Q1

1.

```
. sum yield

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
       yield |     14,171    433.6169    395.1982          0   36461.98

. sum rental_in

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
   rental_in |     14,171    .5031543    4.466769          0        200

. sum rental_out

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
  rental_out |     14,171    .2164561    1.489389          0         60
```

(1) The standard deviation of variable yield is large, which means the output of each households' unit of land is very different.
(2) On average, land rented in is more than land rented out by the households.
(3) The difference in land rented in among each household is greater than the difference in land rented out.
(4) At least one household does not rent in land and at least one household does not rent out land.
(5) The household who rents in most rents in 200 units of land and the household who rents out most rents out 60 units of land.

2. I think β_1 will be negative and β_2 will be positive.
   Because when a household has more land, every unit of land will possibly get less attended and less resources (water, fertilizer, etc.). Thus, the output per unit of land will decrease.
   However, when a household has less land, every unit of land will possibly get more attended and more resources. Thus, the output per unit of land will increase.

3. Yes, the estimated values agree with my prediction.

```
. reg yield rental_in

      Source |       SS           df       MS      Number of obs   =    14,171
-------------+----------------------------------   F(1, 14169)     =      1.10
       Model |  172005.965         1  172005.965   Prob > F        =    0.2940
    Residual |  2.2129e+09    14,169  156180.516   R-squared       =    0.0001
-------------+----------------------------------   Adj R-squared   =    0.0000
       Total |  2.2131e+09    14,170  156181.633   Root MSE        =     395.2

------------------------------------------------------------------------------
       yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   rental_in |  -.7799978   .7432502    -1.05   0.294    -2.236866    .6768703
       _cons |   434.0094   3.340807   129.91   0.000      427.461    440.5578
------------------------------------------------------------------------------

. //predict hat_yield_1
. //predict res_yield_1, residuals
. //sum res_yield_1
.
. reg yield rental_out

      Source |       SS           df       MS      Number of obs   =    14,171
-------------+----------------------------------   F(1, 14169)     =      0.09
       Model |  13342.3767         1  13342.3767   Prob > F        =    0.7701
    Residual |  2.2131e+09    14,169  156191.714   R-squared       =    0.0000
-------------+----------------------------------   Adj R-squared   =   -0.0001
       Total |  2.2131e+09    14,170  156181.633   Root MSE        =    395.21

------------------------------------------------------------------------------
       yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
  rental_out |   .6515137   2.229133     0.29   0.770    -3.717879    5.020907
       _cons |   433.4759   3.354809   129.21   0.000     426.9001    440.0518
------------------------------------------------------------------------------
```

4.

β_1-hat: Holding ceteris paribus, when a household rents in one more unit of land, the output per unit of its land will decrease by about 0.7799978 units.

β_2-hat: Holding ceteris paribus, when a household rents out one more unit of land, the output per unit of its land will increase by about 0.6515137 units.

However, these relationships are not statistically significant, because their p-values are obviously greater than 0.05. So **we may fail to reject the null hypothesis**.

5.

```
. predict hat_yield_1
(option xb assumed; fitted values)

. predict res_yield_1, residuals

. sum res_yield_1

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
 res_yield_1 |     14,171   -6.09e-07    395.1829  -434.0094   36027.97
```

```
. predict hat_yield_2
(option xb assumed; fitted values)

. predict res_yield_2, residuals

. sum res_yield_2

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
 res_yield_2 |     14,171    1.10e-06     395.197  -436.8911    36028.5
```

We can see that the residuals approximately sum to 0.

6.

We can see in sub-question 3 that the R^2 of rental_in is 0.0001 (about 0.01% is explained) and the R^2 of rental_out is 0.0000 (about 0.00%is explained).

7.

```
. reg yield rental_in_share

      Source |       SS           df       MS      Number of obs   =    14,171
-------------+----------------------------------   F(1, 14169)     =      0.71
       Model |  111081.624         1  111081.624   Prob > F        =    0.3991
    Residual |  2.2130e+09    14,169  156184.816   R-squared       =    0.0001
-------------+----------------------------------   Adj R-squared   =   -0.0000
       Total |  2.2131e+09    14,170  156181.633   Root MSE        =     395.2

------------------------------------------------------------------------------
       yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
rental_in_share |  .0540097   .0640427     0.84   0.399    -.0715225    .1795418
       _cons |     433.37   3.332743   130.03   0.000     426.8374    439.9026
------------------------------------------------------------------------------

. reg yield rental_out_share

      Source |       SS           df       MS      Number of obs   =    14,171
-------------+----------------------------------   F(1, 14169)     =     24.04
       Model |  3747795.36         1  3747795.36   Prob > F        =    0.0000
    Residual |  2.2093e+09    14,169  155928.149   R-squared       =    0.0017
-------------+----------------------------------   Adj R-squared   =    0.0016
       Total |  2.2131e+09    14,170  156181.633   Root MSE        =    394.88

------------------------------------------------------------------------------
       yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
rental_out_share |  .2350118   .0479362     4.90   0.000     .1410505    .3289732
       _cons |   431.6123   3.342233   129.14   0.000     425.0611    438.1635
------------------------------------------------------------------------------
```

