Technical Memorandum TRAC-F-TM-0390
May 1990

THE APPLICATION OF EXPLORATORY DATA ANALYSIS METHODS IN COMPUTING SCREENING INTERVALS FOR SELECTED STUDY MEASURES OF EFFECTIVENESS

Fort Leavenworth

U.S. ARMY
TRADOC ANALYSIS COMMAND-FORT LEAVENWORTH
OPERATIONS DIRECTORATE
FORT LEAVENWORTH, KANSAS 66027-5200

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION NCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for Public Release; Distribution unlimited
4. PERFORMING ORGANIZATION REPORT NUMBER(S) RAC-F-TM-0190	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION SA TRADOC Analysis Command-ort Leavenworth 6b. OFFICE SYMBOL (If applicable) ATRC-FOQ	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) TTN: ATRC-FOQ ort Leavenworth, KS 66027-5200	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING / SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
he Application of Exploratory Data Analysis Methods in Computing Screening Intervals or Selected Study Measures of Effectiveness

12. PERSONAL AUTHOR(S)
r. Rudolph J. Pabon

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
inal	FROM May 89 TO May 90	90 May 30	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Exploratory data analysis, batch, screening intervals, prediction intervals, MOE, outliers,

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report presents a methodology for the construction of screening intervals for elected study measures of effectiveness (MOE): force loss exchange ratio, helicopter, and ank system exchange ratios. Using exploratory data analysis techniques, measures of ocation and scale are derived from study MOE data.

Two methodologies are presented: A test for specious data and a test for comparing two atches of data. The first methodology is based on the biweight estimator of location and he second is based on the median of the batch.

The first methodology is to be incorporated into an expert system currently under development at the TRADOC Analysis Command-Fort Leavenworth (TRAC-FLVN). The expert system is esigned to fill a quality control need for comparing emerging study results with past-elated study MOE data to insure validity of study results. Prediction intervals are roposed for use as a screening tool to determine the "acceptability" of new MOE data. ~~(Over)~~

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code) (816)684-2460 22c. OFFICE SYMBOL ATRC-FOQ

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE lock 19. (Continued)

his will provide statistical validity and quality control to emerging study results. he second methodology is for use when two batches of data are to be compared. Both methodologies will serve as screening tools for the study analyst.

Technical Memorandum TRAC-F-TM-0390
MAY 1990

TRADOC Analysis Command-Fort Leavenworth (TRAC-FLVN)
Operations Directorate
Fort Leavenworth, Ks 66027-5200

The Application of Exploratory Data Analysis Methods in Computing Screening Intervals for Selected Study Measures of Effectiveness

by

Rudolph J. Pabon

ACN 9100

CERTIFIED BY:

ERNEST D. BOEHNER
Director, OD

APPROVED BY:

ROBERT LA ROCQUE
Director, TRAC-FLVN

Distribution Statement:
Approved for public release; distribution is unlimited

ACKNOWLEDGMENTS

I would like to express my appreciation to Mr. Howard Haeker and Mr. Ron Magee whose combined experience and guidance assisted greatly in focusing my attention in discovering the need for a quality control tool in the study analysis area.

Also, my thanks to Dr. Leon Godfrey who also had a role in articulating the study analyst requirements and provided a another view of the quality control need. Finally, to Mr. Robert Brown who has taken the decision logic and programed them into the AI system.

TABLE OF CONTENTS

TABLES

FIGURES

ABSTRACT

This paper presents a methodology for the construction of screening intervals for selected study measures of effectiveness (MOE): force loss exchange ratio, helicopter, and tank system exchange ratios. Using exploratory data analysis techniques, measures of location and scale are derived from study MOE data.

Two methodologies are presented: a test for specious data and a test for comparing two batches of data. The first methodology is based on the biweight estimator of location and the second is based on the median of the batch.

The first methodology is to be incorporated into an expert system currently under development at the TRADOC Analysis Command - Fort Leavenworth (TRAC-FLVN). The expert system is designed to fill a quality control need for comparing emerging study results with past-related study MOE data to insure validity of study results. Prediction intervals are proposed for use as a screening tool to determine the "acceptability" of new MOE data. This will provide statistical validity and quality control to emerging study results. The second methodology is for use when two batches of data are to be compared. Both methodologies will serve as screening tools for the study analyst.

1. Introduction.

a. One of the primary missions of the TRADOC Analysis Command-Fort Leavenworth (TRAC-FLVN) is to provide analytical support to the conduct of Department of the Army (DA) studies that assess present and future army warfighting capabilities for a given scenario. (A scenario dictates the geographical location of the battle, the deployment and intended mission(s) of opposing forces, the year and respective force strengths, e.g., numbers and types of weapon systems, and describes the weather and terrain environment.) The purpose of these studies is to evaluate the cost-effectiveness of candidate "improvements" in doctrine, training, leadership, organization, and materiel. The decision maker uses the analysis findings to assist him in reaching a decision on the benefits that these changes will provide the U.S. Army in accomplishing its mission.

b. The Vector-In-Commander (VIC) corps-level computer combat model is one of the primary analysis tools used by TRAC-FLVN to model and assess the corps-level combat effectiveness of opposing forces. The measures of effectiveness (MOE) that are used to answer specific study questions, i.e., essential elements of analysis, are often identical from study to study. Selected weapon system MOE from past VIC supported studies can be compiled for establishing a library of past study results, providing the analyst an historical reference of, for example, weapon system contribution to force effectiveness. More importantly, selected MOE can be used for quality control and quality assurance purposes.

c. A TRADOC Analysis Command memorandum, reference 14, defines quality control as the process of guaranteeing technical soundness and integrity in individual studies as well as consistency among related studies. Quality assurance is viewed as the process of ensuring quality control with the goal of satisfying study requirements, i.e., the completed study product satisfies the study objectives. To this end, study MOE can be used as a basis for developing a quality control screening tool. Such a screening tool would enable the study analyst to evaluate the reasonableness and acceptability of emerging study MOE results by comparing them with past similar scenario study MOE results.

d. The Programming and Quality Assurance Division (PQAD) of Operations Directorate, TRAC-FLVN has the primary responsibility of performing quality assurance for all TRAC-FLVN's studies. This responsibility is performed prior to the execution of the study, through a review of the study plan, and upon completion of the study. When a study is completed, PQAD personnel are required to review the draft study report and evaluate, along with other aspects of the study, the study MOE results for reasonableness and consistency with past studies. This review is conducted by experienced analysts who rely on their study experience and expertise to judgementally perform quality assurance.

2. Problem. Study team personnel are charged with the responsibility of conducting internal quality control during the course of the study. Unfortunately, study team personnel are more often than not under very short milestone schedules; consequently, their concern for internal quality control is often overlooked. The conduct of an extensive literature search for specific MOE results generated in previous and related studies is very time consuming. As a result, the quality control activity of comparing emerging study results with previous study efforts is usually omitted; as a consequence, incorrect analysis sometimes follows. Conducting quality control at or near the completion of the study is very inefficient and leads either to a correction of study analysis already completed or a product that is less than required to meet the study objectives. In either situation, the expenditure of scarce resources, e.g., manpower and computer utilization time, to correct the analysis is counter-productive and not cost-effective.

3. Background. TRAC-FLVN Operations Directorate personnel are applying artificial intelligence (AI) technology to develop an expert system for use as a quality control tool using study MOE. Mr. Robert Brown, Computer Systems Division, Operations Directorate, is the individual who is programming the expert system. The resident data base of this system is to consist of past study force and system category-specific study MOE. The purpose of the AI system is to provide the study analyst a user-friendly interactive tool for use as a comparative benchmark of past study MOE data values. In support of this AI effort, PQAD personnel have started to collect selected MOE scenario and study data for the purpose of compiling a data base of VIC model generated MOE data. The analyst will be able to investigate the range of selected MOE data values for a given set of combat conditions. The expert system is being designed to provide the user the maximum and minimum values of past study MOE data contained in the system. The benefits of the AI system can be significantly improved with the formulation of analytical underpinnings to provide the user a statistical basis for comparing new study MOE values with past MOE data.

4. Objectives. The objective of this paper is to develop the mathematical algorithms for use in constructing screening intervals for MOE data. These screening intevals will be incorporated into the rule-based AI system under development. The AI system is being designed to cite the specific data and

model areas that the study analyst should review for a suspicious MOE result. These intervals will provide the user statistically-derived interval estimates that will serve as an analytical basis for quality control. The benefits to be gained from this paper are:

a. Analytically-based algorithms that will provide the expert system user with statistically based and study credible guides for screening data.

b. The expert system-based algorithms will, in turn, increase the overall quality of TRAC-FLVN studies.

5. Solution Approach.

a. A literature search of past VIC supported TRAC-FLVN studies was conducted. (VIC is the community accepted corps-division level combat development (CD) model and is TRAC-FLVN's primary study analysis support tool.) The MOE identified for the data base collection effort are identified at Appendix A. This paper will develop the screening interval algorithms and apply them to three sets of MOE data. The three MOE are force loss exchange ratio (LER) and helicopter and tank system exchange ratios (SER). The SER is computed as follows:

$$\frac{\text{Total Red system losses due to Blue helicopter (tank) system}}{\text{Total Blue helicopter (tank) system losses}}$$

The LER is computed as follows:

$$\frac{\text{Total Red weapon system losses}}{\text{Total Blue weapon system losses}}$$

Other MOE definitions are contained in Appendix B. The SER and LER MOE were selected because of their ready availability from past and ongoing study and scenario development efforts. These MOE are not overly sensitive to particular study issues and thus can be aggregated across studies.

b. A literature search of statistical reference material was also conducted to identify techniques available for constructing screening intervals using study MOE data. The underlying premise is that each of the selected MOE has a common distribution for given and similar sets of scenario conditions. That is, scenario conditions define a unique distribution of data values for selected MOE according to the study experience of senior analysts. The distribution of MOE output data may be considered to be bounded both below and above; thus, the distribution of MOE data possesses a mean, median, and a variance. Many classical techniques of statistical inference such as parameter estimation and hypothesis testing require, among other conditions, that the data represent random samples from a population having a known or postulated underlying distribution. However, VIC is a deterministic model and the model inputs are selectively determined, based on the study purpose and scenario, to produce "changes" in MOE. Consequently, the MOE data generated from VIC do not represent random samples. Therefore, the literature search was focused on the robust methods of distribution-free reference material not requiring the taking of a random sample. It was discovered that the theoretical concepts of Exploratory Data Analysis (EDA) techniques provide the basis for formulating screening comparison intervals for MOE data.

c. EDA statistical literature was explored to identify a summary statistic that provided a "good" measure of central tendency or location of the data, and another that provided a "good" measure of spread or variance for the MOE data. Screening interval algorithms are to be developed for the selected MOE and will be interfaced with the expert system to provide a significant enhancement to resident expert system software.

d. Specifically, the screening intervals will assist the analyst in identifying when an MOE value or set of MOE values differ from the underlying distribution of prior values for the MOE in question. They alert the analyst to several possibilities: first, the model input data related to this MOE may have been in error; second, the system or process, or the representation of it may have undergone changes which may have altered the distribution of the MOE in question; third, the set of scenario criteria is not sufficiently defined to describe a unique distribution for this MOE; and fourth, any combination of the above. In any event the analyst is forewarned that his study MOE data are suspect in terms of similar past study results.

e. The rule-based A1 system under development is designed to cite the specific data and model areas that the study analyst should review for a suspicious MOE result. See appendix C for samples of the AI decision logic and rules.

