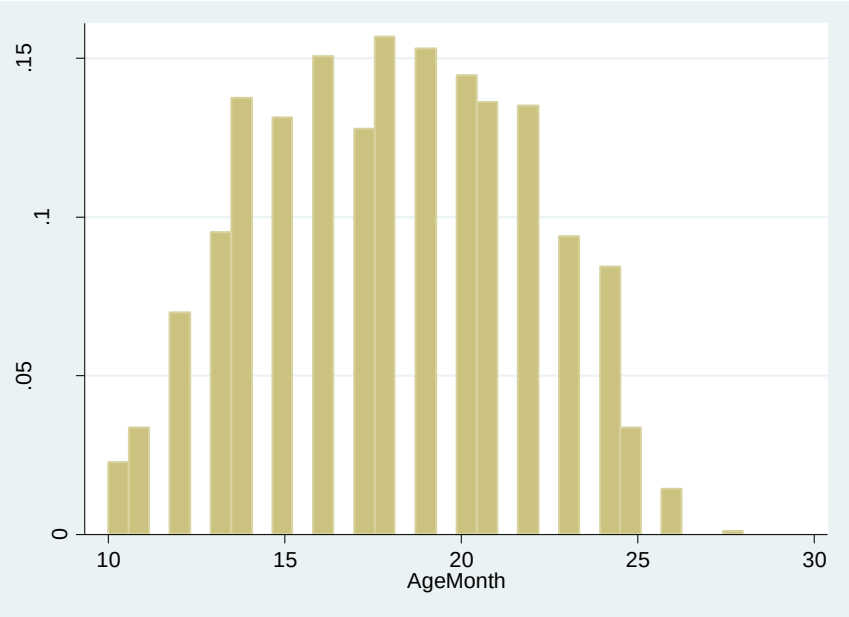
```
------------------------------------------------------------------------------
-------------------------------------------------------
      name:  <unnamed>
       log:  C:\Users\carmelia\Desktop\stndzxage\ado\log2.log
  log type:  text
 opened on:   7 Mar 2019, 14:02:16

. do "C:\Users\carmelia\Desktop\stndzxage\ado\stndzxage tutorial.do"

. *stndzxage tutorial
. *by Sarah Reynolds
. *2-27-19
.
. *The file checks how the command stndzxage differs from zscore
. *The file illustrates how to use the command
.
. clear all

. set more off

. cd "C:\Users\carmelia\Desktop\stndzxage\ado"
C:\Users\carmelia\Desktop\stndzxage\ado

. use "stndzxage_sample_data.dta", clear

.
. count
  1,429

. *1,429 children in the data
.
. count if TestScore~=.
  1,420

. *1,420 were tested
.
. hist AgeMonth
(bin=31, start=10, width=.58064516)
```




```
. *ages concentrated in the center

.
. stndzxage TestScore AgeMonth

. sum stx_TestScore

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
stx_TestSc~e |      1,332    4.73e-08    .9954819  -3.847551   3.815243

. *mean about 0 & standard deviation about 1, as expected
. *however, there are fewer observations
.
. *Do a loop to check standardization with stata command
. levelsof AgeMonth, local(ages)
10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 28

. gen Z_TestScore=.
(1,429 missing values generated)

. foreach age of local ages {
  2.         zscore TestScore if AgeMonth==`age'
  3.         replace Z_TestScore=z_TestScore if AgeMonth==`age'
  4.         drop z_TestScore
  5.         }
z_TestScore created with 1410 missing values
(19 real changes made)
z_TestScore created with 1401 missing values
(28 real changes made)
z_TestScore created with 1373 missing values
(56 real changes made)
z_TestScore created with 1350 missing values
(79 real changes made)
```

