

Statistical Tolerance Interval Applications and Challenges at NASA

Peter A. Parker, Ph.D., P.E.

National Aeronautics and Space Administration

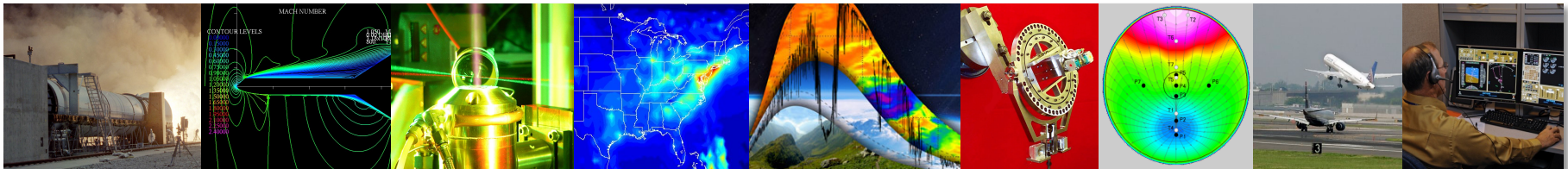
Langley Research Center

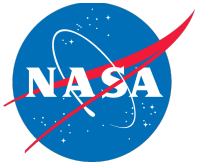
Hampton, Virginia

June 14th, 2022

Quality and Productivity Research Conference

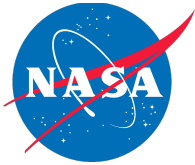
San Francisco, CA





Objectives of this Presentation

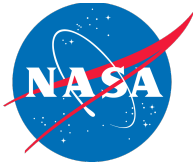
- **Introduce statistical tolerance interval applications at NASA**
 - **Nondestructive inspection capability**
 - **Material allowables**
- **Highlight the challenges arising in practice**
 - **Small sample sizes to estimate intervals on extreme quantiles**
 - **Choice of distributional models**
 - **Available software tools for interval estimation**
 - **Interpreting tolerance intervals for subject-matter experts**



Introduction

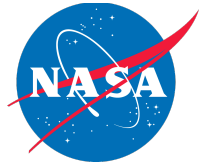
- **In aerospace research and development, it is often required to estimate a limiting capability from a small sample of test data**
 - **A quantity that bounds a large proportion of the population of interest at a stated confidence level**
- **For example, we may need to estimate a limiting strength of a component that is expected to be exceeded by at least 90% of similar components with 95% confidence, known as a material allowable.**
 - **Estimated quantities are critical in evaluating aerospace system performance margins and mission risk**
- **A statistical tolerance interval is the classical approach to estimate these limiting quantities of interest**
- **Usage of statistical tolerance intervals is often mandated/codified in NASA Standards**
 - **Leads to questions of “how” to estimate not “what” to estimate for a statistician, but explaining the “what” and the “how” are significant challenges**

