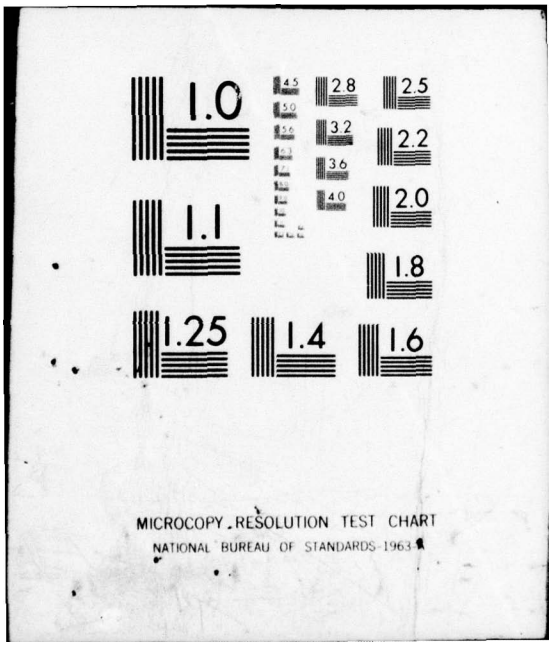
AD-A039 238 NAVAL POSTGRADUATE SCHOOL MONTEREY CALIF F/G 19/5
STATISTICAL ASPECTS OF THE AN/TPQ-27 PSVT DESIGN.(U)
MAR 77 D R BARR
UNCLASSIFIED NPS55-77-13 NL

| OF |
AD
A039238

END

DATE
FILMED
5-77


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963

NPS55-77-13

NAVAL POSTGRADUATE SCHOOL

Monterey, California

STATISTICAL ASPECTS OF THE AN/TPQ-27 PSVT DESIGN

Donald R. Barr

March 1977

Approved for public release; distribution unlimited

Prepared for:

NAVELEX 54031
Naval Electronics Systems Engineering Center
Vallejo, California

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral Isham Linder
Superintendent

Jack R. Borsting
Provost

Reproduction of all or part of this report is authorized.

This report was prepared by:

Donald R. Barr
Professor
Department of Operations Research

Reviewed by:

Michael G. Sovereign, Chairman
Department of Operations Research

Robert R. Fossum
Dean of Research

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: NPS55-77-13
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Statistical Aspects of the AN/TPQ-27 PSVT Design.
5. TYPE OF REPORT & PERIOD COVERED: Technical rept.
6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Donald R. Barr
8. CONTRACT OR GRANT NUMBER(s): N63274-76-PO 0061
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Postgraduate School, Monterey, California 93940
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 33 P.
11. CONTROLLING OFFICE NAME AND ADDRESS: NAVELEX 54031, ATTN: W.F. Trisler, Code 041, AV 222-8924, Naval Electronics Systems Engineering Center, Vallejo, California
12. REPORT DATE: Mar[redacted]77
13. NUMBER OF PAGES: 28
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)
18. SUPPLEMENTARY NOTES
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

AN/TPQ-27
Test & Evaluation
CEP Test

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

The AN/TPQ-27 currently under development for the Marine Corps by NAVELEX is scheduled to undergo a test series known as the PSVT. A primary goal of the PSVT is to determine whether the system is operating with acceptable accuracy. Several sequential and fixed sample size tests are discussed which could be used for this purpose. Several aspects of the statistical design of the experiment are discussed, and recommendations regarding secondary analysis are made. A method of reducing sample size requirements for the nonparametric accuracy tests is presented.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

251 450

STATISTICAL ASPECTS OF THE AN/TPQ-27

PSVT DESIGN

DONALD R. BARR

Naval Postgraduate School
Monterey, California 93940

March 1977

STATISTICAL ASPECTS OF THE AN/TPQ-27 PSVT DESIGN

Donald R. Barr
Naval Postgraduate School
Monterey, California 93940

ABSTRACT

The AN/TPQ-27 currently under development for the Marine Corps by NAVELEX is scheduled to undergo a test series known as the PSVT. A primary goal of the PSVT is to determine whether the system is operating with acceptable accuracy. Several sequential and fixed sample size tests are discussed which could be used for this purpose. Several aspects of the statistical design of the experiment are discussed, and recommendations regarding secondary analysis are made. A method of reducing sample size requirements for the nonparametric accuracy tests is presented.

