STAT 226 Lecture 22-24
Generalized Linear Models For Count Data

Yibi Huang

Outine

- Review of Poisson Distributions
- Section 3.3 GLMs for Poisson Response (Counts) Data
- Section 7.6.1 Models for Rates
- Section 3.3.4 Overdispersion
  Section 7.6.3 Negative Binomial Regression

Review of Poisson Distributions

Review of Poisson Distributions

A random variable Y has a Poisson distribution with parameter $\lambda > 0$ if

$$P(Y = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots$$

denoted as

$$Y \sim \text{Poisson}(\lambda).$$

One can show that

$$E[Y] = \lambda, \quad \text{Var}(Y) = \lambda \Rightarrow \text{SD}(Y) = \sqrt{\lambda}.$$

Poisson Approximation to Binomial

If $Y \sim \text{binomial}(n, p)$ with *huge* n and *tiny* p such that np moderate, then

$$Y \text{ approx.} \sim \text{Poisson}(np).$$

The following shows the values of $\mathrm{P}(Y = k)$, $k = 0, 1, 2, 3, 4, 5$ for

$$Y \sim \text{Binomial}(n = 50,\ p = 0.03),\ \text{and}$$
$$Y \sim \text{Poisson}(\lambda = 50 \times 0.03 = 1.5).$$

```
dbinom(0:5, size=50, p=0.03)          # Binomial(n=50, p=0.03)
[1] 0.21807 0.33721 0.25552 0.12644 0.04595 0.01307
dpois(0:5, lambda = 50*0.03)          # Poisson(lambda = 50*0.03)
[1] 0.22313 0.33470 0.25102 0.12551 0.04707 0.01412
```

Example (Fatalities From Horse Kicks)

The # of deaths in a year resulted from being kicked by a horse or mule was recorded for each of 10 corps of Prussian cavalry over a period of 20 years, giving 200 corps-years worth of data[1].

# of Deaths (in a corp in a year)	0	1	2	3	4	Total
Frequency	109	65	22	3	1	200

The count of deaths due to horse kicks in a corp in a given year may have a Poisson distribution because

- $p = P(\text{a soldier died from horsekicks in a given year}) \approx 0$;
- $n =$ # of soldiers in a corp was large (100's or 1000's);
- whether a soldier was kicked was (at least nearly) independent of whether others were kicked

[1] p.45, John Rice, *Mathematical Statistics and Data Analysis*, 3ed

Example (Fatalities From Horse Kicks — Cont'd)

- Suppose all 10 corps had the same n and p throughout the 20 year period. Then we may assume that the 200 counts all have the Poisson distn. with the same rate $\lambda = np$.

Example (Fatalities From Horse Kicks — Cont'd)

- Suppose all 10 corps had the same n and p throughout the 20 year period. Then we may assume that the 200 counts all have the Poisson distn. with the same rate $\lambda = np$.
- How to estimate λ?

Example (Fatalities From Horse Kicks — Cont'd)

- Suppose all 10 corps had the same n and p throughout the 20 year period. Then we may assume that the 200 counts all have the Poisson distn. with the same rate $\lambda = np$.
- How to estimate λ?
- MLE for the rate λ of a Poisson distribution is the **sample mean** $\overline{Y}$.

Example (Fatalities From Horse Kicks — Cont'd)

- Suppose all 10 corps had the same n and p throughout the 20 year period. Then we may assume that the 200 counts all have the Poisson distn. with the same rate $\lambda = np$.
- How to estimate λ?
- MLE for the rate λ of a Poisson distribution is the **sample mean** $\overline{Y}$.
- So for the horsekick data:

# of Deaths (in a corp in a year)	0	1	2	3	4	Total
Frequency	109	65	22	3	1	200

the MLE for λ is

$$\widehat{\lambda} = \frac{0 \times 109 + 1 \times 65 + 2 \times 22 + 3 \times 3 + 4 \times 1}{200} = 0.61$$

Example (Fatalities From Horse Kicks — Cont'd)

The fitted Poisson probability to have k deaths from horsekicks is

$$\mathrm{P}(Y = k) = e^{-\widehat{\lambda}}\frac{\widehat{\lambda}^k}{k!} = e^{-0.61}\frac{(0.61)^k}{k!}, \quad ,k = 0, 1, 2, \ldots$$

k	Observed Frequency	Fitted Poisson Freq. $= 200 \times P(Y = k)$
0	109	108.7
1	65	66.3
2	22	20.2
3	3	4.1
4	1	0.6
Total	200	199.9

```
200*dpois(0:4, 0.61)
[1] 108.6702  66.2888  20.2181   4.1110   0.6269
```

When Do Poisson Distributions Come Up?

Variables that are generally Poisson:

- # of misprints on a page of a book
- # of calls coming into an exchange during a unit of time (if the exchange services a large number of customers who act more or less independently.)
- # of people in a community who survive to age 100
- # of customers entering a post office on a given day
- # of vehicles that pass a marker on a roadway during a unit of time (for light traffic only. In heavy traffic, however, one vehicle's movement may influence another)

GLMs for Poisson Response Data

GLMs for Poisson Response Data

Assume the response $Y \sim \text{Poisson}(\mu(x))$, where x is an explanatory variable.

Commonly used link functions for Poisson distributions are

- identity link: $\mu(x) = \alpha + \beta x$
  - sometimes problematic because $\mu(x)$ must be > 0, but $\alpha + \beta x$ may not
- log link: $\log(\mu(x)) = \alpha + \beta x \quad \Longleftrightarrow \quad \mu(x) = e^{\alpha + \beta x}$.
  - $\mu(x) > 0$ always
  - Whenever x increases by 1 unit, $\mu(x)$ is multiplied by e^{β}

Loglinear models use Poisson with log link

Inference of Parameters and Goodness of Fit

- Wald, LR tests and CIs for β's work as in logistic models
- **Goodness of fit** (*Grouped data* only):

$$\text{Deviance} = G^2 = 2\sum_i y_i \log\left(\frac{y_i}{\widehat{\mu}_i}\right) = -2(L_M - L_S)$$

$$\text{Pearson's chi-squared} = X^2 = \sum_i \frac{(y_i - \widehat{\mu}_i)^2}{\widehat{\mu}_i}$$

G^2 and X^2 are approx. $\sim \chi^2_{n-p}$, when all $\widehat{\mu}_i$'s are large (≥ 10), where

- n = num. of rows (different for grouped & ungrouped data)
- p = num. of parameters in the model.

