

IPUMS Data Training Exercise

An Introduction to IPUMS Time Use

(Exercise 1 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_SPORTS, BLS_SOCIAL_RELIG, EDUC, and EMPSTAT; 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 <i>ingredients</i> and then <i>stir</i> and then <i>cook</i> ").
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 Select Samples, select the year 2011, keep the default "ATUS respondents" and click Submit sample selections.
- Click on the "Time Use" dropdown menu, and select "Activity coding structure". Click on the plus sign next to ACT_SPORTS to select the variable and add it to the cart.

Request the Data

- Click again on the "Time Use" dropdown menu, and select "BLS published tables". Select the variable BLS_SOCIAL_RELIG (close to the bottom of the page) to add to our data cart by clicking the plus sign next to the variable.
- Select the Person dropdown menu, click on Technical Person, and select the DAY variable.
- Select the variables
 - ACT_SPORTS: Sports, exercise, and recreation
 - BLS_SOCIAL_RELIG: Religious and spiritual activities
 - EDUC: Highest level of education attained
 - DAY: Day of the week of interview
 - SEX: Sex
 - EMPSTAT: Employment status

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. If you would like to use a different stats package, see:

<https://ipums.org/support/exercises>

Download the data

- Go to <https://www.atusdata.org/atus/> and click on Download or Revise Extracts.
- Right-click on the Data link next to the extract you created.
- Choose "Save Target As..." (or "Save Link As...").
- Save into "Documents" (Documents should pop up as the default location).
- Do the same 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 menubar, 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 have 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 package, 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)
```

- 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-labels vignette in the R package. From R, run command:

```
vignette("value-labels", package = "ipumsr")
```



Analyze the Sample

Part I: Generate a New Variable

1. On the website, go to “Build an Extract,” which you'll find on the left hand side of the homepage. Click on Education in the Person dropdown menu under Select Variables. Click on EDUC and find the codes for educational attainment
 - a. Create a variable that combines the codes into four categories.

```
ipums_val_labels(data$EDUC)
data <- data %>%
  mutate(
    EDUC_CAT = EDUC %>%
      lbl_na_if(~.lbl == "NIU (Not in universe)") %>%
      lbl_relabel(
        lbl(1, "Less than HS") ~ .val %in% 10:19,
        lbl(2, "HS Degree") ~ .val %in% 20:29,
        lbl(3, "Some college") ~ .val %in% 30:39,
        lbl(4, "College degree +") ~ .val %in% 40:49
      )
  )
```

2. Is there a difference in the average number of minutes spent doing religious activities reported by individuals of different education levels in this sample? _____

```
data %>%
  group_by(EDUC_CAT = as_factor(EDUC_CAT)) %>%
  summarize(BLS_SOCIAL_RELIG = mean(BLS_SOCIAL_RELIG))
```



Part II: Using Weights (WT06)

Note on Weights

The ATUS sample design requires use of weights to provide an 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.

3. Now answer question 2 using weights. _____

```
data %>%  
  
group_by(EDUC_CAT = as_factor(EDUC_CAT)) %>%  
summarize(BLS_SOCIAL_RELIG = weighted.mean(BLS_SOCIAL_RELIG,  
WT06))
```

Part III: Relationships in the Data

Go to the homepage and choose "Time Use Variables" on the left column under Data. Click ACT_SPORTS and then the link to the description.

4. Is physical exercise the only thing that falls under this category? _____

5. What is the average number of minutes spent doing activities under the ACT_SPORTS category for each day of the week? _____

```
data %>%  
  
group_by(DAY = as_factor(DAY)) %>%  
summarize(ACT_SPORTS = weighted.mean(ACT_SPORTS, WT06))
```



6. What is the average number of minutes spent on these activities for each day of the week if the time was more than zero? Do you have a theory why weekends might differ generally from weekdays? Do these averages seem reasonable? _____

