

ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ - ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
ФАКУЛЬТЕТ ЭКОНОМИКИ
КАФЕДРА СТАТИСТИКИ
2005-2006 учебный год

ЭКОНОМИЧЕСКАЯ И СОЦИАЛЬНАЯ СТАТИСТИКА

Тема 1. Теория статистического наблюдения

Лекция 3. Этапы статистического исследования: представление полученного статистического материала

После завершения сводки и группировки собранного статистического материала переходят к этапу визуализации, или представления данных.

Представление собранного статистического материала может осуществляться в различном виде. Наиболее часто в практических исследованиях используются:

- неупорядоченный ряд наблюдений – непосредственный результат проведённых измерений,
- упорядоченный ряд наблюдений,
- ряд распределения,
- кумулятивный (накопленный) ряд распределения,
- таблица распределения;
- другие таблицы;
- графики;
- числовые характеристики – обобщающие статистические показатели.

Неупорядоченный ряд наблюдений – это матрица размерности $m \times n$, которая представляет собой прямой результат фиксирования значений набора m признаков для каждой из n единиц наблюдения на основании информации, предоставляемой статистическими единицами на этапе производства данных. Иными словами, неупорядоченный ряд наблюдений – это произвольный список результатов статистического измерения.

Размещение единиц наблюдения в порядке возрастания или убывания значений определённого варьирующего признака приводит к построению отсортированного, или упорядоченного списка. Следует отметить, что упорядочивание статистического ряда возможно при измерении признака в шкале не слабее порядковой.

Сортировка производится для облегчения сводки и группировки данных, то есть подсчёта количества единиц совокупности, распределённых по группам относительно значений того или иного варьирующего признака. Подсчёт единиц по выделенным группам приводит к построению так называемых рядов распределения. Для расчета обобщенных характеристик вариационных рядов можно пользоваться как частотами, так и частостями.

Ряды распределения — это последовательности пар значений признака и чисел, характеризующих, как распределяются единицы некоторой совокупности по тому или иному атрибутивному или количественному признаку, то есть частоты или частости распределения.

Ряды, построенные по атрибутивному признаку, называют *атрибутивными рядами распределения*. Элементами атрибутивного ряда распределения являются варианты атрибутивного признака и численности единиц, относящихся к каждой группе частоты или частости атрибутивного ряда. Ряды распределения, построенные по количественному признаку, называются *вариационными рядами*.

Вопросы формирования пространственных рядов распределения были рассмотрены в связи с изучением метода группировок.

Особое место среди различных видов статистических рядов занимают **динамические ряды**.

Ряд динамики представляет собой числовые значения определенного статистического показателя в последовательные моменты или периоды времени. Числовые значения того или иного статистического показателя, составляющие динамический ряд, называются *уровнями* ряда.

Каждый динамический ряд состоит из n -го числа варьирующих во времени показателей. Обычно первый член ряда называют начальным уровнем (y_0 или y_1), а последний — конечным (y_n). Каждый уровень ряда соответствует определённому моменту или периоду времени t . При проведении процедур первичной и вторичной обработки динамических рядов следует всегда помнить о свойствах шкалы разностей, в которой измеряется признак времени.

Закономерность в изменении уровней ряда в одних случаях проявляется довольно наглядно, в других она может затушевываться колебаниями, вызываемыми случайными или другими причинами. Во всех случаях одной из первых задач исследования является выявление основной тенденции в изменении уровней, именуемой трендом.

Таким образом, ряды динамики тех или иных изучаемых показателей могут отражать различные процессы изменения характеристик единиц наблюдения. Уровни

любого ряда являются результатом взаимодействия самых различных факторов, одни из которых могут действовать длительно, другие— кратковременно, одни являются главными, определяющими тенденцию изменений, а другие — случайными затушевывающими ее и т. д.

Поэтому, чтобы сделать правильные выводы о закономерностях развития того или иного показателя, надо суметь отделить главную тенденцию изменения от колебаний, вызванных влиянием случайных кратковременных причин, для чего каждый динамический ряд следует подвергнуть тщательному анализу и в случае необходимости математической обработке.

В зависимости от вида показателей уровней ряда и ряды динамики обычно подразделяют на ряды *абсолютных (измеренных в абсолютной шкале результатов наблюдения), относительных и средних* показателей. Как будет показано позже, при этом ряды динамики абсолютных величин должны рассматриваться как исходные, а ряды относительных и средних величин - как производные. Динамические ряды относительных и средних величин могут быть построены только на основе рядов абсолютных величин.

