2014-05-13

Measures to Ensure Assessment Consistency

Kawash, Jalal

Taylor Institute Teaching Community

Department of Computer Science

Ensuring Assessment Consistency

Dr. Jalal Kawash and Dr. Robert Collier
May 13, 2014

Ensuring Assessment Consistency

- CPSC203; Introduction to Problem Solving
- Targeted at Undergraduate Non-Majors*
- Large, Multi-Section Course

Number of Students

min: 197 (W2014)	mean: 526 (st. dev.: 161)	max: 814 (F2008)

Number of Tutorial Sections

min: 12 (W2013)	mean: 25 (st. dev.: 6.4)	max: 35 (F2008)

Ensuring Assessment Consistency

- Unify Lecture Topics across Sections
- Sections Teach the Same Material at the Same Pace
- Sections Use/Provide the Same Resources
- Requires Co-operation between Instructors
- Necessary Step for Standardizing Other Materials

Ensuring Assessment Consistency

- Unify Assignment Specifications across Sections

- Large Cohorts often entail Overlapping Submissions
- Can Obscure cases of Academic Misconduct

- Open-Ended Assignments Reintroduce Variation
- Requires a Framework for Valid Submissions
- Can Complicate Evaluation and Marking

Ensuring Assessment Consistency

- Unify Examination Specifications across Sections

- Mutiple Sections entail Several Exam Periods
- Can Complicate Invigilation

- Randomly Generated Initial Data Sets are Useful
- Equivalently Difficult Operations can be Exchanged
- Can Complicate Evaluation

Ensuring Assessment Consistency

- (Operationally) Define Inter-Rater Reliability:

  "The Degree to which a Set of Evaluators Agree upon the Mark that should be Assigned"

- Often require Evaluators mark the Same Materials
- Inefficient Use of Resources (in this Context)

- "Spot-Checking" Techniques can be Useful* (**assuming experienced evaluators are available*)

Ensuring Assessment Consistency

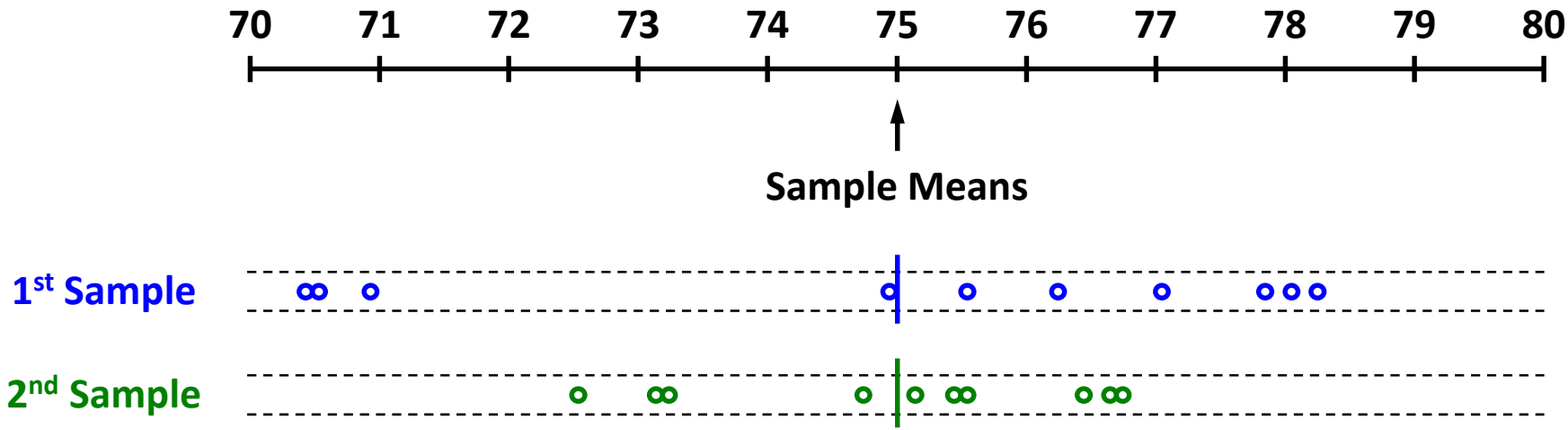
- Assessment Consistency → Detection and Response

- Response is Often as Simple as Remarking
- Extensive Retraining is Rarely Necessary
- Experienced Evaluators are an Asset

- Detection is Far More Challenging
- Outlying Data Points must be Carefully Considered
- Several Simple Statistical Techniques are Available

Ensuring Assessment Consistency

- Comparison of Means is Insufficient
- Variance in Evaluation Criteria is Obscured



- Mean w/ Standard Deviation is Minimum Required

Ensuring Assessment Consistency

- (Operationally) Define Standard Deviation:

"The Square Root of the Mean of the Squared Deviation of Each Grade from the Mean"

- Squaring Each Deviation Exaggerates Differences
- Measures Spread (and Suggests Evalutor Problems)

- Confidence Intervals are Easier to Interpret
- 95% Confidence → Estimate $\pm$ 2 * Standard Error

Ensuring Assessment Consistency

- (Operationally) Define 95% Confidence Interval:

