

## Тема: «ОБРАБОТКА СТАТИСТИЧЕСКОГО МАТЕРИАЛА»

На этап обработки (разработки) статистического материала исследователь «приходит» с собранной документацией (анкетами, опросными листами, статистическими картами, картами выкопировки, дневниками испытуемых, документами официальной статистики и т.д.). «Уйти» с этого этапа работы исследователь должен, имея компьютерную базу данных.

На этапе обработки (разработки) статистического материала последовательно выполняется:

1. Механическая и логическая проверка собранных учетных документов
2. Группировка и сводка материала
3. Составление макетов статистических таблиц
4. Кодирование значений учитываемых признаков и создание компьютерной базы данных

Далее подробно рассмотрим технологию выполнения этих работ.

### 1. МЕХАНИЧЕСКАЯ И ЛОГИЧЕСКАЯ ПРОВЕРКА УЧЕТНЫХ ДОКУМЕНТОВ

**Механической** называется проверка полноты оформления собранной документации (статистических карт, карт выкопировки, опросных листов, анкет и пр.). Исследователь проверяет, все ли необходимые ему сведения по этой конкретной единице наблюдения отражены в учетном документе.

- Если отсутствуют очень важные для данного исследования сведения, причем получить их дополнительно нет никакой возможности, то эту единицу наблюдения не берут в дальнейшую разработку (т.е. этот учетный документ выбрасывают в мусорную корзину).
- Если отсутствуют какие-либо второстепенные сведения, не имеющие очень важного значения для данного исследования, то эту единицу берут в дальнейшую разработку, но в последующем анализе используют только имеющуюся информацию о ней.

**Логической** называется проверка соответствия друг другу сведений, полученных о каждой изучаемой единице. Несоответствие сведений может быть, например, следствием невнимательности специалиста, собирающего материал, в результате чего в документе появляется непредумышленная описка или опечатка. Например, в учетном документе указаны следующие сведения об испытуемом:

Ф.И.О.: Руставели Г.Д.

Пол: женский

Диагноз: Х-сцепленный ихтиоз

Однако Х-сцепленный ихтиоз встречается только у мужчин, а определить пол испытуемого по фамилии в данном случае не удастся.

Действия исследователя по результатам логической проверки собранных документов в целом будут такими же, как и при механической проверке. Если есть возможность установить истину, то это необходимо сделать. Если такой возможности нет, а несоответствие сведений является принципиально важным, то данную единицу исключают из дальнейшей разработки.

### 2. ГРУППИРОВКА И СВОДКА МАТЕРИАЛА

**Группировка данных** – это очень важная часть статистической разработки первичного материала. Она обеспечивает его систематизацию и **определяет весь ход последующего статистического анализа**. Подходы к выполнению группировки продумывают еще на организационном этапе исследования, а сбор материала осуществляют уже с учетом этих, заранее определенных, подходов. Например, признак «Возраст испытуемого» можно учитывать либо как количественный признак, либо как описательный порядковый признак.

- Если возраст учтен как *количественный признак* («..... лет»), то возможными методами анализа<sup>1</sup> данных будут являться:
  - ✓ определение среднего возраста испытуемых (мода, медиана, среднее арифметическое),
  - ✓ оценка характера распределения (симметричное или несимметричное, соответствует нормальному или же нет),
  - ✓ оценка разнообразия возраста в группе испытуемых (по амплитуде, по дисперсии, по среднеквадратическому отклонению, по коэффициенту вариации),
  - ✓ сравнение среднего возраста в изученной группе лиц со средним возрастом другой группы по параметрическим критериям (если доказана нормальность распределения),
  - ✓ анализ корреляционной связи между возрастом испытуемых и каким-либо другим количественным признаком по коэффициенту Пирсона.
- Если возраст учтен как *описательный порядковый признак* (со сгруппированными значениями «до 20 лет», «20-29 лет», «30-39 лет», «40 лет и старше»), то методы последующего статистического анализа будут совсем иными:
  - ✓ описание структуры группы по возрасту (т.е. расчет относительных экстенсивных показателей),
  - ✓ сравнение повозрастной структуры изучаемой группы с повозрастной структурой другой интересующей исследователя группы по непараметрическим критериям (например, «хи-квадрат»),
  - ✓ анализ корреляционной связи между возрастом испытуемых и каким-либо другим количественным или качественным признаком по коэффициенту Спирмена.

В ходе статистической обработки материала исследователь обычно использует следующие варианты группировок:

**1) Объединение мелких однородных групп в более крупные группы.** Этот вариант группировки применяется уже после сбора материала, в процессе статистической обработки данных, если выясняется несостоятельность мелких групп (малое число наблюдений в каждой мелкой группе, нечетко выраженный характер распределения и т.п.). Например, признак «Сопутствующие заболевания» часто имеет огромное множество разнообразных значений, каждое из которых встречается в изученной совокупности 1-2 раза. Чтобы четче представить общую картину сопутствующих заболеваний в группе испытуемых, можно их объединить в группы, соответствующие классам Международной классификации болезней (МКБ), например: «Болезни нервной системы», «Болезни органов дыхания», «Болезни органов пищеварения» и пр.

