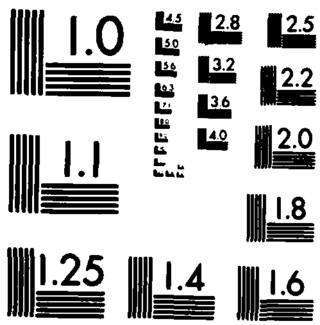
AD-A146 675 A STATISTICAL ANALYSIS PLAN TO SUPPORT THE JOINT FORWARD AREA AIR DEFENSE TEST(U) MILITARY ACADEMY WEST POINT N Y DEPT OF ENGINEERING G H PARLIER 02 AUG 84 1/1

UNCLASSIFIED F/G 15/3 NL



'OPY RESOLUTION TEST CHART

A

STATISTICAL ANALYSIS PLAN TO SUPPORT THE JOINT FORWARD AREA AIR DEFENSE TEST

By

Captain Greg H. Parlier
Department of Engineering
United States Military Academy
West Point, New York 10996
2 August 1984

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A146675
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) A STATISTICAL ANALYSIS PLAN TO SUPPORT THE JOINT FORWARD AREA AIR DEFENSE TEST	5. TYPE OF REPORT & PERIOD COVERED Final Technical Report
	6. PERFORMING ORG. REPORT NUMBER NA
7. AUTHOR(s) Captain Greg H. Parlier	8. CONTRACT OR GRANT NUMBER(s) NA
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Engineering US Military Academy West Point, NY 10996	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NA
11. CONTROLLING OFFICE NAME AND ADDRESS Office of the Under Secretary of Defense for Research and Engineering, Joint Forward Area Air Defense, P.O. Box 16030, Fort Bliss, TX 79906	12. REPORT DATE 2 August 1984
	13. NUMBER OF PAGES 45
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE NA

16. DISTRIBUTION STATEMENT (of this Report)

Unlimited Distribution

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

NA

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Aircraft identification
Airspace management
Air defense command and control
Analysis
Forward area air defense
Joint Forward Area Air Defense
Non-parametric statistics
Parametric statistics
Statistics
Test design
Test and evaluation

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

The Forward Area Air Defense Joint Test Force (JFAAD) was chartered to improve forward area air defense performance in the division area, and reduce friendly air casualities due to ground-based air defense. An analytical approach to accomplishing these purposes was published 30 March 1984 in the JFAAD Test Program Definition (TPD). This document provides an investigative methodology and statistical analysis plan to spport the JFAAD TPD. The methodology is developed in three parts:

- Theoretical statistical issues are discussed in the context of the JFAAD test issues.
- A general statistical methodology is proposed, and detailed statistical analysis plans are outlined in flow diagram form.
- Conclusions are drawn from the study effort and recommendations are offered.

ACKNOWLEDGEMENT

In addition to the entire Analysis Directorate at JFAAD all of whom treated me as one of their own during my short stay, I especially thank Major Ken Matwiczak, Left Fielder, for affording me the opportunity and Major Kevin Dolan, Resident Statistician, for sharing his cell.

Also at West Point, Captain Con Crane, Department of History; Major Steve Williams, Department of Engineering; Lieutenant Colonel Jerome Adams, Science Research Lab; and especially to Major Rick Kolb, Permanent Associate Professor, Department of Mathematics, for illuminating the promise of multi-dimensional contingency tables. A special thanks to Ms Betty Melick for typing the manuscript, as usual, on short notice.

TABLE OF CONTENTS

I. INTRODUCTION

A. PURPOSE

The purpose of this paper is to provide an investigative methodology and useful statistical analysis plan to support the Test Program Definition (TPD) recently developed by the Joint Forward Area Air Defense (JFAAD) Test Force, [Ref. 7]. The two primary test Force objectives are to improve forward area air defense performance, reduce friendly air casualties due to ground-based air defense and also to identify joint tactical, doctrinal, and procedural changes which assist in attaining these objectives. These objectives necessitate an in-depth examination of three major issues:

1. To what degree do the collective means of aircraft identification influence the effectiveness of the forward area air defense systems?
2. To what degree do projected C^3I capabilities support JFAAD elements?
3. How does airspace management and control affect the mission accomplishment of FAAD systems and friendly aircraft?

Detailed "Patterns of Analysis" for each of these issues are contained in the TPD.

B. OBJECTIVES

The primary goal of this effort is the development of an investigative methodology that contains specific statistical analysis plans for each major issue. It is desired that such plans be:

1. Theoretically plausible - Analysis plans should be based upon sound theoretical considerations that provide not only credibility of statistical test results, but also

provide for maximum power in announcing such results.

2. Flexible - The analysis plan should be sufficiently flexible to provide multiple statistical procedures for consideration yet allow the final choice of a specific statistical test, or tests, to be suggested by the actual data generated from the testbed simulation.

After developing an analysis approach, including comparisons of its advantages and disadvantages with other statistical approaches, the statistical plans will be implemented by performing a "pilot test" on manually developed, "dummy" data. It is anticipated that such an effort will provide useful results for the JFAAD Test Force and also provide some guidance in identifying options and potential problem areas as the Analysis Directorate develops, refines, and executes its Master Analysis Plan. At best or at worst, respectively, it is offered as a means of providing either:

1. The primary statistical analysis methodology to support the JFAAD TPD, or

2. A theoretically valid and reasonable alternative, available as a back-up or adjunct analysis plan if needed.

C. CONTENT

The development of the statistical analysis methodology has been partitioned as follows:

1. Theoretical issues involved in statistical analysis are discussed in the context of the JFAAD issues:

- Type I and Type II errors are explained,
- Random sampling with and without replacement in a discrete counting process are reviewed,
- The two major categories of statistical tests (Parametric and Nonparametric) are compared and contrasted,
- The importance of sample size and its relation to power efficiency and asymptotic relative efficiency (ARE) is reviewed.

2. A general statistical analysis methodology is proposed in consonance with the theoretical issues previously emphasized. The advantages and disadvantages of the methodology are cited. Detailed statistical analysis plans are then outlined in flow diagram form to support the patterns of analysis for each of the three test issues:

- Aircraft ID Statistical Analysis Plan,
- C^3I Statistical Analysis Plan,
- Airspace Management Statistical Analysis Plan.

3. Conclusions are drawn from this study effort and recommendations offered.

II. STATISTICAL ANALYSIS METHODOLOGY

A. SOME THEORETICAL CONSIDERATIONS

1. Random Sampling With and Without Replacement.

One of the difficulties encountered in developing a statistical analysis methodology to support the JFAAD TPD is the form of the underlying distribution of data. Classical parametric statistics assume the distributional form of the random variable to be both continuous and normal. However, in most cases the data will not be in this form but rather is obtained from a discrete counting process generated by a dichotomous (binary) random variable (e.g., number of detections, number of engagements, etc). While one might immediately surmise the applicability of the binomial distribution as a valid probability model, consideration must be given to the sampling process. In examining the early warning MOP, for example, the number of detections(c) out of n engagement opportunities may be viewed as sampling with replacement. In this instance, use of the binomial distribution to model number of detections (X) can possibly be justified on theoretical grounds, assuming that detection from one weapon system (or crew) to the next can be considered independent:

$$P(X = c) = \binom{n}{c} p^c (1 - p)^{n-c}$$

$$\mu = np$$

$$\sigma = np\,(1 - p)$$

However, in examining the number of aircraft killed, the binomial is no longer fully justified. Sampling is now occurring without replacement, hence the appropriate distribution to consider is the hypergeometric with its finite population correction factor (N -n/N -1):

$$P(X = c) = \frac{\binom{Np}{c}\binom{N(1-p)}{n-c}}{\binom{N}{n}}$$

where n is the sample size
N is the total population size

$$\mu = np$$

$$\sigma^2 = np\,(1-p)\,\left(\frac{N-n}{N-1}\right)$$

Fortunately, both the binomial and hypergeometric distributions can be approximated by the normal distribution. The approximation improves for increasingly larger sample sizes and provides acceptable results conditional upon $np > 5$, $n(1-p) > 5$, and, for the hypergeometric, $n/N < .1$.

For values of $np < 5$, especially when p is very small and n is large, it is suggested that the Poisson distribution be used to approximate the binomial, with $\lambda = np$.

$$P(X = c) = \frac{e^{-\lambda}\lambda^c}{c!}$$

$$\mu = \sigma^2 = \lambda = np$$

2. Statistical Independence

In a discrete counting process, such as those modeled by the binomial, hypergeometric, and Poisson distributions, a second important theoretical consideration involves the fundamental concept of independence which is a major assumption in statistical theory. Two events A and B are said to be statistically independent if the occurrence or nonoccurrence of A has no efffect on the probability of B and vice versa. The existence, or assumed existence, of independent Bernoulli trials is another condition, in addition to random sampling with replacement, that should be met when using the binomial distribution to model a counting process. In the JFAAD context, the indiscriminate use of the binomial distribution, and tests of proportion based upon normal approximations to the binomial, must be guarded against because the assumption of independence is rarely likely to be valid, depending upon the fidelity of the testbed simulation. For example, the likelihood of Stinger Team B visually detecting an aircraft that has just overflown Stinger Team A's adjacent position may be considered to be independent if communications between the teams are nonexistent. However, if communications do exist, Team B's visual sighting probability becomes conditional upon whether or not Team A visually detects the target and transfers early warning information. Thus, Team B's detection event is dependent, to some degree, upon Team A's success in its target detection event. This sort of conditional dependence, or correlation, among events, in addition to the complexity of the situation within the JFAAD region, is what necessitates performing replications of a large scale simulation.

3. Parametric and Nonparametric Statistics

Two general categories of statistical testing procedures are available: parametric and nonparametric (or distribution-free) tests. The theory for classical parametric procedures is well-established and allows a wide range of hypotheses to be tested. The use of such tests is very common, especially for large-scale experiments requiring an examination of the effects of multiple independent variables upon one or more response variables (e.g., ANOVA, MANOVA, fractional factorial designs, etc.). Although ANOVA techniques are relatively robust parametric tests they do require several assumptions, some more critical than others. When critical assumptions are not sufficiently met ANOVA tests may provide biased, even erroneous results. In contrast, nonparametric techniques require fewer assumptions and, in cases where assumptions essential for ANOVA cannot be sufficiently met, provide unbiased and more powerful tests than their parametric equivalent.