β_1-hat: Holding ceteris paribus, when the proportion of rental_in to total land area increases by 1% (rental_in_share increases by 1 unit), the output per unit of land will increase by about 0.0540097 units.
β_2-hat: Holding ceteris paribus, when the proportion of rental_out to total land area increases by 1% (rental_out_share increases by 1 unit), the output per unit of land will increase by about 0.2350118 units.
However, the relationship between yield and **rental_in_shar**e may not be statistically significant, because its p-value is obviously greater than 0.05. So **we may fail to reject the null hypothesis**.
8.

```
.
. gen ln_yield = log(yield+1)

. reg ln_yield rental_in_share

      Source |       SS           df       MS      Number of obs   =    14,171
-------------+----------------------------------   F(1, 14169)     =      1.94
       Model |  .689144057         1  .689144057   Prob > F        =    0.1642
    Residual |  5045.22149    14,169  .356074634   R-squared       =    0.0001
-------------+----------------------------------   Adj R-squared   =    0.0001
       Total |  5045.91063    14,170  .356098139   Root MSE        =    .59672

------------------------------------------------------------------------------
    ln_yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
rental_in_share |  .0001345   .0000967     1.39   0.164     -.000055    .0003241
       _cons |   5.940788   .0050321  1180.57   0.000     5.930924    5.950651
------------------------------------------------------------------------------

. reg ln_yield rental_out_share

      Source |       SS           df       MS      Number of obs   =    14,171
-------------+----------------------------------   F(1, 14169)     =     19.59
       Model |  6.96783523         1  6.96783523   Prob > F        =    0.0000
    Residual |  5038.94279    14,169  .355631505   R-squared       =    0.0014
-------------+----------------------------------   Adj R-squared   =    0.0013
       Total |  5045.91063    14,170  .356098139   Root MSE        =    .59635

------------------------------------------------------------------------------
    ln_yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
rental_out_share |  .0003204   .0000724     4.43   0.000     .0001785    .0004623
       _cons |   5.938669   .0050475  1176.56   0.000     5.928776    5.948563
------------------------------------------------------------------------------
```

β_1-hat: Holding ceteris paribus, when the proportion of rental_in to total land area increases by 1%, that is, rental_in_share increases by 1 unit, the output per unit of land will increase by about 5.940788%.

β_2-hat: Holding ceteris paribus, when the proportion of rental_out to total land area increases by 1%, that is, rental_out_share increases by 1 unit, the output per unit of land will increase by about 5.938669%.

However, the relationship between ln_yield and **rental_in_share** may not be statistically significant, because its p-value is obviously greater than 0.05. So **we may fail to reject the null hypothesis**.

9.

I prefer model (3) in q8, because the log transformation rescales the values. In this way, we can observe the relationship between the percentage change in yield and the percentage change in the rental_in and rental_out relatively.

10.

Under this context, the assumptions are:

(1) The value of rental_in contains no information about the mean of the unobserved factors.
(2) The value of rental_out contains no information about the mean of the unobserved factors.

Possible reasons:

(1) In reality, the relationship between yield & rental_in and yield & rental_out may not be linear.
(2) There is endogeneity. When we regress yield to rental_in or rental_out, there may be some important factors ignored.
    For example, households who are more skilled (we did not include this into our model) may produce higher yields. Meanwhile, they would like to rent (in) more land for more total output. On the other hand, households not that skilled may rent out some land for not being able to handle so much land. Meanwhile, they also produce lower yield.
    In this case, the variables rental_in and rental_out actually correlate to households' skills, which also influence the yield. In short, the explanatory variables contain some information about the mean of the unobserved factors.

11.

$\beta_1 < 0$, $\beta_2 > 0$

Mechanism 1:

A household may not have enough energy, resources and willingness to attend to too much land. Thus, they may rent out some land to attend the rest more carefully. On the contrary, when a household rents in some land, every unit of land may get less attended so that the output per unit of land will decrease.

