SCHOOL OF COMPUTING
Faculty of Engineering

Subject: PROBABILITY & STATISTICAL DATA ANALYSIS

PROJECT: -2

WORLD VACCINATION PROGRESS

GROUP NAME: GROUP G

1		Name : NABIL RAYHAN Matric Number : A20EC9107 Phone Number: +60187731706 E-mail :nabilrayhan17@gmail.com
2		Name : NAZMUS SAKIB Matric Number : A20EC4046 Phone Number :+60172081293 E-mail:sakib.n@graduate.utm.my
3		Name : FARHANA TAHSEEN Matric Number : A20EC4015 Phone Number:+60193272280 E-mail :farhanatahseen98@gmail.com
4.		Name :MD.MOHAIMINUL ISLAM FAYSAL Matric Number : A20EC9105 Phone Number:+601127739204 E-mail :zxfaysalgrp@gmail.com
5.		Name : MUHAMMAD ROBY HERFIAN Matric Number : A20EC0321 Phone Number: 628127632798 E-mail :herifian@graduaate.utm.my

Introduction

Countries are starting to convey COVID-19 vaccines, carrying new desire to the battle against the worldwide pandemic. COVID-19 vaccines are being prepared by partnering organizations such as WHO, UNICEF, Gavi, and many others. Countries can prepare to ensure foundation and specialized assistance is in place to deliver COVID-19 vaccines securely by taking care of diligent arranging. Evenhanded admittance to protected and compelling immunizations is basic to finishing the COVID-19 pandemic, so it is colossally reassuring to see such countless vaccines demonstrating and going into advancement. In the coming years, we will need to keep looking for and developing successful and protected vaccines: for now, we should wear coveralls, clean our hands, make sure we have good ventilation inside, and be open to separation and avoiding swarms.

Objectives

There are many objectives of this Project. Such as:

- We know that how many vaccine does administered in a country
- How many people take vaccines?
- How many people fully vaccinated?
- Which country take which vaccine?
- Which organization made the Covid-19 vaccine
- How many Country are taking the Covid-19 vaccine

DATASET

Link of our Dataset: https://www.kaggle.com/kingabzpro/world-vaccine-progress

The data is:

- Word_Vaccination_Progress(.csv)

In the World Vaccination Progress datasets we have 178 countries and their ratio of taking vaccine and which vaccine being used in a country.

Number of Variable: 5

There are a lots of different type of variable. They are:

- Nominal: (Country), (Vaccine being used in a country)
- Ratio: (Doses Administered), (Doses per 1000), (Fully Vaccinated Population)

DATA ANALYSIS

Based on our project-2, we have selected a secondary dataset. In the dataset, we have 10 columns and 178 rows in total. Then we have used Microsoft Excel for visualization. We mostly use R for data analysis, It is an open-source graphics analysis and computing platform built on a mathematics programming language. Then, we imported our Excel data set into RStudio using csv files converted from our raw Excel response files. To complete the process, we have used R script to create all charts and graphs.

HYPOTHESIS TEST (ONE SAMPLE)

In insights, a hypothesis is a case or proclamation about a property of a population. A hypothesis test is a standard technique for testing a case about a property of a population.

In our test,

Hypothesis statement: -

H_0: $\mu = 500$

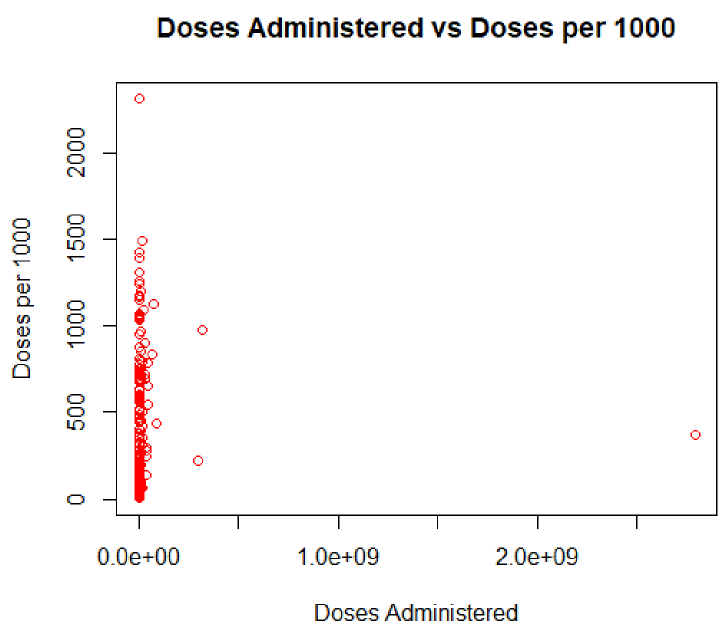
H_1: $\mu > 500$

Significance level, $\alpha = 0.05$

Since we can see that the p-value is less than 0.05, we reject H_0 from one sample t-test.

