

TMA4315 Generalized linear models H2018

Module 4: Count and continuous positive response data (Poisson and gamma regression)

Mette Langaas, Department of Mathematical Sciences, NTNU
– with contributions from Ingeborg Hem

27.09.2018 and 04.10.2018 [PL], 28.09.2018 and 05.10.2018
[IL]

(Latest changes: 06.10: solutions added. 01.10: small changes for second week. 27.09: added one Problem for ILw1, moved stuff to w2, added a few dimensions to score test.)

Overview

Learning material

- ▶ Textbook: Fahrmeir et al (2013): Chapter 5.2, 5.3.
- ▶ Classnotes 27.09.2018
- ▶ Classnotes 04.10.2018

Topics

First week

- ▶ examples of count data
- ▶ the Poisson distribution
- ▶ regression with count data
- ▶ Poisson regression with log-link
- ▶ parameter estimation (ML): log-likelihood, score vector, information matrix to give iterative calculations
- ▶ asymptotic MLE properties
- ▶ confidence intervals and hypothesis tests (Wald, score and LRT)

Second week

- ▶ Count data with Poisson regression (continued)
- ▶ deviance, model fit and model choice
- ▶ overdispersion
- ▶ rate models and offset
- ▶ Modelling continuous response data: lognormal and gamma
- ▶ the gamma distribution
- ▶ the gamma GLM model
- ▶ gamma likelihood and derivations thereof
- ▶ dispersion parameter: scaled and unscaled deviance

FIRST WEEK

Examples of count data

- ▶ the number of automobile thefts pr city worldwide
- ▶ the number of UFO sightings around the world
- ▶ the number of visits at web pages
- ▶ the number of male crabs (satellites) residing nearby a female crab
- ▶ the number of goals by the home team and the number of goals for the away team in soccer
- ▶ the number of newspapers sold at newsagents
- ▶ the number of vampires living on Sesame Street

Sales of newspapers

Response data: number of newspapers (delivered) sold at each outlet. Covariate data: type of outlet, but mainly calendar information= weekday, month, season, public holidays, winter/autumn/easter/xmas, ...

The aim: predict the number of newspapers sold at each outlet (11,000!) in Norway on any day

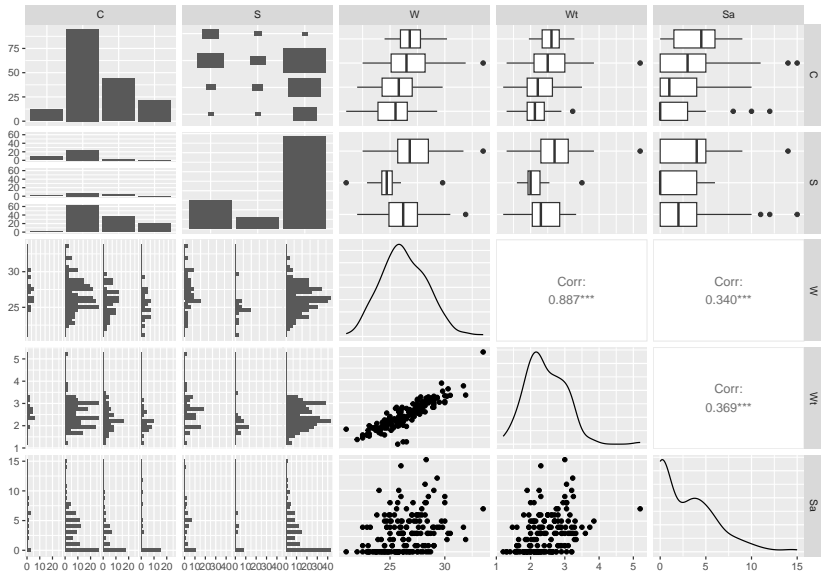
Can use this to optimise the number of newspapers printed and delivered

Female horseshoe crabs with satellites

The study objects were female horseshoe crabs. Each female horseshoe crab had a male attached to her in her nest. The objective of the study was to investigate factors that affect whether the female had any other males, called satellites, residing near her. The following covariates were collected for 173 female horseshoe crabs:

- ▶ C: the color of the female horseshoe crab (1=light medium, 2=medium, 3=dark medium, 4=dark)
- ▶ S: spine condition (1=both good, 2=one worn or broken, 3=both worn or broken)
- ▶ W: width of carapace (cm)
- ▶ Wt: weight (kg)

The response was the number of satellites, S_a = male horseshoe crabs residing nearby.



