

## **Chapter 3**

### **Data Analysis**

In this chapter, the data analyses are presented in two parts. The first part consists of descriptions of the general characteristics of the patients who come for treatment to the Thai Traditional Medicine Clinic in Khokpho Hospital. The second part consists of a description of the associations between the outcome variables and each determinant.

Thai Traditional Medicine (TTM) Clinic was started in October 1999. Over the four year period until December 2003, on average there were 368 patients each year and the number of visits per year was 931. Our study focused on the 327 patients who visited the clinic in 2003. Their total number of visits, including visits going back as far as October 1999, was 1319. There were 205 (62.7%) patients visiting more than once.

#### **3.1 Descriptions of the Variables**

The roles of the variables may be classified as determinants, intervening and outcome variables. These variables and their roles and data types are listed in Table 3.1.

As shown in Table 3.1, there are 6 determinants, one variable (sex) is binary, one variable (age) is ordinal, and four (religion, residence, occupation and card type) are nominal.

We classified the role of eight of the variables as intervening. Two (symptoms and diagnostic group) are multinomial, and six (duration day, average period, cost per day, massage cost, massage cost and sauna cost) are continuous. The outcomes of interest in the study are the number of visits (ordinal), and the total cost (continuous).

<i>Variable</i>	<i>Role</i>	<i>Type</i>
Subject number	identifier	
<b>Demographic characteristics</b>		
Sex	determinant	binary
Age	determinant	ordinal (5)
Religion	determinant	nominal (3)
Residence	determinant	nominal
<b>Socio-economic characteristics</b>		
Occupation	determinant	nominal (7)
Card type	determinant	nominal (9)
<b>Health Problem</b>		
Patient type	intervening	binary
Symptoms	intervening	multinomial
Diagnosis group	intervening	multinomial
<b>Treatment</b>		
Duration (days)	intervening	continuous
Average period	intervening	continuous
Cost per day	intervening	continuous
Massage cost	intervening	continuous
Compress cost	intervening	continuous
Sauna cost	intervening	continuous
Total cost	outcome	continuous
Number of visits	outcome	ordinal (4)

*Table 3.1: Variables and their roles and data types*

### **3.2 Frequency Distributions of the Determinants**

#### ***Demographic characteristics***

Figure 3.1 shows the distributions of the categorical demographic determinants of the patients.

Most of the patients in our sample visiting the Thai Traditional Massage Clinic are female (64.2%). Their age distribution is approximately uniform, with 24.5% in the 45-55 age group and only 13.5% aged below 35. Although most (54%) residents of Khopho district are Muslim, only 13.5% of the patients visiting the clinic are of the Islamic religion, most of the others are Buddhists, with the exception of one Christian.

sex	327		210 (64.22)	female
			117 (35.78)	male
age	327		44 (13.46)	0-5
			64 (19.57)	6-14
			100 (30.58)	15-24
			71 (21.71)	25-34
			108 (32.68)	35+
religion	327		44 (13.46)	1:islam
			283 (86.54)	2:other

Figure 3.1: Frequency distributions of demographic determinants

Figure 3.2 shows the distribution of residence of the patients. Most of patients, 264 (80.7%) come from Pattani Province, with 27 (8.3%) from Songkla Province, 16 (4.9%) from Yala Province, and 7 (2.1%) from other provinces.

Most of the patients (73.7%) come from Khokpho district in Pattani Province, with Taepa district of Songkhla (5.5%) and Yala city (4.6%) providing the next largest sources of patients.

tumbon	327		0 (0.31)	a neu ru
			2 (0.61)	bangkhp
			0 (0.00)	bangnode
			0 (0.00)	boa thong
			0 (0.00)	boa yang
			1 (0.31)	changhatok
			0 (0.00)	datao
			0 (0.00)	khaon
			2 (0.61)	khaon no ri
			0 (0.00)	khaopleng
			22 (6.73)	khokpho
			10 (3.06)	lammai
			0 (0.00)	laili
			0 (0.00)	lumpili
			10 (3.06)	masrood
			17 (5.17)	nakade
			10 (3.06)	naradoo
			0 (0.00)	paborn
			1 (0.31)	para
			2 (0.61)	parklao
			1 (0.31)	plean
			0 (0.00)	pulohpuyo
			0 (0.00)	rusamilae
			1 (0.31)	sabarang
			1 (0.31)	sakhae
			1 (0.31)	sataeng
			1 (0.31)	tha reoa
			5 (1.53)	thamung
			1 (0.31)	thung pla
			0 (0.00)	yamoo
			1 (0.31)	other
			22 (6.73)	unknown
amphur	327		24 (7.34)	1:had yai
			10 (3.06)	2:kokpho
			1 (0.31)	3:maelan
			1 (0.31)	4:muang pattani
			1 (0.31)	5:muang songkla
			1 (0.31)	6:muang yala
			1 (0.31)	7:nongchik
			1 (0.31)	8:panare
			1 (0.31)	9:sabayoi
			1 (0.31)	10:taepa
			0 (0.00)	11:thong yang daeng
			0 (0.00)	12:yala
			1 (0.31)	13:other
			1 (0.31)	14:unknown
province	327		264 (80.73)	1:Pattani
			27 (8.26)	2:Songkla
			16 (4.89)	3:Yala
			7 (2.14)	4:other
			1 (0.31)	5:unknown

Figure 3.2: Frequency distribution of place of residence, grouped by tumbon (subdistrict), amphur (district) and province

**Socio-economic characteristics**

Figure 3.3 gives the frequency distributions of socio-economic characteristics.

