DJ Journal of Engineering and Applied Mathematics, Vol. 4(1) 2018, pp. 20-28

RESEARCH ARTICLE

Exam Results Fit to a Normal Distribution

*Suresh Akella[1], P.M Diaz[2], B Suresh Babu[3]
[1]Superintend of Exams, Sreyas Institute of Engineering & Technology, Hyderabad, India.
[2]Professor, Department of Mechanical Engineering, Sreyas Institute of Engineering & Technology, Hyderabad, India.
[3]Associate Professor, Department of Mathematics Sreyas Institute of Engineering & Technology, Hyderabad, India.

Received- 9 May 2017, Revised- 27 July 2017, Accepted- 19 September 2017, Published- 9 October 2017

ABSTRACT

Students of many engineering colleges in Telangana are affiliated to universities. Examination branch collects data of internal evaluation and university external exam, and finally these two results are added to determine the semester academic evaluation and to know the grade of position in the academic credits rating. In this paper, these three steps of results are fit to normal distribution. Further, transformation to standard normal are done. The advantages of obtaining future predictions of performance of this batch are possible and remedial classes or improvement plans for estimated students can be planned. This distribution can also be used for normalizing difficult exams to improve the results and see changes in the distribution. Simulation is done with excel calculations.
Keywords: Result Analysis, Distribution of marks, Normal distribution, Normalized results, Fitting distributions.

1. INTRODUCTION

In today's world of open access information, with total learning possible on any subject, a qualification certificate with the time, money and effort should be questioned by students, parents and the college or university providing the course. The advantage of a programme is the rigor of the structured teaching, relative evaluation of a large group of students evaluated, reported and credited for their performance over a period of a semester. This evaluation needs to be accurate, precise, transparent & relative to give credits and marks indicating a group's performance.

As an effort in this area, UK NARIC [1] launches increasing transparency of access qualifications for higher education program in 2004 at Cheltenham, England. The aim of this program was to evaluate European higher education system, to include as many countries, around the world, as possible for continuous improvements of the output for society performance and improvement. The pilot project was meant to drive other nations to provide data based evaluation and comparison with statistical basis, to assist admissions to the universities across Europe and avoid migration to other countries and to evaluate the European credit and grade system. Distribution, was used to evaluate grade system, and evaluation of removing lower grade affect was seen to be worth considering. In this report it was proposed that, AAA in top grade boundary has

1
2
3
4
UEE – 99.48%
5

By removing the fail grade '5' of qualification X, the number of students falls by 10%. Using the constant 99.48 from the base data, it is possible to divide the total percentage of students who achieve the minimum required A-level grades (99.48) by the total percentage of students who pass qualification X 99.48, 1.1053333…=90. Also, it was concluded that the distribution study revealed very little variation in the results as similarity in cumulative distributions were maintained. The authors of [2], have created an algorithm for students exam result data analysis. They have incorporated data from admission, results through the programme and beyond. They have also

*Corresponding author. Tel.: +919849628282
Email address: principal@sreyas.ac.in (S.Akella)
Double blind peer review under responsibility of DJ Publications
https://dx.doi.org/10.18831/djmaths.org/2018011002

referred [3-6] the algorithms used for school results analysis in different places like USA and UK. Many colleges & universities like in [7], evaluate the results for basic comparisons, & segregate using histograms, data tables and graphs for quick review for reports and decision making. Schools in China [8] with largest number of students population also provide analysis related to this.

