Allison Upchurch and Sean Higgins
Michael Perry
STAT 303: Final Project

The Opioid Crisis

I. Intro

October, 2017, President Donald Trump declared the Opioid crisis a national emergency. After big pharmaceuticals began to push the sale and use of drugs like Fentanyl or Oxycodone, users became dependent, and when some attempted to get more, they were often given other disguised opioids such as heroin. Because of the user's addiction, they would take anything to get high; many would take foreign pills with an unknown dosage, often resulting in an overdose or death. This has spread throughout the nation affecting millions and leaving the addicts with few solutions. Withdrawals of opioids are considered one of the worst experiences, because of how long and painful it is. There is very little to no cure available for the majority of the nation, and if this trend continues, more of our loved ones may be lost to this national emergency. This data set from the CDC includes each county per state and their respective opioid prescribing rates. We will be looking to analyze the relationships between the negative impact of the change of prescribing rates on opioid overdose deaths, opioid prescribing rates on overdose death rates, and the average overdose prescribing rates of each region in the U.S. We hypothesize that a decrease in opioid prescribing rates from 2013 to 2015 leads to a lower number of drug overdose deaths in counties.

II. ANOVA

In order to determine which policy – increasing, decreasing, or maintaining opioid prescribing rates – had the most negative impact on opioid overdose deaths, we decided to perform an ANOVA test. First we had to divide the counties into three groups by the change in opioid prescribing rates. We did this by creating a new variable, group, and assigned each county as "Increase", "Decrease", or "Maintain" based on their change in opioid prescribing rates. We defined the "Increased" group as counties with a change greater than 0.15%, "Decrease" as counties with a change less than -0.15%, and "Maintain" as counties whose change was between -0.15% and 0.15%. We did this because there were only 28 counties which had 0 change in prescribing rate. Therefore, we decided to expand this category to include rates between -0.15% and 0.15% so this group would have 207 observations rather than 28. Although the groups sizes were still unequal, as the "Increased" group had 96 observations and the "Decrease" group had 644, doing this makes the group sizes slightly more similar.

For our ANOVA test, the null hypothesis was that the mean rate of opioid overdose deaths was the same for counties which increased, decreased, or maintained opioid prescribing rates. The alternative hypothesis was that at least one of mean overdose rates was different for the three groups of counties.

The F-test statistic was 5.29 and the p-value was 0.0052. Therefore, at $\alpha=.05$ we can reject the null hypothesis and conclude that there is statistically significant evidence to show that at least one of the three means is different.

Opioid Overdose Death Rate for Three Prescribing Rate Changes

The GLM Procedure

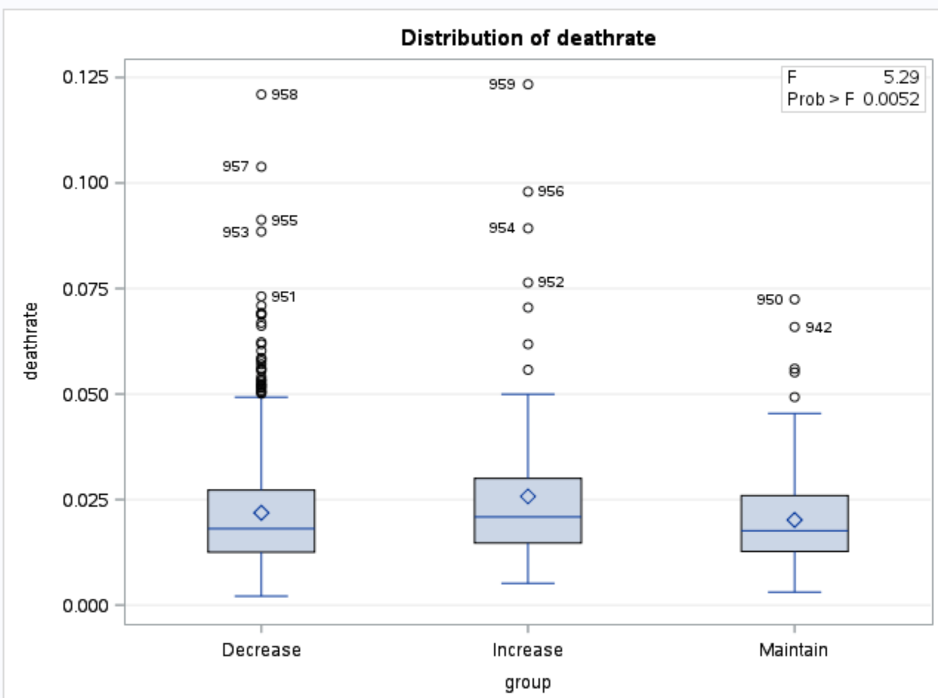
Dependent Variable: deathrate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.00208926	0.00104463	5.29	0.0052
Error	956	0.18870524	0.00019739		
Corrected Total	958	0.19079450			

