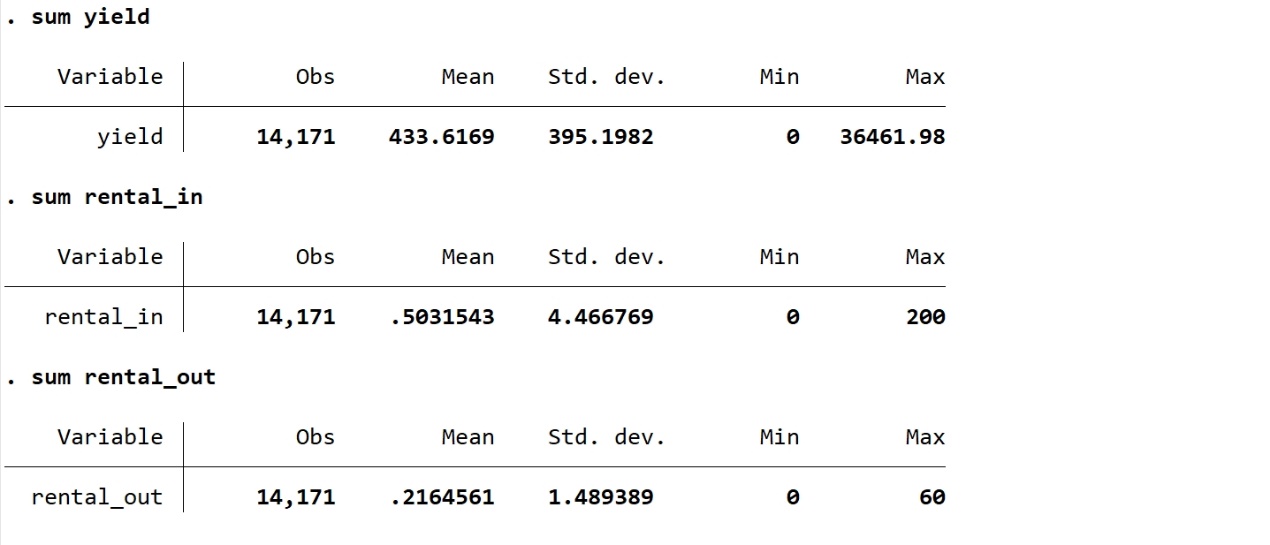
Q1



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.
6. I think β1 will be negative and β2 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.

1. 手机屏幕截图

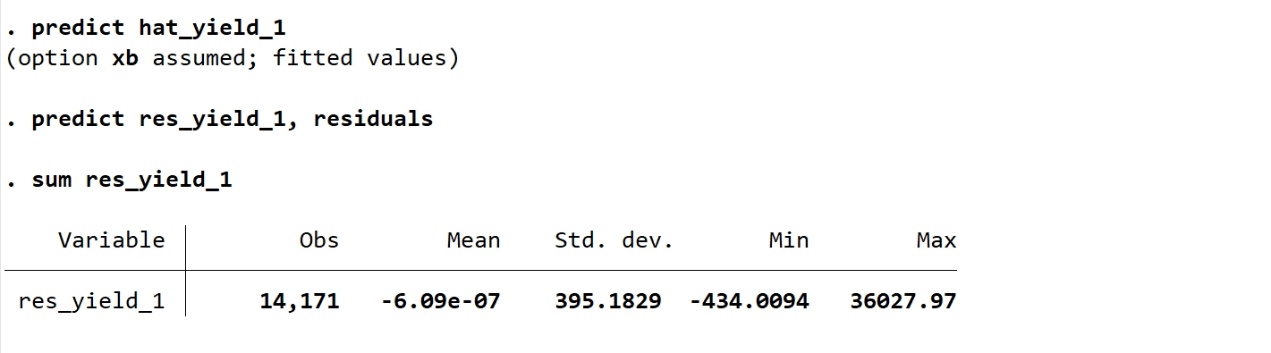
   Description automatically generatedYes, the estimated values agree with my prediction.

β1-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.

