Uncertainty in Test and Measurement

Some History from an

Applied Statistics Viewpoint

CIRMS 2006

Neil R. Ullman, P.E.

Ullman Associates

Overview

- Statistical Approaches Evolve
  - First formal appearance in 1700’s
  - Industrial applications – 1930’s
  - Active period of the 1960’s
  - Interlaboratory approaches introduced, especially in ASTM
- Metrological Needs
  - The GUM appears
  - ASTM addresses the GUM
- Some Comparison and Contrasts

Early Ideas of “Quality”

Workmanship meant that the closeness of any result depended on the skill and training of the crafter or scientist:

- A good carpenter or machinist could always do superior consistent work
- Astronomer - training could produce a better observer

- But this concept was being challenged

Thomas Simpson's - 1755

- One practical problem involved observing the time a star past a crosshair in a telescope or similar problem
- Paper Presented to Royal Society
  - On the Advantage of Taking the Mean of a Number of Observations, in Practical Astronomy

Excerpts

- **It is well known to your Lordship, that the method practised by astronomers, in order to diminish the errors arising from the imperfections of instruments, and of the organs of sense, by taking the Mean of several observations, has not been so generally received, but that some persons of considerable note, have been of opinion, and even publickly maintained, that one single observation, taken with due care, was as much to be relied on as the Mean of a great number.**

The Principle

- **Taking an average of several readings comes closer to the "true" value than typical individual readings**
- **This concept was to become the basis for Confidence Intervals**

Thomas Bayes (also 1755)

- **In responding to Simpson's letter, he observed that:**
  - **"…the more observations you make with an imperfect instrument the more certain it seems to be that the error in your conclusion will be proportional to the imperfection of the instrument made use of…"**
- **This introduced the concern for BIAS**

The quite period

- **Much theoretical and applications in behavioral studies but little specific activity related to engineering and science**

Bell System in the 1920's

- Walter Shewhart writes "Economic Control of Quality of Manufactured Product"
  - Here he Introduces the Control Chart
  - Describes the "state of statistical control"
    - in both level AND variability
  - Recognizes Random and Special Causes

Societies take note

- A subcommittee of ASTM Committee E-1 on Method of Test prepared:
  - STP-15 ASTM Manual on Presentation of Data (1933)
  - In 1935 Supplement A was added: “Presenting ± Limits of Uncertainty of an Observed Average”
- Formal use of Term Uncertainty
  - Average ± a σ

STP-15

- **Limitations:**
  - **Only applies to Observations taken under same conditions**
- **Currently ASTM Manual 7**
  - **Confidence Interval expressions**
  - **Extended to examine standard deviations and proportions.**

The Active 60’s

- Important work by NBS and ASTM E11
- Precision Measurement and Calibration, Special Publication 300 NBS contains numerous important papers such as:
  - Eisenhart - Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems

NBS Handbook 91

- Natrella, “Experimental Statistics” 1963
- Chapter titled “Expression of Uncertainties of Final Results”
  - Uncertainty of a reported value is indicated by giving “credible limits to its likely inaccuracy.”
  - BUT stressed “No single form of expression for these limits is universally satisfactory.”

ASTM E177

- Currently “Practice for Use of the Terms Precision and Bias in ASTM Test Methods”
  - Originally “…precision and accuracy as applied to measurement of a property of a material”
- Precision and imprecision
- Bias
- Accuracy
- Uncertainty

ASTM Consensus Standards

- WHERE TO APPLY THESE PRINCIPLES
- ***guide*****, an organized collection of information or series of options that does not recommend a specific course of action.**
- ***practice*****, a set of instructions for performing one or more specific operations that** ***does not produce a test result*****.**
- ***test method,*** **a definitive procedure that** ***produces a test result*****.**
- **Also Specifications, classifications, terminology**

Test Result

- **The Value of a Characteristic obtained by carrying out a specified test method**
  - **Both ASTM E2282 and ISO 3534-2 (ISO TC69)**
  - **NOTE The test method specifies that one or a number of individual observations be made, and their average or another appropriate function (such as the median or standard deviation) be reported as the test result. It can also require standard corrections to be applied, such as correction of gas volumes to standard temperature and pressure. Thus a test result can be a result calculated from several observed values. In the simple case, the test result is the observed value itself.**
  - **Differs from a Measurement Result (VIM)**

