21st Australasian Transport Research Forum
Adelaide, September 1997

Interpreting Commercial Vehicle Survey Data

Samantha Y Taylor
Institute of Transport Studies
Monash University

Ken W Ogden
Group Manager Public Policy
Royal Automobile Club of Victoria

Abstract:

Analytical rigour in statistical analysis is often taken for granted. Although every effort may be made in the initial phases, there may be several factors which influence the resulting accuracy of data. This paper describes some fundamental post-survey statistical analyses, with particular reference to commercial vehicle surveys. The greater Sydney region Commercial Vehicle Survey is used to illustrate how data can be misinterpreted if errors are not quoted or calculated properly. Stratification of samples, response rates, means, medians, data distribution and sampling errors using the relative standard error, are all discussed with emphasis on the role and calculation of each, and how easily the data and statistics can be misinterpreted.

Contact Author:

Samantha Taylor
Institute of Transport Studies
Department of Civil Engineering
Monash University
Clayton VIC 3168

telephone: (03) 9905 9696 fax: (03) 9905 4944
email: Samantha.Taylor@eng.monash.edu.au

Introduction

Most people agree that statistical analysis, survey methodology, and experimental design go hand in hand. But why is it that we often pay only cursory attention to quoting errors for data estimates? If errors are not quoted, the common perception is that the sampling error is so small it is not worthy of concern. However, in practical transport and traffic problems, there are only very few situations where this is the case. If published estimates are to be compared, it is essential to know the sampling error so that conclusions about the data can be drawn with confidence. It is necessary to know if the apparent differences are due to real differences or if they are due to variability within the data.

While a specified accuracy may be one of the outputs of a survey there are many factors in the survey process which contribute to the final results. Objectives must be clearly and well defined, the design of the survey needs to take account of fatigue and ability of respondents, the sample design needs to consider variability within the population, and thought needs to be given to the likelihood of non-response

Due to the sparseness of expertise and experience in area-wide commercial vehicle surveys it is important to ensure that what is learnt in one study is shared with others. To this end the CVS has allowed the identification of shortfalls and possible improvements for future data collection and analysis

It is essential to clearly define OBJECTIVES

As with all investigations it is vital that the objectives of a study are well defined from the outset, which is not always easy. Although you may think this is obvious, it is surprising how frequently only the bare minimum of time and effort is given to this process. Objectives define why the study is being undertaken, and in most cases they should have been defined well before the decision to undertake the study was made.

In situations where there is little existing knowledge, it becomes very difficult to set meaningful and detailed objectives. By nature, some knowledge is required before the outputs of a study can be defined - i.e. knowledge of how the information will assist in finding a 'solution'. Therefore a certain amount of trial and error is inevitable as we search to understand specific issues in more depth. The urban freight area could arguably be used as an example where (in Australia) there is often a degree of trial and error associated with urban freight data collection. The down side of this is that for obvious reasons professionals don't want to promulgate their 'errors', even though often far more is learnt from 'mistakes' than from successes

However, it is absolutely critical that the objectives be defined as clearly as possible to avoid the propensity toward error! Clearly defined objectives also enable priorities to be defined and informed decisions made about trade-offs between:

- data quality,
- data quantity and
- resources.

Objectives which are too broad can yield data which are also broad and thus do not satisfy needs for detailed information. On the other hand objectives which are very detailed can result in inefficient data collection where the data only satisfy immediate needs. This is likely to be an issue if large amounts of capital are invested in a study. A little forward and informed thinking expressed in more informative objectives could have yielded other useful data with minimal additional resources.

What size sample?

The single factor which impacts most on cost and affects the results most significantly is sample size. If the sample size isn't adequate the data are likely to be of very poor quality.

When defining the objectives, it is good practice to define the minimum accuracy required in the output. Other factors which need to be considered in determining sample size are (Australian Bureau of Statistics, 1996):

- size of the population,
- variability within the population; if there is high variability then a larger sample is required or stratification may be used,
- proportion of the population with the attribute being measured,
- sample design,
- resources (cost),
- level of accuracy required,
- level of detail required,
- likely level of non-response, and
- respondent burden.

There is a point where the sample size becomes too small to provide meaningful representation of the estimated variable. The error is too large. It is therefore necessary to establish a critical minimum sample size by mathematical manipulation. The formula for minimum sample size depends on the desired relative standard error[1], the population

[1] Relative Standard Error (RSE) or sampling error is defined as the difference between the sample mean and population mean. The mathematical formula is:

size and the proportion of the population which constitutes the sample. This can be expressed as:

$$n \geq \frac{N}{1 + N\left[\frac{p \cdot RSE^2}{1-p}\right]}$$

where:
n = sample size
N = population size
p = expected sample proportion
RSE = desired Relative Standard Error

Response rate is critical

Data quality

Even if every effort is made in the steps prior to the survey process, non-response could mean that accuracy is compromised because the sample size is too small or biased.

