IPUMS Data Training Exercise

An Introduction to IPUMS Time Use

(Exercise 2 for R)

Learning goals

- Create and download a ATUS data extract
- Decompress data file and read data into R
- Analyze the data using sample code

Summary

In this exercise, you will gain an understanding of how the Time Use dataset is structured and how it can be leveraged to explore your research interests. This exercise will use the Time Use dataset to explore basic frequencies of income, purchases, caring for children and time spent preparing food. You will create data extracts that include the variables ACT_PURCH, FAMINCOME, FOODPREP, and CHILDCARE; then, you will use a sample code to analyze these data.

R Code to Review

This tutorial's sample code and answers use the so-called "tidyverse" style, but R has the blessing (and curse) that there are many different ways to do almost everything. If you prefer another programming style, please feel free to use it. But, for your reference, these are some quick explanations for commands that this tutorial will use:

Code	Purpose
%>%	The pipe operator helps make code with nested function calls easier to read. When reading code, it can be read as "and then". The pipe makes it so that code like "ingredients %>% stir() %>% cook()" is equivalent to cook(stir(ingredients)) (read as "take *ingredients* and then *stir* and then *cook*").
as_factor	Converts the value labels provided for IPUMS data into a factor variable for R
summarize	Summarize a dataset's observations to one or more groups
group_by	Set the groups for the summarize function to group by
filter	Filter the dataset so that it only contains these values
mutate	Add on a new variable to a dataset
weighted.mean	Get the weighted mean of the variable

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 "<-" (or "="). Use "==" to test for equality.

   Note: In this exercise, for simplicity we will use "weighted.mean". For analysis where variance estimates are needed, use the survey or srvyr package instead.

Registering with ATUS

Go to http://www.atusdata.org/, click on Login at the top, and apply for access. On login screen, enter email address and password and submit it!

Make an Extract

- On the homepage, go to Build an Extract (on the left column)
- Click on the "Change Samples" box, and select years 2009 and 2011. Keep the defaults "ATUS respondents" and select "Submit sample selections".
- Under the "Time Use" dropdown menu, select "Activity coding structure". Click on the plus sign next to the variable ACT_PURCH to select the variable and add it to our data cart.
- Click on the "Create time use variable" box at the top. Select "Load" next to ACT_HHACT, then the diamond sign next to Household Activities to expand the category.
- Unselect all subcategories except for "Food and Drink Preparation, Presentation, and Clean-up", and click "Save time use variable" at the bottom.
- Name your new variable "foodprep" and select "Save time use variable". This selects the time use variable we just created, and adds it to our data cart.

Request the Data

- Click on the "Create time use variable" box again, and this time select the box at the top "Create variable from scratch".
- Select the box next to All, then click on the "Secondary Activity" box at the top.
- Under Secondary Activity, select "Duration of time spent during activity on secondary child care of all children" and then "Save time use variable".
- Name this new variable "childcare" and label it "Secondary childcare". Then select "Save time use variable" again to select it and add it to the cart.
- Under the Household dropdown menu, click on Geographic and then select the variable REGION.
- Using the drop down menu or search feature, select the following variables and add them to your data cart using the plus symbol to the left of the variables:
  - ACT_PURCH: Consumer purchases
  - FAMINCOME: Yearly family income
  - AGE: Age
  - SEX: Sex

- o FOODPREP: Created variable for time spent preparing food
- o CHILDCARE: Created variable for time with childcare as a secondary activity

Review and submit your extract

- Choose the green "View Cart" at the top.
- Click on the green "Create data 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/Revise Extracts link on the homepage.

Getting the data into your statistics software

The following instructions are for R.

Download the Data

- Go to http://www.atusdata.org/ and click on Download/ Revise Extracts
- Right-click on the "data" link next to extract you created, under "files"
- 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 DDI link next to the extract
- (Optional) Do the same thing for the R script
- You do not need to decompress the data to use it in R

Install the ipumsr package

- Open R from the Start menu
- If you haven't already installed the ipumsr package, in the command prompt, type the following command:

```
install.packages("ipumsr")
```

Read the Data

- Set your working directory to where you saved the data above by adapting the following command (Rstudio users can also use the "Project" feature to set the

working directory. In the menu bar, select File -> New Project -> Existing Directory and then navigate to the folder):

```
setwd("~/") # "~/" goes to your Documents directory on most
computers
```

- Run the following command from the console, adapting it so it refers to the extract you just created (note the number may not be the same depending on how many extracts you've already made):

```
library(ipumsr)

ddi <- read_ipums_ddi("atus_000##.xml")

data <- read_ipums_micro(ddi)

# Or, if you downloaded the R script, the following is equivalent:

#    source("atus_00001.R")
```

- This tutorial will also rely on the dplyr, tidyr and ggplot2 packages, so if you want to run the same code, run the following command (but if you know other ways better, feel free to use them):

```
library(dplyr)

library(ggplot2)

library(tidyr)
```

- To stay consistent with the exercises for other statistical packages, this exercise does not spend much time on the helpers to allow for translation of the way IPUMS uses labelled values to the way base R does. You can learn more about these in the value-labes vignette in the R package. From R run command: vignette("value- labels", package = "ipumsr").

