Notes:
1. Unicode is supported; see help unicode_advice.

```
 1 . use "C:\Users\perpe\OneDrive - students.jkuat.ac.ke\Desktop\Peeu\NHANES_dataset for ANALYSI
   > S PROJECT.dta"

 2 . **** Methods

 3 .
 4 . * Merging the depression dataset to NHANES dataset

 5 .
 6 . merge 1:1 seqn using Depression_data

    Result                           # of obs.
    -----------------------------------------
    not matched                         6,595
        from master                     6,595  (_merge==1)
        from using                          0  (_merge==2)

    matched                             8,965  (_merge==3)
    -----------------------------------------

 7 .
 8 .
 9 .
10 . * Keep only observations where age is between 30 and 60

11 .
12 . keep if ridageyr >= 30 & ridageyr <= 60
   (10,896 observations deleted)

13 .
14 .
15 .
16 . * Depression Variable
```

```
17 .
18 . egen Depression = rowtotal(dpq010 dpq020 dpq030 dpq040 dpq050 dpq060 dpq070 dpq080 dpq090),
   >  missing
   (614 missing values generated)

19 .
20 .
21 .
22 . * Recode Depression variable

23 .
24 . recode Depression (0 = 0) (1/max = 1), gen(Depression_recode)
   (2197 differences between Depression and Depression_recode)

25 .
26 . * Label the recoded variable

27 .
28 . label define Depression_label 0 "No" 1 "Yes"

29 .
30 . label values Depression_recode Depression_label

31 .
32 . tabulate Depression_recode
```

RECODE of Depression	Freq.	Percent	Cum.
No	**1,361**	**33.60**	**33.60**
Yes	**2,689**	**66.40**	**100.00**
Total	**4,050**	**100.00**	

```
33 .
34 .
35 .
36 . * Depression

37 .
38 . summarize Depression,detail
```

Depression

Percentiles		Smallest		
1%	**0**	**0**		
5%	**0**	**0**		
10%	**0**	**0**	Obs	**4,050**
25%	**0**	**0**	Sum of Wgt.	**4,050**
50%	**2**		Mean	**3.354074**
		Largest	Std. Dev.	**4.365219**
75%	**5**	**25**		
90%	**9**	**26**	Variance	**19.05514**
95%	**13**	**26**	Skewness	**1.920013**
99%	**20**	**26**	Kurtosis	**7.027943**

```
39 .
40 .
41 .
42 . * Decriptive statistics

43 .
44 . * Gender

45 .
46 . tabulate riagendr
```

Gender	Freq.	Percent	Cum.
1	**2,211**	**47.41**	**47.41**
2	**2,453**	**52.59**	**100.00**
Total	**4,664**	**100.00**	

```
47 .
48 .
49 .
50 . * Age

51 .
52 . * Recode Age variable

53 .
54 . recode ridageyr (min/44 = 0) (45/max = 1), gen(ridageyr_recode)
(4664 differences between ridageyr and ridageyr_recode)

55 .
56 .
57 .
58 . * Label the Age variable

59 .
60 . label define ridageyr_label 0 "Young Population" 1 "Old Population"

61 .
62 . label values ridageyr_recode ridageyr_label

63 .
64 . tabulate ridageyr_recode
```

RECODE of ridageyr (Age in years at screening)	Freq.	Percent	Cum.
Young Population	**2,154**	**46.18**	**46.18**
Old Population	**2,510**	**53.82**	**100.00**
Total	**4,664**	**100.00**	

```
65 .
66 .
67 .
68 . * RIDRETH3 (Race-Ethnicity for oversample of Asian Americans)

69 .
70 . label define race_ethnicity_label 1 "Mexican American" 2 "Hispanic" 3 "Non-Hispanic White"
   > 4 "Non-Hispanic Black" 6 "Non-Hispanic Asian" 7 "Non-Hispanic Multiracial "

71 .
72 . label values ridreth3 race_ethnicity_label

73 .
74 . tabulate ridreth3
```

Race/Hispanic origin w/ NH Asian	Freq.	Percent	Cum.
Mexican American	**595**	**12.76**	**12.76**
Hispanic	**479**	**10.27**	**23.03**
Non-Hispanic White	**1,400**	**30.02**	**53.04**
Non-Hispanic Black	**1,262**	**27.06**	**80.10**
Non-Hispanic Asian	**689**	**14.77**	**94.88**
Non-Hispanic Multiracial	**239**	**5.12**	**100.00**
Total	**4,664**	**100.00**	

```
75 .
76 .
77 .
78 . * DMDEDUC2 (Highest level of education completed)

79 .
80 . label define education_lbll 1 "Less 9th grade" 2 "9-11th grade" 3 "High school graduate/GED
   > " 4 "Some college or AA degree" 5 "College graduate or higher" 7 "Refused" 9 "Don't Know"

81 .
82 . label values dmdeduc2 education_lbll

83 .
84 . * Set Refused and Don't Know responses to missing values

85 .
86 . recode dmdeduc2 (7 9 = .)
(dmdeduc2: 2 changes made)

87 .
88 .
89 .
90 . * indfmpir (Ratio to Family Income)

91 .
92 . gen family_income = .
(4,664 missing values generated)
```

```
 93 .
 94 . replace family_income = 1 if indfmpir < 1
    (771 real changes made)

 95 .
 96 . replace family_income = 2 if indfmpir == 1
    (14 real changes made)

 97 .
 98 . replace family_income = 3 if indfmpir > 1
    (3,879 real changes made)

 99 .
100 . label define family_income_label 1 "< 1" 2 "= 1" 3 "> 1"

101 .
102 . label values family_income family_income_label

103 .
104 . tabulate family_income
```

family_income	Freq.	Percent	Cum.
< 1	**771**	**16.53**	**16.53**
= 1	**14**	**0.30**	**16.83**
> 1	**3,879**	**83.17**	**100.00**
Total	**4,664**	**100.00**	

```
105 .
106 .
107 .
108 . * Main exposure (Length of time lived in the US)

109 .
110 . gen dmdyrusz_filled = .
    (4,664 missing values generated)

111 .
112 . replace dmdyrusz_filled = 1 if dmdyrusz <= 5 & !missing(dmdyrusz)
    (1,534 real changes made)

113 .
114 . replace dmdyrusz_filled = 2 if dmdyrusz > 5
    (3,130 real changes made)

115 .
116 . label define length_US_label 1 "<= 5" 2 "> 5"

117 .
118 . label values dmdyrusz_filled length_US_label
```

```
119 .
120 . tabulate dmdyrusz_filled
```

