Statistical Intervals Vive La Différence!

William Q. Meeker

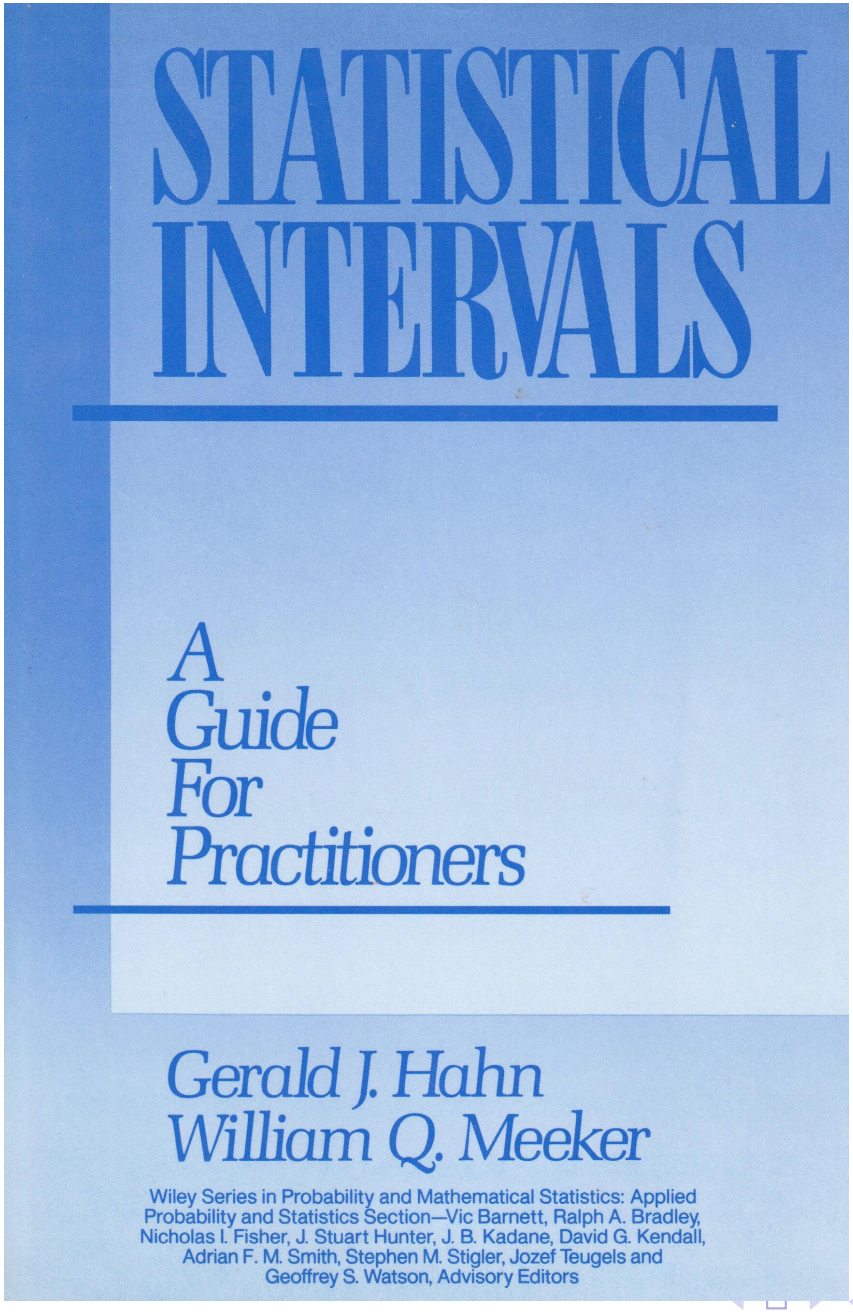
Department of Statistics
Center for Nondestructive Evaluation
Iowa State University
Ames, IA, USA

Quality and Productivity Research Conference
University of Connecticut
Storrs, CT
15 June 2017

