CHAPTER I

PREDICTION OF COLLEGE ACHIEVEMENT IN ENGLISH LANGUAGE

Introduction

The importance of language as one of the basic factors of college success, has been repeatedly stressed through factor analytic studies. Thurstone, who applied an extensive battery of tests to university students, as also Burt, who made factor analysis of college examination marks found verbal ability as one of the essential factors of scholastic achievement. Even Spearman's g factor is often found mixed with verbal saturations, which shows that verbal component is closely related with intelligence. It may be further noted that whereas correlations of intelligence tests such as abstract reasoning tests with the college grades are found to be low, verbal tests correlate sufficiently well at this level also. But the correlation approach is confined to study the linear relationship only and it may be interesting to examine further if improvement can be made using polynomial regressions for predicting college achievement.

Problem

The purpose of this study is, therefore, to seek answer to the problem:

Whether linear mathematical function is adequate for predicting language achievement at college and whether or not polynomial regression improves the prediction?

From the data mentioned earlier, the marks in SSCE English and PScE English were taken up for analysis. The SSCE marks serve as predictor variables to forecast the criterion of PScE English marks. The data was set up into a convenient bivariate distribution (Appendix 1) and the least square method was used for the analysis. The statistical compulations necessary for the analysis are shown below step by step.

Step 1. CALCULATION OF DATA REQUIRED FOR FITTING POLYNOMIAL OF THE THIRD DEGREE

1	2	3	4	5	6	7	8	9	10	11	12
x	Totals y Arrays	xy	x^2y	x^3y	f_x	$f_x x$	$f_x x^2$	$f_x x^3$	$f_x x^4$	$f_x x^5$	$f_x x^6$
-4	-61	244	-976	3904	14	-56	224	-896	3584	-14336	57344
-3	-32	96	-288	864	13	-39	117	-351	1053	-3159	9477
-2	-28	56	-112	224	18	-36	72	-144	288	-576	1152
-1	-64	64	- 64	64	56	-56	56	- 56	56	- 56	56
0	-31	0	0	0	67	0	0	0	0	0	0
1	18	18	18	18	51	51	51	51	51	51	51
2	70	140	280	560	30	60	120	240	480	960	1920
3	52	156	468	1404	19	57	171	513	1539	4617	13851
4	33	132	528	2112	9	36	144	576	2304	9216	36864
5	6	30	150	750	1	5	25	125	625	3125	15625
Σ	-37	936	4	9900	278	22	980	58	9980	-158	136340

Step 2. SOLUTION OF SIMULTANEOUS EQUATIONS FOR FITTING A POLYNOMIAL

	1 N	2 x	3 x^2	4 x^3	5 y
1	278	22	980	58	-37
x		980	58	9980	936
x^2			9980	-158	4
x^3				136340	9900
	278	22	980	58	-37
	1.0	0.079136691	3.525179856	0.208633094	-0.133093525
		978.2590	-19.5540	9975.4101	938.9281
		1.0	-0.019988572	10.197105368	0.959795003
			6524.9329	-163.0662	153.1995
			1.0	-0.024991246	0.023479092
				34603.5162	337.1993
				1.0	0.009744654

Step 3. FITTING A CUBIC EQUATION

$\therefore$ $d = 0.009744654$

$c + d\,(-0.0024991246) = 0.023479092$

$c = 0.023722623$

$b + c\,(-0.019988572) + d(10.197105368) = 0.959795003$

$b = 0.860901921$

$a + b\,(0.079136691) + c(3.525179856) + d(0.208633094) = -0.133093525$

$a = -0.286882024$

$\therefore$ Fitted Cubic Equation is given as-

$y = 0.286882024 + 0.860901921x + 0.023722623x^2 + 0.009744654x^3$

Where $y = \frac{Y - 72}{5}$ and $x = \frac{X - 57}{5}$

Step 4. FITTING A QUADRATIC EQUATION

$c = 0.023479092$
$b + c\ (-0.019988572) = 0.959795003$
$b = 0.960264317$
$a + b\ (0.079136691) + c(3.525179856) = -0.133093525$
$a = -0.291853688$

