DATA621 Extended LMR Ex 3.2

Chun Yip

2022/4/3

R Markdown

The salmonella data was collected in a salmonella reverse mutagenicity assay. The predictor is the dose level of quinoline and the response is the numbers of revertant colonies of TA98 salmonella observed on each of three replicate plates. Show that a Poisson GLM is inadequate and that some overdispersion must be allowed for. Do not forget to check out other reasons for a high deviance.

```
data(salmonella, package="faraway")
head(salmonella)
```

```
##   colonies dose
## 1       15    0
## 2       21    0
## 3       29    0
## 4       16   10
## 5       18   10
## 6       21   10
```

```
summary(salmonella)
```

```
##     colonies          dose
##  Min.   :15.00   Min.   :   0.0
##  1st Qu.:20.25   1st Qu.:  10.0
##  Median :27.00   Median :  66.5
##  Mean   :29.11   Mean   : 246.0
##  3rd Qu.:36.75   3rd Qu.: 333.0
##  Max.   :60.00   Max.   :1000.0
```

I always started with linear model as the reference. Poisson model is also shown below.

```
lmod <- lm(colonies ~ dose, salmonella)
summary(lmod)
```

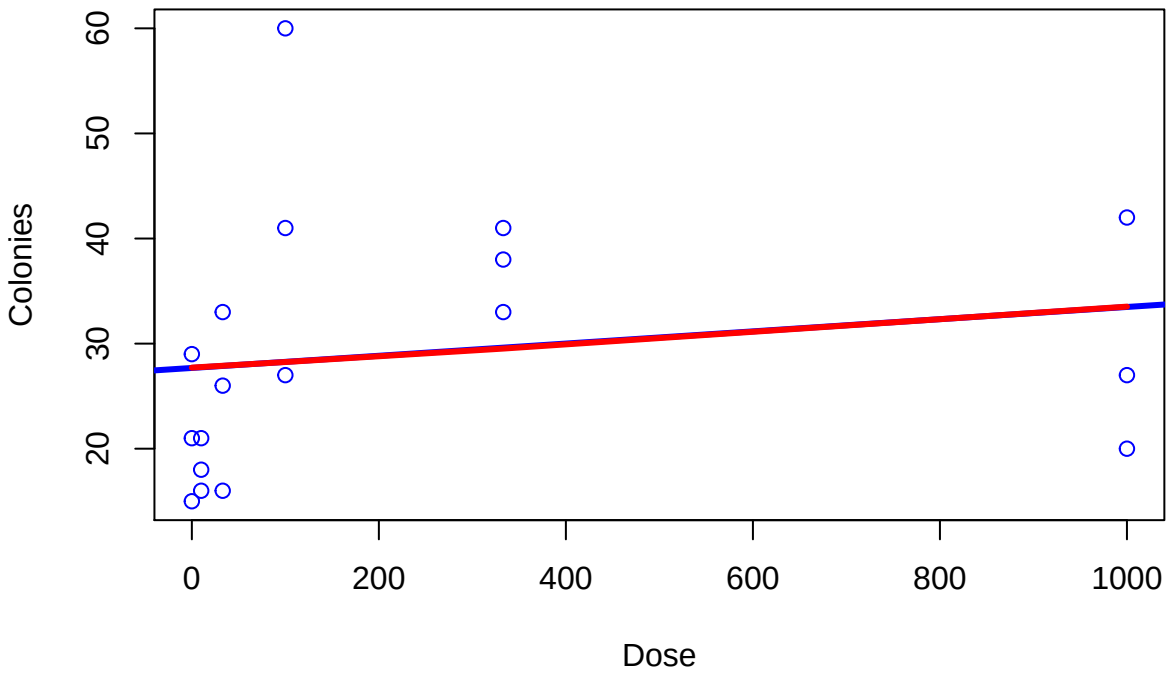
```
##
## Call:
## lm(formula = colonies ~ dose, data = salmonella)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -13.488  -8.991  -1.569   7.569  31.736
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 27.683154   3.462169   7.996 5.58e-07 ***
## dose         0.005805   0.008006   0.725    0.479
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.08 on 16 degrees of freedom
## Multiple R-squared:  0.03181,    Adjusted R-squared:  -0.0287
## F-statistic: 0.5257 on 1 and 16 DF,  p-value: 0.4789
```

