

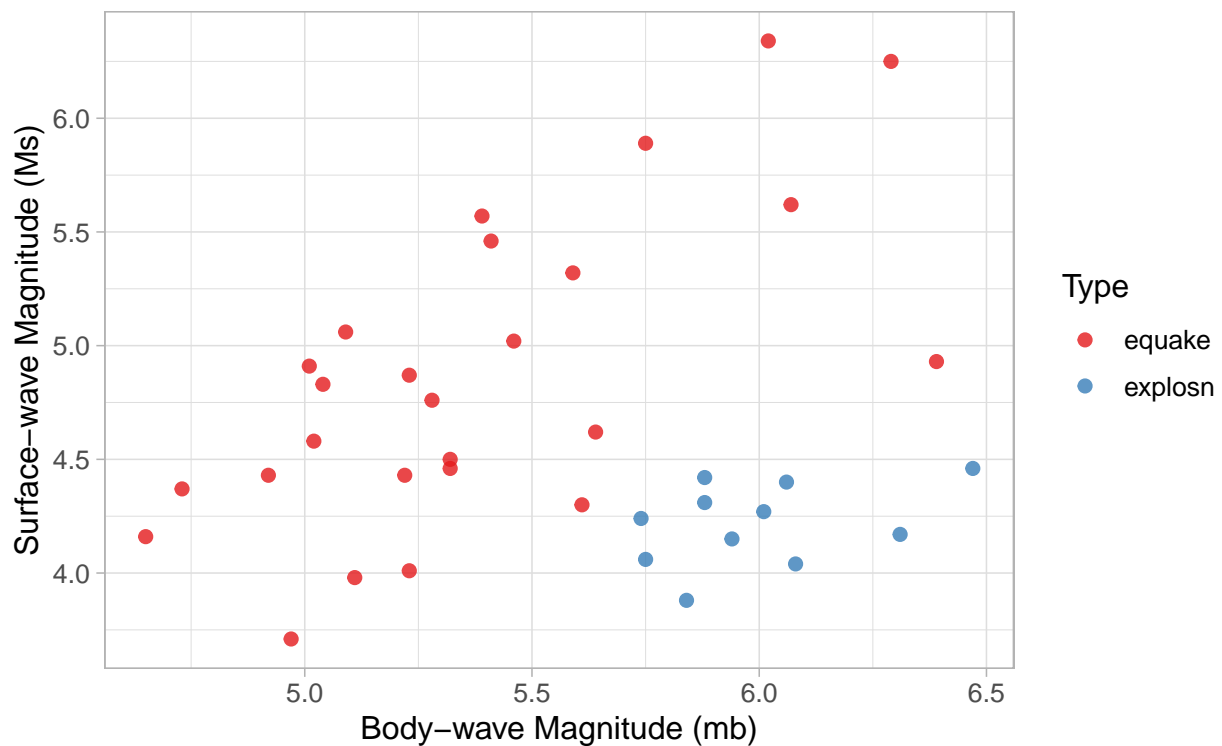
# MATH501 Coursework Report Submission

10883408

2024-04-08

## Machine Learning Part (a)

Body-wave Magnitude vs. Surface-wave Magnitude  
Comparing earthquake types



The result plot by this code would have each earthquake or explosion event as a point in a space defined by its body-wave and surface-wave magnitudes. The colours are deviating the seismic events and this may be very critical for recognition of the patterns or clusters which are earthquake-related rather than explosions.

### Clustering

In case there are evident clusters of distinct areas dominated by one type of event it may suggest that these two features are good for discrimination of earthquakes and explosions.

### Overlaps

If the colours are overlapping significantly then such two features on their own are insufficient to separate between event types without some additional information or more complex modelling.

## Contextual Relevance

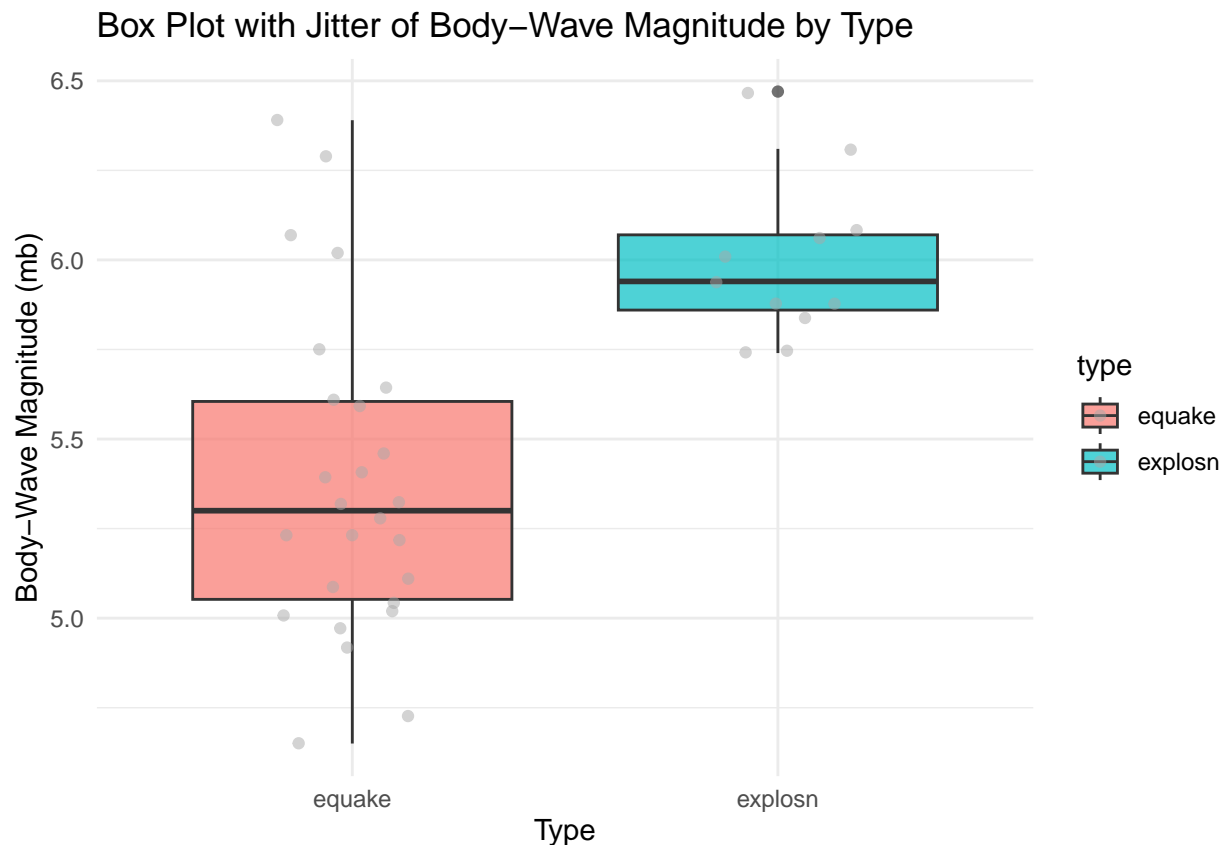
In the monitoring of unauthorized nuclear tests, this makes the visualization process a tool used in a rapid assessment. If there exist some clear and distinctive seismic readings patterns that may show nuclear activities. Efficient discrimination between natural seismic events (earthquakes) and man-made seismic events (nuclear explosions) is very important in the area of global security and control of compliance with the international treaties including the Comprehensive Nuclear-Test-Ban Treaty (CTBT).

## Numerical Summaries

Though the given code emphasizes on the visual analysis, numerical summaries (such as mean, median, variance, and histograms) are also needed for carrying out the data exploration process. Therefore, the MB and Ms summary (i.e. the summary of mb and Ms for each type) would complete this by defining the central tendencies and dispersion. This could also help in a statistical understanding of the magnitudes of each type significantly.

## Justification

The use of a scatter plot is supported since it enables stakeholders to see the relationship between two continuous variables across categories. For such high stakes in nuclear monitoring, instant visual, and vulnerability were of paramount concern assessment coupled with thorough statistical analysis is mandatory. This is made easy by the plot which gives a brief, instant visual representation of the data in mentioned characteristics.



## Explanation of the Plot

The jitter with box plot allows to see how body wave magnitudes are distributed within each one type of seismic event. Furthermore, the box plot component represents the median (the middle line in the box),

the 25th and 75th percentile whose positions determine the hinges of the box, and potential outliers (points that fall further than 1.5 times IQR from the hinges). Points jittered represent individual data points and provide a very fine level analyse the distribution of data and the possible anomalies or outliers.

## **Justification of the Statements**

### **Distribution Insight**

The plot helps in rapid detection of any significant differences in the magnitudes of the body-waves among various types of seismic events. For instance, if a certain type of event generally provides higher magnitudes reading, this implies a different energy release feature.

### **Outlier Detection**

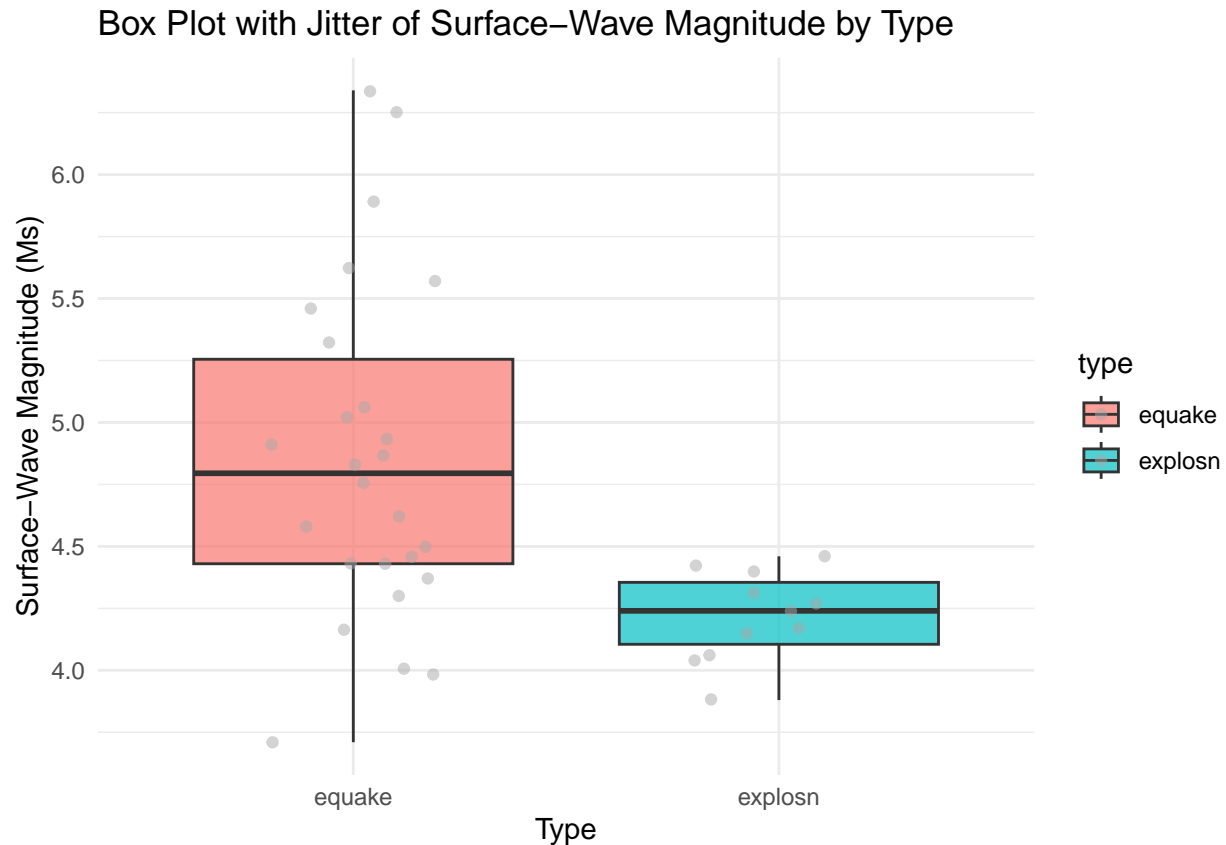
Through also showing the summary statistics and the actual data points, this plot assists in identifying the outliers or unusual observations that may need more attention.