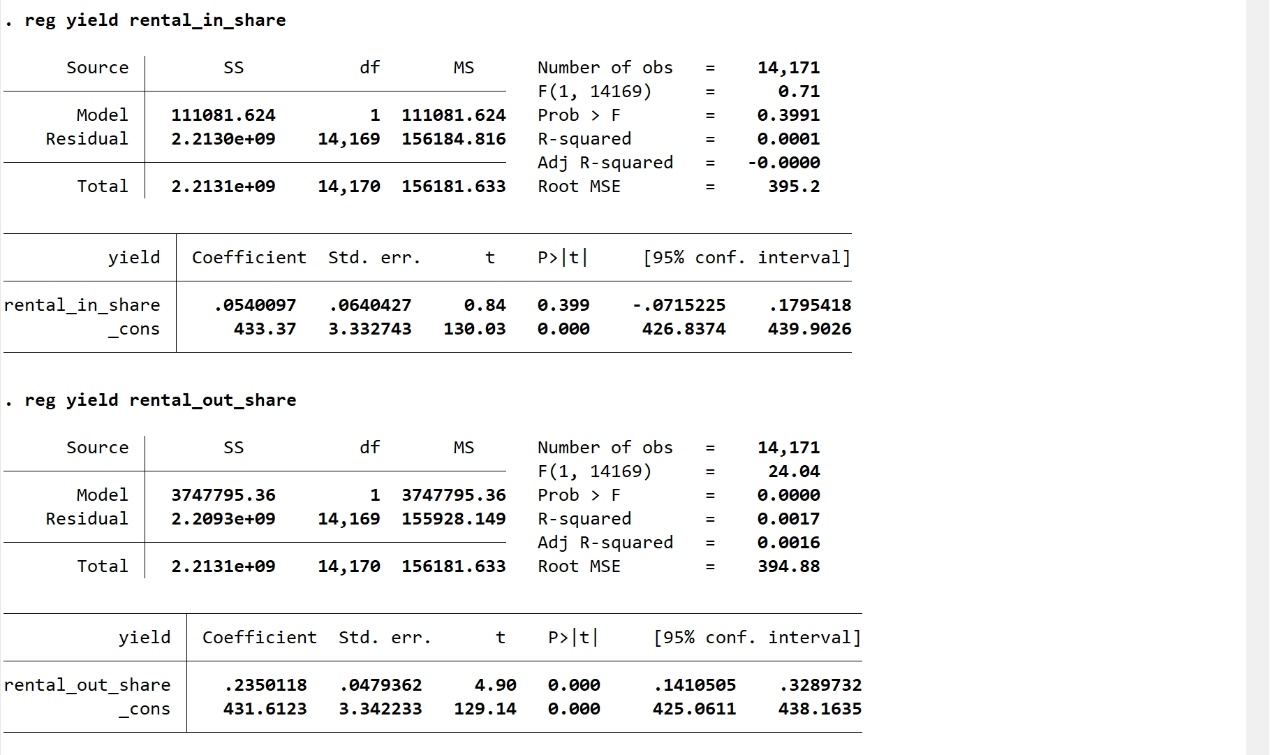
β2-hat: Holding ceteris paribus, when a household rents out one more unit of land, the output per unit of its land will increase by about 0.6515137 units.

**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**.

表格

Description automatically generatedWe can see that the residuals approximately sum to 0.

We can see in sub-question 3 that the R2 of rental\_in is 0.0001 (about 0.01% is explained) and the R2 of rental\_out is 0.0000 (about 0.00%is explained).



β1-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.

β2-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**.

手机屏幕截图

Description automatically generated

β1-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%.

β2-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**.

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.

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.

