Министерство науки и высшего образования Российской Федерации

Уральский федеральный университет имени первого Президента России Б.Н.Ельцина

О Кеткина

Общая Теория Статистики Электронные Варианты Оценочных Работ

Решения и Варианты Открытые Широкому Кругу Студентов

Электронный ресурс содержит теоретический материал, примеры с решением и варианты заданий для выполнения самостоятельной работы. Материалы ресурса могут быть использованы при выполнении индивидуальной самостоятельной работы по курсу статистика, а также при подготовке к семинарским занятиям и при самостоятельном изучении предмета.

Издание подготовлено для студентов, обучающихся по программе бакалавриата по направлению подготовки 38.03.05 «Бизнес-информатика».

Подготовлено кафедрой «Моделирования управляемых систем»

Екатеринбург
2023

Оглавление

ОБЩАЯ ТЕОРИЯ СТАТИСТИКИ 4

ОБЩИЕ СЕДЕНИЯ 4

АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ В СТАТИСТИКЕ 5

ПРИМЕРЫ ДЛЯ САМОСТОЯТЕЛЬНОГО ВЫПОЛНЕНИЯ 7

КРАТКОЕ ОПИСАНИЕ 8

МУЛЬТИПЛИКАТИВНЫЙ И Н Д Е К С 8

МУЛЬТИПЛИКАТИВНЫЙ И Н Д Е К С 9

МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ВЫПОЛНЕНИЮ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ПО ТЕМЕ ГРУППИРОВКА ДАННЫХ 17

§ 1. Построение вариационных рядов 17

§ 2. Вычисление Моды и Медианы. 23

§ 3. Построение дискретного ряда, нахождение *Mo* и *Me*. 29

§ 4. Нахождение среднего, дисперсии и квадратического отклонения дискретного и интервального рядов. 31

ВАРИАНТЫ ДОМАШНЕГО ЗАДАНИЯ. 35

ВВЕДЕНИЕ

Данный электронный ресурс разработан для студентов, обучающихся по программам бакалавриата. В процессе изучения курса студенты выполняют индивидуальную самостоятельную работу, целью которой является выявление степени усвоения изученного материала по предмету, и демонстрация умения применять на практике изученные приемы обработки данных. Электронный текстовый ресурс призван донести необходимую информацию и помочь студентам верно выполнить индивидуальную самостоятельную работу по курсу.

При написании электронного текстового ресурса использовались отечественные и иностранные учебники и статьи. Список использованной литературы приведен в конце пособия.

Данный вариант электронного текстового ресурса является рабочей версией, дополнения и комментарии приветствуются, и могут быть направлены автору по электронной почте.

ОБЩАЯ ТЕОРИЯ СТАТИСТИКИ

ОБЩИЕ СЕДЕНИЯ

В данном электронном ресурсе мы раскроем понятия нескольких тем предмета, что необходимы для выполнения индивидуального самостоятельного проекта. Среди тем, следующие:

Этапы организации статистического исследования

Абсолютные и относительные величины в статистике

Индексы

Группировка данных

ЭТАПЫ ОРГАНИЗАЦИИ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ

Статистика – это способ изучения статистических закономерностей, которые проявляются во множестве случаев. Предметом изучения статистики выступает статистическая совокупность – множество явлений, объединенных общим качеством, но различающихся (варьирующих) между собой. Статистическая совокупность состоит из отдельных явлений – единиц совокупности. Единица совокупности – частный случай проявления статистической закономерности. *Признак* – качественная особенность единицы совокупности. *Статистический показатель* – обобщение признака по многим единицам.

Таким образом, основными категориями статистики являются: статистическая закономерность, статистическая совокупность, единица совокупности, признак единицы совокупности, и статистический показатель.

Рассмотрим на примере. Пусть предметом исследования является следующая статистическая закономерность: среди граждан РФ пожилого возраста, от 60-ти и старше, число осведомленных о наличии возможности обратиться в колл центр экстренной психологической помощи МЧС в случае сложной жизненной ситуации в среднем в 10 раз меньше чем среди молодежи, лиц от 18 до 25 лет. В данном случае статистической совокупностью будут являться граждане РФ в возрасте от 60-ти и старше и от 18 до 25 лет. Единицей совокупности будет каждый отдельный гражданин РФ данного возраста. Признаки, что мы будем уточнять у граждан, для выявления закономерности следующие: возраст, осведомлен или нет гражданин о наличии возможности обратиться в колл центр ФГБУ ЦЭПП МЧС России в случае возникновения сложной ситуации в жизни; дополнительно можно уточнить другие характеристики, что могут оказывать влияние на осведомленность граждан, например, пол, населенный пункт проживания, уровень образования, профессию и др. Статистический показатель, что мы будем вычислять – доли осведомленных и неосведомленных в зависимости от возраста опрашиваемых. В качестве частных совокупностей отдельно также можно рассмотреть мужчин и женщин, возможно во всех возрастных категориях среди женщин будет большее количество осведомленных, чем среди мужчин или наоборот, и так далее.

ПРИМЕРЫ ДЛЯ САМОСТОЯТЕЛЬНОГО ВЫПОЛНЕНИЯ.

1. Укажите основные категории статистики для изучения следующей статистической закономерности. Среди граждан, проинформированных о наличии в ЦЭПП МЧС России колл центра для обращения граждан, попавших в чрезвычайную ситуацию, готовы обратиться в данный центр в случае затруднений около 20%, остальные предпочтут справиться с ситуацией самостоятельно. Среди тех, кто готов обратиться за помощью в колл центр около 70% женщины, среди них 30% в возрасте до 35 лет, остальные более старшего возраста.

2. Для изучения нижеприведенной статистической закономерности укажите основные категории статистики. В России каждый 2-ой не сможет назвать телефонный номер, который предназначен для использования в экстренных ситуациях и для получения консультаций по вопросам безопасности и способам защиты от чрезвычайных ситуаций. Среди пожилых граждан, от 60-ти и старше, данный телефонный номер не сможет может назвать ни один.

АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ В СТАТИСТИКЕ

ОСНОВНЫЕ СВЕДЕНИЯ.

Абсолютная величина отражает уровень развития явления. В статистике все абсолютные величины измеряются в конкретных единицах (рублях, штуках, киловатт-часах, килограммах и т.д.) и могут принимать как положительные, так и отрицательные значения (прибыль/убытки, рост/снижение, выигрыш/потери т.д.). Различают моментные и интервальные абсолютные величины: моментные показывают фактическое наличие или уровень явления на определенный момент, дату (например, численность проживающих на дату, температуру в конкретный момент времени); интервальные показывают итоговый накопленный результат за период в целом (например, объем произведенной продукции за год, прирост/убыль населения за период).

Сама по себе абсолютная величина не дает полного представления об изучаемом явлении, не показывает его структуру, соотношение между отдельными частями, развитие во времени. В ней не показаны соотношения с другими абсолютными показателями. Эти функции выполняют относительные величины. Относительная величина – это обобщающий показатель, который дает числовую меру соотношения двух сопоставляемых абсолютных величин. Кроме того, относительные величины могут быть определены через другие относительные величины. Относительные величины определяются в форме коэффициентов, процентов, и т.д.

Выделяют следующие типы относительных величин:

1. Относительная величина динамики – характеризует изменение уровня развития какого-либо явления во времени.

2. Относительная величина структуры – характеризуют доли, удельные веса составных элементов в общем итоге (вычисляются в долях или в %).

3. Относительные величины сравнения - сопоставляют размеры одноименных абсолютных величин, относящихся к одному и тому же периоду или моменту времени, но к различным объектам или территориям.

Рассмотрим *относительные величины динамики*. Пример, численность психологов МЧС России на конец 2011 года составляла около 700 человек (Y_0), в конце 2023 года данный показатель вырос до 900 человек (Y_1). Относительная величина динамики (иначе коэффициент роста или индекс роста) составит:

$$i = \frac{Y_1}{Y_0} = \frac{900}{700} = 1{,}286$$

Этот же показатель в процентах 128,6% называют темп роста. Т. о. численность за 12 лет выросла на 28,6%. В среднем каждый год по сравнению с предыдущим годом численность сотрудников изменяется в $\sqrt[12]{1{,}286} = 1{,}021$ раза (среднегодовой коэффициент или среднегодовой индекс роста), т.е. возрастает на 2,1% (среднегодовой темп роста).

Другой пример, вычислите динамику (коэффициент роста или индекс роста) численности жителей небольшого поселка во время зимнего периода, если известно, что во время летнего сезона количество жителей в населенном пункте увеличивается за счет отдыхающих до 1 400 человек (Y_0), а в остальные периоды в поселке проживают 750 человек (Y_1)[1]. Искомый показатель равен:

$$i = \frac{Y_1}{Y_0} = \frac{750}{1\,400} = 0{,}536$$

Таким образом, численность в течение года (с летнего в зимний период) сокращается на 46,4%.

Вычислим *относительные величины сравнения*.

При известных коэффициентах роста (i_t^A и i_t^B) и начальном соотношении уровней (Y_A и Y_B) можно записать условие равенства уровней в предстоящем периоде t:

$$Y_A \cdot i_A^t = Y_B \cdot i_B^t$$

Откуда получить формулу для расчета t, которое показывает через какой период времени уровень на объекте A сравняется с уровнем на объекте B:

$$t = \frac{lg\left(\frac{Y_A}{Y_B}\right)}{lg\left(\frac{i_B}{i_A}\right)}$$

---

[1] Презентация на тему «Waste sorting acceptability with information provision», на конференции «The 2nd ESDCA Conference», Smolensk, Russia, 15-17 февраля, 2022 г..

Пример. В 2023 году количество обращений в ЦЭПП МЧС в рамках интернет-консультирования достигло 11,5 тыс., и 1,5 тыс. граждан обратились за помощью по телефону. Предположим, в центре появится телефон для бесплатных звонков со всей территории России и в последующие пять лет, 2024 – 2028 гг. среднегодовые темпы прироста обращений по телефону составят 15,5%, а в рамках интернет-консультирования (-4,5%). Через сколько лет, сохраняя данные темпы роста, количество обращений по телефону и через интернет сравняются? [2]

В данном случае Y_A = 11,5 тыс. обращений, Y_B = 1,5 тыс. обращений, i_t^A = 100% - 4,5% = 95,5%, i_t^B = 100% + 15,5% = 115,5%. Период времени, через который количество обращений по телефону и через интернет сравняется, определим из соотношения:

$$t = \frac{lg\left(\frac{Y_A}{Y_B}\right)}{lg\left(\frac{i_B}{i_A}\right)} = \frac{lg(11{,}5\ /\ 1{,}5)}{\lg\,(1{,}155/0{,}955)} = 10{,}712 \text{ лет}$$

При условии, что темпы прироста обращений каждым из рассматриваемых способов (по телефону и через интернет) сохранятся в течение последующих 12 лет, можно ожидать, что ежегодное количество обращений через интернет сравняется с ежегодным количеством обращений по телефону приблизительно через 12 лет. При этом суммарное количество обращений, по телефону и через интернет, составит около 14,25 тыс. обращений в год.

ПРИМЕРЫ ДЛЯ САМОСТОЯТЕЛЬНОГО ВЫПОЛНЕНИЯ

1. Вычислите всевозможные относительные показатели изменения численности сотрудников центра экстренной психологической помощи (ЦЭПП) МЧС России за последние 10 лет.

2. Представьте статистику и полученные вами ответы на нескольких видах графиков из следующих: ленточные, полосовые, секторные, квадратные, круговые, фигурные диаграммы.

3. Начиная с 2011 года Путин призывает развивать целлюлозно-бумажную промышленность в России. Используя показатели, характеризующие среднее изменение за период, темп роста, прироста и другие по вашему усмотрению, опишите динамику производства бумаги в России и в Европе начиная с 2011 года.

4. Используя показатели, характеризующие среднее изменение за период, темп роста, прироста и другие по вашему усмотрению, опишите динамику производства ткани в России, Китае, Индии.

5. Представьте статистику по численности занятых, самозанятых, безработных и пр. в Дальневосточном федеральном округе за 10 лет. Опишите динамику показателей.

[2] Также любопытно было бы оценить как изменится аудитория, обратившихся в центр, с использованием каждого из видов связи (интернет и телефон). На текущий момент такие данные не доступны.

6. Изучите динамику рыбной промышленности Дальневосточного федерального округа, с момента вступления в должность текущего полномочного представителя президента РФ, Трутнева. Сопоставьте полученные показатели с таковыми граничащих с данным округом соседних государств, имеющих выход к морю.

7. Представьте статистику и полученные вами ответы на нескольких видах графиков – ленточные, полосовые, секторные, квадратные, круговые, фигурные диаграммы.

КРАТКОЕ ОПИСАНИЕ

МУЛЬТИПЛИКАТИВНЫЙ ИНДЕКС

В общем смысле индекс – это показатель сравнения двух состояний одного и того же явления. Наиболее часто понятие индекс используется для обозначения относительного изменения какого-либо показателя во времени. Такое понимание индекса мы и будем использовать в рамках данного пособия.

