Modelling the ethanol-induced sleeping time in mice through a zero inflated model

Njinju Tongwa Fogap

promotor :
Prof. dr. Roel BRAEKERS

Eindverhandeling voorgedragen tot het bekomen van de graad
Master of science in Applied Statistics

Universiteit Maastricht

universiteit hasselt

Universiteit Hasselt

Center for Statistics

Modeling the ethanol-induced sleeping time in mice through a zero inflated model

Presented by NJINJU TONGWA FOGAP

Supervisor: Roel Braekers

Thesis submitted in partial fulfillment of the degree of masters of Science in Applied statistics.

CERTIFICATION

This project entitled ‘Modeling the ethanol-induced sleeping time in mice through a mixture-model’ presented by Njinju Tongwa Fogap has been read and approved as fulfilling part of the requirement for the award of a master degree in applied statistics; Hasselt University and therefore approved for its contribution to scientific knowledge and literature presentation.

_____ Date: _____

Professor Reol Breakers

(Supervisor)

DEDICATION

This piece of work is dedicated to the almighty God for being my source of inspiration and divine guidance all through my life. I shall also, in a special way dedicate this piece of work to my parents Mr. and Mrs. Nkemnjinju Peter Fogap for their unflinching moral and financial support.

ACKNOWLEDGEMENT

Special thanks to my supervisor Prof Roel Braekers who continually guided me through this work, chastising me when necessary and supplying thoughtful suggestions for improving the text which is deemed necessary for publication to be used by Statisticians and Health experts.

TABLE OF CONTENT

LISTS OF TABLES

LISTS OF FIGURES

CONTENT PAGE

LISTS OF ABBREVIATION

ZIP- zero-inflated poisson

SL1- Sleeping time of mice for the first trial

SL2- Sleeping time of mice for the second trial

SL – Average sleeping time

COA- Mice coat colour

WG1- Weight of mice prior to the first trial

Tr1 – Trial one

Mat- mating pair

CRO-Crosses

LDL- Lower Detection Limit

Abstract

In the analysis of data in statistics, it is imperative to select most suitable models. Wrong choice of model selection leads to bias parameter estimates and standard errors. In the ethanol anesthesia data set used in this thesis, we observe more than expected zero counts, usually termed zero-inflation. Traditional application of Poisson and negative binomial distributions for model fitting may not be adequate due to the presence of excess zeros. This zero-inflation comes from two sources; a proportion of mice with an actual zero minute sleeping time and another proportion of mice that had their observations left censored with a fixed detection limit of one minute.

The purpose of this thesis is to illustrate how zero-inflated poisson model is used to model the ethanol anesthesia data set. Such a model accounts for the excess zero in the data. With the zero-inflated model we could calculate the probability of non-sleeping mice in the population and further observe the distribution of sleeping mice.

After fitting the ZIP model and using maximum log likelihood estimating functions, we found out that the probability of observing a non-sleeping mouse is 1.03%, which is extremely low. We could further calculate the probabilities of observing non-sleeping mice under specific covariates. For example we could calculate the probability of observing a female albino non-sleeping mice.

We further checked for the significant effects of our covariates on our response variable. Based on our ZIP model, environmental factors like gender, albinism, birthday, and trial day had a significant influence on the sleeping time after the first trial. The variable, trial day 1 however had a boarder line significant influence (p – value =0.0462) on the sleeping time. The only genetic factor with a significant effect was the chromosome variable.

Chapter 1

1.0 INTRODUCTION

Alcoholic beverages date back to the very early part of man's history. Many archaeologists believe that wines made from grapes have existed for more than 10,000 years and that drinks such as mead and beer have existed for even longer. Throughout its history, alcohol has been used socially for many diverse purposes, such as calming feuds, giving courage in battle, sealing pacts, celebrating festivals, and seducing lovers. Historians speculate that prehistoric nomads may have made beer from grain & water before learning to make bread. The Celts, Ancient Greeks, the Norse, Egyptians, and Babylonians all have records of production and consumption of alcoholic drinks. Alcohol was included in the Egyptian burial provisions for the journey to the afterlife.

Alcohol, specifically ethanol, is a potent central nervous system depressant, with a range of side effects. The amount and circumstances of consumption play a large part in determining the extent of intoxication; for example environmental factors such as socio-economic status, profession and circumstances like the amount of food consumed prior to alcohol consumption should be taking into account.
Alcohol has a biphasic effect on the body, which is to say that its effects change over time. Initially, alcohol generally produces feelings of relaxation and cheerfulness, but further consumption can lead to blurred vision and coordination problems like vomiting, dehydration, hangover etc

Individual reactions to alcohol vary, and are influenced by many factors, including but not limited to age, gender, race and ethnicity, amount of food consumed etc

