SSC Annual Meeting, June 2000
Proceedings of the Survey Methods Section, pp. 217-220

ESTIMATION IN A TRAFFIC NETWORK SURVEY

A. Demnati, J. Lawson, B. Harbour and J. Young[1]

ABSTRACT

Roadside surveys such as the National Roadside Study conduct roadside observations and interviews to produce a profile of the volume and characteristics of the activity on a road network. The data collected at each site consists, in part, of a random sample of interviews and observations of the trips, and in another part, of a series of counts of trips passing the site during the given time period to capture the size of the site. Data collected from different sites are integrated in a single data set. This document shows the steps involved in this integration.

KEY WORDS: National Roadside Study; Origin-Destination survey; multiple frames; multiplicity; calibration.

RÉSUMÉ

Les enquêtes au bord des routes comme l'étude nationale au bord de la route mènent des observations et des interviews pour produire un profile du volume et des caractéristiques de l'activité sur un réseau de routes. Les données recueillies dans chaque site consistent d'une part en un échantillon aléatoire d'interviews et d'observations, et d'autres part en une série de comptes des voyages qui passent le site durant la période du temps donné pour saisir la taille du site. Les données recueillies dans chaque site sont intégrées en un seul ensemble de données. Ce document montre les étapes impliquées dans cette intégration.

MOTS CLÉS: Étude Nationale au bord de la route; enquête sur l'origine et la destination; bases multiples; multiplicité; calage.

1. INTRODUCTION

We consider a fixed period of traffic during which a collection of trips moves through a network. Each trip originates at one node in the network and travels to another node along a path. The intention of the survey is to produce a profile of the volume and characteristics of the network by taking random samples on each directed link. For example, the National Roadside Study (1995 and 1999 NRS) conducts roadside observations and interviews to produce a profile of the volume and characteristics of the trucking activity on a road network in Canada. The data collected at each site (directed link) consists, in part, of a random sample of interviews and observations of the trips, and in another part, of a series of count of trips passing the site during the given time period to capture the size of the site (number of trucks passing the site for the given period).

Data collected from different directed links are integrated into a single data set. The purpose of this document is to show, in section 2, the steps involved in this integration. Under the assumptions included below, the document also show, in section 2, how the methodology takes into account :

- the sampling scheme used in each site,

1 A. Demnati, Social Survey Methods Division, Statistics Canada. R. H. Coats Bldg, 15th Floor, Ottawa, Ontario, Canada, K1A0T6. J. Lawson and J. Young, Macroeconomics Branch(ACACD), Transport Canada. Place de Ville, Tower "C", Floor 25, 330 Sparks Street, Ottawa, Ontario, Canada, K1A 0N6. Bill Harbour, Motor Carrier Policy (ACGC), Transport Canada. Place de Ville, Tower "C", Floor 27, 330 Sparks Street, Ottawa, Ontario, Canada, K1A 0N6.

- the multiplicity of each of the trips and
- the original site counts.

The 1999 NRS survey collects data on more than 250 sites in a network which covers more than 25 thousand kilometres of roads. The network is mostly the National Highway System, augmented by routes of regional importance to trucking. The survey period is one week in order to capture day and hour variation under the following two assumptions,

Assumption 1. *The week of surveying each site is the same.*

In fact, for cost reasons, the survey week is not the same for every site. To reduce the effect of this cost constraint, we assume that the week variation is very small during the survey period. This is a reasonable assumption during a small period of time.

Assumption 2. *The trips window is covered by the survey window.*

In fact, some trips windows may not be contained by the survey window. The trips window was not captured for cost reasons too, but the intention is to do it for future surveys.

2. THE METHODOLOGY

As in the usual finite population sampling situation, we denote the set of all trips during a fixed period of time(one week in case of the NRS) as a finite population $P=\{1,\ldots,N\}$ of size N. Elements of the population (trips) are labelled $i=1,\ldots,N$. Let Y denote the variable of interest, and y_i be the value for Y for the i^{th} population element. Suppose that the network is composed of a set of C nodes, labelled as $1,\ldots,C$. Given C nodes in the network, there are $R=C(C-1)$ possible origin-destination (OD) pairs. Let G be the number of directed links in the network and $N_{g=kl}$ represents the traffic count on directed link g from node k to node l (say). In the census situation in which case all trips are observed, the list of all trips in site g (directed link g) may be viewed as a sampling frame F_g. The frame is most probably incomplete but the union of the G frames, corresponding to the G directed links, cover the entire population. The G frames overlap.

