Assignment 3

Q1: Heteroskedasticity is one of the assumption for MLR, and is used to find the OLS estimation. without this, the $\beta_0, \beta_1, \ldots$, etc. will not be efficient [$Var(U|X \ldots X_k) = \sigma^2$. Also, the explanatory variables should be included to make the OLS unbiased, in order to separate them with the error term. If our OLS has sample correlation coefficient of 0.95 between 2 independent variables, it does not wrong since it does not violate any MLR assumption

Q2: (i) Null hypothesis that, after controlling for sales and roe, ros has no effect on CEO salary

→ $H_0: \beta_3 = 0$

The alternative that better stock market performance increase CEO's salary

→ $H_a: \beta_3 > 0$

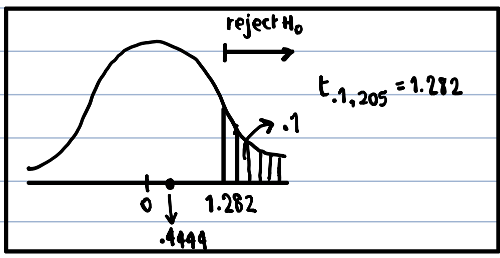
(ii) The proportionate affect of ros on salary if ros increase by 50 points is $.00024(50) = .012$. To get the percentage effect, we can just simply multiply it by 100 and we get 1.2%. Ros does not have a practically large effect on salary as a large change is Ros gives small effect on salary.

(iii) $H_0: \beta_3 = 0$ $\quad\quad$ $\alpha = .1$

$H_0: \beta_3 > 0$

$n = 209$, $d.f. = n-k-1 = 209-3-1 = 205$

→ we can use z-table as $d.f. > 30$ → $z = \dfrac{\hat{\beta}_3 - \beta_3}{s.e.\,\hat{\beta}_3} = \dfrac{0.00024 - 0}{0.00054} = 0.4444$



The z-value does not fall into the rejection region.

→ Does reject H_0 at .1 level of significance.

∴ ros has no effect on salary #

(iv) No, since we already test that ros does not have any effect on salary at .1 level of significance. Eventhough if we add, it will raise r^2, but it will worsen the variance. We only want to put in what really explained the model #

Natchaya Laokham
6104641425

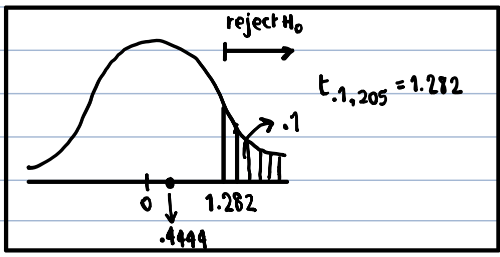
Assignment 3

Q1: Heteroskedasticity is one of the assumption for MLR, and is used to find the OLS estimation. without this, the $\beta_0, \beta_1, \dots$, etc. will not be efficient $[Var(U \mid X \dots X_K) = 6^2$. Also, the explanatory variables should be included to make the OLS unbiased, in order to separate them with the error term. If our OLS has sample correlation coefficient of 0.95 between 2 independent variables, it does not wrong since it does not violate any MLR assumption

Q2: (i) Null hypothesis that, after controlling for sales and roe, ros has no effect on CEO salary
$\rightarrow H_0: \beta_3 = 0$
The alternative that better stock market performance increase CEO's salary
$\rightarrow H_a: \beta_3 > 0$

(ii) The proportionate affect of ros on salary if ros increase by 50 points is $.00024(50) = .012$. To get the percentage effect, we can just simply multiply it by 100 and we get 1.2%. Ros does not have a practically large effect on salary as a large change is Ros gives small effect on salary.

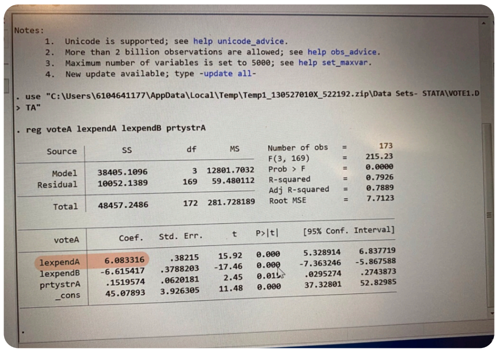
(iii) $H_0: \beta_3 = 0$ $\qquad d = .1$
$H_0: \beta_3 > 0$
$n = 209$, $d.f. = n-k-1 = 209-3-1 = 205$
$\rightarrow$ we can use z-table as $d.f. > 30$ $\Rightarrow$ $z = \dfrac{\hat{\beta}_3 - \beta_3}{s.e.\,\hat{\beta}_3} = \dfrac{0.00024 - 0}{0.00054} = 0.4444$



