

Suggested approach for weighting sample data and calculation of statistics

Version 4.4 – 28 June 2024

Suggested approach for weighting sample data and calculation of statistics

Please refer to this report as follows:

Elgner, J., Andrášik, R., Bijleveld, F., Boets, S. and Van den Berghe, W. (2024). Suggested approach for weighting sample data and calculation of statistics. Note produced as part the Trendline project, supported by the European Union.

Project details:

Project start date:	15/10/2022
Duration:	36 months
Project name:	TRENDLINE
Coordinator:	Wouter Van den Berghe & Agnieszka Stelling SWOV Institute Research for Road Safety Henri Faasdreef 312 – 2492 JP The Hague, The Netherlands
	Co-funded by the European Union

Deliverable details:

Version:	4.4
Dissemination level:	Public
Submission date:	28/06/2024

Authors:

Elgner, J., Andrášik, R. (CDV – Transport Research Centre), Czechia
Bijleveld, F., Van den Berghe, W. (SWOV – Institute for Road Safety Research), The Netherlands
Boets, S. (Vias institute), Belgium

Legal Disclaimer

All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user, therefore, uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission and INEA has no liability in respect of this document, which is merely representing the authors' view.

Table of contents

About TRENDLINE	4
A. Introduction	5
B. Sampling approach	6
C. Data description.....	7
D. The case of several vehicle types, road users or further breakdowns within the stratum	10
E. Weighting the data	11
F. Calculation of KPI characteristics	13
G. Providing the roadside measurement data	15
G.1. <i>Determining the weight of the sessions in the stratum.....</i>	<i>15</i>
G.2. <i>Determining the weight of the strata in the total sample.....</i>	<i>17</i>
H. Using a survey package – example in R.....	19
I. Summary	22
Further reading	23

About TRENDLINE

Trendline brings together 29 European countries (25 EU Member States and 4 countries as observers) for data collection, data analysis, delivery of road safety KPIs and for using these within road safety policies. Trendline is co-funded by the European Union and builds on the experience gained in the Baseline project. KPIs – Key Performance Indicators – are indicators that provide information about factors that are associated with crash and injury risks. At the core of Trendline project are eight KPIs:

Indicator	Definition
Speed	Percentage of vehicles travelling within the speed limit
Safety belt	Percentage of vehicle occupants using the safety belt or child restraint system correctly
Protective equipment	Percentage of riders of powered two wheelers and bicycles wearing a protective helmet
Alcohol	Percentage of drivers driving within the legal limit for blood alcohol content (BAC)
Distraction	Percentage of drivers NOT using a handheld mobile device
Vehicle safety	Percentage of new passenger cars with a Euro NCAP safety rating equal or above a predefined threshold
Infrastructure	Percentage of distance driven over roads with a safety rating above an agreed threshold
Post-crash care	Time elapsed in minutes and seconds between the emergency call following a collision resulting in personal injury and the arrival at the scene of the collision of the emergency services

These 8 KPIs originate from the Commission Staff Working Document 'EU Road Safety Policy Framework 2021-2030 - Next steps towards "Vision Zero" SWD (2019) 283 final.' In addition, some new experimental and complementary indicators will be tested within Trendline (provisional names):

- Driving under the influence of drugs
- Share of 30km/h road lane lengths in urban zones
- Red-light negotiations by road users
- Compliance with traffic rules at intersections
- Helmet wearing of PMD (Personal Mobility Devices) riders
- Self-reported risky behaviour
- Attitudes towards risky behaviour
- Use of lights by cyclists in the dark
- Enforcement of traffic regulations
- Alternative speeding indicators.

For each of the original eight KPIs and the experimental KPIs, a 'KPI Expert Group' (abbreviated as KEG) has been established. Their main role is to draft the common methodological guidelines, to give feedback on questions, and to review the report of the KPI which they are covering.

Website Trendline: <https://www.trendlineproject.eu/>

A. Introduction

Within Trendline, several of the “KPIs” (Key Performance Indicators) refer to the relative number of vehicles or road users that respect certain legal limits and rules. These are sometimes called the “behavioural” KPIs. They refer to speeding, driving under the influence of alcohol, use of protective equipment, wearing a seatbelt or distraction. There are also a number of behavioural KPIs among the new experimental indicators.

In general, it is impossible to measure the performance of all vehicles or road users. Therefore, the KPI values need to be estimated based on a sample of vehicles and/or road users observed or surveyed. The main aim of these estimates is to estimate the percentage of kilometers driven on the entire road network (over a period, which could, for instance, be set to one year) by vehicles respecting the legal limits and rules. In terms of the sampling approach, this means that the statistical population to be considered is the total traffic volume (typically expressed in kilometers driven) of moving vehicles over a certain geographical area (i.e. country or region) over a certain period (e.g. one year). Estimates are made by sampling individual vehicles (or road users) at particular locations and moments in time.

For many KPIs within the Trendline project, data is being collected during observations at different locations (e.g., for distraction by mobile phone) or surveys (e.g., for driving under the influence of alcohol). Sampling on three different road types is required (motorways, rural roads, urban roads) for all behavioural KPIs. For some KPIs sampling of different time periods and/or vehicle types is also required.

