Table 3 Sample calculation of a residual sum of squares (RSS) value

(A) i	(B) x_0	(C) $x_1 = V$	(D) $x_2 = f$	(E) T (min.)	(F) $y = T^{(-0.2)}$	(G) $\hat{y} = \hat{T}^{(-0.2)}$	(H) $R = y - \hat{y}$	(I) R^2
1	1	700	0.0157	1.75	2.538	2.865	-0.327	0.1069
2	1	"	"	1.85	2.775	"	-0.090	0.0081
3	1	"	"	2.0	3.103	"	0.238	0.0566
4	1	"	"	2.2	3.497	"	0.632	0.3994
5	1	500	"	8.8	8.454	9.333	-0.879	0.7726
6	1	"	"	11.0	9.131	"	-0.202	0.0408
7	1	"	"	11.75	9.326	"	-0.007	0.0001
8	1	"	"	19.0	10.668	"	1.335	1.7822
9	1	800	0.01725	1.0	0.0	- 1.452	1.452	2.1083
10	1	"	"	0.9	- 0.510	"	.942	0.8874
11	1	"	"	0.74	- 1.488	"	-0.036	0.0013
12	1	"	"	0.66	- 2.077	"	-0.625	0.3906
13	1	700	"	1.0	0.0	1.782	-1.782	3.1755
14	1	"	"	1.2	0.858	"	-0.924	0.8538
15	1	"	"	1.5	1.867	"	0.085	0.0072
16	1	"	"	1.6	2.150	"	0.368	0.1354
17	1	600	"	2.35	3.765	5.016	-1.251	1.5650
18	1	"	"	2.65	4.245	"	-0.771	0.5944
19	1	"	"	3.0	4.728	"	-0.228	0.0829
20	1	"	"	3.6	5.417	"	0.401	0.1608
21	1	500	"	6.4	7.434	8.250	-0.816	0.6659
22	1	"	"	7.8	8.075	"	-0.175	0.0306
23	1	"	"	9.8	8.784	"	0.534	0.2852
24	1	"	"	16.5	10.287	"	2.037	4.1494
25	1	400	"	21.5	10.993	11.484	-0.491	0.2411
26	1	"	"	24.5	11.327	"	-0.157	0.0246
27	1	"	"	26.0	11.477	"	-0.007	0.0001
28	1	"	"	33.0	12.058	"	0.574	0.3295
29	1	600	0.022	1.2	0.858	1.699	-0.841	0.7073
30	1	"	"	1.5	1.867	"	0.168	0.0282
31	1	"	"	1.6	2.150	"	0.451	0.2034
32	1	"	"	1.6	2.150	"	0.451	0.2034
33	1	450	"	4.0	5.804	6.550	-0.746	0.5565
34	1	"	"	4.7	6.381	"	-0.169	0.0286
35	1	"	"	5.3	6.798	"	0.248	0.0615
36	1	"	"	6.0	7.219	"	0.669	0.4476

RSS = ΣR^2 = 21.093

APPENDIX B

Sample Calculation of a RSS Value

The basis of the RSS contour diagram shown in the body of the paper is a grid of RSS values, where each value is determined by fitting equation (7), i.e.,

$$E[T^{(\lambda)}] = \beta_0 + \beta_1 V^{\alpha_1} + \beta_2 f^{\alpha_2} \tag{28}$$

For instance, if $\alpha_2 = 1$, $\alpha_1 = 1$, and $\lambda = -0.2$, then

$$E[T^{(-0.2)}] = E\left[\frac{T^{-0.2} - 1.0}{(-0.2)(\dot{T})^{-1.2}}\right] = \beta_0 + \beta_1 V + \beta_2 f \tag{29}$$

The coefficients β_0, β_1, and β_2 are estimated by the method of least squares, i.e.,

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{30}$$

where $\mathbf{b}$ is the vector of estimated values of $\boldsymbol{\beta}$

$\mathbf{X}$ is the matrix of transformed independent variables
$\mathbf{X}'$ is the transpose of $\mathbf{X}$
$(\mathbf{X}'\mathbf{X})^{-1}$ is the inverse of $(\mathbf{X}'\mathbf{X})$

and $\mathbf{y}$ is the vector of transformed observations of tool life, i.e., $\mathbf{y} = T^{(\lambda)}$.

