

Lecture 17

Peter Shaffery

3/3/2021

Poisson Distribution

The Poisson distribution describes random numbers of *counts*. Unlike the binomial distribution, which modelled counts of *successes* out of total *trials*, a Poisson random variable can go infinitely high. The PMF of a Poisson distribution is:

$$f(y) = \frac{\lambda^y e^{-\lambda}}{y!}, \quad y = 0, 1, 2, 3, \dots$$

The Poisson distribution has a single parameter λ , known as the **rate** parameter. When discussing this parameter, it is very important to pay attention to the units. For example, if you were modeling car crashes you then you may see λ defined as the mean number of crashes, mean crashes per 100,000 miles driven, or mean crashes per 100,000 miles driven per year.

An unusual feature of the Poisson distribution is that $E[y] = \text{Var}[y] = \lambda$.

Poisson Regression

The Poisson regression model is actually quite simple, there's really only one choice of link function that sees any use:

$$g(E[y_i]) = \log(E[y_i])$$

Hence the Poisson regression model can be written:

$$E[y_i] = \exp(\vec{x}_i \vec{\beta})$$

Deviance and Residuals

Recalling that the Pearson residuals are defined:

$$\chi_i = \frac{y_i - \hat{y}_i}{\sqrt{\text{Var}[y_i|\hat{\lambda}_i]}}$$

We see that, since our estimate of $\hat{y}_i = E[y_i|\lambda_i] = \hat{\lambda}_i$ (where $\hat{\lambda}_i = \exp \vec{x}_i^T \hat{\beta}$) and furthermore that $\text{Var}[y_i|\hat{\lambda}_i] = \hat{\lambda}_i$, then the Pearson residuals are simply:

$$\chi_i = \frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i}} = \frac{y_i - \hat{\lambda}_i}{\sqrt{\hat{\lambda}_i}}$$

Similarly the deviance can be written:

$$\begin{aligned} D &= 2 \sum_i y_i \log(y_i/\hat{y}_i) - (y_i - \hat{y}_i) \\ &= 2 \sum_i y_i \log(y_i/\hat{\lambda}_i) - (y_i - \hat{\lambda}_i) \end{aligned}$$

Which gives us deviance residuals:

$$D_i = \text{sign}(y_i - \hat{y}_i) \sqrt{2y_i \log(y_i/\hat{y}_i) - (y_i - \hat{y}_i)}$$

$$= \text{sign}(y_i - \hat{\lambda}_i) \sqrt{2y_i \log(y_i/\hat{\lambda}_i) - (y_i - \hat{\lambda}_i)}$$

Example: Household Size

The Philippine Statistics Authority (PSA) surveys households in the Phillipines every three years. One set of data collected is the number of individuals living in a household, other than the “head of household” (HoH). The PSA would like to understand how the age of the head of the household relates to the number of individuals living there.

```
library(tidyverse)
library(magrittr)

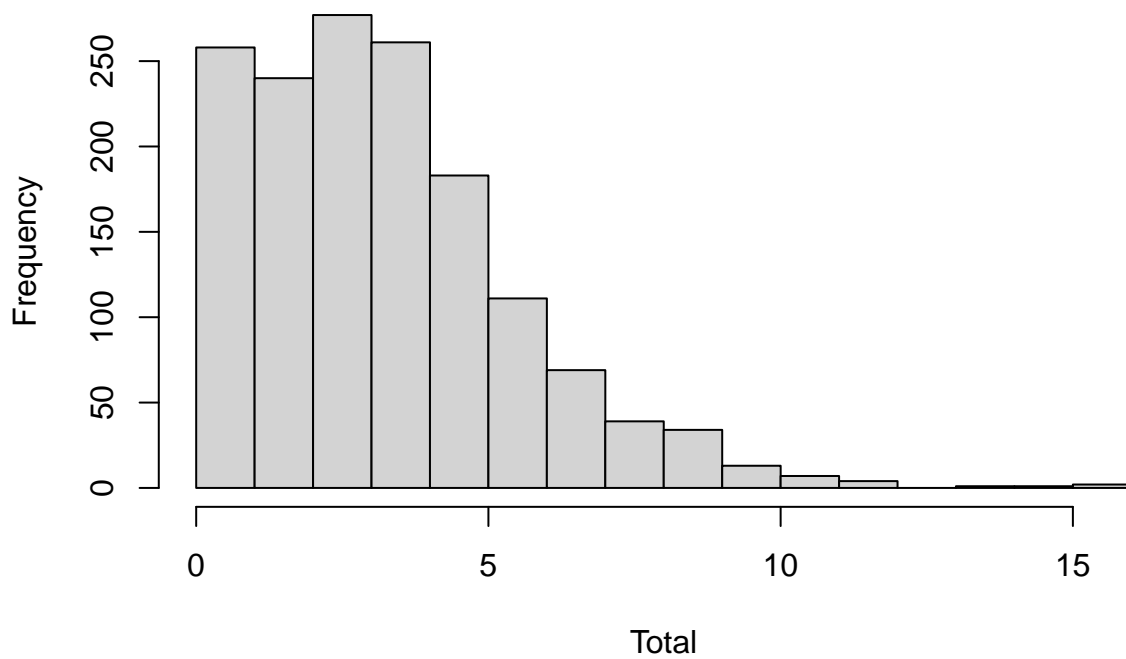
dat = read.csv('../data/households.csv')
dat %>% head
```

```
##      location age total numLT5      roof
## 1 Centralluzon 65      0      0 Predominantly Strong Material
## 2 MetroManila 75      3      0 Predominantly Strong Material
## 3 DavaoRegion 54      4      0 Predominantly Strong Material
## 4      Visayas 49      3      0 Predominantly Strong Material
## 5 MetroManila 74      3      0 Predominantly Strong Material
## 6      Visayas 59      6      0 Predominantly Strong Material
```

Variable	Def'n
LOCATION	Region where house is located
AGE	The age of the HoH
TOTAL	Number of individuals living in house other than HoH
NUMLT5	Number of individuals under 5 years of age
ROOF	Type of roof on the house (Light or Strong, proxy for household wealth)

```
hist(dat$total, xlab='Total')
```

