Analytics Recommendations for HINTS 5, Cycle 1 Data

May, 2020

Table of Contents

Overview of HINTS

The Health Information National Trends Survey (HINTS) is a nationally-representative survey which has been administered every few years by the National Cancer Institute since 2003. The HINTS target population is all adults aged 18 or older in the civilian non-institutionalized population of the United States. The HINTS program collects data on the American public's need for, access to, and use of health-related information and health-related behaviors, perceptions and knowledge. (Hesse, et al., 2006; Nelson, et al., 2004). Previous iterations include HINTS 1 (2003), HINTS 2 (2005), HINTS 3 (2007/2008), HINTS 4 Cycle 1 (2011/2012), HINTS 4 Cycle 2 (2012/2013), HINTS 4 Cycle 3 (Late 2013), HINTS 4, Cycle 4 (2014), and HINTS-FDA, Cycle 1 (2015).

HINTS 5

The HINTS 5 administration includes four data collection cycles over four years starting in 2017. The first of these cycles (HINTS 5, Cycle 1) was conducted from January through May 2017, and is the focus of this report. HINTS 5 draws upon the lessons learned from prior iterations of HINTS. A single-mode mail survey was implemented for HINTS 5, Cycle 1. For more extensive background about the HINTS program and previous data collection efforts, see Finney Rutten et al. (2012).

Methodology

Data collection for Cycle 1 of HINTS 5 was initiated in January 2017 and concluded in May 2017. HINTS 5, Cycle 1 was a self-administered mailed questionnaire. The sampling frame of addresses, provided by Marketing Systems Group (MSG), was grouped into three strata: 1) addresses in areas with high concentrations of minority population; 2) addresses in areas with low concentrations of minority population; and 3) addresses located in counties comprising Central Appalachia regardless of minority population. All non-vacant residential addresses in the United States present on the MSG database, including post office (P.O.) boxes, throwbacks (i.e., street addresses for which mail is redirected by the United States Postal Service to a specified P.O. box), and seasonal addresses, were subject to sampling. The protocol for mailing the questionnaires involved an initial mailing of the questionnaire, followed by a reminder postcard, and up to two additional mailings of the questionnaire as needed for non-responding households. Most households received one survey per mailing (in English), while households that were potentially Spanish-speaking received two surveys per mailing (one in English and one in Spanish). The second-stage of sampling consisted of selecting one adult within each sampled household using the Next Birthday Method. In this method, the adult who would have the next birthday in the sampled household was asked to complete the questionnaire. A $2 monetary incentive was included with the survey to encourage participation. Refer to the HINTS 5, Cycle 1 Methodology Report for more extensive information about the sampling procedures.

Sample Size and Response Rates

The final HINTS 5, Cycle 1 sample consists of 3,285 respondents. Note that 94 of these respondents were considered partial completers who did not answer the entire survey. A questionnaire was considered to be complete if at least 80% of Sections A and B were answered. A questionnaire was considered to be partially complete if 50% to 79% of the questions were answered in Sections A and B. Household response rates were calculated using the American Association for Public Opinion Research response rate 2 (RR2) formula. The overall household response rate using the Next Birthday method was 32.39%.

Analyzing HINTS Data

If you are solely interested in calculating point estimates (means, proportions etc.), either weighted or unweighted, you can use programs including SAS, SPSS, STATA, and Systat. If you plan on doing inferential statistical testing using the data (i.e., anything that involves calculating a p value or confidence interval), it is important that you utilize a statistical program that can incorporate the replicate weights that are included in the HINTS database. The issue is that the standard errors in your analyses will most likely be underestimated if you don't incorporate the jackknife replicate weights; therefore, your p-values will be smaller than they "should" be, your tests will be more liberal, and you are more likely to make a type I error. Statistical programs like SUDAAN, STATA, SAS and Wesvar can incorporate the replicate weights found in the HINTS database.

With the latest release of HINTS 5, Cycle 1, the SPSS dataset contains variance codes that will allow for inferential statistical testing using Taylor Series Linearization along with the Complex Samples module. Please see the “Important Analytic Variables in the Database” section for more information about the variance codes, and the “Variance Estimation Methods: Replicate vs. Taylor Linearization” section for more information about the two variance estimation methods.

Note that analyses of HINTS variables that contain a large number of valid responses usually produce reliable estimates, but analyses of variables with a small number of valid responses may yield unreliable estimates, as indicated by their large variances. The analyst should pay particular attention to the standard error and coefficient of variation (relative standard error) for estimates of means, proportions, and totals, and the analyst should report these when writing up results. It is important that the analyst realizes that small sample sizes for particular analyses will tend to result in unstable estimates.

Important Analytic Variables in the Database

Note: Refer to the HINTS 5, Cycle 1 Methodology Report for more information regarding the weighting and stratification variables listed below.

PERSON_FINWT0: Final sample weight used to calculate population estimates. Note that estimates from the 2015 American Community Survey (ACS) of the US Census Bureau were used to calibrate the HINTS 5, Cycle 1 control totals with the following variables: Age, gender, education, marital status, race, ethnicity, and census region. In addition, variables from the 2016 National Health Interview Survey (NHIS) were used to calibrate HINTS 5, Cycle 1 data control totals regarding: Percent with health insurance and percent ever had cancer.

PERSON_FINWT1 THROUGH PERSON_FINWT50: Fifty replicate weights that can be used to calculate accurate standard error of estimates using the jackknife replication method. More information about how these weights were created can be found in the “HINTS 5, Cycle 1 Methodology Report” included in the data download, or see Korn and Graubard (1999).

VAR_STRATUM: This variable identifies the first-stage sampling stratum of a HINTS sample for a given data collection cycle. It is the variable assigned to the STRATA parameter when specifying the sample design to compute variances using the Taylor Series linearization method. It has two values: high minority (HM) and low minority (LM).

VAR_CLUSTER: This variable identifies the cluster of sampling units of a HINTS sample for a given data collection cycle used for estimating variances. It is the variable assigned to the CLUSTER parameter when specifying the sample design to compute variances using the Taylor Series linearization method. It has values ranging from 1 to 50.

STRATUM: This variable codes for whether the respondent was in the Low or High Minority Area sampling stratum.

HIGHSPANLI: This variable codes for whether the respondent was in the High Spanish Linguistically Isolated stratum (Yes or No).

HISPSURNAME: This variable codes for whether there was a Hispanic surname match for this respondent (Yes or No).

HISP_HH: This variable codes for households identified as Hispanic by either being in a high linguistically isolated strata, or having a Hispanic surname match, or both.

APP_REGION: This variable codes for Appalachia subregion.

FORMTYPE: This variable codes for the type of survey completed (Long or Short form).

LANGUAGE_FLAG: This variable codes for language the survey was completed in (English or Spanish).

QDISP: This variable codes for whether the survey returned by the respondent was considered Complete or Partial Complete. A complete questionnaire was defined as any questionnaire with at least 80% of the required questions answered in Sections A and B. A partial complete was defined as when between 50% and 79% of the questions were answered in Sections A and B. There were 148 partially complete questionnaires. Fifty-one questionnaires with fewer than 50% of the required questions answered in Sections A and B were coded as incompletely-filled out and discarded.

INCOMERANGES_IMP: This is the income variable (INCOMERANGES) imputed for missing data. To impute for missing items, PROC HOTDECK from the SUDAAN statistical software was used. PROC HOTDECK uses the Cox-Iannacchione Weighted Sequential Hot Deck imputation method as described by Cox (1980). The following variables were used as imputation classes given their strong association with the income variable: Education (O6), Race/Ethnicity (RaceEthn), Do you currently rent or own your house? (O15), How well do you speak English? (O9), and Were you born in the United States? (O7).

Variance Estimation Methods: Replicate vs. Taylor Linearization

Variance estimation procedures have been developed to account for complex sample designs. Taylor series (linear approximation) and replication (including jackknife and balanced repeated replication, BRR) are the most widely used approaches for variance estimation. Either of these techniques allow the analyst to appropriately reflect factors such as the selection of the sample, differential sampling rates to subsample a subpopulation, and nonresponse adjustments in estimating sampling error of survey statistics. Both procedures have good large sample statistical properties, and under most conditions, these procedures are statistically equivalent. Wolter (2007) is a useful reference on the theory and applications of these methods.

The HINTS 5, Cycle 1 datasets include variance codes and replicate weights so analysts can use either Taylor Series or replication methods for variance estimation. The following points may provide some guidance regarding which method will best reflect the HINTS sample design in your analysis.

TAYLOR SERIES	REPLICATION METHODS
▪ Most appropriate for simple statistics, such as means and proportions, since the approach linearizes the estimator of a statistic and then uses standard variance estimation methods.	▪ Useful for simple statistics such as means and proportions, as well as nonlinear functions. ▪ Easy to use with a large number of variables. ▪ Better accounts for variance reduction procedures such as raking and post-stratification. However, the variance reduction obtained with these procedures depends on the type of statistic and the correlation between the item of interest and the dimensions used in raking and post-stratification. Depending on your analysis, this may or may not be an advantage.

The Taylor Series variance estimation procedure is based on a mathematical approach that linearizes the estimator of a statistic using a Taylor Series expansion and then uses standard variance methods to estimate the variance of the linearized statistic.

The replication procedure, on the other hand, is based on a repeated sampling approach. The procedure uses estimators computed on subsets of the sample, where subsets are selected in a way that reflect the sample design. By providing weights for each subset of the sample, called replicate weights, end users can estimate the variance of a variety of estimators using standard weighted sums. The variability among the replicates is used to estimate the sampling variance of the point estimator.

An important advantage of replication is that it provides a simple way to account for adjustments made in weighting, particularly those with variance-reducing properties, such as weight calibration procedures. (See Kott, 2009, for a discussion of calibration methods, including raking, and their effects on variance estimation). The survey weights for HINTS were raked to control totals in the final step of the weighting process. However, the magnitude of the reduction generally depends on the type of estimate (i.e., total, proportion) and the correlation between the variable being analyzed and the dimensions used in raking.

Although SPSS's estimates of variance based on linearization take into account the sample design of the survey, they do not properly reflect the variance reduction due to raking. Thus, when comparing across Taylor series and replicate methods, analyses with Taylor series tend to have larger standard errors and generally provide more conservative tests of significance. The difference in the magnitude of standard errors between the two methods, however, will be smaller when using analysis variables that have little to no relationship with the raking variables.

Denominator Degrees of Freedom (DDF)

Replicate Weights: The HINTS 5, Cycle 1 database contains a set of 50 replicate weights to compute accurate standard errors for statistical testing procedures. These replicate weights were created using a jackknife minus one replication method; when analyzing one iteration of HINTS data, the proper denominator degrees of freedom (ddf) is 49. Thus, analysts who are only using the HINTS 5, Cycle 1 data should use 49 ddf in their statistical models. HINTS statistical analyses that involve more than one iteration of data will typically utilize a set of 50*k replicate weights, where they can be viewed as being created using a stratified jackknife method with k as the number of strata, and 49*k as the appropriate ddf. Analysts who were merging two iterations of data and making comparisons should adjust the ddf to be 98 (49*2) etc.

Taylor Series: The HINTS 5, Cycle 1 database contains two variables that can be used to calculate standard errors using the Taylor series, namely VAR_STRATUM and VAR_CLUSTER (see VAR_STRATUM and VAR_CLUSTER variables in the previous section for strata definitions.). The degrees of freedom for the Taylor series, 98, is based on 50 PSUs in each of the two sampling strata (#psus - #strata = 50*2 – 2 = 98).

References

Cox, B. G. (1980). “The Weighted Sequential Hot Deck Imputation Procedure”. Proceedings of the American Statistical Association, Section on Survey Research Methods.

Finney Rutten, L. J., Davis, T., Beckjord, E. B., Blake, K., Moser, R. P., & Moser, R. P. (2012) Picking Up the Pace: Changes in Method and Frame for the Health Information National Trends Survey (2011 – 2014). Journal of Health Communication, 17 (8), 979-989..

Hesse, B. W., Moser, R. P., Rutten, L. J., & Kreps, G. L. (2006). The health information national trends survey: research from the baseline. *J Health Commun, 11 Suppl 1*, vii-xvi.

Korn, E. L., & Graubard, B. I. (1999). Analysis of health surveys. New York: John Wiley & Sons. Kott, P.S. (2009). Calibration Weighting: Combining Probability Samples and Linear Prediction Models. Chapter 25 in Pfeffermann, D. and Rao, C.R. (eds.) Handbook of Statistics Vol. 29B: Sample Surveys: Inference and Analysis. Elsevier: Amsterdam

Nelson, D. E., Kreps, G. L., Hesse, B. W., Croyle, R. T., Willis, G., Arora, N. K., et al. (2004). The Health Information National Trends Survey (HINTS): development, design, and dissemination. *J Health Commun, 9*(5), 443-460; discussion 481-444.

Wolter, K. (2007). Introduction to Variance Estimation. 2nd edition. Springer-Verlag: New York

Appendix

The following appendices A through D provide some coding examples using SAS, SPSS, SUDAAN, and STATA for common types of statistical analyses using HINTS 5, Cycle 1 data. For SAS and STATA, you'll see two sets of code: one when using replicate methods for variance estimation, and one for Taylor Series linearization. For replicate methods, these examples will incorporate both the final sample weight (to get population estimates) and the set of 50 jackknife replicate weights to get the proper standard error. For Taylor Series, the code will incorporate the final sample weight and the two variance codes to compute variance estimates. Although these examples specifically use HINTS 5, Cycle 1 data, the concepts used here are generally applicable to other types of analyses. We will consider an analysis that includes gender, education level (edu) and two questions that are specific to the HINTS 5, Cycle 1 data: seekcancerinfo & generalhealth.

Appendices E and F provide SAS code and Appendices G and H provide SPSS code to combine HINTS 4, Cycle 4 and HINTS 5, Cycle 1 survey iterations and HINTS 5, Cycle 1 and HINTS FDA, Cycle 2 survey iterations. The provided code will generate one final sample weight for population point estimates and 100 replicate weights to compute standard errors.

- **Appendix A:** Analyzing data using SAS
- **Appendix B:** Analyzing data using SPSS
- **Appendix C:** Analyzing data using SUDAAN
- **Appendix D:** Analyzing data using STATA
- **Appendix E:** Merging HINTS 4, Cycle 4 and HINTS 5, Cycle 1 using SAS
- **Appendix F:** Merging HINTS 5, Cycle 1 and HINTS FDA, Cycle 2 using SAS
- **Appendix G:** Merging HINTS 4, Cycle 4 and HINTS 5, Cycle 1 using SPSS
- **Appendix H:** Merging HINTS 5, Cycle 1 and HINTS FDA, Cycle 2 using SAS

Appendix A: Analyzing data using SAS

This section gives some SAS (Version 9.3 and higher) coding examples for common types of statistical analyses using HINTS 5, Cycle 1 data. We begin by doing data management of the HINTS 5 data in a SAS DATA step. We first decided to exclude all “Missing data (Not Ascertained)” and “Multiple responses selected in error” responses from the analyses. By setting these values to missing (.), SAS will exclude these responses from procedures where these variables are specifically accessed. For logistic regression modeling within the PROC SURVEYLOGISTIC procedure, SAS expects the response variable to be dichotomous with values (0, 1), so this variable will also be recoded at this point. It is better to use dummy variables instead of categorical variables in SAS survey procedures, such as PROC SURVEYREG. We use dummy variables for gender and education level in both PROC SURVEYLOGISTIC and PROC SURVEYREG procedures. When recoding existing variables, it is generally recommended to create new variables, rather than over-writing the existing variables. Note: New variables should always be compared to original source variables in a SAS PROC FREQ procedure to verify proper coding.

SAS Data Management Code: Recoding Variables and Creating and Applying New Formats

```
options fmtsearch=(hints5c1); *This is used to call up the formats;
*substitute your library name in the parentheses;

proc format;  *First create some temporary formats;

Value Genderf
1 = "Male"
2 = "Female";

Value Educationf
1 = "Less than high school"
2 = "12 years or completed high school"
3 = "Some college"
4 = "College graduate or higher";

value seekcancerinfof
1 = "Yes"
0 = "No";

Value Generalf
1 = "Excellent"
2 = "Very good"
3 = "Good"
4 = "Fair"
5 = "Poor";

run;

data hints5cycle1;
set hints5c1.hints5_cycle1_public;

/*Recode negative values to missing*/
if genderc = 1 then gender = 1;
if genderc = 2 then gender = 2;

if genderc in (-9, -6) then gender = .;
```

```
/*Recode education into four levels, and negative values to
missing*/ if education in (1, 2) then edu = 1;
if education = 3 then edu = 2;
if education in (4, 5) then edu = 3;
if education in (6, 7) then edu = 4;
if education = -9 then edu = .;

/*Recode seekcancerinfo to 0- 1 format for proc rlogist procedure,
and negative values to missing */
if seekcancerinfo = 2 then seekcancerinfo = 0;
if seekcancerinfo in (-9, -6, -2, -1) then seekcancerinfo = .;

/*Recode negative values to missing for proc regress procedure*/
if generalhealth in (-5, -9) then generalhealth = .;


/*Create dummy variables for proc surveylogistic and proc
surveyreg procedures*/
if gender = 1 then
      Female = 0;
 else if gender = 2 then
      Female = 1;

 if edu = 1 then
      do;
            HighSchool = 0;
            SomeCollege = 0;
            CollegeorMore = 0;
      end;
else if edu = 2 then
      do;
            HighSchool = 1;
            SomeCollege = 0;
            CollegeorMore = 0;
  end;

else if edu = 3 then
      do;
            HighSchool = 0;
            SomeCollege = 1;
            CollegeorMore = 0;
  end;

else if edu = 4 then
      do;
            HighSchool = 0;
            SomeCollege = 0;
            CollegeorMore = 1;
      end;

/*Apply formats to recoded variables */
format gender genderf. edu educationf. seekcancerinfo
seekcancerinfof. generalhealth generalf.;
run;
```

Replicate Weights Variance Estimation Method

Frequency Table and Chi-Square Test

We are now ready to begin using SAS 9.3 to examine the relationships among these variables. Using **PROC SURVEYFREQ**, we will first generate a cross-frequency table of education by gender, along with a (Wald) Chi-squared test of independence. Note the syntax of the overall sample weight, PERSON_FINWT0, and those of the jackknife replicate weights, PERSON_FINWT1—PERSON_FINWT50. The jackknife adjustment factor for each replicate weight is 0.98. This syntax is consistent for all procedures. Other data sets that incorporate replicate weight jackknife designs will follow a similar syntax.

```
proc surveyfreq data = hints5cycle1 varmethod = jackknife;
      weight person_finwt0;
      repweights person_finwt1-person_finwt50 / df = 49 jkcoefs =
      0.98; tables edu*gender / row col wchisq;
run;
```

The *tables* statement defines the frequencies that should be generated. Stand-alone variables listed here result in one-way frequencies, while a "*" between variables will define cross- frequencies. The *row* option produces row percentages and standard errors, allowing us to view stratified percentages. Similarly, the *col* option produces column percentages and standard errors, allowing us to view stratified percentages. The option *wchisq* requests Wald chi-square test for independence. Other tests and statistics are also available; see the [SAS 9.3 Product Documentation Site] for more information.

