



Methods of systematization of quantitative information in historical research

O. D. Dashkovskaya¹

¹P. G. Demidov Yaroslavl State University, 14 Sovetskaya str., Yaroslavl 150003,
Russian Federation

DOI: 10.18255/1996-5648-2021-4-506-515

Research article
Full text in Russian

The article examines various methods of systematization of quantitative information used in historical research: summary, grouping, graphical representation of the results obtained. The classification and features of each method are considered. The stages of building groupings and ways to determine the number of groups are highlighted. The rules for constructing tables and the advantages of using graphs and diagrams are analyzed.

Keywords: systematization; summary; grouping; interval; factor; result; table; chart; cartogram

INFORMATION ABOUT AUTHORS

Dashkovskaya, Olga D. | E-mail: dashod@mail.ru
Cand. Sc. (History), Associate Professor



Методы систематизации количественной информации в исторических исследованиях

О. Д. Дашковская¹

¹Ярославский государственный университет им. П. Г. Демидова, ул. Советская, 14,
Ярославль, 150003, Российская Федерация

DOI: 10.18255/1996-5648-2021-4-506-515
УДК 930.24, 303.05

Научная статья
Полный текст на русском языке

В статье изучаются различные методы систематизации количественной информации, применяемые в исторических исследованиях: сводка, группировка, графическое отображение полученных результатов. Рассматриваются классификация и особенности каждого метода. Выделяются этапы построения группировок и способы определения числа групп. Анализируются правила построения таблиц и преимущества использования графиков и диаграмм.

Ключевые слова: систематизация; сводка; группировка; интервал; фактор; результат; таблица; диаграмма; картограмма

ИНФОРМАЦИЯ ОБ АВТОРАХ

Дашковская, Ольга Дмитриевна | E-mail: dashod@mail.ru
Доцент, кандидат исторических наук

Историки, занимающиеся изучением проблем социально-экономического развития, часто сталкиваются с необходимостью систематизации собранной статистической информации. Ученые осмысливают большие объемы числового материала, чаще всего представленного в виде статистических отчетов, но иногда это может быть и текстовое описание, содержащее количественные данные. Систематизация позволяет более глубоко исследовать закономерности исторического процесса или явления.

Цель данной работы – проанализировать существующие способы систематизации количественной информации, применяющиеся в исторических исследованиях. В большинстве сочинений приемы систематизации первичных данных представлены в виде сводки, группировки и/или графического оформления полученных результатов (построение таблиц и диаграмм).

Методом первичной систематизации количественных данных является сводка. Сводка способствует приведению в определенный порядок собранных статистических сведений, которые описывают отдельные единицы наблюдения. В результате сводки появляется возможность отразить всю генеральную совокупность (т. е. все имеющиеся в распоряжении исследователя данные) или ее часть.

В зависимости от глубины обработки имеющегося материала сводку можно классифицировать как простую или сложную. В первом случае подводятся общие итоги по всей совокупности единиц наблюдения. Сложная сводка одновременно содержит и группировку, а, значит, итоги вычисляются не только по всей генеральной совокупности, но и отдельно по выделенным группам. Результаты и простой и сложной сводки обычно представлены в виде таблиц.

По организационному признаку обычно различают централизованную и децентрализованную сводку. Централизованная сводка направлена на обработку всего собранного первичного материала в одном учреждении. Обычно она используется в ходе единовременных статистических исследований. При децентрализованной сводке обработка материала производится поэтапно. Например, в 1834 году во всех губернских городах Российской империи были созданы статистические комитеты, председателями которых являлись губернаторы. Губернские статистические комитеты занимались сбором первичных данных на местах, проверяли их, заносили в таблицы, создавали на основании этих материалов описание губернии в целом или анализировали отдельные отрасли промышленности, состояние торговли, сельского хозяйства. Подготовленные на местном уровне отчеты поступали «для дальнейших соображений» в Статистическое отделение Совета Министерства внутренних дел.

В зависимости от техники исполнения сводка может быть ручной и автоматизированной. При ручной сводке все итоги подсчитываются вручную, и именно с ней чаще всего сталкиваются историки. В настоящее время использование информационных технологий привело к широкому распространению автоматизированной сводки, при которой для обработки и подсчета данных используются электронно-вычислительные машины.

