**Questions**

**Please select any 3 data analysis algorithms that you learn from the class to develop the model**

The selected algorithms are:

1. Logistic regression
2. Naïve bayes
3. Support vector machine

**Please explain why you select such methods?**

1. Logistic Regression: Logistic regression is a popular statistical algorithm used to model binary outcomes, such as the presence or absence of scoliosis progression. It is easy to interpret and can handle both continuous and categorical predictors.
2. Naive Bayes: Naive Bayes is a probabilistic algorithm that can be used to predict the likelihood of an event based on the presence or absence of certain features. It is often used for text classification but can also be applied to other types of data. Naive Bayes assumes that the features are conditionally independent, which may not always be the case in real-world scenarios.
3. Support Vector Machine: Support Vector Machine (SVM) is a machine learning algorithm that is commonly used for classification tasks. It works by finding the hyperplane that separates the data points into different classes. SVM can handle both linear and non-linear data and is known for its ability to handle high-dimensional data.

**Indicate any relations between parameters**

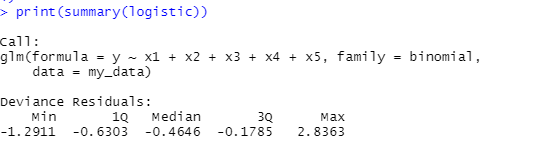
The 5 predictors directly determine whether one is diagnosed with the disease or not. Since they are all interrelated, it is reasonable to use a machine learning algorithm for this.

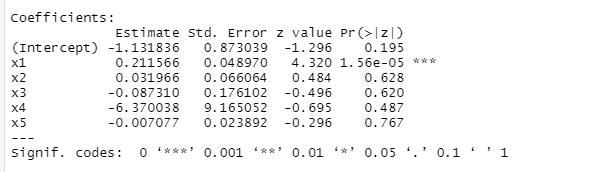
**Provide the result of your models**

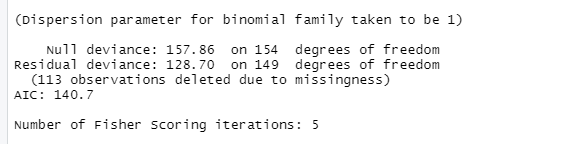
***Logistic regression***

The “Output (Progression (P)/Nonprogression (NP)” was re-written by assigning a value of 1 for entries with “P” and a value of 0 for entries with “NP”. A new column was created with these entries.

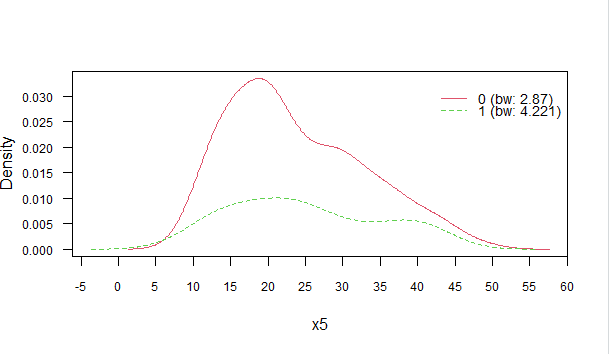
The logistic regression results are shown below:





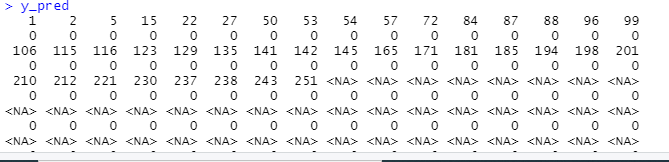


***Naïve Bayes Regression***

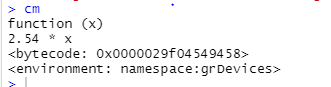


***Support vector machine***

The predicted values are shown below:



Confusion matrix:



**Which model is better?**

The logistic regression model.

**How accurate is your results?**

The results predicted by logistic were very accurate, with an AIC of 140.