PROJECT REPORT ON INFERENTIAL STATISTICS ON POSSIBLE ASTEROID IMPACTS WITH EARTH

TEAM MEMBERS :

AITHA TARUN (18MIS7007)

BHARGAV KUMAR YAMJALA (18MIS7282)

JASWANTH (18MIS7059)

NEKKANTI SAI VENKATESH (18MIS7132)

V SRIDATTA VARMA (18MIS7125)

P.V THARUN (18MIS7028)

COURSE CODE : MAT1014

SLOT : TE2

INTRODUCTION

Sentry is a highly automated collision monitoring system that continually scans the most current asteroid catalogue for possibilities of future impact with Earth over the next 100 years. Whenever a potential impact is detected it will be analysed and the results immediately will be added to the dataset, except in unusual cases where we seek independent confirmation.

Objects normally appear to risk because their orbits can bring them close to the Earth's orbit and the limited number of available observations do not yet allow their trajectories to be well-enough defined. In such cases, there may be a wide range of possible future paths that can be fit to the existing observations, sometimes including a few that can intersect the Earth.

Every day, observations and orbit solutions for Near-Earth Asteroids (NEAs) are received from the Minor Planet Centre (MPC) in Cambridge, Massachusetts. Once classified as an NEA, the asteroid is thereafter given automatic orbit updates within the Sentry system. A new orbit solution for an NEA is computed whenever new optical or radar observations for that object become available. Some high-priority objects are observed daily, while other objects go unobserved for days or weeks, even though they may still be bright enough to be seen. Optical observations cease when an object recedes from the Earth (becoming too faint to be seen even with moderate-size telescopes), or when the object moves into the daytime sky. Similarly, radar observations are possible only when the object is near enough to the Earth for the echo of a radar bounce to be detected. Once all the observations for an object have been collected, an orbit determination process is used to find the orbit that best fits all the observations.

About Data Set :

Object Designation :

Given name or number to the specific celestial object.

Period Start :

Starting year of object impact detected.

Period Start :

Ending year of object impact detected.

Potential Impacts :

Number of dynamically distinct potential impacts that have been detected by Sentry.

Cumulative Impact Probability :

Sum of the impact probabilities from all detected potential impacts.

Asteroid Velocity(km/s) :

Velocity of the asteroid relative to the Earth.

Asteroid Magnitude (mag) :

Absolute Magnitude, a measure of intrinsic brightness. It is the apparent magnitude of the object when it is 1 au from both the sun and the observer, and at full phase for the observer.

Asteroid Diameter (km) :

Estimated diameter of the asteroid.

Cumulative Palermo Scale :

Cumulative hazard rating according to the Palermo technical impact hazard scale, based on the tabulated impact date, impact probability and impact energy.

Maximum Palermo Scale :

Maximum hazard rating according to the Palermo technical impact hazard scale, based on the tabulated impact date, impact probability and impact energy.

Maximum Torino Scale :

Maximum detected hazard rating according to the Torino impact hazard scale, based on the tabulated impact

probability and impact energy. The Torino scale is defined only for potential impacts less than 100 years in the future.

METHODOLOGY

Statistical methodology, with deep roots in probability theory, provides quantitative procedures for extracting scientific knowledge from astronomical data and for testing astrophysical theory. Nonparametric methods are valuable when little is known about the behaviour of the astronomical populations or processes. Data smoothing can be achieved with kernel density estimation and nonparametric regression. Samples measured in many variables can be divided into distinct groups using unsupervised clustering or supervised classification procedures.

Regression analysis is a set of statistical methods used for the estimation of relationships between a dependent variable and one or more independent variables. It can be utilized to assess the strength of the relationship between variables and for modelling the future relationship between them.

The day to day business of science consists of gathering data and consolidating existing knowledge. A hypothesis is a proposed explanation for a set of measurements. The mathematical formulation of the hypothesis is called a model. Scientists test hypotheses by acquiring data of more and more scope and accuracy.

Asteroid impact prediction is the prediction of the dates and times of asteroids impacting Earth, along with the locations and severities of the impacts.

