UNIVERSITY OF STRATHCLYDE

DEPARTMENT OF

MATHEMATICS AND STATISTICS

The Drugs that are leading to the drug mortality rate in Scotland.

By

Morgan Jo Tonner

201706965

BSc Hons Mathematics and Statistics

2021/22

Statement of work in project

The work contained in this project is that of the author and where material from other sources has been incorporated full acknowledgement is made.

Signed 

Print Name MORGAN JO TONNER……

Date 26th November 2021 …….

Supervised by Dr Bingzhang Chen

Abstract

This project is aimed at modelling the drug-related deaths in Scotland by the type of drug that caused death. As well as comparing the trends of drugs to each other against year. Looking at the mortality occurrences of drug- induced death for the years from 2008 to 2020.

The dataset this report is drawn from is the National Record of Scotland website which is shows the drug-related deaths in Scotland. The three separate datasets are merge together to make one larger dataset. The detail within the dataset are the 13 drugs, year, and the death numbers from each drug. The data is also made into another dataset with three columns of drug, death numbers and year. This is explored showing trends between drug and year.

Linear and polynomial regression modelling are used to model and compare drug-related deaths in the 13-year span. The models are also compared to each other to see which one best fits the data. Using R Studio to generate plot, lm and anova tables to shows the mortality between 2008 to 2020.

Constructing a more confident regression model for the other dataset. With year and drug as a factor. Visualising the data in boxplots to see if there are trends and outliers within the data. Using statistical techniques to conducted linear and polynomial modelling and anova hypotheses to look at trends. Using R Studio to generate boxplot, lm, anova tables and residual plot to shows mortality by drugs between 2008 to 2020.

Appling the statistical technique of Linear Regression and Polynomial Regression to the two separate datasets within this project. Built and fit model plots to visualise the calculations and compare slopes and curves to see what one best fit the drug death figures. Look at trends with the 13 drugs and see if illegal drugs cause more deaths than legal drugs.

Contents

List of Tables

List of Figures

The Drugs that are leading to the drug mortality rate in Scotland.

1 - Introduction

1.1 – What is a drug?

A drug is a medicine or another substance that has a physiological effect on you when it's ingested into your body. You can ingest some drugs into your body by swallowing, inhaling, or putting them directly into a vein. Most if not all drugs can alter a person's thinking and judgment and lead to serious health risks. These health risks can include addiction, drunk driving, infectious disease, and may have adverse effects on a woman who is pregnant. [1]

1.2 – Types of drugs

It can be extremely difficult to know what the different drugs do, and there are so many different types of drugs out there. All drugs in the UK can fit into three simple groups. These three groups are stimulants, depressants, and hallucinogens. They all have different side effects. [2]

Stimulants drugs all accelerate your central nervous system. The central nervous system controls most of your body and mind functions. [D] Drugs with stimulant's cause an increase in heart rate, breathing and affect the whole digestive system. Most of the drugs in this group also increase your confidence levels and make you think that your abilities have improved when they have not. These drugs are not addictive in the physical sense, but people like the feeling of added confidence. Therefore, this can cause a habit to form. Stimulant drugs are cocaine, ecstasy to name a few. [2]

Depressant drugs depress the central nervous system, but they do not make you feel depressed. Drugs in this group are hazardous to the central nervous system as they keep your heart going and keep you breathing, but if it gets depressed beyond a critical point, it can stop your breathing and then your heart will stop. More danger comes from taking drugs in this group. The real danger is when you mix two depressants together as you are likely to feel four times the effect than if you only take one. Depressant drugs are alcohol, methadone, morphine, to name a few. [2]

Hallucinogens drugs are slightly different from the other two groups as there are no physical addiction problems. These drugs can give people very scary and terrifying experiences. When taking drugs in this group, you are advised that you take them with people you trust and in a safe place to avoid harm to yourself and others. Hallucinogens drugs are magic mushrooms and ketamine, just to name a few. [2]

Not only can drugs be separated into three simple groups they can all be categorised into two other sub-categories these are legal and illegal. Drugs that are legal are most commonly drugs that have been prescribed to you by your doctor or pharmacist. An over-the-counter drug like Paracetamol are legal and these can be purchased from pharmacies or supermarkets. Illegal drugs are most commonly sold on the street and also come with the add risk if caught with them you could face hefty penalty. This penalty can range dependent on whether the drug is in one of the three class A, B or C and it is possession or supply and production. [3]

1.3 – Punishment for Drug offences

There are different punishments for drug offences in the United Kingdom. Taking, carrying, making, or even suppling drugs can all come with a fine and/or a prison sentence. Each of these offences carries its own punishment. Taking and carrying drugs carry a lesser sentence and maybe a small fine these are known as possession in the eyes of the law. Making and suppling carries a more severe sentence and a possible large fine. Most cases will receive both a fine and a prison sentence, this is known as supply and production . [4]

There are different type classifications of drugs that carry sentences and/or fines. Class A drug come with heavier sentence in comparison to class C. Drugs within class A are cocaine, ecstasy, heroin, class B drugs are amphetamines, cannabis, codeine, and class C drugs are diazepam, anabolic steroids, gamma hydroxybutyrate (GHB) these are just a few drugs within each class. As well as the three class there is a temporary class that the government can and have the right to ban for 1 year this will also come with punishment. The government debate and decide what classification the drug will be classed into. [4]

The jail term for the class A crimes is up to 7 years for possession and up to life for supply and production. For class B crimes is up to 5 years for possession and up to 14 years for supply and production and for class C crime is up to 2 years for possession and up to 14 years for supply and production. All the classes have an unlimited fine as it is up to the judge on all case to set the amount the person will have to pay. The temporary class that the government also has a prison sentence for people who are caught for supply and production with a jail time of up to 14 years were as if you are caught for possession in this class the police will just take the drug off you and give you a warning. [4]

1.4 – Addiction

Addiction is a common problem; it is defined by not having the ability to control behaviour or moods. Taking or using something to a point where the action you take can be harmful to you and/or others. Addiction in United Kingdom are mainly associated with drugs, alcohol, smoking and gambling but it is very possible to be addicted to something else. Other addicted include shopping, internet or even work but some people do not see these as an addiction. Any addiction should be taken seriously. [5]

These are lots of reasons why an addiction can start. In the case of drugs, alcohol and smoking people enjoy how the substance affect them both physically and mentally. Whereas gambling can have the same mental affect after winning as this would start the feeling of trying again to win so that they can recreate that winning feeling. This can start the habit which will be hard to stop if continued. After you are addicted to something it becomes extremely hard to stop. People that are addicts will all experience withdrawal symptoms when that are not doing the action. The withdrawal can be very unpleasant, so it is extremely easy to just give into the craving and carry on what you are doing so the whole cycle continues. [5]

Addiction can cause strain to the addict's personal life in many ways. The strain can be seriously damaging to a point where the addiction just takes over. In many case the substance of the addiction can cause damage to health not just the mental high the person can get. The physical heath of a person is important, but addiction can cause major health effects such as

liver failure, lung disease and even cancer. Some addiction will also lead to death by drug or alcohol, or the disease started because of the addiction. [5]

Many addictions can lead a person to become unemployed as a result of the damage done by the addiction. It can also lead to relationship breaking down and have an effect on children if they are in your life. As well as relationship and unemployment your addiction could make you homeless too. Not only could you be homeless, unemployed and no relationship you may also have dealings with the police, dealers, or debt collector due to your addiction. [5]

1.5 – Getting Help

Help is available for everyone that has an addiction. An addiction is a treatable condition after seeking help, there are lots of ways to get the help you need. By speaking to family and friends, seeing your GP, and asking for advice or contacting an organisation that specialises in helping people with specific addictions. You can find help and advice on the National Health Service (https://www.scot.nhs.uk/) (NHS) website on how to deal with your addiction. As well as help finding treatment services that the NHS offer in your area. [5]

1.6 – Scottish Government Mission

Although my study is from 2008 to 2020 the Scottish Government (https://www.gov.scot/) announced at the beginning of 2021, a National Mission to reduce the unacceptable drug death rate in Scotland. Part of their mission is their commitment to improve the data and the surveillance of the deaths by drug misuse. In the report made by the Scottish Government, they wished to focus on management information from Police Scotland on any suspected drug deaths to provide as timely an indication of current trends in drug death rates in Scotland as soon as possible. As well as the statistics from the National Records of Scotland (NRS) that are used presented the wider context of drug deaths in Scotland. The Police Scotland report is quarterly and the NRS report is annually. Both reports are discussed in the Scottish Parliament with the focus on helping and preventing the unacceptable death rates in Scotland. When the annual report from the NRS is published it has quite a lot of media attention on the data. [6]

1.7 – Aim

Aiming to compare the trends of drug-induced mortality occurrences among different drugs in Scotland from 2008 to 2020 and to see what drugs cause the most deaths in Scotland. Estimating trends with year and drugs to variables and look at legal and illegal drugs to see if the illegal drug causes more deaths than legal.

2 - Methodology

2.1 – Linear Regression

Regression analysis is a statistical method that is widely used in many areas such as statistics, economics, technology, social sciences, and finance. The statistical model of regression represents the relationships between a predictor and a response variable. A linear regression model is constructed to show the relationship between a predictor (x) and a response variable (y). [7]

The simple linear regression model or line of best fit can be found using

$$y = \beta_0 + \beta_1 x + \varepsilon \quad \text{(E1)}$$

The linear model involves a standard error term that is represented by ε. The error term in the model is used to detail its variability in y that cannot be interpreted by the linear relationship between x and y. If epsilon ε did not exist, then that would mean that knowing what x is would provide enough information to determine the value of y. There are also other parameters that are represented by the populations in the study of the model. The parameters in this model are β_0 and β_1. The linear regression equation is a straight line on the graph where β_0 is the y-intercept of the regression line and β_1 is the slope line in the model. [7]

A linear regression model can display one of the three relationships. A positive, negative, or even no relationship at all. No relationship means that the graphic line is flat as there is no relationship between the predictor and response variable. A positive relationship means that the graphic slope points upwards with the lower end of the line at the y-axis of the graph and the upper end of the slope extends upward into the graphic field, away for the x-axis. Positive linear relationship between the predictor and response variable both values increase. Negative relationship means that the slope points downward with the upper end of the line at the y-axis and the lower end of the line extends downward in the graphic field, going towards the x-axis. The negative relationship between the predictor and response variable both the values decrease. [7]

For results of linear regression model to be valid and dependable, the model has to check four assumptions are met for the model to be good. Firstly, there exists a linear relationship between the predictor and response variable. Secondly the residuals are independent and there is no correlation between any of the residuals. Thirdly the residuals have a constant variance at any and all levels of predictor variable. Finally, the residuals of the model are normally distrusted on response variable. If one or more of these assumptions are not met or violated, then the result of the regression model can be seen as misleading or unreliable. [8]

2.2 –Polynomial Regression

Simple linear regression equation only work when the relationship between the two variables in the data are linear but suppose that the data is non-linear then the linear regression is not capable of drawing the best fit line and therefore fails to meet such condition. If we consider that the data has non-linear relationship, but the slope of the linear regression is not close to the reality of the model. Introduction of polynomial regression to overcome this problem will help identify a curve relationship between the two variables. [9]

Polynomial regression is a form of a linear regression which the relationship between predictor (x) and response (y) variables are modelled to an nth degree polynomial. This regression model fits non-linear relationship between the value of x and the mean of y. We can model the expected value of y to is nth degree polynomial. Polynomial regression model can contain more than one predictor variable x. [10]

The equation used in polynomial regression can have each of its predictor variables present in it with as many various of powers:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n + \varepsilon \quad \text{(E2)}$$

The polynomial regression model involves a standard error term that is represented by ε. With y being the predictor and x the response variable in the equation. All the β being the parameter of the model. [10]

Considering a polynomial regression model of degree two, the equation is:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon \quad \text{(E3)}$$

This model is called second order as the predictor is expressed in the polynomial model to first and second power. With all the β being the parameters of the model and y being the response variable. Again, this another regression model that involves an error term of ε. [10]

Assumptions of a second-order polynomial regression model are that the independent variables are independent of one another. As well as the behaviour of a predictor variable can be explained by linear or curvilinear regression and in an additive relationship between the predictor and the set of independent variables. The relationship between both variables predictor and response is curvilinear of the regression. The standard error term is also independent, normally distributed with a mean value zero and a constant variance. [11]

2.3 – Regression Hypotheses

Linear and polynomial regression are techniques that can be used to understand relationship between one or more predictors variables and a response variable. For only one predictor and one response variable, we use equation (E1) to estimate the relationship between predictor and response. So, for simple linear regression we have a null hypothesis of:

$$H_0: \beta_1 = 0 \quad \text{(H1)}$$

where the coefficient β_1 is equal to zero. Therefore, no significant relationship can be found between the predictor and response variable. The alternative hypothesis:

$$H_A: \beta_1 \neq 0 \quad \text{(H2)}$$

where the coefficient β_1 is not equal to zero. Therefore, there is a significant relationship that can be found between the predictor and response variable. [12]

