

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ОРЕНБУРГСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ»**

Кафедра статистики и экономического анализа

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ
ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ**

Статистика (продвинутый уровень)

Направление подготовки (специальность) Экономика

Профиль подготовки (специализация) Экономическая безопасность

Квалификация (степень) выпускника магистр

Форма обучения очная

СОДЕРЖАНИЕ

1. КОНСПЕКТ ЛЕКЦИЙ.....	3
ЛЕКЦИЯ 1: СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ	3
ЛЕКЦИЯ 2: АНАЛИЗ РЯДОВ ДИНАМИКИ.....	10
ЛЕКЦИЯ 3: ПАРНАЯ КОРРЕЛЯЦИЯ	15
ЛЕКЦИЯ 4: МНОГОФАКТОРНЫЙ КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ.....	15
2. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРОВЕДЕНИЮ ПРАКТИЧЕСКИХ ЗАНЯТИЙ.....	22
ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 1: АНАЛИЗ РЯДОВ ДИНАМИКИ (ч.1).....	23
ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 2: АНАЛИЗ РЯДОВ ДИНАМИКИ (ч.2) Ошибка! Закладка не определена.	
ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 3: ПАРНАЯ КОРРЕЛЯЦИЯ	24
ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 4: МНОГОФАКТОРНЫЙ КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ	25

1. КОНСПЕКТ ЛЕКЦИЙ

ЛЕКЦИЯ 1: СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ

Вопросы лекции

1. Этапы статистического исследования.
2. Организация статистического наблюдения.
3. Формы, виды и способы наблюдения.
4. Методы контроля данных наблюдения.
5. Понятие о выборочном наблюдении.
6. Способы отбора единиц из генеральной совокупности.
7. Ошибка выборки.

Краткое содержание вопросов

1. Этапы статистического исследования

Использование различных способов и приемов статистической методологии предполагает наличие полной и достоверной информации об изучаемом объекте. Исследование массовых общественных явлений включает 3 основных этапа (стадии).

Этапы (стадии) статистического исследования

1. *Статистическое наблюдение* - сбор первичных данных (исходной статистической информации). Статистическое наблюдение позволяет охарактеризовать все многообразие условий и способов проявления общественных закономерностей и получить характеристику процесса в целом.

2. *Сводка, группировка данных*. Собранные данные подвергаются систематизации и группировке. На этой стадии совокупность делится по признакам различия и объединяется по признакам сходства, подсчитываются суммарные показатели по группам и в целом. С помощью метода группировок изучаемые явления делятся на важнейшие типы, характерные группы и подгруппы по существенным признакам. С помощью группировок ограничивают качественно однородные совокупности, что позволяет применять обобщающие показатели.

3. *Анализ статистических фактов и обнаружение закономерностей в изучаемых явлениях*. На данном этапе применяются все методы статистики: с помощью обобщающих показателей рассчитываются относительные и средние величины, дается сводная оценка вариации признаков, характеризуется динамика явлений, применяются индексы, балансовые построения, рассчитываются показатели, характеризующие тесноту связей признаков. Выводы и сам анализ излагаются текстом и, для большей наглядности, сопровождаются таблицами и графиками. Основой анализа статистических данных служит *выяснение экономической сущности* изучаемых явлений и процессов.

Все этапы взаимосвязаны. Если при сборе первичных статистических данных допущена ошибка или материал оказался недоброкачественным, это повлияет на правильность и достоверность как теоретических, так и практических выводов. Поэтому статистическое наблюдение от начальной до завершающей стадии должно быть тщательно продуманно и четко организовано.

2. Организация статистического наблюдения

Статистическое исследование начинают с точной формулировки его *цели, задач*, и *тех сведений*, которые могут быть получены в процессе наблюдения. Затем определяются *объект и единица наблюдения*, разрабатывается *программа*, выбираются *вид и способ* наблюдения.

Цель наблюдения должны быть сформулирована четко, ясно и развернуто, с указанием конкретных задач. Например, цель всероссийской переписи населения заключалась в получении данных о численности населения, его размещения по регионам, о составе населения по полу, возрасту, национальности, уровню образования, о

распределении по источникам средств существования и т.д.

Единица наблюдения - явление, признаки которого, подлежат регистрации.

Например, при переписи населения единицей наблюдения было домохозяйство. В зависимости от цели и задач исследования единицей наблюдения могут быть: предприятие, студент и т.д.

Объект наблюдения — совокупность единиц наблюдения. Объектом исследования может быть вид экономической деятельности (растениеводство, животноводство). При переписи населения объектом является население. Но необходимо уточнить, какое именно население подлежит регистрации — наличное, т. е. фактически находящееся в данной местности в момент переписи, или постоянное, т. е. живущее в данной местности постоянно. При обследовании промышленности необходимо точно установить, какие предприятия будут отнесены к промышленным. Для четкого ограничения наблюдаемой совокупности надо указать существенные признаки объекта наблюдения, отличающие его от других объектов.

В ряде случаев для отграничения объекта наблюдения пользуются тем или иным *цензом*. *Ценз* – значение ограничительного признака (признаков), позволяющее отделить единицы наблюдения от других явлений. Цензу должны удовлетворять все единицы изучаемой совокупности. Так, при переписи производственного оборудования нужно строго определить, что отнести к производственному оборудованию, а что к ручному инструменту, какое оборудование подлежит переписи — только действующее или также и находящееся в ремонте, на складе, резервное.

Наряду с определением единицы и объекта наблюдения важную сторону статистического исследования составляет разработка программы статистического наблюдения. *Программа наблюдения* — это перечень вопросов, по которым собираются сведения, либо перечень признаков и показателей, подлежащих регистрации. Программа наблюдения оформляется в виде бланка (анкеты, формуляра), в который заносятся первичные сведения. Необходимым дополнением к бланку является инструкция (или указания на самих формулярах), разъясняющая смысл вопроса. Состав и содержание вопросов программы наблюдения зависят от задач исследования и от особенностей изучаемого общественного явления.

Основные принципы составления программы:

1. Программа должна содержать только такие вопросы, которые безусловно необходимы для данного статистического исследования. Не следует загромождать программу излишними деталями. Чем обширнее проводимое исследование, тем короче должна быть программа.

2. В программу следует включать лишь те вопросы, на которые можно получить точные ответы. Часто для того, чтобы обеспечить единообразное толкование, пояснить вопрос отвечающему, дают подсказку.

3. Нельзя включать в программу вопросы, способные вызвать подозрение, что ответы на них могут быть использованы во вред опрашиваемым.

4. Программу наблюдения целесообразно строить так, чтобы ответами на одни вопросы можно было контролировать ответы на другие.

Организационные вопросы статистического наблюдения включают в себя определение места, времени, формы и способа наблюдения.

Время наблюдения (регистрации) – это то время, к которому относятся собираемые данные. Время регистрации для всех единиц наблюдения устанавливается единое – для предупреждения неполного учета или повторного счета, и для обеспечения сопоставимости. Момент времени, к которому приурочены регистрируемые сведения, называют *критическим моментом наблюдения*. Например, критическим моментом Всероссийской переписи населения 2010 г. был 0 часов 14 октября.

К статистическому наблюдению предъявляются требования:

1) полнота статистических данных (полнота охвата единиц изучаемой

совокупности, а также полнота охвата во времени);

2) достоверность данных;

3) сопоставимость данных.

3. Формы, виды и способы наблюдения

В статистической практике используются две организационные формы наблюдения — отчетность и специальное статистическое обследование.

Отчетность — форма организации сбора данных, при которой единицы наблюдения представляют сведения о своей деятельности в виде формуляров регламентированного образца. Отчетность присуща только государственной статистике. Она проводится в соответствии с федеральным планом статистических работ. Особенность отчетности состоит в том, что она обязательна, документально обоснована и юридически подтверждена подписью руководителя.

Специальные статистические обследования бывают сплошные и выборочные. Например, переписи населения — сплошное обследование. В СССР было осуществлено семь таких переписей: в 1920, 1926, 1939, 1959, 1970, 1979 и 1989 гг. В рыночной России — в 2002 и в 2010 гг. (планируется следующая перепись населения в 2020 г.). В результате обобщения данных переписей получали сведения о численности, размещении, составе населения по различным признакам. Эти данные важны для изучения вопросов социального и экономического развития, они также необходимы для анализа демографических процессов.

Выборочные обследования — проводятся Росстатом регулярно (например, бюджетные обследования домашних хозяйств 1 раз в квартал, ежемесячные обследования на рынке труда и др.).

Статистическое наблюдение подразделяется на виды — по времени наблюдения и по охвату единиц наблюдения.

