

Models for Count Data

FW8051 Statistics for Ecologists

Department of Fisheries, Wildlife and Conservation Biology



Learning Objectives

- Be able to fit regression models appropriate for count data in R and JAGS
 - Poisson regression models
 - Quasi-Poisson (R only)
 - Negative Binomial regression
- Be able to evaluate model fit
 - Residual plots
 - Goodness-of-fit tests
- Use deviances and AIC to compare models.
- Use an offset to model rates and densities, accounting for variable survey effort
- Be able to describe statistical models and their assumptions using equations and text and match parameters in these equations to estimates in computer output.

Deviance

```
fit.pois<-glm(slugs~field, data=slugs, family=poisson())  
summary(fit.pois)
```

Call:

```
glm(formula = slugs ~ field, family = poisson(), data = slugs)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.1331	-1.5969	-0.9519	0.4580	4.8727

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.2429	0.1400	1.735	0.082744 .
fieldRookery	0.5790	0.1749	3.310	0.000932 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance:	224.86	on 79	degrees of freedom
Residual deviance:	213.44	on 78	degrees of freedom
AIC:	346.26		

Number of Fisher Scoring iterations: 6

Deviance

The **deviance** is a measure of model fit when using maximum likelihood.

Deviance

The **deviance** is a measure of model fit when using maximum likelihood.

$$\text{Residual deviance} = 2(\log L(y|\hat{\theta}_s) - \log L(y|\hat{\theta}_0))$$

- $\log L(\hat{\theta}_s|y)$ = log-likelihood for a “saturated” model (one with a parameter for each observation).
- $\log L(\hat{\theta}_0|y)$ = log-likelihood for our model of interest evaluated at the MLEs ($\hat{\theta}_0$)

Deviance

The **deviance** is a measure of model fit when using maximum likelihood.

$$\text{Residual deviance} = 2(\log L(y|\hat{\theta}_s) - \log L(y|\hat{\theta}_0))$$

- $\log L(\hat{\theta}_s|y)$ = log-likelihood for a “saturated” model (one with a parameter for each observation).
- $\log L(\hat{\theta}_0|y)$ = log-likelihood for our model of interest evaluated at the MLEs ($\hat{\theta}_0$)

Null deviance = residual deviance for a model that only contains an intercept

It may be helpful to think of the **null deviance** and **residual deviance** as maximum likelihood equivalents to total and residual sums of squares, respectively.

Deviance

There is no R^2 for generalized linear models, but sometimes see % deviance explained (**pseudo R^2**):

$$100 \times \frac{\text{null deviance} - \text{residual deviance}}{\text{null deviance}}$$

Deviance

There is no R^2 for generalized linear models, but sometimes see % deviance explained (**pseudo R^2**):

$$100 \times \frac{\text{null deviance} - \text{residual deviance}}{\text{null deviance}}$$

- Null deviance \approx maximum likelihood equivalent of total sum of squares.
- Residual deviance \approx maximum likelihood equivalent of residual sum of squares

For a discussion of different pseudo- R^2 measures, see:

https://web.archive.org/web/20130701052120/http://www.ats.ucla.edu/80/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm

Model comparisons

For large samples, the difference in deviances for nested models should be $\sim \chi^2$ with df = difference in number of parameters between the two models.

$$D_2 - D_1 \sim \chi_{df}^2$$

Model comparisons

For large samples, the difference in deviances for nested models should be $\sim \chi^2$ with df = difference in number of parameters between the two models.

$$D_2 - D_1 \sim \chi^2_{df}$$

- Can use `drop1(model, test="Chi")` (equivalent to a likelihood ratio test) or `Anova` in `car` package

Model comparisons

For large samples, the difference in deviances for nested models should be $\sim \chi^2$ with df = difference in number of parameters between the two models.

$$D_2 - D_1 \sim \chi^2_{df}$$

- Can use `drop1(model, test="Chi")` (equivalent to a likelihood ratio test) or `Anova` in `car` package
- Can use forward, backwards, stepwise selection (with the same dangers/caveats related to overfitting)

Model comparisons

For large samples, the difference in deviances for nested models should be $\sim \chi^2$ with df = difference in number of parameters between the two models.

$$D_2 - D_1 \sim \chi^2_{df}$$

- Can use `drop1(model, test="Chi")` (equivalent to a likelihood ratio test) or `Anova` in `car` package
- Can use forward, backwards, stepwise selection (with the same dangers/caveats related to overfitting)

Alternatively, can use $AIC = -2(\log L(y|\theta)) + 2p$ for model selection

Model comparisons

For large samples, the difference in deviances for nested models should be $\sim \chi^2$ with df = difference in number of parameters between the two models.