dmdyrusz_filled	Freq.	Percent	Cum.
<= 5	**1,534**	**32.89**	**32.89**
> 5	**3,130**	**67.11**	**100.00**
Total	**4,664**	**100.00**	

```
121 .
122 .
123 .
124 . * Histograms for continuous variables

125 .
126 . histogram Depression, title("Distribution of Depression in the US") xtitle("Depression") yl
> abel(#10, nogrid) xlabel(#10, nogrid)
(bin=36, start=0, width=.72222222)

127 .
128 .
129 .
130 . *Identifying the Outliers

131 .
132 . extremes Depression, iqr(3)
```

obs:	iqr:	Depres~n
419.	**3.000**	**20**
2445.	**3.000**	**20**
2499.	**3.000**	**20**
4351.	**3.000**	**20**
4489.	**3.000**	**20**
4555.	**3.000**	**20**
4602.	**3.000**	**20**
308.	**3.200**	**21**
994.	**3.200**	**21**
1709.	**3.200**	**21**
1884.	**3.200**	**21**
2246.	**3.200**	**21**
2260.	**3.200**	**21**
2329.	**3.200**	**21**
2662.	**3.200**	**21**
2715.	**3.200**	**21**
3989.	**3.200**	**21**
4042.	**3.200**	**21**
4466.	**3.200**	**21**
115.	**3.400**	**22**
891.	**3.400**	**22**
986.	**3.400**	**22**
1129.	**3.400**	**22**
1522.	**3.400**	**22**
2376.	**3.400**	**22**
2614.	**3.400**	**22**
3232.	**3.400**	**22**
4556.	**3.400**	**22**

317.	3.600	23
1638.	3.600	23
1807.	3.600	23
2935.	3.600	23
2988.	3.600	23
4158.	3.600	23
4294.	3.600	23
1157.	3.800	24
2918.	3.800	24
4161.	4.000	25
4233.	4.000	25
357.	4.200	26
2850.	4.200	26
4554.	4.200	26

```
133 .
134 .
135 .
136 . * Replace the outliers

137 .
138 . winsor2 Depression, replace cut(0,90)

139 .
140 .
141 .
142 .
143 .
144 . * Drop the outliers

145 .
146 . winsor2 Depression, replace cut(0,90)

147 .
148 .
149 .
150 . * Find Missing values

151 .
152 . mdesc
```

Variable	Missing	Total	Percent Missing
seqn	0	4,664	0.00
ridstatr	0	4,664	0.00
riagendr	0	4,664	0.00
ridageyr	0	4,664	0.00
ridreth1	0	4,664	0.00
ridreth3	0	4,664	0.00
dmdborn4	0	4,664	0.00
dmdyrusz	3,058	4,664	65.57
dmdeduc2	2	4,664	0.04
dmdmartz	0	4,664	0.00
indfmpir	690	4,664	14.79
alq111	605	4,664	12.97
alq121	961	4,664	20.60
alq130	1,620	4,664	34.73
alq142	1,620	4,664	34.73
bmxbmi	368	4,664	7.89
bpxosy1	765	4,664	16.40
bpxodi1	765	4,664	16.40

smq020	0	4,664	0.00
smq040	2,785	4,664	59.71
hiq011	0	4,664	0.00
bpq020	0	4,664	0.00
dpq010	614	4,664	13.16
dpq020	615	4,664	13.19
dpq030	615	4,664	13.19
dpq040	615	4,664	13.19
dpq050	615	4,664	13.19
dpq060	616	4,664	13.21
dpq070	616	4,664	13.21
dpq080	616	4,664	13.21
dpq090	617	4,664	13.23
_merge	0	4,664	0.00
Depression	614	4,664	13.16
Depression~e	614	4,664	13.16
ridageyr_r~e	0	4,664	0.00
family_inc~e	0	4,664	0.00
dmdyrusz_f~d	0	4,664	0.00

```
153 .
154 .
155 .
156 . * Exclude missing values from the analysis

157 .
158 . tabulate dmdeduc2 if dmdeduc2 < .
```

Education level - Adults 20+	Freq.	Percent	Cum.
Less 9th grade	331	7.10	7.10
9-11th grade	506	10.85	17.95
High school graduate/GED	1,028	22.05	40.00
Some college or AA degree	1,525	32.71	72.72
College graduate or higher	1,272	27.28	100.00
Total	4,662	100.00	

```
159 .
160 .
161 .
162 . * EXPOSURE (LENGTH OF TIME IN THE US)

163 .
164 . tabulate dmdyrusz_filled riagendr
```

dmdyrusz_filled	Gender 1	2	Total
<= 5	700	834	1,534
> 5	1,511	1,619	3,130
Total	2,211	2,453	4,664

```
165 .
166 . tabulate dmdyrusz_filled riagendr, cell nofreq
```

dmdyrusz_filled	Gender 1	2	Total
<= 5	**15.01**	**17.88**	**32.89**
> 5	**32.40**	**34.71**	**67.11**
Total	**47.41**	**52.59**	**100.00**

```
167 .
168 . tabulate dmdyrusz_filled riagendr, chi2
```

dmdyrusz_filled	Gender 1	2	Total
<= 5	**700**	**834**	**1,534**
> 5	**1,511**	**1,619**	**3,130**
Total	**2,211**	**2,453**	**4,664**

