Thriving on Change
16th General Insurance Seminar
Institute of Actuaries of Australia
9-12th Nov 2008
Hyatt Regency Coolum

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…

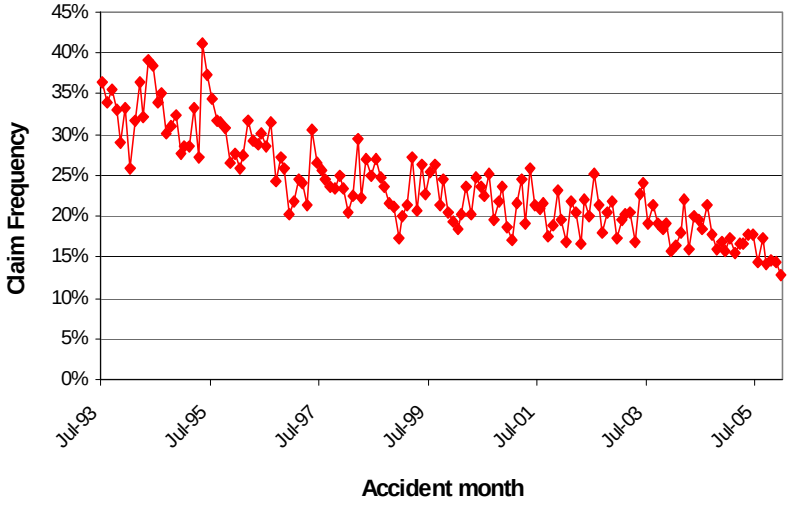
Nor are we talking about…

We discuss Australian (Perth) roads, drivers and conditions

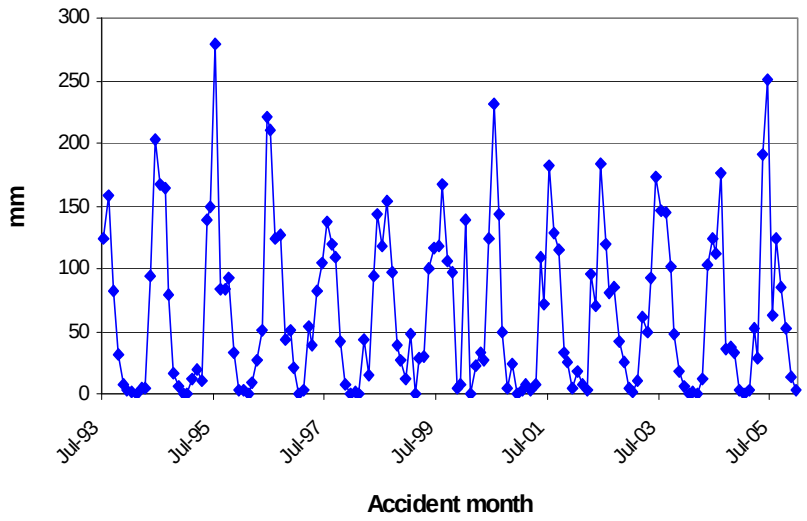
Data

- 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

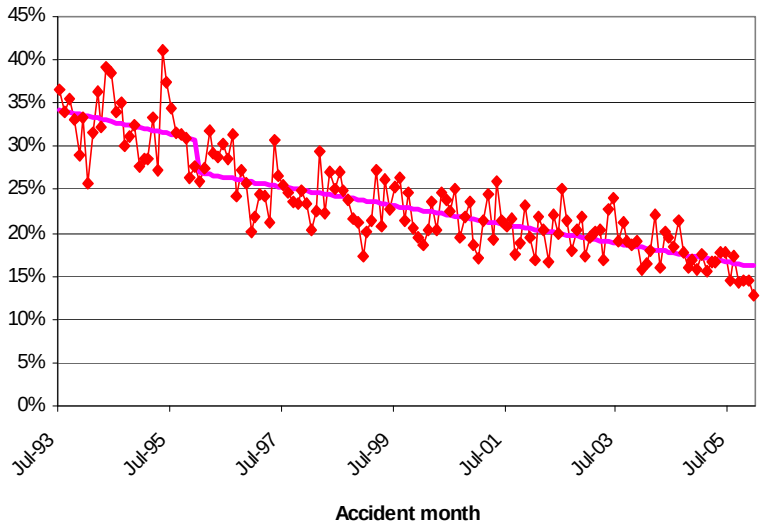


Monthly claim frequency

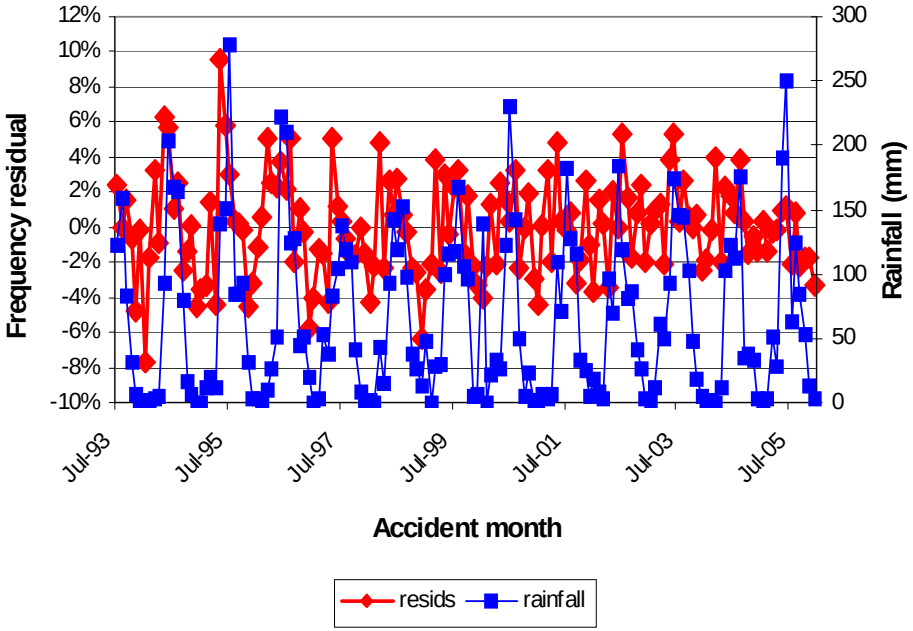


Rainfall

Results after removal of trend



Fitting of trend



Residuals

Regression on rainfall

Quantity	Estimate	Std error	t-value	Significant?
Intercept	-0.0132	0.0029	-4.5413	***
