

Problem Set #1

MACS 30200, Tong Ju

Apr. 19 2017

Data Section –Current Population Survey (CPS) Data (Jan. 2016)

The Current Population Survey (CPS), which is aimed to collect the information about the nation-wide unemployment and labor force participation in U.S., is a monthly survey of approximately 60,000 U.S. households conducted monthly by the United States Census Bureau for the Bureau of Labor Statistics (BLS). In this monthly survey, the data, such as employment status, work experience, occupational mobility, job tenure, educational attainment, income and so on are collected for the civilian non-institutional population (16 years and over) by age, sex, race, Hispanic origin, marital status, and family relationship. The BLS is responsible for curating the CPS data, which is available through CPS homepage¹.

Drawing upon the CPS data, early each month, BLS announces the total number of employed and unemployed people in the United States for the previous month, along with many characteristics about them. These figures, particularly the unemployment rate, informing the public the percentage of the labor force that is unemployed, receive wide coverage in the media. Besides this, the CPS data is also widely used by scholars to address various research questions. For example, researchers use the CPS data to investigate the social cause and effect of unemployment (Meisenheimer , 2000; Borbely, 2010; Ilg, 2010; Sok, 2010), income inequality (Cutler, 1992; Atkinson, 2001; Burkhauser, 2000 and 2008) and labor force characteristics (Card, 1990; Cohany and Emy, 2007; Hirsch, 2003). In addition, when combining with other data set, the CPS data can be used to answer the questions, such as, how the employment status relates to health condition (Bartley, 1996) and

¹ <https://www.census.gov/programs-surveys/cps.html>

medication use (Kaufman, 2002).

As for the process of data collection, the CPS is conducted by a rotation scheme of “4-8-4”, in which a new rotation groups enters the CPS every month and each group is first interviewed for four consecutive months, temporarily dropped for eight months, and then re-interviewed for four consecutive months. Under this sampling frame, each month sample consists of eight rotation groups, in any two-month period there are six overlapping rotation groups, and for any rotation group the second wave of four interview occurs, one year later, in the same calendar months as the first wave. Therefore, if the sample size was constant and there was no attrition, three fourths of the sample in any two consecutive months could be matched (Peracchi, 1992).

This study employs the CPS data on January of 2016, which originally contains 151, 010 individuals. In order to explore how the income and working hour can be influenced by the demographic factors, 9 variables were chosen from the data set, which include individual income², gender, race, age, education attainment, marital status, region, and weekly working hours. After removing the missing data, 60, 655 unique individual samples are finally used to do the analysis. The descriptive statistics is summarized in table 1. The average age of this population is 43 with an education attainment of high school or above. The average annual income of individual is 27, 493 dollars, and on average the respondents in this survey work for 35.55 hours each week. In the Figure 1 and 2, the histogram plot of the annual income and weekly work hours are shown. It is obvious that the annual income follows the truncated normal distribution, while the distribution of weekly work hours concentrates around 35 to 40 hours.

² In this data set, the individual income is calculated by dividing the household income by the household size.

Table 1: Descriptive Statistics for some of the key variables

Variable	Percentage	Mean	Standard Deviation
Sample = 60, 655			
Gender			
<i>Male</i>	51.68		
<i>Female</i>	48.32		
age		43.09	14.4
Education Level ³		3.878	1.321
Marital Status			
<i>Married</i>	56.4		
<i>Single</i>	43.6		
Geographic Features			
<i>Midwest</i>	20.78		
<i>Southern</i>	35.02		
<i>Northeast</i>	17.12		
West	27.08		
Annual Income		27493	21608
Weekly Work Hours		35.55	15.34

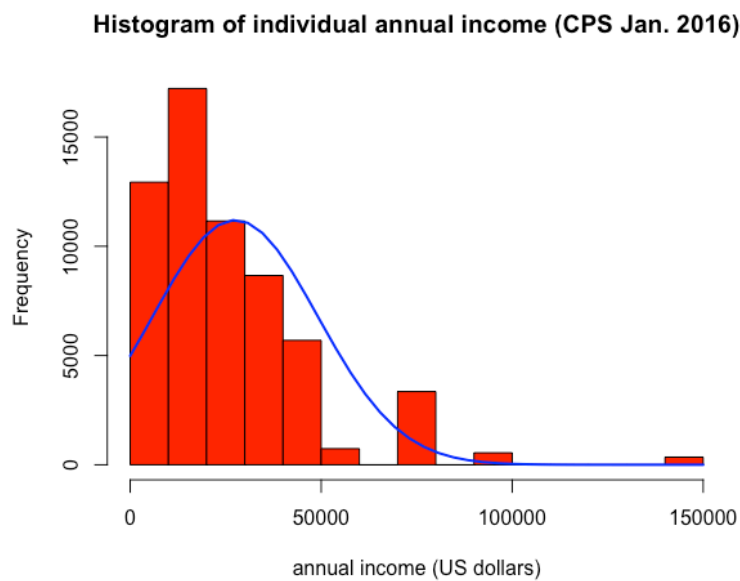


Figure 1

³ The scales are coded as following: 1 = 8th grade or less; 2 = 9th to 12th grade without diploma; 3 = high school graduate; 4 = college credit without degree; 5 = Associate degree; 6 = Bachelor's degree; 7 = Master's degree; 8 = Doctorate or professional degree

