

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.

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 SPSS
- 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
- BPLCOUNTRY: Country of birth
- AGE: Age
- SEX: Sex
- SERVANTS: Number of servants in the household
- URBAN: Urban/rural status
- MARST: Marital status

SPSS Code to Review

Code	Purpose
means tables	Displays the mean of a variable over a categorical variable
crosstabs	Displays a cross-tabulation for up to 2 variables and a control
~=	Not equal to

Review Answer Key (page 8)

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

...

Step 2

Request the Data

- 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

- 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 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? _____

crosstabs
/tables = religion by year
/cells = count.

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

graph
/histogram = sample.

Note: You'll need to enter religion = 1100 in the Select Cases feature.

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? _____

weight by perwt.
graph
/bar (grouped) = pct by sample.

...

Note on Weights

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? _____

```
weight by perwt.  
crosstabs  
/tables = sex by sample  
/cells = count.
```

Hint: You'll need to use the Select Cases feature to filter for migrant = 3 & bplcntry = 42120. Find the weighted populations of men and women in each Canadian census whose birthplace is the United Kingdom and MIGRANT = 3. Divide the number of men by the number of women for each sample.

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

Note: Select Cases are (migrant = 1 or migrant = 2) and bplcntry = 42120

```
crosstabs  
/tables = sex by sample  
/cells = count.
```

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? _____

```
crosstabs  
/tables = marst by migrant  
/cells = count.
```

Analyze the Sample – Part II Relationships in the Data

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? _____

```
crosstabs
/tables = marst by migrant
/cells = count column.
```

Note: Select Cases should be age >=18

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

```
means tables = age by migrant
/cells mean count stddev.
```

Note: The missing code for age is 999, so we need to exclude missing values to prevent a biased estimate. Select Cases should be: sample =1243 & age<200

Analyze the Sample - Part III Relationships in the Data

Section 1

Graph 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.

```
graph
```

```
/bar (grouped) = mean(servants) by 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. Select cases should be sample = 8261 and pernum = 1, and use the same for the next question.

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

```
graph
```

```
/bar (grouped) = mean(servants) by urban by centrygb.
```

...

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.**

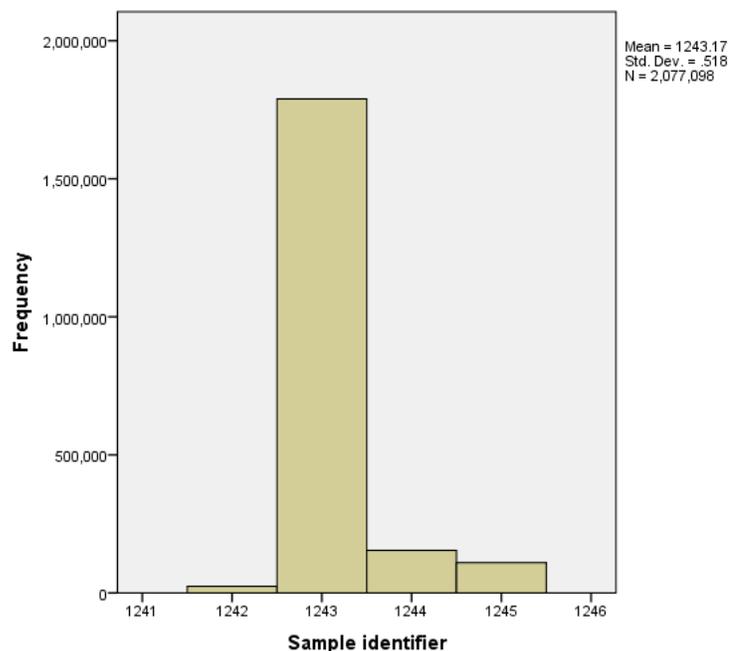
```
crosstabs
/tables = religion by year
/cells = count.
```

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

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.**