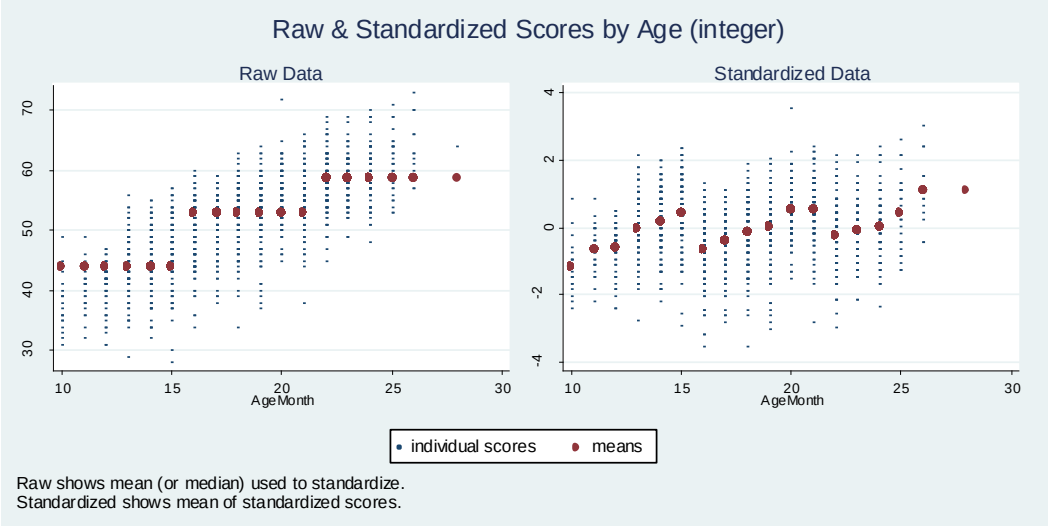
```
z_TestScore created with 1316 missing values
(113 real changes made)
z_TestScore created with 1320 missing values
(109 real changes made)
z_TestScore created with 1304 missing values
(125 real changes made)
z_TestScore created with 1324 missing values
(105 real changes made)
z_TestScore created with 1301 missing values
(128 real changes made)
z_TestScore created with 1303 missing values
(126 real changes made)
z_TestScore created with 1309 missing values
(120 real changes made)
z_TestScore created with 1317 missing values
(112 real changes made)
z_TestScore created with 1318 missing values
(111 real changes made)
z_TestScore created with 1351 missing values
(78 real changes made)
z_TestScore created with 1359 missing values
(70 real changes made)
z_TestScore created with 1401 missing values
(28 real changes made)
z_TestScore created with 1417 missing values
(12 real changes made)
z_TestScore created with 1429 missing values
(0 real changes made)

. sum Z_TestScore

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 Z_TestScore |      1,419    9.17e-10    .9943422  -3.847551   3.815243

. *mean about 0 & standard deviation about 1, as expected
. *however, there are more observations, equal to the
. *number of children who took the test - 1
.
. sum AgeMonth if Z_TestScore==. & TestScore~=.

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
    AgeMonth |          1          28           .         28         28

. *The - 1 corresponds to the child who was the only one of thier age
.
. *Check to see how well they line up if there are both standardization
variables
. scatter Z_TestScore stx_TestScore
```


```
.
. tab AgeMonth if stx_TestScore~=.

   AgeMonth |      Freq.     Percent        Cum.
------------+-----------------------------------
         12 |         56        4.20        4.20
         13 |         79        5.93       10.14
         14 |        113        8.48       18.62
         15 |        109        8.18       26.80
         16 |        125        9.38       36.19
         17 |        105        7.88       44.07
         18 |        128        9.61       53.68
         19 |        126        9.46       63.14
         20 |        120        9.01       72.15
         21 |        112        8.41       80.56
         22 |        111        8.33       88.89
         23 |         78        5.86       94.74
         24 |         70        5.26      100.00
------------+-----------------------------------
      Total |      1,332      100.00

. tab AgeMonth if Z_TestScore~=.

   AgeMonth |      Freq.     Percent        Cum.
------------+-----------------------------------
         10 |         19        1.34        1.34
         11 |         28        1.97        3.31
         12 |         56        3.95        7.26
         13 |         79        5.57       12.83
```

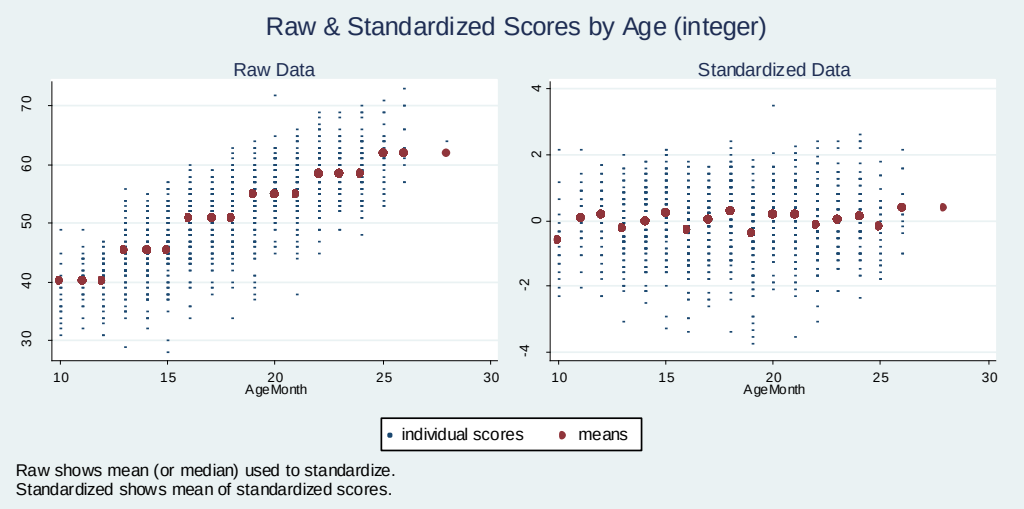
```
         14 |        113        7.96       20.79
         15 |        109        7.68       28.47
         16 |        125        8.81       37.28
         17 |        105        7.40       44.68
         18 |        128        9.02       53.70
         19 |        126        8.88       62.58
         20 |        120        8.46       71.04
         21 |        112        7.89       78.93
         22 |        111        7.82       86.75
         23 |         78        5.50       92.25
         24 |         70        4.93       97.18
         25 |         28        1.97       99.15
         26 |         12        0.85      100.00
------------+-----------------------------------
      Total |      1,419      100.00

. *mismatch in missings because stndzxage has 30 observations minimum
.
. *find out how many are in each month to re-standardize the
. *using the smallest number of observations!
. tab AgeMonth

   AgeMonth |      Freq.     Percent        Cum.
------------+-----------------------------------
         10 |         19        1.33        1.33
         11 |         28        1.96        3.29
         12 |         58        4.06        7.35
         13 |         79        5.53       12.88
         14 |        114        7.98       20.85
         15 |        109        7.63       28.48
         16 |        125        8.75       37.23
         17 |        106        7.42       44.65
         18 |        130        9.10       53.74
         19 |        127        8.89       62.63
         20 |        120        8.40       71.03
         21 |        113        7.91       78.94
         22 |        112        7.84       86.77
         23 |         78        5.46       92.23
         24 |         70        4.90       97.13
         25 |         28        1.96       99.09
         26 |         12        0.84       99.93
         28 |          1        0.07      100.00
------------+-----------------------------------
      Total |      1,429      100.00

. stndzxage TestScore AgeMonth, minbinsize(12)

. assert stx_TestScore==Z_TestScore
1,299 contradictions in 1,429 observations
assertion is false
r(9);

end of do-file
```

```
r(9);

. do "C:\Users\carmelia\AppData\Local\Temp\STD00000000.tmp"

. *This error turns out to be from rounding
. gen stx_round=round(stx_TestScore, 0.0001)
(10 missing values generated)

. gen Z_round=round(Z_TestScore, 0.0001)
(10 missing values generated)

. assert stx_round==Z_round

.
. *****Validation complete********
.
.
. ****Exploring options****
.
. *GRAPHING
. stndzxage TestScore AgeMonth, graph
(9 observations deleted)
```

