

CHAPTER 4

MULTIPLE REGRESSION ANALYSIS

In Chapter 3, the preliminary analysis presented the distribution of each variable and indicated that gender, age group, family status, residence, faculty, entrance mode, university entrance examination score, and school GPA are statistically significantly associated with achievement, and there may be an interaction between gender and age group.

In this chapter we develop a predictive model for achievement, using multiple regression analysis. The achievement scores from 627 students were investigated. The determinants are (1) gender, (2) age group, (3) religion, (4) family status, (5) family income, (6) father's education, (7) father's occupation, (8) mother's education, (9) mother's occupation, (10) method of entrance, (11) university entrance examination score, (12) type of school, (13) province of residence, (14) school GPA, (15) faculty, and (16) type of basic education. The main determinants are (10), (11), (12), and (14).

In Figure 9, the full model of multiple regression analyses with all the predictor variables included gives a goodness-of-fit, measured by the r-squared statistic, of 22.5%, and the residual standard deviation is 0.3937. The model indicates that the variables gender, family status, entrance, school GPA, and faculty are all statistically significant.

linear regression analysis: response = GPA

predictor	coeff	St.Error	p-value
constant	2.44	0.1219	0
gender	(0)		0.02369
male	0.08941	0.03942	0.02369
female			
age group	(0)		0.09405
16-17	-0.02578	0.05731	0.6532
18	-0.05209	0.05924	0.3876
19	-0.1816	0.08032	0.02411
20+			
religion	(0)		0.5363
not muslim	0.03716	0.06006	0.5363
muslim			
family status	(0)		0.01533
couple	0.1147	0.05563	0.03972
single parent	-0.1351	0.05985	0.02386
separated			
family income	(0)		0.4248
<5k b/mo	0.04584	0.03517	0.193
5k-15k b/mo	0.02334	0.08522	0.7843
>15k b/mo			
father educ	(0)		0.9402
no degree	0.003646	0.04859	0.9402
degree			
father occup	(0)		0.7954
govt	0.03505	0.05568	0.5293
priv emp	0.01284	0.04864	0.792
priv owner	-0.01486	0.08752	0.8653
others			
mother educ	(0)		0.5661
no degree	0.03044	0.05301	0.5661
degree			
mother occup	(0)		0.2863
govt	0.1444	0.08128	0.07621
priv emp	0.08238	0.07064	0.2453
priv owner	0.1111	0.07055	0.1159
others			
entrance mode	(0)		0.0000419
pooled	0.198	0.04796	0.0000419
direct			
entrance score	(0)		0.6643
<200 marks	0.01429	0.07138	0.8414
200-300 marks	-0.0282	0.09045	0.7553
>300 marks			
school	(0)		0.06936
formal	0.1236	0.06795	0.06936
nonformal			
residence	(0)		0.865
local	0.0004123	0.04114	0.992
near local	0.02384	0.04629	0.6068
others			
school GPA	(0)		7.939e-013
≤2	0.1334	0.0456	0.003569
2-3	0.5147	0.06934	3.997e-013
>3			
faculty	(0)		0.004969
education	-0.1177	0.04344	0.006939
humanities	-0.1802	0.06303	0.004392
sci&tech	-0.1111	0.08828	0.2088
islamic			
basic educ	(0)		0.2711
science	-0.04488	0.04558	0.3253
language	0.03194	0.04867	0.5119
lang+maths	-0.1318	0.128	0.3037
others			

r-sq: 0.2253 rss: 92.25 df: 595 sd: 0.3937 p-value: 0

Figure 9: Full model of multiple regression analysis

After using a stepwise procedure to eliminate redundant predictors, the r-squared is 19.8%. The standard deviation is reduced from 0.3937 to 0.3935. The result of fitting the model is shown in Figure 10.

