ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

Diagnosis of ADHD using Statistical measures

[1] E.Radhamani, [2] K.Krishnaveni
[1] Research Scholar, [2] Associate Professor
[1][2] Sri S.R.N.M.College, Sattur, Tamil Nadu, India

Abstract: - **Diagnosis of ADHD using Statistical measures is proposed in this paper. An ADHD data set with 105 samples based on different ages ranging from 6 to 9 and gender (Male & Female) is collected and classified into noADHD, moderateADHD and highADHD by using statistical parameters like mean and standard deviation. The ADHD may be caused by different factors like age, gender, and brain damage, etc. In this research work, only two factors namely age and gender are taken to find the goodness of fit for the data collected and to examine that whether these parameters have any significance for the cause of ADHD. The Tests of Independence between these variables are carried out by Chi-Square Test and the results are discussed and analyzed. From this test it has been concluded that age has no significance for the cause of ADHD but Gender has significance ie, AGE and ADHD are independent but Gender and ADHD are dependent variables.**

Keywords: **ADHD, Chi-Square Test, Mean, Standard Deviation.**

I. INTRODUCTION

Attention Deficit Hyperactivity Disorder (ADHD) is a chronic mental disorder which causes various hyperactive and disruptive behaviors. Symptoms of ADHD often include difficulty focusing, sitting still, and staying organized. Many children exhibit signs of this disorder before age 7, but some may not be diagnosed until adulthood. The symptoms of ADHD will differ child to child, age to age and gender to gender. People with ADHD often struggle with school, work, and relationships. They're also more likely to develop other conditions, including anxiety, depression, and learning disabilities. It can also help prevent other disorders from developing in the future. The boys and girls affected with ADHD may differ in a number of different ways. The boys with ADHD usually show externalized symptoms, such as running and impulsivity. Girls with ADHD, on the other hand, typically show internalized symptoms. These symptoms include inattentiveness and low self-esteem. Boys also tend to be more physically aggressive, while girls tend to be more verbally aggressive. Since girls with ADHD often display fewer behavioral problems and less noticeable symptoms, their difficulties are often overlooked. As a result, they aren't referred for evaluation or treatment. This can lead to additional problems in the future. The undiagnosed ADHD can have a negative impact on girls' self-esteem. It can even affect their mental health. Boys with ADHD typically externalize their frustrations. But girls with ADHD usually turn their pain and anger inward which leads to an increased risk for depression, anxiety, and eating disorders. Girls with undiagnosed ADHD are also more likely to have problems in school, social settings, and personal relationships than other girls. Besides that the brain structure and its functions are also be complicated and brain size will be different due to brain damage or abnormality in both boys and girls. In girls with ADHD, the differences appeared in the prefrontal regions of the brain, which control motivation and ability to regulate emotions. In contrast, for boys with ADHD, the differences appeared in primary motor cortex of the brain, which control basic motor function. While the symptoms of ADHD may present differently in boys and girls, it's critical for them to be treated. The symptoms of ADHD do tend to lessen with age, but they can still affect many areas of life. So it is essential to be analyzed initially that the existence of ADHD depends on age and gender or not. In our previous research work [1], ADHD data samples are collected and classified into HIGH ADHD, MODERATE ADHD and NOADHD using Data Mining classifiers. Since the symptoms of ADHD vary with Age and Gender, these two factors are taken in this paper to examine that whether they have any significance for the influence of ADHD. The Chi-Square Test of Independence is proposed to determine the significant relationship between Age and ADHD and Gender and ADHD. The Chi-Square statistic is most commonly used to evaluate Tests of Independence when using a cross tabulation. For example, to examine the relationship between gender (male vs. female) and ADHD (high vs. moderate vs. No) Cross tabulation presents the distributions of these two categorical variables simultaneously, with the intersections of the categories of the variables appearing in the cells of the table. The Test of Independence assesses whether an association exists between the two variables by carefully examining the pattern of responses in the cells; calculating

