Evaluation of Probability that a Bridge Collpases in Omaha, NE

Tyler Davis

Table of contents

Project Description

You are a data scientist for a mid-sized business, in a small group of 3-4 data scientists. You've been tasked with creating a report evaluating a scenario for your business. Your colleagues will also be evaluating the same scenario, and your reports will be used in aggregate to determine a consensus (or lack thereof) on the company's action. The reports will also be used to inform downsizing that is rumored to be coming - you want to ensure your report is better than your peers so that you aren't as easy to cut.

You may talk to your peers who are assigned the same scenario, but you do not want to collaborate too closely, lest you both become targets of the rumored layoffs.

I've scaffolded this report for you to make this process easier - as we talk about different sections of a report in class and read about how to create similar sections, you will practice by writing the equivalent section of your report.

The basic steps for this task are as follows:

- Identify the research question from the business question
- Identify data set(s) which are (1) publicly available (you don't have a budget to pay for private data) and (2) relevant to your task
  - (HW Week 6) Document your data sets in `draft-data-doc.qmd`
- Conduct a statistical analysis to support your answer to your research and business questions
  - Write a methods section for your business report corresponding to your statistical analysis
  - (HW Week 9) Draft of results section of business report with relevant graphics/visual aids in `draft-results.qmd`
- Write your report
  - (HW Week 10) Draft of Intro/Conclusion sections in `draft-intro-conclusions.qmd`

  - (HW Week 11) Draft of Executive summary section in `draft-exec-summary.qmd`
- Revise your report
  - (HW Week 12 – not turned in) Revise your report
  - (HW Week 13) - Rough draft of report due. Create one or more qmd files for your report (you can overwrite or delete intro.qmd and summary.qmd), include the names of each file (in order) in `_quarto.yml`. You should use references (edit references.bib and use pandoc citations). Make sure your report compiles and looks reasonable in both html and pdf.
  - Develop a presentation to go along with your report (Week 13). Create slides for your report using quarto.
- Peer revise reports
  - Peer revise reports
  - (HW Week 14) - Make edits to your report from comments received from peer review
- Final report & presentation due

1 Introduction

Have you ever been driving over a bridge and wondering how stable it is ? This is a question that has created a lot of buzz recently especially with the state of the infrastructure in the United States. There is a data set that contains all of the information regarding federal, state, and local bridges in the United States, and it is the National Bridge Inventory. This data set contains all of the information needed to estimate and analyze the risk of a bridge collapsing. The data set also contains information to focus on individual state or even latitude and longitude data to focus on certain cities. This brings us to the question at hand and that is, what is the chance that you are driving over a bridge at risk of collapse in Omaha, NE. This question can be answered usign the National Bridge Inventory data set and then focusing in on the Omaha, NE area. This will allow us to gain a greater understanding of the bridge condition in the Omaha area and hopefully allows us a greater understanding of the bridges that are at most risk of collapse in Omaha, NE.

2 Methods

When thinking about how to conduct the analysis on percent chance a bridge collapses in Omaha, NE, we first need to find a data set that contains all of the information that we need. We would like to find information on how the bridge is structurally, how old it is, and how much traffic the bridge has, for example. The data set that was found is the NTAD National Bridge Inventory, which contains a majority of the information stated above and much more. This data set ,described in the data documentation, allows us to attempt to figure out the main question at hand what is the percent chance that a bridge will collapse in Omaha, NE. This data set was filtered to contain all bridges in Iowa and Nebraska by state code, it was then filtered again to focus in on the Omaha area. This was done by filter statements in R, using state codes, latitudes, and longitudes. This subset of the NTAD data was then used for the analysis.

?@fig-omaha-map1 above shows how the data contains all of the bridges in the Omaha, NE area. This is crucial as our task is limited to this area. This map also gives us a broader idea of how bridges are distributed across Omaha, NE allowing us to assess the risk for each area. The next step in our analysis is to further understand what variables within the data set are useful to us. This decision was made by analyzing the code book and understanding which variables give us key insight into the structure, age, traffic, and condition of the bridges. These variables will hopefully lead us to be able to draw a conclusion that meet our goal of predicting a bridge collapse.

