Measurement Error

What Happens When You Can't Measure Things Perfectly?

Suppose we measure x with error?

Instead of observing x, we observe $x' = x + e$

(e is random with mean $\bar{e}$ and variance v_e)

$\therefore$ instead of doing the regression

$$y = \alpha + \beta x + \varepsilon,$$

we do the regression

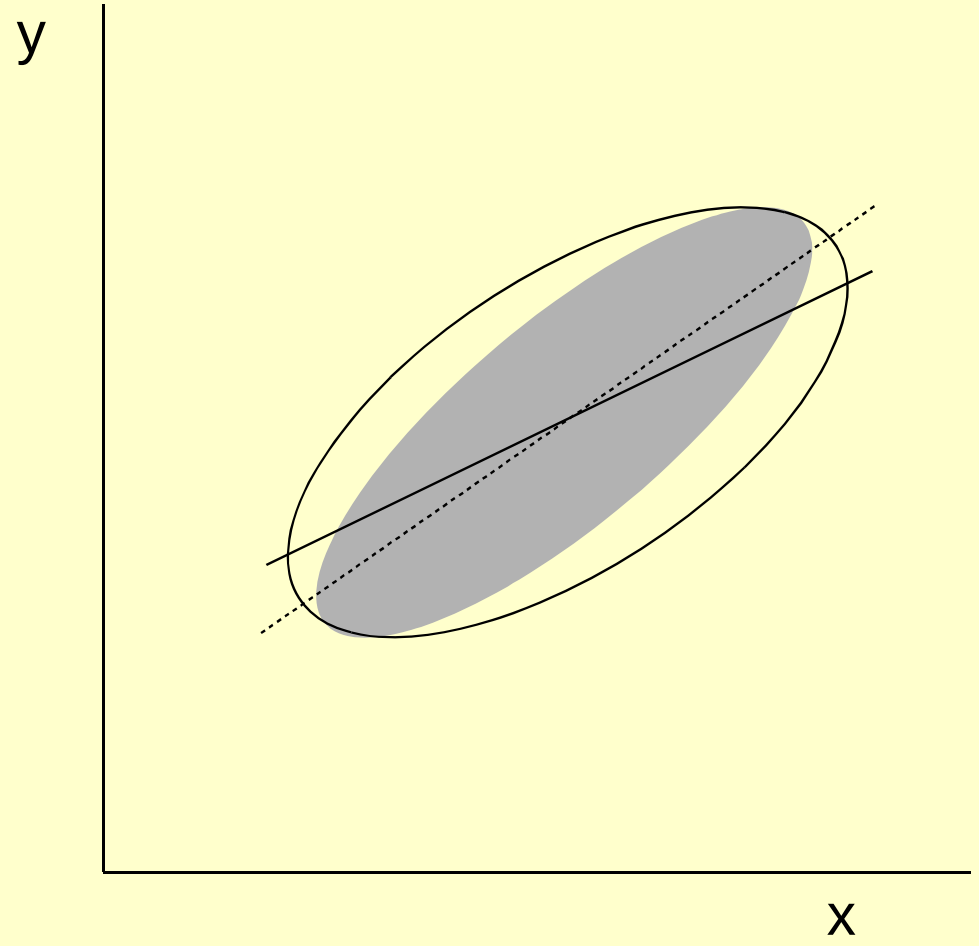
$$y = \alpha + \beta' x' + \varepsilon.$$

What is the relationship between β and β' ?

Answer

$$\beta' = \frac{\mathrm{cov}(x, y)}{\mathrm{var}(x) + \mathrm{var}(e)}$$

Errors in Independent Variables: The Picture



Suppose we measure y with error

Instead of observing y, we observe $y' = y + e$

(e is random with mean $\bar{e}$ and variance v_e)

$\therefore$ instead of doing the regression

$y = \alpha + \beta x + \varepsilon,$

we do the regression

$y' = \alpha + \beta' x + \varepsilon.$

What is the relationship between β and β' ?

