

CHAPTER 4

LOGISTIC REGRESSION ANALYSIS

In this chapter, the main results of the study are presented, by modelling the incidence of HIV/AIDS cases separately for males and females with the different years and different provinces combined using logistic regression. The logistic model is used because of its ability to handle any number covariates simultaneously. The results for females based on the model with no interactions are given in Figure 13

In the first step, we fit a model with all explanatory variables, consisting of four years, five provinces and 15 age groups. The bottom line in Figure 13 contains the number of degrees of freedom, which is equal to the number of cells minus the number of predictors, assuming a constant is included among the predictors in the model.

The table shows how the odds ratios for each year in the four years in this study are compared (female cases before the year 1993 were insufficient to analyse). The years and provinces are thus compared after adjusting for age group. Although age is not a confounder, it is still a strong risk factor, and thus should be included in the model.

The result also includes estimates for the odds ratios of interest and their associated confidence intervals. The combination of determinants with 0 coefficients is referred to as the referent category. Thus each comparison is with year 1993 and Songkhla province in age group 0-4 years. The result shows, as expected, that the age groups 20-24, 25-29, 30-34, 35-39 years each have a higher risks of disease than the other age groups.

The deviance from this model is not high, the p-value indicating a good fit. The standardised residuals are plotted against normal scores, with the unit slope line superimposed. This plot would be linear if the normality assumption were met. However, many of the counts are small, particularly in the early years and the younger

age group. Some of these residuals appear to be unusually high, and will be investigated further.

logistic regression analysis: female hiv/aids incidence 1993-96

factor	coeff	St.Error	p-value	Odds ratio	95% CI	
count 170	-9.631	0.2224	0	0.0001	0	0.0001
province	(0)		0			
Songkhla	-0.2941	0.1951	0.1316	0.7452	0.5084	1.092
Battani	-0.8666	0.1743	0	0.4204	0.2987	0.5915
Yala	-0.94	0.2078	0	0.3906	0.2599	0.587
Narathiwat	-1.291	0.198	0	0.2749	0.1865	0.4053
year	(0)		0			
0	0.3234	0.1337	0.0156	1.382	1.063	1.796
1	-0.1249	0.1495	0.4035	0.8826	0.6584	1.183
2	-1.307	0.2243	0	0.2706	0.1743	0.42
3						
age group	(0)		0			
0-4	-1.545	0.4899	0.0016	0.2132	0.0816	0.557
5-9	-2.031	0.611	0.0009	0.1311	0.0396	0.4344
10-14	-1.172	0.4276	0.0067	0.3098	0.134	0.7162
15-19	-1.262	0.2295	0	3.534	2.253	5.541
20-24	1.646	0.2226	0	5.184	3.351	8.02
25-29	-1.124	0.2406	0	3.218	1.821	4.933
30-34	0.8088	0.2619	0.002	3.218	1.821	4.933
35-39	-0.5033	0.4063	0.2148	0.6042	0.2725	1.34
40-44	-0.7233	0.4899	0.1398	0.4851	0.1857	1.267
45-49	-0.3983	0.4546	0.3803	0.671	0.2753	1.636
50-54	-1.0859	1.02	0.0409	0.1243	0.0168	0.9174
55-59	-3.482	327.1	0.3672	0	0	3.781e+272
60-64	-0.835	0.7349	0.2559	0.4339	0.1028	1.832
65-69	-13.44	475.9	0.9775	0	0	inf
70-74	-13.47	412.7	0.9739	0	0	inf
75+						

