ADA058102

AFML-TR-78-27

LEVEL II

A TEST OF HOMOGENEITY BASED ON MAXIMUM LIKELIHOOD ESTIMATES

AD No. ___
DDC FILE COPY

AVENUE d'ALBIGNY, 9 BIS
74000 ANNECY
FRANCE

APRIL 1978

TECHNICAL REPORT AFML-TR-78-27
Final Report for Period 15 December 1973 – 15 January 1974

AIR FORCE MATERIALS LABORATORY
AIR FORCE WRIGHT AERONAUTICAL LABORATORIES
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This report has been reviewed by the Information Office (OI) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

Robert C. Donat

ROBERT C. DONAT
Metals Behavior Branch
Metals and Ceramics Division

FOR THE COMMANDER

LAWRENCE N. HJELM
Acting Chief
Metals Behavior Branch
Metals and Ceramics Division

"If your address has changed, if you wish to be removed from our mailing list, or if the addressee is no longer employed by your organization please notify AFML/LLN, WPAFB, OH 45433 to help us maintain a current mailing list."

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFML-TR-78-27
2. GOVT ACCESSION NO:
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): A TEST OF HOMOGENEITY BASED ON MAXIMUM LIKELIHOOD ESTIMATES
5. TYPE OF REPORT & PERIOD COVERED: Final Report, for Period 15 Dec 1973 - 15 Jan 1974
6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Waloddi Weibull
8. CONTRACT OR GRANT NUMBER(s): F44620-73-C-0066
9. PERFORMING ORGANIZATION NAME AND ADDRESS: AFML/LLN, Wright-Patterson AFB, OHIO 45433
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 735106, 62102F, 7351
11. CONTROLLING OFFICE NAME AND ADDRESS:
12. REPORT DATE: April 1978
13. NUMBER OF PAGES: 8
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 12p
15. SECURITY CLASS. (of this report): UNCLASSIFIED
15a. DECLASSIFICATION DOWNGRADING SCHEDULE:
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):
18. SUPPLEMENTARY NOTES:
19. KEY WORDS (Continue on reverse side if necessary and identify by block number):
structural failure probability
cumulative fatigue damage
fatigue failure distribution
structural reliability
20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

A sample is said to be homogeneous, if all its elements are drawn from the same population, otherwise heterogeneous. It is self-evident that there is no sense in estimating the parameters of an assumed distribution from heterogeneous samples. Considering the fact that samples of fatigue test data are frequently composed of elements drawn from two or even three populations, it is an important rule, frequently violated, to start statistical analyses of a given sample by stating whether it is homogeneous or not.

DD FORM 1473, 1 JAN 73 — EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

322 000

A simple test based on the alternative maximum likelihood estimates of the complete sample and of the sample more or less truncated is presented.

FOREWORD

The research work reported herein was conducted by Prof. Dr. Waloddi Weibull, Avenue d'Albigny, 9 bis, 74000 Annecy, France, under USAF Contract No. F44620-73-C-0066. This contract, which was initiated under Project No. 7351, "Metallic Materials", Task 735106, "Behavior of Metals", was administered by the European Office of Aerospace Research. The work was monitored by the Metals and Ceramics Division, Air Force Materials Laboratory, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio 45433, under the direction of Mr. W. J. Trapp, AFML/LL.

This report covers work conducted during the period 15 December 1973 to 15 January 1974. The manuscript was submitted by the author for publication in March 1974.

TABLE OF CONTENTS

PRECEDING PAGE BLANK

1. INTRODUCTION

A sample is said to be homogeneous, if all its elements are drawn from the same population, otherwise heterogeneous. It is self-evident that there is no sense in trying to estimate the parameters of an assumed distribution function from heterogeneous samples. Considering the fact that samples of fatigue test data are quite frequently composed of elements drawn from two or even three different populations, it is an important rule, frequently violated, to start the statistical analysis of a given set of test data by stating whether the underlying samples are homogeneous or not.

