CIHT SoRSA

Road Crash Prediction Statistical Modelling Approaches: what could go wrong

Dr Jwan Kamla

About Me

Where I came from originally - Shaqlawa / Kurdistan

My current home town- Nottingham

Jwan Kamla

PhD, MSc, BSc, CEng, FCIHT, MSoRSA, NH CoC

Senior Road Safety Engineer at Jacobs

SoRSA Committee member

CIHT East Midland's Vice Chair

CIHT Mentor and Reviewer

Member of Professional Review Panel (CIHT)

CIHT

Agenda

- History of statistical modelling and what could go wrong
- Case study
- Statistical measurements

Why Collision Prediction Modelling

Who, What, Where, When and Why

Fridstrom et al. (1995) state that it is impossible to predict collisions (where, when, and by whom a collision occurred)because in nature collisions are random.

- Proactive rather than reactive [Safe System Approach]

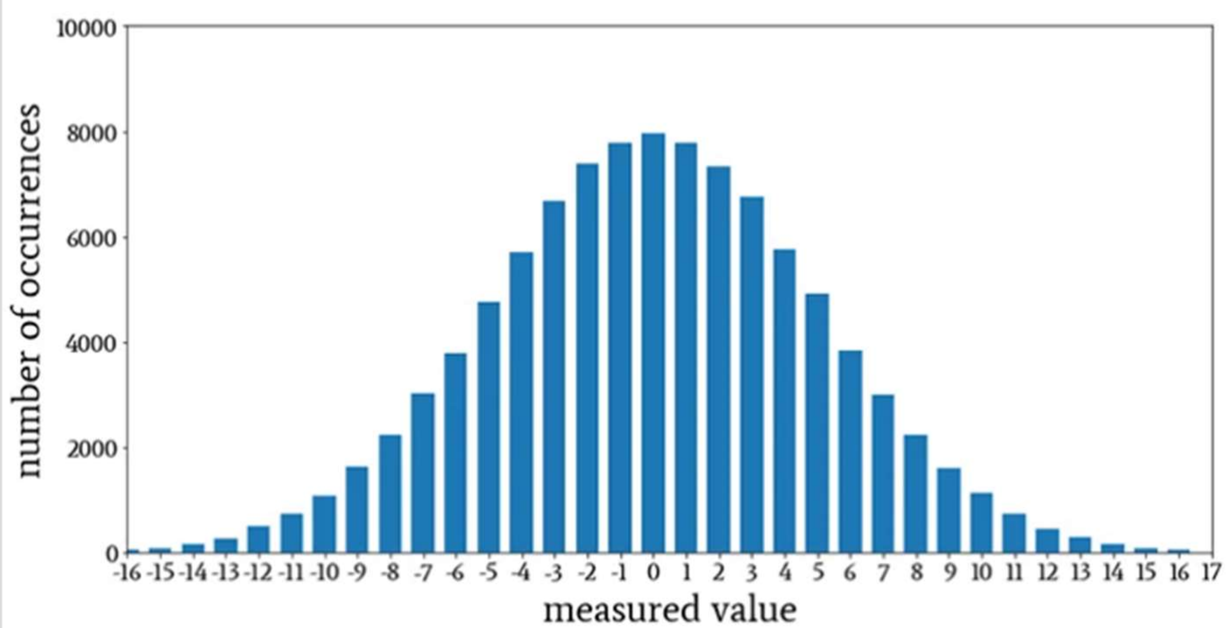
Source: FRIDSTRØM, L., IFVER, J., INGEBRIGTSEN, S., KULMALA, R. & THOMSEN, L. K. 1995. Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. Accident Analysis & Prevention, 27, 1-20.

History of Collision Prediction

- Linear Regression

$$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$$

Linear constituent



Source: https://www.allaboutcircuits.com/technical-articles/normal-distribution-understanding-histograms-probability/

Collision data is a variable that is "sporadic, discrete, and non-negative", and the data distribution is more similar to a Poisson distribution

Continued

- Poisson Regression
  - **Variance equal to mean for incident data**

- **Negative Binomial (NB) Distribution, zero inflated NB, Gamma NB, etc..**
  - **Variance is either greater or less than mean for incident data**

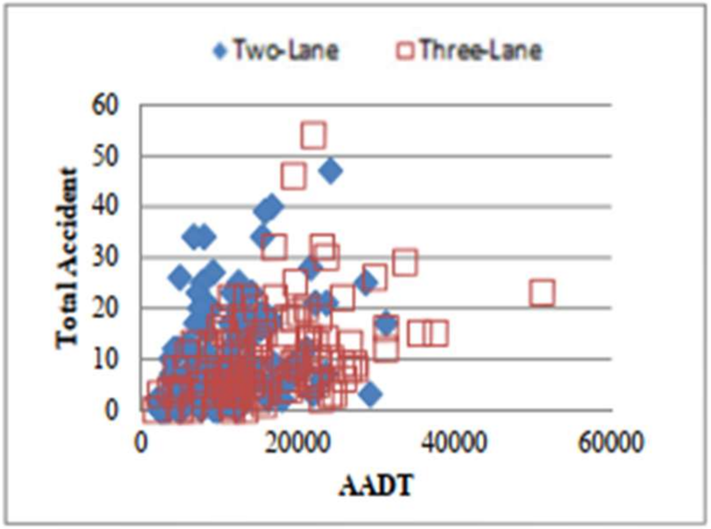
Variable	Min	Max	Mean	Variance
Accident and HBI characteristics (Dependent variable)				
11 -year total accidents numbers	5	170	60.50	2061
11 -year Truck accidents numbers	0	54	14.10	199
2 -year HBIs numbers	0	764	152.6	32472

Source: Kamla J. 2016. Analysing Truck Position Data to Study Roundabout Accident Risk. PhD Thesis, University of Nottingham.