Table 1.Final marks of FLAT course

No	Final marks	No	Final marks	No	Final marks	No	Final marks	No	Final marks	No	Final marks
1	9	45	60	89	55	133	32	177	40	221	57
2	21	46	46	90	41	134	60	178	61	222	26
3	45	47	59	91	55	135	27	179	43	223	32
4	55	48	48	92	44	136	54	180	60	224	48
5	45	49	47	93	70	137	51	181	27	225	14
6	34	50	65	94	63	138	24	182	40	226	46
7	6	51	43	95	74	139	47	183	64	227	25
8	52	52	47	96	40	140	65	184	19	228	35
9	32	53	29	97	79	141	31	185	59	229	53
10	12	54	48	98	46	142	64	186	49	230	32
11	27	55	43	99	41	143	20	187	51	231	21
12	56	56	62	100	40	144	60	188	41	232	40
13	28	57	24	101	46	145	63	189	18	233	21
14	16	58	46	102	23	146	52	190	19	234	29
15	29	59	56	103	67	147	46	191	26	235	14
16	51	60	56	104	72	148	72	192	40		
17	15	61	45	105	42	149	44	193	65		
18	27	62	52	106	44	150	24	194	19		
19	44	63	33	107	30	151	61	195	86		
20	34	64	54	108	49	152	65	196	40		
21	35	65	65	109	30	153	26	197	74		
22	56	66	55	110	46	154	43	198	49		
23	31	67	46	111	53	155	64	199	12		
24	52	68	30	112	59	156	57	200	49		
25	57	69	31	113	21	157	51	201	69		
26	16	70	66	114	51	158	58	202	48		
27	41	71	33	115	33	159	36	203	64		
28	14	72	30	116	33	160	48	204	44		
29	63	73	14	117	47	161	56	205	45		
30	31	74	41	118	51	162	50	206	40		
31	69	75	30	119	52	163	30	207	68		
32	53	76	51	120	22	164	48	208	27		
33	42	77	51	121	47	165	46	209	44		
34	63	78	21	122	58	166	59	210	57		
35	23	79	43	123	32	167	47	211	67		
36	43	80	58	124	52	168	28	212	60		
37	16	81	51	125	48	169	30	213	19		
38	64	82	28	126	52	170	32	214	19		
39	47	83	70	127	56	171	47	215	62		
40	51	84	81	128	38	172	43	216	44		
41	46	85	22	129	56	173	32	217	47		
42	61	86	44	130	52	174	51	218	12		
43	60	87	41	131	35	175	24	219	49		
44	52	88	52	132	44	176	52	220	45		

Author evaluated under graduate engineer's course on materials research methods and the difficulty of the question paper with respect to the results. Analysis had histograms for each level of difficulty.

In this study, an engineering college affiliated to an university results were evaluated and reported. The results of a course, FLAT, of second year students of 235 strength are reported. This is a 2^{nd} year II semester course for CSE, Computer Science Engineers. There are two internal midterm tests, with university objective papers, internal subjective papers and assignments, all adding to internal marks of 25. Next, the students write an university exam for 75 marks, this data is referred to as external exam results. In the external exam, student should get minimum 26 out of 75 to be eligible to pass the course. Finally, university adds these test results and gives a final marks for the course for 100 marks, the pass mark is 40 out of 100. For an undergraduate course, FLAT, for 2^{nd} year 2^{nd} semester computer science engineers data of the class of 235 students final marks is shown in table 1. All three, internal, external & final results marks of random variables, x, fit into normal, Gaussian distribution. Next, it is shown that the transformed data into standard form, of Z variables also form normal distribution as expected. Symmetry of normal distribution & inflexion points are shown. The final results if university needs to normalize, by adding constant marks say, 5 or 10, it is added to the final marks which is shown that the distribution remains same. This work will be useful for many institutes & universities to analyze, and see probability of pass students for each course program in statistically valid method. Further work is going on in applying central limit theorem & test of hypothesis to validate predictions for sub groups in the class. All the calculations are done by simulation in excel.