Data quality is a function of several factors including questionnaire design, collection difficulties, processing procedures and data capture. Some of these factors can be estimated using mathematical methods (sampling error) while others are more difficult to estimate (non-sampling error) and often depend on the skills and experience of the collecting agency (Australian Bureau of Statistics 1991)

Non-sampling error

Non-sampling error refers to errors which cannot be measured mathematically. They could be due to non-response, badly designed questionnaires, respondent bias and processing errors.

No matter how good the questionnaire or the interviewers, errors can be introduced either consciously or unconsciously by the respondent. Fatigue can result if a high level of commitment is required by the respondent, and this varies between respondents. For example, commercial vehicle drivers may have a significant amount of paperwork to complete on their travel (log book records) and activities (documentation to be completed for each delivery and pick-up), and therefore consider a 'voluntary' survey as a low priority task and hence become tired and impatient with it quickly. Additionally, they may have trouble envisaging how the survey will help them in the future.

$$RSE(\%) = \left(\sqrt{\frac{\text{var}(sample) \times (1 - n/N)}{n}}\right) \div mean(sample) \times 100$$

CVS survey process

The commercial vehicle survey (CVS) undertaken by the Transport Data Centre in Sydney, Australia, used mail-out mail-back questionnaires to collect data on the activities of commercial vehicles registered in Greater Sydney. The CVS had an initial sample size of 30,000 vehicles from the Roads and Traffic Authority's registration database. Since the survey was to run over 12 months, to minimise sample loss 7,500 vehicles were chosen each quarter of the 12 month period, and cross checks were done to ensure a vehicle was not selected twice. The sample frame was all registered commercial vehicles within the greater Sydney region (Sydney, Illawarra, The Hunter Valley, Blue Mountains) (see Figure 1). Before a survey was sent to the address, initial contact was made by telephone to establish the eligibility of the vehicle, and the location of the owner. As a result of this process on average about 25% of the sample was ineligible.

Recruiting commercial vehicle owners

Once the vehicle was confirmed as eligible, the owner was asked to participate in the survey, and also asked to pass the survey form onto the driver of the vehicle (if the driver was different to the owner). At this point about half of the eligible vehicles were actually recruited, and this varied between vehicle strata. Only 35% of the light commercial vehicles were recruited, while 52% and 59% of the rigid and articulated vehicles were recruited (see below for details on stratification).

It is interesting to speculate as to why the light commercials had such a low recruitment level. Perhaps the expected long term benefits in terms of improvements to the network through understanding commercial vehicle movements were less apparent or important to them. In addition, or alternatively, couriers in the light commercial category may have been reluctant to commit themselves to fill out a trip diary if they did over 30 trips per day.

On the other hand the higher participation rate of rigid and particularly articulated vehicles may be due to owners either being more aware of the inadequacies of the road network, or directed by their employer to participate.

This reluctance of drivers to participate in self-completion surveys is one of the recurring battles with commercial vehicle data collection. One of the emerging issues is how, in the future, new technology can be used to reduce driver burden and thus increase the accuracy of surveys.

Three types of questionnaires

Recruited vehicles were sent one of three types of questionnaires. Each requested information on trip details, time of day of movements, and commodities delivered or picked-up. The survey for light commercial vehicles didn't ask for commodity information while the survey for vehicle with a trailer (including all articulated vehicles) requested data on configuration changes.

Reducing errors through STRATIFICATION

Stratification is the process of dividing a population into groups (or strata) so that similar units are placed in the same group, which differs from units outside the group. Each strata (group) is then treated and analysed as a population. The theory is that by stratifying a population, the variability in each strata is reduced and therefore the error is reduced since sampling error is directly related to variability and sample size.

In summary, the major advantages of stratification are that (Australian Bureau of Statistics 1996):

- the standard error can be reduced,
- the strata are treated as separate populations and this means that different selection methods may be applied and different information obtained for individual strata, and
- to ensure groups of special interest are represented in the sample.

In some cases stratification may not yield better results, but if stratification is used the variable by which the population is grouped needs careful consideration to ensure that variability is in fact reduced. For example, the Sydney CVS was stratified by vehicle type; light commercial, rigid and articulated vehicles. In practice the results of the survey indicated that light vehicles (2 axle vehicles; 4WD, pick-up/utility) included couriers and trades-people which cover both high frequency deliveries and day-long service (trade) calls. In other words high variability in number of trips per day and hence higher error. In the same survey the rigid and articulated strata exhibited less variation and therefore the aim of stratification was achieved and the error reduced.

