RCodeAndTasksChapter5.Rmd

Witold Wolski

December 27, 2016

Chapter 5 Poisson regression
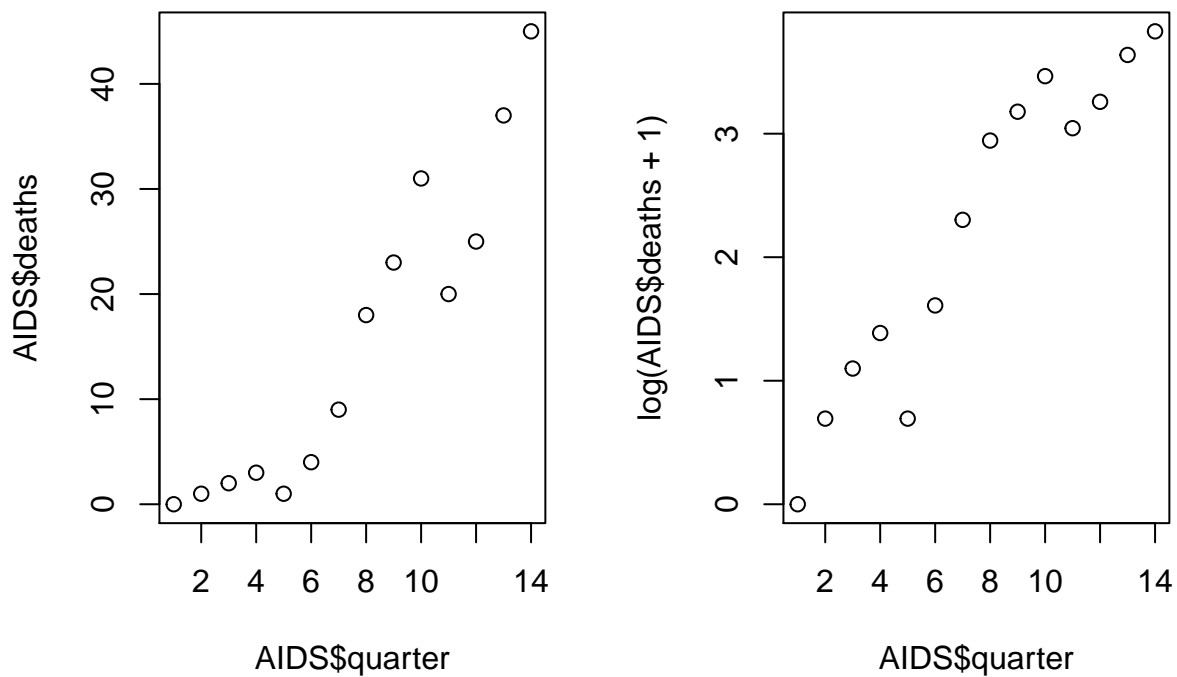
5.1 Introduction

pdf of poisson

$$\frac{\lambda^k e^{-\lambda}}{k!}$$

1.1 Example : AIDS deaths over time

1 plot :

```r
rm(list=ls())
load("data/MAS367-GLMs.RData", envir = e <- new.env())

AIDS <- e$AIDS
par(mfrow=c(1,2))
plot(AIDS$quarter, AIDS$deaths)
plot(AIDS$quarter, log(AIDS$deaths+1))
```

2 fit poisson with log link

```
glm.lin <- glm(deaths ~ quarter, data=AIDS, family=poisson(link='log'))
qchisq(0.95,glm.lin$df.residual)
```

```
## [1] 21.02607
```

3 adding a quadratic term

```
glm.quad <- glm(deaths ~ quarter + I(quarter^2), data=AIDS, family=poisson(link='log'))
summary(glm.quad)
```

```
##
## Call:
## glm(formula = deaths ~ quarter + I(quarter^2), family = poisson(link = "log"),
##     data = AIDS)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -1.7708  -0.9385   0.1304   0.8190   1.4421
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.713375   0.733108  -2.337 0.019432 *
## quarter       0.746031   0.153391   4.864 1.15e-06 ***
## I(quarter^2) -0.025836   0.007751  -3.333 0.000859 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 207.272  on 13  degrees of freedom
## Residual deviance:  16.371  on 11  degrees of freedom
## AIC: 75.298
##
## Number of Fisher Scoring iterations: 4
```

```r
qchisq(0.95,glm.lin$df.residual)
```

```
## [1] 21.02607
```