Индексы вычисляются путем деления значения показателя в текущий (отчетный) период времени на значение этого же показателя в базисный период времени (база сравнения). Для удобства используют подстрочные обозначения: 0 для базисного периода и 1 для текущего периода.

В данном пособии мы рассмотрим индивидуальные и мультипликативные индексы.

Относительная величина называется *индивидуальным индексом*, если исследователь сравнивает необобщенные величины. Индивидуальные индексы принято обозначать прописной буквой *i,*

i_x где x – индексируемая величина; показатель, изменение которого характеризует индекс.

Индивидуальные индексы показывают соотношение между отчетным и базисным показателями, обычно выраженное в процентах; или характеризуют во сколько раз уменьшился или увеличился показатель за рассматриваемый период.

Например, мы сравниваем цену продукции отчетного года p_1 с ценой продукции базисного года p_0. В итоге получаем индивидуальный индекс цены (i_p):

$$i_p = \frac{p_1}{p_0},$$

i_p показывает во сколько раз увеличилась (или уменьшилась) цена продукции в текущем году (p_1) по сравнению с ценой продукции базисного года (p_0).

Аналогично можно получить, индивидуальный индекс себестоимости производства продукции (i_z):

$$i_z = \frac{z_1}{z_0},$$

i_z будет показывать во сколько раз увеличилась (или уменьшилась) себестоимость производства продукции в текущем году (z_1) по сравнению с себестоимостью производства продукции в базисном году (z_0);

Аналогично вычисляется индивидуальный индекс количества проданного товара (i_q):

$$i_q = \frac{q_1}{q_0},$$

i_q будет показывать во сколько раз увеличилось (или уменьшилось) количество продаж некоторого товара в текущем году (q_1) по сравнению с количеством продаж данного товара в базисном году (q_0).

В качестве примера расчета индивидуального индекса рассмотрим статистику о количестве людей, которым потребовалась экстренная психологическая помощь в течение двух рассматриваемых периодов[3]. Так в 2020 году в России экстренная психологическая помощь потребовалась 16 865 людям (округлим до тысяч, около 17 000 человек), в 2021 году данная цифра составила 13 000.

Вычислим индивидуальный индекс количества экстренной психологической помощи оказанной гражданам в течение года силами ЦЭПП МЧС России:

$$\mu_{\text{ЦЭПП МЧС}} = \frac{13\ 000}{17\ 000} = 0{,}7647$$

значение индекса 0,7647 указывает на снижение на 13,53% показателя числа людей, которым была необходима экстренная психологическая помощь в 2021 по сравнению с годом ранее.

МУЛЬТИПЛИКАТИВНЫЙ ИНДЕКС

Индексный метод в целом предполагает, что связь между признаками жестко определена и выражается в виде уравнения связи:

аддитивного

$y = x_1 + x_2 + \ldots + x_k$

либо мультипликативного

$y = x_1 \cdot x_2 \cdot \ldots \cdot x_k$

Рассмотрим мультипликативную модель с двумя факторами. Пусть общее количество консультаций (Q), предоставленных колл центром экстренной психологической помощи МЧС зависит от количества сотрудников колл центра (N) и от количества звонков, отвеченных каждым сотрудником (M).

Тогда общее количество консультаций (Q) составит $Q = N \cdot M$.

Сравнивая количество консультаций в отчетном периоде (Q_1) с количеством предоставленных консультаций в базисном периоде (Q_0), получаем:

---

[3] Официальный сайт экстренной психологической помощи МЧС России. https://psi.mchs.gov.ru/psihologicheskaya-sluzhba/o-nas#

$$i_Q = \frac{Q_1}{Q_0} = \frac{N_1 \cdot M_1}{N_0 \cdot M_0} = i_N \cdot i_M$$

т.е. динамика общего количества консультаций (i_Q) определяется динамикой численности сотрудников ЦЭПП МЧС (i_N) и динамикой численности звонков, отвеченных каждым сотрудником (i_M).

Т.е. $Q_1 = Q_0 \cdot i_N \cdot i_M$ (1)

и мы можем разложить на части изменение количества консультаций (ΔQ) в абсолютном выражении следующим образом:

$\Delta Q = Q_1 – Q_0 = \Delta Q(\Delta N) + \Delta Q(\Delta M)$,

где $\Delta Q(\Delta N)$ – изменение количества консультаций, вызванное изменением численности сотрудников, $\Delta Q(\Delta M)$ – изменение количества консультаций, вызванное изменением продуктивности каждого сотрудника, т.е. увеличением или уменьшением количества консультаций, предоставляемых каждым сотрудником.

В частности, изменение количества консультаций в связи с изменением численности сотрудников можно записать как:

$\Delta Q(\Delta N) = Q_0 \cdot (i_N – 1)$ (2)

т.е. общее количество консультаций в связи с изменением численности сотрудников определятся базисным (исходным) значением количества консультаций умноженным на величину прироста / или снижения количества сотрудников, определяемую показателем ($i_N – 1$).

Изменение общего количества консультаций в связи с изменением продуктивности сотрудников составит:

$\Delta Q(\Delta M) = Q_0 \cdot i_N \cdot (i_M – 1)$ (3)

т.е. общее количество консультаций в связи с изменением продуктивности сотрудников определятся базисным (исходным) значением общего количества консультаций, скорректированным на динамику изменения численности сотрудников, умноженным на величину прироста / или снижения количества консультаций, предоставляемых каждым сотрудником ЦЭПП МЧС, определяемую показателем ($i_M – 1$).

Итоговое изменение общего количества консультаций можно записать в виде

$\Delta Q = \Delta Q(\Delta N) + \Delta Q(\Delta M) = Q_0 \cdot (i_N – 1) + Q_0 \cdot i_N \cdot (i_M – 1) = Q_0 \cdot (i_N \cdot i_M – 1)$

Немного усложним модель, рассмотрим мультипликативную модель с тремя факторами. Пусть общая сумма затрат на приобретение материала определенного вида (*M* – материал) зависит от объема производства продукции из этого материала (*q*), от расхода данного материла на единицу продукции – удельного расхода (*n* – количество материала на одно изделие) и от цены единицы данного и материала (*p*).

Тогда общая сумма затрат на приобретение материала составит $M = q \cdot n \cdot p$

$$\begin{matrix} \text{Сумма затрат на} & & \text{объем производства} & & \text{расход материала} & & \text{цена} \\ \text{приобретение} & = & \text{продукции из} & \cdot & \text{на единицу} & \cdot & \text{единицы} \\ \text{материалов } (M) & & \text{данного материала } (q) & & \text{продукции } (n) & & \text{материала } (p) \end{matrix}$$

Сравнивая сумму затрат на приобретение материала в отчетном периоде (M_1) с суммой затрат на материал в базисном периоде (M_0), получаем:

$$i_M = \frac{M_1}{M_0} = \frac{q_1 \cdot n_1 \cdot p_1}{q_0 \cdot n_0 \cdot p_0} = i_q \cdot i_n \cdot i_p$$

т.е. динамика общей суммы затрат на материал (i_M) определяется динамикой производства продукции (i_q), динамикой удельного расхода материала на единицу продукции (i_n) и изменением цены на материал (i_p).

Т.е. $M_1 = M_0 \cdot i_q \cdot i_n \cdot i_p$ (4)

и мы можем разложить на части изменение суммы затрат на материал (ΔM) в абсолютном выражении следующим образом:

$\Delta M = M_1 - M_0 = \Delta M(\Delta q) + \Delta M(\Delta n) + \Delta M(\Delta p)$,

где $\Delta M(\Delta q)$ – изменение затрат на материал, вызванное изменением объема производства продукции из данного материала, $\Delta M(\Delta n)$ – изменение затрат на материал, вызванное изменением расхода материала на единицу продукции, $\Delta M(\Delta p)$ – изменение затрат на материал, вызванное изменением цен на материал.

В частности, изменение суммы затрат в связи с изменением объема производства продукции можно записать как: $\Delta M(\Delta q) = M_0 \cdot (i_q - 1)$ (5)

т.е. сумма затрат в связи с изменением производства определятся базисным (исходным) значением затрат умноженным на величину прироста / или снижения объема производства, определяемую показателем ($i_q - 1$).

Изменение суммы затрат в связи с изменением удельного расхода материала (на производство одного изделия) составит:

$\Delta M(\Delta n) = M_0 \cdot i_q \cdot (i_n - 1)$ (6)

т.е. сумма затрат в связи с изменением удельного расхода материала определятся базисным (исходным) значением затрат, скорректированным на динамику изменения объема выпуска, умноженным на величину прироста / или снижения удельного расхода материала, определяемую показателем ($i_n - 1$).

Изменение суммы затрат в связи с изменением цены материала равно

$\Delta M(\Delta p) = M_1 - M_0 - \Delta M(\Delta q) - \Delta M(\Delta n) = M_1 - M_0 - [M_0 \cdot (i_q - 1)] - [M_0 \cdot i_q \cdot (i_n - 1)] = M_0 \cdot i_q \cdot i_n \cdot i_p - M_0 - M_0 \cdot i_q + M_0 - M_0 \cdot i_q \cdot i_n + M_0 \cdot i_q = M_0 \cdot i_q \cdot i_n \cdot (i_p - 1)$.

Пример 1. Рассмотрим численный пример.

Пусть M_0 = 16 млн руб., M_1 = 26,4 млн руб., i_q = 1,2; i_n = 1,25; i_p = 1,1.

Общая сумма затрат на производство в абсолютном выражении возросла с базисного по отчетный период на 10,4 млн руб.

$\Delta M = M_1 - M_0 = \Delta M(\Delta q) + \Delta M(\Delta n) + \Delta M(\Delta p) = 26{,}4 - 16 = 10{,}4$ млн руб.

Выясним за счет какого параметра (q, n или p) произошел рос затрат и на сколько.

Как обычно принято в индексном анализе полагаем, что сначала изменяются количественные показатели (q и n), а затем качественный показатель (p). Т.е. первым у нас будет меняться q – объем производства продукции ($q_0 \to q_1$, при этом n и p пока не изменились, т.е. зафиксированы на базисных уровнях, n_0 и p_0) и мы выясняем вклад в общий прирост суммы затрат на материал только количественного фактора – q (объема производства продукции);

Тогда величина $i_M(\Delta q) = \frac{q_1 \cdot n_0 \cdot p_0}{q_0 \cdot n_0 \cdot p_0}$ – покажет изменение суммы затрат на материал в связи с изменением объема производства;

На втором шаге у нас меняется следующий количественный показатель – n (расход материала на производство одного изделия), т.е. n переходит из $n_0 \to n_1$; при этом, q – уже изменилось, т.е. объем производства уже новый, составляет q_1, а цены прежние p_0). Здесь мы выясняем вклад в общий прирост суммы затрат только количественного фактора – n (объем материала на производство одной единицы изделия);

Тогда величина $i_M(\Delta n) = \frac{q_1 \cdot n_1 \cdot p_0}{q_1 \cdot n_0 \cdot p_0}$ – покажет изменение суммы затрат на материал в связи с изменением удельного расхода материала на производство одной единицы изделия;

И в заключение меняется цена единицы материала – p, т.е. p переходит из $p_0 \to p_1$; при этом, q – уже изменилось, т.е. объем производства уже новый – q_1, и расход материала на производство одного изделия также уже новый – n_1. Здесь мы выясняем вклад в общий прирост суммы затрат только качественного фактора – p (цена единицы материала);

Тогда величина $i_M(\Delta p) = \frac{q_1 \cdot n_1 \cdot p_1}{q_1 \cdot n_1 \cdot p_0}$ – покажет изменения суммы затрат на материал в связи с изменением цены единицы материала.

Чтобы проверить, что мы разложили индекс динамики общей суммы материальных затрат на верные три индекса, отражающие три разных эффекта (Δq, Δn и Δp), необходимо проверить, что произведение трех индексов дает нам индекс динамики общей суммы затрат.

Проверка:

$$i_M = \frac{M_1}{M_0} = \frac{q_1 \cdot n_1 \cdot p_1}{q_0 \cdot n_0 \cdot p_0} = i_{M(\Delta q)} \cdot i_{M(\Delta n)} \cdot i_{M(\Delta p)} = \frac{q_1 n_0 p_0}{q_0 n_0 p_0} \cdot \frac{q_1 n_1 p_0}{q_1 n_0 p_0} \cdot \frac{q_1 n_1 p_1}{q_1 n_1 p_0} = \frac{q_1 n_1 p_1}{q_0 n_0 p_0}$$

т.е. разложение индекса i_M верное, теперь можно вычислить все три индекса $i_{M(\Delta q)}$, $i_{M(\Delta n)}$, $i_{M(\Delta p)}$ и измерить влияние каждого фактора (q, n, p) в отдельности на динамику изменения общих суммарных затрат на производство.

