Статистика

Слово «статистика» происходит от латинского status — состояние дел. В науку термин «статистика» ввел немецкий ученый Готфрид Ахенваль в 1746 году, предложив заменить название курса «Государствоведение», преподававшегося в университетах Германии, на «Статистику», положив тем самым начало развитию статистики как науки и учебной дисциплины. Несмотря на это, статистический учет вёлся намного раньше: проводились переписи населения в Древнем Китае, осуществлялось сравнение военного потенциала государств, велся учет имущества граждан в Древнем Риме и т. п.[2]

Статистика разрабатывает специальную методологию исследования и обработки материалов: массовые статистические наблюдения, метод группировок, средних величин, индексов, балансовый метод, метод графических изображений и другие методы анализа статистических данных.

Курс «Статистика» имеет целью дать студентам представление о содержании статистики как научной дисциплины, познакомить с ее основными понятиями, методологией и методиками расчета важнейших статистических аналитических показателей. В соответствии с этим данное учебное пособие охватывает самые общие начальные элементы статистической науки, и прежде всего важнейшие направления анализа социально-экономических процессов. В дальнейшем на базе курса «Статистика» изучаются конкретные статистические дисциплины: теория статистического наблюдения, анализ и прогнозирование временных рядов, классификации и группировки, многомерные статистические методы, экономическая и отраслевые статистики, анализ хозяйственной и финансовой деятельности. Что касается общеэкономических специальностей, данный курс служит основой для разработки и совершенствования методов экономического анализа.

Адрес электронной почты в интернете для связи с преподавателем: PopovaMarina1973@yandex.ru

История, пути и направления статистической науки

Термин "статистика" появился в середине 18 века. Означал "государствоведение". Получил распространение в монастырях. Постепенно приобрел собирательное значение. С одной стороны, статистика – это совокупность числовых показателей, характеризующих общественные явления и процессы (статистика труда, статистика транспорта).

С другой – под статистикой понимается практическая деятельность по сбору, обработке, анализу данных по различным направлениям общественной жизни. С третьей стороны, статистика – это итоги массового учета, опубликованные в различных сборниках.

Наконец, в естественных науках статистикой называются методы и способы оценки соответствия данных массового наблюдения математическим формулам. Таким образом, **статистика – это** общественная наука, изучающая количественную сторону массовых общественных явлений в неразрывной связи с их качественной стороной.

Ученые, внесшие вклад в развитие статистики

- Уильям Петти – основатель статистики. Его заслуга в том, что он впервые применил числовой метод для анализа закономерностей общественной жизни. Работа – "Политическая арифметика".
- Адольф Кетле – бельгийский статистик. Доказал, что даже кажущиеся случайности общественной жизни обладают внутренней закомерностью и необходимостью.
- К.Ф. Герман – русский статистик ("Всеобщая теория статистики").
- В.И. Ленин – теория группировок, теория статистического наблюдения.
- Целый ряд других ученых.

Предмет статистики

Статистика изучает количественно определенные качества массовых социально-экономических

Статистика

явлений.   1   2   3

Существует несколько точек зрения на статистику как на науку:

(1) Статистика – это **универсальная наука**, изучающая массовые явления природы и общества.

(2) Статистика – это **методологическая наука**, разрабатывающая методы исследования для других наук.

(3) Статистика – это **общественная наука**.

Явления общественной жизни – это сложное сочетание различных элементов.

– Общественные явления обладают вполне конкретными размерами.

– Общественным явлениям присущи определенные количественные соотношения, и существуют они независимо от того, изучает ли их статистика или нет.

Размеры и соотношения количества и качества отдельных явлений статистика выражает при помощи определенных понятий, статистических показателей. Числовое значение показателя, относящееся к определенному месту и времени, называют величиной показателя.

Отрасли статистики

Общая теория статистики – это лишь фундамент. В любой своей части она связана с другими науками.

<table>
<tr><th colspan="13">Общая теория статистики</th></tr>
<tr><td rowspan="3">Демографическая статистика</td><td colspan="9">Экономическая статистика</td><td rowspan="3">Статистика образования</td><td rowspan="3">Медицинская статистика</td><td rowspan="3">Спортивная статистика</td></tr>
<tr><td rowspan="2">Статистика труда</td><td rowspan="2">Статистика заработной платы</td><td rowspan="2">Статистика мат.-техн. снабжения</td><td rowspan="2">Статистика транспорта</td><td rowspan="2">Статистика связи</td><td colspan="4">Статистика финансового кредита</td></tr>
<tr><td>Высшие финансовые вычисления</td><td>Статистика денежного обращения</td><td>Статистика валютных курсов</td><td>Прочие</td></tr>
</table>

Статистика также разрабатывает теорию наблюдения.

Метод статистики

Метод статистики предполагает следующую последовательность действий:

– разработка статистической гипотезы,
– статистическое наблюдение,
– сводка и группировка статистических данных,
– анализ данных,

– интерпретация данных.

Прохождение каждой стадии связано с использованием специальных методов, объясняемых содержанием выполняемой работы.

Закон больших чисел

Массовый характер общественных законов и своеобразие их действий предопределяет необходимость исследования совокупных данных.

Закон больших чисел порожден особыми свойствами массовых явлений. Последние в силу своей индивидуальности, с одной стороны, отличаются друг от друга, а с другой – имеют нечто общее, обусловленное их принадлежностью к определенному классу, виду. Причем единичные явления в большей степени подвержены воздействию случайных факторов, ежели их совокупность.

Закон больших чисел в наиболее простой форме гласит, что количественные закономерности массовых явлений отчетливо проявляются лишь в достаточно большом их числе.

Таким образом, сущность его заключается в том, что в числах, получающихся в результате массового наблюдения, выступают определенные правильности, которые не могут быть обнаружены в небольшом числе фактов.

Закон больших чисел выражает диалектику случайного и необходимого. В результате взаимопогашения случайных отклонений средние величины, исчисленные для величины одного и того же вида, становятся типичными, отражающими действия постоянных и существенных фактов в данных условиях места и времени.

Тенденции и закономерности, вскрытые с помощью закона больших чисел, имеют силу лишь как массовые тенденции, но не как законы для каждого отдельного случая.

Статистическая закономерность

Статистические закономерности изучают распределение единиц статистического множества по отдельным признакам под воздействием всей совокупности факторов.

Статистическая закономерность выступает как объективная закономерность сложного массового процесса и является формой причинной связи. Она обнаруживается в итоге массового статистического наблюдения. Этим обуславливается ее связь с законом больших чисел.

Статистическая закономерность с определенной вероятностью гарантирует устойчивость средних величин при сохранении постоянного комплекса условий, порождающих данное явление.

Задачи статистики

(1) Разработка системы гипотез, характеризующих развитие, динамику, состояние социально-экономических явлений.

(2) Организация статистической деятельности.

(3) Разработка методологии анализа.

(4) Разработка системы показателей для управления хозяйством на макро- и микроуровне.

(5) Популяризовать данные статистического наблюдения.

Организация государственной статистики в РФ

Принципы:

(1) централизованное руководство,

(2) единое организационное строение и методология,

(3) неразрывная связь с органами государственного управления.

Система государственной статистики имеет иерархическую структуру. Эта структура имеет федеральный, республиканский, краевой, областной, окружной, городской и районный уровни.

Статистика

Госкомстат имеет управления, отделы, вычислительный центр.

Ряды распределения

Рядами распределения называются группировки особого вида, при которых по каждому признаку, группе признаков или классу признаков известны численность единиц в группе либо удельный вес этой численности в общем итоге. Ряды распределения могут быть построены или по количественному, или по атрибутивному признаку.

Ряды распределения, построенные по количественному признаку, называются **вариационными рядами**. Ряд распределения может быть построен по ***непрерывно варьирующему признаку*** (когда признак может принимать любые значения в рамках какого-либо интервала) и по ***дискретно варьирующему признаку*** (принимает строго определенные целочисленные значения). Непрерывно варьирующий признак изображается графически при помощи ***гистограммы***. Дискретный же ряд распределения графически представляется в виде ***полигона распределения***

Контрольные вопросы:

1. Понятие статистики
2. Определение и основные черты предмета статистики
3. Теоретические основы статистики как науки
4. Методы статистики
5. Основная задача и принципы организации государственной статистики в РФ

Понятие статистического наблюдения

Статистическое наблюдение – это сбор необходимых данных по явлениям, процессам общественной жизни. Но это не всякий сбор данных, а лишь планомерный, научно организованный, систематический и направленный на регистрацию признаков, характерных для исследуемых явлений и процессов. От качества данных, полученных на первом этапе, зависят конечные результаты исследования.

Формы статистического наблюдения

Различают две основные формы статистического наблюдения – отчетность и специально организованное наблюдение.

Отчетность – это такая форма наблюдения, при которой предприятия, организации представляют в статистические и вышестоящие органы постоянные сведения, характеризующие их деятельность. Отчетность предоставляется по заранее определенной программе в строго определенные сроки и содержит важнейшие показатели, необходимые в процессе ежедневной работы.

Специально организованное наблюдение – такое наблюдение, которое организуется со специальной целью на определенную дату для получения данных, которые в силу различных причин не собираются статистической отчетност, а также с целью проверки данных статистической отчетности.

Виды статистического наблюдения

По ***времени регистрации фактов*** статистическое наблюдение может быть непрерывным, периодическим и единовременным.

Непрерывное (текущее) наблюдение – ведется систематически (т.е. регистрация фактов производится по мере их свершения). *Пример – ЗАГС.*

Периодическое наблюдение – повторяется через определенные равные промежутки времени. *Пример – перепись населения.*

Единовременное наблюдение – производится по мере надобности без соблюдения определенной периодичности. *Пример – оценка и переоценка основных фондов.*

По ***охвату единиц совокупности*** выделяют сплошное и несплошное наблюдение.

Статистика

Сплошным называется наблюдение, при котором исследованию подвергаются все единицы изучаемой совокупности.

Несплошным называется такое наблюдение, при котором исследованию подвергается только часть единиц изучаемой совокупности, отобранная определенным образом.

Виды несплошного наблюдения

– Анкетный способ

Исследуются какие-то осредненные показатели и распространяются на всю совокупность.

– Метод основного массива

Исследуются наиболее крупные единицы изучаемого явления.

– Метод направленного долевого отбора

– Выборочный метод

Его основой является случайный отбор. Результат гарантируется с определенной вероятностью ***p***.

– Монографический метод

Подвергаются тщательному исследованию отдельные единицы совокупности, обычно представители новых типов, либо самые лучшие (худшие) единицы. Результаты переносятся на всю совокупность. Позволяет выявить тенденции.

Способы статистического наблюдения

Основанием для регистрации фактов могут служить либо документы, либо высказанное мнение, либо хронометражные данные. В связи с этим различают наблюдение:

– непосредственное (сами измеряют),
– документально (из документов),
– опрос (со слов кого-либо).

В статистике применяются следующие способы сбора информации:

– корреспондентский (штат добровольных корреспондентов),
– экспедиционный (устный, специально подготовленные работники)
– анкетный (в виде анкет),
– саморегистрация (заполнение формуляров самими респондентами),
– явочный (браки, дети, разводы) и т.д.

Программно-методологические вопросы статистического наблюдения

Каждое наблюдение проводится с конкретной целью. При его проведении необходимо установить, что подлежит обследованию. Надо решить следующие вопросы:

Объект наблюдения – совокупность предметов, явлений, у которых должны быть собраны сведения. При определении объекта указываются его основные отличительные черты (признаки). Всякий объект массовых наблюдений состоит их отдельных единиц, поэтому надо решить вопрос о том, каков тот элемент совокупности, который послужит единицей наблюдения.

<u>Единица наблюдения</u> – это составной элемент объекта, который является носителем признаков, подлежащих регистрации и основой счета.

<u>Ценз</u> – это определенные количественные ограничения для объекта наблюдения.

<u>Признак</u> – это свойство, которое характеризует определенные черты и особенности, присущие единицам изучаемой совокупности.

<u>Программа наблюдения</u> – это перечень признаков, подлежащих регистрации. Программа находит отражение в ***формуляре наблюдения***. Выделяются организационные вопросы: перечень мероприятий, обеспечивающих правильность наблюдения, а также ***оргплан***, где учитываются органы наблюдения, время наблюдения, порядок приема и сдачи материала, порядок получения информации.

<u>Период наблюдения</u> – время, в течение которого должна быть осуществлена регистрация.

<u>Критическая дата наблюдения</u> – дата, по состоянию на которую сообщаются сведения.

<u>Критический момент</u> – момент времени, по состоянию на который производится регистрация наблюденных фактов.

Контрольные вопросы:

1. Понятие, этапы проведения статистического наблюдения
2. Цели, задачи и объекты статистического наблюдения
3. Основные формы статистического наблюдения
4. Основные виды статистического наблюдения
5. Основные способы статистического наблюдения
6. Точность наблюдения

Статистическая сводка – это операция по обработке собранных данных, которые выражаются в виде показателей, относящихся к каждой единице объекта статистического наблюдения. В результате сводки эти данные превращаются в систему статистических таблиц и промежуточных итогов. По результатам сводки можно выявить наиболее типичные черты и закономерности изучаемых явлений.

Предварительно составляется программа и план сводки.

В программе определяется подлежащее и сказуемое сводки. *Подлежащее* составляет вся совокупность группы или части, на которые разбивается совокупность. *Сказуемое* – это те показатели, которые характеризуют каждую группу, часть или всю совокупность в целом.

План сводки – содержит организационные вопросы.

<u>Статистическая группировка</u>

Статистическая группировка – это метод исследования массовых общественных явлений путем выделения и ограничения однородных групп, через которые раскрываются существенные черты и особенности состояния и развития всей совокупности.

<u>Основные задачи, которые решаются с помощью группировок</u>:

(1) выделение социально-экономических типов,

(2) изучение структуры социально-экономических явлений,

(3) выявление связи между явлениями.

<u>Важнейшие проблемы</u>:

(1) <u>Определение группировочного признака (основания группировки).</u>

Группировочный признак – это признак, по которому происходит определение единиц в группе. Его выбор зависит от цели группировки и существа данного явления.

(2) Выделение числа групп.

Число групп определяется с таким расчетом, чтобы в каждую группу попало достаточно большое число единиц.

(3) Интервалы

Интервалы могут быть равными и неравными. Последние в свою очередь делятся на равномерно возрастающие и равномерно убывающие.

Виды группировок

(1) Типологические группировки

Их задача – выявление социально-экономических типов или однородных в существенном отношении групп.

№ п/п	Социально-экономические типы	*Мужчины*		Женщины	
		1980	1992	1980	1992
1.	Работники	–	–	–	–
2.	Крестьяне	–	–	–	–
3.	Служащие	–	–	–	–

(2) Структурные группировки

Их задача – изучение состава отдельных типических групп при помощи объединения единиц совокупности, близких друг к другу по величине группировочного признака.

№ п/п	Количество посадочных мест	*Количество столов*	*Число занятых*	Товарооборот на 1 место
1.	до 25	–	–	–
2.	16 – 50	–	–	–
3.	51 – 70	–	–	–
4.	71 – 100	–	–	–

(3) Аналитические группировки

Их задача – выявления влияния одних признаков на другие (выявить связь между социально-экономическими явлениями).

№ п/п	Группы магазинов по числу рабочих мест	Число магазинов	Товарооборот	
			на 1 работника	на 1 раб. место
1.	до 5	100	12,0	13,0
2.	6 – 10	50	14,0	16,0
3.	11 – 15	10	15,0	17,0
4.	16 – 20	4	30,0	39,0
5.	21 – 25	2	31,0	42,0

(4) Комбинационные группировки

В них производится разделение совокупности на группы по двум или более признакам. При этом группы, образованные по одному признаку, разбиваются на подгруппы по другому признаку.

Такие группировки дают возможность изучить структуру совокупности по нескольким признакам одновременно.

№ п/п	Группы предприятий по объему основных фондов	*Оплата труда в рублях*	*Пол*	Количество единиц
1.	до 200	100 – 120	**М**	–
			Ж	–
		120 – 140	**М**	–
			Ж	–
		140 – 160	**М**	–
			Ж	–
2.	200 – 400	100 – 120	**М**	–
			Ж	–
		120 – 140	**М**	–
			Ж	–
		140 – 160	**М**	–
			Ж	–
3.	400 – 600	100 – 120	**М**	–
			Ж	–
		120 – 140	**М**	–
			Ж	–
		140 – 160	**М**	–
			Ж	–
4.	600 – 800	100 – 120	**М**	–
			Ж	–
		120 – 140	**М**	–
			Ж	–
		140 – 160	**М**	–
			Ж	–

Система группировок

Социально-экономический анализ предполагает использование системы простых и комбинационных группировок.

Также очень часто прибегают к вторичной группировке – перегруппировка уже сгруппированных данных. Вторичная группировка может быть проведена методом простого укрупнения интервала.

Часто также используется процентная перегруппировка.

Пример: **Группировка фермерских хозяйств по наличию скота**.

Исходные данные:

№ п/п	Группы хозяйств по числу голов	*% фермерских хозяйств*	*% поголовья*	% по всему кол-ву скота
1.	без голов	26,4	2,8	9,9
2.	с 1-й головой	20,3	9,5	8,9
3.	с 2-мя головами	14,6	11,8	11,1

4.	с 3-мя — " —	9,3	10,5	9,8
5.	с 4-мя — " —	8,3	12,1	11,2
6.	с 5-ю — " —	21,1	53,3	56,1
	Всего:	***100***	***100***	***100***

Процентная перегруппировка

№ п/п	**Группы хозяйств по уровню развития**	*% фермерских хозяйств*	*% поголовья*	**% по всему кол-ву скота**
1.	Низкий	50	**14,9**	**21,3**
2.	Средний	30	**34,6**	**32,5**
3.	Высокий	20	**50,5**	**53,2**
	Всего:	***100***	***100***	***100***

Расчеты:

1. 26,4 + 20,3 = 46,7
2. 50 – 46,7 = 3,3
3. 3,3 / 14,6 = *0,226*
4. 0,226 * 11,8 = 2,6   0,226 * 11,1 = 2,5
5. 2,8 + 9,5 + 2,6 = **14,9**   9,9 + 8,9 + 2,5 = **21,3**

6. 11,3 + 9,3 + 8,3 = 28,9
7. 30 – 28,9 = 1,1
8. 1,1 / 21,1 = *0,052*
9. 0,052 * 53,3 = 2,8   0,052 * 56,1 = 2,9
10. (11,8 – 2,6) + 10,5 + 12,1 + 2,8 = **34,6**   (11,1 – 2,5) + 9,8 + 11,2 + 2,9 = **32,5**

11. 53,3 – 2,8 = **50,5**   56,1 – 2,9 = **53,2**

Контрольные вопросы:

1. Что такое сводка? В чем её экономический смысл?
2. Что такое Группировочные признак? Какие виды вы знаете?
3. Дайте определение вариационного ряда распределения.
4. Какие элементы вариационного ряда? Название и определение.
5. Как графически можно отразить дискретный и интервальный вариационные ряды.