The z-value does not fall into the rejection region.
$\Rightarrow$ Does reject H_0 at .1 level of significance.
$\therefore$ ros has no effect on salary #

(iv) No, since we already test that ros does not have any effect on salary at .1 level of significance. Eventhough if we add, it will raise r^2, but it will worsen the variance. We only want to put in what really explained the model #

C1. $\text{vote}A = \beta_0 + \beta_1 \log(\text{expand }A) + \beta_2 \log(\text{expand }B) + \beta_3 \text{prtystr}A + u,$



```
Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  More than 2 billion observations are allowed; see help obs_advice.
      3.  Maximum number of variables is set to 5000; see help set_maxvar.
      4.  New update available; type -update all-

. use "C:\Users\6104641177\AppData\Local\Temp\Temp1_130527010X_522192.zip\Data Sets- STATA\VOTE1.D
> TA"

. reg voteA lexpendA lexpendB prtystrA

      Source |       SS           df       MS      Number of obs   =       173
-------------+----------------------------------   F(3, 169)       =    215.23
       Model |  38405.1096         3  12801.7032   Prob > F        =    0.0000
    Residual |  10052.1389       169   59.480112   R-squared       =    0.7926
-------------+----------------------------------   Adj R-squared   =    0.7889
       Total |  48457.2486       172  281.728189   Root MSE        =    7.7123

------------------------------------------------------------------------------
       voteA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lexpendA |   6.083316     .38215    15.92   0.000     5.328914    6.837719
    lexpendB |  -6.615417   .3788203   -17.46   0.000    -7.363246   -5.867588
    prtystrA |   .1519574   .0620181     2.45   0.01      .0295274    .2743873
       _cons |   45.07893   3.926305    11.48   0.000     37.32801    52.82985
------------------------------------------------------------------------------

.
```

i) $B_1 = 6.083316$ means an additional increase in $\log(\text{expand }A)$ would increase vote A by 6.083316 units.

ii) $H_0 = \beta_2 = -\beta_1 \quad , \beta_2 + \beta_1 = 0$
$H_a = \beta_2 \neq -\beta_1 \quad , \beta_2 + \beta_1 \neq 0$

iii) $\text{Vote }A = 45.07893 + 6.083316 \log(\text{expand }A) - 6.615417 \log(\text{expand }B) + 0.1519574\ \text{prtystr}A + u$

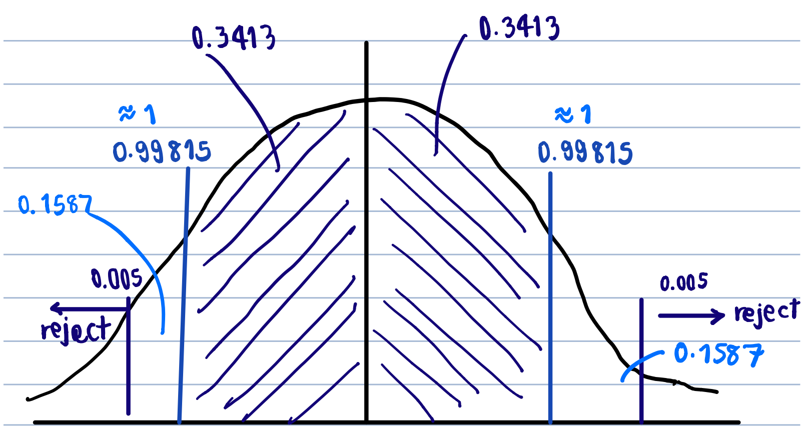
$\therefore$ A's expenture has a positive effect on vote A. However, B's expenditure has a negature effect on Vote A. $\therefore$ we can't use these result to test.

