ORIGINAL ARTICLE

Arvand Journal of Health & Medical Sciences

The effect of choosing different tuning constant in Huber function on robust parameter estimates

Farzaneh Boroumand[1], Alireza Akbarzade Baghban[2], Farid Zayeri[3], Hedyeh Faghir Ghanesefat[4]

[1]*School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[2]*Proteomics Research Center, Department of Basic Sciences, School of Rehabilitation Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.*

[3]*Proteomics Research Center, Department of Biostatistics, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[4] *Department of Pharmacology, Faculty of Medicine, Tehran University of Medical Sciences, Tehran, Iran*

✉ **Reprint or Correspondence**: Alireza Akbarzade Baghban; PhD Proteomics Research Center, Department of Basic Sciences, School of Rehabilitation Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
@✉ akbarzade@sbmu.ac.ir.

ABSTRACT

The presence of outliers appears as an unavoidable main problem in experimental studies. Outliers can greatly distort parameter estimates and subsequent standard errors. Consequently, inferences about the parameters are misleading. In this case, applying outlier robust statistical procedures should be considered. Robust estimation often relies on a dispersion function that is more slowly varying at large values than the square function. However, the choice of tuning constant in dispersion functions may impact the estimation efficiency to a great extent. The data used in this paper is part of a concentration–response study that shows the contraction of corpus cavernosum induced by phenylephrine in the organ bath. Because of the existence of an outlier to achieve robust estimations, M-estimation method and Huber function as a dispersion function are used. Three different measures for tuning constant were considered. (0.1, 1.5, 4). Based on the negative log likelihood, robust model had the best fit when c=1.5. In addition, because of the presence outlier, the Population Average (PA) in common model considerably underestimates the mean response in the upper asymptote. As result, using Huber function when c=1.5 in the robust model to apply the data was led to these results, cumulative administration of phenylephrine (0.1µM - 300µM) caused concentration-dependent contractions in strips of rat corpus cavernosum (-Log EC50 was 5 ± 0.31, 95% CI= 5.92 to 4.21). To estimate parameters of the model because of existence of an outlier in dataset, M-estimation method and Huber function as a dispersion function has been applied. The appropriate choice of tuning constant can be led to accurate results.

Keywords: Outlier, Concentration–response, M-estimation, Huber function.

***Received*: 16 November 2015 *Accepted*: 9 January 2016**

Introduction

In experimental studies, the existence of outliers appears as an unavoidable main problem. Thus, in order to reduce the disadvantages of outliers on statistical analysis, outliers must be

Please cite as: **Boroumand F, Akbarzade Baghban A, Zayeri F, Faghir Ghanesefat H. The effect of choosing different tuning constant in Huber function on robust parameter estimates. Arvand J Health Med Sci 2016;1(1):50-55.**

identified and treated (1). Outlier diagnosis methods are classified in two main categories, first of which is titled as intuitive method, where outliers are diagnosed using graphs. In the second category, analyzing the residuals can be led for diagnosing outliers. The later approach is entitled as inferential method (2). Outliers can greatly distort parameter estimates and subsequent standard errors. Consequently, inferences about the parameters are misleading. In this case, applying outlier robust statistical procedures should be considered. In these methods, the influence of outliers in estimating the model parameters is adjusted (3). As a result, the parameters estimates are more accurate. One of the robust approaches is the so-called M-estimation approach, which relies on minimizing a dispersion function that is slowly varying instead of the squared residuals. Therefore, efficient estimation is possible only if a dispersion function with an appropriate resistance level is chosen(4). Little work has been done on how to choose such a dispersion function for a given dataset. From a likelihood perspective, the "best" loss function would be the negative log-likelihood function(5). The most widely used dispersion function is the Huber function. However, the choice of tuning constant in Huber function may impact the estimation efficiency to a great extent. Wang in his article, suggested obtaining the "best" tuning constant from the data so that the asymptotic efficiency is maximized (4).

The four-parameter logistic (4pl) regression with a random term is a common model to fit the concentration-response curve to the data(6). A concentration-response curve describes the relationship between response to drug treatment and drug dose or concentration. Sensitivity to a drug acting at a specific, saturable receptor typically spans a large concentration range, so dose-response curves are usually semi-logarithmic; the amount of drug is plotted as the log of drug concentration, giving them their familiar sigmoidal shape (6, 7).

The aim of this article is considering the effect of different choices of tuning constant in Huber function on robust parameters estimate. To do this, the concentration-response data examining the effect of phenylephrine in rat corpus cavernosum, has been used.