1.1 Ethanol-induced anesthesia data

The ethanol-induced anesthesia data set was created between 1991 and 1993, and contains repeated measurements on ethanol-induced sleeping time for four different populations of mice, divided over three generations. The first generation consists of two inbred populations, ILS and ISS, which were created by sib mating and selection for length of sleeping time from standard laboratory mice (± 30 generations). In the ILS strain, the mice have a long sleeping time of more than 3 h, while in the ISS strain the mice have a short sleeping time of less than 10 min. Due to the inbreeding, we assume that the alleles which code for a long (resp. short) sleeping time are only found in the ILS (resp. ISS) strain and that these strains are homozygous for these alleles. These strains were crossed, in the way that ILS females were mated with ISS males and vice versa, to produce both reciprocal F1 strains, L/S and S/L respectively. Each animal in the F1 population received one gamete from each of its parents and will therefore be heterozygous for the genes determining the sleeping time. The F1 strains were afterwards crossed in all combinations to produce four F2 strains L/S×L/S, L/S×S/L, S/L×L/S and S/L×S/L. Due to this construction of the F2 population; we know that it consists of mice with three different genotypes. Some mice are homozygous in the short or long sleep alleles while other animals are heterozygous. Therefore we will call the F2 population a segregating population. This in contrast with the ILS, ISS and F1 populations in which all the animals have the same genotype. We call these populations isogenic populations. In the ethanol-induced anesthesia data set we cannot divide the F2 population according to the genotypes since this information is not available. We will assume that the Mendelian proportions hold for this population. This means that a quarter of the F2 animals are homozygous in the short sleep alleles, another quarter is homozygous in the long sleep alleles and the

remaining half is heterozygous. All mice were produced in a specific pathogen-free (SPF) colony and were weaned at 25 days of age. At approximately 50 days they were removed to a non-SPF environment where they remained for 1 week before testing. The mice were maintained on a 12 h light cycle (07.00 to 19.00) and were allowed food (Wayne diet by Teklad) and water ad lib. In this thesis, we look at the sleeping time of the first trial (SL1), in the F2 population. The objective of this project is to model a zero inflated model in the ethanol anesthesia data set. This model shall be used to determine the probability of non sleepers on one hand and to determine the distribution of sleepers on the other hand. The design of this study is a repeated measurements setting. The mice were initially injected at 55-65 days of age (Trial 1) and again 7-10 days later (Trial 2). In the afternoon prior to the day of testing, the animals were weighted and moved into the testing room, where they remained until completion of a trial. The mice were injected intraperitoneally (into the abdomen) with a 4.1 g/kg dose of ethanol (20% w/v in saline) between 09.00 and 13.00. The order of individual injections was retained from the first trial to the next. After an injection the animal was placed on its back in a plexiglass trough and was considered to be anesthetized if it did not turn over more than three times within the first minute. Repeated attempts to observe this behavior were made within 15 min after the injection. Anesthetic recovery was indicated when an individual turned over three times within 1 min after being anesthetized. The sleeping time of a mouse was measured as the time interval between observed anesthesia and the final minute of recovery. We note that in the assessment of sleeping time, the recording is left-censored by a fixed detection limit of 1 min. For a mouse in which repeated attempts to place it on it's back failed, we know that the animal probably did not fall asleep nor had a sleeping time of less than 1 min. In the following analyses, we consider these mice as left-censored observations. Furthermore we note that all the mice were allowed to sleep until recovery, so we

do not have any right- or interval-censored observations in this data set. Our population of interest in this thesis shall be the F2 population (the isogenic population) and we shall concentrate on the first trial (tr1) and our response variable shall be the first sleeping time (sl1).

Next to the sleeping time, we recorded in each animal several covariates. In this thesis we will focus on only a few of those, like sex, coat color (coa), cross (44 = L/S x L/S F2, 45 = L/S x S/L F2, 53 = S/L x L/S F2, 54 = S/L x S/L F2), mating pair (mat), weight for trial 1 (WG1), birthday (bir) and trial day for trial 1(tr1). The variable coat color of a mouse is dichotomized in the analyses and will be used to study whether an albino mouse reacts differently to alcohol than a non-albino mouse. Markel and Corley (1994) found that the gene coding for albinism (Tyr) had an effect on the ethanol-induced sleeping time. They posed that either this gene or a gene closely linked to it, is important for sleeping time.

1.2 Motivation

In this thesis, we investigate the influence (effect) of an overdose of alcohol on the ethanol induced sleeping time in mice in the F2 Population. However due to the construction of the different mice families, we observed that some of the mice did not fall asleep for a given dose. This may be due to the fact that the method to assess whether a mouse was asleep or not did not work well in this area of observation and led to a detection limit.

The ethanol-induced anesthesia data set has already been analysed by Markel and Corley (1994), markel et al (1995a and 1995b), and Roel Braekers (2004).

In the first two articles written by Markel and Corley (1994), Markel et al. (1995a and 1995b), the censoring aspect in this data set was neglected and the censored

observations were treated as non-sleepers, that is, these animals received a zero score for their sleeping time and was after wards analysed as uncensored observations. Here, the main objective was on the determination of environmental factors for the sleeping time in the F2 population.