Using the carbide tool-life data [3] shown in Table 3, the calculation of the b's is illustrated as follows:

1 The matrix of independent variables $\mathbf{X}$ consists of the values given by columns (B), (C), and (D) in Table 3.

2 The observation vector $\mathbf{y}$ consists of the values given in column (F), and are calculated from

$$y_i = T_i^{(-0.2)} = \frac{T_i^{-0.2} - 1.0}{(-0.2)(\dot{T})^{-1.2}} \qquad i = 1, \ldots, n.$$

where T_i is the observed tool life shown in column (E), and where

$$\dot{T} = (T_1 \times T_2 \times T_3 \times \ldots \times T_n)^{\frac{1}{n}}, \text{ i.e.,}$$

$$\dot{T} = [(1.75)(1.85)(2.0)\ldots(6.0)]^{\frac{1}{36}} = 3.692$$

3 By equation (30)

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 36.466 \\ -\ \ 0.032 \\ -698.289 \end{bmatrix}$$

and the predicting equation is

$$\hat{y} = \hat{T}^{(-0.2)} = 36.466 - 0.032V - 698.289f \tag{31}$$

where $\hat{y}$ is the predicted transformed tool life.

The values of $\hat{y}$ at the various cutting conditions are calculated by equation (31) and are shown in column (G).

Finally, the residuals R_i,

$$R_i = (y_i - \hat{y}_i) \qquad i = 1, \ldots, n.$$

are given in column (H), the R_i^2 are given in column (I), and thus the value of the residual sum of squares is:

$$\text{RSS} = \sum_{i=1}^{n} R_i^2 = 21.093.$$

This particular RSS is shown in Fig. 6($\alpha_2 = 1$) at the point ($\lambda = -0.2$, $\alpha_1 = 1.0$).

By using a high-speed computer, the complete RSS grid in the $\lambda - \alpha_1$ plane can be determined and hence RSS contours can be drawn.

DISCUSSION

William G. Hunter[2]

The paper under discussion [7][3] raises many interesting and useful points, the primary one being that power transformations are potentially useful for fitting tool-life data. Because of its exploratory nature some questions are, of course, left unresolved. One hopes that they will be discussed in future work.

The power transformations discussed in [7] remain to be tried

[2] On leave of absence from the University of Wisconsin; presently, Imperial College, London, England.

[3] Numbers in brackets designate Additional References at end of this discussion.