CORRELATION TEST

We have taken Our Variable as: Doses Administered and Doses per 1000



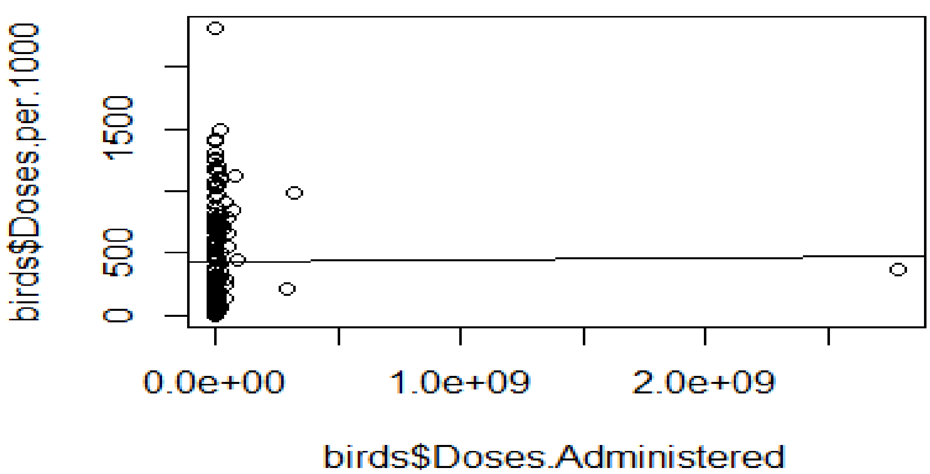
In correlation test, we have taken the two variable and relation among Doses administered and Doses per 1000. Based on the shown graph, the relation between Doses administered and Doses per 1000 is that the positive linear.

And we get the correlation between Doses administered and Doses per 1000 is 0.0075518. And the value of linear regression model y={(4.263e+02) + (1.457e-08)}

```
Call:
lm(formula = y ~ x)

Residuals:
   Min     1Q Median     3Q    Max 
-425.5 -353.8 -118.4  264.2 1886.1 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 4.263e+02  3.103e+01   13.74   <2e-16 ***
x           1.457e-08  1.459e-07    0.10    0.921    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 409.9 on 175 degrees of freedom
Multiple R-squared:  5.703e-05,	Adjusted R-squared:  -0.005657 
F-statistic: 0.009981 on 1 and 175 DF,  p-value: 0.9205
```

REGRESSION

First, we have converted the main Dataset, from xlsx to csv. Then we imported it to R. After that we read the file and gave it a random name as we like. In this case it is birds. We checked the variables. Also checked the first 6 rows of data. Based on the information we declared variable x & y for scatter plot. Finally, this is how we got the scatter plot.



```
Call:
lm(formula = birds$Doses.per.1000 ~ birds$Doses.Administered)

Coefficients:
             (Intercept)  birds$Doses.Administered
               4.263e+02                 1.457e-08
```

Form this we have learned how to find exact variables from a dataset and based on that dataset how we can find scatter plot.

GOODNESS OF FIT TEST

Our Dataset conducted on 178 countries vaccination progress. Like which country how many people does administered the vaccine, Doses per 1000, and how many people fully Vaccinated and which type of vaccine they used in their country.

First test claims that the Doses per 1000 of people, from the data we take 10 data from the 10 different country.

So,

The proportion of the does for 1000 is equal for the 10 different value.

H0: p1=p2=p3=p4=p5=p6=p7=p8=p9=p10
H1: At least 1 out of 10 proportions is different

So, let the significance level= 0.05

Chi-squared test for given probabilities

data: doesper100

X-squared = 1656, df = 9, p-value < 2.2e-16

We get x^2=1656, for p-value=2.2e-16. For critical value we get critical x^2 (9,0.05) =16.919**.**

Here as the p-value is not very small and $x^2 > x^2$critical. So We can reject the null hypothesis.

