**Лекция 6. Кластерная выборка.**

Кластеризация позволяет включать в выборку респондентов, проживающих на небольшом расстоянии друг от друга, сохраняя при этом случайный механизм их отбора. Это достигается путем объединения людей в группы, которые участвуют в отборе как самостоятельные единицы. Такие

группы называются **кластерами**. Чаще всего в качестве кластеров используют различные территориальные образования. Это могут быть административные районы, населенные пункты, городские микрорайоны, городские кварталы, территории избирательных округов или избирательных участков и т. п. В роли кластеров могут также выступать предприятия при опросе рабочих и служащих, учебные заведения при опросе учащихся, магазины при опросе продавцов.

Для получения выборки надо сначала отобрать нужное число кластеров, а затем в каждом из отобранных кластеров отобрать нужное число респондентов, т. е. отбор надо проводить в два этапа. На первом этапе в отборе участвуют кластеры, на втором – люди.

Прежде чем приступить к отбору кластеров, надо составить их полный список. Каждый человек, входящий в изучаемую совокупность,

должен быть отнесен к какому-либо кластеру, причем только к одному. Составление полного списка кластеров представляет гораздо меньше

проблем, чем составление полного списка людей. Особенно тогда, когда кластерами служат единицы административно-территориального деления. Например, списки всех административных районов, а также всех городов и поселков городского типа России (с указанием числа жителей) ежегодно публикуются Федеральной службой госстатистики. Их вполне можно использовать в качестве кластеров при опросе населения.

Одна из задач кластеризации состоит в том, чтобы сократить время и затраты на перемещение интервьюера от респондента к респонденту в пределах кластера. Желательно, чтобы это время не превышало 10–15 минут. Если отобранные кластеры имеют слишком большую территорию и не обеспечивают выполнения данного требования, приходится проводить еще один этап или ступень отбора. При этом кластеры, которые отбираются сначала, на первой ступени, называются первичными единицами отбора (ПЕО).

Внутри них формируются более мелкие кластеры, которые называются единицами отбора второй ступени или вторичными единицами отбора (ВЕО). Вторая ступень отбора проводится только в тех кластерах, которые были отобраны на первой ступени. Например, если на первой ступени проводился отбор административных районов России, то на второй ступени могут отбираться населенные пункты районов, попавших в выборку.

В кластерах, отобранных на второй ступени, можно провести отбор еще более мелких кластеров. Например, в городах можно провести отбор

микрорайонов, кварталов или избирательных участков. Это будет третья ступень отбора. На последней ступени отбираются люди (или другие элементы, из которых состоит изучаемая совокупность и которые являются объектом исследования). В зависимости от числа ступеней отбора выборка будет называться двухступенчатой, трехступенчатой и т.д.

Выборка, в которой на начальных этапах отбираются кластеры, а на последнем этапе – люди (представители совокупности), называется **многоступенчатой** или **кластерной**. В некоторых изданиях на русском языке кластеры называются гнездами, а кластерная выборка – гнездовой.

Использование кластерной выборки избавляет исследователя от необходимости составлять полный список всех представителей совокупности. Вместо этого составляются списки кластеров: первичных единиц отбора – для всей совокупности, вторичных единиц отбора – для тех

ПЕО, которые попали в выборку на первой ступени, и т.д. Списки людей нужны только для проведения последней ступени отбора. Они составляются для тех небольших по размеру кластеров, которые были отобраны на предпоследней ступени. При опросах по месту жительства списки людей заменяются списками домохозяйств. Эти списки могут быть получены на основе домовых книг жилищно-эксплуатационных организаций или сельских администраций, а могут быть составлены интервьюером непосредственно на местности, что достаточно просто для небольших кластеров.

Таким образом, у кластерной выборки отсутствуют два главных недостатка простой случайной выборки – не требуется список всех представителей совокупности и интервьюер имеет возможность опросить нескольких человек, проживающих на небольшом расстоянии друг от друга.

Кластерная выборка является случайной, т. е. для каждого человека (элемента совокупности) должна быть обеспечена определенная (желательно – равная) вероятность попасть в выборку.

Для этого кластеры должны отбираться с вероятностью, пропорциональной числу элементов в кластере. Такой способ отбора часто называют **ВПР\_отбором** (по первым буквам слов “вероятность, пропорциональная размеру”) или **PPS\_отбором** (от аналогичного английского выражения “probability proportional to the size”). Число

элементов совокупности в кластере называют **размером** кластера.

Если отбирать кластеры с вероятностью, пропорциональной размеру, а людей внутри кластера – с равной вероятностью, то для любого человека из изучаемой совокупности будет обеспечена одинаковая вероятность попадания в выборку.

Кластерная выборка получается дешевле простой случайной. За уменьшение стоимости приходится платить увеличением статистической погрешности. Потеря точности кластерной выборки происходит из-за того

же, из-за чего уменьшается ее с т о и м о с т ь , а именно из-за группировки респондентов внутри кластеров. Респонденты, живущие недалеко друг от друга, часто дают похожие или даже одинаковые ответы на вопросы анкеты.