**2) Объединение единиц статистической совокупности в группы по значениям факторных признаков.** Чаще всего факторными признаками являются пол, возраст, уровень доходов, уровень образования, условия проживания и ряд других. Например, все единицы наблюдения изученной совокупности можно объединить в следующие группы:

- ✓ мужчины и женщины (по фактору пола);
- ✓ лица моложе трудоспособного возраста; лица трудоспособного возраста; лица старше трудоспособного возраста (по фактору возраста);
- ✓ семьи с доходами до 10 тысяч рублей на человека в месяц; семьи с доходами от 10 до 30 тысяч рублей на человека в месяц; семьи с доходами свыше 30 тысяч рублей на человека в месяц (по фактору ежемесячного подушевого дохода семьи);
- ✓ студенты, проживающие с родителями; студенты, проживающие в общежитии; студенты, проживающие на съемной квартире (по фактору условий проживания).

Подходы к такой группировке исследователь продумывает еще на организационном этапе при составлении программы исследования.

<sup>1</sup> Перечисленные методы анализа запоминать не нужно. Здесь они приведены только в качестве примера.

**3) Комплексная группировка, которая обеспечивает формирование комплексных оценок на основе нескольких учитываемых признаков (однородных или разнородных).** Такая группировка представляет собой весьма сложный процесс и нередко является самостоятельной задачей исследования. Для получения комплексной оценки используются, как правило, описательные порядковые признаки. Результат комплексной оценки обычно выражается в баллах.

Алгоритм комплексной оценки следующий:

1. Определить перечень учитываемых признаков, которые будут использованы в формировании комплексной оценки.
2. Каждому из возможных значений этих признаков экспертным путем присвоить оценку в баллах, учитывая направление увеличения признака.
3. Определить минимально возможное значение и максимально возможное значение комплексной оценки.
4. Интервал между минимальным значением и максимальным значением экспертным путем разделить на три диапазона: низких значений, средних значений и высоких значений.
5. Для каждой единицы наблюдения преобразовать фактические значения признаков в баллы.
6. Суммировать баллы по каждой единице наблюдения.
7. Сумму баллов преобразовать в значение диапазона.

Пример выполнения комплексной оценки представлен в таблице 1.

Таблица 1.

**Комплексная оценка выраженности табакокурения у студентов**

| Учитываемый признак и его значения                                                                                                                                   | Оценка<br>в баллах    | Студенты        |                 |               |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|-----------------|-----------------|---------------|
|                                                                                                                                                                      |                       | А               | В               | С             |
| 1. Возраст начала табакокурения:<br>✓ не курит<br>✓ до 10 лет,<br>✓ 10-14 лет,<br>✓ 15-17 лет,<br>✓ 18 лет и старше                                                  | 0<br>5<br>4<br>3<br>2 | 4               | 4               | 2             |
| 2. Частота табакокурения:<br>✓ не курит<br>✓ очень редко (несколько раз в год)<br>✓ редко (несколько раз в месяц)<br>✓ часто (несколько раз в неделю)<br>✓ ежедневно | 0<br>2<br>3<br>4<br>5 | 3               | 5               | 2             |
| 3. Количество сигарет, выкуриваемых за неделю:<br>✓ не курит<br>✓ менее одной<br>✓ 1-10 сигарет<br>✓ 10-20 сигарет<br>✓ Более 20 сигарет                             | 0<br>1<br>3<br>5<br>7 | 3               | 7               | 1             |
| Минимальное значение:<br>Максимальное значение:                                                                                                                      | 0<br>17               |                 |                 |               |
| Низкая выраженность табакокурения:<br>Средняя выраженность табакокурения:<br>Высокая выраженность табакокурения:                                                     | 0-5<br>6-11<br>12-17  |                 |                 |               |
| Результат комплексной оценки<br>выраженности табакокурения:                                                                                                          |                       | 10<br>(средняя) | 16<br>(высокая) | 5<br>(низкая) |

И еще несколько важных замечаний относительно группировок. При статистической разработке материала **необходимо учитывать существующие правила и стандарты определения границ групп**. Несоблюдение правил формирования группировок ведет к снижению ценности полученных результатов.

*Во-первых, объединение в стандартные группы обусловлено определенными мотивами, игнорирование которых может лишить научной содержательности все исследование.* Например, группировка людей по возрастным группам может производиться по-разному в исследованиях разной направленности:

- в клинических – с учетом физиологических особенностей развития организма человека,
- в медико-юридических – с учетом действующего трудового и пенсионного законодательства,
- в медико-экономических – в соответствии с практикой экономического анализа,
- в медико-демографических – в соответствии с практикой выполнения демографических исследований и т.п.

*Во-вторых, может оказаться невозможным сравнение данных, полученных вами, с данными других исследований.* Например, для общей оценки возрастного состава в демографической статистике часто применяются такие группы как «0-14 лет», «15-59 лет», «60 лет и старше». Если в вашем исследовании использовались группы «0-20 лет», «21-40 лет» и «старше 40 лет», то полученные вами данные сравнить с результатами других демографических исследований будет уже затруднительно.

Если для вашего исследования нет никаких строгих принципов группировки по возрасту, то **обычно используют пятилетние или десятилетние интервалы**:

- пятилетние интервалы: «до 20 лет», «20-24 года», «25-29 лет», «30-34 года» и т.д.
- десятилетные интервалы: «до 20 лет», «20-29 лет», «30-39 лет» и т.д.

