```
-------------------------------------------------------------------------------
      name:  <unnamed>
       log:  C:\Users\brocc\Documents\Coding\code sample\smoking\newlog.lo
> g
  log type:  text
 opened on:  19 Jun 2023, 20:32:47

. do "C:\Users\brocc\OneDrive\Desktop\d.do"

.
. ****************************************************************
. *** DESCRIPTION OF THE DATASET & OBJECTIVE
. ****************************************************************
.
. *       (a) "smoking rate_raw.csv": data on the adult cigarette smoking
> rate by state for years 2018 and 2020.
. *               Source: American Lung Association (2018 data); BRFSS (20
> 20 data).
. *       (b) "income_raw.csv": data on the real median household income b
> y state, for years 2018 and 2020.
. *               Source: FRED.
. *       (c) "how to quit smoking_raw.csv": data on relative interest by
> state (vis-a-vis other states) in searching the term 'how to quit smokin
> g' on Google, for years 2018 and 2020. More specifically, according to G
> oogle, values are "calculated on a scale from 0 to 100, where 100 is the
>  location with the most popularity as a fraction of total searches in th
> at location, a value of 50 indicates a location which is half as popular
> . A value of 0 indicates a location where there was not enough data for
> this term."
. *               Source: Google Trends.
.
. *       (1) Objective 1: What is the effect of the motivation to quit sm
> oking on the smoking rate? The search interest in "how to quit smoking"
> could be an imperfect proxy for motivation if smokers motivated to quit
> smokingconduct this search on Google. Below, we check how the two variab
> les are related: (i) with and without controlling for median income, and
>  (ii) controlling for factors that differ by state but did not change wi
> thin each state between 2018 and 2020 (through regressions on changes in
>  the variables).
. *
. *       (2) Objective 2: What is the effect of income on the smoking rat
> e? It is generally understood that smoking rate is negatively associated
>  with income. We check whether and to what extent this relationship stil
> l holds when we control for state-fixed factors.
.
.
. ****************************************************************
. *** CLEANING & MERGING DATASETS
. ****************************************************************
.
. *** set working directory and import data on smoking rate.
. clear

. cd "C:\Users\brocc\Documents\Coding\code sample\smoking"
C:\Users\brocc\Documents\Coding\code sample\smoking

. import delimited using "https://drive.google.com/u/0/uc?id=1mQKCKHzx43HS
> oIwlq8qGuw2jNv4IVhqf&export=download"
(3 vars, 102 obs)
```

```
.
. *** rename variables.
. rename v1 state

. rename v2 smoke

. rename v3 year

.
. *** eliminate special characters in front of 'Alabama'.
. replace state = substr(state, length(state) - 6, .) if strpos(state, "Al
> abama") > 0
(1 real change made)

.
. *** save as dta file.
. save "smoking rate.dta", replace
file smoking rate.dta saved

.
. *** clean in the same manner for the data on search interest and income.
. clear

. import delimited using "https://drive.google.com/u/0/uc?id=1ac-PzLJ00btf
> D_MrBtLZGACVQ9H2Ef9k&export=download"
(3 vars, 102 obs)

.
. rename v1 state

. rename v2 quitsmoke

. rename v3 year

. save "how to quit smoking.dta", replace
file how to quit smoking.dta saved

.
. clear

. import delimited using "https://drive.google.com/u/0/uc?id=18GB67CPDbmGe
> 1RHKcHgGFu-ykDhvHYNt&export=download"
(3 vars, 173 obs)

. rename v1 state

. rename v2 income

. rename v3 year

. drop if income==.
(71 observations deleted)

. replace state = substr(state, length(state) - 6, .) if strpos(state, "Al
> abama") > 0
(1 real change made)

. save "income.dta", replace
file income.dta saved
```

```
.
. *** merge datasets.
. merge 1:1 state year using "smoking rate.dta", nogen

    Result                           # of obs.
    -----------------------------------------
    not matched                             0
    matched                               102
    -----------------------------------------

. merge 1:1 state year using "how to quit smoking.dta", nogen

    Result                           # of obs.
    -----------------------------------------
    not matched                             0
    matched                               102
    -----------------------------------------

.