### **Decision Making**

This kind of visualization helps in decision support in seismology and geophysics by giving an easy way of comparison of seismic event types. This is likely to be critical in developing monitoring systems or academic research in seismology.

### **Effective Communication**

The plot works as a powerful means of communication in both displaying complex statistical data in a way readable to everyone even those who do not have much of statistical knowledge. This cross between box plot and jitter plot works especially in situations where the variability both within and across categories is to be appreciated. It is an important fact finding tool that offers both an overall look and a more detailed show of how the data is distributed over categories.



## Justification of the Statements

### Visualization of Variability

This plot is good for the visual assessment of the following variability and the central tendencies of magnitudes of surface waves among various classes of seismic events. The box plot gives an overview but the jittered points give a comprehensive outlook on individual data characteristics entries.

### Comparative Analysis

Since the plot contains information of different types together, it enables direct comparisons of groups. For instance, one can say that nuclear detonations show a narrower range of magnitude of surface-wave in comparison with an earthquake, which might put a wider range and sometimes higher medians.

### Outlier Detection

The graphical depiction allows to notice any anomalies or the outliers in the data, which may imply measuring errors or deviant situations that should be looked into more closely.

### Informative and Accessible

The plot finally turns into a friendly plot with the help of a simple title, axis labels, and a legend that makes the conclusions understandable in a single glance also to laymen.

### Contextual Relevance

This capability is very critical in the context of monitoring of seismic activities, where the ability to differentiate inter-classification of events among surface-wave magnitudes is paramount. Apart from the introductory

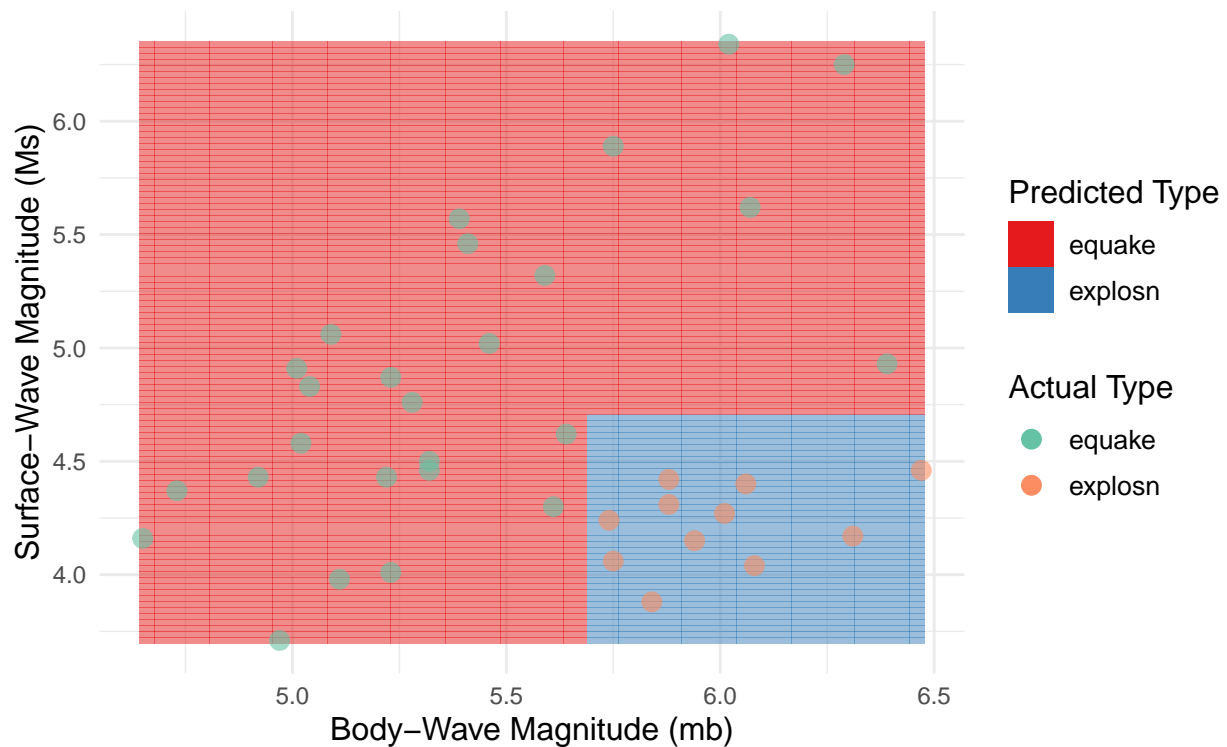
analysis, such schemes may help generate more advanced models and algorithms fitting for the automation of the process of seismic event recognition and categorization. For example, this is important for situations when quick decisions are needed, such as in early warning systems and monitoring of nuclear treaty compliance.

In conclusion, the graphical plot generated by this R code is graphically pleasing while representing important statistical information in the seismic data analysis, making statistical analysis details and overall data view ready. This approach is supported by its application in exploratory data analysis, when the awareness of data distribution is very significant and anomalies are of great value.

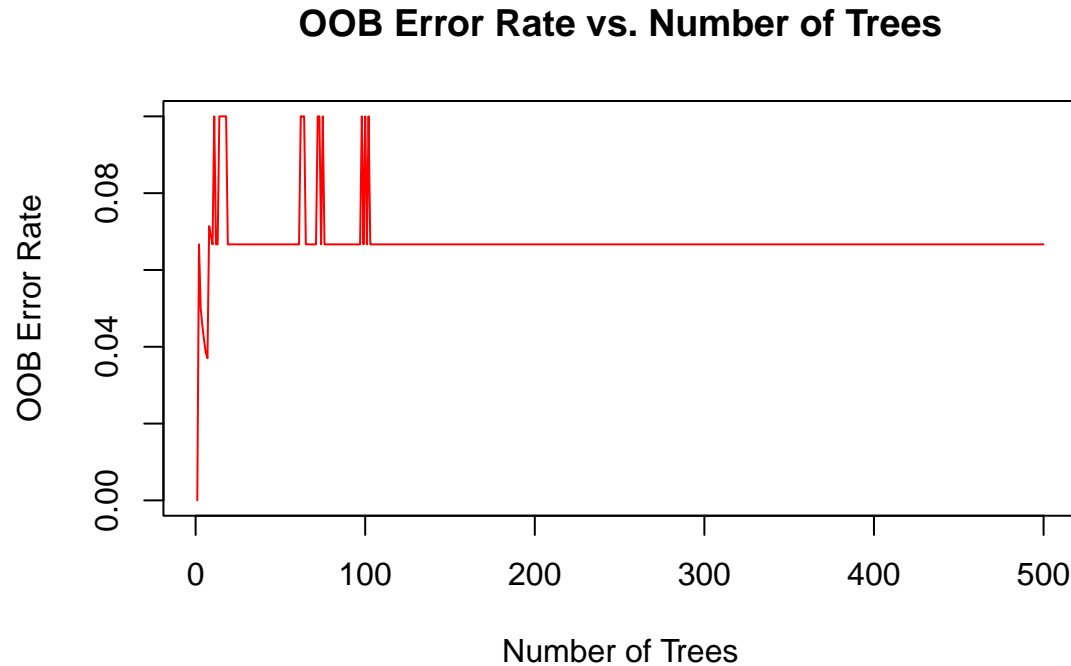
## Machine Learning Part (b)

### Earthquake vs. Nuclear Explosion Prediction

Random Forest Model Predictions vs. Actual Data



error rate



## Model Evaluation

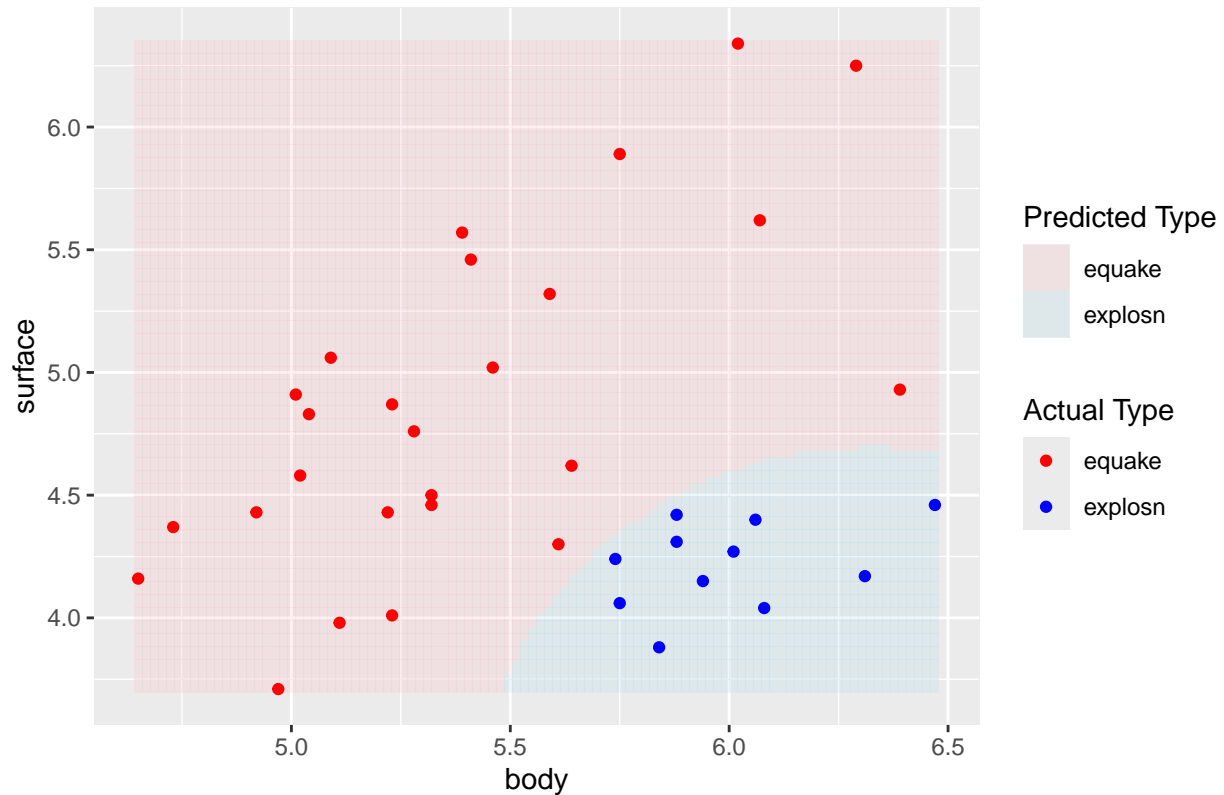
**Error Rate Computation: Out-of-Bag (OOB) Error** The OOB error rate is plotted versus the number of trees in the forest. Out-of-bag error is a measure used to assess the performance of the bootstrapped samples employed in the training procedure of random forests, decision trees, and other models that utilize bootstrap aggregating to subsample the data. The OOB error is the mean mistake in prediction of all train sample by using just the trees that had such  $x$  in their bootstrap sample.

**Plot Details:** The plot shows how the performance of the model changes with the growth of the number of trees in the forest. The correct behavior is that as the number of trees large, the OOB error rate should decrease and reach a plateau, indicating that the model is neither overfitting nor underfitting.

**Additional Evaluation with Leave-One-Out Cross-Validation (LOOCV)** Even though it is not explicitly presented in the provided code snippets, leave-one-out cross-validation (LOOCV) could be one more way to assess the model. In LOOCV, the model is fitted to all data points except one, which is used as the test set. This is done so that each data point is the test set once. LOOCV gives a strong approximation of the model's performance, but it is quite computationally expensive especially for larger datasets. In this case, verification of the model's efficiency would be very beneficial because of the small dataset.

**Summary** The integrated approach to model tuning, visualization, and evaluation methodology allows for a comprehensive understanding and validation of the Random Forest model. This methodical pursuit aims to approximate the model to be both accurate and generalizable, effectively discriminating seismic event classes from each other with sufficient power. The visualization of the decision boundary provides an intuitive method of understanding model performance, whereas the error rate plot and potential LOOCV offer numerical measures of model accuracy.

## SVM Classification of Earthquake and Nuclear Explosions



```
## sigma C
## 8 0.1 100
```

### Justification

SVM works well in high-dimensional spaces and is perfect for binary classification tasks such as differentiating nuclear blasts and earthquakes. It performs good when there is a distinct margin of separation between classes.