Histogram of dat\$total

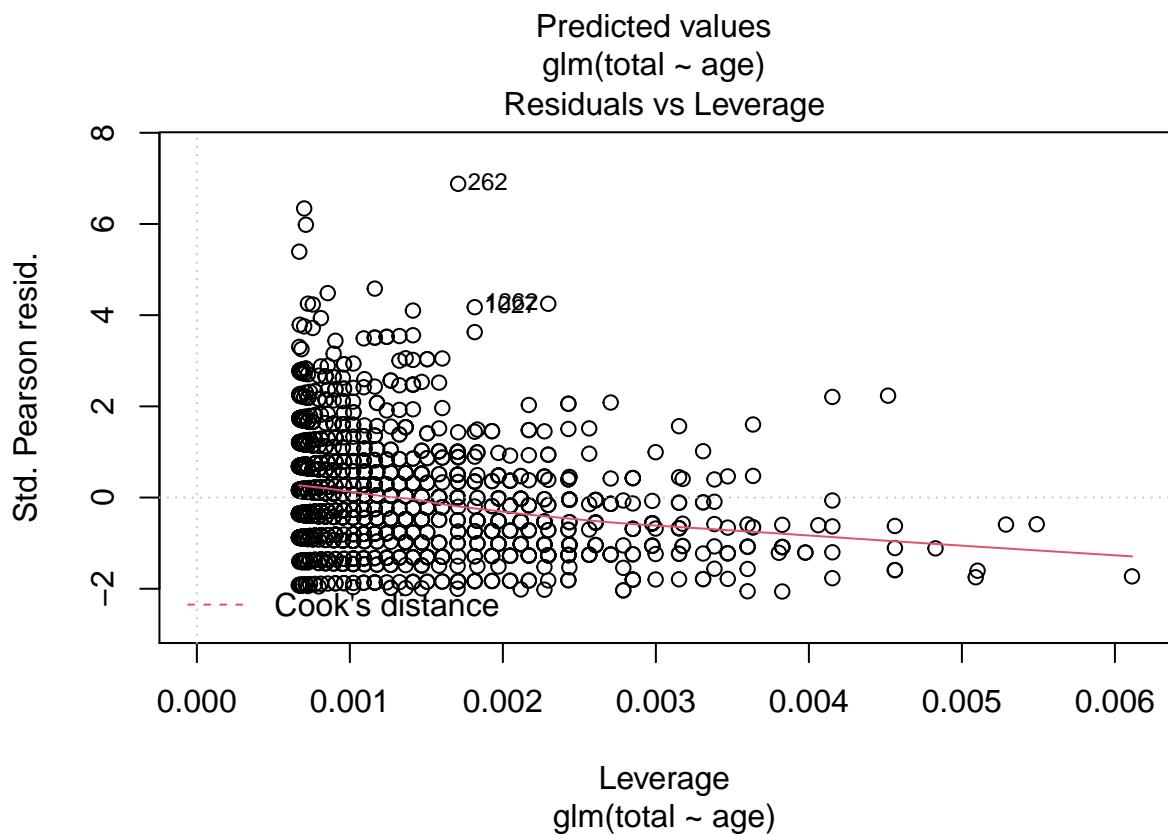
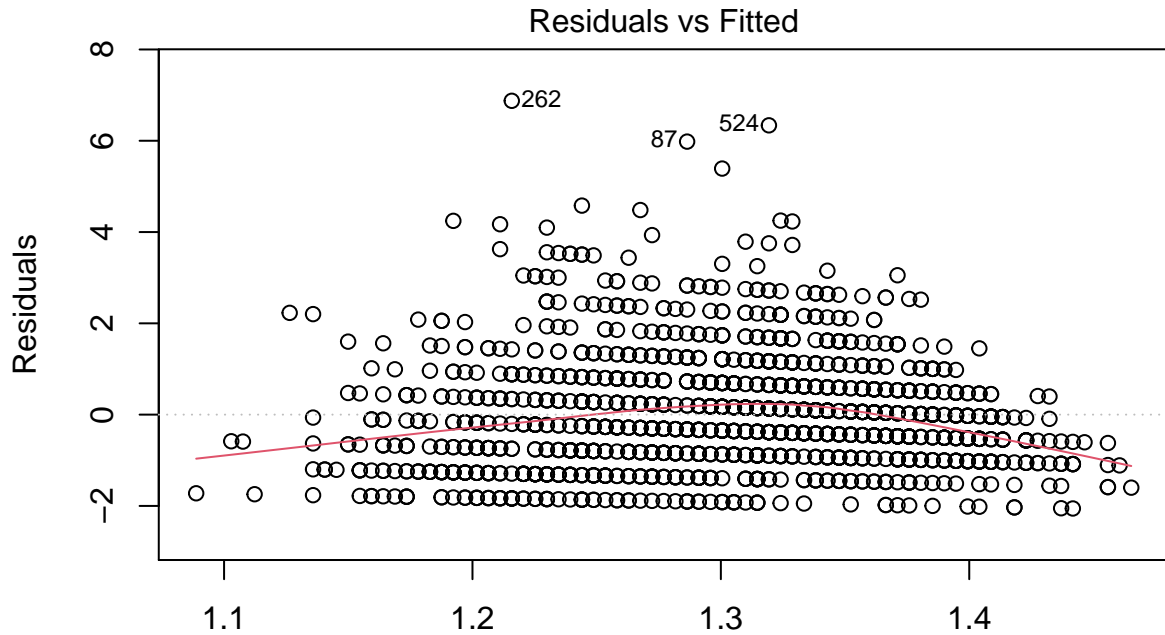


Let's fit a quick model, just relating AGE to TOTAL:

```
mod = glm(total~age, data=dat, family=poisson)
summary(mod)
```

```
##
## Call:
## glm(formula = total ~ age, family = poisson, data = dat)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.9079  -0.9637  -0.2155   0.6092   4.9561
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.5499422  0.0502754  30.829  < 2e-16 ***
## age         -0.0047059  0.0009363  -5.026  5.01e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 2362.5  on 1499  degrees of freedom
## Residual deviance: 2337.1  on 1498  degrees of freedom
## AIC: 6714
##
## Number of Fisher Scoring iterations: 5
```

```
plot(mod,which=c(1,5))
```



As with logistic regression, the coefficient has a special interpretation, known as the **rate ratio**:

