To be presented at the National Conference of
The Society of Aeronautical Weight Engineers, Inc.
10-13 May 1954
Lord Baltimore Hotel, Baltimore, Maryland

STATISTICAL CONSTANTS IN PREDICTING EQUATIONS

Joseph B. Tysver
Associate Research Engineer

2063-4-J

Based Upon Work Done Under USAF Contract
AF 13(616)-199. Published by Permission
of the Wright Air Development Center.
ABSTRACT

A linear predicting equation is presented for which the coefficients are obtained by minimizing the sum of the squares of the percentage errors. Comparison with the ordinary least-squares regression equation (based on actual errors) shows that the two equations can yield substantially different estimates when the range of the variables in the sample to be fitted is large. If percentage errors are more important to the experimenter than actual errors, and if the range of the variables is large, the predicting equation presented in this paper is preferable to the ordinary regression equation.
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Abstract</td>
<td>1</td>
</tr>
<tr>
<td>I</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>II</td>
<td>The Error of Estimation</td>
<td>3</td>
</tr>
<tr>
<td>III</td>
<td>Least-Square Linear Prediction Based on Actual Error</td>
<td>5</td>
</tr>
<tr>
<td>IV</td>
<td>Least-Square Linear Prediction Based on Relative Error</td>
<td>6</td>
</tr>
<tr>
<td>V</td>
<td>Application</td>
<td>9</td>
</tr>
<tr>
<td>VI</td>
<td>Conclusions</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>12</td>
</tr>
<tr>
<td>VII</td>
<td>Appendix</td>
<td>15</td>
</tr>
</tbody>
</table>
INTRODUCTION

In aircraft weight estimation three approaches are possible: analytical, statistical, or a combination of both. No purely analytical approach is possible at the present time because of two factors:

1) Complete analytic treatment has not been developed and will be quite complicated when it does become available.

2) The presence of differences (and trends) in fabrication procedures introduces differences from an analytic model which cannot be ignored.

On the other hand, the extrapolation which is, and will be, required introduces errors in a purely statistical approach which also cannot be ignored. Thus it appears that the only satisfactory method available for aircraft weight estimation will be based on analytic formulas with statistical constants. The analytic components will be determined by availability of theoretical development, adaptability to calculation, applicability to extrapolation, and acceptability of errors obtained; the statistical constants will be chosen to improve the fit of the analytic components to the weights to be estimated. Some of the aspects of fitting of statistical constants are discussed in this paper.

The equation to be used in weight estimation contains variables $x_i$ ($i=1, ..., m$) which may represent parameters of the aircraft or analytically derived weight components, which themselves are functions of aircraft parameters; it also contains constants $a_j$ ($j=1, ..., k$) which are to be determined from a sample of several aircraft so that the equation has a "best fit" to known weights. The definition of a "best fit" to be used here is the usual one of a least-squares fit. That is, the sum of the squares of the errors shall be a minimum. The basic form of the estimation is assumed to be:

$$Y = a_0 + a_1 x_1 + ... + a_m x_m = a_0 + \sum_{i=1}^{m} a_i x_i$$  \hspace{1cm} (1)

Two other specific forms which can be transformed to this form and are of special interest are:

$$Y' = b_0 x_1 x_2 ... x_m = b_0 \prod_{i=1}^{m} x_i$$  \hspace{1cm} (2)
\[ Y' = a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m = \sum_{i=0}^{m} a_i x^i \] (3)

for which the respective transformations are:

\[ x_1 = \log x \] and \[ Y = \log Y' \] (2')

\[ x_1 = x^1 \] and \[ Y = Y' \] (3')


II

THE ERROR OF ESTIMATION

The choice of the quantity which is used to represent the error of estimation will depend on what the estimator wishes to guard against and what assumptions he is willing to make. Let us denote the actual weight of the jth object in a sample of n objects by \( y_j \), and the estimated weight of the jth object by \( Y_j \).

