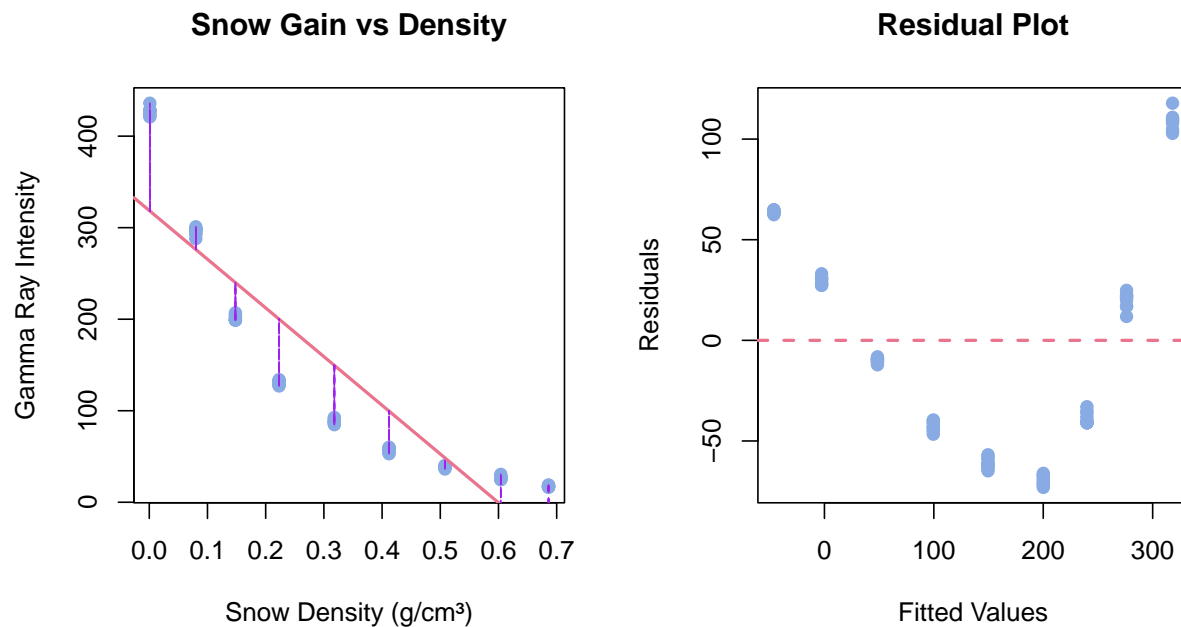


Question 1: Fit a Regression Line

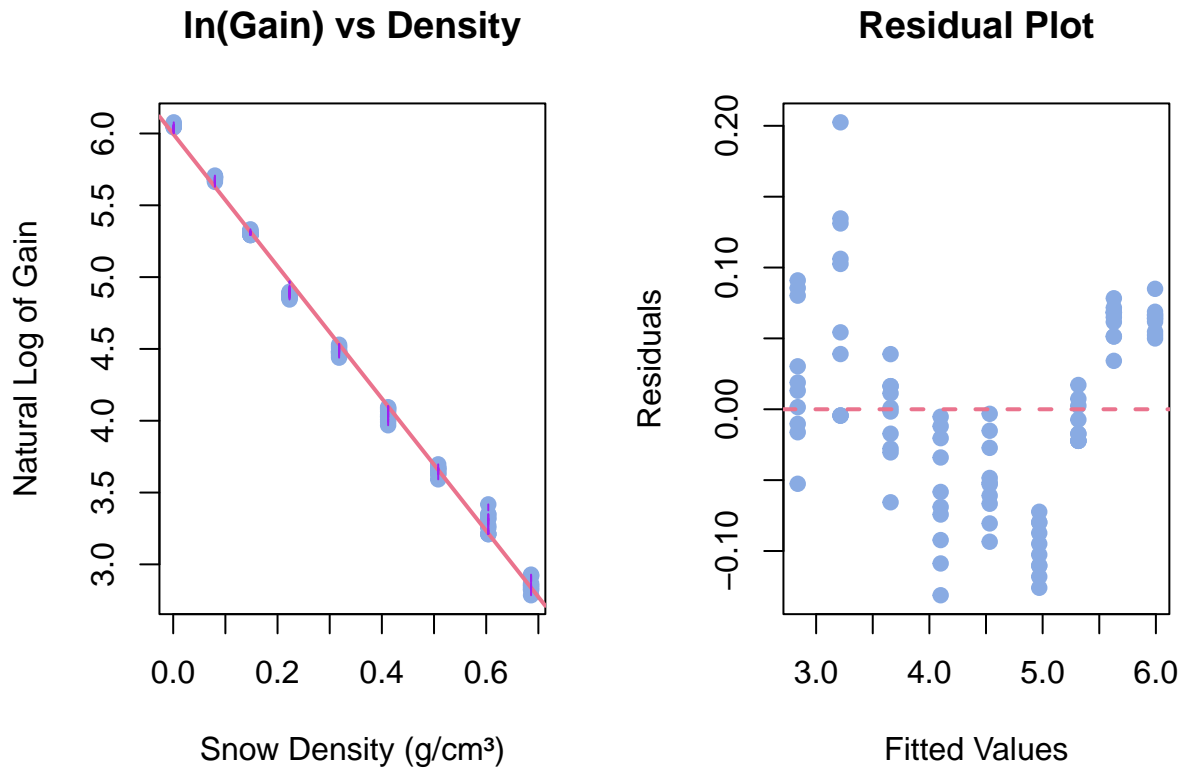


We implemented a regression line along with a residual graph for the snow gain and density measured to determine what type of relationship could be found when analyzing graphically. The regression graph does not look like it follows a line of best fit, potentially showing that the relationship between these two variables is not linear. Additionally, when plotting the residual plot, the same conclusion can be reached. Since the residual plot does not look randomly distributed along the zero line, as a linear relationship would typically be, it shows that our line of best fit doesn't capture the relationship between our variables correctly. Since the line follows an almost parabolic curve, this indicates that there's a good chance that the initial relationship between the two variables could be quadratic, meaning the data needs a transformation or a different model to test something that can accurately capture their relationship.

Additionally, our R^2 of the model turns out to be 0.816, which means that only 81.6% of the variance of the gain can be explained by the variance in the snow density, meaning there is a likely chance that this model is not one that is extremely predictive of the data.

Question 2: Transform Data and Fit

Based on the theoretical exponential relationship between density and gain, we applied a natural logarithm transformation to gain and looked for a linear relationship.



```
