A Comparison of Approximate Bayesian Bootstrap and Weighted Sequential Hot Deck for Multiple Imputation

Darryl V. Creel
RTI International
Why do this presentation?

- Better Understand Performance of Weighted Sequential Hot Deck Alone and in Comparison with Approximate Bayesian Bootstrap Using the Multiple Imputation Variance Estimator (Rubin 1987)
- Address Andridge and Little (2010) Comment: “The weighted sequential hot deck does not appear to have been widely implemented.”
- Graphical Presentation of Monte Carlo Results (which may more easily interpretable than a numeric table)
Outline

- Missing Data
- Approximate Bayesian Bootstrap
- Weighted Sequential Hot Deck
- Monte Carlo Simulation
- Results
Missing Data

- What to do about missing data?
  - Ignore (generally a bad idea)
  - Weight (unit nonresponse)
  - Impute (item nonresponse)
- Multiple Imputation (Rubin, 1987)

\[
\hat{V} = \hat{W} + \left( \frac{M + 1}{M} \right) \hat{B}
\]
Approximate Bayesian Bootstrap

- Approximate Bayesian Bootstrap (Rubin and Schenker, 1986)
- Let \( r \) be the Number of Respondents
- Let \( m \) be the Number of Nonrespondents
- Procedure within an imputation class
  - Select \( r \) Units With Replacement from the Respondents to Create the Donor Pool (Potential Donors)
  - Select \( m \) Units With Replacement from the Donor Pool to be Actual Donors
  - Repeat \( b \) number of times
Kim’s Adjustment

- Kim (2002) investigates ABB and shows multiple imputation variance estimator has a downward bias that “is not negligible for moderate sample sizes.”
- He proposes to reduce the size of the donor pool to minimize bias.
Parzen, Lipsitz, and Fitzmaurice’s Adjustment

- Parzen, Lipsitz, and Fitzmaurice (2005) reviewed Kim’s (2002) paper and suggested “an alternative to reducing bias via a simple correction factor applied to the standard multiple imputation variance estimate.”
- More Easily Implemented
- More Efficient (less variability in the variance estimates)

- $n_r$ is the number of item respondents (5)
- $w_h$ is the sample weight for the $h^{th}$ respondent
- $n_m$ is the number of item nonrespondents (3)
- $s_i$ is the scaled weight for the $i^{th}$ nonrespondent
Weighted Sequential Hot Deck

- Same Donor Pool
  - Repeat WSHD $b$ Times on Respondents
  - Not Proper Multiple Imputation
- Bootstrap Donor Pool
  - Essentially the First With Replacement Sample in the ABB Process
  - ABB: With Replacement Sample to Create Donor Pool, With Replacement Sample to Select Donors, Repeat $b$ Times
  - WSHD: With Replacement Sample to Create Donor Pool, WSHD to Select Donors, Repeat $b$ Times
- WSHD Implemented using SUDAAN® (RTI International 2008)
Monte Carlo Simulation

- Extends Kim’s (2002) which was used by PLF (ignorable nonresponse within one imputation class. Also used by Demirtas, Arguelles, Chung, and Hedeker 2007)
- 2 Sample Sizes
  - 20
  - 100
- 2 Distributions of the Analytic Variable
  - Normal with mean 5 and variance 1
  - Chi-Square with 5 degrees of freedom
- 3 Response Rates
  - 40%
  - 60%
  - 80%
- 2 Values for the Number of Multiple Imputations
  - 3
  - 10
Monte Carlo Simulation

- Each Combination of Factors 10,000 Replications
- 5 Imputation Methods
  - ABB
  - Kim – modifies donor pool size
  - PLF – variance correction factor
  - WSHD – same donor pool
  - WSHDB – bootstrap to create donor pool
- Comparison
  - Relative Bias of the Variance Estimators
Relative Bias of the Variance Estimates

Normal, n = 20, RR 40%

Number of Imputations

Relative Bias, 100(V_hat - S)/S

3 10

-70 -50 -30 -10 10

Normal, n = 20, RR 80%

Number of Imputations

Relative Bias, 100(V_hat - S)/S

3 10

-70 -50 -30 -10 10

Normal, n = 100, RR 40%

Number of Imputations

Relative Bias, 100(V_hat - S)/S

3 10

-70 -50 -30 -10 10

Normal, n = 100, RR 80%

Number of Imputations

Relative Bias, 100(V_hat - S)/S

3 10

-70 -50 -30 -10 10

Legend

ABB
Kim
PLF
WSHDB
WSHD
Relative Bias of the Variance Estimates

ChiSq, n = 20, RR 40%

ChiSq, n = 20, RR 80%

ChiSq, n = 100, RR 40%

ChiSq, n = 100, RR 80%

Legend

ABB
Kim
PLF
WSHDB
WSHD
Future Work

- Additional Empirical Investigations
  - More Complex Ignorable Nonresponse
- WSHD Multiple Imputation
References

More Information

Darryl V. Creel
Senior Statistician
301.770.8229
dcreel@rti.org