Rainfall	0.0002	0.0000	6.3744	***
R^2	22%			
df	148			
F-value	40.6330			***

Rainfall has small but significant effect – higher rainfall → 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

Quantity	Estimate	Std error	t-value	Significant?
Intercept	0.1213	0.0267	4.5345	***
Rainfall	3.E-05	5.E-05	0.6765	
Daylight	-0.0102	0.0020	-5.0554	***
R^2	33%			
df	147			
F-value	36.4664			***

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 1st March have to do with accidents on 31st March?
   - Why doesn't rain on 31st March have any bearing on accidents on 1st 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)

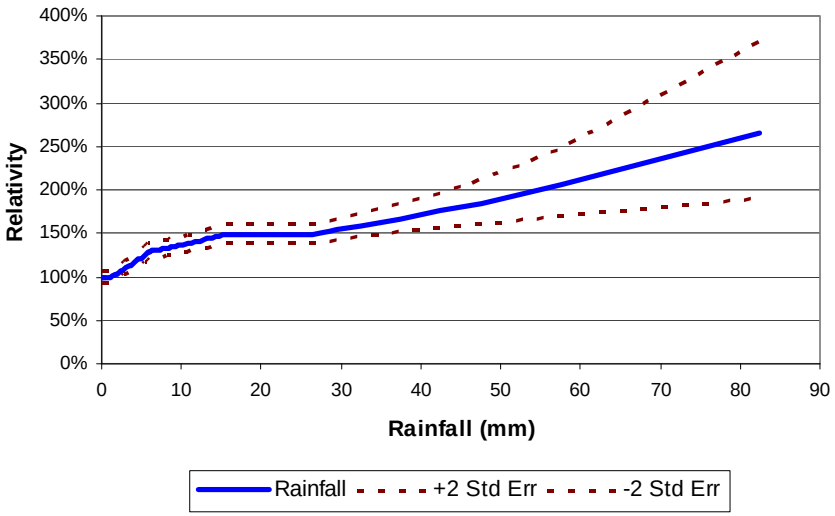
Food for thought – Eisenberg (2004)