What I often see

A Figure showing linear relationship between e.g., number of vehicles using emergency services vs AADT

- or Visibility at junction's vs collision data
- or Accident vs AADT , etc..

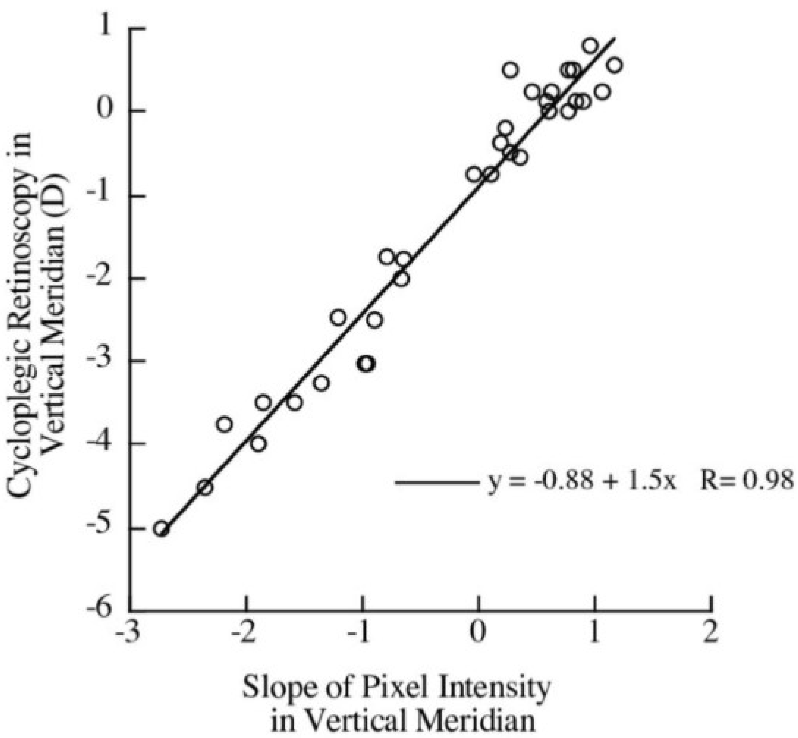


R^2=0.14 two-lane, R^2=0.17 three-lane

Source: Kamla J. 2016. Analysing Truck Position Data to Study Roundabout Accident Risk. PhD Thesis, University of Nottingham.

CIHT

What could go wrong



- Application of linear regression to model count data (incidents) in the transport- related phenomenon is not appropriate.
- Count data is non-negative and linear regression can predict values with negative outcome
- On the left is an example where Y has a negative value
  - ✓ **We cant have negative collision numbers**

Source: Elise N. Harb, 2005, CHARACTERISTICS OF ACCOMMODATIVE BEHAVIOR DURING SUSTAINED PERIODS OF NEAR WORK IN HUMANS AND PRIMATES, MSc thesis, The New England College of Optometry.

Continued

We cannot have a model with just geometric variables and another model with just traffic variables (although there are people that do this, it is not correct because you can easily show that parameter estimates will be biased making any conclusions incorrect)

Example:
where Q is flow and A is accident

$$A_w = 1.577 \times 10^{-2} \times Q_w^{0.76}$$

$$A_c = 9 \times 10^{-4} \times Q_w^{0.912}$$

$$A_a = 0.83 \times 10^{-2} \times Q_a^{0.75}$$

Source: Kamla J. 2016. Analysing Truck Position Data to Study Roundabout Accident Risk. PhD Thesis, University of Nottingham.

Literature NB Regression

"Maycock and Hall (1984), Hauer et al. (1988), Brude and Larsson (1993), Bonneson and McCoy (1993), Miaou (1994), Shankar et al. (1995), Poch and Mannering (1996), Milton and Mannering (1998), Karlaftis and Tarko (1998), Carson and Mannering (2001), Miaou and Lord (2003), Lord et al. (2005), El-Basyouny and Syed (2006), Lord (2006), Kim and Washington (2006), Lord and Mahlawat (2009), Malyshkina and Mannering (2010), Daniels et al. (2010), Cafiso et al. (2010), Geedipally and Lord (2010), Lao et al. (2011), Geedipally and Lord (2011), Lord and Kuo (2012), Meng and Qu (2012), Ye et al. (2013), Aryuyuen and Bodhisuwan (2013), Qin et al. (2013), Vangala et al. (2015), Rahman Shaon and Qin (2016), Naznin et al. (2016), Qin et al. (2016)"

Source: Lord D, Qin X, Greedipally R.S 2021. Highway Safety Analytics and Modelling, Elsevier. ISBN: 978-0-12-816818-9.

