Weighting Large Datasets with Complex Sampling Designs: Choosing the Appropriate Variance Estimation Method

Sara Mann
University of Guelph, smann@uoguelph.ca

James Chowhan
McMaster University, chowhan@mcmaster.ca

Weighting Large Datasets with Complex Sampling Designs:
Choosing the Appropriate Variance Estimation Method

Sara Mann
University of Guelph,
Guelph, Ontario Canada

James Chowhan
McMaster University,
Hamilton, Ontario Canada

Using the Canadian Workplace and Employee Survey (WES), three variance estimation methods for weighting large datasets with complex sampling designs are compared: simple final weighting, standard bootstrapping and mean bootstrapping. Using a logit analysis, it is shown - depending on which weighting method is used - different predictor variables are significant. The potential lack of independence inherent in a multi-stage cluster sample design, as in the WES, results in a downward bias in the variance when conducting statistical inference (using the simple final weight), which in turn results in increased Type I errors. Bootstrap methods can account for the survey's design and adjust the variance so that it is inference appropriate and corrected for downward bias. The WES provides mean, as opposed to standard, bootstrap weights with the data; thus, a further adjustment to account for the reduced variation inherent when information is grouped is required. Failure to use mean bootstrap weights appropriately leads to biased standard errors and inappropriate inference.

Key words: Bootstrap, variance estimation, complex sampling design.

Introduction

Choosing the appropriate variance estimation method when weighting large datasets with complex sampling designs has important implications for researchers. Using the Canadian Workplace and Employee Survey (WES), three variance estimation methods for weighting large datasets with complex sampling designs are compared: simple final weight, standard boot-strapping and mean bootstrapping. This study

uses a logit analysis to show that, depending on which weighting method was used, different predictor variables are significant. Failure to use the mean bootstrap weights appropriately can lead to both biased standard errors and inappropriate inference.

Survey Instrument

A national workplace survey, the WES, conducted over a five year period from 1999-2003 by Statistics Canada was used in this study. Stratified sampling was used for WES and up to twenty four employees were surveyed within each workplace, depending on the establishment's size. In-person interviewers collected the workplace survey data and telephone interviews were conducted with the employees. The WES is unique in that employers and employees are linked at the micro data level and employees are selected from within sampled workplaces (Statistics Canada, 2003). The number of employers included in the sample was 6,322 in 1999, 6,068 in 2000, 6,207 in 2001, 5,818 in 2002 and 6,565 in 2003. The number of employees in the sample was 23,540 in 1999, 20,167 in 2000, 20,352 in 2001, 16,813

Sara Mann is an Assistant Professor of Strategic Human Resource Management and Organizational Behavior in the Department of Business. Her research interests include HRM practices such as performance appraisal and selection, specifically for lower wage workers and those in the agriculture sector. Email her at: smann@uoguelph.ca. James Chowhan is a Ph.D. Candidate in the DeGroote School of Business at McMaster University, and has a Master's degree in Economics. Email him at: chowhan@mcmaster.ca.

in 2002 and 20,834 in 2003 (with a survey response rate of 95% for locations and 83% of workers in 1999). Employers are followed for five years and employees are followed for two years.

The WES uses a multi-stage cluster design to select a sample of respondents. This results in respondents being sampled from the same cluster implying that they are not necessarily independent. This potential lack of independence results in a downward bias in the variance when conducting statistical inference (using the simple final weight). Downward bias results in an increase of Type I errors, rejecting the null when it is true. Bootstrap methods can account for the survey's design and adjust the variance so that it is inference appropriate and corrected for the downward bias. The WES provides mean bootstrap weights with the data, as opposed to standard bootstrap weights; thus, a further adjustment to account for the reduced variation inherent when information that is grouped is required.

The results presented in this article are from a study conducted by Mann & Latham (2008), which examined the predictors of the receipt of a performance appraisal. The variables that were significant predictors differed depending on which variance estimation method was used.