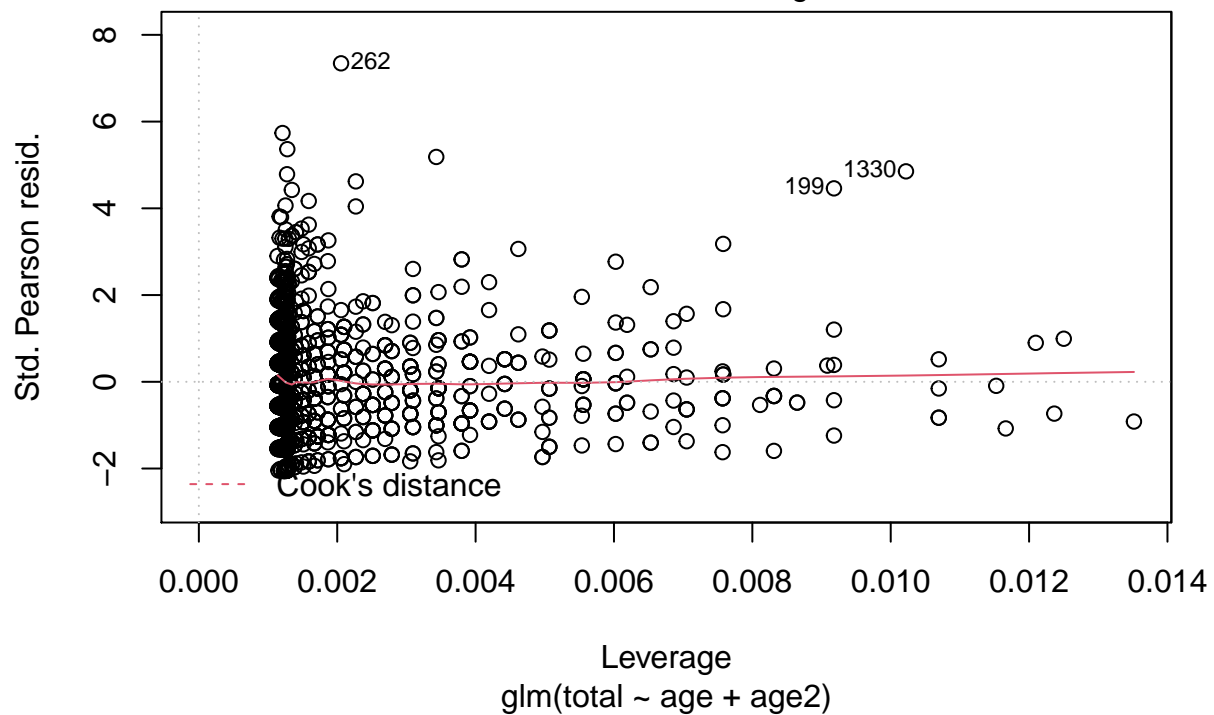
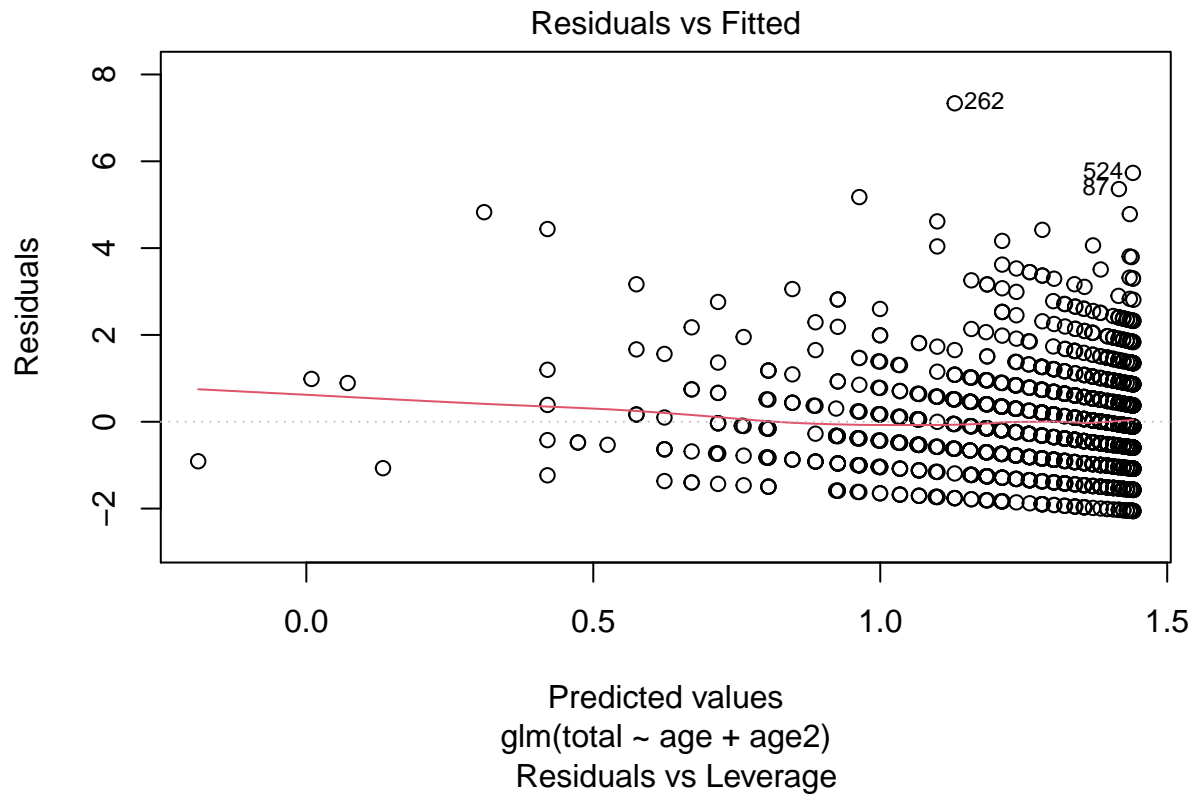
$$RR = \frac{E[y_i | AGE=a+1]}{E[y_i | AGE=a]} = \exp \beta_{AGE}$$

In this example, $\beta_{AGE} \approx -.005$, so the rate ratio is $RR = \exp(-.005) = .995$. Hence, for every additional year of HoH age, the average household size decreases by about half a percent (.005). Put another way, we expect a .5% *increase* in household size for every *decrease* of HoH age by one year.

Obviously we don't expect an infant to have the largest possible house size, so this prompts us to consider a *transformation* of the AGE variable. Specifically, we will include a *quadratic* term, so that $E[y_i]$ will be maximized for some middle value of AGE:

```
dat %<>% mutate(age2 = age^2)
mod = glm(total~age+age2, data=dat, family=poisson)
summary(mod)

##
## Call:
## glm(formula = total ~ age + age2, family = poisson, data = dat)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.9068  -0.9261  -0.1048   0.5773   5.1731
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.325e-01  1.788e-01  -1.859   0.063 .
## age          7.089e-02  6.890e-03  10.288 <2e-16 ***
## age2        -7.083e-04  6.406e-05 -11.058 <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: 2362.5  on 1499  degrees of freedom
## Residual deviance: 2200.9  on 1497  degrees of freedom
## AIC: 6579.8
##
## Number of Fisher Scoring iterations: 5
plot(mod,which=c(1,5))
```



Although the regression table has performed a Wald test for us, let's perform a deviance hypothesis test anyways to determine if our new quadratic model is an improvement over the original linear model.

Treating our linear model as the “null” model, we have $D_0 = 2337$ and $D_1 = 2200$, giving us a $\Delta D = 137$. Furthermore, the degrees of freedom of our null model is 1498, and of our alternate model is 1497, hence under the null hypothesis we would expect $\Delta D \sim \chi^2(1)$. Using $\alpha = .05$:

```
rejection.threshold = qchisq(.95,1)
delta.d = 137
print(delta.d > rejection.threshold)
```

