Use of Statistical Methodology based on Weighted Median Approach to analyze Loudness Discomfort Data

By Mansour A. Ataa*

Abstract :

A real-life application related to data for estimating the Loudness Discomfort from Acoustic Thresholds is presented to demonstrate the use of statistical methodology based on the l_1-norm as a robust alternative to the usual least squares. In this paper, a robust procedure using the l_1-norm estimator was considered to estimate the loudness discomfort level (LDL) from the acoustic threshold level (ART). The estimation results have shown that LDL can conveniently be predicted from ART. To have more confidence in the results and conclusions, the weighted median of l_1-norm was utilized to compute intervals for the values of the response (LDL) variable. Such intervals would provide a critical examination of l_1-regression fit and the data set as well.

***Keywords*: Least absolute deviations, asymptotic tests, Least squares, LAD, l_1-norm, Marginal medians, Defining observations, Non-defining observations.**

* Assistant Prof. Of Statistics at Sanaa University.
Address for correspondence:
Sanaa University, P. O. Box 14670. Sanaa. Republic of Yemen.

1. Introduction

The current paper is concerning with the analysis of real-life application associated with data sets related to medical problem. The estimation results to these data sets using the least squares method was previously discussed in Mansour A. Ataa (200). It was concluded that, the predictions of loudness discomfort level (LDL) from acoustic reflex thresholds (ART) by least squares method were in general unsatisfactory. The estimation results have also indicated that acoustic reflex thresholds correlate too poorly with the loudness discomfort level. Accordingly, the acoustic reflex thresholds may not be appropriate to be used as an objective measure of loudness discomfort level. The failure of least squares method in detecting the relationship between ART and LDL might be due to the fact that these data sets have actually been associated with some outlying data points. Accordingly, the use of l_1-norm method was recommended to analyze these data sets. For more details about the least squares analysis, see, Mansour A. Ataa (2000).

The l_1-norm method or least absolute deviations (or LAD), is extensively being adopted in statistical estimation and numerical approximation (Narula, 1987). In terms of reducing the effect of abnormal errors or heavy-tailed error distributions, the l_1-norm is more robust than the least squares. The inference procedures and hypothesis tests based on l_1-norm method have been developing. For literature on such developments, the reader is referred to Basset and Koenker (1978), Koenker (1987), McKean and Schradder (1987), Rao (1988), and Bai, Rao and Yin (1990).

2. The Data Problem

Three groups (Normal, Conductive and Perceptive) of data sets are available for analysis. These groups were associated with 20, 30 and 42 subjects, respectively. The clinicians of the ANT department of the Republican Teaching Hospital have collected these three data groups. For each of these three groups, three variables were considered: Acoustic Reflex Threshold level (ART), Loudness Discomfort level (LDL) and Most Comfortable level (MCL). For more details about these data sets, the reader is referred to Mansour Ataa, (2000). In the current study only two groups with two variables were considered. These groups are: Conductive and Perceptive groups with the Loudness Discomfort level (LDL) representing the dependent variable and Acoustic Threshold level (ART) representing the independent variable.

2.1 The objectives

The following objectives have been sought:

1. To find accurate relationships (if any) between the Loudness Discomfort Level (LDL) and Acoustic Thresholds (ART). That is, to estimate LDL from ART using weighted l_1-norm approach.
2. To represent these relationships in the forms of linear models which may be used in applied circumstances.
3. To determine intervals for the values of the dependent variable using weighted median approach based on the l_1-norm.

2.2 Modeling The Data

The relationship between the dependent variable LDL and the independent variable ART can take the following simple linear form:

$$LDL = \alpha_0 + \alpha_1 ART + \text{Error term} \quad (1)$$

Where the parameter estimates α_0, α_1 in (1) are to be estimated by $\hat{\alpha}_0, \hat{\alpha}_1$. Denoting by Y to LDL and by X to ART, the above linear forms can be rewritten in the following linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1,2,\Lambda\ ,n \quad (2)$$

