Statistical significance, confidence, uncertainty

Tressa L. Fowler

Accounting for Uncertainty

- Observational
- Model
  - Model parameters
  - Physics
  - Verification scores

- Sampling
  - Verification statistic is a realization of a random process
  - What if the experiment were re-run under identical conditions? Would you get the same answer?

Uncertainty estimates are among a long list of important verification practices

- Well defined questions or goals.
- Large, representative, (identical?) sample.
- Consistent, independent observations.
- Appropriate methods and statistics.
- Uncertainty estimates.
- Spatial, temporal, and conditional differences evaluated.
- User relevant results.
- Thoroughly tested software.

You can’t fix by analysis what you bungled by design. - Light, Singer and Willett.

Define question(s) first.
Then the confidence interval is around the right statistic.

- Which model is best?
- Is my model upgrade an improvement?
- How frequently are ceilings in the correct category?

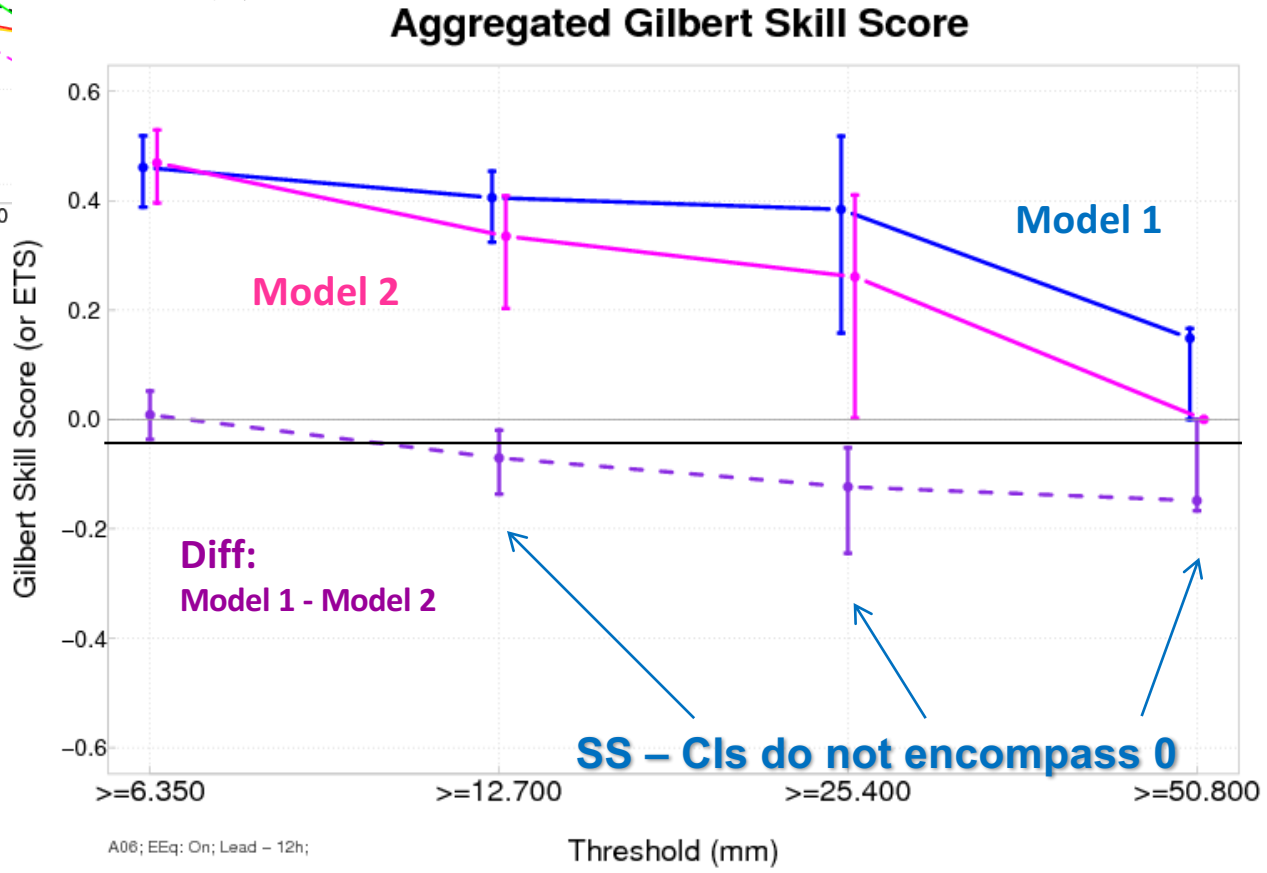
Practical vs. statistical significance

- May not be the same. Why?
  - Failure to use significant figures.
  - Very large sample sizes.
  - Stats assumes independent samples, but weather rarely delivers.
- Which do you need? Both!

Two ways to examine scores

CI about Actual Scores
may be difficult to differentiate model performance differences

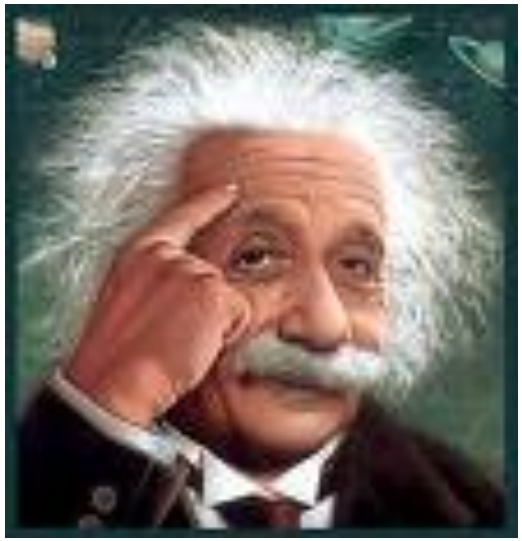
CI about Pairwise Differences
may allow for better differentiation of model performance



Confidence Intervals (CIs)

"If we re-run the experiment N times, and create N (1-α)100% CI's, then *we expect the true value of the parameter to fall inside (1-α)100 of the intervals."*

