Section 3.3 Generalized Linear Models For Count Data

Outline
- Review of Poisson Distributions
- GLMs for Poisson Response Data
- Models for Rates
- Overdispersion and Negative Binomial Regression

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

Poisson Approximation to Binomial

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

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

The following shows the values of \( P(Y = k) \), \( k = 0, 1, 2, \ldots, 8 \) for

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

Example (Fatalities From Horse Kicks)

The number of fatalities in a year that 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\(^1\).

<table>
<thead>
<tr>
<th># of Deaths (in a corp in a year)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>109</td>
<td>65</td>
<td>22</td>
<td>3</td>
<td>1</td>
<td>200</td>
</tr>
</tbody>
</table>

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

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

\(^1\) von Bortkiewicz (1898) \textit{Das Gesetz der Kleinen Zahlen}. Leipzig: Teubner.
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 $\lambda$?

- MLE for the rate $\lambda$ of a Poisson distribution is the sample mean $\bar{Y}$.

- So for the horsekick data:

<table>
<thead>
<tr>
<th># of Deaths (in a corp in a year)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>109</td>
</tr>
<tr>
<td>1</td>
<td>65</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
</tr>
</tbody>
</table>

the MLE for $\lambda$ is

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

When 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)

The fitted Poisson probability to have $k$ deaths from horsekicks is

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

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: $\log(\mu(x)) = \alpha + \beta x$ ⇐⇒ $\mu(x) = e^{\alpha + \beta x}$
  - $\mu(x) > 0$ always
  - Whenever $x$ increases by 1 unit, $\mu(x)$ is multiplied by $e^\beta$

Loglinear models use Poisson with log link.
Inference of Parameters

- Wald, LR tests and CIs for $\beta$'s work as in logistic models
- Goodness of fit:

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

Pearson’s chi-squared is:

$$X^2 = 2 \sum_i \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$$

$G^2$ and $X^2$ are approx. $\chi^2_{n-p}$, when all $\hat{\mu}_i$’s are large ($\geq 10$), where $n =$ num. of observations, and $p =$ 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$^2$:

- Response: number of successful matings in the 8 years of 41 male elephants.
- Predictor: estimated ages of the male elephants at beginning of the study.

<table>
<thead>
<tr>
<th>Age</th>
<th>Matings</th>
<th>Age</th>
<th>Matings</th>
<th>Age</th>
<th>Matings</th>
<th>Age</th>
<th>Matings</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>0</td>
<td>30</td>
<td>1</td>
<td>36</td>
<td>5</td>
<td>43</td>
<td>3</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>32</td>
<td>2</td>
<td>36</td>
<td>6</td>
<td>43</td>
<td>4</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>33</td>
<td>4</td>
<td>37</td>
<td>1</td>
<td>43</td>
<td>9</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>33</td>
<td>3</td>
<td>37</td>
<td>1</td>
<td>44</td>
<td>3</td>
</tr>
<tr>
<td>28</td>
<td>3</td>
<td>33</td>
<td>3</td>
<td>37</td>
<td>6</td>
<td>45</td>
<td>5</td>
</tr>
<tr>
<td>29</td>
<td>0</td>
<td>33</td>
<td>3</td>
<td>38</td>
<td>2</td>
<td>47</td>
<td>7</td>
</tr>
<tr>
<td>29</td>
<td>0</td>
<td>33</td>
<td>2</td>
<td>39</td>
<td>1</td>
<td>48</td>
<td>2</td>
</tr>
<tr>
<td>29</td>
<td>0</td>
<td>34</td>
<td>1</td>
<td>41</td>
<td>3</td>
<td>52</td>
<td>9</td>
</tr>
<tr>
<td>29</td>
<td>2</td>
<td>34</td>
<td>1</td>
<td>42</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>2</td>
<td>34</td>
<td>2</td>
<td>43</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>2</td>
<td>34</td>
<td>3</td>
<td>43</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Example (Elephant)

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

Model 1: $\mu = \alpha + \beta$Age (identity link)

On the plot, ‘3’ means there are 3 points at the same location.
Example (Elephant)

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

```r
> eleph.log = glm(Matings ~ Age, family=poisson(link="log"))
> summary(eleph.log)
```

Coefficients:

```
              Estimate Std. Error z value Pr(>|z|)  
(Intercept)  -1.58201   0.54462  -2.905  0.00368 **  
Age           0.06869   0.01375   4.997  5.81e-07 ***
```

Null deviance: 75.372 on 40 degrees of freedom  
Residual deviance: 51.012 on 39 degrees of freedom  
AIC: 156.46

Fitted model 2: $\log(\hat{\mu}) = -1.582 + 0.0687 \text{Age}$  
\[ \hat{\mu} = \exp(-1.582 + 0.0687 \text{Age}) = 0.205(1.071)^{\text{Age}} \]

