STATS 201 Assignment 4

LiYuzhuo-2019210146

Due Date: 19/12/2021

## Loading required package: s20x

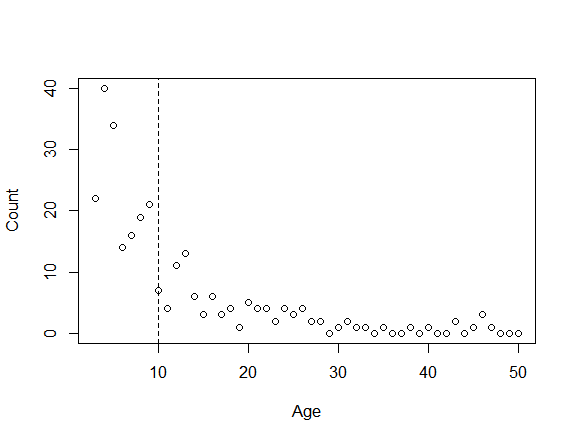
# Question 1

## 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)

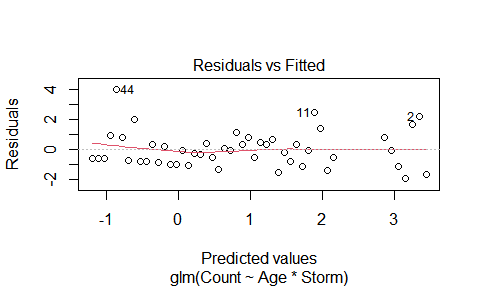


## Comment on plot(s) and/or summary statistics

There appears to be a downward curvilinear trend in the relationship between age and the number of geoducks in that age group. The number of geoducks decreased considerably after the storm. The number of geoducks between the ages of 20 and 50 seems to be practically constant.

## Fit and check model

geoduck.fit1 = glm(Count ~ Age \* Storm, poisson(link="log"), Geoduck.df)  
plot(geoduck.fit1, which = 1)



1 - pchisq(60.985,44)

## [1] 0.04567957

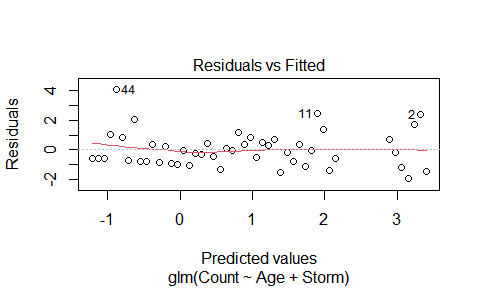
summary(geoduck.fit1)

##   
## Call:  
## glm(formula = Count ~ Age \* Storm, family = poisson(link = "log"),   
## data = Geoduck.df)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.0799 -1.0981 -0.1192 0.5128 2.5695   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 3.72147 0.23386 15.913 <2e-16 \*\*\*  
## Age -0.09560 0.03925 -2.436 0.0149 \*   
## Storm -0.74423 0.33266 -2.237 0.0253 \*   
## Age:Storm 0.01216 0.04068 0.299 0.7650   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (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

geoduck.fit2 = glm(Count ~ Age \* Storm, quasipoisson(link="log"), Geoduck.df)  
summary(geoduck.fit2)

##   
## Call:  
## glm(formula = Count ~ Age \* Storm, family = quasipoisson(link = "log"),   
## data = Geoduck.df)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.0799 -1.0981 -0.1192 0.5128 2.5695   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.72147 0.28203 13.195 <2e-16 \*\*\*  
## Age -0.09560 0.04734 -2.020 0.0495 \*   
## Storm -0.74423 0.40118 -1.855 0.0703 .   
## 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

geoduck.fit3 = glm(Count ~ Age + Storm, quasipoisson(link="log"), Geoduck.df)  
plot(geoduck.fit3, which = 1)



1 - pchisq(61.075, 45)

## [1] 0.0553605

summary(geoduck.fit3)