df: 297 deviance: 209.5 p = 1 number of iterations: 14

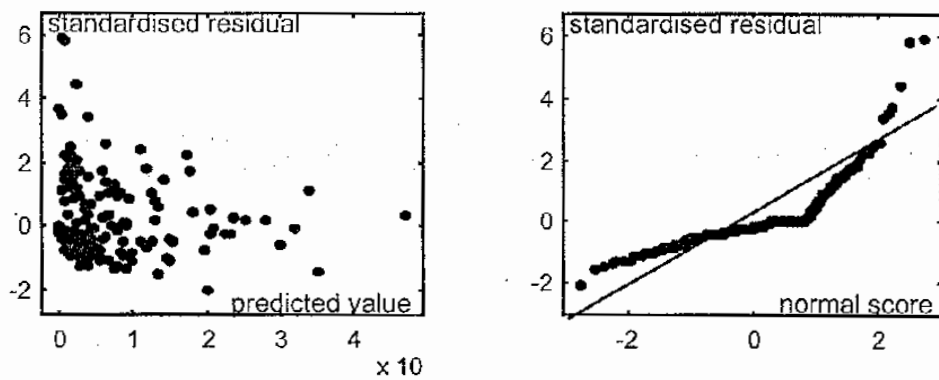


Figure 13: Logistic regression model for females, with no province-year interaction

Figure 14 shows the corresponding results with the interactions between year and province included in the model. These interactions are coded so that, for example, "42" means province 4 (Pattani) and year 1994.

