```
    --------------------------------------------------------------------------------
> --------------------------------------------------------------------------------
> -----------------
      name:  <unnamed>
       log:  C:\Users\saiomkark\OneDrive - The University of Chicago\AdvStats\PS7\Sai_
> Omkar_K_PS7_Q4_Q5.log
  log type:  text
 opened on:   8 Dec 2021, 16:43:38


.
. *Loading data and removing first line using names
.
. use "C:\Users\saiomkark\OneDrive - The University of Chicago\AdvStats\PS7\Homework_7
> .dta"


.
.
.
.
. *analyze the data using summarize to check for NAs etc..
. summarize

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        year |     13,401        2019           0       2019       2019
      sample |     13,401      201901           0     201901     201901
      serial |     13,401     1338026     10413.11    1318154    1355966
     cbserial |    13,401     2.02e+12     2236478    2.02e+12   2.02e+12
        hhwt |     13,401    100.1622     81.10885          1       1069
-------------+--------------------------------------------------------
     cluster |     13,401     2.02e+12    104131.1    2.02e+12   2.02e+12
     met2013 |     13,401       47260           0      47260      47260
      strata |     13,401     5163409    125302.8    5114551    5500251
          gq |     13,401    1.158645    .6454618          1          4
      pernum |     13,401    1.926274    1.203637          1         13
-------------+--------------------------------------------------------
       perwt |     13,401    104.0375    85.72637          1       1163
         sex |     13,401     1.51429    .4998144          1          2
         age |     13,401    42.18126    23.31533          0         94
        race |     13,401    1.745616    1.647789          1          9
       raced |     13,401    175.7108    168.3765        100        990
-------------+--------------------------------------------------------
      hispan |     13,401           0           0          0          0
     hispand |     13,401           0           0          0          0
        educ |     13,401     6.58712    3.176341          0         11
       educd |     13,401    68.27946    31.67752          1        116
      inctot |     13,401     1516745     3532754      -2000    9999999
-------------+--------------------------------------------------------
      incwage |    13,401    185282.5    355839.9          0     999999
     qincbus |     13,401     .161779    .7880291          0          4
     qincinvs |    13,401     .311917    1.072596          0          4
     qincothe |    13,401    .2647564    .9944865          0          4
     qincreti |    13,401    .2913215     1.03947          0          4
-------------+--------------------------------------------------------
      qincss |     13,401    .3077382     1.06599          0          4
     qinctot |     13,401    .7002463    1.520136          0          4
     qincwage |    13,401    .3751959    1.166239          0          4
     qincwelf |    13,401    .2563988    .9797583          0          4


.
```

```
. *Inctot has negative values, which it cannot be. Hence keeping only those where inct
> ot is greater than equal to 0
. keep if inctot>= 0
(12 observations deleted)

.

.
. *check the data if changes are applied
. summarize

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        year |     13,389        2019           0       2019       2019
      sample |     13,389      201901           0     201901     201901
      serial |     13,389     1338026      10411.9    1318154    1355966
     cbserial |    13,389     2.02e+12     2236096    2.02e+12   2.02e+12
        hhwt |     13,389    100.1834    81.13079          1       1069
-------------+--------------------------------------------------------
     cluster |     13,389     2.02e+12      104119    2.02e+12   2.02e+12
     met2013 |     13,389       47260           0      47260      47260
      strata |     13,389     5163393    125284.8    5114551    5500251
          gq |     13,389    1.158563    .6452679          1          4
      pernum |     13,389    1.926432    1.203971          1         13
-------------+--------------------------------------------------------
       perwt |     13,389    104.0596    85.75109          1       1163
         sex |     13,389    1.514303    .4998141          1          2
         age |     13,389    42.17529    23.31629          0         94
        race |     13,389     1.74606     1.64841          1          9
       raced |     13,389    175.7563    168.4401        100        990
-------------+--------------------------------------------------------
      hispan |     13,389           0           0          0          0
     hispand |     13,389           0           0          0          0
        educ |     13,389    6.586452     3.17709          0         11
       educd |     13,389    68.27276     31.6854          1        116
      inctot |     13,389     1518106     3534044          0    9999999
-------------+--------------------------------------------------------
     incwage |     13,389    185448.5    355956.1          0     999999
     qincbus |     13,389    .1616252    .7876703          0          4
    qincinvs |     13,389    .3118978    1.072566          0          4
    qincothe |     13,389    .2646949    .9943792          0          4
    qincreti |     13,389    .2912839    1.039408          0          4
-------------+--------------------------------------------------------
      qincss |     13,389    .3074165    1.065479          0          4
     qinctot |     13,389    .7002763    1.520162          0          4
    qincwage |     13,389    .3752334    1.166291          0          4
    qincwelf |     13,389    .2563298    .9796356          0          4

.
