УНИВЕРСИТЕТ ИТМО

Т.М. Сизова

СТАТИСТИКА для бакалавров

Часть II

Санкт-Петербург
2016

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

УНИВЕРСИТЕТ ИТМО

Т.М. СИЗОВА

СТАТИСТИКА ДЛЯ БАКАЛАВРОВ

Учебное пособие: часть II

Санкт-Петербург
2016

ББК 65.052.627

СИЗОВА Т.М.
Статистика для бакалавров: Учебное пособие. ЧастьII – СПб: Университет ИТМО, 2016– 70с.

Учебное пособие предназначено для использования в учебном процессе студентами Факультета технологического менеджмента и инноваций Университета ИТМО, обучающихся по направлениям 38.03.01 «Экономика» и 38.03.02 «Менеджмент» и изучающих дисциплину «Статистика».

Одобрено на заседании Ученого Совета Факультета технологического менеджмента и инноваций 18.10.2016, протокол №2.

УНИВЕРСИТЕТ ИТМО

Университет ИТМО – ведущий вуз России в области информационных и фотонных технологий, один из немногих российских вузов, получивших в 2009 году статус национального исследовательского университета. С 2013 года Университет ИТМО – участник программы повышения конкурентоспособности российских университетов среди ведущих мировых научно-образовательных центров, известной как проект «5 в 100». Цель Университета ИТМО – становление исследовательского университета мирового уровня, предпринимательского по типу, ориентированного на интернационализацию всех направлений деятельности.

Содержание

Введение

Во второй части учебного пособия «Основы статистического анализа» рассмотрены основные методы статистического анализа массовых явлений и процессов. В главе, посвященной анализу частотных распределений, излагается методика эмпирического и теоретического исследования рядов распределения; в главе, посвященной динамике социально-экономических процессов рассмотрены вопросы оценки интенсивности развития, моделирования рядов динамики, выявления и описания основных его компонент (тенденции, сезонности) и вопросы статистического прогнозирования. Кроме этого, рассмотрены методы индексного анализа и основы теории выборочного исследования, как эффективного метода получения и анализа статистической информации.

Раздел II. Основы статистического анализа

1. Анализ частотных распределений

1.1. Ряды распределения

Результаты статистических сводок и группировок могут быть представлены в виде *рядов распределения.*

Ряд распределения, представляет собой систематизированную последовательность статистических единиц, сгруппированных по конкретному признаку. Он характеризует состав изучаемого явления, позволяет судить об однородности совокупности, закономерности распределения статистических единиц. Обычно ряд распределения представляет собой результат структурной группировки.

Ряд распределения считается построенным, если известно, каким образом меняются в совокупности значения признака, как часто встречаются отдельные значения признака.

Для различных статистических признаков строятся ряды распределения разного типа:

- **атрибутивные –** строятся по описательным признакам в порядке возрастания или убывания значений признака; примером атрибутивных рядов могут служить распределения населения по национальности, по профессиям, по полу; распределение предприятий по формам собственности;
- **вариационные** - строятся по количественным признакам, например, распределение рабочих по уровню квалификации, по заработной плате, распределение студентов по успеваемости.

Вариационные ряды делятся на дискретные и интервальные.

В дискретных рядах признак принимает только целые значения, например, размер семьи, тарифный разряд.

Интервальные ряды основаны на непрерывных признаках, принимающих любые, в том числе и дробные значения. В зависимости от того, какая структурная группировка лежит в основе интервального ряда, различают *равно интервальные и неравно интервальные* ряды.

В *равно интервальных рядах* ширина интервала является величиной постоянной, в *неравно интервальных* – она различна для разных групп.

Основными элементами рядов распределения являются:

- **значения признака (варианты):**

- x_i - дискретное в дискретных рядах;

- $x_i^н$ - $x_i^в$ - интервал для интервальных рядов;

- **частота** n_i - ***число единиц совокупности, обладающих данным значением признака.*** Частота показывает, сколько раз данное значение признака встречается в совокупности; сумма всех частот всегда равна объему статистической совокупности, т. е. $\sum_1^m n_i = N$.

Исследование рядов распределения осуществляется в два этапа:

- *эмпирическое исследование,* целью которого является получение обобщающих характеристик изучаемой совокупности;
- *теоретическое исследование* с целью выявления закономерности данного распределения и его теоретического описания.

Эмпирическое исследование начинается с определения частотных характеристик ряда распределения.

1.2. Частотные характеристики рядов распределения и их графическое представление

Исходной частотной характеристикой любого ряда распределения является **частота** n_i. На ее основе можно рассчитать следующие характеристики:

- **Частость** – *удельный вес (доля) единиц совокупности, имеющих определенное значение признака,* т. е. это частота, выраженная в виде относительной величины (доли единицы или процента):

$$q_i = \frac{n_i}{N}, \quad i = \overline{1,m}, \quad \sum_{i=1}^{m} q_i = 1.$$

Эта характеристика имеет важное значение при исследовании рядов распределения, так как позволяет связать показатели рядов распределения с соответствующими показателями и аппаратом теории вероятностей. В теории вероятностей q_i есть вероятность того, что данное значение признака встретится в совокупности. Частость используется для сопоставления рядов распределения, содержащих разное число статистических единиц.

- **Накопленная частота** – число единиц совокупности, у которых значение признака не превышает данного x^*, т. е. это частота нарастающим итогом:

$$N_{x^*} = \sum_{i=1}^{m^*} n_i, \quad N_{x_m} = N.$$

x^* - данное значение признака в i-ой группе, для которой рассчитывается накопленная частота.

По накопленным частотам можно построить *кумулятивный ряд распределения* – ряд значений числа единиц совокупности с меньшими и равными верхней границе соответствующего интервала значениями признака.

- **Накопленная частость** – удельный вес (доля) единиц, у которых значение признака не превосходит данное x^*, т. е. это частость нарастающим итогом:

$$Q_{x^*} = \sum_{i=1}^{m^*} q_i, \quad Q_{x_m} = 1;$$

- **Плотность распределения** – универсальная частотная характеристика, позволяющая перейти от эмпирического к теоретическому распределению. Для рядов с неравными интервалами только эта характеристика дает правильное представление о характере распределения. Плотность распределения рассчитывается в 2-х вариантах:

- как *абсолютная плотность распределения* φ_i, показывающая число единиц совокупности, приходящихся на единицу ширины интервала значения признака $\varphi_i = \frac{n_i}{a_i}$; - как *относительная плотность распределения* φ_i', показывающая удельный вес единиц совокупности, приходящихся на единицу ширины интервала $\varphi_i' = \frac{q_i}{a_i}$.

Плотность распределения обеспечивает сопоставимость различных рядов распределения.

Разные ряды распределения характеризуются разным набором частотных характеристик: минимальным – атрибутивные ряды (частота n_i, и частость q_i), для дискретных используются четыре характеристики (частота n_i, частость q_i, накопленная частота N_i, накопленная частость Q_i), для интервальных – все пять (частота n_i, частость q_i, накопленная частота N_i, накопленная частость Q_i, абсолютная - φ_i и относительная - φ_i' плотности распределения).

Для наглядности ряды распределения часто представляют графически . Для изображения рядов применяются линейные графики и плоскостные диаграммы, построенные в прямоугольной системе координат.

Для графического представления атрибутивных рядов распределения используются различные диаграммы: столбиковые, линейные, круговые, фигурные, секторные и т. д.

Для дискретных вариационных рядов основным графиком является полигон распределения.

Полигоном распределения называется ломаная линия, соединяющая точки с координатами $\{x_i ; n_i\}$ или $\{x_i ; q_i\}$,где x_i - дискретное значение признака, n_i - частота, q_i - частость.

График строится в принятом масштабе. Вид полигона распределения приведен на рис.1.1.

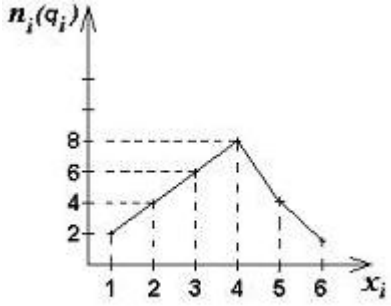

Рис.1.1. Полигон распределения

Для изображения интервальных вариационных рядов применяют **гистограммы,** *представляющие собой* ступенчатые фигуры, состоящие из прямоугольников, основания которых равны ширине интервала a_i, а высота - частоте n_i (частости q_i) равноинтервального ряда или плотности распределения неравноинтервального φ_i, φ_i'. Построение гистограммы аналогично построению столбиковой диаграммы. Примерный вид гистограммы приведен на рис. 1.2.

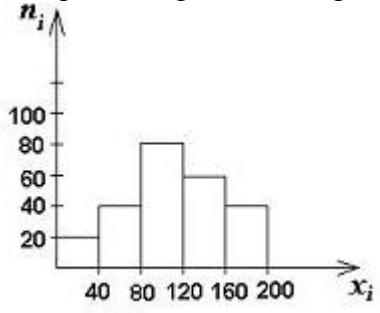

Рис.1.2. Гистограмма распределения

Для графического представления вариационных рядов может использоваться также **кумулята (график накопленных частот)** – ломаная линия, составленная по накопленным частотам (частостям). Накопленные частоты наносятся в виде ординат; соединяя вершины отдельных ординат отрезками прямой, получаем ломаную линию, имеющую неубывающий вид. Координатами точек на графике для дискретного ряда являются $\{x_i ; N_i\}$; для интервального ряда - $\{x_i^e ; N_i\}$. Начальная точка графика имеет координаты $\{x_1^н ; 0\}$, самая высокая точка - $\{x_m^e ; N\}$. Общий вид кумуляты приведен на

рис.1.3. Использование кумуляты особенно удобно при проведении сравнений вариационных рядов.

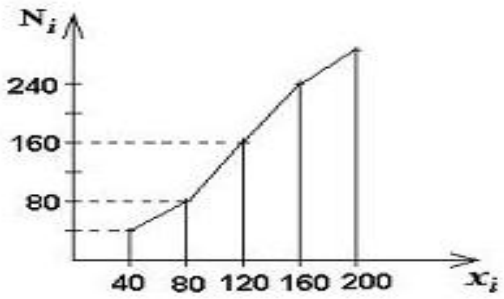

Рис. 1.3. График кумулятивного распределения частот

При построении графиков рядов распределения *большое значение имеет соотношение масштабов по оси абсцисс и оси ординат*. В этом случае необходимо руководствоваться «правилом золотого сечения», в соответствии с которым *высота графика должна быть примерно в два раза меньше его основания*.

1.3. Эмпирическое исследование рядов распределения

При проведении эмпирического исследования ряда распределения рассчитываются и анализируются следующие группы показателей:

- *показатели положения центра распределения;*
- *показатели степени его однородности;*
- *показатели формы распределения.*

1.3.1. *К показателям положения центра распределения* относятся ***степенная средняя в виде средней арифметической и структурные средние – мода и медиана.***

Расчет этих показателей рассмотрен в предыдущей главе. Кроме того

моду можно определить графически по полигону распределения в дискретных рядах, по гистограмме распределения – в интервальных, а медиану - по кумуляте.

Для нахождения моды в интервальном ряду правую вершину модального прямоугольника нужно соединить с правым верхним углом предыдущего прямоугольника, а левую вершину – с левым верхним углом последующего прямоугольника. Абсцисса точки пересечения этих прямых и будет модой распределения.

Для определение медианы высоту наибольшей ординаты кумуляты, соответствующей общей численности совокупности, делят пополам. Через полученную точку проводят прямую, параллельную оси абсцисс, до пересечения ее с кумулятой. Абсцисса точки пересечения является медианой.

Таким образом, для характеристики положения центра ряда распределения можно использовать 3 показателя: **среднее значение признака**, **мода, медиана**. При выборе вида и формы конкретного показателя центра распределения необходимо исходить из следующих рекомендаций:

- для устойчивых социально-экономических процессов в качестве показателя центра используют среднюю арифметическую. Такие процессы характеризуются симметричными распределениями, в которых $\bar{x} = Me = Mo$;
- для неустойчивых процессов положение центра распределения характеризуется с помощью Mo или Me. Для асимметричных процессов предпочтительной характеристикой центра распределения является медиана, поскольку занимает положение между средней арифметической и модой и не чувствительна к крайним значениям признака в совокупности.

1.3.2. Для оценки **однородности ряда распределения** используются рассмотренные ранее абсолютные и относительные показатели. К ним относятся: размах вариации R, дисперсия σ^2, среднеквадратичное отклонение σ, коэффициент вариации $V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100\%$.

1.3.3. Выяснение общего характера распределения предполагает не только оценку степени его однородности, но и *исследование формы распределения, т.е. оценку симметричности и эксцесса*.

Из математической статистики известно, что при увеличении объема статистической совокупности ($N \to \infty$) и одновременном уменьшении интервала группировки ($x_i \to 0$), полигон либо гистограмма распределения все более и более приближаются к некоторой плавной кривой, являющейся для указанных графиков пределом. Эта кривая называется **эмпирической кривой распределения** и представляет собой **графическое изображение в виде непрерывной линии изменения частот, функционально связанного с изменением вариант.**

В статистике различают следующие **виды кривых распределения**:

- *одновершинные кривые;*
- *многовершинные кривые.*

Однородные совокупности описываются одновершинными кривыми. Многовершинность распределения свидетельствует о неоднородности изучаемой совокупности или о некачественном выполнении группировки.

Одновершинные кривые распределения делятся на симметричные, умеренно асимметричные и крайне асимметричные.

Распределение называется симметричным, если частоты любых 2-х вариантов, равноотстоящих в обе стороны от центра распределения, равны между собой. В таких распределениях $\bar{x} = Mo = Me$.

На практике чаще всего приходится работать с *асимметричными распределениями*, в которых частоты при удалении от центра распределения убывают неодинаково. Для характеристики асимметрии (несимметричности распределения) используют специальные показатели - *коэффициенты асимметрии*.

Наиболее часто используются следующие из них:

Коэффициент асимметрии Пирсона

$$As = \frac{\bar{x} - Mo}{\sigma}.$$

В симметричных распределениях As=0. *При As<0* наблюдается *отрицательная (левосторонняя) асимметрия (рис. 6.4.)*, для которой характерно следующее соотношение между показателями центра распределения: $Mo > Me > \bar{x}$.

Чем ближе по модулю *As* к 1, тем асимметрия существеннее:

- *если |As|<0,25, то асимметрия считается незначительной;*
- *если 0.5 < |As |<0.25 то асимметрия считается умеренной;*
- *если |As|>0,5 – асимметрия значительна.*

Коэффициент асимметрии Пирсона характеризует асимметрию только в центральной части распределения, поэтому более распространенным и более точным является **коэффициент асимметрии**, ***рассчитанный на основе центрального момента 3-его порядка:***

$$As = \frac{\mu_3}{\sigma^3},$$

где μ_3 - центральный момент третьего порядка;

σ^3 - среднее квадратическое отклонение в третьей степени.

Центральным моментом в статистике называется среднее отклонение индивидуальных значений признака от его среднеарифметической величины.

Центральный момент k-ого порядка рассчитывается как:

$$\mu_k = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^k}{n}$$ - для несгруппированных данных;

$$\mu_k = \frac{\sum_{i=1}^{m}(x_i - \bar{x})^k}{i=1\sum^{m} n_i}$$ - для сгруппированных данных.

Соответственно формулы для определения центрального момента третьего порядка имеют следующий вид:

$$\mu_3 = \frac{\sum(x_i - \bar{x})^3}{n}$$ - для несгруппированных данных;

$$\mu_3 = \frac{\sum(x_i - \bar{x})^3 \cdot n_i}{\sum n_i}$$ - для сгруппированных данных.

Для оценки существенности рассчитанного вторым способом коэффициента асимметрии определяется его средняя квадратическая ошибка:

$$\sigma_{As} = \sqrt{\frac{6 \cdot (N-1)}{(N+1) \cdot (N+3)}}.$$

Если $\frac{|AS|}{\sigma_{AS}} > 3$, асимметрия является существенной.

Для одновершинных распределений рассчитывается еще один показатель оценки его формы – ***эксцесс.*** Эксцесс *является показателем островершинности распределения*. Он рассчитывается для симметричных распределений на основе центрального момента 4-ого порядка μ_4:

$$Ex = \frac{\mu_4}{\sigma^4} - 3,$$

где μ_4 - центральный момент 4-го порядка.

$$\mu_4 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^4}{N}$$ - для несгруппированных данных;

$$\mu_4 = \frac{\sum_{i=1}^{m}(x_i - \bar{x})^4 \cdot n_i}{\sum_{i=1}^{m} n_i}$$ - для сгруппированных данных.

При симметричных распределениях Ex=0 .Если Ex>0, то распределение относится к ***островершинным****, если Ex<0 –* ***к плосковершинным.***

1.4. Теоретические распределения в анализе вариационных рядов

Эмпирические кривые распределения, построенные на основе, как правило, небольшого числа наблюдений очень трудно описать аналитически, поэтому для выявления статистических закономерностей, сравнения и обобщения различных совокупностей аналогичных данных используются теоретические распределения.

Теоретические распределения – это хорошо изученные в теории *распределения, представляющие собой зависимости между плотностями распределения и значениями признака, отражающие закономерности распределения.* Они описываются

статистическими функциями, параметры которых вычисляются по статистическим характеристикам изучаемой совокупности.

Исследование формы распределения предполагает замену эмпирического распределения известным теоретическим, близким ему по форме. При замене необходимо соблюдать условие: *различия между эмпирическим и теоретическим распределениями должны быть минимальными*. Это означает, что ***сумма частот эмпирического распределения должна соответствовать сумме частот теоретического распределения, т.е.*** $\sum_{i=1}^{m} n_i \approx \sum_{i=1}^{m} n_{i_T}$, где n_{i_T} - частота теоретического распределения.

Теоретическое распределение в этом случае является некоторой идеализированной моделью эмпирического распределения, и анализ вариационного ряда сводится к сопоставлению эмпирического и теоретического распределений и определению различий между ними.

В статистической практике наиболее широко используют **нормальное распределение (распределение Гаусса).** Оно *применяется для описания распределения признаков, на которые действуют множество независимых факторов, среди которых нет доминирующих.*

Функция нормального распределения имеет вид:

$$\varphi'(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\bar{x})^2}{2\sigma^2}},$$

где $\varphi'(x)$ - относительная плотность распределения (ордината кривой нормального распределения);

π =3,14, e = 2,72 - математические константы;

$\bar{x}$ - среднее значение признака в распределении;

σ- среднее квадратическое отклонение.

Для конкретного распределения среднее значение признака $\bar{x}$ и среднее квадратическое отклонение σ являются постоянными величинами. Графически нормальное распределение может быть представлено в виде симметричной колоколообразной кривой. К **основным свойствам** кривой нормального распределения относятся:

- *кривая распределения является одновершинной*; координаты вершины - $\{\bar{x}; \frac{1}{\sigma \cdot \sqrt{2\pi}}\}$;
- *кривая распределения симметрична относительно оси, проходящей через центр распределения* $\bar{x} = Mo = Me$;
- *кривая имеет три точки перегиба*: в вершине, на левой ветви $\{\bar{x} - \sigma; \frac{1}{\sigma \cdot \sqrt{2\pi \cdot e}}\}$, и на правой - $\{\bar{x} + \sigma; \frac{1}{\sigma \cdot \sqrt{2\pi \cdot e}}\}$;
- *кривая имеет две ветви, асимптотически приближающиеся к оси абсцисс*, продолжаясь до бесконечности;
- *если меняется значение* $\bar{x}$*, кривая перемещается вдоль оси ординат*, при этом форма кривой не меняется;
- *если меняется значение* σ*, меняется форма распределения* при неизменном положении центра распределения: при уменьшении σ - уменьшается вариация, кривая становится более пологой, увеличивается эксцесс; при увеличении σ - увеличивается вариация, эксцесс уменьшается;
- *площадь, ограниченная кривой сверху и осью абсцисс снизу, характеризует вероятность появления определенных значений признака*: если всю её принять за 100%, то

в пределах $\bar{x} \pm \sigma$ *находится 68,3% всех значений признака, в пределах* $\bar{x} \pm 2\sigma$ *- 95,44% значений, в пределах* $\bar{x} \pm 3\sigma$ *- 99,73% значений признака.* Этот вывод называется правилом **“трех сигм”**, в соответствии, с которым можно считать, что все возможные значения нормально распределенного признака укладываются в интервал $\bar{x} \pm 3\sigma$.

Пользоваться функцией нормального распределения в её первоначальном виде сложно, так как для каждой пары значений $\bar{x}$ и σ необходимо создавать свои таблицы значений. Поэтому функцию стандартизируют и затем используют для обработки рядов распределения, для чего вводится *понятие стандартного отклонения* t_i:

$$t_i = \frac{x_i - \bar{x}}{\sigma}.$$

тогда:

$$\varphi'(x) = \frac{1}{\sigma} \cdot \left(\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} \right).$$

Выражение $\varphi'(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}$**,** состоящее из констант и не содержащее параметров, **называется стандартизованной функцией нормального распределения.** Для неё разработаны специальные таблицы, позволяющие находить конкретные значения $\varphi'(t)$ при различных значениях аргумента.

Исходная функция нормального распределения связана со стандартизированной соотношением:

$$\varphi'(x) = \frac{1}{\sigma} \cdot \varphi'(t).$$

Стандартизованная функция является четной, т.е. $\varphi'(-t) = \varphi'(t)$.