R-Square	Coeff Var	Root MSE	deathrate Mean
0.010950	64.22488	0.014050	0.021876

Source	DF	Type I SS	Mean Square	F Value	Pr > F
group	2	0.00208926	0.00104463	5.29	0.0052

Source	DF	Type III SS	Mean Square	F Value	Pr > F
group	2	0.00208926	0.00104463	5.29	0.0052



We then performed a Bonferroni test in order to determine which mean was different. Based on the Bonferroni test, the differences between the “Increase” and “Decrease” groups, and the “Increase” and “Maintain” groups were significant. Therefore, we can conclude that the opioid overdose death rate for counties which increased opioid prescribing rates is significantly different than the mean death rates for counties which decreased or maintained opioid prescribing rates. Furthermore, the difference between the “Increase” and “Decrease” groups means was 0.0055786 and the difference between “Increase” and “Maintain” groups was 0.0038835. Thus, we can conclude that mean opioid overdose death rate for counties which increased opioid prescribing rates is significantly higher than the overdose rate in counties which decreased or maintained prescribing rates. These results support our hypothesis that increasing opioid prescribing rates lead to a higher opioid overdose death rate.

Bonferroni (Dunn) t Tests for deathrate

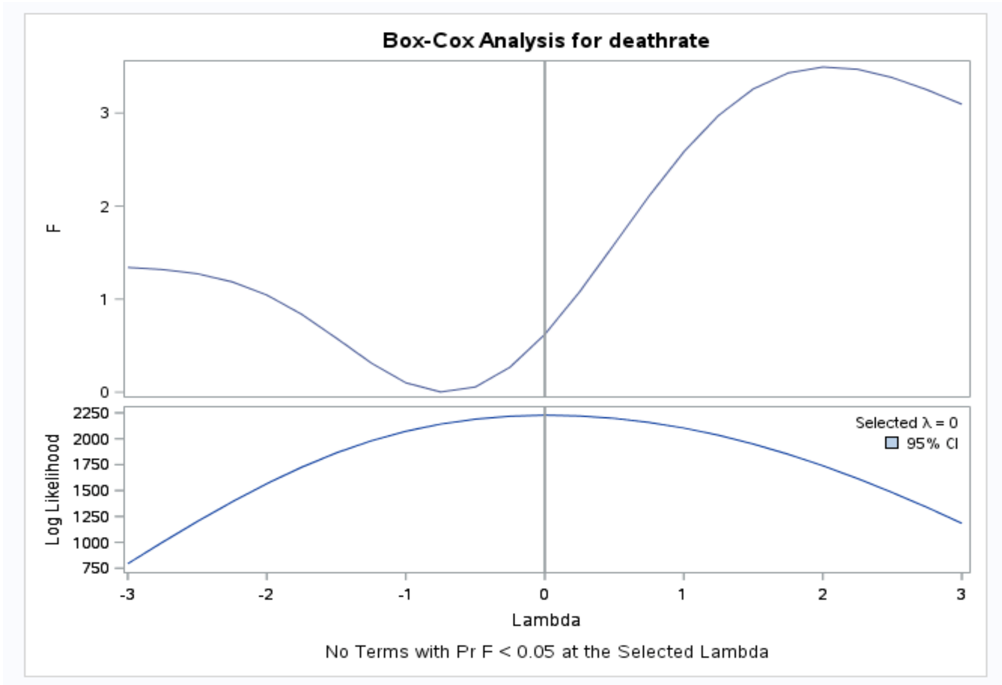
Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than Tukey's for all pairwise comparisons.

