pH of Soil Near Cattle

Sam Viet Huynh and Lauryn Harvey

03/20/2020

Introduction

The cattle industry in the United States is a $76.4 billion dollar business where 91% of beef cattle operations are family or individually owned [2]. If someone were to start their own cattle farm, it is important for them to consider all of the factors that involve caring their cattle.

The cattle industry relies heavily on the use of antibiotics to ensure the health of its cattle. However, the use of antibiotics can be a problem. Overuse of antibiotics can result in strains of antibiotic resistant bacteria thriving [1]. One paper's finding on antibiotic resistant bacteria in the meat industry "support for the adoption of guidelines for the prudent use of antibiotics in food animals and for a reduction in the number of pathogens present on farms and in slaughterhouses" [6]. If antibiotics are ineffective on certain strains of bacteria, how can we control the amount of bacteria present in our food?

Our goal is to investigate an alternative method for limiting the amount of antibiotic resistant bacteria on cattle farms. We will focus on the composition of the soil in the areas where cattle graze or feed. Previous research suggests the pH of the soil may be related to the amount of antibiotic resistant bacteria in the soil [3]. A pH of 6.5-7 is favorable for bacteria growth [4]. Then what factors affect the pH of the soil? Specifically, we will be investigating the relationship between the milligrams of ammonium per kilogram of soil, milligrams of nitrate per kilogram of soil, the location of the sample (grazing or feeding area), and the pH of the soil [5]. The reasoning behind why we chose nitrate and ammonium is that there is a chemical compound called ammonium nitrate that is used as a fertilizer. Specifically, we will propose a model for the pH of the soil based on the amount of ammonium (NH_4), nitrate (NO_3), and the location of the sample. Each of our chosen predictor variables can be reasonably managed by cattle farm owners. If these variables are found to yield a significant impact on the pH, then these variables can be controlled to mitigate the growth of bacteria without the use of antibiotics.

Methods

The data set we will be using was collected by USDA researchers in Kentucky. This USDA team published a paper using this data in 2014 which focused on the growth of different strands of anti-biotic resistant bacteria. The data set included a set of 87 samples of soil on cattle farms. From these samples 39 are from feeding areas while 48 are from grazing area.

For our linear model, we will use RStudio to estimate our regression coefficients. Our response variable is the pH level of the ground where the explanatory variables is the amount of ammonium and nitrate in the ground and whether or not the sample was collected in a feeding or grazing area. We will assess our model by evaluating if our residuals have a normal distribution, an expected value of zero, and a constant variance which are the assumptions of any linear model. Additionally, we will use plots to assess whether there is any collinearity between ammonium and nitrate. In forming our final model, we will look at the significance of partial regression coefficients along with the coefficient of determination.

Results

First, we will look at the plots of ammonium, nitrate and location against pH level. From the first scatter plot, we first notice that there isn't an apparent increase or decrease in pH as we increase the amount of ammonium. Note that the range of the pH levels are mostly between six and nine. There are two outliers where they are noticeably outside the average range. Additionally, we see that there was a lot of observations that were taken between zero and ten milligrams of ammonium.

From the second scatter plot, we see that there isn't a clear relationship between increasing