Для того чтобы оценить близость указанного ряда распределения к нормальному, необходимо рассчитать частоты теоретического ряда распределения n_{i_T} .Для их расчета определяются стандартные отклонения $t = \frac{x - \bar{x}}{\sigma}$, затем по таблицам значений функции Лапласа находятся значения $\varphi'(t)$.

Для получения частот теоретического распределения n_{i_T} необходимо воспользоваться зависимостью относительной плотности распределения $\varphi'(x)$ с частотой n_i, и ее связью со стандартизованной функцией нормального распределения $\varphi'(t)$. Эти зависимости имеют вид:

$$\varphi'(x) = \frac{q_{i_T}}{a_i}, \quad q_{i_T} = \frac{n_{i_T}}{N}, \textit{следовательно}, \quad \varphi'(x) = \frac{n_{i_T}}{N \cdot a_i}.$$

С другой стороны, $\varphi'(x) = \frac{1}{\sigma} \cdot \varphi'(t)$, таким образом, имеет место равенство:

$\frac{n_{i_T}}{N \cdot a_i} = \frac{1}{\sigma} \cdot \varphi'(t)$, отсюда $n_{i_T} = \frac{a_i \cdot N}{\sigma} \cdot \varphi'(t)$;

где a_i - ширина интервала,

N – объем статистической совокупности,

σ - среднее квадратическое отклонение,

$\varphi'(t)$ - стандартизованная функция нормального распределения.

Полученные значения n_{i_T} округляются до целых значений в соответствии со смыслом характеристики частоты.

Для определения близости эмпирического и теоретического распределений, можно построить эмпирическую и теоретическую кривые распределения. Их сопоставление позволяет оценить степень расхождения между ними.

Визуальное сопоставление эмпирической и теоретической кривых распределения позволяет получить субъективную оценку их близости. Сравнивая графики, можно утверждать, что наблюдается довольно большая близость фактических и теоретических частот распределения. Следовательно, можно сделать вывод о том, что исследуемый ряд подчиняется закону нормального распределения. Для получения объективной оценки расхождения между эмпирической и теоретической кривыми распределения используются специальные статистические показатели – критерии согласия.

1.5. Оценка близости эмпирического и теоретического распределений

Эмпирическое распределение отличается от теоретического тем, что на значения признака в нем влияют случайные факторы. С увеличением объема статистической совокупности влияние случайных факторов ослабевает, и эмпирическое распределение все менее отличается от теоретического.

Для оценки близости распределений применяют особые показатели – **критерии согласия,** основанные на использовании различных мер расстояний между эмпирическим и теоретическим распределением.

Наиболее часто на практике используются следующие критерия согласия:

- **«хи-квадрат»- критерий (критерий Пирсона);**
- **«лямбда»- китерий» (критерий Колмогорова).**

1.5.1. «Хи-квадрат» - критерий является случайной величиной, имеющей распределение, близкое к распределению «хи-квадрат». Его величина определяется по формуле:

$$\chi_p^2 = \sum_{i=1}^{m} \frac{(n_i - n_{i_T})^2}{n_{i_T}}.$$

Чем меньше эмпирические и теоретические частоты в отдельных группах отличаются друг от друга, тем меньше эмпирическое распределение отличается от теоретического, то есть тем в большей степени эмпирическое и теоретическое распределения согласуются между собой.

Для оценки существенности расчетной величины «хи-квадрат» - критерия производится ее сравнение с табличным (критическим) значением χ_k^2, определяемым по статистическим таблицам значений χ^2- критерия. χ_k^2 определяют в зависимости от уровня значимости α и параметра k, определяющего число степеней свободы распределения. k=m- m_1-1, где α - вероятность ошибки, m_1 - число оцененных параметров теоретического распределения по наблюдаемым значениям признака.

Уровень значимости выбирается таким образом, что $P(\chi_p^2 > \chi_к^2) = \alpha$. Обычно α принимается равным 0,05 или 0,01, что соответствует вероятности 95% или 99%.

Если $\chi_p^2 \le \chi_k^2$, то считают, что распределения близки друг другу, различия между ними несущественны.

χ^2- критерий может применяться для оценки близости эмпирического распределения к теоретическому при соблюдении следующих условий:

- статистическая совокупность состоит из 50-ти и более единиц;
- теоретические частоты $n_{i_T} \ge 5$,- если это условие не соблюдается, то следует объединить интервалы.

Рассчитаем в таблице 6.7. значения отклонений ($n_i - n_{i_т}$) и фактическое значение χ^2-критерия. По расчету $\chi_p^2 = 1{,}66$. Это значение сравнивается с табличным, определенном при числе степеней свободы k=4 и уровне значимости $\alpha = 0{,}05$. Оно равно $\chi_к^2$=9,49.

Таким образом $\chi_p^2 < \chi_к^2$; эмпирическое и теоретическое распределения признаются близкими друг другу с вероятностью 95%, расхождения между ними - несущественными, вызываемыми случайной вариацией признака в совокупности.

- *На основе «хи-квадрат» - критерия может быть рассчитан ещё один критерий согласия* – **критерий Романовского:**

$$C = \frac{\chi_p^2 - (m-3)}{\sqrt{2 \cdot (m-3)}}.$$

Эмпирическое и теоретическое распределения признаются близкими друг другу, если C<3.

1.5.2. Критерий согласия Колмогорова («лямбда-критерий») основан на другой мере близости распределений. Для оценки близости эмпирического распределения к нормальному используется максимальная разница между накопленными эмпирическими и накопленными теоретическими частотами. Расчетное значение **«лямбда»- критерия»** определяется по формуле:

$$\lambda_p = \frac{Д}{\sqrt{\sum_{i=1}^{m} n_i}} = Д : \sqrt{N},$$

,

где $Д = \max_{i=1,m} \{N_i - N_{i_Т}\}$

N_i - накопленная эмпирическая частота

$N_{i_Т}$ - накопленная теоретическая частота.

По рассчитанному значению λ_p по специальной таблице вероятностей «лямбда»-критерия» определяется вероятность того, что рассматриваемое эмпирическое распределение подчиняется закону нормального распределения.

Для рассматриваемого примера Д=2 - в соответствии с расчетом, приведенным в таблице 6.7.

Тогда $\lambda_p = \frac{2}{\sqrt{50}} = \frac{2}{7{,}07} = 0{,}283$.

По таблице вероятностей $P(\lambda)$ определяем, что такому значению λ соответствует вероятность P(λ), близкая к 1.

Полученное значение вероятности свидетельствует о том, что *расхождение между эмпирическим и теоретическим распределениями несущественны, вызваны случайной вариацией признака в статистической совокупности*. В основе эмпирического распределения рабочих по стажу лежит закон нормального распределения.

Контрольные вопросы и задания

1. Для чего нужны ряды распределения?
2. Из каких элементов состоят ряды распределения?
3. Дайте классификацию рядов распределения.
4. Перечислите частотные характеристики рядов распределения.
5. Как графически могут быть представлены интервальные ряды распределения?
6. Какие показатели определяют положение центра распределения?
7. Как выбирается показатель центра распределения в неоднородных вариационных рядах?
8. Как оценить однородность ряда распределения?

9. Перечислите основные показатели, характеризующие форму распределения.

10.Каким образом можно оценить симметричность распределения?

11.Как интерпретируются показатели эксцесса?

12. С какой целью при анализе рядов распределения используются теоретические кривые распределения?

13.Каким образом оценивается близость построенного эмпирического ряда распределения к нормальному распределению?

2. Статистические связи и их исследование

2.1. Понятие статистической и корреляционной связи, методы их исследования

Одной из важнейших задач статистики является изучение объективно существующих связей между явлениями. **Для описания причинно-следственной связи между явлениями и процессами используется деление статистических признаков,** отражающих отдельные стороны взаимосвязанных явлений, **на факторные и результативные.** *Факторными считаются признаки, обуславливающие вариацию других, связанных с ними признаков*, являющихся причинами и условиями таких изменений. *Результативными являются признаки, варьирующие под воздействием факторных.*

Формы проявления существующих взаимосвязей могут быть разными . В качестве самых общих их видов рассматривают *функциональную и статистическую связи.*

Функциональной *называют связь, при которой определённому значению факторного признака соответствует одно и только одно значение результативного.* Такая связь возможна при условии, что на поведение одного признака (результативного) влияет только факторный признак.

Функциональные связи являются абстракциями, в реальной жизни они встречаются редко, но находят широкое применение в точных науках и в первую очередь, в математике. *Функциональная связь проявляется во всех случаях наблюдения и для каждой конкретной единицы изучаемой совокупности.*

В массовых явлениях проявляются **статистические связи*, при которых строго определённому значению факторного признака ставится в соответствие множество значений результативного.*** Такие связи имеют место, если на результативный признак действуют несколько факторных, а для описания связи используется один или несколько определяющих (учтённых) факторов.

Строгое различие между функциональной и статистической связью можно получить при их математической формулировке.

Функциональную связь можно представить уравнением: $y_i = f(x_i)$**,**

где y_i результативный признак (i=1,…n);

$f(x_i)$ - функция связи результативного и факторного признаков;

x_i - факторный признак.

***Статистическая связь может быть представлена уравнением следующего вида*:**

$\tilde{y}_i = f(x_i) + \varepsilon_i$**,**

где $\tilde{y}_2$ - расчётное (среднее) значение результативного признака;

$f(x_i)$ - часть значения результативного признака, сформировавшаяся под воздействием учтённых факторов;

ε_i - часть значения результативного признака, возникающая вследствие действия неконтролируемых факторов или ошибок измерения.

Примером статистической связи может служить зависимость себестоимости единицы продукции от уровня производительности труда: чем выше производительность труда, тем ниже себестоимость. Но на себестоимость единицы продукции помимо

производительности труда влияют и другие факторы: стоимость сырья, материалов, топлива, общепроизводственные и общехозяйственные расходы и т.д. Поэтому нельзя утверждать, что повышение производительности труда на 5% приведет к аналогичному снижению себестоимости. Может наблюдаться и обратная картина, если на себестоимость будут влиять в бóльшей степени другие факторы, - например, резко возрастут цены на сырье и материалы.

Любую статистическую связь можно представить в виде набора локальных распределений результативного признака при фиксированных значениях факторного:

$x_1 : y_{1,1,} y_{1,2} \ldots y_{1,j} \ldots y_{1,m}$

$x_2 : y_{2,1,} y_{2,2} \ldots y_{2,j} \ldots y_{2,m}$

……

$x_n : y_{n,1,} y_{n,2} \ldots y_{n,j} \ldots y_{n,m}$,

где $i = \overline{1,n}$, $j = \overline{1,m}$.

Каждое локальное распределение результативного признака можно описать на эмпирическом уровне, рассчитав такие его характеристики как *локальная средняя результативного признака* $\tilde{y}_i$, характеризующая положение центра распределения, и *среднеквадратическое отклонение результативного признака* σ_i , характеризующее форму локального распределения.

Если при изменении значений факторного признака x_i *будут смещаться центры локальных распределений (меняться значение локальных средних* $\tilde{y}_i$*), но не будет меняться форма локальных распределений (значения средних квадратических отклонений* σ_i *), то можно говорить о наличии между признаками* ***корреляционной связи.***

Корреляционная связь является обобщающим случаем статистической связи. ***При корреляционной связи с изменением значения факторного признака*** x_i ***закономерно изменяется среднее значение результативного признака*** $\tilde{y}_i$**,** в то время как в каждом отдельном случае факторный признак может принимать множество различных значений.

Корреляционная связь может быть представлена уравнением: $\tilde{y}_i$**=F(** x_i**),**

где F(x_i) – функция связи среднего значения результативного признака с факторным. Она *проявляется только на уровне всей статистической совокупности*, а не в каждом отдельном случае, *так* как *только при достаточно большом числе случаев каждому случайному значению факторного признака будет соответствовать распределение средних значений случайного признака* ***y.***

По направлению корреляционные связи делятся на прямые и обратные. *При прямой связи* результативный признак растёт с увеличением факторного, *при обратной* - рост факторного признака приводит к снижению значений результативного признака. Например, чем больше стаж работы, тем выше производительность труда – прямая связь, а чем выше производительность труда, тем ниже себестоимость единицы продукции – обратная связь.

По форме (аналитическому выражению) **связи делятся на линейные** (прямолинейные) **и нелинейные** (криволинейные) связи. *Линейные связи выражаются уравнением прямой, а нелинейные – уравнением параболы, гиперболы, степенной и т. п.*

По количеству взаимодействующих факторов *связи делятся на* ***парную*** *(однофакторную) и* ***множественную*** *(многофакторную) связи. При парной связи на результативный признак действует один факторный, при множественной – несколько факторных признаков.*

Исследование статистической связи проводится в следующем порядке:

- *Качественный анализ связи* - определение состава признаков, предварительный анализ формы связи;
- *сбор данных на основе статистического наблюдения*;
- *корреляционный анализ*;
- *регрессионный анализ* (аналитическое описание связи).

Основным методом изучения статистической взаимосвязи является **ее моделирование на основе корреляционного и регрессионного анализа**.

К задачам корреляционного анализа относится *количественное определение тесноты связи* между двумя признаками при парной связи или между результативным и несколькими факторными при множественной связи.

Регрессионный анализ *заключается в определении аналитического выражения связи в виде уравнения регрессии.* **Регрессией называется зависимость среднего значения случайной величины результативного признака от величины факторного**, а ***уравнением регрессии – уравнение описывающее корреляционную зависимость между результативным признаком и одним или нескольким факторными.***

2.2. Парная корреляция

Наиболее полно в статистике разработана методология парной корреляции, рассматривающей влияние вариации одного факторного признака на вариацию результативного.

Исследование парной корреляции осуществляется на основе корреляционного анализа, который предполагает последовательное решение ряда задач:

- выявление связи;
- описание связи в табличной и графической формах;
- измерение тесноты связи;
- формулировка выводов о характере существующей связи.

2.2.1. Задача выявления связи между факторным и результативным признаками может быть решена при помощи следующих приёмов:

- *визуализации связи* (построение и визуальный анализ корреляционного поля);
- *использования результатов аналитической группировки* и др.

Корреляционное поле *представляет собой точечный график в системе координат {x,y}*. Каждая точка соответствует единице совокупности. Положение точек на графике определяется величиной двух признаков – факторного и результативного. Точки корреляционного поля могут располагаться на графике хаотично, без всякой закономерности - в этом случае делается вывод об отсутствии связи между признаками; или определённым образом вдоль некоторой гипотетической линии – это свидетельствует о существовании связи между признаками.

При втором способе – *использовании результатов аналитической группировки* **связь считается установленной, если группировка показывает изменение среднего значения результативного признака в группах при изменении факторного признака** (основания группировки).

2.2.2. Описание выявленной связи при проведении корреляционного анализа *проводится в двух формах – табличной и графической*.

При табличном описании связи *статистические единицы группируются по значению факторного признака (располагаются в порядке его возрастания или убывания)*:

Таблица2.1

Табличное описание связи

x_i	x_1	…….	x_{n-1}	x_n
$\tilde{y}_i$	$\tilde{y}_1$	……	$\tilde{y}_{n-1}$	$\tilde{y}_n$

Графическое описание связи *заключается в построении* **линии эмпирической регрессии** – ***ломаной линии, соединяющей на корреляционном поле точки, абсциссами которых являются значения факторного признака*** (индивидуальные значения или групповые значения), а ***ординатами – средние значения результативного признака.***

Линия эмпирической регрессии отражает основную тенденцию рассматриваемой зависимости. Если по своему виду она приближается к прямой линии, то можно предположить наличие прямолинейной связи между признаками.

2.2.3. Теснота связи *показывает меру влияния факторного признака на общую вариацию результативного признака*.

Для описания корреляционной связи используется зависимость $\tilde{y} = F(x)$, которая проявляется только на уровне статистической совокупности. Так как на результат всегда действует множество факторов, то для каждой отдельной единицы наблюдения значение результативного признака состоит из двух частей:

$y_i = \tilde{y}_i + \varepsilon_i$,

где $\tilde{y}_i$ – локальная средняя, характеризующая значение результативного признака, сформированное под воздействием только данного фактора x_i;

$\varepsilon_i = (y_i - \tilde{y}_i)$ - отклонение, характеризующее вариацию результативного признака под влиянием неучтённых факторов.

Таким образом, *теснота связи является характеристикой соотношения между локальной средней* $\tilde{y}_i$ *и отклонением* ε_i. ***Через тесноту связи определяется, в какой степени влияют на результат учтённые и неучтённые факторы.***

На эмпирическом уровне, при проведении корреляционного анализа теснота связи измеряется с помощью интегральных показателей, построенных на правиле «сложения дисперсии», в соответствии с которым **общая дисперсия результативного признака разлагается на внутригрупповую и межгрупповую**:

$\sigma_y^2 = \overline{\sigma_i^2} + \delta^2$,

где $\overline{\sigma_i^2}$ - средняя из внутригрупповых дисперсий;

δ^2 - межгрупповая дисперсия.

Через соотношение дисперсий определяются показатели, измеряющие степень тесноты связи между результативными и факторными признаками: коэффициент детерминации η^2 и эмпирическое корреляционное отношение η.

- **Коэффициент детерминации** рассчитывается по формуле:

$$\eta^2 = \frac{\delta^2}{\sigma_y^2} = 1 - \frac{\overline{\sigma_i^2}}{\sigma_y^2}.$$

Приведенное отношение *определяет удельный вес вариации, объясняемой влиянием учтенного фактора на результат, в общей вариации результативного признака*. Показатель изменяется в диапазоне от 0 до 1. **При** $\eta^2 = 0$ межгрупповая дисперсия δ^2 =0, - это означает, что локальные средние во всех распределениях результативного признака строго одинаковы, центры распределений не смещаются; **связь между признаками отсутствует**. **При** $\eta^2 = 1$ межгрупповая дисперсия δ^2 равна общей

дисперсии результативного признака $\delta^2 = \sigma_y^2$; следовательно, $\overline{\sigma_i^2} = 0$, и внутригрупповые значение результативного признака не варьируют, то есть $y_i = \tilde{y}_i$. Это означает, что на значения результативного признака влияют только учтенные факторы, и **связь между признаками является функциональной**: значению факторного признака соответствует единственное значение результативного.

Коэффициент детерминации сложно интерпретируется, поэтому на его основе рассчитывается ещё один показатель тесноты связи – *эмпирическое корреляционное отношение* η.

- **Эмпирическое корреляционное отношение** рассчитывается по формуле:

$$\eta = \sqrt{\eta^2} = \sqrt{1 - \frac{\overline{\sigma_i^2}}{\sigma_y^2}}.$$

Диапазон изменения этого показателя: $\eta = \{0 \div 1\}$. *Нулевое значение эмпирического корреляционного отношения означает отсутствие связи между результативным и факторным признаками, при* $\eta = |1|$ *связь классифицируется как функциональная.*

Используя численное значение эмпирического корреляционного отношения η, связь можно классифицировать по шкале Чеддока, приведенной в таблице 2.2.:

Таблица 2.2.

Шкала Чеддока

η	$0 \div 0{,}1$	$0{,}11 \div 0{,}3$	$0{,}31 \div 0{,}5$	$0{,}51 \div 0{,}7$	$0{,}71 \div 0{,}9$	$0{,}91 \div 0{,}99$	$0{,}991 \div 1$
Характеристика связи	Отсутствует	Слабая	умеренная	заметная	тесная	сильная	Функциональная

Если известно, что между результативным и факторным признаком существует линейная связь, то для оценки её тесноты используется **линейный коэффициент корреляции** $r_{y,x}$, рассчитываемый по формуле:

$$r_{y.x} = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right] \cdot \left[\sum y^2 - \frac{\left(\sum y\right)^2}{n}\right]}} = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sigma_x \cdot \sigma_y}.$$

Значение линейного коэффициента корреляции имеет важное значение для исследований, в которых распределение признака близко к нормальному. Его значение меняется в интервале $-1 \le r_{y,x} \le +1$. Отрицательные значения $r_{y,x}$ свидетельствуют о наличии обратной связи между признаками, положительные – о прямой связи. При $r_{y,x}$=0 связь между признаками отсутствует. Для классификации связи по значению линейного коэффициента корреляции используется шкала Чеддока.

2.2.4. Выводы по результатам корреляционного анализа включают в себя констатацию факта наличия связи, определение её направления, предварительную оценку формы связи по линии эмпирической регрессии и классификацию связи по степени её тесноты.

2. 3. Парная регрессия на основе метода наименьших квадратов

Парная регрессия характеризует связь между двумя признаками: факторным и результативным.

Задача построения уравнения регрессии для одного факторного и одного результативного признака формулируется следующим образом:

Пусть имеется набор значений двух переменных: результативного признака y_i и факторного признака x_i. Между этими переменными существует объективная связь вида: $y_i = f(x_i) + \varepsilon_i$. Необходимо по данным наблюдения (y_i, x_i, i=1,n) подобрать функцию $\hat{y} = F(x)$, наилучшим образом описывающую существующую связь.

При подборе функции последовательно решаются две задачи:

- **Определяется вид функциональной зависимости, то есть проводится спецификация модели;**
- **Рассчитываются значения параметров уравнения регрессии**.

