An analysis of what factors have relationship with bridge condition index

Ruizhe Huang - 1006444331

December 17, 2021

Introduction

A paper indicates that concrete, steel are the most common bridge materials. Concrete versus steel has a slightly better condition, lower durability, and better longevity, rate, pattern, and equivalent performances (Farhey, 2017). The Ministry of Transportation has trained inspectors to review and rate each bridge component when inspected in detail every two years in Ontario. These ratings are used to determine the current value of the bridge on the Bridge Condition Index(BCI). The Ministry of Transportation uses the BCI to plan maintenance and repairs.
The research question is **What variables could have a relationship with the current bridge condition index(BCI)**. The importance of the research is that it can give some insights for the designers into what variables are most likely to result in the bridge needing maintenance more often because future inspection and maintenance should also be considered at the design stage. (Gerard, Ryall & Harding, 1990)

Terminology

The current BCI is the bridge condition index in 2021, with higher BCI means maintenance is less often required. The age of the bridge is calculated from the year built to 2021. Length and width are the Deck/Culverts length and the total width of the bridge measured in meters.

Methods

Variable selection/ Model selection

There are nine variables in the full model, where *current BCI* is the response variable. I keep some of the critical variables such as *Age* and *Material* based on the practical rationale and background research.
I will select the model(a subset of the full model) by partial F-test if p-value > 0.05 then the reduced model is better, and choose the model with higher adjusted R^2, lower AIC and BIC.

Model Validation

I randomly divided the original dataset into two independent datasets from the sample: a training dataset, which accounts for 80% of the sample, and a test dataset, which accounts for 20% of the sample. I will use the training dataset to create the models and use the test dataset to validate the model.
To validate the model, we hope to see similar estimated regression coefficients and adjusted R^2 between train and test dataset, the same predictors appearing significant in the train and test dataset. We do not want to see any new or worse model violations in the test dataset. If they meet these conditions, then the model is likely validated.

Model Violations and Diagnostics

If there is no pattern between actual current BCI(response) and fitted current BCI(response) with fanning pattern, condition 1 likely fails. If there is a non-linear pattern between predictors, then condition 2 likely fails. In these cases, by finding a function that undo the pattern, do the transformation or remove some predictors to improve the conditions. If we meet condition 1 and condition 2, then we can plot the residual plots to check the linearity, constant variance and uncorrelated error. If there is a discernible and fanning pattern with clusters of residuals separate from the rest, these assumptions likely fail. We can also plot the normal Q-Q plot to check the normality assumption. If there is a severe deviation from the normal Q-Q line, the normality assumption likely fails. In these cases, we could see whether the plots improved after transformation. For the testing dataset, I will apply the same transformation.
After all, if we found variance inflation factor(VIF)>5 after transformation, there may be severe multi-collinearity. We can remove the variables with the highest VIF one by one until all VIF<5. If $h_{ii} > 2\frac{p+1}{n}$, where p is the number of predictors. Then it will be a leverage point. As $r_i = \frac{e_i}{s\sqrt{1-h_{ii}}}$, if $|r_i| > 4$, then it will be a outlier point. If $D_i > F_{p+1,n-p-1}(50\%)$, then it will be a influential point. However, if they are reasonable, I will not remove them and acknowledge their existence in the limitation. Eventually, we will get the final model.

Results

Description of Data

Table 1: Compare the numerical summaries of the numerical variables in different dataset

Variables	Mean	Median	Min	Max	IQR
Original Current BCI	77.63	74.62	13.33	100	11.21
Cleaned Current BCI	77.95	74.77	13.33	100	11.37
Train Current BCI	77.98	74.80	13.33	100	11.23
Test Current BCI	77.86	74.70	29.50	100	11.99
Original Bridge's width	11.39	10.50	0	114.91	9.77
Cleaned Bridge's width	11.72	10.97	0	114.91	9.93
Train Bridge's width	11.8	11.0	0.1	114.9	9.9
Test Bridge's width	11.43	10.87	1.80	49.79	9.79
Original Bridge's length	64.28	40.20	1.20	24400.00	46.97
Cleaned Bridge's length	64.85	40.34	3.20	24400.00	47
Train Bridge's length	66.32	40.12	3.20	24400.00	46.81
Test Bridge's length	58.97	42.45	6.00	999.00	47.38

From table 1, it is clear that there is no significant difference for all numerical variables after removing the missing values. Therefore, we have reasonable ground to remove the missing values of the chosen variables. All the statistics between the four datasets except the test dataset are very similar. Some of the statistics for the test dataset are much different from those in the other datasets. For example, the sample mean for Bridge's length in the test dataset is 58.97m. However, in other datasets, the sample mean Bridge's length is around 64-67m. It implies a foreseeable drawback: the test dataset may not be very similar to the training dataset, thus may impacting the model validation.