Confidence intervals can be *parametric* or *non-parametric*…



Types of Confidence Intervals

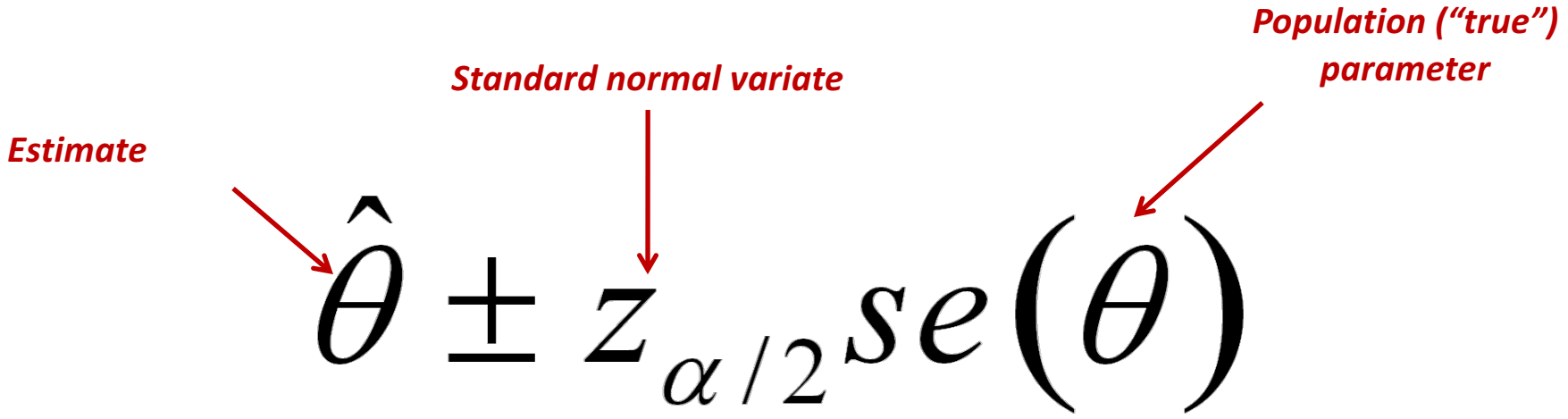
Bootstrap

- Available for almost any statistic.
- More robust to outliers.
- Sensitive to lack of continuity, small samples.

Parametric (normal)

- Sensitive to departures from assumed distribution.
- Often sensitive to outliers.
- Not available for some statistics.

Normal Approximation CI's

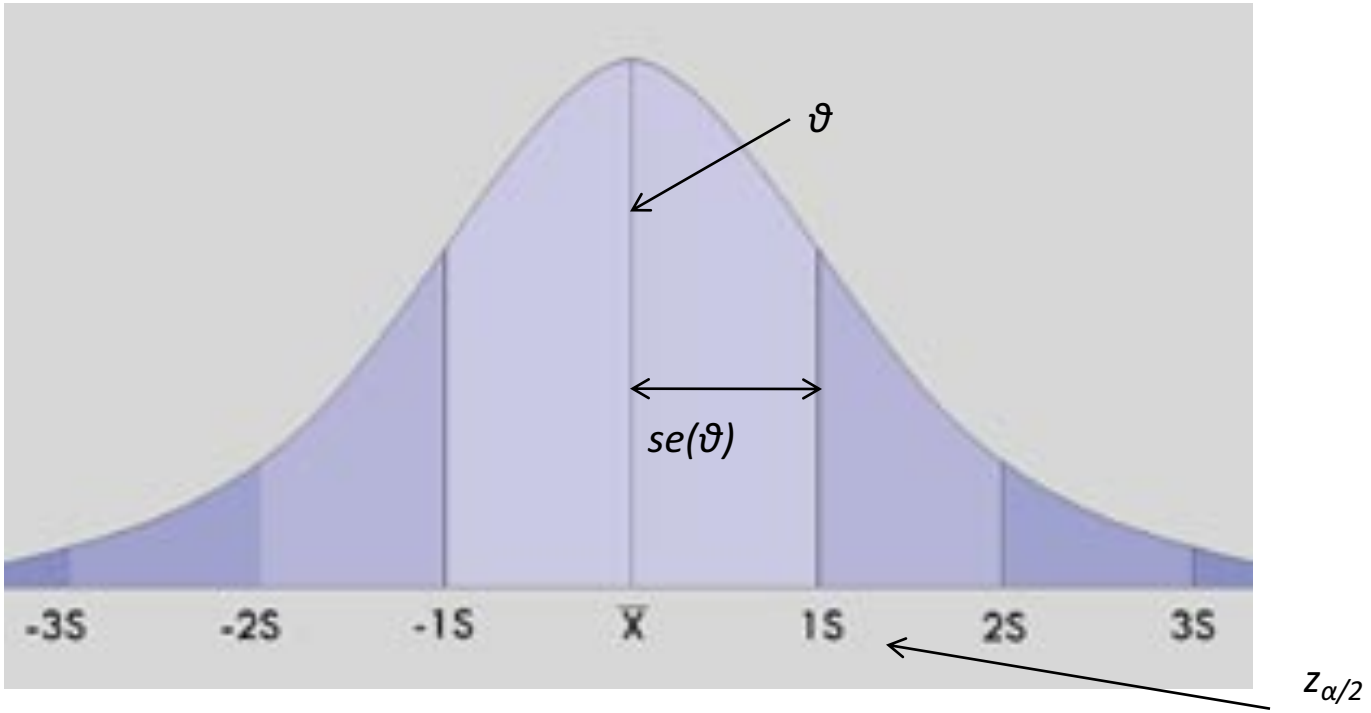


Is a (1-α)100% Normal CI for Θ, where

- Θ is the statistic of interest (e.g., the forecast mean)
- se(Θ) is the standard error for the statistic
- z_v is the v-th quantile of the standard normal distribution where v= α/2.
- A typical value of α is 0.05 so (1-α)100% is refered to as the 95th percentile Normal CI

