Tutorial for working with BRFSS online data tools and downloading data to run analysis in statistical software
April 2020
INTRODUCTION

- This tutorial will cover topics that relate to use of data collected through Behavioral Risk Factor Surveillance system.

- BRFFS is the system of health-related telephone surveys that collect state data about non-institutionalized, adult U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services.

- The states use a standardized core questionnaire, optional modules, and state-added questions administering the surveys continuously throughout the year.
INTRODUCTION

- The survey is conducted using Random Digit Dialing (RDD) techniques on both landlines and cell phones using two samples.

- Disproportionate stratified sampling (DSS) is used for landline sample i.e. dividing telephone numbers into two density-based strata. The landline sampling ratio of high to medium density is 1:1.5.

- The cellular telephone sample is randomly generated from a sampling frame of confirmed cellular area code and prefix combinations.
INTRODUCTION

- For landline interviews, individual respondents are randomly selected from all adults, living in a household.

- Cellular interviews are treated as one-person households.

- BRFSS data and comprehensive documentation are freely available to download. There is a detailed section in this tutorial that will walk you through this process step-by-step.

- Detailed methodology of the survey has been explained in the data user guide.
# TUTORIAL CONTENTS

<table>
<thead>
<tr>
<th>Online Analysis Tool</th>
<th>Slide Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalence and Trends Data Tool</td>
<td>7</td>
</tr>
<tr>
<td>Chronic Disease Indicators (CDI) Tool</td>
<td>20</td>
</tr>
<tr>
<td>Web Enabled Analysis Tool (WEAT) for Cross Tabulation and Logistic Regression</td>
<td>27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Using data with software</th>
<th>Slide Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downloading BRFSS Data</td>
<td>55</td>
</tr>
<tr>
<td>Getting started with SAS</td>
<td>59</td>
</tr>
<tr>
<td>Getting started with SPSS</td>
<td>68</td>
</tr>
<tr>
<td>Getting started with GIS Data</td>
<td></td>
</tr>
</tbody>
</table>
Prevalence And Trends
Data Tool
PREVALENCE AND TRENDS DATA TOOL

• Quick tool to explore prevalence using a wide range of variables

• Produces charts and maps for individual states or the nation by health topics.

• Particularly useful to produce trends over multiple years.

• Trends are produced as line graphs and single year estimates are produced as maps and bar charts.
Go to the BRFSS data website here: https://www.cdc.gov/brfss/data_tools.htm

Click “Prevalence and Trends Data”
Starts with selection of geographical location and produces trends or single year prevalence

Starts with selection of a health topic and produces only single year prevalence
PREVALENCE OF BINGE DRINKING

- This section will instruct you on how to get prevalence for a specific topic

- For this example, let us see how to get:
  - Age-adjusted prevalence of binge drinking in Nebraska in 2018
  - Trend of age-adjusted prevalence of binge drinking in Nebraska for all available years (2011-2018)
  - Prevalence of binge drinking in Nebraska by gender in 2018 and trend over years (only gives crude prevalence)
• Drop the menu down
• Select Nebraska and click “GO”
• Under “class” select “Alcohol Consumption”
• Under “topic” select “Binge Drinking”
• Under “Year” select “2018”
• Click “go”
• Crude results are displayed below the selection tabs

• Drop down the “Data type” tab and select “Age-adjusted Prevalence”
The results show that in 2018, the age-adjusted prevalence of binge drinking in Nebraska was 22.3%.

**Data Source:** Behavioral Risk Factor Surveillance System (BRFSS)
• Click the gear icon to save the results as PDF or CSV file.
• The outputs are shown
• Click “All Available Years” to get a trend of binge drinking prevalence as shown on the right.

• The age-adjusted prevalence of binge drinking in Nebraska has followed nearly a constant trend from 2011 to 2018.
Drop down “view by” and select gender to get the results on the right. Only crude prevalence is displayed.

