

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

ЭКОНОМИЧЕСКАЯ И СОЦИАЛЬНАЯ СТАТИСТИКА

Тема 1. Теория статистического наблюдения

Лекция 2. Этапы статистического исследования: первичная обработка результатов статистического наблюдения

После получения в процессе статистического наблюдения матрицы исходной статистической информации переходят к следующему этапу статистического исследования – первичной обработке полученных данных. В первую очередь, осуществляется **логический и содержательный контроль**.

Как бы тщательно не были продуманы план, статистического наблюдения и программа наблюдения и как бы точно не руководствовались всеми указаниями инструкции лица, осуществляющие сбор сведений, при любом статистическом наблюдении могут возникнуть **ошибки** (погрешности). Эти ошибки наблюдения могут возникнуть по разным причинам: за счет описок, оговорок, округлений, неправильного заполнения формуляра, забывчивости тех, кто отвечает, или их стремления скрыть или исказить факты; при непосредственном наблюдении (например, при пересчёте избирательных формуляров) ошибки могут возникать из-за усталости счётчиков или ограниченности времени работы и т. п.

Все ошибки можно разделить на *преднамеренные* и *непреднамеренные*. Непреднамеренные ошибки, в свою очередь, могут носить *случайный* или *систематический* характер.

Случайные ошибки (погрешности) наблюдения, возникающие и по вине отвечающего, и по вине регистраторов, в результате описок, оговорок, незнания и т. п., не столь опасны для результатов наблюдения, как ошибки систематические. Случайные ошибки одинаково часто могут встретиться и в сторону преуменьшения, и в сторону преувеличения, а при большом числе наблюдений они взаимно погашаются, нейтрализуются.

Непреднамеренные систематические ошибки (погрешности) возникают главным образом при опросе, за счет округлений количественных показателей (округление

возраста, стажа работы, дохода и т. п.) или за счет неточностей измерительных приборов (при непосредственном наблюдении).

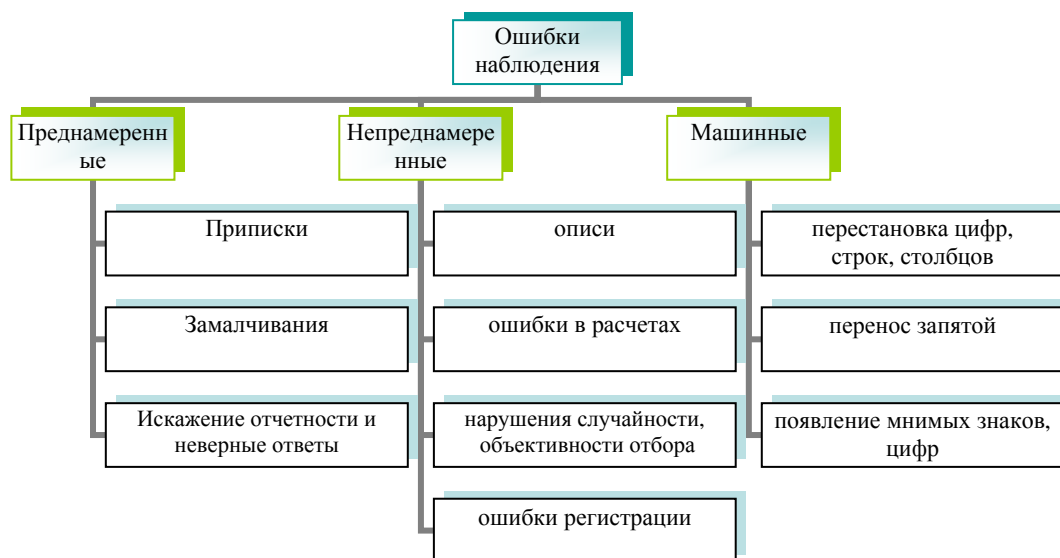


Схема 1. Ошибки статистического наблюдения.

Так, например, замечено, что при регистрации возраста путем опроса наиболее часто возраст округляется вокруг чисел, оканчивающихся на 0 и 5. В результате, например, получается, что 40-летних оказывается по записям значительно больше, чем 39- и 41-летних. Это явление получило в демографической статистике название аккумуляции возрастов.

Такие погрешности приходится исправлять по результатам логического и содержательного контроля при первичной обработке собранного статистического материала.

Преднамеренные ошибки, как говорит само название, возникают в силу сознательного стремления лиц, дающих сведения, исказить истину: уменьшить или увеличить величину того или иного показателя. Ясно, что преднамеренные ошибки искажают сведения в одном направлении (либо преуменьшают, либо преувеличивают). Этот род ошибок наиболее опасен для статистического исследования, и надо всегда приложить максимум усилий, чтобы выявить эти ошибки и устранить. Официальная статистика самым суровым образом борется с преднамеренными ошибками, привлекая виновных к ответственности.

Все указанные выше ошибки могут возникнуть как при сплошном, так и при несплошном статистическом наблюдении в процессе регистрации самих фактов. Отсюда и их название — *ошибки регистрации*.

Так как в процессе наблюдения всегда могут возникнуть ошибки, то, естественно, весь собранный материал должен быть подвергнут контролю с целью устранения обнаруженных ошибок.

Проверка правильности зафиксированных в статистических формулярах сведений должна производиться с точки зрения логического и арифметического контроля.

Логический контроль ставит своей целью определить соответствие ответа поставленному вопросу или соответствие между ответами на разные вопросы программы. Например, если на вопрос «возраст» обнаружен ответ «русский», то ясно, что ответ в данном случае не соответствует вопросу, что это ошибка, вызванная записью ответа не в той строке или графе.

Если же на вопрос «возраст» стоит ответ «3 года», а на вопрос «состоит ли в браке» - ответ «да», то каждый ответ здесь соответствует вопросу, но эти ответы не соответствуют друг другу. Чтобы установить, в каком же ответе содержится ошибка, следует рассмотреть ответы на другие взаимно контролируемые вопросы. Так, если в рассматриваемом случае в графе «место работы» записано наименование определенного предприятия и в графе «образование» указано «среднее» или «высшее», то ясно, что допущена ошибка в возрасте.

Можно установить логическую неточность в ответах, сопоставляя фактические показатели с плановыми, с показателями за предшествующие периоды, сопоставляя показатели по районам, находящимся в одинаковых природных условиях, сопоставляя показатели, относящиеся к одному и тому же явлению, но полученные по нескольким источникам, и т. п.

