Chapter 4

Statistical Modeling

In this chapter we use a statistical model to describe the diabetes mortality incidence rate in Southern Thailand during 1996 to 2006. Multiple regressions may be used to model the data of diabetes mortality incidence rates, provided distribution assumptions are met. The Poisson distribution is also the natural model to describe the distribution of count data occurring dependently in variable and can be used to calculate risks associated with diabetes mortality incidence rate.

Graphs and maps were used to show results from fitting the models. Sum contrasts were used to obtain confidence intervals for comparing the adjusted incidence rates within each factor with the overall incidence rate. We used this contrasts to create schematic maps of provinces according to their estimated diabetes mortality incidence rates.

4.1 Statistical Models

Linear regression was the simplest model to start with for fitting mortality incidence rates by age group, year and province for each gender. However, mortality incidence rates have the right skewed distribution therefore they need to be transformed. We found that logarithm is the appropriate transformation function for diabetes mortality incidence rates. The problem occurs when the mortality incidence rate is zero and logarithm of zero would be encounter, and thus a constant need to be added with incidence rates.

We calculated mortality incidence rates per 1000 population according to age group, year and province for each gender and transformed with logarithm. We used one as the constant for avoiding logarithm of zero. After fitting the model, R-squares were obtained from linear models for males 51.09% and 58.39% for females, respectively. However, the linear regression models do not fit well with the data. Graphs of residuals versus normal quantiles for assessing the models are shown in the top panels of Figure 4.1.

Poisson regression would be a better choice. We used number of deaths as its outcome variable with corresponding population per 1000 as its offset. The residual deviance obtained after fitting the Poisson model for males 2111.7 and for females 2332.3 with 2578 degrees of freedom in both gender. Graphs of residuals versus normal quantiles for assessing the models are shown in the bottom panels of Figure 4.1. The Poisson models give an inappropriate fit with the data. Therefore, we separated the data into two groups according to age. Then, the models were fitted separately to the data for age less than 30 years and age 30 years or more.

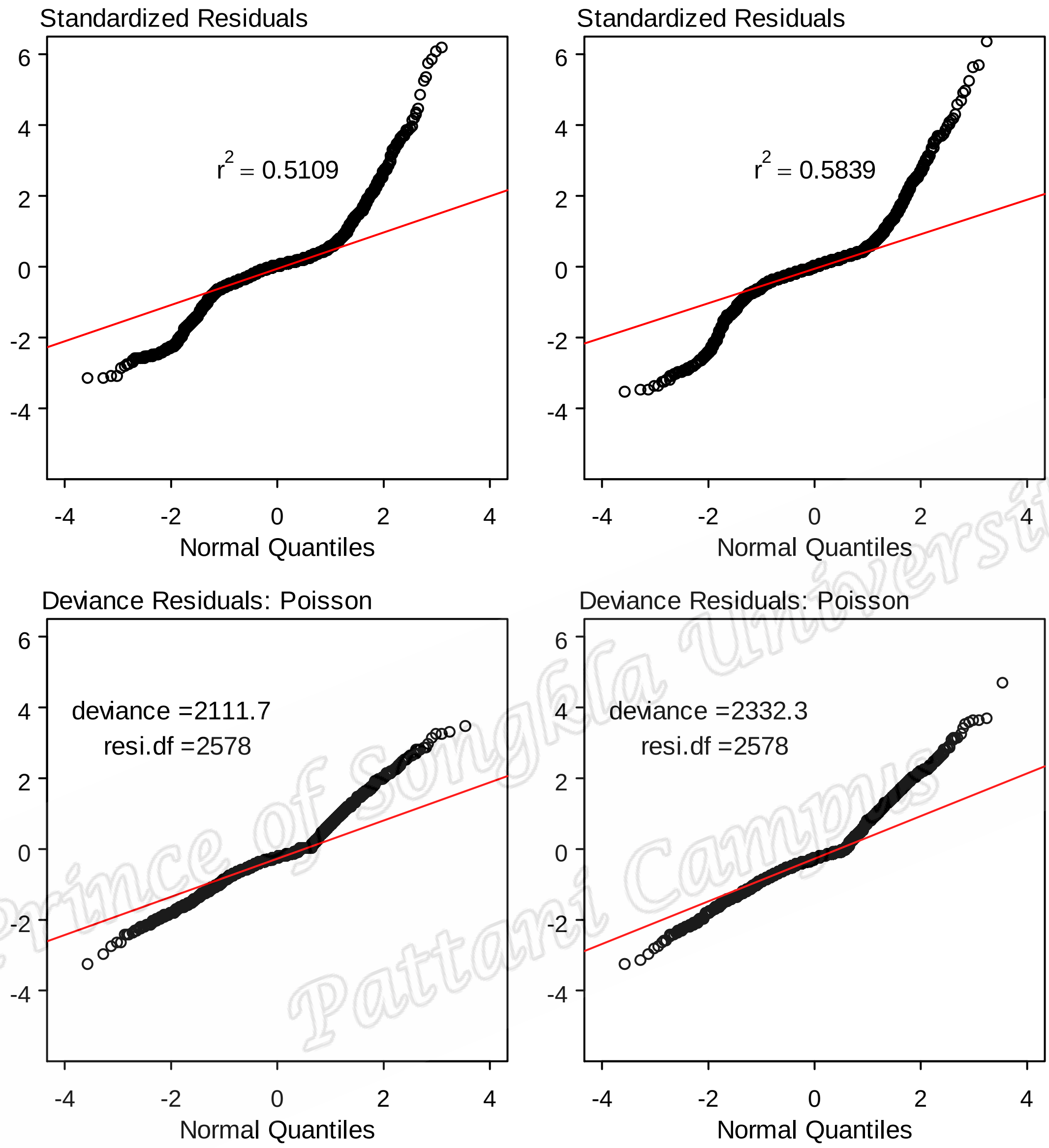

Figure 4.1: Diagnostic plots of linear models on top panels and Poisson models on bottom panels

