Replicate Variance Estimation For Combined Data of Independent Surveys[1]

Donsig Jang and Amang Sukasih, Mathematica Policy Research. Inc.
Donsig Jang, 600 Maryland Ave., SW, Suite 550, Washington, D.C. 20024

Key words: SESTAT, replication method, combined dataset

Several survey data are often combined into one dataset to represent the population covered by those surveys. In many cases each survey component has its own replication method for variance estimation which accounts for its sample design and other adjustment procedures. However, replication methods used for each component survey are not necessarily same. Consequently a way to combine replicate weights from each survey is desirable.

The Scientist and Engineers Statistical Data System (SESTAT) is an integrated data system combining three independent surveys. We constructed combined replicate weights for the SESTAT based on individual survey specific replicate weights and evaluated variance estimates based on the combined replicate weights. Empirical results show that in practice this approach is preferable over the approach that approximates the replicate weights by using unified replication method.

1. Introduction

This paper describes the construction of the 1993 SESTAT replicate weights, followed by an empirical comparison of variance estimates under several different methods. Jang and Sukasih (2002) performed empirical comparisons of several variance estimation methods for each of the three SESTAT component surveys: National Survey of College Graduates (NSCG), National Survey of Recent College Graduates (NSRCG), and Survey of Doctorate Recipients (SDR). Such empirical comparisons indicate that no single method is predominantly better than any other for *all* three surveys. Instead, one or more specific methods can be chosen for each survey, depending on the underlying sample design and weight adjustment procedures used. In this paper, we are to investigate the performance of different replication methods along with Taylor series linearization method for a combined survey dataset using the 1993 SESTAT extract file.

Four variance estimation methods were considered for this investigation: the Taylor series linearization (TS) method and three replication methods. For the replication methods, we constructed three sets of SESTAT replicate weights: (1) the jackknife method, (2) the successive difference method, and (3) the combination of the original survey-specific replicate weights (jackknife weights plus successive difference replicate weights).

2. Construction of Design Variables and Replicate Weights for the 1993 SESTAT

2.1. Design Variables for the TS Method

We assumed the following sample design and chose the following variables for the TS method:

NSCG. The 1993 NSCG sample was based on a two-phase design: phase 1 was a 1990 decennial census long form sample, and phase 2 was the 1993 NSCG sample from long form respondents. Because of the Census Bureau's privacy legislation, Title 13, phase 1 design information was not available to the outside of Census. Accordingly, the 1993 NSCG is assumed to use a single phase stratified sample design in which the stratification variable was the original sampling cells (863 strata) used for the 1993 NSCG.

SDR. The 1993 SDR is a longitudinal survey first administered in 1973. Consequently, this survey sample can be regarded as a multi-phase sample. However, because of the complexity and incomplete information on previous design variables, the 1993 SDR was assumed to be a single-stage stratified sample design for variance estimation. The stratification variable was thus the original sampling cells (240 strata) used for the 1993 SDR.

NSRCG. The 1993 NSRCG used a two stage sample design. The first stage sample consists of certainty and non-certainty PSUs. A systematic probability proportional to size sample design was used for non-certainty PSU selection. A design approximation of two PSUs per stratum was used for the jackknife variance estimation method. There are 50 jackknife strata and thus 100 PSUs.

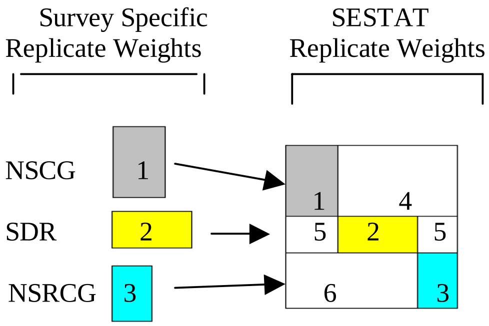
2.2. Construction of Replicate Weights

The three sets of SESTAT replicate weights were constructed: based on the jackknife method, based on the successive difference replication method, and based on the combination of the jackknife and successive difference replication methods. The jackknife and the successive difference replication

[1] The authors would like to thank the National Science Foundation (NSF). This work was motivated by a contract work for the NSF to develop SESTAT variance estimation procedures. The authors take the sole responsibility for the interpretations and conclusions in this article.

methods were chosen here because the replicate weights based on these methods have been available for the NSRCG (jackknife) and NSCG (successive method). The construction of the three sets of weights follows the same algorithm as described in Figure 1 and thereafter.