For the purposes of computing appropriate degrees of freedom for the estimator of the HINTS 5, Cycle 1 differences, we can assume, as an approximation, that the sample is a simple random sample of size 50 (corresponding to the 50 replicates: each replicate provides a 'pseudo sample unit') from a normal distribution. The denominator degrees of freedom (df) is equal to 49*k, where k is the number of iterations of data used in this analysis.

Variance Estimation	
Method	Jackknife
Replicate Weights	hints5cycle1
Number of Replicates	50

edu	gender	Frequency	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent	Column Percent	Std Err of Col Percent
Less than high school	Male	81	4.2623	0.6712	51.1862	4.4709	8.6849	1.3678
	Female	129	4.0647	0.5158	48.8138	4.4709	7.9821	1.0114
	Total	210	8.3271	0.9407	100			
12 years or completed high school	Male	208	10.9489	0.7187	47.5949	2.0853	22.3097	1.4589
	Female	395	12.0555	0.6198	52.4051	2.0853	23.6739	1.2118
	Total	603	23.0044	0.9449	100			

edu	gender	Frequency	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent	Column Percent	Std Err of Col Percent
Some college	Male	391	17.2157	0.6606	52.2525	1.0496	35.0789	1.3492
	Female	542	15.7314	0.3513	47.7475	1.0496	30.8924	0.6849
	Total	933	32.9470	0.7853	100			
College graduate or higher	Male	593	16.6500	0.1397	46.6107	0.2859	33.9264	0.2365
	Female	799	19.0714	0.1379	53.3893	0.2859	37.4515	0.2811
	Total	1392	35.7215	0.1885	100			
Total	Male	1273	49.0770	0.1598			100	
	Female	1865	50.9230	0.1598			100	
	Total	3138	100					

Frequency Missing =147

Wald Chi-Square Test	
Chi-Square	150.3189
F Value	50.1063
Num DF	3
Den DF	49
Pr > F	<.0001
Adj F Value	48.0611
Num DF	3
Den DF	47
Pr > Adj F	<.0001
Sample Size = 3138	

The weighted percentages above show that a greater proportion of women have at least a college degree compared to men, 19.07% vs 16.65%. The Chi-squared test of independence indicates that there is a significant difference between the educational distribution in these two groups (p-value <0.05).

Logistic Regression

This example demonstrates a multivariable logistic regression model using **PROC SURVEYLOGISTIC**; recall that the response should be a dichotomous 0-1 variable.

```
/*Multivariable logistic regression of gender and
education on SeekCancerInfo*/
proc surveylogistic data= hints5cycle1 varmethod=jackknife;
      weight person_finwt0;
      repweights person_finwt1-person_finwt50 / df=49 jkcoefs=0.98;
      model seekcancerinfo (descending) = Female HighSchool
SomeCollege CollegeorMore / tech=newton xconv=1e-8;
      contrast 'Overall model' intercept 1,
                        Female 1,
                        HighSchool 1,
                        SomeCollege 1,
                        CollegeorMore 1;
   contrast 'Overall model minus intercept' Female 1,
                        HighSchool 1,
                        SomeCollege 1,
                        CollegeorMore 1;
      contrast 'Gender' Female 1;
      contrast 'Education overall' HighSchool 1,
                        SomeCollege 1,
                        CollegeorMore 1;
run;
```

The response variable should be on the left hand side (LHS) of the equal sign in the model statement, while all covariates should be listed on the right hand side (RHS). The *descending* option requests the probability of seekcancerinfo="Yes" to be modeled. The "Male" is the reference group for gender effect while "Less than high school" is the reference group for education level effect. The option *tech=newton* requests the Newton-Raphson algorithm. The option xconv=1e-8 helps to avoid early termination of the iteration.

Variance Estimation	
Method	Jackknife
Replicate Weights	hints5cycle1
Number of Replicates	50

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.6849	0.2409	8.0856	0.0045
Female	1	0.3031	0.1290	5.5215	0.0188
HighSchool	1	0.2565	0.2219	1.3371	0.2476
SomeCollege	1	0.7525	0.2716	7.6762	0.0056
CollegeorMore	1	1.3068	0.2061	40.2209	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Female	1.354	1.052	1.743
HighSchool	1.292	0.837	1.996
SomeCollege	2.122	1.246	3.614
CollegeorMore	3.694	2.467	5.533

Contrast Test Results

Contrast	DF	Wald Chi-Square	Pr > ChiSq
Overall model	5	85.5802	<.0001
Overall model minus intercept	4	72.7390	<.0001
Gender	1	5.5215	0.0188
Education overall	3	69.3104	<.0001

To identify levels/variables that display a significant difference in response, the rule of thumb is to examine odds ratios where the confidence interval does not contain 1 (by default, SAS will use alpha=.05 to determine statistical significance; this value can be changed by the user using code). However, significance may also be garnered from the test of whether the associated beta parameter is equal to 0 (see first regression table above). According to this model, those with at least a college degree appear to be statistically more inclined to search for cancer information.

Linear Regression

This example demonstrates a multivariable linear regression model using **PROC SURVEYREG**; recall that the response should be a continuous variable. For the purposes of this example, we decided to use an outcome with five levels as a continuous variable (GENERALHEALTH). Note that higher values on GENERALHEALTH indicate poorer self-reported health status.

```
/*Multivariable linear regression of gender and education on GeneralHealth*/
proc surveyreg data= hints5cycle1 varmethod=jackknife; weight person_finwt0;
repweights person_finwt1-person_finwt50 / df=49 jkcoefs=0.98;
model generalhealth = Female HighSchool SomeCollege CollegeorMore;
contrast 'Overall model' intercept 1,
            Female 1,
            HighSchool 1,
            SomeCollege 1,
            CollegeorMore 1;
contrast 'Overall model minus intercept' Female 1,
            HighSchool 1,
            SomeCollege 1,
            CollegeorMore 1;
contrast 'Gender' Female 1;
contrast 'Education overall' HighSchool 1,
            SomeCollege 1,
            CollegeorMore 1;
```

```
run;
```

Variance Estimation	
Method	Jackknife
Replicate Weights	hints5cycle1
Number of Replicates	50

Analysis of Contrasts

Contrast	Num DF	F Value	Pr > F
Overall model	5	2002.87	<.0001
Overall model minus intercept	4	20.08	<.0001
Gender	1	0.00	0.9904
Education overall	3	26.01	<.0001

Note: The denominator degrees of freedom for the F tests is 49.

From the above table, we can see that gender is **not** associated with general health, but Edu is associated, adjusting for all variables in the model.

Estimated Regression of Coefficients

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	3.0893755	0.10034758	30.79	<.0001
Female	0.0007469	0.06170539	0.01	0.9904
HighSchool	-0.2810533	0.09438473	-2.98	0.0045
SomeCollege	-0.4841933	0.10992542	-4.40	<.0001
CollegeorMore	-0.7618133	0.09508149	-8.01	<.0001

From the above table, compared to those respondents with less than a high school education, those with a high school education have a significantly inverse linear association with general health (i.e. better reported health), controlling for all variables in the model. This association also applies to those with some college, and college or higher. We don't interpret the Gender variable because it is non-significant.

Taylor Series Linearization Variance Estimation Method

Frequency Table and Chi-Square Test

We are now ready to begin using SAS 9.3 to examine the relationships among these variables. Using **PROC SURVEYFREQ**, we will first generate a cross-frequency table of education by gender, along with a (Wald) Chi-squared test of independence. Note the syntax of the strata VAR_STRATUM, cluster VAR_CLUSTER, and overall sample weight TG_all_FINWT0 (no group differences). This syntax is consistent for all procedures. Other analyses that use Taylor Series approximation will follow a similar syntax.

```
proc surveyfreq data = hints5cycle1
     varmethod = TAYLOR;
     strata VAR_STRATUM;
     cluster VAR_CLUSTER;
     weight PERSON_finwt0;
     tables edu*gender / row col wchisq;
run;
```

The *tables* statement defines the frequencies that should be generated. Standalone variables listed here result in one-way frequencies, while a "*" between variables will define cross-frequencies. The *row* option produces row percentages and standard errors, allowing us to view stratified percentages. Similarly, the *col* option produces column percentages and standard errors, allowing us to view stratified percentages. The option *wchisq* requests Wald chi-square test for independence. Other tests and statistics are also available; see the SAS 9.3 Product Documentation Site for more information.

Variance Estimation	
Method	Taylor Series
Number of Strata	2
Number of Clusters	100

edu	gender	Frequency	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent	Column Percent	Std Err of Col Percent
Less than high school	**Male**	81	4.2623	0.6229	51.1862	5.1087	8.6849	1.2280
	Female	129	4.0647	0.5632	48.8138	5.1087	7.9821	1.0927
	Total	210	8.3271	0.8267	100.0000			
12 years or completed high school	**Male**	208	10.9489	0.9928	47.5949	2.9720	22.3097	1.9321
	Female	395	12.0555	0.8943	52.4051	2.9720	23.6739	1.5794
	Total	603	23.0044	1.3125	100.0000			
Some college	**Male**	391	17.2157	1.3501	52.2525	2.7950	35.0789	2.3162
	Female	542	15.7314	0.9351	47.7475	2.7950	30.8924	1.6626
	Total	933	32.9470	1.3860	100.0000			
College graduate or higher	**Male**	593	16.6500	0.8742	46.6107	1.7725	33.9264	1.9181

edu	gender	Frequency	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent	Column Percent	Std Err of Col Percent
	Female	799	19.0714	0.8330	53.3893	1.7725	37.4515	1.3581
	Total	1392	35.7215	1.1525	100.0000			
Total	**Male**	1273	49.0770	1.3513			100.0000	
	Female	1865	50.9230	1.3513			100.0000	
	Total	3138	100.0000					

Frequency Missing = 147

Wald Chi-Square Test	
Chi-Square	3.3479
F Value	1.1160
Num DF	3
Den DF	98
Pr > F	0.3464
Adj F Value	1.0932
Num DF	3
Den DF	96
Pr > Adj F	0.3559
Sample Size = 3138	

The weighted percentages above show that a greater proportion of women have at least a college degree compared to men, 19.07% vs 16.65%. The Chi-squared test of independence indicates that there is a significant difference between the educational distribution in these two groups (p-value <0.05). The Chi-squared test of independence statistic and associated p value suggest that one should accept the null hypothesis that the two variables are not associated, which indicates that there is not a significant difference between the distributions of educational attainment for these two groups.

Logistic Regression

This example demonstrates a multivariable logistic regression model using **PROC SURVEYLOGISTIC**; recall that the response should be a dichotomous 0-1 variable.

```
/*Multivariable logistic regression of gender and education on SeekCancerInfo*/
proc surveylogistic data= hints5cycle1 varmethod=TAYLOR;
  strata VAR_STRATUM;
  cluster VAR_CLUSTER;
  weight PERSON_finwt0;
  class edu (ref="Less than high school")
        gender (ref="Male")/param=REF;
  model seekcancerinfo (descending) = gender edu /tech=newton xconv=1e-8 CLPARM
  EXPB;
run;
```

The response variable should be on the left-hand side (LHS) of the equal sign in the model statement, while all covariates should be listed on the right-hand side (RHS). The *descending* option requests the probability of seekcancerinfo="Yes" to be modeled. The "Male" is the reference group for gender effect, while "Less than high school" is the reference group for education level effect. The option *tech=newton* requests the Newton-Raphson algorithm. The option *xconv=1e-8* helps to avoid early termination of the iteration.

Variance Estimation	
Method	Taylor Series
Number of Strata	2
Number of Clusters	100

Type 3 Analysis of Effects				
Effect	**F Value**	**Num DF**	**Den DF**	**Pr > F**
gender	6.07	1	98	0.0155
edu	24.51	3	96	<.0001

Analysis of Maximum Likelihood Estimates

Parameter		**Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**95% Confidence Limits**	
Intercept		-0.6849	0.2396	-2.86	0.0052	-1.1603	-0.2095
gender	**Female**	0.3031	0.1230	2.46	0.0155	0.0589	0.5473
edu	**12 years or completed high school**	0.2565	0.2302	1.11	0.2679	-0.2004	0.7134
edu	**College graduate or higher**	1.3068	0.2249	5.81	<.0001	0.8605	1.7531
edu	**Some college**	0.7525	0.2690	2.80	0.0062	0.2187	1.2863

Odds Ratio Estimates

Effect	**Point Estimate**	**95% Confidence Limits**	
Female	1.354	1.061	1.729
HighSchool	1.292	0.818	2.041
Some College	2.122	1.244	3.619
College or More	3.694	2.364	5.772

To identify levels/variables that display a significant difference in response, the rule of thumb is to examine odds ratios where the confidence interval does not contain 1 (by default, SAS will use alpha=.05 to determine statistical significance; this value can be changed by the user using code). However, significance may also be garnered from the test of whether the associated beta parameter is equal to 0 (see "Analysis of Maximum Likelihood Estimates" table above). According to this model, those with at least a college degree appear to be statistically more inclined to search for cancer information.

Linear Regression

This example demonstrates a multivariable linear regression model using **PROC SURVEYREG**; recall that the response should be a continuous variable. For the purposes of this example, we decided to use an outcome with five levels as a continuous variable (GENERALHEALTH). Note that higher values on GENERALHEALTH indicate poorer self-reported health status.

```
/*Multivariable linear regression of gender and education on GeneralHealth*/
proc surveyreg data= hints5cycle1 varmethod=TAYLOR;
  strata VAR_STRATUM;
  cluster VAR_CLUSTER;
  weight PERSON_FINWT0;
  class edu (ref="Less than high school") gender (ref="Male");
  model generalhealth = edu gender/solution;
run;
```

Variance Estimation	
Method	Taylor Series
Number of Strata	2
Number of Clusters	100

Test of Model Effects

Effect	Num DF	F Value	Pr > F
Model Minus Intercept	4	26.89	<.0001
Intercept	1	5571.94	<.0001
Education	3	32.33	<.0001
Gender	1	0.00	0.9891

Note: The denominator degrees of freedom for the F tests is 98.

From the above table, we can see that gender is **not** associated with general health, but Edu is associated, adjusting for all variables in the model.

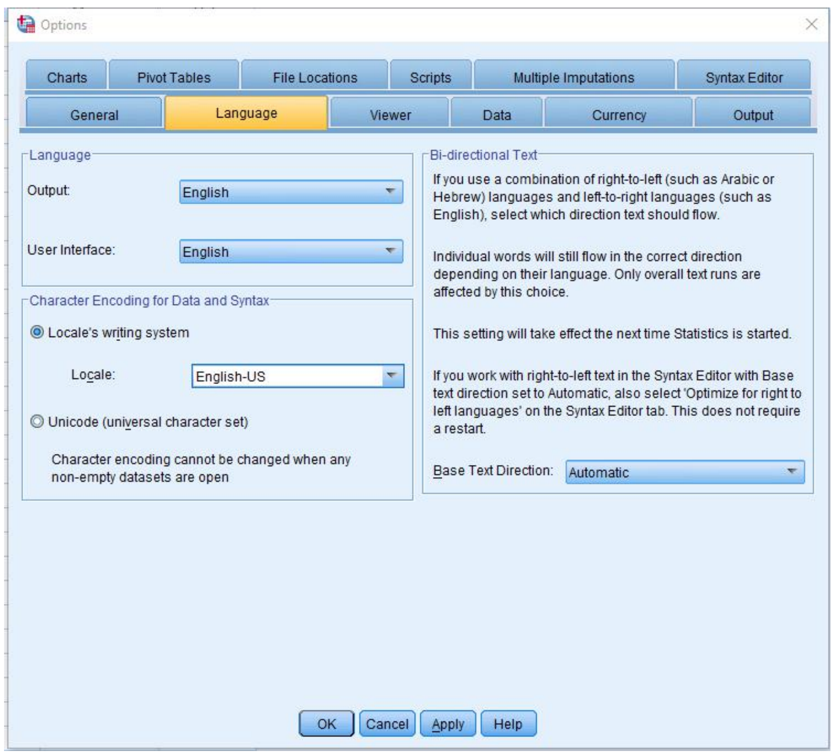
Estimated Regression of Coefficients

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	3.0893755	0.10509687	29.40	<.0001
High School	-0.2810533	0.10191831	-2.76	0.0069
Some College	-0.4841933	0.11045901	-4.38	<.0001
College or More	-0.7618133	0.09543618	-7.98	<.0001
Female	0.0007469	0.05466453	0.01	0.9891

From the above table, compared to those respondents with less than a high school education, those with a high school education have a significantly inverse linear association with general health (i.e. better reported health), controlling for all variables in the model. This association also applies to those with some college, and college or higher. We don't interpret the Gender variable because it is non-significant.

Appendix B: Analyzing data using SPSS—Taylor Series

Prior to opening the HINTS 5, Cycle 3 SPSS data, it is important to ensure that your SPSS environment is set up to be compatible with the dataset. Specifically, the language encoding (i.e., the way that character data are stored and accessed) must match between your environment and the dataset. We recommend locale encoding in U.S. English over Unicode encoding. To ensure compatibility, you must update the language encoding manually through the graphic user interface (GUI). In a new SPSS session, from the empty dataset window, select “Edit” > “Options…” from the menu bar. In the pop-up box, select the “Language” tab. In this tab, look for the “Character Encoding for Data and Syntax” section. Select the “Locale’s writing system” option and English-US or en-US from the “Locale:” dropdown list. “English-US” and “en-US” from the drop down are the common aliases used by SPSS to describe U.S. English encoding; if you do not see these specific aliases verbatim, choose the English alias that is most similar. Click “OK” to save your changes. You may now open the HINTS SPSS data without compatibility issues.



This section gives some SPSS (Version 25 and higher) coding examples for common types of statistical analyses using HINTS 5 Cycle 1 data. These examples will incorporate the stratum variable, VAR_STRATUM, and the cluster variable VAR_CLUSTER. Although these examples specifically use HINTS 5, Cycle 1 data, the concepts used here are generally applicable to other types of analyses. We will consider an analysis that includes gender, education level (edu) and two questions that are specific to the HINTS 5, Cycle 1 data: seekcancerinfo & generalhealth.

