- This talks is about regression methods in which the dependent variable takes nonnegative integer or count values.
- The dependent variable is usually the number of times an event occurs.

Overview

Some examples of event counts are:

- number of claims per year on a particular car owners insurance policy,
- number of workdays missed due to sickness of a dependent in a one-year period,
- number of papers published per year by a researcher.

Poisson Distribution

- The number of persons killed by mule or horse kicks in the Prussian army per year.
- Ladislaus Bortkiewicz collected data from 20 volumes of Preussischen Statistik.
- These data were collected on 10 corps of the Prussian army in the late 1800s over the course of 20 years, giving a total of 200 observations of one corps for a one year period. The period or module of observation is thus one year.

Poisson Distribution: Prussian Cavalary

- The total deaths from horse kicks were 122, and the average number of deaths per year per corps was thus $122/200 = 0.61$.
- In any given year, we expect to observe, well, not exactly 0.61 deaths in one corps
- Here, then, is the classic Poisson situation: a rare event, whose average rate is small, with observations made over many small intervals of time.

```
rpois(200,lambda=0.61)
> X
[1]    1 2 0 1 0 3 0 0 1 0 0 4 0 0 0 1 0 1 0 2
[21]   0 0 0 2 2 0 0 0 1 0 0 0 0 1 0 0 0 1 2 0
[41]   0 0 1 0 1 0 1 0 0 1 1 0 1 0 0 1 0 0 3 1
.......
.......
[141] 0 0 0 0 1 2 0 1 0 1 0 0 0 0 0 0 0 1 0 0
[161] 1 0 1 0 0 0 0 1 0 0 0 0 0 1 1 1 0 2 0 1
[181] 0 0 2 0 2 0 0 1 0 0 3 1 0 0 0 1 1 0 0 0
>
> mean(X)
[1] 0.53
> var(X)
[1] 0.5317588
```

Overview

- Poisson regression is main technique used to model count variables.
- Poisson Distribution : Mean and Variance are equal
$$\mathrm{E}(X) = \mathrm{Var}(X)$$
- Sometimes conventional Poisson Regression is not an appropriate technique, and alternative or variant techniques are used instead.
- For example, Negative Binomial regression is for modelling count variables, usually for over-dispersed count outcome variables.

Generalized Linear Models

- In statistics, the problem of modelling count variables is an example of generalized linear modelling.
- Generalized linear models are fit using the `glm()` function.
- The form of the `glm` function is

```
glm(formula, family=familytype(link=linkfunction),
  data=dataname)
```

Generalized Linear Models

Family	Default Link Function
binomial	`(link = "logit")`
gaussian	`(link = "identity")`
Gamma	`(link = "inverse")`
inverse.gaussian	(link = "$1/mu^2$")
poisson	`(link = "log")`
quasibinomial	`(link = "logit")`
quasipoisson	`(link = "log")`

Texts on GLMs

- Dobson, A. J. (1990) An Introduction to Generalized Linear Models. (*London: Chapman and Hall.*)
- Hastie, T. J. and Pregibon, D. (1992) Generalized linear models. Chapter 6 of Statistical Models in S eds J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole.
- McCullagh P. and Nelder, J. A. (1989) Generalized Linear Models. (*London: Chapman and Hall.*)
- Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S. *New York: Springer.*

VGAM: Vector Generalized Linear and Additive Models

Author(s): Thomas W. Yee (t.yee@auckland.ac.nz)

License: GPL-2

URL: http://www.stat.auckland.ac.nz/~ yee/VGAM

Vector generalized linear and additive models, and associated models (Reduced-Rank VGLMs, Quadratic RR-VGLMs, Reduced-Rank VGAMs).

This package fits many models and distribution by maximum likelihood estimation (MLE) or penalized MLE. Also fits constrained ordination models in ecology.