

# HW4

2024-11-19

## Contribution Statement

Student 1: Drafted the Introduction. Did the methods, analysis, and conclusions for questions 1, 2, 3, and 4.

Student 2: Did the methods, analysis, and conclusions for questions 5 and 6. Did the advanced analysis. Did the conclusions and discussion. Finished up the introduction.

## Introduction

Monitoring water supply in the Sierra Nevada mountains of Northern California is critical, particularly during winter when snowpacks serve as a major source of water. To measure snow density, a gamma transmission snow gauge is used, which indirectly estimates density by detecting gamma ray emissions. The calibration process involves placing polyethylene blocks of known densities between the gauge poles and recording the corresponding gamma ray intensity, referred to as the “gain.” This study seeks to establish a reliable relationship between measured gamma ray intensities and snow density, enabling accurate calibration of the snow gauge for practical use.

The relationship between density and gamma ray intensity follows an exponential decay pattern, described by the equation  $g = Ae^{\beta d}$ , where  $g$  is the gain,  $d$  is the density, and  $A$  and  $\beta$  are parameters determined during calibration. This analysis addresses several key questions: (1) Is the exponential model an appropriate fit for the data, and how does it compare to a linear model? (2) How robust is the model when specific density data points are omitted? (3) How does the model perform across the range of densities, particularly at extreme and intermediate values? Additionally, the study explores the sensitivity of the model to parameter variations and identifies potential limitations in its practical application.

The analysis addresses several questions:

1. Can the raw data be modeled effectively with a linear regression, and does the residual analysis suggest a need for a transformation?
2. What transformation best describes the relationship, and how does the final model perform based on theoretical and empirical justifications?
3. How robust is the model to errors in reported density values, and what impact does this have on predictions?
4. How accurately can the model predict gain values and their uncertainty intervals for specific densities (e.g., 0.508 and 0.001), and are certain gains more predictable than others?
5. How well does the model perform in reverse prediction, mapping gain to density, and how does this compare to true values? Are some densities harder to predict?
6. How does cross-validation, where specific density values are excluded from the fitting process, affect the model’s accuracy in forward and reverse predictions?

