Speed Measurement Variations in Time and Space

Annica Isaksson
Department of Mathematics
Linköping University, Sweden

Abstract

A variance analytical model is formulated for observations on average speed at different locations during different twenty-four-hour periods. The components of the variance of such an observation are estimated from some experimental data. The results suggest that the location-to-location variation is much larger than the variation due to twenty-four-hour periods.

1 Introduction

Since 1996, a survey on vehicle speeds in urban areas is conducted annually (during summer) by the Swedish National Road Administration (SNRA). The survey on vehicle speeds (SVS) is part of an evaluation of a Government-initiated program for safe road traffic in Sweden. The parameter of prime interest is the average speed on the local authority road network during the time period of study. A three-stage sampling design is employed. The primary sampling units are population centres and the second-stage units are small areas. The roads in selected small areas are divided into three strata. From each road stratum, a single one-metre road section – a "measurement location" – is selected for observation during a twenty-four-hour period (randomly selected from the time period of study). Hence, the ultimate sampling units (the population elements) are measurement locations during twenty-four-hour periods.

In order to improve the precision of the survey estimates, one may want to increase the sample size in the final selection stage. Two practical strategies for doing this are the following. Either several measurement locations are selected within each road stratum and observed during the same twenty-four-hour period, or the sample of measurement locations is kept unchanged, but each location is observed during more than one twenty-four-hour period. Which of these strategies that results in the largest gain in precision in the survey estimates obviously depends on what source of variability that is the largest – the location-to-location variability within twenty-four-hour periods or the variability between twenty-four-hour periods within locations.

To investigate this, we formulate a variance analytical model for the observations on average speed at different locations during different twenty-four-hour periods. The components of the variance of such an observation are estimated, using data collected in connection with the SVS 1996.

This study is part of a research project, initiated in 1998 and financed by the SNRA, which aims at developing an error profile for the SVS. An error profile "documents what is known in a given survey about each survey operation as a source of error" [3, p. 639]. For details on the SVS, see earlier work by the author [1].

2 Sampling design of the SVS

The SVS is conducted in a sample of measurement locations selected from a master frame of local authority roads in urban areas. The master frame was constructed at the SNRA during 1995-1996 in order to enable a cost-efficient evaluation of the road traffic safety work. The idea of a master frame can briefly be described as follows (see also [2, pp. 478-480]). Initially, a so called master sample of sampling units is selected. For each sampled unit, a frame is prepared. The sample for a particular survey is then selected from this frame, that serves for a longer time period.

A two-stage sampling design was used to select the master sample. In the first stage, a stratified sample of population centres was selected with probabilities proportional to size (*pps*) with replacement. From each population centre in the resulting sample, a stratified sample of small areas was selected. The concept "small area" refers to a specific partition of the Swedish territory into homogeneous geographical areas. From each stratum, one small area was selected either with simple random sampling or with *pps*. The resulting total sample of small areas is the master sample. For each of these areas, a frame of the local authority road network was prepared. The measurement locations (the one-metre road sections) included in these frames constitute the master frame of local authority roads.

For the SVS, the measurement locations in each master sample area were divided into three road strata: main roads with speed limit 50 kilometers per hour (km/h), main roads with speed limit 70 km/h and other roads. In 1996, one measurement location was randomly selected from each stratum.

The resulting total sample included about 850 locations. Each location was observed during one twenty-four-hour period randomly selected from the time period of study.

3 Data collection and processing

In each selected location, measurements are made using two pneumatic tubes stretched across the road in parallel, with a fixed distance, and connected to a traffic analyzer. When a wheel crosses a tube, this changes the air pressure in the tube. The times of such events, the so called pulses, are registered by the traffic analyzer. The analyzer further combines the pulses into vehicles and calculates their travel direction, speed and type (such as passenger car or lorry with or without trailer). In the combination process, some new pulses are fabricated and some registered pulses eliminated. The variables of study, the number of passing vehicles and the average time used for their passages, are calculated from the vehicle data produced by the traffic analyzer.

This method of measuring traffic is well-tried at the SNRA and known to work well under most circumstances. There are however certain traffic situations, such as vehicles lining up, that have the potential to cause measurement problems. Quality measures such as the 'measurement efficiency' (the proportion of registered pulses that have been combined into vehicles) are then useful for detecting data of bad quality.