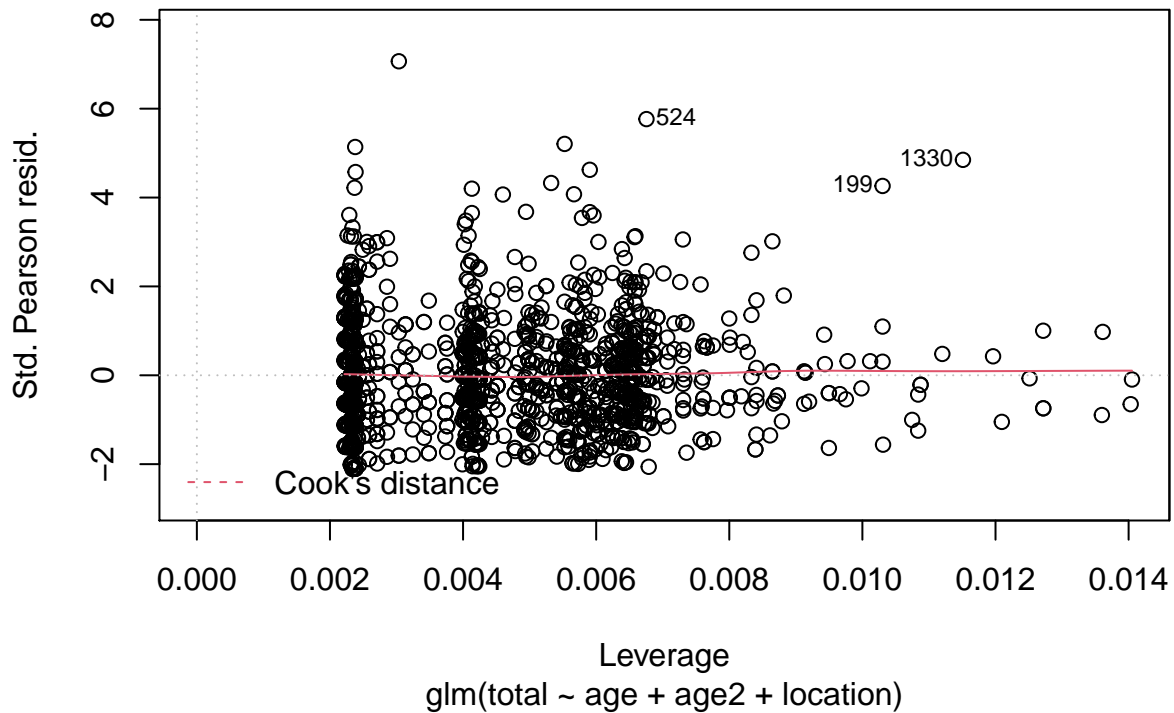
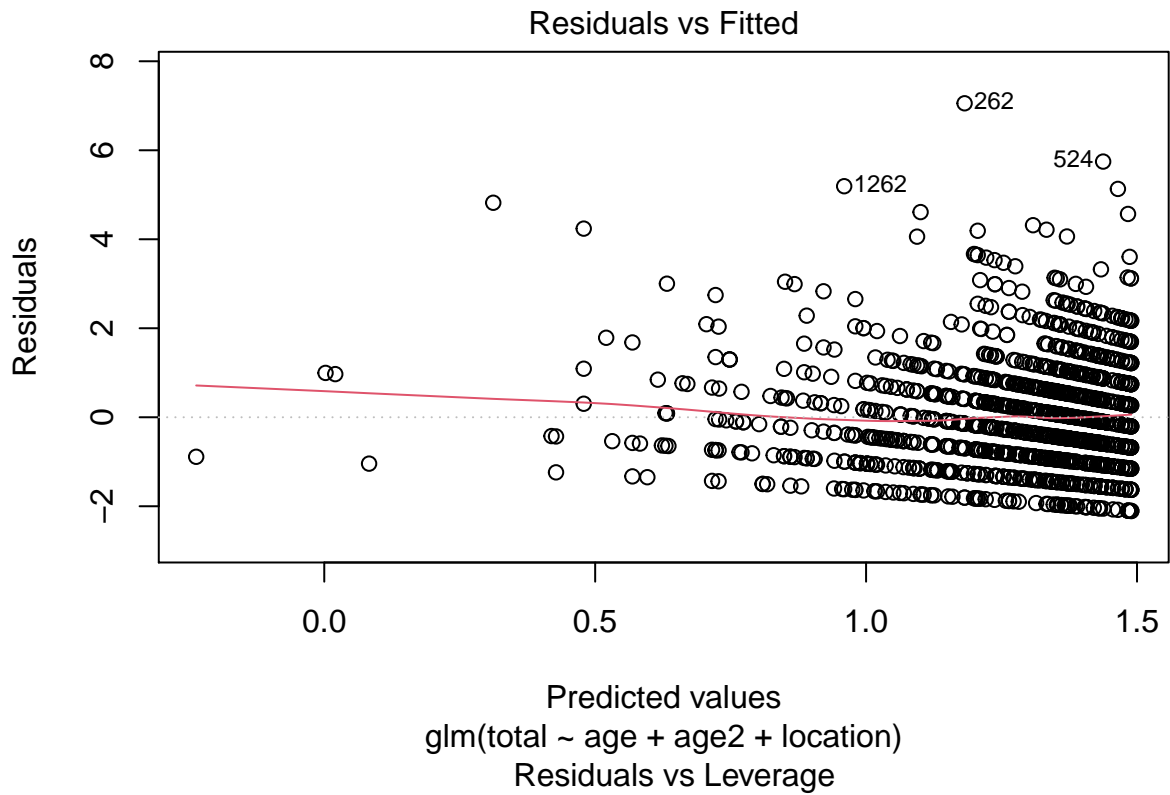
```
## [1] TRUE
```

And so we reject the linear model, in favor of the quadratic model (and by quite a lot).

Let's do one more for good measure, now checking whether the addition of the LOCATION variable is of any use:

```
dat %<>% mutate(age2 = age^2)
mod = glm(total~age+age2+location, data=dat, family=poisson)
summary(mod)
```

```
##
## Call:
## glm(formula = total ~ age + age2 + location, family = poisson,
##      data = dat)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.9781  -0.9278  -0.1149   0.5924   5.0424
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.3843338   0.1820919  -2.111  0.03480 *
## age             0.0703628   0.0069051  10.190 < 2e-16 ***
## age2          -0.0007026   0.0000642 -10.944 < 2e-16 ***
## locationDavaoRegion -0.0193872   0.0537827  -0.360  0.71849
## locationIlocosRegion  0.0609820   0.0526598   1.158  0.24685
## locationMetroManila   0.0544801   0.0472012   1.154  0.24841
## locationVisayas      0.1121092   0.0417496   2.685  0.00725 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 2362.5  on 1499  degrees of freedom
## Residual deviance: 2187.8  on 1493  degrees of freedom
## AIC: 6574.7
##
## Number of Fisher Scoring iterations: 5
plot(mod,which=c(1,5))
```



Because LOCATION is a dummy variable we now have no option to go by the Wald test, and have to use a deviance test instead.

For the quadratic-only model we have $D_0 = 2200$ with $df = 1497$, and for the quadratic+location model we have $D_1 = 2187$ with $df = 1493$. Hence under the null hypothesis we would expect $\Delta D = 13 \sim \chi^2(4)$. The p-value for this is:


```
pchisq(13,4,lower.tail = FALSE)
```

```
## [1] 0.01127579
```

Which is below our choice of $\alpha = .05$, so we reject the quadratic-only model in favor of the quadratic+location model.

Exposure

There is an interesting relationship between the binomial and Poisson distributions. Say that we have a binomial distribution with *very* large n (the number of trials) and *very* small p (the probability of success). It turns out that if $y \sim \text{Binom}(p, n)$, then y is approximately Poisson distributed with rate parameter $\lambda = np$. As n grows larger, and p grows smaller, this approximation becomes more and more exact.

In practice, this relationship can be interpreted through **exposure**. Say that we are recording the occurrence of a tree disease in different forest transects. You might imagine that larger transects have more incidence of disease. Even if this relationship isn't precisely binomial (in the sense that there is not a well-defined number of "trials" occurring), an *exposure model* allows us to account for this in our regression:

$$E[y_i] = n_i \theta_i$$
$$\theta_i = \exp(\vec{x}_i^T \vec{\beta})$$

Or, equivalently:

$$\log E[y_i] = \log n_i + \vec{x}_i^T \vec{\beta}$$

Where the $\log n_i$ term is referred to as the **offset**.