К количественным ответам, полученным как сумма, разность, произведение или часть других показателей, всегда следует применять наряду с логическим и арифметический контроль. Целью арифметического контроля является проверка правильности вычислений

Все обнаруженные ошибки по возможности должны быть исправлены. Для этого часто приходится производить контрольные опросы, запросы почтой, по телефону, телеграфу.

Наряду с ошибками регистрации при несплошном наблюдении в силу его специфики могут возникнуть расхождения между показателями несплошного наблюдения и показателями для всей совокупности при условии сплошного наблюдения. Эти возможные расхождения между показателями несплошного и сплошного наблюдения в статистике именуют *ошибками репрезентативности*. Эти ошибки тоже могут носить случайный характер (в силу несплошного наблюдения) и систематический.

Случайные ошибки, в частности при выборочном наблюдении неизбежны, но они легко поддаются учету, и при правильно организованном случайном отборе всегда можно

определить величину таких ошибок и пределы, в которых может заключаться величина изучаемого показателя во всей совокупности.

Систематические же ошибки репрезентативности, как правило, возникают при неправильном отборе, т. е. при нарушении принципа случайности отбора единиц из так называемой генеральной совокупности (например, если специально, целенаправленно отбираются единицы с заведомо завышенными или заниженными значениями признака).

Собранный в процессе статистического наблюдения статистический материал нуждается в определенной обработке, сведении разрозненных данных воедино. Научно организованная обработка материалов наблюдения (по заранее разработанной программе), включающая в себя, кроме обязательного логического и содержательного контроля собранных данных, систематизацию, классификацию (группировку) материала, составление таблиц, получение итогов и производных показателей (средних, относительных величин). Таким образом, подсчёт единиц в совокупности в целом или в группах и подгруппах с заданным значением варьирующего признака именуется в статистике *сводкой* и завершает собой *третий этап* статистического исследования.

В результате сводки достигается возможность по данным, относящимся к отдельным единицам наблюдения, охарактеризовать совокупность в целом. Так, например, на основе отчетов отдельных промышленных предприятий получают сведения о многих показателях отдельных отраслей и всей промышленности в целом: о выпуске продукции, числе занятых, производительности труда и т. п.

Получение различного рода структурных характеристик (например, определение вклада отдельных форм собственности в общий объем промышленного производства) также возможно только на основе данных, полученных в результате сводки.

Таким образом, целью сводки является сведение воедино материалов статистического наблюдения и получение обобщающих статистических показателей, отражающих сущность социально-экономических явлений и определенные статистические закономерности.

Статистическая сводка проводится по определенной программе. Причем эта программа должна быть разработана еще до сбора статистических данных, т. е. практически одновременно с составлением плана и программы статистического наблюдения.

Разработать **программу сводки** - значит определить, какие группы и подгруппы будут выделены в изучаемой совокупности, какие показатели в виде итогов, средних или

относительных величин должны быть подсчитаны для выделенных групп и в целом по совокупности, в каких таблицах будет оформлен результат сводки.

Причем эти вопросы должны решаться не механически, а с учетом цели исследования и особенностей изучаемой совокупности. Выделение тех или иных групп должно быть обоснованным, а не формальным. Сводка кроме получения итоговых и групповых показателей дает основу для последующего анализа и выявления различного рода закономерностей. Она как бы упорядочивает статистический материал, полученный при наблюдении, классифицирует и систематизирует его.

Упорядоченные в результате сводки статистические совокупности выражаются часто в виде *статистических рядов*.

В зависимости от того, по какому принципу группируются единицы статистической совокупности, статистические ряды могут носить разный характер. Если единицы совокупности систематизируются по какому-либо имманентному их признаку, то они образуют *ряды распределения* (например, распределение населения по полу, возрасту, национальности и т. п., распределение промышленных предприятий по отраслям, размерам производства формам собственности и т. п.). К рядам распределения близко примыкают и так называемые *территориальные*, или *географические*, ряды, характеризующие распределение какой-либо совокупности по отдельным территориальным единицам.

Если единицы совокупности наблюдаются в течение длительного периода, то они могут быть систематизированы по времени. Получаемые в результате такой систематизации данные именуют *рядами динамики*.

Анализ различного рода рядов распределения и рядов динамики составляет основу статистического анализа, направленного на выявление статистических закономерностей.

Организация сводки. По своей организации сводка может быть централизованной и децентрализованной. При централизованной: сводке весь материал наблюдения сосредоточивается в одном центральном органе (например, ФСГС РФ) и там обрабатывается. При децентрализованной сводке тот или иной первичный материал подвергается обработке на нескольких этапах. Например, отчеты промышленных предприятий сводятся в пределах административных районов и передаются в областные центры, где, в свою очередь, эти итоги сводят по областям и передают далее. Наконец, итоговые данные поступают в ФСГС РФ, где сводятся по стране в целом. Как правило, отчетность проходит децентрализованную сводку. Централизованная же сводка более применима и эффективна для разработки материалов больших специальных обследований и переписей. Каждая из этих сводок имеет свои положительные стороны. Так, децентра-

лизованная сводка позволяет получать более оперативно сводные сведения для руководящих органов по отдельным административным районам, позволяет быстрее уточнять те или иные сомнительные сведения. Централизованная же сводка, когда в одном центре сосредоточивается масса материала, допускает более эффективное использование, что ведет к удешевлению обработки и сокращению общих сроков сводки. Кроме того, при централизованной сводке обеспечивается единая методология детализации разработки, что является немаловажным моментом для повышения качества полученной информации.

Техника сводки. По технике, или способу, выполнения сводка может быть ручной и компьютеризированной. В современных условиях, бесспорно, механизированная сводка является доминирующей. Ручная сводка применяется в основном для небольших массивов данных и для пилотных обследований.

Ручная сводка начинается с шифровки, или разметки, материала. Цель этой шифровки — при помощи условных знаков, которыми отмечаются карточка или отдельные ответы, определить, в какую из выделяемых групп должна быть отнесена каждая единица. После шифровки формуляры (карточки) раскладываются по определенным группам и подсчитываются их число и другие показатели. Если статистический формуляр носит списочную форму, то сведения из списка вначале переносятся на безадресные карточки-фишки, которые затем уже раскладываются.

