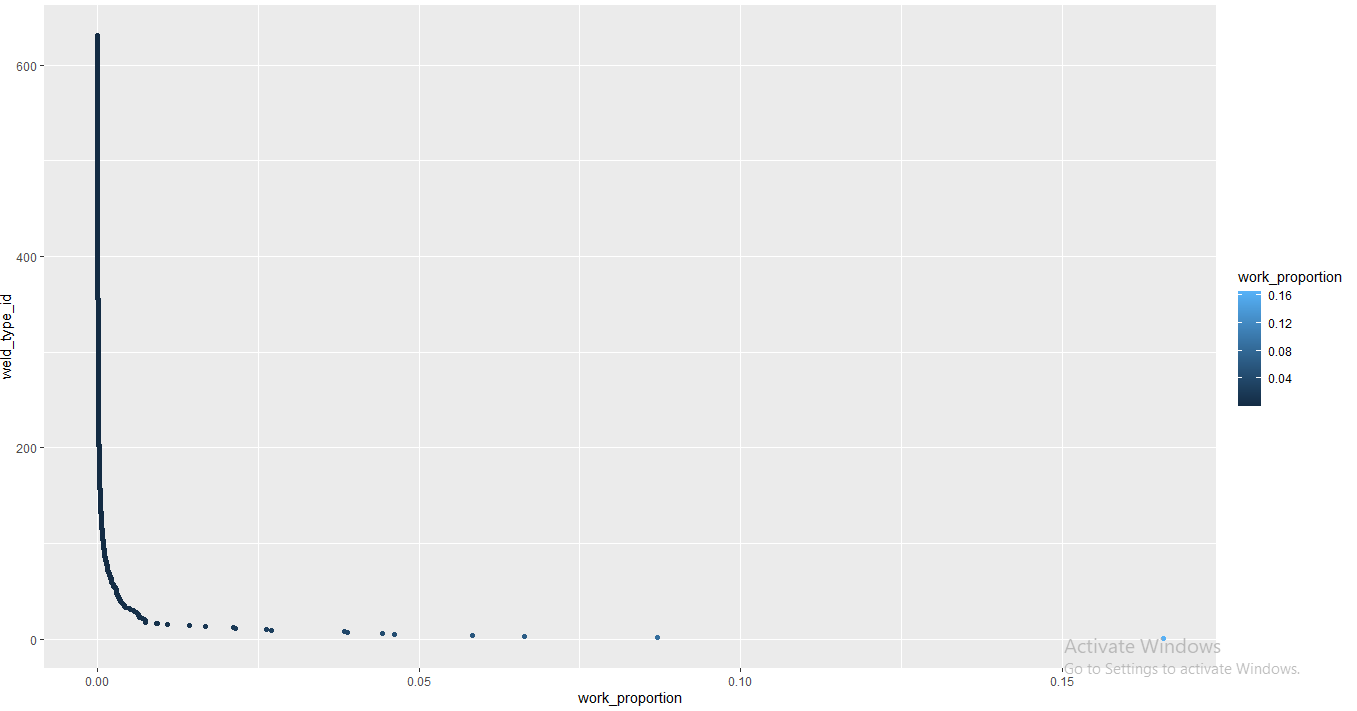
CEIE 474/574 Construction Computer Application and Informatics

Assignment 5 Solution

**STEP 1**

1. To account for the 80 percent of the business we have almost 631 weld types with their different work proportions and their own failed, inspection and repair rate. Each weld type has different purpose in the field of the construction.
2. To visualize the weld types we have to import the dataset first in the R and then we will use the library ggplot to visualize the dataset. The visualization of the dataset will look like below;

