A Statistical Sampling Methodology for eSCM-SP Assessments

David Northcutt
IBM

Mark Paulk
Carnegie Mellon University
IT Services Qualification Center

The Problem

- Performing assessments in large complex organizations quickly becomes cost prohibitive if every component (entity) must be assessed for each criterion
- Sampling Methodology Goals:
  - Reduce the number of entities that must be assessed, especially for assessments with a large number of entities
  - Keep the risk of incorrectly certifying an organization that is not really compliant at or below 10% under most circumstances
Assumptions

- The ability to detect a single entity that is not compliant was not a design goal
- Incorrectly passing a single Practice, while undesirable, is not as serious as incorrectly passing an organization at any given Level
- Every sampled entity that is not compliant will be detected correctly (“perfect teams”)

Sampling Method

- Uses a variation of Stratified Random Sampling – entities are placed in similar strata by Practice (84 Practices in eSCM-SP)
- Stratifying at the Practice level allows us to differentiate while still achieving maximum efficiency for Practices that are implemented across large parts of the organization (e.g., HR, Contracting, etc.)
Example Stratification for 63 Entities

- 63 Entities
- 28 Practices
- 17 Practices
- 39 Practices

Sampling

The sampling uses a table that was designed to have a 70% or better chance of detecting at least one noncompliant entity within a Practice, given that 2 or more noncompliant entities exist in that Practice.
Underlying Method

- Detecting failures at an eSCM-SP Level is a function of Practice detection at that Level
- Detection probabilities are based on independent combinatorial models
- The table is ultimately designed so that a sample would have at least a 90% probability of detecting at least one noncompliant practice for a Level, if there are two or more noncompliant practices at that Level

Failed Level Detection

Because a single failed practice is sufficient to fail at any Level, detecting Level failure is the same as detecting Practice failure.
Typical Results

- With a full census, the earlier example would have required 5,292 Entity-Practice pairs to be reviewed.
- With the sampling plan, only 2,307 need to be reviewed – a 56.4% reduction.

Larger Populations

- For $N \geq 50$, the sample size was chosen so that there was no worse than a 90% chance of detection when 5% of the entities were noncompliant.
- At this point, the sample size is near constant, regardless of the number of entities.
Conclusions

- It is possible to significantly reduce the effort required to evaluate organizations with multiple entities while still controlling for the number of false positives
- While the reduction is significant in relative terms, the remaining effort for large organizations is still substantial in absolute terms

Future Work

- Investigate whether the reduced set of entity/Practice pairs really is feasible in practice—it may still be too big for large organizations
- Investigate further reductions by relaxing the independence assumptions—there are many related practices in eSCM-SP, and it may be possible to exploit this to make further reductions