Homework 4

2024-11-12

Problem Description

In the Sierra Nevada mountains of Northern California, monitoring water supply is crucial, and a gamma transmission snow gauge is used to measure snow density. However, this gauge measures density indirectly through gamma ray emissions.

The calibration process involves placing polyethylene blocks (which simulate snow) of known densities between the gauge poles. The gauge then measures gamma ray intensity, which we call the "gain." Multiple measurements are taken for each density to ensure accuracy.

The physics behind this process follows a predictable pattern. When gamma rays pass through material, denser substances allow fewer rays to reach the detector. This relationship isn't linear but follows an exponential decay, expressed mathematically as $g = Ae^{\beta d}$.

The fundamental goal is to establish a reliable relationship between these gamma measurements and the actual density of the material.

To be used in practice, the snow gauge needs to map the measured gamma ray intensity to snow density. However, the experiment is done in reverse. The calibration process goes as follows.

1. The experiment measures gamma ray intensity as a function of the density of the polyethylene blocks.
2. From the data, a function is determined that maps density to gamma ray intensity.
3. The inverse of the above function is used to map gamma ray intensity to density.

Dataset

The file gauge.txt contains 2 features density and gain. 10 measurements were taken for each of 9 densities (in grams per cubic centimeter of polyethylene).

Questions to be aswered for Analysis Section of the Report

1. **[Raw data]** Fit a regression line to the data and plot the fit. Examine the residual plot and explain why a transformation may be necessary.

2. **[Transformed data]** Determine an appropriate transformation and fit the model to the transformed data. Plot the new fit and examine the residuals. Justify your final model using both theoretical and empirical arguments.

3. **[Robustness]** Suppose the densities of the polyethylene blocks are not reported exactly. How might this affect the fit? Use a simulation to answer this question.

4. **[Forward prediction]** Produce point estimates and uncertainty bands for predicting the gain (in the original scale untransformed scale) as a function of the measured density. Can some gains be predicted more accurately than others? Consider specific prediction intervals for densities of 0.508 and 0.001 and compare these intervals to the range of measured gains for those densities.

5. **[Reverse prediction]** The average measured gains for the density values of 0.508 and 0.001 are 38.6 and 426.7, respectively. Invert the forward prediction line and uncertainty bands to produce point estimates and prediction intervals for the density that correspond to the gain measurements 38.6 and 426.7. How do the reverse predictions compare to the true density values? Are some densities harder to predict than other densities?

6. **[Cross-Validation]** The reverse prediction for density values 0.508 and 0.001 may be influenced by the fact that the measurement corresponding to the densities 0.508 and 0.001 were included in the fitting. To avoid this, omit the set of measurements corresponding to the block of density 0.508, apply your estimation/calibration procedure (forward fit and reverse prediction) to the remaining data, and provide an interval estimate for the density of a block with an average reading of 38.6. Where does the actual density fall in the interval? Try the same test, for the set of measurements at the 0.001 density.