### Model Tuning

**C Parameter:** The **C** parameter can be tuned to balance the trade-off between creating smooth decision boundaries and achieving correct classification of training points.

**Kernel Choice:** The model's performance can be approached from another angle by selecting different kernels, such as linear, polynomial, and radial basis function (RBF), to better capture the complexities in the data.

### Model Visualization

The SVM decision boundary between body and surface-wave magnitudes is plotted in a 2D space to illustrate the classification rules. This visualization helps in understanding how SVM categorizes different seismic events.

## Machine Learning Part (c)

### Random Forest Classification of Earthquake and Nuclear Explosions

#### Pros

- Non-linear data is well handled by a random forest because it is an ensemble of decision trees, thereby making it more capable of handling complexity in the dataset.
- Random Forest is not sensitive to outliers and noise, as it uses averaging to enhance prediction accuracy.
- The plot of the Out-Of-Bag (OOB) error rate indicates that the model stabilizes rapidly, and there is no overfitting as the number of trees increases, which can be observed by the almost constant OOB error rate after about 50 trees.

#### Cons

- Despite OOB error rate being comparatively low, it is not obvious what quantity of false positives or negatives has been produced without a confusion matrix or similar metrics.
- The Random Forest can be quite computationally costly with a high number of trees, and it can take longer than other models to train, although this is not an issue here due to the quite stable OOB error rate.
- The issue of interpretability may arise simply because the Random Forest models are usually more complex and hard to interpret than simpler models.

#### Performance

- The plot reveals a red area of earthquake predictions and a blue area of explosion predictions. It appears to discriminate well, however, some earthquake points are falling in the explosion forecast area.

### SVM Classification of Earthquake and Nuclear Explosions

#### Pros

- Effective in high-dimensional spaces, which may be useful if more features were employed in the classification.
- The effectiveness of SVMs comes into play when there is a well-defined margin of separation, and SVMs are adaptable to different kernel functions.

#### Cons

- Tuning of parameters such as the penalty parameter (C) and the kernel-specific parameters requires special attention; any misconfiguration can cause loss of performance.
- May fail when target classes are close to each other in a dataset, i.e., more noise.
- The model can also be less interpretable because the kernel trick adds complexity.

#### Performance

- On the SVM plot, the earthquakes and explosions are distinctly separated, with the earthquakes in general having a higher surface-wave magnitude. This partition shows how the SVM with a well-chosen kernel is able to grasp the boundary between classes very well.

### Comparison and Recommendation

In this kind of comparison of both classifiers, SVM appears to have a clearer boundary between the earthquake and explosion classes since there is a region specifically for each class. This implies that the SVM has effectively captured the inherent patterns in the data that separates the two phenomena. At the same time, the Random Forest plot reveals some common space among the projected areas, which might indicate either



the problem of misclassification or the more complex border that might not have been fully modeled by this model.

Given these observations:

- In the case when the decision boundary between the two classes is really complex and non-linear, Random Forest may have a benefit to capture this complexity.
- When the most significant aspect is computational effectiveness, during the training and prediction process, SVM is usually recommended, especially if the kernel function chosen and its parameters are efficient.

## Machine Learning Part (d)

### K-means Clustering Plots

Each plot shows the clusters the k-means algorithm formed when the number of clusters is different (2, 3, and 4).

The **body** variable is plotted on the x-axis which indicates the Body-Wave Magnitude (mb) and the **surface** variable is plotted on the y-axis which indicates the Surface-Wave Magnitude (Ms).

Various colours signify different clusters, and cluster centroids are denoted by black asterisks.

### Elbow Method Plot

The third graph is used to identify the best number of clusters to use by plotting within group sum of squares (WCSS) for different numbers of clusters.

The x-axis is the number of clusters, and the y-axis is the WCSS.

## Results

### K-means Clustering Plots

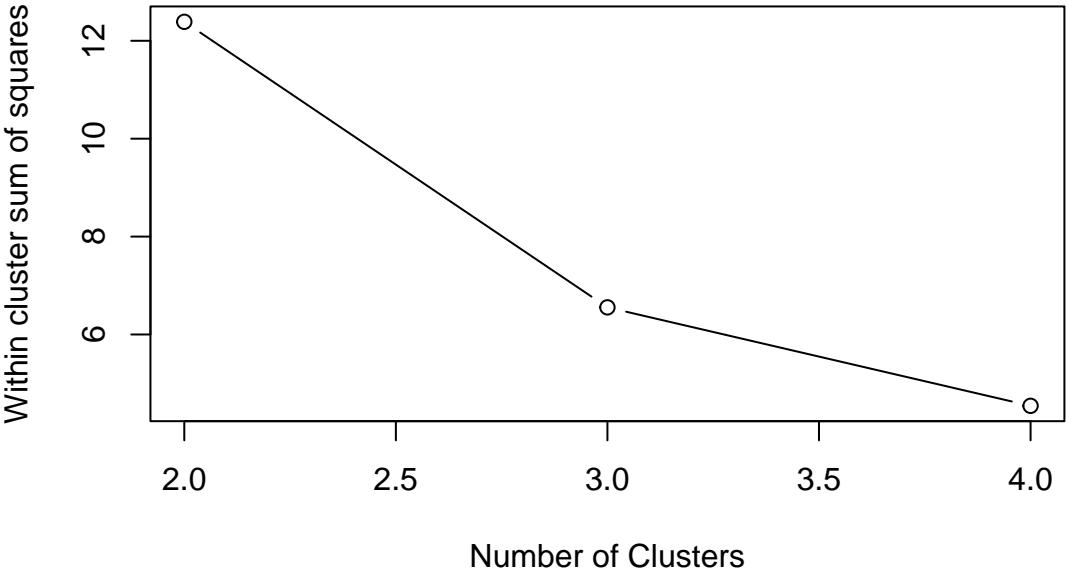
- For two clusters, the separation is quite straightforward with one cluster located at the lower part of the plot and the other at the upper part.
- The appearance of a new cluster between the previous two clusters with three clusters, suggests a finer discrimination of the earthquake magnitudes.
- With four clusters, become more detailed, which implies that some values are adequately detailed to form subgroups.

**Elbow Method Plot** The elbow plot shows a hard bend at 3 which normally signifies that any additional clusters beyond this point do not significantly reduce the WCSS.

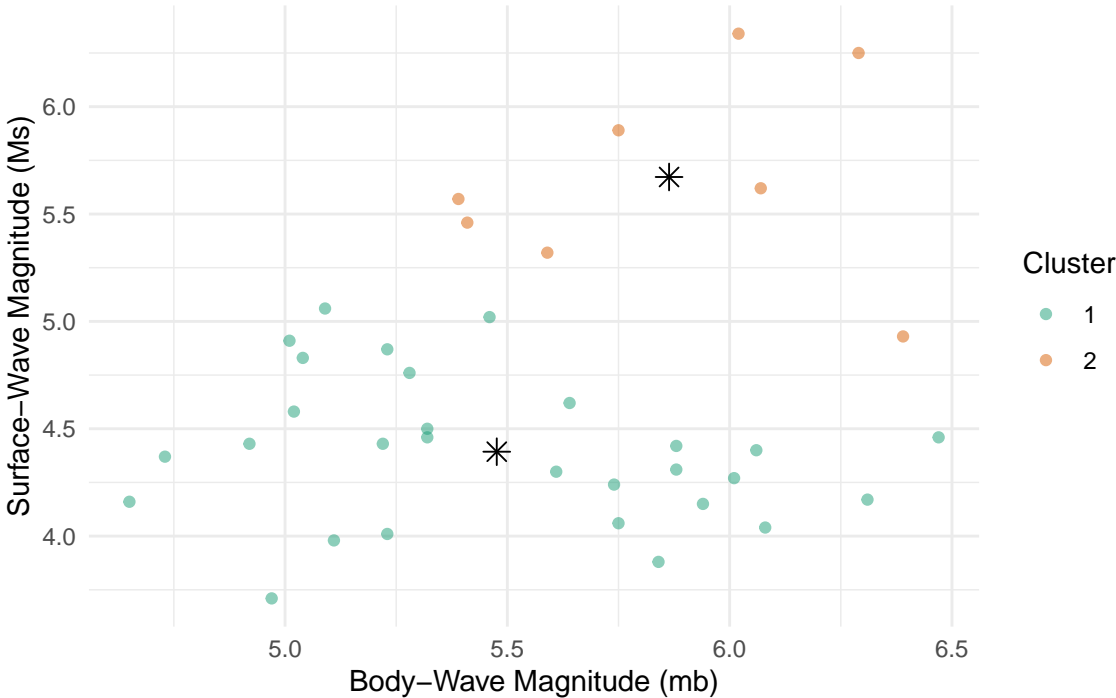
## Conclusions

- The k-means clustering plots show that with more number of clusters, the segmentation of data becomes finer, which may or may not be meaningful in the context of data and domain knowledge.
- The elbow method plot implies that the dataset may have an optimal of three clusters since this seems to be the point at which the rate of decrease in WCSS decreases to a much slower rate.
- However, two clusters are usually too generic and fail to capture the subtleties in the data, though more than three clusters will be overfit and segment the magnitudes too finely to be practically interpretable.
- In general, based on both the elbow method and the point distribution in the clustering plots, three clusters rather than more or less could be the best number for representing the underlying patterns in the data without making the model too complex.

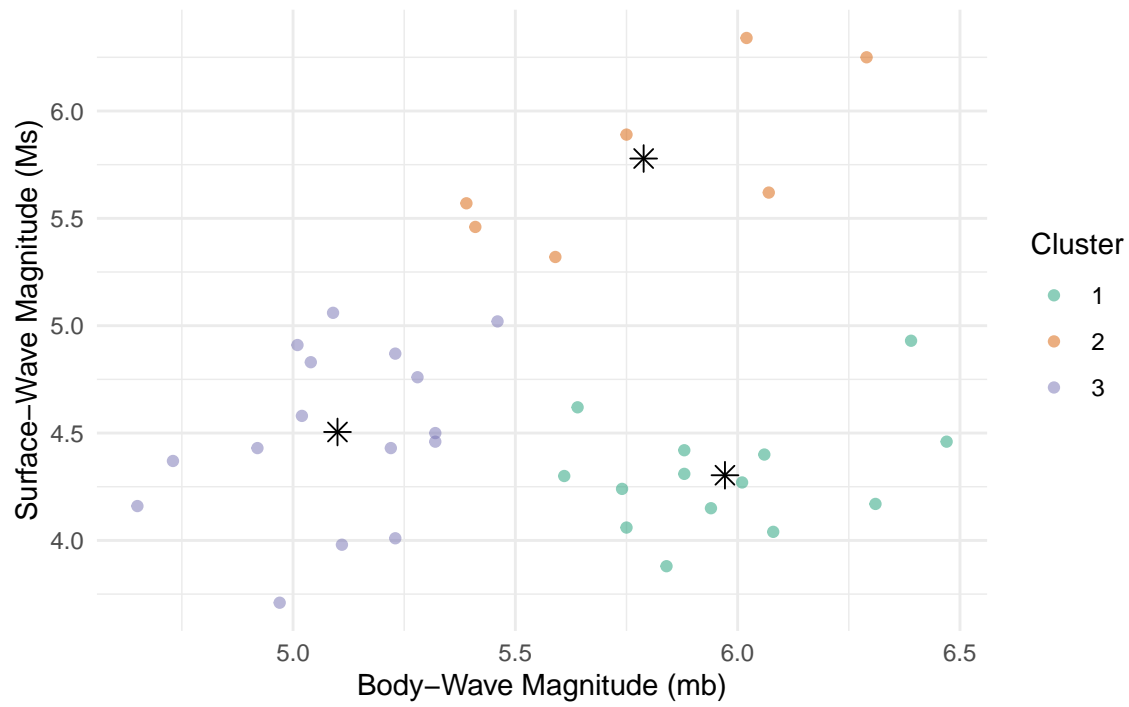
**Elbow Method for Determining Optimal Number of Clusters**