Кроме того, уровни (показатели) рядов динамики могут относиться к определенным моментам или интервалам времени. В зависимости от этого в статистике различают: **моментные ряды и интервальные ряды.** **Моментным** называется ряд, уровни которого формируются показателями типа запаса и характеризуют величину явления по состоянию **на определенные моменты** времени. **Интервальным** называется ряд, уровни которого формируют показатели типа потока и характеризуют величину изучаемого явления **за определенный период** времени. Отличительной особенностью интервальных рядов абсолютных величин является то, что уровни их можно дробить и складывать. Так, зная выпуск продукции по месяцам, можно сложить эти показатели, и полученная сумма будет характеризовать выпуск продукции за год. Если же каждый месячный уровень разделить на продолжительность месяца, то новые уровни дадут нам примерное представление о среднесуточном выпуске продукции по месяцам. Подобные действия с уровнями моментного ряда лишены смысла.

Возможность суммирования уровней интервальных рядов абсолютных величин позволяет строить динамические ряды с нарастающими итогами.

При изучении социально-экономических процессов и явлений в статистике приходится иметь дело с различными видами динамических рядов. Однако, с какими бы рядами динамики ни приходилось работать, основным требованием, предъявляемым к анализируемым рядам, является их динамическая однородность, то есть **сопоставимость**

уровней. **Несопоставимость** уровней может возникнуть по различным **причинам**, в числе основных из которых можно назвать:

- 1) изменение территории, к которой отнесены те или иные показатели;
- 2) изменение даты учета (например, если учет основных средств на предприятии в течение ряда лет проводился по состоянию на 1 октября, а затем — на 1 января, то соединение в один ряд показателей стоимости основных средств за ряд лет с разной датой даст несопоставимые уровни);
- 3) изменение методологии учета или расчета показателей (например, если в одни годы среднюю производительность труда считали по товарной продукции, а в другие — по валовой, то такие уровни будут несопоставимы, или же, если в одни годы производительность труда рассчитывалась на одного рабочего, а в другие — на одного работника промышленно-производственного персонала, то соединить такие данные в один ряд нельзя — они также несопоставимы);
- 4) изменение цен (для стоимостных показателей);
- 5) различная продолжительность периодов, к которым относятся уровни.

Могут быть и другие причины несопоставимости. При этом, говоря об изменении территории, к которой относятся показатели за разное время, следует иметь в виду, что **вопрос о сопоставимости может решаться по-разному в зависимости от целей исследования**. Например, с изменением границ какой-то области часть населения и часть предприятий отошли к другой области. В этом случае, если ставится задача изучения изменения численности населения (или предприятий, или объема промышленного производства) в связи с изменением границ области, не только можно, но и должно сопоставлять данные в разных границах. Если же ставится задача изучения темпов естественного прироста населения или темпов развития промышленности, то сравниваемые показатели должны относиться к одним и тем же территориальным границам.

Следовательно, прежде чем анализировать динамический ряд надо, исходя из цели исследования, убедиться в сопоставимости уровней ряда и, если сопоставимость отсутствует, добиться ее дополнительными расчетами. Решение вопроса о сопоставимости уровней особенно важно при так называемом смыкании рядов. Под **смыканием** рядов динамики понимают объединение в один ряд (более длинный) двух или нескольких рядов, уровни которых исчислены по разной методологии или в разных границах. При этом для осуществления смыкания необходимо, чтобы для одного из периодов (переходного) имелись данные, исчисленные по разной методологии (или в разных границах). Для этого находим соотношение между несопоставимыми показателями переходного периода.

Умножая на полученный коэффициент, приводим изучаемые ряды динамики в сопоставимый вид.

Результаты обработки статистических рядов (как пространственных, территориальных, так и динамических), как правило, оформляются в виде статистических таблиц. Табличная форма позволяет изложить материал наиболее удобно, компактно, наглядно и рационально. В каждой таблице имеется ряд горизонтальных строк и вертикальных граф (столбцов), имеющих заголовки, объясняющие, что характеризуется и какими показателями. Строки и графы при пересечении образуют клетки, которые заполняются статистическими данными.

