

IEEE STANDARDS
ARCHIVES
DO NOT REMOVE

AIEE
No. 1F
June 1958

Report on Guide for
STATISTICAL ANALYSIS OF TEST DATA

(Published for trial use)



AIEE No. 1F

Published by
AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS
33 West Thirty-Ninth Street, New York 18, N. Y.

ACKNOWLEDGEMENT

The Institute wishes to acknowledge its indebtedness to those who have so freely given of their time and knowledge and conducted experimental work on which many of its publications are based.

This Guide was prepared by the Working Group on Statistical Methods of Standards Coordinating Committee No. 4.

The personnel of the Working Group was:

J. H. Davidson, <i>Chairman</i>	T. W. Dakin
P. L. Alger	W. H. Horton
E. J. Croop	C. L. Matz
C. G. Currin	R. H. Rech

CONTENTS

Introduction	3
Estimation of Average Life and Precision of Measurement at One Temperature	3
Quick Estimation of the Relationship Between Life and Temperature	4
Formal Estimation of the Relationship Between Life and Temperature	5
Calculation of Regression Line	6
Accuracy of Regression Line	6
Linearity of Regression	7
Comparing Two Regressions	7
References	7

REPORT ON

GUIDE FOR THE STATISTICAL ANALYSIS OF TEST DATA

1F-1 Introduction

As is stated in AIEE No. 1, accelerated life tests are recommended to determine the probable life expectancy of electrical insulating materials and of complete insulation systems under different conditions of service. Recommended procedures for conducting life tests at various temperatures to accelerate insulation degradation are outlined in AIEE 1D and 1E. AIEE publications Nos. 57, 65, 510 and 511 are examples of test procedures developed for this purpose.

After life data from such tests have been obtained at different temperatures, the problem is to evaluate the resultant data so that the life expectancy of insulation can be estimated. For this purpose, life expectancy is considered as a function of temperature. The most probable form of this relationship, and its reliability, are determined by the statistical procedures outlined herein.

It has been found that the Arrhenius equation, which describes the temperature dependence of the "velocity coefficient" of chemical reactions, can be used to approximate the relationship between insulation life and temperature. This equation indicates that the logarithm of insulation life should be a linear function of the reciprocal of absolute temperature. Regression analysis is the technique used to determine the best estimate of the slope and intercept of the straight line that relates the logarithm of insulation life to the reciprocal of absolute temperature.

Dakin¹ found such relationships fitted the data very well in many cases. The Arrhenius equation for reaction rate is given by:

$$k = Ae^{-E/RT}$$

where:

k = specific reaction rate
 E = activation energy (relatively constant for small temperature range)
 R = gas constant
 T = absolute temperature
 A = frequency factor (constant)

The Arrhenius equation can be written by taking logarithms in the following form:

$$\ln k = \ln A - E/RT$$

If one now lets:

$$\begin{aligned} Y &= \ln k \\ a &= \ln A \\ b &= -E/R \end{aligned}$$

and,

$$X = 1/T^*,$$

one has

$$Y = a + bX \quad (1)$$

*It is often convenient when using a desk calculator to use $X = 1000/T$, rather than $1/T$, as then most values of X and Y will lie between 0 and 10.

which relates the two variables Y and X in the form of a simple linear equation, assuming a and b are constants. These constants can then be obtained from experimental data by the method of least squares, which determines a line, defined as the regression line, such that the sum of squares of the deviations of the observed Y 's from the Y 's estimated by the line shall be a minimum. This line gives the best fit in the sense that the root-mean-square variations from it to all observed points is less than that given by any other straight line that can be drawn.

The first step in finding the relationship between X and Y is to make a plot of the experimental data using graph paper with logarithm of insulation life as the ordinate and the reciprocal of the absolute temperature as the abscissa. For convenience, the horizontal scale is often marked in degrees Centigrade also.

The test data normally consist of a series of values of the observed life of a test sample at different exposure temperatures. Generally, the results are grouped around 3 or 4 temperatures, 20C or more apart.

For each set of test data, there is plotted a central line, defined by the regression equation, which represents the most probable relationship of life versus temperature that can be derived from the test data. Also, there are drawn two limit lines, which define the 95 percent confidence limits of the data. That is, on the basis of the test data, there is a 95 percent certainty that the actual average life of the insulation under the test conditions will lie inside the zone between these two limit lines.

