

CHAPTER SIX

CONCLUSION

6.1 Summary

This dissertation introduces the use of a mixture approach to analyze zero heavy toxicity data. In this dissertation, the focus is on a mechanism which creates distinct subpopulation of zeroes, and hence produces a high proportion of zeroes in the data set. There are cases where these zeroes would mix with zeroes produced by other mechanisms. For example, in chronic toxicity tests, the zeroes are created by both mortality and non-reproductive animals. In order to incorporate the mortality kind of extra zeroes, the idea of a mixture model approach is applied.

Since different data are created based on different mechanisms, it is important for researchers to understand the underlying process before starting to model the data. In Chapter Three, a mixture model is proposed for chronic toxicity testing in which the interest is fecundity. Since the responses are egg counts, a mixture model that combines a Poisson distribution with a zero mass is used to model the responses. This model, namely, the zero inflated Poisson model (ZIP) is compared with the regular Poisson model. A likelihood ratio test shows that the ZIP model is better than the regular Poisson model in terms of the *Ceriodaphnia dubia* test. A method of potency estimation is introduced in Chapter Three to compare the effect of different effluent toxicants. This inhibition concentration estimate (IC) is based on the mixture model approach that combines the reproduction and mortality effect.

The ZIP model is further extended to a multivariate case to analyze longitudinal toxicity data where each subject is measured repeatedly in Chapter Four. A

GEE-ZIP model is developed based on the idea of generalized estimating equations (GEE). The advantage of this model is the ability to analyze longitudinal data with covariates that change over time. It also has the flexibility to include a correlation structure among the repeated measures within each subject to give a more appropriate estimate of the regression coefficients than the regular regression model. Toxicity test has been the focus of the discussion to demonstrate the application of the mixture approach since the motivation for this dissertation comes from problems of chronic toxicity testing. However, one should remember that the use of the mixture approach is also applicable in many other situations. Examples of utilizing this approach for economic and manufacturing data are discussed in Chapter One.

The problem of zero-heavy data also exists in environmental studies. This often gives rise to multivariate data because multiple variables are measured at the same time. The study of the information in the correlation structure is of major interest in these studies because the inter-relationships between different variables are imbedded in the correlation structure. This dissertation further investigates the effect of zeroes on the correlation structure of multivariate data by carrying out simulation studies of principal component analysis. Under the assumption of the mixture approach, an observation can be considered as a product of two variables, and the correlation structure of the non-zero data can be recovered by using an EM algorithm routine.

6.2 Further Research

The ZIP and GEE models proposed in this dissertation are based on the assumption of independence of the two underlying mechanisms. For example, in the toxicity study, the biological mechanism that causes death is assumed to be independent of the mechanism that affects reproductivity. Relaxing this assumption is beyond the scope of the statistical theory. Nevertheless, models that relate the two mechanisms can be used.

The performance of the models can also be improved by applying a better experimental design. Since the model involves Poisson and the logistic regressions, further research may concentrate on developing designs that can accommodate the two parts together, or examine which part is of more importance.

More research can be done in the modeling and analysis of toxicity longitudinal data. The mixed effect model is another type of model that can be applied to toxicity longitudinal data. It can be considered as a generalized estimating equation model which allows correlation between the subjects. In contrast to the population-averaged model, it is sometimes called the subject specific model. It would be interesting to examine how to apply the use of the mixture approach in the mixed effect model.

The principal component analysis of zero-heavy data is still in its infancy, the work presented in this dissertation is rather exploratory. Clearly, additional research can be done in the future to develop the theoretical background using the mixture model idea.