## Question 1

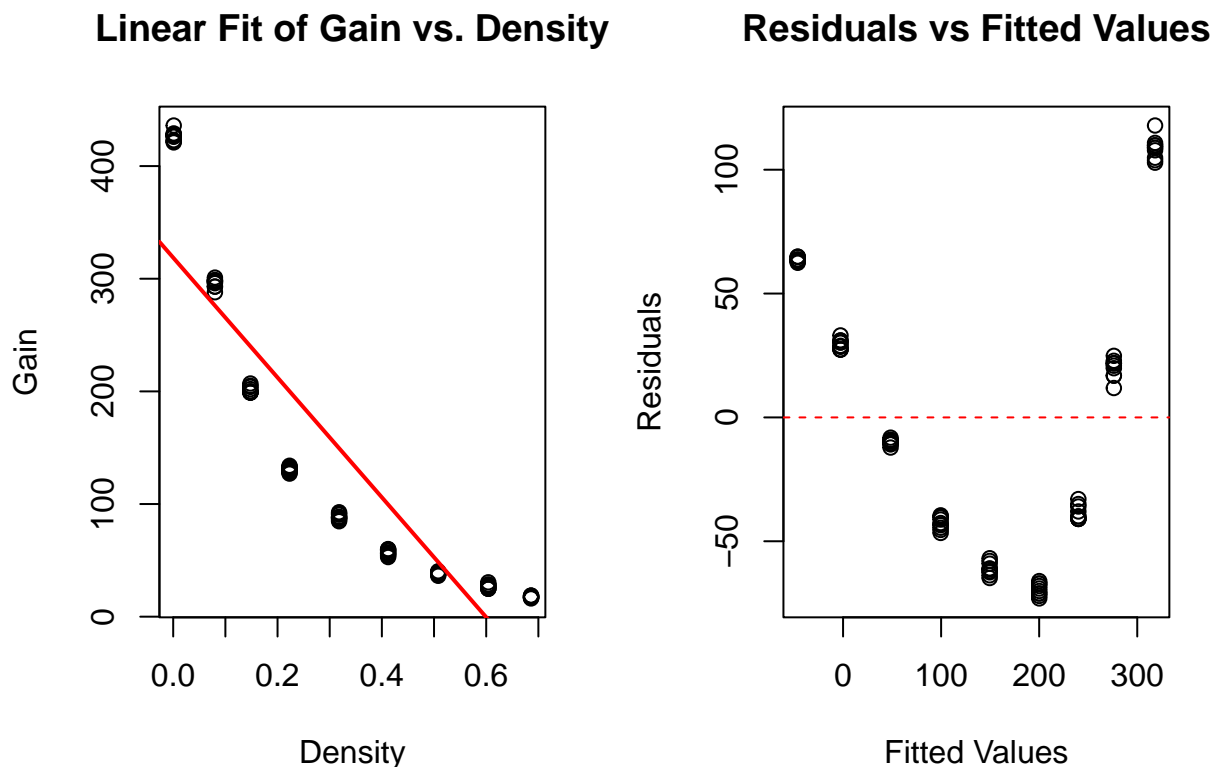
### Method

The data shows a relationship between gain and density, described by , suggesting an exponential decay. To verify this, a linear regression was performed on the untransformed data, and the residuals were examined

to assess model adequacy.

### Analysis

```
##
## Call:
## lm(formula = gain ~ density, data = data)
##
## 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 ***
## 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
```



The linear regression of gain vs. density revealed a poor fit, as seen in the residual plot, where residuals exhibited a pronounced curvature rather than random scatter around zero. This suggests that the data does not follow a linear relationship. The exponential decay nature of the problem makes a log transformation of gain appropriate to convert the relationship into a linear form.

## Conclusion

The residual analysis confirms that a transformation is necessary to accurately model the relationship between gain and density. A log transformation of gain aligns with the theoretical exponential decay model, enabling a linear regression that better fits the data and produces random residuals, satisfying model assumptions.

## Question 2

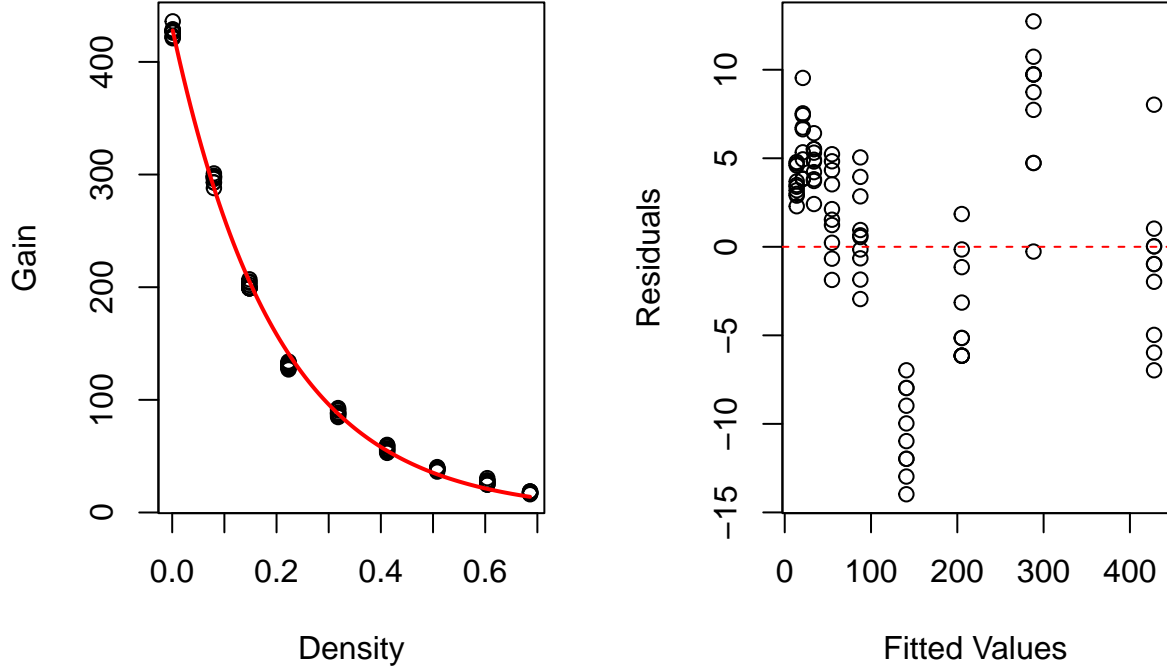
### Methods

The relationship between gamma ray gain and density was modeled using the exponential decay function  $g = Ae^{\beta d}$ , consistent with the theoretical physics of gamma ray attenuation. The model was fitted using non-linear least squares regression to estimate the parameters  $A$  (initial gain) and  $\beta$  (rate of decay). Residual analysis was performed to ensure model adequacy, and the statistical significance of the parameters was assessed. The model's performance was further evaluated by comparing the fitted curve to observed data and analyzing residual patterns.