Example: Campus Crime

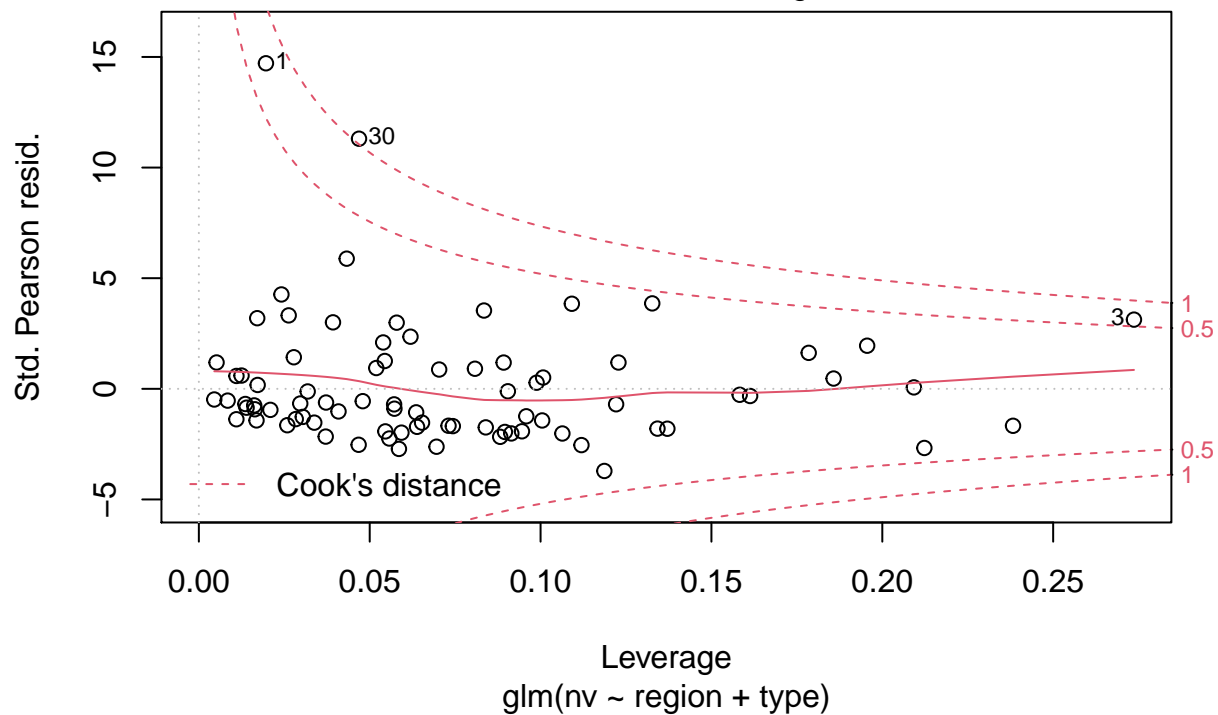
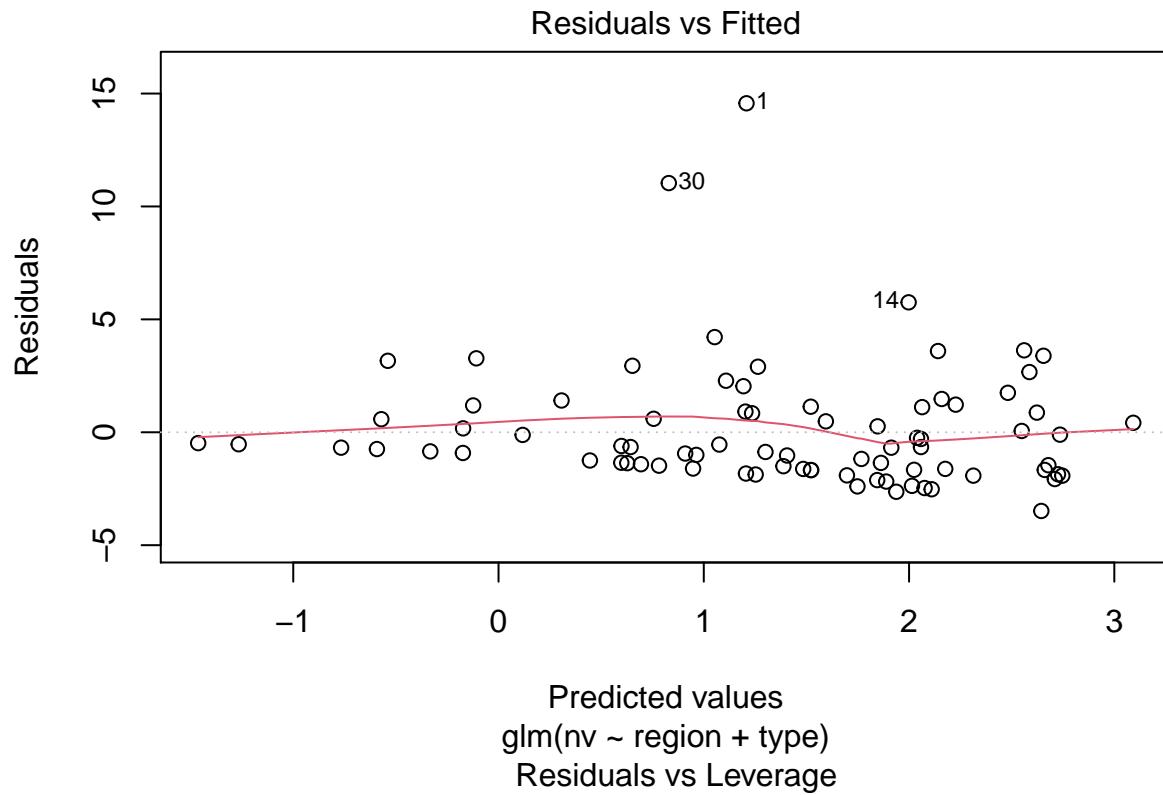
```
dat = read.csv('../data/campus_crime.csv')
dat$region = ifelse(dat$region=='SE', 'S', dat$region)
dat$region = ifelse(dat$region=='SW', 'S', dat$region)
dat %>% head
```

```
##   type nv enroll1000 region
## 1    U 30      5.590      S
## 2    C  0      0.540      S
## 3    U 23     35.747      W
## 4    C  1     28.176      W
## 5    U  1     10.568      S
## 6    U  0      3.127      S
```

Variable	Def'n
TYPE	College or University
NV	Number of violent crimes
ENROLL1000	University enrollment (in 1000s of students)
REGION	Location of school in US (C = Central, MW = Midwest, NE = Northeast, SE = Southeast, SW = Southwest, and W = West)

```
mod = glm(nv ~ region + type, offset=log(enroll1000), data=dat, family=poisson )
summary(mod)
```

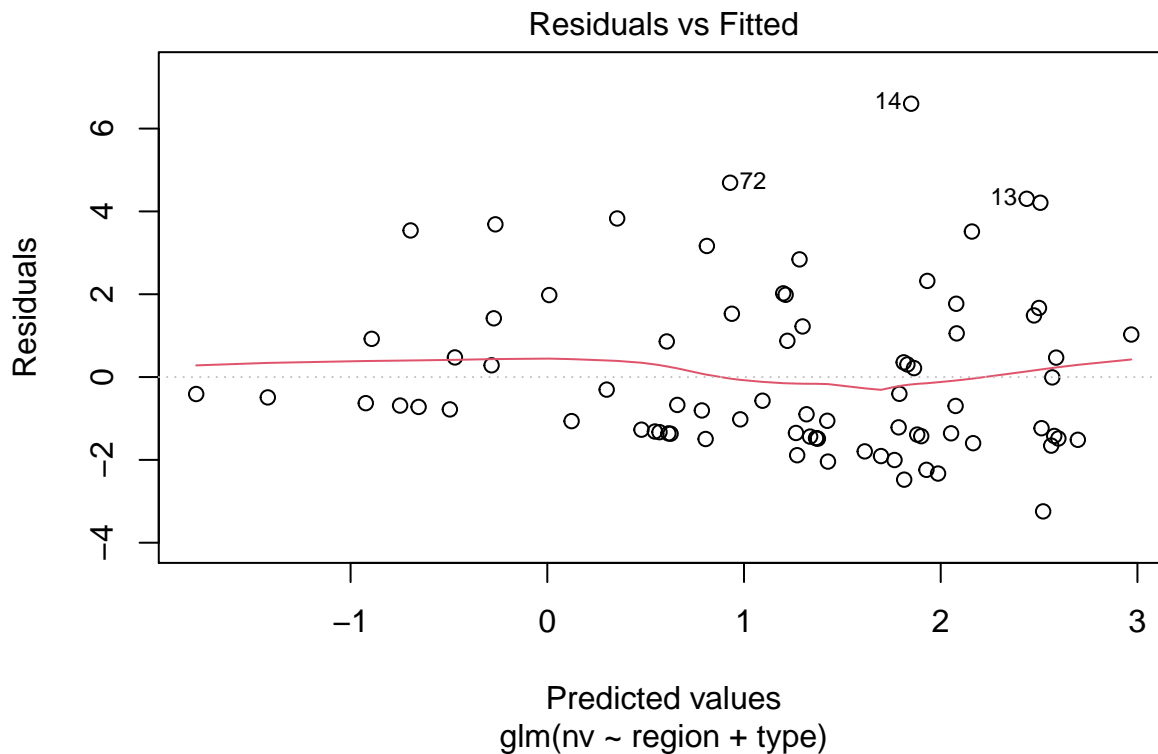
```
##
## Call:
## glm(formula = nv ~ region + type, family = poisson, data = dat,
##      offset = log(enroll1000))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.5671  -1.9636  -0.7205   1.0170   8.8476
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.59629    0.17115  -9.327  < 2e-16 ***
## regionMW     0.09939    0.17752   0.560  0.5756
## regionNE     0.78081    0.15305   5.102 3.37e-07 ***
## regionS      0.74926    0.14503   5.166 2.39e-07 ***
## regionW      0.27223    0.18742   1.453  0.1464
## typeU        0.33415    0.13235   2.525  0.0116 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 491.00  on 80  degrees of freedom
## Residual deviance: 431.45  on 75  degrees of freedom
## AIC: 661.32
##
## Number of Fisher Scoring iterations: 6
plot(mod,which=c(1,5))
```

