$  is an indicator variable that takes the value 0 if age is less than 10 years (before the storm) and 1 if 10 or more. Note: could use  $E[Count_i]$  for  $\mu_i$  as have the same definition.

### **Executive Summary**

We want to model the population of geoducks from the west coast of the North Island taking into account their age and an extreme storm event 10 years prior to the sample being taken.

Both age and the storm event have a significant effect on expected count of geoduck. There was no evidence that these effects interacted, i.e., no evidence that the shape of the relationship between age and the expected count of geoducks differed for geoducks born before and after the storm.

For every year increase in age (notwithstanding the storm event), the expected count of geoduck decreases by between 5.9% and 10.4%.

For any age class of geoducks old enough to be effected by the storm, there was a further decrease in their expected count of between 19% and 68%.

### **Question 2**

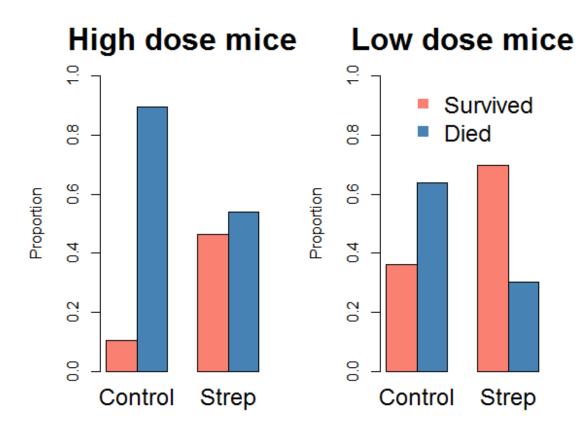
#### **Questions of interest**

Is Streptomycin an effective radiation treatment? If so, does the effectiveness depend on the level of radiation dose?

### Read in and inspect the data:

```
mice.df <- read.csv("mice.csv")</pre>
mice.df
    Treatment Dose Died Survived n
##
                                         Propn
## 1
                     95 11 106 0.1037736
      Control high
## 2
      Control low
                     99
                              56 155 0.3612903
## 3
        Strep high
                     57
                              49 106 0.4622642
## 4
        Strep low
                     47
                             108 155 0.6967742
par(mfrow=c(1,2))
barplot(matrix(c(mice.df$Propn, 1-mice.df$Propn)[mice.df$Dose=="high"],
byrow=T, nrow=2), col=c("salmon", "steelblue"), beside=T, ylim=c(0, 1),
main="High dose mice",
names=mice.df$Treatment[mice.df$Dose=="high"],ylab="Proportion",
cex.names=1.5, cex.main=2)
barplot(matrix(c(mice.df$Propn, 1-mice.df$Propn)[mice.df$Dose=="low"],
byrow=T, nrow=2), col=c("salmon", "steelblue"), beside=T, ylim=c(0, 1),
main="Low dose mice", names=mice.df$Treatment[mice.df$Dose=="low"],
ylab="Proportion",cex.names=1.5, cex.main=2)
```





## Comment on plot(s) and/or summary statistics

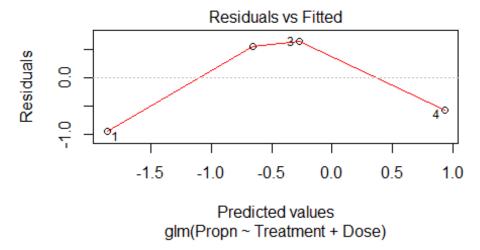
We can see that the proportion of mice that survived was higher in the Streptomycin treated groups compared to the control group and that this was true for both high and low doses of radiation. The survival rate also seems better for both treament groups when lower doses of radiation were used. The barplots doesn't show a strong association between treatment and radiation dosage.

