

ЛАБОРАТОРНАЯ РАБОТА №1

ГРУППИРОВКА И ПРЕДСТАВЛЕНИЕ СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ

1. ЦЕЛЬ И ЗАДАЧИ ИССЛЕДОВАНИЯ

Изучение и приобретение навыков группировки и отображения статистических данных.

2. ОСНОВЫ ТЕОРИИ

СТАТИСТИЧЕСКАЯ СВОДКА И ГРУППИРОВКА

Статистическая сводка представляет собой комплекс последовательных операций по обобщению конкретных единичных фактов, образующих совокупность, для выявления типичных черт и закономерностей, присущих изучаемому явлению в целом. Простой сводкой называется операция по подсчету общих итогов по совокупности единиц наблюдения. Сложная сводка представляет собой комплекс операций, включающих группировку единиц наблюдения, подсчет итогов по каждой группе и по всему объекту и представление результатов группировки и сводки в виде статистических таблиц.

Основные единицы статистической совокупности объединяются в группы при помощи метода группировки. *Группировка* – расчленение множества единиц совокупности на группы по определенным признакам; объединение изучаемых единиц в частные совокупности по существенным для них признакам.

Задачи статистической группировки. Виды группировок

Выделяют следующие задачи, решаемые с помощью метода статистических группировок:

- 1) образование социально - экономических типов явлений;
- 2) изучение строения изучаемых явлений и структурных изменений, происходящих в них;
- 3) выявление связи между изучаемыми признаками.

Для решения этих задач соответственно применяют типологические, структурные и аналитические группировки.

В ходе *типологической* группировки исследуемая качественно-разнородная совокупность делится на социально - экономические типы. При использовании метода типологических группировок важное значение имеет правильный выбор группировочного признака. При атрибутивном признаке число групп определяется свойствами изучаемого явления: группировка населения по странам проживания, коммерческих предприятий по формам собственности и т.д.

Выделение типов на основе количественного признака состоит в определении групп с учетом значений изучаемых признаков. При этом очень важно правильно установить интервал группировки, на основе которого количественно различаются одни группы от других.

В *структурной* группировке производится расчленение однородной в качественном отношении совокупности единиц на группы, характеризующие строение совокупности, ее структуру, т.е. по варьирующему признаку. Например, семьи можно группировать по числу их членов и в динамике проследить тенденции изменения удельного веса семей разной величины в их общей совокупности.

В *аналитической* группировке изучается функциональная связь между отдельными

признаками явления. Если предполагается, что признак В зависит от признака А (А - причина, В - следствие), то признак А называется факторным, признак В - результативным.

Есть большое разнообразие взаимосвязей между признаками. Из них можно выделить следующие:

1) когда факторным выступает количественный признак, а результативным - качественный. Например, стаж работы и квалификация работника;

2) когда факторный признак - качественный, а результативный - количественный. Например, место работы и зарплата;

3) когда факторный и результативный признаки - качественные. Например, образование и выполняемая работа;

4) когда факторный и результативный признаки - количественные. Например, производительность труда и заработная плата.

Комбинационной называется группировка, в которой расчленение совокупности производится по двум и более признакам, взятым в сочетании. При построении комбинационной группировки, как правило, сначала производят группировку по атрибутивным признакам, значения которых имеют ярко выраженные качественные различия.

Выбор группировочного признака. Образование групп и интервалов

При группировке встает вопрос выбора признака, который соответствует цели исследования и характеру исходной информации. Из множества признаков необходимо выбрать определяющие, т.е. те, которые наиболее полно и точно характеризуют изучаемый объект. Например, для торгового предприятия определяющим признаком является объем товарооборота.

После определения группировочного признака важным шагом является распределение единиц совокупности по группам. Встает вопрос о количестве групп и величине интервала изменения признака в каждой группе. Количество групп во многом зависит от того, какой признак служит основанием группировки. Часто атрибутивные (не имеющие количественного значения) признаки предопределяют число групп (группировка работников по образованию). Количественные признаки, принимающие только целочисленным значением, причем в небольшом диапазоне (например, число членов семьи, число секций в магазине) также предопределяют число групп.

Интервалы групп устанавливаются при значительном колебании дискретного признака (число работников) и при непрерывно изменяющемся количественном признаке (возраст). Под величиной интервала понимают разность между максимальными и минимальными значениями признака в каждой группе. Правильное установление величины интервала имеет первостепенное значение для образования качественно однородных групп. Нельзя объединять в одну группу явления, которые относятся к разным частным совокупностям. Например, при статистическом исследовании производственных предприятий не следует включать в одну группу предприятия, выполнившие обязательство на 95 - 105 %, нужно образовывать 2 группы 95 - 100 %, 101 - 105 %.

В зависимости от характера распределения статистической совокупности устанавливаются равные или неравные интервалы. Число групп тесно связано с объемом совокупности. Нет строго научных приемов, позволяющих определить число групп, эта задача всякий раз решается с учетом конкретных обстоятельств. Однако при равенстве интервалов для ориентировки используется формула Стерджесса, с помощью которой можно наметить число групп при известной численности совокупности N:

$$n = 1 + 3,322 * \lg N, \quad (1.1)$$

где **n** - число групп.

Зная размах колебания значений изучаемого признака во всей совокупности и

намечаемое число групп, величину равного интервала i можно определить по формуле

$$i = \frac{R}{n} = \frac{X_{\max} - X_{\min}}{n}, \quad (1.2)$$

где $R = X_{\max} - X_{\min}$ - размах.

В экономической практике в основном применяются неравные интервалы, последовательно возрастающие или убывающие. Например, будет неправильным применять равновеликий интервал для группировки по доходам. Нужны интервалы более короткие для малообеспеченных и более крупные для обеспеченных слоев населения.

Для лучшей характеристики изучаемого явления (первоначальная группировка не позволяет выявить характер распределения единиц совокупности), либо для приведения к сопоставимому виду группировок с целью проведения сравнительного анализа применяется *вторичная группировка*.

Вторичная группировка предполагает образование новых групп на основе ранее осуществленной группировки. В одном случае осуществляется изменение (укрупнение) первоначальных интервалов, в другом – образование новых групп на основе закрепления за каждой группой определенной доли единиц совокупности.

Статистические ряды распределения

Результаты сводки и группировки материалов статистического наблюдения оформляются в виде статистических рядов распределения и таблиц. *Статистические ряды распределения* представляют собой упорядоченное расположение единиц изучаемой совокупности по группировочному признаку.

Ряды распределения, образованные по качественным признакам, называют *атрибутивными*. Пример такого ряда приведен в табл. Если такую группировку составить за два периода по данному вузу, то можно выяснить структурные изменения и качественные сдвиги в составе преподавателей.

Таблица

Пример атрибутивного ряда

Категория преподавателей	Число преподавателей	В % к итогу
Стажеры	20	5
Ассистенты	100	25
Старшие преподаватели	40	10
Доценты	160	40
Профессора	40	10
Итого	400	100

При группировке ряда по количественному признаку получают вариационные ряды. Вариационные ряды бывают дискретными, основанными на дискретном признаке, и интервальными, основанными на непрерывно изменяющемся значении признака. Применяются также интервальные ряды распределения. При их построении возникают вопросы о числе групп, величине интервала, его границе.

Дискретные вариационные ряды. Пусть $a_1, \dots, a_k$ - возможные значения дискретного признака, при этом $a_1 < a_2 < \dots < a_k$. Исследуемая статистическая совокупность содержит статистические единицы. В этой совокупности n_1 единиц имеют значение признака a_1 , n_2 единиц - a_2 , ..., n_k единиц - a_k . Числа $n_1, \dots, n_k$ называются абсолютными частотами.

Если N - общее количество единиц совокупности, то

$$0 < n_j < n, N = \sum_{j=1}^k n_j.$$

Обычно при статистическом исследовании вместо абсолютных рассматриваются относительные частоты:

$$f_j = \frac{n_j}{N} \cdot 100\%.$$

Для них выполняются соотношения:

$$0 \leq f_j \leq 100\%, \sum_{j=1}^k f_j = 100\%.$$

Интервальные вариационные ряды. Пусть исследуемый признак может принимать значения из непрерывного промежутка $[a, b]$. Произведем группировку данных. Выберем интервалы $(a..a_1), (a_1..a_2), \dots, (a_{n-1}, b)$. Обычно заранее решают, к какой именно группе относить границы интервалов $a_1, a_2, \dots, a_{n-1}$ (например, пусть интервал включает свою верхнюю границу, тогда получаем интервалы $[a..a_1], [a_1..a_2], \dots, [a_{n-1}, b]$). Подсчитывается число объектов, попавших в каждый из интервалов, и рассчитываются относительные частоты.

Интервалы группировок выделяют из логических соображений. Иногда используются "принцип равных интервалов", о котором говорилось выше, и "принцип равных частот", по которому интервалы выбираются таким образом, чтобы относительные частоты были примерно одинаковы.

ИЗУЧЕНИЕ ЗАКОНОМЕРНОСТЕЙ НА ОСНОВЕ РЯДОВ РАСПРЕДЕЛЕНИЯ

Закономерности распределений и задачи их анализа

Частоты в вариационных рядах изменяются закономерно в соответствии с изменением варьирующего признака. Такие закономерности изменения частот в вариационных рядах называются *закономерностями распределения*. Одна из важных целей статистического изучения вариационных рядов состоит в том, чтобы выявить закономерность распределения и определить ее характер. Закономерности распределения выражают свойства явлений, общие условия, влияющие на формирование вариации признака.

Так, например, распределение предприятий области по формам собственности отражает условия, определяющие развитие тех или иных форм на данном этапе. Эти условия выражаются, в частности, в степени разработанности законодательной базы, структуре экономики и ее состоянии в отдельных отраслях, социальном составе населения, уровне квалификации занятых и т.д. В меньшей степени для данного процесса характерно действие случайных причин, хотя полностью абстрагироваться от их проявления невозможно.

Когда мы говорим о характере, типе закономерностей распределения, то имеем в виду отражение в них общих условий, определяющих распределение. При этом следует учитывать, что речь идет о распределениях, отражающих однородные явления. Если рассматривать вместе распределение предприятий по формам собственности в областях с различными социально-экономическими условиями, то сильно уменьшается действие общих факторов, определяющих его характер. В этом случае распределение может не проявиться отчетливо или проявиться как бимодальное (двухвершинное) или даже многовершинное.