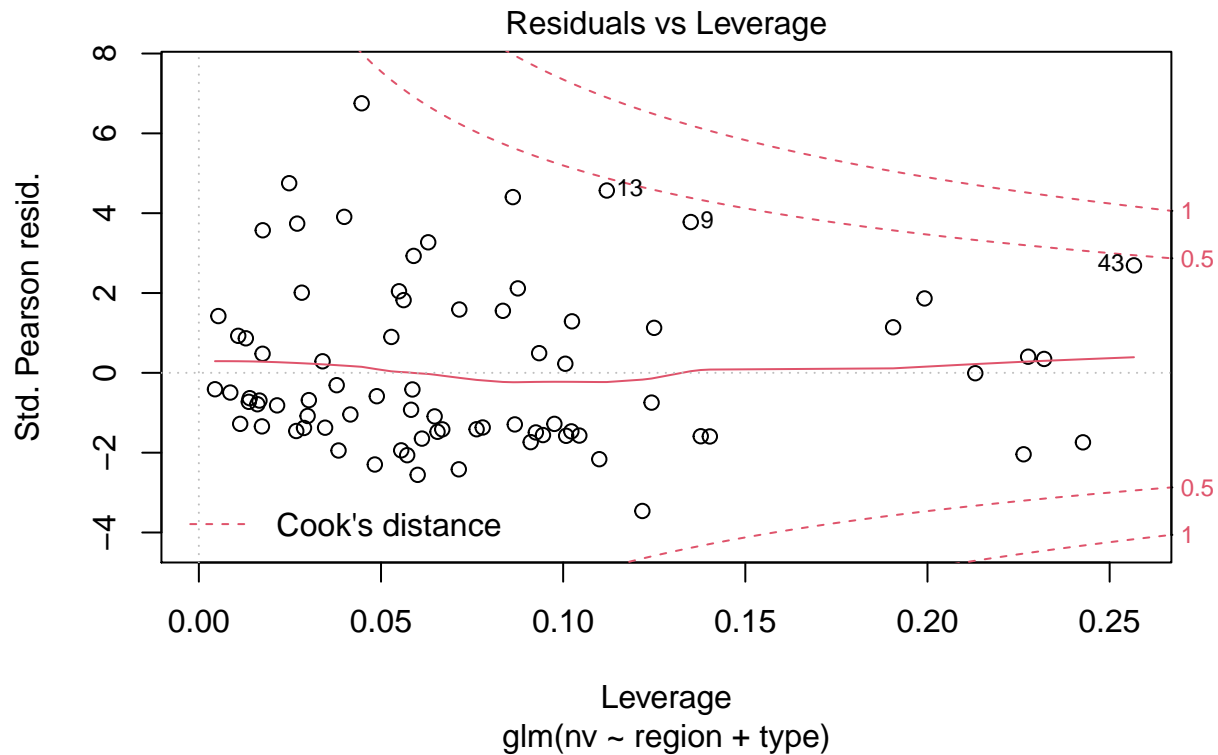


```
# Points 1,3, and 30 are highly influential, so cut
dat = dat[-c(1,3,30),]
mod = glm(nv ~ region + type, offset=log(enroll1000), data=dat, family=poisson )
summary(mod)
```

```
##
```

```
## Call:
## glm(formula = nv ~ region + type, family = poisson, data = dat,
##      offset = log(enroll1000))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.223  -1.751  -0.817   1.088   5.086
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.7521     0.1875  -9.347  < 2e-16 ***
## regionMW       0.1002     0.1775   0.564  0.57255
## regionNE       0.6397     0.1576   4.059  4.93e-05 ***
## regionS        0.5834     0.1489   3.916  8.99e-05 ***
## regionW        0.0274     0.2265   0.121  0.90369
## typeU          0.5075     0.1550   3.274  0.00106 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 336.76  on 77  degrees of freedom
## Residual deviance: 289.72  on 72  degrees of freedom
## AIC: 504.58
##
## Number of Fisher Scoring iterations: 5
plot(mod,which=c(1,5))
```





```
rate.ratios = exp(mod$coefficients)
rate.ratios
```

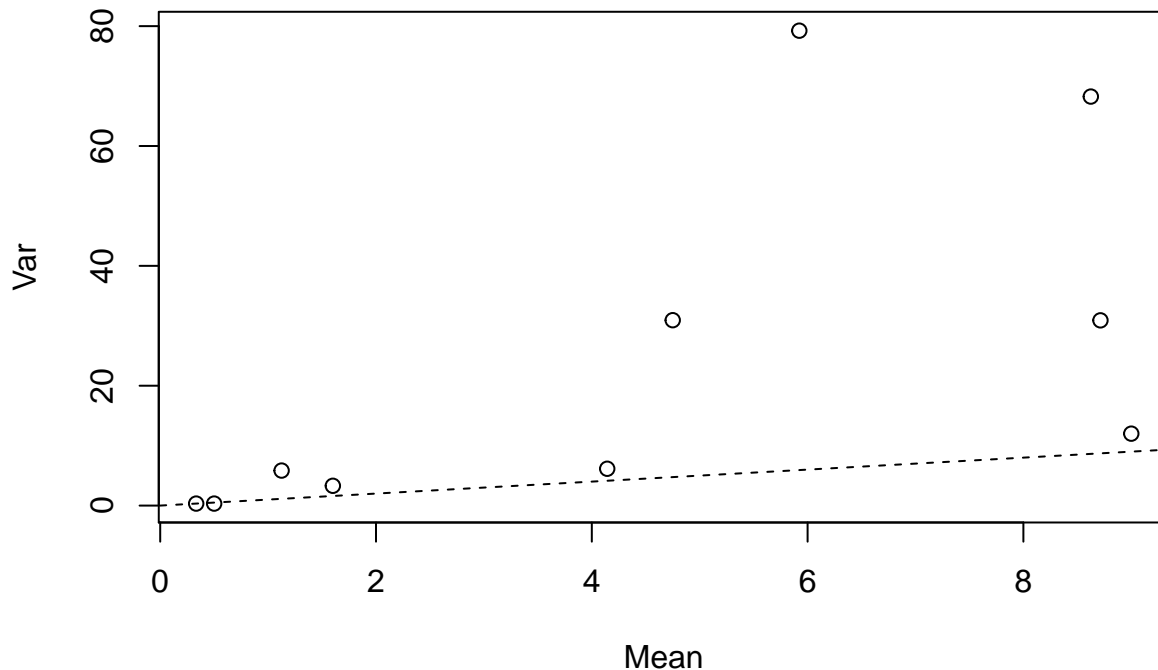
```
## (Intercept)    regionMW    regionNE    regionS    regionW    typeU
##  0.1734033    1.1053645    1.8958994    1.7920621    1.0277787    1.6610875
```

Overdispersion

Recall that the Poisson model assumes that the mean and variance of the y_i are equal. As with logistic regression, it is quite common for count data to be **overdispersed**, that is for the actual $\text{Var}[y_i]$ to be greater than $E[y_i] = \lambda_i$.