The patient type is classified as either “repeat visit”, if the patient visited the clinic in the four years prior to 2003, or “first visit only”, if there is no record of such a visit.

Most patients (62.7%) visited more than once.

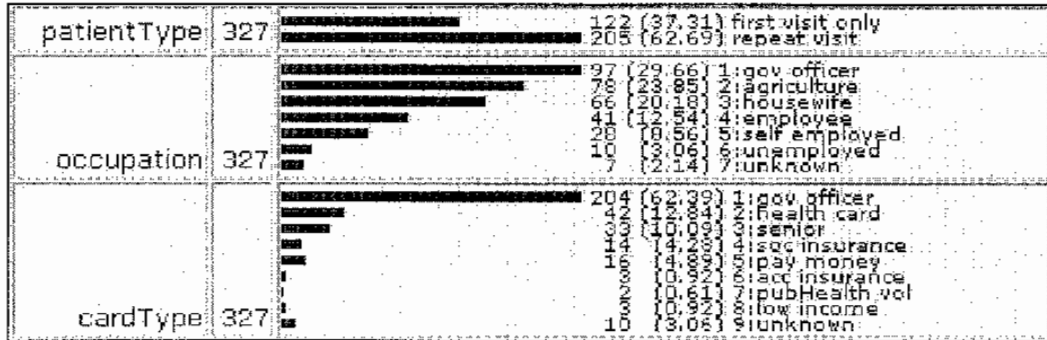


Figure 3.3: Frequency distributions of socio-economic characteristics

With respect to occupation, 29.7% are government officers, 23.8% are employed in agriculture, and 20.2% are housewives.

The type of card refers to the method of payment for the visit. Only 16 (4.9%) pay cash for their treatment and 42 (12.9%) use a health (gold) card.

**Health Problems**

Figure 3.4 shows frequency distributions of the health problems of the patients over all visits. For most visits (74.9%) the patients do not stay overnight at the hospital.

The most popular reason for a visit (28.7%) is arthritis (“worn out and full of aches and pains”), followed by back pain (14.9%), leg pain (13.3%) and shoulder pain (9.0%). The most of the visits (90.9%) involve problems with the musculoskeletal system.

Department	1319	311 (23.58)	in patient
		988 (74.91)	out patient
		20 (1.52)	unknown
Symptom	1319	378 (28.66)	01: arthritis
		196 (14.86)	02: back
		176 (13.34)	03: leg
		119 (9.02)	04: shoulder
		98 (7.43)	05: knee
		66 (5.03)	06: foot
		60 (4.55)	07: waist
		50 (3.79)	08: CVA
		43 (3.26)	09: arm
		25 (1.90)	10: hip
		21 (1.59)	11: neck
		11 (0.83)	12: head
		7 (0.53)	13: URI
		6 (0.45)	14: hand
		2 (0.15)	15: dizziness
		58 (4.40)	16: unknown
Diagnosis	1319	1199 (90.90)	1: musculoskeletal
		47 (3.56)	2: nervous
		10 (0.76)	3: respiratory
		6 (0.45)	4: skin
		4 (0.30)	5: circulatory
		58 (4.25)	6: unknown

Figure 3.4: Frequency distributions of Health Problems

### Treatments

Figure 3.5 shows the distributions of the continuous outcomes. Duration means the number of days from the patient's first visit until their last visit in 2003. The average duration is 126.6 days, ranging from 1 to 1541 days with standard deviation 301.0.

Average period is the duration divided the number of visits. The mean of the average period is 22.6 days, ranging from 1 to 343 days, with standard deviation 52.38.

The cost per day is the total cost of treatment for a patient divided by the duration. The mean of cost per day is 109.4 baht, ranging from 0 to 400 baht with standard deviation 107.0 baht. The minimum is zero because one patient, a senior government officer, failed to complete the registration but was still provided with treatment as a special favour by the staff in the clinic.

Variable name	Size	Graph	Minimum	Maximum	Mean	StDev	Skewness	Kurtosis
durationDays	327		1	1,541	126.581	301.032	3.000	8.329
avePeriod	327		1	343.000	22.614	52.378	3.276	11.501
costPerDay	327		0	400.000	109.355	107.018	1.217	1.047

Figure 3.5: Frequency distributions of continuous determinant

Figure 3.6 shows the distributions of the treatment costs for the 1,319 patient visits to the Thai Traditional Massage clinic from 1999 to 2003. There are three types of treatment, massage, compress, and sauna, and a patient can have one, two, or all three types during the same visit. If a patient does not have a particular type of treatment on a visit, the cost is taken to be zero.

