Survival Analysis of Bridge Deck

Chandan Saha{chandan.saha@rutgers.edu }

1 Data Cleaning

Data contains inspection information about Bridge Deck elements. In this report I am trying to analyze NBE data to find out survival probability of a bridge on given data.
DataClean.csv data describes bridge condition information and other related information such as skew, year built, ADT, etc. Identical StructureNumber shows the information of same structure.
Loading Data is done by,

```
DataClean <- read.table("DataClean.csv", header= TRUE, sep= ",")
```

```
SurvivalData <- read.csv('Survival.csv')
```

1.1 Filter Deck Structure

1) only include data when ElementNumber is 12, which means, deck structure ??
We are only considering rows with ElementNumber equal to 12. Code is given below,

```
Data_DeckStructure <-DataClean[which(DataClean$ElementNumber==12),]
```

Output : The dataset has 24624 records with bridge deck and out of which 4670 rows has missing values.Percentage of missing values based on features is given by,

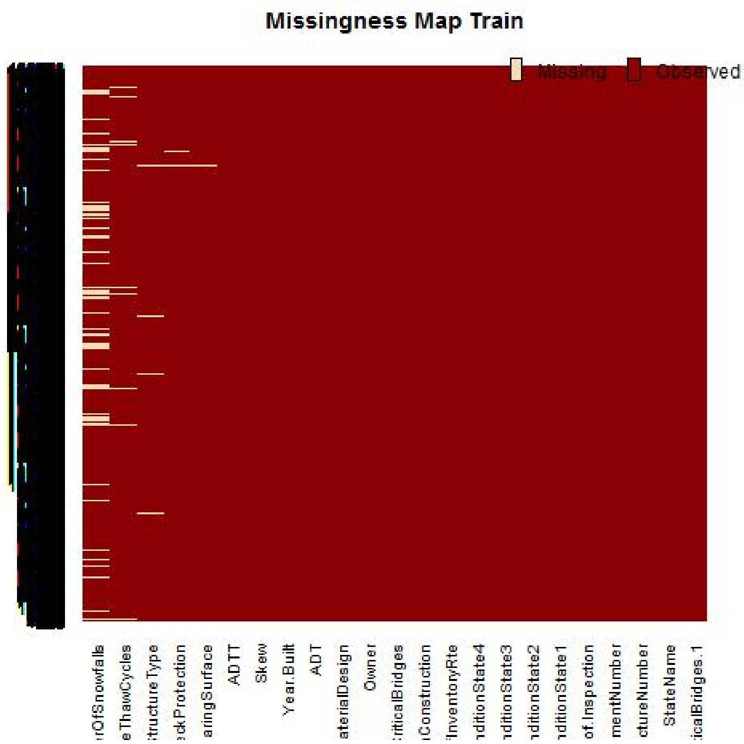

Figure 1: Missing Field map

So, NumberofSnowfall has maximum missing values(Around 16%) but we are not considering those as none of the condition state has missing values according data.

1.2 Delete/Drop NAN Values

2) Drop or delete all the data of a single bridge if any of the data entry of that bridge has Null input of condition state data Check on condition state field,

```
# Check N/A for Condition State Coloumns
sum(is.na(Data_DeckStructure$ConditionState1))
sum(is.na(Data_DeckStructure$ConditionState2))
sum(is.na(Data_DeckStructure$ConditionState3))
sum(is.na(Data_DeckStructure$ConditionState4))
```

Output:

```
> sum(is.na(Data_DeckStructure$ConditionState1))
[1] 0
> sum(is.na(Data_DeckStructure$ConditionState2))
[1] 0
> sum(is.na(Data_DeckStructure$ConditionState3))
[1] 0
> sum(is.na(Data_DeckStructure$ConditionState4))
[1] 0
```