В статистических таблицах различают *подлежащее* и *сказуемое*. *Подлежащим* в таблице являются те объекты или их части, а также отдельные периоды времени и территории, которые характеризуются рядом показателей. *Сказуемым* в таблице являются все те показатели, которыми характеризуется подлежащее. Статистические таблицы могут быть *простыми* и *сложными*. К *простым*, относят перечневые таблицы, в которых подлежащим служит перечень отдельных объектов (домашних хозяйств, министерств, районов, политических партий и т. п.), и динамические таблицы, где подлежащим являются отдельные годы, месяцы или иные периоды времени. У *сложных* таблиц подлежащее представляет собой совокупность, расчлененную на группы по одному или нескольким признакам. Таблицы, в подлежащем которых имеет место группировка. По одному признаку, именуют *групповыми*. При наличии в подлежащем группировки по двум и более признакам таблица именуется *комбинационной*. К числу сложных таблиц относят и корреляционные, и балансовые таблицы, которые мы подробно будем рассматривать позже. Деление таблиц на простые, групповые и комбинационные основано на степени расчленения подлежащего. Однако и сказуемое в таблицах может быть разработано по-разному. Если все показатели сказуемого характеризуют подлежащее отдельно, независимо друг от друга, то такая разработка сказуемого называется простой. Если же в сказуемом один признак комбинируется с другим, то такая разработка сказуемого называется сложной. Составляя таблицы, следует помнить, что таблица должна компактно и наглядно изложить цифровой материал. И если при излишней детализации подлежащего и сказуемого чтение таблицы затрудняется, то лучше вместо одной громоздкой таблицы строить несколько таблиц более простых.

Рассмотрим макет таблицы со сложной разработкой сказуемого, предусмотренный для разработки материалов Всероссийской переписи населения 2002 г.

Таблица Распределение населения по полу, возрасту и состоянию в браке

	Мужчины					Женщины				
	Всего	в том числе				Всего	в том числе			
		состоящие в браке	не состоящие в браке	вдовые	разведенные		не указавшие состояния в браке	состоящие в браке	не состоящие в браке	вдовые
Все население в том числе										
до 15 лет										
15—19 »										
20—24 »										
25—29 »										
и т. д.										
65—69 »										
70 лет и старше										

Статистические таблицы самых различных форм и видов прочно вошли в практику обработки и выдачи результатов статистических исследований, сводки данных. Впервые статистические таблицы (простые и групповые) были применены при изложении статистических данных в 1727 г. в России И. К. Кириловым в работе «Цветущее состояние Всероссийского государства». Применение комбинационных таблиц относится к более позднему периоду (1882). Приоритет в построении комбинационных таблиц и использование их в анализе принадлежат русским земским статистикам.

В заключение отметим несколько **технических моментов**, связанных с построением таблиц.

При построении таблиц необходимо обращать внимание на **четкость заголовков** в отдельных строках и графах, не говоря уже о заголовке самой таблицы. Следует всегда указывать, в каких единицах даются показатели в отдельных графах. Повторяющиеся термины надо выносить в общие заголовки. Графы и строки полезно нумеровать. В групповых и комбинационных таблицах всегда надо давать итоговые графы и строки.

Округление чисел во всех графах и строках следует проводить с **одинаковой точностью** (или до целого знака, или до десятого, или до сотого и т. д.).

При заполнении таблицы числовым материалом отдельные клетки могут оставаться незаполненными. При чтении такой таблицы важно знать причину отсутствия тех или иных данных. Чтобы была ясна причина незаполненности той или иной клетки, установлены определенные **условности объяснения незаполненных клеток таблицы**. Так, если клетка вообще **не подлежит заполнению**, то обычно в ней ставится крест «X». Если **отсутствуют сведения**, то ставятся три точки (...) или записывается «нет св.» — нет сведений. Если **отсутствует само явление**, то ставится **прочерк** (—). Если в

таблице стоит число 0, 0, это значит, что при округлении с большей точностью, например до сотых, здесь может появиться какая-нибудь значащая цифра.

Правильно и грамотно составленные таблицы: являются важным средством изложения обработанного статистического материала и основой проведения анализа.

Кроме таблиц для наглядности и компактности статистические данные могут быть изображены графически, т.е. представлены с помощью различных условных знаков - символов (точек, линий, фигур и т.п.).

Графики могут быть использованы для изображения динамики явления; сравнения показателей, относящихся к одному времени, но разным объектам, определения состава (структуры) совокупности, выявления зависимости одних показателей от других; определения степени распространения какого-либо явления, контроля над выполнением бизнес-планов предприятия и др.

Другими словами, графики помогают выявлять закономерности развития, соотношений, взаимосвязи, пространственного распределения и пр. Как видно, задачи, решаемые при помощи графиков, различны. Вместе с тем для решения определенной задачи, например сравнения или динамики, могут использоваться разные графики, и, наоборот, один и тот же график может быть использован для решения разных задач.

Основными элементами графика являются:

- 1) графический образ, т.е. знаки-символы (линии, фигуры и пр.), с помощью которых изображаются статистические величины,
- 2) поле графика—место, где размещены те или иные графические образы;
- 3) пространственные ориентиры, определяющие расположение графических образов на поле;
- 4) масштабные ориентиры, дающие количественную определенность знакам-символам,
- 5) экспликация (словесное пояснение) графика, включающая точное его название и пояснения к отдельным его частям.