Consequently, it appears desirable to develop a statistical analysis plan that allows either approach to be used. Such an approach encourages selection of a particular test, either parametric or nonparametric (or perhaps both), to be based upon the distributional form of the data to be analyzed. This approach provides substantial flexibility and offers several advantages:

- The decision to use either parametric or nonparametric procedures can be suggested by the data thus enabling selection of the most powerful test available.
- In those instances where reasons for selecting one procedure over the other are not particularly compelling, then both types of tests can be performed and the results from the two can be compared and contrasted; perhaps the results will be mutually supportive.
- Regardless of which approach is ultimately selected, both approaches contain specific tests which answer essentially equivalent hypotheses.

4. Statistical Error and Power Efficiency

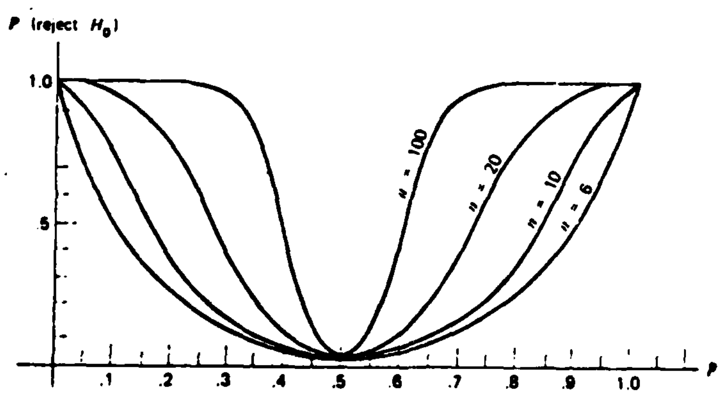
The two types of statistical error that can occur are illustrated in the chart below [Ref. 2, pg. 29]:

		The Decision	
		Accept H_0	Reject H_0
The true situation	H_0 is true	Correct decision probability = $1-\alpha$	Type I error probability = α (level of significance)
	H_0 is false	Type II error probability = β	Correct decision probability = $1-\beta$ (power)

Figure 1

Normally, α (probability of a Type I error) is controlled by establishing a specific significance level prior to performing the statistical test (traditionally α levels are set at .01 or .05). What is often neglected however, is β (probability of Type II error). In many, if not most, of the JFAAD issues, the commission of a Type II error is more severe than commission of a Type I error. For example, failing to detect a difference in early warning (EW) techniques when in fact there is a difference (i.e., a Type II error), can certainly be more disastrous than declaring that a difference exists when in fact there really is none (a Type I error). Hence, while α can be controlled by pre-selection, β is often ignored and, for small sample sizes, may be unacceptably high even though it is of substantially more importance in the context of the JFAAD issues and questions requiring resolution.

Unfortunately, a decrease in α usually results in an undesirable increase in β. For constant α levels, the power (1 - β) of a statistical test can be increased by increasing the sample size of the test. [Ref. 2, pg. 87].



A comparison of several power functions.

Figure 2

Obviously, given a choice between two statistical tests one would prefer to choose that test which achieves maximum power for a specific sample size and α level. Power efficiency, or more commonly "asymptotic relative efficiency" (ARE), is a measure of comparison between two tests (a and b) and indicates which test requires the smaller sample size (n) to achieve specified α and power levels:

$$ARE = \lim_{n_a \to \infty} \left[\frac{n_b}{n_a}\right]$$

Comparisons between parametric and nonparametric procedures have been performed in an effort to determine which procedure offers the more powerful (or equivalently, more efficient) test results. These results are provided below. [Ref. 9, pg. 87]:

Design	Test	Distribution: Normal	Uniform	Logistic	Double exponential	Limiting (lower limit)
One-sample	Sign test	$\frac{2}{\pi} = 0.637$	0.333	0.750	2.000	0.333
	Wilcoxon matched-pair	$\frac{3}{\pi} = 0.955$	1.000	1.047	1.500	0.864
Two-sample (independent)	Median test	$\frac{2}{\pi} = 0.637$	0.333	0.750	2.000	0.333
	Mann-Whitney	$\frac{3}{\pi} = 0.955$	1.000	1.047	1.500	0.864
	Normal scores	1.000			$\frac{4}{\pi} = 1.273$	1.000
Many-sample (independent)	Median test	$\frac{2}{\pi} = 0.637$	0.333	0.750	2.000	0.333
	Kruskal-Wallis	$\frac{3}{\pi} = 0.955$	1.000	1.047	1.500	0.864
	Normal scores	1.000			$\frac{4}{\pi} = 1.273$	1.000

Figure 3

Due to the distribution-free assumption, exact power calculations for nonparametric 2-sample and many-sample tests are not available. However, for parametric ANOVA, procedures and tables exist to determine not only the power of a particular test, but more importantly, sample size necessary to achieve a desired power level prior to conducting the test. Such an investigation will aid immeasurably in developing a more powerful experimental design. Refer to Annex C.

5. Which Approach? A Comparison of Parametric vs. Nonparametric Statistical Tests

In an attempt to compare and contrast the relative advantages and disadvantages of parametric with nonparametric statistics, a thorough analysis of the assumptions required for each approach is necessary. Especially valuable is a recognition of the effects if any, violations of such assumptions will have. Such an analysis should prove beneficial in the attempt to identify the "best" statistical test for a particular hypothesis.

In general violations of assumptions in parametric tests affect both the sensitivity and the significance level of the test. For example, violations of assumptions in the one-way ANOVA usually cause the F test to become less efficient in detecting differences and to announce too many significant differences. Appropriate parametric tests (e.g., t-tests and one-way ANOVA) designed to detect differences among alternative treatments (e.g., different types of early warning or identification procedures) require data from the treatment groups to be:

- normally distributed,
- homoscesdastic (of equal variance), and
- independent among experimental units.

The second assumption requiring constant variance is analogous to asserting that the means and variances of the treatments must be independent. Although data expressed as proportions, or percentages, violate both the normality (unless the sample size is sufficiently large to invoke the Central Limit Theorem) and constant variance assumptions (e.g., for the binomial distribution, the variance, σ^2 = npq is a function of the mean = np), the arcsin, or angular, transformation is regarded as an appropriate way to convert the data such that parametric tests then become theoretically plausible even for small sample sizes:

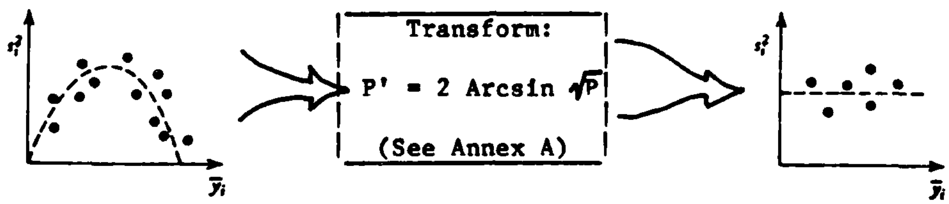


Figure 4

The effect of data transformation, which corrects for the original lack of normality and variance stability in the data, coupled with the known "robustness" of the variance-ratio F tests used in the one-way ANOVA, then allow such parametric tests to be used without fear of attaining erroneous results (See also, The Analysis of Binary Data, by D.R. Cox, on use of the empirical logistic transform).

However, when advancing beyond the one-way ANOVA to the higher-order factorial designs which are the commonly used parametric tests for large-scale experiments (e.g., two-way ANOVA, blocked designs, fractional factorials, etc.), an additional, and much more critical assumption must be considered: The requirment for additivity of treatment and interaction effects in a linear model. In the case of the two-way ANOVA with multiple observations per cell, the linear model is:

$$X_{ijk} = \mu + \tau_i + \Theta_j + \psi_{ij} + \varepsilon_{ijk}$$

where the assumption is made that each observation may be expressed as the algebraic sum of:

1. an overall mean, u
2. a "row effect", τ_i
3. a "column effect", Θ_j
4. an interaction effect, ψ_{ij}, and
5. an experimental error (residual), $\varepsilon_{ijk} \sim N(0,\sigma^2)$

The additional assumption involved, under the null hypotheses, is that of additivity among all effect terms. In the case of factorial type experiments, the assumption of additivity is rarely realistic as evidenced by the numerous interaction terms that are realistically significant. In the case of the two-way ANOVA, if the interaction term is significant, one may conclude that row and column factors are affecting the observations, and explore these effects through other procedures (e.g., one-way ANOVA's on each row and column). A significant interaction suggests that the effects of row and column effects are not additive, and accordingly the two-way ANOVA model is no longer appropriate in terms of testing for pure row and column effects. The magnitude of the errors that would result from continuing to test for row and column effects is not known. In practice such tests are frequently performed even though it is known that the additive model is incorrect. One should always test interaction terms first: if interaction term(s) are insignificant, then a pooled mean square error term can be used to test for main effects; if interactions are significant, then only those main effects not involved in significant interactions should subsequently be tested. Transformations also exist to reduce the effects of nonadditivity (so-called "transformable non-additivity"). However the art of transforming data in an effort to achieve additivity is far less developed than similar methods used to achieve normality and variance stabilization. Such attempts might be practical for small factorial designs (such as a two-way ANOVA with only one interaction term) but rapidly become impractical for large factorial designs involving numerous interaction terms.

Many of the assumptions required by parametric ANOVA appear to be difficult, at best, to meet:

1. Continuity of data (much of the data is binary giving rise to the analysis or proportions),
2. Normality and variance stability (proportions from a binomial distribution are normally distributed in the limit only, and the binomial distribution exhibits dependence between the mean and its variance).
3. Additivity in the linear model (numerous interactions can realistically be expected thus decreasing precision in attempts to measure differences in the main effects which are of primary concern to the JFAAD issues), and
4. Confidence intervals on estimates can be decreased only by increasing the sample size (large sample sizes may be cost prohibitive).