```
graph
/histogram = sample.
```

Note: You'll need to enter religion = 1100 in the Select Cases feature.



ANSWERS - Analyze the Sample - Using weights (PERWT)

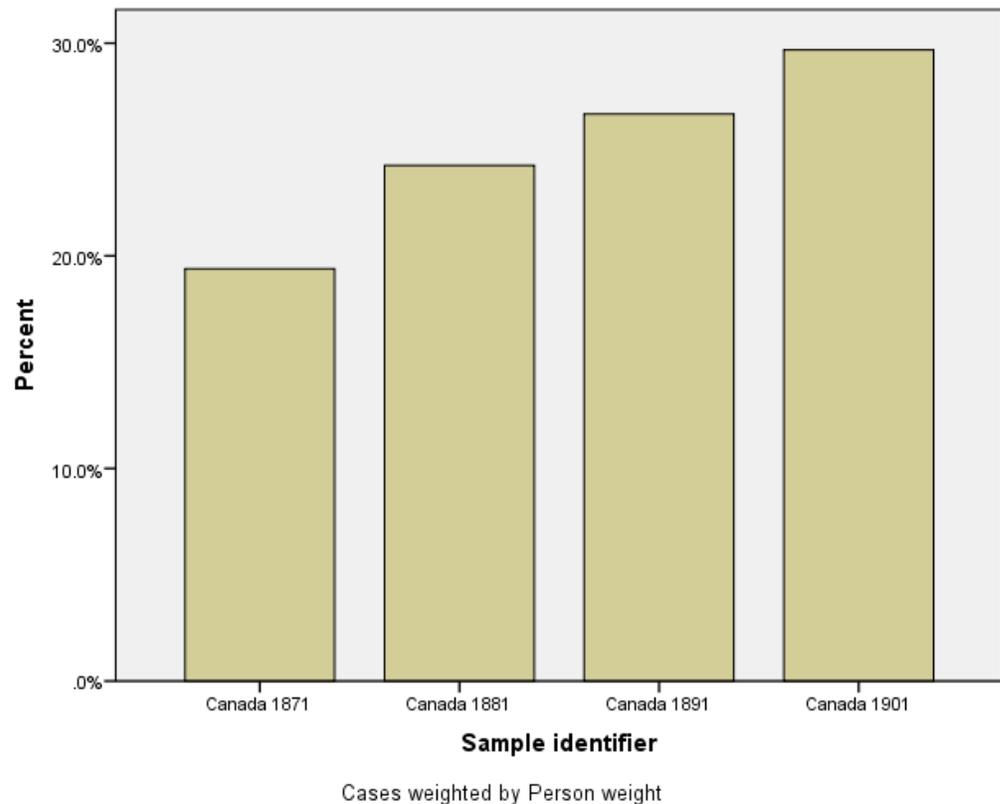
Section 1

Note on Weights

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;**

weight by perwt.
graph
/bar (grouped) = pct by sample.



Analyze the Data

ANSWERS – Analyze the Data – 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 the United Kingdom to Canada in the 19th century? **1871: 1.106; 1881: 1.261; 1891: 1.359**

```
weight by perwt.  
crosstabs  
/tables = sex by sample  
/cells = count.
```

Hint: You'll need to use the Select Cases feature to filter for migrant = 3 & bplcntry = 42120. Find the weighted populations of men and women in each Canadian census whose birthplace is the United Kingdom and MIGRANT = 3. Divide the number of men by the number of women for each sample.

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**

```
crosstabs  
/tables = sex by sample  
/cells = count.
```

Note: Select Cases are (migrant =1 or migrant=2) and bplcntry = 42120

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.**

```
crosstabs  
/tables = marst by migrant  
/cells = count.
```

Analyze the Data

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

Section 1

Analyze the Data

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.**

```
crosstabs
/tables = marst by migrant
/cells = count 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? **Yes, now there are fewer never married/single, and at least half the population in each migrant group is recorded as married.**

Note: In Select Cases, add age >=18

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

```
means tables = age by migrant
/cells mean count stddev.
```

Note: The missing code for age is 999, so we need to exclude missing values to prevent a biased estimate. Select Cases should be: sample = 1243 & age < 200

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.

ANSWERS - Analyze the Sample- Part III Relationships in the Data

Section 1

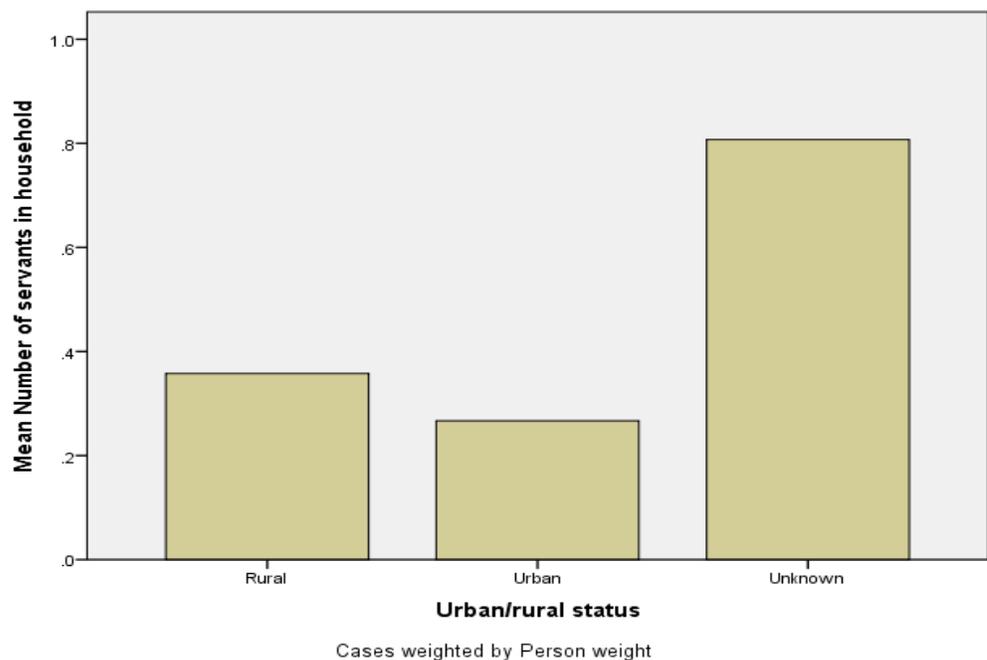
Graph the Data

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.

graph

/bar (grouped) = mean(servants) by 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. Select cases should be sample = 8261 and pernum = 1, and use the same for the next question.



ANSWERS - Analyze the Sample- Part III Relationships in the Data

Section 1

Graph the Data

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 (grouped) = mean(servants) by urban by cntrygb.

