

HYPOTHESIS TESTING & CONFIDENCE INTERVAL ESTIMATION VIA THE LEAST ABSOLUTE VALUE (LAV) APPROACH

Tom G. Filloon, Procter & Gamble, Cincinnati, Ohio

ABSTRACT

We often analyze data sets which are heavy-tailed or contain extreme values that cause one to be wary of Least Squares (LS) analyses. In such cases, one would like to perform a more "robust" analysis of the data to either confirm or replace the LS results. One robust alternative to Least Squares is to use the Least Absolute Value (LAV) estimation approach (a.k.a. L1 or Median regression). This approach is highly robust to extreme data as it is simply estimating the median of a conditional distribution. The LAV approach works for any general linear models setting (e.g. 1-way ANOVA, ANCOVA, regression,...) and estimates parameters by minimizing the sum of absolute residuals, instead of the sum of squared residuals. Point estimation of parameters via the LAV approach has been around for a long time and is given as an example in the SAS/IML manual. What is needed to perform statistical inference (hypothesis testing/confidence intervals) with the LAV approach is a method for estimating the parameter covariance matrix. This has been the sticking point because the usual formulas necessitate the awful task of estimating the density of the model residuals. A simpler yet asymptotically valid approach is to use a function of order statistics of the model residuals to estimate this covariance matrix. Thus, using SAS IML, one is able to calculate parameters and their covariance estimate in order to perform statistical inference via the LAV approach. This provides a nice adjunctive analysis to LS when the residuals are well-behaved and a nice alternative analysis otherwise.

1. INTRODUCTION

Most statisticians have been brought up in the "Least Squares" school of statistics. We are taught about how to compare means and regression lines all using the recurrent normal theory. Furthermore, we are versed in the ways of transforming data and possibly, performing outlier deletion in the hopes of achieving our "normal" data which can then be analyzed by "normal" methods. Then usually after this learning, one finds out about more robust methods that can be used when we do not have or can't massage our data to be "normal". Most approaches to robust estimation tend to be confusing and/or user-friendly software is not readily available. So one usually tends to resort to LS methods because we know how to analyze "normal" data and the computational effort needed is trivial.

But to understand all of this robust estimation more fully, let's get to the fundamental problem of estimating model parameters. We will start off by defining a general linear models setting but will come back later to simple examples to illustrate the methods. Let $Y = X\beta + \varepsilon$, where Y is the response, X is the model matrix, β are the model parameters we are interested in estimating, and ε is the vector of errors (assumed to be independent, identically distributed $(0, \sigma^2)$). Now if we know that our error terms are normally distributed, then via maximum likelihood estimation, we know that least squares estimation (due to the $\exp(-0.5x^{**2})$ nature of the normal likelihood) is best. Hence, for least squares estimation, one is estimating the model parameters by minimizing $\sum_i (Y_i - X_i\beta)^2 = \sum_i r_i^2$ with respect to β . Thus

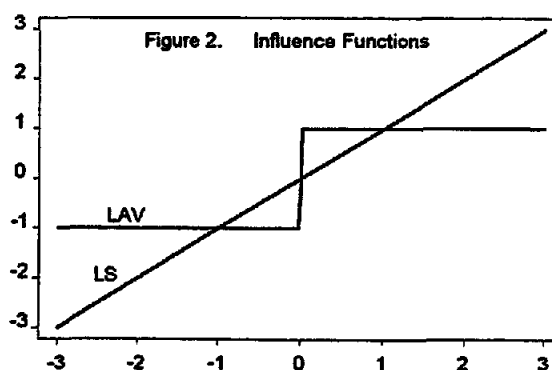
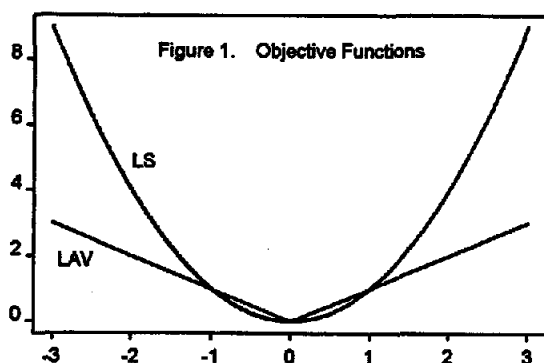
one can see we are just trying to get small residuals via some global objective function, which in this case is a sum of squares function. The real goal of all of this is to generate reliable parameter estimates, with small standard errors. One drawback with this particular sum of squares function is that large residuals (which can occur with "real" data either from heavy-tailed or contaminated error distributions) will tend to have an inordinate amount of influence on the estimation and will yield biased parameter estimates and inflated parameter standard errors. The problem with least squares estimation is that the amount of influence than an individual observation has (i.e., influence function) can be unbounded and such extreme observations can "break down" the estimators. Robust estimators are those that have bounded influence functions and hence, are advantageous in that they are robust/resistant to extreme observations (sometimes called outliers). For this article, we will consider one type of robust estimation that handles extreme data in the Y responses. We will not consider the handling of extreme data in the