An alternative to the parametric approach is offered by various nonparametric techniques. The major advantage offered by nonparametrics is the lack of any distributional assumptions, hence the phrase "distribution-free" and "assumption-free" statistics. The major disadvantage of most of the nonparametric tests, with the exception of those based upon "normal scores", is the small power relative to comparable parametric tests since only part of the information contained in the data (usually based upon ranks) is utilized in the statistical decision. Unlike the relative plethora of parametric ANOVAs that abound in experimental design, nonparametric procedures have not been developed to support analysis of large factorial designs. However, as the ARE entries in Figure 3 show, some of the one-way ANOVA equivalents (2-sample and many-sample tests) possess AREs with lower bounds of 1.0. This indicates that these nonparametric tests have the same asymptotic efficiency as their parametric counterparts when the population is really normal and even larger asymptotic efficiences when the population is non-normal. Thus, when normality assumptions cannot be satisfied these tests provide more powerful results than comparable parametric tests such as the t- and F-tests.

B. ANALYSIS METHODOLOGY

Although a large factorial design approach with multiple (> 2) levels for each treatment factor (MOE and MOP) appears to offer a comprehensive and efficient "macroscopic" view of all the factors involved in an entire issue, or even multiple issues, such an approach is conditional upon the validity of the previously discussed assumptions involved in any parametric experimental design, especially the assumption of linear additivity. The large quantity of interaction terms, many of which should realistically reveal themselves to be statistically significant, jeopardize precise judgements of the main effects. Additionally, providing a reasonable interpretation to the meaning of many of the higher-order interaction terms may prove futile; for example, the highest order interaction term in a factorial design for the identification issue will be a cross product term involving seven elements (since there are seven factors in the ID issue ranging from the flight profile MOP up through the dendrite to the ID system issue at the top of the pattern of analysis). As a consequence of the vast amount of "background noise" (significant interactions) it becomes virtually impossible to detect any main effect differences, even if they do exist. The magnitude of the experimental design consequently becomes counterproductive to the statistical analysis effort because the main effect factors (MOP and MOE) have become inextricably entwined in such a large number of interactions (recall that any main effect involved in any significant interaction term cannot subsequently be tested with precision for possible differences among its levels).

The large factorial design's inability to address the crucial main effects questions with any precision clearly argues for a more satisfactory approach in the analysis effort. An alternative approach, admittedly "microscopic" and thus more tedious, involves sequential iterations of two-way and one-way ANOVAs which successively examine each layer of MOPs and MOEs within the pattern of analysis for each major issue. This methodical approach allows main effects (the MOEs and MOPs at each "tier" in the pattern of analysis) to be addressed with much greater precision and for confidence bands to be established using multiple comparison tests in those instances where significant main effects are identified. When interactions between progressively lower MOPs are identified, differences within each MOP can still be explored through one-way ANOVAs with precision. Hence, the problems posed by significant interaction terms (which precluded tests on main effects in the large factorial design) can now be circumvented. Perhaps the greatest benefit offered by this "one-step-at-a-time" approach, is the opportunity to also employ nonparametric tests. If data analysis reveals that parametric ANOVA assumptions are questionable or clearly invalid, the alternative use of nonparametric analogues for one-way ANOVAs now exists. In fact, the high ARE for the 2- and many-sample comparison tests (refer to Figure 3) strongly encourages their use regardless of the adequacy of comparable parametric tests. This approach, illustrated below in "generic" form, allows either classical parametric ANOVA, nonparametric tests of independence (which are equivalent to "independence," or lack of interaction, between row and column effects in the two-way ANOVA) and tests of comparison, or both parametric and nonparametric tests to be used.

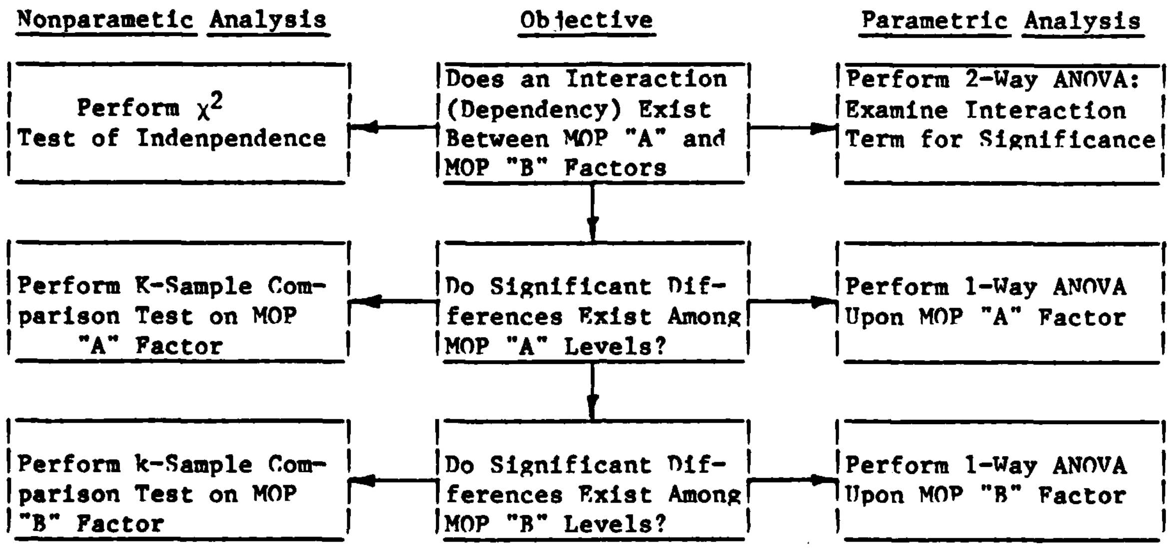


Figure 5

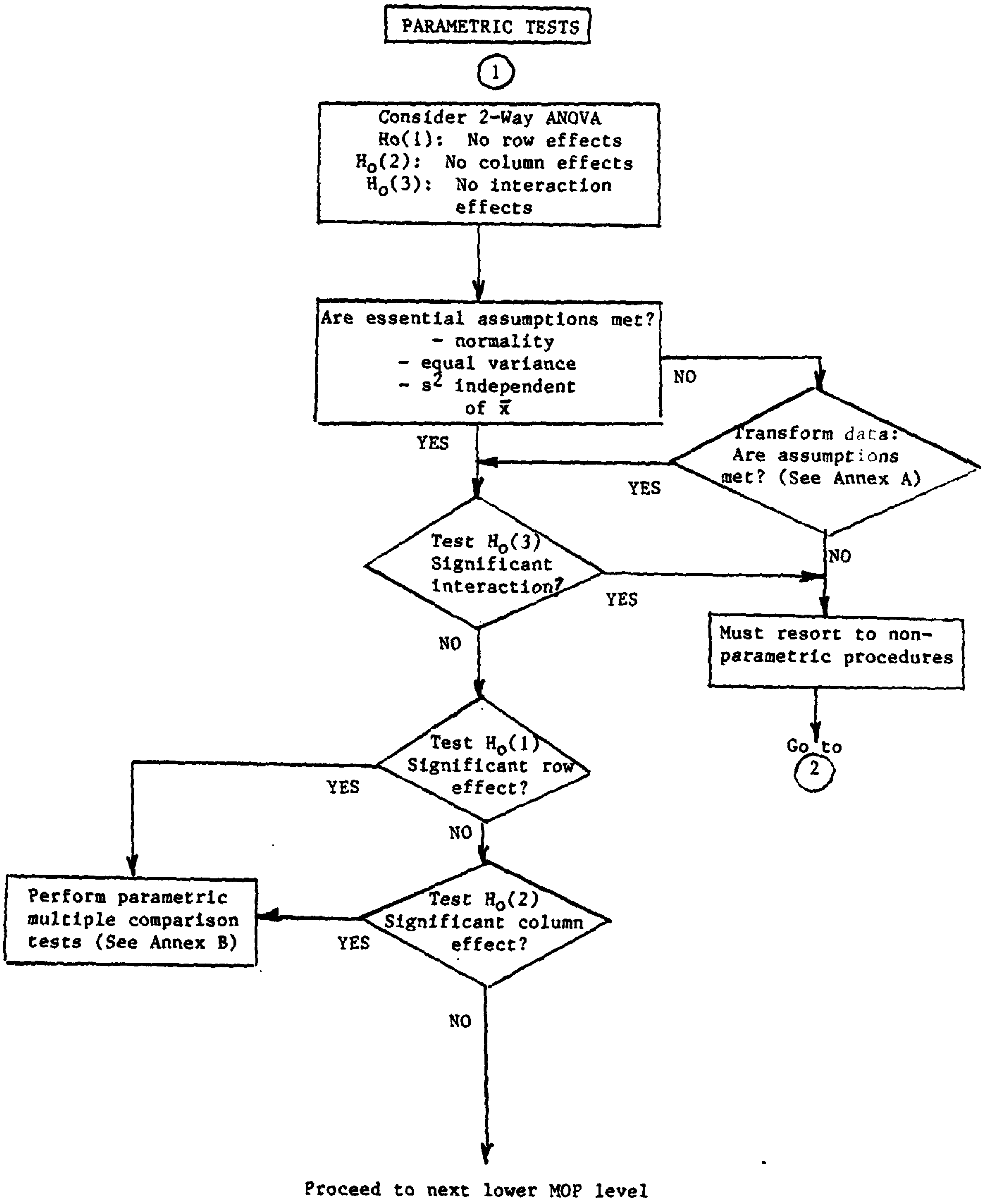
This approach appears to provide the opportunity to use both parametric and nonparametric statistics:

1. The use of known data transformations (e.g., logistic and angular transforms) provide a promising means to partially compensate for small sample sizes while simultaneously converting the data such that it sufficiently approximates the parametric assumptions of continuity, normality, variance stability, and linear additivity.