По *времени* различают:

- непрерывное (текущее). Оно ведется систематически, постоянно, непрерывно по мере возникновения явления. Например, регистрируются в загсе рождения и смерти, браки и разводы, на предприятиях учитываются выпуск продукции, явки и неявки работников, расчеты с дебиторами и кредиторами, поступление денег в кассу и денежные выплаты и т. п.

- *периодическое*. При таком наблюдении регистрация проводится через определенные, обычно одинаковые промежутки времени, например учет успеваемости студентов по данным экзаменационных сессий.

- *единовременное наблюдение* проводится один раз для решения какой-либо задачи или повторяется через неопределенные промежутки времени по мере надобности, например перепись жилого фонда, школьная перепись и т. д.

Применение на практике того или иного вида наблюдения зависит от специфики исследуемого объекта. Так, функционирование общественного производства носит непрерывный характер: ежедневно производится и потребляется множество различных видов продукции, изменяются их запасы и т. д. Обеспечение бесперебойного производства требует непрерывного систематического учета затрат на производство и его результатов. Иной характер носят изменения в составе населения по социальному или национальному признаку, образованию и пр. В обычных условиях для больших групп населения эти признаки несущественно изменяются в короткие промежутки времени, поэтому нет надобности в непрерывной их регистрации.

По *охвату* фактов статистическое наблюдение может быть сплошным и несплошным. *Сплошное наблюдение* представляет собой полный учет всех без исключения единиц изучаемой совокупности (например, при переписи населения). *Несплошное наблюдение* организуют как учет части единиц совокупности, на основе которой можно получить обобщающую характеристику всей совокупности. С развитием многоукладной экономики данный вид наблюдения получил широкое распространение.

Несплошное наблюдение делится на 3 вида:

1) *способ основного массива*. При этом способе отбирают наиболее крупные единицы наблюдения, которые вносят наибольший вклад в изучаемое (например, обследование конъюнктуры торговых оборотов и цен на городских рынках).

2) *выборочное наблюдение*. При выборочном наблюдении обследованию подвергается отобранная в определенном порядке часть единиц совокупности, а получаемые результаты распространяются на всю совокупность. При этом получают информацию о всей совокупности, изучив лишь ее часть. Чтобы понять, хороший квас или нет, не обязательно выпить целую бочку, то же можно сказать в отношении проверки качества любой продукции. В решении такого рода задач, да и во многих других случаях, может помочь только выборка. Выборочный метод играет все большую роль в отечественной статистике. (Например, обследование малых предприятий.) Возникающие же случайные ошибки выборки можно определить с помощью теорем закона больших чисел и надлежащей организацией наблюдения свести их к допустимому минимуму.

3) *Монографическое наблюдение*. Оно позволяет подробно описывать отдельные единицы совокупности в целях их углубленного изучения, которое не может быть столь же детальным при массовом наблюдении. Главное внимание обращается на качественные стороны явления, его поведение, ориентацию, перспективы развития и т. д. Примерами монографических обследований являются этнографические обследования, когда изучается образ жизни семьи или нескольких семей.

Статистическое наблюдение может основываться на непосредственном учете фактов в процессе обследования, либо на документальном учете, либо на опросе респондентов.

При *непосредственном учете фактов* сведения получают путем личного учета единиц совокупности: пересчета, взвешивания, измерения и т. д.

Документальный способ сбора статистической информации базируется на систематических записях в первичных документах, подтверждающих тот или иной факт.

В ряде случаев для заполнения статистических формуляров прибегают к *опросу населения*.

При *экспедиционном способе* специально подготовленный счетчик опрашивает людей и с их слов заполняет бланк обследования. Работа счетчиков гарантирует единообразное понимание вопросов и максимальную правильность ответов.

При *анкетном наблюдении* определенному кругу лиц вручают специальные анкеты. Заполнение анкет носит добровольный характер и осуществляется анонимно. Это снижает полноту и достоверность получаемой информации. Поэтому данный способ наблюдения применяется в обследованиях, где не требуется высокая точность, а нужны приближенные результаты, например в социологических обследованиях и т. п.

При *корреспондентском способе* наблюдения рассылаются бланки обследования и указания к их заполнению с просьбой ответить на поставленные вопросы. После заполнения бланка анкеты организация или отдельное лицо высылают ее в адрес статистической организации, которая их рассылала.

Суть *способа саморегистрации* состоит в том, что обследуемому лицу вручают бланк обследования и разъясняют вопросы, бланк же обследуемое лицо заполняет самостоятельно. В назначенный день специально подготовленный работник посещает обследуемое лицо, получает заполненный бланк и проверяет полноту и правильность ответов.

4. Ошибки наблюдения. Методы контроля данных наблюдения

Проверка результатов начинается с:

- 1) проверки полноты охвата единиц наблюдением.
- 2) полноты заполнения формуляра наблюдения.

Все ошибки регистрации делятся на: случайные и систематические.

Методы контроля данных: счетный, логический.

5. Понятие о выборочном наблюдении

Тема «Выборочное наблюдение» взаимосвязана с такими темами, как «Статистическое наблюдение», «Статистические показатели» и др. Основываясь на законе больших чисел, выборочное наблюдение связано с математической статистикой и теорией вероятностей.

Статистическая методология исследования массовых явлений различает два способа наблюдения в зависимости от полноты охвата объекта: сплошное и несплошное. Разновидностью несплошного наблюдения является *выборочное*.

Под *выборочным наблюдением* понимается такое несплошное наблюдение, при котором статистическому обследованию подвергаются единицы изучаемой совокупности, отобранные случайным способом.

При выборочном наблюдении подлежащая изучению статистическая совокупность, из которой производится отбор, называется *генеральной совокупностью*. Отобранная из генеральной совокупности некоторая часть единиц, подвергающаяся обследованию, называется *выборочной совокупностью*.

Значение выборочного метода 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. При выборочном методе обследованию подвергается 5-10% единиц генеральной совокупности.

Таблица 1 - Обозначения статистических величин для выборки и генеральной совокупности

Название	Генеральная совокупность	Выборка
Число единиц наблюдения	N	n
Доля единиц, обладающих изучаемым признаком	p	W
Средняя величина	\bar{x}	x
Коэффициент корреляции	ρ	r
Дисперсия	σ^2	s^2

Примечание: в генеральной совокупности доля единиц, обладающих изучаемым признаком, называется генеральной долей, в выборке - выборочной долей, средняя величина называется генеральной средней, в выборке - выборочной средней.

Основная задача выборочного наблюдения в том, чтобы на основании характеристик выборочной совокупности w и x получить достоверные сведения о показателях p и \bar{x} в генеральной совокупности. Качество результатов выборочного наблюдения зависит от того, насколько выборка репрезентативна (представительна).

6. Способы отбора единиц из генеральной совокупности

Для обеспечения репрезентативности выборки необходимо соблюдение принципа случайности отбора единиц. Данный принцип предполагает, что все единицы имеют равную вероятность попасть в выборку. На включение объекта в выборку не может повлиять какой-либо иной фактор, кроме случайного.

На практике используют следующие *способы отбора единиц из генеральной совокупности*.

Собственно случайный отбор, или случайная выборка, осуществляется с помощью жеребьевки либо по таблице случайных чисел. Процесс формирования случайных чисел и определения номера отбираемой единицы продолжается до тех пор, пока не будет получен заданный объем выборочной совокупности.

Прежде, чем производить случайный отбор, необходимо убедиться, что все без исключения единицы генеральной совокупности имеют абсолютно равные шансы попадания в выборку. Следует также установить четкие границы генеральной совокупности таким образом, чтобы включение или не включение в нее отдельных единиц не вызывало сомнений. Так, например, при обследовании студентов необходимо указать,

будут ли приниматься во внимание лица, находящиеся в академическом отпуске, студенты негосударственных вузов, военных училищ и т.п.

На практике обычно применяют *механическое* формирование выборочной совокупности. При этом способе отбирается каждый (n/N) -й элемент генеральной совокупности. Например, если имеется совокупность из 100 тыс. ед. и требуется выборка в 1000, то в нее попадет каждый сотый элемент. Если единицы в совокупности не ранжированы относительно изучаемого признака, то первый элемент выбирается наугад, произвольно, а если ранжированы, то из середины первой сотни. При достаточно большой совокупности этот способ отбора близок к собственно случайному, при условии, что применяемый список не составлен таким образом, чтобы какие-то единицы совокупности имели больше шансов попасть в выборку.

Типическая выборка применяется, если генеральная совокупности неоднородна. При *типической выборке* генеральная совокупность разделяется на однородные типические группы с помощью типологической группировки, что обеспечивает репрезентативность. Из каждой группы собственно-случайной или механической выборкой производится индивидуальный отбор единиц в выборочную совокупность.