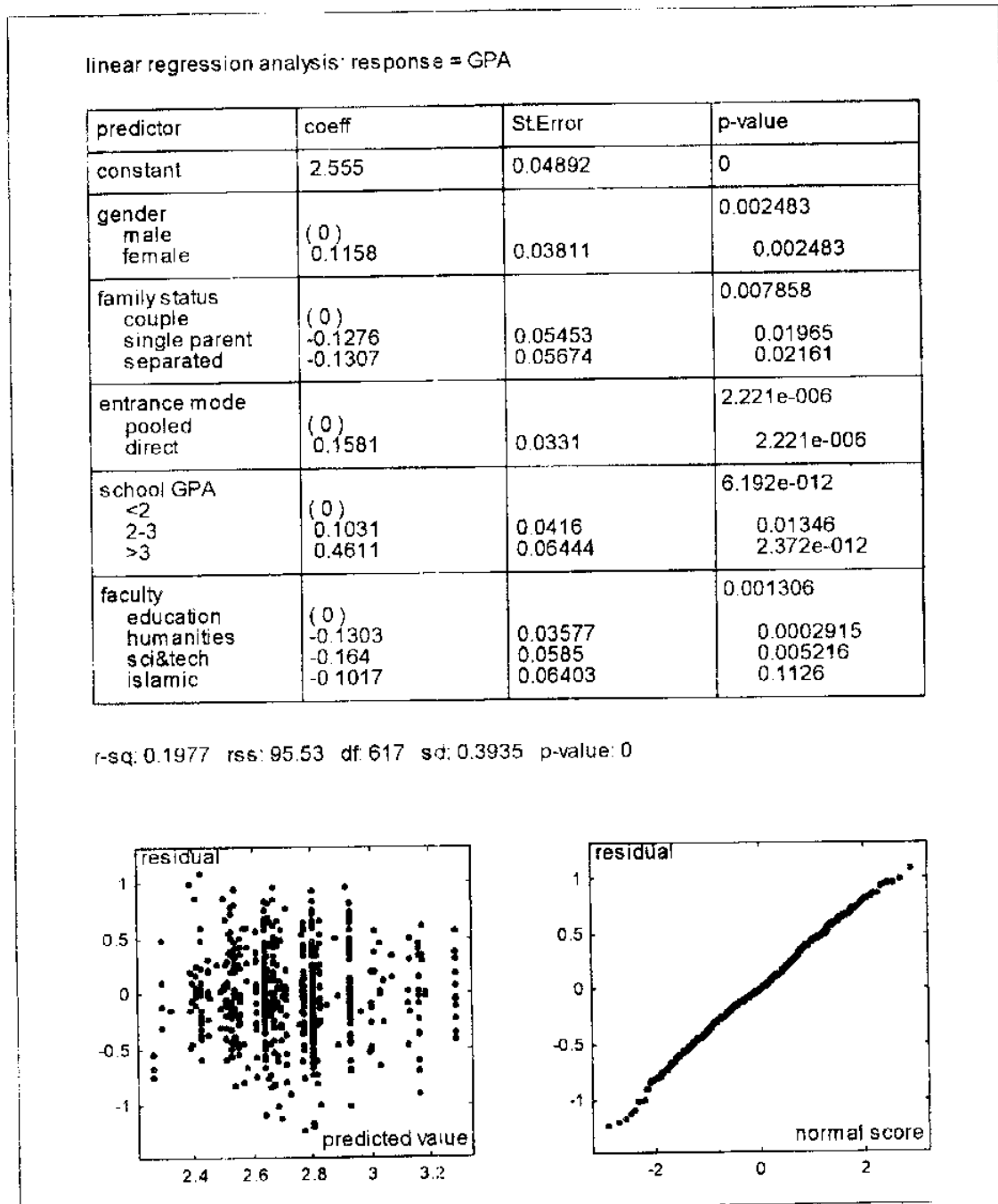


Figure 10: Multiple regression analysis of 627 Pattani undergraduates in academic year 1993

The plot of predicted values against residuals in the bottom left of Figure 10, shows that the relationship between the outcome and determinants is linear, and that the homogeneity assumption is reasonable. The normal scores plot (bottom right), suggests that the normality assumption is reasonable for these data. This model has statistically highly significant predictors. The regression coefficients of age group, family status, and faculty are negative. The coefficients for single parent and separated families are nearly the same. These two groups could be combined. So we repeat the analysis with this combined group, and the printout is shown in Figure 11.