Абсолютные статистические величины

Абсолютные статистические величины показывают объем, размеры, уровни различных социально-экономических явлений и процессов. Они отражают уровни в физических мерах объема, веса и т.п. В общем, абсолютные статистические величины – это именованные числа. Они всегда имеют определенную размерность и единицы измерения. Последние определяют сущность абсолютной величины.

<u>Типы абсолютных величин</u>

(1) Натуральные – такие единицы, которые отражают величину предметов, вещей в физических мерах (вес, объем, площадь и т.д.).

(2) Денежные (стоимостные) – используются для характеристики многих экономических показателей в стоимостном выражении.

(3) Трудовые – используются для определения затрат труда (человеко-час, человеко-день)

(4) Условно-натуральные –единицы, которые используются для сведения воедино нескольких разновидностей потребительных стоимостей (т.у.т = 29,3 МДж/кг; мыло 40 % жирности).

<u>Виды абсолютных величин</u>

– Индивидуальные – отражают размеры количественных признаков у отдельных единиц изучаемой совокупности.

– Общие – выражают размеры, величину количественных признаков у всей изучаемой совокупности в целом.

Абсолютные величины отражают наличие тех или иных ресурсов, это основа материального учета. Они наиболее объективно отражают развитие экономики.

Абсолютные величины являются основой для расчета разных относительных статистических показателей.

Относительные статистические величины

Относительные статистические величины выражают количественные соотношения между явлениями общественной жизни, они получаются в результате деления одной абсолютной величины на другую.

Знаменатель (основание сравнения, база) – это величина, с которой производится сравнение.

Сравниваемая (отчетная, текущая) величина – это величина, которая сравнивается.

Относительная величина показывает, во сколько раз сравниваемая величина больше или меньше базисной или какую долю первая составляет по отношению ко второй. В ряде случае относительная величина показывает, сколько единиц одной величины приходится на единицу другой.

Важное свойство – относительная величина абстрагирует различия абсолютных величин и позволяет сравнивать такие явления, абсолютные размеры которых непосредственно несопоставимы.

<u>Форма выражения относительных величин</u>

В результате сопоставления одноименных абсолютных величин получают ***неименованные*** относительные величины. Они могут выражаться в виде долей, кратных соотношений, процентных соотношений, в виде промилле и т.д.

Результатом сопоставления разноименных величин являются именованные относительные величины. Их название образуется сочетанием сравниваемой и базисной абсолютных величин.

Выбор формы зависит от характера аналитической задачи, которая состоит в том, чтобы с наибольшей ясностью выразить соотношение.

<u>Виды относительных величин</u>

Все применяемые на практике относительные статистические величины подразделяются на следующие виды.

<u>Относительная величина динамики</u>

Статистика

Достигнутый показатель / базисный показатель.

Относительная величина планового задания

Плановый показатель / базисный показатель.

Относительная величина выполнения плана

Достигнутый показатель / плановый показатель.

Относительная величина структуры

Отношение частей и целого.

Относительная величина координации

Соотношение частей целого между собой.

Относительная величина интенсивности

Характеризует распределение явления в определенной среде (насыщенность каким-либо явлением). Это всегда соотношение разноименных величин.

Относительная величина уровня социально-экономического явления

Характеризует размеры производства различных видов продукции на душу населения.

Относительная величина сравнения

Представляет собой отношение одноименных величин, относящихся к различным объектам.

Контрольные вопросы:
1.Абсолютные величины. Область применения.
2.Условно-натуральные. Условные единицы измерения. Область применения.
3.Какие виды относительных величин Вы знаете?
4.Что означает понятие «структура совокупности»?

Большое распространение в статистике имеют средние величины. Средние величины характеризуют качественные показатели коммерческой деятельности: издержки обращения, прибыль, рентабельность и др.

Средняя - это один из распространенных приемов обобщений. Правильное понимание сущности средней определяет ее особую значимость в условиях рыночной экономики, когда средняя через единичное и случайное позволяет выявить общее и необходимое, выявить тенденцию закономерностей экономического развития.

Средняя величина - это обобщающие показатели, в которых находят выражение действия общих условий, закономерностей изучаемого явления.

Статистические средние рассчитываются на основе массовых данных правильно статистически организованного массового наблюдения (сплошного и выборочного). Однако статистическая средняя будет объективна и типична, если она рассчитывается по массовым данным для качественно однородной совокупности (массовых явлений). Например, если рассчитывать среднюю заработную плату в кооперативах и на госпредприятиях, а результат распространить на всю совокупность, то средняя фиктивна, так как рассчитана по неоднородной совокупности, и такая средняя теряет всякий смысл.

При помощи средней происходит как бы сглаживание различий в величине признака, которые возникают по тем или иным причинам у отдельных единиц наблюдения.

Например, средняя выработка продавца зависит от многих причин: квалификации, стажа, возраста, формы обслуживания, здоровья и т.д.

Средняя выработка отражает общее свойство всей совокупности.

Средняя величина является отражением значений изучаемого признака, следовательно, измеряется в той же размерности, что и этот признак.

Каждая средняя величина характеризует изучаемую совокупность по какому-либо одному признаку. Чтобы получить полное и всестороннее представление об изучаемой совокупности по ряду существенных признаков, в целом необходимо располагать системой средних величин, которые могут описать явление с разных сторон.

Существуют различные средние:
* средняя арифметическая;
* средняя геометрическая;
* средняя гармоническая;
* средняя квадратическая;
* средняя хронологическая.

Рассмотрим некоторые виды средних, которые наиболее часто используются в статистике.

Средняя арифметическая

Средняя арифметическая простая (невзвешенная) равна сумме отдельных значений признака, деленной на число этих значений.

Отдельные значения признака называют вариантами и обозначают через x ($x_1, x_2, x_3, \ldots, x_n$); число единиц совокупности обозначают через n, среднее значение признака - через $\bar{x}$. Следовательно, средняя арифметическая простая равна:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum x}{n}$$

По данным дискретного ряда распределения видно, что одни и те же значения признака (варианты) повторяются несколько раз. Так, варианта x встречается в совокупности 2 раза, а варианта x-16 раз и т.д.

Число одинаковых значений признака в рядах распределения называется частотой или весом и обозначается символом n.

Вычислим среднюю заработную плату одного рабочего $\bar{x}$ в руб.:

$$\bar{x} = \frac{110*2 + 130*6 + 160*16 + 190*12 + 220*14}{50} = \frac{8920}{50} = 178{,}4$$

Фонд заработной платы по каждой группе рабочих равен произведению варианты на частоту, а сумма этих произведений дает общий фонд заработной платы всех рабочих.

В соответствии с этим, расчеты можно представить в общем виде:

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + x_3 n_3 + \ldots + x_n n_n}{n_1 + n_2 + n_3 + \ldots + n_n}$$

$$\bar{x} = \frac{\sum x_i n_i}{\sum n_i}$$

Полученная формула называется средней арифметической взвешенной.

Статистический материал в результате обработки может быть представлен не только в виде дискретных рядов распределения, но и в виде интервальных вариационных рядов с закрытыми или открытыми интервалами.

Исчисление средней по сгруппированным данным производится по формуле средней арифметической взвешенной:

В практике экономической статистики иногда приходится исчислять среднюю по групповым средним или по средним отдельных частей совокупности (частным средним). В таких случаях за варианты (x) принимаются групповые или частные средние, на основании которых исчисляется общая средняя как обычная средняя арифметическая взвешенная.

Основные свойства средней арифметической.

Средняя арифметическая обладает рядом свойств:

1. От уменьшения или увеличения частот каждого значения признака x в n раз величина средней

арифметической не изменится.

Если все частоты разделить или умножить на какое-либо число, то величина средней не изменится.

2. Общий множитель индивидуальных значений признака может быть вынесен за знак средней:

$\overline{Kx} = K\overline{x}$

3. Средняя суммы (разности) двух или нескольких величин равна сумме (разности) их средних:

$\overline{x \pm y} = \overline{x} \pm \overline{y}$

4. Если x = c, где c - постоянная величина, то $\overline{x} = \overline{c} = c$.

5. Сумма отклонений значений признака X от средней арифметической x равна нулю:

$$\sum \left| \chi - \overline{\chi} \right| = 0$$

Средняя гармоническая.

Наряду со средней арифметической, в статистике применяется средняя гармоническая величина, обратная средней арифметической из обратных значений признака. Как и средняя арифметическая, она может быть простой и взвешенной.

Характеристиками вариационных рядов, наряду со средними, являются мода и медиана.

Мода - это величина признака (варианта), наиболее часто повторяющаяся в изучаемой совокупности. Для дискретных рядов распределения модой будет значение варианта с наибольшей частотой.

Для интервальных рядов распределения с равными интервалами мода определяется по формуле:

$$Mo = x_{Mo} + i_{Mo} * \frac{f_{Mo} - f_{Mo-1}}{\left(f_{Mo} - f_{Mo-1}\right) + \left(f_{Mo} - f_{Mo+1}\right)}.$$

где x_{Mo} - начальное значение интервала, содержащего моду;

i_{Mo} - величина модального интервала;

f_{Mo} - частота модального интервала;

f_{Mo-1} - частота интервала, предшествующего модальному;

f_{Mo+1} - частота интервала, следующего за модальным.

Медиана - это варианта, расположенная в середине вариационного ряда. Если ряд распределения дискретный и имеет нечетное число членов, то медианой будет варианта, находящаяся в середине упорядоченного ряда (упорядоченный ряд - это расположение единиц совокупности в возрастающем или убывающем порядке).

Контрольные вопросы:

1.Какие виды средних величин применяются в экономических исследования данных, представленных рядами распределения?
2.Как исчисляются средние арифметические: простая и взвешенная?
3.В каких случаях применяется средняя гармоническая?
4.Для каких целей рассчитываются показатели моды и медианы?
5.Как определяются мода и медиана для дискретных и интервальных вариационных рядов распределения.
6.Каким образом графически определить моду и медиану.

Различие индивидуальных значений признака внутри изучаемой совокупности в статистике называется **вариацией признака**.

Она возникает в результате того, что его индивидуальные значения складываются под совокупным влиянием разнообразных факторов, которые по-разному сочетаются в каждом отдельном случае.

Средняя величина — это абстрактная, обобщающая характеристика признака изучаемой совокупности, но она не показывает строения совокупности, которое весьма существенно для ее познания. Средняя величина не дает представления о том, как отдельные значения изучаемого признака группируются вокруг средней, сосредоточены ли они вблизи или значительно отклоняются от нее. В некоторых случаях отдельные значения признака близко примыкают к средней арифметической и мало от нее отличаются. В таких случаях средняя хорошо представляет всю совокупность.

В других, наоборот, отдельные значения совокупности далеко отстают от средней, и средняя плохо представляет всю совокупность.

Колеблемость отдельных значений характеризуют показатели вариации.

Термин "вариация" произошел от латинского variatio –"изменение, колеблемость, различие". Однако не всякие различия принято называть вариацией. Под вариацией в статистике понимают такие количественные изменения величины исследуемого признака в пределах однородной совокупности, которые обусловлены перекрещивающимся влиянием действия различных факторов. Различают вариацию признака: случайную и систематическую.

Анализ систематической вариации позволяет оценить степень зависимости изменений в изучаемом признаке от определяющих ее факторов. Например, изучая силу и характер вариации в выделяемой совокупности, можно оценить, насколько однородной является данная совокупность в количественном, а иногда и качественном отношении, а следовательно, насколько характерной является исчисленная средняя величина. Степень близости данных отдельных единиц xi к средней измеряется рядом абсолютных, средних и относительных показателей.

Для характеристики совокупностей и исчисленных величин важно знать, какая вариация изучаемого признака скрывается за средним.

Для характеристики колеблемости признака используется ряд показателей. Наиболее простой из них - размах вариации.

<u>Размах вариации</u> - это разность между наибольшим (x_{mak}) и наименьшим ($x_{\min}$) значениями вариантов.

$$R = x_{mak} - x_{min}$$

Чтобы дать обобщающую характеристику распределению отклонений, исчисляют среднее линейное отклонение d, которое учитывает различие всех единиц изучаемой совокупности.

Среднее линейное отклонение определяется как средняя арифметическая из отклонений индивидуальных значений от средней, без учета знака этих отклонений:

$$d = \frac{\sum / x - \bar{x} /}{n} = \frac{/ x_1 - \bar{x} / + / x_2 - \bar{x} / + ... + / x_n - \bar{x} /}{n}.$$

Порядок расчета среднего линейного отклонения следующий:

1) по значениям признака исчисляется средняя арифметическая:

$$\bar{x} = \frac{\sum x}{n};$$

2) определяются отклонения каждой варианты x_i от средней $x_i - \bar{x}$;

3) рассчитывается сумма абсолютных величин отклонений: $\sum / x_i - \bar{x} /$;

4) сумма абсолютных величин отклонений делится на число значений:

$$\frac{\sum / x_i - \bar{x} /}{n}.$$

Если данные наблюдения представлены в виде дискретного ряда распределения с частотами, среднее линейное отклонение исчисляется по формуле средней арифметической взвешенной:

Статистика

$$d = \frac{\sum / x_i - \bar{x} / n_i}{\sum n_i} = \frac{/ x_1 - \bar{x} / n_1 + / x_2 - \bar{x} / n_2 + \ldots + / x_n - \bar{x} / n_n}{n_1 + n_2 + \ldots + n_n}$$

Порядок расчета среднего линейного отклонения взвешенного следующий:

1) вычисляется средняя арифметическая взвешенная:

$$\frac{\sum xn}{\sum n};$$

2) определяются абсолютные отклонения вариант от средней / $x_i - \bar{x}$ /;

3) полученные отклонения умножаются на частоты $/ x_i - \bar{x} / n_i$;

4) находится сумма взвешенных отклонений без учета знака:

$$\sum / x_i - \bar{x} / n_i;$$

5) сумма взвешенных отклонений делится на сумму частот:

$$\frac{\sum / x_i - \bar{x} / n_i}{\sum n}.$$

Основными обобщающими показателями вариации в статистике являются дисперсии и среднее квадратическое отклонение.

Дисперсия - это средняя арифметическая квадратов отклонений каждого значения признака от общей средней. Дисперсия обычно называется средним квадратом отклонений и обозначается S^2. В зависимости от исходных данных дисперсия может вычисляться по средней арифметической простой или взвешенной:

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{\sum n}$$ — дисперсия невзвешенная (простая);

$$S^2 = \frac{\sum (x_i - \bar{x})^2 n_i}{\sum n_i}$$ — дисперсия взвешенная.

Среднее квадратическое отклонение представляет собой корень квадратный из дисперсии и обозначается S:

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$ — среднее квадратическое отклонение невзвешенное;

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2 n_i}{\sum n_i}}$$ — среднее квадратическое отклонение взвешенное.

Среднее квадратическое отклонение - это обобщающая характеристика абсолютных размеров вариации признака в совокупности. Выражается оно в тех же единицах измерения, что и признак (в метрах, тоннах, процентах, гектарах и т.д.).

Среднее квадратическое отклонение является мерилом надежности средней. Чем меньше среднее квадратическое отклонение, тем лучше средняя арифметическая отражает собой всю представляемую совокупность.

Вычислению среднего квадратического отклонения предшествует расчет дисперсии.

Порядок расчета дисперсии взвешенную:

1) определяют среднюю арифметическую взвешенную

$$\frac{\sum xn}{\sum n};$$

2) определяются отклонения вариант от средней $\left(x_i - \bar{x}\right)$;

3) возводят в квадрат отклонение каждой варианты от средней $\left(x_i - \overline{x}\right)^2$;

4) умножают квадраты отклонений на веса (частоты) $\left(x_i - \overline{x}\right)^2 n_i$;
5) суммируют полученные произведения
$\sum\left(x_i - \overline{x}\right)^2 n_i$;
6) Полученную сумму делят на сумму весов
$$\frac{\sum\left(x_i - \overline{x}\right)^2 n_i}{\sum n_i}.$$

Свойства дисперсии.

Уменьшение или увеличение весов (частот) варьирующего признака в определенное число раз дисперсии не изменяет.

Уменьшение или увеличение каждого значения признака на одну и ту же постоянную величину А дисперсии не изменяет.

Уменьшение или увеличение каждого значения признака в какое-то число раз к соответственно уменьшает или увеличивает дисперсию в k^2 раз, а среднее квадратическое отклонение - в к раз.

Дисперсия признака относительно произвольной величины всегда больше дисперсии относительно средней арифметической на квадрат разности между средней и произвольной величиной: $S^2 = S_A^2 - (\overline{x} - A)^2$. Если А равна нулю, то приходим к следующему равенству: $S^2 = \overline{x^2} - \overline{x}^2$, т.е. дисперсия признака равна разности между средним квадратом значений признака и квадратом средней.

Каждое свойство при расчете дисперсии может быть применено самостоятельно или в сочетании с другими.

Порядок расчета дисперсии простой:

1) определяют среднюю арифметическую $\overline{x} = \frac{\sum x}{n}$;

2) возводят в квадрат среднюю арифметическую $\overline{x}^2 = \left(\frac{\sum x}{n}\right)^2$;

3) возводят в квадрат каждую варианту ряда x_i^2;

4) находим сумму квадратов вариант $\sum x_i^2$;

5) делят сумму квадратов вариант на их число, т.е. определяют средний квадрат $\overline{x^2} = \frac{\sum x_i^2}{n}$;

6) определяют разность между средним квадратом признака и квадратом средней $\overline{x^2} - \overline{x}^2$.

Порядок расчета дисперсии взвешенной (по формуле $S^2 = \overline{x^2} - \overline{x}^2$):

1) определяют среднюю арифметическую $\overline{x} = \frac{\sum xn}{\sum n}$;

2) возводят в квадрат полученную среднюю $\left(\overline{x}\right)^2$;

3) возводят в квадрат каждую варианту ряда x_i^2;

4) умножают квадраты вариант на частоты $x_i^2 n_i$;

5) суммируют полученные произведения $\sum x_i^2 n_i$;

6) делят полученную сумму на сумму весов и получают средний квадрат признака $\overline{x^2} = \frac{\sum x_i^2 n_i}{\sum n_i}$;

7) определяют разность между средним значением квадратов и квадратом средней арифметической, т. е. дисперсию $S^2 = \overline{x^2} - \bar{x}^2$.