Основная задача анализа вариационных рядов - выявление подлинной закономерности распределения путем исключения влияния второстепенных, случайных для данного распределения факторов - достигается увеличением объема исследуемой совокупности при одновременном уменьшении интервала ряда.

Графическое представление рядов распределений

Анализ рядов распределения удобно проводить на основе их графического изображения. При этом для изображения интервального ряда используется полигон, гистограмма, огива и кумулята распределения.

Полигон используется для изображения дискретных вариационных рядов. Для его построения в прямоугольной системе координат по оси абсцисс в одинаковом масштабе откладываются ранжированные значения варьирующего признака, а по оси ординат наносится шкала для выражения величины частот (рис.).

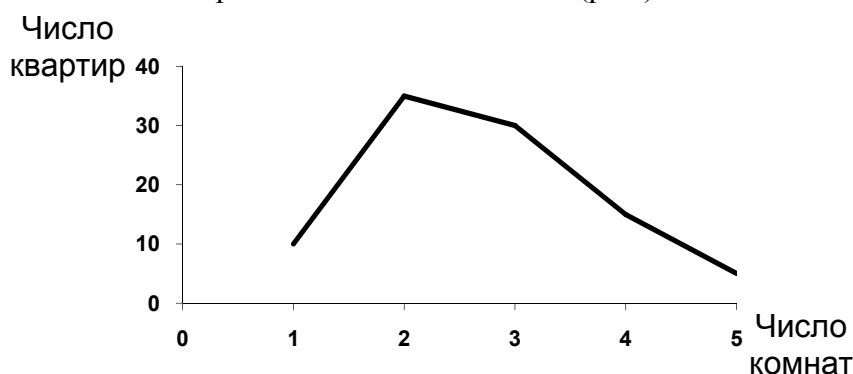


Рис. Полигон распределения жилого фонда по типу квартир

Гистограмма применяется для изображения интервального вариационного ряда. При построении диаграммы на оси абсцисс откладываются величины интервалов ($c_1, c_2, \dots, c_n$), а частоты изображаются прямоугольниками, построенными на соответствующих интервалах. Высота прямоугольников в случае равных интервалов должна быть пропорциональна частотам (f_i/c_i). Т.к. площадь каждого прямоугольника равна относительной частоте интервала, то общая площадь гистограммы равна 1 (100 %). Линия, ограничивающая гистограмму сверху, называется распределением частот.

Гистограмма может быть преобразована в полигон следующим образом: находятся середины сторон прямоугольников и полученные точки соединяются прямыми линиями (рис.).

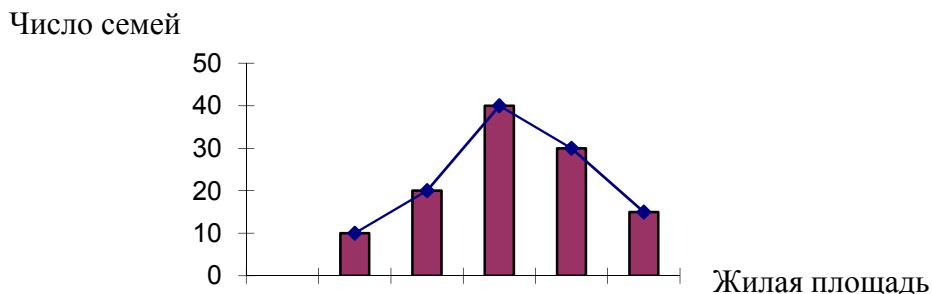
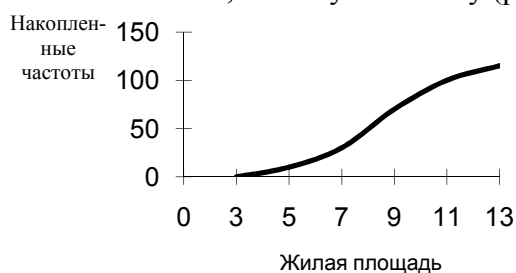


Рис. Гистограмма распределения семей по размеру жилой площади, приходящейся на одного человека

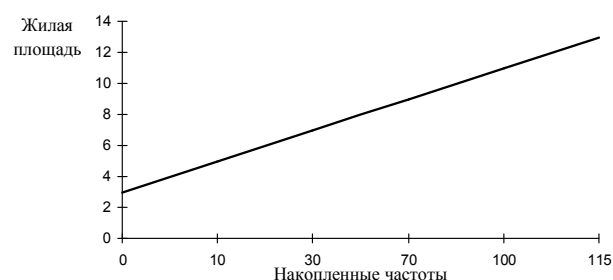
Кумулята изображает ряд накопленных частот. При построении кумуляты интервального ряда по оси абсцисс откладываются варианты ряда, а по оси ординат

накопленные частоты, которые наносят на поле графика в виде перпендикуляров к оси абсцисс в верхних границах интервалов. Затем эти перпендикуляры соединяют и получают ломаную линию – кумуляту (рис. а).

Если при графическом изображении вариационного ряда в виде кумуляты оси поменять местами, то получим огиву (рис.б).



а)



б)

Рис. Кумулята (а) и огива (б) распределения семей по размеру жилой площади, приходящейся на одного человека

ПРЕДСТАВЛЕНИЕ СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ

Построение статистических таблиц

Результаты сводки и группировки материалов наблюдения могут быть представлены в виде статистических таблиц. Это наиболее рациональная форма представления результатов статистической сводки. *Статистическая таблица* – это форма краткого и рационального изложения сводной числовой характеристики исследуемой совокупности по одному или нескольким существующим признакам, взаимосвязанным логикой экономического анализа.

Таблицей называется такая форма представления числовой информации, при которой число располагается на пересечении четко сформулированного заголовка по вертикальному столбцу (графу) и названия по соответствующей горизонтальной полосе (строке). Внешне таблица представляет собой пересечение граф и строк, которые формируют остов таблицы. Составленную таблицу, но не заполненную цифрами, называют *макетом таблицы* или *пустографиком*, в котором мысленно определяются в деталях цель обследования и объем разработки материалов сводки.

Статистическая таблица содержит три вида заголовков: общий, верхние, боковые. Общий заголовок отражает содержание всей таблицы, располагается над макетом таблицы по центру и является внешним заголовком. Верхние заголовки характеризуют содержание граф (заголовки сказуемого), боковые – строк (заголовки подлежащего) и являются внутренними заголовками.

Каждая статистическая таблица содержит подлежащее и сказуемое. *Подлежащим* таблицы называется объект, отдельные единицы или группы, которые характеризуются соответствующими показателями. *Сказуемое* таблицы – это показатели, с помощью которых изучается подлежащее таблицы. В основном в сказуемом отражаются численные значения и характеристики изучаемого явления. Подлежащее таблицы обычно составляет название ее строк и располагается в левой части таблицы. Сказуемое формирует заголовки граф (колонок) и составляет их содержание. Иногда, в целях получения более компактной таблицы подлежащее и сказуемое меняют местами, т.е. подлежащее указывают по графам, а сказуемое – по строкам.

Сказуемое статистических таблиц бывает простым и сложным. Показатели *простого* сказуемого располагаются последовательно один за другим. Распределяя показатели на группы по одному или нескольким признакам в определенном сочетании, получают сложное сказуемое. Содержание и характер информации таблицы определяются

не столько числом показателей сказуемого, сколько правильной их комбинацией и расположением.

Соблюдение правил и последовательности работы со статистическими таблицами помогает исследователю осуществить научно обоснованный экономико-статистический анализ объектов и процессов.

Графическое отражение информации

Выразительность, доходчивость, лаконичность, универсальность, обозримость графических изображений сделали их незаменимыми в исследовательской работе и в сравнениях и сопоставлениях экономических явлений.

Графическое изображение обобщает исходную информацию и позволяет осуществить контроль достоверности статистических показателей, т.к. представленные графически они наиболее ярко отражают неточности, обусловленные либо наличием ошибок наблюдения, либо сущностью изучаемого явления. С помощью графика возможны изучение закономерностей развития явления, установление существующих взаимосвязей, выявление причинных связей. Графики широко используются для изучения структуры явлений, их изменения во времени и размещения в пространстве. В них выразительно проявляются сравниваемые характеристики и отчетливо видны основные тенденции развития и взаимосвязи, присущие изучаемому явлению или процессу. График должен быть достаточно наглядным, выразительным, доходчивым и понятным.

Статистический график – это чертеж, на котором статистические совокупности, характеризующиеся определенными показателями, описываются с помощью условных геометрических образов и знаков. Статистические графики классифицируются по форме применяемых графических образов, способу построения и характеру решаемых задач. Особым видом статистических графиков считаются диаграммы распределения: гистограмма, полигон и кумулята.

В статистическом графике различают следующие основные элементы: поле графика, графический образ, пространственные и масштабные ориентиры, экспликация графика.

Поле графика – это часть плоскости, на которой он выполняется (листы бумаги, географические карты, планы местности и т.п.). Поле графика характеризуется его форматом – размерами и пропорциями сторон, а размер поля графика зависит от его назначения. Стороны поля графика обычно находятся в определенной пропорции. Принято считать, что наиболее оптимальным для зрительного восприятия является график, выполненный на поле прямоугольной формы с соотношением сторон от 1:1,3 до 1:1,5. Иногда используется и квадратное поле графика, еще реже круговое.

Графический образ – это символические знаки, с помощью которых изображаются статистические данные (линии, точки, плоские геометрические фигуры – прямоугольники, квадраты, круги и т.д.). В качестве графического образа выступают и объемные фигуры. Иногда в графиках используются негеометрические фигуры в виде силуэтов или рисунков предметов.

Пространственные ориентиры определяют размещение графических образов на поле графика. Они задаются координатной сеткой или контурными линиями и делят поле графика на части, соответствующие значениям изучаемых показателей. В статистических графиках чаще всего применяется система прямоугольных координат, но могут быть построены графики в полярных координатах. Они необходимы для наглядного изображения циклического движения во времени.