В парной регрессии выбор вида математической функции может быть осуществлён разными методами:

- *аналитическим, исходя из материальной природы связи;*
- *графическим, на основе линии эмпирической регрессии;*
- *на основе показателей качества уравнения регрессии.*

Показателем качества уравнения регрессии является величина остаточной дисперсии:

$$\sigma^2_{y-\hat{y}} = \frac{\sum_{i=1}^{n}(y-\hat{y})^2}{n}.$$

Остаточная дисперсия рассчитывается для уравнений регрессии, построенных по разным математическим функциям. *Лучшим по качеству является уравнение, для которого* $\sigma^2_{y-\hat{y}} \to \min$.

При построении уравнений парной регрессии чаще всего используют следующие уравнения:

- прямой $\hat{y} = a + bx$,
- параболы второго порядка $\hat{y} = a + bx + cx^2$,
- гиперболы $\hat{y} = a + \frac{b}{x}$,
- степенной $\hat{y} = a \cdot x^b$,
- показательной $\hat{y} = a \cdot b^x$,
- логистической кривой $\hat{y} = \frac{a}{1 + bc^{-CX}}$ и т.д.

Оценка параметров уравнений регрессии может быть проведена разными методами.

Классический подход к оцениванию параметров основан на методе наименьших квадратов (МНК).

Метод наименьших квадратов *позволяет получить такие оценки параметров уравнения регрессии, которые минимизируют функционал вида* $S = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \to \min$;

Применение метода наименьших квадратов для расчёта параметров уравнения регрессии рассмотрим на примере прямолинейной зависимости $\hat{y} = a + bx$.

Подставим аналитическое выражение прямолинейной функции $\hat{y} = a + bx$ в функционал S:

$$S = \sum (y - a - bx)^2 \to \min.$$

Для нахождения минимума функции двух переменных a и b необходимо взять частные производные по каждому параметру и приравнять их к нулю:

$\frac{dS}{da} = 0; \quad \frac{dS}{db} = 0$.

В результате получаем систему нормальных уравнений:

$$\begin{cases} na + b \cdot \sum x = \sum y; \\ a \cdot \sum x + b \cdot \sum x^2 = \sum xy. \end{cases}$$

Решение системы уравнений даёт оценки параметров a и b:

$$a = \frac{\sum y - b \cdot \sum x}{n}; \quad b = \frac{\frac{\sum x \cdot \sum y}{n} - \sum xy}{\left(\sum x\right)^2 - \sum x^2};$$

В линейном уравнении регрессии параметр a показывает усреднённое влияние на результативный признак неучтённых факторов. Формально a= $\bar{y}$ при x=0. ***Интерпретация параметра a*** *как среднего значения результативного признака возможно лишь при условии, что среди наблюдаемых значений факторного признака есть значения, равные или близкие к 0.* ***Параметр b*** *в уравнении линейной регрессии называется* ***коэффициентом регрессии.*** **Коэффициент регрессии** ***показывает, на сколько в среднем изменится значение результативного признака при увеличении факторного на единицу собственного измерения.***

Для получения качественного уравнения регрессии необходимо чтобы данные наблюдения соответствовали следующим требованиям:

- **число наблюдений должно в 6-7 раз превышать число рассчитываемых параметров при переменных** x**.** Таким образом, искать линейную регрессию имея менее 6 наблюдений не имеет смысла;
- **распределение единиц наблюдения по факторному признаку должно быть однородным и подчиняться нормальному закону распределения**.

Аналогичным образом на основе МНК рассчитываются параметры нелинейной регрессии.

Для параболы второго порядка: $\hat{y} = a + bx + cx^2$ получаем систему нормальных уравнений следующего вида:

$$\begin{cases} \sum y = n \cdot a + b \cdot \sum x + c \cdot \sum x^2 \\ \sum x \cdot y = a \cdot \sum x + b \cdot \sum x^2 + c \cdot \sum x^3 \\ \sum x^2 \cdot y = a \cdot \sum x^2 + b \cdot \sum x^3 + c \sum x^4 \end{cases}$$

Для показательной функции $\hat{y} = a \cdot b^x$ предварительно необходимо выполнить процедуру линеаризации, то есть привести функцию к линейному виду. Это можно сделать, прологарифмировав обе части уравнения:

$\ln \hat{y} = \ln a + x \ln b.$

Введём следующие обозначения: $\ln \hat{y} = Y$; $\ln a = A$; $\ln b = B$. В этом случае уравнение регрессии принимает вид: Y=A+B◎x, то есть приводится к линейному уравнению.

По полученному уравнению регрессии можно построить *теоретическую линию регрессии.* Она должна проходить через точки, ***абсциссами которых являются значения факторного признака*** (индивидуальные значения или групповые значения), а ***ординатами –*** *соответствующие им* ***теоретические значения результативного признака,*** *рассчитанные по уравнению регрессии****.***

Следует отметить, что парная регрессия хорошо изучена и входит в стандартные пакеты программ (например, “Excel”).

2.4. Оценка существенности парной корреляционной связи

Для проверки существенности парной корреляционной связи, то есть соответствия полученной модели данным наблюдения используется следующий подход: ***модель признаётся значимой, если таковыми являются параметры модели или показатели тесноты связи***. При этом выясняется, не являются ли вычисленные значения параметров регрессии случайными величинами?

- **Значимость параметров линейной модели определяется с помощью t-критерия Стьюдента.**

Для каждого из параметров уравнения регрессии вычисляются расчетные (фактические) значения t-критерия:

для параметра a: $t_{a=0} = a \cdot \frac{\sqrt{n-2}}{\sigma_{y-\hat{y}}}$;

для параметра b: $t_b = b \cdot \frac{\sqrt{n-2}}{\sigma_{y-\hat{y}}} \cdot \sigma_x$

где n – число наблюдений;

$\sigma_{y-\hat{y}} = \sqrt{\frac{\sum (y-\hat{y})^2}{n}}$ - остаточное среднее квадратическое отклонение результативного признака *y* от выравненных значений $\hat{y}$, рассчитанных по модели;

$\sigma_x = \sqrt{\frac{\sum (x-\bar{x})^2}{n}}$ - среднее квадратическое отклонение факторного признака x_i от общей средней $\bar{x}$.

Вычисленные значения t-критериев сравниваются с критическими значениями $t_{кр}$, определёнными по таблице распределения Стьюдента с учётом принятого уровня значимости α и числа степеней свободы вариации $\nu = n-2$.

Параметр признаётся значимым, ***если расчетное значение t-критерия не меньше критического*** $t_{кр}$**.** В этом случае найденные значения параметров не являются случайными, а уравнение регрессии признаётся существенным.

Значимость линейной регрессии можно оценить по линейному коэффициенту корреляции. ***Модель признаётся значимой, если расчётное значение t-критерия для линейного коэффициента корреляции превышает табличное, то есть выполняется неравенство:*** $t_{r_{yx}} > t_{\alpha,\nu}$.

Расчётное значение t-критерия для линейного коэффициента корреляции определяется по формуле:

$$t_{r_{yx}} = r_{yx} \cdot \sqrt{\frac{n-2}{1-r_{yx}^2}}.$$

- *Для* ***нелинейных моделей их существенность проверяется с помощью F-критерия (****критерия* ***Фишера).***

Если расчетное значение ***F-критерия*** *превышает его критическое значеение, т.е. выполняется условие* $F_{факт} > F_{кр}$ то корреляционная модель признается надежной. Фактический уровень критерия Фишера рассчитывается по формуле:

$$F_{факт} = \frac{\eta_т^2}{1-\eta_т^2} \cdot \frac{n-m}{m-1};$$

где $\eta_т^2$ - теоретический коэффициент детерминации,

m - количество параметров уравнения регрессии.

Теоретический коэффициент детерминации $\eta_т^2$ является показателем тесноты связи результативного и факторного признака в уравнении регрессии. Рассчитывается $\eta_т^2$ на основе правила сложения дисперсий.

При наличии уравнения регрессии, описывающего существующую связь, степень влияния факторного признака на результативный признак может быть выражена следующим образом:

$y_i = \hat{y}_i + \varepsilon_i$,

где $\hat{y}_i$ - теоретическое (сглаженное) значение результативного признака, просчитанное по уравнению регрессии.

Соответственно, ***дисперсия результативного признака σ_y^2 должна включить в себя дисперсию теоретических значений результативного признака (объяснённую) $\sigma_{\hat{y}}^2$ и дисперсию отклонений эмпирических (наблюдаемых) значений результативного признака от теоретических $\sigma_{y-\hat{y}}^2$ (остаточную).***

Таким образом, $\sigma_y^2 = \sigma_{\hat{y}}^2 + \overline{\sigma_{y-\hat{y}}^2}$,

где $\sigma_y^2 = \dfrac{\sum(y-\bar{y})^2}{n}$ - общая дисперсия результативного признака;

$\sigma_{\hat{y}}^2 = \dfrac{\sum(\hat{y}-\bar{y})}{n}$ - объяснённая дисперсия результативного признака4

$\sigma_{y-\hat{y}}^2 = \dfrac{\sum(y-\hat{y})^2}{n}$ - остаточная дисперсия результативного признака.

Объяснённая дисперсия $\sigma_{\hat{y}}^2$ *характеризует влияние фактора, включённого в модель, на общую вариацию результативного признака.*

Остаточная дисперсия $\sigma_{y-\hat{y}}^2$ *характеризует влияние факторов, не включённых в уравнение регрессии, на вариацию результативного признака.*

Теоретический коэффициент детерминации определяется через соотношение объясненной и общей дисперсии результативного признака.

$\eta_T^2 = \dfrac{\sigma_{\hat{y}}^2}{\sigma_y^2}$, так как $\sigma_{\hat{y}}^2 = \sigma_y^2 - \sigma_{y-\hat{y}}^2$, то $\eta_т^2 = 1 - \dfrac{\sigma_{y-\hat{y}}^2}{\sigma_y^2}$, где

$\sigma_{y-\hat{y}}^2 = \dfrac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}$ - остаточная дисперсия результативного признака

$\sigma_y^2 = \dfrac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}$ - общая дисперсия результативного признака.

Критическое значение *F*-критерия выбирается по таблицам распределения

Фишера- Снедекора (*F*-распределения) при числе степеней свободы $V_1 = m$-1 и $V_2 = n$-*m* и уровне значимости α = 0,05.

Если $F_p > F_{кр}$, выбранного по таблицам распределения Фишера-Снедекора (F–распределения) при числе степеней свободы $V_1 = m–1$ и $V_2 = n – m$ и заданном уровне значимости , связь признается достоверной. Таблица распределения Фишера приведена в Приложении 4.

2.5. Множественная корреляция

Двухмерные корреляционные модели (парная корреляция) используются в случаях, когда среди факторов, влияющих на результативный признак, есть доминирующий. Таких связей немного, чаще встречаются зависимости результативного признака от нескольких факторных, так как экономические явления находятся под влиянием значительного числа одновременно и совокупно действующих факторов. *Для описания совместного влияния одновременно действующих факторов на результат используют множественные корреляционные модели вида*: $\hat{n} = f(x_1, x_2, ...x_K)$

Модели подобного класса используются при изучении спроса, функции потребления, доходности акций, оценке рисков и т.д.

Задача множественного корреляционно-регрессионного анализа в общем виде формулируется следующим образом**:**

Пусть некоторая статистическая совокупность, состоящая из n единиц наблюдения обладает определённым набором признаков, один из которых играет роль результативного y, а остальные – факторных $(x_1, x_2, ...x_K)$***. На основе наблюдаемых значений всех признаков требуется выявить и описать связь между ними в виде множественной корреляционной модели вида:*** $\hat{n} = f(x_1, x_2, ...x_K)$**.**

Решение данной задачи требует последовательного выполнения следующих этапов исследования множественной корреляционной связи:

- предварительный отбор факторов, включаемых в модель;
- предварительное описание связи;
- уточнение модели на основе анализа корреляционной матрицы; определение тесноты связи;
- оценка надёжности множественной корреляционной модели; интерпретация модели.

2.5.1. Предварительный отбор факторов

Изучение множественной регрессии (корреляции) требует измерения не только прямого воздействия каждого фактора на результат, но и учёта влияния факторов друг на друга, то есть *учета наличия межфакторных связей*. ***Общее число связей*** всегда значительно больше числа факторов, включаемых в модель. Оно определяется выражением:

$$l = \frac{k(k+1)}{2};$$

где k – количество факторных признаков, включенных в модель.

В общем случае, при большом числе учитываемых факторов необходимо строить сложные модели, требующие проведения сложных расчётов; модели получаются громоздкими. С другой стороны,- чем большое количество факторов учитывается, тем адекватнее построенная модель.

Для разрешения указанного противоречия предварительно ограничивается число учитываемых факторов. ***Целесообразность*** их ***включения в модель определяется следующими соображениями:***

- *они должны быть соизмеримы, иметь количественное выражение;*
- *между факторами не должно быть тесной связи*;
- *они должны объяснять вариацию результативного признака*.

При включении в модель коррелированных факторов невозможно определить изолированное влияние таких факторов на результативный показатель, а оценки параметров уравнения множественной регрессии будут ненадёжными, зависимыми от наблюдений.

2.5.2. Предварительное описание множественной корреляционной связи (МКС) осуществляется через построение соответствующего уравнения регрессии. Практика показывает, что можно использовать следующие пять функций, так как они описывают все реально существующие зависимости между социально-экономическими явлениями:

1. линейная $\hat{y} = a_0 + a_1 x_1 + a_2 x_2 + ... + a_k b_k$;
2. степенная $\hat{y} = a_0 \cdot x_1 a_1 \cdot x_2 a_2 \cdot ..... \cdot x_k a_k$;
3. показательная (экспотенциональная) $\hat{y} = e^{a_0 + a_1 X_1 + a_2 x_2 + ... a_K X_K}$;
4. параболическая $\hat{y} = a_0 + a_1 x^2{}_1 + a_2 x_2{}^2 + ... + a_k b_k^2$;
5. гиперболическая $\hat{y} = a_0 + \frac{a_1}{x_1} + \frac{a_2}{x_2} + ... + \frac{a_k}{x_k}$.

Работать с нелинейными функциями сложно, поэтому ***основное значение имеют линейные модели*** в силу их простоты и логичности экономической интерпретации. Нелинейные формы всегда можно привести к линейной, используя известный в математике приём «линеаризации функций».

Величина каждого параметра в уравнении прямой может быть определена по методу наименьших квадратов. Для этого в функционал $S = \sum (y - \hat{y})^2 \to \min$ подставим выражение прямолинейной функции. В результате подстановки получаем:

$$\sum (y - a_0 - a_1 x_1 - a_2 x_2 - ..... - a_K x_K)^2 \to \min$$

Продифференцировав полученное выражение по каждому из параметров, получаем систему нормальных уравнений вида:

$$\begin{cases} \sum y = a_0 \cdot n + a_1 \cdot \sum x_1 + a_2 \cdot \sum x_2 + ... + a_k \cdot \sum x_k^2 \\ \sum y \cdot x_1 = a_0 \cdot \sum x_1 + a_1 \cdot \sum x_1^2 + a_2 \cdot \sum x_1 \cdot x_2 + ... + a_k \cdot \sum x_k^2 \\ \sum y \cdot x_2 = a_0 \cdot \sum x_2 + a_1 \sum x_1 x_2 + a_2 \cdot \sum x_2 \cdot x_k + ... + a_k \sum x_k^2 \\ ..... \\ \sum y \cdot x_k = a_0 \cdot \sum x_k + a_1 \sum x_1 \cdot x_k + a_2 \cdot \sum x_2 \cdot x_k + ... + a_k \sum x_k^2 \end{cases}$$

Число уравнений в системе равно числу параметров, поэтому задача их нахождения является разрешимой.

При выборе формы уравнения множественной регрессии необходимо иметь в виду:

1. Чем сложнее функция, тем хуже интерпретируются параметры модели.
2. Сложные функции (полиномы) с большим количеством факторов требуют большого числа наблюдений (на каждый параметр не менее 6 наблюдений)

2.5.3. Окончательный отбор факторов

Окончательный отбор факторов, то есть уточнение корреляционной модели проводится на основе анализа корреляционной матрицы. **Корреляционная матрица** состоит из *парных линейных коэффициентов корреляции* r_{yx}, отражающих тесноту связи результативного и факторного признака и коэффициентов r_{xixj}, отражающих тесноту связи между i-м и j-м факторными признаками, рассчитываемых по известной формуле:

$$r_{y,x} = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sqrt{\left[x^2 - \frac{(\sum x)^2}{n}\left[y^2 - \frac{(\sum x)^2}{n}\right]\right]}} \quad ; \qquad r_{y,x} = r_{i,j} \ .$$

Общий вид корреляционной матрицы приведен в таблице 2.3.

Таблица 2.3

Общий вид корреляционной матрицы

	y	x_1	x_2	…..	x_j	……	x_p
y	1	$r_{y,x1}$	$r_{y,x2}$	……	$r_{y,xj}$		$r_{y,xp}$
x_1	$r_{y,x1}$	1	$r_{x1,x2}$		$r_{x1,xj}$		$r_{x1,xp}$
x_2	$r_{y,x2}$	$r_{x1,x2}$	1		$r_{x2,xj}$		$r_{x2,xp}$
…				…...		……	
x_j	$r_{y,xj}$	$r_{xj,x1}$	$r_{xj,x2}$		1		$r_{xj,x2}$
…				…...		……	
x_p	$r_{y,xp}$	$r_{x1,xp}$	$r_{x2,xp}$	…….	$r_{xj,xp}$	……	1

Факторы, теснота связи между которыми оценивается как высокая, считаются коллинеарными. ***Окончательный отбор факторов заключается во включении в модель независимых (неколлинеарных) факторов.*** *Процедура отбора может быть осуществлена способом шаговой регрессии*:

- *Для обоснования включения факторов в модель* оценивается первая строка матрицы, отражающая связь факторов с результатом. В модель включаются факторы, оказывающие наибольшее влияние на результат (с максимальными линейными коэффициентами корреляции).
- *Оценивается теснота межфакторной связи*. Если она высока, то между данными факторами существует тесная зависимость, то есть факторы коллинеарны, а коллинеарность (тесная зависимость между факторами) существенно искажает результаты исследования. ***Связь относится к коллинеарной, если:*** $\left|r_{ij}\right| \geq 0.8$**.** Один из коллинеарных факторов необходимо исключить из модели. Исключается фактор с меньшим значением линейного коэффициента корреляции $r_{y,x}$.
- Для включения недостающих факторов в модель рассматриваются факторы, не вошедшие в модель на первом этапе. Из них выбирается фактор с максимальным значением линейного коэффициента корреляции. Он добавляется к уже отобранным факторам; проводится проверка нового фактора на коллинеарность с уже отобранными и т. д.

2.5.4. Оценка тесноты связи

Оценка тесноты множественной корреляционной связи проводится на основе двух показателей: множественного коэффициента детерминации $R^2_{yx1...xk}$ и множественного коэффициента корреляции $R_{yx1...xk}$.

Сложность расчёта этих показателей связана с необходимостью учёта межфакторных связей. Гипотетически данные показатели рассчитываются по формулам:

$$R^2_{yx1...xk} = 1 - \frac{\sigma^2_{Y-\hat{Y}}}{\sigma^2_Y};$$

$$R_{yx1...xk} = \sqrt{R^2_{yx1...xk}}.$$

На практике множественный коэффициент корреляции R рассчитывается на основе определителей корреляционной матрицы:

$$R_{yx1...xk} = \sqrt{1 - \frac{\Delta r}{\Delta r_{xixj}}};$$

где Δr - общий определитель корреляционной матрицы;

Δr_{xixj} - определитель матрицы межфакторной корреляции.

Для двухфакторной модели множественный коэффициент корреляции определяется по формуле:

$$R_{yx1x2} = \sqrt{\frac{r^2_{yx1} + r^2_{yx2} - 2 \cdot r_{yx1} \cdot r_{yx2} r_{x1x2}}{1 - r^2_{x1x2}}}.$$

Диапазон изменения множественного коэффициента корреляции $R_{yx1...xk} = |0 \div 1|$. «0» означает отсутствие связи, «1» - наличие функциональной множественной связи между признаками. *Для классификации тесноты связи используется шкала Чеддока*.

2.5.5. Оценка надёжности модели

Для оценки надёжности выявленной связи сравнивается множественный коэффициент корреляции с линейными корреляционными коэффициентами корреляции между результатом и факторными признаками, включёнными в модель.

Связь признаётся надёжной, если $R_{yx1...xk} \geq \max\{r_{yxj}\}$.

2.5.6. Интерпретация параметров модели

Завершающим этапом множественной корреляции является интерпретация параметров построенной корреляционной модели. ***Чем больше величина этих параметров*** (коэффициентов регрессии), ***тем значительнее влияние данных факторов на результат***. *Важное значение имеют знак перед коэффициентами регрессии*: ***знак «+»*** свидетельствует о росте результата при увеличении факторного признака, ***знак «-»*** о снижении значения результативного признака при росте факторного.

Контрольные вопросы и задания

1. Чем отличается статистическая связь от функциональной?
2. Дайте определение корреляционной связи.
3. Сформулируйте задачу корреляционного анализа?
4. Каковы задачи регрессионного анализа?
5. Какие методы применяются при выявлении факта наличия связи между статистическими признаками?
6. Что понимается под теснотой связи и как она оценивается?
7. Как формулируется основная задача регрессионного анализа парной корреляционной связи?

8. В чем суть метода наименьших квадратов?

9. Как оценивается надежность парной линейной корреляционной модели?

10. Каковы особенности оценки надежности парной нелинейной модели?

