Imputing Missing Income Data and Weighting Data with Imputed Income

Bidisha Mandal and Elizabeth A. Stasny
Ohio State University

Abstract— Income is an important demographic variable in social science research. But income data is more difficult to obtain compared to other demographic information, since it is typically considered to be private and personal.

The focus of this research is to try several imputation schemes to account for missing income data and then to use income for weighting the survey data. Specifically, hot deck imputation, mean imputation within adjustment cells, and regression imputation for missing income values are explored. Then weighting using each imputation scheme is carried out, and the resulting estimates for a key survey variable are compared. Weighting is also carried out without income as one of the weighting variables. The purpose is to see how successful imputation has been in replacing missing income data, and how good imputed income is as a weighting variable.

Keywords – Hot deck imputation, Mean imputation, Regression imputation, RDD survey

Introduction

Income is often a key demographic variable in social science research and many researchers would like their samples to agree with population income distribution. However, income data are very difficult to obtain. Many respondents consider the topic sensitive and personal. Moore, Stinson, and Welniak (1999), citing data from the Current Population Survey (CPS), report income nonresponse rates ranging from 20% up to almost 50%. Körmendi (1998) demonstrates that conducting the interview over the telephone can contribute to problems when collecting income information. There is reluctance to share income information with others, especially interviewers. Income is a major component in defining one's social class and standing. Consequently, people with low incomes are reluctant to divulge this information, fearing that it will reflect badly on them. On the other hand people with high incomes may be concerned about envy, may be modest, or may be embarrassed to discuss their wealth (Smith, 1991). Olson et al. (1999) note that lower income respondents have more difficulty in providing exact income information. They also suggest that even when respondents are willing to disclose their total family income, they may not know what it is, or they may be frustrated by the effort required to come up with the information. Also, people may be unwilling to discuss

income because their true income differs from their declaration on income tax forms, loan/credit/scholarship applications, government benefit statements, or other income based records (Smith, 1991). Garner and Blanciforti (1994) suggest that a respondent's characteristics, such as age, race, and education level, may affect the reporting of income data. More recently, Moore, Stinson, and Welniak (1997) reviewed the literature and summarized research on the cognitive factors which affect income reporting and the nature of the resulting errors in survey measurement of income.

When the item is as analytically important as income, lower item nonresponse makes the data more useful. By imputation one can fill in the missing values and then analyze the 'completed' data using standard methods. Imputation also helps to retain all reported data in multivariate analysis. But there are some demerits of imputation too. Imputation may distort distribution of the variables and alter associations between income and other variables. Standard analyses after imputation will typically overestimate the precision of survey variables. Also, imputation could be computationally intensive. Although many statistical packages, like SPSS, include imputation as one of the features, but if one wants to be sure about what goes on behind the execution of a command, then one might need to get more involved in all the steps, resulting in a time-consuming procedure.

Missing data arising from item nonresponse can lead to biased estimates if the analysis is restricted to the records with complete information for the items in question. Much survey analysis is multivariate in nature, and even low item nonresponse rates for several items together may result in a sizeable proportion of records with missing data for a particular analysis. If the missing data are not missing at random, then there could be problems even with a survey analysis which is univariate in nature. This is definitely true for income data as has been mentioned above.

This article studies the effectiveness of three imputation schemes to account for the missing income data, namely, hot deck imputation, mean imputation within adjustment cells, and regression imputation. Also it addresses the quality of analyses when the imputed income is used as a weighting variable.

Data

Our goal is to see if income categories can be used for weighting the sample to agree with a population income distribution. The data we used to address the above mentioned research questions are mainly from the Buckeye State Poll (BSP) economic surveys. The BSP is a Random Digit Dialing

(RDD) survey, started in November 1996 and carried out each month through April 2002 by the Center for Survey Research, The Ohio State University. The BSP surveys have a total of 39,313 cases. A subset of the data that has complete information on each of the demographic variables – age, education, gender, have at least one child (child is defined as a non-adult person in the household) or not, race and income – is treated as the 'Population'.

As is common in RDD surveys, the BSP asked the demographic questions, including the income question, at the end of the survey. Respondents were initially asked for their exact household income. Many respondents find this question sensitive and/or difficult to answer. If a respondent is hesitant and/or refuses to answer the exact income question, then the question is skipped and the income ladder question is asked. This question provides the respondent with various income ranges and the respondent is expected to choose the range containing the value of the exact household income. While using income ranges does lose some information, respondents appear to be more willing to place themselves in a broad category of incomes than they are to report specific amounts.