Масштабные ориентиры графика придают графическим образам количественную значимость, которая задается с помощью масштаба и системы масштабных шкал. Масштаб графика – это мера перевода численной величины в графическую (например, 1 см соответствует 1000 \$.).

Масштабной шкалой является линия, отдельные точки которой читаются (в соответствии с принятым масштабом) как определенные числа. В масштабной шкале различают: линию – носитель информации, являющуюся опорой шкалы; помеченные на ней черточками точки; цифровые обозначения чисел, соответствующие отдельным точкам. Как правило, цифровое обозначение имеют не все помеченные точки, а лишь некоторые из них, расположенные в определенном порядке, и его необходимо помещать строго против соответствующих точек.

Шкала графика может быть прямолинейной и криволинейной (например, окружность, содержащая 360). Различаются шкалы равномерные и неравномерные. Шкала является *равномерной*, если равным графическим отрезкам соответствуют равные числовые величины. Когда равным числовым интервалам соответствуют неравные графические и наоборот, то шкала называется *неравномерной* (например, логарифмическая шкала).

Классификация статистических графиков

Статистические графики классифицируются по способу построения, форме применяемых графических образов, характеру решаемых задач.

По характеру решаемых задач различают диаграммы сравнения, динамики и структурные (рис.). Особым видом считаются диаграммы распределения, к которым относятся гистограмма, полигон и кумулята.

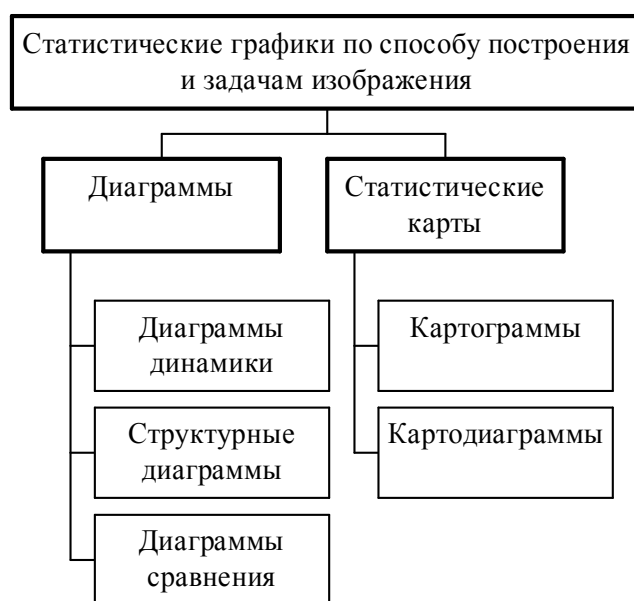


Рис. Классификация графиков по способу построения и задачам изображения

В зависимости от формы применяемых графических образов статистические графики могут быть линейными, плоскостными и объемными (рис.).

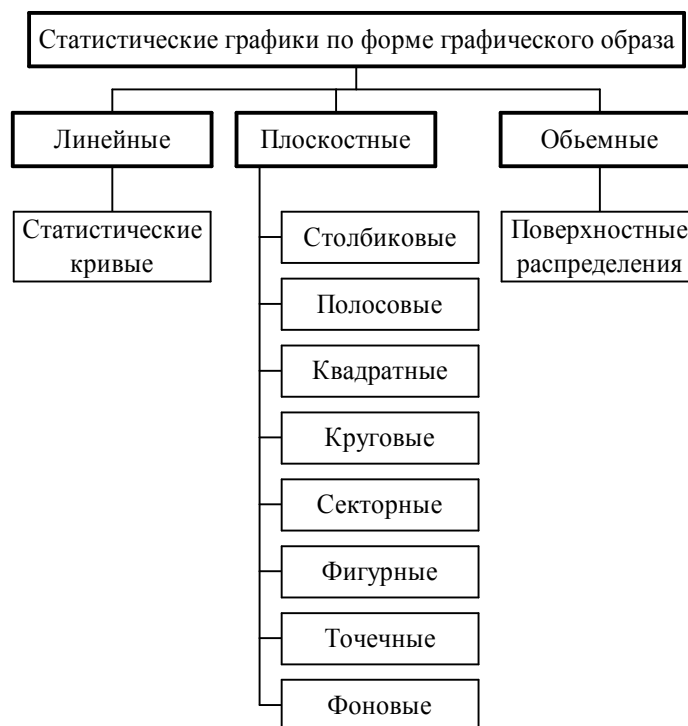


Рис. Классификация графиков по форме графического образа

Основные типы диаграмм

Наибольшее распространение получили линейные диаграммы. Для их построения обычно применяется система прямоугольных координат. На оси абсцисс откладываются величины факторного признака или показатели времени, а по оси ординат - величины результативного признака. По точкам на обеих осях координат определяется положение каждого уровня на поле графика. Последовательно соединяя точки отрезками линий, получают эмпирическую линию графика, так называемую статистическую кривую (рис.).

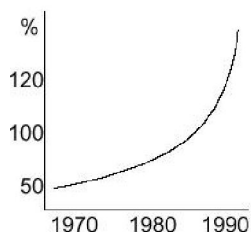


Рис. Статистическая кривая

При построении линейной диаграммы важно правильно выбрать масштаб как по оси абсцисс, так и по оси ординат. От оптимального соотношения этих масштабов зависит наглядность статистического графика и возможные выводы, которые можно будет сделать из него. Существует несколько способов более существенного увеличения линии на графике:

- перенесение исходной точки графика ближе к началу координат;
- отсутствие числовых значений по оси Y;
- увеличение масштаба по оси Y;
- тенденциозный подбор данных, когда опускаются данные, выпадающие из общей структуры.

Для исключения зрительных иллюзий, искажающих реальное представление о действительном развитии явления, целесообразно строить координатную сетку, при которой формат поля графика выбирается в соотношении от 1:1,3 до 1:1,5.

Важным достоинством линейных графиков является то, что на одном и том же поле можно изобразить несколько показателей. При этом каждую кривую надо изображать

отдельной формой линии (сплошная, пунктирная и т.д.) или окрашивать разными цветами.

Координатную сетку, в которой по оси абсцисс нанесена шкала в равномерном масштабе, принято называть арифметической. Линейные графики иногда также строятся с логарифмической шкалой по оси ординат.

Разновидностью линейных диаграмм являются радиальные диаграммы, которые строятся в полярных координатах. Радиальные диаграммы изображают определенные ритмические движения во времени и иллюстрируют сезонные колебания. Обычно в основе радиальных диаграмм лежат повторяющиеся годовые циклы с помесечными или поквартальными данными. При изучении годового цикла с помесечными данными окружность делят радиусами на 12 равных частей. Каждому радиусу дается название месяца года, а располагают их подобно циферблату часов. На каждом радиусе в соответствии с установленным масштабом наносятся точки, соответствующие изучаемым за каждый месяц данным. В результате получается спиралеобразная линия, характеризующая внутригодовые циклы.

Различают кольцевые (замкнутые) диаграммы (рис.), отражающие внутригодовой цикл какого-либо одного года, и спиральные (рис.), отражающие внутригодовой цикл динамики за ряд лет.

При построении кольцевых диаграмм точка на радиусе, определяющем значение за декабрь текущего года, соединяется с точкой на радиусе, определяющей значение за январь этого же года.

При построении спиральных диаграмм точка на радиусе, определяющем значение за декабрь текущего года, соединяется с точкой на радиусе, определяющем значение за январь следующего года. Это и дает возможность изобразить весь ряд динамики в виде спирали.

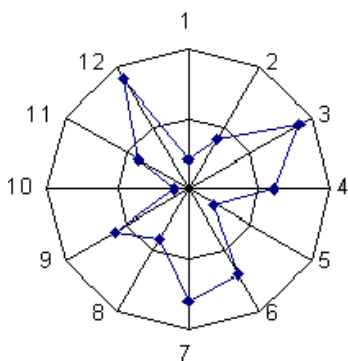


Рис. Кольцевая диаграмма

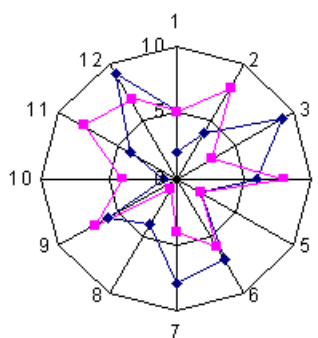


Рис. Спиральная диаграмма

Широкое применение в статистике находят диаграммы, позволяющие отразить структуру совокупности, ее изменение и динамику численности данной совокупности (рис.). Для построения таких диаграмм ряд абсолютных показателей совокупности заменяется рядом относительных величин. Наиболее распространенным видом таких диаграмм являются секторные диаграммы. В этих диаграммах площадь окружности принимается за величину всей изучаемой статистической совокупности, а площади отдельных секторов отображают удельный вес (долю) ее составных частей. Так как площади секторов пропорциональны центральным углам, то сумма всех углов (360 градусов) распределяется пропорционально удельным весам отдельных частей изучаемой совокупности.

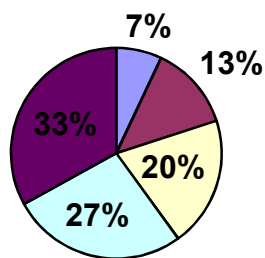


Рис. Секторная диаграмма:

■ - сыр; ■ - творог; ■ - масло; ■ - сметана; ■ - молоко.

К недостаткам секторной диаграммы следует отнести малую эффективность при большом числе частей совокупности и снижение наглядности при незначительных изменениях структуры изображаемых совокупностей.

Другим видом диаграмм, отображающих структуру совокупности, являются столбиковые и полосовые диаграммы, в качестве графического образа в которых используются прямоугольники (рис.). Каждая из полос диаграммы будет иметь одинаковую длину, так как при переходе к относительным величинам погашаются различия в абсолютных размерах совокупностей.

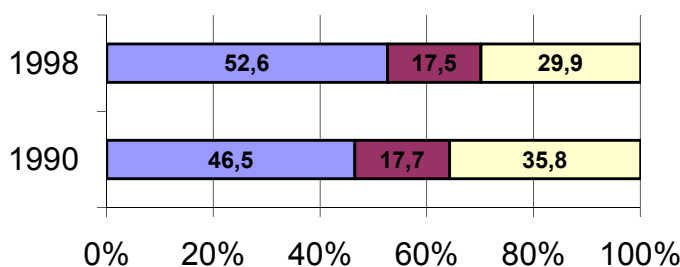


Рис. Полосовая (линейчатая) диаграмма:

■ - будильник; ■ - настенные часы; ■ - наручные часы.