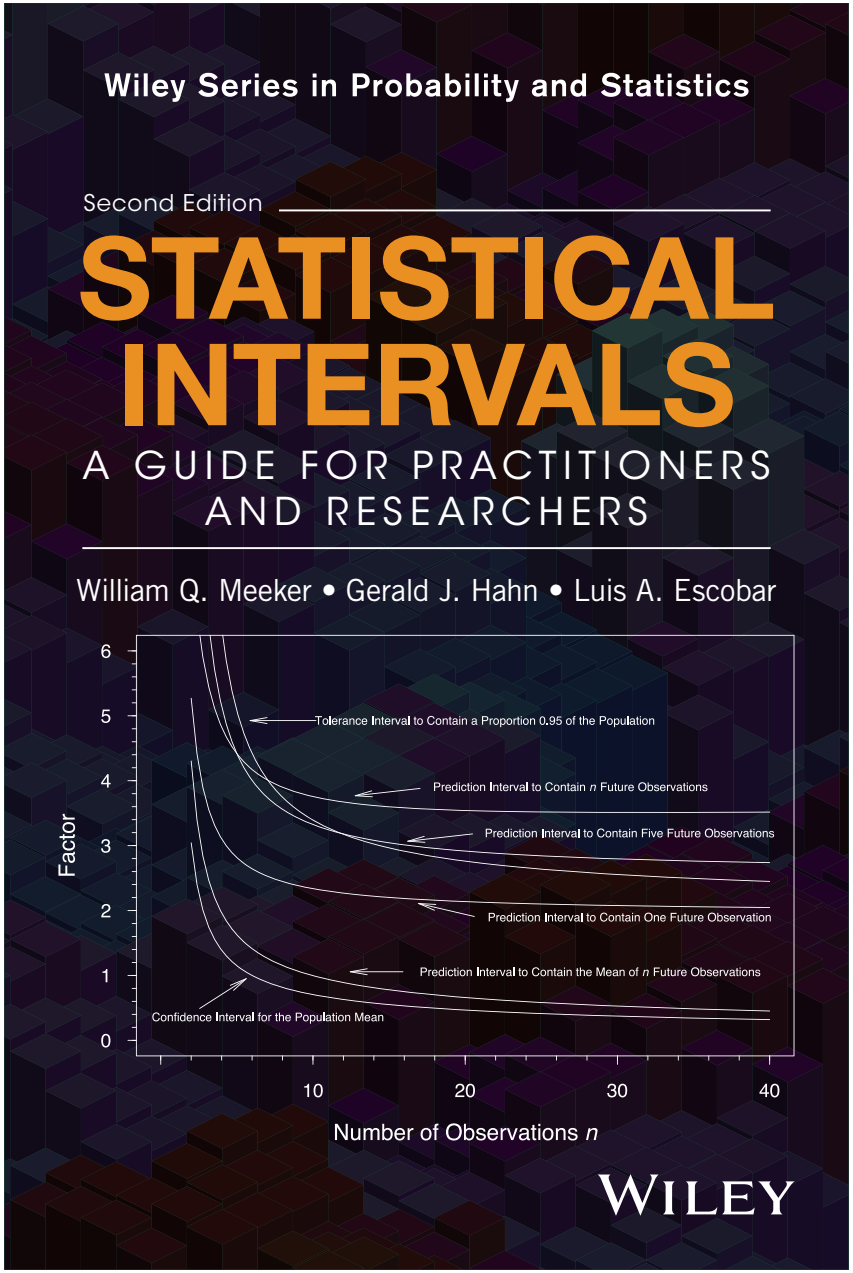
The Different Kinds of Statistical Intervals

- Confidence Intervals
- Tolerance Intervals
- Prediction Intervals

Statistical Intervals Wiley 1991



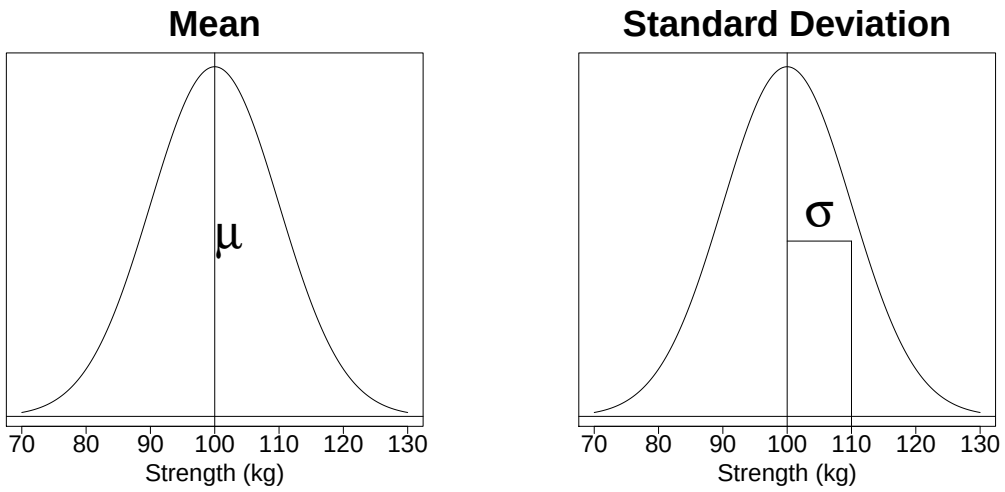
Statistical Intervals Second Edition Wiley 2017