Case study

The University of Nottingham

Analysing Truck Position Data to Study Roundabout Accident Risk

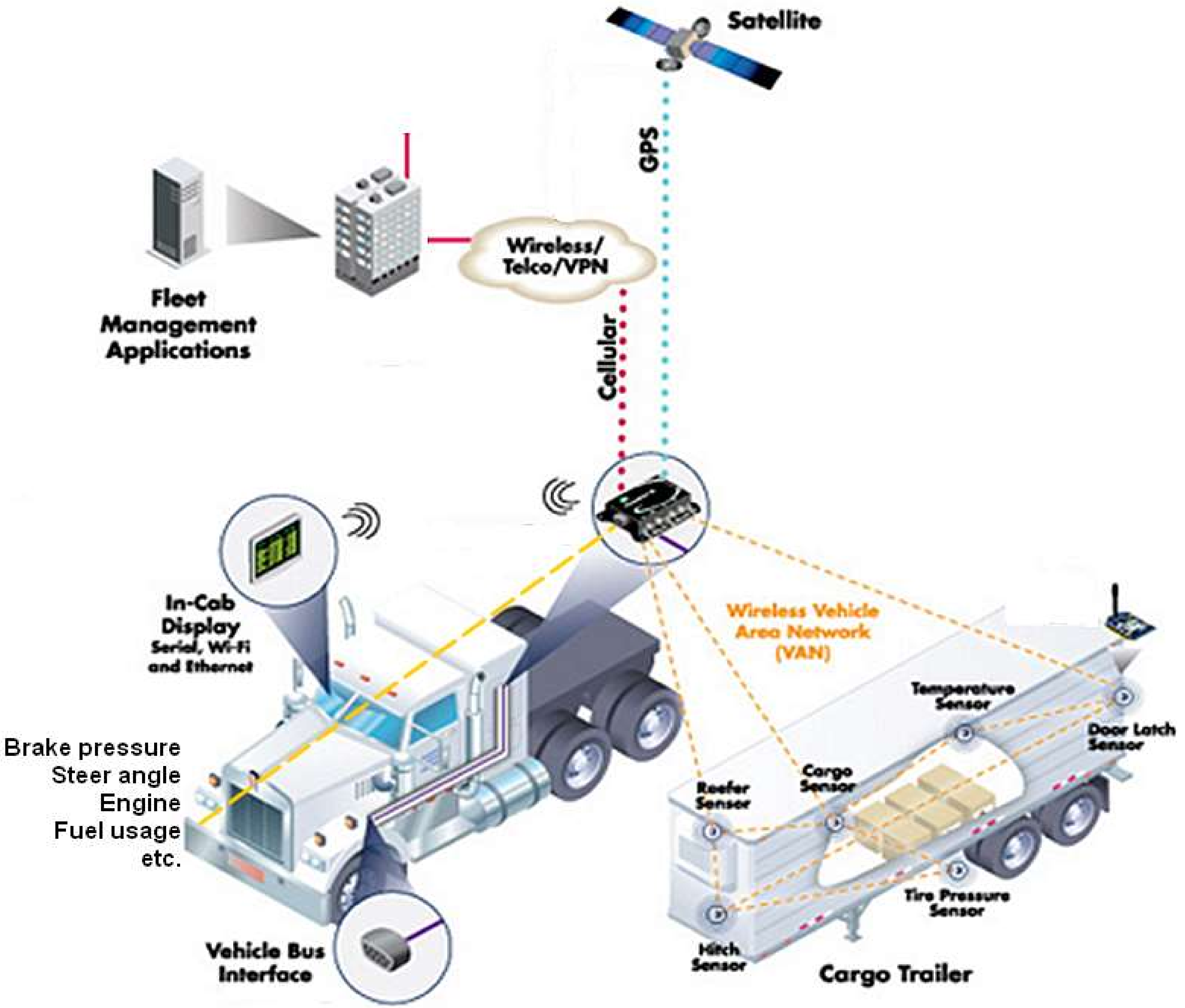
By

Jwan Jameel Shekh Mohammed Kamla

A Doctoral Thesis

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

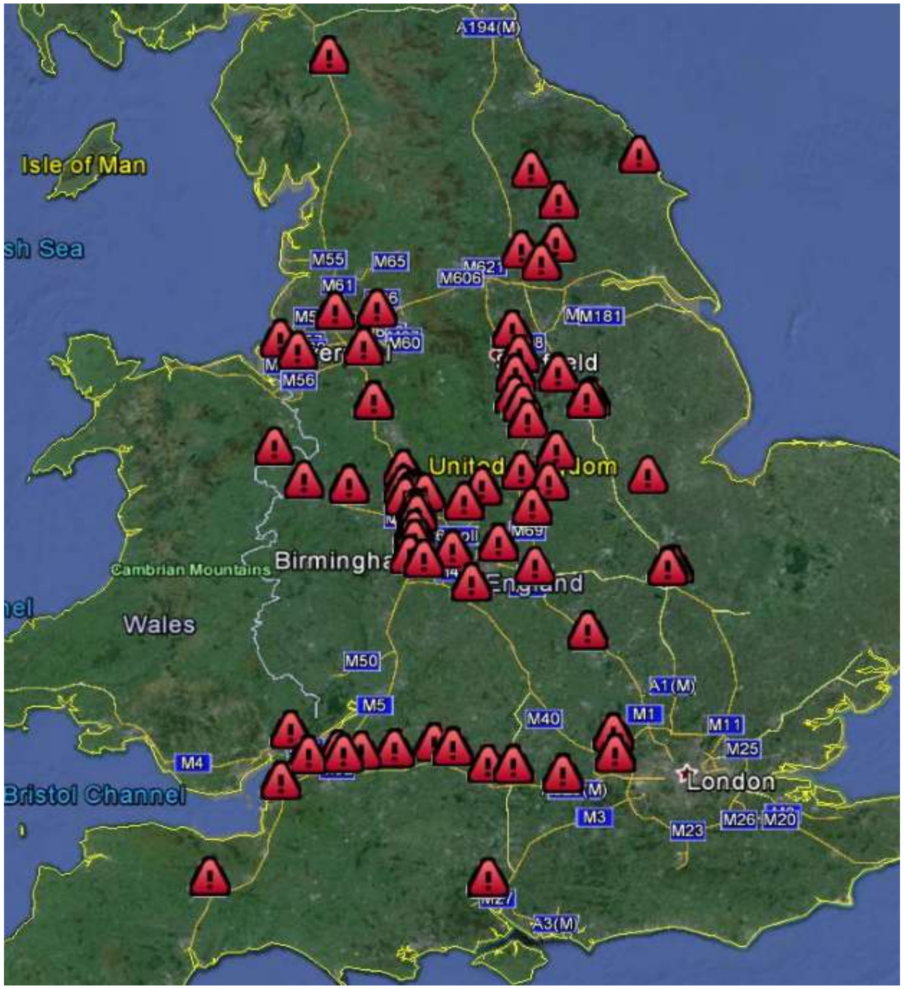
October 2016



Modern truck CAN bus collects and can transmit multiple sensor data via cell-phone network to fleet managers

CIHT

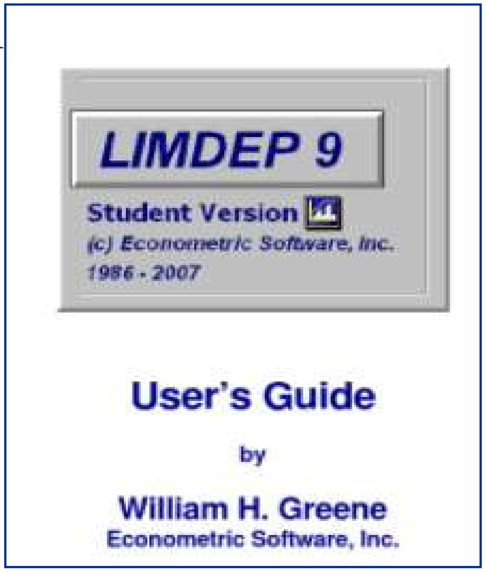
Introduction



- Selected 70 roundabouts (284 approaches)
- Selected Harsh Braking Incidents (HBIs) and Collisions within a 350m radius from the center of the roundabouts
- Used Random Parameter Negative Binomial Model

Method

- ✓Both random- and fixed-parameters NB approach used to predict HBIs, and collisions



- ✓Random parameters used to account for
- ✓ unobserved heterogeneity
  - ✓ i.e. independent variables that may change across the road segment or intersections including roundabouts
- ✓Marginal effects were computed
  - ✓ to give the change in the number of collisions given a unit change in any independent variable

