UNIVERSITY OF KALYANI

NAME-KIRONMOY ROY

SESSION- 2019-2020

REGISTRATION NO-015712

ROLL NO-2116117-1915691

PAPER- STATDSE-4

GROUP MEMBERS - KIRONMOY ROY AND ARNAB GIRI

TOPIC:-

STATISTICAL DATA ANALYSIS PROJECT ON

"RELATIONSHIP BETWEEN DEPRESSION AND ANXIETY"

ACKNOWELEDGEMENT

I take this opportunity to express a deep sense of gratitude towards my supervisor Prof. Kamal Krishna Khanda for providing excellent guidance, encouragement and inspiration through out the project work.My sincere thanks are also to respected principal Dr. Runu Das and Departmental Head Prof. Kakali Dutta for this kind,endless cooperation

and constant help.I would also like to thank project partner for his valuable suggestions and helpful discussions.Finally, I want to express my gratitude to my parents and other family members for their constant support and motivation through out my lifre.

CONTENT

INTRODUCTION

World Health Organization (WHO) considered mental health as an important component of human health. Although the onset of depression can be at any stage of life, the prevalence of major depression is increasing during **adolescence and young adulthood**.

The adverse effect of **psychological distress** among people reduces their **self-esteem** which might cause many problems at both personal and professional levels leading to college dropout, impaired ability to work effectively, poor academic, disturbed relationship and **suicide.**

Psychological distress such as **stress, anxiety and depression** are frequent among people of different age group which can affect their daily work, physical health and psychosocial well-being.

Some research says unhealthy bad habits/lifestyle plays a major role on depression and anxiety such as **irregular sleeping habit**, **Fast food consumption**, etc. It is also seen **that anxiety can causes depression** in many situation.

So, my target is to find correlation among depression and anxiety ans to **test that all the anxiety related factors have gives same effect on depression or they have different significant effect.**

ABSTRACT

As my objective is to **predict the level of depression and level of anxiety on the basis of unhealthy bad habits or lifestyle & find the correlation between depression and anxiety** ,so I have to use an "amount of depression" measuring scale and an "amount of anxiety" measuring scale PHQ-9 and GAD-7 scale.

The PHQ-9 is the 9 -question depression scale of PHQ.
The **Patient Health Questionnaire** (**PHQ**) is a multiple-choice, self-report inventory that is used as
a screening and diagnostic tool for mental health disorders of depression, anxiety, eating, etc.

Generalized Anxiety Disorder 7 (GAD-7) is a self-reported questionnaire for screening and severity measuring of generalized anxiety disorder (GAD).

So, **I find correlation coefficient and scatter diagram between PHQ-9 score and GAD-7 score**.

Finally I fit a model to test the PHQ-9 score based on the effect of Anxiety. It is quite possible to take just only the GAD-7 Score as independent variable but my interest is to know the exact effect of the each and every anxiety related issues.

I'm further discussing about the **model** and **tesing procedure** below.

METHODOLOGY

Data source:-

Snowball sampling technique was used to collect the data in this sample survey.

A survey was conducted by me by Google-Form where some basic questions are asked about his/her lifestyle to make the form interesting and most importantly some anxiety and depression orientated questions were asked to collect the data about the research interest. A total of 151 responses were received.

Measures:-

To collect data for scoring **PHQ-9 score** there were 9 specific questions were asked to know the **true state of depression** of that person and here I used four options in each question & give 4 scores 0,1,2,3 according to their choosen option in each question. “0” indicates low depressive syndrome and “3” indicates high depressive syndrome. The total sum of the responses suggests varying levels of depression Scores range from 0 to 27.From research it is known that a total of 10 or above is suggestive of the presence of depression.

To collect data for scoring **GAD-7 score** there are 7 specific questions were asked to know the **true state of anxiety** of that person and here I used four options in each questions & give 4 scores 0,1,2,3 according to their choosen option in each question.

"0" indicates low anxiety syndrome and "3" indicates high anxiety syndrome. The total sum of the responses suggests varying levels of anxiety Scores range from 0 to 21. The GAD-7 is a valid and efficient tool for screening for GAD and assessing its severity in clinical practice and research.

Statistical Analysis:-

Total statistical data analysis was divided into two type of analysis. One is Descriptive analysis & another is Regression Analysis.

.

In the beginning all the descriptive measures were tabulated for visual representation of the data. Data was collected and tabulated using **Microsoft excel** and all the work of Data Analysis was done using a popular software package **SPSS** .

1] Frequencies were calculated for all quantitative measures
2] Mean and standard deviation were calculated for qualitative measures.
3)A scatter diagram between the **PHQ-9 score** and **GAD-7 score** was represented.

MODEL:-

For regression analysis **Multiple linear regression Model** was taken. In general there are several Regression model Linear or Non-linear ,but in the time of analyzing the data I get **Minimized error sum of square(SSE) or Residual sum of square(RSS)** was significantly low in that particular model in comparing to other models.At last an **ANOVA** table was represented.

