



## ИССЛЕДОВАНИЕ СПОСОБОВ ВЫБОРОЧНЫХ ОБСЛЕДОВАНИЙ ТРАНСПОРТНЫХ ПОТРЕБНОСТЕЙ НАСЕЛЕНИЯ И ИХ МИНИМАЛЬНО ДОПУСТИМЫХ ВЕЛИЧИН

Статья посвящена исследованию способов определения минимально допустимого выборочного объема статистического опроса населения в целях установления общих закономерностей его распределения в транспортных системах городов.

**Ключевые слова:** транспортное планирование, выборочная совокупность, население, корреспонденция, статистика, социологическое исследование, город.

*This article is about the study of ways to determine the minimum sample volume of data in planning the statistical survey of population in urban transport systems.*

**Keywords:** transport planning, sampling frame, population, correspondence, statistics, case study, city.

**Е. И. Енина,**

научный сотрудник  
ООО «Национальный центр  
транспортных исследований»  
(ООО «НЦТИ»)

**Д. В. Енин,**

директор ООО «Институт  
прикладных транспортных  
исследований» (ООО «ИПТИС»),  
доцент кафедры логистики МАДИ,  
кандидат техн. наук

При разработке документов транспортного планирования возникает потребность сбора исходных данных, к числу которых относятся потребности населения в передвижениях: пешеходных, с использованием различных видов пассажирских транспортных средств, смешанных, нереализованных. Среди статистиков и транспортных специалистов используются различные методы и способы исследований, направленные на решение подобных задач. При этом статистические методы опроса остаются важнейшим инструментом для определения корреспонденций населения, а тенденция сокращения физических, временных и финансовых затрат на сбор исходных данных не потеряет актуальности еще долгое время. Проблема заключается в достоверности исследований, основные принципы которых нарушаются еще на стадии их планирования.

Какой же объем выборки необходим, чтобы количество опрошенных респондентов было достаточным и достоверным для целей анализа и планирования, например, маршрутной сети? Из уст специалистов нередко



можно услышать разные ответы: 160, 500, 1600, 2600 и даже 42000 чел. В практической деятельности преобладающее большинство исследований не превышает 1500 респондентов. Наиболее запоминающимся случаем в жизни был доклад одного из специалистов НИПИ Генплана г. Москвы, в котором докладчик с особым энтузиазмом отмечал, что в 10-миллионном городе было опрошено 500 чел. для целей определения транспортных корреспонденций населения. Давайте попробуем разобраться в справедливости мнений различных специалистов.

Следуя закону больших чисел математической статистики и типовым методам расчета минимальной выборочной совокупности, многие транспортные специалисты пользуются общеизвестной формулой [1]:

$$n_o = \frac{t^2 \cdot p(1-p)}{\Delta^2},$$

где  $t$  – стандартное отклонение (определяется по справочным статистическим таблицам, исходя из величины риска отклонения эмпирической характеристики от теоретической);

$p$  – вероятность возникновения события;

$p(1-p)$  – степень вариации распределения;

$\Delta$  – предельно допустимая ошибка выборки.

Для целей дальнейшего изучения минимальной выборочной совокупности рассмотрим поочередно каждую составляющую этой формулы.

Вопреки сложившемуся мнению многих отраслевых ученых о том, что стандартное отклонение для целей исследований транспортных потребностей населения следует принимать с учетом технических систем (т.е. не более 5%), на самом деле пассажиры, водители, пешеходы, население в целом относятся к социальной системе. Для социальных систем более правильно придерживаться правила Парето, согласно которому нормальным считается отклонение 20%. В дополнение, существует

и невозможность ее применения для многомерных городских систем.

С учетом изложенного для однородной группы населения минимальный объем выборочной совокупности может составить:

$$n_o = \frac{1,64^2 \cdot 0,5(1-0,5)}{0,1^2} \approx 68 \text{ чел.}$$

Однако города неоднородны с точки зрения транспортных потребностей населения и их распределения в пространстве и во времени. Поэтому в социальных системах редко применим прямой математический подход. Более значимую роль в обосновании минимальной выборочной совокупности играет логически правильный алгоритм стратификации объекта исследования.