$$D_2 - D_1 \sim \chi^2_{df}$$

- Can use `drop1(model, test="Chi")` (equivalent to a likelihood ratio test) or `Anova` in `car` package
- Can use forward, backwards, stepwise selection (with the same dangers/caveats related to overfitting)

Alternatively, can use $AIC = -2(\log L(y|\theta)) + 2p$ for model selection

- Allows comparisons among non-nested models
- `stepAIC` in `MASS` library (for backwards selection)

Residuals

Deviance residuals = $\text{sign}(y_i - \mu_i)\sqrt{d_i}$, where:

- d_i is the contribution of the i^{th} observation to the residual deviance (may be useful for spotting outliers/influential points)
- $\text{sign} = 1$ if $y_i > \mu_i$ and -1 if $y_i < \mu_i$.

Residuals

Deviance residuals = $\text{sign}(y_i - \mu_i) \sqrt{d_i}$, where:

- d_i is the contribution of the i^{th} observation to the residual deviance (may be useful for spotting outliers/influential points)
- $\text{sign} = 1$ if $y_i > \mu_i$ and -1 if $y_i < \mu_i$.

Pearson residuals = $\frac{Y_i - E[Y_i|X_i]}{\sqrt{\text{Var}[Y_i|X_i]}} = \frac{Y_i - \lambda_i}{\sqrt{\lambda_i}}$ for Poisson

Residuals

Deviance residuals = $\text{sign}(y_i - \mu_i) \sqrt{d_i}$, where:

- d_i is the contribution of the i^{th} observation to the residual deviance (may be useful for spotting outliers/influential points)
- $\text{sign} = 1$ if $y_i > \mu_i$ and -1 if $y_i < \mu_i$.

Pearson residuals = $\frac{Y_i - E[Y_i|X_i]}{\sqrt{\text{Var}[Y_i|X_i]}} = \frac{Y_i - \lambda_i}{\sqrt{\lambda_i}}$ for Poisson

- If the Poisson model is appropriate, these residuals should have constant variance

Residuals

Deviance residuals = $\text{sign}(y_i - \mu_i)\sqrt{d_i}$, where:

- d_i is the contribution of the i^{th} observation to the residual deviance (may be useful for spotting outliers/influential points)
- $\text{sign} = 1$ if $y_i > \mu_i$ and -1 if $y_i < \mu_i$.

Pearson residuals = $\frac{Y_i - E[Y_i|X_i]}{\sqrt{\text{Var}[Y_i|X_i]}} = \frac{Y_i - \lambda_i}{\sqrt{\lambda_i}}$ for Poisson

- If the Poisson model is appropriate, these residuals should have constant variance

Often Deviance and Pearson residuals are similar.

Residual Plots

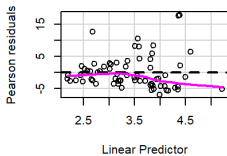
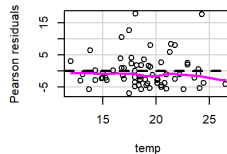
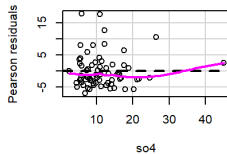
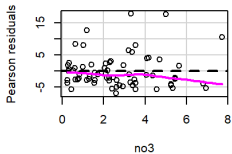
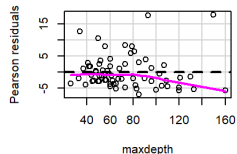
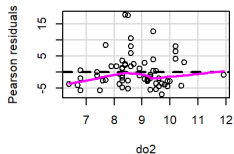
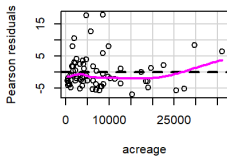
- Residuals versus fitted values
- Residuals versus predictors
- Residuals over time or space (to diagnose possible spatial/temporal correlation)

Residual Plots

- Residuals versus fitted values
- Residuals versus predictors
- Residuals over time or space (to diagnose possible spatial/temporal correlation)

See `residualPlots` in `car` library for useful plots.

Residual plots



Goodness-of-fit: Is the Poisson distribution appropriate?

How can we assess overall model fit?

Goodness-of-fit: Is the Poisson distribution appropriate?

How can we assess overall model fit?

Think “Bayesian p-value”...

Goodness-of-fit: Is the Poisson distribution appropriate?

How can we assess overall model fit?

Think “Bayesian p-value”...

- Account for uncertainty in $\hat{\beta}$ by drawing values of β from $N(\hat{\beta}, \widehat{Cov}(\hat{\beta})^2)$

Goodness-of-fit: Is the Poisson distribution appropriate?

How can we assess overall model fit?

Think “Bayesian p-value”...