Особая форма составления выборки предполагает *серийный, или гнездовой*, отбор, при котором в порядке случайной или механической выборки выбирают не единицы, а определенные районы, серии (гнезда), внутри которых производится сплошное наблюдение.

Рассмотренные способы обычно применяются не в чистом виде, а комбинируются. Например, генеральная совокупность разбивается сначала на серии, далее в сериях производится случайный отбор.

Особенности обследуемых объектов определяют два метода отбора единиц в выборочную совокупность:

- повторный (отбор по схеме возвращенного шара);
- бесповторный (отбор по схеме невозвращенного шара).

При *повторном отборе* каждая попавшая в выборку единица или серия возвращается в генеральную совокупность и имеет шанс вторично попасть в выборку. При этом вероятность попадания в выборочную совокупность всех единиц генеральной совокупности остается одинаковой.

Бесповторный отбор означает, что каждая отобранная единица (или серия) не возвращается в генеральную совокупность и не может подвергнуться вторичной регистрации, а потому для остальных единиц вероятность попасть в выборку увеличивается.

Бесповторный отбор дает более точные результаты по сравнению с повторным, так как при одном и том же объеме выборки наблюдение охватывает больше единиц генеральной совокупности. Поэтому он находит более широкое применение в статистической практике. И только в тех случаях, когда бесповторный отбор провести нельзя, используется повторная выборка (при обследовании потребительского спроса, пассажирооборота и т. п.).

7. Ошибка выборки (ошибка репрезентативности)

Так как изучаемая совокупность состоит из единиц с варьирующими признаками, то состав выборочной совокупности отличается от состава генеральной совокупности. Это расхождение между характеристиками выборки и генеральной совокупности составляет *ошибку выборки*. Она зависит от:

- степени вариации изучаемого признака,
- численности выборки,
- методов отбора единиц в выборочную совокупность,
- принятого уровня достоверности.

Выборочная средняя и выборочная доля являются случайными величинами, которые могут принимать разные значения в зависимости от того, какие единицы

совокупности попали в выборку. Следовательно, ошибки выборки также являются случайными величинами и могут принимать различные значения. Поэтому определяют среднюю из возможных ошибок (μ):

$$\text{Для средней: } \mu_x = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma_x}{\sqrt{n}},$$

$$\text{Для доли: } \mu_w = \sqrt{\frac{p(1-p)}{n}}$$

В этих формулах σ^2 и $p(1-p)$ являются характеристиками генеральной совокупности, которые при выборочном наблюдении неизвестны. На практике для определения средней ошибки выборки обычно используют дисперсии выборочной совокупности s^2 . Так как, согласно закону больших чисел, выборочная совокупность при достаточно большом объеме достаточно точно воспроизводит характеристики генеральной совокупности. Следовательно, *средние ошибки выборки* можно представить следующим образом:

При повторном отборе:

$$\text{Для средней: } \mu_x = \sqrt{\frac{s^2}{n}} = \frac{s_x}{\sqrt{n}},$$

$$\text{где } s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n} \text{ или } s_x^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i} - \text{ по сгруппированным данным.}$$

$$\text{Для доли: } \mu_w = \sqrt{\frac{w(1-w)}{n}}.$$

При бесповторном отборе подкоренное выражение умножается на величину $(1 - n/N)$. В тех случаях, когда доля выборки незначительна и множитель $(1 - n/N)$ близок к единице, поправкой можно пренебречь.

При бесповторном отборе:

$$\text{Для средней: } \mu_x = \sqrt{\frac{s^2}{n} \left(1 - \frac{n}{N}\right)}.$$

$$\text{Для доли: } \mu_w = \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}.$$

Одно из возможных значений генеральной средней определяется по формуле:

$$\bar{x} = \tilde{x} \pm \mu_x$$

Для решения практических задач кроме средней ошибки рассчитывают *предельную ошибку выборки*, которая связана с уровнем вероятности. Уровень вероятности определяет величина нормированного отклонения t . Значения t даются в таблицах нормального распределения вероятностей.

Формулы для расчета предельной ошибки выборки

Метод отбора	Предельные ошибки индивидуального отбора	
	для средней	для доли
Повторный	$\Delta = t \sqrt{\frac{\tilde{\sigma}^2}{n}}$	$\Delta = t \sqrt{\frac{W(1-W)}{n}}$
Бесповторный	$\Delta = t \sqrt{\frac{\tilde{\sigma}^2}{n} \left(1 - \frac{n}{N}\right)}$	$\Delta = t \sqrt{\frac{W(1-W)}{n} \left(1 - \frac{n}{N}\right)}$

Формулы, приведенные в таблице, используются при определении ошибок выборки, осуществляемой собственно случайным или механическим методами.

ЛЕКЦИЯ 2: АНАЛИЗ РЯДОВ ДИНАМИКИ

Вопросы лекции

1. Понятие о динамике и динамических рядах.
2. Показатели, характеризующие динамику.
3. Методы выявления типа тенденции динамики.
4. Прогнозирование на основе уравнения тренда.

Краткое содержание вопросов

1. Понятие о динамике и динамических рядах.

Процесс развития, изменения во времени социально-экономических явлений в статистике называют *динамикой*. Изучение динамики дает возможность экономисту, менеджеру исследовать и прогнозировать поведение управляемого объекта во времени.

Ряды динамики получают в результате сводки и обработки материалов периодического статистического наблюдения. Повторяющиеся во времени (по отчетным периодам) значения одноименных показателей в ходе статистической сводки систематизируются в хронологической последовательности.

Значения показателя, составляющие ряд динамики, называются *уровнями ряда*. Чаще уровни ряда обозначают y, y_1, y_2, \dots, y_n .

В зависимости от вида исходных данных ряды динамики подразделяются на ряды *абсолютных, относительных и средних величин*.

Ряд динамики состоит из двух строк или столбцов. В зависимости от того, как уровни ряда выражают состояния явления от времени, различают *моментные и интервальные* ряды динамики.

Моментные ряды представляются в виде последовательности показателей, относящихся к конкретным моментам времени: на 1 января, на 1 июля и т. д.

Интервальные ряды представляются в виде последовательности значений показателей за определенный интервал времени: за январь, за день, за 1 квартал, за год и т.п.

Сравнивая уровни разных лет, мы замечаем, что в население области изменяется из года в год. Но за длительный период численность населения области возросла. Однако нередко уровень показателя следующего года оказывается ниже предыдущего. Иногда рост по сравнению с предыдущим годом велик, а иногда мал. Следовательно, рост численности населения наблюдается лишь в среднем, как *тенденция*. В отдельные же годы уровни испытывают *колебания*, отклоняясь от основной тенденции, что вызвано влиянием определенных факторов в разные годы.

Если рассматривать динамические ряды ежемесячных уровней производства молока, ряды объема продажи зимней одежды и обуви выявятся регулярно повторяющиеся из года в год *сезонные колебания* уровней.

При статистическом изучении динамики необходимо четко разграничить два основных элемента - тенденцию и колеблемость. *Тенденция* динамики связана с действием *долговременно* существующих причин и условий развития. *Колебания* же связаны с действием *краткосрочных* или *циклических* факторов, влияющих на отдельные уровни динамического ряда, и отклоняющих уровни от тенденции то в одном, то в другом направлении. Например, тенденция динамики урожайности связана с прогрессом агротехники, с укреплением экономики данной совокупности хозяйств, совершенствованием организации производства. Колеблемость урожайности вызвана чередованием благоприятных по погоде и неблагоприятных лет, циклами солнечной активности, колебаниями в развитии вредных насекомых и болезней растений.

Графическое изображение тенденции и колеблемости рядов динамики.

По оси X всегда отражается время, по оси Y – уровни ряда. Необходимо строго соблюдать масштаб, иначе характер динамики будет искажен.

Правила построения динамических рядов

1. Периодизация развития, т. е, расчленение изучаемого явления во времени на однородные этапы, однородные исторические периоды, в пределах которых показатель подчиняется одному закону развития.

2. Статистические данные должны быть сопоставимы по территории, кругу охватываемых объектов, единицам измерения, времени регистрации, **ценам**, методологии расчета.

3. Величины временных интервалов **должны соответствовать интенсивности** изучаемых процессов. Чем больше вариация уровней во времени, тем чаще следует делать замеры. Для стабильных процессов интервалы можно увеличить.

4. Числовые уровни рядов динамики должны быть **упорядоченными во времени**. Не допускается анализ рядов с пропусками отдельных уровней, если же такие пропуски неизбежны, то их восполняют условными расчетными значениями.

