

# NAPP

# Extraction and

# Analysis

## 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 historical demographic characteristics of Iceland.

## Research Questions

What were the most common occupations in Iceland in 1801 and 1901? Were farm households more likely to have more generations living together? How frequent was immigration into Iceland between 1801 and 1901?

## Objectives

- Create and download a NAPP data extract
- Decompress data file and read data into SPSS
- Analyze the data using sample code
- Validate data analysis work using answer key

## NAPP Variables

- OCCHISCO: HISCO occupation classification
- FARMIPUM: Farm household by 19<sup>th</sup> century definition
- NUMGEN: The number of generations in the household
- NAPPSTER: NAPP country of birth
- YRIMMIG: Year of immigration to Iceland
- COUNTYIS: Iceland county

## SPSS Code to Review

Code	Purpose
compute	Creates a new variable
freq	Displays a simple tabulation and frequency of one variable
crosstabs	Displays a cross-tabulation for up to 2 variables and a control
~=	Not equal to

## Review Answer Key (page 7)

### Common Mistakes to Avoid

- 1 Excluding cases you don't mean to. Avoid this by turning off weights and select cases after use, otherwise they will apply to all subsequent analyses
- 2 Terminating commands prematurely or forgetting to end commands with a period (.) Avoid this by carefully noting the use of periods in this exercise

## Registering with NAPP

Go to <http://www.nappdata.org/napp/>, click on User Registration & Login, and apply for access. On login screen, enter email address and password and submit it!

### Step 1

#### Make an Extract

- Go back to homepage and go to Select Data
- Click the Select Samples box. Check the boxes for the 1801 and 1901 Iceland samples. Click the Submit sample selections box
- Using the drop down menu or search feature, select the following variables:

OCCHISCO: HISCO occupation classification

FARMIPUM: Farm household by 19<sup>th</sup> century definition

NUMGEN: The number of generations in the household

NAPPSTER: NAPP country of birth

YRIMMIG: Year of immigration to Iceland

COUNTYIS: Iceland county

...

### Step 2

#### Request the Data

- Click the green VIEW CART button under your data cart
- Review variable selection. Click the green Create Data Extract button
- Review the 'Extract Request Summary' screen, describe your extract and click Submit Extract
- You will get an email when the data is available to download.
- To get to the page to download the data, follow the link in the email, or follow the Download and Revise Extracts link on the homepage.

## Getting the data into your statistics software

The following instructions are for SPSS.

### Step 1

#### Download the Data

...

### Step 2

#### Decompress the Data

...

### Step 3

#### Read in the Data

Go to <http://www.nappdata.org/napp/> and click on Download or Revise Extracts

- Right-click on the data link next to extract you created
- Choose "Save Target As..." (or "Save Link As...")
- Save into "Documents" (that should pop up as the default location)
- Do the same thing for the SPSS link next to the extract

- Find the "Documents" folder under the Start menu
- Right click on the ".dat.gz" file
- Use your decompression software to extract here
- Double-check that the Documents folder contains three files starting "napp\_000..."
- Free decompression software is available at <http://www.irnis.net/soft/wingzip/>

- Double click on the ".sps" file, which should automatically have been named "napp\_000...."

- The first two lines should read:

```
cd ".".
```

```
data list file = 'napp_000.../'
```

- Change the first line to read: cd (location where you've been saving your files). For example:

```
cd "C:\Documents".
```

- Change the second line to read:

```
data list file = "C:\Documents\napp_000...dat"/
```

Under the "Run" menu, select "All" and an output viewer window will open

## Analyze the Sample – Part I Frequencies of OCCHISCO

### Section 1

### Analyze the Data

...

### Note on Weights

A) On the website, find the codes page for the OCCHISCO variable. Go to the Comparability tab and find how individuals were coded who were too young to work in Iceland 1801 and 1901.

---

B) What were the 3 most common occupations in Iceland in 1801? \_\_\_\_\_

C) What about 1901? \_\_\_\_\_

```
frequencies variables = occhisco  
/format = dfreq  
/order = analysis.
```

*Note: The "format = dfreq" option orders the results by descending frequency.*

### Using weights (PERWT)

In other data projects, you might be familiar with using weights to make your analysis more representative of the entire population. However, because the Iceland samples are already a 100 percent sample, the weight for each person is always one. If you compare multiple countries, however, you'll need to use the PERWT weight. To learn more about using weights, see the NAPP data exercise 2.

## Analyze the Sample – Part II Relationships in the Data

### Section 1

### Analyze the Data

A) Go to the codes page for the variable FARMIPUM. What are the codes for this variable? \_\_\_\_\_

B) Which two counties in Iceland have the lowest proportion of farm households in 1901? \_\_\_\_\_

```
crosstabs
/tables = countyis by farmipum by year
/cells = count.
```

C) What is the average number of generations in an Icelandic household in 1901? \_\_\_\_\_

```
means tables = numgen by year
/cells mean count stddev.
```

Now we'll generate a new variable to recode FARMIPUM into a binary variable. It will have a value of 0 for non-farm, and a value of 1 for a farm household. Then we'll use two different ways of testing whether farm households tend to have more generations.