Итак, сводка дает представление о размерах и общих свойствах совокупности, но не позволяет качественно оценить явление, выявить исторические закономерности. Для этих целей статистика использует метод группировки, суть которого заключается в разбивке совокупности на части по признакам сходства или отличия. В результате образуются группы, различающиеся по качественному и/или количественному признаку.

Построение группировок в статистике проходит несколько последовательных этапов:

- 1) выбирается группировочный признак;
- 2) определяется количество групп, на которые будет разбита генеральная совокупность;

3) рассчитывается величина интервала группировки;

4) единицы наблюдения распределяются по группам, подводятся итоги как по выделенным типам, так и по всей генеральной совокупности.

Итак, на первом этапе определяется группировочный признак (основание группировки), по которому и происходит распределение единиц наблюдения по создаваемым группам.

В зависимости от применяемого критерия систематизации группировочные признаки классифицируют следующим образом:

– *по форме выражения* признаки подразделяют на атрибутивные (качественные) и количественные. Первые не отображаются с помощью количественных показателей и обозначаются словом (пол, семейное положение, профессия). Количественные признаки имеют числовое выражение (возраст, доход, объем выпускаемой продукции). При этом количественные признаки могут быть как дискретными, т. е. выражаться только целыми числами (количество крестьян, число промышленных предприятий), так и непрерывными, т. е. принимать как целые, так и дробные значения (размер земельных угодий, заработная плата);

– *в зависимости от характера колеблемости* признаки могут иметь множество значений или быть альтернативными. Последние имеют только два варианта значений, поэтому и получили такое наименование. Примером альтернативного признака может быть «пол», имеющий варианты «мужской» и «женский». Также это могут быть характеристики, которыми одни единицы генеральной совокупности обладают, а другие нет. При этом образуются две группы с противоположными признаками (например, семьи с детьми и бездетные);

– *по роли в группировке* признаки подразделяют на факторные и результативные. Факторные признаки оказывают воздействие на другие, а результативные изменяются под влиянием факторных признаков. Так, квалификация работника – это факторный признак, а его заработная плата – результативный. Такие признаки принимаются за основу аналитических группировок, которые подробнее будут охарактеризованы далее.

На втором этапе построения группировок определяется количество частей, на которые следует разделить изучаемую совокупность. Число образуемых групп зависит от многих показателей: цели исследования, особенностей изучаемого явления, объема статистической совокупности, наконец, от выбранного для классификации группировочного признака. Так, при группировке по атрибутивному признаку количество градаций будет соответствовать числу значений признака. Например, при распределении населения по семейному положению традиционно выделяют следующие типы: холост/не замужем, состоит в браке, вдовец/вдова, разведен/разведена.

При определении числа групп в совокупности, основанием для деления которой является количественный признак, число образуемых частей может быть очень разнообразно. Во многом оно зависит от цели работы. Так,

по признаку «доход» можно выделить группы с зарплатой населения ниже и выше средней по стране, построить интервальный ряд с различными диапазонами доходности. При небольшом объеме генеральной совокупности не нужно формировать много групп, так как они будут нерепрезентативными и не позволят объективно оценить явление. При значительном объеме классифицируемого материала необходимо учитывать количество единиц наблюдения и размах вариации группировочного признака: чем он больше, тем большее количество частей обычно выделяют. Только при соблюдении данных правил обобщающие характеристики каждой группы будут устойчивыми и закономерными.

Определение числа групп тесно взаимосвязано с третьим этапом построения группировки, на котором проводится расчет величины (ширины) интервала. Вполне понятно, что при большой величине интервала получается меньше групп, и наоборот. В зависимости от ширины интервалы могут быть равными и неравными. Неравные интервалы используются, если это определяется изначально атрибутивным признаком. Так, возрастные периоды жизни человека (молодость, средний возраст, пожилой человек) вне зависимости от существующих споров об их верхних и нижних границах, представлены диапазонами неодинаковой ширины. Также неравные интервалы применяются в экономических исследованиях и образуются в тех случаях, когда группировочный признак изменяется неравномерно или в больших пределах, является количественным, а не атрибутивным.

Группировка с равными интервалами строится при не очень значительной вариации признака в совокупности.

Величина равного интервала вычисляется по формуле:

$$K = (X_{\max} - X_{\min})/n,$$

где X_{\max} и X_{\min} – наибольшее и наименьшее значения признака; n – число групп.

