AD

AD737196

REPORT NO. RD-TR-71-26

FREQUENT MISCONCEPTIONS FROM THE USE OF CONFIDENCE STATEMENTS AS THE LONE CONSIDERATION IN RELIABILITY REQUIREMENTS

by
Dr. Raymond H. Myers
Department of Statistics
Virginia Polytechnic Institute
and
Nancy R. Rich

November 1971

Approved for public release; distribution unlimited.

U.S. ARMY MISSILE COMMAND

Redstone Arsenal, Alabama

R17

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Aeroballistics Directorate Directorate for Research, Development, Engineering and Missile Systems Laboratory U.S. Army Missile Command Redstone Arsenal, Alabama 35809	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP NA

3. REPORT TITLE

FREQUENT MISCONCEPTIONS FROM THE USE OF CONFIDENCE STATEMENTS AS THE LONE CONSIDERATION IN RELIABILITY REQUIREMENTS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

Dr. R. H. Myers
Nancy R. Rich

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
22 November 1971	17	0

8a. CONTRACT OR GRANT NO. b. PROJECT NO. (DA) 1X263302D212 AMC Management Structure Code No. c. 523B.12.17100 d.	9a. ORIGINATOR'S REPORT NUMBER(S) RD-TR-71-26 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AD________

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
None	Same as No. 1

13. ABSTRACT

This report presents a discussion of the disadvantages of the use of such phrasing as "the system must demonstrate 95 percent reliability with 90 percent confidence" when the required reliability for the system is 95 percent.

Several examples are offered to illustrate how such wording can be misleading. A brief summary of elementary estimation theory is presented.

The recommendation is given that such phrases as "95 percent reliability with 90 percent confidence" be stricken from all Army documents.

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

DISPOSITION INSTRUCTIONS

Destroy this report when it is no longer needed. Do not return it to the originator.

DISCLAIMER

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

TRADE NAMES

Use of trade names or manufacturers in this report does not constitute an official indorsement or approval of the use of such commercial hardware or software.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Confidence statements						
Reliability requirements						
Elementary estimation theory						
Point and interval estimation						

22 November 1971 Report No. RD-TR-71-26

FREQUENT MISCONCEPTIONS FROM THE USE OF CONFIDENCE STATEMENTS AS THE LONE CONSIDERATION IN RELIABILITY REQUIREMENTS

by

Dr. Raymond H. Myers
Department of Statistics
Virginia Polytechnic Institute
and
Nancy R. Rich

DA Project No. 1X263302D212
AMC Management Structure Code No. 523B.12.17100

Approved for public release; distribution unlimited.

Aeroballistics Directorate
Directorate for Research, Development, Engineering
and Missile Systems Laboratory
U.S. Army Missile Command
Redstone Arsenal, Alabama 35809

CONTENTS

ABSTRACT

This report presents a discussion of the disadvantages of the use of such phrasing as "the system must demonstrate 95 percent reliability with 90 percent confidence" when the required reliability for the system is 95 percent.

Several examples are offered to illustrate how such wording can be misleading. A brief summary of elementary estimation theory is presented.

The recommendation is given that such phrases as "95 percent reliability with 90 percent confidence" be stricken from all Army documents.

1. Introduction

In recent years the application of probability and statistics in the development of Army weapon systems has grown at a rapid pace. Unfortunately, the increase in the number of statisticians employed by the Army has not been correspondingly high. In the scramble to fill the void, management, engineers, and other personnel have been understandably guilty of instances of improper or deceptive use of statistical terms. For example, the strong emphasis in statistics books on the necessity of giving confidence intervals or some other measure of variability along with point estimates of a parameter likely has been innocently but mistakenly taken to mean the point estimate should be disregarded altogether. The result of this misinterpretation is that certain documents relating to system requirements contain such wording as "the system must demonstrate 80 percent reliability with 90 percent confidence". The results of techniques for reporting or testing based upon such phrasing can be very misleading. After a brief summary of elementary estimation theory, this document presents the disadvantages associated with this concept.

2. Point and Interval Estimation

a. The Use of Sample Statistics

Suppose that it is important to know one or more characteristics of a population of items but it is impossible or impractical to measure that characteristic for each member of the population. For example, the hit probability of a missile system is an important parameter; however, it would be unreasonable to fire every missile round manufactured to ascertain exactly what percentage of the rounds hit the target. To obtain some information, albeit incomplete, of a population parameter, a sample is drawn from the population and measurements are made on the sample.

A statistic is a function of the data from a sample. The functional form of the statistic is determined, the sample data are gathered, and the value of the statistic for that sample is calculated. If a different sample were to be drawn from the same population, the new calculated value of the statistic might be different from the first value.