For polynomial models of degree two with only one predictor and one response variable, use in equation (E3) to estimate the relationship between predictor and response variable. The null hypothesis:

$$H_0: \beta_1 = \beta_2 = 0 \quad \text{(H3)}$$

where both coefficients have to equal zero. Hence, there is no significant relationship between the predictor and response variable. The alternative hypothesis:

$$H_A: \beta_1 = \beta_2 \neq 0 \quad \text{(H4)}$$

where at least one coefficient is not equal to zero. Therefore, there is a significant relationship that can be found between the predictor and response variable. [12]

2.4 – Analysis of Variance (ANOVA)

Analysis of variance commonly known as ANOVA is a statistical test used to analyses differences between the mean of two or more groups. [13] A one-way ANOVA is used to determine if there are any statistically significant differences between the means of the independent unrelated groups or factors. One-way ANOVA compares the means between the groups or factors that you are interested in. To determine if any of those means are statistically significantly different from one another. [14]

For a one-way ANOVA test you would have a null hypothesis of:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_n \quad \text{(H5)}$$

where μ is the mean of the group and n is the number of groups you wish to test. For only two groups your null hypothesis would be:

$$H_0: \mu_1 = \mu_2 \quad \text{(H6)}$$

You will accept the null hypothesis if the mean group one is equal to the mean group two. If the two groups mean are different, you fail to accept the null hypothesis, and therefore, you will consider that there is a difference between the means of the groups. That you would have in your alternative hypothesis. An alternative hypothesis would be:

$$H_A: \mu_1 \neq \mu_2 \quad \text{(H7)}$$

If there is statistical difference in test it is like that you will reject the null in favour of the alternative hypothesis. [14]

There is assumption needed to look at for testing for ANOVA and they are that the data is normally distributed, testing homogeneity of variance and that the group are independent of each other. [14]

An ANOVA table for one-way test:

Table 1: ANOVA Table for One-way.

Source	Degree of Freedom	Sum of Squares	Mean Sum of Squares	F-test
Treatment	k -1	Sum of Squares Treatment (SSTr)	Mean Sum of Square Treatment (MSTr) = SSTr / (k – 1)	F = MSTr / MSE
Error	n – k	Sum of Square Error = SST - SSTr	Mean Sum of Square Error (MSE) = SSE / (N – k)	
Total	n – 1	Sum of Squares Total		

Where n represent the number of observations in the response and treatment groups and k is the number of observations in the treatment group.

2.5 - Source data

The dataset that forms this report is gathered from the National Records of Scotland (https://www.nrscotland.gov.uk/) (NRS). NRS is a non-ministerial department in the Scottish Government that is responsible for civil registration, the census in Scotland, demography and statistics, family history, and the national archives and historical records.

The NRS dataset contains many different sets from drug-related deaths in Scotland such as gender, age, council area, NHS heath boards, drug cause of death, method cause of death and many more. These datasets are update every year by the NRS with the most recent years data. Plus, they release an annual drug-related death report that comments and shows all this data.

The Scottish Government release the report that the NRS publish on an annual base, and they also publish the data form Police Scotland on a quarterly base. These datasets and reports are all part of the Scottish Government plan and mission to reduce the unacceptable drug-related deaths in Scotland in the next five years.

This project report will focus on the drugs that cause deaths in Scotland between 2008 to 2020. It will merge three datasets into one for the NRS that has been released. As well as looking at whether the drug is legal or illegal. The objective of the project is to constructed linear and polynomial regression model for all the drugs in the dataset. Fit the model to the given data and plot the slope and curve on the one plot to visualise it. Then to investigate what model best describes and fits the data. As well as comparing both model to each other for all the drugs.

Drugs within the dataset are split into three categories benzodiazepines, opioids, and substances. The 4 drugs within benzodiazepines they are prescription, street, diazepam and etizolam. Methadone, codeine, dihydrocodeine and heroin/morphine are the 4 opioids drugs in the dataset. There are 5 drug contains in substances they are alcohol, cocaine, ecstasy, amphetamines and gabapentin/pregabalin.

Working with a significant level of 5% all models.

For all linear regression model anova tables the null hypotheses are that there is no significant relationship between year and the drug deaths numbers. The alternative hypothesis is that there is a significant relationship between year and drug deaths numbers.

For all polynomial regression model anova tables the null hypotheses are that there is no significant relationship between year and the drug deaths numbers. The alternative hypothesis is that there is a significant relationship between year and drug deaths numbers.

3 – Results

3.1 – Data Visualisation

The data that is used came for 3 datasets that I merged into one big dataset. Figure 1 below shows all the drugs within the dataset as well as the total number of drug death in Scotland between 2008 to 2020. Over the 13-year period it is clear from the plot that the total number of drug death in Scotland are increasing. There are 13 drugs that are contains in the dataset. The highest drug in Scotland that cause the most death is Street Benzodiazepines were as the lowest is the substance Ecstasy.

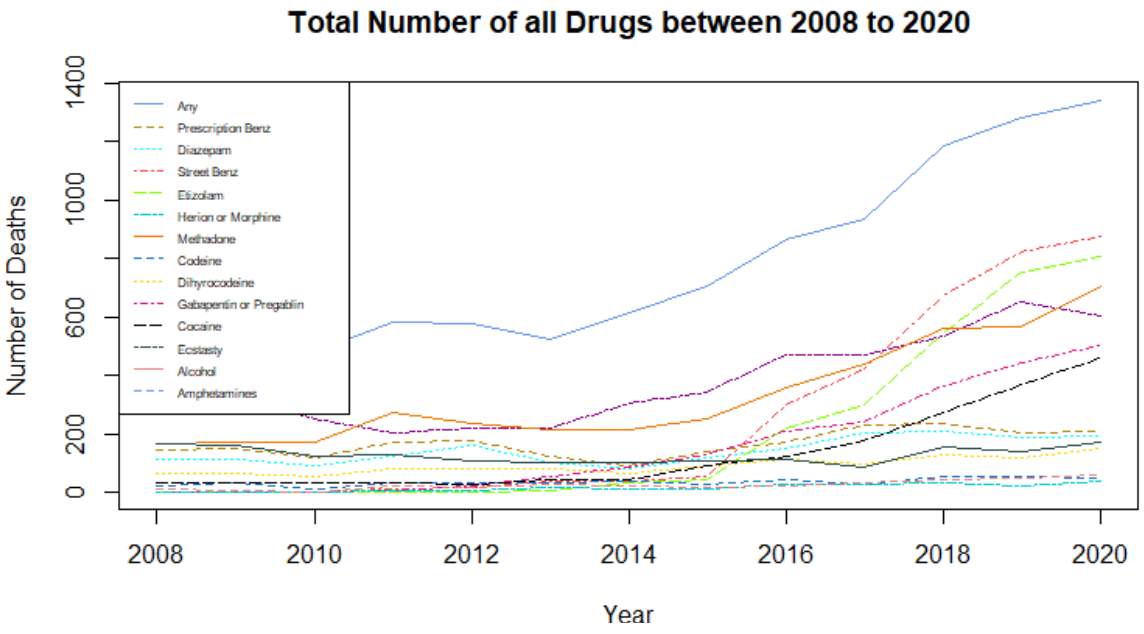

Figure 1: All Drug Deaths in Data set against year

Figure 2 shown below illustrates the legal and illegal drugs between 2008 to 2020. The legal drugs represented in green are lower than most but not all illegal drugs. Illegal drugs shown in pink have increased more over the 13 years, with a significant increase after 2015. With must the drugs death rates follow the overall death number shown in blue. Both the highest and lowest death figures in 2020 are illegal drugs.

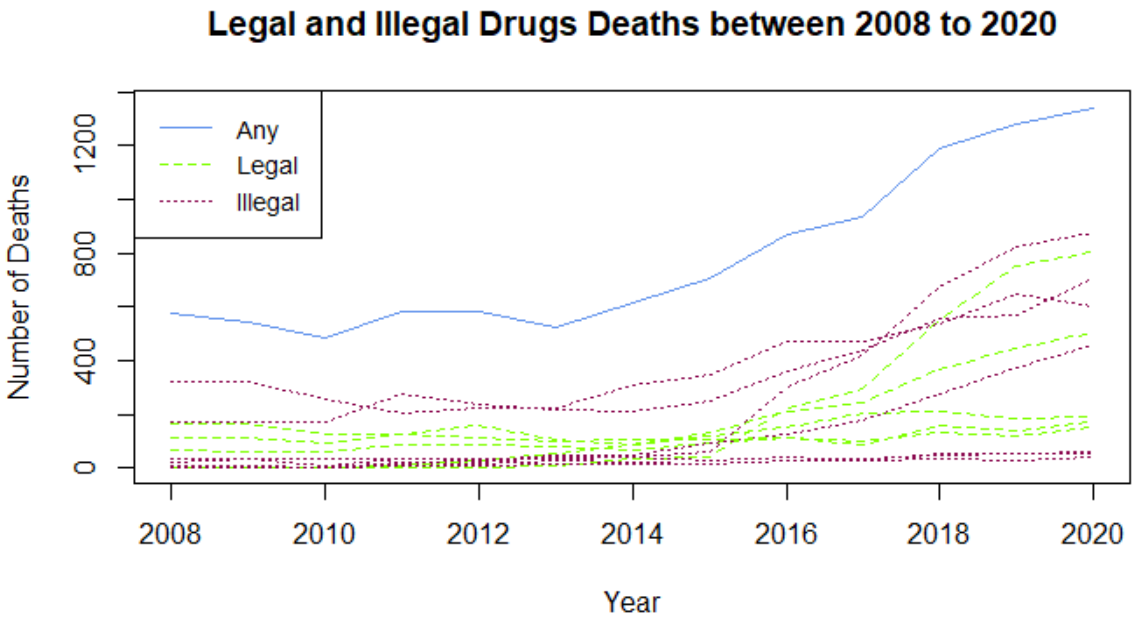

Figure 2: Legal and Illegal Drug Death set against year.

Blue line is total death by drugs. Legal drugs are green lines and Illegal drugs are pink. lines.

3.2 – Boxplot for regression model on new data

Figure 3 shows that drug deaths increase over the 12- years period. The boxplot shows the median of the drug deaths by year is low in all the year boxes. Looking at the boxplot the median is at its highest is 2018 and lowest in 2012. The data spread in 2014 is low in comparison 2019 and 2020. There are 5 outliners out the 169 entries in the data there does not seem to bad in the data. These outliners come in 2013, 2014, 2015 and 2016.

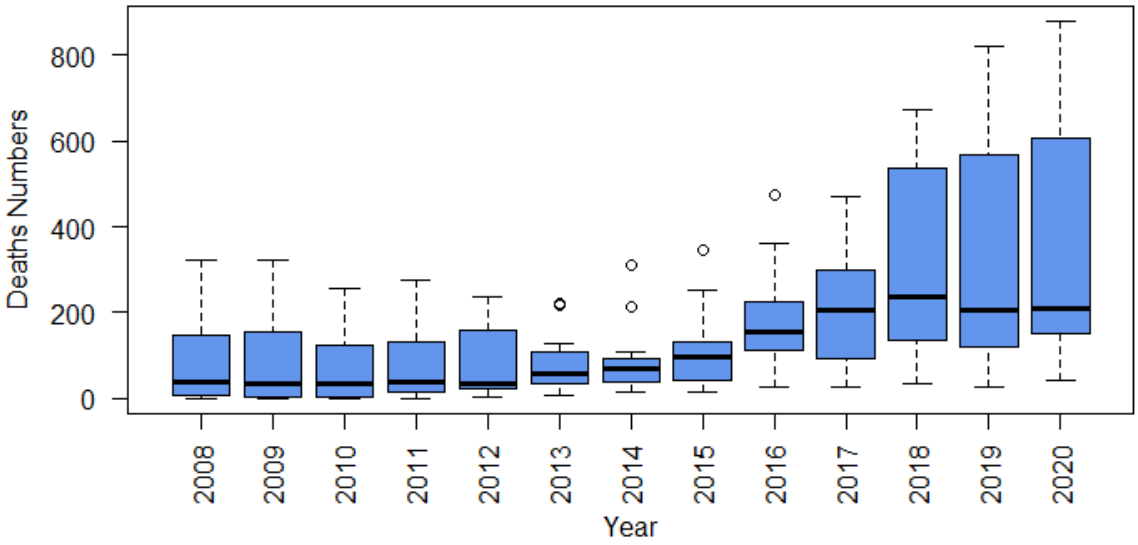

Figure 3: Boxplot of Death Number ~ Year

Figure 4 shows the drug deaths by drug. The boxplot shows the median of the drug deaths by drug to do different in all the drugs. Likely the boxplot in figure 10 there are 5 outliners out the 169 entries in the data there does not seem too bad in the data. These outliners come in amphetamines, cocaine, codeine and etizolam. Looking at the boxplot the median for alcohol is in the middle of the box making it likely the median of alcohol is possible the mean to and prescription benzodiazepines median is also in the middle. Amphetamine's box is small in size, so it is unclear if the median is low or high or in middle of the box this is the same for codeine and ecstasy drug unclear median base on small box. Cocaine, diazepam, etizolam, gabapentin or pregabalin, heroin or morphine, methadone and street benzodiazepines median are low in box. Dihydrocodeine median is not low but not in the middle but is close to the middle, so the median is pretty close the mean for the drug.