In the third articles, the information of all four different populations ILS, ISS, F1 and F2 was used to find estimate for the heritability of ethanol-induced sleeping time and for the number of genes which determine this trait. The Cohen corrected estimate for mean and variance was used to handle the censoring of the data. This method had its draw backs. However, solutions to these drawbacks were given in the thesis written by Roel Braekers (2004).

In this project, we study a zero-inflated model for the ethanol-induced sleeping time for trial 1 (SL1). Such a model shall allow us to estimate on one hand the probability for a non-sleeping mouse and on the other hand find the distribution for the mice that sleep in the F2 population. The first main problem that we need to handle is that we cannot exactly observe the sleeping times due to a detection limit. We observe for some mice an upper bound on the sleeping time. In this setting, we are interested in building a model that solves the above problem and also find out how this model change under covariates.

Chapter 2

2.0 METHODOLGY

2.1 Ethanol-Induced Anesthesia data and left censored data

In the ethanol-induced anesthesia data the observations were left censored. When the data from the sleeping time are left censored, the lower detection limit is known and used to substitute a value for the censored observation.

In this situation, there might be far more observations lying below the lower detection limit (LDL) than would be expected according to any of the standard parametric distributions. Thus in our data, we might observe Zero-inflation of our zero counts due to the left censoring of our data. In this situation, a proportion of the population of mice which actually did not sleep would receive a zero count, we shall call this the 'certain zero' proportion, while another portion of mice would be given a zero count due to the left censoring of the data. This left censoring of the data was due to the fact that the method to access whether a mouse was asleep or not, did not work well in this area of observation and introduced a detection limit. Below this detection limit, we are unable to distinguish between non-sleeping mouse and a very short sleeping mouse. To be more explicit, this technique establishes a detection limit, below which an exact measurement of the sl1 can not be reported, although the LDL will be a function of the technique that is employed.

2.2 Zero Inflated Models

In this chapter, we study a zero inflated model which allows us to estimate the probability of non-sleeping mouse and also to find the distribution for the mice that sleep. This method of approach is suitable because there exist several false zeros in the data set. With false zeroes, not only can the proportion of sites truly occupied be underestimated, but the relationship with explanatory variables (the predictive model, usually the aim of the exercise) may also be biased. Furthermore, the chances of a false zero may not be constant but rather a function of time or sample site characteristics, making comparisons among samples questionable.

The first concept of a zero-inflated distribution originated from the work of Rider (1961) and Cohen (1963), who examined the characteristics of mixed Poisson distributions. Mixed Poisson distributions are characterized by data that have been mixed with two Poisson distributions in the proportions α and 1- α respectively.

The probability density function (pdf) of such a mixed distribution is

$$P(n) = \alpha \frac{\lambda_1^n e^{-\lambda_1}}{n!} + (1-\alpha) \frac{\lambda_2^n e^{-\lambda_2}}{n!}$$

Where $\lambda_2 > \lambda_1$ (the means of the two distributions) and n is the observed count data (0, 1, 2... N). Both Rider (1961) and Cohen (1963) have proposed different approaches using the method of moments for estimating the parameter α. Cohen further described an approach for estimating the parameter α with zero sample frequency. Johnson and Kotz (1968) were the first to explicitly define a modified Poisson distribution (known as Poisson with added zeroes) that explicitly accounted for excess zeroes in the data. The modified distribution is the following:

$$P(n) = \alpha + (1-\alpha)e^{-\lambda}; \qquad n = 0$$
$$P(n) = (1-\alpha)\frac{e^{-\lambda}\lambda^{n}}{n!}; \qquad n \geq 1$$

Johnson and Kotz (1968) proposed a similar procedure to the one suggested by Cohen (1963) for estimating the parameter α. Based on the work of Yoneda (1962), they also developed a general modified Poisson distribution that accounts for any kind of excess in the frequency of the data. Under this distribution, n = 0, 1, 2… K are inflated counts while the rest of the distribution

K +1, K +2…N follows a Poisson process.

The concept of the mixed Poisson distribution introduced by the previous authors has been particularly useful to describe data characterized with a preponderance of zeros. For this type of data, more zeros are observed than what would have been predicted by a normal Poisson or Poisson-Gamma process. It is generally believed that data with excess zeros come from two sources or two distinct distributions, hence the apply-named dual state process. The underlying assumption for this system is that the excess zeros solely explain the heterogeneity found in the data (if we make abstraction of the ZINB) and each observation has the same mean λ. Two different types of regression or predictive

Models have been proposed in the literature for handling this type of data. The first type is known as the hurdle model (Cragg, 1971; Mulhully, 1986). This type of model has not been extensively applied in the statistical field and is therefore not the focus of this paper. The reader is referred to Cragg (1971), Mulhully (1986) and Schmidt and Witte (1989) for additional information about hurdle models. The zero-inflated count models (also called zero-altered probability or count models with added zeroes) represent an alternative way to handle data with a preponderance of zeros. Since their formal introduction by Lambert (1992) (who

expanded the work of Johnson and Kotz (1968)), the use of these models has grown almost boundlessly and can be found in numerous fields, including traffic safety. For example, these models have been applied in the fields of manufacturing (Lambert, 1992; Li et al., 1999), economics (Green, 1996), epidemiology (Heilbron, 1994), sociology (Land et al., 1996), trip distribution (Terza and Wilson, 1990) and political science (Zorn, 1996) among others.