Consider a **multiple linear regression model with k predictor variables**:

$$Yi=\beta0+\beta1x1i+\beta2x2i+...+\beta4x4i+\epsilon i$$

Let each of the 4 predictor variables, X1, X2... X4, have 152 levels. Then xij represents the i th level of the j th predictor variable xj.

Observations, Y1, Y2... Y152, recorded for each of these 152 levels can be expressed in the following way:

The system of 152 equations shown previously can be represented in matrix notation as follows:

$$y=X\beta+\epsilon$$

ϵ be the random error in this model.

ϵi's are independent and indentically normally distributed with mean zero and constant variance.

Errors are uncorrelated to each other.

Testing procedure:-

As our target is to check Whether all **the levels of anxiety** has **equal effect or not**. So out testing objective was,

H0 : $\beta 1 = \beta 2 = \cdots = \beta 4 = 0$

i.e. all the anxiety related issues gives same effect on depression.

H1 : $\beta j \neq 0$ for at least one j, j = 1, . . . , 4

i.e. At least one anxiety related issue gives different significant effect on depression.

Rejection of H0 implies that at least one of the predictors, x1, x2, . . . , x4, contributes significantly to the model.

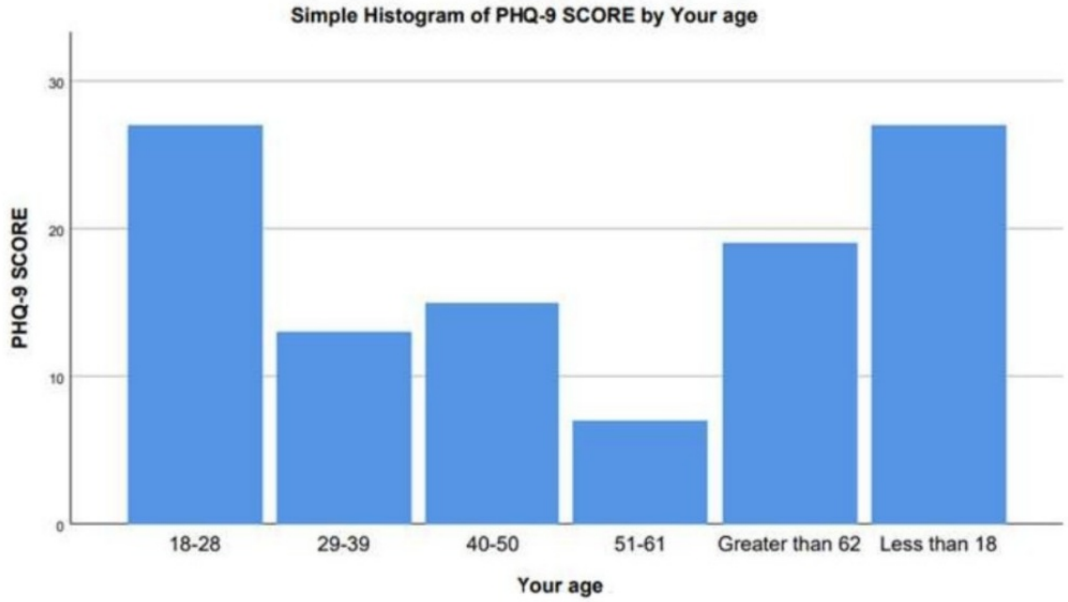
Here the testing will be done by taking level of significance $\alpha=0.01$

DATA ANALYSIS

Descriptive Analysis:-

It is the table showing the level of depression according to PHQ-9 score.

PHQ-9 Score	Depression severity
0-4	None-minimal
5-9	Mild
10-14	Moderate
15-19	Moderately Severe
20-27	Severe



From here we can say the

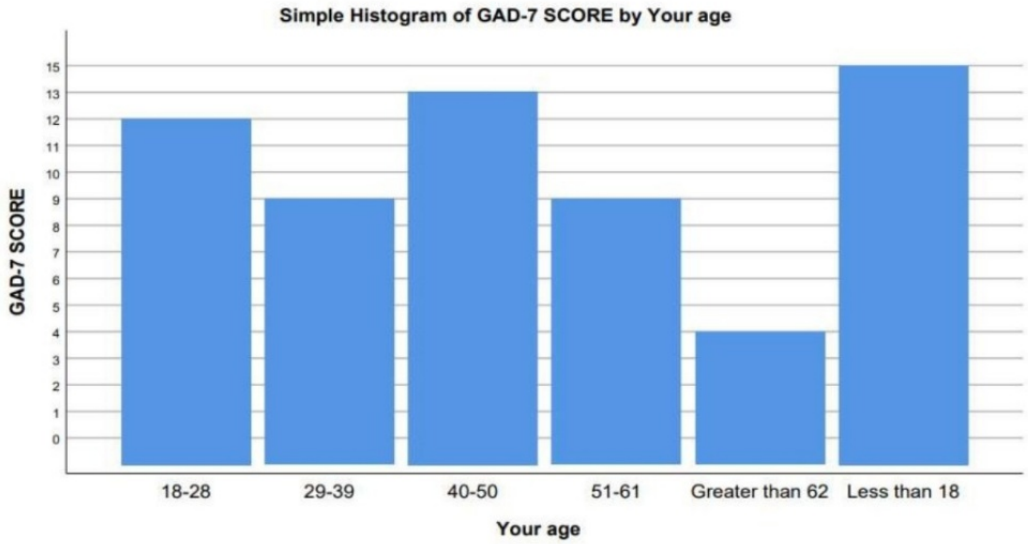
1)Age group 18-28 and Less than 18 years old people are having **serious depressive syndrome(SEVERE**)