Pearson chi2(**1**) = **2.8830** Pr = **0.090**

```
169 .
170 .
171 .
172 . tabulate dmdyrusz_filled ridageyr_recode
```

dmdyrusz_filled	RECODE of ridageyr (Age in years at screening) Young Pop	Old Popul	Total
<= 5	**695**	**839**	**1,534**
> 5	**1,459**	**1,671**	**3,130**
Total	**2,154**	**2,510**	**4,664**

```
173 .
174 . tabulate dmdyrusz_filled ridageyr_recode, cell nofreq
```

dmdyrusz_filled	RECODE of ridageyr (Age in years at screening) Young Pop	Old Popul	Total
<= 5	**14.90**	**17.99**	**32.89**
> 5	**31.28**	**35.83**	**67.11**
Total	**46.18**	**53.82**	**100.00**

```
175 .
176 . tabulate dmdyrusz_filled ridageyr_recode, chi2
```

dmdyrusz_filled	RECODE of ridageyr (Age in years at screening)		
	Young Pop	Old Popul	Total
<= 5	**695**	**839**	**1,534**
> 5	**1,459**	**1,671**	**3,130**
Total	**2,154**	**2,510**	**4,664**

Pearson chi2(**1**) = **0.7076** Pr = **0.400**

```
177 .
178 .
179 .
180 . tabulate dmdyrusz_filled ridreth3
```

dmdyrusz_filled	Race/Hispanic origin w/ NH Asian						
	Mexican A	Hispanic	Non-Hispa	Non-Hispa	Non-Hispa	Non-Hispa	Total
<= 5	**367**	**307**	**75**	**128**	**631**	**26**	**1,534**
> 5	**228**	**172**	**1,325**	**1,134**	**58**	**213**	**3,130**
Total	**595**	**479**	**1,400**	**1,262**	**689**	**239**	**4,664**

```
181 .
182 . tabulate dmdyrusz_filled ridreth3, cell nofreq
```

dmdyrusz_filled	Race/Hispanic origin w/ NH Asian						
	Mexican A	Hispanic	Non-Hispa	Non-Hispa	Non-Hispa	Non-Hispa	Total
<= 5	**7.87**	**6.58**	**1.61**	**2.74**	**13.53**	**0.56**	**32.89**
> 5	**4.89**	**3.69**	**28.41**	**24.31**	**1.24**	**4.57**	**67.11**
Total	**12.76**	**10.27**	**30.02**	**27.06**	**14.77**	**5.12**	**100.00**

```
183 .
184 . tabulate dmdyrusz_filled ridreth3, chi2
```

dmdyrusz_filled	Race/Hispanic origin w/ NH Asian						
	Mexican A	Hispanic	Non-Hispa	Non-Hispa	Non-Hispa	Non-Hispa	Total
<= 5	**367**	**307**	**75**	**128**	**631**	**26**	**1,534**
> 5	**228**	**172**	**1,325**	**1,134**	**58**	**213**	**3,130**
Total	**595**	**479**	**1,400**	**1,262**	**689**	**239**	**4,664**

Pearson chi2(**5**) = **2.3e+03** Pr = **0.000**

```
185 .
186 .
187 .
188 . tabulate dmdyrusz_filled dmdeduc2
```

dmdyrusz_filled	Education level - Adults 20+					
	Less 9th	9-11th gr	High scho	Some coll	College g	Total
<= 5	**261**	**192**	**257**	**288**	**535**	**1,533**
> 5	**70**	**314**	**771**	**1,237**	**737**	**3,129**
Total	**331**	**506**	**1,028**	**1,525**	**1,272**	**4,662**

```
189 .
190 . tabulate dmdyrusz_filled dmdeduc2, cell nofreq
```

dmdyrusz_filled	Education level - Adults 20+					
	Less 9th	9-11th gr	High scho	Some coll	College g	Total
<= 5	**5.60**	**4.12**	**5.51**	**6.18**	**11.48**	**32.88**
> 5	**1.50**	**6.74**	**16.54**	**26.53**	**15.81**	**67.12**
Total	**7.10**	**10.85**	**22.05**	**32.71**	**27.28**	**100.00**

```
191 .
192 . tabulate dmdyrusz_filled dmdeduc2, chi2
```

dmdyrusz_filled	Education level - Adults 20+					
	Less 9th	9-11th gr	High scho	Some coll	College g	Total
<= 5	**261**	**192**	**257**	**288**	**535**	**1,533**
> 5	**70**	**314**	**771**	**1,237**	**737**	**3,129**
Total	**331**	**506**	**1,028**	**1,525**	**1,272**	**4,662**

Pearson chi2(**4**) = **535.6670** Pr = **0.000**

```
193 .
194 .
195 .
196 . tabulate dmdyrusz_filled family_income
```

dmdyrusz_filled	family_income			
	< 1	= 1	> 1	Total
<= 5	**236**	**3**	**1,295**	**1,534**
> 5	**535**	**11**	**2,584**	**3,130**
Total	**771**	**14**	**3,879**	**4,664**

```
197 .
198 . tabulate dmdyrusz_filled family_income, cell nofreq
```

dmdyrusz_filled	family_income < 1	= 1	> 1	Total
<= 5	**5.06**	**0.06**	**27.77**	**32.89**
> 5	**11.47**	**0.24**	**55.40**	**67.11**
Total	**16.53**	**0.30**	**83.17**	**100.00**

```
199 .
200 . tabulate dmdyrusz_filled family_income, chi2
```

dmdyrusz_filled	family_income < 1	= 1	> 1	Total
<= 5	**236**	**3**	**1,295**	**1,534**
> 5	**535**	**11**	**2,584**	**3,130**
Total	**771**	**14**	**3,879**	**4,664**

Pearson chi2(**2**) = **3.0801** Pr = **0.214**

```
201 .
202 .
203 .
204 . * Conduct ANOVA

205 .
206 . anova Depression dmdyrusz_filled
```

Number of obs = **4,050** R-squared = **0.0170**
Root MSE = **3.07197** Adj R-squared = **0.0168**

Source	Partial SS	df	MS	F	Prob>F
Model	**662.04531**	**1**	**662.04531**	**70.15**	**0.0000**
dmdyrusz_~d	**662.04531**	**1**	**662.04531**	**70.15**	**0.0000**
Residual	**38200.957**	**4,048**	**9.4369954**		
Total	**38863.003**	**4,049**	**9.5981731**		

```
207 .
208 .
209 .
210 . tabulate Depression_recode dmdyrusz_filled
```

RECODE of Depression	dmdyrusz_filled <= 5	> 5	Total
No	**552**	**809**	**1,361**
Yes	**736**	**1,953**	**2,689**
Total	**1,288**	**2,762**	**4,050**

```
211 .
212 . tabulate Depression_recode dmdyrusz_filled, cell nofreq
```

