Stat Methods Appl (2009) 18:1–11
DOI 10.1007/s10260-008-0100-1

ORIGINAL ARTICLE

Bounds for how much influence an observation can have

Ingram Olkin · Adi Raveh

Accepted: 16 April 2008 / Published online: 17 May 2008
© Springer-Verlag 2008

Abstract That outliers or influential observations can affect the results in a regression is well-known. But it is not clear how much influence a specific observation can have on other statistics. In time series, especially in predictive situations, the effect of additional observations is of singular importance. We here examine bounds for the effect of an additional observation on the mean, variance, Mahalanobis distance, product moment correlation, and coefficients of linearity and monotonicity.

Keywords Variance · Mahalanobis distance · Product-moment correlation · Durbin–Watson statistic · Coefficient of linearity · Coefficient of monotonicity

1 Introduction

There is much research on the detection of influential observations in regression analysis. To cite but a few references, see Belsley et al. (1980), Atkinson et al. (2004), or Maronna et al. (2006). Indeed, a variety of statistics have been developed to detect particularly influential or aberrant observations in a set of data. These statistics have been incorporated in almost every software program. However the determination of bounds for various statistics such as percentiles, order statistics, confidence intervals, and so on, can be very helpful in anticipating an effect. In time series, especially in predictive models, the effect of additional observations is of singular importance. An understanding of the effect of a single observation can be useful when drawing

I. Olkin (✉)
Stanford University, Stanford, CA, USA
e-mail: iolkin@stat.stanford.edu

A. Raveh
The Hebrew University of Jerusalem, Jerusalem, Israel
e-mail: msraveh@mscc.huji.ac.il

conclusions in stages. For example, a change in public policy can create an influx of new observations. When the American Diabetes Association lowered the threshold from 125 to 100 as denoting a potential diabetes problem, the status of individuals with measures between 100 and 125 was altered from "no problem" to "potential problem".

In an educational context, the federal program known as "No Child Left Behind" stipulates that schools must show improvement in order to qualify for federal funds. Furthermore, schools are required to accept transfer students who are dissatisfied with their current school. The transfer of students from one district to another can create a change in the achievement data.

These changes suggest an investigation of the effect that additional observations can have on various statistics or on confidence intervals that are reported. The present study aims at providing some insights in the change effected by the addition of one observation on the mean, variance, correlation, Mahalanobis distance, and in several regression models. Note that the results for the deletion of an observation can also be obtained from the present results. Although the results obtained can be derived from other types of analysis, the fact that we have explicit expressions helps to interpret the data.

The general topic of bounds for how deviant an observation can be, and its effect, has a long history. For a sample $x_1, \ldots, x_n$ with mean $\bar{x}$ and variance $\upsilon = \sum(x_i - \bar{x})^2/n$, the inequality $(x_j - \bar{x})/\sqrt{\upsilon} \leq \sqrt{n-1}$, $j = 1, \ldots, n$, bounds the maximum deviation for any single observation in terms of the standard deviation. This inequality has been traced to Laguerre (1880), and later to Thompson (1935). However, it was rediscovered by Samuelson (1968), and is often referred to as Samuelson's inequality or the Laguerre–Samuelson inequality. The publication of Samuelson's paper stimulated a host of papers. An extensive history is provided by Jensen and Styan (1999) that lists over 250 references. Several recent papers are by Trenkler and Puntanen (2005), Joarder and Laradji (2005), and Goroncy and Rychlik (2006). Niezgoda (2007) provides generalizations in terms of group theory, many of which are based on convexity and majorization.

2 Univariate statistics

Let $x_1, x_2, \ldots, x_n$ be a random sample, and denote by z an additional observation from the same distribution. Let

$$\bar{x}_n = \sum x_i/n, \quad \upsilon_n = \sum(x_i - \bar{x})^2/n \tag{2.1}$$

denote the sample mean and sample variance. For simplicity of notation, we define the sample variance as in (2.1), and note that the usual unbiased estimator is $s_n^2 = n\upsilon_n/(n-1)$.

