**Write-up of Stata Sample: Exercises with Moderators**

The broad theoretical framework pursued in the questions asked here is based on the mounting consensus that social disadvantage is not only related to negative health outcomes but that social disadvantage also makes the effect of environmental and social correlates of health worse. Exercises 1 and 2 use data from the [UK Biobank](https://byu.box.com/s/tn4lp8aacn0jvv3sd57ri3nmc60dtqea) dataset.

1. There is clear evidence that air pollution damages the cardiovascular system but recent efforts have started to examine whether air pollution damages the brain. I test the following hypotheses:
   * 1. H(0): Educational attainment does not moderate the relationship between air pollution (particulate matter 2.5 in μg/m3) and thalamus volume (in mm3).
     2. H(a): Educational attainment does moderate the relationship between air pollution (particulate matter 2.5 in μg/m3) and thalamus volume (in mm3).

The following regression model tests the null hypothesis. The volume of the left side of the thalamus is the dependent variable. I control for total brain size, age, gender, race, income, body-mass index, and smoking status.

eststo m1: reg thlmsL c.pm2\_5##coldeg total age white inc bmi ib0.smoker female if touse

Source | SS df MS Number of obs = 17,125

-------------+---------------------------------- F(11, 17113) = 775.22

Model | 3.1965e+09 11 290590619 Prob > F = 0.0000

Residual | 6.4148e+09 17,113 374848.72 R-squared = 0.3326

-------------+---------------------------------- Adj R-squared = 0.3321

Total | 9.6113e+09 17,124 561275.575 Root MSE = 612.25

-------------------------------------------------------------------------------

thlmsL | Coefficient Std. err. t P>|t| [95% conf. interval]

--------------+----------------------------------------------------------------

pm2\_5 | -18.34178 6.894005 -2.66 0.008 -31.85473 -4.82882

|

coldeg |

College de.. | -140.2531 92.27669 -1.52 0.129 -321.1249 40.61865

|

coldeg#|

c.pm2\_5 |

College de.. | 24.1305 9.273321 2.60 0.009 5.953838 42.30716

|

total | .0030986 .0000781 39.68 0.000 .0029456 .0032517

age | -17.84789 .7875256 -22.66 0.000 -19.39152 -16.30426

white | 324.1811 28.73465 11.28 0.000 267.8582 380.5039

inc | 23.2106 4.651475 4.99 0.000 14.09323 32.32797

bmi | -2.684027 1.082966 -2.48 0.013 -4.806751 -.5613037

|

smoker |

Previous | -30.87326 10.18275 -3.03 0.002 -50.83249 -10.91403

Current | -41.99053 24.96104 -1.68 0.093 -90.91672 6.935663

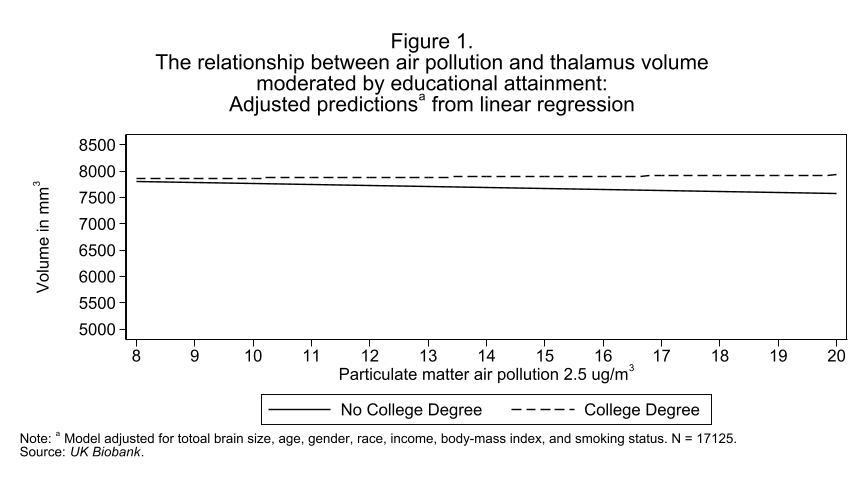
|

female | 682.0416 9.698259 70.33 0.000 663.0321 701.0512

\_cons | 3080.964 175.218 17.58 0.000 2737.519 3424.41

-------------------------------------------------------------------------------

Here is a depiction of the data:



**Interpretation of data**:

* When respondents do not have a college degree, their expected thalamus volume is 3,080.96 mm³. For those with a college degree, the expected thalamus volume is 2,940.71 mm³.
* Among respondents without a college degree, a 1 µg/m³ increase in air pollution is associated with an 18.34 mm³ decrease in thalamus volume, holding all other variables constant. For respondents with a college degree, the same increase in air pollution corresponds to a 5.8 mm³ decrease in thalamus volume.
* In Figure 1, the slope of the line for respondents with a college degree starts further from zero, suggesting that the relationship between thalamus volume and air pollution is minimal in this group. In contrast, the slope for those without a college degree indicates a slight effect of educational attainment on the relationship between thalamus volume and air pollution.
* There is a 12.54 mm³ difference in the relationship between thalamus volume and air pollution when a participant has a college degree. The interaction coefficient is 24.13. At p < 0.05, we reject the null hypothesis and conclude that educational attainment moderates the relationship between thalamus volume and air pollution.