Implementing Tolerance Intervals

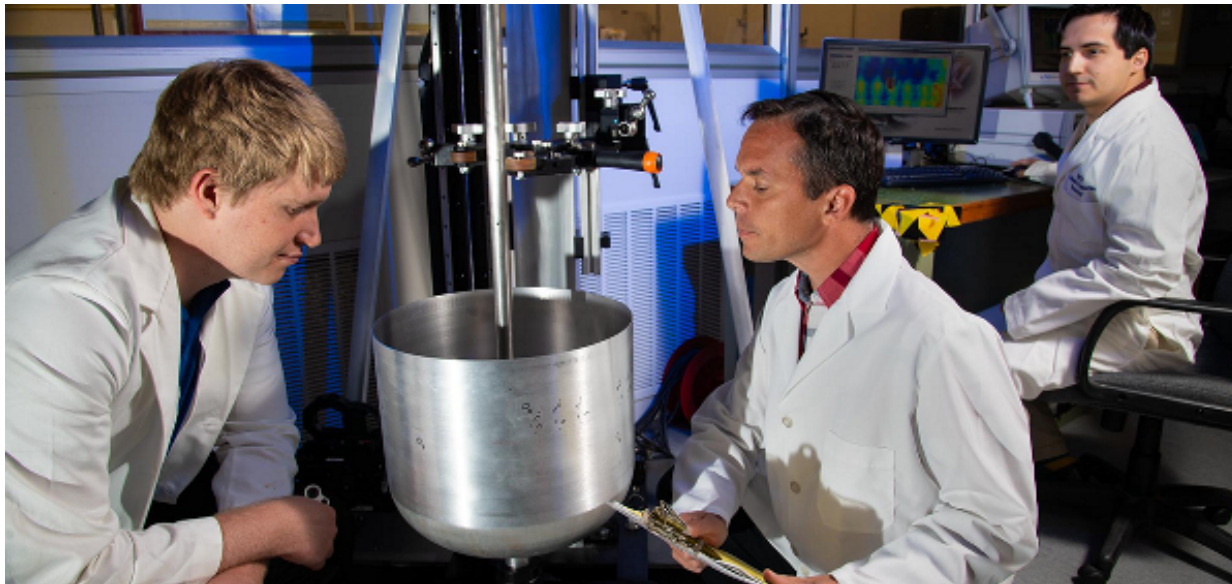


- **While a statistical tolerance interval is not a new concept, they are often unfamiliar to subject-matter experts and even to many applied statisticians (*that was me*).**
 - **Can be confusing in an engineering context, in fact, the term “tolerance” is often confused with a dimensional tolerance**
- **At NASA, tolerance interval estimation is becoming more common, critical, and challenging with a desire to rely on increasingly smaller sample sizes and reduce margins to increase system performance**
 - **Sample sizes driven by cost, very costly in space systems**
 - **Reducing margins often is related to reducing mass, which is king**

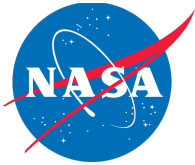
Nondestructive Evaluation Capability



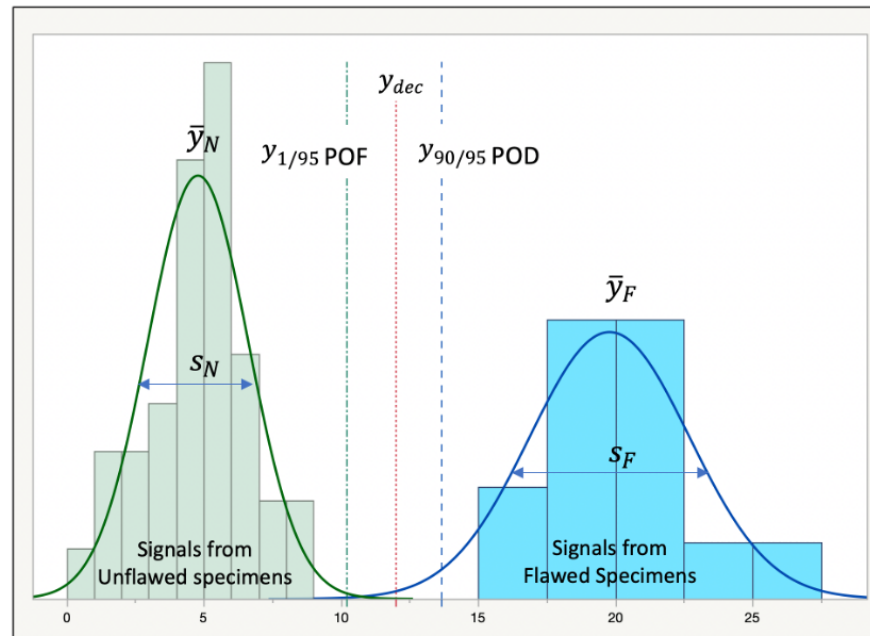
- **Nondestructive evaluation (NDE) methods include radiography, ultrasound, similar to medical applications – noninvasive**
- **The capability of a NDE method is characterized by its probability of detection (POD) and probability of a false positive (POF)**
- **This defines a critical flaw size that is assumed to exist in the spaceflight component and supports operational lifetime analysis**



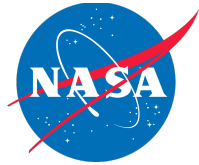
Probability of Detection Demonstration



- **POD study is conducted with multiple flawed and unflawed specimens and response signals (y) from the NDE method are recorded**
 - **Flawed specimens used to estimate a POD threshold**
 - **Unflawed to estimate a probability of a false call (POF) threshold**
- **At least 90% of the signals exceed the threshold with 95% confidence, when a flaw of a specified size is present**