Thus the massage cost associated with each visit ranges from 0 to 300 baht, with mean 150.5 baht and standard deviation 89.7 baht. However, after excluding the 0s to give the actual cost of the massage itself, the mean cost of a massage treatment is 159.5 baht and the standard deviation is 84.2 baht.

The compress cost associated with each visit ranges from 0 to 100 baht, with mean 52.4 baht and standard deviation 28.1 baht. After excluding the 0s to give the actual cost of the compress, the mean cost of a compress treatment is 60.6 baht and the standard deviation is 20.4 baht.

The sauna cost associated with each visit ranges from 0 to 100 baht, with mean 2.6 baht and standard deviation 13.4 baht. After excluding the 0s to give the actual cost of the sauna, the mean cost of a sauna treatment is 69.0 baht and the standard deviation is 16.9 baht.

The total cost is defined as the total cost each time the patient visits the Thai Traditional Medicine Clinic. The total cost associated with each visit ranges from 0 to 500 baht, with mean 205.5 baht and standard deviation 103.1 baht. After excluding the 0s corresponding to the five visits where patients got freebies, the average total cost of treatment is 206.3 baht and the standard deviation is 102.5 baht.


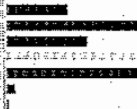

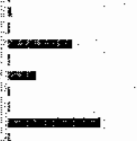
Variable name	Size	Graph	Minimum	Maximum	Mean	StDev
MassageCost	1319		0	300	150.531	89.657
CompressCost	1319		0	100	52.426	28.121
SaunaCost	1319		0	100	2.563	13.442
TotalCost	1319		0	500	205.519	103.073

Figure 3.6: Frequency distributions of treatment component costs

### 3.3 Distributions of Outcome Variables

Figure 3.7 shows frequency distribution of the outcome variable (total cost per patient) before and after a data transformation of the form natural logarithm (cost). The transformation is effective in reducing the heavy skewness and kurtosis in the distribution, although some skewness remains.

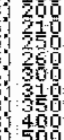

Variable name	Size	Graph	Minimum	Maximum	Mean	StDev	Skewness	Kurtosis
totalCost	327		50	31,850	828.073	2,222.547	9.756	121.682
totalCostLog	327		3.912	10.369	5.965	1.027	0.944	1.377

Figure 3.7: Frequency distributions of total cost and total cost log

Figure 3.8 shows the frequency distribution of the number of visits. The mean number of visits is 4.0 with standard deviation 7.8, but the distribution is very much skewed, with 130 patients (39.8%) visiting only once, and one patient making 100 visits.

Since such a skewed distribution cannot be symmetrised using any transformation, to facilitate the statistical analysis we regrouped the number of visits into four categories with 1, 2, 3-5 and 6 or more visits, respectively.

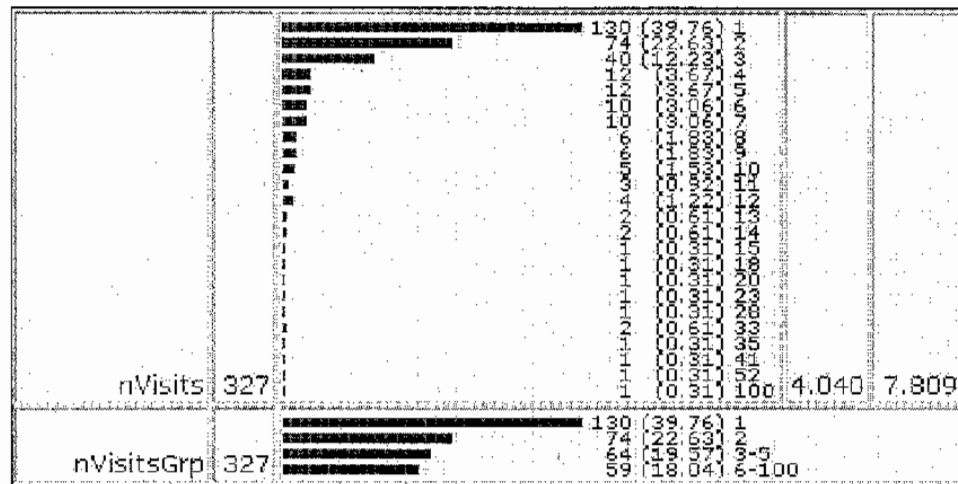


Figure 3.8: Frequency distribution of number of visits

### 3.4 Associations between Total Cost and Determinants

Figure 3.9 shows the association between total cost and age group.