Такие же операции осуществляются при компьютеризированной сводке, но на основе специальных пакетов статистических программ. Применение автоматизированной обработки с помощью ЭВМ закладывает определенные требования к упорядочению первичной статистической документации, созданию общероссийских регистров и классификаций самых различных категорий и их кодированию.

Как уже отмечалось ранее, при обработке статистических материалов возникает необходимость выделения однородных групп, типов, а затем уже описание этих групп определенными количественными характеристиками.

Расчленение совокупности на группы, более однородные по какому-либо признаку, называется группировкой. Группировка является центральным моментом после проведения любой сводки. Именно благодаря группировкам материал наблюдения принимает систематизированный вид.

Признаки, положенные в основу группировки, называются группировочными, а группировка единиц совокупности по тому или иному признаку приводит к формированию рядов распределения.

Группировочные признаки могут носить различный характер. Одни признаки могут иметь количественное выражение (например, возраст, заработная плата, число детей в семье, урожайность отдельных сельскохозяйственных культур и т. п.). Эти признаки называют количественными, а ряды распределения, построенные по этим признакам, называют *вариационными рядами*.

Другие признаки не имеют количественного выражения. Они отражают определенные свойства, качества единиц совокупности. Эти признаки условно называют качественными (например, пол, национальность, семейное положение и т. п.). Группировки, построенные по таким качественным признакам, называют *атрибутивными рядами распределения*.

Если единицы совокупности группируются по территориальному признаку, то такие ряды именуют *географическими*, или территориальными, рядами. Такие ряды дают представление о размещении или степени распространения тех или иных явлений в пространстве. Получение итоговых сведений в территориальном разрезе обеспечивается у нас самой организацией государственной статистики, при которой, в соответствии с *принципами территориальной децентрализации и предметной централизации*, отчетные данные обрабатываются в децентрализованном порядке.

Расчленяя совокупность на части и определяя численность по группам, при помощи группировок можно решать различные задачи:

- 1) показать структуру совокупности,
- 2) выделить основные типы и формы явления,
- 3) выявить взаимосвязь между явлениями,
- 4) охарактеризовать развитие явления с течением времени.

Статистическую группировку можно строить как по одному, так и по нескольким признакам. Группировка по одному признаку называется простой. Группировка по нескольким признакам называется комбинационной. В комбинационной группировке порядок признаков обосновывается экономически, но он может быть легко изменен, а при наличии итоговой группы по каждому из них комбинационная группировка может быть свернута в любом направлении путем исключения признаков.

Комбинационные группировки приобретают особое значение в тех случаях, когда для выделения определенных групп (особенно социально-экономических типов и форм явлений) одного признака бывает недостаточно. Тогда приходится один признак брать в сочетании с другим.

В зависимости от задачи, которую собираются решить при помощи группировки, различают различные виды группировок. Группировки, при помощи

которых решается первая задача, часто называют **структурными**. Примером такой группировки может быть следующая группировка семей из 3 человек по составу в г. Москве (в %, по состоянию на 1 января):

Группировка семей, состоящих из 3 человек:	2003	2002
3 трудоспособных	18	15
2 трудоспособных и 1 ребенка	25	22
2 трудоспособных и 1 пенсионера	16	18
1 трудоспособного и 2 детей	15	17
1 трудоспособного, 1 пенсионера и 1 ребенка	26	28

Группировки, при помощи которых решается вторая задача — выделение основных типов и форм явления, называют **типологическими**. Примером такой группировки может быть распределение населения по социальным или профессиональным группам.

Структурные и типологические группировки, как правило, проводят на основе признаков, значения которых измеряются в шкале не сильнее порядковой. В этом случае удобно проводить **выделение приоритетов и доминирования** групп. **Приоритетной** называется группа, в которой сосредоточена наибольшая часть элементов совокупности. Так, в приведённой группировке семей и в 2003, и в 2002 году приоритетной является группа семей, состоящих из 1 трудоспособного, 1 пенсионера и 1 ребенка (соответственно, 26 и 28 процентов единиц). Выделение приоритетной группы даёт представление о наиболее часто встречающемся, так называемом **модальном типе** элементов совокупности. Следовательно, за год в наблюдаемой совокупности не произошло изменения приоритетов: в течение всего периода наблюдения наиболее часто в совокупности семей из 3 человек наблюдались неполные семьи, представленный тремя поколениями родственников. **Доминантной** называется группа, в которой сосредоточено от 60 до 80 процентов единиц наблюдаемой совокупности. Состав доминантной группы определяется накопленным итогом, начиная от приоритетной группы, по мере убывания частот (или частостей) соответствующих значениям признака. Так, в 2003 году доминантная группа включает 69% совокупности (семьи, состоящие из 1 трудоспособного, 1 пенсионера и 1 ребенка, из 2 трудоспособных и 1 ребенка, из 3 трудоспособных). В 2002 году доминантная группа включает 68% совокупности (семьи, состоящие из 1 трудоспособного, 1 пенсионера и 1 ребенка, из 2 трудоспособных и 1 ребенка, из 2 трудоспособных и 1 пенсионера). Следовательно, в совокупности семей из трёх человек вес доминантной группы почти не изменился, однако изменился (помолодел) её состав. **Малозначимой** называется группа, в которой сосредоточена наименьшая часть

элементов совокупности. Так, в приведённой группировке семей в 2003 году малозначимой является группа семей, состоящих из 1 трудоспособного и 2 детей, а в 2002 году малозначимой была группа семей, состоящих из 3 трудоспособных, которая в 2003 году вошла в состав доминантной группы.

Группировки, при помощи которых выявляется взаимосвязь между явлениями, называют **аналитическими**. При построении таких группировок прежде всего из двух (или нескольких) взаимосвязанных показателей один рассматривается как результат влияния остальных, прочие – как факторы (т. е. влияющие на результат), определяющие вариацию результирующего показателя. Однако следует иметь в виду, что **понятие факторного и результативного признаков рассматривается для каждого конкретного случая особо, так как тот признак, который служит факторным в одном случае, может выступать в качестве результативного признака в другом случае.**

Чтобы при помощи аналитической группировки выявить зависимость между показателями, необходимо разгруппировать единицы совокупности по факторному признаку и для каждой выделенной группы рассчитать среднее значение результативного показателя, а затем проследить за изменениями последнего от группы к группе.

