

Chapter 4

Statistical Modelling

In chapter three we used some basic statistical methods for analysing the data. In this chapter we use linear regression to fit a model with predictors of each outcome.

4.1 Linear Regressions

Linear regression analysis is used to fit a model predicting the mean of each outcome, given the determinants. There are nine determinants in this study namely: sex, grade, program, GPA, school size, father's occupation, mother's occupation, family income and number of brothers and sisters. The referent categories for each determinant are as follows: male, grade 4, science-mathematics program, $GPA \geq 3$, large school, agriculture occupation of father, agriculture occupation of mother, less than 10,000 baht/month family income, and one sibling.

If it is reasonable to assume that the *collaborative*, *like to learn*, *independent*, *hate to learn* and *not creative* outcome factors have normal distributions, linear regression analysis is justified, and a backward elimination procedure is used to find the best regression model.

Table 4.1 gives numerical summaries of the five outcome factors. We see that the *not creative* factor has a greater mean than the other factors.

Variable	Size	Mean	StDev	StErr	95% C.I.	
Collaborative	196	3.473	0.515	0.037	3.400	3.546
Like to learn	196	3.565	0.498	0.036	3.495	3.635
Independence	196	2.645	0.601	0.043	2.559	2.738
Hate to learn	196	2.927	0.602	0.043	2.842	3.011
Not creative	196	3.787	0.534	0.038	3.712	3.863

Table 4.1 Description of five factors

Figures 4.1, 4.2 and 4.3 show the normal scores plot for each factor. We see that the normal scores plot for the *collaborative* and *like to learn* factors show a straight line trend, suggesting that the normality assumption is reasonable for these data.

The normal scores plots for *independent*, *hate to learn* and *not creative* show a rough linear trend, suggesting that the normality assumption also might be reasonable for these data.