In essence, zero-inflated regression models are characterized by a dual-state process, where the observed count can either be located in a perfect state or in an imperfect state with a mean μ.

This type of model is suitable for data generated from two fundamentally different states.

In situations where a recorded absence may in fact represent a failure to detect what is actually there (a false zero) erroneous inferences may result with naïve application of procedures such as logistic regression (MacKenzie and Kendall 2002; Tyre et al. 2003; Wintle et al. 2004).

2.3 ZIP FOR THE ETHANOL INDUCED ANAESTHESIA DATA SET

$Y_{/y>0}$ = Sleep ~ F (x), x > 0

And

$Y_{/y=0} = 0 \sim \pi$ is the distribution of non sleeping mice.

The indicator function of sleeping mice is given by

$Z = \max (Y, 1), \delta = I (Y \geq 1), (Z, \delta)$

To calculate the probability of observing non-sleeping mice based on probability theorem, we have

$P (Y \leq y) = P (Y \leq y / Y=0) P(Y=0) + P (Y \leq y \setminus Y > 0) p (Y > 0)$

$P (Y \leq y) = \pi + (1- \pi) f(y)$

$$P (Z \leq z) = \begin{cases} 0, z < 1 \\ \pi + (1 - \pi) f (z), z \geq 1 \end{cases}$$

Taking sub distributions of the above distribution, we then have**.**

$$H^{U}(z) = p (z \leq z, \delta = 1) = \begin{cases} 0, z < 1 \\ (1 - \pi) f (z) - f(1), z \geq \mathbf{1} \end{cases}$$

$$H^{0}(z) = p(z \leq z, \delta = 0) = \begin{cases} \pi + (1 - \pi) f(1), & z \geq 1 \\ 0, & z < 1 \end{cases}$$

Combining the above two probabilities leads to the likelihood for a single function as indicated below:

$$\mathbf{L = d H^{u}(z_i)^{\sigma i} d H^{0}(z_i)^{1-\sigma i}}$$

$$\mathbf{(1- \pi) f(z)^{\sigma i} [\pi + (1- \pi) f(1)]^{1-\sigma i}}$$

$$\mathbf{L = \prod_{i=1}^{n} [(1 - \pi) f(z)]^{\sigma i} [\pi + (1- \pi) f(1)]^{1-\sigma i}}$$

Where δ_i is an indicator variable, $\delta_i = 0$ when $Y_i \leq 1$ and $\delta_i = 1$ when $Y > 1$

The linear predictor for the inflation probability is given by the following expression:

Linpinfl = ao + a1*sex + a2*albino+a3*bir + a4*tr1 +a5*wg1 + a6*cro

The inflation probability of zero (non sleepers) shall be transformed using the logarithm transformation. The reason for the choice of the exponential function rather than the normal function is that the normal function takes on all real numbers on the number line, while our distribution of the sleeping time of trial 1 indicates only positive numbers. It is also worth noting that the formal and informal test for normality indicated that sleeping time 1 was not normal, hence assuming a normal distribution for our response variable, SL1 is inappropriate.

The inflation probability is then expressed as:

Infprob = 1/ [1 + exp (- (ao + a1*sex + a2*albino+a3*bir + a4*tr1 +a5*wg1 + a6*cro))]

The poisson mean can be expressed as:

λ = exp (b0 + b1*sex + b2*albino + b3*bir + b4*tr1 + b5*wg1 + b6*cro)

Building the ZIP log likelihood in SAS is as follows:

If sl1 $\leq$ 1 or sleep = 0, then

L = log [infprob + (1 – infprob) * exp (- lamda)]

With this log likelihood estimator, we could find the probability of non-sleeping mice in the population and observe how the change with certain covariates.

Chapter 3

3. O Exploratory Data Analysis

In this chapter we perform a first exploration of the ethanol-induced anaesthesia data set to gain more insight and to verify the distribution of the sleeping time of the first trial (SL1).The first sleeping time shall be considered as our response variable. The average sleeping time (SL) is not suitable as our response variable; SL is the average sleeping time of both trials, and since the second sleeping time (SL2) was left censored, calculating the average sleeping time, we could not get a true indication of the real average value.
We shall plot a histogram to see the frequency distribution of the sleeping time of the first trial (SL1).