6. Methodology formulation.

a. Literature search.

(1) Review of the classical statistical literature regarding the concept of screening intervals primarily surfaced the notion of confidence intervals for estimation of population parameters using sample data. Also, the concept of predicting future data values through the use of prediction intervals was discovered.

(2) For a sample from a Gaussian distribution the confidence interval estimate, at a given α -level, for the population mean is the well known formula:

$$\bar{x} + or - t_{\alpha/2,\, n-1\, df} * s/\sqrt{n}$$

However, this standard statistical technique is based on the premise that the data represent a sample, i.e., independently and identically distributed, and come from a known class of distributions. Clearly, the data arena of VIC study MOE results do not conform to this. MOE data represent outcomes that are generated due to purposely (not randomly) changed inputs. Consequently, the use of classical statistics to estimate parameters of MOE data and subsequently make precise probability statements is not possible. Therefore, another statistical concept was needed to identify, develop, and apply a methodology to construct screening intervals for selected study and scenario MOE data.

(3) John W. Tukey and other authors have been credited with introducing a new approach to statistical analysis which has been labeled exploratory data analysis (EDA). The tools and techniques of EDA provide the means of discovering interesting or unsuspected behavior in the data. A first or exploratory phase is characterized by flexibility, both in tailoring the analysis to the data and in responding to features or patterns that successive techniques may reveal. It may be followed by a second phase, of confirmatory analysis, which is more in line with classical techniques by providing statements of significance or confidence. The steps in this confirmatory phase may involve incorporating closely related past results while analyzing new data.

(4) The purpose and objectives of a screening interval for study MOE results are consistent with the concept underlying EDA. Consequently, the empirical and heuristic methods of EDA are to be used as the basis for constructing MOE screening intervals. Tukey coined the term <u>batch</u> in his EDA text, reference 13; he used this term to describe data as "...any set of similar values, obtained however they may have been." Hoaglin, Mosteller, and Tukey, reference 7, state that the term "batch" does not include the assumptions of independence and identical distribution associated with the term "sample." Another term that is used in the EDA literature and requires mentioning is the term "estimator." In the statistical literature an estimator is a statistic used to estimate a parameter of the underlying distribution. In the EDA literature, in particular reference 7, the meaning is much broader as

follows: an estimator is a numerical function evaluated on a batch and used as a measure of some property of the source(s) of the batch.

(5) The literature search was, therefore, focused on distribution-free statistical techniques with the emphasis on the robust methods of EDA. Distribution-free techniques, according to reference 8, comprise those techniques whose validity does not depend on the underlying distribution's form and parameters. Robustness is accepted by reference 8 as the insensitivity of statistical procedures to departures from the assumptions which underlie them. Consequently, it is the intent of this paper to apply robust EDA methods to devise approximate statistical methods for use as screening tools.

(6) A first step in any analysis is the characterization of the data through the use of one or more summary statistics. The mean and median are two summary statistics that describe the center or central tendency of location of a distribution of data values. EDA techniques rely on the use of a measure of location that is resistant. Resistance is defined as the insensitivity of the estimator to one or more deviate and/or outlier data points.

(7) Barnett and Lewis, reference 1, define an outlier in a set of data to be an observation (or set of observations) which appears to be inconsistent with the remainder of that set of data. The phrase "appears to be inconsistent" is crucial. The question arises whether the deviate observation is truly a member of the underlying distribution or instead a member of a different distribution. In other words, is the apparent existence of outlier observations attributable to either an error in recording, reporting, or some other cause or is the suspicious observation actually a data point from another underlying distribution? Examination of the total process that generated the suspicious observation(s) is a logical next step.

(8) The computation of confidence intervals based on resistent location measures is addressed in the EDA statistical literature. The mean is a useful measure of location and the best one for Gaussian type distributions, e.g., short-tailed distributions, and poor for long-tailed distributions. This is because the mean is too sensitive to deviate data points. The median, however, is best for long-tailed distributions and not necessarily the best for short-tailed distributions, but adequate for symmetric distribution. The median is determined from only one or two data points; thus, it ignores some information contained in the batch. For short-tailed distributions, the median is too sensitive to small deviations in the data and not sensitive enough to each data point. (According to experienced senior study analysts, the distribution of study MOE that are

addressed in this paper are characterized by either short-tailed distributions or long-tailed distributions.) Ideally, one would like a measure that is "best" for either type of distribution.

(9) The median is widely known as possessing the desired resistance characteristic when applied to long-tailed distributions. Since the median is that value that marks the middle in rank of the batch after ordering, it is relatively simple to calculate and has become an accepted summary statistic along with the mean -- a classical parametric summary statistic. The data may be ordered in either ascending or descending order. The median can be computed in terms of its <u>depth</u>, which is defined as the relative position of data points in an ordered batch starting from either the lowest or highest end. Thus, the depth of the median, or d(m), is $d(m) = (n+1)/2$. That is, the median value occupies the d(m) position in the ordered batch. However, if the batch has an even count this depth falls between the two middle ranks of the batch, each of which has depth n/2, and which are therefore averaged to arrive at the median.

(10) <u>Hinges</u> are defined as the middle of each half of the batch. The depth of the hinge is $d(h) = ([d(m)] + 1)/2$, where the [] brackets are defined as the "integer part of". Again, a fraction of 1/2 means that the two data values surrounding the depth should be averaged. Approximately, one-fourth of the data in the batch lies below the lower hinge and one-fourth lies above the upper hinge. Note that the lower and upper hinges are similar to the 1st and 3rd quartiles, the difference being that hinges are computed from the depth of the median whereas quartiles are computed from the batch of data directly, without taking the integer part. This has the effect of placing the quartiles at most one rank further away from the median than the hinges. The hinge is preferred to the quartile in EDA as its meaning ties in with the methodology and concept of EDA summary statistics.

(11) The difference between the upper and lower hinges is commonly known as the hinge-spread, i.e., <u>H-spread</u>, which will be the choice for the measure of dispersion of a batch. (The hinges are one pair of many statistics termed <u>summary statistics</u>. Other summary statistics are eighths, sixteenths, etc., each possessing a corresponding spread. See Hoaglin and Velleman, reference 6.)

(12) Hoaglin, Mosteller, and Tukey, reference 7, and others refer to a <u>rule of thumb</u> in the EDA literature that states any value in a group of data values that lies below:

$$\text{lower hinge} - 1.5 * (\text{H-spread})$$

or above

upper hinge + 1.5 * (H-spread)

is termed an "outside observation" and requires an inquiry as to why it lies there. Although these observations may not, in fact, be "outliers" they do require close scrutiny. If the distribution of the data can be described as skewed or more long-tailed than short-tailed then the rule of thumb should identify as outside observations many points that are typical of the distribution and are not outliers. This rule of thumb technique will be used below in the initial calibration of the biweight measure for a new batch of data.

(13) The <u>biweight square</u> or <u>biweight</u> measure for the "center" of a batch is a weighted mean, more resistent than the mean, more sensitive than the median. The literature, in particular Barnett and Lewis, reference 1, refer to the biweight as a member of the class of <u>M-estimators</u> which is so designated because it is derived using maximum likelihood statistical techniques. The procedure used to compute it is much more involved than the computation of either the median or the classical mean. It is an iterative process that alternatively computes the respective weights for each of the data points in the batch and the value of the estimator x', which minimizes the function:

$$\sum_{i=1}^{n} w_i * (x_i - x')^2, \text{ where } x_i \text{ are the data values}$$

and w_i are the respective weights, also to be determined in the process. The solution is the biweight estimator x'.

$$x' = \sum_{i=1}^{n} (w_i * x_i) \,/\, \sum_{i=1}^{n} w_i$$

The weights are calculated:

$$w_i = (1 - ((x_i - x')/cS)^2)^2, \text{ when } ((x_i - x')/cS)^2) < 1$$

$$0, \qquad \text{otherwise}$$

The computational procedure starts with initial values for x' and the scale measure S and iterates until the value for x' converges (see the computations in Appendix D).

(14) Unlike S, which is recomputed at each iteration, c is a constant; cS is the cut-off point. The literature generally recommends for c a value of 4 for short-tailed distributions and a value of 9 for long-tailed distributions. McNeil, reference 11, suggests that the smaller the value of c the more protection the estimator has against the influence of outliers. If there are no outliers he recommends using a moderate to large value of c in the range of 6 to 10. (He states that a value of 10 roughly corresponds to using the mean as a measure of location.) The absence of "outside observations" according to the rule of thumb could be taken to suggest the absences of outliers, allowing the rule a role in the selection of a value for c.

(15) The scale measure, S, can be any one of several common measures. The standard deviation, the H-spread, the mean deviation, or the median absolute deviation (MAD) are just a few of these. The literature, by-in-large, uses the MAD as the scale measurement of choice in calculating the biweight measure. This paper will do likewise. As used here, the MAD is the median of the absolute deviations from the current value of x' :

$$MAD = median_i \{ |x_i - x'| \}$$

It is recomputed at each iteration.