Statistical Tolerance Interval Defined



- **“...an interval that one can claim to contain at least a specified proportion of the distribution with a specified confidence...”**

Meeker, Hahn, Escobar (2017)

- **“...a way to cover a fixed proportion of the population with a stated confidence.”**

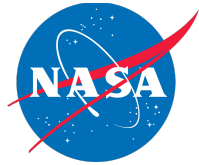
NIST/Sematech e-Handbook of Statistical Methods

- **“...an interval, based on a random sample, that is expected to contain a specified proportion or more of the sampled population.”**

Krishnamoorthy, Mathew (2009)

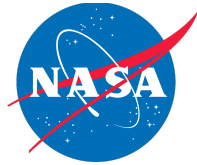
- **In plain language, an estimated value that bounds a large proportion of the population of interest at a stated confidence level**
- **For most NASA applications, we are interested in a one-sided tolerance bound, also known as a one-sided confidence bound on a distribution quantile.**

Strength of Fabric-Type Components



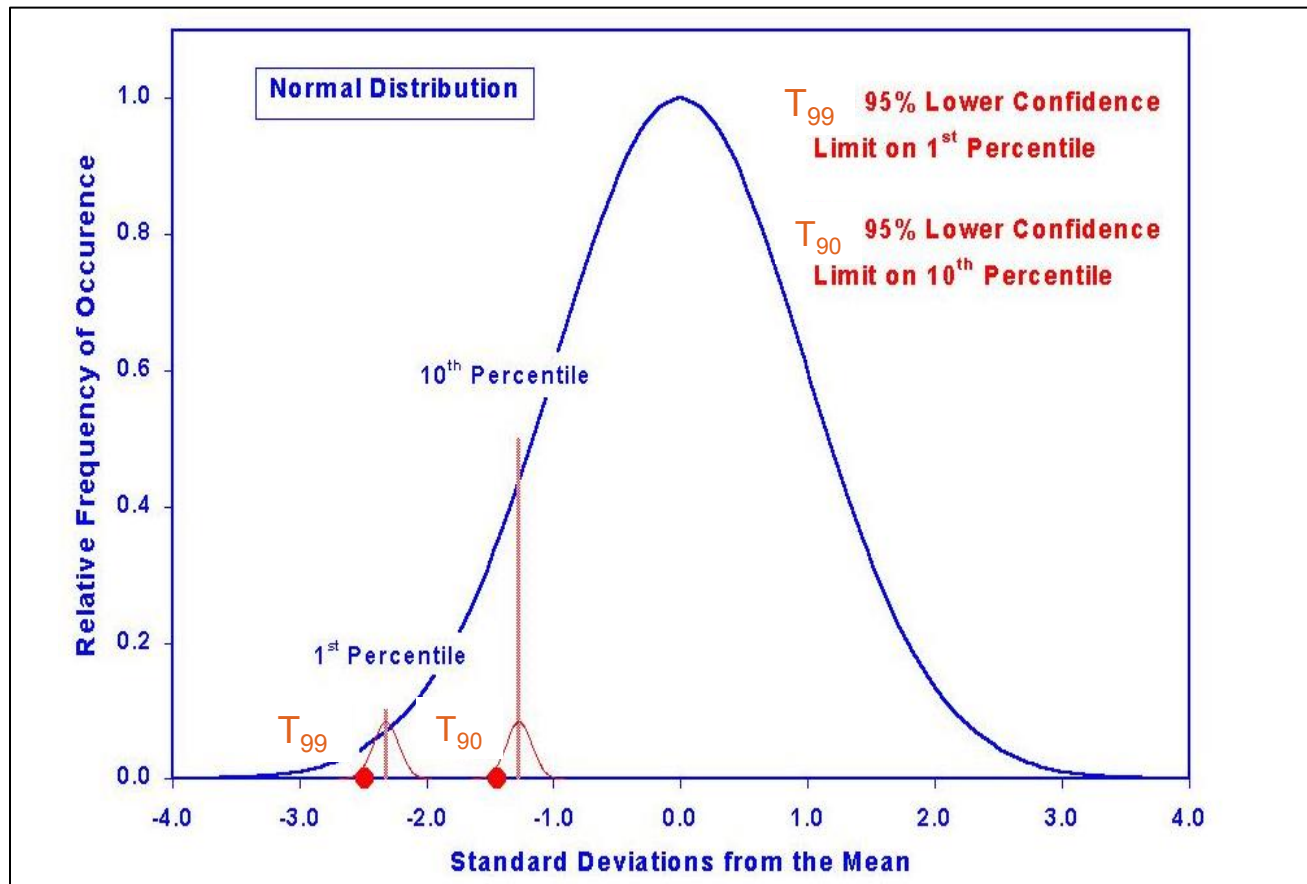
- **A small, random sample of fabric-type components are tested to failure and their breaking strength is recorded**
 - **From this sample, we want to infer a strength that we expect 99% of population of components to exceed with 95% confidence**
- **This limiting strength is known as an A-basis material allowable**
- **Aerospace system margin of safety (risk) is based on the estimated allowable in system design phase.**