We first examine the effect of a single observation on the mean and the variance. For a new observation z to change the mean from $\bar{x}_n$ to a targeted average, A (for average), we require that

$$\frac{\sum_1^n x_i + z}{n+1} = A, \tag{2.2}$$

so that

$$z = n(A - \bar{x}_n) + A. \tag{2.3}$$

It is intuitive and readily seen that any target value A can be achieved by a z as in (2.3), and that to maintain the same sample mean $A = \bar{x}_n$ requires that $z = \bar{x}_n$.

If the population is normal with mean μ and variance σ^2, and z is chosen by (2.3), then

$$P\{z > t\} = P\left\{\bar{x}_n < \frac{(n+1)A - t}{n}\right\} = \Phi\left(\frac{(n+1)A - n\mu - t}{\sigma\sqrt{n}}\right), \tag{2.4}$$

where Φ is the cumulative distribution of the standard normal distribution.

The effect of an additional observation on the median clearly will be minimal. If $n = 2m + 1$ and $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$ are the ordered sample values, then the median is $x_{(m)}$, and an additional observation yields a median in the interval $[x_{(m-1)}, x_{(m+1)}]$. If $n = 2m$, then an additional observation yields a median in the interval $[x_{(m)}, x_{(m+1)}]$.

It is generally desirable to have a small variance. This raises the question of how much influence a single observation can have in decreasing the variance. Let z denote the additional observation, then

$$\begin{aligned} \upsilon_{n+1} &= \left[\sum_1^n x_i^2 + z^2 - (n+1)\left(\frac{n\bar{x}_n + z}{n+1}\right)^2\right] \Big/ (n+1) \\ &= \frac{n}{n+1}\upsilon_n + \frac{n}{(n+1)^2}(\bar{x}_n - z)^2, \end{aligned} \tag{2.5}$$

or equivalently,

$$s_{n+1}^2 = \frac{n-1}{n}s_n^2 + \frac{1}{(n+1)}(\bar{x}_n - z)^2. \tag{2.6}$$

Consequently, the variance (2.5) is minimized if the additional observation z is at the mean $\bar{x}_n$, in which case the amount of decrease is the proportion $n/(n+1)$ of the variance. A similar calculation with the addition of k observations $z_1, \ldots, z_k$ yields

$$\upsilon_{n+k} = \frac{n}{n+k}\upsilon_n + \frac{k}{n+k}\upsilon_k(z) + \frac{nk}{(n+k)^2}(\bar{z} - \bar{x}_n)^2, \tag{2.7}$$

where $\bar{z} = \sum z_i/k$, $\upsilon_k(z) = \sum_1^k (z_i - \bar{z})^2/k$, and $\upsilon_1(z) = 0$. Clearly, the variance is minimized by choosing $\bar{z} = \bar{x}_n$.

Table 1 Width of 95% confidence interval for mean as a proportion of the sample variance

n	t_n	t_{n-1}	$\frac{t_{n-1}}{\sqrt{n}}$	$t_n\sqrt{(n-1)/n(n+1)}$
10	2.228	2.262	0.715	0.637
15	2.131	2.145	0.554	0.515
20	2.086	2.093	0.468	0.444
30	2.042	2.045	0.373	0.361

Finally, notice that if

$$\bar{z} = \bar{x}_n \pm \sqrt{\upsilon_n(n+1)/n}$$

in (2.7), then $\upsilon_{n+1} = \upsilon_n$.

When the population is $\mathcal{N}(\mu, \sigma^2)$ and σ is known, then a new observation at the mean does not alter the mean, and has the effect of changing the confidence interval for μ from

$$(\bar{x}_n \pm c\,\sigma/\sqrt{n}) \quad \text{to} \quad (\bar{x}_n \pm c\,\sigma/\sqrt{n+1}), \tag{2.8}$$

which has a small effect for moderate n.