Для характеристики меры колеблемости изучаемого признака исчисляются показатели колеблемости в относительных величинах. Они позволяют сравнивать характер рассеивания в различных распределениях (различные единицы наблюдения одного и того же признака в двух совокупностях, при различных значениях средних, при сравнении разноименных совокупностей). Расчет показателей меры относительного рассеивания осуществляют как отношение абсолютного показателя рассеивания к средней арифметической, умножаемое на 100%.

1. **Коэффициент осцилляции** отражает относительную колеблемость крайних значений признака вокруг средней.

$$K_o = \frac{R}{\bar{x}} * 100\% \qquad (1)$$

2. **Относительное линейное отклонение** характеризует долю усредненного значения абсолютных отклонений от средней величины.

$$K_o = \frac{\bar{d}}{\bar{x}} * 100\% \qquad (2)$$

3. **Коэффициент вариации**.

$$V = \frac{S}{\bar{x}} * 100\% \qquad (3)$$

Учитывая, что среднеквадратическое отклонение дает обобщающую характеристику колеблемости всех вариантов совокупности, коэффициент вариации является наиболее распространенным показателем колеблемости, используемым для оценки типичности средних величин. При этом исходят из того, что если V больше 40 %, то это говорит о большой колеблемости признака в изучаемой совокупности.

Контрольные вопросы.

1.Для каких целей рассчитываются показатели моды и медианы?

2.Как определяются мода и медиана для дискретных и интервальных вариационных рядов распределения.

3.Каким образом графически определить моду и медиану.

.

Контрольные вопросы:

1.Что означает показатель "колеблемость" применительно к вариационному ряду распределения?

2.Как подразделяются показатели вариации?

3.Что такое размах вариации, дисперсия, среднее квадратическое отклонение?

4.Что такое относительные показатели вариации и как они рассчитываются?

5.Какие способы расчёта дисперсии Вы знаете?

6.Как формулируется правило сложения дисперсий?

7.Какие показатели вариации с Вашей точки зрения являются наиболее существенными при экспресс-анализе?

Ряд распределения или вариационный ряд – упорядоченное распределение единиц совокупности по возрастающим или по убывающим значениям признака и подсчет единиц с тем или иным значением признака. Построение рядов распределения (структурной группировки) является первым этапом изучения вариации и осуществляется с целью выделения характерных свойств и закономерностей изучаемой совокупности. В зависимости от того, какой признак (количественный или качественный) взят за основу группировки данных, различают типы рядов распределения.

Если за основу группировки взят качественный признак, то такой ряд распределения называют атрибутивным (распределение по видам труда, по полу, по профессии, по религиозному признаку, национальной принадлежности и т.д.).

Если ряд распределения построен по количественному признаку, то такой ряд называют вариационным. Построить вариационный ряд - значит упорядочить количественное распределение единиц совокупности по значениям признака, а затем подсчитать числа единиц совокупности с этими значениями (построить групповую таблицу).

Выделяют три формы вариационного ряда: ранжированный ряд, дискретный ряд и интервальный ряд.

Ранжированный ряд - это распределение отдельных единиц совокупности в порядке возрастания или убывания исследуемого признака.

Другие формы вариационного ряда - групповые таблицы, составленные по характеру вариации значений изучаемого признака. По характеру вариации различают дискретные (прерывные) и непрерывные признаки.

Дискретный ряд - это такой вариационный ряд, в основу построения которого положены признаки с прерывным изменением (дискретные признаки). К последним можно отнести тарифный разряд, количество детей в семье, число работников на предприятии и т.д. Эти признаки могут принимать только конечное число определенных значений.

Если признак имеет непрерывное изменение (размер дохода, стаж работы, стоимость основных фондов предприятия и т.д., которые в определенных границах могут принимать любые значения), то для этого признака нужно строить интервальный вариационный ряд.

Величина интервала определяется по формуле $i_x = \frac{x_{max} - x_{min}}{к}$, где $x_{max,\,min}$ - максимальное и минимальное значение признака, к – число групп.

Частота (частота повторения) - число повторений отдельного варианта значений признака, обозначается f_i, а сумма частот, равная объему исследуемой совокупности, обозначается $\sum_{i=1}^{k} f_i$, где к – число вариантов значения признака.

Частоты ряда f могут заменяться частостями w, выраженными в относительных числах (долях или процентах). Они представляют собой отношения частот каждого интервала к их общей сумме, т.е.: $w_i = \frac{f_i}{\sum f_i}$, при этом $\sum w_i = 1$

Основной целью анализа вариационных рядов является выявление закономерности распределения, исключая при этом влияние случайных для данного распределения факторов. Этого можно достичь, если увеличивать объем исследуемой совокупности и одновременно уменьшать интервал ряда.

В практике статистических исследований наиболее часто используются следующие закономерности распределения: нормальное распределение и распределение Пуассона.

Нормальное распределение зависит от двух параметров: средней арифметической и среднего квадратического отклонения. Его кривая выражается уравнением $y = \frac{1}{2\pi} e^{-\frac{t^2}{2}}$

где y - ордината кривой нормального распределения; $t = (x - \bar{x})/\sigma$ - стандартизованные отклонения; е и π - математические постоянные; х - варианты вариационного ряда; $\bar{x}$ - их средняя величина; σ - среднее квадратическое отклонение.

Теоретические частоты при нормальном распределении определяются по формуле: $f' = \frac{Nh}{\sigma}\frac{1}{2\pi}e^{-\frac{t^2}{2}}$, где N = Σf – сумма всех эмпирических частот вариационного ряда; h – величина интервала в группах.

При помощи этой формулы мы получаем теоретическое (вероятностное) распределение, заменяя им эмпирическое (фактическое) распределение, по характеру они не должны отличаться друг от друга.

Если вариационный ряд представляет собой распределение по дискретному признаку, где при увеличении значений признака х частоты начинают резко уменьшаться, а средняя арифметическая, в свою очередь, равна или близка по значению к дисперсии ($\bar{x} = \sigma^2$), такой ряд выравнивается по кривой Пуассона.

Кривую Пуассона можно выразить отношением $P_x = \frac{a^x e^{-a}}{x!}$, где P_x - вероятность наступления отдельных значений х; $a = \bar{x}$ - средняя арифметическая ряда.

Теоретические частоты при распределении Пуассона определяют по формуле: $f' = N P_x$, где N – общее число единиц ряда.

Для расчета обобщающих показателей и для графического изображения вариационных рядов с неравными интервалами используют плотность распределения, которая определяется по формулам:

$$f'_j = \frac{f_j}{i_j} \text{ или } w'_j = \frac{w_j}{i_j},$$

где f'_j - абсолютная плотность распределения в j-м интервале, w'_j - относительная плотность распределения в j-м интервале; i_j – величина интервала.

Объективная характеристика соответствия теоретических и эмпирических частот может быть получена при помощи специальных статистических показателей, которые называют критериями согласия.

Асимметрия распределения определяется на основе расчета коэффициента асимметрии, котрый является мерой несимметричности распределения. Если этот коэффициент отчетливо отличается от 0, распределение является асимметричным. Плотность нормального распределения симметрична относительно среднего.

Для оценки близости эмпирических и теоретических частот применяются критерий согласия Пирсона, критерий согласия Романовского, критерий согласия Колмогорова.

Наиболее распространенным является критерий согласия К. Пирсона, который можно представить как сумму отношений квадратов расхождений между f' и f к теоретическим частотам: $\chi^2 = \sum \frac{(f - f')^2}{f'}$

Вычисленное значение критерия $\chi^2_{расч}$ необходимо сравнить с табличным (критическим) значением $\chi^2_{табл}$. Табличное значение определяется по специальной таблице, оно зависит от принятой вероятности Р и числа степеней свободы k (при этом k = m - 3, где m - число групп в ряду распределения для нормального распределения). При расчете критерия согласия Пирсона должно соблюдаться следующее условие: достаточно большим должно быть число наблюдений (n   50), при этом если в некоторых интервалах теоретические частоты меньше 5, то интервалы объединяют для условия больше 5.

Если $\chi^2_{расч}$ $\chi^2_{табл}$, то расхождения между эмпирическими и теоретическими частотами распределения могут быть случайными и предположение о близости эмпирического распределения к нормальному не может быть отвергнуто.

Статистика

В том случае, если отсутствуют таблицы для оценки случайности расхождения теоретических и эмпирических частот, можно использовать критерий согласия В.И. Романовского (*$K_{Ром}$)*, который, используя величину χ^2, предложил оценивать близость эмпирического распределения кривой нормального распределения при помощи отношения: $K_{Ром} = \frac{\chi^2 - к}{\sqrt{2к}}$, где m - число групп; k = (m - 3) - число степеней свободы при исчислении частот нормального распределения.

Если вышеуказанное отношение < 3, то расхождения эмпирических и теоретических частот можно считать случайными, а эмпирическое распределение - соответствующим нормальному. Если отношение > 3, то расхождения могут быть достаточно существенными и гипотезу о нормальном распределении следует отвергнуть.

Критерий согласия А.Н. Колмогорова используется при определении максимального расхождения между частотами эмпирического и теоретического распределения, вычисляется по формуле: $\lambda = \frac{D}{\sqrt{\sum f}}$, где D - максимальное значение разности между накопленными эмпирическими и теоретическими частотами; $\sum f$ - сумма эмпирических частот.

По таблицам значений вероятностей λ-критерия можно найти величину λ, соответствующую вероятности P. Если величина вероятности P значительна по отношению к найденной величине , то можно предположить, что расхождения между теоретическим и эмпирическим распределениями несущественны.

Необходимым условием при использовании критерия согласия Колмогорова является достаточно большое число наблюдений (не меньше ста).

При анализе вариационного ряда и его свойств используют графические методы. Интервальный ряд изображаю столбиковой диаграммой или гистограммой, в которой основания столбиков, расположенные на оси – абсцисс – это интервалы значений варьирующего признака, а высоты столбиков – частоты.

Если имеется дискретный вариационный ряд или используются середины интервалов, то графическое изображение такого ряда называют полигоном.

Преобразованной формой вариационного ряда является ряд накопленных частот. Это ряд значений числа единиц совокупности с меньшими или равными нижней границе соответствующего интервала значениями признака. Такой ряд называют кумулятивным. Можно построить кумулятивное распределение «не меньше, чем» – кумулята, и «больше, чем» – огива.

Контрольные вопросы:

1. Понятие о закономерностях распределения и их формы
2. Структурные характеристики вариационного ряда распределения
3. Понятие и классификация рядов динамики
4. Сопоставимость уровней ряда и смыкание рядов динамики
5. Показатели изменения уровней ряда динамики
6. Элементы прогнозирования и интерполяции

<u>Основы выборочного метода</u>

Выборочное наблюдение – одно из наиболее современных видов статистического наблюдения. Выборочное наблюдение – это такое наблюдение, при котором обследованию подвергается часть единиц изучаемой совокупности, отобранных на основе научно разработанных принципов, обеспечивающих получение достаточного количества достоверных данных, для того чтобы охарактеризовать всю совокупность в целом.

Средние и относительные показатели, полученные на основе выборочных данных, должны достаточно полно воспроизводить или репрезентатировать соответствующие показатели совокупности в целом.

Логика выборочного наблюдения

(1) определение объекта и целей выборочного наблюдения;

(2) выбор схема отбора единиц для наблюдения;

(3) расчет объема выборки;

(4) проведение случайного отбора установленного числа единиц из генеральной совокупности;

(5) наблюдение отобранных единиц по установленной программе;

(6) расчет выборочных характеристик в соответствии с программой выборочного наблюдения;

(7) определение ошибки, ее размера;

(8) распространение выборочных данных на генеральную совокупность;

(9) анализ полученных данных.

Основные преимущества

(1) Выборочное наблюдение можно осуществить по более широкой программе.

(2) Выборочное наблюдение более дешевое с точки зрения затрат на его проведение.

(3) Выборочное наблюдение можно организовать тогда и в тех случаях, когда отчетностью мы воспользоваться не можем.

Основные недостатки

(1) Полученные данные всегда содержат в себе ошибку, о результатах наблюдения можно судить лишь с определенной степенью достоверности. Но по сравнению с другими видами наблюдения это достоинство выборочного метода.

(2) Для его проведения требуются квалифицированные кадры.

Вся совокупность единиц, из которых производится отбор, называется генеральной. Совокупность единиц отобранных называется выборочной.

Ошибки выборки

Чтобы оценить степень точности выборочного наблюдения, необходимо оценить величину ошибок, которые могут возникнуть в процессе проведения выборочного наблюдения.

Основное внимание уделяется случайным ошибкам репрезентативности.

Статистическое исследование может осуществляться по данным несплошного наблюдения, основная цель которого состоит в получении характеристик изучаемой совокупности по обследованной ее части. Одним из наиболее распространенных в статистике методов, применяющих несплошное наблюдение, является **выборочный метод.**

Под выборочным понимается метод статистического исследования, при котором обобщающие показатели изучаемой совокупности устанавливаются по некоторой ее части на основе положений случайного отбора. При выборочном методе обследованию подвергается сравнительно небольшая часть всей изучаемой совокупности (обычно до 5 — 10%, реже до 15 — 25%). При этом подлежащая изучению статистическая совокупность, из которой производится отбор части единиц, называется **генеральной совокупностью.** Отобранная из генеральной совокупности некоторая часть единиц, подвергающаяся обследованию, называется **выборочной совокупностью** или просто **выборкой.**

Значение выборочного метода состоит в том, что при минимальной численности обследуемых единиц проведение исследования осуществляется в более короткие сроки и с минимальными затратами труда и средств. Это повышает оперативность статистической информации, уменьшает ошибки регистрации.

Статистика

В проведении ряда исследований выборочный метод является единственно возможным, например, при контроле качества продукции (товара), если проверка сопровождается уничтожением или разложением на составные части обследуемых образцов (определение сахаристости фруктов, клейковины печеного хлеба, установление носкости обуви, прочности тканей на разрыв и т.д.).

Проведение исследования социально — экономических явлений выборочным методом складывается из ряда последовательных этапов:

1) обоснование (в соответствии с задачами исследования) целесообразности применения выборочного метода;

2) составление программы проведения статистического исследования выборочным методом;

3) решение организационных вопросов сбора и обработки исходной информации;

4) установление доли выборки, т.е. части подлежащих обследованию единиц генеральной совокупности;

5) обоснование способов формирования выборочной совокупности;

6) осуществление отбора единиц из генеральной совокупности для их обследования;

7) фиксация в отобранных единицах (пробах) изучаемых признаков;

8) статистическая обработка полученной в выборке информации с определением обобщающих характеристик изучаемых признаков;

9) определение количественной оценки ошибки выборки;

10) распространение обобщающих выборочных характеристик на генеральную совокупность.

В генеральной совокупности доля единиц, обладающих изучаемым признаком, называется **генеральной долей** (обозначается р), а средняя величина изучаемого варьирующего признака — **генеральной средней** (обозначается $\overline{x}$).

В выборочной совокупности долю изучаемого признака называют выборочной долей, или частостью (обозначается ω), а среднюю величину в выборке — выборочной средней (обозначается $\tilde{x}$).

Пример.

При контрольной проверке качества хлебобулочных изделий проведено 5%-ное выборочное обследование партии нарезных батонов из муки высшего сорта. При этом из 100 отобранных в выборку батонов 90 шт. соответствовали требованиям стандарта. Средний вес одного батона в выборке составлял 500,5 г при среднем квадратическом отклонении $\pm 15{,}4$ г.

На основе полученных в выборке данных нужно установить возможные значения доли стандартных изделий и среднего веса одного изделия во всей партии.

Прежде всего устанавливаются характеристики выборочной совокупности. Выборочная доля, или частость, ω определяется из отношения единиц, обладающих изучаемым признаком m, к общей численности единиц выборочной совокупности n:

$$\omega = \frac{m}{n}$$

Поскольку из 100 изделий, попавших в выборку n, 90 ед. оказались стандартными m, то показатель частости равен: ω = 90:100=0,9.

Средний вес изделия в выборке x = 500,5 г определен взвешиванием. Но полученные показатели частости (0,9) и средней величины (500,5 г) характеризуют долю стандартной продукции и средний вес одного изделия лишь в выборке. Для определения соответствующих показателей для всей партии товара надо установить возможные при этом значения ошибки выборки.

Ошибка выборки — это объективно возникающее расхождение между характеристиками выборки и генеральной совокупности. Она зависит от ряда факторов: степени вариации изучаемого признака, численности выборки, методом отбора единиц в выборочную совокупность, принятого уровня достоверности результата исследования.

Определение ошибки выборочной средней.

При случайном повторном отборе **средняя ошибка** выборочной средней рассчитывается по формуле:

$$\mu = \sqrt{\frac{s^2}{n}},$$

где μ — средняя ошибка выборочной средней;

s^2 — дисперсия выборочной совокупности;

n — численность выборки.

При бесповторном отборе она рассчитывается по формуле:

$$\mu = \sqrt{\frac{s^2}{n}\left(1-\frac{n}{N}\right)},$$

где N — численность генеральной совокупности.

Определение ошибки выборочной доли.

При повторном отборе средняя ошибка выборочной доли рассчитывается по формуле:

$$\mu = \sqrt{\frac{\varpi(1-\varpi)}{n}},$$

где $\omega = \frac{m}{n}$ — выборочная доля единиц, обладающих изучаемым признаком;

m — число единиц, обладающих изучаемым признаком;

n — численность выборки.

При бесповторном способе отбора средняя ошибка выборочной доли определяется по формулам:

$$\mu = \sqrt{\frac{\omega(1-\varpi)}{n}\left(1-\frac{n}{N}\right)}$$

Предельная ошибка выборки Δ связана со средней ошибкой выборки μ отношением:

$\Delta = t * \mu$.

При этом t как коэффициент кратности средней ошибки выборки зависит от значения вероятности P, с которой гарантируется величина предельной ошибки выборки.

Предельная ошибка выборки при бесповторном отборе определяется по следующим формулам:

$$\Delta_\omega = t\sqrt{\frac{\omega(1-\varpi)}{n}\left(1-\frac{n}{N}\right)},$$

$$\Delta_x = t\sqrt{\frac{s_x^2}{n}\left(1-\frac{n}{N}\right)}.$$

Предельная ошибка выборки при повторном отборе определяется по формуле:

$$\Delta_\omega = t\sqrt{\frac{\omega(1-\varpi)}{n}},$$

$$\Delta_x = t\sqrt{\frac{s_x^2}{n}}.$$

Малая выборка.

При контроле качества товаров в экономических исследованиях эксперимент может проводиться на основе малой выборки.

