Chapter 4

Statistical Modeling

In this chapter, we describe statistical modeling used for estimating drowning death rates in Thailand during 2000 to 2009. Negative binomial regression was chosen for modeling and identifying the strength of association between outcome and determinants as it provided the best fit. Sum contrasts were used to obtain confidence intervals for each level of each factor enabling comparison with the overall mean.

4.1 Model fitting

Drowning death rates per 100,000 population were calculated before modeling the data. A linear model was first considered because it is a straightforward model for continuous outcome. Gender, age group, PHA and year were in the model as determinants. Interaction terms were examined. A significant interaction between gender and age group was found and these two variables were combined to form a new variable named gender-age group. Since the assumptions of linear regression were violated, downing death rates were transformed by taking natural logarithm. Zero counts were replaced by 0.5 to avoid taking the logarithm of zero. The r-squared of log-linear model was 84%. However the residuals plot shows few residual values depart from the diagonal line. Thus the normality assumption of residuals was not satisfied. This means that the model does not fit well with the data. Log-linear model was no longer appropriate for this data.

Poisson model was then considered. This model is a count model. Outcome was number of deaths with corresponding population per 100,000 as its offset. After fitting the model, residual deviance was 3446.7 on 1527 degrees of freedom. This model gave over-dispersion result. Thus Poisson regression model was also no longer appropriate for this data.

Negative binomial regression was further model considered for fitting model. The residual deviance of this model was 1630.5 on 1527 degrees of freedom with no over-dispersion. Therefore negative binomial regression was considered to be the best model for fitting the data as it is a better choice.

4.2 Model diagnostic

Figure 4.1 shows plots of residuals versus the normal quartile. The standardized residuals plot for the log linear regression is on left panel, the deviance residuals plot for Poisson regression is on right panel, and negative binomial regression is on the bottom. As shown in this Figure, negative binomial model is more appropriate for modeling drowning death than log linear regression model and Poisson regression model with less residual values depart from diagonal line resulted in more normally distributed residuals.