Variables

Five organization-level predictor variables were included in the analysis: size (operationalized as the number of employees), industry (service = 1, manufacturing = 0), for-profit (for-profit = 1, not-for-profit = 0), unionized (yes=1, no=0) and an in-house HR department (yes=1, no=0).

Several job-level predictor variables were also included in the analysis: hourly wage, four dummy variables representing whether the job is full-time or part-time and permanent or temporary (full-time/permanent, full-time/temporary, part-time/permanent, part-time/temporary), dummy variables representing occupation (professional, manager, technical/trades, marketing, clerical and production) and a dichotomous variable measuring the use of a computer in the job (yes=1, no=0).

Individual-level variables were also included in the analysis: age, gender (1=female, 0=male), recent immigrant (within the last 5 years) (1=yes, 0=no), and disability (1=yes, 0=no). One dependent variable was used, the receipt of a performance appraisal (1=yes, 0=no). To reduce common method bias, the organization variables were drawn from the employer survey, while the job- and individual-level variables, as well as the dependent variables were drawn from the employee survey.

Methodology

Descriptive statistics and correlations were presented in the Mann & Latham (2008) study but, because they are not relevant to this study, they are not discussed. The stepwise logit regression that was conducted (with the organization-level variables included in the first step, and the job-level and individual-level variables added in the next two steps) is of particular interest to this study. Three different regressions were conducted using different types of variance estimation methods: simple final weighting, standard bootstrapping and mean bootstrapping. The choice of method has important implications for the inference of the significance of the predictor variables.

Mean Bootstrap Comparison

Bootstrap weights are used to make use of complex survey design information and to calculate reliable design-based variance estimates. Generally, Statistics Canada uses a multistage, stratified, randomly selected cluster sample or complex design to draw a representative sample of respondents. Within the WES, workplaces from business locations operating in Canada are selected from relatively homogeneous strata (industry, region and size groupings). In addition, employers must have paid employees; exceptions are locations in the Yukon, Nunavut, the Northwest Territories, agricultural operations, private households, religious organizations or public administration. This results in respondents not necessarily being independent – respondents in the same cluster may share similar economic characteristics as a group relative to the population as a whole. This is a disadvantage of cluster sampling and results

in less efficient estimates (Satin & Shastry, 1993).

To enable researchers to correct for this downward bias in the variance, Statistics Canada has included bootstrap weights with the WES. Bootstrap techniques have been used to generate a set of 5,000 bootstrap weights, which give a more reasonable estimate of variance than estimation that does not account for the complex design of the survey. Statistics Canada generates bootstrap weights by randomly drawing samples with replacement from each stratum of primary sampling units. The size of each sample drawn is equal to the sample size of the data set. Using the same clustering and sample design the weights are assigned to each unit in the selected random draws; selected units receive a positive bootstrap weight and units not selected receive a weight of zero.

For WES this sampling is replicated 5,000 times to generate a set of bootstrap weights large enough to be consistent and allow for the calculation of average bootstrap weights. Further, for the WES data, the bootstrap weights that have been provided are average bootstrap weights. In other words, a set of 100 (B) average bootstrap weights have been calculated over groups of 50 (C) from the original set of 5,000 bootstrap weights. Average bootstrap weights were calculated to preserve the confidentiality of workplace's responses. Using the WES mean bootstrap weights requires a further adjustment to account for the reduced variation inherent when using grouped information. The variance estimator used to calculate the design-based variance estimate with mean bootstrap weights is:

$$v_{\bar{B}}\left(\hat{\theta}\right)=\frac{C}{B}\sum_b\left(\hat{\theta}^*_{(b)}-\hat{\theta}^*_{(.)}\right)^2$$

where

$$\hat{\theta}^*_{(.)}=\left(\frac{1}{B}\right)\sum_b\hat{\theta}^*_{(b)}. \quad (1)$$

Each bth average bootstrap sample set of weights is equal to the means of C bootstrap weights. In this specification, the term $\hat{\theta}^*_{(b)}$ is obtained using the bth mean bootstrap weight variable (Buckley & Chowhan, 2005).

