**Лекция 12. Репрезентативные выборки**

Термин “репрезентативная выборка” применяется довольно часто, однако разные люди вкладывают в это понятие различный смысл. Но практически все придают ему положительный оттенок, подразумевая, что репрезентативная выборка – это, прежде всего, хорошая, правильная выборка.

Если переводить этот термин буквально, то он означает выборку, репрезентирующую, то есть представляющую, изучаемую совокупность. Причем представляющую правильно, в отличие от нерепрезентативной выборки, которая не представляет изучаемую совокупность или представляет ее неправильно. Правильное представление совокупности означает в первую очередь отсутствие смещений, поэтому термин “репрезентативность” ближе всего примыкает к понятию несмещенности выборки. Ведь результаты опросов, проведенных по несмещенной выборке, можно смело распространять на всю совокупность. Именно это имеют в виду люди, когда говорят, что выборка репрезентирует все население или какую-то его часть – сельских жителей, молодежь, автовладельцев и т. п.

Еще один смысл, который вкладывают в понятие “репрезентативность”, заключается в том, что в выборке должны быть представлены все категории респондентов, что ни одна из категорий не должна быть пропущена при отборе. Такое понимание репрезентативности очень близко к определению случайной (вероятностной) выборки как выборки, в которой каждый представитель совокупности имеет известную

ненулевую вероятность быть отобранным. Использование термина в таком контексте, на первый взгляд, нисколько не противоречит пониманию репрезентативности в смысле несмещенности выборки, поскольку случайные выборки являются несмещенными, – но все же некоторые различия есть. Например, в непропорциональной стратифицированной выборке отношение между отдельными категориями людей может не совпадать с их пропорциями в совокупности. Размер выборки по некоторым категориям может быть сознательно увеличен или уменьшен, а несмещенность средних оценок обеспечивается за счет использования при расчетах весов страт. Некоторые авторы называют такие выборки нерепрезентативными, хотя и признают необходимость и обоснованность нарушения пропорций между разными категориями респондентов.

Отсутствие в выборке представителей отдельных малочисленных категорий людей, доля которых в совокупности не превышает величину статистической погрешности, тоже порой интерпретируется как нерепрезентативность выборки, хотя с точки зрения теории вероятности это вполне допустимо. Таким образом, в число нерепрезентативных во втором смысле попадают некоторые несмещенные случайные выборки, что свидетельствует о различии между таким пониманием репрезентативности и ее пониманием в смысле несмещенности.

Бывает, что после слов “репрезентативная выборка” идет уточнение, по каким именно параметрам она репрезентативна. Например, выборка репрезентативна по полу и возрасту, по типу места жительства, по национальному составу респондентов и т. д. Независимо от того, что имеет в виду автор, эти слова часто понимаются буквально. А именно, что пропорции в выборке и во всей совокупности по указанным параметрам совпадают. Нетрудно заметить, что это не совсем то же самое, что несмещенность выборки по этим параметрам. В простой случайной выборке возможны небольшие отклонения по любому параметру в пределах статистической погрешности. Так величина погрешности для доли мужчин в выборке из 1000 человек равна ± 2,9%. Отклонения в этих пределах не означают смещенности выборки, однако могут восприниматься отдельными людьми как отсутствие ее репрезентативности по полу, то есть как признак некачественной выборки. И наоборот, репрезентативность по ряду параметров, в которой можно легко убедиться самостоятельно, часто воспринимается как надежное доказательство несмещенности выборки по всем параметрам, что, как известно, справедливо не всегда. Поэтому “репрезентативность по параметру” отличается от рассмотренных ранее понятий репрезентативности и не позволяет провести четкую границу между смещенными и несмещенными выборками, случайными и неслучайными.

Иногда говорят о статистической репрезентативности выборки, имея в виду, что ее точность должна соответствовать целям исследования. Для решения одних задач статистическая погрешность должна быть небольшой, например, не более 3%, для других задач приемлемая погрешность может составлять 10% и даже больше.

Когда величина погрешности отвечает целям исследования, выборку называют статистически репрезентативной. Такое определение применимо только к случайным выборкам, для которых можно вычислить статистическую погрешность.

Случайные выборки, имеющие недостаточную точность, попадают в категорию статистически нерепрезентативных.

Таким образом, понятие “репрезентативная выборка” имеет много значений, не совпадающих одно с другим. Бывает трудно понять, в каком смысле этот термин употребляется. Порой плохие выборки могут быть названы репрезентативными, и наоборот, хорошие, правильные выборки могут попасть в категорию нерепрезентативных.

Наверное, если подсчитать, какие слова чаще всего встречаются вместе со словом выборка, то сочетание “репрезентативная выборка” займет одно из первых мест. По распространенности его сможет опередить разве что сочетание “случайная выборка”. Возможно, такое широкое употребление этих словосочетаний объясняется многообразием значений, которые в них вкладываются. Об этом важно помнить, встречая или используя эти выражения.