Стратификация населения города по однородным признакам является обязательным условием получения достоверного результата. Недопустимо оценивать все население без такого деления! Это является серьезным нарушением логических правил, а выборочную совокупность нельзя будет признать достоверной. Нужно понять, какие группы населения обладают общими характеристиками с точки зрения постоянства их транспортных потребностей и с какими целями они совершают или планируют свою поездку.

Отечественные и зарубежные транспортные специалисты выделяют следующие основные признаки:

- 1) гендерный;
- 2) социально-демографический (возрастной или по принадлежности к учебе, труду);
- 3) цель поездки.

Первый признак  $n_g$  в классической постановке имеет два значения: мужчины и женщины. В транспортных исследованиях этим признаком часто пренебрегают. Для некоторых задач такое отступление не является критичным, для других – серьезная ошибка. Нам известны примеры, приведенные несколько лет назад великолепным отраслевым ученым, к.т.н. А. И. Стрельниковым. Ему удалось наглядно показать, как недоучет гендерного признака оказывает влияние на некорректность определения транспортных корреспонденций в городах. Например, в моногородах это может быть связано с различиями трудовой миграции мужчин и женщин к разным местам приложения труда, в других городах – с целями поездки (например, женщины чаще сопровождают ребенка в детский сад или школу, чем мужчины) или возрастной структурой населения. В любом случае отказ от этого признака должен быть обоснован.

Второй признак  $n_{с.д.}$  может рассматриваться в различных формах и с различной степенью стратификации. Однако минимально необходимый состав состоит из трех показателей:

- учащиеся средних специальных и высших учебных заведений – это способные самостоятельно передвигаться по территории города дети старших возрастов и люди трудоспособного возраста, получающие образование;
- трудящиеся;
- несамодеятельное население – пожилые люди, инвалиды, безработные люди трудоспособного возраста, дети и другие социальные группы.



высокая вариативность принятия решения людьми в открытых системах, и многолетние наблюдения авторов позволяют с осторожностью говорить о реальной величине  $20 \pm 10\%$ . Значение  $\pm 10\%$  следует рассматривать как допустимую ошибку  $\Delta$ .

В классических методах математической статистики вероятность события есть величина бинарная, однозначно характеризующая систему в виде ответа «да» или «нет». Максимальная вариация распределения признака в таких системах достигается при значении вероятности 50%. Применительно к транспортной системе – это вероятность потребности респондента в поездке. Но для целей транспортного планирования важен вопрос, куда именно поедет этот человек. На этом остановимся, приняв за аксиому бинарную природу системы оценки

Третий признак  $n_d$  также может быть многообразен, но в числе минимально необходимого перечня показателей также можно выделить три основных:

- учебные цели;
- трудовые цели;
- культурно-бытовые цели (условно – все прочие).

Транспортные специалисты обычно хорошо понимают важность второго и третьего признаков при распределении населения по времени суток, дням недели, периодам года. Однако преобладающее большинство из них упускают еще один существенный дополнительный признак – территориальный, поскольку люди пользуются услугой на ближайших к ним остановочных пунктах.

Выделяя четвертый признак  $n_T$ , по нашему мнению, важно обеспечить его привязку не к административным границам районов города или планировочным красным линиям микрорайонов, а к расчетным транспортным районам, располагаемым, как правило, вокруг остановочного пункта и имеющим границы в радиусе нормативной пешеходной доступности. Общее число расчетных транспортных районов для каждого города различно. Оно зависит от площади его селитебной территории и плотности застройки. В крупных городах, например, оно может составлять от 120 до 250 районов.

В случае зависимой выборочной совокупности минимальный объем выборки определяется произведением значений:

$$n = n_o \cdot n_T \cdot n_{с.д} \cdot n_d \cdot n_T (n_T - 1) = 68 \cdot 2 \cdot 3 \cdot 3 \cdot 200 \cdot 199 = 1224 \cdot 39800 = 48715200 \text{ чел.,}$$

где  $n_o$  – математически обоснованный минимальный объем выборочной совокупности для однородной группы;

$n_T$  – число показателей гендерного признака;

$n_{с.д}$  – число показателей социально-демографического признака;

$n_d$  – число показателей признака цели поездки;

$n_T$  – число расчетных транспортных районов отправления;

$(n_T - 1)$  – число расчетных транспортных районов прибытия.