The analysis herein used Stata 9, and specifically, the survey suite of commands to estimate the results. The advantage of these commands is that the final weight (used to generate the point estimates or parameters), the bootstrap weights (used to generate the standard errors), and the variance estimation method (balanced repeated replication) can all be specified using the svyset command. When each piece of analysis is run the adjustment for the mean bootstrap is made by using the fay(.85857864376269) option. This adjustment comes from the Fay's variance estimator, where K could be set equal to

$$K=1-\sqrt{\frac{1}{C}},$$

which is a transformation of equation 1, given that the Fay's variance estimator is as follows

$$v_{Fay}\left(\hat{\theta}\right)=\frac{1}{T(1-K)^2}\sum_t\left(\hat{\theta}^*_{(t)}-\hat{\theta}^*_{(.)}\right)^2$$

where

$$\hat{\theta}^*_{(.)}=\left(\frac{1}{T}\right)\sum_t\hat{\theta}^*_{(t)}.$$

The use of this adjustment re-introduces the variability that had been removed when the average bootstrap weights were generated.

Results

Descriptive statistics are presented in Table 1 and the predictors of the receipt of a performance appraisal are presented in Table 2. For a discussion on the predictors of the receipt of a performance paper, see the Mann & Latham (2008) paper; this study is only concerned with the variance estimation method used to produce the results.

Comparing the three variance estimation methods in Table 2 different predictor variables were significant depending on which method was used. The results under column 3 should be used as they present the findings from the analysis using the mean bootstrapping method and produce the most accurate, unbiased standard errors. Due to this finding, this article adds a significant methodological contribution to

Table 1: Means and Standard Deviations

	n	Mean	SD
Organization-Level Variables			
Industry	20834	.67	.47
In-house HR Dept	20362	.39	.49
Number of Employees	20362	414.23	1085.35
For-Profit	20362	.80	.40
Unionized	20834	.26	.44
Job-Level Variables			
Wage	20619	20.60	12.75
Full-time/perm	20619	.59	.49
Full-time/temp	20619	.02	.13
Part-time/perm	20619	.34	.47
Part-time/temp	20619	.05	.22
Use Computer	20834	.65	.48
Occupation			
Professional	20834	.16	.37
Manager	20834	.13	.33
Technical/Trade	20834	.41	.49
Marketing	20834	.08	.27
Clerical	20834	.15	.35
Production	20834	.07	.25
Individual-Level Variables			
Gender	20834	.53	.50
Age	20834	40.24	11.52
Recent Immigrant	20834	.03	.15
Disability	20834	.09	.29
Dependent Variables			
Receipt of PA	20834	.60	.49

our field by comparing these different weighting methods.

Although the coefficients are the same for all three methods presented in Table 2, which predictors are significant differs depends on which method is presented. Three related approaches are shown: column (1) shows significance levels when only the final weight is used to generate the standard errors, column (2) illustrates how the standard errors are downward bias when the mean bootstrap weights are not accounted for, and column (3) presents the reliable standard errors from the correct use of mean bootstrap weights and the appropriate adjustments.

It is important to note that column 2's standard errors are generally the lowest, followed by columns 1 and 3. Column 3 presents the reliable design-based variance