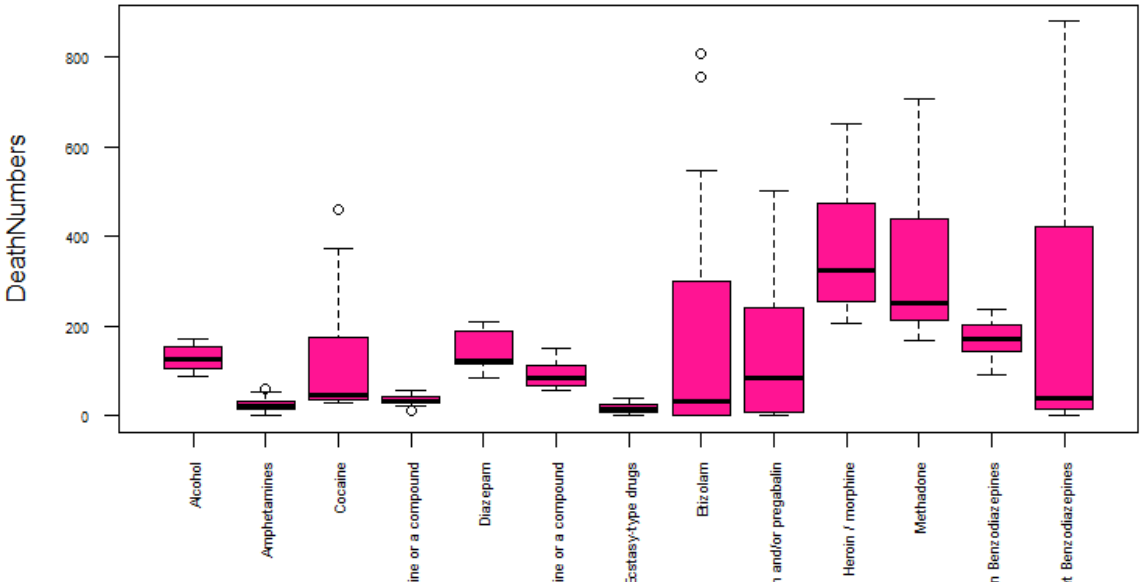

Figure 4: Boxplot of Death Number ~ Drugs

3.3 – Regression Model and Analysis of Variance Table for All Drug Deaths

Figure 5 shows the all-drug deaths model regression plot. The number of total deaths in Scotland has increase over the last 13-year period having hit it all time high last year in 2020. Looking at the plot you see that 2010 has the lowest death numbers in this period of time. The linear regression line doesn't best fit the data with only 2010 and 2011 years on the line or near the line. The curve line fit the model data much better than the linear line. It is not exactly the best fit but in comparison the straight line is works better. All the year points are very close if not on the curve line making the polynomial model the best for this data.

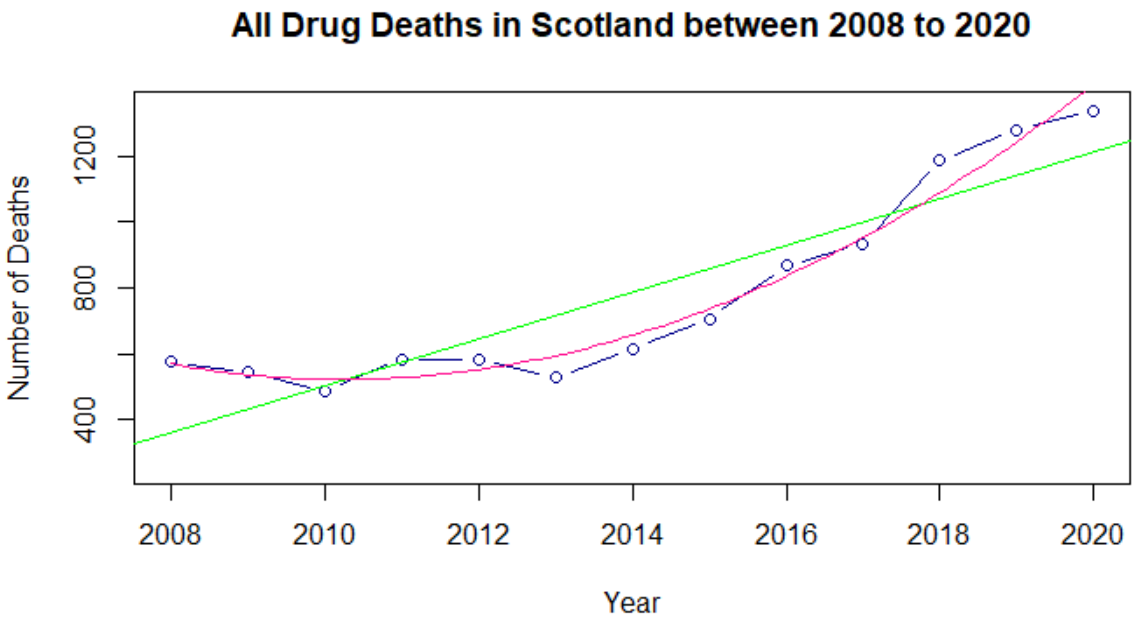

Figure 5: All Drug Deaths Model plot

The linear model shown with the green line and polynomial model shown with the pink curve

The ANOVA table for the linear regression model shown in table 2. The table shows the p-value to be 2.444e-05 this the very low. So, the p-value is less than the significant level of 5% so in this case you would reject the null hypothesis in favour of the alternative.

Table 2: ANOVA Table for Linear Model

Source	df	Sum Sq.	Mean Sq.	F value	Pr (> F)
Year	1	910942	910924	48.21	2.444e-05
Residuals	11	207850	18895		

Figure 6 show the four residuals' plots. Looking at the plot you see that the dataset does not have a lot of entries. The normal Q-Q plot shows the majority data points are below the normal line on the plot and the do not make a straight it is more a snake line. So, the residuals vs fitted plot show an approximately quadratic relationship between year and all drug deaths on the plot. There is no clear patter among the residuals and that the points are randomly scatter around the red line. Plus, the red line is horizontal across the plot then we can assume

that the homoscedasticity is satisfied for the given linear model. Looking at the residual's vs leverage I see that one-point falls outside the cook's distance making an influential observation to the data.

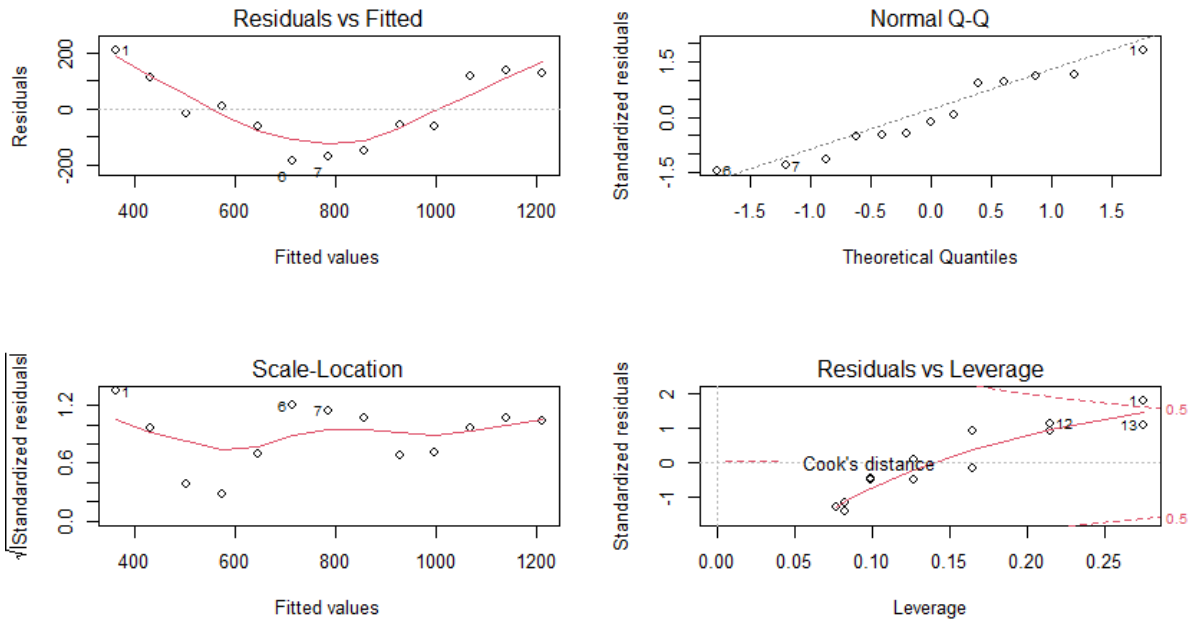

Figure 6: The linear regression residual plot for all drug deaths

The ANOVA table for the polynomial regression model shown in table 3. The table shows the two p-value one for year to be 1.01e-08 and the other of year square to be 2.04e-05. Both p-value are lower than the significant level of 5%. So, for the polynomial model you would reject the null hypothesis in favour of the alternative.

Table 3: ANOVA Table for Polynomial Model

Source	df	Sum Sq.	Mean Sq.	F value	Pr (>F)
Year	1	910942	910942	290.978	1.01e-08
I(Year^2)	1	176543	176534	56.393	2.04e-05
Residuals	10	31306	3131		

Figure 7 shows the four residuals' plots. The plots do not show the dataset does not have a lot of entries. The normal Q-Q plot shows the data points in a straight line but spread apart evenly. The residuals vs fitted plot show that a number of points lay before 800 on the x-axis. There is no clear patter among the residuals and that the points are randomly scatter around the red line. Plus, the red line is horizontal across the plot then we can assume that the homoscedasticity is satisfied for the given linear model. Looking at the residual's vs leverage I see that one-point falls outside the cook's distance making an influential observation to the data.

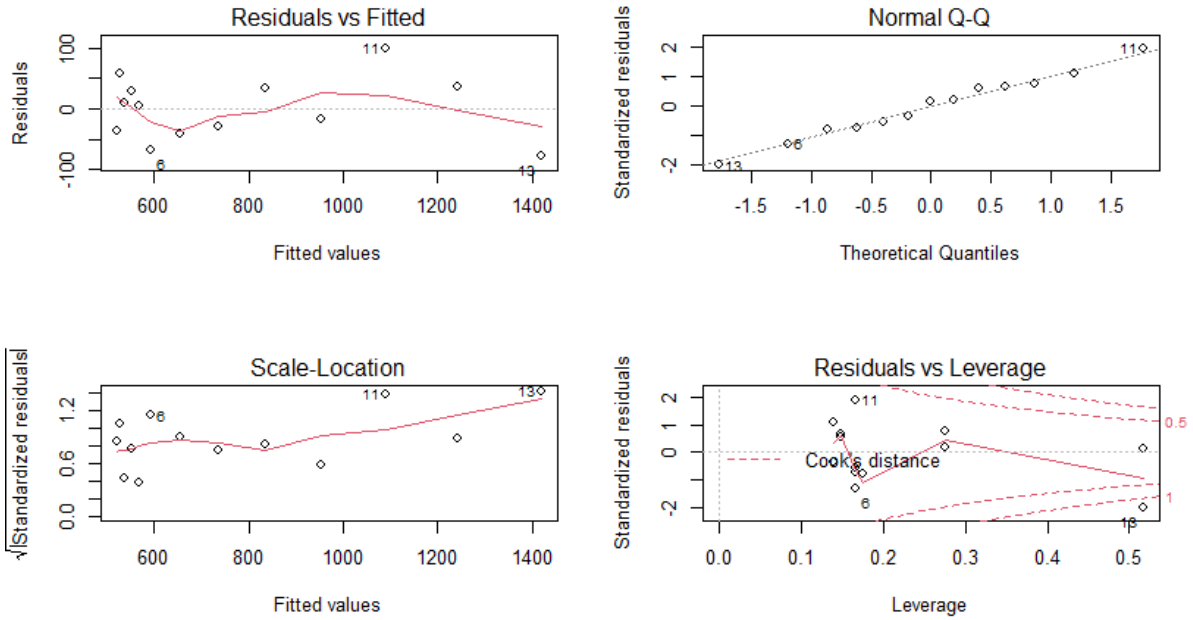

Figure 7: The polynomial regression residual plot for all drug deaths

3.4 – Comparing drug deaths

Comparing benzodiazepines drugs shown in figure 8. Looking at the plot we see that drugs are increasing over the 13-year period. There is a big increase for street and etizolam drugs in 2015 these two drugs follow a similar trend as the lines are similar in shape. Etizolam drugs do not cause as many deaths as compared to street. Both street and etizolam are illegal drugs shows that for benzodiazepine illegal drugs cause more death then legal. Prescription and diazepam do not cause as many deaths in comparison to the other two. These two drugs start higher than illegal drugs but do not increase as fast. Both lines on the plot follow a similar shape.