2)Age group "51-61"having **less depressive syndrome(MILD)**

Depression level of the remaining age group are easy to see from this histogram and the above table.

It is the table showing the level of depression according to GAD-7 SCORE

Score	Symptom Severity
0-4	Minimal
5-9	Mild
10-14	Moderate
>15	Severe

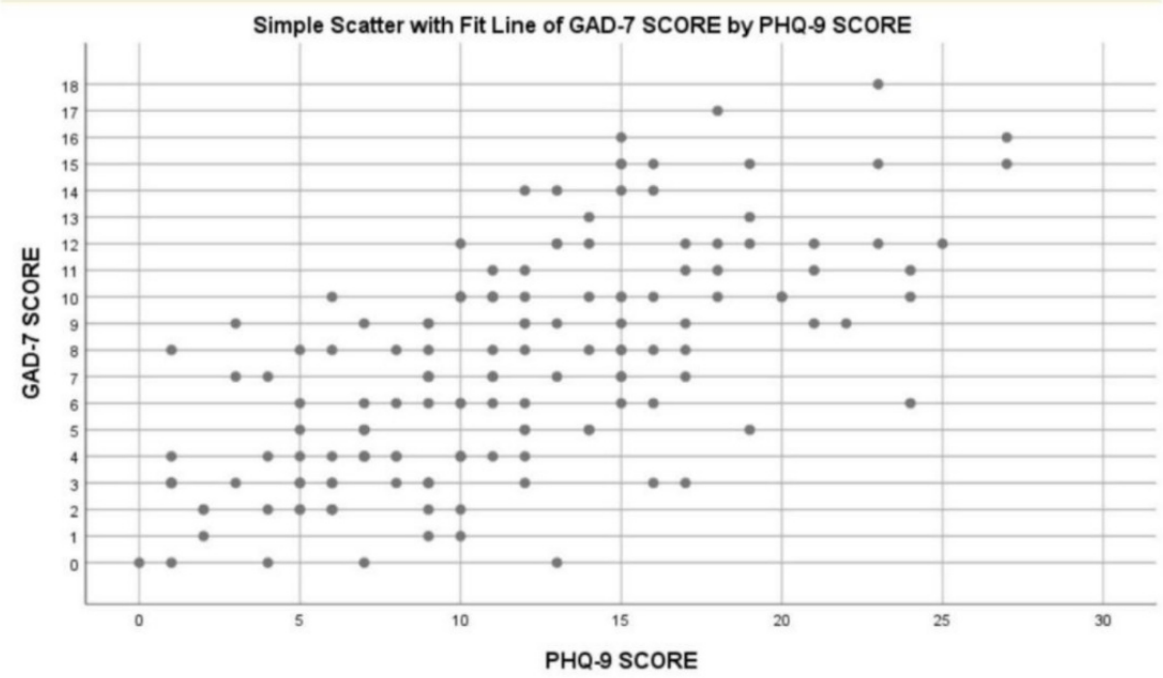


1)Age group “Less than 18”,”18-28”,”40-50” having Moderate level of Anxiety.

2)Age group”29-39”,”51-61” having Mild level of anxiety.

3)“Greater than 62”age group having Minimal level of anxiety.

Scatter Plot Between PHQ-9 score and GAD-7 score:-



From the Scatter plot it is obvious that there is a linear association between GAD-7 score and PHQ-9 score. Also it is seen that there is positive correlation between two

measures.But correlation coefficient gives extend/ degree of linearity in the data.

