Sensitivity Coefficients for the Effects of Errors in the Independent Variables in a Linear Regression.

by

G. W. Stewart

Abstract

This paper is concerned with errors in the observed values of the independent variables of a linear regression. We propose sensitivity coefficients to measure the effects of these errors and show that they can easily be computed from quantities ordinarily calculated in performing the regression.

This work was supported in part by the Office of Naval Research under Contract No. N00014-76-C-0391.
Sensitivity Coefficients
for the Effects of Errors in
the Independent Variables in
a Linear Regression

G. W. Stewart

1. Introduction

In this paper we shall be concerned with the regression problem

\[
\text{minimize } \|y - X\beta\|^2,
\]

where \(X\) is an \(n \times p\) matrix of rank \(p\), \(y\) is an \(n\)-vector, and \(\| \cdot \|\)
denotes the usual Euclidean vector norm.* The problem has the unique
solution

\[
\beta = X^+ y
\]

where \(X^+ = (X'X)^{-1}X'\) is the pseudo-inverse of \(X\).

Although classical regression theory concerns itself with the sta-
tistical analysis of errors in the vector \(y\), it frequently happens that
the design matrix \(X\) is itself contaminated with errors, so that one
is effectively working with a perturbed matrix \(X + \epsilon\). For example, the
columns of \(X\) may be measured by means of some instrument for which the
originator of the problem can only give crude error estimates.
In this case the data analyst is faced with the problem of deciding when
the effects of the errors can be ignored. The problem is especially

* In the sequel \(\| \cdot \|\) will also denote the spectral matrix norm defined by
\[\|A\| = \sup \{\|Ax\| : \|x\| = 1\}.\] See [3] for details.
critical in general purpose regression routines, where it is desirable to provide the user with a set of easily interpretable numbers that indicate the magnitude of the effects of the errors.

A partial solution is provided by the perturbation theory for the regression problem (for a survey of this theory see [4]). This theory bounds perturbations in $\beta$ in terms of $\|\varepsilon\|$ and of the "condition number" $\kappa = \|X\|\|X^+\|$. Although the results of this theory shed considerable light on the behavior of regression problems under perturbations in $X$, they are unsatisfactory in practice for two reasons. First they bound only the norm of the perturbation in $\beta$, so that a large perturbation in one component can conceal the fact that the others have small perturbations. More important, they are not scale invariant; changing the scale of the columns of $X$ will change $\kappa$, even though the statistical problem is essentially unaltered. This phenomenon makes the results quite difficult to interpret.

Taking a different approach, Beaton, Rubin, and Barone [1] have derived measures of sensitivity that to some extent answer the above objections. However, it is assumed that the errors are unbiased and $n$ is large. Swindel and Bauer [5] have derived a useful bound for the relative bias in $\hat{\beta}$, which measures the relative effects of perturbations in $X$ compared to the usual statistical errors in $y$. In this paper we shall derive coefficients $\gamma_{ij}$ that measure the sensitivity of $\beta_i$ to changes in column $j$ of $X$. Specifically $\gamma_{ij}$ is the norm of the Frechet
derivative of $\beta_i$ regarded as a function of the $j$-th column of $X$.

If $\varepsilon$ is the norm of the perturbation of the $i$-th column of $X$, then $\gamma_{ij}\varepsilon$ will be an asymptotic bound on the perturbation induced in $\beta_j$.

2. Derivation of the Coefficients

Although it is in principle possible to calculate the required derivatives directly from the normal equations $(X'X)\beta = X'y$, we prefer to approach the problem through a first order perturbation theorem that is useful in its own right.

Theorem 2.1. In the notation of the last section, let $\beta$ be the solution of the regression problem (1.1), and let $r = y - X\beta$ be the corresponding residual vector. Let $E$ be an $n \times p$ matrix. If

$$\|X^+E\| < 1,$$

then $X + E$ has rank $p$ so that there is a unique solution $\tilde{\beta}$ of the problem

$$\text{minimize } \|y - (X+E)\tilde{\beta}\|^2.$$  

Moreover, as $E$ approaches zero

$$\tilde{\beta} = \beta - X E \beta + (X'X)^{-1}E' r + O(\|E\|^2).$$

Proof. For a proof that (2.1) implies that $X + E$ has full rank, see [4]. This implies that for all sufficiently small $E$, $\tilde{\beta}$ exists and is given by

Here, and throughout this note, the term asymptotic refers to behavior for small $\varepsilon$, not large $n$. 

Now it is well known [3] that if \( P \) is sufficiently small, then \( I + P \) is nonsingular and

\[
(I + P)^{-1} = I - P + O(\|P\|^2).
\]

It follows that

\[
\begin{aligned}
[(X+E)'(X+E)]^{-1} &= (X'X + E'X + X'E + E'E)^{-1} \\
&= [I - (X'X)^{-1}(X'E + E'X)](X'X)^{-1} + O(\|E\|^2).
\end{aligned}
\]

Hence

\[
\tilde{\beta} = (X'X)^{-1}X'\chi + (X'X)^{-1}E'\chi - (X'X)^{-1}X'E(X'X)^{-1}X'\chi \\
&\quad - (X'X)^{-1}E'X(X'X)^{-1}X'\chi + O(\|E\|^2)
\]

\[
= \beta - X^+E\chi + (X'X)^{-1}E'(X\chi).
\]

\[
= \beta - X^+E\chi + (X'X)^{-1}E^T E^{-1}.
\]

Theorem 2.1 immediately gives an expression for the Frechet derivative of \( \beta_1 \) regarded as a function of the \( j \)-th column of \( X \).