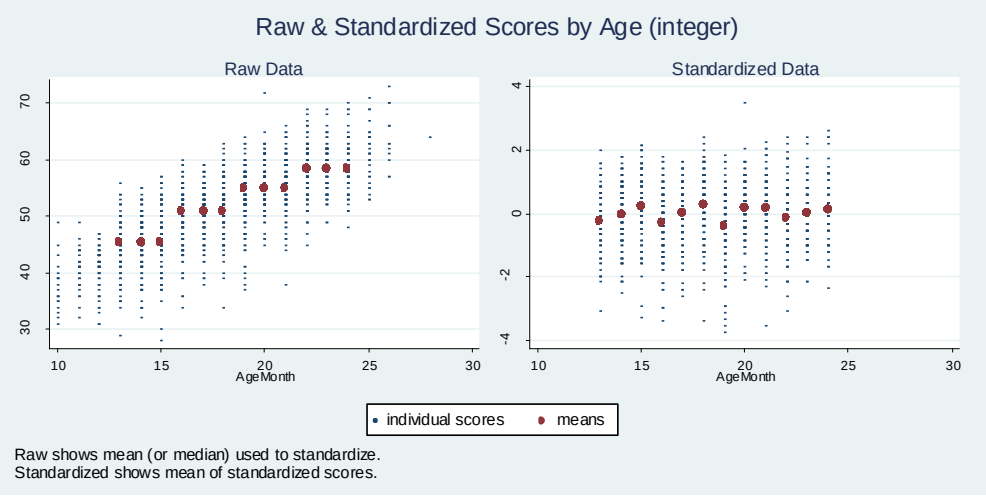

```
. *Notice there are more ages with raw data points than have means
. *These ages had too few observations (default minbinsize is 30)

. *BIN WIDTH
. *let's widen the age bins so more ages are grouped together, resulting
in
. *a larger number of observations in each bin
. stndzxage TestScore AgeMonth, binwidth(6) graph
(9 observations deleted)
```



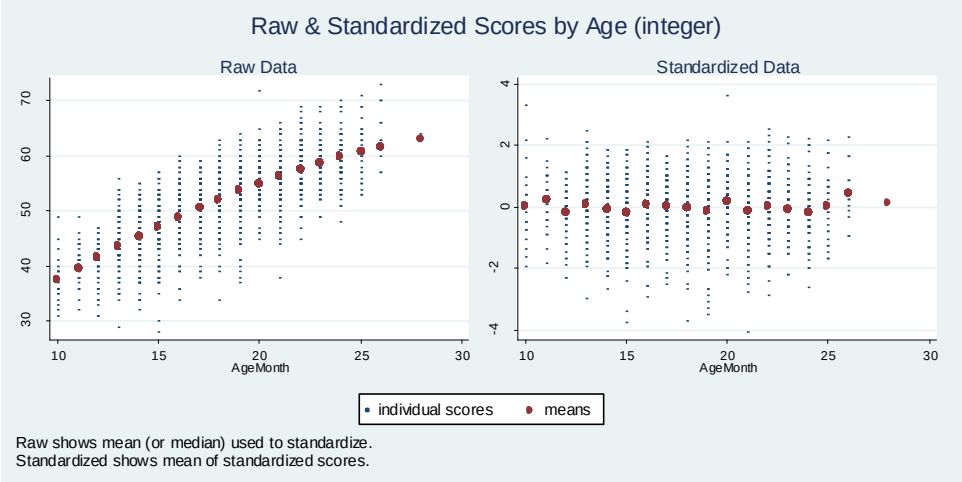

```
. *the waves in the standardized data indicate bins are probably too wide
. stndzxage TestScore AgeMonth, binwidth(3) graph
(9 observations deleted)
```




```
. *still some age dependency but not so much
. *note the last bin included 4 ages (see help file chart about bin
grouping)
.
. *MININIMUM BIN SIZE
. *let's increase the minimum number of observations allowed in each bin
. stndzxage TestScore AgeMonth, binwidth(3) minbinsize(150) graph
(9 observations deleted)
```

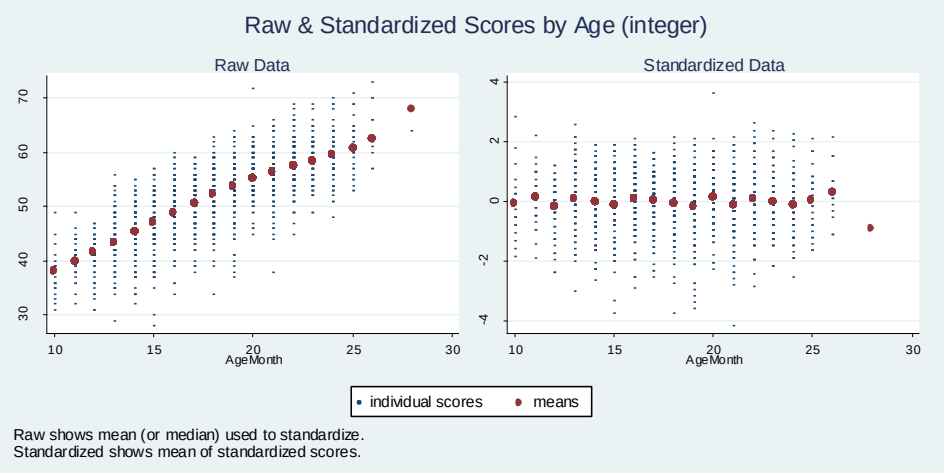



```
.
. *CONTINUOUS
. *continuous standardization is a good option when data density has gaps
(in tails)
. stndzxage TestScore AgeMonth, continuous graph
(9 observations deleted)
```