And again I am testing the goodness of fit test taking the whole variable data from the data set.

From taking the full data (Doses per 1000, Doses administered and fully vaccinated).

We can get the same p-value.

Pearson's Chi-squared test

data: myalldata

X-squared = 40869931, df = 352, p-value < 2.2e-1

Here also, we reject the hypothesis null.

CONCLUSION

In conclusion, based on the available evidence, the mean dose per 1000 persons is less than 500. As a result, we got a correlation of 0.0075518 between Doses administered and Doses in per 1000. And the value of linear regression model y={(4.263e+02) + (1.457e-08)}. Throughout this, we have learned how to extract exact variables from a dataset and find scatter plots based on that dataset. We can reject the null hypothesis because some will not follow the hypotheses from goodness of fittest.

VIDEO LINK:
https://youtu.be/BCzPvsDYEKk

APPENDIX

Doses Administered	Doses per one thousand
2789178281	367.3
319872053	977.7
293370338	216.9
91593968	437.3
74638083	1122.6
69516383	838.3
47722029	789.7
44951168	546.1
43803521	653.9
36998897	138.2
36983641	293.1
36044137	247
35788723	282.8
33585366	906.3
32282553	690.9
27467853	723.2
20403835	1089.4
18646784	419.1
18438254	357.1
18251766	506.6
15298930	308.1
14735666	1495.9
13734246	797.1
13484364	63.5
10671175	1201.2
10082705	62.5
9748714	853.5
9424351	964.7
8756809	449.7
8407342	78.8
7774300	756.1
7722093	719.8
7570988	712.5
7219668	104
7172478	674.9
7125249	805.4
7099237	697.2
6860276	274.5
6378584	392.5
6301727	195.5
5785134	679.3
5153895	161.1

5146241	737.1
4611224	252.3
4553653	54.2
4180186	721
4011742	727
4010467	39.2
3660589	1061.3
3656752	688.1
3645291	1149.9
3407068	604.2
3399122	624
3348198	442.9
3324349	153.4
3272579	288.6
3231849	115.1
3197278	321.1
3100000	638.7
3100000	749.3
3080340	309.8
2926066	1051.9
2788245	163.2
2626337	27.5
2586522	402.8
2379658	581.9
2331030	466.3
2231841	196.6
2229165	50
2160457	774.5
2030063	10.4
1978826	1260.8
1871987	32.4
1746130	74
1695298	241.4
1642744	49.8
1583153	136.9
1441439	345.1
1413358	683.6
1380174	44.8
1302992	45.1
1232876	41.4
1190889	23.2
1183137	162.6
1175505	171.6
1146378	79.4
1092435	567

1019364	208.7
914509	692.3
890052	310.5
765890	20.6
757974	148.6
728458	575.1
685500	72.3
662804	95.3
661648	38.4
655537	184.9
632648	39.9
631446	1305.9
613840	16
610924	14.6
605845	125.4
598863	473.3
491716	809.1
489415	949
488859	51
482058	231.4
425051	23.4
394930	13.4
384974	31
368066	1041
347246	44
322976	86.6
315780	95
307506	486.8
294053	377.5
279639	449.3
274008	310.1
247228	188.9
235400	169.4
186207	8.3
182331	316.6
178264	11.9
176038	9.2
173700	27.5
165266	1033.9
147115	8.5
147036	513
137676	1422.8
136003	30.9
133868	72.5
129743	348.3

125484	451.9
124856	1179.6
124460	50.8
104683	359.3
98546	1172.1
97315	12.7
97245	3.9
94574	877.3
89619	1396.5
84360	101.4
77969	2312.4
77881	1242
71916	187.7
68821	14.3
67511	0.8
66564	1040.6
63606	114.4
62343	134.2
58074	603.1
51563	283.5
45178	230.3
44379	1313.6
44305	39
42644	9.1
41993	7
41737	19.7
40810	17.9
37126	518.3
36759	17.4
33965	878.1
33425	596.6
31054	583.5
29980	283.6
29410	2.7
26624	2.3
25780	680
24634	605.9
23918	113.3
21469	32.9
19064	247.6
13235	732.6
8834	558.8
7391	275.7
4322	1070.1
1306	243.7