- Account for uncertainty in $\hat{\beta}$ by drawing values of β from $N(\hat{\beta}, \widehat{Cov}(\hat{\beta})^2)$
- Use the values of β above to estimate λ (`lambda.hat`)
- Simulate new data (e.g., using `rpois(n, lambda.hat)`)

Goodness-of-fit: Is the Poisson distribution appropriate?

How can we assess overall model fit?

Think “Bayesian p-value”...

- Account for uncertainty in $\hat{\beta}$ by drawing values of β from $N(\hat{\beta}, \widehat{Cov}(\hat{\beta})^2)$
- Use the values of β above to estimate λ (`lambda.hat`)
- Simulate new data (e.g., using `rpois(n, lambda.hat)`)
- Calculate a measure of fit for both real and simulated data

Goodness-of-fit: Is the Poisson distribution appropriate?

How can we assess overall model fit?

Think “Bayesian p-value”...

- Account for uncertainty in $\hat{\beta}$ by drawing values of β from $N(\hat{\beta}, \widehat{Cov}(\hat{\beta})^2)$
- Use the values of β above to estimate λ (`lambda.hat`)
- Simulate new data (e.g., using `rpois(n, lambda.hat)`)
- Calculate a measure of fit for both real and simulated data
- Rinse, repeat.

Goodness-of-fit: Is the Poisson distribution appropriate?

How can we assess overall model fit?

Think “Bayesian p-value”’...

- Account for uncertainty in $\hat{\beta}$ by drawing values of β from $N(\hat{\beta}, \widehat{Cov}(\hat{\beta})^2)$
- Use the values of β above to estimate λ (`lambda.hat`)
- Simulate new data (e.g., using `rpois(n, lambda.hat)`)
- Calculate a measure of fit for both real and simulated data
- Rinse, repeat.

If the model is appropriate: “goodness-of-fit” statistics for real and simulated data should be similar

Goodness of fit test

H_0 : the data were generated by the assumed model

H_A : the data were not generated by the assumed model

Goodness of fit test

H_0 : the data were generated by the assumed model

H_A : the data were not generated by the assumed model

- The GOF statistics for the simulated data tell us what we might expect to see if the Null hypothesis is true.

Goodness of fit test

H_0 : the data were generated by the assumed model

H_A : the data were not generated by the assumed model

- The GOF statistics for the simulated data tell us what we might expect to see if the Null hypothesis is true.
- P-value = proportion of time the GOF statistic for **our observed data** is as or more extreme than GOF statistic for the simulated data

Goodness of fit test

H_0 : the data were generated by the assumed model

H_A : the data were not generated by the assumed model

- The GOF statistics for the simulated data tell us what we might expect to see if the Null hypothesis is true.
- P-value = proportion of time the GOF statistic for **our observed data** is as or more extreme than GOF statistic for the simulated data
- If p-value is small, then we have evidence that the model is not appropriate.

Goodness of fit test

H_0 : the data were generated by the assumed model

H_A : the data were not generated by the assumed model

- The GOF statistics for the simulated data tell us what we might expect to see if the Null hypothesis is true.
- P-value = proportion of time the GOF statistic for **our observed data** is as or more extreme than GOF statistic for the simulated data
- If p-value is small, then we have evidence that the model is not appropriate.
- If the p-value is not small, then. . .

Goodness of fit test

H_0 : the data were generated by the assumed model

H_A : the data were not generated by the assumed model

- The GOF statistics for the simulated data tell us what we might expect to see if the Null hypothesis is true.
- P-value = proportion of time the GOF statistic for **our observed data** is as or more extreme than GOF statistic for the simulated data
- If p-value is small, then we have evidence that the model is not appropriate.
- If the p-value is not small, then. . . we do not have enough evidence to suggest the model is not appropriate

What should we use as a measure of fit?

What should we use as a measure of fit?

One option (see Kery Ch 13), Pearson χ^2 statistic:

$$\chi^2_{n-p} = \sum_{i=1}^n \frac{(Y_i - \hat{E}[Y_i|X_i])^2}{\widehat{Var}[Y_i|X_i]}$$

What should we use as a measure of fit?

One option (see Kery Ch 13), Pearson χ^2 statistic:

$$\chi^2_{n-p} = \sum_{i=1}^n \frac{(Y_i - \hat{E}[Y_i|X_i])^2}{\widehat{Var}[Y_i|X_i]}$$

Loop:

- Simulate a random vector of β 's using $MVN(\hat{\beta}, \widehat{Cov}(\hat{\beta})^2)$
- Use these β 's to form λ_i 's = $\exp(\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i})$
- Simulate new data using these λ_i 's and `rpois()`.
- Calculate χ^2_{n-p} for real and simulated data (plugging in $\hat{E}[Y_i|X_i] = \widehat{Var}[Y_i|X_i] = \hat{\lambda}_i$).