Table 2: LOGIT Results: Predictors of the Receipt of a Performance Appraisal*

Variables	Coeff.	Column 1: Weighting Method		Column 2: Bootstrapping Method		Column 3: Mean Bootstrapping Method	
		Std. Error	Sig.	Std. Error	Sig.	Std. Error	Sig.
Org Variables							
In House HR Dept	.53	.0713	p<.001	.0147	p<.001	.0140	p<.001
Number of Employees	.00	.0000		.0000	p<.001	.0000	
Unionized	-.25	.0797	p<.05	.0148	p<.001	.1048	p<.05
Service Industry	.12	.0787		.0134	p<.001	.0949	
For-Profit	-.10	.0994		.0182	p<.001	.1284	
Job Variables							
Wage	.01	.0037	p<.05	.0006	p<.001	.0041	p<.05
Full-time/Permanent	.66	.1711	p<.001	.0217	p<.001	.1538	p<.001
Full-time/Temporary	.06	.2562		.0351	p<.10	.2480	
Part-time/Permanent	.61	.1721	p<.001	.0241	p<.001	.1705	p<.001
Use Computer	.64	.0820	p<.001	.0127	p<.001	.0901	p<.001
Professional	.41	.2084	p<.05	.0323	p<.001	.2282	p<.10
Manager	.09	.2124		.0295	p<.05	.2084	
Technical/Trades	.14	.1853		.0259	p<.001	.1828	
Marketing	-.16	.2433		.0290	p<.001	.2053	
Clerical	-.08	.2027		.0295	p<.05	.2085	
Individual Variables							
Gender (Female)	.07	.0761		.0113	p<.001	.0797	
Age	.01	.0190		.0033	p<.10	.0236	
Age^2	.00	.0002		.0000	p<.001	.0003	
Recent Immigrant	.27	.2549		.0299	p<.001	.2114	
Disability	-.11	.1138		.0154	p<.001	.1089	
Constant	-1.07	.4593	p<.05	.0761	p<.001	.5378	p<.05
R^2	.06						

*N=20,834; Reference Groups: Part-time/Temporary and Production

estimates. In column 3, the explanatory variables (in-house HR department, unionization, hourly wage, being full-time/permanent, being part-time/permanent, the use of a computer and professional occupation) are statistically significant at the 95% level. Compared to column 1 the significance of professional occupation as a predictor is more accurate and,

without the appropriate variance estimation the conclusions drawn from the inference, would have been inaccurate. Further, when no adjustments are made for the WES provided mean bootstrap weights (column 2) all variances are underestimated by a factor of C resulting in output that leads to inappropriate inference for all variables. All predictors are significant using

this method, when only those predictors significant in column 3 should be interpreted as such. Thus, failing to use bootstrap weights and the mean bootstrap weights appropriately lead to biased standard errors and inappropriate inference.

Conclusion

The results of this study portend a significant methodological contribution with respect to choosing the appropriate variance estimation method when using a large dataset, such as the WES. Although the beta coefficients are the same for all three methods, which predictors are significant differs depending on the method used. When presenting findings from a large dataset and a complex sampling design, the variance estimation method that was used should be acknowledged. Readers should be aware that different results can be presented depending on the method selected. This suggests that researchers should be cautious when choosing a weighting method and be aware of the biased standard errors that are produced when the inappropriate method is used. This study showed the practical implication of choosing an appropriate, unbiased weighting method when analyzing a large dataset with complex sampling design.

References

Buckley, N., & Chowhan, J. (2005). Using mean bootstrap weights in Stata: A BSWREG revision. *The Research Data Centres Information and Technical Bulletin*, *Spring* (*2*), 23-37. Statistics Canada Catalogue No. 12-002-XIE.

Mann, S. L. & Latham, G. P. (2008, May). –Who receives a performance appraisal and does it matter? –An empirical investigation of the determinants of the receipt of a performance appraisal and its effect on job satisfaction. Paper presented at the Annual Meeting of the Administrative Sciences Association of Canada. Halifax, Canada.

Satin, A., & Shastry, W. (1993). *Survey sampling: A non-mathematical guide.* Ottawa: Ministry of Industry, Statistics Canada, Social Survey Methods Division. Catalogue No. 12-602-XPE.

Statistics Canada. (2003). Guide to the analysis of workplace and employee survey 2001. *Business and Labour Market Analysis Division & Labour Statistics Division*, *June*. Ottawa: Canada.

Yeo, D., Mantel, H., & Liu, T. (1999). Bootstrap variance estimation for the national population health survey. *American Statistical Association: Proceedings of the Survey Research Methods Section.* Baltimore, MD, August.