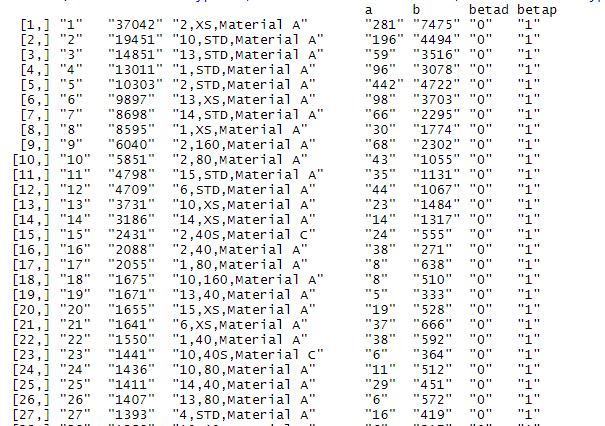


In the above plot we can clearly see that weld types are decreasing from 600 to 0 which shows that the lower the use of them in the construction while the decreasing weld type also shows increased number of work proportions.

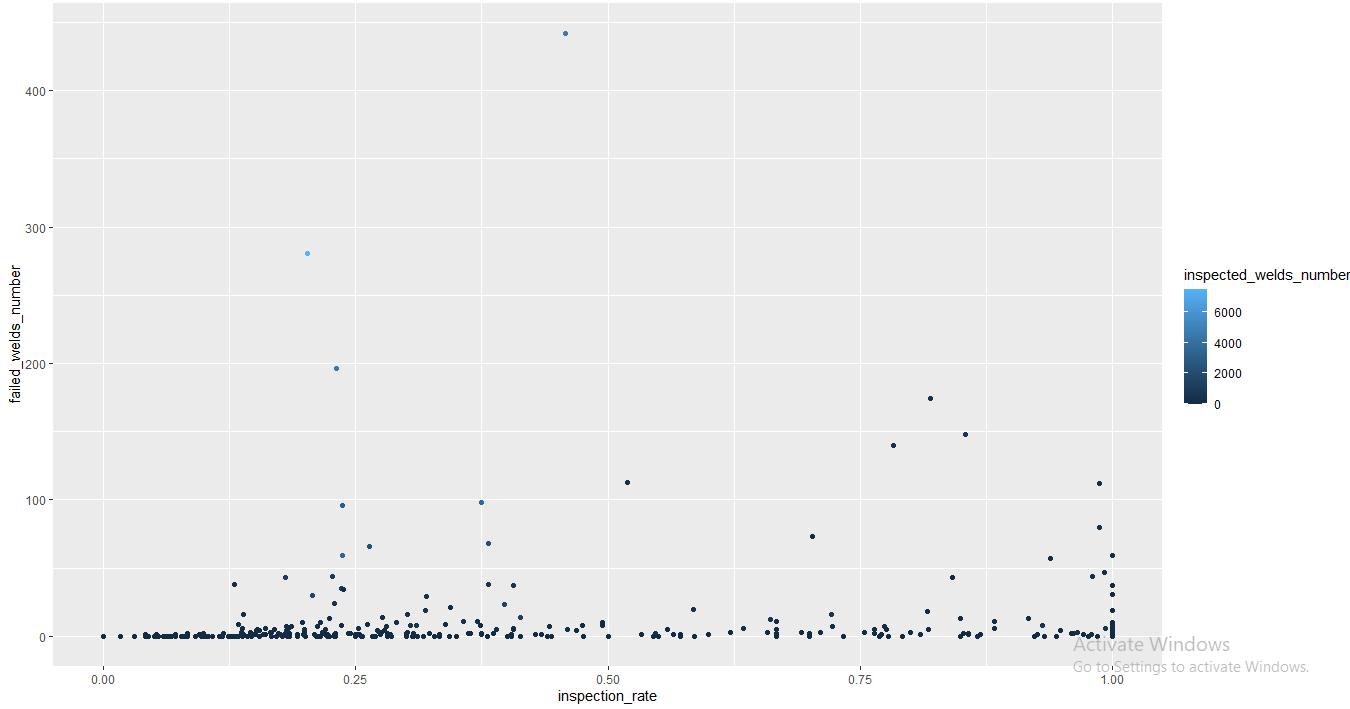
Hence we can say that the number shows the intensity of weld flame the higher the number means higher the risk to eye and other delicate parts of the human body and that is why they are least used.