Изменение суммы затрат в связи с изменением объема производства (Δq) определим используя индекс $i_M(\Delta q) = \frac{q_1 \cdot n_0 \cdot p_0}{q_0 \cdot n_0 \cdot p_0}$

Изменение в абсолютном выражении вычислим как:

$$\Delta M(\Delta q) = q_1 n_0 p_0 - q_0 n_0 p_0 = q_1 n_0 p_0 \frac{q_0}{q_0} - q_0 n_0 p_0 = M_0 \frac{q_1}{q_0} - M_0 = M_0 \left(i_q - 1\right)$$

Для нашей задачи

$$\Delta M(\Delta q) = 16 \cdot \left(\frac{q_1}{q_0} - 1\right) = 16 \cdot \left(i_q - 1\right) = 16 \cdot (1{,}2 - 1) = 16 \cdot 0{,}2 = 3{,}2 \text{ млн руб.}$$

т.е. суммарные затраты на материал возросли на 3,2 млн руб. в связи с изменением объема производства продукции.

Изменение суммы затрат в связи с изменением удельного расхода материала (на производство одного изделия), т.е. $\Delta M(\Delta n)$ определяет индекс $i_M(\Delta n) = \frac{q_1 n_1 p_0}{q_1 n_0 p_0}$. В абсолютном выражении данный показатель равен

$$\Delta M(\Delta n) = q_1 n_1 p_0 - q_1 n_0 p_0 = q_1 n_1 p_0 \cdot \frac{q_0}{q_0} \cdot \frac{n_0}{n_0} - q_1 n_0 p_0 \cdot \frac{q_0}{q_0} = M_0 \cdot i_q i_n - M_0 \cdot i_q = M_0 \cdot i_q \cdot (i_n - 1)$$

Для условий задачи

$$\Delta M(\Delta n) = M_0 \cdot i_q \cdot (i_n - 1) = 16 \cdot 1{,}2 \cdot (1{,}25 - 1) = 4{,}8 \text{ млн руб.}$$

Изменение суммы затрат в связи с изменением цены материала определяет индекс $i_M(\Delta p) = \frac{q_1 n_1 p_1}{q_1 n_1 p_0}$

В абсолютном выражении данный показатель равен

$$\Delta M(\Delta p) = q_1 n_1 p_1 - q_1 n_1 p_0 = M_1 - q_1 n_1 p_0 \cdot \frac{p_1}{p_1} = M_1 - M_1 \cdot \frac{1}{i_p} = M_1 \cdot \left(1 - \frac{1}{1{,}1}\right) = 26{,}4 \cdot \left(1 - \frac{1}{1{,}1}\right)$$
$$= 2{,}4 \text{ млн руб.}$$

Желающие могут проверить, что $\Delta M(\Delta p) = M_0 \cdot i_q \cdot i_n \cdot \left(i_p - 1\right) = M_1 \cdot \left(1 - \frac{1}{1{,}1}\right) =$ 2,4 млн руб.

Суммарно $\Delta \text{M} = M_1 - M_0 = \Delta M(\Delta q) + \Delta M(\Delta n) + \Delta M(\Delta p) = 26{,}4 - 16 = 3{,}2 + 4{,}8 + 2{,}4 =$ 10,4 млн руб.

Пример 2 для самостоятельного решения. Общий объем выпуска продукции в месяц предприятием (Q – объем выпуска) зависит от количества сотрудников предприятия (M – число сотрудников), количества рабочих дней в месяце (D – количество рабочих дней в месяце), количества часов работы в среднем в день (P – среднее количество часов работы в день), объема производства продукции в час (H – объем выпуска продукции в час одним сотрудником).

т.е. $Q = M \cdot D \cdot P \cdot H$

Сравнивая объем выпуска материальных затрат в отчетном периоде (Q_1) с суммой материальных затрат базисного периода (Q_0), получаем:

$$i_Q = \frac{Q_1}{Q_0} = \frac{M_1 \cdot D_1 \cdot P_1 \cdot H_1}{M_0 \cdot D_0 \cdot P_0 \cdot H_0} = i_N \cdot i_D \cdot i_P \cdot i_H$$

т.е. динамика общего объема выпуска (i_Q) определяется динамикой изменения числа сотрудников предприятия (i_M), динамикой изменения количества рабочих дней в месяце (i_D), динамикой изменения среднего количеств часов работы в день (i_P) и изменением объема выпуска продукции в час одним сотрудником (i_H).

Тогда,

$Q_1 = Q_0 \cdot i_N \cdot i_D \cdot i_P \cdot i_H$

и мы можем разложить на составляющие изменение объема выпуска:

$\Delta Q = Q_1 – Q_0 = \Delta Q(\Delta M) + \Delta Q(\Delta D) + \Delta Q(\Delta P) + \Delta Q(\Delta H)$.

Задайте численные значения величин Q_1 и Q_0, i_M, i_D, i_P, i_H и вычислите значения $\Delta Q(\Delta M)$, $\Delta Q(\Delta D)$, $\Delta \mathrm{Q}(\Delta P)$, $\Delta \mathrm{Q}(\Delta H)$.

ПРИМЕРЫ ДЛЯ САМОСТОЯТЕЛЬНОГО ВЫПОЛНЕНИЯ.

МУЛЬТИПЛИКАТИВНАЯ МОДЕЛЬ

Пример 3 для самостоятельного решения. Общий объем производства продукции в месяц предприятием (Q) зависит от количества сотрудников предприятия (N), среднего количества рабочих дней в месяце (D), доли фактически отработанных дней в месяце (D_1 с учетом отпусков, прогулов, больничных, отпусков за свой счет и пр.), средней продолжительности рабочего дня (h – среднее количество часов работы в день), объема производства продукции в час одним станком (q), доли фактического объема производства продукции в час одним станком (q_1, с учетом простоев из-за ремонта и отсутствия электричества), доли брака в фактически произведенном объеме продукции (q_2), количества станков, обслуживаемых одним сотрудником (K_1).

Сравните общий объем производства продукции без брака за два периода, длину периода уточните сами. Задайте численные значения и разложите на составляющие общее изменение объема производства (т.е. за счет каких переменных произошло изменение и на какую величину).

Пример 4.

При аддитивной связи признаков

$y = x_1 + x_2 + \ldots + x_k$

индексный анализ проводится по следующей формуле (рассмотрим случай из четырех признаков t, a, m, p):

$y_1 = t_1 + a_1 + m_1 + p_1$

$y_0 = t_0 + a_0 + m_0 + p_0$

$$I_y = \frac{y_1}{y_0} = \frac{t_1 + a_1 + m_1 + p_1}{t_0 + a_0 + m_0 + p_0} = \frac{t_1}{t_0} \cdot \frac{t_0}{t_0 + a_0 + m_0 + p_0} + \frac{a_1}{a_0} \cdot \frac{a_0}{t_0 + a_0 + m_0 + p_0} +$$

$$+ \frac{m_1}{m_0} \cdot \frac{m_0}{t_0 + a_0 + m_0 + p_0} + \frac{p_1}{p_0} \cdot \frac{p_0}{t_0 + a_0 + m_0 + p_0} = I_{t_{1/0}} \cdot d_{t_0} + I_{a_{1/0}} \cdot d_{a_0} + I_{m_{1/0}} \cdot d_{m_0} + I_{p_{1/0}} \cdot d_{p_0}$$

Таким образом, общее изменение результата зависит от изменения каждого фактора и доли этого фактора в базисной величине результата. Рассмотрим пример. Общее изменение объема продаж может быть представлено как результат изменения продаж товаров четырех категорий и доли данных категорий товаров в общем объеме продаж базисного периода:

Объемы продаж предприятия

Период	Всего (млн руб.)	В том числе			
		Категория 1	Категория 2	Категория 3	Категория 4
Базисный	1 200	300	200	600	100
Отчетный	1 500	400	150	250	700

$$I_y = \frac{1500}{1200} = 1{,}25 = \frac{400}{300} \cdot \frac{300}{1200} + \frac{150}{200} \cdot \frac{200}{1200} + \frac{250}{600} \cdot \frac{600}{1200} + \frac{700}{100} \cdot \frac{100}{1200}$$
$$= 0{,}33(3) + 0{,}125 + 0{,}208 + 0{,}58(3) = 1{,}25$$

За рассматриваемый период изменилась структура продаж, наибольший рост объема продаж (на 600%) произошел в 4-ой категории товаров, что привело к росту доли продаж с 8,33(3)% в базисном периоде до 46,66(6)% в отчетном периоде. Товары категории 3 напротив стали продаваться в меньшем объеме (снижение на 58,33(3)% относительно базисного периода), в результате доля продаж товаров данной категории сократилась с 50% в базисном периоде до 17% в отчетном периоде.

Несмотря на значительный рост объема продаж товаров 4-ой категории (+600%) общий объем продаж увеличился лишь на 25%. Причина крайне малая доля товаров 4-ой категории в общем объеме продаж в базисном периоде (8,333%), а также значительное снижение объемов продаж товаров 3-ей категории (на 58,33(3)%). По двум другим категориям товаров значительных изменений объемов продаж не произошло.

ПРИМЕРЫ ДЛЯ САМОСТОЯТЕЛЬНОГО ВЫПОЛНЕНИЯ.

АДДИТИВНАЯ МОДЕЛЬ

Пример 7. Проведите индексный анализ следующей аддитивной модели состоящей из пяти признаков (признаки – типы операционных расходов организации):

$y_1 = u_1 + q_1 + s_1 + c_1 + w_1$

Общее изменение операционных расходов организации может быть представлено как результат изменения расходов в пяти группах расходов:

Операционные расходы производственной организации

Период	Всего (тыс. руб.)	В том числе				
		З/п	IT и безопасность	Аренда, проектные расходы	Реклама и прочее	Налоги, Аудит
Базисный	4 700	2 100	1 200	750	400	250
Отчетный	5 400	2 750	990	800	540	320

Пример 8. Проведите индексный анализ следующей аддитивной модели состоящей из шести признаков (признаки – чистые активы компаний холдинговой группы):

$y_1 = u_1 + q_1 + s_1 + c_1 + w_1 + f_1$

Общее изменение чистых активов группы компаний может быть представлено как результат изменений чистых активов компаний, входящих в группу:

Консолидированные чистые активы компаний холдинговой группы

Период	Всего (млн руб.)	В том числе					
		Группа компаний ФИС	Дочерняя компания «Веник»	Дочерняя компания «Авангард»	Дочерняя компания «Яшма»	Дочерняя компания «Клубок»	Дочерняя компания «Кубок»
Базисный	7 800	3 500	1 700	550	690	1 010	350
Отчетный	8 200	3 700	1 600	660	820	1 020	400

МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ВЫПОЛНЕНИЮ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ПО ТЕМЕ ГРУППИРОВКА ДАННЫХ

§ 1. Построение вариационных рядов.

Единицы изучаемой совокупности обладают некоторыми признаками, и эти признаки принимают различные значения для разных единиц совокупности, т.е. в совокупности существует вариация по значениям признаков.

Признаки бывают атрибутивными и количественными. Примером атрибутивных признаков служит, например, пол респондента, уровень образования, профессия. Количественные признаки выражаются в виде чисел, например, уровень заработной платы, возраст и т.д.

Построение вариационного ряда (группировка статистических данных) – построение упорядоченного по возрастанию (или убыванию) распределения единиц совокупности с подсчетом числа единиц с тем или иным значением признака.

Таким образом, для единиц совокупности $i = 1,...n$ вариационный ряд состоит из двух столбцов:

значение варьирующего признака (x_i – варианта);

частоты (f_i – абсолютное число случаев данного варианта x_i) или частости (w – относительная доля каждой частоты в общей сумме частот).

Вариационный ряд бывает трех видов:

ранжированный;

дискретный;

интервальный.

Ранжированный ряд – это перечень отдельных единиц совокупности в порядке возрастания (или убывания) изучаемого признака.

Например, ниже приведены сведения о размере активов пяти крупных компаний города *N* по состоянию на 01.01.2020 г.

Таблица 1.

Распределение компаний по размеру активов (данные условные)

Название компании	**Активы, в млн. руб.**
Компания 1	3 256
Компания 2	2 105
Компания 3	1 129
Компания 4	957
Компания 5	729

В данном примере пять компаний приведены в порядке убывания величины активов (варьирующего признака x_i). Это пример ранжированного по убыванию ряда.

Когда варьирующий признак принимает большое количество значений, невозможно образовать группу для каждого из них и представление данных в виде ранжированного ряда становится громоздким, в этом случае вариационный ряд строят с помощью группировки. Число групп не должно превышать 12 – 15 (при достаточно большом числе наблюдений, например свыше 500).

Для большого числа единиц совокупности, имеющих небольшое число значений признака, строят дискретный вариационный ряд. В качестве примера такого ряда можно рассмотреть распределение точности попадания стрелка в цель мишени (таб. 2).

