Chapter 4

Statistical Modeling

In this chapter we use a statistical model to describe the occurrences of terrorist event data in the three southernmost provinces and four bordering districts of Songkla for 2004 to 2005. While the Poisson distribution is the natural model to start with to describe the distribution of count data occurring independently in a region, where there is over-dispersion or clustering due to failure of the independence assumption, the negative binomial distribution is more general, and can be used to calculate risks associated with locations where there such clustering exists. Such clustering is evident in the preliminary results shown in Chapter 3. The results from this method are compared and discussed, and an interpretation is given to the final result.

As described in Chapter 2, we compute the risk rate of events for each region of interest in a specified period of time by dividing the total number of events recorded in the region over the period by the corresponding population resident in the region in 1000s according to the 2000 Population Census of Thailand.

In Chapter 3 we used the sub-district as the region of interest and we displayed the results as grid maps to smooth out local random fluctuations, enabling the viewer to focus on the broad patterns of terrorism events. In this chapter, where we focus on the statistical comparison of risks across the target population, we use the district as the region of interest. This is done because the number of subdistricts (290) is too large for effective statistical comparisons to be made. To compare every subdistrict with every other subdistrict would require 290x289/2 (41,905) pair wise comparisons. However, with only 37 districts it is feasible to make such comparisons.

Thus in this chapter we model the risk rate of events for each district in successive time periods by dividing the total number of events recorded in the district over the period by the corresponding population resident in the district in 1000s according to the 2000 Population Census of Thailand.

4.1 Statistical Modeling

The preliminary analysis shows large variations in the numbers of events occurring in successive time periods. The daily incidences were highly volatile, and aggregating the data by month failed to smooth out these fluctuations. However, the spatial variations appear to be more regular. The objective in the statistical modeling is to combine these temporal and spatial effects and thus arrive at a formula that can be used to describe, at least approximately, the rate at which events occur in a specified place over a specified period of time. This model thus involves parameters associated with these spatial and time effects, and these parameters should be geared to reflect the corresponding variation in event occurrences.

Given the relative variations in these event occurrences shown in Table 3.3 and Figures 3.2 and 3.3, the model we consider in this section (specified in Equation 2.1) uses two-month periods as the time periods and districts to describe the spatial location.

The residual deviance obtained after fitting the Poisson model is 914.0 with 396 degrees of freedom, indicating substantial over dispersion. The negative binomial model with θ estimated as 7.81 (standard error 1.16) gives the much reduced residual deviance of 494.1, and thus provides a better fit to the data.

Figure 3.4 shows various plots based on fitting the negative binomial model to the observed event rates for 2-month periods in districts for 2004-2005. The top left plot shows observed counts versus model-fitted counts, both expressed as logarithms. The top right plot shows deviance residuals versus normal scores. The lower plots show 95% confidence intervals for the period effects and the district effects, each expressed on a log scale as differences from the baseline values (Jan-Feb 2004 and District 3 in Songkla, respectively). The overall mean is shown as a dotted line in these plots.

The residual plot shows some departure from the straight line indicating a perfect fit, due to a small number of high outliers. These correspond to the clustering of events already identified in Figure 3.1. The district effects are specified in Table 4.1. These confirm the pattern seen in the left panel of Figure 3.3. Four district in Songkla has a comparatively low value as expected, given that on 14 events were reported in this district during the 2-year period: the next highest is three district in Songkla with 74 reported events.