```
data %>%  
  filter(ACT_SPORTS > 0) %>%  
  group_by(DAY = as_factor(DAY)) %>%  
  summarize(ACT_SPORTS_IF_ANY = weighted.mean(ACT_SPORTS, WT06))
```

7. How many people reported exercise or watching sports on the day of their interview? _____

```
data %>%  
  group_by(ANY_SPORTS = ACT_SPORTS > 0) %>%  
  summarize(ACT_SPORTS_ANY_NUM = n()) %>%  
  mutate(ACT_SPORTS_ANY_PCT = ACT_SPORTS_ANY_NUM /  
  sum(ACT_SPORTS_ANY_NUM))
```

8. How many observations does this year's sample have for ACT_SPORTS? What percentage of people reported time spent doing ACT_SPORTS activities? _____

9. What is the percent of people employed in the sample? _____

```
data %>%  
  group_by(EMPSTAT = as_factor(EMPSTAT)) %>%  
  summarize(n = n()) %>%  
  mutate(pct = n / sum(n))
```



10. Find the average amount of time participating in sports according to employment status for women. _____

```
data %>%  
filter(SEX == 2) %>%  
group_by(EMPSTAT = as_factor(EMPSTAT)) %>%  
summarize(ACT_SPORTS = weighted.mean(ACT_SPORTS, WT06))
```

11. Find the average amount of time participating in sports according to employment status for men. What differences do you see? _____

```
data %>%  
filter(SEX == 1) %>%  
group_by(EMPSTAT = as_factor(EMPSTAT)) %>%  
summarize(ACT_SPORTS = weighted.mean(ACT_SPORTS, WT06))
```



Answers

Part I: Generate a New Variable

1. On the website, go to “Build an Extract”...
2. Is there a difference in the average number of minutes spent doing religious activities reported by individuals of different education levels in this sample? Less than HS diploma: 17.16 min; HS Diploma: 14.69; Some College: 12.73; College Degree+: 12.17

Part II: Using Weights (WT06)

Note on Weights

3. Now answer question 2 using weights. Less than HS diploma: 12.26 min; HS Diploma: 9.35; Some College: 8.58; Graduate 8.08

Part III: Relationships in the Data

4. Is physical exercise the only thing that falls under this category? No, ACT_SPORTS includes physical exercise, sports recreation, and watching sports. To find only time spent on physical exercise, you would need to Create your own time variable. See Exercise 2 for instructions for creating new variables in the extract builder.
5. What is the average number of minutes spent doing activities under the ACT_SPORTS category for each day of the week? Sunday 22; Monday 18; Tuesday 19; Wednesday 20; Thursday 17; Friday 17; Saturday 28
6. What is the average number of minutes spent on these activities for each day of the week if the time was more than zero? Do you have a theory why weekends might differ generally from weekdays? Do these averages seem reasonable? Sunday 128; Monday 89; Tuesday 93; Wednesday 99; Thursday 84; Friday 104; Saturday 136; Weekend days might have greater time spent exercising



because the person does not have work, or watching professional sports games that are traditionally held on weekends.

7. How many people reported exercise or watching sports on the day of their interview? *Note: See the output from question 3 and sum the_N of each day.* 2319 People
8. How many observations does this year's sample have for ACT_SPORTS? What percentage of people reported time spent doing ACT_SPORTS activities? 12,479 people; 2,319/12,479 = 18.6%
9. What is the percent of people employed in the sample? 57.62%
10. Find the average amount of time participating in sports according to employment status for women.

<u>Average number of minutes in sports participation</u>		
	Women	Men
Employed - at work	15.03	21.66
Employed - absent	19.68	19.65
Unemployed - on layoff	19.26	24.36
Unemployed - looking	19.00	49.24
Not in labor force	13.76	31.13

11. Find the average amount of time participating in sports according to employment status for men. What differences do you see? The average time spent participating in sports doesn't seem to differ across employment status for women, but it does for men.

