Minnesota Population Center
Training and Development

NAPP Extraction and Analysis

Exercise 2

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 shifts in Great Britain and Canada.

1/8/2013

NAPP Training and Development

Research Questions

How did religious composition change over time in Canada? What were the demographic characteristics of migrants in Canada in the 19th century? Which households in Great Britain were more likely to have servants?

Objectives

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

NAPP Variables

- RELIGION: First stated religion
- YEAR: Year of census sample
- MIGRANT: Migration status
- BPLCNTRY: Country of birth
- AGE: Age
- SEX: Sex
- SERVANTS: Number of servants in the household
- URBAN: Urban/rural status
- MARST: Marital status

Stata Code to Review

Code	Purpose
mean	Displays a simple tabulation and frequency of one variable
tabulate	Displays a cross-tabulation for up to 2 variables
!=	Not equal to

Review Answer Key (page 7)

Common Mistakes to Avoid

1 Not changing the working directory to the folder where your data is stored

2 Mixing up = and = = ; To assign a value in generating a variable, use "=". Use "= =" to specify a case when a variable is a desired value using an *if* statement.

3 Forgetting to put [weight=*weightvar*] into square brackets

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 Canadian historical samples from 1871, 1881, 1891, and 1901. Click the Submit sample selections box
- Using the drop down menu or search feature, select the following variables:

RELIGION: First stated religion
YEAR: Year of census sample
MIGRANT: Migration status
BPLCNTRY: Country of birth
AGE: Age
SEX: Sex
SERVANTS: Number of servants in the household
URBAN: Urban/rural status
MARST: Marital status

• • •

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 Stata.

Step 1

Download 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 Stata link next to the extract

•••

Step 2

Decompress the Data

- 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/

•••

Step 3

Read in the Data

- Open Stata from the Start menu
- In "File" menu, choose "Change working directory..."

  Select "Documents", click "OK"
- In "File" menu, choose "Do..."

  Select the *.do file
- You will see "end of do-file" when Stata has finished reading in the data.

Analyze the Sample – Part I Frequencies of RELIGION

Section 1

Analyze the Data

A) On the website, find the codes page for the SAMPLE and RELIGION variables. Find the codes for each Canadian sample and for Roman Catholics in RELIGION. Write them down.

_____

B) Is RELIGION available for every Canadian historical sample? What about Great Britain? _____

C) What was the first year that an individual gave Buddhism as a response? _____

```
tab religion year
```

• • •

D) What is the trend in the population of Roman Catholics in Canada over time in the census samples? Is this a realistic result? _____

```
histogram sample if religion == 1100
```

Note on Weights

Using weights (PERWT)

Because the 1881 Canada sample is the only 100% sample for Canada, the population of Roman Catholics in 1881 appears to skyrocket, and then decrease again in the 1891 sample. In order to find a representative population from the 5 or 9 percent samples from other years, we will need to use a weight.

E) Using weights, what percentage of the population were Roman Catholics for each sample? _____

```
histogram sample if religion == 1100 [fweight = perwt], discrete percent addlabel
```

Analyze the Sample – Part II Relationships in the Data

Section 1

Analyze the Data

A) Go to the codes page for the variable MIGRANT. What is the code for "International Migrant from one NAPP country to another"? _____

B) What is the male to female ratio of migrants from the United Kingdom to Canada in the 19th century? _____

```
tab sample sex if migrant ==3 & bplcntry == 42120 [fweight=perwt]
```

Hint: Find the weighted populations of men and women in each Canadian census whose birthplace is the United Kingdom and MIGRANT code is 3. Divide the number of men by the number of women for each sample year.

C) What is the male to female ratio in Great Britain in the 19th century samples who are not migrants?

_____

```
tab sample sex if (migrant ==1|migrant==2) & bplcntry == 42120 [fweight=perwt]
```

D) Now compare the ratios of the Canadian sample in 1881 and the Great Britain sample in 1881. What hypothesis could you draw from the differences you see? _____

E) If we thought that marital status might be significantly different across migrant status, and this could have something to do with our results above, we can test out the hypothesis. Across all samples, are migrants more likely to be married, spouse absent or never married?_____

```
tab marst migrant, column
```

F) Check the universe for MARST on the website. Does this mean we will have to exclude people under 18 to get a more realistic estimate of Never married/Single? Does excluding children change the table? _____

```
tab marst migrant if age >=18, column
```

Analyze the Sample – Part II Relationships in the Data

G) What is the mean age of individuals in Canada in 1881 by migrant status? _____

```
mean age if sample ==1243 & age<200, over(migrant)
```

Note: The missing code for age is 999, so we need to exclude missing values to prevent a biased estimate.

Part III Relationships in the Data

A) Using a graph, show if there is a difference in the average number of servants by urban/rural status in Great Britain in 1851.

Section 1

Graph the Data

```
graph bar (mean)  servants [weight = hhwt] if sample ==
        8261 & pernum==1, over(urban)
```

Note: Because SERVANTS is a household level variable, you will need to select only one person to represent each household and weight by HHWT.