We begin by creating an analysis plan using the Complex Samples analysis procedures to specify the sample design; PERSON_FINWT0 is the sample weight variable (the final weight for the composite sample, no group differences found), VAR_STRATUM is the stratum variable, and VAR_CLUSTER is the cluster variable. The subcommand SRSESTIMATOR specifies the variance estimator under the simple random sampling assumption. The default value is WR (with replacement), and it includes the finite population correction in the variance computation. The subcommand PRINT is used to control output from CSPLAN, and the syntax PLAN means to display a summary of plan specifications. The subcommand DESIGN with keyword STRATA identifies the sampling stratification variable, and the keyword cluster CLUSTER identifies the grouping of sampling units for variance estimation. The subcommand ESTIMATOR specifies the variance estimation method used in the analysis. The syntax TYPE=WR requires the estimation method of selection with replacement.

```
* Analysis Preparation Wizard.

*substitute your library name in the parentheses of /PLAN FILE=.
CSPLAN ANALYSIS
 /PLAN FILE='(sample.csaplan)'
 /PLANVARS ANALYSISWEIGHT=PERSON_FINWT0
 /SRSESTIMATOR TYPE=WOR
 /PRINT PLAN
 /DESIGN STRATA=VAR_STRATUM CLUSTER=VAR_CLUSTER
 /ESTIMATOR TYPE=WR.
```

We completed data management of the HINTS 5 Cycle 3 data in a SPSS RECODE step. We first decided to exclude all “Missing data (Not Ascertained)” and “Multiple responses selected in error” responses from the analyses. By setting these values to missing (SYSMIS), SPSS will exclude these responses from procedures where these variables are specifically accessed. For logistic regression modeling in the CSLOGISTIC procedure, SPSS by default always uses the last (highest) level of category of the covariates as the reference, similar to SAS. Users in SPSS cannot define the reference category by themselves unless they reorder the categories to create the desired value as the reference, such as using reverse coding (see example below). To make SPSS results comparable with SAS, we reverse coded the variables in SPSS. When recoding existing variables, it is generally recommended to create new variables, rather than over-writing the existing variables. Note: New variables should always be compared to original source variables in a SPSS CROSSTABS procedure to verify proper coding.

```
*Recode negative values to missing.
DATASET ACTIVATE DataSet1.
RECODE GenderC (1=1) (2=2) (ELSE=SYSMIS) INTO gender.
VARIABLE LABELS gender 'gender'.
EXECUTE.

*Recode education into four levels, and negative values to missing.
RECODE Education (3=2) (1 thru 2=1) (4 thru 5=3) (6 thru 7=4) (ELSE=SYSMIS) INTO edu.
VARIABLE LABELS edu 'edu'.
EXECUTE.

*Recode seekcancerinfo to 0- 1 format for CSLOGISTIC procedure, and negative values to missing.
RECODE SeekCancerInfo (2=0) (1=1) (ELSE=SYSMIS) INTO seekcancerinfo_recode.
VARIABLE LABELS seekcancerinfo_recode 'seekcancerinfo_recode'.
EXECUTE.

*Recode negative values to missing for CSGLM procedure.
RECODE GeneralHealth (1 thru 5=Copy) (ELSE=SYSMIS) INTO genhealth_recode.
VARIABLE LABELS genhealth_recode 'genhealth_recode'.
EXECUTE.

*Reverse coding.
RECODE gender (1=2) (2=1) (ELSE=Copy) INTO flippedgender.
VARIABLE LABELS flippedgender 'flippedgender'.
```

```
EXECUTE.

*Reverse coding.
RECODE edu (1=4) (2=3) (3=2) (4=1) (ELSE=Copy) INTO flippededu.
VARIABLE LABELS flippededu 'flippededu'.
EXECUTE.

*Add value labels to recoded variables.
VALUE LABELS gender 1 "Male" 2 "Female".
VALUE LABELS flippedgender 2 "Male" 1 "Female".
VALUE LABELS edu 1 "Less than high school" 2 "12 years or completed high school" 3 "Some college" 4 "College graduate or higher".
VALUE LABELS flippededu 4 "Less than high school" 3 "12 years or completed high school" 2 "Some college" 1 "College graduate or higher".
VALUE LABELS seekcancerinfo_recode 1 "Yes" 0 "No".
VALUE LABELS genhealth_recode 1 "Excellent" 2 "Very good" 3 "Good" 4 "Fair" 5 "Poor".
```

Frequency Table and Chi-Square Test

We are now ready to begin using SPSS v25 to examine the relationships among these variables. Using **CSTABULATE**, we will first generate a cross-frequency table of education by gender. Note that we specify the file that contains the sample design specification using the subcommand PLAN. This syntax is consistent for all procedures. Other analyses using the same sample design will follow a similar syntax.

```
* Complex Samples Crosstabs.
CSTABULATE
/PLAN FILE=”(plan filename)”
/TABLES VARIABLES=edu BY gender
/CELLS POPSIZE ROWPCT COLPCT TABLEPCT
/STATISTICS SE COUNT
/TEST INDEPENDENCE
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

The TABLES subcommand defines the tabulation variables, where the syntax “BY” indicates the two-way crosstabulation. The CELLS subcommand specifies the summary value estimates to be displayed in the table. The *POPSIZE* option produces population size estimates for each cell and marginal. The *ROWPCT* option produces row percentages and standard errors. Similarly, the *COLPCT* option produces column percentages and standard errors. The *TABLEPCT* option produces table percentages and standard errors for each cell. The STATISTICS subcommand specifies the statistics to be displayed with the summary value estimates. The *SE* option produces the standard error for each summary value, and the *COUNT* option produces unweighted counts. The TEST subcommand specifies tests for the table. The *INDEPENDENCE* option produces the test of independence for the two-way crosstabulations. The MISSING subcommand specifies how missing values are handled. The *SCOPE* statement specifies which cases are used in the analyses. The *TABLE* option specifies that cases with all valid data for the tabulation variables are used in the analyses. The *CLASSMISSING* statement specifies whether user-defined missing values are included or excluded. The *EXCLUDE* option specifies user-defined missing values to be excluded in the analysis.

edu * gender

Edu			Gender: Male	Female	Total
Less than high school	Population Size	Estimate	10222693.573	9748881.807	19971575.380
		Standard Error	1524808.996	1387791.571	2078846.610
		Unweighted Count	81	129	210
	% within edu	Estimate	51.2%	48.8%	100.0%
		Standard Error	5.1%	5.1%	0.0%
		Unweighted Count	81	129	210
	% within gender	Estimate	8.7%	8.0%	8.3%
		Standard Error	1.2%	1.1%	0.8%
		Unweighted Count	81	129	210
	% of Total	Estimate	4.3%	4.1%	8.3%
		Standard Error	0.6%	0.6%	0.8%
		Unweighted Count	81	129	210
12 years or completed high school	Population Size	Estimate	26259901.960	28913863.037	55173764.998
		Standard Error	2516881.650	2091357.442	3297762.436
		Unweighted Count	208	395	603
	% within edu	Estimate	47.6%	52.4%	100.0%
		Standard Error	3.0%	3.0%	0.0%
		Unweighted Count	208	395	603
	% within gender	Estimate	22.3%	23.7%	23.0%
		Standard Error	1.9%	1.6%	1.3%
		Unweighted Count	208	395	603
	% of Total	Estimate	10.9%	12.1%	23.0%
		Standard Error	1.0%	0.9%	1.3%
		Unweighted Count	208	395	603
Some college	Population Size	Estimate	41290009.462	37730087.988	79020097.450
		Standard Error	3834893.037	2306432.685	4442396.848
		Unweighted Count	391	542	933
	% within edu	Estimate	52.3%	47.7%	100.0%
		Standard Error	2.8%	2.8%	0.0%
		Unweighted Count	391	542	933
	% within gender	Estimate	35.1%	30.9%	32.9%
		Standard Error	2.3%	1.7%	1.4%
		Unweighted Count	391	542	933
	% of Total	Estimate	17.2%	15.7%	32.9%
		Standard Error	1.4%	0.9%	1.4%
		Unweighted Count	391	542	933

College graduate or higher	Population Size	Estimate	39933429.705	45740870.640	85674300.345
		Standard Error	2240922.453	2072513.387	3090631.068
		Unweighted Count	593	799	1392
	% within edu	Estimate	46.6%	53.4%	100.0%
		Standard Error	1.8%	1.8%	0.0%
		Unweighted Count	593	799	1392
	% within gender	Estimate	33.9%	37.5%	35.7%
		Standard Error	1.9%	1.4%	1.2%
		Unweighted Count	593	799	1392
	% of Total	Estimate	16.7%	19.1%	35.7%
		Standard Error	0.9%	0.8%	1.2%
		Unweighted Count	593	799	1392
Total	Population Size	Estimate	117706034.701	122133703.472	239839738.173
		Standard Error	5332594.927	3545134.506	6371356.251
		Unweighted Count	1273	1865	3138
	% within edu	Estimate	49.1%	50.9%	100.0%
		Standard Error	1.4%	1.4%	0.0%
		Unweighted Count	1273	1865	3138
	% within gender	Estimate	100.0%	100.0%	100.0%
		Standard Error	0.0%	0.0%	0.0%
		Unweighted Count	1273	1865	3138
	% of Total	Estimate	49.1%	50.9%	100.0%
		Standard Error	1.4%	1.4%	0.0%
		Unweighted Count	1273	1865	3138

Tests of Independence

	Chi-Square	Adjusted F	df1	df2	Sig.
Pearson	7.999	1.036	2.856	279.908	.375
Likelihood Ratio	8.002	1.037	2.856	279.908	.374

The weighted percentages above show that a greater proportion of women have at least a college degree compared to men, 19.1% vs 16.7%. The Chi-squared test of independence indicates that there is not a significant difference between the educational distribution in these two groups (p-value > .05).

The results of these tests conducted in SPSS based on Taylor Series linearization contradict the results conducted in SAS using replication shown in Appendix A. (In SAS, the distributions of educational attainment between men and women were determined to be statistically different.) This is a good example of how the variance estimation method used can affect the outcome of a statistical test. Both education and gender are variables used in the raking process as part of the HINTS weighting procedure. As a result, the standard errors based on replication are much smaller than those based on Taylor Series linearization, which in turn results in significant differences in SAS but not in SPSS.

Note that the CSTABULATE procedure provides results for the Pearson Chi-square and Likelihood Ratio tests, but not for the Wald Chi-square test of independence. To get the results for the Wald Chi-square test of independence, users can conduct a logistic regression model in the CSLOGISTIC procedure in which the type of Chi-square test can be specified.

Logistic Regression

This example demonstrates a multivariable logistic regression model using **CSLOGISTIC**; recall that the response should be a categorical variable.

```
*Multivariable logistic regression of gender and education on SeekCancerInfo.
CSLOGISTIC  seekcancerinfo_recode (LOW) BY flippedgender flippededu
 /PLAN FILE='(sample.csaplan)'
 /MODEL flippedgender flippededu
 /CUSTOM  Label = 'Overall model minus intercept'
  LMATRIX = flippedgender 1/2 -1/2;
         flippededu 1/3 1/3 1/3 -1;
        flippededu 1/3 1/3 -1 1/3 ;
        flippededu 1/3 -1 1/3 1/3;
        flippededu -1 1/3 1/3 1/3
  /CUSTOM  Label = 'Gender'
 LMATRIX =  flippedgender 1/2 -1/2
  /CUSTOM  Label = 'Education overall'
  LMATRIX = flippededu 1/3 1/3 1/3 -1;
        flippededu 1/3 1/3 -1 1/3 ;
        flippededu 1/3 -1 1/3 1/3;
        flippededu -1 1/3 1/3 1/3
 /INTERCEPT INCLUDE=YES SHOW=YES
 /STATISTICS PARAMETER SE CINTERVAL TTEST EXP
 /TEST TYPE=CHISQUARE PADJUST=LSD
 /ODDSRATIOS FACTOR=[flippedgender(HIGH)]
 /ODDSRATIOS FACTOR=[flippededu(HIGH)]
 /MISSING CLASSMISSING=EXCLUDE
 /CRITERIA MXITER=100 MXSTEP=50 PCONVERGE=[1e-008 RELATIVE] LCONVERGE=[0] CHKSEP=20
CILEVEL=95
 /PRINT SUMMARY COVB CORB VARIABLEINFO SAMPLEINFO.
```

The response variable should be on the left-hand side of the BY statement, while all covariates should be listed on the right-hand side. The (LOW) option indicates that the lowest category is the reference category, thus requests the probability of seekcancerinfo="Yes" to be modeled. The "Male" is the reference group for gender effect, while "Less than high school" is the reference group for education level effect. The subcommand MODEL specifies all variables in the model. The CUSTOM subcommand allows users to define custom hypothesis tests. The LMATRIX statement specifies coefficients of contrasts, which are used for studying the effects in the model. The INTERCEPT subcommand specifies whether to include or show the intercept in the final estimates. The STATISTICS subcommand specifies the statistics to be estimated and shown in the final result, where the syntax PARAMETER indicates the coefficient estimates, EXP indicates the exponentiated coefficient estimates, SE indicates the standard error for each coefficient estimate, CINTERVAL indicates the confidence interval for each coefficient estimate. The TEST subcommand specifies the type of test statistic and the method of adjusting the significance level to be used for hypothesis tests that are requested on the MODEL and CUSTOM subcommands, where the syntax CHISQUARE indicates the Wald chi-square test, and LSD indicates the least significant difference. The ODDSRATIOS subcommand estimates odds ratios for certain factors. The subcommand MISSING specifies how to handle missing data. The subcommand CRITERIA offers controls on the iterative algorithm that is used for estimations. The option PCONVERGE= [1e-008 RELATIVE] helps to avoid early termination of the iteration. The subcommand PRINT is used to display optional output.

Sample Design Information

		N
Unweighted Cases	Valid	3123
	Invalid	162
	Total	3285
Population Size		239114487.375
Stage 1	Strata	2
	Units	100
Sampling Design Degrees of Freedom		98

Parameter Estimates

seekcancerinfo_recode	Parameter	B	Std. Error	95% Confidence Interval Lower	95% Confidence Interval Upper	Hypothesis Test t	Hypothesis Test df	Hypothesis Test Sig.	Exp(B)	95% Confidence Interval for Exp(B) Lower	95% Confidence Interval for Exp(B) Upper
Yes	(Intercept)	-.685	.239	-1.160	-.210	-2.861	98	.005	.504	.313	.811
	Female	.303	.123	.059	.547	2.465	98	.015	1.354	1.061	1.728
	College graduate or higher	1.307	.225	.861	1.753	5.815	98	.000	3.694	2.365	5.771
	Some college	.753	.269	.219	1.286	2.799	98	.006	2.122	1.245	3.618
	12 years or completed high school	.257	.230	-.200	.713	1.115	98	.268	1.292	.819	2.040

Dependent Variable: seekcancerinfo_recode (reference category = No)

Model: (Intercept), flippedgender, flippededu

a. Set to zero because this parameter is redundant.

Odds Ratios

		Odds Ratio	95% Confidence Interval Lower	Upper
flippedgender	Female vs. Male	1.354	1.061	1.728
flippededu	College graduate or higher vs. Less than high school	3.694	2.365	5.771
	Some college vs. Less than high school	2.122	1.245	3.618
	12 years or completed high school vs. Less than high school	1.292	.819	2.040

Overall Model Minus Intercept

df	Wald Chi-Square	Sig.
4.000	75.759	.000

Gender

df	Wald Chi-Square	Sig.
1.000	6.075	.014

Education Overall

df	Wald Chi-Square	Sig.
3.000	75.163	.000

To identify levels/variables that display a significant difference in response, the rule of thumb is to examine odds ratios where the confidence interval does not contain 1 (by default, SPSS will use alpha=.05 to determine statistical significance; this value can be changed by the user using code). However, significance may also be garnered from the test of whether the associated beta parameter is equal to 0 (see “Parameter Estimates” table above). According to this model, those with at least a college degree appear to be statistically more inclined to search for cancer information.

Note that in SPSS we cannot get the overall model effect, even if we used the CUSTOM subcommand to conduct custom hypothesis tests.

Linear Regression

This example demonstrates a multivariable linear regression model using **CSGLM**; recall that the response should be a continuous variable. For the purposes of this example, we decided to use an outcome with five levels as a continuous variable (GENERALHEALTH). Note that higher values on GENERALHEALTH indicate poorer self-reported health status.

```
* Multivariable linear regression of gender and education on GeneralHealth.
CSGLM genhealth_recode BY flippedgender flippededu
 /PLAN FILE='(sample.csaplan)'
 /MODEL flippededu flippedgender
 /CUSTOM  Label = 'Overall model minus intercept'
  LMATRIX = flippedgender 1/2 -1/2;
          flippededu 1/3 1/3 1/3 -1;
          flippededu 1/3 1/3 -1 1/3 ;
          flippededu 1/3 -1 1/3 1/3;
          flippededu -1 1/3 1/3 1/3
  /CUSTOM  Label = 'Gender'
 LMATRIX =  flippedgender 1/2 -1/2
  /CUSTOM  Label = 'Education overall'
  LMATRIX =  flippededu 1/3 1/3 1/3 -1;
            flippededu 1/3 1/3 -1 1/3 ;
            flippededu 1/3 -1 1/3 1/3;
            flippededu -1 1/3 1/3 1/3
 /INTERCEPT INCLUDE=YES SHOW=YES
 /STATISTICS PARAMETER SE CINTERVAL TTEST
 /PRINT SUMMARY VARIABLEINFO SAMPLEINFO
 /TEST TYPE=F PADJUST=LSD
 /MISSING CLASSMISSING=EXCLUDE
 /CRITERIA CILEVEL=95.
```

Sample Design Information

		N
Unweighted Cases	Valid	3116
	Invalid	169
	Total	3285
Population Size		238728802.297
Stage 1	Strata	2
	Units	100
Sampling Design Degrees of Freedom		98

Parameter Estimates

Parameter	Estimate	Std. Error	95% Confidence Interval Lower	Upper	Hypothesis Test t	df	Sig.
(Intercept)	3.089	.105	2.881	3.298	29.414	98.000	.000
College graduate or higher	-.762	.095	-.951	-.573	-7.988	98.000	.000
Some college	-.484	.110	-.703	-.265	-4.386	98.000	.000
12 years or completed high school	-.281	.102	-.483	-.079	-2.759	98.000	.007
Female	.001	.055	-.108	.109	.014	98.000	.989

Compared to those respondents with less than a high school education, those who completed 12 years of school or completed high school on average reported significantly better general health (i.e., the negative beta coefficient indicates that the average health score is lower among those with some college, and the health variable is coded such that lower scores correspond to better health), controlling for all variables in the model. This association also applies to those who have completed some college and those with a college degree or higher. We do not interpret the estimates for females because the corresponding p-value is greater than .05.

Overall Model Minus Intercept

df1	df2	Wald F	Sig.
4.000	95.000	26.096	.000

Gender

df1	df2	Wald F	Sig.
1.000	98.000	.000	.989

Education

df1	df2	Wald F	Sig.
3.000	96.000	31.707	.000

From the above table, we can see that education, but not gender, is significantly associated with general health.

Appendix C: Analyzing data using SUDAAN

This document gives some SUDAAN (Version 10.0.1 and higher) coding examples for common types of statistical analyses using HINTS 5 Cycle 1 data. These examples will incorporate both the final sample weight (to get population estimates) and the set of 50 jackknife replicate weights to get the proper standard error. Although these examples specifically use HINTS 5, Cycle 1 data, the concepts used here are generally applicable to other types of analyses. We will consider an analysis that includes gender, education level (edu) and two questions that are specific to the HINTS 5, Cycle 1 data: seekcancerinfo & generalhealth.

We begin by doing data management of the HINTS 5, Cycle 1 data in a SAS DATA step. We first decided to exclude all “Missing data (Not Ascertained)” and “Multiple responses selected in error” responses from the analyses. By setting these values to missing (.), SAS will exclude these responses from procedures where these variables are specifically accessed. For logistic regression modeling within the PROC RLOGIST procedure, SUDAAN expects the response variable to be dichotomous with values (0, 1), so this variable will also be recoded at this point. When recoding existing variables, it is generally recommended to create new variables of rather than over-writing the existing variables. Note: New variables should always be compared to original source variables in a SAS PROC FREQ procedure to verify proper coding.

