8th WCTR

THE REALITY OF SURVEY RESULTS; AN URBAN GOODS MOVEMENT CASE STUDY

SAMANTHA YVONNE TAYLOR
Institute of Transport Studies, Department of Civil Engineering
Monash University
Clayton 3168, Vic, Australia

KENNETH WADE OGDEN
Royal Automobile Club of Victoria
550 Princes Hwy
Noble Park, 3174, Vic, Australia

Abstract

Results from a commercial vehicle survey undertaken in Sydney, Australia shows relative standard errors varying from 5% to 33%. When looking at similar survey data from around the world, very few explicitly state the accuracy of final results, which can be misleading to users. When errors are not quoted, many users assume the sampling error is so small it is not worthy of concern. However, in practical transport and traffic problems, there are very few situations where this is the case. In order to compare published results it is essential to know the sampling error so conclusions can be drawn with confidence.

Accuracy depends a lot on the survey methodology, response rate and rigour of the subsequent statistical analysis. It also depends on the sample size and variability of the subject. This paper briefly describes the greater Sydney (Australia) commercial vehicle survey, it discusses the statistical analyses and highlights some important issues to consider in future surveys.

INTRODUCTION

Information is fundamental to understanding and data is fundamental to information. The transport profession has been lamenting for years that there is a much more limited understanding of commercial vehicle movement than people movement despite the fact that they are of comparable economic significance (Ogden 1992). One of the biggest hurdles in the past has been the high cost of collecting data, and the inescapable reliance on driver and commercial cooperation. While today there are increasing numbers of examples where vehicle tracking technology and electronic data collection have been exploited to reduce the cost of collection, we are yet to fully realise the benefits of this technology (Ogden 1993, Battelle Transportation Division 1997, Williams 1997, Taylor 1997b). No doubt our understanding will increase dramatically over the next five to ten years.

In interpreting data from commercial vehicle surveys or any other data for that matter, it is essential to have knowledge of the assumptions and objectives of the data collection. These must be considered if the resulting information is to have any chance of being reliable and indicative of the real world. Questions such as: Is the data representative? What was the nature of non-response? Is there a bias in the data? What is the sampling error? and Are the differences between indices significant? must all be asked to ensure that the nature of the resulting information is understood, and more importantly that its limitations are understood.

This paper uses the Greater Sydney Metropolitan Region, Commercial Vehicle Survey (CVS)(1991/92) to illustrate how easily data can be misinterpreted to form spurious information. It discusses relative standard errors (RSEs) and box-and-whisker plots (stem-and-leaf plots) which help prospective users to understand data. It also identifies areas where the results of the CVS can be used to refine the methodology and assumptions in future surveys.

Definitions

For the purposes of this paper, some definitions are given below:

- *articulated vehicles* refer to vehicles which have an articulation point (also called semi-trailers, or tractor semi-trailer combinations),
- *rigid vehicles* are vehicles which have no articulation point even if the vehicle is towing a trailer (also called straight trucks, and roughly equivalent to medium sized trucks in Europe),
- *light commercial vehicles* (LCVs) are light delivery and service vehicles with comparable characteristics to those of passenger cars; they include 4-wheel vans, pick-up trucks, and other utility vehicles.

Analyses in this paper are 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.

THE CVS SURVEY METHODOLOGY IN BRIEF

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. Commercial vehicles are those registered (licensed) for use for commercial purposes, as distinct from private purposes. 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 included all registered commercial vehicles with addresses within the greater Sydney region (a coastal strip about 250 km in length, with a population of some 4.4 million, including the sub-regions of Sydney, Illawarra, The Hunter Valley, Blue Mountains - see Figure 2). 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.

The sample frame was a probability sample stratified by vehicle class and 13 geographical areas.

In theory a probability sample allows inferences to be made about the target population. By knowing the selection probability for each unit, selections can then be made to ensure a more representative sample. However, this method assumes either full response or random non-response which is rarely, if ever, the case in commercial vehicle surveys. Even though cooperation between commercial organisations, government and survey administrators has improved, there remains a reluctance by drivers to participate in surveys. For example, Lau (1995) in USA reported that owner drivers were more likely to be subjects of non-response than drivers who were directed by the their employers to participate. Overall, the nature of non-response in commercial vehicle surveys is something we know very little about.

