

Homework

Lluís Ramon

20 de abril de 2015

1 Introduction

Cattle are an important economic resource and also a major health factor in many countries. Therefore preventing diseases in a herd is vital.

In this study we focus our interest on the parasite of trypanosomosis which can lead to the death of a cow. This disease, transmitted by the tsetse flies, causes an infection characterized by fever, loss of appetite and anemia, which can lead to death depending on different factors.

One medicine, the Berenil, is used to cure the infected cattle. The aim of this research is to determine the efficiency of different doses of Berenil. Finding the most efficient dose, if such a dose exists, is critical when it comes to save both cattle and money. Here lies all the interest of our study.

To determine if a cow is ill or not a binary variable will be studied. This indicator can be 1 for a healthy cow or 0 for a sick one.

2 Objective

The aim of this study is to assess the efficacy of different doses of Berenil in cattle infected with the trypanosomosis parasite.

3 Dataset

A cohort of 10 different cows infected by trypanosomosis parasite was selected for the study. Each Berenil dose (low, medium and high) was administrated three times (time 1, 2 or 3) for each animal. PCV was reported each time as well as the number of calves it had before being infected.

The variables reported for this study are presented below.

- Id: Each cow has its own id. From 1 to 10. (*id*)
- PCV: Binary variable. (*pcv.b*)
- Dose: H High, M Medium L low. (*dose*)
- Time: From 1 to 3. (*time*)
- Number of birth: From 2 to 8. (*nbirth*)

Since the gathering process of the data was unknown, several assumptions were needed.

- Each time PCV is obtained before the treatment. Therefore the effect of the third dose could not be evaluated.
- Dose is assigned randomly in time to the cow.
- For a given cow, the previous treatments (high, medium or low) do not affect the following ones.
- Time intervals are the same and fixed.

4 Outcome categorization

TODO(Mathieu): Improve text

The binary response was related with a cow being healthy or unhealthy. Taking this into consideration, the outcome was categorized using the following criteria:

- [Literature review](#)*: A healthy cow is estimated to have a PCV value ranging from 24 to 46.
- Practical Modeling: As the binary response should be modeled in following sections, a suitable one was searched. To this end, a trial and error with a cutoff ranging from 20 to 24 was explored.

The threshold between healthy and unhealthy cow was set at a PCV value of 20. If the PCV value was bigger it was categorized as healthy, if it was lower or equal it was categorized as unhealthy.

Table 1 shows a contingency table for the dichotomised response variable. Healthy category increases in time while unhealthy category decreases. A Missing value category for healthy/unhealthy cow i is also included. It can be seen that the number of missing values increases with time. This topic will be studied in further detail in section 7 Missing Data Analysis.

	Time 1	Time 2	Time 3
Unhealthy	28	15	2
Healthy	1	12	18
Missing value	1	3	10

Table 1: Contingence table for the dichotomised response variable (rows) and times.

5 Statistical methods

The methods used in the statistical analyses are detailed in this section. First a Generalized Estimating Equations (GEE) and latter a Generalized Linear Mixed Model. An exploratory data analysis was performed before those regression methods.

5.1 GEE

The response variable was the binary PCV to detect healthy and unhealthy cows. As we wanted to estimate the effect of dose in healthy/unhealthy cows, these two variables were included in the model.

From this initial model, a forward step-wise method was carried for the model selection, including additional covariates or interactions between them. The models were compared quasilikelihood ratio test when nested and with QuasiLikelihood Information Criteria (Pan 2001) when non-nested.

A classification table with predicted values and original data was created to asses how the model performed.

5.2 GLMM

TODO(Mathieu)

IN order to take into account the specific effect due to the cow itself we had to introduce random effects in our model. To do that we performed a Generalised Linear Mixed Model. We kept the fixed effects found in the previous part and add random effects on the intercept and the time. Since we have two way of grouping our data we have a random effect for the id and among the values of a given cow we have random effects for the kind of dose that was used.

5.3 Transition Models

TODO(Lluis): Add model information. TODO(All): Include pros/cons with our data. Decide if our data could be modeled using this kinds of models.

6 Results

6.1 GEE

TODO(Lluis)

As explained in the section 5.1, a starting model for PCV was fitted with dose.

$$\text{logit}(p) = \beta_{\text{dose}} + \beta_{\text{time}}$$

Table X shows estimates for the selected model.

Table X shows classification table for the choosed model.

6.2 Random effects model

TODO(Mathieu)

7 Missing Data Analysis

TODO(Gerard): Missing exploration, in processs

In this section, the patterns of missing data will be discussed and analyzed, in order to assess its influence in the analyses performed in previous sections.

Figure 1 and Table 2 show the missing data distribution over time and for each dose. Clearly, the fact that an observation is missing is associated with the dose and the time (fisher tests p-values= 0.005 and 0.007, respectively). Actually, there are more missings in low and medium dose than in high dose, and also, there are more missings as time increases.

Dose	Time	Missings	Not.Missing
L	1	0	10
L	2	2	8
L	3	6	4
M	1	1	9
M	2	1	9
M	3	4	6
H	1	0	10
H	2	0	10
H	3	0	10

Table 2: Number of missings in the outcome.

8 Limitations and Further research

- Jackknife estimators for small sample better than sandwich estimator

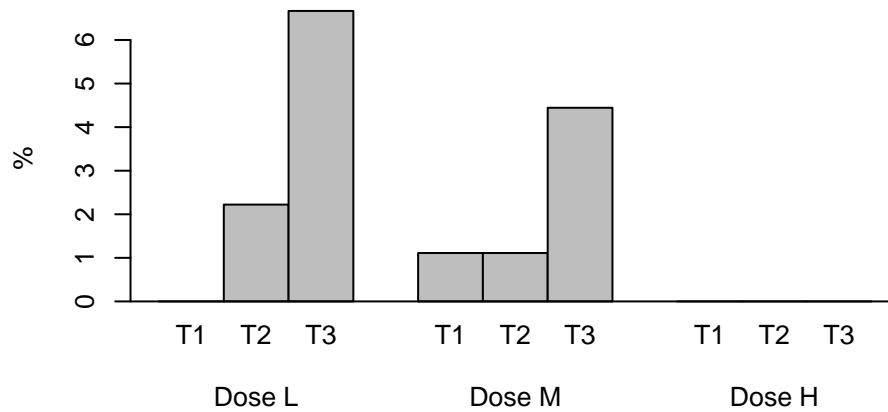


Figure 1: Percentage of missings for each time and dose with respect to the whole sample.

Jackknife variance estimators are preferable to the sandwich estimator in case of a small number of clusters. (Højsgaard, Halekoh, Yan 2006)

- Compare a new factor L and M dose against H dose
- Perfect Separation Problem with gee

It appeared when using a 22 PCV value threshold for healthy or unhealthy cow.

- Compare GEE model with GLMM. Coefficients and SE.

9 Bibliography

TODO(Gerard): Normalize citations

- Pan 2001, Biometrics, Akaike's Information Criterion in Generalized Estimating Equations.
- Højsgaard, S., Halekoh, U. & Yan J. (2006) The R Package geepack for Generalized Estimating Equations Journal of Statistical Software, 15, 2, pp1–11
- Hardin & Hilbe 2002, Generalized Estimating Equations
- Kamil Barton (2015). MuMIn: Multi-Model Inference. R package version 1.13.4. <http://CRAN.R-project.org/package=MuMIn>