## Logarithm Regression Results
```

```
##
## Call:
## lm(formula = log_gain ~ density, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.131216 -0.052396 -0.004436  0.054607  0.202447
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.99727    0.01274   470.8  <2e-16 ***
## density     -4.60594    0.03182  -144.8  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06792 on 88 degrees of freedom
## Multiple R-squared:  0.9958, Adjusted R-squared:  0.9958
## F-statistic: 2.096e+04 on 1 and 88 DF, p-value: < 2.2e-16
```

We compare the above regression results to the regression results before applying the transformation.

```
## Regression Results Before Transformation
```

```
##
## Call:
## lm(formula = data$gain ~ data$density)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -73.08 -44.29  -9.72   30.82 117.83
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    318.70     10.79   29.54  <2e-16 ***
## data$density  -531.95     26.95  -19.73  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 57.54 on 88 degrees of freedom
## Multiple R-squared:  0.8157, Adjusted R-squared:  0.8136
## F-statistic: 389.5 on 1 and 88 DF,  p-value: < 2.2e-16
```

According to physics, the relationship between density and gain is exponential, not linear. Transforming the data with a natural logarithm reveals a linear relationship because the theoretical relationship is

$$g = Ae^{Bd}$$

, which transforms into

$$\ln(g) = \ln(A) + Bd$$

, establishing a theoretical linear relationship between density and the logarithm of gain.