1.3 Condition Status field Consistency

3) Drop or delete the all the data of a single bridge if the overall Condition State quantity of elements for Deck 12 changes during any year of inspection for that specific bridge ??

```
Data_DeckStructure$TotalElement = Data_DeckStructure$ConditionState1 +
Data_DeckStructure$ConditionState2 + Data_DeckStructure$ConditionState3
+Data_DeckStructure$ConditionState4
head(Data_DeckStructure$TotalElement)

UniqueStructure <- as.data.frame(unique(Data_DeckStructure
[c("StructureNumber", "TotalElement")]))
head(UniqueStructure)


results <- as.data.frame(table(UniqueStructure$StructureNumber))
head(results)
ConstCondState <- results[ which(results$Freq=='1'),]


newdata <- Data_DeckStructure[Data_DeckStructure$StructureNumber %in%
ConstCondState$Var1, ]
head(newdata)
```

Output: There are total 1737 unique structures and total of 11119 rows in the dataset after cleaning up all duplicate records . Output of clean dataset is stored as output.csv file in results folder.

Feature Selection

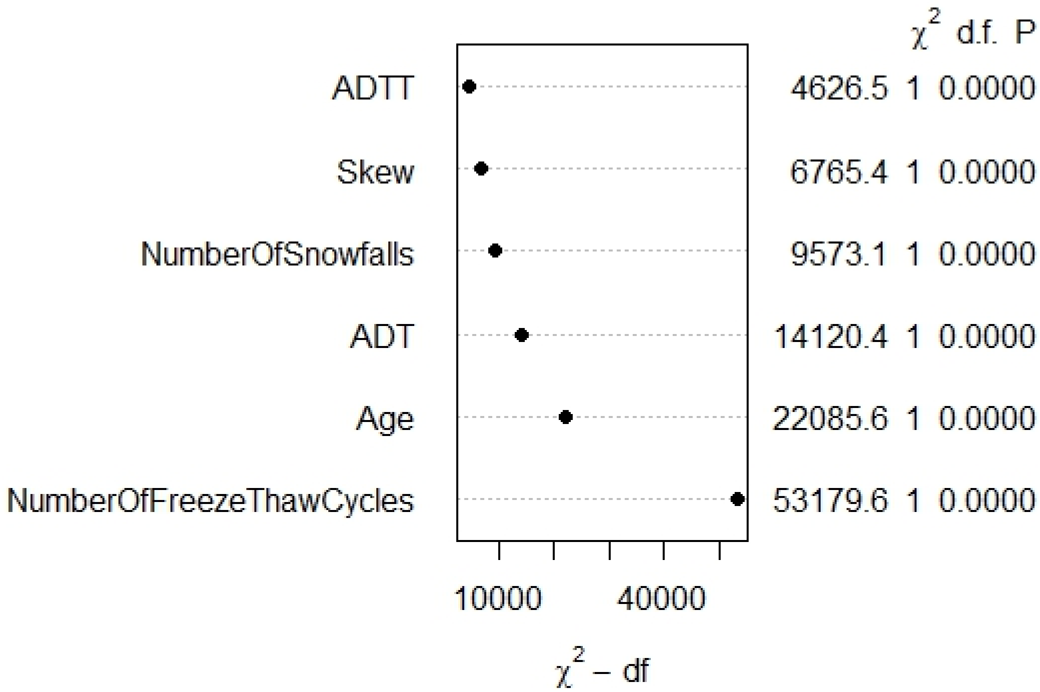

Figure 2: Feature Selection

Based on feature selection using anova , all features looks important and not heavily correlated(Failed hypothesis test for correlation). I have added features based on Akaike information criterion.

2 Survival Analysis

Reliability index is common approach is most commonly utilized method for bridge's future service life prediction. The extent of deterioration renders the bridge not serviceable. Therefore, survival analysis may be more appropriate as the contributing factors are studied without a direct causal relationship between one factor (load) and failure. In survival analyses, the time to any event is considered a random variable T, (survival time) with probability density function f(t), where t is time. Survival function, S(t), and hazard function, h(t) are defined below (Elsayed, 2012):

$$S(t) = P(T \geq t) = 1 - F(t)$$

Where F(t) is defined as:

$$F(t) = P(T < t) = \int_0^t f(t)dt$$

$$h(t) = \lim_{\delta t \to 0} \{ \frac{P(t \leq T \leq t + \delta t | T \geq t)}{\delta t} \}$$

Hazard function, survival function, and probability function are related to each other as below:

$$h(t) = \frac{f(t)}{s(t)}$$

Hazard is referred to as probability of failure per unit time at any given time assuming survival up to that time. Weibull , log-normal , log-logistic are most commonly used models for survival analysis. **Weibull modeling** can be done in two ways.One is to include weights(which is quantity in our dataset) , on other hand modeling can be done with out considering weights as well. I will analyzed both models based on our dataset.