This document aims at guiding the Trendline participants through the different steps of weighting the roadside measurement data for calculating KPI values and confidence intervals (CIs), preparing them for complex survey analysis and assigning weights to each observation. Section B mentions the sampling approach, including the selection of locations and the basic stratum for a given road type and time period. Attribute descriptions for data preparation are further detailed in Section C. Detailed attribute descriptions and the weight algorithm are covered in Section D, with essential distinctions between input attributes. The algorithm for computing weights is described in section E. Section F focuses on theoretical aspects of the complex survey approach, and on computing expected KPI values and CIs. Section G connects templates with an example dataset, while Section H demonstrates data analysis using the Survey package in R, as introduced in Section F. Finally, Section I offers a concluding summary.

B. Sampling approach

Sampling is done in 2 steps:

- 1) Random selection of observatiuon locations for each road type considered. Most Trendline beneficiaries use the same number of locations for each road type.
- 2) Random selection of vehicles/road users (nested) in each session.

The minimum number of locations for observations or surveys in Trendline is 10 per road type. A set of observations at a particular location and at a particular moment of time is called an "observation session" or a "session". At a given location, there may be several observation sessions. If different time periods (e.g. days and nights) are required in the sampling, then time periods should be linked to locations in a balanced way; in such cases, a minimum of 10 locations per time period is required as well as minimum 2 locations for each combination of road type and time period. The data collected during sessions allows to calculate a KPI value for that session and, if sufficient data are available, also for subcategories (e.g., male/female; position in the car, type of vehicle). Moreover, for every session at least the road type is coded:

- motorways,
- urban roads,
- and rural roads.

These are the generally required minimum sampling strata for the behavioural KPIs.

For most behavioural KPIs also a time period is coded for the observation session, specifically:

- Weekday
- Weekend

For the KPI alcohol, four time periods are considered (Weekday daytime, Weekday nighttime, Weekend daytime, Weekend nighttime). For some KPIs (e.g., distraction), only one time period is considered (weekday daytime).

Each combination of road type and time period should be considered as a separate sampling stratum: a combination of 3 road types and 2 time periods would lead to 3 x 2 or 6 strata.

Calculating KPIs for crossed strata of road type x time period is generally not a minimum requirement, but it is recommended, in particular, if these categories have been a part of a sampling strategy. For such strata to include sufficient and sufficiently reliable data, a minimum requirement is that for each stratum (combination of road type and time period) a minimum of 2 different locations are used (but more are recommended).

C. Data description

The following steps should be followed for each of the strata (6 in the example of three road types and two time periods). Let us suppose that there are K survey sessions in the stratum. For instance, there may be 6 observation sessions for observations on urban roads during weekdays. In that case, $K = 6$ for that stratum.

For each survey session k (with k varying between 1 and K), the traffic count(s) need to be determined. The traffic count obtained may concern all vehicles (or vehicles of a certain type) that passed by during the entire observation session, or for a fraction of the period (e.g. for 20 minutes in the middle of the session or for 10 minutes before and 10 minutes after the session; in the KPI guidelines the minimum requirement is to have at least a 10 minutes count per session, but 20 minutes is preferable). The duration of the counting is important. Please register both the actual count of the number of relevant vehicles and the time used to count. In case you have grounds to believe that the traffic density during the observation/survey session is quite different from the density during the counting session (e.g. because there was a sudden traffic jam causing much fewer vehicles to pass by during the observation, or because there was a bridge opening during the counting session), it is also useful to make an estimate of the number of relevant vehicles that passed by during the survey session. This estimate is somewhat redundant but would allow for unique unexpected situations.

Often, it is planned that all observations or survey sessions have the same duration (e.g., 60 minutes). This can be considered as the "standard duration" of a session. So, for the session k in the stratum i the following data is recorded:

Table 1 Description of variables selected for weighting

Variables to be collected at session level	
Duration (sec) of the period used to count passing vehicles, $t(k) = \sum_{l=1}^L t_l(k)$, where l indicates counting periods during 1 observation session e.g. $L = 2$ if there is a counting period before and a counting period after the observation session.	$t(k)$, $t_l(k)$
Number of passing (relevant) vehicles of j vehicle type counted during the counting period, $n_j(k) = \sum_{l=1}^L n_{jl}(k)$, where l indicates counting periods during 1 observation session, e.g. $L = 2$ if there is a counting period before and a counting period after the observation session.	$n_j(k)$, $n_{jl}(k)$
Duration of the observation session (sec)	$T(k)$
Estimated total number of (relevant) passing vehicles by vehicle type j during the observation session, usually ¹ this is equal to $n_j(k) \times T(k) / t(k)$	$ETN_j(k)$
Number of (relevant) vehicles of vehicle type j surveyed during the observation session	$N_j(k)$

¹ In exceptional cases where the traffic during the counting session is not representative for the traffic during the observation session, one should use the best estimate $N_{ik}(k)$ (i.e. estimate of the total number of (relevant) passing vehicles 'per hour' during the observation session).

Variables to be collected at stratum level	
The characteristic at the stratum level by vehicle type j of stratum i	Sl_{ij}
Relative traffic volume by vehicle type j of stratum i	$TR_j(i)$
Total road length (km) of stratum i	$RL(i)$
Average number of vehicles per hour (derived from count period data by vehicle type j of stratum i)	$NS_j(i)$
Relative length of the considered time periods of stratum i	$TP(i)$
Fixed speed limit (km/h) of stratum i	$FSL(i)$

In Table 1 capital letters mainly refer to the observation session whereas lowercase letters denote the counting period (calibration). It is important to have an accurate estimate of the total number of vehicles that passed this survey location during a session (this is $ETN_j(k)$). Otherwise, we do not know what share the individual survey sessions have within the stratum.