Out of the total of 39,313 cases in the BSP surveys, income information is missing for 5,993 respondents. Missing data for any variable means information is missing due to either respondent refusal or uncertainty. Also data may be suspect for some respondents who reported $0, $12 etc. as annual income. Potential explanations are - persons who have businesses can break even, so their total income is zero; persons with losses but some income from other sources can have odd-looking incomes; those whose main source of income is from something illegal (prostitution, drugs, etc.) may not want to tell their interviewers about it; a female head can have a boyfriend who pays her expenses, but maybe he isn't giving her any actual cash--so technically she has no income of her own. So we dropped all those cases with reported income of less than $5,000 per year. There were 392 such cases. Since our aim is to impute a value for missing income, all those cases that did not respond to the exact-income question but did respond to the income-range question are also left out. There were 90 such respondents out of the 5,993 initially dropped for non-response to the exact income question. Fortunately there were no noticeable trends among the cases left out, that is, these people were not from any particular demographic category or from any particular month(s). Finally, the number of nonrespondents is 5,993 plus 392, that is 6,385, and thus, the nonresponse rate for income is 16.24%.

There is some missing data for age, education, have kids or not, and race too. The numbers of nonrespondents for age, education, have kids or not, and race are 415, 111, 119 and 383 respectively, and the corresponding nonresponse rates are 1.06%, 0.28%, 0.30%, and 0.97% respectively. For

building the 'Population' these nonrespondents were also left out. So there are 31,900 cases in the 'Population'.

Respondents were not asked their gender in the BSP telephone surveys, rather the interviewers filled in this information themselves. Table 1 below gives the combined nonresponse rates for income and the five demographic variables, age, education, gender, have kids or not, and race. Table 1 shows that whenever income data are taken under consideration, the combined nonresponse rate shoots up irrespective of the other variable in the pair.

Table 1

Combined Nonresponse Rates

Weighting Variable	Percent with Variable Missing	Percent with Variable and/or Income Missing
Age	1.06%	13.23%
Education	0.28%	13.78%
Have kids	0.30%	13.69%
Gender	0.00%	13.94%
Race	0.97%	13.47%

Figure 1 in the appendix graphs the nonresponse rates of age, education, gender, have kids or not and race, along with the combined nonresponse rates of these demographic variables and income.

The 'Population' also has complete information with respect to another variable of interest. It is the Ohio Debt Stress Index (ODSI) variable. The ODSI measures the level of stress created for a household by the total amount of debt it has undertaken. It rises when the worry/stress/concern of the household rises. The index takes values on a 0-100 scale, with high scores indicating high stress; those who reported no debt were assigned ODSI value equal to 0. This measure was developed by Dunn, Stec, Lavarakas and Kim (2000). All 31,900 cases in our 'Population' answered the ODSI questions. Appendix A1 gives the calculation of the ODSI.

For the purpose of this study we created missing income data in two ways – income data missing at random and greater percentage of missing income data for lower and higher income respondents. From

now on these two datasets will be called Random and Extreme respectively. The nonresponse rate for income for the BSP dataset before any manipulation is closer to the lower value reported in the literature, and hence we create 20% missing income data using the 'Population'. For the Random dataset about 20% income data is deleted at random. For the Extreme dataset about 40% income data is deleted for both lower (income less than $10,000) and higher (income more than $125,000) income respondents, and about 18% for the rest, so that the rate of missing income data is about 20% for the entire dataset. Graph of the income distribution for our 'Population' is given in Figure 2 in the appendix.

Appendix A2 gives the population income distribution of Ohio residents for the year 2000, as obtained from U.S. Current Population Survey (CPS) March Supplement.

The 'Population' consists of 31,900 cases. Imputations using all three above mentioned methods are done for the entire 'Population', but the complete findings of this research are given for the data from January 2001 to April 2001 only, for the ease of calculations. We will call this the 'Target' dataset. The 'Target' dataset consists of 1865 cases, 440 cases from January, 428 from February, 474 from March, and 523 from April. The findings are given for the combined 'Target' dataset, and also for each of the four months separately.

For random hot deck imputation and mean imputation within adjustment cells we created 512 **imputation cells** using the following demographic variables:

- Age – 4 categories
- Education – 4 categories
- Gender – 2 categories
- Race – 2 categories
- Work status – 8 categories

These five variables were selected since our data showed more difference in income across these groups. The exact categories for these and other important demographic variables for the BSP surveys are provided in the appendix A3.