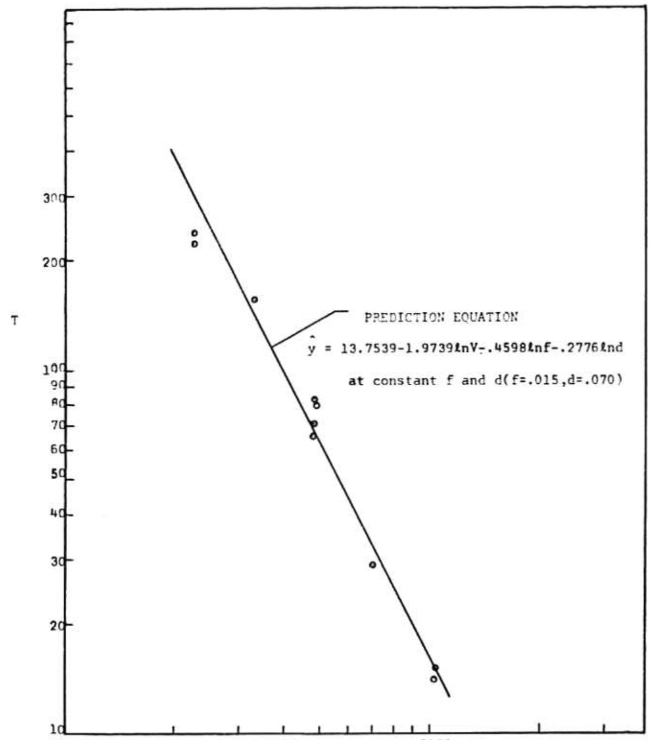

Fig. 12

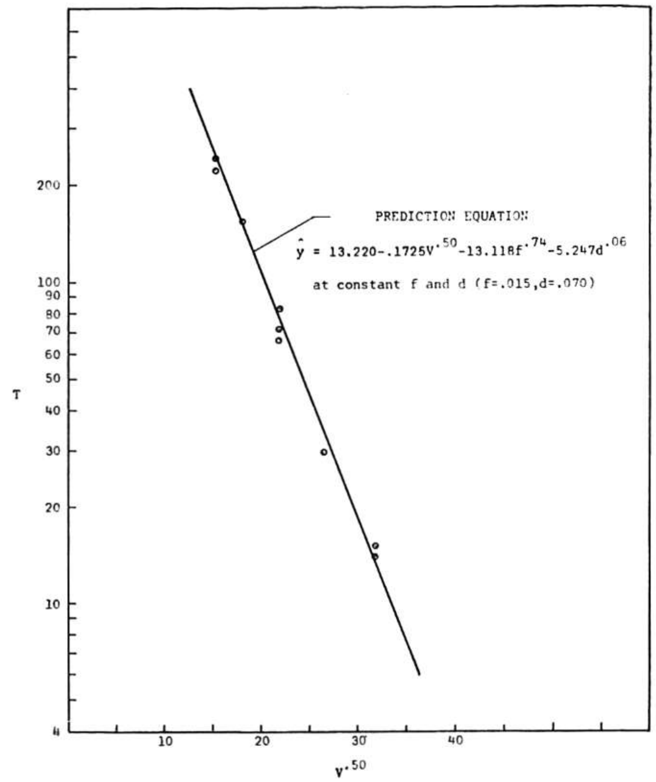

Fig. 14

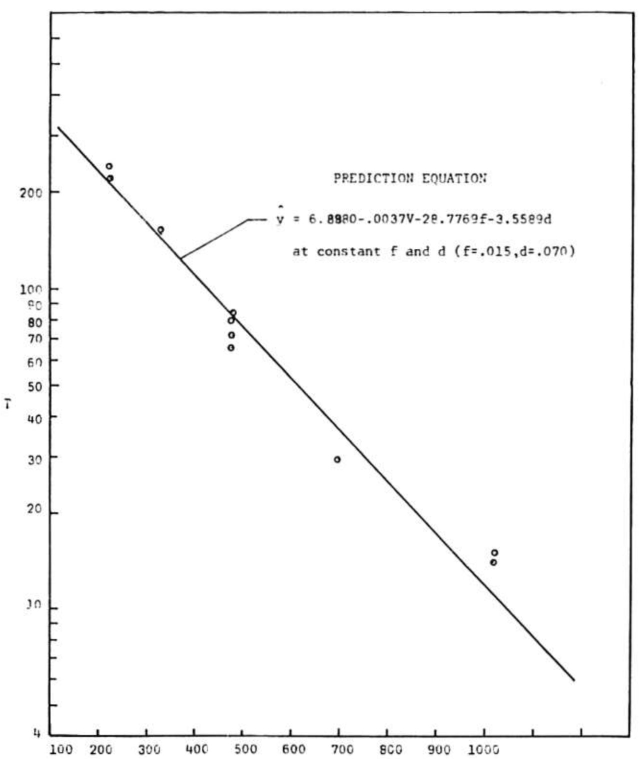

Fig. 13

on many more sets of tool-life data. It will then be possible to assess their general usefulness. Perhaps, in certain circumstances, other classes of transformations will be found superior to the power transformations. But the development of these alternative transformations is somewhere in the future. For the moment power transformations seem to be quite flexible and the immediate task is to try these statistical techniques to see how they work on practical problems. It is likely that they will find wide application in tool-life testing work. In these comments the nomenclature of Wu, Ermer, and Hill [7] is followed.