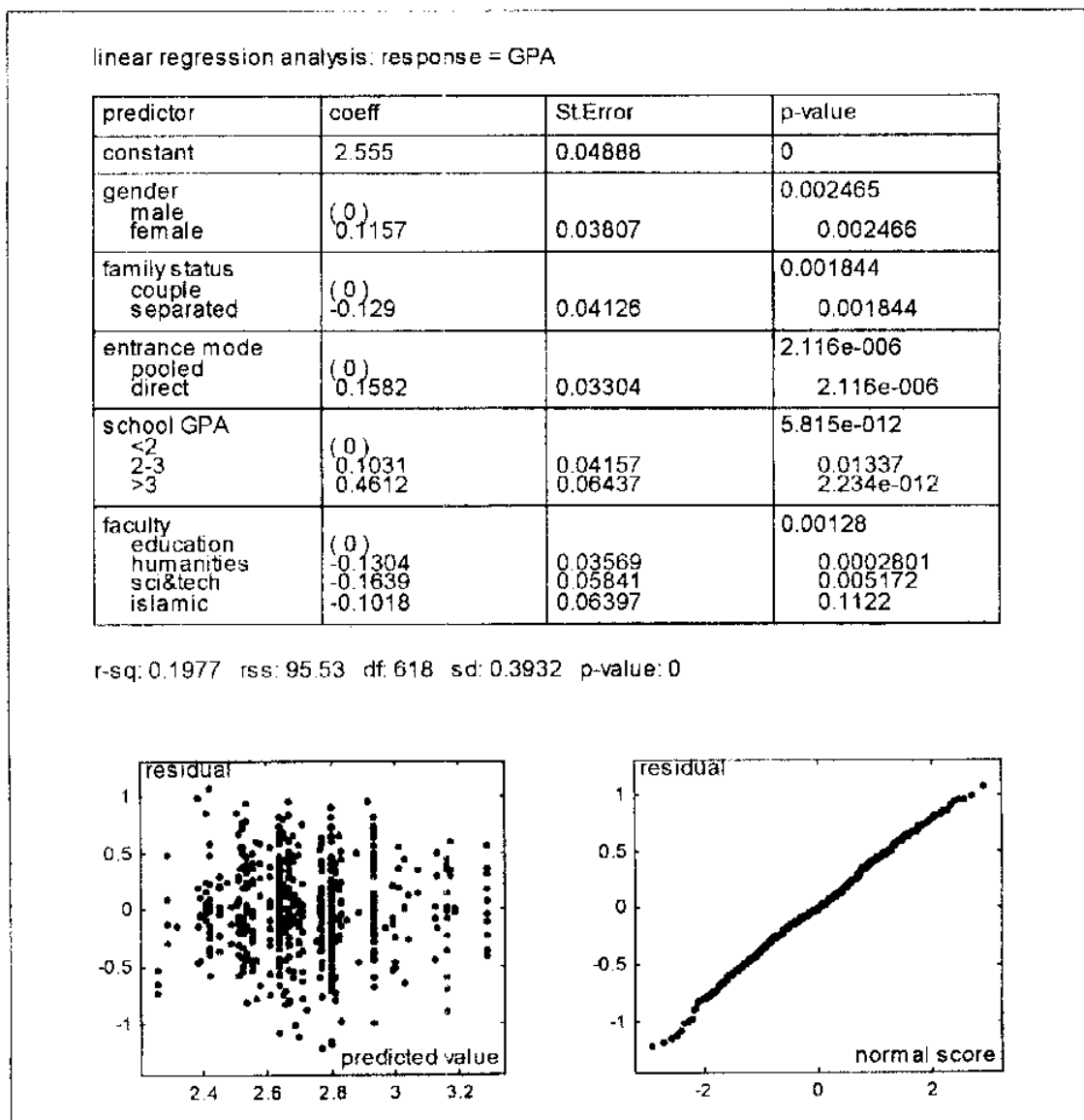


Figure 11: Model for achievement with family status recoded

Comparing the models in Figures 10 and 11, the r-squared does not change. The standard errors, coefficients, and p-values of the predictor variables change slightly. To check confounding, we need to compare the results before and after omitting the possible confounder. In this model gender is a predictor variable and a possible confounder. If it is omitted, the printout is as indicated in Figure 12.

linear regression analysis: response = GPA

predictor	coeff	St.Error	p-value
constant	2.62	0.04426	0
family status couple separated	(0) -0.1317	0.04152	0.001588 0.001588
entrance mode pooled direct	(0) 0.1628	0.03323	1.231e-006 1.231e-006
school GPA <2 2-3 >3	(0) 0.1302 0.484	0.04087 0.06436	1.198e-012 0.001519 1.936e-013
faculty education humanities sci&tech islamic	(0) -0.1244 -0.1878 -0.1387	0.03588 0.05826 0.06323	0.00074 0.0005631 0.001328 0.02864

r-sq: 0.1857 rss: 96.96 df: 619 sd: 0.3958 p-value: 0

Figure 12: Model with gender omitted

The regression coefficients all increase slightly when gender is omitted, but these increases are not substantial. They are mostly less than 10%, so gender is not a confounder.

