

1.4. ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ, ВЫБОРКА И ВЫБОРОЧНЫЙ МЕТОД

Поскольку охватить всю совокупность наблюдений за социально-экономическими явлениями невозможно, прогнозисту приходится использовать только некоторую часть из них. Или, иначе говоря, из всего множества имеющихся наблюдений он использует только некоторую часть из них, которая некоторым образом выбрана им из всего множества. Значит ли это, что прогнозист о всей совокупности наблюдений может говорить как о генеральной совокупности, а о некоторой части этой совокупности – как о выборочной? Ответ на этот вопрос является очень важным, поскольку под понятиями «генеральная совокупность» и «выборочная совокупность» скрываются вполне определённые и чётко очерченные в границах математической статистики характеристики соответствующих объектов. Поэтому, прежде чем использовать их, необходимо уяснить, что же они собой представляют. Тогда прогнозист не допустит терминологической путаницы и не ошибётся в выборе инструмента прогнозирования.

О том, что простого ответа на поставленный вопрос нет, может свидетельствовать следующая цитата: «Анализируемые ряды динамики являются почти всегда выборками из более длинных рядов»¹. А для обработки выборочных значений используется выборочный метод, а он подразумевает использование методов математической статистики. Но в предыдущем параграфе было показано, что, например, для эволюционных процессов методы математической статистики неприемлемы. Следовательно, далеко не всегда прогнозист может говорить о выборке и выборочном методе.

Итак, прежде всего, рассмотрим такое понятие, как «генеральная совокупность». Определения типа: «та совокупность, из которой проводится отбор, называется генеральной совокупностью; отобранные данные составляют выборочную совокупность»² рассматривать не будем, поскольку в таких определениях нет существенных свойств определяемых понятий.

«Генеральной совокупностью называют совокупность всех мыслимых наблюдений (или всех мысленно возможных объектов интересующего нас типа, с которых «снимаются» наблюдения), которые могли бы быть произведены при данном реальном комплексе условий. Поскольку в определении речь идет о всех мысленно возможных наблюдениях (или объектах), то понятие генеральной совокупности есть понятие условно-математическое, абстрактное и его не следует смешивать с реальными совокупностями, подлежащими статистическому исследованию. Так, обследовав даже все предприятия подотрасли с точки зрения регистрации значений характеризующих их технико-экономических показателей, мы можем рассматривать обследованную

¹ Вайну Я.Я.-Ф. Корреляция рядов динамики. – М.: Статистика, 1977. - С. 90.

² Елисеева И.И., Юзбашев М.М. Общая теория статистики: Учебник / Под ред. И.И.Елисеевой. – М.: Финансы и статистика, 2004. – С. 216; Справочник по математике для экономистов: Учеб. пособие / Пед ред. проф. В.И.Ермакова. – М.: ИНФРА-М, 2007. - С. 372; Доугерти К. Введение в эконометрику. – М.: ИНФРА-М, 1997. - С. 4 и др.

совокупность лишь как представителя гипотетически возможной более широкой совокупности предприятий, которые могли бы функционировать в рамках того же самого реального комплекса условий»¹.

Авторы этого определения постарались дать его как можно более понятным, но, как следует из приведённого текста, в погоне за понятностью потеряли чёткость определения. Если, например, представить, что за всю многовековую историю существования Руси, а затем и России социально-экономические показатели этого объекта исследования представляют собой все мысленно возможные состояния, то это – генеральная совокупность или нет? А если к этому множеству добавить и все возможные состояния социально-экономических показателей России в необозримом будущем, то тогда полученное множество является генеральной совокупностью или нет? Из приведённого выше определения однозначный ответ на этот вопрос получить невозможно.

В одной из фундаментальных работ по теории вероятностей и математической статистике об этом говорится так: «Пусть имеется многочисленная совокупность однородных элементов (объектов), каждый из которых может обладать или не обладать каким-либо признаком; неизвестная нам доля тех из них, которые обладают этим признаком, и подлежит определению. Наше испытание заключается в том, что мы выбираем наугад один элемент из множества элементов, отмечаем, обладает или нет этот элемент данным признаком и возвращаем его обратно в совокупность. При выборе элемента из совокупности принимаются меры к тому, чтобы вероятность быть выбранным была одинакова для всех элементов. Тогда имеющееся множество элементов, называется *генеральной* совокупностью. Группа из n элементов, наблюдаемых при повторных испытаниях, называется *случайной выборкой*, число отобранных элементов — объемом выборки, а описанный процесс отбора элементов — простым случайным выбором. Определив частоту признака среди отобранных в выборке объектов, мы можем по ней, опираясь на теорему Лапласа, приближенно оценить долю признака в генеральной совокупности: в самом деле, эта доля в данных условиях играет роль неизменной вероятности появления объекта, обладающего интересующим нас признаком, а отклонения частоты от вероятности приближенно следуют нормальному закону. При правильной организации отбора выборки мы должны обеспечить каждому объекту генеральной совокупности равную вероятность попадания в выборку.