Таблица 2.

Результаты одного круга стрельбы одного из спортсменов
матчей Чемпионата мира по стрельбе 2021 года

Число попаданий в соответствующее значение мишени, (X_i, значение признака)	Частота (f_i)
10	3
9	3
8	13
7	4
6	12
5	10
4	12
3	2
2	2
1	1
0	1

В частности, данный дискретный ряд наглядно демонстрирует, что в течение Чемпионата мира 2021 года спортсмен попал мимо мишени 1 раз, 10 раз поразил цель с результатом 5 баллов, 13 раз поразил цель с результатом 8 баллов и только 3 раза попал точно в цель.

Если число единиц совокупности достаточно велико и значения признака разнообразны, то представление данных в виде ранжированного или дискретного ряда становится неудобным. В этом случае, а также если варьирующий признак является непрерывной величиной, строят интервальный ряд путем группировки единиц совокупности по значениям изучаемого признака.

Интервальный ряд представляет собой таблицу, состоящую из двух столбцов – интервалов варьирующего признака (x_i), и числа единиц совокупности, попадающих в данный интервал (частот, f_i). Вместо частот могут быть указаны частости (w_i) – доли числа единиц совокупности от общей численности совокупности.

При построении интервального ряда необходимо определить какими будут интервалы. Интервалы бывают:

равные (с равной величиной / шагом h интервалов);

неравные (с неравной величиной / шагом h интервалов);

Пример, неравных интервалов - *равнонаполненные* интервалы (не обязательно будут равными, т.к. основной критерий – равное число единиц совокупности в каждом интервале; границы интервала соответствуют фактическим значениям признака в каждой группе). Равнонаполненные ряды строят для признака со значительной вариацией и для совокупности далекой от нормального распределения.

Кроме того, интервалы могут быть закрытые и открытые. Закрытые интервалы имеют нижнюю и верхнюю границы, открытые имеют только одну границу: либо верхнюю, либо нижнюю. Наличие открытых интервалов нежелательно, но их часто применяют для компактности ряда, объединяя в одну группу крайние случаи совокупности. Например, при построении дискретного ряда по количеству членов семьи, логичным представляется выделить самостоятельные группы по 1, 2, 3, 4, и 5 человек, а редко встречающиеся семьи, включающие от шести человек и более объединить в группу «6 и более».

При построении интервального ряда следует помнить, что единиц в группе (в каждом интервале) должно быть не менее 3-5, тогда в сводных / групповых показателях погашаются случайные и проявляются закономерные / типичные характеристики.

При определении интервалов следует обращать внимание на то, чтобы в каждом интервале содержалось не более пятидесяти процентов всех единиц совокупности.

При подсчете единиц в группе / интервале принято считать, что если единица обладает значением признака, равным величине верхней границы интервала, то ее следует отнести к следующему интервалу. Хотя можно относить данную единицу и к текущему интервалу, но данный подход используется реже; потому, необходимо заранее предупредить читателя об использовании такого способа подсчета единиц в группе / интервале.

Если изучаемая совокупность достаточно однородна или близка к нормальному распределению, то ряд удобно разбить на равные интервалы.

Для построения равноинтервального вариационного ряда (ряда распределения), используют технический подход, основанный на формуле Стерджесса. Опишем последовательные шаги данного подхода ниже, используя в качестве примера данные о размере кредитного портфеля 55 средних банков России по состоянию на 01 января 2021 года (см. таб. 3). Для построения равноинтервального ряда рассмотрим банки с размером совокупного кредитного портфеля без учета МБК от 118 до 999 млн руб. (таб. 3).

На первом шаге необходимо упорядочить значения признака (x_i) в порядке возрастания или убывания. Мы упорядочим наши данные по возрастанию и привели в таблице 3 первые 11 банков с названиями, а остальные просто пронумеровали.

Таблица 3.

Совокупный кредитный портфель без учета МБК на 01.01.2021, млн руб.

Наименование банка		№п/п		№п/п		№п/п		№п/п	
ООО КБ "Уралфинанс"	118	12	299	23	493	34	647	45	839
АО "ИК Банк"	161	13	300	24	519	35	654	46	850

АО "НДБанк"	217	14	302	25	531	36	665	47	870
АО "Банк БЖФ"	218	15	312	26	547	37	701	48	870
Банк "КУБ" (АО)	224	16	372	27	579	38	703	49	890
ООО "АЛТЫНБАНК"	235	17	376	28	597	39	719	50	890
ПАО КБ "Сельмашбанк"	238	18	404	29	599	40	731	51	937
ООО КБ "Столичный Кредит"	260	19	415	30	603	41	731	52	981
КБ "Новый век" (ООО)	265	20	447	31	620	42	734	53	981
АО НКБ "СЛАВЯНБАНК"	266	21	471	32	622	43	754	54	994
ООО КБ "Дружба"	276	22	471	33	639	44	793	55	999

Следующим шагом определим количество интервалов, на которые целесообразно разделить рассматриваемую совокупность. Число интервалов следует выбирать таким образом, чтобы разнообразие значений признака (x_i) в совокупности было отражено в полной мере. В тоже время случайные колебания должны сглаживаться в интервальном ряду. Таким образом, при недостаточном количестве числа групп не проявятся закономерности вариации в распределении. При излишнем числе групп случайные колебания / скачки частот исказят форму распределения. Число интервалов / групп можно определить экспертно, или используя способ последовательно возрастающих интервалов, или на основании формулы Стерджесса[4]. Чаще всего при построении вариационного ряда число групп / интервалов (n) определяют по формуле Стерджесса:

$n = 1 + 3{,}322 \cdot lgN$, где

$lg\ N = log_{10}N$ (логарифм по основанию 10),

N – число единиц совокупности (число значений изучаемого признака X_i), в нашем случае N = 55.

Число интервалов для N = 55 равно $lg\ N = log_{10}64 = 6{,}7815 \approx 7$.

На следующем шаге определяем величины равного интервала (h) используя формулу:

$h = (x_{min} – x_{\max}) / n$, где

x_{max}, x_{min} – максимальное и минимальное значение признака x_i;

$x_{max} – x_{min} = R$ называют размах интервала.

Минимальное и максимальное значение признака рассматриваемой совокупности, x_{min} = 118 млн руб., x_{max} = 999 млн руб.

Тогда величина равного интервала (h) равна:

$h = (x_{min} – x_{\max}) / n = (999 – 118) / 7 = 125{,}857$ млн руб.

[4]В процессе построения ряда количество групп, изначально определенное по формуле Стерджесса, может изменятся; например, в случае если в группу попадают менее 3-х значений признака количество групп следует сократить.

Для построения ряда удобно использовать шаг интервала в целых единицах, более того, шаг и границы интервалов принято округлять. Округлим нижнюю границу интервала до 110 млн руб., а шаг до 130 млн руб. [5]

При подсчете единиц в интервале договоримся, что будем следовать наиболее часто используемому правилу, что если единица обладает значением признака, равным величине верхней границы интервала, то ее следует отнести к следующему интервалу.

В соответствии с данным подходом мы получим следующее распределение (интервальный ряд) 55 банков по размеру кредитного портфеля (см. таб. 4).

Таблица 4.

Распределение 55 средних банков России по совокупному кредитному портфелю без учета МБК на 01.01.2021, млн руб.

Размер кредитного портфеля, в млн руб.	Число банков, f_i	Частости, w_i	Накопленные частоты, S_i	Накопленные частости, w_i
110 – 240	7	0,127	7	0,127
240 – 370	8	0,145	15	0,272
370 – 500	8	0,145	23	0,417
500 – 630	9	0,164	32	0,581
630 – 760	11	0,200	43	0,781
760 – 890	7	0,127	50	0,908
890 – 1020	5	0,092	55	1
Итого	55	1		

Изобразим полученный нами интервальный ряд в виде гистограммы. Для этого на оси X отложим отрезки, равные длине интервалов. Над этими отрезками построим прямоугольники высотой равные частоте (можно также использовать частости).

[5] При расчете величины первого интервала в некоторых источниках размер первого интервала может быть вычислен по формуле $h + x_{min}$ [Елисеева И.И., С.144 - 147]. В других источниках рекомендуют отступить от минимального значения примерно на половину длины интервала, и не вычислять размер данного интервала как $h + x_{min}$ [Громыко Г.Л., С.84].

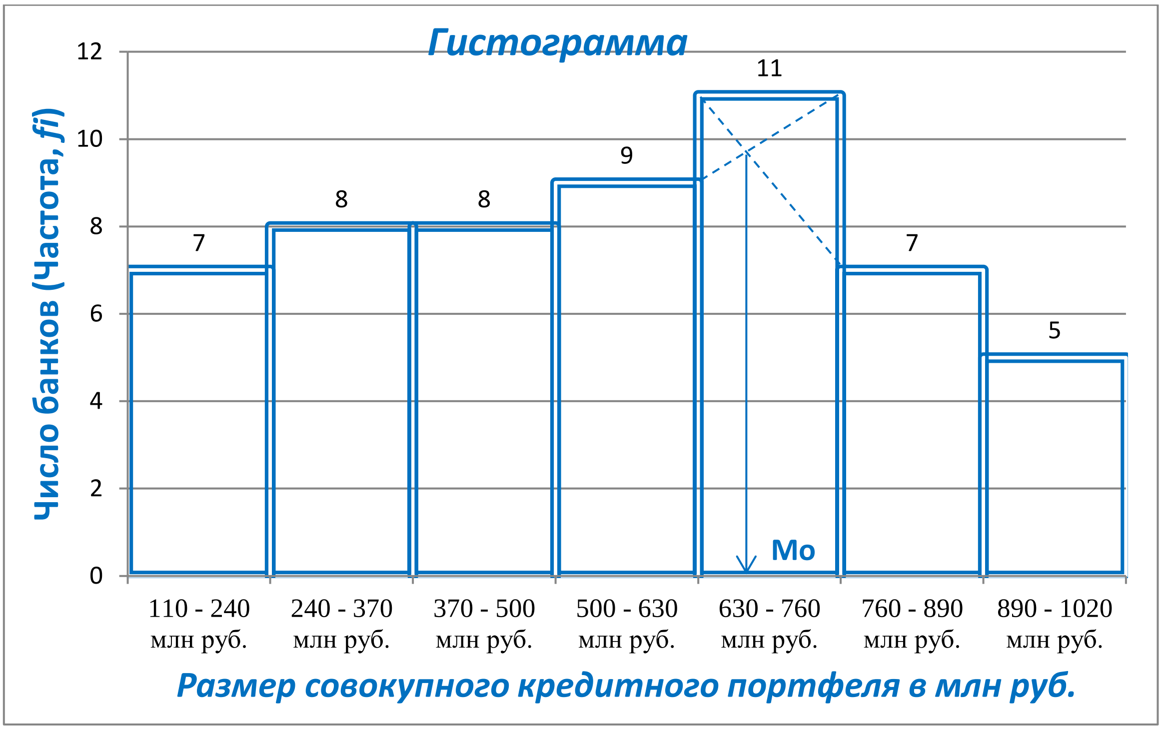

Рис. 1. Распределение банков по размеру совокупного кредитного портфеля

Построим кумулятивную кривую (кумуляту) для нашего ряда распределения. Для этого по оси *X* отложим верхние границы интервалов, а по оси *Y* накопленные частоты (S_i). Так точка (110; 0) на рисунке 2 говорит о том, что в изучаемой совокупности нет банков с размером кредитного портфеля 110 млн руб. или менее. Точка (240; 7) означает, что 7 банков имеют размер совокупного кредитного портфеля от 110 до 240 млн руб.

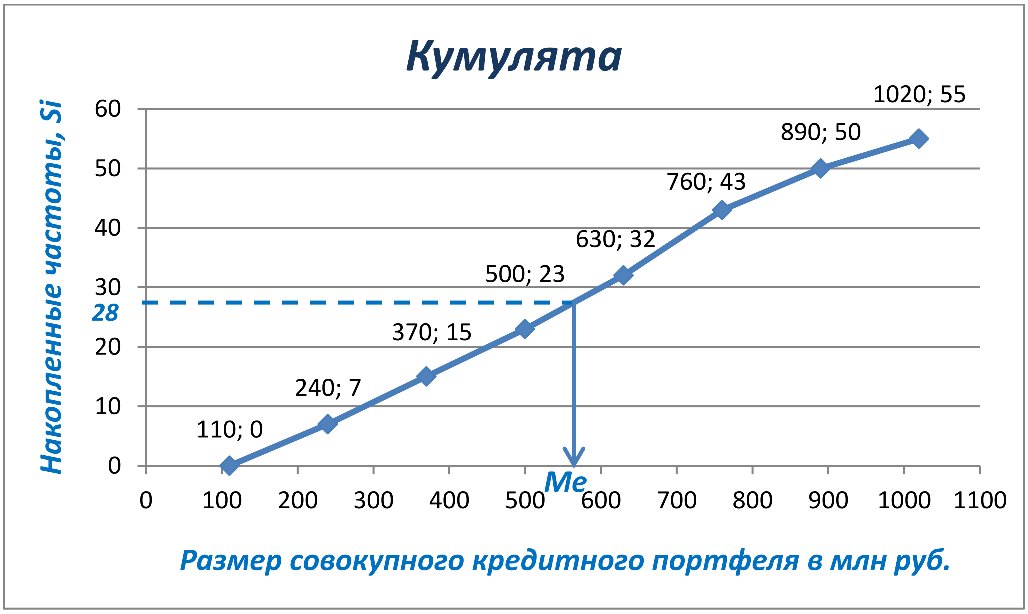

Рис. 2. Кумулятивная кривая

Поменяв местами оси *X* и *Y* в кумулятивной кривой мы получим огиву.