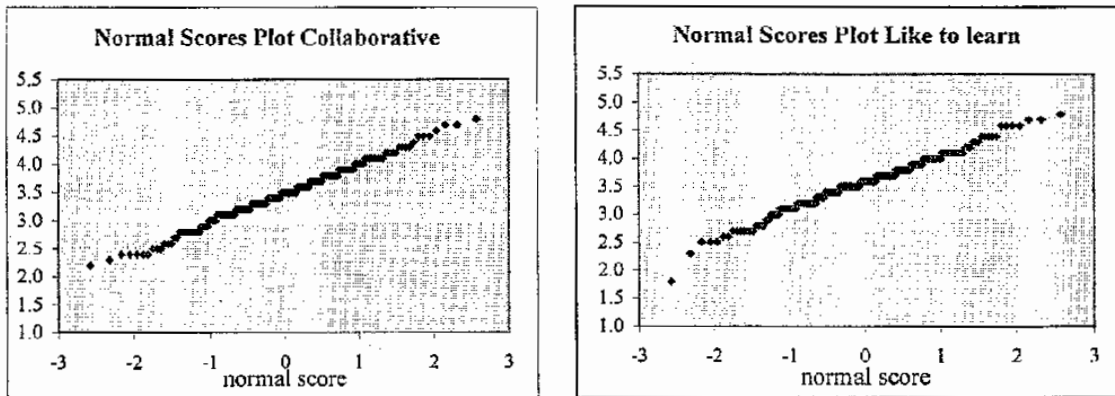


Figure 4.1: The normal scores plots for collaborative, like to learn

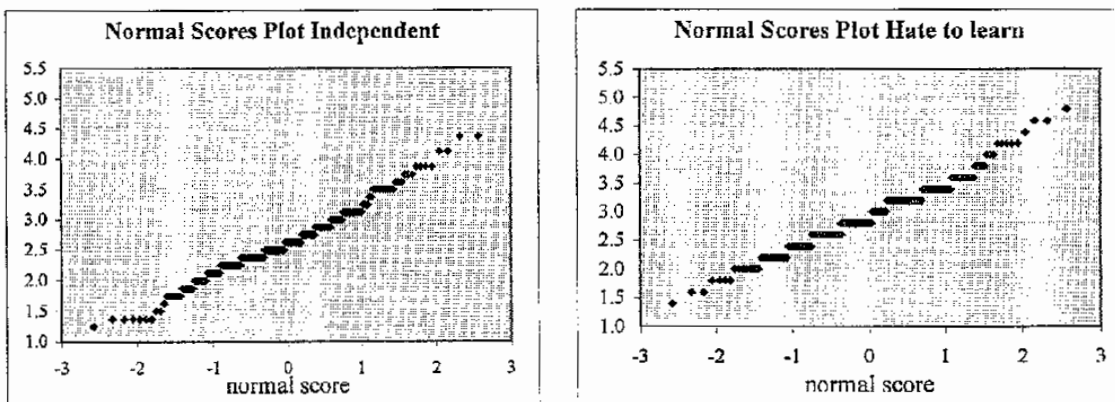


Figure 4.2: The normal scores plots for independence, hate to learn

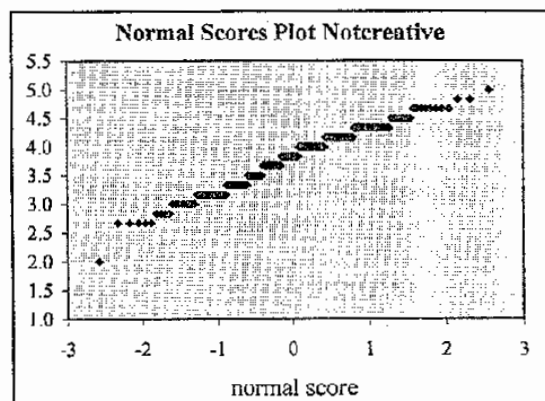


Figure 4.3: The normal scores plot for not creative

Figure 4.4 shows the result of fitting a multiple linear regression model, containing all nine determinants, to the factor "collaborative". There is not determinant statistically significant in this model.

linear regression analysis: response = Collaborative			
predictor	coeff	St.Error	p-value
constant	3.3577	0.22714	0
sex	(0)		0.62834
male	-0.044511	0.091788	0.62834
female			
grade	(0)		0.53972
grade 4	0.033422	0.091611	0.7157
grade 5	-0.071558	0.098349	0.46786
grade 6			
program	(0)		0.087371
sci/math	-0.073872	0.10363	0.47692
eng/math	0.13262	0.1329	0.31972
eng/fren	-0.24398	0.10849	0.02579
general			
GPA	(0)		0.82087
3+	0.038205	0.10814	0.7243
2.00-2.99	-0.01708	0.12939	0.89513
2.00-			
SchoolSize	(0)		0.19974
large	0.14519	0.11626	0.21345
med-large	0.053998	0.11617	0.64267
average	0.36175	0.18544	0.052727
small			
FOccup	(0)		0.31552
agric	-0.12331	0.14837	0.40708
labour	-0.17195	0.17683	0.33222
comm	-0.21921	0.23535	0.35294
SOEs	-0.31581	0.14912	0.035639
offic			
MOccup	(0)		0.15262
agric	0.13943	0.15927	0.38256
labour	0.1943	0.16938	0.25295
comm	-0.2835	0.41838	0.49695
SOEs	0.33384	0.1526	0.03005
offic			
FamInc	(0)		0.20325
10k-	0.14229	0.094705	0.13483
10-29k	0.12684	0.12002	0.29207
30-49k	0.29808	0.15137	0.050548
50k+			
NSibs	(0)		0.36551
0	0.070212	0.12836	0.58508
1	-0.049613	0.12721	0.89702
2+			

r-sq: 0.15116(0.032026) rrs: 43.8567 df: 171 sd: 0.50643 p-value: 0.19211

Figure 4.4: Regression model for "collaborative" factor containing all predictors

Figure 4.5 shows the final reduced model, using backward elimination. The final model from *collaborative* outcome has only one variable that is significant. We can see in this model, the students in the english-french program have a greater *collaborative* score than the students in the science-mathematics program and students in the english-mathematics and general programs have lower *collaborative* score than students in the science/mathematics program. The r-squared is 0.034 which means about 3 % of the *collaborative* variation is explained by the model.

The model takes the form

$$\text{Collaborative} = 3.554 - 0.089 \text{ engl/math} + 0.033 \text{ engl/fren} - 0.267 \text{ general}$$

linear regression analysis: response = Collaborative			
predictor	coeff	St.Error	p-value
constant	3.5542	0.059627	0
program			0.022174
sci/math	(0)		
engl/math	-0.088949	0.095501	0.35282
engl/fren	0.0325	0.10995	0.76786
general	-0.26667	0.094279	0.0051735
r-sq: 0.048705(0.033841) rss: 49.1503 df: 192 sd: 0.50596 p-value: 0.022174			

Figure 4.5: Final reduced model for "collaborative" factor

Table 4.6 shows the result of fitting a multiple linear regression model, containing all nine determinants to the factor "like to learn". There is only one variable significant in this model.

