Statistics/Data analysis

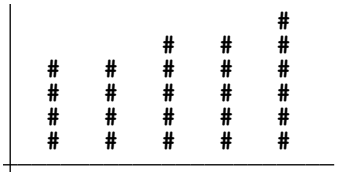
```
      name:  <unnamed>
       log:  C:\Users\kishika\OneDrive - The University of Chicago\IDinsight_Technical
>  Assignment - 1_log file.smcl
  log type:  smcl
 opened on:  22 Feb 2024, 18:46:09

1 . do "C:\Users\kishika\OneDrive - The University of Chicago\IDinsight_Technical Assign
> ment - 1_do file.do"

2 . *Changing the working directory
3 . cd "C:\Users\kishika\OneDrive - The University of Chicago"
C:\Users\kishika\OneDrive - The University of Chicago

4 .
5 . *Loading the dataset
6 . use "Coding Exercise - Regression Analysis Data"

7 .
8 . *Generating and viewing the dummy variables for the ward
9 . tab id_ward, gen(dummy_ward)
```

Ward	Freq.	Percent	Cum.
1	276	6.39	6.39
2	101	2.34	8.72
3	216	5.00	13.72
4	145	3.35	17.08
5	260	6.02	23.09
6	364	8.42	31.51
7	176	4.07	35.59
8	221	5.11	40.70
9	369	8.54	49.24
10	304	7.03	56.27
11	217	5.02	61.29
12	363	8.40	69.69
13	294	6.80	76.49
14	167	3.86	80.36
15	148	3.42	83.78
16	412	9.53	93.31
17	289	6.69	100.00
Total	4,322	100.00	

```
10. ds dummy_ward*
dummy_ward1   dummy_ward4   dummy_ward7   dummy_ward10  dummy_ward13  dummy_ward16
dummy_ward2   dummy_ward5   dummy_ward8   dummy_ward11  dummy_ward14  dummy_ward17
dummy_ward3   dummy_ward6   dummy_ward9   dummy_ward12  dummy_ward15

11.
12. *Generating and applying the weights while accounting for the clustering on the basi
> s of the villages
13. gen inv_prob_weight = 1 / sample_prob

14. svyset village [pweight=inv_prob_weight]

Sampling weights: inv_prob_weight
             VCE: linearized
     Single unit: missing
        Strata 1: <one>
 Sampling unit 1: village
           FPC 1: <zero>
```

```
15.
16. *Including the instrumental variable, the other covariates and the dummy variables
17. ivregress 2sls diarrhoea_self_outcome (Implementation = Treatment) dummy_ward1   dum
  > my_ward5 dummy_ward9   dummy_ward13  dummy_ward17  dummy_ward2   dummy_ward6   dummy
  > _ward10  dummy_ward14 dummy_ward3   dummy_ward7   dummy_ward11  dummy_ward15 dummy_w
  > ard4   dummy_ward8 dummy_ward12  dummy_ward16 edu age water_access health_access hea
  > lth_quality sannut_groups ppi_pct100 caregiver_experience
  note: dummy_ward16 omitted because of collinearity.

  Instrumental variables 2SLS regression          Number of obs   =      3,708
                                                  Wald chi2(25)   =      67.93
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.0178
                                                  Root MSE        =     .34134

------------------------------------------------------------------------------
diarrhoea_self_out~e | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
---------------------+--------------------------------------------------------
      Implementation |  -.0379379   .0115371    -3.29   0.001    -.0605501   -.0153257
         dummy_ward1 |   .0553207   .0294905     1.88   0.061    -.0024797    .1131211
         dummy_ward5 |   .0665972   .0300768     2.21   0.027     .0076477    .1255468
         dummy_ward9 |   .0795076   .0271037     2.93   0.003     .0263853    .1326298
        dummy_ward13 |    .070322   .0296728     2.37   0.018     .0121643    .1284797
        dummy_ward17 |   .0359362   .0288458     1.25   0.213    -.0206006     .092473
         dummy_ward2 |   .0767146   .0416226     1.84   0.065    -.0048641    .1582933
         dummy_ward6 |   .0543109   .0272178     2.00   0.046     .000965    .1076569
        dummy_ward10 |  -.0002265   .0284186    -0.01   0.994     -.055926     .055473
        dummy_ward14 |   .1155498   .0349898     3.30   0.001     .0469711    .1841285