11. Перечислите основные этапы построения множественных уравнений регрессии?

3. Ряды динамики

3.1. Понятие и классификация рядов динамики

В статистике динамикой принято называть процесс развития, движения социально-экономических явлений во времени. Для отображения таких процессов строятся **ряды динамики** (хронологические, временные, динамические ряды), **представляющие собой последовательность упорядоченных во времени значений статистического показателя**. Любой *ряд динамики состоит из двух элементов:*

1. показатель времени t_i, *под которым понимается момент или период времени, к которым относятся числовые значения показателей;*

2. уровень ряда y_i, *под которым понимается значение статистического показателя, относящееся к определенному моменту или периоду времени.*

Каждый ряд динамики может быть представлен в табличной форме - в виде пар значений t_i и y_i (таблица 3.1); и в графической форме - в виде линейной диаграммы

Таблица 3.1.

Ряд динамики

t_1	t_2	t_i	t_n
y_1	y_2	y_i	y_n

При обработке статистических данных *используются ряды динамики, различающиеся по следующим признакам: по времени, форме представления уровней, числу показателей, по расстоянию между датами или интервалами.*

По времени различают **моментные и интервальные ряды динамики**.

В **моментных** *рядах уровни выражают состояние явления на критический момент времени* – начало месяца, квартала, года и т.д. Например, численность населения, численность работающих и т.д. В таких рядах каждый последующий уровень полностью или частично содержит значение предыдущего уровня, поэтому суммировать уровни нельзя, так как это приводит к повторному счету.

В **интервальных** – *уровни отражают состояние явления за определенный период времени* – сутки, месяц, год и т.д; это ряды показателей объема производства, объема продаж по месяцам года, количества отработанных человеко-дней и т.д.

По *форме представления уровней* различают **ряды абсолютных, относительных и средних величин.**

По числу показателей выделяют **изолированные и комплексные ряды** динамики (многомерные).

Изолированный ряд строится по отдельному показателю, комплексный – по системе взаимосвязанных показателей.

По расстоянию между датами или интервалами ряды динамики делятся на ряды с **равноотстоящими и неравноотстоящими уровнями.**

В рядах с равноотстоящими уровнями расстояние между датами или периодами одинаково, в рядах с равноотстоящими уровнями – оно различно.

Чтобы ряды динамики давали правильное представление о процессах, которые они представляют, при их составлении *необходимо соблюдать определенные требования,* основными из которых являются:

1. Обеспечение сопоставимости уровней – использование единых методик расчета показателей, одинаковых единиц измерения, единого круга объектов наблюдения, единых территориальных границ, единого содержания показателей.

2. Систематизация уровней в хронологическом порядке - в рядах динамики не должно быть пропусков отдельных уровней. Если данных не хватает, то их восполняют условными расчетными значениями уровней.

С помощью рядов динамики в статистике решают следующие ***задачи***:

- **получение характеристик интенсивности изменения явления во времени и характеристик отдельных уровней;**
- **выявление и количественная оценка основной долговременной тенденции развития явления;**
- **изучение периодических и сезонных колебаний явления;**
- **экстраполяция и прогнозирование**.

3.2 .Система характеристик динамического ряда

Система характеристик динамического ряда включает в себя:

- ***индивидуальные (частные) характеристики интенсивности;***
- ***сводные (обобщающие) характеристик интенсивности.***

К индивидуальным показателям интенсивности изменения явления относятся:

- *абсолютный прирост* Δy_i *;*

- *темп роста* T_i *(коэффициент роста* K_i *);*

- *темп прироста* $T_i^{'}$ *(коэффициент прироста* $K_i^{'}$ *);*

- *абсолютное значение одного процента прироста* A_i

- *пункт роста* P_i *.*

Первые три из перечисленных характеристик можно рассчитать двумя способами в зависимости от применяемой базы сравнения. База сравнения может быть постоянной или переменной. Соответственно, можно рассчитать *базисные или цепные характеристики динамического ряда*.

Абсолютный прирост Δy_i *характеризует размер увеличения (уменьшения) уровня ряда по сравнению с выбранной базой*:

- **цепной абсолютный прирост** показывает, на сколько изменилось значение данного уровня по сравнению с предыдущим, то есть приращение уровня по сравнению с предыдущим:

$$\Delta y_{ц_i} = y_i - y_{i-1}, \quad i = \overline{2,n}.$$

- **базисный абсолютный прирост** показывает, на сколько изменилось значение данного уровня по сравнению с исходным (начальным) уровнем:

$$\Delta y_{б_i} = y_i - y_1, \quad i = \overline{2,n},$$

где y_1 - начальный уровень ряда.

Между базисными и цепными абсолютными приростами существует взаимосвязь: сумма всех цепных абсолютных приростов равна базисному приросту конечного уровня:

$$\sum_{i=2}^{n} \Delta y_{ц_i} = \Delta y_{б_n},$$

где y_n - конечный уровень ряда.

Коэффициент роста (относительный прирост) *характеризует интенсивность изменения уровней ряда (скорость изменения уровней).* Он показывает, *во сколько раз уровень данного периода выше или ниже базисного уровня*. Этот показатель как

относительная величина, выраженная в долях единицы, называется **коэффициентом (индексом) роста**; выраженная в процентах, называется **темпом роста.**

- **Цепной коэффициент роста** показывает, во сколько раз текущий уровень выше или ниже предыдущего:

$$K_{ц_i} = \frac{y_i}{y_{i-1}}, \quad i = \overline{2,n};$$

- **базисный коэффициент роста** показывает, во сколько раз текущий уровень выше или ниже начального уровня:

$$K_{б_i} = \frac{y_i}{y_1}, \quad i = \overline{2,n}.$$

Между базисными и цепными темпами (коэффициентами) роста имеется зависимость: *произведения последовательных цепных коэффициентов роста равно базисному коэффициенту роста за весь промежуток времени*:

$$K_{ц_2} \cdot K_{ц_3} \cdot ..... \cdot K_{ц_n} = K_{б_n};$$

а *частное от деления текущего базисного коэффициента роста на предыдущий базисный коэффициент роста равно текущему цепному коэффициенту роста*:

$$K_{ц_i} = \frac{K_{б_i}}{K_{б_{i-1}}}, \quad i = \overline{2,n}.$$

Коэффициент роста всегда есть положительная величина, область его допустимых значений- $(0 - + \infty)$.

Коэффициент прироста *характеризует относительную скорость изменения уровня ряда в единицу времени. Показывает, на какую долю единицы (или процент) уровень данного периода или момента времени выше или ниже базисного уровня.*

- **Цепной коэффициент прироста** рассчитывается по формуле:

$$K'_{ц_i} = \frac{\Delta y_{ц_i}}{y_{i-1}} = \frac{y_i - y_{i-1}}{y_{i-1}};$$

Цепной темп прироста равен: $T_{ц_i} = K_{Ц_i} \cdot 100\%$. Он показывает, на сколько процентов уровень текущего периода выше или ниже предыдущего уровня.

- **Базисный коэффициент прироста** равен:

$$K'_{б_i} = \frac{\Delta y_{б_i}}{y_1} = \frac{y_i - y_1}{y_1};$$

а *базисный темп прироста* - $T'_{б_i} = \frac{\Delta y_{б_i}}{y_1} \cdot 100\%$. $T'_{б_i}$ показывает, на сколько процентов уровень текущего периода выше или ниже начального уровня ряда.

Между коэффициентом (темпом) роста и коэффициентом (темпом) прироста существует зависимость:

$K'_i = K_i - 1$ или $T'_i = T_i - 100\%$.

Абсолютное значение одного процента прироста *используется для оценки значения полученного темпа прироста. Он показывает, какое абсолютное значение соответствует одному проценту прироста.* Показатель считается по цепным характеристикам:

$$A_i = \frac{\Delta y_{ц_i}}{T_{ц_i}} = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100} = \frac{y_{i-1}}{100}.$$

Пункты роста *используется в тех случаях, когда сравнение производится с отдалением периода времени, принятого за базу*. Они представляют собой разность базисных темпов роста двух смежных периодов:

$$P_i = T_{б_i} - T_{б_{i-1}} = \frac{y_i}{y_1} - \frac{y_{i-1}}{y_1} = \frac{y_i - y_{i-1}}{y_1} = \frac{\Delta y_{ц_i}}{y_1}.$$

Пункты роста можно суммировать, в результате получаем базовый темп прироста последнего периода:

$$\sum_{i=2}^{n} P_i = T_{б_n}.$$

Вторая часть системы характеристик динамического ряда состоит из обобщающих характеристик, к которым относятся его средние показатели и характеристики вариации уровней:

- *средний уровень ряда* $\overline{y}$;
- *общий абсолютный прирост* Δ;
- *средний абсолютный прирост* $\overline{\Delta}$;
- *средний темп роста* $\overline{T}$ ($\overline{K}$);
- *средний темп прироста* $\overline{T}'$ ($\overline{K}'$);
- *дисперсия и среднее квадратическое отклонение уровней ряда* σ_y^2, σ_y;
- *коэффициент вариации уровней ряда* V_y.

- *Расчет среднего уровня ряда динамики определяется видом ряда и величиной интервала, соответствующего каждому уровню*. **Средний уровень характеризует наиболее типичную величину уровней, центр ряда**.

В интервальных рядах с равноотстоящими интервалами средний уровень ряда определяется по *формуле средней арифметической простой*:

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}.$$

В интервальных рядах *с неравноотстоящими уровнями используется формула средней арифметической взвешенной*:

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i \cdot t_i}{\sum_{i=1}^{n} t_i};$$

t_i - длительность интервала.

В моментных рядах при определении среднего уровня ряда *используется формула средней хронологической:*

$$\overline{y} = \frac{\frac{y_1}{2} + y_2 + ..y_{n-1}. + \frac{y_n}{2}}{n-1}.$$

- **Средний абсолютный прирост** является обобщающим показателем изменения явления во времени. Он **показывает, на сколько в среднем за единицу времени изменяется уровень ряда,** и рассчитывается как простая средняя арифметическая из показателей абсолютных цепных приростов:

$$\overline{\Delta} = \frac{\sum_{i=2}^{n} \Delta y_{ц_i}}{n-1} = \frac{\Delta y_{б_n}}{n-1}.$$

- **Средний коэффициент роста (средний относительный прирост)** *показывает, во сколько раз в среднем за единицу времени изменился уровень динамического ряда*. Эта

характеристика имеет важное значение при выявлении и описании основной долговременной тенденции развития, используется в качестве обобщенного показателя интенсивности развития явления за длительный период времени.

Средний коэффициент роста вычисляется по формуле простой средней геометрической:

$\overline{K} = \sqrt[n-1]{K_{ц_2} \cdot K_{ц_3} \cdot .... \cdot K_{ц_n}} = \sqrt[n-1]{K_{б_n}}$;

- **Средний коэффициент прироста** *характеризует среднюю относительную скорость изменения уровней в единицу времени*. Он определяется на основе среднего темпа роста:

$\overline{K}' = \overline{K} - 1$;.

Средний коэффициент прироста показывает, на какую долю единицы в среднем изменяется уровень ряда за единичный промежуток времени.

Средний темп прироста показывает, на сколько процентов в среднем за единицу времени изменяется уровень ряда. Он рассчитывается на основе среднего темпа роста:

$\overline{T}' = \overline{T} - 100\%$

- **Дисперсия уровней динамического ряда σ_y^2, среднее квадратическое отклонение σ_y и коэффициент вариации V_y** используются для оценки уровня вариации уровней.

Дисперсия уровней динамического ряда рассчитывается по формуле:

$$\sigma_y^2 = \frac{\sum_{i=1}^{n} (y_i - \overline{y})^2}{n}.$$

Среднее квадратическое отклонение как абсолютный показатель колеблемости уровней ряда равно: $\sigma_y = \sqrt{\sigma_y^2}$, а *коэффициент вариации как относительный показатель уровней ряда* -

$$V_y = \frac{\sigma_y}{\overline{y}}.$$

3.3.Модели разложения рядов динамики

Уровни любого ряда динамики формируются под совместным влиянием факторов, различных как по характеру, так и силе воздействия. В первую очередь необходимо выделить **факторы эволюционного характера**, оказывающие постоянное воздействие и определяющие **общее направление развития явления,** его долговременную эволюцию. Такие изменения динамического ряда называют **основной тенденцией развития или трендом.**

Вторую группу факторов составляют **факторы осциллятивного характера**, оказывающие периодическое воздействие. Они вызывают *циклические и сезонные колебания* уровней динамического ряда.

Циклическими (или периодическими) долговременными колебаниями называются *регулярные колебания, вызываемые постоянно действующими причинами*, например, циклы экономической конъюнктуры, циклы гарвардской школы. Схематично циклические колебания можно представить в виде синусоиды $y_i = \sin t_i$ (значение признака вначале возрастает, достигает определенного max, затем снижается, достигает своего min, вновь возрастает и т.д.).

Сезонные колебания – *колебания, периодически повторяющиеся в некоторое определенное время каждого года, в определенные дни каждого месяца или в*

определенные часы суток. Они могут вызываться природно-климатическими условиями, действием экономических, культурных и иных факторов.

Последней группой факторов, влияющих на ряд динамики, являются **факторы, вызывающие нерегулярные колебания уровней**. *Эти факторы подразделяются* в свою очередь на:

- **вызывающие спорадические изменения уровней** (война, экологические катастрофы, эпидемии и т.д.),
- **случайные, слабо воздействующие, второстепенные факторы** вызывающие случайные разнонаправленные изменения уровней.

Таким образом, уровни ряда динамики подвержены разным воздействиям, и теоретически ряд динамики может быть представлен как функция следующих компонент:

$y = f(T, R, S, E)$,

где Т – тренд;

R – циклические колебания;

S - сезонные колебания;

Е – случайные колебания.

Так как каждый фактор вызывает повышение или понижение уровней, то каждую компоненту и исходный динамический ряд можно представить в векторной форме:

$\vec{y} = f(\vec{T}, \vec{R}, \vec{S}, \vec{\varepsilon})$.

В зависимости от связи компонент между собой можно построить две модели ряда динамики:

- **аддитивная модель:** $\vec{y} = \vec{T} + \vec{R} + \vec{S} + \vec{\varepsilon}$ - характеризуется тем, что характер циклических и сезонных колебаний остается постоянными,

- **мультипликативная модель**: $\vec{y} = \vec{T} \cdot \vec{R} \cdot \vec{S} \cdot \vec{\varepsilon}$ - если характер циклических и сезонных колебаний остается постоянным только по отношению к тренду.

3.4. Методы выявления тренда

Первая задача, которая возникает при анализе рядов динамики, заключается в выявлении и описании основной тенденции развития изучаемого явления (тренда).

Трендом называется плавное и устойчивое изменение уровней явления во времени, свободное от случайных колебаний.

Изучение тренда включает в себя два этапа:

- *проверка ряда на наличие тренда;*
- *выравнивание ряда динамики и непосредственное выделение тренда.*

Проверка ряда на наличие тренда проводится разными методами, самыми простыми из которых является **метод средних** и **графический метод.** Суть ***метода средних*** заключается в следующем: изучаемый ряд динамики разбивается на несколько интервалов (чаще всего на два), для каждого из которых определяется средняя величина - $\bar{y}_1$ и $\bar{y}_2$. Выдвигается гипотеза о существенном различии рассчитанных средних. Если выдвинутая гипотеза принимается, то признается наличие тренда.

При графическом методе используется графическое представление ряда динамики и его визуальный анализ, позволяющий подтвердить наличие или отсутствие тренда.

Если основная тенденция выражена неявно, то ее наличие может быть подтверждено более сложными аналитическими методами, Например, *методом серий.*

Для непосредственного выявления тренда используют следующие методы механического и аналитического выравнивания (сглаживания), предполагающих наличие в исходном ряду динамики только одной компоненты – тренда.

К методам механического выравнивания относятся:

- **метод укрупнения интервалов;**
- **метод скользящей средней.**

3.4.1. Метод укрупнения интервалов является одним из наиболее простых методов непосредственного выявления основной тенденции. При использовании этого метода *ряд динамики, состоящий из мелких интервалов, заменяется рядом, состоящим из более крупных интервалов*. Так как на каждый уровень исходного ряда влияют факторы, вызывающие их разнонаправленное изменение, то это мешает видеть основную тенденцию. При укрупнении интервалов влияние факторов нивелируется, и основная тенденция проявляется более отчетливо. *Расчет среднего значения уровня по укрупненному интервалу осуществляется по формуле простой средней арифметической. Недостатком* этого метода является сокращение числа уровней ряда, а это не позволяет учитывать изменения внутри укрупненного интервала. К его *преимуществам* можно отнести сохранение природы явления.

3.4.2. Метод скользящей средней *предполагает замену исходного ряда динамики теоретическим, уровни которого рассчитываются по формуле скользящей средней*. Скользящая средняя относится к подвижным динамическим средним, вычисляемым по ряду при последовательном перемещении на один интервал. При этом, как и в предыдущем методе, происходит укрупнение интервалов. *Число уровней, по которым укрупняется интервал, называется диапазоном укрупнения, интервалом или периодом сглаживания* α. Период сглаживания может быть нечетным (α=3; 5; и т.д.) и четным (α=2; 4; и т.д.).

При нечетном периоде сглаживания полученное среднее значение уровня $\hat{y}_i$ закрепляется за серединой расчетного интервала. При α=3 формула имеет вид:

$$\hat{y}_i = \frac{y_{i-1} + y_i + y_{i+1}}{3}, \quad i = 2, n-1.$$

При четном периоде сглаживания возникает проблема центрирования, для решения которой необходимо осуществить сдвиг сглаженных уровней.

При использовании этого метода получают укороченный теоретический ряд. Число уровней сокращается при этом при α=3 на 2 уровня (крайних), при α=5 соответственно - на 4 и т.д., а это приводит к потере информации.

Рассмотренные методы дают возможность определить общую тенденцию развития явления, освобожденную от случайных и волнообразных колебаний, но не позволяют получить количественного описания тренда исследуемого ряда. Для получения обобщенной статистической модели тренда применяют метод аналитического выравнивания.

3.4.3. Методы аналитического выравнивания

Основная тенденция развития рассчитывается как временная функция $\hat{y}_i = f(t_i)$, где $\hat{y}_i$ - теоретические уровни (уровни динамического ряда, вычисленные по соответствующему аналитическому уравнению на момент времени t_i) т.е. развитие явления рассматривается в зависимости только от течения времени. Отклонения эмпирических уровней ряда y_i от уровней, соответствующих общей тенденции $\hat{y}_i$ объясняются действием случайных или циклических факторов. В результате получаем *трендовую модель* вида:

$$\hat{y}_i = f(t_i) + \varepsilon_i,$$

где ε_i - случайное и циклическое отклонение от тенденции.

Целью аналитического выравнивания динамического ряда является *определение аналитической или графической зависимости* $f(t_i)$. Функция $f(t_i)$ выбирается таким образом, чтобы она давала содержательное объяснение изучаемого процесса.

Подбор функции обычно осуществляется методом наименьших квадратов (МНК), в соответствии с которым наилучшим образом тренд описывает временная функция, обеспечивающая минимальную величину суммы квадратов отклонений эмпирических уровней ряда от соответствующих уровней теоретического ряда:

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \to \min,$$

где y_i - фактические уровни;

$\hat{y}_i$ - выровненные по временной функции уровни ряда.

Наиболее часто в анализе рядов динамики при выравнивании используются следующие зависимости:

- *линейная* $\hat{y} = a + b \cdot t$ *;*
- *параболическая* $\hat{y} = a + b \cdot t + c \cdot t^2$ *;*
- *показательная функция* $\hat{y} = a \cdot b^t$.

Линейная зависимость $\hat{y} = a + b \cdot t$ выбирается в тех случаях, когда в исходном ряду наблюдаются в среднем *постоянные абсолютные цепные приросты* $\Delta_{u_i} \approx const$.

Параметры уравнения a и b находятся по методу наименьших квадратов, в соответствии с которым получают систему нормальных уравнений:

$$\begin{cases} n \cdot a + b\sum t = \sum y, \\ a \cdot \sum t + b\sum t^2 = \sum yt \end{cases};$$

где y – фактические (эмпирические) уровни ряда;

t – хронологические показатели времени (порядковый номер периода или момента времени).

Для решения системы можно использовать любой известный метод, но предварительно необходимо решить проблему замены показателей времени, что позволит значительно упростить расчет параметров. *Хронологические показатели заменяются числовыми аналогами* таким образом, чтобы сумма новых показателей времени по ряду $\sum_{i=1}^{n} t_i = 0$:

- *при нечетном числе уровней* (например, - 7) за начало отсчета времени (t=0) принимается центральный интервал:

2009г.	2010г.	2011г.	2012г.	2013г.	2014г.	2015г.
-3	-2	-1	0	+1	+2	+3;

- *при четном числе уровней* (например, - 6) значения условных показателей времени будут выглядеть следующим образом:

2010г. .	2011г.	2012г.	2013г.	2014г.	2015г.
-3	-2	-1	+1	+2	+3.

Применение условных показателей времени позволяет привести систему нормальных уравнений к виду:

$$\begin{cases} n \cdot a = \sum y \\ b\sum t^2 = \sum yt \end{cases}.$$

Из первого уравнения $a = \dfrac{\sum y}{n}$.

Из второго уравнения $b = \frac{\sum yt}{\sum t^2}$.