Linear Regression Vs Negative Binomial

Using Random Parameter Negative Binomial model:

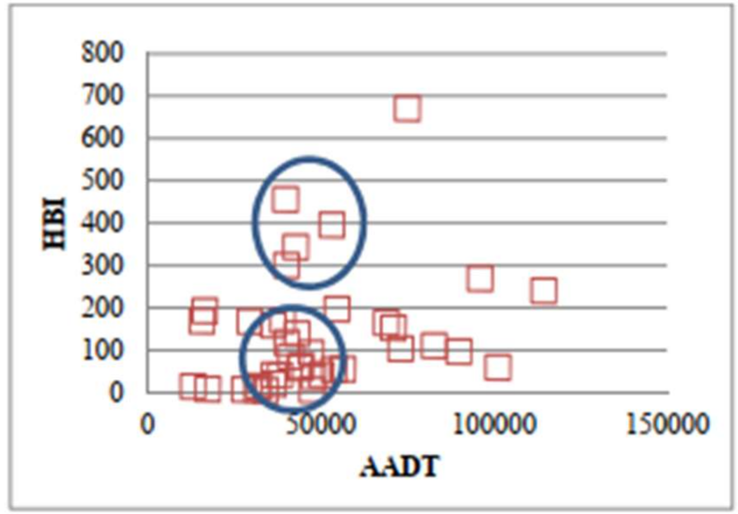
Traffic variables was found to have significant positive influence on the number of e.g. harsh braking incidents similar results identified for collision data

Truck HBI Model Estimation Results.

Roundabout category	Variables	NB Random-parameters model Coefficient	t-stat
Whole roundabout	**Constant**	−11.36	−4.80***
	Geometric characteristics		
	Arm number (1 if 3 arm;0 otherwise)	0.064	0.224
	SD	*1.117*	*3.982****
	Circulatory lane width (m)	−0.182	−2.912***
	Entry width (m)	0.213	2.937***
	Traffic signal (1 if signal;0 otherwise)	−0.145	−0.492
	SD	*0.945*	*5.818****
	Traffic signal (1 if un-signal;0 otherwise)	−0.017	−0.069
	SD	*0.842*	*4.574****
	Traffic Characteristics		
	ln(AADT)	1.37	6.112***
	Percentage AADTT	0.14	4.463***
	Dispersion parameter	1.81	5.448***
	Observation numbers	70	
	Log-likelihood at constant only	−407.4612	
	Log-likelihood at convergence	−396.8231	

