

Лекция № 8

Выборочный метод в статистике

Описанные в предыдущем разделе методы сбора и обработки статистических данных подразумевали, что собрана информация обо всех единицах статистической совокупности. Однако в подавляющем большинстве случаев размер всей совокупности очень велик, и для ее исследования организуется выборочное наблюдение, при котором обследуются не все единицы, а только их часть. При этом вся совокупность единиц, из которой осуществляется отбор, называется *генеральной совокупностью*, а единицы, отобранные для непосредственного наблюдения, - *выборочной совокупностью*, или просто *выборкой*. Отбор из генеральной совокупности производится таким образом, чтобы на основании выборки можно было получить достаточно точное представление об основных параметрах совокупности в целом. Главное требование, которому должна отвечать выборочная совокупность - это требование репрезентативности (представительности).

Выборочный метод наблюдения широко используется на практике и имеет ряд достоинств:

1. Обеспечивает значительную экономию материальных и финансовых ресурсов при проведении наблюдения, что позволяет расширить программу обследования и повысить ее оперативность;
2. Обеспечивает более высокую достоверность получаемой информации, так как при относительно небольшом объеме выборки можно организовать эффективный контроль качества и снизить вероятность появления и необнаружения ошибок регистрации;
3. В ряде случаев, когда наблюдение связано с уничтожением или порчей обследуемых единиц (например, проверка качества продуктов питания), возможно только выборочное наблюдение.

Точность оценок, полученных на основе выборочного метода, зависит не от доли обследованных единиц, а от их числа. Если объем генеральной совокупности достаточно велик, то доля отобранных для наблюдения единиц может быть очень небольшой, а точность при этом - высокой. Например, выборочное обследование по проблемам занятости в России охватывает около 0.2% населения в возрасте от 15 до 72 лет, но обеспечивает высокую точность. Вместе с тем при выборочном наблюдении определенные отклонения выборочных значений параметров от значений для генеральной совокупности неизбежны. Эти отклонения называются *ошибками репрезентативности* и подразделяются на систематические и случайные. *Систематические* ошибки возникают в связи с принятым способом отбора или нарушением его правил. *Случайные* ошибки при выборочном обследовании неизбежны и связаны с невозможностью полностью адекватного воспроизведения генеральной совокупности в выборке. Оценка

случайных ошибок - одна из задач статистики. Стремление максимально уменьшить случайную ошибку приводит к росту выборки, а большая ошибка ставит под сомнение возможность практического использования результатов.

При организации выборочного обследования нужно соблюдать принцип случайности отбора. Каждая единица должна иметь равную вероятность попасть в выборку. Отбор, организованный по этому принципу, называется *случайным*. На практике не всегда удается обеспечить его выполнение, так как для этого необходимо учесть все элементы генеральной совокупности. Другими методами формирования выборок является механический, типический и гнездовой отбор, рассматриваемые в следующем параграфе. В примерах, демонстрирующих применение различных алгоритмов отбора, с целью компактности представления будут рассматриваться генеральные совокупности и выборки с очень малым числом единиц.

2.1 Методы формирования выборок

Собственно случайный отбор

Рассмотрим наиболее важный способ формирования выборочной совокупности - *собственно случайный отбор*. Собственно случайный отбор может быть *повторным* и *бесповторным*. При повторном отборе каждая единица, отобранная в случайном порядке из генеральной совокупности, после проведения наблюдения возвращается в совокупность и может быть повторно подвергнута обследованию. На практике такой вид отбора встречается редко. Гораздо больше распространен бесповторный отбор, при котором обследованные единицы в совокупность не возвращаются. При повторном отборе вероятность попадания в выборку для каждой единицы совокупности остается неизменной. При бесповторном отборе она меняется, но для всех единиц, оставшихся в генеральной совокупности после отбора одной или нескольких единиц, вероятность попадания в выборку одинакова.