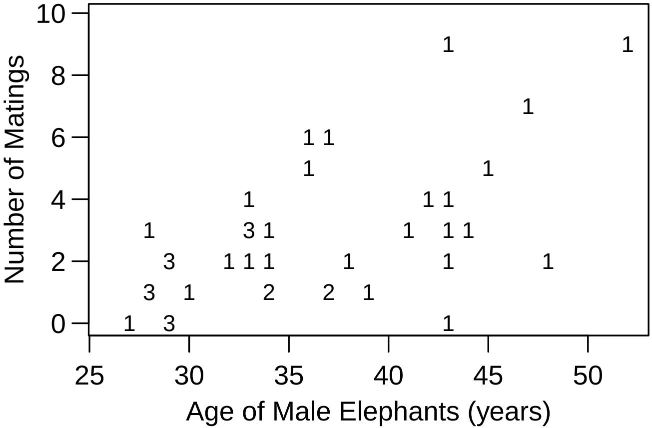
Example (Mating and Age of Male Elephants)

Joyce Poole studied a population of African elephants in Amboseli National Park, Kenya, for 8 years[2].

- `Matings` = # of successful matings in the 8 years of 41 male elephants
- `Age` = estimated age of the male elephant at beginning of the study

Age	Matings	Age	Matings	Age	Matings	Age	Matings
27	0	30	1	36	5	43	3
28	1	32	2	36	6	43	4
28	1	33	4	37	1	43	9
28	1	33	3	37	1	44	3
28	3	33	3	37	6	45	5
29	0	33	3	38	2	47	7
29	0	33	2	39	1	48	2
29	0	34	1	41	3	52	9
29	2	34	1	42	4		
29	2	34	2	43	0		
29	2	34	3	43	2		

[2] p.673, F. Ramsey & D. Schafer, *The Statistical Sleuth*



On the plot, “3” means there are 3 points at the same location.

Example (Elephants) — Identity Link

Let Y = number of successful matings $\sim$ Poisson(μ);

$$\text{Model 1} : \mu = \alpha + \beta\text{Age} \qquad \text{(identity link)}$$

```r
Age = c(27,28,28,28,28,29,29,29,29,29,29,30,32,33,33,33,33,33,34,34,
      34,34,36,36,37,37,37,38,39,41,42,43,43,43,43,43,44,45,47,48,52)
Matings = c(0,1,1,1,3,0,0,0,2,2,2,1,2,4,3,3,3,2,1,1,2,3,
            5,6,1,1,6,2,1,3,4,0,2,3,4,9,3,5,7,2,9)
eleph.id = glm(Matings ~ Age, family=poisson(link="identity"))
```

```r
coef(summary(eleph.id))
            Estimate Std. Error z value     Pr(>|z|)
(Intercept)  -4.5520    1.33916  -3.399 0.0006758549
Age           0.2018    0.04023   5.016 0.0000005289
```

Fitted model 1: $\widehat{\mu} = \widehat{\alpha} + \widehat{\beta}\text{Age} = -4.55 + 0.20$ Age

About $\widehat{\beta} = 0.20$ more matings on average if the male was 1 year older

Example (Elephants) — Log Link

$$\text{Model 2}: \log(\mu) = \alpha + \beta\text{Age} \qquad \text{(log link)}$$

```
eleph.log = glm(Matings ~ Age, family=poisson(link="log"))
coef(summary(eleph.log))
             Estimate Std. Error z value     Pr(>|z|)
(Intercept) -1.58201    0.54462  -2.905 0.0036750516
Age          0.06869    0.01375   4.997 0.0000005812
```

Fitted model 2: $\log(\widehat{\mu}) = -1.582 + 0.0687\text{Age}$

$$\widehat{\mu} = \exp(-1.582 + 0.0687\text{Age}) = 0.205(1.071)^{\text{Age}}$$

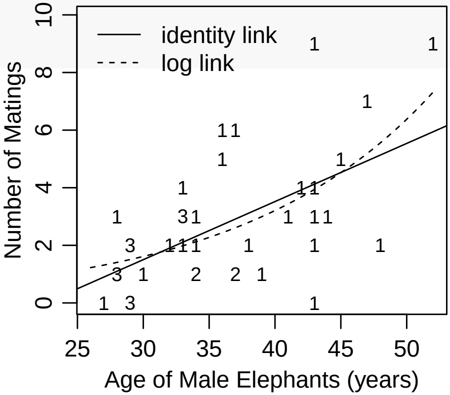
- expected 7.1% increase in number of matings for every extra year in age
- for 40 yr-old males, the expected number of matings is $\widehat{\mu} = \exp(-1.582 + 0.0687(40)) \approx 3.2.$

Which Model Better Fits the Data?

Based on log-likelihood, Model `eleph.id` seems slightly better.

```
logLik(eleph.id)
'log Lik.' -75.75 (df=2)
logLik(eleph.log)
'log Lik.' -76.23 (df=2)
```

- Goodness of fit tests are not appropriate for ungrouped data
- Based on scatter plot...

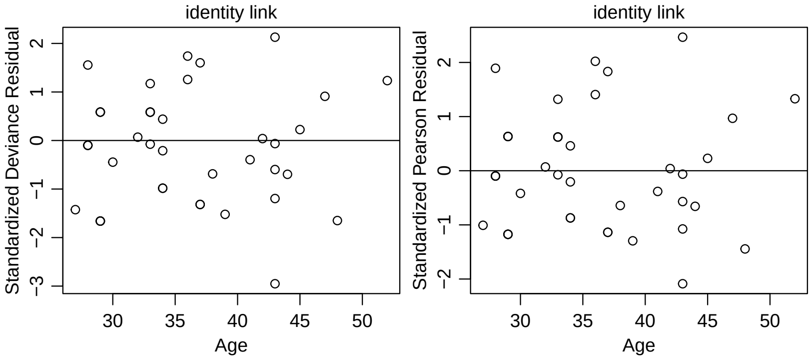


Residuals

- Deviance residual: $d_i = \text{sign}(y_i - \widehat{\mu}_i) \sqrt{2\left[y_i \log(y_i/\widehat{\mu}_i) - y_i + \widehat{\mu}_i\right]}$
- Pearson's residual: $e_i = \dfrac{y_i - \widehat{\mu}_i}{\sqrt{\widehat{\mu}_i}}$
- Standardized Pearson's residual = $e_i / \sqrt{1 - h_i}$
- Standardized Deviance residual = $d_i / \sqrt{1 - h_i}$
  where h_i = leverage of ith observation
- potential outlier if |standardized residual| > 2 or 3
- R function `residuals()` gives deviance residuals by default, and Pearson residuals with option `type="pearson"`.
- R function `rstandard()` gives standardized deviance residuals by default, and standardized Pearson residuals with option `type="pearson"`.

