Journal of Neuroscience Methods 274 (2016) 131–140

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Original research paper

A comparison of methods for the analysis of binomial clustered outcomes in behavioral research

Alberto Ferrari*, Mario Comelli

University of Pavia, Department of Behavioral and Brain Sciences, Medical Statistics Section, Italy

HIGHLIGHTS

- We performed a comparison of statistical methods for the analysis of clustered binary outcomes in behavioral research with small sample sizes.
- Beta-binomial regression performed accurate and powerful hypothesis testing, outperforming even Generalized Linear Mixed Models in a range of scenarios.
- A misspecified linear model, in some circumstances, can represent a reasonable compromise between technical approachability and accuracy when dealing with proportion data.
- Poisson regression should not be applied straight away to modeling of proportion data.

ARTICLE INFO

Article history:
Received 6 July 2016
Received in revised form 5 October 2016
Accepted 7 October 2016
Available online 14 October 2016

Keywords:
Proportions
Generalized Linear Models
Generalized Linear Mixed Models
Clustered data
Beta-binomial regression
Behavioral science

ABSTRACT

Background: In behavioral research, data consisting of a per-subject proportion of "successes" and "failures" over a finite number of trials often arise. This clustered binary data are usually non-normally distributed, which can distort inference if the usual general linear model is applied and sample size is small. A number of more advanced methods is available, but they are often technically challenging and a comparative assessment of their performances in behavioral setups has not been performed.
Method: We studied the performances of some methods applicable to the analysis of proportions; namely linear regression, Poisson regression, beta-binomial regression and Generalized Linear Mixed Models (GLMMs). We report on a simulation study evaluating power and Type I error rate of these models in hypothetical scenarios met by behavioral researchers; plus, we describe results from the application of these methods on data from real experiments.
Results: Our results show that, while GLMMs are powerful instruments for the analysis of clustered binary outcomes, beta-binomial regression can outperform them in a range of scenarios. Linear regression gave results consistent with the nominal level of significance, but was overall less powerful. Poisson regression, instead, mostly led to anticonservative inference.
Comparison with existing methods: GLMMs and beta-binomial regression are generally more powerful than linear regression; yet linear regression is robust to model misspecification in some conditions, whereas Poisson regression suffers heavily from violations of the assumptions when used to model proportion data.
Conclusions: We conclude providing directions to behavioral scientists dealing with clustered binary data and small sample sizes.

1. Introduction

Most physiological parameters studied by biomedical researchers are continuous variables whose distribution approximates well normality; some examples of this are weight, height, blood pressure, hormone levels. For this reason, parametric methods assuming normal distribution of the response variable are the most widely used statistical instruments in biomedicine. Data showing strong departure from normality, on the other hand, are usually dealt with by transforming them to achieve better Gaussian approximation, or resorting to the use of nonparametric methods.

* Corresponding author.
E-mail addresses: alberto.ferrari04@universitadipavia.it, aferrari34@yahoo.com (A. Ferrari).

http://dx.doi.org/10.1016/j.jneumeth.2016.10.005
0165-0270/

Nonparametric tests, though, suffer from decreased power and difficulty in dealing with interaction effects; these limitations suggest the use of more powerful instruments when they are available. Furthermore, in some fields of research it is not uncommon to see variables arise whose behavior, while not approximating normality, is well described by other known probability distributions.

In behavioral sciences these non-Gaussian behaviors arise naturally quite often, due to the peculiar nature of the measured responses. One such example is the outcomes of decision making tasks, in which the subject has to choose among two or more different behaviors, with one response being considered a "success" and the other(s) a "failure"; e.g. the Iowa Gambling Task (Stockard et al., 2007) or the Game Dice Task (Brand et al., 2005), used in the study of pathological gambling. In these cases, the outcome of interest is the ratio of correct choices on the number of trials.

The distributions of proportions usually do not approximate normality; they are generally asymmetrical and only admit a range of values from 0 to 1. One way to approach them is to apply the arcsine square root transformation to the data, but this has been shown to provide only minor or no improvements in power over the analysis of untransformed data, thus its use is not advised (Jaeger, 2008). Since the outcomes of behavioral experiments such as the decision making tasks we mentioned are a series of Bernoulli trials, logistic regression can be proposed as a more formal solution to their analysis. Despite this, the very common problem of overdispersion, i.e. the excess variance not accounted for by the model (usually due to overlooked sources of variation, such as inter-individual variability), can make the choice of the appropriate analytical instrument and experimental design very challenging. Simpler analytic approaches such as classical linear regression have the advantage of being very easy to apply, but at the same time must rely on their robustness to violations of some of the model's assumptions. More refined instruments are available, but they are often technically challenging even to statisticians, and their misuse can have harmful consequences. Therefore, in everyday practice, researchers must often face a tradeoff between correct model specification and technical approachability.