Q: Discuss what you see. Any potential covariates to influence Sa? Which distribution can Sa have?

Modelling counts with the Poisson distribution

The Poisson process

We observe events that may occur within a time interval or a region.

1. The number of events occurring within a time interval or a region, is independent of the number of events that occurs in any other disjoint (non-overlapping) time interval or region.
2. The probability that a single event occurs within a small time interval or region, is proportional to the length of the interval or the size of the region.
3. The probability that more than one event may occur within a small time interval or region is negligible.

When all of these three properties are fulfilled we have a *Poisson process*. This leads to three distributions

- ▶ The number of events in a Poisson process follows a Poisson distribution.
- ▶ Time between two events in a Poisson process follows an exponential distribution.
- ▶ Time between many events in a Poisson process follows a gamma distribution.

We will first study the Poisson distribution - and link it to a regression setting.

The Poisson distribution

We study a Poisson process within a time interval or a region of specified size. Then, the number of events, Y , will follow a *Poisson distribution* with parameter λ

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

Here the parameter λ is the proportionality factor in the requirement 2 (above) for the Poisson process. Another popular parameterization is μ , or given some interval λt , but we will stick with λ . In R we calculate the Poisson point probabilities using `dpois`.

Expected value and variance

Let Y follow a Poisson distribution with parameter λ . Then

$$E(Y) = \lambda \text{ and } \text{Var}(Y) = \lambda$$

(proofs in notes)

Properties of the Poisson distribution

- ▶ A sum of n independent Poisson distributed random variables, Y_i with means λ_i are Poisson distributed with mean $\sum_{i=1}^n \lambda_i$.
- ▶ When the mean increases the Poisson distribution becomes more and more symmetric and for large λ the Poisson distribution can be approximated by a normal distribution.

Exponential family

In Module 1 we introduced distributions of the Y_i , that could be written in the form of a *univariate exponential family*

$$f(y_i | \theta_i) = \exp \left(\frac{y_i \theta_i - b(\theta_i)}{\phi} \cdot w_i + c(y_i, \phi, w_i) \right)$$

where we said that

- ▶ θ_i is called the canonical parameter and is a parameter of interest
- ▶ ϕ is called a nuisance parameter (and is not of interest to us=therefore a nuisance (plage))
- ▶ w_i is a weight function, in most cases $w_i = 1$
- ▶ b and c are known functions.

Exponentially Poisson

The log-likelihood for a Poisson is

$$l(\theta) = \ln L(\theta) = \sum_{i=1}^n \ln L_i(\beta) = \sum_{i=1}^n l_i(\beta) = \sum_{i=1}^n [y_i \ln(\lambda_i) - \lambda_i - \ln(y!)]$$

So, comparing with

$$l(\theta_i; y_i) = \frac{y_i \theta_i - b(\theta_i)}{\phi} \cdot w_i + c(y_i, \phi, w_i)$$

we get

- ▶ $\theta_i = \ln(\lambda_i)$ is the canonical parameter
- ▶ $\phi = 1$, no nuisance
- ▶ $w_i = 1$
- ▶ $b(\theta_i) = \exp(\theta_i)$
- ▶ $\mu_i = E(Y_i) = \lambda_i$

For a GLM with linear predictor η_i - to have a canonical link we need

$$\theta_i = \eta_i$$

Since $\eta_i = g(\mu_i) = g(\lambda_i)$ this means to us that we need

$$g(\mu_i) = g(\lambda_i) = \theta_i$$

saying that with the Poisson the canonical link is $\ln(\lambda_i)$.