The quantity to represent the error is then

\[ E_j = \pm (y_j - Y_j) \]

where the choice of the plus or minus sign will depend upon the estimator's point of view, but does not affect the mathematical treatment of \( E_j^2 \). The statistician uses the plus sign and considers \( E_j \) as the error of observation. If it can be assumed that the predicting variables (\( x_1 \)) are without error, then this choice of \( E_j \) leads to the ordinary linear regression theory found in standard statistical texts (Ref. 1) which is outlined in the next section of this paper.

If the predicting variables cannot be assumed to be without error we are led to an orthogonal linear regression theory which also can be found in statistical texts (Ref. 2). The difference between these two theories is illustrated in Figure 1 for one predicting variable.

\[ E_j = \text{Actual Error} \]
\[ E_j' = \text{Orthogonal Error} \]

**FIG. 1** ERROR OF ESTIMATION
In some applications the relative (percentage) error is more important than the actual error. For example an error of 500 pounds may be acceptable in an observed value of 40,000 pounds, but would be intolerable in an observed value of 1,000 pounds. If the range of values is large it may be advisable to determine statistical constants which minimize the sums of squares of relative errors rather than of actual errors. A method for accomplishing this is presented in Section IV of this paper.

Four possible definitions of the relative error $e$ are presented below

**TABLE I**

**RELATIVE ERROR, $e$**

<table>
<thead>
<tr>
<th>Numerator</th>
<th>$y - Y$</th>
<th>$y - y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denominator</td>
<td>$y - (a + bx)$</td>
<td>$(a + bx) - y$</td>
</tr>
<tr>
<td>$Y$</td>
<td>$\frac{y - (a + bx)}{a + bx}$</td>
<td>$\frac{(a + bx) - y}{a + bx}$</td>
</tr>
<tr>
<td>$y$</td>
<td>$1 - a\left(\frac{1}{y}\right) - b\left(\frac{x}{y}\right)$</td>
<td>$a\left(\frac{1}{y}\right) + b\left(\frac{x}{y}\right) - 1$</td>
</tr>
</tbody>
</table>

On the basis of logical signs and simplicity of calculation the definition selected here is

$$e = a\left(\frac{1}{y}\right) + b\left(\frac{x}{y}\right) - 1$$
III

LEAST-SQUARE LINEAR PREDICTION BASED ON ACTUAL ERROR

For purposes of review and/or comparison, the ordinary least-squares development is outlined here with one predictor $x$. A predicting equation

$$ Y = a + bx $$

is to be obtained from a sample of observed values $(x_j, y_j)$ with $j = 1, 2, \ldots, n$. The values of $x_j$ are assumed to be without error. The coefficients, $a$ and $b$, are to be determined so that the sum of the squares of the actual errors, defined as:

$$ E_j = (y_j - \hat{y}_j), $$

is a minimum. The values of the coefficients which satisfy this condition are determined from the two equations:

$$ \frac{\partial}{\partial a} \sum E_j^2 = 0 \quad \text{and} \quad \frac{\partial}{\partial b} \sum E_j^2 = 0 $$

or

$$ \sum y_j - na - b \sum x_j = 0 \quad \text{and} \quad \sum x_j y_j - a \sum x_j^2 - b \sum x_j^2 = 0 $$

giving

$$ b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} $$

$$ a = \frac{\sum y - b \sum x}{n} $$

It should be noted that the first equation gives the condition $\sum E_j = 0$; hence this choice of coefficients gives errors with zero mean and minimum standard deviation.
IV

LEAST-SQUARE LINEAR PREDICTION BASED ON RELATIVE ERROR

First we shall consider a predicting equation of the form

\[ Y = a + bx \]

The coefficients \( a \) and \( b \) are to be determined so as to minimize the sum of the squares of the relative errors

\[ e_j = \frac{y_j - y_j}{y_j} \]

for a sample of \( n \) pairs \((x_j, y_j)\). The two conditions used to determine the coefficients are

\[ \frac{\partial}{\partial a} \Sigma e_j^2 = 0 \quad \text{and} \quad \frac{\partial}{\partial b} \Sigma e_j^2 = 0. \]

Setting

\[ u_j = x_j/y_j \quad \text{and} \quad v_j = 1/y_j \]

we have

\[ e_j = a v_j + b u_j - 1. \] (6)