Where the unknown parameters β_0, β_1 are required to be estimated and where ε_i, $i = 1,2,\Lambda\ ,n$ are the random error variables. The linear model (2) was fitted to the data sets using the l_1-norm giving the results in Table (2.1). Moreover, to test for model adequacy, the test statistics is defined as $T_{l_1} = \dfrac{2B}{[w/(n-m)]} \xrightarrow{d} \chi^2_{(q)}$ where q is the rank of l_1-solution and where $B = LAD_r - LAD_f$ and $W = LAD_f(\hat{\beta})$. The quantities LAD_r and LAD_f are, respectively, the LAD values of the reduced and the full models (see, M. Ataa, 2000). As an example, the result for testing the model $\hat{LDL}_1 = 93.00 + 0.20\, ART_1$ produced by the un-weighted l_1-norm shown in Table (2.1) (Perceptive group) is given by: $T_{l_1} = 318.94 > \chi^2(2) = 5.99$ which is highly significant at 0.05 levels of significance.

Table (2.1): Parameter estimates using l_1-norm for Fitting the Linear Model:

$$\hat{LDL}_r = \hat{\beta}_0 + \hat{\beta}_1 \, ART_r , \quad r = 1,2,\Lambda \ ,8$$

Group	Conductive		Perceptive	
Stimulus	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
1	105.33	0.13	93.00	0.20
2	90.00	0.30	91.67	0.17
3	105.33	0.67	89.17	0.17
4	88.67	0.07	71.25	0.25
5	72.50	0.25	75.00	0.20
6	98.00	0.00	63.13	0.38
8	76.00	0.20	80.00	0.11
9	63.33	0.33	65.00	0.25

3. Computing Intervals for LDL using Weighted Median Approach

To allow the user a more complete analysis of the l_1-norm regression fit, it would be very useful to determine intervals for the response (here LDL) and/or the predictor (here ART) variables. In this section, we only compute the intervals for the response (LDL) variable. This type of analysis will allow the user, in applied circumstances, a clear, simple and straightforward way to examine a data set and the l_1-norm regression fit critically and thus to have more confidence in the results and conclusions. As illustration, consider the data sample, in Table (3.1), taken from the conductive group of subjects corresponding to stimulus 2. Using the linear model (2), the l_1-norm regression line fitted to the data in Table (3.1) is given by

$$\hat{y}_i = 90.0 + 0.3x_i , \ i = 1,2,\Lambda \ ,19 \qquad (3)$$

Table (3.1): $ART_i = x_i$, $LDL_i = y_i$, $\hat{y}_i = \hat{LDL}_i$ **values and** e_i **is the** i**th** l_1**-residual corresponding to stimulus 2 taken from fitting of equation (3) to Conductive Group of subjects.**

i	x_i	y_i	$\hat{y}_i$	e_i	i	x_i	y_i	$\hat{y}_i$	e_i
1	105	113	121.5	-8.50	11	105	125	121.5	3.50
2	100	115	120.0	-5.00	12	110	125	123.0	2.00
3	110	125	123.0	2.00	13	100	125	120.0	5.00
4	115	125	124.5	0.50	14	105	125	121.5	3.50
5	105	125	121.5	3.50	15	110	123	123.0	0.00
6	110	125	123.0	2.00	16	100	120	120.0	0.00
7	115	120	124.5	-4.50	17	105	115	121.5	-6.50
8	105	115	121.5	-6.50	18	115	115	124.5	-9.50
9	100	115	120.0	-5.00	19	120	125	126.0	-1.00
10	95	115	118.5	-3.50					

So, the objective is to use the weighted median of the l_1 norm to determine intervals for the values of the LDL variable for the non-defining and defining observations. The non-defining observations within which the l_1-norm regression does not change are defined to be the non-zero l_1-residuals $(e_i \neq 0)$. On the other hand, the defining observations are those observations corresponding to the zero l_1-residuals $(e_i = 0)$. The rule of defining and non-defining observations in constructing inte rvals for the response and/or predictor variables have been discussed in Narula and Wellington (1985) and Narula, Sposito and Wellington (1993). To calculate the intervals for the LDL, the weights w_{ij} were computed based on weighted median approach using the method of l_1-norm as follows: the linear model (2) was fitted to the data set given in Table (3.1) using the usual method of l_1-norm yielding the result in equation (3). Note that, some the observed residuals $e_i = y_i - \hat{y}_i$, $i = 1,2,\Lambda\ ,n$ will be equal to 0