Под **малой выборкой** понимается несплошное статистическое обследование, при котором выборочная совокупность образуется из сравнительно небольшого числа единиц генеральной совокупности. Объем малой выборки обычно не превышает 30 единиц и может доходить до 4 — 5 единиц.

Средняя ошибка малой выборки $\mu_{M.B}$ вычисляется по формуле:

$$\mu_{M.B} \approx \sqrt{\frac{s^2_{M.B}}{n}},$$

где $s^2_{M.B}$ — дисперсия малой выборки.

При определении дисперсии s^2 число степеней свободы равно n-1:

$$s^2_{M.B} = \frac{\sum (x_i - \tilde{x})^2}{n-1}.$$

Предельная ошибка малой выборки $\Delta_{M.B}$ определяется по формуле $\Delta_{M.B} = t\mu_{M.B}$

При этом значение коэффициента доверия t зависит не только от заданной доверительной вероятности, но и от численности единиц выборки n. Для отдельных значений t и n доверительная вероятность малой выборки определяется по специальным таблицам Стьюдента (Табл. 9.1.), в которых даны распределения стандартизированных отклонений:

$$t = \frac{\tilde{x} - \bar{x}}{s_{M.B}}.$$

Поскольку при проведении малой выборки в качестве доверительной вероятности практически принимается значение 0,59 или 0,99, то для определения предельной ошибки малой выборки $\Delta_{M.B}$ используются следующие показания распределения Стьюдента:

n	S_t	
	0,95	0,99
4	3,183	5,841
5	2,777	4,604
6	2,571	4,032
7	2,447	3,707
8	2,364	3,500
9	2,307	3,356
10	2,263	3,250
15	2,119	2,921
20	2,078	2,832

Способы распространения характеристик выборки на генеральную совокупность.

Выборочный метод чаще всего применяется для получения характеристик генеральной совокупности по соответствующим показателям выборки. В зависимости от целей исследований это осуществляется или прямым пересчётом показателей выборки для генеральной совокупности, или посредством расчёта поправочных коэффициентов.

Способ прямого пересчёта. Он состоит в том, что показатели выборочной доли ω или средней $\tilde{x}$ распространяется на генеральную совокупность с учётом ошибки выборки.

Так, в торговле определяется количество поступивших в партии товара нестандартных изделий. Для этого (с учётом принятой степени вероятности) показатели доли нестандартных изделий в выборке умножаются на численность изделий во всей партии товара.

Способ поправочных коэффициентов. Применяется в случаях, когда целью выборочного метода является уточнение результатов сплошного учета.

В статистической практике этот способ используется при уточнении данных ежегодных переписей скота, находящегося у населения. Для этого после обобщения данных сплошного учета практикуется

10%-ное выборочное обследование с определением так называемого “процента недоучета”.

Способы отбора единиц из генеральной совокупности.

В статистике применяются различные способы формирования выборочных совокупностей, что обусловливается задачами исследования и зависит от специфики объекта изучения.

Основным условием проведения выборочного обследования является предупреждение возникновения систематических ошибок, возникающих вследствие нарушения принципа равных возможностей попадания в выборку каждой единицы генеральной совокупности. Предупреждение систематических ошибок достигается в результате применения научно обоснованных способов формирования выборочной совокупности.

Существуют следующие способы отбора единиц из генеральной совокупности:

1) индивидуальный отбор — в выборку отбираются отдельные единицы;

2) групповой отбор — в выборку попадают качественно однородные группы или серии изучаемых единиц;

3) комбинированный отбор — это комбинация индивидуального и группового отбора.

Способы отбора определяются правилами формирования выборочной совокупности.

Выборка может быть:

— собственно-случайная;

— механическая;

— типическая;

— серийная;

— комбинированная.

Собственно-случайная выборка состоит в том, что выборочная совокупность образуется в результате случайного (непреднамеренного) отбора отдельных единиц из генеральной совокупности. При этом количество отобранных в выборочную совокупность единиц обычно определяется исходя из принятой доли выборки.

Доля выборки есть отношение числа единиц выборочной совокупности n к численности единиц генеральной совокупности N, т.е.

$$K_B = \frac{n}{N}.$$

Так, при 5%-ной выборке из партии товара в 2 000 ед. численность выборки n составляет 100 ед. (5*2000:100), а при 20%-ной выборке она составит 400 ед. (20*2000:100) и т.д.

Механическая выборка состоит в том, что отбор единиц в выборочную совокупность производится из генеральной совокупности, разбитой на равные интервалы (группы). При этом размер интервала в генеральной совокупности равен обратной величине доли выборки.

Так, при 2%-ной выборке отбирается каждая 50-я единица (1:0,02), при 5%-ной выборке — каждая 20-я единица (1:0,05) и т.д.

Таким образом, в соответствии с принятой долей отбора, генеральная совокупность как бы механически разбивается на равновеликие группы. Из каждой группы в выборку отбирается лишь одна единица.

Важной особенностью механической выборки является то, что формирование выборочной совокупности можно осуществить, не прибегая к составлению списков. На практике часто используют тот порядок, в котором фактически размещаются единицы генеральной совокупности. Например, последовательность выхода готовых изделий с конвейера или поточной линии, порядок размещения единиц партии товара при хранении, транспортировке, реализации и т.д.

Типическая выборка. При типической выборке генеральная совокупность вначале расчленяется на однородные типические группы. Затем из каждой типической группы собственно-случайной или механической выборкой производится индивидуальный отбор единиц в выборочную совокупность.

Типическая выборка обычно применяется при изучении сложных статистических совокупностей. Например, при выборочном обследовании производительности труда работников торговли, состоящих из отдельных групп по квалификации.

Статистика

Важной особенностью типической выборки является то, что она дает более точные результаты по сравнению с другими способами отбора единиц в выборочную совокупность.

Для определения средней ошибки типической выборки используются формулы:

повторный отбор

$$\mu_{\omega} = \sqrt{\frac{\overline{\omega}(1-\overline{\omega})}{n}}, \qquad \mu_{\chi} = \sqrt{\frac{s_{x^2}}{n}}$$

бесповторный отбор

$$\mu_{\omega} = \sqrt{\frac{\overline{\omega}(1-\overline{\omega})}{n}\left(1-\frac{n}{N}\right)}, \qquad \mu_{\chi} = \sqrt{\frac{\overline{s}_{x^2}}{n}\left(1-\frac{n}{N}\right)}$$

Дисперсия определяется по следующим формулам:

$$\overline{s}_{\omega^2} = \frac{\sum \omega_i \left| 1-\omega_i \right| n_i}{\sum n_i}, \qquad \overline{s}_{\chi^2} = \frac{s_i^2 n_i}{\sum n_i}$$

При **одноступенчатой** выборке каждая отобранная единица сразу же подвергается изучению по заданному признаку. Так обстоит дело при собственно-случайной и серийной выборке.

При **многоступенчатой** выборке производят подбор из генеральной совокупности отдельных групп, а из групп выбираются отдельные единицы. Так производится типическая выборка с механическим способом отбора единиц в выборочную совокупность.

Комбинированная выборка может быть двухступенчатой. При этом генеральная совокупность сначала разбивается на группы. Затем производят отбор групп, а внутри последних осуществляется отбор отдельных единиц.

Контрольные вопросы:

1. Понятие, значение, характеристики выборочного наблюдения
2. Основные способы формирования выборочной совокупности
3. Определение необходимого объема выборки

Различают два типа связи между различными явлениями и их признаками: функциональную или жестко детерминированную и статистическую или стохастически детерминированную с другой стороны.

Если с изменением одной из переменных вторая изменяется строго определенным образом, т.е. значению одной переменной обязательно соответствует одно или несколько точно заданных значений другой переменной, связь между ними является функциональной.

При стохастически детерминированной связи (статистической) с изменением значения одной переменной вторая может в определенных пределах принимать любые значения с некоторыми вероятностями, но ее среднее значение или иные статистические (массовые) характеристики изменяются по определенному закону, т.е. разным значениям одной переменной соответствуют разные распределения значений другой переменной.

Частным случаем статистической связи является корреляционная связь.

Корреляционная связь - это связь, где воздействие отдельных факторов проявляется только как тенденция (в среднем) при массовом наблюдении фактических данных.

Наиболее простым вариантом корреляционной зависимости является парная корреляция, т.е. зависимость между двумя признаками (результативным и факторным или между двумя факторными). Математически эту зависимость можно выразить как зависимость результативного показателя у от факторного показателя х. Связи могут быть прямые и обратные. В первом случае с увеличением признака х увеличивается и признак у, при обратной связи с увеличением признака х уменьшается признак у.

Статистика

Методы изучения статистической связи.

Важнейшей задачей является определение формы связи с последующим расчетом параметров уравнения, или, иначе, нахождение уравнения связи (уравнения регрессии).

Могут иметь место различные формы связи:

$\bar{y}_x = a_0 + a_1 x$;

прямолинейная

линейные связи являются основными и применяются также и при многофакторном анализе.

$\bar{y}_x = a_0 + a_1 x + a_2 x^2$,

криволинейная в виде:

параболы второго порядка (или высших порядков)

параболической связью описывается взаимосвязь при которой характер связи между факторным и результативным признаком может измениться на противоположный при прохождении некоторого оптимального значения.

$\bar{y}_x = a_0 + \frac{a_1}{x}$,

гиперболы

гиперболические зависимости характерны для связей, в которых результативный признак не может варьироваться неограниченно, его вариация имеет односторонний предел.

$\bar{y}_x = a_0 \; a_1^x$

показательной функции

$n\,a_0 + a_1 \sum x = \sum y$,

Параметры для всех этих уравнений связи, как правило, определяют из системы нормальных уравнений, которые должны отвечать требованию метода наименьших квадратов (МНК):

$a_0 \sum x + a_1 \sum x^2 = \sum xy$.

$\eta = \frac{\delta}{\sigma} = \sqrt{\frac{\delta^2}{\sigma^2}}$,

Другая важнейшая задача - измерение тесноты зависимости - для всех форм связи может быть решена при помощи вычисления эмпирического корреляционного отношения:

где -

Статистика

$$\delta^2 = \frac{\sum(\bar{y}_x - \bar{y})^2}{n}$$

дисперсия в ряду выравненных значений результативного показателя ; -

$$\sigma^2 = \frac{\sum(y - \bar{y})^2}{n}$$

дисперсия в ряду фактических значений у.

Для определения степени тесноты парной линейной зависимости служит линейный коэффициент корреляции r, для расчета которого можно использовать следующие формулы:

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

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{n\sigma_x\sigma_y} \ ;$$

Линейный коэффициент корреляции может принимать значения в пределах от -1 до + 1 или по модулю от 0 до 1. Чем ближе он по абсолютной величине к 1, тем теснее связь. Знак указывает направление связи: «+» - прямая зависимость, «-» имеет место при обратной зависимости.

Общий вид многофакторного уравнения регрессии имеет вид:

$y = a + b_1x_1 + .. + b_kx_k$

Многофакторная система требует не одного, а множества показателей тесноты связей. Основой измерения связей является матрица коэффициентов корреляции. На основе этой матрицы судят о тесноте связи факторов с результативным признаком и между собой. Не рекомендуется включать в уравнение регрессии факторы слабо связанные с результативным признаком, но тесно связанные с другими факторами. Множественный коэффициент корреляции определяется как отношение части вариации результативного признака, объясняемой за счет вариации входящих в уравнение факторов, к общей вариации результативного признака за счет всех факторов. Под вариацией понимается сумма квадратов отклонений индивидуальных значений от расчетных по уравнению регрессии (объясненная вариация) или от общей средней величины признака (общая вариация).

Для случая двух факторов коэффициент множественной детерминации вычисляется по формуле из парных коэффициентов корреляции::

$$R_{yx_1x_2} = \frac{r_{yx_1}^2 + r_{yx_2}^2 - r_{yx_1}^2 r_{yx_2}^2 r_{x_1x_2}^2}{1 - r_{x_1x_2}^2}$$

Коэффициент частной детерминации фактора x_m – это доля вариации y, не объясненной ранее включенными факторами. Если обозначить частный коэффициент детерминации для фактора x_m как $r^2_{yx_m.x_1...x_{m-1},x_{m+1}..x_k}$ Тогда $r^2_{yx_m.x_1...x_{m-1},x_{m+1}..x_k} = \frac{R^2_y - R^2_{yx_1...x_{m-1},x_{m+1}..x_k}}{1 - R^2_{yx_1...x_{m-1},x_{m+1}..x_k}}$

Основные задачи применения корреляционно-регрессионного анализа.

В соответствии с сущностью корреляционной связи ее изучение имеет две цели: 1) измерение параметров уравнения, выражающего связь средних значений зависимой переменной со значениями независимой переменной; 2) измерение тесноты связи двух (или большего числа признаков) между собой

Задачи корреляционно-регрессионного анализа:

1. Задачи выделения важнейших факторов, влияющих на результативный признак (т.е. вариацию его значений в совокупности). Эта задача решается на базе мер тесноты связи факторов с результативным признаком.

2. Задачи оценки хозяйственной деятельности по эффективности использования факторов производства. Эта задача решается путем расчета для каждой единицы совокупности тех величин результативного признака, которые были получены при средней по совокупности эффективности использования факторов и сравнивания их с фактическими результатами производства.

3. Задача прогнозирования возможных значений результативного признака при задаваемых значениях факторных признаков. Такая задача решается путем подстановки ожидаемых, или планируемых, или возможных значений факторных признаков в уравнении связи и вычисления ожидаемых значений результативного признака.

4. Задача подготовки данных, необходимых в качестве исходных для решения оптимизационных задач.

При решении каждой из названных задач нужно учитывать особенности и ограничения корреляционно-регрессионного метода. Всякий раз необходимо специально обосновать возможность причинной интерпретации уравнения как объясняющего связь между вариацией фактора и результата. Трудно обеспечить раздельную оценку влияния каждого из факторов.

Непараметрические методы определения тесноты связи.

В статистической практике могут встречаться такие случаи, когда качества факторных и результативных признаков не могут быть выражены численно. Поэтому для измерения тесноты зависимости необходимо использовать другие показатели. Для этих целей используются так называемые непараметрические методы.

Наибольшее распространение имеют ранговые коэффициенты корреляции, в основу которых положен принцип нумерации значений статистического ряда. При использовании коэффициентов корреляции рангов коррелируются не сами значения показателей х и у, а только номера их мест, которые они занимают в каждом ряду значений. В этом случае номер каждой отдельной единицы будет ее рангом.

$$p = 1 - \frac{6\sum d^2}{n(n^2-1)},$$

Коэффициент корреляции рангов Спирмэна (р) основан на рассмотрении разности рангов значений результативного и факторного признаков и может быть рассчитан по формуле

где d = Nx - Ny , т.е. разность рангов каждой пары значений х и у; n - число наблюдений.

К непараметрическим методам исследования можно отнести коэффициент ассоциации $К_{ас}$ и коэффициент контингенции $К_{кон}$, которые используются, если, например, необходимо исследовать тесноту зависимости между качественными признаками, каждый из которых представлен в виде альтернативных признаков.

Для определения этих коэффициентов создается расчетная таблица (таблица «четырех полей»), где статистическое сказуемое схематически представлено в следующем виде:

Признаки	А(да)	А(нет)	Итого

В(да)	a	b	a + b
В(нет)	c	d	c + d
Итого	a + c	b + d	n

Здесь a, b, c, d - частоты взаимного сочетания (комбинации) двух альтернативных признаков ; n - общая сумма частот.

$$K_{ас} = \frac{ad - bc}{ad + bc} \ .$$

Коэффициент ассоциации можно рассчитать по формуле

$$K_{кон} = \frac{ad - bc}{\sqrt{(a+b)(b+d)(a+c)(c+d)}} \ .$$

Коэффициент контингенции рассчитывается по формуле

Нужно иметь в виду, что для одних и тех же данных коэффициент контингенции (изменяется от -1 до +1) всегда меньше коэффициента ассоциации.

Наконец, следует упомянуть коэффициент Фехнера, характеризующий элементарную степень тесноты связи, который целесообразно использовать для установления факта наличия связи, когда существует небольшой объем исходной информации. Данный коэффициент определяется по формуле

$$K_ф = \frac{n_a - n_b}{n_a + n_b} \ ,$$

где n_a - количество совпадений знаков отклонений индивидуальных величин от их средней арифметической; n_b - соответственно количество несовпадений. Коэффициент Фехнера может изменяться в пределах -1,0 $K_ф$ +1,0.

Целью применения корреляционно-регрессионного метода является построение такого уравнения регрессии, которое включает основные факторы, влияющие на вариацию результативного признака, обладающего высоким (не ниже 0,5) коэффициентом детерминации и коэффициентами регрессии, интерпретируемыми в соответствие с теоретическим знанием о природе связей в изучаемой системе.

При использовании корреляционно-регрессионного метода при анализе социально-экономических явлений необходимо учесть следующие ограничения.

Интерпретировать корреляционные показатели строго следует лишь в терминах вариации (различий в пространстве) отклонений от средней величины. Если задача состоит в изменении связи не между вариацией двух признаков в совокупности, а меду изменениями признаков объекта во времени, то корреляционно-регрессионный анализ требует значительных изменений.

Корреляционно-регрессионный метод основан на том, что группировка совокупности по одному факторному признаку при условии, что все другие не связаны с изучаемым, а случайные отклонения и ошибки взаимопогасились в большой совокупности. Если же фактор связан с другими факторами, то будет получена не чистая характеристика влияния.

При этом относительная простота и применение компьютерной техники позволяет достаточно широко и эффективно применять данных метод на практике.

Параметрический методы определения тесноты связи состоит в расчете F критерия Фишера, который рассчитывается по формуле:

Статистика

$$F = \frac{r^2 / \kappa}{1 - r^2}(n - \kappa - 1)$$

где r^2 – коэффициент корреляции, n – число единиц в совокупности, к- число степеней свобожы.

Для оценки надежности результатов уравнения регрессии F сравнивают с $F_{табл}$ при заданном уровне вероятности. Если полученное значение больше табличного, то можно говорить о высокой степени надежности результатов регрессионного моделирования. Если ниже – то полученные оценки коэффициентов регрессии статистически незначимы.

Коэффициент конкордации

$$W = \frac{12}{m^2(n^3 - n)} S$$

$$S = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} R_{ij} - \frac{m(n+1)}{2}\right)^2$$

где n - количество анализируемых объектов, m - количество экспертов, R_{ij} - ранг j-го объекта, который присвоен ему i-ым экспертом.

Следует обратить внимание на отличие значений коэффициента конкордации от коэффициента корреляции, так как он существует в пределах от 0 до 1. Если мнения экспертов полностью противоположны, коэффициент конкордации равен нулю (W = 0), а коэффициент корреляции в этом случае будет равен -1.