Interlaboratory Testing

- In this period of the 60’s and into the 70’s NBS and ASTM were developing important approaches to interlaboratory testing recognizing ways to examine
  - Within Laboratory variation
  - Between Laboratory variation
- ASTM committees wrote several standards
  - In 1979 Committee E11 prepared E691

Form & Style of ASTM Standards

- **Test Methods - Section A21**
  - **"Every test method shall contain a statement *(1)* regarding the precision of test results obtained in the same laboratory under specifically defined conditions of within-laboratory variability (repeatability conditions), and *(2)* regarding the precision of test results obtained in different laboratories (reproducibility conditions)."**
  - **"Precision shall be estimated in accordance with the interlaboratory test program prescribed in ASTM Practice E 691, for Conducting an Interlaboratory Study to Determine the Precision of a Test Method … [or an equivalent method]"**
  - **Statement on Bias is also mandatory**

Purpose of Repeatability & Reproducibility

- **Informs the user of TYPICAL Variation to be anticipated**
- **Repeatability suggests a “best case” - smallest variation in short term**
- **Reproducibility captures Typical laboratories with different operators, equipment, training, etc., etc.**
  - **Also usually is affected by systematic “relative biases between different labs**

Caveats

- ILS programs
  - Usually only a small number of labs
  - Often hand-picked and data specially treated
  - Desire to throw out "outliers" which minimizes reality
  - Sometimes when results are "poor" the ILS is ignored
- What about Proficiency Programs
  - Labs pay to participate
  - Wide range of actual labs using the method
  - Usually the results are not made a part of standards

Metrological Concerns

- **In late 1970's international metrological organizations (CIPM and BIPM)**
  - **Identified need for measurement uncertainty**
  - **Primary concern was comparing national standards in international studies**
  - **May be at the limits of ability to measure**

An Alternate version of Uncertainty emerges

- **By 1994 - The Guide to the Expression of Uncertainty in Measurement (GUM) was completed and adopted by a number of international organizations**
- **In addition:**
  - **NIST published Tech Note 1297 requiring all NIST measurements to contain uncertainty statements.**
  - **ISO 17025 General Requirement for the competence of testing and calibration laboratories spelled out need for uncertainty**
- **These have led to conflict especially between the test and measurement communities**

ASTM Becomes Engaged

- **In 1994 ASTM Committee E11 became aware through inquiries about substituting uncertainty for P&B**
- **Initial Activities Included:**
  - **Workshops with NIST, Task Groups formed and**
  - **Revisions to ASTM E1488 Guide for statistical procedures to use in Developing and Applying Test Methods -**
    - **Recognized Uncertainty as part of validation of test method**

Contrast in Definitions of Uncertainty

- **ASTM**
  - **E456 (as early as 1972)**
- **An *indication* of the variability associated with a measured value that takes into account two major components of error: (1) bias, and (2) the random error attributed to the imprecision of the measurement process.**

- **GUM**
- ***Parameter*, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand**

E456 Definition Discussion

- Quantitative measures of uncertainty generally require descriptive statements of explanation because of differing traditions of usage and because of differing circumstances. For example: (1) the bias and imprecision may both be negligible; (2) the bias may not be negligible while the imprecision is negligible; …

Intermediate Precision

- Between Repeatability and Reproducibility
- Within a SINGLE Lab
    - Same method, material,
- Incorporates changing conditions
    - Different operators
    - Different Measuring equipment, calibrations, …
    - Location within a lab
    - Environmental conditions
    - Over time

ASTM Addresses Policy

- Much internal dialogue between a Task Group of the ASTM Committee on Standards and Committee E11 on Quality and Statistics
- Ultimately resulted in a resolution adopted within the society that acknowledged that Uncertainty differed from the P&B requirement

Measurement Uncertainty

- A22 of ASTM Form & Style (Test Methods)
  - Measurement uncertainty is an estimate of the magnitude of systematic and random measurement errors that may be reported along with the measurement result. An uncertainty statement relates to a particular result obtained in a laboratory carrying out the test method, as opposed to precision and bias statements which are mandatory parts of the method itself and normally derived from an interlaboratory study conducted during development of the test method.