##   
## Call:  
## glm(formula = Count ~ Age + Storm, family = quasipoisson(link = "log"),   
## data = Geoduck.df)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.0980 -1.0881 -0.1776 0.5151 2.5915   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.65761 0.11672 31.338 < 2e-16 \*\*\*  
## Age -0.08428 0.01243 -6.782 2.15e-08 \*\*\*  
## Storm -0.66339 0.23351 -2.841 0.00673 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (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

confint(geoduck.fit3)

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 3.4265266 3.88472275  
## Age -0.1096566 -0.06081664  
## Storm -1.1290889 -0.21234019

100\*(1-exp(confint(geoduck.fit3)))

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) -2976.95824 -4765.345148  
## Age 10.38582 5.900423  
## Storm 67.66723 19.131046

## Method and Assumption Checks

The number of geoducks suggests that we can use a generalized linear model (GLM) with a Poisson response distribution for the expected counts at each age. We modified the first model slightly to include the interaction effect of the storm.

The deviation residual plots looks good. The interaction term is not significant, however, the test for overdispersion is significant, so a Poisson-like response distribution is used and the interaction term still not significant in the quasipossion model.

Although the overdispersion test shows that we have significant results , and thus can continue with the assumption that the dispersion ratio is 1, we have decided to remain with the quaispoisson response distribution for greater confidence in our estimated values.

Our model is:

is 1 if Age is 10 or greater, and 0 if otherwise.

## Executive Summary

We are interested in modelling the population of geoducks from the west coast of the North Island taking into account their age and an extreme storm event 10 years before.

We have a strong evidence that there is a decreasing relationship between age and the expected population of geoducks .

And for a 1-year increase in age, the population count of geoducks decreases by between 5.9% and 10.4%, on average.

We also have a strong evidence that storm had a negetive effect on the population count of geoducks .

The storm lowered the population count of geoducks by between 19.1% and 67.7%, on average.

And there is no interaction between storm and age.

# Question 2

## Questions of interest

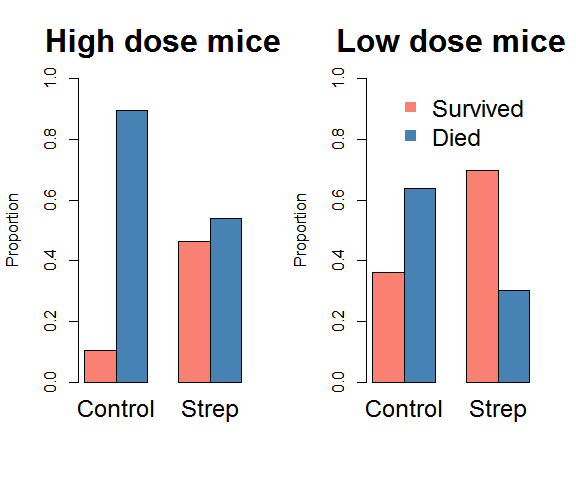
Is Streptomycin an effective radiation treatment? If so, does the effectiveness depend on the level of radiation dose?

## Read in and inspect the data:

mice.df <- read.csv("mice.csv")  
  
mice.df

## Treatment Dose Died Survived n Propn  
## 1 Control high 95 11 106 0.1037736  
## 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)  
  
legend("topright",c("Survived", "Died"),pch=15,col=c("salmon","steelblue"),bty="n", cex=1.5)

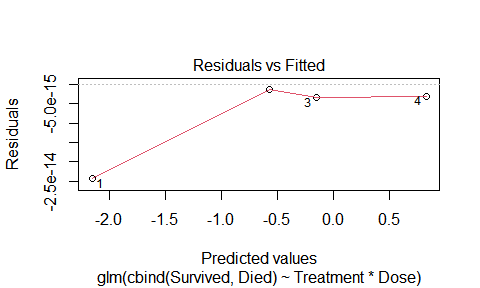


## Comment on plot(s) and/or summary statistics

We can see that in the two control groups, the died mice are more than survived mice, and for the mice get a high dose of radiation, the number of died mice and survived mice basically the same. For the mice get a low dose of radiation, the number of survived mice is more than died mice.

