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 non-concurring 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)

4-18

Completed Income Reported (Y/N) using excel command “=IF(AK2>0,1,0”

Data Cleaning complete, real work begin

gen EDU\_AveYears= ( (18\*EDUcollDegreeprcnt/100) + (13.5\* EDUsomeCollegeprcnt/100) + (12\* EDUonlyHSprcnt/100) + (10\* EDUlessthanHSprcnt/100))

gen AGE\_roughcompositete= ( (8\* AGE19\_64/100) + (38\* AGE19\_64/100) + (71\* AGE65\_84/100) + (88\* AGEover\_85/100))

Define categories of interest: Service Availability, Negative Choices and GeneralInitiative

Services Available:

Center YN

Healthcare Shortage Area YN

Dentist Rate

Physician Rate

Personal Choice:

No\_Exer

Few\_FriutVeg

Smoke

Uninsured

Initiative:

EDU\_

MedianInc

gen EDU\_AveYears= ( (18\*EDUcollDegreeprcnt/100) + (13.5\* EDUsomeCollegeprcnt/100) + (12\* EDUonlyHSprcnt/100) + (10\* EDUlessthanHSprcnt/100))

tab EDU\_AveYears,m

histogram AGE65\_84, normal

gen AGE\_roughcomposite= ( (8\* AGE19\_64/100) + (38\* AGE19\_64) + (71\* AGE65\_84/100) + (88\* AGEover\_85/100))]

gen AGE\_roughcomposite= ( (8\* AGE19\_64/100) + (38\* AGE19\_64) + (71\* AGE65\_84/100) + (88\* AGEover\_85/100))

tab AGE\_roughcomposite, m

tab AGE19\_64, m

AGE19\_64/100

tab AGE19\_64/100

drop AGE\_roughcomposite

gen AGE\_roughcomposi AGE\_roughcompositete= ( (8\* AGE19\_64/100) + (38\* AGE19\_64/100 AGE\_roughcomposite AGE\_roughcomposite) + (71\* AGE65\_84/100) + (88\* AGEover\_85/100))

AGE\_roughcompositete= ( (8\* AGE19\_64/100) + (38\* AGE19\_64/100) + (71\* AGE65\_84/100) + (88\* AGEover\_85/100))

gen AGE\_roughcompositete= ( (8\* AGE19\_64/100) + (38\* AGE19\_64/100) + (71\* AGE65\_84/100) + (88\* AGEover\_85/100))

tab AGE\_roughcompositete

aorder

corr Obesity\_Rate CenterYN HP\_ShortageYN DENT\_Rateper100 PHYS\_Rateper100 poverty

factor Obesity\_Rate CenterYN HP\_ShortageYN DENT\_Rateper100 PHYS\_Rateper100 poverty

rotate

corr CenterYN HP\_ShortageYN DENT\_Rateper100 PHYS\_Rateper100 poverty

egen Serv\_Avail= rowmean ( Obesity\_Rate CenterYN HP\_ShortageYN DENT\_Rateper100 PHYS\_Rateper100 poverty )

tab Serv\_Avail, m

regress Obesity\_Rate Serv\_Avail

corr FEW\_FruitVeg\_Rate MAJ\_Depress\_Rate MammogrmRate No\_EXER\_Rate Pap\_SmearRate ProctoscopRate SMOKER\_Rate

factor FEW\_FruitVeg\_Rate MAJ\_Depress\_Rate MammogrmRate No\_EXER\_Rate Pap\_SmearRate ProctoscopRate SMOKER\_Rate

rotate

corr Obesity\_RateFEW\_FruitVeg\_Rate MAJ\_Depress\_Rate MammogrmRate No\_EXER\_Rate Pap\_SmearRate ProctoscopRate SMOKER\_Rate

corr Obesity\_Rate FEW\_FruitVeg\_Rate MAJ\_Depress\_Rate MammogrmRate No\_EXER\_Rate Pap\_SmearRate ProctoscopRate SMOKER\_Rate

factor Obesity\_Rate FEW\_FruitVeg\_Rate MAJ\_Depress\_Rate MammogrmRate No\_EXER\_Rate Pap\_SmearRate ProctoscopRate SMOKER\_Rate

rotate

corr Obesity\_Rate MAJ\_Depress\_Rate MammogrmRate No\_EXER\_Rate Pap\_SmearRate ProctoscopRate SMOKER\_Rate

factor Obesity\_Rate MAJ\_Depress\_Rate MammogrmRate No\_EXER\_Rate Pap\_SmearRate ProctoscopRate SMOKER\_Rate

rotate

egen PersonalChoice= rowmean (factor Obesity\_Rate MAJ\_Depress\_Rate MammogrmRate No\_EXER\_Rate Pap\_SmearRate ProctoscopRate SMOKER\_Rate)