2. Показатели, характеризующие динамику

Для изучения развития явления во времени применяют систему показателей динамики. Рассмотрим показатели интенсивности изменения во времени:

- 1) абсолютное изменение;
- 2) ускорение;
- 3) темпы изменения;
- 4) темпы прироста;
- 5) абсолютное значение одного процента прироста.

Если сравнение проводится с начальным в ряду динамики периодом (моментом) времени, получают *базисные показатели*. Если сравнение производят с предыдущим уровнем (моментом), получают *цепные показатели*.

Абсолютное изменение уровней:

$$\text{цепное: } \Delta_i = y_i - y_{i-1}; \quad \text{базисное: } \Delta_i = y_i - y_1,$$

где $i = 1, \dots, n$.

Если абсолютное изменение отрицательно, его следует называть *абсолютным сокращением (абсолютной убылью)*. Абсолютное изменение имеет ту же единицу измерения, что и уровни ряда с добавлением единицы времени, за которую определено изменение: например, 22 тысячи тонн в год. Без указания единицы времени, за которую произошло измерение, абсолютный прирост нельзя правильно интерпретировать.

Ускорение - это разность между абсолютным изменением за данный период и абсолютным изменением за предыдущий период одинаковой длительности:

$$\Delta'_i = \Delta_i - \Delta_{i-1}.$$

Показатель абсолютного ускорения применяется только в цепном варианте, но не в базисном. Отрицательная величина ускорения говорит о замедлении роста или об ускорении снижения уровней ряда.

Темп изменения – это отношение сравниваемого уровня (более позднего) к уровню, принятому за базу сравнения (более раннему). Он говорит о том, во сколько раз больше (меньше) или сколько процентов составляет сравниваемый уровень по отношению к предыдущему:

$$\text{цепной темп изменения: } T_i = \frac{y_i}{y_{i-1}}; \text{ базисный: } T_i = \frac{y_i}{y_1}.$$

Темп изменения - величина всегда положительная.

Темп прироста: $T_{пр} = T - 1$ или $T - 100$.

Интерпретация: цепной $T_{пр} = 20\%$ означает «прирост показателя за год составил 20%».

$T_{пр}$ может иметь как положительные значения, так и отрицательные.

Если уровень ряда динамики принимает положительные и отрицательные значения, например финансовый результат от реализации продукции предприятием может быть прибылью (+), а может быть убытком (-), тогда темп изменения и темп прироста применять нельзя. В этом случае такие показатели теряют смысл и не имеют экономической интерпретации.

Соотношения между цепными и базисными показателями:

1) сумма цепных абсолютных изменений равна базисному абсолютному изменению:

$$\sum \Delta_{(цепн)} = \Delta_{(баз)},$$

2) произведение цепных темпов изменения равно базисному темпу изменения

$$\prod_{(i)} k_{(цепн)} = k_{(баз)},$$

Абсолютное значение 1% прироста равно сотой части предыдущего уровня:

$$Абс1\% i = y_{i-1} / 100.$$

Единицы измерения $Абс1\%$ те же, что и у изучаемого признака.

Особенности расчета показателей динамики для рядов, состоящих из относительных уровней

Если исходные данные выражены в % (структура, рентабельность, курс акций) то абсолютный прирост выражается в *процентных пунктах*.

Пример. Рентабельность продукции в хозяйстве:

2012 - 20%, 2013 - 17%.

$\Delta_{цеп} = 17\% - 20\% = -3$ п.п. (пункта). Это значит, рентабельность продукции в 2013 году по сравнению с 202 годом снизилась на 3 процентных пункта.

Средние показатели тенденции динамики

1. Средний уровень ряда,
2. Средний абсолютный прирост,
3. Средний темп роста,
4. средний темп прироста.

Средний уровень ряда - обобщающий показатель динамического ряда.

Для интервальных рядов с равными интервалами:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n},$$

где n - число уровней ряда.

Для интервальных рядов с неравными интервалами:

$$\bar{y} = \frac{\sum_{i=1}^n y_i \times t_i}{\sum_{i=1}^n t_i},$$

где t_i – продолжительность интервала времени.

Для моментных рядов внутри календарного периода используется *средняя хронологическая*. Например, для расчета среднегодового значения показателя по данным на первое число каждого месяца:

$$\bar{y} = \frac{y_{я}^t + y_{ф}^t + y_{м}^t + \dots + y_{д}^t + y_{я}^{t+1}}{12},$$

Где я, ф, м, ..., д – январь, февраль, март, ..., декабрь;

t – принадлежность показателя к текущему году, t+1 – значение показателя на 1 января следующего года.

Хронологическая средняя используется для расчета среднегодовой стоимости основных средств, оборотных средств и т.п.

Средний абсолютный прирост:

$$\bar{\Delta} = \frac{\sum \Delta_{ицце}}{n-1} = \frac{\Delta_{баз}}{n-1}.$$

Средний темп роста рассчитывается как корень (n-1)-й степени из произведения цепных темпов роста или из базисного темпа роста:

$$\bar{T} = \sqrt[n-1]{\prod T_{ицце}} = \sqrt[n-1]{T_{баз}}.$$

где n – число уровней ряда динамики.

Результат вычисления можно умножить на 100%.

Средний темп прироста (%) определяется по единственной методологии:

$$T_{пр} = T_p - 100.$$

3. Методы выявления типа тенденции динамики

Важной задачей, возникающей при анализе рядов динамики, является определение основной тенденции в развитии изучаемого явления.

Всякий ряд динамики теоретически может быть представлен в виде составляющих:

- 1) тренд—основная тенденция развития динамического ряда (к увеличению либо снижению его уровней);
- 2) циклические (периодические) колебания, в том числе сезонные;
- 3) случайные колебания.

Тенденция динамики связана с действием *долговременно* существующих причин и условий развития. Понятие «тренд» было введено английским ученым Гукером в 1902 году. Он предложил называть трендом уравнение, выражающее основную тенденцию динамического ряда.

Колебания же связаны с действием *краткосрочных* или циклических факторов, влияющих на отдельные уровни динамического ряда, и отклоняющих уровни от тенденции то в одном, то в другом направлении. Например, тенденция динамики урожайности связана с прогрессом агротехники, с совершенствованием организации производства. Колеблемость урожайности вызвана чередованием благоприятных по погоде и неблагоприятных лет, циклами солнечной активности, колебаниями в развитии вредных насекомых и болезней растений.

Для выявления типа уравнения тренда применяют ряд методов. Рассмотрим 2 метода.

1. *Сглаживание временного ряда с помощью скользящей средней.* Суть сглаживания сводится к замене фактических уровней временного ряда расчетными, которые в меньшей степени подвержены колебаниям. *Алгоритм сглаживания:*

1. Определяют длину интервала сглаживания ($l < n$). Удобнее, если l – нечетное число (обычно $l=3, 5, 7$). Чем шире интервал сглаживания, тем в большей степени поглощаются колебания и тенденция носит более плавный характер.
2. Весь период наблюдения разбивают на участки и интервал сглаживания «скользит» по ряду с шагом = 1.
3. Рассчитывают средние арифметические из уровней ряда на каждом участке:

$$\bar{y}_1 = \frac{y_1 + y_2 + y_3}{3}, \bar{y}_2 = \frac{y_2 + y_3 + y_4}{3}, \bar{y}_3 = \frac{y_3 + y_4 + y_5}{3} \text{ и т.д.}$$

4. Фактические значения ряда, стоящие в центре каждого участка, заменяют на полученные средние значения.

Таким образом, при вычислении этим способом средних уровней (звеньев скользящей средней) «скользят» по ряду динамики, при этом каждый раз отбрасывается один уровень в начале и добавляется следующий.

Пример.

Таблица 1 - Динамика потребления растительного масла в регионе

Годы	Тыс. кг	Средняя из суммы 3-х уровней ряда	Сглаженный уровень
1	175	175+263+326=764/3=254,7	-
2	263		-
3	326		254,7
4	297	263+326+297=886/3=295,3	295,3
5	247	326+297+247=870/3=290	290,0
6	298	842/3=280,7	280,7
7	366	911/3=	303,7
8	341	1005/3=	335,0
9	420	1127/3=	375,7
10	441	1202/3=	400,7
11	453	1314/3=	438,0
12	399	1293/3=	431,0
13	426	1278/3=	426,0
14	449	1274/3=	424,7
15	482	1357/3=	452,3
16	460	1391/3=	463,7

Недостаток данного метода состоит в условности определения сглаженных уровней для точек в начале и конце ряда. Необходимо иметь длинный ряд динамики.

