```
      name:  <unnamed>
       log:  /Users/carsoncrenshaw/Library/CloudStorage/OneDrive-Universityof
> Virginia/RMDA I/RMDA1_Homework5_log.smcl
  log type:  smcl
 opened on:   3 Nov 2023, 15:56:21

 1 .
 2 . *********
 3 . *       1: Homework Code
 4 . *********
 5 .
 6 .
 7 . *QUESTION 1
 8 .
 9 . use "${classpath}nlsw88.dta", clear
   (NLSW, 1988 extract)

10 . *1A
11 . reg wage ttl_exp // reg Y(outcome of treatment) X

      Source │       SS           df       MS      Number of obs   =     2,24
> 6
─────────────┼──────────────────────────────────   F(1, 2244)      =    170.1
> 4
       Model │  5241.29609         1  5241.29609   Prob > F        =    0.000
> 0
    Residual │  69126.6713     2,244   30.805112   R-squared       =    0.070
> 5
─────────────┼──────────────────────────────────   Adj R-squared   =    0.070
> 1
       Total │  74367.9674     2,245  33.1260434   Root MSE        =    5.550
> 2

─────────────┬────────────────────────────────────────────────────────────────
> ─
        wage │ Coefficient  Std. err.      t    P>|t|     [95% conf. interval
> ]
─────────────┼────────────────────────────────────────────────────────────────
> ─
     ttl_exp │   .3314291   .0254087    13.04   0.000     .2816021    .381256
> 2
       _cons │   3.612492   .3393469    10.65   0.000     2.947026    4.27795
> 9
─────────────┴────────────────────────────────────────────────────────────────
> ─
```

```
12 .
13 . *1B
14 . generate ttl_exp_decade = ttl_exp / 10 // Changing units proved by creating
   >  a new variable

15 . reg wage ttl_exp_decade

      Source │       SS           df       MS      Number of obs   =     2,24
   > 6
   ──────────┼──────────────────────────────────   F(1, 2244)      =    170.1
   > 4
       Model │  5241.29609         1  5241.29609   Prob > F        =    0.000
   > 0
    Residual │  69126.6713     2,244   30.805112   R-squared       =    0.070
   > 5
   ──────────┼──────────────────────────────────   Adj R-squared   =    0.070
   > 1
       Total │  74367.9674     2,245  33.1260434   Root MSE        =    5.550
   > 2

   ──────────┼──────────────────────────────────────────────────────────────────
   > ──
        wage │ Coefficient  Std. err.      t    P>|t|     [95% conf. interv
   > al]
   ──────────┼──────────────────────────────────────────────────────────────────
   > ──
   ttl_exp_decade │   3.314291   .2540872    13.04   0.000     2.816021    3.812
   > 562
          _cons │   3.612492   .3393469    10.65   0.000     2.947026    4.277
   > 959
   ──────────┼──────────────────────────────────────────────────────────────────
   > ──
```

```
16 .
17 . *1C
18 . generate wage_weekly = wage*40 // 8 hours a day, 5 days = 40hr work week

19 . reg wage_weekly ttl_exp

      Source │       SS           df       MS      Number of obs   =     2,24
> 6
─────────────┼──────────────────────────────────   F(1, 2244)      =    170.1
> 4
       Model │  8386073.73         1  8386073.73   Prob > F        =    0.000
> 0
    Residual │   110602676     2,244  49288.1801   R-squared       =    0.070
> 5
─────────────┼──────────────────────────────────   Adj R-squared   =    0.070
> 1
       Total │   118988750     2,245  53001.6703   Root MSE        =    222.0
> 1

─────────────┼──────────────────────────────────────────────────────────────
> ─
 wage_weekly │ Coefficient  Std. err.      t    P>|t|     [95% conf. interval
> ]
─────────────┼──────────────────────────────────────────────────────────────
> ─
     ttl_exp │   13.25717   1.016349    13.04   0.000     11.26408    15.2502
> 5
       _cons │   144.4997   13.57388    10.65   0.000      117.881    171.118
> 4
─────────────┼──────────────────────────────────────────────────────────────
> ─

20 .
21 .
```

```
22 . *QUESTION 2
23 .
24 . use "${classpath}tracking.dta", clear

25 . *2A
26 . describe, short

Contains data from /Users/carsoncrenshaw/Library/CloudStorage/OneDrive-Univer
> sityofVirginia/RMDA I/tracking.dta
 Observations:         6,431
    Variables:            13                  21 Oct 2022 19:58
Sorted by:

27 .