2. NORMAL DISTRIBUTION

Gauss has seen the frequency distribution of many naturally varying parameters like height, weight of humans etc. and classified it as an exponential function. [9] This distribution was symmetric about the mean, had the shape of a bell and was also popular as "Bell Curve". If random variable, x has a mean, μ and a standard deviation, σ and variance σ is 2, the function of the frequency of occurrence or the probability density, f(x) is defined as:

$$f(x,\mu,\sigma^2) = \frac{1}{\sqrt{2\mu\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{1}$$

In the present study, x is the marks of each of the 235 students of CSE department, 2^{nd} year 2^{nd} semester. Marks are for a course called FLAT. When this data of marks for internal, external and final were evaluated, the data forms the normal distribution. The x axis of this is variable x, y axis is the frequency distribution, f(x) and the shape of the distribution is a Bell curve of Gaussian natural variation, which are also defined here in the form of normal distribution.

3. NORMAL DISTRIBUTION FOR X OR STANDARD NORMAL FOR Z

Further, the normal distribution variable of x may be transformed, to a variable Z where Z is a function of variable x, it's mean μ and standard deviation σ.

$$Z = \frac{(x-\mu)}{\sigma} \tag{2}$$

This form of transformation is useful in calculating the probability. In the standard normal distribution, the mean gets shifted to zero, $\mu = 0$ and standard deviation to one, $\sigma = 1$. The probability function now becomes $\phi(x)$ for this distribution. The advantage of this form is that standard tables are available and can be used for calculating any value. In this study for example, we can estimate the number of students who have attained more than 14/25 in internal exams; more than 26/75 in external exams or more than 40/75 in final exams. The general value of x is between $-\infty$ and ∞. However for the data selected, the values of x(internal marks) is 0 to 25; the values of x(exernal marks) is 0 to 75 and the values of x(final marks) is 0 to 100

$$\emptyset(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{3}$$

Table 2.Probability from the distribution curve

Distance Normal	Distance in standard Normal	% Probability of occurrence (Area under curve)
μ - σ to μ +σ	-1 to 1	68.26
μ -2σ to μ +2 σ	-2 to 2	95.44
μ -3σ to μ +3 σ	-3 to 3	99.73

The probability distribution of $\emptyset$(z) vs z will give the probability distribution of the standard random variable, z. The area under the curve $-\infty < z < \infty$ is equal to 1. As it is symmetric, it is distributed 0.5 on each side. The area in % can be given as in table 2.

The requirements for the exam analysis is to find out how many are above 14 marks in internal exams, above 26 in external exams & above 40 in final exams.

4. INFLEXION POINT IN BELL CURVE

For standard normal distribution, inflexion calculation is taken as x instead of z. For the Bell curve, the slope of probability density curve comes down very fast up to z =1=σ and then there is inflexion, change. For a normal distribution, if mean μ is the central point, then $\mu-\sigma$ and $\mu+\sigma$ are the inflexion points, where the slope of the continuous curve changes. Inflexion can be calculated for normal frequency distribution with variable x or for the standard normal distribution $\emptyset$(x), with μ =0 and σ=1, which is done in the equations below. For inflexion point, $\emptyset^2$(x)=0, and $\emptyset^3$(x) not equal to zero.

$$\frac{d\emptyset(x)}{dx} = \frac{1}{2\pi}e^{-\frac{x^2}{2}}\left(\frac{2x}{2}\right) = -\frac{x}{2\pi}e^{-\frac{x^2}{2}} \tag{4}$$

$$\frac{d^2\emptyset(x)}{dx^2} = \frac{1}{2\pi}e^{=\frac{x^2}{2}}(-x)+e^{-\frac{x^2}{2}}\left(\frac{-1}{2\pi}\right) \tag{5}$$

$$\frac{d^2\emptyset(x)}{dx^2} = \frac{1}{2\pi}\left[x^2e^{-\frac{x^2}{2}} - e^{-\frac{x^2}{2}}\right] \tag{6}$$

$$\frac{d^2\emptyset(x)}{dx^2} = \frac{-e^{-\frac{x^2}{2}}}{2\pi}[x^2-1] \tag{7}$$

From equation (7), to obtain the inflexion point, equate the RHS to zero.