. *filtering data to required age
.
. keep if age >= 25 & age <= 55
(8,312 observations deleted)


.
. * Divide people into six educational categories: 1) less than high school including
> GED recipients. 2) exactly a high school degree. 3) some college but no degree 4) as
> sociates degree 5)
>  exactly bachelors degree 6) more than a bachelors degree.
.
. codebook educd,tab(25)


--------------------------------------------------------------------------------------
> -------------------------------------------------------------------------------------
> -----------------
educd
>                                                                educational attainment
> [detailed version]
--------------------------------------------------------------------------------------
> -------------------------------------------------------------------------------------
> -----------------
```

```
                type:   numeric (int)
                label:   EDUCD

                range:   [2,116]                        units:  1
        unique values:   21                          missing .:  0/5,077

            tabulation:   Freq.    Numeric   Label
                             34         2   no schooling completed
                              1        11   nursery school, preschool
                              1        14   grade 1
                              1        17   grade 4
                              3        22   grade 5
                              3        23   grade 6
                              3        25   grade 7
                             11        26   grade 8
                             23        30   grade 9
                             42        40   grade 10
                             72        50   grade 11
                             57        61   12th grade, no diploma
                            809        63   regular high school diploma
                            237        64   ged or alternative credential
                            374        65   some college, but less than 1
                                            year
                            784        71   1 or more years of college
                                            credit, no degree
                            590        81   associate's degree, type not
                                            specified
                          1,208       101   bachelor's degree
                            644       114   master's degree
                            111       115   professional degree beyond a
                                            bachelor's degree
                             69       116   doctoral degree

. gen edu_cat = "less than high school including GED recipients"

. replace edu_cat = "exactly a high school degree" if educd == 63
(809 real changes made)

. replace edu_cat = "some college but no degree" if educd == 64 | educd == 65 | educd
> == 71
(1,395 real changes made)

. replace edu_cat = "associates degree" if educd == 81
(590 real changes made)

. replace edu_cat = "exactly bachelor's degree" if educd == 101
(1,208 real changes made)

. replace edu_cat = "more than a bachelor's degree" if educd == 114 | educd == 115 | e
> ducd == 116
(824 real changes made)

. br edu_cat educd

.
.
. *Arranging the data according to the racial groups into White -> 1, African American
>  -> 2, Other -> 3
. gen race_cat = "Other race"
```

```
. replace race_cat = "White" if (race == 1)
(3,404 real changes made)

. replace race_cat = "African American" if (race == 2)
variable race_cat was str10 now str16
(1,224 real changes made)

. br race_cat race

.

.