Figure 1: EDA plots including the boxplots, barplot, scatterplots for traning and testing set

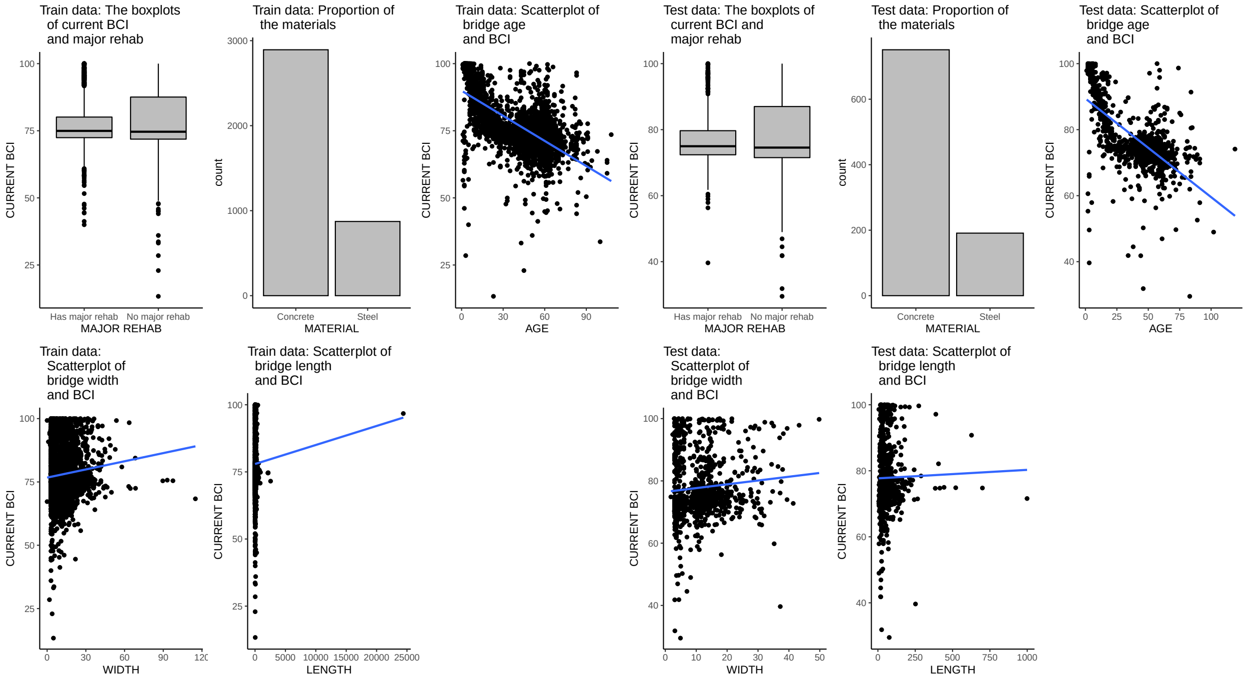


Figure 1 is the graphical summaries for the training and test dataset. The boxplot of current BCI and major rehab and the barplot of the materials' proportion in the training and test dataset are very similar. The current BCI shows a negative linear relationship with the bridge's age in the training and test dataset. The current BCI shows a positive linear relationship with the bridge's width in the training and test dataset. However, the relationship between the current BCI and the bridge's length is unclear as several potential problematic points in the training and test dataset may affect the regression line.

Process of Obtaining Final Model

For the full model, condition 1 and 2 is reasonably satisfied based on the method. However, the residual plots are not entirely satisfied. It may violate the linearity and constant variance assumption. Moreover, the normal Q-Q plot is a little bit heavy-tailed. Normality assumption may need to be improved. Thus, we may need a transformation to improve the violations.

I performed the transformation on both predictors and responses. However, the violation is not improved at all, making the model more complicated. Therefore, I decided to transform only the square root response as shown in the figure 3 in the appendix, making the model easier to interpret.