4 A variance analytical model

A prime goal of the SVS is to estimate the average speed for the population of study, U, defined as the Swedish local authority road network during the time period of study. We think of the road network as (hypothetically) partitioned into one-metre road sections or measurement locations, and the time period of study as partitioned into twenty-four-hour periods. The average speed, S, is defined as the ratio of two population totals,

$$S = \frac{\sum_U d_{ij}}{\sum_U t_{ij}} \tag{1}$$

where d_{ij} equals the number of vehicles that pass measurement location i during twenty-four-hour period j, and t_{ij} equals the time used for their passages.

Now consider the ratio of the variable values for element (i, j) within road stratum h; $y_{hij} = d_{hij}/t_{hij}$. This is the average speed on the "element-level" and can be shown to equal the harmonic mean of the passing vehicles' speeds. A statistical model for the observed average speed in measurement location i within road stratum h during twenty-four-hour period j, Y_{hij}, can be written as

$$Y_{hij} = \mu + \gamma_h + \tau_i + \beta_j + (\gamma\tau)_{hi} + (\gamma\beta)_{hj} + (\tau\beta)_{ij} + (\gamma\tau\beta)_{hij} + \epsilon_{hij} \quad (2)$$

where μ is an overall mean, γ_h is the effect due to the hth road stratum, τ_i is the effect due to the ith measurement location, β_j is the effect due to the jth twenty-four-hour period, $(\gamma\beta)_{hj}$, $(\gamma\tau)_{hi}$, $(\tau\beta)_{ij}$ and $(\gamma\tau\beta)_{hij}$ are the two-factor and three-factor interaction effects, and ϵ_{hij} is an experimental error. We assume that γ_h, τ_i, β_j, $(\gamma\tau)_{hi}$, $(\gamma\beta)_{hj}$, $(\tau\beta)_{ij}$, $(\gamma\tau\beta)_{hij}$ and ϵ_{hij} are independent random variables with zero means and variances σ^2_γ, σ^2_τ, σ^2_β, $\sigma^2_{\gamma\tau}$, $\sigma^2_{\gamma\beta}$, $\sigma^2_{\tau\beta}$, $\sigma^2_{\gamma\tau\beta}$ and σ^2 respectively. Since y_{hij} now is regarded as a random variable, the y is capitalized.

We will focus on two alternative formulations of Model (2). The first reformulation is

$$Y_{hij} = \mu + \gamma_h + \tau_{i(h)} + \beta_{j(hi)} + \epsilon_{(hij)} \quad (3)$$

where $\tau_{i(h)} = \tau_i + (\gamma\tau)_{hi}$ is the effect due to the ith measurement location nested within stratum h and $\beta_{j(hi)} = \beta_j + (\gamma\beta)_{hj} + (\tau\beta)_{ij} + (\gamma\tau\beta)_{hij}$ is the effect due to the jth twenty-four-hour period nested within stratum h and location i.

The second reformulation is

$$Y_{hij} = \mu + \gamma_h + \beta_{j(h)} + \tau_{i(hj)} + \epsilon_{(hij)} \quad (4)$$

where $\beta_{j(h)} = \beta_j + (\gamma\beta)_{hj}$ is the effect due to the jth twenty-four-hour period nested within stratum h and $\tau_{i(hj)} = \tau_i + (\gamma\tau)_{hi} + (\tau\beta)_{ij} + (\gamma\tau\beta)_{hij}$ is the effect due to the ith measurement location nested within stratum h and twenty-four-hour period j.

We are interested in comparing the magnitudes of $\sigma^2_{\tau+\gamma\tau+\tau\beta+\gamma\tau\beta}$ and $\sigma^2_{\beta+\gamma\beta+\tau\beta+\gamma\tau\beta}$. Their proportion tells us which source of variation that has the largest impact on the total variation of speed measurements within road strata: the location-to-location variation within twenty-four-hour periods or the variation between twenty-four-hour periods within measurement locations. In the next section, we will describe the design of an experiment conducted in order to estimate these variances.

Remark 1 *If the total sample design of the SVS (as described in Section 2) is taken under consideration, the road stratum effect γ_h in Model (2) can be* *viewed as representing the sum of several effects; some random (the effects due to the random selection of population centres and small areas), other fixed (the effects due to stratification in each selection stage). The remaining effects included in Model (2) are nested within population centre stratum, population centre, small area stratum and small area. However, to keep things simple, we regard γ_h as a single random effect and omit the nesting indices.*

5 Experimental design

In most road traffic safety surveys conducted by the SNRA, estimates are wanted not only for the whole population of study but also for the seven SNRA regions. Therefore, these are treated as strata in the first selection stage of the master sample. Our experimental data was collected within the southernmost region, the Skåne region, in connection with the SVS 1996. Three experiments, hereafter referred to as Experiment 1 up to 3, were conducted. The design of each of them is described below.

