Boehringer Ingelheim
University of Nebraska
Department of Statistics

Evaluating Tolerance Interval Estimates: To Capture or Not to Capture

Michelle Quinlan[1], James Schwenke[2], Walt Stroup[1]

[1]University of Nebraska-Lincoln
[2]Boehringer Ingelheim Pharmaceuticals, Inc.

Outline

- Background: Interval estimates
- What is a tolerance interval?
- Other interval estimates
- Comparing interval estimates
- Simulation results
- Conclusions

Background: Interval Estimates

- Statistical interval estimates constructed to:
  - Estimate parameters
  - Quantify characteristics of population
- Not always clear what each interval is estimating
  - Confidence, prediction intervals are well understood
  - Definition of tolerance interval varies in literature
    - "practical guidelines to calculate and use tolerance intervals in real-world applications are lacking" (*Gryze, et al* 2007)

What is a Tolerance Interval?

- Consistent definition of tolerance interval needs to be established
  - Uses: setting acceptance limits, determining if values lie within specification
  - Interest not in location/spread of distribution but in specific region of distribution
- Many definitions claim TI captures proportion of distribution
  - Not specified whether upper, lower, middle, or asymmetric proportion is captured

What is a Tolerance Interval?

- Definitions include:
  - Bound covering at least (100-α)% of measurements with (100-γ)% confidence (*Walpole & Myers)*
  - At given confidence level, simultaneously for every x, at least *p%* of distribution of Y is contained *(Lieberman & Miller)*
  - Interval including certain percentage of measurements with known probability (*Mendenhall & Sincich)*
    - TI is CI, except it captures a proportion, not a parameter

What is a Tolerance Interval?

- Definitions (cont.):
  - TI with 95% confidence for 90% of distribution implies 95% of intervals cover *at least* 90% of distribution (*Hauck*)
  - TI is extension of PI to multiple responses; where certain percent of responses lie with certain confidence *(Gryze)*
    - 95% PI does *not* contain 95% of population given x
    - As degrees of freedom increase, TI approaches PI

What is a Tolerance Interval?

- Original definition of TI: 2-sided interval estimate on lower/upper *percentile (*not *percentage*) of distribution
  - Formula using non-central t-distribution of percentile:

$$TI = [\bar{X} - (1\text{-}\alpha/2, n\text{-}1, \text{-}\delta))\frac{s}{\sqrt{n}},\ \bar{X} + t(1\text{-}\alpha/2, n\text{-}1, \delta)\frac{s}{\sqrt{n}}] \text{ where } \delta = \Phi(\text{percentile})\sqrt{n}$$

  - *Note*: percentile is value below which exactly 100*p%* of population is located; value below which a random observation lies with probability *p*
  - *Hahn* uses same formula for CI on p^{th} percentile

What is a Tolerance Interval?

- Intervals estimate parameters; parameter estimated by TI is percentile
- 1-sided TI is interval estimate of percentile; interval estimate on percentile defines upper/lower proportion estimated
  - Same cannot be said for 2-sided TI
  - Interval is exact – no need to say “at least” a certain proportion is covered

What is a Tolerance Interval?

- μ and σ are known: TI is exact and calculated directly using area under normal curve
- μ and σ are unknown: TI must account for variability in simultaneously estimating mean, variance
  - Leads to interval containing "at least" a proportion of distribution with certain probability
- As sample size increases, TI approaches interval containing *exactly* proportion *p* of normal distribution

Other Interval Estimates

- *Simultaneous/Non-simultaneous tolerance interval:*
  - TI computed for more than one/single x value
- *Two 1-sided tolerance interval:*
  - 1-sided (1-α)100% TI on lower (1-*p*)/2 with 1-sided (1-α)100% TI on upper (1+*p*)/2
  - Simultaneous 2-sided interval estimate of upper/lower percentiles