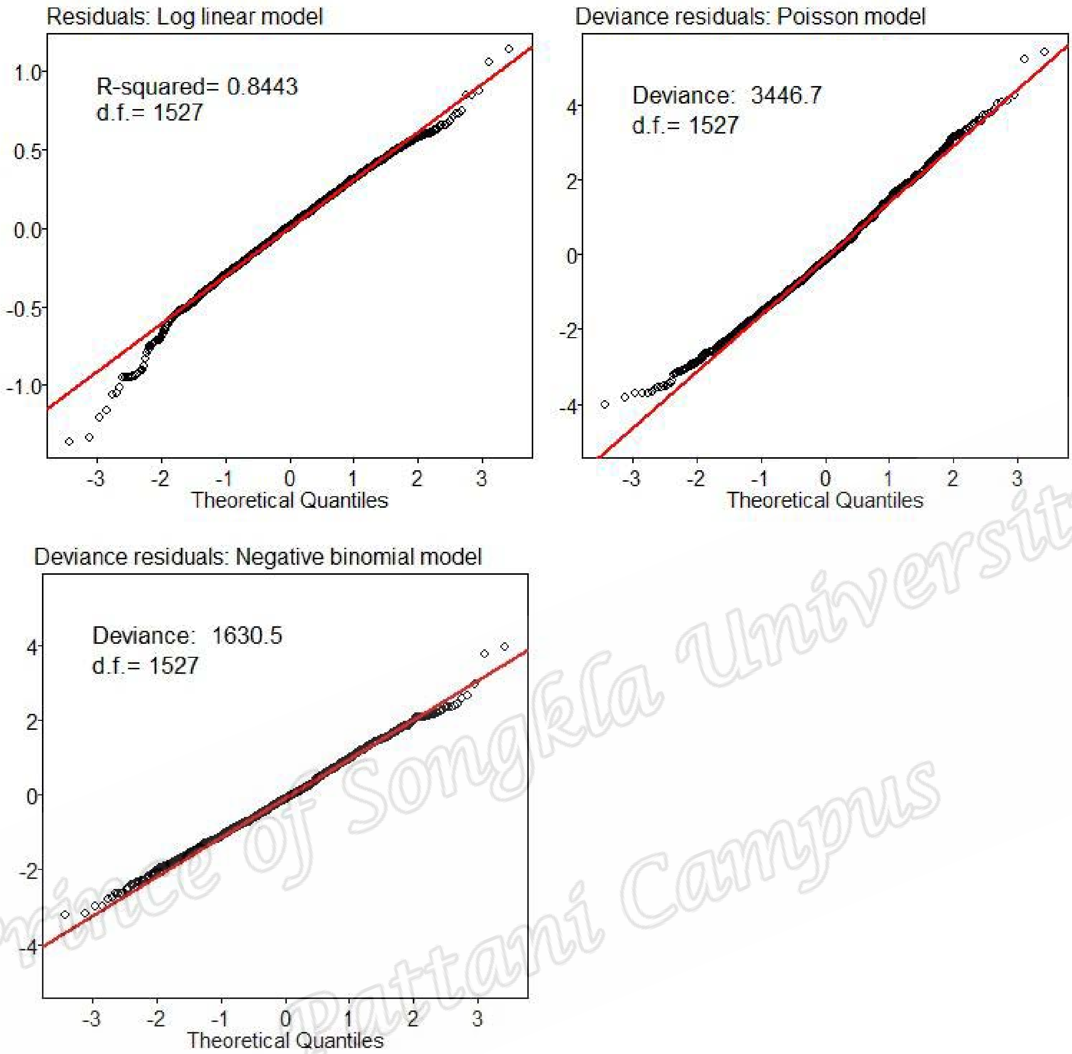


Figure 4.1: Residuals plot for log linear model (left panel), Poisson model (right panel), and negative binomial model (bottom panel)

Figure 4.2 shows plots of observed counts (left panel) and observed death rates (right panel) per 100,000 population against fitted values obtained from negative binomial model. As shown in this Figure, negative binomial model fit with the data quite well. However, one outlier appeared in the plot which is female aged 45-59 years in PHA 12 and year 2006.

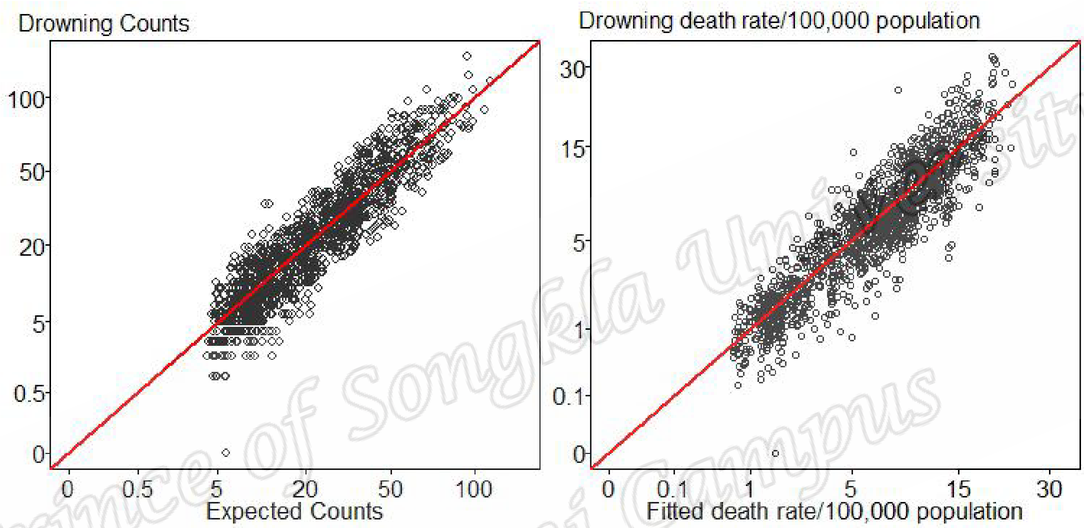


Figure 4.2: Plot of observed counts (left panel) and observed death rate (right panel) against fitted values from negative binomial model

Table 4.1 shows the coefficients of parameters and their standard errors from Negative binomial model using sum contrasts. The results showed that gender-age group, PHA, and year were statistically significantly associated with drowning death rate.