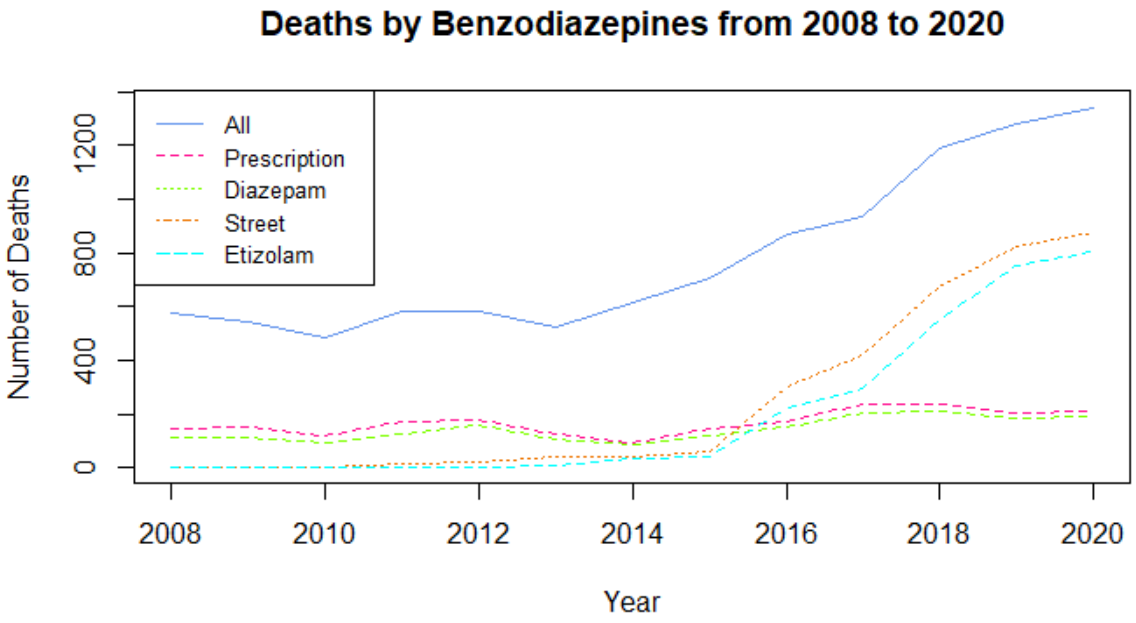

Figure 8: Benzodiazepine's plot

Figure 9 shows opioid drug deaths between 2008 to 2020. Heroin and/or morphine start off the highest of the four drugs but decrease between 2009 to 2014 before rising again till last year when it falls slightly. Methadone stays even at the beginning of the data then increase in 2011 to decrease the following year. It again increases in 2015 before it reaches its highest in 2020 this caused the most death in the year. Heroin, morphine, and methadone are all illegal and this causes more deaths the legal drugs for opioids in the last 13-years. Codeine and dihydrocodeine are legal drug and do not cause as many deaths. Both these drugs are very low on the plot in comparison to the other two. Legal drugs on the plot are very similar in shape and follow the same trend were as for heroin, morphine and methadone do not follow similar shape lines on the plot.

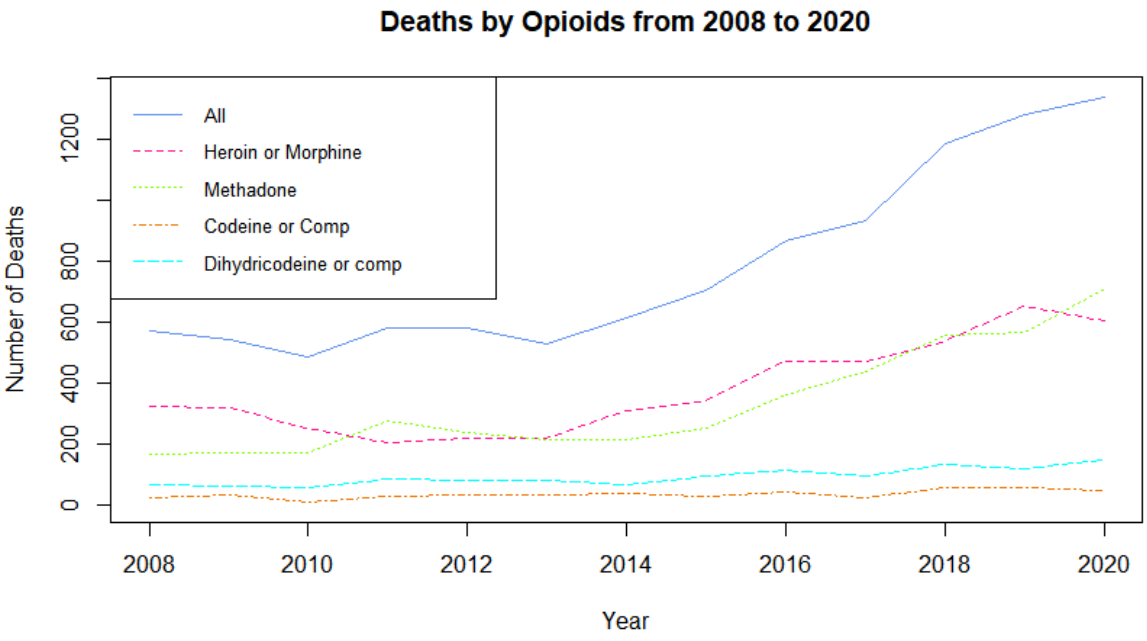

Figure 9: Opioid's plot

Figure 10 shows the substance that cannot be categorised into benzodiazepines and opioids. These drugs do not have as many deaths. Alcohol deaths were the highest at the start of the data but decrease gradually until 2017 before increasing. Gabapentin and/or pregabalin like the other drugs on the plot started off very low but increase over the 13-years and have become the most common in the other substance categories. The other three drug in the plot follow very similar trends and death figures are less than 450 between them. Making the legal substance drugs cause more deaths than illegal.

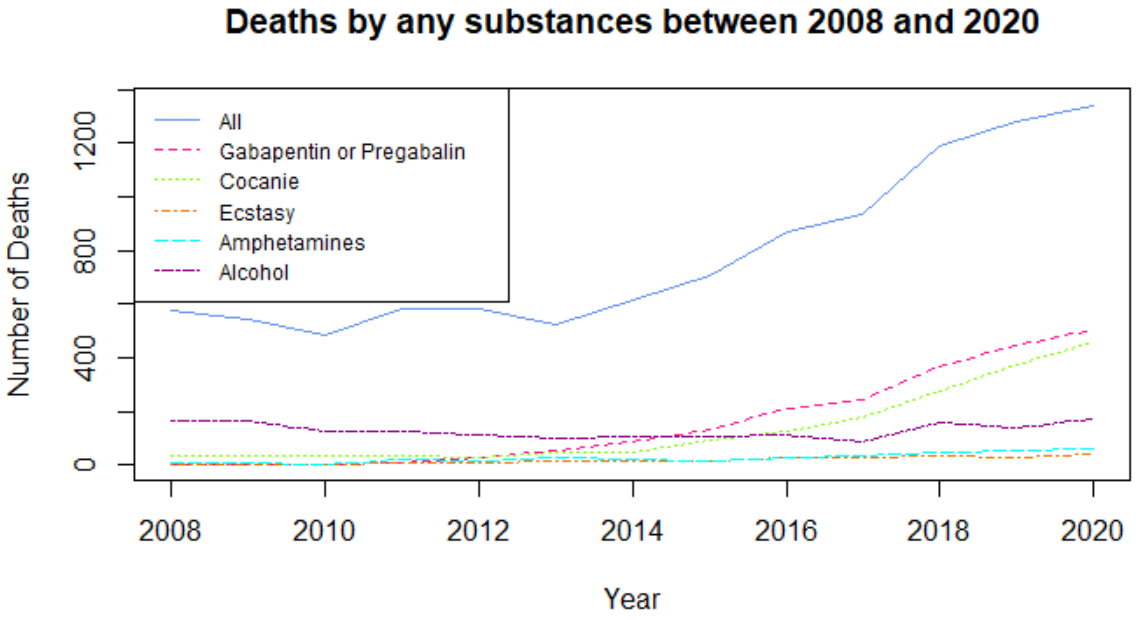

Figure 10: Substance's plot

3.5 – Linear and Polynomial Regression Model

Figure 11 shows the four plots of both the linear and polynomial regression models the drugs in these plots are benzodiazepines. There are four benzodiazepine drugs within the dataset. The line for the linear model as seen on the plots does not best fit the data in comparison to the polynomial model curve. The curve in all four plot fits the model data better than the line but this is not the best fit data for prescription benzodiazepines and diazepam. Etizolam and street benzodiazepine curve line does best fit the data. There is a significant drop in 2014 for prescription benzodiazepines and diazepam there is also a drop in 2014 street benzodiazepines and etizolam but in not as significant as the other three.

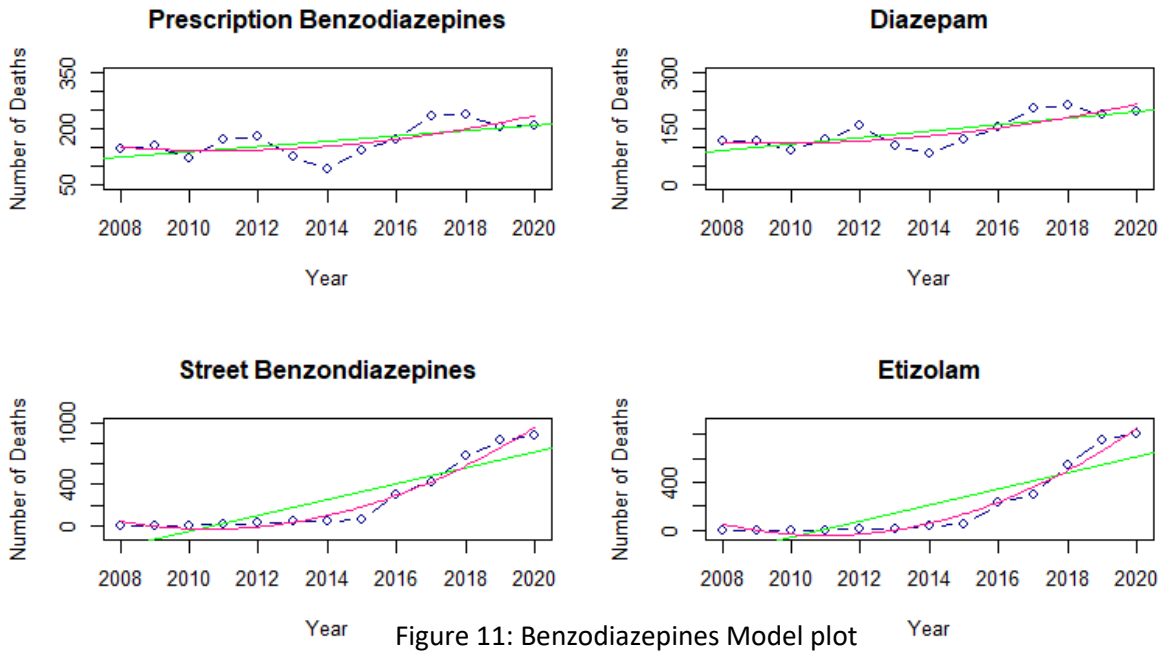

Figure 11: Benzodiazepines Model plot

The linear model shown with the green line and polynomial model shown with the pink curve

Figure 12 shows four plots with both linear and polynomial model on them. All drugs in these plots are of opioids there are four drugs in the dataset. Looking at the plots I can see that the linear line is not the best fit for methadone, heroin, and morphine in comparison to the curve line for polynomial regression fit the model much better than the linear one. The curve looks as if it goes through must of the points if not all of them. Both models best fit the data for codeine as the points are close to the line and the curve line is not as curve as the round as the other model plots.