Other Interval Estimates

- *β-expectation tolerance interval (Mee's definition):*
  - Interval containing approximately 100β% of distribution:

$$E_{\hat{\mu},\hat{\sigma}_x}\{Pr_X[\hat{\mu} - k\hat{\sigma}_x < X < \hat{\mu} + k\hat{\sigma}_x | \hat{\mu},\hat{\sigma}_x]\} = \beta$$

  - Expected value of proportion covered is β
- β-expectation interval equals PI for single observation
  - (1-α)100% PI to capture future observation equivalent to TI containing on average (1-α)100% of population

Other Interval Estimates

- *β-content tolerance interval:*

$$Pr_{\hat{\mu},\hat{\sigma}_x}\{Pr_X[\hat{\mu} - k\hat{\sigma}_x < X < \hat{\mu} + k\hat{\sigma}_x | \hat{\mu},\hat{\sigma}_x] \geq \beta\} = \gamma$$

  - Interval containing at least 100β% of population with given confidence γ (*Mee*)
    - Computed using Normal, Chi-squared distributions
  - Confidence level for coverage probability of one observation
  - SAS® Proc Capabilities Method 3 computes *approximate* statistical TI containing at least proportion *p* of population

$$\bar{X} \pm z_{\frac{1+p}{2}}(1+1/2n)s^*\sqrt{\frac{n-1}{\chi^2_{\alpha}(n-1)}}$$ [1]

[1] Same formula given by *Hahn* for TI when μ, σ are unknown

Comparing Interval Estimates

- Simulation conducted to determine whether:
  - Confidence, prediction, tolerance (1, 2-sided) intervals capture other parameters than what they are designed to
  - 2-sided PI captures next k random obs., middle proportion of distribution
  - 1-sided PI captures upper(lower) percentile of distribution
  - 2-sided PI for next k random obs. captures middle proportion of distribution

Comparing Interval Estimates

- Simulation conducted to determine whether:
  - 1-sided TI captures next random and next k random obs.
  - Is it always CI < PI < TI?
  - Two 1-sided TI captures middle proportion of distribution (whether Bonferroni adjustment necessary), next random and next k random obs.
  - SAS® Method 3 captures middle proportion of distribution, next random and next k random obs.

Simulation Results

α = 0.05, percentile =.95

- ❑ Is the next random observation captured?

Interval	Simulation 1	2	3	4	5
→ 2-sided PI	0.95	0.95	0.96	0.96	0.95
→ 1-sided lower PI	0.95	0.93	0.96	0.95	0.95
→ 1-sided upper PI	0.94	0.96	0.95	0.96	0.96
2-sided PI for next k random obs.	0.99	0.99	0.99	0.99	1.00
1-sided lower TI	0.98	0.97	0.99	0.97	0.98
1-sided upper TI	0.97	0.98	0.98	0.99	0.98
2 1-sided TI	0.98	0.98	0.98	0.98	0.98
SAS Method 3	0.98	0.98	0.98	0.98	0.98

On average, at least 95% of time the next random observation is captured for all intervals

Simulation Results

α = 0.05, percentile =.95

- ❑ Are the next k = 5 random observations captured?

Interval	Simulation 1	2	3	4	5
2-sided PI	0.77	0.78	0.77	0.80	0.81
→ 2-sided PI for next k random obs.	0.94	0.94	0.95	0.96	0.96
1-sided lower TI	0.90	0.87	0.90	0.90	0.91
1-sided upper TI	0.88	0.91	0.88	0.90	0.91
2 1-sided TI	0.89	0.90	0.90	0.90	0.93
SAS Method 3	0.88	0.89	0.89	0.89	0.92

Most intervals do fairly well except for 2-sided PI

Simulation Results

α = 0.05, percentile =.95

- Is the upper(lower) percentile of the distribution captured?