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

the Chi-Square statistic and comparing it against a critical value from the Chi-Square distribution allows the researcher to assess whether the association seen between the variables in a particular sample is likely to represent an actual relationship between those variables in the population. If the null hypothesis is accepted there would be no relationship between the variables (gender, age) and ADHD. If the null hypothesis is rejected, the implication would be that there is a relationship between gender and ADHD. This paper is organized into three sections. Section II discusses the data samples, classification using statistical parameters and Statistical Analysis using Chi-Square test. Experimental Results and analysis on ADHD data set using Chi-Square test statistical tool is presented in Section III. t merits and demerits and finally presenting the conclusion.

II. METHODOLOGY

The methodology of ADHD diagnosis using statistical measures is provided in this section. The theoretical parameters (mean and standard deviation) are applied to classify the data samples and Chi-square Statistical test is proposed to examine the level of significance between ADHD and the factors Age & Gender.

a. Data Samples

The ADHD research study was carried out among 105 school going children aged 5–9 years in the southern region of Tamil Nadu, India 2016. Out of the data samples, 47 children were male and 58 children were female. A sample questionnaire with three parts each consisting of ten questions related to the symptoms of different types of ADHD is prepared. To measure the severity of ADHD each question is rated by a 4-point scale as given in Table I and the total score of the questionnaire for each group ranges from 0 to 30.

TABLE I: ADHD RATING SCALE

S.No	Scale	Score
1	Never	0
2	Frequently	1
3	Often	2
4	Very Often	3

b. Data Analysis and Classification

The aim of this research work is to initially classify the data samples into two groups namely, children with or without ADHD and then further classify them into high, moderate and No ADHD. Various statistical measures which are proposed in this paper to identify the level of ADHD are given below.

(i)Arithmetic mean $(A) = \frac{1}{n}\sum_{i=1}^{n} x_i$ ---------- **(1)**

(ii)Standard Deviation $(S.D) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - A)^2}$------ **(2)**

Where,

n = Number of items

x_i = Value or Rating of each individual item

A = Arithmetic mean

For the input data samples, the mean and standard deviation values are 23 & 17. Using these values, the data set is classified into highADHD, moderateADHD and noADHD as below.

highADHD = A + S.D = 23 + 17=40
noADHD =A - S. D =23 – 17=6
moderateADHD = Score between High level & Low level

In ADHD data set, the sample is classified into highADHD if the score >= 40, moderateADHD if the score is 6<score<= 40 and noADHD if the score <= 6. The overall classification of the children is presented in Table II. From the table it has been identified that out of 105 samples 21 (20%) children are affected with highADHD, 60 (57%) are affected with moderateADHD and 24(23%) are affected with noADHD. The pie chart representation of this ADHD classification is given in Fig 1.

TABLE II: ADHD CLASSIFICATION USING MEAN & STD. DEVIATION

S.No	ADHD Measure	No. of Children's	Percentage (%)
1	High	21	20
2	Moderate	60	57
3	No	24	23
	Total	105	100

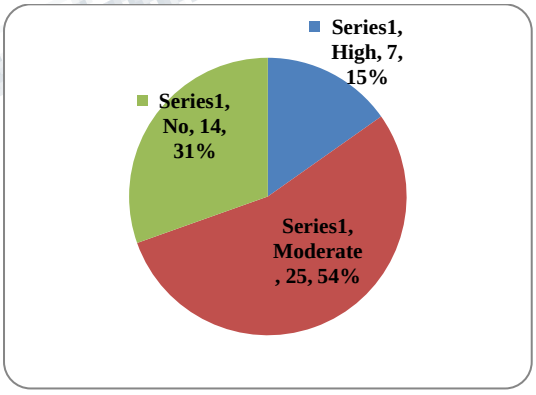

Fig 1 ADHD Measure

From Fig 1, it is surmised that out of 105 respondents 20% has highADHD, 57% has moderateADHD and 27 % has noADHD.

c. Statistical Analysis using Chi-Square test

Chi-square is a statistical test used to test the goodness of fit to verify the distribution of observed data with assumed theoretical distribution. It is a measure to study the divergence of actual and expected frequencies. If

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

there is no difference between the actual and expected frequencies, the Chi-square value (χ^2) is zero.