Alpha	0.05
Error Degrees of Freedom	956
Error Mean Square	0.000197
Critical Value of t	2.39820

Comparisons significant at the 0.05 level are indicated by ***.				
group Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
Increase - Decrease	0.0038835	0.0002138	0.0075532	***
Increase - Maintain	0.0055786	0.0014663	0.0096910	***
Decrease - Increase	-0.0038835	-0.0075532	-0.0002138	***
Decrease - Maintain	0.0016951	-0.0009451	0.0043353	
Maintain - Increase	-0.0055786	-0.0096910	-0.0014663	***
Maintain - Decrease	-0.0016951	-0.0043353	0.0009451	

III. Linear Regression

In order to determine the strength of the relationship between opioid prescribing rate and the opioid overdose death rate, we decided to perform a linear regression on the opioid overdose death rate in a county in 2016 by the opioid prescribing rates in 2015. This way we could analyze the strength of the relationship between prescribing rates in 2015 and the overdose death rate in 2016, and extend these conclusions to the relationship between opioid overdose death rate in 2015 and 2016, and the change in opioid prescribing rates from 2013 to 2015. However, the overdose death rate was non-linear, so we decided to perform a transformation on the data before performing another linear regression. In order to determine the best transformation to use, we first performed a Box-Cox Transformation, which gave us the result lambda=0.



Therefore, we went ahead with a natural log transformation on the death rate. This gave us an adjusted R-squared value of 0.0033, meaning .33% of the overdose death rate data can be explained by the relationship between the death rate in 2016 and the prescribing rate for 2015. This result suggests that the prescribing rate for a county has basically no impact on the overdose death rate for the next year.

The REG Procedure
Model: MODEL1
Dependent Variable: lndeathrate

Number of Observations Read	502
Number of Observations Used	502

Analysis of Variance

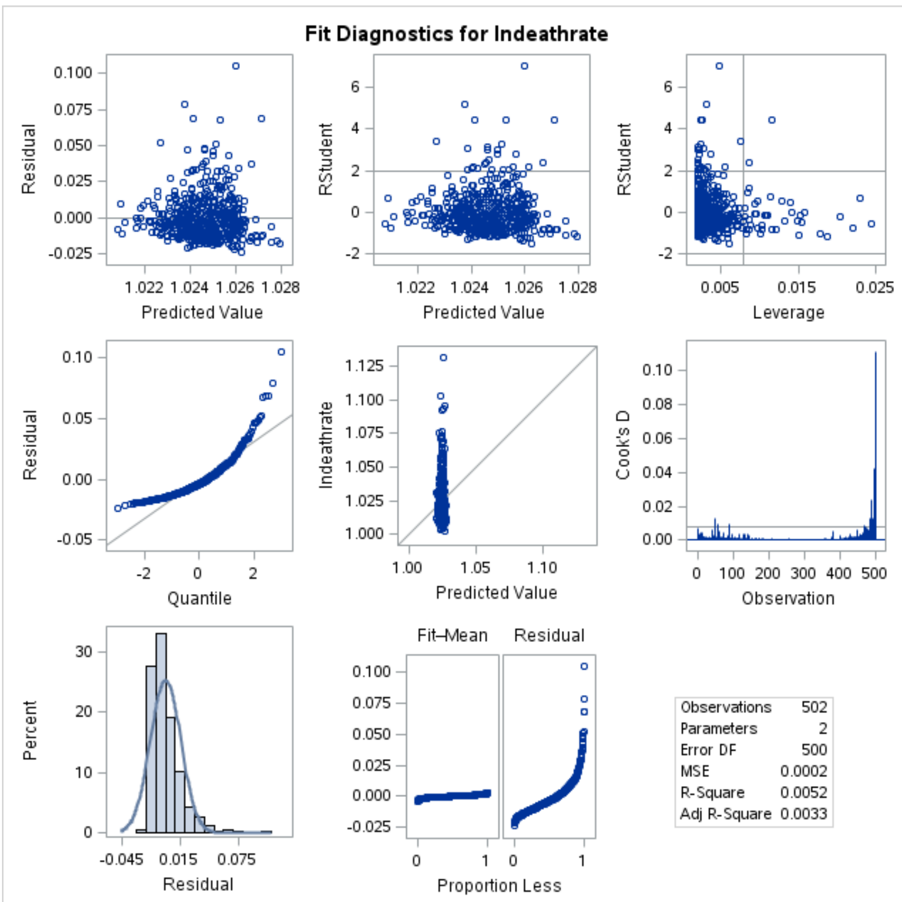
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.00065425	0.00065425	2.63	0.1052
Error	500	0.12417	0.00024835		
Corrected Total	501	0.12483			