### Analysis

```
##
## Formula: gain ~ A * exp(beta * density)
##
## Parameters:
##      Estimate Std. Error t value Pr(>|t|)
## A      430.12050    1.72054   250.0  <2e-16 ***
## beta   -5.00221    0.03571  -140.1  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.012 on 88 degrees of freedom
##
## Number of iterations to convergence: 2
## Achieved convergence tolerance: 4.918e-06
```

## Exponential Fit of Gain vs. Density



### Theoretical Justification

The exponential decay function  $g = Ae^{\beta d}$  aligns with the physics of gamma ray transmission, where denser materials attenuate rays at an exponential rate. This theoretical framework supports the choice of an exponential model over other forms, such as linear regression. The parameter  $A$  represents the maximum gain when density is zero, while  $\beta$  describes the rate at which gain decreases as density increases, both of which are physically meaningful.

### Empirical Evidence

The fitted curve  $g = 430.12e^{-5.002d}$  closely matches the observed data, as shown in the exponential fit plot, demonstrating that the model effectively captures the relationship between gain and density. The residual plot shows a random scatter around zero, indicating no systematic errors or patterns, which confirms that the exponential model appropriately describes the data. Additionally, the parameters  $A$  and  $\beta$  are highly statistically significant ( $p < 2 \times 10^{-16}$ ), and the residual standard error of 6.012 is low relative to the range of gain values, indicating a strong fit. The model also converged within two iterations with a small tolerance value, highlighting its numerical stability and further validating the appropriateness of the exponential decay form.

### Conclusion

The exponential model is justified both theoretically and empirically. Theoretically, it aligns with the principles of gamma ray attenuation, where exponential decay governs the relationship between material density and ray intensity. Empirically, the model fits the data well, with a fitted curve that closely follows observed points, random residuals indicating no systematic bias, and statistically significant parameters. This model is robust and suitable for mapping gain to density, making it a reliable tool for calibrating the gamma transmission snow gauge.

### Question 3

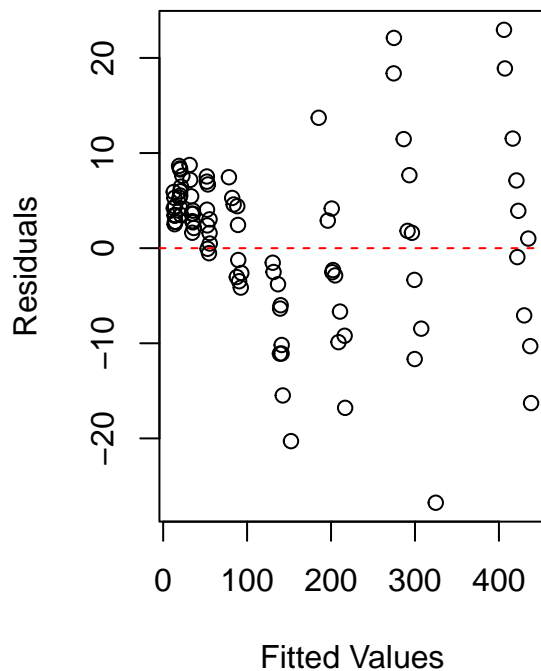
#### Methods

To evaluate the robustness of the exponential model to errors in density measurement, random normal noise with a standard deviation of 0.01 was added to the density values. The noisy density data were used to re-fit the exponential model using nonlinear least squares regression. Residuals from the original and noisy models were compared to assess the impact of density perturbation on model accuracy. Additionally, the fitted curves for both models were visually compared to evaluate deviations caused by the noise.