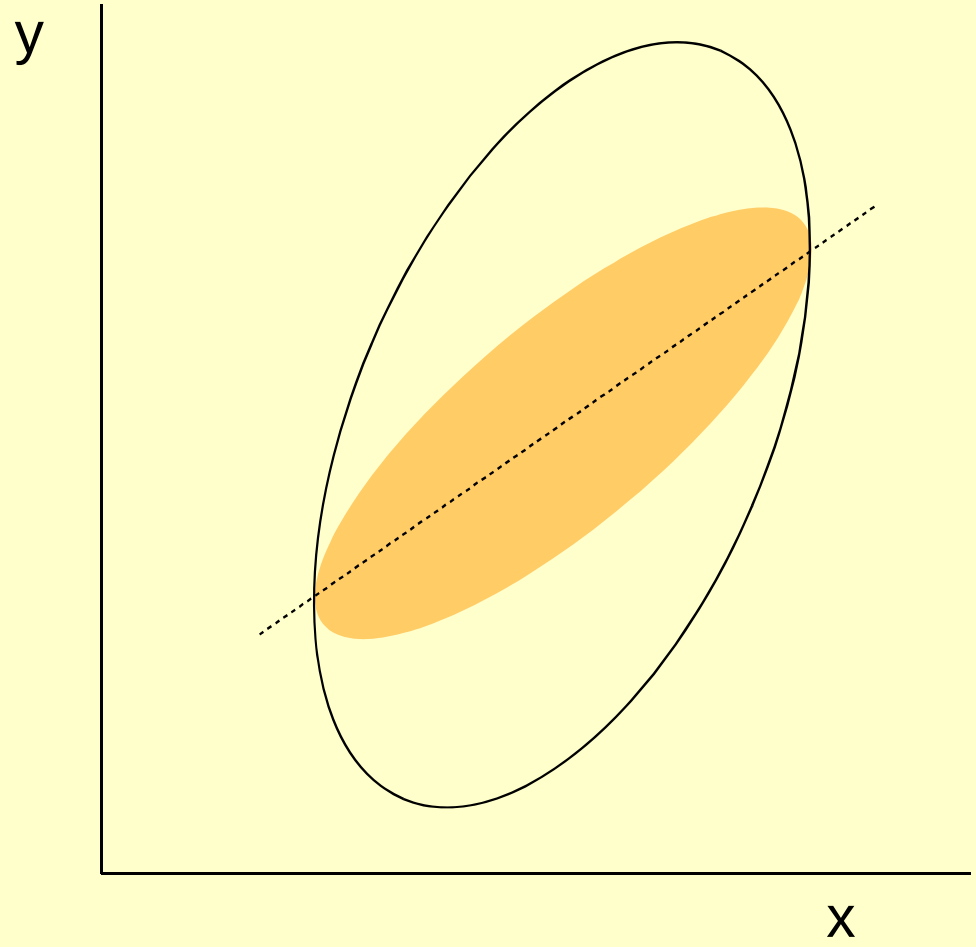
The answer

$$\beta' = \frac{\mathrm{cov}(x, y)}{\mathrm{var}(x)} = \beta$$

But…

- Standard errors and s.e.r. inflated
- R^2 deflated

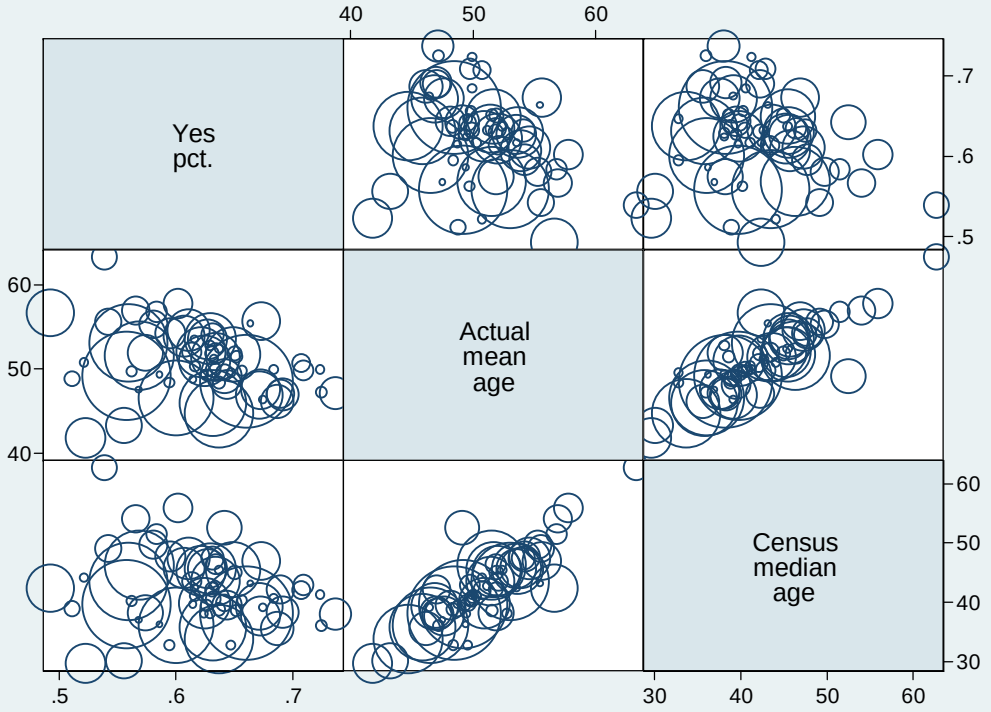
Errors in Dependent Variables: The Picture