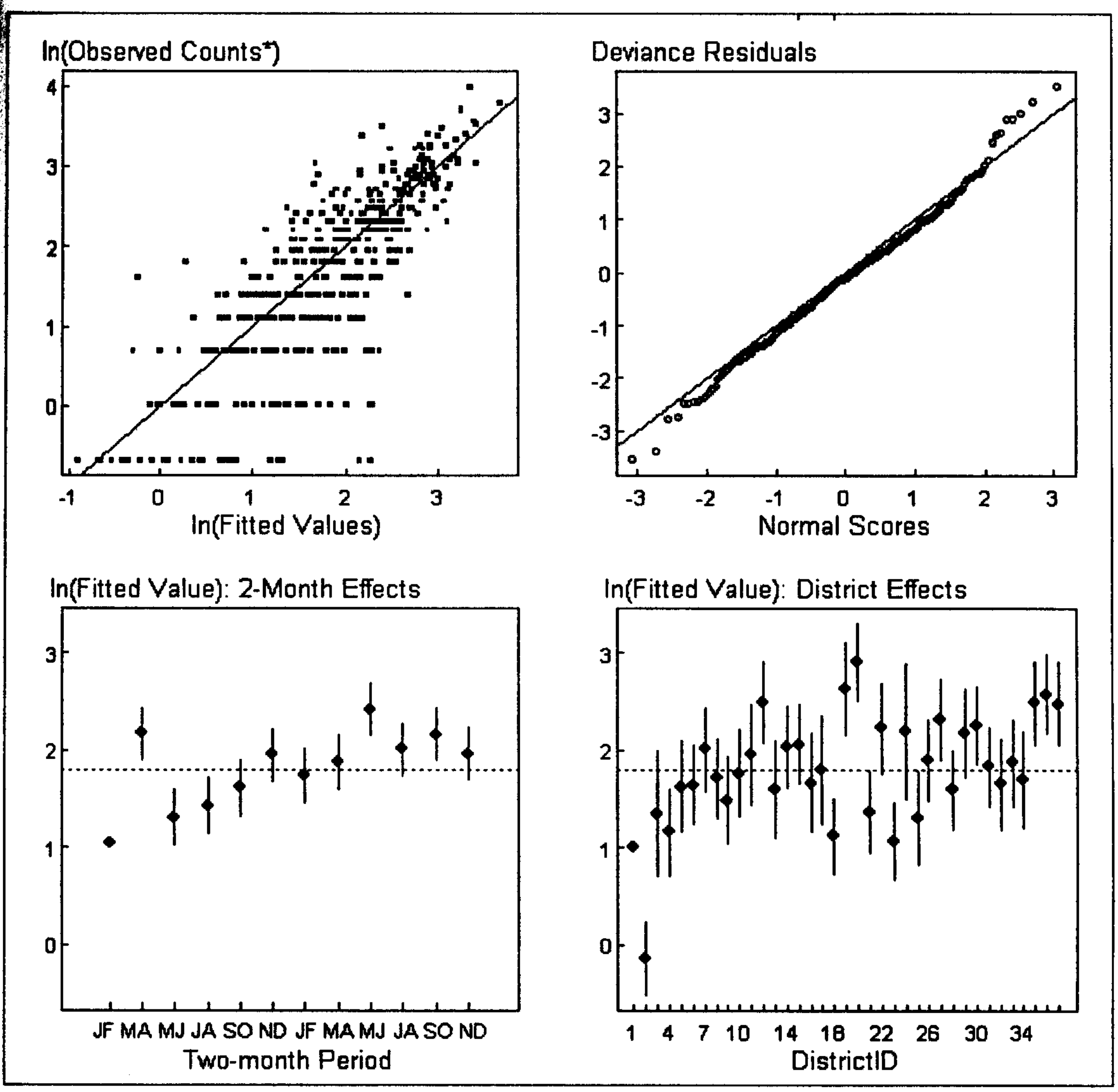

Figure 3.4: Results from fitting negative binomial model to event rates in districts in 2-month periods (* Zero counts are replaced by 0.5 to avoid logarithms of 0)

The period effects are listed in Table 4.2. They show a steadily increasing trend in the rate during 2004 stabilizing in 2005, with the exception of a high peak in March-April 2004 and a less pronounced peak in May-June 2005. The district effects confirm the pattern seen in the left panel of Figure 3.3. Four district in Songkla has a comparatively low value as expected, given that on 14 events were reported in this district during the 2-year period: the next highest is three district in Songkla with 74 reported events.

District ID	District Name	Estimate	St. Error
9003	Chana	0	-
9004	Na Thawi	-1.149	0.328
9005	Thepha	0.339	0.223
9006	Saba Yoi	0.148	0.236
9401	Muang Pattani	0.621	0.205
9402	Khok Pho	0.640	0.216
9403	Nong Chik	0.998	0.206
9404	Panare	0.709	0.227
9405	Mayo	0.476	0.227
9406	Thung Yang Daeng	0.754	0.264
9407	Sai Buri	0.952	0.211
9408	Mai Kaen	1.482	0.254
9409	Yaring	0.593	0.213
9410	Yarang	1.021	0.204
9411	Ka Pho	1.056	0.260
9412	Mae Lan	0.658	0.285
9501	Muang Yala	0.786	0.199
9502	Betong	0.114	0.241
9503	Bannangsata	1.621	0.204
9504	Than To	1.902	0.216
9505	Yaha	0.356	0.237
9506	Raman	1.219	0.202
9507	Ka Bang	0.062	0.356
9601	Muang Narathiwat	1.189	0.239
9602	Tak Bai	0.297	0.212
9603	Bacho	0.889	0.211
9604	Yi-ngo	1.299	0.209
9605	Rangae	0.587	0.231
9606	Ruso	1.161	0.202
9607	Si Sakhon	1.252	0.205
9608	Waeng	0.825	0.239
9609	Sukhirin	0.647	0.227
9610	Su Ngai Ko Lok	0.860	0.248
9611	Sungai Padi	0.693	0.214
9612	Chanae	1.476	0.203
9613	Cho-I-rong	1.570	0.213

Table 4.1: District effects and standard errors based on negative binomial generalized linear model (Four district of Songkla taken as baseline level)

Period	Estimate	St. Error
Jan-Feb 2004	0	-
Mar-Apr 2004	1.118	0.136
May-Jun 2004	0.260	0.148
Jul-Aug 2004	0.380	0.146
Sep-Oct 2004	0.560	0.143
Nov-Dec 2004	0.899	0.138
Jan-Feb 2005	0.678	0.141
Mar-Apr 2005	0.821	0.139
May-Jun 2005	1.369	0.133
Jul-Aug 2005	0.955	0.137
Sep-Oct 2005	1.110	0.136
Nov-Dec 2005	0.914	0.138

Table 4.2: Period effects and standard errors based on negative binomial generalized linear model (Jan-Feb 2004 taken as baseline level)