HOW TO ANALYZE SAMHDA DATA

Step-by-step instructions on how to perform logistic regression using SAMHDA data
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Topic</th>
<th>Slide Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Example</td>
<td>3</td>
</tr>
<tr>
<td>Navigate to Site</td>
<td>4-6</td>
</tr>
<tr>
<td>Choose Survey</td>
<td>7-9</td>
</tr>
<tr>
<td>Select Variables</td>
<td>11-24</td>
</tr>
<tr>
<td>Recode Variables</td>
<td>25-34</td>
</tr>
<tr>
<td>Characteristics of sample</td>
<td>35-44</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>44-54</td>
</tr>
</tbody>
</table>
• **Project:** To examine the probability of veterans of abusing substances in 2018.

• **Background:** Veterans experience mental health and substance use disorders at a higher rate than the population at large. The examination of specific subpopulations of veterans will allow for targeted prevention and policies to aid in their long-term health.
The first steps will demonstrate how to navigate to public access data through SAMHDA.
Select “Public-use Data Analysis System (PDAS)”. 
SURVEY

- The next steps will show how to select the survey of interest for this project.
Select “NSDUH”.

SPRINT
Survey:
1. Open the Substance Abuse and Mental Health Data Archive (SAMSHA) website.
2. Navigate to the “Recent Analyses” section.
3. Select the “SURVEYS” option.
4. From the list, choose “NSDUH” to proceed with the analysis.
Select “National Survey on Drug Use and Health, 2018”.
SELECT VARIABLES

• The following steps will demonstrate how to select each variable we are interested in based on the availability within the dataset and the project description.

• Slides 13-25 will repeat the steps from 10 and 11 using different categories and variables.

• One can also use the search bar above the favorites tab to find the variables of interest.
SELECT VARIABLES

To begin to gather the sample, select “Demographics”.

National Survey on Drug Use and Health, 2018

VARIABLES
- Recoded Consumption in Accrual
- Marijuana Purchases
- Interview Information
- Demographics
- Imputed Demographics
- Recoded Demographics
- Education
- Recoded Education
- Employment
- Imputed Employment
- Household Composition (Roster)
- Proxy Information
- Health Insurance
- Imputed Health Insurance
- Recoded Health Insurance
- Income

CROSSTAB CREATOR

Row Variable
- Row Variable is required

Column Variable
- Control Variable

Weight Variable
- ANALWT_C: Fin Prs-Level Simple Wgt

Run Crosstab

CROSSTAB RESULTS

VIEW OPTIONS
- Crosstab Table(s)
- Chi-Square Test
- Chart(s)

TABLE DISPLAY OPTIONS
- (show/hide)
  - Weighted Count
  - Unweighted Count
  - Total %
  - Row %
  - Column %
  - Confidence Intervals
  - Standard Errors

No results to display. To generate results, create and run a crosstab in the Crosstab Creator section above.
Click the star to add the variable to the “Favorites” tab. Include “ACTDEVER”.
INCLUDE VARIABLES

When scrolling to the top of the variables tab, you will see each of the starred demographic variables listed.
SELECT VARIABLES

National Survey on Drug Use and Health, 2018

CROSSTAB CREATOR

Generate a crosstab from data available in the survey National Survey on Drug Use and Health, 2018. Begin by selecting a variable from the variable list on the left. A row variable is required for analysis. Additionally, you can use the checkboxes in the Results section below to select which views you would like to generate and which table options you would like to display.

Weight Variable: ANALWT_C: Fin Prsn-Level Simple Wgt

Row Variable is required

Click on “Imputed Demographics”.

CROSSTAB RESULTS

No results to display. To generate results, create and run a crosstab in the Crosstabs Creator section above.
SELECT VARIABLES

Select “Imputation Revised Gender”.

National Survey on Drug Use and Health, 2018

CROSSTAB CREATOR

Generate a crosstab from data available in the survey National Survey on Drug Use and Health, 2018. Begin by selecting a variable from the variable list on the left. A row variable is required for analysis. Additionally, you can use the checkboxes in the Results section below to select which views you would like to generate and which table options you would like to display.

Row Variable

Column Variable

Select “Imputation Revised Gender”.

CROSSTAB RESULTS

VIEW OPTIONS

- Crosstab Table(s)
- Chi-Square Test
- Chart(s)

TABLE DISPLAY OPTIONS

- Weighted Count
- Unweighted Count
- Total %
- Row %
- Column %
- Standard Errors

No results to display. To generate results, create and run a crosstab in the Crosstab Creator section above.
SELECT VARIABLES

Click on “Recoded Demographics.”
SELECT VARIABLES

Select “NEWRACE2”.

Generate a crosstab from data available in the survey National Survey on Drug Use and Health, 2018. Begin by selecting a variable from the variable list on the left. A row variable is required for analysis. Additionally, you can use the checkboxes in the Results section below to select which views you would like to generate and which table options you would like to display.
Select Variables

Scroll and click on the “County” tab.
SELECT VARIABLES

Select "COUTYP4".
SELECT VARIABLES

Click “Recoded Drug Use”.
SELECT VARIABLES

Select “MRJYR”.
SELECT VARIABLES

Scroll down in the same category and select “OPINMYR”.
SELECT VARIABLES

Select "Recode Substance Dependence and Abuse".
SELECT VARIABLES

Scroll down and click “ABODALC”.
Recode Variables

- Data within the variables include a series of responses that are not a useful measure.
- For example, the answers “REFUSED” “SKIP” and “BAD DATA” are all responses in the dataset.
- The next steps demonstrate how to recode these variables to “MISSING”.
In the Favorites section, click the blue arrow-head on “ACTDEVER”.