I also suspect that the left thalamus of people becomes more sensitive to pollution as people age. Here are my hypotheses:

* 1. H(0): Age does not moderate the relationship between air pollution (particulate matter 2.5 in μg/m3) and thalamus volume (in mm3).
  2. H(a): Age moderates the relationship between air pollution (particulate matter 2.5 in μg/m3) and thalamus volume (in mm3).

This regression model controls for total brain size, gender, race, educational attainment, income, body-mass index, and smoking status.

eststo m2: reg thlmsL c.pm2\_5##c.age total coldeg female inc bmi ib0.smoker white if touse

Source | SS df MS Number of obs = 17,125

-------------+---------------------------------- F(11, 17113) = 774.32

Model | 3.1940e+09 11 290364061 Prob > F = 0.0000

Residual | 6.4173e+09 17,113 374994.348 R-squared = 0.3323

-------------+---------------------------------- Adj R-squared = 0.3319

Total | 9.6113e+09 17,124 561275.575 Root MSE = 612.37

-------------------------------------------------------------------------------

thlmsL | Coefficient Std. err. t P>|t| [95% conf. interval]

--------------+----------------------------------------------------------------

pm2\_5 | 8.427482 39.21353 0.21 0.830 -68.43506 85.29003

age | -15.64852 6.281288 -2.49 0.013 -27.96048 -3.336549

|

c.pm2\_5#c.age | -.2203796 .6291039 -0.35 0.726 -1.453488 1.012729

|

total | .0030977 .0000781 39.66 0.000 .0029446 .0032508

coldeg | 98.52985 9.678017 10.18 0.000 79.55995 117.4998

female | 681.3468 9.701844 70.23 0.000 662.3302 700.3634

inc | 23.38734 4.6523 5.03 0.000 14.26836 32.50633

bmi | -2.71842 1.0832 -2.51 0.012 -4.841603 -.5952362

|

smoker |

Previous | -31.01935 10.18461 -3.05 0.002 -50.98223 -11.05646

Current | -41.33074 24.96487 -1.66 0.098 -90.26445 7.602958

|

white | 323.4812 28.74633 11.25 0.000 267.1354 379.8269

\_cons | 2818.242 421.3354 6.69 0.000 1992.382 3644.103

A diagram of a graph

AI-generated content may be incorrect.

**Interpretation of data**:

* When air pollution and age are both 0, the expected thalamus volume is 2,818.24 mm³.
* When age is 0, the slope of air pollution is 8.43, indicating a positive relationship. When air pollution is 0, the slope of age is -15.65, indicating a negative relationship. For each 1-year increase in age, the slope of air pollution decreases by 0.22.
* In Figure 2, the lines representing different age values start at different points but appear relatively parallel, suggesting little to no moderating effect. The interaction coefficient is -0.22, and the p-value is 0.726, which is greater than 0.05. Since p = 0.726, we cannot reject the null hypothesis and conclude that age does not significantly moderate the relationship between air pollution (PM2.5 in µg/m³) and thalamus volume (in mm³).

The final exercises uses data from the [NHANES](https://byu.box.com/s/tkbn1e993z2gw2wgi3o3wez2om5hmkus) dataset.

Toxoplasma Gondii (*T. Gondii*) is a parasite whose definitive host (i.e., can only reproduce there) is the Felidae family (i.e., cats). Humans can be infected with *T. Gondii* to get toxoplasmosis from a variety of sources including ingesting cat feces and undercooked meats. Toxoplasmosis can cause serious health problems in people who have compromised immune systems but has been thought to be harmless to those who are otherwise healthy. However, some recent research suggests that toxoplasmosis is related to cognitive deficits. My hypotheses are:

1. H(0): Being a racial-ethnic minority does not moderate the relationship between toxoplasmosis and performance on the serial digit learning test.
2. (H(a): Being a racial-ethnic minority does moderate the relationship between toxoplasmosis and performance on the serial digit learning test.

