SECTION H1

STATISTICAL METHODS

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF SYMBOLS COMMONLY USED IN STATISTICS

Symbol	Definition	
$C\{x\}$	Coefficient of variation $= \frac{\sigma\{x\}}{M\{x\}}$	[1]
C.D.F.	Cumulative distribution function	[2]
Cov	Covariance	[2]
C_r^n	$\frac{n!}{(n-r)!\,r!} = \binom{n}{r}$	[2]
D.F.	Degrees of freedom	[2]
df	Distribution function/degrees of freedom	
Exp X	e^x	[2]
"F"	F ratio or variance ratio	[2]
f	(N-1) Degrees of freedom; fraction of parts sampled; frequency	[2]
k	A standardized variable expressing the dispersion about the mean in terms of σ	[3]
k_P	A constant for a specified probability level which correlates population density with standard deviation; sometimes called a standardized variable, applies to normal distributions	[3]
$k_{P\gamma}$	Composite probability factor; a standardized variable relating P, γ, and n to some limiting value of the variable, applied to normal distributions	[3]

LIST OF SYMBOLS COMMONLY USED IN STATISTICS (Continued)

Symbol	Definition	
$M\{x\}$	The mean of a stochastic variable x	[1]
M.G.F.	Moment generating function	[2]
Mo	Mode	[2]
m	Population mean	[4]
N	Sample size [2]; number of loading cycles to failure [3]	
N_e	An arbitrary lifetime to which fatigue test data are to be extrapolated	[3]
N_i	Same relationship to N as x_i has to x	[3]
$N_{p\gamma}$	Lower limiting life above which any test value can be expected to fall with a probability of P and a confidence of γ	[3]
$\bar{N}_s$	Arithmetic mean of test sample lives	[3]
$\bar{N}_u$	Arithmetic mean life that could be expected if an infinite number of samples could be tested	[3]
N_4	Number of loading cycles at which weakest of four specimens fails	[3]
$\bar{N}_4$	Arithmetic mean of least-of-four test sample lives	[3]

LIST OF SYMBOLS COMMONLY USED IN STATISTICS (Continued)

Symbol	Definition	
n	Number of independent single test specimens	[3]
$n!$	Factorial $n = 1 \times 2 \times \ldots \times (n-1) \times n$	[2]
n_4	Number of least-of-four test points	[3]
$\binom{n}{r}$	$C\binom{n}{r} = \frac{n!}{(n-r)!\, r!}$	[2]
P	Probability; the percent of a group of specimens expected to fall within a certain range	[3]
$P(x)$	Probability of an event x occurring	[2]
P_4	Probability of four consecutive test values exceeding some specified limiting value	[3]
R.M.S.	Root mean square	[2]
r	Correlation coefficient	[2]
S	Loading stress level at which failure occurs at some number of cycles (N) in fatigue work	[3]
S_a	A value of stress on the estimated average S-N curve corresponding to some arbitrary test point life, N_i	[3]
S_D	Standard deviation of a sample	[2]
S_e	A value of stress on the estimated average S-N curve corresponding to the lifetime (N_e) to which fatigue data are to be extrapolated	[3]

LIST OF SYMBOLS COMMONLY USED IN STATISTICS (Continued)

Symbol	Definition	
S_i	An arbitrary value of stress occurring in a particular problem at lifetime N_i	[3]
S_{ie}	A derived stress value at lifetime (N_e) corresponding to an observed stress value S_i at lifetime N_i	[3]
S_s	Standard error of the standard deviation	[2]
S^2	Variance of a sample	[2]
"t"	Student's "t" statistic	[2]
u	Standardized variable $= \frac{x - \xi}{\sigma}$	[1]
V	Coefficient of variation	[2]
w	Often used for range	[2]
x	A random statistical variable	[3]
$\bar{x}$	Arithmetic mean of x-values regardless of the number of values involved	[3]
x_i	Any arbitrary value of x occurring in a specified problem	[3]
x_P	A variable expressed as a function of P and σ	[3]
$x_{P\gamma}$	A limiting value of x dependent on p, γ, n, and σ	[3]

LIST OF SYMBOLS COMMONLY USED IN STATISTICS (Continued)