Контрольные вопросы:

1. Дайте определение функциональной,стахостической и корреляционной связи.
2. Дайте отределение факторного и результативного признака.
3. Что означает прямая и обратная связь между признаками?
4. Какие задачи решаются с помощью корреляционно-регрессионного анализа?

Изменение социально-экономических явлений во времени изучается статистикой методом построения и анализа динамических рядов. Ряды динамики - это значения статистических показателей, которые представлены в определенной хронологической последовательности.

Каждый динамический ряд содержит две составляющие: 1) показатели периодов времени (годы, кварталы, месяцы, дни или даты); 2) показатели, характеризующие исследуемый объект за временные периоды или на соответствующие даты, которые называют уровнями ряда.

Уровни ряда выражаются как абсолютными, так и средними или относительными величинами. В зависимости от характера показателей строят динамические ряды абсолютных, относительных и средних величин. Ряды динамики из относительных и средних величин строят на основе производных рядов абсолютных величин. Различают интервальные и моментные ряды динамики.

Динамический интервальный ряд содержит значения показателей за определенные периоды времени. В интервальном ряду уровни можно суммировать, получая объем явления за более длительный период, или так называемые накопленные итоги.

Динамический моментный ряд отражает значения показателей на определенный момент времени (дату времени). В моментных рядах исследователя может интересовать только разность явлений, отражающая изменение уровня ряда между определенными датами, поскольку сумма уровней здесь не имеет реального содержания. Накопленные итоги здесь не рассчитываются.

Важнейшим условием правильного построения динамических рядов является сопоставимость уровней рядов, относящихся к различным периодам. Уровни должны быть представлены в однородных величинах, должна иметь место одинаковая полнота охвата различных частей явления.

Для того, чтобы избежать искажения реальной динамики, в статистическом исследовании проводятся предварительные расчеты (смыкание рядов динамики), которые предшествуют статистическому анализу динамических рядов. Под смыканием рядов динамики понимается объединение в один ряд двух и более рядов, уровни которых рассчитаны по разной методологии или не соответствуют территориальным границам и т.д. Смыкание рядов динамики может предполагать также приведение абсолютных уровней рядов динамики к общему основанию, что нивелирует несопоставимость уровней рядов динамики.

Для характеристики интенсивности развития во времени используются статистические показатели, получаемые сравнением уровней между собой, в результате чего получаем систему абсолютных и относительных показателей динамики: абсолютный прирост, коэффициент роста, темп роста, темп прироста, абсолютное значение 1% прироста. Для характеристики интенсивности развития за длительный период рассчитываются средние показатели: средний уровень ряда, средний абсолютный прирост, средний коэффициент роста, средний темп роста, средний темп прироста, среднее абсолютное значение 1% прироста.

Базисные показатели характеризуют итоговый результат всех изменений в уровнях ряда от периода базисного уровня до данного (i-го) периода. Рассчитываются как отношение i-го уровня к базисному (начальному).

Цепные показатели характеризуют интенсивность изменения уровня от одного периода к другому в пределах того промежутка времени, который исследуется. Рассчитываются как отношение i-го к предшествующему уровню.

Абсолютный прирост выражает абсолютную скорость изменения ряда динамики и определяется как разность между данным уровнем и уровнем, принятым за базу сравнения.

$$\Delta_{(Б)} = y_i - y_0 ,$$

Абсолютный прирост (базисный)

где yi - уровень сравниваемого периода; y0 - уровень базисного периода.

Абсолютный прирост с переменной базой (цепной), который называют скоростью роста,

$$\Delta_{(Ц)} = y_i - y_{i-1} ,$$

где yi - уровень сравниваемого периода; yi-1 - уровень предшествующего периода.

Коэффициент роста K_i определяется как отношение данного уровня к предыдущему или базисному, показывает относительную скорость изменения ряда. Если коэффициент роста выражается в процентах, то его называют темпом роста.

$$K_{(Б)} = \frac{y_i}{y_0} .$$

Коэффициент роста базисный

$$K_{(Ц)} = \frac{y_i}{y_{i-1}} .$$

Статистика

Коэффициент роста цепной

$$T_р = K \cdot 100\% .$$

Темп роста

Темп прироста ТП определяется как отношение абсолютного прироста данного уровня к предыдущему или базисному.

$$T_{п(Б)} = \frac{y_i - y_0}{y_0} \cdot 100\% .$$

Темп прироста базисный

$$T_{п(Ц)} = \frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100\% .$$

Темп прироста цепной

Темп прироста можно рассчитать и иным путем: как разность между темпом роста и 100 % или как разность между коэффициентом роста и 1 (единицей):

1) $T_п = T_р$ - 100%; 2) $T_п = K_i$ - 1.

$$A_i = \frac{y_i - y_{i-1}}{T_{п_{i/(i-1)}}} = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100\%} = \frac{y_{i-1}}{100} = 0{,}01 y_{i-1} .$$

Абсолютное значение одного процента прироста A_i . Этот показатель служит косвенной мерой базисного уровня. Представляет собой одну сотую часть базисного уровня, но одновременно представляет собой и отношение абсолютного прироста к соответствующему темпу роста. Данный показатель рассчитывают по формуле

Для характеристики динамики изучаемого явления за продолжительный период рассчитывают группу средних показателей динамики. Можно выделить две категории показателей в этой группе: а) средние уровни ряда; б) средние показатели изменения уровней ряда.

Средние уровни ряда рассчитываются в зависимости от вида временного ряда.

Для интервального ряда динамики абсолютных показателей средний уровень ряда рассчитывается по формуле простой средней арифметической:

$$\bar{y} = \frac{\sum y}{n} ,$$

где n - число уровней ряда.

Средний уровень моментного ряда с равными интервалами рассчитывается по формуле

Статистика

средней хронологической:

$$\bar{y} = \frac{\frac{y_1}{2} + y_2 + y_3 + \ldots + y_{n-1} + \frac{y_n}{2}}{n-1},$$

где n - число дат.

Средний уровень моментного ряда с неравными интервалами рассчитывается по формуле средней арифметической взвешенной, где в качестве весов берется продолжительность промежутков времени между временными моментами изменений в уровнях динамического ряда:

$$\bar{y} = \frac{\sum y \cdot t}{\sum t},$$

где t - продолжительность периода (дни, месяцы), в течение которого уровень не изменялся.

Средний абсолютный прирост (средняя скорость роста) определяется как средняя арифметическая из показателей скорости роста за отдельные периоды времени:

$$\bar{\Delta} = \frac{\sum \Delta}{n-1}, \text{ или}$$

$$\bar{\Delta} = \frac{y_n - y_1}{n-1},$$

где y_n - конечный уровень ряда; y_1 - начальный уровень ряда.

Средний коэффициент роста рассчитывается по формуле средней геометрической из показателей коэффициентов роста за отдельные периоды:

$$\bar{K}_p = \sqrt[n-1]{K_{p1} K_{p2} \ldots K_{p\,n-1}},$$

$$\bar{K}_p = \sqrt[n-1]{\frac{Y_n}{Y_1}}.$$

где K_{p1}, K_{p2}, ..., $K_{p\,n-1}$ - коэффициенты роста по сравнению с предыдущим периодом; n - число уровней ряда.

Средний коэффициент роста можно определить иначе:

Средний темп роста, %. Это средний коэффициент роста, который выражается в процентах:

Статистика

$$\bar{T}_р = \bar{K}_р \cdot 100 \ .$$

Средний темп прироста , %. Для расчета данного показателя первоначально определяется средний темп роста, который затем уменьшается на 100%. Его также можно определить, если уменьшить средний коэффициент роста на единицу:

$$\bar{T}_п = \bar{T}_р - 100; \quad \bar{T}_п = \left(\bar{K}_р - 1\right) \cdot 100 \ .$$

$$\bar{A} = \frac{\bar{\Delta}}{\bar{T}_п} \ .$$

Среднее абсолютное значение 1% прироста можно рассчитать по формуле

В ходе обработки динамического ряда важнейшей задачей является выявление основной тенденции развития явления (тренда) и сглаживание случайных колебаний. Для решения этой задачи в статистике существуют особые способы, которые называют методами выравнивания.

Выделяют три основных способа обработки динамического ряда: а) укрупнение интервалов динамического ряда и расчет средних для каждого укрупненного интервала; б) метод скользящей средней; в) аналитическое выравнивание (выравнивание по аналитическим формулам).

Укрупнение интервалов - наиболее простой способ. Он заключается в преобразовании первоначальных рядов динамики в более крупные по продолжительности временных периодов, что позволяет более четко выявить действие основной тенденции (основных факторов) изменения уровней. По интервальным рядам итоги исчисляются путем простого суммирования уровней первоначальных рядов. Для других случаев рассчитывают средние величины укрупненных рядов (переменная средняя). Переменная средняя рассчитывается по формулам простой средней арифметической.

Скользящая средняя - это такая динамическая средняя, которая последовательно рассчитывается при передвижении на один интервал при заданной продолжительности периода. Если, предположим, продолжительность периода равна 3, то скользящие средние рассчитываются следующим образом:

$$\bar{y}_1 = \frac{y_1 + y_2 + y_3}{3} \ ;$$

$$\bar{y}_2 = \frac{y_2 + y_3 + y_4}{3} \ ;$$

и т.д.

Первую рассчитанную среднюю относят ко второму периоду, вторую - к третьему, третью - к четвертому и т.д. По сравнению с фактическим сглаженный ряд становится короче на (m - 1)/2, где m - число уровней интервала.

Важнейшим способом количественного выражения общей тенденции изменения уровней динамического ряда является аналитическое выравнивание ряда динамики, которое позволяет получить описание плавной линии развития ряда. При этом эмпирические уровни заменяются уровнями, которые рассчитываются на основе определенной кривой, где уравнение рассматривается как функция времени. Вид уравнения зависит от конкретного характера динамики развития. Его можно определить как теоретически, так и практически. Теоретический анализ основывается на рассчитанных показателях динамики. Практический анализ - на исследовании линейной диаграммы.

Задачей аналитического выравнивания является определение не только общей тенденции развития явления, но и некоторых недостающих значений как внутри периода, так и за его пределами. Способ определения неизвестных значений внутри динамического ряда называют интерполяцией. Эти неизвестные значения можно определить: 1) используя полусумму уровней, расположенных рядом с интерполируемыми; 2) по среднему абсолютному приросту; 3) по темпу роста. В результате аналитического выравнивания получают следующую трендовую модель:

$$У_t = f(t) + \varepsilon_t.$$

где f(t) – уровень, определяемый тенденцией развития; e_t – случайное и циклическое отклонение от тенденции.

Целью аналитического выравнивания динамического ряда является определение аналитической или графической зависимости f(t). На практике по имеющемуся временному ряду задают вид и находят параметры функции f(t), а затем анализируют поведение отклонений от тенденции. Функцию f(t) выбирают таким образом, чтобы она давала содержательное объяснение изучаемого процесса.

Чаще всего при выравнивании используются следующие зависимости:

линейная $f(t) = a_0 + a_1 t;$

параболическая $f(t) = a_0 + a_1 t + a_2 t^2,$

экспоненциальные $f(t) = \exp(a_0 + a_1 t)$

или $f(t) = \exp(a_0 + a_1 t + a_2 t^2).$

Линейная зависимость выбирается в тех случаях, когда в исходном временном ряду наблюдаются более или менее постоянные абсолютные цепные приросты, не проявляющие тенденции ни к увеличению, ни к снижению.

Параболическая зависимость используется, если абсолютные цепные приросты сами по себе обнаруживают некоторую тенденцию развития, но абсолютные цепные приросты абсолютных цепных приростов (разности второго порядка) никакой тенденции развития не проявляют.

Экспоненциальные зависимости применяются, если в исходном временном ряду наблюдается либо более или менее постоянный относительный рост (устойчивость цепных темпов роста, темпов прироста, коэффициентов роста), либо, при отсутствии такого постоянства, – устойчивость в изменении показателей относительного роста (цепных темпов роста цепных же темпов роста, цепных коэффициентов роста цепных же коэффициентов или темпов роста и т.п.).

Оценка параметров (a0, a1, a2, ...) осуществляется следующими методами:
1) методом избранных точек,
2) методом наименьших расстояний,
3) методом наименьших квадратов (МНК).

В большинстве расчетов используют метод наименьших квадратов, который обеспечивает наименьшую сумму квадратов отклонений фактических уровней от выравненных:

$$\min \sum (У_t - f(t))^2.$$

Для линейной зависимости ($f(t)=a_0+a_1t$) параметр a_0 обычно интерпретации не имеет, но иногда его рассматривают как обобщенный начальный уровень ряда; a_1 – сила связи, т.е. параметр, показывающий, насколько изменится результат при изменении времени на единицу. Таким образом, а можно представить как постоянный теоретический абсолютный прирост.

$$F_{\text{факт}} = \frac{\frac{1}{k-1}\sigma^2_{\text{факт}}}{\frac{1}{n-k}\sigma^2_{\text{ост}}}, \quad F_{\text{факт}} = \frac{\sigma^2_{\text{факт}}(n-k)}{\sigma^2_{\text{ост}}(k-1)},$$

Построив уравнение регрессии, проводят оценку его надежности. Оценка надежности параметров уравнения проводится на основании анализа случайной компоненты. Это делается посредством критерия Фишера (F). Фактический уровень ($F_{\text{факт}}$) сравнивается с теоретическим (табличным) значением:

где k – число параметров функции, описывающей тенденцию;

$$\sigma^2_{\text{ост}} = \frac{\sum(y - f(t))^2}{n};$$

$$\sigma^2_{\text{факт}} = \sigma^2_y - \sigma^2_{\text{ост}} = \frac{\sum(f(t) - \bar{y})^2}{n};$$

$$\sigma^2_y = \frac{\sum(y - \bar{y})^2}{n} = \sigma^2_{\text{факт}} + \sigma^2_{\text{ост}}.$$

n – число уровней ряда;

$F_{\text{факт}}$ сравнивается с $F_{\text{теор}}$ при v1 = (k-1), v2 = (n-k) степенях свободы и уровне значимости а (обычно а = 0,05). Если Fфакт > Fтеор, уравнение регрессии значимо, т.е. построенная модель адекватна фактической временной тенденции.

Способ определения количественных значений за пределами ряда называют экстраполяцией. Экстраполирование используется для прогнозирования тех факторов, которые не только в прошлом и настоящем обусловливают развитие явления, но и могут оказать влияние на его развитие в будущем.

Экстраполировать можно по средней арифметической, по среднему абсолютному приросту, по среднему темпу роста.

$$r_a = \frac{\overline{y_t y_{t-1}} - \bar{y}_t \cdot \bar{y}_{t-1}}{\sigma_{y_t}\sigma_{y_{t-1}}}.$$

При аналитическом выравнивании может иметь место автокорреляция, под которой понимается зависимость между соседними членами динамического ряда. Автокорреляцию можно установить с помощью перемещения уровня на одну дату. Коэффициент автокорреляции вычисляется по формуле

Автокорреляцию в рядах можно устранить, коррелируя не сами уровни, а так называемые остаточные величины (разность эмпирических и теоретических уровней). В этом случае корреляцию между остаточными величинами можно определить по формуле

$$r = \frac{\sum(x - \bar{x}_t)(y - \bar{y}_t)}{\sqrt{\sum(x - \bar{x}_t)^2 \sum(y - \bar{y}_t)^2}}.$$

Анализ рядов динамики предполагает и исследование сезонной неравномерности (сезонных колебаний), под которыми понимают устойчивые внутригодовые колебания, причиной которых являются многочисленные факторы, в том числе и природно-климатические. Сезонные колебания измеряются с помощью индексов сезонности, которые рассчитываются двумя способами в зависимости от характера динамического развития.

При относительно неизменном годовом уровне явления индекс сезонности можно рассчитать как процентное отношение средней величины из фактических уровней одноименных месяцев к общему среднему уровню за исследуемый период:

$$r = \frac{\sum(x - \bar{x}_i)(y - \bar{y}_i)}{\sqrt{\sum(x - \bar{x}_i)^2 \sum(y - \bar{y}_i)^2}} .$$

В условиях изменчивости годового уровня индекс сезонности определяется как процентное отношение средней величины из фактических уровней одноименных месяцев к средней величине из выровненных уровней одноименных месяцев:

$$И_с = \frac{\bar{y}_i}{\hat{y}_i} \cdot 100 .$$

В простейших случаях для характеристики взаимосвязи двух или более рядов их приводят к общему основанию, для чего берут в качестве базисных уровни за один и тот же период и исчисляют коэффициенты опережения по темпам роста или прироста. Коэффициенты опережения по темпам роста – это отношение темпов роста (цепных или базисных) одного ряда к соответствующим по времени темпам роста (также цепным или базисным) другого ряда. Аналогично находятся и коэффициенты опережения по темпам прироста.

Временной лаг - экономический показатель, показывающий отставание или опережение одного экономического явления по сравнению с другим, связанным с ним явлением.

Контрольные вопросы:

1.Какие абсолютные и относительные показатели анализа рядов динамики Вы знаете?

2.Что такое темп роста и темп прироста? Как они рассчитываются и какая между ними взаимосвязь.

3.Что показывает темп наращивания?

4.Какая разница между периодическим и моментным рядами динамики?

5.Как рассчитываются средние уровни моментных и периодических рядов динамики.

Индексами называют сравнительные относительные величины, которые характеризуют изменение сложных социально-экономических показателей (показатели, состоящие из несуммируемых элементов) во времени, в пространстве, по сравнению с планом.

Индекс - это результат сравнения двух одноименных показателей, при исчислении которого следует различать числитель индексного отношения (сравниваемый или отчетный уровень) и знаменатель индексного отношения (базисный уровень, с которым производится сравнение). Выбор базы зависит от цели исследования. Если изучается динамика, то за базисную величину может быть взят размер показателя в периоде, предшествующем отчетному. Если необходимо осуществить территориальное сравнение, то за базу можно принять данные другой территории. За базу сравнения могут приниматься плановые показатели, если необходимо использовать индексы как показатели выполнения плана.

Признак изменение которого характеризует индекс называется индексируемым.

Признак-вес выполняет функцию веса по отношению к индексируемому признаку.

При построении индексов решают следующие вопросы: 1) определение вида индекса и вида показателей с помощью которых строится индекс; 2) выбор базы (а) данные по той же совокупности и по тому же признаку за предшествующий период; б) плановое задание; в) данные по какой-либо другой совокупности, сходной по характеру с изучаемой).

При установлении базы необходимо соблюдать следующие правила: сопоставимость базисных и отчетных данных; обеспечить типичность базовых данных.