From a policy perspective, it is often useful to aggregate the results from the strata. This may be necessary, for instance, to calculate a national indicator considering all road types, regions (optional), time periods (if several are considered) and vehicle types (if several are considered). This is often required within Trendline.

In order to aggregate the strata, the relative importance of each stratum needs to be assessed. Within Trendline, the relative importance is based on the (estimated) traffic volume in each stratum. If the first stratum represents (or is representative for) 50% of traffic volume, the second represents 30% and the third 20%, then the aggregated KPI value should be based on 50% of the KPI value obtained for the first stratum, plus 30% of the KPI value of the second stratum and 20% of the KPI value of the third stratum.

It is recommended to ask country experts (e.g. Ministry of Transport) for estimates of traffic volume by region, road type and time period for each considered vehicle type (ideally estimates for all crossings of the considered strata). There are two possible ways to account for the relative importance of traffic volume and, hence, to determine or estimate this:

- **Version 1 - traffic volume data (recommended)**

National data on traffic volume (vehicle kilometers driven) by type of vehicle (j), type of road and time period (if different periods are considered) is represented by $TR_j(i)$ (Table 5). In the ideal situation national traffic volume data is available for all considered strata but possibly this information has to come from combining different sources. It is also possible that data is only available for certain, but not all strata (e.g., no indication of national traffic volume according to the considered time periods).

Information on traffic volume can come from different sources, such as national counts on roads for proportions on the road types. Representative online mobility survey data may be available for the relative proportions according to time period (e.g. consultation of the Ministry of Transport). If distributed traffic volume data are available by road type, then the distribution of traffic volume over different considered time periods can be estimated by using as a proxy the relative proportion or number of hours of the time period (e.g. weekday daytime: 48%, weekday night: 19%, weekend day: 19% and weekend night: 14%). These proportions should then be combined in a logical way (e.g. multiplication) with the traffic volume proportions by

road type to calculate a percentage of the traffic volume for each crossed stratum, all summing up to 100%.

If no traffic volume data for the considered strata are readily available, it is advised to consult an expert for estimates (e.g. Ministry of Transport).

- **Version 2 - combinations of other variables**

If traffic volume information is not available (nor can be estimated) but a reliable estimate of the length of the roads of each road type is available (e.g. based on expert consultation: GIS department or Ministry of Transport), one could alternatively use the traffic counts from the sessions in the stratum to make an estimate of the hourly number of vehicles at the survey locations ($=NS_j(i)$). If the locations are randomly selected, this average (time-standardized) vehicle count estimates the average hourly vehicle count of all locations in the stratum i . This value, multiplied by the estimate of the length of the roads in the stratum i – and, if different time periods are considered, the number of hours in the time period considered – should give some estimate of the traffic volume in the stratum. These values could then be used to weight the strata. The validity of the counts for traffic volume estimates increases if: (1) the locations are randomly selected (including also low volume locations), covering the geographical area of the country, and (2) if sessions are planned at mixed time intervals during the considered time period strata, e.g. including peak as well as off-peak hours on weekdays, to cover the variability of traffic volume within the relative length of considered time periods.

Note that $NS(i)$ is the average number of passing vehicles per hour on the road type (e.g., urban roads) and within the time period (e.g., weekdays) the stratum (i) represents. $NS_j(i)$ is equal to the mean of all $NS_j(k)$ – for K sessions – in the stratum i . Examples of the described figures are available in tables 6–8 (template sheets *4.Roadside_RawData-3level* and *5.Roadside_RawData-1level*).

When reporting results to the project coordinator, Trendline beneficiaries need to report, for each stratum used in the analysis, an estimate of the traffic volume (or at least percentual share of it), since this is a key element in assuring respect for minimal requirements for weighting and to assure internationally comparable results. The KPI guidelines and templates include the informative variable “weighted proportion”, which is not part of the weight algorithm. Weighted proportion indicates the distribution of the volume of collected vehicles according to the road length and time period per stratum, allowing for data verification.

Important: if no vehicle counts or no road length information are available, and no otherwise obtained (actual or estimated) traffic volume information can be provided, one should only treat the strata separately and defer from aggregation. In such cases, some of the minimum required KPIs in Trendline cannot be delivered.

D. The case of several vehicle types, road users or further breakdowns within the stratum

For some KPIs it is desirable or even required to distinguish between several vehicle types and/or road user groups (j). The logic discussed above should be applied to each considered subpopulation.

However, this supposes that you can also count these different subpopulations during the traffic count in each session. If that is not possible, then you could assume that the distribution of the subpopulations in the traffic passing by is the same as that of the vehicles or road users observed/surveyed. This assumption is usually justified as the general rule during the fieldwork is to observe (or survey) the first arriving vehicle after coding the former one (random sampling - no deliberate over- or under-sampling of a specific vehicle/road user type).

This means that $ETN_j(k)$ has to be adapted accordingly and a value of $N_j(k)$ by considered vehicle type j has to be used.

Other variables like age category and gender are generally not specified sampling strata in behavioural measurements on the road but variables of the surveyed (sub)populations². If there is a need, for instance, to distinguish between male and female drivers, then the same assumption applies: the relative number of females in the set of the observed vehicles is the same in the set of the vehicles passing by.

² In questionnaire surveys age and gender are sampling strata - so there it makes sense to weight according to population statistics. But this is not the case in roadside surveys.