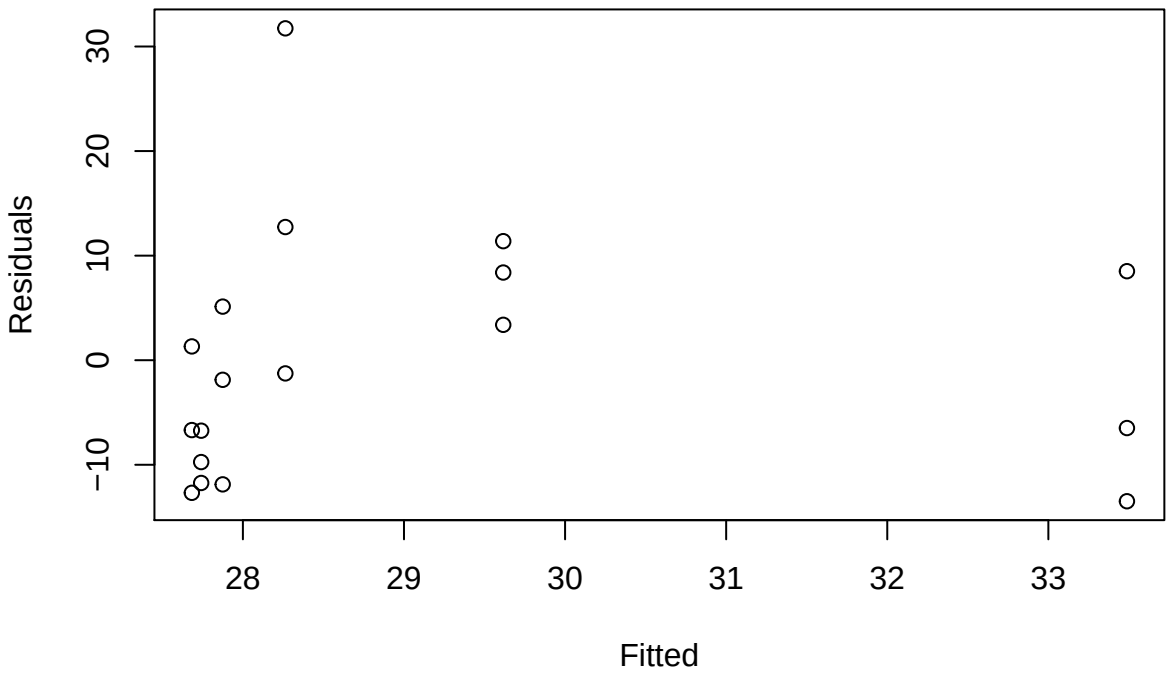
```
modp <- glm(colonies ~ dose, family="poisson", salmonella)
summary(modp)
```

```
##
## Call:
## glm(formula = colonies ~ dose, family = "poisson", data = salmonella)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6482  -1.8225  -0.2993   1.2917   5.1861
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 3.3219950  0.0540292  61.485   <2e-16 ***
## dose        0.0001901  0.0001172   1.622    0.105
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 78.358  on 17  degrees of freedom
## Residual deviance: 75.806  on 16  degrees of freedom
## AIC: 172.34
##
## Number of Fisher Scoring iterations: 4
```

```
plot(salmonella$dose, salmonella$colonies, col="blue", xlab="Dose", ylab="Colonies")
abline(lmod, col="blue", lwd=3); lines(salmonella$dose, modp$fitted, col="red", lwd=3)
```




```
plot(predict(lmod), residuals(lmod), xlab="Fitted", ylab = "Residuals")
```



The poisson distribution is very similar to linear regression. Both models don't really fit the data. The assumption behind poisson is for small success probabilities and large totals. At low value of Dose, there are several colonies. There are less colonies at high dose.