Extension of Present Research

One obvious extension of the present work is to consider depth of cut in addition to speed and feed. In the summer of 1964, Mr. W. J. Hill and I did some preliminary work along these lines. The methods of Box and Tidwell [8] were applied to the data given by Wu [9], twenty-four observations of tool life T, the variables being speed V, feed f, and depth of cut d. Fitting ln T one finds the values of the transforming parameters which yield the best fit are $\hat{\alpha}_1 = 0.50$, $\hat{\alpha}_2 = 0.74$, and $\hat{\alpha}_3 = 0.06$.

The best fitting first-order prediction equation is

$$\hat{y} = 13.2200 - 0.1725V^{0.50} - 13.1180f^{0.74} - 5.2472d^{0.06} \quad (32)$$

where $\hat{y}$ is the predicted value of ln T. The residual sum of squares $R(\alpha_1, \alpha_2, \alpha_3)$ is 0.4339 in this case with $(\alpha_1, \alpha_2\ \alpha_3) = (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3) = (0.50, 0.74, 0.06)$, i.e., $R(0.50, 0.74, 0.06) = 0.4339$. This value is less than that obtained by simply fitting to the logarithms of the independent variable, $(\alpha_1, \alpha_2, \alpha_3) = (0, 0, 0)$, from which one obtains the residual sum of squares $R(0, 0, 0) = 0.6702$. An interesting point to be noted in passing is that by fitting to the *untransformed* independent variables, $(\alpha_1, \alpha_2, \alpha_3) = (1, 1, 1)$, one obtains a residual sum of squares slightly less than $R(0, 0, 0)$. In fact, $R(1, 1, 1) = 0.6598$.

The prediction equation for the case where $(\alpha_1, \alpha_2, \alpha_3) = (0, 0, 0)$ is

$$\hat{y} = 13.7539 - 1.9739 \ln V - 0.4598 \ln f - 0.2776 \ln d \quad (33)$$

and the prediction equation for the case where $(\alpha_1, \alpha_2, \alpha_3) = (1, 1, 1)$ is

$$\hat{y} = 6.8880 - 0.0037V - 28.7769f - 3.5589d \tag{34}$$

A visual comparison of the three prediction equations (32), (33), and (34) can be obtained from Figs. 12, 13, and 14.

In summary the best fit is obtained in neither the case where the fitting is done to the experimental variables themselves (equation (34)) nor to the logarithms of the variables (equation (33)). Instead equation (32) gives the best fit but it is more complicated and not as easily appreciated as either equations (33) or (34). Perhaps for these reasons equation (32), although giving the best fit, would not be as useful in practice. A close approximation to this prediction equation can be obtained, however, by using $(\alpha_1, \alpha_2, \alpha_3) = (\frac{1}{2}, \frac{3}{4}, \frac{1}{20}) = (0.50, 0.75, 0.05)$. In this form it may be slightly more convenient to use. A summary of the results obtained from the four prediction equations is given in Table 4.

Table 4 Results obtained by fitting set of twenty-four tool-life runs with four prediction equations

Values of transforming parameters α_1, α_2, α_3	Variables used	Residual sum of squares
1.00, 1.00, 1.00	V, f, d	0.6598
0.00, 0.00, 0.00	$\ln V$, $\ln f$, $\ln d$	0.6702
0.50, 0.74, 0.06	$V^{0.50}$, $f^{0.74}$, $d^{0.06}$	0.4339
0.50, 0.75, 0.05	$V^{0.50}$, $f^{0.75}$, $d^{0.05}$	0.4382

The mean square for pure error is 0.027531 (see Appendix C of [9]) and this is associated with eight degrees of freedom. Lack of fit tests indicate that all the above transformations are adequate.

When the number of transforming parameters exceeds two or three the calculations become more difficult and time-consuming. Also there are problems of presenting the results of such calculations so that they can be readily understood and appreciated. Handling these matters can be facilitated by the use of an automatic plotting device which can be employed in direct conjunction with an electronic computer, e.g., the Calcomp plotter or the Stromberg Carlson 4020 Computer Recorder.