Overview

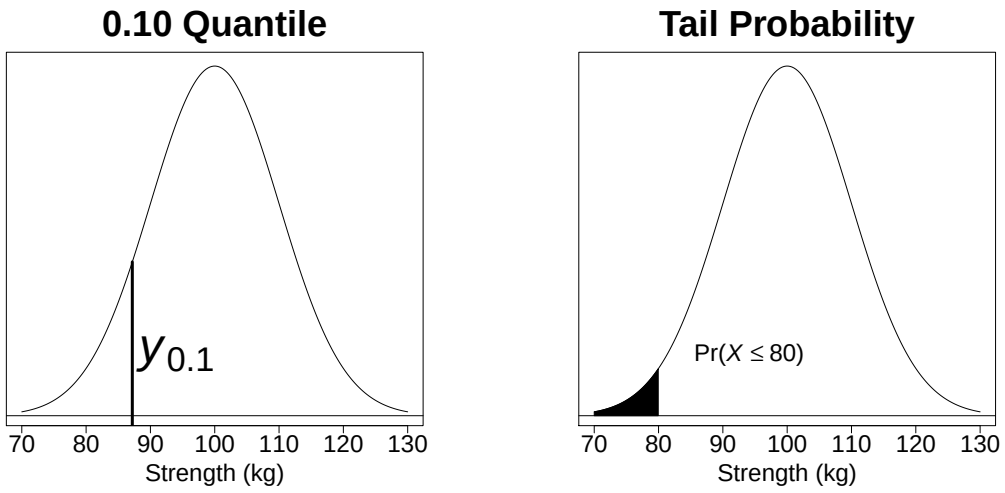
- Confidence Intervals and the “Other Intervals”
- What was in Hahn and Meeker 1991?
- Statistical Intervals: What has changed in the past 25 years?
- Statistical Intervals: Exact, Conservative, and Approximate
- Why did we write *Statistical Intervals*, Second Edition?
- Advances described in *Statistical Intervals*, Second Edition
  - General methods for computing statistical intervals
  - Generalized pivotal quantities
  - Advanced case studies further illustrating general methods
  - Technical appendices
- Predictions about future developments for statistical intervals (Third edition!?)
- Concluding remarks

Confidence Intervals



- Confidence intervals on normal distribution mean and standard deviation routinely covered in elementary text books

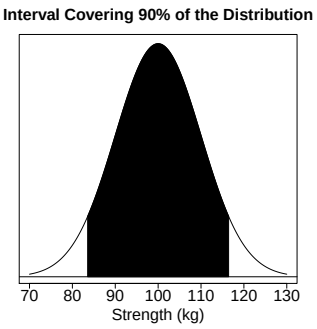
The Other Confidence Intervals



Confidence intervals on **quantiles** and **tail probabilities**

- Frequently needed in applications
- Usually not covered in text books! **Why?**

Tolerance Intervals to Characterize a Distribution



- Tolerance intervals discussed only in a few texts—but frequently needed in practice.
- When distribution parameters are **known**, it is easy to compute a **Probability Interval** that will contain a specified proportion (e.g., $\beta = 0.90$) of the distribution.
- When the parameters are **unknown**, one can compute a statistical **Tolerance Interval** to contain at least a proportion β with $100(1-\alpha)\%$ confidence (e.g., to contain a proportion 0.90 with 95% confidence).
- A **one-sided tolerance bounds** is equivalent to a **one-sided confidence bound** on a distribution **quantile**.

Prediction Intervals

- Prediction Intervals are used to quantify the uncertainty when predicting a **future value of a random variable.**
- Prediction Intervals are well known (e.g., covered in text books) in regression and in time series.
- Prediction Intervals are not so well known in other area of application and infrequently considered in texts, despite their wide applicability. (Consequence: CIs are often incorrectly calculated when PIs are required)
- **Simultaneous** Prediction Intervals to contain k-out-of-m future random variables are sometimes needed.
- When distribution parameters are known, it is easy to compute a **Probability Interval** that will contain a future observation with a specified probability. Otherwise a **statistical** approach is required.

