

Economics 142 Problem Set #1
Due Thursday Sept. 1 in class

Note: It is recommended that you use the R programming language for assignments in this class. You are welcome to use other programming languages (e.g., Matlab and Python) or software packages (e.g., Stata). However, the problem sets will contain instructions assuming you are using R.

For problem sets that require coding (probably all of them), you are required to submit your **reproducible code separately** along with a write up with answers to the questions. Online submissions are encouraged.

1. Set up an R program to conduct a simulation of drawing a sample size of n from a Bernoulli distribution with mean p . In each "replication" r you will draw a sample, construct the estimated mean from that replication (which we will denote as \bar{Y}_r), and calculate the 95% confidence interval $(\bar{Y}_r - 1.96 s_r / n^{1/2}, \bar{Y}_r + 1.96 s_r / n^{1/2})$ where s_r is the estimated standard deviation in that replication, $s_r = (\bar{Y}_r (1 - \bar{Y}_r))^{1/2}$. In each replication, record the length of the confidence interval, and whether or not the true mean is inside the interval.

For given (n, p) , conduct 1,000 replications and report the following statistics:

- a) the mean estimate of p
- b) the mean estimate of the true standard deviation
- c) the fraction of time that the confidence interval contains the true p . This is called the "coverage rate"

Conduct the analysis for the cases $n=30$ using $p=0.05$ and $p=0.25$, and again for $n=60$ using $p=0.05$ and $p=0.25$ (a total of 4 cases). It is often claimed that n of 30 or larger is enough to ensure that asymptotic confidence intervals work well. Do you agree or not?

2. Go to the course bcourses page and download the data set "atus_sum_2013.csv" (Files -> ps1). This is a cleaned version of the 2013 American Time Use Survey. It has 11,386 rows (one for each person) and 413 columns. The first few columns are variables that describe the person – see atus-column-names.pdf for a brief description, and atusintcodebk13.pdf for a more complete codebook.

The 2013 Activity Summary file contains data for the total number of minutes that each respondent spent doing each 6-digit activity on a chosen day. These are recorded in the variables in columns Z ... OX (a total of 389 separate columns). You will see that most of these are 0 for any one person.

The variable names (i.e., column headings) identify activities by their 6-digit codes. The letter "t" precedes the 6-digit codes. The first and second digits of the 6-digit code correspond to the variable tutier1code; the third and fourth digits of the 6-digit code correspond to the variable tutier2code; and, the fifth and sixth digits of the 6-digit code correspond to the variable tutier3code.

For example, the variable "t040102" refers to the activity with TUTIER1CODE=04, TUTIER2CODE=01, and TUTIER3CODE=02. As documented in the ATUS 2013 Coding Lexicon (<http://www.bls.gov/tus/lexiconwex2013.pdf>), activity code "040102"

corresponds to "Caring for and helping nonhousehold members/Caring for and helping nonhousehold children/Reading to or with nonhousehold children." See Chapter 5 of the ATUS User's Guide (<http://www.bls.gov/tus/atususersguide.pdf>) for more information about the ATUS Coding Lexicon. All "t" activity variables on this file correspond to an activity that at least one survey respondent reported doing for one minute or more.

The New York Times used the 2013 ATUS to analyze how American men and women spend their weekdays. You are encouraged to read the entire article.¹ Here is a paragraph summarizing some findings:

"Nonworkers spend much more time doing housework. Men without jobs, in particular, spend more time watching television, while women without jobs spend more time taking care of others. And the nonemployed of both sexes spend more time sleeping than their employed counterparts."

The goal of this course is not only to teach you econometric methods, but also how to present findings in ways that are accessible to others. In this question, we will aim to create a much simpler version of the interactive graph in the article.

- a) Subset the Summary file into four separate data frames for employed males, employed females, nonemployed males and nonemployed females. (Hint: use the subset function).
- b) Using the data frames you created above, find and store the **average amount of time** spent performing the following tasks for each of the four groups:
 - a. t050101 – "work, main job"
 - b. t120101 – "socializing and communicating with others"
 - c. t110101 – "eating and drinking"
 - d. t120303 – "television and movies (not religious)"
 - e. t010201 – "washing, dressing and grooming oneself"
 - f. t120307 – "playing games"
 - g. t020101 – "interior cleaning"
 - h. t010101 – "sleeping"

In addition, compute the amount of time left over after accounting for all these activities. You should end up with 4 vectors of length 9. For example, here is what you should find for employed males (in hours):

t050101	t120101	t110101	t120303	t010201	t120307	t020101	t010101	Rem
4.6979851	0.6177969	1.1639767	2.4984760	0.5761328	0.1677304	0.2130920	8.4554362	5.6093740

Double check that these vectors add up to the length of a day (24 hours in this case).

- c) Combine the 4 vectors you created in b) into a matrix or data frame with each row corresponding to one of the 4 groups. Use this to graphically summarize the different time use patterns as best as you can. Provide a short written analysis of the patterns you see.

¹ "How Nonemployed Americans Spend Their Weekdays: Men vs. Women", Josh Katz, NYT (Jan 6 2015). Available online at <http://nyti.ms/1MM1aHW>.

Below is an example of a well-labeled, visual analysis. The graph was produced using the ggplot2 package. You will receive partial credit for this part if you produce a table instead of a graph.