По степени охвата элементов явления индексы делят на индивидуальные и общие (сводные).

Индивидуальные индексы (i) - это индексы, которые характеризуют изменение только одного элемента совокупности.

Общий (сводный) индекс (I) характеризует изменение по всей совокупности элементов сложного явления. Если индексы охватывают только часть явления, то их называют групповыми.

В зависимости от способа изучения общие индексы могут быть построены или как агрегатные (от лат. aggrega - присоединяю) индексы, или как средние взвешенные индексы (средние из индивидуальных).

Способ построения агрегатных индексов заключается в том, что при помощи так называемых соизмерителей можно выразить итоговые величины сложной совокупности в отчетном и базисном периодах, а затем первую сопоставить со второй.

Если индексы можно рассчитать на основе сравнения двух сумм, полученных, например, путем умножения среднесписочной численности работников в базисном и отчетном периоде (по каждому j предприятию, структурному подразделению и т.д.) t_{0j} и t_{1j} и средней заработной - z_{0j} и z_{1j} , то такие индексы называют агрегатными. Таким образом, общие индексы могут быть рассчитаны не только через осреднение индивидуальных индексов, а и на основе сравнения двух сумм (агрегатов). Агрегатные индексы считаются основной формой индексов. Они выполняют две функции: синтетическую и аналитическую.

Первая функция обеспечивается тем, что в одном индексе обобщаются (синтезируются) непосредственно несоизмеримые явления, когда мы записываем $I_z = \frac{\sum z_0 t_1}{\sum z_0 t_0}$ (где z – средняя заработная плата, а t – среднесписочная численность работников), то благодаря использованию денежного соизмерителя можно агрегировать данные по различным категориям работников (несопоставимым по натуральным измерителям).

Аналитическая функция вытекает из взаимосвязи индексов, т.к. практически каждый индекс можно рассматривать как составляющую некой системы индексов, в которой его роль сводится к измерению одного из факторов общего изменения сложного явления и вклада этого фактора в соответствующее изменение.

Так, например, индекс цен можно рассматривать как показатель влияния изменения средней заработной платы на фонд оплаты труда, что основано на следующей связи признаков: среднесписочная численность * средняя заработная плата = фонд оплаты труда или tz = w. Системе признаков соответствует система индексов.

$$I_t = \frac{\sum z_0 t_1}{\sum z_0 t_0} \quad \text{и} \quad I_z = \frac{\sum z_1 t_1}{\sum z_0 t_1}$$

Когда мы указывает индекс среднесписочной численности работников или индекс средней заработной платы, мы имеем в виду изменение фонда оплаты труда за счет изменения среднесписочной численности работников или средней заработной платы.

При построении агрегатных индексов пользуются такими понятиями, как индексируемый признак и признак-вес. Индексируемый признак – это признак, изменение которого характеризует данный индекс. Например, в I_t – это t. Значение индексируемого признака изменяется, т.е. отчетное значение сопоставляется с базисным.

Признак-вес выполняет функцию веса по отношению к индексируемому признаку, его значение в индексе принимается постоянным, т.к. он не должен искажать оценку изменения индексируемого признака. Например, в I_t – это z.

Если индексы рассматриваются в системе, то должна обеспечиваться взаимосвязь между ними: $I_t * I_z = I_w$

Статистика

Расчет среднего индекса применяется при определении общего индекса или общего изменения состояния изучаемого объекта. Так как расчет среднего индекса как отношения суммы индивидуальных признаков в текущем и базисном периоде ($\frac{z_{11} + z_{12} + \ldots\ldots + z_{1n}}{z_{01} + z_{02} + \ldots\ldots + z_{0n}}$) или как простой средней из индивидуальных индексов ($\frac{\frac{z_{11}}{z_{01}} + \frac{z_{12}}{z_{02}} + \ldots + \frac{z_{1n}}{z_{0n}}}{n}$), т.е. невзвешенных средних арифметических не учитывает объемов и структуры изучаемого объекта, то применяют взвешенную среднюю.

Для расчета среднего индекса может использоваться другие формы средних величин.

Средняя геометрическая: $I_z = \sqrt[n]{i_{z1} \cdot i_{z2} \cdot \ldots\ldots \cdot i_{zn}}$

Средняя гармоническая невзвешенная рассчитывается по формуле:

$$I_z = \frac{n}{\frac{1}{i_{z1}} + \frac{1}{i_{z2}} + \ldots + \frac{1}{i_{zn}}}$$

Индексы с постоянными и переменными весами и метод выявления роли факторов динамики сложных явлений.

При построении агрегатных индексов веса могут быть закреплены на базисном, отчетном или смешанном уровнях. При закреплении весов только на базисном или только на отчетном уровне, постоянных весов, равенство

I_t * I_z = I_w не выполняется. Например, $\frac{\sum t_1 z_0}{\sum t_0 z_0} \cdot \frac{\sum t_0 z_1}{\sum t_0 z_0} \neq \frac{\sum t_1 z_1}{\sum t_0 z_0}$

Только когда взаимосвязанные индексы строятся с весами разных периодов, увязка их в системе выполняется. Например, $\frac{\sum t_1 z_0}{\sum t_0 z_0} \cdot \frac{\sum t_1 z_1}{\sum t_1 z_0} = \frac{\sum t_1 z_1}{\sum t_0 z_0}$. В приведенном примере индексы первичных признаков стоятся на весах базисного периода, вторичных – на весах отчетного периода. Отечественная статистика в своей практике придерживалась именно такого подхода. Но при таком подходе значение полученных индексов при изменении последовательности признаков различаются, т.е. если в модели tz = w t и z поменять местами значения полученных индексов будут иметь расхождения.

Различие между индексами с разными весами можно объяснить при помощи уравнения В.И. Борткевича (1868 – 1931):

$$\frac{\sum t_1 z_1}{\sum t_1 z_0} : \frac{\sum t_0 z_1}{\sum t_0 z_0} = 1 + r_{i_z i_t} v_{i_t} v_{i_z}$$

, где $r_{i_z i_t}$ - корреляция между изменением цен и объемом продаж на отдельные товары, $v_{i_t} v_{i_z}$ - темпы изменения объемов реализованных товаров и цен соответственно.

Таким образом, из формулы видно, что индексы с отчетными и базисными весами будут равны, если выполняется хотя бы одно из условий: $r_{i_z i_t} = 0$, $v_{i_z} = 0$, $v_{i_t} = 0$. Чем больше величина сравниваемого периода, тем сильнее проявляется различие.

Однако на практике, как правило, стремятся получить однозначное решение тем или иным способом. Первый способ заключается в получении средних оценок изменений, либо путем построения индексов на средних весах $I_z = \frac{\sum z_1 \frac{t_1 + t_0}{2}}{\sum z_0 \frac{t_1 + t_0}{2}}$

либо через осреднение равновзвешенных индексов. При этом предпочтение отдается средней

Статистика

геометрической. $$I_z = \sqrt{\frac{\sum t_1 z_1}{\sum t_1 z_0} \cdot \frac{\sum t_0 z_1}{\sum t_0 z_0}}$$

Второй путь основан на предпочтении какого-то одного варианта построения взаимосвязанных индексов, применялся в отечественной практике.

В статистике имеют большое значение индексы переменного и фиксированного состава, которые используются при анализе динамики средних показателей.

Индексом переменного состава называют отношение двух средних уровней.

$$I_{пер.сост} = \frac{\sum z_1 t_1}{\sum t_1} : \frac{\sum z_0 t_0}{\sum t_0}$$

Индекс фиксированного состава есть средний из индивидуальных индексов. Он рассчитывается как отношение двух стандартизованных средних, где влияние изменения структурного фактора устранено, поэтому данный индекс называют еще индексом постоянного состава.

$$I_{фикс.сост} = \frac{\sum z_1 t_1}{\sum t_1} : \frac{\sum z_0 t_1}{\sum t_1}$$

$$I_{структ} = I_{фикс.сост} : I_{пост.сост}$$

В зависимости от характера и содержания индексируемых величин различают индексы количественных (объемных) показателей и индексы качественных показателей.

К индексам количественных (объемных) показателей относятся такие индексы, как индексы физического объема производства продукции, затрат на выпуск продукции, стоимости продукции, а также индексы показателей, размеры которых определяются абсолютными величинами. Используются различные виды индексов количественных показателей.

Индекс физического объема продукции (ФОП) отражает изменение выпуска продукции.

Индивидуальный индекс ФОП отражает изменение выпуска продукции одного вида и определяется по формуле

$$i_{q_{1/0}} = \frac{q_1}{q_0},$$

где q_1 и q_0 - количество продукции данного вида в натуральном выражении в текущем и базисном периодах.

Агрегатный индекс ФОП отражает изменение выпуска всей совокупности продукции, где индексируемой величиной является количество продукции q, а соизмерителем - цена p:

$$I_{q_{1/0}} = \frac{\sum q_1 p_0}{\sum q_0 p_0},$$

где q_1 и q_0 - количество выработанных единиц отдельных видов продукции соответственно в отчетном и базисном периодах; p_0 - цена единицы продукции (отдельного вида) в базисном периоде.

При вычислении индекса ФОП в качестве соизмерителей может выступать также себестоимость продукции или трудоемкость.

Средние взвешенные индексы ФОП используются в том случае, если известны индивидуальные индексы объема по отдельным видам продукции и стоимость отдельных видов продукции (или затраты) в базисном или отчетном периоде.

Средний взвешенный арифметический индекс ФОП определяется по формуле

$$I_{q_{1/0}} = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0},$$

Статистика

где i_q - индивидуальный индекс по каждому виду продукции; q0 p0 - стоимость продукции каждого вида в базисном периоде.

Средний взвешенный гармонический индекс ФОП

$$I_{q1/0} = \frac{\sum q_1 p_1}{\sum \frac{1}{i_q} q_1 p_1},$$

где q_1 p_1 - стоимость продукции каждого вида в текущем периоде.

Аналогично рассчитывается индекс затрат на выпуск продукции, который отражает изменение затрат на производство и может быть как индивидуальным, так и агрегатным.

Между индексами существует также взаимосвязь и взаимозависимость, как и между самими экономическими явлениями, что позволяет проводить факторный анализ. Благодаря индексному методу можно рассматривать все факторы независимо друг от друга, что дает возможность определить размер абсолютного изменения сложного явления за счет каждого фактора в отдельности.

Предположим, что результативный признак зависит от трех факторов и более. В этом случае результативный индекс примет вид

$$I = \frac{\sum a_1 b_1 c_1 \ldots n_1}{\sum a_0 b_0 c_0 \ldots n_0}.$$

Изменение результативного индекса за счет каждого фактора может быть выражено следующим образом:

$$I_n = \frac{\sum a_1 b_1 c_1 \ldots n_1}{\sum a_1 b_1 c_1 \ldots n_0},$$

$$I_c = \frac{\sum a_1 b_1 c_1 \ldots n_1}{\sum a_1 b_1 c_0 \ldots n_0},$$

$$I_b = \frac{\sum a_1 b_1 c_0 \ldots n_1}{\sum a_1 b_0 c_0 \ldots n_0},$$

$$I_a = \frac{\sum a_1 b_0 c_0 \ldots n_0}{\sum a_0 b_0 c_0 \ldots n_0},$$

Для выявления роли каждого фактора в отдельности индекс сложного показателя разлагают на частные (факторные) индексы, которые характеризуют роль каждого фактора. При этом используют два метода: метод обособленного изучения факторов; последовательно-цепной метод.

При первом методе сложный показатель берется с учетом изменения лишь того фактора, который взят в качестве исследуемого, все остальные остаются неизменными на уровне базисного периода.

Последовательно-цепной метод предполагает использование системы взаимосвязанных индексов, которая требует определенного расположения факторов. Как правило, на первом месте в цепи располагают качественный фактор. При определении влияния первого фактора все остальные сохраняются в числителе и знаменателе на уровне базисного периода, при определении второго факторного индекса первый фактор сохраняется на уровне базисного периода, а третий и все последующие - на уровне отчетного периода, при определении третьего факторного индекса первый и второй факторы сохраняются на уровне базисного периода, четвертый и все остальные - на уровне отчетного периода и т.д.

Территориальные индексы.

Индексы могут быть использованы не только как показатели сравнения состояний изучаемого явления во времени, но и в пространстве, между отдельными территориями. Индексы позволяющие сравнивать различные территориальные образования между собой носят название территориальных индексов. При построении территориальных индексов применяются те же правила, что при сравнении явлении во времени, только в территориальных индексах в качестве весов используются показатели численности населения, доли в общих доходах населения от заработной платы и т.д. Кроме того, при сравнении разных территорий за один период значки «0» и «1» не используются. Использование индексов при анализе различий между территориями обусловлено следующим: индексы позволяют сопоставить территории с разным уровнем социально-экономического развития, с разным уровнем развития производства, с разной структурой потребительского рынка и доходов и т.д.

Контрольныу вопросы:

1. В чём разница между индивидуальными и общими индексами?
2. Каков экономический смысл применение агрегатных индексов физического объёма по ценам базисным? По текущим ценам?
3. Что означает компонентная зависимость между агрегатными индексами цен и физического объёма?
4. Какие, конкретно взаимосвязаны индексы?
5. В чём суть факторного анализа по индексному методу?
6. Чем объясняется разница в результатах расчёта агрегатных индексов Пааше и Ласпейреса? Какова область применения в анализе методике Пааше и Ласпейреса?

Дисперсионный анализ.

Целью дисперсионного анализа является проверка статистической значимости различия между средними (для групп или переменных). Эта проверка проводится с помощью разбиения суммы квадратов на компоненты, т.е. с помощью разбиения общей дисперсии (вариации) на части, одна из которых обусловлена случайной ошибкой (то есть внутригрупповой изменчивостью), а вторая связана с различием средних значений. Последняя компонента дисперсии затем используется для анализа статистической значимости различия между средними значениями. Если это различие *значимо*, нулевая гипотеза *отвергается* и принимается альтернативная гипотеза о существовании различия между средними.

Разбиение суммы квадратов. Для выборки объема n выборочная дисперсия вычисляется как сумма квадратов отклонений от выборочного среднего, деленная на n-1 (объем выборки минус единица). Таким образом, при фиксированном объеме выборки n дисперсия есть функция суммы квадратов (отклонений). В основе дисперсионного анализа лежит разделение дисперсии на части или компоненты, т.е. выборка разбивается на две части в которых вычисляются среднии и сумма квадратов отклонений. Расчет тех же показателей по выборки в целом дает большее значение дисперсии, что объясняется расхождение между групповыми средними. Таким образом, дисперсионный анализ позволяет объяснить внутригрупповую изменчивость, которая при исследовании всей группы в целом не может быть изменена.

Проверка значимости в дисперсионном анализе основана на сравнении компоненты дисперсии, обусловленной межгрупповым и компоненты дисперсии, обусловленной внутригрупповым разбросом (называемой средним квадратом ошибки). Если верна нулевая гипотеза (равенство средних в двух популяциях), то можно ожидать сравнительно небольшое различие выборочных средних из-за чисто случайной изменчивости. Поэтому, при нулевой гипотезе, внутригрупповая дисперсия будет практически совпадать с общей дисперсией, подсчитанной без учета групповой принадлежности. Полученные внутригрупповые дисперсии можно сравнить с помощью F-критерия, проверяющего, действительно ли отношение дисперсий значимо больше 1.

Преимущества: 1) дисперсионный анализ существенно более эффективен и, для малых выборок, т.к. более информативен; 2)дисперсионный анализ позволяет обнаружить эффекты *взаимодействия* между факторами и, поэтому, позволяет проверять более сложные гипотезы

Метод главных компонент состоит в линейном понижении размерности, в котором определяются попарно ортогональные направления максимальной вариации исходных данных, после чего данные проектируются на пространство меньшей размерности, порожденное компонентами с наибольшей вариацией.

Метод главных компонент является частью факторного анализа, который состоит в том, что две коррелированные переменные объединены в один фактор. Если пример с двумя переменными распространить на большее число переменных, то вычисления становятся сложнее, однако основной принцип представления двух или более зависимых переменных одним фактором остается в силе.

При сокращении числа переменных решение о том, когда следует остановить процедуру выделения факторов, главным образом зависит от точки зрения на то, что считать малой "случайной" изменчивостью. При повторных итерациях выделяются факторы с все меньшей и меньшей дисперсией.

Центроидный метод определения факторов.

Центроидный метод используется при кластерном анализе. В этом методе расстояние между двумя кластерами определяется как расстояние между их центрами тяжести при не взвешенном центроидном методе..

Взвешенный центроидный метод (медиана) идентичен не взвешенному, за исключением того, что при вычислениях используются веса для учёта разницы между размерами кластеров (т.е. числами объектов в них). Поэтому, если имеются (или подозреваются) значительные отличия в размерах кластеров, этот метод оказывается предпочтительнее предыдущего.

Кластерный анализ.

Термин кластерный анализ в действительности включает в себя набор различных алгоритмов классификации. Общий вопрос, задаваемый исследователями во многих областях, состоит в том, как организовать наблюдаемые данные в наглядные структуры, т.е. определить кластеры схожих объектов. Фактически, кластерный анализ является не столько обычным статистическим методом, сколько "набором" различных алгоритмов "распределения объектов по кластерам". Существует точка зрения, что в отличие от многих других статистических процедур, методы кластерного анализа используются в большинстве случаев тогда, когда вы не имеете каких-либо априорных гипотез относительно классов, но все еще находитесь в описательной стадии исследования. Следует понимать, что кластерный анализ определяет "наиболее возможно значимое решение".

Алгоритм древовидной кластеризации. Назначение этого алгоритма состоит в объединении объектов в достаточно большие кластеры, используя некоторую меру сходства или расстояние между объектами. Типичным результатом такой кластеризации является иерархическое дерево, которое представляет собой диаграмму. Диаграмма начинается с каждого объекта в классе (в левой части диаграммы). Теперь представим себе, что постепенно (очень малыми шагами) вы "ослабляете" ваш критерий о том, какие объекты являются уникальными, а какие нет. Другими словами, вы понижаете порог, относящийся к решению об объединении двух или более объектов в один кластер. В результате, вы связываете вместе всё большее и большее число объектов и агрегируете (объединяете) все больше и больше кластеров, состоящих из все сильнее различающихся элементов. Окончательно, на последнем шаге все объекты объединяются вместе. На этих диаграммах горизонтальные оси представляют расстояние объединения (в вертикальных древовидных диаграммах вертикальные оси представляют расстояние объединения). Так, для каждого узла в графе (там, где формируется новый кластер) вы можете видеть величину расстояния, для которого соответствующие элементы связываются в новый единственный кластер. Когда данные имеют ясную "структуру" в терминах кластеров объектов, сходных между собой, тогда эта структура, скорее всего, должна быть отражена в иерархическом дереве различными ветвями. В результате успешного анализа методом объединения появляется возможность обнаружить кластеры (ветви) и интерпретировать их.