D) Is the mean of NUMGEN different between farms and non-farms in 1901? \_\_\_\_\_

```
compute farm = 0.
if (farmipum=2) farm = 1.
means tables = numgen by farm by year
/cells mean count.
```

E) Does being a farm household make a family more likely to live with more generations in 1901? Is this significantly significant?

```
regression
/select = year eq 1901
/dependent numgen
/method = enter farm.
```

## Analyze the Sample – Part III Frequencies in the Data

### Section 1

#### Analyze the Data

...

Complete!  
Check  
your  
Answers!

A) Go to the Universe tab for YRIMMIG. What is the universe for YRIMMIG in Iceland 1901? \_\_\_\_\_

B) What are the codes for "Unknown" and "Not in Universe"? To whom does "Not in Universe" apply?  
\_\_\_\_\_

C) How many people were immigrants from Norway and Denmark living in Iceland in 1901?

```
freq nappster.
```

*Note: You'll have to select only cases from 1901. Go to the Data menu, choose "Select Cases", then "If...", and enter "year = 1901". Click Continue, then Ok.*

D) What years did the majority of these immigrants move to Iceland? \_\_\_\_\_

```
crosstabs  
/tables = yrimmig by nappster  
/cells = count column.
```

*Note: Return to Select Cases and enter "(nappster = 4 or nappster = 7)".*

## ANSWERS: Analyze the Sample – Part I Frequencies of OCCHISCO

### Section 1

### Analyze the Data

...

### Note on Weights

A) On the website, find the codes page for the OCCHISCO variable. Go to the Comparability tab and find how individuals were coded who were too young to work in Iceland 1801 and 1901.

**Unknown/No occupation**

B) What were the 3 most common occupations in Iceland in 1801?

**General Farmers, Servants nfs, Farmer and Fisherman**

C) What about 1901? **Servants nfs, Farm workers, Fishermen**

```
frequencies variables = occhisco  
/format = dfreq  
/order = analysis.
```

*Note: The "format = dfreq" option orders the results by descending frequency.*

### Using weights (PERWT)

In other data projects, you might be familiar with using weights to make your analysis more representative of the entire population. However, because the Iceland samples are already a 100 percent sample, the weight for each person is always one. If you compare multiple countries, however, you'll need to use the PERWT weight. To learn more about using weights, see the NAPP data exercise 2.



## ANSWERS: Analyze the Sample – Part II Relationships in the Data

### Section 1

#### Analyze the Data

A) Go to the codes page for the variable FARMIPUM. What are the codes for this variable? **Non-farm: 1; Farm: 2**

B) Which two counties in Iceland have the lowest proportion of farm households in 1901? **Gullbringusýsla 20.42%, Reykjavíkurborg 3.65%**

```
crosstabs
/tables = countyis by farmipum by year
/cells = count.
```

C) What is the average number of generations in an Icelandic household in 1901? **2.06 generations**

```
means tables = numgen by year
/cells mean count stddev.
```

D) Is the mean of NUMGEN different between farms and non-farms in 1901? **Yes, the difference is 0.149 generations.**

```
compute farm = 0.
if (farmipum=2) farm = 1.
means tables = numgen by farm by year
/cells mean count.
```

E) Does being a farm household make a family more likely to live with more generations in 1901? Is this significantly significant?

**According to the regression, the difference in means 0.149 generations, and is significant at the 0.001 level.**

```
regression
/select = year eq 1901
/dependent numgen
/method = enter farm.
```

## ANSWERS: Analyze the Sample – Part III Frequencies in the Data

### Section 1

#### Analyze the Data

- A) Go to the Universe tab for YRIMMIG. What is the universe for YRIMMIG in Iceland 1901? **All foreign-born persons.**
- B) What are the codes for "Unknown" and "Not in Universe"? To whom does "Not in Universe" apply? **Unknown: 0000; NIU: 9999. NIU applies to anyone born in Iceland, or not-foreign born.**
- C) How many people were immigrants from Norway and Denmark living in Iceland in 1901? **207 from Norway, 110 from Denmark**

```
freq nappster.
```

*Note: You'll have to select only cases from 1901. Go to the Data menu, choose "Select Cases", then "If...", and enter "year = 1901". Click Continue, then Ok.*

- D) What years did the majority of these immigrants move to Iceland? **1901 for Norway (95 people); 1892/1901 for Denmark (5)**

```
crosstabs  
/tables = yrimmig by nappster  
/cells = count column.
```

*Note: Return to Select Cases and enter "(nappster = 4 or nappster = 7)".*