Figure 1. Construction of the SESTAT replicate weights



1 = NSCG Replicate Weights
2 = SDR Replicate Weights
3 = NSRCG Replicate Weights
4 = SESTAT Final Weights (NSCG component)
5 = SESTAT Final Weights (SDR component)
6 = SESTAT Final Weights (NSRCG component)

Each weight was put into the data set as a form of block diagonal (#1, #2, #3), and the SESTAT final weights were put into the dataset for off-diagonal (#4, #5, #6). This construction was based on the assumption that the populations covered by three component surveys are mutually exclusive and that each survey sample was constructed separately. Given these assumptions, each survey can be regarded as a "super stratum." The use of the final SESTAT weights as the replicate weights for off-diagonal units in the dataset does not introduce variability into the variance estimates for domains that are completely nested in a component survey. In fact, the replicate estimates from the off-diagonal are equal to the full sample estimate, because the replicate weights on the off-diagonal are same as the final SESTAT weight. Thus, the values of variance factor $(\hat{\theta}_r - \hat{\theta})^2$, i.e., the difference between the estimates calculated from the replicate r and that calculated on the basis of full sample, is zero for replicates with off-diagonal units only.

2.2.a. Method 1: Jackknife Replicate Weights

We used the regular jackknife method, in which a unit or a cluster of units is deleted within stratum one at a time.

NSCG. the original 1993 NSCG sample design was approximated by a two PSUs per stratum design. We developed the 125 replicate weights on the basis of a sampling design having two PSUs per stratum. The 125 weighting cells (used in developing the 1993 NSCG final weights) were used as jackknife strata, two pseudo-PSUs were created within each jackknife stratum, and then the Jackknife procedure deleted one PSU at a time, keeping only half of the respondents from the jackknife stratum with deleted pseudo-PSU. The weights were adjusted for nonresponses and poststratification.

SDR. As in the 1993 NSCG, we approximated the 1993 SDR sample design as a two PSUs per stratum sample design. Specifically, we developed 99 replicate weights on the basis of this design, in which the 240 original sampling cells (used in the 1993 SDR) were first collapsed into 99 jackknife strata, two pseudo-PSUs were created within each jackknife stratum, and the jackknife procedure was then used to delete one PSU at a time. The weights were adjusted for nonresponses through the same nonresponse adjustment procedures.

NSRCG. The 50 replicate weights have been available, which are the original jackknife replicate weights provided in the 1993 NSRCG public use data. These weights were developed on the basis of a two PSUs per stratum design.

2.2.b. Method 2: Successive Difference Replicate Weights

This method assumes that each survey sample was selected systematically. For NSCG and SDR, a systematic sample selection was used within each sampling stratum. Consequently, a successive difference replication method would be applicable to these two survey components. On the other hand, the NSRCG used a two stage sample design. A design option adapted for the jackknife method by the survey contractor was a two PSUs per stratum design with 100 PSUs. For successive difference replication method application, we assume that those 100 PSUs were selected systematically. In the successive difference replicate construction algorithm, the order of the Hadamard matrix needs to be prime number plus one so that two row combinations can be unique regardless of the distance of two rows considered. See York 2001 for the details on the Hadamard matrix row assignment algorithm.

NSCG. The 160 replicate weights have been available, which are the original successive difference replicate weights constructed by the Census Bureau.

SDR. The 104 replicate weights were developed on the basis of the assumption that the sample was selected systematically from the frame sorted by sampling stratum variables. Because the SDR sample is about two-thirds the size of the 1993 NSCG, we determined 104 replicate weights that is about two-thirds of the NSCG replicates.