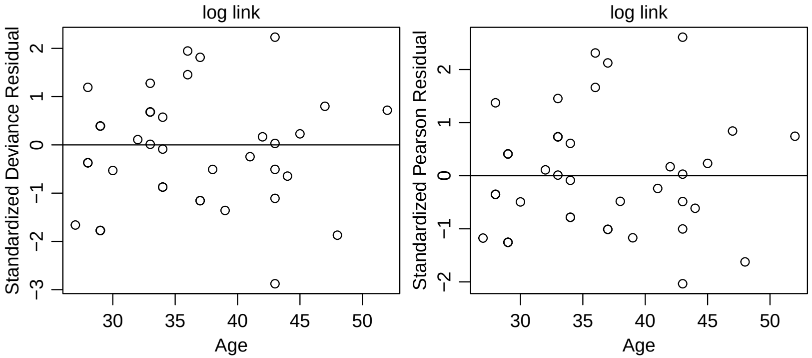
Residual Plots for Model w/ Identity Link

```
plot(Age, rstandard(eleph.id),
     ylab="Standardized Deviance Residual", main="identity link")
abline(h=0)
plot(Age, rstandard(eleph.id, type="pearson"),
     ylab="Standardized Pearson Residual", main = "identity link")
abline(h=0)
```



Residual Plots for Model w/ Log-Link

```
plot(Age, rstandard(eleph.log),
     ylab="Standardized Deviance Residual", main="log link")
abline(h=0)
plot(Age, rstandard(eleph.log, type="pearson"),
     ylab="Standardized Pearson Residual", main = "log link")
abline(h=0)
```



Section 7.6.1 Count Regression Modeling of Rate Data

Section 7.6.1 Count Regression Modeling of Rate Data

When events occur *over time, space*, or some other index of size, models can focus on the *rate* rather than the *count* at which the events occur.

Example: Which city is safer to live?

- City A: 200 homicides last year, 1 M population
- City B: 300 homicides last year, 2 M population

Section 7.6.1 Count Regression Modeling of Rate Data

When events occur *over time, space*, or some other index of size, models can focus on the *rate* rather than the *count* at which the events occur.

Example: Which city is safer to live?

- City A: 200 homicides last year, 1 M population
- City B: 300 homicides last year, 2 M population

It makes more sense to compare *rates* than *counts*:

- City B had more homicide cases
- City B had fewer homicides per million of population

Log-Linear Models for Rate Data

Let y = count of homicides a year in a city with $\boxed{\text{population } t}$.

Assume $y \sim \text{Poisson}(\mu)$

Instead of modeling the mean count of homicides μ, better modeling how the *rate μ/t*

$$\text{rate} = \frac{\mu}{t} = \frac{\text{expected \# of homicides}}{\text{population size}}$$

changes with the explanatory variable $x =$ unemployment rate, etc.

Here $t =$ population size is called the *index*.

Loglinear model:

$$\log\left(\frac{\mu}{t}\right) = \alpha + \beta x \quad \Rightarrow \quad \log(\mu) = \log(t) + \alpha + \beta x$$

$\log(t)$ is an *offset*, which means a term in the model with a *known coefficient 1*.

Example (British Train Accidents in 1975-2003)

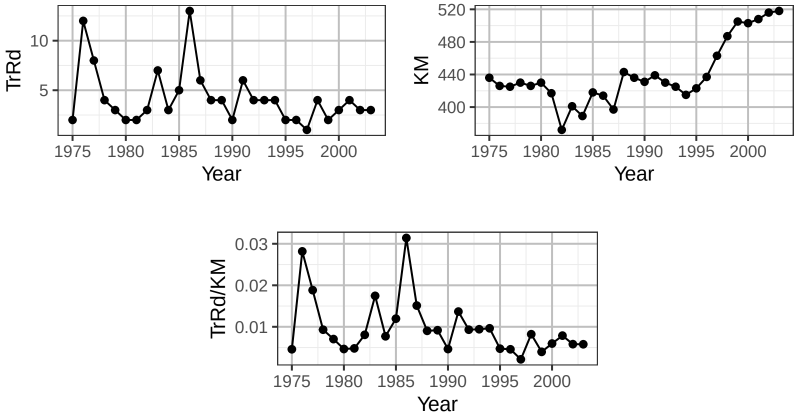
```
trains = read.table(
  "http://www.stat.uchicago.edu/~yibi/s226/traincollisions.dat",
  header=T)
```

```
   Year  KM Train TrRd
1  2003 518     0    3
2  2002 516     1    3
3  2001 508     0    4
...
28 1976 426     2   12
29 1975 436     5    2
```

Variables

- `TrRd` = # of collisions betw. trains and road vehicles that year
- `KM` = total mileage traveled by trains during the year in millions of kilometers

Have collisions between trains and road vehicles become more prevalent over time?



- Total number of train-km (in millions) varies from year to year.
- Model annual rate of train-road collisions per million train-km with the *index* t = `KM` = annual number of train-km, and x = `Year`

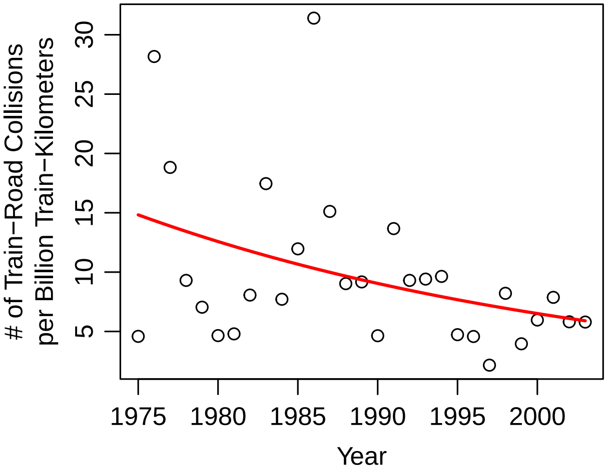
```r
trains1 = glm(TrRd ~ Year, offset = log(KM),
              family=poisson, data=trains)
summary(trains1)$coef
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 60.80160   21.38001   2.844 0.004457
Year        -0.03292    0.01076  -3.060 0.002217
```

