```
------------------------------------------------------------------------------
-------------------------------------
      name:  WS
       log:  C:\Users\Will\Desktop\Econ
321\Strata\shao_project_2_HTV2.log
  log type:  text
 opened on:   7 Mar 2021, 23:03:25

.
.
/*===================================================================
=======
> The purpose of this assignment is to show how education affects wages.
> The data set HTV2.dta is an observational data set on a random
> sample of adults in the U.S. in 1991. Open the HTV2.dta data set.
>
===================================================================
=====*/
.
. use "C:\Users\Will\Desktop\Econ 321\Strata\HTV2.dta"

.
.
```

```
/*==================================================================
=======
> (Q1): Summarize and describe the data set.
> (a) How many observations are in the data set?
> (b) How many variables are in the data set?
> (c) How are the wage and educ variables measured?
>
==================================================================
=====*/
.
. summarize

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        wage |      1,193    13.23942    9.116401   1.023529   91.30922
        educ |      1,193    13.03437    2.346208          6         20
          ne |      1,193     .210394    .4077594          0          1
          nc |      1,193    .3730092    .4838073          0          1
        west |      1,193    .1684828    .3744514          0          1
-------------+---------------------------------------------------------
       south |      1,193     .248114    .4320995          0          1
       exper |      1,193    10.72842    3.105527          1         19

. describe, f

Contains data from C:\Users\Will\Desktop\Econ 321\Strata\HTV2.dta
  obs:         1,193
 vars:             7                          20 Sep 2020 17:40
------------------------------------------------------------------------
----------------------------------------
              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------
----------------------------------------
wage            float   %9.0g                 hourly wage in dollars
educ            byte    %9.0g                 years of education
ne              byte    %9.0g                 =1 if person lives in the
Northeast
nc              byte    %9.0g                 =1 if person lives in the
Midwest
west            byte    %9.0g                 =1 if person lives in the
West
south           byte    %9.0g                 =1 if person lives in the
South
exper           byte    %9.0g                 years of work experience
------------------------------------------------------------------------
----------------------------------------
Sorted by:

.
. /* a) There are 1193 observations in the data set
> b) There are 7 variables in the data set
> c) Wages are measured in hourly wage in dollars and the educ variable
is
```

```
> measured in years of education.
> */
.
.
```

```
/*==================================================================
=======
> (Q2): Estimate a bivariate regression relating education to wages.
> Assume homoskedasticity is true. You will consider whether it is
actually true
> in (Q6).
>
==================================================================
=====*/
.
. reg wage educ

      Source |       SS           df       MS      Number of obs   =
1,193
-------------+----------------------------------   F(1, 1191)      =
174.98
       Model |  12689.9011         1  12689.9011   Prob > F        =
0.0000
    Residual |   86375.759     1,191  72.5237271   R-squared       =
0.1281
-------------+----------------------------------   Adj R-squared   =
0.1274
       Total |  99065.6602     1,192  83.1087753   Root MSE        =
8.5161

------------------------------------------------------------------------
-----
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------
-----
        educ |   1.390671   .1051321    13.23   0.000     1.184407
1.596936
       _cons |  -4.887101   1.392335    -3.51   0.000    -7.618804   -
2.155398
------------------------------------------------------------------------
-----

.
.
```

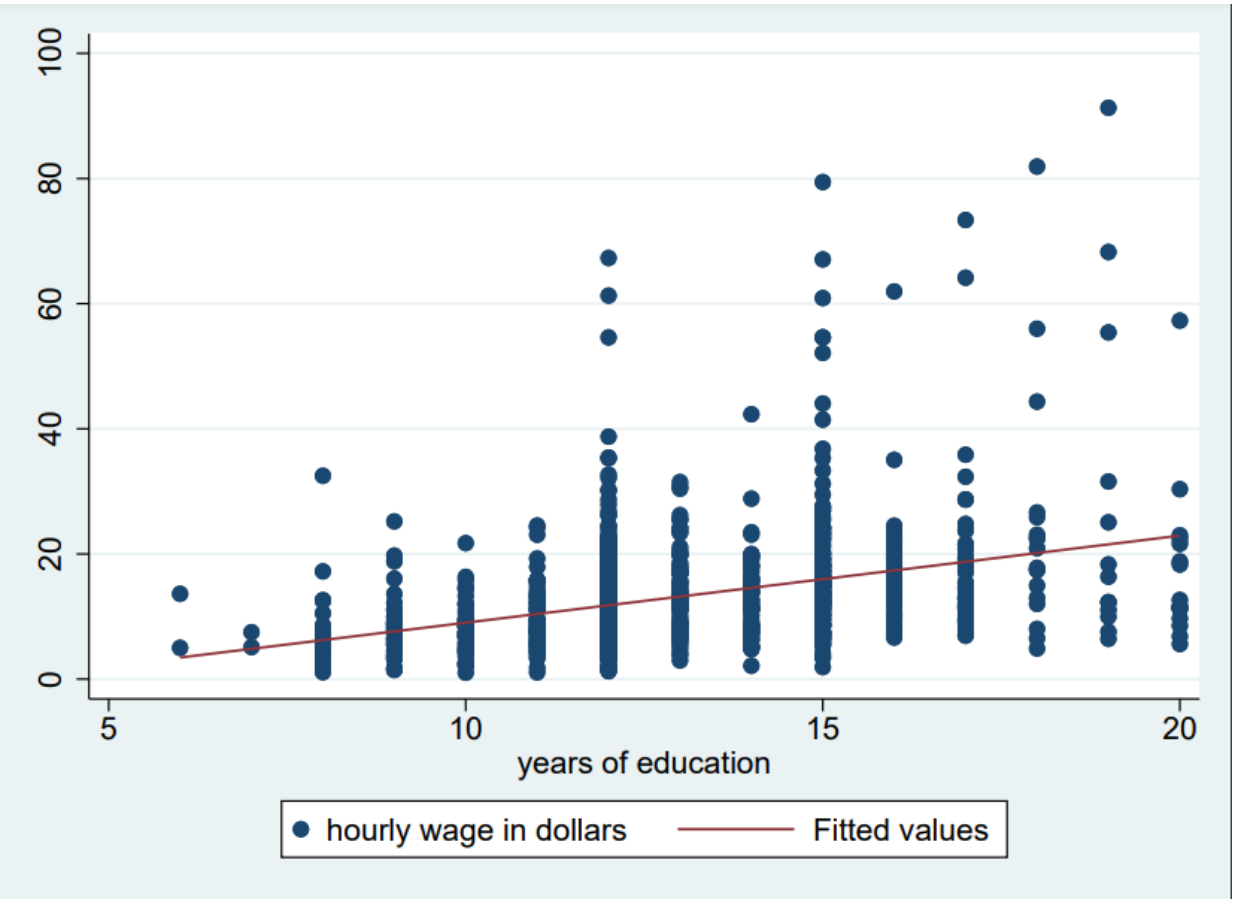
```
/*=========================================================================
=======
> (Q3): Consider the three assumptions that are necessary to achieve
unbiased
> and consistent estimators. Does the model in (Q2) satisfy the first
assumption?