In 2018, males had a higher prevalence (27.6%) of binge drinking compared to females (14.9%) in Nebraska (Hover over the bars or convert into tabular form to see the numbers).
• Click “All Available Years” to get trends of binge drinking by gender (crude prevalence)
• Among males, the prevalence of binge drinking in Nebraska shows a decrease in 2013 followed by a steady trend
• The prevalence of binge drinking among females has followed a steady trend from 2011 to 2018
PREVALENCE AND TRENDS DATA TOOL

- Quick tool to explore prevalence using a wide range of variables
- Produces charts and maps for individual states or the nation by health topics.
- Particularly useful to produce side by side tables of user selected indicators.
- Has a wider variety of variables compared to the ‘prevalence and trends data’ tool.
Go to the BRFSS data website here: https://www.cdc.gov/brfss/data_tools.htm

Click “Web Enabled Analysis Tool”
1. To start with a geographic location, drop down ‘Location’, select the state and click ‘Go’.

2. To start with a topic drop down ‘Category’, select topic and click ‘Go’.

3. Use ‘View Indicator Definition’ to see how an indicator has been defined

We will start with location. Select Nebraska and click ‘Go’
Go to the BRFSS data website here: https://www.cdc.gov/brfss/data_tools.htm

Click “Chronic Disease Indicators”

Select alcohol and click ‘Go’
Shows multiple alcohol indicators side by side.

Scroll down your browser to see a total of eight indicator at a time.
1. Click ‘Select indicators to view’

2. Select 8 indicators of your choice and click, ‘OK’.

The 8 indicators will be displayed
Use to see results by gender or race

Use to select year

Click to see data in tabular form

Click the map icon to see US state-level prevalence map for a selected variable

Select ‘All available years’ to see the trend
WEB ENABLED ANALYSIS TOOL (WEAT)
WEB ENABLED ANALYSIS TOOL (WEAT)

• Enables users to create custom crosstabulation tables for variables within selected states

• Can be used to run logistic regression
Go to the BRFSS data website here: https://www.cdc.gov/brfss/data_tools.htm

Click “Web Enabled Analysis Tool”
ORAL HEALTH AND TOBACCO USE

- This section will instruct you on how to use WEAT tool to explore if an association exists between oral health and tobacco use.

- For this example, let us see how to get:
  - A cross-tabulation between current tobacco use and permanent teeth extraction stratified by gender among Nebraska residents in 2018.
  - Gender-adjusted measure of association (odds ratio in this case) between current tobacco use and permanent teeth extraction among Nebraska residents in 2018 (as the outcome variable is binary, we will use logistic regression).
1. Drop down "Select Year" and click "2018"
2. Click next

Using these options, years can be combined or compared. We will not use these in this example.
1. Drop down “Select Location” and click “Nebraska”
2. Click next

Using these options, locations can be combined or compared. We will not use these in this example
We will use our exposure (current tobacco use) as the row variable.

Drop down “Topic” and select “Tobacco use”

Different tobacco related variables will appear.
Select “Calculated variable for adults who are current smokers”
Click “Next”
We will use our outcome (permanent tooth extraction) as the column variable.

Drop down “Topic” and select “Oral Health”

Different oral health related variables will appear.
Select “calculated variable for adults aged 18+ who have had permanent teeth extracted”

Click “Next”
As we want to see results stratified by gender, we will select gender as our control variable.

Click “Include a Control Variable”

Drop down “topic” and select “Demographic Information”

We can select two controls per analysis but for this one we will select only one.
Select “Gender (SEX1)”
Click “Next”
These will be automatically calculated

You can choose the statistic you want

Click “Run Report”
Multiple result tables are produced.

The first table, as shown here, shows overall sample without stratification

In 2018, among former/never smokers living in Nebraska, the prevalence of permanent tooth extraction was 32.5%

In 2018, among current smokers living in Nebraska, the prevalence of permanent tooth extraction was 51.7%

The Chi-Square p-value (≤0.0001) suggests a significant association between smoking status and permanent tooth extraction
The second table, as shown here, shows results only for males

In 2018, among male former/never smokers living in Nebraska, the prevalence of permanent tooth extraction was 35.3%

In 2018, among male current smokers living in Nebraska, the prevalence of permanent tooth extraction was 50.5%

The Chi-Square p-value (<0.0001) suggests a significant association between smoking status and permanent tooth extraction in males
The third table, as shown here, shows results only for females.