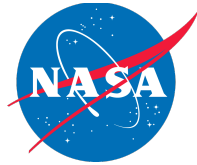


Statistical Design Allowable

- Allowables are routinely used for metallic components, which are more homogenous in their failure modes and breaking strength
- Diagram from the standard materials handbook is provided below



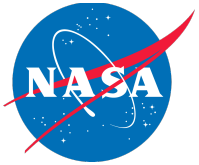
Guidance on Developing Statistical Design Allowables



- **A-basis:** The lower of either a statistically calculated T99 or specification minimum. T99 is the value at which at least 99% of population is expected to equal or exceed with 95% confidence.
- **B-basis:** Same as T90; at least 90% of population is expected to equal or exceed with 95% confidence.

NOTE: A- and B-basis for static design properties are based on minimum population of 100

- **A-basis is a 0.99/95% and B-basis is a 0.90/95% one-sided tolerance interval, although not explicitly stated**
- **For fabric-type components, it's very rare to have 100 test values, usually less than 20 are initially available.**
- **Fabric components tend exhibit multiple modes of failure, more complex than metallic materials, superimposed distributions**
- **Estimating a relatively extreme tail probability, from a small sample is precarious – and, yet it is required...**



Some Statistical Challenges

Small sample sizes for estimating intervals on extreme quantiles

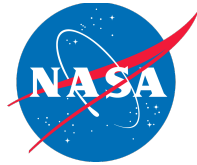
- May involve significant extrapolation, said another way, we're unlikely to see extreme test values in a small sample