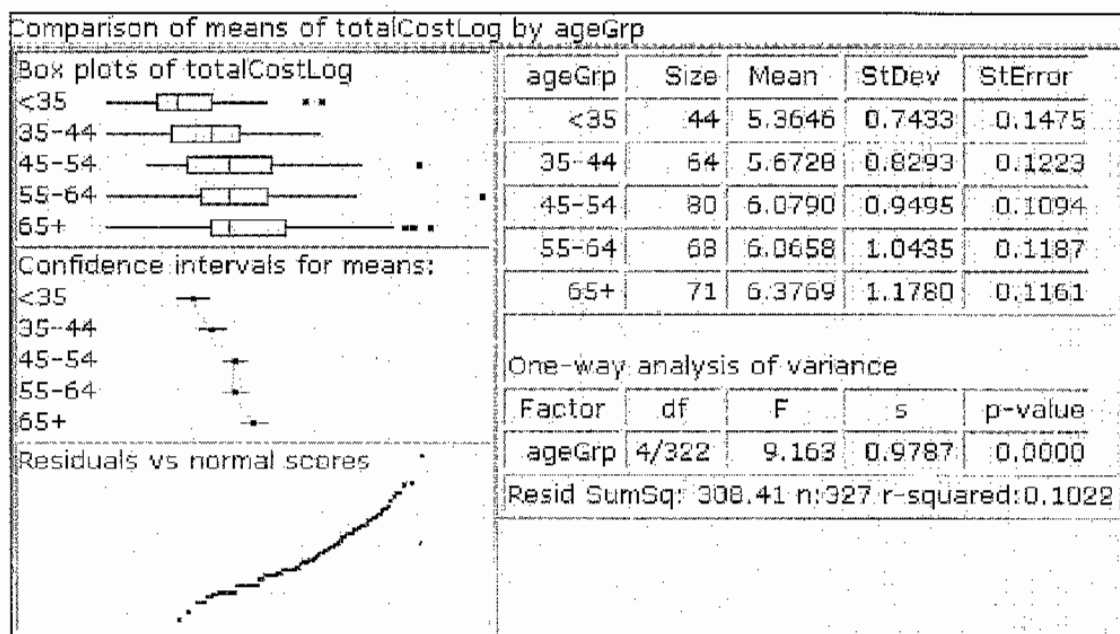


Figure 3.9: Comparison between Total Cost Log and age



This association between age and cost is statistically highly significant ( $p\text{-value} < 0.00005$ ). The average cost increases with age. Transforming the logged cost back to actual baht and using the fact that the mean of a log-normal distribution is  $\exp(m+s^2/2)$ , where  $m$  is the mean and  $s$  is the standard deviation of the underlying normal distribution, the average cost for patients aged below 35 is estimated to be  $\exp(5.365+0.979^2/2) = 345.2$  baht, whereas the estimated average cost for patients aged 65 or more is  $\exp(6.377+0.979^2/2) = 949.8$  baht.

Figure 3.10 shows the association between total cost and occupation. This association is statistically significant ( $p\text{-value} = 0.0012$ ). Housewives incur the highest costs on average and unemployed persons incur the lowest costs.

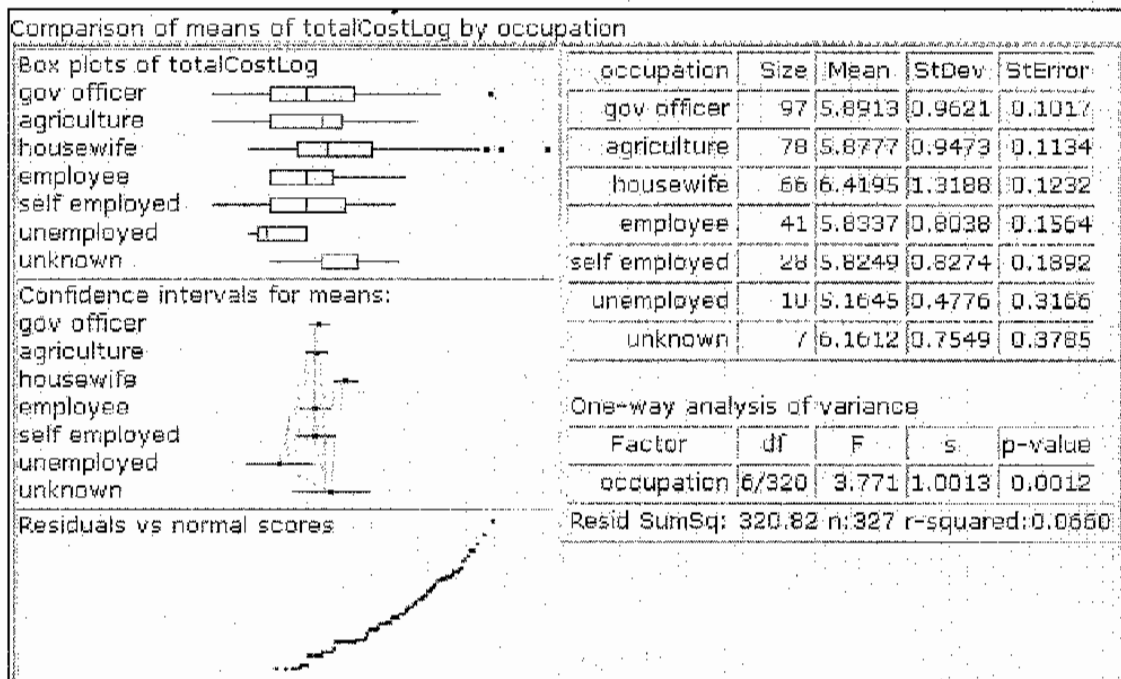


Figure 3.10: Comparison between Total Cost Log and occupation

Figure 3.11 shows the association between total cost and card type. This association is statistically significant ( $p\text{-value} = 0.0078$ ). Cash payers incur the lowest costs and those using their low income status incur the highest costs.