for at least q observations where $\hat{y}_i$ is the predicted value of the response variable of the ith observation. Now, the problem formulation can be summarized as follows: for model (2), the l_1-norm problem can be written as shown by Josvanger and Sposito (1983) and Narula, Sposito and Wellington (1993) by:

$$\min_{\beta} \sum_{i \neq j}^{n} w_{ij} \left| s_{ij} - \beta_1 \right| \qquad (4)$$

where $w_{ij} = \left| x_i - x_j \right|$ and where $s_{ij} = (y_i - y_j)/(x_i - x_j)$.

Since $w_{ij} \neq 1$ for all i and j. Problem formulation (4) is a weighted median problem. Let $w_{(ij)}$ denote the weights of the corresponding ordered s_{ij}. To determine the optimal weighted median solution to problem (4) with respect to (x_j, y_j), we first determine r_j such that

$$\sum_{i=1}^{r_j} w_{(ij)} \geq \sum_{i=r_j+1}^{n} w_{(ij)}$$

and

$$\sum_{i=1}^{r_j-1} w_{(ij)} < \sum_{i=r_j}^{n} w_{(ij)}$$

(5)

If the observation at i corresponds to r_j then $\hat{\beta}_1 = s_{ij}$. As noted by Josvanger and Sposito (1983) and Narula, Sposito and Wellington (1993), if the pth observation is one of the defining observation, then the solution to the weighted problem will lead to an observation (x_p, y_q) such that the pth and qth observations define a regression line based on the l_1-norm: observations associated with $e_i = 0$.

3.1 Interval Calculations: Non-defining observations for (LDL) variable

To calculate intervals for the values of the response (LDL), the data of LDL and ART and the corresponding l_1-residuals produced by equation (3) are shown in Table (3.1). It is useful to layout the tables of w_{ij} and S_{ij} for the defining observations p and q as shown in Table (3.2). For our example, these values for the defining observations $q = 15$ and $p = 16$ were given in Table (3.2). Let y_i denote the new value of LDL variable for the ith observations, $i \neq p, q$. From the weighted median formulation of the problem we observe that y_i should be such that

$$\frac{y_i - y_p}{x_i - x_p} \geq \hat{\beta}_1, \text{ if } S_{ip} > \hat{\beta}_1 \qquad (6a)$$

or

$$\frac{y_i - y_p}{x_i - x_p} \leq \hat{\beta}_1, \text{ if } S_{ip} < \hat{\beta}_1 \qquad (6b)$$

For our example, with $i = 3$, we have

$$(y_3 - 115)/(110 - 105) \geq 0.30 \text{ or } y_3 \geq 116.5$$

Table (3.2): Values of w_{i15}, S_{i15} and w_{i16}, S_{i16} for defining observations

$(x_{15}, y_{15}) = (110,123)$ **and** $(x_{16}, y_{16}) = (100,120)$

i	$w_{i15} = \lvert x_i - 105 \rvert$	$y_i - 115$	$w_{i15} = \lvert x_i - 115 \rvert$	$y_i - 120$	S_{i15}	*Rank*	S_{i16}	*Rank*
1	5	-10	5	-7	2.00	1	1.40	1
2	10	-8	0	-5	1.60	8	1.00	8
3	0	2	10	5	1.60	17	1.00	10
4	5	2	15	5	1.60	18	1.00	17
5	5	2	5	5	0.80	2	0.33	18
6	0	2	10	5	0.80	9	0.00	7
7	5	-3	15	0	0.60	7	-0.25	$r_{16} = 19$
8	5	-8	5	-5	0.53	$r_{15} = 10$	-0.30	15
9	10	-8	0	-5	0.30	16	-0.33	4
10	15	-8	5	-5	-0.20	13	-0.50	3
11	5	2	5	5	-0.20	19	-0.50	6
12	0	2	10	5	-0.40	4	-0.50	12
13	10	2	0	5	-0.40	5	-1.00	5
14	5	2	5	5	-0.40	11	-1.00	11
15	0	0	10	3	-0.40	14	-1.00	14
16	10	-3	0	0		3		2
17	5	-8	5	-5		6		9
18	5	-8	15	-5		12		13
19	10	2	20	5		15		16