After transforming the full model, I first select the manually selected model by the significance. Thus all the predictors are significant variables(age and material) in the manually selected model. Moreover, I used the automated selection method to select the model. The model selected by stepwise selection contains one predictor that is not very significant.

Since the p-value of partial F-test for the automated selected model and transformed full model is 0.46. Also, for the manually selected model and transformed model, the p-value is 0.22. Thus, the results of the partial F-test for both models suggest that the reduced models are better than the original. It means that the automated and manually selected models are better than the transformed full model.

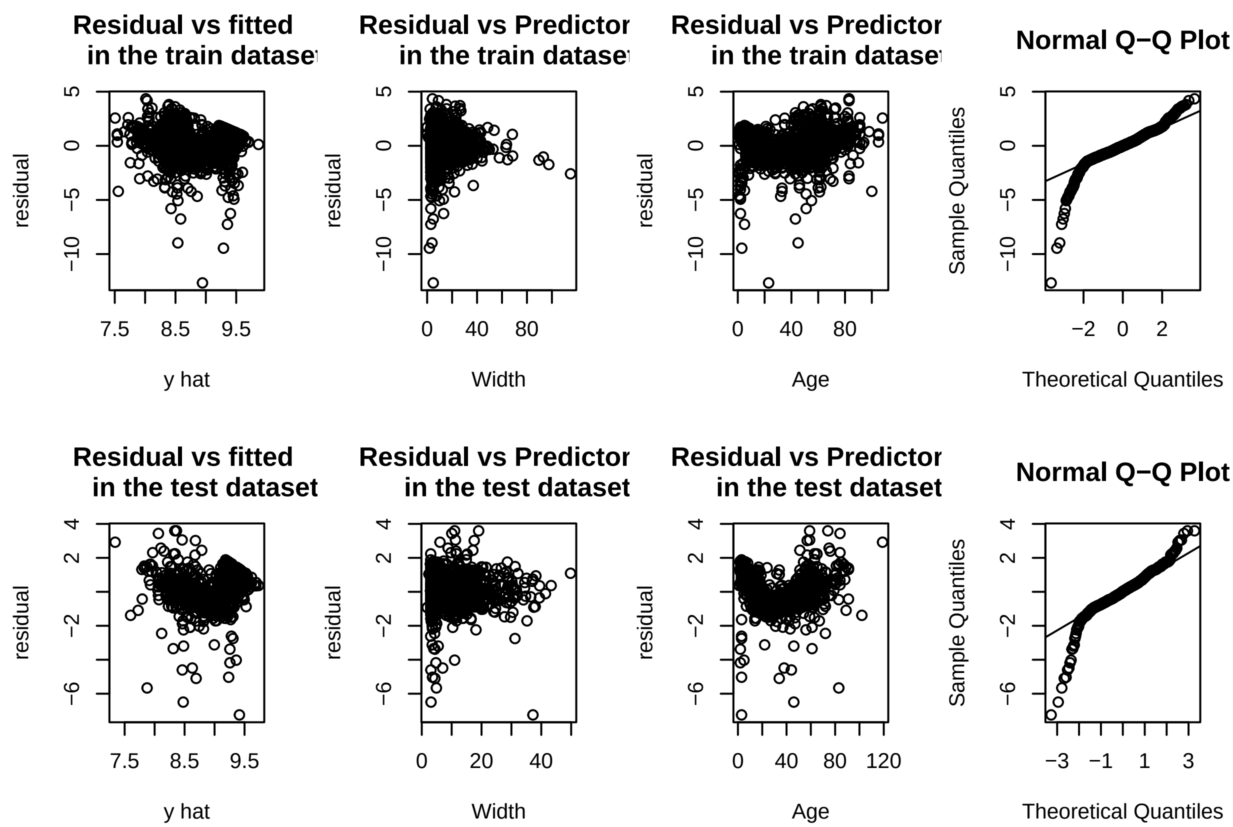
Furthermore, the VIF of all predictors is smaller than 5 in both automated and manually selected models, implying no severe multicollinearity among the predictors.