There is no evidence of any relation between the total cost and sex, religion or residence.

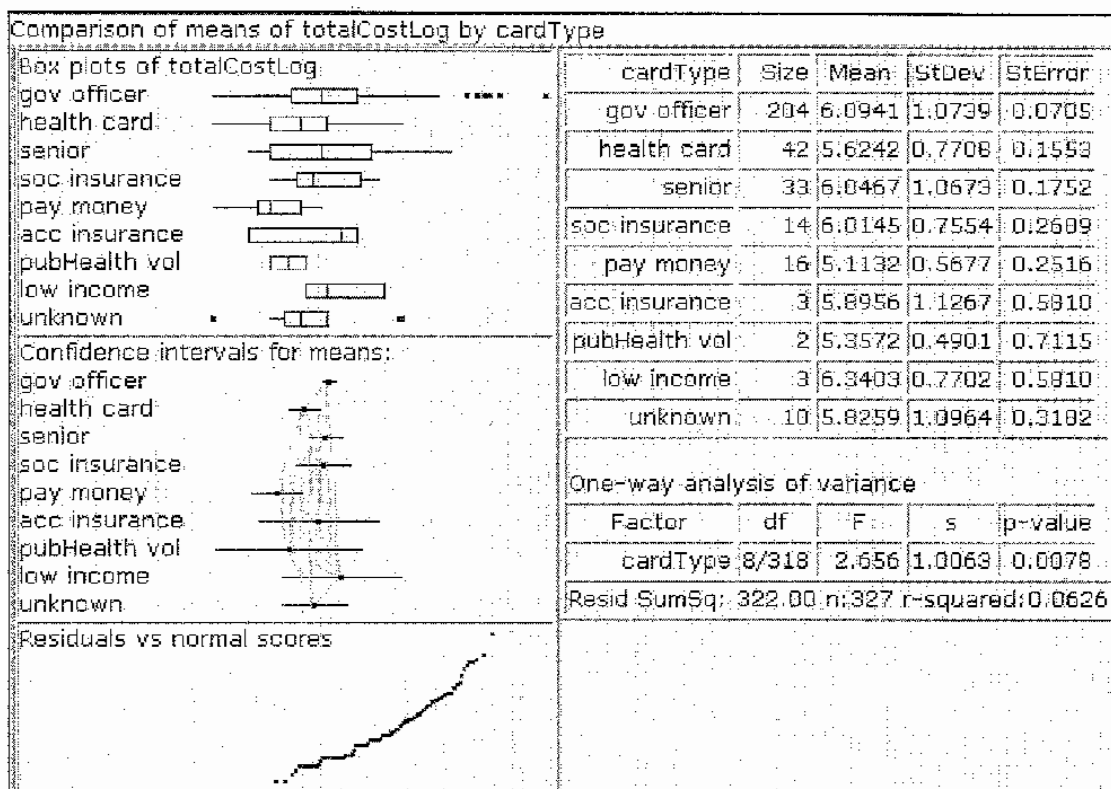


Figure 3.11: Comparison between total cost and card type

### 3.5 Associations between Number of Visits and Determinants

Figure 3.12 shows association between the number of visits and sex. There is no evidence of an association between the number of visits and sex ( $p\text{-value} = 0.18$ ).

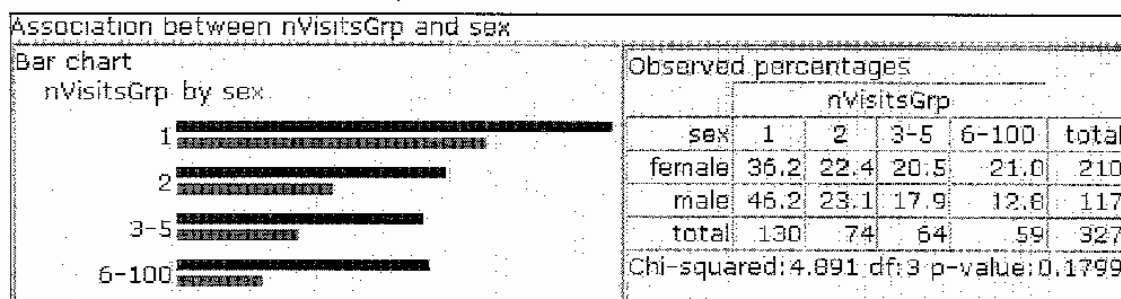


Figure 3.12: Association between number of visits and sex

Figure 3.13 shows the association between number of visits and age. There is no evidence of an association between the number of visits and age (p-value = 0.29).