. *** drop percentage characters in the smoke variable and destring.
. split smoke, parse(%) gen(smoke_n)
variable created as string:
smoke_n1

. destring smoke_n, replace
smoke_n1: all characters numeric; replaced as double

. drop smoke

. rename smoke_n1 smoke

.
. *** label variables.
. label var income "Real Median Household Income"

. label var quitsmoke "Relative Interest in Searching 'how to quit smoking
> '"

. label var smoke "Adult Smoking Rate"

.
. *** save as merged dataset.
. save "merged.dta", replace
file merged.dta saved

.
.
. ****************************************************************
. *** DATA ANALYSIS & VISUALIZATION
. ****************************************************************
.
. *** re Objective 1, regress smoke on quitsmoke, without controls.
. reg smoke quitsmoke if (year==2018), r

Linear regression                               Number of obs     =
>   42
                                                F(1, 40)          =
> 4.36
                                                Prob > F          =     0.
> 0432
                                                R-squared         =     0.
> 2109
                                                Root MSE          =     3.
> 3233
```

```
------------------------------------------------------------------------------
> ----
             |               Robust
       smoke |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Inter
> val]
-------------+----------------------------------------------------------------
> ----
   quitsmoke |   .1246767   .0597103     2.09   0.043     .0039977    .245
> 3557
       _cons |    11.3885   2.669797     4.27   0.000     5.992634    16.7
> 8436
------------------------------------------------------------------------------
> ----

. dis "Adjusted Rsquared = " e(r2_a)
Adjusted Rsquared = .19113226

.
. reg smoke quitsmoke if (year==2020), r

Linear regression                               Number of obs     =
>   42
                                                F(1, 40)          =      3
> 3.83
                                                Prob > F          =     0.
> 0000
                                                R-squared         =     0.
> 3216
                                                Root MSE          =     3.
> 1147

------------------------------------------------------------------------------
> ----
             |               Robust
       smoke |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Inter
> val]
-------------+----------------------------------------------------------------
> ----
   quitsmoke |   .1695875   .0291584     5.82   0.000     .1106562    .228
> 5188
       _cons |   8.134909   1.405894     5.79   0.000     5.293491    10.9
> 7633
------------------------------------------------------------------------------
> ----

. dis "Adjusted Rsquared = " e(r2_a)
Adjusted Rsquared = .30468591

.
. *** re Objective 1, regress smoke on quitsmoke, controlling for income.
. reg smoke quitsmoke income if (year==2018), r

Linear regression                               Number of obs     =
>   42
                                                F(2, 39)          =      2
> 9.16
                                                Prob > F          =     0.
> 0000
                                                R-squared         =     0.
> 6155
                                                Root MSE          =     2.
> 3494
```

```
------------------------------------------------------------------------------
> ----
             |               Robust
       smoke |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Inter
> val]
-------------+----------------------------------------------------------------
> ----
   quitsmoke |   .0842838    .045586     1.85   0.072    -.0079226    .176
> 4903
      income |  -.0002029   .0000281    -7.22   0.000    -.0002597    -.00
> 0146
       _cons |   27.33031   2.973456     9.19   0.000     21.31593    33.3
> 4469
------------------------------------------------------------------------------
> ----

. dis "Adjusted Rsquared = " e(r2_a)
Adjusted Rsquared = .59575328

.
. reg smoke quitsmoke income if (year==2020), r

Linear regression                               Number of obs     =
>   42
                                                F(2, 39)          =      6
> 8.75
                                                Prob > F          =     0.
> 0000
                                                R-squared         =     0.
> 7645
                                                Root MSE          =     1.
> 8586

------------------------------------------------------------------------------
> ----
             |               Robust
       smoke |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Inter
> val]
-------------+----------------------------------------------------------------
> ----
   quitsmoke |   .1062139   .0196313     5.41   0.000     .0665058    .145
> 9219
      income |  -.0001981   .0000194   -10.20   0.000    -.0002374   -.000
> 1588
       _cons |   25.18969   1.712767    14.71   0.000     21.72529    28.6
> 5408
------------------------------------------------------------------------------
> ----

. dis "Adjusted Rsquared = " e(r2_a)
Adjusted Rsquared = .7524217

.