In this regression model, I control for age, gender, educational attainment, poverty-to-income ratio, and health.

eststo m3: reg sdl toxo##race age female ib1.edu pir health if touse

Source | SS df MS Number of obs = 4,169

-------------+---------------------------------- F(13, 4155) = 133.47

Model | 30599.9399 13 2353.84153 Prob > F = 0.0000

Residual | 73274.1489 4,155 17.6351742 R-squared = 0.2946

-------------+---------------------------------- Adj R-squared = 0.2924

Total | 103874.089 4,168 24.9218063 Root MSE = 4.1994

-------------------------------------------------------------------------------

sdl | Coefficient Std. err. t P>|t| [95% conf. interval]

--------------+----------------------------------------------------------------

toxo |

Positive | -.2885734 .2746894 -1.05 0.294 -.8271116 .2499648

|

race |

Black | 1.540506 .185523 8.30 0.000 1.176782 1.904231

Hispanic | 2.272872 .1956651 11.62 0.000 1.889263 2.65648

Other | 1.154927 .4478998 2.58 0.010 .276804 2.03305

|

toxo#race |

Positive #|

Black | 1.1476 .3979232 2.88 0.004 .3674572 1.927742

Positive #|

Hispanic | 1.113606 .4159812 2.68 0.007 .2980602 1.929152

Positive #|

Other | 3.013253 .7570917 3.98 0.000 1.528948 4.497557

|

age | .11079 .0064968 17.05 0.000 .0980528 .1235272

female | .1442433 .131448 1.10 0.273 -.113465 .4019516

|

edu |

High school | -1.85435 .1749713 -10.60 0.000 -2.197387 -1.511312

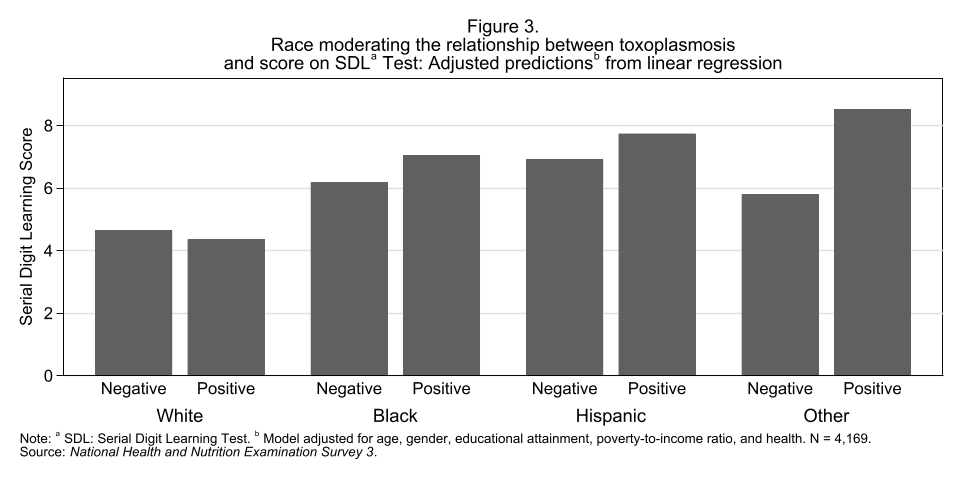
More than .. | -3.363874 .1902215 -17.68 0.000 -3.73681 -2.990938

|

pir | -.3878561 .043954 -8.82 0.000 -.4740296 -.3016827

health | -.3302748 .0703337 -4.70 0.000 -.4681665 -.192383

\_cons | 4.458536 .4051224 11.01 0.000 3.664279 5.252793



**Interpretation of data**:

* In the regression analysis testing whether racial-ethnic minority status moderates the relationship between toxoplasmosis and performance on the Serial Digit Learning Test, the interaction coefficients are 1.15 for Blacks, 1.11 for Hispanics, and 3.01 for the "Other" category.
* The p-values for these interaction terms are all less than 0.05, allowing us to reject the null hypothesis. This suggests that racial-ethnic minority status does moderate the relationship between toxoplasmosis and test performance. Moreover, the effect of toxoplasmosis appears to be worse for racial-ethnic minorities than for Whites.
* In Figure 3, Whites who test positive for toxoplasmosis show a slight decrease in Serial Digit Learning Test scores. However, for Blacks, Hispanics, and individuals in the "Other" category, test scores increase when they are positive for toxoplasmosis.

**Conclusion**:

The theoretical framework that guided these analyses is that social disadvantage is not only related to negative health outcomes but that social disadvantage also makes the effect of environmental and social correlates of health worse. The first test of the moderating effect of educational attainment on thalamus volume and air pollution suggests that those without a college degree are more negatively affected by air pollution than those with a college degree. The second test outcomes suggest that age does not moderate the effect of air pollution and thalamus volume, since the p-value was not significant. The third test suggests that the effect of toxoplasmosis is worse for racial-ethnic minorities than Whites. The first and third test support the theoretical framework that social disadvantage is related to negative health outcomes and that social disadvantage also makes the effect of environmental and social correlates of health worse.