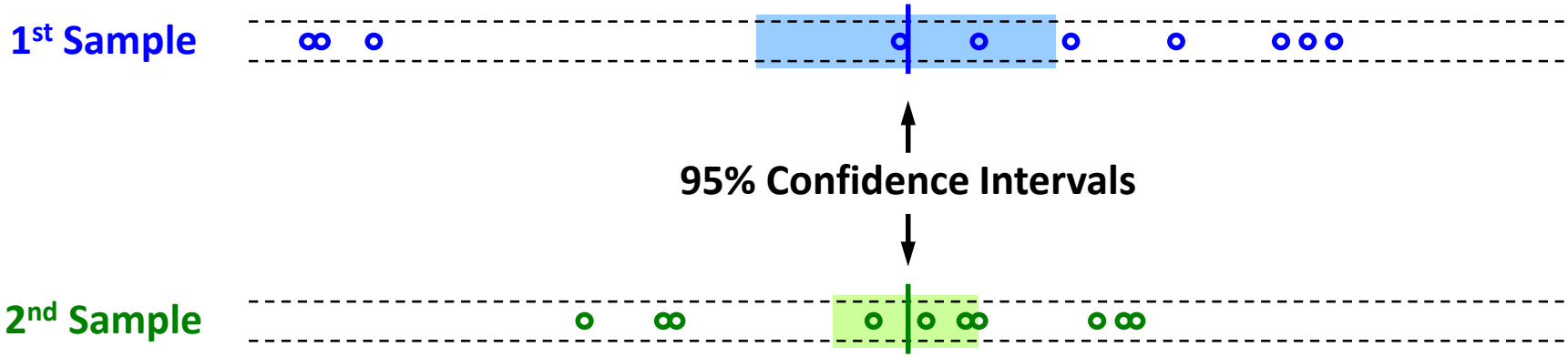
"The Probability that the True Mean lies within the Specified Range around the Sample Mean is 0.95"

- If Each Evaluator Marked Infinite Submissions, Then the Average Grades Assigned should be Equal

- Disjoint Intervals Suggest Significant Differences (n.b., overlapping intervals do not)

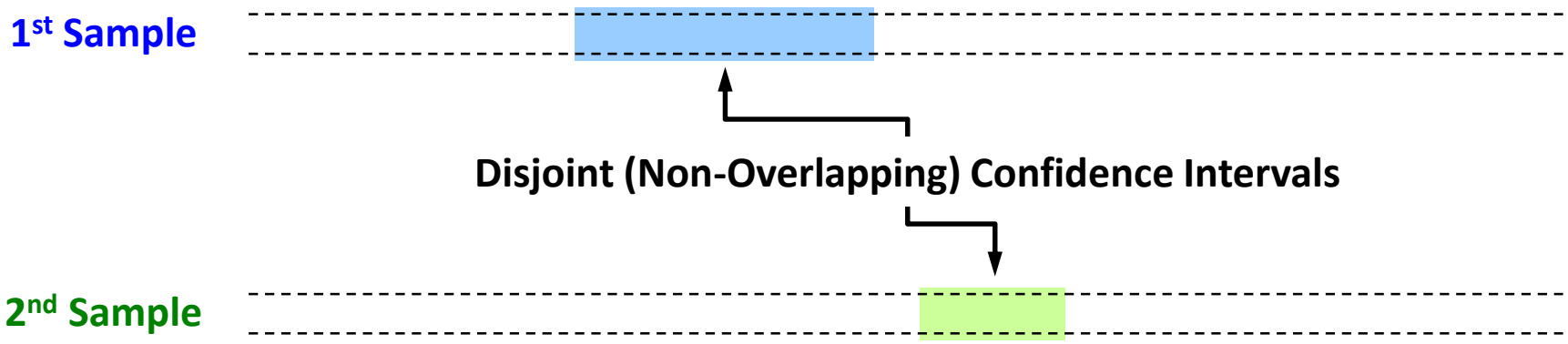
Ensuring Assessment Consistency

- Disjoint Intervals may Indicate Evaluator Problems
- Outlying Data Points may also cause Disjointedness



Ensuring Assessment Consistency

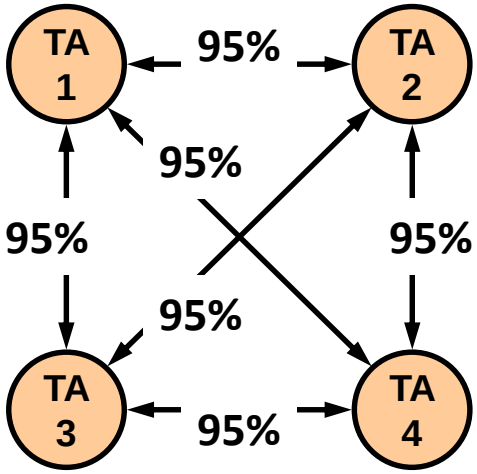
- Disjoint Intervals may Indicate Evaluator Problems
- Outlying Data Points may also cause Disjointedness



- ~90% Confidence Observed Difference is Significant
- ($0.95^2 = 0.9025 \rightarrow$ 90.25% chance intervals correct)

Ensuring Assessment Consistency

- Multiple Comparisons still Reduce Confidence
- 95% → 5% Chance of an Incorrect Conclusion



4 Evalutors → 6 Comparisons

$(0.95)^6 = 0.735$

26.5% Chance of an Incorrect Conclusion

- Applies to Student T-Tests of Significance as well

Ensuring Assessment Consistency

- Confidence Levels can be Adjusted to Correct this
- Bonferroni's Correction is Simple and Effective
- Use Confidence Level of $1.00 - (\alpha / \text{Number of Comparisons})$

For the Example with 6 Comparisons...

$\alpha = 0.0500 \rightarrow (0.9500)^6 \approx 0.735$ 26.5% Error

$\alpha = 0.0083... \rightarrow (0.9917)^6 \approx 0.951$ 4.9% Error

Activity 1 of 5

List the Most Significant Obstacles to Maintaining Assessment Consistency

Activity 2 of 5

Group these Obstacles into Categories and Name Each Category

Activity 3 of 5

Rank these Categories from Most Important to Least Important

Activity 4 of 5

Propose Solution Techniques to Address the Most Important Categories

Activity 5 of 5

Evaluate the Proposed Solutions

Select the Best Solution for Each Category