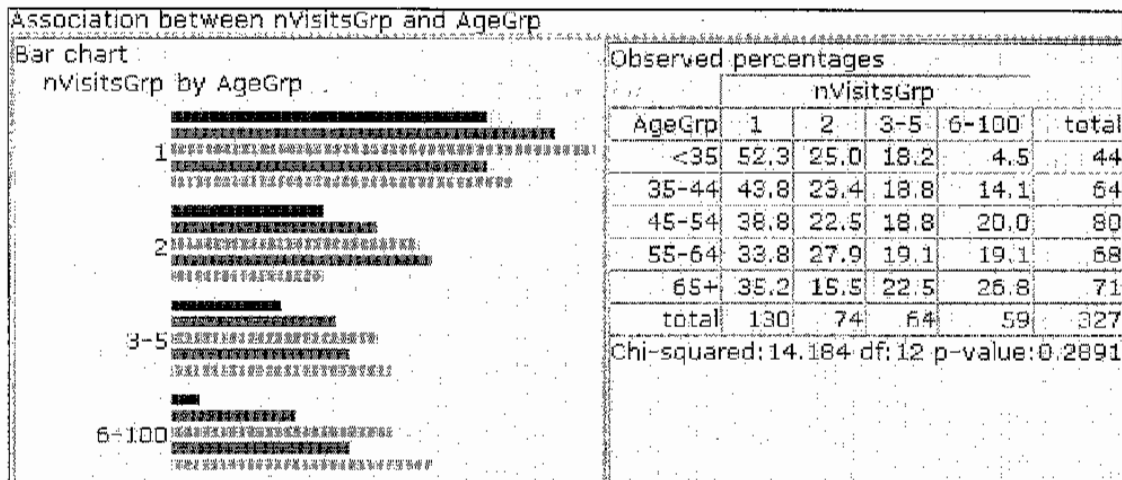


Figure 3.13: Association between number of visits and age

Figure 3.14 shows association between number of visits and religion. There is no evidence of an association between the number of visits and religion (p-value = 0.10).

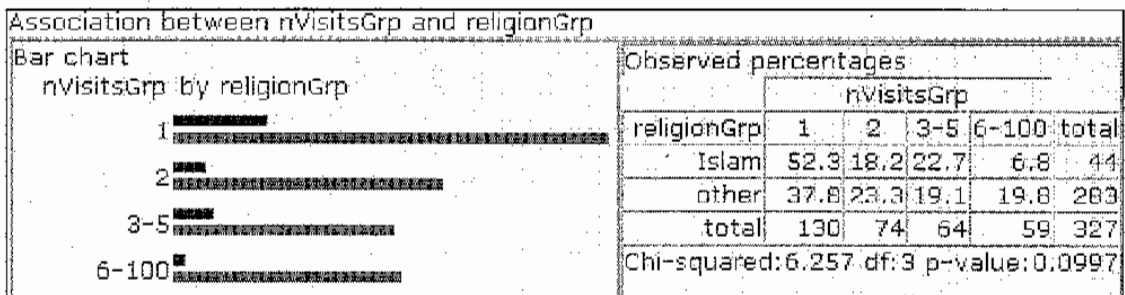


Figure 3.14: Association between number of visits and religion

Figure 3.15 shows the association between number of visits group and tumbon. There is no evidence of an association (p-value = 0.62).

Figure 3.16 shows the association between number of visits group and occupation.

There is no evidence of an association (p-value = 0.45).