Relationship Between Tolerance and Prediction Intervals

- Tolerance and Prediction Intervals (and when to use them) are **sometimes confused**.
- A k-out-of-m **Simultaneous Prediction Interval** with large m can be approximated by a **tolerance interval** to contain a proportion $\beta = k/m$ of the distribution.
- Tolerance Intervals are appropriate when one wants to describe a distribution.
- Prediction Intervals are appropriate when one wants an interval to contain one or a small number of future random outcomes (e.g., a consumer who buys a single refrigerator).

Outline
Statistical Intervals First Edition

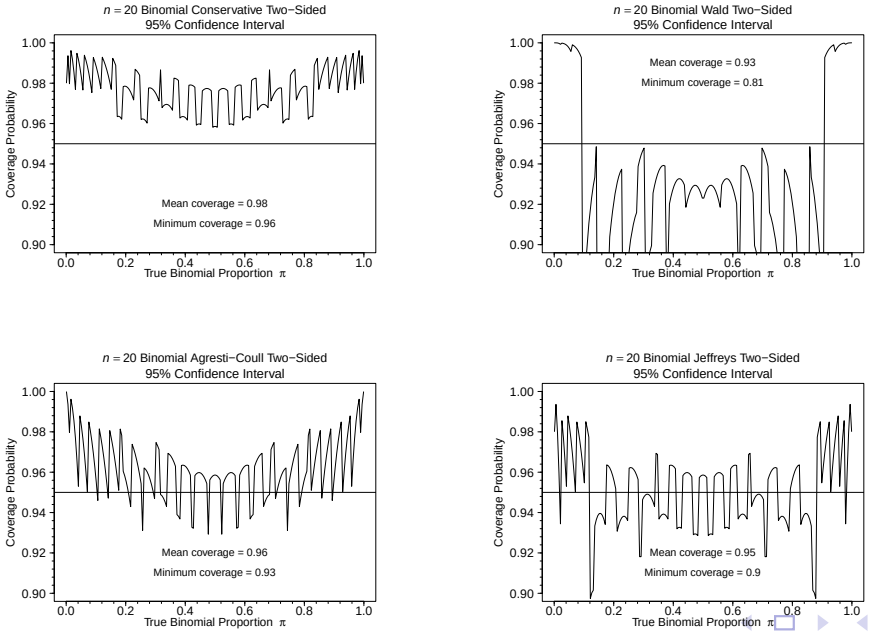
- Background and Assumptions (Chapters 1 and 2)
- Confidence, tolerance, and prediction intervals and examples for
  - Normal distribution (Chapters 3 and 4)
  - Binomial distribution (Chapter 6)
  - Poisson distribution (Chapter 7)
  - Distribution-free intervals (Chapter 5)
- Sample size determination for statistical intervals (Chapters 8-10)
- Basic case studies

Statistical Intervals: What has Changed in the Past 25 Years?

- Recognition that the commonly used text book formulas for Binomial and Poisson distributions confidence intervals are seriously flawed and the development of improved methods
- Wide recognition that likelihood-based intervals are better than Wald (a.k.a., normal-approximation) intervals (and implementation in more commercial statistical software)
- Wider use of bootstrap and simulation-based interval procedures (e.g., general procedures now implemented in JMP Pro)
- A revolution in the use of Bayesian method (and associated intervals) in many practical applications
- Vastly more computational power and ease of doing computations (e.g., using R)