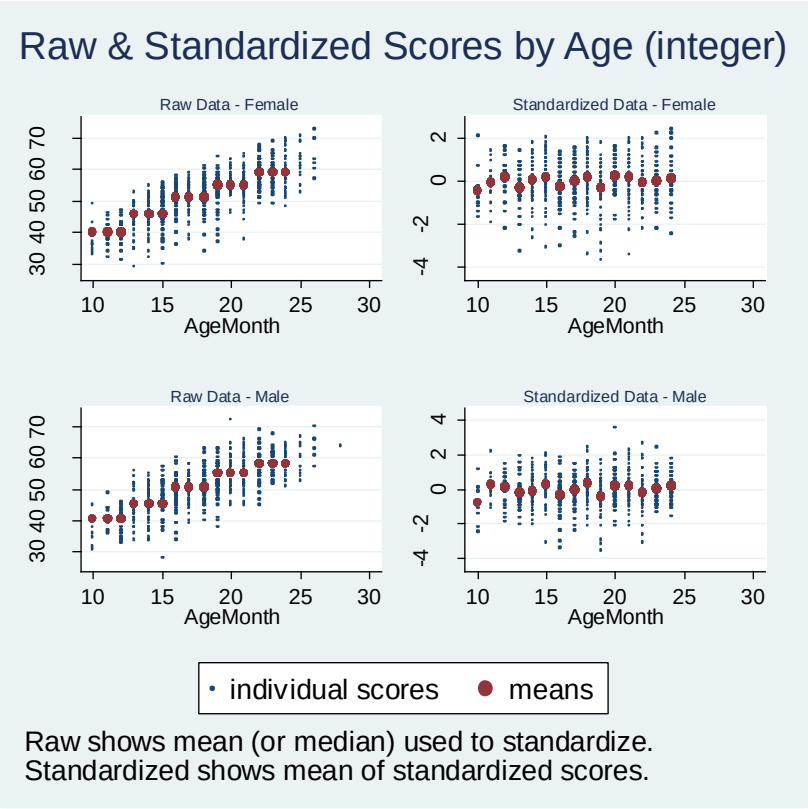

```
. sum stx_TestScore

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
stx_TestSc~e |      1,420    .0001268    1.001219  -4.085064    3.63665

. *note all observations are standardized
. stndzxage TestScore AgeMonth, continuous poly(1) graph // linear
(9 observations deleted)
```


```
. stndzxage TestScore AgeMonth, continuous poly(5) graph // a bit more
curvature
(9 observations deleted)
```



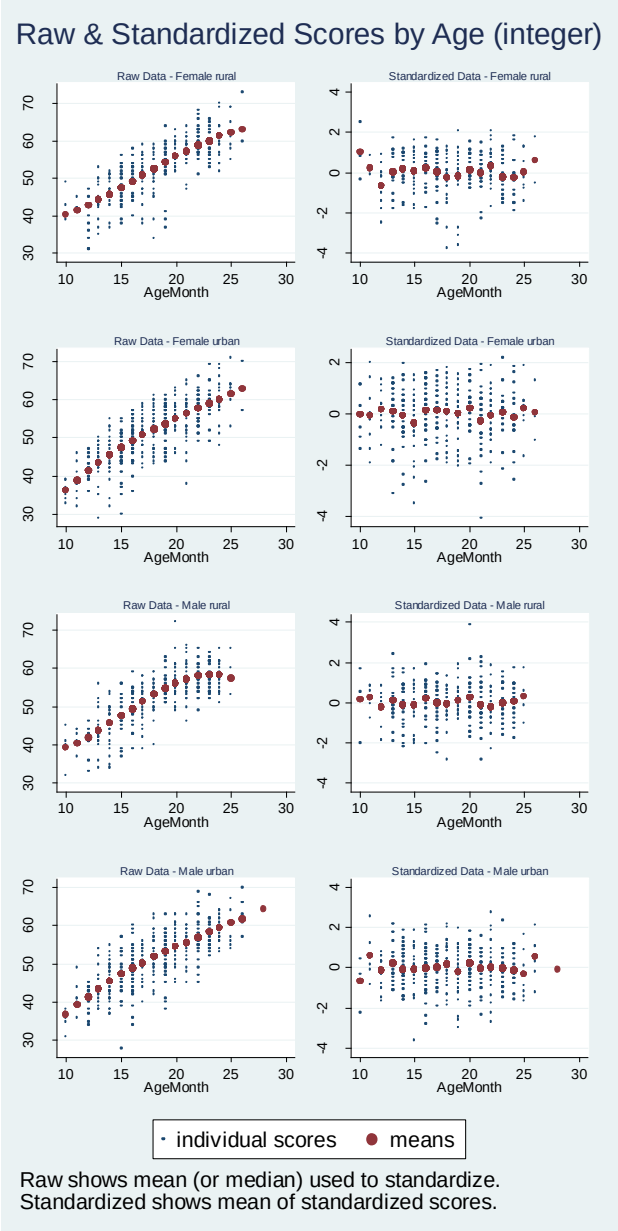

```
.
. *STANDARDIZING OVER ADDITIONAL VARIABLES
. *you can use if to standardize a single subgroup
. stndzxage TestScore AgeMonth if Male==1, binwidth(3)

. tab Male, sum(stx_TestScore)

           |    Summary of Stndz TestScore by
           |    AgeMonth  ~N(0,1) binwidth 3
      Male |        Mean   Std. Dev.       Freq.
-----------+------------------------------------
      Male |  -7.448e-09   .99707603         686
-----------+------------------------------------
     Total |  -7.448e-09   .99707603         686

. stndzxage TestScore AgeMonth if Male==0, binwidth(3)

. tab Male, sum(stx_TestScore)
```

```
           |    Summary of Stndz TestScore by
           |    AgeMonth  ~N(0,1) binwidth 3
      Male |        Mean   Std. Dev.       Freq.
-----------+------------------------------------
    Female |   1.224e-07   .99710566         693
-----------+------------------------------------
     Total |   1.224e-07   .99710566         693

. *but below is more efficient
.
. *standardize by age & gender
. stndzxage TestScore AgeMonth Male, binwidth(3) graph
(9 observations deleted)
```



