**Appendix (Statistical Analysis):**

This section will focus on the analysis of the models to further explore how agricultural profits are determined in Ghana. During our study, we focused on how agricultural profit in Ghana is affected by variables collected during the GLSS 4 survey.

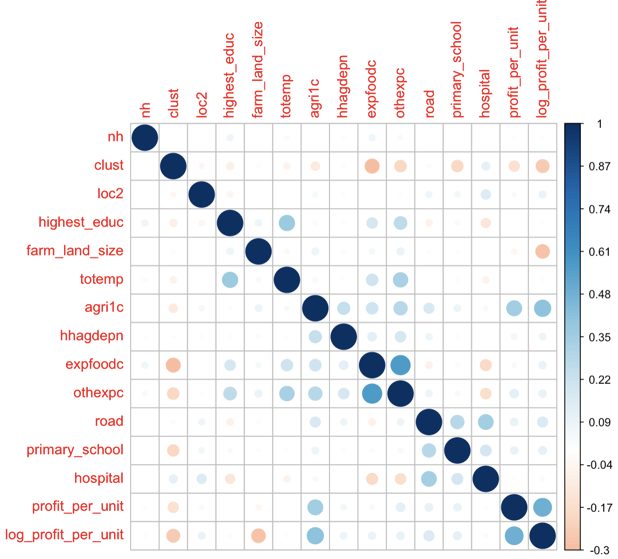
**Hypothesis Testing:**

Ho: Educational attainment does not affect agricultural profit per area unit

HA: Educational attainment does affect agricultural profit per area unit

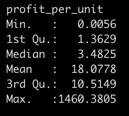
**Correlation Matrix:**

The figure below presents a visual of correlation between the variables that we used in our Base Model.

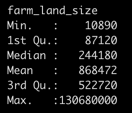


**Summary statistics for key variables:**

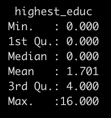
Profit per Unit (dependent variable):



Farm Land Size (key explanatory variable):



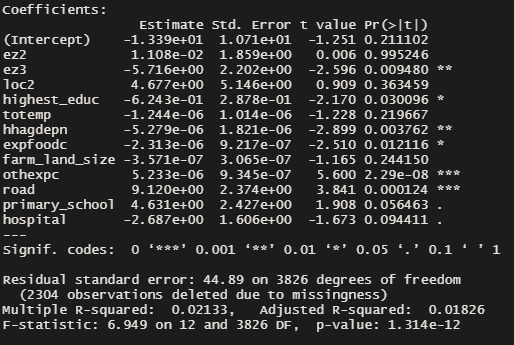
Highest Education(key explanatory variable):



**Regression 1:**

We referred to ‘regression\_1’ as our ‘Base Model’. Our aim for this model was to find variables with statistical significance based on our initial assumptions. Multiple steps were taken to check the validity of our base model using the Residual Standard Error, Multiple R-Squared, and p-value. We ensured that our correlation matrix had no NA’s included. Through this base model, we found six variables with statistical significance. These variables were:

Regression results for regression\_1:



Dependent variable:

profit\_per\_unit: We used ‘profit\_per\_unit’ as our dependent variable since this is what represents profit to us. We measured against the other explanatory variables to see the effect, if any, they had on profit. We calculated the profit per unit by dividing agricultural profit (agri1c) by the farm\_land\_size.

Explanatory variables with statistical significance:

ez3 (ecological zone 3): Ecological Zone 3 was statistically significant at the 1% level with a p-value of 0.0095. This tells us that the ‘Savannah’ ecological zone is less profitable than ‘Coastal’ since the effect is -5.176.

highest\_educ (highest education): Our analysis shows that the households with higher levels of education have a negative effect of -0.624 on agricultural profitability. Our model shows this to be true at the 5% level.

hhagdepn (Depreciation of farm equipment): Quality gear is paramount for any profession. In order to be successful, you must be able to depend on the tools that you help you accomplish your objective. A higher rate of depreciation of farm equipment would indicate higher expenditures on farm equipment for replacements, maintenance, or upgrades. We included this variable to see if investing more money in higher quality farm equipment would affect profitability form the agricultural land. ‘hhagdepn’ was found to be statistically significant at the 1% level of significance.

expfoodc (Expenditure on food by households): There is a decent effect on profit from the amount of money spent on food for the household. ‘expfoodc’ was found to be statistically significant at the 5% level of significance.

othexpc (Other Expenditures): Other Expenditures was found to have a slightly positive effect, significant at the 1% level.

road (Presence of a road in the town): ‘road’ was found to be statistically significant at the 1% level of significance.