Наглядным изображением статистической информации являются столбиковые диаграммы (рис.). При построении столбиковых диаграмм также используется прямоугольная система координат. Каждое значение изучаемого показателя изображается в виде вертикального столбика. По оси абсцисс размещается основание столбиков. Их ширина может быть произвольной, но обязательно одинаковой для каждого столбика. Высота столбиков (в соответствии с принятым по оси ординат масштабом) должна строго соответствовать изображаемым данным.

Количество столбцов определяется числом данных. Расстояние между столбиками должно быть одинаковым. У основания столбиков дается название изучаемого показателя. Значения показателей помещаются внутри каждого столбика.

В статистике находят применение и так называемые ленточные (полосовые) графики (рис.). В этих диаграммах основания располагаются вертикально, а масштабная шкала наносится на горизонтальную ось. Ленточная диаграмма представляет собой ряд простирающихся по оси абсцисс лент одинаковой ширины. При построении ленточных диаграмм соблюдаются те же требования, что и при построении столбиковых графиков (одинаковая ширина полос, начало шкалы от нулевой отметки и др.).

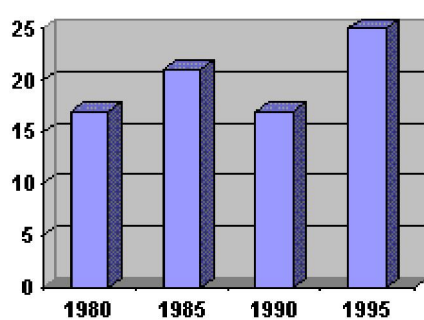


Рис. Столбиковая диаграмма

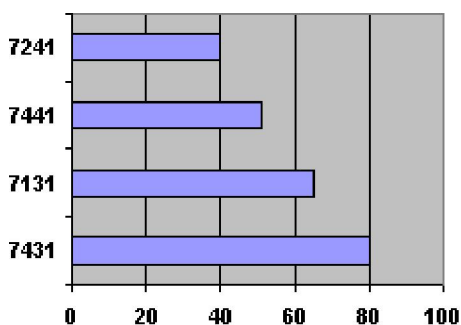


Рис. Полосовая диаграмма

Линейные, столбиковые и полосовые диаграммы имеют широкое применение в изображении статистической информации. При этом линейным и столбиковым диаграммам отдается предпочтение, если в изучаемых показателях проявляется общая тенденция роста.

Горизонтальные полосы нагляднее, если изображаемые показатели отображают результат функционирования того или иного процесса.

Для простого сравнения независимых друг от друга показателей могут использоваться диаграммы, принцип построения которых основан на том, что сравниваемые величины изображаются в виде правильных геометрических фигур (квадрат, круг, прямоугольник), площадь которых характеризует величину изучаемого признака. Эти диаграммы получили названия квадратных, круговых и знаков Варзара (рис.).

Для построения квадратных и круговых диаграмм необходимо из сравниваемых величин извлечь квадратный корень. Затем на основе полученных результатов определить сторону квадрата или радиус круга соответственно принятому масштабу.

Квадраты или круги следует располагать на одинаковом друг от друга расстоянии. Также допускается построение таких диаграмм путем расположения один в другом квадратов или кругов и использования различной штриховки или закрашки фигур. В каждой фигуре следует указывать числовое значение, которое она изображает, не приводя масштаба измерения.

Для графического изображения статистических показателей применяются и так называемые знаки Варзара. В этих диаграммах используются прямоугольники для графического изображения тех показателей, один из которых является произведением двух других. В каждом таком прямоугольнике основание пропорционально одному из показателей-сомножителей, а высота его соответствует второму показателю-сомножителю. Площадь прямоугольника соответствует величине третьего показателя, являющегося произведением двух первых. Располагая рядом несколько прямоугольников, относящихся к разным показателям, можно сравнивать не только размеры показателя - произведения, но и значения показателей-сомножителей.

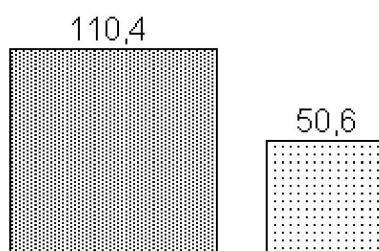


Рис. Квадратная диаграмма



Рис. Круговая диаграмма

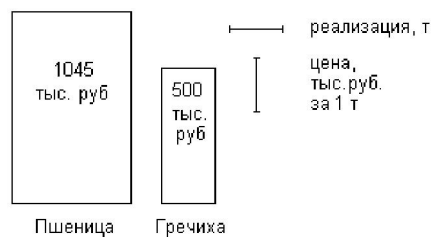


Рис. Знаки Варзара

В статистике, прежде всего для рекламных целей, применяются фигурные диаграммы. При их построении статистические данные изображаются рисунками, т.е. символами, которые в наибольшей степени соответствуют существу отображаемых явлений. Эти диаграммы более выразительны, зрительно легко воспринимаются.

3. ЗАДАНИЕ

Используя средства OpenOffice.Calc решить следующие задачи

1. По данным выборочного обследования рабочих строительного треста в соответствии с вариантом 1-30 (данные см. ниже)
 - 1.1 По данным ряда 1 построить дискретный ряд распределения, полигон, кумуляту; сформулировать вывод.
 - 1.2 По данным ряда 2 построить интервальный ряд распределения (количество групп определить на основе формулы Стерджесса), гистограмму, кумуляту; сформулировать вывод.
 - 1.3 Для выявления зависимости между рядом 1 и рядом 2 построить аналитическую группировку. Сформулировать вывод.
 - 1.4 Отобразите графически данные, представленные в ряду 1 и 2. Тип диаграммы выберите самостоятельно.
2. По данным, характеризующим денежные доходы населения в Российской Федерации по месяцам за 2015 - 2017 гг. в расчете на душу населения в долларах США, постройте диаграммы динамики (столбиковые, спиральные).

Годы \	2015	2016	2017
Январь	4,3	20,5	64,8
Февраль	7,3	23,7	72,8
Март	11,2	26,3	79,8
Апрель	12,7	26,7	88,4
Май	16,3	26,4	82,6
Июнь	21,4	30,6	93,0
Июль	25,8	39,0	99,3
Август	24,2	50,1	104,3
Сентябрь	21,5	54,3	101,4
Октябрь	17,1	54,5	86,2
Ноябрь	17,2	65,9	89,4
Декабрь	26,8	93,8	114,6

3. Имеются данные о посевной площади, валовом сборе и урожайности отдельных зерновых культур (во всех категориях хозяйств):

	2000 г.	2010 г.
--	---------	---------

Валовой сбор зерновых культур, млн. т	186,8	140,1
в том числе:		
овес	14,2	12,5
просо	2,10	1,13
гречиха	1,08	0,49
Урожайность зерновых культур, ц с 1 га	15,6	10,9
в том числе:		
овес	15,3	10,3
просо	7,8	3,9
гречиха	5,7	3,3
Посевная площадь под зерновыми культурами, млн. Га	119,3	127,9
в том числе:		
овес	9,2	12,1
просо	2,7	2,8
гречиха	1,9	1,5

Изобразите приведенные в таблице данные при помощи диаграмм структуры, динамики и сравнения. Отобразите данные с помощью знаков Варзара. Самостоятельно определите, по каким показателям какие диаграммы построить.

ВАРИАНТЫ ЗАДАНИЙ

Вариант	Ряд 1		Ряд 2	
	признак	номера	признак	номера
1.	Стаж работы	1-40	Зарплата за январь	1-40
2.	Тарифный разряд	1-60	Возраст рабочих	1-60
3.	Число членов семьи	1-80	% выполнения норм	1-80
4.	Стаж работы	21-60	Возраст рабочих	21-60
5.	Тарифный разряд	21-80	% выполнения норм	21-80
6.	Число членов семьи	21-100	Средний тариф.разряд	21-100
7.	Стаж работы	41-80	Возраст рабочих	41-80
8.	Тарифный разряд	41-100	% выполнения норм	41-100
9.	Число членов семьи	41-120	Средний тариф.разряд	41-120
10.	Стаж работы	51-100	% выполнения норм	51-100
11.	Тарифный разряд	61-100	Средний тариф.разряд	61-100
12.	Число членов семьи	71-140	Зарплата за январь	71-140
13.	Стаж работы	71-120	% выполнения норм	71-120
14.	Тарифный разряд	81-140	Средний тариф.разряд	81-140
15.	Число членов семьи	91-160	Возраст рабочих	91-160
16.	Стаж работы	101-140	Средний тариф.разряд	101-140
17.	Тарифный разряд	101-150	Зарплата за январь	101-150
18.	Число членов семьи	111-160	Возраст рабочих	111-160
19.	Стаж работы	121-160	Средний тариф.разряд	121-160
20.	Тарифный разряд	121-170	Возраст рабочих	121-170
21.	Число членов семьи	121-180	% выполнения норм	121-180
22.	Стаж работы	131-180	Зарплата за январь	131-180
23.	Тарифный разряд	141-190	Возраст рабочих	141-190
24.	Число членов семьи	151-200	% выполнения норм	151-200
25.	Стаж работы	131-200	Возраст рабочих	131-200
26.	Тарифный разряд	141-200	% выполнения норм	141-200
27.	Число членов семьи	161-200	Средний тариф.разряд	161-200
28.	Стаж работы	11-80	Возраст рабочих	11-80
29.	Тарифный разряд	11-100	% выполнения норм	11-100
30.	Число членов семьи	11-60	Средний тариф.разряд	11-60

ИСХОДНЫЕ ДАННЫЕ ДЛЯ ВЫПОЛНЕНИЯ ЗАДАНИЙ ПО ВАРИАНТАМ

Данные обследования рабочих строительного треста, принятых на работу за последние 10 лет.