*** At 99% significance level.

Linear Regression model



CIHT

Some Results

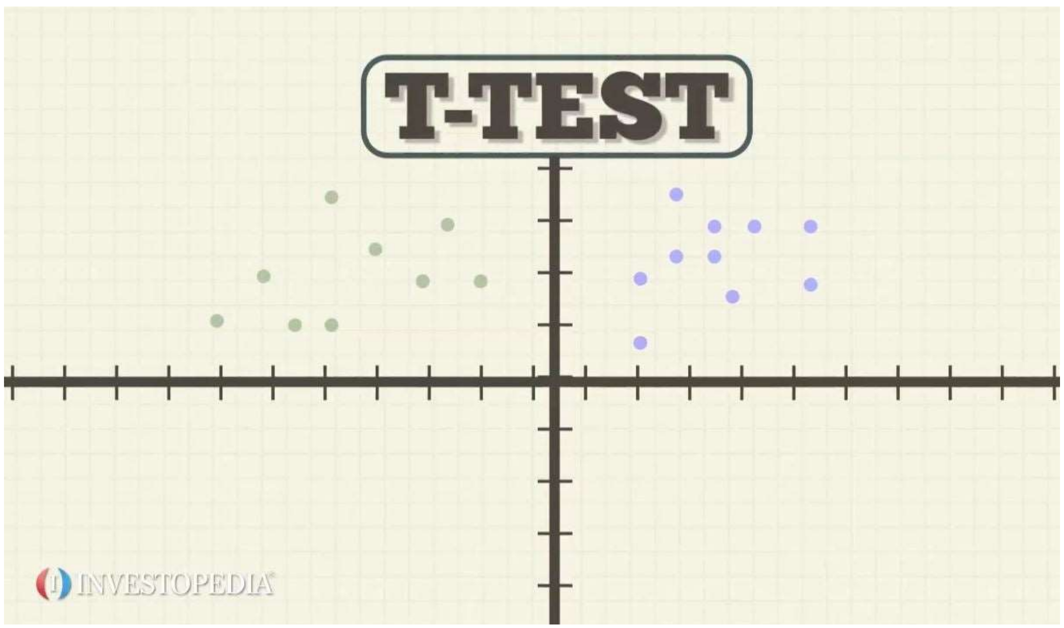
Variables that found to have ***negative*** *influence on incidents:*

- *Circulatory roadway width*
- *Three-arm roundabouts compared to six*

Variables that found to have ***positive*** *influence on incidents:*

- *AADT*
- *% of HGVs*
- *Partially signalised roundabouts compared to signalised and un-signalised*
- *Inscribed circle diameter*
- *Entry width*
- *Two-lane approaches compared to three lane*

Statistical Measurements



$$R^2$$

$$\chi^2$$

Source: Google

Regression Coefficient R-squared test (R^2)

Table 4-16 ANOVA Results for Total and Truck Accidents with AADT Based on Different Geometric Factors within Circulatory Lanes

Factor	Total Accident with AADT			Truck Accident with AADT		
	R^2	*p*-value	Sig	R^2	*p*-value	Sig
Two-lane	0.39	0.000	yes	0.25	0.001	yes
Three-lane	0.03	0.325	no	0.04	0.273	no
Signalised	0.03	0.460	no	0.28	0.586	no
Un-signalised	0.04	0.275	no	0.011	0.583	no
Partially signalised	0.11	0.166	no	0.016	0.019	yes
Grade-separated	0.12	0.014	yes	0.10	0.022	yes
At-grade	0.09	0.211	no	0.22	0.041	yes

Conclusions can't be made based on low R^2

CIHT

How to compare before and after studies?

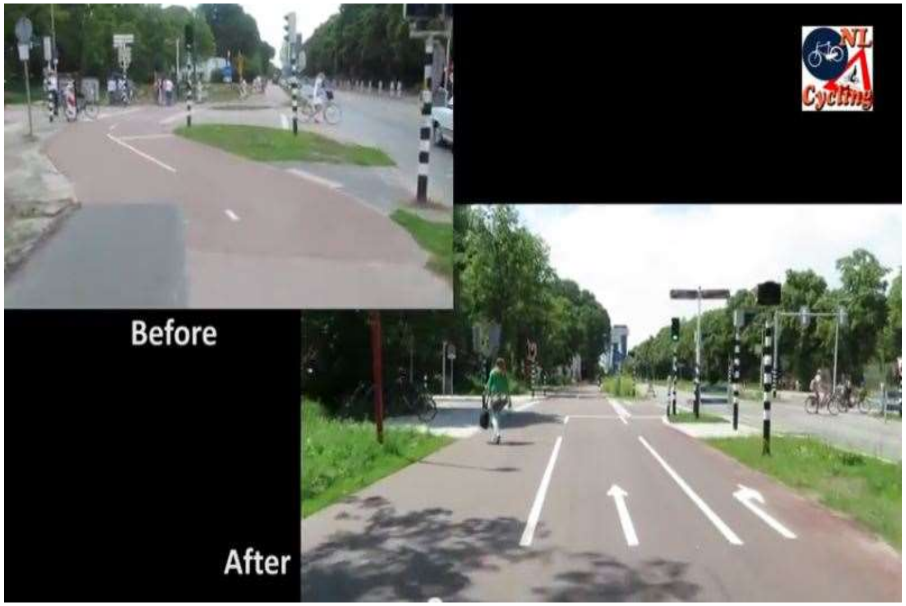
Source: Google

Answer

- You may have seen or have been trained to use chi-square test χ^2
- A chi-square test is used to help determine if observed results are in line with expected results .. usually useful for survey or questionary data (categorical variables).
- For before and after engineering studies after applying treatment measures best to use