What should we use as a measure of fit?

One option (see Kery Ch 13), Pearson χ^2 statistic:

$$\chi_{n-p}^2 = \sum_{i=1}^n \frac{(Y_i - \hat{E}[Y_i|X_i])^2}{\widehat{Var}[Y_i|X_i]}$$

Loop:

- Simulate a random vector of β 's using $MVN(\hat{\beta}, \widehat{Cov}(\hat{\beta})^2)$
- Use these β 's to form λ_i 's = $\exp(\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i})$
- Simulate new data using these λ_i 's and `rpois()`.
- Calculate χ_{n-p}^2 for real and simulated data (plugging in $\hat{E}[Y_i|X_i] = \widehat{Var}[Y_i|X_i] = \hat{\lambda}_i$).

p-value = proportion of times $\chi_{sim}^2 \geq \chi_{real}^2$ (want large values)

What should we use as a measure of fit?

One option (see Kery Ch 13), Pearson χ^2 statistic:

$$\chi_{n-p}^2 = \sum_{i=1}^n \frac{(Y_i - \hat{E}[Y_i|X_i])^2}{\widehat{Var}[Y_i|X_i]}$$

Loop:

- Simulate a random vector of β 's using $MVN(\hat{\beta}, \widehat{Cov}(\hat{\beta})^2)$
- Use these β 's to form λ_i 's = $\exp(\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i})$
- Simulate new data using these λ_i 's and `rpois()`.
- Calculate χ_{n-p}^2 for real and simulated data (plugging in $\hat{E}[Y_i|X_i] = \widehat{Var}[Y_i|X_i] = \hat{\lambda}_i$).

p-value = proportion of times $\chi_{sim}^2 \geq \chi_{real}^2$ (want large values)

Lets do this!

Overdispersion, $Var(Y|X) > E(Y|X)$

Reasons data may be overdispersed:

Overdispersion, $Var(Y|X) > E(Y|X)$

Reasons data may be overdispersed:

- Omitted variables

Overdispersion, $Var(Y|X) > E(Y|X)$

Reasons data may be overdispersed:

- Omitted variables
- Explanatory and response variables may be measured with error

Overdispersion, $Var(Y|X) > E(Y|X)$

Reasons data may be overdispersed:

- Omitted variables
- Explanatory and response variables may be measured with error
- Model may be mis-specified (relationship between $\log(\mu)$ and x may be non-linear)

Overdispersion, $Var(Y|X) > E(Y|X)$

Reasons data may be overdispersed:

- Omitted variables
- Explanatory and response variables may be measured with error
- Model may be mis-specified (relationship between $\log(\mu)$ and x may be non-linear)
- Outliers

Overdispersion, $Var(Y|X) > E(Y|X)$

Reasons data may be overdispersed:

- Omitted variables
- Explanatory and response variables may be measured with error
- Model may be mis-specified (relationship between $\log(\mu)$ and x may be non-linear)
- Outliers
- Spatial, temporal, within-individual clustering (repeated measures)

Overdispersion, $Var(Y|X) > E(Y|X)$

Reasons data may be overdispersed:

- Omitted variables
- Explanatory and response variables may be measured with error
- Model may be mis-specified (relationship between $\log(\mu)$ and x may be non-linear)
- Outliers
- Spatial, temporal, within-individual clustering (repeated measures)
- Response may be due to a mixture of random processes
 - Presence/absence determined by suitable habitat
 - Counts | suitable habitat may be Poisson
 - Leads to “zero-inflation” models

Testing for Overdispersion: Residual Deviance and Pearson's χ^2

Some compare:

$$\text{Residual deviance} \quad \text{or} \quad \sum_{i=1}^n \frac{(Y_i - E[Y_i|X_i])^2}{\text{Var}[Y_i|X_i]}$$

to a χ^2 distribution with $n - p$ degrees of freedom.

Testing for Overdispersion: Residual Deviance and Pearson's χ^2

Some compare:

$$\text{Residual deviance} \quad \text{or} \quad \sum_{i=1}^n \frac{(Y_i - E[Y_i|X_i])^2}{\text{Var}[Y_i|X_i]}$$

to a χ^2 distribution with $n - p$ degrees of freedom.

Large values can be caused by:

- Mis-specified model (missing predictors)
- Mis-specified distribution
- Outliers
- Small numbers of observations for any set of unique covariate values

Testing for Overdispersion: Residual Deviance and Pearson's χ^2

Some compare:

$$\text{Residual deviance} \quad \text{or} \quad \sum_{i=1}^n \frac{(Y_i - E[Y_i | X_i])^2}{\text{Var}[Y_i | X_i]}$$

to a χ^2 distribution with $n - p$ degrees of freedom.