NSRCG. The 52 replicate weights were developed based on the assumption that PSUs were selected systematically. We chose 52 because 51 is the closet prime number to the number of survey-specific replicate weights created by the survey contractor (i.e., 50 jackknife replicate weights).

2.2.c. Method 3: Combined Replicate Weights

Combined replicate weight method is based on the assumption that the variance estimation method within each survey is the most appropriate for its own survey. For the 1993 survey cycle, the replicate weights calculated by the survey contractors have been available for the NSCG and NSRCG. The balanced repeated replication (BRR) replicate weights for the SDR are not available for this study. Following what the Census Bureau did for the 1999 and 2001 SDRs, we created SDR replicate weights based on the successive different replication method.

NSCG. The 160 successive difference replicate weights are the original replicate weights constructed and provided to us by the Census Bureau.

SDR. Because no replicate weights for the SDR were available from the survey contractor, we considered the successive difference replicate weights as original weights. As discussed above in section 2.1.b, 104 replicate weights were developed.

NSRCG. The 50 jackknife replicate weights are the original replicate weights given in the 1993 NSRCG public use data.

3. Variance Estimation Procedures

3.1. Taylor Series Linearization Method

In order to use the TS method in survey analysis software like SUDAAN or Stata, different sample design features used in the SESTAT component surveys must be integrated into one design. The 1993 NSCG and 1993 SDR are assumed to be a stratified sampling design without clusters, whereas the 1993 NSRCG is assumed to be a two-stage sampling design with clusters of individuals. We, therefore, want to create common design variables, such as sampling strata and PSUs, for each of three surveys. For the NSRCG, survey-specific design variables for the strata and PSUs can be used for SESTAT-level design variables without any modification. On the other hand, for the NSCG and SDR, we created a dummy variable indicating PSU that contains only one observation within each PSU. Thus, given the common design variables, variances can be estimated by specifying a two-stage cluster sample design.

3.2. Replication Method

The following formula was used with all three replication methods to estimate variances:

$$v\left(\hat{\theta}\right)=\sum_{r=1}^{R} c_r\left(\hat{\theta}_r-\hat{\theta}\right)^2$$

where $\hat{\theta}$ is obtained from the full sample, $\hat{\theta}_r$ is from the *r*-th replicate, and the multiplier c_r related to the *r*-th replicate is determined based on the replication method used. The three replication variance estimators with appropriate values of c_r are presented below.

3.2.a. Jackknife Variance Estimator

The general formula for the jackknife variance estimator in SUDAAN (RTI 2002) can be expressed as:

$$v_{Jack}\left(\hat{\theta}\right)=\sum_h \frac{N_h-D_h}{D_h R_h}\sum_i\left(\hat{\theta}_{hi}-\hat{\theta}\right)^2$$

where

N_h is the number of PSUs or clusters within the stratum *h*,

D_h is the number of PSUs or clusters deleted in creating the replicate,

R_h is the number of replicates selected,

$\hat{\theta}_{hi}$ is the estimate of the parameter θ from the *i*-th replicate of the *h*-th stratum,

$\hat{\theta}$ is the estimate based on the entire sample.

As described in Section 2, jackknife replicate weights were constructed as if the 1993 SESTAT sample were based on the two-PSUs-per-stratum design with 274 strata. From each stratum, one PSU was randomly deleted to construct a replicate. Consequently, the multiplier for SESTAT Jackknife variance estimation can be $(N_h-D_h)/D_hR_h=(2-1)/(1\times1)=1$ for all 274 strata. The simplified jackknife variance estimator can thus be expressed as:

$$v_{Jack}\left(\hat{\theta}\right)=\sum_{r=1}^{R}\left(\hat{\theta}_r-\hat{\theta}\right)^2$$

where R denotes the number of jackknife replicates constructed for the 1993 SESTAT data, i.e. R = 274.