В ситуации, когда необходимо вычислить не только величину интервала, но и определить количество групп, применяется формула, предложенная американским статистиком Г. Стерджессом [1, с. 312]:

$$K = (X_{\max} - X_{\min}) / (1 + 3,2 \lg N),$$

K – величина интервала;

X_{\max} – наибольшее значение признака;

X_{\min} – наименьшее значение признака;

N – число элементов совокупности.

Например, на фабрике трудятся 100 рабочих со стажем работы от 5 до 40 лет. Найдем величину интервала:

$$K = (40 - 5) / (1 + 3,2 \lg 100) = 35 / 7,4 \approx 4,7.$$

Итак, ширина интервала составляет $\approx 4,7$ а полученная группировка выглядит следующим образом:

Стаж	5–9,7	9,7–14,4	14,4–19,1	...
------	-------	----------	-----------	-----

При образовании интервалов необходимо точно определять их границы. Если признаки дискретные и выражаются только целыми числами, то границы интервалов устанавливаются без совпадения крайних показателей смежных интервалов. Один показатель не может находиться одновременно в двух интервальных группировках. Например, при классификации предприятия по числу рабочих группировка может выглядеть следующим образом: 1–100, 101–200, 201–300...

Для непрерывных признаков применяется обязательное совпадение смежных границ интервалов. В примере выше мы это и наблюдаем. Но в какую группу будет отнесено значение, которое является границей интервала, например 14,4? В данном случае в статистике руководствуются следующим правилом: если единица имеет значение признака, равное величине верхней границы интервала, то ее следует причислять к следующему интервалу. Таким образом, значение 14,4 относится к третьему интервалу.

На четвертом этапе построения группировок происходит распределение единиц совокупности по группам и подведение итогов по группам и совокупности в целом.

Обычно историки применяют три вида группировок: типологические, структурные, аналитические [4, с. 20].

Типологические группировки способствуют выявлению однородных групп в качественно разнородной совокупности. Например, это деление предприятий по формам собственности на государственные, частные и кооперативные. При этом требуется очень тщательно подбирать качественный признак, который и закладывается в основу группировки.

Структурные группировки направлены на изучение состава и структуры генеральной совокупности. В их основе находится количественный признак (деление населения по возрасту, уровню заработной платы, группировка заводов по численности рабочих).

Аналитические (факторные) группировки способствуют выявлению взаимосвязи между явлениями. При этом один из группировочных признаков является результатом, другой – фактором. Группировка происходит по факторному признаку. Если при возрастании значения факторного признака увеличивается значение результативного, то связь между признаками считается прямой, если же значение результативного фактора убывает при росте факторного, то – обратной. Примером прямой связи является общий трудовой стаж и время работы сотрудника на каком-либо предприятии.

В зависимости от поставленных задач и степени сложности изучаемого явления единицы совокупности можно систематизировать как по одному, так и по нескольким группировочным характеристикам. Если группы сформированы по одному признаку, то группировка считается простой.

Если же разделение происходит по двум и более признакам, то такая группировка называется комбинационной. Обычно одновременно сочетают не более двух-трех признаков, так как из-за излишней детализации группировка может потерять свое аналитическое значение.

Пример комбинационной группировки по двум признакам (пол и возраст) приведен в таблице 1.

Таблица 1

**Половозрастная структура рабочих Европейской России
и Бакинской губернии в 1901 г. [3, с. III]**

Распределение рабочих по возрасту	Количество рабочих (чел.)		
	мужского пола	женского пола	оба пола
малолетние	20308	13738	34046
подростки	90435	55742	146177
взрослые	1127261	383872	1511133
итого	1238004	453352	1691356

Наконец, существует и еще одна классификация группировок – деление их на первичные и вторичные. Первичные группировки производятся на основе еще не обработанного статистического материала, при этом требуется выбор группировочного признака, определение числа групп и величины интервалов. Но историкам часто приходится иметь дело со вторичными группировками (уже готовыми таблицами) и образовывать новые группы без обращения к первичным данным, применять перегруппировку на основе изменения имеющихся интервалов.

Особым видом группировки считается классификация, под которой подразумевается устойчивое разграничение объектов, каждый из которых может быть отнесен лишь к одной группе (например, классификации форм собственности, основных отраслей экономики). Классификация всегда основывается на существенных признаках и используется в течение длительного времени. Таким образом, классификация – это общепринятая, нормативная группировка.