Weibull Model With Weights

Code looks like , Here we are building model with all fields in the survival dataset including weights.We are only considering time in state greater than zero as negative values are invalid in our scenario.

```
#Weibull Model Fit - With Weight
weibull <- survreg(Surv(TIS, Survival)~ Age + ADTT+ ADT+
Skew+ NumberOfSnowfalls+ NumberOfFreezeThawCycles,
weights = Quantity, data = SurvivalData[SurvivalData$TIS >0,],
dist = "weibull")

summary(weibull)
```

Output: Weibull accelerate failure time (AFT) model

```
> summary(weibull)

Call:
survreg(formula = Surv(TIS, Survival) ~ Age + ADTT + ADT + Skew +
    NumberOfSnowfalls + NumberOfFreezeThawCycles, data = SurvivalData[SurvivalData$TIS >
    0, ], weights = Quantity, dist = "weibull")
                            Value Std. Error      z p
(Intercept)              3.74e+00   3.43e-03 1090.9 0
Age                      1.07e-02   7.21e-05  148.6 0
ADTT                     8.04e-05   1.18e-06   68.0 0
ADT                     -2.24e-05   1.89e-07 -118.8 0
Skew                     4.78e-03   5.82e-05   82.3 0
NumberOfSnowfalls        1.05e-02   1.07e-04   97.8 0
NumberOfFreezeThawCycles -7.61e-03   3.30e-05 -230.6 0
Log(scale)               8.48e-02   9.29e-04   91.3 0

Scale= 1.09

Weibull distribution
Loglik(model)= -4185013   Loglik(intercept only)= -4232015
        Chisq= 94005.2 on 6 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 7
n=1177 (31 observations deleted due to missingness)
```

Figure 3: Weibull accelerate failure time (AFT) model

Weighbull Model without Weights

```
#Weibull Model Fit - Without Weight
weibull_nonweight <- survreg(Surv(TIS, Survival)~ Age + ADTT+ ADT+
Skew+ NumberOfSnowfalls+ NumberOfFreezeThawCycles,
data = SurvivalData[SurvivalData$TIS >0,],
dist = "weibull")

summary(weibull_nonweight)
```

Output: There are total 1737 unique structures and total of 11119 rows in the dataset after cleaning up all duplicate records . Output of clean dataset is stored as output.csv file in results folder.

```
> summary(weibull_nonweight)

Call:
survreg(formula = Surv(TIS, Survival) ~ Age + ADTT + ADT + Skew +
    NumberOfSnowfalls + NumberOfFreezeThawCycles, data = SurvivalData[SurvivalData$TIS >
    0, ], dist = "weibull")
                            Value Std. Error      z         p
(Intercept)              2.75e+00   1.13e-01 24.260 5.16e-130
Age                      7.62e-03   2.28e-03  3.339  8.39e-04
ADTT                     1.25e-05   5.59e-05  0.224  8.23e-01
ADT                     -1.86e-05   9.62e-06 -1.929  5.38e-02
Skew                     1.13e-03   2.15e-03  0.523  6.01e-01
NumberOfSnowfalls        3.40e-03   4.11e-03  0.826  4.09e-01
NumberOfFreezeThawCycles -2.69e-03   1.11e-03 -2.424  1.53e-02
Log(scale)              -5.64e-02   3.69e-02 -1.532  1.26e-01

Scale= 0.945

Weibull distribution
Loglik(model)= -2097.5   Loglik(intercept only)= -2113.3
        Chisq= 31.58 on 6 degrees of freedom, p= 2e-05
Number of Newton-Raphson Iterations: 5
n=1177 (31 observations deleted due to missingness)
```

Figure 4: Weibull accelerate failure time (AFT) model - Excluding Weight

3 Prediction Of survival probability

3.Please predict the survival probability of a bridge deck elements in condition state 1, using the model you developed in task 2, with age=26.68, ADT=5267, ADTT=763.37, Skew=13.39, NumberOfSnowfalls=14.88, NumberOfFreezeThawCycles=56.34. Plot the survival probability (it is equivalent to the percentage of total elements in condition state 1) for 50 years (starting from first inspection year) and calculate the survival probability when year t=25.