> Why or why not?
>
*=========================================================================
=====*/
. /*
> The fist assumption Coditional mean assumption, which states that the
error term
> has a conditional mean of 0 given regressor X which implies
unbiasedness. The
> model in (Q2) cannot achieve the first assumption since its based on a
> observational data and there will be errors with in data further
leading to bias
> results.
> */
.
.
```

```
/*=====================================================================
=======
> (Q4):  Consider the three assumptions that are necessary to achieve
unbiased
> and consistent estimators. Does the model in (Q2) satisfy the second
assumption?
> Why or why not?
>
*=====================================================================
=====*/
. /*
> The second assumption is that (Xi,Yi) are i.i.d, or independently and
> identitically distributed. The model in (Q2) does satisfy the second
asummption
> because the observational data was collected by taking a random
sampling of
> adults.
> */
.
.
```

```
/*=================================================================================
=======
> (Q5): Consider the three assumptions that are necessary to achieve
unbiased
> and consistent estimators. Does the model in (Q2) satisfy the third
assumption?
> Why or why not?
>
*=================================================================================
=====*/
.
. scatter wage educ || lfit wage educ

.
. graph export "C:\Users\Will\Desktop\Econ
321\Strata\project_2_graph.pdf", replace
(file C:\Users\Will\Desktop\Econ 321\Strata\project_2_graph.pdf written
in PDF format)

.
. *The third assumption states that large outliers are unlikely. The
model in (Q2)
. *does not satisfy the third assumption which can be seen on the graph.
.
.
```



```
/*=========================================================================
=======
> (Q6): Consider the homoskedasticity assumption. Do you think the model
in (Q2)
> exhibits homoskedasticity or heteroskedasticty? Why?
>
*=========================================================================
=====*/
. /*
> The graph that was provided in question 5 allows me to see that there
are
> varying variances with different widths throught the distrubutions
along the
> population regression line, therefore the model in question 2 exhibits
> heteroskedasticty.
> */
.
.
```

```
/*=====================================================================
=======
> (Q7): Consider the normality of the errors assumption. Do you think the
errors
> in the model in (Q2) follow a normal distribution? Why or why not?
Suppose
> assumptions (1)-(4) are true, and the errors follow a normal
distribution. Why
> should we care?
>
*=====================================================================
=====*/
.
. /*
> The model in Q2 doesn't follow a normal distribution since in the graph
there
> is a noticably more data on the right side of the graph. If we are
asumming that
> the errors folow a normal distribution that means that the data
collected in the
> sample will result in a none bias result.
> */
.
.
```

```
/*=========================================================================
=======
> (Q8): Return to your results from (Q2). Interpret beta1hat in a
sentence.
> Round to two decimal places.
>
*=========================================================================
=====*/
.
. *For an additional year of education, the hourly wage increases $1.39
.
.
```

```
/*=============================================================================
=======
> (Q9): Is beta1 statistically significant when alpha=0.01? Use the t-
statistic
> to justify your answer.
>
*=============================================================================
=====*/
.
. *Since it's greater than the critical value for alpha, which is greater
than the two sided *critical 2.58, it'
> s statistically significant.
.
.
.
```

```
/*=============================================================================
=======
> (Q10): Test the null hypothesis that beta1=0 vs. the alternative that
beta1>0.
> Calculate the p-value for this hypothesis test. What do you conclude?
Use
> alpha=0.01.
>
*=============================================================================
=====*/
.
. /*H_0: beta1 = 0
>   H_1: beta1 > 0
>   alpha = 0.01 = 2.58
>   9.27 > 2.58
>   You can conclude that beta1hat is greater than 0*/
.
.
.
```

```
/*=============================================================================
=======
> (Q11): Construct a 90% confidence interval for beta0. Interpret your
confidence
> interval in a sentence. Round beta0hat and its standard error to two
decimal
> places.
>
*=============================================================================
=====*/
.
. *true value of beta0 lies between -7.18 and -2.60, meaning that without
any
. *years of education, wages fall between -$7.18 and -$2.60.
.
.
/*=============================================================================
=======
>
=============================================================================
=====*/
. cap log close _all
```