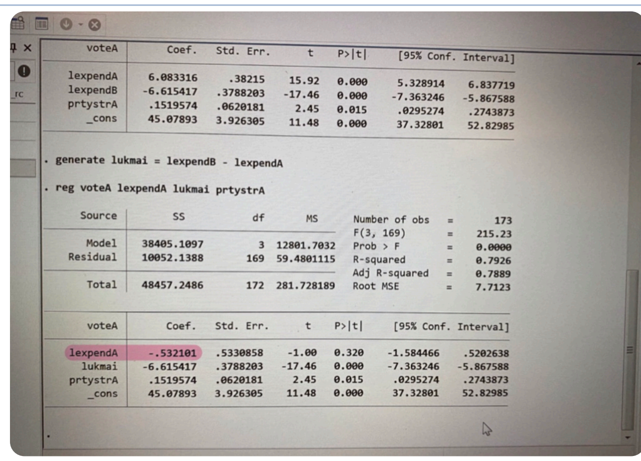
iv) $H_0: \beta_2 + \beta_1 = 0$
$H_a: \beta_2 + \beta_1 \neq 0$
Let $\hat{\beta}_2 + \hat{\beta}_1 = \hat{\theta} \rightarrow \beta_1 = \theta - \beta_2$
From $t_{\beta_j} = \dfrac{\hat{\beta}_j - \beta_j}{s.e.\,\hat{\beta}_j} = \dfrac{(\hat{\beta}_2 + \hat{\beta}_1) - 0}{s.e.(\hat{\beta}_2 + \hat{\beta}_1)} = \dfrac{\hat{\theta} - 0}{s.e.\,\hat{\theta}}$

Sub in the equation
$\text{Vote }A = \beta_0 + \beta_1 \log(\text{expand}A) + \beta_2 \log(\text{expand}B) + \beta_3 \text{prtystr}A + u$
$\text{Vote }A = \beta_0 + (\theta - \beta_2) \log(\text{expand}A) + \beta_2 \log(\text{expand}B) + \beta_3 \text{prtystr}A + u$
$= \beta_0 + \theta \log(\text{expand}A) - \beta_2 \log(\text{expand}B) + \beta_3 \text{prtystr}A + u$
$\beta_0 + \theta \log(\text{expand}A) - \beta_2 (\log(\text{expand}A) - \log(\text{expand}B)) + \beta_3 \text{prtystr}A + u$

Do regression again

$t = \dfrac{\hat{\theta} - 0}{s.e.\,\hat{\theta}} = \dfrac{0.532101}{0.5330858} = 0.99815$



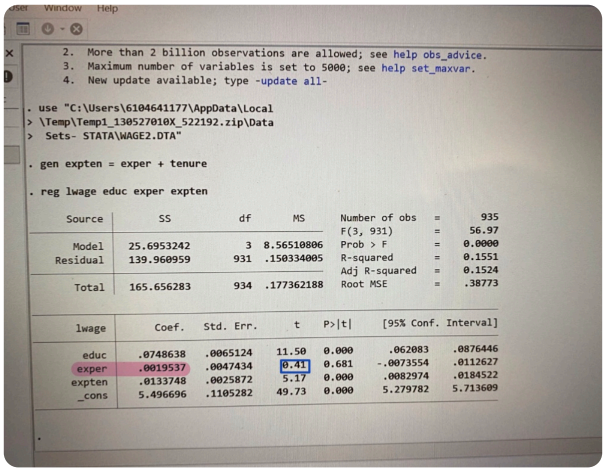


```
      voteA  |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lexpendA |   6.083316     .38215    15.92   0.000     5.328914    6.837719
    lexpendB |  -6.615417   .3788203   -17.46   0.000    -7.363246   -5.867588
    prtystrA |   .1519574   .0620181     2.45   0.015     .0295274    .2743873
       _cons |   45.07893   3.926305    11.48   0.000     37.32801    52.82985
------------------------------------------------------------------------------

. generate lukmai = lexpendB - lexpendA

. reg voteA lexpendA lukmai prtystrA

      Source |       SS           df       MS      Number of obs   =       173
-------------+----------------------------------   F(3, 169)       =    215.23
       Model |  38405.1097         3  12801.7032   Prob > F        =    0.0000
    Residual |  10052.1388       169  59.4801115   R-squared       =    0.7926
-------------+----------------------------------   Adj R-squared   =    0.7889
       Total |  48457.2486       172  281.728189   Root MSE        =    7.7123

------------------------------------------------------------------------------
       voteA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lexpendA |   -.532101   .5330858    -1.00   0.320    -1.584466    .5202638
      lukmai |  -6.615417   .3788203   -17.46   0.000    -7.363246   -5.867588
    prtystrA |   .1519574   .0620181     2.45   0.015     .0295274    .2743873
       _cons |   45.07893   3.926305    11.48   0.000     37.32801    52.82985
------------------------------------------------------------------------------

.
```