28 . *2B
29 . codebook schoolid

--------------------------------------------------------------------------------
schoolid                                                  ID of primary school
--------------------------------------------------------------------------------

                  Type: Numeric (float)

                 Range: [430,1020]                    Units: 1
         Unique values: 111                       Missing .: 0/6,431

                  Mean: 776.686
             Std. dev.: 174.801

           Percentiles:     10%       25%       50%       75%       90%
                            485       687       790       938       974

30 . distinct schoolid

             |        Observations
             |       total   distinct
-------------+----------------------
    schoolid |        6431        111
```

```
31 .
32 . *2C
33 . bysort schoolid: gen order = _n //should generate the same number for all v
   > alues of schoolid

34 .
35 . sum schoolid if order == 1
```

Variable	Obs	Mean	Std. dev.	Min	Max
schoolid	**111**	**777.8559**	**173.8787**	**430**	**1020**

```
36 .
37 . *    Variable |        Obs        Mean    Std. dev.       Min        Max
38 . *-------------+---------------------------------------------------------
39 . *    schoolid |        111    777.8559    173.8787        430       1020
40 .
41 . tab tracked if order == 1 //tab track should then show school level
```

school assigned to track students	Freq.	Percent	Cum.
0	**51**	**45.95**	**45.95**
1	**60**	**54.05**	**100.00**
Total	**111**	**100.00**	

```
42 .
43 . *     school |
44 . *assigned to |
45 . *      track |
46 . *   students |      Freq.     Percent        Cum.
```

```
47 . *------------+-----------------------------------
48 . *          0 |         51       45.95       45.95
49 . *          1 |         60       54.05      100.00
50 . *------------+-----------------------------------
51 . *      Total |        111      100.00
52 .
53 . *2D
54 . reg female tracked, cluster(schoolid) // reg baseline characteristic (Y) x
   > variable (X), clustered at the school level

   Linear regression                               Number of obs     =      6,43
   > 1
                                                   F(1, 110)         =       0.1
   > 7
                                                   Prob > F          =     0.684
   > 4
                                                   R-squared         =     0.000
   > 0
                                                   Root MSE          =     .5000
   > 2

                               (Std. err. adjusted for 111 clusters in schoolid
   > )
   -------------+-------------------------------------------------------------
   > -
                |               Robust
         female | Coefficient  std. err.      t    P>|t|     [95% conf. interval
   > ]
   -------------+-------------------------------------------------------------
   > -
        tracked |   .0051266   .0125799     0.41   0.684    -.0198039    .030057
   > 1
          _cons |     .48989   .0092499    52.96   0.000     .4715589    .508221
   > 2
   -------------+-------------------------------------------------------------
   > -
```

```
55 . // can also look at the difference without clustering: reg female tracked
56 .
57 .
58 . *QUESTION 3
59 .
60 . *3A
61 . // variable that represents missing values as 1
62 . generate missing_scores = 0

63 . replace missing_scores = 1 if score_lit_30months ==. | score_math_30months
   > ==.
   (1,408 real changes made)

64 . tab missing_scores

   missing_sco |
           res |      Freq.     Percent        Cum.
   ------------+-----------------------------------
             0 |      5,023       78.11       78.11
             1 |      1,408       21.89      100.00
   ------------+-----------------------------------
         Total |      6,431      100.00

65 .
66 . reg missing_scores tracked

         Source |       SS           df       MS      Number of obs   =     6,43
   > 1
   -------------+----------------------------------   F(1, 6429)      =      0.1
   > 2
          Model |  .021316488         1  .021316488   Prob > F        =    0.724
   > 1
       Residual |  1099.71185     6,429  .171054884   R-squared       =    0.000
   > 0
   -------------+----------------------------------   Adj R-squared   =   -0.000
   > 1
          Total |  1099.73317     6,430  .171031597   Root MSE        =    .4135
   > 9
```

```
------------------------------------------------------------------------------
missing_sc~s | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     tracked |  -.0036692   .0103941    -0.35   0.724    -.0240451    .016706
       _cons |   .2210004   .0077897    28.37   0.000       .20573    .236270
------------------------------------------------------------------------------

. 
. // variable that represents missing values as 0
. generate complete_scores = 1

. replace complete_scores = 0 if score_lit_30months ==. | score_math_30months
>  ==.