Correlations

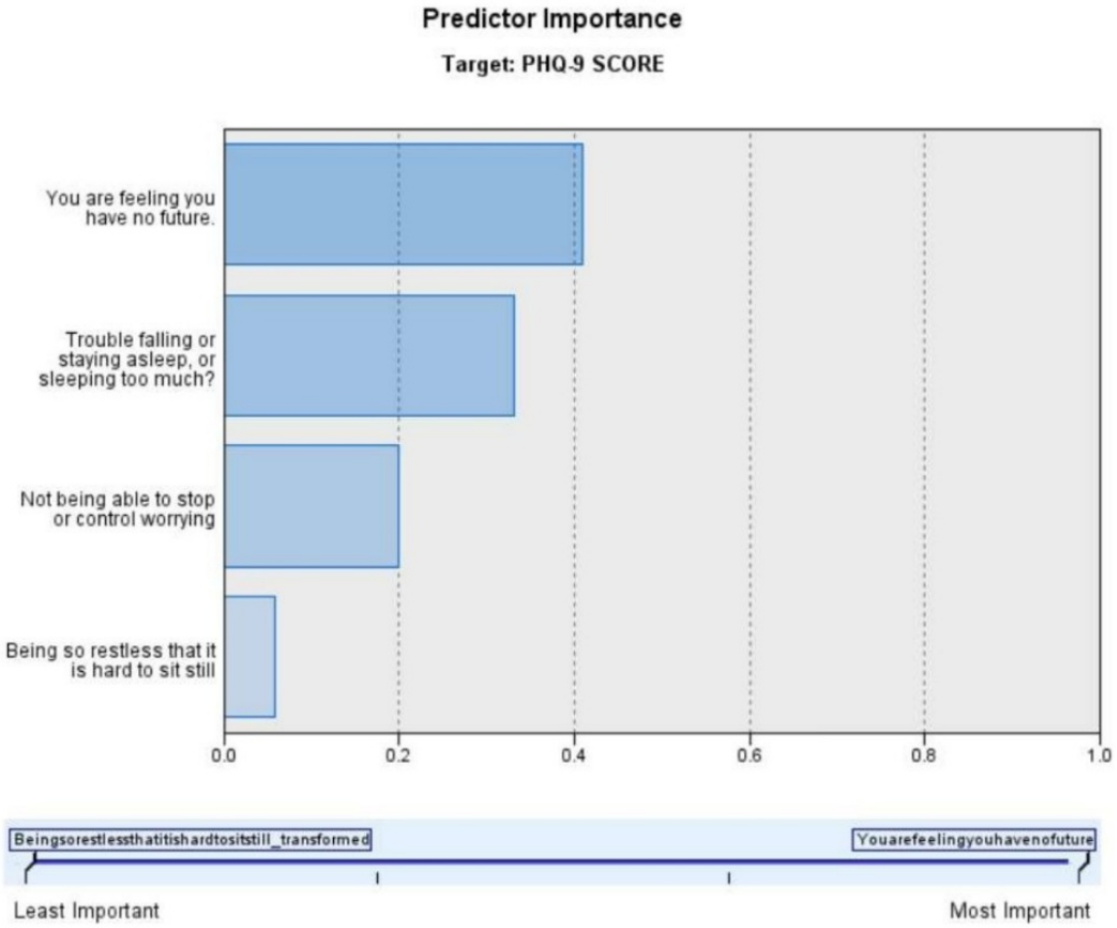
		GAD-7 SCORE	PHQ-9 SCORE
GAD-7 SCORE	Pearson Correlation	1	.295**
	Sig. (2-tailed)		.000
	N	151	151
PHQ-9 SCORE	Pearson Correlation	.295**	1
	Sig. (2-tailed)	.000	
	N	151	151

Product moment correlation coefficient between PHQ-9 Score and GAD-7 score is **0.295**.So,there is **low positive correlation** between measure of anxiety and measure of depression.

Regression Analysis:-

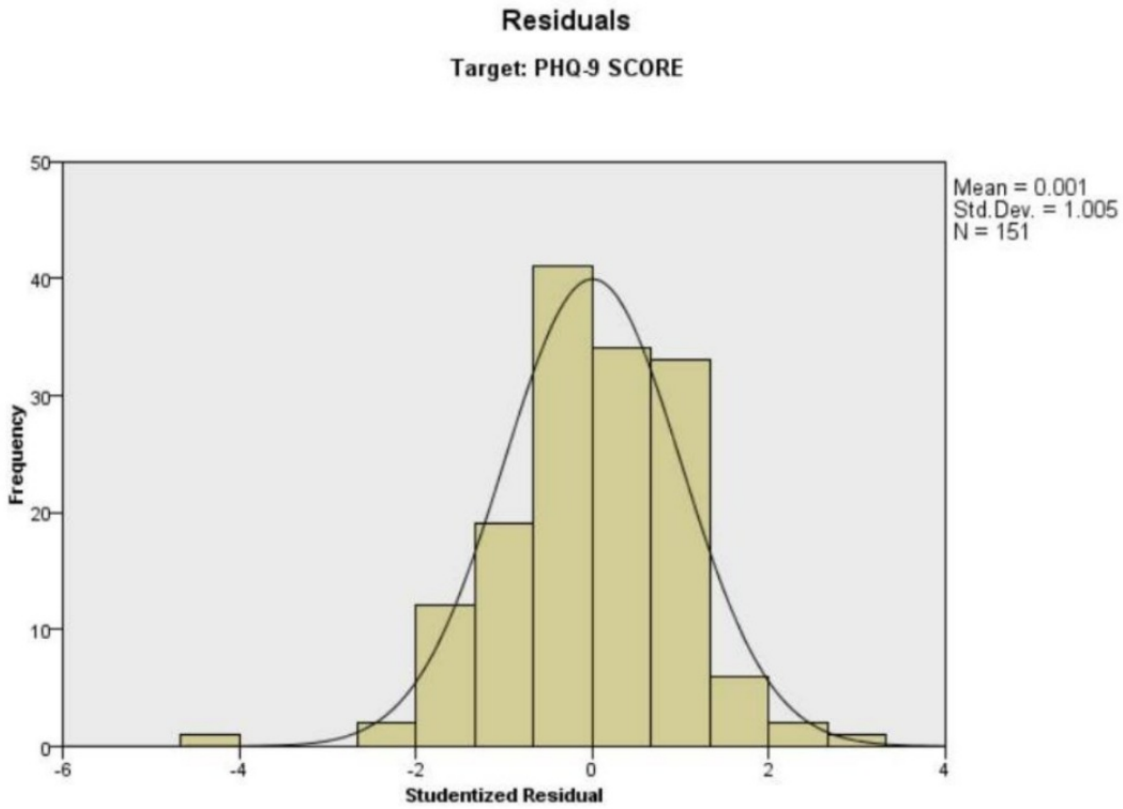
In the above section I have mentioned the model (Multiple linear regression model) and hypothesis.In SPSS all the predictor related to anxiety are taken suitably and Sum of square due to Error & Sum of sqare due to factors of anxiety are found.By orthogonal splitting it is known that