Large values can be caused by:

- Mis-specified model (missing predictors)
- Mis-specified distribution
- Outliers
- Small numbers of observations for any set of unique covariate values

So, best to test using (predictive) simulation techniques discussed earlier (e.g., using the Pearson χ^2 statistic).

Consequences of Overdispersion

We may obtain reasonable estimates of β , but:

- SE may be too small
- We may select overly complex models

Consequences of Overdispersion

We may obtain reasonable estimates of β , but:

- SE may be too small
- We may select overly complex models

Before 'correcting for overdispersion' , consider whether:

- You may have left out important predictors
- If you need to allow for non-linear relationships (residual plots).

How to deal with overdispersion

Model comparisons

Relaxing the Poisson Assumption

Poisson regression:

- $\log(\lambda_i) = \log(\mu_i) = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}$
- $f(y_i) \sim \text{Poisson}(\lambda_i)$

Relaxing the Poisson Assumption

Poisson regression:

- $\log(\lambda_i) = \log(\mu_i) = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}$
- $f(y_i) \sim \text{Poisson}(\lambda_i)$
- $E[Y_i|X_i] = \lambda_i = e^{\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}}$
- $\text{Var}[Y_i|X_i] = \lambda_i = e^{\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}}$

Relaxing the Poisson Assumption

Poisson regression:

- $\log(\lambda_i) = \log(\mu_i) = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}$
- $f(y_i) \sim \text{Poisson}(\lambda_i)$
- $E[Y_i|X_i] = \lambda_i = e^{\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}}$
- $\text{Var}[Y_i|X_i] = \lambda_i = e^{\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}}$

What if $E[Y_i|X_i] = \exp(\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i})$ is appropriate, but the Poisson distribution is not?

Relaxing the Poisson Assumption

Poisson regression:

- $\log(\lambda_i) = \log(\mu_i) = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}$
- $f(y_i) \sim \text{Poisson}(\lambda_i)$
- $E[Y_i|X_i] = \lambda_i = e^{\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}}$
- $\text{Var}[Y_i|X_i] = \lambda_i = e^{\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}}$

What if $E[Y_i|X_i] = \exp(\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i})$ is appropriate, but the Poisson distribution is not?

In particular, what if $\text{Var}[Y_i|X_i] > E[Y_i|X_i]$?

Relaxing the Poisson Assumption

Poisson regression:

- $\log(\lambda_i) = \log(\mu_i) = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}$
- $f(y_i) \sim \text{Poisson}(\lambda_i)$
- $E[Y_i|X_i] = \lambda_i = e^{\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}}$
- $\text{Var}[Y_i|X_i] = \lambda_i = e^{\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}}$

What if $E[Y_i|X_i] = \exp(\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i})$ is appropriate, but the Poisson distribution is not?

In particular, what if $\text{Var}[Y_i|X_i] > E[Y_i|X_i]$?

Option 1: Bootstrap!

Option 2: Variance Inflation

Another option: add a scale parameter to inflate variances.

- $E[Y_i|X_i] = \mu = e^{\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}}$
- $Var[Y_i|x_i] = \phi\mu = \phi e^{\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}}$

Option 2: Variance Inflation

Another option: add a scale parameter to inflate variances.

- $E[Y_i|X_i] = \mu = e^{\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}}$
- $Var[Y_i|x_i] = \phi\mu = \phi e^{\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}}$

Estimate of overdispersion, $\hat{\phi}$ by either:

- $\hat{\phi} = \frac{\text{Residual deviance}}{(n-p)}$
- $\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^n \frac{(Y_i - E[Y_i|X_i])^2}{Var[Y_i|X_i]}$ (what R does, probably better)

Option 2: Variance Inflation

Another option: add a scale parameter to inflate variances.

- $E[Y_i|X_i] = \mu = e^{\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}}$
- $Var[Y_i|x_i] = \phi\mu = \phi e^{\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}}$

Estimate of overdispersion, $\hat{\phi}$ by either:

- $\hat{\phi} = \frac{\text{Residual deviance}}{(n-p)}$
- $\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^n \frac{(Y_i - E[Y_i|X_i])^2}{Var[Y_i|X_i]}$ (what R does, probably better)

Zuur et al. recommend:

- If $\phi > 1.5$ should adjust for overdispersion
- If greater than 15 or 20, consider alternative methods (Negative Binomial, zero-inflation models, Poisson-Normal model)

Variance Inflation: Quasilikelihood

In R, can use `glm` with `family = quasipoisson()`

Variance Inflation: Quasilielihood

In R, can use `glm` with `family = quasipoisson()`

Will estimate an inflation factor using:

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^n \frac{(Y_i - E[Y_i|X_i])^2}{\text{Var}[Y_i|X_i]}$$

Variance Inflation: Quasilielihood

In R, can use `glm` with `family = quasipoisson()`

Will estimate an inflation factor using:

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^n \frac{(Y_i - E[Y_i|X_i])^2}{\text{Var}[Y_i|X_i]}$$

- $\hat{\beta}$ will be unchanged, but SE will be larger by a factor of $\sqrt{\hat{\phi}}$.