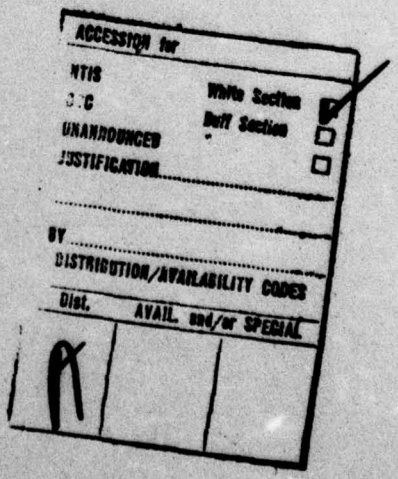


TABLE OF CONTENTS

1. INTRODUCTION

During the period 14 February 1977-16 February 1977, meetings were held at the Naval Postgraduate School (NPS) to discuss the TPQ-27 PSVT. Participating in this meeting were Major Earl Peete (MAD, Pt. Mugu), Major Dave Allen (MCTSSA, Camp Pendleton), Capt. Jerry Paccassi and myself (both at NPS). Also in attendance were Mike Patrow and Mike Lowe, students at NPS. A test concept was developed which called for bomb drops with 18 cells in a "base line" group, together with additional "demonstration" drops, conducted under eight additional combinations of conditions. These combinations are shown in Figure 1. Within each cell of the design for baseline drops, a test is to be made of whether contract specified CEP's have been met.

In what follows, we discuss the design, certain aspects of performing the trials in the field, and an outline of the Analysis procedure proposed for testing CEP's and making other inferences from the test data. Some of these comments came out of discussions at the NPS meeting, and others are suggestions and observations by the author.

2. THE STATISTICAL DESIGN

It is desirable to test the TPQ-27 over a wide range of levels of the variables involved, in order to facilitate inference about performance characteristics of this system and its sensitivity and response to variations in drop conditions.

Baseline drops

Altitude (k-feet)	Mode	Ranges (mi) 5	20	55	106
2	Auto	X			
10	Auto	X	X,0	X	
10	Voice Vector		X,0		
10	Skin Track	X	X		
20	Auto	X	X,0		
20	Voice Vector	X	X,0		
30	Auto			X	X
40	Auto				X

NOTE: "X" Denotes drops at 500 kts; "0" denotes 350 kts.

Demonstration drops

FGC/STICK:	2 sticks at 20 mi, 10k ft, 500 kts
SEMIAUTO:	5 mi, 20k, 500 kts
	20 mi, 20k, 500 kts
	55 mi, 20k, 500 kts
MANUAL:	20 mi, 20k, 500 kts
WIND AT ALT:	20 mi, 20k, 500 kts
	55 mi, 20k, 500 kts
RDL:	20 mi, 20k, 500 kts

FIGURE 1. Combination of conditions under which drops are planned in the PSVT.

However, this testing, involving dropping bombs on an instrumented range, is expensive. Thus there is also a conflicting desire to hold the sample size as low as possible, consistent with achieving reasonable confidence in the tests and in the inferences to be made. For this reason, a sample of baseline conditions was established, in which most of the drops are to be made. The baseline cases were selected so as to cover a fairly large portion of operationally realistic conditions. The data resulting from these baseline drops will allow testing against contract specified CEP's in each cell, as well as subsequent analyses such as testing whether there are significant differences due to the factors range, altitude, range × altitude interaction, speed, mode, speed × mode interaction and speed × altitude interaction. In addition, estimates of the type and amount of response to changes in the main effects (for Auto mode) can be made. For the demonstration cells, tests against contract specified CEP's can also be made.

The nature of the tests of CEP has not been completely determined at this time, but appears to have been narrowed down to several candidates. Sequential testing within each cell of the design appears attractive because of the expected savings in numbers of bomb drops. In Section 4 below we outline two possible sequential procedures (called "sequential Rayleigh" and "sequential nonparametric") as well as two fixed sample size procedures (called "fixed Rayleigh" and "fixed nonparametric"). Sample size characteristics of the sequential and fixed-sample size tests are shown in Tables 1 and 2.

α	β	CEP_1/CEP_0	Min Accept	Min Reject	Max E(N)	Typical N	N Fixed	Slope	Accept Intercept	Max $3\sigma_N$
0.1	0.1	1.41	5	7	26	19.3	35	.393	$\pm$2.49	78
0.2	0.2	1.41	3	4	11	9.1	16	.393	$\pm$1.57	32
0.1	0.1	1.1	17	19	293	213.9	1824	.468	$\pm$8.55	879
0.2	0.2	1.1	11	12	117	101.2	787	.468	$\pm$5.39	350
0.01	0.1	1.41	5	12.984	58	32.9	70	.393	2.601	173
0.1	0.1	2	2	5	9	6.3	12	.312	$\pm$1.32	27
0.2	0.2	2	2	3	4	3.0	6	.312	$\pm$ .833	12
0.01	0.1	2	3	8.656	18	10.5	24	.312	1.377	53
0.1	0.2	2	2	4.000	6	4.4	9	.312	.903	16
0.05	0.1	2	2	5.560	11	7.7	16	.312	1.352	32
0.05	0.15	2	2	5.450	9	6.6	13	.312	1.109	27
0.05	0.20	2	2	5.333	8	5.7	11	.312	0.936	22
0.05	0.05	2	3	6	15	9.4	20	.312	$\pm$1.769	44

TABLE 1. Sequential and Fixed Non-Parametric Characteristics for Several α, β, c_1 Combinations.