Symbol	Definition	
$\bar{x}_s$	Arithmetic mean of values from a limited sample size	[3]
$\bar{x}_u$	Arithmetic mean that could be expected from an infinite number of specimens	[3]
y	Frequency of occurrence of test points in given intervals of the variable (Δx)	[3]
y'	Percentage frequency; percentage of total number of test points in a given variable increment	[3]
α	α risk, Type I error	[2]
α_{P_1}	Limiting value of standardized variable for a given probability and an unknown distribution	[3]
α_γ	Limiting value of standardized variation of the mean for a given confidence and an unknown distribution	[3]
β	β risk, Type II error; also equals $(1 - \alpha)$	[2]
γ	Confidence; the percentage of sample mean values falling within a given range of the universe mean [3]; associated with the tolerance limit tables [2]	
μ	Population mean	[2]
ξ	Mean of a stochastic variable $x = M\{x\}$	[1]
σ	Standard deviation of variable x about $\bar{x}$	[3]

LIST OF SYMBOLS COMMONLY USED IN STATISTICS (Concluded)

Symbol	Definition	
σ_e	Standard deviation of a limited number of derived stress points about S_e	[3]
$\sigma_{\bar{N}_4}$	Standard error of least-of-four mean ($\bar{N}_4$)	[3]
σ_s	Standard error of the standard deviation [2]; standard deviation computed from a limited sample size [3]	
σ_u	Unbiased standard deviation of the universe; corresponds to an infinite sample size	[3]
σ_{u_4}	Standard deviation of universe of least-of-four points	[3]
$\sigma_{\bar{x}}$	Standard deviation of $\bar{x}_s$, about $\bar{x}_u$; sometimes called the standard error of the mean	[3]
σ_4	Standard deviation of least-of-four test failure points	[3]
σ^2	Variance of a population	[2]
χ^2	Chi-square	[2]

LIST OF DEFINITIONS COMMONLY USED IN STATISTICS

	Definition
Alternative Hypothesis	Possible true alternate answer to the hypothesis being statistically tested. The larger the sample size, the greater the possibility of rejecting a hypothesis when an alternate answer is true [2].
AOQL (Average Outgoing Quality Limit)	Upper limit on outgoing quality that may be expected in the long run, when all rejected lots are subjected to 100 percent inspection, with all defective articles removed and replaced by good articles [2].
A Posteriori Probability	If in a number of trials an event has occurred N times and failed M times, the probability of its occurring in the next trial is $\frac{N}{M + N}$ [2].
A Priori Probability	Let N be the number of exhaustive, mutually exclusive, and equally likely cases of an event under a given set of conditions. If M of these cases are known as the event A, then the mathematical, or a priori probability of event

LIST OF DEFINITIONS COMMONLY USED IN STATISTICS (Continued)

	Definitions
	A occurring under the given set of conditions is M/N [2].
AQL (Acceptable Quality Level	Percentage of defective items in an inspection lot that a sampling plan will accept with (in the usual case) an associated α Risk of 0.05 [2].
Biased Sample	If some individuals in the Universe are more likely to be chosen than others, the sample is said to be biased.
Class Interval	When the number of observations is large, the range of the data can be broken into a limited number of segments of equal length. The segments are known as class intervals or cells [2].
Confidence Intervals	This provides a method of stating how close an estimate is to the true value [2]. It is the interval associated with a prescribed confidence coefficient. The confidence coefficient is the proportion of samples of size n for

	Definition
	which intervals computed by the prescribed methods may be expected to bracket a value [4].
Continuous Distribution	One in which the only limit to size of intervals measured is the sensitivity of the measuring apparatus [2].
Cumulative Distribution	Indicates by its magnitude the proportion of the Universe (or sample) to the left of that point [2].
Curve Fitting	Method utilizing computed statistics as approximate parameters for theoretical distributions [2].
Degree of Freedom	Number of free variables (unrestricted and independent in the sense of random sampling) entering into a statistic. In the case of a sample of size N, from a Universe, there are N-1 degrees of freedom [2].
Discrete Distribution	If a random variable has only a finite number of possible values, then it will form a discrete distribution [2].

LIST OF DEFINITIONS COMMONLY USED IN STATISTICS (Continued)

	Definition
Double Sampling	Involves the possibility of putting off the disposition of an inspection lot until a second test sample is taken. A lot will be accepted on the basis of the first sample if the results are very good or will be rejected if the results are very poor. If the results from the first sample are of a borderline nature (between good and poor), a second sample must be taken. On the basis of the results of the combined first and second samples, the lot is either accepted or rejected [2].
Error	In statistics there are two types of error. If we reject the null hypothesis when it is true, then we make an Error of the First Kind. If we fail to reject a null hypothesis when it is false, then we make an Error of the Second Kind [4].
"F" Distribution	Sampling distribution of the variance [2].