Аналогичные пяти- или десятилетние интервалы группировок обычно используются при изучении стажа чего-либо (работы на данном производстве, состояния в браке, курения и пр.). Например: «до 5 лет», «5-9 лет», «10-14 лет», «15-19 лет» и т. д. (либо «до 10 лет», «10-19 лет», «20-29 лет» и т.д.)

Иногда границы соседних групп могут быть одинаковыми. В таких случаях интервал включает меньшую границу, но не включает большую. Например:

- дети до 7 лет (это дети в возрасте от 0 лет до 6 лет 11 месяцев 29 дней);
- дети от 7 до 15 лет (это дети в возрасте от 7 полных лет до 14 лет 11 месяцев 29 дней);
- подростки, от 15 до 18 лет (это лица в возрасте от 15 полных лет до 17 лет 11 месяцев 29 дней).

Группировка данных по своей сути представляет собой процесс классификации, т.е. установление принадлежности изученных единиц к определенным классам. В государственной (рутинной) статистике для этого используются различные **классификаторы** – специальные справочники, инструкции и указатели в виде алфавитных и систематических словарей, дополняемых стандартным перечнем объектов и их групп. Использование классификаторов в официальной статистике является обязательным. В научных исследованиях использование государственных и международных классификаторов состояний, болезней, причин смерти не является строго обязательным. Однако их применение создает возможность сопоставить статистические сведения о заболеваемости и смертности населения разных категорий на различных территориях.

После группировки выполняется сводка материала. Статистическая сводка позволяет получить обобщающую информацию обо всей изученной совокупности в целом. Эта работа представляет собой подсчет числа единиц в каждой из выделенных групп и определение групповых итогов (в абсолютных величинах). Сводную информацию

можно представить в форме статистических таблиц. В настоящее время сводка статистического материала обычно выполняется с помощью компьютерной техники.

### 3. СОСТАВЛЕНИЕ МАКЕТОВ СТАТИСТИЧЕСКИХ ТАБЛИЦ

Результаты научного исследования часто представляют в форме статистических таблиц. Технология их построения тесно связана с работой по группировке материала, поэтому виды таблиц и правила их оформления мы рассмотрим в рамках данной темы.

Возникновение статистики как науки тесно связано с появлением статистических таблиц. Статистические таблицы принципиально отличаются от перечневых таблиц. **Перечневые** или описательно-информационные таблицы представляют собой простой перечень данных. К таковым можно отнести, например, таблицу умножения, таблицы логарифмов, расписание движения поездов или автобусов и т. п. Данные, приводимые в **статистической** таблице, сгруппированы особым способом, что дает возможность не просто их систематизировать, но и проводить анализ.

На рисунке 1 представлена схема оформления статистической таблицы.

Рисунок 1

#### Схема оформления статистической таблицы

Номер таблицы

Название таблицы (единицы измерения данных)

| Наименование статистического подлежащего (верхний внутренний боковой заголовок) | Наименования статистических сказуемых (верхние внутренние подзаголовки) |                  |   |   |       |
|---------------------------------------------------------------------------------|-------------------------------------------------------------------------|------------------|---|---|-------|
|                                                                                 | А                                                                       | В                | С | Д | Итого |
| 1                                                                               | 2                                                                       | 3                | 4 | 5 | 6     |
| 1.                                                                              |                                                                         |                  |   |   |       |
| 2. Статистические                                                               |                                                                         | Ячейки с числами |   |   |       |
| 3. подлежащие (боковые                                                          |                                                                         |                  |   |   |       |
| 4. внутренние подзаголовки)                                                     |                                                                         |                  |   |   |       |
| 5.                                                                              |                                                                         |                  |   |   |       |
| Всего:                                                                          |                                                                         |                  |   |   |       |

Статистическую таблицу можно сравнить с грамматическим предложением – в ней так же есть подлежащее и сказуемое.

**Статистическое подлежащее** – это те признаки единиц наблюдения совокупности, которые характеризуются в таблице. Статистические подлежащие размещаются в левом боковом столбце таблицы и обычно являются *результативными признаками*.

**Статистическое сказуемое** – это характеристики статистического подлежащего. Сами характеристики представлены в форме чисел. Они располагаются в ячейках в строках, соответствующих тому или иному подлежащему. Сущность этих числовых характеристик (т.е. названия статистических сказуемых) располагаются в заголовках столбцов. Эти названия, как правило, являются *факторными признаками*.

*Макетом статистической таблицы* называется пустая (т.е. без чисел) таблица, показывающая расположение ее конструктивных элементов, содержащая названия всех строк и столбцов. Словом «*таблица*» обычно называют макет, заполненный числовыми сведениями.

Статистическая таблица должна быть по возможности небольшой, чтобы ее было легче анализировать. Лучше построить две-три небольшие таблицы, чем одну большую.

## Правила оформления статистических таблиц

Соблюдение этих правил позволяет представить полученные сведения в максимально удобной для читателя форме.