X factor space/leverage points (Least Median of Squares or Least Trimmed Squares - Rousseeuw & Leroy (1987). Other robust methods (R, M) are available but will not be discussed further here (Hettmansperger 1984; Huber 1981; Fillooon & Drago 1995).

2. LAV ESTIMATION

One might consider minimizing the sum of absolute residuals in order to yield parameter estimates. This results in Least Absolute Value (LAV) estimation, also known as LAD, LAE, L1 or Median regression (Bloomfield & Steiger 1983). Notice that this minimizes $\sum_i |r_i|$ with respect to β , which results in median-

like estimation in that it causes the number of positive residuals and negative residuals to be equal. Notice that this is performing MLE for double-exponentially distributed errors. But regardless of data distribution, one could validly estimate model parameters via minimizing an LAV objective function of residuals (it just might not be the best approach). This provides the general linear models generalization of estimating the median for a 1-sample location problem. An interesting feature of this approach is that by weighting the positive and negative residuals differently, one can be estimating other regression percentiles besides the median (Koenker & Bassett 1978). Figures 1 & 2 show the objective and influence functions respectively for the LS and LAV estimation. One can see graphically the bounded nature of the LAV approach (vs. the unbounded LS method).



3. ASYMPTOTICS & INFERENCE

Now that we know how (in theory) to compute our LAV estimator, what next? As statisticians, we should be concerned with estimating standard errors (i.e., covariance matrix) of our parameter estimates by which to carry out inference (estimation and/or testing). Let's go back to LS estimation for review. We know that, for the general linear models setting, our variance-covariance matrix takes the form

$Var(\hat{\beta}_{LS}) = \sigma^2 (X^T X)^{-1}$ if the errors are iid($0, \sigma^2$) and furthermore that $\hat{\beta}_{LS}$ is asymptotically normally distributed. For real data, one replaces σ^2 by its estimator s^2 and uses a t distribution approach to account for the uncertainty in the variance estimator. For LAV estimation in the linear models setting, it turns out that the parameter estimator is asymptotically normal and that the corresponding variance-covariance matrix is surprisingly similar to that of the LS method. For LAV-estimation, one obtains

$Var(\hat{\beta}_{LAV}) = \omega^2 (X^T X)^{-1}$. In practice, one replaces ω by $\hat{\omega}$, where $\omega^{-1} = 2 * f(\text{median})$ with $f()$ representing the probability density function for the error distribution. Estimation of the scale parameter ω can be accomplished more simply (Birkes & Dodge 1993) by equating the length of the large sample confidence interval for the median with its equivalent, exact binomial order statistic counterpart (Hahn & Meeker 1991) using model residuals.

4. COMPUTING

In terms of computing, Least Squares estimation is obviously the simplest and most ubiquitous of the methods. It is available in most statistical software packages and is at the heart of all major software packages. Most robust software available does not allow for complete general linear models inference. For performing LAV-estimation, which is actually a linear programming problem, SAS[®] IML Example 11 (Regression Quantiles) provides code for estimation but not parameter standard errors. Construction of the scale parameter estimator is performed via the following "l1scale" SAS IML module per the Birkes & Dodge approach.

```
start l1scale(r,n,p,scale);
  m=n-p;
  k1=round((m+1)/2-sqrt(m));
  k2=round((m+1)/2+sqrt(m));
  rnew=repeat(0,m,1);
  count=0;
  do i=1 to n;
    if fuzz(r[i]) = 0 then r[i]=0;
    if count < p then do;
      if r[i] ^= 0 then rnew[i-count]=r[i];
      else count=count+1;
    end;
    else rnew[i-p]=r[i];
  end;
  temp=rnew;
  rnew[rank(rnew),]=temp;
  scale=sqrt(m)/4*(rnew[k2]-rnew[k1]);
  print ,, 'L1 Scale Estimate', scale, r;
finish l1scale;
```