α	β	CEP_1/CEP_0	Min Accept	Low Reject	Max E(N)	Typical N	N Fixed	Slope	Acc Int	Rej Int	max $3\sigma_N$
0.1	0.1	2	2	3	2	1.9	5	2.667	-4.227	4.227	9
0.2	0.2	2	1	2	2	0.0	2	2.667	-2.667	2.667	4
0.01	0.1	2	2	7	6	2.9	7	2.667	-4.410	8.656	16
0.1	0.2	2	2	3	2	1.3	3	2.667	-2.893	4.000	6
0.05	0.1	2	2	5	4	2.3	5	2.667	-4.331	5.560	11
0.05	0.15	2	2	5	3	1.9	4	2.667	-3.551	5.450	9
0.05	0.20	2	2	4	3	1.6	4	2.667	-2.997	5.333	7
0.05	0.05	2	3	5	5	2.9	7	2.667	-5.664	5.664	15
0.1	0.1	1.41	4	-	11	7.4	15	2.000	-6.340	6.340	31
0.2	0.2	1.41	2	-	5	3.5	14	2.000	-4.000	4.000	14
0.01	0.1	1.41	4	-	22	12.0	60	2.000	-6.614	12.984	65
0.1	0.1	1.1	12	-	134	98.0	500	1.585	-18.265	18.265	400
0.2	0.2	1.1	8	-	54	45.8	70	1.585	-11.524	11.524	160

TABLE 2. Sequential and Fixed Rayleigh Characteristics for Several α, β, c_1 Combinations.

The column heads in Tables 1 and 2 are as follows:

α	:	probability the test rejects H_0:CEP = C_0 in favor of H_1:CEP = C_1, when in fact the system has CEP = C_0.
β	:	probability the test accepts H_0 when in fact the system has CEP = C_1.
CEP_1/CEP_0	:	ratio of minimum unacceptable CEP to contract specified CEP.
min accept	:	the smallest possible sample size at termination with acceptance of H_0 (i.e., the sample size required to accept even a perfect system).
min reject	:	the smallest sample size possible for rejecting H_0 (for nonparametric sequential procedure only--for the sequential Rayleigh procedure, the min reject number is 1 for all cases). NOTE: for the nonparametric case, round up to integer values where necessary.
Low reject	:	the sample size required for rejection if all radial misses fell at distance CEP_1 from the target (for sequential Rayleigh only).

Max E(N) : the worst case expected sample size for the sequential procedures (this occurs for some true system CEP between CEP_0 and CEP_1).

Typical N : average of sequential tests expected sample sizes under H_0 and under H_1.

N fixed : sample size required by the fixed-sample size procedures.

slope : slope of lines forming boundaries of the continuation region for sequential procedures.

Accept intercept : the y-intercept of the boundary line defining the accept region for sequential procedures.

Reject intercept: the y-intercept of the boundary line defining the reject region for sequential procedures. **NOTE:** for the sequential nonparametric procedures, the y-intercepts are symmetric if $\alpha = \beta$; otherwise the x-intercept of the rejection line is given under "Min reject."

Max $3\sigma_N$: three times the max E(N). This is roughly two standard deviations above the expected sample size--virtually none of the tests should continue beyond this value.

The values shown in Tables 1 and 2 pertaining to sequential tests were obtained using Walds' approximations, and are therefore slightly conservative. Exact stopping bounds are available for these tests (for example those prepared by Leo A. Aroian at TRW Systems, Redondo Beach, California), and they should be used if the sequential approach to testing CEP is adopted. Truncation of the sequential test was considered, but it appears undesirable for several reasons: 1) truncation increases average sample sizes, 2) truncation complicates the computation of acceptance and rejection bounds (although, again, tables may be available covering most of our cases), and 3) the terminal decision for cases reaching the truncation point is somewhat arbitrary. In addition, for the α, β, CEP_1/CEP_0 combinations, we can realistically anticipate (see Tables 1 and 2 with α and β on the order of 0.10 and CEP_1/CEP_0 about 2, for example), max $3\sigma_N$ (which is essentially an upper bound on sample size N) is not unacceptably large, in view of the fact that over the many cells of the design, with an individual sequential test being performed in each cell, the overall average sample size per cell will almost certainly fall below max E(N). Consequently, it is felt that trunction would only cause unnecessary increase in overall drop requirements for the entire test sequence.