We shall then observe plots illustrating how some of our covariates change with our response (SL1).This shall give us an idea on what we expect to see when fitting our zero inflated model. Our covariates of interest were selected based on existing literature review and previous studies carried out on this same data set. Based on this therefore, sex, albino, trial 1 (tr1), mat (mating pair), weight for trial 1 (Wg1), and birthday were choosing as our covariates of interest.
Figure 1.0 below shows the histogram and density function of the dependent variable, Sl1

Fig 1.0

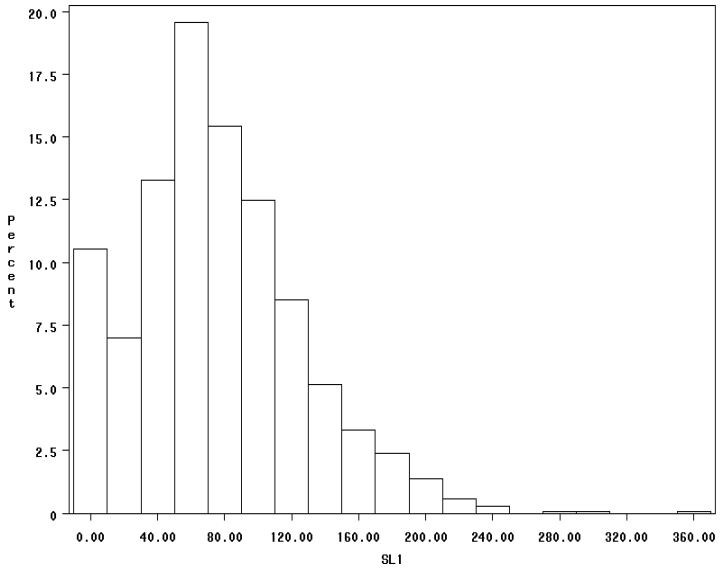


Fig 1.0: Histogram and density estimate for the sleeping time in trial 1 for the F2 population in the ethanol induced data set.

From the above figure we see the distribution of the dependent variable, sl1. The distribution is not normal. We also notice inflation in the zero counts in the sleeping time of the first trial for the F2 population. If you observe the distribution above, we notice that the zero counts are higher than expected. This zero-inflation is contributed by two different factors; those mice that did not actually sleep or had a true sleeping time of zero and those mice in which repeated attempts to place it on its back failed, and the animal was considered by the experimental as those mice that probably did not fall asleep or had a sleeping time of less than 1 min, this might be due to the wrong method to access the sleeping time by the experimental. Based on this zero-inflation therefore, we shall use the zero-inflated poisson model to model our dependent variable. This is a statistical method that is most suitable to model count data with many zeros. With this zero-inflated poisson model, we can determine the probability of non sleeping mice and also look at the distribution of the mice that sleep in the F2 population. From the above figure it seems likely that

the distribution of the mice that sleep follow a weibull distribution. Appendix I in the appendix also shows the inflation of the zero count and the distribution of the dependent variable.

Fig 2.1

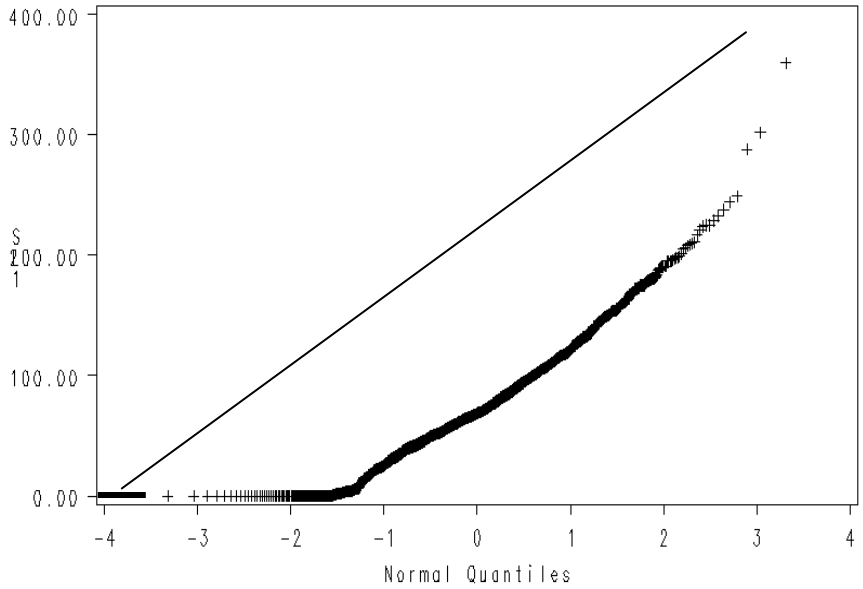


Fig 2.0 Quantile-quantile plot of the dependent variable, sl1. The sl1 observations for the anaesthesia data were censored at 1.0 minute; the shift away from the line of equality represents heaviness in the left tail of the observed distribution, due to a larger than expected number of censored observations.