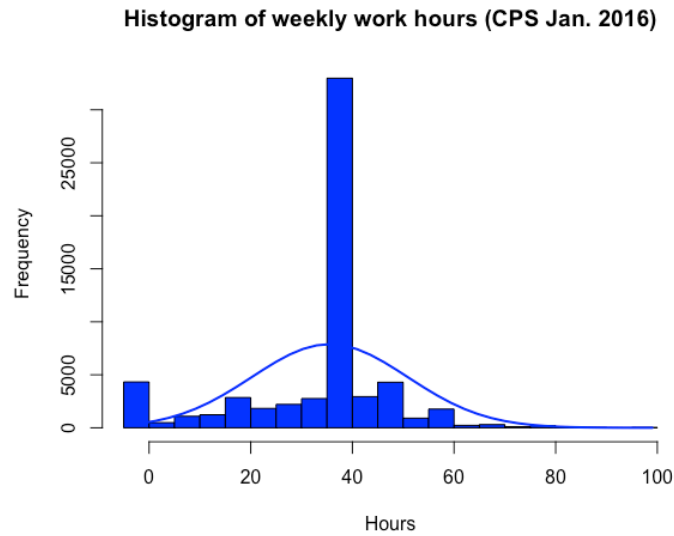


Figure 2

In the following sections, I focus on how the annual income distributes across gender, age, education attainment, region, and race. On average, male earns more annual income than female by around 1200 dollars (Figure 3). Not surprisingly, the annual income gradually increases until the respondents reach their age of retirement (60-65 years old, Figure 4). In addition, there is a strong positive relationship between the education attainment and income, shown in Figure 5.

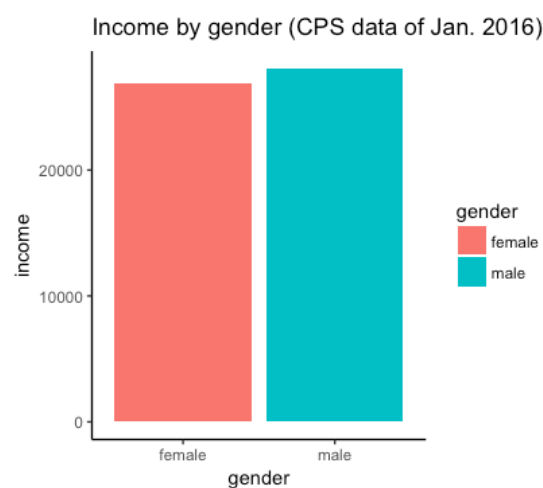


Figure 3.

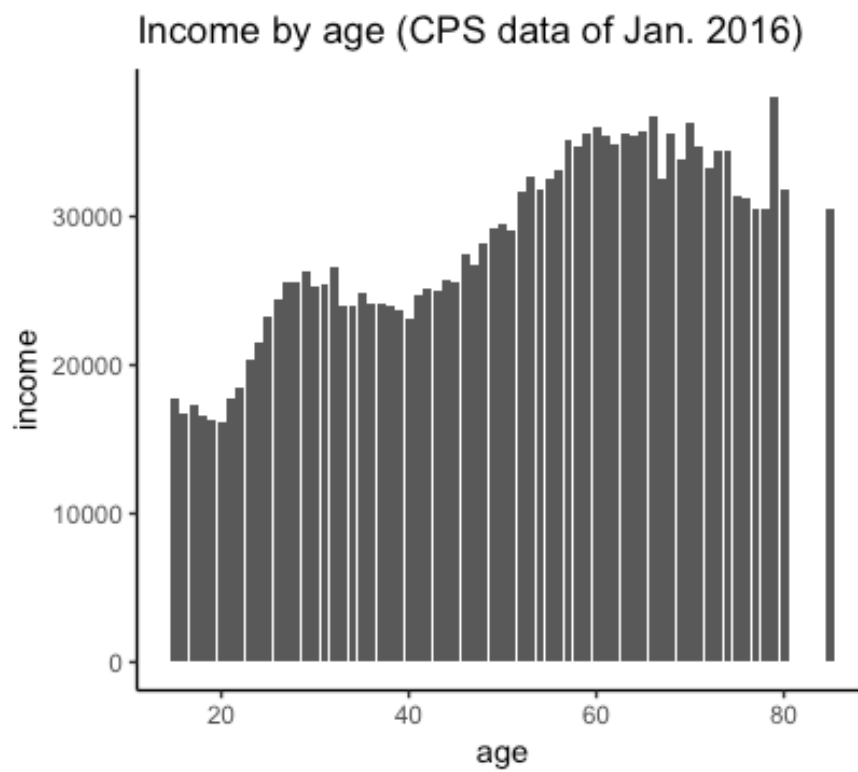


Figure 4.