- No longer “maximum likelihood”
- **quasilielihood** (more on this later)
- Modeling the first two moments of Y ($E[Y|X]$, $\text{Var}[Y|X]$)

Poisson Model with Overdispersion parameter

Fit to slug data:

- $\hat{\beta}$ does not change
- SE's inflated by $\sqrt{\hat{\phi}}$

Poisson Model with Overdispersion parameter

Fit to slug data:

- $\hat{\beta}$ does not change
- SE's inflated by $\sqrt{\hat{\phi}}$

Lets do this!

Option 3: Use a different distribution than Poisson

Negative Binomial: $X \sim \text{NegBin}(\mu, \theta)$

$$P(Y = y) = \binom{y+\theta-1}{y} \left(\frac{\theta}{\mu+\theta}\right)^\theta \left(\frac{\mu}{\mu+\theta}\right)^y$$

- $E[Y] = \mu$
- $\text{Var}(Y) = \mu + \frac{\mu^2}{\theta}$

Option 3: Use a different distribution than Poisson

Negative Binomial: $X \sim \text{NegBin}(\mu, \theta)$

$$P(Y = y) = \binom{y+\theta-1}{y} \left(\frac{\theta}{\mu+\theta}\right)^\theta \left(\frac{\mu}{\mu+\theta}\right)^y$$

- $E[Y] = \mu$
- $\text{Var}(Y) = \mu + \frac{\mu^2}{\theta}$
- In r: ***nbinom**, with parameters (`prob = p`, `size = n`) or (`mu = \mu`, `size = \theta`)
- JAGS: **dnegbin** with parameters (p, θ)

Option 3: Use a different distribution than Poisson

Negative Binomial: $X \sim \text{NegBin}(\mu, \theta)$

$$P(Y = y) = \binom{y+\theta-1}{y} \left(\frac{\theta}{\mu+\theta}\right)^\theta \left(\frac{\mu}{\mu+\theta}\right)^y$$

- $E[Y] = \mu$
- $\text{Var}(Y) = \mu + \frac{\mu^2}{\theta}$
- In r: ***nbinom**, with parameters (`prob = p`, `size = n`) or (`mu = μ` , `size = θ`)
- JAGS: **dnegbin** with parameters (p, θ)

Overdispersed relative to Poisson ($\text{Var}(x)/E[x] = 1 + \frac{\mu}{\theta}$) versus 1 for Poisson

Option 3: Use a different distribution than Poisson

Negative Binomial: $X \sim \text{NegBin}(\mu, \theta)$

$$P(Y = y) = \binom{y+\theta-1}{y} \left(\frac{\theta}{\mu+\theta}\right)^\theta \left(\frac{\mu}{\mu+\theta}\right)^y$$

- $E[Y] = \mu$
- $\text{Var}(Y) = \mu + \frac{\mu^2}{\theta}$
- In r: ***nbinom**, with parameters (`prob = p`, `size = n`) or (`mu = μ` , `size = θ`)
- JAGS: **dnegbin** with parameters (p, θ)

Overdispersed relative to Poisson ($\text{Var}(x)/E[x] = 1 + \frac{\mu}{\theta}$) versus 1 for Poisson

Poisson is a limiting case (when $\theta \rightarrow \infty$)

Negative Binomial Models in R

Can fit negative binomial models in R using the `glm.nb` function in the MASS library

- $\log(E[Y_i|X_i]) = \log(\mu_i) = \beta_0 + \beta_1 X_{1,i} + \dots \beta_p X_{p,i}$
- $Y_i|X_i \sim \text{NegBin}(\mu_i, \theta)$

```
glm.nb(y ~ x, data=)
```

Negative Binomial Models in R

Can fit negative binomial models in R using the `glm.nb` function in the MASS library

- $\log(E[Y_i|X_i]) = \log(\mu_i) = \beta_0 + \beta_1 X_{1,i} + \dots \beta_p X_{p,i}$
- $Y_i|X_i \sim \text{NegBin}(\mu_i, \theta)$

```
glm.nb(y ~ x, data=)
```

Lets do this!