На практике выбор из генеральной совокупности производится различными способами, в частности, выбор подразделяется на следующие разновидности:

1) *Выбор с возвратом* или повторением, или, иначе говоря, простой случайный выбор, о котором говорилось выше.

¹ Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. Учебник для вузов. - М.: ЮНИТИ, 1998. - С. 194.

2) *Выбор без возврата* или без повторения, когда каждый отобранный индивидуум перед выбором следующего индивидуума обратно в генеральную совокупность не возвращается.

Полученная выборка называется репрезентативной (представительной), если она достаточно хорошо представляет пропорции генеральной совокупности»¹.

Этот отрывок также сложно назвать определением, но, во всяком случае, в нём есть те свойства, которые присущи генеральной совокупности и выборке из неё. Относительно генеральной совокупности сущностные свойства это - совокупность всех *однородных* элементов. Относительно выборки – *случайность* выбора, *одинаковая вероятность* для каждого элемента быть выбранным, способность выборки *характеризовать свойства* генеральной совокупности.

Генеральная совокупность может быть конечной, а может быть и бесконечной. В первом случае множество однородных элементов, хотя и очень велико, но всё же конечно, как например, количество подберёзовиков в лесах Поволжья осенью текущего года. Во втором случае множество этих элементов не ограничено, поскольку новые наблюдения за ними дают нам всё новые и новые значения этой совокупности, например, количество подберёзовиков в Поволжье с момента их появления в этом регионе до настоящего времени. Поскольку «настоящее время», как верхняя временная граница этого множества меняется со временем, то и генеральная совокупность элементов является бесконечной.

Важно, что генеральная совокупность представляет собой именно совокупность однородных элементов – элементов одного рода. В этом понятии ключ к толкованию того, какую совокупность элементов можно отнести к генеральной, а какую - нельзя.

Чётких определений понятия «однородность» не встречается. Говорят об однородной функции, об упруго-однородном теле и т.п. Поэтому, отмечая сложность попытки дать однозначное толкование термину «однородная выборка», некоторые учёные вводят собственные определения, удобные для практического применения. Например: однородной называется такая совокупность, элементы которой формируются под воздействием общих основных причин и условий, а их законы распределения имеют простую структуру². Так, например, если динамика некоторого показателя X описывается моделью постоянного прироста со случайными отклонениями

$$X_t = X_{t-1} + a + e_t, \quad (1.4.1)$$

где a - постоянный прирост, e_t - не зависящие друг от друга случайные отклонения, имеющие нормальный закон распределения с нулевым математическим ожиданием и дисперсией, не зависящей или зависящей от t , то все элементы совокупности X_1, X_2, \dots, X_T имеют различные законы

¹ Дунин-Барковский И.В., Смирнов Н.В. Теория вероятностей и математическая статистика в технике. - М.: Гос. изд-во технико-теоретической литературы, 1955. – С. 179.

² Смоляк С.А., Титаренко Б.П.. Устойчивые методы оценивания (статистическая обработка неоднородных совокупностей). - М.: Статистика, 1980. – С. 14.

распределения, но сама совокупность, в соответствии с указанным определением, будет однородной.

Сложность использования такого «рабочего» определения однородности совокупности заключается в том, что понятия «общие причины и условия» недостаточно конкретны. Например, рассматривая динамику развития промышленности России на протяжении тридцати лет, можно говорить, что основной причиной формирования этого динамического ряда является спрос на товары промышленности, а условием - «дальнейшее повышение благосостояния народа»¹. На основе этого можно утверждать об однородности совокупности. Если же говорить о том, что сама причина - спрос на продукцию непрерывно меняется и качественно, и количественно, да и структура промышленного производства претерпела кардинальные изменения, то необходимо будет признать, что совокупность всё же является неоднородной. Принципиально важной характеристикой однородности процесса является её инвариантность времени или порядку наблюдения. Поэтому *совокупность элементов будет являться однородной только в том случае, когда ее элементы $Y_i, i=1,2,3,\dots, t,\dots,N$ формируются под воздействием общих основных неизменных причин и условий X_i так, чтобы при возникновении для $i=k$ условий и причин, равных $X_k = X_m$ элемент Y_k будет равен*

$$Y_k = Y_m + \varepsilon_k, \quad (1.4.2)$$

где ε_k - не зависящие друг от друга случайные отклонения, имеющие нормальный закон распределения с нулевым математическим ожиданием и конечной дисперсией.