Experiment 1 was designed mainly to enable estimation of the variance due to twenty-four-hour periods within road strata and measurement locations, $\sigma^2_{\beta+\gamma\beta+\tau\beta+\gamma\tau\beta}$. The experiment comprised 29 locations, subsampled from the SVS sample of measurement locations in Skåne. Each subsampled location was observed twice: during the twenty-four-hour period that was scheduled for the SVS and again during another twenty-four-hour period, randomly chosen from the time period of study. The statistical model for this design equals Model (3), except that γ_h and $\tau_{i(h)}$ are undistinguishable. This follows since only one location was observed within each road stratum.

Experiment 2 was designed mainly to enable estimation of the variance due to measurement locations within road strata and twenty-four-hour periods, $\sigma^2_{\tau+\gamma\tau+\tau\beta+\gamma\tau\beta}$. The experiment comprised 18 twenty-four-hour periods, subsampled from the set of twenty-four-hour periods that were scheduled for observation within the SVS in Skåne. During each subsampled twenty-four-hour period, observations were made at two measurement locations: at the SVS measurement location and at another random location within the same small area and road stratum. It was demanded that both locations had the same speed limit. The statistical model for this design equals Model (4), except that γ_h and $\beta_{j(h)}$ are undistinguishable. This follows since only one twenty-four-hour period was observed within each road stratum.

Experiment 3 was aimed for confirmation. It was designed mainly to enable estimation of the total variance within road strata, $\sigma^2_{\tau+\beta+\gamma\tau+\gamma\beta+\tau\beta+\gamma\tau\beta}$. The experiment comprised 25 road strata in the Skåne region. Within each of these strata, two observations were made: one at the SVS measurement

Source of Variation	Degrees of Freedom	Sum of Squares	Expected Mean Square
A	$a-1$	$2\sum_{i=1}^{a}\left(\bar{Y}_{i.}-\bar{Y}_{..}\right)^2$	$\sigma^2+\sigma^2_{B(A)}+2\sigma^2_A$
B within A	a	$\sum_{i=1}^{a}\sum_{j=1}^{2}\left(\bar{Y}_{ij}-\bar{Y}_{i.}\right)^2$	$\sigma^2+\sigma^2_{B(A)}$
Error	0	0	0

Table 1: ANOVA Table for a Two-Stage Nested Design.

location during its scheduled twenty-four-hour period and the other at another random location within the same small area and road stratum during another twenty-four-hour period, randomly chosen from the time period of study. It was demanded that both locations had the same speed limit. The statistical model for this design equals Model (3), except that $\tau_{i(h)}$ and $\beta_{j(hi)}$ are undistinguishable. This follows since each location and twenty-four-hour period was observed only once.

The data collection method used for the experiment is the same as the one used in the regular survey. Measurements with a 'measurement efficiency' below 80 percent were excluded from the experiment. Measurements with an 'imputation rate' below 80 percent were excluded as well (the imputation rate being defined as the proportion of observed vehicles that have been formed entirely or partly from fabricated pulses). When a measurement was excluded in accordance with these rules, the entire "observation pair" was excluded.

6 Data analysis

We are interested in estimating the variances associated with each experimental model. We will use a simple estimation procedure, the "analysis of variance method", which consists of equating the expected mean squares to their observed values in the ANOVA table and solving for the variances. No normality assumptions are required.

Note that the design of each experiment can be viewed as a two-stage nested design. The ANOVA table for such a design, with a levels of a random factor A, two levels of a random factor B nested under each level of A and *one replicate*, is given in Table 1. The table can be made applicable to each experiment by the use of Table 2. The ANOVA tables for our data are given in Appendix A. The data include all types of observed vehicles.