When σ is unknown and $z = \bar{x}_n$ then $s_{n+1} = \sqrt{(n-1)/n}\, s_n$ and the confidence interval for μ changes from

$$\left(\bar{x}_n \pm t_{n-1}\frac{s_n}{\sqrt{n}}\right) \quad \text{to} \quad \left(\bar{x}_n \pm t_n\frac{s_{n+1}}{\sqrt{n+1}}\right) = \left(\bar{x}_n \pm t_n\sqrt{\frac{n-1}{n}}\frac{s_n}{\sqrt{n+1}}\right), \tag{2.9}$$

so the difference in the width of the confidence interval depends on the relation between $t_{n-1}/\sqrt{n}$ and $\sqrt{\frac{n-1}{n}}t_n/\sqrt{n+1}$, where t_m is obtained from the upper tail of the t-distribution with m degrees of freedom.

To illustrate this difference, Table 1 provides comparisons between $t_{n-1}/\sqrt{n}$ and $t_n\sqrt{(n-1)/n(n+1)}$ for the case of a 95% confidence interval. For n large, the two confidence intervals in (2,9) become close to one another. The difference in the confidence intervals is greatest for small sample sizes, or for heterogeneous data.

As an example, suppose that for a random sample of 10 men, the mean height is $\bar{x} = 68.2$ inches, with a standard deviation of $s = 2.9$ inches. A 95% confidence interval for the mean μ (assuming an underlying normal distribution) is $68.2 \pm 2.262(2.9)/\sqrt{10}$ or $(66.13, 70.27)$. With an additional observation taken at the mean, the confidence interval becomes $68.2 \pm 2.228(2.9)\sqrt{10/(11 \times 12)}$ or $(66.43, 69.97)$, which is a 17% decrease in the length of the interval.

3 Multivariate statistics

In the multivariate case we observe a sample $(x_{i1}, x_{i2}, \ldots, x_{ip})$, $i = 1, \ldots, n$ from a p-variate normal distribution $\mathcal{N}(\mu, \Sigma)$.

The confidence ellipsoid for the mean vector μ based on n observations is

$$(\bar{x}_n - \mu)' S_n^{-1} (\bar{x}_n - \mu) \leq \frac{(n-1)p}{n(n-p)} F_{p,n-p} \equiv c(n, p). \tag{3.1}$$

Suppose that we add a new p-dimensional observation $z = \bar{x}_n$ so that the mean vector does not change. Then the covariance matrix changes by a factor of $(n-1)/n$, that is, the covariance matrix S_{n+1} based on $n+1$ observations is equal to $[(n-1)/n]S_n$. To see this, note that

$$nV_n = X\left(I - \frac{J_n}{n}\right)X' = XX' - n\bar{x}_n\bar{x}_n',$$

where $X = (x_{ij})$ is the $p \times n$ data matrix, $J_n = e_n e_n'$, and $e_n' = (1, \ldots, 1)$. As before, we distinguish the sample covariance matrix by V_n, which has n in the denominator, and S_n which has $(n-1)$ in the denominator. Clearly, $S_n = [n/(n-1)]V_n$. With the new observation z,

$$\begin{aligned}
(n+1)V_{n+1} &= (X, z)\left(I_{n+1} - \frac{J_{n+1}}{n+1}\right)(X, z)' \\
&= XX' + zz' - \frac{(X, z)e_{n+1}e_{n+1}'(X, z)'}{n+1} \\
&= (nV_n + n\bar{x}_n\bar{x}_n') + zz' - \frac{(n\bar{x}_n + z)(n\bar{x}_n + z)'}{n+1} \\
&= nV_n + \frac{n}{n+1}(\bar{x}_n - z)(\bar{x}_n - z)'.
\end{aligned}$$