An alternative to untruncated sequential testing is to use fixed sample size tests. This has the effect of

balancing the number of drops in the various cells of the design matrix, which is desirable for the subsequent analyses concerning differences due to the various factors. However, as with truncation of the sequential procedures, the overall sample size requirements are larger for fixed sample size tests. It is our feeling that the balance in design achieved by fixed sample size testing is far outweighed by its disadvantage with respect to overall sample size requirement. As is discussed in the succeeding section, the way in which the field tests may be carried out will tend to balance the design even with sequential testing in each cell, and this further points to superiority of using sequential testing.

3. OUTLINE OF FIELD TEST PROCEDURES

In order to avoid loosing efficiency in the PSVT, it is desirable that drops be conducted in such a way as to avoid (as much as possible) confounding factors suspected to affect system performances, and to provide "insurance" against bias in results due to unknown causes. Ideally this would be in part accomplished by scheduling individual drops over the various cells of the design using a formal randomization procedure. This might mean, for example, that a single flight (operation) would call for first dropping a bomb at 300 kts, 20k ft altitude at 20 mi range in Auto mode, next dropping a bomb at 500 kts, 10k altitude at 55 mi range in Auto mode,

and so on for the remaining bombs to be dropped in this operation. Clearly such a schedule may not be practical, so constraints must be imposed on the scheduling process. The author is not in a position to assess what constraints are necessary, but he wishes to point out the desirability of imposing as little constraint as possible.

In order to gain appreciation of the possible effects of confounding mentioned above, consider an example test schedule in which the first group of operations are all conducted at 500 kts, 10k altitude, 20 mi range, Auto mode. These drops might be followed by operations all at 500 kts, 20k, 20 mi, auto, etc. Suppose, moreover, each individual operation (consisting of eight bombs) is constrained such that all eight bombs are dropped under the same conditions (in the same cell of the design). Then factors having to do with each individual operation (such as radar alignment, pilot effect, wind profile errors, etc.), whose effects for the given operation may be unknown or only partially known (even using ARIS), cannot be "balanced out"; rather they may cause bias of an amount undeterminable by the experimenter and analyst. Simiarly, conducting operations all with fixed combinations of speed, range, etc. close together in time would preclude balancing out unknown long term trend effects (if any).

There is another reason why allowing drops in different cells in a single operation would be desirable. If a sequential test plan is adopted for CEP testing, forcing observations to

be made in batches of eight (say) in a given cell of the design rather than one at a time (i.e., no closer together in time than the miss distance determination turn-around time) will generally lead to larger than necessary sample sizes--perhaps substantially larger. As a rough assessment of the effect of such "batch" testing relative to ordinary sequential testing, consider the nonparametric sequential test with $\alpha = \beta = .1$ and $CEP_1/CEP_0 = 2$. Then the "typical" expected sample size is about 6.3. Imagine for the moment sample size N is roughly exponentially distributed (which is certainly an oversimplification but is consistent with the observation that in many cases the mean and standard deviation of N are about the same, and is adequate for the present discussion). With batches of size eight, one batch would be required with a probability on the order of .7, two batches with probability about .2 and three batches with probability roughly .1. Thus the expected number of batches required would be about 1.4, or roughly 11 drops per cell on the average. Thus the effect of batch arrivals of observations in each cell is an increase in total drops for the experiment, perhaps by as much as 75%.

In summary, the implication of the foregoing discussion is that it may well be worth expending test resources to allow individual drops in more than one cell within a given operation. In addition, variables such as aircraft heading, time of day, order within the overall test sequence, weather, etc. should

"be varied" as much as practicable within a given cell of the design (that is, have as many variations and combinations of levels as practical associated with the drops in each given cell). This may be viewed as "buying insurance" against unforeseen effects of unknown causes in the experiment; in addition, such an approach may allow deduction of probable causes of system misbehavior in some cases of importance, should such difficulty be experienced in the PSVT.

Final comments on the field conduct that the author would like to mention are that there should be no possibility of specialized "tweeking" of the system (by either test personnel or the contractor) to alter its performance in any way for the tests. This may involve careful monitoring of any software changes, for example. Secondly, if the sequential approach to CEP testing is to be adopted, there should be a mechanism for assessing each drop miss distance (or hit-miss ourcome) in a period of time which is short relative to the following time interval standards. If individual drops are continued within a cell with a given operation only until sequential termination, the standard is the operation duration (hours?). If batch testing is used within each cell of the design, the standard would be time between operations (days?). If the individual drops within each operation are allocated to various cells of the design (which I recommend if at all possible), the standard is the time spent at a given range

(weeks?). Thus in the latter case there is perhaps not a measurement "turn-around time" problem at all, an additional bonus in taking this approach.