5. ASSUMPTION CHECKING

The usual approach to data analysis is to perform a Least Squares (LS) analysis and then to evaluate the subsequent model residuals as to validity of assumptions. The assumptions of normality and homogeneous variance (homoscedasticity) can be tested for by such things as the Shapiro-Wilk normality test and Bartlett's variance test. However, these assumptions are only necessary for getting p-values via the "normal" approach. LS analysis could still be validly performed without these assumptions if one gets p-values by using a randomization/permutation approach (more computing but valid).

The more important question is: Are the residuals symmetric and are variances homogeneous? One would like to be working with symmetric data, for if the data are symmetric, then the mean, median, etc. are all estimating the same location parameter and it makes some sense to compare these location parameters. Similarly one can check for homoscedasticity of treatment groups robustly via an L1 approach (Levene 1960) to determine if the constant scale assumption is reasonable. Thus one should really be testing for symmetry, instead of normality, to determine if these estimators are sensible at all. Then assuming symmetry and verifying constancy of the scale parameter, one can use the asymptotic results to get p-values for treatment comparisons. However, symmetry and homoscedasticity are really only needed for validation of calculating p-values via the asymptotic normal approach (as one could determine permutation p-values instead). The symmetry test of choice is the "triples" test as developed by Randles et al (1980), which is derived from U-statistics, and so it is both robust and free of normality assumptions.

6. COMPARING METHODS

Let's suppose that we now know how to compute this robust estimator (and its standard error) and that symmetry (and homoscedasticity) assumptions have been satisfied. How do we compare the estimators (LS vs. L1)? We will compare precision of the estimators via their variance-covariance matrices, which

reverts to comparison of their respective estimated scale parameters (s^2 vs. $\hat{\omega}^2$). Hence, if one had to choose the "best" estimator among several estimators, one would choose the one that yields the smallest estimated scale parameter. It is known that the robust estimators will be less efficient than the least squares method if the data are truly normal (L1 - 64% efficiency). But these are asymptotic efficiencies under an assumed model, so it is best to evaluate these various methods on a dataset by dataset basis in order to be performing the most efficient/precise estimation. Furthermore, it is probably more important that one use some type of robust analysis in concert with LS in order to gain some qualitative robustness of conclusions than to try to worry about which method is best (either for a particular case or in general).

7. EXAMPLES

i. Example 1. Cholesterol-lowering animal experiment (2-sample comparison)

The first example comes from McKean & Vidmar(1994). In this data set, a treated group of quail is to be compared to a placebo group of quail with regards to LDL (Low Density Lipoprotein) Cholesterol to determine if the test drug given has the potential for cholesterol-lowering. From this boxplot comparison, the boxplots are at least in the right direction for a possible cholesterol lowering effect by the test drug. There is one extremely high response in the treated group of quail that obviously will have a huge impact on a LS analysis. For comparison, McKean & Vidmar compared the LS- and R (Rank-based)-estimation procedures to see what inferences result from the different methods. To modify this comparison, we have performed the corresponding LAV analysis (estimating the difference in location between the two groups) and included it in the following table.

