Reference: Agresti, Section 7.3

Negative binomial models

- The negative binomial model is useful when overdispersion occurs with Poisson GLMs, i.e., when the variance of the data exceeds the mean.

Slide 1

- A common reason for overdispersion is heterogeneity that may arise from omitted covariates.

- A mixture model is a flexible way to account for overdispersion

$$
\begin{aligned}
Y|\lambda &\sim \text{Poisson}(\lambda) \\
\lambda &\sim \text{Gamma}(\alpha, \beta),
\end{aligned}
$$

where $\text{Gamma}(\alpha, \beta)$ is the gamma distribution with mean $\alpha\beta$

---

and variance $\alpha\beta^2$, with density

$$
f(\lambda) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\lambda/\beta), \quad \text{for } \lambda > 0.
$$

- The marginal distribution of Y is then negative binomial

$$
f(y) = \frac{\Gamma(\alpha + y)}{\Gamma(\alpha) y!} \left(\frac{\beta}{1+\beta}\right)^y \left(\frac{1}{1+\beta}\right)^\alpha, \quad y = 0, 1, 2, \ldots
$$

Slide 2

- This distribution has mean

$$
E[Y] = E[E[Y|\lambda]] = E[\lambda] = \alpha\beta
$$

and variance

$$
\begin{aligned}
Var(Y) &= E[Var(Y|\lambda)] + Var(E[Y|\lambda]) \\
&= E[\lambda] + Var(\lambda) = \alpha\beta + \alpha\beta^2.
\end{aligned}
$$

Slide 3

- We set $\mu = \alpha\beta$ and $\gamma = 1/\alpha$, so that

$$E[Y] = \mu \quad \text{and} \quad Var(Y) = \mu + \gamma\mu^2,$$

  γ is called the **dispersion parameter**.

- The distribution of Y is then

$$f(y) = \frac{\Gamma(\gamma^{-1} + y)}{\Gamma(\gamma^{-1})\Gamma(y+1)} \left(\frac{\mu}{\gamma^{-1} + \mu}\right)^y \left(\frac{\gamma^{-1}}{\gamma^{-1} + \mu}\right)^{1/\gamma}.$$

- As $\gamma \to 0$, $Var(Y) \to \mu$ and the negative binomial distribution converges to the Poisson distribution with mean μ.

- The greater γ, the greater the overdispersion compared to the ordinary Poisson GLM.

Slide 4

- Usually, the dispersion parameter γ is unknown and needs to be estimated.

- For independent observations, the ML estimate of μ is the sample mean, but ML estimation for γ requires iterative methods.

- The negative binomial can accommodate overdispersion but not underdispersion with respect to the Poisson model.

Slide 5

- For the negative binomial GLM, we assume

$$Y_i \sim \ \mathrm{NegBin}(\mu_i, \gamma)$$

and apply a log link, so that

$$\log \mu_i = \eta_i = x_i'\beta.$$

Or we can use

$$\log \mu_i = \eta_i = o_i + x_i'\beta,$$

if an offset is needed.

Slide 6

Maximum Likelihood estimation

The log-likelihood function is given by

$$\begin{aligned}
\log \mathcal{L}(\boldsymbol{\beta}, \gamma) \ &= \ \sum_{i=1}^{N} \left[\log \Gamma \left(y_i + \frac{1}{\gamma} \right) - \log \Gamma \left(\frac{1}{\gamma} \right) - \log \Gamma(y_i + 1) \right] \\
&+ \ \sum_{i=1}^{N} \left[y_i \log \left(\frac{\gamma \mu_i}{1 + \gamma \mu_i} \right) - \frac{1}{\gamma} \log \left(1 + \gamma \mu_i \right) \right]
\end{aligned}$$

- ML fitting can use a Newton-Raphson algorithm on all the parameters simultaneously;

- Another approach iterates until convergence between
  (1) using IRWLS to solve the equations for $\boldsymbol{\beta}$, for fixed $\hat{\gamma}$ and
  (2) using Newton-Raphson to estimate γ, for fixed $\hat{\boldsymbol{\beta}}$.