The process of impact prediction follows three major steps:

I. **Discovery of an asteroid and initial assessment of its orbit which is generally based on a short observation arc of less than 2 weeks.**
II. **Follow up observations to improve the orbit determination**
III. **Calculating if, when and where the orbit may intersect with Earth at some point in the future.**

Sample data :

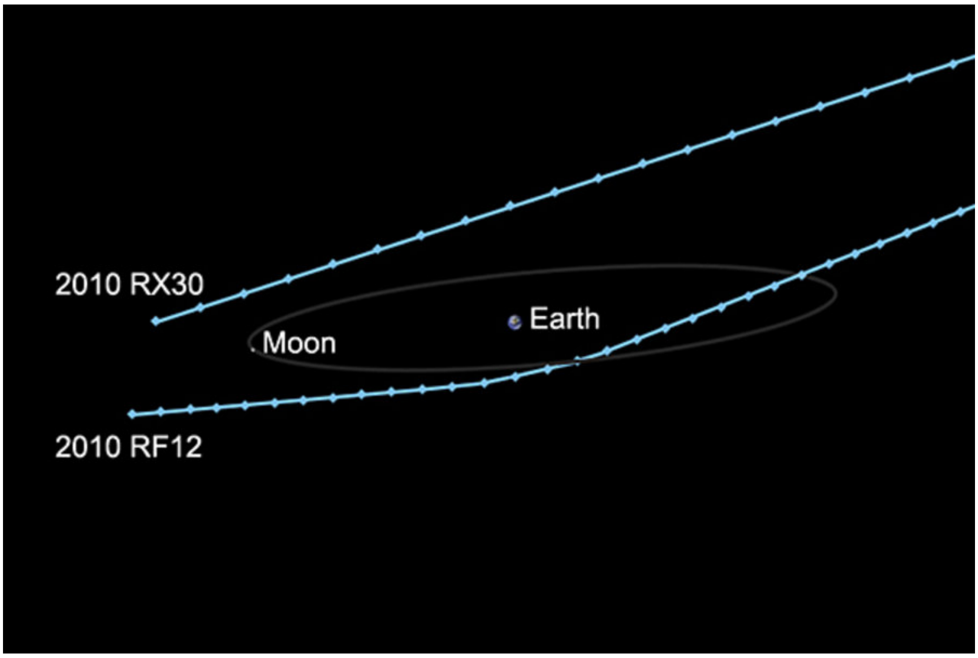
	Object.Name	Period.Start	Period.End	Possible.Impacts	Cumulative.Impact.Probability	Asteroid.Velocity	Asteroid.Magnitude	Asteroid.Diameter..km.	Cumulative.Palermo.Scale	Maximum.Palermo.Scale	Maximum.Torino.Scale
1	2006 WP1	2017	2017	1	5.20000e-09	17.7700	28.3000	0.00700000	-8.31000	-8.31000	0
2	2013 YB	2017	2046	23	7.60000e-05	8.98000	31.4000	0.00200000	-6.60000	-6.96000	0
3	2008 US	2017	2062	30	1.60000e-05	18.3300	31.4000	0.00200000	-6.48000	-6.87000	0
4	2010 VR139	2017	2076	24	2.00000e-07	4.99000	26.7000	0.0160000	-6.83000	-6.95000	0
5	2015 ME131	2017	2096	85	2.30000e-08	19.4600	19.2000	0.497000	-3.85000	-4.30000	0
6	2010 XB73	2017	2110	55	2.80000e-07	5.98000	22.5000	0.110000	-5.03000	-5.51000	0
7	2005 TM173	2017	2111	123	9.00000e-07	8.79000	24.1000	0.0520000	-5.41000	-6.42000	0
8	2006 SF281	2017	2111	514	1.90000e-06	4.49000	26.2000	0.0200000	-5.91000	-7.58000	0
9	2010 VP139	2017	2112	350	1.80000e-05	2.04000	28.5000	0.00700000	-6.51000	-7.54000	0
10	2014 HR197	2017	2112	187	7.90000e-08	17.0200	26.9000	0.0140000	-7.18000	-7.97000	0
11	2015 HV182	2017	2113	509	7.10000e-07	7.78000	21.7000	0.153000	-4.14000	-5.45000	0
12	2008 SH148	2017	2113	163	8.10000e-07	9.47000	26.1000	0.0200000	-6.57000	-7.46000	0
13	2008 XK	2017	2113	171	6.10000e-07	9.03000	27.2000	0.0120000	-6.85000	-7.50000	0
14	2010 XC	2017	2113	332	3.30000e-06	3.96000	29.1000	0.00500000	-7.25000	-8.05000	0
15	2010 MY112	2017	2114	440	7.00000e-06	2.10000	25.5000	0.0290000	-4.90000	-5.95000	0
16	2014 MO68	2017	2114	262	1.50000e-06	8.36000	23.5000	0.0670000	-5.06000	-6.04000	0
17	2009 FZ4	2017	2114	434	2.90000e-06	10.1100	25.2000	0.0310000	-5.21000	-6.03000	0
18	2008 VS4	2017	2114	300	5.90000e-07	8.31000	24.1000	0.0510000	-5.44000	-6.60000	0
19	2009 VZ39	2017	2114	924	7.10000e-06	6.33000	27.9000	0.00900000	-6.22000	-7.61000	0
20	2014 JT79	2017	2114	861	1.00000e-06	12.0200	26.6000	0.0160000	-6.23000	-7.31000	0
21	2008 VL	2017	2114	615	1.30000e-05	9.68000	27.9000	0.00900000	-6.33000	-7.46000	0
22	2010 WW8	2017	2114	92	5.70000e-07	4.45000	26.9000	0.0150000	-7.02000	-7.73000	0
23	2008 EM68	2017	2115	1144	1.30000e-05	14.5400	27.8000	0.0100000	-5.31000	-5.65000	0
24	2015 HW182	2017	2115	174	5.70000e-08	10.7600	25.2000	0.0310000	-6.69000	-7.17000	0
25	2012 BP123	2017	2115	228	5.90000e-07	4.60000	27.1000	0.0130000	-7.24000	-8.11000	0