Hypothesis:

Stating hypothesis about the probability distribution of one or more data samples is an important procedure of this method. The main function of this testing is to choose two challenging hypotheses. In this research work, two factors (age &gender) are selected for the analysis of ADHD using Chi-Square test. The hypotheses formulated for these factors are given below.

AGE factor

- H_1: There is no significant difference in ADHD due to Age.

GENDER factor

- H_2: There is no significant difference in ADHD due to Gender.

Observed Frequency (O):

Consider the 2x2 contingency table.

TABLE III: OBSERVED VALUES

	Y1	Y2	
X1	a_{11}	a_{12}	a_{11+} $a_{12=A}$
X2	a_{21}	a_{22}	a_{21+} $a_{22=B}$
	a_{11+} $a_{12=C}$	a_{11+} $a_{12=D}$	N=A+B+C+D

In this table III, the observed values are presented in four cells. X represents the independent variable and Y represents the response variable. The marginal values are presented by A, B, C and D. In each row and column the marginal values present the summation of that row or column. Here, the observed variables (a_{11}, a_{12}, a_{21}, a_{22}) are given in Table III.

Expected Frequency (E):

The expected values are computed using (3) and provided in Table IV.

$$E = \text{Expected Frequency} = \frac{\text{Row Total} * \text{Column Total}}{\text{Total Number of Respondents}} \quad \textbf{(3)}$$

TABLE IV: EXPECTED VALUES

	Y1	Y2	
X1	(A*C)/N	(A*D)/N	A
X2	(B*C)/N	(B*D)/N	B
	C	D	N

Chi-square value (χ^2):

The value of Chi-square is calculated as below:

$$\text{Chi-square Test } (\chi^2 = \frac{\sum (O-E)^2}{E} \quad \text{---------- (4)}$$

Where, O = Observed Frequency & E = Expected Frequency

Degree of Freedom (DF):

The Degree of Freedom represents the number of classes to which values can be assigned at will, without violating restrictions.

$V = n-k$

Where V - Degree of Freedom, n – Number of frequency classes, k – Number of independent constraints.

For 2 X 2 contingency table, the Degree of Freedom

$DF = (r-1) * (c-1)$ ---------- **(5)**

where,

r = Number of Rows

c = Number of Columns

ie, DF = (2-1) * (2-1) = 1

Chi-Square Distribution Table:

The Chi-Square distribution values for different level of significances (P) are provided in Table V.

TABLE V: CHI-SQUARE DISTRIBUTION TABLE

DF	Probability (P)		
	P=0.05	P=0.001	P=0.0001
1	3.84	6.64	10.83
2	5.99	9.21	13.82
3	7.82	11.35	16.27
4	9.49	13.28	18.47
5	11.07	15.09	20.52
6	12.59	16.81	22.46
7	14.07	18.48	24.32
8	15.51	20.09	26.13
9	16.92	21.67	27.88
10	18.31	23.21	29.59

The procedure for Chi-square test is given below.

1. A hypothesis is established along with the significance level.
2. Create a contingency table with Observed values (O).
3. Calculate the Expected values (E).
4. Compute the deviations between Observed values and Expected values (O-E).
5. Find the Chi-square value (χ^2) .
6. Find the Degrees of freedom (DF).
7. Find the value of χ^2 from χ^2 Distribution table at certain level of significance, usually 5% level.