Let J_{gi} the frame or site membership indicators where

$$J_{gi}=\begin{cases}1 & if\ i\in F_g\\ 0 & if\ i\notin F_g\end{cases}$$

with $J_{g+}=N_g$ the size of the frame g, F_g.

The population total of the characteristic Y can be written as

$$Y=\sum_{g=1}^{G}\sum_{i}^{N} J_{gi}\ y_i\ p_{gi}$$

where p_{gi} are constants satisfying the constraints $\sum_{g=1}^{G} J_{gi}\ p_{gi}=1$. For example $p_{gi}=m_i^{-1}=J_{+i}^{-1}$ where $m_i=J_{+i}$ is the multiplicity of the unit i, i.e. the total number of frames reporting the unit i. Note that this parameter requires the multiplicity of every unit in the sample, but does not require matching of units to eliminate duplicate records.

Random samples are drawn from each frame (from each directed link). The basic weight is simply the inverse of the sampling inclusion probabilities

$$d_{gi}^{(0)}=J_{gi}\ I_{gi}\ \pi_{gi}^{-1}$$

where

I_{gi} is the sample membership indicator for unit i in frame F_g and $\pi_{gi}=\Pr(I_{gi}=1/J_{gi}=1)$ is the conditional probability of selection of unit i from frame g.

At the survey sites where sampling is not conducted throughout the entire survey period, an adjustment is made to represent the total traffic of the survey period

$$d_{gi}^{(1)}=d_{gi}^{(0)}\frac{N_g}{\hat{N}_g}$$

where $\hat{N}_g=\sum_i d_{gi}^{(0)}$.

As in the census case, a second adjustment has to be made to account for the multiplicity of the trips when aggregating individual sites.

$$d_{gi}^{(2)}=d_{gi}^{(1)}\ p_{gi}$$

where for example $p_{gi}=m_i^{-1}$ in the case of the simple multiplicity estimator (Sirken, 1972) and

$p_{gi} = (J_{gi}\,\pi_g) / (\sum_g J_{gi}\,\pi_g)$ in the case where we want to take the sampling fraction of each site into account (Kalton and Anderson, 1986).

For any order pairs of nodes (i, j) and path p, let Z_{ijp} stand for the number of trips originating at node i and terminating at node j using the path p. Z_{ijp} is the unobservable ODp count between these two nodes using path p. The ODp triples Z_{ijp} and the traffic counts N_g are related through the routine matrix A,

$$N = AZ$$

where A is a $(G \times R)$ zero-one matrix, its rows correspond to the directed links, its columns correspond to the ODp triples and its entries are 1 or 0 according to whether the link does or does not belong to the directed path of the ODp. This relation expresses each link count N_g as the sums of the ODp counts for all routes that include directed link g,

$$\sum_i^N J_{gi} = J_{g+} = N_g$$

This relation may not hold when substituting parameters Z by estimates $\hat{Z}$, which use the weights $d_{gi}^{(2)}$,

$$\sum_s^G \sum_i^N J_{gi}\, d_{si}^{(2)} = \tilde{N}_g \neq N_g$$

where the sum is done over all the set of trips that cross the site g regardless of the site from which the trips were selected. The third and the final adjustment changes the weights $d^{(2)}$ as little as possible to have new weight $d^{(3)}$ while respecting the counts for each site $g, g = 1,\ldots,G$,

$$\sum_s^G \sum_i^N J_{gi}\, d_{si}^{(3)} = N_g$$

The new weights can be expressed as

$$d_{gi}^{(3)} = d_{gi}^{(2)}\, F(x_{gi}'\,\hat{\lambda})$$

where in the case of the raking ratio $F(x) = e^x$.

The LaGrange multipliers λ are determined by solving the calibration equations

$$\sum_s^G \sum_i^N J_{gi}\, d_{si}^{(2)}\, F(x_{si}'\,\hat{\lambda}) = N_g$$

$$\text{for } g = 1, \ldots, G$$

3. CONCLUSIONS

We have presented a new look to weighting traffic network surveys. We used multiple frames methodology. Classical calibration methods are under investigation.

ACKNOWLEDGEMENTS

The authors would like to thanks several people from Transport Canada, Provinces, Territories, Waterloo University, Université Laval, U.S. Border States and Statistics Canada for their valuable contributions to the project.

REFERENCES

Kalton G. and Anderson D. (1986), Sampling Rare Populations , *J. R. Statist. Soc.* A, 149, Part 1, pp. 65-82.

Sirken, M. G. (1972), Stratified Sample Surveys with Multiplicity, *Journal of the American Statistical Association*, 67, 224-227.