The choice $z = \bar{x}_n$ "minimizes" the covariance matrix in the Loewner ordering. (For a discussion of this ordering, see, e.g., Marshall and Olkin 1979, p 463 or Zhan 2002, p 1.) Consequently, $V_{n+1} = nV_n/(n+1)$, and the confidence ellipsoid becomes

$$(\bar{x}_n - \mu)' S_{n+1}^{-1} (\bar{x}_n - \mu) \leq \frac{np}{(n+1)(n-p+1)} F_{p,n-p+1} \equiv c(n+1, p). \tag{3.2}$$

To compare the two confidence ellipsoids, Table 2 gives values of $c(n, p)$ and $c(n+1, p)$. As expected, the confidence ellipsoid is somewhat smaller with the choice of the additional observation at the mean. It is surprising that, with more variables, the decrease can be considerable for small sample sizes.

Table 2 95% confidence ellipsoid values for $p = 3$ and $p = 8$ variables, and sample sizes $n = 15$ and $n = 30$

p	n	$c(n, p)$	$c(n+1, p)$
3	15	0.80	0.74
	30	0.40	0.32
8	15	3.99	3.23
	30	0.84	0.80

4 Simple regression

Consider a sample (x_i, y_i), $i = 1, \ldots, n$ to which we fit a simple linear regression

$$\hat{y} = b_0 + b_1 x, \tag{4.1}$$

where

$$b_1 = \sum_1^n (x_i - \bar{x}_n)(y_i - \bar{y}_n) \Big/ \sum (x_i - \bar{x}_n)^2, \tag{4.2}$$

$$b_0 = \bar{y}_n - b_1 \bar{x}_n. \tag{4.3}$$

If we add a new observation at $(\bar{x}, \bar{y})$ then the regression line remains the same, and we do not expect a large effect on the analysis.

The standard error of b_1 is $s_{y.x}/s_x\sqrt{n-1}$, where $s_{y.x}^2 = \frac{(n-1)}{(n-2)} s_y^2 (1 - r_n^2)$. The additional observation at $(\bar{x}, \bar{y})$ does not alter $(n-1)s_x$, $(n-1)s_y$, or r_n, so that the ratios of standard errors for n and $n+1$ observations is $\sqrt{(n-1)/(n-2)}$, which has a minimal effect.

5 Simple correlation

Simple correlation analysis is a basic analysis in both the social and physical sciences. Here we observe a sample (x_i, y_i), $i = 1, \ldots, n$ for which the data has been standardized, so the means $\bar{x}_n$ and $\bar{y}_n$ are equal to 0 and the variances s_x^2 and s_y^2 are equal to 1. The "best" predictor $\hat{y}_{n+1}$ for a new observation x_{n+1} is

$$\hat{y}_{n+1} = r_n x_{n+1}, \tag{5.1}$$

where r_n is the Pearson product–moment correlation based on a sample of size n. With the choice of $(x_{n+1}, \hat{y}_{n+1})$ as in (5.1), the absolute value of the correlation r_{n+1} increases so that

$$|r_{n+1}| \geq |r_n|. \tag{5.2}$$

This is intuitive in that by adding a point for which the y-value is on the regression line we create a better fit as measured by the correlation.

We now focus on choosing a y-value so that $r_{n+1} = r_n$ (for $|r_n| \neq 0$ or 1). When the growth of x_n is such that $x_{n+1}^2/(n+1) \to 0$, the required y-value, y_{n+1}, is approximately

$$y_{n+1} \approx \left(\frac{1 \pm \sqrt{1-r_n^2}}{r_n}\right) x_{n+1}, \tag{5.3}$$

which is independent of n. Using this value of y_{n+1} for prediction is a compromise between classical regression and the arithmetic mean. (See the Appendix for the derivation.) When $r > 0$, because $[1 - \sqrt{1-r^2}]/r \leq r$, the predicted value y_{n+1} is dampened as n increases. However, because $[1 + \sqrt{1-r^2}]/r \geq 1$, y_{n+1} grows as n increases.