According to the results of the regression, the logarithm transformation greatly improves the strength of correlation detected. After performing the logarithm, the

$$R^2$$

increased from **0.816** to **0.996**. This means 99.6% of the variance in the logarithm of gain can be explained by variance in the snow density. Looking at the actual values of residuals, they are in the range of **-0.15 and 0.2**, compared to the pre-transformation values of **-80 and 130**. This also shows a massive improvement in the regression after applying the logarithm transformation.

The shape of the residual plot still has a very slight parabolic shape but it is much less clear than the shape of the pre-transformation residual plot. Additionally, the dramatic decrease in residual values resulting from the transformation is evidence for its validity.

Question 3: Effect of Random Error

Question 4: Forward Prediction

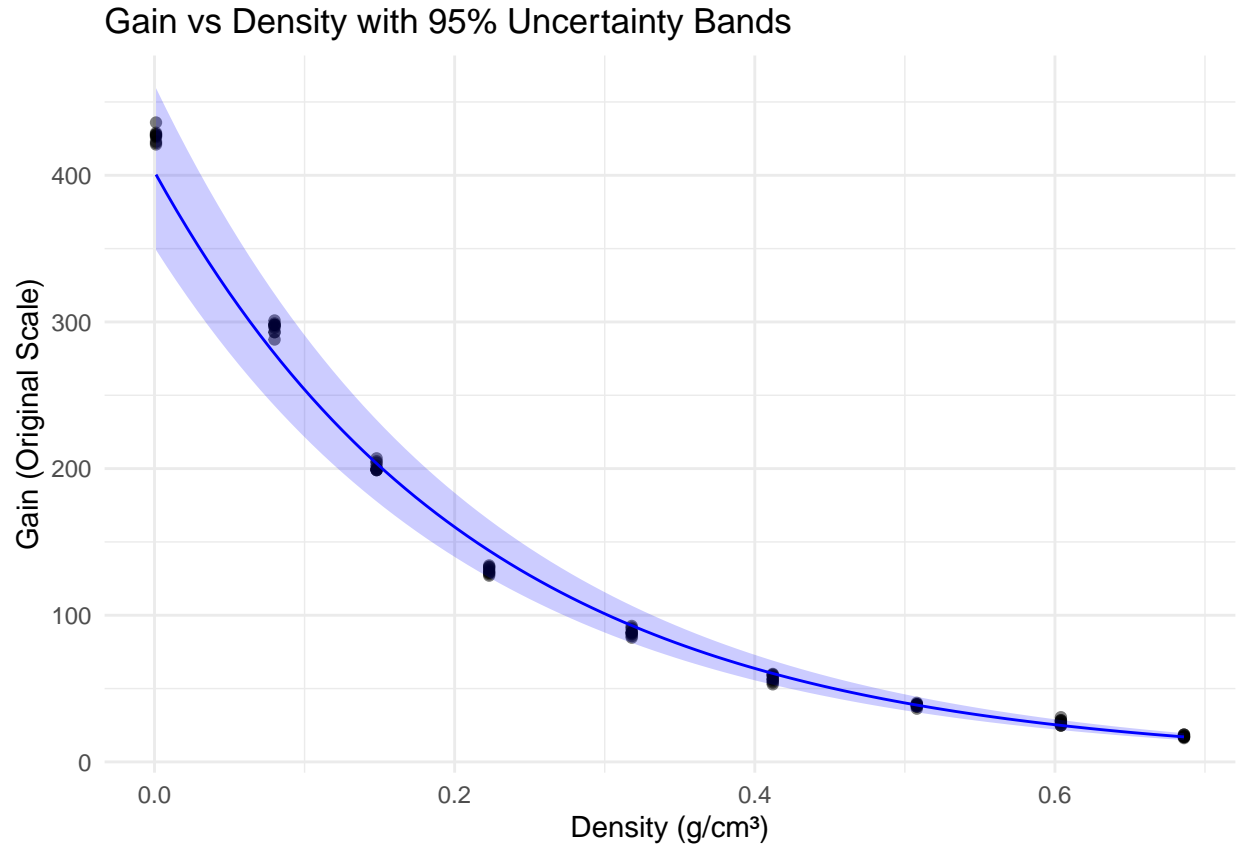


Table 1: Prediction Intervals (Original Scale)

Density	Predicted Value	95% Uncertainty Bands	Actual Range
0.001	400.48	349.09 to 459.43	421 to 436
0.080	278.33	242.78 to 319.08	288 to 301
0.148	203.48	177.57 to 233.18	199 to 207
0.223	144.05	125.74 to 165.01	127 to 134
0.318	93.00	81.2 to 106.52	84.7 to 92.7
0.412	60.32	52.66 to 69.09	52.9 to 60
0.508	38.76	33.83 to 44.42	36.3 to 40.3
0.686	17.07	14.88 to 19.59	16.2 to 18.7

In order to do a forward prediction with the logarithmic model we created earlier, we have to untransform its predictions after predicting gain from density. In order to do that, we exponentiated with the natural base in order to get predicted values in our original untransformed scale. Second, we produced point estimates and uncertainty bands (with a 95% confidence interval) for predicting the gain, where the returned output of our function returns the interval as well as the point estimate of the predicted function.

