Chapter 4

Statistical Analysis

In this chapter we use linear regression to fit a model for ln(BMI-8), using the determinants found in Chapter 3 to be separately related to this outcome. These determinants are age group, sex, religion group, birthplace and parents' combined salary.

First, we impute values for the parents' combined salaries coded as zeroes, again using a linear regression model based on relevant demographic predictors. Then we fit the regression model to ln(BMI-8), using the complete set of parents' combined salaries including the imputed values instead of the zeroes. To avoid any assumption that the age group, sex, and religion group effects in the linear regression model are additive, we create a variable comprising all combinations of these demographic variables, called AgeSexRelig. Finally, we use the coefficients in this model to graph both the Age-Sex-Religion patterns of BMI and its variation with respect to birthplace.

4.1 Imputing the Missing Parental Combined Salaries

In Chapter 3 we found that age group and religion group are both associated with parents' combined salary, but not sex. There is an association between ln(parentSalary) and birth place, but for simplicity we use only age group and religion group as predictors. Table 4.1 shows the result of fitting a multiple linear regression model to the variable ln(parentSalary) after excluding the cases with 0-coded salaries, with both predictors included in the model.

Linear Regression Analysis: Outcome = lnParentSalary1				
Determinant	Coefficient	StError	95% CI	p-value
Constant	8.3734	0.0240	8.3263, 8.4205	0.0000
ageGrp	baseline: 13			0.0000
14	0.0170	0.0282	-0.0383, 0.0722	0.5473
15	0.0520	0.0292	-0.0052, 0.1091	0.0747
16	0.1254	0.0328	0.0610, 0.1897	0.0001
17	0.1506	0.0332	0.0854, 0.2157	0.0000
18	0.0667	0.0338	0.0005, 0.1330	0.0482
religGrp	baseline: Islam			0.0000
other	0.5282	0.0188	0.4913, 0.5650	0.0000
r-sq: 0.0866 df: 8461 RSS: 6,046.5847 s: 0.8454				

Table 4.1: Linear regression for imputing log-transformed parents' combined salary

Based on the results in Table 4.1, the formula for imputing the missing salaries corresponding to students in the first two age groups gives the following results.

Muslim aged 12-13: $\ln(\text{parentSalary}) = 8.3734 + 0 + 0$

$= 8.3734;$

Other religion aged 12-13: $\ln(\text{parentSalary}) = 8.3734 + 0 + 0.5282$

$= 8.9016;$

Muslim aged 14: $\ln(\text{parentSalary}) = 8.3734 + 0.0170 + 0$

$= 8.3904;$

Other religion aged 14: $\ln(\text{parentSalary}) = 8.3734 + 0.0170 + 0.5282$

$= 8.9186.$

The results for the other age groups may be imputed similarly.

Figure 4.1 shows the distribution of ln(parentSalary) before and after imputing the missing salaries.

Variable name	Size	Graph	Minimum	Maximum	Mean	StDev	Skewness	Kurtosis
lnParentSalary1	9393		0	12.612	7.872	2.734	-2.232	3.872
lnParentSalary	9393		5.017	12.612	8.736	0.843	0.079	0.485

Figure 4.1: Log-transformed parents' combined salaries before and after imputation

4.2 The Linear Regression Model

We create a new variable by joining the three variables age group, sex, and religion group. Figure 4.2 shows the distribution of this new variable, using the coding system AgeSexRelig = 100*(age group) + 10*sex + religion group. This means, for example, that students who are 12-13 years old, of female sex and religion group Islam are coded as 1301.

Variable name	Size	Graph	Minimum	Maximum
ageSexRelig	9393			316 (3.36) 1301 720 (7.67) 1302 362 (3.85) 1311 755 (8.04) 1312 377 (4.01) 1401 608 (6.47) 1402 394 (4.19) 1411 685 (7.29) 1412 375 (3.99) 1501 557 (5.93) 1502 335 (3.57) 1511 483 (5.14) 1512 344 (3.66) 1601 373 (3.97) 1602 235 (2.50) 1611 240 (2.56) 1612 346 (3.68) 1701 360 (3.83) 1702 230 (2.45) 1711 199 (2.12) 1712 302 (3.22) 1801 287 (3.06) 1802 317 (3.37) 1811 193 (2.05) 1812

Figure 4.2: The distribution of age group, sex and religion group

Next, we fit a linear regression model to ln(BMI-8) using the variables (1) (age, sex, religion) group, (2) birthplace, and (3) parents' combined salary with 0s imputed. Table 4.2 shows the results.