From the exact solution given by (6.15), if x_n grows such that $x_{n+1}/\sqrt{n+1} \to b \neq 0$, then

$$y_{n+1} = x_{n+1} \left\{ \frac{1 \pm \sqrt{1-r^2}\sqrt{1+b^2}}{r\left[1 - \frac{1-r^2}{r^2} b^2\right]} \right\} \equiv B x_{n+1}. \tag{5.4}$$

Here $B = 1$ for $r = 1$; $B = -1$ for $r = -1$; and $\lim_{r \to 0} B = 0$. None of these cases leads to an oscillating series.

6 Time series

In a time series, z_i denotes the observation at time t_i, $i = 1, \ldots, n$, from which a concern is to determine the relation between z_{i+1} and z_i. A graphical display might show linearity which is equivalent to constant first differences. A weaker comparison between successive observations is that the first differences are positive, in which case the data exhibits monotonicity. Clearly, there may be other comparisons. We here discuss linearity and monotonicity.

6.1 Coefficient of linearity

Consider a sequence $z_1, \ldots, z_n$, as in a time series, for which we wish to provide a measure of linearity. Using the fact that linearity is equivalent to zero second differences, Raveh and Schwarz (1996) proposed a coefficient of linearity:

$$\alpha_n = \frac{1}{2} + \frac{\sum_1^{n-1} d_i d_{i+1}}{\sum_1^{n-1} d_i^2 + \sum_2^n d_i^2}, \tag{6.1}$$

where

$$d_i = z_{i+1} - z_i \tag{6.2}$$

are first differences. The coefficient α_n is equal to 1 when the series is linear, and $\alpha_n = 0$ for a series of period 2.

The coefficient of linearity is similar to the Durbin–Watson statistic which measures the departure from an autocorrelated series.

Define

$$r_n = 2\alpha_n - 1 = \frac{\sum_1^{n-1} d_i d_{i+1}}{\frac{1}{2}\left(\sum_1^{n-1} d_i^2 + \sum_2^n d_i^2\right)}, \tag{6.3}$$

so that the similarity to the Durbin–Watson statistic becomes more apparent. In this form, r_n is approximately the correlation between the vectors $(d_1, \ldots, d_{n-1})$ and $(d_2, \ldots, d_n)$, in which the arithmetic mean in the denominator of (6.3) replaces the geometric mean. It follows from (6.2) and (6.3) that r_n and α_n are invariant with respect to change of scale and addition of constants.

To maintain the same correlation, or coefficient of linearity, with the addition of a new observation z_{n+1}, we require

$$\begin{aligned} r_n &= \frac{\sum_1^{n-1} d_i d_{i+1}}{\frac{1}{2}\left(\sum_1^{n-1} d_i^2 + \sum_2^n d_i^2\right)} \\ &= \frac{\sum_1^{n-1} d_i d_{i+1} + d_n d_{n+1}}{\frac{1}{2}\left(\sum_1^{n-1} d_i^2 + \sum_2^n d_i^2 + d_n^2 + d_{n+1}^2\right)} = r_{n+1}. \end{aligned} \tag{6.4}$$

Denote by A the numerator of the left-hand side of (6.4), and by B the denominator, in which case (6.4) can be written as

$$r = \frac{A}{B} = \frac{A + d_n d_{n+1}}{B + \frac{1}{2}\left(d_n^2 d_{n+1}^2\right)},$$

where r is the common value. It follows that

$$r = \frac{d_n d_{n+1}}{\frac{1}{2}\left(d_n^2 + d_{n+1}^2\right)}. \tag{6.5}$$

For positive d_n and d_{n+1}, this is the ratio of the geometric to the arithmetic mean of d_n^2 and d_{n+1}^2. This leads to the equation

$$d_{n+1} = C\, d_n, \tag{6.6}$$

where $C = C_+ = [1 + \sqrt{1 - r^2}]/r$ or $C = C_- = [1 - \sqrt{1 - r^2}]/r$. For all $r > 0$, $C_+ > 1$ and $C_- < 1$, so that C_+ leads to an explosive series, whereas C_- leads to a dampened series. Because $d_{n+1} = z_{n+1} - z_n$, $d_n = z_n - z_{n-1}$,