1. **Таблицы необходимо нумеровать.** Это делают с помощью надписи «Таблица №...», которую выравнивают по правому краю текста. Номер таблице присваивают для удобства последующего описания материала: чтобы обратить внимание читателя на какую-то закономерность, отраженную в таблице, не нужно писать в тексте ее длинное название, достаточно указать только ее номер. Нумерация таблиц в работе может быть сплошной (1, 2, 3, 4 и т.д.), а может отражать структуру работы: номера 1.1, 1.2, 1.3 и т.д. – для первой главы, номера 2.1, 2.2, 2.3 и т.д. – для второй главы и т.п.

В настоящее время можно встретить другой вариант оформления номера и названия таблицы: пишут «Таблица №...», далее ставят тире, после чего с заглавной буквы пишут название таблицы (точки не ставят). Этот заголовок выравнивают по центру строки. Например:

Таблица 1 – Название таблицы

2. **Общий заголовок (название таблицы)** должен лаконично отражать содержание таблицы. Обычно название таблицы начинается со слов «Уровни...», «Частота...», «Распространенность...», «Структура...», «Распределение...», «Динамика...». Если таблица приводится одна, вне текста или без пояснений в сопровождающем тексте, то обязательно указывается место и время отражаемого в ней явления. Для улучшения восприятия содержимого таблицы желательно указать в заголовке вывод, вытекающий из содержания таблицы. Например: заголовок «Заболеваемость населения К-ского района в 2010-2018 гг.» следует заменить на «Рост заболеваемости...». Если в таблицу включаются данные из официального источника, либо из опубликованной ранее научной работы, либо данные других авторов, то в названии следует указать источники получения информации. Кроме того в названии приводится расшифровка сокращений, используемых далее в таблице. Например: «Заболеваемость ишемической болезнью сердца (ИБС) населения города К. по данным 2018 года (‰)».

3. Если для всех ячеек таблицы (всех столбцов и всех строк) **единицы измерения** одинаковы, то они указываются в конце заголовка в круглых скобках («‰», «‰», «на 100 опрошенных», «на 1000 населения» и т.п.). Если единицы измерения в столбцах различаются, то они приводятся в названии каждого столбца. Если каждая строка таблицы имеет свою особую единицу измерения, то для их обозначения следует отводить отдельный столбец (следующий за статистическим подлежащим).

4. **Строки и столбцы таблицы желательно пронумеровать.** В левом боковом столбце перед названием каждого статистического подлежащего ставят числа (1, 2, 3 и т.д.). Чтобы пронумеровать столбцы, ниже заголовков столбцов выделяют специальную строку. Шрифт в ней обычно делают мельче, чем во всей таблице. Номера столбцов в этой строке выравнивают по центру ячейки. Если таблица большая, и на одной странице не помещается, то ее можно «разорвать» и разместить на нескольких страницах. В этом случае на каждой странице следует повторить шапку таблицы (названия столбцов и их номера). Нумерация строк и столбцов облегчает пользование таблицей, повышает удобство работы с ней. При необходимости сослаться на табличный материал в тексте можно указать только номер строки или номер столбца без их длинных названий, например: «Сравнение сведений, указанных в столбцах 4 и 7 таблицы 15, позволяет заметить, что...». Кроме того, нумерация строк и столбцов дает возможность показать способ расчета производных показателей. Это пишется в названии строки или в названии столбца. Например, строка содержит величины, которые являются суммой величин, указанных в нескольких выше расположенных строках. Тогда в названии этой строки после текста, в скобках будет написано «(строка 3 + строка 5 + строка 7)».

5. Не следует строить громоздкие таблицы с большим числом строк и столбцов. Иногда можно некоторые строки или столбцы **объединить под названием «Прочие».**

Однако не рекомендуется, чтобы эта группа охватывала более 10% всех единиц совокупности.

6. **Итоги в статистической таблице** имеют ключевое значение. Без них статистическая таблица не считается законченной. Отсутствие итогов является грубейшей методологической ошибкой. Ознакомление с данными таблицы следует начинать именно с итогов. **Вертикальные итоги (итоги по вертикали)** получаются в результате суммирования чисел столбцов, **горизонтальные итоги (итоги по горизонтали)** – суммирования чисел строк. Если таблица групповая или комбинационная (см. далее), то обязательно вычисляются промежуточные итоги в анализируемых подгруппах. Подведение итогов имеет особое значение для показателей структуры, поскольку позволяет увидеть, от каких величин ведется отсчет долей, что именно автор принял за 100%. Итоговые строки и столбцы часто обозначают словами «Итого» и «Всего». Эти названия по смыслу несколько различаются. Словом «Итого» обычно называют итог для некоторой части совокупности, словом «Всего» - итог для всей представленной в таблице совокупности. Однако рекомендуется по возможности избегать таких слов в названии итогов и заменять их на более информативные формулировки, например, «Все опрошенные», «Врачи всех специальностей», «Все заболевания» и т.п.

7. Строки подлежащего и столбцы сказуемого обычно располагаются **по принципу от частного к общему**, т. е. сначала показываются слагаемые, составные части, а в конце подлежащего или сказуемого подводят итоги. Если же приводятся не все слагаемые, а только выделяются наиболее важные из них, то поступают наоборот: сначала показывают общие итоги, затем пишут «в том числе», после чего выделяют наиболее важные составляющие части.

8. **Цифровые данные следует округлять единообразно** для всех ячеек столбца. При этом однотипные показатели приводятся с одинаковой точностью.