Q: Why may we want to choose a canonical link?

Regression with count data

Aim

1. Construct a model to help understand the relationship between a count variable and one or many possible explanatory variables. The response measurements are counts.
2. Use the model for understanding what can explain count, and for prediction of counts.

(you can use OLS, but you might have to transform the response to stabilise the variance. If the counts are all large, OLS will work fine)

The log-linear Poisson model

Assumptions:

1. $Y_i \sim \text{Poisson}(\lambda_i)$, with $E(Y_i) = \lambda_i$, and $\text{Var}(Y_i) = \lambda_i$.
2. Linear predictor: $\eta_i = \mathbf{x}_i^T \beta$.
3. Log link

$$\eta_i = \ln(\lambda_i) = g(\lambda_i)$$

and (inverse thereof) response function

$$\lambda_i = \exp(\eta_i)$$

Assumptions 1 and 3 above can be written as

$$Y_i \sim \text{Poisson}(\exp(\eta_i)), \quad i = 1, \dots, n$$

Interpreting parameters in the log-linear Poisson model

In the log-linear model the mean, $E(Y_i) = \lambda_i$ satisfy an exponential relationship to covariates

$$\lambda_i = \exp(\eta_i) = \exp(\mathbf{x}_i^T \beta) = \exp(\beta_0) \cdot \exp(\beta_1)^{x_{i1}} \cdots \exp(\beta_k)^{x_{ik}}.$$

Let us look in detail at β_1 with covariate x_{i1} for observation i .

1. If x_{i1} increases by one unit to $x_{i1} + 1$ then the mean $E(Y_i)$ will in our model change by a factor $\exp(\beta_1)$.
2. If $\beta_1 = 0$ then $\exp(\beta_1) = 1$, so that a change in x_{i1} does not change $E(Y_i)$.
3. If $\beta_1 < 0$ then $\exp(\beta_1) < 1$ so if x_{i1} increase then $E(Y_i)$ decrease.
4. If $\beta_1 > 0$ then $\exp(\beta_1) > 1$ so if x_{i1} increase then $E(Y_i)$ increase.

Thus, the covariates have a multiplicative effect on the rate λ_i .

Example: interpreting parameters for the female crabs with satellites

We fit a log-linear model to Sa, assuming the number of satellites follows a Poisson distribution with log-link, and use S (spine condition) as a covariate.

```
## Intercept +  
## (Intercept)          S2          S3  
##    1.2943569   -0.6012097   -0.2612018  
## Intercept+W, exp  
## (Intercept)          S2          S3  
##    3.6486486    0.5481482    0.7701255
```

(1=both good, 2=one worn or broken, 3=both worn or broken)

So S2 roughly halves the number of satellite males

Parameter estimation with maximum likelihood

We would like to estimate β by maximizing the likelihood -

This is essentially the same as for Module 3: Binary regression - with “Poisson and log” instead of “Bernoulli and logit”.

Likelihood $L(\beta)$

We assume that pairs of covariates and response are measured independently of each other: (\mathbf{x}_i, Y_i) , and Y_i follows the distribution specified above, and \mathbf{x}_i is fixed.

$$L(\beta) = \prod_{i=1}^n L_i(\beta) = \prod_{i=1}^n f(y_i; \beta) = \prod_{i=1}^n \frac{\lambda_i^{y_i}}{y_i!} \exp(-\lambda_i)$$

Note: still a slight misuse of notation - where is β ?

Loglikelihood $l(\beta)$

$$l(\beta) = \ln L(\beta) = \sum_{i=1}^n \ln L_i(\beta) = \sum_{i=1}^n l_i(\beta) = \sum_{i=1}^n [y_i \ln(\lambda_i) - \lambda_i - \ln(y!)]$$

Observe that the log-likelihood is a sum of individual contributions for each observation pair i . We often omit the last term since it is not a function of model parameters, only data.

