*Journal of Pharmaceutical & Biomedical Analysis*
Vol. 11, No. 9, pp. 843–845, 1993
Printed in Great Britain

0731–7085/93 $6.00 + 0.00


# Estimating shelf-life using $L_1$ regression methods*

TAPON ROY

*Boehringer Ingelheim Pharmaceuticals, Inc., 900 Ridgebury Road, Ridgefield, CT 06877, USA*

**Abstract**: The shelf-life of a drug is usually determined by fitting the assayed values of drug potency versus time to a linear least-squares model. In least-squares models, the parameters are estimated by minimizing the sum of squares of the residuals. Shelf-life is estimated as the time corresponding to the intersection point of the fitted line and the minimum acceptable per cent or proportion of the initial drug potency. The least-squares method can be seriously affected by outliers, leading to erroneous shelf-life estimates. In this paper, an alternative method, based on minimizing the sum of *absolute deviations*, was applied to the shelf-life determination problem. The resistance of these $L_1$ based estimates to outliers was demonstrated using a typical stability dataset.

**Keywords**: *Stability; shelf-life;* $L_1$*; least squares.*

## Introduction

When the usual assumptions of normality and independence for the residuals hold, and outliers are not present, the classical least-squares methodology yields estimators that are optimal under many criteria. It is well established that alternative fitting methods such as the $L_1$ and the $L_\infty$ or Chebyshev techniques are effective when the usual assumptions do not hold or outliers are present, particularly in the responses [1]. These alternative criteria are not often used, partly due to computational considerations and our lack of knowledge regarding their sampling properties. The techniques used are conceptually clear and applicable to linear and multiple regression models. Since outliers are frequently encountered in stability data, these robust methods are especially important.

## Methods

The general linear model is:

$$\mathbf{y} = \mathbf{X}\underline{\beta} + \mathbf{e}, \tag{1}$$

where y is an $n$-vector of random variables, $\mathbf{X}$ is an $n \times m$ matrix of regressor variables, $\underline{\beta}$ is an $m$-vector of unknown parameters and $\mathbf{e}$ is an $n$-vector of errors.

The linear $L_p$-norm estimation problem is characterized as: find the parameter vector $\hat{\underline{\beta}}$ minimizing

$$\sum_{i=1}^{n} |y_i - \mathbf{x}_i \hat{\underline{\beta}}|^p = \sum_{i=1}^{n} |\hat{e}_i|^p. \tag{2}$$

If we let $\hat{e}_i = u_i - v_i$, where $u_i, v_i \geq 0$, represent positive and negative deviations, respectively, the general $L_p$-norm problem then becomes:

$$\text{minimize} \sum_{i=1}^{n} (u_i^p + v_i^p) \tag{3}$$

$$\text{with constraints} \quad \mathbf{x}_i \hat{\underline{\beta}} + u_i - v_i = y_i \tag{4}$$

$$\text{and} \quad u_i, v_i \geq 0 \tag{5}$$

$(i = 1, \ldots, n)$

here $\hat{\underline{\beta}}$ is unconstrained.

For the cases $p = 1$ and $p \to \infty$, linear programming procedures can be used. When $p = 2$, one has the classical least-squares problem solved using the normal equations. Unconstrained minimization techniques can be used for other values of $p$ (which need not be integer). Barrodale and Roberts [2], suggested that the convex simplex or Newton's method for $p > 1$, and a modified simplex method for $0 < p < 1$ be used.

For the $L_1$-norm problem, ($p = 1$) the linear programming implementation is in $2n + m$ variables with $n$ constraints. In ref. 3 an efficient algorithm for the $L_1$-norm problem has been developed that has been used in the supplemental SAS® Version 5 procedure LAV utilized in this paper [4]. (In SAS Version 6, a

* Presented at the Seventh Annual Meeting and Exposition of the American Association of Pharmaceutical Scientists (Analysis and Pharmaceutical Quality Section), San Antonio, TX, USA, 15–19 November 1992.

SAS/IML® program can be used for $L_1$ estimation [5].)

For the linear model, define $\hat{\beta}$ as the estimate of $\beta$ such that

$$S_1(\hat{\beta}) = \sum_{i=1}^{n} |y_i - \mathbf{x}_i \hat{\beta}| \qquad (6)$$

is minimized.

The standard error for an element of $\beta$ is then

$$\mathrm{SE}(\hat{\beta}_j) = \lambda \sqrt{(\mathbf{X}'\mathbf{X})^{jj}}, \qquad (7)$$

where $(\mathbf{X}'\mathbf{X})^{jj}$ denotes the $j$th diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$. The quantity $\lambda^2/n$ is the variance of the median of residuals from which $\lambda$ is obtained. Unfortunately, $\lambda$ is unknown, and methods for its estimation rely on evaluation of the probability density function (pdf) of the error distribution evaluated at zero. Estimation of the pdf requires choosing boundary values of ordered residuals, the selection of which can greatly affect the standard errors [6]. Thus, bootstrap methods [7] will be used to estimate the standard errors of the parameters.

The fitted regression line is:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \qquad (8)$$

where $\hat{y}_i$ represent the fitted responses, $\hat{\beta}_0$ is the estimated intercept, and $\hat{\beta}_1$ is the estimated slope.

The shelf-life is then:

$$\text{shelf-life} = (y - \hat{\beta}_0)/\hat{\beta}_1, \qquad (9)$$

where $y$ is the minimum acceptable potency.

Higher order polynomials are sometimes employed to model the potency versus time relationship. The methods described above are applicable in those cases, and shelf-life would then be determined using standard root-finding methods. In certain cases, a nonlinear (in the parameters) model is the most appropriate, and related methods, described below, can be used.