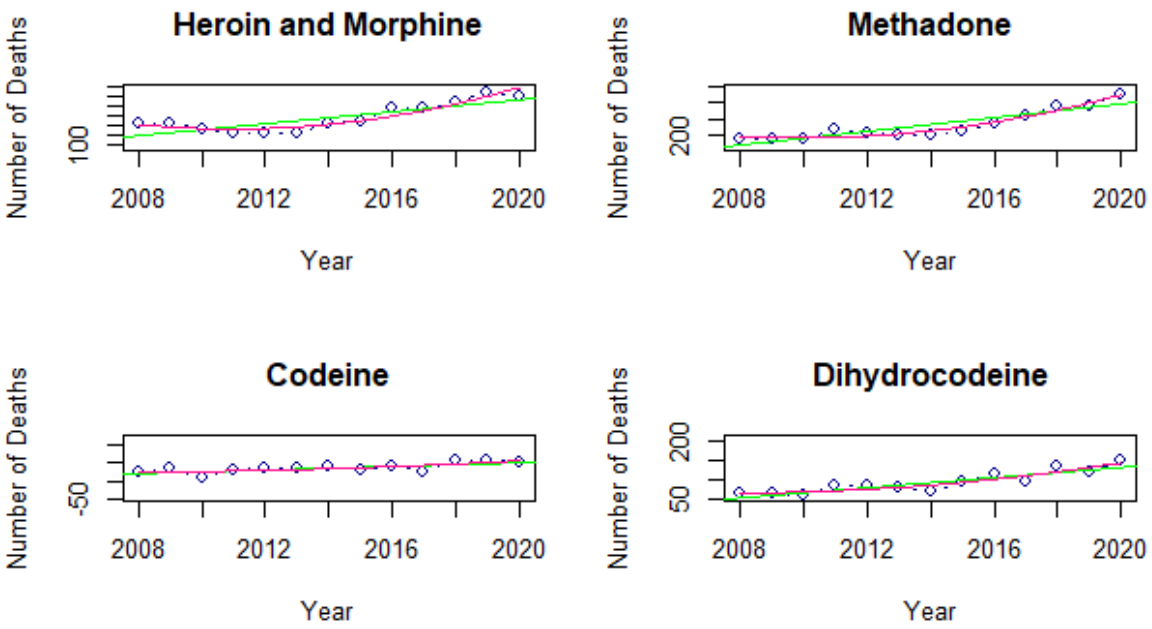

Figure 12: Opioids Model plot

The linear model shown with the green line and polynomial model shown with the pink

Figures 13 shows both linear and polynomial regression model for four of the five other substances within the dataset. The linear line does not best fit the data for cocaine, gabapentin, and pregabalin were as ecstasy and amphetamines data does fit the linear line. Ecstasy and amphetamines polynomial regression model curve is very close to being a straight line like the linear line on their plot. For these two drugs there year points are very close to the model lines and some of the points are on the lines. Polynomial regression for cocaine, gabapentin and pregabalin data best fits the curve line. Both cocaine, gabapentin and pregabalin data points lay on or close to the curve making the model a perfect fit of the data in these cases.

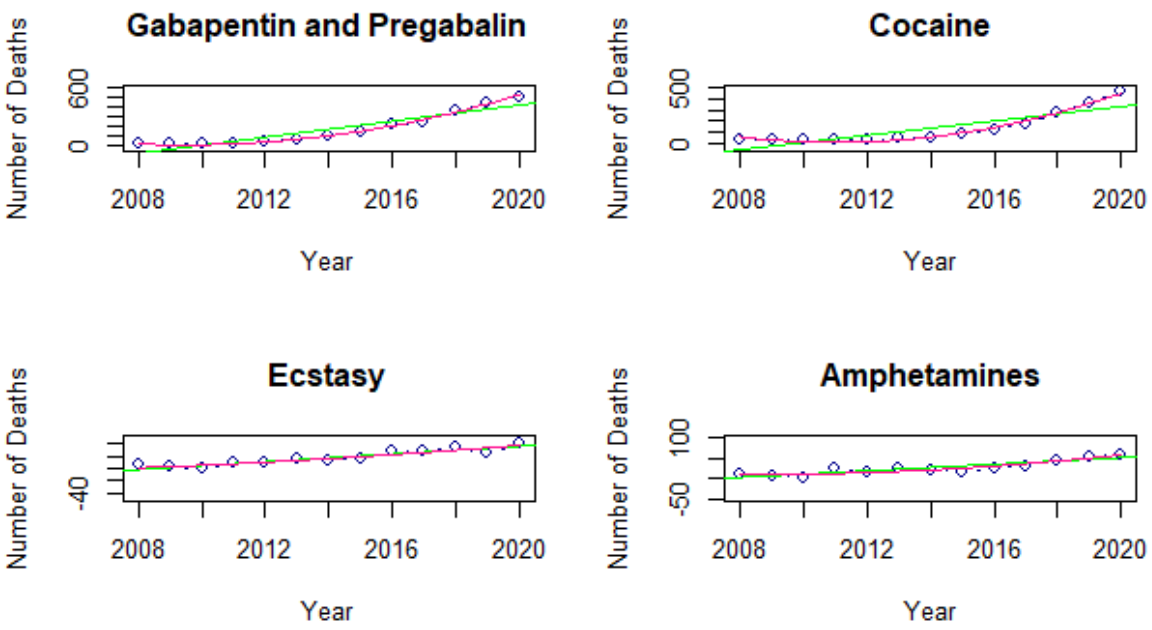

Figure 13: Any other Substance Model plot

The linear model shown with the green line and polynomial model shown with the pink curve

Figure 14 show below is model plot of alcohol. Again, looking at the plot the linear line does not best fit the model. The model linear line is a horizontal line with a decreased in 2014. This decrease come as alcohol related death was decreasing every year until it increasing quite significantly in 2018. In 2020 the death rate was just as high as it was at the begin of the 12-year period. The curve line fit most of the data points in this model as the curve show the trend of the alcohol death going down and then back up. The polynomial model best fits the data for alcohol.

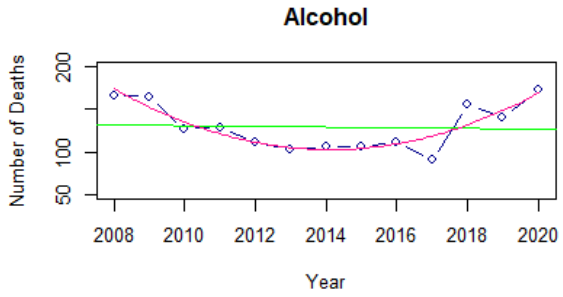

Figure 14: Alcohol Model plot

The linear model shown with the green line and polynomial model shown with the pink curve

3.6 – Analysis of Variance (ANOVA) Table for Regression Model

In this section we will look at six ANOVA tables for both linear and polynomial models. The six drugs picked base of the plots section 3.3. The drug in this section are prescription benzodiazepines, street benzodiazepines, methadone, codeine, cocaine, and alcohol.

The ANOVA table for the linear regression model shown in table 4. The table shows the p-value to be 0.02394 this the very low. So, the p-value is less than the significant level of 5% so in this case you would reject the null hypothesis in favour of the alternative for prescription benzodiazepines.

Table 4: ANOVA Table for Linear Model – Prescription Benzodiazepines

Source	df	Sum Sq.	Mean Sq.	F value	Pr (>F)
Year	1	8960	8960.0	6.8613	0.02394
Residuals	11	14389	1307.8		

The ANOVA table for the polynomial regression model shown in table 5. The table shows the two p-value one for year to be 0.02098 and the other of year square to be 0.18588. The p-value for year only is lower than the significant level of 5% but the p-value for year squared higher than the 5% significant level. So, for the polynomial model for prescription benzodiazepines you accept the null hypothesis that there is no significant difference in the means of the two groups.

Table 5: ANOVA Table for Polynomial Model – Prescription Benzodiazepines

Source	df	Sum Sq.	Mean Sq.	F value	Pr (>F)
Year	1	8960	8960.0	7.4853	0.02098
I(Year^2)	1	2415.4	2415.4	2.0178	0.18588
Residuals	10	11970.3	1197.0		

The ANOVA table for the linear regression model shown in table 6. The table shows the p-value to be 6.135e-05 this the very low p-value. So, as the p-value is less than the significant level of 5% you would reject the null hypothesis in favour of the alternative that there is a difference for street benzodiazepines.

Table 6: ANOVA Table for Linear Model – Street Benzodiazepines

Source	df	Sum Sq.	Mean Sq.	F value	Pr (>F)
Year	1	1059914	1059914	39.243	6.135e-05
Residuals	11	297102	27009		

The ANOVA table for the polynomial regression model shown in table 7. The table shows the two p-value one for year to be 2.513e-08 and the other of year square to be 1.869e-05. Both p-value are lower than the significant level of 5%. So, for the polynomial model you would reject the null hypothesis in favour of the alternative there is a difference between groups in the data for street benzodiazepines.

Table 7: ANOVA Table for Polynomial Model – Street Benzodiazepines

Source	df	Sum Sq.	Mean Sq.	F value	Pr (>F)
Year	1	1059914	1059914	240.99	2.513e-08
I(Year^2)	1	253119	253119	57.55	1.869e-05
Residuals	10	43982	4398		

The ANOVA table for the linear regression model shown in table 8. The table shows the p-value to be 2.735e-05 this the very low. So, as the p-value is less than the 5% significant level you would reject the null hypothesis in favour of the alternative. That there is a difference for methadone.

Table 8: ANOVA Table for Linear Model – Methadone

Source	df	Sum Sq.	Mean Sq.	F value	Pr (>F)
Year	1	311130	311130	47.032	2.735e-05
Residuals	11	72769	6615		

The ANOVA table for the polynomial regression model shown in table 9. The table shows the two p-value one for year to be 2.046e-07 and the other of year square to be 0.0004438 these p-vales are low meaning that they are less than the 5% significant level. So, you would reject the null hypothesis for polynomial model for methadone in favour of the alternative.

Table 9: ANOVA Table for Polynomial Model - Methadone

Source	df	Sum Sq.	Mean Sq.	F value	Pr (>F)
Year	1	311130	311130	155.305	2.046e-07
I(Year^2)	1	52735	52735	26.323	0.0004438
Residuals	10	20033	2003		

The ANOVA table for the linear regression model shown inn table 10. The table shows the p-value to be 0.00243 is a low p-value. The p-value is less that the 5% significant level then you would reject the null hypothesis in favour of the alternative. That there is a significant difference for codeine.

Table 10: ANOVA Table for Linear Model – Codeine

Source	df	Sum Sq.	Mean Sq.	F value	Pr (>F)
Year	1	1208.58	1208.58	15.296	0.00243
Residuals	11	869.12	79.01		

The ANOVA table for the polynomial regression model shown in table 11. The table shows the two p-value one for year to be 0.003091 and the other of year square to be 0.394870. Both p-value are lower than the significant level of 5%. So, for the polynomial model you would reject the null hypothesis in favour of the alternative.

Table 11: ANOVA Table for Polynomial Model - Codeine

Source	df	Sum Sq.	Mean Sq.	F value	Pr (>F)
Year	1	1208.58	1208.58	15.0049	0.003091
I(Year^2)	1	63.66	63.66	0.7904	0.394870
Residuals	10	805.45	80.55		

The ANOVA table for the linear regression model shown inn table 12. The table shows the p-value to be 0.0001356 is a very low p-value. The p-value in the linear model is less that the significant level of 5% then you would reject the null hypothesis in favour of the alternative. That there is a significant difference for cocaine.

Table 12: ANOVA Table for Linear Model – Cocaine

Source	df	Sum Sq.	Mean Sq.	F value	Pr (>F)
Year	1	188036	188036	32.638	0.0001356
Residuals	11	63373	5761		

The ANOVA table for the polynomial regression model shown in table 13. The table shows the two p-value one for year to be 7.449e-10 and the other of year square to be 1.926e-07. Both p-value are very low and therefore, the p-value is less than the 5% significant level. So, for the polynomial model you would reject the null hypothesis in favour of the alternative for cocaine.

Table 13: ANOVA Table for Polynomial Model - Cocaine

Source	df	Sum Sq.	Mean Sq.	F value	Pr (>F)
Year	1	188036	188036	496.43	7.449e-10
I(Year^2)	1	59585	59585	157.31	1.926e-07
Residuals	10	3788	379		

The ANOVA table for the linear regression model shown inn table 14. The table shows the p-value to be 0.8344 is a very high p-value. The p-value in the linear model is greater that the significant level of 5% then you would accept the null hypothesis. That there is no significant difference for alcohol.

