

CONSIDERATIONS WHEN USING CONFIDENCE INTERVALS TO COMPARE FLIGHT TEST DATA SETS

Author

Elwood T. Waddell, Jr. Capt, USAF, Technical Support Deputy Director, USAF TPS Instructor

Timothy R. Jorris, Lt Col, USAF, PhD, USAF TPS Instructor

David L. Vanhoy, Technical Director, USAF TPS Instructor

ABSTRACT

Flight test data is often used to validate a prediction model or compare results to previous data. Thus, from the initial design and modeling through fielding and system upgrades, the very nature of flight testing requires comparisons of data sets to be made.

Most generally, the determinations of whether or not two data sets are the same or different is made by comparing the confidence intervals, or error bars, associated with the data sets.

Unfortunately, the details of how the comparison is done can vary from organization to organization, and even from person to person. Further, the conclusions drawn with these methods are affected by a number of factors which are often not taken into account during the comparison.

This paper will compare the results of conclusions obtained using various confidence interval overlap rules to those obtained by using the two tailed t-test. The t-test is known to be the uniformly most powerful unbiased test for determining differences in population means, and as such, is the "gold standard" against which such comparisons should be gauged.

Factors such as confidence levels, sample sizes, and standard deviations will be explored, and a set of guidelines will be presented which will describe when overlap methods can be trusted, and when they should be considered suspect.