```
proc format;  *First create some temporary formats;

Value Genderf
1 = "Male"
2 = "Female";

Value Educationf
1 = "Less than high school"
2 = "12 years or completed high school"
3 = "Some college"
4 = "College graduate or higher";

value seekcancerinfof
1 = "Yes"
0 = "No";

Value Generalf
1 = "Excellent"
2 = "Very good"
3 = "Good"
4 = "Fair"
5 = "Poor";

run;


data hints5cycle1;
set hints5c1.hints5_cycle1_weighted_public;

/*Recode negative values to missing*/
if genderc = 1 then gender = 1;
```

```
if genderc = 2 then gender = 2;
if genderc = -9 then gender = .;

/*Recode education into four
levels, and negative values to
missing*/ if education in (1, 2)
then edu = 1;
if education = 3 then edu = 2;
if education in (4, 5) then edu = 3;
if education in (6, 7) then edu = 4;
if education = -9 then edu = .;

/*Recode seekcancerinfo to 0-1 format, and negative values to missing for
proc rlogist procedure*/
if seekcancerinfo = 2 then seekcancerinfo = 0;
if seekcancerinfo = -9 then seekcancerinfo = .;

/*Recode negative values to missing for proc regress procedure*/
if generalhealth in (-5, -9) then generalhealth = .;

/*Apply formats to recoded variables */
format gender genderf. edu educationf. seekcancerinfo seekcancerinfof.
generalhealth generalf.;

run;
```

Crosstabs Procedure

This syntax is consistent for all procedures. Other data sets that incorporate the final sample weight and the 50 jackknife replicate weights will utilize the same three lines of code.

```
proc crosstab data=hints5cycle1 design=jackknife ddf = 49;
weight person_finwt0;
jackwgts person_finwt1-person_finwt50 / adjjack=.98;
class gender edu;
tables edu*gender;
test chisq;
run;
```

Since this procedure is mainly for categorical variables, each variable should be specified as such by inclusion in the class statement (which is ubiquitous in all SUDAAN procedures). The tables statement defines the frequencies that should be generated. Stand-alone variables listed here result in one-way frequencies, while a "*" between variables will define cross-frequencies. In general, the CROSSTAB procedure may be used to investigate n-way variable frequencies, along with their relationships. This is accomplished by the test statement, which defines various types of independence tests: here a Chi-Squared test is implemented. Other tests and statistics are also available; see the SUDAAN site link at the end of this document for more information.

The HINTS 5, Cycle 1 database for a single iteration contains a set of 50 replicate weights to compute accurate standard errors for statistical testing procedures. These replicate weights were created using a jackknife minus one replication method. Thus, the proper denominator degrees of freedom (ddf) should be 49 when one iteration of HINTS data is being analyzed. Thus, analysts who are only using the HINTS 5 Cycle 1 data should use 49 ddf in their statistical models.

HINTS 5 databases with more than one iteration of data will contain a set of 50*k replicate weights, where they can be viewed as being created using a stratified jackknife method with k as the number of strata and 49*k as the appropriate ddf. Analysts who were merging two iterations of data and making comparisons these should adjust the ddf to be 98 (49*2) etc.

Variance Estimation Method: Replicate Weight Jackknife
By: EDU, GENDER

		Are you male or female?		
What is the highest grade or level of schooling you completed?		**Total**	**Male**	**Female**
Total	**Sample Size**	3138	1273	1865
	Col Percent	100.00	100.00	100.00
	Row Percent	100.00	49.08	50.92
Less than HS	**Sample Size**	210	81	129
	Col Percent	8.33	8.68	7.98
	Row Percent	100.00	51.19	48.81
12 years or completed HS	**Sample Size**	603	208	395
	Col Percent	23.00	22.31	23.67
	Row Percent	100.00	47.59	52.41
Some college	**Sample Size**	933	391	542
	Col Percent	32.95	35.08	30.89
	Row Percent	100.00	52.25	47.75
College graduate or higher	**Sample Size**	1392	593	799
	Col Percent	35.72	33.93	37.45
	Row Percent	100.00	46.61	53.39

Variance Estimation Method: Replicate Weight Jackknife
Chi Square Test of Independence for EDU and GENDER

ChiSq Test Value	50.1063
P-value for ChiSq	0.0000
Degress of Freedom ChiSq	3

Logistic Regression

This example demonstrates a multivariable logistic regression model using **PROC RLOGIST** (*RLOGIST* is used to differentiate it from the SAS procedure, PROC LOGISTIC, and is used with SAS-callable SUDAAN); recall that the response should be a dichotomous 0-1 variable.

```
/*Multivariable logistic regression of gender and education on
SeekCancerInfo*/
```

```
proc rlogist data = hints5cycle1 design = jackknife ddf = 49;
weight person_finwt0;
jackwgts person_finwt1-person_finwt50 / adjjack = 0.98;
class gender edu;
model seekcancerinfo = gender edu;
reflev gender=1 edu=1;
run;
```

The response variable should be on the left hand side (LHS) of the equal sign in the model statement, while all covariates should be listed on the right hand side (RHS). Categorical variables should also be included in the class statement. By default, the reference level of each categorical variable is that of the highest numeric level. This may be changed by using the *reflevel* statement to explicitly define another reference level.

Variance Estimation Method: Replicate Weight Jackknife
Working Correlations: Independent
Link Function: Logit
Response variable SEEKCANCERINFO: A8. Have you ever looked for information about cancer from any source?
by: Independent Variables and Effects.

Independent variables and effects	Beta Coeff.	SE Beta	T-test B=0	P-value T-Test B=0
Intercept	-0.68	0.24	-2.84	0.0065
Gender				
Male	0.00	0.00	.	.
Female	0.30	0.13	2.35	0.0229
Education Level				
Less than HS	0.00	0.00	.	.
12 years or HS completed	0.26	0.22	1.16	0.2532
Some College	0.75	0.27	2.77	0.0079
College graduate or higher	1.31	0.21	6.34	0.0000

Contrast	Wald F	P-value Wald Chi-Sq
Overall Model	17.12	0.0000
Model minus intercept	18.18	0.0000
Intercept	.	.
Gender	5.52	0.0229
Edu	23.10	0.0000

Variance Estimation Method: Replicate Weight Jackknife
Working Correlations: Independent
Link Function: Logit
Response variable SEEKCANCERINFO: A8. Have you ever looked for information about cancer from any source?
by: Independent Variables and Effects.

Independent variables and effects	Odds Ratio	Lower 95% Limit OR	Upper 95% Limit OR
Intercept	0.50	0.31	0.82
Gender			
Male	1.00	1.00	1.00
Female	1.35	1.04	1.75
Education Level			
Less than HS	1.00	1.00	1.00
12 years or HS completed	1.29	0.83	2.02
Some College	2.12	1.23	3.66
College graduate or higher	3.69	2.44	5.59

To identify levels/variables that display a significant difference in response, the rule of thumb is to examine odds ratios where the confidence interval does not contain 1 (by default, SUDAAN will use alpha=.05 to determine statistical significance; this value can be changed by the user using code). However, significance may also be garnered from the test of whether the associated beta parameter is equal to 0 (see first regression table above). According to this model, women and those with at least some college education appear to be statistically more inclined to search for cancer information (compared with men and those who did not graduate from high school, respectively).

Linear Regression

This example demonstrates a multivariable linear regression model using **PROC REGRESS** (*REGRESS* is used to differentiate it from the SAS procedure, PROC REG, and is used with SAS-callable SUDAAN); recall that the response should be a continuous variable. For the purposes of this example, we decided to use an outcome with five levels as a continuous variable (GENERALHEALTH). Note that higher values on GENERALHEALTH indicate poorer self-reported health status.

```
/*Multivariable linear regression of gender and education on GeneralHealth*/
proc regress data = hints5cycle1 design = jackknife ddf = 49;
weight person_finwt0;
jackwgts person_finwt1-person_finwt50 / adjjack = 0.98;
class gender edu;
model generalhealth = gender edu ;
reflev gender=1 edu=1;
run;
```

Variance Estimation Method: Replicate Weight Jackknife
Working Correlations: Independent
Link Function: Identity
Response variable GENERALHEALTH: D1. In general, would you say your health is…
by: Contrast.

Contrast	Wald F	P-value Wald F
Overall Model	2002.87	0.0000
Model minus intercept	20.08	0.0000
Intercept	.	.
Gender	0.00	0.9904
Edu	26.01	0.0000

From the above table, we can see that Gender is not associated with the outcome, but Edu is associated, adjusting for all variables in the model.

Variance Estimation Method: Replicate Weight Jackknife
Working Correlations: Independent
Link Function: Identity
Response variable GENERALHEALTH: D1. In general, would you say your health is…
by: Independent Variables and Effects.

Independent variables and effects	Beta Coeff.	SE Beta	T-test B=0	P-value T-Test B=0
Intercept	3.09	0.10	30.79	0.0000
Gender				
Male	0.00	0.00	.	.
Female	0.00	0.06	0.01	0.9904
Education Level				
Less than HS	0.00	0.00	.	.
12 years or HS completed	-0.28	0.09	-2.98	0.0045
Some College	-0.48	0.11	-4.40	0.0001
College graduate or higher	-0.76	0.10	-8.01	0.0000

From the above table, it can be seen that, compared to those respondents with Less than High School education, those with Some College have a significantly negative linear association with the outcome (i.e., better reported health), controlling for all variables in the model. This association also applies to those with a College Degree or Higher. We don't interpret the Gender variable because it is non-significant.

Appendix D: Analyzing data using STATA

This section gives some Stata (Version 10.0 and higher) coding examples for common types of statistical analyses using HINTS 5, Cycle 1 data. We begin by doing data management of the HINTS 5 data. We first decided to exclude all “Missing data (Not Ascertained)”, “Multiple responses selected in error”, “Question answered in error (Commission Error)” and “Inapplicable, coded 2 in SeekHealthInfo” responses from the analyses. By setting these values to missing (.), Stata will exclude these responses from analysis commands where these variables are specifically accessed. For logistic regression modeling within the **svy: logit** command, Stata expects the response variable to be dichotomous with values (0, 1), so this variable will also be recoded at this point. When recoding existing variables, it is generally recommended to create new variables of rather than over-writing the existing variables. Note: New variables should always be compared to original source variables in a Stata **tabulate** command to verify proper coding.

STATA Data Management Code

```
use “file path\hints5_cycle1_public.dta”
*Recode negative values to missing

recode genderc (1=1 "Male") (2=2 "Female") (nonmissing=.),

generate(gender) label variable gender "Gender"

*Recode education into four levels, and negative values to missing

recode education (1/2=1 "Less than high school") (3=2 "12 years or
completed high school") (4/5=3 "Some college") (6/7=4 "College graduate or
higher") (nonmissing=.), generate(edu)

label variable edu "Education"



*Recode seekcancerinfo to 0-1 format, and negative values to missing for
svy: logit


replace seekcancerinfo = 0 if seekcancerinfo == 2

replace seekcancerinfo = . if seekcancerinfo == -1 | seekcancerinfo == -2
| seekcancerinfo == -6

label define seekcancerinfo 0 "No" 1 "Yes"

label val seekcancerinfo seekcancerinfo



*Recode negative values to missing for svy: regress
```

```
replace generalhealth = . if generalhealth == -5 | generalhealth== -9
```

Replicate Weights Variance Estimation Method

Declare survey design

Stata requires declaring the survey design for the data set globally before any analysis. The declared survey design will be applied to all future survey commands unless another survey design is declared. Other data sets that incorporate the final sample weight and the 50 jackknife replicate weights will utilize the same code.

```
* Declare survey design for the data set

svyset [pw=person_finwt0], jkrw(person_finwt1-person_finwt50,
multiplier(0.98)) vce(jack) mse
```

Cross-tabulation

```
* cross-tabulation

svy: tabulate edu gender, column row format(%8.5f) percent wald noadjust
```

The **svy: tabulate** command defines the frequencies that should be generated. Single variables listed in **svy: tabulate** results in one-way frequencies, while two variables will define cross-frequencies. The options **column** and **row** request column and row frequencies, respectively. The option **percent** requests the frequencies are displayed in percentage. The options **wald** and **noadjust** together request unadjusted Wald test for independence. Stata recommends default pearson test for independence. Other tests and statistics are also available; see the Stata website for more information: http://www.stata.com/

For the purposes of computing appropriate degrees of freedom for the estimator of the HINTS 5, Cycle 1 differences, we can assume as an approximation that the sample is a simple random sample of size 50 (corresponding to the 50 replicates: each replicate provides a 'pseudo sample unit') from a normal distribution. The denominator degrees of freedom (df) is equal to 49*k, where k is the number of iterations of data used in this analysis. Stata uses the number of replicates minus one as the denominator degrees of freedom and does not provide the option for user to specify the denominator degrees of freedom.

knife *: for cell counts

```
Number of strata =     1              Number of obs  =    3,138
Population size      =
239,839,738
                                      Replications  =      50
                                      Design df     =      49
```

Education	Gender		Total
	Male	Female	
Less than HS	51.18622	48.81378	1.0e+02
	8.68494	7.98214	8.32705
12 years or HS completed	47.59491	52.40509	1.0e+02
	22.30973	23.67394	23.00443
Some college	52.25254	47.74746	1.0e+02
	35.07892	30.89245	32.94704
College grad or higher	46.61075	53.38925	1.0e+02
	33.92641	37.45147	35.72148
Total	49.07695	50.92305	1.0e+02
	1.0e+02	1.0e+02	1.0e+02

Key: row percentages
column percentages

Wald (Pearson):
Unadjusted chi2(3) = 150.3198
Unadjusted F(3, 49) = 50.1066 P = 0.0000
Adjusted F(3, 47) = 48.0614 P = 0.0000

Logistic Regression

This example demonstrates a multivariable logistic regression model using **svy: logit** (to get parameters) and **svy, or: logit** (to get odds ratios); recall that the response should be a dichotomous 0-1 variable.

```
* Define reference group for categorical variables for both svy: logit and svy: regress

char gender [omit] 1

char edu [omit] 1


* Multivariable logistic regression of gender and education on seekcancerinfo

xi: svy: logit seekcancerinfo i.gender i.edu

test _Igender_2 _Iedu_2 _Iedu_3 _Iedu_4 _cons, nosvyadjust

test _Igender_2 _Iedu_2 _Iedu_3 _Iedu_4, nosvyadjust

test _Igender_2, nosvyadjust

test _Iedu_2 _Iedu_3 _Iedu_4, nosvyadjust

xi: svy, or: logit seekcancerinfo i.gender i.edu
```

The **char** command defines categorical variable with reference group. The “Male” is the reference group for gender effect while the “Less than high school” is the reference group for education level effect. These definitions will be applied to future commands until another **char** command re-defines the reference group. The xi command will create proper dummy variables for i.gender and i.edu variables in the analysis commands. The response variable should be the first variable in **svy: logit** command and be followed by all covariates. The **test** command tests the hypotheses about estimated parameters.

```
i.gender    _Igender_1-2    (naturally coded; _Igender_1 omitted)

i.edu       _Iedu_1-4       (naturally coded; _Iedu_1 omitted)

(running logit on estimation sample)
```

Jackknife replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
.... 50

Survey: Logistic regression

Number of strata = 1

Number of obs = 3,138
Population size = 239,839,738
Replications = 50
Design df = 49
F(4, 46) = 17.01
Prob > F = 0.0000

seekcancer~o	Coef.	Jknife * Std. Err.	t	P>|t|	[95% Conf. Interval]	
_Igender_2	.3096428	.1294447	2.39	0.021	.049514	.5697717
_Iedu_2	.2594209	.2216962	1.17	0.248	-.1860943	.704936
_Iedu_3	.7613234	.2714073	2.81	0.007	.2159099	1.306737
_Iedu_4	1.310878	.2063649	6.35	0.000	.8961718	1.725583
_cons	-.6882554	.2412734	-2.85	0.006	-1.173112	-.2033983

Unadjusted Wald test

(1) [seekcancerinfo]_Igender_2 = 0
(2) [seekcancerinfo]_Iedu_2 = 0
(3) [seekcancerinfo]_Iedu_3 = 0
(4) [seekcancerinfo]_Iedu_4 = 0
(5) [seekcancerinfo]_cons = 0

F(5, 49) = 17.17

Prob > F = 0.0000

Unadjusted Wald test

(1) [seekcancerinfo]_Igender_2 = 0
(2) [seekcancerinfo]_Iedu_2 = 0
(3) [seekcancerinfo]_Iedu_3 = 0 (4) [seekcancerinfo]_Iedu_4 = 0

F(4, 49) = 18.12

Prob > F = 0.0000

Unadjusted Wald test

(1) [seekcancerinfo]_Igender_2 = 0

F(1, 49) = 5.72
Prob > F = 0.0206

Unadjusted Wald test

(1) [seekcancerinfo]_Iedu_2 = 0
(2) [seekcancerinfo]_Iedu_3 = 0
(3) [seekcancerinfo]_Iedu_4 = 0
F(3, 49) = 23.05

Prob > F = 0.0000

i.gender _Igender_1-2 (naturally coded; _Igender_1 omitted)
i.edu _Iedu_1-4 (naturally coded; _Iedu_1 omitted)

(running logit on estimation sample)

Jackknife replications (50)

----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5

.... 50

Survey: Logistic regression

Number of strata = 1
Number of obs = 3,138
Population size = 239,839,738
Replications = 50
Design df = 49
F(4, 46) = 17.01
Prob > F = 0.0000

seekcancer~o	Odds Ratio	Jknife *	t	P>|t|	[95% Conf. Interval]

		Std. Err.				
_Igender_2	1.362938	.1764251	2.39	0.021	1.05076	1.767863
_Iedu_2	1.296179	.287358	1.17	0.248	.8301953	2.023717
_Iedu_3	2.141108	.5811124	2.81	0.007	1.240991	3.6941
_Iedu_4	3.709428	.7654957	6.35	0.000	2.450205	5.615796

To identify levels/variables that display a significant difference in response, the rule of thumb is to examine odds ratios where the confidence interval does not contain 1 (by default, Stata will use alpha=.05 to determine statistical significance; this value can be changed by the user using code). However, significance may also be garnered from the test of whether the associated beta parameter is equal to 0 (see first regression table above). According to this model, those with a college education or more appear to be statistically more inclined to search for cancer information compared with those who did not graduate from high school.

Linear Regression

This example demonstrates a multivariable linear regression model using **svy: regress**; recall that the response should be a continuous variable. For the purposes of this example, we decided to use an outcome with five levels as a continuous variable (generalhealth). Note that higher values on generalhealth indicate poorer self-reported health status.

