

Chapter 11: Introduction to Poisson regression

Contents

1	Introduction	1
2	The Poisson regression model	4
2.1	Poisson model for counts	4
2.2	Interpretation of parameter estimates	4
2.3	Parameter estimation	5
2.4	Illustration in R	5

1 Introduction

Explanatory variables	Response variable	Method
Continuous	Continuous	Regression
Categorical	Continuous	ANOVA
Continuous	Dummy (0 or 1)	Logistic regression
Continuous	Count	Poisson regression

Poisson regression is used when the **Response variable** is a count of something per unit or per time interval.

Examples:

- number of people in an organization
- number of visits to a physician
- number of arrests in the past year

Poisson regression model assumes that the response variable Y has a **Poisson distribution** (with parameter λ): $Y \sim \text{Poisson}(\lambda)$.

We saw in chapter 3: "We often model counts per unit or counts per interval by a Poisson distribution. Let Y be the random variable associated with such a count, and let λ be the appropriate expected rate of occurrences. The probability function for Y is

$$p(y) = \frac{\lambda^y}{y!} \exp(-\lambda) \quad \text{for } y = 0, 1, 2, \dots \text{ and } \lambda > 0.$$

For $Y \sim \text{Poisson}(\lambda)$, we can show that

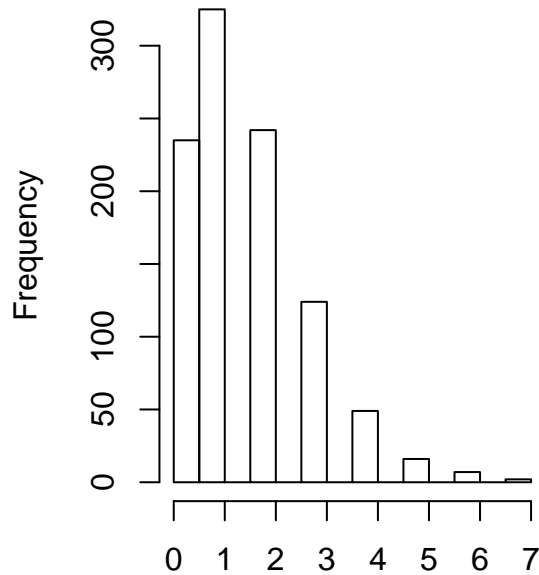
$$\mu = E(Y) = \lambda$$

$$\sigma^2(Y) = \lambda$$

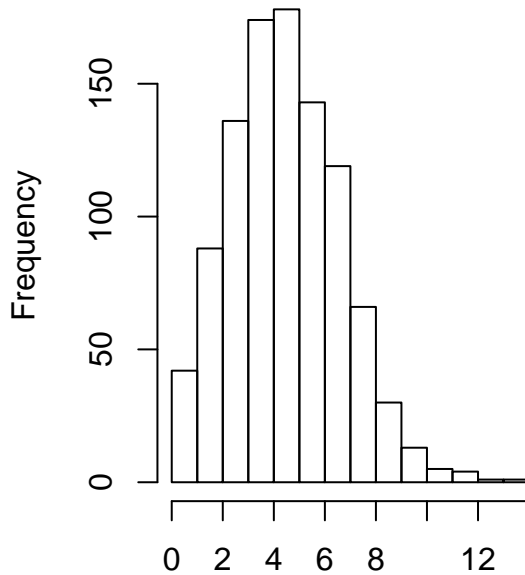
$$\sigma(Y) = \sqrt{\lambda}$$

```
lam_1.5 <- rpois(1000, 1.5)
lam_5 <- rpois(1000, 5)
par(mfrow = c(1,2))
hist(lam_1.5, breaks = 15, main = "Histogram of Poisson, lambda = 1.5", xlab="")
hist(lam_5, breaks = 15, main = "Histogram of Poisson, lambda = 5", xlab="")
```

Histogram of Poisson, lambda = 1



Histogram of Poisson, lambda =



Properties:

1. As λ increases, the mode moves away from 0.
2. $E(Y) = Var(Y) = \lambda$

In Poisson regression, one wants to express how the parameters λ depends on the explanatory variables.

Hence:

1. We write λ_i to allow the parameter to vary across individuals.
2. Because $\lambda > 0$, we usually express the $\log(\lambda)$ as a linear function of the explanatory variables:

$$\log(\lambda_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}$$

Example *horseshoe crabs*

We consider the data of the horseshoe crabs (see Agresti (1996)) with $n = 173$.