So, we do not reject H_0 b/c P-value is more than 0.005 (significant level)
$\therefore$ Increase in an additional unit in $-\beta_2$ would cause an additional negative value to β_1 and vise versa

(C 6.) i) $\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u.$

$H_0: \beta_2 = \beta_3$



```
2.  More than 2 billion observations are allowed; see help obs_advice.
3.  Maximum number of variables is set to 5000; see help set_maxvar.
4.  New update available; type -update all-

. use "C:\Users\6104641177\AppData\Local
> \Temp\Temp1_130527010X_522192.zip\Data
>  Sets- STATA\WAGE2.DTA"

. gen expten = exper + tenure

. reg lwage educ exper expten

      Source |       SS           df       MS      Number of obs   =       935
-------------+----------------------------------   F(3, 931)       =     56.97
       Model |  25.6953242         3  8.56510806   Prob > F        =    0.0000
    Residual |  139.960959       931  .150334005   R-squared       =    0.1551
-------------+----------------------------------   Adj R-squared   =    0.1524
       Total |  165.656283       934  .177362188   Root MSE        =    .38773

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0748638   .0065124    11.50   0.000      .062083    .0876446
       exper |   .0019537   .0047434     0.41   0.681    -.0073554    .0112627
      expten |   .0133748   .0025872     5.17   0.000     .0082974    .0184522
       _cons |   5.496696   .1105282    49.73   0.000     5.279782    5.713609
------------------------------------------------------------------------------
```

from $H_0: \beta_2 = \beta_3$, $\beta_2 - \beta_3 = 0$

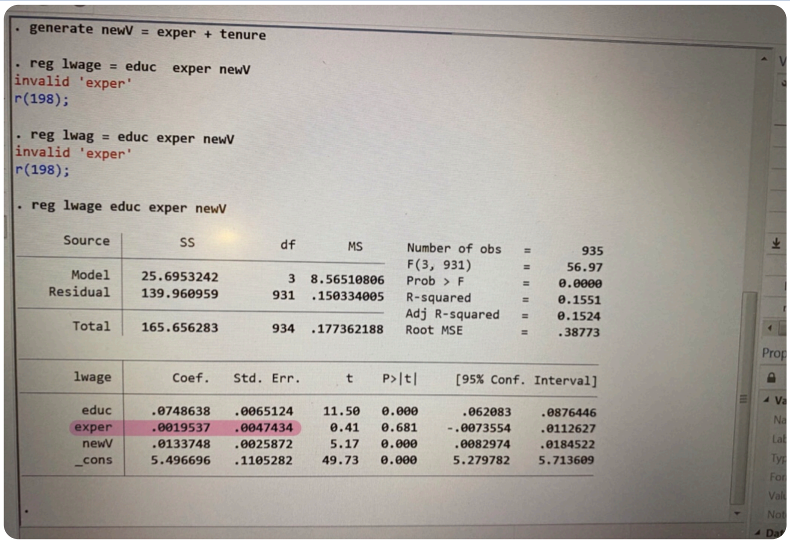
$H_a: \beta_2 \neq \beta_3$, $\beta_2 - \beta_3 \neq 0$

Let $\hat{\theta} = \hat{\beta}_2 - \hat{\beta}_3 \rightarrow \beta_2 = \theta + \beta_3$