- expected number of matings increase by 7.1% for every extra year of age  
- for a 40 year-old male, the expected number of matings is \[ \hat{\mu} = \exp(-1.582 + 0.0687(40)) \approx 3.2. \]

Residuals

- Deviance residual:  
  \[ d_i = \text{sign}(y_i - \hat{\mu}_i) \sqrt{2 \left[ y_i \log(y_i/\hat{\mu}_i) - y_i + \hat{\mu}_i \right]} \]
- Pearson’s residual:  
  \[ e_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}} \]
- Standardized Pearson’s residual = \[ \frac{e_i}{\sqrt{1 - h_i}} \]
- Standardized Deviance residual = \[ \frac{d_i}{\sqrt{1 - h_i}} \]

where \( h_i \) = leverage of \( i \)th observation  
- potential outlier if \(|\text{standardized residual}| > 2\) 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"`.  

Which Model Better Fits the Data?

<table>
<thead>
<tr>
<th>Model 1 (identity link)</th>
<th>AIC</th>
<th>Deviance</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>155.50</td>
<td>50.058</td>
<td>39</td>
</tr>
</tbody>
</table>

- Based on AIC, Model 1 fits better  
- Goodness of fit tests are not appropriate because...  
- Based on scatter plot...
Residual Plots

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)

Models for Rates

Sometimes \( y_i \) have different bases (e.g., number murders for cities with different pop. sizes)

Let \( y \) = count with base \( t \). Assume \( y \sim \text{Poisson}(\mu) \), where

\[
\mu = \lambda t
\]

more relevant to model rate \( \lambda \) at which events occur.

Loglinear model:

\[
\log \lambda = \log(\mu/t) = \alpha + \beta x
\]

i.e.,

\[
\log(\mu) - \log(t) = \alpha + \beta x
\]

\( \log(t) \) is an offset.

See pp. 82-84 of text for discussion.

Example (British Train Accidents over Time)

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 base \( t = \) annual number of train-km, and \( x = \) num. of years since 1975

\[
\begin{array}{cccc}
\text{Year} & \text{KM} & \text{Train} & \text{TrRd} \\
1 & 2003 & 518 & 0 & 3 \\
2 & 2002 & 516 & 1 & 3 \\
3 & 2001 & 508 & 0 & 4 \\
4 & 2000 & 503 & 1 & 3 \\
5 & 1999 & 505 & 1 & 2 \\
\ldots & & & \\
27 & 1977 & 425 & 1 & 8 \\
28 & 1976 & 426 & 2 & 12 \\
29 & 1975 & 436 & 5 & 2 \\
\end{array}
\]

Fitted Model: \( \log(\hat{\lambda}) = \log(\hat{\mu}/t) = -4.21 - 0.0329x \)

\[
\hat{\lambda} = \frac{\hat{\mu}}{t} = e^{-4.21-0.0329} = e^{-4.21}(e^{-0.0329})^x = (0.0148)(0.968)^x
\]

Rate estimated to decrease by 3.2% per yr from 1975 to 2003.

- Est. rate for 1975 (\( x = 0 \)) is 0.0148 per million km (15 per billion).
- Est. rate for 2003 (\( x = 28 \)) is 0.0059 per million km (6 per billion).
There were 13 train-road collisions in 1986, a lot higher than the fitted mean 4.3 for that year.

Models for Rate Data With Identity Link

For \( y \sim \text{Poisson}(\mu) \) with base \( t \), where

\[
\mu = \lambda t
\]

the loglinear model

\[
\log \lambda = \log(\mu/t) = \alpha + \beta x
\]

assumes the effect of the explanatory variable on the response to be multiplicative. Alternatively, if we want the effect to be additive,

\[
\lambda = \mu/t = \alpha + \beta x
\]

\[ \Leftrightarrow \mu = \alpha t + \beta tx \]

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

base $t = \text{annual num. of train-km}, \ x = \text{num. of years since 1975}$

```r
> trains2 = glm(TrRd ~ -1 + KM + I(KM*(Year-1975)), 
  family=poisson(link="identity"), data=trains)
> summary(trains2)
```

```
Estimate Std. Error z value Pr(>|z|)
KM  1.426e-02  1.888e-03  7.557   4.14e-14 ***
I(KM * (Year - 1975)) -3.239e-04  9.924e-05  3.264   0.0011 **
---
Null deviance: Inf on 29 degrees of freedom
Residual deviance: 37.287 on 27 degrees of freedom
AIC: 132.95
```

Fitted Model: \( \hat{\lambda} = \hat{\mu} / t = 0.0143 - 0.000324x \)

- Estimated rate decreases by 0.00032 per million km (0.32 per billion km) per yr from 1975 to 2003.
- Est. rate for 1975 ($x = 0$) is 0.0143 per million km (14.3 per billion km).
- Est. rate for 2003 ($x = 28$) is 0.0052 per million km (5.2 per billion km).

