

```

-----
> -----
> -----
      name: <unnamed>
      log: C:\Users\saiomkark\OneDrive - The University of Chicago\AdvStats\PS7\Stat
> a_Sai_Omkar_K_PS7_Q1.log
      log type: text
      opened on: 8 Dec 2021, 16:45:18

```

```

.
. *Loading data and removing first line using names
.
. insheet using "C:\Users\saiomkark\OneDrive - The University of Chicago\AdvStats\PS7\
> ppha312x2021.csv", names clear
(14 vars, 10,149 obs)

```

```

.
.
. label data "Data is from IPUMS-USA restricted to Albuquerque,New Mexico (2018)"
.
. *use ppha312x2021
.
. *analyze the data using summarize to check for NAs etc..
. summarize

```

Variable	Obs	Mean	Std. Dev.	Min	Max
year	10,149	2019	0	2019	2019
statefip	0				
met2013	0				
perwt	10,149	115.5946	104.4547	2	1977
sex	0				
age	0				
race	0				
hispan	0				
bpl	0				
educd	0				
empstat	0				
uhrswork	0				
inctot	10,149	1743385	3748625	-6900	9999999
incwage	10,149	209523.8	379486.3	0	999999

```

.
. *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
(9 observations deleted)

```

```

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

```

Variable	Obs	Mean	Std. Dev.	Min	Max
year	10,140	2019	0	2019	2019
statefip	0				
met2013	0				
perwt	10,140	115.6266	104.4945	2	1977
sex	0				
age	0				
race	0				
hispan	0				
bpl	0				
educd	0				
empstat	0				
uhrswork	0				
inctot	10,140	1744935	3749928	0	9999999

```

incwage |      10,140      209709.4      379603.5          0      999999

.
. *Encoding the string variables so that we can analyze them in STATA
.
. encode sex, gen(sex_cat)
. encode race, gen(race_cat)
. encode empstat, gen(empstat_cat)
. encode age, gen(age_cat)
.
.
. codebook race_cat
-----
> -----
> -----
race_cat
>
>      (unlabeled)
-----
> -----
> -----

              type:  numeric (long)
              label:  race_cat

              range:  [1,9]              units:  1
unique values:  9              missing .:  0/10,140

tabulation:  Freq.   Numeric  Label
              9         1  American Indian or Alaska Native
              3,899     2  Black/African American/Negro
              9         3  Chinese
              3         4  Japanese
              61        5  Other Asian or Pacific Islander
              93        6  Other race, nec
              15        7  Three or more major races
              136       8  Two major races
              5,915     9  White

. gen is_AfricanAmerican = 0

. replace is_AfricanAmerican =1 if (race_cat == 2)
(3,899 real changes made)

.
. codebook sex_cat
-----
> -----
> -----
sex_cat
>
>      (unlabeled)
-----
> -----
> -----

              type:  numeric (long)
              label:  sex_cat

              range:  [1,2]              units:  1
unique values:  2              missing .:  0/10,140

```

```

      tabulation:   Freq.    Numeric    Label
                   5,317         1 Female
                   4,823         2 Male

. gen is_female = 0

. replace is_female = 1 if (sex_cat == 1)
(5,317 real changes made)

. codebook empstat_cat

-----
> -----
> -----
empstat_cat
> 
> (unlabeled)
-----
> -----
> -----

           type: numeric (long)
        label: empstat_cat

       range: [1,4]                units: 1
unique values: 4                  missing : 0/10,140

      tabulation:   Freq.    Numeric    Label
                   4,662         1 Employed
                   1,874         2 N/A
                   3,350         3 Not in labor force
                    254         4 Unemployed

. gen is_employed = 0

. replace is_employed = 1 if (empstat_cat == 1)
(4,662 real changes made)

. *limiting the sample to African American Women respondents, who are employed
. keep if (is_employed == 1 & is_female == 1 & is_AfricanAmerican == 1)
(9,213 observations deleted)

. *Using bootstrap to determine standard error of the correlation coefficient between
> wages and worker ages. 10000 replications with seed 110821

. drop if age_cat == .
(0 observations deleted)

. drop if incwage == .
(0 observations deleted)

. bootstrap r(rho), nodots nowarn seed(110821) reps(10000): correlate incwage age_cat

Bootstrap results                                Number of obs    =          927
                                                Replications     =        10,000

command: correlate incwage age_cat
       _bs_1: r(rho)

-----
| Observed | Bootstrap | z | P>|z| | Normal-based |
| Coef.    | Std. Err. |   |     | [95% Conf. Interval] |
-----+-----
_ bs_1 | .1495144 | .0278748 | 5.36 | 0.000 | .0948808 | .204148
```

```
. log close
  name: <unnamed>
  log: C:\Users\saiomkark\OneDrive - The University of Chicago\AdvStats\PS7\Stat
> a_Sai_Omkar_K_PS7_Q1.log
  log type: text
  closed on: 8 Dec 2021, 16:45:31
-----
> -----
> -----
```