Результаты проведения аналитической группировки могут быть представлены в форме таблицы, у которой подлежащее формируют градации результирующего признака, а сказуемое – градации признаков факторов. Клетки аналитической, или корреляционной таблицы заполняются частотами или частостями, характеризующими распределение элементов совокупности в зависимости от сочетания вариантов значений исследуемых признаков. В каждой клетке аналитической таблицы частость распределения представляет собой условную вероятность появления заданного значения результирующего признака (при условии, что факторы зафиксированы на определённом уровне). Тогда итоговый столбец таблицы покажет общее для совокупности в целом, то есть безусловное распределение частостей по градациям результирующего признака (*независимо* от того, какие значения принимали признаки-факторы у соответствующего элемента совокупности). Если факторные признаки не оказывают существенного влияния на вариацию результирующего признака, то распределение условных вероятностей в каждом столбце аналитической таблицы будет пропорционально независимому распределению вероятностей в итоговом столбце. Следовательно, оценив статистическую значимость отклонения условных вероятностей от соответствующих независимых вероятностей, можно проверить гипотезу о наличии или отсутствии существенного влияния факторов на результирующий показатель.

Возможно также представление результатов аналитической группировки в форме так называемой групповой таблицы. Для её построения, после выделения групп элементов совокупности по комбинации значений факторов и результирующего признака, для каждой группы по признаку-фактору вычисляется среднее значение признака результата. Сравнение полученных групповых средних (при условии достаточно большого числа наблюдений) даёт возможность сформулировать гипотезу о существующей статистической закономерности влияния факторов на результат.

Метод аналитических группировок широко используется на практике для статистического исследования зависимостей, как между количественными, та и между атрибутивными показателями.

Построение аналитической группировки на основе комбинации атрибутивных показателей приводит к построению корреляционных таблиц, которые называются **таблицами сопряжённости** и которые представляют собой единственную реальную основу выявления взаимосвязей в условиях непараметрической информации.

Группировки, которые служат для решения задач четвёртого типа, называются **динамическими**. При проведении таких группировок данные о совокупности, относящиеся к разным моментам или периодам времени, обобщённо представляются в виде **таблицы, или матрицы переходов (матрицы мобильности)**. Подлежащее и сказуемое таблицы переходов формируют градации значений варьирующего признака, соответственно, на базисном и на сравниваемом уровне фиксирования информации. В клетках таблицы переходов фиксируются частоты или частоты, характеризующие распределение элементов совокупности по комбинации значений варьирующего признака и признака времени. Таким образом, на пересечении, например, первой строки и второго столбца матрицы переходов указывается абсолютное или относительное число элементов совокупности, которые на базисном уровне времени имели первое значение варьирующего признака, а на сравниваемом уровне времени имели второе значение варьирующего признака. Если количество выделенных групп значений варьирующего признака k с течением времени не изменяется, то мы получаем квадратную матрицу переходов, которую часто называют «шахматной таблицей». Матрица переходов отражает число объектов, перешедших (или не перешедших) за исследуемый период из одной группы в другую. Анализ матрицы переходов удобнее проводить на основе относительных величин – частостей.

Частость представляет количество элементов совокупности, которые имели значение варьирующего признака на заданном уровне, оцененное в соотношении с общим количеством элементов совокупности. Следовательно, в каждый момент времени частость

можно интерпретировать как вероятность появления соответствующего варианта значения исследуемого признака. Таким образом, **строки матрицы переходов** образованы условными вероятностями: частотами для соответствующих градаций варьирующего признака на сравниваемом временном уровне (*при условии*, что на базисном уровне варьирующий признак имел определённое значение). Тогда как в **итоговой строке** мы фиксируем распределение частот по градациям варьирующего признака на сравниваемом временном уровне **независимо** от того, какое значение этот признак принимал у соответствующего элемента совокупности на базисном временном уровне. Следовательно, независимое распределение вероятностей в итоговой строке матрицы переходов можно рассматривать в качестве «теоретического», «идеального» распределения элементов совокупности по вариантам исследуемого признака: если предшествующие значения вариантов не влияют на будущие значения вариантов, то распределение частостей в каждой строке матрицы переходов будет пропорционально их распределению в итоговой строке. Это означает, что для **оценки зависимости значений признака от предшествующих временных уровней необходимо провести динамическую группировку, построить матрицу переходов, проанализировать распределение частот или частостей в клетках таблицы и сравнить их (на основе соответствующих показателей, которые будут рассмотрены позже) с частотами или частостями в итоговой строке.**

Аналогичные рассуждения можно было бы привести и **для столбцов матрицы переходов**, но шкала измерения времени не допускает сортировки временных уровней.

Прежде чем рассмотреть пример построения динамической группировки результатов наблюдения, следует напомнить о необходимости соблюдения **требования однородности рассматриваемых данных**. В случае изучения данных в динамике это означает, что в последовательные моменты или периоды времени статистический мониторинг должен обеспечивать сопоставимость проводящихся измерений, то есть программа обследований должна носить **лонгитюдный характер** (относительно методов формирования совокупности, конкретизации признаков для статистического применения, выбора единиц измерения и соизмерителей и т. п.).

Пример.

Имеются данные о динамике минимального уровня заработной платы по рабочим профессиям химического производства предприятий Санкт-Петербурга:

Профессии	апрель 2003	апрель 2004
Аппаратчики разложения	2485,0	2478,6
Аппаратчики синтеза	2402,0	2480,4

Аппаратчики электролиза	2421,0	2407,5
Аппаратчики гидрохлорирования	2875,0	3429,7
Аппаратчики хемосорбции	2801,0	2906,3
Операторы дистанционного пульта управления в химическом производстве	2538,0	2550,0
Операторы технологических установок	4123,0	4140,7
Машинисты оборудования распределительных нефтебаз	2999,0	3780,3
Машинисты технологических насосов	3483,0	5022,0
Машинисты технологических компрессоров	4291,0	4116,3
Помощники мастера	3309,0	2978,0
Шихтовщики	2486,0	2578,4

Проведите динамическую группировку данных об уровне минимальной зарплаты в химическом производстве, выделив две группы рабочих профессий по уровню оплаты труда. Оцените изменение общей ситуации с оплатой труда на химических предприятиях Санкт-Петербурга в течение периода наблюдения.