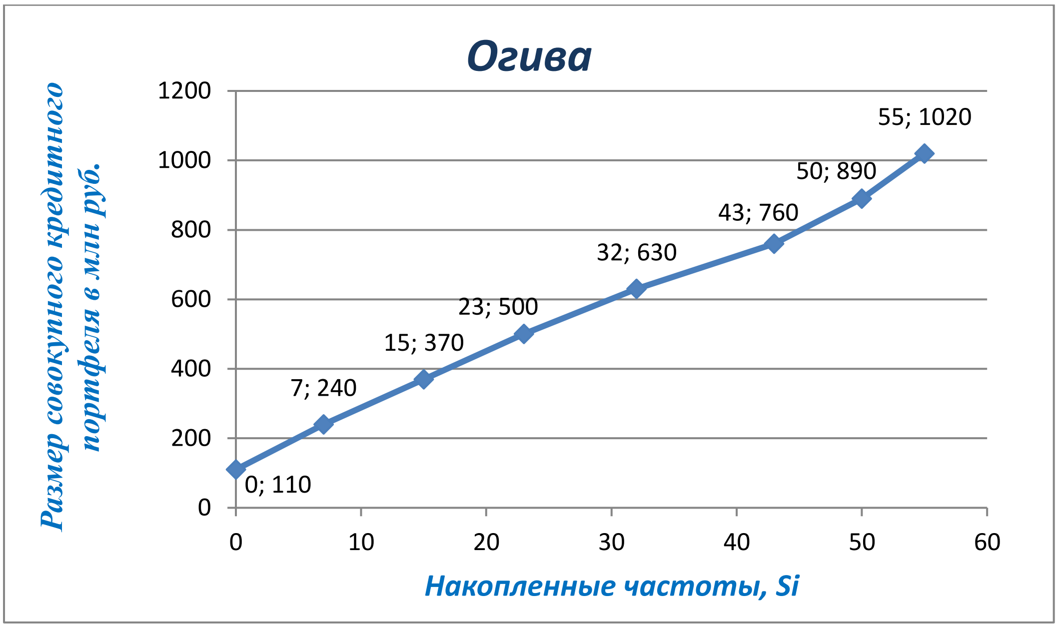

Рис. 3. Огива.

§ 2. Вычисление Моды и Медианы.

Вычислим для данного ряда *моду* (*Мо*) и *медиану* (*Ме*). Предварительно заметим, что графически *моду* наглядно можно изобразить на гистограмме (см. рис. 1), а *медиану* на кумуляте (см. рис.2).

Мода по определению это значение признака наиболее часто встречающееся в вариационном ряду. Обычно встречаются распределения с одним модальным значением. Порой в распределении бывает два или несколько равных (и даже несколько различных, но больших, чем соседние) значений признака, в этом случае распределение считается бимодальным либо мультимодальным. Наличие нескольких мод говорит о неоднородности изучаемой совокупности, возможно состоящей из нескольких совокупностей с разными модами. Обратите на это внимание при выполнении домашнего задания. *Медианой* называют значение признака, которое делит упорядоченную совокупность x_i на две равные по численности части. В итоге у одной половины единиц совокупности значение признака не превышает медианный уровень, а у другого – превышает медианный.

В случае интервального ряда для равных интервалов моду определяют по формуле:

$$Mo = x_{Mo} + h_{Mo}\frac{(f_{Mo}-f_{Mo-1})}{(f_{Mo}-f_{Mo-1})+(f_{Mo}-f_{Mo+1})} \qquad (1.1)$$

где x_{Mo} – нижняя граница модального интервала,

h_{Mo} – ширина модального интервала,

f_{Mo} – частота в модальном интервале,

f_{Mo-1} – частота в интервале, предшествующем модальному,

f_{Mo+1} – частота в интервале, который следует за модальным.

Для случая неравных интервалов моду вычисляют по формуле:

$$Mo = x_{Mo} + h_{Mo} \frac{(f'_{Mo} - f'_{Mo-1})}{(f'_{Mo} - f'_{Mo-1}) + (f'_{Mo} - f'_{Mo+1})} \quad (1.2)$$

где x_{Mo} – нижняя граница модального интервала,

h_{Mo} – ширина модального интервала,

f'_{Mo} – плотность в модальном интервале,

f'_{Mo-1} – плотность в интервале, предшествующем модальному,

f'_{Mo+1} – плотность в интервале, который следует за модальным,

а плотность это отношение частоты к ширине соответствующего интервала $\left(f' = \frac{f}{h}\right)$.

На первом шаге определим тот интервал, в котором содержится мода. В равноинтервальном ряду *мода* содержится в интервале с наибольшей частотой. Как мы рассуждаем? Мы не знаем в каком интервале равноинтервального ряда больше одинаковых значений, но можем сказать, что вероятность найти одинаковые значения больше в том интервале, где больше самих значений, т.е. в интервале с наибольшей частотой. В нашем примере наибольшая частота по числу банков $- f_{max} = 11$, т.е. модальным является интервал 630 – 760. Его нижняя граница $x_{Mo} = 630$, ширина $h_{Mo} = 130$, частота в модальном интервале $f_{Mo} = 11 = f_{max}$, частота в интервале, предшествующем модальному $f_{Mo-1} = 9$, частота в интервале, который следует за модальным $f_{Mo+1} = 7$.

$$Mo = 630 + 130 \cdot \frac{(11-9)}{(11-9)+(11-7)} = 630 + 130 \cdot \frac{2}{6} = 673{,}3(3) \text{ млн руб.}$$

Таким образом, в данном равноинтервальном ряду наиболее часто встречаются банки с размером совокупного кредитного портфеля равным 673,3(3) млн руб.

Для расчета медианы не имеет значения являются интервалы равными или нет, так как медиана определяется как центр упорядоченного ряда. Для ее нахождения нужно всю совокупность поделить пополам и найти в каком интервале содержатся центральное/-ные (при нечетном количестве элементов ряда / при четном количестве элементов ряда) элементы ряда. При расчете медианы интервального ряда используют универсальную формулу:

$$Me = x_{Me} + h_{Me} \cdot \frac{0{,}5 \cdot \sum f_i - S_{Me-1}}{f_{Me}} \quad (1.3)$$

x_{Mo} – нижняя граница интервала, в котором находится медиана,

h_{Me} – величина медианного интервала,

f_i – число единиц изучаемого множества,

$S_{Me\text{-}1}$ – накопленная чистота в интервале, предшествующем медианному,

f_{Me} – частота в медианном интервале.

При расчете медианы важно не забывать соблюдать условие ранжированности (упорядоченности по возрастанию или убыванию) элементов ряда. В нашем случае ряд упорядочен по возрастанию признака x_i (размер совокупного кредитного портфеля).

Для вычисления медианы определим медианный интервал. Для этого воспользуемся накопленными частотами (или частостями). Для наших 55 банков медианным, стоящим в середине, является (N + 1) / 2 = (55 + 10) / 2 = 28-ой банк (x_i)[6]. Из столбца накопленных частот (S_i), таблицы 4, видим, что значение 28 попадает в четвертый интервал (500 – 630)[7]. Если исходить из частостей, то медианным является интервал, в который попадает значение накопленных частостей равное 0,5, также четвертый интервал (500 – 630).

Нижняя граница медианного интервала x_{Me} = 500, ширина медианного интервала h_{Me} = 130, число единиц изучаемого множества Σf_i = 55, накопленная чистота в интервале, предшествующем медианному S_{Me-1} = 23, частота в медианном интервале f_{Me} = 32. Подставляем значения в формулу (1.3), получаем:

$$Me = 500 + 130 \cdot \frac{0{,}5 \cdot 55 - 23}{9} = 565 \text{ млн руб.}$$

Таким образом, банк с размером совокупного кредитного портфеля 565 млн руб. делит рассматриваемую нами совокупность банков на две части, банки с размером активов менее 565 млн руб., и банки с размером активов более 565 млн руб.

Мы построили равноинтервальный ряд, в нем нет интервалов с частотами менее 3-х, что позволяет сгладить случайные характеристики и выявить закономерные (типичные) характеристики изучаемого ряда.

Выбранный нами подход построения равноинтервального ряда в данном случае позволил построить хорошее распределение, единицы совокупности распределены достаточно равномерно по интервалам ряда. Показатели *Mo* и *Me*, построенного нами интервального ряда, также отражают значения данных показателей исходного ряда, что свидетельствует в пользу построенного нами интервального ряда. Количество интервалов выбрано удачно, благодаря чему в достаточной мере отражается разнообразие значений признака и вместе с тем не искажается случайными колебаниями частот форма распределения.

Следует отметить, что технический подход построения интервального ряда работает не всегда. Рассмотрим пример построения равноинтервального ряда для совокупности из 114 банков с уровнем

---

[6] Для 100 наблюдений медианными, стоящими в середине были бы (N + 1) / 2 = (100 + 1) / 2 = 50,5, т.е. 50-ое и 51-ое наблюдения.

[7] Размер кредитного портфеля от 110 до 500 млн руб. имеют 23 банка, а от 110 до 630 уже 32 банка. Таким образом, 28-ой банк попадает в интервал 500 – 630 млн руб.

просроченной задолженности от 5% до 67%. Изучаемый признак – уровень просроченной задолженности. Упорядоченный по значению признака (x_i) ряд представлен в таблице 5.

Таблица 5.

Уровень просроченной задолженности в розничном кредитном портфеле на 01.01.2021, в %

Наименование банка	в %	№п/п	в %	№п/п	в %	№п/п	в %	№п/п	в %	№п/п	в %
АО КБ «ИС Банк»	5,0	20	6,0	39	7,7	58	10,6	77	18,3	96	31,1
ББР Банк (АО)	5,0	21	6,0	40	7,8	59	10,7	78	19,7	97	31,6
ПАО «Банк	5,0	22	6,0	41	7,9	60	11,1	79	19,8	98	32,3
АО «УРАЛПРОМБАНК»	5,1	23	6,1	42	8,0	61	11,1	80	20,0	99	34,0
ПАО «БыстроБанк»	5,2	24	6,5	43	8,0	62	11,3	81	20,7	100	35,6
АО Банк «Венец»	5,2	25	6,5	44	8,3	63	11,4	82	21,5	101	36,9
АО «Банк ДОМ.РФ»	5,3	26	6,5	45	8,6	64	12,6	83	21,9	102	38,4
ПАО Банк ЗЕНИТ	5,3	27	6,6	46	8,7	65	12,7	84	22,6	103	38,7
ПАО «АК БАРС» БАНК	5,4	28	6,6	47	8,7	66	13,2	85	23,0	104	41,3
АО «Газнефтьбанк»	5,4	29	6,6	48	9,0	67	13,4	86	23,3	105	43,0
АО КИБ «ЕВРОАЛЬЯНС»	5,4	30	6,9	49	9,1	68	14,2	87	24,4	106	44,3
АКБ «ТЕНДЕР-БАНК» (АО)	5,6	31	7,0	50	9,3	69	14,4	88	25,8	107	47,7
ООО Банк Оранжевый	5,7	32	7,0	51	9,3	70	15,2	89	26,0	108	48,0
КБ «ЭНЕРГОТРАНСБАНК»	5,8	33	7,1	52	9,5	71	15,5	90	26,7	109	57,7
АО «АЛЬФА-БАНК»	5,9	34	7,2	53	9,5	72	15,5	91	27,3	110	59,0
КБ «ЛОКО-Банк» (АО)	5,9	35	7,3	54	9,9	73	16,5	92	27,9	111	60,5
АО «Автоградбанк»	5,9	36	7,3	55	10,0	74	16,6	93	29,5	112	60,8
ПАО Банк «Кузнецкий»	5,9	37	7,4	56	10,2	75	17,2	94	29,6	113	63,6
АО «КАБ «Викинг»	5,9	38	7,4	57	10,4	76	17,3	95	29,7	114	67,0

В таблице 5 приведены названия 19 банков, остальные просто пронумерованы.

Число интервалов, рассчитанных по формуле Стерджесса равно:

$n = 1 + 3{,}322 \cdot lgN = 1 + 3{,}322 \cdot log_{10}114 = 7{,}833 \approx 8$.