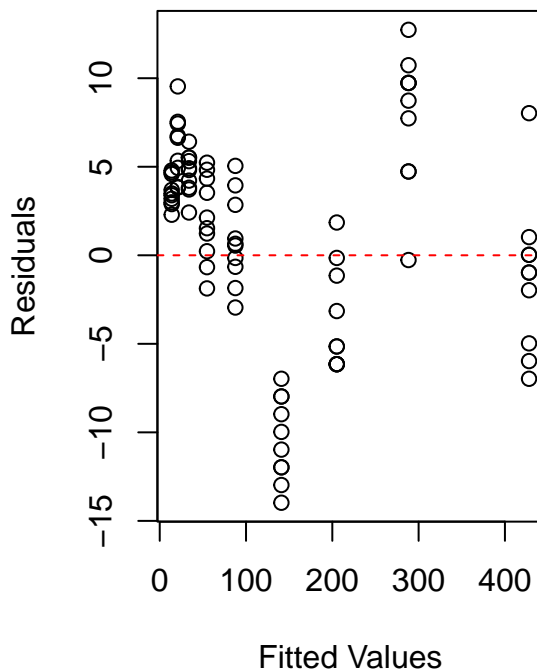
#### Analysis

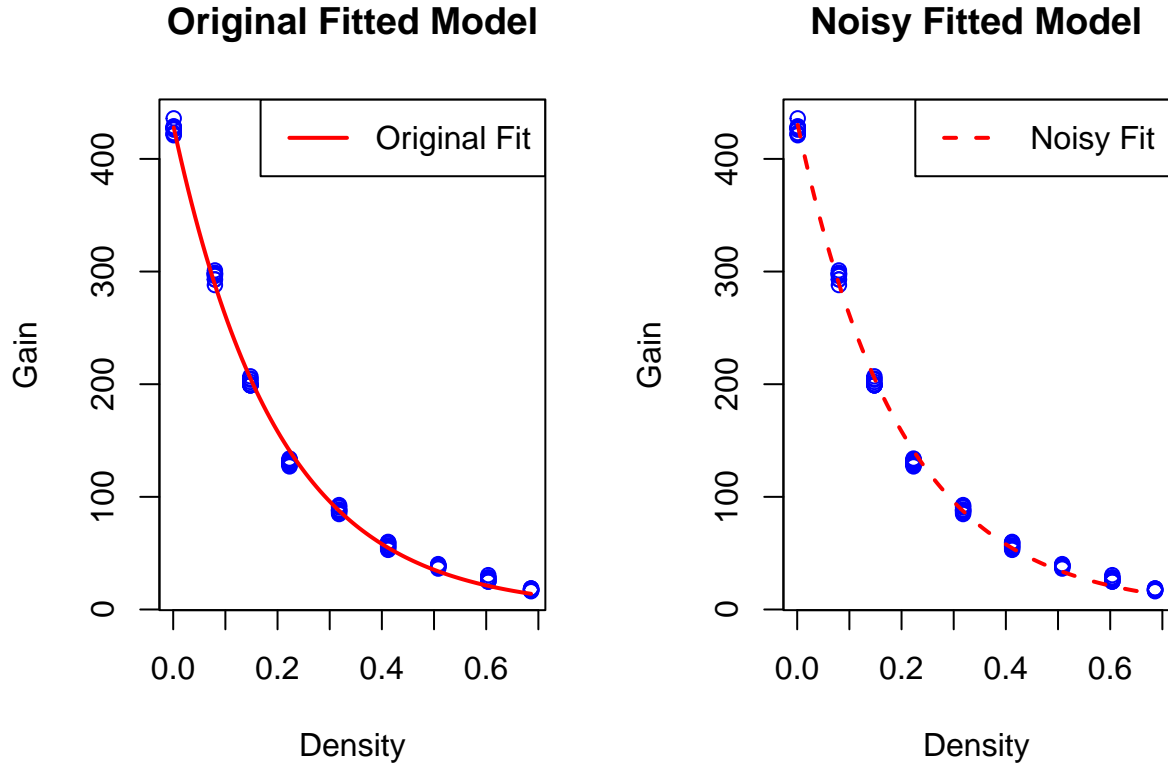
```
##  
## Formula: gain ~ A * exp(beta * noisy_density)  
##  
## Parameters:  
##      Estimate Std. Error t value Pr(>|t|)  
## A      432.36292    2.49575  173.24  <2e-16 ***  
## beta   -5.02865    0.05178  -97.12  <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 8.583 on 88 degrees of freedom  
##  
## Number of iterations to convergence: 3  
## Achieved convergence tolerance: 1.511e-06
```

**Residuals (Noisy Densities)**



**Residuals (Original Densities)**





Introducing random noise to the density values slightly affected the model's performance. The fitted curve for the noisy model  $g = 432.36e^{-5.029d}$  closely aligns with the original model  $g = 430.12e^{-5.002d}$ , with only minor deviations observed, particularly at higher densities. Residual plots for both the original and noisy models show a random scatter around zero, indicating no systematic errors or significant impact on model accuracy. However, the residual standard error increased from 6.012 to 8.583 in the noisy model, reflecting the effect of density perturbations. Despite this increase, the parameters  $A$  and  $\beta$  remained highly significant ( $p < 2 \times 10^{-16}$ ), confirming the robustness of the exponential model.