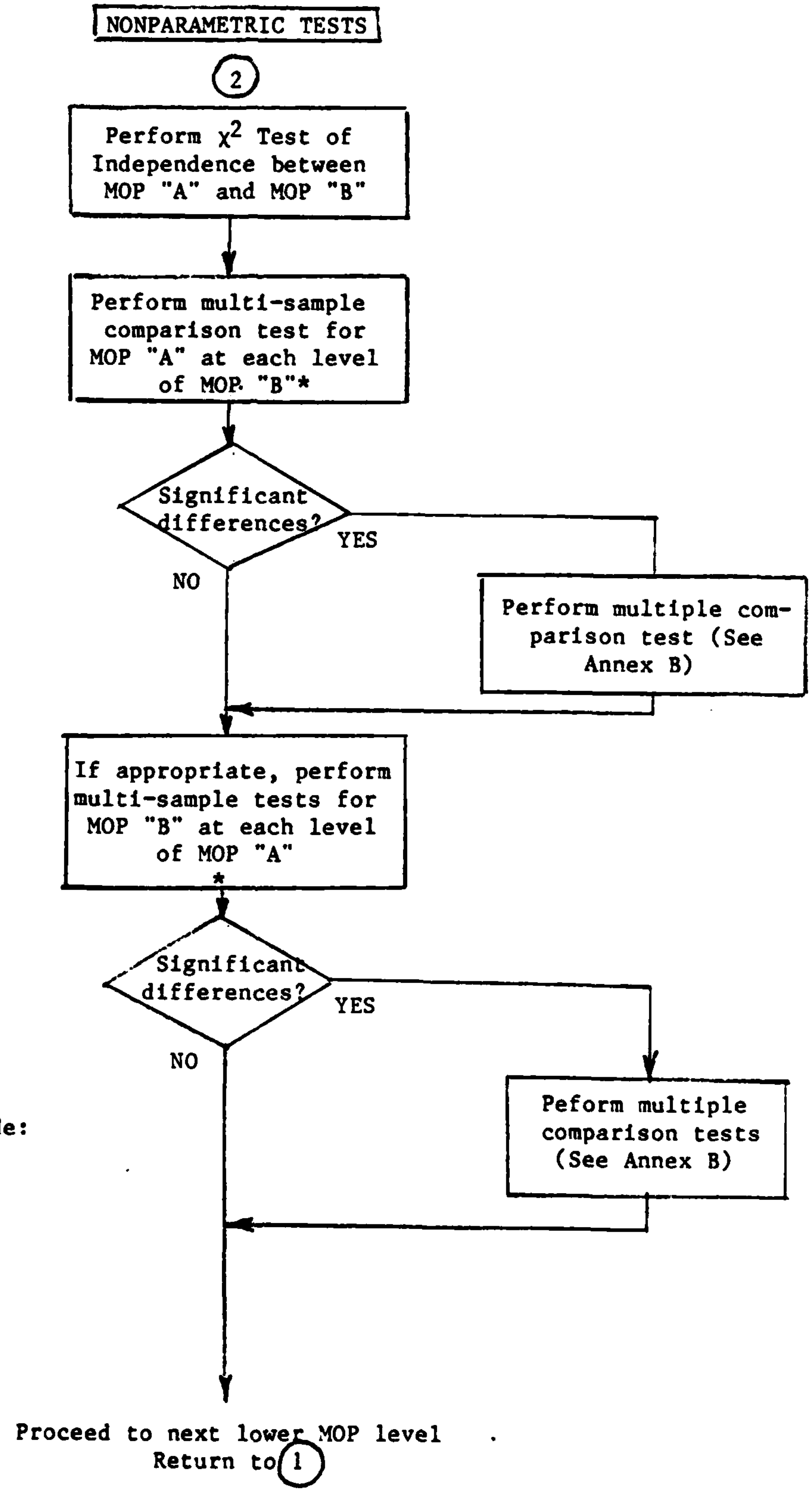
2. The high ARE nonparametric, multi-sample tests become available for use and provide test results at least as powerful as parametric ANOVA.

An appealing attribute of this approach, not available in large scale parametric designs, is the opportunity to compare and contrast decisions on statistical hypotheses obtained from both parametric and nonparametric statistical tests. When both types of tests can be used with sound theoretical justification, identical decisions to either reject or not reject a null hypothesis will provide mutual reinforcement to each other lending greater credibility to the accuracy of the decision. When the two procedures yield different results, the reason for the disparity can be examined in an attempt to determine which result provides greater credibility (e.g., a significant difference in a Kruskal-Wallace test may conflict with a non-significant main effect result from a two-way ANOVA test; closer scrutiny may reveal a marginally non-significant interaction term which suggests that the additivity is questionable in the two-way ANOVA, hence greater credibility should be attached to the Kruskal-Wallace results than those obtained from ANOVA).

Specific statistical tests for both parametric and nonparametric analysis are outlined in flow diagram form on the following pages. The statistical test logic is applicable to all three patterns of analysis.



PARAMETRIC TESTS
1
Consider 2-Way ANOVA
Ho(1): No row effects
$H_o(2)$: No column effects
$H_o(3)$: No interaction effects
Are essential assumptions met?
- normality
- equal variance
- s^2 independent of $\bar{x}$
NO
YES
Transform data: Are assumptions met? (See Annex A)
YES
NO
Test $H_o(3)$ Significant interaction?
YES
NO
Must resort to non-parametric procedures
Go to
2
Test $H_o(1)$ Significant row effect?
YES
NO
Perform parametric multiple comparison tests (See Annex B)
Test $H_o(2)$ Significant column effect?
YES
NO
Proceed to next lower MOP level

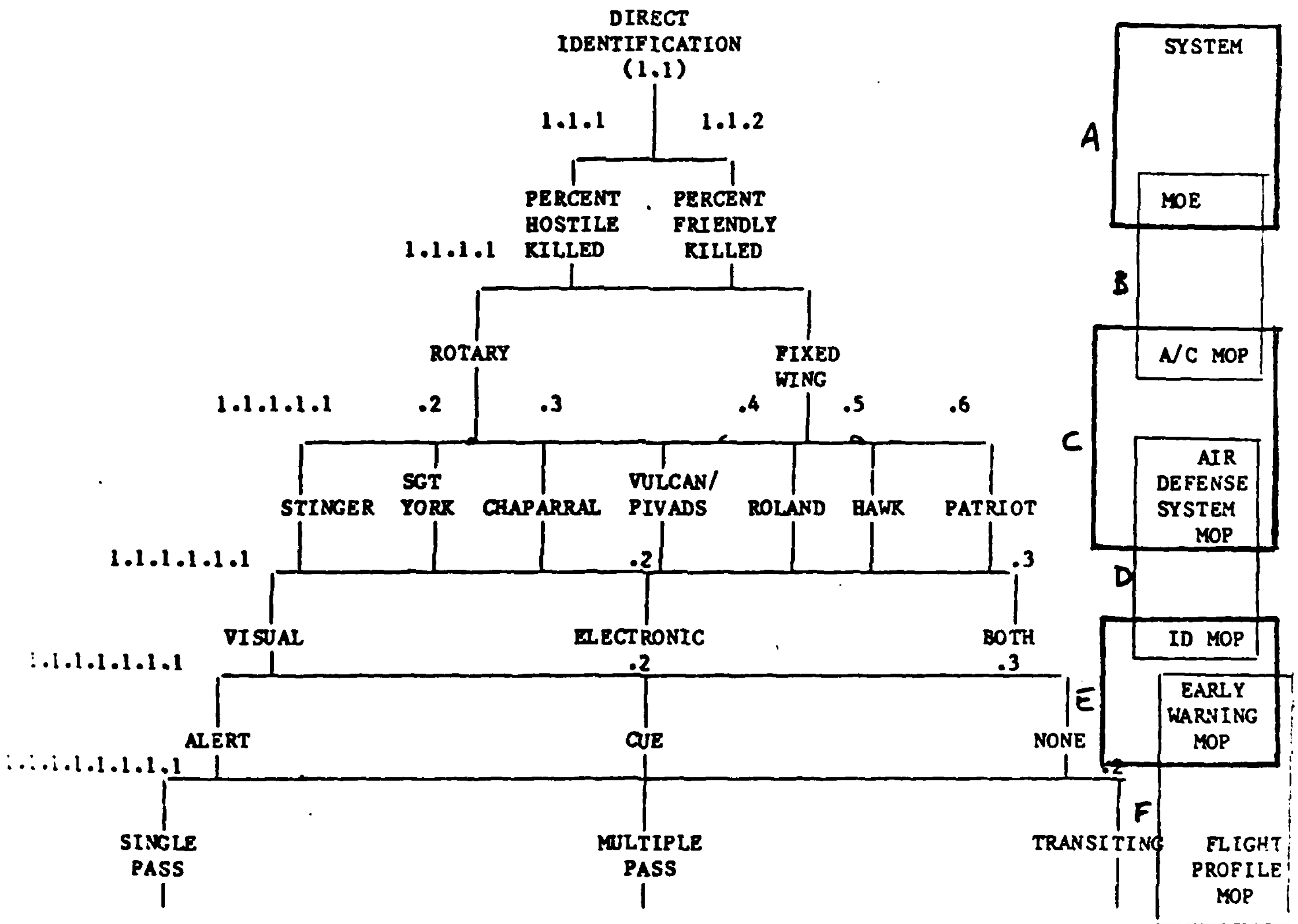


*Available nonparametric multi-sample tests include:

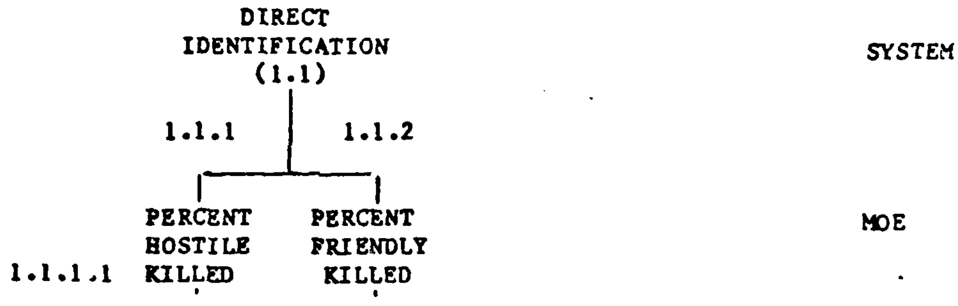
<u>2 - Sample:</u>
- median test
- Mann-Whitney
- Normal scores