Hence, for $y_3^* \in [116.5, \infty)$ the current l_1-regression line will not change. In Table (3.3), we give the intervals, which correspond to the non-defining observations within which the l_1-regression line for the data in Table (3.1) continues to be given by equation (3). If the response value for the defining observations is changed, the l_1-regression for the data is changed. However, if this change is limited to certain intervals, the defining

observations for the l_1-regression for the data will remain the same. Using the above notes, the intervals for the values of the response (LDL) variable were given in Table (3.2). If the computed intervals of LDL are to be compared with the predicted values using the linear model (3), Table (3.1) exhibits LDL values and $\hat{LDL}$ predicted values. As can be seen, the predicted values of LDL are within the intervals of either that of non-defining observation 16 or of non-defining observation 15 displayed in Table (3.3). For example, $\hat{y}_1 \in [115, \infty)$, $\hat{y}_2 \in [115.5, \infty)$, $\hat{y}_3 \in [116.5, \infty)$, $\hat{y}_4 \in [118, \infty)$, etc.

Table (3.3): Intervals for the values of the response (LDL) Variable for the non-defining observations

Non-defining observation 15				**Non-defining observation 16**			
	Interval for				***Interval for***		
i	y_i	i	y_i	i	y_i	i	y_i
1	$[115, \infty)$	11	$[115, \infty)$	1	$[117, \infty)$	11	$[117, \infty)$
2	$(-\infty, 113.5]$	12	$[116.5, \infty)$	2	$[115.5, \infty)$	12	$[118.5, \infty)$
3	$[116.5, \infty)$	13	$(-\infty, 113.5]$	3	$[118.5, \infty)$	13	$[115.5, \infty)$
4	$[118, \infty)$	14	$[115, \infty)$	4	$(-\infty, 120]$	14	$[117, \infty)$
5	$[115, \infty)$			5	$[117, \infty)$		
6	$[116.5, \infty)$			6	$[118.5, \infty)$		
7	$[118, \infty)$	17	$[115, \infty)$	7	$(-\infty, 120]$	17	$[117, \infty)$
8	$[115, \infty)$	18	$[118, \infty)$	8	$[117, \infty)$	18	$(-\infty, 120]$
9	$(-\infty, 113.5]$	19	$[119, \infty)$	9	$[115.5, \infty)$	19	$(-\infty, 121.5]$
10	$(-\infty, 112]$			10	$[114, \infty)$		

3.2 Defining observations for (LDL) variable

Let y_p represent the new value of the defining observation p from the weighted median formulation of l_1-problem. Since the value of y_p does not affect equation (5), the pth observation will continue to be a defining observation if y_p^* is such that

$$S_q(r_q - 1) \le \frac{y_p - y_q}{x_p - x_q} \le S_q(r_q + 1) \qquad (7)$$

where $S_i(j) = S_{ij}$. For our example with $p = 16$, from Table (3.2), we have

$$0.6 \le \frac{y_{16}^* - 123}{100 - 110} \le 0.3 \text{ or } 120 \le y_{16}^* \le 117$$

i.e., if $y_{16}^* \in [117, 120]$, then observation 16 will continue to be defining observations. The interval calculations for defining observations are given in Table (3.4).

Table (3.4): Intervals for the values of the response (LDL)

Variable for the defining observations	
Defining observation 15	**Defining observation 16**
Interval for y_{15}	**Interval for** y_{16}
$123 \le y_{15} \le 120$	$120 \le y_{16} \le 117$

5. Discussion

The results can be summarized in the following concluding remarks:

(1) The linear models fitted to these data sets using l_1-regression have indicated that acoustics reflex thresholds (ART) can be used as an objective measure

of loudness discomfort (LDL); that is, the LDL can conveniently be predicted from ART.