The two conditions yield the equations

\[ a \Sigma v_j^2 + b \Sigma u_j v_j - \Sigma v_j = 0 \]

and

\[ a \Sigma u_j v_j + b \Sigma u_j^2 - \Sigma u_j = 0 \]

and the solutions

\[ b = \frac{\Sigma u_j^2 \Sigma v_j - \Sigma u_j \Sigma u_j v_j}{\Sigma u_j^2 \Sigma v_j^2 - (\Sigma u_j v_j)^2} \] (7)

\[ \]
and
\[ a = \frac{\sum v_j \sum u_j^2 - \sum u_j \sum u_j v_j}{\sum u_j^2 \sum v_j^2 - (\sum u_j v_j)^2} \]  

(8)

Note that the condition
\[ \sum e_j = 0 \]
does not necessarily hold for this form. It can be shown that
\[ \sum e_j = -\sum e_j^2. \]

This theory may be generalized for application to a sample of \( n \) observations on \( m \) independent variables \( (x_{ij}) \) and one dependent variable \( y \) with the estimation equation
\[ Y_j = a_0 + a_1 x_{1j} + \ldots + a_m x_{mj} \quad j = 1, \ldots, n \]

and the relative errors
\[ e_j = \frac{Y_j - \bar{y}_j}{\bar{y}_j} = \sum_{i=0}^{m} a_i u_{ij} - 1 \]

where
\[ u_{ij} = \begin{cases} 
1/y_j & \text{for } i = 0 \\
x_{ij}/y_j & \text{for } i = 1, \ldots, m 
\end{cases} \]

The requirement that
\[ \sum_{j=1}^{n} e_j^2 = \text{minimum} \]
yields \( m + 1 \) linear equations

\[
\frac{\partial}{\partial a_i} \sum e_j^2 = 0 \quad i = 0, 1, \ldots, m
\]

which can be solved for the \( m + 1 \) coefficients \( a_i \). A computational form is presented in the Appendix.

Other variations of these results can be obtained by specifying such conditions as:

1) \( \sum e_j^2 = 0 \) (the mean relative error is zero).
2) \( a_o = 0 \) (the estimation equation passes thru the origin).
The analytically derived components of structural wing weight presented in Reference 3 will be used to illustrate the method of determining statistical constants. The four analytic components $x_1$, $x_2$, $x_3$, and $x_4$ are used to estimate structural wing weights for 12 fighters, 13 bombers, and 15 cargo aircraft with an equation of the form

$$w_{we_1} = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4.$$  \(9\)

From the analytic derivation the physical interpretations will be maintained only if $x_1$ and $x_3$ have positive coefficients and $x_2$ and $x_4$ have negative coefficients. These signs are not obtained in all of the three samples (fighters, bombers, and cargo aircraft) so that physical interpretation is lost in making the statistical fit.

A second attempt is made using the equation

$$w_{we_2} = a_0 + a_1 (x'_1) + a_2 (x'_2).$$  \(10\)

where $x'_1 = x_1 - x_2$ and $x'_2 = x_3 - x_4$. For this equation it is assumed that the poorer fit will be offset by a gain in physical interpretation and (it is hoped) a better applicability to extrapolation. The coefficients $a_1$ and $a_2$ should be positive to allow physical interpretation. The statistical fit of the samples again violates these sign requirements.

A further forcing of the sign requirements is made to give the equation

$$w_{we_3} = a_0 + a_1 x'.$$  \(11\)

where

$$x' = x'_1 + x'_2.$$  

In this case the coefficient $a_1$, as determined by a statistical fit, is positive so that physical interpretation is maintained.
Since the right side of the predicting formula has been reduced to one predictor, other formulas using \( x_1 \) and \( x_2 \) respectively may be of interest. These are

\[
W_{we_4} = a_0 + a_1 x_1,
\]

and

\[
W_{we_5} = a_0 + a_1 x_2.
\]