We base the number of imputation cells on work done by the US Bureau of the Census for the CPS. With around 100,000 observations, CPS hot deck imputation has several thousand cells for the weekly earnings hot deck. The cells are defined by age, education, gender, race, usual hours and occupation.

In our case, random hot deck imputation and mean imputation within adjustment cells will be carried out using the 512 imputation cells. Regression imputation will be carried out using number of adults in the household, age, education, gender, have kids or not, marital status, party affiliation, race, region, work status as the predictor variables. Although the entire 'Population' is used to build the imputation cells and the regression models, imputation will be done only for the 'Target' dataset. We expect that paired t-tests, where a pair consists of the real income and the imputed income, and correlation between the sets of real income values and imputed income values will tell us something about how good each imputation method is. Finally we will weight the data using the WesVar statistical software to obtain raked weights and replicate weights. These are necessary to calculate the estimate and standard error of any variable.

We will be comparing the estimates and standard errors of the variable ODSI (Ohio Debt Stress Index) to find out how good income is as a weighting variable. Initially, the 'Target' dataset will be weighted using age, education, gender, have kids or not, race and region. These are the standard weighting variables used in the BSP surveys. Hence we will call the estimates and standard errors so obtained as the standard or benchmark values. Next, weighting will be done using imputed income for each dataset. There are six such datasets – three Random datasets and three Extreme datasets. The three Random datasets are the Random datasets where income imputation is done using random hot deck imputation, mean imputation within adjustment cells, and regression imputation. Similarly, there are three Extreme datasets. We will compare the estimates and standard errors of ODSI when imputed income is used as a weighting variable with the benchmark values.

Random Hot deck Imputation

In this method a missing value (or record) is replaced with a value from a similar responding unit. In our case, the similar responding unit is chosen from the appropriate imputation cell. For example, if we have a missing income value for a white male who is 35 year old, has a HS degree, and works full time, we randomly choose an income value from all the available income values in the corresponding imputation cell, which is the class consisting of whites, males, 30-44 years olds, HS degree holders, and full time workers. One obvious problem is that a household which actually earns $10,000 per year could get an imputed income value of $100,000 per year provided there is at least one such donor value in the corresponding imputation class with an annual income of $100,000. Since each imputation cell has donor values ranging from less than $5,000 to more than $125,000 per year, this is a valid

concern. An example of such a wide ranged imputation cell from our dataset is the cell corresponding to age – 18-29 years; education – HS degree; gender – male; race – white; work status – in school. The income range of the donor values for this particular imputation cell is $5,000 - $125,000.

Mean Imputation within adjustment cells

Similar to the hot deck imputation, the 'Population' is divided into 512 subsets. Mean imputation is carried out by replacing a missing value by the mean of all the donor values of the corresponding imputation cell. Thus, all the imputed income values are same in any subset. Naturally, this distorts the distribution of the variables, and the distribution peaks at the cell means. Another problem is similar to the one faced using hot deck imputation, i.e., if the mean of a particular cell is $50,000, then a person whose actual income is $45,000 will get this imputed value, and so will a person whose real income is $150,000. An example of such a wide ranged imputation cell from our dataset is the cell corresponding to age – 18-29 years; education – some college; gender – male; race – not white; work status – in school. The income range of the donor values for this particular imputation cell is $5,000 - $200,000, and the mean of the donor values is $52,380.

Regression Imputation

Here predictions from a regression model based on observed values and predictors of the value of interest (in our case income) are used to replace missing values. For regression imputation we consider the respondent's age, education, gender, have kids or not, race, region, marital status, party affiliation, and work status as predictor variables. Two quantitative variables - number of adults in the household and number of children in the household, were included as well. However, the later, number of children in the household, never came up as an important predictor variable during the regression analysis and thus we do not include it in our model. So, in total, we have ten predictor variables to estimate the missing income values. Once again, the 'Population' is created by removing all those cases with incomplete information on age, education, gender, have kids or not and race. Now, it is quite possible that one or more of the four predictors, marital status, party affiliation, region, and work status is missing. There are sixteen such combinations in which these variables can be missing, ranging from none missing to all missing. So we divide the 'Population' into sixteen corresponding subsets. Hence, if work status is the only variable missing, then it can never be a predictor for income

for that subset, and so on. However, for the 'Target' dataset only three of the sixteen such possibilities are present. These three cases are (1) no information missing; (2) information on only marital status missing; (3) information on only party missing. For different subsets, different categories of the demographic variables turn out to be important as predictor variables. We used best subsets regression and stepwise regression to obtain the statistically significant variables for each of the sixteen models. Appendix A4 gives detailed information on the variables used in the three regression models for both Random and Extreme datasets.

