



# ТЕОРИЯ СТАТИСТИКИ И ЭФФЕКТИВНОСТЬ РЕКЛАМНОГО ДЕЛА

**В.А. Чвьякин**

Учебник



[www.scipro.ru](http://www.scipro.ru)

# ТЕОРИЯ СТАТИСТИКИ И ЭФФЕКТИВНОСТЬ РЕКЛАМНОГО ДЕЛА

В.А. Чвякин

Учебник

Нижний Новгород  
2024

УДК 005.8  
ББК 65.05  
DOI 10.54092/9785907607569

**Главный редактор:** Краснова Наталья Александровна – кандидат экономических наук, доцент, руководитель НОО «Профессиональная наука»

**Технический редактор:** Канаева Ю.О.

**Рецензенты:**

*Мальшаков Григорий Викторович*, кандидат технических наук, доцент кафедры информационных систем и цифровых технологий, Московский государственный университет технологий и управления им. К.Г. Разумовского (ПКУ)

*Ксенофонтова Халидя Зейнатулловна*, кандидат социологических наук, доцент Центра проектной деятельности, доцент кафедры рекламы и связей с общественностью в медиаиндустрии, Московский политехнический университет

**Автор:**

*Чвякин Владимир Алексеевич*, доктор философских наук, профессор кафедры рекламы и связей с общественностью в медиаиндустрии, Московский политехнический университет

Теория статистики и эффективность рекламного дела [Электронный ресурс]: учебник – Эл. изд. - Электрон. текстовые дан. (1 файл pdf: 110 с.). - В.А. Чвякин. 2024. – Режим доступа: [http://scipro.ru/conf/advertising1\\_24.pdf](http://scipro.ru/conf/advertising1_24.pdf). Сист. требования: Adobe Reader; экран 10'.

ISBN 978-5-907607-56-9

Учебник «Теория статистики и эффективность рекламного дела» дает объем знаний и представлений о возможностях использования теории статистики для оценки эффективности рекламной деятельности применительно к направлениям подготовки «Реклама и связи с общественностью»: бакалавриат 42.03.01 и магистратура 42.04.01.

Учебник обсужден и рекомендован к печати кафедрой рекламы и связей с общественностью в медиаиндустрии Московского политехнического университета, протокол № 4 от 8 декабря 2023 г.

ISBN 978-5-907607-56-9



© В.А. Чвякин. 2024  
© Оформление: издательство НОО Профессиональная наука, 2024

# Содержание

|   |           |
|---|-----------|
| Введение .....  | 5         |
| <b>Раздел 1. Методы оценки эффективности рекламной кампании .....</b>   | <b>6</b>  |
| Тема. 1. Общие вопросы эффективности рекламы и методов ее экономической оценки .....  | 6         |
| Тема 2. Методы и способы оценки эффективности рекламы .....   | 12        |
| <b>Раздел 2. Методы обобщения статистической информации .....</b>   | <b>23</b> |
| Тема 3. Общетеоретическое понимание основных категорий и понятий статистики. Формы представления статистических данных .....  | 23        |
| Тема 4. Построение ряда распределения. Расчет структурных характеристик ряда распределения .....  | 29        |
| Тема 5. Методы выявления основной тенденции (тренда) в рядах динамики. Оценка адекватности тренда и прогнозирование .....   | 34        |
| Тема 6. Расчет показателей размера и интенсивности вариации. Расчет моментов распределения и показателей его формы .....  | 39        |
| Тема 7. Проверка соответствия ряда распределения нормальному. Проверка соответствия ряда распределения закону Пуассона .....  | 44        |
| Тема 8. Абсолютные и относительные показатели изменения структуры. Ранговые показатели изменения структуры .....  | 52        |
| Тема 9. Понятие выборочного наблюдения. Способы формирования выборки. Средняя ошибка выборки. Предельная ошибка выборки. Необходимая численность выборки .....  | 58        |
| Тема 10. Понятие корреляционной зависимости. Методы выявления и оценки корреляционной связи. Коэффициенты корреляции рангов. Особенности коррелирования рядов динамики. Показатели тесноты связи между качественными признаками ..... | 61        |
| <b>Раздел 3. Статистические инструменты макроэкономического анализа .....</b>   | <b>82</b> |
| Тема 11. Система национальных счетов (СНС) – инструмент макроэкономического анализа .....   | 82        |
| Тема 12. Методы исчисления ВВП и взаимосвязь основных макроэкономических агрегатов СНС .....  | 84        |
| Заключение .....  | 88        |
| Глоссарий по теории статистики .....  | 89        |
| Тестовые задания по теории статистики (учебно-ознакомительный тест с ответами) .....  | 96        |
| Литература .....  | 107       |

## Введение

Проведение рекламных мероприятий всегда сопряжено с необходимостью оценки их эффективности. Экономическая и коммуникативная эффективность рекламы требует обязательного статистического учета многочисленных критериев, которые могут убедительно свидетельствовать о разумности расходования средств на рекламу. Поэтому важно понимать потенциал статистики как оценочного инструментария рекламы. Специалисты в сфере рекламной деятельности используют различные статистические программы, которые позволяют получить многочисленные количественные показатели и это очень важно. Но среди этих показателей имеются такие, оценочный смысл которых может быть не совсем понятен или, когда этот смысл становится очевидным только в совокупности с другими показателями.