<u>Multi-sample:</u>
- Quade
- Kruskal-Wallace
- Friedman
- Van Der Waerden

1. TOP-DOWN, SEQUENTIAL ANALYSIS OF VARIANCE APPROACH APPLIED TO THE IDENTIFICATION ISSUE PATTERN OF ANALYSIS



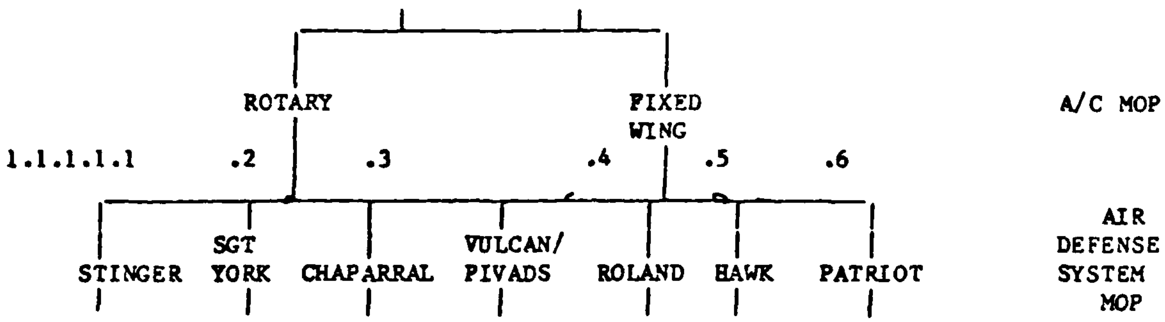
Pattern of Analysis for Direct Identification System.



<u>ANALYSIS PLAN: A</u>

1. Which ID system type (indirect or direct) kills more hostile A/C?
2. Which ID system results in lowest fratracide?

Techniques outlined in II.B are directly applicable

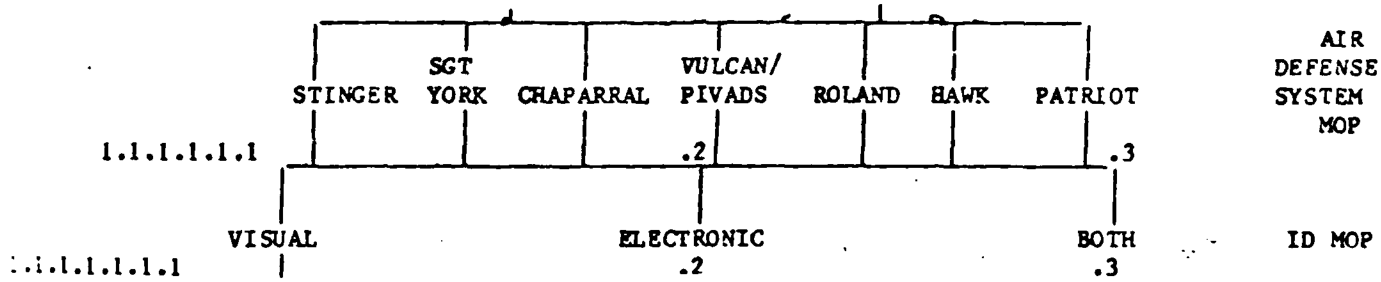


ANALYSIS PLAN: C

Questions:

1. Are different AD systems more effective against rotary wing than fixed wing? (interaction)
2. Is there an overall difference between A/C MOP for each ID system?
3. Which AD systems contribute most to the MOEs?
4. Does ID system effect AD system performance? If so, how? (e.g., do SHORAD systems account for more, or less, of the MOEs?) (cross system investigation)

Techniques outlined in II.B are directly applicable

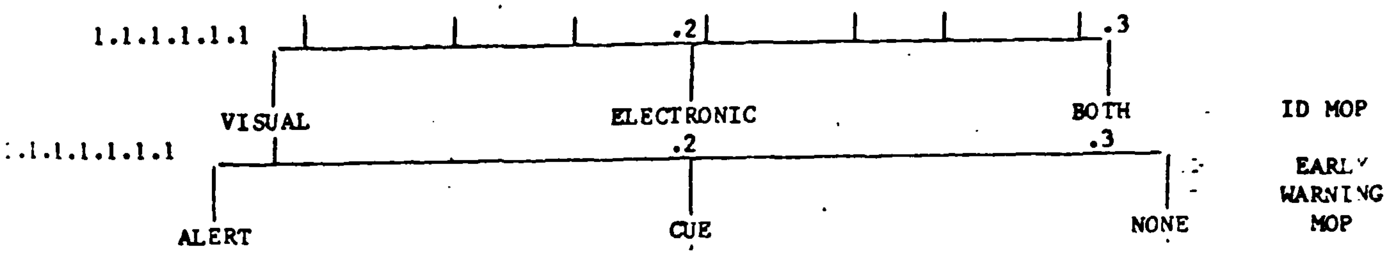


ANALYSIS PLAN: D

Questions:

1. Is visual ID more or less effective than electronic ID for SHORAD systems?
2. Should IFF be allowed for fixed wing engagements but visual ID required for rotary? (within ID Issue investigation)
3. Does electronic ID significantly reduce fratracide for non-radar assisted systems? (cross system investigation)

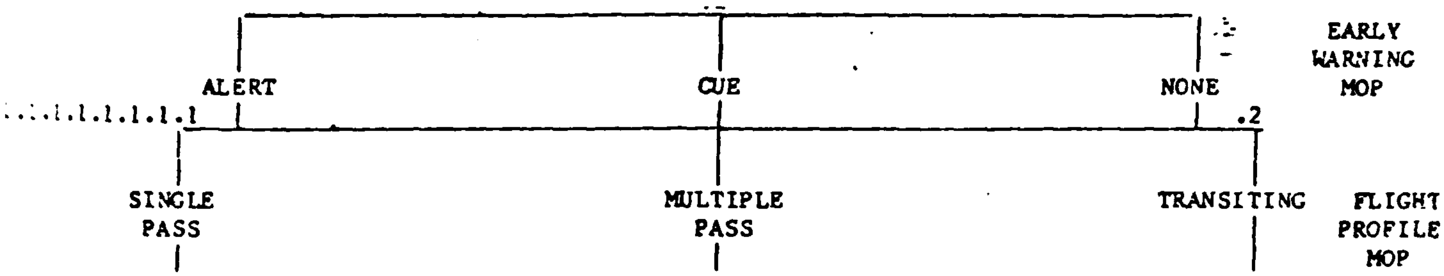
Techniques outlined in II.B are directly applicable



ANALYSIS PLAN: E

1. Does the degree of early warning have any effect upon A/C identification? (interaction)
2. Are there differences among ID techniques? EW?
3. If an indirect ID procedure is used by visually sighted SHORAD systems, does cueing significantly improve the MOEs? (cross system investigation)

Techniques outlined in II.B are directly applicable



ANALYSIS PLAN: F

1. Does the amount of early warning make any difference across different flight profiles? (interaction)
2. What effect does EW have? Is cueing substantially better than alerting information?
3. Do flight profiles have any effect upon the MOEs?

Techniques outlined in II.B are directly applicable

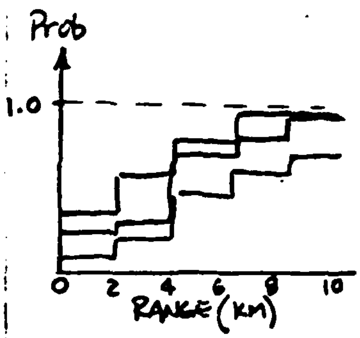
TABLE A-1. SINGLE PASS

RANGE BAND	DETECTED IN/OUTBOUND	CORRECTLY IDENTIFIED IN/OUTBOUND	ENGAGED IN/OUTBOUND	KILLED IN/OUT BOUND
0-1				
1-2				
2-3				
3-4				
4-5				
5-6				
6-7				
TOTAL				

ANALYSIS PLAN: RANGE BAND EXAMINATION

Nonparametric

Mann-Whitney and k-sample Smirnov-type tests:



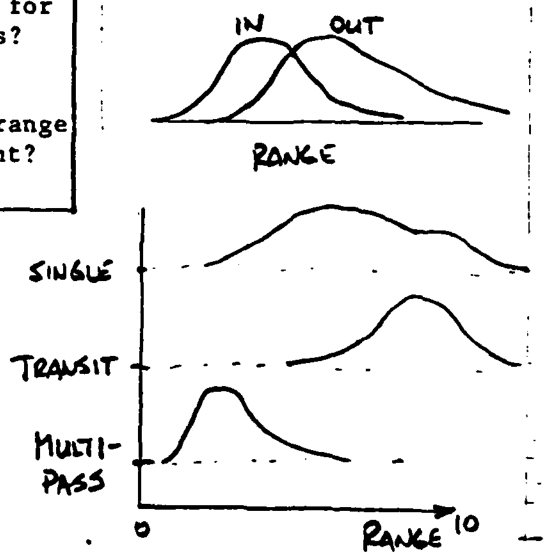
Hypotheses

1. Does A/C direction affect range at which detection, ID, engagement, and kill occur for single pass A/C? multi-pass? transiting?