4. STATISTICAL ANALYSIS PLAN

There are two levels of analysis in the PSVT plan. The primary goal is to test whether system performance in each cell of the design is within design specifications. The secondary analyses concern determining which factors have significant effect, and what the effects are.

For the primary tests of CEP, there appear to be several alternatives: sequential parametric test (SPT), sequential nonparametric test (SNT), fixed-sample size parametric test (FPT) and fixed-sample size nonparametric test (FNT). The SPT and SNT are discussed in an earlier report [1] and we thus give only a very brief comment on them here. The FPT and FNT are discussed below. All of the tests involve testing whether the system displays accuracy (in each given cell of the design) to within the contract specified CEP, say CEP_0, or whether it has performance worse than some minimally acceptable performance (CEP_1). Thus the tests may be developed as tests of $H_0: M = CEP_0$ vs $H_a: M = CEP_1$, where M denotes the true (population) median radial miss distance of the system under the conditions of the given cell of the design. Both of the sequential test procedures are applications of Wald's Sequential Probability Ratio Test (SPRT). One, the SPT, is based on sequentially observing

(within each cell of the design) observed radial miss distances, and assuming a Rayleigh distribution model. The SNT is based on observing only whether each drop falls within CEP_0 and assuming a binomial distribution model. The SPT requires smallest average sample size, but its validity depends on whether the Rayleigh assumption is tenable (the latter assumption is implied by the assumption impact on the target plane follow a circular normal distribution, for example). The SNT requires somewhat larger samples on the average than does the SPT, but the binomial model involved is far less open to criticism on the grounds of invalidity due to assumption of distribution of radial miss distance.

The fixed sample size procedures are also based on the respective stochastic models (Rayleigh and binomial). If we consider the equivalent hypotheses about median squared radial miss distance and measure the squared radial miss distance of each drop, the Rayleigh model transforms to a chi-squared model which is somewhat more tractable computationally. In what follows we describe the FPT in these terms.

Suppose R is distributed Rayleigh so R^2 is distributed exponential with mean $C^2/\ln 2$, where C^2 is squared CEP. The likelihood ratio test of H_0:median$(R^2) = C_0^2$ vs H_a:median$(R^2) = C_1^2$ is based on the test statistic $T = \sum_{i=1}^{N} R_i^2$. H_0 is rejected whenever the calculated T is sufficiently large. Under H_0, $[(2 \ln 2)/C_0^2]T$ is distributed chi-squared

with 2N degrees of freedom

$$\left(\frac{2\ \ell n\ 2}{C_0^2}\ T \sim \chi^2_{(2N)}\right) .$$

Thus the FPT procedure for each cell of the design is to observe N radial miss distances, $R_1, R_2, \ldots, R_N$. Calculate the sum of squares, $T = \sum_{i=1}^{N} R_i^2$. Reject H_0 if T exceeds

$$\frac{C_0^2}{2\ \ell n\ 2}\ \chi^2_{(1-\alpha;2N)} ,$$

where $\chi^2_{(1-\alpha;2N)}$ is the $(1-\alpha)100\%$ point in the $\chi^2_{(2N)}$ tables. For example, with $C_0^2 = 1$ (i.e., R_i measured in C_0-units), $N = 9$ and $\alpha = 0.10$, this critical value is 18.747.

The FNT is a test of hypotheses about a binomial parameter p; $H_0: p \geq 1/2$ vs $H_a: p < 1/2$, where p represents the probability an impact falls within the contract specified CEP, say CEP_0. Assuming independence among the bomb impacts (see comments in Section 3 above), the number X of "hits" (impacts with $R_i \leq CEP_0$) in N drops is binomially distributed; further, under the null hypothesis it is binomial with parameter 1/2 ($X \sim b(N,1/2)$). The null hypothesis should be rejected if the observed value of X is on or below $b_{\alpha,N}$, where $b_{\alpha,N}$ is the largest value (obtained from the b(N,1/2) tables) such that $P[X \leq b_{\alpha,N}] \leq \alpha$. For example, with

$\alpha = 0.10$ and $N = 12$, this critical value is 3. Note: due to the discreteness of the binomial distribution, this procedure is somewhat conservative, in that the actual type-1 error probability for this example is .073, rather than the desired value, 0.10. If an exact test is desired, a randomized decision rule can be used (see E. Lehmann [3] for details).