Дискриминантный анализ используется для принятия решения о том, какие переменные различают (дискриминируют) две или более возникающие совокупности (группы). Наиболее общим применением дискриминантного анализа является включение в исследование многих переменных с целью определения тех из них, которые наилучшим образом разделяют совокупности между собой. Другими словами, вы хотите построить "модель", позволяющую лучше всего предсказать, к какой совокупности будет принадлежать тот или иной образец. В следующем рассуждении термин "в модели" будет использоваться для того, чтобы обозначать переменные, используемые в предсказании принадлежности к совокупности; о неиспользуемых для этого переменных будем говорить, что они "вне модели".

В пошаговом анализе дискриминантных функций модель дискриминации строится по шагам. Точнее, на каждом шаге просматриваются все переменные и находится та из них, которая вносит наибольший вклад в различие между совокупностями. Эта переменная должна быть включена в модель на данном шаге, и происходит переход к следующему шагу.

Можно также двигаться в обратном направлении, в этом случае все переменные будут сначала включены в модель, а затем на каждом шаге будут устраняться переменные, вносящие малый вклад в предсказания. Тогда в качестве результата успешного анализа можно сохранить только "важные" переменные в модели, то есть те переменные, чей вклад в дискриминацию больше остальных.

Эта пошаговая процедура "руководствуется" соответствующим значением F для включения и соответствующим значением F для исключения. Значение F статистики для переменной указывает на ее статистическую значимость при дискриминации между совокупностями, то есть, она является мерой вклада переменной в предсказание членства в совокупности.

Для двух групп дискриминантный анализ может рассматриваться также как процедура множественной регрессии. Если вы кодируете две группы как 1 и 2, и затем используете эти переменные в качестве зависимых переменных в множественной регрессии, то получите результаты, аналогичные тем, которые получили бы с помощью дискриминантного анализа. В общем, в случае двух совокупностей вы подгоняете линейное уравнение следующего типа:

Группа = a + b1*x1 + b2*x2 + ... + bm*xm

где a является константой, и b1...bm являются коэффициентами регрессии. Интерпретация результатов задачи с двумя совокупностями тесно следует логике применения множественной регрессии: переменные с наибольшими регрессионными коэффициентами вносят наибольший вклад в дискриминацию.

Если имеется более двух групп, то можно оценить более, чем одну дискриминантную функцию подобно тому, как это было сделано ранее. Например, когда имеются три совокупности, вы можете оценить: (1) - функцию для дискриминации между совокупностью 1 и совокупностями 2 и 3, взятыми вместе, и (2) - другую функцию для дискриминации между совокупностью 2 и совокупности 3. Например, вы можете иметь одну функцию, дискриминирующую между теми выпускниками средней школы, которые идут в колледж, против тех, кто этого не делает (но хочет получить работу или пойти в училище), и вторую функцию для дискриминации между теми выпускниками, которые хотят получить работу против тех, кто хочет пойти в училище. Коэффициенты b в этих дискриминирующих функциях могут быть проинтерпретированы тем же способом, что и ранее.

Каноническая корреляция.

Канонический анализ предназначен для анализа зависимостей между списками переменными. Если говорить точнее, он позволяет исследовать зависимость между двумя множествами переменных. При вычислении канонических корней подсчитывают собственные значения матрицы корреляций. Эти значения равны доле дисперсии, объясняемой корреляцией между соответствующими каноническими переменными. При этом полученная доля вычисляется относительно дисперсии канонических переменных, т.е. взвешенных сумм по двум множествам переменных; таким образом, собственные значения не показывают абсолютного значения, объясняемого в соответствующих канонических переменных.

Если извлечь квадратный корень из полученных собственных значений, получим набор чисел, который можно проинтерпретировать как коэффициенты корреляции. Поскольку они относятся к каноническим переменным, их также называют каноническими корреляциями. Как и собственные значения, корреляции между последовательно выделяемыми на каждом шаге каноническими переменными, убывают. Однако другие канонические переменные также могут быть значимо коррелированы, и эти корреляции часто допускают достаточно осмысленную интерпретацию.

Критерий значимости канонических корреляций сравнительно несложен. Во-первых, канонические корреляции оцениваются одна за другой в порядке убывания. Только те корни, которые оказались статистически значимыми, оставляются для последующего анализа. Хотя на самом деле вычисления происходят немного иначе. Программа сначала оценивает значимость всего набора корней, затем значимость набора, остающегося после удаления первого корня, второго корня, и т.д.

Исследования показали, что используемый критерий обнаруживает большие канонические корреляции даже при небольшом размере выборки (например, n = 50). Слабые канонические корреляции (например, R = .3) требуют больших размеров выборки (n > 200) для обнаружения в 50% случаев. Отметим, что канонические корреляции небольшого размера обычно не представляют практической ценности, поскольку им соответствует небольшая реальная изменчивость исходных данных.

Канонические веса. После определения числа значимых канонических корней возникает вопрос об интерпретации каждого (значимого) корня. Напомним, что каждый корень в действительности представляет две взвешенные суммы, по одной на каждое множество переменных. Одним из способов толкования "смысла" каждого канонического корня является рассмотрение весов, сопоставленных каждому множеству переменных. Эти веса также называются каноническими весами.

При анализе, обычно, пользуются тем, что чем больше приписанный вес (т.е., абсолютное значение веса), тем больше вклад соответствующей переменной в значение канонической переменной.

Если вы знакомы с множественной регрессией, вы можете применить для канонических весов интерпретацию, использованную для бета - весов в уравнении множественной регрессии. Канонические веса, в некотором смысле, аналогичны частным корреляциям переменных, соответствующих каноническому корню. Таким образом, рассмотрение канонических весов позволяют понять "значение" каждого канонического корня, т.е. увидеть, как конкретные переменные в каждом множестве влияют на взвешенную сумму (т.е. каноническую переменную).

Параметрические и непараметрические методы оценки результатов.

Параметрические методы, основанные на выборочном распределении определенной статистики. Говоря кратко, если вы знаете распределение наблюдаемой переменной, то можете предсказать, как в повторных выборках равного объема будет "вести себя" используемая статистика - т. е. каким образом она будет распределена.

В практике использование параметрических методов ограничено из-за объема или размера выборки доступной для анализа; проблем с точным измерением признаков наблюдаемого объекта

Таким образом, возникает необходимость в наличие процедур, позволяющих обрабатывать данные "низкого качества" из выборок малого объема с переменными, про распределение которых мало что или вообще ничего не известно. Непараметрические методы как раз и разработаны для тех ситуаций, достаточно часто возникающих на практике, когда исследователь ничего не знает о параметрах исследуемой популяции (отсюда и название методов - непараметрические). Говоря более специальным языком, непараметрические методы не основываются на оценке параметров (таких как среднее или стандартное отклонение) при описании выборочного распределения интересующей величины. Поэтому эти методы иногда также называются свободными от параметров или свободно распределенными.

По существу, для каждого параметрического критерия имеется, по крайней мере, один непараметрический аналог. Эти критерии можно отнести к одной из следующих групп:

Статистика

критерии различия между группами (независимые выборки);

критерии различия между группами (зависимые выборки);

критерии зависимости между переменными.

Различия между независимыми группами. Обычно, когда имеются две выборки (например, мужчины и женщины), которые вы хотите сравнить относительно среднего значения некоторой изучаемой переменной, вы используете t-критерий для независимых. Непараметрическими альтернативами этому критерию являются: критерий серий Вальда-Вольфовица, U критерий Манна-Уитни и двухвыборочный критерий Колмогорова-Смирнова. Если вы имеете несколько групп, то можете использовать дисперсионный анализ. Его непараметрическими аналогами являются: ранговый дисперсионный анализ Краскела-Уоллиса и медианный тест.

Различия между зависимыми группами. Если вы хотите сравнить две переменные, относящиеся к одной и той же выборке (например, математические успехи студентов в начале и в конце семестра), то обычно используется t-критерий для зависимых выборок. Альтернативными непараметрическими тестами являются: критерий знаков и критерий Вилкоксона парных сравнений. Если рассматриваемые переменные по природе своей категориальны или являются категоризованными (т.е. представлены в виде частот попавших в определенные категории), то подходящим будет критерий хи-квадрат Макнемара. Если рассматривается более двух переменных, относящихся к одной и той же выборке, то обычно используется дисперсионный анализ (ANOVA) с повторными измерениями. Альтернативным непараметрическим методом является ранговый дисперсионный анализ Фридмана или Q критерий Кохрена (последний применяется, например, если переменная измерена в номинальной шкале). Q критерий Кохрена используется также для оценки изменений частот (долей).

Зависимости между переменными. Для того, чтобы оценить зависимость (связь) между двумя переменными, обычно вычисляют коэффициент корреляции. Непараметрическими аналогами стандартного коэффициента корреляции Пирсона являются статистики Спирмена R, тау Кендалла и коэффициент Гамма Если две рассматриваемые переменные по природе своей категориальны, подходящими непараметрическими критериями для тестирования зависимости будут: Хи-квадрат, Фи коэффициент, точный критерий Фишера. Дополнительно доступен критерий зависимости между несколькими переменными так называемый коэффициент конкордации Кендалла. Этот тест часто используется для оценки согласованности мнений независимых экспертов (судей), в частности, баллов, выставленных одному и тому же субъекту.

Если данные не являются нормально распределенными, а измерения, в лучшем случае, содержат ранжированную информацию, то вычисление обычных описательных статистик (например, среднего, стандартного отклонения) не слишком информативно. Например, в психометрии хорошо известно, что воспринимаемая интенсивность стимулов (например, воспринимаемая яркость света) представляет собой логарифмическую функцию реальной интенсивности (яркости, измеренной в объективных единицах - люксах). В данном примере, обычная оценка среднего (сумма значений, деленная на число стимулов) не дает верного представления о среднем значении действительной интенсивности стимула. (В обсуждаемом примере скорее следует вычислить геометрическое среднее.) Непараметрическая статистика вычисляет разнообразный набор мер положения (среднее, медиану, моду и т.д.) и рассеяния (дисперсию, гармоническое среднее, квартильный размах и т.д.), позволяющий представить более "полную картину" данных.

Контрольные вопросы:

1.Поясните смысл коэффициентов частной и множественной корреляции.

2.Каковы условия построения уравнения множественной регрессии?

3.Каковы направления анализа на основе уравнения регрессии?

4. Как использовать уравнение регрессии для прогноза?

1. **единица [объект]**

Статистика

То, что можно рассмотреть и описать индивидуально.

Примечание - Единицей может, например, быть:

- изделие;

- определенное количество материала;

- услуга, действие или процесс;

- организация или человек;

- некоторая их комбинация.

2. **признак**

Свойство, которое помогает идентифицировать или различать единицы данной генеральной совокупности.

Примечание - Признак может быть количественным или качественным (альтернативным).

3. **(генеральная) совокупность**

Множество всех рассматриваемых единиц.

Примечание - Для случайной величины распределение вероятностей рассматривают как определение совокупности этой случайной величины.

4. **рамки отбора**

Список, заполняемый для выборочных целей, в котором отмечают те единицы, которые надо отобрать и исследовать.

5. **подсовокупность**

Определенная часть генеральной совокупности.

6. **наблюдаемое значение**

Значение данного признака, полученного в результате единичного наблюдения (см. п . 3.6).

7**. класс**

а) Для качественного признака - Определенные группы объектов, каждые из которых имеют отдельные общие признаки, взаимно исключают друг друга, исчерпывая все объекты.

b) Для количественного признака - Каждый из последовательных взаимоисключающих интервалов, на которые разделен весь интервал варьирования.

2.8. **границы класса;** *пределы класса*

Значения, определяющие верхнюю и нижнюю границы класса.

Примечания

1. Следует уточнить, какую из двух границ считают принадлежащей классу.

Статистика

2. Если возможно, надо чтобы граница класса не совпадала с возможным значением.

9. **середина класса**

Среднее арифметическое верхней и нижней границ класса для количественного признака.

10. **интервал класса**

Разница между верхней и нижней границами класса для количественного признака.

11. **частота**

Число наступлений события данного типа или число наблюдений, попавших в данный класс.

12. **накопленная кумулятивная частота**

Число наблюдений из множества, имеющих значения, которые меньше заданного значения или равны ему.

Примечание - Для данных, объединенных в классы, кумулятивную частоту можно указать только в границах класса.

13. **относительная частота**

Частота, деленная на общее число событий или наблюдений.

14. **кумулятивная относительная частота**

Кумулятивная частота, деленная на общее число наблюдений.

15. **распределение частот**

Эмпирическое отношение между значениями признака и его частотами или его относительными частотами.

Примечание - Это распределение можно представить графически в виде гистограммы, столбиковой диаграммы, полигона кумулятивных частот или как таблицу сопряженности двух признаков.

16. **одномерное распределение частот**

Распределение частот для единственного признака.

17. **гистограмма**

Графическое представление распределения частот для количественного признака, образуемое соприкасающимися прямоугольниками, основаниями которых служат интервалы классов, а площади пропорциональны частотам этих классов.

18. **столбиковая диаграмма**

Графическое представление распределения частот для дискретной случайной величины, образуемое набором столбцов равной ширины, высоты которых пропорциональны частотам.

19. **полигон кумулятивных частот**

Ломаная линия, получаемая при соединении точек, абсциссы которых равны верхним границам классов, а ординаты - либо кумулятивным абсолютным частотам, либо кумулятивным относительным частотам.

20. **двумерное распределение частот**

Эмпирическое отношение между парами значений или классами признаков с одной стороны, и их частотами с другой - для двух признаков, рассматриваемых одновременно.

21. **диаграмма разброса [рассеяния]**

Графическое представление множества точек, координаты которых x и y в обычной прямоугольной системе координат - это значения признаков X и Y.

Примечания

1. Множество из n элементов таким образом дает n точек, которые наглядно показывают зависимость между X и Y.

2. Концепцию диаграммы разброса можно распространить на более чем два признака.

22. **таблица сопряженности двух признаков**

Таблица, используемая для представления распределения двух признаков, в строках и столбцах которой указывают, соответственно, значения или классы первого и второго признаков, при этом на пересечении строки и столбца появляется частота, соответствующая данной комбинации значений или классов.

Примечание - Это понятие можно распространить на число признаков более двух.

23**. многомерное распределение частот**

Эмпирическое отношение между совместными наборами значений или классов признаков с одной стороны и их частотами с другой - для нескольких признаков, рассматриваемых одновременно.

24. **маргинальное распределение частот**

Распределение частот подмножества $k_1 < k$ признаков из многомерного распределения частот k признаков, когда остальные (k - k_1) переменных принимают любые значения из своих областей значений.

Примечания

1. Для k = 2 признаков маргинальное распределение частот можно получить, добавляя к каждому значению или классу значений рассматриваемого признака соответствующие частоты или относительные частоты остальных признаков.

2. В распределении частот трех признаков X, Y и Z существуют:

- три двумерных маргинальных распределения частот, то есть распределения пар (X, Y), (X, Z), (Y, Z);

- три одномерных маргинальных распределения частот, то есть распределения X, Y и Z.

25. **условное распределение частот**

Распределение частот k_1 < 1 признаков из многомерного распределения частот, когда остальные (k - k_1) признаков фиксированы.

Статистика

Примечания

1. Для k = 2 признаков условные распределения частот считывают непосредственно из строк и столбцов таблицы сопряженности двух признаков. Условное распределение относительных частот получают делением чисел в каждой строке (столбце) на общее число в соответствующей строке (столбце).

2. В распределении частот двух признаков X и Y:

- условное распределение частот X; конкретные распределения выражают как распределение X при $Y = y$;

- условное распределение частот Y; конкретные распределения выражают как распределение Y при $X = x$.

26. **среднее арифметическое**

Сумма значений, деленная на их число.

Примечания

1. Термин «среднее» обычно используют, когда имеют в виду параметр совокупности, а термин «среднее арифметическое» - когда имеют в виду результат вычислений по данным, полученным из выборок.

2. Среднее арифметическое простой случайной выборки, взятой из совокупности, - это несмещенная оценка арифметического среднего генеральной совокупности. Однако другие формулы для оценки, такие как геометрическое или гармоническое среднее, медиана или мода, иногда тоже используют.

27. **взвешенное среднее арифметическое**

Сумма произведений каждого значения на его вес, деленная на сумму весов, где веса - неотрицательные коэффициенты, связанные с каждым значением.

28. **выборочная медиана**

Если n случайных значений упорядочены по возрастанию и пронумерованы от 1 до n , то, если n нечетно, выборочная медиана принимает значение с номером $\left(\frac{n+1}{2}\right)$; если n четно, медиана лежит между $\frac{n}{2}$ -м и $\left(\frac{n}{2}+1\right)$ -м значениями и не может быть однозначно определена.

Примечание - При отсутствии других указаний и четном n за выборочную медиану можно принять среднее арифметическое этих двух значений.

29. **середина размаха (выборки)**

Среднее арифметическое между наибольшим и наименьшим наблюденными значениями количественного признака.

30. **размах (выборки)**

Разность между наибольшим и наименьшим наблюденными значениями количественного признака в выборке.

Статистика

31. **средний размах (выборок)**

Среднее арифметическое размахов множества выборок одинакового объема.

32. **среднее отклонение (выборки)**

Среднее арифметическое отклонение от начала координат, когда все отклонения имеют положительный знак.

Примечание - Обычно выбранное начало отсчета представляет собой среднее арифметическое, хотя среднее отклонение минимизируется, когда за начало отсчета принимают медиану.

33. **выборочная дисперсия**

Одна из мер рассеяния, представляющая собой сумму квадратов отклонений наблюдений от их среднего арифметического, деленная на число наблюдений минус единица.

Примечания

1. Для серии из n наблюдений $x_1, x_2, ..., x_n$ со средним арифметическим

$$\bar{x} = \frac{1}{n} \sum_i x_i$$

выборочная дисперсия

$$s^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2 .$$

2. Выборочная дисперсия - это несмещенная оценка дисперсии совокупности.

3. Выборочная дисперсия - это центральный момент второго порядка, кратный $n/(n-1)$ (п. 2.39 , примечание).

34. **выборочное стандартное отклонение**

Положительный квадратный корень из выборочной дисперсии.

Примечание - Выборочное стандартное отклонение - это смещенная оценка стандартного отклонения совокупности.