Объектом статистического исследования является совокупность рабочих химического производства на предприятиях Санкт-Петербурга. определение границ совокупности производится на основе признака занятости рабочего в химическом производстве на предприятиях Санкт-Петербурга. Единица наблюдения - рабочий химического производства на предприятии Санкт-Петербурга. Статистическая единица – бухгалтерия предприятия Санкт-Петербурга. Время проведения наблюдений – апрель 2003 и апрель 2004 года. В условиях предложенного задания предполагаем, что мониторинг проводится по лонгитюдной программе. Для каждой единицы наблюдения фиксируются статистические признаки: атрибутивный признак профессии количественный признак уровень заработной платы. Профессия положена в основу типологической группировки. Для каждой профессиональной группы определён статистический показатель минимального уровня заработной платы. Результаты типологической группировки для каждого уровня лонгитюдного мониторинга представлены в таблице выше.

По различным профессиям значения показателя меняются в различных направлениях и на различную величину, поэтому сделать однозначный вывод об общем изменении ситуации с оплатой труда непосредственно по исходным данным даже по столь малой выборке невозможно. Необходимо провести динамическую группировку.

Для выделения двух групп профессий по уровню оплаты труда, прежде всего, следует определить группировочный признак и его градации. В качестве группировочного признака будем использовать минимальный уровень заработной платы. Признак измеряется в абсолютной шкале, следовательно, можно перейти к интервальной шкале как более слабой. Градации группировочного признака (низкооплачиваемые профессии, высокооплачиваемые профессии) можно определить по-разному в зависимости от выбора критерия среднего значения.

Т.к. за исследуемый период изменение значений показателя не обнаруживает определённой тенденции, можно границу между группами выбрать на основе половины размаха вариации и не менять её в различные периоды.

В апреле 2003 минимальное и максимальное значение показателя в совокупности составляли, соответственно, 2402 рубля и 4291 рубль; в апреле 2004 - 2407,5 рубля и 4116

рублей. Округляя граничное значение группировочного признака, и учитывая средний размер прожиточного минимума в Санкт-Петербурге в рассматриваемый период времени, принимаем: в группу высокооплачиваемых включаем профессии с минимальным уровнем оплаты труда не ниже 3000 рублей.

Построим динамическую группировку и оформим её результаты в виде шахматной таблицы.

2003	2004	низкооплачиваемые профессии	высокооплачиваемые профессии	всего профессий в группе
низкооплачиваемые профессии		6	2	8
высокооплачиваемые профессии		1	3	4
всего профессий в группе		7	5	12

Переходим к интерпретации полученной матрицы перехода.

Из 12 рабочих профессий на химических предприятиях Санкт-Петербурга 6 следовало отнести к низкооплачиваемым и в 2003 и в 2004 году. У 1 профессии (помощников мастера) в апреле 1998 года уровень оплаты труда существенно снизился. Однако 5 рабочих профессий в апреле 2004 года относились к категории высокооплачиваемых, при этом по 2 из них минимальный уровень оплаты труда за рассматриваемый период существенно повысился. В целом, не проводя расчёт специальных показателей, по распределению частот шахматной таблицы, можно отметить, что общий уровень оплаты труда рабочих на химических предприятиях Санкт-Петербурга за исследуемый период незначительно повысился.

Нередко исследователь, не имея в своем распоряжении первичных данных, а, располагая уже обработанными, сгруппированными данными, вынужден **перегруппировывать материал** в соответствии с задачами своего исследования. Такая перегруппировка уже сгруппированного материала, т. е. образование новых групп на основе ранее произведенной группировки, называется **вторичной группировкой**.

К вторичной группировке прибегают: когда из большого числа первоначально образованных групп надо получить меньшее число более крупных, более характерных групп; когда в целях сравнения нужно привести в сопоставимый вид по-разному сгруппированный материал. Наиболее простым способом вторичной группировки является такой, когда новые группы образуются путем укрупнения интервалов, т. е. путем объединения в одну группу нескольких мелких групп, полученных при первичной группировке. При дроблении отдельных групп (интервалов) соответственно дробят и суммарные показатели, характеризующие эти группы. Рассмотрим группировку аграрных предприятий одного из районов (см. таблицу ниже).

Таблица 1
Группировка хозяйств по числу дворов

Группы хозяйств по числу дворов	Число хозяйств	Посевная площадь, га	Валовой сбор зерновых, ц	Урожайность, ц/га
До 40	1	900	6 750	7,5
41—60	2	2 100	16 800	8,0
61—150	7	8 000	70 400	8,8
151—300	12	32 000	288 000	9,0
301—500	10	30 000	285 000	9,5
Свыше 500	3	10 000	100 000	10,0
Итого	35	83 000	766 950	9,2

Предположим далее, что в других районах группировка хозяйств по числу дворов проведена с выделением таких групп: до 50: 51—100, 101—200, 201—500, свыше 500.

Тогда, чтобы данные табл. 5 стали сопоставимыми с показателями других районов, они должны быть перегруппированы в соответствии с интервалами, общими для всех районов. Пересчитаем все показатели для выделяемых групп.

В группу хозяйств с числом дворов до 50, очевидно, войдут все хозяйства с числом дворов до 40 и половина хозяйств из групп с числом дворов от 41 до 60, т. е. общая численность первой группы составит 2. Посевная площадь зерновых культур в этой группе будет получена как сумма посевной площади хозяйств первой группы и половины площади хозяйств из второй группы, т. е. $900 + 0,5 * 2100 = 1950$ га. Аналогично рассчитывается и валовой сбор зерновых в этой группе хозяйств $6750 + 0,5 * 16800 = 15150$ ц.

В группу хозяйств с числом дворов от 51 до 100 отнесем оставшуюся половину хозяйств второй группы и $4/9$ хозяйств третьей группы (так как 100 делит интервал 60—150 в соотношении 4 к 9). Отсюда число хозяйств в образуемой группе оценим как $1 + 4/9 * 7 = 4$, их посевная площадь составит $(2100 - 1050) + 4/9 * 8000 = 4605$ га, а валовой сбор зерновых, соответственно, $(16800 - 8400) + 4/9 * 70400 = 8400 + 31290 = 39690$ ц.

В следующую группу хозяйств с числом дворов от 101 до 200 следует отнести оставшиеся $5/9$ хозяйств третьей группы и $1/3$ хозяйств с числом дворов 151—300, общая численность хозяйств этой группы составит $4 + 1/3 * 12 = 4 + 4 = 8$, а их посевная площадь $(8000 - 3555) + 1/3 * 32000 = 4445 + 10667 = 15112$ га и валовой сбор зерновых $(70400 - 31290) + 1/3 * 288000 = 39110 + 96600 = 135710$ ц. Последующие группы образуются аналогично.