After backward elimination, the reduced model has only one variable that is significant. It is found that the coefficients of Mattayomsuksa 5 and Mattayomsuksa 6 are negative. Students in Mattayomsuksa 4 have greater *like to learn* learning style than Mattayomsuksa 5 and Mattayomsuksa 6, as shown in the Figure 4.7. The r-squared is 0.024, which means that about 2% of the variation in *like to learn* learning style is explained by the model.

The model takes the form

$$\text{Like to learn} = 3.687 - 0.159 \text{ grade5} - 0.220 \text{ grade6}$$

linear regression analysis: response = like to learn

predictor	coeff	St.Error	p-value
constant	3.6998	0.21589	0
sex	(0)		0.22152
male	0.10705	0.087243	0.22152
female			
grade	(0)		0.057524
grade 4	-0.17196	0.087075	0.049892
grade 5	-0.20414	0.093479	0.030338
grade 6			
program	(0)		0.79396
sci/math	0.040385	0.0985	0.68232
engl/math	0.072245	0.12632	0.56812
engl/fren	-0.052083	0.10311	0.61414
general			
GPA	(0)		0.094212
3+	-0.2236	0.10278	0.030968
2.00-2.99	-0.16344	0.12298	0.18562
2.00-			
SchoolSize	(0)		0.16981
large	0.19155	0.11051	0.08483
med-large	0.098888	0.11042	0.37272
average	0.34102	0.17626	0.054572
small			
FOccup	(0)		0.30422
agric	-0.033675	0.14102	0.81155
labour	-0.32115	0.16808	0.057712
comm	-0.2714	0.2237	0.22671
SOEs	-0.22108	0.14174	0.12067
offic			
MOccup	(0)		0.040614
agric	0.19741	0.15138	0.19396
labour	0.22759	0.16099	0.15928
comm	-0.41363	0.39768	0.29975
SOEs	0.37778	0.14504	0.010009
offic			
FamInc	(0)		0.34397
10k-	0.037041	0.090016	0.68123
10-29k	0.00083546	0.11407	0.99416
30-49k	0.25116	0.14388	0.082659
50k+			
NSibs	(0)		0.13475
0	-0.093221	0.122	0.44586
1	-0.21259	0.12091	0.080498
2+			

r-sq: 0.17976(0.064636) rrs: 39.621 df: 171 sd: 0.48135 p-value: 0.054858

Figure 4.6: Regression model for "like to learn" factor containing all predictors

linear regression analysis: response = like to learn			
predictor	coeff	St.Error	p-value
constant	3.6868	0.059619	0
grade	(0)		0.03454
grade 4	-0.1586	0.083419	0.058766
grade 5	-0.2201	0.088288	0.01351
grade 6			
r-sq: 0.034276(0.024268) rss: 46.6484 df: 193 sd: 0.49163 p-value: 0.03454			

Table 4.7: Final reduced model for "like to learn" factor

In the full model of the multiple regression analysis with all predictor variables for the *independent* factor, no variable is statistically significant.

Next, for selecting the best regression equation, backward elimination was used. The final model for the *independent* outcome consists of one variable that is significant. The coefficient of 0.450 suggests that students with GPA < 2 have greater *independent* learning style scores than students with higher GPAs, as shown in Figure 4.8. The r-squared is 0.076, which means that about 8% of the variation in *independent* learning style is explained by the model.

The model takes the form

$$\text{Independence} = 2.330 + 0.380 (\text{GPA } 2.00\text{-}2.99) + 0.499(\text{GPA} < 2)$$

linear regression analysis: response = Independence			
predictor	coeff	St.Error	p-value
constant	2.3304	0.086121	0
GPA	(0)		0.00018564
3+	0.38036	0.10608	0.00042608
2.00-2.99	0.44971	0.11239	8.9727e-005
2.00-			
r-sq: 0.085185(0.075705) rss: 64.4151 df: 193 sd: 0.57772 p-value: 0.00018564			

Figure 4.8: Final reduced model for "independent" factor

The model containing all predictors is shown in Figure 4.9.