Table 4.1: Coefficients, standard errors and p-values based on the negative binomial regression model fitted to drowning mortality rate in Thailand

Factors	Coefficient	SE	P-value
Constant	1.719	0.009	< 0.001
Gender: Age group			
Male:0-4	0.973	0.026	< 0.001
Male:5-14	0.601	0.024	< 0.001
Male:15-29	0.135	0.024	< 0.001
Male:30-44	0.516	0.023	< 0.001
Male:45-59	0.560	0.024	< 0.001
Male:60+	0.837	0.025	< 0.001
Female:0-4	0.240	0.031	< 0.001
Female:5-14	0.094	0.027	< 0.001
Female:15-29	-1.479	0.036	< 0.001
Female:30-44	-1.343	0.034	< 0.001
Female:45-59	-0.865	0.033	< 0.001
Female:60+	-0.270	0.031	< 0.001
Public Health Area (PHA)			
PHA1	0.020	0.030	0.506
PHA2	0.332	0.029	< 0.001
PHA3	0.316	0.027	< 0.001
PHA4	0.112	0.029	< 0.001

Table 4.1: (cont.)

Factors	Coefficient	SE	P-value
PHA5	0.150	0.025	< 0.001
PHA6	-0.054	0.026	0.040
PHA7	0.034	0.026	0.200
PHA8	0.236	0.029	< 0.001
PHA9	0.196	0.028	< 0.001
PHA10	-0.054	0.029	0.060
PHA11	-0.228	0.031	< 0.001
PHA12	-0.477	0.032	< 0.001
PHA13	-0.582	0.030	< 0.001
Year			
2000	-0.022	0.025	0.382
2001	-0.014	0.025	0.567
2002	0.014	0.025	0.572
2003	-0.026	0.025	0.298
2004	-0.044	0.025	0.076
2005	0.076	0.024	0.002
2006	0.116	0.024	< 0.001
2007	-0.014	0.024	0.569
2008	-0.042	0.025	0.091
2009	-0.045	0.025	0.066

4.3 Confidence interval

The 95% confidence interval graph of drowning death rate per 100,000 population for each factor from negative binomial regression model using sum contrasts are shown in Figures 4.3-4.5. The dotted horizontal lines in each graph represent the overall mean of drowning death rate which was 6.3 per 100,000 population.

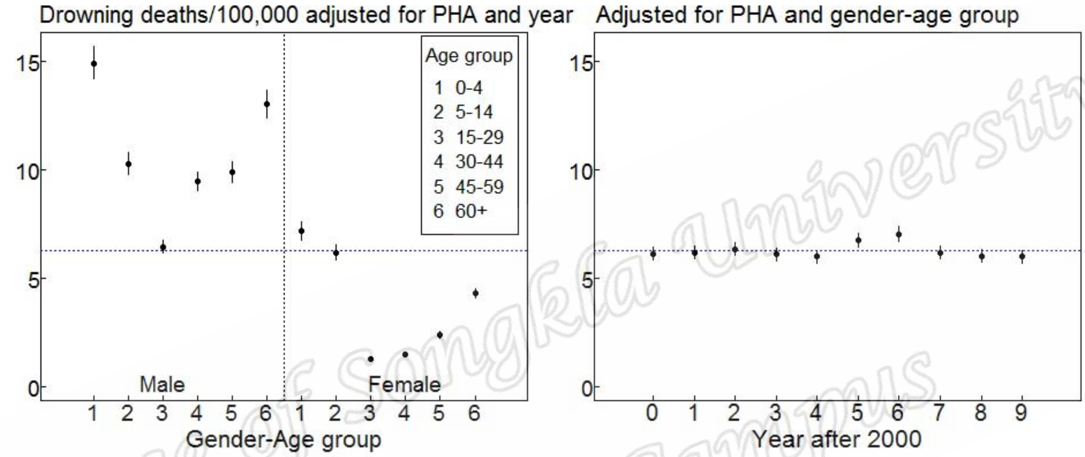


Figure 4.3: Confidence intervals for gender-age group (left panel) and year (right panel) of drowning death rates per 100,000 population adjusted for PHA