$$z_{n+1} = (C + 1)z_n - C z_{n-1}. \tag{6.7}$$

When $r = 1$, $A_+ = A_- = 1$ and $z_{n+1} = 2z_n - z_{n-1}$; this case represents linearity. When $r = 0$, $A_+ \to \infty$ and $A_- \to 0$, so that in the second case, $z_{n+1} = z_n$ which is the case of random walk. When $r = -1$, $A_+ = A_{-1} = -1$ and $z_{n+1} = z_{n-1}$; this case represents an oscillating series $a, b, a, b, \ldots$

6.2 Coefficient of monotonicity

The notion of nonmetric measures such as monotonicity date to Guttman (1968, 1977) in which the aim was to quantify the implication that if X increases then Y increases. Earlier, similar measures of association in the context of contingency tables were studied by Goodman and Kruskal (1954). Monotonicity and association have similarities, and actually coincide in special cases. Raveh (1986) provides a discussion of various measures of monotonicity that relate to measures of association; see also Raveh (1989).

A series $z_1, \ldots, z_n$ is defined to be *positive monotone* if $z_i > z_j$ for all $i > j$. A measure μ_n of positive monotonicity is

$$\mu_n = \frac{\sum_{i>j}^{n} \left(z_i - z_j\right)}{\sum_{i>j}^{n} |z_i - z_j|}. \tag{6.8}$$

If $z_1 \le z_2 \le \cdots \le z_n$, then there are no reversals and $\mu_n = 1$, whereas if $z_1 \ge z_2 \ge \cdots \ge z_n$, then $\mu_n = -1$.

With an additional point z_{n+1}, the equality $\mu_{n+1} = \mu_n \equiv \mu$ implies that

$$\mu = \frac{\displaystyle\sum_{k=1}^{n} (z_{n+1} - z_k)}{\displaystyle\sum_{k=1}^{n} |z_{n+1} - z_k|}. \tag{6.9}$$

Let T_+ denote the sum of all positive differences and T_- the sum of all negative differences. Then

$$\mu = (T_+ - T_\_)/(T_+ + T_\_) \tag{6.10}$$

so that

$$T_\_ = \frac{1-\mu}{1+\mu}\, T_+. \tag{6.11}$$

This means that z_{n+1} must be chosen judicially. Because $-1 \le \mu \le 1$, and because of symmetry between T_+ and T_-, we need only consider $0 \le \mu \le 1$. If $\mu = 0$, then $z_{n+1} = \bar{z}_n$ yields $T_\_ = T_+$. The case $\mu = 1$ is achieved by any value of $z_{n+1} \ge \max\{z_1, \ldots, z_n\}$. For any $0 < \mu < 1$, z_{n+1} will be a weighted average of the means of the positive and negative sums of differences.

Appendix

The following provides a discussion of the approximation of (5.3). To simplify the algebra, without loss of generality, normalize by setting $\bar{x}_n = \bar{y}_n = 0$, $v_n(x) = v_n(y) = 1$ so that $\sum_1^n x_i^2 = \sum_1^n y_i^2 = n$ and $\sum_1^n x_i y_i = nr_n$. Then $\bar{y}_{n+1} = y_{n+1}/(n+1)$, and by a straightforward evaluation we obtain

$$
\begin{aligned}
(n+1)v_n(y) &= \sum_1^{n+1}(y_i - \bar{y}_{n+1})^2 = n\left\{1 + \frac{y_{n+1}^2}{n+1}\right\},\\
(n+1)v_n(x) &= \sum_1^{n+1}(x_i - \bar{x}_{n+1})^2 = n\left\{1 + \frac{x_{n+1}^2}{n+1}\right\},\\
\sum_1^{n+1}(x_i - \bar{x}_{n+1})&(y_i - \bar{y}_{n+1}) = nr_n + \frac{n}{n+1}x_{n+1}y_{n+1},
\end{aligned}
\tag{6.12}
$$

so that $r_{n+1} = r_n \equiv r$ yields the equation

$$
r = \frac{r + x_{n+1}y_{n+1}/(n+1)}{\sqrt{1 + y_{n+1}^2/(n+1)}\sqrt{1 + x_{n+1}^2/(n+1)}}. \tag{6.13}
$$