Table 2 gives the correlations between real and imputed income values, as well as the average absolute deviation of the imputed income from the real income for the 'Target' Dataset. 'Target' dataset is the combination of January to April 2001 surveys, and the sample size is 1865.

Table 2

Correlations and Average Absolute Deviations for 'Target' Dataset

	Correlation	Average Absolute Deviation
Random		
Hot deck	0.919	6394.32
Mean	0.926	6081.46
Regression	0.951	4732.62
Extreme		
Hot deck	0.731	7980.42
Mean	0.774	6678.48
Regression	0.782	6321.14

Correlation between the actual and imputed income values for each scheme is high, implying that our imputation procedure is quite successful. Average absolute deviations between the actual and imputed income values imply that regression imputation is best irrespective of whether income is missing at random or greater percentage missing for higher and/or lower income values.

Tables 3a, 3b, 3c and 3d below give the paired t-tests, corresponding p-values and correlations between the actual income and imputed income for the months of January, February, March and April 2001 respectively. Sample sizes are 440, 428, 474, and 523 respectively for January, February, March, and April 2001 BSP surveys. The correlations are high for each imputation scheme for all four months. Significant p-values for paired t-tests are marked with asterisk.

Table 3a

	t-value	p-value	correlation
Random			
Random Hot deck	1.996	0.047*	0.831
Mean	0.423	0.672	0.873
Regression	0.251	0.802	0.886
Extreme			
Random Hot deck	1.995	0.047*	0.863
Mean	1.044	0.297	0.884
Regression	1.322	0.187	0.897

Table 3b

	t-value	p-value	correlation
Random			
Random Hot deck	2.045	0.041*	0.844
Mean	0.932	0.352	0.918
Regression	0.262	0.793	0.899
Extreme			
Random Hot deck	0.803	0.423	0.833
Mean	-0.636	0.525	0.911
Regression	-0.210	0.834	0.906

Table 3c

	t-value	p-value	correlation
Random			
Random Hot deck	1.415	0.158	0.915
Mean	0.085	0.932	0.938
Regression	-0.648	0.517	0.954
Extreme			
Random Hot deck	2.277	0.023*	0.860
Mean	0.821	0.412	0.905
Regression	0.623	0.534	0.911

Table 3d

	t-value	p-value	correlation
Random			
Random Hot deck	1.723	0.085	0.960
Mean	0.403	0.687	0.977
Regression	-0.037	0.971	0.979
Extreme			
Random Hot deck	2.579	0.010*	0.612
Mean	1.930	0.054	0.653
Regression	2.037	0.042*	0.665

Hence, when income is missing at random, regression imputation works best, as shown by both large p-values and high correlations. When income is missing with greater percentage for lower and/or higher value, mean imputation within adjustment cells does slightly better than regression imputation for April 2001. For the other three months, again regression imputation performs better.

Weighting data

The next aim is to see how good income with imputation is as a weighting variable. Sample units are often weighted to make the sample compatible with the characteristics of the population. We obtained the population values for age, education, gender, have kids or not, race and region from the U.S. Current Population Survey (CPS) 2000.

First, the 'Target' dataset is weighted using age, education, gender, have kids or not, race, and region as weighting variables. These are the usual weighting variables for the BSP and many other social surveys. We use the new raked weights and the replicate weights to calculate the estimate and standard error of the variable ODSI (Ohio Debt Stress Index). These are our standard or benchmark values. The statistical software WesVar is used for the weighting and variance estimation. Next, we weight both the Random 'Target' dataset and Extreme 'Target' dataset using imputed income as the weighting variable. Table 4a gives the estimates of ODSI and the standard errors (given in parenthesis) for the Random 'Target' dataset, i.e. when income is missing at random, as well as for the four months separately. The standard values are obtained when the weighting variables are age, education, gender, have kids or not, race and region. The other values are obtained when weighting variable is imputed income only.