Среди графиков различают: *диаграммы* — графическое изображение статистических величин с помощью различных геометрических фигур или знаков, *картограммы* — изображение величины того или иного показателя на географической карте с помощью графических символов (штриховки, расцветки, точек) и *картодиаграммы* — сочетание картограммы с диаграммой, т.е. диаграммы на географической карте.

В зависимости от применяемых графических образов, среди диаграмм различают столбиковые, плоскостные, объемные, линейные и др.

Пространственные ориентиры для большинства диаграмм обычно задаются в виде системы прямоугольных (декартовых) координат или координатных сеток. Для картограмм и картодиаграмм пространственными ориентирами являются географические ориентиры (границы или контуры различных районов, рек и т. п.).

Количественная определенность графических образов задается масштабными шкалами. Масштабной шкалой называется линия, на которую нанесены точки, соответствующие определенным числовым значениям показателя. Отрезок шкалы, соответствующий определенной числовой единице, называется масштабом графика. Другими словами, масштаб графика является условной мерой перевода числовых значений показателя в графические.

Масштабные шкалы могут быть прямолинейными и криволинейными (например, круговые).

Если равным отрезкам на шкале соответствуют равные числовые интервалы, то шкала называется равномерной. Если равным отрезкам на шкале соответствуют неравные числовые интервалы, то шкала называется неравномерной. Из неравномерных наиболее часто применяется логарифмическая шкала. Когда для построения графика нужно использовать два масштаба для двух показателей, то применяют систему прямоугольных координат или координатные сетки. Примером условных координатных сеток может быть график Лоренца.

Масштабные шкалы могут быть непрерывными и прерывными. Последние используются для изображения статистических величин, имеющих значения лишь с определенного уровня. Например, если значения показателя варьируют в пределах от 140 до 160, то нет необходимости строить шкалу со всеми делениями от 0 до 160. Приняв, например, условную величину отрезка за 5 единиц, можно от 5 до 140 шкалу прервать.

Как уже указывалось, каждый график должен иметь экспликацию, т. е. словесные пояснения к отдельным элементам графика, четкое и ясное название, содержащее сведения о характере изображаемых показателей, месте и времени, к которым они относятся.

К числу наиболее простых графиков относятся столбиковые диаграммы. Они представляют собой ряд прямоугольников (столбиков) с одинаковым основанием, высота которых пропорциональна числовым значениям изображаемых показателей, т. е. для всех столбиков принимается один и тот же масштаб. Все столбики должны строиться на одной базовой линии. О соотношении между величинами изображаемых показателей судят по высоте столбиков, поэтому разрыв шкалы для них не допускается. Столбиковые диаграммы обычно используются для сравнения данных, относящихся к разным объектам,

или для отражения динамики. Если данные относятся к ряду объектов, то рекомендуется располагать столбики от большего к меньшему или наоборот. Прямоугольники можно располагать и горизонтально — тогда диаграмма называется *ленточной*.

Разновидностью ленточных диаграмм являются иллюстрированные, или «фигурные», диаграммы. В этом случае «ленты» строятся из фигур-знаков, соответствующих изображаемому числовым показателям

К числу *плоскостных* относят диаграммы, изображающие величину показателя площадью различных фигур. Наиболее часто из плоскостных диаграмм применяют *квадратные* и *круговые*.

Для построения и квадратной и круговой диаграммы необходимо, прежде всего, извлечь корень квадратный из каждой изображаемой числовой величины, а затем по полученным данным в принятом масштабе строить квадрат или круг.

Разновидностью плоскостных диаграмм являются и картинные (фигурные) диаграммы, в которых величина изображаемого показателя пропорциональна площади фигуры, изображающей данное явление. Например, представление о производстве автомашин (или другой продукции) в разных странах или в одной стране за ряд лет можно дать с помощью изображения автомашин разного размера, площадь которых пропорциональна изображаемому числовым показателям. Конечно, такие диаграммы особенно нуждаются в числовых подписях, так как сопоставлять площади таких фигур «на глаз» трудно.

К числу плоскостных диаграмм относятся и так называемые *знаки Варзара* — прямоугольники, предложенные русским статистиком К.Е.Варзаром для графического изображения трех взаимосвязанных показателей, из которых один является результатом взаимодействия двух других. Знак Варзара — это своего рода комбинация линейной и плоскостной диаграммы. В прямоугольниках Варзара основание отражает величину одного признака (a), высота—величину второго признака (b), площадь—величину третьего показателя (c), являющегося результатом произведения двух первых ($c=ab$). Знаки Варзара особенно удобны для выявления роли отдельных факторов в изображаемых сложных явлениях.