Fitted Model: $\log\left(\frac{\widehat{\mu}}{t}\right) = 60.8016 - 0.0329$ Year

$$\text{rate} = \frac{\widehat{\mu}}{t} = \exp(60.8016 - 0.0329\text{Year})$$

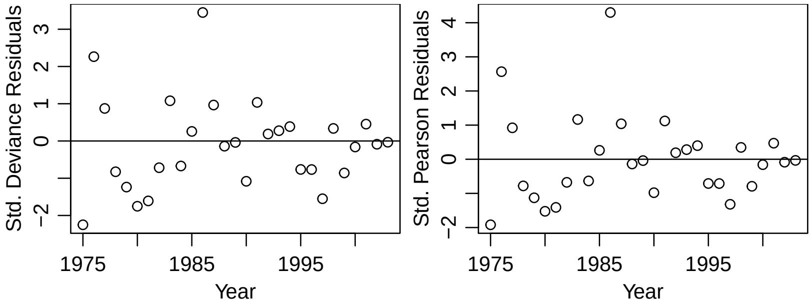
- $\exp(-0.0329) \approx 0.9676$
  $\Rightarrow$ Rate estimated to decrease by 3.2% per yr in 1975-2003
- Est. rate for $x = 1975$ is $e^{60.8016-0.0329(1975)} \approx 0.0148$ per million train-km (15 per billion train-km).
- Est. rate for $x = 2003$ is $e^{60.8016-0.0329(2003)} \approx 0.0059$ per million train-km (6 per billion train-km).

```r
plot(trains$Year, 1000*trains$TrRd/trains$KM,xlab="Year",
     ylab="# of Train-Road Collisions\nper Billion Train-Kilometers")
curve(1000*exp(trains1$coef[1]+trains1$coef[2]*x), add=T, col="red")
```



Train Data — Standardized Deviance & Pearson Residuals

```
plot(trains$Year, rstandard(trains1),
     xlab="Year", ylab="Std. Deviance Residuals")
abline(h=0)
plot(trains$Year, rstandard(trains1,type="pearson"),
     xlab="Year", ylab="Std. Pearson Residuals")
abline(h=0)
```



There were 13 train-road collisions in 1986, far above the fitted mean of 4.3 for that year.

Linear (Additive) Models for Rate Data

For $y \sim$ Poisson(μ) with *index t*, the loglinear model

$$\log\left(\frac{\mu}{t}\right) = \alpha + \beta x$$

assumes the effect of the explanatory variable x on the response to be **multiplicative**.

Alternatively, if we want the effect to be **additive**,

$$\frac{\mu}{t} = \alpha + \beta x$$
$$\Leftrightarrow \quad \mu = \alpha t + \beta t x$$

we may fit a GLM model with *identity link*, using t and tx as explanatory variables and with *no intercept* or offset terms.

Train Data — Identity Link

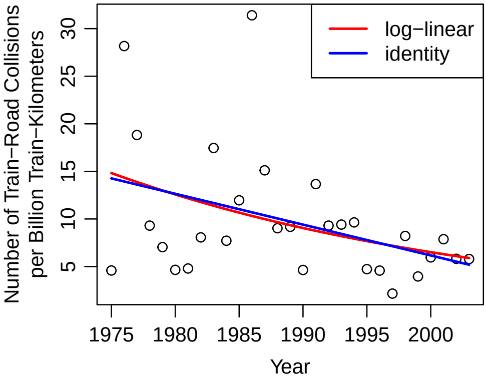
index t = KM = annual num. of train-km, x = year

```
trains2 = glm(TrRd ~ -1 + KM + I(KM*Year),
              family=poisson(link="identity"), data=trains)
summary(trains2)$coef
                Estimate Std. Error z value  Pr(>|z|)
KM             0.6539613 0.19770270   3.308 0.0009403
I(KM * Year) -0.0003239 0.00009924  -3.264 0.0010997
```

Fitted Model: $\widehat{\text{rate}} = \frac{\widehat{\mu}}{t} = \frac{\widehat{\mu}}{\text{KM}} \approx 0.654 - 0.000324\text{Year}$

- Estimated rate decreases by 0.00032 per million km (0.32 per billion km) per yr from 1975 to 2003.
- Est. rate for 1975 is $0.654 - 0.0003239 \times 1975 \approx 0.0143$ per million km (14.3 per billion km).
- Est. rate for 2003 is $0.654 - 0.0003239 \times 2003 \approx 0.0052$ per million km (5.2 per billion km).

```
plot(trains$Year, 1000*trains$TrRd/trains$KM,xlab="Year",
  ylab="Number of Train-Road Collisions\nper Billion Train-Kilometers")
curve(1000*exp(trains1$coef[1]+trains1$coef[2]*x),add=T, col="red")
curve(1000*trains2$coef[1]+1000*trains2$coef[2]*x,add=T, col="blue")
legend("topright", c("log-linear","identity"), lty=1:2)
```



The loglinear fit and the linear fit (identity link) are nearly identical.

Overdispersion & Negative Binomial Regression

Section 3.3.4 Overdispersion: Greater Variability than Expected

- One of the defining characteristics of Poisson regression is its lack of a parameter for **variability**:

$$E(Y) = \text{Var}(Y),$$

  and no parameter is available to adjust that relationship
- In practice, when working with Poisson regression, it is often the case that the variability of y_i about $\widehat{\mu}_i$ is larger than what $\widehat{\mu}_i$ predicts
- This implies that there is more variability around the model's fitted values than is consistent with the Poisson distribution
- This phenomenon is **overdispersion**.

Example (Known Victims of Homicide)

A recent General Social Survey asked subjects,

> *"Within the past 12 months, how many people have you known personally that were victims of homicide?"*

Number of Victims	0	1	2	3	4	5	6	Total
Black Subjects	119	16	12	7	3	2	0	159
White Subjects	1070	60	14	4	0	0	1	1149

If fit a Poisson distribution to the data from blacks, MLE for the Poisson mean λ is the sample mean

$$\widehat{\lambda} = \frac{0 \cdot 119 + 1 \cdot 16 + 2 \cdot 12 + \cdots + 6 \cdot 0}{159} = \frac{83}{159} \approx 0.522$$

Fitted $P(Y = k)$ is $e^{-\frac{83}{159}} \left(\frac{83}{159}\right)^k / k!,\ k = 0, 1, 2, \ldots.$

```
159*dpois(0:6, lambda = 83/159)
```

```
[1] 94.34 49.25 12.85  2.24  0.29  0.03  0.00
```

Example (Known Victims of Homicide)

Num. of Victims	0	1	2	3	4	5	6	Total	Mean	Variance
Black	119	16	12	7	3	2	0	159	0.522	1.150
White	1070	60	14	4	0	0	1	1149	0.092	0.155

Likewise, MLE of λ for whites is

$$\widehat{\lambda} = \frac{0 \cdot 1070 + 1 \cdot 60 + 2 \cdot 14 + \cdots + 6 \cdot 1}{1149} = \frac{106}{1149} \approx 0.092$$

Fitted $P(Y = k)$ is $e^{-\frac{106}{1149}} \left(\frac{106}{1149}\right)^k / k!,\ k = 0, 1, 2, \ldots.$

```
round(1149*dpois(0:6, lambda = 106/1149), 3) # fitted Poisson counts.
[1] 1047.743   96.659    4.459    0.137    0.003    0.000    0.000
```

- Too many 0's and too many large counts for both races than expected if the data are Poisson
- Poor Poisson fits are NOT surprising from the large discrepancies between sample mean and sample variance.

