Queensland University of Technology
Brisbane Australia

This is the author's version of a work that was submitted/accepted for publication in the following source:

Barnett, Adrian
(2014)
Analysis of death data during the Morwell mine fire.
[Working Paper]
(Unpublished)

This file was downloaded from: http://eprints.qut.edu.au/76230/

© Copyright 2014 Please consult the author

Notice: *Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source:*

Analysis of death data during the Morwell mine fire

Introduction

This document explains my analysis of the Morwell mine fire data. I have tried to give as much technical detail as possible whilst still making it understandable to the non-specialist reader.

I am happy for this document to be freely shared. I am also happy to answer further questions via e-mail: a.barnett@qut.edu.au.

Methods

Data

The data were monthly numbers of deaths from 2009 to 2014 for the months of January to June. The deaths were split by four postcodes (3840, 3842, 3825 and 3844) according to usual place of residence. The six years, six months and four postcodes gives 144 observations. In total there were 1,811 deaths.

Statistical model

I used a regression model to examine the key hypothesis of whether deaths rates were higher during the two months of the fire.

I give the model as an equation below and then explain each line of the equation.

$$\begin{aligned}
d_{i,t} &\sim \text{Poisson}(\mu_{i,t}), \qquad i = 1, \dots, 4,\ t = 1, \dots, 36, \\
\log(\mu_{i,t}) &= \log(\text{pop}_t/10000) + \alpha_0 + \text{trend}_t + \text{season}_t + \text{postcode}_i + \text{fire}_t, \\
\text{trend}_t &= \alpha_1 t, \\
\text{season}_t &= \alpha_2 \cos\left(\frac{2\pi(\text{month}_t - 1)}{12}\right) + \alpha_3 \sin\left(\frac{2\pi(\text{month}_t - 1)}{12}\right), \\
\text{postcode}_i &\sim N(0, \sigma^2) \\
\text{fire}_t &= \begin{cases} \alpha_4, & \text{if year} = 2014 \text{ and month} = 2, 3, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}$$

The first line says that the deaths from postcode i at time t are modelled as a Poisson distribution, which is the most appropriate distribution for count data. There are four postcodes and 36 times.

The second line is the regression model, it includes the population at time t (divided by 10,000) as an offset which is used to account for the region's growing population. This population data is for LaTrobe City Council which includes other postcodes outside the four in the death data. Ideally I would have had population data for each individual postcode, but I've assumed that the influx and outgoings of people in these four postcodes over time mirrors the patterns for the wider council area. In a sensitivity analysis I removed the population data and it had little impact on the results.

The regression equation uses a log-link which means the model is multiplicative and hence gives results as death rates rather than numbers. The overall mean death rate is modelled by α_0 (labelled as the intercept in the tables below). A linear trend in death rates is modelled by α_1 to control for the expected small reduction in death rates over 2009 to 2014.

Deaths in Australia are strongly seasonal with a winter peak. To model this I have include a annual sinusoidal model based on the month in time t.

To adjust for any differences in death rates between postcodes I included a random effect using a Normal distribution with a zero mean. This allows deaths rates to be higher or lower in some postcodes and constrains the differences to follow a Normal distribution.

The effect of the fire is modelled using a simple change in death rates during February and March 2014 compared with all other months.

The absolute number of deaths was estimated using: $\overline{d}[\exp(\alpha_4) - 1]$, which is the mean number of monthly deaths per postcode multiplied by the relative change in deaths.

In an alternative model I included a term for temperature: $\alpha_5\text{temperature}_t$, where temperature_t is the maximum monthly temperature from the Bureau of Meteorology. This adjustment is added because we know that high temperatures increase the risk of death. Ideally I would have used daily temperature data to give a finer adjustment, but this would also require daily death data.

The model was fitted using a Bayesian paradigm as this allowed me to easily estimate the probability that there was an increase in the death rate: $Pr(\alpha_4 > 0)$.

The plots and tables were created using the R software (`www.r-project.org`) and the Bayesian model was fitted using JAGS (`mcmc-jags.sourceforge.net`).

Results

Plots

Looking at the total figures, the deaths in 2014 in February and March do appear to be high. Another year with high deaths rates is 2009 and this may be due to bushfires and extreme heat that summer.

The differences in numbers on the y-axes between panels are because some suburbs are larger than others.