Negative Binomial in JAGS

JAGS: `dnegbin` specified in terms of parameters (p, θ)

Negative Binomial in JAGS

JAGS: **dnegbin** specified in terms of parameters (p, θ)

We will specify the model in terms of μ and θ , then solve for p :

```
log(mu[i]) <- alpha + beta*IRook[i]  
p[i] <- theta/(theta+mu[i])  
slugs[i] ~ dnegbin(p[i],theta)
```


Negative Binomial in JAGS

JAGS: **dnegbin** specified in terms of parameters (p, θ)

We will specify the model in terms of μ and θ , then solve for p :

```
log(mu[i]) <- alpha + beta*IRook[i]  
p[i] <- theta/(theta+mu[i])  
slugs[i] ~ dnegbin(p[i],theta)
```

- $E[Y_i|X_i] = \mu_i = e^{\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}}$

Negative Binomial in JAGS

JAGS: **dnegbin** specified in terms of parameters (p, θ)

We will specify the model in terms of μ and θ , then solve for p :

```
log(mu[i]) <- alpha + beta*IRook[i]  
p[i] <- theta/(theta+mu[i])  
slugs[i] ~ dnegbin(p[i],theta)
```

- $E[Y_i|X_i] = \mu_i = e^{\beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}}$
- $Var[Y_i|X_i] = \mu_i + \frac{\mu_i^2}{\theta}$

Practice Opportunity: Fit this model to slug data (and use the above mean/variance equations) to fill in the code needed to calculate the Bayesian p-value.

Poisson-normal model

$$\begin{aligned} \log(\lambda_i) = \log(\mu_i) &= \beta_0 + \beta_1 X_{1,i} + \dots \beta_p X_{p,i} + \epsilon_i \\ \epsilon_i &\sim N(0, \sigma^2) \end{aligned}$$

Poisson-normal model

$$\begin{aligned}\log(\lambda_i) &= \log(\mu_i) = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i} + \epsilon_i \\ \epsilon_i &\sim N(0, \sigma^2)\end{aligned}$$

- $E[Y_i | X_i, \epsilon_i = 0] = \mu = \exp^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p} \neq E[Y_i | X_i]$

Poisson-normal model

$$\log(\lambda_i) = \log(\mu_i) = \beta_0 + \beta_1 X_{1,i} + \dots \beta_p X_{p,i} + \epsilon_i$$
$$\epsilon_i \sim N(0, \sigma^2)$$

- $E[Y_i|X_i, \epsilon_i = 0] = \mu = \exp^{\beta_0 + \beta_1 x_1 + \dots \beta_p x_p} \neq E[Y_i|X_i]$
- $E[Y_i|X_i] = e^{\beta_0 + \beta_1 X_{1,i} + \dots \beta_p X_{p,i} + \frac{1}{2}\sigma^2}$

Poisson-normal model

$$\log(\lambda_i) = \log(\mu_i) = \beta_0 + \beta_1 X_{1,i} + \dots \beta_p X_{p,i} + \epsilon_i$$
$$\epsilon_i \sim N(0, \sigma^2)$$

- $E[Y_i|X_i, \epsilon_i = 0] = \mu = \exp^{\beta_0 + \beta_1 x_1 + \dots \beta_p x_p} \neq E[Y_i|X_i]$
- $E[Y_i|X_i] = e^{\beta_0 + \beta_1 X_{1,i} + \dots \beta_p X_{p,i} + \frac{1}{2}\sigma^2}$
- $Var[Y_i|X_i] = \mu_i + (e^{\sigma^2} - 1)\mu_i^2$

Poisson-normal model

Practice Opportunity:

- Fit this model in JAGS, by adapting your old code
- Add appropriate code for the goodness of fit test

Or, have a look at Rglms.html file with solutions.

How to deal with overdispersion

Model comparisons

AIC

$$\text{AIC} = -2 \times \log L(\hat{\theta}|y) + 2 \times \text{number of parameters}$$

- measure of “fit” with “penalty” for model complexity

AIC

$$\text{AIC} = -2 \times \log L(\hat{\theta}|y) + 2 \times \text{number of parameters}$$

- measure of “fit” with “penalty” for model complexity
- the larger log-likelihood, the smaller AIC

AIC

$$\text{AIC} = -2 \times \log L(\hat{\theta}|y) + 2 \times \text{number of parameters}$$

- measure of “fit” with “penalty” for model complexity
- the larger log-likelihood, the smaller AIC
- for similar likelihoods, AIC will be smaller for simpler models

AIC

$$\text{AIC} = -2 \times \log L(\hat{\theta}|y) + 2 \times \text{number of parameters}$$

- measure of “fit” with “penalty” for model complexity
- the larger log-likelihood, the smaller AIC
- for similar likelihoods, AIC will be smaller for simpler models
- → smaller AIC is better

Model Comparisons

DIC = **Deviance Information Criterion** = is often used to compare models.

Deviance: $D = -2\log L(\theta|y)$ (has a posterior distribution)

Model Comparisons

DIC = **Deviance Information Criterion** = is often used to compare models.