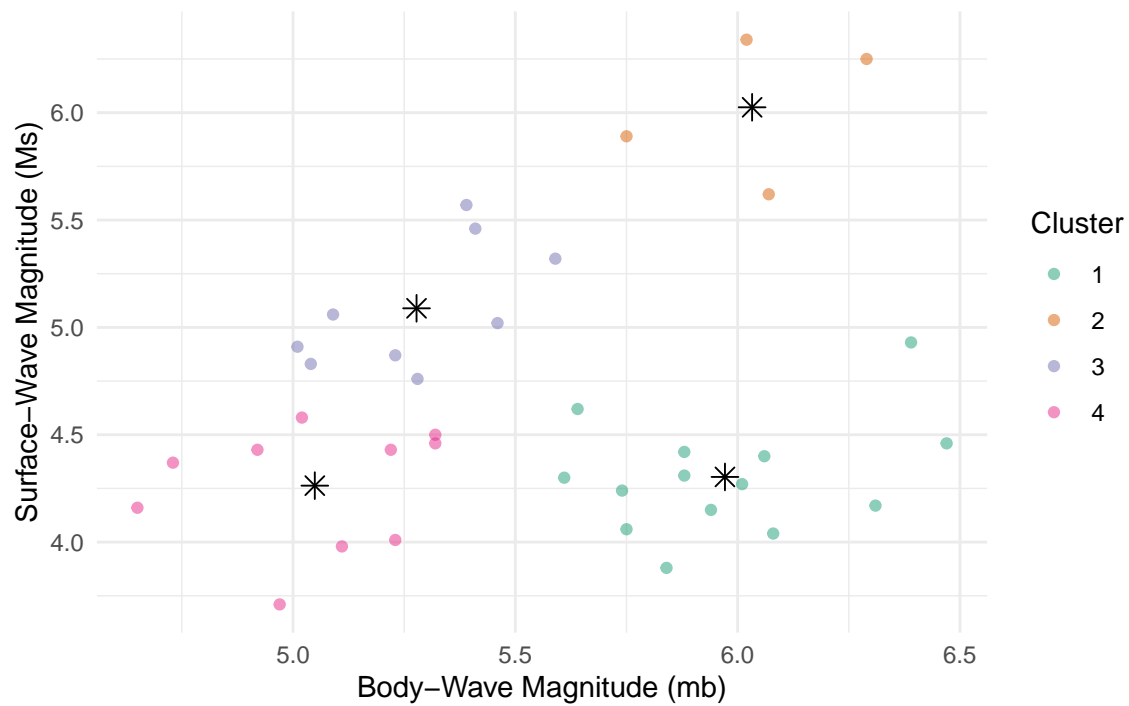
**K-means Clustering with 2 Clusters**



K-means Clustering with 3 Clusters

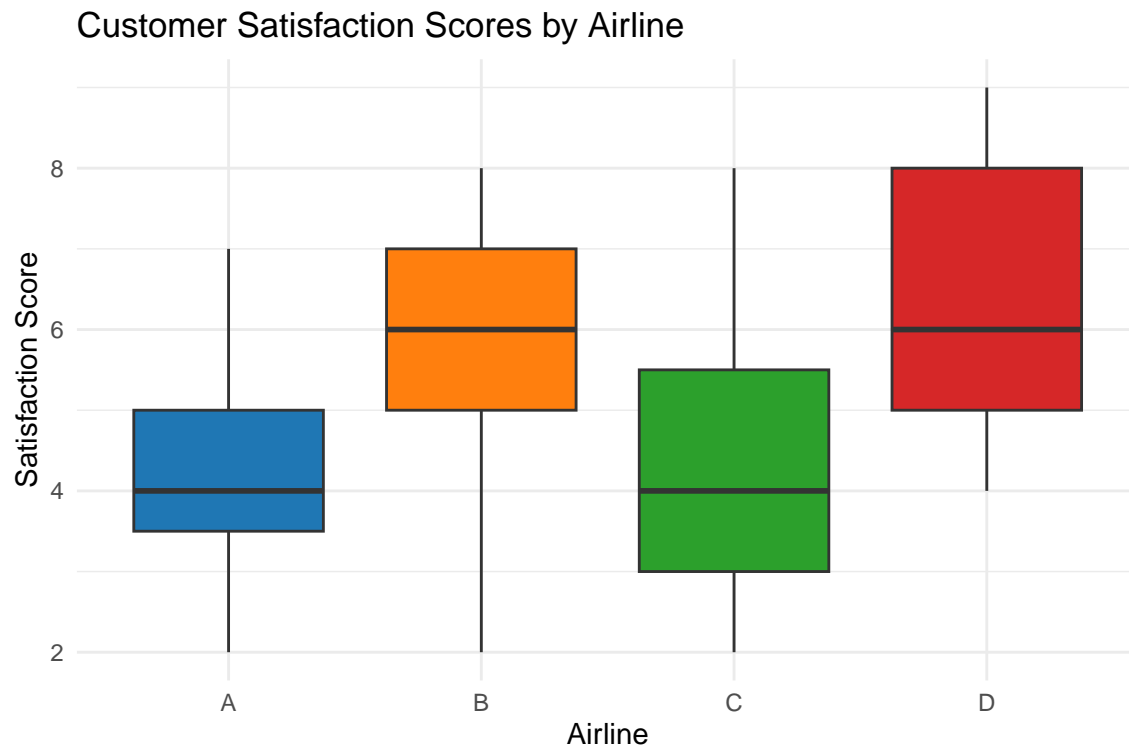


K-means Clustering with 4 Clusters



# Bayesian Statistics Task - Customer Satisfaction Scores by Airline

## Bayesian Statistics Part (a)



### Customer Satisfaction Scores by Airline

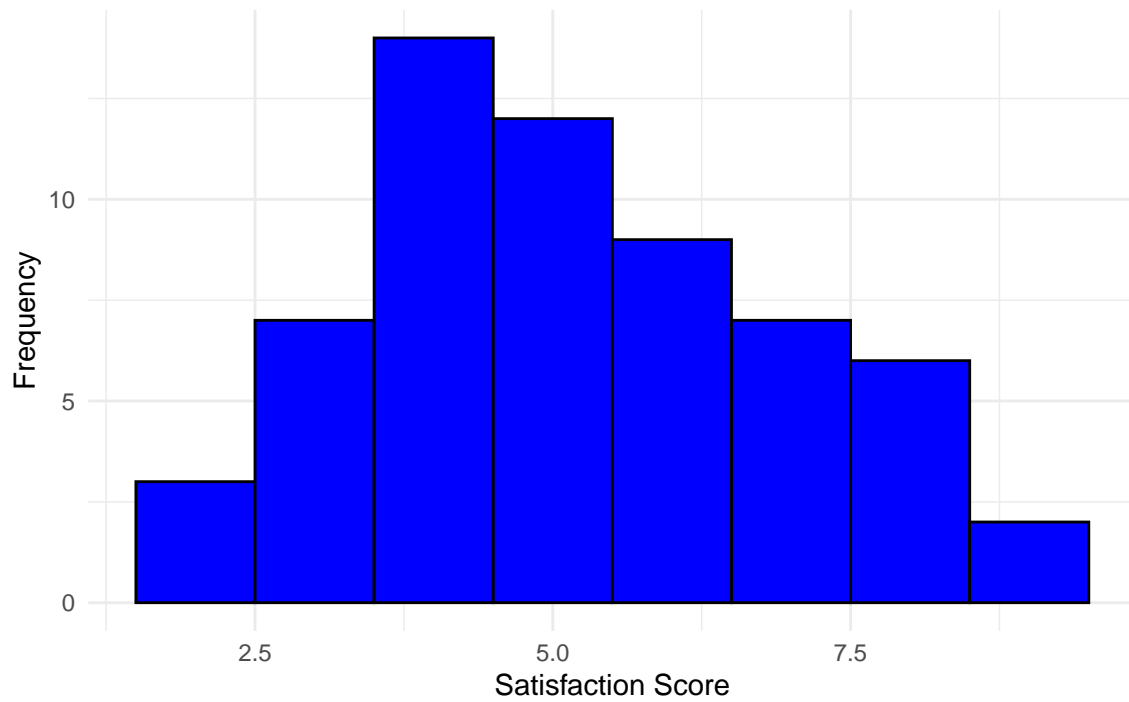
**Central Tendency** :Within each box, the median line represents the central tendency of satisfaction scores of each airline. A by-line comparison can be used to find out which airlines have the highest or lowest median satisfaction scores.

**Spread and Variability** :Each box's height shows the Interquartile Range (IQR), which is a measure of the mid-50% spread. A lesser box height means that the satisfaction scores are clustered evenly around the median, indicating more homogeneous service quality. Bigger boxes indicate more variability, which suggests no consistent passenger experience.

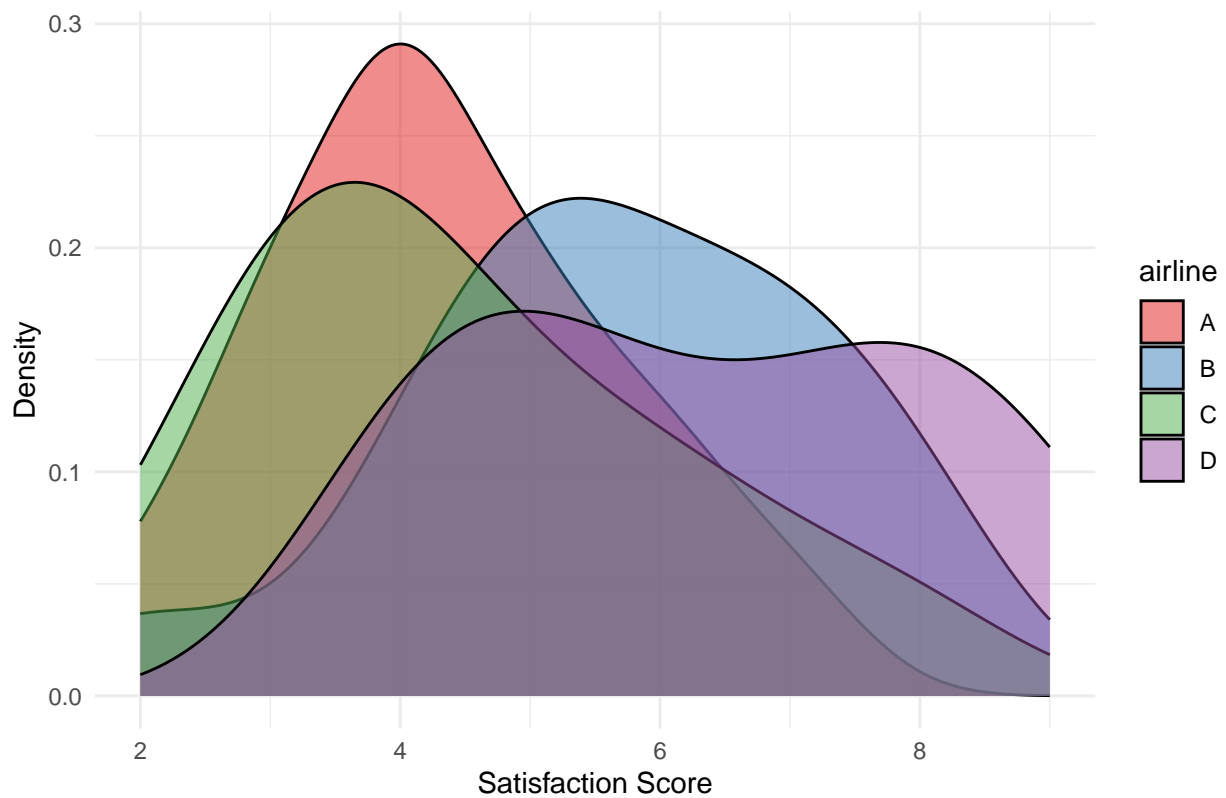
**Outliers** :The outliers of the boxplot are represented by the dot-like points outside the main box of the boxplot. These are anomalies that are way higher or lower than the other parts of the data. Outliers may also represent very good or very bad experiences that are atypical of the average customer.

**Comparison Across Airlines** :If the median of one airline is much higher than those of others, it reflects that this airline generally provides superior customer service. On the negative end, the presence of outliers is a challenging issue for the airlines as it clearly shows that some of the customers are more dissatisfied.