RECODE of Depression	dmdyrusz_filled <= 5	> 5	Total
No	**13.63**	**19.98**	**33.60**
Yes	**18.17**	**48.22**	**66.40**
Total	**31.80**	**68.20**	**100.00**

```
213 .
214 . tabulate Depression_recode dmdyrusz_filled, chi2
```

RECODE of Depression	dmdyrusz_filled <= 5	> 5	Total
No	**552**	**809**	**1,361**
Yes	**736**	**1,953**	**2,689**
Total	**1,288**	**2,762**	**4,050**

Pearson chi2(**1**) = **72.4601** Pr = **0.000**

```
215 .
216 .
217 .
218 .
219 .
220 . * OUTCOME (DEPRESSION) CATEGORICAL

221 .
222 . tabulate Depression_recode riagendr
```

RECODE of Depression	Gender 1	2	Total
No	**754**	**607**	**1,361**
Yes	**1,174**	**1,515**	**2,689**
Total	**1,928**	**2,122**	**4,050**

```
223 .
224 . tabulate Depression_recode riagendr, cell nofreq
```

RECODE of Depression	Gender 1	2	Total
No	**18.62**	**14.99**	**33.60**
Yes	**28.99**	**37.41**	**66.40**
Total	**47.60**	**52.40**	**100.00**

```
225 .
226 . tabulate Depression_recode riagendr, chi2
```

RECODE of Depression	Gender 1	2	Total
No	**754**	**607**	**1,361**
Yes	**1,174**	**1,515**	**2,689**
Total	**1,928**	**2,122**	**4,050**

Pearson chi2(**1**) = **49.9423** Pr = **0.000**

```
227 .
228 .
229 .
230 . tabulate Depression_recode ridageyr_recode
```

RECODE of Depression	RECODE of ridageyr (Age in years at screening) Young Pop	Old Popul	Total
No	**586**	**775**	**1,361**
Yes	**1,259**	**1,430**	**2,689**
Total	**1,845**	**2,205**	**4,050**

```
231 .
232 . tabulate Depression_recode ridageyr_recode, cell nofreq
```

RECODE of Depression	RECODE of ridageyr (Age in years at screening) Young Pop	Old Popul	Total
No	**14.47**	**19.14**	**33.60**
Yes	**31.09**	**35.31**	**66.40**
Total	**45.56**	**54.44**	**100.00**

```
233 .
234 . tabulate Depression_recode ridageyr_recode, chi2
```

RECODE of Depression	RECODE of ridageyr (Age in years at screening) Young Pop	Old Popul	Total
No	**586**	**775**	**1,361**
Yes	**1,259**	**1,430**	**2,689**
Total	**1,845**	**2,205**	**4,050**

Pearson chi2(**1**) = **5.1612** Pr = **0.023**

```
235 .
236 .
237 .
238 . tabulate Depression_recode ridreth3
```

RECODE of Depression	Race/Hispanic origin w/ NH Asian: Mexican A	Hispanic	Non-Hispa	Non-Hispa	Non-Hispa	Non-Hispa	Total
No	**188**	**142**	**359**	**375**	**251**	**46**	**1,361**
Yes	**333**	**285**	**895**	**716**	**297**	**163**	**2,689**
Total	**521**	**427**	**1,254**	**1,091**	**548**	**209**	**4,050**

```
239 .
240 . tabulate Depression_recode ridreth3, cell nofreq
```

RECODE of Depression	Race/Hispanic origin w/ NH Asian: Mexican A	Hispanic	Non-Hispa	Non-Hispa	Non-Hispa	Non-Hispa	Total
No	**4.64**	**3.51**	**8.86**	**9.26**	**6.20**	**1.14**	**33.60**
Yes	**8.22**	**7.04**	**22.10**	**17.68**	**7.33**	**4.02**	**66.40**
Total	**12.86**	**10.54**	**30.96**	**26.94**	**13.53**	**5.16**	**100.00**

```
241 .
242 . tabulate Depression_recode ridreth3, chi2
```

RECODE of Depression	Race/Hispanic origin w/ NH Asian: Mexican A	Hispanic	Non-Hispa	Non-Hispa	Non-Hispa	Non-Hispa	Total
No	**188**	**142**	**359**	**375**	**251**	**46**	**1,361**
Yes	**333**	**285**	**895**	**716**	**297**	**163**	**2,689**
Total	**521**	**427**	**1,254**	**1,091**	**548**	**209**	**4,050**

Pearson chi2(**5**) = **64.8045** Pr = **0.000**

```
243 .
244 .
245 .
246 . tabulate Depression_recode dmdeduc2
```

RECODE of Depression	Education level - Adults 20+: Less 9th	9-11th gr	High scho	Some coll	College g	Total
No	**109**	**131**	**302**	**400**	**419**	**1,361**
Yes	**157**	**293**	**596**	**961**	**681**	**2,688**
Total	**266**	**424**	**898**	**1,361**	**1,100**	**4,049**

```
247 .
248 . tabulate Depression_recode dmdeduc2, cell nofreq
```

RECODE of Depression	Education level - Adults 20+					
	Less 9th	9-11th gr	High scho	Some coll	College g	Total
No	**2.69**	**3.24**	**7.46**	**9.88**	**10.35**	**33.61**
Yes	**3.88**	**7.24**	**14.72**	**23.73**	**16.82**	**66.39**
Total	**6.57**	**10.47**	**22.18**	**33.61**	**27.17**	**100.00**

```
249 .
250 . tabulate Depression_recode dmdeduc2, chi2
```

RECODE of Depression	Education level - Adults 20+					
	Less 9th	9-11th gr	High scho	Some coll	College g	Total
No	**109**	**131**	**302**	**400**	**419**	**1,361**
Yes	**157**	**293**	**596**	**961**	**681**	**2,688**
Total	**266**	**424**	**898**	**1,361**	**1,100**	**4,049**

Pearson chi2(**4**) = **28.6282** Pr = **0.000**

```
251 .
252 .
253 .
254 . tabulate Depression_recode family_income
```