(2) Unlike the least square estimates, the l_1-norm estimates of model (2) do not change if the value of the response variable lies within a certain specified interval. Moreover, observations with the largest residual need not be influential, i.e. the l_1-norm methodology for statistical data analysis is insensitive to certain types of discrepancy in a data set.

(3) In this paper we have only constructed intervals for the values of the response (LDL) variable at stimulus 2. The derived intervals for values of the response (LDL) variable allows the user to examine a data set and the l_1-norm regression fit critically and thus to have more confidence in the results and conclusions. It is straightforward to calculate similar intervals for LDL at the other stimuli.

In conclusion, the l_1-norm methodology in statistical data analysis could be the most appropriate method for analyzing abnormal data.

References

1) Bai, Z. D., Rao, C. R., and Yin, Y. Q., 1990. "Least absolute deviations
2) Analysis of variance". Sankhya, **52** (A), pp. 166-177
3) Bai, Z. D., Chen, X. R., Wu, Y., and Zhao, L. C., 1987. "Asymptotic
4) Normality of the minimum l_1-norm estimates in linear models".
5) Technical report 87-35, Center for Multivariate analysis, University
6) of Pittsburgh.
7) Barrodale, I., and Roberts, F. D., 1974. "Solution of an over determined
8) System of equations in l_1-norm ". J. ACM., **17**, pp. 319-320.
9) Bassett, G. W., and Koenker, R. W., 1978. "The asymptotic distribution
10) of the least absolute error estimator" J. Amer. Statist. Assoc., **73**,
11) pp. 618-620.
12) Mansour A. Ataa (2000). Constrained l_1-norm for analyzing Categorical or Ordinal data. J. Yemeni Sci. (Conference Press.).
13) Mansour A. Ataa and Mohammed F. Alazazi, (2000). Estimation of Loudness Discomfort and Most Comfortable Levels from Acoustic Thresholds using Least Squares method (in press).
14) Mansour A. Ataa (2000). Robust Statistical Approach For Analyzing The Loudness Discomfort and Acoustic Reflex Thresholds Data. J. Yemeni Sci. (Conference Press.).

15) Narula, S.C. (1987). The minimum sum of absolute errors regression. J. Qual. Technol., **19**, 37-45.

16) Narula, S.C., Sposito, V. A. and Wellington, J. F. (1993). Intervals which leave the minimum of absolute errors regression unchanged. Appl. Statist., **42**, No. **2**, 369-378.

17) Rao, C. R., 1988. "Methodology based on the l_1-norm in Statistical

18) Inference". Sankhyā , **50**, Series A, pp. 289-213.

19) Sposito, V. A. (1989). Linear programming with Statistical Applications, p. 139. Ames: Iowa State University Press.

استخدام الطرق الإحصائية طبقاً إلى الوزن المتوسط لرؤيا تحليلية في إستخدام الأوزان الصالحة طبقاً إلى المعطيات

منصور عطا *

خلاصة :

تطبيق حقيقي متصل ببيانات لتقدير إزعاج الصخب الصوتي من العتبة السمعية ، مقدم لاظهار منهجية إحصائية تعتمد على معيار l_1 كبديل قوي لطريقة التربيعات الصغرى للقيم المعتادة . في هذا البحث ، إجراء قوي يستعمل ميعار l_1 اعتمد لتقدير مستوى إزعاج الصخب الصوتي (LDL) من مستوى العتبة السمعية (ART) .

بينت نتائج التقدير أن LDL يمكن التنبؤ به وبشكل مريح من مستوى العتبة السمعية ART ، وللحصول على ثقة أكبر بالنتائج والإستنتاجات ، فان المتوسط الموزون لمعيار l_1 قد استخدم لحساب الفترات الفاصلة لقيم متغير الاستجابة LDL . إن مثل هذه الفترات الفاصلة تعطي الفحص الحرج للقراءات المتغيرة ولمجموعة البيانات كذلك .

* جامعة صنعاء الجمهورية اليمنية ص . ب (١٤٦٧٠) .