```R
# toxo.R

> toxo <- read.table(file="N:\\courses\\stat8620\\Fall 08\\toxo.dat",header=T)
> #toxo <- read.table(file="C:\\Documents and Settings\\dhall\\My Documents\\Dan's Work Stuff\\courses\\STAT8620\\Fall 08\\toxo.dat",header=T)
> toxo$rain1000 <- toxo$rainf/1000
> toxo$ypos <- round(toxo$ppos*toxo$n)
> toxo$yneg <- toxo$n-toxo$ypos
> toxo$samplogit <- log((toxo$ypos+0.5)/(toxo$n-toxo$ypos+0.5))
> toxo[1:3,]

  rainf   ppos   n   rain1000   ypos   yneg  samplogit
1  1735  0.500  4  1.735      2  2  0.0000000
2  1800  0.600  5  1.800      3  2  0.3364722
3  2050  0.292 24  2.050      7 17 -0.8472979

> plot(toxo$rain1000,toxo$ppos,main="Prop positive versus rainfall (in 1000's)")
```

Prop positive versus rainfall (in 1000's)
> plot(toxo$rain1000, toxo$samplogit, main="Samp log odds positive versus rainfall (in 1000's)"

Samp log odds positive versus rainfall (in 1000's)

> m1 <- glm(cbind(ypos, yneg) ~ poly(rain1000, 5), data = toxo,  
+   family = binomial(link = "logit"))
> summary(m1)

Call:
  glm(formula = cbind(ypos, yneg) ~ poly(rain1000, 5), family = binomial(link = "logit"),  
     data = toxo)

Deviance Residuals:
   Min       1Q   Median       3Q      Max
-2.9829  -1.2096  -0.4572   0.4160   2.8846

Coefficients:  
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.02505     0.07709  0.325  0.74524
poly(rain1000, 5)1 -0.24223     0.48608 -0.498  0.61825
poly(rain1000, 5)2 -0.23450     0.49023 -0.478  0.63240
poly(rain1000, 5)3  1.46167     0.43170  3.386  0.00071 ***
poly(rain1000, 5)4 -0.23823 0.47500 -0.502 0.61599
poly(rain1000, 5)5 0.51553 0.46234 1.115 0.26484

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 74.212  on 33  degrees of freedom
Residual deviance: 61.196  on 28  degrees of freedom
AIC: 163.89

Number of Fisher Scoring iterations: 3

>
> m2 <- glm(cbind(ypos,yneg)~poly(rain1000,3),data=toxo,
+     family=binomial(link="logit"))
> summary(m2)

Call:
glm(formula = cbind(ypos, yneg) ~ poly(rain1000, 3), family = binomial(link = "logit"),
    data = toxo)

Deviance Residuals:
       Min      1Q  Median      3Q     Max
-2.7620 -1.2166 -0.5079  0.3538  2.6204

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.02427    0.07693  0.315   0.7524
poly(rain1000, 3)1 -0.08606    0.45870 -0.188   0.8512
poly(rain1000, 3)2 -0.19269    0.46739 -0.412   0.6801
poly(rain1000, 3)3  1.37875    0.41150  3.351   0.0008 ***

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 74.212  on 33  degrees of freedom
Residual deviance: 62.635  on 30  degrees of freedom
AIC: 161.33

Number of Fisher Scoring iterations: 3

> anova(m2,m1,test="Chisq")
Analysis of Deviance Table

Model 1: cbind(ypos, yneg) ~ poly(rain1000, 3)
Model 2: cbind(ypos, yneg) ~ poly(rain1000, 5)
### Analysis of Deviance Table

**Model 1:** `cbind(ypos, yneg) ~ 1`

**Model 2:** `cbind(ypos, yneg) ~ poly(rain1000, 3)`

---

```r
> m0 <- glm(cbind(ypos, yneg)~1, data=toxo, family=binomial(link="logit"))
> anova(m0, m2, test="Chisq")
```

---

```r
> # deviance of model m2 is GOF statistic:
> deviance(m2)
> [1] 62.6346
> # Pearson X^2 statistic:
> sum(resid(m2, type="pearson")^2)
> [1] 58.21314
> m2q <- glm(cbind(ypos, yneg)~poly(rain1000, 3), data=toxo, 
+ family=quasibinomial(link="logit"))
> summary(m2q)
```

---

**Call:**

```
glm(formula = cbind(ypos, yneg) ~ poly(rain1000, 3), family = quasibinomial(link = "logit"),
    data = toxo)
```

**Deviance Residuals:**

```
    Min      1Q  Median      3Q     Max
-2.7620  -1.2166  -0.5079   0.3538  2.6204
```

**Coefficients:**

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.02427    0.10716   0.226   0.8224
poly(rain1000, 3)1 -0.08606    0.63897  -0.135   0.8938
poly(rain1000, 3)2 -0.19269    0.65108  -0.296   0.7693
poly(rain1000, 3)3  1.37875    0.57321   2.405  0.0225 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 . ‘.’ 0.1 ‘ ’ 1
```

(Dispersion parameter for quasibinomial family taken to be 1.940446)

---

Null deviance: 74.212 on 33 degrees of freedom
Residual deviance: 62.635 on 30 degrees of freedom
AIC: NA
Number of Fisher Scoring iterations: 3

```r
> r0 <- seq(from=min(toxo$rain1000),to=max(toxo$rain1000),length=100)
> expit <- function(x) {1/(1+exp(-x))}
> pred.m2 <- predict(m2,data.frame(rain1000=r0),se.fit=T,type="link")
> L <- expit(pred.m2$fit-1.96*pred.m2$se.fit)
> U <- expit(pred.m2$fit+1.96*pred.m2$se.fit)
> plot(toxo$rain1000,toxo$ppos,type="p",xlab="Rainfall/1000", ylab="Prop positive for toxoplasmosis", main="Fitted probability from model m2")
> lines(r0,expit(pred.m2$fit))
> lines(r0,L,lty=4)
> lines(r0,U,lty=4)
> legend(locator(1),lty=c(1,4),legend=c("Fitted probability","Approx 95% conf. limits"))
```

**Fitted probability from model m2**

```
> pred.m2q <- predict(m2q,data.frame(rain1000=r0),se.fit=T,type="link")
> L <- expit(pred.m2q$fit-1.96*pred.m2q$se.fit)
> U <- expit(pred.m2q$fit+1.96*pred.m2q$se.fit)
```
> plot(toxo$rain1000, toxo$ppos, type="p", xlab="Rainfall/1000",
+ ylab="Prop positive for toxoplasmosis",
+ main="Fitted probability from model m2q")
> lines(r0, expit(pred.m2q$fit))
> lines(r0, L, lty=4)
> lines(r0, U, lty=4)
> legend(locator(1), lty=c(1, 4), legend=c("Fitted probability", "Approx 95% conf. limits"))

**Fitted probability from model m2q**