FINDINGS

We have used CRAN R to analyse and get results from the dataset provided by sentry on asteroid impacts.

From the given dataset of impacts of celestial objects on earth we can find that top 3 asteroids to impact the earth by the given probability : 2010 RF 12 with a probability of 0.065, 2006 JU26 with a probability of 0.0097, and 2000 SG344 with a probability of 0.002. Also, astronomers used this data to also find the impact dates which includes news like; The European Space Agency (ESA) Has Detected An Asteroid Named As '2010 RF12' Which Is All Set To Fly Dangerously Close To Earth On September 23, 2022, Scientists have put a date to Armageddon. It will occur on 21 September 2030, when earth is in danger of being hit by an asteroid.

The newly discovered threat to global civilisation is called 2000 SG344, and it could strike our planet with a force 100 times greater than that released by the atom bomb that destroyed Hiroshima, astronomers have calculated.



Also, we can find the asteroids which have a greater number of impacts with earth for the next 100 years, from the data we can find that asteroids 2008 EM68, 2009 VZ39, 2014 JT79 are top 3 asteroids with the greatest number of impacts with earth. But with less probability of impacts. **Although the odds of any one particular asteroid ever impacting Earth are quite low, it is still likely that one day our planet will be hit by another asteroid.**

When it comes to diameters of asteroids some of asteroids can be a length of more than 900 kms like ceres in the asteroid belt between Mars and Jupiter which can impact earth and can wipe out a large amount of landmass on earth, on this basis we can find the large asteroids which have probability to impact earth, they includes asteroids like 2011 SR52 of diameter 2.5 km, 2010 AU118 of diameter 1.9 km and 29075 (1950 DA) of diameter 1.3 km.

Also, we can know the near future asteroids to impact earth from this given data , by analysing asteroids 2001 YN2, 2001 UO and 2008 VM can have a probable impact with earth in the current year 2020.

With correlation concept we can find how strongly the variables of dataset are related . In the preferred dataset we can find correlation of the variables using R to analyse how strongly they are related to each other. In dataset we found the correlations between Impact Probability, Number Of Impacts, Asteroid Diameter and its Velocity. And found that every variable in the given dataset is weekly related to each other. That is absolute of correlation co-efficient is in between 0 and 0.5.

With linear regression we can model the relationship between dependent variable and independent variable. In CRAN R we can find linear model with function lm() and we can predict values of a variable with respective to another variable with predict function.

When performing any regression methods, we can find co-efficient Pr(>|t|) found in the model output relates to the probability of observing any value equal or larger than *t*. Where co-efficient t-value is a measure of how many standard deviations our coefficient estimate is far away from 0. We want it to be far away from zero as this would indicate we could reject the null hypothesis - that is, we could declare a relationship between the variables. Typically, a p-value of 5% or less is a good cut-off point. A small p-value for the intercept and the slope indicates that we can reject the null hypothesis which allows us to conclude that there is a relationship between the variables.

And the response variables with less significant value(Pr(>|t|)) that is equal or less than 0.05 posses much impact on the prediction of predictor variables. In the given data Number of Impacts can be predicted with Asteroid velocity, Asteroid diameter can be predicted by Asteroid velocity and Asteroid velocity can be predicted by both possible impacts and asteroid diameter.

We have done prediction of possible impacts with linear regression with asteroid velocity of user input, for example we could consider

it has a value of 11 and we get predicted number of impacts is 44.5354.

For predicting asteroid diameter, we could use linear regression with asteroid velocity value form user, for example we could consider it has a value of 11 and we get predicted asteroid diameter is 0.04729264 km.