         dummy_ward3 |   .0396374   .0318399     1.24   0.213    -.0227676    .1020424
         dummy_ward7 |    .014783   .0334024     0.44   0.658    -.0506845    .0802505
        dummy_ward11 |   .0520126   .0317233     1.64   0.101    -.0101639    .1141892
        dummy_ward15 |   .0667244   .0374623     1.78   0.075    -.0067004    .1401492
         dummy_ward4 |    .062265   .0367002     1.70   0.090    -.0096661    .1341961
         dummy_ward8 |   .0643208   .0321885     2.00   0.046     .0012326     .127409
        dummy_ward12 |   .0565659   .0276699     2.04   0.041     .0023338     .110798
        dummy_ward16 |          0  (omitted)
                 edu |  -.0137709   .0060775    -2.27   0.023    -.0256825   -.0018593
                 age |  -.0026297   .0008215    -3.20   0.001    -.0042399   -.0010196
        water_access |   .0000653   .0000602     1.08   0.278    -.0000526    .0001832
       health_access |   .0001398   .0001075     1.30   0.194     -.000071    .0003505
      health_quality |  -.0222102   .0187061    -1.19   0.235    -.0588735     .014453
       sannut_groups |   .0174054   .0366331     0.48   0.635    -.0543942     .089205
          ppi_pct100 |   .0003321   .0003101     1.07   0.284    -.0002757    .0009399
caregiver_experience |  -.0012165   .0026934    -0.45   0.652    -.0064956    .0040625
               _cons |   .2574678   .0466225     5.52   0.000     .1660893    .3488463
------------------------------------------------------------------------------
  Endogenous: Implementation
  Exogenous:  dummy_ward1 dummy_ward5 dummy_ward9 dummy_ward13
              dummy_ward17 dummy_ward2 dummy_ward6 dummy_ward10
              dummy_ward14 dummy_ward3 dummy_ward7 dummy_ward11
              dummy_ward15 dummy_ward4 dummy_ward8 dummy_ward12 edu
              age water_access health_access health_quality
              sannut_groups ppi_pct100 caregiver_experience
              Treatment

18.
19. *Inspecting to see the variables with missing values
20. inspect

  hhid:  Household ID                          Number of observations
  -----------------------                    ------------------------------
                                                    Total      Integers   Nonintegers
  |  #                #          Negative            -             -             -
  |  #   #   #   #    #          Zero                -             -             -
  |  #   #   #   #    #          Positive        4,322         4,322             -
  |  #   #   #   #    #                        ---------     ---------     ---------
  |  #   #   #   #    #          Total           4,322         4,322             -
  |  #   #   #   #    #          Missing             -
  +-------------------+                        ---------
  1                  4322                      4,322
  (More than 99 unique values)
```

id_ward: Ward



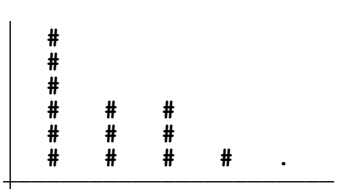
	Number of observations		
	Total	Integers	Nonintegers
Negative	-	-	-
Zero	-	-	-
Positive	**4,322**	**4,322**	-
Total	**4,322**	**4,322**	-
Missing	-		

1 **17**

4,322

(17 unique values)

village: Village



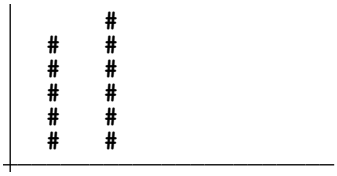
	Number of observations		
	Total	Integers	Nonintegers
Negative	-	-	-
Zero	-	-	-
Positive	**4,322**	**4,322**	-
Total	**4,322**	**4,322**	-
Missing	-		

531 **4931**

4,322

(More than 99 unique values)

Treatment: Sannut Assignment



	Number of observations		
	Total	Integers	Nonintegers
Negative	-	-	-
Zero	**2,084**	**2,084**	-
Positive	**2,238**	**2,238**	-
Total	**4,322**	**4,322**	-
Missing	-		

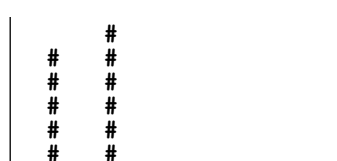
0 **1**

4,322

(2 unique values)

Treatment is labeled and all values are documented in the label.