In the preliminary results reported in Chapter 3, an interaction between gender and age group was found. The joint effect of two variables can be modeled in two ways, (a) as a main effect and interaction, (b) as all combinations of the two variables. Models (a) and (b) are shown in Figure 13 and 14, respectively.

linear regression analysis: response = GPA

predictor	coeff	St.Error	p-value
constant	2.42	0.1336	0
gender	(0)		0.01648
male	0.3386	0.1408	0.01648
female			
age group	(0)		0.0002241
16-17	0.2728	0.1396	0.05116
18	0.1144	0.1394	0.412
19	-0.1151	0.1506	0.4449
20+			
family status	(0)		0.001443
couple	-0.1312	0.04098	0.001443
separated			
entrance mode	(0)		4.814e-006
pooled	0.1533	0.03322	4.814e-006
direct			
school GPA	(0)		1.958e-012
<2	0.1049	0.04118	0.01108
2-3	0.468	0.06393	7.772e-013
>3			
faculty	(0)		0.0007265
education	-0.1378	0.03573	0.0001275
humanities	-0.1569	0.05808	0.007108
sci&tech	-0.04958	0.06493	0.4454
islamic			
gender.age group	(0)		0.001469
0	-0.3806	0.1511	0.01203
female*18	-0.2022	0.1511	0.1812
female*19	0.03114	0.1782	0.8613
female*20+			

r-sq: 0.2269 rss: 92.05 df: 612 sd: 0.3878 p-value: 0

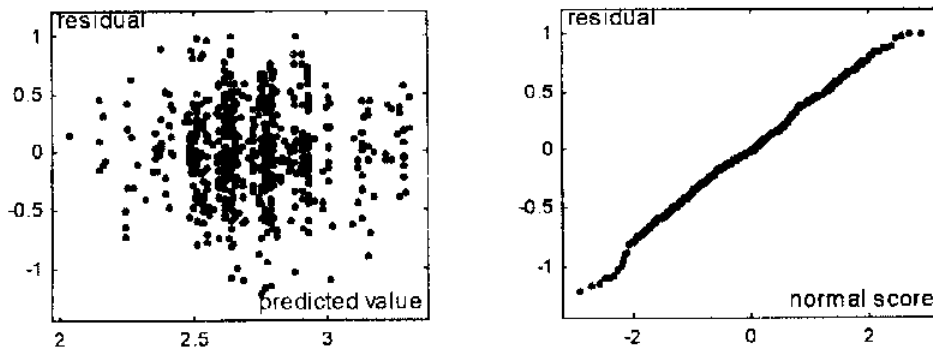


Figure 13: Model with main effect and interaction

The model shown in Figure 13 has a component for each combination of gender and age group. Gender, age group, and the interaction terms in Figure 12 are replaced by a set of eight interaction terms with girls aged 18 chosen as the reference group because this group had the highest number of students (see Table 5 of Chapter

3). The interaction terms in Figure 14 are labeled so that g18 refers to females aged 18 years, m18 refers to males aged 18 years, etc.