Величина равного интервала (h):

$h = (x_{min} – \mathrm{x}_{\mathrm{max}}) / n = (67 – 5) / 8 = 7{,}75$

Для построения ряда удобно использовать шаг интервала в целых единицах, более того, шаг и границы интервалов принято округлять. Округлим шаг интервала до 8 %.

При подсчете единиц в интервале договоримся, что будем следовать наиболее часто используемому правилу, что если единица обладает значением признака, равным величине верхней границы интервала, то ее следует отнести к следующему интервалу.

Используя данный подход для построения равноинтервального ряда для совокупности из 114 банков с уровнем просроченной задолженности от 5% до 67% на первом шаге мы получим следующий равноинтервальный ряд:

Таблица 6.

Распределение 114 банков России по уровню просроченной задолженности на 01.01.2021, в %

Уровень просроченной задолженности, в %	Число банков, f_i	Частости, w_i	Накопленные частоты, S_i
5 – 13	65	0,5702	65
13 – 21	16	0,1404	81
21 – 29	11	0,0965	92
29 – 37	9	0,0789	101

37 – 45	5	0,0439	106
45 – 53	2	0,0175	108
53 – 61	4	0,0351	112
61 – 69	2	0,0175	114
Итого	114	1	

Заметим, что данный ряд содержит интервалы, включающие менее трех значений признака, это 6-ой и 8-ой интервалы. Вместе с тем, первый интервал содержит более 50% значений признака (а именно 57%, 65 из 114,). Оба показателя говорят о том, что выбранный нами подход построения равноинтервального ряда не позволяет нам построить хорошее распределение.

Изменяя количество интервалов, мы не сможем преодолеть обе эти проблемы. С одной стороны, уменьшая количество интервалов до 7 и более мы добьёмся результата при котором в каждом интервале будет от 3-х значений признака и более, но при этом размах первого интервала также будет увеличиваться и доля значений признака, включенных в данный интервал будет увеличиваться далее. Что при ведет к построению ряда с недопустимыми характеристиками (один из интервалов содержит более 50% значений признака). Увеличивая количество интервалов мы сможем добиться результата при котором в первом интервале будет содержаться менее 50% всех значений признака, но при этом останутся интервалы содержащие менее 3-х значений признака. Что также приведет к построению ряда с недопустимыми характеристиками (один из интервалов содержит менее 3-х значений признака).

Желающие могут убедиться, что постепенно уменьшая количество интервалов с 8 до 5, и увеличивая размах интервала, мы построим равноинтервальный ряд и в нем не будет интервалов с частотами менее трех.

Это интервал с шагом:

$h = (x_{min} - x_{max}) / n = (67 - 5) / 5 = 12{,}4$ %, округлим данное значение до 13 %, получаем распределение табл. 7.

Таблица 7.

Распределение 114 банков России по уровню просроченной задолженности на 01.01.2021, в %

Уровень просроченной задолженности, в %	Число банков, f_i	Частости, w_i	Накопленные частоты, S_i
5 – 18	76	0,666(6)	76
18 – 31	19	0,166(6)	95
31 – 44	10	0,0877	105
44 – 57	3	0,0263	108
57 – 70	6	0,0526	114
Итого	114	1	

Вместе с тем стоит отметить, что выбранный нами подход построения равноинтервального ряда в данном случае не позволил построить хорошее распределение, так около 67% единиц всей совокупности оказались в первом интервале. Показатели *Mo* и *Me* (см. значения ниже по тексту), построенного нами интервального ряда, также значительно отличаются от истинных значений данных показателей исходного ряда, что свидетельствует не в пользу построенного нами равноинтервального ряда.

$$Mo = 5 + 13 \cdot \frac{(76 - 0)}{(76 - 0) + (76 - 19)} = 5 + 13 \cdot \frac{76}{133} = 12{,}429\ \%$$

$$Me = 5 + 13 \cdot \frac{0{,}5 \cdot 114 - 0}{76} = 14{,}75\ \%$$

В исходном распределении (см. таб. 5) значение *моды* 5,9%, а *медиана* принимает значение 10,5%.

В целом, разбиение на равные интервалы дает хорошее результаты если изучаемая совокупность достаточно однородна или близка к нормальному распределению. Наши исходные данные (см. таб. 5) таковыми не являются.

В случае если подход, основанный на формуле Стерджесса, не дает хорошего результата, для разбиения совокупности на интервалы строят ряд с последовательно возрастающими интервалами или используют экспертный подход определения количества интервалов ряда. Эти подходы позволяют получить приемлемое распределение для совокупностей имеющих значительный разброс значений. Построим ряд с возрастающими интервалами по варьирующему признаку x_i – размер просроченной задолженности в розничном кредитном портфеле для совокупности 114 банков России. Для нашей совокупности банков значения x_i изменяются от 5% до 67%. В данном случае совокупность банков можно разделить на следующие интервалы (см. таб. 8).

Таблица 8.

Распределение 114 банков России по уровню просроченной задолженности на 01.01.2021, в %

Уровень просроченной задолженности, в %	Число банков, (частоты), f_i	Шаг интервала, h	Частости, w_i	Накопленные частоты, S_i	Накопленные частости, w_i	Плотность распределения, f'_i
5 – 6	19	1	0,166(6)	19	0,166(6)	19
6 – 7	11	1	0,0965	30	0,2632	11
7 – 9	17	2	0,1491	47	0,4123	8,5
9 – 12	16	3	0,1404	63	0,5526	5,(3)
12 – 20	16	8	0,1404	79	0,6930	2,0
20 – 30	16	10	0,1404	95	0,833(3)	1,6
30 – 50	13	20	0,1140	108	0,9474	0,65
50 – 70	6	20	0,0526	114	1	0,3
Итого	114		1			

В данном распределении шаг интервала последовательно возрастает от 1 до 20. Частоты варьируются незначительно, т.е. данные распределены по интервалам достаточно равномерно. Плотность распределения говорит о наполненности каждого интервала ряда.

Вычислим *моду* данного распределения. Напомним, что для случая неравных интервалов *моду* вычисляют по формуле (1.2), заменяя в исходной формуле (1.1) частоты плотностями распределения. В нашем случае, в силу того, что модальным является первый интервал [5 – 6) и для первых двух интервалов частоты и плотности совпадают, две формулы становятся эквивалентными.

$$Mo = x_{Mo} + h_{Mo} \cdot \frac{(f'_{Mo} - f'_{Mo-1})}{(f'_{Mo} - f'_{Mo-1}) + (f'_{Mo} - f'_{Mo+1})} == 5 + 1 \cdot \frac{(19-0)}{(19-0)+(19-11)} = 5 + 1 \cdot \frac{19}{27}$$
$$= 5{,}704\ \%$$

Медиану вычислим по формуле (1.3):

$$Me = 9 + 3 \cdot \frac{0{,}5 \cdot 114 - 47}{16} = 9 + 3 \cdot \frac{10}{16} = 10{,}875\ \%$$

Таким образом, в данном интервальном ряду наиболее часто встречаются банки с уровнем просроченной задолженности 5,704%. А медианным является банк с размером просроченной задолженности 10,875%.

Наш ряд с последовательно возрастающими интервалами не содержит интервалов с частотами менее 3-х, и интервалов, содержащих значительное количество значений признака (от 50% и более), что позволяет сгладить случайные характеристики и выявить закономерные / типичные характеристики изучаемого ряда. Значения признака достаточно равномерно распределены по интервалам ряда. Показатели *Мо* и *Ме* близки к соответствующим значениям исходного ряда – *мода* 5,9%, а *медиана* 10,5%, что свидетельствует в пользу построенного нами интервального ряда. Данный подход – построение ряда с последовательно возрастающими интервалами позволил построить хорошее распределение для изучаемой совокупности данных.

§ 3. Построение дискретного ряда, нахождение *Mo* и *Me*.

Построить дискретный ряд означает провести группировку статистических данных и представить их в виде таблицы, в одном столбце которой указаны значения варьирующего признака (x_i – варианта), в другом частоты (f_i – абсолютное число случаев данного варианта x_i) или частости (w_i – относительная доля каждой частоты в общей сумме частот).

Построим дискретный ряд распределения, используя в качестве вариантного признака (x_i) данные о доле рынка по размеру активов (см. таб. 9). Значения варьирующего признака возьмем в интервале от 0,001% до 0,010%, всего 114 банков.

Таблица 9.

Доля рынка по размеру активов 114 банков России по состоянию на 01.01.2021, в %

Наименование банка	в %	№п/п	в %	№п/п	в %	№п/п	в %	№п/п	в %	№п/п	в %
Банк «Снежинский» АО	0,010	20	0,006	39	0,004	58	0,003	77	0,002	96	0,001
«ЗИРААТ БАНК (МОСКВА)» (АО)	0,009	21	0,006	40	0,004	59	0,002	78	0,002	97	0,001
КБ «Гарант-Инвест» (АО)	0,009	22	0,006	41	0,004	60	0,002	79	0,002	98	0,001
ПАО УКБ «Новобанк»	0,009	23	0,006	42	0,004	61	0,002	80	0,002	99	0,001
ПАО «Томскпромстройбанк»	0,008	24	0,006	43	0,004	62	0,002	81	0,002	100	0,001
АО «Солид Банк»	0,008	25	0,006	44	0,004	63	0,002	82	0,002	101	0,001
ООО «Хакасский муниципальный банк»	0,008	26	0,005	45	0,004	64	0,002	83	0,002	102	0,001
«Северный Народный Банк» (ПАО)	0,008	27	0,005	46	0,004	65	0,002	84	0,002	103	0,001
АО УКБ «Белгородсоцбанк»	0,008	28	0,005	47	0,004	66	0,002	85	0,002	104	0,001
«СИБСОЦБАНК» ООО	0,007	29	0,005	48	0,003	67	0,002	86	0,002	105	0,001
АКБ «Трансстройбанк» (АО)	0,007	30	0,005	49	0,003	68	0,002	87	0,002	106	0,001
АО «Банк «Агророс»	0,007	31	0,005	50	0,003	69	0,002	88	0,002	107	0,001

АО «Кузнецкбизнесбанк»	0,007	32	0,005	51	0,003	70	0,002	89	0,002	108	0,001
ПАО Ставропольпромстройбанк	0,007	33	0,004	52	0,003	71	0,002	90	0,002	109	0,001
АО «Углеметбанк»	0,007	34	0,004	53	0,003	72	0,002	91	0,002	110	0,001
ПАО Банк «Кузнецкий»	0,007	35	0,004	54	0,003	73	0,002	92	0,001	111	0,001
ПАО «Бест Эффортс Банк»	0,007	36	0,004	55	0,003	74	0,002	93	0,001	112	0,001
ООО Банк Оранжевый	0,006	37	0,004	56	0,003	75	0,002	94	0,001	113	0,001
ООО КБ «РостФинанс»	0,006	38	0,004	57	0,003	76	0,002	95	0,001	114	0,001

В таблице 9 приведены названия 19 банков, остальные просто пронумерованы. Дискретный ряд данной совокупности имеет вид:

Таблица 10.

Дискретное распределение 114 банков России по *доле рынка по размеру активов*, по состоянию на 01.01.2021 года

Доля рынка, в % (x_i, значение признака)	Частота, (f_i)	Накопленные частоты, (f_i)
0,010	1	1
0,009	3	4
0,008	5	9
0,007	8	17
0,006	8	25
0,005	7	32
0,004	15	47
0,003	11	58
0,002	**33**	91
0,001	23	114
Итого	114	

Принято считать, что число групп в дискретном ряду не должно превышать 12 – 15, иначе распределение становится громоздким. Для рядов с большим количеством значений признака (x_i) рекомендуется строить интервальный ряд.

Определим *моду* данного дискретного ряда. *Мода* – наиболее часто встречающееся значение признака (наиболее часто встречающаяся варианта x_i). Находим ее по определению. В нашем дискретном ряду наибольшая частота равна 33 (f_{max} = 33), соответственно наиболее часто встречающееся значение признака x_i = 0,002 %, т.е. *Mo* = 0,002 %. Таким образом, среди небольших банков России, с долей рынка от 0,001% до 0,01%, наиболее часто встречаются банки с долей рынка равной 0,002%. Из 114 рассмотренных нами банков, 33 обладают долей рынка 0,002%.

Найдем *медиану* полученного дискретного ряда распределения. *Медианой* называют значение признака, что расположено в центре упорядоченного по возрастанию (или убыванию) ряда.

Номер *медианы* ряда можно определить по формуле:

$$N_{Me} = \frac{N+1}{2}$$

где *N* – количество наблюдений в ряду.