Based on the results in Table 4.2, the formula for predicting ln(BMI-8) is as follows. Suppose a student is in (age, sex, religion) group i and birth place j. The expected value of ln(BMI-8) for this student is given by the formula

$$\mathrm{E}[\ \ln(\text{BMI-8})\ ] = 2.1157 + 0.0220 \times \ln(\text{parentSalary}) + a_i + b_j\ , \qquad (4.1)$$

where the coefficients a_i (i = 1, 2, …, 24) and b_j (j = 1, 2, …, 24) are listed in Table 4.2. The expected value of the BMI is then calculated by assuming that ln(BMI-8) is normally distributed with mean m and standard deviation s, say, and consequently BMI-8 has a lognormal distribution with mean $\exp(m+s^2/2)$. Thus, substituting the estimate of the standard deviation s = 0.2755 from Table 4.2, we obtain

$$\mathrm{E}[\ \text{BMI}\ ] = 8 + \exp(2.1157 + 0.0220 \times \ln(\text{parentSalary}) + a_i + b_j + 0.2755^2/2). \qquad (4.2)$$

Consider, for example, Muslim females aged 12-13 years old (i = 1, a_1 = 0), born in Sabarang subdistrict (j = 1, b_1 = 0) and whose parents' combined salary is such that its logarithm is the average value of ln(parentSalary), namely 8.736. Equation (4.1) gives

$$\mathrm{E}[\ \ln(\text{BMI-8})\ ] = 2.1157 + 0.0220 \times 8.736 + 0 + 0 = 2.3079,$$

and then Equation (4.2) gives, for her expected body mass index,

$$\mathrm{E}[\ \text{BMI}\ ] = 8 + \exp(2.1157 + 0.0220 \times 8.736 + 0 + 0 + 0.2755^2/2) = 18.44.$$

Linear Regression Analysis: Outcome = lnBMIminus8				
Determinant	Coefficient	StError	95% CI	p-value
Constant	2.1157	0.0370	2.0432, 2.1883	0.0000
lnParentSalary	0.0220	0.0038	0.0146, 0.0293	0.0000
ageSexRelig	baseline:	1301		0.0000
1302	0.0158	0.0189	-0.0212, 0.0529	0.4024
1311	0.0123	0.0213	-0.0294, 0.0539	0.5642
1312	0.0205	0.0188	-0.0163, 0.0574	0.2745
1401	0.0389	0.0210	-0.0024, 0.0801	0.0646
1402	0.0910	0.0194	0.0529, 0.1291	0.0000
1411	-0.0063	0.0209	-0.0472, 0.0346	0.7626
1412	0.0534	0.0191	0.0159, 0.0908	0.0052
1501	0.0701	0.0211	0.0287, 0.1114	0.0009
1502	0.1040	0.0198	0.0653, 0.1427	0.0000
1511	0.0074	0.0217	-0.0351, 0.0498	0.7337
1512	0.0605	0.0203	0.0208, 0.1002	0.0028
1601	0.0984	0.0215	0.0562, 0.1405	0.0000
1602	0.1191	0.0214	0.0771, 0.1611	0.0000
1611	0.0844	0.0238	0.0378, 0.1310	0.0004
1612	0.1204	0.0239	0.0735, 0.1673	0.0000
1701	0.1241	0.0215	0.0819, 0.1662	0.0000
1702	0.1104	0.0216	0.0681, 0.1527	0.0000
1711	0.0931	0.0239	0.0462, 0.1400	0.0001
1712	0.1247	0.0252	0.0752, 0.1742	0.0000
1801	0.1194	0.0223	0.0757, 0.1631	0.0000
1802	0.1549	0.0228	0.1103, 0.1995	0.0000
1811	0.1196	0.0221	0.0764, 0.1629	0.0000
1812	0.1327	0.0256	0.0826, 0.1828	0.0000
birthPlace	baseline:	1		0.0000
2	-0.0281	0.0195	-0.0663, 0.0100	0.1483
3	-0.0285	0.0280	-0.0834, 0.0265	0.3099
4	-0.0296	0.0213	-0.0714, 0.0123	0.1660
5	-0.0226	0.0209	-0.0635, 0.0182	0.2776
6	-0.0477	0.0228	-0.0924, -0.0030	0.0364
7	-0.0045	0.0110	-0.0260, 0.0170	0.6828
8	-0.0275	0.0156	-0.0580, 0.0031	0.0778
9	-0.0705	0.0134	-0.0967, -0.0443	0.0000
10	-0.0273	0.0167	-0.0600, 0.0054	0.1015
11	-0.0092	0.0282	-0.0645, 0.0462	0.7455
12	-0.0170	0.0144	-0.0452, 0.0112	0.2371
13	-0.1552	0.0254	-0.2050, -0.1054	0.0000
14	-0.0444	0.0149	-0.0735, -0.0152	0.0029
15	0.0009	0.0158	-0.0301, 0.0320	0.9523
16	0.0118	0.0252	-0.0376, 0.0612	0.6395
17	-0.0449	0.0266	-0.0970, 0.0071	0.0908
18	-0.0019	0.0106	-0.0227, 0.0188	0.8550
19	0.0053	0.0146	-0.0234, 0.0340	0.7160
20	0.0137	0.0210	-0.0274, 0.0548	0.5131
21	-0.0367	0.0265	-0.0887, 0.0153	0.1662
22	-0.0395	0.0233	-0.0852, 0.0062	0.0900
23	0.0188	0.0222	-0.0248, 0.0623	0.3980
24	-0.0035	0.0202	-0.0431, 0.0361	0.8634
r-sq: 0.0463 df: 9345 RSS: 709.1103 s: 0.2755				