(16) The iterative procedure is followed until the x' measure values converge. The final value is a weighted average of the data where the points close to the center have large weights and those further from the center have reduced weight. Those values at distance cS or greater would have zero weight. Hoaglin, Mosteller, and Tukey, reference 7, state that for a Gaussian type distribution the expected value of MAD is approximately $2/3 * \sigma$. Therefore a c-value of 6 tends to give zero weight to those observations that are 4=2/3 * 6, standard deviations away from the median.

(17) Mosteller and Tukey, reference 12, provide a comparison of the resistance and the efficiency of the median and biweight measures relative to the mean, using two types of distributions, namely Gaussian (short-tailed) and stretched-tailed (long-tailed). Roughly speaking, when two estimators of the same quantity have unequal variances the estimator with the larger variance is said to be less efficient, and the ratio of the smaller variance to the larger variance is termed <u>relative efficiency</u>. Efficiencies around 90% are deemed

"very good," with differences in efficiencies of a single percentage point "never detectable in practice".

(18) In addition to resistance, robustness is another desirable property of estimators. Robustness in this context is the characteristic of having a high degree of efficiency in a wide range of possible situations. Table 1 has been extracted from reference 12. As can be seen from this table the biweight measure is the most efficient except for quite small samples. For the smallest samples of size three, four, and five, the median may be the better choice.

(19) Thus far, two resistant measures of location have been presented and discussed. Their variance measures and the construction of screening comparison intervals using them are presented in the following sections.

b. Screening interval: a test for specious data.

(1) For the purpose of determining the "acceptability" of a new MOE data value, the test for specious data will involve the use of an MOE prediction interval constructed from a batch of previous similar runs and centered at the biweight location measure x' . The previous sections show that x' has the desired properties of resistance and rebustness when applied to random sample data. However, as discussed earlier, the batch of previously generated MOE data is not truly a random sample; consequently, precise probability statements cannot be attached to the intervals to be presented. Therefore, the term "screening" interval is used to describe the proposed interval to purposely differentiate it from the well-known term "confidence" interval which is associated with precise probability statements. The width of the screening interval must take into consideration two sources of variability, that of x' , and that of the batch distribution.

(2) Estimating the variability of the batch is the crux of the matter. One can use past MOE model data or that which is on hand, i.e., the observed batch. The use of an historical data base might be preferred; however, TRAC-FLVN is only now attempting to document and accumulate VIC study MOE data. Therefore, this paper will only use the observed data. This initial estimate should be subsequently updated as the batch of MOE becomes available.

Table 1. Comparison of Estimator Resistance and Robustness

	Sample size	Resistant?	Gaussian efficiency	Stretched-tail efficiency	Robustness of efficiency
Arithmetic Mean	Small	No	100%	Poor	Poor
	Large	No	100%	Very poor	Very Poor
Median	Small	Yes	High	Higher	High
	Large	Yes	62%	Higher	Moderate
Biweight	Small	Reasonably	Highish	Higher	Higher
(c = 6 or 9)	Large	Yes	>90%	>90%	High

Reference: Mosteller, Fredrick and Tukey, John W., Data Analysis and Regression, Reading, Massachusetts, Addison-Wesley, 1977

The cited reference does not assign numerical values to the ordinal table entries; however, the rankings of highish, high, and higher probably fall within the 70 to 90% range. The rankings of very poor, poor, and moderate probably fall within the 10 - 50% range.

(3) Hoaglin and Velleman, reference 6, approach the estimate of variability by imagining a batch of data distributed according to the standard Gaussian distribution, e.g., mean 0 and variance 1. That is, they compute summary batch statistics for astandard Gaussian distribution and calculate the corresponding spreads. Because the hinges are the most resistant of the summary statistics, reference 6 uses them to estimate batch variability. For a standard Gaussian distribution, the hinges are calculated to be + and - 0.6745 with a corresponding hinge spread of 1.349. Thus, the general value of a Gaussian H-spread is $1.349 * \sigma$. For a batch of data that are exactly Gaussian but not standard, the ratio H-spread/1.349 yields the standard deviation. The value H-spread/1.349 is not overly-sensitive to the actual shape of the batch and provides a robust measure of variability for a batch based on a symmetric distribution. Should the distribution deviate from symmetry the measure still provides a meaningful approximation. Consequently, a plausible measure of variability of the batch is obtained by the following expression:

standard deviation(batch) = (H-spread)/1.349

This measure of batch variability is but one component of the overall variability of the screening interval to be proposed. The other component is the variance of the biweight location measure x'. The literature, in particular Hoaglin, Mosteller, and Tukey, reference 7, proposes the following formulation for this variance estimate. For a given value of x, as calculated from c, S, and weights

$$w_i = \begin{cases} ((1 - (r_i/cS)^2)^2, & if\ (r_i/cS)^2 < 1 \\ 0 & , \text{otherwise} \end{cases}$$

where $r_i = x_i - x'$, they obtain

$$var(x') = n * \frac{\sum_{i=1}^{n} w_i^2 * r_i^2}{(\sum_{i=1}^{n} (w_i) * (1-5*r_i/cS)^2))^2}$$

we will take the standard deviation (x') to be its square root.

(4) Consequently, a measure of total variability for screening, i.e., prediction, purposes can be estimated by summing

the two components of variability. That is the screening interval proposed for implementation is:

$$x' + or - [standard\ deviation(x') + standard\ deviation(batch)]$$

The proposed screening tool provides the analytical underpinnings to support the implementation of a prediction interval in the expert system under development. It lacks precise probabilistic interpretation, but it rests on a resistant and robust estimator and is sensitive to variability in both the estimator and the batch itself. By drawing on historical data or multiplying the term in square brackets by a constant, the user can adjust the length to fit existing circumstances. More to the point, this technique achieves the desired goal of a quality control tool. That is, it alerts the analyst that something is "out-of-line" early on in the analysis rather than later-on when the cost to correct is exceedingly much more expensive.

c. Screening interval: A test for comparing batches.

(1) The following derivation for batch comparison is based on Velleman and Hoaglin's, reference 6, and McGill, Tukey, and Larsen's, reference 10. It will address the situation when two or more batches of data are to be compared using intervals about the median. As a framework, let such an interval about the median be expressed as follows:

$$median + or - Z_{\alpha/2} * standard\ deviation(median)$$

The derivation and computation of standard deviation(median) and $Z_{\alpha/2}$ are the thrust of the remainder of this section.

(2) Reference 6 states for a Gaussian distribution the variance of the median is approximately $\pi/2$ (or 1.571) times the variance of the mean. Thus, the constant 1.253 times the standard deviation of the mean approximates the standard deviation of the median. This is true for large samples, say 20 or more, and provides a good estimate for a wide variety of distributions. For smaller even-numbered batch sizes, Kendall and Stuart, reference 8, provide appropriate constants that the standard deviation of the mean should be multiplied by to arrive at a standard deviation of the median. Interpolation of these constants is recommended for odd-numbered batch sizes.

(3) The need now exists to estimate batch variability about the mean. As stated in the last section, Hoaglin and Velleman, reference 6, accomplished this using the H-spread of a standard Gaussian distribution to arrive at an estimate of variability, i.e., H-spread/1.349. This value is not

overly-sensitive to the actual shape of the batch and provides a robust estimator of variability for the batch based on a symmetric distribution. Consequently, the standard deviation of the median, for large n, can be computed as:

$$1.253 * (H\text{-}spread) / (1.349 * \sqrt{n})$$

McGill, Tukey, and Larsen, reference 10, state that this standard error estimate of the median applies for any distribution that is approximately Gaussian in the middle, which is a common situation (Winsor's principle). Consequently, the proposed screening interval about the median for batch comparison is computed as:

$$median + or - Z_{\alpha/2} * (1.253 * (H\text{-}spread) / (1.349 * \sqrt{n}))$$

where the value of $Z_{\alpha/2}$ is based on the following derivation by Hoaglin and Velleman, reference 6.

(4) They use the mean of the Gaussian distribution as a model for their discussions of the median of a batch. As a reminder, recall that the usual 95% confidence interval for the mean of a Gaussian distribution with known variance is

$\bar{x} + or - 1.96 * \sigma_{\bar{x}}, with\ \sigma_{\bar{x}} = \sigma_x/\sqrt{n}$, where n is the size of the sample.

(5) For the median, the analogous interval based on a Gaussian distribution uses $Z_{\alpha/\alpha}$ =1.96, and is centered at the median of the batch. That gives the interval

$$median + or - 1.96 * 1.253 * (H\text{-}spread) / (1.349 * \sqrt{n}))$$

or

$$median + or - 1.821 * ((H\text{-}spread) / \sqrt{n})$$

The medians of the distributions from which two batches, of size n1 and n2, respectively, may be termed significantly different at a 0.05 level if the intervals for the two batches do not overlap, or equivalently if

$$|median2 - median1| > 1.821 * (H\text{-}spread1/ \sqrt{n1} + H\text{-}spread2/ \sqrt{n2})$$

(6) However, as reference 6 points out, the resulting interval may be far too large if the two values analogous to $\sigma_{\bar{x}}$ are not far apart. Consider first the comparison of means from a Gaussian distribution. To compare two batches with equal values

for $\sigma_{\bar{x}}$, the statistic is $|\bar{x}_2 - \bar{x}_1|/\sqrt{2} * \sigma_{\bar{x}}$ which is a Z-value and should be compared to + or - 1.96. Equivalently, if

$|\bar{x}_2 - \bar{x}_1| - 1.96 * \sqrt{2} * \sigma_{\bar{x}} > 0$ the means are significantly

different at an α- level of 0.05. The corresponding inequality for medians can be written as

$$|median2 - median1| > 1.821 * \sqrt{2} * (1/2) * [H\text{-}spread1/\sqrt{n1} + H\text{-}spread2/\sqrt{n2}]$$

because the two terms in the square brackets are equal. Equivalently,

$$|median2 - median1| > (1.821/\sqrt{2}) * [H\text{-}spread1/\sqrt{n1} + H\text{-}spread2/\sqrt{n2}]$$

Thus the candidate multiplier in this case is not 1.821 but $1.821/\sqrt{2}$. Following reference 6, we average these two candidates:

$$1/2 * (1.821 + 1.821/\sqrt{2}) = 1.554$$

and obtain

$$median + or - 1.554 * (H\text{-}spread)/\sqrt{n}$$

as the 95% screening comparison interval. These screening intervals can be used for pairwise batch comparison. Consequently, if the intervals for two batches do not overlap, we have 95% confidence that the two Gaussian-type batches have different medians.