**Results and Discussion**

To illustrate the methodology, the stability data in Table 1 was analysed using least-squares and $L_1$ techniques. In Table 1, the $x$ values are storage times in months, and the $y$ values represent product strength. These are real stability data exhibiting a pattern consistent with good stability, but high assay variability. Factors such as dosage form should not affect the results. The regression results are shown in Table 2, and indicate a substantially lower shelf-life from the $L_1$ analysis. The least-squares analysis seems affected much more than the $L_1$ analysis by the nonmonotonicity of the decreasing trend over time, shifting the least-squares line upwards.

**Table 1**
Stability dataset

| $x$ | $y$ |
|---|---|
| 0 | 0.3000 |
| 3 | 0.3003 |
| 6 | 0.3062 |
| 12 | 0.2930 |
| 18 | 0.3092 |
| 24 | 0.2760 |
| 30 | 0.2840 |
| 36 | 0.2890 |

When a nonlinear regression model is used, nonlinear programming methods are used to solve the $L_1$-norm problem. Sequential quadratic programming [8], Levenberg–Marquardt methods [9], and unconstrained minimization methods [10] are most often used, though currently available algorithms tend to converge slowly [11].

When the regression relationship is well determined, confidence bands around the line or curve can be used to determine shelf-life or expiration date. When the line or curve is not well estimated, as in the least-squares case with outliers or a large variability in response, the confidence bands can be very wide or oddly shaped rendering them meaningless for in-

**Table 2**
Comparison of regression methods

| Parameter estimate | Least-squares (SE) | $L_1$ (SE)* |
|---|---|---|
| intercept, $\hat{\beta}_0$ | 0.303414 (0.005592) | 0.302111 (0.005812) |
| slope, $\hat{\beta}_1$ | −0.0005397 (0.0002760) | −0.0006037 (0.0002864) |
| shelf-life† | 61.9 months | 53.2 months |

*Bootstrap standard errors based on $B = 2000$ replications.
†For a minimum acceptable assay value of 0.27.

ferential purposes [12, 13]. In these cases, a robust method such as $L_1$ can be used, and the shelf-life determined from the fit of the appropriate model. Model selection is an important issue, in [13] it is clearly demonstrated that segmented, nonlinear, and polynomial models can accurately characterize stability profiles. Either $L_1$ or least-squares methods can be used to estimate the fits of these models, with shelf-life obtained directly from the curve.

For issues involving the pooling of multiple batches, analysis of covariance models to test the equality of batch lines are often used. If pooling is not allowed (batch lines unequal), then the most conservative estimate of shelf-life among the batches is taken to give the overall shelf-life. $L_1$ techniques for analysis of variance and covariance can be used for these procedures [14].

Since stability data often have unusual observations, including those that do not follow the overall trend, it is especially important to use a method that is resistant to outliers. Since the $L_1$ and other robust techniques are becoming increasingly available, their use in stability studies can be expected to grow. Robust methods are usually more computationally intensive than conventional methods. With modern computational facilities, this should be a minor factor in the application of the methodology.

*Acknowledgement* — The author wishes to thank the BIPI Document Processing Group for their expert word processing of the manuscript.

## References

[1] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel, *Robust Statistics*. John Wiley, New York (1986).

[2] I. Barrodale and F.D.K. Roberts, in *Nonlinear Programming* (J.B. Rosen, O.L. Mangasarian and K. Ritter, Eds), pp. 447–463. Academic Press, New York (1970).

[3] I. Barrodale and F.D.K. Roberts, *Commun. Assoc. Comp. Mach.* **17**, 319–320 (1974).

[4] SAS Institute, Inc., *SUGI Supplemental Library User's Guide, Version 5 Edition*, pp. 257–260. SAS Institute, Inc., Cary, NC (1986).

[5] SAS Institute, Inc., *SAS/IML Software: Usage and Reference, Version 6, First Edition*, pp. 146–150. SAS Institute, Inc., Cary, NC (1989).

[6] K. Cogger, University of Kansas (1979) cited in: T.E. Dielman and R.C. Pfaffenberger, *Commun. Stat. — Comp. Sim.* **17**, 843–856 (1988).

[7] B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans*. SIAM, Philadelphia (1982).

[8] R.H. Bartels and A.R. Conn, in *Numerical Analysis: Lecture Notes in Mathematics* (J.P. Hennart, Ed.), pp. 48–58. Springer-Verlag, New York (1982).

[9] R.A. McLean and G.A. Watson, in *Numerical Methods of Approximation Theory* (L. Collatz, G. Meinardus and H. Werner, Eds), pp. 169–183. Birkhäuser Verlag, Basel (1980).

[10] R.A. El-Attar, M. Vidyasagar and S.R.K. Dutta, *SIAM J. Num. Anal.* **16**, 70–86 (1979).

[11] R. Gonin and A.H. Money, *Nonlinear* $L_p$*-Norm Estimation*. Marcel Dekker, New York (1989).

[12] N.R. Draper and H. Smith, *Applied Regression Analysis*, 2nd edn, pp. 47–51. John Wiley, New York (1981).

[13] J.T. Carstensen, in *Proceedings of Stability Guidelines for Testing Pharmaceutical Products: Issues and Alternatives*, 11–12 Dec. 1989, Arlington, VA, AAPS/FDA.

[14] J.W. McKean and R.M. Schrader, in *Statistical Data Analysis Based on the* $L_1$*-Norm and Related Methods* (Y. Dodge, Ed.), pp. 297–305. North-Holland, Amsterdam (1987).

[*Received for review 30 October 1992; revised manuscript received 10 February 1993*]