Materials and Methods

The data used in this paper is part of a concentration–response study showing the contraction of corpus cavernosum induced by phenylephrine in the organ bath. Eight different doses of phenylephrine were used. Three experimental groups were used in this study; each group consisted of eight rats. The concentration–response curves to phenylephrine (0.1μM to 300μM) were obtained by the cumulative addition of phenylephrine to the chamber.

The 4pl regression model with a random term can be written as follows:

$$y_{ij} = (A + a_i) + \frac{D-(A+a_i)}{1+\left(\frac{x_{ij}}{C}\right)^B} + \epsilon_{ij}$$

(1)

Where y_{ij} is the jth measured response of the subject exposed to X_{ij} dose, A is the upper asymptote parameter, D is the lower asymptote parameter, C is the ED50 parameter (the dose is required to elicit 50% response), and B is the rate parameter. Since the A, B, C, and D parameters are fixed effects, and the parameter α is a random term, model1 is a nonlinear mixed model.

In order to diagnose the outliers, a plot of the response against the log dose was used. Because of the existence of an outlier, M-estimation method and Huber function as a dispersion function are used to achieve robust estimations(3).

Huber function can be written as follows:

$$\rho_{H(r_i)} = \begin{cases} \frac{1}{2} r_i^2 & |r_i| \leq c \\ c\,|r_i| - \frac{1}{2}(c)^2 & o.w \end{cases}$$

(2)

The constant *c* must be pre-specified. In fact, the constant *c* regulates the amount of robustness. The choice of *c* can have a great impact on the estimation efficiency. The choice of *c* should reflect the possible proportion of outliers in the data. It is therefore sensible to adjust the *c* value accordingly based on the distribution of the data (4).

Results

Eight different doses of phenylephrine in three clusters were used in this study. 24 observations were measured. A plot of the response (the contraction of corpus cavernosum) against the logdose for each cluster is given in Figure1.

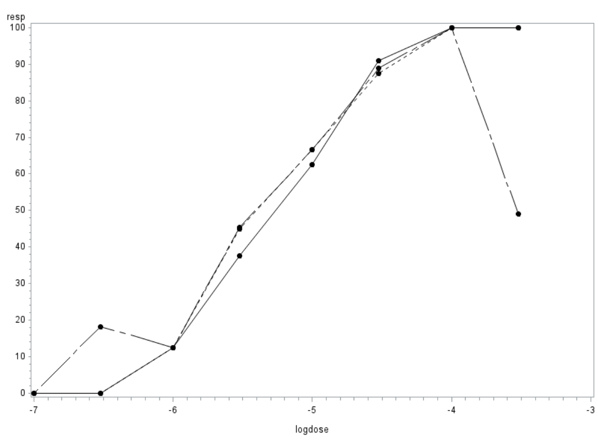

Figure 1. The percent of corpus cavernosum contraction against the logarithm of the dose in each cluster.

As shown in figure1, there is an outlier in the eighth cluster, eighth dose, so robust estimation should be considered. As mentioned earlier the choice of *c* in Huber function is very important and it should be chosen based on the distribution of the data. The comparison of non-robust PA curve and robust PA curve, for different measures of *c* was shown in figure2 to figure 4.

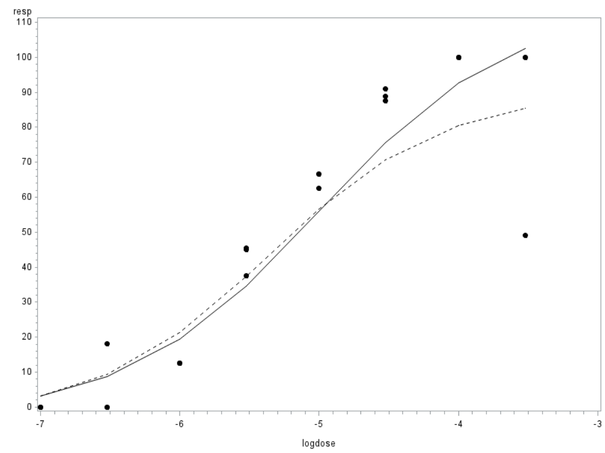

Figure 2. Comparison of non-robust PA curve and robust PA curve to concentration-response data, *c*=0.1