$\emptyset''(x) = 0$ gives us $x^2-1=0$, which gives $x=-1$ and +1. For standard normal, $Z=-\sigma$ and $+\sigma$ are the inflexion points where the slope changes and becomes slower as in a Bell.

5. ANALYSIS & DISCUSSION

The class of 235 students of 2^{nd} year 2^{nd} semester in computer science engineering is taken for internal exams conducted by institute & university evaluated by institute staffs; external exam conducted by university and evaluated by staffs from all affiliated colleges; final marks are an addition of these two exam results. Table 1 gives the 235 students final marks, similarly, internal and external marks are available for a CSE course FLAT.

The data of all students, N=234, internal marks evaluated for a total of 25, are in excel to obtain mean, $\mu=17.56$ and standard deviation, $\sigma=3.55$. These values are evaluated for each student marks, x, to obtain the frequency distribution, f(x) given in equation (1). The distribution gives a normal distribution as shown in figure 1. A student, if he gets more than 14 marks out of 25 he has a possibility to pass as he needs to obtain a minimum of 26 marks out of 75 from external marks to finally fetch a sum of 40 out of 100 marks of total final results.

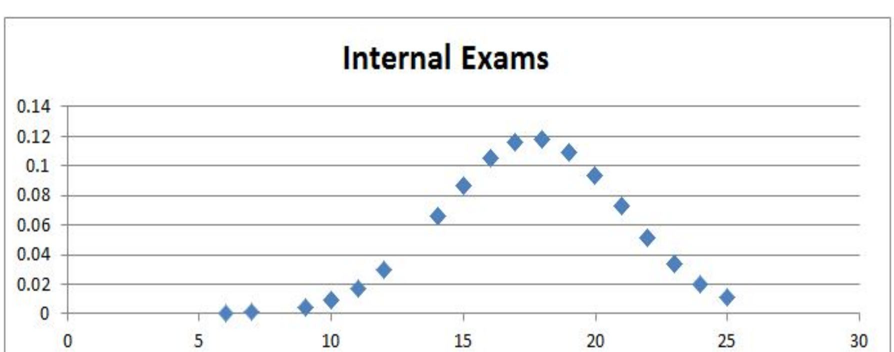

Figure 1.Distribution of internal marks

External exams are conducted by the university for 75 marks, examination paper circulated to different colleges where exams are held. These papers are commonly evaluated by staffs who taught this subject and reported. The data of all students, N=235, external marks evaluated for a total of 75, are in excel to obtain mean, $\mu = 26.4$ and standard deviation, $\sigma = 13.86$. These values are as done for internal marks. The distribution gives a normal distribution as shown in figure 2. A student if he gets more than 26 marks out of 75 he has a possibility to pass as he needs to obtain a minimum sum of 40 out of 100 marks of total final results.

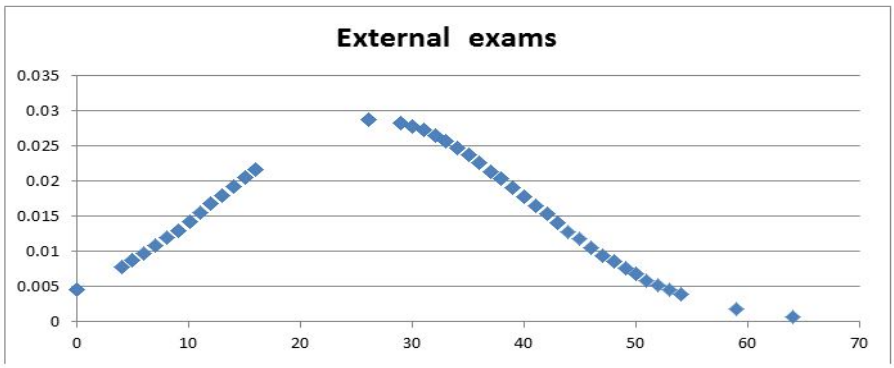

Figure 2.Distribution of external marks

Final marks for all 235 students is obtained by summing the internal and external marks. These are shown in table 1. Similar tables were available for both internal and external marks. Again the same procedure to fit the data to normal distribution of final marks is done. From the data of all students, N=235, mean, $\mu = 43.96$ and standard deviation, $\sigma = 15.75$. The frequency distribution gives a normal distribution as shown in figure 3. A student if he gets more than 40 out of 100 marks of total final results, would be declared pass.