#### Fit and check model

```
mice.bino <-
glm(Propn~Treatment*Dose,data=mice.df,weights=n,family="binomial")
anova(mice.bino,test="Chisq")

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Propn</pre>
```

```
##
## Terms added sequentially (first to last)
##
##
                  Df Deviance Resid. Df Resid. Dev
##
                                                    Pr(>Chi)
## NULL
                                      3
                                           103.122
                                      2
## Treatment
                   1
                       64.813
                                            38.309 8.236e-16 ***
## Dose
                       36.407
                                      1
                                             1.903 1.602e-09 ***
## Treatment:Dose 1
                       1.903
                                      0
                                             0.000
                                                      0.1678
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
mice.bino2 <-
glm(Propn~Treatment+Dose, data=mice.df, weights=n, family="binomial")
summary(mice.bino2)
##
## Call:
## glm(formula = Propn ~ Treatment + Dose, family = "binomial",
       data = mice.df, weights = n)
##
##
## Deviance Residuals:
                           3
## -0.9461
             0.5406
                      0.6268 -0.5677
##
## Coefficients:
##
                  Estimate Std. Error z value Pr(>|z|)
                               0.2113 -8.832 < 2e-16 ***
## (Intercept)
                  -1.8660
                                        7.941 2.01e-15 ***
## TreatmentStrep 1.5924
                               0.2005
## Doselow
                    1.2054
                               0.2071
                                        5.821 5.85e-09 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Dispersion parameter for binomial family taken to be 1)
##
       Null deviance: 103.1219 on 3 degrees of freedom
## Residual deviance:
                        1.9026
                                on 1 degrees of freedom
## AIC: 27.907
## Number of Fisher Scoring iterations: 4
1-pchisq(1.9026,1)
## [1] 0.1677876
plot(mice.bino2, which=1)
```



```
confint(mice.bino2)
## Waiting for profiling to be done...
##
                       2.5 %
                                 97.5 %
## (Intercept)
                  -2.2932780 -1.464011
## TreatmentStrep
                  1.2050074
                              1.992005
## Doselow
                   0.8052496
                              1.618074
exp(confint(mice.bino2))
## Waiting for profiling to be done...
##
                      2.5 %
                                97.5 %
## (Intercept)
                  0.1009351 0.2313066
## TreatmentStrep 3.3367839 7.3302136
## Doselow
                  2.2372548 5.0433667
```

## Model building and assumption checking

We have a response variable with two outcomes, survived or died, so have fitted a logistic regression model. We have two categorical (grouping) explanatory variables so have fitted these along with their interaction. The interaction was not significant (P-value=0.17), so we dropped the interaction term and just have a main effects model. The check for residual deviance has a P-value of 0.168 and hence the Binomial assumption is adequate. The residual plot shows no problem (which with only 4 points is not at all surprising).

```
Our model is log(Odds_i) = \beta_0 + \beta_1 \times Doselow_i + \beta_2 \times TreatmentStrept_i
```

where the odds are the odds of a mouse with given value of explanatory variables surviving.  $Doselow_i$  is 1 if a mouse in the ith group got a low dose of radiation and 0 otherwise, and  $TreatmentStrept_i$  is 0 if a mouse was in the control group and 1 if the mouse was in the group treated with Streptomycin.

Alternatively:  $Y_i$  is the number of mice in the *ith* group that survived,  $n_i$  is the total number of mice in the *ith* group,  $Y_i \sim Binomial(n_i, p_i)$ ,  $odds_i = \frac{p_i}{(1-p_i)}$ 

Alternatively: Our model is  $\log(Odds_{ij}) = \alpha_i + \beta_j$  where  $\alpha_i$  = the effect of the ith dose given to a group and  $\beta_i$  = the effect of the jth treatment given to a group.

### **Executive summary**

An experiment was conducted to investigate the utility of Streptomycin as a means of treating radiated mice.

There was no evidence that the effectiveness of the streptomycin depended on the dosage level of the radiation.

Mice treated with streptomycin are more likely to survive. For mice with a given dose of radiation, we estimate the odds that a mouse that was treated with streptomycin survived the first 10 days after exposure are between 3.3 and 7.3 times the odds that a mouse that wasn't treated survived.

Not surprisingly, it was found that mice subjected to lower doses of radiation were more likely to survive. For mice with a specified treatment, we estimate that the odds of a mouse that was exposed to a low dose of radiation surviving the first 10 days after exposure are between 2.2 and 5.0 times of the odds of a mouse exposed to high levels of radiation surviving.