Table 14: ANOVA Table for Linear Model – Alcohol

Source	df	Sum Sq.	Mean Sq.	F value	Pr (>F)
Year	1	38.8	38.77	0.0458	0.8344
Residuals	11	9308.0	846.18		

The ANOVA table for the polynomial regression model shown in table 15. The table shows the two p-value one for year to be 0.6576 and the other of year square to be 8.544e-05. The p-value for year only is high but the p-value for year squared is low. So, as the squared p-value is less than the significant level of 5%. So, for the polynomial model you would reject the null hypothesis in favour of the alternative for the case of alcohol.

Table 15: ANOVA Table for Polynomial Model - Alcohol

Source	df	Sum Sq.	Mean Sq.	F value	Pr (>F)
Year	1	38.8	38.8	0.2087	0.6576
I(Year^2)	1	7450.1	7450.1	40.0988	8.544e-05
Residuals	10	1857.9	185.8		

In the case for alcohol the linear best fits the data compare to the polynomial model.

3.7 – Regression Modelling on New data

Table 16 shows p-values less than 2.2e-16 for all their sources of drugs, year and the interaction of drugs and year. The p-values are all less than a 5% significant level. Therefore, you would reject null hypotheses. With the f-values of 26.257, 237.062 and 19.899 these are high values.

Table 16: ANOVA Table for linear regression model

Source	df	Sum Sq.	Mean Sq	F value	P value
Drugs	12	1942699	161892	26.257	< 2.2e-16
Year	1	1461650	1461650	237.062	< 2.2e-16
Drug: Year	12	1472315	122693	19.899	< 2.2e-16
Residuals	143	881693	6166		

Figure 15 shows the four residuals plot for the linear factorial design. Looking at the plot you see that the dataset has quite a lot of entries. The normal Q-Q plot shows the data points to be straight line between -1 and 1 but it curves down before -1 and curve up after 1. Making it seem that the data is not normally distributed. The residuals vs fitted plot shows data points to be around the 0 mark and on the red line. For the scale-location plot there is no clear patter among the residuals and that the points are randomly scatter around the red line. With a few wild points far away for the line. Plus, as the red line is horizontal across the plot then we can assume that the homoscedasticity is satisfied for the given linear model. Looking at the residual's vs leverage that there is not data point outside of the cook's distance.

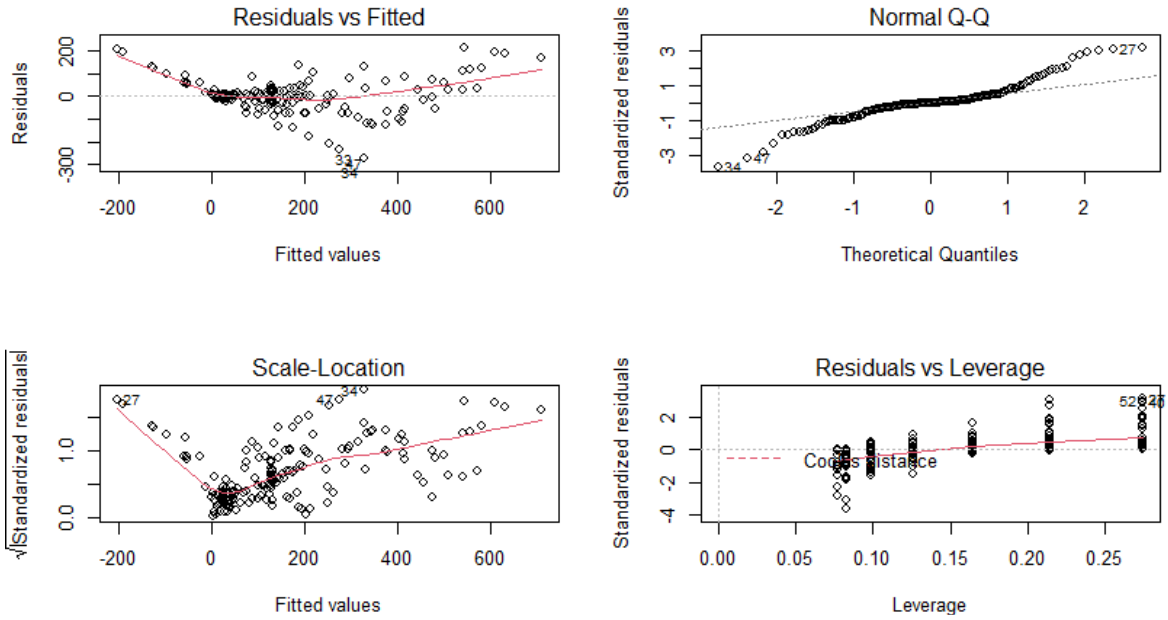

Figure 15: Residuals Plots for linear regression model

Table 17 again shows a p-values less than 2.2e-16 for all their sources of drugs, year, year squared, an interaction of drugs and year, and an interaction between drugs and year squared. Just like table 16 the p-values are all less than a 5% significant level. Therefore, you would reject null hypotheses. Both tables have very similar outputs with the only a few changes come in the f-value column of the tables. Plus, the residual value has changes with the squaring of terms to the model equation.

Table 17: ANOVA Table for polynomial regression model

Source	df	Sum Sq.	Mean Sq	F value	P value
Drugs	12	1942699	161892	136.644	< 2.2e-16
Year	1	1461650	1461650	1233.697	< 2.2e-16
I(Year^2)	1	357072	357072	301.385	< 2.2e-16
Drug: Year	12	1472315	122693	103.558	< 2.2e-16
Drug: I(Year^2)	12	370600	30883	26.067	< 2.2e-16
Residuals	130	154020	1185		

Figure 16 shows the four residuals plot for the first polynomial factorial design. Just like figure 16 there is quite a lot of entries in the dataset. The normal Q-Q plot shows the data points to

be straight line between -1 and 1 but it curves down before -1 and curve up after 1. This making the data seem as it is not normally distributed. The residuals vs fitted plot shows all of data points together at 0 with not a lot of point after 200. Looking at the plot I see there is no clear patter among the residuals and that the points are randomly scatter around the red line. With a few wild points far away for the line. Plus, as the red line is horizontal across the plot then we can assume that the homoscedasticity is satisfied for the given polynomial model. Looking at the residual's vs leverage that there is not data point outside of the cook's distance.

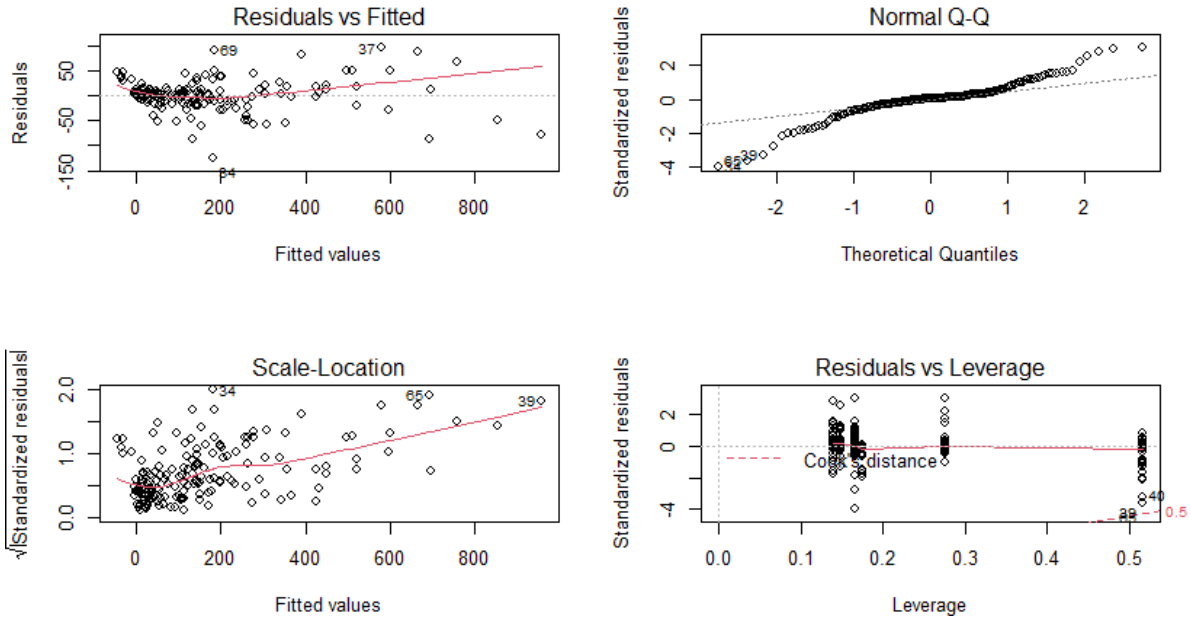

Figure 16: Residuals Plots for polynomial regression model

4 – Discussion

4. 1 – Summary of Main Finding

The first concerns in the findings are more deaths come from illegal drug over legal. This is concerning as it is unknown what is in illegal drugs. This can also cause more pain to the loved ones of the deceased person as illegal drugs will require a more in-depth post-mortem as the actual cause of death is required for death certificate.

Another observation is the data is spread randomly in all the drugs in most case. There is no general trend between any the drugs, each drug has its own independent trend. The only thing most of the drugs have in common is they all have increased in the last 13-years.

A finding in the regression models is that most of the linear model were not the best fit of the data. The polynomial model was the best fit on most of the data. This was not surprising due to the fact that the drug death has increased over the last 13 years. There was still a decrease

between 2015 to 2018. The 2020 death figures are highest, but this could be to the COVID-19 pandemic.

Another find is there is no interaction between the drug take and the year of death. This is shown in all the ANOVA table in the last sub-section of the results. This also shows the data is not normally distributed with fails one the model assumption that the data is normally distributed.

4.2 – Comparison with Literature

The published report forms the NRS looked at the drugs in three groups opioids, benzodiazepine, and substance. This document also tells us that 93% of all drug death in Scotland in 2020, they found more than one drug present in the body. With opioids drugs causing the most deaths with 89% of the total death for that year. Benzodiazepine is still high at 73% of the total deaths. These drugs are more common found based on the report. [15]

The basis of the NRS's statistics is that the drugs were implicated in or contributed to the cause of death of an individual death. Opioids where implicated deaths have remained pretty similar over the last 25 years, with the rise in deaths coming with the rise in total number of drug-relation deaths. In recent years, there has been an increase in the number of deaths in which heroin or morphine were implicated or contributed to from 345 in 2015 to 351 in 2019 and then a fall to 605 in 2020. There has also been an increase in methadone implicated deaths from 251 in 2015 to 708 in 2020. [15]

The death implicated or contributed by benzodiazepines have risen just like opioids in the last five years. From as few as 200 death per year in 2016 to nearly a stagnant 1000 death per year in 2020. This increase has come about by the street benzodiazepines being implicated in the cause of death. Street benzodiazepines cause or implicate 66% of all drug-related deaths in the last year. This has rose largely since 2015 with 58 deaths to 879. [15]

There are number of other drug-related deaths that are cause by substances that cannot be categorised into opioids or benzodiazepines. In recent years deaths implicated by gabapentin and/or pregabalin has rapidly increases from 131 in 2015 to 502 in 2020 there is 37% of all drug-related deaths. Not only has gabapentin and/or pregabalin risen cocaine deaths have also risen from 93 to 459 in the five years. This contributes to 34% of all drug-related drugs in Scotland. [15]

4.3 – National Mission

The national mission was announced by the first minister Nicola Sturgeon in January 2021. This mission is to reduce drug related deaths and harms it is supported by an additional £50 million funding per year for the next five years. The aim of the mission is to save and improve the lives through Scotland. The money will be used for faster and appropriate services for treatment and support through any services not just the NHS. As well as the improving in the frontline help for drug service third sector is also included in this. For the support to be maintain as long as you need it. Money for safe places and for working together to react immediately to overdose to help stop the death. Increase in capacity in rehabilitation facilities and a more join-up approach with Police Scotland to help reduce these deaths. With the help

from Police Scotland the Scottish Government also hope to find the underlying issue so they can address this. [16]

Actions to date that have happened include an £18 million fund that is aimed to support the national mission to reduce drug deaths in Scotland. As well as access, choice and support on the medication assisted treatment standards. The government has also detailed how the additional fund for services is going to coming and be used in 2021 and beyond. A target of £32 million of the £50 million per year will be targeted at the front-line services in the addition support fund. The final action to date is the use of digital technology to help prevent drug related deaths. The digital technology is to help keep drug-user service to lifeline connection. As social media is be used to help and prevent drug death in Scotland in the young and even older generation as social media is getting used more and more. The Scottish government are hopeful all measures will reduce the drug related death in the next five years. [16] / [17]

4.4 – Strength and Limitations

The fact that the data in this project is founded and drawn from trusted source like the national record of Scotland (NRS) and the Scottish Government this is consider a strength. As the data is also updated every year as part of the Scottish Government to cut down the drug death total in Scotland.