Table 2: Compare the models by adjusted R^2, AIC, AIC corrected and BIC

Model	Adjusted R^2	AIC	AIC corrected	BIC
Transformed full model	46.16%	4144.52	4144.63	4231.79
Transformed automated selected model	46.17%	4141.11	4141.18	4209.68
Transformed manual selected model	46.14%	4142.31	4142.37	4204.65

From table 2, the adjusted R^2, AIC, AIC corrected, and BIC is very similar among the transformed models. Based on the adjusted R^2, AIC and AIC corrected, the transformed automated selected model is preferred. However, based on the BIC, the transformed manual model is preferred.
I found 111 leverage points in the automated selected model and 73 leverage points in the manual selected model. The automated has more leverage points than the manual selected model. Moreover, both models have 20 outliers and none of the influential points. Based on the analysis of problematic points, there are fewer leverage points in the manual selected model. Thus, a manually selected model is preferred in terms of the leverage points.

Figure2: Model Violations and Diagnostics plots include the residual plots and normal Q-Q plot



Goodness of final model

Since the adjusted R^2, AIC, AIC corrected, BIC, the model assumptions and the model validation are very similar among all models. Thus, the final model is the manually selected model. It has fewer leverage points, and all the predictors are significant and are important based on the literature.
Moreover, I checked condition 1 and condition 2 as shown in the appendix. They are reasonably satisfied. From figure 2, the residual plots of the final model are not satisfied in both train and test dataset. There is a discernible linear pattern in the residual versus fitted plot, and there is a fanning pattern in the residual versus the predictor's plot. Therefore, the linearity and constant variance assumption may not hold even after transforming the response or both response and predictors. I also checked the normality assumption is not improved in the final model even though I have tried transform on response or both response and predictors, it is still heavy-tailed. However, it is hard to get the perfect normality. Since from the background research and the partial F-test, the predictors are all important. I will leave the model as it is.
I validate the models in the test dataset, the regression coefficients and the adjusted R^2 of final model between train and test dataset are similar. However, the width was significant in the training dataset. However, it appears non-significant in the test dataset. There are no new or worse model violations in the test dataset. Furthermore, there are no influential points in all training and testing datasets models, as everything else is very similar in the train and test dataset. Thus, it is validated.

Discussion

Final Model Interpretation and Importance

Table 3: The summary table of the final model

term	estimate	std.error	statistic	p.value
(Intercept)	9.5457814	0.0263330	362.503251	0.0000000
MATERIALSteel	-0.0589399	0.0171127	-3.444214	0.0005790
REGIONEastern	-0.1162739	0.0210132	-5.533362	0.0000000
REGIONNortheastern	-0.0697158	0.0206815	-3.370921	0.0007567
REGIONNorthwestern	-0.0621612	0.0283678	-2.191257	0.0284944
REGIONWest	-0.0651251	0.0198453	-3.281637	0.0010415
`MAJOR REHAB`No major rehab	-0.0902664	0.0157707	-5.723694	0.0000000
WIDTH	0.0053378	0.0009120	5.852659	0.0000000
AGE	-0.0179649	0.0003361	-53.444948	0.0000000

From table 3, the estimated final model is:

$$
\begin{aligned}
Current\ BCI^{0.5} = {} & 9.5457814 - 0.0589399 * I(Material = Steel) \\
& - 0.1162739 * I(Region = Eastern) \\
& - 0.0697158 * I(Region = North\ eastern) \\
& - 0.0621612 * I(Region = North\ western) \\
& - 0.0651251 * I(Region = West) \\
& - 0.0902664 * I(Major\ Rehab = No\ major\ rehab) \\
& + 0.0053378 * Width \\
& - 0.0179649 * Age
\end{aligned}
$$

For example, for the numerical variable Age, by holding other predictors constant, as the age of the bridge increase by 1 year, the average square root of current BCI will decrease by around 0.02. For the categorical variable Material, by holding other predictors constant. The steel bridges, on average, have less around 0.06 in current BCI than the concrete bridges. Thus, the result suggests that the current BCI would drop with

the ageing bridge.
Moreover, the result suggests that the concrete has a slightly better condition than steel. These results agree with the literature in background research. Furthermore, based on the final model, the material, region, whether it has major rehab, width and age have a relationship with the current BCI.

Limitations of Analysis

As the final model somehow violate the linearity and constant variance assumption, many decisions depend on the assumptions holding. So it's possible that the final model isn't necessarily the best one. Moreover, the p-value may not that reliable as the normality assumption is not that satisfied.
Even though the final model has the least leverage points in the candidate models, it still has 73. Many leverage points can potentially affect the estimated regression line.