Особый вид плоскостных (круговых) диаграмм представляют собой *секторные диаграммы*, используемые для изображения структуры (состава) совокупности. Для построения секторной диаграммы вычерчивается круг произвольного радиуса, принимаемый за всю совокупность и приравненный 100%. Секторы круга либо по-разному заштриховывают, либо окрашивают в разные цвета.

Если нужно показать не только структуру, но и различие в величине изображаемых совокупностей, то секторные диаграммы сочетают с круговыми — плоскостными, т. е. строят круги различных радиусов и, принимая каждый круг за 100%, делят его в соответствии с удельным весом отдельных групп в совокупности.

Для изображения структуры совокупности кроме секторных диаграмм можно воспользоваться и столбиковыми. В этом случае каждый столбик (прямоугольник) делят на части в соответствии с долей отдельных групп в совокупности.

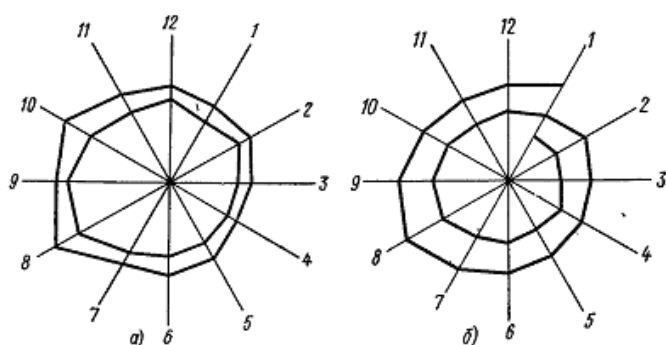
Аналогично тому, как в плоскостных диаграммах величина изображаемого показателя отражается площадью тех или иных фигур, в *объемных* диаграммах величина показателя изображается объемом того или иного геометрического тела. Чаще других в качестве объемных диаграмм используются кубы. Для изображения статистических данных с помощью кубов из числовых показателей извлекаются кубические корни и по найденным значениям, принятым за высоту ребра, в определенном масштабе строятся кубы.

Если линейный график используется для изображения двух взаимосвязанных показателей, то на горизонтальной оси строится шкала одного признака — факторного (x), а на вертикальной оси — второго — результативного (y).

Для характеристики изменения явлений во времени, выявления зависимости между двумя показателями и некоторых других задач удобно пользоваться *линейными графиками*, строящимися при помощи прямоугольной системы координат. При этом на горизонтальной оси строится шкала времени, а на вертикальной — шкала уровней изображаемого показателя. Соединив концы ординат, получаем ломаную линию, которая и является линейным графиком. Строя линейный график, характеризующий динамику явления, следует помнить, что на горизонтальной оси равным периодам времени должны соответствовать равные отрезки. Градуируя вертикальную масштабную шкалу, следует помнить, что нельзя в начале координат ставить какое-либо числовое значение результативного- показателя кроме нуля. Если до определенного значения нет показателей, то следует вертикальную шкалу сделать прерывной. Линейные графики удобны тем, что на одном графике можно строить несколько кривых (ломаных). Так, например, на одном графике можно изобразить динамику определенного показателя в разных странах или же различных показателей в одной стране. Изображая графически динамику нескольких разноименных показателей, необходимо вертикальную шкалу строить не по абсолютным, а по относительным величинам (в процентах).

Для графического изображения некоторых показателей в качестве шкал удобнее пользоваться не осями прямоугольной системы координат, а радиусами круга. В

частности, при графическом изображении помесечных данных за год (или несколько лет) можно воспользоваться 12 равноудаленными друг от друга радиусами концентрических окружностей. «Закрепив» за каждым радиусом определенный месяц (январь, февраль, март и т. д.), откладывают на них в принятом масштабе отрезки, пропорциональные изображаемым числовым показателям за эти месяцы, а затем концы отрезков соединяют последовательно. Причем, если соединением концов отрезков замыкается круг одного года, то такая радиальная диаграмма называется *замкнутой*. Если же при изображении помесечных данных за ряд лет концы отрезков все время» последовательно соединяются, не замыкаясь, то диаграмма называется *спиральной*.



Радиальные диаграммы: а) замкнутая; б) спиральная

Радиальные диаграммы являются разновидностью линейных диаграмм. На рисунке показаны радиальные диаграммы: а) замкнутая и б) спиральная.