3.2.b. Successive Difference Replication Variance Estimator

The successive difference replication variance estimator, given R replicate weights (Fay and Train 1995), can be expressed as:

$$v_{SDRM}\left(\hat{\theta}\right)=\frac{4}{R}\sum_{r=1}^{R}\left(\hat{\theta}_r-\hat{\theta}\right)^2$$

where R is the number of replicates,

$\hat{\theta}_r$ is the estimate of the parameter θ from the r-th replicate,

$\hat{\theta}$ is the estimate based on the entire sample.

Since the numbers of replicates are different across the three SESTAT surveys, the variance estimator for SESTAT-level estimates can be calculated as:

$$v_{SDRM}\left(\hat{\theta}\right)=\sum_{i=1}^{3}\frac{4}{R_i}\sum_{r=1}^{R_i}\left(\hat{\theta}_{ir}-\hat{\theta}\right)^2$$

where the multiplier $4/R_i$ for each survey is calculated as follows:

- NSCG part: $4/R_1=4/160=0.025$
- SDR part: $4/R_2=4/104=0.03846$
- NSRCG part: $4/R_3=4/52=0.07692$.

3.2.c. Combined Replication Method Variance Estimator

As presented in Section 2, replicate weights for this method were constructed by combining the jackknife replicate weights (50 replicates) from the NSRCG and the successive difference replicate weights from the NSCG (160 replicates) and the SDR (104 replicates). This process resulted in 314 combined replicate weights. The variance estimator can be expressed as:

$$v_{Comb}=\sum_{r=1}^{160}k_1\left(\hat{\theta}_r-\hat{\theta}\right)^2+\sum_{r=161}^{264}k_2\left(\hat{\theta}_r-\hat{\theta}\right)^2+\sum_{r=265}^{314}k_3\left(\hat{\theta}_r-\hat{\theta}\right)^2$$

where the multiplier is calculated as follows:

-NSCG part: $k_1=4/R_1=4/160=0.025$

-SDR part: $k_2=4/R_2=4/104=0.03846$

-NSRCG part: $k_3=1$.

All four variance estimates presented above can be obtained from SUDAAN by specifying the appropriate design option, weight, and multipliers. Our empirical results in Section 5 are obtained from SUDAAN.

4. Types of Statistics and Variables Considered

For the empirical evaluation of the variance estimates presented above, we considered five key characteristics—total number of scientists and engineers (S&E), proportion of S&Es employed or unemployed, mean salary, median salary, and regression coefficients for predicting salary—for various groups.

A. Total number of S&Es: all, by handicap status, by race, and by degree level
B. Proportion of S&Es who were employed or unemployed: unemployed, looking for Job and employed by race
C. Mean salary: all, by handicap status, by race, and by degree level
D. Median salary: all, by handicap status, by race, and by degree level
E. Regression coefficients to predict salary on degree field/occupation group, degree level and job-education match

For all estimates of these characteristics, we estimated the variance using the methods described in Section 3.

5. Performance of the Variance Estimation Methods

Using the 1993 SESTAT dataset, we calculated point estimates and standard errors for all variables presented in Section 4. To evaluate the performance of the four variance estimation methods, we also calculated the estimate of the coefficient variations (CVs):

$$\text{CV}_i(\hat{\theta})=\frac{se_i(\hat{\theta})}{\hat{\theta}}\times 100\%$$

where index *i* indicates the TS, jackknife, successive difference replication, and combined replication method, respectively. The estimate of the coefficient of variation is scale-free and quantifies the precision of a point estimate (in terms of its standard error) in relation to the point estimate itself. Thus, it can be used to compare variance estimates (standard errors) across different groups/domains and variables.

In addition to CV, we calculated the relative difference (RD) of each estimate from the TS-based variance estimate:

$$\text{RD}_i(\hat{\theta})=\frac{se_i(\hat{\theta})-se_{TS}(\hat{\theta})}{se_{TS}(\hat{\theta})}\times 100\%$$

where index *i* indicates the jackknife, successive difference replication, and combined replication method, respectively. Under the condition that the sampling design assumption used for the Taylor Series calculation is appropriate, TS-based variance estimate can be used as the baseline for the relative differences calculation. This strategy is analogous to the customary practice that the TS variance estimator is usually used in comparisons when a resampling method is analytically evaluated (Shao 1996).