Interval	Simulation 1	2	3	4	5
1-sided lower PI	0.58	0.56	0.58	0.56	0.56
1-sided upper PI	0.57	0.56	0.60	0.57	0.54
→ 1-sided TI lower	0.97	0.94	0.94	0.95	0.95
→ 1-sided TI upper	0.95	0.95	0.95	0.95	0.96
→ 2-sided TI lower	0.96	0.94	0.95	0.95	0.95
→ 2-sided TI upper	0.95	0.95	0.94	0.95	0.96

1, 2-sided TI perform best

Simulation Results

α = 0.05, percentile =.95

- Is the middle 95% of the distribution captured?

Interval	Simulation 1	2	3	4	5
2-sided PI	0.43	0.38	0.43	0.40	0.38
2-sided PI for next k random obs.	0.99	0.98	0.99	0.98	0.99
2 1-sided TI	0.92	0.91	0.90	0.91	0.92
→ 2 1-sided TI (Bonferroni)	0.96	0.94	0.95	0.95	0.96
SAS Method 3	0.89	0.87	0.87	0.87	0.88

2 1-sided TI (Bonferroni) does best, 2-sided PI for next k random obs. too wide, SAS® Method 3 too narrow

Simulation Results

α = 0.05, percentile =.95

- ❑ Is the middle 95% of the distribution of future obs. captured?

Interval	Simulation 1	2	3	4	5
2-sided PI	0.39	0.35	0.40	0.37	0.35
2-sided PI for next k random obs.	0.99	0.98	0.99	0.98	0.99
2 1-sided TI	0.91	0.89	0.89	0.89	0.91
→ 2 1-sided TI (Bonferroni)	0.95	0.93	0.94	0.95	0.96
SAS Method 3	0.87	0.85	0.85	0.85	0.86

2 1-sided TI (Bonferroni) does best, 2-sided PI for next k random obs. too wide, SAS® Method 3 too narrow

Simulation Results

- ❑ Effect of changing percentile
  - Capturing next random observation:
    - o Coverage decreases for most intervals as percentile decreases
  - Capturing next k random observations:
    - o Coverage decreases significantly for most intervals as percentile decreases
  - Capturing upper(lower) percentile of distribution:
    - o Coverage increases for 1-sided PI as percentile decreases

Simulation Results

- Effect of changing percentile (cont.)
  - Capturing middle *p%* of distribution:
    - Coverage generally improves as percentile decreases
    - Coverage for SAS® Method 3 decreases as percentile decreases
  - Capturing middle *p%* of distribution of future obs.:
    - Similar results for capturing middle *p%* of distribution

Simulation Results

- Effect of changing α-level
  - Results similar for $\alpha = 0.05$; conservative intervals become more conservative as α decreases
- CI < PI < TI for *p* larger than approx. 0.9467728 ($\alpha = 0.05$)
- Two 1-sided TI (Bonferroni correction) performs best for capturing middle *p%* of distribution
  - Interval does not relate directly to specific hypothesis test
  - Useful describing spread of distribution or setting equivalence deltas for assessing parallelism (see *Hauck)*

Conclusions

- Various definitions for TI
  - Inconsistency between percentiles/proportions
  - Lower, middle, upper or asymmetric proportion covered?
- 1-sided TI for percentile (and proportion) constructed using non-central t-distribution
  - Percentile defines proportion; proportion does not define percentile
  - Definitions involving "at least *p%*" incorporate uncertainty when simultaneously estimating μ and σ

Conclusions

- Simulation results
  - 2-sided PI, SAS® Method 3 do not capture middle percent of distribution
  - Two 1-sided TI (Bonferroni correction) covers middle percent of distribution
- Future work:
  - Investigate two 1-sided TI relationship with equivalence testing
  - Simulate models with more than one variance component
  - Research intervals containing *entire* regression line