(7) The choice of $z_{\alpha/2} = 1.96$ for constructing a 95% screening interval can be modified later according to the experience of the user. In the general case for a (1- α) percent screening interval we obtain

$$median + or - z_{\alpha/2} * (0.793 * (H\text{-}spread)/\sqrt{n})$$

where the corresponding z_α value is obtained from the Gaussin distribution, and 0.793 = 1.544/1.96. It is left to the user to determine whether the preference is for an interval that is "long" or "short".

(8) Although these intervals lack the probabilistic meaning of the traditional confidence interval, they have both an historical and analytical basis. From an historical basis these intervals will continue to be updated with observed data as the batch of MOE accumulates. Analytically, the interval is derived based on accepted robust and resistant techniques of EDA.

7. Batch Symmetry Considerations.

a. A basic assumption of this methodology is the requirement of the batch to be characterized by a shape that is roughly symmetric. Senior analysts with much more experience have stated this appears to be a reasonable characterization of MOE distributions. Further, not only are these measures supported as robust and resistant but the literature, references 2 and 7, have either referenced or involked "Winsor's principle" which states, in essence, that most distributions tend to be Gaussian in the middle. Threfore, this basic assumption although critical appears to be a plausible one for study MOE.

b. In the event the batch does not conform to the "Winsor principle" assumption, reexpression of the data is an alternative technique that can applied to achieve the desired effect. A measure of skewness that McNeil, reference 11, describes is S = (median - lower hinge)/(upper hinge - lower hinge). It is applied to the present data in paragraph 8b(6), below. A value close to 1/2 indicates symmetry. Reference 7 provides an extensive discussion on a family of transformations that can be applied to a set of data to achieve symmetry, but these will not be considered here.

8. Implementation.

a. Scenario situation.

(1) The scenario criteria that define a distribution of ratio type MOE data are: 1) the theater of conflict, 2) the respective missions of the opposing forces, 3) and starting force ratios. The MOE data represent a European theater conflict where the Blue corps-level forces conduct a defense with the option for a counterattack. Red conducts a main attack with a force that presents approximately a 5:1 Red to Blue beginning force ratio. MOE that are in form of percentages are extremely sensitive to the days of conflict. Consequently, for these MOE another criterion would be the number of days of conflict.

(2) It is the long term objective of the Programming and Quality Assurance Division to collect the MOE data outlined in appendix A for other VIC-generated scenarios and studies. Other scenario theaters of conflict are to include southeast Asia, southwest Asia, and Latin American. Starting force ratios in these areas can range from 0.5:1 to 2:1 Red to Blue forces. The approximate numbers of weapon system types for the European scenario of interest are provided in table 2. The LER and SER MOE data are contained in table 3.

b. Screening interval Computation.

(1) Figures 1 through 3 provide a plot of the helicopter, tank SER and force LER MOE data, respectively. These data points represent VIC study computer run results from the "same" scenario but from several different studies. The computations to compute both the prediction interval based on the biweight estimator and the screening interval for batch comparison are provided in appendices D and E, respectively. The batch comparison screening intervals were constructed at a 95% level of significance for a Gaussian distribution.

(2) When applied to the helicopter data the rule of thumb computation for outside observations, presented on page 15, provided a lower value of 4.93 and an upper value of 14.0. Because the SER data did not appear to contain any outside observations, a moderate c-value of 6 was used. The biweight measure for the helicopter SER data was computed to be 8.97. The standard deviation of the biweight measure was calculated to be 1.13 with a batch standard deviation of 1.68; therefore, the proposed prediction screening interval for new helicopter SER data values is 6.16 - 11.78. Applying this interval to the initial batch of data reveals that three data values are outside this interval; consequently, the conditions that generated these data values should be investigated further.

(3) The "outside observation" lower and upper values for the tank SER data were computed to be -0.67 and 8.58, respectively. In the data arena of study MOE a negative value is not possible; therefore, as a practical matter zero is the lower value. The tank SER data did not appear to contain any suspicious observations and again a moderate c-value of 6 was used. The tank data biweight measure was computed to be 3.73. The biweight standard deviation was computed to be 1.69 with a batch standard deviation of 1.71; consequently, the tank SER prediction screening interval is 0.33 - 7.13. The entire batch of tank SER MOE data values lie within this interval.

(4) Finally, the "outside observation" computation for the force LER data provided a lower value of 0.29 and an upper value of 8.30. Although the LER value of 8.8 appeared to be

Table 2. Red and Blue forces

	Red	Blue
Tanks	5000	1000
AFVs[1]	8000	1500
Artillery	4000	800
Helicopters	500	600
Total	17500	3900

[1]AFV: armor fighting vehicles

Table 3. LER and SER MOE values

Helo SER	Tank SER	LER
5.51	1.20	2.35
7.80	1.40	3.23
8.15	2.70	3.35
8.51	2.80	3.60
8.72	2.80	3.94
8.74	2.91	4.18
9.27	3.30	5.27
9.27	3.47	5.30
9.60	3.70	8.80
9.76	4.27	
11.44	4.95	
13.19	5.11	
13.57	5.50	
	5.92	
	6.00	