Для обеспечения случайности используются два основных способа. При небольшом объеме совокупности все ее элементы перенумеровываются и записываются, например, на карточках, из которых затем случайным образом «вслепую» выбирается столько карточек, сколько элементов предполагается включить в выборку. При большом количестве единиц в совокупности для случайного отбора используются таблицы равномерно распределенных *случайных чисел*. Основным свойством случайных чисел является то, что любая цифра от 0 до 9 в любой позиции этих чисел встречается с равной вероятностью $1/10$. Это не означает, что в любом наборе из 10 цифр каждая цифра встретится один раз, но если взять большое количество таких наборов и посчитать, сколько раз каждая цифра встретится в среднем, то результат будет приближенно равен $1/10$ от общего количества. Например, если в каком-то из наборов цифра 2 встретила 3 раза, то в дальнейшем обязательно будут наборы, где 2 вообще

не будет. В основе генерации случайных последовательностей лежат довольно сложные математические алгоритмы.

Пример. Построить с помощью таблицы случайных чисел выборку из 10 элементов для генеральной совокупности 48 городов России, табл. 2.1.

Таблица 2.1. Список городов России

1. Барнаул	13. Калининград	25. Омск	37. Тамбов
2. Великий Новгород	14. Киров	26. Оренбург	38. Тверь
3. Владивосток	15. Комсомольск-на- А.	27. Пенза	39. Тольятти
4. Владимир	16. Кострома	28. Пермь	40. Томск
5. Волгоград	17. Краснодар	29. Петрозаводск	41. Тюмень
6. Воронеж	18. Магнитогорск	30. Псков	42. Ульяновск
7. Екатеринбург	19. Москва	31. Ростов-на-Дону	43. Уфа
8. Иваново	20. Набережные Челны	32. Рязань	44. Улан-Удэ
9. Ижевск	21. Нальчик	33. Самара	45. Хабаровск
10. Иркутск	22. Нижний Новгород	34. Санкт-Петербург	46. Чебоксары
11. Кав.Мин.Воды	23. Новокузнецк	35. Саратов	47. Челябинск
12. Казань	Новосибирск	36. Стерлитамак	48. Ярославль

Таблица 2.2. Таблица двузначных равномерно распределенных случайных чисел

7	76	56	79	91	83	70	6	35	31
9	19	98	71	77	17	15	52	73	65
78	89	33	80	15	64	90	69	17	91
5	36	54	47	55	28	43	68	16	47
26	73	49	9	24	96	25	16	33	41
84	32	57	78	41	35	49	84	25	25
87	81	79	30	37	49	100	73	27	83
31	53	65	14	34	54	16	75	93	8
35	27	83	99	88	71	42	26	5	100
7	77	20	16	61	6	57	66	70	91

Решение. Пронумеруем все города в некотором порядке, табл. 2.1. Произвольно выберем в таблице 2.2 равномерно распределенных случайных чисел некоторое число в качестве начального (например, закрыв глаза, укажем пальцем в произвольную точку таблицы). Предположим, что это оказалось число 41. Выпишем это и следующие за ним 9 чисел и начнем выбирать соответствующие им по номеру города из таблицы городов:

41 Тюмень	83
37 Тамбов	17 Краснодар
34 С.-Петербург	64

88	28 Пермь
61	96

При этом ряд чисел оказались больше, чем число городов, и поэтому реально удалось выбрать только пять городов. Выбираем далее из таблицы 2.2 еще 5 подходящих (не превышающих 48) чисел 35, 49, ~~54~~, ~~71~~, 6, ~~70~~, 15, ~~90~~, 43, 25 и добавляем в список пять городов:

41 Тюмень	35 Саратов
37 Гольягти	6 Воронеж
34 С.-Петербург	15 Комсомольск на А.
17 Краснодар	43 Уфа
28 Пермь	25 Омск

Выборка сформирована.

Недостатком случайной схемой отбора является то, что пронумеровать все единиц очень большой совокупности достаточно сложно. В ряде случаев пронумерованный список может уже существовать. Например, если исследуется совокупность всех работников предприятия, то их список наверняка есть в отделе кадров или бухгалтерии.