RECODE of Depression	family_income			
	< 1	= 1	> 1	Total
No	**182**	**2**	**1,177**	**1,361**
Yes	**497**	**12**	**2,180**	**2,689**
Total	**679**	**14**	**3,357**	**4,050**

```
255 .
256 . tabulate Depression_recode family_income, cell nofreq
```

RECODE of Depression	family_income			
	< 1	= 1	> 1	Total
No	**4.49**	**0.05**	**29.06**	**33.60**
Yes	**12.27**	**0.30**	**53.83**	**66.40**
Total	**16.77**	**0.35**	**82.89**	**100.00**

```
257 .
258 . tabulate Depression_recode family_income, chi2
```

RECODE of Depression	family_income			
	< 1	= 1	> 1	Total
No	**182**	**2**	**1,177**	**1,361**
Yes	**497**	**12**	**2,180**	**2,689**
Total	**679**	**14**	**3,357**	**4,050**

Pearson chi2(**2**) = **19.6072** Pr = **0.000**

```
259 .
260 .
261 .
262 .
263 .
264 . * OUTCOME (DEPRESSION) Continuous

265 .
266 . tabulate riagendr
```

Gender	Freq.	Percent	Cum.
1	**2,211**	**47.41**	**47.41**
2	**2,453**	**52.59**	**100.00**
Total	**4,664**	**100.00**	

```
267 .
268 . summarize Depression riagendr
```

Variable	Obs	Mean	Std. Dev.	Min	Max
Depression	**4,050**	**2.872346**	**3.098092**	**0**	**9**
riagendr	**4,664**	**1.525943**	**.49938**	**1**	**2**

```
269 .
270 . summarize Depression riagendr, detail
```

Depression

Percentiles		Smallest		
1%	**0**	**0**		
5%	**0**	**0**		
10%	**0**	**0**	Obs	**4,050**
25%	**0**	**0**	Sum of Wgt.	**4,050**
50%	**2**		Mean	**2.872346**
		Largest	Std. Dev.	**3.098092**
75%	**5**	**9**		
90%	**9**	**9**	Variance	**9.598173**
95%	**9**	**9**	Skewness	**.8594211**
99%	**9**	**9**	Kurtosis	**2.402751**

Gender

Percentiles		Smallest		
1%	**1**	**1**		
5%	**1**	**1**		
10%	**1**	**1**	Obs	**4,664**
25%	**1**	**1**	Sum of Wgt.	**4,664**
50%	**2**		Mean	**1.525943**
		Largest	Std. Dev.	**.49938**
75%	**2**	**2**		
90%	**2**	**2**	Variance	**.2493804**
95%	**2**	**2**	Skewness	**-.1039136**
99%	**2**	**2**	Kurtosis	**1.010798**

```
271 .
272 . anova Depression riagendr
```

Number of obs =	**4,050**	R-squared =	**0.0147**
Root MSE =	**3.07554**	Adj R-squared =	**0.0145**

Source	Partial SS	df	MS	F	Prob>F
Model	**573.10971**	**1**	**573.10971**	**60.59**	**0.0000**
riagendr	**573.10971**	**1**	**573.10971**	**60.59**	**0.0000**
Residual	**38289.893**	**4,048**	**9.4589657**		
Total	**38863.003**	**4,049**	**9.5981731**		

```
273 .
274 .
275 .
276 . tabulate ridageyr_recode
```

RECODE of ridageyr (Age in years at screening)	Freq.	Percent	Cum.
Young Population	**2,154**	**46.18**	**46.18**
Old Population	**2,510**	**53.82**	**100.00**
Total	**4,664**	**100.00**	

```
277 .
278 . summarize Depression ridageyr_recode
```

Variable	Obs	Mean	Std. Dev.	Min	Max
Depression	**4,050**	**2.872346**	**3.098092**	**0**	**9**
ridageyr_r~e	**4,664**	**.5381647**	**.4985948**	**0**	**1**

```
279 .
280 . summarize Depression ridageyr_recode, detail
```

Depression

Percentiles		Smallest		
1%	**0**	**0**		
5%	**0**	**0**		
10%	**0**	**0**	Obs	**4,050**
25%	**0**	**0**	Sum of Wgt.	**4,050**
50%	**2**		Mean	**2.872346**
		Largest	Std. Dev.	**3.098092**
75%	**5**	**9**		
90%	**9**	**9**	Variance	**9.598173**
95%	**9**	**9**	Skewness	**.8594211**
99%	**9**	**9**	Kurtosis	**2.402751**

RECODE of ridageyr (Age in years at screening)

Percentiles		Smallest		
1%	**0**	**0**		
5%	**0**	**0**		
10%	**0**	**0**	Obs	**4,664**
25%	**0**	**0**	Sum of Wgt.	**4,664**

50%	1		Mean	.5381647
		Largest	Std. Dev.	.4985948
75%	1	1		
90%	1	1	Variance	.2485968
95%	1	1	Skewness	-.1531053
99%	1	1	Kurtosis	1.023441

```
281 .
282 . anova Depression ridageyr_recode
```

Number of obs =	4,050	R-squared =	0.0001
Root MSE =	3.09825	Adj R-squared =	-0.0001

Source	Partial SS	df	MS	F	Prob>F
Model	5.52209	1	5.52209	0.58	0.4482
ridageyr_~e	5.52209	1	5.52209	0.58	0.4482
Residual	38857.481	4,048	9.59918		
Total	38863.003	4,049	9.5981731		

```
283 .
284 .
285 .
286 . tabulate ridreth3
```

Race/Hispanic origin w/ NH Asian	Freq.	Percent	Cum.
Mexican American	595	12.76	12.76
Hispanic	479	10.27	23.03
Non-Hispanic White	1,400	30.02	53.04
Non-Hispanic Black	1,262	27.06	80.10
Non-Hispanic Asian	689	14.77	94.88
Non-Hispanic Multiracial	239	5.12	100.00
Total	4,664	100.00	

```
287 .
288 . summarize Depression ridreth3
```

Variable	Obs	Mean	Std. Dev.	Min	Max
Depression	4,050	2.872346	3.098092	0	9
ridreth3	4,664	3.560892	1.64894	1	7

```
289 .
290 . summarize Depression ridreth3, detail
```

Depression

Percentiles		Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	4,050
25%	0	0	Sum of Wgt.	4,050