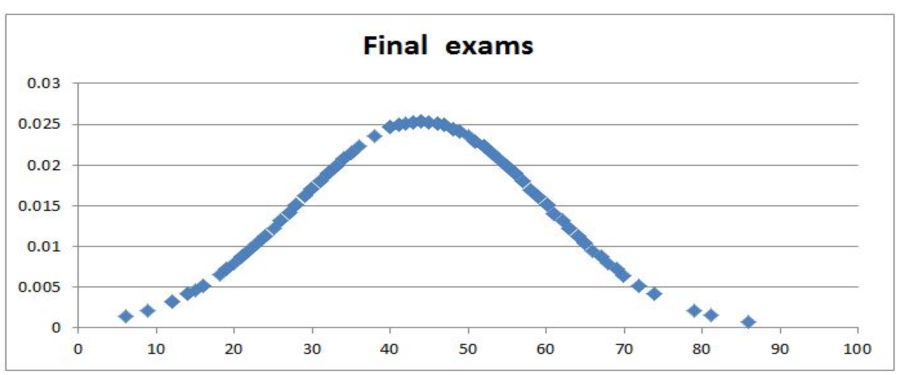

Figure 3.Distribution of final marks

6. STANDARD NORMAL DISTRIBUTION:

Once the data, x, is found to be of normal distribution, it can be transformed to the standard normal distribution, which has $\mu = 0$ and $\sigma = 1$. This transformation is obtained by defining z as in equation (2). Using these values of z, μ & σ, the standard normal function of probability density, $\phi(z)$ is obtained, using equation (3). Again the range of z here can be $-\infty < z < \infty$. Standard tables give the area under the curve which gives the probability of occurrence of that event. Figure 4 gives the standard normal distribution for internal exam marks. Our interest is to find the probability of students who gets more than 14/25. In equation (2), x is used as 14 and z is evaluated as -1.08. From z using equation (3), these results are summarized in table 3. The area under the standard normal from $-\infty$ to -1.08 is .14; that is 14% of students obtained below 14 marks in the internal exams. As total area under the distribution is probability, P, of all events of results, P=1-.14 = 0.86 gives 86% of students have obtained above the defined 14 marks. The data from N=235 students for internal marks gives N*P =201 students obtaining more than 14/25 from standard distribution from actual data it is 220 and the difference % of these two values, defined as error of normal fit to the data will be 8.6%.