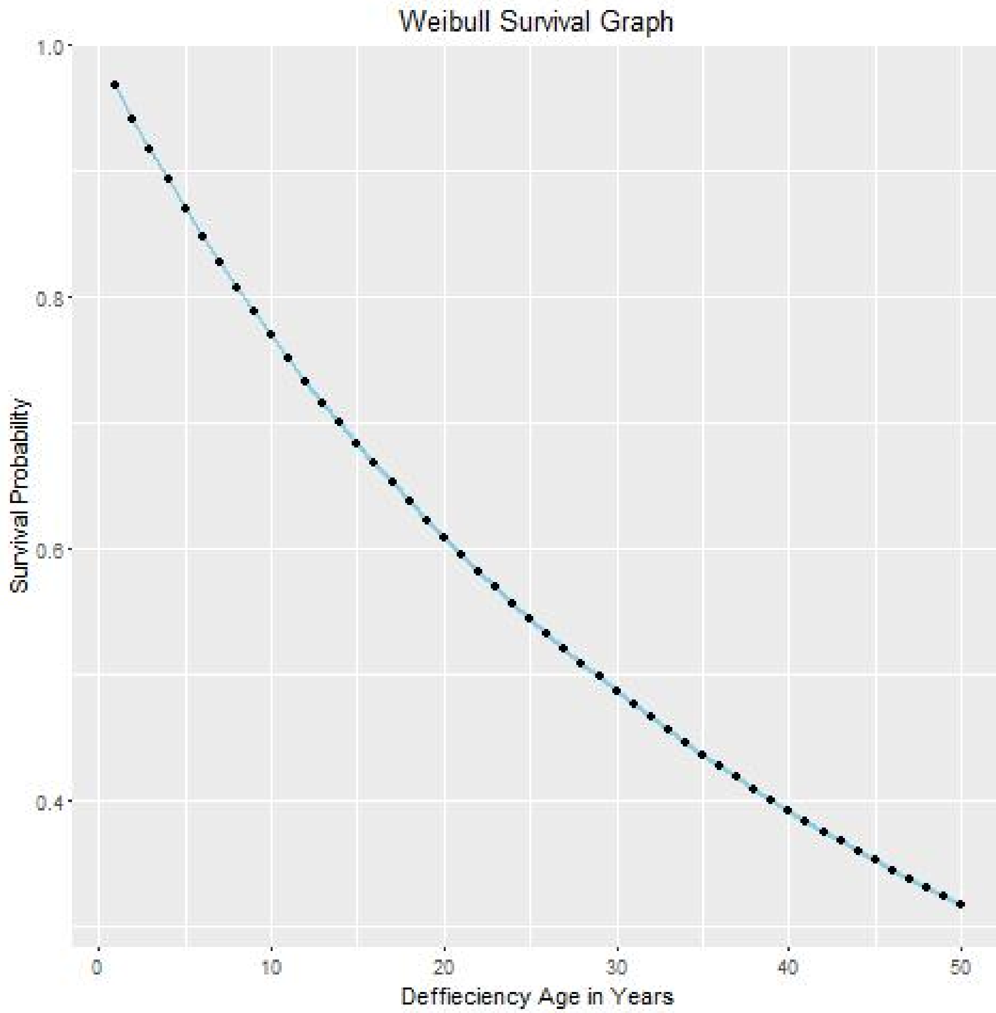

Figure 5: Survival probability Plot- with weight

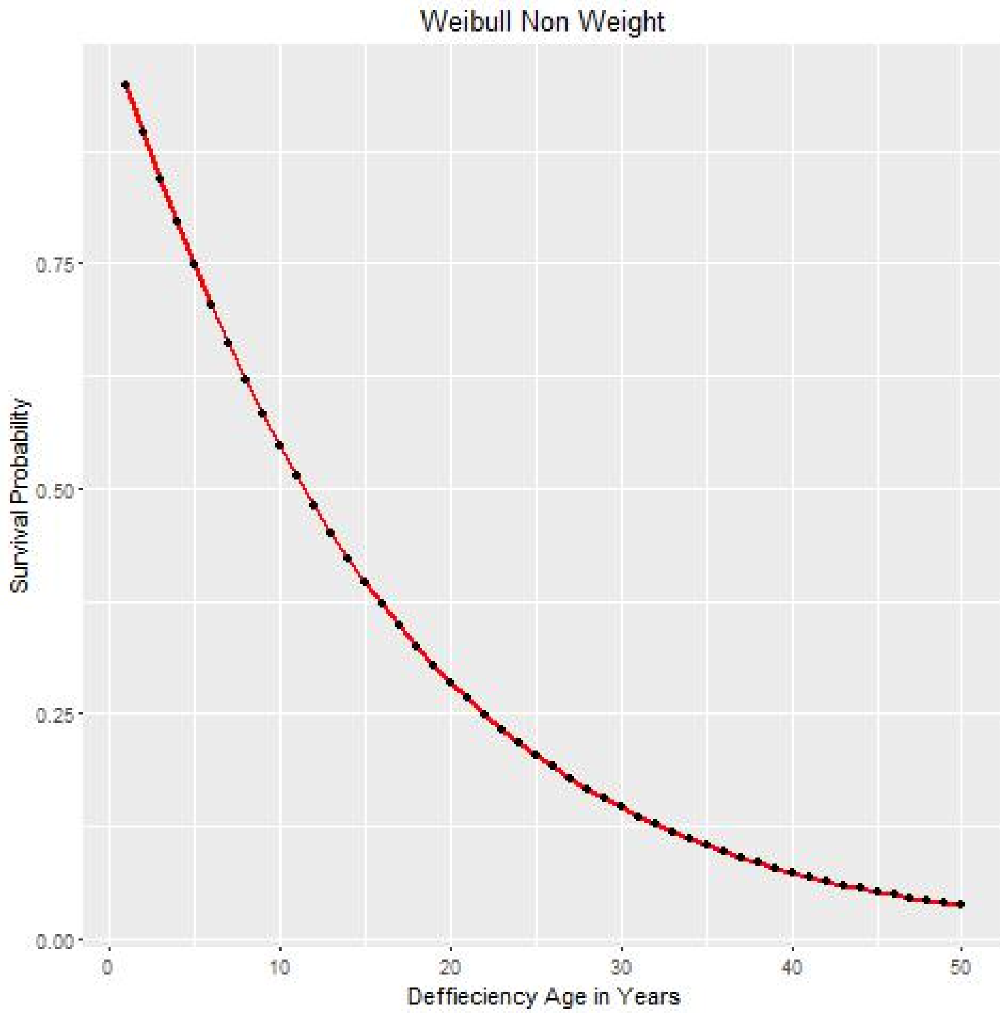

Figure 6: Survival probability Plot- Excluding weight

Prediction of survival probability for Weibull model(with weight) is 54% and sround 20 % without considering weights(Quantity)

Folder Structure

- proj

  Code - Folder Consists of Scripts

  Data - NBI Data file (DataClean.csv , Survival.csv)

Resources - Libraries required for running scripts

Output - results and images of Analysis

- CAIT_Bridge.Rproj

Code files are given below,

- 1package.r script installs packages required for the script.
- Common . r script loads packages required for the script.
- SolutionScript.r script cleans data and builds the model.
- **Only run Run.R file to run all scripts at once.**

4 Conclusion

I have implemented Weibull model for condition state 1 with and without weights as well as using Kaplan-Meier estimate of survival probabilities. All results validates Weibull model's prediction at t=25.

5 References

1. Azam Nabizadehdarabi.Reliability of Bridge Superstructures in Wisconsin .