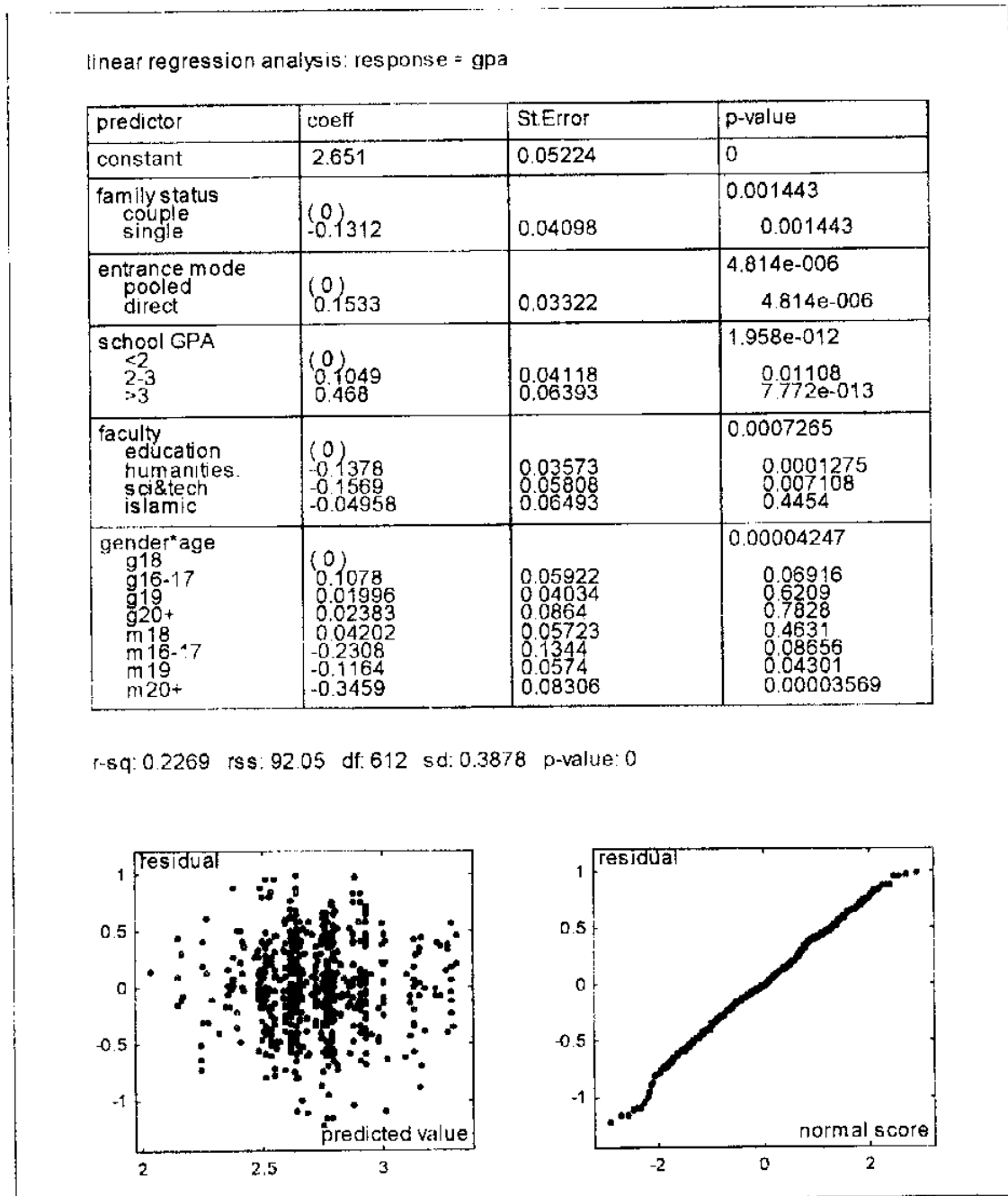


Figure 14: Model with all combinations of gender and age group

The models in Figure 13 and 14 have identical results. The r-squared, residual sum of squares, and residual degrees of freedom are the same. They are simply parameterised differently.

For males, the predicted takes the form:

$$\begin{aligned} \text{Achievement} = & 2.651 - 0.2308 \text{ age } 16-17 + 0.0420 \text{ age } 18 - 0.1164 \text{ age } 19 \\ & - 0.3459 \text{ age } 20^+ - 0.1312 \text{ single parent} + 0.1533 \text{ direct entrance} \\ & + 0.1049 \text{ school GPA } 2-3 + 0.468 \text{ school GPA } >3 \\ & - 0.1378 \text{ humanities} - 0.1569 \text{ science\&technology} - 0.0496 \text{ islamic} \end{aligned}$$

for females:

$$\begin{aligned} \text{Achievement} = & 2.651 + 0.1078 \text{ age } 16-17 + 0.02 \text{ age } 19 + 0.0238 \text{ age } 20^+ \\ & - 0.1312 \text{ single parent} + 0.1533 \text{ direct entrance} \\ & + 0.1049 \text{ school GPA } 2-3 + 0.468 \text{ school GPA } >3 \\ & - 0.1378 \text{ humanities} - 0.1569 \text{ science\&technology} - 0.0496 \text{ islamic} \end{aligned}$$

The interpretation is as follows.