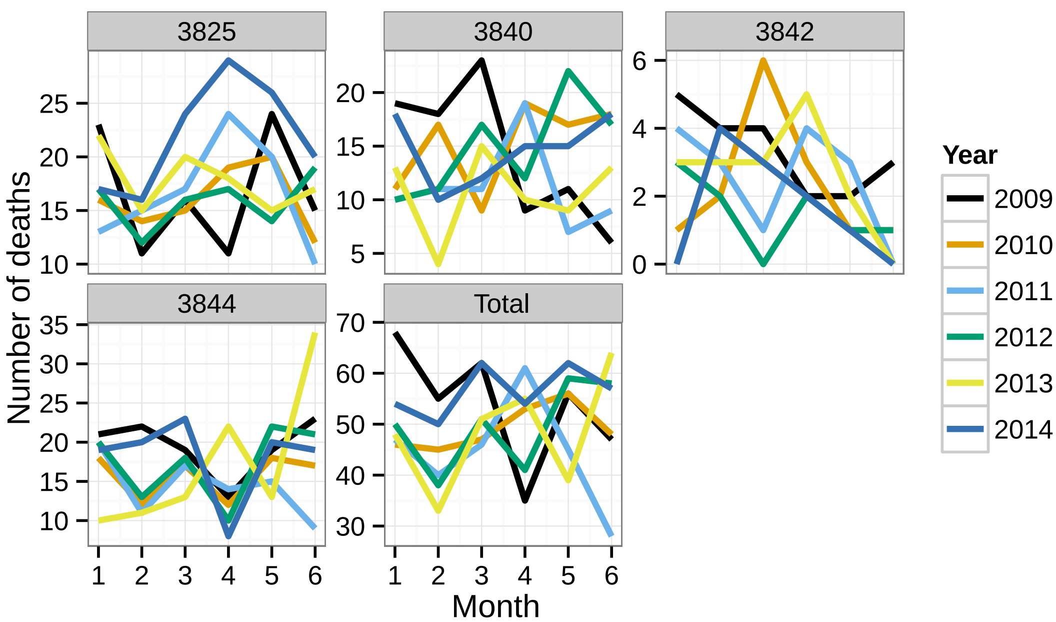

Figure 1: Deaths numbers by month and year in each postcode and the overall number of deaths. The scales on the y-axes differ between postcodes.

Statistical model results

Table 1: Estimates without adjusting for temperature. Statistics are the mean, standard deviation and lower and upper 95% credible interval. Estimates are on a log scale except for the relative risks and absolute number of deaths.

	mean	SD	lower	upper
Intercept	0.30	0.06	0.17	0.42
Trend	0.00	0.01	−0.03	0.03
Postcode 1	0.57	0.04	0.49	0.66
Postcode 2	0.31	0.05	0.22	0.40
Postcode 3	−1.43	0.08	−1.60	−1.27
Postcode 4	0.55	0.04	0.46	0.63
Season, cos	−0.04	0.04	−0.12	0.04
Season, sin	−0.02	0.08	−0.17	0.14
Fire	0.13	0.11	−0.08	0.34
Fire, relative risk	1.14	0.12	0.92	1.41
Absolute deaths	1.82	1.57	−1.02	5.10

The probability that the death rate was higher than the average during the fire is 0.89. This means that the probability that the death rate was not higher than the average during the fire is 0.11. The mean increase in deaths is as a relative risk is 1.14, or 14 as a percentage. The absolute number of deaths per postcode per month is 1.8, which over 4 postcodes and 2 months is 14.4.

Table 2: Estimates after adjusting for monthly temperatures. Statistics are the mean, standard deviation and lower and upper 95% credible interval. Estimates are on a log scale except for the relative risks and absolute number of deaths.

	mean	SD	lower	upper
Intercept	0.30	0.06	0.18	0.42
Trend	0.00	0.01	−0.03	0.03
Postcode 1	0.57	0.04	0.49	0.66
Postcode 2	0.31	0.05	0.22	0.40
Postcode 3	−1.43	0.08	−1.59	−1.27
Postcode 4	0.55	0.04	0.46	0.63
Season, cos	−0.16	0.15	−0.46	0.13
Season, sin	−0.01	0.08	−0.16	0.15
Fire	0.09	0.11	−0.13	0.32
Fire, relative risk	1.11	0.13	0.87	1.37
Absolute deaths	1.34	1.60	−1.58	4.71
Temperature	0.02	0.02	−0.02	0.06

The probability that the death rate was higher than the average during the fire is 0.80. The mean increase in deaths is as a relative risk is 1.11, or 11 as a percentage. The absolute number of deaths per postcode per month is 1.4, which over 4 postcodes and 2 months is 11.2.

EXP.0005.001.0006

The reduction in the risk of the fire and the death numbers after adjusting for temperature is plausible as we know that high temperatures can kill. High temperatures and high levels of air pollution can interact to produce greater combined risks than when only one exposure is present.

The figures in the first released analysis quoted 11 deaths rather than 14. This is because the request to present absolute deaths was made after the request to adjust for temperature.