β1 < 0, β2 > 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

***yieldi***: the output per unit of land of village *i* this year.

***treati***: a boolean variable indicating whether village *i* received their treatment. 1 for yes and 0 for no.

***yieldi = β0 + β1treati + μi***

***yieldi*:** the output per unit of land of village *i* this year.

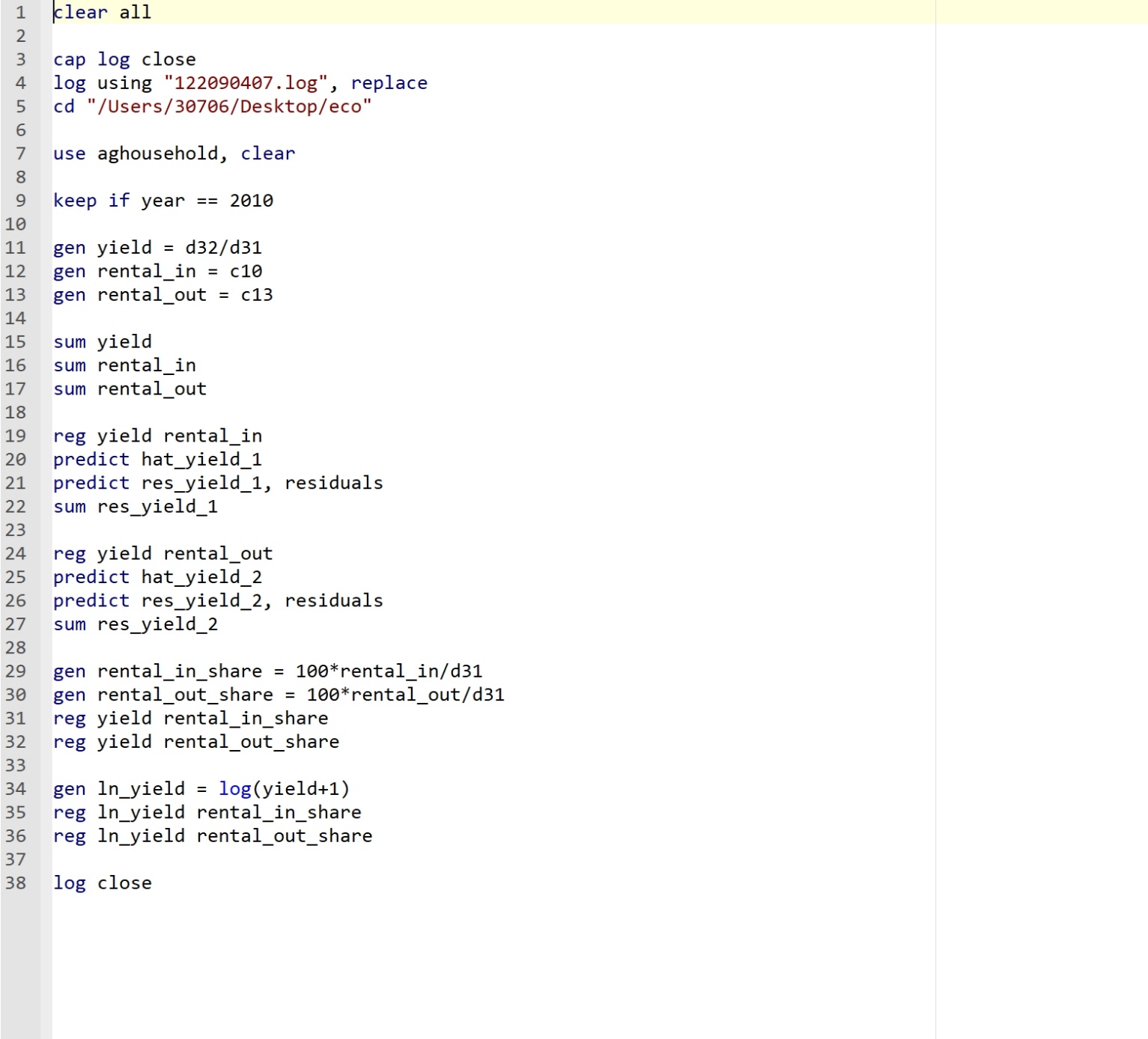
***treati*:** a boolean variable indicating whether village *i* received their treatment. 1 for yes and 0 for no.