Выборка оказалась чрезмерно завышенной для города, состоящего из 200 транспортных районов, и поэтому нереализуемой. Рассмотрим допустимые пути ее снижения на основе комбинированных методов обоснования выборочной совокупности.

1. Значение показателя  $n_o = 68$  принято минимально допустимым. Его целесообразно принимать менее указанного. Однако теоретически существует возможность его снизить за счет значений вероятности (при наличии статистических данных о среднеквадратическом отклонении транспортного спроса). Также существуют методы оценки и восстановления данных, работающие с малыми числами в количестве не менее 20. Это значение установим в качестве критического.
2. Для целей настоящей статьи пренебрежем показателем  $n_T$ . При этом его можно учесть через пропорцию на основе имеющейся в городе ежегодной статистической информации.
3. Снижение числа показателей  $n_{с.д}$  и  $n_d$  невозможно, поскольку выбран их минимально необходимый

перечень. Стратификацию населения по социально-демографическому признаку удобнее всего выполнить на основании существующих статистических данных по городу в целом и по его отдельным районам. В случае отсутствия таких данных можно использовать метод независимой выборки, когда учащиеся могут воспользоваться всеми тремя целями, трудящиеся – только двумя (кроме учебных), нетрудоспособное население – одной (культурно-бытовой). Таким образом, произведение составляющих выборочной совокупности сократится с  $3 \cdot 3 = 9$  до  $1 \cdot 3 + 1 \cdot 2 + 1 \cdot 1 = 6$ .

4. Число расчетных транспортных районов отправления  $n_T$  желательно ограничить исключительно селитебными зонами. Дополнительно возможно объединение в один транспортный район не более трех смежных остановочных пунктов, не имеющих разветвления или альтернативных транспортных связей в пределах километровой пешеходной доступности. Число транспортных районов прибытия можно сократить путем предварительных исследований по распределению мест приложения труда в городе на основе информации статистических служб города. В качестве иного варианта возможно решение обратной настроечной задачи путем проведения опросов в учебных заведениях, местах приложения труда, а также на объектах торговли, в учреждениях социальной сферы и культуры. Для примера расчета сформируем критический минимум только применительно к районам отправления, численностью вдвое меньше заданной – 100 ед. Районами прибытия попробуем пренебречь, полагая, что их количество будет определяться на основе мнения респондентов.

Численность населения в расчетных транспортных районах можно получить на основе статистических данных города, паспортных столов, пенсионного фонда и пр. Также возможен укрупненный подсчет числа жилых домов в каждом транспортном районе и числа квартир в каждом из них, умноженных на среднее число людей, проживающих в каждой квартире (колеблется от 2 до 3).

Следуя вышеприведенным примерам, минимально допустимая выборка для каждого транспортного района составит:  $20 \cdot 6 = 120$  чел. (при минимуме 68 чел. –  $68 \cdot 6 = 408$  чел.), а для всей территории крупного города:  $20 \cdot 6 \cdot 100 = 12000$  чел. (при минимуме 68 чел. –  $68 \cdot 6 \cdot 100 = 40800$  чел.). То есть после всех сокращений до критически допустимых значений математически и логически обоснованный минимальный объем выборки оказался близким к наиболее высоким экспертным оценкам.

Несмотря на полученный результат, некоторым практикующим социологам, к числу которых можно отнести, например, известного отечественного ученого Ф.Э. Шереги, удается доказать, что математика не всемогуща в социальных системах. Как было отмечено выше, такие системы являются открытыми, а потребности каждого человека в поездках и их корреспонденции в течение времени формируются в условиях огромного количества

Таблица

Параметры выборочных совокупностей (пример)