Completing the square leads to the quadratic equation

$$
y_{n+1}^2[x_{n+1}^2(1-r^2) - (n+1)r^2] + y_{n+1}[2(n+1)rx_{n+1}] - (n+1)r^2x_{n+1}^2 = 0, \tag{6.14}
$$

from which we obtain

$$
y_{n+1} = x_{n+1}\left\{\frac{1 \pm \sqrt{1-r^2}\sqrt{1 + \frac{x_{n+1}^2}{n+1}}}{r\left[1 - \frac{1-r^2}{r^2}\frac{x_{n+1}^2}{n+1}\right]}\right\}. \tag{6.15}
$$

If the growth in the x's such that $x_{n+1}/(\sqrt{n+1}) \to 0$ then as $n \to \infty$

$$
y_{n+1} = x_{n+1}\left[\frac{1 \pm \sqrt{1-r^2}}{r}\right]. \tag{6.16}
$$

Acknowledgments The authors wish to acknowledge the support of their work in part by the Recanati Foundation.

References

Atkinson AC, Riani M, Cerioli A (2004) A exploring multivariate data with the forward search. Springer, New York

Belsley DA, Kuh E, Welsch RE (1980) Regression diagnostics: identifying influential data and sources of collinearity. Wiley, New York

Goodman LA, Kruskal WH (1954) Measures of association for cross classifications. J Am Stat Assoc 49:732–764

Goroncy A, Rychlik T (2006) How deviant can you be? The complete solution. Math Inequalities Appl 9:633–647

Guttman L (1968) A general nonmetric technique for finding the smallest coordinate space for a configuration of points. Psychometrika 33:469–506

Guttman L (1977) What is not what in statistics. Statistician 26:81–107

Jensen ST, Styan GPH (1999) GPH Some comments and a bibliography on the Laguerre–Samuelson inequality with extensions and applications in statistics and matrix theory. In: Rassias TM, Srivastava HM (eds) Analytic and geometric inequalities and applications. Kluwer, Dordrecht pp 151–182

Joarder AH, Laradji A (2005) Algebraic inequalities for measures of dispersion. J Prob Stat Sci 3:317–326

Laguerre EN (1880) Sur une méthode pour obtenir par approximation les racines d'une éq algébrique qui a toutes ses racines réelles. Nouv Ann Math (Paris) 2^e Serie 19:161–171; 193–202

Maronna RA, Martin RD, Yohai VJ (2006) Robust statistics: theory and methods. Wiley, Chichester

Marshall AW, Olkin I (1979) Inequalities: theory of majorization and its applications. Academic Press, New York

Niezgoda M (2007) Laguerre–Samuelson type inequalities. Linear Algebra Appl 412:574–581

Raveh A (1986) On measures of monotone association. Am Stat 40:117–123

Raveh A (1989) A new version of structural persistence in prediction. J Time Ser Anal 10:77–93

Raveh A, Schwarz G (1996) Simple diagnostic tools for investigation linear trends in time series. Commun Stat Theory Methods 25:1505–1524

Samuelson PA (1968) How deviant can you be? J Am Stat Assoc 63:1522–1525

Thompson WR (1935) On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. Ann Math Stat 6:214–219

Trenkler G, Puntanen G. (2005) A multivariate version of Samuelson's inequality. Linear Algebra Appl 410:143–149

Zhan X (2002) Matrix inequalities. Lecture Notes in Mathematics, vol 1790. Springer, Berlin