We can quickly see that this is the case with the Campus Crime data:

```
summ = dat %>% group_by(type,region) %>% summarize(mn=mean(nv),vr=var(nv))
plot(summ$mn,summ$vr,xlab='Mean',ylab='Var')
lines(seq(0,12,1),seq(0,12,1),lty=2)
```



One remedy is to use a model which assumes:

$$\text{Var}[y_i] = \phi \lambda_i$$

```
mod.quasipois = glm(nv ~ region + type, offset=log(enroll1000), data=dat, family=quasipoisson )
summary(mod.quasipois)
```

```
##
## Call:
## glm(formula = nv ~ region + type, family = quasipoisson, data = dat,
##      offset = log(enroll1000))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.223  -1.751  -0.817   1.088   5.086
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.7521     0.3866  -4.532 2.27e-05 ***
## regionMW      0.1002     0.3661   0.274  0.7852
## regionNE      0.6397     0.3250   1.968  0.0529 .
## regionS       0.5834     0.3072   1.899  0.0616 .
## regionW       0.0274     0.4670   0.059  0.9534
## typeU         0.5075     0.3196   1.588  0.1168
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 4.253308)
##
##      Null deviance: 336.76  on 77  degrees of freedom
## Residual deviance: 289.72  on 72  degrees of freedom
## AIC: NA
```

```
##
## Number of Fisher Scoring iterations: 5
```

Negative Binomial Model

Unlike with a Binomial model however, in Poisson we have a second (more elegant) option: a negative Binomial model.

The “story” of the Negative Binomial distribution depends a little on who you ask. For the purposes of Poisson regression, it can be thought of as modeling a Poisson random variable, where the rate parameter λ is **also random**. Specifically, if:

$$y|\lambda \sim \text{Pois}(\lambda)$$

$$\lambda \sim \text{Gamma}(r, \frac{1-p}{p})$$

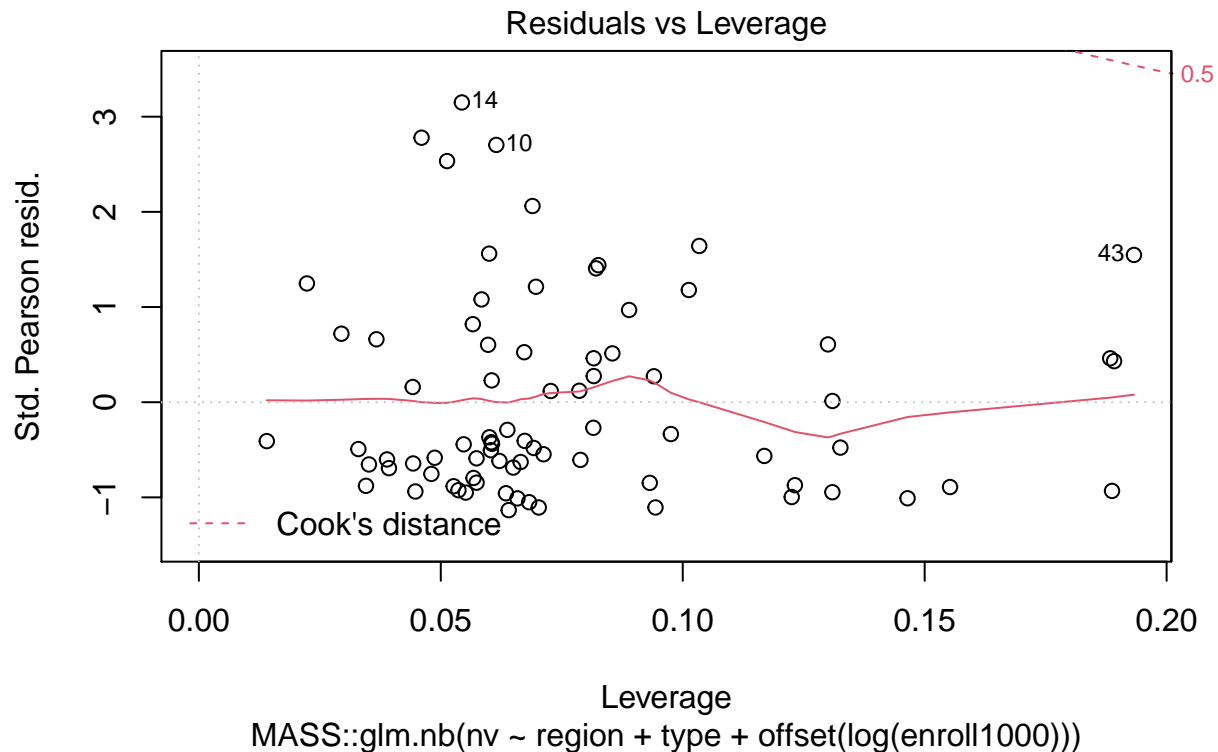
Then $y \sim \text{NegBinom}(r, p)$ where $E[y] = \frac{pr}{1-p}$ and $\text{Var}[y] = \frac{pr}{(1-p)^2} = E[y] + \frac{E[y]^2}{r}$. In this case the addition of the r parameter gives our model the extra flexibility it needs to allow for different variance and mean.

Tragically, R does not include the Negative Binomial as an option for the `glm` function, so we’ll need to use a package:

```
mod.negbinom = MASS::glm.nb(nv ~ region + type + offset(log(enroll1000)), data=dat)
summary(mod.negbinom)
```

```
##
## Call:
## MASS::glm.nb(formula = nv ~ region + type + offset(log(enroll1000)),
## data = dat, init.theta = 1.458107418, link = log)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2563  -1.0642  -0.4933   0.4845   1.9251
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.6131     0.3344  -4.824 1.41e-06 ***
## regionMW       0.1066     0.4011   0.266  0.7905
## regionNE       0.7754     0.3351   2.314  0.0207 *
## regionS        0.5460     0.3269   1.670  0.0949 .
## regionW       -0.1491     0.4696  -0.318  0.7508
## typeU          0.3494     0.2665   1.311  0.1897
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(1.4581) family taken to be 1)
##
##      Null deviance: 95.447  on 77  degrees of freedom
## Residual deviance: 86.179  on 72  degrees of freedom
## AIC: 395.42
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta:  1.458
```

```
##          Std. Err.:  0.356
##
## 2 x log-likelihood: -381.420
plot(mod.negbinom,which=5)
```



Zero Inflation

One more bit of weirdness which you may encounter in a Poisson regression is an overabundance of zero values.