If we want a function of $\eta_i = \ln(\lambda_i)$ or β :

$$l(\beta) = \sum_{i=1}^n [y_i \eta_i - \exp(\eta_i) + C_i] = \sum_{i=1}^n y_i \mathbf{x}_i^T \beta - \sum_{i=1}^n \exp(\mathbf{x}_i^T \beta) + C$$

Score function $s(\beta)$

The score function is a $p \times 1$ vector, $s(\beta)$, with the partial derivatives of the log-likelihood with respect to the p elements of the β vector. Remember, the score function is linear in the individual contributions:

$$s(\beta) = \frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^n \frac{\partial l_i(\beta)}{\partial \beta} = \sum_{i=1}^n s_i(\beta)$$

We use the chain rule to find $s_i(\beta)$.

$$\begin{aligned} s_i(\beta) &= \frac{\partial l_i(\beta)}{\partial \beta} = \frac{\partial l_i(\beta)}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta} \\ &= \frac{\partial [y_i \eta_i - \exp(\eta_i) + C_i]}{\partial \eta_i} \cdot \frac{\partial [\mathbf{x}_i^T \beta]}{\partial \beta} \\ &= [y_i - \exp(\eta_i)] \cdot \mathbf{x}_i \\ &= (y_i - \lambda_i) \mathbf{x}_i \end{aligned}$$

(see Module 3 for rules for partial derivatives of scalar wrt vector)

The score function is given as:

$$s(\beta) = \sum_{i=1}^n s_i(\beta) = \sum_{i=1}^n (y_i - \lambda_i) \mathbf{x}_i$$

So

$$E(s_i(\beta)) = E((Y_i - \lambda_i) \mathbf{x}_i) = 0$$

(because $E(Y_i) = \lambda_i$, thus $E(s(\beta)) = 0$).

Fisher Information

We will also need the Fisher Information, for the estimation (=a numerical optimisation problem), and to estimate the covariances

- ▶ the **expected** Fisher information matrix, $F(\beta) = \text{Cov}(\beta)$
- ▶ the **observed** Fisher information matrix,
 $H(\beta) = -\partial^2 s(\beta) / \partial \beta^T \partial \beta$
- ▶ for us $F(\beta) = H(\beta)$

The expected Fisher information matrix $F(\beta)$

$$F(\beta) = \text{Cov}(s(\beta)) = \sum_{i=1}^n \text{Cov}(s_i(\beta)) = \cdots = \sum_{i=1}^n F_i(\beta)$$

assuming the responses Y_i and Y_j are independent, and that $E(s_i(\beta)) = 0 \ \forall i$.

Remember that $s_i(\beta) = (Y_i - \lambda_i)\mathbf{x}_i$, then:

$$\begin{aligned} F_i(\beta) &= E(s_i(\beta)s_i(\beta)^T) = E((Y_i - \lambda_i)\mathbf{x}_i(Y_i - \lambda_i)\mathbf{x}_i^T) \\ &= \mathbf{x}_i\mathbf{x}_i^T E((Y_i - \lambda_i)^2) \\ &= \mathbf{x}_i\mathbf{x}_i^T \lambda_i \end{aligned}$$

where $E((Y_i - \lambda_i)^2) = \text{Var}(Y_i) = \lambda$ is the variance of Y_i . Thus

$$F(\beta) = \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T \lambda_i.$$

Observed Fisher information matrix $H(\beta)$

Since we use canonical link $H(\beta) = F(\beta)$. But, for completeness, we add the direct derivation of $H(\beta)$.

$$\begin{aligned} H(\beta) &= -\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta^T} = -\frac{\partial s(\beta)}{\partial \beta^T} \\ &= \frac{\partial}{\partial \beta^T} \left[\sum_{i=1}^n (\lambda_i - y_i) \mathbf{x}_i \right] \quad (s(\beta) = \sum_{i=1}^n (y_i - \lambda_i) \mathbf{x}_i) \\ &= \frac{\partial}{\partial \beta^T} \left[\sum_{i=1}^n (\exp(\eta_i) - y_i) \mathbf{x}_i \right] \quad (\lambda_i = \exp(\eta_i)) \end{aligned}$$

$$H(\beta) = \sum_{i=1}^n \frac{\partial}{\partial \beta^T} [\mathbf{x}_i \lambda_i - \mathbf{x}_i y_i] = \sum_{i=1}^n \frac{\partial}{\partial \beta^T} \mathbf{x}_i \lambda_i = \sum_{i=1}^n \mathbf{x}_i \frac{\partial \lambda_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta^T}$$