Теория статистики позволяет понять закономерности динамики количественных показателей в структуре оценочного процесса применительно, например, к рекламной деятельности. Связано это с тем, что теория статистики способствует познанию методологических основ и практическому овладению приемами экономико-статистического анализа, ознакомлению с системой статистических показателей, отражающих состояние и развитие явлений и процессов общественной жизни. Реклама в этом смысле может рассматриваться в качестве показательного примера для изучения не только статистических, но и других экономических дисциплин, использующих статистические методы анализа (финансы, кредит, финансовый и банковский менеджмент, ценные бумаги и других).

Освоение методов получения и обработки статистической информации, наиболее универсальных и распространенных в практике методов статистического анализа, методологии построения и анализа системы статистических показателей, отражающих состояние и развитие явлений и процессов общественной, жизни имеет практическое значение. Содержание данного учебника направлено на освоение таких тем, которые часто представляют сложности для обучающихся по направлениям подготовки «Реклама и связи с общественностью»: бакалавриат (42.03.01) и магистратура (42.04.01). В результате освоения этих тем выпускник должен:

- знать область применения статистического учета; организацию статистического наблюдения и обработку его результатов; методы статистического анализа; основные правила расчета обобщающих статистических показателей; важнейшие показатели, характеризующие различные стороны социально-экономического развития;

- уметь формулировать выводы, необходимые для проведения научных исследований и осуществления практической деятельности; анализировать полученные результаты и представлять информацию для дальнейших исследований прикладных экономических наук; использовать статистическую информацию при дальнейшем освоении общих и специальных экономических дисциплин;

- владеть навыками оценки степени достоверности имеющейся статистической информации, грамотно применять ее для решения конкретных практических и научных задач; навыками работы с базами статистических данных; использования разнообразных источников статистической информации (отечественные, зарубежные, международные); навыками свободного ориентирования в статистическом информационном поле.

Материал учебника «Теория статистики и эффективность рекламного дела» дает объем знаний и представлений о возможностях использования теории статистики для оценки эффективности рекламной деятельности применительно к направлениям подготовки «Реклама и связи с общественностью».

## Раздел 1. Методы оценки эффективности рекламной кампании

### *Тема. 1. Общие вопросы эффективности рекламы и методов ее экономической оценки*

Использование статистических методов наблюдения, анализа, моделирования и прогнозирования различных аспектов рекламной деятельности и воздействия на потребителей позволяет снизить степень риска принимаемых решений, а также составить довольно обоснованное представление о том, какие формы и виды рекламы оказывают наиболее сильное воздействие на потребителя, грамотно оценивать конечные коммерческие и социально-психологические результаты рекламной деятельности. В силу данного обстоятельства рекламная деятельность должна опираться на необходимые статистические показатели, расчеты и прогнозы, позволяющие выявить определенные закономерности. Поэтому рекламную деятельность можно рассматривать как особый объект статистического исследования.

Для оценки эффективности рекламы используются самые разные методы исследования, применяются всевозможные приёмы и инструменты, а полученные результаты интерпретируются как субъективным образом, так и с применением методов анализа, заимствованных из самых разных дисциплин – экономики, социологии, психологии. Тем не менее, результат в настоящий момент можно рассматривать скорее, как экстенсивный: вместо однозначного ответа или хотя бы четко определенного семейства решений на сегодняшний день мы имеем только чрезвычайно богатый набор всевозможных подходов, концепций, точек зрения и т.д. В лучшем случае мы можем говорить о более или менее корректных локальных решениях данной задачи, но речь даже близко не идет о каком-либо общем ответе. Поэтому, как и всякая проблема, которая жизненно волнует очень многих, но долго не имеет четкого решения, вопрос эффективности рекламы начал обрастать различными эзотерическими теориями, мифами и даже анекдотами.

Вопрос об оценке эффективности рекламы до сих пор остаётся открытым. Неразрешённость этой темы обуславливает её актуальность, так как по логике вещей основное требование, предъявляемое к рекламе, - это её эффективность.

Вопрос эффективности рекламы - один из самых сложных и самых важных вопросов, стоящих перед рекламистами. Это является предметом постоянной озабоченности создателей рекламы. Как сделать рекламу максимально эффективной? Однозначного ответа на этот вопрос не дает ни один специалист, поскольку выделить чистый эффект, полученный от рекламы, крайне сложно. Ведь на этот же результат — продажа товара/услуги — работают и остальные составляющие маркетингового комплекса: товар, цена, дистрибуция и вся система маркетинговых коммуникаций, частью которой является и реклама.

