Stats*3510 - Assignment 3
Graham Eckel – 0679576

Q1
library(gmodels)
dir = "E:\\Google Drive\\...\\Assignment 3\\"
file1 = "Elephant_Data.csv"
dfElephants = read.table(file=paste(dir,file1, sep=""), header=TRUE, sep=',')

a)
p.model=glm(Matings~Age,family=poisson, data = dfElephants)
summary(p.model)

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) -1.58201    0.54462  -2.905  0.00368 ** 
Age          0.06869    0.01375   4.997 5.81e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 75.372  on 40  degrees of freedom
Residual deviance: 51.012  on 39  degrees of freedom
AIC: 156.46

Number of Fisher Scoring iterations: 5
```

There is strong evidence (p<5.81e-07) against the null that the mean number of successful matings and age are unrelated. As the age of the elephant increases by 1 unit, the estimated number of successful matings is multiplied by e^Bhat1 = 1.071104.

b)
ci(p.model)

```
               Estimate    CI lower    CI upper Std. Error      p-value
(Intercept) -1.58200796 -2.68360855 -0.48040737 0.54462132 3.675052e-03
Age          0.06869281  0.04088935  0.09649627 0.01374578 5.811590e-07
```

We knew from the p-value of our model that 0 would not be in our 95% CI for the estimated B1 and so the null hypothesis B1 = 0 is not a plausible.

c)
pchisq(51.012, 39, lower.tail = F)
p = 0.094256

Ho: Our model is an adequate fit
Ha: Our mode is not an adequate fit

With a p value of 0.094256 from the chi squared distribution we do not have strong enough evidence to reject the null hypothesis. Our model is an adequate fit.

d)
l.model=lm(Matings~Age, data = dfElephants)
summary(l.model)

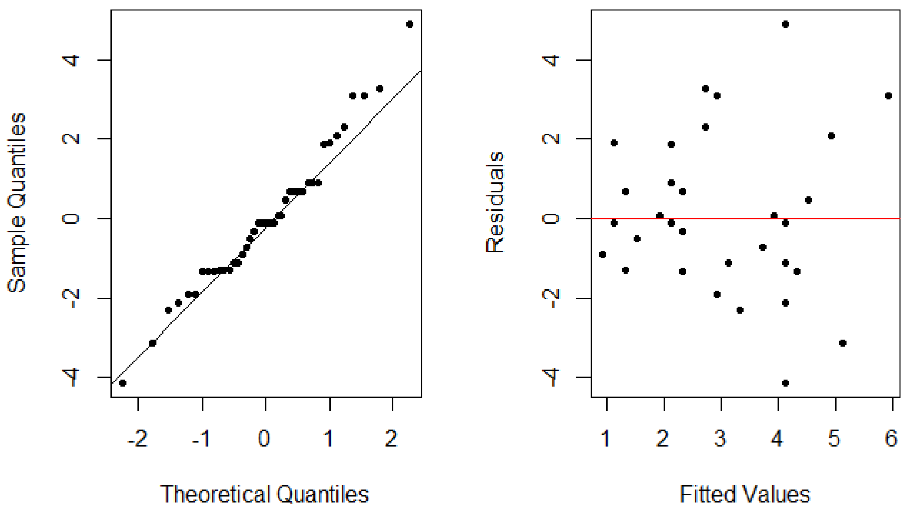
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -4.50589    1.61899  -2.783  0.00826 ** 
Age          0.20050    0.04443   4.513 5.75e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.849 on 39 degrees of freedom
Multiple R-squared:  0.343,	Adjusted R-squared:  0.3262 
F-statistic: 20.36 on 1 and 39 DF,  p-value: 5.749e-05
```

There is strong evidence (p<5.75e-05) against the null that the mean number of successful matings and age are unrelated. As the age of the elephant increases by 1 unit, the estimated number of successful matings increases by 0.2005. The estimated slope of both models is positive while the linear regression model’s estimated slope is greater than that estimated by the Poisson model.

e)

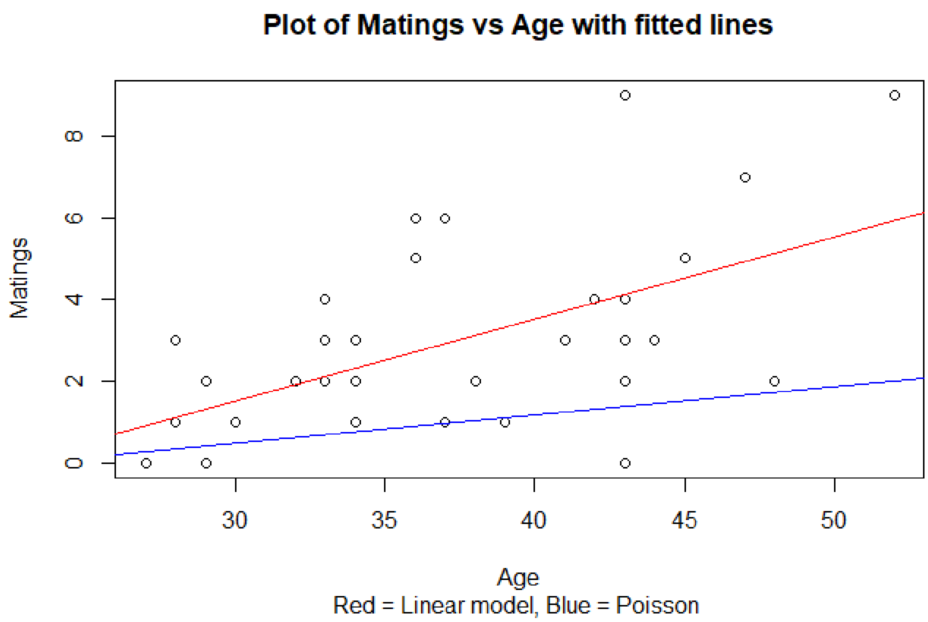
```
par(mfrow=c(1,2))
qqnorm(l.model$residuals, pch = 20, main = " ")
qqline(l.model$residuals)
plot(l.model$fitted.values, l.model$residuals, pch = 20,
    xlab = "Fitted Values", ylab = "Residuals")
abline(h = 0, col = "red")
```



The assumption of normality holds as we have a relatively straight line through the qq plot. In the plot of residuals versus fitted values we seem a relatively even dispersion across 0, so our assumption of linearity holds. Lastly, the assumption of constant variance appears to hold as the residuals evenly spread across the y’s.

f)

```
plot(dfElephants$Age, dfElephants$Matings)
abline(p.model, col = "blue")
abline(l.model, col = "red")
```



Visually, the linear model (red) appears to fit the model better. By inspection, it splits the data more evenly than the poisson model (blue).

g)
yhat = -4.50589 +(37)(0.2005) = 2.9126

Using the linear model, the predicted mean number of mating pairs for elephants that are 37 years old is 2.9126.