For example, let's look at data taken from a survey of college students which included the question "How many alcoholic drinks did you consum last weekend?". This survey was conducted on a "dry campus", where no drinking was allowed even among students of legal drinking age. There will therefore be an unusually large number of students that report 0 drinks, which includes a **mixture** of students that *never* drink (form whom $y_i = 0$ no matter how much we sample), as well as students who just didn't happen to drink this weekend (for whom a different sample might return $y_i > 0$).

```
dat = read.csv('../data/weekend_drinks.csv')
dat %>% head
```

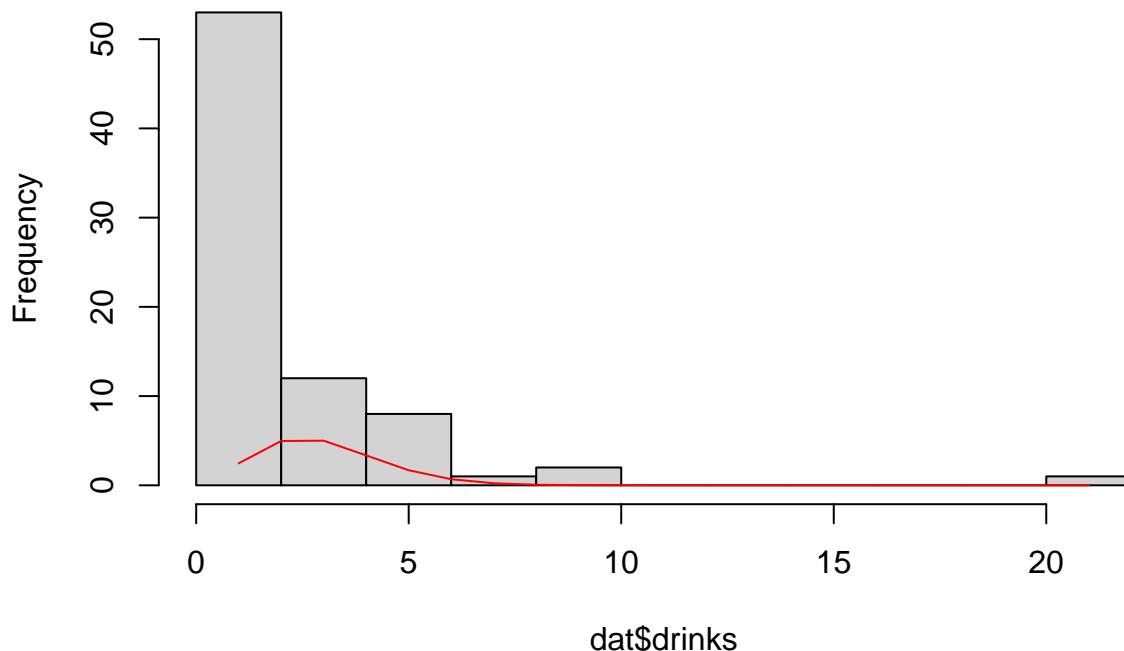
```
##      dorm drinks sex first.year off.campus
## 1  mohn      0  f      TRUE      FALSE
## 2 ytterboe    5  f     FALSE      FALSE
## 3  larson    10  m     FALSE      FALSE
## 4   rand     0  f     FALSE      FALSE
## 5 thorson     0  m     FALSE      FALSE
## 6 hillboe     3  f     FALSE      FALSE
```

```
hist(dat$drinks)
pmf = dpois(seq(0,20,1),mean(dat$drinks))
```



```
pmf = 5*pmf/max(pmf)
lines(pmf,col='red')
```

Histogram of dat\$drinks



To account for the **inflated** number of 0s, we will introduce a new parameter α to our model, representing the proportion of *non-drinkers*, ie. students for whom $y_i = 0$ no matter what. We will attempt to jointly estimate λ (the mean number of drinks *among students who drink*) and α . This type of model is referred to as a **zero-inflated Poisson** (ZIP) model.

Fitting

The ZIP model belongs to a class of models known as **mixture models**. These are models which assume that the data are sampled from *more than one* underlying distributions. Unlike a basic Poisson regression, where all data points are simply drawn from a Poisson distribution, in a mixture model the sampling process looks like:

1. Select an underlying distribution at random from D_1, \dots, D_M with probabilities $\alpha_1, \dots, \alpha_M$ (*does this remind you of a distribution we learned about last lecture?*)
2. Given a choice D_m , draw $y_i \sim D_m$

In our ZIP model, $M = 2$, with D_1 is just a spike at 0 and D_2 is the Poisson distribution. $\alpha_1 = \alpha$ and $\alpha_2 = 1 - \alpha$.

Fitting ZIP models (and mixture models generally) is a little beyond the scope of this course, but the basic idea is fairly simple. Given some estimate of λ , you can make a reasonable guess at the number of zero counts which are due to the drinking population, ie. $1 - \alpha = \text{Pois}(0, \hat{\lambda}_i)$. You can then use this guess of $1 - \alpha$ to construct a likelihood for α . We now subtract α percent of the zeros from the data, re-estimate $\hat{\lambda}_i$, and iterate.

As with the negative binomial model, base R does not give us an option for a ZIP model, so we will use a fitting function provided by the **pscl** package. For this model we actually need to specify **two** models, one for λ (which includes a student's sex and whether they live off campus), and one for α (which here uses

first-year status as a proxy for age)

```
mod.zip = pscl::zeroinfl(drinks ~ off.campus + sex|first.year, data = dat)
summary(mod.zip)
```

```
##
## Call:
## pscl::zeroinfl(formula = drinks ~ off.campus + sex | first.year, data = dat)
##
## Pearson residuals:
##      Min       1Q   Median       3Q      Max
## -1.1118 -0.8858 -0.5290  0.6367  5.2996
##
## Count model coefficients (poisson with log link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.7543    0.1440   5.238 1.62e-07 ***
## off.campusTRUE 0.4159    0.2059   2.020  0.0433 *
## sexm          1.0209    0.1752   5.827 5.63e-09 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.6036    0.3114  -1.938  0.0526 .
## first.yearTRUE 1.1364    0.6095   1.864  0.0623 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 8
## Log-likelihood: -140.8 on 5 Df
```

```
mod.nozip = glm(drinks ~ off.campus + sex, data = dat, family=poisson)
summary(mod.nozip)
```

```
##
## Call:
## glm(formula = drinks ~ off.campus + sex, family = poisson, data = dat)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6350 -1.5086 -0.9119  0.7290  6.6472
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.1293    0.1241   1.041  0.298
## off.campusTRUE 0.8976    0.2008   4.470 7.83e-06 ***
## sexm          1.1154    0.1611   6.925 4.36e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 294.54  on 76  degrees of freedom
## Residual deviance: 230.54  on 74  degrees of freedom
## AIC: 357.09
##
## Number of Fisher Scoring iterations: 6
```

```
exp(mod.zip$coefficients$count)
```

```
##      (Intercept) off.campusTRUE      sexm  
##      2.126070      1.515795      2.775691
```

```
exp(mod.nozip$coefficients)
```

```
##      (Intercept) off.campusTRUE      sexm  
##      1.137996      2.453819      3.050707
```

Note that the ZIP model table returns two sets of coefficients to us. The first are the Poisson regression coefficients, and the second are the coefficients which determine α .

We see that the ZIP model has really changed our estimated rate of drinking. In the base model (without zero inflation), we had assumed the “baseline” rate of drinking was about one drink per weekend, with off campus students drinking at over twice that rate. However, having accounted for zero inflation we now see that the trend is basically reversed. Off campus students appear to be drinking less than on-campus students.

Note that we **cannot** compare these models using a deviance hypothesis test, since the models are non-nested. Other statistical tools exist for this comparison (eg. a Vuong test), but these are beyond the scope of the course.