E. Weighting the data

Once the sampling strategy has been fixed, the question arises as to how the observations have to be weighed in order for the overall average or percentage to reflect the overall percentage of vehicles or road users complying with the rules in the total population.

There is a need to **weight** the results at the observation locations (sessions) **within** the stratum (to arrive at the best estimate for the KPI value within the stratum) but also **across** the strata (to obtain, for example, a value for all considered time periods and/or for all roads together).

Please note that for certain KPIs, additional breakdowns (other than road type or time period) are also possible (or even required), such as region (optional) and vehicle type/road user (e.g. three minimum required vehicle type for KPI distraction). In such cases, the number of strata that can be considered will be higher. However, in general, strata with less than 500 data points should not be considered for calculating KPIs (unless specified differently in the minimum requirements of the methodological guidelines for the KPI), because the number of different observations and/or observation locations is too small and/or confidence intervals will be too wide. When strata with less than 500 observations are obtained and delivered to the Trendline coordination team, they will be treated differently in the tables and graphs of Trendline reports (e.g., shown in another colour or marked with an asterisk). However, in view of reporting, such strata could be combined with or added to other strata to achieve this minimum. For instance, “weekday daytime” and “weekday nighttime” could be combined to “weekday”.

It is considered acceptable to assume that what is observed among the surveyed vehicles – $N_j(k)$ – is representative for all passing vehicles. Therefore, each surveyed vehicle represents $ETN_j(k)/N_j(k)$ vehicles in a session³. Observation weight within session k by vehicle type j in session k of stratum level i is defined as:

$$W_{ij}(k) = \begin{cases} \frac{ETN_j(k)}{N_j(k)} \cdot Sl_{ij} ; N_j(k) > 0 \\ 0 & ; N_j(k) = 0 \end{cases}, i = 1, \dots, M; j = 1, \dots, J; k = 1, \dots, K, \quad (1)$$

where j represents vehicle type (e.g. $j = 1$ for passenger car, $j = 2$ for goods vehicle, etc.), i indicates basic stratum (e.g. region (optional), road type and time period) and Sl_{ij} denotes the stratum level weight characteristic including either traffic volume estimates $TR_j(i)$ (version 1) or a combination of other variables (version 2) by formula:

$$Sl_{ij} = \begin{cases} \frac{TR_j(i)}{\sum_{i=1}^M TR_j(i)} & ; \text{version 1} \\ \frac{RI(i) \cdot NS_j(i) \cdot TP(i)}{\sum_{i=1}^M RI(i) \cdot NS_j(i) \cdot TP(i)} & ; \text{version 2} \end{cases}, i = 1, \dots, M; j = 1, \dots, J. \quad (2)$$

³ If an observed vehicle represents 4 vehicles in the session, we have just one observation, not four, but it ‘weights’ for four vehicles.

Moreover, extreme values of weights could significantly affect the weighting procedure because the weighted average is sensitive to scale. For some combinations of attributes and low sample sizes, an extreme weight value can significantly influence the expected KPI value. Therefore, it is advisable to reduce the weight values by trimming within the boxplot whiskers at the basic stratum level.

$$W_{ij}'(k) = \begin{cases} Q1 - 1.5 \cdot IQR; & W_{ij}(k) \leq Q1 - 1.5 \cdot IQR \\ W_{ij}(k); & Q1 - 1.5 \cdot IQR < W_{ij}(k) < Q3 + 1.5 \cdot IQR, \\ Q3 + 1.5 \cdot IQR; & Q3 + 1.5 \cdot IQR \leq W_{ij}(k) \end{cases} \quad (3)$$

where $Q1$, $Q3$ and IQR are first, third quartile and the interquartile range, respectively, of W_{ij} over all sessions $k = 1, \dots, K$, of stratum i . The weights should add up to the number of vehicles by vehicle type j that passed on all sessions of the stratum i , had they been identical in duration. These weights are assigned to all individual survey observations as W_x , where $x = 1, \dots, n$ and n is number of observations in dataset (Table 2

F. Calculation of KPI characteristics

Weighted analysis is a standard accompaniment to survey datasets. The weights can be constructed in many ways. However, they should be stable and ensure adequate adjustment for individual roadside measurements relative to count period data and other stratum level characteristics at the region (optional), road type, and time period levels.

Let's look closer at the defining characteristics of the dataset collected from roadside measurements, where the dependent variable of interest ($Compliant_x$, e.g. "use of seatbelt") follows a Bernoulli distribution. The expected value of positive counts – in other words, the KPI value – is calculated as weighted average

$$\hat{p} = \frac{\sum_{x=1}^n W_x \cdot I(Compliant_x)}{\sum_{x=1}^n W_x}, \quad I(Used_x) = \begin{cases} 1; & Compliant_x = TRUE, \\ 0; & Compliant_x = FALSE \end{cases} \quad (4)$$

The expected value \hat{p} represents the KPI value, which is calculated at each level of stratification using formula (4).

The calculation of confidence intervals (CIs) for the data collected in Trendline through observation sessions is far from trivial. The statistical reference works considered do not precisely cover the sampling problem considered and the methods discussed. Some Trendline beneficiaries appeared to have used in Baseline gaussian approximations to statistics to aggregate over sample sessions within strata and to aggregate over strata, but there were also some who used statistical software to take the complex sampling design into account. The most important consideration is that the confidence intervals to be calculated should consider the variation between locations, not between individual vehicles.