2. *Аналитическое выравнивание.* Под этим понимают определение основной тенденции развития изучаемого явления. В итоге выравнивания временного ряда получают наиболее общий, суммарный, проявляющийся во времени результат действия всех причинных факторов. Отклонение конкретных уровней ряда от уровней, соответствующих общей тенденции, объясняют действием факторов, проявляющихся случайно или циклически. В результате приходят к трендовой модели:

$$Y_t = f(t) + \varepsilon_t, \text{ где}$$

$f(t)$ - теоретический уровень, определяемый тенденцией развития,

ε_t - случайное и циклическое отклонение от тенденции.

Целью аналитического выравнивания динамического ряда является определение аналитической или графической зависимости $f(t)$. Функцию $f(t)$ выбирают таким образом, чтобы она давала содержательное объяснение изучаемого процесса.

Данный метод предпочтительнее остальных, так как он определяет закон, по которому можно достаточно точно спрогнозировать значения уровней ряда.

Чаще всего при выравнивании используются следующие зависимости:

линейная $f(t) = a_0 + a_1t$;

параболическая $f(t) = a_0 + a_1t + a_2t^2$,

экспоненциальные $f(t) = e^{a_0 + a_1t}$ или $f(t) = e^{a_0 + a_1t + a_2t^2}$.

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.

Экспоненциальные зависимости применяются, если в исходном временном ряду наблюдается либо более или менее постоянный относительный рост, либо устойчивость в изменении показателей относительного роста (цепных темпов роста цепных же темпов роста).

Определение параметров тренда производится по методу наименьших квадратов (МНК).

Уравнение тренда еще называют уравнением *авторегрессии*, т.е. зависимости уровней от изменения одного параметра - времени.

Интерпретация параметров уравнения тренда

Для линейной зависимости $f(t) = a_0 + a_1t$:

параметр a_0 интерпретации не имеет, но иногда его рассматривают как обобщенный начальный уровень ряда; a_1 — среднегодовой абсолютный прирост, показывающий, на сколько изменится результат при изменении времени на единицу.

Если $a_1 > 0$ значит наблюдаем тенденцию возрастания уровней ряда на изучаемом отрезке времени. Если $a_1 < 0$ - снижение уровней.

При параболической зависимости: $f(t) = a_0 + a_1t + a_2t^2$

Если $a_2 < 0$ ветви параболы направлены вниз, значит со временем рост уровней приостановится и даже будет снижаться со все большей скоростью (такой характер развития свойственен производству устаревшей продукции). Необходимо проверить, преодолел ли изучаемый уровень вершину параболы (приравнять первую производную к нулю).

Оценка надежности уравнения тренда

Это делается с помощью критерия Фишера (F-критерия) или коэффициента достоверности аппроксимации (R^2).

$$F_{расч} = \frac{s_t^2 (n - p)}{(1 - s_t^2) * (p - 1)},$$

где n - число наблюдений, p - число параметров уравнения тренда;

s_t - среднее квадратическое отклонение фактических уровней от линии тренда.

4. Прогнозирование на основе уравнения тренда

Статистическое прогнозирование по уравнению тренда основано на *экстраполяции*, т.е. на предположении, что параметры тренда сохраняются до прогнозируемого периода. Такая экстраполяция справедлива, если система развивается в относительно стабильных условиях, без резких изменений тенденции.

Прогнозирование по тренду производят в два этапа:

1) вычисляется «точечный прогноз»;

2) определяется доверительный интервал прогноза с заданной вероятностью.

«Точечный прогноз» - значение тренда (y) при подстановке в уравнение номера прогнозируемого года. Т.е. если мы строили уравнение тренда на основе ряда из 10 уровней (с 2004 по 2013 гг.), то чтобы получить точечный прогноз на 2014 год, достаточно

подставить в уравнение тренда вместо t порядковый номер 2014 года в данном ряду (11). Однако параметры тренда, полученные по ограниченному числу уровней ряда - лишь выборочные средние оценки. При изменении длины ряда можно получить иное уравнение. Поэтому прогноз необходимо дополнить доверительным интервалом.

Статистический прогноз с учетом доверительного интервала выглядит так: «точечный прогноз» $\pm \alpha$, где α - доверительный интервал прогноза.

$$\alpha = m * t_{\text{Стьюдента}}$$

упрощенная формула: $m = \frac{S_t}{\sqrt{n}}$

для однократного выравнивания: $m = S_t * \sqrt{\frac{1}{N} + \frac{t_k^2}{\sum_{i=1}^N t_i^2}}$, где

S_t - среднее квадратическое отклонение фактических значений от тренда.

t_k - номер года прогноза, t_i - по всей длине ряда N (с учетом прогнозного года).

t -критерий Стьюдента определяется по таблице с указанной исследователем вероятностью (обычно 90, 95 или 99%) и при числе степеней свободы, равном $(n-p)$.

ЛЕКЦИЯ 3: ПАРНАЯ КОРРЕЛЯЦИЯ

Вопросы лекции

1. Сущность и задачи корреляционно-регрессионного анализа.
2. Параметрические показатели связи.
3. Статистическая оценка надежности параметров парной линейной корреляции.
4. Непараметрические показатели связи.

Краткое содержание вопросов

1. Сущность и задачи корреляционно-регрессионного анализа

Исследуя природу, общество, экономику, необходимо считаться со взаимосвязью наблюдаемых процессов и явлений. Оценка наиболее существенных из них, а также силы воздействия одних факторов на другие является одной из основных задач статистики.

Формы проявления взаимосвязей разнообразны. Выделяют функциональную (полную) и корреляционную (неполную) связи. При функциональной зависимости величине факторного признака строго соответствует одно или несколько значений функции (формулы из физики, химии, в экономике - прямо пропорциональная зависимость между производительностью труда и увеличением производства продукции).

Корреляционная связь (или неполная, или статистическая) проявляется в среднем, для массовых наблюдений, когда значениям зависимой переменной соответствует некоторый ряд вероятных значений независимой переменной (переменных). Объяснение тому – сложность взаимосвязей между анализируемыми факторами, на взаимодействие которых влияют неучтенные случайные величины. Поэтому связь между признаками проявляется лишь в среднем, в массе случаев. При корреляционной связи каждому значению аргумента соответствуют случайно распределенные в некотором интервале значения функции.

Например, в сельском хозяйстве это может быть связь между урожайностью и количеством внесенных удобрений. Очевидно, что удобрения участвуют в формировании урожая. Но для каждого конкретного поля, участка одно и то же количество внесенных удобрений вызовет разный прирост урожайности, так как во взаимодействии находится еще целый ряд факторов (погода, состояние почвы и др.), которые и формируют конечный результат. Однако в среднем такая связь наблюдается – увеличение массы внесенных удобрений ведет к росту урожайности.

Слово *корреляция* ввел в употребление в статистику английский биолог и статистик Френсис Гальтон в конце XIX в. Тогда оно писалось как «*corelation*» (соответствие), но не

просто «связь» (*relation*), а «как бы связь», т. е. связь, но не в привычной в то время функциональной форме.

Признак, характеризующий *следствие* называют *результативным*, признак, характеризующий *причину* - *факторным*.

По направлению связи бывают *прямыми* (*положительными*), когда зависимая переменная (результативный признак) растет с увеличением факторного признака, и *обратными* (*отрицательными*), когда рост факторного признака вызывает уменьшение функции.

По своей аналитической форме связи *бывают линейными и нелинейными*.

Если характеризуется связь двух признаков, то ее называют *парной*. Если изучаются более чем две переменные — *множественной*.

По силе различаются *слабые, средние и тесные* связи.

Задачи статистики в области изучения взаимосвязей:

- 1) измерение параметров уравнения, выражающего статистическую связь;
- 2) измерение тесноты связи признаков между собой.

Для решения задач применяются две группы методов, одна из которых включает в себя методы корреляционного анализа, а другая - регрессионный анализ. В то же время ряд исследователей объединяет эти *методы* в *корреляционно-регрессионный анализ*, что имеет под собой некоторые основания: наличие целого ряда общих вычислительных процедур, взаимодополнения при интерпретации результатов и др.

Последовательность (методика) проведения корреляционно-регрессионного анализа

1. Установление причинно-следственной связи (экономическое обоснование выбора факторов).
2. Оценка тесноты связи между признаками с помощью коэффициентов связи.
3. Установление математической формы связи (выбор вида уравнения регрессии).
4. Решение уравнения связи и определение его параметров.
5. Экономическая интерпретация результатов. При выводах следует обращать внимание на возможность появления *ложной корреляции*.

Следует заметить, что традиционные методы корреляции и регрессии широко представлены в статистических пакетах программ для персональных компьютеров. Исследователю остается только правильно подготовить информацию, выбрать удовлетворяющий требованиям анализа пакет программ и быть готовым к интерпретации полученных результатов.