When an investigator desires to do the work with a slide rule for approximate or "quick and rough" purposes, the less exact procedure given in Section 1F-2 may be used.

However, for purposes of record, and as a basis for determining standards or classifying new materials, the regression analysis procedure should be carried through in a formal way as indicated under Section 1F-4. The methods outlined in this section are arranged for convenient use when computing machines are employed in making numerical calculations. The work is blocked out in tabular form, so that it can be performed by an analyst without special knowledge of statistical theory.

1F-2 Estimation of Average Life and Precision of Measurement at One Temperature

For convenience and a quick preliminary estimate, the data may be plotted on arithmetic probability paper. On this paper, the ordinate is marked according to the arithmetic scale while the abscissa is scaled according to the normal distribution so that data from a normal curve would fall on a straight line. Figure

1F-2 shows a plot, on such paper, of the data given in Tables 1F-3a through 1F-3d in Section 1F-3. In this graph, the vertical scale represents life in hours and the horizontal scale indicates the failure probability. The horizontal scale is compressed at the center and expanded at the ends, in accordance with the normal distribution law. Therefore, if the test data follow the normal law, the consecutive points for each insulation component should lie upon a straight line. This is a good quick check for normality.

The three lines drawn through the points representing insulation failures of ground, between phase and between conductors, all fit the data closely showing that these tests do not show marked deviations from the normal distribution. From a theoretical standpoint, it must be recognized that the logarithms are normally distributed rather than the values themselves. However, for the smaller ranges of values usually encountered, the arithmetic plot is generally sufficient for a rough check.

In Figure 1F-2, the average life is given by the ordinate at the center of the horizontal scale, or 50 percent point. In these tests, this is found to be 368, 358 and 312 hours at 180C for the three elements of the insulation, as indicated in the figure. The standard deviation of the life is measured by the slope of the life line. For example, since a range of 2 standard deviations on either side of the mean covers 95 percent of the total tests, under the normal law, the standard deviation for the conductor insulation is found by reading the ordinate at the 2.5 percent point on Figure 1F-2, subtracting this from the ordinate at the 50 percent point, and dividing the result by 2, giving 58 hours. Similarly, the standard deviations for ground and phase insulations are found to be 75 and 71 hours, as indicated on the figure.

The standard deviation of any measured quantity, X , can be obtained more exactly by using the following formula:

$$\text{Standard Deviation (S)} = \frac{\text{the square root of the sum } (x - \bar{x})^2}{\sqrt{n - 1}}$$

$$= \left[\frac{n \sum x^2 - (\sum x)^2}{n(n - 1)} \right]^{1/2}$$

where:

Standard Deviation (S) = the root-mean-square of the deviations of the observed times to failure from their average
 x = each individual value recorded
 $\bar{x}$ = average of the values at a given temperature
 n = number of test points

Lines drawn through these calculated values will represent the data more accurately than a line drawn "by eye" through the individual test points.

1F-3 Quick Estimation of the Relationship Between Life and Temperature

The next step in the approximate method is the plotting of the averages of life against the corresponding temperatures. To do this, the ordinate is used to represent log life and the abscissa represents the reciprocal of the absolute temperature. Plotting in this manner, a straight line should result in accordance with the Arrhenius equation. A straight line may be drawn through the points (one at each of the temperatures) fitting the points by "eye". A transparent ruler is helpful and can be adjusted so that the differences seem to balance out on either side of the line.

As an alternate to drawing the line by "eye", one may use the "two-point formula". This is the method proposed by A. Hald and is outlined in Work Sheet 1F-3.

When there is an even number of temperatures, let $k = 2m$. Compute the means for the first m points and similarly for the second m points and draw the line through the two points so obtained. For only 4 temperatures, the calculations of means may be dispensed with, and the two points determined by eye from the graph.