Looking through the graph that we created, it can be seen that the 95% confidence interval uncertainty bands that we created worked well, encompassing almost all of the observed values. However, it can be seen

that some gain values are more easily predicted than others. For instance, if we look at the given densities of 0.508 g/cm^3 and 0.001 g/cm^3 , it's seen that the predicted value and uncertainty band of the gain at 0.508 is very much in line with the actual range of the observed values, as the predicted value of 38.76 gain is well within the actual range of 36.3 to 40.3 gain in the observed data. This gives us a sense that at higher values of density within the data like 0.508 g/cm^3 , our model is more accurate. On the other hand, at density 0.001 g/cm^3 , while our uncertainty band does capture the actual range of the observed data, with the actual range of gain of 421 to 436 gain within our uncertainty band of $349 - 459$ gain, our predicted value does not. Our predicted value of 400.48 is not in the actual range of $421 - 436$ gain present in the observed data, implying that while our model does a decent job of generalization and prediction of the model, it struggles with the very small value of 0.001 g/cm^3 at the beginning of the observed densities, which can also be noticed with the decreasingly small size of the density bands as the density increases. Some important parts to notice: the model predicts best at densities 0.148 , 0.508 , and at 0.686 , where the predicted value falls within the actual range of the data. On the other hand, densities of 0.001 , 0.080 , 0.223 , 0.318 and 0.412 have uncertainty bands that capture the actual range but fail to predict a value that falls within the actual range. Overall, an observation to notice is that as the density increases within our predicted model, the uncertainty band decreases in size, potentially exhibiting that gain converges to smaller and more tightly observed values easily as density increases.