Normal Approximation CI's

$$\hat{\theta} \pm z_{\alpha/2} se(\theta)$$



Application of Normal Approximation CI's

- **Independence assumption** (i.e., "iid") – temporal and spatial
  - Should check the validity of the independence assumption
  - *MET accounts for first order temporal correlation*
- **Normal distribution assumption**
  - Should check validity of the normal distribution (e.g., qq-plots, other methods)
  - MET does not do this – should be done outside of MET
  - However… MET applies appropriate approaches to verification statistics
- **Multiple testing**
  - When computing many confidence intervals, the true significance levels are affected (reduced) by the number of tests that are done.

Normal Approximation CI's

- Normal approximation is appropriate for numerous verification measures

  Examples: *Mean error, Correlation, ACC, BASER, POD, FAR, CSI*

- Alternative CI estimates are available for other types of variables

  Examples: forecast/observation *variance*, *GSS, HSS, FBIAS, Brier Score*

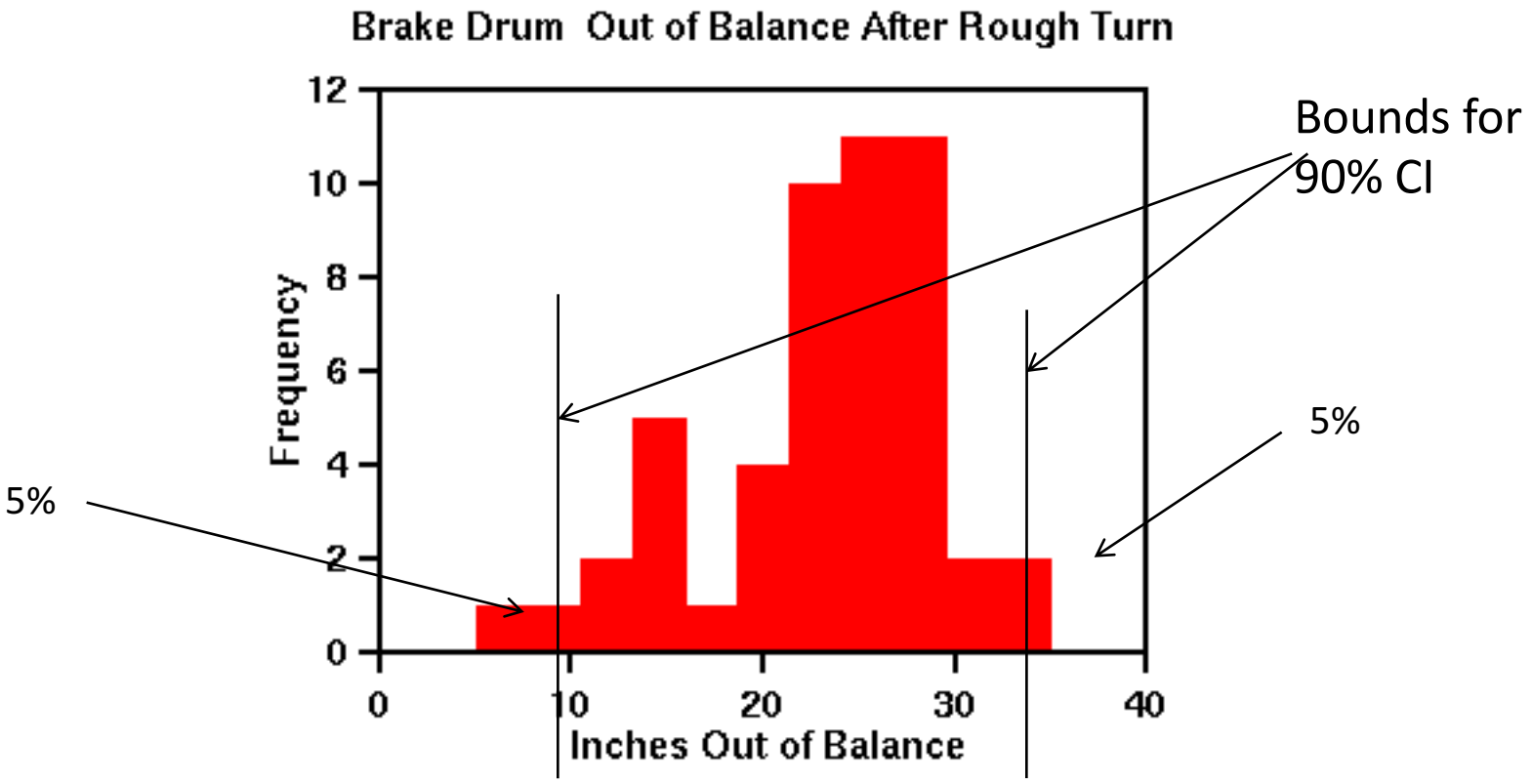
- All approaches expected the sample values to be independent and identically distributed.