Common Causes of Overdispersion

- Subject heterogeneity
  - subjects have different μ
    e.g., people of the same race might have different mean in the # of known victims of homicide as crime rate may vary from region to region.
  - there are important predictors not included in the model
- Observations are not independent – clustering

Negative Binomial Distributions

If Y has a negative binomial distribution, with mean μ and dispersion parameter $D = 1/\theta$, then

$$P(Y = k) = \frac{\Gamma(k + \theta)}{k!\Gamma(\theta)} \left(\frac{\theta}{\mu + \theta}\right)^{\theta} \left(\frac{\mu}{\mu + \theta}\right)^{k}, \quad k = 0, 1, 2, \ldots$$

One can show that

$$\mathrm{E}[Y] = \mu, \quad \mathrm{Var}(Y) = \mu + \frac{\mu^2}{\theta} = \mu + D\mu^2.$$

- As $D = 1/\theta \downarrow 0$, negative binomial $\rightarrow$ Poisson.
- Negative binomial is a gamma mixture of Poissons, where the Poisson mean varies according to a gamma distribution.
- MLE for μ is the sample mean.
  MLE for θ has no close form formula.

Poisson and Neg. Bin Models for Homicide Data

Data:	$Y_{b,1}, Y_{b,2}, \ldots, Y_{b,159}$	answers from black subjects
	$Y_{w,1}, Y_{w,2}, \ldots\ldots, Y_{w,1149}$	answers from white subjects

Poisson Model: $Y_{b,j} \sim \text{Poisson}(\mu_b), \quad Y_{w,j} \sim \text{Poisson}(\mu_w)$

Neg. Bin. Model: $Y_{b,j} \sim \text{NB}(\mu_b, \theta), \quad Y_{w,j} \sim \text{NB}(\mu_w, \theta)$

Goal: Test whether $\mu_b = \mu_w$.

Equivalent to test $\beta = 0$ in the log-linear model.

$$\log(\mu) = \alpha + \beta x, \quad x = \begin{cases} 1 \text{ if black} \\ 0 \text{ if white,} \end{cases}$$

Note $\mu_b = e^{\alpha+\beta}$, $\mu_w = e^{\alpha}$. So $e^{\beta} = \mu_b/\mu_w$.

Poisson and Neg. Bin Models for Homicide Data

Can fit Negative binomial regression models using `glm.nb()` in the `MASS` package.

```
nvics = c(0:6,0:6)
race = c(rep("Black", 7),rep("White",7))
freq = c(119,16,12,7,3,2,0,1070,60,14,4,0,0,1)
```

```
   nvics  race freq
1      0 Black  119
2      1 Black   16
3      2 Black   12
... (omit) ...
13     5 White    0
14     6 White    1
```

```
race = factor(race, levels=c("White","Black"))
hom.poi = glm(nvics ~ race, weights=freq, family=poisson)
library(MASS)
hom.nb = glm.nb(nvics ~ race, weights=freq)
```

Example (Known Victims of Homicide) — Poisson Fits

```
summary(hom.poi)
```

```
Call:
glm(formula = nvics ~ race, family = poisson, weights = freq)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.3832     0.0971   -24.5   <2e-16
raceBlack     1.7331     0.1466    11.8   <2e-16

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 962.80  on 10  degrees of freedom
Residual deviance: 844.71  on  9  degrees of freedom
AIC: 1122

Number of Fisher Scoring iterations: 6
```

Example (Known Victims of Homicide) — Neg. Binomial

```r
summary(hom.nb)
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -2.383      0.117  -20.33  < 2e-16
raceBlack      1.733      0.238    7.27  3.7e-13
---
    Null deviance: 471.57  on 10  degrees of freedom
Residual deviance: 412.60  on  9  degrees of freedom
AIC: 1002

Number of Fisher Scoring iterations: 1
              Theta:  0.2023
          Std. Err.:  0.0409
 2 x log-likelihood:  -995.7980
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -2.383      0.117  -20.33  < 2e-16
raceBlack      1.733      0.238    7.27  3.7e-13
```

- Fitted values given by the Neg. Bin model are simply the sample means — $\exp(-2.383) = 0.0922$ (= $\frac{106}{1149}$) for whites and $\exp(-2.383 + 1.733) = 0.522$ (= $\frac{83}{159}$) for blacks.
- Estimated common dispersion parameter is $\widehat{\theta} = 0.2023$ with SE = 0.0409.
- Fitted $P(Y = k)$ is

$$\frac{\Gamma(k+\widehat{\theta})}{k!\Gamma(\widehat{\theta})}\left(\frac{\widehat{\theta}}{\widehat{\mu}+\widehat{\theta}}\right)^{\theta}\left(\frac{\widehat{\mu}}{\widehat{\mu}+\widehat{\theta}}\right)^{k}, \text{ where } \widehat{\mu} = \begin{cases} \frac{83}{159} & \text{for blacks} \\ \frac{106}{1149} & \text{for whites.} \end{cases}$$

- Textbook uses $D = 1/\theta$ as the dispersion parameter, estimated as $\widehat{D} = 1/\widehat{\theta} = 1/0.2023 \approx 4.94$.

Black Subjects

Num. of Victims	0	1	2	3	4	5	6	Total
observed freq.	119	16	12	7	3	2	0	159
relative freq.	0.748	0.101	0.075	0.044	0.019	0.013	0	1
poisson fit	0.593	0.310	0.081	0.014	0.002	0.000	0.000	1
neg. bin.fit	0.773	0.113	0.049	0.026	0.015	0.009	0.006	0.991

White Subjects:

num. of victims	0	1	2	3	4	5	6	Total
observed freq.	1070	60	14	4	0	0	1	1149
relative freq.	0.931	0.052	0.012	0.003	0.000	0.000	0.001	0.999
poisson fit	0.912	0.084	0.004	0.000	0.000	0.000	0.000	1
neg. bin.fit	0.927	0.059	0.011	0.003	0.001	0.000	0.000	1.001

```r
# neg. bin fit
round(dnbinom(0:6, size = hom.nb$theta, mu = 83/159),3)  # black
[1] 0.773 0.113 0.049 0.026 0.015 0.009 0.006
round(dnbinom(0:6,size = hom.nb$theta, mu=106/1149),3)   # white
[1] 0.927 0.059 0.011 0.003 0.001 0.000 0.000
```