Результаты сводки и группировки материалов можно представить в различной форме: в виде текстового описания, статистических таблиц и графиков. Наиболее часто цифровое выражение итоговой характеристики оформляется в виде таблицы, достоинствами которых является их наглядность и компактность. К статистической таблице можно отнести не каждую. Если таблица не содержит результатов подсчетов (таблица умножения, опросные листы), она не причисляется к исследуемой категории.

По внешнему виду статистическая таблица представляет комбинацию вертикальных и горизонтальных строк, содержащую верхние и боковые заголовки. Подлежащее и сказуемое являются основными элементами та-

блицы. Подлежащее представляет собой объект изучения, оно показывает, о чем идет речь в таблице. Это могут быть социальные группы населения, виды предприятий, территории. Обычно подлежащее находится в левой части таблицы и составляет содержание строк. В таблице 1 заголовком подлежащего является «Распределение рабочих по возрасту», содержанием подлежащего – наименование строк («малолетние», «подростки», «взрослые»). Совокупность числовых показателей, характеризующих изучаемый объект, называется сказуемым таблицы. Заголовки сказуемого в таблице 1 – это заголовки столбцов «Количество рабочих (чел.)», «мужского пола», «женского пола», а его содержание – цифры, при помощи которых характеризуются выделенные в подлежащем группы.

Статистические таблицы, используемые историком при обработке «чернового» материала, могут быть самыми разнообразными и иногда занимают несколько страниц. Но при публикации результатов исследования следует соблюдать определенные правила в оформлении таблиц:

1. Каждая таблица должна иметь заголовок, отражающий ее основное содержание, место и время, к которым относятся приводимые данные. В заголовке можно указывать единицу измерения, если она одна. Над верхним правым углом заголовка таблицы размещается надпись «Таблица...» с указанием ее номера: Таблица 1 и т. д.;

2. После заголовка таблицы должна находиться сноска, указывающая, на основании каких данных составлена таблица или из какого источника информации она взята;

3. Заголовки подлежащего и сказуемого следует формулировать точно и лаконично. Здесь же при необходимости указываются единицы измерения;

4. Таблицы не должны быть очень громоздкими и содержать много признаков, иначе снижается их наглядность;

5. Значения одних и тех же показателей следует заносить в таблицу с одинаковой степенью точности (например, все числа округляются до десятых);

6. Раньше при отсутствии отдельных данных историки ставили многоточие или прочерк в соответствующей ячейке таблицы. Сейчас этого делать не рекомендуется, так как большинство таблиц составляется и обрабатывается с помощью программы MS Excel. Чтобы результаты были подсчитаны автоматически, лучше оставлять такие ячейки пустыми;

7. Итоги должны быть рассчитаны по подгруппам, группам и в целом, что также нашло отражение в приведенной ранее таблице 1.

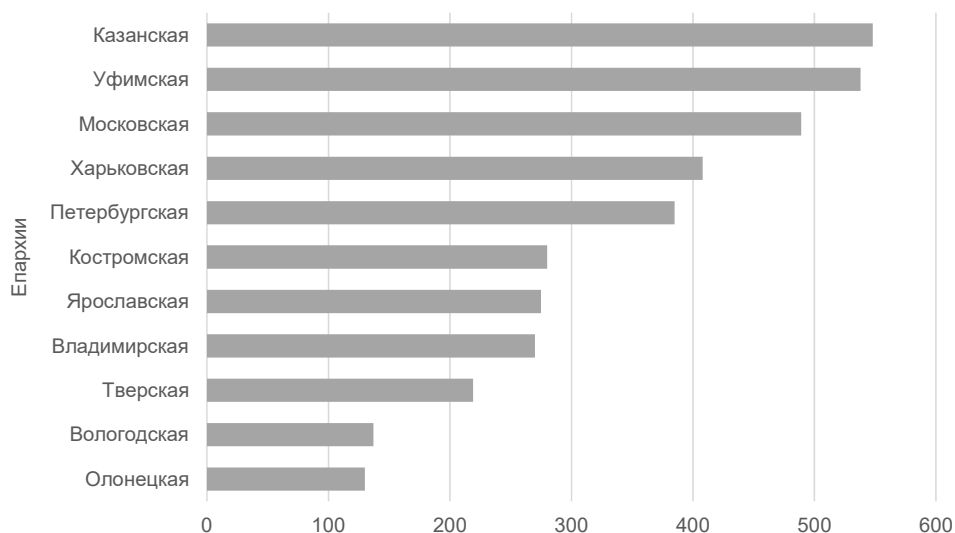
Более наглядными методами систематизации данных являются формы графического изображения, представленные графиками. Они логически продолжают статистические таблицы в процессе обобщения и анализа информации. Графиком называется чертеж, на котором статистические совокупности, характеризующиеся определенными показателями, описываются с помощью условных геометрических образов

или знаков. В зависимости от способа построения графики подразделяются на диаграммы, картограммы и картодиаграммы.