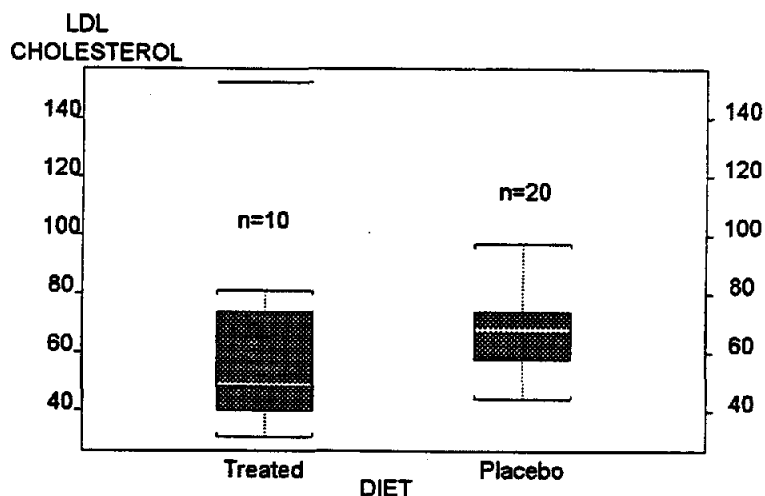


Figure 3. Boxplots for Quail Data

Table 1. Analyses for the Quail Example

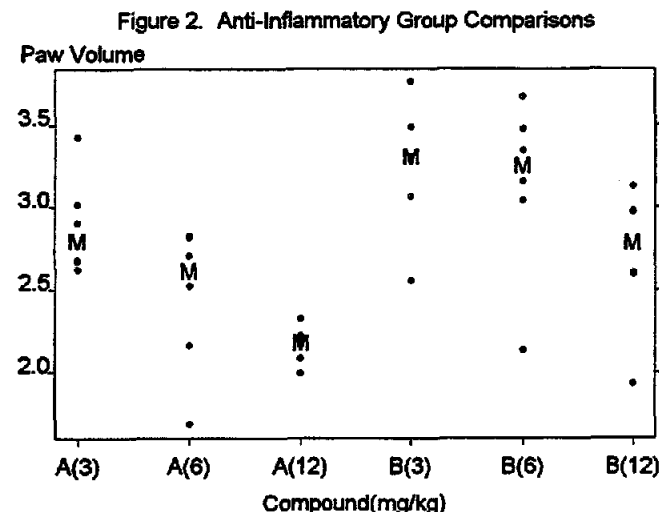
Analyses	Estimator	Std. Error	t-statistic	p value (1-sided)
LS	5.0	8.96	0.56	0.290
L1	18.0	8.19	2.20	0.018

Notice that the one large observation in the treated group tends to mislead the LS analysis, whereas the more robust LAV method suggests that a difference exists between the two groups. Regarding assumption checking, although the LS residuals are not normal (Shapiro-Wilks test), both of the methods yield symmetric errors (Triples test). Thus, one feels comfortable in comparing location/center differences between the two groups. Even though the LS residuals are not normal, the LS estimator itself is known to be asymptotically normal so that its p-value can be approximated (this would be exact if

one knew the data/errors were really normal) via t-methods. This shows how the LS approach, which is sadly non-robust, can "break down" with just one extreme observation and give misleading results.

ii. Example 2: Anti-Inflammatory animal assay (1-way ANOVA)

In this animal experiment, two compounds (each evaluated at 3 dose levels: 3, 6 and 12 mg/kg) are



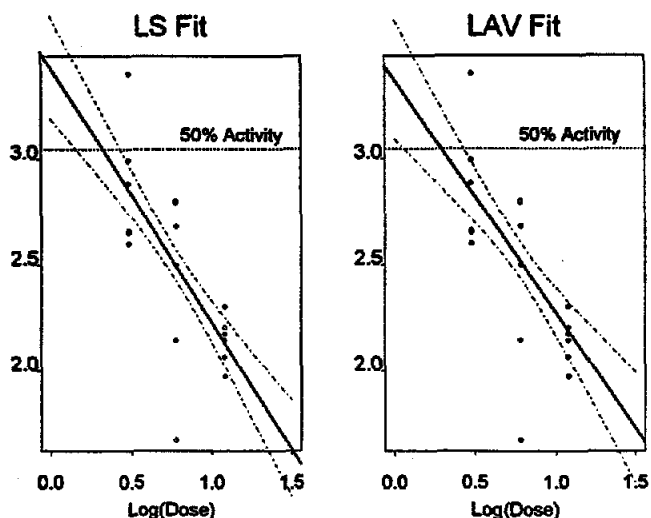
being compared (Figure 4) with regards to their anti-inflammatory responses. The letter M denotes group median. Efficacy by a compound will result in lower paw volumes and hence, response should decrease as dose level increases. For the moment, we will concentrate on comparing all of these doses via a 1-way analysis of variance model. In the next example, we will evaluate the dose response nature of one of these compounds via simple linear regression. Hence, one can perform 1-way ANOVA on this dataset using each of our various estimators. One can summarize and compare the various analyses with the following table.