Import data set *crab.txt* as **crab** in R

```
crab <- read.table(file = file.choose(), header = TRUE)
names(crab)
head(crab, n = 6)
```

```
## [1] "obs" "C"   "S"   "Wt"  "W"   "Sa"

##   obs C S   Wt   W Sa
## 1   1 2 3 28.3 3.05 8
## 2   2 3 3 26.0 2.60 4
## 3   3 3 3 25.6 2.15 0
## 4   4 4 2 21.0 1.85 0
## 5   5 2 3 29.0 3.00 1
```

```
## 6 6 1 2 25.0 2.30 3
```



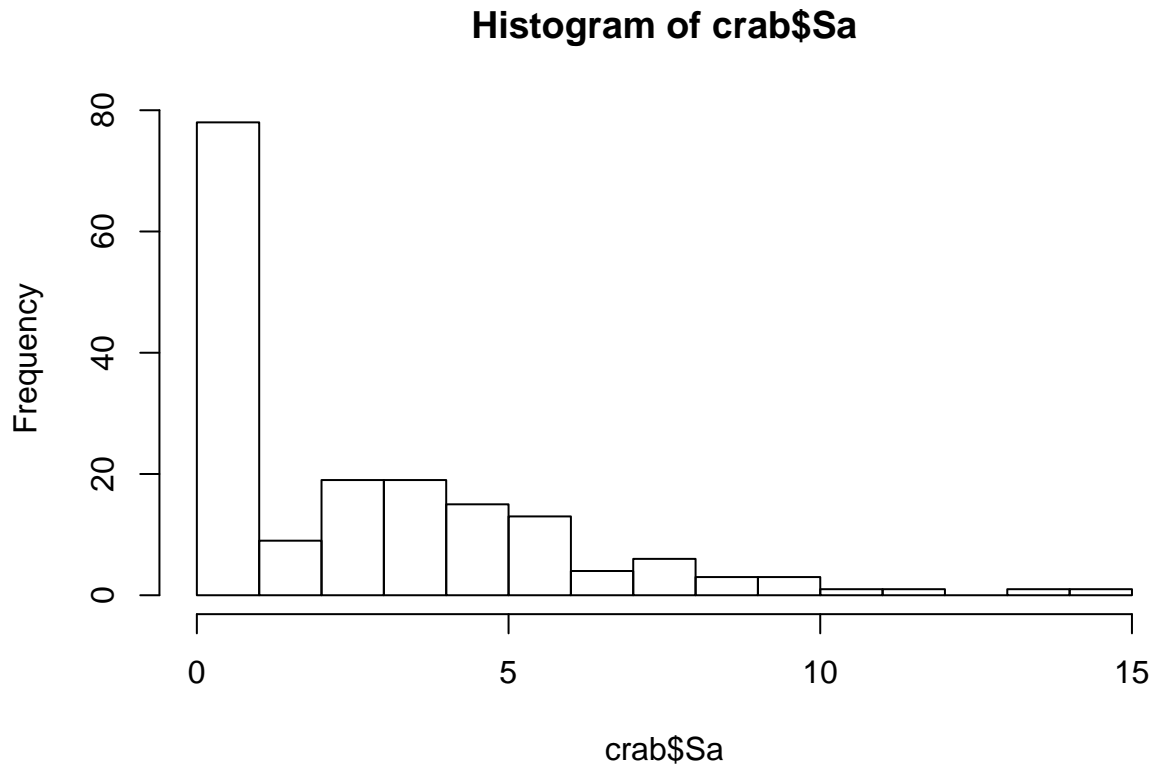
Each female horseshoe crab in the study had a male crab attached to her in her nest. The study investigates factors that affect whether the female crab had any other males, called satellites, residing near her.

The response outcome for each female crab is her number of satellites (Sa).

Explanatory variables that are thought to affect this are

- The female crab's color (C)
- Spine condition (S),
- Weight (Wt),
- Carapace width (W).

```
hist(crab$Sa, breaks = 15)
```



Descriptive statistics

```
list(mean = mean(crab$Sa), var = var(crab$Sa))
```

```
## $mean  
## [1] 2.919075  
##  
## $var  
## [1] 9.912018
```

2 The Poisson regression model

2.1 Poisson model for counts

In Poisson regression, we suppose that the expected count $E(Y) = \lambda$ can be determined by a set of explanatory variables:

$$\log(\lambda) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

Here, we will only illustrate the simple Poisson regression model with one explanatory variable, we write:

$$\log(\lambda) = \beta_0 + \beta_1 x_1.$$

This is equivalent to:

$$\lambda = \exp(\beta_0 + \beta_1 x) = \exp(\beta_0) \exp(\beta_1 x)$$

2.2 Interpretation of parameter estimates

- $\exp(\beta_0)$ = effect on the mean of Y , (that is λ), when $x = 0$.