Для дискретного ряда с нечетным количеством элементов центральным значением будет один элемент ряда. Например, для дискретного ряда, состоящего из 49 элементов, центральным / медианным элементом будет

$$N_{Me} = \frac{N+1}{2} = \frac{49+1}{2} = 25\text{-ый}.$$

Дискретный ряд с четным числом элементов разделит пополам не один, а два элемента ряда. Для дискретного ряда, состоящего из 114 банков, N_{Me} = 115 / 2 = 57,5. Таким образом, медианное значение данного ряда находится между 57 и 58 значениями упорядоченного ряда. В соответствии с накопленными частотами (см. таб. 10) для нашего распределения 57-е и 58-е значения ряда имеют значение 0,003, как результат, Ме = (0,003 + 0,003) / 2 = 0,003 %.

Изобразим полученный дискретный ряд в виде *полигона*. Для этого на оси *x* отложим значения признака (x_i), а на оси *y* значения частот (f_i). Затем соединим полученные отрезки прямой, а из первой и последней точки опустим перпендикуляры на ось *x*.

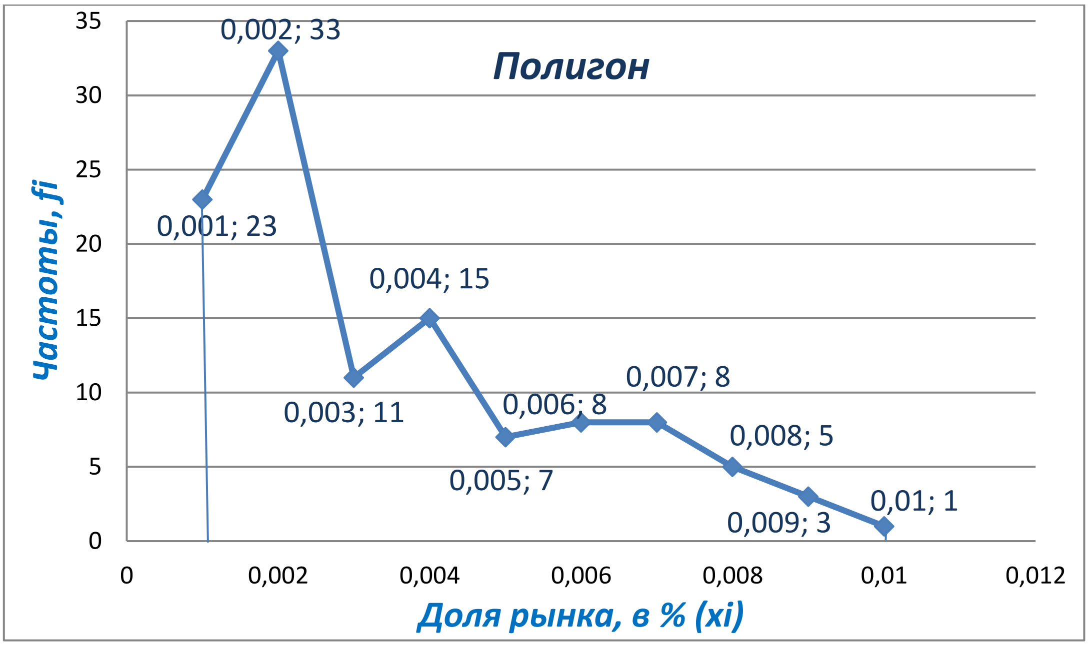

Рис. 4. Полигон распределения.

§ 4. Нахождение среднего, дисперсии и квадратического отклонения дискретного и интервального рядов.

Среднее значение дискретного ряда определяют по формуле средней арифметической взвешенной.

$$\bar{X} = \frac{\sum x_i f_i}{\sum f_i},$$

где x_i – значение осредняемого признака;

f_i – частота, показывающая сколько раз встречается *i*-е значение осредняемого признака.

Для дискретного ряда (таб. 10) среднее значение равно:

$\bar{X} = \frac{\sum x_i f_i}{\sum f_i} =$

$= \frac{0{,}01 \cdot 1 + 0{,}009 \cdot 3 + 0{,}008 \cdot 5 + 0{,}007 \cdot 8 + 0{,}006 \cdot 8 + 0{,}005 \cdot 7 + 0{,}004 \cdot 15 + 0{,}003 \cdot 11 + 0{,}002 \cdot 33 + 0{,}001 \cdot 23}{1 + 3 + 5 + 8 * 2 + 7 + 15 + 11 + 33 + 23}$

$= \frac{0{,}398}{114} = 0{,}00349$

Таким образом, среднее значение *доли рынка по размеру активов* составляет 0,00349%.

Для интервального ряда среднее значение вычисляется по формуле:

$$\bar{X} = \frac{\sum x'_i f_i}{\sum f_i},$$

где x'_i – среднее значение интервала;

f_i – частота, показывающая сколько раз встречается i-е значение осредняемого признака в интервале.

Для интервального ряда (таб. 8) уровень просроченной задолженности составляет:

$$\bar{X} = \frac{\sum x'_i f_i}{\sum f_i} = \frac{5{,}5 \cdot 19 + 6{,}5 \cdot 11 + 8 \cdot 17 + 10{,}5 \cdot 16 + 16 \cdot 16 + 25 \cdot 16 + 40 \cdot 17 + 60 \cdot 6}{114} = \frac{2\,016}{114}$$

$$= 17{,}684\ \%$$

Среднее значение интервального ряда (таб. 8) составляет 17,684%.

Дисперсия (σ^2) ряда может быть вычислена двумя способами, на основе формулы квадратической степенной средней:

$$\sigma^2 = \frac{\sum (x_i - \bar{X})^2}{n},$$

при повторяемости отдельных значений x_i формула примет вид:

$$\sigma^2 = \frac{\sum (x_i - \bar{X})^2 \cdot f_i}{\sum f_i}$$

или на основе формулы преобразованной методом моментов:

$$\sigma^2 = \overline{x^2} - (\bar{X})^2$$

где $\overline{x^2}$ – среднее значение квадратов признака (или начальный *момент второго порядка*),

$\bar{X}$ – среднее значение признака (или начальный *момент первого порядка*).

В случае интервального ряда формулы для расчета дисперсии принимают вид:

$$\sigma^2 = \frac{\sum (x'_i - \bar{X})^2}{n},$$

при повторяемости отдельных значений x_i формула примет вид:

$$\sigma^2 = \frac{\sum (x'_i - \bar{X})^2 \cdot f_i}{\sum f_i}$$

или на основе формулы преобразованной методом моментов:

$$\sigma^2 = \overline{(x')^2} - (\bar{X})^2$$

где x'_i – среднее значение интервала;

Среднее квадратическое отклонение равно корню квадратному из дисперсии, $\sigma = \sqrt{\sigma^2}$.

Для дискретного ряда (таб. 10) дисперсия равна:

$$\sigma^2 = \frac{\sum(x_i - \bar{X})^2 \cdot f_i}{\sum f_i} =$$

$$= \frac{\begin{array}{c}(0{,}01 - 0{,}003491)^2 \cdot 1 + (0{,}009 - 0{,}003491)^2 \cdot 3 + (0{,}008 - 0{,}003491)^2 \cdot 5 + (0{,}007 - 0{,}003491)^2 \cdot 8 + \\ (0{,}006 - 0{,}003491)^2 \cdot 8 + (0{,}005 - 0{,}003491)^2 \cdot 7 + (0{,}004 - 0{,}003491)^2 \cdot 15 + (0{,}003 - 0{,}003491)^2 \cdot 11 \\ +(0{,}002 - 0{,}003491)^2 \cdot 33 + (0{,}001 - 0{,}003491)^2 \cdot 23\end{array}}{1 + 3 + 5 + 8 \cdot 2 + 7 + 15 + 11 + 33 + 23}$$

$$= \frac{622{,}4912}{114} = 5{,}46 \cdot 10^{-6}$$

А среднеквадратическое отклонение $\sigma = \sqrt{\sigma^2} = \sqrt{5{,}46 \cdot 10^{-6}} = 2{,}337 \cdot 10^{-3}$ %.

Для интервального ряда (таб. 8) дисперсия составляет:

$$\sigma^2 = \frac{\sum(x'_i - \bar{X})^2 \cdot f_i}{\sum f_i} =$$

$$= \frac{\begin{array}{c}(5{,}5 - 17{,}684)^2 \cdot 19 + (6{,}5 - 17{,}684)^2 \cdot 11 + (8 - 17{,}684)^2 \cdot 17 + \\ +(10{,}5 - 17{,}684)^2 \cdot 16 + (16 - 17{,}684)^2 \cdot 16 + (25 - 17{,}684)^2 \cdot 16 + (40 - 17{,}684)^2 \cdot 13 \\ +(60 - 17{,}684)^2 \cdot 6\end{array}}{19 + 11 + 17 + 16 + 16 + 16 + 13 + 6}$$

$$= \frac{24\,736{,}13}{114} = 216{,}984$$

А среднеквадратическое отклонение $\sigma = \sqrt{\sigma^2} = \sqrt{216{,}984} = 14{,}73\%$

Расчет дисперсии методом моментов по формуле $\sigma^2 = \overline{x^2} - (\bar{X})^2$ для дискретного ряда дает следующие результаты:

Таблица 11.

Дискретное распределение 114 банков России по *доле рынка по размеру активов*, по состоянию на 01.01.2021 года

Доля рынка, в % (x_i, значение признака)	Частота, (f_i)	Накопленные частоты, (f_i)	Значение признака в квадрате, x_i^2
0,010	1	1	$0{,}1 \cdot 10^3$
0,009	3	4	$0{,}081 \cdot 10^3$
0,008	5	9	$0{,}064 \cdot 10^3$
0,007	8	17	$0{,}049 \cdot 10^3$
0,006	8	25	$0{,}036 \cdot 10^3$
0,005	7	32	$0{,}025 \cdot 10^3$
0,004	15	47	$0{,}016 \cdot 10^3$
0,003	11	58	$0{,}009 \cdot 10^3$
0,002	**33**	91	$0{,}004 \cdot 10^3$
0,001	23	114	$0{,}001 \cdot 10^3$
Итого	114		

$$\overline{x^2} = \frac{\begin{gathered}0{,}1 \cdot 10^3 \cdot 1 + 0{,}081 \cdot 10^3 \cdot 3 + 0{,}064 \cdot 10^3 \cdot 5 + 0{,}049 \cdot 10^3 \cdot 8 + 0{,}036 \cdot 10^3 \cdot 8 + 0{,}025 \cdot 10^3 \cdot 7 \\ +0{,}016 \cdot 10^3 \cdot 15 + 0{,}009 \cdot 10^3 \cdot 11 + 0{,}004 \cdot 10^3 \cdot 33 + 0{,}001 \cdot 10^3 \cdot 23\end{gathered}}{1 + 3 + 5 + 8 \cdot 2 + 7 + 15 + 11 + 33 + 23}$$

$$= \frac{0{,}002012}{114} = 1{,}7649 \cdot 10^{-5}$$

$$\sigma^2 = \overline{x^2} - (\bar{X})^2 = 1{,}7649 \cdot 10^{-5} - (0{,}00349)^2 = 5{,}46 \cdot 10^{-6}$$

А среднеквадратическое отклонение $\sigma = \sqrt{\sigma^2} = \sqrt{5{,}46 \cdot 10^{-6}} = 2{,}337 \cdot 10^{-3}$ %.

Мы убедились, что для дискретного ряда оба способа расчета дисперсии (на основе формулы квадратической степенной средней и формулы, преобразованной методом моментов) совпадают.

Вычислим дисперсию методом моментов для интервального ряда.

Таблица 12.

Распределение 114 банков России по уровню просроченной задолженности на 01.01.2021, в %

Уровень просроченной задолженности, в %	Число банков, (частоты), f_i	Накопленные частоты, S_i	Плотность распределения, f'_i	Середина интервала, x'_i	Значение признака в квадрате, $(x'_i)^2$
5 – 6	19	19	19	5,5	30,25
6 – 7	11	30	11	6,5	42,25
7 – 9	17	47	8,5	8	64
9 – 12	16	63	5,(3)	10,5	110,25
12 – 20	16	79	2,0	16	256
20 – 30	16	95	1,6	25	625
30 – 50	13	108	0,65	40	1 600
50 – 70	6	114	0,3	60	3 600
Итого	114				

$$\overline{(x')^2} = \frac{30{,}25 \cdot 19 + 42{,}25 \cdot 11 + 64 \cdot 17 + 110{,}25 \cdot 16 + 256 \cdot 16 + 625 \cdot 16 + 1\,600 \cdot 13 + 3\,600 \cdot 6}{19 + 11 + 17 + 16 + 16 + 16 + 13 + 6}$$

$$= \frac{60\,387{,}5}{114} = 529{,}715\%$$

$$\sigma^2 = \overline{x^2} - (\bar{X})^2 = 529{,}715 - (17{,}684)^2 = 216{,}984$$

А среднеквадратическое отклонение $\sigma = \sqrt{\sigma^2} = \sqrt{216{,}984} = 14{,}735\%$

Полученные значения дисперсии и среднеквадратического отклонения совпадают с этими значениями, вычисленными на основе формулы квадратической степенной средней.