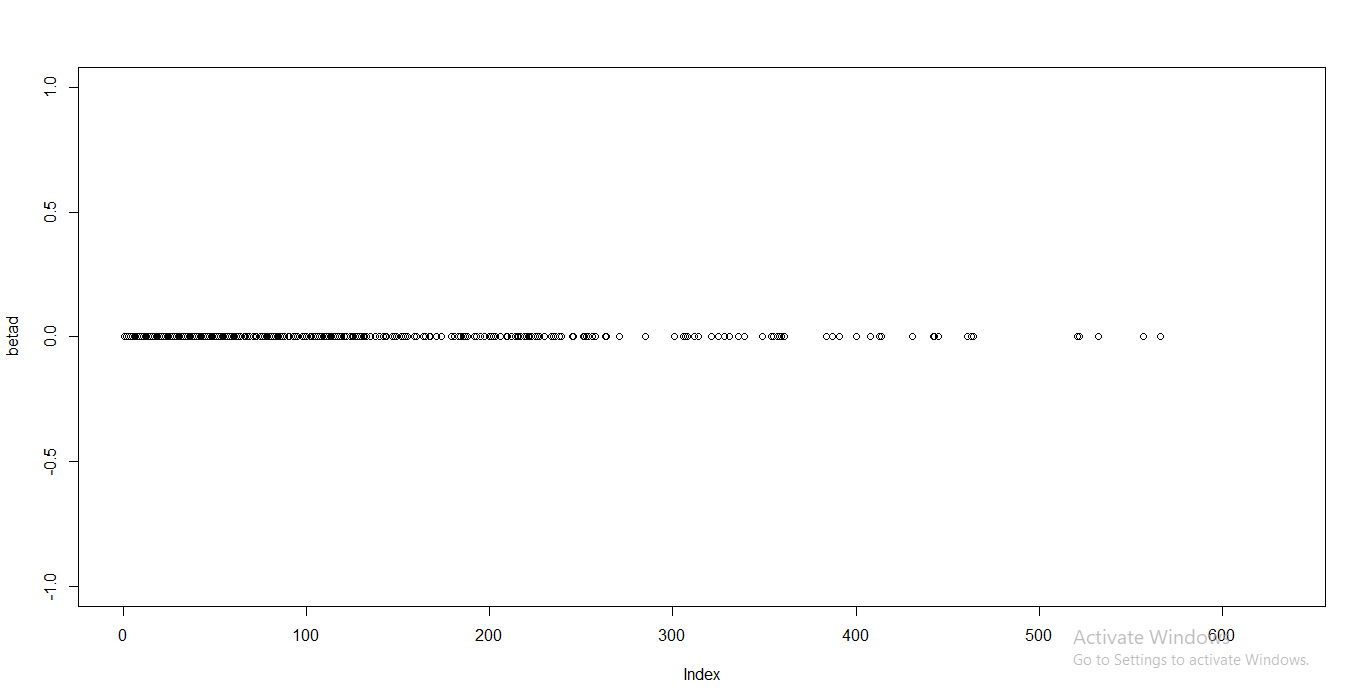
**STEP 2**

1. The beta distribution calculation contains some parameters of the tables like number of welds and failed welds and inspected welds number etc. after collecting all the data simply run the pbeta and qbeta command also to create the table we have to combine the values of weld id, weld type and the values of a and b in one table which looks like below.



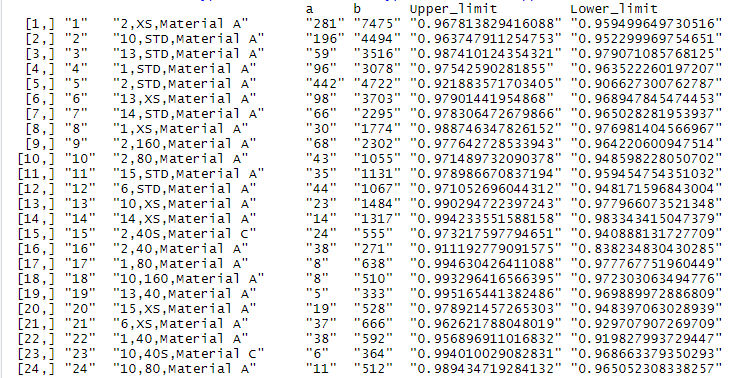
1. To visualize the data of the top welds we simply have to run the plot command for the values we have calculated for the beta distribution and it will give the graph which looks like below;

 and the plot for the beta distribution will look like below;