The following table gives the values of \( e_{rms} \) for the different equations as applied to a sample of 12 fighter aircraft where

\[
e_{rms} = \sqrt{\frac{\sum e_j^2}{n}} = \text{root mean square relative error}
\]

### TABLE II

**RELATIVE ERROR (rms) FIGHTER AIRCRAFT (12)**

<table>
<thead>
<tr>
<th>Formula</th>
<th>( e_{rms} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( W_{we_1} = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 )</td>
<td>0.122</td>
</tr>
<tr>
<td>( W_{we_2} = a_0 + a_1 (x_1 - x_2) + a_2 (x_3 - x_4) )</td>
<td>0.149</td>
</tr>
<tr>
<td>( W_{we_3} = a_0 + a_1 [(x_1 - x_2) + (x_3 - x_4)] )</td>
<td>0.150</td>
</tr>
<tr>
<td>( W_{we_4} = a_0 + a_1 (x_1 - x_2) )</td>
<td>0.157</td>
</tr>
<tr>
<td>( W_{we_5} = a_0 + a_1 (x_3 - x_4) )</td>
<td>0.152</td>
</tr>
</tbody>
</table>

Note the increase in \( e_{rms} \) is comparatively small when the variables are combined to force the signs for physical interpretation.
A comparison of the formulas based on actual errors with those based on relative errors is presented for Equation (12). An over-all fit to all categories of aircraft (i.e. including the three samples of 12 fighters, 13 bombers, and 15 cargo aircraft in one sample of 40 aircraft) is included. Values of $e_{rms}$ are presented in Table III.

Figures 2 and 3 present estimated wing weights versus actual wing weights for relative error fit and actual error fit, applied to the over-all category. Note the large relative errors at the lower range of Figure 3.

### TABLE III

**RELATIVE ERROR ($e_{rms}$) OF WING WEIGHT ESTIMATES**

\[
W_{we} = a_0 + a_1 x_1
\]

<table>
<thead>
<tr>
<th>Category</th>
<th>Relative Error Fit</th>
<th>Actual Error Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>By Category</td>
<td>All Categories</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fighters (12)</td>
<td>0.157</td>
<td>0.174</td>
</tr>
<tr>
<td>Bombers (13)</td>
<td>0.175</td>
<td>0.195</td>
</tr>
<tr>
<td>Cargos (15)</td>
<td>0.092</td>
<td>0.144</td>
</tr>
<tr>
<td>All (40)</td>
<td>0.143</td>
<td>0.171</td>
</tr>
</tbody>
</table>
VI

CONCLUSIONS

In fitting an equation to data statistically the experimenter must first determine what type error he wishes to minimize. When percentage error is of primary importance, as it is in aircraft weight estimation, the method developed in this paper is appropriate for fitting a predicting equation to the data. This approach requires a slight increase in computation over that required for the equation based on actual errors, but it reduces the dominating role of the errors in the larger values of a sample. If the range of values in the sample is large, the two bases can yield substantially different percentage errors.

REFERENCES

1. P. G. Hoel, Introduction to Mathematical Statistics, Chapter V


3. Spath, R. M., "A Simple Wing Weight Estimation Equation", University of Michigan, Willow Run Research Center Memorandum 2063-3-J.
FIG. 2 ESTIMATED WING WEIGHT USING RELATIVE ERROR FIT
ALL CATEGORIES
FIG. 3  ESTIMATED WING WEIGHT USING ACTUAL ERROR FIT ALL CATEGORIES
VII

APPENDIX

LINEAR REGRESSION WITH LEAST-SQUARES RELATIVE ERROR

Observed Sample Values: \((x_{ij}, y_j)\) \(i = 1, \ldots, m; \ j = 1, \ldots, n\)

Equations: \(Y_j = a_0 + a_1 x_{1j} + a_2 x_{2j} + \ldots + a_m x_{mj}\)

Transformation:

\[
u_{ij} = \begin{cases} 
1/y_j & i = 0 \\
 x_{ij}/y_i & i = 1, 2, \ldots, m
\end{cases}
\]

Relative Error:

\(e_j = \frac{y_j - y_j}{y_j} = \sum_{i=0}^{m} a_i u_{ij} - 1\)