Table 1 presents point estimates, standard errors, coefficients of variation, and the corresponding relative differences for total number of S&Es, proportion of S&Es employed or unemployed, mean salary, and median salary. This table provides the comparison based on descriptive analysis. Table 2 presents the same comparison for multivariate analysis through the regression coefficients.

From Tables 1 and 2, the values of coefficient of variation (CV) estimates are all small, i.e., less than < 10%, except for estimates for the small domain “Other Race” and some of regression coefficients. In addition, CV estimates across the four methods are not significantly different although the jackknife method produced somewhat smaller CVs for some number of variables/estimates. Based on these observations, all four methods seem to be acceptable for the use of SESTAT variance estimation.

However, the tables show that the jackknife method seems to produce variance estimates that deviate significantly from those produced by TS and the two other replication methods. That is, the relative difference measures of jackknife estimates are markedly different from successive difference replication estimates and combined replication estimates. On the other hand, relative difference measures from the successive difference replication and combined replication methods are close to each other. This empirical result strongly indicates that these two replication methods and the TS method would produce variance estimates based on the sample design assumed. Note that two of three component surveys used the successive difference replicate weights and more than 80% of the SESTAT units come from these two components. This explains why the successive difference replication method gives results similar to the combined replication method. On the other hand, the jackknife method produces estimates that deviate from the other three mainly because the design approximation required for the jackknife method might be farther from the original sample design than the other two replication methods or the TS method.

For the estimate of median salary, none of the replication methods reviewed seems to be appropriate. This is consistent both with conclusions in the literature and with our previous comparison results based on SESTAT specific-survey variables.

6. Recommendations

We recommend the combined replication method for estimating variances in the 1993 SESTAT. This method rests on the assumption that the original replicate weights constructed by NSF's survey contractors for each SESTAT component survey are the appropriate, if not the best, weights for their respective sample survey design.

The construction of SESTAT replicate weights through the combined method can be easily justified theoretically for the linear estimators, although this needs analytical justification for nonlinear estimators. However, our work empirically shows that this method works reasonably well for SESTAT variance estimation. Another advantage of the use of combined survey specific replicate weights is that the survey contractors will not need to change their replicate weight construction methods to fit in SESTAT variance estimation. This in turn will save the time and money for the development of the direct variance estimation method for SESTAT.

Given the combined replicate weights in the dataset, users need to know the survey component of each unit and replicate multipliers for variance estimation (see section 3.2c). Then the variance computation can be easily performed using a simple routine/macro. For analysts who have access to SUDAAN software, these multiplier values can be easily specified as an option in variance estimation using replicate weights.

REFERENCES

Fay, R.E., and G.F. Train. 1995. "Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties (Disc: p172-174)." *Proceedings of the American Statistical Association*. Section on Government Statistics: 154-159.

Jang, D. and Sukasih, A. 2002. "An Empirical Investigation of the Taylor Series Linearization Method, Jackknife Replication Method, and Successive Difference Replication Method for Variance Estimation. Washington, DC: Mathematica Policy Research, Inc

Kovar, J.G., J.N.K. Rao, and C.F.J. Wu. 1988. "Bootstrap and Other Methods to Measure Errors in Survey Estimates." *Canadian Journal of Statistics* 16, Supplement: 25-45.

Krewski, D., and J.N.K. Rao. 1981. "Inference from Stratified Samples: Properties of the Linearization, Jackknife and Balanced Repeated Replication Methods." *The Annals of Statistics* 9: 1010-1019.

Research Triangle Institute. 2001. *SUDAAN User's Manual, Release 8.0.* Research Triangle Park, NC: Research Triangle Institute, 2001.

Shao, J. 1996. "Resampling Methods in Sample Survey." *Statistics* 27: 203-254.