9. **Текстовые сокращения** в названиях строк и столбцов используются также единообразно, на основе общепринятых правил или стандартов. Как говорилось выше, расшифровка используемых сокращений может быть приведена в общем заголовке таблицы.

10. **Для выделения** особо важных показателей или итогов применяются различные технические приемы акцентирования (изменение шрифта, выделение строк и граф двойной или жирной линией).

11. **В статистической таблице не должно быть пустых ячеек.** Это очень важное правило, соблюдение которого позволяет автору таблицы показать ее законченность, завершенность. В каждой ячейке должно стоять либо число, либо определенный символ, который показывает причины отсутствия числа в ячейке.

- При *отсутствии сведений (невозможности их получить)* можно ставить три точки «...» или написать «н/д», т.е. «нет данных».
- При *невозможности появления признака у единиц наблюдения данного типа* (например, у женщин не может быть хронического простатита), ставится тире «—» или знак «X».
- Если *изучаемое явление наблюдается в очень малых размерах* и составляет менее половины последней значащей цифры в условиях принятой точности, в клетке выставляется ноль «0,0». Это говорит о том, что при меньшем округлении какая-либо значащая цифра может появиться.

12. **Можно сопровождать таблицу примечаниями и сносками.** Примечания относятся обычно ко всей таблице и содержат, как правило, сведения о методах получения исходной информации. Сноски относятся либо к отдельным клеткам, либо к графам или строкам таблицы. Отсутствие сноски или примечаний там, где они необходимы, может приводить к серьезным погрешностям в трактовке табличных данных.

По структуре выделяют:

- Простые таблицы
- Групповые таблицы
- Комбинационные таблицы

В **простой таблице** каждому статистическому подлежащему соответствует только одно статистическое сказуемое, т.е. представляется распределение только в одной группе, по одному признаку (рисунок 2).

Рисунок 2.

#### Макет простой таблицы

Таблица 1

*Распределение рабочих предприятия К. по кратности острых заболеваний в 2018 году (%)*

| Кратность заболеваний   | Удельный вес |
|-------------------------|--------------|
| 1                       | 2            |
| 1. Не болели            |              |
| 2. Болели 1 раз         |              |
| 3. Болели 2 раза        |              |
| 4. Болели 3 раза        |              |
| 5. Болели 4 и более раз |              |
| Всего:                  | 100,0        |

В данной таблице «Кратность заболеваний» – это общее название статистического подлежащего, название результативного признака. Значения этого признака указаны в первом столбце. Чтобы заполнить такую таблицу по каждой единице наблюдения (по каждому рабочему предприятия К.) нужно знать только один учитываемый признак – кратность острых заболеваний. С помощью простой таблицы можно выполнить самый примитивный анализ – в данном примере сравнить доли рабочих с разной кратностью заболеваний.

Теперь предположим, что у каждого рабочего зарегистрирован не только результативный признак «Кратность заболевания», но и два факторных признака «Пол» (значения «мужской» и «женский») и «Возраст». Количественный признак «Возраст» был сгруппирован, т.е. преобразован в описательный порядковый со значениями «до 44 лет» (по классификации ВОЗ – «молодой возраста») и «44 года и старше» (по той же классификации – «средний возраст»). В этом случае объем имеющейся у нас информации о здоровье рабочих предприятия К. увеличивается, и мы можем провести углубленный анализ заболеваемости. Выполнить сводку имеющихся сведений о заболеваемости рабочих разного пола и разного возраста нам помогут групповая и комбинационная таблицы.

В **групповой таблице** каждому статистическому подлежащему соответствует два или более сказуемых, не связанных между собой.

Рисунок 2

#### Макет групповой таблицы

Таблица 2

*Распределение рабочих предприятия К. по кратности острых заболеваний  
в 2018 году (%)*

| Кратность заболеваний                                                                              | Пол     |         | Возраст   |                  | Итого |
|----------------------------------------------------------------------------------------------------|---------|---------|-----------|------------------|-------|
|                                                                                                    | мужской | женский | до 44 лет | 44 года и старше |       |
| 1                                                                                                  | 2       | 3       | 4         | 5                | 6     |
| 1. Не болели<br>2. Болели 1 раз<br>3. Болели 2 раза<br>4. Болели 3 раза<br>5. Болели 4 и более раз |         |         |           |                  |       |
| Всего:                                                                                             | 100,0   | 100,0   | 100,0     | 100,0            | 100,0 |

В данном примере название таблицы не изменилось, т.к. не изменились ее содержание, место и время. В названиях столбцов появились заголовки статистических сказуемых, которые являются факторными признаками. Однако эти сказуемые между собой не связаны: все рабочие предприятия К. сначала были сгруппированы по признаку пола, а потом точно так же все были сгруппированы по признаку возраста. Групповая таблица уже позволяет сравнить кратность острых заболеваний у мужчин и женщин, у лиц молодого и среднего возраста. Может оказаться, что доля не болевших мужчин больше доли не болевших женщин, а доля болевших четыре и более раз в год среди лиц 44 лет и старше больше, чем среди лиц в возрасте до 44 лет.

