

KONSTRUKCE VÝBĚROVÉHO SOUBORU PRŮZKUMŮ DOPRAVNÍHO CHOVÁNÍ PRO ÚČELY MĚSTSKÉHO PLÁNOVÁNÍ

Petr Kouřil, Vít Gabrhel, Michal Šimeček, Daniel Szabó, Marek Tögel

Kvalitní plánování městské dopravy se neobejde bez znalosti dopravního chování jeho obyvatel. K jeho zjišťování slouží dotazníková šetření, zjišťující odkud, kam, pomocí jakých dopravních prostředků a za jakým účelem lidé cestují. Klíčovým faktorem kvalitního šetření je konstrukce výběrového souboru, jehož cílem je za pomoci omezeného počtu vybraných jednotek co nejlépe reprezentovat celou populaci. Článek popisuje problematiku konstrukce vzorku pro městská dotazníková šetření a předkládá doporučení s ohledem na nejčastější využití dat o dopravním chování.

Klíčová slova: průzkumy dopravního chování, modelování poptávky po dopravě, plán udržitelné městské mobility, výběrový soubor

Úvod

Průzkumy dopravního chování (dále jen PDCH) představují klíčový zdroj informací o poptávce po dopravě [Šenk a kol, 2001]. Odpovídají na otázky po tom odkud, kam, kdy, jakými dopravními prostředky či za jakými účely lidé cestují. Tyto informace pak nacházejí využití v oblasti dopravního plánování – ať už je to na úrovni národní, regionální či městské. V České republice bylo dosud realizováno několik desítek takových šetření, řádově jednotky z nich pak na základě mezinárodně standardizované metodiky, zajišťující potřebnou kvalitu výstupních dat (např. Pardubice v letech 2013 a 2017, Uherské Hradiště v roce 2013, Jihomoravský kraj 2013, Olomouc 2016, Litoměřice 2017, celostátní průzkum Česko v pohybu v letech 2017–2019).

Získávané informace

PDCH zjišťují informace o čtyřech základních zkoumaných jednotkách. Jedná se o a) domácnosti, b) osoby, které v domácnosti žijí, c) dopravní prostředky, které mají osoby v domácnosti k dispozici a v neposlední řadě o d) cesty podniknuté členy domácnosti ve zkoumaném období.

Informace o podniknutých cestách se většinou sbírají za jeden (tzv. rozhodný) den, byť může být dotazované období i delší. Vzhledem k vysoké zátěži

kladené na respondenty průzkumu se nicméně často volí jednodenní varianta a vybírá se běžný pracovní den, úterý až čtvrtek, tedy s výjimkou dní následujících či předcházejících svátečním dnům. Podobné zjednodušení se často provádí i z hlediska ročního období, kdy se sbírá v pouze v období mezi březnem a květnem a mezi zářím a říjnem. Těchto pět měsíců je považováno za z dopravního hlediska běžná období. Vhodnější či robustnější variantou je kontinuální sběr, kdy dochází ke sběru dat rovnoměrně časově i geograficky rozloženému do období celého roku.

Využití PDCH při plánování udržitelné městské mobility

Z hlediska dopravního chování ve městě nacházejí data z PDCH uplatnění v rámci plánů udržitelné městské mobility. Plány udržitelné mobility jsou strategické dokumenty vycházející z klimatických a energetických cílů Evropské unie. Tyto plány se v evropských zemích postupně stávají povinnými pro větší města, která mají v úmyslu čerpat finanční prostředky ze strukturálních a investičních fondů. Například v České republice je pro období 2014–2020 Sustainable Urban Mobility Plan (SUMP; v češtině se používá termín Plán udržitelné městské mobility), nebo jeho zjednodušená verze Sustainable Urban Mobility Framework (SUMF), podmínkou pro čerpání z Operačního programu Doprava

(OPD) a Integrovaného regionálního operačního programu (IROP) pro města nad 50 000 obyvatel.

Hlavním cílem SUMP je zajistit naplňování dopravních potřeb obyvatel a podniků při zlepšování kvality života, životního prostředí a městské ekonomiky. SUMP tak usiluje o proměnu měst, městské a příměstské dopravy a městského plánování v různých oblastech: například v oblasti dopravy jde o snížení kongescí, podporu veřejné, nemotorové a nízko- a bezemisní dopravy, podporu inteligentních dopravních systémů a technologických inovací, o redukci nároků a dopadů dopravních systémů na životní prostředí, zlepšení dostupnosti a sociální rovnosti a zlepšení bezpečnosti a spolehlivosti městské dopravy.