Trendline beneficiaries should use a method for calculating CIs that takes the sampling design method into account, in particular the fact that observations are nested in sessions. Trendline beneficiaries need to indicate in the metadata how they calculated the CIs. Since approximations that assume simple random sampling of vehicles clearly lead to unrealistically small confidence intervals, approximations using simple random sampling of vehicles are not acceptable for calculating confidence intervals.

Parker JD et al. (2017, p. 3), propose presentation standards for proportions: "The Clopper-Pearson CI (2) [adapted for complex surveys by Korn and Graubard (3) when applicable] should be used to determine whether or not the CI of a proportion meets the presentation criteria.". Effectively this means using the method proposed in Korn and Graubard (1998) for estimating the confidence interval of proportions.

This confidence interval construction is based on an incomplete beta function, with effective samples size based on the estimated variance of proportion. In applications with a known sampling design, various modifications of CI are possible, e.g. truncating the degrees of freedom adjusted effective sample size. The justification of these truncation procedure is that the design effect may be estimated to be less than one because of instability of the variance estimator. For more details, one can consult Korn and Graubard (1998).

For a $1 - \alpha$ level confidence interval based on a sample of size n , first define the effective sample size by (closely following Korn and Graubard, 1998, p. 194)

$$n. eff = \frac{\hat{p} \cdot (1 - \hat{p})}{\widehat{var}(\hat{p})}, \quad (5)$$

where $\widehat{var}(\hat{p})$ is a variance estimator of \hat{p} . Moreover, the confidence interval is based on the degrees of freedom adjusted effective sample size by

$$n. df = \frac{\hat{p} \cdot (1 - \hat{p})}{\widehat{var}(\hat{p})} \cdot \left(\frac{t_{n-1} \left(1 - \frac{\alpha}{2}\right)}{t_d \left(1 - \frac{\alpha}{2}\right)} \right)^2, \quad (6)$$

whereas $t_d(v)$ is quantile of Student distribution with sample size d and significance level v . When n is large, the $1 - \frac{\alpha}{2}$ quantile of a Normal distribution can be used in place of $t_d(v)$ in (6). The confidence interval is based on beta distribution

$$CI = \left(Q\left(\frac{\alpha}{2}, n. df \cdot \hat{p}, n. df \cdot (1 - \hat{p}) + 1\right), Q\left(1 - \frac{\alpha}{2}, n. df \cdot \hat{p}, n. df \cdot (1 - \hat{p}) + 1\right) \right), \quad (7)$$

where $Q(p, sh1, sh2)$ is p -th quantile of a Beta distribution with two shape parameters (Korn and Graubard, 1998).

It is advised to use dedicated survey software, as readily available in R and other software packages. Tables 2-8 that will be introduced in section G and all other variables needed for the weighting will serve as input to these procedures. Packages that can be considered are:

- R Survey Package <https://cran.r-project.org/web/packages/survey/index.html>
- STATA Analysis of Complex Survey Data in Stata e.g. https://www.stata.com/meeting/mexico10/mex10sug_canette.pdf
- SPSS: <https://www.ibm.com/products/spss-statistics/complex-samples>
- SAS: <https://support.sas.com/documentation/cdl/en/statug/63033/PDF/default/statug.pdf> (extensive document including documentation on survey methods)

Please note that some observation methods allow to observe all passing vehicles, for example when measuring speed by using induction loops or radars, or when the traffic volume is low enough to count all passing vehicles (e.g. helmet use by motorcyclists). In such cases the formulations can be simplified, because there is only one stage of sampling, i.e. stratified sampling of observation locations. However, one needs to be 100% sure that all traffic has been counted, and also in this case it is strongly recommended to use a statistical package that can accommodate the sampling design, rather than using the complex formula for calculation of confidence intervals (which can lead to typing or programming errors).

G. Providing the roadside measurement data

Trendline beneficiaries have the option to submit raw data using templates. If raw data is provided, the Trendline coordination team will process the results and ensure consistency, thereby saving effort for Trendline beneficiaries (who do not need to fill out the template sheets "6.Roadside_Results-aggr" and "7.Roadside_Results-semiaggr" which is optional). Otherwise, Trendline beneficiaries must follow the algorithms outlined in this document according to specific KPI guidelines and minimum requirements. In any case, meticulous completion of the templates and Metadata sheet is important.

The templates to be used for the collection of raw roadside measurement data contain five sheets, along with two additional sheets related to results at the aggregate or semi-aggregate level. Now it is possible to create a database table or a spreadsheet with columns including attributes mentioned in KPI guidelines (*Table 2*), which meet at least minimum requirements. *Tables 3-8* contain supplementary to the database table, regarding count period data (calibrations) and other characteristics at the basic stratum level, which are used in the weight calculation. Data can be partially anonymized through the attributes Region ID (optional), Session ID, Location ID, and Vehicle ID, where a unique code linked to all source tables is sufficient. For the source data, it is not necessary to provide the exact time of passed vehicles.

G.1. Determining the weight of the sessions in the stratum

The first sheet includes fields for raw data entry, and the other four sheets contain supplementary tables for estimating weights according to the weight algorithm. For instance, the observations of using a seatbelt in a survey could be ordered in the way as indicated in *Table 2* below (the other variables would concern the position of the person, whether he/she is driver or not, sex, ...).