Условия применения и ограничения корреляционно-регрессионного метода

1. Наличие данных по достаточно большой совокупности. Считают, что число наблюдений должно быть не менее, чем в 5-6, а лучше – в 10 раз больше числа факторов. Чем больше наблюдений, тем лучше: закон больших чисел обеспечивает погашение случайных отклонений от средней величины.

2. Качественная однородность совокупности. При этом закономерность надежно проявляется в средней величине, что позволяет выявить закономерность корреляционной связи.

2. Параметрические показатели связи

В корреляционном анализе установление тесноты связи производится параметрическими и непараметрическими показателями. Рассмотрим *параметрические* показатели связи.

1. *Эмпирическое корреляционное отношение* – рассчитывается по результатам группировки:

$$\eta = \sqrt{\frac{\sigma_{\text{факт}}^2}{\sigma_{\text{общ}}^2}},$$

где $\sigma_{\text{факт}}^2$ - факторная (межгрупповая) дисперсия;

$\sigma_{\text{общ}}^2$ - общая дисперсия.

Правило сложения дисперсий: $\sigma_{\text{общ}}^2 = \sigma_{\text{факт}}^2 + \sigma_{\text{ост}}^2$, где

$\sigma_{\text{ост}}^2$ - остаточная дисперсия.

Показатель изменяется в интервале (0;1), может применяться для измерения нелинейной зависимости между признаками. Его недостаток - он не показывает направление связи, как коэффициент линейной корреляции.

2. Коэффициент линейной (парной) корреляции. Предложен англичанином Карлом Пирсоном:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}},$$

где x_i - индивидуальное значение факторного признака,

y_i - индивидуальное значение результативного признака,

\bar{x}, \bar{y} - средние значения факторного и результативного признаков.

Коэффициент корреляции не зависит от принятых единиц измерения, следовательно он сравним для любых признаков. Коэффициент корреляции еще называют стандартизированным коэффициентом регрессии.

Интервалы значений r :

$r \in (-1, 1)$

Если $r > 0,7$ - связь сильная (тесная),

При $0,5 < r < 0,7$ - связь средней тесноты,

При $0,3 < r < 0,5$ связь слабая,

При $r < 0,3$ связь очень слабая, статистически несущественная.

Значение r надо округлять до 2-3 цифр после запятой.

3. Коэффициент детерминации - квадрат коэффициента корреляции:

$R^2 \in (0; 1)$

Коэффициент детерминации показывает долю вариации результативного признака, вызванную вариацией факторного признака. Например, если r^2 между урожайностью зерновых и количеством внесенных удобрений = 0,61 (или 61%), это означает, что вариация урожайности на 61 % обусловлена влиянием дозы удобрений.

3. Статистическая оценка надежности параметров парной корреляции

Показатели корреляционной связи, вычисленные по ограниченной совокупности (по выборке), являются лишь оценками статистической закономерности изучаемого процесса. Поэтому необходима статистическая оценка степени точности и надежности параметров корреляции. Под надежностью понимается вероятность того, что значение проверяемого параметра не равно нулю.

Надежность линейного коэффициента корреляции (r) проверяется с использованием t -критерия Стьюдента:

$$t_{\text{расч}} = \frac{r}{m_r}, \text{ где } m_r - \text{средняя случайная ошибка коэффициента корреляции.}$$

Средняя случайная ошибка коэффициента корреляции:

$$m_r = \frac{1 - r^2}{\sqrt{n - 2}}, \text{ где } n - \text{число наблюдений.}$$

Расчетный t -критерий необходимо сравнить с табличным значением ($t_{\text{табл}}$) при выбранном вами уровне значимости (0,01, 0,05, 0,1).

Если $t_{\text{расч}} > t_{\text{табл}}$, то коэффициент линейной корреляции достоверен с определенной вероятностью.

Если $t_{\text{расч}} < t_{\text{табл}}$, то говорят: "связь надежно не установлена".

4. Непараметрические показатели связи

Непараметрические показатели рассчитаны на основе определения рангов (ранжирования), в зависимости от вариации каждого из признаков. Такие показатели используются при изучении взаимосвязи качественных, атрибутивных или дискретных признаков. К ним относятся:

- коэффициент корреляции знаков Фехнера;
- коэффициент ассоциации и коэффициент контингенции (только для альтернативных признаков);
- коэффициенты взаимной сопряженности Пирсона и Чупрова (для качественных и количественных признаков, если признак может принимать несколько значений);
- коэффициенты рангов Спирмена и Кендэла;
- коэффициент конкордации, основанный на личном мнении, используется при экспертной оценке и характеризует степень согласия экспертов.

Рассмотрим некоторые из перечисленных показателей подробнее.

1. Коэффициент Фехнера:

$$K_f = (C - H) / n,$$

Где С, Н - число наблюдений, у которых по паре признаков X и Y наблюдается совпадение (С) или несовпадение (Н) знаков отклонения от средних уровней.

Kf показывает наличие и направление связи. Если $K_f > 0$, то связь прямая, если $K_f < 0$, то связь обратная.

1. Коэффициент корреляции рангов.

2.1. по формуле Спирмена (для линейной зависимости):

$$K_c = 1 - \frac{\sum_{i=1}^n d_i^2}{n^3 - n}, \text{ где } d_i = \text{ранг}(x_i) - \text{ранг}(y_i).$$

Порядок расчета Kc:

1) *поранжировать значения X и Y в одном и том же порядке (либо по возрастающей, либо по убывающей);*

2) *найти разности рангов d_i ;*

3) *получить результат по формуле.*

1.2. по формуле Кэндалла (для нелинейной зависимости):

$$K_k = \frac{2S}{n(n-1)}, \text{ где } S - \text{фактическая сумма рангов.}$$

Если у нескольких единиц наблюдения величина признака будет одинаковой, то их нумеруют подряд и присваивают среднее из этих рангов.

ЛЕКЦИЯ 4: МНОГОФАКТОРНЫЙ КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ

Вопросы лекции

1. Уравнение регрессии.
2. Корреляционно-регрессионные модели.
3. Стандартизированные коэффициенты регрессии.

Краткое содержание вопросов

1. Уравнение регрессии

Уравнение регрессии может быть *парным* или *множественным*. Если изучается влияние одного фактора на результативный признак, то получают уравнение *парной* регрессии. При изучении влияния множества факторов на результативный признак получают уравнение *множественной* регрессии.

Регрессия бывает *линейной* (выражается уравнением прямой) или *криволинейной* (выражается уравнениями параболы, гиперболы и т.д.).

Линейная регрессия применяется чаще всего.

Уравнение парной линейной регрессии: $Y = a + bx$

Уравнение множественной регрессии: $Y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$.

где y – теоретическое (расчетное) значение результативного признака,

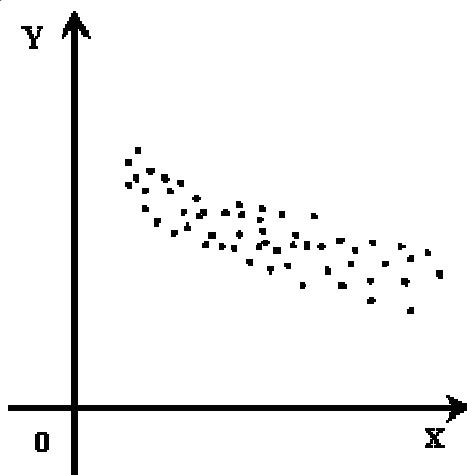
a, b_1, b_2, \dots, b_n – параметры уравнения регрессии,

b_1, b_2, \dots, b_n – коэффициенты чистой регрессии.

Коэффициенты чистой регрессии измеряют среднее отношение отклонения результативного признака от его средней величины к отклонению факторного признака от его средней величины на одну единицу его измерения. *Коэффициенты регрессии* выражаются в тех же единицах измерения, что и признаки-факторы. *Параметр a* показывает остаточное влияние факторов, не включенных в уравнение регрессии. Переменные x_1, x_2, x_n – значения факторного признака.

Для того, чтобы правильно *выбрать форму уравнения регрессии* необходимо:

1. Изобразить графически изучаемое распределение X и Y (т.е. построить корреляционное поле) – рисунок.



2. Путем перебора различных видов уравнений, выбрать тот вид уравнения регрессии, для которого сумма квадратов отклонений фактических значений от теоретических минимальна: $\sum(y_i - \hat{y})^2 \rightarrow \min$ (методом наименьших квадратов).

3. Оценить надежность полученного уравнения.