Порядковый номер рабочего	Стаж работы на данном предприятии	Тарифный разряд	Профессия, полученная в ПТУ	Средний тарифный разряд работ в январе	Зарплата за январь (р.)	% выполнения норм в январе	Возраст рабочих (в годах)	Число членов семьи
1	2	3	4	5	6	7	8	9
1	5	5	каменщик	5,0	27600	186	25	1
2	3	4	монтажник	4,7	20100	181	22	2
3	4	4	каменщик	3,8	24400	190	34	4
4	5	3	монтажник	4,8	25300	181	28	2
5	3	4	монтажник	4,6	23600	181	22	1
6	3	4	монтажник	4,7	27100	200	40	4
7	5	4	каменщик	4,2	21000	176	38	3
8	5	5	монтажник	5,2	21600	140	32	2
9	3	4	монтажник	4,5	17800	144	30	2
10	5	5	монтажник	6,0	21300	136	23	2
11	4	5	бетонщик	4,9	24000	164	25	2
12	5	3	монтажник	4,0	21800	170	25	2
13	1	3	каменщик	3,5	22000	169	27	3
14	2	4	монтажник	5,0	21600	155	23	2
15	2	1	монтажник	4,0	19400	158	26	1
16	2	3	монтажник	4,0	17400	156	28	2
17	4	4	монтажник	3,6	20900	168	29	2
18	1	4	монтажник	4,9	16300	115	20	1
19	2	4	монтажник	3,7	14900	116	20	1
20	3	3	монтажник	4,1	16000	119	22	1
21	2	3	монтажник	4,3	16200	119	21	1
22	2	4	монтажник	4,3	17300	125	22	1
23	2	4	монтажник	4,0	16100	127	22	1
24	3	4	монтажник	5,0	18700	128	23	2
25	1	4	плотник	4,0	16300	129	19	1
26	1	4	плотник	4,5	16800	129	25	1
27	4	5	плотник	5,0	21000	140	28	2
28	1	4	бетонщик	4,2	15100	113	22	1
29	2	4	бетонщик	4,0	13700	110	20	1
30	2	4	бетонщик	3,8	14200	112	21	1
31	1	3	плотник	3,9	14500	109	20	1
32	3	4	плотник	4,9	18000	132	23	1
33	5	4	монтажник	4,9	18720	135	26	3
34	3	1	плотник	4,8	17440	132	25	1
35	5	4	монтажник	4,8	20100	132	24	3
36	3	3	плотник	4,9	18170	132	21	1
37	2	5	монтажник	4,4	20260	132	22	1
38	1	4	бетонщик	4,2	14130	113	19	1
39	3	4	бетонщик	5,0	14540	100	24	1
40	1	4	бетонщик	4,0	10470	80	20	1
41	3	3	монтажник	4,1	14680	104	22	2
42	3	3	плотик	4,1	13800	104	21	1
43	1	4	плотник	5,0	15400	107	21	1
44	4	5	плотник	4,0	27040	230	24	1
45	2	3	каменщик	3,0	11680	96	2:	1
46	5	4	монтажник	4,8	25910	190	36	5
47	3	4	монтажник	4,9	30220	232	32	4
48	2	5	монтажник	4,8	30440	195	31	3
49	3	3	каменщик	5,0	27550	186	36	4
50	4	4	монтажник	4,5	23330	180	28	3
51	3	5	Плотник	3,2	13510	105	36	2
52	1	3	каменщик	4,2	12380	90	20	1
53	5	4	монтажник	5,0	28990	203	33	3
54	5	3	каменщик	4,3	29940	153	35	2
55	5	4	монтажник	4,0	20400	158	34	4
56	2	4	монтажник	4,0	19080	146	22	1
57	3	4	бетонщик	4,0	13080	100	24	1
58	3	4	бетонщик	3,9	16120	125	23	2
59	4	4	монтажник	5,0	22200	159	26	2
60	2	3	плотник	4,1	17110	123	27	1

61	4	4	монтажник	4,6	20670	159	33	4
62	3	4	бетонщик	5,5	16180	100	24	1
63	3	5	каменщик	4,9	27840	190	26	3
64	1	4	каменщик	4,1	27350	199	27	3
65	3	4	монтажник	4,8	19650	139	26	1
66	2	5	монтажник	4,8	27020	190	38	4
67	1	5	монтажник	5,2	13840	83	20	1
68	1	3	бетонщик	4,0	12380	98	19	1
69	2	4	плотник	4,1	18000	139	20	2
70	1	4	монтажник	4,8	10920	85	20	1
71	4	5	монтажник	4,8	20240	146	30	3
72	5	4	монтажник	4,7	25240	183	31	4
73	3	4	бетонщик	3,8	12300	90	21	1
74	3	4	монтажник	4,8	17660	146	25	2
75	2	3	бетонщик	3,5	12080	95	20	1
76	1	4	плотник	4,9	14440	110	19	1
77	1	4	плотник	4,9	15880	113	19	1
78	1	4	монтажник	4,9	13550	113	22	2
79	2	4	монтажник	4,0	14440	111	24	2
80	1	4	каменщик	3,8	14440	110	19	1
81	1	4	монтажник	4,2	16550	128	19	1
82	2	4	каменщик	4,2	16340	128	25	2
83	5	4	монтажник	4,3	20660	169	28	3
84	5	5	монтажник	5,0	33000	190	31	5
85	5	5	монтажник	5,3	30770	190	29	3
86	5	5	каменщик	6,0	34560	190	25	3
87	1	5	плотник	5,0	18440	127	22	2
88	5	4	монтажник	4,8	10890	80	28	1
89	2	5	каменщик	5,0	21770	140	23	2
90	1	3	монтажник	3,9	10120	87	43	1
91	2	3	монтажник	3,6	13000	106	22	1
92	3	4	каменщик	4,6	17770	123	22	1
93	4	5	каменщик	6,0	31550	189	25	2
94	3	4	каменщик	4,6	24550	180	21	2
95	4	3	монтажник	4,0	14440	168	26	2
96	2	4	каменщик	3,8	14110	120	23	2
97	1	4	монтажник	4,6	18550	144	28	3
98	3	4	монтажник	4,5	18770	139	28	2
99	2	4	плотник	4,4	18040	132	25	1
100	1	4	бетонщик	4,0	12550	104	23	1
101	2	5	монтажник	6,0	36880	231	36	5
102	1	3	каменщик	3,6	26120	200	33	4
103	2	5	каменщик	4,6	19550	137	32	3
104	5	4	монтажник	4,5	27220	210	30	3
105	1	4	плотник	4,5	16770	129	22	2
106	4	5	каменщик	5,0	19440	132	23	1
107	1	3	монтажник	6,0	18770	92	20	1
108	2	3	бетонщик	4,6	12980	92	20	1
109	2	5	монтажник	5,1	21020	144	22	2
110	1	4	монтажник	4,0	16120	125	23	1
111	3	4	монтажник	4,8	19220	168	24	2
112	3	4	монтажник	4,5	17220	144	20	2
113	4	5	монтажник	5,2	23000	160	38	4
114	1	4	каменщик	4,9	14880	116	19	1
115	3	4	монтажник	3,6	21660	168	36	3
116	1	3	каменщик	5,0	14550	107	20	1
117	3	5	каменщик	5,0	25550	186	23	2
118	2	4	монтажник	4,8	18660	132	22	2
119	3	1	монтажник	4,0	21640	168	31	3
120	3	4	монтажник	4,0	20880	153	32	4
121	1	3	плотник	4,0	14880	110	20	1
122	1	3	каменщик	4,1	13040	100	19	1
123	1	3	монтажник	3,0	13040	100	19	1
124	1	3	монтажник	4,1	11660	83	19	1
125	1	3	плотник	5,0	12330	76	20	1
126	1	3	монтажник	4,8	11080	80	19	1
127	1	3	каменщик	4,8	14090	110	22	1
128	2	4	монтажник	4,0	16770	121	23	1
129	3	3	каменщик	4,0	11880	111	24	2
130	2	4	каменщик	4,0	14880	111	23	2
131	2	4	монтажник	4,0	15110	111	21	2
132	2	3	каменщик	3,6	14990	115	22	1
133	2	3	каменщик	3,6	14990	115	21	1
134	2	4	бетонщик	4,9	17020	115	23	1
135	2	4	каменщик	3,6	14880	115	23	1

136	3	3	монтажник	5,0	15030	119	24	1
117	3	4	монтажник	5,0	17340	124	29	3
138	3	4	монтажник	5,0	15770	124	24	2
139	4	4	монтажник	4,3	16110	125	23	1
140	2	4	монтажник	5,3	18770	128	22	2
141	1	4	монтажник	4,0	14770	111	23	1
142	2	3	монтажник	3,6	15220	115	21	1
143	2	3	монтажник	3,6	15240	115	24	2
144	2	4	каменщик	4,9	17080	115	25	2
145	3	4	каменщик	3,9	15560	115	22	1
146	3	3	бетонщик	4,0	18110	130	28	3
147	3	4	монтажник	5,0	16000	124	25	2
148	3	4	монтажник	5,0	17770	124	23	1
149	3	4	монтажник	5,0	17770	124	21	1
150	3	4	монтажник	5,3	18550	128	23	2
151	3	3	монтажник	4,4	22550	162	24	4
152	3	4	монтажник	4,8	18550	132	25	3
153	2	3	каменщик	4,9	23000	160	31	5
154	1	4	каменщик	3,8	11550	96	30	1
155	1	4	монтажник	4,0	19770	152	22	3
156	4	4	монтажник	3,8	18440	142	22	2
157	5	3	монтажник	4,3	27550	203	35	4
158	1	3	плотник	4,4	10880	84	20	1
159	2	3	плотник	4,4	10880	84	21	1
160	1	4	каменщик	4,8	11330	80	19	1
161	1	4	монтажник	4,5	9500	68	19	1
162	1	3	каменщик	3,9	10100	87	20	1
163	1	3	каменщик	3,9	10100	87	19	1
164	1	3	каменщик	3,9	10100	87	20	1
165	2	3	монтажник	3,6	13000	106	32	2
166	2	3	бетонщик	4,1	13770	104	31	3
167	1	1	каменщик	4,6	28880	205	41	4
168	3	4	монтажник	4,6	25990	187	40	5
161	3	1	монтажник	4,6	18770	137	32	3
170	4	5	монтажник	5,0	18770	130	28	2
171	4	5	монтажник	5,3	20000	132	33	3
172	5	5	монтажник	5,0	28220	190	35	4
173	5	4	монтажник	4,8	24550	181	42	5
174	5	5	бетонщик	6,0	29220	190	31	4
175	1	3	каменщик	4,1	12660	92	19	1
176	1	4	бетонщик	5,0	17130	107	19	1
177	4	5	монтажник	4,9	25330	169	23	2
178	4	4	монтажник	4,9	26660	164	22	2
179	3	4	монтажник	4,9	26660	164	23	1
180	4	4	каменщик	4,9	24140	164	30	5
181	5	4	монтажник	5,0	26000	190	28	2
182	3	3	каменщик	3,0	12300	115	27	1
183	3	4	монтажник	5,0	19440	132	23	2
184	3	5	монтажник	5,0	20220	130	23	3
185	2	5	бетонщик	5,3	20770	130	30	2
186	1	3	каменщик	4,6	15880	123	19	1
187	4	5	каменщик	5,0	25880	132	23	2
188	5	4	монтажник	4,8	26220	203	29	4
189	2	4	каменщик	5,3	18770	128	23	2
190	3	4	каменщик	5,8	20110	128	25	2
191	5	3	монтажник	4,2	30330	232	33	5
192	3	4	монтажник	4,4	17110	132	26	2
193	4	5	каменщик	5,2	19000	127	22	2
194	4	5	монтажник	6,0	31550	190	29	3
195	2	4	монтажник	3,6	21660	168	23	2
196	2	4	каменщик	3,6	21640	168	21	1
197	1	4	каменщик	3,6	21640	168	21	1
198	1	4	бетонщик	4,0	13660	110	20	1
199	1	4	бетонщик	4,6	10330	80	20	1
200	2	4	монтажник	4,0	19770	153	20	2