Where has for predicting asteroid velocity, we should use multiple regression with values of possible impacts and asteroid diameter from user, for example here we can consider it values 43 and 0.04 respectively and we get predicted asteroid velocity is 11.40301.

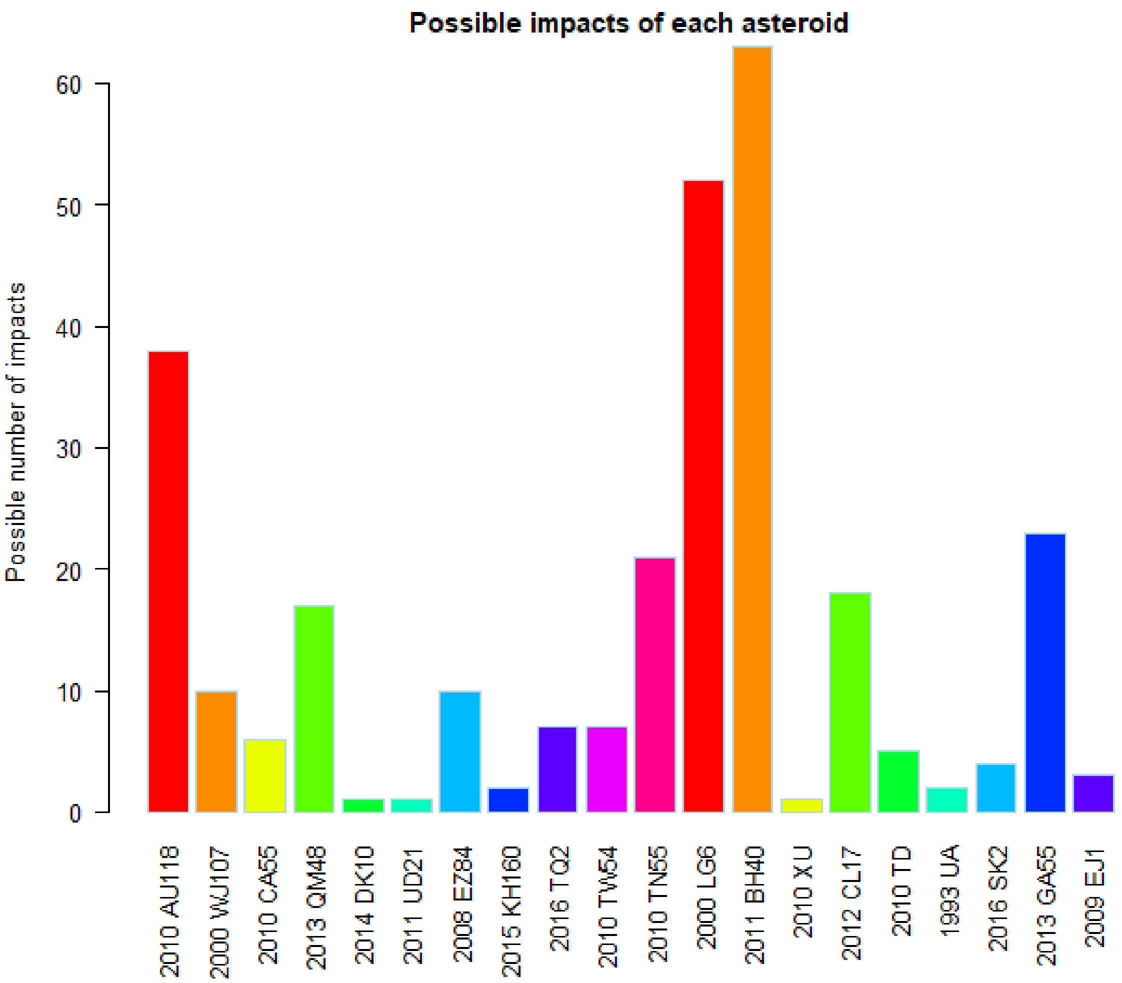
Exploratory Data Analysis :

With exploratory data analysis we can analyse the data with plot and graphs for better understanding and we can easily make hypothesis by observing these plots.