Параметр a в линейной трендовой модели обычно интерпретации не имеет, но иногда его рассматривают как обобщенный начальный уровень ряда. ***Параметр*** *b* в трендовом уравнении ***называется коэффициентом регрессии***. Он определяет направление развития явления: при $b > 0$ – уровни ряда динамики равномерно возрастают, при $b < 0$ – равномерно снижаются. *Коэффициент регрессии показывает, насколько в среднем изменится уровень ряда при изменении времени на единицу*. Это означает, что параметр *b* можно рассматривать как средний абсолютный прирост с учетом тенденции к равномерному росту (росту в арифметической прогрессии).

Парабола второго порядка используется для описания рядов динамики, в которых меняется направление развития: со снижения показателей на их рост и наоборот.

Трендовое уравнение имеет вид: $\hat{y} = a + b \cdot t + c \cdot t^2$**.**

Параметр с называется коэффициентом регрессии и характеризует изменение интенсивности развития в единицу времени. При c>0 наблюдается ускоренное развитие, при c<0 – замедленное.

Показательная функция $\hat{y} = a \cdot b^t$ *применяется для описания динамических рядов со стабильными цепными темпами роста:* $T_{u_i} = const$. Такие динамические ряды отражают развитие в геометрической прогрессии. *Параметр b называется коэффициентом регрессии*, интерпретируется как средний темп роста изучаемого явления в единицу времени.

Для нахождения параметров модели функцию предварительно логарифмируют:

$\ln y = \ln a + t \cdot \ln b$.

На практике выбор формы кривой может быть основан на анализе графического изображения уровней ряда динамики (линейной диаграммы). При этом целесообразно использовать графическое изображение сглаженных уровней, в которых погашены случайные колебания.

Для оценки близости трендового уравнения эмпирическому ряду динамики применяется критерий Фишера (F).

Для динамических рядов, имеющих небольшую длину и подверженных значительным колебаниям, использовать метод аналитического выравнивания с помощью временной функции не рекомендуется, так как аппроксимация практически не адаптируется к изменяющимся условиям формирования уровней, при появлении новых данных нужно строить новые модели. Для сглаживания таких рядов динамики используются методы адаптивного моделирования и прогнозирования. В основе указанных методов лежит модель экспоненциального сглаживания. Временной ряд сглаживается с помощью взвешенной скользящей средней, в которой веса распределяются по экспоненциальному закону.

3.5. Анализ сезонных колебаний

Сезонными называют периодические колебания, возникающие под влиянием смены времени года и других причин природного или социально-культурного порядка. Они имеют устойчивый характер, повторяются регулярно с интервалом в один год.

Их роль велика в агропромышленном комплексе, строительстве, транспорте, здравоохранении, торговле и т.д. При этом сезонные колебания в одних отраслях экономики вызывает соответствующие колебания в других. Таким образом, проблема сезонности носит общий характер для экономики страны. Как правило, сезонность отрицательно влияет на результаты работы, поскольку приводит к неравномерному использованию рабочей силы, производственных мощностей, материальных ресурсов.

Поэтому хозяйствующие субъекты принимают меры для смягчения сезонности или стараются учитывать её влияние на свою деятельность.

Для выявления и измерения сезонных колебаний используются различные статистические методы, такие как, например, *построение модели сезонной волны и гармонический анализ.*

Метод построения «сезонной волны» заключается в расчете специальных показателей, которые называются **индексами сезонности** I_s^i**.** *Совокупность индексов сезонности отражают* **сезонную волну.**

Индексами сезонности *называется процентные отношения фактических (эмпирических) внутригрупповых уровней к теоретическим уровням, рассчитанным по трендовому уравнению, либо к средним уровням.*

Для выявления устойчивой сезонной волны, на которой не отражаются случайные условия одного года, *индексы сезонности рассчитываются за период не менее чем 3 года, распределенный по месяцам или кварталам.*

Расчет индексов сезонности выполняют двумя методами в зависимости от характера динамики:

- **если тренд неявно выражен**, то есть годовой уровень явления из года в год остается относительно неизменным, то *индексы сезонности рассчитываются методом постоянной средней.* Они рассчитываются по формуле:

$$\bar{I}_s^i = \frac{\bar{y}_s^i}{\bar{y}} \cdot 100\%$$

где i – номер одноименного периода (сезона);

$\bar{y}_s^i$ - средняя из фактических уровней одноименных периодов (месяцев или кварталов), вычисляется по формуле:

$$\bar{y}_s^i = \frac{\sum_{i=1}^{n} y_s^i}{n};$$

y_s^i - фактический уровень одноименного периода;

$\bar{y}$ - средний уровень ряда за исследуемый период.

Индексы сезонности рассчитываются в такой последовательности:

- рассчитываются средние уровни для каждого одноименного периода по данным за все годы наблюдения $\bar{y}_s^i$.

- определяется общая средняя $\bar{y}$ за весь период наблюдения.

- вычисляется индекс сезонности по приведенной выше формуле.

- **Если тренд явно выражен,** то для исчисления индексов сезонности *используется метод переменной средней,* в соответствии с которым их расчет проводится по формуле:

$$\bar{I}_s^i = \frac{\sum_{i=1}^{n} i_s^i}{n} \cdot 100\%;$$

где $i_s^i = \frac{y_i}{\hat{y}_i} \cdot 100\%$ - индивидуальный индекс сезонности одноименных периодов,

n – число лет наблюдения.

Применение переменной средней позволяет исключить влияние имеющейся тенденции на индексы сезонности.

Совокупность средних индексов сезонности одноименных периодов составляет модель сезонной волны.

Если при построении модели сезонной волны случайные колебания гасятся полностью, то сумма средних индексов сезонности одноименных периодов = 1200%, если

уровни брались за месяц, и 400%, если уровни были квартальными. Если это условие не выполняется, то проводится корректировка модели. Для этого рассчитывается поправочный коэффициент:

$$П = \frac{1200(400)}{\sum \bar{I}_s^i}.$$

На величину поправочного коэффициента корректируются все рассчитанные средние индексы сезонности $\bar{Y}_{s_{лор}}^{i} = \bar{Y}_s^i \cdot П$.

Кроме указанного способа для выявления сезонных колебаний можно использовать метод скользящих средних, гармонический анализ.

При применении гармонического анализа ряд динамики представляется как совокупность колебательных процессов, описываемых с помощью гармонического ряда Фурье.

Модель сезонных колебаний на основе гармоник Фурье имеет вид:

$$\hat{y}_i = a_0 + \sum_{i=1}^{m}(a_1 \cdot \cos kt_i + b \cdot \sin kt_i),$$

k – номер гармоники, определяющий степень адекватности модели ($k = 1 \div 4$),

a_0, a, b - параметры уравнения, определяются по МНК:

$$a_0 = \frac{\sum y}{n}; \qquad a_1 = \frac{2}{n} \cdot \sum y \cos kt; \quad b = \frac{2}{n} \cdot \sum y \sin kt.$$

При k=1 модель принимает вид: $\hat{y}_i = a_0 + a_1 \cdot \cos kt + b \cdot \sin kt$.

При изучении внутригодовой динамики n принимается равным 12, а показатели времени переводятся в условные, как части окружности. Для перевода можно использовать данные таблицы 3.2.

Таблица 3.2

Перевод хронологических показателей времени в условные

t_i	1	2	3	4	5	6	7	8	9	10	11	12
$t_i^{усл}$	0	$\frac{1}{6}\pi$	$\frac{1}{3}\pi$	$\frac{1}{2}\pi$	$\frac{2}{3}\pi$	$\frac{5}{6}\pi$	π	$\frac{7}{6}\pi$	$\frac{4}{3}\pi$	$\frac{3}{2}\pi$	$\frac{5}{3}\pi$	$\frac{11}{6}\pi$

3.6. Экстраполяция в рядах динамики и прогнозирование

Полученные при анализе динамических рядов характеристики используются для получения **статистических прогнозов**, под которыми понимаются **статистические оценки состояния явления в будущих периодах**.

Статистическое прогнозирование основано на предположении, что закономерность развития, основная тенденция, действующая в прошлом (внутри ряда динамики), сохранится и в будущем. Такое предположение называется **экстраполяцией**. Теоретической основой распространения тенденции на будущее является инерционность социально-экономических явлений.

Следует иметь в виду, что экстраполяция в рядах динамики носит приближенный характер. *Точность прогноза зависит от сроков прогнозирования:* чем они короче, тем надежнее результат экстраполяции, так как за короткий период времени не успевают

значительно измениться условия развития явления и характер его динамики. Обычно рекомендуется, чтобы срок прогноза не превышал 1/3 длительности базы расчета тренда.

С помощью метода экстраполяции получают два вида прогноза: *точечные и интервальные*. **Точечный прогноз** представляет собой *конкретное численное значение уровня в прогнозируемый период (момент) времени*. интервальный **прогноз –** *диапазон численных значений, предположительно содержащий прогнозируемое значение уровня.*

В зависимости от того, какие принципы и исходные данные положены в основу прогноза, выделяют следующие ***методы экстраполяции (прогнозирования)***:

- *на основе среднего абсолютного прироста* $\overline{\Delta}$,
- *на основе среднего коэффициента роста* $\overline{K}$,
- *на основе аналитического выравнивания ряда.*

Метод прогнозирования на основе среднего абсолютного прироста $\overline{\Delta}$ применяется в том случае, если уровни ряда динамики изменяются равномерно (линейно).

Прогнозируемое значение уровня определяется по формуле:

$\hat{y}_{n+l} = y_n + \overline{\Delta} \cdot l$;

где y_{n+l} - экстраполируемый уровень;

y_n - конечный уровень ряда динамики;

l - период упреждения прогноза (срок экстраполяции).

Прогнозирование по среднему коэффициенту роста $\overline{K}$ применяется, если общая тенденция характеризуется экспотенциальной кривой. В этом случае экстраполируемый уровень определяется по формуле:

$\hat{y}_{n+l} = y_n \cdot (\overline{K})^l$.

Прогнозирование на основе аналитического выравнивания является наиболее распространенным методом прогнозирования. Для получения прогноза используется аналитическое выражение тренда. Чтобы получить прогноз, достаточно в модели продолжить значение условного показателя времени t_i до t_{n+l}.

Интервальные прогнозы имеют значительные преимущества перед точечными – они учитывают вероятность свершения прогноза, соответствуют всем требованиям качества статистических оцен6ок. Для их получения необходимо построить доверительный интервал.

Величина доверительного интервала определяется в общем виде как $\hat{y}_{n+l} \pm t_\alpha \cdot \sigma_{y_i - \hat{y}_i}$,

где t_α - коэффициент доверия по распределению Стьюдента;

$\sigma_{y_i - \hat{y}_i}$ - средняя квадратическая ошибка тренда, рассчитываемая по формуле:

$$\sigma_{y_i - \hat{y}_i} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n - m}};$$

n – число уровней исходного ряда,

m – число параметров трендового уравнения.

Коэффициент доверия t_α выбирается по таблице распределения Стьюдента.

Таким образом, при использовании интервального прогноза прогнозируемый уровень ряда динамики находится в границах:

$\hat{y}_{n+l} - t_\alpha \cdot \sigma_{y_i - \hat{y}_i} \le \hat{y}_{n+l} \le \hat{y}_{n+l} + t_\alpha \cdot \sigma_{y_i - \hat{y}_i}$.

Контрольные вопросы и задания

1. Дайте определение рядов динамики. Какие задачи решаются с их помощью?

2. Какие виды рядов динамики Вам известны?
3. Опишите систему динамических характеристик ряда
4. Что показывает индивидуальный коэффициент роста? Какими способами он может быть рассчитан?
5. Что характеризует средний абсолютный прирост?
6. Перечислите основные компоненты уровня ряда динамики.
7. Какие модели разложения рядов динамики вам известны? В чем их отличие?
8. Дайте определение тренда.
9. Назовите методы выявления тренда.
10. Расскажите о сезонных колебаниях.
11. Для чего рассчитываются индексы сезонности и что они показывают?
12. Чем отличается точечный прогноз от интервального?
13. Какие методы прогнозирования Вам знакомы?

4.Экономические индексы

4.1. Индексы и их использование в экономико-статистических исследованиях

Индексы используются в качестве обобщающих характеристик изучаемых явлений. В переводе с латинского “index” означает указатель, показатель.

Индексы являются относительными величинами, характеризующими изменение уровней простых или сложных социально-экономических явлений во времени, пространстве или по сравнению с планом, то есть это, соответственно, относительные показатели динамики (индексы динамики), относительные показатели сравнения (территориальные индексы) и относительные показатели плана и выполнения плана.

От обычных относительных показателей индексы отличаются тем, что характеризуют изменение не только простых, но и сложных явлений. Сложные явления состоят из непосредственно несоизмеримых элементов, а простые – только из однородных элементов.

Показатель, для которого рассчитывается индекс, называется *индексируемой величиной.* Так, в индексе себестоимости индексируемой величиной является себестоимость, в индексе физического объема – объем выпуска в натуральном выражении.

С помощью индексов решаются следующие задачи:

- Оценка изменений сложных явлений и отдельных их частей (например, на сколько в текущем периоде изменился объем продаж по сравнению с предыдущим).
- Определение влияния отдельных факторов на общую динамику сложного явления (например, влияние изменения цен на объем продаж), для чего используется индексный анализ.

В практической деятельности используются разнообразные индексы, которые можно классифицировать по следующим основаниям:

- *содержание изучаемых объектов (характер);*
- *степень охвата элементов совокупности;*
- *методы расчета.*

По содержанию и характеру изучаемых показателей различают два вида индексов:

- *индексы количественных показателей* (объемных)

К ним относятся индексы физического объема произведенной продукции, физического объема потребления и т.д. *Индексируемой величиной* в таких индексах является объемный показатель, измеряемый в натуральных единицах.

- *индексы качественных показателей*

Эти индексы используются для измерения изменения показателя, рассчитываемого на единицу совокупности. Такие показатели называются качественными и характеризуют интенсивность изучаемого явления или процесса. *Индексируемой величиной* в индексах качественных показателей является уровень явления в расчете на единицу совокупности.

К индексам качественных показателей относятся индекс цен, себестоимости единицы продукции, трудоемкости, производительности труда и т.д.

По степени охвата элементов совокупности выделяют три формы индексов:

- *индивидуальные индексы* характеризуют изменение отдельных элементов, входящих в состав сложного явления. Это простая форма индексов (например, индивидуальный индекс цен отдельного вида товара).

- *сводные индексы* характеризуют изменение всего сложного явления, выражаемого сложным показателем. В таком явлении его элементы являются величинами несопоставимыми. Для решения проблемы несопоставимости индексируемых величин используются специальные показатели, называемые *соизмерителями индексируемых величин (статическими весами)*.

- *групповые индексы (субъиндексы)* рассчитываются для определенной части элементов совокупности. Например, индекс физического объема по отдельным отраслям или территориям.

По методам расчета классифицируются только общие индексы. Они делятся на *агрегатные и средние*.

В агрегатных индексах числитель и знаменатель (величина и база сравнения) представляют собой набор или агрегат разнородных элементов («aggregatus» - складываемый, суммируемый).

Средние индексы используются в тех случаях, когда данных для построения агрегатных индексов недостаточно. Они рассчитываются на основе индивидуальных индексов и делятся на средние арифметические и средние гармонические индексы.

Для удобства применения индексов используется определенная символика и специальная терминология.

Каждая ***индексируемая величина имеет свое обозначение:***

q – количество продукции одного вида в натуральном выражении,

p – цена единицы продукции,

z – себестоимость единицы продукции,

w – выработка продукции на 1-ого работника или в единицу времени,

t – трудоемкость единицы продукции.

Индивидуальные индексы обозначаются следующими символами:

i_q - индивидуальный индекс физического объема,

i_p - индивидуальный индекс цен,

i_z - индивидуальный индекс себестоимости и т.д.

Общие (сводные) индексы имеют обозначения:

I_q - общий индекс физического объема,

I_p - общий индекс цен,

I_z - общий индекс себестоимости и т.д.

При расчете индексов используются **два вида данных**:

- **данные базисного уровня** – уровня, с которым производится

сравнение; для их обозначения к символу соответствующего показателя добавляется «0».

- **данные текущего уровня** – уровня, который сравнивается

обозначаются добавлением «1» к символу соответствующего показателя.

В соответствии с принятыми обозначениями *индивидуальный индекс физического объема рассчитывается как* $i_q = \frac{q_1}{q_0}$, а *сводный индекс физического объема в агрегатной форме как* $I_q = \frac{\sum q_1 \cdot p_0}{\sum q_0 \cdot p_0}$ или $I_q = \frac{\sum q_1 \cdot p_1}{\sum q_0 \cdot p_1}$.

Индексы могут рассчитываться в виде коэффициентов или процентов.

4.2. Виды и формы индексов

Виды индексов определяются *видом индексируемой величины.* Различают **индексы физического показателя (объемные) индексы** и **индексы качественного показателя.**

Индексы физического показателя применяют для измерения изменения объемных показателей (объема продаж, численности работающих и т. п.).

Индексы качественного показателя используются для измерения изменений качественных показателей (цены, себестоимости единицы продукции и т. п.).

Формы индексов выделяются по степени охвата элементов совокупности. Элементами совокупности считаются её разнородные части. Например, предприятие выпускает несколько видов продукции. Каждый вид продукции – это отдельный элемент совокупности.

В практической деятельности применяют *три формы индексов: индивидуальные, общие (свободные) и групповые (субъиндексы)*.

Самая простая форма индексов – **индивидуальные**, так как они **являются обычными относительными величинами и представляют собой соотношение двух уровней индексируемой величины**.

Например, индивидуальный индекс физического объема $i_q = \frac{q_1}{q_0}$, где q_1, q_0 - количество произведенной одноименной продукции в текущем (отчетном) периоде и базисном. Этот индекс показывает, во сколько раз больше (меньше) в текущем периоде было произведено продукции по сравнению с базисным.

Индивидуальный индекс цен $i_p = \frac{p_1}{p_0}$, где p_1, p_0 - цена единицы продукции отчетного и базисного периодов, показывает, во сколько раз цена единицы продукции отчетного периода выше (ниже) соответствующей цены базисного периода.

Индивидуальный индекс стоимости $i_{pq} = \frac{p_1 \cdot q_1}{p_0 \cdot q_0}$, где p_1, q_1 - стоимость одноименной продукции отчетного периода, p_0, q_0 - стоимость одноименной продукции базисного периода, показывает, во сколько раз стоимость продукции отчетного периода больше (меньше) стоимости этой же продукции в базисном периоде.

Таким образом, *индивидуальный индекс показывает, во сколько раз индексируемая величина изменилась в отчетном (текущем) периоде по сравнению с базисным периодом*.

Сводные (общие) индексы характеризуют изменение всех элементов сложного явления.

Методика их расчета зависит от характера индексируемого показателя, качества исходных данных и целей исследования.

Сводные индексы рассчитываются двумя способами:

- как агрегатные;
- как средние из индивидуальных.

Средние индексы, в свою очередь, *рассчитываются как средние арифметические и средние гармонические*.

Из 2-х форм сводных индексов **основной является агрегатная форма**.

В числителе и знаменателе агрегатных индексов представлены несопоставимые элементы индексируемой величины. Для обеспечения сопоставимости при расчете используются специальные показатели–*соизмерители или веса индексов.*

Таким образом, агрегатный индекс строится как отношение сумм произведений индексируемой величины на показатель–соизмеритель, то есть по формуле:

$$I_x = \frac{\sum_{i=1}^{n} x_i^1 \cdot \omega_i}{\sum_{i=1}^{n} x_i^0 \cdot \omega_i},$$

где x_i^1, x_i^0 - текущее и базисное значение индексируемой величины i-ого элемента,

ω_j - показатель-соизмеритель явления i-ого элемента,

n – число элементов явления,

x_j

$x_i \cdot \omega_i$ - результативный показатель для j-ого элемента.

Показатель-соизмеритель может относится либо к текущему периоду, либо к базисному.

Если в качестве показателя-соизмерителя используется показатель текущего периода (отчетного), то формула для расчета агрегатного индекса выглядит следующим образом:

$$I_x = \frac{\sum_{i=1}^{n} x_i^1 \cdot \omega_1}{\sum_{i=1}^{n} x_i^0 \cdot \omega_1}.$$

Такая формула расчета была предложена в 1874 году Г.Пааше.

Если в качестве показателя-соизмерителя выступает показатель базисного периода, то формула для расчета принимает вид:

$$I_x = \frac{\sum_{i=1}^{n} x_i^1 \cdot \omega_0}{\sum_{i=1}^{n} x_i^0 \cdot \omega_0}.$$

Эту форму называют агрегатной формой индекса Э. Ласпейреса. Она была предложена в 1864 году.

При выборе формы агрегатного индекса необходимо решить три вопроса:

- *выбрать индексируемую величину.*
- *определить состав разнородных элементов, по которым рассчитывается индекс.*
- *выбрать показатель – соизмеритель индексируемой величины (её вес).*

Выбор показателя-соизмерителя индексируемой величины определяется её характером (содержанием).

При построении агрегатного индекса количественного (объемного) показателя-соизмерителем выступает качественный показатель; при построении агрегатного индекса качественного показателя-соизмерителем является количественный (объемный) показатель. Это означает, что агрегатные индексы качественных и количественных показателей рассчитываются по-разному.

4.3. Агрегатные индексы количественных показателей

К агрегатным индексам количественных показателей относятся **агрегатные индексы стоимости продукции** или товарооборота I_{pq} и **агрегатные индексы физического объема** I_q.