Wolter, K.M. 1984. "An Investigation of Some Estimators of Variance for Systematic Sampling." *Journal of the American Statistical Association* 79: 781-790.

Wolter, K. M. 1985. *Introduction to Variance Estimation*, New York: Springer-Verlag.

York, B.L. 2001. "Generalized Variance Parameter for the 1999 National Survey of College Graduate." Internal Census Bureau Memorandum to Chester E. Bowie from Alan R. Tupek, April 24, 2001.

Table 1. Standard errors, Coefficient of variations, and relative differences for several estimates: 1993 SESTAT

Statistic	Variable	Sample size	Point estimate	Standard error				Coefficient variation (%)				Relative difference (%)			
				Taylor Series	Jackknife	Successive	Combined Weights	Taylor Series	Jackknife	Successive	Combined Weights	Taylor Series	Jackknife	Successive	Combined Weights
Total	ALL	126,721	11,615,160	52,119.958	47,129.589	57,251.699	44,636.406	0.45	0.41	0.49	0.38	0.00	-9.57	9.85	-14.36
Total	Handicap	8,039	806,369	13,723.416	9,440.660	15,238.478	15,184.336	1.70	1.17	1.89	1.88	0.00	-31.21	11.04	10.65
Total	Hispanic	5,929	352,005	5,918.430	5,590.622	6,381.207	5,940.053	1.68	1.59	1.81	1.69	0.00	-5.54	7.82	0.37
Total	White	97,780	9,896,201	50,484.867	44,296.111	52,983.746	41,915.564	0.51	0.45	0.54	0.42	0.00	-12.26	4.95	-16.97
Total	Black	7,283	573,821	8,454.139	6,456.838	9,868.608	8,621.495	1.47	1.13	1.72	1.50	0.00	-23.63	16.73	1.98
Total	Asian	14,933	757,897	8,634.919	5,980.789	8,481.087	8,372.786	1.14	0.79	1.12	1.10	0.00	-30.74	-1.78	-3.04
Total	Native American	753	29,684	1,431.601	1,190.818	1,488.015	1,464.973	4.82	4.01	5.01	4.94	0.00	-16.82	3.94	2.33
Total	Other Race	43	5,551	1,106.289	752.467	1,024.906	1,024.906	19.93	13.55	18.46	18.46	0.00	-31.98	-7.36	-7.36
Total	Bachelor	52,717	6,974,978	41,541.392	40,270.261	48,867.054	35,335.608	0.60	0.58	0.70	0.51	0.00	-3.06	17.63	-14.94
Total	Master	27,701	3,011,721	26,468.292	38,374.090	24,295.885	23,867.721	0.88	1.27	0.81	0.79	0.00	44.98	-8.21	-9.83
Total	Doctorate	41,253	706,701	6,361.178	8,878.675	5,897.349	5,898.524	0.90	1.26	0.83	0.83	0.00	39.58	-7.29	-7.27
Total	Professional Degree	5,050	921,760	16,222.821	10,006.706	15,163.410	15,163.454	1.76	1.09	1.65	1.65	0.00	-38.32	-6.53	-6.53
Proportion	Unemployed	126,721	0	0.002	0.001	0.002	0.002	1.09	0.85	1.00	0.99	0.00	-22.22	-8.19	-8.77
Proportion	Unemployed Looked for Work	2,995	1	0.000	0.000	0.000	0.000	0.00	0.00	0.00	0.00				
Mean	Salary, All Level	97,689	51,784	162.993	221.465	182.548	164.809	0.31	0.43	0.35	0.32	0.00	35.87	12.00	1.11
Mean	Salary, Bachelor	39,907	44,951	189.113	110.210	210.509	185.592	0.42	0.25	0.47	0.41	0.00	-41.72	11.31	-1.86
Mean	Salary, Master	20,901	53,172	287.763	422.022	262.914	261.897	0.54	0.79	0.49	0.49	0.00	46.66	-8.64	-8.99
Mean	Salary, Doctorate	32,547	62,857	333.622	266.613	335.033	335.068	0.53	0.42	0.53	0.53	0.00	-20.09	0.42	0.43
Mean	Salary, Professional Degree	4,334	83,549	789.184	1,601.878	796.916	796.894	0.94	1.92	0.95	0.95	0.00	102.98	0.98	0.98
Median	Salary, ALL	97,689	44,992	22.775	1.590	3.682	3.601	0.05	0.00	0.01	0.01	0.00	-93.02	-83.83	-84.19
Median	Salary, Bachelor	39,907	39,999	0.273	1.918	0.304	0.257	0.00	0.00	0.00	0.00	0.00	602.16	11.30	-5.79
Median	Salary, Master	20,901	47,998	104.879	1.600	1.267	1.265	0.22	0.00	0.00	0.00	0.00	-98.47	-98.79	-98.79
Median	Salary, Doctorate	32,547	57,188	423.367	658.786	646.016	646.018	0.74	1.15	1.13	1.13	0.00	55.61	52.59	52.59
Median	Salary, Professional Degree	4,334	77,993	1,122.485	1,591.550	1,114.075	1,114.074	1.44	2.04	1.43	1.43	0.00	41.79	-0.75	-0.75