Однако групповая таблица не сможет ответить на многие вопросы. Например, среди кого больше доля не болевших – среди мужчин в возрасте 44 лет и старше или среди женщин в возрасте до 44 лет? То есть для углубленного анализа явления часто бывает необходимо связать между собой несколько факторных признаков, например, пол и возраст. Надо поделить изучаемую совокупность на группы сначала по признаку пола, а затем каждую из образовавшихся групп поделить по признаку возраста. В этом случае нам придется построить **комбинационную таблицу**, в которой каждому статистическому подлежащему соответствует два или более статистических сказуемых, связанных друг с другом (рисунок 4).

Рисунок 4

#### Макет комбинационной таблицы

Таблица 3

*Распределение рабочих предприятия К. по кратности острых заболеваний  
в 2018 году (%)*

| Кратность заболеваний                                                                              | Пол и возраст рабочих |                  |                |           |                  |                |           |                  |                |
|----------------------------------------------------------------------------------------------------|-----------------------|------------------|----------------|-----------|------------------|----------------|-----------|------------------|----------------|
|                                                                                                    | Мужчины               |                  |                | Женщины   |                  |                | Оба пола  |                  |                |
|                                                                                                    | До 44 лет             | 44 года и старше | Всех возрастов | До 44 лет | 44 года и старше | Всех возрастов | До 44 лет | 44 года и старше | Всех возрастов |
| 1                                                                                                  | 2                     | 3                | 4              | 5         | 6                | 7              | 8         | 9                | 10             |
| 1. Не болели<br>2. Болели 1 раз<br>3. Болели 2 раза<br>4. Болели 3 раза<br>5. Болели 4 и более раз |                       |                  |                |           |                  |                |           |                  |                |
| Всего:                                                                                             | 100,0                 | 100,0            | 100,0          | 100,0     | 100,0            | 100,0          | 100,0     | 100,0            | 100,0          |

Существенным недостатком комбинационных таблиц является необходимость многократно делить на группы данные по изучаемой статистической совокупности. Поэтому чтобы обнаружить статистически значимые различия, совокупность должна быть достаточно большой.

#### 4. КОДИРОВАНИЕ ЗНАЧЕНИЙ УЧИТЫВАЕМЫХ ПРИЗНАКОВ И СОЗДАНИЕ КОМПЬЮТЕРНОЙ БАЗЫ ДАННЫХ

В настоящее время обработка и анализ статистического материала выполняется с применением компьютерной техники и соответствующего программного обеспечения. Для этих целей часто применяются табличные процессоры, например, Microsoft Excel. В программе MS Excel имеется специальный пакет функций, которые позволяют создавать обширные базы данных, производить из них отбор необходимой информации, проводить ее сводку в статистические таблицы, вычислять производные величины и другие характеристики, необходимые для статистического анализа. Познакомимся с тем, как создается база данных.

Информация в любой базе данных всегда располагается строго в определенном порядке. Каждый столбец электронной таблицы в базе данных называется **«ПОЛЕ»** (иногда говорят «Переменная»). Каждый столбец содержит информацию *о значениях конкретного учитываемого признака* у каждой единицы наблюдения. Каждая строка электронной таблицы в базе данных называется **«ЗАПИСЬ»** (иногда говорят «Случай»). Каждая строка содержит сведения *об одной единице наблюдения*, т.е. значения всех переменных, всех учитываемых признаков.

Ячейка базы данных может содержать:

- текст,
- дату,
- число,
- формулу.

В первой строке электронной таблицы обязательно указывают **наименования полей**, т.е. наименования учитываемых признаков. Последующие строки будут представлять собой записи об отдельных единицах наблюдения, например, испытуемых. Так, вторая строка будет являться записью о пациенте И.И. Иванове (т.е. будет содержать его значения учитываемых признаков), третья строка – записью о пациенте П.П. Петрове, четвертая строка – записью о пациенте Н.Н. Николаеве и т.д.

Если групп испытуемых несколько, то базу данных по каждой группе можно создать на отдельных листах MS Excel. Последний лист книги («книгами» называются файлы, создаваемые в MS Excel) посвящают формулам для расчетов. Можно создавать всю базу данных только на одном листе MS Excel. Но в этом случае должна существовать переменная «Номер группы».

Первая переменная (т.е. столбец А) – это всегда «Код единицы наблюдения» (или «Номер по порядку»). Следом за ней обычно указывают основные идентификационные признаки единицы наблюдения: «Ф.И.О.» (если эти сведения нужны), «Пол», «Возраст», «Диагноз» и др. Часто фамилию, имя и отчество испытуемого в базе данных не указывают в целях сохранения врачебной тайны.

Если значение признака у данной единицы наблюдения получить не удалось, то ячейку оставляют пустой. Отдельное поле (столбец) можно посвятить причинам отсутствия этой информации.

В каждой ячейке базы данных должно быть **только одно значение**. Не нужно его сопровождать никакими сопутствующими символами (например, содержимое ячейки «1,55 к+» будет неверным). Всю пояснительную информацию лучше завести в базу как отдельную переменную, выделив для этого отдельное поле.

Значение количественного признака вводится как число, которое было получено в результате измерения признака. Самостоятельно округлять количественные данные не

надо (различие в тысячных долях при округлении исчезнет, и результат вычислений может быть смещенным).