linear regression analysis: response = Independence

predictor	coeff	St.Error	p-value
constant	2.5655	0.25578	0
sex	(0)		0.56713
male	-0.059266	0.10336	0.56713
female			
grade	(0)		0.21367
grade 4	-0.062106	0.10316	0.54785
grade 5	-0.28532	0.11075	0.084076
grade 6	-0.19244		
program	(0)		0.1428
sci/math	0.11381	0.1167	0.3308
eng/math	-0.28532	0.14965	0.058253
eng/fren	0.078369	0.12216	0.52206
general			
GPA	(0)		0.12592
3+	0.24461	0.12177	0.04613
2.00-2.99	0.24423	0.1457	0.095515
2.00-			
SchoolSize	(0)		0.044999
large	-0.064179	0.13092	0.62462
med-large	-0.099705	0.13082	0.44702
average	-0.57666	0.20883	0.0063839
small			
FOccup	(0)		0.60649
agric	-0.038601	0.16708	0.81756
labour	-0.2182	0.19913	0.27472
comm	-0.12385	0.26502	0.64087
SOEs	-0.24438	0.16792	0.14742
offic			
MOccup	(0)		0.65526
agric	-0.058666	0.17935	0.74398
labour	0.1836	0.19074	0.33711
comm	0.37936	0.47114	0.42183
SOEs	0.11228	0.17184	0.51438
offic			
FamInc	(0)		0.29849
10k-	-0.079198	0.10665	0.45872
10-29k	-0.13133	0.13515	0.33255
30-49k	-0.18978	0.17048	0.26712
50k+			
NSibs	(0)		0.22776
0	0.1327	0.14454	0.35986
1	0.23345	0.14325	0.10501
2+			

r-sq: 0.2102(0.099346) rss: 55.6127 df: 171 sd: 0.57028 p-value: 0.010307

Figure 4.9: Regression model for "independent" factor containing all predictors

In Figure 4.10, the full model of multiple regression analysis with all the predictor variables included for the *hate to learn* factor gives a goodness-of-fit measured by the r-squared statistic of 5.9%, and the standard deviation is 0.583. The model indicates that the variables program and GPA are statistically significant.

linear regression analysis: response = Hate to learn

predictor	coeff	St.Error	p-value
constant	2.9183	0.26159	0
sex	(.0)		0.43345
male			
female	0.082999	0.10571	0.43345
grade	(.0)		0.70343
grade 4			
grade 5	0.081489	0.10551	0.44097
grade 6	0.076069	0.11327	0.50275
program	(.0)		0.0017435
sci/math			
eng/math	-0.39763	0.11935	0.0010581
eng/fren	-0.30811	0.15306	0.045675
general	-0.36313	0.12494	0.004141
GPA	(.0)		0.01476
3+			
2.00-2.99	0.30422	0.12454	0.01559
2.00-	0.42769	0.14901	0.004621
SchoolSize	(.0)		0.35767
large			
med-large	-0.11583	0.1339	0.38823
average	-0.18022	0.1338	0.17978
small	-0.35645	0.21357	0.096949
FOccup	(.0)		0.083034
agric			
labour	0.4126	0.17087	0.016804
comm	0.21715	0.20365	0.2878
SOEs	0.095095	0.27105	0.72614
offic	0.027606	0.17174	0.87249
MOccup	(.0)		0.1947
agric			
labour	-0.35193	0.18342	0.056695
comm	-0.29592	0.19507	0.13113
SOEs	0.13262	0.48186	0.78347
offic	-0.37353	0.17575	0.03499
FamInc	(.0)		0.21406
10k-			
10-29k	0.049336	0.10907	0.6516
30-49k	0.28571	0.13822	0.040238
50k+	0.082376	0.17433	0.63715
NSibs	(.0)		0.89559
0			
1	-0.00019509	0.14783	0.99895
2+	-0.04301	0.14651	0.78844

r-sq: 0.17538(0.059643) rss: 58.1704 df: 171 sd: 0.58325 p-value: 0.067869

Figure 4.10: Regression model for "hate to learn" factor containing all predictors

After using a backward procedure to eliminate redundant predictors, the r-squared is 5%, the standard deviation is 0.5859. The regression coefficients for program are negative, which means that the students in the science-mathematics program have greater *hate to learn* learning style than the others. The coefficients for GPA value is positive, which means the students having GPAs of 3 or more have less hatred of learning than other students, as shown in Figure 4.11. The r-squared is 0.051, which

means that about 5% of the variation in *hate to learn* learning style is explained by the model.

The model takes the form

$$\text{Hate to learn} = 2.914 - 0.344 \text{ engl/math} - 0.160 \text{ engl/fren} - 0.338 \text{ general} + 0.214 (\text{GPA } 2.00 - 2.99) + 0.323 (\text{GPA } <2)$$

linear regression analysis: response = Hate to learn			
predictor	coeff	St>Error	p-value
constant	2.9138	0.10493	0
program	(0)		0.0066622
sci/math	-0.34412	0.11388	0.0028592
engl/math	-0.15976	0.13295	0.231
engl/fren	-0.33804	0.11365	0.0033174
GPA	(0)		0.049216
3+	0.21483	0.11536	0.064112
2.00-2.99	0.32264	0.13111	0.014758
2.00-			
r-sq: 0.075419(0.051088) rss: 65.2219 df: 190 sd: 0.58589 p-value: 0.010298			

Figure 4.11: Final reduced model for "hate to learn" factor

Figure 4.12 shows the full model of the multiple regression analysis with all predictor variables for the *not creative* factor. There is only one statistically significant variable. It is school size.