Implementation: Sannut Implementation



	Number of observations		
	Total	Integers	Nonintegers
Negative	-	-	-
Zero	**2,034**	**2,034**	-
Positive	**2,288**	**2,288**	-
Total	**4,322**	**4,322**	-
Missing	-		

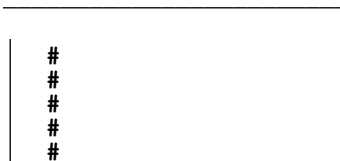
0 **1**

4,322

(2 unique values)

Implementation is labeled and all values are documented in the label.

diarrhoea_self_outcome: Average self-re



	Number of observations		
	Total	Integers	Nonintegers
Negative	-	-	-
Zero	**3,419**	**3,419**	-
Positive	**811**	**532**	**279**
Total	**4,230**	**3,951**	**279**
Missing	**92**		

0 **1**

4,322

(6 unique values)

ppi_pct100: Likelihood of being below t Number of observations

	Total	Integers	Nonintegers
Negative	-	-	-
Zero	1	1	-
Positive	4,321	503	3,818
Total	4,322	504	3,818
Missing	-		
0	96.4	4,322	

(18 unique values)

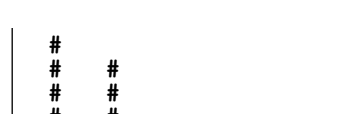
edu: Education Number of observations

	Total	Integers	Nonintegers
Negative	-	-	-
Zero	-	-	-
Positive	4,262	4,262	-
Total	4,262	4,262	-
Missing	60		
1	5	4,322	

(5 unique values)

edu is labeled and all values are documented in the label.

age: Age of Respondent Number of observations



	Total	Integers	Nonintegers
Negative	-	-	-
Zero	-	-	-
Positive	4,215	4,215	-
Total	4,215	4,215	-
Missing	107		
15	95	4,322	

(66 unique values)

water_access: Time taken to fetch water Number of observations

	Total	Integers	Nonintegers
Negative	-	-	-
Zero	97	97	-
Positive	4,198	4,198	-
Total	4,295	4,295	-
Missing	27		
0	900	4,322	

(54 unique values)

health_access: Time taken to reach heal Number of observations

	Total	Integers	Nonintegers
Negative	-	-	-
Zero	-	-	-
Positive	3,935	3,935	-
Total	3,935	3,935	-
Missing	387		
1	420	4,322	

(51 unique values)

health_quality: Respondent received sou Number of observations

		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	744	744	-
#	Positive	3,578	3,578	-
#				
#	Total	4,322	4,322	-
# #	Missing	-		

0 1 4,322

(2 unique values)

sannut_groups: Member of Sanitation and Number of observations

		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	4,222	4,222	-
#	Positive	100	100	-
#				
#	Total	4,322	4,322	-
# .	Missing	-		

0 1 4,322

(2 unique values)

caregiver_experience: Total number of c Number of observations

		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	-	-	-
#	Positive	4,322	4,322	-
#				
#	Total	4,322	4,322	-
# # . . .	Missing	-		

1 32 4,322

(26 unique values)

sample_prob: Probability of being sampl Number of observations

		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	-	-	-
#	Positive	4,322	1,322	3,000
#				
# # #	Total	4,322	1,322	3,000
. # # # #	Missing	-		

.1015 1 4,322

(More than 99 unique values)

dummy_ward1: id_ward== 1.0000 Number of observations

		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	4,046	4,046	-
#	Positive	276	276	-
#				
#	Total	4,322	4,322	-
# .	Missing	-		