Use that

$$\frac{\partial \eta_i}{\partial \beta^T} = \frac{\partial \mathbf{x}_i^T \beta}{\partial \beta^T} = \left(\frac{\partial \mathbf{x}_i^T \beta}{\partial \beta} \right)^T = \mathbf{x}_i^T$$

and

$$\frac{\partial \lambda_i}{\partial \eta_i} = \frac{\partial \exp(\eta_i)}{\partial \eta_i} = \exp(\eta_i) = \lambda_i$$

And thus

$$H(\beta) = \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T \lambda_i.$$

How to solve it

To find the maximum likelihood estimate $\hat{\beta}$ we solve the set of p non-linear equations:

$$s(\hat{\beta}) = 0$$

And, as before we do that using the Newton-Raphson or Fisher Scoring iterative methods, so we need the derivative of the score vector (our Fisher information).

Parameter estimation - in practice

To find the ML estimate $\hat{\beta}$ we need to solve

$$s(\hat{\beta}) = 0$$

We have that the score function for the log-linear model is:

$$s(\beta) = \sum_{i=1}^n \mathbf{x}_i (y_i - \lambda_i) = \sum_{i=1}^n \mathbf{x}_i (y_i - \exp(\mathbf{x}_i^T \beta)).$$

Observe that this is a non-linear function in β , and has no closed form solution (except for a few special cases).

Fisher scoring

To solve this we use the Fisher scoring algorithm, where we at iteration $t + 1$ have

$$\beta^{(t+1)} = \beta^{(t)} + F(\beta^{(t)})^{-1} s(\beta^{(t)})$$

Remark: what do we need to do to use the Newton-Raphson method instead? Well, replace F with H , but for canonical link (which is the log-link for the Poisson) $F = H$.

Requirements for convergence

For the Fisher scoring algorithm the expected Fisher information matrix F needs to be invertible. For this we need:

- ▶ $\lambda_i > 0$ for all i
- ▶ design matrix (X) has full rank (p) .

For the log link $\lambda_i = \exp(\mathbf{x}_i^T \beta)$ which is always positive.

Note, with the linear link $\lambda_i = \eta_i$ this might be a challenge, and restrictions on β must be set.

The algorithm might not converge, if the data are evil enough, particularly with small samples.

Statistical inference

Asymptotic properties of ML estimates

We repeat what we found for Module 3: Under some (weak) regularity conditions:

Let $\hat{\beta}$ be the maximum likelihood (ML) estimate in the GLM model. As the total sample size increases, $n \rightarrow \infty$:

1. $\hat{\beta}$ exists
2. $\hat{\beta}$ is consistent (convergence in probability, yielding asymptotically unbiased estimator, variances goes towards 0)
3. $\hat{\beta} \approx N_p(\beta, F^{-1}(\hat{\beta}))$

Observe that this means that asymptotically $\text{Cov}(\hat{\beta}) = F^{-1}(\hat{\beta})$: the inverse of the expected Fisher information matrix evaluated at the ML estimate.

In our case we have

$$F(\beta) = \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T \lambda_i = \mathbf{X}^T \mathbf{W} \mathbf{X},$$

where $\mathbf{W} = \text{diag}(\lambda_i)$. This means $\text{Cov}(\hat{\beta}) = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}$ (remember that $\hat{\beta}$ comes in with $\hat{\lambda}_i$ in \mathbf{W}).

Let $\mathbf{A}(\beta) = F^{-1}(\beta)$, and $a_{jj}(\beta)$ is diagonal element number j .

For one element of the parameter vector:

$$Z_j = \frac{\hat{\beta}_j - \beta_j}{\sqrt{a_{jj}(\hat{\beta})}}$$

is standard normal, which can be used to make confidence intervals
- and test hypotheses.