The Transformed Design Matrix

In some cases it may be meaningful to ask what happens to the design matrix as a result of transformation. Suppose a statistically designed experiment has been run and values of the transforming parameters have been determined which yield a first-order relationship between the expected value of the *transformed* observations (dependent variable) and the *transformed* (independent) variables. That is, a relationship of the following form has been found to fit the data adequately:

$$E[y^{(\lambda)}] = \beta_0 + \sum_{i=1}^{k} \beta_i U_i \tag{35}$$

The untransformed design matrix $D(x)$ for N runs with k variables is the $N \times k$ matrix of x's giving the standardized levels of the variables to be run. The levels of x are often -1, 0, and $+1$. There will be a standardizing formula for each variable. For example, for each tool-life testing variable V, f, and d there would be such a formula, say, $x_1 = g_1(V)$, $x_2 = g_2(f)$, and $x_3 = g_3(d)$. (See equations (2) in reference [10].)

The reason for selecting one particular design $D(x)$ for the experimental plan as preferable to all the others that might have been chosen is that it has certain properties that the experimenter deems desirable. For example, he may wish to obtain uncorrelated estimates of the effects of the variables and for this reason he selects an orthogonal design. (Certain mild assumptions about the nature of the experimental errors are required to make this logic exactly correct.) An orthogonal design matrix is one for which all the possible sums of cross-products for the columns are zero, i.e., $\sum_{u=1}^{N} x_{iu}x_{ju} = 0$ for all $i \neq j$, where $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, k$. It is extremely unlikely, of course, that the *transformed* design matrix $D(\dot{x})$ will possess this desirable property; on the contrary, it will almost always be true that

$$\sum_{u=1}^{N} \dot{x}_{iu}\dot{x}_{ju} \neq 0, \quad \text{where} \quad \dot{x}_l = g_l(U_l), \qquad l = 1, 2, \ldots, k.$$

The experimenter may be willing to run a few additional experiments (say n) if it is possible to "repair the damage" to the design matrix that has resulted from transformation. In other words, it might be worthwhile to perform n more runs if it is possible to achieve the orthogonality condition $\sum_{u=1}^{N+n} \dot{x}_{iu}\dot{x}_{ju} = 0$. In considering this situation a number of questions arise. When will it be possible to "repair the damage" in this way? We must remember that we are not completely free to choose any levels of the variables for these additional runs; there will be certain constraints imposed because of the physical nature of the situation. However, given these constraints and the results from the first N experiments the problem of minimizing some meaningful measure of nonorthogonality can be solved mathematically for fixed n. But at present it is not known which is the best way to proceed. Another related question is: In a given situation what is the smallest number of addition runs required to satisfy the orthogonality condition, i.e., what is the minimum value for n?

In certain circumstances, I think, experimenters might be interested in answers to questions of this kind. I would be interested in knowing whether the authors agree. The property of orthogonality was merely chosen for illustrative purposes. There are many other properties and which particular properties are most important in a given situation depends, of course, upon the situation. With suitable modification, however, the above remarks still apply, whichever properties are most pertinent. Perhaps the design matrix to consider after transformation should be defined in some alternative way.

Lack of Fit Tests

It is not always true that all values contained in a calculated 95 percent confidence region fit the data adequately. In some cases all such values may give an inadequate fit because the model itself is inadequate. As illustrated so well in [9] and [10] a lack of fit test is necessary. A statistical analysis is incomplete if one does not attempt to check the adequacy of the mathematical model (and, for that matter, all the other assumptions that are tentatively made at the outset of the analysis). In the examples in this paper what do each of the lack of fit tests reveal?

In this regard it would be helpful to publish the data that were used. Granted there are serious space limitations in scientific journals, but in general, data are not published as often as they should be.