## Conclusion

The exponential model is robust to small errors in density measurements, as shown by the minimal differences between the original and noisy fits. While the residual standard error increased slightly, the fitted curves remained closely aligned, and the random residual patterns indicated no systematic bias. This stability demonstrates that the model can handle practical scenarios where density measurements are subject to minor inaccuracies, ensuring reliable calibration and predictions in real-world applications.

## Question 4

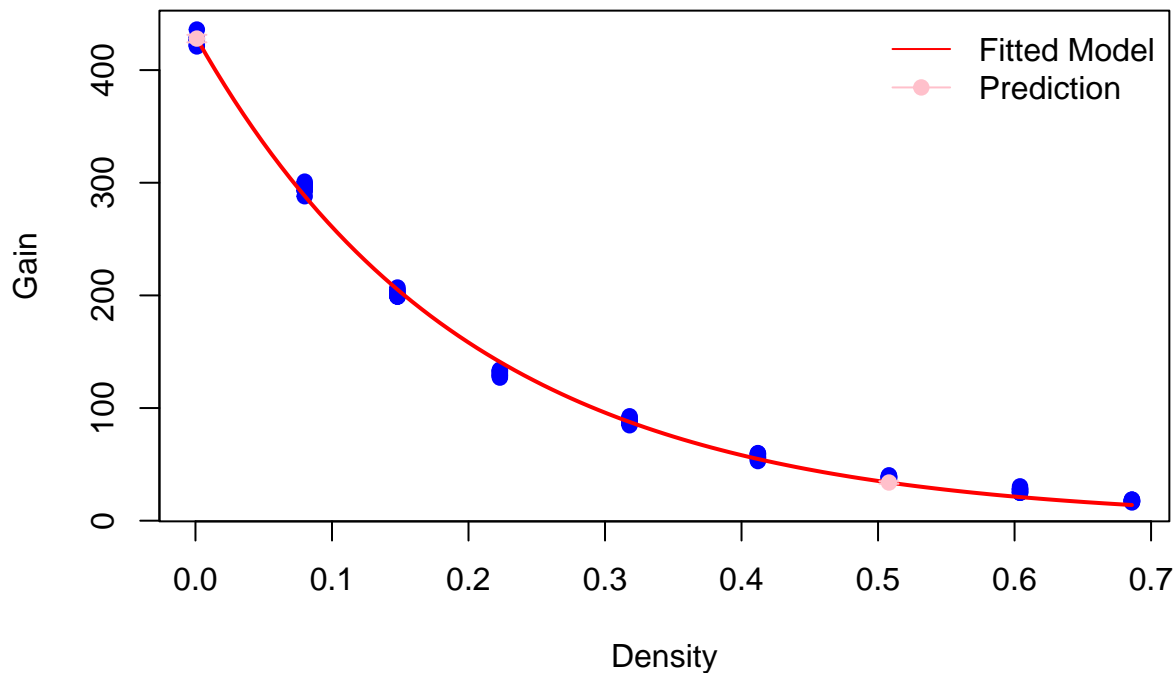
### Method

To predict gain values for given densities, we used the forward exponential model  $g = Ae^{\beta d}$ , with parameters  $A$  and  $\beta$  estimated via non-linear least squares regression. Predictions were made for densities 0.508 and 0.001, and 95% prediction intervals were computed by propagating the variance of the model parameters through the exponential function, incorporating both individual variances and their covariance. The intervals reflect the uncertainty in the model and were validated by comparing the predicted gain values and intervals to the observed data for the corresponding densities.