For plotting graphs, we could select a sample from the population with function sample in R. The selected sample for this analysis is :

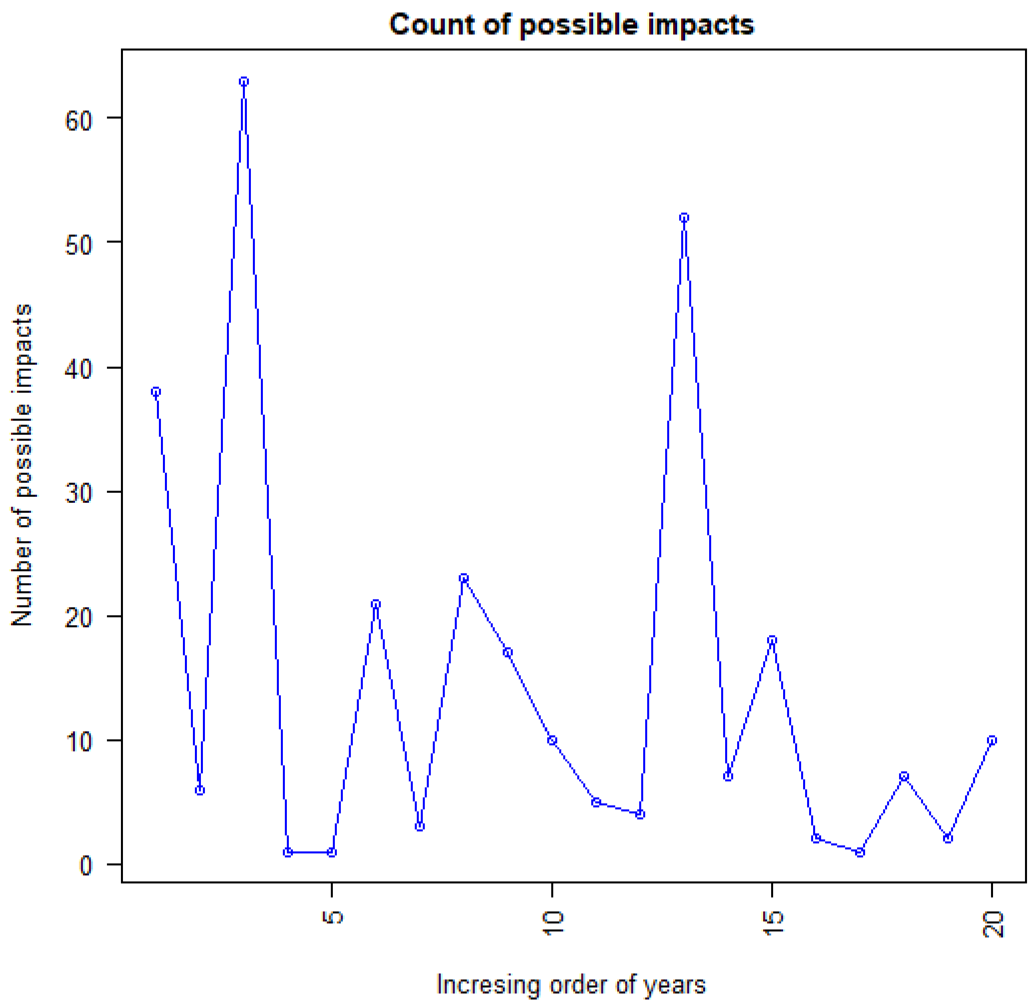
	Object.Name	Period.Start	Period.End	Possible.Impacts	Cumulative.Impact.Probability	Asteroid.Velocity	Asteroid.Magnitude	Asteroid.Diameter..km.	Cumulative.Palermo.Scale	Maximum.Palermo.Scale	Maximum.Torino.Scale
1	2010 AU118	2020	2112	38	1.80000e-08	25.2200	16.2000	1.90000	-2.73000	-3.16000	0
2	2000 WJ107	2114	2114	10	2.20000e-06	12.7800	23.1000	0.0810000	-5.02000	-5.63000	0
3	2010 CA55	2021	2087	6	6.30000e-08	24.0300	22.0000	0.140000	-4.64000	-4.73000	0
4	2013 QM48	2071	2115	17	4.00000e-06	7.47000	27.4000	0.0110000	-6.93000	-7.50000	0
5	2014 DK10	2060	2060	1	3.60000e-08	11.8300	27.8000	0.0100000	-8.71000	-8.71000	0
6	2011 UD21	2105	2105	1	9.30000e-06	1.29000	28.5000	0.00700000	-7.21000	-7.21000	0
7	2008 EZ84	2072	2114	10	0.000220000	7.91000	26.2000	0.0200000	-4.39000	-4.68000	0
8	2015 KH160	2097	2107	2	1.80000e-08	29.2000	26.4000	0.0180000	-8.24000	-8.25000	0
9	2016 TQ2	2110	2115	7	3.80000e-06	15.2100	26.4000	0.0170000	-6.29000	-6.56000	0
10	2010 TW54	2092	2107	7	4.20000e-06	7.77000	27.6000	0.0100000	-6.94000	-7.43000	0
11	2010 TN55	2066	2111	21	2.80000e-06	25.3700	27.0000	0.0140000	-6.33000	-6.88000	0
12	2000 LG6	2075	2115	52	0.000990000	2.11000	29.0000	0.00500000	-5.37000	-5.86000	0
13	2011 BH40	2023	2115	63	1.30000e-07	5.31000	25.1000	0.0320000	-6.80000	-7.47000	0
14	2010 XU	2065	2065	1	1.50000e-10	6.22000	25.4000	0.0290000	-10.1700	-10.2000	0
15	2012 CL17	2095	2112	18	1.40000e-06	21.5400	25.7000	0.0240000	-6.15000	-6.75000	0
16	2010 TD	2072	2091	5	2.70000e-06	18.8600	26.9000	0.0140000	-6.34000	-6.75000	0
17	1993 UA	2111	2111	2	7.10000e-07	9.40000	25.4000	0.0280000	-6.72000	-6.88000	0
18	2016 SK2	2073	2081	4	2.10000e-06	13.6400	25.3000	0.0290000	-5.88000	-5.91000	0
19	2013 GA55	2070	2114	23	3.90000e-05	9.34000	27.7000	0.0100000	-6.04000	-6.21000	0
20	2009 EJ1	2067	2086	3	1.00000e-07	10.6300	28.2000	0.00800000	-8.64000	-8.78000	0