$\therefore$ The fitted Quadratic Equation is given as-

$$y = -0.291853688 + 0.960264317x + 0.023479092x^2$$

Where, as before, $y = \dfrac{Y - 72}{5}$ and $x = \dfrac{X - 57}{5}$

Transferred to original units, the quadratic equation is given as-

$$Y = 0.004695818X^2 + 0.424941065X + 31.062378173$$

Step 5. FITTING A LINEAR EQUATION

$b = 0.959795003$

$a + b\ (0.079136691) = -0.133093525$

$a = -0.209048526$

$\therefore$ $y = 0.959795003x - 0.209048526$, is the fitted linear equation

In original units, this becomes

$$Y = 0.959795003X + 16.246442199$$

Step 6. TESTING THE SIGNIFICANCE OF THE INCREASE IN THE SUMS OF SQUARES FOR REGRESSION DUE TO EACH ADDITIONAL DEGREE OF FITTING

S.S. due to linear regression $= (978.2590)\ (0.959795003)^2$
$= 901.1785$

S.S. due to quadratic regression $= 901.1785 + (6524.9329)\ (0.023479092)^2$
$= 901.1785 + 3.5970$
$= 904.7755$

S.S. due to cubic regression $= 904.7755 + (34603.5162)\ (0.0097446542)^2$
$= 904.7755 + 3.2859$
$= 908.0614$

Results

The following results are obtained as shown in the analysis of variance (Table 1.1) for testing the significance of the increase in the sums of squares for regression due to each additional degree of fitting.

Table 1.1 Testing the Significance of Each Additional Degree in Fitting of the Polynomial Regression of PScE English Marks on SSCE English Marks

Source	SS	DF	MS
Linear regression	901.18	1	901.18
Excess due to quadratic	3.60	1	3.60
Excess due to cubic	3.28	1	3.28
Residual	1260.02	274	4.26
Total	2168.08	277	

Evidently there is significant linear regression, but the fitting of the quadratic or that of cubic, does not make a significant improvement.

After the degree of the polymial was ascertained, the results were also obtained on ungrouped data for more exactness. The fitted linear and the quadratic equations are:

$$Z_1 = 0.993079075X_1 + 14.370632326$$

$$Z_1 = 0.005954238X_1^2 + 0.321149525X_1 + 32.811090668$$

Table 1.2 Testing the Significance of Linear and Quadratic Regressions of PScE English Marks on SSCE English Marks (Un-grouped)

ANALYSIS OF VARIANCE

Source	SS	DF	MS	F
Linear Regression	23606.71	1		
Quadratic Regression	23745.10	2		
Difference	138.39	1	138.39	1.22
Residual	31117.29	275	113.15	
Total	54862.39	277		

Next we consider the implications of a fundamental assumption on the results - the assumption of normality in the data. Normality in the distribution of marks is desirable for two reasons:

1. For application of the most of statistical methods, normality is a basic requirement, the fulfilment of which gives better results.

2. Secondly, the mental abilities, like other measurable characteristics on this universe, tend to be normally distributed.

Analysis on Normalized Standard Scores

When a fairly stable system of examinations is working, we do not expect major changes in the results except that the distribution of marks may be somewhat skewed differently from year to year. Under this situation, the method of changing raw scores to normalized standard scores by making use of percentile ranks seems to be more appropriate so that results will be more comparable from year to year. Hence the marks in SSCE English were converted to normalized standard scores as shown below:

Table 1.3 SSCE English Marks Converted to Normalized Standard Scores

Raw Marks in SSCE English Intervals	f	Percentile rank	Normalized standards scores
80 - 84	1	99.82	2.91
75 - 79	9	98.02	2.06
70 - 74	19	92.99	1.48
65 - 69	30	84.17	1.00
60 - 64	51	69.60	0.52
55 - 59	67	48.38	-0.04
50 - 54	56	26.26	-0.64
45 - 49	18	12.95	-1.13
40 - 44	13	7.37	-1.45
35 - 39	14	2.52	-1.96

The analysis on these normalized standard scores was done in the same manner as described above, and the results were obtained as shown in the following analysis of variance table:

Table 1.4 Testing the Significance of Each Additional Degree in Fitting of the Polynomial Regression of PScE English Marks on the Normalized SSCE English Marks

Source	SS	DF	MS
Due to linear	895.11	1	895.11
Excess due to quadratic	1.12	1	1.12
Excess due to cubic	4.45	1	4.45
Residual	1267.40	274	4.63
Total	2168.08	277	

Again we find linear regression is significant and addition of quadratic or cubic does not make any significant contribution.