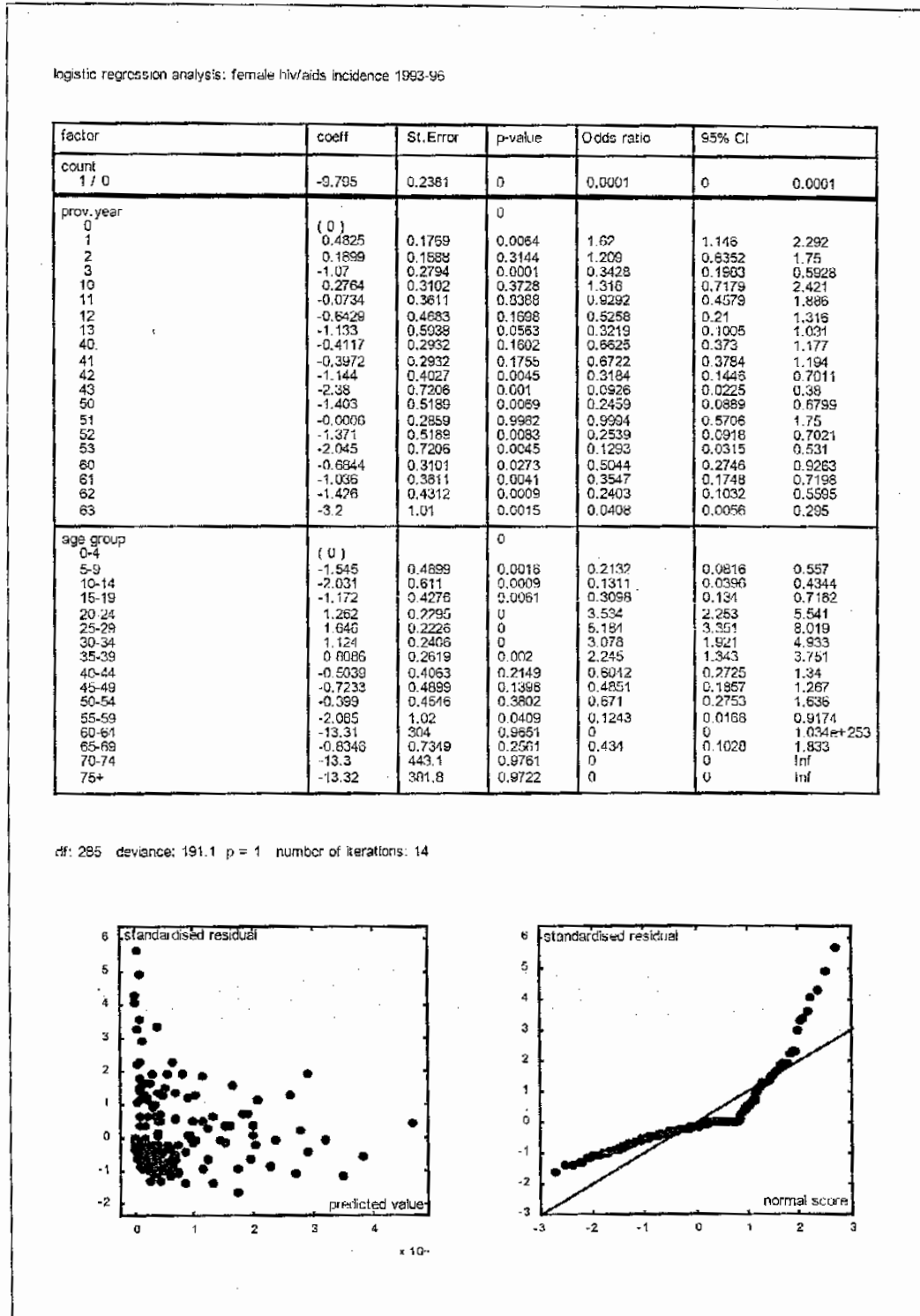


Figure 14: Logistic regression model for females, with province-year interaction

Figure 14: when fitting the model containing the year-province interactions, the deviance has decreased from 209.5 to 191.1, a decrease of 18.4, after including 12 additional parameters. This decrease is not quite statistically significant ($p=0.104$). The residuals in this more complex model are similar to those in the model with no interactions. The cells corresponding to the largest seven standardised residuals are shown in Table 8.

Table 8: The largest standardised residuals corresponding for females

age group	province	year	residual	number of cases
0-4	Songkhla	1993	5.63	4
40-44	Satun	1993	4.91	1
10-14	Pattani	1993	4.29	1
5-9	Narathiwat	1994	4.07	1
45-49	Yala	1996	3.60	1
45-49	Yala	1995	3.34	2
10-14	Yala	1996	3.27	1

From this table, it can be seen that the large residuals do not correspond to unusual observations. With the possible exception of the data for age group 0-4 in Songkhla in 1993, they indicate poor approximations to normality in the normal scores plot rather than any deficiency in the model.

Finally, the coefficients for the province-year interaction effects are plotted in Figure 15 these may be compared with the odds ratios plotted in Chapter 3 (see Figure 12)