The aim of this paper is to review and compare some of the most relevant methods available for the analysis of proportions in the usual behavioral setup, in order to evaluate in a simulation study the robustness of hypothesis testing to model misspecification, and to provide guidelines to the reader for the choice of adequate analytic instruments and sample sizes.

2. Dealing with proportions

2.1. Linear regression and binomial model

Let us consider a behavioral experiment in which N subjects are exposed to a predetermined number n of trials, and in each trial they are required to choose between two different possible responses. Some examples of this kind of experiments are questionnaires; escape tests, in which the subject has to choose the appropriate response to avoid an aversive stimulus; or risky decision making tasks used to assess the preference of the subject for "safe" versus "risky" rewards, such as the already mentioned Iowa Gambling Task. We may want to assess whether an experimental variable, such as a genetic trait or a drug treatment, has a significant effect on the propensity of the subject towards one of the two choices.

As we noted before, the most widespread method used to deal with such results in behavioral science is linear approximation: subject becomes the statistical unit and the number of successes, or the ratio of successes to failures, is the measured outcome. As long as subject is the only grouping factor in experimental design, observations are independent and clustering is basically removed from the picture (a more refined approach consists in regressing single outcomes on the covariates in a mixed model with subject as random effect, but this is of use mainly when within-subjects fixed effects are of interest (Stockard et al., 2007)).

This approach is not a priori inacceptable, and has a very clear advantage in practical terms, i.e. it is very easy to apply and to interpret; yet, when clustered binary data are collapsed into counts or ratios of successes, violations of the assumptions of linearity and homoscedasticity are expected and their effects on inference should be evaluated.

A more formal approach, that takes into greater account the nature of the data generating process, consists in considering each of the $n \times N$ trials as a Bernoulli process with two possible outcomes, "success" and "failure", with probability of success $\boldsymbol{\pi}$ and probability of failure $1 - \pi$; in this case the number of correct responses y is a random variable with a binomial probability distribution of parameter π (Jaeger, 2008).

This approach to the data requires us to perform the analysis using Generalized Linear Models (GLMs), that allow us to model relations between the covariates and the response variable when the latter's distribution is described by a noted non-Gaussian probability function.

2.2. Poisson regression

One possible alternative to linear regression that takes more into account the data generating process is Poisson regression. Indeed, the so called "law of rare events" states that, when n is large compared to π, i.e. successes are "rare", the binomial distribution approximates the Poisson distribution.

Poisson regression is the optimal solution to deal with count data that can be interpreted as the outcome of a binomial process with an infinite number of trials and a finite number of successes; e.g., when considering the number of occurrences of a certain event in a given amount of time (Cameron and Trivedi, 2013).

In our hypothetical experiment, the single subject is the statistical unit, and the raw number of correct choices it makes is the response to be analyzed through Poisson regression. In this case, the model will be expressed in the form:

$$\log(\mathrm{E}(y|x)) = \beta_0 + \boldsymbol{\beta}'\mathbf{x} \quad (1)$$

Where y is the count outcome vector, β_0 is the intercept and $\boldsymbol{\beta}'$ the vector of fixed effect coefficients. This method is very easy to apply in most statistical software and in particular R; plus, it can be used also to model situations in which n varies from cluster to cluster in the N clusters by including an $offset = \log(n_i)$ into the model. Nevertheless, the efficiency of Poisson regression in a context in which n is limited is hindered by the upper bound on the number of possible correct responses, since Poisson distribution allows for all integer values in the range going from 0 to $+\infty$. Therefore, the Poisson model is also misspecified for proportion data. We can expect Poisson approximation to work well only when we have large n and comparatively low π; in fact it has been shown that it can be a powerful alternative to linear regression even in experimental conditions where an upper bound is present (Lazic, 2015), and it has been applied to the study of complex decision making (Giang and Donmez, 2015; Paserman, 2016), gambling (James et al., 2016) and perseverative behavior (Lazic, 2015).

Another issue the experimenter might meet when applying Poisson regression is the inflation of Type I error rate in presence of overdispersion; indeed, inference in Poisson regression is heavily dependent on the assumption of equality of mean and variance. However, this problem can be fixed by applying robust sandwich