Though in India, admissions are given on the basis of SSCE marks, in countries like U.S.A. many colleges and universities administer aptitude tests for admission purposes. The question then arises what function will be appropriate for the aptitude scores to predict college examination marks. The data on a standardized aptitude test in English was therefore further analysed and the results examined.

The test was constructed as a part of the composite scholastic aptitude test, standardized by Examination reform and research unit, M.S. University of Baroda. The test, consisting of 30 items, was administered to university students admitted to PSc course in June 64, and these students appeared for PSc examination in April 1965. The English test scores, and PScE English marks (1965) were analysed and the following results obtained.

Table 1.5 Testing the Significance of Each Additional Degree in Fitting of the Polynomial Regression of PScE English Marks on the Standardized English Test Scores

Source	SS	DF	MS
Linear regression	204.67	1	204.67
Excess due to quadratic	1.95	1	1.95
Excess due to cubic	0.13	1	0.13
Residual	652.97	348	1.88
Total	859.72	351	

It is evident that there is significant regression, but the fitting of quadratic or cubic does not come out significant.

The results obtained on ungrouped data are presented below:

The fitted linear and quadratic equations are:

$$P_E = 1.087589729T_e + 42.429741429$$

$$P_E = 0.029079465T_e^2 + 0.385880991T_e + 46.276484702$$

The significance of the above equations was tested by following analysis of variance(given in Table 1.6):

Table 1.6 Testing the Significance of Linear and Quadratic Regressions of PScE English Marks on Standardized English Test Scores (Un-grouped)

ANALYSIS OF VARIANCE

Source	SS	DF	MS	F
Linear Regression	5467.27	1	[illegible]	[illegible]
Quadratic Regression	5550.08	2		
Difference	82.81	1	82.81	1.79
Residual	16152.90		46.28	
Total	21702.98			

The graphs of regression lines for SSCE English marks and aptitude scores are plotted in figure 1.

Conclusions

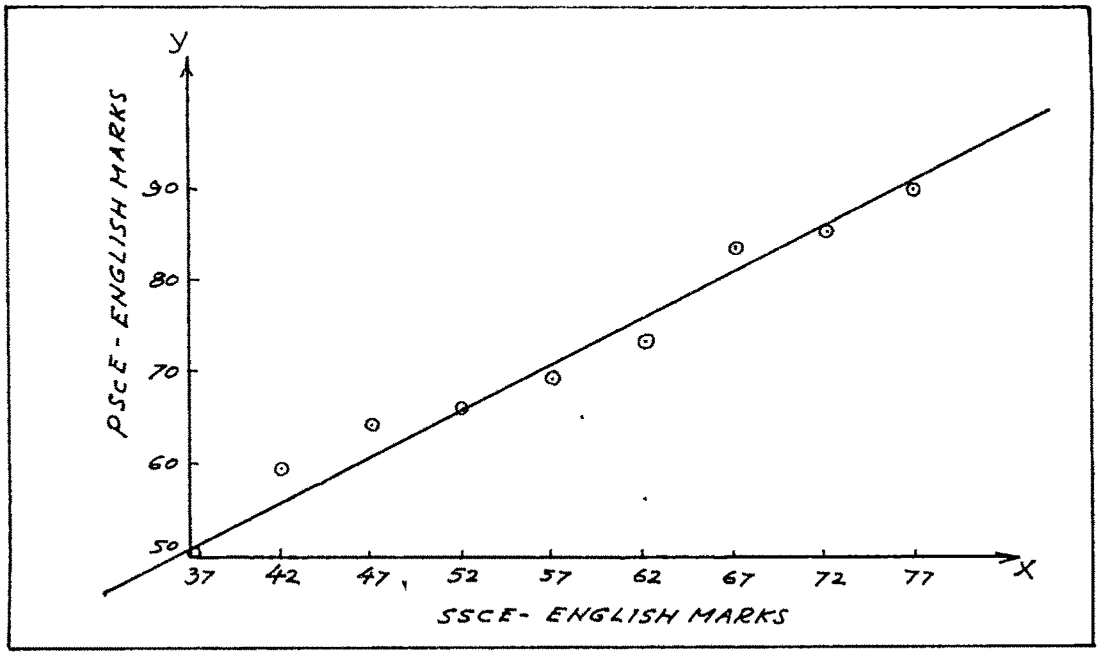
The findings of this study can be conclusively put as follows:

1. The empirical mathematical functions of educational measurements - achievement as well as aptitude measures for predicting college achievement in language, are found to hold linear relationship significantly.

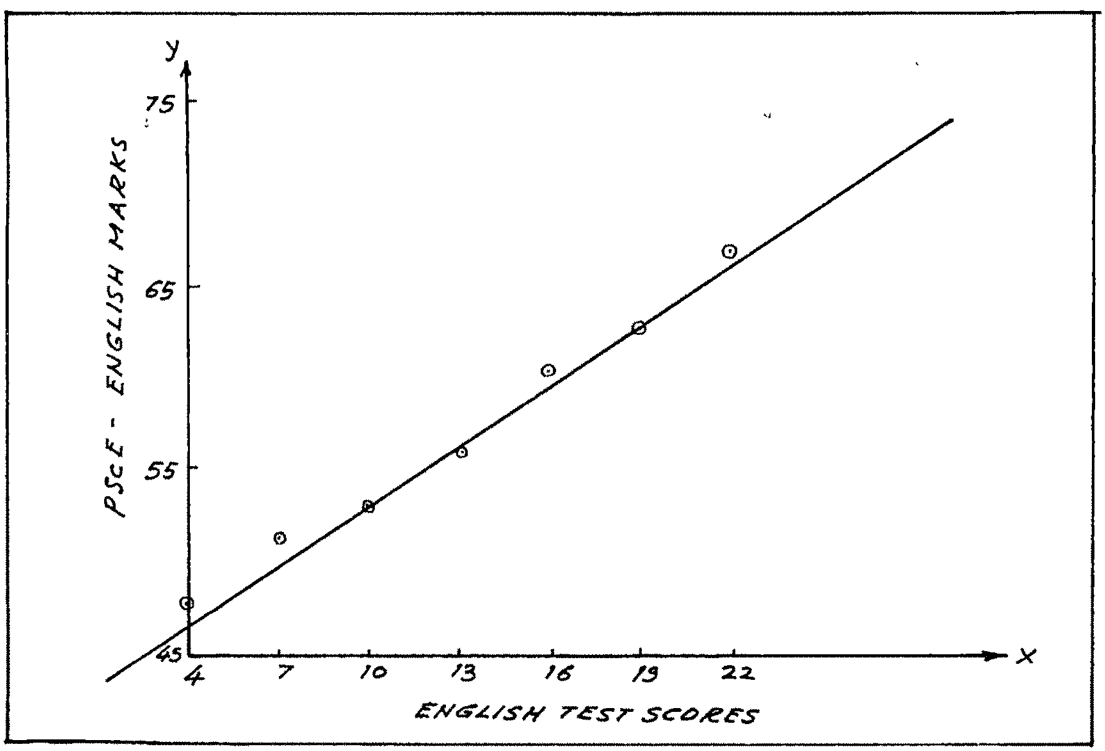
2. Addition of quadratic or cubic does not make any significant contribution.

3. Both the educational measures show parallel trends in regressions for forecasting college achievement (Fig. 1).

4. Converting SSCE English marks to normalized standard scores, does not affect the results.



(i)



(ii)

FIG. 1

GRAPHS SHOWING THE RELATIONSHIP OF THE CRITERION VARIABLE PScE ENGLISH WITH (i) SScE ENGLISH AND (ii) ENGLISH TEST SCORES