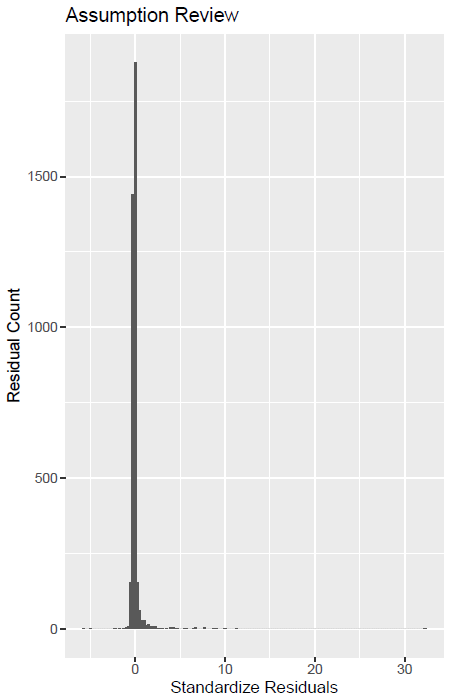
primary\_school: The completion of primary school was statistically significant at the 10% level. There was a positive effect of 4.63 on agricultural profit.

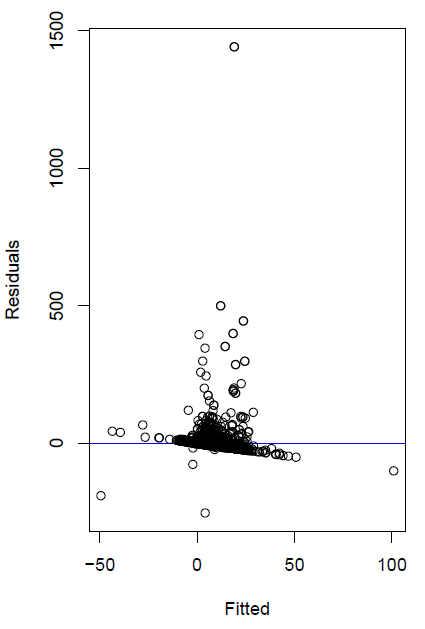
hospital: The presence of a hospital to a locality was statistically significant at the 10% level. Surprisingly, this had a negative effect on profit at -2.687.

Model validation:

Regression 1 Histogram and Fitted vs. Residual Plot:

Looking at both our plots for regression\_1, we can infer that there are issues with our data, resulting in a skewed histogram and fitted chart.

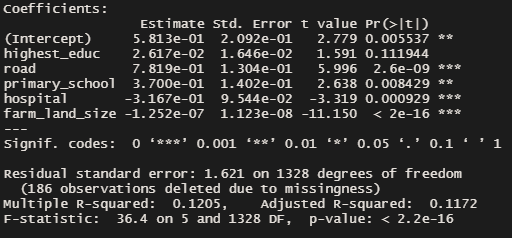




**Regression 2:**

After seeing the results of our base model, we made adjustments to normalize the profit per unit by taking the log of this variable. We had to create a new data frame to get the log of profit for all positive profit household records since we were unable to take the log of negative profits. By taking these steps, we came up with four variables with statistical significance.

Regression results for regression\_2:



Dependent variable:

log\_profit\_per\_unit: this was taken by filtering out data points < 0. We created a new data frame to apply this filter and created a new variable which takes the log of our original dependent variable ‘profit\_per\_unit’.

Explanatory variables with statistical significance:

road (road): We discovered that areas with roads that connect the community had a positive effect on profit of agricultural land. This makes sense since transportation of goods is essential to meet the demands of the community. ‘road’ was found to be statistically significant at the 1% level of significance.

primary\_school (Completion of Primary School in household): We were especially interested in what effect household educational attainment had on profit of agricultural land. Our assumption was that higher education attainment would lead to higher profits but were surprised to see that the relationship shows just a slightly positive trend of 0.37. ‘primary\_school’ was found to be statistically significant at the 1% level of significance.