Table 4.2: The linear regression model for ln(BMI-8) outcome

Similarly, for non-Muslim males aged 12-13 years ($i = 4$, $a_4 = 0.0205$), also born in Sabarang subdistrict ($j = 1$, $b_1 = 0$) and whose parents' combined salary is 20,000 baht per month, Equation (4.1) gives

$$E[\ln(BMI-8)] = 2.1157 + 0.0220\times\ln(20{,}000) + 0.0205 + 0 = 2.3541,$$

and then Equation (4.2) gives, for his the expected body mass index,

$$E[BMI] = 8 + \exp(2.1157 + 0.0220\times\ln(20{,}000) + 0.0205 + 0 + 0.2755^2/2)$$
$$= 18.9356.$$

And similarly, for Muslim males aged 15 years ($i = 11$, $a_{11} = 0.0074$), born in Arnohru subdistrict ($j = 2$, $b_2 = -0.0281$) and whose parents' combined salary is also given by the mean of ln(parentSalary), namely 8.736, Equation (4.1) gives

$$E[\ln(BMI-8)] = 2.1157 + 0.0220\times8.736 + 0.0074 + (-0.0281) = 2.2872,$$

and Equation (4.2) gives the expected BMI,

$$E[BMI] = 8 + \exp(2.1157 + 0.0220\times8.736 + 0.0074 + (-0.0281) + 0.2755^2/2)$$
$$= 18.2281.$$

And for non-Muslim females aged 15 years ($i = 10$, $a_{10} = 0.1040$), born in Arnohru subdistrict ($j = 2$, $b_2 = -0.0281$) and whose parents' combined salary is 5,000 baht, Equation (4.1) gives

$$E[\ln(BMI-8)] = 2.1157 + 0.0220\times\ln(5{,}000) + 0.1040 + (-0.0281) = 2.3790,$$

and Equation (4.2) gives

$$E[BMI] = 8 + \exp(2.1157 + 0.0220\times\ln(5{,}000) + 0.1040 + (-0.0281) + 0.2755^2/2)$$
$$= 19.2114.$$

The results for others may be computed similarly.

We use the equations above with parents' combined salary taking the mean of ln(parentSalary) to impute the BMIs of students separated by age group, sex, religion group and the median coefficient of birth place, namely Rusamilae subdistrict with the value of coefficient equal to -0.0226.