The data presented in Table 3 show some peculiarities. Apparently there are nine distinct sets of conditions, each of which is repeated four times. But within each set there is a consistent trend of results for the tool life T. Are the results shown in the order in which they were taken? If so, what is the cause of the trends? Were there variables in addition to speed and feed? If there were no extra variables and the model is correct, then the trends are only manifestations of experimental error. But such a chance happening (that is, consistent trends within each set of four replicates) is very rare, occurring in fact fewer than once in a million trials. Perhaps the results were taken in random order and rearranged for purposes of presentation. Is this the case? If so, does each of the nine sets consist of four runs which are genuine replicates?

Questions such as these are relevant to a proper assessment of the model. Incidentally, in answering them one might shed some light on the supposed ambiguity in the carbide tool-life data. If the model represented by equation (35) is inadequate then it is of no practical use to report the best transformations; they are simply not good enough. If, on the other hand, the model is adequate for both sets of data (i.e., the 36 observations at 200 Bhn with a C-6 carbide tool and the 30 observations at 300 Bhn with a C-7 carbide tool) then perhaps there is not *one* best transformation that is appropriate for both. There is the possibility that two transformations are necessary, one for each set of data.

There is a tendency in some other fields of engineering research for experimenters, who are searching for "the best" model to fit their data, to examine only the values of the residual sum of squares for the various models and to report the *single* model which yields the minimum residual sum of squares. In general, such practice is of questionable practical value, especially since experimental error is rarely taken properly into account. If experimental error is taken into account (this will be possible if there is replication), one can see that the data often do not support the conclusion that a single model is correct but rather support alternative conclusions, e.g., all models so far considered are inadequate or more than one model adequately fit the data.

Readers may get the impression from [7] that lack of fit tests are unnecessary and I feel that the authors do not want to give this impression. For the proper assessment of Tables 1 and 2, for instance, it is necessary to know what experimental error is. By merely selecting that transformation which gives the minimum residual sum of squares an experimenter can sometimes be misled.

Confidence Regions

Equations (25) to (27) in [7] are incorrect because $\hat{\alpha}_2$ has not been used as specified in equation (23) which is correct. The approximate 95 percent confidence regions in all the figures should therefore be rechecked. The value of $\hat{\alpha}_2$ should be reported to the same number of decimals as $\hat{\lambda}$ and $\hat{\alpha}_1$.

Additional References

7 S. M. Wu, D. S. Ermer, W. J. Hill, "An Exploratory Study of Taylor's Tool-Life Equation by Power Transformations," a paper by S. M. Wu, D. S. Ermer, and W. J. Hill presented at the 1965 ASME Metals Engineering-Production Engineering Conference, Berkeley, California, June 10, 1965.

8 Box and Tidwell, "Transformation of the Independent Variables," *Technometrics*, vol. 4, 1962, pp. 531–550.

9 S. M. Wu, "Tool-Life Testing by Response Surface Methodology," Journal of Engineering for Industry, Trans. ASME, Series B, vol. 86, no. 2, part 2, 1964, pp. 105–116.

10 S. M. Wu, part 1 (1964)—same as above.

Authors' Closure

The authors appreciate Dr. Hunter's comments and his interest in this investigation. The paper was written from the viewpoint of engineering applications of power transformations. However, a discussion of the use of power transformations from a statistical viewpoint is very helpful. Since the present paper is only an exploratory study, we expect further tool-life data to be analyzed by power transformations as well as other alternative methods.

It is obvious that experimenters will be interested in techniques which will enable them to obtain a transformed design matrix with desirable properties, provided the experimenters had started with a statistically designed experiment. However, the data presented in this paper, except for the third part of the study, were not obtained using statistical experimental designs. The data in Table 3 were obtained from a graph in a previously published government report and the authors do not have any additional information about the order in which the data were taken.

Theoretically, the adequacy of the postulated model should be checked by a lack of fit test. In this instance, however, previous research has shown that a linear model based on a logarithmic transformation is adequate. Since the purpose of the study was to search for transformations to linearize the tool-life data and give a better fit than a logarithmic transformation, there was no reason to doubt the adequacy of the linear models presented.

Finally, the authors want to thank Dr. Hunter for pointing out that $\hat{\alpha}_2$ instead of α_2 should be used in equations (25) to (27).