Two approaches to Uncertainty

- GUM
- Bottom Up
  - Define All factors
  - Determine their contributions
  - Add up the components of variation

- ASTM
- Top Down
  - Strictly data driven
  - Subject test method to various conditions
  - Evaluate the test results as one group

Some Pros and Cons

- **GUM**
- **Pros**
  - **Allows all possible factors to be considered**
- **Cons**
  - **Models may not be valid**
  - **Missing factors**
  - **Bias is questionable**
  - **Requires sophisticated math understanding**

- **ASTM**
- **Pros**
  - **Simple extension of control charting**
  - **Can be easy to implement**
- **Cons**
  - **May be difficult to provide control materials**
  - **Introducing new factors takes effort**

ASTM E11 approach

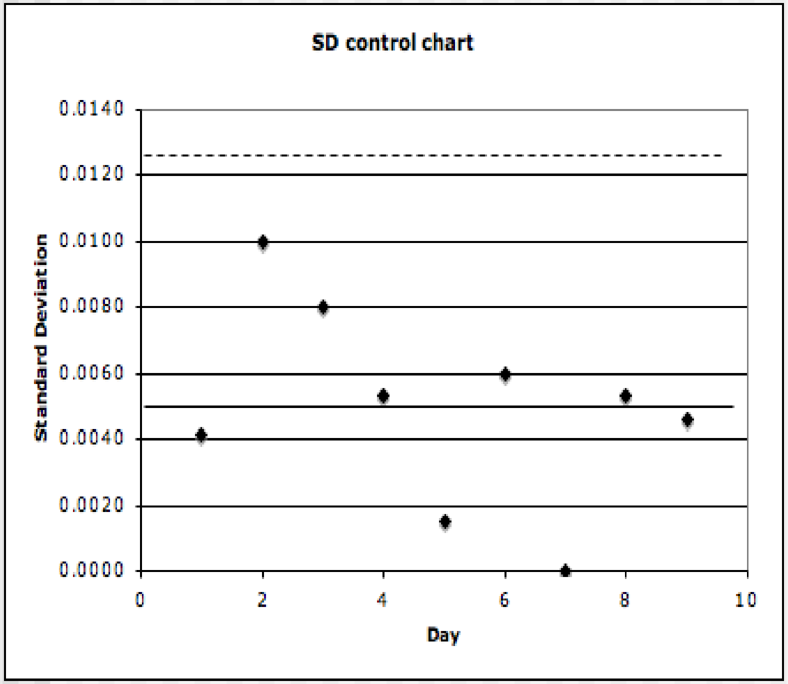
- Use of a Control Chart
  - Shewhart Control charts
    - Use small samples to gauge “repeatability”
    - Use that variability to establish limits for means
  - Eisenhart 1962
    - Measurement is a “manufacturing” process
    - State of COMPLEX Statistical Control
      - Between times may have additional variability
      - But could be stable within some larger limits
  - New standard balloted in Subcommittee E11.20