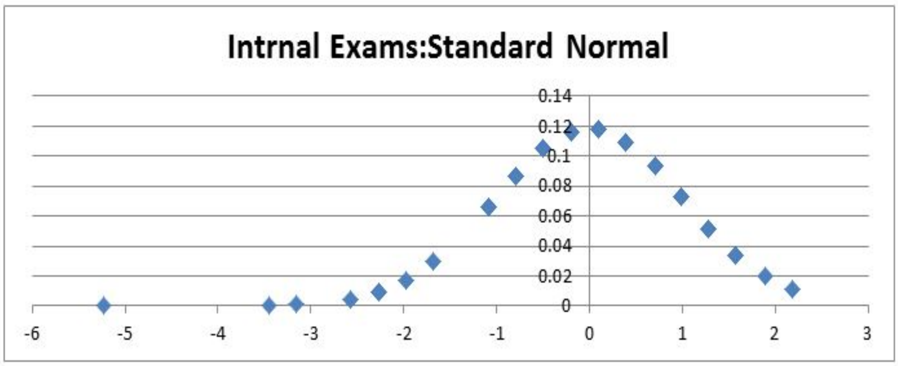

Figure 4.Standard normal distribution of internal marks

Figure 4 gives the standard normal distribution for internal exam marks. As our interest is to find the probability of students getting more than 14/25. In equation (2), x is used as 14 and z is evaluated as -1.08. From equation (3) Z is calculated and all values are obtained from standard normal table, or Z table, these results are summarized in table 3. The area under the standard normal from $-\infty$ to -1.08 is .14; that is 14% of students obtained below 14 marks in the internal exams. As total area under the distribution is probability, P, of all events of results, P=1-.14 = 0.86 gives 86% of students have obtained above the defined 14 marks. The data from N=235 students for internal marks gives N*P =201 students obtaining more than 14/25 from standard normal distribution. From actual data it is 220 and the difference % of these two values, defined as error of normal fit to the data will be 8.6%.

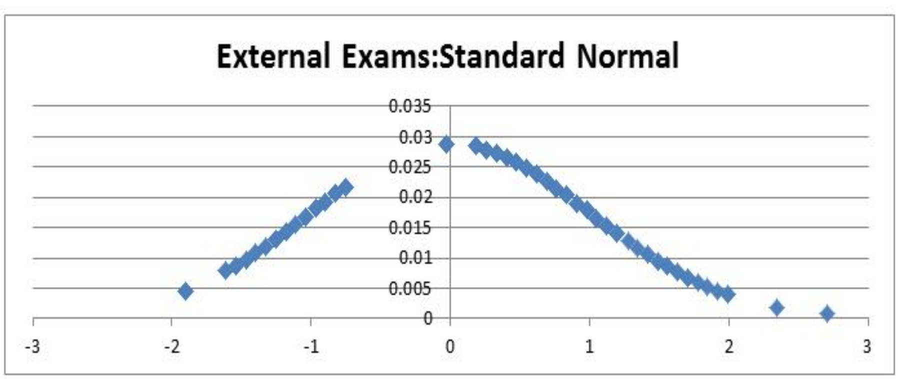

Figure 5.Standard normal distribution of external marks

Figure 5 gives the standard normal distribution for external exam marks. Our aim is to find the probability of students getting more than 26/75. In equation (2), x is used as 26 and z is evaluated as -0.29. Negative sign indicates that less than 50% of students have obtained below 26 marks and more than 50% have obtained above

26 marks. Using equation (3) value of Z is obtained, these results are summarized in table 3. The area under the standard normal from $-\infty$ to -.029 is .488; that is 48.8% of students obtained below 26 marks in the external exams. As total area under the distribution is probability, P, of all events of results, P=1-.488= 0.512 gives 51.2% of students have obtained above the required 26 marks. The data from N=235 students for internal marks gives N*P =120 students obtaining more than 26/75 from standard distribution. From actual data, it is 159 and the difference % of these two values, defined as the error of normal fit to the data will be 24.5%. This fit has high error and the students have not performed to fit properly to a normal distribution in the external exam.

