Group 28 - Andrew Ma, Hiro Schmidt

Replication Study: Housing, Health, and Happiness

For our final project, we will be conducting a replication study of Matias D. Cattaneo et.al "Housing, Health, and Happiness" 2009. We will begin by introducing the paper's main research questions and conclusions along with basic data exploration and a summary statistics table of important covariates.

In a sentence, Matias D. Cattaneo and his peers study the health effects of the *Piso Firme* project, a government-subsidized "slum upgrade" involving the installation of cement floors in households near the Torreon region of Mexico. In particular, they compare outcome variables of parasitic infestations, diarrhea, anemia, height, weight, and cognitive development against the untreated Durango region across state lines. Moreover, the regression model includes 41 covariates, including both numerical and dummy variables such as parasite count, total income, e.t.c.

To ensure causality between the Piso Firme program and an increase in household health/happiness, households in the study undergo a pre-intervention check, to ensure that we are using comparable households from both neighborhoods. This is done through comparing variables such as father presence, presence of diseases, respiratory issues, intellectual scores, e.t.c. - there are no significant differences in mean values between control and treatment groups. This implies that improvements can be attributed to the Piso Firme program and the installation of a cement floor.

With respect to data cleaning, there seems to be the presence of missing/null values within all rows of both datasets; researchers seem to impute missing values with 'zeros' as many of them are dummy variables and the missing values seem missing at random; ie, imputing with zeroes is a quick and easy solution to a nuanced and complicated data-completeness issue. Finally, you may find below a table of the covariates and outcome variables of the regression model implemented by Cattaneo et al.; other variables were used in pre-intervention studies for causality purposes and not shown below.

Covariates:

```
Column type frequency:
  numeric                  36
_________________________
Group variables            None

── Variable type: numeric ──────────────────────────────────────────────────────
   skim_variable        complete_rate     mean      sd hist
 1 dpisofirme                   1      4.90e-1 5.00e-1 ▇▁▁▁▇
 2 idcluster                    0.990  2.09e+8 1.37e+8 ▇▁▁▁▇
 3 coord_x                      0.929 -1.03e+2 5.52e-2 ▆▇▃▂▁
 4 coord_y                      0.929  2.56e+1 3.71e-2 ▁▂▇▅▂
 5 idmun                        1      2.10e+1 1.37e+1 ▇▁▁▁▇
 6 idmza                        1      9.84e+1 2.59e+2 ▇▁▁▁▁
 7 S_age                        1      1.34e+1 1.43e+1 ▇▅▂▁▁
 8 S_childma                    0.611  9.65e-1 1.83e-1 ▁▁▁▁▇
 9 S_childmaage                 0.635  2.94e+1 9.43e+0 ▆▇▃▁▁
10 S_childmaeduc                0.635  6.69e+0 2.84e+0 ▃▇▇▃▁
11 S_childpa                    0.611  7.80e-1 4.14e-1 ▂▁▁▁▇
12 S_childpaage                 0.492  3.23e+1 1.02e+1 ▃▇▅▂▁
13 S_childpaeduc                0.490  6.74e+0 3.20e+0 ▃▇▇▅▂
14 S_HHpeople                   1      5.63e+0 2.17e+0 ▂▇▃▁▁
15 S_rooms                      1      2.04e+0 1.09e+0 ▇▅▂▁▁
16 dtriage_5_02_male            1      1.05e-2 1.02e-1 ▇▁▁▁▁
17 dtriage_5_35_male            1      1.55e-2 1.24e-1 ▇▁▁▁▁
18 dtriage_5_68_male            1      1.67e-2 1.28e-1 ▇▁▁▁▁
19 dtriage_5_911_male           1      1.66e-2 1.28e-1 ▇▁▁▁▁
20 dtriage_5_02_female          1      9.26e-3 9.58e-2 ▇▁▁▁▁
21 dtriage_5_35_female          1      1.64e-2 1.27e-1 ▇▁▁▁▁
22 dtriage_5_68_female          1      1.51e-2 1.22e-1 ▇▁▁▁▁
23 dtriage_5_911_female         1      1.51e-2 1.22e-1 ▇▁▁▁▁
24 S_waterland                  1      9.73e-1 1.63e-1 ▁▁▁▁▇
25 S_waterhouse                 1      5.23e-1 4.99e-1 ▇▁▁▁▇
26 S_electricity                1      9.90e-1 1.02e-1 ▁▁▁▁▇
27 S_milkprogram                1      6.89e-2 2.53e-1 ▇▁▁▁▁
28 S_foodprogram                1      2.88e-2 1.67e-1 ▇▁▁▁▁
29 S_seguropopular              1      1.42e-2 1.18e-1 ▇▁▁▁▁
30 S_hasanimals                 1      4.93e-1 5.00e-1 ▇▁▁▁▇
31 S_animalsinside              1      1.93e-1 3.94e-1 ▇▁▁▁▂
32 S_garbage                    1      8.19e-1 3.85e-1 ▂▁▁▁▇
33 S_washhands                  1      3.74e+0 1.49e+0 ▃▇▃▁▁
34 S_incomepc                   0.999  1.06e+3 3.47e+3 ▇▁▁▁▁
35 S_cashtransfers              0.999  1.31e+1 3.42e+1 ▇▁▁▁▁
36 S_assetspc                   1.00   2.14e+4 7.28e+3 ▅▇▁▁▁
```

Outcomes:

```
── Data Summary ────────────────────────
                           Values
Name                       indiv
Number of rows             6693
Number of columns          89
___________________________
Column type frequency:
  numeric                  7
___________________________
Group variables            None

── Variable type: numeric ──────────────────────────────────────────────────────
  skim_variable complete_rate   mean     sd hist
1 S_parcount            0.467   0.302  0.625 ▇▁▁▁
2 S_diarrhea            0.609   0.133  0.340 ▇▁▁▁
3 S_anemia              0.561   0.386  0.487 ▇▁▁▅
4 S_mccdts              0.0898 15.4   20.1   ▇▁▁▁
5 S_pbdypct             0.237  31.9   25.4   ▇▃▂▁
6 S_haz                 0.591  -0.602  1.11  ▁▃▇▂
7 S_whz                 0.594   0.131  1.13  ▁▃▇▁
```

References

1. Matias D. Cattaneo, Sebastian Galiani, Paul J. Gertler,Sebastian Martinez and RocioTitiunik ,“Housing, Health, and Happiness”, *American Economic Journal: Economic Policy*, 2009