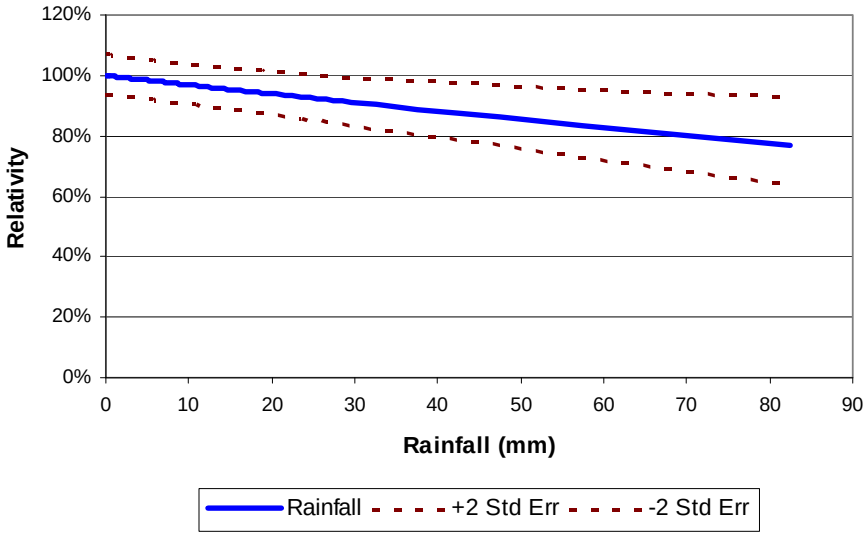
- Analysis based on American motor accident data
- Monthly analysis showed an inverse relation with rain
  - More rain, less accidents
- Daily analysis demonstrated two opposing rainfall effects
  - Primary: rain on a particular day leads to more accidents that day
  - Secondary: rain on previous days 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

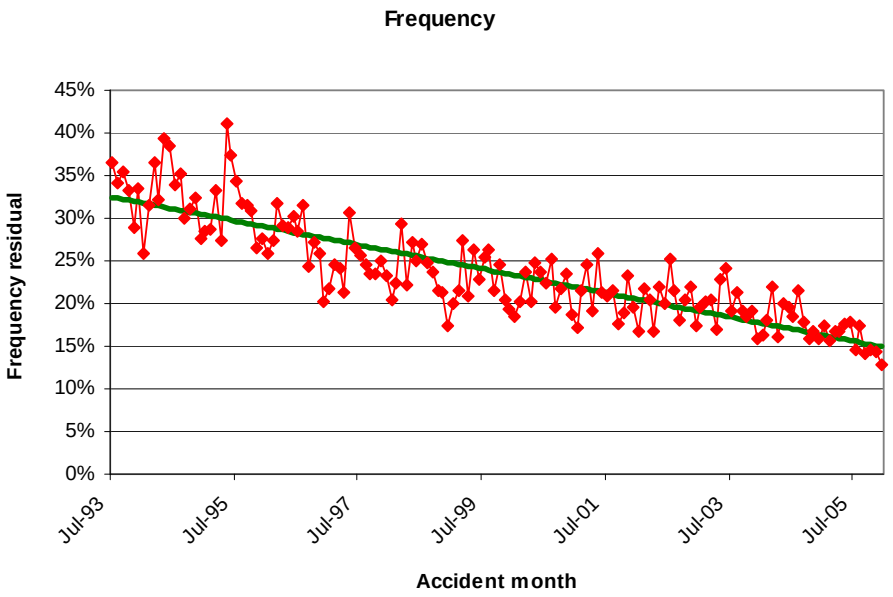
Secondary

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

- Davies, R., Winn, R. and Jiang, J. (2004). Determinants of claim frequency in CTP schemes. Accident Compensation Seminar, 2004, Institute of Actuaries of Australia.
- Eisenberg, D. (2004). The mixed effects of precipitation on traffic crashes. Accident Analysis and Prevention, 36, 637-647.

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