Another strength is that most of the information gather has come for trusted source. That fact that not source was biased in any way makes this a good source. All sources have the backing for both Scottish and UK government. Plus, the fact that I gather my support for addiction information form the NHS.

The linear and polynomial regression model in the project is a recognised statistical technique. This can all identify p-values, F-values, sum of squared, mean sum of squared and degrees of freedom this all come for the anova table. All the R code that I used is all statistical tools that are strength in the project.

The data that I used come for a big source of mini dataset but a limitation of the mini dataset in the project was the fact set was small with only 13 entries for each drug. I feel that if I have more data the results that I got could have been stronger. Another limitation is that the dataset relies on the death certificate being correct.

4.5 – Further Development

In this project I focus on linear and polynomial regression models for both datasets. For the new data that had three columns I could have used a different model. I could have used a mixed-effect model over the regression models. The mixed-effect model could have changed the outcome of the result and analysis. The model would have had two factors year and drug. Year is a time variable and is fixed were as drug is a random effect. So, a mixed effect model could have also been used to show drug-related deaths between 2008 to 2020.

4.6 – Future Investigation

Future investigation that could happen about drug-related drugs in Scotland is that you can look into gender, local or age in relation to the drug cause of deaths. There is also an interesting dataset on the cause of death such likes of accidental poisoning, drug abuse, intentional self-poisoning, and undetermined internet.

Next year data could also change the model and outcome this report as the national mission is full under way. So, 2021 drug-related data could be significantly lower than 2020. This is what the government wants so that they can shows that the funds that are given are working to reduce the unnecessary deaths in Scotland.

You could also compare the drug deaths in Scotland to the rest of the United Kingdom or Europe or even the rest of the World. As it stands today Scotland is the drug capital of the World. This is not something that the Scottish Government want and the whole aim of the national mission is so that Scotland can lose this title.

5 – Conclusion

The purpose of the study was to look at drug-related death in Scotland by using the statistical technique of regression model for both linear and polynomial. Looking at the drugs that cause the deaths in Scotland and built the model around the 13 drugs within the dataset. Compare the trends between all the drugs within the dataset. As well as comparing both models and looking at the slopes of the data on plots.

References

[1] "Commonly Used Drugs Charts", August 20, 2020 [Online Source]. Website Address: https://www.drugabuse.gov/drug-topics/commonly-used-drugs-charts. Accessed: October 2021

[2] "Types of drugs" [Online Source]. Website Address: https://www.catalystsupport.org.uk/types-of-drugs/. Accessed: October 2021

[3] "How the spinal cord works" [Online Source]. Website Address: https://www.christopherreeve.org/living-with-paralysis/health/how-the-spinal-cord-works#:~:text=Intervention%20strategies-,What%20is%20the%20central%20nervous%20system%3F,of%20control%20over%20body%20movement. Accessed: October 2021

[4] "Drugs penalties" [Online Source]. Website Address: https://www.gov.uk/penalties-drug-possession-dealing. Accessed: October 2021

[5] "Addiction: what is it?", 9 June 2021 [Online Source]. Website Address: https://www.nhs.uk/live-well/healthy-body/addiction-what-is-it/ . Accessed: October 2021

[6] "Suspected drug deaths in Scotland: April to June 2021", 14 September 2021 [Online Source]. Website Address: https://www.gov.scot/publications/suspected-drug-deaths-scotland-april-june-2021/ . Accessed: September 2021

[7] G. Devault, “What A Simple Linear Regression Model Is and How It Works,” 9 January 2020. [Online Source]. Website Address: https://www.thebalancesmb.com/what-is-simple-linear-regression-2296697. Accessed: October 2021

[8] Zach, “Introduction to Simple Linear Regression.”, 10 December 2018 [Online Source]. Website Address: https://www.statology.org/linear-regression/ . Accessed: October 2021

[9] R. Agrawal, “All you need to know about Polynomial Regression,” 9 July 2021 [Online Source]. Website Address: https://www.analyticsvidhya.com/blog/2021/07/all-you-need-to-know-about-polynomial-regression/ . Accessed: October 2021

[10] I. M. Azubuike, “Second Order Regression with Two Predictor Variables Centered on Mean in an Ill Conditioned Model”, 2019 [Online Source]. Website Address: http://article.sapub.org/10.5923.j.statistics.20190904.01.html . Accessed: November 2021

[11] Abhigyan, “Understanding Polynomial Regression!!!”, 2 August 2020 [Online Source]. Website Address: https://medium.com/analytics-vidhya/understanding-polynomial-regression-5ac25b970e18. Accessed: November 2021

[12] Zach, “Understanding the Null Hypothesis for Linear Regression”, 14 May 2021. [Online Source]. Website Address: https://www.statology.org/null-hypothesis-for-linear-regression/. Accessed: November 2021

[13] R. Bevans, “An introduction to the two-way ANOVA”, 7 January 2021 [Online Source]. Website Address: https://www.scribbr.com/statistics/two-way-anova/. Accessed: November 2021

[14] “One-Way ANOVA” [Online Source]. Website Address: https://statistics.laerd.com/statistical-guides/one-way-anova-statistical-guide.php. Accessed: November 2021

[15] F Dixon, “Drug-related death in Scotland 2020”, 30 July 2021, © Crown Copyright [Online Report]. Link to Report: https://www.nrscotland.gov.uk/files//statistics/drug-related-deaths/20/drug-related-deaths-20-pub.pdf. Accessed: November 2021

[16] “Alcohol and drugs” [Online Source]. Website Addressed: https://www.gov.scot/policies/alcohol-and-drugs/national-mission/. Accessed: November 2021

[17] “National drugs mission: additional funds for services – 2021 to 2022”, 17 June 2021 [Online Source]. Website Address: https://www.gov.scot/publications/national-drugs-mission-additional-funds-for-services-2021-to-2022/. Accessed: November 2021