```
* Multivariable linear regression of gender and education on

generalhealth xi: svy: regress generalhealth i.gender i.edu

test _Igender_2 _Iedu_2 _Iedu_3 _Iedu_4 _cons,

nosvyadjust test _Igender_2 _Iedu_2 _Iedu_3 _Iedu_4,

nosvyadjust

test _Igender_2, nosvyadjust

test _Iedu_2 _Iedu_3 _Iedu_4, nosvyadjust
```

```
i.gender      _Igender_1-2      (naturally coded; _Igender_1 omitted)

i.edu         _Iedu_1-4         (naturally coded; _Iedu_1 omitted)

(running regress on estimation sample)


Jackknife replications (50)
```

----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5

.... 50

Survey: Linear regression

Number of strata = 1

Number of obs = 3,116
Population size = 238,728,802
Replications = 50
Design df = 49
F(4, 46) = 18.85
Prob > F = 0.0000
R-squared = 0.0620

generalhea~h	Coef.	Jknife * Std. Err.	t	P>|t|	[95% Conf. Interval]	
_Igender_2	.0007469	.0617054	0.01	0.990	-.1232547	.1247485
_Iedu_2	-.2810533	.0943847	-2.98	0.005	-.4707265	-.0913801
_Iedu_3	-.4841933	.1099254	-4.40	0.000	-.7050967	-.2632899
_Iedu_4	-.7618133	.0950815	-8.01	0.000	-.9528868	-.5707399
_cons	3.089375	.1003477	30.79	0.000	2.887719	3.291032

Unadjusted Wald test

(1) _Igender_2 = 0
(2) _Iedu_2 = 0
(3) _Iedu_3 = 0
(4) _Iedu_4 = 0
(5) _cons = 0

F(5, 49) = 2002.87
Prob > F = 0.0000

Unadjusted Wald test

(1) _Igender_2 = 0
(2) _Iedu_2 = 0
(3) _Iedu_3 = 0
(4) _Iedu_4 = 0

F(4, 49) = 20.08
Prob > F = 0.0000

Unadjusted Wald test

(1) _Igender_2 = 0

F(1, 49) = 0.00
Prob > F = 0.9904

Unadjusted Wald test

(1) _Iedu_2 = 0
(2) _Iedu_3 = 0
(3) _Iedu_4 = 0

F(3, 49) = 26.01
Prob > F = 0.0000

From the above table, it can be seen that, compared to those respondents with less than a high school education, those with a high school education, some college, or a college degree or higher have a significantly negative linear association with the outcome (i.e., better reported health), controlling for all variables in the model. We don't interpret the gender variable because it is non- significant.

Taylor Series Linearization Variance Estimation Method

Declare Survey Design

Stata requires that the survey design be declared for the dataset globally before any analysis. The declared survey design will be applied to all future survey commands unless another survey design is declared. Other datasets that incorporate the final sample weight and stratum and cluster variables will utilize the same code.

```
* Declare survey design for the data set
  svyset var_cluster [pw=person_finwt0], strata(var_stratum)
```

Cross-tabulation

```
* cross-tabulation
svy: tabulate edu gender, column row format(%8.5f) percent wald noadjust
```

The **svy: tabulate** command defines the frequencies that should be generated. Single variables listed in **svy: tabulate** results in one-way frequencies, while two variables will define cross-frequencies. The options column and row request column and row frequencies, respectively. The option percent requests the frequencies and are displayed in percentages. The options wald and noadjust together request the unadjusted Wald test for independence. Stata recommends the default Pearson test for independence.

Other tests and statistics are also available; see the Stata website for more information: http://www.stata.com.

Number of strata = 2 Number of obs = 3,138
Number of PSUs = 100 Population size = 239,839,738
Design df = 98

Education	Gender		Total
	Male	Female	
Less than HS	51.18622	48.81378	1.0e+02
	8.68494	7.98214	8.32705
12 years or HS completed	47.59491	52.40509	1.0e+02
	22.30973	23.67394	23.00443
Some college	52.25254	47.74746	1.0e+02
	35.07892	30.89245	32.94704
College grad or higher	46.61075	53.38925	1.0e+02
	33.92641	37.45147	35.72148
Total	49.07695	50.92305	1.0e+02
	1.0e+02	1.0e+02	1.0e+02

Key: row percentages
column percentages

Wald (Pearson):
Unadjusted chi2(3) = 3.3479
Unadjusted F(3, 98) = 1.1160 P = 0.3464
Adjusted F(3, 96) = 1.0932 P = 0.3559

Logistic Regression

This example demonstrates a multivariable logistic regression model using **svy: logit** (to get parameters) and **svy, or: logit** (to get odds ratios); recall that the response should be a dichotomous 0-1 variable.

```
* Define reference group for categorical variables for both svy: logit and
svy: regress
char gender [omit] 1

char edu [omit] 1
```

```
* Multivariable logistic regression of gender and education on seekcancerinfo

xi: svy: logit seekcancerinfo i.gender i.edu

test _Igender_2 _Iedu_2 _Iedu_3 _Iedu_4 _cons, nosvyadjust
test _Igender_2 _Iedu_2 _Iedu_3 _Iedu_4, nosvyadjust
test _Igender_2, nosvyadjust
test _Iedu_2 _Iedu_3 _Iedu_4, nosvyadjust
xi: svy, or: logit seekcancerinfo i.gender i.edu
```

The **char** command defines categorical variable with reference group. The “Male” is the reference group for gender effect, while the “Less than high school” is the reference group for education level effect. These definitions will be applied to future commands until another char command redefines the reference group. The xi command will create proper dummy variables for i.gender and i.edu variables in the analysis commands. The response variable should be the first variable in **svy: logit** command and be followed by all covariates. The **test** command tests the hypotheses about estimated parameters.

i.gender _Igender_1-2 (naturally coded; _Igender_1 omitted)

i.edu _Iedu_1-4 (naturally coded; _Iedu_1 omitted)

(running logit on estimation sample)

Survey: Logistic regression

Number of strata = 2 Number of obs = 3,138
Number of PSUs = 100 Population size = 239,839,738
Design df = 98
F(4, 95) = 18.37
Prob > F = 0.0000

seekcancer info	Coef.	Linearized Std. Err.	t	P>t	[95% Conf. Interval]	
_Igender_2	0.309643	0.1231996	2.51	0.014	0.065157	0.554129
_Iedu_2	0.259421	0.2299473	1.13	0.262	-0.1969	0.715744
_Iedu_3	0.761323	0.2688377	2.83	0.006	0.227824	1.294823
_Iedu_4	1.310878	0.2251864	5.82	0	0.864002	1.757753
_cons	-0.68826	0.239752	-2.87	0.005	-1.16404	-0.21248

Unadjusted Wald test

(1) [seekcancerinfo]_Igender_2 = 0
(2) [seekcancerinfo]_Iedu_2 = 0
(3) [seekcancerinfo]_Iedu_3 = 0
(4) [seekcancerinfo]_Iedu_4 = 0
(5) [seekcancerinfo]_cons = 0

F(5, 98) = 22.47
Prob > F = 0.0000

Unadjusted Wald test

(1) [seekcancerinfo]_Igender_2 = 0
(2) [seekcancerinfo]_Iedu_2 = 0
(3) [seekcancerinfo]_Iedu_3 = 0
(4) [seekcancerinfo]_Iedu_4 = 0

F(4, 98) = 18.95
Prob > F = 0.0000

Unadjusted Wald test

(1) [seekcancerinfo]_Igender_2 = 0

F(1, 98) = 6.32
Prob > F = 0.0136

Unadjusted Wald test

(1) [seekcancerinfo]_Iedu_2 = 0
(2) [seekcancerinfo]_Iedu_3 = 0
(3) [seekcancerinfo]_Iedu_4 = 0

F(3, 98) = 25.05
Prob > F = 0.0000

i.gender _Igender_1-2 (naturally coded; _Igender_1 omitted)
i.edu _Iedu_1-4 (naturally coded; _Iedu_1 omitted)
(running logit on estimation sample)

Survey: Logistic regression

Number of strata = 2 Number of obs = 3,138
Number of PSUs = 100 Population size = 239,839,738
Design df = 98
F(4, 95) = 18.37
Prob > F = 0.0000

seekcancerinfo	Odds Ratio	Linearized Std. Err.	t	P>t	[95% Conf. Interval]	
_Igender_2	1.362938	0.1679135	2.51	0.014	1.067327	1.740423
_Iedu_2	1.296179	0.2980529	1.13	0.262	0.821271	2.045708
_Iedu_3	2.141108	0.5756105	2.83	0.006	1.255864	3.65035
_Iedu_4	3.709428	0.8353128	5.82	0	2.372638	5.79939
_cons	0.502452	0.1204639	-2.87	0.005	0.312224	0.808580

Note: _cons estimates baseline odds.

To identify levels/variables that display a significant difference in response, the rule of thumb is to examine odds ratios where the confidence interval does not contain 1 (by default, Stata will use alpha=.05 to determine statistical significance; this value can be changed by the user using code). However, significance may also be garnered from the test of whether the associated beta parameter is equal to 0 (see first regression table above). According to this model, those with a college education or more appear to be statistically more inclined to search for cancer information compared with those who did not graduate from high school.

Linear Regression

This example demonstrates a multivariable linear regression model using **svy: regress**; recall that the response should be a continuous variable. For the purposes of this example, we decided to use an outcome with five levels as a continuous variable (generalhealth). Note that higher values on generalhealth indicate poorer self-reported health status.

```
* Multivariable linear regression of gender and education on

generalhealth

xi: svy: regress generalhealth i.gender i.edu

test _Igender_2 _Iedu_2 _Iedu_3 _Iedu_4 _cons, nosvyadjust

test _Igender_2 _Iedu_2 _Iedu_3 _Iedu_4, nosvyadjust

test _Igender_2, nosvyadjust

test _Iedu_2 _Iedu_3 _Iedu_4, nosvyadjust
```

i.gender _Igender_1-2 (naturally coded; _Igender_1 omitted)
i.edu _Iedu_1-4 (naturally coded; _Iedu_1 omitted)
(running regress on estimation sample)

Survey: Linear regression

Number of strata = 2
Number of PSUs = 100

Number of obs = 3,116
Population size = 238,728,802
Design df = 98
F(4, 95) = 26.10
Prob > F = 0.0000

R-squared = 0.0620

generalhea~h	Coef.	Linearized Std. Err.	t	P>t	[95% Conf. Interval]	
_Igender_2	0.000747	0.054629	0.01	0.989	-0.10766	0.109157
_Iedu_2	-0.28105	0.101853	-2.76	0.007	-0.48318	-0.07893
_Iedu_3	-0.48419	0.110388	-4.39	0	-0.70325	-0.26513
_Iedu_4	-0.76181	0.095375	-7.99	0	-0.95108	-0.57255
_cons	3.089375	0.105029	29.41	0	2.880948	3.297803

Unadjusted Wald test

(1) _Igender_2 = 0
(2) _Iedu_2 = 0
(3) _Iedu_3 = 0
(4) _Iedu_4 = 0
(5) _cons = 0

F(5, 98) = 1941.26
Prob > F = 0.0000

Unadjusted Wald test

(1) _Igender_2 = 0
(2) _Iedu_2 = 0
(3) _Iedu_3 = 0
(4) _Iedu_4 = 0

F(4, 98) = 26.92
Prob > F = 0.0000

Unadjusted Wald test

(1) _Igender_2 = 0

F(1, 98) = 0.00
Prob > F = 0.9891

Unadjusted Wald test

(1) _Iedu_2 = 0
(2) _Iedu_3 = 0
(3) _Iedu_4 = 0

F(3, 98) = 32.37
Prob > F = 0.0000

From the above table, it can be seen that, compared to those respondents with less than a high school education, those with a high school education, some college, or a college degree or higher have a significantly negative linear association with the outcome (i.e., better reported health), controlling for all variables in the model. We don't interpret the gender variable because it is non-significant.

Appendix E: Merging HINTS 4, Cycle 4 and HINTS 5, Cycle 1 using SAS

This section provides SAS (Version 9.3 and higher) code for merging the HINTS 4, Cycle 4 and HINTS 5, Cycle 1 iterations. It first creates a temporary format for a new ‘survey’ variable that will distinguish between the two iterations. The code then creates two temporary data files and adds the new ‘survey’ variable to each dataset. Next, the two files are merged into one. It will match up variables that have the same name and format and create a merged data file (n = 6,962) that contains one final sample weight (for population point estimates) and 100 replicate weights (NWGT1 TO NWGT100; to compute standard errors).

```
/*FIRST CREATE THE FORMAT FOR THE SURVEY VARIABLE*/
proc format;

value survey
1="HINTS 4 CYCLE 4"
2="HINTS 5 CYCLE 1"
;
run;

/**************************************************************************/
/*CREATE TWO SEPARATE TEMPORARY DATA FILES THAT CONTAIN THE NEW ‘SURVEY’
VARIABLE.

options fmtsearch=(HINTS4C4); /*PUT NAME OF LIBRARY WHERE HINTS 4
                                CYCLE 4 FORMATS ARE STORED*/
data tempHINTS4CYCLE4;
set HINTS4C4.hints4cycle4_09062017_public; /*PUT NAME OF LIBRARY AND
                     NAME OF EXISTING HINTS 4 CYCLE 4 DATA FILE*/
survey=1;
format survey survey.;
run;

options fmtsearch=(HINTS5C1); /* PUT NAME OF LIBRARY WHERE HINTS 5 CYCLE 1
FORMATS ARE STORED*/

data tempHINTS5CYCLE1;
set HINTS5C1.hints5_cycle1_weighted_public; /*PUT NAME OF LIBRARY AND
                     NAME OF EXISTING HINTS 5 CYCLE 1 DATA FILE*/
survey=2;
format survey survey.;
run;

/**************************************************************************/
```

SAS Code to Set Up Final and Replicate Weights for the Replicate Variance Estimation Method

```
/*THIS CODE MERGES THE TWO TEMPORARY DATA SETS CREATED ABOVE. IT ALSO
CREATES ONE FINAL SAMPLE WEIGHT (NWGT0) AND 100 REPLICATE WEIGHTS
(NWGT1 THRU NWGT100)*/
```

```
data mergeHINTS4C4_HINTS5C1;
set tempHINTS4CYCLE4 tempHINTS5CYCLE1;

/*Create Replicate Weights for trend tests*/

      **Replicate Weights;
      array hints4wgts[50] person_finwt1-person_finwt50;
      array hints5wgts[50] person_finwt1-person_finwt50;
      array newWghts[100] nwgt1-nwgt100;


**Adjust Final And Replicate Weights;
      if survey eq 1 then do i=1 to 50;  *HINTS 4 CYCLE 4;
            nwgt0=person_finwt0;
            newWghts[i]=hints4wgts[i];
            newWghts[50+i]=person_finwt0;
            end;
      else if survey eq 2 then do i=1 to 50; *HINTS 5 CYCLE 1;
            nwgt0=person_finwt0;
            newWghts[50+i]=hints5wgts[i];
            newWghts[i]=person_finwt0;
            end;
run;


/*********************************************************/

/*YOU CAN USE THE CODE BELOW TO RUN SIMPLE FREQUENCIES ON TWO COMMON
VARIABLES, 'SEEKHEALTHINFO' AND 'CHANCEASKQUESTIONS'*/


/*SAS CODE*/
proc surveyfreq data = mergehints4c4_hints5c1 varmethod = jackknife;
weight nwgt0;
repweights nwgt1-nwgt100 / df = 49 jkcoefs = 0.98;
tables seekhealthinfo chanceaskquestions;
run;


/*SUDAAN CODE*/

proc crosstab data=mergefda1_fda2 design=jackknife ddf = 49;
weight nwgt0;
jackwgts nwgt1-nwgt100 / adjjack=.98;
class ecig_quit ecig_chemicals;
tables seekhealthinfo chanceaskquestions;
run;
```

SAS Code to Set Up Final and Replicate Weights for the Taylor Series Estimation Method

```
/*THIS CODE MERGES TWO TEMPORARY HINTS DATA SETS CREATED USING THE TAYLOR SERIES
LINEARZATION METHOD. PLEASE NOTE, THIS CODE IS BASED ON THE ASSUMPTION THAT THE
DATA SETS HAVE THE CORRECT VARIANCE CODES AND HHID VARIABLES MATCH*/
```

```
/*FIRST CREATE THE FORMAT FOR THE SURVEY VARIABLE*/
proc format;

value survey
1="HINTS 4 CYCLE 4"
2="HINTS 5 CYCLE 1"
;
run;

/*************************************************************************/

/*CREATE TWO SEPARATE TEMPORARY DATA FILES THAT CONTAIN THE NEW 'SURVEY'
VARIABLE.

options fmtsearch=(HINTS4C4); /*PUT NAME OF LIBRARY WHERE HINTS 4
                                CYCLE 4 FORMATS ARE STORED*/
data tempHINTS4CYCLE4;
set HINTS4C4.hints4cycle4_09062017_public; /*PUT NAME OF LIBRARY AND
                      NAME OF EXISTING HINTS 4 CYCLE 4 DATA FILE*/
survey=1;
format survey survey.;
run;

options fmtsearch=(HINTS5C1); /* PUT NAME OF LIBRARY WHERE HINTS 5 CYCLE 1
FORMATS ARE STORED*/

data tempHINTS5CYCLE1;
set HINTS5C1.hints5_cycle1_weighted_public; /*PUT NAME OF LIBRARY AND
                      NAME OF EXISTING HINTS 5 CYCLE 1 DATA FILE*/
survey=2;
format survey survey.;
run;

/*********************************************************************/

data mergeHINTS4C4_HINTS5C1;
      set tempHINTS4CYCLE4 tempHINTS5CYCLE1;
run;

/*******************************************************/
/*YOU CAN USE THE CODE BELOW TO RUN SIMPLE FREQUENCIES ON TWO COMMON
VARIABLES, 'SEEKHEALTHINFO' AND 'CHANCEASKQUESTIONS'*/

/*SAS CODE*/
proc surveyfreq data = MergeHints4C4_Hints5C1 varmethod = TAYLOR;
      strata VAR_STRATUM;
      cluster VAR_CLUSTER;
      weight MERGED_FINWT0;
      tables seekhealthinfo chanceaskquestions / row col;
run;
```

Appendix F: Merging HINTS 5, Cycle 1 and HINTS FDA, Cycle 2 using SAS

This section provides SAS (Version 9.3 and higher) code for merging the HINTS 5, Cycle 1 and HINTS FDA, Cycle 2 iterations. It first creates a temporary format for a new 'survey' variable that will distinguish between the two iterations. The code then creates two temporary data files and adds the new 'survey' variable to each dataset. Next, the two files are merged into one. It will match up variables that have the same name and format and create a merged data file (n = 5,021) that contains one final sample weight (for population point estimates) and 100 replicate weights (NWGT1 TO NWGT100; to compute standard errors).