Confidence intervals

In addition to providing a parameter estimate for each element of our parameter vector β we should also report a $(1 - \alpha)100\%$ confidence interval (CI) for each element.

Let $z_{\alpha/2}$ be such that $P(Z_j > z_{\alpha/2}) = \alpha/2$. We then use

$$P(-z_{\alpha/2} \leq Z_j \leq z_{\alpha/2}) = 1 - \alpha$$

insert Z_j and solve for β_j to get

$$P(\hat{\beta}_j - z_{\alpha/2} \sqrt{a_{jj}(\hat{\beta})} \leq \beta_j \leq \hat{\beta}_j + z_{\alpha/2} \sqrt{a_{jj}(\hat{\beta})}) = 1 - \alpha$$

A $(1 - \alpha)\%$ CI for β_j is when we insert numerical values for the upper and lower limits.

Example: Female crabs with satellites

```
model2 = glm(Sa ~ W, family = poisson(link = log), data = c
# summary(model2)
lower = model2$coefficients - qnorm(0.975) * sqrt(diag(vcov
upper = model2$coefficients + qnorm(0.975) * sqrt(diag(vcov
cbind(lower, upper)
confint(model2)
```

##	lower	upper
## (Intercept)	-4.3675312	-2.2419833
## W	0.1249137	0.2031764
##	2.5 %	97.5 %
## (Intercept)	-4.3662326	-2.2406858
## W	0.1247244	0.2029871

Hypothesis testing

There are three methods that are mainly used for testing hypotheses in GLMs

- ▶ Wald test,
- ▶ likelihood ratio test and
- ▶ score test.

The Wald and likelihood ratio test are the same as in Module 3...

Hypotheses

$$H_0 : \mathbf{C} = \mathbf{d} \text{ vs. } H_1 : \mathbf{C} \neq \mathbf{d}$$

We specify \mathbf{C} to be a $r \times p$ matrix and \mathbf{d} to be a column vector of length r ,

and/or where we define

- ▶ A: the larger model and
- ▶ B: the smaller model (under H_0), and the smaller model is nested within the larger model (i.e. $B \subset A$).

The Wald test

The Wald test statistic is:

$$w = (\mathbf{C}\hat{\beta} - \mathbf{d})^T [\mathbf{C}F^{-1}(\hat{\beta})\mathbf{C}^T]^{-1} (\mathbf{C}\hat{\beta} - \mathbf{d})$$

it measures the distance between the estimate $\mathbf{C}\hat{\beta}$ and \mathbf{d}

Under the null follows a χ^2 distribution with r degrees of freedom (where r is the number of hypotheses tested).

P -values are calculated in the upper tail of the χ^2 -distribution.

The likelihood ratio test

A: the larger model and B: the smaller model (under H_0)

The likelihood ratio statistic is defined as

$$-2 \ln \lambda = -2(\ln L(\hat{\beta}_B) - \ln L(\hat{\beta}_A))$$

under the null is asymptotically χ^2 -distributed with degrees of freedom equal the difference in the number of parameters in A and B.

p -values are calculated in the upper tail of the χ^2 -distribution.

The score test

The *score statistic* is based on the *score function*, and measures the distance to the score function at the maximum likelihood for model A (which is 0) and scales with the covariance to form the test statistic.

- ▶ Under the null hypothesis investigated let $\tilde{\beta}$ be the ML estimate (that is, model B, the smaller model) - that means that this is a restricted ML estimate, and
- ▶ under H_1 we have the larger model (A) with maximum likelihood $\hat{\beta}$.

The score statistics is:

$$U = (s(\tilde{\beta}) - \mathbf{0})^T \mathbf{F}^{-1}(\tilde{\beta})(s(\tilde{\beta}) - \mathbf{0})$$

- ▶ $s(\tilde{\beta})$ is a subvector of the score function, for the elements in H_1 and not H_0
- ▶ dimension is the difference in number of parameters between A and B models
- ▶ the score function is evaluated based on parameter estimates under H_0 (i.e. the value of $\hat{\lambda}$ in the score and expected Fisher information is based on $\tilde{\beta}$).