Appendix

```
# Set Working Directory

#Open data and view

drug.deaths <- read.csv("Drug Related Death Data/Used/Drug-related deaths in total with all data.csv")

View(drug.deaths)

summary(drug.deaths)

#plot another graph with all drug on it

plot(drug.deaths$Year, drug.deaths$All.drug.related.deaths, type = "l",

    lty = 1, col = "cornflowerblue", ylim = c(0, 1350),

    xlab = "Year", ylab = "Number of Deaths",

    main = "Total Number of all Drugs between 2008 to 2020")

lines(drug.deaths$Year, drug.deaths$Prescription.Benzodiazepines, type = "l",

     lty = 2, col = "darkgoldenrod")

lines(drug.deaths$Year, drug.deaths$Diazepam, type = "l",

     lty = 3, col = "cyan1")

lines(drug.deaths$Year, drug.deaths$Street.Benzodiazepines, type = "l",

     lty = 4, col = "brown1")

lines(drug.deaths$Year, drug.deaths$Etizolam, type = "l",

     lty = 5, col = "chartreuse")

lines(drug.deaths$Year, drug.deaths$Heroin...morphine, type = "l",

     lty = 6, col = "darkmagenta")

lines(drug.deaths$Year, drug.deaths$Methadone, type = "l",

     lty = 7, col = "darkorange2")

lines(drug.deaths$Year, drug.deaths$Codeine.or.a.compound, type = "l",

     lty = 8, col = "dodgerblue3")

lines(drug.deaths$Year, drug.deaths$Dihydrocodeine.or.a.compound, type = "l",

     lty = 9, col = "gold")

lines(drug.deaths$Year, drug.deaths$Gabapentin.and.or.pregabalin, type = "l",

     lty = 10, col = "deeppink2")
```

```
lines(drug.deaths$Year, drug.deaths$Cocaine, type = "l",
    lty = 11, col = "gray0")
lines(drug.deaths$Year, drug.deaths$Ecstasy.type.drugs, type = "l",
    lty = 12, col = "darkturquoise")
lines(drug.deaths$Year, drug.deaths$Alcohol, type = "l",
    lty = 13, col = "darkslategrey")
lines(drug.deaths$Year, drug.deaths$Amphetamines, type = "l",
    lty = 14, col = "lightcoral")
legend("topleft",
     legend=c("Any", "Prescription Benz", "Diazepam", "Street Benz",
           "Etizolam", "Herion or Morphine", "Methadone",
           "Codeine", "Dihyrocodeine", "Gabapentin or Pregablin",
           "Cocaine", "Ecstasty", "Alcohol", "Amphetamines"),
     col=c("cornflowerblue", "darkgoldenrod", "cyan1", "brown1",
         "chartreuse","darkturquoise", "darkorange2", "dodgerblue3",
         "gold", "deeppink2", "gray0", "darkslategrey", "lightcoral"),
     lty = 1:14, cex=0.5)
# plot the drugs that are Benzodiazepines
plot(drug.deaths$Year, drug.deaths$All.drug.related.deaths, type = "l",
   lty = 1, col = "cornflowerblue", xlab = "Year", ylab = "Number of Deaths",
   main = "Deaths by Benzodiazepines from 2008 to 2020", ylim = c(0, 1350))
lines(drug.deaths$Year, drug.deaths$Prescription.Benzodiazepines, type = "l",
    lty = 2, col = "deeppink")
lines(drug.deaths$Year, drug.deaths$Diazepam, type = "l",
    lty = 2, col = "chartreuse")
lines(drug.deaths$Year, drug.deaths$Street.Benzodiazepines, type = "l",
    lty = 3, col = "darkorange2")
lines(drug.deaths$Year, drug.deaths$Etizolam, type = "l",
    lty = 2, col = "cyan1")
legend("topleft",
     legend=c("All", "Prescription", "Diazepam", "Street", "Etizolam"),
```

```
      col=c("cornflowerblue", "deeppink", "chartreuse","darkorange2",
          "cyan1"), lty = 1:4, cex=0.8)
# plot the drugs that are Optioids
plot(drug.deaths$Year, drug.deaths$All.drug.related.deaths, type = "l",
    lty = 1, col = "cornflowerblue", xlab = "Year", ylab = "Number of Deaths",
    main = "Deaths by Optioids from 2008 to 2020", ylim = c(0, 1350))
lines(drug.deaths$Year, drug.deaths$Heroin...morphine, type = "l",
     lty = 2, col = "deeppink")
lines(drug.deaths$Year, drug.deaths$Methadone, type = "l",
     lty = 3, col = "chartreuse")
lines(drug.deaths$Year, drug.deaths$Codeine.or.a.compound, type = "l",
     lty = 4, col = "darkorange2")
lines(drug.deaths$Year, drug.deaths$Dihydrocodeine.or.a.compound, type = "l",
     lty = 5, col = "cyan1")
legend("topleft",
      legend=c("All", "Heroin or Morphine", "Methadone", "Codeine or Comp",
           "Dihydricodeine or comp"),
      col=c("cornflowerblue", "deeppink", "chartreuse",
          "darkorange2", "cyan1"), lty = 1:5, cex=0.8)
# plot the graph of substances
plot(drug.deaths$Year, drug.deaths$All.drug.related.deaths, type = "l",
    lty = 1, col = "cornflowerblue", xlab = "Year",
    ylab = "Number of Deaths",
    main = "Deaths by any substances between 2008 and 2020", ylim = c(0,1350))
lines(drug.deaths$Year, drug.deaths$Gabapentin.and.or.pregabalin,
     type = "l" , lty = 2, col = "deeppink")
lines(drug.deaths$Year, drug.deaths$Cocaine, type = "l", lty = 3,
     col = "chartreuse")
lines(drug.deaths$Year, drug.deaths$Ecstasy.type.drugs, type = "l", lty = 4,
     col = "darkorange2")
lines(drug.deaths$Year, drug.deaths$Amphetamines, type = "l", lty = 5,
```

```
    col = "cyan1")
lines(drug.deaths$Year, drug.deaths$Alcohol, type = "l",
    lty = 6, col = "darkmagenta")
legend("topleft",
     legend=c("All", "Gabapentin or Pregabalin", "Cocanie", "Ecstasy",
          "Amphetamines", "Alcohol"),
     col = c("cornflowerblue", "deeppink", "chartreuse",
         "darkorange2", "cyan1", "darkmagenta"),
     lty = 1:6, cex = 0.75)
# plot the graph identifying legal and illegal
plot(drug.deaths$Year, drug.deaths$All.drug.related.deaths, type = "l",
   lty = 1, col = "cornflowerblue", xlab = "Year",
   ylab = "Number of Deaths", ylim = c(0, 1350),
   main = "Legal and Illegal Drugs Deaths between 2008 to 2020")
lines(drug.deaths$Year, drug.deaths$Prescription.Benzodiazepines, type = "l",
    lty = 2, col = "charteuse")
lines(drug.deaths$Year, drug.deaths$Diazepam, type = "l",
    lty = 2, col = "chartreuse")
lines(drug.deaths$Year, drug.deaths$Street.Benzodiazepines, type = "l",
    lty = 3, col = "deeppink4")
lines(drug.deaths$Year, drug.deaths$Etizolam, type = "l",
    lty = 2, col = "chartreuse")
lines(drug.deaths$Year, drug.deaths$Heroin...morphine, type = "l",
    lty = 3, col = "deeppink4")
lines(drug.deaths$Year, drug.deaths$Methadone, type = "l",
    lty = 3, col = "deeppink4")
lines(drug.deaths$Year, drug.deaths$Codeine.or.a.compound, type = "l",
    lty = 3, col = "deeppink4")
lines(drug.deaths$Year, drug.deaths$Dihydrocodeine.or.a.compound, type = "l",
    lty = 2, col = "chartreuse")
lines(drug.deaths$Year, drug.deaths$Gabapentin.and.or.pregabalin, type = "l",
```

```
    lty = 2, col = "chartreuse")
lines(drug.deaths$Year, drug.deaths$Cocaine, type = "l",
    lty = 3, col = "deeppink4")
lines(drug.deaths$Year, drug.deaths$Ecstasy.type.drugs, type = "l",
    lty = 3, col = "deeppink4")
lines(drug.deaths$Year, drug.deaths$Alcohol, type = "l",
    lty = 2, col = "chartreuse")
lines(drug.deaths$Year, drug.deaths$Amphetamines, type = "l",
    lty = 3, col = "deeppink4")
legend("topleft",
      legend=c("Any", "Legal", "Illegal"),
      col=c("cornflowerblue", "chartreuse", "deeppink4"),
      lty = 1:3, cex=0.9)
# Linear model All drugs deaths and plot
z1 <- lm(All.drug.related.deaths ~ Year, data = drug.deaths)
summary(z1)
z1a <- lm(All.drug.related.deaths ~ Year + I(Year^2), data = drug.deaths)
summary(z1a)
anova(z1)
anova(z1a)
plot(drug.deaths$Year, drug.deaths$All.drug.related.deaths, type = "b",
    col = "darkblue", xlab = "Year", ylab = "Number of Deaths",
    main = "All Drug Deaths in Scotland between 2008 to 2020",
    ylim = c(250,1350))
abline(z1, col = "green")
newx <- seq(2008, 2020, by = .1)
newy1 <- predict(z1a, data.frame(Year = newx))
lines(newx, newy1, col = "deeppink")
# new par
par(mfrow = c(2,2))
# plot the residual plot for both models
```

```
plot(z1)
plot(z1a)
# Linear Model Prescription Benzodiazepines
z2 <- lm(Prescription.Benzodiazepines ~ Year, data = drug.deaths)
summary(z2)
z2a <- lm(Prescription.Benzodiazepines ~ Year + I(Year^2), data = drug.deaths)
summary(z2a)
anova(z2)
anova(z2a)
plot(drug.deaths$Year, drug.deaths$Prescription.Benzodiazepines, type = "b",
    col = "darkblue", xlab = "Year", ylab = "Number of Deaths",
    main = "Prescription Benzodiazepines", ylim = c(50,350))
abline(z2, col = "green")
newy2 <- predict(z2a, data.frame(Year = newx))
lines(newx, newy2, col = "deeppink")
# Linear model Diazepam
z3 <- lm(Diazepam ~ Year, data = drug.deaths)
summary(z3)
z3a <- lm(Diazepam ~ Year + I(Year^2), data = drug.deaths)
summary(z3a)
anova(z3, z3a)
plot(drug.deaths$Year, drug.deaths$Diazepam, type = "b", col = "darkblue",
    xlab = "Year", ylab = "Number of Deaths", main = "Diazepam",
    ylim = c(0, 300))
abline(z3, col = "green")
newy3 <- predict(z3a, data.frame(Year = newx))
lines(newx, newy3, col = "deeppink")
# Linear Model Street Benzodiazepines
z4 <- lm(Street.Benzodiazepines~ Year, data = drug.deaths)
summary(z4)
z4a <- lm(Street.Benzodiazepines ~ Year + I(Year^2), data = drug.deaths)
```

```
summary(z4a)
anova(z4)
anova(z4a)
plot(drug.deaths$Year, drug.deaths$Street.Benzodiazepines, type = "b",
    col = "darkblue", xlab = "Year", ylab = "Number of Deaths",
    main = "Street Benzondiazepines", ylim = c(-100, 1000))
abline(z4, col = "green")
newy4 <- predict(z4a, data.frame(Year = newx))
lines(newx, newy4, col = "deeppink")
# Linear Model Etizolam
z5 <- lm(Etizolam ~ Year, data = drug.deaths)
summary(z5)
z5a <- lm(Etizolam ~ Year + I(Year^2), data = drug.deaths)
summary(z5a)
anova(z5, z5a)
plot(drug.deaths$Year, drug.deaths$Etizolam, type = "b", col = "darkblue",
    xlab = "Year", ylab = "Number of Deaths",
    main = "Etizolam", ylim = c(-50, 900))
abline(z5, col = "green")
newy5 <- predict(z5a, data.frame(Year = newx))
lines(newx, newy5, col = "deeppink")
# Linear Model Heroin and/or Morphine
z6 <- lm(Heroin...morphine ~ Year, data = drug.deaths)
summary(z6)
z6a <- lm(Heroin...morphine ~ Year + I(Year^2), data = drug.deaths)
summary(z6a)
anova(z6)
anova(z6a)
plot(drug.deaths$Year, drug.deaths$Heroin...morphine, type = "b",
    col = "darkblue", xlab = "Year", ylab = "Number of Deaths",
    main = "Heroin and Morphine", ylim = c(50, 700))
```

```
abline(z6, col = "green")
newy6 <- predict(z6a, data.frame(Year = newx))
lines(newx, newy6, col = "deeppink")
# Linear Model Methadone
z7 <- lm(Methadone ~ Year, data = drug.deaths)
summary(z7)
z7a <- lm(Methadone ~ Year + I(Year^2), data = drug.deaths)
summary(z7a)
anova(z7)
anova(z7a)
plot(drug.deaths$Year, drug.deaths$Methadone, type = "b",
    col = "darkblue", xlab = "Year", ylab = "Number of Deaths",
    main = "Methadone", ylim = c(50, 790))
abline(z7, col = "green")
newy7 <- predict(z7a, data.frame(Year = newx))
lines(newx, newy7, col = "deeppink")
# Linear Model Codeine
z8 <- lm(Codeine.or.a.compound ~ Year, data = drug.deaths)
summary(z8)
z8a <- lm(Codeine.or.a.compound ~ Year + I(Year^2), data = drug.deaths)
summary(z8a)
anova(z8)
anova(z8a)
plot(drug.deaths$Year, drug.deaths$Codeine.or.a.compound, type = "b",
    col = "darkblue", xlab = "Year", ylab = "Number of Deaths",
    main = "Codeine", ylim = c(-50, 120))
abline(z8, col = "green")
newy8 <- predict(z8a, data.frame(Year = newx))
lines(newx, newy8, col = "deeppink")
# Linear Model Dihydrocodeine
z9 <- lm(Dihydrocodeine.or.a.compound ~ Year, data = drug.deaths)
```

```
summary(z9)
z9a <- lm(Dihydrocodeine.or.a.compound ~ Year + I(Year^2), data = drug.deaths)
summary(z9a)
anova(z9, z9a)
plot(drug.deaths$Year, drug.deaths$Dihydrocodeine.or.a.compound, type = "b",
    col = "darkblue", xlab = "Year", ylab = "Number of Deaths",
    main = "Dihydrocodeine", ylim = c(50, 210))
abline(z9, col = "green")
newy9 <- predict(z9a, data.frame(Year = newx))
lines(newx, newy9, col = "deeppink")
# Linear Model Gabapentin and Pregabalin
z10 <- lm(Gabapentin.and.or.pregabalin ~ Year, data = drug.deaths)
summary(z10)
z10a <- lm(Gabapentin.and.or.pregabalin ~ Year + I(Year^2), data = drug.deaths)
summary(z10a)
anova(z10, z10a)
plot(drug.deaths$Year, drug.deaths$Gabapentin.and.or.pregabalin, type = "b",
    col = "darkblue", xlab = "Year", ylab = "Number of Deaths",
    main = "Gabapentin and Pregabalin", ylim = c(-50, 600))
abline(z10, col = "green")
newy10 <- predict(z10a, data.frame(Year = newx))
lines(newx, newy10, col = "deeppink")
# Linear Model Cocaine
z11 <- lm(Cocaine ~ Year, data = drug.deaths)
summary(z11)
z11a <- lm(Cocaine ~ Year + I(Year^2), data = drug.deaths)
summary(z11a)
anova(z11)
anova(z11a)
plot(drug.deaths$Year, drug.deaths$Cocaine, type = "b",
    col = "darkblue", xlab = "Year", ylab = "Number of Deaths",
```

```
    main = "Cocaine", ylim = c(-50, 500))
abline(z11, col = "green")
newy11 <- predict(z11a, data.frame(Year = newx))
lines(newx, newy11, col = "deeppink")
#Linear Model Ecstasy
z12 <- lm(Ecstasy.type.drugs ~ Year, data = drug.deaths)
summary(z12)
z12a <- lm(Ecstasy.type.drugs ~ Year + I(Year^2), data = drug.deaths)
summary(z12a)
anova(z12, z12a)
plot(drug.deaths$Year, drug.deaths$Ecstasy.type.drugs, type = "b",
    col = "darkblue", xlab = "Year", ylab = "Number of Deaths",
    main = "Ecstasy", ylim = c(-50, 50))
abline(z12, col = "green")
newy12 <- predict(z12a, data.frame(Year = newx))
lines(newx, newy12, col = "deeppink")
# Linear Model Amphetamines
z13 <- lm(Amphetamines ~ Year, data = drug.deaths)
summary(z13)
z13a <- lm(Amphetamines ~ Year + I(Year^2), data = drug.deaths)
summary(z13a)
anova(z13, z13a)
plot(drug.deaths$Year, drug.deaths$Amphetamines, type = "b",
    col = "darkblue", xlab = "Year", ylab = "Number of Deaths",
    main = "Amphetamines", ylim = c(-50, 100))
abline(z13, col = "green")
newy13 <- predict(z13a, data.frame(Year = newx))
lines(newx, newy13, col = "deeppink")
# Linear Model Alcohol
z14 <- lm(Alcohol ~ Year, data = drug.deaths)
summary(z14)
```

```
z14a <- lm(Alcohol ~ Year + I(Year^2), data = drug.deaths)
summary(z14a)
anova(z14)
anova(z14a)
plot(drug.deaths$Year, drug.deaths$Alcohol, type = "b", col = "darkblue",
    xlab = "Year", ylab = "Number of Deaths",
    main = "Alcohol", ylim = c(50, 200))
abline(z14, col = "green")
newy14 <- predict(z14a, data.frame(Year = newx))
lines(newx, newy14, col = "deeppink")
# par back to (1,1)
par(mfrow = c(1,1))

# Open three column dataset
drug.deaths.3 <- read.csv("Drug Related Death Data/Used/Drug Death Data.csv")
View(drug.deaths.3)
summary(drug.deaths.3)
# Factor Drug
drug.deaths.3$Drugs <- factor(drug.deaths.3$Drugs)
summary(drug.deaths.3)
# contrasts sum
contrasts(drug.deaths.3$Drugs) <- contr.sum(length(levels(drug.deaths.3$Drugs)))
drug.deaths.3$Drugs
# boxplot
boxplot(DeathNumbers ~ Year, data = drug.deaths.3, xlab = "Year",
      ylab = "Deaths Numbers", col = (c("cornflowerblue")), las = 2)
boxplot(DeathNumbers ~ Drugs, xlab = '', data = drug.deaths.3,
      col = (c("deeppink")), las = 2, cex.axis = 0.55)
# Two-way Model
z0.1 <- lm(DeathNumbers ~ Drugs + Year + Drugs:Year, data = drug.deaths.3)
summary(z0.1)
```

```
anova(z0.1)

z0.2 <- lm(DeathNumbers ~ Drugs + Year + Drugs:Year + I(Year^2), data = drug.deaths.3)

summary(z0.2)

anova(z0.2)

z0.3 <- lm(DeathNumbers ~ Drugs + Year + Drugs:Year + I(Year^2) + Drugs:I(Year^2), data =
drug.deaths.3)

summary(z0.3)

anova(z0.3)

# Plot Residuals of the three model

par(mfrow = c(2,2))

plot(z0.1)

plot(z0.2)

plot(z0.3)
```