

STATA Assignment I

Due June 28 (by 12:00 pm)

The Current Population Survey (CPS) is a monthly survey of about 50,000 households conducted by the Bureau of the Census for the Bureau of Labor Statistics (more info on the CPS can be found at <http://www.census.gov/cps/>). The data set “CPSdata.dta” is a subset of the CPS from year 2004 for male workers from the state of North Carolina.

Use a Stata log file to record your results, and print out all STATA output. Feel free to write out the answers to the short answer questions on the Stata Log printout if you can do so legibly; otherwise you can cut and paste the log file into a Word document and type out your answers where appropriate.

I. Data Management

- (1) Download “CPSdata.dta” from the course website. Use the ‘describe’ command to familiarize yourself with the variables contained in the data.
- (2) Drop the variable ‘recnum’ from the data set
- (3) Rename the variable ‘interest_income’ as ‘intinc’. Assign the variable label “Interest income” to the variable ‘intinc’
- (4) Generate a new variable ‘annhrs’, where $\text{annhrs} = \text{hours} * \text{wkslyr}$
- (5) Generate a new variable ‘wage’, where $\text{wage} = \text{earnings} / \text{annhrs}$
- (6) Generate a new variable ‘agegrp’, where $\text{agegrp} = 1$ if $\text{age} \geq 25$ & $\text{age} \leq 34$; $\text{agegrp} = 2$ if $\text{age} \geq 35$ & $\text{age} \leq 44$; and $\text{agegrp} = 3$ if $\text{age} \geq 45$ and $\text{age} \leq 54$

II. Describing Data

- (1) Get summary statistics of all the variables in the data set.
 - ⇒ What is the minimum and maximum age in the data?
 - ⇒ Is the maximum value of hours an outlier? Compute the z-score
- (2) Get detailed summary statistics for the variable ‘earnings’.
 - ⇒ What is the mean value for earnings?
 - ⇒ What is the median value of earnings?
 - ⇒ What is the 90th percentile value for earnings?
- (3) Create a histogram for the variable ‘earnings’. Is the distribution bell shaped?

- (4) Look at the relationship between earnings and hours.
- ⇒ Create a scatterplot of the relationship between earnings and hours. Does there appear to be a positive, negative, or zero relationship between the two variables?
 - ⇒ What is the correlation coefficient between ‘earnings’ and ‘hours’? Does this confirm your intuition from the scatterplot?
- (5) Does the wage rate differ by demographic characteristics?
- ⇒ Compute mean wage rate by marital group. Which group has the highest mean wage rate?
 - ⇒ Create a bar graph that shows the mean wage rate for each education group. Does wage appear to be increasing or decreasing in educational attainment?
 - ⇒ Create a bar graph that shows the median wage rate for each education group.
- (6) Interest income is the money that a respondent receives from interest on assets (savings, stocks, etc.).
- ⇒ How does mean interest income vary by age group?
 - ⇒ Get detailed summary statistics on interest income (the variable ‘intinc’) for men ages 45 to 54 only. Within this group, how does median interest income compare with mean interest income among college graduates? Does this suggest the distribution of interest income among men ages 45 to 54 is symmetrical, or skewed to the right or left?