## Fit and check model

mice.fit1 = glm(cbind(Survived,Died) ~ Treatment \* Dose, family = binomial(link="logit"), mice.df)  
plot(mice.fit1, which = 1)



anova(mice.fit1, test="Chisq")

## Analysis of Deviance Table  
##   
## Model: binomial, link: logit  
##   
## Response: cbind(Survived, Died)  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)   
## NULL 3 103.122   
## Treatment 1 64.813 2 38.309 8.236e-16 \*\*\*  
## Dose 1 36.407 1 1.903 1.602e-09 \*\*\*  
## Treatment:Dose 1 1.903 0 0.000 0.1678   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(mice.fit1)

##   
## Call:  
## glm(formula = cbind(Survived, Died) ~ Treatment \* Dose, family = binomial(link = "logit"),   
## data = mice.df)  
##   
## Deviance Residuals:   
## [1] 0 0 0 0  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.1560 0.3185 -6.769 1.29e-11 \*\*\*  
## TreatmentStrep 2.0048 0.3733 5.370 7.89e-08 \*\*\*  
## Doselow 1.5862 0.3597 4.410 1.04e-05 \*\*\*  
## TreatmentStrep:Doselow -0.6030 0.4448 -1.356 0.175   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1.0312e+02 on 3 degrees of freedom  
## Residual deviance: 3.8192e-14 on 0 degrees of freedom  
## AIC: 28.005  
##   
## Number of Fisher Scoring iterations: 3

mice.fit2 = glm(cbind(Survived,Died) ~ Treatment + Dose, family = binomial(link="logit"), mice.df)  
anova(mice.fit2, test="Chisq")

## Analysis of Deviance Table  
##   
## Model: binomial, link: logit  
##   
## Response: cbind(Survived, Died)  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)   
## NULL 3 103.122   
## Treatment 1 64.813 2 38.309 8.236e-16 \*\*\*  
## Dose 1 36.407 1 1.903 1.602e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(mice.fit2)

##   
## Call:  
## glm(formula = cbind(Survived, Died) ~ Treatment + Dose, family = binomial(link = "logit"),   
## data = mice.df)  
##   
## Deviance Residuals:   
## 1 2 3 4   
## -0.9461 0.5406 0.6268 -0.5677   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.8660 0.2113 -8.832 < 2e-16 \*\*\*  
## TreatmentStrep 1.5924 0.2005 7.941 2.01e-15 \*\*\*  
## 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

exp(confint(mice.fit2))

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 0.100935 0.2313066  
## TreatmentStrep 3.336784 7.3302128  
## Doselow 2.237255 5.0433677

## Model building and assumption checking

Our data was composed of two categorical predictors： Treatment and Dose . Our response was the log odds of surviving. The properties of our data meant that a logistic regression was chosen to model it. We initially fit an model with interaction items using the logit link to transform the linear predictor to give us the log odds of surviving. The deviance residual check shows residuals of zero, since all four observations have been used to fit for our four parameters. So this is not very helpful.

Results of the anova analysis shows that the interaction of Treatment and Dose is non-significant. We then simplified our model to a main effects only model. Testing for the suitability of a quasibinomial dsitribution returned non-significant results hence we remained with the binomial distribution.

Our final model is:

* is 1 if the th mouse was in the Strep group, and 0 otherwise.
* is 1 if the th mouse was in the low Dose group, and 0 otherwise.
* is the odds of the th mouse surviving.

## Executive summary

Our analysis shows that Streptomycin is an effective treatment for radiation. We found that keeping radiation dose constant, the odds of survival for mice that receive streptomycin are 3.3 to 7.3 times greater than mice in the control group.

However, Strepomycin did not have an effect on the relationship between dose and odds of survival . Streptomycin’s effectiveness did not depend on the level of radiation dose. Keeping treatment constant, we found the odds of survival for mice that assigned a low dose of radiation are 2.2 to 5 times greater than mice that assigned a high dose.