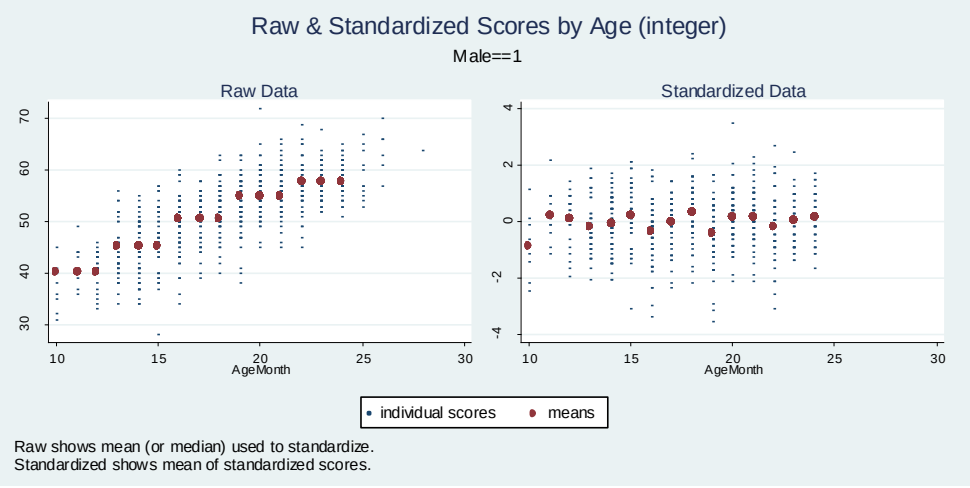

```
. tab Male, sum(stx_TestScore)

           |    Summary of Stndz TestScore by
           |  AgeMonth Male ~N(0,1) binwidth 3
      Male |        Mean   Std. Dev.       Freq.
-----------+------------------------------------
    Female |   1.224e-07   .99710566         693
      Male |  -7.448e-09   .99707603         686
-----------+------------------------------------
     Total |   5.780e-08   .99672906       1,379

. *note means & s.d. are 0 in both cases
.
. *standardize by age, gender, and urban
```

```
. stndzxage TestScore AgeMonth Male Urban, continuous graph
(9 observations deleted)
```




```
. tab Male Urban, sum(stx_TestScore)

                Means, Standard Deviations and Frequencies
              of Stndz TestScore by AgeMonth Male Urban ~N(0,1)

           |        Urban
      Male |     rural      urban |     Total
-----------+----------------------+----------
    Female | .00198964 -.00020896 | .00066925
           | 1.0023932  .99970047 | 1.0000767
           |       286        430 |       716
-----------+----------------------+----------
```

```
      Male |-.00063704 -.00151477 |-.00113326
           | 1.0038777  1.0026333 | 1.0024605
           |       306        398 |       704
-----------+----------------------+----------
     Total | .00063193 -.00083663 |-.00022439
           | 1.0023127  1.0005059 | 1.0009067
           |       592        828 |      1420

.
. *STANDARDIZING WTIH REGARDS TO A REFERENCE GROUP
. stndzxage TestScore AgeMonth, binwidth(3) reference(Male) graph
(9 observations deleted)
```