B) Does this relationship change if you panel this by country of Great Britain? _____

```
graph bar (mean)  servants [weight = hhwt] if sample ==
        8261 & pernum==1, over(urban) by(cntrygb)
```

• • •

Complete! Check your Answers!

ANSWERS: Analyze the Sample – Part I Frequencies of RELIGION

Section 1

Analyze the Data

A) On the website, find the codes page for the SAMPLE and RELIGION variables. Find the codes for each Canadian sample and for Roman Catholics in RELIGION. Write them down. **1241: Canada 1852; 1242: Canada 1871; 1243: Canada 1881; 1244: Canada 1891; 1245: Canada 1901. Roman Catholic: 1100**

B) Is RELIGION available for every Canadian historical sample? What about Great Britain? **RELIGION was asked for every Canadian sample, but it is not available for the Great Britain samples.**

```
tab religion year
```

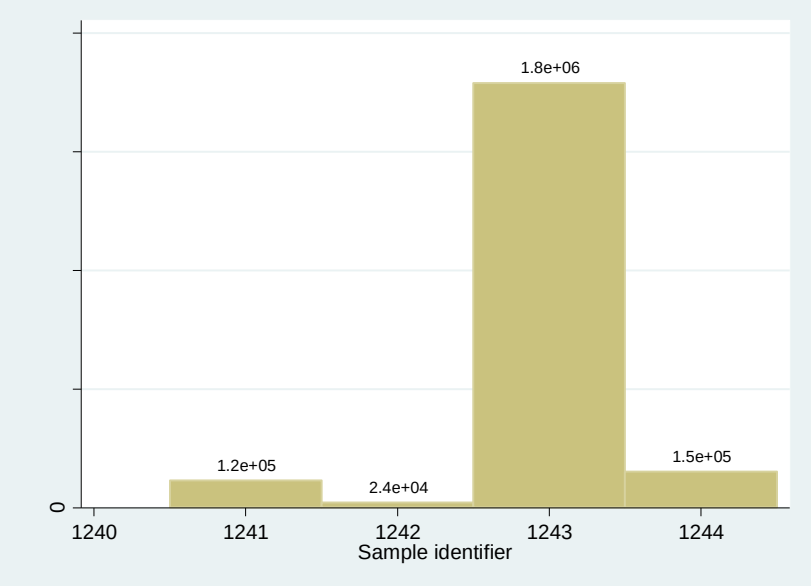
C) What was the first year that an individual gave Buddhism as a response? **1881**

```
histogram sample if religion ==1100, discrete frequency addlabel
```

• • •

Note on Weights

D) What is the trend in the population of Roman Catholics in Canada over time in the census samples? Is this a realistic result? **The population jumps from the tens of thousands to more than a million in 1881. This is unrealistic because 1881 is simply a 100 percent sample, whereas the other samples are no more than 9 percent.**



Section 1

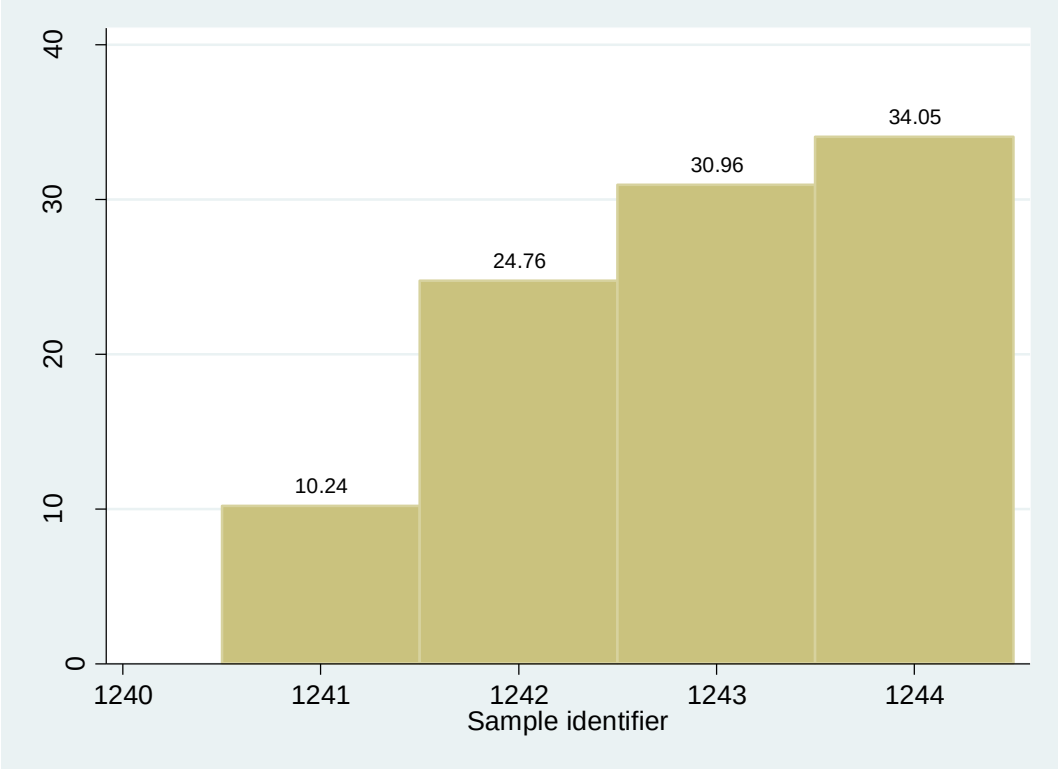
Analyze the Data

Using weights (PERWT)