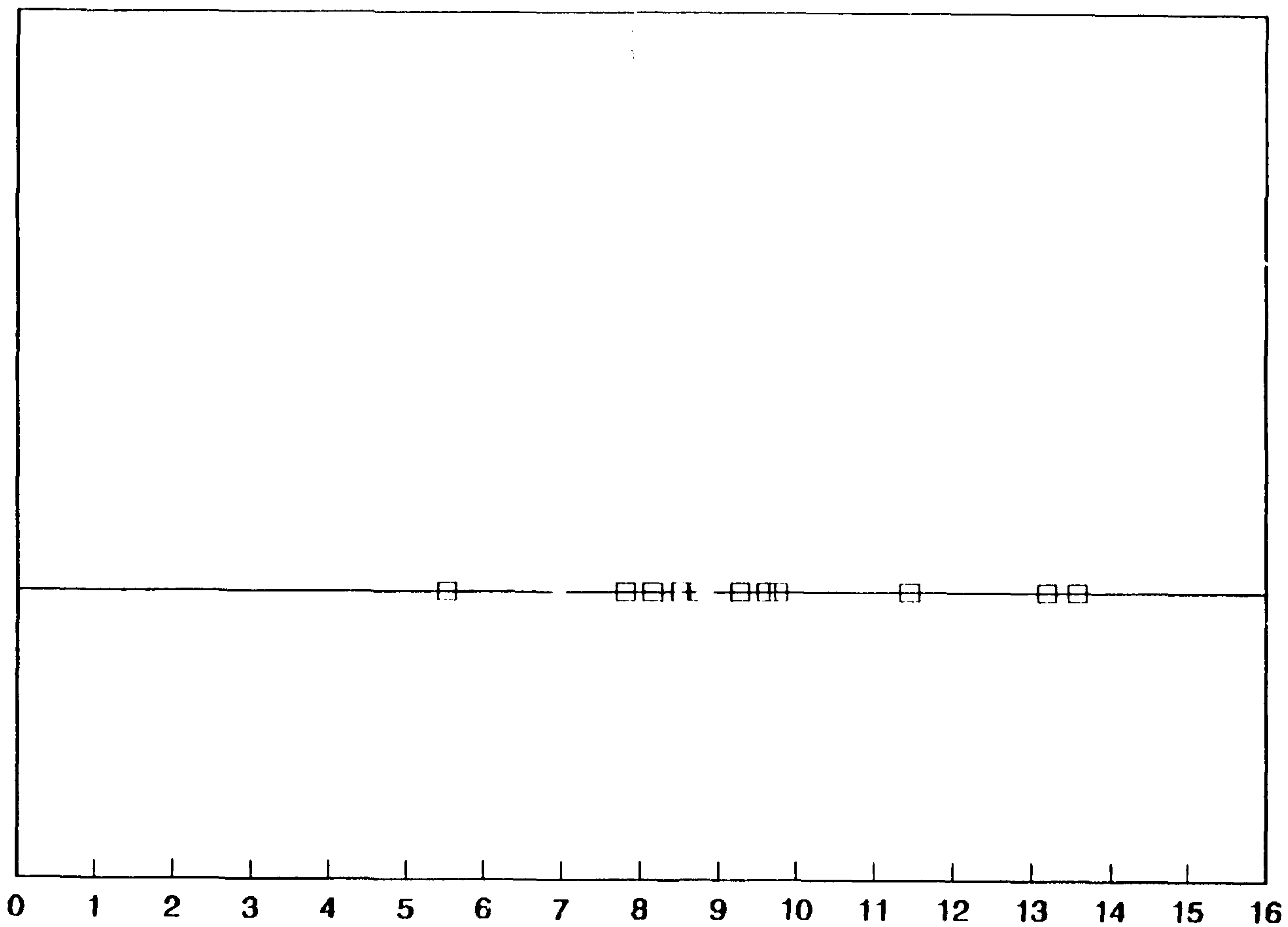

Figure 1. Helicopter SER Data

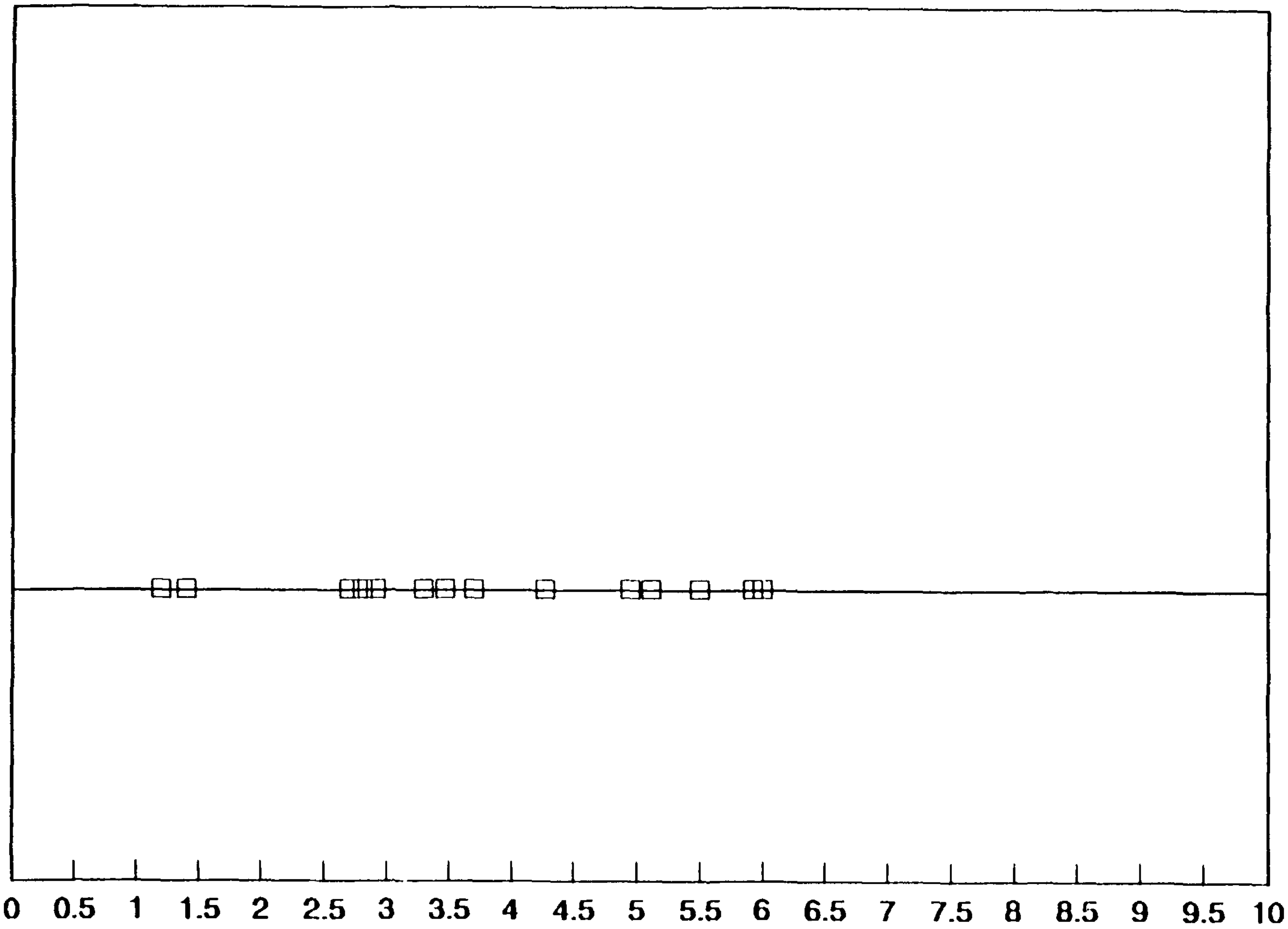

Figure 2. Tank SER Data

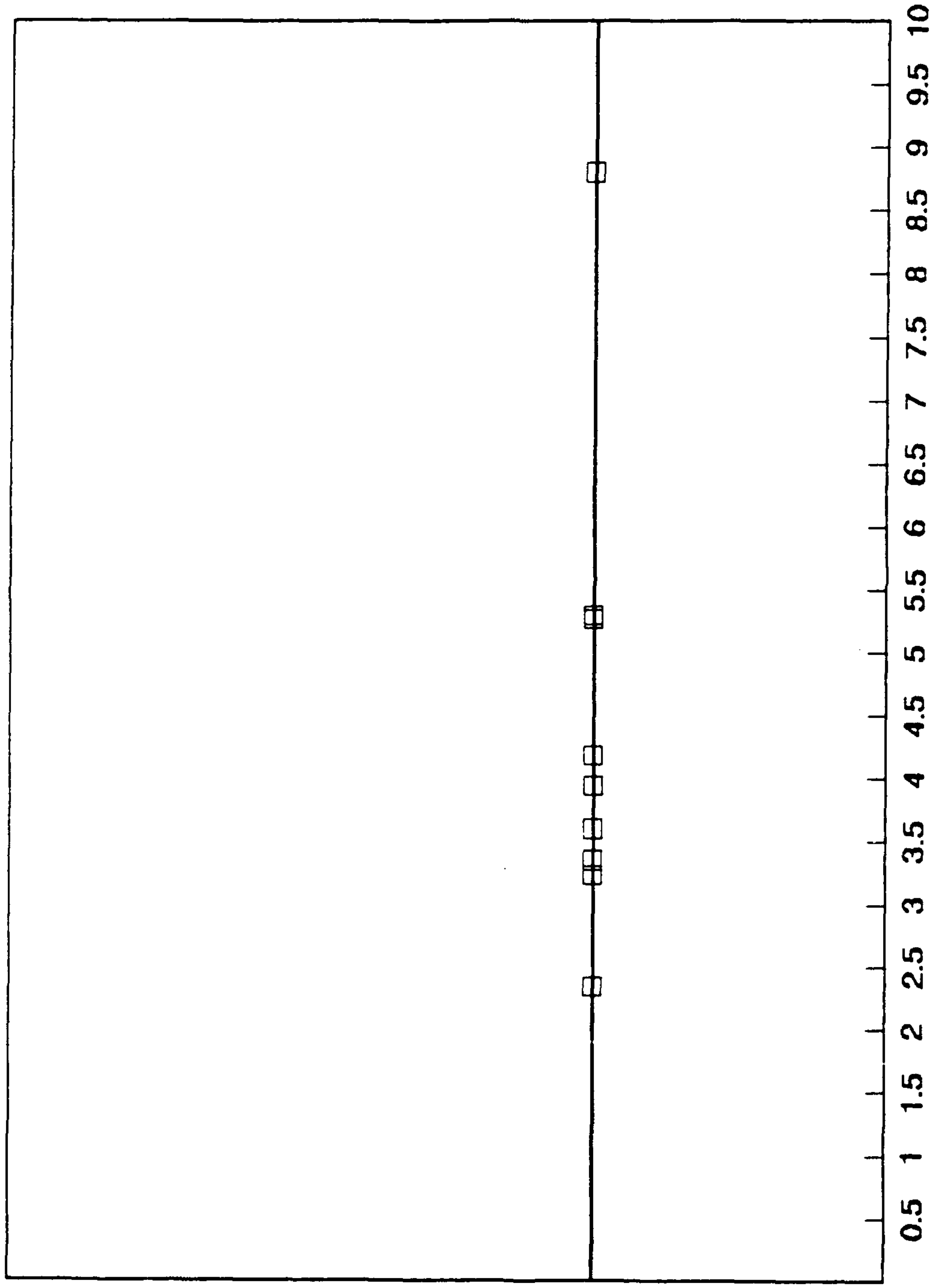

Figure 3. Force LER Data

slightly on the large side, it was not considered necessary to change the c-value of 6. The biweight measure was computed to be 3.88 with a standard deviation of 1.14. The batch standard deviation was calculated to be 1.48; therefore, the prediction screening interval is 1.26 - 6.50. Only one data point falls outside this interval and it requires investigation.

(5) Screening interval computations based on the biweight measure are appropriate when the analyst is interested in comparing new MOE data values to an existing batch of accumulated data base of study MOE data. However, when the analyst desires to compare two separate batches of MOE data, i.e., test the hypothesis that the two batch medians are equal, he should implement the batch comparison methodology. The 95% screening interval for batch comparisons using the helicopter data was computed to be 8.29 - 10.25 with a median value of 9.27. The tank SER data 95% screening interval is 2.54 - 4.40 with a median value of 3.47. The force LER data 95% screening interval is 2.90 - 4.98 with the computed median of 3.94.

(6) Applying McNeil's test for symmetry to the helicopter batch results in a value of 0.42; this value appears close to the desired 0.50 value. The test for symmetry on the tank SER and LER data results in values of 0.29 and 0.33, respectively. These two data batches are candidates for reexpression. An expression that results in a favorable value for the McNeil test is the inverse transformation of the data values. Such a transformation results in values of 0.57 for the two batches. Consequently, for the batch comparison test one might consider the inverse transformation of the data prior to calculating comparison intervals.

9. Conclusions.

a. Using past study and scenario MOE data, the methodologies developed in this paper provide an analytical procedure to implement quality control and quality assurance. These methodologies make no assumptions of the underlying distribution of the data.

b. These intervals are, in fact, screening tools to be used to identify inconsistent results based on past study expectations. Further, the expert system will guide the study analyst by focusing his search for possible errors in the data base and the model. If a thorough search of these areas does not reveal any miscues or problems, then chances are that the MOE is a valid occurrence. This result would then provide additional supporting analysis that the candidate force "improvement" is significant from an effectiveness standpoint.

c. This methodology used in conjunction with an expert system will provide a responsive and interactive tool to guide the analyst during the conduct of a study by providing a frame of reference for MOE results. The MOE data used to compute the prediction intervals need to be updated with future study data to increase the size of the historical data bases. As this MOE data bases accumulate, it is necessary to recompute the MOE prediction intervals.

10. Recommendations. This methodology can and should be expanded to address other combat development models in the community. Consideration should also be given to collecting MOE data on an echelon basis for each model. Ultimately, screening intervals for selected MOE by echelon for each combat model will provide a means, for similar scenarios, to compare model consistency. Also, this methodology could be extended to the development of a table look-up model for gross estimates and quick reaction to questions of force and/or system effectiveness for a particular scenario.

APPENDIX A

STUDY MOE REQUIREMENTS

STUDY MOE REQUIREMENTS

The following MOE data requirements have been requested of study team personnel for the purpose of establishing an initial data base of selected MOE.