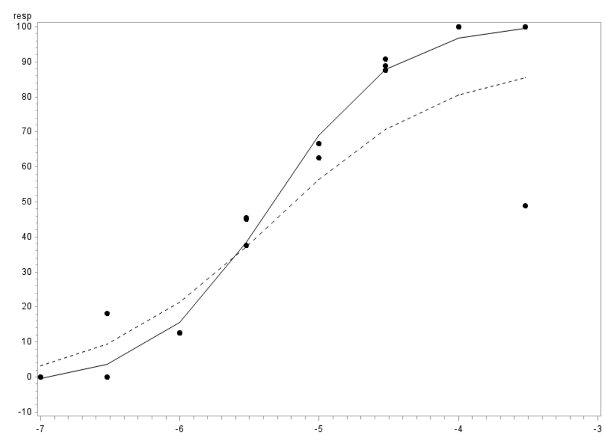

Figure 3. Comparison of non-robust PA curve and robust PA curve to concentration-response data, *c*=1.5.

As can be seen when c=1.5, the model has better fit than the other two values for *c*. in addition, for comparing these three robust models with different measures of *C*, the Negative Log Likelihood was obtained. The Negative Log Likelihood statistic is 164.5 and 84.409 and 163.84 for the first robust model (c=0.1) and for the second robust model (*c*=1.5) and for the third robust model (*c* =4), which means the second robust model had better fit than the others. The parameters estimates of these three models were given in table1.

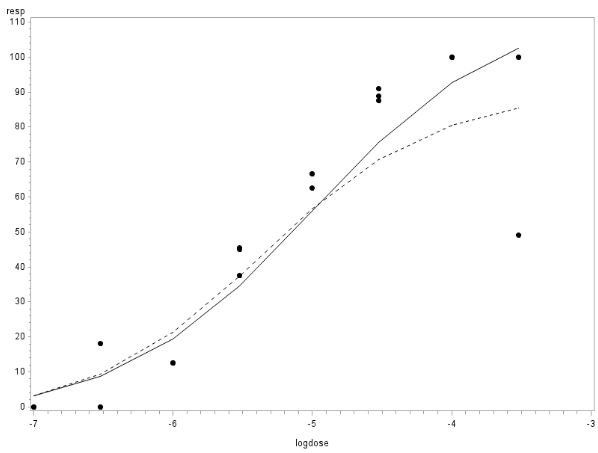

Figure 4. Comparison of non-robust PA curve and robust PA curve to concentration-response data, *c*=4.

Consequently, considering c=1.5, the 4pl with a random term parameters estimates, in columns (2) and (3) and the robust model parameters estimates in columns (4) and (5) have been displayed in table1.

For comparing the robust model and the common model, the Akaike Information Criterion (AIC) was obtained. The AIC statistic is 180.818 and 255.2 for robust model and common model which means robust model had better fit than the common model. According to the results of robust model in table1 the contraction of corpus cavernosum induced by phenylephrine in the organ bath. Cumulative administration of phenylephrine (0.1µM - 300µM) caused concentration-dependent contractions in strips of rat corpus cavernosum (-Log EC_{50} was 5 ± 0.31, 95% CI= 5.92 to 4.21). The contraction of corpus cavernosum started in the concentration of 0.3µM and then gradually increased in a dose-dependent manner till it reached a plateau in 100µM.

As displayed in figure3, the parameter A estimate in the common model is much lower than this in robust model. This underestimating

Table 1. Comparison of three robust models output with different measures of *c*

	c=0.1		*c*=1.5		*c*=4	
Parameter(1)	Estimate(2)	S.E(3)	Estimate(4)	S.E(5)	Estimate(6)	S.E(7)
A	100	4.14	100	0.95	100	1.26
B	0.63	0.05	1.125	0.03	1.02	0.05
C	9.993E-6	1.0789E-6	9.7008E-6	0.31	4.539E-6	28841E-7
D	0	2.43	0	1.18	0	1.41
σ^2	2.3	16.63	1.76	1.07	2.310E-8	1.34
σ_a^2	1.4	3.02	1.6	0.56	4.99	1.57

Table 2. Comparison of non-robust output and robust output

	Non-robust model		Robust model	
(Parameter(1	(Estimate(2	(S.E(3	(Estimate(4	(S.E(5
A	90.00	12.83	100	0.95
B	0.70	0.05	1.125	0.03
C	4.659E-6	0.48	9.7008E-6	0.31
D	2	1.81	0	1.18
σ^2	30.98	3.05	1.76	1.07
σ_a^2	48.72	90.30	1.6	0.56

just happens because of the excessive influence of the outlier in the upper asymptote area. The Population Average (PA) in common model considerably underestimates the mean response in the upper asymptote.