Используя наше определение для случая (1.4.1), можно убедиться в том, что все элементы совокупности X_1, X_2, \dots однородны. Действительно, при достижении $X_t = X_\tau$, величина X будет определяться как

$$X_\tau = X_{\tau-1} + a + e_\tau. \quad (1.4.3)$$

Теперь можно дать определение понятия генеральной совокупности, которое будем использовать в дальнейшем в качестве основного: *генеральная совокупность – это множество всех возможных однородных элементов*. Из этого определения с учётом введённого понятия однородности со всей очевидностью следует, что элементы генеральной совокупности обладают характеристиками генеральной совокупности, но в силу воздействия на них множества случайных факторов эти характеристики в каждом элементе наблюдаются с ошибкой.

Теперь можно ответить на проверочный вопрос, сформулированный ранее: социально-экономические показатели за всю многовековую историю существования Руси, а затем и России представляют собой все мысленно возможные состояния; это – генеральная совокупность или нет? Если предложить, например, российской промышленности выпустить продукцию в количестве и качестве, соответствующим 1960-му году, то количество потребленной, например, при этом электроэнергии будет значительно меньше

¹ Светуныков С.Г., Литвинов А.А. Конкуренция и предпринимательские решения. – Ульяновск: Изд-во Корпорации технологий продвижения, 2000. – С. 186.

электропотребления, соответствовавшего 1960-му году, поскольку технологии промышленного производства существенно изменились. Если в 1960 году на промышленных предприятиях наряду с электрическим приводом машин встречались паровые, газовые и гидравлические приводы, то к настоящему времени имеются в подавляющем большинстве электрические приводы, да ещё и с программным управлением. Но, что ещё более важно, сегодня промышленность России просто не в состоянии выпустить ту номенклатуру изделий, которая была в 1960 году – разрушены оснастка, технологические линии, уничтожена проектная документация, потеряны навыки производства продукции, которая давно не выпускается. Если по каким-то причинам будет всё же необходимо в точности повторить эту номенклатуру и количество, то условия производства, затраты сырья, труда, денежных средств и т.п. будут существенно отличаться от того, что было в 1960 году.

Это значит, что социально-экономические показатели Руси и России в разные периоды времени отражали разные условия неоднородной динамики этой системы. Поэтому их ни в коем случае нельзя назвать генеральной совокупностью! Множество наблюдений социально-экономического развития России за многие века, пусть даже упорядоченное во времени, остаётся только множеством и никогда не превратится в генеральную совокупность.

Теперь следует определить понятие «выборочного множества». Понятно, что оно связано с генеральной совокупностью и выборочная совокупность представляет собой её часть. Обратимся вновь к определениям этого понятия.

«Если ограничиться числом наблюдений, которое в той или иной степени меньше общего числа элементов исследуемой совокупности, то будем говорить, что используется выборка... В широком смысле выборка – это любая часть всех элементов совокупности. Обычно эта часть извлекается таким образом, чтобы она «представляла» совокупность»¹.

В этом определении не совсем ясно, о какой совокупности идёт речь – о генеральной совокупности или вообще любом множестве каких-то элементов, представленном в виде некоторой совокупности. Но в этом определении есть одна важная характеристика, которая обязательно встречается при констатации основных характеристик выборочной совокупности – она должна отражать свойства генеральной совокупности. Если выборка не будет обладать таким свойством, то её исследование бессмысленно, а такую совокупность называть выборочной нельзя.

«Случайная выборка есть совокупность единиц (выборочная совокупность), отобранных из генеральной совокупности случайно, т.е. таким образом, что все единицы генеральной совокупности имеют равные шансы быть отобранными (попасть в выборку)»².

В этом определении указывается на наличие источника выборки (генеральная совокупность), способ получения выборки (случайная выборка),

¹ Джессен Р. Методы статистических обоснований. – М.: Финансы и статистика, 1985. – С. 21-22.

² Головач А.В., Ерина А.М., Трофимов В.П. Критерии математической статистики в экономических исследованиях. – М.: Статистика, 1973. – С. 5.

но нет указания на то, что выборка должна представлять основные характеристики генеральной совокупности.

Этот же недостаток содержится в определении, сформулированном в одном из последних учебных пособий, изданный в России: «Выборкой размера n из распределения F называется случайный вектор $\{\xi_1, \dots, \xi_n\}$, компоненты которого независимы и одинаково распределены с функцией распределения $F(x)$ »¹. Впрочем, необходимо отметить, что здесь появилась ещё одна существенная характеристика выборки, а именно – независимость элементов выборки.

«Выборка из данной генеральной совокупности – это результаты ограниченного ряда наблюдений x_1, x_2, \dots, x_n случайной величины ξ ... Если... ряд наблюдений x_1, x_2, \dots, x_n образует последовательность независимых и одинаково распределённых случайных величин, то выборка называется случайной»².