Table 2 Source data to be collected per observation, example on KPI seatbelts & CRS
(template sheet 1.Roadside_RawData-7level)⁴

Region ID (optional)	Road type	Time period	Session ID	Location ID	Vehicle ID	Vehicle type	Seat Position	Gender	Age group	Used
ro00001	rural road	weekday/daytime	s000001	l000001	v000001	goods vehicle	Driver	male	25-64	TRUE
ro00001	rural road	weekday/daytime	s000001	l000001	v000002	passenger car	Driver	female	18-24	TRUE
ro00001	rural road	weekday/daytime	s000001	l000001	v000003	goods vehicle	Rear passenger	male	18-24	FALSE
...										
ro00001	Motorway	weekend/daytime	s000120	l000119	v034126	passenger car	Driver	female	65+	TRUE
ro00001	Motorway	weekend/daytime	s000120	l000119	v034127	passenger car	Driver	male	25-64	FALSE
ro00001	Motorway	weekend/daytime	s000120	l000119	v034127	passenger car	Front passenger	male	18-24	FALSE
...										

Per session the KPI value is calculated as the average proportion from all observations using the formula (4).

Table 3 is extended with the weight attribute based on Vehicle type ($n_j(k)$): number of vehicles of vehicle type j counted during session k . Table 4 contains the time duration of observational and count period measurements for each session. The variables mentioned in Tables 3 and 4 are further used to estimate of the total number of vehicles ($ETN_j(k)$) that passed the survey location during a session. See section C for more details

Table 3 Example number of vehicle divided by vehicle type in each session from count period data
(template sheet 2.Roadside_RawData-6level_1)

Region ID (optional)	Road type	Time period	Session ID	Location ID	Vehicle type	$n_j(k)$
ro00001	rural road	weekday/daytime	s000001	v000001	passenger car	93
ro00001	rural road	weekday/daytime	s000001	v000001	goods vehicle	3
...						
ro00001	motorway	weekend/daytime	s000120	l000119	passenger car	210
ro00001	motorway	weekend/daytime	s000120	l000119	goods vehicle	0
...						

⁴ The sheet title in parentheses for tables 1-7 corresponds to the template in excel file.

Table 4 Example of time duration in each session from observational and count period data (template sheet 3.Roadside_RawData-6level_2)

Region ID	Road type	Time period	Session ID	Location ID	T(k)	t(k)
r000001	rural road	weekday/daytime	5000001	v000001	3599	1156
...						
r000001	motorway	weekend/daytime	5000120	l000119	1863	1194
...						

G.2. Determining the weight of the strata in the total sample

As mentioned in Section E, the weight formula includes a stratum level component (Sl_{ij}), which can be estimated through two alternatives – see page 8. If Trendline beneficiaries decide for the first alternative, they will complete Table 5, which includes relative proportion of traffic volume (estimates) by considered (mixed) sampling strata: Region ID (optional), Road type, Time period, Vehicle type j of stratum i ($TR_j(i)$).

Given that version 1 is recommended, please consult country experts (e.g., Ministry of Transport) for estimates.

Table 5 Example of relative traffic volume (template sheet 4.Roadside_RawData-3level)

Region ID (optional)	Road type	Time period	Vehicle type	TR _j (i)
r000001	motorway	weekday/ daytime	passenger car	0.441313
r000001	motorway	weekday/ daytime	goods vehicle	0.038375
...				
r000001	urban road	weekend/ daytime	passenger car	0.030938
r000001	urban road	weekend/ daytime	goods vehicle	0.00269
...				

Alternatively, version 2 is based on calculating the stratum level weight component using variables such as Road length ($RI(i)$), relative length of the considered time periods ($TP(i)$), and the average number of vehicles per hour (or any other standard duration) of stratum i ($NS_j(i)$) as detailed in Tables 6-8. If Trendline beneficiaries provide raw data, $NS_j(i)$ estimation is derived from count period data. Completing Tables 6 and 7 in the template is therefore sufficient.

Table 6 Example of road length (template sheet 5.Roadside_RawData-1level)

Region ID (optional)	Road type	RI
r000001	motorway	1326
r000001	rural road	36592
r000001	urban road	58427

Table 7 Example of relative length of considered time period proportion
(template sheet 5.Roadside_RawData-1level)

Time period	TP
weekday/daytime	0,6897
weekend/daytime	0,3103

Table 8 Example of number of vehicles adjusted per hour duration from count period data
(template sheet 5.Roadside_RawData-1level)

Region ID (optional)	Road type	Vehicle type	NSj(i)
r000001	motorway	passenger car	1612
r000001	motorway	goods vehicle	314
r000001	rural road	passenger car	1488
r000001	rural road	goods vehicle	265
r000001	urban road	passenger car	523
r000001	urban road	goods vehicle	34

Regarding the KPI Speed, the Fixed speed limit table (template sheet 5.Roadside_RawData-1level) specifies the speed limits used for different types of vehicles to be used in filling in the Results sheet in templates (6.Roadside_Results-aggr and 7.Roadside_Results-semiaggr). Please note that it is important to carefully complete the Metadata sheet, where is the selected variant for stratum level component. In the case of KPI Seatbelt & CRS, the chosen variant of collected age group attribute might be selected.