Sampling error is also a function of sample size; the larger the sample size the smaller the error, ceteris paribus (see earlier section on sample size).

Interpreting the CVs results

Background information on the CVS

A total of 9,946 questionnaires were returned providing data on 24,882 trips across three vehicle classes; light, rigid and articulated vehicles. Further information on the CVS methodology and analysis can be found in Maldonado and Akers (1992) and Taylor (1997) respectively.

For the purposes of this paper analysis is restricted to internal-internal trips, or trips with both ends within the greater Sydney region, and a trip is defined as a one way movement from origin to destination.

Data were provided by the drivers and included origin and destination, travel time, distance travelled, idle time, time of day, and so on. Data were generally highly variable and the spread of trips was very large indeed (see Figure 2). This was mostly due to the inherently variable nature of commercial travel but is also due to the low response rate in parts of the survey, which is typical of mail-out mail-back freight movement surveys (Lau 1995).

Accuracy of survey results

An example of the large spread of data is shown by the box and whisker plot in Figure 2 (Taylor 1997). The top and bottom of the rectangle indicates the lower and upper quartiles (25%le and 75%le respectively) while the vertical lines (whiskers) which extend from the ends of the rectangle depend on the interquartile range (1.5×IQR). The median is marked inside the rectangle with a bold line while the mean is shown by a dashed line. Usually the individual data points which do not fall within the box and whisker range are also plotted, but for reasons of scale these values are not shown here, suffice to say that values were reported over the whole permissible range.

A low response rate leads to a smaller sample size and combined with high variability this can lead to lower confidence in the results. To estimate the error in the individual characteristics reported, the sampling error or standard error (SE) is used, and is expressed as a percentage of the associated estimate (or as a Relative SE, RSE). The sampling error is defined as the difference between the population mean and the sample mean[2]. Where RSEs are high, comparisons between estimates need to be made

[2] To estimate the population mean, the Central Limits Theorem states that if infinite number of samples were drawn from the population then the average of these sample means would be the population mean. Therefore, on average the sample mean is a good estimator of the population mean. Similarly, the variance of the sample means would be the error (or SE). And by mathematical manipulation, the variance of the sample means is $\mathrm{var}(means) = \dfrac{\mathrm{var}(pop)\times(1-n/N)}{n}$. The Central Limits Theorem states that on average the sample variance will yield the population variance, providing we use (n-1) on the denominator in the usual sample variance equation, therefore the sample variance is used in place of the population variance in the variance equation above. The SE is then equal to the square root of the var(means):

cautiously as some differences may be due more to data variability than to actual differences between categories (Table 1 and 2) For example the difference between total and rigid daily distance travelled might be interpreted as a real difference if the RSEs are not taken into account. When the RSEs are expanded to produce upper and lower bounds you can see that the total and rigid values overlap, indicating that there is no difference, or the apparent difference is due to data variability. If you now look at the difference between the daily idle time of rigid vehicles and articulated vehicles, taking into account the RSEs, you can conclude that the difference is real, ceteris paribus (see Table 2 for upper and lower bounds).

Although prima facie this data may at times have higher than desirable uncertainty associated with the results (ie RSEs associated with light vehicles), it is not uncommon for the transport profession to pay only cursory attention to errors associated with surveys such as this one and not publish errors. It is important to identify areas where high variability exists because this information is extremely useful for both future surveys, and for establishing a level of robustness or context for conclusions and discussion. In terms of future surveys, existing data can be used for survey design using 'optimal allocation'. [Optimal allocation is a statistical process which can only be used when there is previous data available and uses revealed variability of stratum to allocate appropriate sample sizes.]

Conclusions

- Publishing results of surveys, particularly where there is a dearth of information, is vital if we are to progress the state of knowledge in urban freight data collection. Even if results are less than ideal, the profession should make an effort to publish. What is critical is that the results are transparent and explicit - so they can then be used in context!
- Objectives must be clearly defined, explicit and recorded together with any modifications which result as the study progresses. As well as establishing explicit criteria by which the study can be evaluated, it provides a reality check on expectations. This assists discussion of issues which may arise when (or if) the data is used for purposes other than those it was originally intended for.
- Optimal allocation should be used where possible in future surveys. The experience of the CVS indicates that light commercial vehicles (4WDs and utilities) have quite different survey response rates and characteristics than heavier vehicles. On one hand, light vehicles seem to exhibit characteristics which are more comparable with passenger vehicles than trucks, particularly with regard to trade and some service vehicles. On the other hand, couriers which undertake 40 and 50 stops are different to both trucks and private passenger cars.