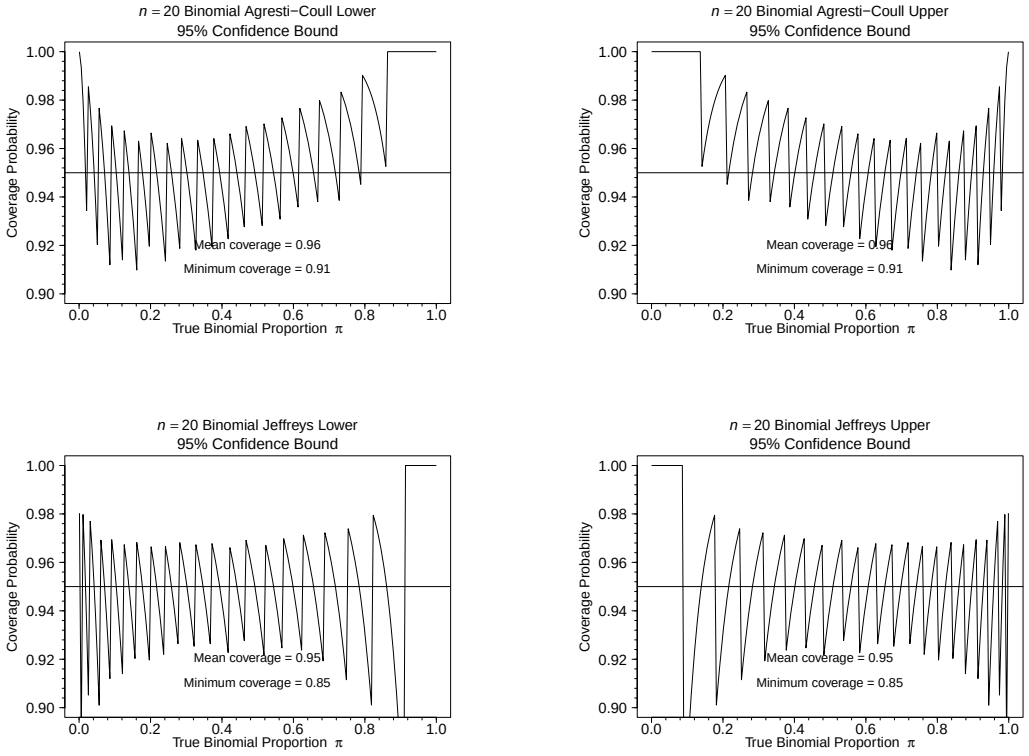
Statistical Intervals: Exact, Conservative, Approximate

- Outside of simple situations, **exact** statistical intervals are usually not available.
- A statistical interval procedure should be evaluated relative to its **coverage probability** (how close it is to the **nominal confidence level using separate evaluations for each tail**) and (secondarily) the expected width or other measure of precision.



Upper/Lower Balance in Error Probabilities is Desirable

Coverage probabilities for lower and upper one-sided confidence bounds



Why Did We Write *Statistical Intervals* Second Edition?

- Bring discussion of statistical interval procedures up to date
- Present **general** methods for computing statistical intervals
- Provide **technical justification** for all intervals (mostly in technical appendices)
- More applications

Advances
Statistical Intervals Second Edition

- New methods for obtaining binomial, Poisson and distribution-free intervals (Chapters 5 to 7 completely re-written)
- Five completely new chapters on **general methods**
  - Likelihood and Wald methods
  - Bootstrap and simulation methods (also pivotal quantities and generalized pivotal quantities)
  - Bayesian methods
  - Hierarchical Models via Bayesian methods
  - Advanced Case Studies and examples
- References and further information are presented in a Bibliographic Notes section at the end of each chapter
- Book is software neutral, but uses R as an advanced calculator to compute certain intervals

Pivotal Quantity (PQ): Confidence Interval for the Mean of a Normal Distribution

The pivotal quantity random variable

$$\frac{\sqrt{n}(\bar{X}-\mu)}{S}$$

has a distribution that does not depend on unknown parameters.

$$\Pr\left[-t_{(1-\alpha/2;n-1)} \leq \frac{\sqrt{n}(\bar{X}-\mu)}{S} \leq t_{(1-\alpha/2;n-1)}\right] = 1-\alpha$$

Solving for μ, gives

$$\Pr\left[\bar{X} - t_{(1-\alpha/2;n-1)})\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{(1-\alpha/2;n-1)}\frac{S}{\sqrt{n}}\right] = 1-\alpha$$

Thus

$$[\underset{\sim}{\mu}, \quad \tilde{\mu}] = \bar{x} \mp t_{(1-\alpha/2;n-1)}\frac{s}{\sqrt{n}}$$

is a $100(1-\alpha)\%$ confidence interval for μ.