If the calculated value of χ^2 is greater than the table value of χ^2, at certain level of significance, reject the hypothesis. If the computed value of χ^2 value is zero, then the observed value and expected value completely coincide. If the calculated value of χ^2 is less than the table value of χ^2, at certain level of significance, it is said to be

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

non-significant. This implies that the discrepancy between the observed and expected frequencies may be due to fluctuations in simple sampling.

III. EXPERIMENTAL ANALYSIS & RESULTS

Once the ADHD data samples are classified into high, moderate and noADHD, statistical analysis is carried out with the two factors age and gender to check whether they have any significance to the cause of ADHD.

A. Hypothesis 1

H_0: Age has no significance for the cause of ADHD (ADHD & AGE are independent).

Step 1: Compute the *Observed Frequencies (O).*

TABLE VI: OBSERVED VALUES OF AGE AND LEVEL OF ADHD CHILDREN

ADHD Measure	Age (years)			Total
	6	7	8	
High	7	5	9	21
Moderate	25	17	18	60
No	14	5	5	24
Total	46	27	32	105

From Table VI, out of 46 respondents under the age group of 6 years, 7 have highADHD, 25 have moderateADHD and the remaining 14 have noADHD. Out of 27 respondents in the age group of 7 years, 5 have highADHD, 17 have moderateADHD and the remaining 5 have noADHD. Out of 31 respondents is age group of 8 years, 9 have high ADHD, 18 have moderateADHD and the remaining 5 have no ADHD. The relationship between the age and the level of ADHD is given in fig 2.

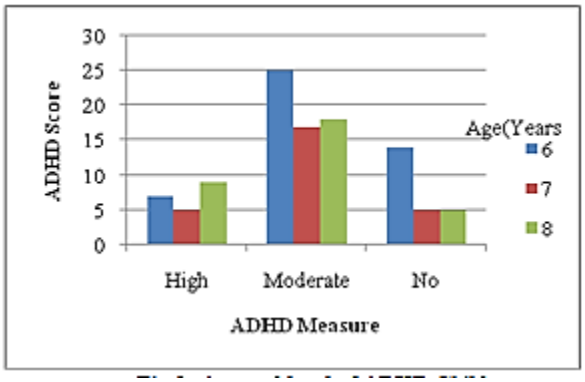

Fig 2: Age and level of ADHD Children

Step 2: Calculate the Expected frequencies (E).

The expected values for the table of observed frequencies are calculated and listed in Table VII.

TABLE VII: EXPECTED VALUES OF AGE AND LEVEL OF ADHD

ADHD Measure	Age (years)			Total
	6	7	8	
High	9.2	5.4	6.4	21
Moderate	26.29	15.43	18.29	60
No	10.51	6.17	7.31	24
Total	46	27	32	105

Step 3: Find the *Chi-square Value* (χ^2).

To find out the discrepancy between age and the level of ADHD the Chi-Square value is calculated using the observed and expected frequencies.

TABLE VIII: CHI SQUARE TEST OF H_0

Cell	O	E	C(chi-square value)		
			O-E	$(O-E)^2$	$(O-E)^2/E$
R_1C_1	7	9.2	-2.2	4.84	0.5261
R_1C_2	5	5.4	-0.4	0.16	0.0296
R_1C_3	9	6.4	2.6	6.76	1.0563
R_2C_1	25	26.29	-1.29	1.6641	0.0633
R_2C_2	17	15.43	1.57	2.4649	0.1597
R_2C_3	18	18.29	-0.29	0.0841	0.0046
R_3C_1	14	10.51	3.49	12.1801	1.1589
R_3C_2	5	6.17	-1.17	1.3689	0.2219
R_3C_3	5	7.31	-2.31	5.3361	0.7300
Total					**3.9504**

Step 4: Find the *Degrees of Freedom (DF)*.

In this Hypothesis, Observed values are presented in 9 cells that are 3 columns and 3 rows. From equation (5) the Degrees Of Freedom is calculated as,

Degrees of Freedom = (c-1) (r-1)

= (3-1) (3-1)
= 2*2
= 4

Step 5: Select The *Table Value*.