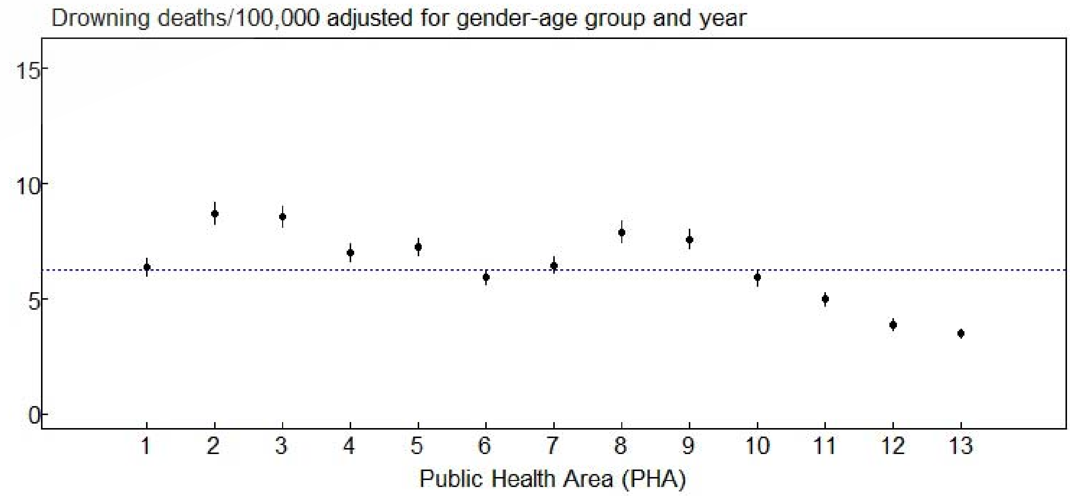


Figure 4.4: Confidence intervals for PHA of drowning death rates per 100,000 population adjusted for gender-age group and year

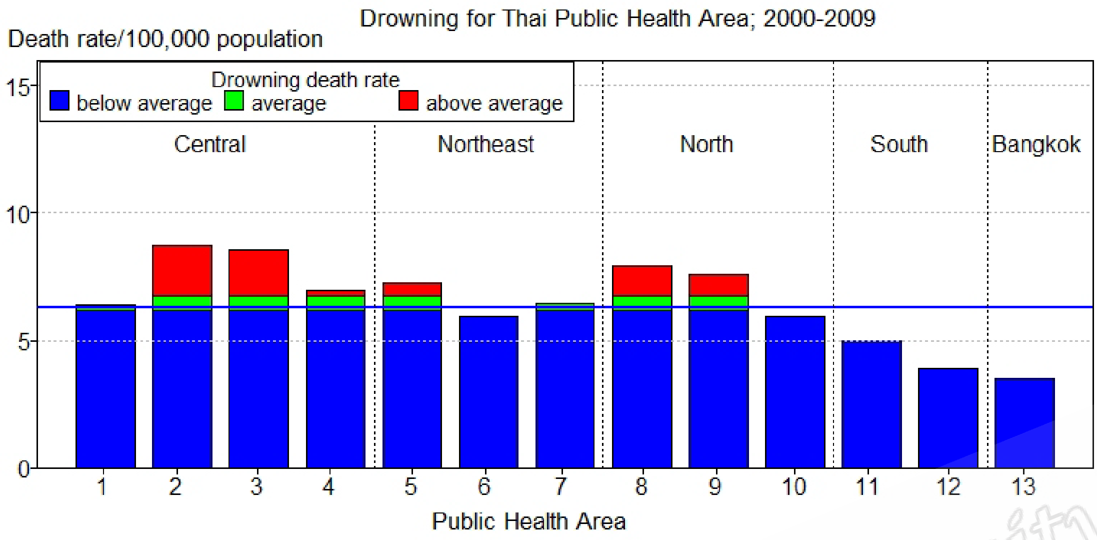
Figure 4.3 and 4.4 shows the 95% confidence intervals of drowning death rates per 100,000 population based on the negative binomial model separated by gender-age group (left panel), year (right panel) and PHA (Figure 4.4). Each graph adjusted for the effects of the other factors in the model.

Drowning death rates by gender-age group in males were consistently higher than those in females across all age groups. A decreasing trend from age 0-4 to 15-29 years was found in both sexes and increasing trend for both sexes occurred after aged 15-29 until aged 60 years and over. Males aged 0-4 years had the highest drowning death rate while females aged 15-29 years had lowest drowning death rate. Males in age groups 0-4, 5-14, 30-44, 45-59, and 60 and over and females in age group 0-4 years had higher drowning death rates than the average rate whereas females aged 15+ had lower death rate than the average rate.

Drowning death rates in year 2005 and 2006 were statistically significant higher than the average with a peak in 2006.