Параметры уравнения регрессии определяются с помощью *метода наименьших квадратов (МНК)*. МНК заключается в решении системы нормальных уравнений. Число уравнений в системе зависит от числа параметров уравнения регрессии. Для линейной регрессии:

$$\begin{cases} an + b\sum x_i = \sum y_i \\ a\sum x_i + b\sum x_i^2 = \sum y_i x_i \end{cases}$$

Необходимо определить a и b , решив систему уравнений.

Оценку правильности выбора вида уравнения регрессии и характеристику значимости всего уравнения получают с помощью F -критерия Фишера, который представляет собой отношение большей дисперсии к меньшей, рассчитанных на одну степень свободы, или по формуле:

$$F_{\text{расч}} = [R^2(n-m)] / [(1-R^2)(m-1)], \text{ где}$$

n – число наблюдений, m – число параметров уравнения регрессии.

Если $F_{\text{расч}} > F_{\text{табл}}$, то вид уравнения выбран верно.

Если $F_{\text{расч}} < F_{\text{табл}}$, то следует пересмотреть форму уравнения, перечень переменных и т.д.

Оценку значимости параметров уравнения регрессии проводят также, как для коэффициента линейной корреляции (с помощью t-критерия Стьюдента).

Экстраполяция регрессионных уравнений - расчет ожидаемых (прогнозируемых) значений результативного признака.

2. Корреляционно-регрессионные модели

Корреляционно-регрессионная модель - аналитическая форма множественной или парной корреляции.

Корреляционно-регрессионной моделью считается такое уравнение регрессии, которое включает основные факторы, влияющие на вариацию результативного фактора, обладает высоким (не ниже 0,5) коэффициентом детерминации и статистически значимыми и экономически обоснованными коэффициентами регрессии.

Рекомендации для построения корреляционно-регрессионной модели

1. Признаки-факторы должны находиться в причинной связи с результативным признаком. Поэтому, недопустимо, например, в модель себестоимости вводить в качестве одного из факторов коэффициент рентабельности, хотя включение такого фактора значительно повышает коэффициент детерминации.
2. Признаки-факторы не должны быть составными частями результативного признака или его функциями (в этом случае коэффициент корреляции равен единице). Например, при анализе производительности труда в промышленности нельзя как признак-фактор использовать индексы физического объема промышленной продукции, так как они входят в формулу для расчета производительности труда.
3. Не рекомендуется включать в уравнение факторы слабо связанные с результативным признаком, но тесно связанные между собой. Например, если $r_{yx1} = 0,55$, $r_{yx2} = 0,67$, $r_{x1x2} = 0,88$, то в регрессионное уравнение следует включить фактор x_2 , а фактор x_1 не включать, так как он тесно связан с x_2 , и его корреляция с y слабее, чем корреляция x_2 .
4. Нельзя включать в модель факторы разных уровней иерархии, т. е. фактор ближайшего порядка и его субфакторы. Например, в моделях себестоимости зерна не следует включать и урожайность зерновых культур, и дозу удобрений под них или затраты на обработку гектара, показатели качества семян, плодородия почвы, т. е. субфакторы самой урожайности.
5. Необязательное, но желательное условие, чтобы между результативным и факторными признаками соблюдалось единство единицы совокупности, к которой они отнесены. Т.е. если результативный признак рассчитан в целом по области, то все факторы также должны относиться к областному уровню. Однако, при анализе влияния факторов на розничный товарооборот области допустимо рассматривать фактор задолженности по заработной плате в промышленности. Так, размер среднемесячной начисленной заработной платы в промышленности Оренбургской области в среднем в два раза больше среднемесячной начисленной заработной платы в целом по области. Невыплата заработной платы на промышленных предприятиях отражается на величине розничного товарооборота.
6. При выборе между двумя корреляционно-регрессионными моделями предпочтительнее модель с меньшим числом факторов при том же коэффициенте детерминации или даже при несущественно меньшем коэффициенте. Но величина коэффициента детерминации в полученной модели должна быть не менее 0,5.

Для изучения множественной корреляции и отбора факторов в модель используется *матрица парных коэффициентов корреляции*. При этом необходимо иметь в виду, что в экономических исследованиях часто проявляется эффект взаимодействия (взаимозаменяемости) факторов, т.е. мультиколлинеарность (или по двум факторам - коллинеарность). Поэтому факторы, имеющие $r > 0,8$ в модель не включаются (см. п. 3 Рекомендаций построения КРМ).

Чаще всего уравнение множественной регрессии имеет *линейную* форму.

$$Y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n = a + \sum b_jx_j, \text{ где}$$

j - число факторных признаков.

3. Стандартизированные коэффициенты регрессии

Так как *коэффициенты чистой регрессии* выражаются в тех же единицах измерения, что и признаки-факторы, невозможно узнать, какой же фактор сильнее воздействует на результат. Поэтому необходимо *коэффициенты чистой регрессии* выразить в стандартизированной форме: в виде β -коэффициентов и коэффициентов эластичности (\mathcal{E}).

Расчет β -коэффициентов:

$$\beta_1 = b_1 * \frac{\sigma_1}{\sigma_y}; \beta_2 = b_2 * \frac{\sigma_2}{\sigma_y}; \dots$$

β -коэффициент показывает, на сколько средних квадратических отклонений изменится в среднем результативный признак, если соответствующий фактор изменится на свое среднее квадратическое отклонение.

Расчет коэффициентов эластичности:

$$\mathcal{E}_1 = b_1 * \frac{x_1}{x_y}; \mathcal{E}_2 = b_2 * \frac{x_2}{x_y}; \dots$$

Коэффициентов эластичности показывает, на сколько процентов изменится в среднем результативный признак, если фактор изменится на один процент.

Соотношение между β_i и \mathcal{E}_i : $\beta_i / \mathcal{E}_i = v_{x_i} / v_y$.

β -коэффициенты всегда меньше коэффициентов эластичности.

2. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРОВЕДЕНИЮ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 1: СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ

1. Задания для работы

Вопросы к занятию:

1. Требования, предъявляемые к исходным данным.
2. Программа статистического наблюдения.
3. Организационные формы статистического наблюдения
4. Виды статистического наблюдения.
5. Подготовка статистического наблюдения.
6. Ошибки статистического наблюдения.

Типовые тесты (для контроля усвоения знаний)

1. Особая форма сбора данных, присущая только государственной статистике:
 - а) анкетирование;
 - б) статистическая отчетность;
 - в) способ основного массива;
 - г) монографическое обследование.
2. Непрерывным (текущим) наблюдением является:
 - а) перепись населения;
 - б) учет успеваемости студентов по итогам сессий;
 - в) сельскохозяйственная перепись;
 - г) регистрация браков и разводов в ЗАГСе.
3. Перепись населения проводилась с 9 по 16 октября. Критическим моментом регистрации было 0 часов 9 октября. Переписчик пришел в семью Петровых 12 октября. В этой семье 11 октября родился ребенок. Переписчик должен:
 - а) Не вносить сведения о родившемся ребенке в переписной лист;

- б) Внести сведения с пометкой о дате рождения;
- в) Внести сведения без пометки о дате рождения.
- 4. Критический момент наблюдения - это:
 - а) время, в течение которого происходит заполнение статистических формуляров;
 - б) конкретный день года, час дня, по состоянию на который должна быть проведена регистрация признаков по каждой единице исследуемой совокупности;
 - в) время, в течение которого происходит обработка информации.
- 5. Последовательность этапов статистического исследования:
 - а) сбор первичной статистической информации;
 - б) анализ статистической информации;
 - в) сводка и группировка первичной информации;
 - г) рекомендации на основе анализа данных.

2. Краткое описание проводимого занятия

1. Познакомиться с программой статистических работ Росстата, организацией статистических наблюдений в системе государственной статистики России.
2. Изучить бланки федеральных статистических наблюдений.
3. С помощью устного опроса и (или) тестирования оценить уровень усвоения студентами изученного материала.

3. Результаты и выводы

Усвоение студентами знаний по теме практического занятия.

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 2: АНАЛИЗ РЯДОВ ДИНАМИКИ

1. Задания для работы

Вопросы к занятию

1. Понятие о динамике и динамических рядах.
2. Показатели, характеризующие динамику.