The tests of CEP within each cell, discussed above, constitute the primary goal of the PSVT. Secondary goals include analyses of effects of various factors included in the design. An analysis of variance (AOV) is planned, using data from the baseline trials. These types of cells in the design received relatively greater numbers of drops, and form a factorial arrangement (with some unbalance in sample size). It is anticipated that the analysis of variance will be based on $(\log R_i)$ data, the log transformation being used to stabilize variance over the cells, a condition required in analysis of variance. To see the appropriateness of this transformation, consider the type of distribution that is likely to be sampled through observing radial miss distances $R_1, R_2, \ldots, R_N$ within a cell of the design. We anticipate that $R_i^2 \sim k \cdot \chi^2_{(2)}$, where "$\sim$" means "approximately distributed as" and k is a constant proportional to CEP^2. Then $E(R_i^2) \approx 2k$ and $V(R_i^2) \approx 4k^2$ so the standard deviation in a given cell is approximately proportional to the mean, i.e., $\sigma = k\mu = h(\mu)$, where h is linear. Then the transformation g given by

$$g(r^2) = \int \frac{1}{h(r^2)}\, dr^2 = \int \frac{1}{r^2}\, dr^2 = \ell n\ r^2$$

is commonly used to make σ constant over varying values of μ (see Curtiss [2], for example). But $\ell n\ r^2 \propto \ell n\ r$, hence analysis of variance can be performed on $\log R_i$ data. Appropriateness of this transformation can be assessed once the experimentation data are available.

If the speed and altitude levels actually attained in the trials vary substantially (say more than 10%) from the levels specified in the design matrix, one or both of these factors may be incorporated as covariates in an Analysis of Covariance (AOC), rather than the analysis of variance described above. Again, determination of whether this is necessary or desirable can be made once the experimentation data are available. For this purpose, the data arising from each drop should be in a format which includes measured values of speed and altitude.

In addition to the AOV or AOC, secondary analysis may include fitting a response surface to the observed drop data. This could be done using regression (perhaps weighted to accommodate inhomogeneity of variance) to estimate a surface giving system accuracy as a function of the variables altitude range and possibly speed, for the system in the Auto mode. Terms in the model should be selected so known and anticipated physical system characteristics and target/range characteristics are likely to be adequately represented. Although the dependent variable could be taken to be sample CEP in each cell, a better

model might result from modeling squared radial miss distance via the regression, then transforming predictions with this model to CEP predictions, if desired, using the Rayleigh-based relationship.

Finally, additional analyses (such as pairwise comparisons, cell CEP estimates, patterns of trial "aborts" and "outlier" rejections, etc.) and presentations of summary data should be undertaken. The precise nature of these analyses has not been explored as yet, and to a large extent will depend on the data obtained. Close coordination with test personnel should also be maintained by the analyst, in order to assist in determining what additional analyses would be appropriate.

It is planned to use the ARIS system to assist in determining causes for observed large misses. This procedure constitutes an "outlier" rejection rule, which could bias the experiment, as follows. If only large miss drops are subjected to the ARIS screening, the overall effect will be to possibly eliminate some of the large misses, which in turn makes the remaining drops appear more accurate. Such screening may be appropriate; however, we suggest two actions which may assist in determining whether biasing has occurred. First, records of any such eliminated drops should be kept, for possible subsequent analysis. Second, the ARIS screen should be applied formally to a sample of "good" drops, using the same rejection criteria as for the outlier cases. Records should be kept of the results of such screening of "good" drops. These can

be used to help assess the degree of bias that may have been induced by elimination of drops with large miss distances that were not actually outliers.

With the large number of individual tests being performed with the primary analysis (i.e., one CEP test in each cell of the design), it is likely that there will be a mixture of rejections and acceptances of the contract specified CEP's. There will occur, therefore, the problem of making an overall assessment of whether the system is sufficiently accurate. It would be a good idea to explore this problem with the decision maker, and to indicate how changing the Type I and Type II error rates (α and β, respectively) affect the accept/reject patterns that may be encountered. Perhaps the significance of the observed number of rejections can be assessed in terms of physical explanation of system patterns, as well as the conditions anticipated in actual operational use of the system. The binomial distribution may be of some use in determining whether the number of rejections is significant, or perhaps Fisher's method [4] of combining experimental results can be used.

It should be borne in mind that theoretically the secondary analyses may be affected by the stopping rule used in the primary tests. If sequential tests are used for the primary analysis, the data in each cell are, in a mild sense, conditional, given the data obtained led to acceptance or

rejection, as the case may be. It is not anticipated that this simultaneous inference effect will be great enough to cause difficulties from a practical point of view, however.