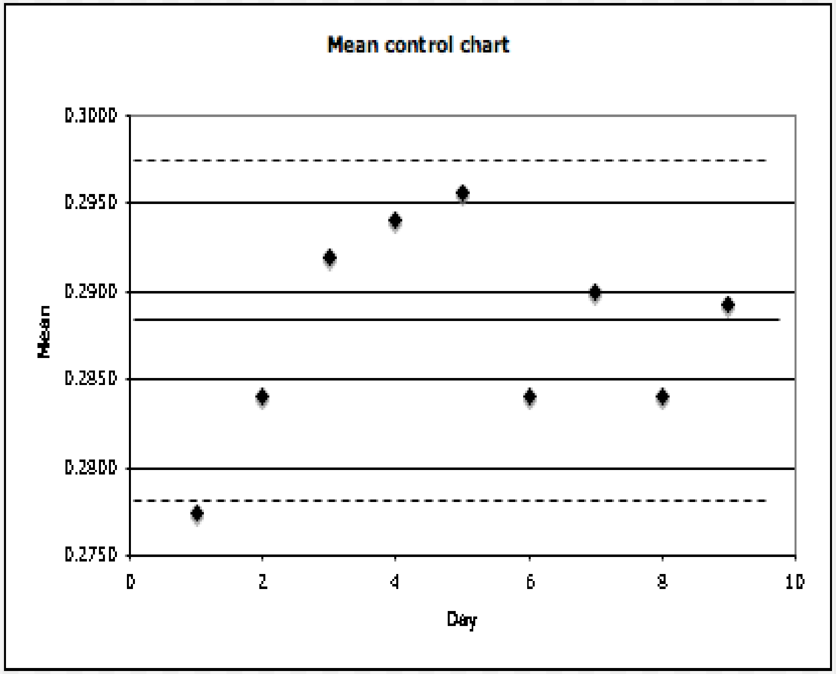
ASTM approach

- **ASTM E51707 2002 (Appendix A2)**
- **Dosimeters irradiated over different days, samples of 3** **(Each day 3 different dosimeters were irradiated)**

Day	Test 1	Test 2	Test 3	Ave	sd
1	0.282	0.274	0.276	0.2773	0.0042
2	0.294	0.274	0.284	0.2840	0.0100
3	0.300	0.284	0.292	0.2920	0.0080
4	0.290	0.300	0.292	0.2940	0.0053
5	0.296	0.294	0.297	0.2957	0.0015
6	0.290	0.278	0.284	0.2840	0.0060
7	0.290	0.290	0.290	0.2900	0.0000
8	0.278	0.288	0.286	0.2840	0.0053
9	0.284	0.292	0.292	0.2893	0.0046

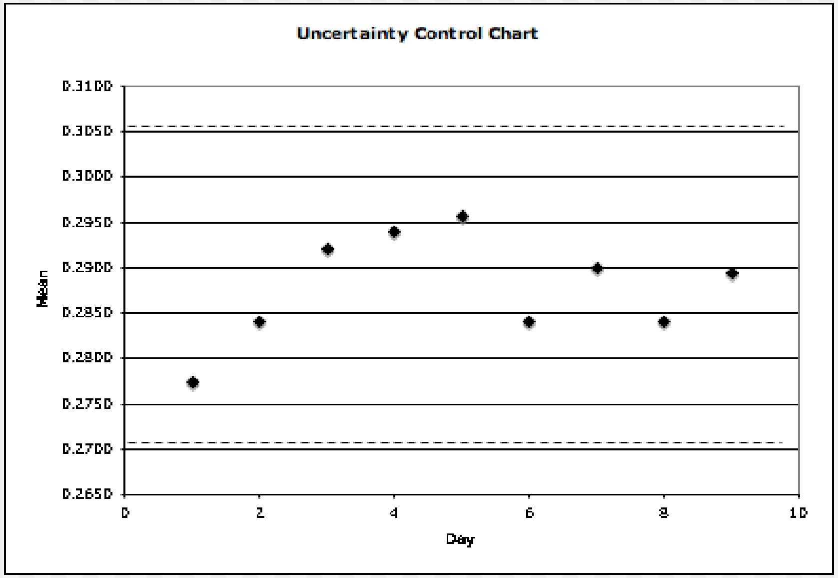
Control Charts





An “Uncertainty” Control Chart

- This is an “initial” chart
  - Used to obtain preliminary estimate of uncertainty sd
  - Can monitor the uncertainty
  - Can revise the estimate



An academic experiment

Torsion Pendulum
Intro to Theory of Error, Beers, 1957

- Gum
- As described in the book:
  - Analyze dimensional measurements of diameters, calculate moment of inertia, period of swing - arrive at contributions of error and combine all using propagation of error methods

- ASTM type approach
- Lab has collection of groups performing the actual experiment
  - Have each group perform the complete exercise several times to establish “final” results
  - Bring results from all of the groups together to obtain overall view of variation between and among lab groups

Conclusion

- **No single approach**
- **Still need for dialogue and development of approaches especially:**
  - **that can be implemented in environments where the users are not sophisticated mathematically**
  - **and/or the test/measurement method is not easy to represent theoretically**