Дробление показателей во всех случаях проводилось в предположении, что нарастание признака среди единиц каждой группы происходит равномерно. Это предположение обоснованно, так как исходным требованием к формированию статистической совокупности является её однородность. Средняя урожайность в каждой группе хозяйств определяется путем деления валового сбора на посевную площадь.

Результаты вторичной группировки представлены в таблице ниже.

Таблица 2

Группировка хозяйств по числу дворов

Группы хозяйств по числу дворов	Количество хозяйств	Посевная площадь зерновых га	Валовой сбор зерновых, ц	Урожайность зерновых, ц/га
До 50	2	1 950	15 150	7,8
51—100	5	4 605	39 690	8,6
101—200	8	15 112	135 110	8,3
201—500	18	51 333	477 000	9,1
Свыше 500	3	10 000	100 000	10,0
Итого*	35	83 000	766 950	9,2

Рассмотренный выше пример иллюстрировал перегруппировку данных путем укрупнения интервалов. Есть другой способ вторичной группировки — по удельному весу намечаемых к образованию групп. Суть этого способа заключается в том, что на основе предварительного изучения первично сгруппированного материала устанавливается определенный удельный вес (доля) отдельных групп единиц совокупности и все показатели, составляющие 100% или подлежащие суммированию, перегруппировываются соответственно удельному весу намеченных групп.

В результат сводки и группировки данных формируются ряды распределения — ряды чисел, характеризующие, каким образом распределяются единицы некоторой совокупности по тому или иному атрибутивному или количественному признаку.

Ряды, построенные по атрибутивному признаку, называют атрибутивными рядами распределения. Группировка собранного материала **по значениям признака, измеренного в абсолютной шкале**, приводит к формированию разновидности ряда распределения, которая называется **вариационным рядом**.

Элементами любого ряда распределения являются варианты значений признака и численности единиц, относящихся к каждой группе. Абсолютные численности единиц, относящихся к каждой группе, называются частотами ряда распределения. Относительные численности, характеризующие удельный вес каждой группы в общей численности единиц совокупности, называются частостями распределения. Для расчета обобщенных характеристик вариационных рядов можно пользоваться как частотами, так и частостями распределения.

Различия единиц совокупности в отношении некоторого измеряемого признака называют **вариацией** этого признака. Уровень вариации определяется для совокупности в целом и для каждой из выделенных однородных групп. Вариация признака внутри группы, не зависящая от вариации группировочного признака, называется **случайной**.

Вариация изучаемого признака, зависящая от значений признака, положенного в основу выделения более однородных групп в пределах менее однородной совокупности, называется **объясняемой**. Случайная и объясняемая вариация представляют собой составляющие **общей** вариации признака в совокупности.

Вариация количественных признаков может быть **дискретной (прерывной) или непрерывной**. В случае **дискретной вариации** величина количественного признака у единиц совокупности может принимать только вполне определенные значения, отличающиеся друг от друга на одну или несколько единиц. Так, количество детей в семье может выражаться только целыми числами; количество осей в вагонах может выражаться только четными целыми числами. В случае **непрерывной вариации** величина признака у единиц совокупности может принимать в определенном численном промежутке любые значения, как угодно мало отличающиеся друг от друга, так, доход одного человека может отличаться от дохода другого на сколь угодно малую величину.

При построении вариационного ряда непрерывного признака **невозможно указать каждое значение варианта**, поэтому совокупность распределяется по интервалам его значений, то есть переходить к интервальной шкале значений. Анализ особенностей единиц, сосредоточенных в одном интервале позволяет в дальнейшем исследовании структуры совокупности перейти к порядковой шкале, определив основное качественное различие между группами. В зависимости от характера вариации при формировании вариационного ряда интервалы можно брать как равные, так и неравные. **Формирование равных интервалов возможно только** в том случае, когда величина разницы между максимальным и минимальным значением признака в совокупности (так называемый размах вариации) позволяет пренебречь уменьшением относительной значимости единичного различия между соседними значениями признака при возрастании вариантов значений в упорядоченном (ранжированном) ряду. Во всех остальных случаях следует укрупнять интервалы по мере возрастания вариантов. **Обязательным требованием к проведению группировки является наличие возможности для каждой единицы наблюдения быть отнесённым к той или иной группе вариантов значений признака, причём только к одной единственной из них. Кроме того, необходимо добиться отсутствия незаполненных групп.** Максимальное количество выделяемых групп k определяется в зависимости от количества произведённых наблюдений n по формуле Стерджесса:

$$k = 1 + 3,322 * \lg n$$

Для каждого из интервалов указываются частоты или частости, т. е. абсолютное или относительное число единиц, у которых значение варианта находится в пределах границ заданного интервала.

Верхний и нижний интервалы рядов во многих случаях формируются открытыми, то есть для первого интервала указывается только верхняя граница, а для последнего — только нижняя. Использование открытых интервалов удобно, когда в совокупности встречается незначительное число единиц с очень малыми или очень большими значениями вариантов, резко отличающимися от всех остальных значений. Чтобы закрыть открытый интервал, полезно привлекать дополнительную теоретическую информацию о характере вариации признака в совокупности, о наличии нормативно, экспертно или эмпирически установленных границ возможной вариации признака.

Интервальные вариационные ряды строятся и для признаков с дискретной вариацией. Такие ряды были представлены в таблицах 1 и 2. Число дворов, объединенных в одно хозяйство, строго говоря, дискретный признак. Но указывать отдельно каждое его возможное значение нецелесообразно, так как это не только затруднило бы рассмотрение его вариации, но и могло бы исказить полученные результаты. Поэтому хозяйства распределены по группам, объединяющим все попавшие в принятый интервал значений дискретно варьирующего признака.

При построении интервального ряда по дискретному признаку границы смежных интервалов не повторяют друг друга: следующий интервал начинается со следующего по порядку дискретного значения признака, как это показано в таблицах 1 и 2.