There are 130 variables in the NTAD Bridge Inventory data set, some of variables that we will be using are: bridge condition, substructure and superstructure condition, average daily traffic, and year built. These variables give us a good picture of what it means to be a stable and healthy bridge and also what it means to be a bridge on the verge of collapse. We next analyzed the variable bridge condition, which is a variable that describes the overall condition of the bridge, which has rating of "P" for poor, "G" for good, and "F" for fair. Bridge condition rating is determined by, " the lowest rating of National Bridge Inventory (NBI) condition ratings for Item 58 (Deck), Item 59 (Superstructure), Item 60 (Substructure), or Item 62 (Culvert). If the lowest rating is greater than or equal to 7, the bridge is classified as Good; if it is less than or equal to 4, the classification is Poor. Bridges rated 5 or 6 are classified as Fair.", Weseman (n.d.). This allows us to looks at the overall condition of the bridge by looking primarily at one variable.

In (**fig_brid?**) we can see the overall breakdown of how the bridge condition variable. When first looking at this graph we can see that a majority of the bridges in Omaha are in good or

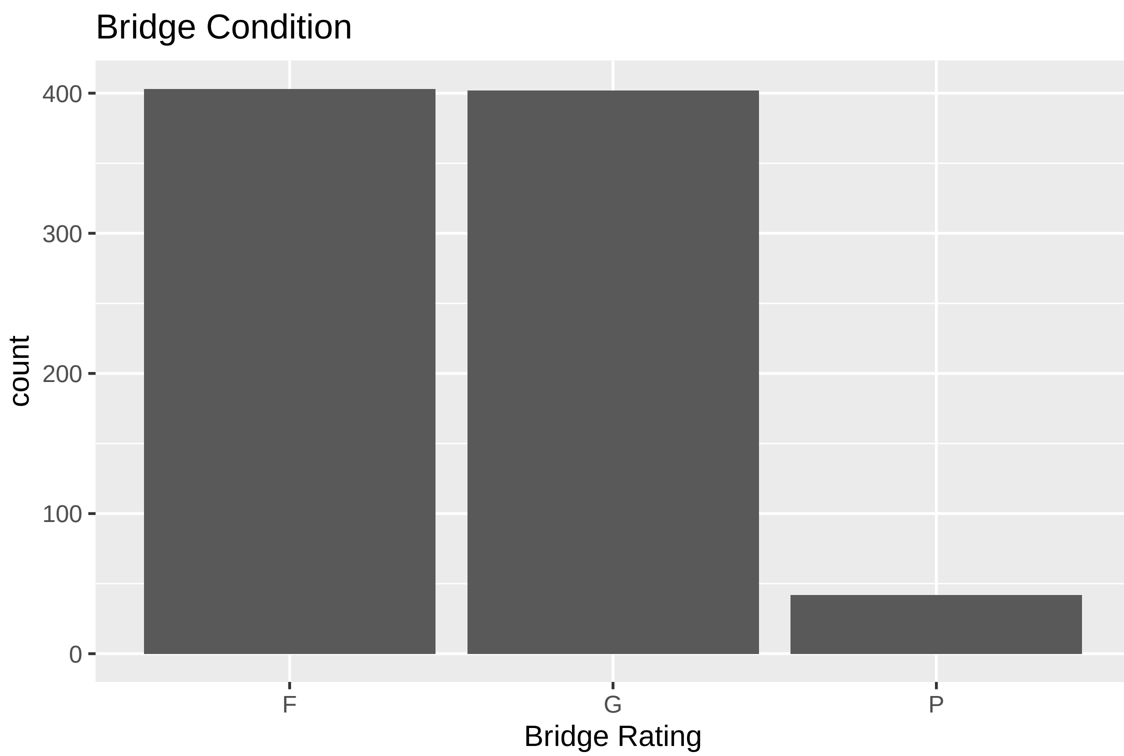

Figure 2.1: Bridge Condtion Ratings Across Omaha, NE

fair condition. This is great but, we now want to take a closer look into the bridges that have a poor rating.