Table 2. Mean & Standard Error for each Treatment Group (for each analysis)

Analysis		A (3)	A (6)	A (12)	B (3)	B (6)	B (12)
	Scale Estimate						
LS	0.405	2.88 (0.17)	2.45 (0.17)	2.16 (0.17)	3.23 (0.18)	3.13 (0.17)	2.70 (0.17)
LAV	0.471	2.68 (0.19)	2.52 (0.19)	2.16 (0.19)	3.30 (0.21)	3.15 (0.19)	2.60 (0.19)

For this dataset, LS-estimation yielded the more precise analysis (smaller scale estimate) but both analyses tend to give similar evidence of differences between the two compounds.

iii. Example 3: TAAA dose response (simple linear & inverse regression)



The following example is just a subset of the previous example, where we use just the Compound A dose groups and perform a simple linear regression analysis of the data. Note: It is typical to perform the linear regression using the log10(dose) parametrization. Figure 3 shows the various line fits with their corresponding standard error limits. In our pre-clinical evaluation of compounds, it is of interest to estimate a measure of the compound's potency in a given animal experiment. Here we are interested in determining the compound's ED50 dose (i.e., the dose of the compound that is needed to yield a 50% reduction in inflammatory response as determined from healthy & placebo animals). Thus, we are not interested in comparing slope & intercept parameters per se, but are interested in

determining where the regression line (and its SE lines) cross the horizontal 50% activity line. For a more detailed account of this inverse regression (i.e., linear calibration) problem, one should see Snedecor & Cochran 1989). From the figure, we see that the line fits (and SE's) are very similar, except that the L1 standard error limits are somewhat larger (less precise). Correspondingly, the ED50 estimators are very similar, with the LAV approach yielding larger ED50 SE's. Here we get a sense of the qualitative robustness of having another estimator for confirming the LS results.

Table 3. Linear Regression analyses for anti-inflammatory dose-response data

Analyses	Scale Estimate	Intercept (SE)	Slope (SE)
LS	0.310	3.433 (0.243)	-1.201 (0.298)
L1	0.360	3.373 (0.282)	-1.096 (0.345)

8. SUMMARY

Computational software is currently available so that robust methods can be used relatively easily. When analyzing real data, robust analyses (such as LAV) can sometimes yield more precise analyses than LS. Robust methods are not restricted to regression contexts but can be performed in any general linear models settings. Asymptotic normality of the robust methods allow for complete inference (estimation/testing). Robust methods still require assumption checking (symmetry, homoscedasticity), even if one is no longer worrying about normality testing. Comparison of the estimation approaches can be carried out via their respective scale estimators. It is highly advisable to use at least 1 "robust" method (such as LAV) for cross-validation of LS results and in many instances, the more precise robust method should serve as the primary analysis of a dataset.

Correspondence can be made to Tom Filloon at email: filloontg@pg.com or by phone: (513) 627-1581.

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration

REFERENCES

- Snedecor, G.W. and Cochran, W.G. (1989, 8th ed.) Statistical Methods, Iowa State University Press.
- Birkes, D. and Dodge, Y. (1993) Alternative Methods for Regression, John Wiley & Sons.
- Bassett, G. & Koenker, R. (1978) "Regression Quantiles", Econometrica, 46, 33-50.
- McKean, J.W. and Vidmar, T.J. (1994) "A Comparison of Two Rank-Based Methods for the Analysis of Linear Models", The American Statistician 48, 220-229.
- Filloon, T.G. & Drago, E.C. (1995) "Robust Analysis Adds Muscle to Data Analyses", SUGI20 Proceeding
- Levene, H. (1960) "Robust tests for equality for variances" In: Contributions to Probability and Statistics, Olkin et al (ed.), 278-292, Stanford University.
- SAS Institute Inc. (1990) SAS/IML Software: Usage & Reference, version 6, SAS Institute Inc., Cary, N.C.
- Bloomfield, P. and Steiger, W.L. (1983) Least Absolute Deviations, Boston: Birkhäuser
- Hettmansperger, T.P. (1984) Statistical Inference on Ranks, Krieger Publishing
- Huber, P.J. (1981) Robust Statistics, John Wiley & Sons
- Hahn, G.J. & Meeker, W.Q. (1991) Statistical Intervals: A Guide for Practitioners, John Wiley & Sons
- Rousseeuw, P.J. & Leroy, A.M. (1987) Robust Regression & Outlier Detection, John Wiley & Sons.
- Randles, R.H., Fligner, M.A., Policello, G.E. and Wolfe, D.A. (1980) "An Asymptotically Distribution Free Test for Symmetry Versus Asymmetry", JASA 75, 168-172.