ЛАБОРАТОРНАЯ РАБОТА №2

АНАЛИЗ СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ НА ОСНОВЕ ОБОБЩАЮЩИХ ПОКАЗАТЕЛЕЙ

4. ЦЕЛЬ И ЗАДАЧИ ИССЛЕДОВАНИЯ

Изучение и приобретение навыков расчета и анализа обобщающих показателей.

5. ОСНОВЫ ТЕОРИИ

ОСНОВНЫЕ СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ

Средние величины

Средние величины - это обобщающие показатели, в которых находит выражение действие общих условий, закономерность изучаемого явления. Статистические средние рассчитываются на основе массовых данных статистически правильно организованного массового наблюдения (сплошного или выборочного). Статистическая средняя будет объективна и типична, если она рассчитывается по массовым данным для качественно однородной совокупности.

Существует несколько типов средних величин, из которых - степенные средние. Исчисляется степенная средняя по общей формуле:

$$\bar{x} = \sqrt[m]{\frac{\sum x^m}{n}},$$

где m - степень;

$\bar{x}$ - средняя величина;

x - величина признака у конкретной величины совокупности, часто называемой вариантой;

n - количество единиц (вариант) в изучаемой совокупности.

Если одно и то же значение повторяется несколько раз, то степенная средняя определяется по формуле:

$$\bar{x} = \sqrt[m]{\frac{\sum x^m f}{\sum f}},$$

где f - частота (вес) признака.

В рамках одного типа степенных величин различают несколько видов средних величин. Как правило, их применение ограничено четырьмя видами средних:

- 1) средней арифметической;
- 2) средней гармонической;
- 3) средней квадратической;
- 4) средней геометрической.

Различаются они по показателю средней степени m (табл. 1).

Наибольшее применение имеют средние арифметические, несколько меньше - средняя гармоническая. Средняя геометрическая применяется при оценке динамики статистических явлений. Наиболее редко применяется средняя квадратическая.

Таблица 1

Виды степенных средних величин

средняя	вид средней	простая средняя	взвешенная средняя
---------	-------------	-----------------	--------------------

степень			
-1	Средняя гармоническая	$\bar{x} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$	$\bar{x} = \frac{\sum_{i=1}^n f_i}{\sum_{i=1}^n \frac{f_i}{x_i}}$
0	Средняя геометрическая	$\bar{x} = \sqrt[n]{\prod_{i=1}^n x_i}$	$\bar{x} = \sqrt[n]{\prod_{i=1}^n (x_i)^{f_i}}$
1	Средняя арифметическая	$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$	$\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i}$

Средняя арифметическая - наиболее распространенный тип средней. Она используется, когда определяющий показатель $g(x_1, \dots, x_n) = x_1 + \dots + x_n$.

Варианты x , по которым вычисляется взвешенная средняя арифметическая, могут быть отдельными значениями признака, но могут и сами быть средними величинами, что бывает в двух случаях:

1) варианты признака являются групповыми средними величинами. При вычислении общей средней арифметической из групповых средних каждая групповая средняя рассматривается как вариант, а число единиц в группе - как частота;

2) варьирующие значения признака даны в интервалах. Для вычисления средней величины в каждом интервале определяется серединное значение. Иногда задача вычисления средней осложняется тем, что неизвестны крайние границы начального и конечного интервалов. В этом случае предполагается, что расстояние между границами данного интервала такое же, как и в соседнем интервале.

Средняя гармоническая используется в следующих случаях:

1) когда заданы две величины, связанные обратно пропорциональной зависимостью, например, производительность труда и затраты времени на изготовление одной единицы продукции ($y = 1/x$), скорость движения и затраты времени на единицу пути;

2) когда один из двух заданных показателей включает другой показатель в виде произведения, например, надо определить среднюю цену товара, используя цену и сумму реализации по группам.

Область применения *средней геометрической* – статистический анализ динамики процессов, в частности, определение темпов роста.

Показатели вариации

Абсолютные показатели - величины именованные, имеющие ту же единицу измерения, что и варьирующий признак. К абсолютным относят размах вариации, среднее линейное отклонение, дисперсию и среднеквадратическое отклонение (СКО).

Дисперсия характеризует распределение признака внутри совокупности:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2,$$

$$\sigma^2 = \frac{1}{\sum f_i} \sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i.$$

Дисперсия характеризует отклонения отдельных значений случайной величины от средней арифметической. Чем меньше дисперсия, тем более тесно концентрируются отдельные значения случайной величины вблизи средней.

Дисперсия оказывается в ряде случаев неудобной для практического использования, так как имеет размерность квадрата величины. Поэтому в качестве характеристики рассеивания часто применяют корень квадратный из дисперсии, получивший название среднеквадратического отклонения (СКО).

Наибольшее применение на практике из относительных показателей нашел коэффициент вариации по СКО:

$$K_{\sigma} = \sigma / \bar{x}.$$

Коэффициент является наиболее значимым при оценке однородности изучаемой совокупности. Чем он больше, тем неоднороднее группа.

В качестве критерия однородности исследуемой совокупности принято значение показателя вариации 0.33 (для нормального распределения). Если изучаемая совокупность имеет показатель вариации ≤ 0.33 , то она считается однородной.

РАСЧЕТ ПОКАЗАТЕЛЕЙ НА ОСНОВЕ ВЫБОРОЧНЫХ ИССЛЕДОВАНИЙ

Средняя и предельная ошибки выборки

В процессе проведения выборочного наблюдения, как и вообще при анализе данных любого обследования, статистика выделяет два вида ошибок: регистрации и репрезентативности. Ошибки репрезентативности, органически присущие выборочному наблюдению, возникают в силу того, что выборочная совокупность не воспроизводит генеральную. Избежать ошибок репрезентативности нельзя, однако, пользуясь методами теории вероятностей, основанными на использовании предельных теорем закона больших чисел, эти ошибки можно свести к минимальным значениям, границы которых устанавливаются с достаточно большой точностью.

Основные характеристики параметров генеральной и выборочной совокупностей обозначаются определенными символами (табл. 2).

Таблица 2

Основные характеристики параметров
генеральной и выборочной совокупностей

Характеристика	Генеральная совокупность	Выборочная совокупность
Объем совокупности (численность единиц)	N	n
Среднее значение	$\bar{x}$ (истинное значение)	$\hat{\bar{x}}$ (оценка)

Ошибка выборочного наблюдения - это разность между величиной параметра в генеральной совокупности и его величиной, вычисленной по результатам выборочного наблюдения. Для среднего значения ошибка будет определяться так:

$$\Delta_{\bar{x}} = \left| \hat{\bar{x}} - \bar{x} \right|.$$

Величина $\Delta_{\bar{x}}$ называется предельной ошибкой выборки. Предельная ошибка выборки величина случайная. Предельная ошибка выборки определяется через среднюю ошибку по формуле $\Delta_{\bar{x}} = t\mu$, т.е. она равна t-кратному числу средних ошибок выборки. Величина t определяется исходя из уровня заданной доверительной вероятности P. Так, при доверительной вероятности P=0,997 t=3, при P=0,954 t=2, при P=0,663 t=1.

В свою очередь, величина μ , выражающая среднее квадратическое отклонение выборочной средней от генеральной средней, зависит от колеблемости признака в генеральной совокупности σ и числа отобранных единиц n . Эта зависимость выражается формулой для повторной и бесповторной выборки соответственно:

$$\mu = \frac{\sigma}{\sqrt{n}}, \quad \mu = \frac{\sigma}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$$

где μ - средняя ошибка выборки;

σ^2 - генеральная дисперсия.

Зная выборочную среднюю величину признака ($\bar{x}$) и предельную ошибку выборки, можно определить границы (пределы), в которых заключена генеральная средняя: $\bar{x} - t_{\gamma}\mu_{\bar{x}} \leq \bar{X} \leq \bar{x} + t_{\gamma}\mu_{\bar{x}}$.

ИЗУЧЕНИЕ ФОРМЫ РАСПРЕДЕЛЕНИЯ

Структурные характеристики совокупности

Для характеристики структуры совокупности применяются особые показатели, которые называют структурными. Для оценки среднего значения совокупности используется мода и медиана.

Модой (M_o) называется значение признака, имеющее наибольшую частоту. Мода представляет наиболее часто встречающееся или типичное значение. Мода широко используется в коммерческой практике при изучении покупательского спроса, например, при определении размеров одежды и обуви, которые пользуются наибольшим спросом.