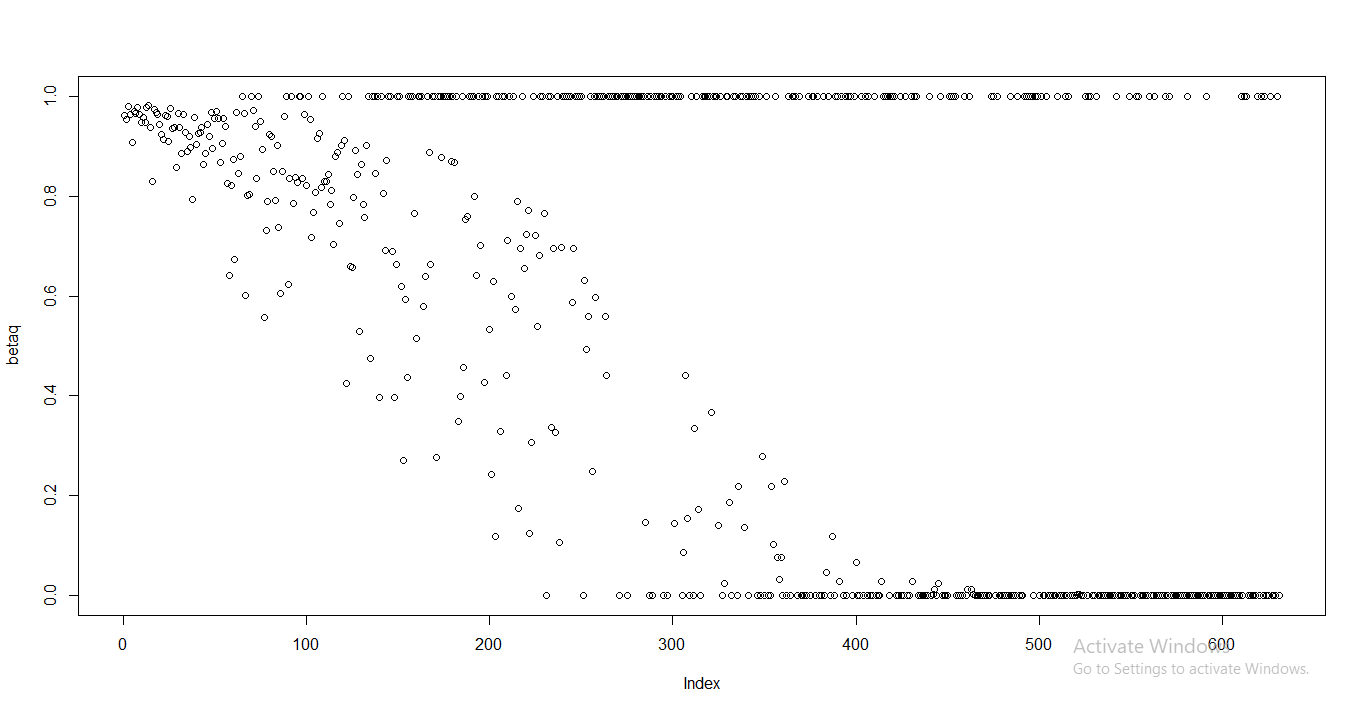


**STEP 3**

1. To calculate the confidence interval we first have to define the upper and lower limit of the given dataset which can be calculated by the qbeta function after finding out the values the we simply has to calculate the error and then add it with upper limit while subtract with the lower limit and hence we can easily find out the confidence interval of the dataset. The table looks like below after combining the values.