- (a) The strongest predictor of university achievement is school GPA. Students with a school GPA of 3 or more achieve a score of 0.47 greater than those with school GPAs below 2, and 0.11 greater than those with school GPAs between 2 and 3.
- (b) Students entering by the direct method achieve a score 0.16 greater than other students.
- (c) The effects of age at entry and gender are not independent, and can be described in Table 7.

Table 7: The Coefficients of gender and age group

Age group	Gender	
	male	female
Age 16-17	-0.23	0.11
Age 18	0.04	0
Age 19	-0.12	0.02
Age 20+	-0.35	0.02

From this table, female students entering university at age 16 or 17 gain an achievement score of 0.11 greater than males entering at the same age. Male students entering university at age 20 or more do worse. However, there is no substantial difference between males and females entering university at age 18.

- (d) Students whose parents are single or separated achieve a score 0.13 less than other students.
- (e) Students studying in the faculties of Science and Technology and Humanities and Social Science achieve scores 0.16 and 0.13, respectively below students enrolled in the Faculty of Education. However, there is no substantial difference between the scores of students enrolled in the College of Islamic Studies and those in the faculty of Education.

Tables 8 and 9 show the predicted values for achievement separately for males and females in the Faculty of Science and Technology. The maximum and minimum mean achievement scores in this faculty are 2.99 and 2.13, respectively.

Table 8: Mean achievement score for males of the Faculty of Science and Technology

Gender		male							
School GPA	Age group	16-17 years		18 years		19 years		20+ years	
	Entrance	pooled	direct	pooled	direct	pooled	direct	pooled	direct
	Family status:								
	couple	2.26	2.42	2.26	2.42	2.26	2.42	2.26	2.42
<2	single parent	2.13	2.29	2.13	2.29	2.13	2.29	2.13	2.29
	couple	2.37	2.52	2.37	2.52	2.37	2.52	2.37	2.52
2-3	single parent	2.24	2.39	2.24	2.39	2.24	2.39	2.24	2.39
	couple	2.73	2.88	2.73	2.88	2.73	2.88	2.73	2.88
3+	single parent	2.60	2.76	2.60	2.76	2.60	2.76	2.60	2.76

Table 9: Mean achievement score for females of the Faculty of Science and Technology

Gender		female							
School GPA	Age group	16-17 years		18 years		19 years		20+ years	
	Entrance	pooled	direct	pooled	direct	pooled	direct	pooled	direct
	Family status:								
	couple	2.26	2.42	2.37	2.52	2.28	2.44	2.29	2.44
<2	single parent	2.13	2.29	2.24	2.40	2.15	2.31	2.16	2.31
	couple	2.37	2.52	2.48	2.63	2.39	2.54	2.39	2.55
2-3	single parent	2.24	2.39	2.35	2.50	2.26	2.41	2.26	2.42
	couple	2.73	2.88	2.84	2.99	2.75	2.90	2.75	2.91
3+	single parent	2.60	2.76	2.71	2.86	2.62	2.78	2.63	2.78

The predictive variables are gender, age group, family status, entrance mode, school GPA, and faculty. Figure 15 shows histograms and numerical summaries of predicted data. The total number of cases is 384. The outcome variable is mean achievement. The maximum mean and minimum means are 3.39 and 2.03, respectively.