Table 2. Standard errors, coefficient of variations, and relative fifference for regression coefficients: 1993 SESTAT

Variable/Term	Estimated Coefficient	Standard Error				Coefficient Variation (%)				Relative Difference (%)			
		Taylor Series	Jackknife	Successive	Combined	Taylor Series	Jackknife	Successive	Combined Weights	Taylor Series	Jackknife	Successive	Combined Weights
Intercept	42079.9141	481.2093	290.7790	491.8514	491.1094	1.14	0.69	1.17	1.17	0.00	-39.57	2.21	2.06
Major Field for Highest Degree:													
Computer and Math Sciences	1410.5280	542.8930	524.7548	608.7802	599.2193	38.49	37.20	43.16	42.48	0.00	-3.34	12.14	10.38
Life and Related Sciences	-6135.1325	513.3523	427.8378	498.4244	495.7548	-8.37	-6.97	-8.12	-8.08	0.00	-16.66	-2.91	-3.43
Physical and related sciences	2351.9118	647.0566	616.1574	694.6445	688.8437	27.51	26.20	29.54	29.29	0.00	-4.78	7.35	6.46
Social and Related Sciences	-5010.8489	456.2805	228.7671	492.5204	479.0783	-9.11	-4.57	-9.83	-9.56	0.00	-49.86	7.94	5.00
Engineering	7924.6792	448.3208	210.6405	448.0088	441.8934	5.66	2.66	5.65	5.58	0.00	-53.02	-0.07	-1.43
Non-SE degrees (BASELINE)	-	-	-	-	-	-	-	-	-	-	-	-	-
Highest Degree:													
Bachelor (BASELINE)	-	-	-	-	-	-	-	-	-	-	-	-	-
Master	7370.2870	345.6697	384.5989	343.1357	338.8117	4.69	5.22	4.66	4.60	0.00	11.26	-0.73	-1.98
Doctorate	18141.3121	401.4243	276.6666	420.0177	411.3924	2.21	1.53	2.32	2.27	0.00	-31.08	4.63	2.48
Professional Degree	38271.2060	886.6684	1574.5194	905.1669	903.6406	2.32	4.11	2.37	2.36	0.00	77.58	2.09	1.91
Principal Job Related to Highest Degree:													
Job-Edu Closely Related	3241.3204	423.3034	282.2777	408.5177	407.3314	13.06	8.71	12.60	12.57	0.00	-33.32	-3.49	-3.77
Job-Edu Smwhat Related	5245.0054	459.8312	225.0670	487.7206	487.2848	8.77	4.29	9.30	9.29	0.00	-51.05	6.07	5.97
Job-Edu Not Related (BASELINE)	-	-	-	-	-	-	-	-	-	-	-	-	-