To calculate $\mathbf{F}^{-1}(\tilde{\beta})$ this is a submatrix of the full inverted matrix (not invert just the submatrix). The dimension of this matrix is “difference in number of parameters between A and B models” squared.

When the null hypothesis is true $U \sim \chi_r^2$ (asymptotically): r is the difference in number of estimated parameters between the models.

Comments

- ▶ In R the function `statmod::glm.scoretest` the score test for a GLM when the difference between H_0 and H_1 is one parameter. The output is \sqrt{U} (what distribution does this follow?).
- ▶ The score test is called Rao's efficient score test in `add1` in R
- ▶ The score test is very useful for special situations when the smaller model is to be tested towards many larger models, because only the smaller model has to be fitted.
- ▶ The score test is perhaps the the most complex and least studied of the three tests, and in this course the main focus will be on the Wald and LRT tests.
- ▶ It is important for you to have heard of the score test, because in special situation it may be the preferred test.

Example: Female crabs with satellites - the different tests.

We fit a model with two covariates (Width & Colour): C is categorical and we use effect coding. We want to test if we need to add these covariates.

```
model3 = glm(Sa ~ W + C, family = poisson(link = log), data = data)
summary(model3)
```

```
##
```

```
## Call:
```

```
## glm(formula = Sa ~ W + C, family = poisson(link = log),
##      contrasts = list(C = "contr.sum"))
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.92089    0.56010  -5.215 1.84e-07 ***
## W            0.14934    0.02084   7.166 7.73e-13 ***
## C1           0.27085    0.11784   2.298  0.0215  *
## C2           0.07117    0.07296   0.975  0.3294
## C3          -0.16551    0.09316  -1.777  0.0756  .
## ---
```

Type III ANOVA, Score test

```
# possible to use type III anova
```

```
drop1(model3, test = "Rao") #Q:same as summary?
```

```
## Single term deletions
```

```
##
```

```
## Model:
```

```
## Sa ~ W + C
```

```
##           Df Deviance      AIC Rao score  Pr(>Chi)
```

```
## <none>           559.34 924.64
```

```
## W           1    609.14 972.44    51.514 7.108e-13 ***
```

```
## C           3    567.88 927.18     8.579  0.03545 *
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

Type III ANOVA, LRT

```
drop1(model3, test = "LRT") #same as comparing deviances
```

```
## Single term deletions
```

```
##
```

```
## Model:
```

```
## Sa ~ W + C
```

```
##           Df Deviance    AIC    LRT  Pr(>Chi)
```

```
## <none>           559.34 924.64
```

```
## W           1    609.14 972.44 49.794 1.707e-12 ***
```

```
## C           3    567.88 927.18  8.534  0.03618 *
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

Type I ANOVA, LRT

```
anova(model3, test = "LRT")
```

```
## Analysis of Deviance Table
```

```
##
```

```
## Model: poisson, link: log
```

```
##
```

```
## Response: Sa
```

```
##
```

```
## Terms added sequentially (first to last)
```

```
##
```

```
##
```

```
##      Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
```

```
## NULL                172      632.79
```

```
## W      1      64.913      171      567.88 7.828e-16 ***
```

```
## C      3       8.534      168      559.34  0.03618 *
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

Type I ANOVA, Score test

```
anova(model3, test = "Rao")
```

```
## Analysis of Deviance Table
```

```
##
```

```
## Model: poisson, link: log
```

```
##
```

```
## Response: Sa
```

```
##
```

```
## Terms added sequentially (first to last)
```

```
##
```

```
##
```

```
##      Df Deviance Resid. Df Resid. Dev      Rao Pr(>Chi)
```

```
## NULL                172      632.79
```

```
## W      1      64.913      171      567.88 67.474 < 2e-16 **
```

```
## C      3       8.534      168      559.34  8.579  0.03545 *
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

Yeah, let's stop here