Линейные графики могут использоваться, например, для наблюдения и анализа хода выполнения плана. Для этой цели на одном графике строят две линии: плановую и фактическую. Сравнение линий планового и фактического оборота дает возможность наглядно видеть степень выполнения плана за каждый конкретный отрезок времени. Линия планового задания может иметь при этом различный характер в зависимости от того, предполагается равномерный или неравномерный выпуск за все отрезки времени. Среди линейных графиков, используемых для текущего наблюдения за выполнением плана на разных объектах или по разным продуктам, особое место занимают так называемые контрольно-плановые графики, отражающие выполнение плана за каждый день с начала того или иного периода. Строится такой график в виде ленточной диаграммы на специальных таблицах, у которых строки отводятся для наименования отдельных объектов или видов продукции, а графы — для дней недели. Длина графы каждого дня, независимо от того, планируется равномерный или неравномерный выпуск продукции, принимается за 100%. Соответственно каждая пятая часть длины графы приравнивается 20%, каждая десятая часть — 10% и т. д.

Процент выполнения плана за каждый день отмечают на графике тонкой линией, а толстой итоговой линией отмечается выполнение плана с начала периода.

Среди различных видов графиков особое место занимает **график Лоренца**, поскольку он может быть использован для характеристики уровня относительной концентрации тех или иных явлений, для изучения порайонной специализации или локализации и т. п.

При построении графика Лоренца на оси абсцисс откладываются значения накопленных частостей выделенных групп (в форме простого кратного отношения или в процентах), а на оси ординат – значения накопленных долей, рассчитанных на основе вклада значений признака группе в его общий объём (также в форме простого кратного отношения или в процентах). Интерпретация этого графика может быть сведена к следующему. При равномерном распределении явления между единицами изучаемой совокупности должно соблюдаться равенство $x = y$ (10% единиц обеспечивают 10% объёма признака, 20% единиц обеспечивают 20% объёма признака и т. д.). На графике эта зависимость выразится прямой, проходящей через начало координат под углом в 45° , т. е. это будет диагональ квадрата, на котором строится график Лоренца. Следовательно, диагональ квадрата, соединяющая нижний левый угол с верхним правым, может рассматриваться как *линия равномерного распределения*. Всякое отклонение от нее — признак неравномерности распределения. И **чем больше эмпирическая ломаная графика Лоренца отклоняется от диагонали квадрата, тем, следовательно, больше неравномерность распределения и выше уровень относительной концентрации объёма изучаемого явления у отдельных единиц.**

Построение линии Лоренца во всех случаях требует обязательной записи исходных данных в определенном порядке. Предположим, в одном из районов имеются предприятия разных отраслей, и мы располагаем сведениями о числе занятых в каждой из этих отраслей. График Лоренца в данном случае может отразить равномерность или неравномерность распределения занятых по отдельным отраслям. Для его построения необходимо переписать все отрасли по рангу, начиная с отрасли, имеющей наибольшую занятость, и рассчитать кумулятивные итоги для всех записанных по рангу отраслей. А далее строится график по кумулятивным итогам. Пользуясь графиком Лоренца и зная наименования отраслей, записанных по горизонтальной линии квадрата в порядке возрастания в них числа занятых, можно делать вывод о равномерном или неравномерном распределении занятых по отраслям, а, следовательно, и о специализации района. Вместо предприятий разных отраслей одного района можно воспользоваться для построения кривой Лоренца данными о числе занятых или выпуске продукции на предприятиях одной

отрасли, но размещенной в разных районах (1, 2, 3 и т. д.) Построенная по таким данным кривая Лоренца будет характеризовать уровень локализации данной отрасли.

Картограмма—это изображение объёма интересующего нас признака на географической карте с помощью графических символов: штриховки, расцветки, точек. В зависимости от используемых символов различают *фоновые* и *точечные* картограммы.

Для построения фоновой картограммы необходимо произвести группировку изображаемого географического ряда, выделив несколько характерных групп, отличающихся по величине изображаемого признака. Для каждой группы принимается условный символ для изображения либо окраска разных тонов, либо различная штриховка. Для построения картограммы используется карта с очертаниями только границ районов (областей, хозяйств и т. п.), которые и окрашиваются или штрихуются в соответствии с уровнем изучаемого показателя в каждом районе и принятом для его изображения символом.

Если после заштриховки или окраски соответствующих участков карты обнаруживается определенная закономерность в географическом расположении территории с одинаковой величиной изображаемого показателя, то можно говорить о зависимости данного показателя от географического фактора. Если же районы с одинаковой окраской или штриховкой разбросаны в беспорядке на карте, то, очевидно, нет определенной закономерности в пространственном размещении данного показателя, т.е. его распространение или уровень не связаны с географическим положением района.