Because the 1881 Canada sample is the only 100% sample for Canada, the population of Roman Catholics in 1881 appears to skyrocket, and then decrease again in the 1891 sample. In order to find a representative population from the 5 or 9 percent samples from other years, we will need to use a weight.

E) Using weights, what percentage of the population were Roman Catholics for each sample? **Canada 1852: 10.24; Canada 1871: 24.76; Canada 1881:30.96; Canada 1891: 34.05;**

```
histogram sample if religion == 1100 [fweight = perwt], discrete percent addlabel
```



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

Section 1

Analyze the Data

A) Go to the codes page for the variable MIGRANT. What is the code for "International Migrant from one NAPP country to another"? **MIGRANT = 3**

B) What is the male to female ratio of migrants from Great Britain to Canada in the 19th century? **1871:** 1.106; **1881:** 1.261; **1891:** 1.359

```
tab sample sex if migrant ==3 & bplcntry == 42120 [fweight=perwt]
```

Hint: Find the weighted populations of men and women in each Canadian census whose birthplace is the United Kingdom and MIGRANT code is 3. Divide the number of men by the number of women for each sample year.

C) What is the male to female ratio in Great Britain in the 19th century samples who are not migrants? **1851:** 0.949 **1881:** 0.946

```
tab sample sex if (migrant ==1|migrant==2) & bplcntry == 42120 [fweight=perwt]
```

D) Now compare the ratios of the Canadian sample in 1881 and the Great Britain sample in 1881. What hypothesis could you draw from the differences you see? **Over time, the ratio of men to women in Canada was increasing. In Great Britain, women outnumbered men while in Canada, men outnumbered women. The hypothesis that could be drawn is that migration was mostly male-dominated, and men were either unmarried or did not bring their spouses with them.**

E) If we thought that marital status might be significantly different across migrant status, and this could have something to do with our results above, we can test out the hypothesis. Across all samples, are migrants more likely to be married, spouse absent or never married? **Actually, migrants were more likely to be married, spouse present than non-migrants, which disproves our previous hypothesis.**

```
tab marst migrant, column
```

Section 1

Analyze the Data

F) Check the universe for MARST on the website. Does this mean we will have to exclude people under 18 to get a more realistic estimate of Never married/Single? Does excluding children change the table? **Yes, now there are fewer never married/single, and at least half the population in each migrant group is recorded as married.**

```
tab marst migrant if age >=18, column
```

G) What is the mean age of individuals in Canada in 1881 by migrant status? Does this make sense?

```
mean age if sample ==1243 & age<200, over(migrant)
```

Note: The missing code for age is 999, so we need to exclude missing values to prevent a biased estimate.

Resident in state of birth: 20.7

Resident in country of birth: 28.1

International migrant from NAPP country to another: 39.2

International migrant from a non-NAPP country: 46.1

Unclassifiable: 23.5

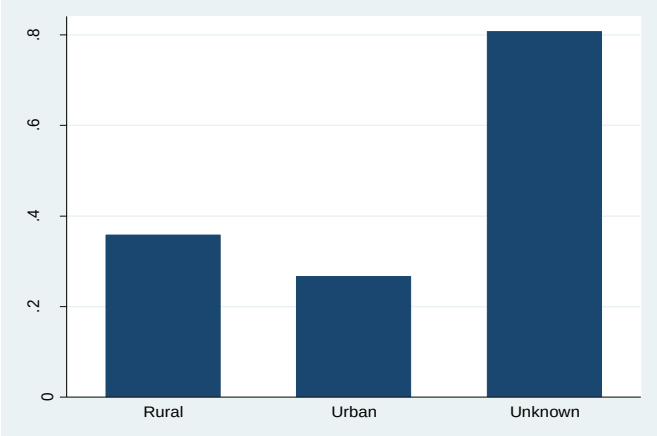
This makes sense because children will lower the average age, and children are more likely to be living in their state or at least country of birth. Also, migrants may be less likely to have young children with them if they are on the move for a period of time.

Section 1

Graph the Data

A) Using a graph, show if there is a difference in the average number of servants by urban status in Great Britain in 1851.

```
graph bar (mean)  servants [weight = hhwt] if sample ==
        8261 & pernum==1, over(urban)
```



B) Does this relationship change if you panel this by country of Great Britain? **No, it appears that the average number of servants is higher in rural areas, perhaps because estates are larger and cover more area.**

```
graph bar (mean)  servants [weight = hhwt] if sample ==
        8261 & pernum==1, over(urban) by(cntrygb)
```