The chi-square test has been used for such investigations, but this test requires very large samples. The present test is quite simple, and it is not restricted with regard to the size of the sample.

2. THE PRINCIPLE OF THE TEST

Consider a sample of size N with its elements $x_1,\ldots,x_N$ rearranged in ascending order of magnitude. If all these order statistics belong to the same population, and the unknown parameters are estimated by use of a reliable method, say, the maximum likelihood method, applied alternatively to the complete sample and to (N - 1) of its elements, for instance, by excluding the largest order statistic x_N, then no significant differences between the two sets of estimates will be expected. If a significant difference appears, then it may be concluded that the value x_N does not belong to the same population as the main part of the other (N - 1) order statistics.

Let $F(x,m,\beta,\mu)$ be the distribution function of the population and let $f(x,m,\beta,\mu)$ be the density function of a single value drawn from it, where m, β, and μ are parameters.

Following formulae presented by Sarhan and Greenberg (1), the density function of the joint distribution of the first k order statistics, which will be called the likelihood equation of the k-truncated sample and denoted by L_k, is given by

$$L_k = \frac{N!}{(N-k)!}\{1 - F(x_k)\}^{N-k} \cdot f(x_1) \ldots f(x_k) \qquad (1)$$

which for k = N takes the form

$$L_N = (N!)\, f(x_1)\ldots f(x_N) \qquad (2)$$

The present test now consists in computing the particular set of estimates $(\hat{m},\hat{\beta},\hat{\mu})$, which maximizes the likelihood equations L_N and L_{N-1}, possibly also L_{N-2}, etc., and stating whether there are significant differences between the alternative sets of estimates. The term "significant" will be defined in the sequel.

The further calculations will be applied to the Weibull distribution.

3. APPLICATIONS TO THE WEIBULL DISTRIBUTION

3.1 General Formulas

In this particular case, we have

$$P = F(x,m,\beta,\mu) = 1 - e^{-(x-\mu)^m/\beta^m} \tag{3}$$

and

$$f(x,m,\beta,\mu) = (m/\beta^m)(x-\mu)^{m-1} \cdot e^{-(x-\mu)^m/\beta^m} \tag{4}$$

Introducing (4) into (1), we have

$$L_k = \frac{N!}{(N-k)!}(m^k/\beta^{km})\,((x_1-\mu)\ldots(x_k-\mu))^{m-1} \cdot$$
$$\cdot e^{-((x_1-\mu)^m + \ldots + (N-k+1)(x_k-\mu)^m)} \tag{5}$$

and

$$\ln L_k = \ln\frac{N!}{(N-k)!} + k(\ln m) - km\,(\ln\beta) + (m-1)(\ln(x_1-\mu) + \ldots$$
$$+ \ln(x_k-\mu)) - ((x_1-\mu)^m + \ldots + (N-k+1)(x_k-\mu)^m)/\beta^m \tag{6}$$

From the condition $\partial L_k/\partial\beta = 0$, we obtain

$$\beta^m = ((x_1-\mu)^m + \ldots + (N-k+1)(x_k-\mu)^m)/k \tag{7}$$

Introducing (7) into (5) we have, neglecting factors depending only on N and k,

$$L_k = \frac{m^k((x_1 - \mu) \ldots (x_k - \mu))^{m-1}}{((x_1 - \mu)^m + \ldots + (N - k + 1)(x_k - \mu)^m)^k} \tag{8}$$

which for k = N takes the form

$$L_N = \frac{m^k((x_1 - \mu) \ldots (x_N - \mu))^{m-1}}{((x_1 - \mu)^m + \ldots + (x_N - \mu)^m)^N} \tag{9}$$

The unknown parameters m and μ may be estimated by equating to zero the partial derivatives of L_k, that is

$$\partial L_k / \partial m = 0, \qquad \partial L_k / \partial \mu = 0 \tag{10}$$

