TECHNICAL REPORT NO. 919

APRIL 1994

CONFIDENCE INTERVAL VERSUS REGIONS

by

Norman R. Draper and Irwin Guttman
CONFIDENCE BOUNDS VERSUS REGIONS

Norman R. Draper and I. Guttman
Statistics Department
University of Wisconsin
1210 West Dayton St.
Madison, WI 53706

Statistics Department
University of Buffalo
3435 Main St.
Buffalo, NY 14214

SUMMARY

In regression problems, individual confidence bands on the parameters typically are examined. Doing this avoids the difficulties of envisaging more accurate ellipsoidal confidence regions. We discuss the relationship between the two and suggest an easy volume calculation that can offer insight.
1. INTRODUCTION

Suppose we fit the \( p \)-parameter model \( Y = X\beta + \epsilon \), where \( X \) is \( n \) by \( p \) and \( \beta = (\beta_1, \beta_2, \ldots, \beta_p)' \) is \( p \) by \( 1 \), to a set of \( n \) regression data values. If it is appropriate to assume that \( \epsilon \sim N(0, I\sigma^2) \) we can use the \( t(v) \) and \( F(p, v) \) distributions where \( v = n - p \) to construct confidence statements for the elements of \( \beta \). Let \( b = (X'X)^{-1}X'Y \) be the least squares estimator of \( \beta \). Then the intervals

\[
    b_i \pm t(v, \alpha/2)(V_i s^2)^{1/2}, \quad i = 1, 2, \ldots, p, \tag{1.1}
\]

where \( (V_{ij}) = V = (X'X)^{-1} \), \( s^2 \) is the residual mean square based on \( v \) degrees of freedom and where \( t \equiv t(v, \alpha/2) \) is the percentage point of the \( t(v) \) distribution that leaves a probability \( \alpha/2 \) in the upper tail, define a \( p \)-dimensional rectangular region whose edges are the marginal confidence intervals (1.1). It is well known that this is not a proper joint confidence region. An ellipsoidal joint region is, however, defined by the interior of

\[
    (\beta - b)'X'X(\beta - b) = ps^2F(p, v, \theta), \tag{1.2}
\]

where \( F \equiv F(p, v, \theta) \) is the percentage point of the \( F(p, v) \) distribution which leaves a probability \( \theta \) in the upper tail.

In most regression books, the difference between (1.1) and (1.2) is explained, and a comparison diagram is often given for \( p = 2 \). Rarely is this comparison diagram drawn accurately for a specific example. Sometimes the two pieces are presented accurately, but not on the same diagram. We speculate that this is not due to any difficulty in drawing the diagram, but to lack of certainty about what values of \( \alpha \) and \( \theta \) to use in the comparison, if it is given accurately. Box, Hunter and Hunter (1978, p. 464), for example, use \( \alpha = \theta = 0.10 \) without explanation, while Draper and Smith (1981, p. 561) use \( \alpha = 0.05, \theta = 0.10 \) on the principal that if the \( b_i \)'s were independent (which is rarely true), the rectangle formed by (1.1) might be thought of as a \((1 - 0.05)^2 = 0.9025 \) region, roughly \( 1 - 0.10 \). What choices of \( \alpha \) and \( \theta \) are sensible? We discuss this in Section 2 and offer a useful volume comparison for the two regions.
2. COMPARING THE REGIONS

We shall compare the regions on the basis of their volumes. The volume of the rectangular region generated by (1.1) is obviously

\[ R \equiv 2^p p^p s^p (V_{11} V_{22} \cdots V_{pp})^{1/2}. \] 

(2.1)

The volume of the ellipsoidal region is given by a constant (depending on the dimension \( p \)) times the product of the semi-axial lengths. In terms of the notation in (1.2), and via manipulations used in the analogous method of canonical reduction, we can show that it is

\[ E \equiv \frac{\pi^{p/2}}{\Gamma\left(\frac{p}{2} + 1\right)} (ps^2 \bar{F})^{p/2} c_p^{1/2} (V_{11} V_{22} \cdots V_{pp})^{1/2}, \] 

(2.2)

where \( \pi = 3.14159 \), and where the gamma functions required satisfy, \( \Gamma(u) = (u - 1)\Gamma(u - 1) \), \( \Gamma(1) = 1 \), \( \Gamma\left(\frac{1}{2}\right) = \pi^{1/2} \) and where \( c_p \) is the determinant of a normalized form of \((X'X)^{-1}\) namely of \( \{V_{ij}/(V_{ii}V_{jj})^{1/2}\} \). Thus \( c_p \) is simply the determinant of the correlation matrix of \( b_1, b_2, \ldots, b_p \). The ratio of the volumes of the two regions is then

\[ \frac{E}{R} = \frac{\pi^{p/2} p^{p/2} \bar{F}^{p/2} c_p^{1/2}}{\Gamma\left(\frac{p}{2} + 1\right) 2^{p+p}}. \] 

(2.3)