Formulawise, b in Equation (1) can be calculated:

$$b = \frac{\bar{Y}'' - \bar{Y}'}{\bar{X}'' - \bar{X}'}$$

where:

$$\bar{X}' = \frac{1}{m} \sum_{i=1}^m X_i$$

$$\bar{X}'' = \frac{1}{m} \sum_{i=m+1}^{2m} X_i$$

and $\bar{Y}'$, $\bar{Y}''$ are the means of the Y_i 's corresponding to the X_i 's. Also,

$$a = \bar{Y}' - b\bar{X}'$$

When there is an odd number of temperatures, the center points are omitted in determining the slope but the regression line is drawn so that it goes through the overall mean point $(\bar{X}, \bar{Y})$. For three temperatures, the regression line will be such that the vertical distance of the line from either extreme temperature is one-half the distance from the center. The computations for this are shown in Work Sheet 1F-3, "Odd Number of Temperatures". Figure 1F-3 indicates the comparison between the regression lines calculated in this manner and the more accurate methods outlined in Section 1F-4.

The points should lie on the same straight line when plotted in this manner since it is expected that the observed aging will follow the Arrhenius equation. When using short-cut methods, confidence limits cannot be established about the regression line nor can analytical tests be made for linearity or for comparing two different lines. However, the fact that test points in general do fit a straight line very closely gives some

confidence in reliability of this method. A marked curve in the temperature-life line on such a chart would suggest that reactions are taking place at the higher test temperatures that are not significant at the lower test temperatures. Such a condition would arise if the lower temperature failures were due to a different cause than the higher temperature failures. This might occur, for example, if different components of the insulation system were responsible for failures at the two ends of the temperature scale.

Table 1F-3a

INSULATION LIFE IN HOURS AT 180C

Number of Unit	Ground	Phase	Between Conductors	
			Top Coil	Bottom Coil
1	240	240	192	240
2	288	288	240	288
3	336	336	288	288
4	336	336	288	288
5	336	336	288	288
6	384	336	288	336
7	384	384	336	336
8	384	384	336	336
9	480	432	336	384
10	480	480	432	432

Table 1F-3b

**CONDUCTOR INSULATION FAILURES
ON MOTORETTES AGED AT 180C**

Number of Failures	Hours Aged to Failure	Cumulative Percentage
1	192	2.5
2	240	7.5
3	240	12.5
4	288	17.5
5	288	22.5
6	288	27.5
7	288	32.5
8	288	37.5
9	288	42.5
10	288	47.5
11	288	52.5
12	336	57.5
13	336	62.5
14	336	67.5
15	336	72.5
16	336	77.5
17	336	82.5
18	384	87.5
19	432	92.5
20	432	97.5

Table 1F-3c

**GROUND INSULATION FAILURES OF
MOTORETTES AGED AT 180C**

Number of Failures	Hours Aged to Failure	Cumulative Percentage
1	240	5
2	288	15
3	336	25
4	336	35
5	336	45
6	384	55
7	384	65
8	384	75
9	480	85
10	480	95

Table 1F-3d

**PHASE INSULATION FAILURES OF
MOTORETTES AGED AT 180C**

Number of Failures	Hours Aged to Failure	Cumulative Percentage
1	240	5
2	288	15
3	336	25
4	336	35
5	336	45
6	336	55
7	384	65
8	384	75
9	432	85
10	480	95

**1F-4 Formal Estimation of the Relationship Between
Life and Temperature**

If the data follow a linear relationship, as indicated by graphical plot, then the regression equation can be written in the form of Equation (1) where:

$Y = \text{logarithmic insulation life}$

$a, b = \text{constants}$

$X = \text{reciprocal of the absolute temperature}$

The problem then is to find the values of a and b which will give the best straight line. "Best" is defined as that line which minimizes the sum of squares of the deviations of observed Y 's from the Y 's estimated by the regression line. This method for finding the best line is referred to as the "least squares" technique. The estimation of the regression line was covered in a clear and concise manner by Horton².

In this section, statistical techniques and formulas are recommended for a detailed analysis of the test data. Since oftentimes it may not be necessary to make a complete analysis, a short-cut technique is presented

which will give approximate results for finding the regression line. This is important because, as illustrated by Graph 1F-3, generally little is lost in accuracy of the line by using approximate methods. The main contribution of statistical methods is the determination of the pertinent variations. The methods to be discussed are grouped under the following headings:

1F-4a Calculation of Regression Line

1F-4b Accuracy of Regression Line

1F-4c Linearity of Relationship

1F-4d Comparing Regression Lines

The statistical techniques are illustrated by application to the data shown in Table 1F-4a*. Detailed work sheets (1F-4a to 1F-4f) are given showing the step-by-step computations to be made.