The reason why we only made one-replicate experiments is that only one observation can be made at a given measurement location at a given time

Experiment	a	A	B within A
1	29	Road strata and locations (within road strata)	24-hour periods (within road strata and locations)
2	18	Road strata and 24-hour periods (within road strata)	Locations (within road strata and 24-hour periods)
3	25	Road strata	24-hour periods and locations (within road strata)

Table 2: The Factors in Each Subexperiment.

point with the data collection method in use. As a consequence, we cannot uniquely estimate σ^2_A, $\sigma^2_{B(A)}$ and σ^2. But, the error variance σ^2 is probably very small compared to the others. Under the assumption that $\sigma^2 = 0$, we get the analysis of variance estimators $\hat{\sigma}^2_A = \left(MS_A - MS_{B(A)}\right)/2$ and $\hat{\sigma}^2_{B(A)} = MS_{B(A)}$. Now, from our data, we obtain the following variance estimates:

$$
\begin{array}{lrclrcl}
\text{Experiment 1:} & \hat{\sigma}^2_{\gamma+\tau+\gamma\tau} & = & 114.4, & \hat{\sigma}^2_{\beta+\gamma\beta+\tau\beta+\gamma\tau\beta} & = & 1.8 \\
\text{Experiment 2:} & \hat{\sigma}^2_{\gamma+\beta+\gamma} & = & 100.2, & \hat{\sigma}^2_{\tau+\gamma\tau+\tau\beta+\gamma\tau\beta} & = & 27.9 \\
\text{Experiment 3:} & \hat{\sigma}^2_{\gamma} & = & 43.0, & \hat{\sigma}^2_{\tau+\beta+\gamma\tau+\gamma\beta+\tau\beta+\gamma\tau\beta} & = & 37.3
\end{array}
$$

We see that $\hat{\sigma}^2_{\tau+\gamma\tau+\tau\beta+\gamma\tau\beta}$ is much larger than $\hat{\sigma}^2_{\beta+\gamma\beta+\tau\beta+\gamma\tau\beta}$ and that $\hat{\sigma}^2_{\tau+\beta+\gamma\tau+\gamma\beta+\tau\beta+\gamma\tau\beta}$ is larger than both of them taken together. Appearantly, within road strata, the location-to-location variation within twenty-four-hour periods is larger than the variation due to twenty-four-hour periods within measurement locations. Further, these two sources of variation make up most of the total variation of speed measurements within road strata.

7 Conclusions and final remarks

Our study suggests that, in order to improve the precision of the survey estimates, one should increase the number of measurement locations within road strata rather than the number of observed twenty-four-hour periods for each location. No costs were taken into account in this study. We suspect however, that the suggested strategy also is the cheaper of the two, since it requires fewer visits by the field staff in each road stratum.

In the SVS, some sampled measurement locations are regarded by the field staff as difficult or impossible to observe. For example, cars parked by the roadsides may obstruct the installation of the pneumatic tubes. Such problematic locations are immediately replaced by geographically neighboring locations and observed during the scheduled twenty-four-hour periods. In other cases, measurements from sampled locations are dismissed due to low 'measurement efficiency'. Such locations are re-measured by a later date. These procedures correspond to two different kinds of so called field substitution. The impact of field substitution on the survey estimates has not yet been investigated. However, the latter type of substitution probably has the least impact on the estimates. This follows since, according to our results, the variation due to twenty-four-hour periods is smaller than the location-to-location variation.

The total variance of the estimator of S can be viewed as composed of the variances arising from the sample selection in time (the selection of twenty-four-hour periods) and in space (the selection of population centres, small areas and measurement locations). One possible continuation to this study is to derive the theoretical expressions for the variances due to time and space.

References

[1] A. Isaksson, *En studie av hastigheter och tidluckor 1996 – Teknisk rapport (A survey on speeds and time gaps, 1996 – Technical report)*, SNRA Report 1997:85, Swedish National Road Administration, SE-781 87 Borlänge, Sweden, 1997. (In Swedish).

[2] L. Kish, *Survey Sampling*, Wiley, New York, 1965.

[3] C.-E. Särndal, B. Swensson, and J. Wretman, *Model Assisted Survey Sampling*, Springer, New York, 1992.

A ANOVA tables for the experimental data

Experiment 1

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Road strata and locations (within road strata)	28	6457.2	230.6
24-hour periods (within road strata and locations)	29	50.7	1.7

Experiment 2

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Road strata and 24-hour periods (within road strata)	17	3880.0	228.2
Locations (within road strata and 24-hour periods)	18	502.1	27.9

Experiment 3

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Road strata	24	2959.1	123.3
24-hour periods and locations (within road strata)	25	932.6	37.3