5. A SAMPLE SIZE REDUCTION METHOD

We have argued elsewhere [1] that the major shortcoming of the Rayleigh model for unguided weapon misses is that in some applications it fails to adequately fit the upper tail of the miss distance distribution. Even in such cases, however, the model may provide useful results for the major portion of the miss distribution short of the very large misses. In what follows we describe such an application of the Rayleigh distribution to reduce sample size required in the primary analyses concerning CEP testing. This approach is applicable to both the SNT and FNT. Throughout, we assume the Rayleigh model provides reasonable fit to the radial miss distribution except possibly for the upper tail region (which we define here as the set of points larger than the upper 95% point in the Rayleigh distribution).

Suppose, then, under fixed conditions the squared radial miss cumulative distribution function is

$$F_{R^2}(y) = 1 - \exp(-y \ln 2/C^2), \quad y \geq 0 ,$$

where C^2 is the median of R^2 (i.e., C^2 is the square of

the system CEP). Let C_0^2 denote the squared CEP under the null hypothesis H_0:CEP = C_0 and assume the alternative hypothesis is H_a:CEP = $C_1 = kC_0$. For convenience in notation, assume miss distances are measured in C_0-units, so $C_0 \equiv 1$, and k represents the C_1/C_0 ratio. Recall both the SNT and FNT are based on the binomial distribution of the number of hits inside a circle of radius 1 (= C_0). Under the null hypothesis the probability of hitting this circle is

$$F_{R^2}(y) = 1 - \exp(-\ln 2/1^2) = 0.5$$

and under H_a the probability of such a hit is

$$F_{R^2}(1) = 1 - \exp(-\ln 2/k^2) .$$

For example, with $k = 2$ this probability is $1 - \exp(\ln 2/2) \approx .2929$.

The basic idea we wish to discuss is that of allowing the definition of "hit" to be associated with circles of radii possibly different from C_0. We shall show that even though we maintain the null and alternate hypotheses about CEP described above, the binomial data to test these hypotheses can be made be far more efficient by defining the hit/miss criterion differently. Let $p_0(C)$ denote the probability under H_0 of observing a miss distance within C units of the target, and similarly let $p_1(C)$ denote that probability under H_a. Then

$$p_0(C) = P[R^2 \leq C^2 | CEP = 1] = 1 - \exp(-C^2 \ln 2) = 1 - 2^{-C^2}$$

and

$$p_1(C) = P[R^2 \leq C^2 | CEP = k] = 1 - \exp(-C^2 \ln 2/k^2)$$

$$= 1 - 2^{-C^2/k^2} = 1 - (1 - p_0(C))^{1/k^2}$$

We wish to determine C so as to minimize the sample size N (or in the sequential case, Expected sample size) required to achieve a test of H_0 vs H_a with preselected operating characteristics α and β. Our procedure is to express N as a function of C, then minimize. For ease of presentation we use the arcsine transformation of binomial random variables to normality [2], and limit ourselves to the Fixed sample size case (although neither of these conveniences is necessary).

With some radius C of the hit circle definition, the test of H_0 vs H_a would be based on $\bar{X}$, the observed relative frequency of hits. The null hypothesis is rejected for $\bar{X}$ sufficiently small. For any selected value of C, let $p(C)$ denote the corresponding probability an individual bomb results in a hit. For even moderate values of N,

$$Y = 2 \sin^{-1} \sqrt{\bar{X}} \sim N(2 \sin^{-1} \sqrt{p}, \frac{1}{N})$$

although the approximation may be quite rough if p is "extreme" (outside the interval (.05, .95) or so). Note: the angle $2 \sin^{-1}\sqrt{\cdot}$ is measured in radians. Now, in terms of

the test statistic Y, because $2 \sin^{-1} \sqrt{\cdot}$ is monotone increasing, H_0 should be rejected if $Y \leq d$, where the critical value d and sample size N are selected so that the desired size and power are attained:

$$P[Y \leq d | C = C_0] = \alpha ,$$

$$P[Y \leq d | C = C_1] = 1 - \beta .$$

Using the arcsine transformation described above, these conditions are met (at least to good approximation) provided

$$d - 2 \sin^{-1} \sqrt{p_0} = z_\alpha / \sqrt{N} ,$$

$$d - 2 \sin^{-1} \sqrt{p_1} = z_{1-\beta} / \sqrt{N} ,$$

where z_δ is the δ^{th} quantile of the standard normal distribution. Thus in order to minimize N subject to meeting the α and β requirements it suffices to maximize

$$f(p_0) = \sin^{-1} \sqrt{p_0} - \sin^{-1} \sqrt{p_1}$$

$$= \sin^{-1} \sqrt{p_0} - \sin^{-1} \sqrt{1 - (1-p_0)^{1/k^2}}$$

This is easily done for various fixed values of the CEP_1/CEP_0 ratio k. Values of $f(p_0)$ can be used to estimate FNT

sample size requirements through the approximation

$$N \approx \left[\frac{z_{1-\beta} - z_{\alpha}}{2f(p_0)}\right]^2 .$$

As an example to demonstrate this idea, suppose $\alpha = \beta = 0.10$, $k = 2$. Then values of $f(p_0)$, the radius C of the "hit" circle, and approximate sample sizes for the FNT are as shown in Table 3.