Slide 7

- The estimated covariance matrix for $\hat{\boldsymbol{\beta}}$ is

$$\widehat{Var}(\hat{\boldsymbol{\beta}}) = \left(\boldsymbol{X}'\hat{W}\boldsymbol{X} \right)^{-1},$$

where W is the diagonal matrix with

$$w_i = \frac{1}{Var(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 = \frac{\mu_i}{1 + \gamma \mu_i}$$

- The deviance for a negative binomial GLM is

$$G^2 = 2 \sum_i \left[y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) - \left(y_i + \frac{1}{\hat{\gamma}} \right) \log \left(\frac{1 + \hat{\gamma} y_i}{1 + \hat{\gamma} \hat{\mu}_i} \right) \right]$$

Slide 8

Example

To see how these methods behave, let's work with some artificial data. I generated $N = 100$ observations from

$$
\begin{aligned}
Y_i &\sim \text{NegBin}(\mu_i, \gamma) \\
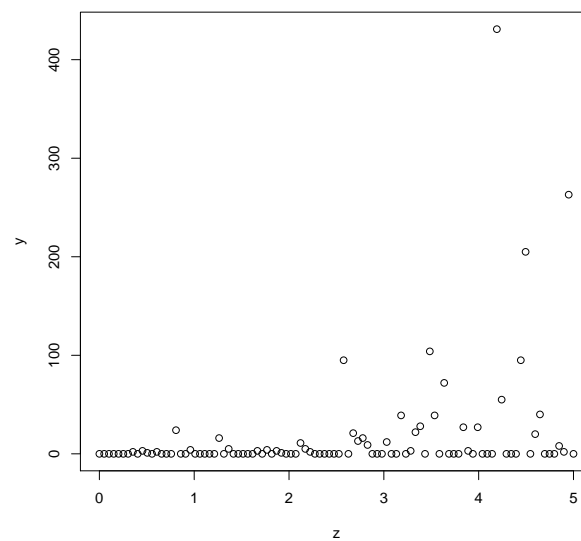\log \mu_i &= \beta_0 + \beta_1 z_i,
\end{aligned}
$$

with $\beta_0 = 0$, $\beta_1 = 1$, $\gamma = 10$. The values of z_i were equally spaced along a grid from 0 to 5.
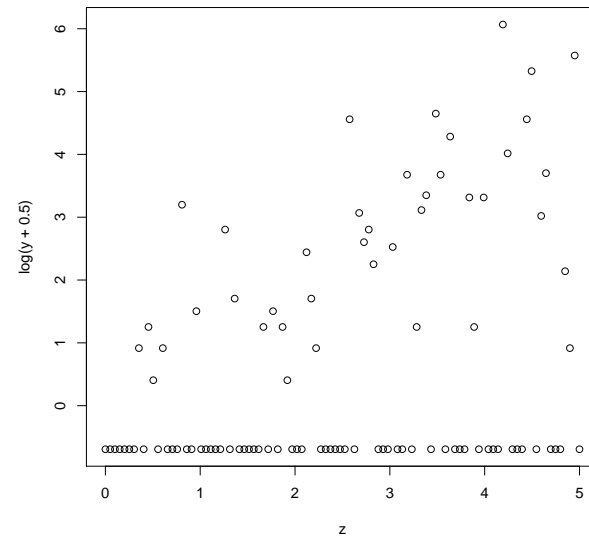
Slide 9

R code for simulating $Y_i \sim$ **NegBin**(μ_i, γ)

```
set.seed(10)
N=100; beta0=0; beta1=1; theta=1/10
z = seq(0, 5, length=N)
mu = exp(beta0+beta1*z)
y = rnegbin(N, mu, theta)
plot(z, y)
plot(z, log(y+0.5))
```

Slide 10

Slide 11

Slide 12

Let's start by fitting a Poisson model

$$Y_i \quad \sim \quad \text{Poisson}(\mu_i)$$
$$\log \mu_i \quad = \quad \beta_0 + \beta_1 z_i.$$

```
fit.poisson = glm(y ~ z, family=poisson)
> summary(fit.poisson)


            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.07171    0.09866  -0.727    0.467
z            0.88168    0.02422  36.396   <2e-16 ***
---
(Dispersion parameter for poisson family taken to be 1)


    Null deviance: 6666.8  on 99  degrees of freedom
Residual deviance: 4731.7  on 98  degrees of freedom
AIC: 4914.2
```
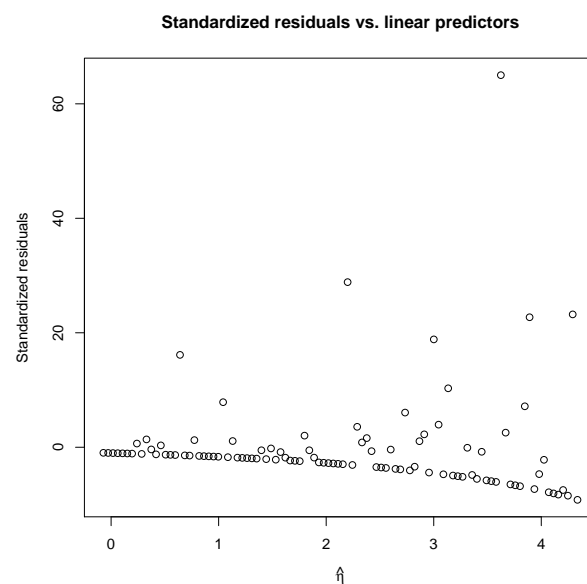
Slide 13