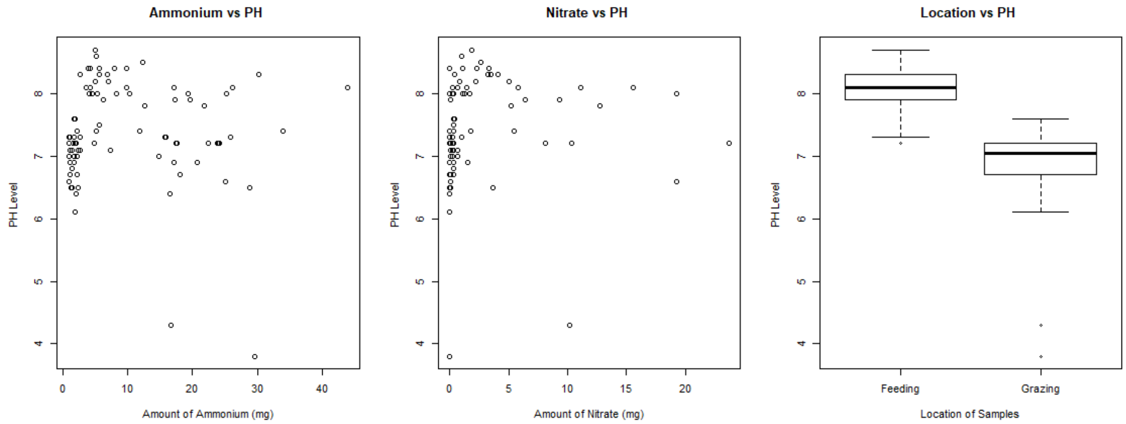

Figure 1: Ammonium, nitrate, and location vs pH

nitrate and level of pH. The range of pH is also between six and nine. There are also two outliers near the bottom of the graph, and the majority of the observations were found between zero and five milligrams.

In the boxplot, we see a clear difference of pH levels for feeding and grazing locations. Observations in feeding locations have noticeably higher pH level than the ones from grazing areas. We see one fairly close low outlier for feeding locations and two large outliers for grazing locations.

We should consider that there may be two distinct patterns within the first two scatter plots. Many of the observed data points appear to be following a positive linear pattern if restricted to lower measurements ammonium and nitrate. Higher measurements of ammonium and nitrate appear to be randomly distributed around some horizontal mean pH value.

Next, we plot predictor variables against one another to assess any possible collinearity.

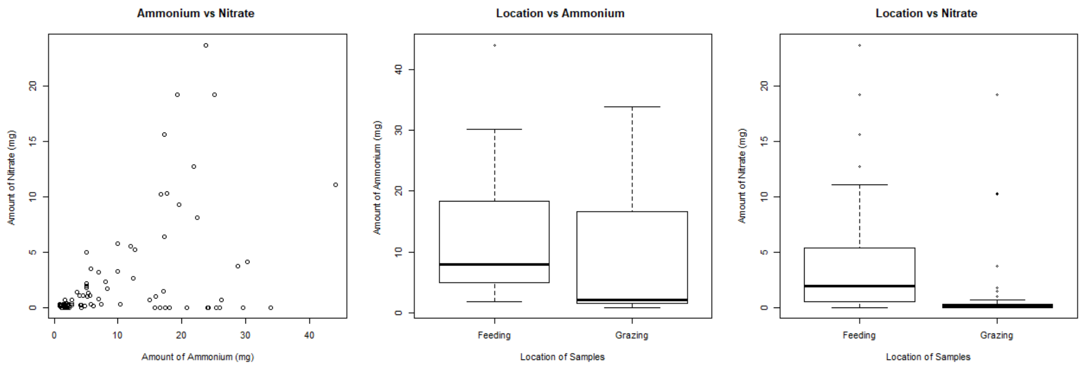

Figure 2: Plots of ammonium vs nitrate, location vs ammonium, and location vs nitrate

In the first plot picturing ammonium versus nitrate, we see a cluster of observations with low ammonium and low nitrate. We see that there is a pattern between an increase in ammonium and an increase in nitrate. As the measurement of nitrate increases the measurement of ammonium also appears to increase. However, this relationship does not appear to be linear across all of the collected observations. The variance between these two predictor variables is not constant.

In the second plot, we can see that the inter-quartile range of feeding and grazing locations appears roughly identical. The median ammonium measurement for the feeding location is noticeably higher than the ammonium measurement for the grazing location. We can see there is one outlier in the left boxplot with a feeding location that has a very high ammonium measurement.

The third plot pictures boxplots of feeding versus nitrate measurement. The nitrate measurements taken from feeding locations had a wider spread pictured via the inter-quartile range. On the other hand, nitrate measurements taken from grazing locations had a relatively smaller inter-quartile range. There appear to be about four outliers with high nitrate measurements for both feeding and grazing locations.

