Problem Set 1
Sushmita V Gopalan
MACS 30200
Dr. Evans
April 19, 2017

The United States' Current Population Survey (CPS) is one of the country's largest surveys, providing vital information on labour, earnings, education, and health. Each month, a supplemental questionnaire is included, covering a range of issues like child support, health insurance coverage, tobacco usage, etc. depending upon the needs of the body sponsoring the supplement. The CPS is conducted by the United States Census Bureau for the Bureau of Labour Statistics. It is hosted at the website for the United States Census Bureau and also in the archives of the National Bureau of Economic Research.

The Current Population Survey is conducted on a probability-selected sample of about 60,000 households. Each household is under survey for four straight months, out of survey for the next eight months and then back in the sample for a final four months. In effect, one-eighth of the sample is new each month. Data collection is carried out using a computerized questionnaire, with enumerators using both phone and in-person survey modes. The week of the 19th each month is when surveying is usually conducted. One adult in the household generally provides information regarding all other household members.

A brief survey of some of the key papers that have used the Current Population Survey illustrates the variety of important ways in which this data is utilized. For instance, Hirch and Macpherson (2003) used CPS data to create a database with estimates of private and public sector union membership, coverage and density estimates[1]. This data is accessible here. Burkhauser, et al. (2000) used monthly CPS data to investigate the effects of increases in federal minimum wages on employment[2]. Mellor and Milyo(2002) looked into the effect of income inequality on health status for poor and non-poor individuals to find no consistent association between income

[1] Hirsch, Barry T., and David A. Macpherson. "Union membership and coverage database from the current population survey: Note." *ILR Review* 56.2 (2003): 349-354.

[2] 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* 18.4 (2000): 653-680.

inequality and individual health status[3]. Frostin(1996) collated summary data on insured and uninsured populations in each state and explored the ways in which health protection changes for the insured and identified some key variables that are related to whether or not an individual is likely to have health insurance[4].

Here are some summary statistics for 8 primary variables in the January 2017 Current Population Survey dataset. As they are mostly categorical variables, summary statistics like means, quartiles or standard deviation are not meaningful. Medians have been included only for ordinal or continuous variables. The sample consists of surveys from 151,010 individuals.

Table 1: Descriptive Statistics (N=151,010)

	No of Missing Obs	% Missing Obs	Most Frequent Response	% Most Frequent Response	Least Frequent Response	% Least Frequent Response	Median
Education[1]	45590	30.19%	High School Diploma	27.95%	Less than First Grade	0.03%	Some College, No Degree
Income[2]	20812	13.78%	$100,000-$150,000	13.12%	$5000-$7499	0.02%	$50,000-$59,000
Age	20812	13.78%	55	1.24%	79	0.03%	39
Sex[3]	20812	13.78%	Female	51.65%	Male	48.35%	-
Race[4]	20812	13.78%	White Only	80.04%	American Indian -Hispanic	0.02%	-
Citizenship[5]	20812	13.78%	Native, Born in the US	87.74%	Native, Born in U.S. Islands	0.01%	-
Marital Status[6]	45590	30.19%	Married- Spouse Present	50.45%	Married, Armed Forces Spouse Present	0.03%	-
Hours Worked/Week	90355	59.83%	40	51.64%	94	0.02%	40

[1]**Education Codes**: 31 = Less Than 1st Grade, 32 = 1st, 2nd, 3rd Or 4th Grade, 33 = 5th Or 6th Grade, 34 = 7th Or 8th Grade, 35 = 9th Grade, 36 = 10th Grade, 37 = 11th Grade, 38 = 12th Grade No Diploma ,39 = High School Grad-diploma Or Equiv (Ged), 40 = Some College But No Degree, 41 = Associate Degree-occupational/Vocational, 42 = Associate Degree-academic Program, 43 = Bachelor's Degree (Ex: Ba, Ab, Bs), 44 = Master's Degree, 45 = Professional School Deg (Ex: Md, Dds, Dvm), 46= Doctorate Degree (Ex: Phd, Edd)
[2]**Income:** 1 = Less Than $5,000, 2 = 5,000 To 7,499, 3 = 7,500 To 9,999, 4 = 10,000 To 12,499, 5 = 12,500 To 14,999, 6 = 15,000 To 19,999, 7 = 20,000 To 24,999, 8 = 25,000 To 29,999, 9 = 30,000 To 34,999, 10 = 35,000 To 39,999, 11 = 40,000 To 49,999, 12 = 50,000 To 59,999, 13 = 60,000 To 74,999, 14 = 75,000 To 99,999, 15 = 100,000 To 149,999, 16 = 150,000 Or More
[3]**Sex:** 1=Male, 2=Female