4 a line predictor on log(x)

```r
glm.logline <- glm(deaths ~ I(log(quarter)), data=AIDS, family=poisson(link='log'))
summary(glm.logline)
```

```
##
## Call:
## glm(formula = deaths ~ I(log(quarter)), family = poisson(link = "log"),
##     data = AIDS)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.08992  -1.07141  -0.04657   0.38956   1.94311
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.9442     0.5116   -3.80 0.000145 ***
## I(log(quarter))   2.1748     0.2150   10.11  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 207.272  on 13  degrees of freedom
## Residual deviance:  17.092  on 12  degrees of freedom
## AIC: 74.019
##
## Number of Fisher Scoring iterations: 4
```

```
qchisq(0.95,glm.logline$df.residual)
```

```
## [1] 21.02607
```

```
plot(log(AIDS$quarter), log(AIDS$deaths+0.5))
```
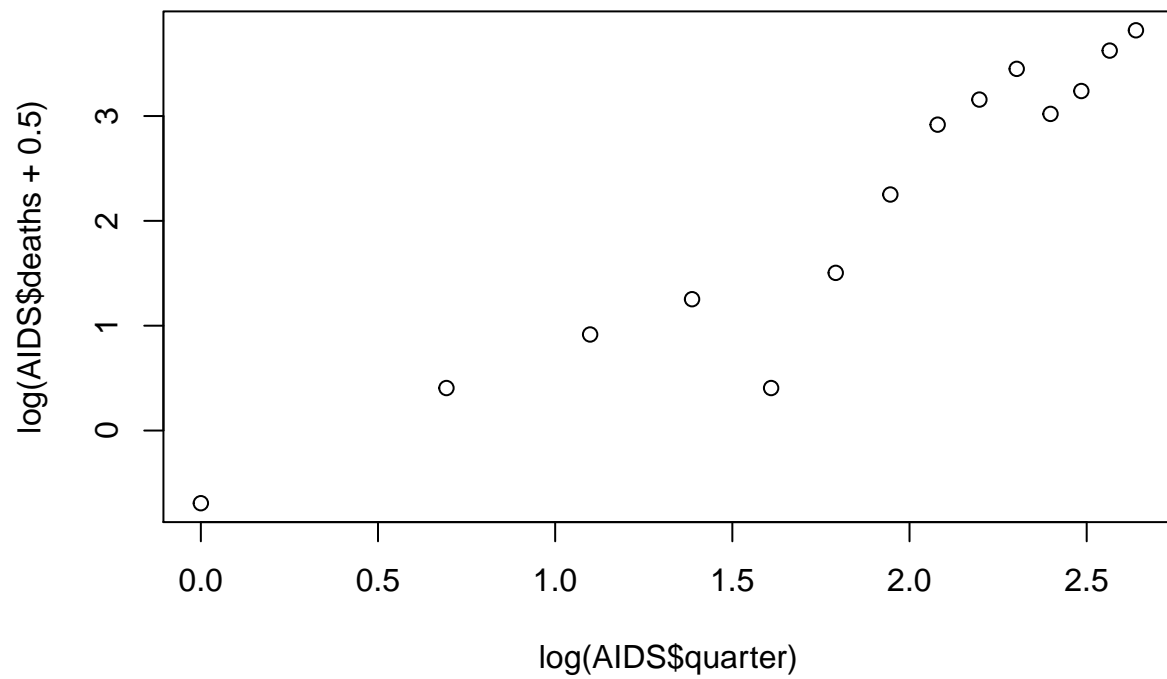

5

Thus possible simple models are a line in logx or a quadratic in x, but there are reservations about both.