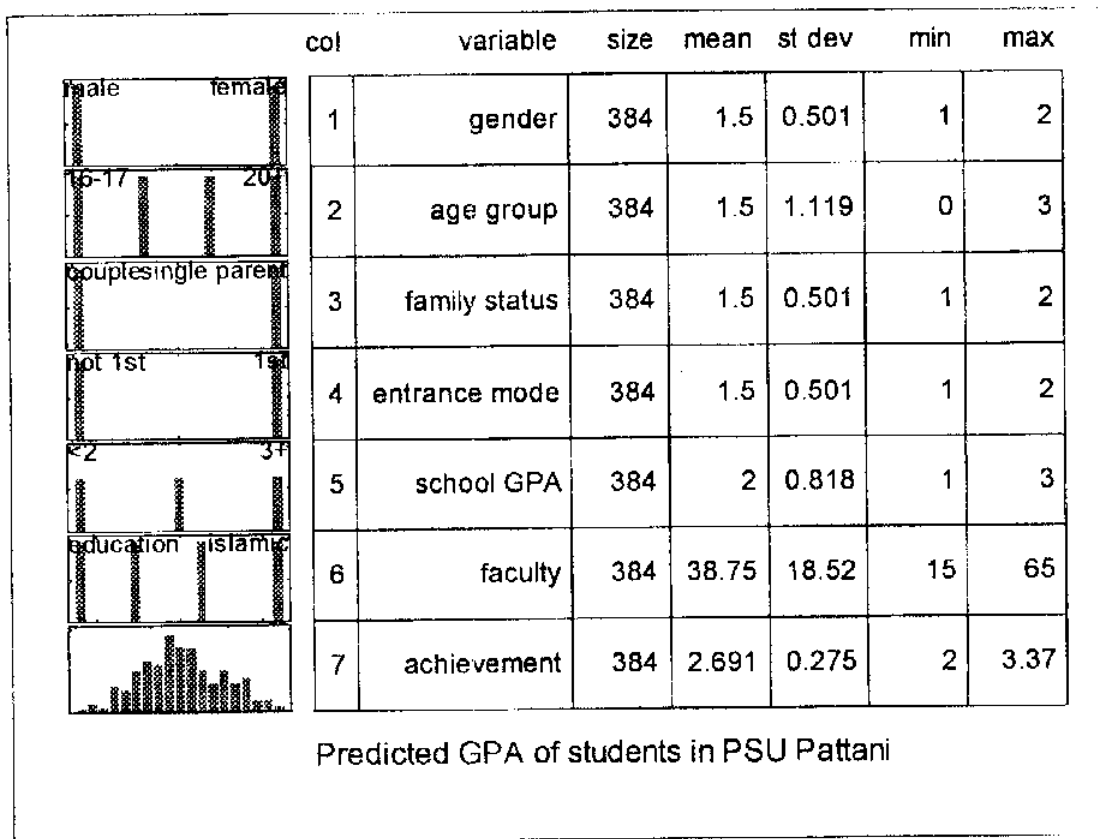


Figure 15: Histograms and numerical summaries of predicted data

Figure 16 shows the effects of age, gender and school GPA on achievement. The mean achievement scores tend to decrease with age. At 18 years of age, males have mean achievement scores higher than females while the rest females have mean achievement scores higher than males. Thus the gender effect also depends on age.

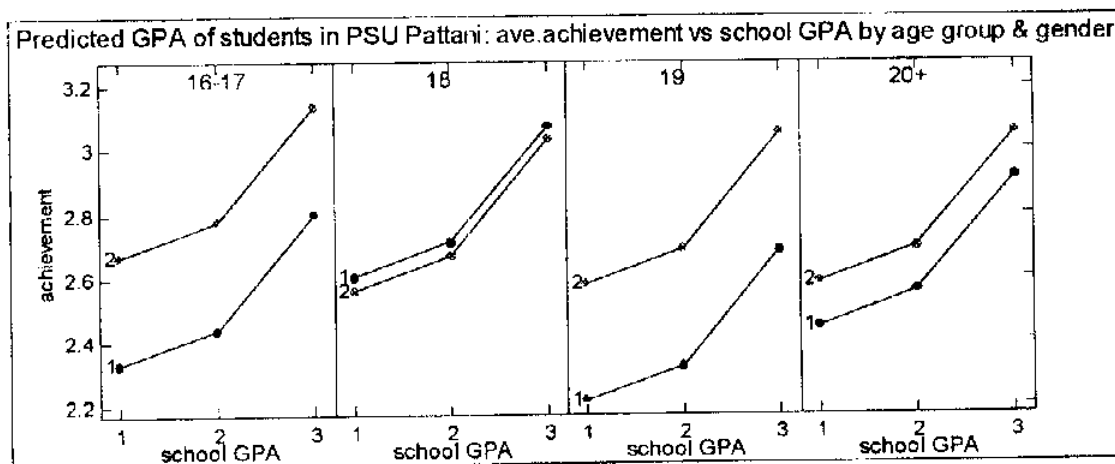


Figure 16: The predictive data adjusted by age group and gender

Figure 17 shows that gender difference is the same for each faculty. The faculty of Science and Technology has an average mean achievement lower than the other faculties.