Mechanism2:

A household may rent out the less productive land for a higher profit rate. In this case, the output per unit of land will increase.

A household may rent in the less productive land from others for higher total output. In this case, the output per unit of land (which is actually an average value), will decrease.

Q2

1.

$\boldsymbol{yield_i}$: the output per unit of land of village *i* this year.

$\boldsymbol{treat_i}$: a boolean variable indicating whether village *i* received their treatment. 1 for yes and 0 for no.

2.

$$\boldsymbol{yield_i = \beta_0 + \beta_1 treat_i + \mu_i}$$

$\boldsymbol{yield_i}$**:** the output per unit of land of village *i* this year.

$\boldsymbol{treat_i}$**:** a boolean variable indicating whether village *i* received their treatment. 1 for yes and 0 for no.

$\boldsymbol{\beta_0}$**:** the intercept indicating the yield (output per unit of land) without treatment (land rental contract law), which may be the average yield of the control group (the remaining 100 villages).

$\boldsymbol{\beta_1}$**:** the coefficient of interest, which indicates the effect of the treatment (holding ceteris paribus).

$\boldsymbol{\mu_i}$**:** the error term for village *i which may influence* $yield_i$.

3.

(1) Maybe the richest villages are rich because basically their yields are higher. This issue may increase the $\boldsymbol{yield_i}$ we observe, which will magnify the treatment effect we observe. In this case, $\boldsymbol{\beta_1}$ may be too large.

(2) If we use the average yield of the control group, who failed to be selected to receive the treatment, as $\boldsymbol{\beta_0}$, then $\boldsymbol{\beta_0}$ may be smaller than our ideal assumption. Because the villages in the control group may be poorer and their average yield may be lower than the population (all the 200 villages) before we exert treatment. In this case, $\boldsymbol{\beta_1}$ may be too small.

(3) The treatment effect, $\boldsymbol{\beta_1}$, may not be applicable for all the villages. It may just be applicable for richer villages. Because the data we acquired is from the richer villages.

Code

```
clear all

cap log close
log using "122090407.log", replace
cd "/Users/30706/Desktop/eco"

use aghousehold, clear

keep if year == 2010

gen yield = d32/d31
gen rental_in = c10
gen rental_out = c13

sum yield
sum rental_in
sum rental_out

reg yield rental_in
predict hat_yield_1
predict res_yield_1, residuals
sum res_yield_1

reg yield rental_out
predict hat_yield_2
predict res_yield_2, residuals
sum res_yield_2

gen rental_in_share = 100*rental_in/d31
gen rental_out_share = 100*rental_out/d31
reg yield rental_in_share
reg yield rental_out_share

gen ln_yield = log(yield+1)
reg ln_yield rental_in_share
reg ln_yield rental_out_share

log close
```