Recruiting commercial vehicle owners and drivers

Once the vehicle was confirmed by telephone 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 stratum. 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 consider possible reasons 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 numerous 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. Typically also, the number of trips to be reported was quite small (see below).

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 increase the accuracy of surveys.

Three types of questionnaire were used

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. For the purposes of this paper analyses is restricted to trip distance, travel time, idle time, frequency, daily travel time, daily distance travelled and daily idle time. Further information on commodity movements and temporal distributions can be found in Taylor (1995, 1997a) while information on methodology can be found in Maldonado and Akers (1992).

INTERPRETING THE CVS DATA

Responses to self-completion questionnaires

A total of 9,946 questionnaires were returned providing data on 24,882 trips across three vehicle classes; light, rigid and articulated vehicles. A total of 5,623 (57.2%) questionnaires reported at least one trip (Table 1). Of the questionnaires containing trip data, 5,294 reported weekday travel data and 331 reported travel on a weekend. Due to the small sample size of weekend data, only weekday data will be considered in this paper.

Data were generally highly variable and the spread of trip indices was very large indeed (see Figure 3 and Figure 4). 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).

No matter how good the questionnaire or the interviewers, errors can be introduced either consciously or unconsciously by the respondent. Fatigue results if a high level of commitment is required by the respondent, which 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.

Statistical analyses

As part of the usual reporting procedures, fundamental statistics such as mean, median and sampling error were produced for the survey results. Table 2 and Table 3 shows that the relative standard error[1]. (RSE) (or the sampling error) varies significantly depending on the vehicle class under investigation. Light Commercial Vehicles (LCVs) have a higher RSE due to a lower response rate and a significantly higher number of them in the population, compared to the other vehicle classes (see Figure 1, Table 2 and Table 4).

Data richness

An example of the large spread of data is shown by the box-and-whisker plot in Figure 3 (Taylor 1997a). 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 (1.5xIQR). 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.

It is clear from the box-and-whisker plot in Figure 3 that there is a wide range of distances over which articulated vehicles travel on an average weekday. This figure illustrates the relative positions of the 25%le, 50%le and 75%le points and shows that the mean value of 175 km falls far short of conveying the full richness of the information. On studying Figure 4, the daily travel time on an average weekday, a similar picture emerges. Again, the articulated vehicles clearly have a wider range of daily travel time values compared with the light and rigid vehicles. This is as expected since travel time and distance are correlated, and the very large area of the Greater Sydney Metropolitan Region (GSMR) means that the study area includes the provincial cities of Newcastle and Wollongong. As a result, many of these trips between Sydney, Newcastle and Wollongong could arguably be considered as inter-regional, but because Australia is such a vast country this is not the case (see Figure 2). In European countries these trips would probably be considered inter-regional.

Accuracy

A low response rate leads to a smaller sample size and combined with high variability this can lead to lower confidence in the results. This is clearly evident in the light commercial class where a low response and high variability has led to a higher relative standard error.

To estimate the error of the mean indices, the sampling error or standard error (SE) is used, and is expressed as a percentage (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 cautiously as some differences may be due more to data variability than to actual differences between categories.

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 (Table 2 and Table 3). When the RSEs are expanded to produce upper and lower bounds it can be seen that the total and rigid values overlap, indicating that there is no difference, or the apparent difference is due to data variability. If we look at the difference between the daily idle time of rigid vehicles and articulated vehicles, taking into account the RSEs, we can conclude that the difference is real, ceteris paribus (see Table 3 for upper and lower bounds).

Comparing some results of the CVS

Taking the RSEs into account, the mean number of trips made by each vehicle class does not vary. Prima facie, this is counter-intuitive as we would expect vehicles which travel longer distances per trip, on average, to make less trips per day than vehicles which travel shorter distances. There are greater economies of scale achieved by one larger vehicle travelling long distances, rather than multiple smaller vehicles travelling the same distances. One explanation for the unexpected result could be the fact that the survey was undertaken during the Australian recession of the early 1990s (Taylor 1997a). The Australian national GDP decreased by almost 1% in 1991 (Australian Bureau of Statistics 1995). Since there is a significant relationship between GDP and the amount of freight moved, a reduction in economic activity leads to a reduction in commercial vehicle trips (Taplin 1983).

Alternatively, a more undesirable explanation could be that the responses are biased toward the lower end of the trip scale. Perhaps the novelty of completing a questionnaire was reduced after a certain number of stops.

