Poisson Regression "Exposure" and offset

- Poisson regression may also be appropriate for rate data, where the rate is a count of events occurring to a particular unit of observation, divided by some measure of that unit's exposure.
- For example, biologists may count the number of tree species in a forest, and the rate would be the number of species per square kilometre.
- Demographers may model death rates in geographic areas as the count of deaths divided by personyears.
- More generally, event rates can be calculated as events per unit time, which allows the observation window to vary for each unit.

Poisson Regression : Exposure and Offset

In these examples, exposure is respectively unit area, personyears and unit time. In Poisson regression this is handled as an offset, where the exposure variable enters on the right-hand side of the equation, but with a parameter estimate (for log(exposure)) constrained to 1.

$$\log\left(\mathrm{E}(Y \mid x)\right) = \log\left(\text{exposure}\right) + \theta' x$$

which implies

$$\log\left(\mathrm{E}(Y \mid x)\right) - \log\left(\text{exposure}\right) = \log\left(\frac{\mathrm{E}(Y \mid x)}{\text{exposure}}\right) = \theta' x$$

Poisson Regression : Exposure and Offset

Offset in the case of a GLM in R can be achieved using the offset() function:

```
glm(y ~ offset(log(exposure)) + x, family=po
```