Does rainfall increase or decrease motor accidents

A reflection on the good, the bad and the ugly (in statistics)

Gráinne McGuire
Taylor Fry Consulting Actuaries
Contents

1. Introduction
2. Data
3. Modelling with rainfall
   - Monthly
   - Daily
4. Reflection on the use of statistics
Introduction

• Does rain increase the number of motor accidents?
• The role of statistics in getting a good answer
• Discussion of both of these points previously presented in Davies et al (2004)
We are not talking about...
Nor are we talking about...
We discuss Australian (Perth) roads, drivers and conditions
Data

• Data
  – CTP claims from accidents in Perth from July 1993 to December 2005
    • Accident date and time
  – Vehicle registrations
  – Monthly and daily rainfall data from Perth weather stations

• Manipulation
  – Match accident and rainfall days
    • Days defined to begin at 9am
Claim frequency – modelling with rainfall

Monthly claim frequency

Rainfall
Results after removal of trend

Fitting of trend

Residuals
Regression on rainfall

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Estimate</th>
<th>Std error</th>
<th>t-value</th>
<th>Significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.0132</td>
<td>0.0029</td>
<td>-4.5413</td>
<td>***</td>
</tr>
<tr>
<td>Rainfall</td>
<td>0.0002</td>
<td>0.0000</td>
<td>6.3744</td>
<td>***</td>
</tr>
<tr>
<td>R^2</td>
<td>22%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>148</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-value</td>
<td>40.6330</td>
<td></td>
<td></td>
<td>***</td>
</tr>
</tbody>
</table>

Rainfall has small but significant effect – higher rainfall $\rightarrow$ higher frequency.

But low R^2
Is rainfall truly an explanatory variable?

• Is it possible that rainfall is significant simply because it acts as a proxy for a seasonal effect
• If we include daylight hours in regression, rainfall is no longer significant

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Estimate</th>
<th>Std error</th>
<th>t-value</th>
<th>Significant?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.1213</td>
<td>0.0267</td>
<td>4.5345</td>
<td>***</td>
</tr>
<tr>
<td>Rainfall</td>
<td>3.0E-05</td>
<td>5.0E-05</td>
<td>0.6765</td>
<td></td>
</tr>
<tr>
<td>Daylight</td>
<td>-0.0102</td>
<td>0.0020</td>
<td>-5.0554</td>
<td>***</td>
</tr>
<tr>
<td>R^2</td>
<td>33%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>147</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-value</td>
<td>36.4664</td>
<td></td>
<td></td>
<td>***</td>
</tr>
</tbody>
</table>
Is this the end?

- The monthly normal linear regression analysis does not support a rainfall effect.
- End of story?
- Maybe not – are we approaching the problem correctly?
- Lies, damn lies and statistics
Good modelling

1. **Why use monthly data?**
   - What does rain on 1\textsuperscript{st} March have to do with accidents on 31\textsuperscript{st} March?
   - Why doesn’t rain on 31\textsuperscript{st} March have any bearing on accidents on 1\textsuperscript{st} April
   - Suggests use of data on a finer scale – e.g. daily

2. **Accidents = count data.**
   - Poisson error distribution is preferable to normal

Analysis is similar to that in Eisenberg (2004)
Food for thought – Eisenberg (2004)

- Analysis based on American motor accident data
- Monthly analysis showed an inverse relation with rain
  - More rain, less accidents
- Daily analysis demonstrated two opposing rainfall effects
  - Primary: rain on a particular day leads to more accidents that day
  - Secondary: rain on previous days means fewer accidents. May be due to cleaner roads or more careful drivers.
- What will analysis of Perth daily data show?
Model setup

- Over-dispersed Poisson GLM
- Covariates include:
  - Accident month: to remove the overall downward trend
  - Month of year: to capture annual seasonal effects
  - Day of week: e.g. Fridays different to Sundays
  - Daily rainfall: both rainfall on the accident day and rainfall in the past (represented here by rain 2 days before the accident)
Rainfall results from daily modelling

Primary
On day of accident

Secondary
Past rainfall
Statistics: the good, the bad and the ugly
The bad

- Trying to explain the overall reducing trend using regression techniques and various explanatory factors.
  - Example here regresses frequency on the author’s age. Pretty good fit but meaningless in terms of causation.
- Similar results from any monotonic sequence including sensible ones like fleet average age, multi-vehicle ownership, but does it mean anything.
- Correlation does not equal causation.
The ugly

• Accident numbers are count data
  – Use of normal error distribution is inappropriate
  – Could model log(accident numbers/frequency); a log normal model. But still incorrect and it requires a bias correction.
The good

- Modelling daily rather than monthly
- Using an appropriate error distribution
- We can be more confident that rainfall is the cause of the “rainfall” effects
  - But never 100% sure……
References


To conclude

- Rainfall results are interesting in themselves
- The problem is a good example of the importance of getting the level of data detail correct

*All models are wrong, some are useful*