Analysing number of impacts for each asteroid :

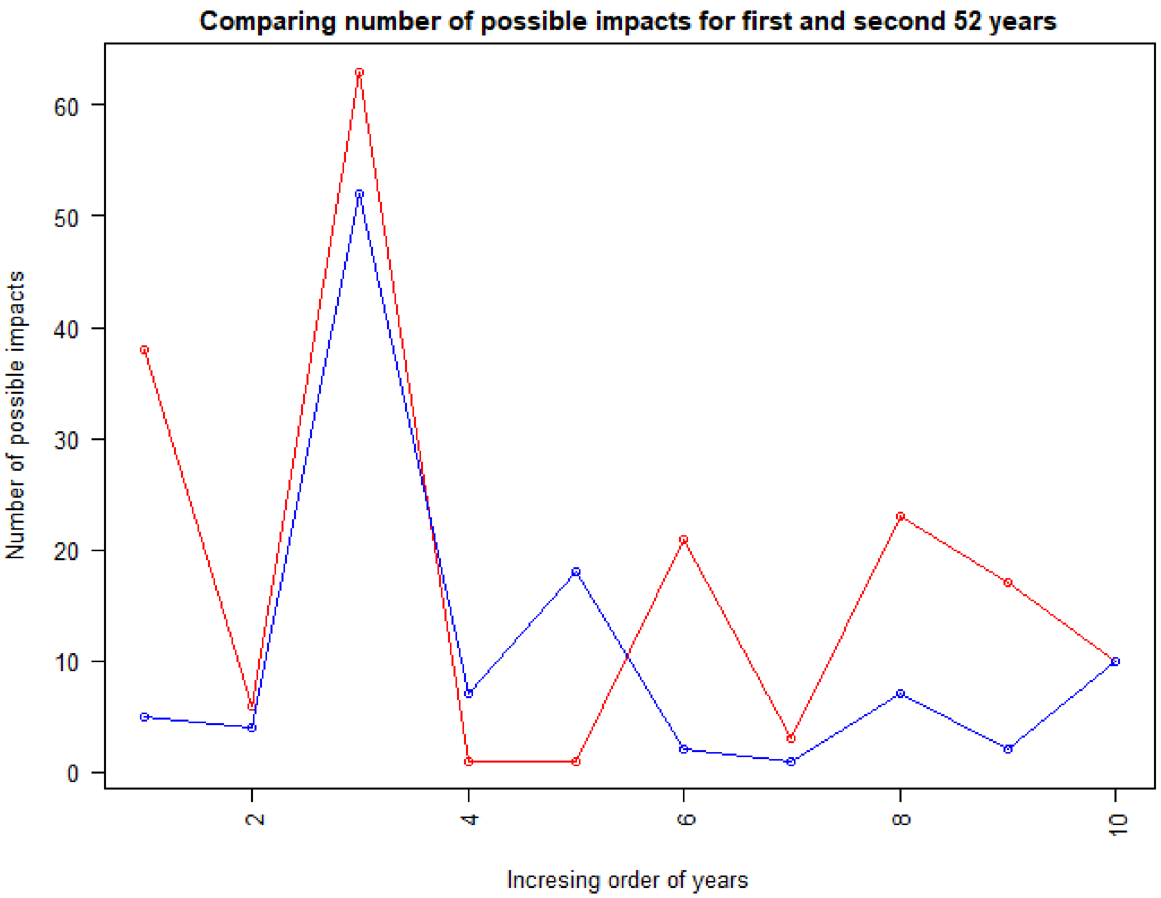


From the above graph we can analyse that asteroid 2011 BH40 has the greatest number of impacts with earth followed by 2000 LG6 and 2010 AU118.