We can safely conclude that light and rigid vehicles, on average, travel significantly less distance and time per trip than articulated vehicles (Table 4 and Table 5). In addition, articulated vehicles travel at a higher average speed than rigid or light vehicles. This higher speed is most likely due to large vehicles travelling more on arterial roads where stops at traffic lights are a smaller percentage of the overall trip and speed limits are higher (Table 4 and Table 5).

Consistent with the trip indices, daily vehicle travel time and distance travelled by articulated vehicles is significantly greater than for light or rigid vehicles (Tables 2 and 3). This is due to economies of scale of larger vehicles over long distances. As expected, daily idle time per vehicle is less for articulated vehicles than for light and rigid vehicles.

There is no conclusive evidence which suggests that light and rigid vehicles have different characteristics. One possible explanation for this is that the rigid vehicle definition is too broad and hence contains too much variability to adequately represent the population. However further research is required to support this hypothesis.

DISCUSSION

Although prima facie the CVS 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. The experience of the CVS indicates that light commercial vehicles (vans and utility vehicles) 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. In terms of future surveys, this kind of information is useful because existing data can be used for survey design using 'optimal allocation'. (Optimal allocation is a statistical process which can only be used when there are previous data available and uses revealed variability of stratum to allocate appropriate sample sizes.)

The CVS provides information specifically provided by the driver of the vehicle and is therefore limited by the driver's knowledge and experience. A driver does not necessarily know the mass of a consignment or the specific attributes of a vehicle (eg Gross Vehicle Mass). This has implications for what a survey can reasonably expect from a driver, notwithstanding fatigue, attitudinal and literacy factors.

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.

Basic statistical tests

As a minimum, Table 6 presents a list of fundamental statistical tests for survey data.

It is often assumed that fundamental statistics such as those listed in Table 6 are relatively simple to calculate. As with many theories and practices, the practice does not always conform to the rationale of the theory, and it is important to have a sound understanding of basic statistics and probability when manipulating, presenting and testing data. For example, developing RSEs for medians is complex and often not done, whereas RSEs for mean values can be calculated relatively routinely (see Section 3.2) providing a useful indication of the quality of data.

In addition it is suggested that a graph such as the box-and-whisker plot (also known as the stem-and-leaf plot) be used to convey and help assimilate the richness of the data. This provides a particularly useful display of mean, median, 25%le and 75%le positions as well as identifying any perceived 'outliers'.

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, practitioners and researchers 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. This also means that where possible it is essential to convey the error associated with published statistics.

While this paper has concentrated on analyses of the traditional pen and paper survey, the authors acknowledge that a new era of survey data collection is developing as applications of advanced technology to transport systems become more widespread. Intelligent Transport Systems (ITS) have potential to impact and significantly improve many of the fundamental problems associated with data collection such as sample size and cost; notwithstanding the likelihood of concurrently providing different challenges in data analyses. The increasing accuracy and affordability of GPS receivers, smartcards, responders, and scanners enables vehicles to be tracked in real time. This has potential for improvements in accuracy, and reducing respondent burden.

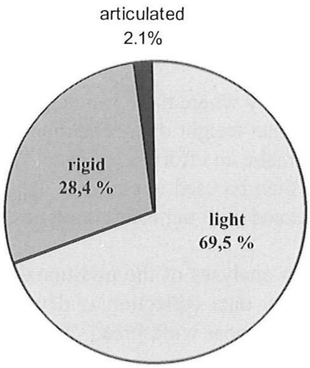

Figure 1 - Graphical representation of vehicle class proportions in population

Note: Wollongong and Newcastle are provincial cities of the State of New South Wales

Figure 2 - Greater Sydney metropolitan region (study area)

