CHAPTER IV

PREDICTION OF OVERALL DEGREE OF COLLEGE SUCCESS

In the first three chapters, we have dealt with the prediction of each of the criterion variables language, mathematics and science separately. The main objective behind these studies was to get some idea to build up a kind of model for predicting the degree of college success. This is dealt with in this chapter.

In previous chapters, we observed that-

1. the variance accounted for by the quadratic term in mathematics variable is significant in many cases and that-

2. the variance accounted for by the quadratic term in English variable though statistically not significant, is observed fairly considerable occasionally.

Problem

We now have the problem: what mathematical function of these variables will be appropriate in predicting the overall degree of academic success as measured by PScE Grand Total percent?

Method

Consistent with the observation made, we start with the most general function of second degree in the two predictor variables-English, (say_x) and mathematics (say_y) for predicting the criterion (say_z) of overall degree of success measured by PScE Grand Total percent.

The most general equation of the second degree in x and y for predicting Z, is given as-

$$Z = ax^2 + 2hxy + by^2 + 2gx + 2fy + c$$
$$= Ax^2 + Hxy + By^2 + Gx + Fy + C \text{ (say)}$$

We therefore first of all, fit this equation to the data and test its significance. If found not significant, we go on fitting alternative models subsequently derived by deteøting the unnecessary term one by one, and test their significance.

The method used is multiple regression method which is applicable in case of quadratic and product term also.

Since PScE mathematics serves as more objective and hence more accurate criterion variable, we study the prediction of this criterion variable also; the whole study is made in four sections - I to IV.

I Prediction of overall degree of success (PScE Grand Total Percent: from (i) SSCE English marks and (ii) SSCE Mathematics marks).

Analysis:

Let Y_e denote PScE Grand Total Percent

x_1 denote SSCE English

x_2 denote SSCE Mathematics

1. The sums of squares and products are-

x	N	x_1	x_2	x_2^2	x_1^2	x_1x_2	y_e
1	278	15899	18758	1320592	933211	1083758	11514
x_1		933211	10833758	77087772	56072195	64277000	671646
x_2			1320592	96530570	64277000	77087772	797514
x_2^2				7293833128	4619816878	5693346842	57554392
x_1^2					3441409147	3902881040	40160830
x_1x_2						4619816878	47012551

The details of calculations are given at the end in (Appendix 3).

2. The fitted regression equations are-

$$y_e=0.000296801x_1^2+0.001015470x_1x_2+0.002156692x_2^2+0.304140489x_1 - 0.057402935x_2+12.696481468$$

$$y_e=0.000670528x_1^2+0.002384344x_2^2+0.331641948x_1-0.029321393x_2+ 10.851602865$$

$y_e=0.002426104x_2^2-0.033995319x_2+0.406727996x_1+8.925260124$

$y_e=0.415435175x_1+0.292338266x_2 - 2.067212463$

$y_e=0.549481648x_1 + 9.992054936$

3... The significance of each of the foregoing equations is tested by following analysis of variance (Table 4.1):

Table 4.1 Testing the Significance of Various Terms in SSCE-English(x_1) and SSCE Mathematics(x_2) for the Prediction of PScE Grand Toral Percent

ANALYSIS OF VARIANCE

Source	S.S.	D.F.	Means S.S.	F
Reg.on $x_1,x_2,x_2^2,x_1^2,x_1x_2$	11590.73	5		
Reg.on x_1,x_2,x_2^2,x_1^2	11587.15	4		
Difference	3.58	1	3.58	0.08
Residual	11670.87	272	42.91	
Reg.on x_1,x_2,x_2^2,x_1^2	11587.15	4		
Reg.on x_1,x_2,x_2^2	11585.47	3		
Difference	1.68	1	1.68	0.04
Residual	11674.45	273	42.76	
Reg.on x_1,x_2,x_2^2	11585.47	3		
Reg.on x_1,x_2	11488.95	2		
Difference	96.52	1	96.52	2.27
Residual	11676.13	274	42.61	
Reg.on x_1,x_2	11488.95	2		
Reg.on x_1	7227.27	1		
Difference	4261.68	1	4261.68	99.55**
Residual	11772.65	275	42.81	
Total	23261.60	277		

Only linear combination of x_1 and x_2 (i.e. Reg. on x_1 and x_2) is found to be significant in the above analysis. All other F-ratios are found to be not significant at 1% or 5% level. Hence in this case we are led to conclude that General College Scholarship as shown by overall degree of college success is more or less a linear function of SSCE achievement Test marks in English and Mathematics. The multiple correlation (R) of PScE Grand Total Percent on SSCE English and Mathematics marks is found to be .70.