egen PersonalChoice= rowmean (Obesity\_Rate MAJ\_Depress\_Rate MammogrmRate No\_EXER\_Rate Pap\_SmearRate ProctoscopRate SMOKER\_Rate)

tab PersonalChoice, m

corr EDU\_AveYears MedianInc

corr Obesity\_RateEDU\_AveYears MedianInc

corr Obesity\_Rate EDU\_AveYears MedianInc

factor Obesity\_Rate EDU\_AveYears MedianInc

rotate

4-25

HP Shortage Area – Interesting findings, opposite of what would be intuitive

use "V:\Users\APST670\Desktop\Obesity\_Data(unclean).dta", clear

do "V:\Users\APST670\Desktop\Obesity\_Data\_Cleaning.do"

do "V:\Users\APST670\Desktop\STATA Run First.do"

do "V:\Users\APST670\Desktop\Eval4.do"

use "V:\Users\APST670\Desktop\Obesity\_Data(unclean).dta", clear

do "V:\Users\APST670\Desktop\Obesity\_Data\_Cleaning.do"

corr EDU\_AveYears MedianInc FEW\_FruitVeg\_Rate MammogrmRate No\_EXER\_Rate Pap\_SmearRate ProctoscopRate SMOKER\_Rate

corr Obesity\_Rate EDU\_AveYears MedianInc FEW\_FruitVeg\_Rate MammogrmRate No\_EXER\_Rate Pap\_SmearRate ProctoscopRate SMOKER\_Rate

factor SMOKER\_Rate No\_EXER\_Rate FEW\_FruitVeg\_Rate

rotate

egen NegBehav = rowmean FEW\_FruitVeg\_Rate No\_EXER\_Rate SMOKER\_Rate

egen NegBehav = rowmean (FEW\_FruitVeg\_Rate No\_EXER\_Rate SMOKER\_Rate)

regress Obesity\_Rate NegBehav

twoway ( lfitci Obesity\_Rate NegBehav)

corr Obesity\_Rate EDU\_AveYears MedianInc MammogrmRate Pap\_SmearRate ProctoscopRate

corr Obesity\_Rate MammogrmRate Pap\_SmearRate ProctoscopRate

factor MammogrmRate Pap\_SmearRate ProctoscopRate

rotate

egen HealthProactv = rowmean ( MammogrmRate Pap\_SmearRate ProctoscopRate)

corr Obesity\_Rate EDU\_AveYears MedianInc

corr Obesity\_Rate UNINSURE\_Rate FEW\_FruitVeg\_Rate No\_EXER\_Rate SMOKER\_Rate

factor Obesity\_Rate UNINSURE\_Rate FEW\_FruitVeg\_Rate No\_EXER\_Rate SMOKER\_Rate

rotate

egen NegBehav = rowmean (FEW\_FruitVeg\_Rate No\_EXER\_Rate SMOKER\_Rate UNINSURE\_Rate)

drop NegBehav

egen NegBehav = rowmean (FEW\_FruitVeg\_Rate No\_EXER\_Rate SMOKER\_Rate UNINSURE\_Rate)

regress Obesity\_Rate NegBehav

twoway ( lfitci Obesity\_Rate NegBehav)

l

corr Obesity\_Rate AGE\_roughcompositete EDU\_AveYears MedianInc UNEMPL\_Rate

corr Obesity\_Rate UNEMPL\_Rate SEV\_Work\_Disabl\_Rate Elder\_Medcar\_Rate

corr Obesity\_Rate CenterYN DENT\_Rateper100 HP\_ShortageYN PHYS\_Rateper100

factor CenterYN DENT\_Rateper100 HP\_ShortageYN PHYS\_Rateper100

rotate

egen = rowmean ( CenterYN DENT\_Rateper100 HP\_ShortageYN PHYS\_Rateper100)

egen ServAvail = rowmean ( CenterYN DENT\_Rateper100 HP\_ShortageYN PHYS\_Rateper100)

regress Obesity\_Rate ServAvail

twoway ( lfitci Obesity\_Rate ServAvail)

recode HP\_ShortageYN (1=2) (2=0)

drop ServAvail

corr Obesity\_Rate CenterYN DENT\_Rateper100 HP\_ShortageYN PHYS\_Rateper100

factor CenterYN DENT\_Rateper100 HP\_ShortageYN PHYS\_Rateper100

rotate

egen ServAvail = rowmean ( CenterYN DENT\_Rateper100 HP\_ShortageYN PHYS\_Rateper100)

twoway ( lfitci Obesity\_Rate ServAvail)

recode HP\_ShortageYN (1=0) (2=0)

clear

use "V:\Users\APST670\Desktop\Obesity\_Data(unclean).dta", clear

do "V:\Users\APST670\Desktop\Obesity\_Data\_Cleaning.do"

tab HP\_ShortageNY, m

do "C:\Users\thenriod\AppData\Local\Temp\STD00000000.tmp"