Table 4a

	'Target' dataset	January 2001	February 2001	March 2001	April 2001
Standard	26.8478 (0.7351)	27.4944 (1.5279)	27.7921 (1.3123)	26.1048 (1.6631)	25.9512 (1.2027)
Hot deck	27.6232 (0.7347)	28.3567 (1.3900)	28.8301 (1.5282)	25.9874 (1.2713)	27.5444 (1.0426)
Mean	27.6425 (0.7462)	28.6455 (1.5924)	29.2006 (1.5145)	26.1196 (1.3336)	27.0553 (1.0413)
Regression	27.8513 (0.7822)	29.0490 (1.6002)	29.3697 (1.5566)	26.2950 (1.3585)	27.3443 (1.0364)

Table 4b gives the estimates of ODSI and the standard errors (given in parenthesis) for the Extreme 'Target' dataset, i.e. when higher and/or lower income value are missing with greater percentage, as well as for the four months separately. The standard values are obtained when the weighting variables are age, education, gender, have kids or not, race and region. The other values are obtained when weighting variable is imputed income only.

Table 4b

	'Target' dataset	January 2001	February 2001	March 2001	April 2001
Standard	26.8478 (0.7351)	27.4944 (1.5279)	27.7921 (1.3123)	26.1048 (1.6631)	25.9512 (1.2027)
Hot deck	27.6662 (0.7494)	28.6990 (1.5540)	29.0893 (1.5079)	25.9886 (1.3394)	27.5323 (1.0356)
Mean	27.6379 (0.7475)	29.1916 (1.5324)	29.4015 (1.5512)	25.5810 (1.3104)	27.2845 (1.0961)
Regression	27.7077 (0.7814)	29.3292 (1.5387)	29.3060 (1.3706)	25.7469 (1.3706)	27.4268 (1.0796)

Figures 3 and 4 in the appendix give the graphical comparison of ODSI estimates of the Random and Extreme 'Target' datasets respectively for the four months.

Conclusions

While the imputed income values obtained by all three imputation schemes – random hot deck imputation, mean imputation within adjustment cells, and regression imputation – were highly correlated to the actual income values, the procedure itself is computationally intensive and time-consuming. We used several software packages to do the imputation, like SPSS and Minitab. We were dealing with a huge dataset of 31,900 cases. On the other hand, if we did not have such a large dataset, we would not have been successful in the imputation due to lack of donor values.

As mentioned earlier, one concern is that each imputation cell had a wide range of income values. Other practical issues concerning hot deck and mean imputations are that though we had a big dataset,

but still many imputation cells were empty. For example, the imputation cell corresponding to age – 60+ years; education – HS degree; gender – male; race – not white; work status – unemployed. So, if new surveys are added to this 'Population' and imputation requires use of the above mentioned cell, then there will not be any donor value. Many imputation cells had only one value. For example, the imputation cell corresponding to age – 18-29 years; education – HS degree; gender – male; race – not white; work status – off work last week. Again, if new surveys require the use of this particular cell for imputation, then the same value will be used multiple times.

Similar concerns are present regarding regression imputation. Though there are sixteen possibilities of information missing for marital status, party affiliation, region, and work status is missing, ranging from none missing to all missing, as mentioned before we only had to deal with three of these: (1) no information missing; (2) information on only marital status missing; (3) information on only party missing. So, we had models with no data, such as with marital status, region observed, and work status, party affiliation unobserved. In our 'Population' we had one case with only region observed and the rest unobserved. Fortunately, this case was not in the 'Target' dataset. But there will be problems if new surveys are added and imputation requires the use of such models with no data or only one datum.

We saw that regression imputation performed best when income was missing at random. The average absolute deviation of imputed income from actual income was $4,732.62, lowest among the three schemes of imputation. Even when income was missing with higher percentage for extreme income values, the average absolute deviation of imputed income from actual income for regression imputation was lowest among the three schemes of imputation at $6,321.14. However, from the results of paired t-tests we noticed that mean imputation within adjustment cells might do better than regression imputation in case of extreme values of income are missing at greater percentage. This is quite expected since when most of the extreme income values are missing, regression models are not as good and predicted or imputed income values are off mark consistently.

Finally, weighted values obtained using the standard weighting variables (age, education, gender, have kids or not, race and region) and imputed income showed that the magnitude of difference between the estimates and standard errors is small.

Thus, this study suggests that the use of imputed income as a weighting variable for social science surveys may not be worth the effort. However, imputed income values were quite close to actual values overall and hence could be used for other kinds of analyses.

ACKNOWLEDGEMENTS

This work was partially supported by a summer research award from the Center for Survey Research at the Ohio State University to the first author. We thank Dr. Gerald Kosicki and the Center for Survey Research for allowing us to use the BSP economic surveys data.