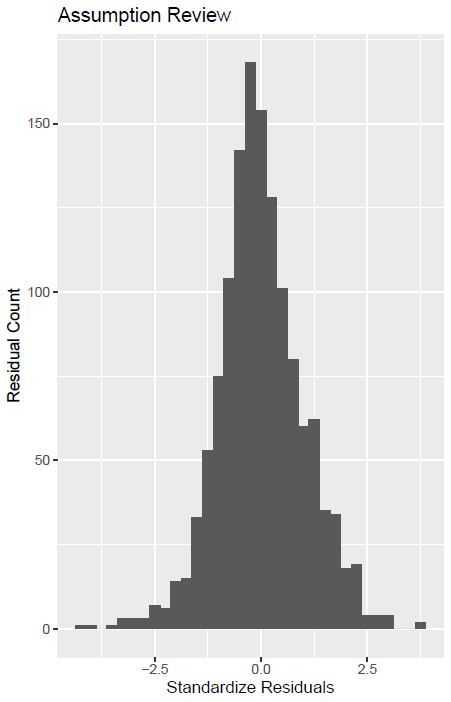
hospital (presence of a hospital): We were interested to see the effect of services offered in different localities, so the presence of a hospital would be integral to sustaining life in a particular area. To our surprise, our model showed a negative effect on profit of -3.17 which was significant at the 1% level.

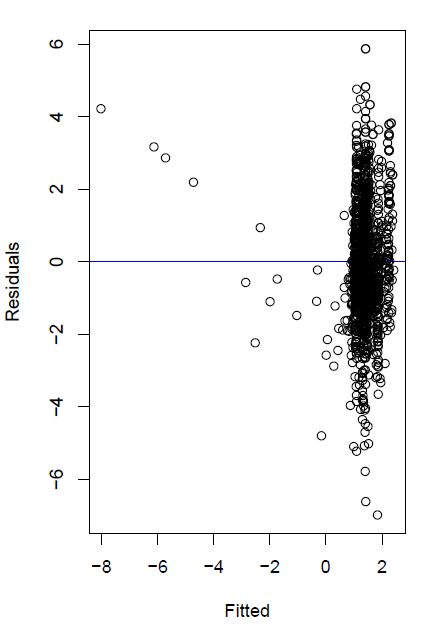
farm\_land\_size (Size of the farm): We originally assumed that bigger farm sizes would yield higher profits, but that was not the case in our model. We found that the size of the farm gave us a slightly negative trend of -0.00000013 which was significant at the 1% level.

Model validation:

Regression 2 Histogram and Fitted vs. Residual Plot:

Looking at both our plots for regression\_2, we can infer that there are issues with our data, resulting in a skewed histogram and fitted chart.

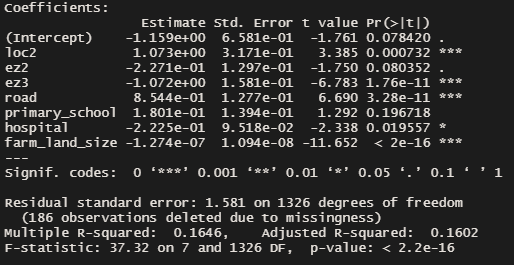




**Regression 3:**

After running our base model (regression\_1) and our model with the log of profits (regression\_2), we wanted to see the effect that locality and ecological zone had on the log of profit per unit. We took the factor of the ‘ez’ variable to distinguish the three different ecological zones of ‘Coastal’, ‘Forest’, and ‘Savannah’. Our summary statistics for ‘loc2’ in this regression show that households in an urban area yield 1.07% more profit than rural areas. Profits showed a negative trend when in the ‘Savannah’ ecological zone (ez3).

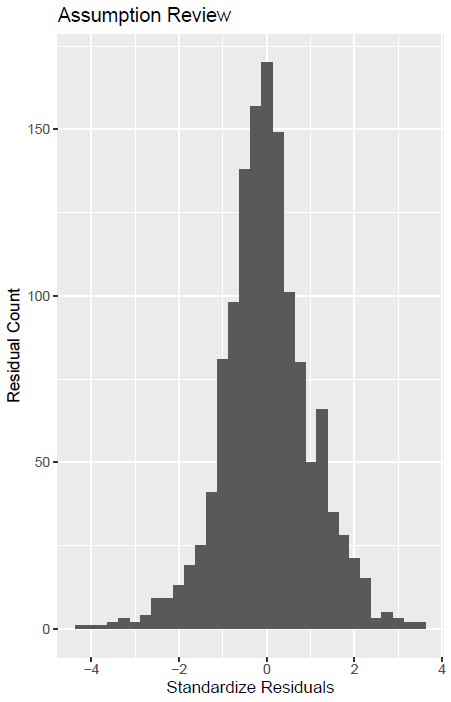
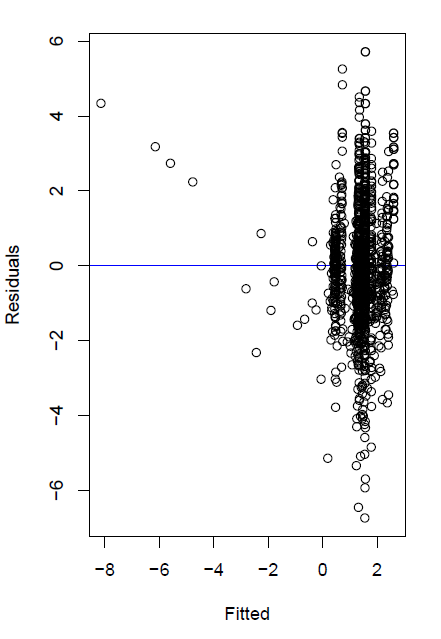
Regression results for regression\_3:



Model validation:

Regression 3 Histogram and Fitted vs. Residual Plot:

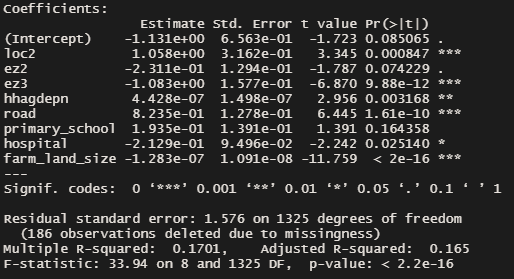
This histogram is also normally distributed, like regression\_2. The Fitted vs. Residual plot begins to show some variation.



**Regression 4:**

After running all our models, we wanted to look at the effect of one more variable that did not deal with locality or education. The variable we chose was the household depreciation of farm equipment (‘hhagdepn’). This variable did have a slightly positive effect on profit at the 1% level with a p-value of 0.0032. We included this variable in our base model and wanted to see the effect when tested against the log of profit per unit.

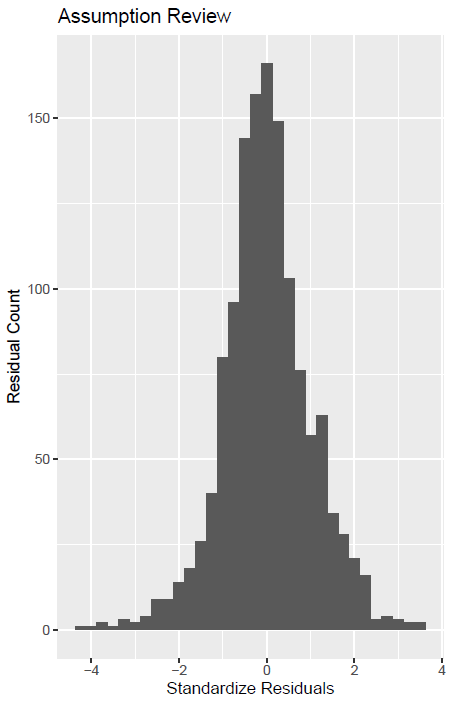
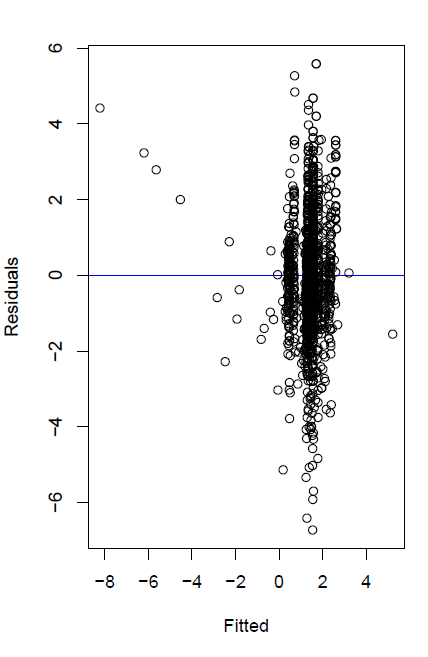
Regression results for regression\_4:



Model validation:

Regression 4 Histogram:

We did not expect a significant difference between regression\_3 and regression\_4 as we only included one extra variable in the model, ‘hhagdepn’. The histogram results remain like the previous two regressions and the Residual vs. Fitted graph shows more variance than previous models.



**Conclusion:**

Based on the analysis of our models, we fail to reject our null hypothesis due to insignificant levels of significance for the highest education explanatory variable. However, an interesting finding would be that if there is a primary education facility in a community, it was found to have a significant effect on agricultural profits. As we refined our models, the R^2 value continued to increase, indicating that predictability increases. In Regression 4, our R^2 value was 0.17 showing that 17% of the explanatory variables impact the dependent variable.