(Nonparametric) Bootstrap CI's

IID Bootstrap Algorithm

1. Resample *with replacement* from the sample (forecast and observation pairs), $x_1, x_2, \ldots, x_n$
2. Calculate the verification statistic(s) of interest from the resample in step 1.
3. Repeat steps 1 and 2 many times, say B times, to obtain a sample of the verification statistic(s) ϑ_B .
4. Estimate (1-α)100% CI's from the sample in step 3.

Empirical Distribution (Histogram) of statistic calculated on repeated samples



Bootstrap CI Considerations

- Number of points impacts speed of bootstrap
  - Grid-based typically uses more points than Point-based
  - THUS: Bootstrap is quicker with Point-based
- Number of resamples impacts speed of bootstrap
  - Recommended value is 1000
  - If you need to reduce – try to determine where solutions converge to pick your value
- Bootstrap can be disabled in MET, if concerned about compute speed - check status in config file before running

METViewer alternatives

- Two types of parametric intervals available where appropriate.
  - Accumulate scores (e.g. overall average), find parametric interval.
  - Summarize scores (e.g. find average or median value of all daily POD values), find interval appropriate for average or median.
- Bootstrap the *statistics* for each field over time.
  - Measures (between-field) uncertainty of the estimates over time, rather than the within field uncertainty.
- Pairwise difference statistics and intervals (with event equalization).
  - Gives more power to detect differences by eliminating case to case variability.

Conclusions

- Uncertainty estimates are an essential part of good verification evaluations.
- All estimates are wrong, some estimates are useful.
- MET and METViewer developers strive to provide the most correct and useful intervals for output statistics.

References and further reading

- Gilleland, E., 2010: Confidence intervals for forecast verification. NCAR Technical Note NCAR/TN-479+STR, 71pp. *Available at:* https://opensky.ucar.edu/islandora/object/technotes%3A491
- Jolliffe and Stephenson (2011): Forecast verification: A practitioner's guide, 2nd Edition, Wiley & sons
- JWGFVR (2009): Recommendation on verification of precipitation forecasts. WMO/TD report, no.1485 WWRP 2009-1
- Nurmi (2003): Recommendations on the verification of local weather forecasts. ECMWF Technical Memorandum, no. 430
- Wilks (2012): Statistical methods in the atmospheric sciences, ch. 7. Academic Press

See also

http://www.cawcr.gov.au/projects/verification/

Appendix C of MET Documentation:
http://www.dtcenter.org/met/users/docs/overview.php