?@fig-omaha-map above shows us the distribution of bridges across Omaha, NE, that have a bridge condition rating of poor. This gives us a good understanding of the bridges at most risk of collapse right now. We can see in **?@fig-omaha-map** that some of the bridges that are at most risk of collapse are in downtown Omaha, which could lead to major economic and logistical issues. The economic impact of a bridge collapse is a important one to consider as it can impact the severity of the incident. Another factor that might play a role on if a bridge collapses would be how much traffic a bridge takes in per day. We then created another subset of data from the bridges that have poor condition, and added that the average daily traffic must be greater than 5,000 cars and that is must have over two lanes. This will give us a good picture of major bridges in Omaha, NE that are at risk of collapse that could impact that most people.

?@fig-omaha-map2 above show us the distributed of bridges with a poor condition that undergo high traffic scenarios each day. These bridges could possibly be at higher risk for collapse due to the increase of traffic. When looking at the map above we can also see that none of these bridges are located over bridges which may mitigate the damage and economic impact of a collapse. We can also see that of the bridges that are at a poor rating and have high traffic none are downtown Omaha, which again could limit the economic impact of a collapse.

Statistical analysis was then conducted, first by analyzing the descriptive statistics of both the bridge condition subset and the high traffic subset. This was then extended by finding the percent of bridges that are at a poor condition and poor condition with high traffic respectively. This gave us a basic understanding of what the overall condition was currently. We then had to create a method of prediction, which was done by fitting a linear model. The missing values in the data set were a concern which lead to imputation methods being used by way of the "VIM" package in R. This then allowed us to analyze the data using a linear model creating a method of prediction of when a bridge would collapse in Omaha, NE.

3 Results

The first step to the formal analysis is to find the descriptive statistics from the bridges that are in poor condition.

Table 3.1: Descriptive statistics of bridges in poor condition

	mean	sd	n	median
YEAR_BUILT_027	1954.714	20.072	42	1957
TRAFFIC_LANES_ON_028A	2.500	0.994	42	2
ADT_029	12244.024	19761.961	42	2027
DECK_COND_058*	3.485	1.176	33	3
SUPERSTRUCTURE_COND_059*	3.545	1.148	33	3
SUBSTRUCTURE_COND_060*	3.606	0.827	33	4
CHANNEL_COND_061*	3.156	1.347	32	3
CULVERT_COND_062*	1.000	0.000	9	1
STRUCTURAL_EVAL_067	3.595	1.515	42	4
SCOUR_CRITICAL_113*	4.031	1.121	32	4
BRIDGE_CONDITION*	1.000	0.000	42	1

From Table 3.1 above we can see that the average year of the bridges built in Omaha, NE is 1955. We can also see fro this data that all of our structural condition ratings are low as the average is below 4 which tells us that the structural condition of the all of these bridges is below par. We can also see that the average daily traffic(ADT_029) median is 12240 cars per day. We will now look at the descriptive statistics for the high traffic data subset.

Table 3.2: High Traffic Bridges with Poor Rating

	mean	sd	n	median	se
YEAR_BUILT_027	1969.167	17.198	12	1965	4.965
TRAFFIC_LANES_ON_028A	3.917	0.515	12	4	0.149
ADT_029	38159.167	20251.467	12	34353	5846.095
DECK_COND_058*	1.000	0.000	9	1	0.000
SUPERSTRUCTURE_COND_059*	2.111	1.167	9	2	0.389
SUBSTRUCTURE_COND_060*	1.222	0.441	9	1	0.147