Агрегатный индекс стоимости продукции рассчитывается по формуле:

$$I_{pq} = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_0},$$

то есть как отношение стоимости продукции текущего периода к стоимости продукции базисного периода.

Агрегатный индекс стоимости I_{pq} *показывает, во сколько раз изменилась* (возросла или уменьшилась) *стоимость продукции или товарооборота отчетного периода по сравнению с базисным периодом.*

Разность (I_{pq} - 100) *показывает, на сколько % изменилась стоимость продукции отчетного периода по сравнению с базисным.*

Разность числителя и знаменателя, т.е. $\Delta_{pq} = \sum p_1 \cdot q_1 - \sum p_0 \cdot q_0$, *показывает абсолютный прирост результативного показателя,* т.е. на сколько денежных единиц изменилась стоимость продукции текущего периода по сравнению с базисным.

Численное значение индекса стоимости определяется двумя факторами:

- изменением количества (объема) товара;
- изменением цен.

Для того, чтобы оценить изменение стоимости только за счет одного фактора, необходимо устранить влияние другого фактора. Это можно сделать, если зафиксировать в формуле данный фактор неизменным, т.е. на уровне одного и того же периода.

Так, *если физический объем продаж оценивать по одним и тем же ценам*, то можно получить *индекс, отражающий изменение только одного фактора – количества товара.*

В этом случае **индексируемой величиной является количество товара** (физический объем), а сам **индекс называется агрегатным индексом физического объема** I_q**.**

При его расчете в качестве статистических весов можно использовать цены базисного или отчетного периодов. Если выбираются цены базисного периода, то получают **агрегатный индекс физического объема в сопоставимых (базисных) ценах – индекс физического объема Ласпейреса:**

$$I_q = \frac{\sum q_1 \cdot p_0}{\sum q_0 \cdot p_0}.$$

Индекс Ласпейреса показывает, *во сколько раз изменился физический объем продукции (товара) в текущем периоде по сравнению с базисным.*

Числитель формулы $\sum q_1 \cdot p_0$ означает расчетную стоимость объема продаж текущего периода в неизменных базисных ценах; знаменатель $\sum q_0 \cdot p_0$ - фактическую стоимость продаж базисного периода.

Разность ($I_q - 1$) *показывает, на сколько % изменилась стоимость объема продаж за счет изменения его физического объема в текущем периоде по сравнению с базисным.*

Разница между числителем и знаменателем $\Delta_{pq_{(ü)}} = \sum q_1 p_0 - \sum q_0 \cdot p_0$ *есть абсолютное изменение стоимости продаж за счет изменения её физических (*натуральных) *объемов.*

При исчислении агрегатного индекса физического объема в качестве соизмерителя индексируемой величины можно использовать цены текущего периода. В этом случае формула принимает вид:

$$I_q = \frac{\sum q_1 \cdot p_1}{\sum q_0 \cdot p_1},$$

где $\sum q_1 \cdot p_1$ - стоимость объема продукции текущего периода в ценах текущего периода;

$\sum q_0 \cdot p_1$ - расчетная стоимость объема продаж базисного периода в ценах текущего периода.

Индекс, рассчитанный по приведенной формуле называется **агрегатным индексом физического объема Пааше,** и *показывает, во сколько раз изменился физический объем в текущем периоде по сравнению с базисным, если в базисном периоде цены были бы равны текущим.*

Индекс физического объема, рассчитанный по формулам Пааше и Ласпейреса, имеет разное значение. Численной значение индекса, рассчитанное по формуле Пааше всегда выше, чем рассчитанное по формуле Ласпейреса. Это связано с тем, что в формуле Ласпейреса при использовании в качестве соизмерителя неизменных цен базисного периода полностью устраняется влияние изменения цен на динамику объема продукции.

В формуле Пааше, т.е. при использовании в качестве соизмерителя нефиксированных цен текущего периода, устранить влияние изменения цен на динамику объема продукции не удается. В связи с этим *использовать формулу Пааше для расчета агрегатного индекса физического объема продукции не рекомендуется***.**

Помимо цен в качестве показателя-соизмерителя при построении индекса физического объема можно использовать трудоемкость и себестоимость единицы продукции.

Индекс, построенный с применением в качестве соизмерителя себестоимости, имеет следующий вид:

$$I_q = \frac{\sum q_1 \cdot z_0}{\sum q_0 \cdot z_0},$$

где $\sum q_1 z_0$ - расчетные издержки производства текущего периода по себестоимости базисного;

$\sum q_0 \cdot z_0$ - издержки производства базисного периода.

Индекс характеризует изменение издержек производства в результате изменения физического объема, а разность между числителем и знаменателем $\Delta q \cdot z_{(q)} = \sum q_1 z_0 - \sum q_0 \cdot z_0$ - абсолютное изменение затрат (издержек производства) за счет изменения физического объема производства.

Аналогичным образом строится индекс физического объема с применением в качестве показателя-соизмерителя трудоемкости единицы продукции.

4.4. Агрегатные индексы качественных показателей

К агрегатным индексам качественных показателей относятся:

I_p - *агрегатный индекс цен;*

I_z - *агрегатный индекс себестоимости;*

I_t - *агрегатный индекс трудоемкости;*

I_w - *агрегатный индекс производительности труда (выработки).*

При построении перечисленных индексов *показателем-соизмерителем является связанный с индексируемой величиной количественный показатель***.**

Агрегатный индекс цен I_p характеризует изменение результирующего показателя (общей стоимости) за счет изменения цен в текущем периоде по сравнению с базисным. При его построении важно устранить влияние изменения количества товара, т.е. физического объема. Для этого в качестве соизмерителя индексируемой величины – цены используется неизменный физический объем либо отчетного, либо базисного периода. Таким образом, ***агрегатный индекс цен можно рассчитать по формуле Пааше и по формуле Ласрейреса:***

$I_p = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1}$ - агрегатный индекс цен Пааше;

$I_p = \frac{\sum p_1 \cdot q_0}{\sum p_0 \cdot q_0}$ - агрегатный индекс цен Ласпейреса.

Рассмотренные индексы позволяют определить относительное изменение цен, но оно *не будет одинаковым*, так как имеет различное экономическое содержание.

Индекс Пааше показывает, во сколько раз изменился уровень цен на продукцию текущего периода, а разность между числителем и знаменателем $\Delta p \cdot q_{(p)} = \sum q_1 \cdot p_1 - \sum q_1 \cdot p_0$ - на сколько изменилась стоимость продукции в текущем периоде за счет изменения цен.

Индекс Ласпейреса *показывает, во сколько раз подорожала бы или подешевела бы продукция базисного периода из-за изменения цен на нее в отчетном периоде***.**

Согласно практике *индекс цен, исчисленный по формуле Пааше, всегда больше индекса, исчисленного по формуле Ласпейреса*. Применение того или иного индекса зависит от цели исследования.

Если целью анализа является определение экономического эффекта (прибыль или убыток) от изменения цен в отчетном периоде по сравнению с базисными, то ***используется индекс Пааше.***

Если целью анализа является прогнозирование объема продаж в связи с возможным изменением цен в предстоящем периоде, то ***используется индекс Ласпейреса***, так как он позволяет определить стоимость продаж одного и того же физического объема базисного периода по новым ценам.

Достаточно часто в экономическом анализе используется ещё один вид общего индекса цен - **индекс Лоу (общий индекс на средних весах)**. В его формуле в качестве соизмерителя используется средний физический объем продаж $\overline{q}$, рассчитанный как простая средняя арифметическая $\overline{q}_i = \frac{q_{0_i} + q_{1_i}}{2}$:

$$I_{\overline{p}} = \frac{\sum p_1 \cdot \overline{q}}{\sum p_0 \cdot \overline{q}}.$$

Индекс Лоу используется в расчетах, связанных с закупкой или реализацией товаров в течение длительного периода (по долгосрочным контрактам). Он показывает, во сколько раз в среднем изменился бы объем продаж за счет изменения цен.

Достоинством индекса Лоу является то, что при его использовании устраняются недостатки индекса Пааше и Ласпейреса.

Кроме перечисленных индексов можно использовать *«идеальный индекс» Фишера.*

Идеальный индекс Фишера рассчитывается как средняя геометрическая из индексов цен Ласпейреса и Пааше:

$$I_{\overline{p}} = \sqrt{\frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1} \cdot \frac{\sum p_1 \cdot q_0}{\sum p_0 \cdot q_0}}.$$

Идеальный индекс Фишера используется при исчислении индексов цен на длительный период времени для сглаживания тенденции в структуре и составе объема продукции, в которых происходят значительные изменения. Его недостаткам является то, что он не имеет экономической интерпретации.

Аналогично строятся индексы других качественных показателей. Например, ***агрегатный индекс себестоимости продукции*** рассчитывается следующим образом:

$$I_z = \frac{\sum z_1 \cdot q_1}{\sum z_0 \cdot q_1},$$

где $\sum z_1 \cdot q_1$ - затраты на производство продукции отчетного периода,

$\sum z_0 \cdot q_1$ - расчетные затраты на производство продукции текущего периода по себестоимости базисного.

Агрегатный индекс себестоимости продукции *показывает, во сколько раз изменился уровень стоимости на продукцию отчетного периода,* а *разница между числителем и знаменателем* $\Delta zq_z = \sum z_1 q_1 - \sum z_0 q_1$ *показывает увеличение или снижение затрат за счет изменения себестоимости единицы продукции.*

Таким образом, индексы качественных и количественных показателей показывают, как меняется результирующий показатель при изменении либо физического объема проданных (реализованных) товаров, либо цен (себестоимости) единицы товара.

Следует учитывать, что изменение цен и изменение физических объемов действуют на результирующий показатель одновременно. При этом направление действия указанных факторов и их интенсивность могут быть различными. Для оценки совместного их влияния на изменение результативного показателя используются системы взаимосвязанных индексов, называемые индексными системами.

4.5. Индексные системы и факторный анализ

В индексных системах отражается взаимосвязь экономических показателей: *если экономические показатели связаны между собой определенным образом, то таким же образом связаны между собой и характеризующие их индексы, т.е. если* $z = x \cdot y$, *то* $y_z = I_x \cdot I_y$.

Индексные системы дают возможность использовать индексный метод для изучения взаимосвязи показателей и проведения факторного анализа с целью определения влияния каждого фактора на результативный показатель.

Построение индексной системы рассмотрим на примере индекса стоимости, индекса цен и индекса физического объема:

I_p и I_q являются факторными по отношению к индексу стоимости продукции.

Индекс стоимости рассчитывается по формуле:

$$I_{pq} = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_0};$$

индекс цен рассчитаем по формуле Пааше:

$$I_p = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1};$$

а индекс физического объема – по формуле Ласпейреса:

$$I_q = \frac{\sum q_1 \cdot p_0}{\sum q_0 \cdot p_0}.$$

Перемножение индекса цен и индекса физического объема дает следующий результат:

$$I_p \cdot I_q = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1} \cdot \frac{\sum p_1 \cdot q_0}{\sum p_0 \cdot q_0} = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_0} = I_p \cdot I_q.$$

Таким образом: $I_p \cdot I_q = I_{pq}$.

Аналогична взаимосвязь других результативных признаков с факторными. Например, индекс объема продукции с индексом численности работающих и индексом производительности труда (выработки) связан таким же образом, как объем производства Q связан с выработкой одного работающего w и численностью работающих *r*.

Если Q = w·r то $I_{wr} = I_w \cdot I_r$.

$$I_{wr} = \frac{\sum w_1 \cdot r_1}{\sum w_0 \cdot r_0} = \frac{\sum w_1 \cdot r_1}{\sum w_0 \cdot r_1} \cdot \frac{\sum r_1 \cdot w_0}{\sum r_0 \cdot w_0} = I_w \cdot I_r ;$$

где I_w - индекс производительности труда, рассчитываемый по формуле Ласпейреса;

I_r - индекс численности работающих, рассчитываемый по формуле Пааше.

Индексные системы используются для определения влияния отдельных факторов на формирование уровня результативного показателя, позволяют по 2-м известным значениям индексов определить значение неизвестного.

Рассмотренные индексные системы относятся к двухфакторным, но результативный признак можно разложить и на большее число факторов и, соответственно, получить многофакторные индексные системы, которые могут разложить изменение результативного показателя на элементы, вызванные влиянием отдельных факторов.

Индексные системы позволяют разложить и абсолютное изменение результативного показателя на составляющие, вызванные влиянием разных факторов, т.е. разложить абсолютное изменение по факторам. Это можно сделать, если результативный показатель представляет собой произведение количественного фактора на качественный.

Абсолютное изменение результативного показателя определяется как разница между числителем и знаменателем формулы расчета индекса стоимости $\Delta pq = \sum p_1 \cdot q_1 - \sum p_0 \cdot q_0$.

Абсолютное изменение результативного показателя за счет изменения цен рассчитывается как

$$\Delta pq_p = \sum p_1 \cdot q_1 - \sum p_0 \cdot q_1 .$$

Абсолютное изменение результативного показателя за счет изменения физического объема составит

$$\Delta pq_q = \sum p_0 \cdot q_1 - \sum p_0 \cdot q_0 .$$

Сложение абсолютного изменения результативного показателя за счет изменения цен и абсолютного изменения результативного показателя за счет изменения физического объема дает следующий результат:

$$\Delta pq_p + \Delta pq_p = \sum p_1 \cdot q_1 - \sum p_0 \cdot q_1 + \sum q_1 \cdot p_0 - \sum q_0 \cdot p_0 = \sum p_1 \cdot q_1 - \sum p_0 \cdot q_0 = \Delta pq .$$

Следовательно:

$$\Delta pq = \Delta pq_p + \Delta pq_q .$$

4.6. Средние индексы

Средние индексы – это вторая форма исчисления общих индексов, применяемая в случаях, когда невозможно вести учет показателей в натуральных измерителях (коммерческие организации, торговля, где в основном ведется стоимостной учет), или в плановых расчетах.

Во всех случаях, когда информация о физических объемах в натуральном исчислении отсутствует, для определения изменения показателей используется средняя форма индексов.

В практических расчетах используются два вида средних индексов:

- **средний индекс качественного показателя,**
- **средний индекс физического объема.**

Каждый из средних индексов может быть рассчитан по формулам средней арифметической взвешенной или средней гармонической взвешенной.

Средний индекс физического объема *используется в тех случаях, когда отсутствует информация об объемах выпуска в натуральных измерителях.*

Средняя арифметическая форма индекса физического объема применяется, когда имеется информация о стоимости реализованной продукции в базисном периоде, и об индивидуальных индексах физического объема i_q.

Формулу среднеарифметического индекса физического объема можно получить на основе агрегатного индекса физического объема Ласпейреса $I_q = \frac{\sum q_1 \cdot p_0}{\sum q_0 \cdot p_0}$, заменив физические объемы текущего периода на их выражение через индивидуальный индекс физического объема i_q.

$i_q = \frac{q_1}{q_0}$, следовательно $q_1 = i_q \cdot q_0$; тогда

$$\bar{I}_q = \frac{\sum i_q \cdot q_0 \cdot p_0}{\sum q_0 \cdot p_0},$$

где i_q - усредняемая величина, а $p_0 \cdot q_0$ - статистический вес.

Полученная формула является формулой среднеарифметического индекса.

Среднеарифметический индекс показывает*, во сколько раз в среднем изменится физический объем в планируемом (предстоящем) периоде*. Таким образом, среднеарифметический индекс физического объема есть средний из индивидуальных индексов физического объема.

Разница между числителем и знаменателем характеризует изменение стоимости продукции в планируемом периоде.

$$\Delta pq_q = \sum i_q \cdot p_0 \cdot q_0 - \sum q_0 \cdot p_0 .$$

Средний индекс физического объема можно рассчитать по формуле средней гармонической взвешенной. Она применяется в случае, если исходная информация представлена индивидуальными индексами физического объема i_q (или их легко рассчитать), или фактической стоимостью продукции текущего периода $q_1 \cdot p_1$.

Формула ***среднего геометрического индекса физического объема*** *может быть получена из агрегатной формы общего индекса физического объема Пааше*:

$I_q = \frac{\sum q_1 \cdot p_1}{\sum q_0 \cdot p_1}$, которая показывает, во сколько раз изменяется стоимость продукции за счет изменения физических объемов.

В указанной формуле физические объемы базисного периода q_0 заменяются их выражением через индивидуальный индекс физического объема:

если $i_q = \frac{q_1}{q_0}$ то $q_0 = \frac{q_1}{i_q}$.

Исходя из такой замены формула для расчета среднего гармонического индекса физического объема будет иметь следующий вид:

$$\bar{I}_q = \frac{\sum q_1 \cdot p_1}{\sum \frac{1}{i_q} \cdot q_1 \cdot p_1},$$

где i_q - усредняемая величина;

$p_1 \cdot q_1$ - статистический вес.

Разница между числителем и знаменателем дает показатель среднего изменения стоимости в текущем периоде за счет изменения физического объема:

$$\Delta pq_p = \sum q_1 \cdot p_1 - \sum \frac{1}{i_q} \cdot q_1 \cdot p_1 .$$

Содержание и расчет **среднего индекса качественного показателя** рассмотрим на примере цен.

Общий индекс цен в средней арифметической форме используется в плановых расчетах (при прогнозировании). *Информация для расчета должна быть представлена в виде индивидуальных индексов цен или планируемых изменений цен и стоимости продукции базисного периода (отчетного).*

Формулу для расчета общего индекса цен в средней арифметической форме легко получить преобразованием формулы агрегатного индекса цен Ласпейреса $I_p = \frac{\sum p_1 \cdot q_0}{\sum p_0 \cdot q_0}$, выразив цены отчетного периода p_1 через индивидуальные индексы цен i_p и цены базисного периода p_0: $i_p = \frac{p_1}{p_0}$, следовательно $p_1 = i_p \cdot p_0$. Тогда **формула для расчета среднего арифметического индекса цен имеет вид**:

$$\bar{I}_p = \frac{\sum i_p \cdot p_0 \cdot q_0}{\sum p_0 \cdot q_0},$$

где i_p - усредняемая величина,

$p_0 \cdot q_0$ - статистический вес усредняемой величины.

Средний арифметический индекс цен показывает, во сколько раз в среднем изменится стоимость продукции предстоящего периода за счет изменения цен. Разность числителя и знаменателя $\Delta pq_p = \sum i_p q_0 \cdot p_0 - \sum q_0 \cdot p_0$ определяет общее изменение стоимости продукции предстоящего (планового) периода за счет изменения цен.

Средняя гармоническая форма общего индекса цен *используется, когда информация представлена в виде индивидуальных индексов цен или их изменений и стоимости продукции текущего периода* $p_1 \cdot q_1$.

Формулу для расчета среднего гармонического индекса цен можно получить преобразованием агрегатного индекса цен Пааше, заменив цены базисного периода p_0 *его выражением через индивидуальный индекс цен* i_p. Так как $i_p = \frac{p_1}{p_0}$, то $p_0 = \frac{p_1}{i_p}$; следовательно

$$\bar{I}_p = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1} = \frac{\sum p_1 \cdot q_1}{\sum \frac{1}{i_p} \cdot p_1 \cdot q_1} .$$

Средний гармонический индекс цен $\bar{I}_p$ показывает, во сколько раз в среднем изменилась стоимость продукции текущего периода за счет изменения цен:

Разность между числителем и знаменателем формулы показывает абсолютное изменение стоимости продукции за счет изменения цен $\Delta p = \sum q_1 \cdot p_1 - \sum \frac{1}{i_p} \cdot q_1 \cdot p_1$.

4.7.Индексный анализ динамики среднего уровня

Индексный метод в некоторых случаях используется и для изучения динамики соизмеримых между собой величин. При этом он позволяет *выявлять*

факторы, влияющие на динамику среднего уровня качественного показателя для продукции одного итого же назначения, отличающейся только ценами за единицу физического объема. Для такой продукции в случае объединения ее физических объемов можно вводить *средние уровни соответствующих качественных признаков* (например, средняя цена цемента разных марок, выпускаемого предприятием).

Средний уровень качественного показателя в отчетном периоде может отличаться от среднего уровня в базисном периоде по двум причинам: *изменений*

- *значения осредняемого показателя;*
- *изменений структуры явления.*

Если анализируется однородная (соизмеримая) продукция одного и того же назначения, то для характеристики изменения среднего уровня качественного показателя рассчитываются следующие индексы:

- ***индекс среднего уровня качественного показателя переменного состава;***
- ***индекс среднего уровня качественного показателя постоянного (фиксированного) состава;***
- ***индекс структурного сдвига.***

Индекс переменного состава характеризует общее изменение среднего уровня качественного показателя (цены) вследствие изменения качественного и количественного (структурного) факторов:

$$I_{p_{перем.}} = \frac{\overline{p}_1}{\overline{p}_0} = \frac{\sum p_{11}q_1 \Big/ \sum q_1}{\sum p_0 q_0 \Big/ \sum q_0}.$$

Он представляет собой отношение двух взвешенных средних с изменяющимися (переменными) весами, и показывает изменение индексируемой средней величины.

Индекс постоянного состава характеризует динамику среднего уровня качественного показателя в связи с изменением качественного показателя при неизменности физических объемов. При этом количественный (неиндексируемый) признак фиксируется на отчетном уровне:

$$I_{\substack{p \\ (пост.сост.)}} = \frac{\sum p_1 q_1 \Big/ \sum q_1}{\sum\limits_i p_0 q_1 \Big/ \sum q_1}.$$

После сокращения на $\sum q_1$ индекс принимает вид $I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}$ (агрегатный индекс цен по формуле Пааше).