Deviance: $D = -2\log L(\theta|y)$ (has a posterior distribution)

DIC is calculated from:

- \overline{D} = the mean of the posterior deviance
- \hat{D} = the deviance calculated using the posterior means of the model parameters

Estimated effective number of parameters in a model, $p_D = \overline{D} - \hat{D}$, and DIC as $\hat{D} + 2p_D$:

$$\text{DIC} = \hat{D} + 2p_D = (\overline{D} - p_D) + 2p_D = \overline{D} + p_D$$

Not without its problems:

- Not always appropriate for certain types of models (problematic if you have latent variables, e.g., mixture models)

DIC

Martyn Plummer (creator of JAGS):

DIC [like AIC] is (an approximation to) a theoretical out-of-sample predictive error.

“The deviance information criterion (DIC) is widely used for Bayesian model comparison, despite the lack of a clear theoretical foundation. . . . valid only when the effective number of parameters in the model is much smaller than the number of independent observations. In disease mapping, a typical application of DIC, this assumption does not hold and DIC under-penalizes more complex models. Another deviance-based loss function, derived from the same decision-theoretic framework, is applied to mixture models, which have previously been considered an unsuitable application for DIC.”

Andrew Gelman: “I don’t really ever know what to make of DIC. On one hand, it seems sensible... On the other hand, I don’t really have any idea what I would do with DIC in any real example. In our book we included an example of DIC—people use it and we don’t have any great alternatives—but I had to be pretty careful that the example made sense. Unlike the usual setting where we use a method and that gives us insight into a problem, here we used our insight into the problem to make sure that in this particular case the method gave a reasonable answer.”

http://andrewgelman.com/2011/06/22/deviance_dic_ai/

Model comparisons

There are other potential options out there (e.g., WIC, cross-validation estimates of predictive error, etc)

Hooten, Mevin B, and N Thompson Hobbs. 2015. "A Guide to Bayesian Model Selection for Ecologists." *Ecological Monographs* 85 (1): 3–28.

Offsets

Count data, Y , are often collected:

- over varying lengths of time
- in sample units that have different areas

Offsets

Count data, Y , are often collected:

- over varying lengths of time
- in sample units that have different areas

We may be interested in modeling rates:

$$E[Y_i|X_i]/\text{Time}_i$$

Offsets

Count data, Y , are often collected:

- over varying lengths of time
- in sample units that have different areas

We may be interested in modeling rates:

$$E[Y_i|X_i]/\text{Time}_i$$

Or densities:

$$E[Y_i|X_i]/\text{Area}_i$$

We may want to account for variable survey effort (varying times or areas)!

Offsets

Poisson and negative binomial models for rate data:

$$\log(E[Y_i|X_i]/\text{Time}_i) = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}$$

Offsets

Poisson and negative binomial models for rate data:

$$\log(E[Y_i|X_i]/\text{Time}_i) = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}$$

$$\Rightarrow \log(E[Y_i|X_i]) - \log(\text{Time}_i) = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}$$

Offsets

Poisson and negative binomial models for rate data:

$$\log(E[Y_i|X_i]/\text{Time}_i) = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}$$

$$\Rightarrow \log(E[Y_i|X_i]) - \log(\text{Time}_i) = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}$$

$$\Rightarrow \log(E[Y_i|X_i]) = \log(\text{Time}_i) + \beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}$$

Offsets

Poisson and negative binomial models for rate data:

$$\log(E[Y_i|X_i]/\text{Time}_i) = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}$$

$$\Rightarrow \log(E[Y_i|X_i]) - \log(\text{Time}_i) = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}$$

$$\Rightarrow \log(E[Y_i|X_i]) = \log(\text{Time}_i) + \beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}$$

$\log(\text{Time}_i)$ is called an **offset** and can be modeled using:

```
glm(y~x + offset(log(time)), data= , family =  
poisson())
```

Offsets

Poisson and negative binomial models for rate data:

$$\log(E[Y_i|X_i]/\text{Time}_i) = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}$$

$$\Rightarrow \log(E[Y_i|X_i]) - \log(\text{Time}_i) = \beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}$$

$$\Rightarrow \log(E[Y_i|X_i]) = \log(\text{Time}_i) + \beta_0 + \beta_1 X_{1,i} + \dots + \beta_p X_{p,i}$$

$\log(\text{Time}_i)$ is called an **offset** and can be modeled using:

```
glm(y~x + offset(log(time)), data= , family =  
poisson())
```

An **offset** is an explanatory variable with a regression coefficient fixed at 1.

See **PoissonOffsetTemplate.R** and **PoissonOffset.R** (in the Generalized linear models folder) for an exercise fitting a Poisson model with an offset.