Paired T-test to identify if results are different from each other

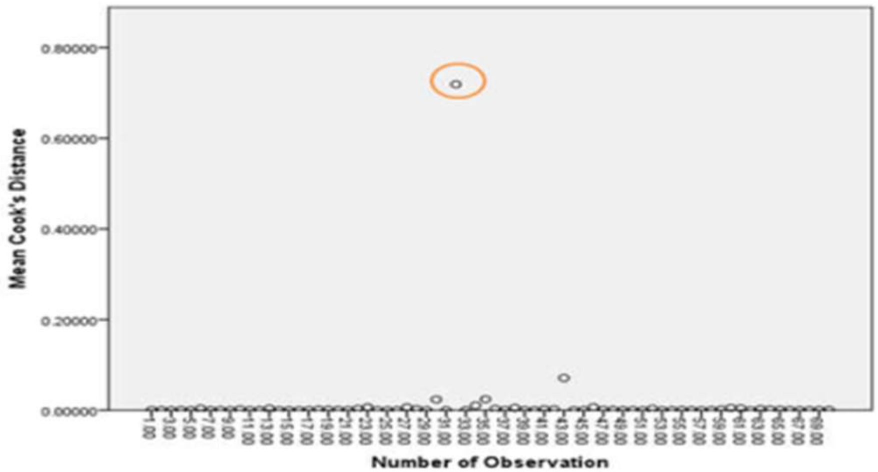
This can be easily identified in Excel

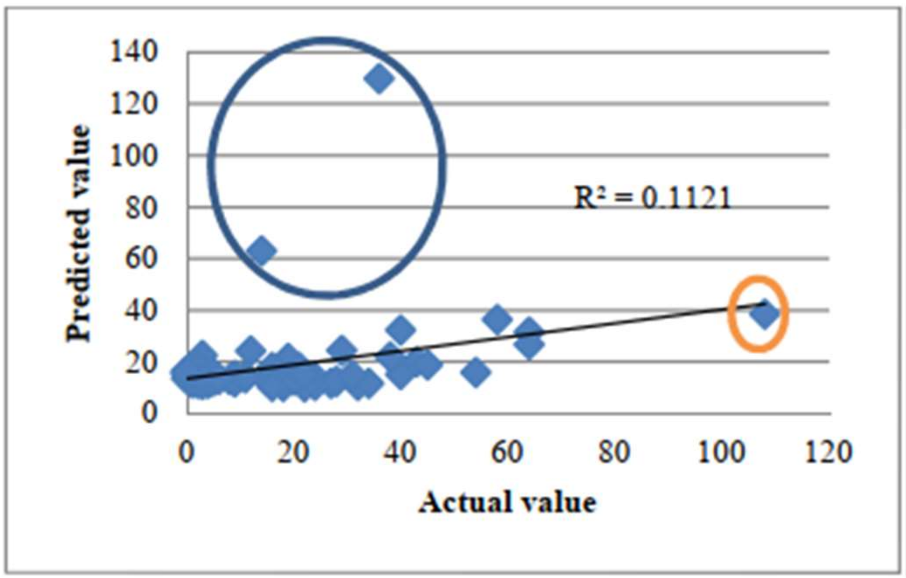


Source: https://bicycledutch.wordpress.com/2014/09/11/utrecht-straightens-out-a-cycle-route/

Outliers: how to test it

"Cook's distance is a measure of the change in the regression coefficients that would occur if this case was omitted, thus revealing which cases are most influential in affecting the regression equation" (Stevens, 1984, p.109)."





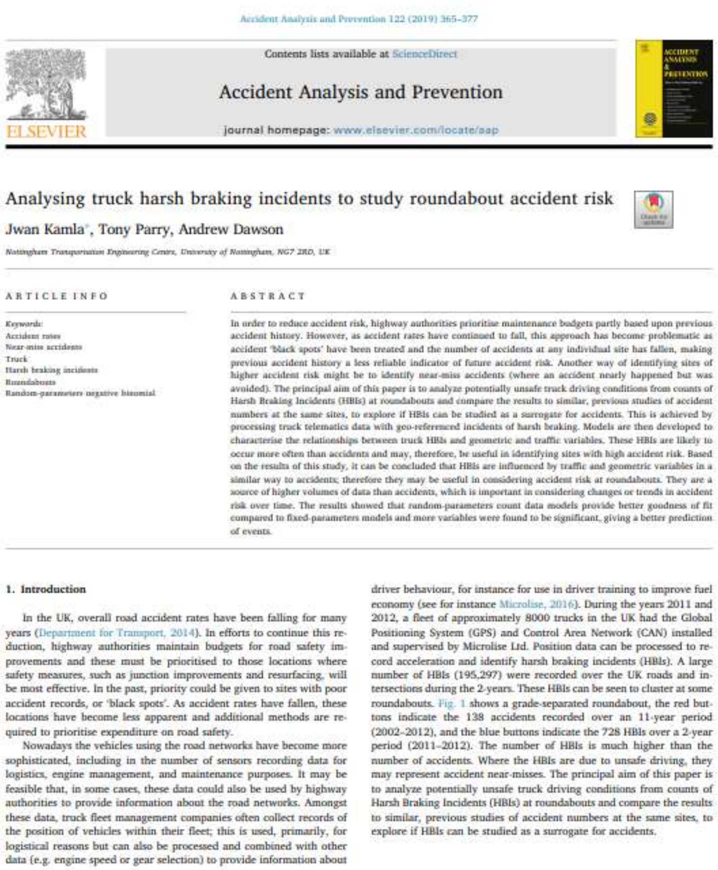
As a rule of thumb if the Cook's distance of the associated value exceeds the cut-off value of (4/number of observations) then it is considered too influential (Nieuwenhuis et al., (2012), Van Der Meer et al. (2010), Belsley et al. (1980)).