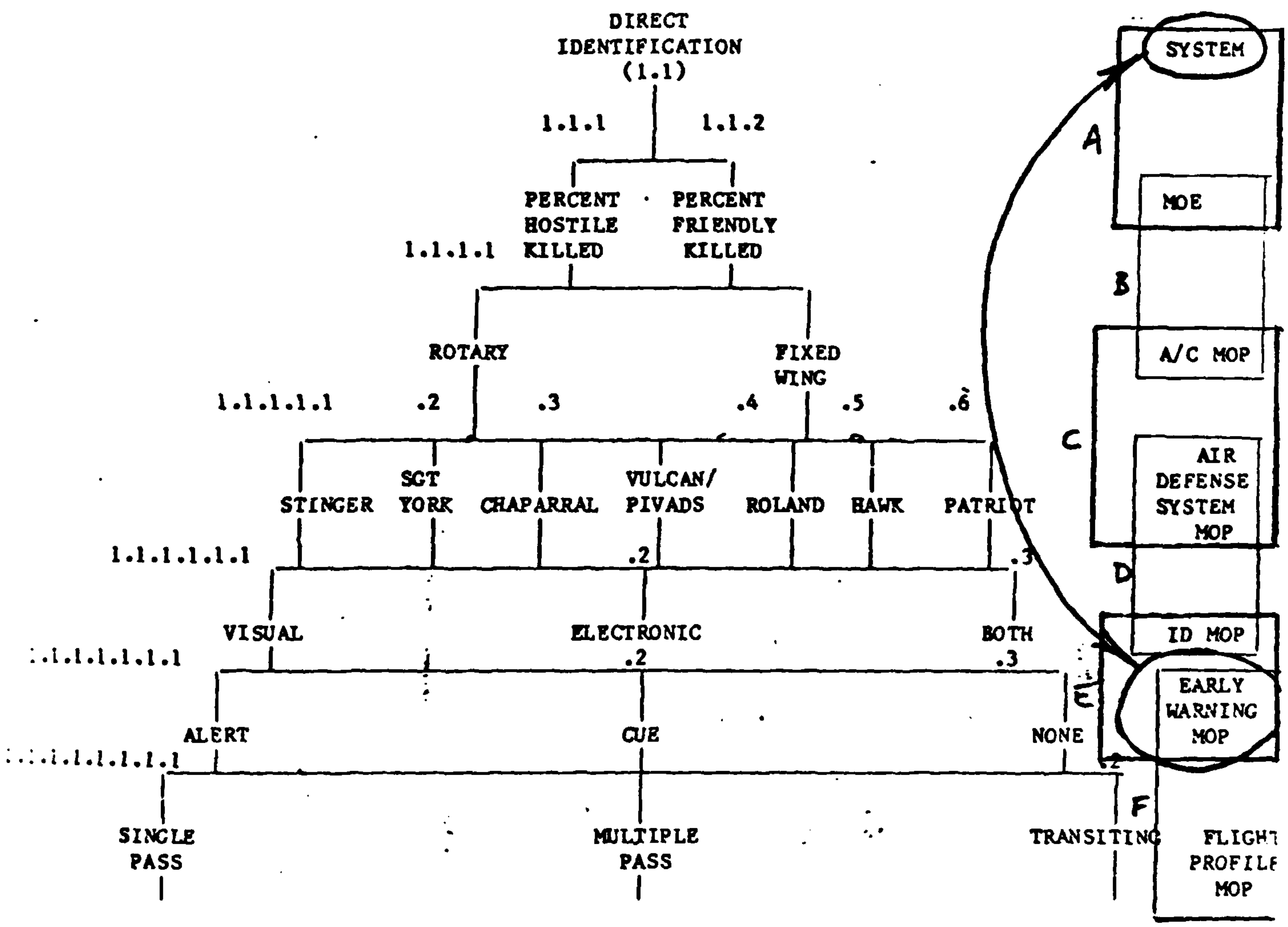
2. Do A/C profiles affect range of detection? ID? engagement? and kill?

Parametric

Standard t-tests and 1-way ANOVA:

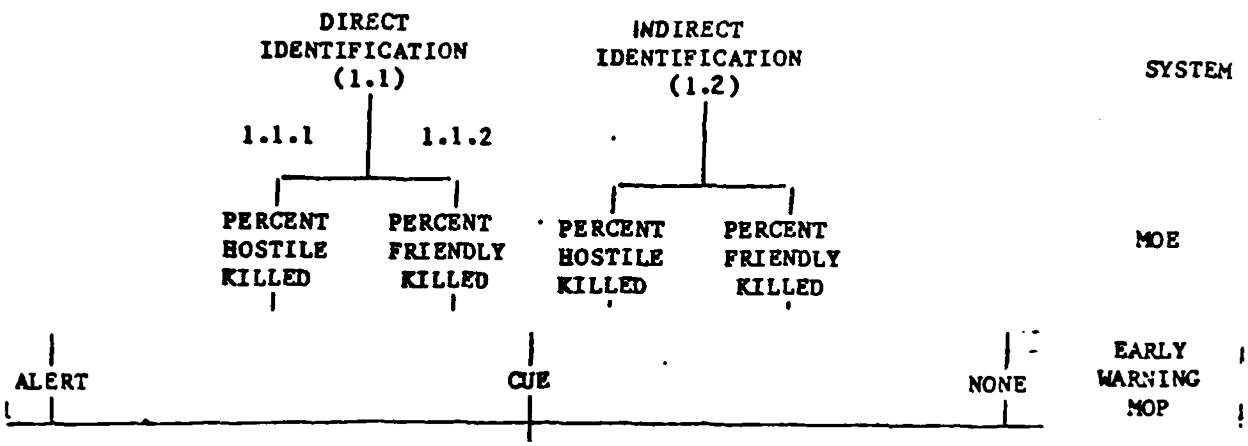


AN EXAMPLE OF CROSS SYSTEM INVESTIGATION OF LOWER LEVEL MOP EFFECTS



Pattern of Analysis for Direct Identification System.

AN EXAMPLE OF CROSS SYSTEM INVESTIGATION OF LOWER LEVEL MOP EFFECTS



Questions:

1. Does EW type have an effect on ID system performance? (interaction significant)
2. Within each ID system, are there differences in fratracide and % hostile killed?
3. Which EW technique is better?

Techniques outlined in II.B are directly applicable

AN EXAMPLE OF CROSS ISSUE INVESTIGATION OF MOEs

DIRECT IDENTIFICATION (1.1)

INDIRECT IDENTIFICATION (1.2)

SYSTEM

PERCENT HOSTILE KILLED

PERCENT FRIENDLY KILLED

MOE

CENTRAL REGION AIRSPACE CONTROL PLAN (3.1)

MINIMUM RISK PASSAGE IN AIR DEFENSE PLAN (3.2)

EXCURSION ASM PLAN (3.3)

SYSTEM

ANALYSIS PLAN

ID Issue

ASM Issue		Direct	Indirect
	Central Region Plan		
	Minimum Risk Plan		All cell entries are MOEs
	Excursion Plan		

Questions:

1. Does the effectiveness of ID procedures vary among different ASM plans? (interaction)
2. Should the ID procedures be standardized across all ASM plans or should specific ID procedures be predicated upon the ASM plan in effect?

Techniques outlined in II.B are directly applicable

III. OBSERVATIONS, FINDINGS AND RECOMMENDATIONS

I have divided my concluding comments into three categories as follows:

1. Observations are offered in the form of major concerns I would have if I were a full-fledged member of the JFAAD Test Force. These concerns evolved as I gradually became more acquainted with JFAAD and, in particular, the analysis plan. As always, context is a crucial consideration and these few observations are based upon the "context" of my experience, to wit:

   a) All of my tactical air defense experience has been devoted to improving effectiveness at and below battery level, especially fire unit procedures. Consequently, I doubt if I fully appreciate the "aggregate" problems that must be contended with at higher command/ staff levels.

   b) My understanding of the actual JFAAD testbed simulation process is incomplete; specifically, how are the distributions for the input variables for each of the required events obtained? (Ref: Chapter 3, JFAAD TPD)

2. Findings are presented which summarize the results of my research effort.

3. Recommendations are offered for consideration.

A. OBSERVATIONS

1. My initial impression has only been reinforced as I've attempted to "get a handle" on the total JFAAD Test Force effort. This is a massive, ambitious undertaking that, at once, due to its magnitude and complexity appears nearly impossible to resolve and yet, for the same reasons, demands such a resolution. If anything, the attempt to seek a resolution is long overdue and, regardless of the eventual outcome, justifies the awesome effort required despite the intimidating and seemingly insurmountable challenge it presents.

2. I am convinced that the greatest danger toward realization of the full potential for JFAAD success lies in a testbed simulation (or "model") that possesses insufficient model fidelity to accurately capture fundamental fire unit level (crew level) activities and events that, when aggregated, could very well yield profound (and not necessarily intuitive) results. For example; 1) the type of search and scan pattern employed by visually-directed FAAD crews has a tremendous effect upon target detection range and time to detection, 2) VACR probabilities are heavily dependent upon both type and number of aircraft that will be operating in the FAAD region, and 3) in many FAAD fire unit situations cueing may actually be counterproductive to the target detection effort. In this regard, I think there are procedural "crew-drill" type activities that will actually have a greater impact upon the MOEs of fratracide and hostile aircraft kills than, for example, system level issues such as EMSCS vs. Objective C^2 or different ASM plans.

B. FINDINGS

1. My initial task was to manually derive "dummy" data, then "roll up" the data through successively higher MOP/MOE levels for each of the three patterns of analysis. This was accomplished, for the most part, without difficulty and shows that the hierarchical patterns of analysis support "bottom-up" data aggregation and, consequently, a "top-down" data analysis approach such as the one detailed in this paper, or a large-scale factorial design as well. I believe that experience allowed me to create fairly reasonable "dummy" data in most cases. The one area where I felt completely incapable of providing semi-realistic data involved communication nodal performance for the C^3I Analysis Plan (pg B-16 of JFAAD TPD). While there is obviously no problem in rolling-up from Table B-1 to Table B-2 (since B-2 is just a summary of B-1), I mention this because my brief research into the commo field while at JFAAD (in an attempt to get realistic data) left me with the impression that there exists a rather large gap in empirical data for this issue (e.g., sensor transmission rates, reliability at the oeprator-machine interface, realistic delay times, etc.). Thus, it would appear difficult to evaluate alternative communication linkages, the effects of sensor netting options, and other excursion options without such data to "drive" the C^3I MOP in such a manner as to provide reasonably accurate results to compare and contrast alternatives.