Discussion

In this paper, the effect of choosing the best tuning constant in Huber function on the outlier robust nonlinear mixed model estimation was examined. To show the significant effect of choosing *c* in parameter estimates, the real data was used.

In real data, the presence of outliers is unavoidable. In this case, to estimate the parameters, the outlier robust procedure must be used (3). In outlier robust statistical procedures the influence of outliers in estimating parameters are adjusted. There are several approaches in robust statistic. Mancini et al. (8) and Muler and Yohia (9) proposed a robust M-estimator that assigns a much lower weight to the outliers than the Gaussian maximum likelihood estimators does. Pinheiro et al. (10) and Staudenmayer et al. (11) introduced robust estimation techniques in which both random effects and errors have multivariate Student-t distributions. Yeap and Davidian (12) proposed a two-stage approach for robust estimation in nonlinear mixed effects when outliers are present within and between individuals. Finally the procedure used in this article was proposed by Williams in 2015. He introduced a one-step approach by utilizing a robust version of the linearized Gaussian likelihood for the nonlinear mixed model(3). As pointed out by Huber(13), "The constant *c* regulates the amount of robustness; good choices are in the range between 1 and 2, say, $c = 1.5$." Other values are also used in the literature, for example, $c = 1.2$ (14), $c = 1.25$ (15). The default value of *c* in R package (rlm function) is 1.345 to achieve about 90% efficiency when the data are normally distributed (4).

As displayed in table 2, the parameter *A* estimate based on two models is different. The outlier made a considerable difference in the estimations .Robust estimations are reliable because the robust model is not affected by outliers. The robust parameter estimations do not systematically underestimate the PA near the upper asymptote area. It is clear from figure3, that the outlier has not influenced the estimates of the robust model as significantly as the estimates of the common model. Most researchers find that in the existence of the outliers, the Gaussian quasi-maximum likelihood estimators are very inaccurate. Accordingly In the presence of the outliers four parameter logistic regression with a random term need to be robust using appropriate technique.

Conclusion

In this paper, in order to consider the clustering feature of dose-response data, the 4pl regression with a random term has been used. To estimate parameters of the model because of existence of an outlier in dataset, M-estimation method and Huber function as a dispersion function has been applied. The appropriate choice of tuning constant can be led to accurate results.

Conflict of Interest

Not Declared.

References

1. Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA. Robust statistics: the approach based on influence functions: John Wiley & Sons, Inc. 2011.

2. Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. Applied linear statistical models: Irwin Chicago; 1996.

3. Williams JD, Birch JB, Abdel-Salam ASG. Outlier robust nonlinear mixed model estimation. Stat Med. 2015;34:1304-16.

4. Wang Y-G, Lin X, Zhu M, Bai Z. Robust estimation using the Huber function with a data-dependent tuning constant. J Comput Graph Stat. 2015;16:468-81.

5. Schrader RM, Hettmansperger TP. Robust analysis of variance based upon a likelihood ratio criterion. Biometrika. 1980;67:93-101.

6. Mailman RB. Toxicant–receptor interactions: fundamental principles. Molecular and Biochemical Toxicology 4th Edition Chichester: John Wiley & Sons, Inc. 2008:359-90.

7. Woolcock AJ, Salome C, Yan K. The Shape of the Dose-Response Curve to Histamine in Asthmatic and Normal Subjects. Am Rev Respir Dis. 1984; 130:71-5.

8. Mancini L, Ronchetti E, Trojani F. Optimal conditionally unbiased bounded-influence inference in dynamic location and scale models. J Am Stat Assoc. 2005;100:628-41.

9. Muler N, Yohai VJ. Robust estimates for GARCH models. J Stat Plan Inference. 2008;138:2918-40.

10. Pinheiro JC, Liu C, Wu YN. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. J Comput Graph Stat. 2001;10:249-76.

11. Staudenmayer J, Lake E, Wand M. Robustness for general design mixed models using the t-distribution. Statistical Modelling. 2009;9:235-55.

12. Yeap BY, Davidian M. Robust Two-Stage Estimation in Hierarchical Nonlinear Models. Biometrics. 2001;57:266-72.

13.Huber PJ. Robust estimation of a location parameter. Ann Math Sta. 1964;35:73-101.

14. Cantoni E, Ronchetti E. Robust inference for generalized linear models. J Am Stat Assoc. 2011; 96:1022-30.

15. Carroll RJ, Ruppert D. Transformation and weighting in regression: CRC Press; 1988.