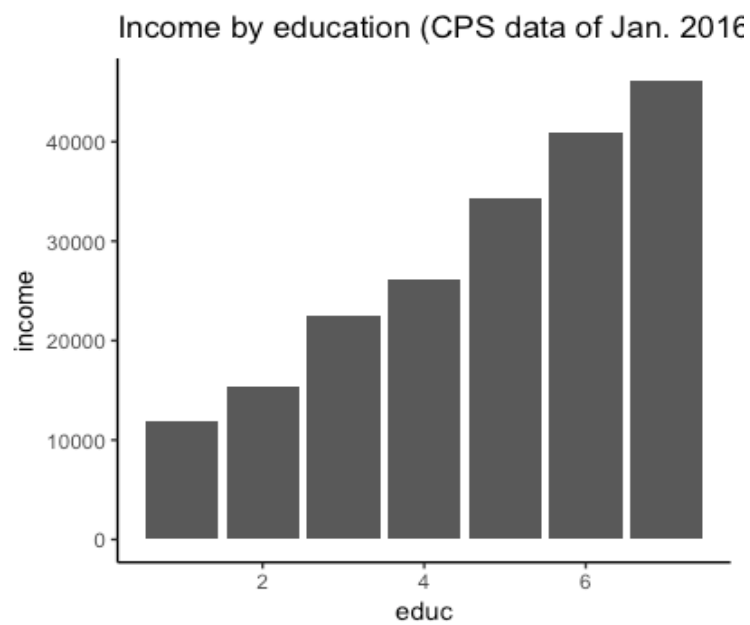


Figure 5

According to Figure 6, it is the white and Asian people who earn more money than other races. The average annual income for white people is around 28,305 dollars, Asian people 28,166 dollars, black people 22,400 dollars, Others 22,039 dollars, while Native American people's annual income is the least in all five racial categories, around 18,183 dollars, which indicates the income inequality across various races. Also, the data shows that people living in the Northeast have higher (approximately 3000 dollars higher) annual income than people living in other regions of U.S. (Figure 6).

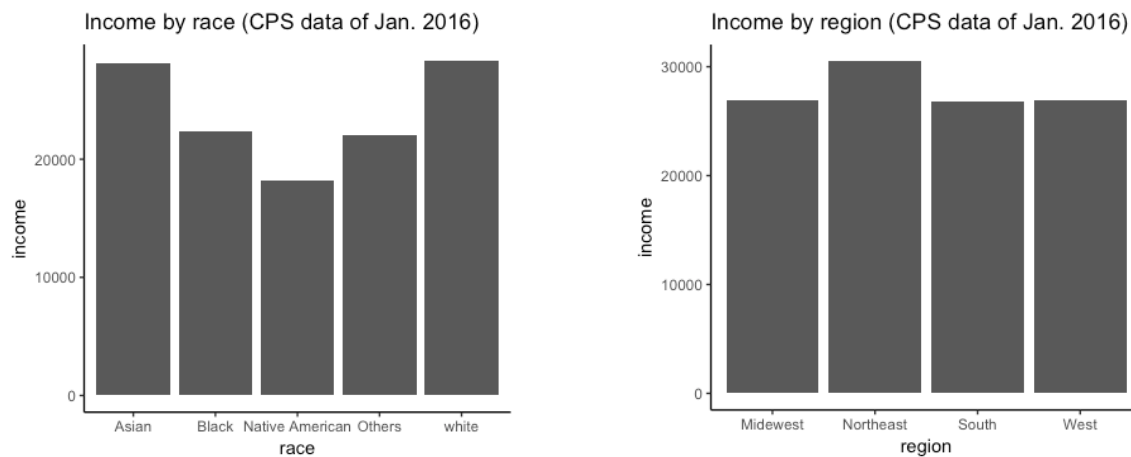


Figure 6

Finally, to explore the relationship between the demographic factors with the annual income and weekly work hour, the OLS regression is deployed, in which, the annual income and weekly work hour serves as dependent variables, while gender, age, race, region, marital status serve as independent variables. The statistics of two models are presented in the table 2 on page 7. The regression result has shown that gender, age, race, education level, regions and marital status all have strong associations with the individual annual income ($p < 0.01$). As the bar charts I have presented above, age and education levels are in positive relation with the annual income. Also, a white male living in Northeast has much higher possibility to earn more money than female in other racial groups living in other places.