Analyze the Sample

Part I: Generate a New Variable

1. Create a variable that distinguishes individuals who reported consumer purchases on the day of their interview.

```
data <- data %>%

mutate(PURCHASE = ACT_PURCH > 0)
```

2. Find a frequency for reported consumer purchases for the sample for each year? _____

```
data %>%

group_by(YEAR) %>%

summarize(PURCHASE = mean(PURCHASE))
```

3. Is there a difference in incidence of consumer purchasing between men and women in 2009? _____

_____

```
data %>%

group_by(YEAR, SEX = as_factor(SEX)) %>%

summarize(PURCHASE = mean(PURCHASE))
```

4. In the sample, when consumer purchases are greater than zero, what is the average amount of time spent on purchases each year? Does it appear the recession had any effect? _____

```
data %>%

group_by(YEAR) %>%

summarize(ACT_PURCH = mean(ACT_PURCH))
```

Part II: Using Weights (WT06)

Note on Weights

The ATUS sample design requires use of weights to provide and accurate representation at the national level. Half of the interview days in the sample are weekdays, while the other half are weekends. The weight WT06 adjusts for the disproportional number of weekend days, and should be used to weight time use variables. More specifically, WT06 gives the number of person-days in the calendar quarter represented by each survey response. Also keep in mind that the "Eating and Health," "Well-Being," and "Employee Leave" Modules have weights unique to them.

5. Using weights, what is the average time spent on purchases? _____
   _____

```
data %>%

    group_by(YEAR) %>%

    summarize(ACT_PURCH = weighted.mean(ACT_PURCH, WT06))
```

Part III: Relationships in the Data

Go to the ATUS homepage and choose Demographic Variables.

6. What is the range of values for this variable? What values indicate family incomes of $35,000 and higher? _____

7. What is the average time spent in food preparation across income groups?

Is there a trend? _____

_____

_____

```
data %>%

group_by(FAMINCOME = as_factor(FAMINCOME, level = "both")) %>%
summarize(foodprep = weighted.mean(foodprep, WT06))
```

8. Does the pattern change when you separate the analysis by year? _____

_____

_____

```
data %>%

group_by(YEAR, FAMINCOME = as_factor(FAMINCOME, level = "both"))
%>%

summarize(foodprep = weighted.mean(foodprep, WT06)) %>%

spread(YEAR, foodprep)
```

9. What could be an explanation for the result in parts 7 and 8? _____

_____

_____

_____

IV. Graph the Data

   10. Graph the results from 8.

```
data_summary <- data %>%

group_by(YEAR, FAMINCOME = as_factor(FAMINCOME)) %>%

summarize(foodprep = weighted.mean(foodprep, WT06))

ggplot(data_summary, aes(x = FAMINCOME, y = foodprep, fill =
factor(YEAR))) +

geom_col(position = "dodge") +

scale_fill_manual(values = c("#7570b3", "#e6ab02")) +

theme(

axis.text.x = element_text(angle = 20, hjust = 1), legend.position
= "bottom"

)
```

V. Relationships in Data Part II

   11. The way the variable CHILDCARE is constructed, what activities will it
       include? _____

       _____

   12. What are the codes for REGION? Find it under Demographic Variables.

       _____

       _____

13. What is the average amount of time for adults to be taking care of children as a secondary activity? _____

```
data %>%

filter(AGE >= 18) %>%

summarize(childcare = weighted.mean(childcare, WT06))
```

14. Are there differences in means across regions in 2011 in time spent in secondary child care? What about between men and women?_____

_____

_____

```
data %>%

filter(AGE >= 18) %>%

group_by(REGION = as_factor(REGION)) %>%

summarize(childcare = weighted.mean(childcare, WT06))


data %>%

filter(AGE >= 18) %>%

group_by(SEX = as_factor(SEX)) %>%

summarize(childcare = weighted.mean(childcare, WT06))
```

Answers:

Part I: Generate a New Variable

1. Create a variable that distinguishes individuals who reported consumer purchases on the day of their interview.

2. Find a frequency for reported consumer purchases for the sample for each year? <u>2009: 41.58%; 2011: 40.82</u>

3. Is there a difference in incidence of consumer purchasing between men and women in 2009? <u>Women: 44.97; Men: 37.08</u>

4. In the sample, when consumer purchases are greater than zero, what is the average amount of time spent on purchases each year? Does it appear the recession had any effect? <u>2009: 25 minutes; 2011: 24.7 minutes. There appears to be no significant difference between the two years.</u>

Part II: Using Weights (WT06)