H. Using a survey package – example in R

Let us illustrate the whole process in an example with the Czech seat belt data. We were interested in the proportion of seat belt use. Six attributes were considered: road type, time period, vehicle type, age group, gender and seat position. The survey design was defined in R using the function `svydesign` as follows:

```
svydesign( ids = ~Session_ID,  
          strata = ~Road_type + Time_period,  
          weights = ~Wx,  
          data = dt )
```

```
> svydesign(ids = ~Session_ID, strata = ~Road_type + Time_period , weights = ~Wx, data = dt)  
Stratified 1 - level Cluster Sampling design (with replacement)  
with (120) clusters.  
svydesign(ids = ~Session_ID, strata = ~Road_type + Time_period,  
          weights = ~Wx, data = dt)
```

where `Session_ID` contains the IDs of the sessions, road type and time period define the strata because they were chosen prior to the observation, W_x are weights defined by formula (3) assigned to each observation and `dt` is a data.frame of all the observations. This data.frame consists of the attributes, the information on the seat belt use called `Used` (TRUE/FALSE), session ID and weight.

```
> head(dt)  
# A tibble: 6 × 9  
  Session_ID Road_type Time_period Vehicle_type Age_group Gender Seat_position Used Wx  
  <chr>      <fct>      <fct>      <fct>      <fct>  <fct> <fct>      <lgl> <dbl>  
1 s000001   rural road weekday/daytime goods vehicle 15+   male driver      TRUE 0.00441  
2 s000001   rural road weekday/daytime passenger car 15+   female driver    TRUE 0.145  
3 s000001   rural road weekday/daytime goods vehicle 15+   male driver      FALSE 0.00441  
4 s000001   rural road weekday/daytime passenger car 15+   female driver    FALSE 0.145  
5 s000001   rural road weekday/daytime passenger car 15+   male front passenger FALSE 0.145  
6 s000001   rural road weekday/daytime passenger car 15+   male driver      TRUE 0.145
```

Considering all the data, the proportion of seat belt use, along with its standard error and confidence interval, was calculated by function `svyciprop` utilizing the above-defined design and estimation method `beta` (Korn and Graubard, 1998):

```
svyciprop( ~Used, design = design, method = "beta" )
```

```
> svyciprop( ~Used, design = design, method = "beta" )  
2.5% 97.5%  
Used 0.976 0.972 0.980
```

Further, applying functions `SE` and `confint` derived the standard error and the confidence interval from the `svyciprop` object.

We were interested in estimating not only the general proportion of seat belt use on the aggregated level, but also in deriving estimates about the proportion on the semi-aggregated levels. This was done by using function `svyciprop` within the `svyby` context while reporting the variability in terms of the standard error and confidence interval (see parameter `vartype`). For clarity, let us denote that `svyby` does not report results for empty subsets. Our approach is demonstrated in three examples:

1. Semi-aggregated level with subsets defined by all six attributes

```
> svyby( ~Used, by = ~Road_type + Time_period + Vehicle_type + Age_group + Gender + Seat_position,
+       design = design, svyciprop, method = "beta", vartype = c("se","ci") )
```

Road_type	Time_period	Vehicle_type	Age_group	Gender	Seat_position	Used	se.as.numeric(Used)	ci_l	ci_u	
motorway	weekday/daytime	passenger	car	15+	male	driver	0.9829919	0.002639286	0.9765130	0.9880710
rural road	weekday/daytime	passenger	car	15+	male	driver	0.9329023	0.011013868	0.9057920	0.9541549
urban road	weekday/daytime	passenger	car	15+	male	driver	0.9289677	0.008543711	0.9088124	0.9458541
motorway	weekend/daytime	passenger	car	15+	male	driver	0.9871096	0.002858789	0.9796070	0.9923913
rural road	weekend/daytime	passenger	car	15+	male	driver	0.9490854	0.007265129	0.9315363	0.9632337
urban road	weekend/daytime	passenger	car	15+	male	driver	0.9395588	0.007170218	0.9226581	0.9537213
motorway	weekday/daytime	goods	vehicle	15+	male	driver	0.8683386	0.014122738	0.8356773	0.8966022
rural road	weekday/daytime	goods	vehicle	15+	male	driver	0.7872093	0.026618491	0.7252490	0.8407854
urban road	weekday/daytime	goods	vehicle	15+	male	driver	0.6408451	0.029812006	0.5750221	0.7029696
motorway	weekend/daytime	goods	vehicle	15+	male	driver	0.9565217	0.023945400	0.8734614	0.9917454

Estimates, standard errors and confidence intervals were calculated for all possible combinations of all the six attributes. The result had 82 rows.

2. Semi-aggregated level with subsets defined by time period, age group, gender and seat position.

```
> svyby( ~Used, by = ~Time_period + Age_group + Gender + Seat_position,
+       design = design, svyciprop, method = "beta", vartype = c("se","ci") )
```