Analysing how number of possible impacts changes has year orders with line graph :

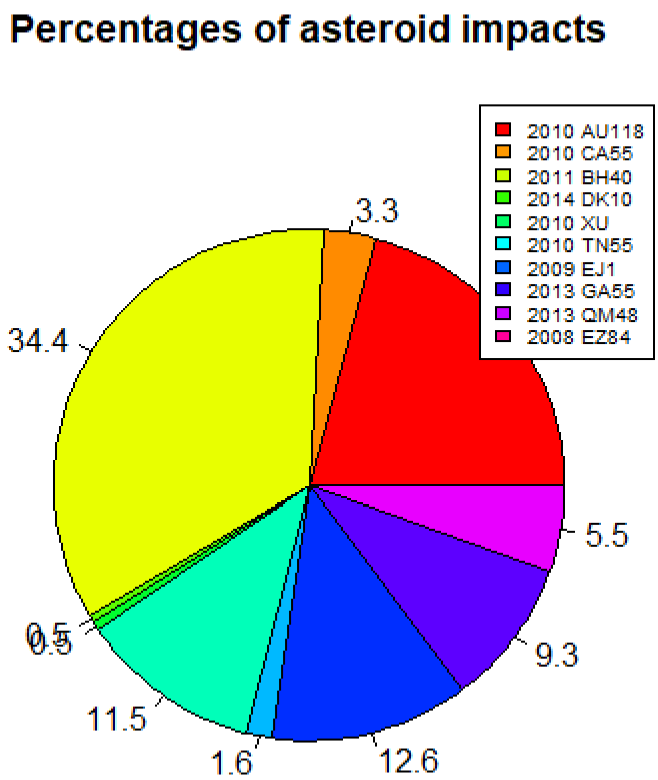


<u>Analysing number of possible impacts changes for first 52 years and second 52 years with multi line graph</u> :



So, from the above graph we can analyse that for the first 52 years asteroids impacts were large that the second 52 years.

<u>Analysing percentages of possible impacts for asteroids with pie graph</u> :



Here we can observe that asteroid 2011 BH40 has greatest percentage to impact earth.