Example drawn from 2012 Florida Constitutional Amendments

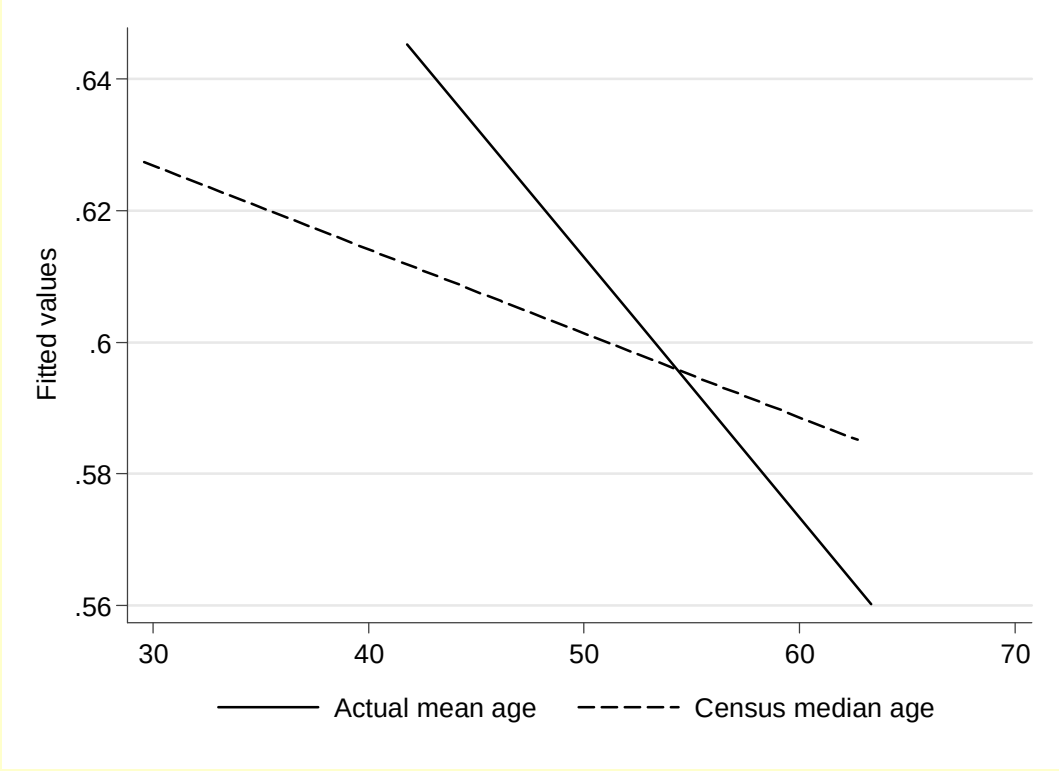
ADDITIONAL HOMESTEAD EXEMPTION; LOW-INCOME SENIORS WHO MAINTAIN LONG-TERM RESIDENCY ON PROPERTY; EQUAL TO ASSESSED VALUE