35. **выборочный коэффициент вариации** (Ндп. *относительное стандартное отклонение*)

Отношение выборочного стандартного отклонения к среднему арифметическому для неотрицательных признаков.

Примечание - Это отношение можно выразить в процентах.

36. **выборочный момент порядка *q* относительно начала отсчета**

Среднее арифметическое наблюдаемых значений в степени q в распределении единственного признака:

$$\frac{1}{n} \sum_i x_i^q ,$$

где n - общее число наблюдений.

Примечание - Момент первого порядка - это среднее арифметическое наблюдаемых значений.

37. **выборочный центральный момент порядка *q***

Среднее арифметическое разностей между наблюдаемыми значениями x_i и их средним арифметическим $\bar{x}$ в степени q в распределении единственного признака:

$$\frac{1}{n}\sum_i (x_i - \bar{x})^q ,$$

где n - число наблюдений.

Примечание - Выборочный центральный момент первого порядка равен нулю.

38. **выборочный совместный момент порядков *q* и *s* относительно начала отсчета**

В совместном распределении двух показателей - среднее арифметическое произведений x_i в степени q и y_i в степени s для всех наблюдаемых пар значений (x_i , y_i)

$$\frac{1}{n}\sum_i x_i^q \; y_i^s ,$$

где n - число наблюдаемых пар.

Примечания

1. Выборочный совместный момент порядков q и s - это один из моментов порядка (q + s).

2. Выборочный момент порядков 1 и 0 - это среднее арифметическое маргинального распределения частот X, а момент порядков 0 и 1 - среднее арифметическое маргинального распределения частот Y.

39. **выборочный совместный центральный момент порядков *q* и *s***

В совместном распределении двух признаков - среднее арифметическое произведений разности между x_i и его средним арифметическим значением $\bar{x}$ в степени q и разности между y_i и его средним арифметическим значением $\bar{y}$ в степени s для всех наблюдаемых пар (x_i , y_i):

$$\frac{1}{n}\sum_i (x_i - \bar{x})^q \; (y_i - \bar{y})^s ,$$

где n - число наблюдаемых пар.

Примечание - Выборочный центральный момент порядков 2 и 0 - это выборочная дисперсия маргинального распределения частот X, умноженная на (n - 1)/ n , а выборочный центральный момент порядков 0 и 2 - выборочная дисперсия маргинального распределения частот Y, умноженная на (n - 1)/ n .

40. **выборочная ковариация**

Сумма произведений отклонений x и y от их соответствующих средних арифметических, деленная на число наблюдаемых пар без единицы:

Статистика

$$S_{xy} = \frac{1}{n-1} \sum_i (x_i - \bar{x})(y_i - \bar{y}),$$

где n - число наблюдаемых пар.

Примечание - Выборочная ковариация - это несмещенная оценка ковариации совокупности.

41. **выборочный коэффициент корреляции**

Частное от деления выборочной ковариации двух показателей на произведение их выборочных стандартных отклонений:

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}},$$

где S_{xy} - выборочная ковариация X и Y;

S_x и S_y - выборочные стандартные отклонения X и Y соответственно.

Примечания

1. Этот коэффициент часто используют как цифровое выражение взаимной зависимости между X и Y в серии парных наблюдений. Для проверки линейности можно строить диаграмму разброса.

2. Его значения всегда лежат между минус 1 и плюс 1. Когда выборочный коэффициент корреляции равен одному из указанных пределов, это означает, что существует точная линейная зависимость в серии парных наблюдений.

3. Этот выборочный коэффициент корреляции применяют для измеряемых признаков; для ранговых данных используют другие коэффициенты корреляции, такие как коэффициенты Спирмена и Кендалла.

42. **кривая регрессии (*Y* по *X* для выборки)**

Для выборки *n* пар наблюдений двух показателей X и Y - кривая регрессии Y от X отображает зависимость функции Y от X .

43. **поверхность регрессии (*Z* по *X* и *Y* для выборки)**

Для выборки п наблюдений каждого из трех показателей X, Y и Z - поверхность регрессии Z от X и Y отображает зависимость функции Z от X и Y .

Примечание - Вышеуказанные определения можно распространить также на случай более трех показателей.

44. **выборочный коэффициент регрессии**

Коэффициент при переменной в уравнении кривой или поверхности регрессии.

45. **статистика**

Функция от выборочных значений.

Статистика

Примечание - Статистика как функция от выборочных значений - случайная величина, которая может принимать различные значения от выборки к выборке. Значение статистики, получаемое при использовании наблюдаемых значений, как их функция может быть использовано при проверке статистических гипотез или как оценка параметра совокупности, например среднего арифметического или стандартного отклонения.

46. **порядковая статистика**

Каждое из упорядоченных выборочных значений, расположенных в неубывающем порядке.

Примечания

1. В более общем выражении всякую статистику, основанную на порядковых статистиках в этом узком смысле, также называют порядковой статистикой.

2. k-е значение в неубывающей последовательности наблюдений $x_{|k|}$ - это значение случайной величины $X_{|k|}$, называемое k-й порядковой статистикой. В выборке объема n наименьшее наблюдаемое значение $x_{|1|}$ и наибольшее значение $x_{|n|}$ - это значения случайных величин $X_{|1|}$ и $X_{|n|}$ - первая и n -я порядковые статистики соответственно. Размах $x_{|n|}$ - $x_{|1|}$ - это значение порядковой статистики $X_{|n|}$ - $X_{|1|}$.

47. **тренд**

Тенденция к возрастанию или убыванию наблюдаемых значений, нанесенных на график в порядке их получения после исключения случайных ошибок и циклических эффектов.

48. **серия**

а) Появление в рядах наблюдений по качественному признаку непрерывающихся рядов одного и того же значения признака.

b) Последовательный набор монотонно возрастающих или монотонно убывающих значений в рядах наблюдений по количественному признаку.

Примечание - Последовательный набор монотонно возрастающих значений называют возрастающей серией, а монотонно убывающих значений - убывающей серией.

49. **оценивание (параметра)**

Операция определения на основе выборочных данных числовых значений параметров распределения, принятого в качестве статистической модели генеральной совокупности, из которой извлечена выборка.

Примечание - Результат этой операции может быть выражен как одним числовым значением, так и доверительным интервалом.

50. **оценка**

Статистика, используемая для оценивания параметра совокупности.

51. **значение оценки**

Значение параметра, полученное в результате оценивания.

52**. погрешность оценки**

Статистика

Разность (*T* - *q*) при оценивании параметра, где *T* обозначает результат оценки, а *q* - оцениваемый параметр.

Примечание - Погрешность при оценивании может включать в себя один или несколько из следующих компонентов:

- погрешность выборочного метода;

- погрешность измерения;

- округление значений или разделение на классы;

- другие погрешности .

53. **погрешность выборочного метода**

Часть погрешности при оценивании, обусловленная только тем, что объем выборки меньше, чем объем генеральной совокупности.

54. **смещение оценки**

Разность между математическим ожиданием оценки и значением оцениваемого параметра.

.55. **несмещенная оценка**

Оценка со смещением, равным нулю.

56. **стандартная ошибка;** *среднеквадратичная ошибка*

Стандартное отклонение оценки.

57. **двусторонний доверительный интервал**

Если T_1 и T_2 - две функции от наблюдаемых значений таких, что для оценки параметра распределения совокупности *q* вероятность $\Pr\left[T_1 \le \theta \le T_2\right]$ равна (1 - *a*), где (1 - *a*) - константа, положительная и меньше 1, то интервал между T_1 и T_2 - это двусторонний доверительный интервал для *q* при доверительной вероятности (1 - *a*).

Примечания

1. Границы T_1 и T_2 доверительного интервала - это статистики (2.45), которые в общих предположениях принимают различные значения от выборки к выборке.

2. В длинном ряду выборок относительная частота случаев, когда доверительный интервал накрывает истинное значение параметра совокупности *q* , больше или равна (1 - *a*).

58. **односторонний доверительный интервал**

Если *T* - функция от наблюдаемых значений такая, что для оценки параметра распределения совокупности *q* вероятность $\Pr\left(T \ge \theta\right)$ или вероятность $\Pr\left(T \le \theta\right)$ равна (1 - *a*), где (1 - *a*) - константа, положительная и меньше 1, то интервал от наименьшего возможного значения *q* до *T* или интервал от *T*до наибольшего возможного значения *q* - это односторонний доверительный интервал для *q* при доверительной вероятности (1 - *a*).

Примечания

Статистика

1. Граница T доверительного интервала - это статистика, которая в общих предположениях принимает различные значения от выборки к выборке.

2. См. п. 2.57 , примечание 2.

59. **доверительная вероятность;** *уровень доверия*

Величина (1 - a) - вероятность, связанная с доверительным интервалом или со статистически накрывающим интервалом.

Примечание - Величину (1 - a) часто выражают в процентах.

60. **доверительная граница**

Каждая из границ, нижняя T_1 , верхняя T_2 для двустороннего доверительного интервала или граница Tдля одностороннего интервала.

61. **толерантный интервал**

Интервал, для которого можно утверждать с данным уровнем доверия, что он содержит, по крайней мере, заданную долю определенной совокупности.

Примечание - Если определены обе границы по статистическим данным, то интервал двусторонний. Если одна из двух границ представляет собой бесконечность или ограничение области определения случайной величины, то интервал односторонний.

62. **толерантные границы**

Для двустороннего статистически накрывающего интервала - нижняя и верхняя границы этого интервала; для одностороннего статистически накрывающего интервала - значение статистики, ограничивающей этот интервал.

63. **критерий согласия распределения**

Мера соответствия между наблюдаемым распределением и теоретическим распределением, выбранным априори либо подобранным по результатам наблюдений.

64. **выбросы**

Наблюдения в выборке, отличающиеся от остальных по величине настолько, что возникает предположение, что они принадлежат другой совокупности или получены в результате ошибки измерения.

65. **статистический критерий**

Статистический метод принятия решений о том, стоит ли отвергнуть нулевую гипотезу в пользу альтернативной или нет.

Примечания

1. Решение о нулевой гипотезе принимают исходя из значений соответствующих статистик, лежащих в основе статистических критериев или рассчитанных по результатам наблюдений. Так как статистики - случайные величины, существует некоторый риск принятия ошибочного решения (п. 2.75 и п. 2.77).

2. Критерий априори предполагает, что проверяют некоторые предположения, например предположение о независимости наблюдений, предположение о нормальности и т.д.

66. **нулевая гипотеза и альтернативная гипотеза**

Утверждения относительно одного или нескольких параметров или о распределении, которые проверяют с помощью статистического критерия.

Примечания

1. Нулевая гипотеза (H_0) - предположение, обычно сложное, относят к утверждению, подвергаемому проверке, в то время как альтернативную гипотезу (H_1) относят к утверждению, которое будет принято, если нулевую гипотезу отвергают.

2. Проверка гипотезы о том, что математическое ожидание m случайной величины X в совокупности не меньше, чем заданное значение m_0:

$$H_0(\mu \geq \mu_0) \quad \text{и} \quad H_1(\mu < \mu_0).$$

3. Проверка гипотезы о том, что доли несоответствующих деталей в двух партиях p_1 и p_2 одинаковы (неодинаковы):

$$H_0(p_1 = p_2) \quad \text{и} \quad H_1(p_1 \neq p_2).$$

4. Проверка гипотезы о том, что случайная величина X имеет нормальное распределение с неизвестными параметрами. Альтернативная гипотеза - распределение не нормально.

67. **простая гипотеза**

Гипотеза, которая полностью задает распределение совокупности.

68. **сложная гипотеза**

Гипотеза, которая не полностью задает распределение совокупности.

Примечания

1. Это обычно гипотеза, которая включает в себя бесконечную систему простых гипотез.

2. В предположении нормального распределения гипотеза $m = m_0$ будет простой, если стандартное отклонение совокупности известно, но она будет сложной, если оно неизвестно.

3. Все гипотезы из примечаний, приведенных в п. 2.66 , сложные.

69. **свободный от распределения критерий**

Критерий, в котором функция распределения статистики, лежащей в основе критерия, не зависит от функции распределения наблюдений

70. **уровень значимости (критерия)**

Заданное значение верхнего предела вероятности ошибки первого рода. Примечание- Уровень значимости обычно обозначают a.

71. **критическая область**

Статистика

Множество возможных значений статистики, лежащей в основе критерия, для которого отвергают нулевую гипотезу.

Примечания

1. Критические области определяют таким образом, что если нулевая гипотеза верна, вероятность ее отбрасывания равна заданному значению *a* , обычно малому, например 5 % или 1 %.

2. Классический способ проверки нулевой гипотезы, относящийся к математическому ожиданию нормального распределения с известным стандартным отклонением *s* , H_0 (*m* ³ m_0) против альтернативы H_1 ($m < m_0$), - использование статистики $\bar{x}$ выборочного среднего арифметического.

Критическая область - это множество значений статистики, меньших чем

$$A = \mu_0 - \mu_{1-\alpha}\, \sigma/\sqrt{n}\ ,$$

где *n* - объем выборки;

m_{1-a} - это квантиль уровня (1 - *a*) стандартизованной нормальной случайной величины.

Если рассчитанное значение $\bar{x}$ меньше *A*, гипотезу H_0 отвергают. В противном случае - H_0 не отвергают (принимают).

72. **критическое значение**

Значение, ограничивающее критическую область.

73. **односторонний критерий**

Критерий, в котором используемая статистика одномерна, а критическая область включает в себя множество значений, меньших критического значения, или множество значений, больших критического значения.

74. **двусторонний критерий**

Критерий, в котором используемая статистика одномерна, а критическая область состоит из множества значений, меньших первого критического значения, и множества значений, больших второго критического значения.

Примечание - Выбор между односторонним и двусторонним критериями определяется альтернативной гипотезой. В примечании, приведенном в п. 2.71 , критерий односторонний, а критическое значение равно А.

75. **ошибка первого рода**

Ошибка, состоящая в отбрасывании нулевой гипотезы, поскольку статистика принимает значение, принадлежащее критической области, в то время как эта нулевая гипотеза верна.

76. **вероятность ошибки первого рода**

Вероятность допустить ошибку первого рода.

Примечания

Статистика

1. Она всегда меньше уровня значимости критерия или равна ему.

2. В примечании 2 к п. 2.71 ошибка первого рода состоит в отбрасывании H_0 ($m < m_0$), потому что $\bar{x}$ меньше *A*, в то время как на самом деле m равно или превышает m_0 . Вероятность такой ошибки равна *a*при $m = m_0$ и уменьшается с увеличением m .

77. **ошибка второго рода**

Ошибка принять нулевую гипотезу, поскольку статистика принимает значение, не принадлежащее критической области, в то время как нулевая гипотеза не верна.

78. **вероятность ошибки второго рода**

Вероятность допустить ошибку второго рода.

Примечание - Вероятность ошибки второго рода, обычно обозначаемая *b* , зависит от реальной ситуации и может быть вычислена лишь в том случае, если альтернативная гипотеза задана адекватно.

79. **мощность критерия**

Вероятность недопущения ошибки второго рода.

Примечания

1. Это вероятность отбрасывания нулевой гипотезы, когда она не верна. Ее обычно обозначают (1 - *b*).

2. В примечании 2 к п. 2.71 ошибка второго рода состоит в принятии гипотезы H_0 ($m\ ^3\ m_0$), поскольку $\bar{x}$ превышает *A*, в то время как на самом деле m меньше m_0 . Вероятность *b* такой ошибки зависит от фактического значения m : чем ближе m к m_0 , тем ближе мощность к 1.

3. В примечании 4 к п. 2.66 проверка нулевой гипотезы H_0 (нормально распределенная совокупность) против альтернативы H_1 (совокупность с ненормальным распределением) невозможно выразить *b* как функцию от альтернативной гипотезы, поскольку она не определена.

80. **функция мощности критерия**

Функция, которая определяет мощность критерия, обычно обозначаемую (1 - *b*) или (1 - *Pa*), при проверке гипотезы относительно значений скалярного параметра.

Примечание - Эта функция, определяемая для значений тех параметров, которые относятся к соответствующим альтернативным гипотезам, представляет собой вероятность отклонения нулевой гипотезы, когда она не верна.

Основные источники:

1. Громыка Г.Л. Теория статистики, М.: Инфра-М , 2001г. -160с.
2. Елисеева И.И. Статистика. М.: «Проспект». 2010г. – 463 с.
3. Елисеева И.И. Общая теория статистики: учебник для вузов / И.И. Елисеева, М.М. Юзбашев; под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2009. – 656 с.

Статистика

4. Ефимова М.Р. Практикум по общей теории статистики: учебное пособие для вузов / М.Р. Ефимова и др. – М.: Финансы и статистика, 2007. – 368 с.
5. Мелкумов Я.С. Социально-экономическая статистика: учебно-методическое пособие. – М.: ИМПЭ-ПАБЛИШ, 2007. – 200 с.
6. Общая теория статистики: Статистическая методология в изучении коммерческой деятельности: учебник для вузов / О.Э. Башина и др.; под ред. О.Э. Башиной, А.А. Спирина. - М.: Финансы и статистика, 2008. – 440 с.
7. Практикум статистика, Под редакцией проф. В.Н. Салина и доцента Е.Н. Шпековской. М.: КНОРУС, 2009г. - 496с.
8. Салин В.Н. Курс теории статистики для подготовки специалистов финансово-экономического профиля: учебник / В.Н. Салин, Э.Ю. Чурилова. – М.: Финансы и статистика, 2007. – 480 с.
9. Социально-экономическая статистика: практикум: учебное пособие / В.Н. Салин и др.; под ред. В.Н. Салина, Е.П. Шпаковской. – М.: Финансы и статистика, 2009. – 192 с.
10. Статистика, под редакцией В.С.Михитаряна, М.: Академия, 2003 г. – 271 с.
11. Статистика: учебник / И.И. Елисеева и др.; под ред. И.И. Елисеевой. – М.: Высшее образование, 2008. - 566 с.
12. Статистика: учебное пособие / А.В. Багат и др.; под ред. В.М. Симчеры. – М.: Финансы и статистика, 2007. – 368 с.
13. Теория статистики: учебник для вузов / Р.А. Шмойлова и др.; под ред. Р.А. Шмойловой. - М.: Финансы и статистика, 2007. – 656 с.
14. Харченко А.П. Статистика, М.: Инфра-М, 1997г. - 310с.
15. Шагойловой Р.А. Практикум по теории статистики М.: Финансы и Статистика, 1999г. - 411с.
16. Шмойлова Р.А. Практикум по теории статистики: учебное пособие для вузов / Р.А. Шмойлова и др.; под ред. Р.А. Шмойловой. - М.: Финансы и статистика, 2007. – 416 с.