Root MSE	0.01576	R-Square	0.0052
Dependent Mean	1.02462	Adj R-Sq	0.0033
Coeff Var	1.53804		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.02912	0.00287	359.19	<.0001
rate2015	X	1	-0.00077016	0.00047450	-1.62	0.1052

Based on Plot 1, the residuals are not randomly distributed, and are instead clustered around 1.025, therefore the assumption of independent residuals is not met. This cluster also shows that the residuals do not have constant variance, so this assumption is not met as well. Plot 4 shows that the residuals are curved and not straight, and in Plot 7 the data is clearly right skewed and not normal. Therefore, none of the residual assumptions are met. Since the assumptions are not met, we cannot use this model to predict the overdose death rate of a county in 2016 based on its overdose prescribing rate in 2015.



IV. Correlation

The extremely low adjusted R-squared found between overdose death rate and opioid prescribing rate made us wonder what variables actually influenced the overdose death rate. Therefore, we decided to investigate the correlation coefficients between 6 different variables: population, overdose opioid deaths, overdose death rate, and the opioid prescribing rates for 2013, 2014, and 2015.

For some of the variables tested, the correlation was easily explained. Obviously there is a strong relationship between population and overdose death count, since counties with more residents are likely to have more deaths of any kind. Similarly, we ignored the correlation between the overdose death count and the overdose death rate, since the death count was used to calculate the death rate. Finally, there was obviously a strong correlation between the three opioid prescribing rates for 2013, 2014, and 2015. After these, the strongest relationship was between overdose death rate and population. The correlation coefficient was -0.23477, meaning there is a weak negative relationship between the overdose death rate and the population of a county. This indicates that counties with a larger population actually had fewer opioid overdose deaths relative to the population.

The CORR Procedure

6 Variables:	pop deaths deathrate rate2013 rate2014 rate2015

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
pop	502	472194	716709	237041536	18646	10137915	
deaths	502	88.79880	110.01505	44577	20.00000	1041	
deathrate	502	0.02420	0.01517	12.14887	0.00247	0.12335	
rate2013	502	6.17004	1.50310	3097	1.68000	11.30000	J
rate2014	502	6.07803	1.48720	3051	1.57000	11.49000	Q
rate2015	502	5.85329	1.48378	2938	1.56000	10.83000	X