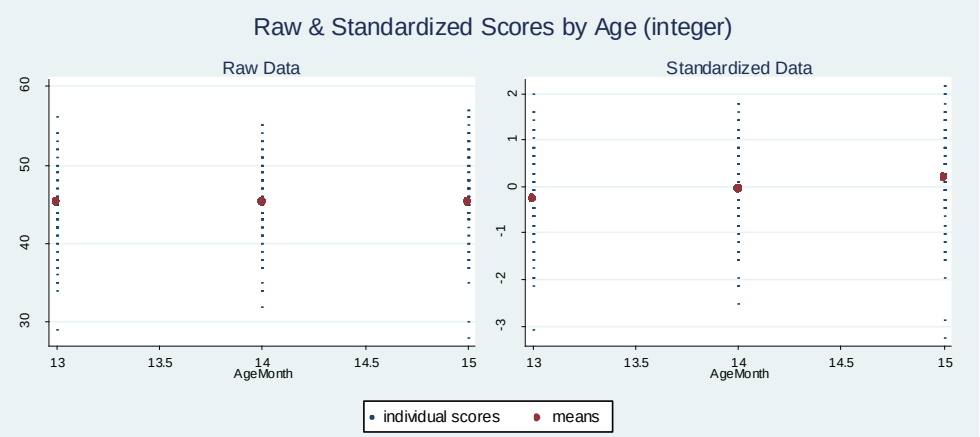

```
. *The graph only illustrates the data for the reference group, which was
used
. *for standardizing
. tab Male, sum(stx_TestScore)

           |    Summary of Stndz TestScore by
           |  AgeMonth  ~N(0,1) binwidth 3; ref
           |             grp Male=1
      Male |        Mean   Std. Dev.       Freq.
-----------+------------------------------------
    Female |   .06922886   1.0157441         693
      Male |  -7.448e-09   .99707603         686
-----------+------------------------------------
     Total |   .03479013   1.0067312       1,379

. *note here the mean & s.d. is ~0 & ~1 for the reference group, but
different for
. *the non reference group
.
. *USING A REFERENCE GROUP & A SUBGROUP
. *can you do it both reference group
. stndzxage TestScore AgeMonth Urban, binwidth(3) minbinsize(30)
reference(Male) graph
(9 observations deleted)
```


```
. tab Male Urban, sum(stx_TestScore)

                Means, Standard Deviations and Frequencies
  of Stndz TestScore by AgeMonth Urban ~N(0,1) binwidth 3; ref grp Male=1

           |        Urban
      Male |     rural      urban |     Total
-----------+----------------------+----------
    Female | .10767812  .07666479 | .08912855
           | 1.1174035  .95998533 | 1.0254123
           |       256        381 |       637
-----------+----------------------+----------
      Male | 5.173e-08 -5.781e-08 |-9.468e-09
           |  .9946476  .99577762 | .99449898
           |       282        357 |       639
-----------+----------------------+----------
     Total |  .0512372  .03957895 | .04449442
           | 1.0552224  .97755028 | 1.0106373
           |       538        738 |      1276


. *USING A DIFFERENT RUNNING VARIABLE
. *Suppose the test was administered with different questions to
different ages
. *Cut the data at the ages for each group
. egen testgroups=cut(AgeMonth), at(10, 13, 16, 19, 25, 30)

. tostring testgroups, replace
testgroups was float now str2
```

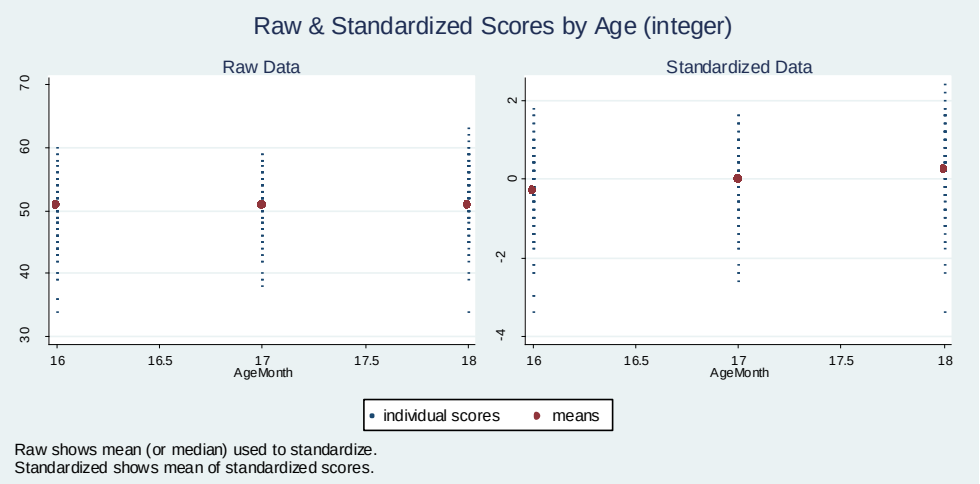
```
. encode testgroups, gen(TestGroups)

. label values TestGroups // remove label from TestGroup2

. stndzxage TestScore TestGroups, graph
(9 observations deleted)
```


```
. rename stx_TestScore testgroups_z

. *This graph has the test groups all lumped together
. *If you want to see the ages graphed also, use the if option.
. *Select the binwidth to be the widest number of ages in a bin.
. levelsof TestGroups, local(groups)
1 2 3 4 5

. gen testgroups_if_z=.
(1,429 missing values generated)

. foreach i of local groups {
  2.         stndzxage TestScore AgeMonth if TestGroups==`i', binwidth(6)
graph
  3.         replace testgroups_if_z=stx_TestScore if TestGroups==`i'
  4.         }
```


```
(1,326 observations deleted)
(103 real changes made)
```

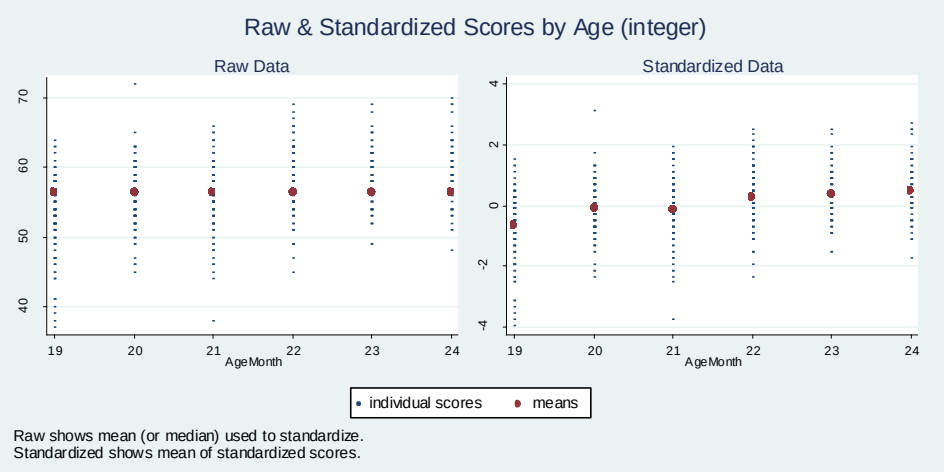



```
(1,128 observations deleted)
(301 real changes made)
```