(1,408 real changes made)

. tab complete_scores

complete_sc |
       ores |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      1,408       21.89       21.89
          1 |      5,023       78.11      100.00
------------+-----------------------------------
      Total |      6,431      100.00

. 
. reg complete_scores tracked

      Source |       SS           df       MS      Number of obs   =     6,431
-------------+----------------------------------   F(1, 6429)      =       0.10
       Model |  .021316488         1  .021316488   Prob > F        =     0.7241
    Residual |  1099.71185     6,429  .171054884   R-squared       =     0.0000
-------------+----------------------------------   Adj R-squared   =    -0.0001
       Total |  1099.73317     6,430  .171031597   Root MSE        =     .4135
```

```
------------------------------------------------------------------------------
> ------------------------------------------------------------------------------
complete_s~s |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
> ]
-------------+----------------------------------------------------------------
> ------------------------------------------------------------------------------
     tracked |   .0036692   .0103941     0.35   0.724    -.0167066    .024045
> 1
       _cons |   .7789996   .0077897   100.00   0.000     .7637293      .7942
> 7
------------------------------------------------------------------------------
> ------------------------------------------------------------------------------

74 .
75 . //can also do it clustered: reg missing_scores tracked, cluster(schoolid)
76 .
77 . *3B
78 . count if complete_scores ==1
     5,023

79 .
80 . distinct schoolid if complete_scores ==1

             |        Observations
             |       total   distinct
-------------+----------------------
    schoolid |        5023        111

81 .
82 . sum score_math_30months if complete_scores ==1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
s~h_30months |      5,023    1.31e-09           1  -2.091329   2.550627
```

```
83 .
84 .
85 . *QUESTION 4
86 .
87 . drop if complete_scores ==0 // remove observations with missing data
   (1,408 observations deleted)

88 . *4A
89 . reg score_lit_30months tracked, cluster(schoolid)

   Linear regression                               Number of obs     =      5,02
   > 3
                                                   F(1, 110)         =       3.9
   > 8
                                                   Prob > F          =     0.048
   > 6
                                                   R-squared         =     0.005
   > 7
                                                   Root MSE          =     .9972
   > 7

                               (Std. err. adjusted for 111 clusters in schoolid
   > )
   ──────────────┬──────────────────────────────────────────────────────────────
   > ─
                 │               Robust
   s~t_30months  │ Coefficient  std. err.      t    P>|t|     [95% conf. interval
   > ]
   ──────────────┼──────────────────────────────────────────────────────────────
   > ─
        tracked  │   .1521231   .0762934     1.99   0.049     .0009274    .303318
   > 7
          _cons  │  -.0845058   .0561763    -1.50   0.135     -.195834    .026822
   > 4
   ──────────────┴──────────────────────────────────────────────────────────────
   > ─
```

```
90 .
91 . *4B
92 . reg score_lit_30months tracked

      Source │       SS           df       MS      Number of obs   =     5,02
   > 3
   ─────────┼──────────────────────────────────   F(1, 5021)      =     28.7
   > 6
       Model │  28.6012552         1  28.6012552   Prob > F        =    0.000
   > 0
    Residual │  4993.59245     5,021  .994541417   R-squared       =    0.005
   > 7
   ─────────┼──────────────────────────────────   Adj R-squared   =    0.005
   > 5
       Total │  5022.19371     5,022  1.00003857   Root MSE        =    .9972
   > 7

   ─────────┼──────────────────────────────────────────────────────────────────
   > ─
   s~t_30months │ Coefficient  Std. err.      t    P>|t|     [95% conf. interval
   > ]
   ─────────┼──────────────────────────────────────────────────────────────────
   > ─
        tracked │   .1521231    .028367     5.36   0.000     .0965113    .207734
   > 8
          _cons │  -.0845058   .0212812    -3.97   0.000    -.1262262   -.042785
   > 5
   ─────────┼──────────────────────────────────────────────────────────────────
   > ─

93 .
94 .
95 . *********
96 . *       2: Close log file
97 . *********
98 .
99 .         log close
         name:  <unnamed>
          log:  /Users/carsoncrenshaw/Library/CloudStorage/OneDrive-Universityof
   > Virginia/RMDA I/RMDA1_Homework5_log.smcl
     log type:  smcl
    closed on:   3 Nov 2023, 15:56:21
   ──────────────────────────────────────────────────────────────────────────────
```