[3] Mellor, Jennifer M., and Jeffrey Milyo. "Income inequality and health status in the United States: evidence from the current population survey." *Journal of Human Resources* (2002): 510-539.

[4] Fronstin, Paul. "Sources of health insurance and characteristics of the uninsured: Analysis of the March 1996 Current Population Survey." (1996).

[4]**Race:** 01 = White Only, 02 = Black Only, 03 = American Indian, Alaskan Native Only, 04 = Asian Only, 05 = Hawaiian/Pacific Islander Only, 06 = White-Black, 07 = White-AI, 08 = White-Asian, 09 = White-HP, 10 = Black-AI, 11 = Black-Asian, 12 = Black-HP, 13 = AI-Asian, 14= AI-HP, 15 = Asian-HP, 16 = W-B-AI, 17 = W-B-A, 18 = W-B-HP, 19 = W-AI-A, 20 = W-AI-HP, 21 = W-A-HP, 22 = B-AI-A, 23 = W-B-AI-A, 24 = W-AI-A-HP, 25 = Other 3 Race Combinations, 26 = Other 4 and 5 Race Combinations
[5]**Citizenship:** 1 = Native, Born In The United States, 2 = Native, Born In Puerto Rico Or Other U.S. Island Areas, 3 = Native, Born Abroad Of American Parent Or Parents, 4 = Foreign Born, U.S. Citizen By Naturalization, 5 = Foreign Born, Not A Citizen Of The United States
[6]**Marital Status:** 1 = Married - Spouse Present, 2 = Married - Spouse Absent, 3 = Widowed, 4 = Divorced, 5 = Separated, 6 = Never Married

We see here that the modal income is $100,000-$149,999. This seems to be inordinately high, given that per capita income in the United States is estimated to be around $53,000 per annum. Here is a histogram of the income categories.

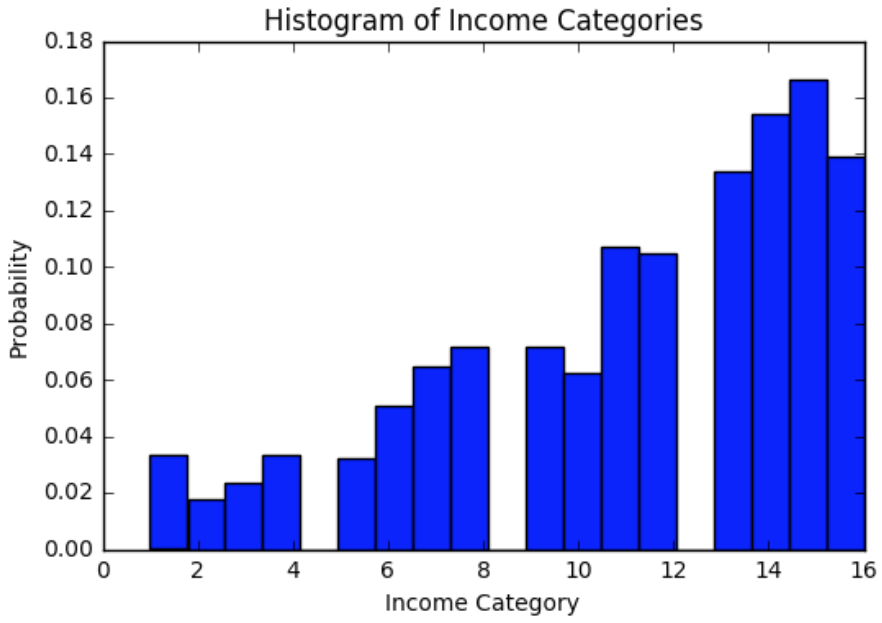

Figure 1: Histogram of Income Categories

There is missing data at categories 5, 9 and 13 and the histogram also confirms the modal observation of 15, i.e. $100,000-$149,999. Closer inspection revealed a warning in the dataset codebook against using the income variable due to an error in data collection.