LIST OF DEFINITIONS COMMONLY USED IN STATISTICS (Continued)

	Definition
Frequency Table	Tabulation of the number of observations that occur in each class interval of a histogram [2].
Histogram	Block representation of data arranged to show the dispersion of the data [2].
Hypothesis	Statement formulated in such a way that it may be refuted through statistics.
Inference	Based on the theory of probability, statistical inference is that mathematical framework which supplies a technique for description, prediction, and rational design decisions despite the complications which arise because of variation [2].
Latin Square	"Analysis of Variance" ordering technique of observed values in an experiment, allowing control of several sources of variability [2].
Least Squares	Method based upon the principle that the best value of a quantity that can be deduced from a set of measurements or observations is that for which the sum of the squares of the deviations of the observations (from it) is a minimum [2].

LIST OF DEFINITIONS COMMONLY USED IN STATISTICS (Continued)

	Definition
Lot	Group of manufactured articles which are essentially alike, such as 1 day's production [2].
LTPD (Lot Tolerance Percent Defective)	Usually refers to the incoming quality, above which there is a small chance that a lot will be accepted. It is usually taken with a consumer's risk of $\beta = 0.10$ [2].
Mean	The arithmetic average of a group of observations [2]. It is the location parameter of a normal distribution locating the "center of gravity" of the distribution [4].
Mean Deviation	Arithmetic mean of the absolute distance of each observation from the mean [2].
Median	Middle value of a group of observations. In the case of an even numbered set of observations, it is the average of the middle pair [2].
Midrange	Arithmetic average of the extreme values of a set of observations [2].
Mode	Most frequent value of a set of observations [2].

LIST OF DEFINITIONS COMMONLY USED IN STATISTICS (Continued)

	Definition
Moments	In statistics, moments are analogous to moments in mechanics in several ways. Just as some bodies are completely characterized by their moments, some probability distributions are completely characterized by their moments. The first moment about the origin is equivalent to the expected value (the mean) of the random variable. The first moment is also the center of gravity of the probability mass. The second moment above the mean is also known as the moment of inertia and variance [2].
Nonparametric Statistics	Statistical techniques developed to test hypotheses without the assumption of normality, or any other assumption, other than that of continuity of a distribution [2].
Normal (Gaussian) Curve	Bell-shaped curve from the Gaussian probability distribution. It is a two parameter distribution requiring the mean and the variance for its description [2].

LIST OF DEFINITIONS COMMONLY USED IN STATISTICS (Continued)

	Definition
Normal Probability Paper	Special graph paper which reduces the cumulative normal curve to a straight line. The log-normal probability paper does the same with log values as the normal paper does with linear values [2].
OC Curves (Operating Characteristics Curves)	Give the chance of accepting a hypothesis [2].
Population	Any set of individuals (or objects) having some common observable characteristic. It is synonymous with Universe [2].
Proof	Differs from mathematical proof as it does not fall within the framework mathematics, but results from experimentation, with an accompanying probability statement [2].
Random Digits	Digits picked in such a way that each digit has an equal chance of occurring at any time [2].
Randomization	Assignment of a sequence of operations to a test program by the use of some technique, such as random tables, to avoid bias in the test results [2].

LIST OF DEFINITIONS COMMONLY USED IN STATISTICS (Continued)

	Definition
Random Sample	Picked in such a way that all members of the population have an equal chance of selection [2].
Range	Absolute difference between the extreme values of a set of observations [2].
Root Mean Square	Square root of the average of the sum of the squares of a set of observations [2]. $$RMS = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N}}$$
Sample	Set of observations chosen from a population [2].
Sampling Distribution	Distribution of a statistic in the set of all samples of a specific size from a given Universe [2].
Sequential Samplings	Acceptance plans permitting from three, up to an unlimited number of samples [2].
Significance Level	This expresses our reluctance to give up or "reject" the null hypothesis and is given by the magnitude of the α Risk. The

LIST OF DEFINITIONS COMMONLY USED IN STATISTICS (Continued)

	Definition
	smaller the magnitude of significance the less we are willing to reject the null hypothesis [4].
Single Sampling	When the decision as to the disposition of an inspection lot is always made on the evidence of only one sample, the acceptance plan is described as a single sampling plan [2].
Standard Deviation	Positive square root of the variance [2].
Statistic	Estimation of a population parameter, computed entirely from a sample [2].
Stochastic Variable	In general, any variable which may have a probability function is a chance or stochastic variable, even though the frequency function is not known [2].
"t" Distribution	Sampling distribution of the mean [2].
Tolerance Limits	Limits such that a certain portion of the population shall lie within or above them with a specified probability [2].
Unbiased Estimate	Estimate of a parameter which has been corrected for sample size effects and is equivalent to the total population parameter [3].

