**Лекция 4. Систематическая выборка.**

***Систематическая* выборка** по качеству часто приближается к простой случайной. Систематическая выборка, как и простая случайная, требует полного списка или заданного упорядочения совокупности. Техника осуществления систематического отбора элементарна: сначала случайным образом отбирается первая единица, затем отбору подлежит каждый *k*-й элемент. Число *k* в данном случае называют шагом отбора. Можно, например, отбирать каждый 25-й или каждый 200-й элемент. Чтобы определить шаг отбора, нужно поделить известный объем генеральной совокупности (*N*) на предполагаемый объем выборки (*n*).

Пусть, например, нужно отобрать 200 человек из 20000 владельцев телефонов:

1. Определим шаг отбора: N/n = 20000/200=100;
2. С помощью таблицы случайных чисел найдем первую выборочную единицу. Если, скажем, выпал номер «053», то из списка владельцев телефонов выпишем того, кто значится под этим номером;
3. С установленным шагом отбираем номера: 153, 253, 353, 453 и т.д. до исчерпания списка.

Иногда генеральная совокупность (и соответственно основа выборки) слишком велика либо исследователю известен не полный список, а лишь правило упорядочения элементов в генеральной совокупности. Предположим, что мы хотим составить представление о весе и формате книг, содержащихся в некой библиотеке, при том, что мы не располагаем полным каталогом, а лишь видим, как книги расставлены на стеллажах. При условии, что объем библиотечного собрания нам приблизительно известен, мы можем воспользоваться процедурой систематического отбора и отобрать, скажем, каждую 55-ю книгу. Очень важно отобрать «стартовую» единицу сугубо случайным образом. Именно в этом пункте кроется основная слабость систематического отбора. Если в способе упорядочения единиц совокупности имеет место некая цикличность, т.е. неизвестная нам «система» (систематический паттерн), а случайность выборе «старта» должным образом не обеспечена, то полученная выборка может также оказаться смещенной (если о систематическом паттерне мы знаем заранее, то он не представляет собой угрозы валидности и может быть учтен входе отбора)

Чем хороша простая случайная выборка?

Тем, что при достаточном размере *n* в ней будут представлены все категории людей, присутствующие в списке, из которого она отбиралась, и примерно в тех же самых пропорциях. А это значит, что исследователю не надо думать о том, сколько надо опросить мужчин и сколько женщин, сколько молодых и сколько пожилых, сколько богатых и сколько бедных. Все эти пропорции будут с большой вероятностью выдержаны в простой случайной выборке.

Например, если вся совокупность, которая насчитывает 10 тысяч человек, на 45% состоит из мужчин и на 55% – из женщин, то в выборке из 1000 человек пропорции мужчин и женщин будут примерно такие же. Конечно, нельзя рассчитывать на то, что мужчин будет ровно 450 человек, а женщин – 550. Возможны случайные отклонения от точных пропорций, но они будут невелики.

При увеличении размера выборки n погрешность. Можно даже подобрать размер выборки *n* так, чтобы погрешность стала такой, какой нужно сследователю. Формулу для вычисления *n* несложно получить из приведенной выше формулы для вычисления Δ.

Простая случайная выборка всегда создает приближенную копию всей совокупности, точность которой возрастает с увеличением размера выборки. Это справедливо для всех параметров, в том числе и для места жительства. Дома и квартиры людей, включенных в выборку, будут равномерно распределены по всей территории, на которой проживают представители изучаемой совокупности. Если это город, то по всем районам города, если это Россия, то по всей территории России. Это обеспечивает хорошую

географическую представительность выборки, но одновременно создает дополнительные сложности при проведении опросов. Чтобы опросить

нескольких человек, интервьюеру придется совершать путешествия из одной части города в другую или из одного населенного пункта в другой. А это сильно повышает стоимость исследования.

Подведем итог. Простая случайная выборка, обладая несомненными достоинствами, такими как простота реализации, хорошее воспроизведение структуры совокупности, возможность вычисления доверительных интервалов, имеет также ряд недостатков:

- для реализации выборки необходимо иметь список всех представителей совокупности;

- стоимость исследования сильно возрастает из-за удаленности респондентов друг от друга;

- статистическая погрешность возникает по всем параметрам выборки, даже по тем, для которых известны истинные пропорции (например,

по полу или возрасту).

Для устранения перечисленных недостатков используются два специальных приема формирования выборки – стратификация и кластеризация, к рассмотрению которых мы сейчас перейдем.