ВАРИАНТЫ ДОМАШНЕГО ЗАДАНИЯ.

По первичным данным, что вы выбираете самостоятельно[8], требуется составить два ряда распределения – один дискретный, другой интервальный. Дискретный ряд строится по первым трем переменным (x_1, x_2, x_3), чьи значения варьируются незначительно (не более 12-и вариант), интервальный ряд строится по оставшимся трем переменным (x_4, x_5, x_6). По каждому ряду необходимо определить среднюю величину признака, по которому построен вариационный ряд, дисперсию (двумя способами), среднее квадратическое отклонение, моду, и медиану. При этом по дискретному ряду необходимо построить график полигона распределения; по интервальному ряду гистограмму, кумуляту и огиву. Получив все характеристики, написать выводы о качественных особенностях ряда, сопоставить Mo, Me и среднее значение изучаемого признака. Вычислить коэффициент асимметрии и сделать вывод об асимметричности построенного распределения, проверить соответствует ли рассчитанный коэффициент асимметрии данному выводу.

Для вариантов 1 – 10, 31 – 40, 61 – 70, 91 – 100 используйте переменную x_4 для построения интервального ряда и x_1 для построения дискретного ряда распределения.

Предварительно отсортируйте данные по переменной x_1 по возрастанию признака. Затем проводите построение рядов на основе данных с порядковыми номерами, соответствующими вашему варианту.

Вариант 1. Построения рядов проведите на основе данных с порядковыми номерами 1 – 142.

Вариант 2. Построения рядов проведите на основе данных с порядковыми номерами 35 – 146.

Вариант 3. Построения рядов проведите на основе данных с порядковыми номерами 35 – 167.

Вариант 4. Построения рядов проведите на основе данных с порядковыми номерами 53 – 167.

Вариант 5. Построения рядов проведите на основе данных с порядковыми номерами 74 – 179.

Вариант 6. Построения рядов проведите на основе данных с порядковыми номерами 91 – 179.

Вариант 7. Построения рядов проведите на основе данных с порядковыми номерами 91 – 191.

Вариант 8. Построения рядов проведите на основе данных с порядковыми номерами 124 – 212.

Вариант 9. Построения рядов проведите на основе данных с порядковыми номерами 143 – 212.

Вариант 10. Построения рядов проведите на основе данных с порядковыми номерами 124 – 221.

Вариант 31. Построения рядов проведите на основе данных с порядковыми номерами 53 – 212.

[8] Согласование с преподавателем потребуется для выбора совместно всей группой данных с количеством наблюдений не менее 233, и количеством переменных не менее 6-ти; три переменные должны иметь не более 12 вариант (различных значений), оставшиеся три пусть варьируются значительно.

Вариант 32. Построения рядов проведите на основе данных с порядковыми номерами 53 – 221.

Вариант 33. Построения рядов проведите на основе данных с порядковыми номерами 74 – 221.

Вариант 34. Построения рядов проведите на основе данных с порядковыми номерами 74 – 212.

Вариант 35. Построения рядов проведите на основе данных с порядковыми номерами 74 – 215.

Вариант 36. Построения рядов проведите на основе данных с порядковыми номерами 180 – 233.

Вариант 37. Построения рядов проведите на основе данных с порядковыми номерами 168 – 233.

Вариант 38. Построения рядов проведите на основе данных с порядковыми номерами 147 – 233.

Вариант 39. Построения рядов проведите на основе данных с порядковыми номерами 143 – 233.

Вариант 40. Построения рядов проведите на основе данных с порядковыми номерами 124 – 203.

Вариант 61. Построения рядов проведите на основе данных с порядковыми номерами 124 – 233.

Вариант 62. Построения рядов проведите на основе данных с порядковыми номерами 124 – 215.

Вариант 63. Построения рядов проведите на основе данных с порядковыми номерами 204 – 231.

Вариант 64. Построения рядов проведите на основе данных с порядковыми номерами 204 – 233.

Вариант 65. Построения рядов проведите на основе данных с порядковыми номерами 168 – 233.

Вариант 66. Построения рядов проведите на основе данных с порядковыми номерами 124 – 231.

Вариант 67. Построения рядов проведите на основе данных с порядковыми номерами 1 – 146.

Вариант 68. Построения рядов проведите на основе данных с порядковыми номерами 1 – 167.

Вариант 69. Построения рядов проведите на основе данных с порядковыми номерами 74 – 233.

Вариант 70. Построения рядов проведите на основе данных с порядковыми номерами 53 – 233.

Вариант 91. Построения рядов проведите на основе данных с порядковыми номерами 147 – 231.

Вариант 92. Построения рядов проведите на основе данных с порядковыми номерами 124 – 231.

Вариант 93. Построения рядов проведите на основе данных с порядковыми номерами 124 – 233.

Вариант 94. Построения рядов проведите на основе данных с порядковыми номерами 53 – 233.

Вариант 95. Построения рядов проведите на основе данных с порядковыми номерами 53 – 215.

Вариант 96. Построения рядов проведите на основе данных с порядковыми номерами 53 – 203.

Вариант 97. Построения рядов проведите на основе данных с порядковыми номерами 53 – 212.

Вариант 98. Построения рядов проведите на основе данных с порядковыми номерами 74 – 191.

Вариант 99. Построения рядов проведите на основе данных с порядковыми номерами 74 – 167.

Вариант 100. Построения рядов проведите на основе данных с порядковыми номерами 74 – 197.

Для вариантов 11 – 20, 41 – 50, 71 – 80 используйте переменную x_5 для построения интервального ряда и x_2 для построения дискретного ряда.

Предварительно отсортируйте данные по переменной x_2 по возрастанию признака. Затем проводите построение рядов на основе данных с порядковыми номерами, соответствующими вашему варианту.

Вариант 11. Построения рядов проведите на основе данных с порядковыми номерами 1 – 81.

Вариант 12. Построения рядов проведите на основе данных с порядковыми номерами 1 – 90.

Вариант 13. Построения рядов проведите на основе данных с порядковыми номерами 1 – 99.

Вариант 14. Построения рядов проведите на основе данных с порядковыми номерами 1 – 113.

Вариант 15. Построения рядов проведите на основе данных с порядковыми номерами 1 – 118.

Вариант 16. Построения рядов проведите на основе данных с порядковыми номерами 1 – 126.

Вариант 17. Построения рядов проведите на основе данных с порядковыми номерами 1 – 129.

Вариант 18. Построения рядов проведите на основе данных с порядковыми номерами 35 – 99.

Вариант 19. Построения рядов проведите на основе данных с порядковыми номерами 35 – 113.

Вариант 20. Построения рядов проведите на основе данных с порядковыми номерами 35 – 118.

Вариант 41. Построения рядов проведите на основе данных с порядковыми номерами 1 – 181.

Вариант 42. Построения рядов проведите на основе данных с порядковыми номерами 1 – 190.

Вариант 43. Построения рядов проведите на основе данных с порядковыми номерами 1 –199.

Вариант 44. Построения рядов проведите на основе данных с порядковыми номерами 1 – 163.

Вариант 45. Построения рядов проведите на основе данных с порядковыми номерами 1 – 148.

Вариант 46. Построения рядов проведите на основе данных с порядковыми номерами 1 – 176.

Вариант 47. Построения рядов проведите на основе данных с порядковыми номерами 1 – 189.

Вариант 48. Построения рядов проведите на основе данных с порядковыми номерами 135 – 299.

Вариант 49. Построения рядов проведите на основе данных с порядковыми номерами 137 – 213.

Вариант 50. Построения рядов проведите на основе данных с порядковыми номерами 137 – 218.

Вариант 71. Построения рядов проведите на основе данных с порядковыми номерами 35 – 124.

Вариант 72. Построения рядов проведите на основе данных с порядковыми номерами 35 – 126.

Вариант 73. Построения рядов проведите на основе данных с порядковыми номерами 35 – 129.

Вариант 74. Построения рядов проведите на основе данных с порядковыми номерами 35 – 131.

Вариант 75. Построения рядов проведите на основе данных с порядковыми номерами 91 – 155.

Вариант 76. Построения рядов проведите на основе данных с порядковыми номерами 91 – 157.

Вариант 77. Построения рядов проведите на основе данных с порядковыми номерами 154 – 176.

Вариант 78. Построения рядов проведите на основе данных с порядковыми номерами 53 – 157.

Вариант 79. Построения рядов проведите на основе данных с порядковыми номерами 53 – 164.

Вариант 80. Построения рядов проведите на основе данных с порядковыми номерами 53 – 167.

Для вариантов 21 – 30, 51 – 60, 81 – 90 используйте переменную x_6 для построения интервального ряда и x_3 для построения дискретного ряда.

Предварительно отсортируйте данные по переменной x_3 по возрастанию признака. Затем проводите построение рядов на основе данных с порядковыми номерами, соответствующими вашему варианту.

Вариант 21. Построения рядов проведите на основе данных с порядковыми номерами 1 – 114.

Вариант 22. Построения рядов проведите на основе данных с порядковыми номерами 1 – 130.

Вариант 23. Построения рядов проведите на основе данных с порядковыми номерами 1 – 132.

Вариант 24. Построения рядов проведите на основе данных с порядковыми номерами 1 – 138.

Вариант 25. Построения рядов проведите на основе данных с порядковыми номерами 24 – 138.

Вариант 26. Построения рядов проведите на основе данных с порядковыми номерами 24 – 132.

Вариант 27. Построения рядов проведите на основе данных с порядковыми номерами 24 – 130.

Вариант 28. Построения рядов проведите на основе данных с порядковыми номерами 24 – 114.

Вариант 29. Построения рядов проведите на основе данных с порядковыми номерами 57 – 138.

Вариант 30. Построения рядов проведите на основе данных с порядковыми номерами 57 – 132.

Вариант 51. Построения рядов проведите на основе данных с порядковыми номерами 57 – 155.

Вариант 52. Построения рядов проведите на основе данных с порядковыми номерами 68 – 155.

Вариант 53. Построения рядов проведите на основе данных с порядковыми номерами 83 – 155.

Вариант 54. Построения рядов проведите на основе данных с порядковыми номерами 83 – 156.

Вариант 55. Построения рядов проведите на основе данных с порядковыми номерами 68 – 160.

Вариант 56. Построения рядов проведите на основе данных с порядковыми номерами 68 – 156.

Вариант 57. Построения рядов проведите на основе данных с порядковыми номерами 57 – 160.

Вариант 58. Построения рядов проведите на основе данных с порядковыми номерами 57 – 156.

Вариант 59. Построения рядов проведите на основе данных с порядковыми номерами 24 – 155.

Вариант 60. Построения рядов проведите на основе данных с порядковыми номерами 24 – 156.

Вариант 81. Построения рядов проведите на основе данных с порядковыми номерами 68 – 138.

Вариант 82. Построения рядов проведите на основе данных с порядковыми номерами 68 – 145.

Вариант 83. Построения рядов проведите на основе данных с порядковыми номерами 57 – 145.

Вариант 84. Построения рядов проведите на основе данных с порядковыми номерами 50 – 138.

Вариант 85. Построения рядов проведите на основе данных с порядковыми номерами 50 – 155.

Вариант 86. Построения рядов проведите на основе данных с порядковыми номерами 29 – 114.

Вариант 87. Построения рядов проведите на основе данных с порядковыми номерами 29 – 132.

Вариант 88. Построения рядов проведите на основе данных с порядковыми номерами 29 – 138.

Вариант 89. Построения рядов проведите на основе данных с порядковыми номерами 49 – 155.

Вариант 90. Построения рядов проведите на основе данных с порядковыми номерами 49 – 156.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

Елисеева, И И Общая теория статистики : Учебник / И.И. Елисеева, М.М. Юзбашев; под ред. И.И. Елисеевой. – М. : Финансы и статистика, 2006. – 656 с.

Общая теория статистики, общие сведения (рабочая версия документа готовится совместно с Затопляевой С, Иванковой Е И, Кеткиной О С, Коноваловым О Н, документ доступен в распечатанном формате, а также по запросу может быть отправлен по электронной почте), 2023. – 34 с.

Харченко, Л П Статистика : учебное пособие / Л.П. Харченко, В Г Долженкова, В Г Ионин и др. ; под ред. В Г Ионина. – М. : ИНФРА-М, 2002. – 384 с.

Электронный текстовый ресурс

Кеткина Ольга Сергеевна

Общая теория статистики прикладной характер

Методические материалы для студентов, обучающихся по программе бакалавриата по направлению подготовки 38.03.05 «Бизнес-информатика»

Подготовка к публикации *О Кеткина*

Компьютерный набор *О Кеткина*

Рекомендовано Методическим советом УрФУ

Разрешено к публикации 21.12.2023

Электронный формат – pdf

Объем 1,5 уч.-изд. л.

620002, Екатеринбург, ул. Гоголя, 25

Информационный портал УрФУ

www.study.urfu.ru