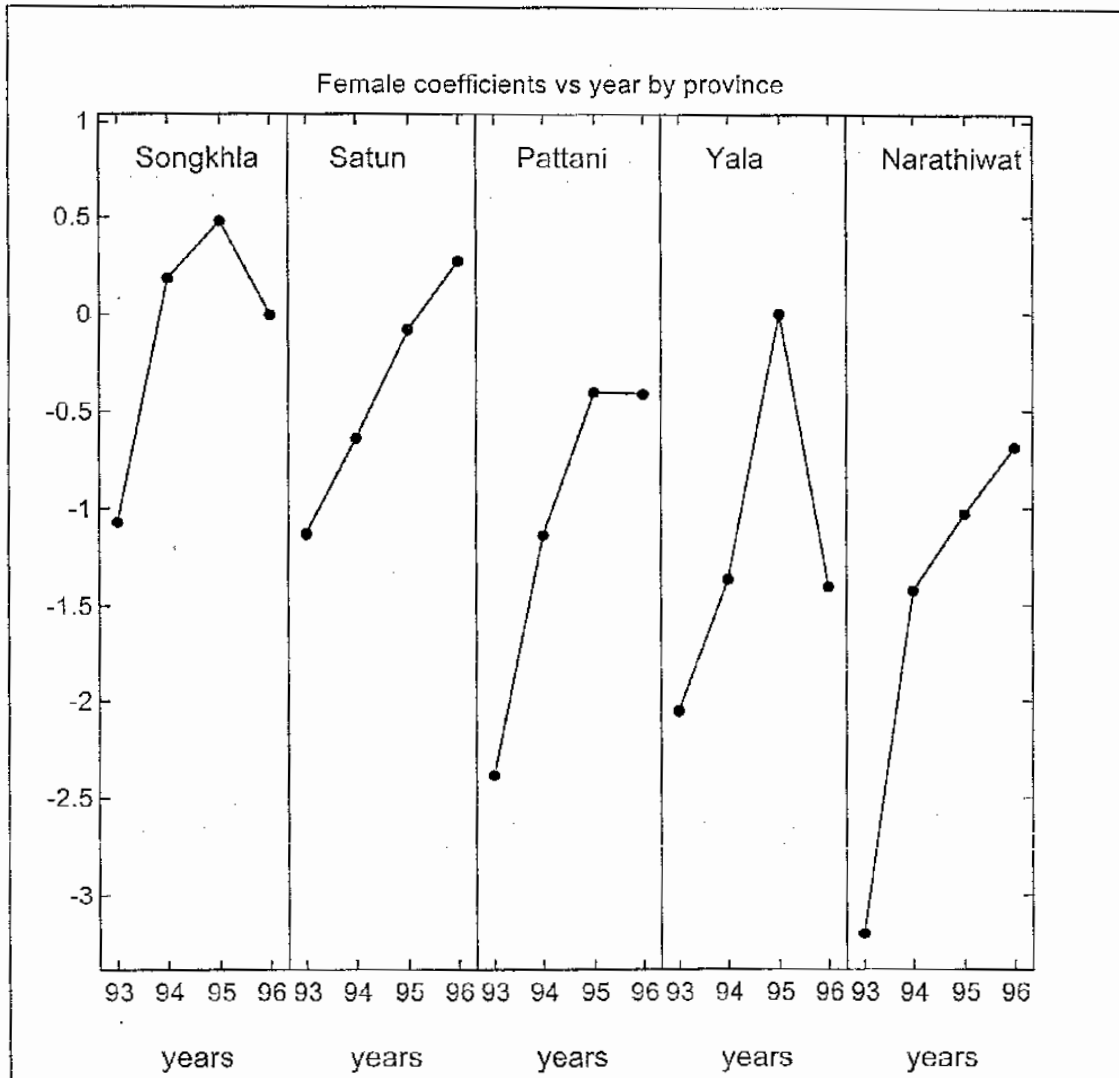


Figure 15: Coefficient plots for female incidence vs year by province

This graph shows clearly that the incidence of HIV/AIDS among females decreased in Songkhla after reaching a peak in 1995, with a similar pattern for Yala, where the risk was generally lower during the same period. In Pattani the risk was straightly lower still, but did not decrease in 1996. However, in Satun and Narathiwat, the incidence rates were still increasing in 1996, with Satun showing the highest rate and Narathiwat somewhat lower.

Figure 16 shows the result of fitting the no-interaction logistic regression model for the male cases. This shows how the odds ratios of each year in six years are compared. The result also includes estimates for odds ratio of interest and their associated confidence intervals, as for females, each comparison is with year 1996 and Songkhla province. The result suggests that age groups 30-34, 25-29 have the highest incidence, and may be compared with the Mantel-Haenszel estimates given in Chapter 3. The deviance given in the bottom line is not high, the p-value indicating a good fit. The standardised residuals are plotted against normal scores with the unit slope line superimposed.