SS(Total)=SS(Due to anxiety)+SS(due to residual). Then degress of freedom are found for both of this.



These are the anxiety related predictors.

In SPSS It is seen that the predictor "You are feeling you have no future " influenced most in the prediction of PHQ-9



It is mentioned above that **errors are taken normally distributed** and also **they are uncorrelated** to each other.This histogram of standardized residuals to a normal distribution.The smooth line represents the normal distribution.The closer the frequencies of the residuals are to this line, the closer the distribution of the residuals is to the normal distribution **.**

ANOVA TABLE:-

Effects

Target: PHQ-9 SCORE

Source	Sum of Squares	df	Mean Square	F
Corrected Model ▼	3,829.662	10	382.966	34.266
Youarefeelingyouhavenofuture	621.322	3	207.107	18.531
Troublefallingorstayingasleeporsleepingtoomuch	503.910	3	167.970	15.029
Notbeingabletostoporcontrolworrying_transformed	302.513	2	151.257	13.534
Beingsorestlessthatitishardtositstill_transformed	88.955	2	44.478	3.980
Residual	1,564.695	140	11.176	
Corrected Total	5,394.358	150		

So ,here **sum of squares due to error is 1564.695** with **140 degrees of freedom** and the **sum of square due to predictors (related to anxiety)is 3829.662 with 10 degress of freedom.** Here also sum of squares of individual predictors with corresponding degrees of freedom is also provided.It is to be noted that all sum of squares are mutually orthogonal to each other. Here it is seen that our test statistics

F= MS(due to predictor)/MS(Due to error)

Now According to our hypothesis we reject H0

If Observed F>Tabulated F

Conclusion:-

Here observed F=34.266

Tabulated F= $F_{\alpha;10,140}$

As mentioned above we have taken level of significance $\alpha=0.01$

We get $F_{0.01;10,140}=2.44997541$

So,Observed F>Tabulated F

So, we reject H0 at level of significance $\alpha=0.01$

i.e.

All the different anxiety related issues gives different significant effect on depression.