The maximum of $f(p_0)$ occurs at $p_0 = .94$ and this theoretically minimizes N. Note, however, that $p_0 = .90$ yields the same savings in sample size and has the advantage of not involving the model so far into the upper tail as does the sample size minimizing value, .94. Note the sample size requirement with $p_0 = .90$ is substantially below the $C = CEP_0$ defined "hit" circle described in Section 4 in connection with the SNT and FNT, where $p_0 = .50$. The relative reduction in approximate sample size requirements for the example discussed above are shown in Table 3.

As mentioned above, this sample size reduction scheme can be used for both the SNT and FNT, although only the FNT case was illustrated by example. The approximation involved through the arcsine transformation is an inessential part of this development; it was used here for simplicity of demonstrating the potential of this approach.

"hit" circle radius C	"hit" probability under H_0 p_0	$f(p_0)$	approximate sample size N	% savings in sample size
.39	.1	.1596	65	
.72	.3	.2832	21	
1.00	.5	.3751	12	0
1.15	.6	.4165	10	17
1.32	.7	.4562	8	33
1.52	.8	.4939	7	42
1.82	.9	.5262	6	50
2.01	.94	.5329	6	50
2.16	.96	.5312	6	50
2.58	.99	.4970	7	42

TABLE 3. Comparison of Sample Sizes for Various "hit" Definitions, for $\alpha = \beta = .1$, $k = 2$.

6. RECOMMENDATION

Based on the information available, the following approach to the PSVT is recommended: use the SNT for primary CEP testing, possibly with reduction in E(N) using the method described in Section 5. However, if this reduction scheme is adopted, the definition of the "hit" circle should not be allowed to involve p_0 values too extreme (i.e., C^2 values too far in the upper tail of the Rayleigh distribution). Probably a reasonable upper bound for p_0 is .90.

The tests should be conducted so as to deliver individual drops in each cell of the design on different days, to the extent possible. Drops should be made in each cell so that Uncontrolled variables (such as day, time of day, heading, pilot, aircraft, weather, etc.) vary over as wide a span as practicable. As pointed out in the preceding, this approach yields the following advantages: (1) it gives observations which are more nearly independent; (2) it provides estimates of CEP which are more realistc; (3) it avoids the increase in sample size with batch testing; and (4) it may give more time to measure miss distances.

Secondary analyses of the radial miss data, including (but not limited to) analysis of variance, analysis of covariance, and multiple regression should be performed. Transformations to stabilize variance and weighted regression should be used if the data suggest there is lack of homogeneity of variance.

REFERENCES

[1] Barr, D.R., "Two Sequential CEP Tests," Naval Postgraduate School Technical Report NPS55BN74061 (1974).

[2] Curtiss, J.H., "On Transformations Used in the Analysis of Variance," Annals of Mathematical Statistics, 14, (1943), 107-122.

[3] Lehmann, E.L., "Testing Statistical Hypotheses", John Wiley and Sons, New York (1959).

[4] Pearson, E.S., "On Questions Raised by the Combination of Tests Based on Discontinuous Distributions," Biometrika, 37, 383-398 (1950).

INITIAL DISTRIBUTION LIST

	No. Copies
Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
Dean of Research Code 023 Naval Postgraduate School Monterey, California 93940	1
Library (Code 0212) Naval Postgraduate School Monterey, California 93940	2
Library (Code 55) Naval Postgraduate School Monterey, California 93940	2
Major Dave Allen MCTSSA, TSCRB MCB, Camp Pendleton, California 92055	1
Major Earl Peete Marine Aviation Detachment Point Mugu, California 93042	1
Mr. W. F. Trisler Code 05431 NAVELEX Washington, D.C. 20360	1
Capt. Jerry Paccassi Code 37 Naval Postgraduate School Monterey, California 93940	1
Major Lou Simpleman 323 Plantation Dr. New Bern, N.C. 28560	1
Major Ted Bean Development Center, C^3 Division MCDEC Quantico, Virginia 22134	1
Department of Operations Research Naval Postgraduate School Monterey, California 93940	
Attn: Professor F. R. Richards	1
Professor D. R. Barr	36