Каждая ячейка должна содержать только одно число (например, «8,2371»). Если чисел в ячейке будет несколько (например, «8,32; 7,19; 9,05» или «3, 4, 7, 8»), то программа прочитает такое содержимое не как число, а как текст, и исключит эту ячейку из математических расчетов.

Значения качественных (описательных) признаков для удобства ввода часто кодируются числами. При этом рекомендуется соблюдать следующие правила:

А) **Дихотомический (альтернативный) признак** кодируется «1» (присутствует) и «0» (отсутствует). Значение «1», как правило, присваивают негативному событию (возникновение заболевания, развитие осложнения, выход на инвалидность, смерть)

Например:

Летальный исход: да – 1, нет – 0

Септический шок: есть – 1, нет – 0

Б) **Полихотомический (номинативный) признак** кодируется произвольно.

Например:

- Пол: мужской – 1, женский – 2
- Место жительства: Ивановский район – 1, Палехский район – 2, Шуйский район – 3 и т.д.

В) **Порядковый (ординальный) признак** кодируется в направлении его нарастания.

Например:

- Состояние пациента: удовлетворительное – 1, средней тяжести – 2, тяжелое – 3.
- Эхогенность образования: низкая – 1, нормальная – 2, высокая – 3.

Если были проведены повторные измерения одного и того же признака, то для каждого из них заводят отдельное поле (столбец).

Например:

- «СОЭ 1» – СОЭ на момент поступления в стационар,
- «СОЭ 2» – повторное исследование СОЭ через неделю
- «СОЭ 3» – СОЭ на момент выписки

Если создается база данных по результатам анкетного опроса, то следует обращать внимание на то, какое количество пунктов может быть выбрано респондентом при ответе на каждый вопрос. Если вопрос в анкете предполагает только один ответ, то для его заведения в базу данных потребуется только одно поле (столбец). Если вопрос в анкете предполагает несколько вариантов ответа, то он должен быть описан несколькими переменными.

Например:

Вопрос: Как вы считаете, зачем нужно вести здоровый образ жизни?

1 – чтобы не болеть,

2 – чтобы жить долго,

3 – чтобы выглядеть красиво

4 – чтобы добиться успеха в жизни

5 – другое (напишите): \_\_\_\_\_

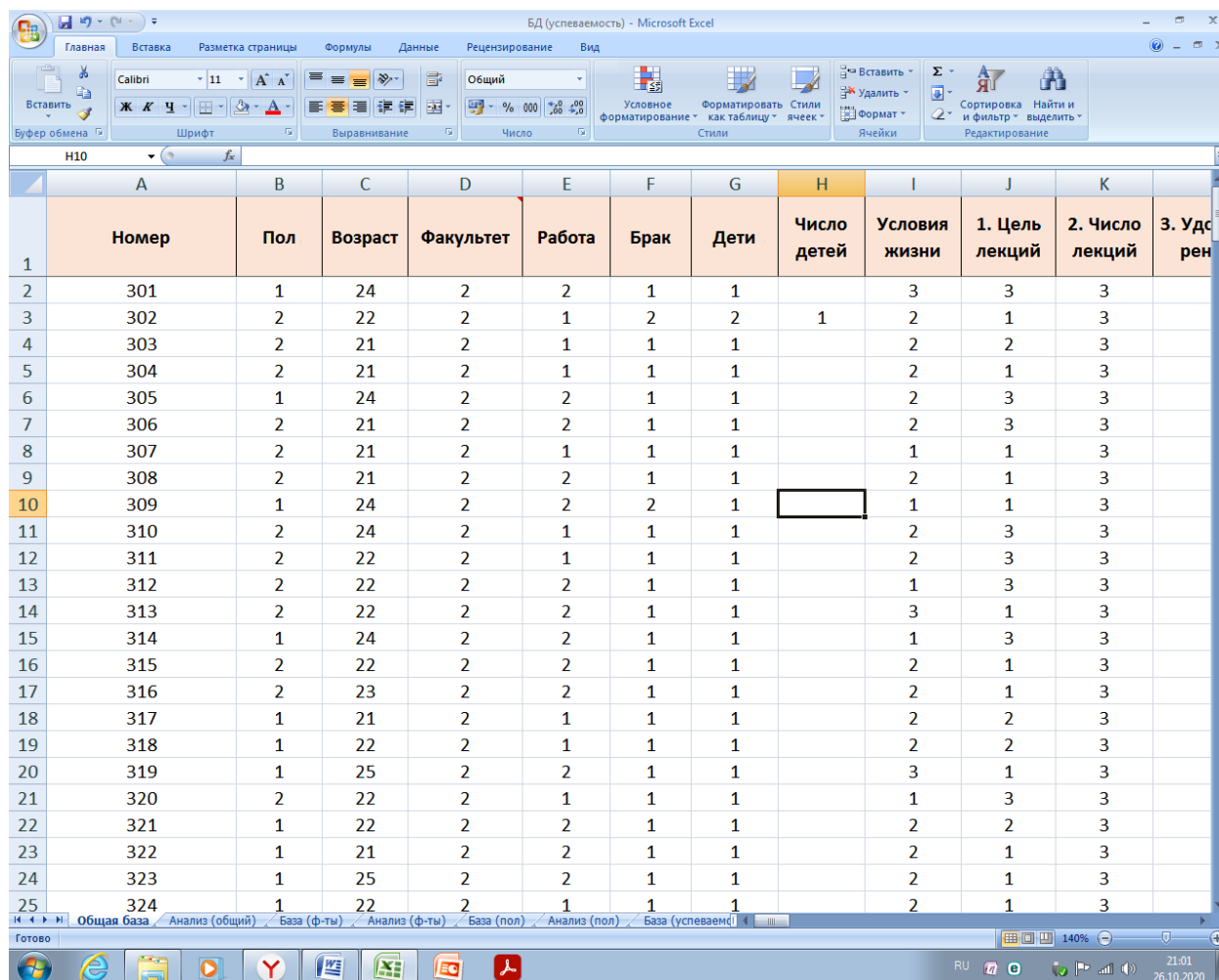
В данном случае надо выделить 5 полей (5 столбцов). В 4 поля будут занесены значения «0» (ответ не выбран) или «1» (ответ выбран). В 5-е поле будет занесен код «0», если ответ не выбран, либо другие коды, если респондент написал свой ответ. Эти ответы должны быть систематизированы, обобщены и закодированы как номинативный признак.