2. The major portion of my research effort was spent in an attempt to find "the best" statistical approach for data analysis. This entailed an examination of the theoretical assumptions that underly many of the "usual" tests that, at first glance, may appear appropriate. To summarize the major results of my research effort:

a) Tests of proportions must be used with caution since the binomial distribution, requiring independent trials and random sampling with replacement, will rarely serve as an appropriate probability model.

b) The occurrence of statistical error reveals that, in many cases, Type II error is more important to control than Type I. Hence the selection of a high power level (1-β) may be more important than a small signficiance level (α). This has special signficance for small sample sizes.

c) Large scale factorial designs are excellent for determining interactions and measuring the magnitude of interactions, however they preclude an examination of main effects.

d) Nonparametric tests are available that exhibit greater ARE than comparable parametric ANOVA.

e) An analysis of the advantages and disadvantages of both parametric and nonparametric statistical tests encourages the use of a dual-approach methodology, allowing the choice of a specific test or tests to ultimately be suggested by the data.

f) Such a statistical methodology allows a flexible "top-down" approach to investigate each major issue. Furthermore it also allows cross-investigation of lower-level MOPs within different systems of a particular issue and also allows cross-investigation of MOEs among different tests conditions (e.g., scenario location, air environment, EW environment, and visibility conditions).

C. RECOMMENDATIONS

1. Software. Presumably, data generated from the testbed simulation will be "tagged" and stored in a database for subsequent retrieval. Hence there is a need for an effective data retrieval system that allows efficient recovery of the specific data to be analyzed. A data retrieval system capable of directly interfacing with a statistical graphics package would appear to offer a flexible statistical analysis package. The statistical routines needed would obviously depend upon the type of analysis ultimately used. SPSS routines could be used for a large-scale factorial experimental design, or OA 3660 (even MINITAB) could be used to automate the methodology I've proposed. The purpose of the graphics package would be to examine the data in an effort to ascertain distributional form thus aiding the selection of appropriate statistical tests, examining the effects of transformations, and presentation of results.

2. Statistical Analysis. The "dual-approach" methodology detailed in this paper is recommended as it appears to offer a flexible, yet powerful and theoretically sound approach, capable of providing detailed answers pertinent to all facets of the JFAAD TPD. This approach enables advantages offered by both parametric and nonparametric statistics to be capitalized upon, thus enhancing the credibility of test results.

3. Additional Research. Toward the end of my research effort, I stumbled across yet another approach that appears to offer a promising alternative. The technique is referred to as analysis of "discrete multivariate data", "multivariate binary data", and "cross-classified categorical data" in different references. While the theory is relatively advanced and I was not able to pursue the subject extensively, it involves the analysis of multi-dimensional contingency tables (cell entries are count data) using a hierarchical loglinear model. The resultant model, developed using an iterative proportional fitting procedure to compute maximum likelihhod estimates (MLEs) for each cell value, appears to be analogous to factorial ANOVA

in both functional form and interpretation. Although the development of the model appears to be quite a lengthy and tedious process, it fortunately has been computerized. The Operational Test and Evaluation Agency (OTEA) has and uses the computer algorithms for large-scale loglinear models. I strongly encourage further investigation into this technique, especially if the software can be transported to JFAAD from OTEA or if the JFAAD data can be transported to OTEA for analysis there.

4. References. I found the following texts to be particularly good in their respective fields and recommend them as valuable references:

Nonparametric and Distribution - Free Methods for the Social Sciences, by L. A. Marascuilo and M. McSweeney.

The Analysis of Cross-Classified Categorial Data, by S. E. Fienberg

Discrete Multivariate Analysis: Theory and Practice, by Y. M. M. Bishop, et. al.

Interactive Data Analysis, by D. R. McNeil

ANNEX A

DATA TRANSFORMATION

The special problems that surface when attempting to analyze proportions, or percentages, derived from a binary response variable can be remedied through various data transformations. The most commonly used are:

1. logistic, $\log\left(\frac{P}{1-P}\right)$
2. linear, P
3. integrated normal (probit)
4. arcsin, $2 \arcsin \sqrt{P}$

Cox (in The Analysis of Binary Data) concludes that all four are in reasonable agreement as long as $0.1 < p < 0.9$. The primary advantage in using the arcsin, or angular, transformation is its ability to remove variance dependence upon the mean in a binomial-type variable. The transformation:

$$\phi = 2 \arcsin \sqrt{Y}\,; \quad 0 < Y < 1$$

yields the following tabulated results (Y is used to denote the proportion P):

$\phi = 2 \arcsin \sqrt{Y}$

Y	φ	Y	φ	Y	φ	Y	φ	Y	φ
.001	.0633	.041	.4078	.36	1.2870	.76	2.1177	.971	2.7993
.002	.0895	.042	.4128	.37	1.3078	.77	2.1412	.972	2.8053
.003	.1096	.043	.4178	.38	1.3284	.78	2.1652	.973	2.8115
.004	.1266	.044	.4227	.39	1.3490	.79	2.1895	.974	2.8177
.005	.1415	.045	.4275	.40	1.3694	.80	2.2143	.975	2.8240
.006	.1551	.046	.4323	.41	1.3898	.81	2.2395	.976	2.8305
.007	.1675	.047	.4371	.42	1.4101	.82	2.2653	.977	2.8371
.008	.1791	.048	.4418	.43	1.4303	.83	2.2916	.978	2.8438
.009	.1900	.049	.4464	.44	1.4505	.84	2.3186	.979	2.8507
.010	.2003	.050	.4510	.45	1.4706	.85	2.3462	.980	2.8578
.011	.2101	.06	.4949	.46	1.4907	.86	2.3746	.981	2.8650
.012	.2195	.07	.5355	.47	1.5108	.87	2.4039	.982	2.8725
.013	.2285	.08	.5735	.48	1.5308	.88	2.4341	.983	2.8801
.014	.2372	.09	.6094	.49	1.5508	.89	2.4655	.984	2.8879
.015	.2456	.10	.6435	.50	1.5708	.90	2.4981	.985	2.8960
.016	.2537	.11	.6761	.51	1.5908	.91	2.5322	.986	2.9044
.017	.2615	.12	.7075	.52	1.6108	.92	2.5681	.987	2.9131
.018	.2691	.13	.7377	.53	1.6308	.93	2.6062	.988	2.9221
.019	.2766	.14	.7670	.54	1.6509	.94	2.6467	.989	2.9315
.020	.2838	.15	.7954	.55	1.6710	.95	2.6906	.990	2.9413
.021	.2909	.16	.8230	.56	1.6911	.951	2.6952	.991	2.9516
.022	.2978	.17	.8500	.57	1.7113	.952	2.6998	.992	2.9625
.023	.3045	.18	.8763	.58	1.7315	.953	2.7045	.993	2.9741
.024	.3111	.19	.9021	.59	1.7518	.954	2.7093	.994	2.9865
.025	.3176	.20	.9273	.60	1.7722	.955	2.7141	.995	3.0001
.026	.3239	.21	.9521	.61	1.7926	.956	2.7189	.996	3.0150
.027	.3301	.22	.9764	.62	1.8132	.957	2.7238	.997	3.0320
.028	.3363	.23	1.0004	.63	1.8338	.958	2.7288	.998	3.0521
.029	.3423	.24	1.0239	.64	1.8546	.959	2.7338	.999	3.0783
.030	.3482	.25	1.0472	.65	1.8755	.960	2.7389		
.031	.3540	.26	1.0701	.66	1.8965	.961	2.7440		
.032	.3597	.27	1.0928	.67	1.9177	.962	2.7492		
.033	.3654	.28	1.1152	.68	1.9391	.963	2.7545		
.034	.3709	.29	1.1374	.69	1.9606	.964	2.7598		
.035	.3764	.30	1.1593	.70	1.9823	.965	2.7652		
.036	.3818	.31	1.1810	.71	2.0042	.966	2.7707		
.037	.3871	.32	1.2025	.72	2.0264	.967	2.7762		
.038	.3924	.33	1.2239	.73	2.0488	.968	2.7819		
.039	.3976	.34	1.2451	.74	2.0715	.969	2.7876		
.040	.4027	.35	1.2661	.75	2.0944	.970	2.7934		

(Ref. 8, pg. 831)

ANNEX B

MULTIPLE COMPARISON (POST-HOC) TESTS

A. PARAMETRIC MULTIPLE COMPARISON TESTS:

In those instances when a null hypothesis for a treatment is rejected (i.e., differences among levels of a treatment do exist), it does not necessarily follow that each level is different from the others. Hence, once a null hypothesis is rejected, it becomes extremely important in the JFAAD context to determine which levels are different and which are not (e.g., differences in the EW MOP between Alerting, Cueing, and no EW). Various multiple comparison tests have been developed, including (see Reference 4, pp 262-271 and Reference 10, pp. 233-238):

1. Fisher's Least Significant Difference
2. Duncan's New Multiple Range Test
3. The Student - Newman - Keuls' Procedure
4. Tukey's Honestly Significant Difference
5. Scheffe's Method

The choice of a particular test is dependent upon which type of error (Type I or Type II) is more serious. In most cases it is desireable to reduce the Type II error (i.e., if differences do exist among levels of a particular MOP, it is important that such differences actually be detected by the test; a test is needed with high power). Thus, in the JFAAD context, Fisher's Test will be appropriate to test for differences among MOP levels when a MOP factor has been declared significant by ANOVA procedures.