APPENDIX

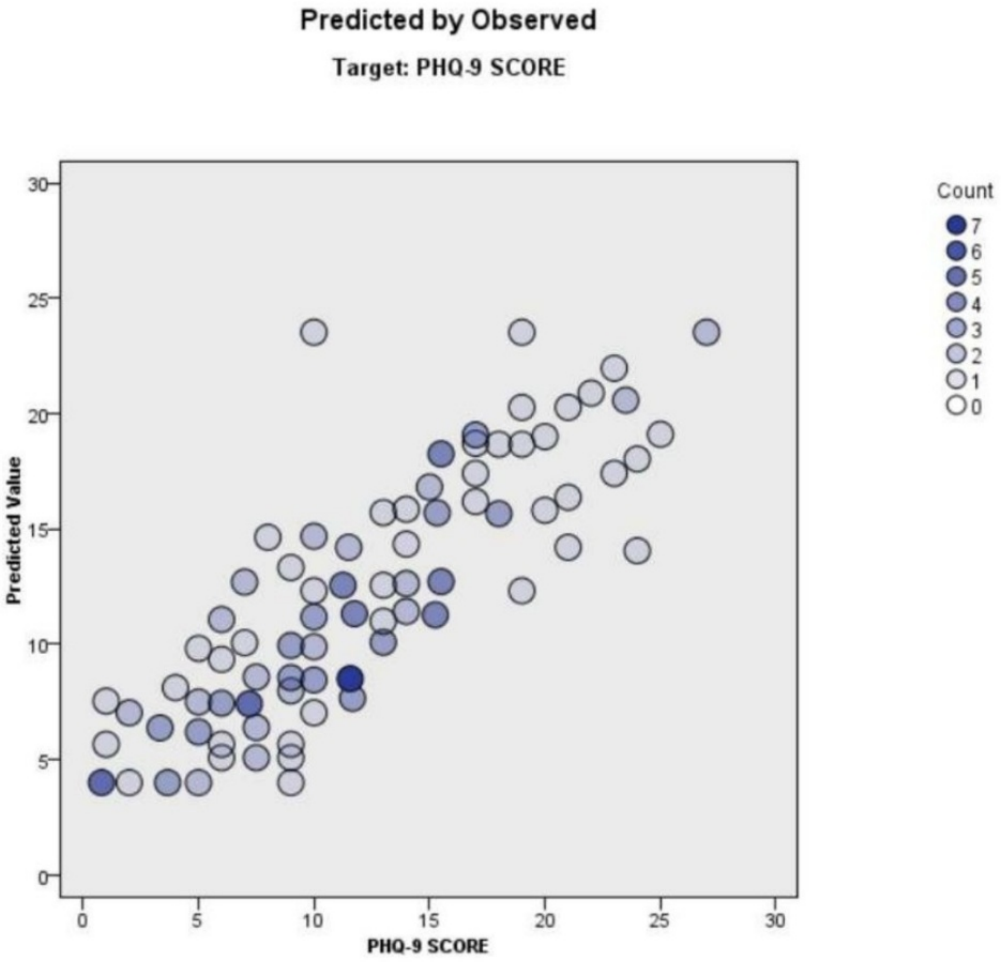
Some other observations which were helpful in this entire project is represented below

1)