sub in equation, we get

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u$$

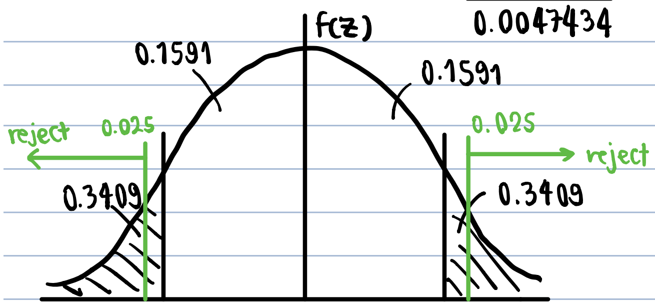
$$\begin{aligned} \log(wage) &= \beta_0 + \beta_2 educ + (\theta + \beta_3) exper + \beta_3 tenure + u \\ &= \beta_0 + \beta_2 educ + \theta\, exper + \beta_3 exper + \beta_3 tenure + u \\ &= \beta_0 + \beta_2 educ + \theta\, exper + \beta_3 \underbrace{(exper + tenure)}_{\text{new V}} + u \end{aligned}$$



```
. generate newV = exper + tenure

. reg lwage = educ  exper newV
invalid 'exper'
r(198);

. reg lwag = educ exper newV
invalid 'exper'
r(198);

. reg lwage educ exper newV

      Source |       SS           df       MS      Number of obs   =       935
-------------+----------------------------------   F(3, 931)       =     56.97
       Model |  25.6953242         3  8.56510806   Prob > F        =    0.0000
    Residual |  139.960959       931  .150334005   R-squared       =    0.1551
-------------+----------------------------------   Adj R-squared   =    0.1524
       Total |  165.656283       934  .177362188   Root MSE        =    .38773

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0748638   .0065124    11.50   0.000      .062083    .0876446
       exper |   .0019537   .0047434     0.41   0.681    -.0073554    .0112627
        newV |   .0133748   .0025872     5.17   0.000     .0082974    .0184522
       _cons |   5.496696   .1105282    49.73   0.000     5.279782    5.713609
------------------------------------------------------------------------------
```

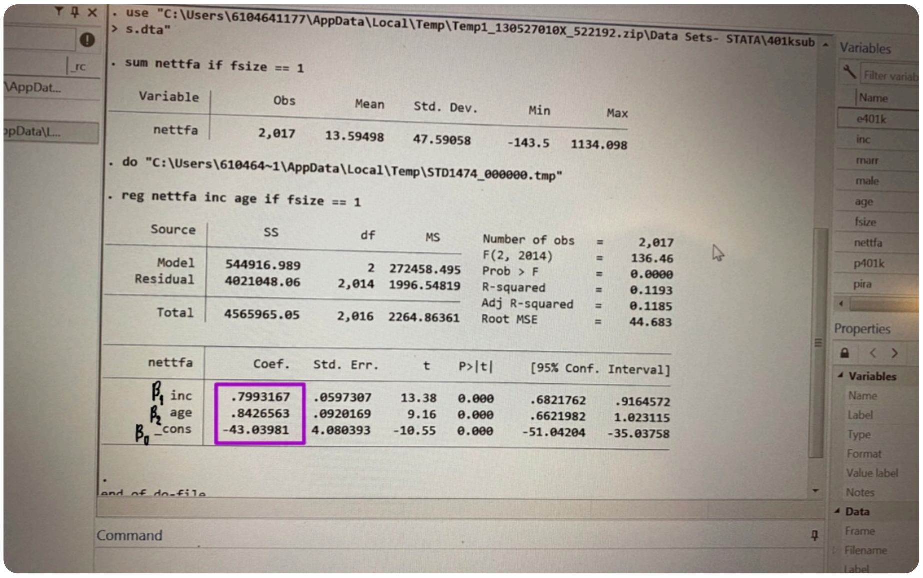
from $t_{\beta_j} = \dfrac{\hat{\beta}_j - \beta_j}{s.e.\ \hat{\beta}_j} = \dfrac{\hat{\theta} - 0}{s.e.\ \hat{\theta}}$

$= \dfrac{0.0019537}{0.0047434} = 0.41178 \approx 0.41$



Since, P-value of 0.41 = 0.3409 is greater than significant level of 5% (0.025). Therefore, we can conclude that we will not reject null hypothesis cause P-value is not fall into rejection area.

C8.