Then, the following tables are the numerical summaries for our variables.

Variable	Min.	Q1	Med.	Mean	Q3	Max
pH	3.800	7.000	7.300	7.393	8.000	8.700
Ammonium	0.900	2.10	5.40	10.01	17.15	44.00
Nitrate	0.000	0.100	0.300	2.493	2.250	23.700

Table 1: Numerical summary for pH, Ammonium, and Nitrate

	Feeding	Grazing
Number of observations	39	48
Mean pH	8.0	6.9

Table 2: Numerical summary for sample locations

From Table 1 we see that the bulk of measured pH values lies between 7 and 8. Additionally, the bulk of measured ammonium levels lie between 2 and 17 mg/kg of soil and measured nitrate levels generally lie between 0 and 2 mg/kg of soil. From Table 2 we see that the mean pH from feeding locations is more basic that the mean pH from grazing locations.

In RStudio, our linear model is coded with the following variables: pH where it's the measure of the acidity and basicity of the soil, NH4 is the ammonium in the soil, NO3 is the nitrate in the soil, and Trt is the dummy variable where it's either the cows grazing from the grass on the or being fed by the farmer with feed. So, we can say our full model is

$$pH = \beta_0 + (\beta_1 \cdot \text{NH4}) + (\beta_2 \cdot \text{Trt}) + (\beta_3 \cdot \text{NO3}) + (\beta_4 \cdot \text{NH4} \cdot \text{Trt}) + (\beta_5 \cdot \text{NO3} \cdot \text{Trt})$$

where Trt has the value of 1 for grazing location and 0 for feeding location. Note that β_{0-4} are the partial regression coefficients.

Now we will look at the summary for the full linear model. In the following tables, we will summarize the regression output and look at the variance inflation factors for each predictor variable.

Coefficients				
	Estimate	Std. Error	T-value	Pr(> |t|)
(Intercept)	8.205	0.1404	58.451	<2e-16 ***
NH4	-0.008	0.0107	-0.756	0.452
TrtGrazing	-1.172	0.175	-6.717	2.36e-09 ***
NO3	-0.015	0.018	-0.830	0.409
NH4: TrtGrazing	-0.008	0.014	-0.547	0.586
TrtGrazing: NO3	-0.017	0.031	-0.550	0.584

Table 3: Coefficients for linear model

Residuals				
Min	1Q	Median	3Q	Max
-2.769	-0.147	0.101	0.279	0.899

Table 4: Residuals for linear model

Residual Standard Error	0.541
R^2	0.582
Adjusted R^2	0.556
Degrees of Freedom	81

Table 5: RSE, R^2, adjusted R^2, and degrees of freedom of the linear model

Variance Inflation Factor					
NH4	Trt	NO3	NH4:Trt	Trt:NO3	Average VIF
3.216	2.243	2.238	3.878	1.875	2.690

Table 6: Variance inflation factors for our variables

As seen in Table 3, the "TrtGrazing" variable was highly significant while "NH4" and "NH3" were not. We can see that there is insufficient evidence to suggest an interaction between nitrate and the location of the sample. Additionally, there is insufficient evidence to suggest an interaction between ammonium and the location of the sample. From Table 4 we can see that the residuals are centered fairly evenly about zero which does not conflict with our regression assumption that the errors should be normally distributed about zero. From Table 5 we see that the full model explains roughly 58% of the variability in pH. From Table 6, we have the variance inflation factor for each coefficient in our model. There is no individual variable for which the variance inflation factor exceeds ten, which means that no individual predictor possesses severe multicollinearity. However, the mean variance inflation factor across all predictors exceeds two. This means that our model as a whole possesses a fairly high amount of multicollinearity.

Next, we will look at the residual diagnostics for our full linear model.