Example (Known Victims of Homicide)

$$\text{Model:} \quad \log(\mu) = \alpha + \beta x, \quad x = \begin{cases} 1 \text{ if black} \\ 0 \text{ if white,} \end{cases}$$

Model	$\widehat{\alpha}$	$\widehat{\beta}$	SE($\widehat{\beta}$)	Wald 95% CI for $e^{\beta} = \mu_B/\mu_A$
Poisson	-2.38	1.73	0.147	$\exp(1.73 \pm 1.96 \cdot 0.147) = (4.24, 7.54)$
Neg. Binom.	-2.38	1.73	0.238	$\exp(1.73 \pm 1.96 \cdot 0.238) = (3.54, 9.03)$

Poisson and negative binomial models give

Example (Known Victims of Homicide)

$$\text{Model:} \quad \log(\mu) = \alpha + \beta x, \quad x = \begin{cases} 1 \text{ if black} \\ 0 \text{ if white,} \end{cases}$$

Model	$\widehat{\alpha}$	$\widehat{\beta}$	SE($\widehat{\beta}$)	Wald 95% CI for $e^{\beta} = \mu_B/\mu_A$
Poisson	-2.38	1.73	0.147	$\exp(1.73 \pm 1.96 \cdot 0.147) = (4.24, 7.54)$
Neg. Binom.	-2.38	1.73	0.238	$\exp(1.73 \pm 1.96 \cdot 0.238) = (3.54, 9.03)$

Poisson and negative binomial models give

- identical estimates for coefficients
  (this data set only, not always the case)

Example (Known Victims of Homicide)

$$\text{Model:} \quad \log(\mu) = \alpha + \beta x, \quad x = \begin{cases} 1 \text{ if black} \\ 0 \text{ if white,} \end{cases}$$

Model	$\widehat{\alpha}$	$\widehat{\beta}$	SE($\widehat{\beta}$)	Wald 95% CI for $e^{\beta} = \mu_B/\mu_A$
Poisson	-2.38	1.73	0.147	$\exp(1.73 \pm 1.96 \cdot 0.147) = (4.24, 7.54)$
Neg. Binom.	-2.38	1.73	0.238	$\exp(1.73 \pm 1.96 \cdot 0.238) = (3.54, 9.03)$

Poisson and negative binomial models give

- identical estimates for coefficients
  (this data set only, not always the case)
- but different SEs for $\widehat{\beta}$ (Neg. Binom. gives bigger SE)

Example (Known Victims of Homicide)

$$\text{Model:} \quad \log(\mu) = \alpha + \beta x, \quad x = \begin{cases} 1 \text{ if black} \\ 0 \text{ if white,} \end{cases}$$

Model	$\widehat{\alpha}$	$\widehat{\beta}$	SE($\widehat{\beta}$)	Wald 95% CI for $e^{\beta} = \mu_B/\mu_A$
Poisson	-2.38	1.73	0.147	$\exp(1.73 \pm 1.96 \cdot 0.147) = (4.24, 7.54)$
Neg. Binom.	-2.38	1.73	0.238	$\exp(1.73 \pm 1.96 \cdot 0.238) = (3.54, 9.03)$

Poisson and negative binomial models give

- identical estimates for coefficients
  (this data set only, not always the case)
- but different SEs for $\widehat{\beta}$ (Neg. Binom. gives bigger SE)

To account for overdispersion, neg. binom. model gives wider Wald CIs (and also wider LR CIs).

Example (Known Victims of Homicide)

$$\text{Model:} \quad \log(\mu) = \alpha + \beta x, \quad x = \begin{cases} 1 \text{ if black} \\ 0 \text{ if white,} \end{cases}$$

Model	$\widehat{\alpha}$	$\widehat{\beta}$	SE($\widehat{\beta}$)	Wald 95% CI for $e^{\beta} = \mu_B/\mu_A$
Poisson	-2.38	1.73	0.147	$\exp(1.73 \pm 1.96 \cdot 0.147) = (4.24, 7.54)$
Neg. Binom.	-2.38	1.73	0.238	$\exp(1.73 \pm 1.96 \cdot 0.238) = (3.54, 9.03)$

Poisson and negative binomial models give

- identical estimates for coefficients
  (this data set only, not always the case)
- but different SEs for $\widehat{\beta}$ (Neg. Binom. gives bigger SE)

To account for overdispersion, neg. binom. model gives wider Wald CIs (and also wider LR CIs).

Remark. Observe $e^{\widehat{\beta}} = e^{1.73} = 5.7$ is the ratio of the two sample means $\bar{y}_{\text{black}}/\bar{y}_{\text{white}} = 0.522/0.092.$

Wald CIs

```
confint.default(hom.poi)
                2.5 % 97.5 %
(Intercept) -2.574 -2.193
raceBlack    1.446  2.020
exp(confint.default(hom.poi))
                 2.5 % 97.5 %
(Intercept) 0.07626 0.1116
raceBlack   4.24557 7.5414
confint.default(hom.nb)
                2.5 % 97.5 %
(Intercept) -2.613 -2.154
raceBlack    1.266  2.201
exp(confint.default(hom.nb))
                 2.5 % 97.5 %
(Intercept) 0.07332 0.1161
raceBlack   3.54571 9.0300
```

Likelihood Ratio CIs

```
confint(hom.poi, "raceBlack")
Waiting for profiling to be done...
 2.5 % 97.5 %
 1.444  2.019
exp(confint(hom.poi, "raceBlack"))
Waiting for profiling to be done...
 2.5 % 97.5 %
 4.236  7.533
confint(hom.nb, "raceBlack")
Waiting for profiling to be done...
 2.5 % 97.5 %
 1.275  2.212
exp(confint(hom.nb, "raceBlack"))
Waiting for profiling to be done...
 2.5 % 97.5 %
 3.578  9.132
```

If Not Taking Overdispersion Into Account . . .

- SEs are underestimated
- CIs will be too narrow
- Significance of variables will be over stated (reported P values are lower than the actual ones)

How to Check for Overdispersion?

- Think about whether overdispersion is likely — e.g., important explanatory variables not available, or dependence in observations.
- Compare the sample variances to the sample means computed for groups of responses with identical explanatory variable values.
- Large deviance relative to its df can be a sign of overdispersion
- Examine residuals to see if a large deviance statistic may be due to outliers
- Large numbers of outliers is usually a sign of overdispersion
- Check standardized residuals and plot them against them fitted values $\widehat{\mu}_i$.

Train Data Revisit

Recall Pearson's residual:

$$e_i = \frac{y_i - \widehat{\mu}_i}{\sqrt{\widehat{\mu}_i}}$$

If no overdispersion, then

$$\mathrm{Var}(Y) \approx (y_i - \widehat{\mu}_i)^2 \approx \mathrm{E}(Y) \approx \widehat{\mu}_i$$

So the size of Pearson's residuals should be around 1.