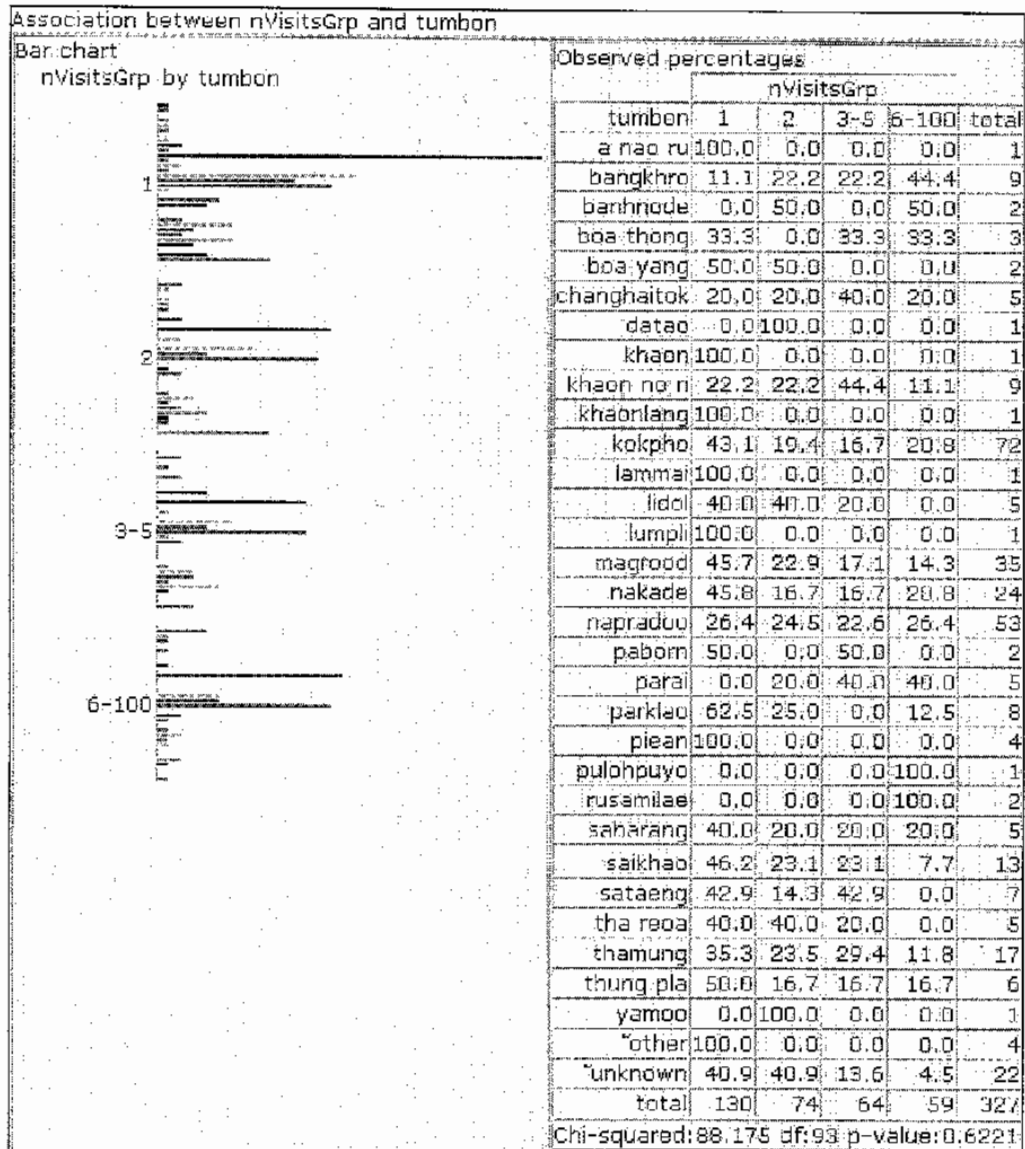


Figure 3.15: Association between number of visits and tumbon

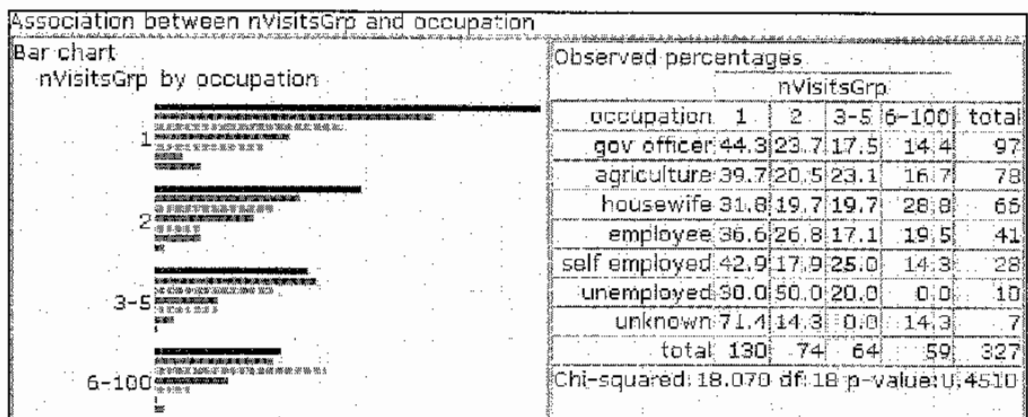


Figure 3.16: Association between number of visits and occupation

Figure 3.17 shows association between number of visits and card type. There is no evidence of an association between the number of visits and card type (p-value = 0.70).