В рассмотренной таблице представлены только двузначные случайные числа. Существуют таблицы случайных чисел и с большим числом десятичных знаков (трех-, четырех- и даже пятизначные), но при их отсутствии для получения, например, четырехзначных случайных чисел можно воспользоваться двумя стоящими в соседних столбиках двузначными числами, табл. 2.2: 75 и 93 ---> 7593.

Механический отбор

Суть механического отбора заключается в том, что все элементы выборки сначала перенумеровываются, а затем из них выбирается, например, каждый третий, четвертый, пятый, в зависимости от того, какое количество единиц нужно включить в выборку.

Пример. Построить методом механического отбора выборку из 10 единиц для совокупности городов из предыдущего примера, табл. 2.1.

Решение. Так как нужно отобрать 10 городов из 48, то, разделив 48 на 10, получаем, что отобран должен быть примерно каждый пятый. Случайным образом, используя таблицу случайных чисел, выберем начальный номер, например, 4. После этого, начиная с 4, отбираем каждый пятый город и включаем его в список: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 и т.д. Последним десятым элементом

должен быть город с номером $44 + 5 = 49$. Так как в списке всего 48 городов, отбор продолжается с начала списка: $49 - 48 = 1$. Выборка построена.

- | | |
|-----------------|------------------|
| 4. Владимир | 29. Петрозаводск |
| 9. Ижевск | 34. С. Петербург |
| 14. Киров | 39. Тольятти |
| 19. Москва | 44. Улан-Удэ |
| 24. Новосибирск | 1. Барнаул |

Достоинствами механической схемы отбора является простота. Недостатком механического отбора является возможность внесения систематической ошибки, связанной со способом сортировки единиц совокупности. Например, если при обследовании семейных пар их данные заносили в список в определенном порядке:

1. Жена
2. Муж
3. Жена
4. Муж,

то механический отбор с четным шагом даст выборку, включающую только женщин или только мужчин.

Типический отбор

При использовании схемы типического (стратифицированного) отбора генеральная совокупность подразделяется на группы по одному или нескольким признакам и затем из каждой группы производится отбор нескольких единиц, чтобы обеспечить ее представленность в выборке.

Пример. Построить для совокупности 20 студентов, табл. 2.3, выборку из 8 единиц, используя схему типического отбора и группировку по признакам "Пол" и "Курс".

Таблица 2.3. Список студентов

№	Фамилия Имя	Пол	Курс	№	Фамилия Имя	Пол	Курс
1	Альтский Петр	М	1	11	Мартыненко Евгения	Ж	1
2	Барков Денис	М	2	12	Мельников Георгий	М	1
3	Бронштейн Тамара	Ж	1	13	Орлов Геннадий	М	2
4	Викторенко Светлана	Ж	1	14	Павлюкевич Татьяна	Ж	2
5	Владиславлева Наталья	Ж	1	15	Петров Анатолий	М	1
6	Дробышев Михаил	М	1	16	Саакян Регина	Ж	1
7	Дроздов Алексей	М	1	17	Тимохин Сергей	М	2
8	Жеглова Екатерина	Ж	2	18	Трофимова Ольга	Ж	2

Решение. Сначала разделим совокупность на две группы, включив в первую всех женщин, во вторую всех мужчин, табл. 2.4.

Таблица 2.4. Список студентов, сгруппированный по признаку "Пол"

Женщины				Мужчины			
№	Фамилия Имя	Пол	Курс	№	Фамилия Имя	Пол	Курс
1	Бронштейн Тамара	Ж	1	1	Альтский Петр	М	1
2	Викторенко Светлана	Ж	1	2	Барков Денис	М	2
3	Владиславлева Наталья	Ж	1	3	Дробышев Михаил	М	1
4	Жеглова Екатерина	Ж	2	4	Дроздов Алексей	М	1
5	Мартыненко Евгения	Ж	1	5	Левицкий Альберт	М	2
6	Павлюкевич Татьяна	Ж	2	6	Лунев Сергей	М	2
7	Саакян Регина	Ж	1	7	Мельников Георгий	М	1
8	Трофимова Ольга	Ж	2	8	Орлов Геннадий	М	2
9	Юрлова Юлия	Ж	2	9	Петров Анатолий	М	1
10	Янчева Александра	Ж	2	10	Тимохин Сергей	М	2

Далее, каждую их двух групп разделим еще на две, отделив первокурсников от студентов второго курса, табл. 2.5.