```
/*FIRST CREATE THE FORMAT FOR THE SURVEY VARIABLE*/
proc format;

value survey
1="HINTS 5 CYCLE 1"
2="FDA-2"
;
run;
```

SAS Code to Set Up Final and Replicate Weights for the Replicate Variance Estimation Method

```
/************************************************************************/

/*CREATE TWO SEPARATE TEMPORARY DATA FILES THAT CONTAIN THE NEW 'SURVEY'
VARIABLE.*/

options fmtsearch=(HINTS5C1); /*PUT NAME OF LIBRARY WHERE HINTS 5
                                 CYCLE 1 FORMATS ARE STORED*/
data tempHINTS5CYCLE1;
set HINTS5C1.hints5_cycle1_public; /*PUT NAME OF LIBRARY AND
                           NAME OF EXISTING HINTS 5 CYCLE 1 FILE*/
survey=1;
format survey survey.;
run;

options fmtsearch=(FDA2); /* PUT NAME OF LIBRARY WHERE HINTS 5 CYCLE 1
FORMATS ARE STORED*/

data tempFDA2;
set FDA2.hints_fda2_public; /*PUT NAME OF LIBRARY AND NAME
                                      OF EXISTING FDA-2 DATA FILE*/
survey=2;
format survey survey.;
run;

/********************************************************************/

/*THIS CODE MERGES THE TWO TEMPORARY DATA SETS CREATED ABOVE. IT ALSO
CREATES ONE FINAL SAMPLE WEIGHT (NWGT0) AND 100 REPLICATE WEIGHTS
```

```
(NWGT1 THRU NWGT100)*/

data mergeHINTS5C1_FDA2;
set tempHINTS5CYCLE1 tempFDA2;

/*Create Replicate Weights for trend tests*/

      **Replicate Weights;
      array hints5wgts[50] person_finwt1-person_finwt50;
      array fda2wgts[50] person_finwt1-person_finwt50;
      array newWghts[100] nwgt1-nwgt100;


**Adjust Final And Replicate Weights;
      if survey eq 1 then do i=1 to 50;  *HINTS 5 CYCLE 1;
            nwgt0=person_finwt0;
            newWghts[i]=hints5wgts[i];
            newWghts[50+i]=person_finwt0;
            end;
      else if survey eq 2 then do i=1 to 50; *FDA-2;
            nwgt0=person_finwt0;
            newWghts[50+i]=fda2wgts[i];
            newWghts[i]=person_finwt0;
            end;
run;


/**********************************************************/

/*YOU CAN USE THE CODE BELOW TO RUN SIMPLE FREQUENCIES ON TWO COMMON
VARIABLES, 'LOTOFEFFORT' AND 'TRUSTDOCTOR'*/


/*SAS CODE*/
proc surveyfreq data = mergeHINTS5C1_FDA2 varmethod = jackknife;
weight nwgt0;
repweights nwgt1-nwgt100 / df = 49 jkcoefs = 0.98;
tables lotofeffort trustdoctor;
run;


/*SUDAAN CODE*/

proc crosstab data=mergefda1_fda2 design=jackknife ddf = 49;
weight nwgt0;
jackwgts nwgt1-nwgt100 / adjjack=.98;
class ecig_quit ecig_chemicals;
tables lotofeffort trustdoctor;
run;
```

SAS Code to Set Up Final and Replicate Weights for the Taylor Series Estimation Method

```
/*THIS CODE MERGES TWO TEMPORARY HINTS DATA SETS CREATED USING THE TAYLOR SERIES
LINEARZATION METHOD. PLEASE NOTE, THIS CODE IS BASED ON THE ASSUMPTION THAT THE
DATA SETS HAVE THE CORRECT VARIANCE CODES AND HHID VARIABLES MATCH*/

/*FIRST CREATE THE FORMAT FOR THE SURVEY VARIABLE*/
proc format;

value survey
1="HINTS 5 CYCLE 1"
2="FDA-2"
;
run;

/*********************************************************************/

/*CREATE TWO SEPARATE TEMPORARY DATA FILES THAT CONTAIN THE NEW 'SURVEY'
VARIABLE.*/

options fmtsearch=(HINTS5C1); /*PUT NAME OF LIBRARY WHERE HINTS 5
                                CYCLE 1 FORMATS ARE STORED*/
data tempHINTS5CYCLE1;
set HINTS5C1.hints5_cycle1_public; /*PUT NAME OF LIBRARY AND
                           NAME OF EXISTING HINTS 5 CYCLE 1 FILE*/
survey=1;
format survey survey.;
run;

options fmtsearch=(FDA2); /* PUT NAME OF LIBRARY WHERE HINTS 5 CYCLE 1
FORMATS ARE STORED*/

data tempFDA2;
set FDA2.hints_fda2_public; /*PUT NAME OF LIBRARY AND NAME
                                      OF EXISTING FDA-2 DATA FILE*/
survey=2;
format survey survey.;
run;

/*****************************************************************/

data mergeHINTS_FDA2_HINTS5C1;
      set tempFDA2 tempHINTS5CYCLE1;
run;

/*****************************************************/
/*YOU CAN USE THE CODE BELOW TO RUN SIMPLE FREQUENCIES ON TWO COMMON
VARIABLES, 'SEEKHEALTHINFO' AND 'CHANCEASKQUESTIONS'*/

/*SAS CODE*/
proc surveyfreq data = mergeHINTS_FDA2_HINTS5C1 varmethod = TAYLOR;
      strata VAR_STRATUM;
      cluster VAR_CLUSTER;
      weight MERGED_FINWT0;
      tables seekhealthinfo chanceaskquestions / row col;
run;
```

Appendix F: Merging HINTS 4, Cycle 4 and HINTS 5, Cycle 1 using SPSS

This section provides SPSS (Version 25) syntax for merging the HINTS 5, Cycle 1 and HINTS 4, Cycle 4 iterations. Within the below example SPSS syntax, a new "survey" variable is created in both datasets that will distinguish between the two iterations once the datasets are merged. Next, the two files are merged into one. It will match up variables that have the same name and format and create a merged data file (n = 6,962).

First, you will need to have **HINTS 5, Cycle 1** data open. The below syntax will first save a copy of HINTS 5, Cycle 1 and rename it as a new file called 'MERGED_H5C1_HFDAC2.sav'. We highly suggest this step for several reasons, mainly being that when SPSS merges datasets the old file may be overwritten. By saving your original datafile, you can always have this available to refer to. Next, the syntax will rename the dataset to help with making sure the correct dataset is active and being edited in later syntax.
Next, the below syntax copies Cycle 1's weighting variable person_finwt0 so that both cycles' weighting variable names match (nwgt0). Finally, the syntax creates a new variable called 'Survey' and gives each participant in HINTS 5 Cycle 1 a "2" so that analysts can easily identify cases from HINTS 5, Cycle 1.

```
SAVE OUTFILE='H:\HINTS\5 Cycle 1\SPSS\MERGED_H5C1_H4C4.sav '
 /COMPRESSED.
DATASET NAME MERGED_DATA.

DATASET ACTIVATE MERGED_DATA.
ALTER TYPE VAR_STRATUM (A10).
COMPUTE MERGED_FINWT0=PERSON_FINWT0.
COMPUTE Survey=2.
EXECUTE.
```

Next, we need to open our HINTS 4 Cycle 4 data and rename our datafile, again to help with keeping files aligned for the merging process below. The following code will open your HINTS 4 Cycle 4 data and rename the dataset as H4C4. The syntax will then create the 'Survey' variable in the HINTS 4 Cycle 4 dataset and give each participant from Cycle 4 a value of "1". Again, this is so that once the datasets are merged, analysts can easily identify which cases were from the HINTS 4 Cycle 4 dataset. Finally, the syntax creates copies the weighting variable Person_FINWT0 and names it MERGED_FINWT0 so that the key weighting variable matches the key weighting variable from our HINTS 5 Cycle 1 dataset
Note, the analyst will need to insert the file path for where HINTS 4 Cycle 4 is saved.

```
**below, you should insert the filepath for your HINTS 4 Cycle 4 data**.
GET FILE='H:\HINTS\HINTS 4 Cycle 4\HINTS4Cycle4_SPSS\ Hints4Cycle4_09062017_public'.
ALTER TYPE VAR_STRATUM (A10).
DATASET NAME H4C4 WINDOW=FRONT.
COMPUTE MERGED_FINWT0=Person_FINWT0.
COMPUTE Survey=1.
EXECUTE.
```

Next, a plan file is required to conduct analyses in SPSS. To create a plan file and subsequently conduct analyses, paste the following syntax in the SPSS Syntax Editor:

```
* Analysis Preparation Wizard.
*INSERT DATH OF PATH TO SAMPLE DESIGN FILE IN /PLAN FILE=.
CSPLAN ANALYSIS
```

```
/PLAN FILE='H:\HINTS\5 Cycle 1\SPSS\MergePlan.csaplan'
/PLANVARS ANALYSISWEIGHT=MERGED_FINWT0
/SRSESTIMATOR TYPE=WOR
/PRINT PLAN
/DESIGN STRATA=VAR_STRATUM CLUSTER=VAR_CLUSTER
/ESTIMATOR TYPE=WR.
```

Once you have your plan file, you can begin the merging process. You should, by this point, have two datasets open: “MERGED_H5C1_HFDAC2.SAV” (which currently contains only HINTS 5 Cycle 1’s data) and “H4C4”. Within your “MERGED_H5C1_HFDAC2.SAV” dataset you will navigate to the “Data” dropdown and select “Merge Files”. You will be given the option to merge by cases or variables. Because we are merging two different cycles with mostly the same variables, we will want to select merge by “Add Cases”. You will then select the hints4_cycle4_public dataset that is open from the window that pops up and click continue. Ensure that the variables you need in the new merged dataset you are creating are in the “Variables in New Active Dataset” box. Once you have verified all your desired variables are in that box, click “OK”.

```
DATASET ACTIVATE MERGED_DATA.
ADD FILES /FILE=*
  /RENAME (AccessedFamRec_MyPwd AccessedFamRec_TheirPwd AccessFamilyMedRec AccessOnlineRecord
    AlcoholConditions_Cancer AlcoholConditions_Cholesterol AlcoholConditions_Diabetes
    AlcoholConditions_HeartDisease AlcoholConditions_LiverDisease AlcoholConditions_Overweight
    AlcoholIncreaseCancer AlcoholReduceHeart CancerFatal Caregiving_AcuteCond Caregiving_Aging
    Caregiving_Alzheimers Caregiving_Cancer Caregiving_Child Caregiving_ChronicCond Caregiving_Family
    Caregiving_Friend Caregiving_HoursPerWeek Caregiving_MentalHealth Caregiving_NeuroDev Caregiving_No
    Caregiving_NotSure Caregiving_OrthoMusc Caregiving_Other Caregiving_Other_OS Caregiving_Parent
    Caregiving_Spouse CaregivingCond_Cat CaregivingWho_Cat ConcernedQuality ConfidentGetHealthInf Deaf
    eCigUse edu Electronic_BuyMedicine Electronic_CompletedForms Electronic_HCPSearch
    Electronic_HealthInfoSE Electronic_MadeAppts Electronic_SelfHealthInfo Electronic_TalkDoctor
    Electronic_TestResults Electronic_TrackedHealthCosts ESent_AnotherHCP ESent_Family ESent_HealthApp
    flippededu flippedgender Frustrated gender GeneticTestUse_Cat GeneticTestUse_DetermineMed
    GeneticTestUse_DeterminePass GeneticTestUse_DetermineRisk GeneticTestUse_DetermineTx
    genhealth_recode HadTest_Ancestry HadTest_BRCA HadTest_Cat HadTest_CFCarrier HadTest_DNAFing
    HadTest_Lynch HadTest_None HadTest_NotSure HadTest_Other HadTest_Other_OS HadTest_Paternity
    HCPEncourageOnlineRec HealthIns_IHS HealthIns_InsuranceEmp HealthIns_InsurancePriv
    HealthIns_Medicaid HealthIns_Medicare HealthIns_Other HealthIns_Other_OS HealthIns_Tricare
    HealthIns_VA HeardDNATest IntRsn_SharedSocNet IntRsn_SupportGroup IntRsn_VisitedSocNet
    IntRsn_WroteBlog IntRsn_YouTube LotOfEffort MostRecentCheckup2 NCHSURCODE2013
    NotAccessed_ConcernedPrivacy NotAccessed_NoInternet NotAccessed_NoNeed NotAccessed_NoRecord
    NotAccessed_Other NotAccessed_Other_OS NotAccessed_SpeakDirectly OfferedAccessHCP2
    OfferedAccessInsurer2 OtherDevTrackHealth PR_RUCA_2010 ProbCare_BringTest ProbCare_ProvideHist
    ProbCare_RedoTest ProbCare_WaitLong ProviderMaintainEMR2 RecordsOnline_AddHealthInfo
    RecordsOnline_Allergies RecordsOnline_ClinNotes RecordsOnline_DownloadHealth
    RecordsOnline_HealthProbs RecordsOnline_Immunizations RecordsOnline_Labs RecordsOnline_MakeAppt
    RecordsOnline_MakeDecision RecordsOnline_Meds RecordsOnline_MessageHCP RecordsOnline_MonitorHealth
    RecordsOnline_Paperwork RecordsOnline_RefillMeds RecordsOnline_RequestCorrection
    RecordsOnline_ViewResults RecordsOnline_VisitSummary SEC_RUCA_2010 seekcancerinfo_recode
    SexualOrientation SexualOrientation_OS SharedHealthDeviceInfo SkinCancerHPExam SkinCancerSelfCheck
    StrongNeedHealthInfo StrongNeedHealthInfo_OS Tablet_AchieveGoal Tablet_DiscussionsHCP
    Tablet_MakeDecision TabletHealthWellnessApps TextFromDoctor TimesStrengthTraining TooHardUnderstand
    TrustCharities TrustDoctor TrustFamily TrustGov TrustInternet TrustNewsMag TrustRadio
    TrustReligiousOrgs TrustTelevision UnderstandOnlineMedRec UsedECigEver UseECigNow
    UsefulOnlineMedRec WhereUseInternet_GamingDevice WhereUseInternet_Home
    WhereUseInternet_MobileDevice WhereUseInternet_PublicPlace WhereUseInternet_School
    WhereUseInternet_Work=d0 d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12 d13 d14 d15 d16 d17 d18 d19 d20 d21
```

```
  d22 d23 d24 d25 d26 d27 d28 d29 d30 d31 d32 d33 d34 d35 d36 d37 d38 d39 d40 d41 d42 d43 d44 d45 d46
  d47 d48 d49 d50 d51 d52 d53 d54 d55 d56 d57 d58 d59 d60 d61 d62 d63 d64 d65 d66 d67 d68 d69 d70 d71
  d72 d73 d74 d75 d76 d77 d78 d79 d80 d81 d82 d83 d84 d85 d86 d87 d88 d89 d90 d91 d92 d93 d94 d95 d96
  d97 d98 d99 d100 d101 d102 d103 d104 d105 d106 d107 d108 d109 d110 d111 d112 d113 d114 d115 d116
  d117 d118 d119 d120 d121 d122 d123 d124 d125 d126 d127 d128 d129 d130 d131 d132 d133 d134 d135 d136
  d137 d138 d139 d140 d141 d142 d143 d144 d145 d146 d147 d148 d149 d150 d151 d152 d153 d154 d155 d156
  d157)
 /FILE='H4C4'
 /RENAME (AccessFamilyInfo AccessOwnInfo AccessPHR AverageDailyTVGames Behaviors_Cancer
  Behaviors_Diabetes Behaviors_HeartDisease Behaviors_HighBP Behaviors_Obesity CancerConcernedQuality
  CancerConfidentGetHealthInf CancerFrustrated CancerLotOfEffort CancerTooHardUnderstand
  CancerTreatmentDecisionLow CancerTreatmentDecisionMod CancerTrustCharities CancerTrustDoctor
  CancerTrustFamily CancerTrustGov CancerTrustInternet CancerTrustNewsMag CancerTrustRadio
  CancerTrustReligiousOrgs CancerTrustTelevision ConfidentControlInfo ConflictingOpinionsTestTx
  DrShouldPSATest DrTakeCareNeeds DrTalkDiffColCaTests DrTalkMammogram DrTalkPapTest
  ExerciseIntention FaxedInfoSafe FruitIntent Genetics_Cancer Genetics_Diabetes Genetics_HeartDisease
  Genetics_HighBP Genetics_Obesity GetHealthInsurance HaveDevice_Cat HaveDevice_None
  HealthApps_AchieveGoal HealthApps_MakeDecision HealthApps_NewQuestions HeardGeneticTest
  HeardLungTest InterestedEInfo_ApptRemind InterestedEInfo_Diagnostics InterestedEInfo_GenHealth
  InterestedEInfo_Images InterestedEInfo_LabResults InterestedEInfo_Lifestyle
  InterestedEInfo_MedRemind InterestedEInfo_Symptoms InterestedEInfo_Vitals MedicalResearchStudy
  MedInfo_App MedInfo_Cat MedInfo_Email MedInfo_Fax MedInfo_None MedInfo_SocMed MedInfo_Text
  MedInfo_Video MedTests_DefiniteCancer MedTests_HarmsOutweigh MedTests_LikelyCancer
  MedTests_MoreTests MostRecentCheckup OfferedAccessHCP OfferedAccessInsurer PCCScale
  ProviderMaintainEMR PTEngage_EverEngaged PTEngage_HeardOf PTEngage_Interested QuittingReduceHarm
  RegExercise_Appearance RegExercise_Enjoyment RegExercise_Guilt RegExercise_Pressure
  RegularSodaIntention Shading_Flag ShareEMR SmokerReduce_Exercise SmokerReduce_FruitVeg
  SmokerReduce_Sleep SmokerReduce_Vitamins SmokingOpinion Sunscreen TabletSmartPh_HealthApps
  Treatment_C4 VegetablesIntent WeightIntention WeightOpinion=d158 d159 d160 d161 d162 d163 d164 d165
  d166 d167 d168 d169 d170 d171 d172 d173 d174 d175 d176 d177 d178 d179 d180 d181 d182 d183 d184 d185
  d186 d187 d188 d189 d190 d191 d192 d193 d194 d195 d196 d197 d198 d199 d200 d201 d202 d203 d204 d205
  d206 d207 d208 d209 d210 d211 d212 d213 d214 d215 d216 d217 d218 d219 d220 d221 d222 d223 d224 d225
  d226 d227 d228 d229 d230 d231 d232 d233 d234 d235 d236 d237 d238 d239 d240 d241 d242 d243 d244 d245
  d246 d247 d248 d249 d250 d251 d252 d253 d254)
 /DROP=d0 d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12 d13 d14 d15 d16 d17 d18 d19 d20 d21 d22 d23 d24
  d25 d26 d27 d28 d29 d30 d31 d32 d33 d34 d35 d36 d37 d38 d39 d40 d41 d42 d43 d44 d45 d46 d47 d48 d49
  d50 d51 d52 d53 d54 d55 d56 d57 d58 d59 d60 d61 d62 d63 d64 d65 d66 d67 d68 d69 d70 d71 d72 d73 d74
  d75 d76 d77 d78 d79 d80 d81 d82 d83 d84 d85 d86 d87 d88 d89 d90 d91 d92 d93 d94 d95 d96 d97 d98 d99
  d100 d101 d102 d103 d104 d105 d106 d107 d108 d109 d110 d111 d112 d113 d114 d115 d116 d117 d118 d119
  d120 d121 d122 d123 d124 d125 d126 d127 d128 d129 d130 d131 d132 d133 d134 d135 d136 d137 d138 d139
  d140 d141 d142 d143 d144 d145 d146 d147 d148 d149 d150 d151 d152 d153 d154 d155 d156 d157 d158 d159
  d160 d161 d162 d163 d164 d165 d166 d167 d168 d169 d170 d171 d172 d173 d174 d175 d176 d177 d178 d179
  d180 d181 d182 d183 d184 d185 d186 d187 d188 d189 d190 d191 d192 d193 d194 d195 d196 d197 d198 d199
  d200 d201 d202 d203 d204 d205 d206 d207 d208 d209 d210 d211 d212 d213 d214 d215 d216 d217 d218 d219
  d220 d221 d222 d223 d224 d225 d226 d227 d228 d229 d230 d231 d232 d233 d234 d235 d236 d237 d238 d239
  d240 d241 d242 d243 d244 d245 d246 d247 d248 d249 d250 d251 d252 d253 d254.
EXECUTE.
/*********************************************************/

**YOU CAN USE THE CODE BELOW TO RUN SIMPLE FREQUENCIES ON TWO COMMON
VARIABLES, ‘seekhealthinfo’ AND ‘chanceaskquestions’*/ /*SPSS CODE***.