logistic regression analysis: male hiv/aids incidence 1991-96

factor	coeff	St.Error	p-value	Odds ratio	95% CI	
count 1 / 0	-9.646	0.1914	0	0.0001	0	0.0001
province	(0)		0			
Songkhla	-0.1771	0.0902	0.0496	0.8377	0.7019	0.9997
Satun	-0.24	0.0654	0.0002	0.7866	0.692	0.8941
Pattani	-0.7508	0.0902	0	0.472	0.3955	0.5633
Yala	-1.556	0.1073	0	0.2109	0.1709	0.2603
Narathiwat						
year	(0)		0			
1996	0.236	0.0627	0.0002	1.266	1.12	1.432
1995	-0.1575	0.0695	0.0235	0.8543	0.7454	0.979
1994	-1.126	0.0966	0	0.3243	0.2684	0.3919
1993	-2.1	0.1476	0	0.1225	0.0917	0.1636
1992	-3.31	0.2624	0	0.0365	0.0218	0.0611
1991						
age group	(0)		0			
0-4	-0.852	0.3432	0.0131	0.4265	0.2177	0.8359
5-9	-3.281	1.017	0.0013	0.0376	0.0051	0.2759
10-14	-0.191	0.2907	0.511	0.8261	0.4673	1.46
15-19	1.999	0.2	0	7.384	4.989	10.93
20-24	3	0.1916	0	20.08	13.79	29.23
25-29	3.047	0.1921	0	21.05	14.44	30.67
30-34	2.695	0.1958	0	14.81	10.09	21.74
35-39	2.08	0.2075	0	8.001	5.327	12.01
40-44	1.621	0.2275	0	5.058	3.238	7.899
45-49	0.9478	0.2701	0.0004	2.58	1.52	4.381
50-54	0.4037	0.3338	0.2265	1.497	0.7784	2.88
55-59	0.8498	0.3055	0.0054	2.339	1.285	4.257
60-64	-0.0196	0.4843	0.9677	0.9806	0.3795	2.533
65-69	-0.5246	0.7311	0.473	0.5918	0.1412	2.48
70-74	0.0821	0.5334	0.8776	1.086	0.3816	3.088
75+						

df: 455 deviance: 368.3 p = 0.9989 number of iterations: 8

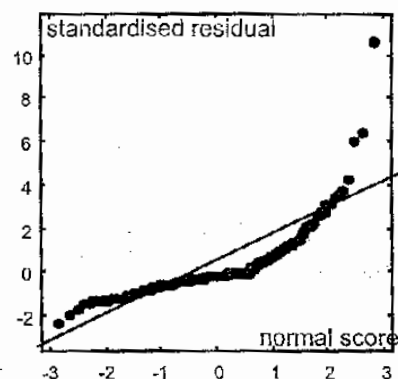
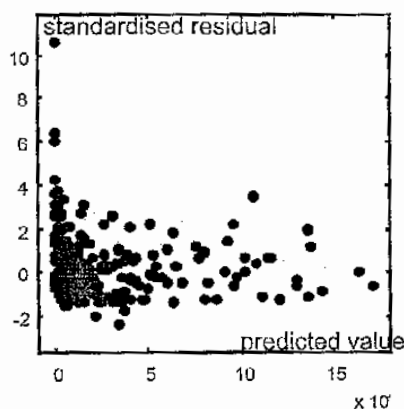


Figure 16: Logistic regression model for males, with no province-year interaction

Figure 17, shows the more extensive model involving individual parameters for each combination of province and year included in the model.