In 2018, among female former/never smokers living in Nebraska, the prevalence of permanent tooth extraction was 35.0%.

In 2018, among female current smokers living in Nebraska, the prevalence of permanent tooth extraction was 53.1%.

The Chi-Square p-value (<0.0001) suggests a significant association between smoking status and permanent tooth extraction in females.
As the variable “Gender” had 4 categories including male, female, don’t know/not sure and refused, we get the cross tabulations for the last two categories.

We can ignore these tables as the sample size is too small for any estimate computation. These categories may be converted to missing data in a proper statistical software. This online analysis tool does not provide an option for recoding the variables.
To run logistic regression, go to the starting webpage at https://nccd.cdc.gov/weat/#/analysis and click “Logistic Regression”
1. Drop down “Select year” and click “2018”
2. Click “Next”
Drop down “Select Location” and click “Nebraska”

Click “Next”
Our dependent variable is permanent teeth extraction

Drop down “Topic” and click “Oral health”
Click “Calculated variable for adults aged 18+ who have had permanent teeth extracted”
Click “Next”
Our primary independent variable (exposure) is current tobacco use and we have age as another independent variable.

Drop down “Topic” and click “Tobacco Use”

Click “Calculated variable for adults who are current smokers”

Click “Add”
1. Drop down “Topic” and click “Demographic Information”
2. Click “Gender”
3. Click “Add”
4. Click “Next”
Change the category for dependent variable to “Yes”

Any of the categories for the independent variables can be selected as references. We will select former smoker and male as reference.

Click “Run Report”

1

2

52
The value of Odds Ratio against “Current smoker” is 1.98 and the p-value is <0.0001.

The odds of permanent teeth extraction is 1.98 times higher for current smokers compared to former/never smokers, after adjusting for gender.

As p-value is less than 0.05, the association is statistically significant.

The odds ratio against “Female” is 1.01 and the p-value is also insignificant (0.8921). So there is no significant difference in permanent tooth extraction for females compared to males after adjusting for smoking status.

As the sample sizes for gender categories of “Don’t know/Not sure” and “Refused” are very small, the results associated with them are not precise. This is evident from the very broad confidence intervals.
DOWNLOADING BRFSS DATA
Go to the webpage, “https://www.cdc.gov/brfss/annual_data/annual_2018.html” and select “2018 BRFSS Data (SAS Transport Format)”
Select destination folder and save the zip file

Right click the zip folder and Extract All
Within the extracted folder you will have this SAS export transport file.
GETTING STARTED WITH SAS
Go again to the webpage, “https://www.cdc.gov/brfss/annual_data/annual_2018.html” scroll down and click, “TRANSPORT.SAS [SAS – 3 KB]”. Save this file in the same folder as your data.
First, we will convert the SAS Xport transport file to permanent SAS data set

Double click on the SAS system program, “TRANSPRT” that we downloaded from the BRFSS website. This will open the SAS program
We will have to make changes to two SAS commands in the already provided SAS codes.

We will specify the path where our SAS Xport transport file is saved along with the name of the file. See the example below the command.

We will specify the path of the folder where we want to save the permanent SAS data set. See the example below the command.
• See the changes that have been made
• Specify your own folder pathways
• Make sure to remove the Asterix
• Make sure to remove <>

Select the code and click the running man. Close the SAS program.
Permanent SAS dataset has been saved to the folder.

Specify your own path.

Name of the permanent dataset.

Open SAS
Use this program to open the permanent data set.
SAS CODE FOR CROSS-TABULATION