. *check the data if changes are applied
. summarize

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        year |      5,077        2019           0       2019       2019
      sample |      5,077      201901           0     201901     201901
      serial |      5,077     1338384    10251.04    1318209    1355966
     cbserial |      5,077     2.02e+12     1880145    2.02e+12   2.02e+12
        hhwt |      5,077    107.7051     87.7203          1       1069
-------------+--------------------------------------------------------
     cluster |      5,077     2.02e+12    102510.4   2.02e+12   2.02e+12
     met2013 |      5,077       47260           0      47260      47260
      strata |      5,077     5167959    130459.8    5114551    5500251
          gq |      5,077    1.099074    .4956798          1          4
      pernum |      5,077    1.558598    .8018666          1         10
-------------+--------------------------------------------------------
       perwt |      5,077    113.9677    92.95438          1       1068
         sex |      5,077    1.514674    .4998339          1          2
         age |      5,077    39.82903    9.222254         25         55
        race |      5,077    1.723065    1.580555          1          9
       raced |      5,077    173.6262    162.2136        100        990
-------------+--------------------------------------------------------
      hispan |      5,077           0           0          0          0
     hispand |      5,077           0           0          0          0
        educ |      5,077    8.046878    2.168386          0         11
       educd |      5,077    82.75458    21.73427          2        116
      inctot |      5,077    51131.76    58796.41          0     946000
-------------+--------------------------------------------------------
      incwage |      5,077    45799.74    52279.28          0     516000
      qincbus |      5,077    .1709671    .8091771          0          4
     qincinvs |      5,077    .3466614    1.125487          0          4
     qincothe |      5,077    .3246011    1.092371          0          4
     qincreti |      5,077    .3285405     1.09839          0          4
-------------+--------------------------------------------------------
      qincss |      5,077    .3182982     1.08264          0          4
     qinctot |      5,077     .768958    1.576396          0          4
     qincwage |      5,077    .4782352    1.297907          0          4
     qincwelf |      5,077    .3049045    1.061543          0          4
      edu_cat |          0
-------------+--------------------------------------------------------
     race_cat |          0

.
. *Question 4A - A Calculate the income from wages and salaries for those with positiv
> e earn-ings by sex and race;
.
. tab sex race_cat, sum(inctot)

    Means, Standard Deviations and Frequencies of total personal income
```

```
           |              race_cat
       sex | African..  Other r..      White |     Total
-----------+--------------------------------+----------
      male | 42003.425   63893.86  69524.396 | 62968.206
           | 49604.631  70234.817   68727.85 |  66080.68
           |       543        215       1706 |      2464
-----------+--------------------------------+----------
    female | 36249.148  34914.744   42159.34 | 39970.256
           | 43607.488  59261.278  48471.595 | 48412.827
           |       681        234       1698 |      2613
-----------+--------------------------------+----------
     Total | 38801.904  48791.158  55874.025 | 51131.759
           | 46432.465  66278.918  61037.065 | 58796.406
           |      1224        449       3404 |      5077


.
. *Question 4B Calculate the income from wages and salaries for those with positive ea
> rn-ings by sex and race using the weights "perwt."
.
. tab sex race_cat [aw=perwt], sum(inctot)

    Means, Standard Deviations, Frequencies and Number of Observations
                     of total personal income

           |              race_cat
       sex | African..  Other r..      White |     Total
-----------+--------------------------------+----------
      male | 45412.393  63820.686  66461.351 | 59704.029
           | 48525.631  75628.249  64838.563 |  62079.28
           |     87693      25069     170197 |    282959
           |       543        215       1706 |      2464
-----------+--------------------------------+----------
    female | 38693.859  33769.194  42426.886 | 40322.411
           | 37456.074  61155.204  48790.921 | 46513.905
           |    104482      26816     164357 |    295655
           |       681        234       1698 |      2613
-----------+--------------------------------+----------
     Total |  41759.65  48289.014  54653.892 | 49800.583
           | 42975.348  70084.303  58750.775 | 55528.658
           |    192175      51885     334554 |    578614
           |      1224        449       3404 |      5077


.
.
. ********************************************************************************
.
.
.