Comparison of Multiple Comparison Procedures

Multiple Comparison Procedure	Power	Type I Error Rate
Fisher's	Highest	Highest
Duncan's	↓ More conservative, less likely to detect real differences	↑ More likely to indicate false differences
Student-Newman-Keuls'		
Tukey's		
Scheffé's	Lowest	Lowest

B. NONPARAMETRIC MULTIPLE COMPARISON TESTS.

Multiple comparison tests are also available for the nonparametric rank and normal-scores tests. These so called "post-hoc" tests, can be found on the below listed pages of the following references:

Test	Conover (Ref. 2)	Marascuilo and McSweeney (Ref. 9)
Kruskal-Wallace	231 .	306-310
Quade	297	-
Friedman	300	362-366
Vander Waerden	319	405-414

ANNEX C

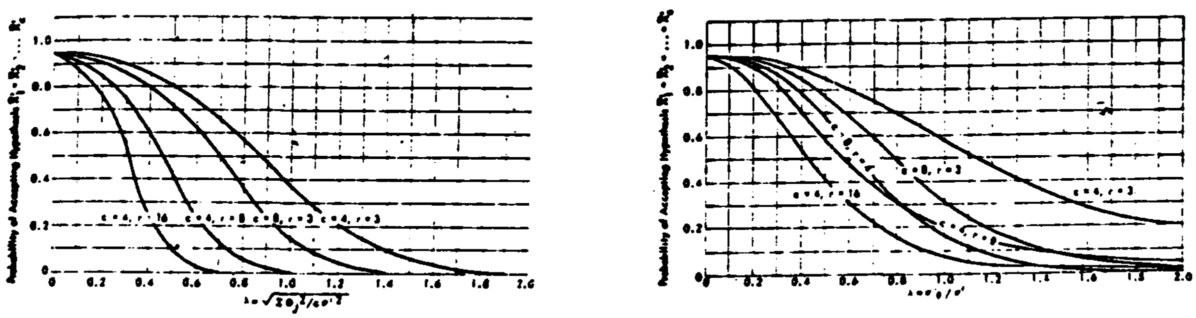
POWER CALCULATIONS IN ANOVA

As discussed earlier, improving power (1-β) in ANOVA can be accomplished by either:

1. Increasing α (probability of Type I error) which can be accomplished by pre-selecting larger α values, for example α = .1 instead of .05 or .01, or

2. Increasing the sample size for a constant α level.

Knowledge of power is useful not only for assessing the discriminating ability of the test, but also for determining the necessary sample size to achieve a desired level of power. Detailed power calculations are provided in Reference 5, pp. 615-619, 660 and Reference 8, pp. 142-145. Power function curves (Operating Characteristic - "OC Curves") have been developed and can be used to assist determination of sample size (see, for example, Figure 29.1, Reference 5, p. 617).



However, a simple procedure, requiring no a priori information of sample variance, is available simply by specifying α, the desired power (1-β), p (the number of levels (columns) in a 1-Way ANOVA), and C (which is a multiple of the unknown residual error, σ_e) and using Table E.15 of Reference 8 (pp. 840-841), reproduced on the next page.

This technique should prove very valuable in estimation of minimum sample sizes necessary to achieve the desired level of power. It illustrates the tradeoffs between α and β, as well as the differential improvement that can be gained in power by increasing the sample size for a given test.

TABLE E.15 Minimum Sample Size Needed to Ensure a Given Power

p	α	1 − β = .70							1 − β = .80						
		C							C						
		1.00	1.25	1.50	1.75	2.00	2.50	3.00	1.00	1.25	1.50	1.75	2.00	2.50	3.00
2	.10	11	7	6	4	4	3	3	14	9	7	5	4	3	3
	.05	14	9	7	6	5	4	3	17	12	9	7	6	4	4
	.01	21	15	11	9	7	5	5	26	17	13	10	8	6	5
3	.10	13	9	7	5	4	3	3	17	11	8	6	5	4	3
	.05	17	11	8	7	5	4	3	21	14	10	8	6	5	4
	.01	25	17	12	10	8	6	5	30	20	14	11	9	7	5
4	.10	15	10	7	6	5	4	3	19	13	9	7	6	4	3
	.05	19	13	9	7	6	4	4	23	15	11	9	7	5	4
	.01	28	19	13	10	8	6	5	33	22	16	12	10	7	5
5	.10	17	11	8	6	5	4	3	21	14	10	8	6	4	4
	.05	21	14	10	8	6	5	4	25	17	12	9	7	5	4
	.01	30	20	14	11	9	6	5	35	23	17	13	10	7	6
6	.10	18	12	9	7	5	4	3	22	15	11	8	7	5	4
	.05	22	15	11	8	7	5	4	27	18	13	10	8	6	4
	.01	32	21	15	12	9	7	5	38	25	18	13	11	8	6
7	.10	19	13	9	7	6	4	3	24	16	11	9	7	5	4
	.05	24	16	11	9	7	5	4	29	19	14	10	8	6	5
	.01	34	22	16	12	10	7	5	39	26	18	14	11	8	6
8	.10	20	13	10	7	6	4	3	25	16	12	9	7	5	4
	.05	25	16	12	9	7	5	4	30	20	14	11	9	6	5
	.01	35	23	17	13	10	7	5	41	27	19	15	12	8	6
9	.10	21	14	10	8	6	4	4	26	17	12	9	7	5	4
	.05	26	17	12	9	8	5	4	31	21	15	11	9	6	5
	.01	37	24	17	13	10	7	6	43	28	20	15	12	8	6
10	.10	22	14	10	8	6	5	4	27	18	13	10	8	5	4
	.05	27	18	13	10	8	6	4	33	21	15	12	9	6	5
	.01	38	25	18	14	11	7	6	44	29	21	16	12	8	6
11	.10	23	15	11	8	7	5	4	28	18	13	10	8	6	4
	.05	28	19	13	10	8	6	4	34	22	16	12	9	7	5
	.01	39	26	18	14	11	8	6	46	30	21	16	13	9	7
13	.10	24	16	11	9	7	5	4	30	20	14	11	8	6	4
	.05	30	20	14	11	9	6	5	36	24	17	13	10	7	5
	.01	42	27	19	15	12	8	6	49	32	22	17	13	9	7

This table is abridged from Tables 1, 2, 3, and 4 in Tables of sample sizes in the analysis of variance. *Journal of Quality Technology*, 1970, 2, 156–164 with permission of the first author, T. L. Bratcher, and the editor.

TABLE E.15 (continued)

p	α	1 − β = .90							1 − β = .95						
		C							C						
		1.00	1.25	1.50	1.75	2.00	2.50	3.00	1.00	1.25	1.50	1.75	2.00	2.50	3.00
2	.10	18	12	9	7	6	4	3	23	15	11	8	7	5	4
	.05	23	15	11	8	7	5	4	27	18	13	10	8	6	5
	.01	32	21	15	12	10	7	6	38	25	18	14	11	8	6
3	.10	22	15	11	8	7	5	4	27	18	13	10	8	6	4
	.05	27	18	13	10	8	6	5	32	21	15	12	9	7	5
	.01	37	24	18	13	11	8	6	43	29	20	16	12	9	7
4	.10	25	16	12	9	7	5	4	30	20	14	11	9	6	5
	.05	30	20	14	11	9	6	5	36	23	17	13	10	7	5
	.01	40	27	19	15	12	8	6	47	31	22	17	13	9	7
5	.10	27	18	13	10	8	5	4	33	22	15	12	9	6	5
	.05	32	21	15	12	9	6	5	39	25	18	14	11	7	6
	.01	43	28	20	15	12	9	7	51	33	23	18	14	10	7
6	.10	29	19	14	10	8	6	4	35	23	16	12	10	7	5
	.05	34	23	16	12	10	7	5	41	27	19	14	11	8	6
	.01	46	30	21	16	13	9	7	53	35	25	19	15	10	8
7	.10	31	20	14	11	9	6	5	37	24	17	13	10	7	5
	.05	36	24	17	13	10	7	5	43	28	20	15	12	8	6
	.01	48	31	22	17	13	9	7	56	36	26	19	15	10	8
8	.10	32	21	15	11	9	6	5	39	25	18	14	11	7	5
	.05	38	25	18	13	11	7	6	45	29	21	16	12	8	6
	.01	50	33	23	17	14	9	7	58	38	27	20	16	11	8
9	.10	33	22	16	12	9	6	5	40	26	19	14	11	8	6
	.05	40	26	18	14	11	8	6	47	30	22	16	13	9	6
	.01	52	34	24	18	14	10	7	60	39	28	21	16	11	8
10	.10	35	23	16	12	10	7	5	42	27	19	15	11	8	6
	.05	41	27	19	14	11	8	6	48	31	22	17	13	9	7
	.01	54	35	25	19	15	10	7	62	40	29	21	17	11	8
11	.10	36	23	17	13	10	7	5	43	28	20	15	12	8	6
	.05	42	28	20	15	12	8	6	50	33	23	17	14	9	7
	.01	55	36	26	19	15	10	8	64	42	29	22	17	12	9
13	.10	38	25	18	13	11	7	5	46	30	21	16	12	8	6
	.05	45	29	21	16	12	8	6	53	34	24	18	14	10	7
	.01	59	38	27	20	16	11	8	68	44	31	23	18	12	9

REFERENCE LIST

A. TEXTS

1. Box, G. P., Hunter, W. C., and Hunter, J. S., Statistics for Experimenters, Wiley, 1978.

2. Conover, W. J., Practical Nonparametric Statistics, Wiley, 1980.

3. Cox, D. R., Analysis of Binary Data, Chapman and Hall, 1970.

4. Dowdy, Shirley, and Wearden, Stanley, Statistics for Research, Wiley, 1983.

5. Duncan, A. J., Quality Control and Industrial Statistics, Irwin, 1974.

6. Fienberg, Stephen E., The Analysis of Cross-Classified Categorial Data, MIT Press, 1977.

7. Joint Forward Area Air Defense Test Program Definition OUSDRE/JFAAD, Analysis Directorate, 30 March 1984.

8. Kirk, Roger E., Experimental Design: Procedures for the Behavioral Sciences, Brooks-Cole, 1982.

9. Marascuilo, L. A., and McSweeney, M., Nonparametric and Distribution-Free Methods for the Social Sciences, Brooks-Cole, 1977.

10. Namboodiri, N. K., Carter, L. F., and Blalock, H. M., Applied Multivariate Analysis and Experimental Designs, McGraw-Hill, 1975.

11. Natrella, M. G., Experimental Statistics, reprint from National Bureau of Standards, Wiley, 1966.

12. Sachs, Lothan, Applied Statistics: A Handbook of Techniques, Springer-Verlag, 1982.

B. ARTICLES

1. Cox, D. R., "The Analysis of Multivariate Binary Data", Applied Statistics, Vol. 21, 1972.

2. Fisher, R. A., "The Analysis of Variance with Various Binomial Transformations (With Discussion)", Biometrics, 10, 1954.

3. Gabriel, K. R., "Analysis of Variance of Proportions with Unequal Frequencies", Journal of the American Statistical Association, 58, 1963.

4. Lindsay, G. F., Two-Way ANOVA With More Than One Item Per Cell, unpublished lecture notes from the course "Test and Evaluation", Naval Postgraduate School, Monterey, CA, 1982.

END

FILMED

11-84

DTIC