Kromě základních zjištění o mobilityním chování obyvatelstva data o dopravním chování nacházejí uplatnění v rámci SUMP při tvorbě dopravních modelů. Dopravní modely je možné konstruovat i pro menší města, a to s podrobností na úrovni základních územních jednotek nebo i sčítacích obvodů. Přestože dopravní modely nejsou povinnou součástí SUMP, pokud jsou vhodně realizovány a udržovány, jejich přínos v procesu tvorby a implementace SUMP je nesporný. Dopravní modely totiž umožňují provádět spolehlivé dopravní analýzy pro současnost i pro dlouhodobé horizonty [JASPERS, 2014]. Ukazují dopady realizace měkkých i tvrdých doprav-

ních opatření, dopravních staveb, změn ve využití území a změn dopravního chování a jejich synergií v různých scénářích vývoje. Poskytují tím zásadní podporu při rozhodování o dopravním a územním rozvoji s ohledem na vyšší strategické cíle. Zároveň umožňují stanovení měřitelných indikátorů, hodnot těchto indikátorů v současném stavu, ale i hodnot predikovaných v rámci stanoveného scénáře. Poskytují tak rámec pro monitoring a zpětnou vazbu mezi plánováním a realizací. Dopravní modely mají v procesu rozhodování a plánování přínos zejména pro:

- poznání skladby cestujících, intenzit dopravy, druhů cest a jejich cílů;
- identifikaci nedostatků a úzkých hrdel infrastruktury a porozumění poptávce pro posouzení potřebnosti rozšíření (redukce) služeb a kapacity infrastruktury;

- predikci dopadu rozvoje dopravní infrastruktury na multimodální dopravní toky a poptávku;
- analýzu vztahu mezi dopravními systémy a sociodemografickým vývojem, ekonomickým rozvojem, vlastnictvím vozidel, rozvojem a změnami využití území atp.