0 1 4,322

(2 unique values)

dummy_ward2: id_ward== 2.0000

		Number of observations		
		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	4,221	4,221	-
#	Positive	101	101	-
#				
#	Total	4,322	4,322	-
# .	Missing	-		
0	1	4,322		

(2 unique values)

dummy_ward3: id_ward== 3.0000

		Number of observations		
		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	4,106	4,106	-
#	Positive	216	216	-
#				
#	Total	4,322	4,322	-
# .	Missing	-		
0	1	4,322		

(2 unique values)

dummy_ward4: id_ward== 4.0000

		Number of observations		
		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	4,177	4,177	-
#	Positive	145	145	-
#				
#	Total	4,322	4,322	-
# .	Missing	-		
0	1	4,322		

(2 unique values)

dummy_ward5: id_ward== 5.0000

		Number of observations		
		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	4,062	4,062	-
#	Positive	260	260	-
#				
#	Total	4,322	4,322	-
# .	Missing	-		
0	1	4,322		

(2 unique values)

dummy_ward6: id_ward== 6.0000

		Number of observations		
		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	3,958	3,958	-
#	Positive	364	364	-
#				
#	Total	4,322	4,322	-
# .	Missing	-		
0	1	4,322		

(2 unique values)

dummy_ward7: id_ward== 7.0000

		Number of observations		
		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	4,146	4,146	-
#	Positive	176	176	-
#				
#	Total	4,322	4,322	-
# .	Missing	-		
0	1	4,322		

(2 unique values)

dummy_ward8: id_ward== 8.0000

		Number of observations		
		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	4,101	4,101	-
#	Positive	221	221	-
#				
#	Total	4,322	4,322	-
# .	Missing	-		
0	1	4,322		

(2 unique values)

dummy_ward9: id_ward== 9.0000

		Number of observations		
		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	3,953	3,953	-
#	Positive	369	369	-
#				
#	Total	4,322	4,322	-
# .	Missing	-		
0	1	4,322		

(2 unique values)

dummy_ward10: id_ward== 10.0000

		Number of observations		
		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	4,018	4,018	-
#	Positive	304	304	-
#				
#	Total	4,322	4,322	-
# .	Missing	-		
0	1	4,322		

(2 unique values)

dummy_ward11: id_ward== 11.0000

		Number of observations		
		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	4,105	4,105	-
#	Positive	217	217	-
#				
#	Total	4,322	4,322	-
# .	Missing	-		
0	1	4,322		

(2 unique values)

dummy_ward12: id_ward== 12.0000

		Number of observations		
		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	3,959	3,959	-
#	Positive	363	363	-
#				
#	Total	4,322	4,322	-
# .	Missing	-		
0	1	4,322		

(2 unique values)

dummy_ward13: id_ward== 13.0000

		Number of observations		
		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	4,028	4,028	-
#	Positive	294	294	-
#				
#	Total	4,322	4,322	-
# .	Missing	-		
0	1	4,322		

(2 unique values)

dummy_ward14: id_ward== 14.0000

		Number of observations		
		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	4,155	4,155	-
#	Positive	167	167	-
#				
#	Total	4,322	4,322	-
# .	Missing	-		
0	1	4,322		

(2 unique values)

dummy_ward15: id_ward== 15.0000

		Number of observations		
		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	4,174	4,174	-
#	Positive	148	148	-
#				
#	Total	4,322	4,322	-
# .	Missing	-		
0	1	4,322		

(2 unique values)

dummy_ward16: id_ward== 16.0000

		Number of observations		
		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	3,910	3,910	-
#	Positive	412	412	-
#				
#	Total	4,322	4,322	-
# .	Missing	-		
0	1	4,322		

(2 unique values)

dummy_ward17: id_ward== 17.0000

		Number of observations		
		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	4,033	4,033	-
#	Positive	289	289	-
#				
#	Total	4,322	4,322	-
# .	Missing	-		
0	1	4,322		

(2 unique values)

inv_prob_weight:

		Number of observations		
		Total	Integers	Nonintegers
#	Negative	-	-	-
#	Zero	-	-	-
#	Positive	4,322	1,322	3,000
#				
#	Total	4,322	1,322	3,000
#	Missing	-		
1	9.852217	4,322		

(More than 99 unique values)

```
21.
22. *Generating variables which show 0 if the value is missing and the value of the vari
  > able if the value is not missing
23.
24. generate edu_filled = edu
  (60 missing values generated)

25. replace edu_filled = 0 if missing(edu)
  (60 real changes made)

26.
27. generate age_filled = age
  (107 missing values generated)

28. replace age_filled = 0 if missing(age)
  (107 real changes made)

29.
30. generate water_access_filled = water_access
  (27 missing values generated)

31. replace water_access_filled = 0 if missing(water_access)
  (27 real changes made)

32.
33. generate health_access_filled = health_access
  (387 missing values generated)

34. replace health_access_filled = 0 if missing(health_access)
  (387 real changes made)

35.
36. *Generating dummy variables to capture the missingness which show the value to be 1
  > if the data is missing and 0 if the data is not missing
```

```
37.
38. generate dummy_missing_edu = 1 if (edu == .)
  (4,262 missing values generated)

39. replace dummy_missing_edu = 0 if !missing(edu)
  (4,262 real changes made)

40.
41. generate dummy_missing_age = 1 if (age == .)
  (4,215 missing values generated)

42. replace dummy_missing_age = 0 if !missing(age)
  (4,215 real changes made)

43.
44. generate dummy_missing_water_access = 1 if (water_access == .)
  (4,295 missing values generated)

45. replace dummy_missing_water_access = 0 if !missing(water_access)
  (4,295 real changes made)

46.
47. generate dummy_missing_health_access = 1 if (health_access == .)
  (3,935 missing values generated)

48. replace dummy_missing_health_access = 0 if !missing(health_access)
  (3,935 real changes made)

49. *Final regression equation keeping in mind all the parameters
50. svy: ivregress 2sls diarrhoea_self_outcome (Implementation = Treatment) dummy_ward1
  >   dummy_ward5 dummy_ward9   dummy_ward13  dummy_ward17  dummy_ward2   dummy_ward6
  > dummy_ward10  dummy_ward14 dummy_ward3   dummy_ward7   dummy_ward11  dummy_ward15 du
  > mmy_ward4   dummy_ward8   dummy_ward12  dummy_ward16 health_quality sannut_groups pp
  > i_pct100 caregiver_experience edu_filled age_filled water_access_filled health_acces
  > s_filled dummy_missing_edu dummy_missing_age dummy_missing_water_access dummy_missin
  > g_health_access
  (running ivregress on estimation sample)

  Survey: Instrumental variables 2SLS regression

  Number of strata =   1                        Number of obs   =      4,230
  Number of PSUs   = 603                        Population size = 6,297.9636
                                                Design df       =        602
                                                F(29, 574)      =       2.40
                                                Prob > F        =     0.0001
                                                R-squared       =     0.0182

------------------------------------------------------------------------------
             |             Linearized
diarrhoea_self_outcome | Coefficient  std. err.      t    P>|t|     [95% conf. in
> terval]
-------------+----------------------------------------------------------------
      Implementation |  -.0294779   .0125727    -2.34   0.019    -.0541696   -.
> 0047863
         dummy_ward1 |   .0348818   .0308315     1.13   0.258    -.0256686    .
> 0954322
         dummy_ward5 |   .0337968   .0364364     0.93   0.354    -.0377611    .
> 1053547
         dummy_ward9 |    .052572   .0297889     1.76   0.078    -.0059307    .
> 1110747
        dummy_ward13 |   .0420831    .027372     1.54   0.125     -.011673    .
> 0958393
        dummy_ward17 |   .0339951   .0366458     0.93   0.354    -.0379741    .
> 1059643
         dummy_ward2 |   .0633811   .0405645     1.56   0.119     -.016284    .
> 1430462
         dummy_ward6 |   .0309164   .0274022     1.13   0.260    -.0228991    .
> 0847319
        dummy_ward10 |  -.0093362   .0286597    -0.33   0.745    -.0656214    .
> 0469489
```

```
                dummy_ward14 |   .1135838   .0441847     2.57   0.010     .0268088    .
> 2003587
                 dummy_ward3 |    .006387   .0299617     0.21   0.831    -.0524552    .
> 0652292
                 dummy_ward7 |   -.027486   .0299995    -0.92   0.360    -.0864024    .
> 0314303
                dummy_ward11 |   .0376223   .0279388     1.35   0.179    -.0172471    .
> 0924917
                dummy_ward15 |   .0439964   .0313982     1.40   0.162    -.0176669    .
> 1056598
                 dummy_ward4 |   .0109946   .0321173     0.34   0.732     -.052081    .
> 0740703
                 dummy_ward8 |   .0359947   .0314966     1.14   0.254    -.0258618    .
> 0978512
                dummy_ward12 |   .0383295   .0274318     1.40   0.163    -.0155441    .
> 0922031
                dummy_ward16 |          0  (omitted)
              health_quality |  -.0195182   .0184229    -1.06   0.290    -.0556992    .
> 0166628
               sannut_groups |    .030631   .0329054     0.93   0.352    -.0339923    .
> 0952543
                  ppi_pct100 |   .0000224    .000307     0.07   0.942    -.0005805    .
> 0006253
      caregiver_experience |  -.0020814   .0022943    -0.91   0.365    -.0065871    .
> 0024244
                edu_filled |  -.0141391   .0065191    -2.17   0.030    -.0269421   -.
> 0013362
                age_filled |  -.0025017   .0007854    -3.19   0.002    -.0040441   -.
> 0009593
       water_access_filled |   .0000308   .0000605     0.51   0.611    -.0000881    .
> 0001496
      health_access_filled |   .0001461   .0001127     1.30   0.196    -.0000753    .
> 0003675
         dummy_missing_edu |  -.0168754    .058488    -0.29   0.773    -.1317407    .
> 0979898
        dummy_missing_age |  -.1270723   .0429952    -2.96   0.003    -.2115111   -.
> 0426335
 dummy_missing_water_access |   .1049132   .0907307     1.16   0.248    -.0732739    .
> 2831003
dummy_missing_health_access |  -.0503056   .0233055    -2.16   0.031    -.0960756   -.
> 0045357
                      _cons |   .2858959   .0466506     6.13   0.000     .1942782    .
> 3775137
------------------------------+---------------------------------------------------------------
```