The statistic is subject to chance variation and is thus a random variable. The probability structure of the statistic, which can often be determined from certain assumptions, dictates the conclusions or inferences that are made concerning the parameter.

The sampling properties of statistics are fundamental in parameter estimation, a subject briefly discussed here. The fundamentals of point and interval estimation as presented and their relevance to statistical problems associated with missile systems are emphasized.

b. Fundamentals

Suppose that the flight time of a certain rocket for a fixed set of conditions (weather, quadrant elevation, etc.) is of interest. A number, N, of the rockets are fired under the same conditions and the flight time is measured for each round. It is desirable to condense the observations from the sample into two statistics (the mean flight time, μ, and the variance, σ^2) which will estimate the two population parameters. The values μ and σ^2 could be found without error only if every rocket produced were fired. However, the sample observations can contribute information which enables the experimenter to learn more about these two important parameters.

An intuitively appealing estimate of the mean is the sample arithmetic average, computed as the sum of the N flight times divided by N. If y_i is the flight time for the i^{th} round, then the sample average $\bar{y}$ is equal to:

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i \quad .$$

As shown in the appendix, this estimator is unbiased. Further, an unbiased estimator for σ^2 is the statistic s^2, found as the sum of squares of deviations about the sample average divided by N-1, or

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(y_i - \bar{y}\right)^2 \quad .$$

Thus, the point estimates for μ and σ^2 can be found by the fomulas for $\bar{y}$ and s^2, respectively.

A point estimate can be compared to a measurement made by a hypothetical machine. The sample observations are fed into the machine and the point estimate is output. As with any measuring device, the accuracy of the measurement is of great importance. It is desirable to be able to make a statement analogous to "the value is given to the nearest tenth". Thus, in addition to the point estimate, it would be helpful to have a range of values, or an interval, which is likely to cover the true parameter. Such an interval is called a confidence interval.

A 100 (1 - α) percent confidence interval estimate on a parameter is one for which the probability is 1 - α that the interval covers the true parameter. The upper and lower confidence limits or bounds are random variables, the values of which depend upon the sample size, the confidence coefficient (or confidence level), and the sample information. Normally the calculation of the confidence interval estimate is accompanied by the calculation of the point estimate. The combination of the point estimate and the confidence interval estimate provides much more information than either one alone.

In the flight time example, suppose that N = 10 rounds had been fired and the estimated values of μ and σ^2 had been

$$\bar{y} = 1.98 \text{ seconds}$$

$$s^2 = 0.010 \text{ second}^2 \quad .$$

Then a 90 percent confidence interval estimate for μ (taken as symmetric about $\bar{y}$) is given by

$$\left(\bar{y} - t_{9,\, 0.05} \frac{s}{\sqrt{N}} \quad , \quad \bar{y} + t_{9,\, 0.05} \frac{s}{\sqrt{N}} \right)$$

where $t_{9,\, 0.05}$ is the fifth percentile point of the Student's t distribution with N - 1 = 9 degrees of freedom. The confidence interval estimate is given by

$$\left(1.98 - 2.262 \frac{\sqrt{0.010}}{\sqrt{10}} \quad , \quad 1.98 + 2.262 \frac{\sqrt{0.010}}{\sqrt{10}} \right)$$

or

$$(1.91, \quad 2.05) \quad .$$

Thus, for the sample of size 10, the point estimate of the mean flight time is 1.98 seconds and a 90 percent confidence interval for the mean flight time is (1.91 seconds, 2.05 seconds).

Now suppose that 90 more rounds were fired. These can be combined with the first run, making the total sample size N = 100. The point estimates would probably change, e.g.,

$$\bar{y} = 1.95 \text{ seconds}$$

$$s^2 = 0.012 \text{ second}^2.$$

The new 90 percent confidence interval would be

$$\left(1.95 - 1.960 \frac{\sqrt{0.012}}{\sqrt{100}} , \quad 1.95 + 1.960 \frac{\sqrt{0.012}}{\sqrt{100}}\right)$$

or

$$(1.93, \quad 1.97) \quad .$$

Notice that the width of this confidence interval is smaller than that of the preceding confidence interval.