*READ THE PERMANENT SAS DATASET INTO SAS;
LIBNAME FILE "C:\Users\zaeem\Documents\_BRFSS\LLCP2018XPT"
*PROC CONTENTS SHOWS ALL THE VARIABLES IN THE DATASET;
PROC CONTENTS DATA = FILE.llcp2018 VARNUM;
RUN;
*CREATING A NEW TEMPORARY DATASET FROM PERMANENT DATA SET AND RECODING TO ASSIGN MISSING VALUES TO MATCH THE RESULTS FROM WEAT TOOL;
DATA BRFSS;
SET FILE.llcp2018;
IF SEX1=9 THEN SEX1=.
IF _RFSMOK3=9 THEN _RFSMOK3=.
IF _EXTETH3 = 9 THEN _EXTETH3=.RUN;
*CROSS TABULATION BETWEEN CURRENT SMOKING AND PERMANENT TEETH EXTRACTION BY GENDER;
PROC SURVEYFREQ DATA = BRFSS;
WHERE _STATE = 31; *SPECIFIES STATE = NEBRASKA (31 IS THE FIPS CODE);
CLUSTER _PSU; *SPECIFIES CLUSTER VARIABLE;
STRATA _ststr; *SPECIFIES THE STRATUM VARIABE;
WEIGHT _llcpwt; *SPECIFIES THE WEIGHT VARIABLE;
TABLE SEX1* RFSMOK3* EXTETH3/CHISQ COL ROW;*PRODUCES THE CROSS-TABULATION BY SEX WITH CHI SQUARE TEST. SEX1 IS SEX, _RFSMOK3 IS CURRENT SMOKING AND _EXTETH3 IS EXTRACTION OF PERMANENT TEETH;
run;
Here is one of the resulting tables. The estimates are for men (SEX=1)

We can compare these results to the results of the table produced by WEAT Tool as they are similar

We use proc surveyfreq instead of the traditional proc freq because it takes strata and clusters into account

The variable names and details can be looked up in the code book for BRFSS 2018 data
SAS CODE FOR LOGISTIC REGRESSION

*READ THE PERMANENT SAS DATASET INTO SAS;
LIBNAME FILE "C:\Users\zaeem\Documents\_BRFSS\LLCP2018XPT";
*PROC CONTENTS SHOWS ALL THE VARIABLES IN THE DATASET;
PROC CONTENTS DATA = FILE.llcp2018 VARNUM;
RUN;
*CREATING A NEW TEMPORARY DATASET FROM PERMANENT DATA SET AND RECODING TO ASSIGN MISSING VALUES TO MATCH THE RESULTS FROM WEAT TOOL;
DATA BRFSS;
SET FILE.llcp2018;
IF SEX1=9 THEN SEX1=.;
IF _RFSMOK3=9 THEN _RFSMOK3=.;
IF _EXTETH3 = 9 THEN _EXTETH3=.;RUN;
*LOGISTIC REGRESSION PERMANENT TEETH EXTRACTION AS DEPENDENT VARIABLE AND CURRENT SMOKING AND GENDER AS INDEPENDENT VARIABLES;
PROC SURVEYLOGISTIC DATA= BRFSS;
WHERE _STATE = 31;
CLUSTER _PSU;
STRATA _STSTR;
WEIGHT _llcpwt;
CLASS _RFSMOK3 (REF="1") SEX1 (REF='1'); *SPECIFYING REFERENCE CATEGORIES. NO CURRENT SMOKING AND MALES GENDER ARE REFERECE CATEGORIES;
MODEL _EXTETH3 (EVENT='2')= _RFSMOK3 SEX1; *MODEL WITH TEETH EXTRACTION AS DEPENDENT VARIABLE AND CURRENT SMOKING AND GENDER AS INDEPENDENT VARIABLES;
RUN;
The results are shown to the left.

We can compare these results to the results of the logistic regression produced by WEAT Tool as they are similar.

P value for gender is different because we removed one category of gender that was included by the WEAT tool (SEX1=9)

We use proc surveylogistic instead of the traditional proc logistic because it takes strata and clusters into account

Similarly, we use proc surveyreg for linear regression
GETTING STARTED WITH SPSS
Please review the section, “Downloading BRFFS DATA” starting sliding number 55 as we will start with opening this data set into SPSS
• Open SPSS
• Click, “file”
• Click, “Open”
• Click, “Data”
• Browse to the file where your data is saved

• Drop down, “Files of types” and select, “SAS”
• Click the data file, “LLCP2018.XPT”
• Click, “Open”
• First, we will restrict data to Nebraska
• Click, “Data”
• Click, “Select Cases”
• Select, “If condition is satisfied”
• Click, “If”
• Select, “STATE FIPS CODE”
• Click the arrow to bring it over
• Put, “=31”. This is FIPS code for Nebraska
• Click, “Continue”
• Click, “OK”
• Second, we will recode the variables sex, smoking and permanent teeth extraction to remove unwanted categories and to match the output from WEAT Tool