50%	**2**		Mean	**2.872346**
		Largest	Std. Dev.	**3.098092**
75%	**5**	**9**		
90%	**9**	**9**	Variance	**9.598173**
95%	**9**	**9**	Skewness	**.8594211**
99%	**9**	**9**	Kurtosis	**2.402751**

Race/Hispanic origin w/ NH Asian

Percentiles		Smallest		
1%	**1**	**1**		
5%	**1**	**1**		
10%	**1**	**1**	Obs	**4,664**
25%	**3**	**1**	Sum of Wgt.	**4,664**
50%	**3**		Mean	**3.560892**
		Largest	Std. Dev.	**1.64894**
75%	**4**	**7**		
90%	**6**	**7**	Variance	**2.719002**
95%	**7**	**7**	Skewness	**.3714498**
99%	**7**	**7**	Kurtosis	**2.50795**

```
291 .
292 . anova Depression ridreth3
```

Number of obs =	**4,050**	R-squared =	**0.0241**
Root MSE =	**3.06242**	Adj R-squared =	**0.0229**

Source	Partial SS	df	MS	F	Prob>F
Model	**936.56981**	**5**	**187.31396**	**19.97**	**0.0000**
ridreth3	**936.56981**	**5**	**187.31396**	**19.97**	**0.0000**
Residual	**37926.433**	**4,044**	**9.3784453**		
Total	**38863.003**	**4,049**	**9.5981731**		

```
293 .
294 .
295 .
296 . tabulate dmdeduc2
```

Education level - Adults 20+	Freq.	Percent	Cum.
Less 9th grade	**331**	**7.10**	**7.10**
9-11th grade	**506**	**10.85**	**17.95**
High school graduate/GED	**1,028**	**22.05**	**40.00**
Some college or AA degree	**1,525**	**32.71**	**72.72**
College graduate or higher	**1,272**	**27.28**	**100.00**
Total	**4,662**	**100.00**	

```
297 .
298 . summarize Depression dmdeduc2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
Depression	**4,050**	**2.872346**	**3.098092**	**0**	**9**
dmdeduc2	**4,662**	**3.622265**	**1.193364**	**1**	**5**

```
299 .
300 . summarize Depression dmdeduc2, detail
```

Depression

Percentiles	Smallest		
1%	**0**	**0**	
5%	**0**	**0**	
10%	**0**	**0**	Obs **4,050**
25%	**0**	**0**	Sum of Wgt. **4,050**
50%	**2**		Mean **2.872346**
	Largest		Std. Dev. **3.098092**
75%	**5**	**9**	
90%	**9**	**9**	Variance **9.598173**
95%	**9**	**9**	Skewness **.8594211**
99%	**9**	**9**	Kurtosis **2.402751**

Education level - Adults 20+

Percentiles	Smallest		
1%	**1**	**1**	
5%	**1**	**1**	
10%	**2**	**1**	Obs **4,662**
25%	**3**	**1**	Sum of Wgt. **4,662**
50%	**4**		Mean **3.622265**
	Largest		Std. Dev. **1.193364**
75%	**5**	**5**	
90%	**5**	**5**	Variance **1.424117**
95%	**5**	**5**	Skewness **-.6272024**
99%	**5**	**5**	Kurtosis **2.531322**

```
301 .
302 . anova Depression dmdeduc2
```

Number of obs = **4,049** R-squared = **0.0229**
Root MSE = **3.06431** Adj R-squared = **0.0219**

Source	Partial SS	df	MS	F	Prob>F
Model	**888.71172**	**4**	**222.17793**	**23.66**	**0.0000**
dmdeduc2	**888.71172**	**4**	**222.17793**	**23.66**	**0.0000**
Residual	**37973.019**	**4,044**	**9.3899652**		
Total	**38861.731**	**4,048**	**9.6002299**		

```
303 .
304 .
305 .
306 . tabulate family_income
```

family_income	Freq.	Percent	Cum.
< 1	771	16.53	16.53
= 1	14	0.30	16.83
> 1	3,879	83.17	100.00
Total	4,664	100.00	

```
307 .
308 . summarize Depression family_income
```

Variable	Obs	Mean	Std. Dev.	Min	Max
Depression	4,050	2.872346	3.098092	0	9
family_inc~e	4,664	2.666381	.7436757	1	3

```
309 .
310 . summarize Depression family_income, detail
```

Depression

Percentiles		Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	4,050
25%	0	0	Sum of Wgt.	4,050
50%	2		Mean	2.872346
		Largest	Std. Dev.	3.098092
75%	5	9		
90%	9	9	Variance	9.598173
95%	9	9	Skewness	.8594211
99%	9	9	Kurtosis	2.402751

family_income

Percentiles		Smallest		
1%	1	1		
5%	1	1		
10%	1	1	Obs	4,664
25%	3	1	Sum of Wgt.	4,664
50%	3		Mean	2.666381
		Largest	Std. Dev.	.7436757
75%	3	3		
90%	3	3	Variance	.5530535
95%	3	3	Skewness	-1.787455
99%	3	3	Kurtosis	4.204759

```
311 .
312 . anova Depression family_income
```

Number of obs =	**4,050**	R-squared =	**0.0256**
Root MSE =	**3.05898**	Adj R-squared =	**0.0251**

Source	Partial SS	df	MS	F	Prob>F
Model	**993.72863**	**2**	**496.86432**	**53.10**	**0.0000**
family_in~e	**993.72863**	**2**	**496.86432**	**53.10**	**0.0000**
Residual	**37869.274**	**4,047**	**9.3573694**		
Total	**38863.003**	**4,049**	**9.5981731**		

```
313 .
314 .
315 .
316 .
317 .
318 . *Linear Regression

319 .
320 . * 1. EXAMINE THE CONTINOUS DEPRESSION SCORE OUTCOME - This is the dependent variable for th
    > e linear regression analysis */