Study Title:
Scenario (hours of combat):
Mission:
Force Years:
Weapon system list/quantity:
Run Description:

1. The macro-level MOE data required are the total force LER. The following MOE percent statistics are also required:

	Percent ACVs Killed	
	Red	Blue
Direct Fire		
Helicopters		
Artillery		
Fixed-Wing		
Other		

	Percent artillery killed	
	Red	Blue
Artillery		

	Percent Helo Killed	
	Red	Blue
ADA		

	Percent Fixed-Wing Killed	
	Red	Blue
ADA		

2. The following weapon system specific MOE data are also required for each Blue weapon/Red target category (see appendix B for a definition of Red target category). *Blue weapon/Red target:

	Number in Force	Percent Red Target Category Killed	SER
a. Direct Fire			
(1) Tanks			
(2) LOS Missiles			
(a) TOW			
(b) AMS-H			
(c) AAWS-M			
(d) KEM			
(e) OTHER(specify)			
(3) NLOS Missiles			
(a) FOG-M			
b. Helicopters			
(1) AH-1S			
(2) AH-64			
(3) OTHER (specify)			
c. Artillery			
(1) Conventional			
(a) 203mm			
(b) 105mm			
(c) 155mm			
(d) MLRS			
(e) OTHER (specify)			
(2) Smart			
(a) 155/Copperhead			
(b) 155/SADARM			
(c) MLRS/TGW			
(d) MLRS/SADARM			
(e) MLRS/ATACMS			
(f) OTHER (specify)			
d. Mines			
e. ADA			
(1) ADATS			
(2) NLOS			
(3) PMS			
(4) Stinger (Manpack)			
(5) OTHER (specify)			
f. TACAIR (All)			

* Percent Red target killed is defined to mean of the total targets killed, what percent was killed by the Blue weapon system.

3. The following weapon system category MOE data are also required for each Red weapon/ Blue target category (see appendix B for the definition of Blue target category). *Red weapon/Blue target:

	Number in Force	Percent Blue target Category Killed	SER
a. Direct Fire			
(1) Tanks			
(2) LOS Missiles			
(3) NLOS Missiles			
b. Helicopters			
c. Artillery			
(1) Conventional			
(2) Smart			
d. Mines			
e. ADA			
f. TACAIR (All)			

* Percent Blue target killed is defined to mean of the total targets killed, what percent was killed by the Red weapon system.

APPENDIX B

MOE DEFINITIONS

MOE DEFINITIONS

This appendix provides the definition of MOE and target categories.

Loss Exchange Ratio (LER):

$$\frac{\text{Total Red weapon system losses}}{\text{Total Blue weapon system losses}}$$

Red target categories:

- ACV (includes tank, BRDM, BMP, and BTR systems)
- Helos
- Air
- ADA
- Artillery
- Other (CSS,C&C sites, etc.)

Blue target categories:

- AFV (includes tank, ITV, CFV, and IFV systems)
- Helos
- Air
- ADA
- Artillery
- (Other CSS, C&C sites, etc.)

Number in force:

Can be weapons count or number of rounds fired (artillery and mines) or number of sorties (TACAIR) flown

SER (system exchange ratio):

$$\frac{\text{Total Red (Blue) system losses due to Blue(Red) system i}}{\text{Total Blue (Red) system losses of system i}}$$

APPENDIX C

AI DECISION RULES

BLUE HELICOPTER

```
                                            (start)
        Is the SER for Blue                Is the percent kill of Red
yes . . helicopters within. . . . yes . . . . ACVs by Blue helicopters
.       the interval                       within the range
.       xx.xx  -  yy.yy?                        x - y%
.              .                                 .
.              .                                 .
.              .                                 .
.             no                          . . . . . . no          Is the number of Red systems          The utilization of Blue
.              .                          .              no . . . killed per sortie by Blue . . . yes . . . helicopters may be in
.              .                          .              .        helicopters within the range          error.  Also, check
.              .           Are Blue helicopter losses    .           x.xx  -  y.yy?                    tactics for validity.
.              . . . . . . per sortie by Blue helicopters. . . .        .                                    .
.                          either above or below, or within              .                                    .
.                          the range x.xx  -  y.yy?                     no                                   .
.                                            .                          .                                    .
.                                            .                          .                                    .
.                                          above                  The lethality data of Blue                 .
.                                            or                   helicopters may be in error.               .
.                                          below                  Also, check the Blue helicopter. . . . . . . . GO TO A
.                                            .                    target acquisition process and
.                                            .                    the magnitude of target density
.                                            .                    as possible contributors to this
.                              Red air defense lethality and/or   erroneous behavior.
. .NEXT SYSTEM CATEGORY. . A . . . range data vs Blue helicopters
                                   may be in error.  Also, check Red
                                   air defense employment parameters.
```

MANEUVER FORCES

```
                                                (start)
          Is the force LER                Is the percent kill of Red
no. . . within the interval . . . yes . . . ACVs within the range of
 .          x.xx  -  y.yy?                      x - y% ?
 .               .                                .
 .               .                                .
 .              yes                               .
 .               .                               no
 .               .                                .
 .   NEXT SYSTEM CATEGORY. . .  A                 .
 .                                                .                          Check the Red ACV and Blue tank
 . . . . . . . . . . . . . . . . . . . . . . . Is the SER for Blue tanks . . . NO . . . lethality data.  Also, check the
                                              within the interval                     tactics employed by Blue tanks.
                                                x.xx  -  y.yy?
                                                  .
                                                  .
                                                 yes
                                                  .
                                                  .                          Check the Red ACV and Blue AFV
                                              Is the SER for Blue AFVs . . . No . . . lethality data.  Also, check the
                                              within the interval                     tactics employed by Blue AFVs.
                                                x.xx  -  y.yy?                                 .
                                                  .                                            .
                                                  .                                            .
                                                 yes                                           .
                                                  .                                            .
                                                  .
                                              Check the tactics employed . . . . . . . . . . . GO TO A
                                              by both Red and Blue ACVs.
```

BLUE ARTILLERY

```
        (start)                                        A problem with Blue conventional
     Is the percent kill of Red                        artillery may exist.  In particular, . . . . . GO TO A
 . . .ACVs by conventional artillery                   check basic loads, resupply rates,
 .        within the range                             and firing rates that may be too high.
 .            x - y% ?                                              .
yes            .                                                  above                 Since the number of rounds per
 .             no                                                   .                   tube per day by Blue conventional
 .             .                                                                        artillery and the number of Red
 .   Is the number of Red ACV's killed                 Is the number of rounds of Blue  ACVs killed per 1000 rounds are
 .   per 1000 thousand rounds of Blue  . . . yes . . .conventional artillery fired per . . within . . within expected ranges, it is
 .   conventional artillery within the                 tube per day below, within, or   unlikely that artillery is
 .        range  x - y% ?                              above the range of xx  -  yy ?   directly responsible for its own
 .             .                                                    .                   unexpected percent kill of Red
 A             no                                                   .                   forces.    .
 .             .                                                  below                            .
 .   A problem may exist with the Blue                              .                              .
 .   conventional artillery lethality                               .                              .
 .   data.  Also, any recent model                     The following factors may be                .
 .   modifications may have adversely                  contributing to the anomalous . . . . . . . . . . . . . GO TO A
 .   affected the play of Blue conventional            results:
 .   artillery.  Check if smart munitions                     - Resupply rates may be too low
 .   are being used at an unexpected rate.                    - Artillery may be moving too much
 .                                                            - Target acquisition problems may exist
Is the percent kill of
Red ACVs by Blue smart . . . . . . .  NEXT SYSTEM CATEGORY
within the range  x - y% ?
 .
 no                                                                  An error may exist with the munition
 .                                 Is the number of kills            type lethality data.  Also, check the
Was munition type used? . .  yes  . . per round within the  . . . no . . . false targeting process as well as the
 .                                 range of x.xx  -  y.yy?           intelligence/acquisition process.
 no                                    .
 . . NEXT MUNITION TYPE                  yes . . . .An error may exist with the munition type resupply rate.
```

AIR DEFENSE

```
                                                      (start)
                                   below . . . Are aircraft losses per sortie . . above
                                       .       below, within, or above the          .
                                       .       range of x.xx - y.yy?                .
Check AD employment. . . no . . Are HIMAD responsible        .        Are HIMAD responsible for. . no. . Check AD employment
paramters; also,                for at least y %           within     at least y % of the losses?      parameters; also,
SHORAD/HIMAD may not            of the losses?               .                   .                    SHORAD maybe to lethal
be lethal enough.                      .                     .                  yes                             .
.                                     yes                    .                   .                              .
.                                      .                     A          Check AD employment parameters;          .
GO TO A . . Check AD employment parameters;                  .          also, HIMAD may be to lethal . . . . . GO TO A
            also, HIMAD may not be lethal                    .
            enough.                   Are helicopter losses per sortie below
                              below . . . within, or above x.xx - y.yy? . . . . . above
Check AD employment                    .                     .                  .                     Check AD employment
parmeters; also,   . . . no . .Are SHORAD responsible for  within     Are SHORAD responsible for . no . parameters; also, HIMAD
SHORAD/HIMAD may not            at least x % of losses?      .        at least x % of losses?          may be too lethal.
be lethal enough                       .          Is the helicopter SER            .                            .
.                                     yes     yes. . . . within the interval        .                            .
.                                      .       .            xx.xx  -  yy.yy?       yes                           .
.                                      .       .                     .              .                            .
.            Check AD employment paramters; .                        no             .                            .
GO TO B . . .also, SHORAD may not be lethal .                        .      Check AD employment parameters;      .
             enough.                        .                        .      also, SHORAD may be too lethal . . .  GO TO B
                                            .                        B
                                            .                        .
                                            .  Are AD loss rate commensurate. . no. .Check lethality data vs AD system; also,
                                            .  with the ACV loss rate?              priority of AD systems as targets.
                                            .                        .                          .
                                            .                       yes                         .
                                            .                        .                          .
                                            . . . . .....NEXT SYSTEM CATEGORY . . . . . . . . . . .
```

APPENDIX D

A TEST FOR SPECIOUS DATA: BATCH BIWEIGHT SCREENING INTERVAL COMPUTATION

A TEST FOR SPECIOUS DATA: BATCH BIWEIGHT SCREENING INTERVAL COMPUTATION

The computation of the helicopter SER data biweight measure x' and the respective weights involve an iterative process. This process continues until the solution x' converges. Recall the expression for X' is:

$$x' = \sum_{i=1}^{n} (w_i * x_i) / \sum_{i=1}^{n} w_i$$

and the weights are computed as follows:

First iteration:

$m_1 = 9.27$, the batch median and
$s_1 = \text{median } |r_i| = 0.76$

x_i	$r_i=x_i-m_1$	$r_i/6s_1$	$(r_i/6s_1)^2$	$w_i=((1-(r_i/6s_1)^2)^2$
5.51	-3.76	-0.825	0.681	0.102
7.80	-1.47	-0.322	0.104	0.011
8.15	-1.12	-0.246	0.061	0.882
8.51	-0.76	-0.167	0.028	0.945
8.72	-0.55	-0.121	0.015	0.970
8.74	-0.53	-0.116	0.014	0.972
9.27	0.00	0.000	0.000	1.000
9.27	0.00	0.000	0.000	1.000
9.60	0.33	0.072	0.005	0.990
9.76	0.49	0.108	0.012	0.976
11.44	2.17	0.476	0.227	0.598
13.19	3.92	0.860	0.740	0.068
13.57	4.30	0.943	0.889	0.012

$$m_2 = \sum_{i=1}^{n} (w_i * x_i) / \sum_{i=1}^{n} w_i = 9.18$$

Second iteration:

$m_2 = 9.18$, where
$s_2 = \text{median } |r_i| = 0.67$

x_i	$r_i=x_i-m_2$	$r_i/6s_2$	$(r_i/6s_2)^2$	$w_i=((1-(r_i/6s_2)^2)^2$
5.51	-3.67	-0.913	0.843	0.028
7.80	-1.38	-0.343	0.118	0.778
8.15	-1.03	-0.256	0.066	0.872
8.51	-0.67	-0.167	0.028	0.945
8.72	-0.46	-0.114	0.013	0.974
8.74	-0.44	-0.110	0.012	0.976
9.27	0.09	0.022	0.001	0.998
9.27	0.09	0.022	0.001	0.998
9.60	0.42	0.105	0.011	0.978
9.76	0.58	0.144	0.021	0.958
11.44	2.26	0.562	0.316	0.468
13.19	4.01	0.998	0.996	0.000
13.57	4.30	0.000	0.000	0.000

$$m_3 = \sum_{i=1}^{n} (x_i * w_i) / \sum_{i=1}^{n} w_i = 9.03$$

Third iteration:

$$m_3 = 9.03$$
$$s_3 = \text{median } |r_i| = 0.73$$

x_i	$r_i=x_i-m_3$	$r_i/6s_3$	$(r_i/6s_3)^2$	$w_i=((1-(r_i/6s_3)^2)^2$
5.51	-3.52	-0.804	0.646	0.125
7.80	-1.23	-0.281	0.079	0.848
8.15	-0.88	-0.201	0.040	0.992
8.51	-0.52	-0.119	0.014	0.972
8.72	-0.31	-0.071	0.005	0.990
8.74	-0.29	-0.066	0.004	0.992
9.27	0.24	0.055	0.003	0.994
9.27	0.24	0.055	0.003	0.994
9.60	0.57	0.130	0.017	0.966
9.76	0.73	0.167	0.028	0.945
11.44	2.41	0.550	0.303	0.486
13.19	4.16	0.950	0.903	0.486
13.57	4.54	0.000	0.000	0.000

$$m_4 = \sum_{i=1}^{n} (x_i * w_i) / \sum_{i=1}^{n} w_i = 8.98$$

Fourth iteration:

$m_4 = 8.98$

$s_4 = \text{median } |r_i| = 0.78$

x_i	$r_i = x_i - m_4$	$r_i/6s_4$	$(r_i/6s_4)^2$	$w_i = ((1 - (r_i/6s_4)^2)^2$
5.51	-3.47	-0.742	0.551	0.202
7.80	-1.18	-0.252	0.064	0.876
8.15	-0.83	-0.177	0.031	0.939
8.51	-0.47	-0.100	0.010	0.980
8.72	-0.26	-0.056	0.003	0.994
8.74	-0.24	-0.051	0.003	0.994
9.27	0.29	0.062	0.004	0.992
9.27	0.29	0.062	0.004	0.992
9.60	0.62	0.133	0.018	0.964
9.76	0.78	0.167	0.028	0.945
11.44	2.46	0.526	0.277	0.523
13.19	4.21	0.900	0.810	0.036
13.57	4.59	0.981	0.962	0.001

$$m_5 = \sum_{i=1}^{n} (x_i * w_i) / \sum_{i=1}^{n} w_i = 8.97$$

Fifth iteration:

$m_5 = 8.97$

$s_5 = \text{median } |r_i| = 0.79$

x_i	$r_i = x_i - m_5$	$r_i/6s_5$	$(r_i/6s_5)^2$	$w_i = ((1 - (r_i/6s_5)^2)^2$
5.51	-3.46	-0.730	0.533	0.218
7.80	-1.17	-0.247	0.061	0.882
8.15	-0.82	-0.173	0.030	0.941
8.51	-0.46	-0.097	0.009	0.982
8.72	-0.25	-0.053	0.003	0.994
8.74	-0.23	-0.049	0.002	0.996
9.27	0.30	0.063	0.004	0.992
9.27	0.30	0.063	0.004	0.992
9.60	0.63	0.133	0.018	0.964
9.76	0.79	0.167	0.028	0.945
11.44	2.47	0.521	0.271	0.531
13.19	4.22	0.890	0.792	0.043
13.57	4.60	0.971	0.943	0.003

$$x' = \sum_{i=1}^{n} (x_i * w_i) / \sum_{i=1}^{n} w_i = 8.97$$

Since this value is the same as that obtained after the fourth iteration, convergence is obtained. To compute the variance of the biweight measure, use the last set of r_i and w_i data values and solve for the variance using the following formula:

$$var(x') = n * \frac{\sum_{i=1}^{n} w_i^2 * r_i^2}{(\sum_{i=1}^{n} (w_i) * (1-5 * (r_i/cS)^2))^2}$$

The variance(biweight) = 1.2666
standard deviation(biweight) = 1.1252

Recall the standard deviation(batch) = (H-spread)/1.349. The lower and upper hinges of the SER data are 8.33 and 10.60, respectively. Consequently,

standard deviation(batch) = 2.27/1.349 = 1.6827

The helicopter SER prediction interval is 8.97 + and - (1.13 + 1.68), i.e., 6.16 - 11.78.

Similarly, the computation of the tank SER data biweight measure x' and the respective weights involve an iterative process. The computation of the biweight measure for the tank SER data is as follows:

First iteration:

m_1 = 3.47, the batch median and

s_1 = median $|r_i|$ = 0.80

x_i	$r_i=x_i-m_1$	$r_i/6s_1$	$(r_i/6s_1)^2$	$w_i=((1-(r_i/6s_1)^2)^2$
1.20	-2.27	-0.423	0.179	0.674
1.40	-2.07	-0.431	0.186	0.663
2.70	-0.77	-0.160	0.027	0.947
2.80	-0.67	-0.140	0.020	0.960
2.80	-0.67	-0.140	0.020	0.960
2.91	-0.56	-0.117	0.014	0.972
3.30	-0.17	-0.035	0.001	0.998
3.47	0.00	0.000	0.000	1.000
3.70	0.23	0.048	0.002	0.996
4.27	0.80	0.167	0.028	0.945
4.95	1.48	0.308	0.095	0.819
5.11	1.64	0.342	0.117	0.780
5.50	2.03	0.423	0.179	0.674
5.92	2.45	0.510	0.260	0.548
6.00	2.53	0.527	0.278	0.521

$$m2 = \sum_{i=1}^{n} (x_i * w_i) / \sum_{i=1}^{n} w_i = 3.62$$

Second iteration:

$m_2 = 3.62$, where

$s_2 = \text{median } |r_i| = 0.92$

x_i	$r_i=x_i-m_2$	$r_i/6s_2$	$(r_i/6s_2)^2$	$w_i=((1-(r_i/6s_2)^2)^2$
1.20	-2.42	-0.438	0.192	0.653
1.40	-2.22	-0.402	0.162	0.702
2.70	-0.92	-0.167	0.028	0.945
2.80	-0.82	-0.149	0.022	0.957
2.80	-0.82	-0.149	0.022	0.957
2.91	-0.71	-0.129	0.017	0.966
3.30	-0.32	-0.058	0.003	0.994
3.47	-0.15	-0.027	0.001	0.998
3.70	0.08	0.015	0.000	1.000
4.27	0.65	0.118	0.014	0.972
4.95	1.33	0.241	0.058	0.887
5.11	1.49	0.270	0.073	0.859
5.50	1.88	0.341	0.116	0.782
5.92	2.30	0.417	0.174	0.682
6.00	2.38	0.431	0.186	0.663

$$m_3 \sum_{i=1}^{n} (x_i * w_i) / \sum_{i=1}^{n} w_i = 3.70$$

Third iteration:

$m_3 = 3.70$
$s_3 = 1.00$

x_i	$r_i = x_i - m_3$	$r_i/6s_3$	$(r_i/6s_3)^2$	$w_i = ((1 - (r_i/6s_3)^2)^2$
1.20	-2.50	-0.417	0.174	0.682
1.40	-2.30	-0.383	0.147	0.728
2.70	-1.00	-0.167	0.028	0.945
2.80	-0.90	-0.150	0.023	0.955
2.80	-0.90	-0.150	0.023	0.955
2.91	-0.79	-0.132	0.017	0.966
3.30	-0.40	-0.067	0.005	0.990
3.47	-0.23	0.038	0.001	0.998
3.70	0.00	0.000	0.000	1.000
4.27	0.57	0.095	0.009	0.982
4.95	1.25	0.208	0.043	0.916
5.11	1.41	0.235	0.055	0.893
5.50	1.80	0.300	0.090	0.828
5.92	2.22	0.370	0.137	0.745
6.00	2.30	0.383	0.147	0.728

$$M_4 = \sum_{i=1}^{n} (x_i * w_i) / \sum_{i=1}^{n} w_i = 3.73$$

Fourth iteration:

$m_4 = 3.73$

$s_4 = \text{median } |r_i| = 1.03$

x_i	$r_i = x_i - m_4$	$r_i/6s_4$	$(r_i/6s_4)^2$	$w_i = ((1 - (r_i/6s_4)^2)^2$
1.20	-2.53	-0.409	0.167	0.694
1.40	-2.33	-0.377	0.142	0.736
2.70	-1.03	-0.167	0.028	0.945
2.80	-0.93	-0.151	0.023	0.955
2.80	-0.93	-0.151	0.023	0.955
2.91	-0.82	-0.133	0.018	0.964
3.30	-0.43	-0.070	0.005	0.990
3.47	-0.26	0.042	0.002	0.996
3.70	0.03	0.005	0.000	1.000

4.27	0.54	0.087	0.008	0.984
4.95	1.22	0.197	0.039	0.924
5.11	1.38	0.223	0.050	0.903
5.50	1.77	0.286	0.082	0.843
5.92	2.19	0.354	0.125	0.766
6.00	2.27	0.367	0.135	0.748

$$x' = \sum_{i=1}^{n} (x_i * w_i) / \sum_{i=1}^{n} w_i = 3.73$$

Since this value is the same as the previous value, convergence is obtained. To compute the variance of the biweight measure, use the r_i and w_i data values and in the formula for variance and obtain:

var(biweight measure) = 2.840
standard deviation = 1.685

The lower and upper hinges of the tank SER data are 2.80 and 5.11, respectively. Consequently,

standard deviation(batch) = 2.31/1.349 = 1.7124

The tank SER prediction interval is 3.73 + and - (1.69 + 1.71), i.e., 0.33 - 7.13.

Finally, the computation of the force SER data biweight measure x' and the respective weights involve an iterative process. The computation of the biweight measure for the force LER data is as follows:

First iteration:

m_1 = 3.94 the batch median and
s_1 = median $|r_i|$ = 0.71

x_i	$r_i=x_i-m_1$	$r_i/6s_1$	$(r_i/6s_1)^2$	$w_i=((1-(r_i/6s_1)^2)^2$
2.35	-1.59	-0.373	0.139	0.741
3.23	-0.71	-0.167	0.028	0.945
3.35	-0.59	-0.139	0.019	0.962
3.60	-0.34	-0.080	0.006	0.988
3.94	0.00	0.000	0.000	1.000
4.18	0.24	0.056	0.003	0.994
5.27	1.33	0.312	0.097	0.815
5.30	1.36	0.319	0.102	0.806
8.00	4.86	0.000	0.000	0.000

$$m_2 = \sum_{i=1}^{n} (x_i * w_i) / \sum_{i=1}^{n} w_i = 3.89$$

Second iteration:

m_2 = 3.89 the batch median and

s_2 = median $|r_i|$ = 0.66

x_i	$r_i = x_i - m_2$	$r_i/6s_2$	$(r_i/6s_2)^2$	$w_i = ((1 - (r_i/6s_2)^2)^2$
2.35	-1.54	-0.389	0.151	0.721
3.23	-0.66	-0.167	0.028	0.945
3.35	-0.54	-0.136	0.019	0.962
3.60	-0.29	-0.073	0.005	0.990
3.94	0.05	0.013	0.000	1.000
4.18	0.29	0.073	0.005	0.990
5.27	1.38	0.349	0.122	0.771
5.30	1.41	0.356	0.127	0.762
8.00	4.91	0.000	0.000	0.000

$$m_3 = \sum_{i=1}^{n} (x_i * w_i) / \sum_{i=1}^{n} w_i = 3.88$$

Third iteration:

m_3 = 3.88 the batch median and

s_3 = median $|r_i|$ = 0.65

x_i	$r_i = x_i - m_3$	$r_i/6s_3$	$(r_i/6s_3)^2$	$w_i = ((1 - (r_i/6s_3)^2)^2$
2.35	-1.53	-0.392	0.154	0.716
3.23	-0.65	-0.167	0.028	0.945
3.35	-0.54	-0.139	0.019	0.962
3.60	-0.28	-0.072	0.005	0.990
3.94	0.06	0.015	0.000	1.000
4.18	0.30	0.077	0.006	0.998
5.27	1.39	0.356	0.127	0.762
5.30	1.42	0.364	0.133	0.752
8.00	4.92	0.000	0.000	0.000

$$x' = \sum_{i=1}^{n} (x_i * w_i) / \sum_{i=1}^{n} w_i = 3.88$$

This value is the same as the previous iteration, i.e. convergence is obtained. To compute the variance of the biweight measure, use the last r_i and w_i data values in the variance formula and obtain:

variance(biweight) = 1.2889
standard deviation = 1.1353

The lower and upper hinges of the LER data are 3.29 and 5.29, respectively. Consequently,

standard deviation(batch) = 2.00/1.349 = 1.4826

The force LER prediction interval is 3.88 + and - (1.14 + 1.48), i.e., 1.26 - 6.50.

APPENDIX E

A TEST FOR COMPARING BATCHES: BATCH MEDIAN SCREENING INTERVAL COMPUTATION

A TEST FOR COMPARING BATCHES: BATCH MEDIAN SCREENING INTERVAL COMPUTATION

The MOE batch data screening interval for batch comparison is computed as follows:

$$median + or - ((0.793) * Z_{\alpha/2} * (H\text{-}spread)/\sqrt{n})$$

The lower and upper hinges for the helicopter SER data are 8.33 and 10.60, repectively. The 95% screening interval is 9.27 + or - (0.793) * (1.96) * 2.27/3.606, i.e., 8.29 - 10.25.

The lower and upper hinges for the tank SER data are 2.80 and 5.11, respectively. The 95% screening interval is 3.47 + or - (0.793) * (1.96) * 2.31/3.873, i.e., 2.54 - 4.40.

The lower and upper hinges for the force LER data are 3.29 and 5.29, respectively. The 95% screening interval is 3.94 + or - (0.793) * (1.96) * 2.00/3.00, i.e. 2.90 - 4.98.

APPENDIX F

REFERENCES

1. Barnett, Victor and Lewis, Toby, Outliers in Statistical Data, New York, New York, John Wiley & Sons, 1984.

2. Cleveland William S., Tools for the Data Analyst, Murray Hill, New Jersey, Bell Laboratories, 1982.

3. Gross, Alan M., Confidence Intervals for Bisquare Regression Estimates, Journal of the American Statistical Association, 1977, Vol 72, Pages 341-354.

4. Gross, Alan M., Confidence Interval Robustness with Long-Tailed Symmetric Distributions, Jouurnal of the American Statistical Association, 1976, Vol 71, Pages 409-416.

5. Hogg, Robert V. and Craig, Allen T., Introduction to Mathematical Statistics, 4th Edition, New Yrok, Macmillan Publishing Co, 1978.

6. Hoaglin, David C. and Velleman, Paul F., Application, Basics, and Computing of Exploratory Data Analysis, Boston, Massachusetts, Duxbury Press, 1981.

7. Hoaglin, David C., Mosteller, Frederick, and Tukey, John W., Understanding Robust and Exploratory Data Analysis, New York, New York, John Wiley & Sons, Inc., 1983.

8. Kendall, M.G. and Stuart, A., The Advanced Theory of Statistics, Vol 1, New York, Hafner Publishing, 1969.

9. Mallows, Colin L., Robust Methods-Some Samples of Their Use, The American Statistician, 1979, No. 33, Pages 179-184.

10. McGill, Robert, Tukey, John W., and Larsen, Wayne A., Variations of Box Plots, The American Statistician, 1978, Vol 32, Pages 12-16.

11. McNeil, Donald R., Interactive Data Analysis, New York, New York, John Wiley & Sons, Inc, 1977.

12. Mosteller, Frederick and Tukey, John W., Data Analysis and Regression, Reading, Massachusetts, Addison-Wesley, 1977.

13. Tukey, John W., Exploratory Data Analysis, Reading Massachusetts, Addison-Wesley, 1977.

14. Memorandum, ATRC-WP, Subject: Management of TRADOC Analysis, 16 June 1987.

DISTRIBUTION LIST

No. of Copies

Commander 1
U. S. Training and Doctrine Command
ATTN: ATCD
Fort Monroe, Virginia 23651-5000

HQDA 1
Deputy Under Secretary of the Army
for Operations Research
ATTN: Mr. Walter W. Hollis
Room 2E660, The Pentagon
Washington, D. C. 20310-0102

Deputy Director for Force Structure 1
R&A, J-8
Office Chief of Staff for Operations and Plans
Room 1E965, The Pentagon
Washington, D. C. 20301-5000

HQDA 1
Office of the Technical Advisor
Deputy Chief of Staff for Operations and Plans
ATTN: DAMO-ZD
Room 3a538, The Pentagon
Washington, D. C. 20310-0401

Commander 1
U.S. Army TRADOC Analysis Command
ATTN: ATRC
Fort Leavenworth, Kansas 66027-5200

Director
U.S. Army TRADOC Analysis Command-Fort Leavenworth
ATTN: ATRC-F 1
ATRC-FOQ (Document Control) 1
Fort Leavenworth, Kansas 66027-5200

Director
U.S. Army TRADOC Analysis Conmmand-White Sands
Missile Range
ATTN: ATRC-W
ATRC-WSL (Technical Library) 1
White Sands Missile Range, New Mexico 88002-5502

Director 2
TRADOC Analysis Command
Requirements & Programs Directorate
ATTN: ATRC-RP

Fort Monroe, Virginia 23651-5143

Director U.S. Army Concepts Analysis Agency 8120 Woodmont Avenue Bethesda, Maryland 20814-2797	1
Director U.S. Army Materiel Systems Analysis Activity ATTN: AMXSY Aberdeen Proving Ground, Maryland 21005-5071	1
U.S. Army Combined Arms Research Library (CARL) ATTN: ATZL-SWS-L Fort Leavenworth, Kansas 66027	1
Defense Technical Information Center ATTN: DTIC, FDAC Building 5, Cameron Station Alexandria, Virginia 22304-6145	2
U.S. Army Library Army Study Documentation and Information Retrieval System (ASDORS) ABNRAL-RS Room: ASDIRS Room 1a518, The Pentagon Washington, D. C. 20310	1