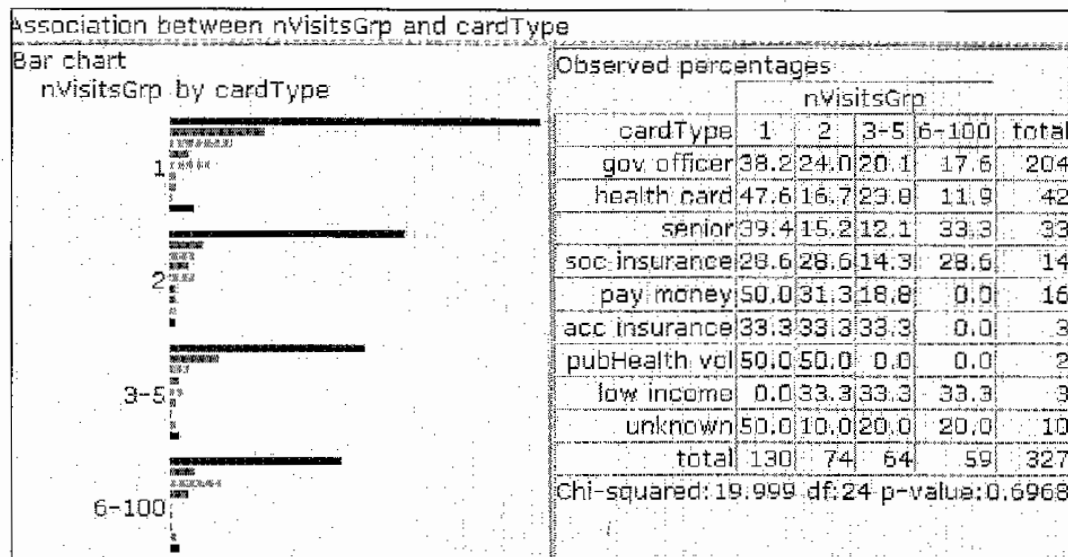


Figure 3.17: Association between number of visits and card type

### 3.6 Multiple Regression Model

In Section 3.4 we reported associations between the total cost outcome and three of the determinants – age group, occupation, and card type. Now we investigate the relation between the total cost and these three variables considered together, using a multiple linear regression model. When all three determinants are included in the model, it turns out that only age group is statistically significant, as Figure 3.18 shown. Since occupation and card type are correlated, it is preferable to omit one of them from the model to get a more useful model for the total cost associated with a patient. Figure 3.19 shows what happens when card type is omitted. Both age group and occupation are jointly associated with the total cost. Comparing this result with those shown in Figures 3.9 and 3.10, the r-squared is 13.7%, which may be compared with an r-squared of 10.2% for age group alone and 6.6% for occupation alone. The age effects shown in Figure 3.18 are similar to those in Figure 3.19, but the occupations effects are different,

indicating that age is a confounder for the association between occupation and treatment cost.

Linear Regression Analysis: Outcome = totalCostLog				
Determinant	Coefficient	StError	95% CI	p-value
Constant	5.4836	0.1851	5.1207, 5.8465	0.0000
ageGrp	baseline: <35			0.0012
35-44	-0.1662	0.2048	-0.2352, 0.5676	0.4177
45-54	0.5886	0.2048	0.1871, 0.9900	0.0048
55-64	0.6227	0.2214	0.1887, 1.0567	0.0052
65+	0.8531	0.2395	0.3838, 1.3225	0.0004
occupation	baseline: gov officer			0.1500
agriculture	-0.0757	0.1860	-0.4402, 0.2889	0.6849
housewife	-0.3842	0.1777	0.0360, 0.7324	0.0314
employee	0.1491	0.2329	-0.3074, 0.6056	0.5226
self employed	0.0331	0.2333	-0.4242, 0.4905	0.8873
unemployed	-0.1421	0.3720	-0.8712, 0.5871	0.7028
unknown	-0.1426	0.4060	-0.9982, 0.6531	0.7257
cardType	baseline: gov officer			0.5001
health card	-0.2595	0.1952	-0.6440, 0.1251	0.1870
senior	-0.3101	0.2043	-0.7106, 0.0904	0.1301
soc insurance	0.1809	0.3311	-0.4681, 0.8298	0.5853
pay money	-0.4934	0.3055	-1.0921, 0.1053	0.1073
acc insurance	0.0476	0.5800	-1.0891, 1.1844	0.9346
pubHealth vol	-0.4281	0.7071	-1.8139, 0.9578	0.5453
low income	0.2689	0.5852	-0.8800, 1.4178	0.6468
unknown	-0.0788	0.3300	-0.7256, 0.5679	0.8114
r-sq: 0.1576 df: 308 RSS: 289.3831 s: 0.9693				

Figure 3.18: Multiple regression model for total cost: three predictors

Linear Regression Analysis: Outcome = totalCostLog				
Determinant	Coefficient	StError	95% CI	p-value
Constant	5.4067	0.1770	5.0597, 5.7537	0.0000
ageGrp	baseline: <35			0.0000
35-44	0.2339	0.1992	-0.1566, 0.6244	0.2413
45-54	0.6876	0.1956	0.3042, 1.0710	0.0005
55-64	0.6983	0.2035	0.2995, 1.0972	0.0007
65+	0.9108	0.2132	0.4929, 1.3287	0.0000
occupation	baseline: gov officer			0.0474
agriculture	-0.2129	0.1566	-0.5198, 0.0940	0.1749
housewife	0.3185	0.1703	-0.0153, 0.6524	0.0624
employee	0.0946	0.1853	-0.2686, 0.4577	0.6100
self employed	-0.1021	0.2106	-0.5148, 0.3106	0.6281
unemployed	-0.4031	0.3421	-1.0736, 0.2675	0.2396
unknown	-0.1259	0.3999	-0.9097, 0.6578	0.7530
r-sq: 0.1374 df: 316 RSS: 296.2977 s: 0.9683				

Figure 3.19: Multiple regression model for total cost: two predictors