LIST OF DEFINITIONS COMMONLY USED IN STATISTICS (Concluded)

	Definition
Universe	Comprised of any set of individuals having some common observable characteristic [2].
Variables Sampling	When a record is made of an actual measured quality characteristic, such as a dimension expressed in thousandths of an inch, it is known as variables sampling [2].
Variance	Sum of the squares of the deviations from the mean divided by the number of observations less than one [2].
α Error	Risk of rejecting a true hypothesis. This is also known as an α Risk, Consumer's Risk, or Type I [2].
β Error	Risk of accepting a false hypothesis. This is also known as β risk, Producer's Risk, or Type II Error [2]. Equals $(1 - \alpha)$.

H1 STATISTICAL METHODS

1.1 INTRODUCTION

"One of the main objectives of statistics is to give a mathematical description of observed data in such a manner that the observed phenomena and the method of observation are characterized by a few numbers" [1]. While a single observation cannot be reproduced, experience has shown that a set of observations, resulting from the repetition of some process, produces certain characteristic features which can be reproduced; and it is these characteristic features that statistics attempt to describe.

Various methods of describing characteristic features have been developed, and one method includes the use of a histogram. Consider a hypothetical set of recorded test values as being laid off in ascending order of magnitude along a horizontal axis to form the abscissa of a graph. Now let the range of values be divided into equal intervals and a count made of the number of test points in each interval. The number of points per interval is known as the frequency. If the various frequencies are now laid off on a vertical scale, a histogram is produced (Fig. H1-1). From the histogram a frequency distribution curve may be obtained by fairing a smooth line through the shape of the histogram (Fig. H1-2). Several frequency distribution curves (Normal, Log-Normal, χ^2, "t", Weibull) have been established, but only two (Normal and Log-Normal) will be discussed here.

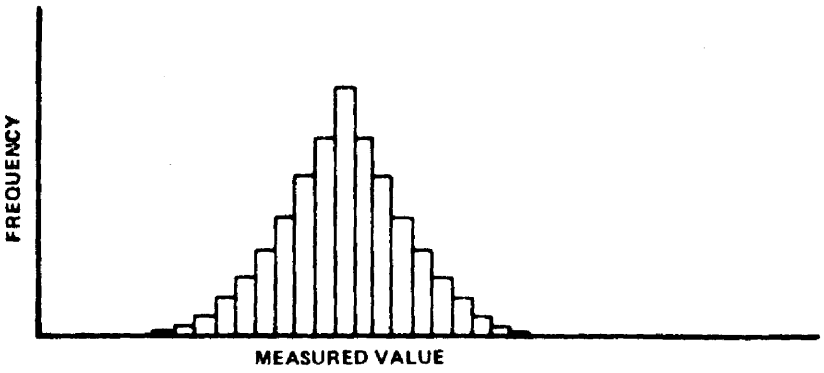

FIGURE H1-1. HISTOGRAM.

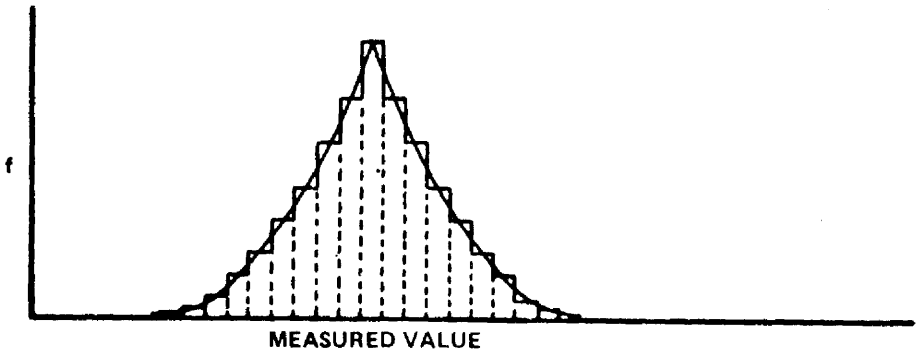

FIGURE H1-2 DISTRIBUTION FUNCTION FROM "FAIRED" HISTOGRAM

The statistical methods discussed will be limited to those methods necessary for evaluating material, fatigue, and fracture mechanics data. It has been found that material data obey normal distribution and that fatigue [3] and fracture [5] data obey log-normal distributions.