After using a backward procedure to eliminate predictors, the model indicates that four variables are in this model and two variables are statistically significant as shown in Figure 4.13. After omitting MOccup and program from the model, the resulting model has no variable that is statistically significant shown in Figure 4.14. To check confounding we need to compare the result before and after omitting the possible confounded.

The conclusion is that further study is needed.

linear regression analysis: response = Not creative

predictor	coeff	St.Error	p-value
constant	3.7365	0.23503	0
sex	(0)		0.38681
male	0.082404	0.094976	0.38681
female			
grade	(0)		0.36875
grade 4	-0.13096	0.094793	0.16893
grade 5			
grade 6	-0.042679	0.10177	0.67546
program	(0)		0.28368
sci/math	0.081179	0.10723	0.45006
engl/math	0.25085	0.13751	0.069873
engl/fren			
general	-0.01132	0.11225	0.91979
GPA	(0)		0.56745
3+	-0.03225	0.11189	0.77353
2.00-2.99			
2.00-	-0.12604	0.13388	0.34781
SchoolSize	(0)		0.01044
large	0.36496	0.1203	0.0027926
med-large	0.15324	0.12021	0.20412
average			
small	0.44774	0.19189	0.020791
FOccup	(0)		0.91695
agric	0.01271	0.15352	0.93411
labour	-0.029194	0.18297	0.87342
comm	-0.1391	0.24352	0.56861
SOEs			
offic	0.058342	0.1543	0.70582
MOccup	(0)		0.28751
agric	-0.0030631	0.1648	0.98519
labour	-0.055741	0.17526	0.75084
comm	-0.52029	0.43292	0.2311
SOEs			
offic	0.15581	0.1579	0.32516
Faminc	(0)		0.028063
10k-	-0.082989	0.097994	0.39825
10-29k			
30-49k	-0.26643	0.12418	0.033329
50k+	0.21442	0.15663	0.17281
NSibs	(0)		0.36442
0	-0.14749	0.13281	0.26834
1			
2+	-0.04117	0.13163	0.75484

r-sq: 0.15633(0.037926) rrs: 46.956 df: 171 sd: 0.52402 p-value: 0.15695

Figure 4.12: Regression model for "not creative" factor containing all predictors

linear regression analysis: response = Not creative

predictor	coeff	St.Error	p-value
constant	3.6127	0.13022	0
program	(0)		0.089032
sci/math	0.015672	0.10062	0.87639
engl/math	0.30965	0.12948	0.017797
engl/fren	-0.05728	0.10555	0.588
SchoolSize	(0)		0.0074267
large	0.36231	0.11525	0.0019495
med-large	0.17687	0.11169	0.11504
average			
small	0.45492	0.18132	0.012981
MOccup	(0)		0.13985
agric	0.0053743	0.11472	0.96269
labour	-0.056905	0.11956	0.63467
comm	-0.56003	0.39083	0.1536
SOEs	0.17016	0.10918	0.12083
FamInc	(0)		0.033163
10k-	-0.097809	0.091522	0.28662
10-29k	-0.2378	0.11818	0.045678
30-49k			
50k+	0.20325	0.14979	0.1765

r-sq: 0.12138(0.058621) rss: 48.9015 df: 182 sd: 0.51835 p-value: 0.028902

Figure 4.13: Final reduced model for "not creative" factor

linear regression analysis: response = Not creative

predictor	coeff	St.Error	p-value
constant	3.778	0.087388	0
SchoolSize	(0)		0.1069
large	0.19563	0.10005	0.05203
med-large	0.024899	0.094964	0.79346
average			
small	0.27121	0.16011	0.091931
Fam Inc	(0)		0.10214
10k-	-0.095449	0.089141	0.28564
10-29k	-0.21133	0.11	0.05621
30-49k			
50k+	0.12306	0.14651	0.402

r-sq: 0.064225(0.034518) rss: 52.0826 df: 189 sd: 0.52495 p-value: 0.048455

Figure 4.14: Final reduced model for "not creative" factor after omitting MOccup and program from the model