Недостатком моды является то, что ее значение может быть определено неоднозначно: максимум f_1 может достигаться при нескольких различных значениях признака. А если признак непрерывный и повторяющихся значений нет (т.е. все $f_1=1$), то моду невозможно выбрать. Поэтому мода обычно определяется для дискретных признаков или для интервальных рядов. Распределение признака часто бывает унимодальным, поэтому в большинстве случаев мода определяется однозначно.

Введем теперь следующий показатель - *медиану*. Пусть объекты в статистической совокупности упорядочены по возрастанию значений признака, т.е. $x_1, x_2, \dots, x_n$. Такой ряд называется ранжированным.

Медиана (M_e) - это величина, для которой число единиц совокупности, у которых значение признака меньше медианы, равно числу единиц, у которых значение признака больше медианы. Если в ранжированном ряду нечетное число членов, т.е. $n = 2k - 1$, то медианой является вариант, расположенная в центре ряда: $M_e = x_k$.

Для ранжированного ряда с четным числом членов $n = 2k$ медианой будет средняя арифметическая из двух смежных вариантов. Медиана более определена, чем мода и ее основное практическое применение связано с тем, что сумма абсолютного отклонения членов ряда от медианы есть величина наименьшая:

$$\sum_{i=1}^n (x_i - M_e) \rightarrow \min.$$

Отметим, что структурные средние не связаны с определяющим показателем. Они не основные, а дополнительные характеристики совокупности.

Помимо моды и медианы, можно определить и другие структурные характеристики статистической совокупности. Например, пусть $k(x)$ - кумулятивная кривая, т.е. доля всех единиц в совокупности, у которых значение признака не превосходит $(n - 1)$. Значения x'_1, x'_2, x'_3 , для которых $k(x'_1)=1/4$, $k(x'_2)=1/2$, $k(x'_3)=3/4$, называются *квартелями* (при этом второй квартиль равен медиане: $x'_2 = M_e$), значения x''_1, x''_2, x''_3 , для которых $k(x''_j) = j/10, j = 1...9$, называются *децилями*, значения $k(x'''_j) = j/100, j = 1...99$ - *процентилями*.

Показатели асимметрии и эксцесса

В статистической практике встречаются разнообразные распределения. Наиболее часто используется нормальное распределение. Всякое искажение формы кривой означает нарушение или изменение нормальных условий. Выяснение общего характера распределения предполагает оценку его однородности, а также вычисление показателей асимметрии и эксцесса.

Для симметричных распределений частоты любых двух вариантов, равноотстоящих в обе стороны от центра распределения, равны между собой. Рассчитанные для таких распределений средняя, мода и медиана также равны. Для симметричных распределений средняя арифметическая и медиана равны между собой. Поэтому простейший показатель асимметрии $\bar{X} - M_e$.

Чем больше величина $|\bar{X} - M_e|$, тем больше асимметрия. Если $\bar{X} - M_e > 0$, то асимметрия правосторонняя, если $\bar{X} - M_e < 0$ - левосторонняя.

Вместо медианы можно использовать моду. Если среднее арифметическое больше моды, то имеем правостороннюю асимметрию, если среднее арифметическое меньше моды - левостороннюю асимметрию (рис.).



Рис. Правосторонняя и левосторонняя асимметрия

1. Коэффициент асимметрии

$$A = \eta_3 / \sigma^3.$$

$$\text{СКО коэффициента асимметрии } \zeta_A = \sqrt{6(n-1) / [(n+1)(n+3)]}.$$

Для нормального распределения эти коэффициенты равны нулю. Асимметрия незначительна, если $|A| < 1,5 \cdot \zeta_A$. Иначе асимметрия существенна.

2. Коэффициент эксцесса

$$E = \eta_4 / \sigma^4 - 3.$$

СКО коэффициента эксцесса

$$\zeta_E = \sqrt{24n(n-2)(n-3) / [(n-1)^2(n+3)(n+5)]}.$$

Для нормального распределения эти коэффициенты равны нулю. Распределение принято считать нормальным, если выполняются условия

$$\left| E - \frac{6}{n+1} \right| \leq 1,5\zeta_E.$$

Если $E > 0$ - пик острый, если $E < 0$ – пологий.

6. ЗАДАНИЕ

Используя средства OpenOffice.Calc решить следующие задачи по данным выборочного обследования рабочих строительного треста (общее число рабочих 1000 человек), варианты 1-30:

1. По исходным данным Ряд 1 и Ряд 2 в соответствии с вариантом определите среднюю арифметическую, ее ошибку для выборки, дисперсию, СКО, степень однородности, а также структурные характеристики: моду, медиану, асимметрию и эксцесс.

2. В соответствии с вариантом по Ряду 1 построить ряд дискретный, по Ряду 2 – интервальный (лабораторная работа №1).

3. По сгруппированным данным определите среднюю арифметическую, ее ошибку для выборки, дисперсию, СКО, степень однородности, а также структурные характеристики: моду, медиану, асимметрию и эксцесс. Определите отличие рассчитанных показателей от п1. Сформулировать выводы.

ВАРИАНТЫ ЗАДАНИЙ

Вариант	Ряд 1		Ряд 2	
	признак	номера	признак	номера
31.	Стаж работы	1-40	Зарплата за январь	1-40
32.	Тарифный разряд	1-60	Возраст рабочих	1-60
33.	Число членов семьи	1-80	% выполнения норм	1-80
34.	Стаж работы	21-60	Возраст рабочих	21-60
35.	Тарифный разряд	21-80	% выполнения норм	21-80
36.	Число членов семьи	21-100	Средний тариф.разряд	21-100
37.	Стаж работы	41-80	Возраст рабочих	41-80
38.	Тарифный разряд	41-100	% выполнения норм	41-100
39.	Число членов семьи	41-120	Средний тариф.разряд	41-120
40.	Стаж работы	51-100	% выполнения норм	51-100
41.	Тарифный разряд	61-100	Средний тариф.разряд	61-100
42.	Число членов семьи	71-140	Зарплата за январь	71-140
43.	Стаж работы	71-120	% выполнения норм	71-120
44.	Тарифный разряд	81-140	Средний тариф.разряд	81-140
45.	Число членов семьи	91-160	Возраст рабочих	91-160
46.	Стаж работы	101-140	Средний тариф.разряд	101-140
47.	Тарифный разряд	101-150	Зарплата за январь	101-150
48.	Число членов семьи	111-160	Возраст рабочих	111-160
49.	Стаж работы	121-160	Средний тариф.разряд	121-160
50.	Тарифный разряд	121-170	Возраст рабочих	121-170
51.	Число членов семьи	121-180	% выполнения норм	121-180
52.	Стаж работы	131-180	Зарплата за январь	131-180
53.	Тарифный разряд	141-190	Возраст рабочих	141-190
54.	Число членов семьи	151-200	% выполнения норм	151-200
55.	Стаж работы	131-200	Возраст рабочих	131-200
56.	Тарифный разряд	141-200	% выполнения норм	141-200
57.	Число членов семьи	161-200	Средний тариф.разряд	161-200
58.	Стаж работы	11-80	Возраст рабочих	11-80
59.	Тарифный разряд	11-100	% выполнения норм	11-100
60.	Число членов семьи	11-60	Средний тариф.разряд	11-60

ИСХОДНЫЕ ДАННЫЕ ДЛЯ ВЫПОЛНЕНИЯ ЗАДАНИЙ ПО ВАРИАНТАМ

Данные обследования рабочих строительного треста, принятых на работу за последние 10 лет.

Порядковый номер рабочего	Стаж работы на данном предприятии	Тарифный разряд	Профессия, полученная в ПТУ	Средний тарифный разряд работ в январе	Зарплата за январь (р.)	% выполнения норм в январе	Возраст рабочих (в годах)	Число членов семьи
1	2	3	4	5	6	7	8	9
1	5	5	каменщик	5,0	27600	186	25	1
2	3	4	монтажник	4,7	20100	181	22	2
3	4	4	каменщик	3,8	24400	190	34	4
4	5	3	монтажник	4,8	25300	181	28	2
5	3	4	монтажник	4,6	23600	181	22	1
6	3	4	монтажник	4,7	27100	200	40	4
7	5	4	каменщик	4,2	21000	176	38	3
8	5	5	монтажник	5,2	21600	140	32	2
9	3	4	монтажник	4,5	17800	144	30	2
10	5	5	монтажник	6,0	21300	136	23	2
11	4	5	бетонщик	4,9	24000	164	25	2
12	5	3	монтажник	4,0	21800	170	25	2
13	1	3	каменщик	3,5	22000	169	27	3
14	2	4	монтажник	5,0	21600	155	23	2
15	2	1	монтажник	4,0	19400	158	26	1
16	2	3	монтажник	4,0	17400	156	28	2
17	4	4	монтажник	3,6	20900	168	29	2
18	1	4	монтажник	4,9	16300	115	20	1
19	2	4	монтажник	3,7	14900	116	20	1
20	3	3	монтажник	4,1	16000	119	22	1
21	2	3	монтажник	4,3	16200	119	21	1
22	2	4	монтажник	4,3	17300	125	22	1
23	2	4	монтажник	4,0	16100	127	22	1
24	3	4	монтажник	5,0	18700	128	23	2
25	1	4	плотник	4,0	16300	129	19	1
26	1	4	плотник	4,5	16800	129	25	1
27	4	5	плотник	5,0	21000	140	28	2
28	1	4	бетонщик	4,2	15100	113	22	1
29	2	4	бетонщик	4,0	13700	110	20	1
30	2	4	бетонщик	3,8	14200	112	21	1
31	1	3	плотник	3,9	14500	109	20	1
32	3	4	плотник	4,9	18000	132	23	1
33	5	4	монтажник	4,9	18720	135	26	3
34	3	1	плотник	4,8	17440	132	25	1
35	5	4	монтажник	4,8	20100	132	24	3
36	3	3	плотник	4,9	18170	132	21	1
37	2	5	монтажник	4,4	20260	132	22	1
38	1	4	бетонщик	4,2	14130	113	19	1
39	3	4	бетонщик	5,0	14540	100	24	1
40	1	4	бетонщик	4,0	10470	80	20	1
41	3	3	монтажник	4,1	14680	104	22	2
42	3	3	плотик	4,1	13800	104	21	1
43	1	4	плотник	5,0	15400	107	21	1
44	4	5	плотник	4,0	27040	230	24	1
45	2	3	каменщик	3,0	11680	96	2:	1
46	5	4	монтажник	4,8	25910	190	36	5
47	3	4	монтажник	4,9	30220	232	32	4
48	2	5	монтажник	4,8	30440	195	31	3
49	3	3	каменщик	5,0	27550	186	36	4
50	4	4	монтажник	4,5	23330	180	28	3
51	3	5	Плотник	3,2	13510	105	36	2
52	1	3	каменщик	4,2	12380	90	20	1
53	5	4	монтажник	5,0	28990	203	33	3
54	5	3	каменщик	4,3	29940	153	35	2
55	5	4	монтажник	4,0	20400	158	34	4
56	2	4	монтажник	4,0	19080	146	22	1
57	3	4	бетонщик	4,0	13080	100	24	1
58	3	4	бетонщик	3,9	16120	125	23	2
59	4	4	монтажник	5,0	22200	159	26	2
60	2	3	плотник	4,1	17110	123	27	1