Table 3.2: High Traffic Bridges with Poor Rating

	mean	sd	n	median	se
CHANNEL_COND_061*	2.400	0.894	5	3	0.400
CULVERT_COND_062*	1.000	0.000	3	1	0.000
STRUCTURAL_EVAL_067	3.833	1.899	12	4	0.548
SCOUR_CRITICAL_113*	1.800	0.837	5	2	0.374
BRIDGE_CONDITION*	1.000	0.000	12	1	0.000

These bridges displayed in Table 3.2 are at risk of collapse and they also have the highest traffic flow among at risk bridges. These bridges are further at risk due to the fact that the average daily traffic is 38160 which is very high compared to most bridges in Omaha, NE. We can see this in the descriptive statistics as the mean of the structural conditions is below 3 which is lower than if the bridges were just at poor condition. We can also see that the average year built of the bridges is slightly newer at 1969. Finally, we can see that median of the average number of lanes is 3.92 which is much higher than the previous subset. The next step is to look at the proportion of bridges in Omaha that have a poor condition and that have a poor condition under high traffic.

Table 3.3: Proportion of Bridges at Risk in all of Omaha, NE

Bridges at Risk	High Traffic Bridges at Risk
0.0496	0.0142

From Table 3.3 we can see that the percent chance that a bridge in Omaha, NE is in poor condition is 4.96 %. Along with that, the percent chance that a bridge in poor condition has high traffic is 1.42%. This tells us that we have a minimal chance of coming across a bridge that is at risk of collapse by structural ratings. This was a solution primarily based on descriptive statistics, we will now create a statistical model to predict a collapse of a bridge. To do this task we must create a collapse variable. This will be done by creating a binary variable that depends on if the bridge condition is poor and is the structural evaluation of the bridge in less than 4. These two variables give us the best insight into if a bridge is at risk of collapse, hence they will be used to create the variable. The data that we have also contains many “NA” values, to take care of this issue will will impute the data, or estimate values for the “NA” values based on information in the data set. The new imputed data set is what we will use to form our model.

The model we will use will be the following:

$$\log\left(\frac{P(\text{collapse\_risk}=1)}{1-P(\text{collapse\_risk}=1)}\right)=\beta_0+\beta_1\cdot\text{DECK\_COND\_058}+\beta_2\cdot\text{OPERATING\_RATING\_064}+\beta_3\cdot\text{YEAR\_}$$

Table 3.4: Regression Model Summary

term	estimate	std.error	statistic	p.value
(Intercept)	76.89777	28.38119	2.709	6.739e-03
as.numeric(DECK_COND_058)	-1.93498	0.33590	-5.761	8.380e-09
OPERATING_RATING_064	-0.05982	0.02013	-2.971	2.965e-03
as.numeric(YEAR_BUILT_027)	-0.03376	0.01462	-2.309	2.092e-02

The model above is a generalized linear model, as our response variable is non - normal (Binomial), this is done using the logit link function. From this model we will then get our regression coefficient estimates to predict to use for prediction when the next bridge will collapse.

From Table 3.4 we can see that all variables are significant (p.value $< .05$) and this should produce accurate results in predicting our response variable on testing data. We will now predict the response on a test data set using this model.

From this table produced we can see that we get a majority of our predicted values to be 0 and only a few to be 1. This represents the relationship between the bridges that are at risk of collapse that bridges that are not as risk of collapse. We can see that we are getting similar numbers in our prediction to the values contained in the data set. We will now test the accuracy of the model.

Table 3.5: Accuracy of Regression Model

Accuracy
0.9762

From Table 3.5 we can see that our model has a 97.6% accuracy in predicting our defined collapse variable. Which shows us that we can predict with good accuracy if a bridge is at risk of collapse in Omaha,NE.

