Discrete Responses and Maximum Likelihood

author: Maria Paniw date: 10.02.2016 width: 1620 height: 1080

Classes of experimental and sampling design

Dependent variable (response Y)	Independent variable (predictor X)	
	Continuous	**Categorical**
Continuous	Regression	ANOVA
Categorical	Generalized linear models	Contingency tables

Having covered regression and ANOVA, we are now in the domain of discrete response variables. What is a discrete response?

> A discrete variable consists of countable values. Abundance and occurrence are good examples of discrete variables.

A discrete variable is obviously not normally distributed, so how do we get parameter estimates to describe the discrete Y as a function of X?

Binomial distribution - logistic regression

When we have a binomial distribution of our response variable, Y can only take two values:

- success/presence: 1
- failure/absence: 0

The probability of success is denoted as p, i.e., *P(Y = 1) = p*

Because we only have two outcomes, the probability of failure is *P(Y = 0) = 1 - p*

Of course, when we collect observations, we don't typically collect 1s and 0s (maybe only in presence - absence studies). Usually, we have 2 columns: one filled with p successes, the other with n total counts.

Examples include p = 15 germinated seeds out of n = 20 totally sown seeds. Or, p = 4 detected fish out of n = 30 total fish in population.

The binomial distribution is then described as $Y \sim \mathcal{B}\rangle\backslash(n, p)$

Linear regression?

One of many reasons why linear regression doesn't work with binomial data is that the residuals will definitely not be normally distributed.

But we still want to explore relationship:

$$Y = \beta_0 + \beta_1 x$$

In the case of a binomial response, Y can be expressed as $E(Y|x)$, or what is the probability of getting Y given my predictor x.

Now, let $\pi(x)$ be the conditional probability of getting Y given x.

Specifically, we get from a 0/1 response to a probability by this formula:

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

From probability to a generalized linear model

A key transformation of $\pi(x)$ is the **logit** transformation because it linearizes the predictor term $\beta_0 + \beta_1 x$.

It is defined as:

$$g(x) = \ln\left\{\frac{\pi(x)}{1-\pi(x)}\right\}$$

but $1-\pi(x) = 1 - \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} = \frac{1}{1+e^{\beta_0+\beta_1 x}}$

Hence,

$$g(x) = \ln\left\{\frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} \Big/ \frac{1}{1+e^{\beta_0+\beta_1 x}}\right\} = \ln\left\{e^{\beta_0+\beta_1 x}\right\} = \beta_0 + \beta_1 x$$

The back-transformation of $g(x)$, to the probability once you have the linear predictor, is:

$$p = \frac{e^{g(x)}}{1+e^{g(x)}}$$

What is maximum likelihood?

In linear regression, we estimated our parameters, β_0 and β_1 by minimizing the residual sum of squares $RSS = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$

This option is not available to us when we model discrete responses. But there is a general method for parameter estimation in Frequentist statistics that actually leads to the least-squares function under the linear regression - **MAXIMUM LIKELIHOOD (MLE)**!

> MLE provides estimates of parameters, given a statistical model, that make the observed data the most likely. In other words, MLE estimates parameters that *maximize* the probability of obtaining the observed data. Now, because the probability of obtaining the observed data is given by the *likelihood function*, MLE maximizes the likelihood function.

Let's look at a specific example.

Binomial likelihood function

y	x
0	12
1	23
0	15
0	4
1	18

What the likelihood function?

we know:

$$p = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

$$1-p = \frac{1}{1+e^{\beta_0+\beta_1 x}}$$

Therefore,

$$l(x|\beta) = \frac{1}{1+e^{\beta_0+\beta_1(12)}} * \frac{e^{\beta_0+\beta_1(23)}}{1+e^{\beta_0+\beta_1(23)}} * \frac{1}{1+e^{\beta_0+\beta_1(15)}} * \frac{1}{1+e^{\beta_0+\beta_1(14)}} * \frac{e^{\beta_0+\beta_1(18)}}{1+e^{\beta_0+\beta_1(18)}}$$

R and other statistical packages use numerical algorithms to estimate β_0 and β_1 that maximize l.

Log-likelihood

It it mathematically easier to estimate the log-likehood, i.e., $ln(l(x|\beta))$. This is because the logarithm removes the exponents and converts multiplications to sums. If you want more information on the theoretical part of likelihood functions, please consult: http://www.unc.edu/courses/2006spring/ecol/145/001/index.html