4.3 Graph and Map of BMI

We create a graph of adjusted mean BMI of the students attending public high schools in Pattani province classified by female Islam, female other, male Islam, and male other religion with any age group as shown in Figure 4.3. The BMI was adjusted for age, sex, religion, parents' combined salary and birth place.

Figure 4.3 shows that the age patterns of BMI differ for the four groups, but they start and end at very similar values. The Muslim females show a steady increase in average BMI from 18.3 at age group 12-13 to 19.6 at age 17 and then remaining constant. The pattern for non-Muslim males is similar, but less steady. In contrast, the average BMI for Muslim males does not begin to increase until 15, and then increases rapidly from age 15 to 16, whereas the non-Muslim females increase their BMI rapidly between ages 12-13 and 14 and then slow down.

Finally, we create a schematic range map of the expected values of BMI in the various birth places of the students attending public high schools in Pattani province. These expected values are based on the average of referent median (age, sex, religion) group, namely 15 year-old students with average parents' combined salary, that is, with the value of ln(parentSalary) equal to 8.736, as shown in Figure 4.4.

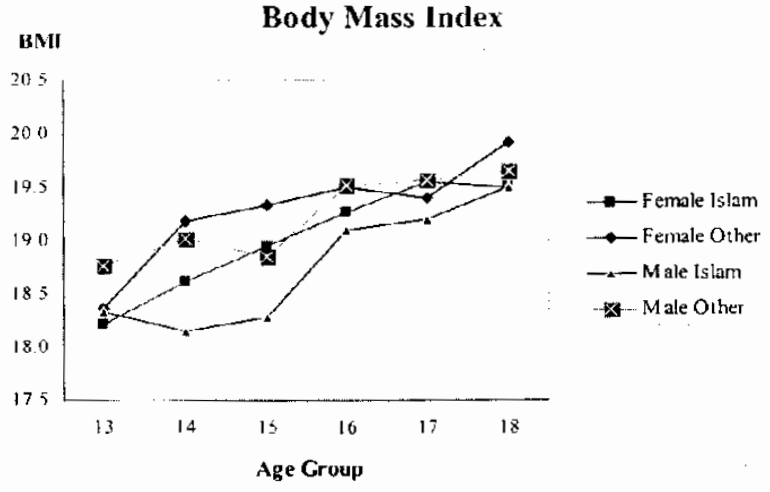

Figure 4.3: Age patterns of body mass index

Figure 4.4 shows that the students who were born in Yarang and Kapor districts, and Sabarang subdistrict have highest BMI between 19.09 and 19.67. The lowest BMI, between 17.42 and 18.46, is located in Panare and MaiKan districts.

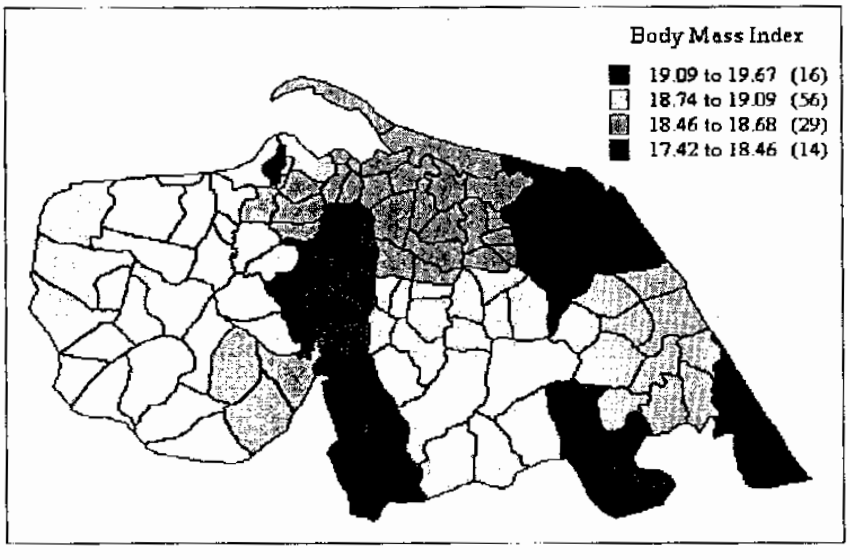

Figure 4.4: The Pattani province's body mass index map