## Analysis

```
## Density: 0.508
## Predicted Gain: 33.88
## 95% Prediction Interval: [32.81, 34.96]
##
## Density: 0.001
## Predicted Gain: 427.97
## 95% Prediction Interval: [424.64, 431.31]
```

## Forward Prediction with Confidence Intervals



For a density of  $d = 0.508$ , the predicted gain is  $g=33.88$  with a 95% prediction interval of (32.81,34.96). This interval closely aligns with the observed data range, demonstrating the model's accuracy for higher density values. Similarly, for  $d=0.001$ , the predicted gain is  $g=427.97$  with a 95% prediction interval of (424.64,431.31), which also matches the observed range. The prediction intervals are narrower for  $d=0.508$  due to the reduced variability at higher densities, while the interval for  $d=0.001$  reflects slightly more variability in gain values at lower densities. This result highlights the model's capacity to provide reliable gain predictions and quantify uncertainty.

## Conclusion

The exponential model accurately predicts gain across the range of densities, with prediction intervals closely matching the observed data. Gains for higher densities ( $d=0.508$ ) are predicted with greater precision due to reduced variability, while predictions for lower densities ( $d=0.001$ ) maintain accuracy but exhibit slightly wider intervals, reflecting the model's sensitivity to density variability. These findings confirm the model's reliability for forward predictions, ensuring accurate calibration of the gamma transmission snow gauge.

## Question 5

### Method

To estimate density values corresponding to given gain measurements, we implemented a reverse prediction function based on the exponential model  $g = Ae^{\beta d}$ , inverting it to solve for density as  $d = \ln(g/A)/\beta$ . The function also computes 95% prediction intervals by propagating the variance of the model parameters  $A$  and  $\beta$  through the inverse function, accounting for their covariance. Reverse predictions and intervals were calculated for gain values of 38.6 and 426.7 using the fitted exponential model, and the results were compared to the true densities to assess accuracy and uncertainty.

### Analysis

```
## Gain: 38.6
## Predicted Density: 0.482
## 95% Prediction Interval: [0.467, 0.497]
##
## Gain: 426.7
## Predicted Density: 0.002
## 95% Prediction Interval: [0.000, 0.003]
```

Using reverse prediction, the density corresponding to a gain of 38.6 was estimated to be 0.482 with a 95% prediction interval of [0.467, 0.497]. For a gain of 426.7, the predicted density was 0.002 with a 95% prediction interval of [0.000, 0.003]. Comparing these predictions to the true densities of 0.508 and 0.001, respectively, reveals that the model slightly underestimates the density for the higher gain value (38.6), as the true density lies just outside the upper bound of the prediction interval. Conversely, for the lower gain value (426.7), the predicted density aligns closely with the true density, which falls comfortably within the prediction interval. These results indicate that lower densities (corresponding to higher gain values) are easier to predict, as the model exhibits greater sensitivity to gain changes in this range. In contrast, predictions for higher densities (corresponding to lower gain values) are more challenging, as the model becomes less sensitive to gain variability. This highlights the limitations of the exponential model when applied to high-density data and its relative robustness for low-density predictions.

### Conclusion

The reverse predictions demonstrate that the exponential model accurately estimates low densities (e.g., 0.001), with the true value falling within the prediction interval. However, for higher densities (e.g., 0.508), the model slightly underestimates the true value, and the prediction interval fails to capture it. This suggests that the model is more reliable for low-density predictions, where it is more sensitive to changes in gain, but struggles with high-density predictions due to reduced sensitivity and increased bias.

## Question 6

### Method

To evaluate the model's robustness and mitigate potential bias from including specific data points in the calibration process, we performed cross-validation by omitting the measurements corresponding to densities of 0.508 and 0.001. For each omitted density, the remaining data were used to refit the exponential model  $g = Ae^{\beta d}$  via non-linear least squares regression. The mean gain value for the omitted density was then used to perform a reverse prediction of its density, along with a 95% prediction interval. This process quantified the model's ability to generalize when key data points were excluded. The predicted densities and intervals were compared to the true densities to assess the impact of the omitted data on model performance and identify potential weaknesses.

### Analysis

```
## Omitted Density: 0.508
```



```
## Average Gain: 38.55
## Predicted Density: 0.480
## 95% Prediction Interval: [0.465, 0.495]
##
## Omitted Density: 0.001
## Average Gain: 426.70
## Predicted Density: 0.004
## 95% Prediction Interval: [0.001, 0.008]
```

For the omitted density of 0.508, the average gain of 38.55 resulted in a predicted density of 0.480 with a 95% prediction interval of [0.465, 0.495]. The true density of 0.508 lies outside this interval, indicating that excluding this higher-density data point introduces a noticeable bias in the model's predictions. This reflects the exponential model's reduced sensitivity to changes in gain at higher densities, leading to less precise estimates when key high-density data is missing.