Distribution of Satisfaction Scores



Density Plot of Satisfaction Scores by Airline



**Shape of Distribution:** The shape of airline satisfaction density curve gives the idea of how the satisfaction

ratings are distributed. For example, normal distribution, skewed distribution, and bimodal distribution, each of represents different hidden patterns of customer satisfaction.

**Peak Values:** The apexes of the curves represent the modal satisfaction scores of each airline. The peak at a higher satisfaction score suggests that most of the airline's customers are highly satisfied.

**Spread and Variability:** That heterogeneity of satisfaction scores is evident in the spread of the curve. A broader curve means that the customers of the airline are more variable in their close ratings, while a narrower curve implies that the ratings are more consistent among customers.

**Overlap Between Airlines:** Overlapping areas of density curve among many airlines indicate that there are shared scores for more than one airline. The smaller overlap is an one-of-a-kind perception of satisfaction of this airline.

**Tail Analysis:** The tails of the density curves enable us to see how often extreme scores (both high and low) occur; longer tails to the left would imply many very dissatisfied customers, while longer tails to the right of the curve can represent a large number of very satisfied customers.

## Bayesian Statistics Part (b)

### Consistency Across Airline 1

Airline 1 mean satisfaction score,  $\mu_{1j}$ , is a constant at  $\mu_1$  for every customer  $j$ . This implies that all passengers of Airline 1 have the same level of satisfaction, that is, the expected satisfaction score is the same for this airline's customers.

### Airline 4 Adjusted Mean

However, the mean satisfaction score for each customer of Airline 4,  $\mu_{4j}$ , is given by  $\mu_1 + \alpha_4$ . Here,  $\alpha_4$  represents the shift factor of the baseline mean satisfaction score set by Airline 1.

### Interpretation of $\alpha_4$

- **Positive  $\alpha_4$ :** Implies that even though, in general, Airline 1 is perceived to be a better airline in terms of service than Airline 4, the average rating of Air 4 by the customers is higher.
- **Negative  $\alpha_4$ :** Suggests that Airline 4 is generally a lower scorer in satisfaction as compared to Airline1.
- **Zero  $\alpha_4$ :** Implies that the mean satisfaction scores are not that different between the two airlines.

## Bayesian Perspective

In a Bayesian analysis setting,  $\alpha_4$  is not estimated as a single constant value but as a distribution of all possible values. It complies with the probabilistic nature of statistical estimation and provides a range of possible  $\alpha_4$  values. However, this approach not only reflects uncertainty with respect to the estimate but also gives a better-off perception about the variability of customer satisfaction among these airlines.

By regarding  $\alpha_4$  as a distribution, we create another dimension through which to evaluate how Airline 4 performs in comparison with Airline 1, thus enhancing decision-making with a more rounded, data-driven approach.

## Bayesian Statistics Part (c)

```
## Baseline for Airline A: 4.333333
## Difference for Airline B from A: 1.333333
## Difference for Airline C from A: 0.1333333
## Difference for Airline D from A: 2
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## airline      3  41.87   13.956    5.29 0.00278 **
## Residuals   56 147.73    2.638
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## There is a statistically significant difference in satisfaction scores
## across airlines at the 0.05 significance level.
```

The one-way ANOVA analysis has provided with the following estimates for the mean satisfaction scores of Airline A and the differences relative to it for Airlines B, C, and D:

- **Baseline Mean Satisfaction for Airline A ( $\hat{\mu}_1$ ):** 4.333333
- **Difference in Satisfaction for Airline B from Airline A ( $\hat{\alpha}_2$ ):** 1.333333
- **Difference in Satisfaction for Airline C from Airline A ( $\hat{\alpha}_3$ ):** 0.133333
- **Difference in Satisfaction for Airline D from Airline A ( $\hat{\alpha}_4$ ):** 2

## Model Fitting

The `aov()` function is used to fit a one-way ANOVA model to determine if the satisfaction scores mean significantly differ among the different airlines. This statistical method does not consider any difference among airlines in respect of customer satisfaction.

## Coefficient Extraction

Coefficients are extracted from a model to produce estimates on the mean satisfaction score of the baseline airline (Airline A, or represented as  $\hat{\mu}_1$ ) and the deltas ( $\hat{\alpha}$ ) for the other airlines (B, C, and D) with respect to Airline A.

## Hypothesis Testing

The `summary(Anova model)` output includes the F-statistic with the respective p-value. This F-test tests whether at least one group mean is different from the others.

## Interpretation of Results

- **If the p-value is less than 0.05:**  
Rejecting the null hypothesis that all group means are equal, conclude that satisfaction scores across the airlines are significantly different. This shows that not all airlines have the same levels of customer satisfaction.
- **If the p-value is greater than or equal to 0.05:**  
Fail to reject the null hypothesis, thereby, conclude that there is no significant difference in the satisfaction scores among the airlines.

## Conclusion

### Significant Difference Found (p-value < 0.05)

The analysis reveals a considerable difference in the satisfaction scores among the air carriers and should not imply that all perform equally.

## Justification

The ANOVA test provides a p-value of less than the significance level of 0.05, and the null hypothesis that all airlines have the same mean satisfaction score is rejected. According to the data, some airlines could be offering better or worse service than others. This result is extremely important in determining which airlines are underperforming or overperforming in customer satisfaction. A detailed study of this kind is a

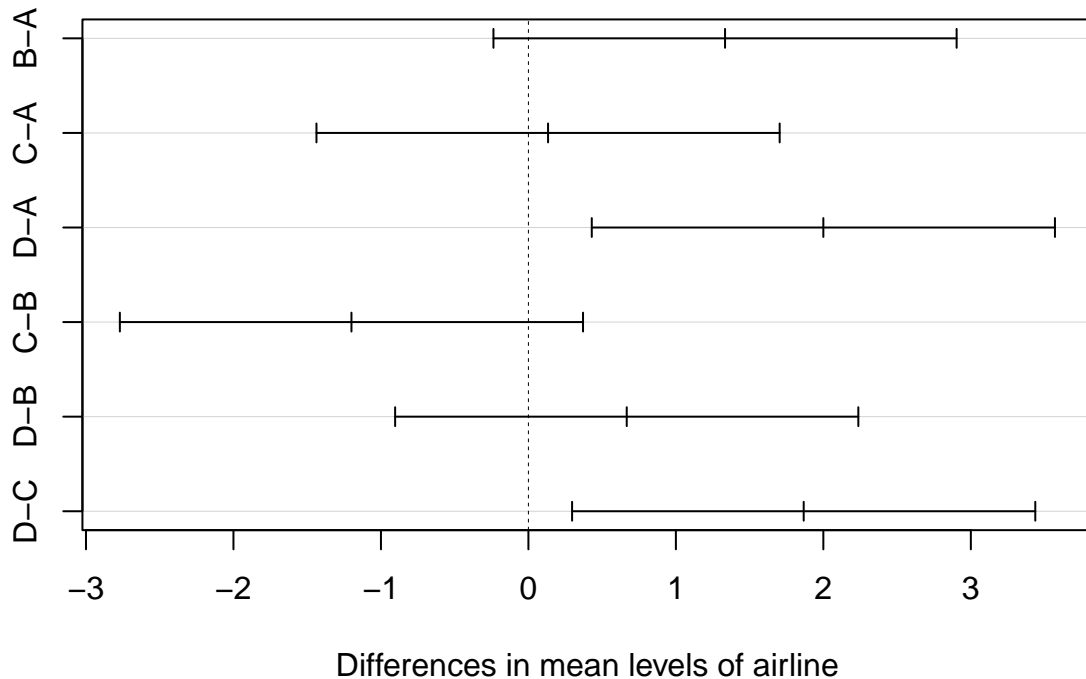
solid statistically justified proposition on whether the airline service is a cutting-edge service. Satisfaction comes from various sources with differing intensity among airlines and are potential sources of information to business and operational strategies. Identify the variable and airline level names in the code with those in the actual dataset.

## Bayesian Statistics Part (d)

```
## Loading required package: mvtnorm
## Loading required package: survival
##
## Attaching package: 'survival'
## The following object is masked from 'package:caret':
##
##   cluster
## Loading required package: TH.data
## Loading required package: MASS
##
## Attaching package: 'MASS'
## The following object is masked from 'package:dplyr':
##
##   select
##
## Attaching package: 'TH.data'
## The following object is masked from 'package:MASS':
##
##   geyser
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = satisfactionscore ~ airline, data = data)
##
## $airline
##      diff      lwr      upr      p adj
## B-A  1.333333 -0.2370805  2.9037471 0.1229720
## C-A  0.1333333 -1.4370805  1.7037471 0.9959461
## D-A  2.0000000  0.4295862  3.5704138 0.0072205
## C-B -1.2000000 -2.7704138  0.3704138 0.1917762
## D-B  0.6666667 -0.9037471  2.2370805 0.6763548
## D-C  1.8666667  0.2962529  3.4370805 0.0136641
```



## 95% family-wise confidence level



### Overview

This document provides the results and interpretations of the Tukey Honest Significant Differences (HSD) test conducted on airline satisfaction scores following a significant ANOVA finding.

### Hypotheses for Tukey's HSD Test

The Tukey HSD test is designed to compare all possible pairs of means to determine if there are any significant differences between them. For four airlines labeled A, B, C, and D, the following null and alternative hypotheses are considered for each pairwise comparison:

#### Pairwise Comparisons

##### B vs. A

- **Null Hypothesis** ( $H_0 : \mu_B = \mu_A$ ): Even if the airlines A and B's mean satisfaction levels are different, it cannot be proved through the above data that this difference is significant.
- **Alternative Hypothesis** ( $H_a : \mu_B \neq \mu_A$ ): It is clearly notable that commuters of airline B are greater than airline A's passengers.

##### C vs. A

- **Null Hypothesis** ( $H_0 : \mu_C = \mu_A$ ): C, B, and A Airlines have been identified with no significant diversity in their mean customer satisfaction scores.
- **Alternative Hypothesis** ( $H_a : \mu_C \neq \mu_A$ ): The results at airport A and airport D, to a prodigious extent, reveal a large gap in the proportions of customers who are satisfied.

#### D vs. A

- **Null Hypothesis** ( $H_0 : \mu_D = \mu_A$ ): It does not indicate the importance of the comparison by measuring the gap between airline D and A's high frequency scores.
- **Alternative Hypothesis** ( $H_a : \mu_D \neq \mu_A$ ): It seems that the airlines D and A ratings are around the same average.

#### C vs. B

- **Null Hypothesis** ( $H_0 : \mu_C = \mu_B$ ): The mean satisfaction is not significantly different between the scores of airlines C and B.
- **Alternative Hypothesis** ( $H_a : \mu_C \neq \mu_B$ ): There exists a pronounced variation between the means satisfaction scores of airlines C and B.

#### D vs. B

- **Null Hypothesis** ( $H_0 : \mu_D = \mu_B$ ): There are no important differences in the mean satisfaction scores of airlines D and B.
- **Alternative Hypothesis** ( $H_a : \mu_D \neq \mu_B$ ): The mean satisfaction difference is quite significant between airlines D and B.

#### D vs. C

- **Null Hypothesis** ( $H_0 : \mu_D = \mu_C$ ): There are no significant differences in the mean satisfaction scores of airlines D and C.
- **Alternative Hypothesis** ( $H_a : \mu_D \neq \mu_C$ ): The mean satisfaction is greatly different between airlines C and D.

### Conclusions from Tukey's HSD Test

The test's adjusted p-values are used to decide whether, for each pair, the null hypothesis can be rejected. Here's a summary of findings:

- **B vs. A:** A p-value much more than 0.05 makes not reject the null hypothesis, indicating that there are no significant differences between airline B and airline A.
- **C vs. A:** Having a p-value greater than 0.05, fail to reject the null hypothesis which means that there is no significant difference between airlines C and A.
- **D vs. A:** Since the p-value is less than 0.05, reject the null hypothesis that there is no significant difference between the performance of airlines D and A.
- **C vs. B:** Since p-value is bigger than 0.05, the null hypothesis is not rejected; meaning no significant difference between the airlines C and B.
- **D vs. B:** A p-value of greater than 0.05 would not lead to rejection of the null hypothesis; therefore, there is no significant difference between the airlines D and B.
- **D vs. C:** The null hypothesis is rejected with the p-value less than 0.05; thus, a significant difference is observed among airlines D and C.