As a second example, consider a reliability test for a certain missile. Suppose that of 75 missiles fired, 61 were reliable. What is the point estimate of reliability and what is a two-sided 90 percent confidence interval? An appealing point estimate of reliability is the ratio of the number of successful (reliable) rounds to the total number of rounds fired. The point estimate in this case is

$$\hat{R} = \frac{61}{75} = 0.81 \quad .$$

Since the sample size is fairly large, the normal approximation to the binomial distribution can be used in the calculation of the 90 percent confidence interval. The 90 percent confidence interval would be

$$\left(\hat{R} - z_{0.05} \frac{\sqrt{\hat{R}(1 - \hat{R})}}{\sqrt{N}} , \quad \hat{R} + z_{0.05} \frac{\sqrt{\hat{R}(1 - \hat{R})}}{\sqrt{N}}\right) ,$$

where $z_{0.05}$ is the fifth percentile point of a standard normal random variable. In this case, the 90 percent confidence interval is

$$\left(0.81 - 1.645 \frac{\sqrt{(0.813)(0.187)}}{\sqrt{75}} \quad , \quad 0.81 + 1.645 \frac{\sqrt{(0.813)(0.187)}}{\sqrt{75}}\right) .$$

Thus, the point estimate of reliability is 0.81 and a 90 percent confidence interval is (0.74, 0.89).

Quite often in reliability applications a one-sided confidence interval is quoted. In this case, the upper bound of reliability would be set to 1.00 and the lower 90 percent bound would be calculated as

$$\hat{R} - z_{0.10} \frac{\sqrt{\hat{R}(1 - \hat{R})}}{\sqrt{N}} = 0.81 - 1.282 \frac{\sqrt{(0.813)(0.187)}}{\sqrt{75}} = 0.76 .$$

Thus, the point estimate is 0.81 and a one-sided 90 percent confidence interval is (0.76, 1.00).

The clear distinction between this and the standard two-sided interval is that the one-sided confidence interval affords protection in only one direction and should be used only when the situation dictates.

For a third example, suppose that a certain radar system is to be used on a continuous basis. It is important to know how often repair action will be required on the system. The radar is set up and the time until the first system failure is measured. After repairs are effected, the system is reactivated and the measuring procedure begins again. This continues until 15 times between failure are measured. An exponential distribution for the time between failure is assumed. The parameter in this distribution is the mean time between failure, θ. An appropriate point estimate for failure is the arithmetic average of the times between failure. For this case, the sample average is

$$\hat{\theta} = 63.4 \text{ hours}.$$

The equation for a 95 percent confidence interval for θ is

$$\left(\frac{2N\hat{\theta}}{\chi^2(2N,\ 0.975)} \quad , \quad \frac{2N\hat{\theta}}{\chi^2(2N,\ 0.025)}\right)$$

where $\chi^2(2N, 0.975)$ is the 97.5 percentile point of a chi-square random variable with 2N degrees of freedom and $\chi^2(2N, 0.025)$ is the 2.5th percentile point. In this case, the lower 95 percent confidence interval is

$$\left(\frac{2(15)(63.4)}{47.0}, \frac{2(15)(63.4)}{16.8}\right) .$$

Thus, the point estimate of the mean time between failure is 63.4 hours and the 95 percent confidence interval estimate is (40.5, 113.2).

3. Disadvantages of Improper Phrasing

The phrasing "B percent reliability with A percent confidence", which is apparently used to a great extent to assess the reliability of systems, can be misleading and deceptive. It should first be emphasized that the use of confidence intervals should not be abandoned. To the contrary, confidence interval estimation is an important form of inference and, in general, its use is certainly preferable to citing point estimates alone. However, this type statement, without the quotation of the point estimate, can make the system appear quite unreliable when in fact the system reliability is indeed acceptable.

The use of the phrase "B percent reliability with A percent confidence" puts no direct emphasis on the point estimate of reliability. Consider a very simple case in point. Assume that it is important to maintain 80 percent reliability for a particular system. Suppose that 10 observations are made on the system with nine successful results. Then the interval (0.80, 1.00) would be the 62 percent one-sided confidence interval. A statement consistent with the phrasing in question is "80 percent reliability has been demonstrated with 62 percent confidence". It would appear from the statement that one is reasonably sure that the system is at best barely acceptable, acceptability being taken as 80 percent; when in fact the data indicated that the best estimate of reliability is R = 9/10 (100) percent or 90 percent. Now the confidence interval is valid but the most direct evidence concerning the reliability, namely the point estimate is not mentioned in the concluding statement. A more informative statement would be "the point estimate of reliability is 0.90 and a 90 percent confidence interval is (0.61, 0.96)".

The confusion involved in quoting a one-sided confidence interval alone is emphasized in the case of small samples. When the sample size is small, the confidence interval is wide and the point estimate becomes remote from the lower confidence bound. For example if 3 out of 4 runs are successful in the sample, the lower 90 percent confidence bound is 0.32. It is truly deceiving to state "with 90 percent confidence there is 32 percent reliability". Certainly it must be stated somewhere that the sample reliability is 0.75. Further information would be revealed by the two-sided 90 percent confidence interval (0.25, 0.90) along with the point estimate of 0.75.