Таблица 2.5. Список студентов, сгруппированный по признакам "Пол" и "Курс"

Группа 1				Группа 2				
1	Бронштейн Тамара	Ж	1	1	Альтский Петр	М	1	
2	Викторенко Светлана	Ж	1	2	Дробышев Михаил	М	1	
3	Владиславлева Наталья	Ж	1	3	Дроздов Алексей	М	1	
4	Мартыненко Евгения	Ж	1	4	Мельников Георгий	М	1	
5	Саакян Регина	Ж	1	5	Петров Анатолий	М	1	
Группа 3				Группа 4				
1	9	Левицкий Альберт	Ж	2	1	Барков Денис	М	2
2		Павлюкевич Татьяна	Ж	2	2	Левицкий Альберт	М	2
3		Трофимова Ольга	Ж	2	3	Лунев Сергей	М	2
4		Юрлова Юлия	Ж	2	4	Орлов Геннадий	М	2
5		Янчева Александра	Ж	2	5	Тимохин Сергей	М	2

Так как всего в выборке должно быть 8 человек, то из каждой группы должно быть отобрано двое. Для этой цели воспользуемся случайными числами, которые оказались (например) следующими:

Группа 1 - Студенты 5 и 4 Группа 2 - Студенты 1 и 3
Группа 3 - Студенты 5 и 2 Группа 4 - Студенты 3 и 4

Выборка сформирована, табл. 2.6.

Таблица 2.6. Выборка, сформированная методом типического отбора

5	Саакян Регина	Ж	1	1	Альтский Петр	М	1
4	Мартыненко Евгения	Ж	1	3	Дроздов Алексей	М	1
5	Янчева Александра	Ж	2	3	Лунев Сергей	М	2
2	Павлюкевич Татьяна	Ж	2	4	Орлов Геннадий	М	2

Разновидностью типического отбора является районированный отбор, при котором отбор единиц для наблюдения осуществляется из групп, соответствующих административно-

территориальному делению.

Основным достоинством типического отбора является то, что он обеспечивает представительность в выборке всех подгрупп совокупности, важных для данного исследования. У метода есть, однако, и два существенных недостатка. Если количество признаков, наблюдаемых у единиц совокупности, велико, то построение соответствующего разбиения на подгруппы становится сложной задачей и требует больших усилий. Кроме этого, если атрибутивный признак является сложно формализуемым, например, характеризует доверие или отношение человека к чему-либо, то разбиение на подгруппы по этому признаку произвести непросто.

Серийный отбор

Этот метод применяется в случае, если генеральная совокупность разбита на группы (серии) еще до начала выборочного исследования. Тогда в выборку включаются не отдельные единицы генеральной совокупности, а целые серии.

Например, генеральной совокупностью является партия товаров, упакованных в коробки. Тогда для выборочного контроля качества можно применить серийный отбор и обследовать *все* единицы из *одной* коробки. Если для аналогичного исследования применить метод случайного отбора, то единицы будут браться из разных коробок и затраты на проведение исследования (на распаковку и упаковку коробок) значительно возрастут.

Другой пример. Предполагается провести обследование зубов у школьников с применением переносного рентгеновского аппарата с целью выяснить, какой процент детей имеет здоровые зубы. В рамках серийного метода формирования выборки, исследователь выбирает один из классов какой-либо школы, приезжает туда с аппаратом и всех обследует. Для того чтобы обследовать такое же количество детей из разных школ и классов пришлось бы затратить намного больше средств и времени для перевозки аппарата.

Приведенные примеры иллюстрируют основные достоинства метода - он уменьшает затраты, упрощает работу «на выезде» и просто удобен. Основным же недостатком является то, что единицы в выбранной серии могут иметь распределение (вариацию) признака, отличающееся от совокупности в целом. Причиной является то, что единицы в серии (классе, клубе, квартале города) являются, как правило, более однородными.