Endogenous: **Implementation**
Exogenous: **dummy_ward1 dummy_ward5 dummy_ward9 dummy_ward13 dummy_ward17 dummy_ward2 dummy_ward6 dummy_ward10 dummy_ward14 dummy_ward3 dummy_ward7 dummy_ward11 dummy_ward15 dummy_ward4 dummy_ward8 dummy_ward12 health_quality sannut_groups ppi_pct100 caregiver_experience edu_filled age_filled water_access_filled health_access_filled dummy_missing_edu dummy_missing_age dummy_missing_water_access dummy_missing_health_access Treatment**

```
51.
52. *Interpreting the results
53.
54. *(a) The treatment effect of compliers rounded to the nearest thousandth place is -.
  > 029.
```

```
55.
56. *(b) The standard error on this coefficient rounded to the nearest thousandth place
  > is 0.126.
57.
58. *(c) To see if we can reject the null hypothesis of no difference at the 5% level, w
  > e need to analyze the p-value associated with this coefficient. Since the p-value is
  >  0.019 which is lesser than the level of significance (alpha = 0.05), we reject the
  > null hypothesis of no difference. This essentially means that there is a statistical
  > ly significant difference in self-reported diarrhoea levels between treatment and co
  > ntrol households.
59.
  end of do-file

60. log close
        name:  <unnamed>
         log:  C:\Users\kishika\OneDrive - The University of Chicago\IDinsight_Technical
  >  Assignment - 1_log file.smcl
    log type:  smcl
   closed on:  22 Feb 2024, 18:46:34
```