APPENDIX

A1: Calculation of the ODSI

Three survey items were used by Dunn, Stec, Lavrakas and Kim (2000) to build ODSI. They are (1) amount of worry about debt; (2) amount of stress from debt; (3) concern that they will never be able to pay off the debt. Each component is measured on a five-point scale from 0 to 4, with 0 indicating no stress and 4 indicating high stress. Each value for the households is then multiplied by 8.3334 to distribute the index across a 0-100 scale, with high scores indicting high stress.

A2: Income distribution for Ohio population (U.S. CPS March Supplement 2000)

<$10000	414793
$10000 to $ 14999	305279
$15000 to $ 19999	414575
$ 20000 to $24999	295205
$25000 to $29999	309145
$30000 to $34999	294608
$35000 to $39999	253618
$40000 to $44999	241773
$45000 to $49999	269707
$50000 to $59999	413921
$60000 to $74999	404378
$75000 to $99999	406099
>$100000	507622

A3: Categories of selected demographic variables

Age: 18-29 yrs 30-44 yrs 45-50 yrs 60+ yrs	**Education**: Less than high school degree High School graduate Some college College graduate
Gender: Female Male	**Have kids or not**: None At least one
Race: White Other	**Party Affiliation**: Democrat Republican Other
Region: Columbus Cleveland Cincinnati Other	**Income**: Less than $20,000 $20,000 - $39,999 $40,000 - $74,999 $75,000 or more
Marital Status: Married Cohabitating Divorced Separated Single or never married Widow or Widower	**Work Status:** Full-time Part-time Off work last week Unemployed Retired In school Keeping house Other

A4: Variables used in Regression Models for 'Target' dataset

	Regression code	Categories of Demographic variables used in the model
Random	(1)	Age, Education, Have kids or not, Race, Marital Status, Work Status, Party Affiliation, Region, Number of adults in the household
	(3)	Age, Race
	(5)	Education, Marital Status, Work Status, Gender, Number of adults in the household
Extreme	(1)	Age, Education, Gender, Have kids or not, Race, Marital Status, Work Status, Party Affiliation, Region, Number of adults in the household
	(3)	Education, Work Status, Party Affiliation
	(5)	Age, Education, Gender, Marital Status, Work Status, Number of adults in the household

(1) = None of the variables region, party affiliation, marital status, work status missing

(3) = Only information on marital status missing

(5) = Only information on party affiliation missing

Out of the 1,865 cases in the 'Target' dataset there were 1814 cases corresponding to regression code (1), 2 cases for (3) and 48 cases for (5)

Fig 1

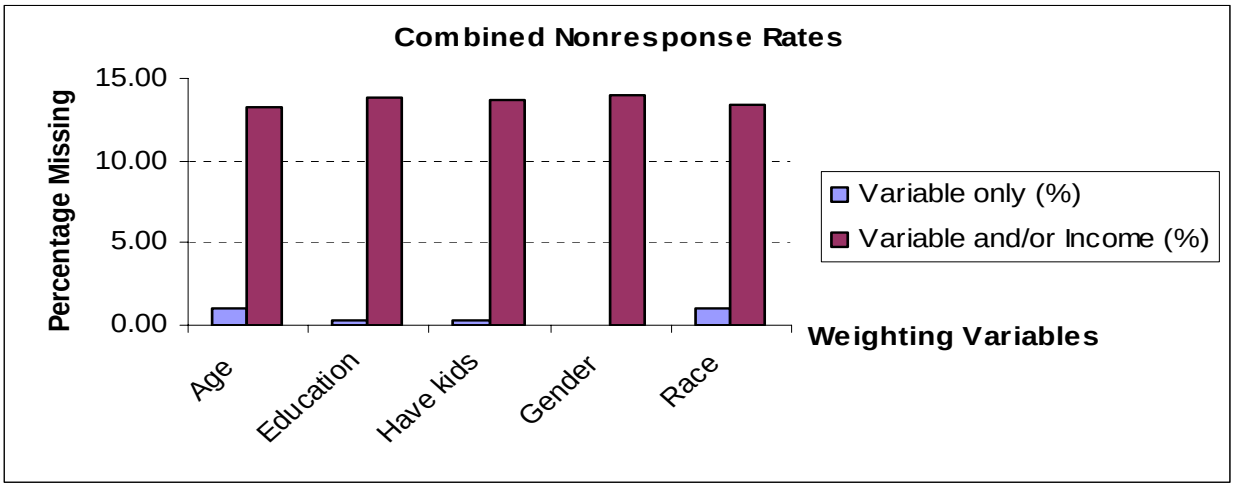

Combined Nonresponse Rates
Percentage Missing
15.00
10.00
5.00
0.00
Age
Education
Have kids
Gender
Race
Weighting Variables
Variable only (%)
Variable and/or Income (%)