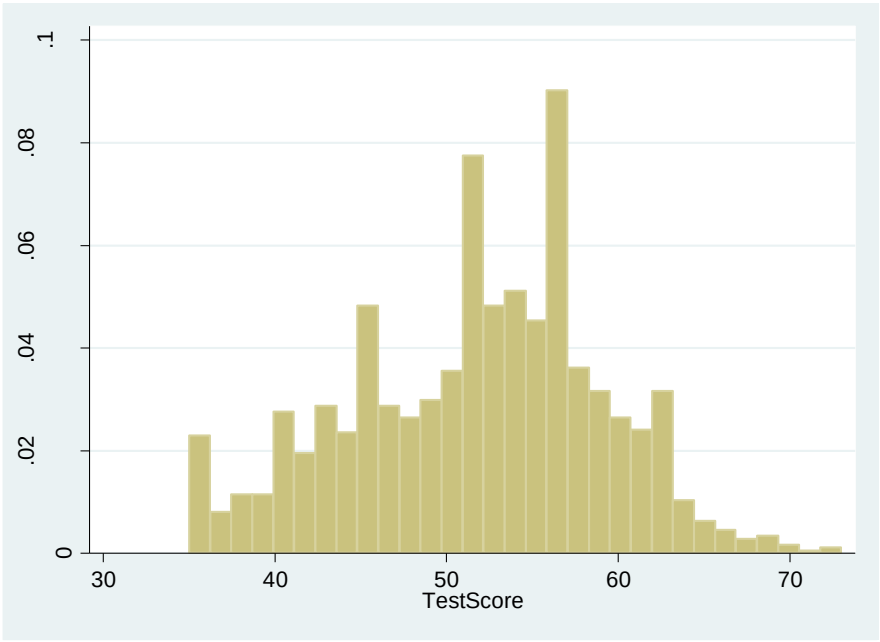

```
(1,071 observations deleted)
```

```
(358 real changes made)
```

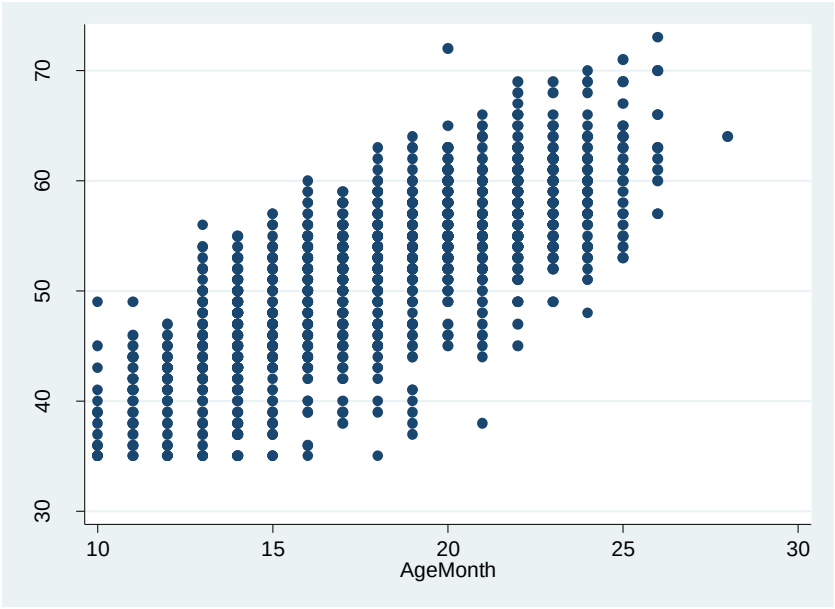



```
(812 observations deleted)
(617 real changes made)
```

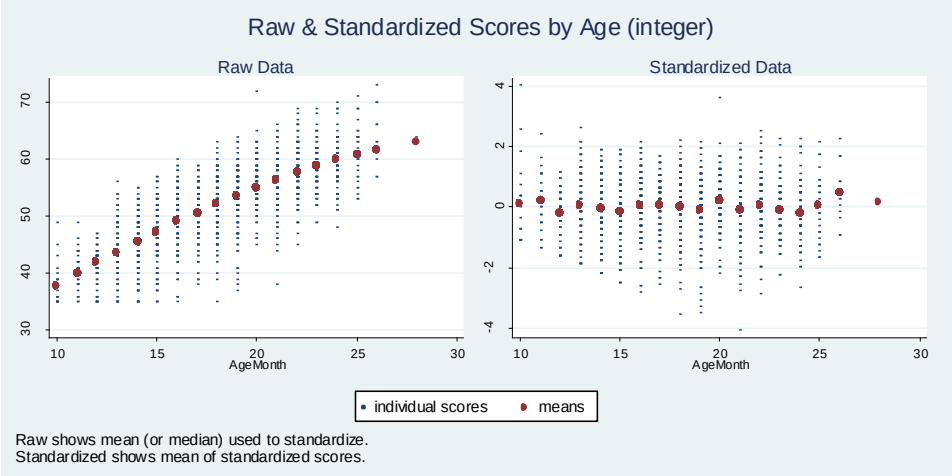

```
(1,388 observations deleted)
(41 real changes made)

. assert testgroups_z==testgroups_if_z

. *Though the syntax below is appealing, it does not work because
. *the ages are divided up by binwidth before the TestGroups
. *       stndzxage TestScore AgeMonth TestGroups, binwidth(6) graph
. *don't use this code!
.
.
. *FLOORS & CEILINGS
. *let's make an artificial floor in this data
. replace TestScore=35 if TestScore<35
(21 real changes made)

. hist TestScore
(bin=31, start=35, width=1.2258065)
```



