

Coding Sample

Minjae Seo

August 10, 2024

1 Stata Code: Data Cleaning

```
1 cd "\\Client\\H$\\Desktop\\GPRL_StataAssessment_2024\\data"
2
3 // Q1: At what level is each dataset uniquely identified?
4
5 /*****
6 * Dataset: Demographics
7 * Description: Each row represents a combination of a household
8   at different wave and demographic information for each member
9   within that household.
10 * Unique Identifiers: hhid (household ID), wave (1=before, 2=after),
11   hhmid (household member ID)
12 *****/
13 use demographics.dta, clear
14 browse in 1/5
15
16 /*****
17 * Dataset: Asset
18 * Description: Each row represents a combination of a household
19   at a specific wave and information about assets owned by the household.
20 * Unique Identifiers: hhid (household ID), wave (1=before, 2=after)
21 *****/
22 use assets.dta, clear
23 browse in 1/5
24
25 /*****
26 * Dataset: Depression
27 * Description: Each row represents a combination of a household
28   at different wave with information for the Kessler Psychological
29   Distress Scale for household heads and their spouses.
```

```

30 * Unique Identifiers: hhid (household ID), wave (1=before, 2=after),
31   hhmid (household member ID)
32 *****/
33 use depression.dta, clear
34 browse in 1/5
35
36 /** Q2: Proxying for household size */
37 use demographics.dta, clear
38
39 keep if wave == 1
40 bysort hhid: egen household_size_proxy_wave1 = count(hhmid)
41 gen household_size_proxy_wave2 = household_size_proxy_wave1
42 sum household_size_proxy_wave1
43 browse hhid hhmid household_size_proxy_wave1 in 1/10
44
45 /** Q3: Impute missing values for currentvalue */
46 use assets.dta, clear
47 egen median_currentvalue = median(currentvalue), by(Asset_Type)
48 replace currentvalue = median_currentvalue if missing(currentvalue)
49 drop median_currentvalue
50 browse hhid wave currentvalue in 1/22
51
52 /** Q4: Calculate total monetary value */
53 gen total_monetary_value = quantity * currentvalue
54 browse hhid wave quantity currentvalue total_monetary_value in 1/22
55
56 /** Q5: Produce household-wave level dataset */
57 destring hhid, replace
58 egen total_value_animals = total(quantity * currentvalue) if Asset_Type == 1, by(hhid wave)
59 egen total_value_tools = total(quantity * currentvalue) if Asset_Type == 2, by(hhid wave)
60 egen total_value_durable_goods = total(quantity * currentvalue) if Asset_Type == 3, by(hhid
    wave)
61 egen total_asset_value = total(currentvalue), by(hhid wave)
62 collapse (first) total_value_animals total_value_tools total_value_durable_goods (first)
    total_asset_value, by(hhid wave)
63 browse hhid wave total_value_animals total_value_tools total_value_durable_goods
    total_asset_value in 1/10
64 save updated_assets.dta, replace
65
66 /** Q6: Kessler-10 scale for mental health */
67 use depression.dta, clear
68 egen kessler_score = rowtotal(tired nervous sonervous hopeless restless sorestless depressed
    everythingeffort nothingcheerup worthless)

```

```

69 gen kessler_categories = ""
70 replace kessler_categories = "no significant depression" if kessler_score >= 10 &
    kessler_score <= 19
71 replace kessler_categories = "mild depression" if kessler_score >= 20 & kessler_score <= 24
72 replace kessler_categories = "moderate depression" if kessler_score >= 25 & kessler_score <=
    29
73 replace kessler_categories = "severe depression" if kessler_score >= 30 & kessler_score <=
    50
74 replace kessler_categories = "." if kessler_score == 0
75 list wave hhid hhmid kessler_score kessler_categories in 1/20
76 sum kessler_score
77 tab kessler_categories
78 save depression.dta, replace
79
80 /** Q7: Combine datasets for analysis */
81 use depression.dta, clear
82 recast double hhid
83 merge 1:m hhid wave hhmid using demographics.dta
84 drop if _merge == 2
85 drop _merge
86 save combined_data.dta, replace
87
88 use updated_assets.dta, clear
89 recast double hhid
90 duplicates report hhid wave
91 use combined_data.dta, clear
92 merge m:1 hhid wave using updated_assets.dta
93 drop if _merge == 2
94 drop _merge
95 save merged.dta, replace

```

2 Stata Code: Analysis

```

1 /**Part 2
2 Exploratory analysis
3 Using Wave 1 data, conduct exploratory analysis to understand the relationship between
4 depression and household and demographic characteristics among individuals in Ghana.
5 **/
6
7 /** Specifically, do the following:
8 Q1: Explore the relationship between depression and:

```

```

9      (1) Household wealth, proxied by total asset value.
10     (2) A household or demographic characteristic that seems interesting to you.
11 Present the results from your exploration through tables, plots, a write-up, or anything
    else you
12 **/
13
14 * Q1: Explore the relationship between depression and:
15 *      (1) Household wealth, proxied by total asset value.
16
17 // import the dataset
18 cd "\\Client\\H$\\Desktop\\GPRL_StataAssessment_2024\\data"
19 use merged.dta, clear
20
21 * Keep only Wave 1 data for the analysis
22 keep if wave == 1
23
24 *Looked at each of factor for each column
25 label list
26
27 * Summary statistics for total_asset_value and depression score
28 summarize total_asset_value kessler_score
29
30 * Scatter plot of depression score by total_asset_value
31 scatter kessler_score total_asset_value, title("Depression score vs Total Asset Value")
    xlabel(, format(%10.0gc)) ylabel(, format(%10.0gc))
32
33 * Correlation between k-10 scale scores and total_asset_value
34 correlate kessler_score total_asset_value
35
36 * Q1: Explore the relationship between depression and:
37 *      (2) A household or demographic characteristic that seems interesting to you.
38
39 // age, gender, treated_household
40
41 // age
42 * Summary statistics for age and kessler_score
43 summarize age kessler_score
44
45
46 // gender
47 *gender(1 - Male, 5 - Female)
48 tabulate gender
49

```

```

50 *summary statistics of depression scores gender
51 by gender, sort: summarize kessler_score
52
53 * Box Plot of Depression Scores by Gender
54 graph box kessler_score, over(gender) title("Depression Score by Gender")
55
56 * Histogram of Depression Scores by Gender
57 histogram kessler_score, by(gender) title("Depression Scores by Gender")
58
59 * Difference in depression scores between genders.
60 ttest kessler_score, by(gender)
61
62
63 // treat_hh(household = 1(when treated) 0 if not)
64 tabulate treat_hh
65
66 *summary statistics of depression scores by treatment indicator
67 by treat_hh, sort: summarize kessler_score
68
69 * Box Plot of Depression Scores by Treatment Status
70 graph box kessler_score, over(treat_hh) title("Depression Score by Treatment Status")
71
72 * Histogram of Depression Scores by Treatment status
73 histogram kessler_score, by(treat_hh) title("Depression Scores(by hhid)")
74
75 * Conduct a t-test to see if the mean depression score differs between treatment groups
76 ttest kessler_score, by(treat_hh)
77
78 * Regression
79 * using robust standard error
80
81 * Regress depression scores on total_asset_value
82
83 reg kessler_score total_asset_value, robust
84 eststo model1
85
86 * Regress depression scores on age
87 reg kessler_score age, robust
88 eststo model2
89
90 * Regress depression scores on gender
91 reg kessler_score gender, robust
92 eststo model3

```

```

93
94 * Regress depression scores on treat_hh
95 reg kessler_score treat_hh, robust
96 eststo model4
97 esttab, r2 ar2 se scalar(rmse)
98
99 * output the regression results
100 outreg2 [model1 model2 model3 model4] using "RegressionResults.doc", replace word title("
    Regression Analysis")
101
102 * outreg2 [model1 model2 model3 model4] using "RegressionResults.tex", replace word title("
    Regression Analysis") tex
103
104 /**
105 Evaluating the RCT
106 Using Wave 2 data to measure outcomes, answer the following questions, explaining any
107 decisions and assumptions you make, and interpret your results. There is no need for you to
    address the validity of the random assignment of the intervention.
108 **/
109
110 * Q2: Were the GT sessions effective at reducing depression?
111
112 // Following assumption for my RCT evaluation
113
114 * To check whether Group Therapy sessions in wave 2 are effective for decreasing depression,
    I will do a statistical analysis compare depression scores(kessler_score) between
    treatment(treated household) and control(controlled household) at wave2
115
116 use merged.dta, clear
117
118 * Keep only Wave 2 data
119 keep if wave == 2
120
121 * summary statistics of kessle and /treat_hh
122 summarize kessler_score treat_hh
123
124 *summary statistics of depression scores by treatment indicator
125 by treat_hh, sort: summarize kessler_score
126
127 * To check randomization
128 tabulate treat_hh
129
130 * Average treatment effect for treated wave 2(after intervention)

```

```

131 regress kessler_score treat_hh, robust
132 eststo fit1
133 esttab, r2 ar2 se scalar(rmse)
134
135 * Save regression results
136 outreg2 fit1 using "GT_effectiveness.doc", replace
137 * outreg2 fit1 using "GT_effectiveness.tex", replace
138
139 /**
140 Q3: Did the effect of the GT sessions on depression vary for men and women? To answer this
      question perform a linear regression of the Kessler Score against a      Woman      binary
      variable, a      Treated      Household      binary variable, and an interaction term      Treated
      Household * Woman , using only wave 2 observations.
141
142 Note: In your write-up for this question, please make sure to explain and interpret all
      coefficients in your specification, keeping in mind units and reference groups.
143 **/
144
145 * To check randomization between gender
146 tabulate gender
147
148 *Summary Statistics of kessler_score by gender
149 sort gender
150 by gender: summarize kessler_score
151
152 *Linear Regression of Differential treatment effect: to measure the average treatment effect
      of GT between two gender
153 regress kessler_score treat_hh gender treat_hh#gender, robust
154 eststo fit2
155 esttab, r2 ar2 se scalar(rmse)
156
157 * Save regression results
158 outreg2 fit2 using "GT_gender.doc", replace
159 * outreg2 fit2 using "GT_gender.tex", replace

```