Fig 2

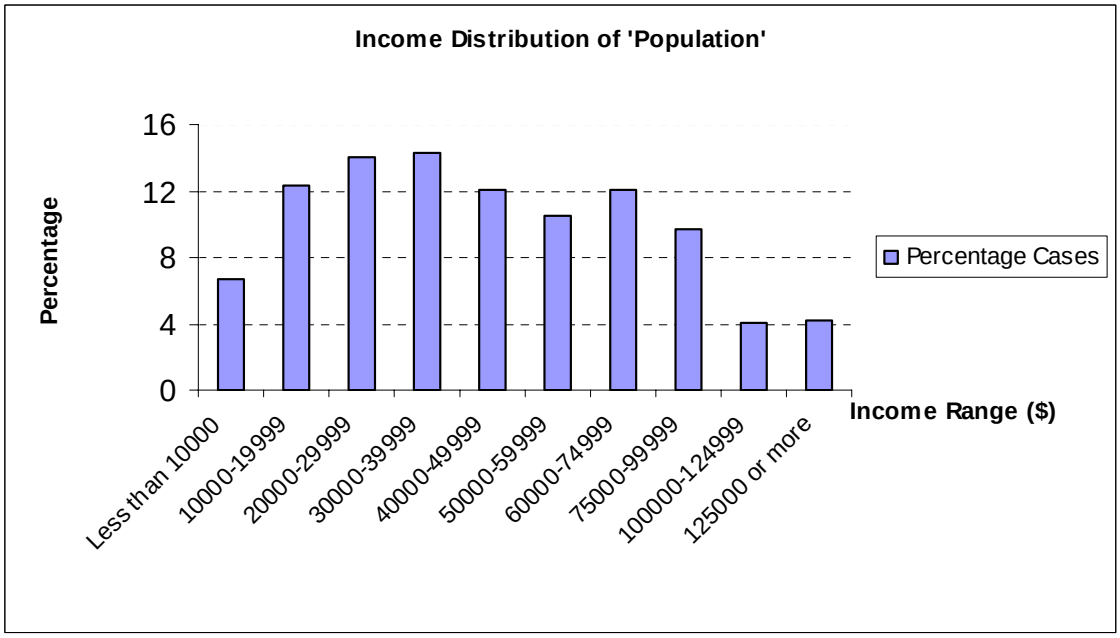

Income Distribution of 'Population'
Percentage
16
12
8
4
0
Less than 10000
10000-19999
20000-29999
30000-39999
40000-49999
50000-59999
60000-74999
75000-99999
100000-124999
125000 or more
Income Range ($)
Percentage Cases