Pearson Correlation Coefficients, N = 502
Prob > |r| under H0: Rho=0

	pop	deaths	deathrate	rate2013	rate2014	rate2015
pop	1.00000	0.76610 <.0001	-0.23477 <.0001	-0.19967 <.0001	-0.19255 <.0001	-0.18632 <.0001
deaths	0.76610 <.0001	1.00000	0.08665 0.0523	-0.19788 <.0001	-0.19146 <.0001	-0.18180 <.0001
deathrate	-0.23477 <.0001	0.08665 0.0523	1.00000	-0.06081 0.1738	-0.05753 0.1981	-0.07169 0.1087
rate2013 J	-0.19967 <.0001	-0.19788 <.0001	-0.06081 0.1738	1.00000	0.97661 <.0001	0.94970 <.0001
rate2014 Q	-0.19255 <.0001	-0.19146 <.0001	-0.05753 0.1981	0.97661 <.0001	1.00000	0.97878 <.0001
rate2015 X	-0.18632 <.0001	-0.18180 <.0001	-0.07169 0.1087	0.94970 <.0001	0.97878 <.0001	1.00000

V. Histogram

Lastly, we decided to make a chart of the average opioid prescribing rate separated by each of the four regions of the United States, Midwest, Northeast, South, and West. After summing the total opioid claims and overall claims, we were able to come up with a new prescribing rate to describe each region, and to more clearly tell which area of the U.S. is affected most. In our findings, the South has the highest prescribing rate, while the Northeast is considerably lower than the rest. Although the number of observations is also lower for the Northeast, we do not believe that is influential as 217 is large enough to represent the population.