Выборка	1		2		3		4		5		6		7		8		9		10	
	чел.	%	число	%	число	%	число	%	число	%	число	%	число	%	число	%	число	%	число	%
10	41	2,4	12	29,3	6	14,6	2	4,9	7	17,1	0	0,0	4	9,8	8	19,5	0	0,0	1	2,4
20	82	3,7	24	29,3	18	22,0	3	3,7	12	14,6	0	0,0	11	13,4	9	11,0	1	1,2	1	1,2
30	123	4,1	39	31,7	21	17,1	6	4,9	14	11,4	7	5,7	16	13,0	9	7,3	4	3,3	2	1,6
40	164	3,7	52	31,7	27	16,5	8	4,9	18	11,0	11	6,7	22	13,4	13	7,9	4	2,4	3	1,8
50	204	2,9	64	31,4	33	16,2	9	4,4	24	11,8	13	6,4	27	13,2	16	7,8	7	3,4	5	2,5
60	245	2,8	77	31,4	39	16,0	11	4,6	28	11,6	15	6,3	33	13,1	20	8,1	9	3,5	6	2,6
70	286	2,9	89	31,0	46	16,2	14	4,8	33	11,7	19	6,5	38	13,3	23	7,7	9	3,3	7	2,6
80	327	3,0	100	30,9	53	16,2	16	4,8	38	11,7	21	6,4	44	13,4	25	7,6	11	3,4	9	2,6
90	368	3,0	116	31,4	59	15,9	18	4,8	42	11,5	23	6,3	49	13,4	28	7,7	12	3,3	10	2,7
100	408	2,9	128	31,4	64	15,8	19	4,7	47	11,5	27	6,5	55	13,5	31	7,7	13	3,3	11	2,7

факторов, учесть которые математически не представляется возможным. В этой связи наиболее достоверными являются выборки, полученные на основе пилотажных социологических исследований.

Для таких исследований необходимо иметь статистические данные о социально-демографической и градостроительной структуре города. Сформировав расчетные транспортные районы, на основе этих данных следует выделить однотипные районы, обладающие близкими характеристиками, и сгруппировать их по этим признакам. В каждой из групп выбирают один-два района для проведения в них пилотажных социологических исследований.

Объем выборки для исследований желательно устанавливать приемлемым, т.е. без снижения значений до «критической массы», поскольку повышенные затраты на пилотажные исследования помогут в дальнейшем сэкономить значительные ресурсы в ходе основного этапа исследования.

В качестве примера реализации такого пилотажного метода можно привести следующий. Предположим, из десяти транспортных районов для исследования выбран один. Нужно определить минимально необходимый объем выборки применительно к остальным однотипным ему районам. Результаты представлены в таблице.

Приемлемые значения показателей таблицы наблюдаются при объеме выборки 50% и выше. Таким образом, для дальнейших исследований можно ограничиться числом респондентов 200 чел. на один транспортный район.

В завершение хочется отметить особенности сбора аналогичных исходных данных при помощи операторов мобильных сетей.

Данное направление является перспективным, оно могло бы значительно упростить задачи сбора информации для целей транспортного планирования. Однако в настоящее время этому способу присущи несколько недостатков:

- 1) для достоверности данных необходим доступ ко всем основным операторам связи региона;
- 2) при получении данных не должно нарушаться право граждан на защиту информации о них;
- 3) треки о путях передвижения человека поступают по координатам с ближайших радиомачт, при этом возможные отклонения наблюдаются в пределах до 200 м, что не позволяет определить объект назначения человека;
- 4) многие жители страны имеют несколько работающих сим-карт, некоторые из них зарегистрированы на одних лиц, но используются другими;
- 5) наиболее существенное замечание – предоставляемые оператором данные не раскрывают гендерных, социально-демографических признаков респондентов, а также целей их передвижения, т.е. не обеспечивают репрезентативность выборки.

Таким образом, социологические исследования остаются на передовой транспортной науки. Они являются хорошим источником данных для корректного восстановления матрицы корреспонденций по данным пассажирских потоков и для определения нереализованного транспортного спроса населения. Аналогичные методы исследования в настоящее время являются актуальными при исследованиях транспортных потребностей маломобильных групп населения и могут быть приняты за основу при более глубокой проработке состава их исследуемых признаков.

**ЛИТЕРАТУРА**

1. Колемаев А. В. Теория вероятностей и математическая статистика / А. В. Колемаев, В. Н. Калинина. – М.: Инфра-М, 2001. – 302 с.