With overdispersion,

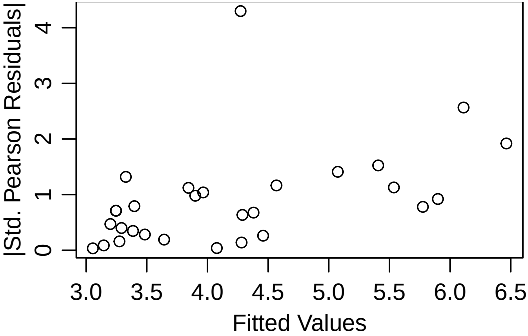
$$\mathrm{Var}(Y) = \mu + D\mu^2$$

then the size of Pearson's residuals may increase with μ.

We may check the plot of the absolute value of (standardized) Pearson's residuals against fitted values $\widehat{\mu}_i$.

Train Data — Checking Overdispersion

```
plot(trains1$fit, abs(rstandard(trains1, type="pearson")),
     xlab="Fitted Values", ylab="|Std. Pearson Residuals|")
```



The size of standardized Pearson's residuals tend to increase with fitted values. This is a sign of overdisperson.

Train Data — Neg. Bin. Model

```r
trains.nb = glm.nb(TrRd ~ Year+offset(log(KM)), data=trains)
summary(trains.nb)
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  62.2960    25.5956    2.43   0.0149
Year         -0.0337     0.0129   -2.61   0.0089
---
    Null deviance: 32.045  on 28  degrees of freedom
Residual deviance: 25.264  on 27  degrees of freedom
AIC: 132.7
Number of Fisher Scoring iterations: 1
              Theta:  10.12
          Std. Err.:  8.00
 2 x log-likelihood:  -126.69
```

For `Year`, the estimated coefficients are similar (0.0337 for neg. bin. fit v.s. 0.032 for Poisson fit), but less significant (P-value = 0.009 in neg. bin. fit v.s. 0.002 in Poisson fit)

Example (British Football Fans)

The table below lists total home field attendance (in 1000s) and the total # of arrests in a season for soccer teams in the Second Division of the British football league.

Team	Attendance	Arrests	Team	Attendance	Arrests
Aston Villa	404	308	Shrewsbury	108	68
Bradford City	286	197	Swindon Town	210	67
Leeds United	443	184	Sheffield Utd	224	60
Bournemouth	169	149	Stoke City	211	57
West Brom	222	132	Barnsley	168	55
Hudderfield	150	126	Millwall	185	44
Middlesbro	321	110	Hull City	158	38
Birmingham	189	101	Manchester City	429	35
Ipswich Town	258	99	Plymouth	226	29
Leicester City	223	81	Reading	150	20
Blackburn	211	79	Oldham	148	19
Crystal Palace	215	78			

Which Teams Had Most Aggressive Fans?

Let Y = # of arrests for a team with total home field attendance t.

If the arrests distributed homogeneously among people attending football games, the expected numbers of arrests $\mu = \mathrm{E}(Y)$ at the home field of a team should be proportional to t = total # of attendance at the home field of that team. Thus, it's reasonable to assume

$$\mu = \mathrm{E}(Y) = \lambda t,$$

where λ = # of arrests per thousand of attendance. Taking logarithm on both sides, we get

$$\log(\mu) = \log(t) + \alpha, \quad \text{where } \alpha = \log(\lambda).$$

Observe that $\log(t)$ is an *offset* in the model equation since its coefficient is a fixed number 1 that we don't have to estimate.

Example (British Football Fans) — Poisson Fit

```
soccer = read.table(
  "http://www.stat.uchicago.edu/~yibi/s226/SoccerGameArrests.dat",
  header=TRUE)
fit.poi = glm(arrests ~ 1, offset = log(attendance), family=poisson,
            data=soccer)
summary(fit.poi)$coef
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.9103    0.02164  -42.07        0
```

Fitted model: $\log(\widehat{\mu}) = -0.9103 + \log(t)$ or

$$\text{Predicted arrests} = \widehat{\mu} = e^{-0.9103} t \approx 0.4024\, t.$$

That is, **there were about 0.4 arrests in every thousands of attendance**.

Example (British Football Fans) — Poisson Fit

```
soccer = read.table(
  "http://www.stat.uchicago.edu/~yibi/s226/SoccerGameArrests.dat",
  header=TRUE)
fit.poi = glm(arrests ~ 1, offset = log(attendance), family=poisson,
              data=soccer)
summary(fit.poi)$coef
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.9103    0.02164  -42.07        0
```

Fitted model: $\log(\widehat{\mu}) = -0.9103 + \log(t)$ or

$$\text{Predicted arrests} = \widehat{\mu} = e^{-0.9103} t \approx 0.4024\, t.$$

That is, **there were about 0.4 arrests in every thousands of attendance**.

Team “Leeds United” had the highest # of attendance $t = 443K$.
Predicted arrests $= \widehat{\mu} \approx 0.4024\, t = 0.4024 \times 443 \approx 178.26$.

```
summary(fit.poi)$coef
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.9103    0.02164  -42.07        0
```

The 95% Wald CI for α is

$$-0.9103 \pm 1.96 \times 0.02164 \approx (-0.9527, -0.8679)$$

and for $\lambda = e^{\alpha}$ is $(e^{-0.9527}, e^{-0.8679}) \approx (0.3857, 0.4198)$.

Interpretation: With 95% confidence, there were about 0.3857 to 0.4198 arrests on average in every thousands of attendance.

Residuals

We can use standardized (deviance or Pearson) residuals to identify teams with far more arrests than expected. (Output in the next 2 pages).

```
cbind(soccer,
      Fit = round(fit.poi$fit,1),
      StdDevRes = round(rstandard(fit.poi),1),
      StdPsnRes=round(rstandard(fit.poi,type="pearson"),1))
```

Several teams have huge std. residuals, over 4, 8, even over 10. Among them, Aston Villa, Bradford City, Bournemouth, West Brom, Hudderfield had far more arrests than expected (residual > 4), and the arrests for Manchester City, Plymouth, Reading, and Oldham were far below expected (residual < -4).