**Corollary 2.2.** Let \( \beta_1 = f_{ij}(x_j) \), where \( x_j \) denotes the \( j \)-th column of \( X \). Then

\[
\frac{Df_{ij}}{x_j} = -\beta_j e_{ij}^T + \frac{e_i'(X'X)^{-1}e_j}{e_j^T X^T}.
\]
Proof. The Frechet derivative $Df_{ij}$ is the unique row vector satisfying

$$f_{ij}(x_j + e) = \beta_i + Df_{ij}e + O(\|e\|^2).$$

Let $e_j$ denote the $j$-th unit vector. Then a perturbation $e$ in $x_j$ amounts to adding to $X$ the matrix $E = ee^T$. Hence from Theorem 2.1

$$f_{ij}(x_j + e) = e_i^T[\begin{bmatrix} \beta_i X \\ ee^T \end{bmatrix} + (X^TX)^{-1}(ee^T)] + O(\|e\|^2)$$

$$= \beta_i - \beta_j e_i^T X + e_i^T X e + o_i^T (XX)^{-1} e_i e + O(\|e\|^2)$$

$$= \beta_i - (\beta_j - e_i^T X e + o_i^T (XX)^{-1} e_i e + O(\|e\|^2),$$

which shows that $f'_{ij}$ defined by (2.2) satisfies (2.3).

Corollary 2.3. Let $C = (X^TX)^{-1}$. Then

$$\|Df_{ij}\| = \gamma_{ij} = \sqrt{\beta_j^2 c_{ii} + \|r\|^2 c_{ij}^2}.$$ 

Proof. The vector $r$ is orthogonal to the column space of $X$, which is the same as the row space of $X^T$. Hence $e_i^TX^T$ and $r$ are orthogonal, and

$$\|Df_{ij}\|^2 = |\beta_j|^2 \|e_iX^T\|^2 + c_{ij}^2 \|r\|^2.$$ 

The proof will be complete if we can show that $\|e_iX^T\|^2 = c_{ii}$. But

$$\|e_iX^T\|^2 = \|e_i^T (X^TX)^{-1} X'\|^2 = e_i^T (X^TX)^{-1} X'X (X^TX)^{-1} e_i = e_i^T (X^TX)^{-1} e_i = c_{ii}.$$
3. Applications

It should be observed that the numbers $\gamma_{ij}$ can be calculated from the quantities that are usually computed in the course of the regression. For example, if sweep techniques (e.g. see [2]) have been used to solve the problem, the numbers $c_{ij}$ will be available, as will $\|x\|^2$, since it is simply the residual sum of squares.

The results are readily interpretable. If a bound $\epsilon_j$ on the norm of the error in the $j$-th column is available, then the error $\delta_i$ induced in $\beta_i$ is asymptotically bounded by $\gamma_{ij}\epsilon_j$. If this number is too large, the problem requires further study. In interpreting the results it should always be borne in mind that $\gamma_{ij}\epsilon_j$ is a first order bound. A large value is a signal that something may be wrong, but if $\epsilon_j$ is so large that the first order approximation is not applicable, then the difficulties may turn out to be illusory.

A particularly attractive feature of the asymptotic bounds is that, since they deal with individual components of $\beta$, their interpretation is independent of the scaling of the columns of $X$. This is particularly apparent if the bounds are cast in terms of relative error in the form

$$\frac{|\delta_i|}{|\beta_i|} \leq \left( \frac{\gamma_{ij}\|x_j\|}{\|\beta_i\|} \right) \frac{\epsilon_j}{\|x_j\|} .$$

Now

$$\gamma_{ij}\|x_j\| \frac{1}{|\beta_i|} = \left[ \left( |\beta_j|^2\|x_j\|^2 \right) \left( \frac{c_{ij}}{|\beta_i|^2} \right) + (\|r\|^2) \left( \frac{|c_{ij}|^2\|x_j\|^2}{|\beta_i|^2} \right) \right]^{1/2} ,$$

(3.1)
and each parenthesized term in the right hand side of (3.1) is easily seen to be invariant under scaling of the columns of $X$. Indeed, in some applications where the $p_i$'s are known to be bounded away from zero it may be more appropriate to report $\gamma_{ij}\|x_j\|/|\beta_i|$ than $\gamma_{ij}$.

In deriving our bounds, we have used the Cauchy-Schwarz inequality $\|x^*y\| \leq \|x\|\|y\|$, an inequality which is usually pessimistic, since it must account for the worst case where $x$ and $y$ are dependent. If we are willing to assume more about the perturbation $\varepsilon$ in $X_j$, then we may be able to say more. For example, we have the following consequence of Corollary 2.3.

**Corollary 3.1.** Let $\varepsilon \in N(0,\sigma^2 I)$. Then $Df_{ij}^*\varepsilon$ is normally distributed with mean zero and standard deviation $\gamma_{ij}\sigma$.

**References**


This paper is concerned with errors in the observed values of the independent variables of a linear regression. We propose sensitivity coefficients to measure the effects of these errors and show that they can easily be computed from quantities ordinarily calculated in performing the regression.