recode hpsa\_ind (1=1) (2=0), gen (HP\_ShortageY0N1)

corr Obesity\_Rate CenterY0N1 DENT\_Rateper100 HP\_ShortageYN PHYS\_Rateper100

corr Obesity\_Rate CenterYN DENT\_Rateper100 HP\_ShortageYN PHYS\_Rateper100

corr Obesity\_Rate CenterYN DENT\_Rateper100 HP\_ShortageY0N1 PHYS\_Rateper100

drop HP\_ShortageY0N1

corr Obesity\_Rate CenterYN DENT\_Rateper100 HP\_ShortageYN PHYS\_Rateper100

recode hpsa\_ind (1=0) (2=1), gen (HP\_ShortageYN)

recode hpsa\_ind (1=1) (2=0), gen (HP\_ShortageY0N1)

corr Obesity\_Rate CenterYN DENT\_Rateper100 HP\_ShortageY0N1 PHYS\_Rateper100

rotate Obesity\_Rate CenterYN DENT\_Rateper100 HP\_ShortageY0N1 PHYS\_Rateper100

corr CenterYN DENT\_Rateper100 HP\_ShortageY0N1 PHYS\_Rateper100

rotate CenterYN DENT\_Rateper100 HP\_ShortageY0N1 PHYS\_Rateper100

factor CenterYN DENT\_Rateper100 HP\_ShortageY0N1 PHYS\_Rateper100

rotate

egen ServAvail = rowmean ( CenterYN DENT\_Rateper100 HP\_ShortageY0N1 PHYS\_Rateper100)

do "C:\Users\thenriod\AppData\Local\Temp\STD00000000.tmp"

corr DIABETES\_Rate Disabl\_Medcar\_Rate Elder\_Medcar\_Rate Hi\_BPRate MAJ\_Depress\_Rate SEV\_Work\_Disabl\_Rate UNEMPL\_Rate)

corr DIABETES\_Rate Disabl\_Medcar\_Rate Elder\_Medcar\_Rate Hi\_BPRate MAJ\_Depress\_Rate SEV\_Work\_Disabl\_Rate UNEMPL\_Rate

factor DIABETES\_Rate Disabl\_Medcar\_Rate Elder\_Medcar\_Rate Hi\_BPRate MAJ\_Depress\_Rate SEV\_Work\_Disabl\_Rate UNEMPL\_Rate

rotate

egen HlthDisadvtg = rowmean ( DIABETES\_Rate Disabl\_Medcar\_Rate Elder\_Medcar\_Rate Hi\_BPRate MAJ\_Depress\_Rate SEV\_Work\_Disabl\_Rate UNEMPL\_Rate)

regress Obesity\_Rate HlthDisadvtg

twoway ( lfitci Obesity\_Rate HlthDisadvtg)

egen HlthDisadvtg = rowmean ( DIABETES\_Rate Disabl\_Medcar\_Rate Elder\_Medcar\_Rate Hi\_BPRate MAJ\_Depress\_Rate SEV\_Work\_Disabl\_Rate UNEMPL\_Rate)

corr AGE\_roughcompositete EDU\_AveYears

corr EDU\_AveYears MedianInc

factor

factor EDU\_AveYears MedianInc

rotate

egen SES = rowmean ( EDU\_AveYears MedianInc)

twoway ( lfitci Obesity\_Rate SES)

clear

use "V:\Users\APST670\Desktop\Obesity\_Data(unclean).dta", clear

do "V:\Users\APST670\Desktop\STATA Run First.do"

do "V:\Users\APST670\Desktop\Eval4.do"

do "C:\Users\thenriod\AppData\Local\Temp\STD00000000.tmp"

use "V:\Users\APST670\Desktop\Obesity\_Data(unclean).dta", clear

do "C:\Users\thenriod\AppData\Local\Temp\STD00000000.tmp"

do "V:\Users\APST670\Desktop\Obesity\_Categorical\_VariableGen.do"

regress Obesity\_Rate HlthDisadvtg

regress Obesity\_Rate HlthDisadvtg

regress Obesity\_Rate NegBehav

regress Obesity\_Rate ProBehav

regress Obesity\_Rate SES

regress Obesity\_Rate ServAvail

twoway ( lfitci Obesity\_Rate SES)

eval4 logit Obesity\_Rate

predict Obesity\_Rate

logit Obesity\_Rate SES

predict probHat2, pr

twoway ( lfitci Obesity\_Rate SES)

eval4 linear Obesity\_Rate "10 15 20 25 30 35 40"

eval4 linear SES "20000 30000 40000 50000 60000"

regress Obesity\_Rate SES

eval4 linear SES "20000 30000 40000 50000 60000"

regress Obesity\_Rate MedianInc EDU\_AveYears

eval4 linear MedianInc "20000 30000 40000 50000 60000"

eval4 linear EDU\_AveYears "10 11 12 13 14 15 16"

4/29