61	4	4	монтажник	4,6	20670	159	33	4
62	3	4	бетонщик	5,5	16180	100	24	1
63	3	5	каменщик	4,9	27840	190	26	3
64	1	4	каменщик	4,1	27350	199	27	3
65	3	4	монтажник	4,8	19650	139	26	1
66	2	5	монтажник	4,8	27020	190	38	4
67	1	5	монтажник	5,2	13840	83	20	1
68	1	3	бетонщик	4,0	12380	98	19	1
69	2	4	плотник	4,1	18000	139	20	2
70	1	4	монтажник	4,8	10920	85	20	1
71	4	5	монтажник	4,8	20240	146	30	3
72	5	4	монтажник	4,7	25240	183	31	4
73	3	4	бетонщик	3,8	12300	90	21	1
74	3	4	монтажник	4,8	17660	146	25	2
75	2	3	бетонщик	3,5	12080	95	20	1
76	1	4	плотник	4,9	14440	110	19	1
77	1	4	плотник	4,9	15880	113	19	1
78	1	4	монтажник	4,9	13550	113	22	2
79	2	4	монтажник	4,0	14440	111	24	2
80	1	4	каменщик	3,8	14440	110	19	1
81	1	4	монтажник	4,2	16550	128	19	1
82	2	4	каменщик	4,2	16340	128	25	2
83	5	4	монтажник	4,3	20660	169	28	3
84	5	5	монтажник	5,0	33000	190	31	5
85	5	5	монтажник	5,3	30770	190	29	3
86	5	5	каменщик	6,0	34560	190	25	3
87	1	5	плотник	5,0	18440	127	22	2
88	5	4	монтажник	4,8	10890	80	28	1
89	2	5	каменщик	5,0	21770	140	23	2
90	1	3	монтажник	3,9	10120	87	43	1
91	2	3	монтажник	3,6	13000	106	22	1
92	3	4	каменщик	4,6	17770	123	22	1
93	4	5	каменщик	6,0	31550	189	25	2
94	3	4	каменщик	4,6	24550	180	21	2
95	4	3	монтажник	4,0	14440	168	26	2
96	2	4	каменщик	3,8	14110	120	23	2
97	1	4	монтажник	4,6	18550	144	28	3
98	3	4	монтажник	4,5	18770	139	28	2
99	2	4	плотник	4,4	18040	132	25	1
100	1	4	бетонщик	4,0	12550	104	23	1
101	2	5	монтажник	6,0	36880	231	36	5
102	1	3	каменщик	3,6	26120	200	33	4
103	2	5	каменщик	4,6	19550	137	32	3
104	5	4	монтажник	4,5	27220	210	30	3
105	1	4	плотник	4,5	16770	129	22	2
106	4	5	каменщик	5,0	19440	132	23	1
107	1	3	монтажник	6,0	18770	92	20	1
108	2	3	бетонщик	4,6	12980	92	20	1
109	2	5	монтажник	5,1	21020	144	22	2
110	1	4	монтажник	4,0	16120	125	23	1
111	3	4	монтажник	4,8	19220	168	24	2
112	3	4	монтажник	4,5	17220	144	20	2
113	4	5	монтажник	5,2	23000	160	38	4
114	1	4	каменщик	4,9	14880	116	19	1
115	3	4	монтажник	3,6	21660	168	36	3
116	1	3	каменщик	5,0	14550	107	20	1
117	3	5	каменщик	5,0	25550	186	23	2
118	2	4	монтажник	4,8	18660	132	22	2
119	3	1	монтажник	4,0	21640	168	31	3
120	3	4	монтажник	4,0	20880	153	32	4
121	1	3	плотник	4,0	14880	110	20	1
122	1	3	каменщик	4,1	13040	100	19	1
123	1	3	монтажник	3,0	13040	100	19	1
124	1	3	монтажник	4,1	11660	83	19	1
125	1	3	плотник	5,0	12330	76	20	1
126	1	3	монтажник	4,8	11080	80	19	1
127	1	3	каменщик	4,8	14090	110	22	1
128	2	4	монтажник	4,0	16770	121	23	1
129	3	3	каменщик	4,0	11880	111	24	2
130	2	4	каменщик	4,0	14880	111	23	2
131	2	4	монтажник	4,0	15110	111	21	2
132	2	3	каменщик	3,6	14990	115	22	1
133	2	3	каменщик	3,6	14990	115	21	1
134	2	4	бетонщик	4,9	17020	115	23	1
135	2	4	каменщик	3,6	14880	115	23	1

136	3	3	монтажник	5,0	15030	119	24	1
117	3	4	монтажник	5,0	17340	124	29	3
138	3	4	монтажник	5,0	15770	124	24	2
139	4	4	монтажник	4,3	16110	125	23	1
140	2	4	монтажник	5,3	18770	128	22	2
141	1	4	монтажник	4,0	14770	111	23	1
142	2	3	монтажник	3,6	15220	115	21	1
143	2	3	монтажник	3,6	15240	115	24	2
144	2	4	каменщик	4,9	17080	115	25	2
145	3	4	каменщик	3,9	15560	115	22	1
146	3	3	бетонщик	4,0	18110	130	28	3
147	3	4	монтажник	5,0	16000	124	25	2
148	3	4	монтажник	5,0	17770	124	23	1
149	3	4	монтажник	5,0	17770	124	21	1
150	3	4	монтажник	5,3	18550	128	23	2
151	3	3	монтажник	4,4	22550	162	24	4
152	3	4	монтажник	4,8	18550	132	25	3
153	2	3	каменщик	4,9	23000	160	31	5
154	1	4	каменщик	3,8	11550	96	30	1
155	1	4	монтажник	4,0	19770	152	22	3
156	4	4	монтажник	3,8	18440	142	22	2
157	5	3	монтажник	4,3	27550	203	35	4
158	1	3	плотик	4,4	10880	84	20	1
159	2	3	плотник	4,4	10880	84	21	1
160	1	4	каменщик	4,8	11330	80	19	1
161	1	4	монтажник	4,5	9500	68	19	1
162	1	3	каменщик	3,9	10100	87	20	1
163	1	3	каменщик	3,9	10100	87	19	1
164	1	3	каменщик	3,9	10100	87	20	1
165	2	3	монтажник	3,6	13000	106	32	2
166	2	3	бетонщик	4,1	13770	104	31	3
167	1	1	каменщик	4,6	28880	205	41	4
168	3	4	монтажник	4,6	25990	187	40	5
161	3	1	монтажник	4,6	18770	137	32	3
170	4	5	монтажник	5,0	18770	130	28	2
171	4	5	монтажник	5,3	20000	132	33	3
172	5	5	монтажник	5,0	28220	190	35	4
173	5	4	монтажник	4,8	24550	181	42	5
174	5	5	бетонщик	6,0	29220	190	31	4
175	1	3	каменщик	4,1	12660	92	19	1
176	1	4	бетонщик	5,0	17130	107	19	1
177	4	5	монтажник	4,9	25330	169	23	2
178	4	4	монтажник	4,9	26660	164	22	2
179	3	4	монтажник	4,9	26660	164	23	1
180	4	4	каменщик	4,9	24140	164	30	5
181	5	4	монтажник	5,0	26000	190	28	2
182	3	3	каменщик	3,0	12300	115	27	1
183	3	4	монтажник	5,0	19440	132	23	2
184	3	5	монтажник	5,0	20220	130	23	3
185	2	5	бетонщик	5,3	20770	130	30	2
186	1	3	каменщик	4,6	15880	123	19	1
187	4	5	каменщик	5,0	25880	132	23	2
188	5	4	монтажник	4,8	26220	203	29	4
189	2	4	каменщик	5,3	18770	128	23	2
190	3	4	каменщик	5,8	20110	128	25	2
191	5	3	монтажник	4,2	30330	232	33	5
192	3	4	монтажник	4,4	17110	132	26	2
193	4	5	каменщик	5,2	19000	127	22	2
194	4	5	монтажник	6,0	31550	190	29	3
195	2	4	монтажник	3,6	21660	168	23	2
196	2	4	каменщик	3,6	21640	168	21	1
197	1	4	каменщик	3,6	21640	168	21	1
198	1	4	бетонщик	4,0	13660	110	20	1
199	1	4	бетонщик	4,6	10330	80	20	1
200	2	4	монтажник	4,0	19770	153	20	2

Используемые функции

Средние величины

- AVERAGE(Диапазон) – среднее арифметическое
- GEOMEAN(Диапазон) – среднее геометрическое

Изучение формы распределения

- TRIMMEAN(диапазон; доля используемых элементов) – среднее арифметическое без учета крайних значений
- MEDIAN(Диапазон) - медиана
- MODE(Диапазон) - мода
- QUARTILE(Диапазон; номер квартиля) - квартиль
- KURT(Диапазон) – эксцесс
- SKEW(Диапазон) - ассиметрия

Показатели вариации:

- VAR(Диапазон) – дисперсия для выборки
- VARP(Диапазон) – дисперсия для генеральной совокупности
- STDEV(Диапазон) - СКО для выборки
- STDEVP(Диапазон) – СКО для генеральной совокупности