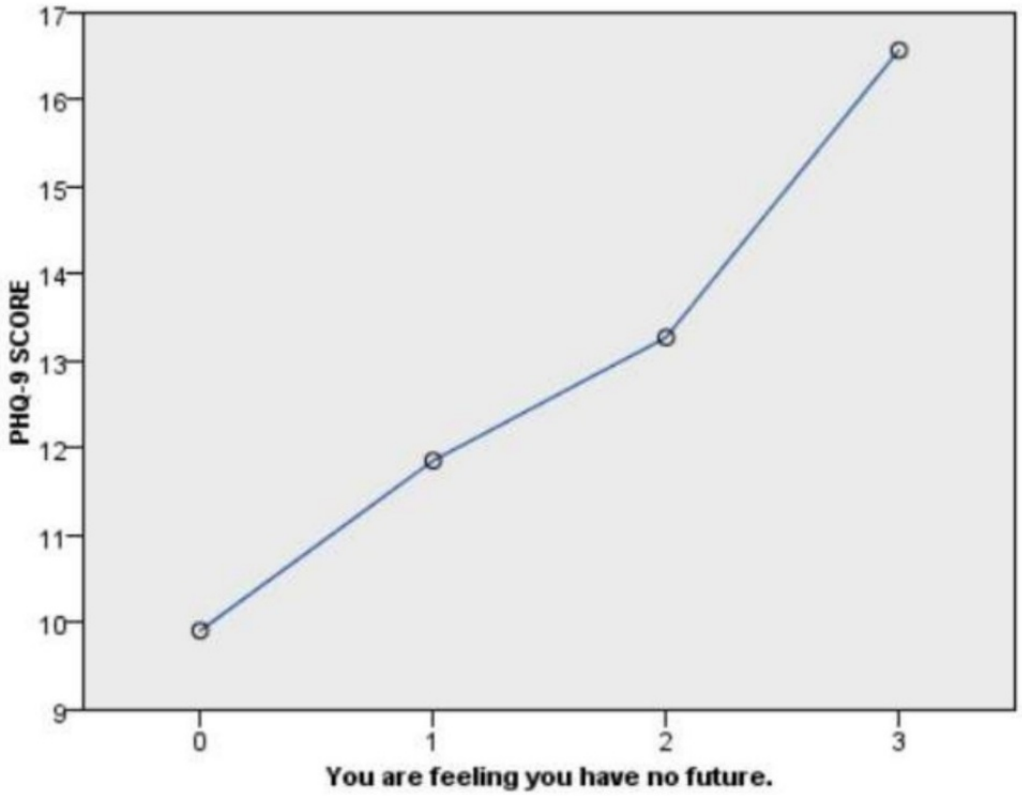
Descriptive Statistics

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
GAD-7 SCORE	151	15	0	15	5.97	.241	2.960	8.759	.281	.197	-.094	.392
Mood Swing (In a Day)	151	3	0	3	1.29	.078	.963	.928	.473	.197	-.683	.392
Lack of wish to talk with friends and others(In a Week)	151	3	0	3	1.24	.095	1.170	1.369	.384	.197	-1.342	.392
PHQ-9 SCORE	151	27	0	27	11.40	.488	5.997	35.962	.386	.197	-.241	.392
Feel like you are wasting your time and harm your health after watching a movie or relaxing or consuming some fast food.	151	3	0	3	1.13	.091	1.124	1.262	.478	.197	-1.185	.392
Not being able to stop or control worrying	151	3	0	3	1.29	.094	1.152	1.328	.151	.197	-1.465	.392
Overthinking about Small incidents in life(In a Week)	151	3	0	3	1.33	.090	1.100	1.210	.377	.197	-1.171	.392
Being so restless that it is hard to sit still	151	3	0	3	.98	.086	1.061	1.126	.650	.197	-.907	.392
Trouble concentrating on things, such as reading the newspaper or watching television?	151	3	0	3	1.08	.095	1.163	1.354	.590	.197	-1.165	.392
Trouble falling or staying asleep, or sleeping too much?	151	3	0	3	1.36	.094	1.152	1.326	.178	.197	-1.408	.392
Thoughts that you would be better off dead, or of hurting yourself in some way?(suicidal thoughts)	151	3	0	3	.39	.066	.816	.666	1.998	.197	2.824	.392

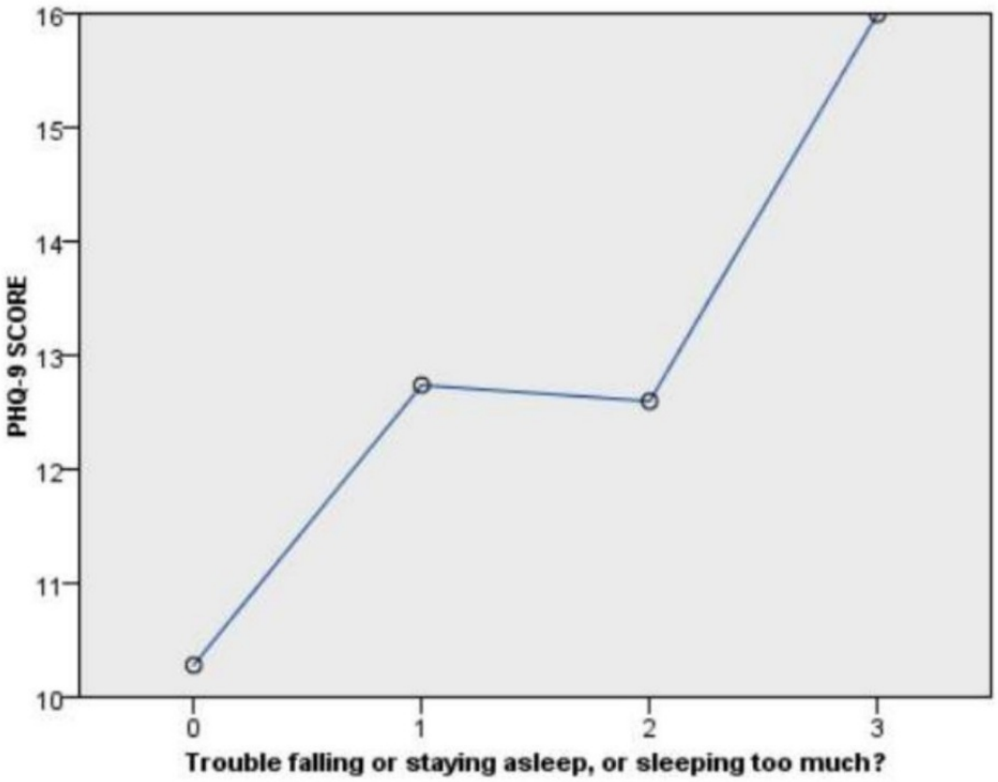
2)


Predicted by Observed
Target: PHQ-9 SCORE
Predicted Value
PHQ-9 SCORE
0
5
10
15
20
25
30
Count
7
6
5
4
3
2
1
0

