**DEALING WITH COUNT DATA**

**Do not log-transform count data**

**O'Hara & Kotze 2010**

* Generalised linear models would be better at dealing with count data
* How do you deal with zeros
  + Normally done by adding an arbitrary integer (usually 1) to the data set to fudge it.
  + zero-inflated models are also a good way of handling lots of zeroes (Zuue, Ieno & Elphick 2010)
* in comparisons, transformations performed poorly, except when dispersion was small and the mean counts were large
  + Quassi-Poisson and negative binomial models consistently performed well, with little bias.
    - These distributions are often specified as an alternative to the traditional poisson distribution when dealing with count data as overdispersion is common, they add an additional parameter that allow the variance to vary independently from the mean.
    - Quassi-poisson assumes
    - negative binomial assumes:
    - and are both overdispersion parameters
    - Read Ver Hoef & Boveng (2007)
* Textbooks on statistical methodology in ecology recommend the use of the SQRT transformation

**Regression models for count data in R  
Zeileis *et al.* 2008**

* The classic poisson regression is not suitable for count data due to over-dispersion and lots of zeroes.
* glm() - the Generalized linear model function
* use the models hurdle() and zeroinfl() in the pscl package
* Although quassi-poisson, negative binomial, and other Generalised Linear models are fine for overdispersion, they don’t deal well with excess zeroes
* *zero-augmented* models capture the zero counts in a second model component
* *Hurdle* models combine a left-truncated count component with a right-censored hurdle component
* *zero inflated* models are mixture models that combine a count component and a point mass at zero.

**Analysing data with clumping at zero An example demonstration**

**Chang & Pocock 2000**

* Uses 2 approaches to analyse the relationship of multiple covariates to an outcome with lots of zeroes
  + Categorise the continuous outcome and fit a proportional odds model
  + Use a logistic regression to model the probability of a zero response and ordinary linear regression to model the non-zero continuous responses
* The data used here appears to be log-normal.