References

1. Government of Ontario, M. of T. (2013, October 25). *Bridge repairs.* Ministry of Transportation. Retrieved December 17, 2021, from http://www.mto.gov.on.ca/english/highway-bridges/ontario-bridges.shtml.
2. Farhey, D. N. (2017). *Structural Performances of Bridge Materials in the U.S. National Bridge Inventory 2013.* Structural Engineering International: Journal of the International Association for Bridge and Structural Engineering (IABSE), 27(1), 101–113. https://doi.org/10.2749/101686617X14676303589039
3. Gerard, P., Ryall, M., & Harding, P.J.E. (Eds.). (1990). *Bridge Management: Inspection, maintenance, assessment and repair (1st ed.).* CRC Press. https://doi-org.myaccess.library.utoronto.ca/10.1201/9781482267105
4. Schaper, D. (2017, August 1). 10 years after bridge collapse, America is still crumbling. NPR. Retrieved December 17, 2021, from https://www.npr.org/2017/08/01/540669701/10-years-after-bridge-collapse-america-is-still-crumbling.

Appendix

Figure 3: BoxCox transformation

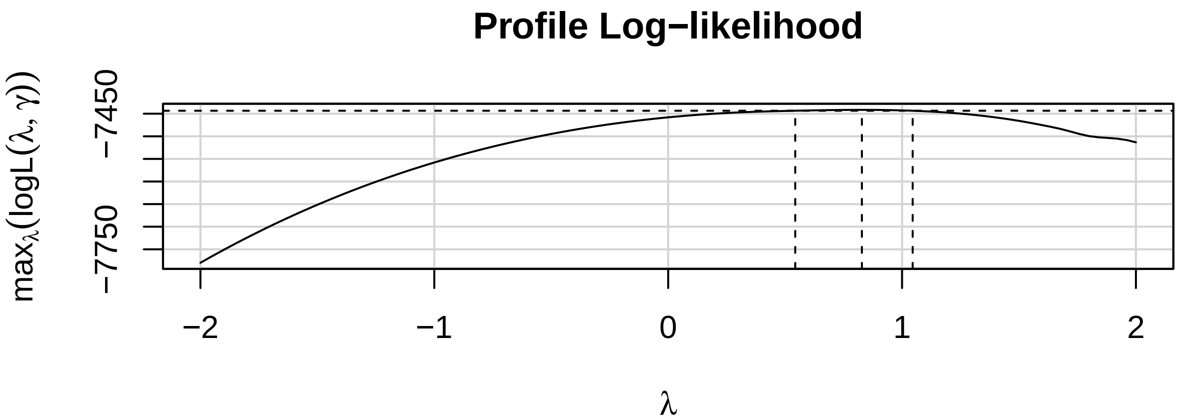


Figure 4: Y Versus Y hat(Condition 1) for final model

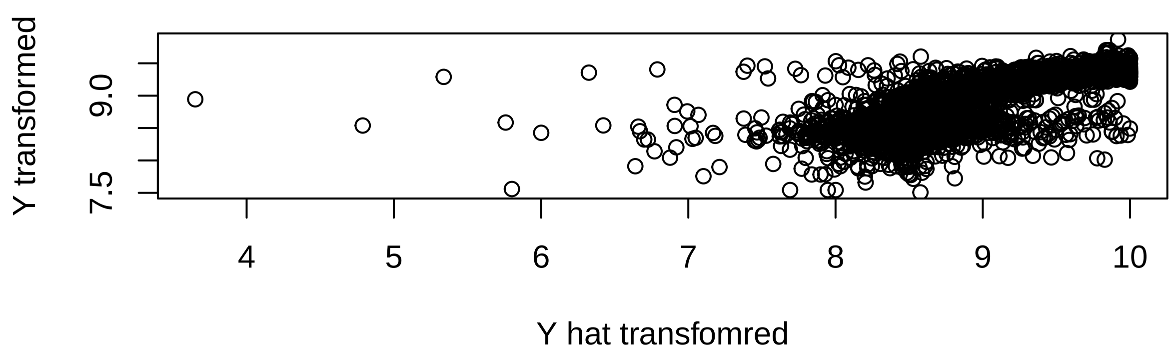


Figure 5: Predictors Versus Predictors(Condition 2) for final model