4 Conclusion

When looking back at the opening question, do you ever think of how stable a bridge is when you are driving over top of it, the answer would be hopefully not. This is because in our analysis we have found that 5% of the bridges in the Omaha,NE area have a bridge rating of “poor”. We have also see that 1.4 % of bridges are considered high traffic bridges and also have a “poor” rating. This tells us that in Omaha, NE if you took a random sample of 100 bridges, 5 would come out with a “poor” rating. This is a major issue highlighting the lack of infrastructure maintenance and repair in our country. This is a future concern and with bridges and roads only getting older it is time to put more money and research into how to maintain our aging infrastructure. This lead to further investigation using the National Bridge Inventory data set which lead us to create a collapse variable based on important variables and then try to predict, from the data we have, which bridges are at most risk of collapse. The model we developed predicted our collapse variable with 98% accuracy showing that with this data set there is the ability to predict which bridges are at most risk of collapse. The predicted percentage of bridges that were at risk of collapse was the a very similar percentage to the percentage in our actual data.

Some issues with the data set are that there are a lot of missing values this was difficult to handle especially with the amount of variables that we have in our data set. This forced us to use imputation methods to estimate our missing values, this allowed us to use much more of our data than if we did not do this. Some future implications of this analysis and data, would be taking similar approaches to see which bridges are in the worst condition. This will allow governments to allocate resources to the proper locations creating more understanding in which bridges need the most work.

5 Summary

This analysis examines the risk of bridge collapse in Omaha, NE, using data from the NTAD National Bridge Inventory. The data set was filtered to focus on bridges within Nebraska and more specifically Omaha. The results show that around 5% of bridges in Omaha, NE are in poor condition hence at risk of collapse. We also see that bridges that are considered high traffic have a 1.4% percent chance of being at poor condition. This is all collaborated by our linear model which uses a formulated collapse variable to select which bridges are at risk of collapse. This model has been shown to be 98% accurate and predicted a similar percentage of bridges at risk of collapse as the data showed.

Given these findings, it is clear that Omaha, NE needs to invest in infrastructure improvement. Local and state government need to take action by improving overall bridge structural condition starting with the bridges in the highest risk category. The traffic patterns of the city should also be analyzed to understand which bridges pose the most risk to commuters on a daily basis. The public can also help this issue by creating awareness around the lack of care into the infrastructure of the city of Omaha.

References

"National Bridge Inventory (NBI)." n.d. https://tvar-hub-usdot.hub.arcgis.com/pages/national-bridge-inventory.

Weseman, William A. n.d. "Recording and Coding Guide."

A Data Documentation

Overview:

The data that I will be using is from the Department of transportation. This data set is the national bridge inventory collected annually by the DOT. This data was updated on 09-30-2024. This data provides information about all bridges that are on public roads and span over 20 feet across the United States. It contains information about bridge structure, bridge location, bridge rating, along with many other data points. I will be focusing on bridge location, bridge rating, along with daily traffic load.

Description listed on website:

"The National Bridge Inventory details bridges on public roads that include interstate highways, U.S. highways, state and county roads, and publicly accessible bridges located on Federal and Tribal lands. The database includes the location, description, classification, and condition of each bridge. The condition of the structural and bridge management elements of each bridge are cataloged for all bridges on the National Highway System (NHS).

The NBI is published annually but its publication location varies. The 2024 (most recent) NBI is published with a map viewer as part of the Bureau of Transportation Statistics National Transportation Atlas Database. The data set is also published through FHWA's Bridges & Structures site.

The National Bridge Inventory can support resilience analysis and detour planning efforts. It provides bridge location, design load and clearance, roadway geometry, rating, posting and structural evaluations, number of lanes, traffic counts, among other variables." "National Bridge Inventory (NBI)" (n.d.)

Detailed information:

This data is located on a csv file in excel. It contains 116 variables describing various aspects of bridges. I will be focusing my analysis on variables 1,4,5,21, 28-29, 57 - 67,109,113-115. These variables and all information regarding analysis and coding can be found in the code book, Weseman (n.d.).

Process:

This data will be used to analyze the question regarding the probability that a bridge will collapse in Omaha, NE. This is the reason certain variables will be analyzed and others will not be as much of a factor. The factors that I will use to make this decision will be location of bridge, amount of lanes, amount of traffic, bridge ratings, daily truck traffic, etc. All of this information and more can be found on the NBI website and data set.