Log

```
--------------------------------------------------------------------------------
      name:  <unnamed>
       log:  C:\Users\30706\Desktop\eco\122090407.log
  log type:  text
 opened on:   6 Oct 2023, 01:45:31

. cd "/Users/30706/Desktop/eco"
C:\Users\30706\Desktop\eco


.
. use aghousehold, clear


.
. keep if year == 2010
(0 observations deleted)


.
. gen yield = d32/d31

. gen rental_in = c10

. gen rental_out = c13


.
. sum yield

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
       yield |     14,171    433.6169    395.1982          0   36461.98

. sum rental_in

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
   rental_in |     14,171    .5031543    4.466769          0        200

. sum rental_out

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
  rental_out |     14,171    .2164561    1.489389          0         60

.
```

```
.
. reg yield rental_in

      Source |       SS           df       MS      Number of obs   =    14,171
-------------+----------------------------------   F(1, 14169)     =      1.10
       Model |  172005.965         1  172005.965   Prob > F        =    0.2940
    Residual |  2.2129e+09    14,169  156180.516   R-squared       =    0.0001
-------------+----------------------------------   Adj R-squared   =    0.0000
       Total |  2.2131e+09    14,170  156181.633   Root MSE        =     395.2

------------------------------------------------------------------------------
       yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   rental_in |  -.7799978   .7432502    -1.05   0.294    -2.236866    .6768703
       _cons |   434.0094   3.340807   129.91   0.000      427.461    440.5578
------------------------------------------------------------------------------

. predict hat_yield_1
(option xb assumed; fitted values)

. predict res_yield_1, residuals

. sum res_yield_1

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
 res_yield_1 |     14,171   -6.09e-07    395.1829  -434.0094   36027.97

.
. reg yield rental_out

      Source |       SS           df       MS      Number of obs   =    14,171
-------------+----------------------------------   F(1, 14169)     =      0.09
       Model |  13342.3767         1  13342.3767   Prob > F        =    0.7701
    Residual |  2.2131e+09    14,169  156191.714   R-squared       =    0.0000
-------------+----------------------------------   Adj R-squared   =   -0.0001
       Total |  2.2131e+09    14,170  156181.633   Root MSE        =    395.21

------------------------------------------------------------------------------
       yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
  rental_out |   .6515137   2.229133     0.29   0.770    -3.717879    5.020907
       _cons |   433.4759   3.354809   129.21   0.000     426.9001    440.0518
------------------------------------------------------------------------------
```

```
. predict hat_yield_2
(option xb assumed; fitted values)

. predict res_yield_2, residuals

. sum res_yield_2

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
 res_yield_2 |     14,171    1.10e-06     395.197  -436.8911    36028.5

.
. gen rental_in_share = 100*rental_in/d31

. gen rental_out_share = 100*rental_out/d31

. reg yield rental_in_share

      Source |       SS           df       MS      Number of obs   =    14,171
-------------+----------------------------------   F(1, 14169)     =      0.71
       Model |  111081.624         1  111081.624   Prob > F        =    0.3991
    Residual |  2.2130e+09    14,169  156184.816   R-squared       =    0.0001
-------------+----------------------------------   Adj R-squared   =   -0.0000
       Total |  2.2131e+09    14,170  156181.633   Root MSE        =     395.2

------------------------------------------------------------------------------
       yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
rental_in_s~e |   .0540097   .0640427     0.84   0.399    -.0715225    .1795418
       _cons |     433.37   3.332743   130.03   0.000     426.8374    439.9026
------------------------------------------------------------------------------
```

```
. reg yield rental_out_share

      Source |       SS           df       MS      Number of obs   =    14,171
-------------+----------------------------------   F(1, 14169)     =     24.04
       Model |  3747795.36         1  3747795.36   Prob > F        =    0.0000
    Residual |  2.2093e+09    14,169  155928.149   R-squared       =    0.0017
-------------+----------------------------------   Adj R-squared   =    0.0016
       Total |  2.2131e+09    14,170  156181.633   Root MSE        =    394.88

------------------------------------------------------------------------------
       yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
rental_out_~e |   .2350118   .0479362     4.90   0.000     .1410505    .3289732
       _cons |   431.6123   3.342233   129.14   0.000     425.0611    438.1635
------------------------------------------------------------------------------

.
. gen ln_yield = log(yield+1)

. reg ln_yield rental_in_share

      Source |       SS           df       MS      Number of obs   =    14,171
-------------+----------------------------------   F(1, 14169)     =      1.94
       Model |  .689144057         1  .689144057   Prob > F        =    0.1642
    Residual |  5045.22149    14,169  .356074634   R-squared       =    0.0001
-------------+----------------------------------   Adj R-squared   =    0.0001
       Total |  5045.91063    14,170  .356098139   Root MSE        =    .59672

------------------------------------------------------------------------------
    ln_yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
rental_in_s~e |   .0001345   .0000967     1.39   0.164     -.000055    .0003241
       _cons |   5.940788   .0050321  1180.57   0.000     5.930924    5.950651
------------------------------------------------------------------------------
```

```
. reg ln_yield rental_out_share

      Source |       SS           df       MS      Number of obs   =    14,171
-------------+----------------------------------   F(1, 14169)     =     19.59
       Model |  6.96783523         1  6.96783523   Prob > F        =    0.0000
    Residual |  5038.94279    14,169  .355631505   R-squared       =    0.0014
-------------+----------------------------------   Adj R-squared   =    0.0013
       Total |  5045.91063    14,170  .356098139   Root MSE        =    .59635

------------------------------------------------------------------------------
    ln_yield | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
rental_out_~e |   .0003204   .0000724     4.43   0.000     .0001785    .0004623
       _cons |   5.938669   .0050475  1176.56   0.000     5.928776    5.948563
------------------------------------------------------------------------------

.
. log close
      name:  <unnamed>
       log:  C:\Users\30706\Desktop\eco\122090407.log
  log type:  text
 closed on:   6 Oct 2023, 01:45:32
-------------------------------------------------------------------------------
```