Let's look at some residuals plots

```
stres.pois = rstandard(fit.poisson, type="pearson")
plot(fit.poisson$linear.predictors, stres.pois,
      xlab=expression(hat(eta)), ylab = "Standardized residuals",
      main="Standardized residuals vs. linear predictors")
plot(fit.poisson$linear.predictors, abs(stres.pois),
      xlab=expression(hat(eta)), ylab = "Absolute standardized residuals",
      main="Absolute value of standardized residuals vs. linear predictors")
```

Slide 14

Slide 15

Absolute value of standardized residuals vs. linear predictors
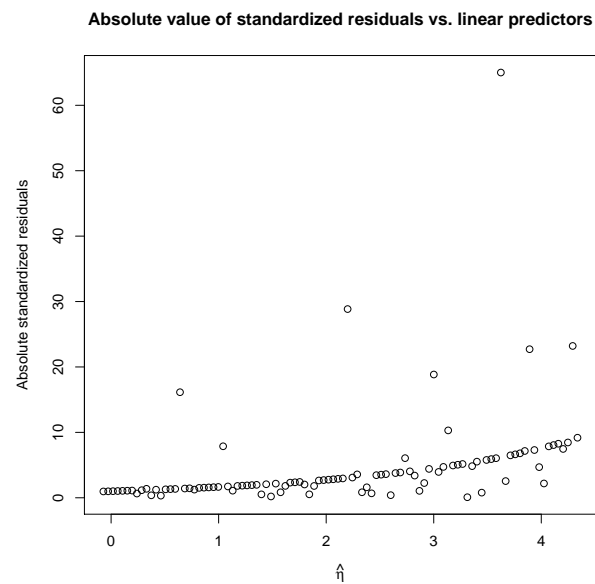
Slide 16

- The model does not fit well.

```
> 1-pchisq(4731.7, 98)
[1] 0
```

- The variance of Y_i appears to be increasing faster than the Poisson model allows.

- The 95% confidence intervals for β_1 does not cover the true value.

Now let's fit the negative binomial model

$$Y_i \quad \sim \quad \text{NegBin}(\mu_i, \kappa)$$
$$\log \mu_i \quad = \quad \beta_0 + \beta_1 z_i.$$

```
library(MASS)
fit.negbin = glm.nb(y ~ z)
```

Slide 17

```
> summary(fit.negbin)

            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.4112     0.5593  -0.735    0.462
z             0.9786     0.1903   5.142 2.72e-07 ***
---
(Dispersion parameter for Negative Binomial(0.138) family taken to be 1)

    Null deviance: 99.437  on 99  degrees of freedom
```

Slide 18

```
Residual deviance: 77.287  on 98  degrees of freedom
AIC: 478.08

            Theta:  0.1380
        Std. Err.:  0.0263
```

Slide 19

- The model fits well (p-value $= 0.14$).

  ```
  > 1-pchisq(77.287, 98)
  [1] 0.9394652
  ```

- The estimated dispersion parameter is $\hat{\gamma} = 1/0.138 = 7.246$, which is close to the true value $\gamma = 10$.

- The estimated coefficient, $\hat{\beta}_1$ is close to its true value.

- The larger standard errors are appropriate, reflecting the fact that there is more uncertainty than the Poisson model allows.

Slide 20

Underdispersion and boundary estimates

- Recall that the negative binomial model allows overdispersion but not underdispersion relative to the Poisson.

- The negative binomial approaches the Poisson as $\gamma \to 0$.

- If we apply the negative binomial model to data that are not overdispersed relative to the Poisson, we will usually get an estimated value of γ equal to zero.

- It seems logical to test H_0 : the data are Poisson, versus H_A : the data are negative binomial, but the situation is non-standard because under H_0, γ is on the boundary of its parameter space.