II - Prediction of overall degree of success (PSc Grand Total Percent) from standardized (i) English test score and (ii) Numerical test score.

Analysis:

Let y_T denote PScE Grand Total Percent

T_e denote English Test Score

T_n denote Numerical Test Score

1. The sums of squares and products are:

	N	T_e	T_n	T_n^2	T_e^2	T_eT_n	y_T
1	352	4107	6296	122722	52541	76948	18246
T_e		52541	76948	1560336	724563	1023242	217622
T_n			122722	2564234	1023242	1560336	335947
T_n^2				56702782	21499022	33840454	6726085
T_e^2					10631837	14576902	2837860
T_eT_n						21499022	4196512

The details of calculations are given at the end (Appendix 4).

2. The fitted regression equations are:

$$y_T=0.063271124T_n^2-0.035156042T_eT_n-0.035813412T_e^2+1.900922869T_e - 1.082079092T_n+39.982330603$$

$$y_T=0.041743291T_n^2-0.038429395T_e^2+1.333862989T_e-0.718546148T_n+40.307073548$$

$$y_T=0.032277443T_n^2+0.406713656T_e-0.370227694T_n+42.458602786$$

$$y_T=0.416726136T_e+0.805045880 \quad +32.573684389$$

$$y_T=1.024363500T_e+39.883349735$$

3. The significance of each of the foregoing equations is tested by following: analysis of variance(Table 4.2):

Table 4.2 Testing the Significance of Various Terms in Standardized English Test Scores (T_e) and Numerical Test Scores (T_n) for the Prediction of PScE Grand Total Percent

ANALYSIS OF VARIANCE

Source	S.S.	D.F.	Mean S.S.	F
Reg.on $T_e,T_n,T_n^2,T_e^2,T_eT_n$	10294.21	5		
Reg.on T_e,T_n,T_n^2,T_e^2	10253.49	4		
Difference	40.72	1	40.72	0.58
Residual	24216 .23	346	69.99	
Reg.on T_e,T_n,T_n^2,T_e^2	10253.49	4		
Reg.on T_e,T_n,T_n^2	10147.89	3		
Difference	105.60	1	105.60	1.51
Residual	24256.95	347	69.90	
Reg.on T_e,T_n,T_n^2	10147.89	3		
Reg.on T_e,T_n	9695.41	2		*
Difference	452.48	1	452.48	6.46
Residual	24362.55	348	70.01	
Reg.on T_e,T_n	9695.41	2		
Reg.on T_e	4850.07	1		* *
Difference	4845.34	1	4845.34	68.15
Residual	24815.03	349	71.10	
Total	34510.44	351		

In the above analysis, we find that the regression in T_e, T_n, T_n^2 comes out significant over and above the linear combination i.e. regression on T_e and T_n. Hence we are led to conclude that in case of prediction of overall degree of success from the standardized aptitude tests, the quadratic term (T_n^2) improves the prediction significantly over and above the multiple linear regression. It may be noted that all F-ratios (except one corresponding to linear combination i.e. last one) in case of standardized tests are also elevated compared to those found in case of prediction from SSCE achievement tests. The multiple correlation (R) of PScE Grand Total Percent on standardized English and Numerical Test scores is found to be .53.

III. Prediction of PScE mathematics from (i) SSCE English marks and (ii) SSCE Mathematics marks-

Analysis:

Let y_d denote PScE Mathematics marks

and as before x_1 denote SSCE English

x_2 denote SSCE Mathematics

Here, also, we start our analysis with the most general equation of the second degree, i.e.

$$y_d = a_1x_1^2 + 2hx_1x_1 + bx_2^2 + 2gx_1 + 2fx_2 + c$$

1. The sums of squares and products are-

	N	x_1	x_2	x_2^2	x_1^2	x_1x_2	y_d
1	278	15899	18758	1320592	933211	1083758	12038
x_1		933211	1083758	77087772	56072195	64277000	707966
x_2			1320592	96530570	64277000	77087772	857995
x_2^2				7293833128	4619816878	5693346842	63533459
x_1^2					3441409147	3902881040	42652590
x_1x_2						4619816878	50974387

The details of calculations are given at the end(Appendix 5).

