Data Sets Used

Variables from Each Data Set

Data Management

2/29

Data was sorted from differing sources small to large by FIPS code and placed side by side in Excel to locate data that did not concur. Once nonconcurrent data was found (in the form of counties or areas), it was deleted from the selection. Then redundant columns (like duplicate FIPS codes) were eliminated.

3/7

Data was copied and pasted from Excel to the Data Editor in STATA. Then some variables were recoded to eliminate missing data, and the commands were saved in the “CurrentProgress3-7.do” Do File, for convenience.

Dichotomous variables were recoded from “1” and “2” to “0” and “1.”

3/14

Learned commands:

histogram \*dependent var\*

histogram \*dependent var\* normal

twoway ( lfitci \*dep\* \*ind\* )

3/28

More data cleaning, reassessed which variables to use or not use. Added Unemployed, Severly Work Disabled and Major Depression numbers (not rates, those were created). Also, Age categories (0-19, 19-65, 65-85, and 85+).

Did something to mess up the cleaning do file 🡺 Must fix

4/4

. corr proctoscopy mammogram pap\_smear

(obs=3141)

| procto~y mammog~m pap\_sm~r

-------------+---------------------------

proctoscopy | 1.0000

mammogram | 0.7965 1.0000

pap\_smear | 0.6191 0.6144 1.0000

corr proctoscopy mammogram pap\_smear

over\_age85orr proctoscopy mammogram pap\_smear povertyRec onlyHSpercent

corr over\_age85 proctoscopy mammogram pap\_smear povertyRec onlyHSpercent

factor proctoscopy mammogram pap\_smear

rotate

sum proctoscopy mammogram pap\_smear

<http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_10_1YR_S1902&prodType=table>

Downloaded and filtered median income data by county

4-11

Matched median income with proper county.

Began new variable: Income reported? (Y/N)