Кроме различной окраски и штриховки при построении картограмм используют и так называемый точечный метод. Суть этого метода сводится к тому, что определенная величина изображаемого признака принимается за одну точку, и тогда в зависимости от общего размера показателя на той или иной территории ставится столько точек, сколько соответствует ему в принятом масштабе. Например, если при изображении размещения признака на той или иной территории одну точку (•) принять за 1000 объектов, то пять точек будут соответствовать пяти тысячам объектов и т. д. Точки по территориям распределяются либо равномерно, либо более концентрированно к местам сосредоточения изображаемого показателя в действительности.

Фоновые картограммы используются, как правило, для изображения средних и относительных показателей (плотность населения, удельный вес продукции животноводства или растениеводства в общем объеме сельскохозяйственной продукции, средняя урожайность зерновых и т. п.). Точечные же картограммы используются для объемных (количественных) показателей (численность населения, поголовье скота и т. д.).

Картодиаграмма представляет собой сочетание картограммы с диаграммой. При построении картодиаграммы на схематической карте на участках отдельных районов (или областей) строятся различные диаграммы: столбиковые, круговые, квадратные, секторные и др., которые и характеризуют (изображают) величину (или структуру) того или иного показателя на данной территории.

При графическом изображении **динамического ряда** на оси абсцисс (значения признака-фактора) строится шкала времени, а на оси ординат — шкала уровней признака, варьирующего в динамическом ряду (арифметическая или, иногда, логарифмическая).

Вариационный ряд можно изобразить, как и любой ряд значений аргумента и функции, используя прямоугольную систему координат и строя точки с координатой (x_1, f_1) ; (x_2, f_2) ; (...); (x_n, f_n) . Для дискретного ряда, если затем соединить последовательно полученные точки отрезками прямой, а из первой и последней опустить перпендикуляры на ось X , получим фигуру, которая называется **полигоном** распределения и графически представляет распределение совокупности по признаку x .

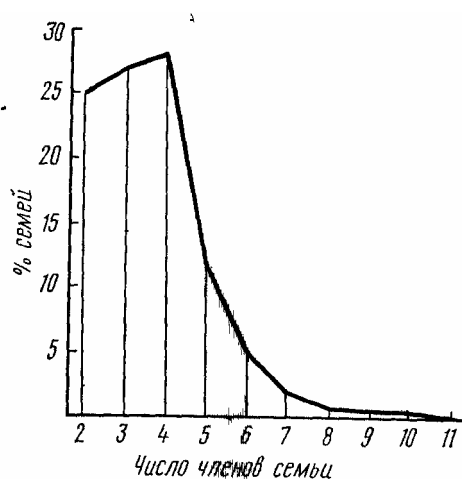


Рис 1

На рис. 1 представлен полигон распределения семей по количеству членов семьи.

Графическое изображение интервального вариационного ряда строится в виде гистограммы. При построении гистограммы **на оси X откладывают границы интервалов** значений признака. Затем, используя отрезки, представляющие интервалы, в качестве основания, для интервального ряда с **равными интервалами** на оси ординат откладывают частоты или частости, то есть строят прямоугольники с высотой, равной **частоте** (или частости) заданного интервала. На рис: 2 представлена гистограмма распределения тракторов по времени службы в годах. На рис. 2 величину последнего открытого интервала при построении гистограммы мы приняли равной величине предшествующего интервала (в данном случае всех предшествующих). Такой способ

определения границ незакрытых интервалов применяется, когда действительная верхняя граница последнего интервала неизвестна даже предположительно. Первый интервал, если он не закрыт, определяется тем же способом: его считают равным, следующему за ним (или его началом принимают «0»).

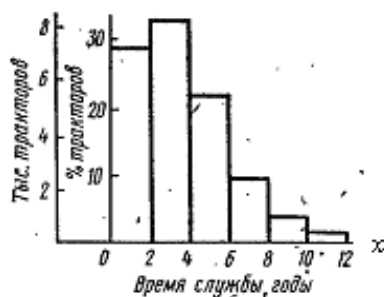


Рис. 2

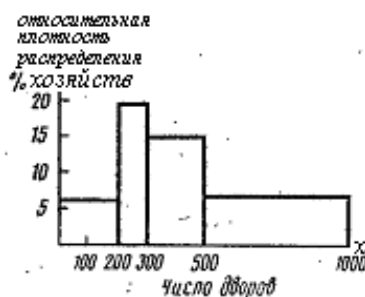


Рис. 3

Для интервальных рядов с неравными интервалами строят гистограмму плотностей распределения, так как в ряду с неравными интервалами именно плотность дает адекватное представление о заполненности каждого интервала.