2. The fitted regression equations are:

$$y_d=0.007834671x_2^2+0.0010746800x_1x_2-0.005277173x_1^2+0.966217581x_1 - 0.371337028x_2 - 10.592569124$$

$$y_d=0.008075597x_2^2-0.004881655x_1^2+0.995322580x_1-0.341618126x_2 - 12.545017883$$

$$y_d=0.007771570x_2^2+0.448672671x_1-0.307590474x_2+1.479359305$$

$$y_d=0.476564491x_1+0.737758124x_2-33.732970310$$

$$y_d=0.814850258x_1 - 3.299655582$$

3. The significance of the each of the above regression equations is tested by following analysis of variance (Table 4.3):

Table 4.3 Testing the Significance of Various Terms in SSCE-English (x_1) and SSCE-Mathematics (x_2) in the Prediction of PScE Mathematics

ANALYSIS OF VARIANCE

Source	S.S.	D.F.	Means S.S.	F
Reg.on $x_1, x_2, x_2^2, x_1^2, x_1x_2$	44118.96	5		
Reg.on x_1, x_2, x_2^2, x_1^2	44114.95	4		
Difference	4.01	1	4.01	0.02
Residual	56197.66	272	206.61	
Reg.on x_1, x_2, x_2^2, x_1^2	44114.95	4		
Reg.on x_1, x_2, x_2^2	44025.79	3		
Difference	89.16	1	89.16	0.43
Residual	56201.67	273	205.87	
Reg.on x_1, x_2, x_2^2	44025.79	3		
Reg.on x_1, x_2	43035.36	2		*
Difference	990.43	1	990.43	4.82
Residual	56290.83	274	205.44	
Reg.on x_1, x_2	43035.36	2		* *
Reg.on x_1	15893.64	1		
Difference	27141.72	1	27141.72	130.30
Residual	57281.26	275	208.30	
Total	1000316.62	277		

In the above analysis, we find that quadratic term x_2^2 improves the prediction of PScE Mathematics over and above the linear combination of the two predictors i.e. Reg. on x_1, x_2, x_2^2 comes out significant over and above the significant Reg. on x_1, x_2.

IV. Prediction of PScE Mathematics from Standardized (i) English test score and (ii) Numerical test score-

We start our analysis with the most general equation of the second degree, i.e.

$$P_m = a_1 T_e^2 + 2hT_eT_n + b_1 T_n^2 + 2gt_e + 2fT_n + c$$

1. The sums of squares and products are:

	N	T_e	T_n	T_n^2	T_e^2	T_eT_n	P_m
1	352	4107	6296	122722	52541	76948	18888
T_e		52541	76948	1560336	724563	1023242	227242
T_n			122722	2564234	1023242	1560336	353445
T_n^2				56702782	21499022	3384454	7187815
T_e^2					10631837	14576902	2980010
T_eT_n						21499022	4453015

The details of calculations are given at the end (Appendix 6).

2. The fitted regression equations are-

$$P_m = 0.231753282T_n^2 - 0.190390137T_eT_n - 0.151970060T_e^2 - 4.783357641T_n + 7.488583174T_e + 35.346538295$$

$$P_m = 0.115167713T_n^2 - 0.166137107T_e^2 - 2.814618155T_n + 4.417628573T_e + 37.105207424$$

$$P_m = 0.074245169T_n^2 - 1.308775540T_n + 0.409397106T_e + 46.406651224$$

$$P_m = 0.43242800T_e + 1.394610368T_n + 23.669179953$$

$$P_m = 1.485060385T_e + 36.331980110$$

3. The significance of each of the above equations is tested by following analysis of variance (Table 4.4):

Table 4.4 Testing the Significance of Various Terms in Standardized English Test Scores (T_e) and Numerical Test Scores (T_n) for the Prediction of PScE Mathematics.