We now link \( \alpha \) and \( \theta \) in the following manner. We specify that when \( c_p = 1 \), that is, when the \( b_i \)'s are uncorrelated, \( E = R \), and so their ratio is 1. This implies that, in the case where the major axes of the ellipsoid are parallel to the axes of the \( \beta \)'s, we would wish to link \( \alpha \) and \( \theta \) so that the ellipsoid and its approximating rectangular block have equal volume. This requires that

\[ F(p, v, \theta) = \left\{ \frac{\Gamma\left(\frac{p}{2} + 1\right) 2^{p+p}}{\pi^{p/2} p^{p/2}} \right\}^{2/p} \] 

\[ = \frac{4\{\Gamma\left(\frac{p}{2} + 1\right)\}^{2/p}}{\pi p} \{t(v, \frac{1}{2}\alpha)\}^{2} \] 

(2.4)

\[ = \frac{4\{\Gamma\left(\frac{p}{2} + 1\right)\}^{2/p}}{\pi p} \{F(1, v, \alpha)\}. \]

Note, as a check, that when \( p = 1 \), we get the obvious \( \theta = \alpha \), for the one parameter case.
Table 1 shows the results of using (2.4) to evaluate $F(p, v, \theta)$, and hence $\theta$, for $p = 2, 3, \ldots, 6$, and $v = 10, 20, 30, \text{ and } 60$. For $p = 2$, for example, the case most often depicted, and for uncorrelated $b_i$'s, we obtain an elliptical confidence region of size equal to the rectangular region based on separate 95% confidence intervals if we use a 91.3% confidence ellipse (approximately). Similar approximate results from other columns are 88.4% for $p = 3$, 86.0% for $p = 4$, 83.9% for $p = 5$, and 82.1% for $p = 6$.

The above calculations lead us to an easy way to assess how well the rectangular block can represent the correct ellipsoidal region in a regression for any value of $p$. Using any selected linked values $\alpha$ and $\theta$ that satisfy (2.4) we see that the right hand side of (2.3) reduces to $c_p^{1/2}$. This value gives the ratio of the volume of the ellipsoidal confidence region compared to the volume of the rectangular block. Note that $0 \leq c_p^{1/2} \leq 1$, the zero corresponding to linearly dependence in the $X$-columns and the 1 to an orthogonal set of $X$-columns. A relative volume calculation can be made from (2.3) even if an ellipsoid other than the one satisfying (2.4) is selected, of course. Note that our calculations can also be applied to the slightly different but closely related suggestions for confidence regions made by Weisberg (1985, pp. 97-99, Equations (4.26) and (4.27)).

**Example.** $p = 2$. Consider the straight line steam-data fit in Draper and Smith (1981, pp. 78 and 561). For this, $p = 2, v = 23$, the right hand side of (2.4) is $2.7241, 1 - \theta = 0.9132$, and the correlation between $b_0$ and $b_1$ is $r = V_{12}/V_{11}V_{22}^{1/2} = (-0.0073535)/[0.4267941(0.0001398)]^{1/2} = -(0.90628)^{1/2} = -0.992$. Thus $c_p^{1/2} = (1 - r^2)^{1/2} = 0.306$ and so the 91.3% ellipse covers about 30.6% of the area of the rectangle. Moreover, the high negative correlation indicates a diagonal upper-left-to-lower-right-lying ellipse. Figure 1 shows that these simple calculations describe the situation well. If the ellipse in Figure 1 were replaced by the 95% ellipse, which surrounds the 91.3% ellipse and juts out somewhat more at the upper left and lower right extremes, the area covered would increase to about 38.4% of the area of the rectangle.
Table 1. Values of $F(p, v, \theta)$ and $\theta$ corresponding to $\alpha = 0.05$, for selected $p$ and $v$, evaluated from (2.4). Of each pair, the top value is $F$, the lower value is $\theta$.

<table>
<thead>
<tr>
<th>$v$ = df of $t$</th>
<th>Value of $t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2.228</td>
<td>3.1602 2.5471 2.2346 2.0436 1.9141</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0864 0.1148 0.1379 0.1573 0.1739</td>
</tr>
<tr>
<td>20</td>
<td>2.086</td>
<td>2.7702 2.2328 1.9588 1.7914 1.6779</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0867 0.1158 0.1399 0.1604 0.1783</td>
</tr>
<tr>
<td>30</td>
<td>2.042</td>
<td>2.5516 2.1396 1.9771 1.7167 1.6079</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0868 0.1160 0.1404 0.1612 0.1794</td>
</tr>
<tr>
<td>60</td>
<td>2.000</td>
<td>2.5465 2.0525 1.8006 1.6468 1.5424</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0868 0.1161 0.1405 0.1616 0.1800</td>
</tr>
<tr>
<td>Approximate $1 - \theta$</td>
<td>0.913 0.884 0.860 0.839 0.821</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Individual 95% confidence bands and a 91.3% joint confidence region for the steam data (Draper and Smith, 1981, p. 9).
ACKNOWLEDGEMENTS

N. R. Draper gratefully acknowledges partial support from the Wisconsin Alumni Research Foundation through the University of Wisconsin Graduate School and from the National Security Agency via Grant No. MDA 904-92-H-3096. We thank Lisa H. Ying who drew the figure.

REFERENCES