Следует заметить, что вариация по одному и тому же признаку может быть представлена разными интервальными рядами, в которых частоты будут зависеть не только от величины вариантов и характера вариации, но и от размеров интервалов: чем больший взят интервал, тем больше попадает в него единиц совокупности. Если построен ряд с равными интервалами, то частоты и частости дают представление о том, как заполнен тот или иной интервал единицами совокупности. Сравнивая частоты ряда с неравными интервалами, еще нельзя судить об относительной заполненности разных интервалов. В этом случае необходимо для сравнения заполненности интервалов рассчитывать особый показатель, называемый плотностью распределения, который показывает, сколько (в абсолютном или относительном выражении) единиц совокупности приходится на единицу изменения варианта в интервале. Таким образом, средняя плотность в интервале — это частное от деления соответствующей частоты или частости на длину интервала. В первом случае получается абсолютная, во втором — относительная плотность распределения.

Любое распределение можно охарактеризовать с помощью накопленных частот или частостей. **Накопленная частота (или частость) показывает, какое число (соответственно, в абсолютном или относительном выражении) единиц совокупности имеет величину варианта, не большую заданной.** Если вместо абсолютных частот взять частости, то аналогично получим накопленные частости. В табл. 1 в последних графах приведены накопленные частости для распределения колхозов по числу дворов.

Накопленная частота (частость) для данного варианта или для верхней границы данного интервала получается суммированием (накапливанием) частот (частостей) всех предшествующих интервалов, включая данный.

Частота, частость, плотность и накопленная частота вариационного ряда - это различные функции от величины варианта.

Если варианты обозначить буквой « x » с субиндексами 1, 2, т. е. $X_1, X_2, X_3, \dots, X_m$, а частоты или частости буквой f с соответствующими номерами, то любой вариационный ряд (дискретный – по значениям вариантов, интервальный – по центральным значениям выделенных однородных интервалов) можно записать таким образом:

Варианты x	Частоты f
x_1	f_1
x_2	f_2
⋮	⋮
x_m	f_m
	Σf

При этом пары значений вариантов и частот можно считать парами значений аргумента и его функции. Для непрерывно варьирующего признака такая трактовка частот (частостей) без формирования интервалов неприемлема, так как эти показатели относятся не к точечным значениям признака, а к интервалам. Накопленная же частота (частость) может рассматриваться как функция значений признака в любом случае.

Один и тот же материал дает диаметрально противоположные выводы при различных приёмах группировок, и правильность выводов, сделанных на основе группировок, зависит от правильности выбора группировочного признака. Поэтому *группировочные признаки не должны быть случайными, формальными, а должны быть главными, существенными, и к выбору их следует подходить на основе проведения самого серьёзного нормативного анализа, дифференцированно в зависимости от места и времени. Одни признаки, существенные для характеристики явления в определенных условиях, могут оказаться формальными, случайными в других условиях.*

При неправильной группировке статистическое исследование не только не описывает закономерности, а затушевывает их, тенденциозно маскируя многие их стороны. При неудовлетворительной сводке, при неправильной или недостаточной группировке может получиться — и постоянно получается при целенаправленной рекламной обработке результатов проводящихся социологических опросов и микропереписей производственных единиц — такой результат, что необыкновенно детальные, великолепные данные, имеющиеся о каждом отдельном элементе совокупности, исчезают, теряются, пропадают и даже искажают обобщающую информацию о совокупности в целом. *Если группировочные признаки выделяемых типов явлений выбраны неправильно, подобраны неполно, то самая лучшая статистическая перепись не может дать адекватной социально-экономической и политической картины действительности.*

Закреплённая формально для всей совокупности группировка называется классификацией. При этом для обеспечения сопоставимости и дифференцированности данных, все классификации подлежат согласованию между статистическими службами и потребителями информации и гармонизируются на международном уровне.

Вспомним, что однородность является основным свойством статистической совокупности. Поэтому в процессе группировки, для получения адекватной и упорядоченной статистической информации, происходит выделение более однородных частей - групп описываемых единиц - в составе менее однородного целого – общего состава совокупности. При этом для обеспечения сопоставимости и дифференцированности данных классификации подлежат согласованию между статистическими службами и потребителями информации и гармонизируются на международном уровне.

Основными задачами единой системы классификации и кодирования технико-экономической и социальной информации определены следующие:

- создание условий для формирования единого информационного пространства на территории Российской Федерации;
- систематизация информации по единым классификационным правилам и их использование при прогнозировании социально-экономического развития страны и ведении учета и отчетности;
- информационное обеспечение налогообложения, лицензирования, квотирования, операций с недвижимостью, социального страхования, финансового посредничества;
- содействие специализации и кооперированию в области производства продукции и оказания услуг;

- упорядочение стандартизации и сертификации выпускаемой продукции и оказываемых услуг;
- создание условий для унификации документации при осуществлении межотраслевого документооборота;
- обеспечение совместимости информационных систем и ресурсов;
- обеспечение межотраслевого обмена информационными ресурсами.

Рассмотрим некоторые важнейшие стандартные системы классификации экономической информации.

Общероссийский классификатор видов экономической деятельности (ОКВЭД).

Условное обозначение: ОК-029-2001 (КДЕС ред. 1).

Объекты классификации: виды экономической деятельности, осуществляемые хозяйствующими субъектами в различных отраслях экономики.

Признаки классификации: сфера деятельности, процесс производства, используемые сырье и материалы и др.

Кодовое обозначение: шестизначный цифровой десятичный код.

Формула (типовая) структуры кода: XX +X+X+X+X.

Форма и пример записи позиций

Код	Наименование
51.47.22	Оптовая торговля газетами и журналами

Пример построения кодов классификационных группировок

Обозначение	Код	Наименование
Раздел	0	Обрабатывающие производства ~~
Подраздел	0А	Производство пищевых продуктов, включая напитки, и табака
Класс	15	Производство пищевых продуктов, включая напитки
Подкласс	15.5	Производство молочных продуктов
Группа	15.51	Переработка молока и производство сыра
Подгруппа	15.51.1	Производство цельномолочной продукции
Вид	15.51.12	Производство сметаны и жидких сливок

В ОКВЭД для уточнения содержания группировок видов экономической деятельности и обеспечения однозначной трактовки и интерпретации используемых понятий приведены текстовые описания группировок, дающие в некоторых случаях отсылки к другим группировкам. Например, группировка 27.43 «Производство свинца, цинка и олова» включает в себя:

- производство свинца, цинка и олова из руды;

- производство свинца, цинка и олова методом электролитического рафинирования отходов и лома свинца, цинка и олова;

- производство сплавов свинца, цинка и олова;

- производство полуфабрикатов из свинца, цинка и олова или их сплавов.