ANALYSIS OF VARIANCE

Source	S.S.	D.F.	Means S.S.	F
Reg.on $T_e, T_n, T_n^2, T_e^2, T_eT_n$	30296.56	5		
Reg.on T_e, T_n, T_n^2, T_e^2	29102.26	4		
Difference	1194.30	1	1194.30	*
Residual	99838.53	346	288.55	4.14
Reg.on T_e, T_n, T_n^2, T_e^2	29102.26	4		
Reg.on T_e, T_n, T_n^2	27128.50	3		
Difference	1973.76	1	1973.76	**
Residual	101032.83	347	291.16	6.78

(contd.)

Source	S.S.	D.F.	Means S.S.	F
Reg.on T_e, T_n, T_n^2	27128.50	3		
Reg.on T_e, T_n	24734.43	2		
Difference	2394.07	1	2394.07	8.09**
Residual	103006.59	348	295.99	
Reg.on T_e, T_n	24734.43	2		
Reg.on T_e	10193.62	1		
Difference	14540.81	1	14540.81	48.15**
Residual	105400.66	349	302.01	
Total	130135.09	351		

It is of interest to note in the foregoing analysis, that all the terms in the general equation of the second degree i.e. T_e, T_n, T_e^2, T_n^2, and T_eT_n come out significant.

RESULTS AND DISCUSSION

Comparing the analysis of variance, we find that, as the criterion y_e - PScE GT percent - is changed to y_d - PScE mathematics, the non-linear component emerge; or in case achievement tests are replaced by objective aptitude tests, the non-linearity is clearly visible, which can be well understood from the following table:

Table 4.5 Testing the Significance of Each Additional Terms in English and Mathematics Variables for Prediction of PScE GT Percent and PScE Mathematics

	Terms in achievement measures	PScGT Percent	PScE Maths.	Terms in Aptitude measures	PScGT Percent	PSc Maths.
		1	2		3	4
i	$x_1, x_2, x_2^2, x_1^2, x_1x_2$	NS	NS	$T_e, T_n, T_n^2, T_e^2, T_eT_n$	NS	S
ii	x_1, x_2, x_2^2, x_1^2	NS	NS	T_e, T_n, T_n^2, T_e^2	NS	S
iii	x_1, x_2, x_2^2	NS	S	T_e, T_n, T_n^2	S	S
iv	x_1, x_2	S	S	T_e, T_n	S	S
v	x_1	S	S	T_e	S	S

Comparing 1 with 3 in the above table, it is observed that the term x_2^2 in (iii) is not significant while T_n^2 is significant. This shows that the achievement measures in combination ultimately tend to be linearly related with the criterion of general college scholarship while aptitude measures, in combination also, tend to have curvilinear relationship. Also the strength of linear relationship of achievement measures is more than that of the aptitude measures.

Whether the differences mentioned above with regard to achievement versus aptitude tests are due to comprehensive nature of the achievement test or whether they are due to achievement measures containing linear component in excess which may be due to such additional abilities as memory, application to studies, cannot be said here. It is beyond the scope of this study and hence needs further investigation.

Conclusions:

From the above study, the only conclusions that can be drawn are:

1. The prediction function of comprehensive SSCE achievement measures-English and Mathematics in combination - is found to be linear for predicting the PScE Grand Total Percent and it is found to be curvilinear (with only quadratic term in SSCE-Mathematics significant) for the prediction of PScE Mathematics.

2. The prediction function of brief aptitude measures-English and Numerical test scores in combination - is found to be curvilinear (with only quadratic term in Numerical variable significant) for predicting the PScE Grand Total Percent, while the

prediction function given by the general equation of the second degree in English and Mathematics Test variables is found to be significant for the prediction of PScE Mathematics.

3. The curvilinear relationship given by the general equation of the second degree raises the level of accuracy of the prediction of PSc Mathematics (from English and Mathematics test measures) significantly well above the linear combination. This result bears much significance in validation research. Yet it is certain that the level of accuracy reached by SSCE achievement measures is far more than that of the aptitude test measures as indicated by multiple R's in the prediction of the PScE Grand Total Percent. In other words, the predictive value of aptitude test scores is limited for this criterion, compared to achievement measures. Hence the implications of the above results should be studied in combination with some auxilary variables, to have the same level of accuracy. For this reason, the analysis with aptitude measures is abandoned in Part II.