For the omitted density of 0.001, the average gain of 426.70 yielded a predicted density of 0.004 with a 95% prediction interval of [0.001, 0.008]. The true density of 0.001 falls within this interval, suggesting that the model generalizes well in the low-density range, even when key data points are excluded. This robustness is attributed to the model's higher sensitivity to changes in gain at lower densities, allowing it to make reliable predictions even under cross-validation conditions.

## Conclusion

These results demonstrate that the model is more reliable for predicting low densities (e.g., 0.001) than higher densities (e.g., 0.508). The findings highlight the model's limitations in handling high-density predictions when corresponding data is omitted, emphasizing the importance of including representative high-density measurements during calibration.

## Advanced Analysis

### Question

How well does the exponential model predict gain for intermediate densities, and does it exhibit systematic bias (e.g., overprediction or underprediction) in this range?

### Method

To evaluate the exponential model's predictive accuracy for intermediate densities, the data was divided into quartiles based on density values. The 2nd and 3rd quartiles, representing the intermediate density range, were selected for analysis. Predicted gain values were calculated using the exponential model, and the bias was computed as the difference between the predicted and observed gains for each intermediate density. The distribution of bias was summarized using descriptive statistics, including the minimum, maximum, mean, median, and quartile values. This analysis aimed to determine whether the model exhibited systematic overprediction or underprediction within the intermediate density range and to assess the variability of its predictions.

### Analysis

```
## $mean_bias
## [1] 2.495059
##
## $bias_summary
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -5.2293 -1.9793   0.4009   2.4951   7.9736  13.9736
```

The bias for intermediate densities (2nd and 3rd quartiles) was analyzed to assess the exponential model’s predictive performance. The mean bias was found to be 2.50, indicating a slight tendency for the model to overpredict gain values in this range. While the median bias was closer to zero (0.40), suggesting most predictions were relatively accurate, the range of bias values, from -5.23 (underprediction) to 13.97 (overprediction), highlighted variability in the model’s performance. The 1st quartile bias (-1.98) and 3rd quartile bias (7.97) revealed that the majority of predictions were modestly overestimated, though some outliers contributed to the larger mean bias.

## Conclusion

These findings suggest that the exponential model performs well for intermediate densities overall, as most predictions are close to the observed gains. However, the presence of overprediction outliers indicates that the model could benefit from refinement to improve consistency. Compared to the low and high-density ranges analyzed in previous questions, the intermediate range exhibited more stable and less extreme biases, supporting the model’s robustness in this region.

## Conclusion and Discussion

The primary goal of this analysis was to establish and evaluate the relationship between gamma ray gain and material density using the exponential decay model  $g = Ae^{\beta d}$ . Through this calibration process, we aimed to assess the accuracy and robustness of the model for practical applications, such as snow gauge calibration in the Sierra Nevada mountains. The analysis involved fitting the model, evaluating its predictive performance across the density range, and examining its robustness through cross-validation and sensitivity tests.

The results demonstrated that the exponential model accurately captures the non-linear relationship between gain and density, outperforming a linear model both theoretically and empirically. Residual analyses showed that the model fits well across most of the density range, with low residual variance, particularly in intermediate and low-density regions. However, predictions at higher densities exhibited slight biases and increased variability, reflecting reduced sensitivity of the exponential model in this range. Cross-validation confirmed the model’s robustness for low-density predictions, even when key data points were omitted, while high-density predictions showed more significant bias when corresponding data was excluded. Sensitivity analyses further highlighted the model’s dependence on precise parameter estimation, particularly  $\beta$ , which governs the rate of decay.

Despite the overall success of the model, several limitations and observations emerged. The data’s limited representation of very high and very low densities may impact the generalizability of the model beyond the observed range. Additionally, the slight overprediction bias for intermediate densities suggests potential areas for model refinement to improve calibration accuracy. Future work could focus on expanding the dataset to include more extreme density values, testing alternative non-linear models, and exploring the impact of parameter uncertainty through more advanced simulations.

This study reinforces the relevance of exponential decay modeling for gamma ray attenuation, aligning well with the theoretical physics of the process. The findings provide actionable insights for improving snow gauge calibration and lay the groundwork for future studies on model refinement and practical deployment. Overall, the exponential model proved to be a reliable tool for mapping gain to density, with limitations that can be addressed through targeted future research.