CIHT

Conclusions

- Linear regression not suitable for count data (incident/collision)
- Poisson only suitable if mean and variance of count data (incident/collision) are equals
- Negative binomial used when variance is not equal to mean
- Low R^2 not an indication of no relationship
- Paired T-test for before and after studies
- Outliers should be tested using proper statistical methods

My publications



Accident Analysis and Prevention 122 (2019) 365–377

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Analysing truck harsh braking incidents to study roundabout accident risk

Jwan Kamla*, Tony Parry, Andrew Dawson

Nottingham Transportation Engineering Centre, University of Nottingham, NG7 2RD, UK

ARTICLE INFO

Keywords:
Accident rates
Near-miss accidents
Truck
Harsh braking incidents
Roundabouts
Random-parameters negative binomial

ABSTRACT

In order to reduce accident risk, highway authorities prioritise maintenance budgets partly based upon previous accident history. However, as accident rates have continued to fall, this approach has become problematic as accident 'black spots' have been treated and the number of accidents at any individual site has fallen, making previous accident history a less reliable indicator of future accident risk. Another way of identifying sites of higher accident risk might be to identify near-miss accidents (where an accident nearly happened but was avoided). The principal aim of this paper is to analyze potentially unsafe truck driving conditions from counts of Harsh Braking Incidents (HBIs) at roundabouts and compare the results to similar, previous studies of accident numbers at the same sites, to explore if HBIs can be studied as a surrogate for accidents. This is achieved by processing truck telematics data with geo-referenced incidents of harsh braking. Models are then developed to characterise the relationships between truck HBIs and geometric and traffic variables. These HBIs are likely to occur more often than accidents and may, therefore, be useful in identifying sites with high accident risk. Based on the results of this study, it can be concluded that HBIs are influenced by traffic and geometric variables in a similar way to accidents; therefore they may be useful in considering accident risk at roundabouts. They are a source of higher volumes of data than accidents, which is important in considering changes or trends in accident risk over time. The results showed that random-parameters count data models provide better goodness of fit compared to fixed-parameters models and more variables were found to be significant, giving a better prediction of events.

1. Introduction

In the UK, overall road accident rates have been falling for many years (Department for Transport, 2014). In efforts to continue this reduction, highway authorities maintain budgets for road safety improvements and these must be prioritised to those locations where safety measures, such as junction improvements and resurfacing, will be most effective. In the past, priority could be given to sites with poor accident records, or 'black spots'. As accident rates have fallen, these locations have become less apparent and additional methods are required to prioritise expenditure on road safety.

Nowadays the vehicles using the road networks have become more sophisticated, including in the number of sensors recording data for logistics, engine management, and maintenance purposes. It may be feasible that, in some cases, these data could also be used by highway authorities to provide information about the road networks. Amongst these data, truck fleet management companies often collect records of the position of vehicles within their fleet; this is used, primarily, for logistical reasons but can also be processed and combined with other data (e.g. engine speed or gear selection) to provide information about driver behaviour, for instance for use in driver training to improve fuel economy (see for instance Microlise, 2016). During the years 2011 and 2012, a fleet of approximately 8000 trucks in the UK had the Global Positioning System (GPS) and Control Area Network (CAN) installed and supervised by Microlise Ltd. Position data can be processed to record acceleration and identify harsh braking incidents (HBIs). A large number of HBIs (195,297) were recorded over the UK roads and intersections during the 2-years. These HBIs can be seen to cluster at some roundabouts. Fig. 1 shows a grade-separated roundabout, the red buttons indicate the 138 accidents recorded over an 11-year period (2002–2012), and the blue buttons indicate the 728 HBIs over a 2-year period (2011–2012). The number of HBIs is much higher than the number of accidents. Where the HBIs are due to unsafe driving, they may represent accident near-misses. The principal aim of this paper is to analyze potentially unsafe truck driving conditions from counts of Harsh Braking Incidents (HBIs) at roundabouts and compare the results to similar, previous studies of accident numbers at the same sites, to explore if HBIs can be studied as a surrogate for accidents.

1) Jwan Kamla (2016). **"Analysing Truck Position Data to Study Roundabout Accident Risk ".** Thesis

2) Kamla J.,Parry T, & Dawson A. (2018). **Analysing Truck Harsh Braking Incidents to Study Roundabout Accident Risk.** Accident Analysis and Prevention**,** DOI10.1016/j.aap.2018.04.03.

3) Kamla J.,Parry T, Dawson A. (2018) **.The Relationship between Truck Harsh Braking Incidents and Truck Accidents at Roundabout Approaches:** Transportation Research Board, 97th Annual Meeting, 10 January 2018, Washington, D.C

4) Kamla J.,Parry T, & Dawson A. (2017). **The influence of road marking, shape of central island, and truck apron on total and truck accidents at roundabouts:** Transport Infrastructure and Systems: Transport Infrastructure and Systems, Dell'Acqua & Wegman (Eds), @2017 Taylor and Francis Group, London, ISBN: 978-1-138-03009-1.

5) Kamla J.,Parry T, Dawson A. (2017) **.Application of Random Parameters Model to Estimate Truck Accidents at Roundabouts:** Transportation Research Board, 96th Annual Meeting, 11 January 2017, Washington, D.C.

6) Kamla J., Dawson A. Parry T. (2017) . **Feasibility of Using Truck Harsh Braking Incidents for Predicting All Vehicle Accidents at Roundabouts:** Transportation Research Board, 96th Annual Meeting, 9 January 2017, Washington, D.C.