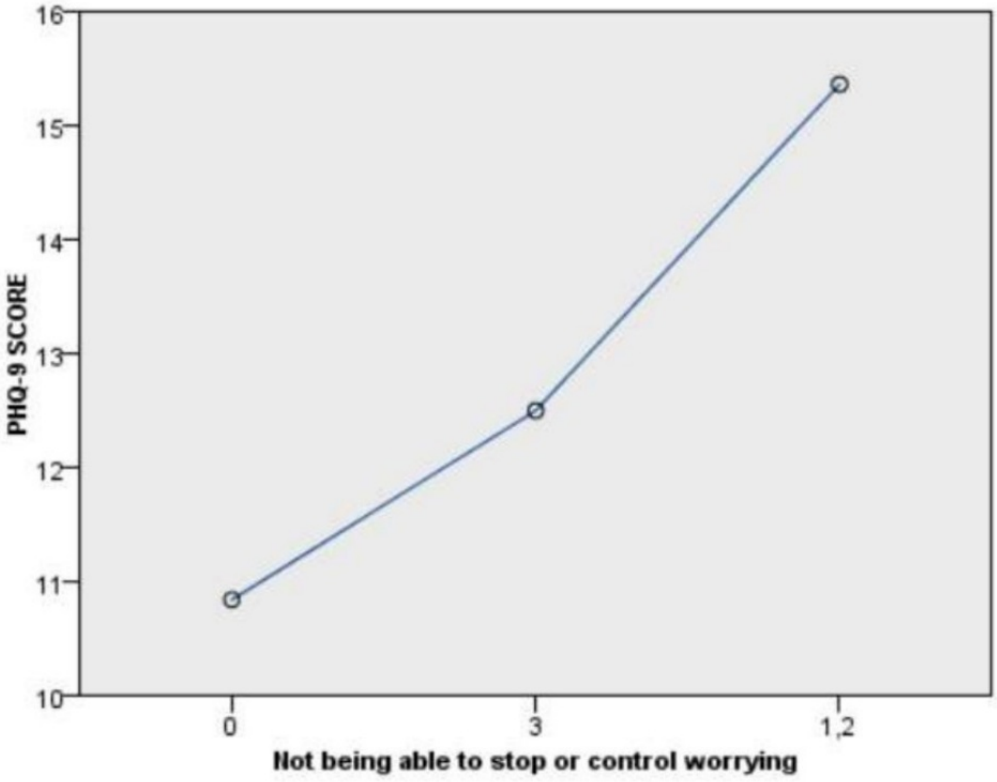
3)


PHQ-9 SCORE
17
16
15
14
13
12
11
10
9
0
1
2
3
You are feeling you have no future.

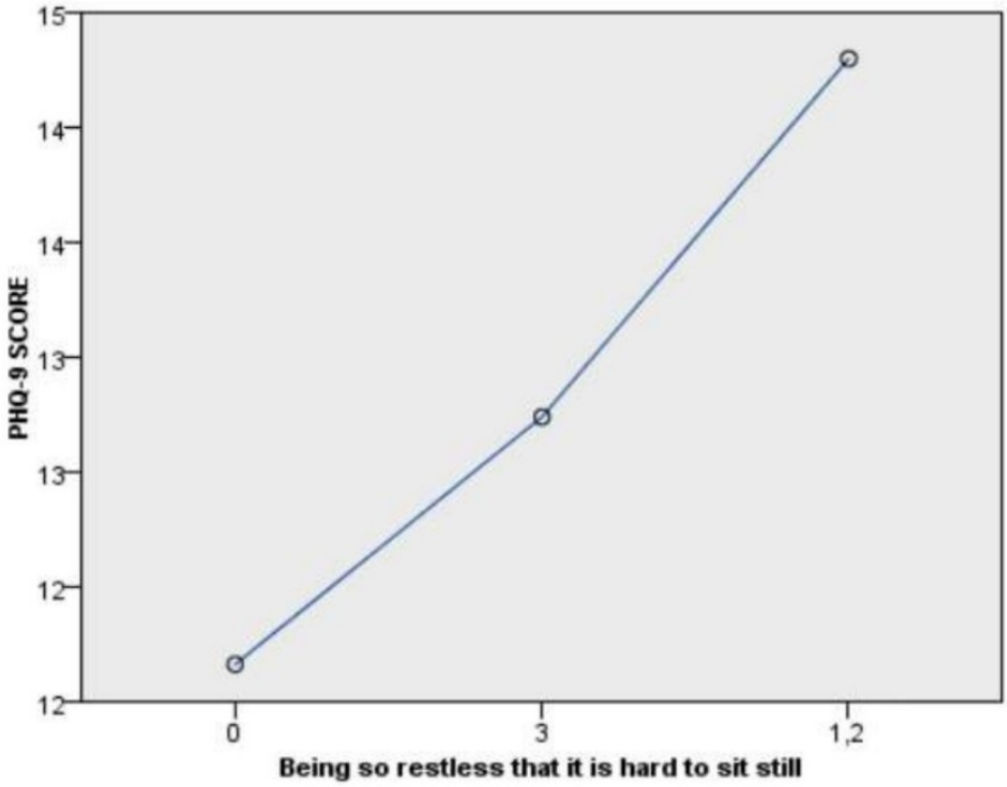
4)


16
15
14
13
12
11
10
PHQ-9 SCORE
0
1
2
3
Trouble falling or staying asleep, or sleeping too much?

5)


16
15
14
13
12
11
10
PHQ-9 SCORE
0
3
1,2
Not being able to stop or control worrying

6)


15
14
14
13
13
12
12
PHQ-9 SCORE
0
3
1,2
Being so restless that it is hard to sit still

References

BOOKS:-

1)Linear Statistical Inferences and its Applications-

C RADHAKRISHNA RAO

2)Fundamental of Statistics vol-1-

A.M.GUN, M.K GUPTA,B.DASGUPTA

3)Statistical Inference –**MANOJ KUMAR SRIVASTAVA**

NAMITA SRIVASTAVA

WEBSITES:- 1)**Generalized Anxiety Disorder 7 - Wikipedia**

2) **PHQ-9 - Wikipedia**

THANK YOU