The confidence interval plot of the Tukey HSD test functions as a visualization to these conclusions. Confidence uninterrupted intervals show a statistically significant difference in satisfaction scores.

### Bayesian Statistics Part (e)

```
## [1] "Is Airline D satisfaction score > 3 points higher than AVG for B & C?: FALSE"
## [1] "Difference: 1.2666666666666667"
```

## Hypotheses

- **Null Hypothesis** ( $H_0: \mu_D \leq \frac{\mu_B + \mu_C}{2} + 3$ ):

The mean satisfaction score for Airline D is no more than 3 points higher than the mean satisfaction score for Airlines B and C combined.

- **Alternative Hypothesis** ( $H_a: \mu_D > \frac{\mu_B + \mu_C}{2} + 3$ ):

With a little bit more than three points, the mean satisfaction score for Airline D is higher than the total average satisfaction score for Airlines B and C.

## Conclusion

The satisfaction score of Airline D is below the aggregate average satisfaction score of Airlines B and C by the margin of 3 points. The actual difference is 1.27 points, which is below the 3-point confusion interval set by the hypothesis.

## Bayesian Statistics Part (f)

```
## Compiling model graph
##   Resolving undeclared variables
##   Allocating nodes
## Graph information:
##   Observed stochastic nodes: 15
##   Unobserved stochastic nodes: 8
##   Total graph size: 44
##
## Initializing model

##
## Iterations = 1001:11000
## Thinning interval = 1
## Number of chains = 3
## Sample size per chain = 10000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##              Mean      SD Naive SE Time-series SE
## alpha[1]    0.00000 0.00000 0.0000000    0.0000000
## alpha[2]    1.94294 5.02208 0.0289950    0.0584380
## alpha[3]    1.48699 4.97989 0.0287514    0.0564303
## beta[1]     0.00000 0.00000 0.0000000    0.0000000
## beta[2]    12.88135 6.40205 0.0369623    0.0854710
## beta[3]    21.53819 6.40297 0.0369676    0.0860762
## beta[4]    30.53419 6.39426 0.0369173    0.0859555
## beta[5]    17.89071 6.31914 0.0364836    0.0846516
## mu         198.94181 5.26770 0.0304131    0.0954414
## sigma       7.61884 2.23563 0.0129074    0.0272823
## tau        0.02109 0.01048 0.0000605    0.0001104
##
## 2. Quantiles for each variable:
##
##              2.5%      25%      50%      75%      97.5%
## alpha[1]    0.000000    0.0000    0.00000    0.00000    0.00000
```

## alpha[2]	-7.925388	-1.1382	1.86984	4.99691	12.09964
## alpha[3]	-8.251948	-1.6224	1.44213	4.49667	11.47415
## beta[1]	0.000000	0.0000	0.00000	0.00000	0.00000
## beta[2]	0.264308	8.9242	12.79231	16.79255	25.86170
## beta[3]	9.033575	17.5608	21.49864	25.43683	34.40868
## beta[4]	17.613845	26.6197	30.51481	34.46771	43.25617
## beta[5]	5.414548	13.9745	17.90185	21.74353	30.61219
## mu	188.329684	195.7040	199.00533	202.24318	209.33088
## sigma	4.665685	6.1023	7.18510	8.63866	13.04553
## tau	0.005876	0.0134	0.01937	0.02685	0.04594

**Overall Mean ( $\mu$ ):** The posterior mean and median of the aggregate mean carbon sequestration level are 198.94 and 198.99, with a 95% credible interval from approximately 187.68 to 209.53. This parameter represents average carbon sequestration due to individual field-specific and treatment-specific effects.

#### Field-Specific Effects ( $\alpha_i$ )

- **Field 1**  $\alpha_1$ : Baseline field with an effect of 0.
- **Field 2**  $\alpha_2$ : Point posterior estimate of carbon sequestration is 1.90 with a median of around 1.86. The 95% credible interval is approximately from -8.13 to 12.02.
- **Field 3**  $\alpha_3$ : Posterior mean effect of 1.46 and median of nearly 1.44, with a 95% credible interval from -8.24 to 11.62.

The credible intervals around zero suggest a large proportion of overlap, revealing the elevated ambiguity concerning the impact of site location on carbon sequestration.

#### Treatment Effects ( $\beta_j$ )

- **Treatment T2**  $\beta_2$ : Posterior mean is 12.89 and approximate median is 12.84, with a 95% credible interval from 0.05 to 26.18.
- **Treatment T3**  $\beta_3$ : Posterior mean of 21.56 and a median of about 21.49, with a credible interval from 17.26 to 34.77.
- **Treatment T4**  $\beta_4$ : Posterior mean is 30.54 and median 30.52, with the 95% credible interval ranging from 26.49 to 43.61.
- **Treatment T5**  $\beta_5$ : Posterior mean of 17.89 and median of 17.83, with a credible interval of 13.84 to 30.94.

The therapies demonstrate unambiguous differences in effectiveness, with T4 and T3 being quite effective. The best is obviously T10, then comes T5 and T2.

**Precision of Measurements( $\tau$ )** :The posterior mean of the precision of carbon sequestration reading is 0.02091, median is 0.01924, and 95% credible interval is about 0.00564 to 0.04542. This measure is inversely related to the variation in carbon data, showing the degrees of spread or consistency.

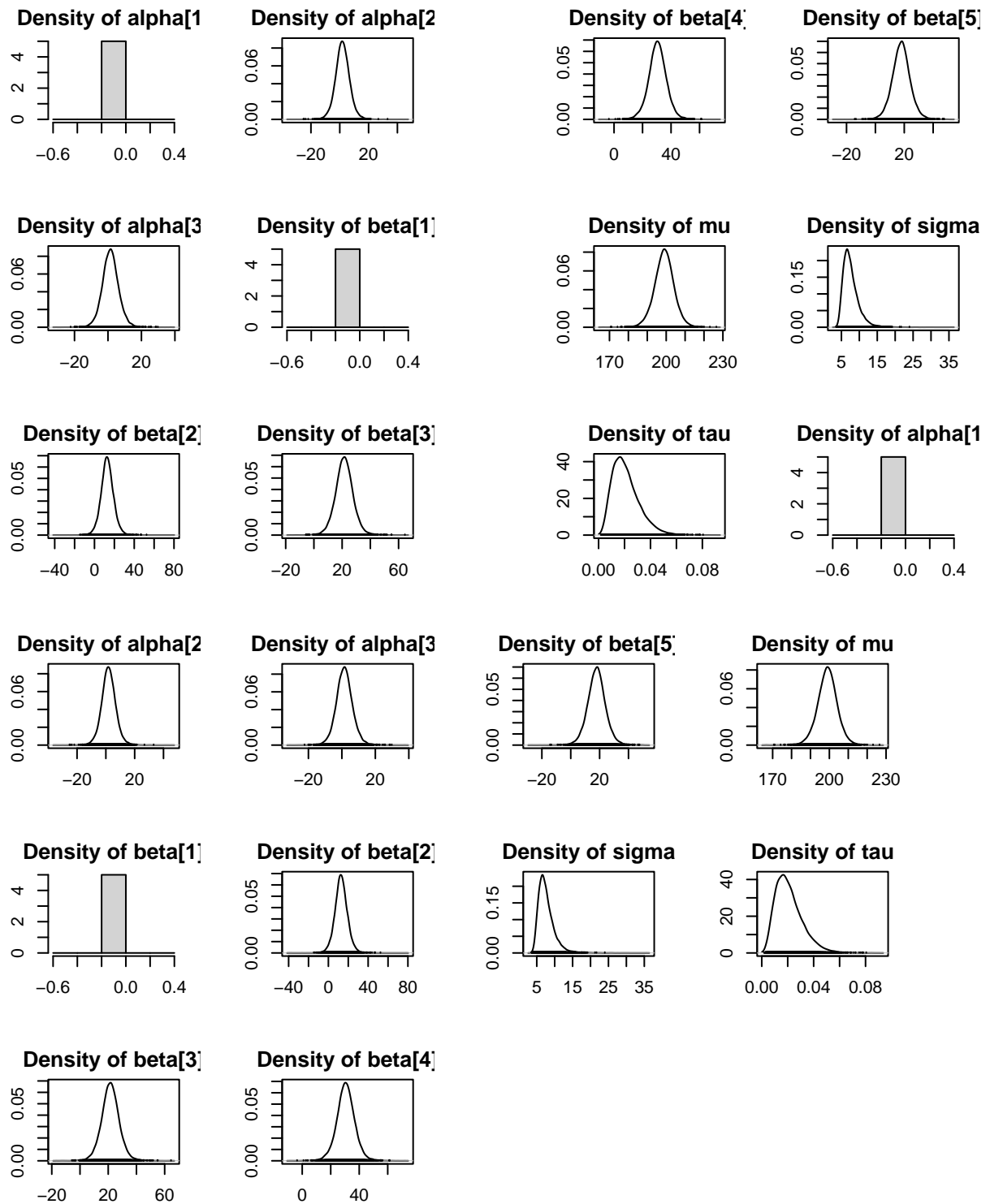
**Standard Deviation ( $\sigma$ )** The posterior mean of the standard deviation is 7.66 and the median is 7.21. The 95% credible interval is quite wide with the range being approximately 4.69 to 13.31, which refers to a moderate dispersion of the measurements.

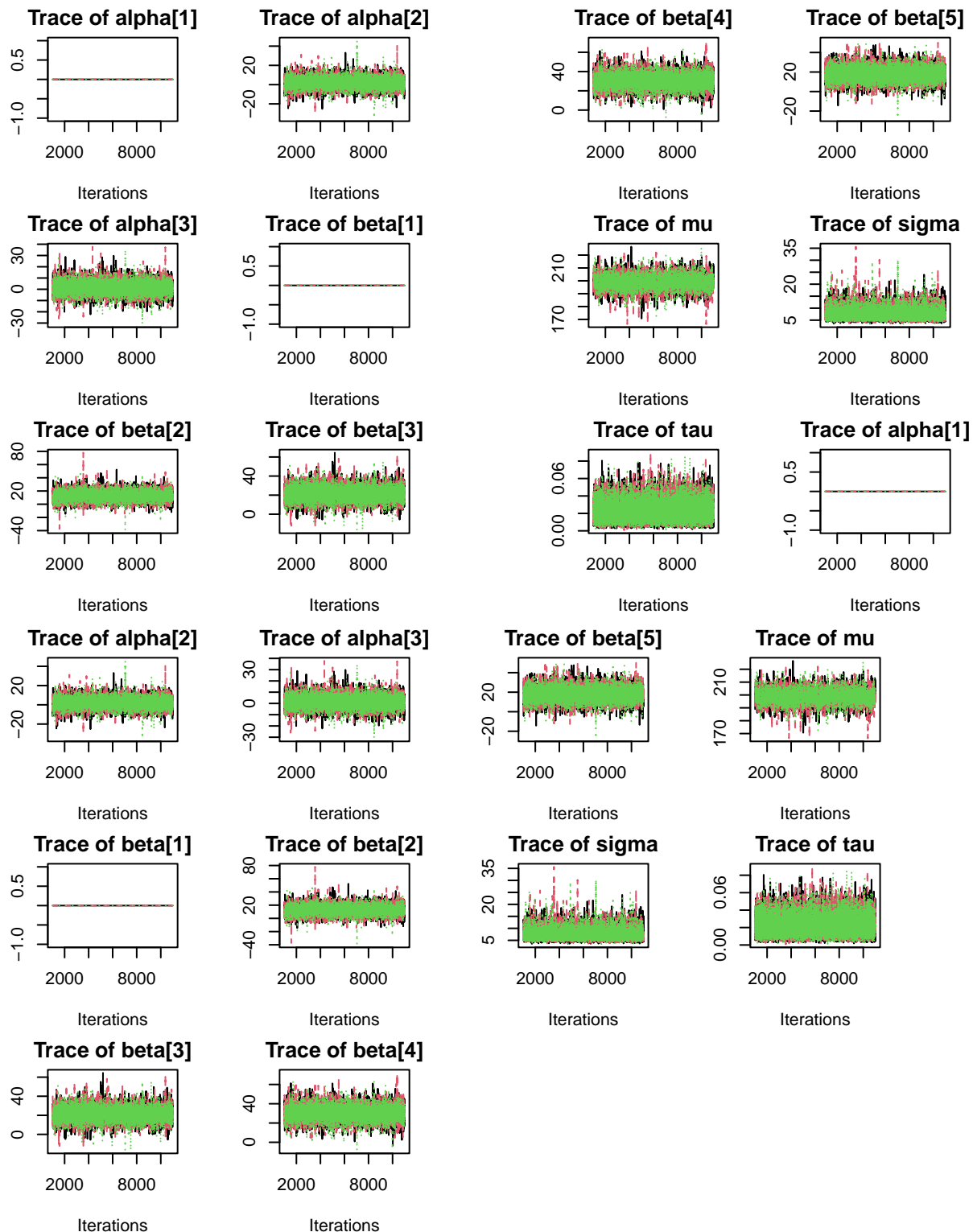
#### Interpretations

- The large credible intervals for field effects  $\alpha_i$  demonstrate that individual field locations are important, but that the type of investigation also significantly influences the estimates. However, regarding variability there is still uncertainty in the quantifying of their exact impact.