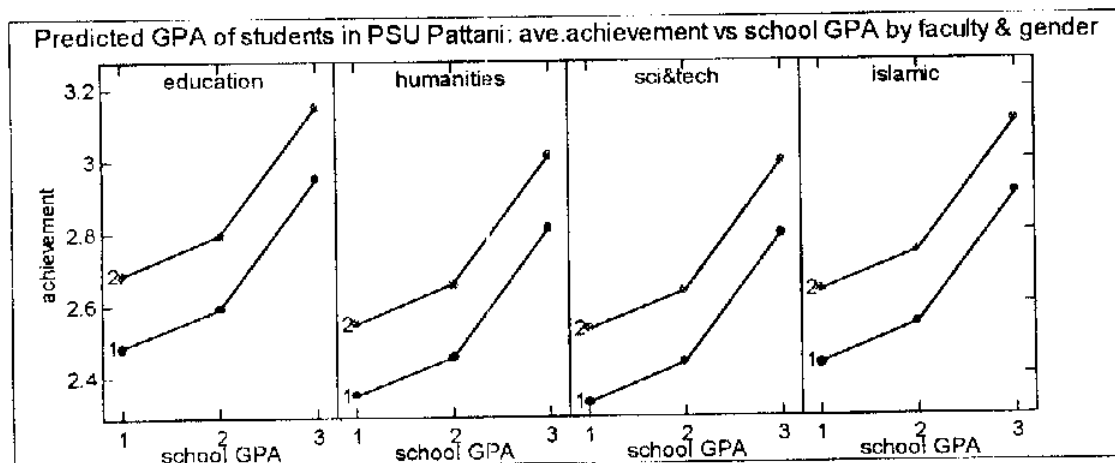


Figure 17: The predictive data adjusted by faculty and gender

The following figure shows a plot of actual score against predicted score for the 627 students in Prince of Songkla University Pattani Campus. The data from the predictive model are used.

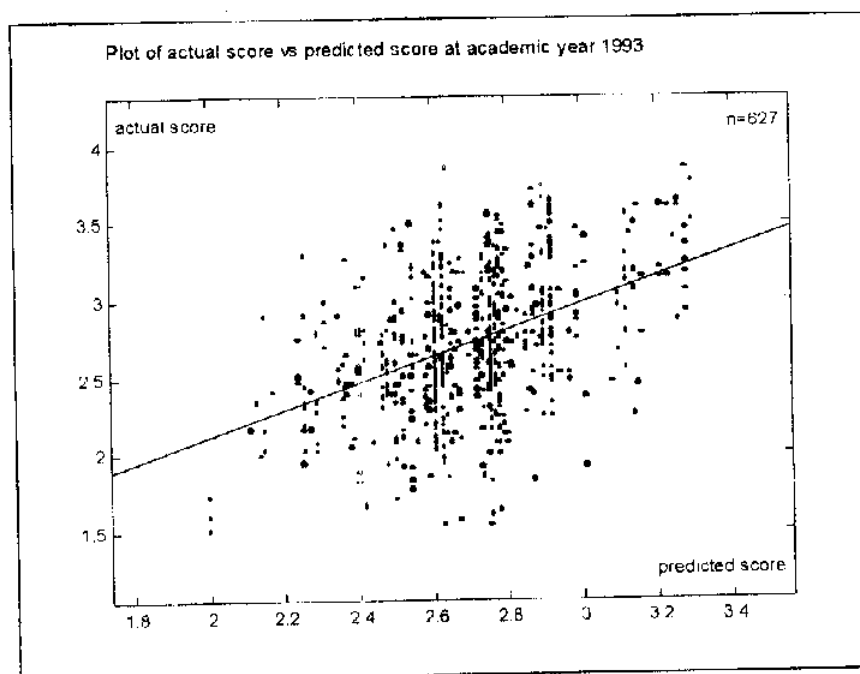


Figure 18: Plot of actual score against predicted score