Slide 21

Example – Agresti Exercise 7.31 (Table 7.5).

In this study, 1308 subjects cross-classified by race were asked how many people they personally knew who have been victims of homicide within the last year.

Response	Black	White
0	119	1070
1	16	60
2	12	14
3	7	4
4	3	0
5	2	0
6	0	1

Slide 22

- The sample mean for the 159 blacks is

$$\frac{\sum_i n_i y_i}{\sum_i n_i} = \frac{83}{159} = 0.522$$

  with a variance of

$$\frac{\sum_i n_i (y_i - \bar{y}_i)^2}{\sum_i n_i - 1} = 1.150$$

- The sample mean for the 1149 whites is 0.092 with a variance of 0.155.

- Thus, in both groups, the sample variances are about twice the sample means.

Slide 23

A natural first choice for modeling count data is a Poisson GLM
with an indicator variable for race

$$\log \mu_i = \alpha + \beta x_i,$$

with $x_i = 1$ for blacks and $x_i = 0$ for whites.

```
homicide = data.frame(response=rep(0:6,each=2), race=rep(c("black", "white"), 7),
                      counts=c(119, 1070, 16, 60, 12, 14, 7, 4, 3, 0, 2, 0, 0, 1))
homicide$race = relevel(homicide$race, ref="white")


hom.pois = glm(response ~ race, family=poisson, weights=counts, data=homicide)


> summary(hom.pois)
             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 ***
---
```

Slide 24

```
(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
```

The model does not give a good fit to the data (p-value ≈ 0).

Heterogeneity among subjects of a given race is plausible due to various demographic factors. Now let's fit the negative binomial model

$$Y_i \sim \text{NegBin}(\mu_i, \gamma)$$
$$\log \mu_i = \beta_0 + \beta_1 x_i$$

```
hom.nb = glm.nb(response ~ race, weights=counts, data=homicide)

> summary(hom.nb)
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.3832     0.1172 -20.335  < 2e-16 ***
raceblack     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
```

Slide 25

```
Residual deviance: 412.60  on  9  degrees of freedom
AIC: 1001.8

            Theta:  0.2023
        Std. Err.:  0.0409

 2 x log-likelihood:  -995.7980
```

Slide 26

Slide 27

- The parameter estimates are the same as before $\hat{\beta}_1 = 1.733$, since both models provide fitted means equal to the sample means
    - $\exp(-2.38) = 0.092$ for whites
    - $\exp(-2.38 + 1.733) = 0.522$ for blacks
- The standard error of $\hat{\beta}$ increases from 0.147 for the Poisson GLM to 0.239 for the negative binomial model.
- The Wald 95% confidence interval for the ratio of means for blacks versus whites is

$$\exp\left(1.733 \pm 1.96 \times 0.239\right) = (3.55, 9.03)$$

- $\widehat{Var}(Y) = \hat{\mu} + \hat{\gamma}\hat{\mu}^2$ is closer to the sample variances
    - $0.092 + 1/0.2023 * 0.092^2 = 0.134$ for whites and
    - $0.522 + 1/0.2023 * 0.522^2 = 1.869$ for blacks.

Slide 28

- There is strong evidence that $\gamma > 0$; $\hat{\gamma} = 1/0.20 = 4.94$ and a 95% CI for γ is

$$1/(0.2023 \pm 1.96 \times 0.0409) = (3.54, 8.19)$$

Slide 29

```
muhat = unique(hom.nb$fitted.values)
n = by(homicide$counts, homicide$race, sum)
nwhite.poi = round(dpois(0:6, lambda=muhat[2])*n[1],3)
nblack.poi = round(dpois(0:6, lambda=muhat[1])*n[2],3)
nwhite.negb = round(dnbinom(0:6, size=fit.negbin$theta, mu=muhat[2])*n[1],3)
nblack.negb = round(dnbinom(0:6, size=fit.negbin$theta, mu=muhat[1])*n[2],3)
```

Slide 30

Observed and expected counts under the Poisson and Negative Binomial models:

	Data		Poisson GLM		Neg. Bin. GLM	
Response	Black	White	Black	White	Black	White
0	119	1070	94.339	1047.743	122.838	1064.904
1	16	60	49.246	96.659	17.910	67.474
2	12	14	12.854	4.459	7.760	12.704
3	7	4	2.237	0.137	4.105	2.921
4	3	0	0.292	0.003	2.369	0.732
5	2	0	0.030	0.000	1.435	0.193
6	0	1	0.003	0.000	0.897	0.1052