Appendix II in the appendix also illustrates the occurrence of the dependent variable with sex variable. There exist an almost equal distribution of the sl1 observations between the male and females, thought the observations are slightly higher in males than females. The sex variable was categorized, with 1 for males and 2 for females. This graphs also suggest that majority of the mice were sleeping with majority of the sl1 observations being greater than 1 minute. There exist very

few observations below the LDL of 1 minute. This indicates that the probability of occurrence non-sleeping males and females should be very small.

Appendix III in the appendix also indicates occurrence of the dependent variable, sl1 with the coat color of the mice. In this experiment, we were interested in the albino mice which has an indicative number of 10 i.e Albino is given by (I=10).A majority of the albino mice slept after being injected with ethanol. The probability of non-sleeping albino mice is expected to be very low.

Chapter 4

4.0 RESULTS

In this chapter we present the results of the statistical analysis on the ethanol-induced anesthesia data set. By using the zero inflated Poisson model introduced in chapter two to describe the sleeping time of trial one, we calculate the probability of non sleepers and also describe the distribution of sleepers in the F2 population.

We shall go further to illustrate with some examples on how to calculate our probability of non-sleepers under specific covariates.

Table 1.0

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
b0	-3.002	0.04818	1384	89.24	<.0001	0.05	4.2056	-3.1947	0.0910
b1	-0.1255	0.008070	1384	-15.55	<.0001	0.05	-0.1413	-0.1097	0.1277
b2	0.0147	0.000901	1384	16.33	<.0001	0.05	0.01295	0.01649	0.7343
b3	-0.0001	0.000013	1384	-13.11	<.0001	0.05	-0.00019	-0.00014	173.80
b4	0.000023	0.000012	1384	2.00	0.0462	0.05	3.98E-7	0.000046	195.58
b5	-0.0001	0.001317	1384	-0.05	0.9625	0.05	-0.00264	0.002521	1.937993
b6	0.0093	0.000447	1384	20.79	<.0001	0.05	0.008408	0.01016	4.5052
a0	0.0367	0.02415	1384	-1.52	0.1286	0.05	-0.08411	0.01066	0.1826
a1	-0.2223	0.1792	1384	-1.24	0.2151	0.05	-0.5739	0.1293	-0.0309
a2	0.0403	0.03205	1384	1.26	0.2086	0.05	-0.02255	0.1032	-0.0093
a3	-0.0003	0.000511	1384	-0.65	0.5136	0.05	-0.00134	0.000669	-1.4136
a4	-0.0008	0.000476	1384	-1.73	0.0834	0.05	-0.00176	0.000109	-1.9416
a5	0.0051	0.02080	1384	0.24	0.8078	0.05	-0.03574	0.04586	-0.0129
a6	-0.0118	0.009563	1384	-1.24	0.2165	0.05	-0.03058	0.006936	-0.0286

Table 1.0 Results of parameter estimates for the ZIP model

From the results obtained in table 1.0, the parameter estimates can be used to calculate the probability function of non sleepers. We shall illustrate this with some examples. There was convergence problem observed when interaction effects were

considered in the fitting of our model, thus all interaction effects were dropped. The parameter sex was seen to have a significant effect on the sleeping time of mice in the F2 population after being induced with ethanol. This implies male and females responded differently when induced with ethanol with regards to their sleeping time after the first trial.

The coat color of the mice also has a significant effect on the sleeping time of the first trial. Thus the findings of Markel and Corley (1994) that gene coding for albinism (Tyr) had an influenced on the ethanol-induced sleeping time, was confirmed by this result. The variable Cro has a significant effect on the sleeping time of mice after the first trial. This can be attributed to the fact that the alleles that code for short sleeping time (ISS) and those for long sleeping time (ILS) were crossed and off springs in the F2 population inherited some of these alleles. In effect parental inheritance had a significant effect on the sleeping time of mice after being induced with ethanol. The trial days had a boarder line significant effect with a p-value of 0.0462. We cannot exactly say that the trial day has an effect on the sleeping time of the ethanol induced mice.

The linear predictor for the inflation probability is given by the following expression:

Linpinfl = 0.0367+ -0.2223*sex + 0.0403*albino+ -0.0003*bir + -0.0008*tr1 + -0.0051*wg1 + -0.0118*cro

The inflation probability is then expressed as:

Infprob = 1/ [1 + exp (-0.0367+ -0.2223*sex + 0.0403*albino+ -0.0003*bir + -0.0008*tr1 + -0.0051*wg1 + -0.0118*cro)]

The poisson mean can be expressed as:

λ = exp (-**4.30**+ -0.1255*sex + 0.0147*albino + -0.0001*bir + 0.000023*tr1 + -0.0001*wg1 + 0.0093*cro)

The log likelihood estimator for the probability of non sleepers can be expressed as:

L = log [infprob + (1 – infprob) * exp (- lamda)], where infprob and lamda (λ) are given by the above expressions.