. *Question 5 5. Use the same data set and setup as in problem 4. Create an age catego
> ry variable with categories, 25 -29, 30 -34, 35 -39, 40 -44, 45 -49, and 50 -55. The
>  variable "qincwag
> e" is equal to four when the data are suppressed. Using our race, age, sex, and educ
> ation categories along with the sample weights, construct IPW weights to account for
>  the imputed incom
> e. Calculate income from wages and salaries by sex and race using the IPW weights. C
> ompare these means to those in problem 4b.
.
.
.
. egen age_cat = cut(age), at(25,30,35,40,45,50,56)
```

```
.
.
. *
. * Construct missing indicator for income
. *
. gen miss = qinctot == 4

. * construct phat
. *
. * creates a unique variable for each age, sex  and education category
. *
. egen x = group(age_cat sex edu_cat)

. *
. * total people in each sex by age and education categories
. *
. egen den = sum(perwt), by(x)

. *
. * people with non-missing data
. *
. egen num =  sum((1-miss)*perwt), by(x)

. *
. * probability of responding
. gen phat = num/den

. *
. * generate new weights called "wt".  Again, note the use of perwts.
. *
. gen wt = perwt/phat if miss == 0
(976 missing values generated)

. *
. * Test to make sure weights work.  Fraction by age_cat should be the same
. *
. * using perwt
. tab age_cat [aw=perwt]

    age_cat |      Freq.      Percent        Cum.
------------+-----------------------------------
         25 |   996.01907        19.62       19.62
         30 |  867.921369        17.10       36.71
         35 |    856.71644       16.87       53.59
         40 | 742.2278583        14.62       68.21
         45 |  716.220489        14.11       82.31
         50 |  897.894774        17.69      100.00
------------+-----------------------------------
      Total |       5,077       100.00

. * using our new wts
. tab age_cat [aw=wt]

    age_cat |      Freq.      Percent        Cum.
------------+-----------------------------------
         25 |  804.544839        19.62       19.62
         30 |  701.072589        17.10       36.71
         35 |  692.021691        16.87       53.59
         40 |  599.542339        14.62       68.21
         45 |  578.534616        14.11       82.31
         50 |  725.283925        17.69      100.00
------------+-----------------------------------
      Total |       4,101       100.00
```

```
. * Test to make sure weights work.  Fraction by sex should be the same
. *
. * using perwt
. tab sex [aw=perwt]

        sex |      Freq.      Percent        Cum.
------------+-----------------------------------
       male |    2,482.8        48.90       48.90
     female |    2,594.2        51.10      100.00
------------+-----------------------------------
      Total |      5,077       100.00

.
. tab sex [aw=perwt], summarize(incwage)

            |        Summary of wage and salary income
        sex |        Mean   Std. Dev.       Freq.         Obs.
------------+-----------------------------------------------
       male |   53193.622   55944.252     282,959        2,464
     female |   36469.683   41128.633     295,655        2,613
------------+-----------------------------------------------
      Total |   44648.174   49641.774     578,614        5,077

. * using our new wts
. tab sex [aw=wt]

        sex |      Freq.      Percent        Cum.
------------+-----------------------------------
       male |  2,005.5077      48.90       48.90
     female |  2,095.4923      51.10      100.00
------------+-----------------------------------
      Total |      4,101       100.00

.
. *
.
.
. * Test to make sure weights work.  Fraction by edu_cat should be the same
. *
. * using perwt
. tab edu_cat [aw=perwt]

                           edu_cat |      Freq.     Percent        Cum.
-----------------------------------+-----------------------------------
                 associates degree | 609.848243       12.01       12.01
        exactly a high school degree | 896.131117       17.65       29.66
           exactly bachelor's degree | 1,137.6381       22.41       52.07
less than high school including GED r.. | 258.845275        5.10       57.17
          more than a bachelor's degree | 734.795928       14.47       71.64
            some college but no degree | 1,439.7413       28.36      100.00
-----------------------------------+-----------------------------------
                             Total |      5,077       100.00

. * using our new wts
. tab edu_cat [aw=wt]

                           edu_cat |      Freq.     Percent        Cum.