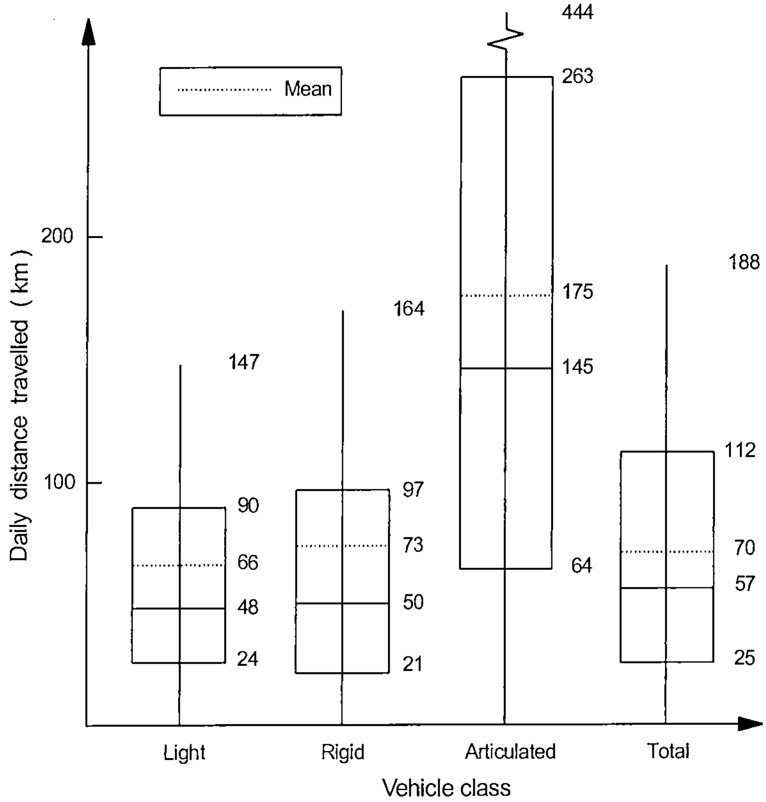

Figure 3 - Box-and-whisker plot: weekday daily distance travelled (km)

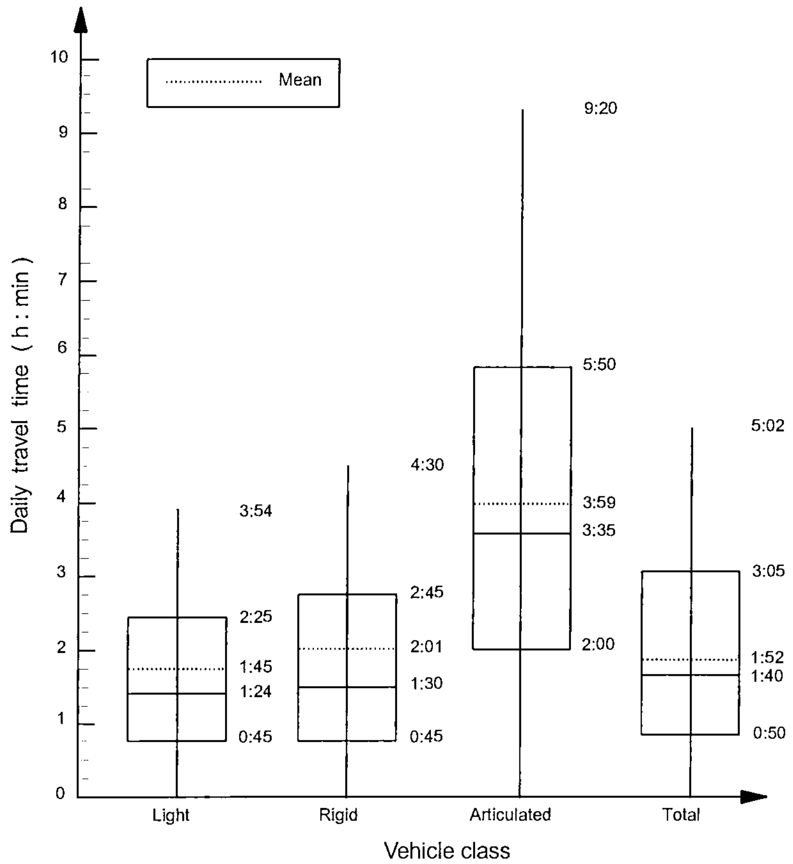

Figure 4 - Box-and-whisker plot: weekday daily travel time (h:min)

Table 1 - Status of returned questionnaires

Status of Returned Q'naires	Light	Rigid	Articulated	Total
Unuseable questionnaires	24	81	18	123
Useable q'naires - no trips	716	2,722	762	4,200
Useable q'naires - trips	1,182 (62%)	3,409 (55%)	1,032 (57%)	5,623 (57%)
Total returned q'naires	1,922	6,212	1,812	9,946

Source: 1991-92 CVS

Table 2 - Mean and median travel per vehicle on an average weekday

	Light	Rigid	Articulated	Total
Trip frequency*				
mean	4.7	4.5	4.5	4.6
median	3.5	3.0	4.0	3.0
RSE (%)	±32%	±6%	±5%	±8%
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%
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 3 - Mean travel per vehicle on an average weekday; including upper and lower bounds

	Light	Rigid	Articulated	Total
Trip frequency*				
lower bound	3.2	4.2	4.3	4.2
mean	4.7	4.5	4.5	4.6
upper bound	6.2	4.8	4.7	5.0
RSE (%)	±32%	±6%	±5%	±8%
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
* Trip frequency is the number of trips within the study area, an average working vehicle makes on an average day.