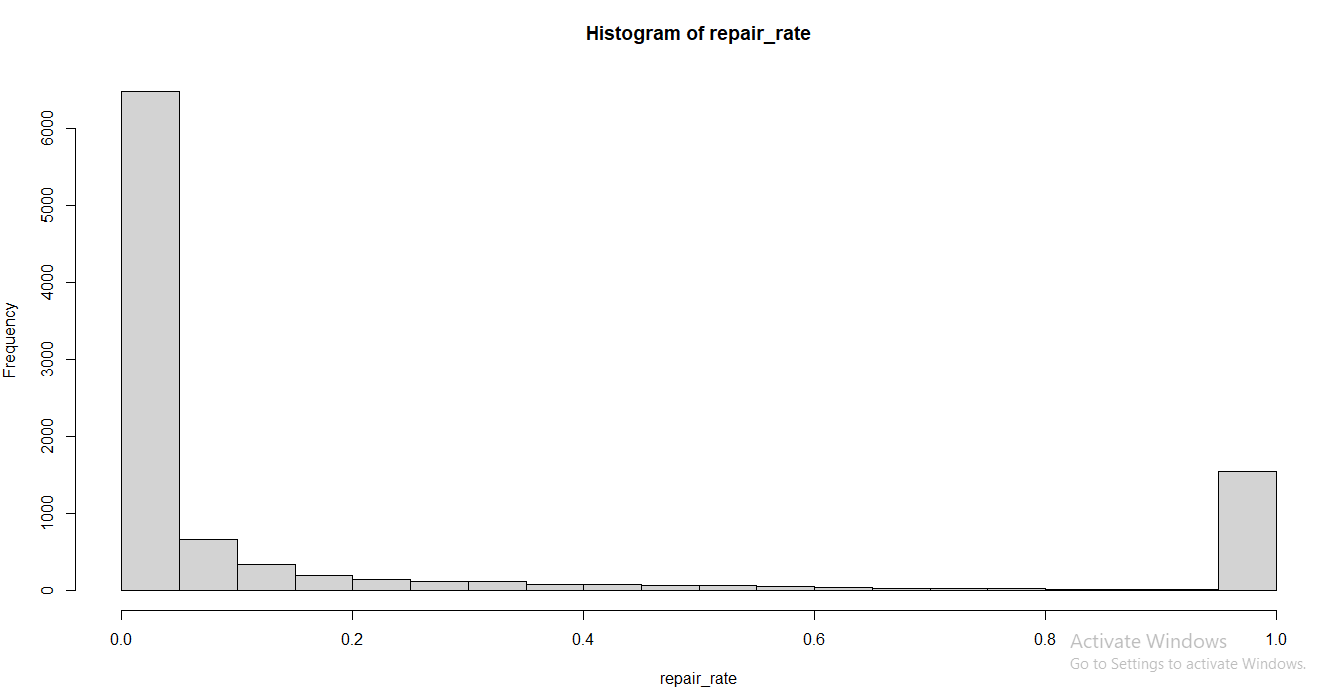


1. The plot of the dataset after the confidence interval is achieved will look like below;

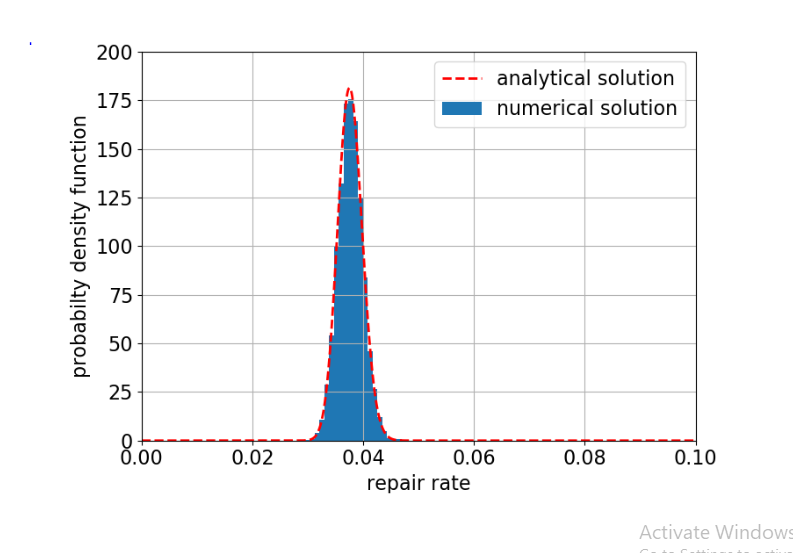


**STEP 4**

1. To plot the histogram first we have to define the repair rate of the data with the 10000 samples then we have to calculate the beta values of the samples and then we can easily plot the graph which looks like below;



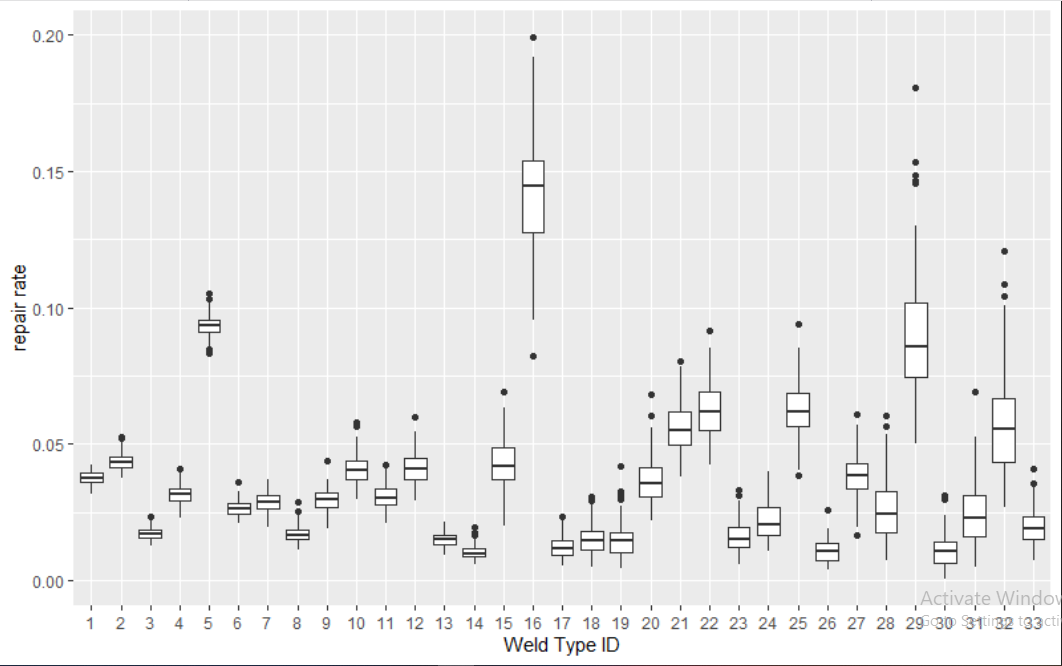
The distributed frequency graph for the dataset with respect to probability distribution will look like below;



From the above histogram we can easily conclude the frequency of the repairing of the welds in using the beta function

**STEP 5**

1. To visualize the boxplot of the repair rate first we have to combine the sample values to the dataset and then we can simply use the ggplot command to plot the visualization of the dataset in the system. The boxplot of the dataset with sample value will look like below;



**STEP 6 (BONUS)**

1. To calculate the confidence interval with 95% significance we simply have to calculate the error values with left and right values in the dataset which could be find out in the weld inspection rates and failed weld numbers, after finding the error we can simply find out the difference between them.





1. The root mean square error also known as rmse works on following basic formula,

Rmse(predicted value, actual value)

In this case the predicted value could be the repair rate as it has taken by the user but it will not going to work in expected way until we find out the root mean square for which first we have to find out the mean of the data values and then we can easily find out the rmse value.



1. To interpret the value of the rmse we have to find the difference between lower limit and the upper limit of the errors thus outcome can interpret the difference between actual and predicted values.