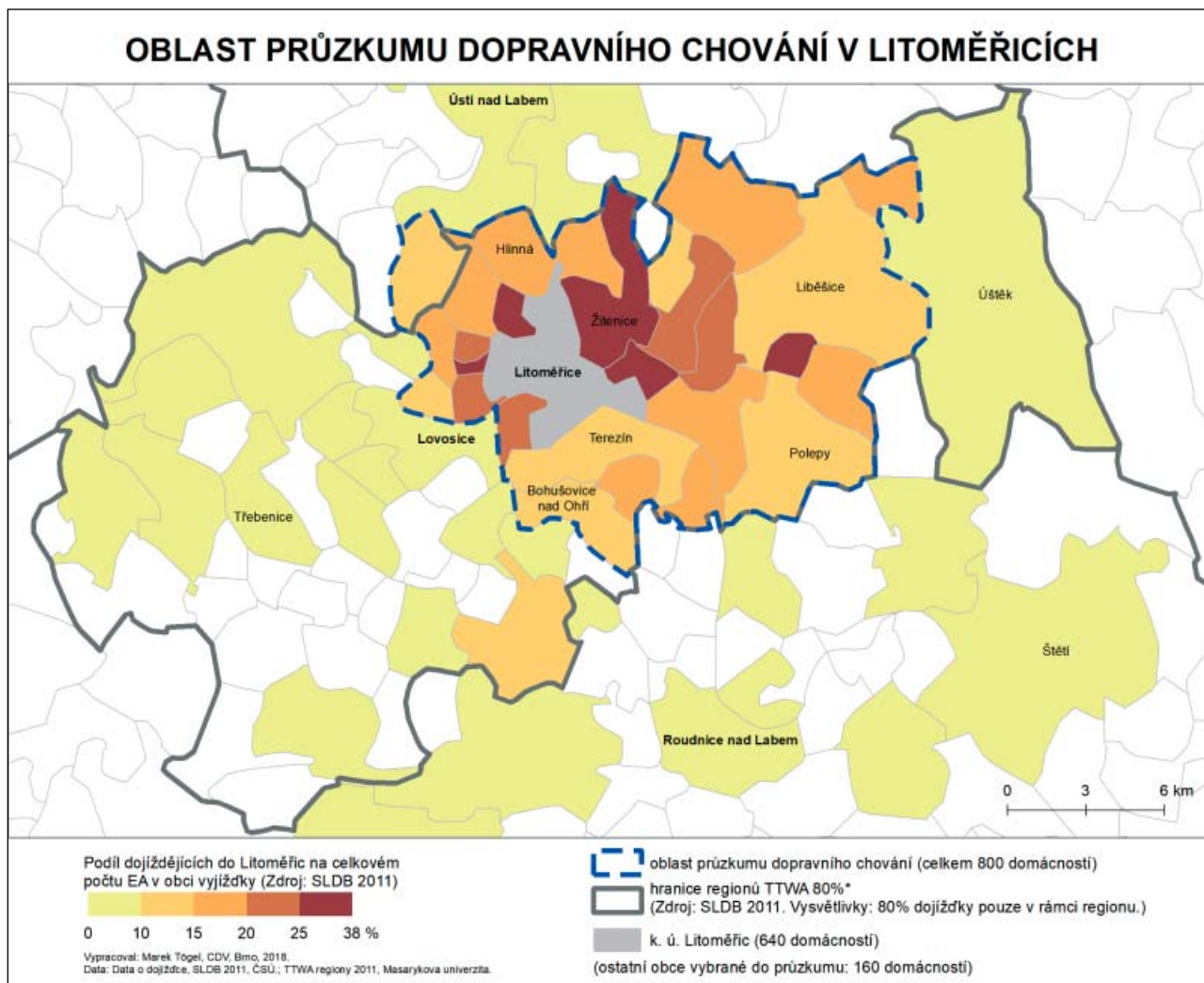
Výběr oblasti pro realizaci PDCH

Jedním z prvních kroků při konstrukci výběrového souboru pro PDCH je definice geografické oblasti, v níž se budou data sbírat. Obyvatelé této oblasti představují tzv. populaci, tj. soubor jednotek, o kterých průzkum vypovídá.

Geografickou oblast pro sběr dat lze vymezit jako tzv. funkční městský region nebo tzv. travel-to-work area (TTWA)

regionů. Použití jedné nebo druhé metody závisí na charakteru sídelní struktury daného území. Vymezit funkční městský region je vhodné v případě, že struktura osídlení je monocentrická (např. Praha, Brno, České Budějovice, Plzeň). Jádrem v takovém případě tvoří jediné velké město, kam směřuje většina pracovní dojížděky, nebo ze kterého vychází převážná část pracovní vyjížděky. Zahnutí obce do funkčního regionu je zpravidla stanoveno kritériem podílu počtu vyjíždějících do centra regionu na celkovém počtu ekonomicky aktivních v dané obci.

K vymezování TTWA regionů se přistupuje v případě, že struktura osídlení v daném území je polycentrická (např. Ústecko, Liberecko, Pardubice a Hradec Králové, Ostravsko, Olomoucko, Zlínsko). Pro tato území jsou charakteristické stejně silné vztahy



Stanovení oblasti průzkumu dopravního chování pro účely plánu udržitelné mobility města Litoměřice

dojížděky směřující do různých center polycentrického regionu, a proto je třeba zahrnout do modelované oblasti všechna tato sídla. V případě nezahrnutí některého z těchto sídel dochází k situaci, kdy v průzkumu a následně i v dopravním modelu není zastoupena populace, která však ve skutečnosti významně ovlivňuje dopravní situaci ve zkoumaném území.

Výběrový soubor

Ať už je geografická oblast definována jakkoliv, z ekonomických a praktických důvodů není při zjišťování dopravního chování možné dotazovat celou zkoumanou populaci, například všechny obyvatele monocentrického osídlení typu Praha či Brno. Průzkum tedy probíhá na tzv. výběrovém souboru, tedy výběru jednotek, který zastupuje sledovanou populaci.