Generalized Pivotal Quantities (GPQ)

Given the data, a GPQ Z_θ is a random variable with a distribution that does not depend on unknown parameters (but may depend on the data through MLEs $\widehat{\mu}$ and $\widehat{\sigma}$. For example, GPQs for μ and σ are

$$Z_\mu = \widehat{\mu} + \left(\frac{\mu - \widehat{\mu}^*}{\widehat{\sigma}^*}\right)\widehat{\sigma}$$

$$Z_\sigma = \left(\frac{\sigma}{\widehat{\sigma}^*}\right)\widehat{\sigma}$$

Lower tail probability: $p = F(x) = F(x; \mu, \sigma) = \Phi\left(\frac{x - \mu}{\sigma}\right)$

$$\widehat{p} = \Phi\left(\frac{x - \widehat{\mu}}{\widehat{\sigma}}\right)$$

Then a GPQ for p is

$$Z_p = \Phi\left[\left(\frac{\widehat{\sigma}^*}{\sigma}\right)\Phi^{-1}(\widehat{p}) + \frac{\widehat{\mu}^* - \mu}{\sigma}\right]$$

GPQ Confidence Interval for a Lower Tail Probability

A GPQ for p is

$$Z_p = \Phi\left[\left(\frac{\widehat{\sigma}^*}{\sigma}\right)\Phi^{-1}(\widehat{p}) + \frac{\widehat{\mu}^* - \mu}{\sigma}\right]$$

Simulate a large number (e.g., 100,000) of realizations Z_p^* from the distribution of Z_p. Without loss of generality, can set $\mu = 0$ and $\sigma = 1$ and simulate from

$$Z_p^* = \Phi\left[\widehat{\sigma}^*\Phi^{-1}(\widehat{p}) + \widehat{\mu}^*\right]$$

A $100(1-\alpha)\%$ confidence interval for p is obtained from the $\alpha/2$ and $1-\alpha/2$ quantiles of the empirical distribution of Z_p^*.

Further Comments about GPQs

- Started with work by Tsui and Weerahandi in 1989
- GPQ intervals may give exact interval procedures, but in general provide interval procedures that are asymptotically exact
- GPQ procedures are especially simple for functions of the parameters of location-scale/log-location-scale distributions.
- Some difficult confidence interval problems easily handled by GPQs are
  - Mean of lognormal or Weibull distributions
  - Functions of variance components in GR&R studies
  - Probability of being in a specified interval
  - Two-sample comparisons with different spread or shape parameters
- GPQ inference is related to a generalized form of **Fisher's fiducial inference** (e.g., Hannig, Iyer, and Patterson 2006 and Hannig 2009), providing a theoretical basis for the methods.

Advanced Case Studies Further Illustrating General Methods

- Proportion of defective integrated circuits (LFP model using Wald, likelihood, and bootstrap methods)
- Components of variance in a measurement process (Gauge R&R using Bayes and GPQ)
- Tolerance interval to characterize the distribution of process output in the presence of measurement error (Bayes and GPQ)
- Estimating the proportion of nonconforming product—probability of being between specification limits (Bayes and GPQ)
- Estimating the treatment effect in a marketing campaign (Bayes and simulation)
- Estimating probability of detection with limited hit-miss data (likelihood and Bayes)
- Using prior information to estimate the service-life distribution of a rocket motor (Bayes)

Technical Appendices

- **A.** Notation and Acronyms
- **B.** Generic Definition of Statistical Intervals and Formulas for Computing Coverage Probabilities
- **C.** Important Probability Distributions
- **D.** General Results from Statistical Theory and Some Methods Used to Construct Statistical Intervals
- **E.** Pivotal Methods for Constructing Parametric Statistical Intervals
- **F.** Generalized Pivotal Quantities
- **G.** Distribution Free Intervals Based on Order Statistics
- **H.** Basic Results from Bayesian Inference Models
- **I.** Probability of Successful Demonstration
- **J.** Tables

Predictions for Future Developments and Material for *Statistical Intervals* **Third Edition**

- Additional computing power (but will growth slow?)
- Likelihood methods (perhaps including second-order corrections) will be widely deployed in statistical software
- Bayesian methods will be widely deployed in some advanced statistical software
- Better methods will be developed to specify diffuse prior distributions in Bayesian methods to provide good frequentist properties (**objective Bayes**)
- Continued development of theory relating GPQ methods and generalized fiducial methods
- Possible theoretical unification of Bayesian and non-Bayesian methods

Concluding Remarks

- It is important to quantify statistical uncertainty and statistical intervals provide the best way to do that
- It is critically important that analysts use **appropriate** statistical intervals
- In the use of any statistical inference method, it is important to pay careful attention to **important assumptions** (both those that can be checked and those that cannot)
- Statistical intervals provide much better insight than inference methods like hypothesis testing and *p*-values (e.g., size effect and information about practical significance). See recent ASA statement on the use of *p*-values.
- Improvements in computing power and theoretical developments have provided a clear path to the construction of **appropriate** Statistical Intervals for almost

The End

Thank You