The Chi-Square distribution value for the significance level of 5% and degree of freedom 4 is obtained from Table V, as 9.49. The calculated value of $\chi^2$is 3.9504, which is less than the table value. Therefore, the hypothesis may be accepted and hence it is concluded that Age has no significance for the cause of ADHD.

b. Hypothesis 2

H_1: Gender has no significance for the cause of ADHD (ADHD & Gender are independent)

Step 1: Compute the *Observed frequencies (O).*

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

TABLE IX: OBSERVED VALUES OF GENDER AND LEVEL OF ADHD CHILDREN

ADHD Measure	Gender		Total
	Male(0)	Female(1)	
High	4	17	21
Moderate	27	33	60
No	16	8	24
Total	47	58	105

From Table VI, out of 47 male children, 4 have highADHD, 27 have moderateADHD and the remaining 16 have no ADHD. Out of 58 female children, 17 have highADHD, 33 have moderateADHD and the remaining 8 have noADHD. The relationship between the gender and the level of ADHD is given in Fig 3.

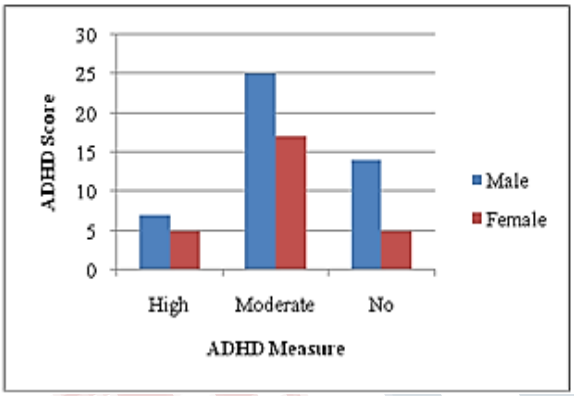

Figure 3: Gender and level of ADHD Children

Step 2: Calculate the *Expected frequencies (E)*.
The expected values for the table of observed frequencies are calculated and listed in Table X.

TABLE X: EXPECTED VALUES OF GENDER AND LEVEL OF ADHD CHILDREN

ADHD Measure	Gender		Total
	Male(0)	Female(1)	
High	9.4	11.6	21
Moderate	26.9	33.1	60
No	10.7	13.3	24
Total	47	58	105

Step 3: Find the *Chi-square value* (χ^2).
To find out the discrepancy between gender and the level of ADHD the Chi-Square value is calculated using the observed and expected frequencies.

TABLE XI: CHI SQUARE TEST OF H_1

Cell	O	E	C(chi-square value)		
			O-E	(O-E)2	(O-E)2/E
R_1C_1	4	9.4	-5.4	29.2	3.1064
R_1C_2	17	11.6	5.4	29.2	2.5172
R_2C_1	27	26.9	0.1	0.01	0.0052
R_2C_2	33	33.1	-0.1	0.01	0.0120
R_3C_1	16	10.7	5.3	28.1	2.6262
R_3C_2	8	13.3	-5.3	28.1	2.1128
Total					10.3798

Step 4: Find the *Degrees of Freedom (DF)*.
In this Hypothesis, Observed values are presented in 6 cells that are 2 columns and 3 rows. From equation (5) the Degrees Of Freedom is calculated as,

Degrees of Freedom = (c-1) (r-1)
= (2-1) (3-1)
= 1 *2
= 2

Step 5: Select The *Table Value*.
The Chi-Square distribution value for the significance level of 5% and degree of freedom 2 is obtained from Table V, as 5.99. The calculated value of $\chi^2$is 10.3798, which is greater than the table value. Therefore, the ***hypothesis may be rejected*** and hence it is concluded that Gender and ADHD are not independent, ie, gender is the significant factor for the cause of ADHD disorder.