Table 4 - Mean and median trip lengths, durations and average speeds on an average weekday

	Light	Rigid	Articulated	Total
Length (km)				
mean	14.0	16.1	38.6	15.0
median	8.0	10.0	25.0	10.0
RSE (%)	±21%	±5%	±5%	±6%
Duration (h:min)				
mean	0:22	0:27	0:53	0:24
median	0:15	0:20	0:40	0:20
RSE (%)	±18%	±4%	±4%	±5%
Average Speed (km/h)				
mean	33	33	38	33
median	30	30	38	31
RSE (%)	±9%	±2%	±2%	±3%
No. of trips in sample	4973	13191	3387	21551

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

Table 5 - Mean trip lengths, durations and average speeds on an average weekday; including upper and lower bounds

	Light	Rigid	Articulated	Total
Length (km)				
lower bound	11.1	15.3	36.7	14.1
mean	14.0	16.1	38.6	15.0
upper bound	16.9	16.9	40.5	15.9
RSE (%)	±21%	±5%	±5%	±6%
Duration (h:min)				
lower bound	0:18	0:26	0:51	0:23
mean	0:22	0:27	0:53	0:24
upper bound	0:26	0:28	0:55	0:25
RSE (%)	±18%	±4%	±4%	±5%
Average Speed (km/h)				
lower bound	30	32	37	32
mean	33	33	38	33
upper bound	36	34	39	34
RSE (%)	±9%	±2%	±2%	±3%
No. of trips in sample	4973	13191	3387	21551

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

Table 6 - Fundamental statistical tests for survey data

Assumption	Statistical Tests
Representativeness	Minimum sample size
Accuracy	Relative Standard Error (RSE)
Comparing Means	One or more of the following: • error bars • t-statistic • paired t-test • z-statistic • analysis of variance (eg ANOVA table)

ENDNOTES

1. Relative Standard Error (RSE) or sampling error is defined as the difference between the sample mean and population mean (Walpole and Myers 1989). The mathematical formula is:

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

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 (Australian Bureau of Statistics 1996, Walpole and Myers 1989). 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 $\text{var}(means) = \frac{\text{var}(pop) \times (1 - n/N)}{n}$. 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):

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

ACKNOWLEDGMENTS

The authors wish 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 (1995) Measuring Australia's Economy. **Catalogue No. 1360.0.** ABS, Canberra, Australia.

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

Battelle Transportation Division (1997) **Lexington Area Travel Data Collection Test**. Final report to Office of Highway Information Management (HPM-40), FHWA, Washington, DC, USA.

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, Australia.

Ogden, K.W. (1992) **Urban goods movement: a guide to policy and planning.** 397p, Ashgate: London, UK.

Ogden, K.W. (1993) IVHS: Applications in freight and trucking operations. **Proc Pacific Rim TransTech Conference Vol 1**, pp 333-339, Department of Transportation: Washington, WA, USA.

Taplin, J.H.E. (1983) A long term look at the development of transport with particular reference to road transport over the next 10 to 20 years. **XVII World Road Congress, Conference Discussion, October, Sydney**, *pp 19-34,* Permanent International Association of Road Congresses: Paris, France.

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, Australia.

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

Taylor, S.Y. (1997b) Technology applications for freight data collection. Proc 19th Conference of Australian Institutes of Transport Research (CAITR), November 1997, Peter Daly (ed). **Working Paper ITS-WP-97-18,** Institute of Transport Studies, Monash, Melbourne, Australia.

Walpole, R.E. and Myers, R.H. (1989) **Probability and Statistics for Engineers and Scientists, 4th ed**. Macmillan Publishing Company, New York.

Williams, B. (1997) **Existing and emerging techniques for the automatic collection of data for transport modelling**. A report by Smith System Engineering for Dept. of Transport, UK, Smith System Engineering, UK.