**Общероссийский классификатор видов экономической деятельности,
продукции и услуг (ОКДП)**

Условное обозначение: ОК 004-93.

Объекты классификации: виды экономической деятельности, осуществляемые во всех отраслях экономики: сельское хозяйство, охота и лесоводство; рыболовство; горнодобывающая промышленность и разработка карьеров; обрабатывающая промышленность и другие; а также продукция и услуги, как результат экономической деятельности.

Признаки классификации: функциональное назначение, направленность использования, характерные свойства, конструктивно- технологические особенности и другие.

Кодовые обозначения: четырехзначный, цифровой, десятичный код для видов экономической деятельности; семизначный, цифровой, десятичный код для группировок продукции и услуг

Формулы (типовые) структуры кодов: XX+X+X - для видов экономической деятельности; XX+X+X+XX+X - для видов продукции и услуг

Форма и пример записи позиций для видов экономической деятельности:

Группа	Подгруппа	Наименование деятельности услуг
192		Производство обуви
	1921	Производство кожаной обуви

Пример построения кодов классификационных группировок:

Раздел А Сельское хозяйство, охота и лесоводство

01 Сельское хозяйство, охота и связанная с этим деятельность по предоставлению услуг

011 Растениеводство: выращивание сельскохозяйственных культур, товарное овощеводство, садоводство

0111 Выращивание зерновых культур

Форма и пример записи позиций для видов продукции и услуг:

Класс	Вид	Наименование видов продукции (услуг)
0111000		Зерновые, основные зерновые культуры
	0111110	Пшеница озимая твердая

	0111111	Зерно пшеницы озимой твердой
--	---------	------------------------------

Пример построения кодов классификационных группировок:

Раздел А. Продукция и услуги сельского хозяйства, охоты и лесоводства

0111 Зерновые, основные зернобобовые культуры

011111 Пшеница озимая твердая

0111111 Зерно пшеницы озимой твердой

Для уточнения содержания группировок видов экономической деятельности и обеспечения однозначной трактовки и интерпретации используемых понятий в ОКДП приведены текстовые описания группировок, например:

Группировка	Описание группировки
2524	Производство тары из пластмасс. Эта подгруппа включает виды деятельности, связанные с производством тары из пластмасс. Номенклатура производимых видов продукции включает мешки, пакеты, бочки, фляги, флаконы, бутылки, банки, ящики, коробки, лотки из полиэтилена, поливинилхлорида, полистирола, полипропилена, фторопласта, полиамида и прочих полимерных материалов

Общероссийский классификатор

органов государственной власти и управления (ОКОГУ)

Условное обозначение: ОК 006-93.

Объектами классификации являются: федеральные органы представительной (законодательной), исполнительной и судебной власти, органы государственной власти субъектов Российской Федерации; органы местного самоуправления; объединения предприятий и организаций, выполняющие крупные экономические функции в народном хозяйстве.

Признаки классификации: организационный уровень и характер выполняемых функций.

Кодовое обозначение: пятизначный, цифровой, десятичный код.

Формула (типовая) структуры кода: X+X+X+XX.

Форма и пример записи позиций:

Код	Наименование	
	полное	сокращенное
10000	Федеральные органы государственной власти	
13000	Исполнительная власть Российской Федерации	
13100	Федеральные министерства	

13101	Министерство Российской Федерации по атомной энергии	Минатом России
-------	--	----------------

Система российских национальных классификаторов гармонизирована с международной системой и учитывает специфические сырьевые, технологические и экономические условия внутреннего рынка. Классифицируются субъекты и объекты экономического оборота. В качестве объектов экономического оборота классификации подлежат продукты, услуги, денежные средства, активы, пассивы или любые другие носители стоимости.

Ниже приведён перечень общероссийских классификаторов технико-экономической и социальной информации.

В соответствии с международной системой кодирования информации классификаторы расположены иерархически, т.е. в порядке убывания значимости с точки зрения социальных процессов (см. номер в обозначении классификатора).

Общероссийские классификаторы технико-экономической и социальной информации.

№ п/п	Наименование классификатора	Аббревиатура	Обозначение
1	Общероссийский классификатор услуг населению	ОКУН	ОК 002-93
2	Общероссийский классификатор информации по социальной защите населения	ОКИСЗН	ОК 003-99
3	Общероссийский классификатор видов экономической деятельности, продукции и услуг	ОКДП	ОК-004-93
4	Общероссийский классификатор продукции	ОКП	ОК 005-93
5	Общероссийский классификатор органов государственной власти и управления	ОКОГУ	ОК 006-93
6	Общероссийский классификатор предприятий и организаций	ОКПО	ОК 007-93
7	Общероссийский классификатор специальностей по образованию	ОКСО	ОК 009-93
8	Общероссийский классификатор занятий	ОКЗ	ОК 010-93
9	Общероссийский классификатор управленческой документации	ОКУД	ОК 011-93
10	Общероссийский классификатор основных фондов	ОКОФ	ОК 013-94
11	Общероссийский классификатор валют	ОКВ	ОК014-2000 (МК 003-97 (ISO4217))
12	Общероссийский классификатор единиц измерения	ОКЕИ	ОК015-94

13	Общероссийский классификатор профессий рабочих, должностей служащих и тарифных разрядов	ОКПДТР	ОК 016-94
14	Общероссийский классификатор специальностей высшей научной квалификации	ОКСВНК	ОК017-94
15	Общероссийский классификатор информации о населении	ОКИН	ОК 018-95
16	Общероссийский классификатор объектов административно-территориального деления	ОКАТО	ОК 019-95
17	Общероссийский классификатор начального профессионального образования	ОКНПО	ОК 023-95
18	Общероссийский классификатор экономических регионов	ОКЭР	ОК 024-95
19	Общероссийский классификатор стран мира	ОКСМ	ОК 025-95
20	Общероссийский классификатор информации об общероссийских классификаторах	ОКОК	ОК 026-95
21	Общероссийский классификатор форм собственности	ОКФС	ОК 027-99
22	Общероссийский классификатор организационно-правовых форм	ОКОПФ	ОК 028-99
23	Общероссийский классификатор видов экономической деятельности	ОКВЭД	ОК-029-2001