RECODE VARIABLES
RECODE VARIABLES

Click “Recode”.
RECODE VARIABLES

Click “+” to create a new category.
RECODE VARIABLES

Rename category “Missing”.
RECODE VARIABLES

Click on categories to move them into the new “Missing” category. Click “BAD DATA”, “DON’T KNOW”, “REFUSED” “BLANK” AND “Legitimate skip”.
RECODE VARIABLES

Click save.
RECODE VARIABLES

Click "+" to add a new category.
RECODE VARIABLES

Rename the category “Yes”, click “Yes” from the left to add it to the recoded category, and click “Save”. 
RECODE VARIABLES

Click “+” to add another new category.
RECODE VARIABLES

Rename the category “No”, click “No” from the left to add it to the recoded category, and click “Save” for the category. Then click “Save” for the variable.
CHARACTERISTICS OF SAMPLE

• Running a crosstab will help us understand the characteristics of our sample.

• Crosstabs using demographic characteristics helps us define the sample (Slides 36-43).

• We will run crosstabs with the outcomes of interest and include a chi square analysis to determine whether there is a relationship between the demographic and outcome variables (Slides 43-45).
CHARACTERISTICS OF SAMPLE

In the Recoded Variables category, click the variable to reveal crosstab options. Click “Set Row”.

Weight Variable: ANALWT_C, Fin Pctn-Level Single Weight
CHARACTERISTICS OF SAMPLE

Under “NEWRACE2”, click “Set Column”.

National Survey on Drug Use and Health, 2018

CROSSTAB CREATOR

Generate a crosstab from data available in the survey National Survey on Drug Use and Health, 2018. Begin by selecting a variable from the variable list on the left. A new variable is required for analysis. Additionally, you can use the checkboxes in the Results section below to select which views you would like to generate and which table options you would like to display.

Ever On Active Duty In Us

Weight Variable ANALWT_Ci Fin Prm-Level Simple Wght

Run Crosstab

CROSSTAB RESULTS

VIEW OPTIONS

- Crosstab Table(s)
- Chart(s)

TABLE DISPLAY OPTIONS

- Weighted Count
- Unweighted Count
- Total %
- Row %
- Column %
- Confidence Intervals
- Standard Errors

EVER ON ACTIVE DUTY IN US MILITARY/RESERV

Total

Weighted Count
Count
Column %
Column % SE

Results
CHARACTERISTICS OF SAMPLE

Under “COUTYP4”, click “Set Control”.
The “Add Filter” button can be used when you only need to view one aspect of the variable. For example, adding a filter to the County/Metro control variable can allow you to view the active duty race demographics for just large metro areas.

View the variables in the Crosstab Creator. Click “Run Crosstab”.

CHARACTERISTICS OF SAMPLE
CHARACTERISTICS OF SAMPLE

Scroll down to view the results of the crosstab. Notice the first group is the “Large Metro” analysis. Below, you will see other levels of metro size, including graphs.
CHARACTERISTICS OF SAMPLE

To download the crosstab, scroll to Export, and click "full crosstab as csv".
Click the up arrow to open or save the crosstab.
Remove “NEWRACE2” from the crosstab.
CHARACTERISTICS OF SAMPLE

Include “MRJYR” as the Column Variable. Select “Run Chi-Square Test?” to see whether there is a relationship between active duty status and marijuana use, when stratifying by metro status. Then click “Run Crosstab” and export the data as before.
LOGISTIC REGRESSION

• We will now run a logistic regression to determine the effect of veteran status on past year substance use.
LOGISTIC REGRESSION

Click the Logistic Regression tab.
In the Logistic Regression Creator, notice the same variables remain in your favorites tab. Select “OPINMYR” and click “Set Dependent”.
Click Add Independent for “MRJYR”, “ABODALC”, “ACTDEVER”, “IRSEX”, “NEWRACE2”, “COUTYP4”. In the Crosstab, you could only add one variable for each aspect of the analysis. Here, you can have one Dependent and multiple Independents.

Also note: You cannot use the recoded variable in the logistic regression. You will get an error and the analysis will not run.
See the variables in the Logistic Regression Creator on the right and click “Run Logistic Regression”.
LOGISTIC REGRESSION

Scroll down to view the results of the logistic regression.
This is what the completed analysis looks like. These values here cannot be exported as they could in the Crosstab.
To analyze similar data over time, use a concatenated dataset from the homepage. Note that the survey may have changed from 2002 to 2018, so some of the variables in the single year analysis might not exist in the concatenated dataset.
If you have any questions related to the content of this presentation, or have noticed that this presentation needs to be updated due to changes in the format of the SAMHDA website, please contact Erin Johnson at erinl.johnson@unmc.edu.