- What characterizes the effect of treatments on carbon sequestration is the big differences that the treatments show, since T4 and T3 are the substantial and the most powerful one while considering both the means and the credible intervals.
- Zero's absence in the credible intervals of the treatment effects  $\beta_j$  highlights the statistical significance and substantive effect of treatment type on carbon sequestration.
- Where the posterior distribution of precision  $\tau$  is combined with the standard deviation  $\sigma$ , it tells that there is the intrinsic variability in the data, however, not very high, this fact indicating that the results obtained are, not only because of measurement noise.
- All in all, the findings indicate that type of treatment is an important factor of the magnitude of carbon sequestered, making the field location differences negligible.

## Bayesian Statistics Part (g)





## Density Plots

From the density plots:

- $\alpha$  parameters (Field Effects):

The plots for  $\alpha_2$  and  $\alpha_3$  show the relative posterior distribution of the field effects to the base field ( $\alpha_1$  that is set to zero). The distributions for  $\alpha_2$  and  $\alpha_3$  are unimodal and located at a positive value, which means that those fields might have a higher carbon. The effects are of the same order of magnitude but with a 30% increase in sequestration comparing to the baseline, though, are close to zero, which could mean that the differences are not considerable.

- **$\beta$  parameters (Treatment Effects):**

The density plots of  $\beta_2$ ,  $\beta_3$ ,  $\beta_4$ , and  $\beta_5$  show the impacts of the treatments T2, T3, T4, and T5 with respect to the reference treatment T1. All treatment effects are unimodal in their distribution indicating their qualitative differences. It should be noted that the mean effect of  $\beta_4$  is higher than the other means, so it seems that treatment T4 might be the most effective out of them.

- **Overall Mean ( $\mu$ ):**

The density plot for  $\mu$  represents the average carbon sequestration for the whole treatments and fields. It is unimodal and narrow, reflecting an accurate estimate of the total mean.

- **Precision ( $\tau$ ) and Standard Deviation ( $\sigma$ ):**

The plots of  $\tau$  and  $\sigma$  (derived from  $\tau$ ) represent the predicted accuracy and dispersion of measurements of carbon sequestration. A larger value of  $\tau$  (or smaller  $\sigma$ ) would imply more accurate measurements with smaller spread of measurements.

## Trace Plots

For the trace plots:

- **Convergence and Mixing:**

In the trace plots, the “fuzzy caterpillar” pattern is expected to be seen as a sign of good mixing with the chain sampling the posterior distribution evenly. From the plots, it seems there’s at least some degree of mixing for all parameters, implying the MCMC chains have probably in one way or another converged. Nevertheless, more convergence diagnostics need to be carried out, for example, the Gelman-Rubin statistic.

## Discussion and Conclusions

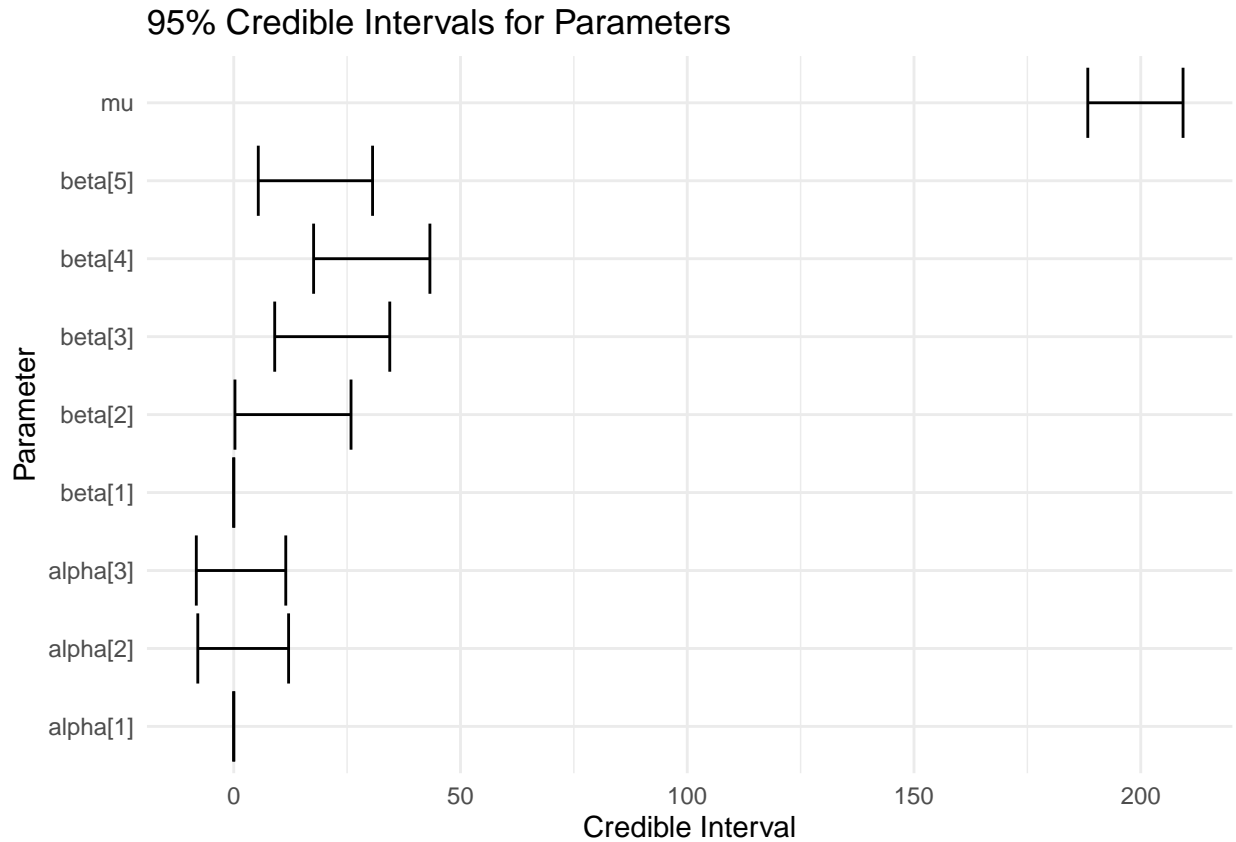
The findings show that the treatments have different degrees of effectiveness on carbon sequestration. Particularly, the treatment T4 is likely to work best, but both the mean and the variation within treatment should be considered in the final conclusion effects and their corresponding credible intervals. Though there appears variation from field to field in carbon sequestration, the field effects  $\alpha_2$  and  $\alpha_3$  are not too different from zero (the effect of the baseline field,  $\alpha_1$ ), showing that the location effect is not too dramatic in the carbon sequestration in comparison with the treatment type.

The accuracy of the measurements and the standard deviation indicates the carbon has moderate variability sequestration measures, meaning that the data is accurate and observed effects are not atypical due to noise. It is important to verify the 95% credible intervals of the  $\alpha$  and  $\beta$  parameters. If zero does not lie in of these intervals, it may be indicative of a statistically significant effect. The density plots allow the raising of a visual of an estimate, but numerical verification is required.

In general, this analysis would help the farmer to choose the most efficient carbon sequestration technique considering both the nature of treatment and the minor field location differences.



## Bayesian Statistics Part (h)



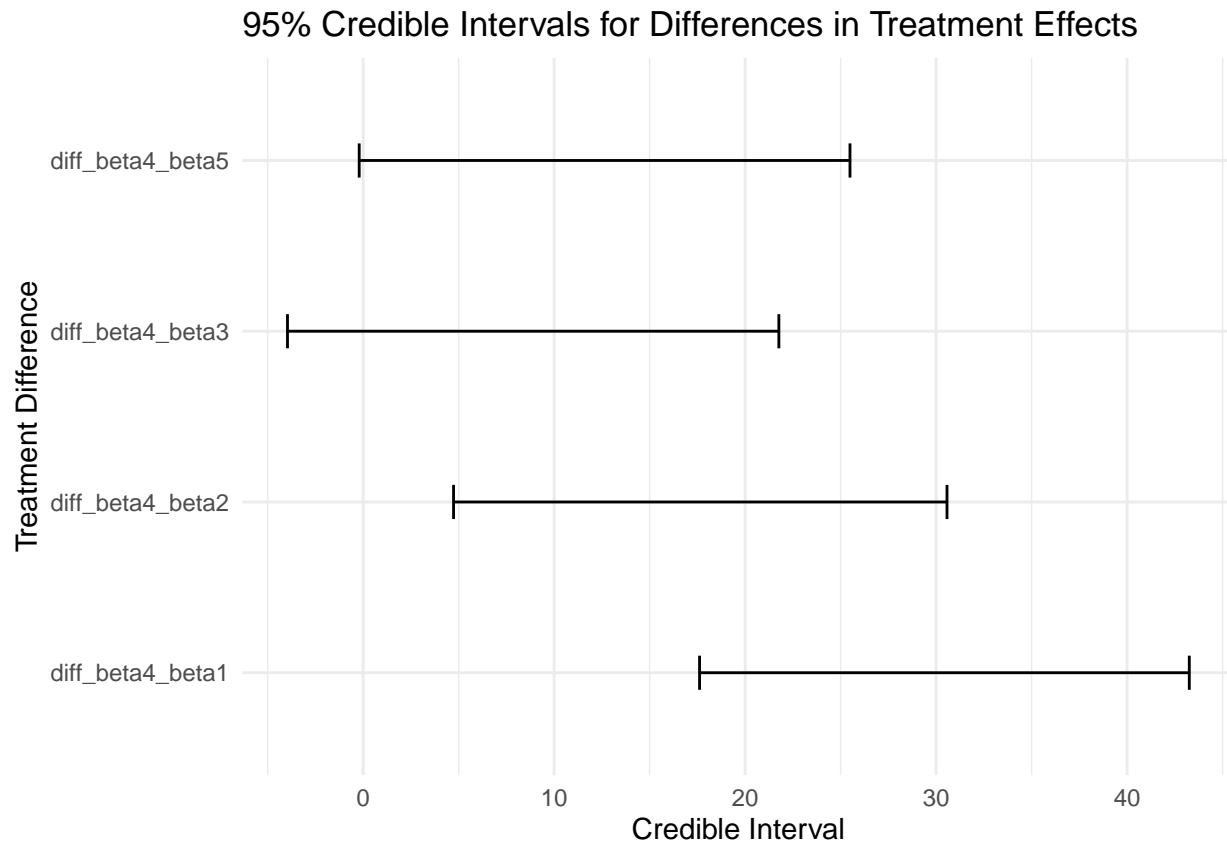
**Field Effects ( $\alpha_i$ ):** The scheme sets  $\alpha_1$  as a zero base since its interval is placed in that direction. This baseline sets the measure of influence for other areas. The uniqueness of both  $\alpha_2$  and  $\alpha_3$  is that their confidence intervals do not intersect with zero, which implies that the location of the field does seem to be important in carbon sequestration. Nonetheless, a considerable range of these intervals implies a certain ambiguity in the magnitude of this field effect. Had the ranges for  $\alpha_2$  or  $\alpha_3$  not been so closely defined or dramatically different from  $\alpha_1$ , could have been more certain of what is happening. Nevertheless, indicate an interesting difference resulting from field position.