Fig 3

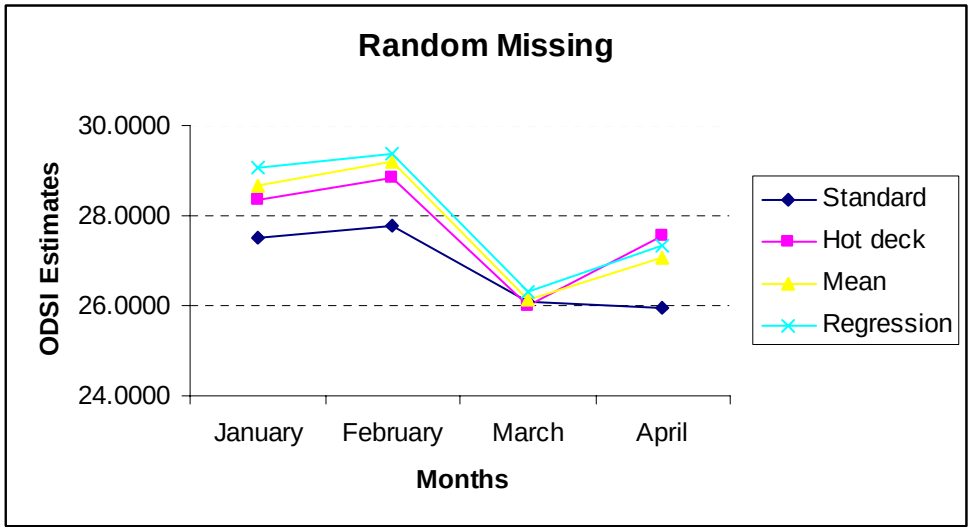
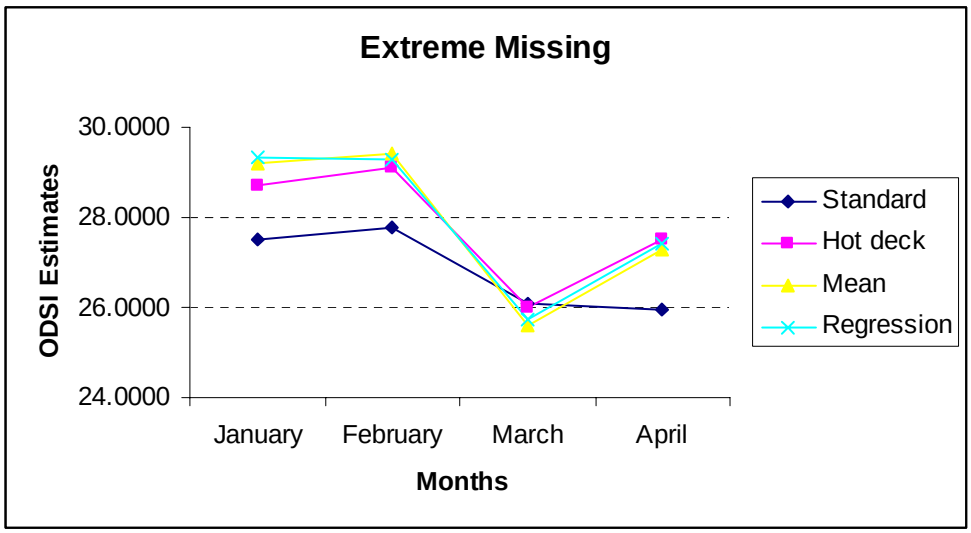

Random Missing
30.0000
28.0000
26.0000
24.0000
ODSI Estimates
January
February
March
April
Months
Standard
Hot deck
Mean
Regression

Fig 4


Extreme Missing
30.0000
28.0000
26.0000
24.0000
ODSI Estimates
January
February
March
April
Months
Standard
Hot deck
Mean
Regression

References

OLSON, L., RODÉN, S., DENNIS, M., CANNAROZI, F., and WRIGHT, R.A., (1999), *Alternative Methods of Obtaining Family Income in RDD Surveys,* ASA Proceedings of the Section on Survey Research Methods, 940-945.

MOYER, L.H., FANSLER, N.E., LEE, M.A., and VON THURN, D., (1997), *How Do People Answer Income Questions,* ASA Proceedings of the Section on Survey Research Methods, 870-874.

ROTHGEB, J., and COHANY, S., (1992), *The Revised CPS Questionnaire: Differences Between the current and the Proposed Questionnaire,* ASA Proceedings of the Section on Survey Research Methods, 649-654.

KÖRMENDI, E., (1988), *The Quality of Income Information in Telephone and Face to Face Surveys,* Telephone Survey Methodology, eds. R. GROVES, et al., New York: Wiley.

MOORE, J., STINSON, L., and WELNIAK, E., (1997), *Income Reporting in Surveys: Cognitive Issues and measurement Error,* Paper presented at the 2nd conference on Cognitive Aspects of Survey Methodology (CASM), 947-952.

DUNN, L., LAVRAKAS, P.J., STEC, J., and KIM, T.H., (2000), *A Debt Stress Index for Measuring the Stress Associated with One's Total Debt,* ASA Proceedings, 78-81.

LITTLE, R.J.A., and SAMUHEL, M.E., (1983), *Alternative Models for CPS Income Imputation,* ASA Proceedings of the Section on Survey Research Methods, 415-420.

SMITH, T.W., (1991), *An Analysis of Missing Income Information on the General Social Surveys,* GSS Methodological Report 71, Chicago: NORC.

Current Population Survey, Technical Paper 63, *Design and Methodology,* U.S. Census Bureau, Bureau of Labor Statistics.

MOORE, J.C., and LOOMIS, L.S., (2000), *Using Alternative Question Strategies to reduce Income Nonresponse,* ASA Proceedings of the Section on Survey Research Methods, 947-951.

LITTLE, R.J.A., and RUBIN, D.B., (1987), *Statistical Analysis with Missing Data,* New York: John Wiley & Sons.

BRICK, J.M., and KALTON, G., (1996), *Handling Missing Data in Survey Research,* Statistical methods in Medical Research, 215-238.