Synthesizing Multiple Evaluative Statements into a Summative Evaluative Conclusion

Cristian Gugiu & Nadini Persaud

April 5, 2006
The ABC Project

The purpose of the evaluation

- Determine the merit, worth, and significance of the ABC College.
- The evaluation was commissioned by the principal of the College.

Timeframe

- Evaluation activities started in January 2005 and continue into the present. Close to 2000 hours have been invested by the co-principal investigators.
- All work has been completed pro bono.
Evaluative Framework

- Value-driven evaluation
  - Five key features that distinguish between evaluation and research.
    - Values
    - Standards
    - Meaningful significance
    - Data synthesis
    - Summative confidence

- Evaluation approaches
  - Collaborative evaluation
  - Goal-free evaluation
  - Needs assessment
  - Summative evaluation
Methodology

Data sources

- Eight surveys were administered to
  - 291 Students (response rate 100%)
  - 28 Instructors (17 lecturers and 11 tutors, response rate 60%)
  - 2 Administrators (response rate 100%)
  - 7 Librarians (response rate 100%)
  - 3 Office staff (response rate 100%)
  - 4 Janitors (response rate 100%)
  - 5 Security guards (response rate 100%)
  - 29 Key stakeholders
- Records (financial records, student records, exam results, legislative, newspapers)
- Interviews with key informants
- Site visit
- The Internet
Identifying relevant values

- **Values**
  - “Evaluative statements consist of fact and value claims intertwined” (House & Howe, 1999).
  - Needs assessment and analysis of qualitative data are an excellent source for identifying relevant values.
  - **Student Survey**

- **Analyzing qualitative data**
  - Development of a relational database
  - Development of a qualitative coding scheme

- **Rating the importance of values**
  - In order to develop a scoring rubric that reflects the values of the stakeholders, it is important to gather data
  - **Values Survey**
Obtaining appropriate standards

- **Standards**
  - Standards refer to the level of performance that demarks acceptable and unacceptable (or excellent and less than excellent) performance for a value.
  
  - Three types of standards
    - Minimum bars
    - High bars
    - Holistic bars
  
  - Setting appropriate bars
    - [Standards Survey](#)

Quantifying qualitative data

- Our 8 surveys generated a total of 1,197 statements from the open-ended questions.
- These responses were then coded into a total of 2,212 categories. 68% Students, 16% Instructors, 5% Librarians, 4% Security guards, 3% Principal, 2% Office staff, and 2% Janitors.
- Calculate inter-rater reliability
  - For binary data, this means a phi coefficient or an interclass correlation.
  - Model type (I, II, III) and ICC
  - We calculated ICCs for Model 1. Our correlations rarely fell below 0.70.
Determining success and failure

One cannot determine whether the evaluand has “passed” or “failed” by simply comparing means. Consider the following example.

Comparing performance to standards

- Minimum Bar
- Performance

Unacceptable
Acceptable

0 20 40 60 80 100
Low High
Comparing performance to standards

- **Standard error of the mean**
  - The $SE_m$ is an estimate of the population mean that would be observed if data were repeatedly sampled from a population and means were calculated for each of the samples taken.
  - Obviously, one cannot repeatedly sample the entire population. Fortunately the $SE_m$ can be estimated by the formula

$$SE_m = \sqrt{\frac{\sigma^2}{N}}$$

$$SE_m = \sqrt{\frac{PQ}{N}}$$
Asymmetric Confidence Intervals

There is no reason to expect that CIs should be symmetric and when dealing with proportions, they are not (except for when the proportion \((p) = 0.50\)).

Asymmetric CIs can be calculated using the following formula provided by Hays (1994):

\[
\frac{N}{N+z^2} \left( P + \frac{z^2}{2N} \pm z \sqrt{\frac{PQ}{N} + \frac{z^2}{4N^2}} \right)
\]
Comparing performance to standards

- **Finite Population Correction factor**
  - The variance of the mean must be adjusted whenever one samples from a finite population where your Sample size $N$ is 5% or more of the total population size $T$.
  - The correction factor $(T-N)/T$ is applied to the variance component in your model. So, in the case of the asymmetric CI it looks like this:

\[
\frac{N}{N + z^2} \left( P + \frac{z^2}{2N} \right) \pm z \sqrt{\left( \frac{T - N}{T} \right) \frac{PQ}{N} + \frac{z^2}{4N^2}}
\]
Comparing performance to standards

Accounting for inter-rater (un)reliability

Although Nadini and I were able to attain fairly high inter-rater reliability, the unreliability within the data, will nevertheless, cause our CIs to expand. The question is by how much?

Unfortunately, I could not locate a formula in the literature. However, I decided to write my own based onto principles.

1. The CIs should expand as a result of “adding” more uncertainty.
2. The CI should be unaffected when $\rho = 1$.
3. Likewise, as $\rho$ approaches Zero, the CI should expand toward $\pm\infty$ (or %100 in the case of proportions).
4. The CI is a function of $\sigma^2$ and $\rho$. More specifically, it is proportional to $\sigma^2$ and $\rho$.

Therefore, I believe the $\text{Var}(\sigma^2 / \rho) = \sigma^2 / \rho$. 
Creating a composite measure

- **Accounting for variance**
  - Whenever a composite measure is created the variance is affected. Because we are interested knowing the standard error of the mean, we need to know how to handle the variance.
  - Case 1: Independent terms
    
    \[
    \text{Var}(\sigma_1 + \sigma_2) = \sigma_1 + \sigma_2
    \]
  - Case 2: Dependent terms
    
    \[
    \text{Var}(\sigma_1 + \sigma_2) = \sigma_1 + \sigma_2 + 2\rho_{12}\sigma_1\sigma_2
    \]
Putting it all together

- Summative Conclusions for ABC
- The case for Summative Confidence
Synthesizing Multiple Evaluative Statements into a Summative Evaluative Conclusion

Cristian Gugiu & Nadini Persaud

April 5, 2006