$$SE = \sqrt{\frac{\mathrm{var}(sample) \times (1 - n/N)}{n}}$$

- The CVS provides information specifically provided by the driver of the vehicle and is therefore limited by the drivers knowledge and experience. This has implications for what a survey can reasonably expect from a driver, notwithstanding fatigue, attitudinal and literacy factors.
- Advances in Intelligent Transport Systems have potential to impact and significantly improve many of the above listed issues. The increasing accuracy and affordability of GPS receivers, smartcards, responders, and scanners enables vehicles to be tracked in real time. This has potential not only for data collection and more explicit accuracy, but also for improvements in total transport system productivity and safety.

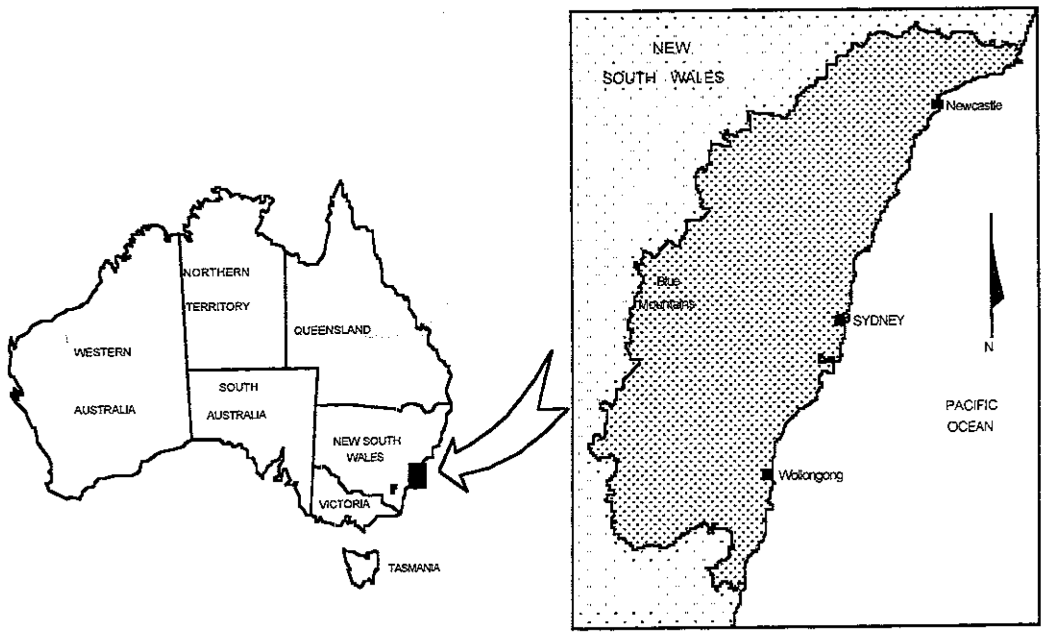

Figure 1 Greater Sydney metropolitan region (study area)