CODE

Used CRAN R 4.0 and PyCharm IDE

This below code consists of all commands which were performed to find the outputs in findings.

```
library("dplyr")

library("plotrix")


data<-read.csv("SentryImpactData.csv",)

print(data)

#' Finding asteroids which have most probability to impact earth.

sortedProbability<-
data[order(data$Cumulative.Impact.Probability,decreasing =
TRUE),]

head(sortedProbability,3)


#' Finding asteroids which have most number of impacts with
earth.

sortedImpacts<-data[order(data$Possible.Impacts,decreasing =
TRUE),]

head(sortedImpacts,3)


#' Finding largest asteroids to impact earth.

sortedDiameters<-
data[order(data$Asteroid.Diameter..km.,decreasing = TRUE),]

head(sortedDiameters,3)


#' Finding near future asteroid impacts.

nearFutureImpacts<-
data[data$Period.Start>=as.numeric(format(Sys.Date(),"%Y")),]

head(nearFutureImpacts)


#' Finding correlation between the variables of dataset.

between(abs(cor(data$Possible.Impacts,data$Cumulative.Impact.Prob
ability)),0,0.5)
```

```
between(abs(cor(data$Possible.Impacts,data$Asteroid.Velocity)),0,
0.5)

between(abs(cor(data$Possible.Impacts,data$Asteroid.Diameter..km.
)),0,0.5)


between(abs(cor(data$Cumulative.Impact.Probability,data$Asteroid.
Velocity)),0,0.5)

between(abs(cor(data$Cumulative.Impact.Probability,data$Asteroid.
Diameter..km.)),0,0.5)


between(abs(cor(data$Asteroid.Velocity,data$Asteroid.Diameter..km
.)),0,0.5)


#' Linear regression with the dataset.

summary(lm(data$Possible.Impacts~data$Cumulative.Impact.Probabili
ty+data$Asteroid.Diameter..km.+data$Asteroid.Velocity))

summary(lm(data$Cumulative.Impact.Probability~data$Possible.Impac
ts+data$Asteroid.Diameter..km.+data$Asteroid.Velocity))

summary(lm(data$Asteroid.Diameter..km.~data$Cumulative.Impact.Pro
bability+data$Possible.Impacts+data$Asteroid.Velocity))

summary(lm(data$Asteroid.Velocity~data$Cumulative.Impact.Probabil
ity+data$Possible.Impacts+data$Asteroid.Diameter..km.))


#' Prediction of Possible Impacts

x<-data$Possible.Impacts

y<-data$Asteroid.Velocity

linearModel<-lm(x~y)

summary(linearModel)

new_Asteroid.Velocity<-data.frame(y=as.numeric(readline("Enter
Asteroid Velocity to Predict Number of impacts :")))

new_Possible.Impacts<-predict(linearModel,new_Asteroid.Velocity)
```

```
cat("Predicted number of asteroid impacts
:",new_Possible.Impacts,"\n")


#' Prediction of Asteroid Diameter

x<-data$Asteroid.Diameter..km.

y<-data$Asteroid.Velocity

linearModel<-lm(x~y)

summary(linearModel)

new_Asteroid.Velocity<-data.frame(y=as.numeric(readline("Enter
Asteroid Velocity to Predict Asteroid Diameter :")))

new_Asteroid.Diameter<-predict(linearModel,new_Asteroid.Velocity)

cat("Predicted asteroid diameter : ",new_Asteroid.Diameter,"\n")


#' Prediction of Asteroid velocity

x<-data$Asteroid.Velocity

y<-data$Possible.Impacts

z<-data$Asteroid.Diameter..km.

model<-lm(x~y+z)

summary(model)

new_Possible.Impacts_And_Asteroid.Diameter<-
data.frame(y=as.numeric(readline("Enter Possible Impacts to
predict Asteroid Velocity :")),z=as.numeric(readline("Enter
Asteroid Diameter to predict Asteroid Velocity :")))

new_Asteroid.Velocity<-
predict(model,new_Possible.Impacts_And_Asteroid.Diameter)

cat("Predicted asteroid velocity : ",new_Asteroid.Velocity,"\n")
```

```
#' Exploratory Data Analysis

sampleData<-sample_n(data,20)

print(sampleData)


#' Analysing the number of impacts of each asteroid

par(las=2) # To align x-label’s vertically

par(mar = c(6, 4, 2, 2)) # To give margins
(bottom,left,top,right)

barplot(names.arg=sampleData$Object.Name,sampleData$Possible.Impa
cts,main="Possible impacts of each asteroid",ylab="Possible
number of impacts", col=rainbow(length(sampleData)),
border="light blue")


#' Analysing how number of possible impacts changes has year
orders with line graph

plotData<-sampleData[order(sampleData$Period.Start),] # Sorting
data with respective to year.

plot(plotData$Possible.Impacts,type
='o',col="blue",xlab="Incresing order of years", ylab="Number of
possible impacts",main="Count of possible impacts")


#' Analysing how number of possible impacts changes for first 52
years and second 52 years with multi line graph :

firstGroup<-plotData[1:10,] # Dividing into first half

secondGroup<-plotData[(11:20),] # Dividing into second half

plot(firstGroup$Possible.Impacts,type='o',col='red',xlab='Incresi
ng order of years',ylab='Number of possible
impacts',main='Comparing number of possible impacts for first and
second 52 years')

lines(secondGroup$Possible.Impacts,type = 'o',col='blue')
```

```
#' Analysing percentages of possible impacts for asteroids with
pie graph:

x<-firstGroup$Possible.Impacts

labels<-firstGroup$Object.Name

piepercent<-round(100*x/sum(x),1)

pie(x,labels =piepercent,main = "Percentages of asteroid
impacts",col=rainbow(length(firstGroup)))

legend("topright",labels,cex = 0.6,fill=rainbow(length(x)))
```

CONCLUSION

With this we can analyse large datasets to get results like predicting variables, finding relations between variables, concluding hypothesis and visualizing data easily with plots and graphs.

With this data provided by sentry we can analyse the faith and situation of earth in future from asteroids, also we can visualize their properties as the year passes.

As we can see asteroids like 2010 RF 12 can posses a threat to earth by passing near, and asteroids of diameters 2.5 km(2011 SR52) can also impact our earth in drastic ways which caused the extinction of dinosaurs.

So, this statistical analysis can be used in many fields to achieve results and predict the future values. If we consider statistical analysis in Weather Forecasts they use to observe and prepare prediction models of climate for future weather conditions.

REFERENCES

- **Kaggle**
- **Probability & Statistics for Engineers & Scientists By Ronald E. Walpole**