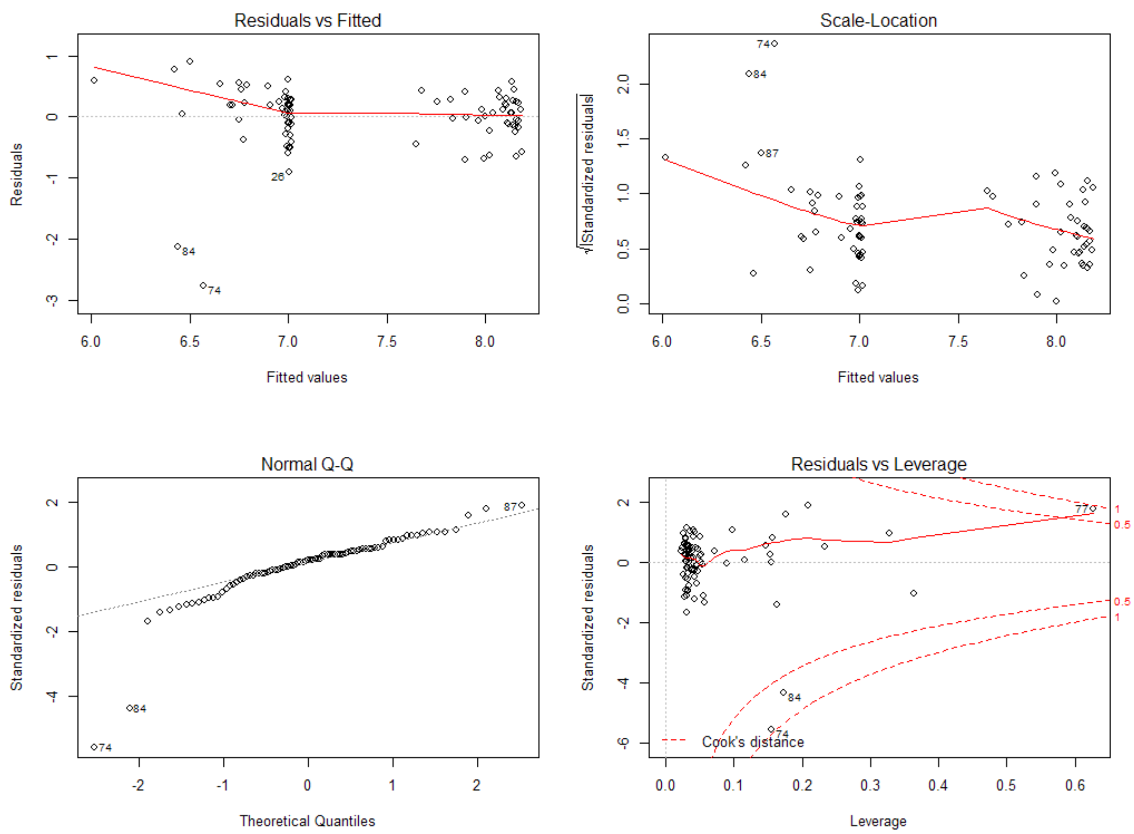

Figure 3: Diagnostic plots of our linear model

From figure 3, we can assess whether or not our model meets the assumptions of a linear model. First off, for the Residuals vs Fitted plot, we see that there is mostly a constant variance for the residuals. We see the residuals evenly spread out around the zero line. However, there are two main outliers that we notice that don't fit with the other points. These have index 84 and 74 in our dataset. Second, in the Scale-Location plot we see the exact same thing as from our previous plot. There seems to be a constant variance and the two outliers are the same. Third, from our Normal Q-Q plot we do see that the residuals generally follow a normal distribution, with some variation at the tails. Again, we see the same two low outliers that are pulling the model line towards them. Finally, for the Residuals vs Leverage plot, we see that the two outliers we had before are still present but we also see a new outlier with index 77 that also has a strong influence on the model. Everything else seems to influence the model equally. Generally, this model appears to fit the assumptions of a linear model. The following table shows a brief summary of the outliers found from these residual diagnostics.