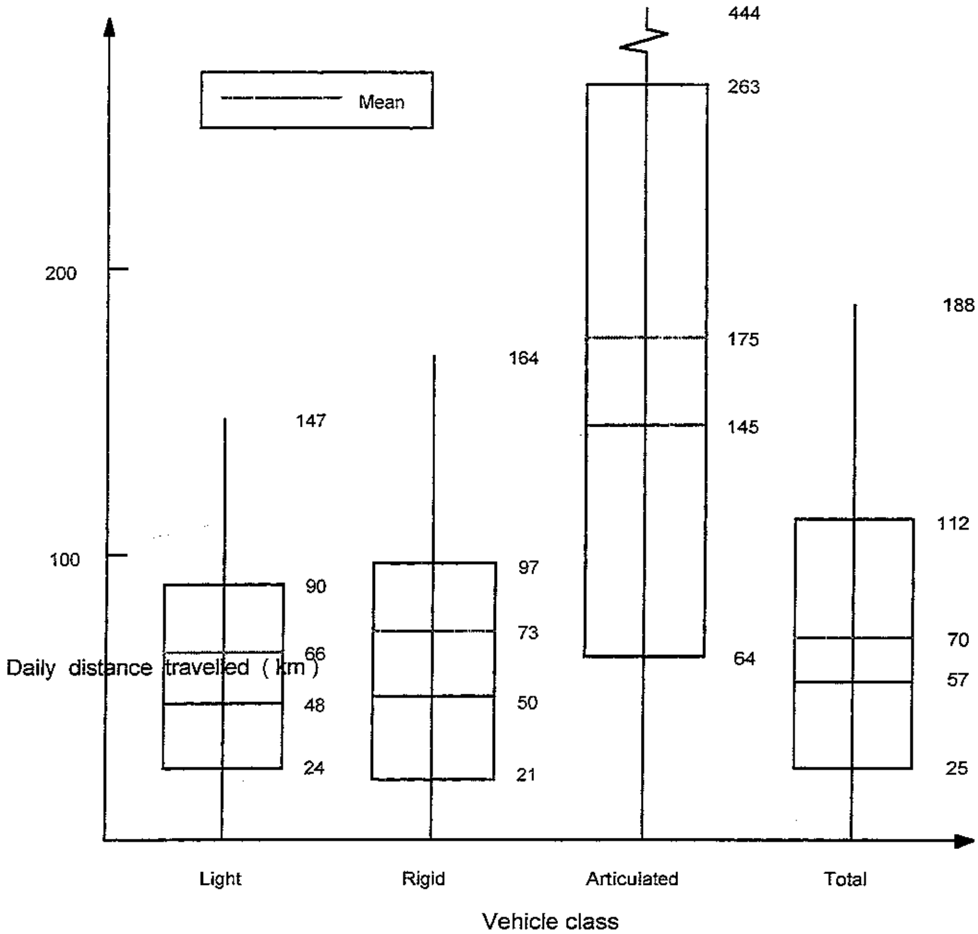

Figure 2 Example of box and whisker plot: daily distance travelled

Table 1 Vehicle usage patterns per vehicle (weekdays)

	Light	Rigid	Articulated	Total
Daily distance travelled (km)				
mean	66	73	175	70
median	48	50	145	57
RSE (%)	±33%	±8%	±7%	±9%
Daily travel time (h:min)				
mean	1:45	2:01	3:59	1:52
median	1:24	1:30	3:35	1:40
RSE (%)	±30%	±6%	±5%	±8%
Daily idle time (h:min)				
mean	5:35	4:32	3:31	5:16
median	6:08	4:15	3:00	4:17
RSE (%)	±24%	±6%	±6%	±7%
Working Vehicles#	64%	58%	52%	62%
No. vehicles in sample	1058	2945	757	4760

Source: 1991/92 CVS
RSE is the relative standard error
\# Working vehicles are the average proportion of vehicles that make at least one trip within the study area, on an average day
* Trip frequency is the number of trips within the study area, an average working vehicle makes on an average day.

Table 2 Mean vehicle usage patterns per vehicle (weekdays); including upper and lower bounds

	Light	Rigid	Articulated	Total
Daily distance travelled (km)				
lower bound	44	67	163	64
mean	66	73	175	70
upper bound	88	79	187	76
RSE (%)	±33%	±8%	±7%	±9%
Daily travel time (h:min)				
lower bound	1:14	1:54	3:47	1:43
mean	1:45	2:01	3:59	1:52
upper bound	2:16	2:08	4:11	2:01
RSE (%)	±30%	±6%	±5%	±8%
Daily idle time (h:min)				
lower bound	4:15	4:16	3:18	4:54
mean	5:35	4:32	3:31	5:16
upper bound	6:55	4:48	3:44	5:38
RSE (%)	±24%	±6%	±6%	±7%

Source: 1991/92 CVS
RSE is the relative standard error

Acknowledgments

The authors would like to gratefully thank the Transport Data Centre (TDC), Department of Transport NSW, for providing access to the commercial vehicle survey data. The views expressed, interpretations and results presented are the authors' and are not necessarily supported by TDC.

References

Australian Bureau of Statistics (1991) Survey of motor vehicle use, 1991 *Cat 9208.0* (ABS, Canberra).

Australian Bureau of Statistics (1996) *Basic Survey Design; Course Notes.* (ABS, Canberra).

Lau S W (1995) *Truck Travel Surveys: A review of the literature and state-of-the-art.* (Metropolitan Transportation Commission, Oakland CA, USA).

Maldonado R D and Akers G (1992) 1991-92 commercial vehicle survey: A presentation and discussion of the survey methodology. *Australasian Transport Research Forum 17(1),* pp 135-148. (Bureau of Transport and Communications Economics, Canberra).

Taylor S Y (1995) *Urban freight Activities in Sydney: Some Analysis of the 1991-92 Commercial Vehicle Survey.* MEngSc Thesis, Department of Civil Engineering, Monash University, Melbourne.

Taylor S Y (1997) A basis for understanding urban freight and commercial vehicle travel. *Research Report ARR 300.* (ARRB Transport Research Ltd, Melbourne).