На рис. 3 изображена гистограмма распределения хозяйств по числу дворов на основе относительной плотности распределения. Эта гистограмма дает ясное представление и о величине интервалов (верхняя граница — 1000 дворов), и о заполненности каждого из них. Следует обратить внимание на то, что площадь отдельных прямоугольников гистограммы пропорциональна частотам или частостям соответствующих вариантов. В самом деле, если вспомнить, что плотность распределения равна отношению частоты (частости) к длине интервала, а площадь прямоугольников гистограммы — это произведение плотности на величину интервала, то площадь прямоугольника равна произведению плотности в интервале на его длину, то есть частоте. Площадь всей гистограммы, таким образом, численно равна сумме частот или численности единиц, в совокупности.

Любой вариационный ряд можно представить графически также и на основе накопленных частот (или частостей) как функции признака. При этом на оси абсцисс откладывают варианты значений признака (границы интервалов), а на оси ординат — соответствующие накопленные частоты или частости. Полученные точки соединяют отрезками и получают график, который называется кумулятой или кумулятивной кривой.

Существует определенная связь между кумулятой и линией плотности распределения. При непрерывной вариации признака (в бесконечно большой

совокупности) плотность распределения является производной от накопленной частоты или частоты.

Пусть имеем интервал от x до $x+\Delta x$. Накопленная частота $F(x)$ при переходе от начала интервала к его концу возрастает на величину частоты в нем признака. Разделив ее на величину интервала, получим среднюю плотность в нем:

$$\bar{f} = \frac{F(x + \Delta x) - F(x)}{\Delta x} = \frac{\Delta F(x)}{\Delta x}.$$

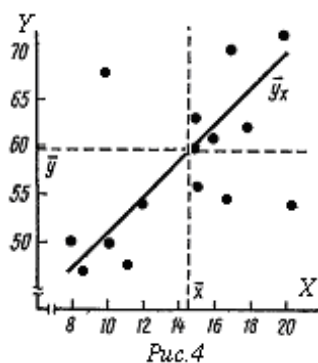
Чтобы от средней плотности в интервале перейти к плотности в точке, надо представить, что Δx — бесконечно мало и перейти к пределу:

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{\Delta F(x)}{\Delta x} = F'(x),$$

т. е. производной от $F(x)$, если она для данного значения x существует.

Таким образом, плотность распределения является производной накопленной частоты. А график накопленных частот — кумулята — является графиком интегральной функции для плотности распределения.

Визуализация статистической информации играет важную роль и при исследовании взаимосвязей между процессами и явлениями. Так, при аналитической группировке единиц наблюдения на основе вариантов значений пары взаимосвязанных признаков, используя данные о комбинации индивидуальных значений фактора и результата, можно построить в прямоугольных координатах точечный график, который называется полем корреляции. Он представляет собой множество точек с координатами, зафиксированными на уровне вариантов признаков (факторного X и результирующего Y), положенных в основу группировки. Если на полученном графике провести 2 оси, соответствующие средним значениям факторного X и результирующего Y признака, то вся плоскость графика окажется поделённой на 4 квадранта.



Равномерность распределения точек по всем квадрантам плоскости графика является основанием для проверки гипотезы об отсутствии статистически значимой связи между факторным X и результирующим Y признаками. И наоборот, концентрация точек в

тех или иных квадрантах позволяет предположить наличие значимого влияния положенного в основу группировки факторного признака на результат (см. рис.4).

Перечисленные графики не исчерпывают все возможности визуализации статистической информации, но являются наиболее аналитически важными и часто употребляемыми.

Визуальный анализ является обязательным подготовительным этапом в процессе решения задач статистического исследования. Он даёт дополнительные возможности для уточнения первоначальных гипотез, для логического и содержательного контроля результатов первичной обработки данных и, следовательно, подготавливает материал для вторичной обработки имеющегося статистического материала, то есть для оценки обобщающих статистических показателей статистическими методами.

Статистический показатель - основная категория статистики. ***Статистический показатель представляет собой результат обобщения, агрегирования информации об индивидуальных значениях признаков отдельных элементов совокупности.*** В отличие от признаков показатель относится не к отдельным элементам, а к совокупности в целом. В самом общем смысле можно сказать, что ***статистический показатель представляет собой некоторую функцию значений индивидуальных признаков, то есть результат вторичной обработки собранного статистического материала. Таким образом, статистический показатель описывает обнаруженную в результате исследования статистическую закономерность.***