321 .
322 .
323 .
324 . regress Depression i.dmdyrusz_filled
```

Source	SS	df	MS		
Model	**662.045311**	**1**	**662.045311**	Number of obs =	**4,050**
Residual	**38200.9574**	**4,048**	**9.43699541**	F(1, 4048) =	**70.15**
Total	**38863.0027**	**4,049**	**9.59817306**	Prob > F =	**0.0000**
				R-squared =	**0.0170**
				Adj R-squared =	**0.0168**
				Root MSE =	**3.072**

Depression	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
dmdyrusz_filled						
> 5	**.8681637**	**.1036513**	**8.38**	**0.000**	**.6649501**	**1.071377**
_cons	**2.28028**	**.0855971**	**26.64**	**0.000**	**2.112462**	**2.448097**

```
325 .
326 . regress Depression i.ridageyr_recode
```

Source	SS	df	MS		
Model	**5.52208998**	**1**	**5.52208998**	Number of obs =	**4,050**
Residual	**38857.4806**	**4,048**	**9.59918**	F(1, 4048) =	**0.58**
Total	**38863.0027**	**4,049**	**9.59817306**	Prob > F =	**0.4482**
				R-squared =	**0.0001**
				Adj R-squared =	**-0.0001**
				Root MSE =	**3.0983**

Depression	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ridageyr_recode						
Old Population	**.0741441**	**.0977557**	**0.76**	**0.448**	**-.1175108**	**.2657991**
_cons	**2.831978**	**.0721305**	**39.26**	**0.000**	**2.690563**	**2.973394**

```
327 .
328 . regress Depression i.dmdeduc2
```

Source	SS	df	MS		
Model	888.711724	4	222.177931	Number of obs	4,049
Residual	37973.0191	4,044	9.38996515	F(4, 4044)	23.66
				Prob > F	0.0000
				R-squared	0.0229
				Adj R-squared	0.0219
Total	38861.7308	4,048	9.60022994	Root MSE	3.0643

Depression	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
dmdeduc2						
9-11th grade	.2464357	.2396805	1.03	0.304	-.2234702	.7163415
High school graduate/GED	-.172656	.2139092	-0.81	0.420	-.5920359	.2467238
Some college or AA degree	-.0290918	.2054262	-0.14	0.887	-.4318404	.3736567
College graduate or hig..	-1.053178	.2093728	-5.03	0.000	-1.463664	-.6426925
_cons	3.180451	.1878846	16.93	0.000	2.812094	3.548808

```
329 .
330 . regress Depression i.family_income
```

Source	SS	df	MS		
Model	993.728634	2	496.864317	Number of obs	4,050
Residual	37869.2741	4,047	9.35736943	F(2, 4047)	53.10
				Prob > F	0.0000
				R-squared	0.0256
				Adj R-squared	0.0251
Total	38863.0027	4,049	9.59817306	Root MSE	3.059

Depression	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
family_income						
= 1	-.107511	.8259326	-0.13	0.896	-1.726793	1.511771
> 1	-1.31735	.1287188	-10.23	0.000	-1.569709	-1.06499
_cons	3.964654	.1173929	33.77	0.000	3.734499	4.194809

```
331 .
332 .
333 .
334 . * Full Model 1

335 .
336 . regress Depression i.dmdyrusz_filled ridageyr_recode i.family_income ib3.dmdeduc2
```

Source	SS	df	MS		
Model	2125.34163	8	265.667703	Number of obs	4,049
Residual	36736.3892	4,040	9.09316564	F(8, 4040)	29.22
				Prob > F	0.0000
				R-squared	0.0547
				Adj R-squared	0.0528
Total	38861.7308	4,048	9.60022994	Root MSE	3.0155

Depression	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
dmdyrusz_filled						
> 5	.8424605	.1083824	7.77	0.000	.6299712	1.05495
ridageyr_recode	.0764126	.0953909	0.80	0.423	-.1106062	.2634313
family_income						
= 1	-.1665716	.8145379	-0.20	0.838	-1.763515	1.430372
> 1	-1.102758	.1320736	-8.35	0.000	-1.361695	-.8438212
dmdeduc2						
Less 9th grade	.5709135	.2194051	2.60	0.009	.1407585	1.001069
9-11th grade	.4267742	.1789001	2.39	0.017	.0760312	.7775171
Some college or AA degree	.2196018	.1306554	1.68	0.093	-.0365548	.4757584
College graduate or hig..	-.5090834	.1395289	-3.65	0.000	-.782637	-.2355299
_cons	3.15316	.1744978	18.07	0.000	2.811048	3.495272

```
337 .
338 .
339 .
340 . * Full Model 2

341 .
342 . regress Depression i.dmdeduc2 i.ridageyr_recode i.dmdyrusz_filled##i.family_income
```

Source	SS	df	MS		
Model	2167.98016	10	216.798016	Number of obs	= 4,049
Residual	36693.7506	4,038	9.08711011	F(10, 4038)	= 23.86
Total	38861.7308	4,048	9.60022994	Prob > F	= 0.0000
				R-squared	= 0.0558
				Adj R-squared	= 0.0534
				Root MSE	= 3.0145

Depression	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
dmdeduc2						
9-11th grade	-.1800455	.2410725	-0.75	0.455	-.6526805	.2925895
High school graduate/GED	-.6084253	.2200662	-2.76	0.006	-1.039876	-.176974
Some college or AA degree	-.3839998	.2150125	-1.79	0.074	-.8055428	.0375433
College graduate or hig..	-1.118723	.2143743	-5.22	0.000	-1.539015	-.6984309
ridageyr_recode						
Old Population	.066512	.0954706	0.70	0.486	-.120663	.253687
dmdyrusz_filled						
> 5	1.268569	.2579711	4.92	0.000	.7628038	1.774335
family_income						
= 1	-1.630203	1.753587	-0.93	0.353	-5.068201	1.807795
> 1	-.7442989	.2347445	-3.17	0.002	-1.204528	-.2840702
dmdyrusz_filled# family_income						
> 5#= 1	1.818953	1.980452	0.92	0.358	-2.063825	5.701732
> 5#> 1	-.5111128	.2772492	-1.84	0.065	-1.054674	.0324486
_cons	3.462714	.2592372	13.36	0.000	2.954466	3.970962

```
343 .
344 .
345 .
346 .
347 .
348 . *Logistic Regression

349 .
350 . * QUESTION 1: What is the prevalence of depression in your sample? */

351 .
352 . logistic Depression_recode i.dmdyrusz_filled
```

Logistic regression — Number of obs = **4,050**; LR chi2(**1**) = **71.17**; Prob > chi2 = **0.0000**; Log likelihood = **-2549.8531**; Pseudo R2 = **0.0138**

Depression_recode	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
dmdyrusz_filled						
> 5	**1.810569**	**.1269778**	**8.46**	**0.000**	**1.578044**	**2.077356**
_cons	**1.333333**	**.0750738**	**5.11**	**0.000**	**1.19402**	**1.488901**

Note: **_cons** estimates baseline odds.