Индекс постоянного состава показывает изменение средней величины качественного показателя в отчетном периоде по сравнению с базисным только за счет изменений самой индексируемой величины, то есть когда когда влияние структурного фактора устранено.

Индекс структурного сдвига характеризует влияние изменения структуры (изменение доли отдельных единиц совокупности, из которой формируются средние, в общей их численности). Этот индекс предполагает учет качественного показателя (неиндексируемого в нем) на базисном уровне:

$$I_{\text{стр сдв}} = \frac{\sum p_0 q_{1i} \Big/ \sum_i q_1}{\sum_i p_0 q_0 \Big/ \sum_{i1} q_0}$$

Между этими индексами существует взаимосвязь:

$$I_{\bar{p}(\text{перем.сост.})} = I_{\bar{p}(\text{пост.сост.})} \cdot I_{\text{стр.сдв}}$$

Контрольные вопросы и задания

1. Какие задачи решаются с помощью индексов?
2. Дайте определение индекса.
3. Какие виды индексов вы знаете?
4. Что такое индексируемая величина?
5. Чем отличается агрегатный индекс от среднего?
6. Как выбираются статистические веса в индексах качественных показателей Ласпереса и Пааше?
7. Что из себя представляет индексная система?
8. Как проводится индексный анализ?
9. Когда возникает необходимость преобразования агрегатного индекса цен в средний (средний арифметический или средний гармонический)? Как осуществляются такие преобразования?
10. Как выглядит агрегатный индекс трудоемкости по формуле Ласпейреса и Пааше? Что он показывает?
11. Какой индекс называется индексом переменного состава? Что он показывает?
12. Что характеризует индекс постоянного состава, Как он рассчитывается?
13. Что называется индексом структурных сдвигов, Как он вычисляется?
14. Как связаны между собой индексы переменного, постоянного составов и структурных сдвигов?

5. Выборочное исследование

5.1. Постановка задачи выборочного исследования

Выборочным называется статистическое исследование, при котором обобщающие показатели изучаемой совокупности устанавливаются по некоторой ее части, сформированной на основе положений случайного отбора.

В основе выборочного исследования лежит *несплошное наблюдение*, при котором обследуются не все единицы совокупности, а лишь определенная их часть.

Выборочное исследование широко применяется на практике, поскольку обладает ***существенными преимуществами*** по сравнению с другими методами получения статистических данных. К ним относятся:

- дотаточно высокая точность результатов обследования благодаря использованию более квалифицированных кадров, что приводит к сокращению ошибок регистрации;

- экономия времени и средств в результате сокращения объема работы, большая оперативность в получении данных о результатах обследования;

- возможность исследования очень больших статистических совокупностей;

- выборочный метод является единственно возможным, если сбор информации связан с разрушением или потерей единиц наблюдения, например, при органалитическом контроле качества продукции;

- возможность исследования полностью недоступных совокупностей.

При выборочном исследовании изучается сравнительно небольшая часть статистической совокупности (5-10%, реже 20-25% объема ее единиц).

Проведение выборочного исследования является достаточно сложным процессом, выполнение которого включает в себя:

обоснование целесообразности применения выборочного метода в данном исследовании;

- *составление* программы исследования;
- *установление* объема выборки – n;
- *основание* способа формирования выборки;
- *отбор единиц* из Генеральной совокупности (формирование выборки);
- *измерение* изучаемых признаков у отдельных единиц;
- *обработка* полученной информации и расчет характеристик выборки;
- *определение* ошибки выборки;
- *распространение* выборочных характеристик на Генеральную совокупность.

Для постановки задачи выборочного исследования *необходимо ввести следующие понятия***:**

- **Генеральная совокупность** – совокупность, содержащая все исследуемые элементы, она может быть конечной (N) или бесконечной (∞).

- **Выборочная совокупность (выборка)** – часть единиц генеральной совокупности, отобранная для изучения (n).

Качество результатов выборочного исследования зависит от того, насколько состав выборки представляет генеральную совокупность, иначе говоря, *насколько выборка репрезентативна*.

Под репрезентативностью выборки понимается *соответствие ее свойств и структуры свойствам и структуре генеральной совокупности*.

Репрезентативность выборки может быть обеспечена только при объективности отбора данных, гарантируемой принципами случайности отбора единиц.

Принцип случайности предполагает, что на включение или исключение статистической единицы из выборки не может повлиять никакой другой фактор, кроме случая.

Этот принцип лежит в основе методов случайного отбора, с помощью которых формируется выборка.

Использование методов случайного отбора при формировании выборки позволяет в дальнейшем при обработке использовать аппарат теории вероятности.

Чаще всего с помощью выборочного исследования определяются следующие характеристики генеральной совокупности.

- **среднее значение признака в совокупности** $\overline{X}$;
- **доля альтернативного признака в совокупности** $\overline{d}$. *Альтернативным* считается признак, принимающий два значения. Если одно из них изменяется как заданное, то *доля альтернативного признака будет характеризовать удельный вес статистических единиц, обладающих заданным значением альтернативного признака*, например, доля брака в изготовленной партии продукции;
- **дисперсия признака в совокупности** σ^2, как показатель вариации.

В общем виде ***задача выборочного исследования*** формулируется следующим образом:

Пусть имеется *некоторая генеральная совокупность* известного объема (N единиц), обладающая ***неизвестными статистическими характеристиками***:

$\bar{d} = \frac{P}{N}$ - ***генеральная доля*** (удельный вес статистических единиц генеральной совокупности, обладающих данным значением признака), где Р- число единиц генеральной совокупности, обладающих данным значением признака.

$\bar{X}$ - ***генеральная средняя*** (среднее арифметическое значение признака в генеральной совокупности).

σ^2 - ***генеральная дисперсия*** (дисперсия исследуемого признака в генеральной совокупности).

σ - ***генеральное среднеквадратического отклонения*** (среднее квадратическое отклонение исследуемого признака в генеральной совокупности).

Для их определения *сформирована выборочная совокупность* объемом n статистических единиц ($n << N$), обладающая аналогичными характеристиками:

ω - ***выборочная доля*** (удельный вес статистических единиц, обладающих данным значением признака в выборочной совокупности).

$\tilde{X}$ - ***выборочная средняя*** (среднее арифметическое значение признака в выборочной совокупности).

S^2- ***выборочная дисперсия*** (дисперсия исследуемого признака в выборочной совокупности).

S – ***выборочное среднее квадратическое отклонение*** (среднее квадратическое отклонение изучаемого признака в выборке).

Необходимо на основе известных характеристик выборки *получить статистические оценки характеристик генеральной совокупности*.

5.2.Статистические оценки параметров (характеристик) генеральной совокупности

Статистической оценкой или статистикой характеристики (параметра) генеральной совокупности *называют приближенное значение искомой характеристики (параметра), полученное по данным выборки*.

В статистике используются *два вида оценок - точечные и интервальные.*

Точечной статистической оценкой параметра генеральной совокупности называется конкретное числовое значение искомой характеристики.

Интервальная оценка представляет собой числовые интервалы, *предположительно* содержащие значение параметра генеральной совокупности.

Качество статистических оценок **определяется следующими их свойствами:**

- **Состоятельность**

Оценка считается состоятельной, если при неограниченном увеличении объема выборки [$n \longrightarrow \infty$ (N)], ее ошибка стремится к 0:

$\lim_{n\to\infty}(\tilde{\alpha} - \alpha) = 0$, т. к. при $n \uparrow$ $\lim_{n\to\infty}\tilde{\alpha} = \alpha$;

где α - значение характеристики генеральной совокупности;

$\tilde{\alpha}$ - значение характеристики выборки;

$\tilde{\alpha} - \alpha$ - ошибка выборки.

- **Несмещенность**

Оценка считается несмещенной, если при данном объеме выборки n математическое ожидание ошибки равно 0.Для несмещенной оценки ее математическое ожидание точно равно математическому ожиданию характеристики выборки:

$M[\tilde{\alpha} - \alpha] = 0$ *или* $M[\tilde{\alpha}] = M[\alpha]$.

Несмещенная оценка не всегда дает хорошее приближение оцениваемого параметра, так как возможные значения получаемой оценки могут быть сильно рассеяны вокруг

своего среднего значения. Поэтому оценка должна соответствовать еще одному требованию – эффективности.

- **Эффективность**

Оценка считается эффективной, если ее ошибка, называемая ошибкой выборки, является величиной минимальной. В математической статистике доказывается, что ошибка выборки определяется как:

$\mu(\tilde{\alpha}) = \sqrt{M^2[\tilde{\alpha} - \alpha] + S^2}$;

где $M^2[\tilde{\alpha} - \alpha]$- квадрат математического ожидания ошибки выборки;

S^2 - выборочная дисперсия.***Оценка эффективна***, если выполняется условие: $\mu(\tilde{\alpha}) \to \min$.

Для точечных оценок справедливы следующие утверждения:

- *точечной оценкой генеральной доли является выборочная доля,* то есть $d \sim \omega$;

- *точечной оценкой генеральной средней является выборочная средняя*, то есть $\bar{x} \sim \tilde{x}$.

Таким образом, заранее известно, что оценки для указанных параметров являются состоятельными и несмещенными.

Для остальных параметров генеральной совокупности это утверждение не является справедливым, то есть $\sigma^2 \neq S^2$, а $\sigma \neq S$.

В математической статистике доказывается, что точечной оценкой генеральной дисперсии является выборочная дисперсия, откорректированная на отношение $\frac{n}{n-1}$, то есть $\sigma^2 = S^2 \times \frac{n}{n-1}$; при увеличении n $\frac{n}{n-1} \to 1$, поэтому в выборках, объемом больше 30 единиц наблюдения, указанным отношением можно пренебречь.

Аналогично, точечной оценкой генерального среднеквадратического отклонения является выборочное среднеквадратическое отклонение, откорректированное на $\frac{n}{n-1}$, то есть $\sigma = S \times \frac{n}{n-1}$.

В этом случае точечные оценки генеральной дисперсии и генерального среднеквадратического отклонения являются состоятельными и несмещенными.

Основным недостатком точечных оценок является то, что они не учитывают ошибки выборки, то есть не являются эффективными. Поэтому более предпочтительными являются интервальные оценки параметров генеральной совокупности, в которых эти ошибки учитываются. Интервальные оценки соответствуют всем трем требованиям качества статистической оценки.

В математической статистике доказывается, что:

- **Интервальной оценкой генеральной доли** является ее выборочная доля с учетом ошибки выборочной доли, то есть $\bar{d} \approx \omega \pm \mu_\omega$, где μ_ω - ошибка выборочной доли.

- **Интервальной оценкой генеральной средней** является выборочная средняя с учетом ошибки выборочной средней, то есть $\bar{x} \approx \tilde{x} \pm \mu_{\tilde{x}}$, где $\mu_{\tilde{x}}$ - ошибка выборочной средней.

Применение интервальных оценок означает, что характеристики генеральной совокупности укладываются в определенный диапазон значений. Чтобы их получить, необходимо рассчитать соответствующие ошибки выборки.

5. 3. Ошибки выборки

При правильном формировании выборки величину ее ошибки можно рассчитать заранее. В общем случае **под ошибкой выборки** понимают объективно возникающее расхождение между характеристиками выборки и генеральной совокупности.

Ошибки выборки подразделяются на *ошибки регистрации и ошибки репрезентативности*.

Ошибки регистрации возникают из-за неправильных или неточных сведений. Их источником является невнимательность регистратора, неправильное заполнение формуляров, описки или же непонимание существа исследуемого вопроса.

Ошибки репрезентативности возникают вследствие несоответствия структуры выборки структуре генеральной совокупности. Источником их существования является разная вариация признака у статистических единиц, в результате которой распределение единиц в выборочной совокупности отличается от распределения единиц в генеральной совокупности.

Ошибки репрезентативности делятся на систематические и случайные.

Систематические ошибки репрезентативности возникают из-за неправильного формирования выборки, при котором нарушается основной принцип научно организационной выборки – принцип случайности.

Случайные ошибки репрезентативности означают, что даже при соблюдении принципа случайности отбора единиц, расхождения между характеристиками выборки и генеральной совокупности все же имеют место. В разных выборках, сформированных по одной и той же генеральной совокупности, выборочная средняя и выборочная доля принимают различные значения в зависимости статистических единиц, попавших в выборку. Это означает, что *выборочная средняя и выборочная доля являются случайными величинами.*

Ошибки выборки также можно считать случайными величинами. Они могут принимать разные значения, поэтому определяют среднюю из возможных ошибок (стандартную).

Величина ошибки выборки зависит от следующих факторов:

- *степень колеблемости признака в генеральной совокупности*

Чем однороднее исследуемая совокупность, тем меньше величина средней ошибки при той же самой численности выборки.

- *объем (численность) выборки*

Увеличивая или уменьшая объем выборки n, можно регулировать величину средней ошибки. Чем больше единиц будет включено в выборку, тем меньше будет величина ошибки, так как тем точнее в выборке будет представлена генеральная совокупность.

- *способ отбора единиц в выборочную совокупность*

Для каждого способа формирования выборки величина ее ошибки определяется по-разному. В практической деятельности используются различные способы формирования выборочной совокупности, но принципиальное значение имеет их деление на *способы случайного повторного и бесповторного отбора*.

При собственно случайном повторном отборе общее число единиц генеральной совокупности в процессе выборке не меняется. *Статистическая единица, попавшая в выборку, после регистрации изучаемого признака возвращается в генеральную совокупность* и может вновь попасть в выборку. Таким образом, **для всех единиц генеральной совокупности обеспечивается равная вероятность отбора**.

В математической статистике доказывается, что средняя ошибка выборки $\mu_{\tilde{\alpha}}$ определяется по формуле:

$$\mu_{\tilde{\alpha}} = \sqrt{\frac{\sigma^2_{\tilde{\alpha}}}{n}},$$

где $\sigma^2_{\tilde{x}}$ - генеральная дисперсия.

Генеральная дисперсия, также как и остальные параметры генеральной совокупности является неизвестной величиной, но известно соотношение между генеральной и выборочной дисперсией: $\sigma^2_{\tilde{x}} \sim S^2_{\tilde{x}} \times \frac{n}{n-1}$; тогда при достаточно большом объеме выборки (n>30), $\frac{n}{n-1}$ является величиной близкой к 1, и можно считать, что $\sigma^2_{\tilde{x}} \sim S^2_{\tilde{x}}$.

В случаях малой выборки при n<30 необходимо учитывать отношение $\frac{n}{n-1}$ и рассчитывать среднюю ошибку малой выборки по формуле:

$$\mu_{\tilde{x}\ м.в.} = \sqrt{\frac{S^2_{\tilde{x}}}{n-1}}\ .$$

Таким образом, **для средней *количественного признака*** **средняя ошибка выборки** $\mu_{\bar{x}}$ равна:

$$\mu_{\tilde{x}} = \sqrt{\frac{S^2_{\tilde{x}}}{n}}\ ;$$

где $S^2_{\tilde{x}} = \frac{\sum_{i=1}^{n}(x_i - \tilde{x})^2}{n}$ - выборочная дисперсия количественного признака.

Средняя ошибка выборки для доли μ_{ω} определяется по формуле:

$$\mu_{\omega} = \sqrt{\frac{S^2_{\omega}}{n}}\ ;$$

где $S^2_{\omega} = \omega \times (1-\omega)$ - выборочная дисперсия доли альтернативного признака.

Применение простой случайной повторной выборки на практике весьма ограниченно. Это связано с тем, что практически нецелесообразно, а иногда и невозможно повторное наблюдение одних и тех же единиц, и поэтому однажды обследованная единица повторному учету не подвергается. В связи с этим чаще на практике применяется бесповторный отбор.

При бесповторном собственно случайном отборе общее количество статистических единиц в генеральной совокупности в процессе формирования выборки меняется, уменьшаясь каждый раз на единицу, попавшую в выборку, поскольку отобранные единицы в генеральную совокупность не возвращаются. Таким образом, *вероятность попадания отдельных единиц в выборку при бесповторном случайном отборе также меняется* (для оставшихся единиц она возрастает). В целом *вероятность попадания любой статистической единицы в выборку* при бесповторном отборе *может быть определена как* $1-\frac{n}{N}$. На эту величину должна быть скорректирована и средняя ошибка выборки при бесповторном отборе.

Таким образом, расчетные **формулы средней ошибки выборки при бесповторном отборе принимают вид:**

- **для средней** количественного признака

$$\mu_{\tilde{x}} = \sqrt{\frac{S^2_{\tilde{x}}}{n} \times (1-\frac{n}{N})}\ ;$$

- **для доли** альтернативного признака

$$\mu_{\omega} = \sqrt{\frac{\omega \times (1-\omega)}{n} \times (1-\frac{n}{N})}\,.$$

На практике при применении выборочного метода *определяются пределы,* за которые не выйдет величина конкретной ошибки выборочного исследования. Величина пределов конкретной ошибки определяется степенью вероятности, с которой измеряется ошибка выборки.

Ошибка выборки, исчисленная с заданной степенью вероятности, называется предельной ошибкой выборки.

Предельная ошибка выборки является максимально возможной при данной вероятности ошибкой. Это означает, что с заданной вероятностью гарантируется, что ошибка любой выборки не превысит предельную ошибку. Такая вероятность называется доверительной.

Предельная ошибка Δ выборки рассчитывается по формуле:

$\Delta_{\tilde{\alpha}} = t \times \mu_{\tilde{\alpha}}$;

где t – коэффициент доверия, значения которого определяются доверительной вероятностью $P(t)$.

Значения коэффициента доверия t задаются в таблицах нормального распределения вероятностей. Чаще всего используются следующие сочетания:

T	$P(t)$
1	0,683
1,5	0,866
2,0	0,954
2,5	0,988
3,0	0,997
3,5	0,999

Так, если t = 1, то с вероятностью 0,683 можно утверждать, что расхождение между выборочными характеристиками и параметрами генеральной совокупности не превысит одной средней ошибки.

Для экономических задач доверительная вероятностьобычно принимается равной 0,95.

Предельные ошибки выборки Δ для разных параметров при разных методах отбора статистических единиц рассчитываются по формулам, приведенным в таблице 5.2.

Таблица 5.2.

Предельные ошибки выборки

Метод отбора	Предельные ошибки выборки	
	Для средней	Для доли
Повторный	$\Delta_{\tilde{x}} = t \times \sqrt{\frac{S^2_{\tilde{x}}}{n}}$	$\Delta_{\omega} = t \times \sqrt{\frac{\omega \times (1-\omega)}{n}}$
Бесповторный	$\Delta_{\tilde{x}} = t \times \sqrt{\frac{S^2_{\tilde{x}}}{n} \times (1-\frac{n}{N})}$	$\Delta_{\omega} = t \times \sqrt{\frac{\omega \times (1-\omega)}{n} \times (1-\frac{n}{N})}$

Зная величину предельной ошибки выборки, можно рассчитать *интервалы для характеристик генеральной совокупности*:

Доверительный интервал для генеральной средней равен $(\tilde{x} \pm \Delta\tilde{x})$; **для генеральной доли** - $(\omega \pm \Delta\omega)$.

5.4. Способы формирования выборочной совокупности

Способы формирования выборки (отбора) влияют на результат выборочного исследования, в частности, на точность статистических оценок параметров генеральной совокупности.

Основное требование к отбору заключается в том, что он должен быть по возможности простым.

Различают два способа формирования выборки:

- ***простой собственно-случайный,***
- ***отбор с предварительным разделением генеральной совокупности на части.***

5.4.1. При простом собственно-случайном отборе на включение или исключение какой-либо статистической единицы в выборку влияет только случай. Это обеспечивает равную вероятность каждой единице попасть в выборку.

Технически собственно-случайный отбор проводят методом жеребьевки или по таблице случайных чисел. При этом можно ожидать, что среди отобранных единиц имеются представители разных состояний, которыми характеризуется признак в общей совокупности. В таком случае среднее значение изучаемого признака окажется представленным достаточно точно.

Собственно-случайный отбор в «чистом виде» применяется редко, но он является исходным для всех других видов отбора.

Случайный отбор может быть повторным или бесповторным.

- **При повторном отборе** статистические единицы, отобранные ранее, возвращаются в генеральную совокупность и могут вновь попасть в выборку.

При этом численность генеральной совокупности при проведении отбора остается постоянной, тем самым обеспечивается каждой статистической единице равная возможность попасть в выборку

- **При бесповторном отборе** единицы не возвращаются обратно в генеральную совокупность, ее численность с каждой единицей сокращается, абсолютно равная возможность каждой статистической единице попасть в выборку полностью не обеспечивается. Но при этом при одном и том же объеме выборки наблюдение охватывает больше единиц генеральной совокупности, что обеспечивает более точные результаты по сравнению с повторным отбором (меньшую ошибку выборки).

Бесповторный отбор находит более широкое применение на практике. Он используется в тех случаях, когда нельзя применить повторную выборку, например, при обследовании потребительского спроса, изучении общественного мнения по какому-либо вопросу и т. п.

Отбор с предварительным разделением генеральной совокупности на части может быть организован различными способами, которым соответствуют свои виды отбора.

В практике выборочных исследований наибольшее распространение получили следующие виды выборки:

- *механическая;*
- *типическая;*
- *серийная;*
- *комбинированная*.