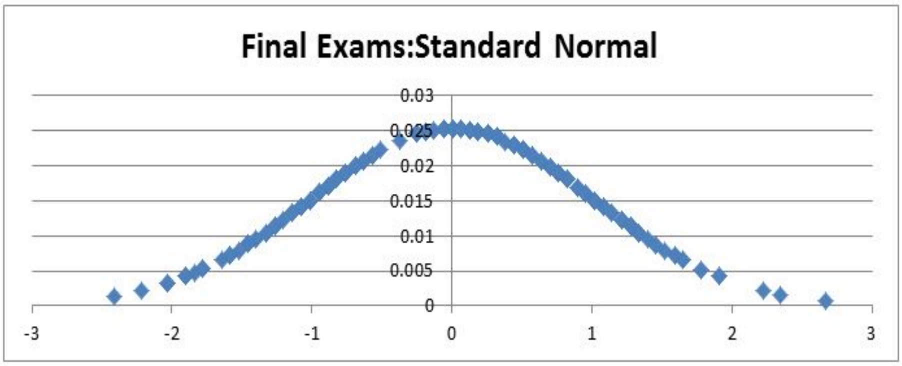

Figure 6.Standard normal distribution of final marks

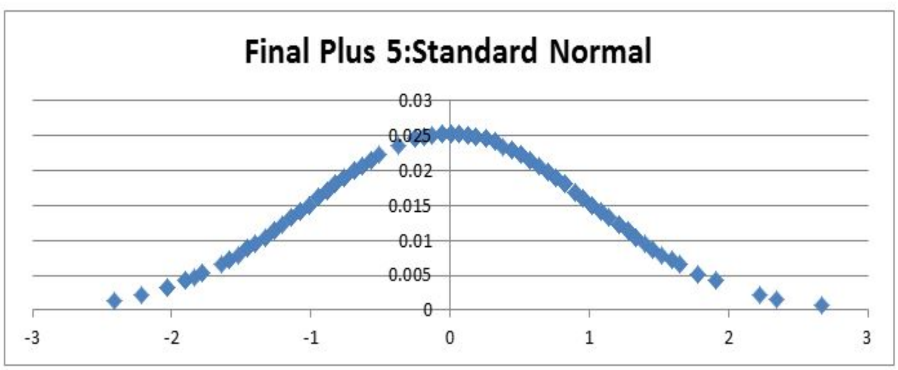

Figure 7.Standard normal distribution of final marks plus 5

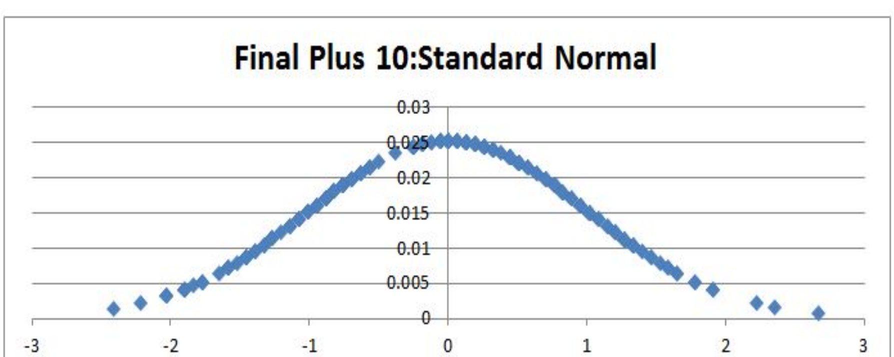

Figure 8.Standard normal distribution of final marks plus 10

Figure 6 gives the standard normal distribution for final marks. Our interest is to find the probability of students getting more than 40/100. In equation (2), x is used as 40 and z is evaluated as -0.25. Using equation (3) value of Z is obtained, these results are summarized in table 3. The area under the standard normal from $-\infty$ to -.029 is .401; that is 40.1% of students obtained below 40 marks in the external exams. As total area under the distribution is probability, P, of all events of results, P=1-.401= 0.598 gives that 59.8% of the students have obtained above the required 40 marks. The data from N=235 students for final marks gives N*P =140 students obtaining more than 40/100 from standard distribution. From actual data it is 158 and the difference % of these two values, defined as error of normal fit to the data, will be 11.4 %. This fit is assumed to be a medium fit of the data to a normal distribution.