Jedním z klíčových předpokladů platnosti závěrů vyvozených z průzkumů dopravního chování je jejich reprezentativita. Tedy jinými slovy to, že osoby tvořící výběrový soubor dobře reprezentují populaci obyvatel žijících na vymezeném území. Navzdory existenci jasných standardů v oblasti metodologie sběru dat [Armoogum, 2014] se nejen v kontextu tvorby plánů udržitelné městské mobility setkáváme s tím, že data o dopravním chování nejsou získávána způsobem, který by právě jejich reprezentativitu zajišťoval. Dopady této skutečnosti přitom mohou vést ke zkresleným závěrům, a ty pak k nežádoucím důsledkům, například k rozhodnutím o výstavbě nevhodné či nepotřebné a nákladné dopravní infrastruktury.

Existují přitom v zásadě dvě cesty, kterými se lze při konstrukci výběrového souboru vydat, totiž pravděpodobnostní a nepravděpodobnostní. Pravděpodobnostní způsoby výběru (např. systematický, stratifikovaný či více-
stupňový) vznikají tak, že z populace jeví, například domácností obývajících dané území, vybereme požadovaný počet jednotek výběru, tedy domácností tak, že každá výběrová jednotka má stejnou či obecněji předem známou míru pravděpodobnosti zahrnutou

do výběru. Proto se tomuto způsobu sběru říká pravděpodobnostní. Oproti tomu nepravděpodobnostní způsob výběru tuto charakteristiku postrádá – výběrové jednotky jsou zahrnuty na základě kritérií, jakými jsou jejich dostupnost (kupř. anketa na internetu) či znalost populačních parametrů u proměnných, které jsou považovány za klíčové (tzv. kvótní výběr). Zatímco pravděpodobnostní způsob výběru umožňuje usuzovat z výběrového souboru na populaci, tedy provádět statistickou inferenci, u nepravděpodobnostních způsobů toto možné není.

Pro konstrukci pravděpodobnostního výběrového souboru je nutné disponovat seznamem jednotek populace, z nichž budeme vybírat. Statistickými jednotkami, které pro účely PDCH představují populaci, jsou buď jednotlivci, nebo domácnosti. Doporučujeme zvolit domácnosti, protože jedině pro ně lze v našich podmínkách získat použitelnou oporu. Není sice k dispozici kontaktní seznam všech českých a moravských domácností, lze však vytvořit anonymní seznam domácností vycházející z údajů o adresních bodech, sesbíraných v rámci posledního cenzu [SLDB 2011]. Jelikož víme, kolik domácností žije na každém adresním bodě, jsme schopni každé domácnosti v populaci zajistit, aby měla stejnou či předem známou pravděpodobnost, že bude zařazena do výběrového souboru.

To, že je opora postavena na adresních bodech, determinuje celou řadu dalších aspektů průzkumu. Pravděpodobnostní „adresní sběr“ totiž patří mezi nejkomplicovanější typy průzkumů, které agentury na českém trhu provádějí. Probíhá tak, že tazatel obdrží seznam adres (tazatelský úkol, většinou okolo 10 adres) a předpis, na jehož základě má oslovit požadovaný počet domácností. Tyto domácnosti musí osobně kontaktovat a tento kontakt v případě neúspěchu i několikrát opakovat, než se vydá za náhradní domácnosti. Je tedy zřejmé, že tento typ průzkumu je extrémně náročný na organizaci a řízení tazatelů, na jejich zkušenost a vytrvalost, ale i na cestovní, personální a především časové náklady.

Všechny tyto aspekty je nutno brát v úvahu při konstrukci výběrového souboru. Nákladnost je v každém případě vyvážena získáním takových dat o výběrovém souboru, která nám umožní činit závěry o celé sledované populaci.

Velikost výběrového souboru

Zásadním krokem při konstrukci výběrového souboru je stanovení jeho velikosti. Jak velký by měl výběrový soubor být, aby dobře reprezentoval zkoumanou populaci? Na tomto místě je potřeba se opět vrátit k účelu průzkumu, k otázkám, na které by nám měl odpovědět. Obecně lze doporučit v první řadě identifikovat nejméně četný jev, který nás zajímá. Zjednodušeně můžeme říci, že čím méně četný jev nás zajímá, tím větší výběrový soubor budeme potřebovat.