Note on Weights

5. Using weights, what is the average time spent on purchases? <u>2009: 22.7 minutes; 2011: 22.2 minutes.</u>
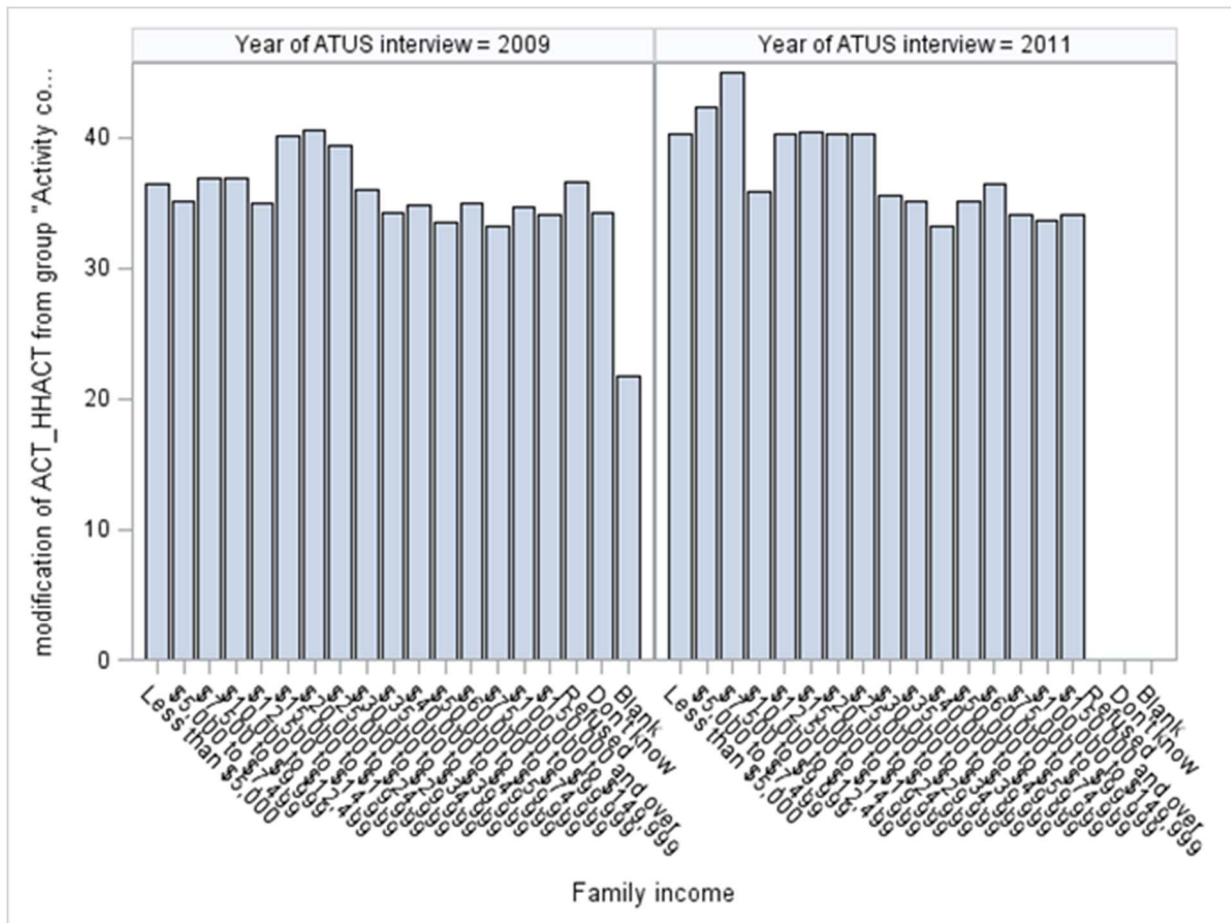
Part III: Relationships in the Data

6. What is the range of values for this variable? What values indicate family incomes of $35,000 and higher? <u>Codes 10 through 16.</u>

7. What is the average time spent in food preparation across income groups? Is there a trend? <u>There appears to be a small peak in income groups 5 through 7, then a slight decline.</u>

8. Does the pattern change when you separate the analysis by year? Not significantly.

9. What could be an explanation for the result in parts 7 and 8? The lowest income group may have slightly lower food prep time because they may work multiple jobs or be single parents with not enough time to dedicate to food preparation, while on the other hand, high paying jobs such as lawyers may have a high opportunity cost of time and also work long hours.

IV. Graph the Data

10. Graph the results from 8.

V. Relationships in Data Part II

11. The way the variable CHILDCARE is constructed, what activities will it include? <u>CHILDCARE should include time in any activity in which the respondent also reported child care at the same time.</u>

12. What are the codes for REGION? Find it under Demographic Variables. <u>1: Northeast; 2: Midwest; 3: South; 4: West</u>

13. What is the average amount of time for adults to be taking care of children as a secondary activity? <u>101.24 minutes a day.</u>

14. Are there differences in means across regions in 2011 in time spent in secondary child care? What about between men and women? <u>The Northeast has the lowest average, while the South has the highest. Women are more likely to be incorporating childcare into other activities (74.9 for men, 125.5 for women).</u>