```
353 .
354 . logistic Depression_recode i.ridageyr_recode
```

Logistic regression — Number of obs = **4,050**; LR chi2(**1**) = **5.17**; Prob > chi2 = **0.0230**; Log likelihood = **-2582.8546**; Pseudo R2 = **0.0010**

Depression_recode	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ridageyr_recode						
Old Population	**.8588282**	**.0575504**	**-2.27**	**0.023**	**.7531248**	**.9793673**
_cons	**2.148464**	**.1074396**	**15.29**	**0.000**	**1.947877**	**2.369707**

Note: **_cons** estimates baseline odds.

```
355 .
356 . logistic Depression_recode i.dmdeduc2
```

Logistic regression — Number of obs = **4,049**; LR chi2(**4**) = **28.50**; Prob > chi2 = **0.0000**; Log likelihood = **-2570.7793**; Pseudo R2 = **0.0055**

Depression_recode	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
dmdeduc2						
9-11th grade	**1.552827**	**.2532123**	**2.70**	**0.007**	**1.128033**	**2.137591**
High school graduate/GED	**1.370144**	**.1963319**	**2.20**	**0.028**	**1.034654**	**1.814418**
Some college or AA degree	**1.667978**	**.2304245**	**3.70**	**0.000**	**1.272331**	**2.186655**
College graduate or hig..	**1.128392**	**.1571619**	**0.87**	**0.386**	**.8588255**	**1.482569**
_cons	**1.440367**	**.1795771**	**2.93**	**0.003**	**1.128106**	**1.839062**

Note: **_cons** estimates baseline odds.

```
357 .
358 . logistic Depression_recode i.family_income
```

Logistic regression — Number of obs = **4,050**; LR chi2(**2**) = **20.50**; Prob > chi2 = **0.0000**; Log likelihood = **-2575.1896**; Pseudo R2 = **0.0040**

Depression_recode	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
family_income						
= 1	**2.197152**	**1.688857**	**1.02**	**0.306**	**.4870549**	**9.911571**
> 1	**.6782582**	**.0636801**	**-4.14**	**0.000**	**.5642579**	**.8152906**
_cons	**2.730769**	**.2365952**	**11.59**	**0.000**	**2.304286**	**3.236186**

Note: **_cons** estimates baseline odds.

```
359 .
360 .
361 .
362 . * Model 1

363 .
364 . logistic Depression_recode i.dmdyrusz_filled ridageyr_recode i.family_income ib3.dmdeduc2
```

Logistic regression — Number of obs = **4,049**; LR chi2(**8**) = **104.65**; Prob > chi2 = **0.0000**; Log likelihood = **-2532.7077**; Pseudo R2 = **0.0202**

Depression_recode	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
dmdyrusz_filled						
> 5	**1.717728**	**.1286524**	**7.22**	**0.000**	**1.483208**	**1.989329**
ridageyr_recode	**.8639245**	**.0587639**	**-2.15**	**0.032**	**.7560967**	**.9871297**
family_income						
= 1	**2.093492**	**1.616449**	**0.96**	**0.339**	**.4609275**	**9.508459**
> 1	**.6936667**	**.068209**	**-3.72**	**0.000**	**.5720727**	**.8411055**
dmdeduc2						
Less 9th grade	**.9590451**	**.1447669**	**-0.28**	**0.782**	**.7134294**	**1.28922**
9-11th grade	**1.171445**	**.1511258**	**1.23**	**0.220**	**.9097251**	**1.50846**
Some college or AA degree	**1.223099**	**.1148552**	**2.14**	**0.032**	**1.017489**	**1.470259**
College graduate or hig..	**.9593293**	**.0936538**	**-0.43**	**0.671**	**.7922638**	**1.161624**
_cons	**1.89713**	**.2379014**	**5.11**	**0.000**	**1.483733**	**2.425707**

Note: **_cons** estimates baseline odds.

```
365 .
366 .
367 .
368 . * Model 2

369 .
370 . logistic Depression_recode i.dmdeduc2 i.ridageyr_recode i.dmdyrusz_filled##i.family_income
```

Logistic regression — Number of obs = **4,049**; LR chi2(**10**) = **105.38**; Prob > chi2 = **0.0000**; Log likelihood = **-2532.3379**; Pseudo R2 = **0.0204**

Depression_recode	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
dmdeduc2						
9-11th grade	**1.228429**	**.2066352**	**1.22**	**0.221**	**.8834224**	**1.708171**
High school graduate/GED	**1.04642**	**.1584943**	**0.30**	**0.765**	**.7776429**	**1.408095**
Some college or AA degree	**1.27993**	**.1899666**	**1.66**	**0.096**	**.9568659**	**1.712068**
College graduate or hig..	**1.004198**	**.1471425**	**0.03**	**0.977**	**.7535199**	**1.338269**
ridageyr_recode						
Old Population	**.8645727**	**.0588753**	**-2.14**	**0.033**	**.7565485**	**.9880211**
dmdyrusz_filled						
> 5	**1.572863**	**.294295**	**2.42**	**0.015**	**1.089994**	**2.269642**
family_income						
= 1	**.945487**	**1.170832**	**-0.05**	**0.964**	**.0834797**	**10.70853**
> 1	**.6497904**	**.1063435**	**-2.63**	**0.008**	**.4714843**	**.8955283**
dmdyrusz_filled# family_income						
> 5#= 1	**3.38374**	**5.505928**	**0.75**	**0.454**	**.1394265**	**82.11993**
> 5#> 1	**1.10572**	**.220684**	**0.50**	**0.615**	**.7477546**	**1.635051**
_cons	**1.919033**	**.3460404**	**3.61**	**0.000**	**1.347701**	**2.73257**

Note: **_cons** estimates baseline odds.

```
371 .
```