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

3.3.4 Overdispersion and Negative Binomial Regression

Overdispersion: Greater Variability than Expected

- One of the defining characteristics of Poisson regression is its lack of a parameter for variability:
  \[ \mathbb{E}(Y) = \text{Var}(Y), \]

  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 $\hat{\lambda}_i$ is larger than what $\hat{\lambda}_i$ 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.
Common Causes of Overdispersion

- Subject heterogeneity
  - subjects have different \( \mu \)
    - e.g., rates of infestation may differ from location to location on the same tree and may differ from tree to tree
  - there are important predictors not included in the model

- Observations are not independent – clustering

---

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?”

<table>
<thead>
<tr>
<th>Number of Victims</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black Subjects</td>
<td>119</td>
<td>16</td>
<td>12</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>159</td>
</tr>
<tr>
<td>White Subjects</td>
<td>1070</td>
<td>60</td>
<td>14</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1149</td>
</tr>
</tbody>
</table>

If fit a Poisson distribution to the data from blacks, MLE for \( \lambda \) is the sample mean

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

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

> `round(dpois(0:6, lambda = 83/159), 3)`

Likewise, if we fit a Poisson distribution to the data from whites, MLE for \( \lambda \) is

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

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

> `round(dpois(0:6, lambda = 106/1149), 3)`

---

Example (Known Victims of Homicide)

<table>
<thead>
<tr>
<th>Num. of Victims</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Total</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>119</td>
<td>16</td>
<td>12</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>159</td>
<td>0.522</td>
<td>1.150</td>
</tr>
<tr>
<td>White</td>
<td>1070</td>
<td>60</td>
<td>14</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1149</td>
<td>0.092</td>
<td>0.155</td>
</tr>
</tbody>
</table>

---

Negative Binomial Distribution

If \( Y \) has a negative binomial distribution, with mean \( \mu \) 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

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

- 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 \( \mu \) is the sample mean. MLE for \( \theta \) has no close form formula.

---

Example (Known Victims of Homicide)

<table>
<thead>
<tr>
<th>Num. of Victims</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Total</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>119</td>
<td>16</td>
<td>12</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>159</td>
<td>0.522</td>
<td>1.150</td>
</tr>
<tr>
<td>White</td>
<td>1070</td>
<td>60</td>
<td>14</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1149</td>
<td>0.092</td>
<td>0.155</td>
</tr>
</tbody>
</table>

Likewise, if we fit a Poisson distribution to the data from whites, MLE for \( \lambda \) is

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

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

> `round(dpois(0:6, lambda = 106/1149), 3)`

Too many 0’s and too many large counts for both races than expected if the samples were drawn from Poisson distributions.

It is not surprising that Poisson distributions do not fit the data because of the large discrepancies between sample mean and sample variance.
Example (Known Victims of Homicide)

Data:
\[ Y_{b,1}, Y_{b,2}, \ldots, Y_{b,159} \] answers from black subjects
\[ Y_{w,1}, Y_{w,2}, \ldots, Y_{w,1149} \] 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 \( \mu_b = e^{\alpha+\beta}, \quad \mu_w = e^\alpha \). So \( e^\beta = \mu_b/\mu_w \).

Poisson - 33

Example (Known Victims of Homicide) — Poisson Fits

```r
> summary(hom.poi)
Call:
  glm(formula = nvics ~ race, family = poisson, data = homicide, 
weights = freq)
Deviance Residuals:
  Min       1Q   Median       3Q      Max
-14.051    0.000    5.257    6.216   13.306

Coefficients:
          Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.38321   0.09713  -24.54   <2e-16 ***
raceBlack   1.73314   0.14657   11.82   <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
```

Poisson - 35

Example (Known Victims of Homicide) — Neg. Binomial

```r
> summary(hom.nb)
Call:
  glm.nb(formula = nvics ~ race, weights = freq, init.theta = 0.2023119205, 
link = log)
Deviance Residuals:
  Min       1Q   Median       3Q      Max
-12.754    0.000    2.086    3.283    9.114

Coefficients:
          Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.6501   0.2077   -3.130  0.00175 **
raceWhite  -1.7331   0.2385  -7.268  3.66e-13 *** 
---
(Dispersion parameter for Negative Binomial(0.2023) family taken to be 1)

Null deviance: 471.57 on 10 degrees of freedom
Residual deviance: 412.60 on 9 degrees of freedom
AIC: 1001.8

Theta: 0.2023
2 x log-likelihood: -995.7980
```