Указанные виды выборки являются дальнейшим развитием и видоизменением собственно-случайного отбора. Их применение вызывается соображениями удешевления или облегчения процесса наблюдения, особым характером объектов наблюдения.

5.4.2. Механический отбор относится к наиболее применяемым способам формирования выборки. При механическом отборе *генеральная совокупность предварительно упорядочивается по несущественному для цели исследования признаку* (списки избирателей, табельные номера работников, различные другие базы данных). Отбор осуществляется *бесповторным способом через равные интервалы*. Из каждого интервала в выборку попадает только одна единица.

При проведении механической выборки необходимо установить *шаг отсчета* (расстояние между отбираемыми единицами) и *начало отсчета* (номер единицы, которая должна быть обследована первой). Шаг отсчета устанавливается, исходя из предполагаемого процента отбора. Например, при 10%-ой выборке отбирается каждая десятая единица, при 20%-ой – каждая двадцатая.

Особенностью механического отбора является то, что при его применении *возможно появление систематических ошибок*, связанное со случайным совпадением выбранного интервала и циклических закономерностей в расположении единиц генеральной совокупности. Чтобы избежать систематических ошибок, следует отбирать статистическую единицу, находящуюся в середине каждого интервала.

Этот способ очень удобен в тех случаях, когда нельзя заранее составить список единиц генеральной совокупности (выборка берется из постоянно формирующейся во времени совокупности). В таком случае, например, при изучении спроса на определенный товар, удобно наблюдать каждого десятого или каждого двадцатого входящего в магазин покупателя; или же при контроле качества продукции – проверять каждое пятое или каждое десятое изделие, сходящее с конвейера.

При определении средней ошибки механической выборки используются формулы средней ошибки при собственно-случайном бесповторном отборе:

для выборочной средней $\mu_{\tilde{x}} = \sqrt{\frac{S^2_{\tilde{x}}}{n} \times (1 - \frac{n}{N})}$;

для выборочной доли $\mu_{\omega} = \sqrt{\frac{\omega \times (1-\omega)}{n} \times (1 - \frac{n}{N})}$.

5.4.3. Расслоенный (стратифицированный) отбор *используется при изучении сложных совокупностей*, которые можно разбить на несколько качественно однородных групп по существенным для целей исследования признакам. *Внутри каждой группы проводится собственно-случайный или механический отбор*. Полученные группы по численности единиц, как правило, не равны между собой, поэтому *отбор единиц осуществляется пропорционально объему группы,* т. е. количество отбираемых в выборку единиц пропорционально удельному весу данной группы по числу единиц в генеральной совокупности. Таким образом, число наблюдений по каждой группе определяется по формуле:

$$n_{iв} = N \cdot \frac{n_i}{N},$$

где $n_{iв}$ - число наблюдений из i-ой группы генеральной совокупности,

N – объем генеральной совокупности,

n_i - объем i-ой группы генеральной совокупности.

Если пропорции между группами в выборке совпадают с пропорциями между группами в генеральной совокупности, то отбор называется **типическим.**

Типическая выборка обеспечивает более точные результаты по сравнению с другими способами отбора единиц в выборочную совокупность, так как позволяет исключить влияние межгрупповой дисперсии δ^2 на среднюю ошибку выборки.

На величину средней ошибки типической выборки влияет только величина средней из внутригрупповых дисперсий.

Типическую выборку можно получить повторным или бесповторным отбором:

Среднюю ошибку типической выборки *при повторном отборе* определяют по формулам:

- *для средней количественного признака:* $\mu_{\tilde{x}} = \sqrt{\frac{\overline{S}_i^2}{n}}$,

где $\overline{S}_i^2$ - средняя из внутригрупповых выборочных дисперсий;

- *для доли альтернативного признака:* $\mu_{\omega} = \sqrt{\frac{\overline{\omega_i \times (1-\omega_i)}}{n}}$,

где $\overline{\omega_i \times (1-\omega_i)}$ - средняя из внутригрупповых дисперсий доли альтернативного признака по выборке.

При бесповторном отборе **среднюю ошибку типической выборки** рассчитывают по следующим формулам:

-*для средней количественного признака:*

$$\mu_{\tilde{x}} = \sqrt{\frac{\overline{s_i^2}}{n} \times (1-\frac{n}{N})},$$

- *для доли альтернативного признака:*

$$\mu_{\omega} = \sqrt{\frac{\overline{\omega_i \times (1-\omega_i)}}{n} \times (1-\frac{n}{N})}.$$

5.4.4. Серийная выборка применяется в тех случаях, когда *единицы статистической совокупности объединены в небольшие группы или серии*. В качестве таких серий могут рассматриваться, например, упаковки с определенным количеством готовой продукции.

Для отбора серий применяют либо собственно-случайную, либо механическую выборку. Наблюдению подвергаются все единицы отобранной серии.

Серийный отбор имеет большое практическое значение, так как обследуется незначительное число серий, и это сокращает расходы на проведение наблюдения; однако при серийном отборе случайная ошибка получается несколько большей, чем при других способах отбора.

При серийном отборе, поскольку внутри серий обследуются все без исключения статистические единицы, *величина средней ошибки зависит только от межгрупповой* (межсерийной) *дисперсии.*

Средняя ошибка серийной выборки при повторном отборе определяется следующим образом:

- ***для средней количественного признака***:

$$\mu_{\tilde{x}} = \sqrt{\frac{\delta_{\tilde{x}}^2}{r}},$$

где $\delta_{\tilde{x}}^2 = \frac{\sum(\tilde{x}_i - \tilde{x})^2}{r}$,

$\tilde{x}_i$ - среднее i-той серии,

$\tilde{x}$ - средняя по всей выборке,

r – число отобранных серий;

- ***для доли альтернативного признака***:

$$\mu_{\omega} = \sqrt{\frac{\delta_{\omega}^2}{r}},$$

где $\delta_{\omega}^2 = \frac{\sum(\omega_i - \overline{\omega})^2}{r}$ - межгрупповая дисперсия доли серийной выборки;

ω_i - доля признака в i-той доли;

$\overline{\omega}$ - средняяя доля альтернативного признака во всей выборке.

При бесповторном отборе средняя ошибка серийной выборки может быть определена:

- ***для средней количественного признака***:

$$\mu_{\tilde{x}} = \sqrt{\frac{\delta_{\tilde{x}}^2}{r} \times (1 - \frac{r}{R})},$$

где R – общее число серий в генеральной совокупности.

- ***для доли альтернативного признака***:

$$\mu_{\omega} = \sqrt{\frac{\delta_{\omega}^2}{r} \times (1 - \frac{r}{R})}.$$

Рассмотренные способы формирования выборки могут применяться в «чистом виде», а могут комбинироваться в различных сочетаниях и последовательности. Использование нескольких методов формирования выборки в одном выборочном исследовании называется **комбинированной выборкой** (отбором).

Такая выборка проводится в несколько этапов, и на каждом из них применяется свой способ отбора.

Например, при обследовании семейных доходов выборочное обследование проводится в такой последовательности:

- устанавливаются населенные пункты, попадающие под обследование.

При этом используется расслоенный отбор, с помощью которого отбираются крупные города, средние города, и другие населенные пун

- в каждом населенном пункте устанавливаются места, где проживают

семьи - улицы, дома (на основе механического отбора по списку улиц и нумерации домов);

- в каждом месте проживания семей отбираются конкретные семьи,

для чего применяется собственно-случайный бесповторный или механический отбор. Для отбора используют перечень номеров квартир или списки семей.

Методы формирования выборки влияют на точность статистических оценок через ошибки выборки), а также на объем выборочной совокупности, на ее численность.

5.5. Численность выборки и способы распространения ее характеристик на Генеральную совокупность

Численность выборки – один из факторов, влияющих на величину ее ошибки: *чем она больше, тем меньше ошибка*. С другой стороны, с объемом выборки связаны затраты на проведение исследования: *чем она больше, тем больше затраты.*

Таким образом, **выборка должна быть оптимальной по численности**, чтобы обеспечить достоверность результатов исследования и не вызвать дополнительных затрат труда и денежных средств.

Численность выборки может быть определена исходя из допустимой ошибки при выборочном наблюдении, способа отбора статистических единиц.

Для определения необходимой численности выборки необходимо задаться предельной ошибкой выборки.

В общем случае предельная ошибка выборки связана с ее численностью следующим соотношением:

$\Delta = t \times \mu_\alpha = t \times \sqrt{\frac{S^2_{\tilde{\alpha}}}{n}}$, следовательно:

$n = \frac{t^2 \times S^2_{\tilde{\alpha}}}{\Delta^2}$.

Приведенная формула показывает, что с увеличением предполагаемой ошибки значительно уменьшается необходимый объем выборки и наоборот.

Для разных характеристик и разных методов формирования выборок формулы для определения необходимой численности выборки приведены в таблице 5.3.

Таблица5.3.

Численность выборки при разных методах отбора

Метод отбора	Формулы определения объема выборки	
	Для средней	Для доли
Повторный	$n = \frac{t^2 \times S^2_{\tilde{x}}}{\Delta^2_{\tilde{x}}}$	$n = \frac{t^2 \times \omega \times (1-\omega)}{\Delta^2_{\omega}}$
Бесповторный	$n = \frac{t^2 \times S^2_{\tilde{x}} \times N}{N \times \Delta^2_{\tilde{x}} + t^2 \times S^2_{\tilde{x}}}$	$n = \frac{t^2 \times \omega \times (1-\omega) \times N}{N \times \Delta^2_{\omega} + t^2 \times \omega \times (1-\omega)}$

На практике определение необходимого объема выборки часто составляет серьезную проблему, связанную с определением показателя вариации изучаемого признака. К началу проведения выборочного наблюдения показатели вариации неизвестны.

Приблизительно *показатель вариации определяют одним из следующих способов*:

- берут из предыдущих исследований;
- по правилу «трех сигм» общий размах вариации R при нормальном распределении укладывается в 6 среднеквадратических отклонений σ: $R \cong 6\sigma$, отсюда $\sigma \approx \frac{R}{6}$; для бóльшей точности R делят на 5;
- если хотя бы приблизительно известна средняя величина изучаемого признака $\bar{x}$, то среднеквадратическое отклонение $\sigma \approx \frac{\bar{x}}{3}$;
- при изучении альтернативного признака, если нет других данных можно брать максимальную величину дисперсии, равную 0,25, то есть $\omega \times (1-\omega) = 0{,}25$.
- проводят «пробную» выборку, по которой рассчитывают показатель вариации, используемый в качестве оценки генеральной совокупности.

Характеристики выборки могут быть распространены на генеральную совокупность *с помощью одного из двух способов распространения выборочных данных:*

1) способа прямого пересчета;
2) способа поправочных коэффициентов.

При первом способе средние величины и доли, полученные по выборке, переносятся на генеральную совокупность. При этом генеральная средняя определяется как $\bar{x} \cong \tilde{x} \pm \Delta\tilde{x}$, а генеральная доля – как $\bar{d} \cong \omega \pm \Delta\omega$.

Способ поправочных коэффициентов применяется, когда целью выборочного исследования является *уточнение результатов сплошного наблюдения.* Для этого после

обобщения данных сплошного наблюдения практикуется 10%-ное выборочное наблюдение с установлением поправочного коэффициента γ, который устанавливает процент расхождений между данными сплошного и выборочного наблюдения.

Контрольные вопросы и задания

1. Даете определение выборочного наблюдения.
2. Какими преимуществами обладает выборочное наблюдение по сравнению со сплошным?
3. Перечислите и сформулируйте цели этапов выборочного наблюдения.
4. Сформулируйте задачу выборочного исследования.
5. Что такое статистическая оценка, каким требованиям она должна соответствовать?
6. Дайте определение ошибки выборки, от чего зависит её величина?
7. Что представляет собой средняя ошибка выборки (для доли и средней)?
8. Чем отличается повторный отбор от бесповторного?
9. Как считаются средние ошибки выборки при повторном и бесповторном отборе?
10. Что представляет собой предельная ошибка выборки; как она вычисляется для средней и для доли?
11. Какие факторы влияют на объем выборки?
12. Как определяется объем выборки при случайном отборе?
13. Какие способы распространения выборочных данных на генеральную совокупность Вам известны? В чем их суть?

Литература

1. Васильева В.К., Лялин В.С. Статистика: учебник, М.:ЮНИТИ, 2012
2. Вуколов Э.А. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и Excel – Форум-Инфра-М, 2014
3. Годин А.М. Статистика: учебник, М.: Дашков и К, 2014
4. Елисеева И.И. Статистика: учебник - Издатель Проспект, 2014
5. Ефимова М.Р., Петрова Е.В., Румянцев В.Н. Общая теория статистики: учебник - ИНФРА-М, 2013
6. Ефимова М.Р. Практикум по общей теории статистики: учебное пособие – М. Финансы и статистика, 2014
7. Назаров М.Г., Минашкин В.Г., Мхитарян В.С. Статистика: учебник для ВУЗов. – Омега-Л, 2011.
8. Минашкин В.Г. Статистика: учебник для бакалавров – М. Юрайт, 2013
9. Просветов Г.И. Статистика: задачи и решения. Учебно-методическое пособие – М. Альфа-Пресс, 2014
10. Статистика: учебник для теория и практика в Excel – М. Финансы и статистика.2010
11. Статистика: учебник для бакалавров, под ред. В.С. Мхитаряна –М. Юрайт, 3013
12. Сизова Т.М. Статистика: учебное пособие – СПб. НИУ ИТМО, 2013
13. Сизова Т.М., Мишура Л.Г. Статистика: практикум – СПб. Университет ИТМО, 2016
14. Электронно-библиотечная система. Издательство «Лань» [Электронный ресурс] Елисеева И.И. Практикум по общей теории статистики – М. Финансы и

статистика,

2008. http://e.lanbook.com.academicnt.ru/books/element.php?pl1_id=53865

15. Электронно-библиотечная система. Издательство «Лань» [Электронный ресурс]Лялин В.С. Статистика: теория и практика в Excel: учебное пособие - М. Финансы и
статистика,2010. http://e.lanbook.com.academicnt.ru/books/element.php?pl1_id=1048

16.Электронно-библиотечная система. Издательство «Лань» [Электронный ресурс] Годин А.М. Статистика: учебник - М.: Дашков и К, 2014 Электронно-библиотечная система. Издательство «Лань»
http://e.lanbook.com.academicnt.ru/books/element.php?pl1_id=56301

Приложение 1

Таблица значений функции $f(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$

t	0	1	2	3	4	5	6	7	8	9
0,0	3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3725	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0855	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0203	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107

Приложение 2

Значения χ^2 - критерия Пирсона

Число степеней свободы К=(м$_1$1)·(м$_2$ -1)	Уровень значимости			Число степеней свободы	Уровень значимости		
	0,10	0,05	0,01		0,10	0,05	0,01
1	2	3	4	1	2	3	4
1	2,71	3,84	6,63	21	29,62	32,67	38,93
2	4,61	5,99	9,21	22	30,81	33,92	40,29
3	6,25	7,81	11,34	23	32,01	35,17	41,64
4	7,78	9,49	13,28	24	33,20	36,42	42,98
5	9,24	11,07	15,09	25	34,38	37,65	44,31
6	10,64	12,59	16,81	26	35,56	38,89	45,64
7	12,02	14,07	18,48	27	36,74	40,11	46,96
8	13,36	15,51	20,09	28	37,92	41,34	48,28
9	14,68	16,92	21,67	29	39,09	42,56	49,59
10	15,99	18,31	23,21	30	40,26	43,77	50,89
11	17,28	19,68	24,72	40	51,80	55,76	63,69
12	18,55	21,03	26,22	50	63,17	67,50	76,15
13	19,81	22,36	27,69	60	74,40	79,08	88,38
14	21,06	23,68	29,14	70	85,53	90,53	100,42
15	22,31	25,00	30,58	80	96,58	101,88	112,33
16	23,54	26,30	32,00	90	107,56	113,14	124,12
17	24,77	27,59	33,41	100	118,50	124,34	135,81
18	25,99	28,87	34,81				
19	27,20	30,14	36,19				
20	28,41	31,41	37,57				

Приложение 3

Значения t -критерия Стьюдента

Число степеней свободы n-1	Уровень значимости			Число степеней свободы	Уровень значимости		
	0,10	0,05	0,01		0,10	0,05	0,01
1	2	3	4	1	2	3	4
1	6,314	12,706	63,657	18	1,734	2,101	2,878
2	2,920	4,303	9,925	19	1,729	2,093	2,861
3	2,353	3,182	5,841	20	1,725	2,086	2,845
4	2,132	2,776	4,604	21	1,721	2,080	2,831
5	2,015	2,571	4,032	22	1,717	2,074	2,819
6	1,943	2,447	3,707	23	1,714	2,069	2,807
7	1,895	2/365	3,499	24	1,711	2,064	2,797
8	1,859	2,306	3,355	25	1,708	2,059	2,787
9	1,833	2,262	3,249	26	1,706	2,055	2,779
10	1,812	2,228	3,169	27	1,703	2,052	2,771
11	1,796	2,201	3,106	28	1,701	2,048	2,763
12	1,782	2,179	3,054	29	1,699	2,045	2,756
13	1,771	2,160	3,012	30	1,697	2,042	2,750
14	1,761	2,145	2,977	40	1,684	2,02	2,704
15	1,753	2,131	2,947	60	1,671	2,000	2,660
16	1,746	2,120	2,921	120	1,658	1,980	2,617
17	1,740	2,110	2,898				

Приложение 4

Значения F-критерия Фишера при уровне значимости 0,05

V_1 / V_2	1	2	3	V_1 / V_2	1	2	3
1	161,00	200,00	216,00	18	4,41	3,55	3,16
2	18,51	19,00	19,16	19	4,38	3,52	3,13
3	10,13	9,55	9,28	20	4,35	3,49	3,10
4	7,71	6,94	6,59	21	4,32	3,47	3,07
5	6,61	5,79	5,41	22	4,30	3,44	3,05
6	5,99	5,14	4,76	23	4,28	3,42	3,03
7	5,59	4,74	4,35	24	4,26	3,40	3,01
8	5,32	4,46	4,07	25	4,24	3,88	2,99
9	5,12	4,26	3,86	26	4,22	3,37	2,98
10	4,96	4,10	3,71	27	4,21	3,35	2,96
11	4,84	3,98	3,59	28	4,20	3,34	2,95
12	4,75	3,88	3,49	29	4,18	3,33	2,93
13	4,67	3,80	3,41	30	4,17	3,32	2,92
14	4,60	3,74	3,34	40	4,08	3,23	2,84
15	4,54	3,68	3,29	50	4,03	3,18	2,79
16	4,49	3,63	3,24	60	4,00	3,15	2,76
17	4,45	3,59	3,20	100	3,94	3,09	2,70

V_1=m-1; V_2=n-m; n-число наблюдений; m-число признаков.

Приложение 5

Распределение вероятности в малых выборках в зависимости от коэффициента доверия *t* и объема выборки *n*

N / *T*	4	5	6	7	8	9	10	15	20	∞
0,5	348	356	362	366	368	370	372	376	378	383
1,0	608	626	636	644	650	654	656	666	670	683
1,5	770	792	806	816	832	828	832	846	850	865
2,0	860	884	908	908	914	920	924	936	940	954
2,5	933	946	955	959	963	966	968	975	978	988
3,0	942	960	970	970	980	938	984	992	992	997

При n=∞ в таблице даны вероятности нормального распределения. Для определения вероятности соответствующие табличные значения следует разделить на 1000.

УНИВЕРСИТЕТ ИТМО

Миссия университета – генерация передовых знаний, внедрение инновационных разработок и подготовка элитных кадров, способных действовать в условиях быстро меняющегося мира и обеспечивать опережающее развитие науки, технологий и других областей для

содействия решению актуальных задач.

КАФЕДРА ФИНАНСОВОГО МЕНЕДЖМЕНТА И АУДИТА

Кафедра финансового менеджмента и аудита (ФМиА) создана в 2015 году на базе трех кафедр: экономики и финансов, экономики и предпринимательской деятельности, финансового менеджмента. Заведующий кафедры ФМиА – профессор, доктор экономических наук Сергеева Ирина Григорьевна. Кафедра является выпускающей кафедрой факультета технологического менеджмента и инноваций. Кафедра проводит обучение бакалавров по направлению 38.03.02 «Менеджмент», программы «Финансовый менеджмент», «Логистика», «Управление малым бизнесом». Кафедрой ФМиА осуществляется подготовка магистров по направлению 38.04.02 «Менеджмент», магистерские программы «Финансовый менеджмент», «Ресурсный менеджмент в инновационной деятельности», «Антикризисное управление и аудит»; по направлению 27.04.05 «Инноватика», магистерская программа «Экономика и управление инновационной деятельностью в областях науки»; по направлению 27.04.02 «Управление качеством», магистерская программа «Аудит качества и стандартизация».

Сизова Тамара Митрофановна

СТАТИСТИКА ДЛЯ БАКАЛАВРОВ
Учебное пособие
Часть II

В авторской редакции
Редакционно-издательский отдел Университета ИТМО
Зав. РИО Н.Ф. Гусарова
Подписано к печати
Заказ №
Тираж 75 экз.
Отпечатано на ризографе

УНИВЕРСИТЕТ ИТМО

Редакционно-издательский отдел
Санкт-Петербургского
национально-исследовательского университета информационных технологий,
механики и оптики