3.2 Criteria for Significant Differences

3.2.1 Parameter μ known

Without loss of generality, we may put μ = 0 and equation (8) then takes the form

$$L_k = \frac{m^k(x_1 \ldots x_k)^{m-1}}{x_1^m + \ldots + (N - k + 1)x_k^m)^k} \tag{11}$$

and

$$\ln L_k = k(\ln m) + (m - 1)(\ln x_1 + \ldots + \ln x_k) - k\ln (x_1^m + \ldots$$

$$+ (N - k + 1)x_k^m) \tag{12}$$

The likelihood L_k will be maximized by the particular value $\hat{m}$ which satisfies the equation

$$\partial \ln L_k / \partial m = 0 \tag{13}$$

that is,

$$\frac{x_1^{\hat{m}}(\ln x_1^{\hat{m}}) + \ldots + (N-k+1)x_k^{\hat{m}}(\ln x_k^{\hat{m}})}{x_1^{\hat{m}} + \ldots + (N-k+1)x_k^{\hat{m}}} - \frac{\ln x_1^{\hat{m}} + \ldots + \ln x_k^{\hat{m}}}{k} = 1 \tag{14}$$

The sampling distribution of the estimate $\hat{m}$ may be determined by generating a large number of random samples from the actual population and computing the corresponding values of $\hat{m}$ for each of them by use of equation (14). Observing that the probability P in equation (3) is uniformly distributed over the interval (0,1), the elements x_i of a sample from a Weibull population with the parameters (m_o, β_o) is obtained by taking a sample of random sampling numbers $r_i(0,1)$ and computing x_i from

$$x_i = \beta_o(-\ln(1 - r_i))^{1/m_o} \tag{15}$$

Introducing (15) into (14), the parameter β_o disappears and $x_i^{\hat{m}}$ will be replaced by

$$x_i^{\hat{m}} = (-\ln(1 - r_i))^{\hat{m}/m_o} \tag{16}$$

which implies that the sampling distribution of $\hat{m}/m_o$ is, for any given value of k, parameter-free and thus uniquely given for each combination of N and k.

Denoting the estimate of m for k = N by $\hat{m}_1$, for k = N - 1 by $\hat{m}_2$, for k = N - 2 by $\hat{m}_3$, etc., it can be concluded that the quotients $\hat{m}_1/\hat{m}_2$, $\hat{m}_2/\hat{m}_3$, etc., have sampling distributions which are independent of the true population parameter m_o and uniquely given for each sample size N.

From these distributions, quotients $\hat{m}_1/\hat{m}_2$, $\hat{m}_2/\hat{m}_3$, corresponding to preassigned levels of significance, say, 2% or 5%, will be determined. The hypothesis of homogeneity will be rejected for samples yielding larger quotients.

3.2.2 All Parameters Unknown

If none of the parameters is known, it has been found convenient to estimate the parameters by computing the values of L_k or L_N in equations (8) and (9) for an appropriate set of pairs (m,u) and to select that particular pair (m,μ) which yields the largest value of L_k or L_N.

The computer program 6/73 has been written for computing L_N. Observing that

$$0 \geq \mu \leq x_1 \qquad (17)$$

this : gram c. ıtes the eleven values of m which maximize L_N for

$$\mu = x_1(i/10) \text{ , } \qquad i = 0(1)10 \qquad (18)$$

and corresponding values of log L_N and $= x_o$.

The computing time of this procedure is about one second for sample size N = 10 with the computer IBM 360, M/75.

This program will now be extended to produce the sam set of data for k = N - 1, N - 2, etc., and to present the eleven quotients $\hat{m}_1/\hat{m}_2$, etc. In cases where all the quotients are larger than the assigned levels of significance, the hypothesis of homogeneity will be rejected. If some of the quotients have acceptable values, the decision depends on the location of these values. If they are far from the place of maximum likelihood, they will motivate a rejection. If not, the hypothesis may be accepted or examined by other tests.

REFERENCES

1. Sarhan, A. E. and Greenberg, B. G., "Contributions to Order Statistics", John Wiley & Sons, New York and London, 1962.

*U.S.Government Printing Office: 1978 — 757-080/48