Time_period	Age_group	Gender	Seat_position	Used	se.as.numeric(Used)	ci_l	ci_u
weekday/daytime	15+	male	driver	0.9786398	0.0031526003	0.9713392	0.9844966
weekend/daytime	15+	male	driver	0.9811566	0.0029360071	0.9743023	0.9865844
weekday/daytime	15+	female	driver	0.9928637	0.0021733561	0.9870124	0.9965346
weekend/daytime	15+	female	driver	0.9957479	0.0011587859	0.9927189	0.9977432
weekday/daytime	15+	male	front passenger	0.9346108	0.0129053919	0.9035067	0.9580953
weekend/daytime	15+	male	front passenger	0.9456488	0.0117308320	0.9168579	0.9667301
weekday/daytime	15+	female	front passenger	0.9914398	0.0038387994	0.9796878	0.9973277
weekend/daytime	15+	female	front passenger	0.9977712	0.0006313523	0.9961029	0.9988508
weekday/daytime	0-14	unknown	front passenger	0.9940438	0.0027593340	0.9854944	0.9982245
weekend/daytime	0-14	unknown	front passenger	0.9977792	0.0011837258	0.9938856	0.9994803
weekday/daytime	15+	male	rear passenger	0.6445574	0.0644403099	0.5014319	0.7709751
weekend/daytime	15+	male	rear passenger	0.7379659	0.0527494804	0.6163297	0.8377457
weekday/daytime	15+	female	rear passenger	0.9427542	0.0303626926	0.8467061	0.9870702
weekend/daytime	15+	female	rear passenger	0.9705556	0.0137666048	0.9283068	0.9914625
weekday/daytime	0-14	unknown	rear passenger	0.9946387	0.0026633674	0.9861441	0.9985764
weekend/daytime	0-14	unknown	rear passenger	0.9816456	0.0063914197	0.9638341	0.9921563
weekday/daytime	0-14	unknown	CRS	0.9851534	0.0139332706	0.9237138	0.9995321
weekend/daytime	0-14	unknown	CRS	0.9967230	0.0025607075	0.9864306	0.9997417

Estimates, standard errors and confidence intervals were calculated for all possible combinations of time period, age group, gender and seat position (i. e. aggregation across road type and vehicle type). The result had 18 rows.

3. Semi-aggregated level with subsets defined by road type and seat position.

```
> svyby( ~Used, by = ~Road_type + Seat_position,
+       design = design, svyciprop, method = "beta", vartype = c("se","ci") )
```

Road_type	Seat_position	Used	se. as.numeric(Used)	ci_l	ci_u
motorway	driver	0.9877223	0.001745476	0.9836520	0.9910007
rural road	driver	0.9473351	0.007036240	0.9311251	0.9606816
urban road	driver	0.9447626	0.004606334	0.9346510	0.9537264
motorway	front passenger	0.9804783	0.004169900	0.9700729	0.9880033
rural road	front passenger	0.9484185	0.006080122	0.9346212	0.9600417
urban road	front passenger	0.9343009	0.006086945	0.9208309	0.9461022
motorway	rear passenger	0.9052754	0.014591317	0.8712646	0.9327791
rural road	rear passenger	0.8901400	0.016244816	0.8525176	0.9209524
urban road	rear passenger	0.8696779	0.013110362	0.8406185	0.8951608
motorway	CRS	0.9911086	0.007456176	0.9596428	0.9995094
rural road	CRS	0.9943600	0.003220324	0.9832750	0.9988876
urban road	CRS	0.9926882	0.003043930	0.9835177	0.9974907

Estimates, standard errors and confidence intervals were calculated for all possible combinations of road type and seat position (i. e. aggregation across time period, vehicle type, age group and gender). The result had 12 rows. If the sample size in the subset does not exceed at least 30 observations (Parker et al., 2017), the result values will be replaced with not available (NA) values.

I. Summary

The document primarily focuses on field measurement within the scope of behavioral KPIs (such as speeding, driving under the influence of alcohol, use of protective equipment, wearing a seatbelt, or distraction), where each observation within one session at a specific location (in a specific time period) is weighted based on the number of vehicles by type, and the observation and counting (calibration) time duration. The resulting data processing involves combinations of stratum attributes specified in the templates.

The weighting formula is designed to respect the specific conditions of each measurement and national indicators describing the traffic characteristics on the given road and chosen period. The weight value varies by session and vehicle type, and these weights are subsequently applied to the original dataset.

The calculation of the expected value of KPI is conducted through a complex survey approach, which includes defining the sample design and computing characteristics based on beta distribution. It is advisable to use the survey package in statistical software (e.g. R) for computing these characteristics.

It is recalled that Trendline beneficiaries have the choice to provide raw data. If raw data is provided, the Trendline coordination team will return the results and ensure consistency, thereby saving the effort of Trendline beneficiaries (they do not need to fill in the template sheets "6.Roadside_Results-aggr" and "7.Roadside_Results-semiaggr" which is optional). If no raw data is provided, Trendline beneficiaries have to use the algorithms described in this document according to the specific KPI guidelines and minimum requirements. When doing so, careful completion of the templates and Metadata sheet is important.

Further reading

- Billingsley, P. (1995). Probability and Measure (3rd ed.). John Wiley & Sons. p. 357-363.
- Cochran, W. G. (1977). Sampling Techniques. Wiley
- Hogg, R., McKean, J., Craig, A. (2019). Introduction to Mathematical Statistics. Eight edition. Boston Pearson
- Korn EL, Graubard BI. (1998) Confidence Intervals For Proportions With Small Expected Number of Positive Counts Estimated From Survey Data. Survey Methodology 23:193-201.
- Parker JD, Talih M, Malec DJ, et al. (2017) National Center for Health Statistics data presentation standards for proportions. National Center for Health Statistics. Vital Health Stat 2(175).
- Skinner, Ch. and Wakefield J. (2017). Introduction to the Design and Analysis of Complex Survey Data. Statistical Science, vol. 32, No.2, 165-175, DOI:10.1214/17-STS614.
- Thompson, S. K. (2012). Sampling. Wiley
- Wu, C., Thompson, M. E. (2020). Sampling Theory and Practice. Springer International Publishing