*INSERT PATH OF TO ANALYSIS PLAN UNDER /PLAN FILE.
CSTABULATE
/PLAN FILE='H:\HINTS\5 Cycle 1\SPSS\MergePlan.csaplan'
/TABLES VARIABLES=  seekhealthinfo chanceaskquestions
```

```
/CELLS POPSIZE TABLEPCT
/STATISTICS SE COUNT
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

Appendix G: Merging HINTS 4, Cycle 4 and HINTS 5, Cycle 1 using SPSS

This section provides SPSS (Version 25) syntax for merging the HINTS 5, Cycle 1 and HINTS 4, Cycle 4 iterations. Within the below example SPSS syntax, a new "survey" variable is created in both datasets that will distinguish between the two iterations once the datasets are merged. Next, the two files are merged into one. It will match up variables that have the same name and format and create a merged data file (n = 6,962).

First, you will need to have **HINTS 5, Cycle 1** data open. The below syntax will first save a copy of HINTS 5, Cycle 1 and rename it as a new file called 'MERGED_H5C1_H4C4.sav'. We highly suggest this step for several reasons, mainly being that when SPSS merges datasets the old file may be overwritten. By saving your original datafile, you can always have this available to refer to. Next, the syntax will rename the dataset to help with making sure the correct dataset is active and being edited in later syntax.

Next, the below syntax copies Cycle 1's weighting variable person_finwt0 so that both cycles' weighting variable names match (MERGED_FINWT0). Finally, the syntax creates a new variable called 'Survey' and gives each participant in HINTS 5 Cycle 1 a "2" so that analysts can easily identify cases from HINTS 5, Cycle 1.

```
SAVE OUTFILE='H:\HINTS\5 Cycle 1\SPSS\MERGED_H5C1_H4C4.sav '
 /COMPRESSED.
DATASET NAME MERGED_DATA.

DATASET ACTIVATE MERGED_DATA.
ALTER TYPE VAR_STRATUM (A10).
COMPUTE MERGED_FINWT0=PERSON_FINWT0.
COMPUTE Survey=2.
EXECUTE.
```

Next, we need to open our HINTS 4 Cycle 4 data and rename our datafile, again to help with keeping files aligned for the merging process below. The following code will open your HINTS 4 Cycle 4 data and rename the dataset as H4C4. The syntax will then create the 'Survey' variable in the HINTS 4 Cycle 4 dataset and give each participant from Cycle 4 a value of "1". Again, this is so that once the datasets are merged, analysts can easily identify which cases were from the HINTS 4 Cycle 4 dataset. Finally, the syntax creates copies the weighting variable Person_FINWT0 and names it MERGED_FINWT0 so that the key weighting variable matches the key weighting variable from our HINTS 5 Cycle 1 dataset

Note, the analyst will need to insert the file path for where HINTS 4 Cycle 4 is saved.

```
**below, you should insert the filepath for your HINTS 4 Cycle 4 data**.
GET FILE='H:\HINTS\HINTS 4 Cycle 4\HINTS4Cycle4_SPSS\ Hints4Cycle4_09062017_public'.
ALTER TYPE VAR_STRATUM (A10).
DATASET NAME H4C4 WINDOW=FRONT.
COMPUTE MERGED_FINWT0=Person_FINWT0.
COMPUTE Survey=1.
EXECUTE.
```

Next, a plan file is required to conduct analyses in SPSS. To create a plan file and subsequently conduct analyses, paste the following syntax in the SPSS Syntax Editor:

```
* Analysis Preparation Wizard.
*INSERT DATH OF PATH TO SAMPLE DESIGN FILE IN /PLAN FILE=.
CSPLAN ANALYSIS
 /PLAN FILE='H:\HINTS\5 Cycle 1\SPSS\MergePlan.csaplan'
 /PLANVARS ANALYSISWEIGHT=MERGED_FINWT0
 /SRSESTIMATOR TYPE=WOR
 /PRINT PLAN
 /DESIGN STRATA=VAR_STRATUM CLUSTER=VAR_CLUSTER
 /ESTIMATOR TYPE=WR.
```

Once you have your plan file, you can begin the merging process. You should, by this point, have two datasets open: “MERGED_H5C1_H4C4.SAV” (which currently contains only HINTS 5 Cycle 1’s data) and “H4C4”. Within your “MERGED_H5C1_H4C4.SAV” dataset you will navigate to the “Data” dropdown and select “Merge Files”. You will be given the option to merge by cases or variables. Because we are merging two different cycles with mostly the same variables, we will want to select merge by “Add Cases”. You will then select the hints4_cycle4_public dataset that is open from the window that pops up and click continue. Ensure that the variables you need in the new merged dataset you are creating are in the “Variables in New Active Dataset” box. Once you have verified all your desired variables are in that box, click “OK”.

```
DATASET ACTIVATE MERGED_DATA.
ADD FILES /FILE=*
  /RENAME (AccessedFamRec_MyPwd AccessedFamRec_TheirPwd AccessFamilyMedRec AccessOnlineRecord
    AlcoholConditions_Cancer AlcoholConditions_Cholesterol AlcoholConditions_Diabetes
    AlcoholConditions_HeartDisease AlcoholConditions_LiverDisease AlcoholConditions_Overweight
    AlcoholIncreaseCancer AlcoholReduceHeart CancerFatal Caregiving_AcuteCond Caregiving_Aging
    Caregiving_Alzheimers Caregiving_Cancer Caregiving_Child Caregiving_ChronicCond Caregiving_Family
    Caregiving_Friend Caregiving_HoursPerWeek Caregiving_MentalHealth Caregiving_NeuroDev Caregiving_No
    Caregiving_NotSure Caregiving_OrthoMusc Caregiving_Other Caregiving_Other_OS Caregiving_Parent
    Caregiving_Spouse CaregivingCond_Cat CaregivingWho_Cat ConcernedQuality ConfidentGetHealthInf Deaf
    eCigUse edu Electronic_BuyMedicine Electronic_CompletedForms Electronic_HCPSearch
    Electronic_HealthInfoSE Electronic_MadeAppts Electronic_SelfHealthInfo Electronic_TalkDoctor
    Electronic_TestResults Electronic_TrackedHealthCosts ESent_AnotherHCP ESent_Family ESent_HealthApp
    flippededu flippedgender Frustrated gender GeneticTestUse_Cat GeneticTestUse_DetermineMed
    GeneticTestUse_DeterminePass GeneticTestUse_DetermineRisk GeneticTestUse_DetermineTx
    genhealth_recode HadTest_Ancestry HadTest_BRCA HadTest_Cat HadTest_CFCarrier HadTest_DNAFing
    HadTest_Lynch HadTest_None HadTest_NotSure HadTest_Other HadTest_Other_OS HadTest_Paternity
    HCPEncourageOnlineRec HealthIns_IHS HealthIns_InsuranceEmp HealthIns_InsurancePriv
    HealthIns_Medicaid HealthIns_Medicare HealthIns_Other HealthIns_Other_OS HealthIns_Tricare
    HealthIns_VA HeardDNATest IntRsn_SharedSocNet IntRsn_SupportGroup IntRsn_VisitedSocNet
    IntRsn_WroteBlog IntRsn_YouTube LotOfEffort MostRecentCheckup2 NCHSURCODE2013
    NotAccessed_ConcernedPrivacy NotAccessed_NoInternet NotAccessed_NoNeed NotAccessed_NoRecord
    NotAccessed_Other NotAccessed_Other_OS NotAccessed_SpeakDirectly OfferedAccessHCP2
    OfferedAccessInsurer2 OtherDevTrackHealth PR_RUCA_2010 ProbCare_BringTest ProbCare_ProvideHist
    ProbCare_RedoTest ProbCare_WaitLong ProviderMaintainEMR2 RecordsOnline_AddHealthInfo
    RecordsOnline_Allergies RecordsOnline_ClinNotes RecordsOnline_DownloadHealth
    RecordsOnline_HealthProbs RecordsOnline_Immunizations RecordsOnline_Labs RecordsOnline_MakeAppt
    RecordsOnline_MakeDecision RecordsOnline_Meds RecordsOnline_MessageHCP RecordsOnline_MonitorHealth
    RecordsOnline_Paperwork RecordsOnline_RefillMeds RecordsOnline_RequestCorrection
    RecordsOnline_ViewResults RecordsOnline_VisitSummary SEC_RUCA_2010 seekcancerinfo_recode
    SexualOrientation SexualOrientation_OS SharedHealthDeviceInfo SkinCancerHPExam SkinCancerSelfCheck
```

```
  StrongNeedHealthInfo StrongNeedHealthInfo_OS Tablet_AchieveGoal Tablet_DiscussionsHCP
  Tablet_MakeDecision TabletHealthWellnessApps TextFromDoctor TimesStrengthTraining TooHardUnderstand
  TrustCharities TrustDoctor TrustFamily TrustGov TrustInternet TrustNewsMag TrustRadio
  TrustReligiousOrgs TrustTelevision UnderstandOnlineMedRec UsedECigEver UseECigNow
  UsefulOnlineMedRec WhereUseInternet_GamingDevice WhereUseInternet_Home
  WhereUseInternet_MobileDevice WhereUseInternet_PublicPlace WhereUseInternet_School
  WhereUseInternet_Work=d0 d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12 d13 d14 d15 d16 d17 d18 d19 d20 d21
  d22 d23 d24 d25 d26 d27 d28 d29 d30 d31 d32 d33 d34 d35 d36 d37 d38 d39 d40 d41 d42 d43 d44 d45 d46
  d47 d48 d49 d50 d51 d52 d53 d54 d55 d56 d57 d58 d59 d60 d61 d62 d63 d64 d65 d66 d67 d68 d69 d70 d71
  d72 d73 d74 d75 d76 d77 d78 d79 d80 d81 d82 d83 d84 d85 d86 d87 d88 d89 d90 d91 d92 d93 d94 d95 d96
  d97 d98 d99 d100 d101 d102 d103 d104 d105 d106 d107 d108 d109 d110 d111 d112 d113 d114 d115 d116
  d117 d118 d119 d120 d121 d122 d123 d124 d125 d126 d127 d128 d129 d130 d131 d132 d133 d134 d135 d136
  d137 d138 d139 d140 d141 d142 d143 d144 d145 d146 d147 d148 d149 d150 d151 d152 d153 d154 d155 d156
  d157)
 /FILE='H4C4'
 /RENAME (AccessFamilyInfo AccessOwnInfo AccessPHR AverageDailyTVGames Behaviors_Cancer
  Behaviors_Diabetes Behaviors_HeartDisease Behaviors_HighBP Behaviors_Obesity CancerConcernedQuality
  CancerConfidentGetHealthInf CancerFrustrated CancerLotOfEffort CancerTooHardUnderstand
  CancerTreatmentDecisionLow CancerTreatmentDecisionMod CancerTrustCharities CancerTrustDoctor
  CancerTrustFamily CancerTrustGov CancerTrustInternet CancerTrustNewsMag CancerTrustRadio
  CancerTrustReligiousOrgs CancerTrustTelevision ConfidentControlInfo ConflictingOpinionsTestTx
  DrShouldPSATest DrTakeCareNeeds DrTalkDiffColCaTests DrTalkMammogram DrTalkPapTest
  ExerciseIntention FaxedInfoSafe FruitIntent Genetics_Cancer Genetics_Diabetes Genetics_HeartDisease
  Genetics_HighBP Genetics_Obesity GetHealthInsurance HaveDevice_Cat HaveDevice_None
  HealthApps_AchieveGoal HealthApps_MakeDecision HealthApps_NewQuestions HeardGeneticTest
  HeardLungTest InterestedEInfo_ApptRemind InterestedEInfo_Diagnostics InterestedEInfo_GenHealth
  InterestedEInfo_Images InterestedEInfo_LabResults InterestedEInfo_Lifestyle
  InterestedEInfo_MedRemind InterestedEInfo_Symptoms InterestedEInfo_Vitals MedicalResearchStudy
  MedInfo_App MedInfo_Cat MedInfo_Email MedInfo_Fax MedInfo_None MedInfo_SocMed MedInfo_Text
  MedInfo_Video MedTests_DefiniteCancer MedTests_HarmsOutweigh MedTests_LikelyCancer
  MedTests_MoreTests MostRecentCheckup OfferedAccessHCP OfferedAccessInsurer PCCScale
  ProviderMaintainEMR PTEngage_EverEngaged PTEngage_HeardOf PTEngage_Interested QuittingReduceHarm
  RegExercise_Appearance RegExercise_Enjoyment RegExercise_Guilt RegExercise_Pressure
  RegularSodaIntention Shading_Flag ShareEMR SmokerReduce_Exercise SmokerReduce_FruitVeg
  SmokerReduce_Sleep SmokerReduce_Vitamins SmokingOpinion Sunscreen TabletSmartPh_HealthApps
  Treatment_C4 VegetablesIntent WeightIntention WeightOpinion=d158 d159 d160 d161 d162 d163 d164 d165
  d166 d167 d168 d169 d170 d171 d172 d173 d174 d175 d176 d177 d178 d179 d180 d181 d182 d183 d184 d185
  d186 d187 d188 d189 d190 d191 d192 d193 d194 d195 d196 d197 d198 d199 d200 d201 d202 d203 d204 d205
  d206 d207 d208 d209 d210 d211 d212 d213 d214 d215 d216 d217 d218 d219 d220 d221 d222 d223 d224 d225
  d226 d227 d228 d229 d230 d231 d232 d233 d234 d235 d236 d237 d238 d239 d240 d241 d242 d243 d244 d245
  d246 d247 d248 d249 d250 d251 d252 d253 d254)
 /DROP=d0 d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12 d13 d14 d15 d16 d17 d18 d19 d20 d21 d22 d23 d24
  d25 d26 d27 d28 d29 d30 d31 d32 d33 d34 d35 d36 d37 d38 d39 d40 d41 d42 d43 d44 d45 d46 d47 d48 d49
  d50 d51 d52 d53 d54 d55 d56 d57 d58 d59 d60 d61 d62 d63 d64 d65 d66 d67 d68 d69 d70 d71 d72 d73 d74
  d75 d76 d77 d78 d79 d80 d81 d82 d83 d84 d85 d86 d87 d88 d89 d90 d91 d92 d93 d94 d95 d96 d97 d98 d99
  d100 d101 d102 d103 d104 d105 d106 d107 d108 d109 d110 d111 d112 d113 d114 d115 d116 d117 d118 d119
  d120 d121 d122 d123 d124 d125 d126 d127 d128 d129 d130 d131 d132 d133 d134 d135 d136 d137 d138 d139
  d140 d141 d142 d143 d144 d145 d146 d147 d148 d149 d150 d151 d152 d153 d154 d155 d156 d157 d158 d159
  d160 d161 d162 d163 d164 d165 d166 d167 d168 d169 d170 d171 d172 d173 d174 d175 d176 d177 d178 d179
  d180 d181 d182 d183 d184 d185 d186 d187 d188 d189 d190 d191 d192 d193 d194 d195 d196 d197 d198 d199
  d200 d201 d202 d203 d204 d205 d206 d207 d208 d209 d210 d211 d212 d213 d214 d215 d216 d217 d218 d219
  d220 d221 d222 d223 d224 d225 d226 d227 d228 d229 d230 d231 d232 d233 d234 d235 d236 d237 d238 d239
  d240 d241 d242 d243 d244 d245 d246 d247 d248 d249 d250 d251 d252 d253 d254.
EXECUTE.
/*********************************************************/

**YOU CAN USE THE CODE BELOW TO RUN SIMPLE FREQUENCIES ON TWO COMMON
```

```
VARIABLES, 'seekhealthinfo' AND 'chanceaskquestions'*/ /*SPSS CODE***.

*INSERT PATH OF TO ANALYSIS PLAN UNDER /PLAN FILE.
CSTABULATE
/PLAN FILE='H:\HINTS\5 Cycle 1\SPSS\MergePlan.csaplan'
/TABLES VARIABLES=  seekhealthinfo chanceaskquestions
/CELLS POPSIZE TABLEPCT
/STATISTICS SE COUNT
/MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

Appendix H: Merging HINTS 5, Cycle 1 and HINTS FDA, Cycle 2 using SPSS

This section provides SPSS (Version 25) syntax for merging the HINTS 5, Cycle 1 and HINTS FDA, Cycle 2 iterations. Within the below example SPSS syntax, a new “survey” variable is created in both datasets that will distinguish between the two iterations once the datasets are merged. Next, the two files are merged into one. It will match up variables that have the same name and format and create a merged data file (n = 5,021).

First, you will need to have **<u>HINTS 5, Cycle 1</u>** data open. The below syntax will first save a copy of HINTS 5, Cycle 1 and rename it as a new file called ‘MERGED_H5C1_HFDAC2.sav’. We highly suggest this step for several reasons, mainly being that when SPSS merges datasets the old file may be overwritten. By saving your original datafile, you can always have this available to refer to. Next, the syntax will rename the dataset to help with making sure the correct dataset is active and being edited in later syntax.

Next, the below syntax copies Cycle 1’s weighting variable person_finwt0 so that both cycles’ weighting variable names match (MERGED_FINWT0). Finally, the syntax creates a new variable called ‘Survey’ and gives each participant in HINTS 5 Cycle 1 a “2” so that analysts can easily identify cases from HINTS 5, Cycle 1.

```
SAVE OUTFILE='H:\HINTS\5 Cycle 1\SPSS\MERGED_H5C1_HFDAC2.sav '
 /COMPRESSED.
DATASET NAME MERGED_DATA.

DATASET ACTIVATE MERGED_DATA.
ALTER TYPE VAR_STRATUM (A10).
COMPUTE MERGED_FINWT0=PERSON_FINWT0.
COMPUTE Survey=2.
EXECUTE.
```

Next, we need to open our HINTS FDA CYCLE 2 data and rename our datafile, again to help with keeping files aligned for the merging process below. The following code will open your HINTS FDA Cycle 2 data and rename the dataset as HFDAC2. The syntax will then create the ‘Survey’ variable in the HINTS FDA, Cycle 2 dataset and give each participant from Cycle 2 a value of “1”. Again, this is so that once the datasets are merged, analysts can easily identify which cases were from the HINTS FDA Cycle 2 dataset. Finally, the syntax creates copies the weighting variable Person_FINWT0 and names it MERGED_FINWT0 so that the key weighting variable matches the key weighting variable from our HINTS 5 Cycle 1 dataset.

Note, the analyst will need to insert the file path for where HINTS FDA Cycle 2 is saved.