Итак, следует сделать такой вывод относительно выборочной совокупности. Она обладает следующими характеристиками, отличающими её от других подмножеств:

- 1) элементы выборочной совокупности формируются из генеральной совокупности;
- 2) формируются они случайным образом так, чтобы каждый элемент генеральной совокупности имел одинаковую вероятность попасть в выборку;
- 3) способ формирования выборки должен обеспечивать независимость элементов выборки друг от друга;
- 4) элементы выборки должны давать представление о свойствах генеральной совокупности.

Тогда, зная эти свойства случайной выборки, можно дать её определение, которое будем использовать в дальнейшем.

Выборочная совокупность – это случайно выбранные элементы генеральной совокупности, дающие представление о ней, и выбранные так, что каждый элемент генеральной совокупности имеет одинаковую вероятность попасть в выборку не зависимо от других элементов.

Теперь, определив понятия генеральной совокупности и выборочной совокупности, можно понять суть выборочного метода, который является основой математической статистики и эконометрии. Он является одним из основных методов реализации индуктивного вывода, когда по изученным свойствам нескольких элементов всего множества делается вывод о свойствах самого множества в целом. Следует напомнить, что метод индуктивного вывода пришёл на смену методу дедуктивного вывода, который более тысячи лет был единственным методом обобщения научных результатов и получения достоверных знаний. Именно применение индуктивного метода, как метода на порядок более дешёвого и требующего значительно меньшего времени,

¹ Лагутин М.В. Наглядная математическая статистика: учебное пособие. – М.: БИНОМ, Лаборатория знаний, 2009. – С. 42.

² Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. Учебник для вузов. - М.: ЮНИТИ, 1998. - С. 197.

послужило бурному росту науки с середины второго тысячелетия нашей эры. Выборочный метод является математическим обоснованием этого индуктивного вывода, даёт возможность математически корректно сформулировать этот вывод, и, что немаловажно, оценить вероятность правильности этого вывода.

Процесс выборочного метода заключается в следующем. Пусть имеется многочисленная совокупность однородных элементов, каждый из которых может обладать каким-либо признаком. Из этой совокупности выбирается наугад один из элементов, причем при выборе элемента из совокупности принимаются все меры к тому, чтобы вероятность быть выбранным была одинаковой для всех элементов. Затем таким же образом выбирается другой элемент и т.д. до тех пор, пока сформулированная выборка не будет удовлетворять некоторым заранее сформулированным требованиям исследователя (размер выборки, стоимость получения выборки и т.п.).

Полученные выборочные значения обрабатываются с помощью методов математической статистики на предмет выявления основных закономерностей генеральной совокупности и оценки их основных характеристик. Чаще всего это – выявление вида и формы некоторого закона изменения элементов генеральной совокупности, оценка параметров этого закона и достоверности полученных значений.

В том случае, когда имеющиеся в распоряжении прогнозиста статистические данные представляют выборочную совокупность, то они, очевидно, являются реализацией стационарного процесса и к ним в полном объёме может быть применен аппарат математической статистики. В том случае, когда имеющиеся данные не представляют выборку, то обрабатывать их с помощью методов математической статистики нет смысла – по этим значениям нельзя судить о процессе в целом, как это следует из выборочного метода. В этом случае прогнозист должен диагностировать тип процесса – эволюционный он или хаотический, а после этого выбрать инструмент обработки данных.

Как определить, что имеющийся в распоряжении прогнозиста ряд статистических значений о некотором социально-экономическом объекте является выборочным множеством, а когда – нет? Это сделать довольно просто. Надо только понять: безграничное расширение числа этих наблюдений позволит уточнить характеристики изучаемого объекта или наоборот – помешают? Если увеличение числа наблюдений действительно улучшит понимание изучаемого процесса и позволит построить более точные модели, то перед нами – выборочная совокупность и для её обработки необходимо использовать методы математической статистики.

Например, если прогнозист строит модель зависимости производительности труда швей-мотористок пятого разряда от уровня оплаты их труда, то для выявления такой зависимости и построения соответствующей модели новые дополнительные данные позволят улучшить результаты этой работы.

Если же ответ будет заключаться в том, что включение в имеющуюся совокупность новых элементов только ухудшит свойства модели, то, очевидно, что прогнозист не имеет дело с выборочной совокупностью. Например, если прогнозист собирается построить модель производственной функции Кобба-Дугласа для промышленности современной России и сделать с её помощью прогнозы о том, как будет меняться в ближайшие годы объёмы промышленного производства, то ему для этого нужны данные, в лучшем случае, с 1999 года по настоящее время. Включение в имеющуюся совокупность данных значения прошлых лет – 1990, 1980, 1960 и т.п., бессмысленно – тогда была другая социально-экономическая система и эти данные, которые расширяют статистическую базу, при их использовании для построения прогнозной модели производственной функции только ухудшат её характеристики.