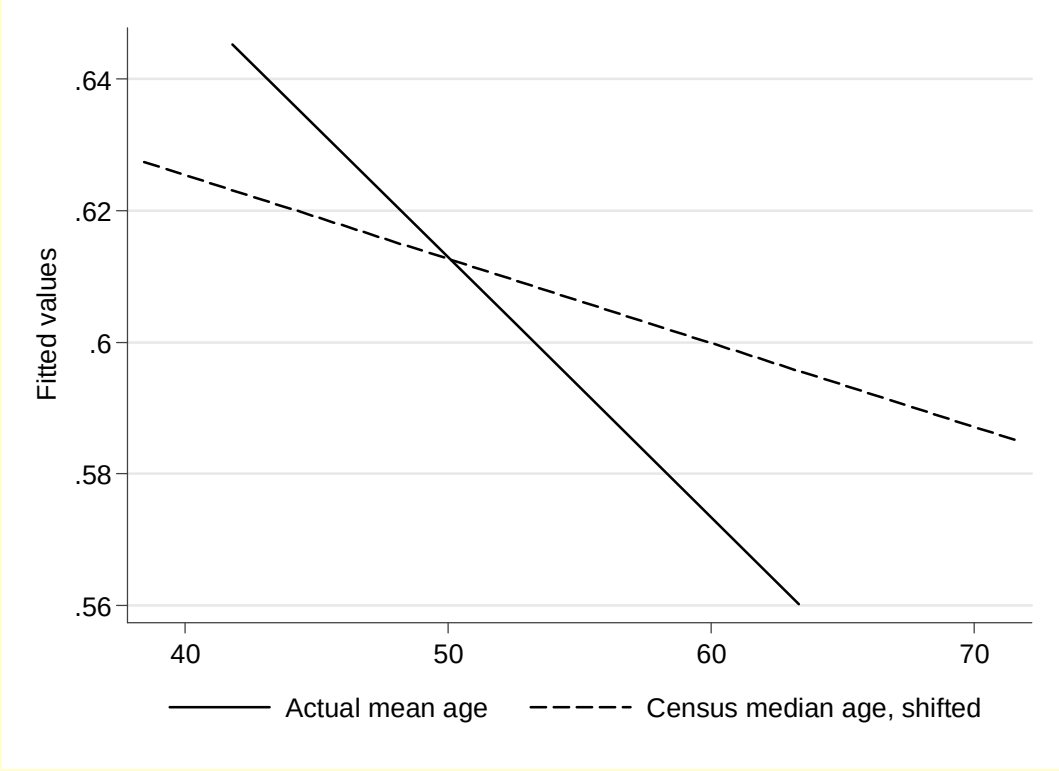

```
. corr yespct age medage [aweight=turnout]
(sum of wgt is   7.7021e+06)
(obs=67)

             |   yespct      age   medage
-------------+---------------------------
      yespct |   1.0000
         age |  -0.2988   1.0000
      medage |  -0.1427   0.8837   1.0000
```



```
. summ age medage [aw=turnout]

    Variable |     Obs      Weight        Mean   Std. Dev.       Min        Max
-------------+--------------------------------------------------------------------
         age |      67     7702097     50.0893   3.740067   41.80486   63.33784
      medage |      67     7702097    41.24497   5.523959       29.6       62.7
```

```
. reg yespct age [aw=turnout]
(sum of wgt is   7.7021e+06)

      Source |       SS       df       MS              Number of obs =      67
-------------+------------------------------           F(  1,    65) =    6.37
       Model |  .014375688     1  .014375688           Prob > F      =  0.0140
    Residual |  .146588661    65   .00225521           R-squared     =  0.0893
-------------+------------------------------           Adj R-squared =  0.0753
       Total |  .160964349    66  .002438854           Root MSE      =  .04749

------------------------------------------------------------------------------
      yespct |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0039461   .0015629    -2.52   0.014    -.0070675   -.0008246
       _cons |    .810193   .0785012    10.32   0.000     .6534153    .9669708
------------------------------------------------------------------------------


. reg yespct medage [aw=turnout]
(sum of wgt is   7.7021e+06)

      Source |       SS       df       MS              Number of obs =      67
-------------+------------------------------           F(  1,    65) =    1.35
       Model |   .00327667     1   .00327667           Prob > F      =  0.2494
    Residual |   .15768768    65  .002425964           R-squared     =  0.0204
-------------+------------------------------           Adj R-squared =  0.0053
       Total |  .160964349    66  .002438854           Root MSE      =  .04925

------------------------------------------------------------------------------
      yespct |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      medage |  -.0012755   .0010975    -1.16   0.249    -.0034675    .0009164
       _cons |   .6651476   .0456661    14.57   0.000      .573946    .7563491
------------------------------------------------------------------------------
```


.64
.62
.6
.58
.56
Fitted values
30
40
50
60
70
Actual mean age
Census median age


.64
.62
.6
.58
.56
Fitted values
40
50
60
70
Actual mean age
Census median age, shifted