. *** re Objective 2, regress smoke on income, without controls.
. reg smoke income if (year==2018), r

Linear regression                               Number of obs     =
>   51
                                                F(1, 49)          =      4
> 9.52
                                                Prob > F          =     0.
> 0000
                                                R-squared         =     0.
> 4990
                                                Root MSE          =      2
> .495
```

```
------------------------------------------------------------------------------
> ----
             |               Robust
       smoke |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Inter
> val]
-------------+----------------------------------------------------------------
> ----
      income |  -.0002204   .0000313    -7.04   0.000    -.0002834   -.000
> 1575
       _cons |   32.59084   2.263648    14.40   0.000     28.04187    37.1
> 3981
------------------------------------------------------------------------------
> ----

. dis "Adjusted Rsquared = " e(r2_a)
Adjusted Rsquared = .48876427

.
. reg smoke income if (year==2020), r

Linear regression                               Number of obs     =
>   51
                                                F(1, 49)          =      9
> 1.64
                                                Prob > F          =     0.
> 0000
                                                R-squared         =     0.
> 6151
                                                Root MSE          =     2.
> 2142

------------------------------------------------------------------------------
> ----
             |               Robust
       smoke |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Inter
> val]
-------------+----------------------------------------------------------------
> ----
      income |  -.0002239   .0000234    -9.57   0.000    -.0002709   -.000
> 1769
       _cons |   32.03611   1.725226    18.57   0.000     28.56914    35.5
> 0308
------------------------------------------------------------------------------
> ----

. dis "Adjusted Rsquared = " e(r2_a)
Adjusted Rsquared = .60728171

.
. *** scatter plots of the univariate regressions. show the fitted values
> with a confidence interval for the mean.
. graph twoway (lfitci smoke quitsmoke if (year==2018)) (scatter smoke qui
> tsmoke if (year==2018)), title("Smoking rate and Search interest, 2018 D
> ata") ytitle(Adult Smoking Rate)

. graph twoway (lfitci smoke quitsmoke if (year==2020)) (scatter smoke qui
> tsmoke if (year==2020)), title("Smoking rate and Search interest, 2020 D
> ata") ytitle(Adult Smoking Rate)

.
```

```
. graph twoway (lfitci smoke income if (year==2018)) (scatter smoke income
>  if (year==2018)), title("Smoking rate and Income, 2018 Data") ytitle(Ad
> ult Smoking Rate)

. graph twoway (lfitci smoke income if (year==2020)) (scatter smoke income
>  if (year==2020)), title("Smoking rate and Income, 2020 Data") ytitle(Ad
> ult Smoking Rate)

.
. *** for a differences comparison, construct data on changes in the varia
> bles between 2018 and 2020.
. preserve

.
. gen dsmoke = smoke-smoke[_n-1]
(1 missing value generated)

. gen dquitsmoke = quitsmoke-quitsmoke[_n-1]
(27 missing values generated)

. gen dincome = income-income[_n-1]
(1 missing value generated)

.
. keep if year==2020
(51 observations deleted)

.
. *** regress the changes in smoking rate on the changes in search interes
> t in how to quit smoking.
. reg dsmoke dquitsmoke, r

Linear regression                               Number of obs     =
>   42
                                                F(1, 40)          =
> 0.98
                                                Prob > F          =     0.
> 3277
                                                R-squared         =     0.
> 0281
                                                Root MSE          =     1.
> 1398

------------------------------------------------------------------------------
> ----
             |               Robust
      dsmoke |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Inter
> val]
-------------+----------------------------------------------------------------
> ----
  dquitsmoke |  -.0158479   .0159949    -0.99   0.328    -.0481748    .016
> 4791
       _cons |  -1.479809   .1808233    -8.18   0.000    -1.845267   -1.11
> 4352
------------------------------------------------------------------------------
> ----

. dis "Adjusted Rsquared = " e(r2_a)
Adjusted Rsquared = .00375718
```

```
.
. *** control for the changes in income.
. reg dsmoke dquitsmoke dincome, r

Linear regression                               Number of obs     =
>   42
                                                F(2, 39)          =
> 0.70
                                                Prob > F          =     0.