1F-4a Calculation of Regression Line

Two alternative methods for calculating the regression line are given in decreasing order of complexity.²

- (a) "Exact"—This method can be applied to the most general type of data, e.g., when each sample unit is tested at a different temperature. The parameters of the line are given by:

$$b = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2} =$$

$$\frac{(X_1 - \bar{X})(Y_1 - \bar{Y}) + (X_2 - \bar{X})(Y_2 - \bar{Y}) + (X_3 - \bar{X})(Y_3 - \bar{Y}) + \dots}{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + (X_3 - \bar{X})^2 + \dots} \quad (2)$$

$$a = \bar{Y} - b\bar{X} \quad (3)$$

where $\bar{X}$ and $\bar{Y}$ are the averages of the X's and Y's, respectively, i.e.,

$$\bar{X} = \frac{1}{N} \sum X \quad (4)$$

$$\bar{Y} = \frac{1}{N} \sum Y \quad (5)$$

where N is the total number of observations.

An alternative form of equation (2) is usually more suitable for computation:

$$b = \frac{N \sum XY - \sum X \sum Y}{N \sum X^2 - (\sum X)^2} \quad (2a)$$

After the values of a and b have been calculated by Equations (3) and (2) or (2a), the regression line can be plotted by selecting arbitrary values X and computing the corresponding values of Y from:

$$Y = a + bX$$

These points, representing corresponding values of X and Y , are plotted and a straight line drawn through them. Work Sheet 1F-4a, "Linear Regression of Paired Data", gives the step-by-step calculation of these quantities and an ex-

ample has been included. On Figure 1F-4, the data and the calculated regression line have been plotted.

- (b) "Equal Testing"—In case the same number n of sample units are tested at each of k temperatures, $X_1, \dots, X_k$, so that $N = kn$, the above formula for b can be simplified.*

$$b = \frac{\sum ZW}{\sum W^2} = \frac{\sum Z(X - \bar{X})}{\sum (X - \bar{X})^2} \quad (2b)$$

where the Z 's are the averages of (log) life at each of the temperatures, and the W 's are deviations from the mean, i.e.,

$$W = X - \bar{X} \quad (6)$$

$$\bar{X} = \frac{1}{k} \sum X \quad (4a)$$

Note that:

$$\bar{Y} = \bar{Z} = \frac{1}{k} \sum Z \quad (7)$$

The detailed method of calculation and an example of "equal testing" is found on Work Sheet 1F-4b, "Linear Regression of Grouped Data."

1F-4b Accuracy of Regression Line

The variance or the square of the standard deviation, S^2 , about the regression line must be obtained and is computed most easily from the formula:

$$S^2 = \frac{1}{N-2} \left[\sum (Y - \bar{Y})^2 - b \sum (X - \bar{X})(Y - \bar{Y}) \right] \quad (8)$$

or (8) can be rewritten in the form:

$$S^2 = \frac{1}{N-2} \left[\sum Y^2 - \bar{Y} \sum Y - b(\sum XY - \bar{X} \sum Y) \right] \quad (8a)$$

This calculation is illustrated in Work Sheet 1F-4a, Items 18 and 20. For "equal testing" in (b) above, (8a) can be written as**

$$S^2 = \frac{n}{nk-2} \left[\frac{1}{n} \sum Y^2 - \bar{Z} \sum Z - b \sum WZ \right] \quad (8b)$$

which is calculated for the example in Work Sheet 1F-4b, Items 18 and 20.

It is rather common to use "sum of square" (SS) terminology, as follows (see Hald³, p. 537):

$$SS \text{ Total} = \sum (Y - \bar{Y})^2$$

$$SS \text{ Regression} = b \sum (X - \bar{X})(Y - \bar{Y})$$

$$SS \text{ Error} = SS \text{ total} - SS \text{ Regression}$$

*The insulation life y , given in the column marked "Hours", is the midpoint of the time interval between time failure observed and time of last previous test.

**If each of the sample units in any group are at a slightly different temperature, the mean temperature could be used. In this section, all summations on W and Z go from 1 to k .

**It is more convenient in the computation sheet to work with S^2/n ; also with $SS' = SS/n$.