Abychom mohli stanovit, jak velký výběrový soubor bychom pro tento nejméně četný jev potřebovali, musíme zvážit, s jakou výběrovou chybou jsme ochotni se smířit. Pro maximální přípustnou velikost chyby neexistuje žádná norma nebo doporučení, opět je nutné vycházet z konkrétních potřeb a důrazu kladeného na kvalitu dat, ať už dopravním modelem nebo potřebami podkladových analýz. Při stanovení přípustné výběrové chyby pro nejméně četný jev je vhodné absolutní výběrovou chybu (E_a , tab. 1, vz. 1) vztáhnout k velikosti tohoto jevu (tab. 1, vz. 2). Získáme tak relativní výběrovou chybu (E_r), která nám při nízké četnosti jevu dá lepší představu o rozměrech chyby.

Při výpočtu velikosti vzorku (tab. 1, vz. 3) je nutné uvědomit si, s četností jakých statistických jednotek zacházíme. Výběrový soubor sice konstruujeme z domácností, ale nejméně četný jev, který nás zajímá, se může vypočítávat z jednotek osob (např. zastoupení některé věkové skupiny), nebo z jednotek cest (například zastoupení některého dopravního modu). Počet osob můžeme zhruba odhadnout, když počet domácností vynásobíme dvěma (obvyklý počet osob z domácnosti, které odpoví na dotazník). Počet cest pak můžeme

(1) $E_a = z \cdot \frac{P_j \cdot (1 - P_j)}{n}$	E_a = absolutní výběrová chyba E_r = relativní výběrová chyba n = velikost výběrového souboru
(2) $E_r = \frac{E_a}{P_j}$	P_j = pravděpodobnost nejméně četného jevu, který nás zajímá z = velikost hranice intervalu spolehlivosti, zde doporučujeme dosadit hodnotu jedné směrodatné odchylky, která odpovídá hladině spolehlivosti 95 %, tedy 1,96
(3) $n = \frac{P_j \cdot (1 - P_j)}{E_a^2} \cdot z^2$	

Tab. 1: Výpočet velikosti výběrového souboru

odhadnout, když počet osob dále vynásobíme 2,5 (obvyklý průměr cest vykonaných za jeden sledovaný den).

Tabulka 2 ukazuje vztah velikosti relativní výběrové chyby k velikosti výběrového souboru a pravděpodobnosti výskytu zkoumaného jevu, v tomto případě pro jednotky cest. S nárůstem velikosti výběrového souboru a růstem pravděpodobnosti výskytu chyba plynule klesá.

Náklady a rizikové faktory

Z praktického hlediska je velikost výběrového souboru determinována třemi faktory. Tím prvním jsou finanční prostředky, které jsou na realizaci průzkumu k dispozici. Pokud nezahrnujeme náklady na metodologickou přípravu, kontrolu průběhu průzkumu a analýzu dat, pak náklady na jednu úspěšně dotázanou domácnost dosahují přibližně 1 000 až 1 200 Kč vč.

DPH. Tyto ceny se mohou poměrně rychle měnit, zejména s ohledem na vývoj pracovního trhu, neboť podstatnou část této ceny, zhruba tři čtvrtiny, tvoří náklady na odměny tazatelům, provádějícím dotazování.

Druhým faktorem je délka sběru. Průzkum dopravního chování prováděný dotazováním přímo v domácnostech je časově velmi náročná aktivita, závislá na poctivé a neúnavné práci tazatelů. V průběhu jednoho měsíce jsou tazatelé schopni úspěšně dotázat 10 až 20 domácností. Tazatelé se v terénu potýkají s celou řadou problémů, jako jsou zamčené vchody do domů, s neochotou či dokonce agresivitou respondentů, či s nemožností domácnost vůbec zastihnout. Na jednu úspěšně dotázanou domácnost tak často připadá několik domácností neúspěšně dotázaných. Na základě dosud provedených průzkumů, např. Olomouc v roce 2016, Litoměřice v roce 2017, nyní běžící celostátní průzkum Česko

v pohybu [Gabrhel a kol, 2016], odhadujeme maximální rychlost sběru na 650 až 750 domácností za měsíc.

Třetí faktor se týká zejména měst s menší velikostí. Je jím poměr velikosti výběrového souboru ku velikosti zkoumané populace, kdy může vzniknout problém přehlacení populace průzkumem. Jelikož na jednu úspěšně dotázanou domácnost připadá několik nedotázaných (třeba i 3), v menších městech bude šance domácnosti na oslovení v rámci průzkumu poměrně vysoká. Přehlacenost může vyústit v negativní přijetí průzkumu obyvatelstvem, v další prudký pokles ochoty k účasti na něm. To se pak pochopitelně může promítnout do kvality dat.