• Click, “Transform”

• Click, “Recode into different variable”
• Click, “SEX1” and bring it “Numeric Variable → Output Variable”
• Name and Label the variable
• Click, “Old and New Values”
• The variable, “SEX1” has four categories; 1, 2, 7 and 9.
• We want to set 9 to missing and keep the rest as is.
• Write 1 under Old value and 1 under new value. This tells SPSS that we want category 1 as 1 for the new variable Gender.
• Click “Add”
• Do the same for 2 and 7
• For “9”, instead of a value click system missing
• Click, “Continue”
• Click, “Change”
• Click, “OK”
• Similarly recode “current smoking computed variable” and “risk factor for having had permanent teeth extraction” variables and name the new variables current_smoking and permanent_teeth respectively
• Now, let us set up the data for complex analysis
• Click, “Analyze”
• Click, “Complex Samples”
• Click, “Prepare for Analysis"
• Make sure “Create a plan” file is selected
• Click, “Browse”
• Name the file
• Click, “Save”
Click, “Next”
• Click, “SAMPLE DESIGN STRATIFICATION VARIABLE” from the variables list and send it under Strata

• Click “Gender of participants” and send it under strata. This is the variable that we have just created setting 9 to missing

• Click, “PRIMARY SAMPLING UNIT” from the variable list and send it under Clusters

• Click “FINAL WIRGHT: LAND AND CELL-PHONE” send it under Sample Weight

• Click “Next”
• Uncheck “Use the finite population correction (FPC) when estimating variance under simple random sampling assumption”

• Click “Next”
• Click, “Next”
• Click, “Finish”
• The complex sample plan file is saved in your specified folder
Now we will run the cross tabulation

- Click, “Analyze”
- Click “Complex Samples”
- Click “Crosstabs”
• Whenever you open SPSS, browse and select the plan file that we just created. This plan file has the information about cluster, strata and weight.

• Click, “Continue”
• Bring “Current_smoking” under Rows
• Bring Permanent_teeth under column
• Bring “Gender of participants” under Subpopulations
• These variables are the ones that we recoded into new variables specifying the missing values
• The recoded variables always appear at the bottom of the “Variables”
• Click, “Statistics”
You can select the options of your choice or follow what has been shown here

Click, “Continue”
• Click “Missing Values”
• Select the options as shown
• Click, “Continue”
• Click, “OK”
These are two tables from the results.

The upper table is the cross tabulation between current smoking and permanent teeth extraction for males.

As you scroll down in your SPSS, you will see results for other gender categories.

The bottom table shows the results of chi-square test.

Note that we have received the same results as those of the WEAT tool and SASS.
• To perform logistic regression and to use gender as a covariate, we cannot include gender under “Strata” as shown on slide number 86

• Create a complex sample plan file as shown in the previous slides and do not include gender under “Strata”
• Click “Analyze”
• Click “Complex Samples”
• Click “Logistic Regression”
• Click “Browse”
• Select your saved complex sample plan file.
• Click open
• Click “Continue”
• Select “permanent_teeth” from variables list and transfer it under “Dependent Variable”
• Select “current_smoking” from variables list and transfer it under “Factors”
• Select “Gender” from variables list and transfer it under “Factors”
• Click “Reference Category”
• Set the reference category to “lowest value”

• The variable “permanent-teeth” has two categories i.e.
  • 1 = not at risk
  • 2 = at risk

• We want to see the results for those who are at risk, thus we will put 1 as reference (lowest value)

• Click Continue
• Click “Odds Ratios”
• On the resulting screen, transfer both, “current_smoking” and “gender” in the factors list to the cell on the right
The options under “Reference Category” one by one and change to lowest as shown in the lower figure.

It is very important to set your references correctly to get the desired measurements.

Click Continue.
• Click “OK”
• See the SPSS results above
• They are similar to what we get from SAS and WEAT tool