> 5008
                                                R-squared         =     0.
> 0421
                                                Root MSE          =     1.
> 1459

------------------------------------------------------------------------------
> ----
             |               Robust
      dsmoke |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Inter
> val]
-------------+----------------------------------------------------------------
> ----
  dquitsmoke |  -.0156099   .0157181    -0.99   0.327    -.0474027    .016
> 1829
     dincome |   .0000397   .0000547     0.73   0.472    -.0000708    .000
> 1503
       _cons |  -1.574963   .2213552    -7.12   0.000    -2.022696    -1.1
> 2723
------------------------------------------------------------------------------
> ----

. dis "Adjusted Rsquared = " e(r2_a)
Adjusted Rsquared = -.0069779

.
. *** regress the changes in smoking rate on the changes in income.
. reg dsmoke dincome, r

Linear regression                               Number of obs     =
>   51
                                                F(1, 49)          =
> 0.14
                                                Prob > F          =     0.
> 7147
                                                R-squared         =     0.
> 0021
                                                Root MSE          =     1.
> 1629

------------------------------------------------------------------------------
> ----
             |               Robust
      dsmoke |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Inter
> val]
-------------+----------------------------------------------------------------
> ----
     dincome |   .0000125   .0000339     0.37   0.715    -.0000557    .000
> 0806
       _cons |  -1.355009   .1924018    -7.04   0.000    -1.741655   -.968
> 3635
------------------------------------------------------------------------------
> ----
```

```
. dis "Adjusted Rsquared = " e(r2_a)
Adjusted Rsquared = -.018288

.
. *** visual display of the univariate regressions.
. graph twoway (lfitci dsmoke dquitsmoke) (scatter dsmoke dquitsmoke), tit
> le("Smoking rate and Search Interest, Differenced Data") xtitle (Changes
>  in Search Interest) ytitle(Changes in Adult Smoking Rate)

. graph twoway (lfitci dsmoke dincome) (scatter dsmoke dincome), title("Sm
> oking rate and Income, Differenced Data") xtitle (Changes in Income) yti
> tle(Changes in Adult Smoking Rate)

.
. restore

.
. *** Comments on findings:
. *
. *       (1) Search interest in "how to quit smoking"
. *
. *               With or without controlling for income, search interest
> in "how to quit smoking" is positively associated with smoking rate, i.e
> ., the greater the proportion of smokers within a state, the greater the
>  interest in searching "how to quit smoking." The positive coefficient o
> n search interest (controlling for income) is significant at the 10% lev
> el for the 2018 data and 1% level for the 2020 data, but there is likely
>  substantial OVB not captured by the income control.
. *
. *               By regressing the changes in smoking rate on the changes
>  in search interest, we control for omitted variables that: (i) influenc
> e smoking rate and are systematically correlated with search interest(no
> t captured by the income control), and (ii) did not change from 2018 to
> 2020 within each state (e.g., general cultural attitudes toward smoking,
>  the presence of a large tobacco industry). Importantly, we do not contr
> ol for omitted variables that changed over time (e.g., those related to
> COVID). The regression results suggest that, holding constantthe state-f
> ixed factors, search interest in how to quit smoking hasa negative assoc
> iation with smoking rate - but the coefficient on theeffect of search in
> terest is not significant at the 10% level.
. *
. *       (2) Income
. *
. *               As expected, the univariate regression of smoke on incom
> e indicates a negative association between smoking rate and income, and
> the coefficient on the effect of income is significant at the 1% level f
> or both years' data.  But there is likely substantial OVB from factors t
> hat influence smoking rate and are correlated with income. When we contr
> ol for factors that remained constant over time within each state(by reg
> ressing the changes in smoking rate on the changes in income), the resul
> ts suggest that income has a weakly positive association with smoking ra
> te - but the coefficient on the effect of changes in income is not signi
> ficant at any conventional significance level.
. *
.
.
.
end of do-file

. log close
      name:  <unnamed>
       log:  C:\Users\brocc\Documents\Coding\code sample\smoking\newlog.lo
> g
  log type:  text
 closed on:  19 Jun 2023, 20:33:16
-------------------------------------------------------------------------------
```