- $\exp(\beta_1)$ = with every unit increase in x , the predictor variable has multiplicative effect of $\exp(\beta_1)$ on the mean of Y , (that is λ).

To see this, consider the following:

– assume $\lambda_1 = \exp(\beta_0) \exp(\beta_1 x)$

– assume $\lambda_2 = \exp(\beta_0) \exp(\beta_1(x + 1))$

→ Hence, $\lambda_2 = \exp(\beta_1) \lambda_1$

* If $\beta_1 = 0$, then $\exp(\beta_1) = 1$, the expected count $\lambda = E(Y) = \exp(\beta_0)$, and Y and x are not related.

* If $\beta_1 > 0$, then $\exp(\beta_1) > 1$, and $\lambda_2 > \lambda_1$.

* If $\beta_1 < 0$, then $\exp(\beta_1) < 1$, and $\lambda_2 < \lambda_1$.

2.3 Parameter estimation

Similar to the case of logistic regression, the **maximum likelihood estimators (MLEs)** for β_0, β_1, \dots etc. are obtained by finding the values that maximizes log-likelihood.

In general, there are no closed-form solutions, so the MLEs are obtained by using iterative algorithms such as *Newton-Raphson* (NR), *Iteratively reweighted least squares* (IRWLS), etc.

2.4 Illustration in R

Example *horseshoe crabs*

- Model 0: intercept only

Intercept-only model: $\log(\lambda) = \beta_0$

Fitting the intercept-only model. This model implies the expected number of satellites per each crab is the same.

```
model.0 <- glm(Sa ~ 1, family = poisson(link = log), data = crab)
summary(model.0)
```

```
##
## Call:
## glm(formula = Sa ~ 1, family = poisson(link = log), data = crab)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4162  -2.4162  -0.5707   1.1045   4.9942
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.0713     0.0445   24.07  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 632.79  on 172  degrees of freedom
## Residual deviance: 632.79  on 172  degrees of freedom
## AIC: 990.09
##
## Number of Fisher Scoring iterations: 5
```

In this case:

$$E(Sa) = \exp(1.0713) = 2.919$$

Then $\log(\lambda) = 1.0713$

or $\lambda = 2.919$.

- Model 1: single explanatory variable

Poisson regression of number of satellites (Sa) on Width (W)

```
model.1 <- glm(Sa ~ W, family = poisson(link = log), data = crab)
summary(model.1)
```

```
##
## Call:
## glm(formula = Sa ~ W, family = poisson(link = log), data = crab)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.9306  -1.9981  -0.5627   0.9299   4.9992
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.4282     0.1789  -2.394  0.0167 *
## W              0.5892     0.0650   9.065  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 632.79  on 172  degrees of freedom
## Residual deviance: 560.84  on 171  degrees of freedom
## AIC: 920.14
##
## Number of Fisher Scoring iterations: 5
```

The estimated model is:

$$\log(\lambda_i) = -0.43 + 0.59 \cdot W_i$$

or

$$\lambda_i = E(Sa_i) = \exp(-0.43) \cdot \exp(0.59 \cdot W_i)$$

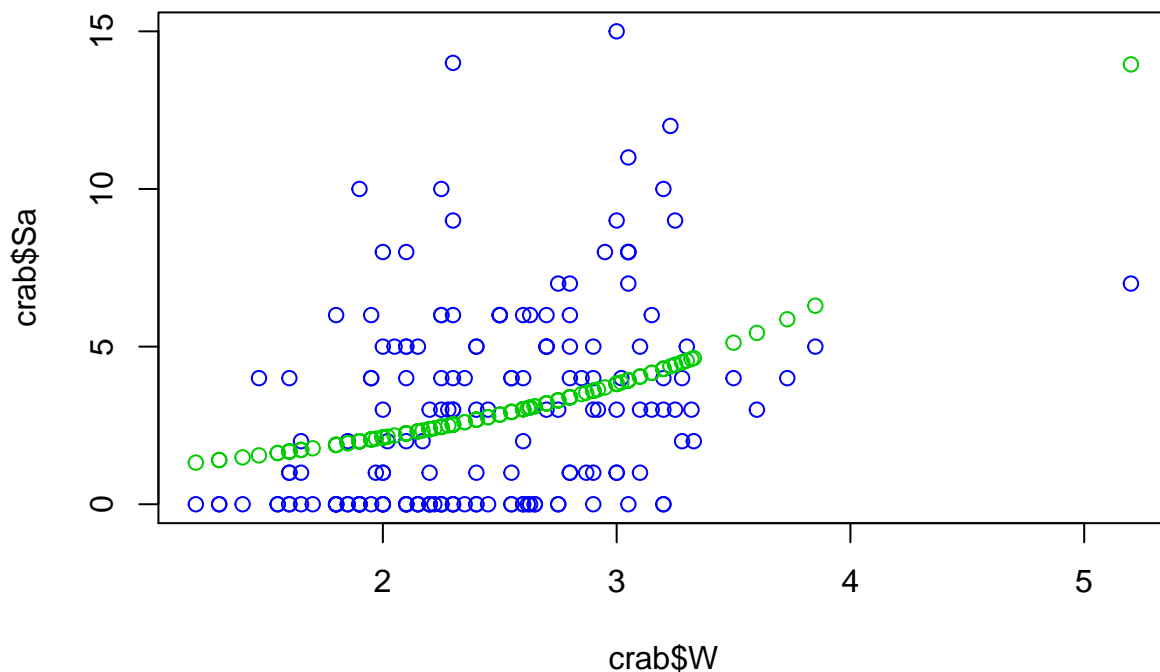
The slope is statistically significant.

Interpretation:

Since the estimate of $\beta_1 > 0$, the wider the female crab, the larger the expected number of male satellites on the multiplicative order as $\exp(0.59) = 1.8$. More specifically, for one unit of increase in the width, the expected number of satellites (Sa) will increase and it will be multiplied by 1.8.

Visualize the Poisson regression model.

```
pred <- fitted(model.1)
plot(crab$W, crab$Sa, col = 4) # Scatter plot of Sa vs W
points(crab$W, pred, col = 3)
```



Is there a good fit?

The deviance is used as overall-goodness-of-fit statistic. This statistic has $n - p$ degrees of freedom (with n the number of observations and p the number of parameters in the model).

Example *horseshoe crabs*

```
anova(model.1)
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: Sa
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev
## NULL                172      632.79
## W      1    71.949      171    560.84
```

In our example of the crab data, we have a residual deviance of 560 with 171 degrees of freedom (df). (with $n = 173$ and $p = 2$).

To compare model 1 with model 0 (the intercept-only model), we can also use the chi-square test for comparing the deviances.

```
anova(model.0, model.1, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: Sa ~ 1
## Model 2: Sa ~ W
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1         172      632.79
## 2         171      560.84   1    71.949 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 1 is significant better than the intercept-only model (= model 0).

Remark:

1. The assumption says that, for a given set of explanatory values, the variance of the response is equal to its mean. Often, the variance is higher than that.
2. Note that the *Deviance* has an approximate χ^2 -distribution with $n - p$ degrees of freedom, where n is the number of observations ($n = 173$ in the crab example) and p is the number of parameters ($p = 2$ in the crab example).
3. The expected value of a χ^2 random variable is equal to the degrees of freedom.

Hence, if our model fits the data well, the ratio of the *Deviance* to *DF* (the degrees of freedom), $\frac{Deviance}{DF}$, should be about one.

The ratio $\frac{deviance}{DF} = \frac{560}{171} = 3.2$ which is much larger than 1.

So if the residual deviance is larger than the residual degrees of freedom, this is an indication of overdispersion. Overdispersion means that observed variance is larger than the assumed variance. Sometimes this can be solved by including extra explanatory variables. Other solutions are running a negative binomial regression model or use of an adjustment for overdispersion.

If you have overdispersion (residual deviance is much larger than degrees of freedom), you may want to use `quasipoisson()` instead of `poisson()`. For a **quasi-Poisson** regression, the variance is assumed to be a linear function of the mean. Sometimes one also use the negative binomial model in those cases.

4. In case your data set has many zero counts, then one might consider a **zero-inflated Poisson regression model**. Then we assume there is one process which determines whether a female crab has satellites or not. There is another process which determines how many satellites a female crab has.