```
**below, you should insert the filepath for your HINTS FDA Cycle 2 data**.
GET FILE='H:\HINTS\FDA Cycle 2\HINTS-FDA_Cycle2_SPSS\hints-FDA_cycle2_public.sav'.
ALTER TYPE VAR_STRATUM (A10).
DATASET NAME HFDAC2 WINDOW=FRONT.
COMPUTE MERGED_FINWT0=Person_FINWT0.
COMPUTE Survey=1.
```

```
EXECUTE.
```

Next, a plan file is required to conduct analyses in SPSS. To create a plan file and subsequently conduct analyses, paste the following syntax in the SPSS Syntax Editor:

```
* Analysis Preparation Wizard.
*INSERT DATH OF PATH TO SAMPLE DESIGN FILE IN /PLAN FILE=.
CSPLAN ANALYSIS
 /PLAN FILE='H:\HINTS\5 Cycle 1\SPSS\MergePlan.csaplan'
 /PLANVARS ANALYSISWEIGHT=MERGED_FINWT0
 /SRSESTIMATOR TYPE=WOR
 /PRINT PLAN
 /DESIGN STRATA=VAR_STRATUM CLUSTER=VAR_CLUSTER
 /ESTIMATOR TYPE=WR.
```

Once you have your plan file, you can begin the merging process. You should, by this point, have two datasets open: “MERGED_H5C1_HFDAC2.SAV” (which currently contains only HINTS 5 Cycle 1’s data) and “hints-FDA_cycle2_public”. Within your “MERGED_H5C1_HFDAC2.SAV” dataset you will navigate to the “Data” dropdown and select “Merge Files”. You will be given the option to merge by cases or variables. Because we are merging two different cycles with mostly the same variables, we will want to select merge by “Add Cases”. You will then select the hints_fda_cycle2_public dataset that is open from the window that pops up and click continue. Ensure that the variables you need in the new merged dataset you are creating are in the “Variables in New Active Dataset” box. Once you have verified all your desired variables are in that box, click “OK”.

```
DATASET ACTIVATE MERGED_DATA.
ADD FILES /FILE=*
  /RENAME (AccessedFamRec_MyPwd AccessedFamRec_TheirPwd AccessFamilyMedRec AccessOnlineRecord
AgeDX
    AgeGrpA AlcoholConditions_Cancer AlcoholConditions_Cholesterol AlcoholConditions_Diabetes
    AlcoholConditions_HeartDisease AlcoholConditions_LiverDisease AlcoholConditions_Overweight
    AlcoholIncreaseCancer AlcoholReduceHeart BMI CaBladder CaBone CaBreast CaCervical CaColon
    CaEndometrial CaHeadNeck CaHodgkins CaLeukemia CaLiver CaLung CaMelanoma Cancer_Cat
    CancerAbilityToWork CancerDeniedCoverage CancerFatal CancerHurtFinances CancerMoreCommon
    CancerTx_Chemo CancerTx_Other CancerTx_Radiation CancerTx_Surgery CancerTxSummary CaNonHodgkin
    CaOral CaOther CaOther_OS CaOvarian CaPancreatic CaPharyngeal CaProstate CaRectal
    Caregiving_AcuteCond Caregiving_Aging Caregiving_Alzheimers Caregiving_Cancer Caregiving_Child
    Caregiving_ChronicCond Caregiving_Family Caregiving_Friend Caregiving_HoursPerWeek
    Caregiving_MentalHealth Caregiving_NeuroDev Caregiving_No Caregiving_NotSure Caregiving_OrthoMusc
    Caregiving_Other Caregiving_Other_OS Caregiving_Parent Caregiving_Spouse CaregivingCond_Cat
    CaregivingWho_Cat CaRenal CaSkin CaStomach ChanceAskQuestions ChanceGetCancer ClinicalTrialCancerTx
    ConfidentGetHealthInf ConfidentInfoSafe Deaf DiscussedClinicalTrial DiscussHPVVaccination12m
    DrTalkLungTest edu EducB ElectInfoSafe Electronic_BuyMedicine Electronic_CompletedForms
    Electronic_HCPSearch Electronic_HealthInfoSE Electronic_MadeAppts Electronic_SelfHealthInfo
    Electronic_TalkDoctor Electronic_TestResults Electronic_TrackedHealthCosts EmotionalSupport
    ESent_AnotherHCP ESent_Family ESent_HealthApp EverHadCancer EverHadPSATest ExplainedClearly
    FamBetween9and27 FamilyEverHadCancer FeelingsAddressed flippededu flippedgender FreqGoProvider
    FreqWorryCancer Fruit gender GeneralHealth GeneticTestUse_Cat GeneticTestUse_DetermineMed
    GeneticTestUse_DeterminePass GeneticTestUse_DetermineRisk GeneticTestUse_DetermineTx
```

```
genhealth_recode HadTest_Ancestry HadTest_BRCA HadTest_Cat HadTest_CFCarrier HadTest_DNAFing
  HadTest_Lynch HadTest_None HadTest_NotSure HadTest_Other HadTest_Other_OS HadTest_Paternity
  HaveDevice_CellPh HaveDevice_SmartPh HaveDevice_Tablet HCPEncourageOnlineRec HealthIns_IHS
  HealthIns_InsuranceEmp HealthIns_InsurancePriv HealthIns_Medicaid HealthIns_Medicare
  HealthIns_Other HealthIns_Other_OS HealthIns_Tricare HealthIns_VA HeardDNATest HeardHPV
  HeardHPVVaccine2 Height_Feet Height_Inches HelpDailyChores HelpUncertainty HookahLessHarm Hopeless
  HowLongFinishTreatment_Cat HowLongModerateExerciseHr HowLongModerateExerciseMn
HPVCauseCancer_Anal
  HPVCauseCancer_Cervical HPVCauseCancer_Oral HPVCauseCancer_Penile HPVMedicalTreatment
  HPVShotPrevent HPVSTD Internet_BroadBnd Internet_Cell Internet_DialUp Internet_WiFi
  InternetCancerInfoSelf IntRsn_VisitedSocNet IntRsn_WroteBlog InvolvedDecisions LittleInterest
  MedConditions_Arthritis MedConditions_Depression MedConditions_Diabetes
  MedConditions_HeartCondition MedConditions_HighBP MedConditions_LungDisease MostRecentCheckup2
  Nervous NotAccessed_ConcernedPrivacy NotAccessed_NoInternet NotAccessed_NoNeed NotAccessed_NoRecord
  NotAccessed_Other NotAccessed_Other_OS NotAccessed_SpeakDirectly OfferedAccessHCP2
  OfferedAccessInsurer2 OtherDevTrackHealth OwnAbilityTakeCareHealth PHQ4 ProbCare_BringTest
  ProbCare_ProvideHist ProbCare_RedoTest ProbCare_WaitLong ProviderMaintainEMR2 QualityCare
  RatherNotKnowChance ReceivedCareVA RecommendHPVShot RecordsOnline_AddHealthInfo
  RecordsOnline_Allergies RecordsOnline_ClinNotes RecordsOnline_DownloadHealth
  RecordsOnline_HealthProbs RecordsOnline_Immunizations RecordsOnline_Labs RecordsOnline_MakeAppt
  RecordsOnline_MakeDecision RecordsOnline_Meds RecordsOnline_MessageHCP RecordsOnline_MonitorHealth
  RecordsOnline_Paperwork RecordsOnline_RefillMeds RecordsOnline_RequestCorrection
  RecordsOnline_ViewResults RecordsOnline_VisitSummary RegularProvider SeekCancerInfo
  seekcancerinfo_recode SharedHealthDeviceInfo SkinCancerHPExam SkinCancerSelfCheck SpentEnoughTime
  StrongNeedHealthInfo StrongNeedHealthInfo_OS Tablet_AchieveGoal Tablet_DiscussionsHCP
  Tablet_MakeDecision TabletHealthWellnessApps TalkHealthFriends TanningBed TextFromDoctor
  TimeSinceDX TimesModerateExercise TimesStrengthTraining TriedQuit TrustInternet TrustNewsMag
  TrustRadio TrustTelevision UndergoCancerTreatment UnderstandOnlineMedRec UnderstoodNextSteps
  UsefulOnlineMedRec UseMenuCalorieInfo Vegetables WeeklyMinutesModerateExercise Weight
  WhenDiagnosedCancer WhenMammogram WhenPapTest WhoLookingFor WithheldInfoPrivacy Worrying=d0 d1 d2
  d3 d4 d5 d6 d7 d8 d9 d10 d11 d12 d13 d14 d15 d16 d17 d18 d19 d20 d21 d22 d23 d24 d25 d26 d27 d28
  d29 d30 d31 d32 d33 d34 d35 d36 d37 d38 d39 d40 d41 d42 d43 d44 d45 d46 d47 d48 d49 d50 d51 d52 d53
  d54 d55 d56 d57 d58 d59 d60 d61 d62 d63 d64 d65 d66 d67 d68 d69 d70 d71 d72 d73 d74 d75 d76 d77 d78
  d79 d80 d81 d82 d83 d84 d85 d86 d87 d88 d89 d90 d91 d92 d93 d94 d95 d96 d97 d98 d99 d100 d101 d102
  d103 d104 d105 d106 d107 d108 d109 d110 d111 d112 d113 d114 d115 d116 d117 d118 d119 d120 d121 d122
  d123 d124 d125 d126 d127 d128 d129 d130 d131 d132 d133 d134 d135 d136 d137 d138 d139 d140 d141 d142
  d143 d144 d145 d146 d147 d148 d149 d150 d151 d152 d153 d154 d155 d156 d157 d158 d159 d160 d161 d162
  d163 d164 d165 d166 d167 d168 d169 d170 d171 d172 d173 d174 d175 d176 d177 d178 d179 d180 d181 d182
  d183 d184 d185 d186 d187 d188 d189 d190 d191 d192 d193 d194 d195 d196 d197 d198 d199 d200 d201 d202
  d203 d204 d205 d206 d207 d208 d209 d210 d211 d212 d213 d214 d215 d216 d217 d218 d219 d220 d221 d222
  d223 d224 d225 d226 d227 d228 d229 d230 d231 d232 d233 d234 d235 d236 d237 d238 d239 d240 d241 d242
  d243 d244 d245 d246 d247 d248 d249)
 /FILE='HFDAC2'
 /RENAME (Addictive_Cigar Addictive_Cigarette Addictive_ECigs Addictive_Hookah Addictive_Pipe
  Addictive_Rollown Addictive_Smokeless AgeGrpC AnyoneRegulateTobacco AwareQuitlineSite
  BehaviorCauseCancer CigarettesHarmHealth CigarStat CigLessHarm DailySmokelessHarm ECig_Cat
  ECig_Chemicals ECig_Cost ECig_Health ECig_HowFree ECig_Instructions ECig_Other ECig_Other_OS
  ECig_Quit ECig_Reviews ECig_WhereBuy ECigInfoSeek ECigInfoSeek_OS ECigTrustDoctor ECigTrustECigCo
  ECigTrustFamily ECigTrustGovHealth ECigTrustHealthOrgs ECigTrustReligiousOrgs ECigTrustTobaccoCo
```

```
  EverCalledQuitlineSite FDAQualifiedTobacco FDARegulateTobacco FewCigarettesHarmHealth
  FriendsUseTobacco FutureCallQuitlineSite Harm10Cigarettes Harmful_Cigar Harmful_Cigarette
  Harmful_ECig Harmful_Hookah Harmful_Pipe Harmful_RollOwn Harmful_Smokeless IntRsn_DietWebsite
  IntRsn_ExchangedSupport IntRsn_HCProviderSearch IntRsn_HealthInfoSE IntRsn_InfQuitSmoking
  IntRsn_PDADownload IntRsn_SelfHealthInfo IntRsn_TrackedPHR ListenRadio LookedECig
  LowNicotineAddictive LowNicotineBelievable LowNicotineHarmful LowNicotineLungCancer
  NicotineAddictionConcern NicotineCauseCancer NicotineWantSmoke NoAdditivesHarmful
  NoticeHealthInfoInternet NumberCigarsSmoked NumberTobaccoUsers OrganicHarmful PersonalInternet
  PolyUserA PolyUserB ReadHealthInfoInternet ReadNewspaper RecentTobacco_Cat RecentTobacco_Chemicals
  RecentTobacco_Cost RecentTobacco_Health RecentTobacco_Instructions RecentTobacco_NewProd
  RecentTobacco_Other RecentTobacco_Other_OS RecentTobacco_Quitting RecentTobacco_ReduceExp
  RecentTobacco_WhereBuy RegulateTobacco_Cat RegulateTobacco_CDC RegulateTobacco_FDA
  RegulateTobacco_FTC RegulateTobacco_NIH RegulateTobacco_None RegulateTobacco_SurgGen
  RegulateTobacco_TobaccoInd SizeCigarsSmoked smokelessstat SmokeNowCigars SmokingOpinion
  SomeDaysSmokelessHarm ThoughtChemicalsTobacco Tobacco_Cat Tobacco_Chemicals Tobacco_Cost
  Tobacco_Health Tobacco_Instructions Tobacco_Never Tobacco_NewProd Tobacco_Other Tobacco_Other_OS
  Tobacco_Quitting Tobacco_ReduceExp Tobacco_WhereBuy TobaccoConfidentGetHealthInf
  TobaccoEffects_Billboard TobaccoEffects_CommunityEv TobaccoEffects_GovWWW TobaccoEffects_HealthWWW
  TobaccoEffects_Magazines TobaccoEffects_Mailings TobaccoEffects_Newspaper TobaccoEffects_NewsWWW
  TobaccoEffects_POS TobaccoEffects_PublicTrans TobaccoEffects_Radio TobaccoEffects_SocialWWW
  TobaccoEffects_TV TobaccoHeard_Cat TobaccoHeard_ECig TobaccoHeard_Hookah TobaccoHeard_NeverHeard
  TobaccoHeard_Pipe TobaccoHeard_RollOwn TobaccoHeard_Snus TobaccoLessAddictive TobaccoLessHarmful
  TobaccoSaferNow TobaccoTried_Cat TobaccoTried_Hookah TobaccoTried_NeverTried TobaccoTried_Pipe
  TobaccoTried_RollOwn TobaccoTried_Snus TobaccoTrustDoctor TobaccoTrustFamily TobaccoTrustGovHealth
  TobaccoTrustHealthOrgs TobaccoTrustReligiousOrgs TobaccoTrustTobaccoCo TobaccoUserInHH
  TobaccoWithoutChemicals TrustHealthOrgs UsedTobacco20Times UseFlavoredTobacco UseTobaccoNow
  UseTobaccoWakeUp WatchTV WeekendInternet WeekendRadio WeekendWatchTV WhenQuitSmoke
  WhereCigaretteChemicals WhereUseInternet_Other=d250 d251 d252 d253 d254 d255 d256 d257 d258 d259
  d260 d261 d262 d263 d264 d265 d266 d267 d268 d269 d270 d271 d272 d273 d274 d275 d276 d277 d278 d279
  d280 d281 d282 d283 d284 d285 d286 d287 d288 d289 d290 d291 d292 d293 d294 d295 d296 d297 d298 d299
  d300 d301 d302 d303 d304 d305 d306 d307 d308 d309 d310 d311 d312 d313 d314 d315 d316 d317 d318 d319
  d320 d321 d322 d323 d324 d325 d326 d327 d328 d329 d330 d331 d332 d333 d334 d335 d336 d337 d338 d339
  d340 d341 d342 d343 d344 d345 d346 d347 d348 d349 d350 d351 d352 d353 d354 d355 d356 d357 d358 d359
  d360 d361 d362 d363 d364 d365 d366 d367 d368 d369 d370 d371 d372 d373 d374 d375 d376 d377 d378 d379
  d380 d381 d382 d383 d384 d385 d386 d387 d388 d389 d390 d391 d392 d393 d394 d395 d396 d397 d398 d399
  d400 d401 d402 d403 d404 d405 d406 d407 d408 d409 d410 d411 d412)
/DROP=d0 d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12 d13 d14 d15 d16 d17 d18 d19 d20 d21 d22 d23 d24
  d25 d26 d27 d28 d29 d30 d31 d32 d33 d34 d35 d36 d37 d38 d39 d40 d41 d42 d43 d44 d45 d46 d47 d48 d49
  d50 d51 d52 d53 d54 d55 d56 d57 d58 d59 d60 d61 d62 d63 d64 d65 d66 d67 d68 d69 d70 d71 d72 d73 d74
  d75 d76 d77 d78 d79 d80 d81 d82 d83 d84 d85 d86 d87 d88 d89 d90 d91 d92 d93 d94 d95 d96 d97 d98 d99
  d100 d101 d102 d103 d104 d105 d106 d107 d108 d109 d110 d111 d112 d113 d114 d115 d116 d117 d118 d119
  d120 d121 d122 d123 d124 d125 d126 d127 d128 d129 d130 d131 d132 d133 d134 d135 d136 d137 d138 d139
  d140 d141 d142 d143 d144 d145 d146 d147 d148 d149 d150 d151 d152 d153 d154 d155 d156 d157 d158 d159
  d160 d161 d162 d163 d164 d165 d166 d167 d168 d169 d170 d171 d172 d173 d174 d175 d176 d177 d178 d179
  d180 d181 d182 d183 d184 d185 d186 d187 d188 d189 d190 d191 d192 d193 d194 d195 d196 d197 d198 d199
  d200 d201 d202 d203 d204 d205 d206 d207 d208 d209 d210 d211 d212 d213 d214 d215 d216 d217 d218 d219
  d220 d221 d222 d223 d224 d225 d226 d227 d228 d229 d230 d231 d232 d233 d234 d235 d236 d237 d238 d239
  d240 d241 d242 d243 d244 d245 d246 d247 d248 d249 d250 d251 d252 d253 d254 d255 d256 d257 d258 d259
  d260 d261 d262 d263 d264 d265 d266 d267 d268 d269 d270 d271 d272 d273 d274 d275 d276 d277 d278 d279
```

```
   d280 d281 d282 d283 d284 d285 d286 d287 d288 d289 d290 d291 d292 d293 d294 d295 d296 d297 d298 d299
   d300 d301 d302 d303 d304 d305 d306 d307 d308 d309 d310 d311 d312 d313 d314 d315 d316 d317 d318 d319
   d320 d321 d322 d323 d324 d325 d326 d327 d328 d329 d330 d331 d332 d333 d334 d335 d336 d337 d338 d339
   d340 d341 d342 d343 d344 d345 d346 d347 d348 d349 d350 d351 d352 d353 d354 d355 d356 d357 d358 d359
   d360 d361 d362 d363 d364 d365 d366 d367 d368 d369 d370 d371 d372 d373 d374 d375 d376 d377 d378 d379
   d380 d381 d382 d383 d384 d385 d386 d387 d388 d389 d390 d391 d392 d393 d394 d395 d396 d397 d398 d399
   d400 d401 d402 d403 d404 d405 d406 d407 d408 d409 d410 d411 d412.
EXECUTE.
/********************************************************/

**YOU CAN USE THE CODE BELOW TO RUN SIMPLE FREQUENCIES ON TWO COMMON
VARIABLES, ‘lotofeffort’ AND ‘trustdoctor’*/ /*SPSS CODE***.

*INSERT PATH OF TO ANALYSIS PLAN UNDER /PLAN FILE.
CSTABULATE
/PLAN FILE='H:\HINTS\5 Cycle 1\SPSS\MergePlan.csaplan'
/TABLES VARIABLES=lotofeffort trustdoctor
/CELLS POPSIZE TABLEPCT
/STATISTICS SE COUNT
        /MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```