7) Kamla J.,Parry T, Dawson A. (2016) . **Roundabout Accident Prediction Model: Random-Parameter Negative Binomial Approach:** Journal of the 604 Transportation Research Board, DOI 10.3141/2585-02, Oct 2016, No. 2585, pp. 11–19.

8) Kamla J.,Parry T, Dawson A. (2016) . **Roundabout accident prediction model: an application of random parameters negative binomial approach to roundabout accidents in the United Kingdom:** Transportation Research Board, 95th Annual Meeting, 10-12 January 2016, Washington, D.C.

9) Kamla J.,Parry T, Dawson A. (2015) **. Roundabout accident prediction model: an application of random parameters negative binomial approach to roundabout accidents in the United Kingdom:** Transportation Research Board, 94th Annual Meeting, 11-15 January 2015, Washington, D.C.

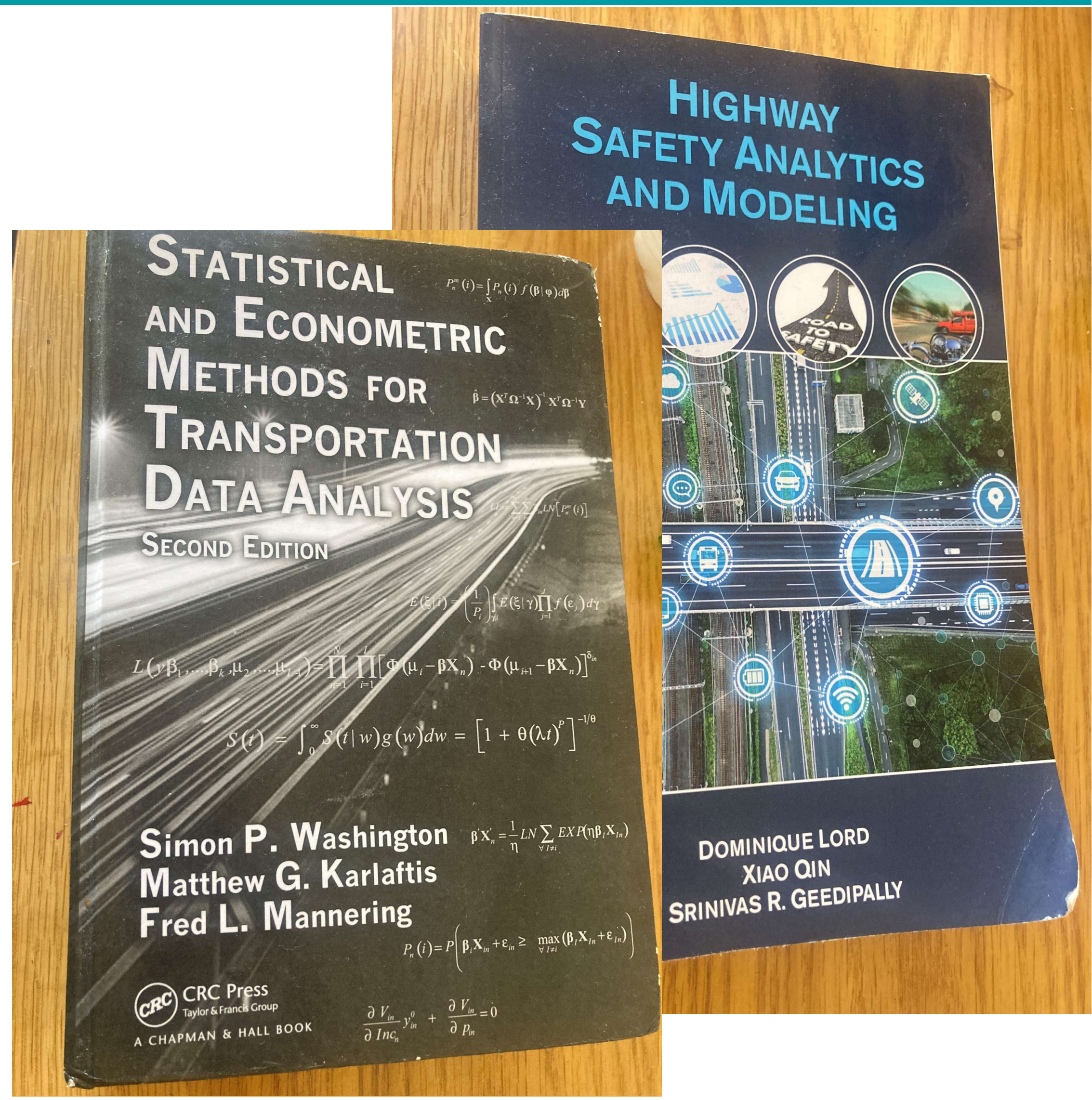
10) Parry T., Kamla J., Dickinson I. (2014). **Feasibility of Using Truck Position Data to Identify Accident Risk.** The 4th International Safer Roads Conference. Cheltenham, UK, 18 - 21 May 2014.

11) Kamla J., Parry T. (2014). **Analysing Near-miss Incidents Using Truck Position Data on M1 Motorway Roundabouts:** Kurdistan Student Conference, hosted jointly by the Kurdistan Society and the International Office of Nottingham University, UK, 15 September 2014**.**

Extra Source

Mannering, F., Bhat, C., 2014. Analytic methods in accident research: Methodological frontier and future directions. Analytic Methods in Accident Research 1, 1-22

Which will get you up-to-date on the latest statistical approaches.



Thanks for your attention

jwan.kamla@jacobs.com