Large numbers of huge std. residuals is a sign of *overdispersion*.

	team	attendance	arrests	Fit	StdDevRes	StdPsnRes	
1	Aston Villa	404	308	162.6	10.5	11.9	<--
2	Bradford City	286	197	115.1	7.1	7.8	<--
3	Leeds United	443	184	178.3	0.4	0.4	
4	Bournemouth	169	149	68.0	8.6	10.0	<--
5	West Brom	222	132	89.3	4.3	4.6	<--
6	Hudderfield	150	126	60.4	7.5	8.6	<--
7	Middlesbro	321	110	129.2	-1.8	-1.7	
8	Birmingham	189	101	76.1	2.8	2.9	
9	Ipswich Town	258	99	103.8	-0.5	-0.5	
10	Leicester City	223	81	89.7	-1.0	-0.9	
11	Blackburn	211	79	84.9	-0.7	-0.7	
12	Crystal Palace	215	78	86.5	-1.0	-0.9	
13	Shrewsbury	108	68	43.5	3.5	3.8	
14	Swindon Town	210	67	84.5	-2.0	-1.9	
15	Sheffield Utd	224	60	90.1	-3.5	-3.2	

	team	attendance	arrests	Fit	StdDevRes	StdPsnRes	
16	Stoke City	211	57	84.9	-3.3	-3.1	
17	Barnsley	168	55	67.6	-1.6	-1.6	
18	Millwall	185	44	74.4	-3.9	-3.6	
19	Hull City	158	38	63.6	-3.5	-3.3	
20	Manchester City	429	35	172.6	-13.3	-10.9	<--
21	Plymouth	226	29	90.9	-7.8	-6.6	<--
22	Reading	150	20	60.4	-6.1	-5.3	<--
23	Oldham	148	19	59.6	-6.2	-5.3	<--

Example (British Football Fans) — Neg. Bin. Fit

```
library(MASS)
fit.nb = glm.nb(arrests ~ 1+ offset(log(attendance)), data=soccer)
summary(fit.nb)
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.905      0.120   -7.55  4.5e-14
--
    Null deviance: 24.15  on 22  degrees of freedom
Residual deviance: 24.15  on 22  degrees of freedom
AIC: 244.2
Number of Fisher Scoring iterations: 1
              Theta:  3.136
          Std. Err.:  0.920
 2 x log-likelihood:  -240.236
```

Example (Football Fans) — NB v.s. Poisson

Model	$\widehat{\alpha}$	SE($\widehat{\alpha}$)	Wald 95% CI for $e^{\alpha} = \lambda$
Poisson	-0.9103	0.0216	$\exp(-0.9103 \pm 1.96 \cdot 0.0216) \approx (0.386, 0.420)$
NB	-0.9052	0.1200	$\exp(-0.9052 \pm 1.96 \cdot 0.1200) \approx (0.320, 0.512)$

Interpretation: With 95% confidence, there were about 0.3197 to 0.5117 arrests on average in every thousands of attendance, based on NB fit.

NB fit gives a much wider CI for e^{α} than Poisson fit.

If we ignore overdispersion, the CI obtained would have a confidence level substantially lower than the nominal 95% level.

Example (Football Fans) — NB v.s. Poisson

Model	$\widehat{\alpha}$	SE($\widehat{\alpha}$)	Wald 95% CI for $e^{\alpha} = \lambda$
Poisson	-0.9103	0.0216	$\exp(-0.9103 \pm 1.96 \cdot 0.0216) \approx (0.386, 0.420)$
NB	-0.9052	0.1200	$\exp(-0.9052 \pm 1.96 \cdot 0.1200) \approx (0.320, 0.512)$

Interpretation: With 95% confidence, there were about 0.3197 to 0.5117 arrests on average in every thousands of attendance, based on NB fit.

NB fit gives a much wider CI for e^{α} than Poisson fit.

If we ignore overdispersion, the CI obtained would have a confidence level substantially lower than the nominal 95% level.

Example (Football Fans) — Evidence of Overdispersion

- Large discrepancy in SE($\widehat{\alpha}$) (0.12 by NB, 0.0216 by Poisson)
- Too many huge standardized residuals
- Huge size of deviance compared to its df: The deviance of the Poisson model is over 30 times of its df 22

```
deviance(fit.poi)
[1] 669.4
df.residual(fit.poi)
[1] 22
deviance(fit.nb)
[1] 24.15
df.residual(fit.nb)
[1] 22
```

- R does not report the estimate and SE of the dispersion parameter D, but of its inverse $\theta = 1/D$, which is 3.136,
  $\Rightarrow \widehat{D} = 1/\widehat{\theta} = 1/3.136 \approx 0.319$.
  R gives SE($\widehat{\theta}$) $= 0.920$, but SE($\widehat{D}$) is not available.

Standardized Deviance & Pearson Residuals For NB Fit

```r
cbind(soccer,
      Fit = round(fit.nb$fit,1),
      StdDevRes = round(rstandard(fit.nb),2),
      StdPsnRes = round(rstandard(fit.nb,type="pearson"),2))
```

```
              team attendance arrests   Fit StdDevRes StdPsnRes
1      Aston Villa        404     308 163.4      1.27      1.59
2    Bradford City        286     197 115.7      1.05      1.26
3     Leeds United        443     184 179.2      0.05      0.05
4      Bournemouth        169     149  68.4      1.59      2.09 <--
5        West Brom        222     132  89.8      0.73      0.84
6      Hudderfield        150     126  60.7      1.48      1.90
7       Middlesbro        321     110 129.8     -0.29     -0.27
8       Birmingham        189     101  76.4      0.52      0.57
9     Ipswich Town        258      99 104.4     -0.09     -0.09
10  Leicester City        223      81  90.2     -0.19     -0.18
```

11	Blackburn	211	79	85.3	-0.14	-0.13	
12	Crystal Palace	215	78	87.0	-0.19	-0.18	
13	Shrewsbury	108	68	43.7	0.84	0.97	
14	Swindon Town	210	67	84.9	-0.41	-0.38	
15	Sheffield Utd	224	60	90.6	-0.68	-0.60	
16	Stoke City	211	57	85.3	-0.67	-0.59	
17	Barnsley	168	55	68.0	-0.36	-0.34	
18	Millwall	185	44	74.8	-0.86	-0.73	
19	Hull City	158	38	63.9	-0.84	-0.72	
20	Manchester City	429	35	173.5	-2.26	-1.43	<--
21	Plymouth	226	29	91.4	-1.70	-1.22	
22	Reading	150	20	60.7	-1.63	-1.18	
23	Oldham	148	19	59.9	-1.68	-1.20	

Almost all teams have std. residuals < 2 after accounting for overdispersion.

- Bournemouth seemed to have too many arrests than expected. Its std. pearson residual is 2.09, but its std deviance residual is only 1.59.
- Manchester City had 39 arrests only, far below its fitted value 173.5. The std deviance residual is -2.26 but its std pearson residual is -1.43.

Neither team seemed to be an extreme outlier for the negative binomial model.