logistic regression analysis: male hiv/aids incidence 1991-96

factor	coeff	St.Error	p-value	Odds ratio	95% CI	
count 1 / 0	-9.723	0.1952	0	0.0001	0	0.0001
prov .year	(0)		0			
0	0.2743	0.0857	0.0014	1.316	1.112	1.556
1	-0.0066	0.092	0.9428	0.9934	0.8295	1.19
2	-0.8425	0.1198	0	0.4306	0.3405	0.5446
3	-2.124	0.2064	0	0.1195	0.0797	0.1791
4	-3.299	0.3594	0	0.0369	0.0182	0.0746
5	0.0769	0.1581	0.6268	1.08	0.7921	1.472
10	0.205	0.1506	0.1735	1.228	0.9138	1.649
11	-0.2961	0.1882	0.1156	0.7437	0.5143	1.076
12	-2.351	0.5041	0	0.0952	0.0355	0.2558
13	-2.613	0.5809	0	0.0733	0.0235	0.229
14	-3.683	1.002	0.0002	0.0251	0.0035	0.1792
15	0.0043	0.1165	0.9703	1.004	0.7993	1.262
40	0.1729	0.1108	0.1187	1.189	0.9567	1.477
41	-0.6129	0.1494	0	0.5418	0.4042	0.7261
42	-1.766	0.2509	0	0.1711	0.1046	0.2798
43	-2.178	0.3083	0	0.1133	0.0619	0.2073
44	-3.452	0.5809	0	0.0317	0.0101	0.0989
45	-0.7374	0.1745	0	0.4784	0.3398	0.6735
50	-0.5928	0.1655	0.0003	0.5528	0.3996	0.7647
51	-0.6441	0.1707	0.0002	0.5251	0.3758	0.7339
52	-1.745	0.2847	0	0.1746	0.0999	0.3051
53	-2.34	0.3834	0	0.0964	0.0455	0.2043
54	-4.259	1.002	0	0.0141	0.002	0.1008
55	-1.57	0.2101	0	0.2081	0.1379	0.3142
60	-1.13	0.1745	0	0.3229	0.2293	0.4546
61	-1.803	0.2383	0	0.1648	0.1033	0.2629
62	-2.645	0.3594	0	0.071	0.0351	0.1437
63	-3.314	0.5041	0	0.0364	0.0135	0.0977
64	-3.98	0.71	0	0.0187	0.0046	0.0751
65						
age group	(0)		0			
0-4	-0.852	0.3432	0.0131	0.4266	0.2177	0.8359
5-9	-3.281	1.017	0.0013	0.0376	0.0051	0.2759
10-14	-0.1861	0.2907	0.5221	0.8302	0.4696	1.468
15-19	1.999	0.2	0	7.385	4.99	10.93
20-24	3	0.1916	0	20.07	13.79	29.22
25-29	3.047	0.1921	0	21.05	14.44	30.67
30-34	2.695	0.1958	0	14.81	10.09	21.74
35-39	2.079	0.2075	0	7.996	5.324	12.01
40-44	1.621	0.2275	0	5.058	3.238	7.899
45-49	0.9478	0.2701	0.0005	2.58	1.52	4.38
50-54	0.4036	0.3338	0.2266	1.497	0.7783	2.88
55-59	0.8498	0.3055	0.0054	2.339	1.285	4.257
60-64	-0.0197	0.4843	0.9675	0.9805	0.3795	2.533
65-69	-0.5255	0.7311	0.4723	0.5913	0.1411	2.478
70-74	0.0821	0.5334	0.8776	1.086	0.3816	3.088
75+						

df: 435 deviance: 330.7 p = 0.9999 number of iterations: 8

Figure 17: Logistic regression model for males, with province-year interaction

The coefficients of interaction model are negative. Comparing the deviance we see the deviance has decreased from 368.3 to 330.7, a decrease of 37.6, after including

12 additional parameters, corresponding to a significant p-value. The residuals versus normal scores, the standardized residuals are shown in follow figure.