PHA13 had the lowest drowning death rate. Death rates in PHA2, PHA3, PHA4, PHA5, PHA8, and PHA9 were higher than the average death rate whereas death rates in PHA6, PHA11, PHA12, and PHA13 were lower than the average.

Figures 4.5 shows the bar chart for drowning death rates per 100,000 population by PHA based on confidence intervals. The graph shows that drowning was found to be the leading cause of death in the Central (PHA 2, PHA3 and PHA4), the Northeast (PHA5) and in the North regions (PHA8 and PHA9).



Figures 4.5: Bar char of drowning death rates per 100,000 population

Figure 4.6 shows schematic map of drowning death rates per 100,000 population by province, based on confidence intervals. The map shows that the PHA1 (Phra Nakhon Si Ayutthaya, and AngThong), PHA2 (Lop Buri, Sing Buri, ChaiNat, Saraburi, Nakhon Nayok and Suphan Buri), PHA3 (Rayong, Chanthaburi, Trat, Chachoengsao, Prachin Buri and SaKaeo), PHA4 (Ratchaburi, Kanchanaburi, Nakhon Pathom, and Samut Songkhram), PHA5 (Nakhon Ratchasima, Buri Ram and Surin), PHA7 (Si Sa Ket and Nakhon Phanom), PHA8 (Nakhon Sawan, Uthai Thani, Kamphaeng Phet, and Sukhothai), and PHA9 (Uttaradit, Phitsanulok, Phichit and Phetchabun) had drowning death rates significantly higher than the overall mean.

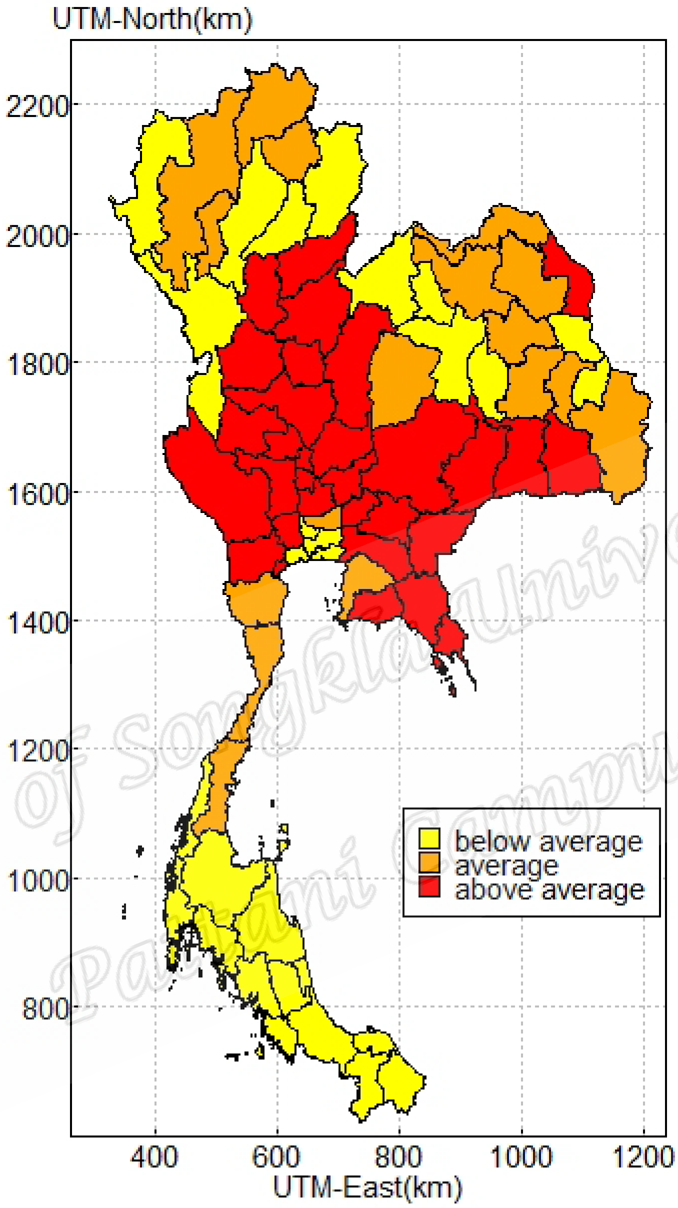


Figure 4.6: Schematic maps of drowning death rate for each province