***β0*:** 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).

***β1*:** the coefficient of interest, which indicates the effect of the treatment (holding ceteris paribus).

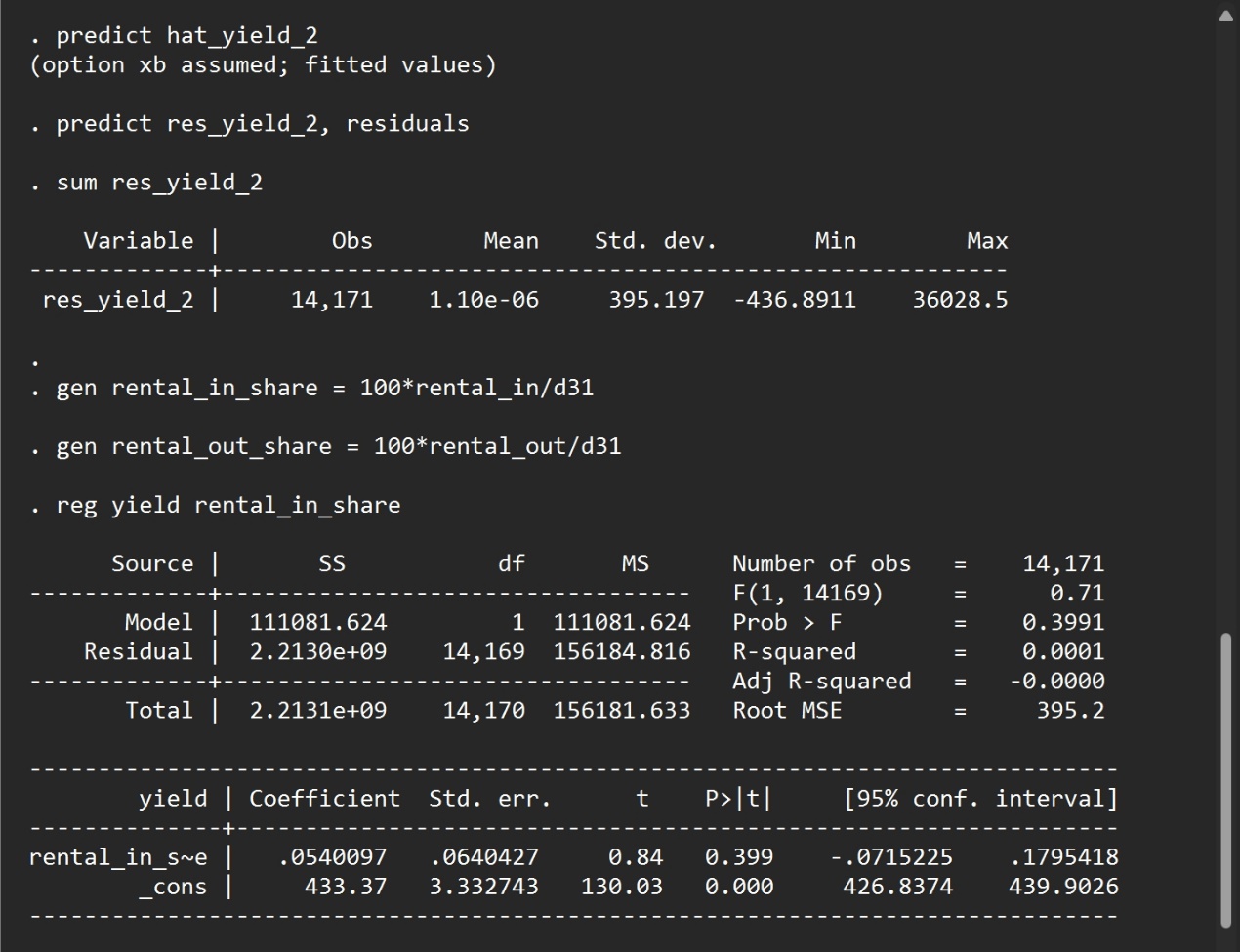
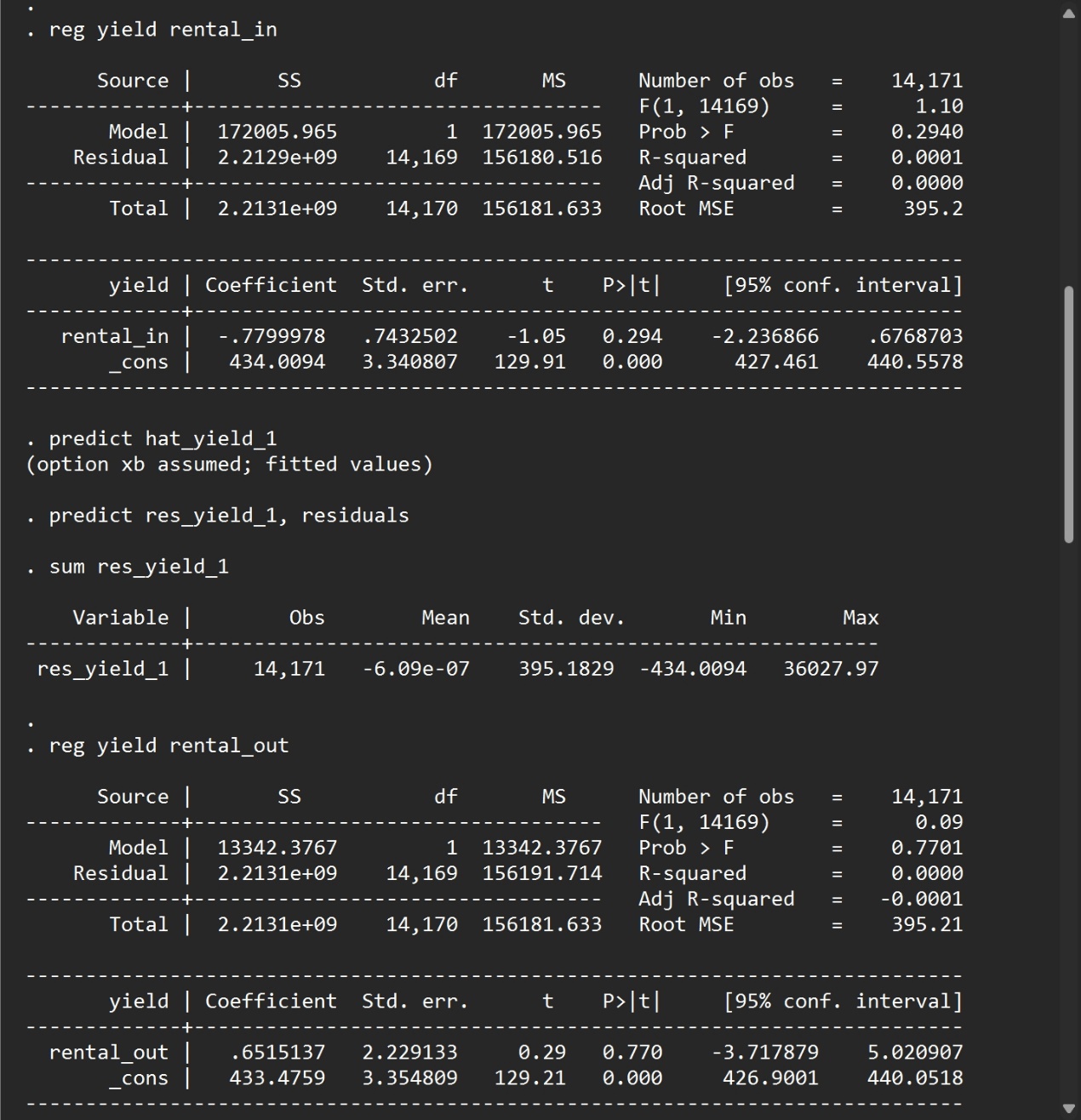
***μi*:** the error term for village *i which may influence yieldi*.

1. Maybe the richest villages are rich because basically their yields are higher. This issue may increase the ***yieldi*** we observe, which will magnify the treatment effect we observe. In this case, ***β1*** 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 ***β0***, then ***β0*** 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, ***β1*** may be too small.
3. The treatment effect, ***β1***, 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**

**Log**

**文本

Description automatically generated文本

Description automatically generated文本

Description automatically generated**