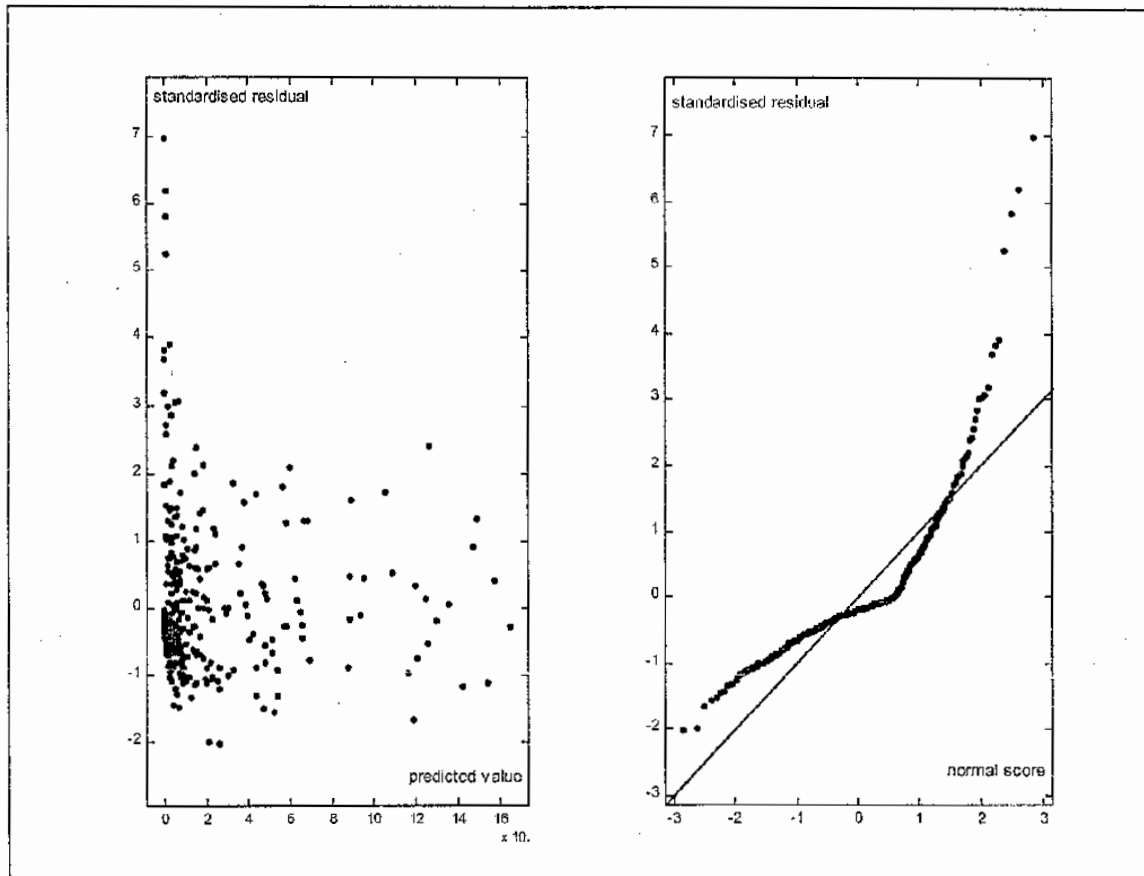


Figure 18: the residuals vs normal scores for males, with province-year interaction

The residuals in this more complex model are different from those in the model with no interactions. The cell corresponding to the largest ten standardized residuals are shown in Table 9.

Table 9: The largest residuals corresponding for males

age group	province	year	residuals	number of cases
55-59	Narathiwat	1991	6.99	1
70-74	Yala	1993	6.21	1
55-59	Satun	1993	5.26	1
60-64	Yala	1993	5.81	1
10-14	Yala	1996	3.91	3
5-9	Songkhla	1993	3.82	1
50-54	Songkhla	1991	3.69	1
55-59	Songkhla	1991	3.18	1
75+	Satun	1996	3.07	1
40-44	Pattani	1990	3.04	1

From this table, it can be seen that the large residuals do not correspond to unusual observations. With the possible exception of the data for age group 10-14 in Yala in 1996, they indicate poor approximations to normality in the normal scores plot rather than any deficiency in the model.