Choice of distributional models

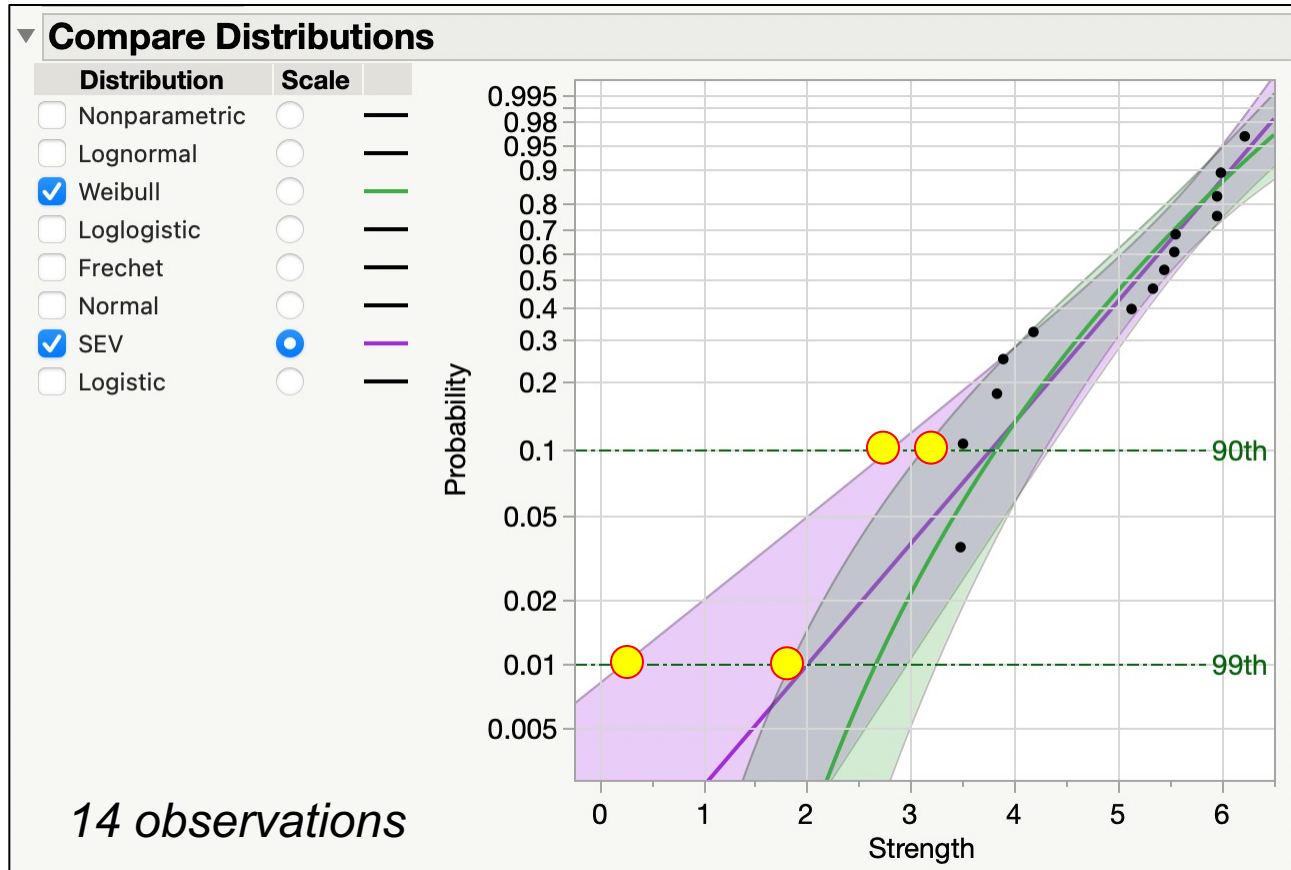
- Results can be very sensitive to the choice of distributional models, particularly when extrapolating
 - Typically interested in one tail, not the entire distribution
- Goodness of fit tests support candidates, but not particularly useful
- Akaike information criterion (AICc) rank candidates, somewhat useful

Available tools for interval estimation

- Custom-built software (spreadsheets) for specific applications that attempt to flow-chart and black-box the calculations – warning!
- Limited statistical software available, particularly for non-normal distributions, with the exception of the R 'tolerance' package

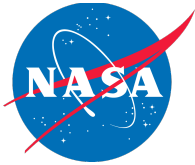


Choosing a Distributional Models

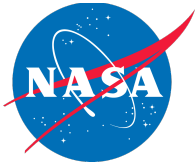


- Which distributional model is a better fit for the left tail at the lower 90th or 99th percentile, for reporting a limiting capability of strength?
- When are we interpolating or extrapolating?
- Conservative choice may not be acceptable or defensible

Tips to Explain Tolerance Intervals



- Start with subject-matter vernacular, appeal to intuition, not statistical theory
 - We're going to test 15 samples that are intended to be representative of strength of all of the flight system components, existing and future
 - We expect variability in breaking strength among our test samples
- Distinguish between “variability” and “uncertainty”
 - If we tested 3 samples, we'd get an idea of the mean (average), but we don't know if we'd observed extreme values, or not, so we're uncertain about the variability in the range of strengths
 - If we tested 20 samples, there would still be inherent variability among the strength values, but our uncertainty would be lower, we're more confident.
- Conceptually, to estimate the tolerance interval, i.e. the allowable
 - From our sample, we estimate the mean (average) and standard deviation, and derive a tolerance factor that is a function of the sample size and a distributional model to estimate the allowable strength



Concluding Remarks

- **The need to infer a limiting quantity from a small sample of test data is commonplace in aerospace research and development**
- **Estimating a statistical tolerance interval is often the most appropriate approach, but they are not well known or understood**
- **Software tools are limited to estimate tolerance intervals, which makes them more difficult to implement in practice**
- **Explaining and interpreting a tolerance interval to subject-matter experts who rely on them for vital decisions regarding a system's margin of performance is often the more challenging than the statistical theory**