Exercise 1

OBJECTIVE: Gain an understanding of how the NAPP dataset is structured and how it can be leveraged to explore your research interests. This exercise will use the NAPP dataset to explore a variety of household characteristics in early 19th century Norway.
Research Questions
What are the patterns of household characteristics in 19th century Norway?

Objectives
- Select datasets and variables of interest
- Analyze the data using sample code
- Validate data analysis work using answer key

IPUMS-USA Variables
- LABFORCE: Labor force participation
- DIFFEYE: Seeing difficulty
- KITCHEN: Whether or not the household has a kitchen
- MARST: Marriage status

SDA Code to Review

<table>
<thead>
<tr>
<th>Field</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row</td>
<td>Represents the primary variable of interest</td>
</tr>
<tr>
<td>Column</td>
<td>Divides the analysis of the variable of interest into categories</td>
</tr>
<tr>
<td>Control</td>
<td>Creates a separate chart for each category of the control</td>
</tr>
<tr>
<td>Selection Filter</td>
<td>Allows you to select cases; ex: year(2000-*) -&gt; all years 2000-onward</td>
</tr>
</tbody>
</table>

Review Answer Key (page 4)

Common Mistakes to Avoid

1 Choosing numerical instead of categorical variables for the Frequencies/Cross Tabulation Program. For these, use the Comparison of Means Program instead.
Getting Started

- Go to http://www.nappdata.org/napp/sda.shtml, and select the Norway 1801 100% sample
- Use your account and password to log in
- The default analysis is frequency/cross-tabulation

- Either browse variables under the Household and Person variables categories, or Search on the main IPUMS-USA site for variables
- When you browse for a variable, click on it, and it will appear in the Selected box. To send the variable to your input section, click the appropriate box (Row, etc)
- Row and Column are the variables of interest that you will perform the cross-tabulation on
- Filters select only specified cases
- A Control creates multiple tables for row and column variables, separated by a third categorical variable. For example, if you include the variable SEX as a control, you will get two frequency tables
- The Weight default is person weight (perwt), which extrapolates the sample to represent the entire population. However, the 1801 sample for Norway is a 100 percent sample, so using the weight is unnecessary.
Part I – Basic Frequencies

Section 1
Analyze the Data

A) What is the rate of labor force participation among men and women in Norway 1801? 

- Row: labforce
- Column: sex

B) What percent of single family households had at least one unmarried daughter? ____________ unmarried son?___________

- Row: unmardau unmarson
- Column: nfams

C) Go to the main website and find the code for the occupation Medical Doctor and Physician in the variable OCCHISCO.______

D) How many medical doctors were in Norway in 1801?__________

- Row: occhisco
- Filter: occhisco(06110)

Part II – Comparison of Means

E) What is the average size of the primary family? ________________

F) Which county has the highest average family size? ________________

- Dependent: prmfamsz
- Row: countyno
ANSWERS: Part I – Basic Frequencies

Section 1
Analyze the Data

A) What is the rate of labor force participation among men and women in Norway 1801? **Men: 45.4%**  **Women: 7.3%**

<table>
<thead>
<tr>
<th>Row: labforce</th>
<th>Column: sex</th>
</tr>
</thead>
</table>

B) What percent of single family households had at least one unmarried daughter? **27.1% (100-72.9)**  son? **27.3% (100-72.7)**

<table>
<thead>
<tr>
<th>Row: unmardau unmarson</th>
<th>Column: nfams</th>
</tr>
</thead>
</table>

C) Go to the main website and find the code for the occupation Medical Doctor and Physician in the variable OCCHISCO. **06110**

D) How many medical doctors were in Norway in 1801? **24**

| Row: occhisco | Filter: occhisco(06110) |

Part II – Comparison of Means

E) What is the average size of the primary family? **5.13**

F) Which county has the highest average family size? **Norde Berg**

| Dependent: prmfamsz | Row: countyno |