Table 2: OLS regression for the income and work hours.

OLS regression for the income and work hour (CPS Jan. 2016)		
	<i>Dependent variable:</i>	
	income (1)	WorkHour (2)
sexmale	2,278.133*** (158.666)	3.832*** (0.123)
raceBlack	-2,594.851*** (442.896)	0.067 (0.343)
raceNative American	-3,950.977*** (816.322)	0.657 (0.632)
raceOthers	-1,488.123** (662.996)	-0.938* (0.513)
racewhite	2,404.139*** (360.855)	-0.566** (0.279)
age	330.194*** (5.786)	-0.001 (0.004)
educ	5,931.071*** (61.095)	1.422*** (0.047)
maritalsingle	5,924.984*** (170.292)	-2.379*** (0.132)
regionNortheast	1,952.075*** (257.857)	-0.368* (0.200)
regionSouth	614.001*** (220.599)	1.185*** (0.171)
regionWest	533.123** (232.514)	0.006 (0.180)
Constant	-15,840.330*** (548.305)	29.247*** (0.425)
Observations	60,655	60,655
Log Likelihood	-684,936.600	-250,443.200
Akaike Inf. Crit.	1,369,897.000	500,910.300
Note:	*p<0.1; **p<0.05; ***p<0.01	

On the other hand, this result also indicates that compared with annual income, the weekly work hours are only associated with some of the demographic variables, such as gender, education level, marital status and region. The independent variables, such as race

and age seems not to be in strong relation with work hours. On average, male has longer work hours than female, and people with higher education living in the South are more likely to work for longer time per week than other people. On the other hand, the singles have more probability to work in less hours than the married people. In conclusion, according to the CPS data, we can find there is strong relationship between the income and work hours with the demographic factors. As I have shown above, the CPS data could be used to investigate the income inequality in future research.

Reference

- Atkinson, Anthony B., and Andrea Brandolini. "Promise and pitfalls in the use of" secondary" data-sets: Income inequality in OECD countries as a case study." *Journal of economic literature* 39, no. 3 (2001): 771-799.
- Bartley, Mel, and Charlie Owen. "Relation between socioeconomic status, employment, and health during economic change, 1973-93." *Bmj* 313, no. 7055 (1996): 445-449.
- Borbely, James M. "Sizing up the 2007–09 recession: comparing two key labor market indicators with earlier downturns." *Issues in Labor Statistics* (2010): 10-11.
- Burkhauser, Richard V., Kenneth A. Couch, and David C. Wittenburg. "A reassessment of the new economics of the minimum wage literature with monthly data from the Current Population Survey." *Journal of Labor Economics* 18, no. 4 (2000): 653-680.
- Burkhauser, Richard V., Shuaizhang Feng, Stephen P. Jenkins, and Jeff Larrimore. Estimating trends in US income inequality using the Current Population Survey: the importance of controlling for censoring. No. w14247. National Bureau of Economic Research, 2008.
- Card, David. "The impact of the Mariel boatlift on the Miami labor market." *ILR Review* 43,

no. 2 (1990): 245-257.

Cohany, Sharon R., and Emy Sok. "Trends in labor force participation of married mothers of infants." *Monthly Lab. Rev.* 130 (2007): 9.

Cutler, David M., and Lawrence F. Katz. *Rising Inequality? Changes in the Distribution of Income and Consumption in the 1980s*. No. w3964. National Bureau of Economic Research, 1992.

Hirsch, Barry T., and David A. Macpherson. "Union membership and coverage database from the current population survey: Note." *ILR Review* 56, no. 2 (2003): 349-354.

Ilg, Randy. "Long-term unemployment experience of the jobless." Washington, DC: Bureau of Labor Statistics. Retrieved July 27 (2010): 2010.

Kaufman, David W., Judith P. Kelly, Lynn Rosenberg, Theresa E. Anderson, and Allen A. Mitchell. "Recent patterns of medication use in the ambulatory adult population of the United States: the Slone survey." *Jama* 287, no. 3 (2002): 337-344.

Meisenheimer, Joseph R. "Looking for a better job: job-search activity of the employed." *Monthly Lab. Rev.* 123 (2000): 3.

Peracchi, Franco, and Finis Welch. "How representative are matched cross-sections? Evidence from the Current Population Survey." *Journal of Econometrics* 68, no. 1 (1995): 153-179.

Sok, Emy. "Record unemployment among older workers does not keep them out of the job market." *Issues in labor statistics* 10 (2010): 04-16.