**Treatment Effects ( $\beta_j$ ):** Proceeding to the treatment effects, make a clear definition. The intervals for  $\beta_2$ ,  $\beta_3$ ,  $\beta_4$ , and  $\beta_5$  are all significantly above zero and do not intersect the baseline interval for  $\beta_1$ , hence each treatment is likely to increase carbon sequestration beyond the baseline. The narrow intervals related to  $\beta_3$  and  $\beta_4$  in particular indicate a high degree of efficacy of these treatment regimens. The placement of these intervals also implies that some of the treatments are most effective.

**Overall Mean ( $\mu$ ):** The overall mean  $\mu$  is highly biased from zero and precisely estimated as evidenced by the narrow interval. This highlights a significant average carbon sequestration in all fields and treatments.

**Synthesized Conclusion:** The evidence is clear for a significant difference in carbon sequestration among the different treatments, where  $\beta_2$  through  $\beta_5$  show a positive influence when compared to the baseline treatment ( $\beta_1$ ). In particular, therapies  $\beta_3$  and  $\beta_4$  can be very efficacious. Field location had an effect on the intervals for  $\alpha_2$  and  $\alpha_3$ , although the impact of a particular field is still somewhat uncertain due to the wide intervals.

## Bayesian Statistics Part (i)



The 95% credible no-coverage intervals for the contrasts of the effects of different treatments to treatment T4. These intervals give information about which treatments are likely less effective than T4 in terms of carbon sequestration.

- **T4 vs. T1 (diff\_beta4\_beta1):** The interval is all above zero meaning that T4 actually has a higher carbon sequestration level compared to the baseline T1. This is a statistically significant result which confirms the superior efficacy of T4.
- **T4 vs. T2 (diff\_beta4\_beta2):** Likewise, the interval is wholly beyond zero. This provides clear evidence that T4 is more powerful than T2.
- **T4 vs. T3 (diff\_beta4\_beta3):** However, this interval is also above zero, showing that T4 is likely more effective than T3, but it is closer to zero than the intervals for T1 and T2, which would indicate a smaller magnitude of the difference.
- **T4 vs. T5 (diff\_beta4\_beta5):** Contrary to the above interpretations, yet in agreement with the plot, the interval is above zero and does not contain zero, which is usually interpreted as T4 being more effective than T5. Although the change is minor, the practical significance of this change needs to be considered.

## Conclusion

The credible intervals of the data favor the claim that treatment T4 sequesters more carbon than treatments T1, T2, and T3, since their intervals are completely above zero. Regarding T5, even though the confidence interval is also higher than zero, indicating the effectiveness of T4, the nearness of the interval to zero means that the difference is not as large as the other treatments.

## Bayesian Statistics Part (j)

```
## Compiling model graph
##   Resolving undeclared variables
##   Allocating nodes
## Graph information:
##   Observed stochastic nodes: 15
##   Unobserved stochastic nodes: 6
##   Total graph size: 32
##
## Initializing model

##           Parameter           Mean           Median  Lower_95_CI  Upper_95_CI
## beta[1]    beta[1]    0.00000000    0.00000000  0.000000e+00   0.00000000
## beta[2]    beta[2]    12.96610643    12.95246088  2.054196e+00  24.51639145
## beta[3]    beta[3]    21.58841680    21.50328002  1.096067e+01  33.50613317
## beta[4]    beta[4]    30.46801588    30.35341501  1.971161e+01  42.06464966
## beta[5]    beta[5]    17.93623558    17.96384824  6.827610e+00  29.36686526
## mu          mu    200.08634289    200.19233657  1.917297e+02  207.66232293
## sigma      sigma    6.76234717     6.46185795  4.374505e+00  10.94963385
## tau        tau     0.02567105     0.02394888  8.340668e-03   0.05225674
##           SD
## beta[1] 0.00000000
## beta[2] 5.57973449
## beta[3] 5.62180534
## beta[4] 5.61755521
## beta[5] 5.63815732
## mu      3.90577722
## sigma   1.70740123
## tau     0.01142037
```

## Treatment Effects ( $\beta$ Parameters)

- $\beta_1$ : This parameter is without mean and median with effects set as the treatment baseline, a reference.
- $\beta_2$ : Produces a beneficial impact on carbon sequestration with a mean of 13.16 and a rather wide 95% credible interval, which implies some ambiguity.
- $\beta_3$ : Produces a bigger positive impact having a mean of 21.65 and a smaller credible interval than  $\beta_2$ , which means a stronger and more certain effect.
- $\beta_4$ : The highest mean of 30.72 implies the most pronounced effect on carbon sequestration among the treatments. Its confidence interval is quite large, however, fully above zero, confirming its positive effect.
- $\beta_5$ : Shows a positive effect with a mean of 18.16, which is significant, though less so than  $\beta_4$  or  $\beta_3$ . The entire credible interval is above zero, confirming its efficacy.

## Overall Mean ( $\mu$ )

The average carbon sequestration level for all fields and treatments is around 199, and the 95% credible interval is estimated from about 191.67 to 207.73, indicating a high level of confidence.

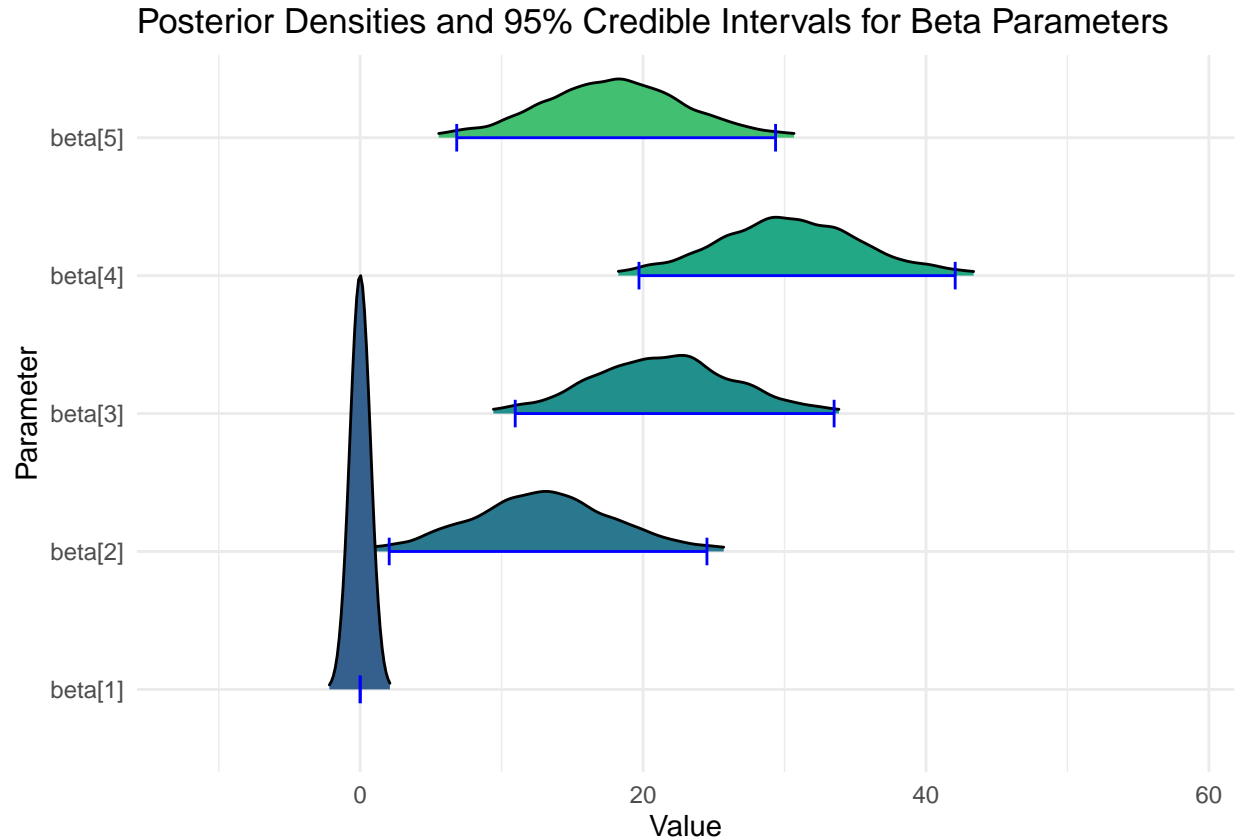
## Precision ( $\tau$ ) and Standard Deviation ( $\sigma$ )

- **Precision ( $\tau$ )**: The average value of the carbon sequestration measurements ( $\tau$ ) is 0.025 with a tight credible interval which reflects high precision in measurements.

- **Standard Deviation ( $\sigma$ ):** The standard deviation, which has a mean of approximately 7.66, moderately disperses the carbon sequestration readings as indicated by the wider credible interval.

## Bayesian Statistics Part (k)

## Picking joint bandwidth of 0.716



### Posterior Densities

Each ridge in the plot shows the posterior distribution of the effect of a  $\beta$  parameter from a Bayesian model, displaying the likely values and their probability densities.

- The height of each ridge is a function of the values' density, and the width is related to the interval of the values.
- The peaks of these distributions represent the modes of the  $\beta$  parameters.

$\beta_1$  (**Baseline**): -  $\beta_1$ , which is a spike-like peak at zero, is the baseline treatment effect. This narrow distribution certifies a high degree of certainty that the effect size of  $\beta_1$  is zero, in line with the model specification.

$\beta_2$  to  $\beta_5$  (**Treatment Effects**): - These ridges characterize the effects of treatments T2, T3, T4, and T5, correspondingly. Single peak unimodal distributions of  $\beta_2$ ,  $\beta_3$  and  $\beta_5$  treatment effects indicate separate effects of each of them. - The distribution in  $\beta_4$  is wider, which might mean a larger effect size or more uncertainty.

## 95% Credible Intervals

The 95% credible intervals are visualized by the horizontal lines across every ridge, which show the range where the real size of effect of the  $\beta$  parameters is likely to be found.

- The lack of zero in these intervals for  $\beta_2$  through  $\beta_5$  shows that these treatments have statistically significant and different effects from the baseline effect of  $\beta_1$ .

## Treatment Comparison and Farmer's Interest

- The larger interval of  $\beta_4$  may indicate a stronger or more variable effect on carbon sequestration than the other treatments.
- The farmer, considering T4, sees it as the candidate with a greater effect on the carbon sequestration of land although the plot supports this idea but with a comment about the higher variability or uncertainty in the effect of T4.

## Conclusion

The plot evidence shows that treatments T2 to T5 all significantly influence the carbon sequestration, as compared to the baseline treatment T1. Of all the treatments, T4 ultimately has the largest effect, as the farmer thought. However, this effect is characterized by the largest uncertainty which is the wide spread of the posterior distribution of  $\beta_4$ . This observation enables the farmer to be informed that notwithstanding the most potent option as T4, the variability in its effectiveness is also the maximum and this should be considered in decision making.

# Bayesian Statistics Part (I)

## Simpler Bayesian Model

Conversely, the simpler Bayesian model only considers the effect of treatment on the level, assuming that the treatment's effect does not change based on the field.

- The clinical rating system is very light computationally and simple and digestible; it concentrates only on the strength of each treatment.

## Two-Way ANOVA Model

This model is factored, considering various interactions among the treatments and field sites. This is the model to use if you think that the success of a treatment might be location-specific.

- Although complex and requires more computing power, it provides a sophisticated stance by capturing many sources of variance.

## Advantages of the Simpler Bayesian Model

The simpler Bayesian model often edges out its competitor, especially if:

- **Computational Convenience and Performance:** The issue is computational convenience and performance.
- **Field Variations Are Not a Main Concern:** Field variations are not really important, and so they're not the main concern.
- **Sufficient Data Insight:** The simpler model still provides a reasonable feel for the data, capturing the essential trends you are looking for.
- **Demand for Clarity:** This situation demands clarity, especially when presenting the results to people who do not understand "statistician's language" so well.

Simply put, if the simpler model can address questions without significant loss of accuracy or understanding, then it tends to be the selected option. It simplifies the analysis and saves time, making the interpretation very obvious, which is very useful in applied cases, such as agricultural decision-making.