Condition: \(\sum_j e_j^2 = \text{minimum} \Rightarrow \frac{\partial}{\partial a_r} \Sigma e_j^2 = 0, \ r = 0, 1, \ldots, m\)

\[
\frac{\partial}{\partial a_r} \Sigma e_j^2 = 2 \Sigma e_j \frac{\partial e_j}{\partial a_r} = 2 \Sigma e_j u_{rj}
\]

\[
\Sigma e_j u_{rj} = \Sigma u_{rj} \sum_{i=0}^{m} a_i u_{ij} - \Sigma u_{rj}
\]

\[
= \sum_{i=1}^{m} \left( \sum_{j=1}^{n} u_{rj} u_{ij} \right) - \sum_{j=1}^{n} u_{rj} = 0
\]
System to be solved for $a_i$, $i = 0, 1, \ldots, m$:

$$a_0\Sigma u_{o,j}^2 + a_1\Sigma u_{o,j}u_{1,j} + a_2\Sigma u_{o,j}u_{2,j} + \ldots + a_m\Sigma u_{o,j}u_{m,j} = \Sigma u_{o,j}$$

$$a_0\Sigma u_{o,j}u_{1,j} + a_1\Sigma u_{1,j}^2 + a_2\Sigma u_{1,j}u_{2,j} + \ldots + a_m\Sigma u_{1,j}u_{m,j} = \Sigma u_{1,j}$$

$$a_0\Sigma u_{o,j}u_{2,j} + a_1\Sigma u_{1,j}u_{2,j} + a_2\Sigma u_{2,j}^2 + \ldots + a_m\Sigma u_{2,j}u_{m,j} = \Sigma u_{2,j}$$

..........................................................

$$a_0\Sigma u_{o,j}u_{k,j} + a_1\Sigma u_{1,j}u_{k,j} + a_2\Sigma u_{2,j}u_{k,j} + \ldots + a_m\Sigma u_{m,j}^2 = \Sigma u_{m,j}$$
### Computational Form

\[
\begin{array}{cccccc}
\Sigma u_{ij}^2 & \Sigma u_{ij}u_{ij}^1 & \cdots & \Sigma u_{ij}u_{mj} & \Sigma u_{ij} \\
\Sigma u_{1j}^2 & \Sigma u_{1j}u_{2j} & \cdots & \Sigma u_{1j}u_{mj} & \Sigma u_{1j} \\
\Sigma u_{2j}^2 & \cdots & \Sigma u_{2j}u_{mj} & \cdots & \Sigma u_{2j} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\Sigma u_{mj}^2 & \cdots & \cdots & \cdots & \cdots & \Sigma u_{mj} \\
\end{array}
\]

\[
\begin{array}{ccc}
U_{11} & U_{12} & \cdots & U_{1m} \\
U_{22} & \cdots & U_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
U_{mm} & \cdots & \cdots & U_{mm} \\
U_{00} & \cdots & \cdots & \cdots \\
\end{array}
\]

\[
u_{rs} = \Sigma u_{ij}^2 \Sigma u_{rj}s_j - \Sigma u_{ij}u_{rj} \Sigma u_{oj}s_j \\
u_{ro} = \Sigma u_{oj}^2 \Sigma u_{rj} - \Sigma u_{oj}u_{rj} \Sigma u_{oj}
\]

\[
u_{rs,k} = \frac{U_{kk}(k-1) U_{rs}(k-1) - U_{ks}(k-1) U_{rk}(k-1)}{U(k-1)(k-1)(k-2)}
\]

\[
a_0, a_1, a_2, \ldots, a_m
\]

\[
a_m = U_{m0}(m-1) / U_{mm}(m-1)
\]

\[
a_{m-j} = U(m-j)0.(m-j-1) \sum_{i=m-j+1}^{m} a_i U(m-j)i.(m-j-1) / U(m-j)(m-j)(m-j-1)
\]

\[
a_o = \Sigma u_{oj}^2 \sum_{i=1}^{m} a_i (\Sigma u_{oj} u_{ij}) / \Sigma u_{oj}
\]