Диаграмма представляет чертеж, на котором информация отображается с помощью геометрических фигур или символических знаков. Наиболее распространенными видами диаграмм являются линейные графики, гистограммы и круговые диаграммы.

Линейные диаграммы чаще всего применяются для изображения изменения процесса во времени и при изучении связи между явлениями. В системе координат на горизонтальной оси X откладываются показатели времени, а на вертикальной оси Y – величины изучаемого показателя. После этого наносят данные и, соединив точки, получают статистическую кривую. Преимуществом линейных графиков является то, что в одном поле можно отобразить несколько показателей, что упрощает сравнение.

Динамика социально-экономических явлений может быть представлена в виде гистограмм. Другое их наименование, столбиковые диаграммы, связано с принципом построения: каждое значение показателя отображается в виде вертикального прямоугольника (столбика). На рис. 1 приведен пример гистограммы, показывающей размер дохода служителей церкви



(причта) от платы за обряды (крещение, венчание, бракосочетание и пр.) в разных епархиях Российской империи в конце XIX в.

Рис. 1. Средний годовой доход причта от платы за церковные обряды в 1880 г. [2, с. 8–9]

Гистограмма показывает, что наименьший доход от совершения обрядов получали служители церкви в малонаселенных северных епархиях. В Москве и столице плата за требы была традиционно выше, чем в провинции. Подобные столбиковые диаграммы очень удобны для визуального сравнения различных показателей. Также их можно использовать для изу-

чения структуры явления и динамики различных процессов, так как гистограмма обладает большими аналитическими характеристиками по сравнению с линейными графиками.

Круговые диаграммы также находят широкое применение в исторических исследованиях. Площадь окружности в них – это вся статистическая совокупность, а площади отдельных секторов отражают удельный вес (долю) ее составных частей. При этом оптимально, чтобы количество секторов было не очень большим, а подписи к ним (категории, доли) располагались рядом с сегментами с помощью «линий выноски», а не на отдельно выделенной легенде. В противном случае снижается наглядность диаграммы.

Интересной формой графического изображения статистических материалов являются картограммы, на которых штриховкой различной густоты, точками или окраской различной степени насыщенности показана сравнительная интенсивность какого-либо показателя в пределах нанесенного территориального деления. На картодиаграмме карта дополняется различными социально-экономическими диаграммами, например показателями развития промышленности и сельского хозяйства в регионах.

Итак, в ходе научной деятельности историки могут использовать различные методы систематизации количественной информации. Построение сводок и группировок тесно взаимосвязано, часто группировка является логическим продолжением сводки, но при этом и сама завершается в итоге сводными показателями. Сводки и группировки оформляются в виде таблиц, которые содержат точные количественные данные, но не всегда удобны для восприятия. Диаграммы строятся на основе табличных данных, они более наглядны, но в то же время не так точны, демонстрируют общие тенденции, без детализации. Часто методы систематизации дополняют друг друга (например, табличные данные одновременно визуализируются диаграммами), и выбор того или иного метода, безусловно, зависит от целей исследования. Наконец, следует отметить, что в последние два десятилетия построение таблиц и диаграмм производится не вручную, а с помощью специального программного обеспечения. Общеизвестной и интуитивно понятной разработкой является редактор электронных таблиц MS Excel, значительно упростивший работу историков с количественными данными.

Ссылки

1. Мазур Л. Н. Методы исторического исследования: учебное пособие. 2-е изд. Екатеринбург: Изд-во Урал. ун-та, 2010. 608 с.
2. О мерах к улучшению материального обеспечения православного сельского духовенства. СПб.: Синодальная типография, 1880. 37 с.
3. Свод отчетов фабричных инспекторов за 1901 г. СПб.: Типография Киришбаума, 1903. 202 с.
4. Федорова Н. А. Математические методы в историческом исследовании: курс лекций. Казань: Форт-Диалог, 1996. 108 с.