

## Methods for Weighting the 2005 Sandusky County Assessment Data

Data from sample surveys have the potential for bias if there are different rates of response for different segments of the population. In other words, some subgroups of the population may be more represented in the completed surveys than they are in the population from which those surveys are sampled. If a sample has 25% of its respondents being male and 75% being female, then the sample is biased towards the views of males (if males respond differently than females). This same phenomenon holds true for any possible characteristic that may alter how an individual responds to the survey items.

In some cases, the procedures of the survey methods may purposefully over-sample a segment of the population in order to gain an appropriate number of responses from that subgroup for appropriate data analysis when investigating them separately (this is often done for minority groups). Whether the over-sampling is done inadvertently or purposefully, the data needs to be weighted so that the proportioned characteristics of the sample accurately reflect the proportioned characteristics of the population. In the 2005 Sandusky County survey, a weighting was applied prior to the analysis that weighted the survey respondents to reflect the actual distribution of Sandusky County based on age, sex, race, and income.

Weightings were created for each category within sex (male, female), race (White, Non-White), Age (7 different age categories), and income (7 different income categories). The numerical value of the weight for each category was calculated by taking the percent of the Sandusky County within the specific category and dividing that by the percent of the sample within that same specific category. Using sex as an example, the following represents the data from the 2005 Sandusky County Survey and actual values based on the 2000 Census.

<u>Sex</u>	<u>2005 Sandusky Survey</u>		<u>2000 Sandusky County Census</u>		<u>Weight</u>
	<u>Number</u>	<u>Percent</u>	<u>Number</u>	<u>Percent</u>	
Male	256	46.97248	30242	48.94161	1.0419210
Female	289	53.02752	31550	51.05839	0.9628658

In this example, it shows that there was a slightly larger portion of females in the sample compared to the actual portion in Sandusky County. The weighting for males was calculated by taking the percent of males in Sandusky County (based on Census information) (48.94161%) and dividing that by the percent found in the 2005 Sandusky County sample (46.97248%) [ $48.94161 / 46.97248\% =$  weighting of **1.0419210** for males]. The same was done for females [ $51.05839 / 53.02752 =$  weighting of **0.9628658** for females]. Thus males' responses are weighted heavier by a factor of 1.0419210 and females' responses weighted less by a factor of 0.9628658.

## Methods for Weighting the 2005 Sandusky County Assessment Data

This same thing was done for each of the 18 specific categories as described above. For example, a respondent who was Male, White, in the age category 35-44, and with a household income in the \$50-\$75k category would have an individual weighting of 1.4312435 [1.041921006 (weight for males) x 1.001164941 (weight for White) x 1.324514139 (weight for age 35-44) x 1.035896888 (weight for income \$50-\$75k)]. Thus, each individual in the 2005 Sandusky County sample has their own individual weighting based on their combination of age, race, sex, and income. See next page for each specific weighting and the numbers from which they were calculated.

Multiple sets of weightings were created and used in the statistical software package (SPSS 12.0) when calculating frequencies. For analyses done for the entire sample and analyses done based on subgroups other than age, race, sex, or income – the weightings that were calculated based on the product of the four weighting variables (age, race, sex, income) for each individual. When analyses were done comparing groups within one of the four weighting variables (e.g., smoking status by race/ethnicity), that specific variable was not used in the weighting score that was applied in the software package. In the example smoking status by race, the weighting score that was applied during analysis included only age, sex, and income. Thus a total of eight weighting scores for each individual were created and applied depending on the analysis conducted. The weight categories were as follows:

- 1) **Total weight** (product of 4 weights) – for all analyses that did not separate age, race, sex, or income.
- 2) **Weight without sex** (product of age, race, and income weights) – used when analyzing by sex.
- 3) **Weight without age** (product of sex, race, and income weights) – used when analyzing by age.
- 4) **Weight without race** (product of age, sex, and income weights) – used when analyzing by race.
- 5) **Weight without income** (product of age, race, and sex weights) – used when analyzing by income.
- 6) **Weight without sex or age** (product of race and income weights) – used when analyzing by sex and age.
- 7) **Weight without sex or race** (product of age and income weights) – used when analyzing by sex and race.
- 8) **Weight without sex or income** (product of age and race weights) – used when analyzing by sex and income.

## Methods for Weighting the 2005 Sandusky County Assessment Data

Category	Sandusky Sample	%	2000 Census Data	%	Weighting Value
<b>Sex:</b>					
Male	256	46.97248	30,242	48.9416106	1.041921006
Female	289	53.02752	31,550	51.0583894	0.962865822
<b>Age:</b>					
20-24	21	3.846154	3,413	7.748013621	2.014483541
25-34	64	11.72161	7,637	17.33711691	1.479072787
35-44	92	16.84982	9,831	22.31782066	1.324514139
45-54	107	19.59707	8,586	19.49148695	0.994612325
55-59	50	9.157509	3,054	6.933030647	0.757086947
60-64	48	8.791209	2,587	5.872871737	0.66803916
65-74	72	13.18681	4,601	10.44494892	0.792075293
75-84	68	12.45421	3,181	7.221339387	0.579831074
85+	24	4.395604	1,160	2.633371169	0.599091941
<b>Race:</b>					
White	501	92.09559	56,974	92.20287416	1.001164941
Other	43	7.904412	4,818	7.797125842	0.986427083
<b>Household Income</b>					
Less than \$10,000	36	7.2	1,533	6.473270839	0.899065394
\$10k-\$15k	38	7.6	1,580	6.671733806	0.877859711
\$15k-\$25k	78	15.6	3,040	12.83675365	0.822868824
\$25k-\$35k	77	15.4	3,644	15.38721392	0.999169735
\$35k-\$50	94	18.8	4,692	19.81251583	1.053857225
\$50k-\$75k	109	21.8	5,348	22.58255215	1.035896888
\$75k or more	68	13.6	3,845	16.2359598	1.193820574
<p>Note: The weighting ratios are calculated by taking the ratio of the proportion of the population of Sandusky County (determined from 2000 Census data) in each subcategory by the proportion of the sample in the Sandusky County survey for that same category.</p>					