The log likelihood estimator can then be used to calculate the probability of non-sleeping mice in the population and also to observe how this probability changes under certain covariates.

For example the probability of non sleeping mice can be calculated as follows:

L = log [infprob + (1 – infprob) * exp (- lamda)], where infprob and lamda (λ) are defined as follows:

Infprob = 1/ [1 + exp (-0.0367)

Infprob = 0.509

Lamda (λ) = $\exp^{-3.002,}$ λ = 0.0497

L = log [0.509 + (0.49 * 1.059)], L = log 1.023

L = 0.01

Thus the probability of non-sleeping mice in the F2 population is 0.01. This value is very small and goes in confirmity with what was observed in the

exploratory phase of our data analysis. We could use the parameter estimates in table 1.0 to calculate the probability of non-sleeping mice under specific covariates. For example, it is easy to calculating the probability of getting non-sleeping female albino mice.

Chapter 5

5.0 Conclusion

As with any analysis, it is important to choose appropriate statistical models. In this thesis, we analyze the ethanol data set given in Markel and Corley (1994) and in Roel Braekers (2004). This data consist of four different mouse populations, but we shall concentrate on the F2 population. The response of interest in the sleeping time of the first trail (sl1), and our covariates are sex, coat color (coa), cross (44 = L/S x L/S F2, 45 = L/S x S/L F2, 53 = S/L x L/S F2, 54 = S/L x S/L F2), mating pair (mat), weight for trial 1 (WG1), birthday (bir) and trial day for trial 1(tr1).
In the exploratory phase of our data analysis, we observe from the distribution of the mice population that there is the occurrence of excess of zeros (zero inflation), thus we used the zero inflated poisson model for the data set. Such a, model accounts for the zero excesses, allow us to calculate the probability of non-sleeping mice on one hand and also to observe the distribution of sleeping mice on the other hand. Some of the observations in this data set were censored to the left by a detection limit of 1 min. Thus mice with very short sleeping time below 1 minute were considered as non sleeping mice. This implies a proportion of the excess zeros is contributed by the left censoring while another proportion is contributed by those mice who actually did not sleep (‘exact zero’) proportion. We build our ZIP model and use maximum Likelihood estimates to calculate the probability of non sleeping mice.

In the exploratory data analysis, we observed from the QQ plot that there is heaviness in the left tail of the observed distribution, due to a larger than expected number of censored observations. We also observed that the probability of the

occurrence of non-sleeping mice was very low compared to mice that actually fell asleep. After calculating the probability of non-sleeping mice, based on our ZIP model, we realized that 1.03 %of the mice population were non-sleepers. Based on our ZIP model, environmental factors like gender, albinism, birthday, and trial day had a significant influence on the sleeping after the first trial. The variable, trial day 1 however had a boarder line significant influence (p –value =0.0462) on the sleeping time. These results are similar to those obtained by Roel Braekers (2004) for the imputation approach.

Genetic factors like Chromosomes (Cro), was also found to have a significant influence on the sleeping time of the mice in the F2 population. This can be explained by the fact that alleles coding for both long and short sleeping times were transferred to the offspring of the F2 generation and thus had an influence on their sleeping time.

Bibliography

Roel Braekers, (2003) modelling ethanol-induced sleeping time in inbred strains of mice: *A repeated measures design with left-censored data*
The history of alcohol (2005) Drug Rehab.org (http://www.drug-rehabs.org/)

Dominique Lord, Simon P. Washington, John N. Ivan. (2004) Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory 83rd Annual Meeting of *Transportation Research Board*

Greene, W. (1994). Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. *Working Paper EC-94-10, Department of Economics,* New York University.

Martin Ridout, Clarice G.B, John Hinde. (1998). Models for count data with many zeros. *International Biometric Conference*. Cape town

P D Markel, B Bennett, M Beeson, L Gordon and T E Johnson . (2007). Confirmation of quantitative trait loci for ethanol sensitivity in long-sleep and short-sleep mice. *Genome Res*. 1997 7: 92-99

P.D. Markel and R. P. Corley. (1994). A multivariate analysis of repeated measures: linkage of the albinism gene (Tyr) to a QTL influencing ethanol-induced anaesthesia in laboratory mice, *Psychiatric Genetics, 4, 205-210*

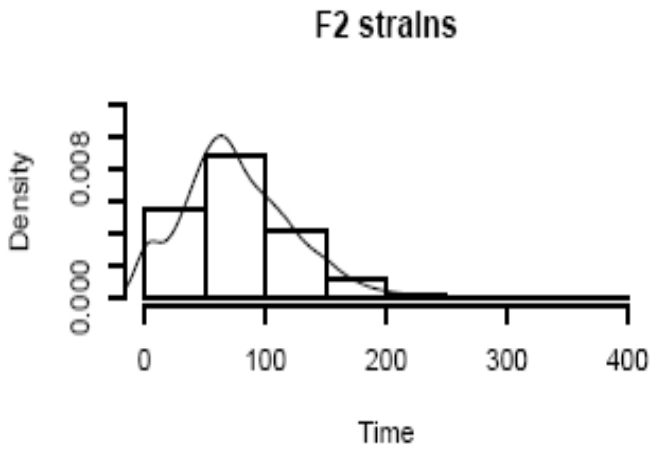
P. D. Markel, J. C. DeFries and T. E. Johnson, (1995a) Use of repeated measures in an analysis of ethanol-induced loss of righting reflex in inbred long-sleep and short-sleep mice, alcoholism: *clinical Experimental Research, 19, 299-304.*