Sum of Opiod Claims and Overall Claims by Region

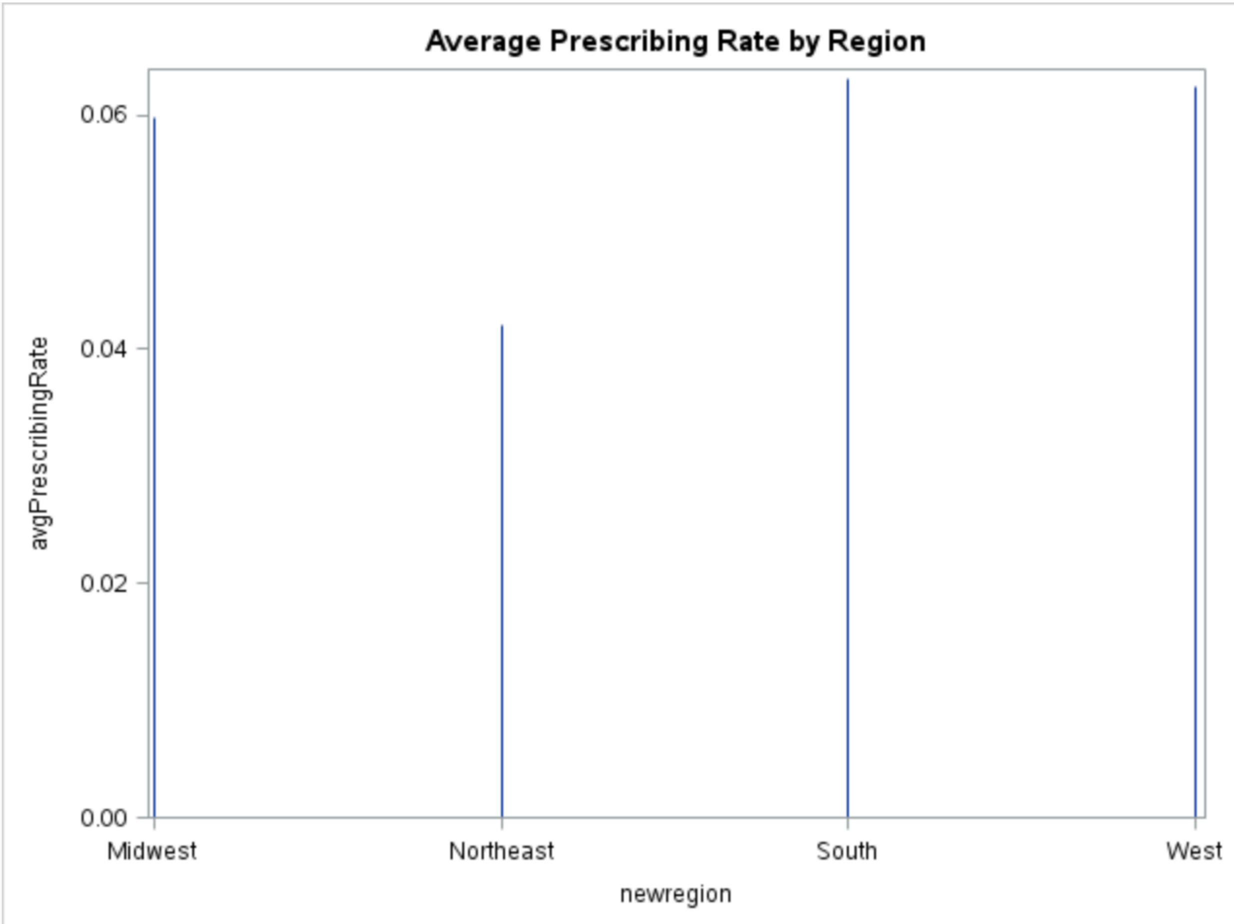
The MEANS Procedure

region	N Obs	Variable	Label	Sum
Midwest	1030	Opioid Claims Overall Claims	Opioid Claims Overall Claims	18040314.00 301464590
North East	217	Opioid Claims Overall Claims	Opioid Claims Overall Claims	10760493.00 256324049
South	1397	Opioid Claims Overall Claims	Opioid Claims Overall Claims	33040889.00 523644893
West	440	Opioid Claims Overall Claims	Opioid Claims Overall Claims	15255457.00 244149270

Prescribing Rates by Region

Obs	newregion	avgPrescribingRate
1	midwest	0.0598
2	northeas	0.0420
3	south	0.0631
4	west	0.0625

From the conclusions we made earlier to support our hypothesis that decreasing prescribing rates leads to a decrease in overdose deaths, this shows us that in the Northeast, compared inversely to the South, there is a lower rate of overdose deaths. The gap of 0.017 between Northeast and Midwest (the next smallest) is significant enough to conclude that there is something different about the Northeast that leads to a lower number of overdose deaths.



VI. Conclusion

From the ANOVA test, we were able to conclude that counties had an increase in opioid prescribing rates greater than 0.15 between 2013 and 2015 saw a significantly higher opioid overdose death rate in 2016 than counties which had a change in prescribing rates less than 0.15. However, between counties which decreased prescribing rates and counties which kept rates fairly similar, there was no significant difference between overdose death rates in 2016. This supports our hypothesis that increasing opioid prescribing rates leads to more overdose deaths. However, when we performed a linear regression between the prescribing rate in 2015 and the death rate in 2016, we found there was almost no relationship between the two variables. This refutes our hypothesis, since it suggests that changing prescribing rates in any way has no effect on the overdose death rate. This conclusion prompted us to find the correlation coefficients between several different variables in order to find what actually influenced the overdose death rate. The strongest relationship we found was between population and the death rate, although this relationship was still weak. However, the R-squared value for population and death rate was 0.0551, which was still 16 times higher than the adjusted R-squared between the prescribing rate in 2015 and the death rate. So population still had a much stronger relationship with the overdose death rate than the opioid prescribing rate did. Lastly, the graph of average prescribing rate by region shows us that there is evidence to show the Northeast region of the U.S. is affected less than the others, but there is not enough evidence to conclude anything about the other three. Finally, because from the ANOVA we can see a positive relationship in overdose death rates, but our other statistical analysis show there is likely a more influential factor in relationship between the two rates. Thus, we conclude that we do reject our hypothesis, a decrease in opioid prescribing rates from 2013 to 2015 leads to a lower number of drug overdose deaths in counties.