The "B percent reliability with A percent confidence" calculation itself is not improper. However, two difficulties cloud the issue:

a) The calculation is not accompanied by the point estimate indicating the sample reliability.

b) The statement, by its very nature generates confusion as to what is meant by "B percent reliability".

To elaborate on point b), the statement should be taken to mean that the reliability parameter is at least B percent. This is consistent with the A percent confidence interval (B, 100). So, for the example given in the previous section, a fairer statement, one which does not depict quite the bleak picture generated by the existing statement, would be as follows: "the sample reliability was found to be 0.90 and there is 62 percent confidence that the true reliability exceeds 0.80".

4. Possible Alternate Approach

There is yet another difficulty which is implicit in the present phrasing and accompanied assessment of a system. Suppose that 80 percent reliability (or better) is desirable for a system. Presumably, under the present system, one would not accept that system unless he was able to "demonstrate" 80 percent reliability (at least) with, e.g., 90 percent confidence. This will certainly insure that very few bad systems (i.e., unreliable ones) are accepted as reliable. Indeed there will be 0.10 probability of accepting a system when in fact the reliability is as low as 80 percent. However, it does not "protect" against erroneously rejecting good systems. For example, suppose that the true reliability is 0.82, certainly an acceptable value. Some elementary compuations can show that using the present scheme with 90 percent confidence as the criterion and a sample size as large as 100, then 78 percent of the time the system will not be accepted. It would be desirable if one could have a 10 percent chance of accepting a system that is barely acceptable and yet a 10 percent chance of rejecting a system that is "good", the latter being defined on the basis of a reliability of, say, 0.82. That is, an unreliable system should not be accepted, but there should be some assurance that a quality system would be accepted.

This approach could be used and the probabilities attained with the control of sample size. The hazard, from the practical point of view, is that it will lead to what is likely an unusually large number of tests. If this sample size is unattainable, there should at least be some documentation regarding the values of the probability that good systems, i.e., ones with true reliability above the minimum, will be rejected.

5. Summary

There are definite disadvantages to the use of such wording as "the system must demonstrate 80 percent reliability with 90 percent confidence", when 80 percent reliability is a requirement. In such phrasing, only a lower confidence limit is considered; the point estimate is ignored altogether. Even with an acceptably reliable system, if the number of tests or test hours is limited, stating only a lower confidence limit likely creates the deceptive impression that the reliability of the system is much less than its true value. Thus, the practice of citing only a lower confidence limit can be quite misleading. For a more representative depiction, the point estimate of reliability and the lower and upper confidence limits should be given. For only if the point estimate is given is it directly specified how the system actually performed in the test.

It is strongly recommended that the usage of such phrases as "80 percent reliability with 90 percent confidence" be purged from all Army documents.

Appendix UNBIASED ESTIMATORS AND THE NORMAL APPROXIMATION TO THE BINOMIAL

1. Unbiased Estimators

The bias of an estimator $\hat{\theta}$ is defined by

$$\text{bias}\ (\hat{\theta}) = E(\hat{\theta}) - \theta$$

where the notation E refers to the expectation or long run average value. The statistic $\hat{\theta}$ is said to be an unbiased estimator for θ if $E(\hat{\theta}) = \theta$. For example, no matter what the population distribution, each observation is an unbiased estimator of the mean, $E(x_i) = \mu$. The sample average is also unbiased for the mean,

$$\text{Sample average} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$E(\text{sample average}) = \mu \quad .$$

2. Normal Approximation to the Binomial

For large samples, i.e. $N > 20$, the normal approximation to the binomial can be used in finding confidence intervals on the reliability parameter p. The usual structure for the 100 $(1 - \alpha)$ percent confidence interval for a population mean from some distribution with variance σ^2 is given by

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$$

where $\bar{x}$ is a sample average of size N and $z_{\alpha/2}$ is the 100 $(1 - \alpha)$ percentile point of the standard normal distribution. For a sample from a binomial distribution (success or failure observations), the estimator for the binomial parameter p (the reliability) is given by

$$R = \frac{\text{No. of successes}}{\text{No. of observations}}$$

i.e., the proportion of successes. The expected value and variance of R are given by

$$E(\hat{R}) = p$$

$$\mathrm{Var}(\hat{R}) = \frac{p(1 - p)}{N}$$

and the normal distribution of $\hat{R}$ is a reasonable approximation for large samples. Thus the structure given for the 100 $(1 - \alpha)$ percent confidence interval is

$$\left(\hat{R} - z_{\alpha/2} \frac{\sqrt{\hat{R}(1 - R)}}{\sqrt{N}} \quad , \quad \hat{R} + z_{\alpha/2} \frac{\sqrt{\hat{R}(1 - R)}}{\sqrt{N}} \right) .$$