**Лекция 10-11. Расчет объема выборки**

В формировании выборочной совокупности важную роль играет определение ее объема и обеспечение репрезентативности.

Если тип выборки говорит о том, как попадают люди в выборочную совокупность, то объём выборки сообщает о том, какое их количество попало сюда. То есть объем выборки – это количество единиц попавших в выборочную совокупность. И очень важно, чтобы выборка была репрезентативной, то есть не искажала представлений о генеральной совокупности в целом.

Требования репрезентативности выборки означают, что по выделенным параметрам (критериям) состав обследуемых должен приближаться к соответствующим пропорциям в генеральной совокупности.

Одна из ключевых проблем, встающих, как правило, перед социологом, решающим: доверять полученным в ходе него данным или нет, это то, сколько же человек должно быть опрошено для того, чтобы получить действительно репрезентативную информацию. К сожалению, единой и четкой формулы, используя которую можно было бы рассчитать оптимальный объем выборочной совокупности, не существует в природе. И объясняется это весьма просто. Дело в том, что определение объема выборочной совокупности – это проблема не столько статистическая, сколько содержательная.

Иными словами, объем выборочной совокупности зависит от множества факторов, основные из них следующие:

1. затраты на сбор информации, включая временные;

2. стремление к определённой статистической достоверности результатов, которую надеется получить исследователь;

3. ценность и новизна информации, получаемой в результате опроса.

Объем выборки обусловлен степенью однородности или неоднородности, генеральной совокупности, количеством характеризующих ее признаков. Однородной считается совокупность, в которой контролируемый признак, например уровень грамотности, распределён равномерно, то есть не образует пустот и сгущений, тогда опросив лишь несколько человек, можно сделать вывод о том, что большинство людей грамотны. Чем более однородна генеральная совокупность, тем меньше объем выборки. Например, «допустим, мы осуществляем отбор из генеральной совокупности в 2000 человек, контролируя состав выборочной совокупности по признаку «пол»»: 70% мужчин и 30% женщин. Согласно теории вероятности, можно предположить, что примерно среди каждых десяти отбираемых респондентов встретятся три женщины. Если мы хотим опросить по крайней мерее 90 женщин, то исходя из вышеупомянутого соотношения, нам необходимо отобрать не менее 300 человек. А теперь предположим, что в генеральной совокупности 90% мужчин и 10% женщин. В этом случае, чтобы в выборочную совокупность попало 90 женщин, необходимо отобрать уже не менее 900 человек». Из примера видно, что объем выборки зависит от разброса признака (дисперсии), и его нужно вычислять по признаку, дисперсия значений которого наибольшая.

Степень однородности социального объекта зависит, в сущности, от того, насколько детально мы намерены его исследовать. Практически любой, самый «элементарный» объект оказывается чрезвычайно сложным. Лишь в анализе мы представляем его как относительно простой, выделяя те или иные его свойства. Чем более основательным и детальным будет анализ, чем больше свойств данного объекта мы намерены принять во внимание в их сочетании, а не изолированно, тем больше должен быть объем выборки.

Существуют, так называемые «правила левой руки» для определения размера выборки :

|  |  |
| --- | --- |
| **Размер выборки растёт** | **Размер выборки уменьшается** |
| - при необходимости опубликовать данные для отдельных подгрупп (размеры подвыборок при этом суммируются, и выборка в целом растёт пропорционально числу подгрупп); | - при исследовании организаций, институтов и прочих «первичных единиц отбора», если сравнительно невелика величина генеральной совокупности, из которой производится отбор(например, совокупности сотрудников рекламных агентств, школьников, пациентов и т.п.); |
| - при проведении общенациональных обследований, когда велика генеральная совокупность; | - при проведении локальных и региональных исследований; |
| - если уже имеющаяся информация  по ключевым вопросам (например, о намерениях избирателей голосовать  за ту или иную партию) явно недостаточна и степень неопределённости значительна. | - если уже существующая информация относительно полна и всё ещё остающаяся степень неопределенности незначительна. |

В репрезентативной выборке все элементы генеральной совокупности представлены в той же пропорции. Но как бы тщательно не соблюдать этот принцип, случайные ошибки все же будут. Мы имеем возможность определять ошибку репрезентативности. Ошибкой репрезентативности, как правило, называют «расхождение между двумя совокупностями – генеральной, на которую направлен теоретический интерес социолога и представление о свойствах которой он хочет получить в конечном итоге, и выборочной, на которую направлен практический интерес социолога, которая выступает одновременно как объект обследования и средство получить информацию о генеральной совокупности».

Важно учитывать, что при помощи выборочного метода никогда нельзя получить абсолютно точную оценку наблюдаемого признака, всегда существует вероятность ошибки, но, если вероятность ошибки мала, то она скорее всего не произойдет.

В отечественной литературе наряду с термином «ошибка репрезентативности» встречается и другой – «ошибка выборки». Обычно они используются как синонимы, но понятие «ошибка выборки» количественно более точное, чем «ошибка репрезентативности». Ошибка выборки – это «отклонение средних характеристик выборочной совокупности от средних характеристик генеральной совокупности. На практике она определяется путём сравнения известных характеристик генеральной совокупности с выборочными средними».

Репрезентативность выборки определяется двумя компонентами: систематическими и случайными ошибками. Случайные ошибки связаны «со статистическими погрешностями (зависят от динамики исследуемых признаков) и непредвиденными нарушениями процедуры сбора информации (процедурные ошибки, допущенные при регистрации признаков)». Случайные ошибки уменьшаются с увеличением объема выборочной совокупности. Случайную ошибку можно измерить методами математической статистики, если при формировании выборочной совокупности соблюдался принцип случайности, обеспечивающийся строго определенными правилами, которые составляют метод формирования выборочной совокупности, и устранить.

В.И. Паниотто приводит следующие расчёты репрезентативной выборки с допущением 5-процентной ошибки (здесь имеются ввиду случайные ошибки) (таблица 2):

Размеры выборки для различных генеральных совокупностей

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Объём генеральной совокупности | 500 | 1000 | 2000 | 3000 | 4000 | 5000 | 10000 | 100000 |
| Объем выборки | 222 | 286 | 333 | 350 | 360 | 370 | 385 | 458 |

Например, если опросив 380 человек в поселении, где общая численность платёжеспособного населения 10 тысяч человек, мы выявили что 36% опрошенных покупателей, предпочитают отечественную продукцию, то с 95-процентной степенью вероятности мы можем утверждать, что отечественную продукцию постоянно покупают 46±5% (то есть от 41 до 51%) жителей этого поселения.

Можно было также воспользоваться расчётами института Гэллапа для оценки соотношения размеров выборки и ошибки выборки:

|  |  |
| --- | --- |
| Размер выборки | Интервал доверия, % |
| 4000 | ±2 |
| 1500 | ±3 |
| 1000 | ±4 |
| 600 | ±5 |
| 400 | ±6 |
| 200 | ±8 |
| 100 | ±11 |