Для каждого актуального диагноза выделяют свою переменную («Диагноз 1», «Диагноз 2» и т.д.), в которую вводятся коды: «0» - нет этого заболевания, «1» - есть это заболевание. Если заболевания встречаются редко и неактуальны для исследования, то

они объединяются в одной переменной (например, «Болезни эндокринной системы»: имеются – «1», нет – «0»). Наоборот, если требуется точное описание какого-либо диагноза, вместо одной переменной «Диагноз» можно использовать несколько переменных, в каждой из которых кодируется та или иная часть диагноза.

Рисунок 5

### Макет базы данных, созданной в табличном процессоре Microsoft Excel



|    | А     | В   | С       | Д         | Е      | Ф    | Г    | Н           | И             | Ј              | К               |            |
|----|-------|-----|---------|-----------|--------|------|------|-------------|---------------|----------------|-----------------|------------|
|    | Номер | Пол | Возраст | Факультет | Работа | Брак | Дети | Число детей | Условия жизни | 1. Цель лекций | 2. Число лекций | 3. Удс рен |
| 1  |       |     |         |           |        |      |      |             |               |                |                 |            |
| 2  | 301   | 1   | 24      | 2         | 2      | 1    | 1    |             | 3             | 3              | 3               |            |
| 3  | 302   | 2   | 22      | 2         | 1      | 2    | 2    | 1           | 2             | 1              | 3               |            |
| 4  | 303   | 2   | 21      | 2         | 1      | 1    | 1    |             | 2             | 2              | 3               |            |
| 5  | 304   | 2   | 21      | 2         | 1      | 1    | 1    |             | 2             | 1              | 3               |            |
| 6  | 305   | 1   | 24      | 2         | 2      | 1    | 1    |             | 2             | 3              | 3               |            |
| 7  | 306   | 2   | 21      | 2         | 2      | 1    | 1    |             | 2             | 3              | 3               |            |
| 8  | 307   | 2   | 21      | 2         | 1      | 1    | 1    |             | 1             | 1              | 3               |            |
| 9  | 308   | 2   | 21      | 2         | 2      | 1    | 1    |             | 2             | 1              | 3               |            |
| 10 | 309   | 1   | 24      | 2         | 2      | 2    | 1    |             | 1             | 1              | 3               |            |
| 11 | 310   | 2   | 24      | 2         | 1      | 1    | 1    |             | 2             | 3              | 3               |            |
| 12 | 311   | 2   | 22      | 2         | 1      | 1    | 1    |             | 2             | 3              | 3               |            |
| 13 | 312   | 2   | 22      | 2         | 2      | 1    | 1    |             | 1             | 3              | 3               |            |
| 14 | 313   | 2   | 22      | 2         | 2      | 1    | 1    |             | 3             | 1              | 3               |            |
| 15 | 314   | 1   | 24      | 2         | 2      | 1    | 1    |             | 1             | 3              | 3               |            |
| 16 | 315   | 2   | 22      | 2         | 2      | 1    | 1    |             | 2             | 1              | 3               |            |
| 17 | 316   | 2   | 23      | 2         | 2      | 1    | 1    |             | 2             | 1              | 3               |            |
| 18 | 317   | 1   | 21      | 2         | 1      | 1    | 1    |             | 2             | 2              | 3               |            |
| 19 | 318   | 1   | 22      | 2         | 1      | 1    | 1    |             | 2             | 2              | 3               |            |
| 20 | 319   | 1   | 25      | 2         | 2      | 1    | 1    |             | 3             | 1              | 3               |            |
| 21 | 320   | 2   | 22      | 2         | 1      | 1    | 1    |             | 1             | 3              | 3               |            |
| 22 | 321   | 1   | 22      | 2         | 2      | 1    | 1    |             | 2             | 2              | 3               |            |
| 23 | 322   | 1   | 21      | 2         | 2      | 1    | 1    |             | 2             | 1              | 3               |            |
| 24 | 323   | 1   | 25      | 2         | 2      | 1    | 1    |             | 2             | 1              | 3               |            |
| 25 | 324   | 1   | 22      | 2         | 1      | 1    | 1    |             | 2             | 1              | 3               |            |