Index	pH	NH4	NO3	Trt
74	3.8	29.6	0	Grazing
77	6.6	25.1	19.2	Grazing
84	4.3	16.7	10.2	Grazing

Table 7: Outliers with index 74, 77, and 84

From table 7, we see that the outliers with index 74 and 84 have a pH level of 3.8 and 4.3. This is much lower that the average pH level for those that are grazing as seen on figure 1. Note that index 74 has no nitrate whatsoever. Index 77 and 84 have nitrate levels that are way above the average as seen in figure 2.

Discussion

When sampling from a grazing area we expect the intercept for our linear model to decrease by approximately 1.17 on the pH scale. This means that we can expect grazing areas to be significantly more acidic than feeding areas.

From the plot of ammonium vs nitrate in figure 2 and from our mean variance inflation factor being slightly greater than two, we see that there is a relationship between some of the predictor variables. One possible explanation for why these variables sometimes have a positive correlation with one another is that they are both common ingredients in fertilizer. However, fertilizers can vary in composition which may explain why there was no clear linear pattern in 2. We chose to remove both ammonium and nitrate in our reduction of the full model. The reduced model included only location as a predictor variable, but still accounted for about 55% of the variability in pH. This is very close to the 58% explained by the full model. Given this set of predictor variables, we find that using only location would be the simplest model that explains a good portion of the variability in pH. However, this reduced model would need to be tested on a new sample in order to ensure it is not over-fitted to this particular sample.

One possible explanation for why ammonium and nitrate were not significant predictors of pH is that they are both relatively neutral on the pH scale. We expect there could be other factors that effect soil pH.

Given our data set, we have insufficient evidence to explain the samples with indices 74, 77, or 84. We suspect there are additional factors which affect pH of the soil that we have not tested in this model. We might also consider whether the data was collected and recorded correctly. We have chosen to keep these outliers in our model given that we have no evidence to suggest they fall outside of our target sample, or that they may have been errors. Further investigation of these outliers is necessary. A possible explanation is that there are different substances in the soil that affect the pH level more significantly than the ammonium and the nitrate. Even though ammonium nitrate is a fertilizer, there could be other fertilizers with different ingredients that can affect the pH levels more.

Conclusion

While ammonium and nitrate were not significant predictors for pH in our full model, we suggest there are additional factors on cattle farms that effect pH of the soil. For future investigation of controlling pH on cattle farms, we may investigate why feeding areas had a less-favorable pH for bacteria growth. Additionally, further investigation of the cattle farms where indices 74, 77, 84 were collected may help us find new predictors for pH. Additionally, we should ensure that cattle on these farms are tolerant enough to withstand pH conditions outside of the 6.5-7 range. Alternative methods for fighting antibiotic resistant bacteria is necessary for the health of cattle and ultimately the people these cattle support.

References

[1] Antibiotic resistance and food safety. https://www.cdc.gov/foodsafety/challenges/antibiotic-resistance.html. Accessed: 2020-03-09.

[2] United states department of agriculture: Publication highlights. https://www.nass.usda.gov/Publications/Highlights/index.php. Accessed: 2020-02-29.

[3] G. E. Agga, K. L. Cook, A. M. Netthisinghe, R. A. Gilfillen, P. B. Woosley, and K. R. Sistani. Persistence of antibiotic resistance genes in beef cattle backgrounding environment over two years after cessation of operation. *PloS one*, 14(2), 2019.

[4] J. Blamire. Effect of ph on growth rate. http://www.brooklyn.cuny.edu/bc/ahp/CellBio/Growth/MGpH.html, 2000.

[5] A. Netthisinghe, K. Cook, C. Rice, R. Gilfillen, and K. Sistani. Soil nutrients, bacteria populations, and veterinary pharmaceuticals across a backgrounding beef feedlot. *Journal of environmental quality*, 42(2):532–544, 2013.

[6] D. G. White, S. Zhao, R. Sudler, S. Ayers, S. Friedman, S. Chen, P. F. McDermott, S. McDermott, D. D. Wagner, and J. Meng. The isolation of antibiotic-resistant salmonella from retail ground meats. *New England journal of medicine*, 345(16):1147–1154, 2001.