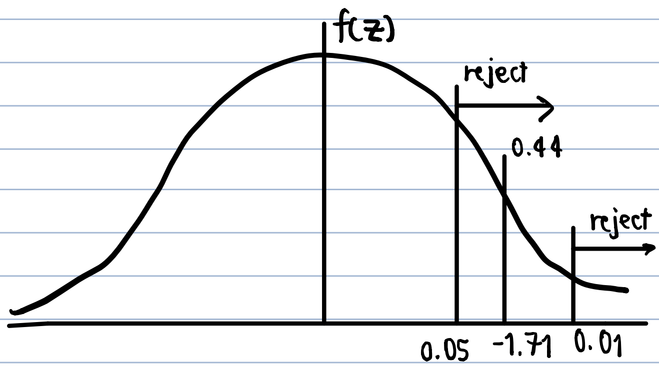
```
. use "C:\Users\6104641177\AppData\Local\Temp\Temp1_130527010X_522192.zip\Data Sets- STATA\401ksub
> s.dta"

. sum nettfa if fsize == 1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      nettfa |      2,017    13.59498    47.59058     -143.5   1134.098

. do "C:\Users\610464~1\AppData\Local\Temp\STD1474_000000.tmp"

. reg nettfa inc age if fsize == 1

      Source |       SS           df       MS      Number of obs   =     2,017
-------------+----------------------------------   F(2, 2014)      =    136.46
       Model |  544916.989         2  272458.495   Prob > F        =    0.0000
    Residual |  4021048.06     2,014  1996.54819   R-squared       =    0.1193
-------------+----------------------------------   Adj R-squared   =    0.1185
       Total |  4565965.05     2,016  2264.86361   Root MSE        =    44.683

------------------------------------------------------------------------------
      nettfa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inc |   .7993167   .0597307    13.38   0.000     .6821762    .9164572
         age |   .8426563   .0920169     9.16   0.000     .6621982    1.023115
       _cons |  -43.03981   4.080393   -10.55   0.000    -51.04204   -35.03758
------------------------------------------------------------------------------

. end of do-file

Command
```


i) 2017

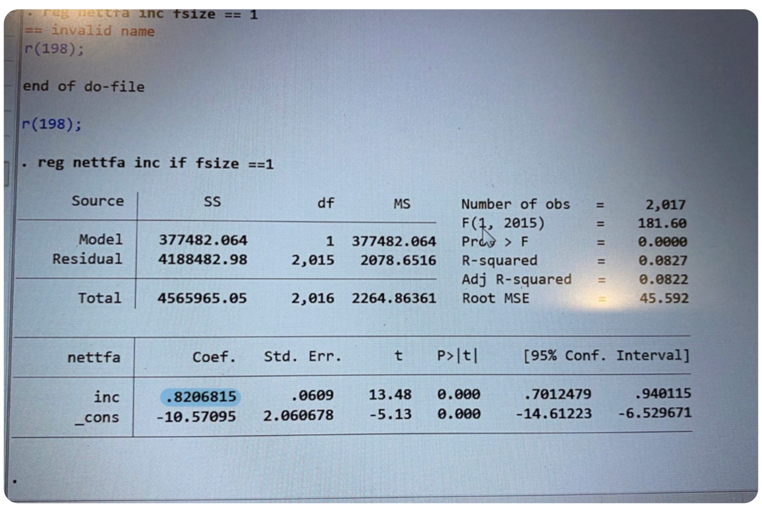
ii) $\hat{\beta}_1$ means a \$1 thousand dollars increase in income corresponds to \$799 increase in nettfa. $\hat{\beta}_2$ as 1 year increase in age corresponds to a \$842 increase in net total financial wealth.

iii) $\beta_0 = -43.04$ ∴ when their income and age equal to 0, so this is the net financial wealth of newborn babies.

iv) $t_{\beta_2} = \dfrac{\hat{\beta}_2 - \beta_2}{s.e.\ \beta_2} = \dfrac{(0.843 - 1)}{0.092017} = -1.71$ ∴ The P-value is $P(T < -1.71) \approx 0.044$ → from the table



→ ∴ we can reject the null hypothesis at 5% level significant but not 1% significant



```
. reg nettfa inc fsize == 1
== invalid name
r(198);

end of do-file

r(198);

. reg nettfa inc if fsize ==1

      Source |       SS           df       MS      Number of obs   =     2,017
-------------+----------------------------------   F(1, 2015)      =    181.60
       Model |  377482.064         1  377482.064   Prob > F        =    0.0000
    Residual |  4188482.98     2,015  2078.6516    R-squared       =    0.0827
-------------+----------------------------------   Adj R-squared   =    0.0822
       Total |  4565965.05     2,016  2264.86361   Root MSE        =    45.592

------------------------------------------------------------------------------
      nettfa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inc |   .8206815      .0609    13.48   0.000     .7012479     .940115
       _cons |  -10.57095   2.060678    -5.13   0.000    -14.61223   -6.529671
------------------------------------------------------------------------------
```

→ we will get coef of $\hat{\beta}_1 = 0.8206815$ which is not so different from $\hat{\beta}_1 = 0.799$ in the previous regression fn. Since, there is an omitted variable in this regression fn. which we found to be 0.039. That's why the coefficient does not charge much.