Типовые тесты (для контроля усвоения знаний)

1. Ряд динамики характеризует:
 - а) структуру совокупности по какому-либо признаку;
 - б) изменение значений признака во времени;
 - в) определенное значение варьирующего признака в совокупности;
 - г) факторы изменения показателя на определенную дату или за определенный период.
2. Моментным рядом динамики является:
 - а) остаток оборотных средств предприятия по состоянию на 1 число каждого месяца;
 - б) производительность труда на предприятии за каждый месяц года;
 - в) сумма банковских вкладов населения на конец каждого года;
 - г) средняя заработная плата рабочих и служащих по месяцам года.
3. По формуле $T_p = \frac{y_n}{y_0}$ определяется:
 - а) базисный темп роста; б) цепной темп роста;
 - в) базисный темп прироста; г) цепной темп прироста;
 - д) абсолютное значение 1% прироста.

$$T_p = \frac{y_t}{y_{t-1}}$$

4. По формуле $T_p = \frac{y_t}{y_{t-1}}$ определяется:
 - а) базисный темп роста; б) цепной темп роста; в) базисный темп прироста;
 - г) цепной темп прироста; д) абсолютное значение 1 % прироста.
5. Периодические колебания, возникающие под влиянием смены времени года называются...:
 - а) хронологическими; б) сезонными; в) тенденцией; г) случайными.

Типовые задачи

1. Данные на начало месяцев (млн. руб.): на I/IV-2015 г. – 500; на I/V-2015г. - 520; на I/VI – 2015г. -510; на I/VII - 2015 г. – 470. Средний остаток оборотных средств за 2 квартал составил ... млн. руб.

2. Используя данные таблицы, рассчитайте абсолютные, относительные и средние показатели динамики. Результаты расчетов представьте в табличной форме. Сделайте выводы.

Таблица 1 – Динамика производства молока, тыс. тонн

	2000 г.	2005 г.	2006 г.	2007 г.	2008 г.	2009 г.	2010 г.	2011 г.	2012 г.
Произведено молока, тыс. т	173,4	399,4	537,7	572,6	571,8	649,3	637	651	498,4

3. Рассчитайте индексы сезонности для января-марта :

Месяц	Выручка, млн. руб.	
	2004	2005 г.
Январь	17,3	16,0
Февраль	15,2	15,8
Март	17,2	18,4
...
Итого за год	213,6	220,4

При решении задач необходимо акцентировать внимание на следующем:

1. Правильное применение формул для расчета показателей динамики.
2. Отработка навыков графического представления динамического ряда.
3. Обязательная формулировка экономически обоснованных выводов по результатам расчетов.

2. Краткое описание проводимого занятия

1. Устный опрос и (или) тестирование по теме занятия.
2. Решение задач по теме занятия.

3. Результаты и выводы

Усвоение обучающимися знаний и закрепление навыков по теме практического занятия.

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 3: ПАРНАЯ КОРРЕЛЯЦИЯ

1. Задания для работы

Вопросы к занятию:

1. Понятие, роль и задачи корреляционно-регрессионного метода анализа.
2. Расчет и интерпретация коэффициентов парной линейной регрессии, корреляции и детерминации.
3. Проверка статистической значимости коэффициентов регрессии и корреляции.

Типовые тесты (для контроля усвоения знаний)

1. Парный коэффициент корреляции показывает тесноту
 - а) линейной зависимости между двумя признаками;
 - б) линейной зависимости между тремя и более признаками;
 - в) нелинейной зависимости между двумя признаками;
 - г) связи между результативным признаком и остальными, включенными в модель.
2. Парный коэффициент корреляции может принимать значения....
 - а) от 0 до 1; б) от -1 до 0; в) от -1 до 1; г) любые положительные;
 - д) любые значения.
3. Установите последовательность этапов проведения корреляционно-регрессионного анализа:
 - а) Экономическая интерпретация результатов.
 - б) Оценка тесноты связи между признаками с помощью коэффициентов связи.
 - в) Установление причинно-следственной связи (экономическое обоснование выбора факторов).
 - г) Установление (аналитической) математической формы связи (выбор вида уравнения

регрессии), решение уравнения связи и определение его параметров.

4. Оценка значимости параметров уравнения регрессии осуществляется на основе:

а) t - критерия Стьюдента; б) F - критерия Фишера; в) средней квадратической ошибки; г) средней ошибки аппроксимации.

5. Параметрическими показателями связи являются:

а) Эмпирическое корреляционное отношение;

б) Коэффициент линейной корреляции К. Пирсона;

в) Коэффициент корреляции знаков Фехнера;

г) Коэффициент корреляции рангов Спирмена.

Типовые задачи

1. Используя данные таблицы 1, рассчитайте показатели связи между признаками: параметрический коэффициент корреляции К. Пирсона и непараметрический коэффициент корреляции рангов Спирмена. Сделайте сравнительные выводы о тесноте и направлении связи признаков.

Таблица 1 – Данные по сельскохозяйственным организациям.

Надой молока, ц	27,2	20,5	24,9	28,1	33,6	25,8	24,5	28,3	47,7	40,1	37,8	43	35,6	53,1	40,9
Расход кормов на корову, ц к.ед.	30,5	25	30,6	34,3	34,6	31,5	27,2	31,7	43,9	45,7	46,1	44,3	44,8	60	49,5

2. Используя данные таблицы 1, выполните оценку статистической значимости показателя связи.

При решении задач необходимо акцентировать внимание на следующем:

1. Правильное применение формул для расчета показателей связи.
2. Экономическое обоснование причинно-следственных связей при изучении социально-экономических процессов.
3. Обязательная формулировка экономически обоснованных выводов по результатам расчетов.

2. Краткое описание проводимого занятия

1. Устный опрос и (или) тестирование по теме занятия.
2. Решение задач по теме занятия.

3. Результаты и выводы

Усвоение студентами знаний и закрепление навыков по теме практического занятия.

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 4: МНОГОФАКТОРНЫЙ КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ

1. Задания для работы

Вопросы к занятию:

1. Понятие множественной корреляции. Этапы МКРА.
2. Расчет и интерпретация парных, частных и множественных коэффициентов корреляции.
3. Проверка статистической значимости коэффициентов корреляции.
4. Практическое использование уравнение регрессии для оценки эффективности деятельности предприятия и для прогнозирования.

Типовые тесты (для контроля усвоения знаний)

1. В результате проведения регрессионного анализа получают функцию, описывающую ... показателей:

а) взаимосвязь; б) соотношение; в) структуру; г) темпы роста; д) темпы прироста.

2. Оценка значимости уравнения регрессии в целом осуществляется на основе:

а) t - критерия Стьюдента; б) F - критерия Фишера;

в) средней квадратической ошибки; г) средней ошибки аппроксимации.

3. По аналитической форме связи бывают...:
- парные и множественные; б) линейные и нелинейные;
 - прямые и обратные; г) слабые и тесные.
4. По числу взаимосвязанных элементов связи бывают...:
- парные и множественные; б) линейные и нелинейные;
 - прямые и обратные; г) слабые и тесные.
5. В линейном уравнении $\bar{Y}_x = a_0 + a_1x$ коэффициент регрессии показывает:
- тесноту связи между признаками;
 - долю дисперсии "Y", зависимую от "X";
 - на сколько в среднем изменится "Y" при изменении "X" на одну единицу;
 - ошибку коэффициента корреляции.

Типовые задачи

1. Используя данные таблицы 1, выполните корреляционный анализ, значения коэффициентов корреляции представьте в виде матрицы. Сделайте вывод о силе и тесноте связи между признаками, обоснуйте выбор факторных и результативного признаков. Выполните оценку статистической значимости полученных коэффициентов корреляции.

Таблица 1 – Данные по сельскохозяйственным организациям

Надой молока, ц	Доля концентрированных кормов	Ср. живая масса, кг	Длительность сухостойного периода, дн.	Расход кормов на корову, ц к.ед.
27,2	32,5	447	91	30,5
20,5	35,3	454	82	25
24,9	33,6	473	72	30,6
28,1	26,1	458	66	34,3
33,6	42,6	462	65	34,6
25,8	32,5	460	82	31,5
24,5	29,4	463	83	27,2
28,3	33,8	463	80	31,7
47,7	36,7	581	62	43,9
40,1	33,3	559	61	45,7
37,8	31,6	552	59	46,1
43	24,3	550	71	44,3
35,6	23,9	547	66	44,8
53,1	29,3	618	66	60
40,9	29,9	565	66	49,5

2. По данным таблицы 1 постройте уравнение регрессии продуктивности животных от выбранных вами факторных признаков. Выполните оценку статистической значимости параметров полученного уравнения и уравнения регрессии в целом. Интерпретируйте полученные результаты с экономической точки зрения.

При решении задач необходимо акцентировать внимание на следующем:

- Правильное применение формул для расчета показателей связей, уравнения регрессии.
- Экономическое обоснование причинно-следственных связей при изучении социально-экономических процессов.
- Обязательная формулировка экономически обоснованных выводов по результатам расчетов.

2. Краткое описание проводимого занятия

- Устный опрос и (или) тестирование по теме занятия.
- Решение задач по теме занятия.

3. Результаты и выводы

Усвоение студентами знаний и закрепление навыков по теме практического занятия.