IV. CONCLUSION

In this paper, the statistical analysis of ADHD in the age group of 6-9 years of school going children is proposed. Totally 105 samples are collected based on different ages ranging from 6 to nine and gender (Male & Female). The statistical parameters like mean and standard deviation are computed to analyze the level of ADHD under various parameters. The ADHD may be caused by different parameters like age, gender, brain damage, etc. In this research work, two factors age and gender are taken and their significance for the cause of ADHD is tested using Chi-Square Analysis Test. From this statistical analysis it has been concluded that Age has no significance for the cause of ADHD but gender is one of the significant factor for the cause of ADHD disorder.

REFERENCES

[1] Radhamani E, Krishnaveni K, "Prognosis of ADHD using R Programming and MATLAB Tools", International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, ISO 9001:2008, Vol 7, 2017.

[2] Radhamani E, Krishnaveni K, "Diagnosis and Evaluation of ADHD using Naïve Bayes and J48 Classifiers", 2016 International Conference on Computing for Sustainable Global Development (INDIACom), 978-9-3805-4421-2/16/$31.00_c 2016 IEEE, 2016.

[3] Radhamani E, Krishnaveni K, "Diagnosis and Evaluation of ADHD using MLP and SVM Classifiers", Indian Journals of Science and Technology, Vol 9(19)/93853, ISSN 0974-5645, 2016.

[4] Regina Bussing, Johanna Meyer, "Childhood ADHD Symptoms: Association with Parental Social Networks and Mental Health Service Use during Adolescence", International Journal of Environmental Research and Public Health, ijerph120911893, ISSN 1660-4601, 2015.

[5] Daniel Minahim, Luis A. Rohde, "Attention deficit hyperactivity disorder and intellectual giftedness: a study of symptom frequency and minor physical anomalies", Revista Brasileira de Psiquiatria 37:289–295, ISSN 1516-4446, 2015.

[6] Kayvan Miri Lavassani, Bahar Movahedi, "Developments in Analysis Of Multiple Response Survey Data in Categorical Data Analysis: The Case Of Enterprise System Implementation In Large North American Firms1", Spring, Vol 4, 2014.

[7] Sureiman Onchiri, "Conceptual model on application of chi-square test in education and social sciences", Educational Research and Reviews, Vol.8 (15), 1231-1241, 2013.

[8] Anuradha J, Jayasri N, "Classification Rules for Attention Deficit Hyperactive Disorder using Decision tree" International Journal of Computer Science & Communication Networks, Vol 2(2), 295-303, ISSN:2249-5789, 2012.

[9] https://www3.nd.edu/~rwilliam/stats1/x51.pdf.

[10] https://www.trentu.ca/academicskills/documents/chi-square instruction.pdf.

[11]

[12] http://en.wikipedia.org/wiki/Attention_deficit_hyperactivity_disorder.

ACKNOWLEDGMENT

E.Radhamani has received her MCA & M.Phil degree in Computer Science from Sri S.Ramasamy Naidu Memorial College, Sattur affiliated to Madurai Kamaraj University, Madurai, Tamil Nadu. She is pursuing Ph.D in Computer Science. She is also working as a Project Fellow for UGC Major Research Project. Her area of interests are Medical Image processing and Data mining.

Dr.K.Krishnaveni, is an Associate Professor and Head of the Department of Computer Science, Sri S.Ramasamy Naidu Memorial College, Sattur, India. She has 25 years of teaching experience. She received her B.E. degree in Computer Science and Engineering from Bharathiar University, Coimbatore, India in 1990 and M.Tech. degree in Computer and Information Technology from Center for Information Technology and Engineering of Manonmaniam Sundaranar University, Tirunelveli, India in 2004 and received Ph.D degree from the same University. Currently her research interests are Medical Image Processing, Data mining, and Mathematical Morphology. She is a member of IEEE and Editor of various reputed journals.