P. D. Markel, J. C. DeFries and T. E. Johnson, (1995b) Ethanol-induced anaesthesia in inbred strains of long-sleep and short-sleep mice: A genetic analysis of repeated measures using censored data, *Behavior Genetics, 25, 67-73.*

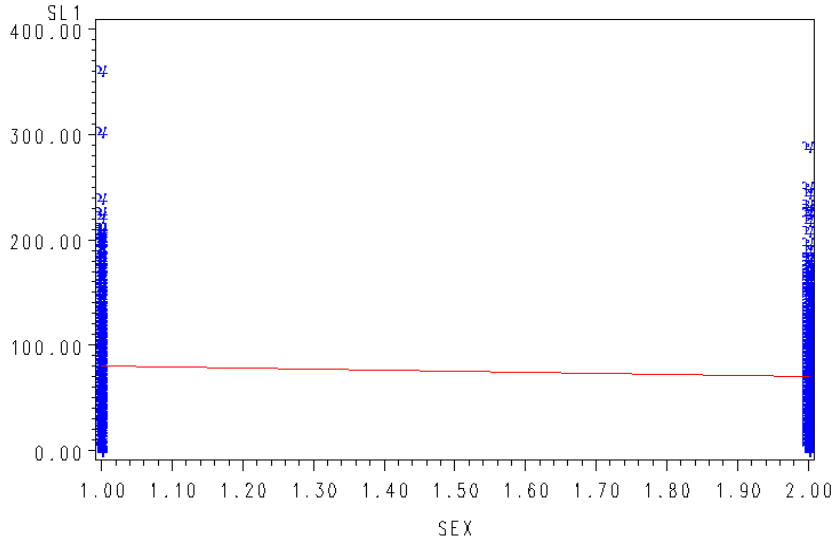
Dominique L, Simon P, John N. Ivan, (2004) Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory, *Center for Transportation Safety, Texas Transportation Institute, Texas A and M University System, 3135 TAMU, College Station, TX 77843-3135, USA*

Appendix

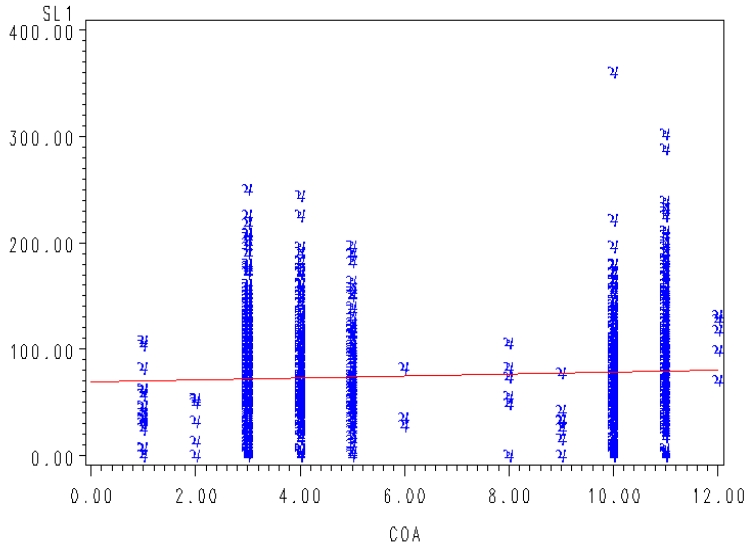
Appendix 1



Appendix II



Appendix III



APPENDIX IV

Tests for Normality

Test	Statistic		p Value	
Shapiro-Wilk	**W**	0.959161	**Pr < W**	<0.0001
Kolmogorov-Smirnov	**D**	0.064441	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	1.291072	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	8.971708	**Pr > A-Sq**	<0.0050

Auteursrechterlijke overeenkomst

Opdat de Universiteit Hasselt uw eindverhandeling wereldwijd kan reproduceren, vertalen en distribueren is uw akkoord voor deze overeenkomst noodzakelijk. Gelieve de tijd te nemen om deze overeenkomst door te nemen, de gevraagde informatie in te vullen (en de overeenkomst te ondertekenen en af te geven).

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Modelling the ethanol-induced sleeping time in mice through a zero inflated model
Richting: **Master of science in Applied Statistics** Jaar: **2007**
in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Ik ga akkoord,

Njinju Tongwa Fogap

Datum: **27.08.2007**