Table 3.Validating data fitment to normal distribution for students N=234

Value	Internal Marks	External Marks	Final Marks	Final Marks +5	Final Marks +10
Required Variable, x	14	26	40	40	40
Mean,	17.64	26.4	43.96	48.96	53.96
Standard Deviation,	3.37	13.84	15.75	15.75	15.75
Standard Variant, Z	-1.08	-.029	-.25	-.56915	-.8686
Probability, P(<z) Table	.14008	.488	.40129	.284	.188
P(>Z)	.86	.512	.59871	.716	.812
Normal Freq., N*P	201	120	140	168	190
Freq. from data	220	159	158	162	187
% Error of freq.	8.6	24.5	11.4	3.6	1.6
Error of data Fit to Normal dist.	Ok	High	Medium	V Good	Excellent

The main aim is to make the data fit better by assuming that the normal distribution is a correct representation of students appearing for exams to test their natural ability to understand, learn and write as per the exam pattern. Also, there could be examination questions which are out of syllabus, set tougher than a standard of university, or due to some external disturbances during the semester. Such causes might require some corrections to normalize the final marks. With these assumptions, two cases were seen adding 5 marks in case 1 and 10 marks in case 2. Results of the analysis are added in table 3. The mean shifts by the added marks, μ of final exam of 43.96 becomes 48.96 and 53.96 respectively. However, the standard deviation, σ, remains unaltered at 15.75 for all cases. It means the shape of the normal distribution remains unaltered as in figures 7 & 8. The fitment of data to normal distribution becomes very good with 3.6% error, when 5 marks are added and becomes an excellent fit with 1.6% error when 10 marks are added.

7. FURTHER WORK

Works have to be made to study skewness and kurtutiosis of the distribtion, to extend the class performance in other courses of same semester, to note semester-wise students performance and to compare groups of girls to boys, hostel to non-hostel and management to conveynor.

A larger population of university data is required to define the overall results and comparisons with different colleges.

Comparison requires test of hypothesis and segregation of data. Central limit theorem applies to smaller samples in a normally distributed population.

8. CONCLUSION

Data of one course FLAT for internal exams, external exams & final exams have shown that the distribution is a proper selection, which fit to normal distribution. Data was transformed to standard normal for obtaining probability estimates from standard tables. Error estimate of fit is 8.6 to 24.5 to 11.4. Addition of marks improved the error of data to normal distribution. Any revision, addition or normalizing is also shown to follow same distribution with a shift in mean and same standard deviation.

9. ACKNOWLEDGEMENT

Data from JNTUH & Sreyas Institute of Engineering & Technology was used for this analysis.

REFERENCES

[1] Increasing Transparency of Access Qualifications for Higher Education in Europe, Report No. 2004-3279/001-001/SO2 61-NAR, UK NARIC, 2004.

[2] Abel U.Osagie, Abu Mallam, Students Record Analysis And Examination Result Computation Algorithm, International Journal of Technology Enhancements and Emerging Engineering Research, Vol. 2, No. 8, pp. 49.

[3] E.O.Ukem and F.A.Ofoegbu, A Software Application for University Student's Results Processing, Journal of Theoretical and Applied Information Technology, Vol. 35, No. 1, 2012, pp. 34-43.

[4] American Association of School Administrators (2002). "Using data to improve schools: Whats working".Available:www.aasa.org/cas/UsingDataToImproveSchools.pdf.

[5] J.C.Wayman, S.Stringfield and M.Yakimowski, Software Enabling School Improvement through Analysis of Student Data, Center for Research on the Education of Students Placed At Risk (CRESPAR), U.S. Department of Education, 2004.

[6] J.H.Chrispeels, Purposeful Restructuring: Creating a Culture for Learning and Achievement in Elementary Schools, Falmer Press, Taylor and Francis, Bristol, PA, 1992.

[7] https://www.gdc-uk.org/api/files/ResultsSeptPart1.pdf.

[8] W.Yuan, C.Deng, H.Zhu and J.Li, The Statistical Analysis and Evaluation of Examination Results of Materials Research Methods Course, Creative Education, Vol. 3, No. 07, 2013, pp. 162-164, http://dx.doi.org/10.4236/ce.2012.37B042.

[9] P.Maya and T.Nicholas, Some Results on Integer Edge Cordial Graph DJ Journal of Engineering and Applied Mathematics, Vol. 3, No. 1, 2017, pp. 1-12, http://dx.doi.org/10.18831/djmaths.org/2017011001.