Question 5: Reverse Prediction

Reverse Prediction of Density for All Gains

Blue: Predicted Density, Red: True Density

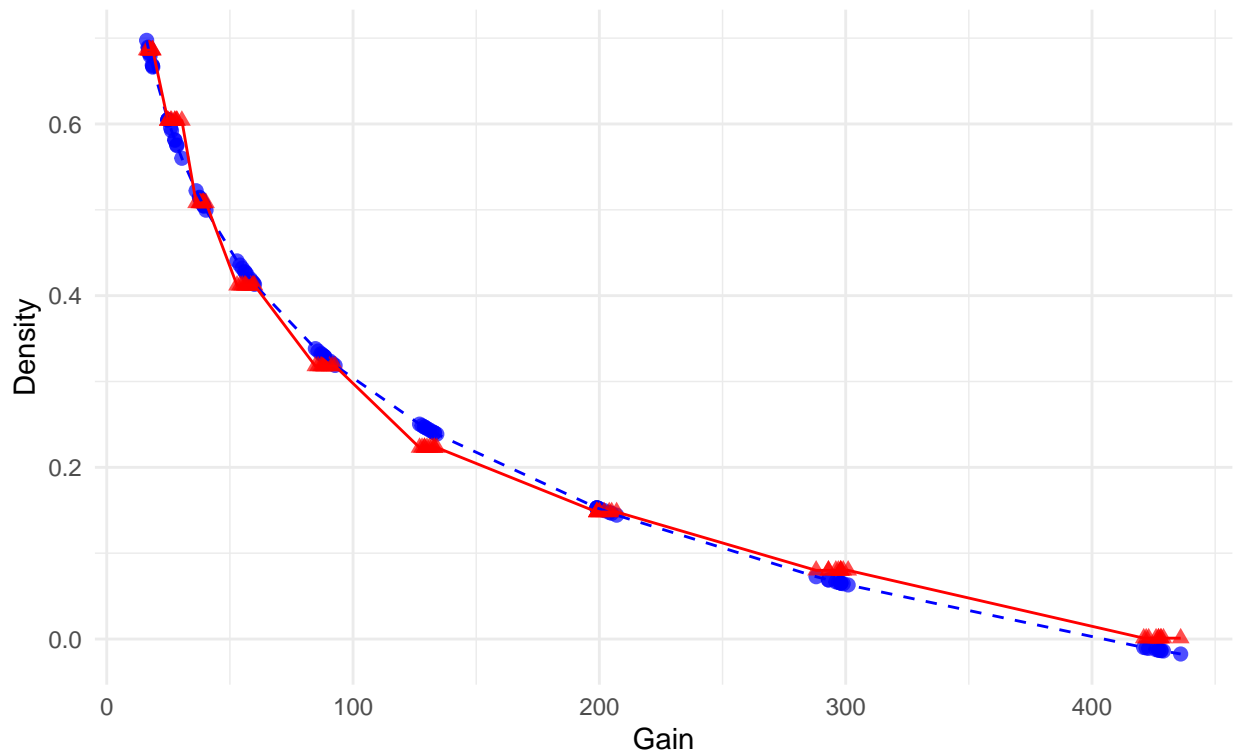


Table 2: Predicted Density From Average Gain

Gain	Predicted Density	Difference in Predicted and True Densities
38.6	0.5089113	0.0009113
426.7	-0.0127695	0.0137695

In order to create a prediction for the averaged measured gains, we had to reverse the prediction from our models using the coefficients from our fitted logarithmic model. After doing this, in order to extrapolate our logarithmic model through all of the data, we created a graphical representation to illustrate the affect of gains and what a predicted density would entail. Doing this, we noticed some values that were easier to predict than others. For instance, at a gain level of ~ 40 and ~ 200 , the predicted and true densities are nearly identical. On the other hand, there are gains where the predicted and true densities are different, as seen in the graph above. For instance, at gains ~ 130 and ~ 425 , the densities are noticably different from each other. In fact, in the table above, it can be seen that the difference between the predicted and true densities in a gain like 38.6 is very minimal, with a 0.0009 difference, meaning our model predicts these gains very well. On the other hand, at a gain of 426.7, the difference in predicted and true density is 0.0137, which is significant, since the density scale only goes up to 0.686. Additionally, looking at our model, it can be observed that typically, the model generalizes better for predictions where the gain is smaller, because when the gain is large, the model can predict negative density values, as it did for the gain of 426.7. This implies that the smaller gains are easier to predict the density of, while the larger gains have densities that are harder to predict.