Now, I recode the 'Education' and 'Race' categories. The 'education' variable is recoded into 1:"No High School Diploma", 2:"High School Diploma", 3:"College, No Degree" and 4:"Bachelor's Degree or Above", while 'race' is recoded into "White", "Black" and "Other. The chart below shows the distribution of the number of hours worked per week, conditional upon education levels. We see that the lowest education category has the lowest proportion of

individuals working >40 hours, and the highest proportion of individuals working <20 hours. The second-highest education level has the highest proportion of individuals working > 40 hours a week, while the highest education level seems relatively more evenly distributed across the three categories.

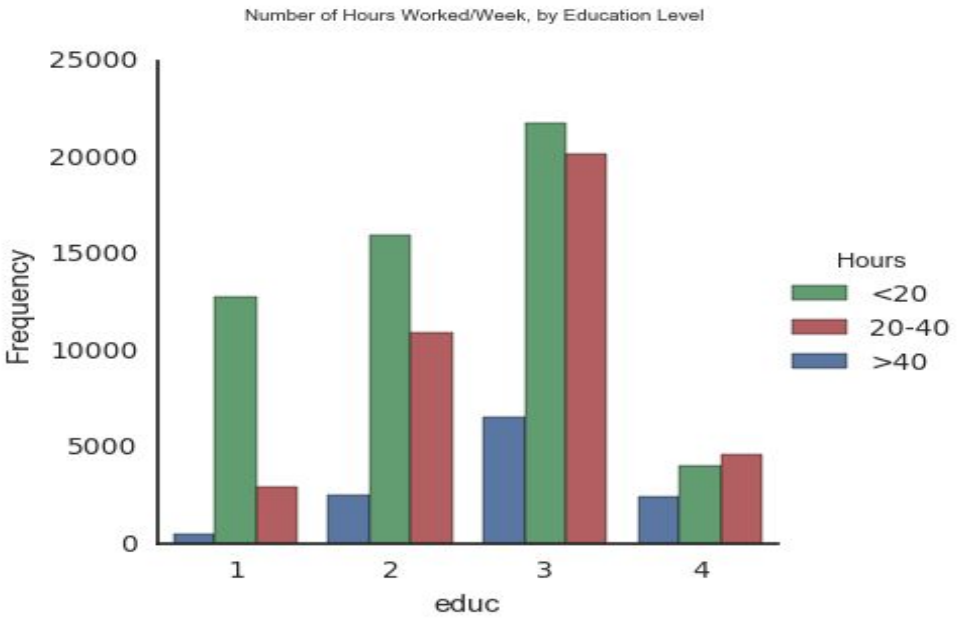

Figure 2: Number of Hours Worked by Education Level

We now look at a cross-tabulation of the variables 'education' and 'race', after recoding. The following table provides an understanding of distribution across education levels, given race. We see that 'Other' races have the highest proportion of 'Bachelor's Degree or Above' while 'Black' individuals have the highest proportion of 'No High School Diploma' and 'High School Diploma. 'White' individuals have the highest proportion of 'College, No Degree' and the second-highest proportion of 'Bachelor's Degree or Above'.

Table 2: Race and Education

	White	**Black**	**Other**
No High School Diploma	14.81%	19.34%	16.98%
High School Diploma	27.83%	32.55%	23.28%
College, No Degree	46.69%	41.00%	45.99%
Bachelor's Degree or Above	10.67%	7.12%	13.75%