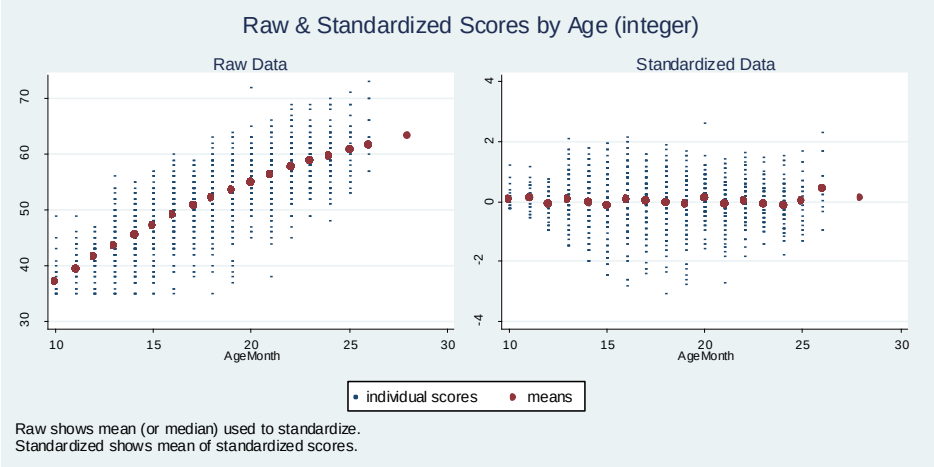

```
. scatter TestScore AgeMonth
```




```
. *If your data actually looked like this, you might be ok with the test
ceiling, but
. *you might want to rethink the appropriateness of the test for the
younger kids:
. *the test best discriminates after about 15 months.
. stndzxage TestScore AgeMonth, continuous graph
(9 observations deleted)
```

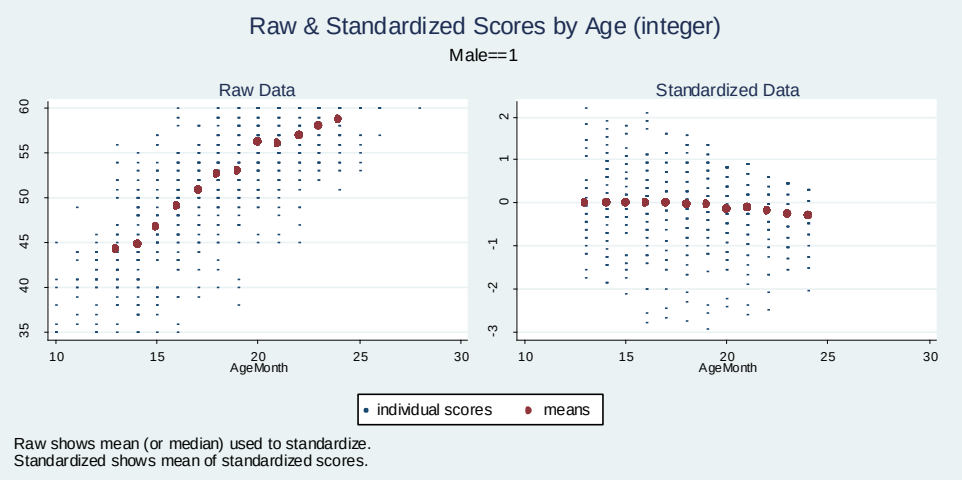



```
. sum stx_TestScore

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
stx_TestSc~e |      1,420    .0005976    1.002678  -4.073903   4.044211

. stndzxage TestScore AgeMonth, continuous floor graph
(9 observations deleted)
```




```
. sum stx_TestScore

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
stx_TestSc~e |      1,420    .0066415    .8071882  -3.050389   2.605103

. *The floor option uses a Tobit adjustment, which assumes a spread
farther below
. *that which is censored.  Censoring pushes the mean up. Without the
adjustment,
. *the mean used to standardize is higher than the mean used to
standardize with a
. *Tobit adjustment. Average standadrdized scores are higher in the Tobit
adjustment
.
```

```
. *We can take ceilings into account as well.
. replace TestScore=60 if TestScore>60 & TestScore~=.
(151 real changes made)

. stndzxage TestScore AgeMonth, floor ceiling minbinsize(30)
reference(Male) graph
(9 observations deleted)
```




```
.
. *USING THE MEDIAN & RESCALING
. *The median can be used for standardizing instead of the mean.
. *A different standard mean/median & standard deviation can be selected
. stndzxage TestScore AgeMonth, sd(15) mean(100) binw(3)

. sum stx_TestScore

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
stx_TestSc~e |      1,420         100    14.97355   39.86069   135.5473

. stndzxage TestScore AgeMonth, median sd(15) mean(100) binw(3)

. sum stx_TestScore

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
stx_TestSc~e |      1,420    97.73348    15.05202   36.84235   133.5137

.
end of do-file

. log close
      name:  <unnamed>
       log:  C:\Users\carmelia\Desktop\stndzxage\ado\log2.log
  log type:  text
 closed on:   7 Mar 2019, 14:05:01
-------------------------------------------------------------------------------
-------------------------------------------------------
```