-----------------------------------+-----------------------------------
                 associates degree | 492.611317       12.01       12.01
        exactly a high school degree | 723.859305       17.65       29.66
           exactly bachelor's degree |918.9391461       22.41       52.07
less than high school including GED r.. | 209.084984        5.10       57.17
          more than a bachelor's degree |593.5391273       14.47       71.64
            some college but no degree |1,162.96612       28.36      100.00
-----------------------------------+-----------------------------------
                             Total |      4,101       100.00
```

```
.
. *Comparing total income with old weights and total income with new weights derived
.
. tab sex [aw=perwt], sum(inctot)

            |     Summary of total personal income
       sex |        Mean    Std. Dev.        Freq.         Obs.
-----------+-------------------------------------------------------
      male |   59704.029    62079.28      282,959        2,464
    female |   40322.411   46513.905      295,655        2,613
-----------+-------------------------------------------------------
     Total |   49800.583   55528.658      578,614        5,077

. tab sex [aw=wt], sum(inctot)

            |     Summary of total personal income
       sex |        Mean    Std. Dev.        Freq.         Obs.
-----------+-------------------------------------------------------
      male |   59285.076   60123.079      282,959        1,958
    female |   39065.801   44753.843      295,655        2,143
-----------+-------------------------------------------------------
     Total |   48953.612   53783.288      578,614        4,101

. tab sex race_cat [aw = perwt], summarize(incwage)

    Means, Standard Deviations, Frequencies and Number of Observations
                    of wage and salary income

            |             race_cat
       sex | African..  Other r..      White |     Total
-----------+----------------------------------+---------
      male | 39401.802   57386.27  59682.225 | 53193.622
           |  42742.53  75792.004  57259.879 | 55944.252
           |     87693      25069     170197 |    282959
           |       543        215       1706 |      2464
-----------+----------------------------------+---------
    female | 35560.856  28470.808  38352.498 | 36469.683
           | 37339.911   33307.85  44306.832 | 41128.633
           |    104482      26816     164357 |    295655
           |       681        234       1698 |      2613
-----------+----------------------------------+---------
     Total | 37313.551  42441.738  49203.528 | 44648.174
           | 39925.072  59583.763  52396.017 | 49641.774
           |    192175      51885     334554 |    578614
           |      1224        449       3404 |      5077

. tab sex race_cat [aw=wt], summarize(incwage)

    Means, Standard Deviations, Frequencies and Number of Observations
                    of wage and salary income

            |             race_cat
       sex | African..  Other r..      White |     Total
-----------+----------------------------------+---------
      male | 36499.868  56757.946  59096.046 | 53145.939
           | 39766.733  77737.458  54006.868 | 54253.197
           | 71930.644  24925.604  186102.75 |    282959
           |       342        169       1447 |      1958
-----------+----------------------------------+---------
    female | 35101.892   27075.94   36396.25 | 35128.939
           | 39224.498  34996.598  37801.494 | 38077.445
           | 92656.723  27333.416  175664.86 |    295655
           |       488        191       1464 |      2143
-----------+----------------------------------+---------
     Total | 35712.858   41233.15   48073.62 | 43939.774
           | 39444.136  61095.684  48190.356 | 47548.392
           | 164587.37   52259.02  361767.61 |    578614
           |       830        360       2911 |      4101
```

```
.
. *The means are lower using IPW weights than using "perwt". Almost similar, but not c
> lose
.
.
. log close
       name:  <unnamed>
        log:  C:\Users\saiomkark\OneDrive - The University of Chicago\AdvStats\PS7\Sai_
> Omkar_K_PS7_Q4_Q5.log
   log type:  text
 closed on:   8 Dec 2021, 16:43:38
--------------------------------------------------------------------------------------
> --------------------------------------------------------------------------------------
> ------------------
```