Figure 4.2 shows the residual plot versus normal quantiles for the two models. The Poisson model for the data with age less than 30 were not fit but the model for the data with age 30 years or more giving the normal quartiles plot of residuals follows the straight line indicating an acceptable fit with the data. The residual deviance obtained after fitting the Poisson models 1766.8 for males and 1876 for females with

1507 degrees of freedom in both gender. Therefore, we omitted the data for age less than 30 from further statistical analysis.

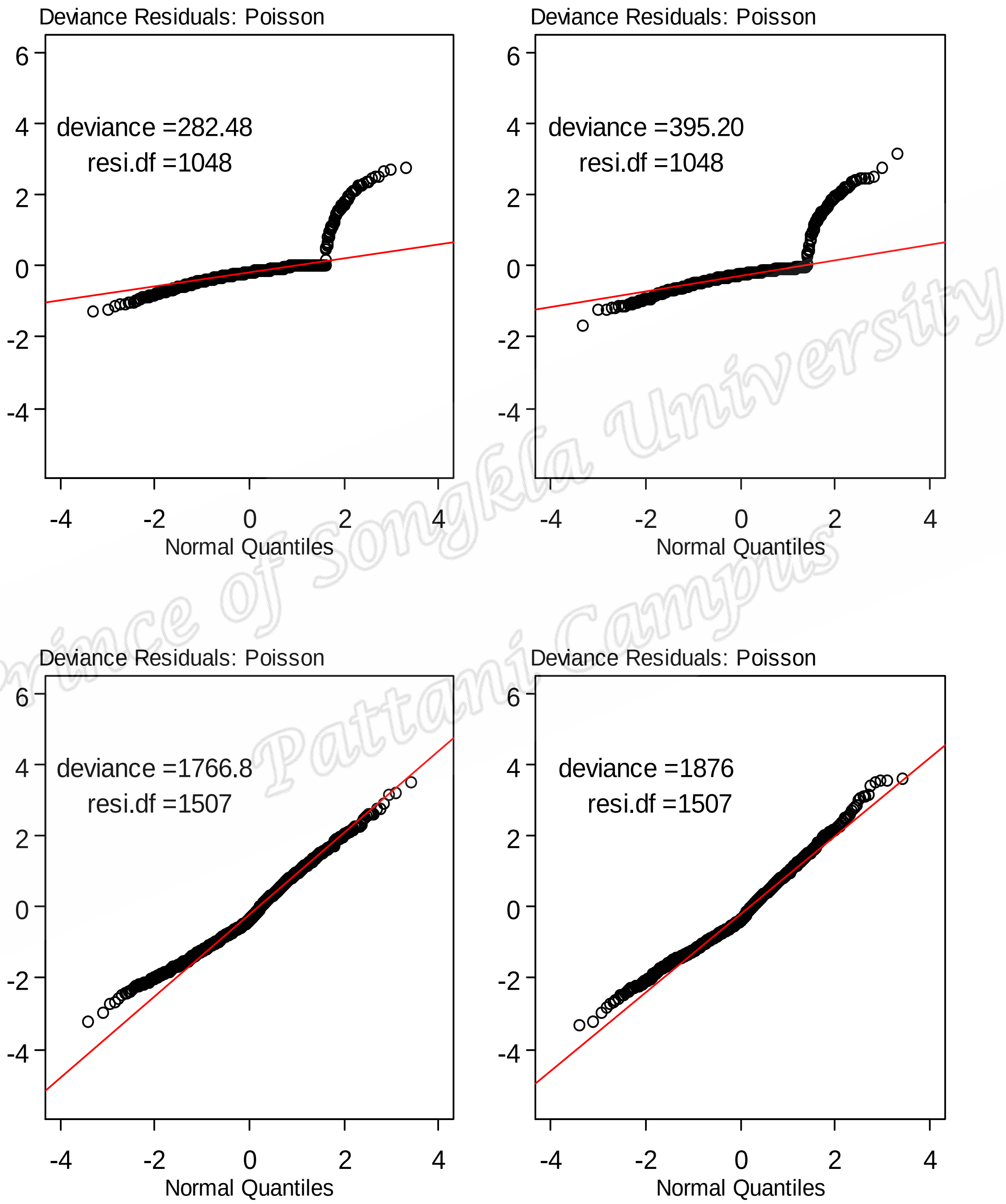

Figure 4.2: Diagnostic plots of Poisson models for age less than 30 on top panels and Poisson models for age more than 30 on bottom panels