Finally, the coefficients for the province-year interaction effects are plotted in Figure 4.3 These may be compared with the odds ratios plotted in the Chapter 3. (Figure 11)

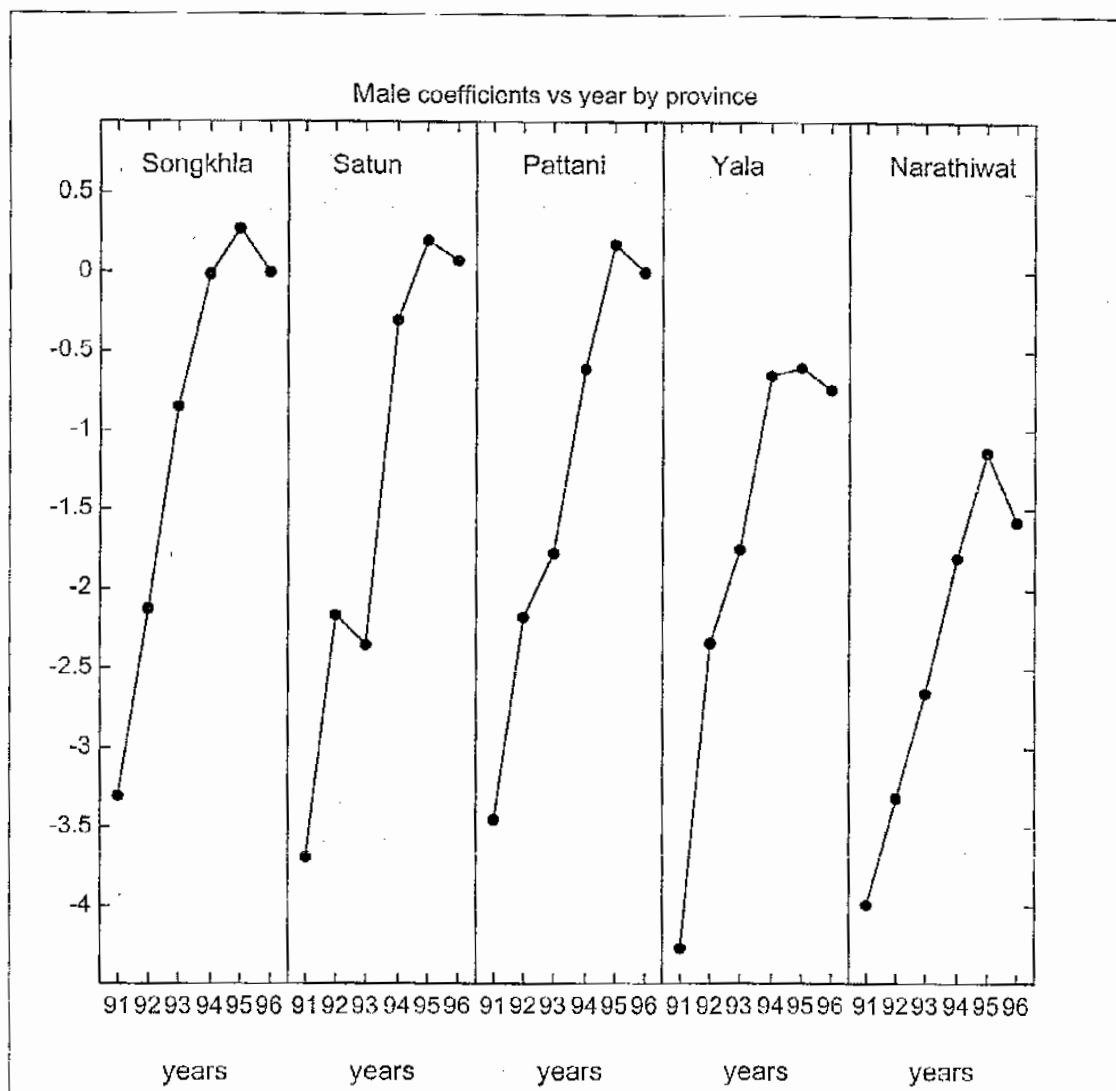


Figure 19: Coefficient plots for males incidence vs year by province

This graph presents clearly that the incidence of HIV/AIDS among males decreased in Songkhla after reaching a peak in 1995, with a similar pattern for every provinces, where the risk was generally lower during the same period. In Yala the risk was straightly lower still.