Závěr

PDCH přináší cenné informace pro dopravní plánování, ať už ve smyslu plánování fyzické infrastruktury, tak tzv. měkkých opatření. S daty z PDCH se pak explicitně pracuje například v plánech udržitelné městské mobility (SUMP). Zde se tato data používají jednak k analýze dopravního chování obyvatel, například ve smyslu vybavenosti domácnosti dopravními prostředky, tak k přípravě datových vstupů do dopravního modelu. Ke zdárné a informačně hodnotné realizaci PDCH je přitom třeba dodržet několik kroků: a) určit z hlediska dopravního chování relevantní území, na kterém bude sběr probíhat; b) zvolit reprezentativní

Velikost výběrového souboru (n)			Pravděpodobnost jevu v populaci cest (Pj)				
Domácností	Osob (odhad)	Cest (odhad)	10 %	20 %	30 %	40 %	50 %
250	500	1 250	±16,6 %	±11,1 %	±8,5 %	±6,8 %	±5,5 %
500	1 000	2 500	±11,8 %	±7,8 %	±6,0 %	±4,8 %	±3,9 %
750	1 500	3 750	±9,6 %	±6,4 %	±4,9 %	±3,9 %	±3,2 %
1 000	2 000	5 000	±8,3 %	±5,5 %	±4,2 %	±3,4 %	±2,8 %
1 250	2 500	6 250	±7,4 %	±5,0 %	±3,8 %	±3,0 %	±2,5 %
1 500	3 000	7 500	±6,8 %	±4,5 %	±3,5 %	±2,8 %	±2,3 %

Tab. 2: Relativní výběrová chyba (Er) pro vybrané velikosti výběrového souboru (n) a četnosti jevu (pro případ cest)

způsob konstrukce výběrového souboru a c) stanovit požadovanou velikost tohoto výběrového souboru. Přípravě sběru dat i jeho samotné realizaci pak musí předcházet rozvaha nad možnými limity, zejména nad těmi časovými a finančními. Byť se jedná o náročnou proceduru, při správném postupu povede k získání hodnotných informací, o které se lze opřít při tvorbě dopravní politiky a její realizaci.

Tento článek byl vytvořen za finanční podpory Ministerstva školství, mládeže a tělovýchovy v rámci programu Národní program udržitelnosti I, projektu Dopravní VaV centrum (LO1610) na výzkumné infrastruktuře pořízené z Operačního programu Výzkum a vývoj pro inovace (CZ.1.05/2.1.00/03.0064).

Použité zdroje:

Armoogum, J. (2014). *Survey Harmonisation with New Technologies Improvement*. Les collections de L'INRETS.

Gabrhel, V., Kouřil, P., Melzer, Z. (2016). „Česká republika v pohybu: Návrh celostátního průzkumu dopravního chování“. *Dopravní inženýrství*, roč. 16, č. 2, s. 20–23. ISSN 1801-8890.

JASPERS (2014). *JASPERS Appraisal Guidance (Transport): The Use of Transport Models in Transport Planning and Project Appraisal*.

Šenk, P., Kouřil, P. (2014). „Průzkumy dopravního chování v ČR a zahraničí. Potřebujeme národní průzkum?“ *Dopravní inženýrství*, č. 1, s. 24–26. ISSN 1801-8890.

Mgr. Petr Kouřil, Mgr. Vít Gabrhel

Mgr. Michal Šimeček, Ph.D.

*Mgr. Daniel Szabó, Mgr. Marek Tögel
Centrum dopravního výzkumu, v.v.i.*

ENGLISH ABSTRACT

Construction of a sample set for urban planning research in transport behaviour, by Petr Kouřil

High-quality planning of urban transportation is not feasible without knowledge of transport behaviour of inhabitants. Transport behaviour is identified by questionnaire surveys which detect from where, to where, by which means of transportation, and for what purpose people travel. A key factor in a reliable investigation is the construction of a sample set representing the whole population based on a limited number of units. This article describes how a sample set for urban questionnaire surveys is constructed and gives recommendations for how data on transport behaviour can commonly be used.