Poisson - 36
Example (Known Victims of Homicide)

Black Subjects

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

White Subjects:

<table>
<thead>
<tr>
<th>Num. of victims</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>observed freq.</td>
<td>1070</td>
<td>60</td>
<td>14</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1149</td>
</tr>
<tr>
<td>relative freq.</td>
<td>0.931</td>
<td>0.052</td>
<td>0.012</td>
<td>0.003</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.999</td>
</tr>
<tr>
<td>poisson fit</td>
<td>0.912</td>
<td>0.084</td>
<td>0.004</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1</td>
</tr>
<tr>
<td>neg. bin. fit</td>
<td>0.927</td>
<td>0.059</td>
<td>0.011</td>
<td>0.003</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>1.001</td>
</tr>
</tbody>
</table>

Example (Known Victims of Homicide)

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

<table>
<thead>
<tr>
<th>Model</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\beta}$</th>
<th>SE($\hat{\beta}$)</th>
<th>Wald 95% CI for $e^\beta = \mu_B/\mu_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson</td>
<td>-2.38</td>
<td>1.73</td>
<td>0.147</td>
<td>$\exp(1.73 \pm 1.96 \cdot 0.147) = (4.24, 7.54)$</td>
</tr>
<tr>
<td>Neg. Binom.</td>
<td>-2.38</td>
<td>1.73</td>
<td>0.238</td>
<td>$\exp(1.73 \pm 1.96 \cdot 0.238) = (3.54, 9.03)$</td>
</tr>
</tbody>
</table>

Poission and negative binomial models give

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

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

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

---

Wald CIs

<table>
<thead>
<tr>
<th>Model</th>
<th>2.5 %</th>
<th>97.5 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-2.573577</td>
<td>-2.192840</td>
</tr>
<tr>
<td>raceBlack</td>
<td>1.445877</td>
<td>2.020412</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>2.5 %</th>
<th>97.5 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.0762623</td>
<td>0.1115994</td>
</tr>
<tr>
<td>raceBlack</td>
<td>4.2455738</td>
<td>7.5414329</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>2.5 %</th>
<th>97.5 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-2.612916</td>
<td>-2.153500</td>
</tr>
<tr>
<td>raceBlack</td>
<td>1.265738</td>
<td>2.00551</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>2.5 %</th>
<th>97.5 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.07332043</td>
<td>0.1160771</td>
</tr>
<tr>
<td>raceBlack</td>
<td>3.54871025</td>
<td>9.0299848</td>
</tr>
</tbody>
</table>
### Likelihood Ratio CIs

```r
> confint(hom.poi)
Waiting for profiling to be done...
2.5 %   97.5 %
(Intercept) -2.579819 -2.198699
raceBlack  1.443698  2.019231

> exp(confint(hom.poi))
Waiting for profiling to be done...
2.5 %   97.5 %
(Intercept) 0.0757877 0.1109474
raceBlack  4.2363330  7.5325339
```

```r
> confint(hom.nb)
Waiting for profiling to be done...
2.5 %   97.5 %
(Intercept) -2.616478 -2.156532
raceBlack  1.274761  2.211746

> exp(confint(hom.nb))
Waiting for profiling to be done...
2.5 %   97.5 %
(Intercept) 0.07305976 0.1157258
raceBlack  3.57784560  9.1316443
```

#### 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 are 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 deviance
- Examine residuals to see if a large deviance statistic may be due to outliers
- Large numbers of outliers are usually signs of overdispersion
- Check standardized residuals and plot them against them fitted values $\hat{\mu}_i$.

### Train Data Revisit

Recall Pearson’s residual:

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

If no overdispersion, then

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

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

With overdispersion,

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

then the size of Pearson’s residuals may increase with $\mu$.

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

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

Poisson - 45

Train Data — Neg. Bin. Model

```r
> trains.nb = glm.nb(TrRd ~ I(Year-1975)+offset(log(KM)), data=trains)
> summary(trains.nb)

Coefficients:

Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.19999  0.19584  -21.446   < 2e-16 ***
I(Year - 1975) -0.03367  0.01288   -2.615  0.00893 **
---
(Dispersion parameter for Negative Binomial(10.1183) family taken to be 1)

Null deviance: 32.045 on 28 degrees of freedom
Residual deviance: 25.264 on 27 degrees of freedom
AIC: 132.69

Theta: 10.12
Std. Err.: 8.00
2 x log-likelihood:  -126.69

For year effect, the estimated coefficients are similar (0.0337 for neg. bin. model compared to 0.032 for Poisson model), but less significant (P-value = 0.009 in neg. bin. model compared to 0.002 in Poisson model)