References

- Boulanger, B., et al. "Statistical Considerations in Analytical Method Validation." SAS book on pre-clinical Statistics.
- Gryze, Steven De, et al. "Using the correct intervals for prediction: A tutorial on tolerance intervals for ordinary least-squares regression." Chemometrics and Intelligent Laboratory Systems. 87 (2007) 147-154.
- Hahn, Gerald J. "Statistical Intervals for a Normal Population, Part I. Tables, Examples and Applications." Journal of Quality Technology. Vol. 2, No. 3, July 1970.
- Hahn, Gerald J. "Statistical Intervals for a Normal Population, Part II. Formulas, Assumptions, Some Derivations." Journal of Quality Technology. Vol. 2, No. 4, Oct. 1970.
- Hauck, Walter W., et al. "Assessing Parallelism Prior to Determining Relative Potency." PDA Journal of Pharmaceutical Science and Technology. March-April 2005, Vol. 59, No. 2, 127-137.
- Kuik, D.J., et al. "Sample Size Computations for Tolerance Region Estimation." Proceedings of the Conference CompStat 2002 Short Communications and Posters.
- Mee, R.W. "β-Expectation and β-Content Tolerance Limits for Balanced One-Way ANOVA Random Model." Technometrics. Aug. 1984, Vol. 26, No. 3.

References

- Mee, R.W. "Estimation of the Percentage of a Normal Distribution Lying Outside a Specified Interval." Commun. Statist.-Theor. Meth., 17(5), 1465-1479 (1988).
- Mee, R.W., and D.B. Owen. "Improved Factors for One-Sided Tolerance Limits for Balanced One-Way ANOVA Random Model." Journal of the American Statistical Association. Dec. 1983, Vol. 78, No. 384.
- Mee, R.W. "Simultaneous Tolerance Intervals for Normal Populations With Common Variance." Technometrics. Feb. 1990, Vol. 32, No. 1.
- Mendenhall, W. and T. Sincich. Statistics for Engineering and the Sciences. 4th Ed. New Jersey: Prentice-Hall, 1995.
- "Methods for Computing Statistical Intervals." SAS Institute Inc. 2004. SAS OnlineDoc® 9.1.3. Cary, NC: SAS Institute Inc.
- Odeh, R.E., et al. "Sample-Size Determination for Two-Sided Expectation Tolerance Intervals for a Normal Distribution." Technometrics. Nov. 1989, Vol. 31, No. 4.
- Odeh, R.E. and D.B. Owen. Tables for Normal Tolerance Limits, Sampling Plans and Screening. New York: Marcel Dekker Inc., 1980.

References

- Owen, D.B. "Control of Percentages in Both Tails of the Normal Distribution." Technometrics. Nov. 1964, Vol. 6, No. 4.
- Patel, J.K. "Tolerance Limits – A Review." Commun. Statist.-Theor. Meth., 15(9), 2719-2762 (1986).
- Rao, J.N.K., et al. "Effect of Non-Normality on Tolerance Limits Which Control Percentages in Both Tails of Normal Distribution." Technometrics. Aug. 1972, Vol. 14, No. 3.
- Satterthwaite, F.E. "An Approximation Distribution of Estimates of Variance Components." Biometrics Bulletin. Dec. 1946, Vol. 2, No. 6.
- Wald, A. and J. Wolfowitz. "Tolerance Limits for a Normal Distribution." The Annuals of Mathematical Statistics. 1946, Vol. 17, No. 2.
- Walpole, R.E. and R.H. Myers. Probability and Statistics for Engineers and Scientists. 5th Ed. New York: Macmillan, 1993.
- Weissberg, A. and G.H. Beatty. "Tables of Tolerance-Limit Factors for Normal Distributions." Technometrics. Nov. 1960, Vol. 2, No. 4.
- Wroughton, Jacqueline. "Techniques and Applications of Interval Estimation." PhD dissertation, 2007.

Acknowledgements

PQRI

Product Quality Research Institute

This research is funded through the

PQRI Stability Shelf Life Working Group

PQRI is the Product Quality Research Institute, Arlington, Virginia