Проблема с определением эффекта, получаемого только от рекламной деятельности, является одной из сложнейших в рекламной практике по ряду причин. Это потому, что:

1. Реклама является одним из многих аргументов, которые определяют конечные маркетинговые результаты, но при этом немаловажную роль играют другие элементы маркетингового комплекса.

2. На конечный результат влияют множество факторов, которые практически невозможно формализовать или удачно смоделировать, в частности, поведение конкурентов.

3. Сложность оценки рекламного воздействия обусловлена поведением покупателя, которого можно сравнить с «черным ящиком» с его непредсказуемыми процессами.

4. Рыночные факторы изобилуют случайными событиями, которые могут определить успех или неуспех товара.

И все же можно говорить об эффективности отдельных показателей рекламной кампании. Эффективность рекламной кампании – максимальное достижение тех целей и задач, которые были поставлены при планировании рекламной кампании.

Цели рекламного обращения, а тем более всей рекламной кампании должны быть установлены четко, обязательно выражены количественно и по срокам. Например, «Добиться к началу зимы того, чтобы 60% потенциальных потребителей услышало о нашем товаре, 30% услышавших попробовало его, а 90% попробовавших вынесли положительный вердикт и сбыт увеличился до 500 упаковок в неделю».

Принято считать, что в отношении рекламной деятельности есть две составляющие ее эффективности - *экономическая* и *коммуникативная*. В рамках каждой из составляющих анализируются статьи рекламного бюджета, а также эффективность задействованных средств рекламы, проводятся тестирования созданных ими рекламных обращений на различных стадиях их готовности. Эти составляющие эффективности рекламы образуют практически единый процесс. Они следуют одна за другой.

Прежде всего всех интересует коммуникативная эффективность рекламы, которая определяется эффективностью контакта с потребителем. Этот показатель планируется рекламистами в течение всего процесса создания и производства рекламного продукта. И уже эта подсчитанная теоретически и заложенная в исходные данные всей рекламной кампании эффективность является основанием для получения в дальнейшем экономической эффективности, то есть эффективности получения прибыли. Этот вид эффективности возникает во время контакта потребителя с товаром в местах продаж.

*Коммуникативная эффективность рекламы* представляет собой оценку рекламной кампании в ее взаимоотношениях с потребителями рекламы. В зависимости от применяемых средств распространения рекламы это могут быть зрители, читатели, пешеходы и водители, покупатели в магазинах и т.д.

Коммуникативная эффективность рекламы даёт возможность оценить:

- долю потребителей, встречавших рекламное сообщение;
- эффективность различных медиаисточников;
- какие из элементов рекламы запомнились потребителю наиболее всего;
- какое рекламное сообщение было понято и интерпретировано потребителем;
- силу связи между рекламным сообщением и рекламируемой маркой;
- уровень влияния рекламы на продажи/потребление продуктов.

Результаты рекламной кампании в первую очередь зависят от правильно выбранных подходов к созданию рекламного обращения: его идеи, воплощения. Чтобы не ошибиться в этих вопросах, рекламисты проводят большую работу по определению оптимальных параметров рекламного обращения.

При всех прочих равных условиях эффективность рекламы на этом этапе может быть обеспечена, если она удовлетворяет нижеперечисленным условиям, то есть рекламное сообщение должно:

- соответствовать имиджу, который рекламодаделец желает придать своему товару;
- ясно выражать заложенные в нем идеи, обращенные к разуму и чувствам потребителя;
- быть доступным целевой аудитории по своему языку, социальному коду и не идти вразрез с ее мнением;
- внушать доверие: не лгать, не преувеличивать, позволять потребителю по возможности проверять сказанное;
- быть оригинальным;
- доносить до потребителя оптимальный объем информации;
- обладать внутренней цельностью и не противоречить другим акциям рекламной кампании;
- быть актуальным в течение долгого времени, чтобы рекламодаделец имел возможность использовать удачную рекламную идею в течение длительного периода.

Изучение коммуникативной (информационной) результативности рекламы дает возможность улучшить качество как содержания, так места и формы подачи информации.

*Экономическая эффективность рекламы* – это экономический результат, полученный от применения рекламного средства или организации рекламной кампании. Экономическую эффективность рекламы можно выразить через эффективность затрат на продвижение.

*Эффективность затрат* – это отношение количественного выражения эффекта к затратам, вызвавшим этот эффект. Основная часть проблемы при оценке эффективности затрат на рекламу заключается в определении числителя дроби.

Иногда этот вид эффективности называют еще *сбытовой, торговой или коммерческой*. Для рекламодателя самым основным вопросом всегда является вопросувеличения прибыли, получаемой в результате проведения рекламных кампаний. Это основной определяющий момент в оценке проведенной кампании. То есть делается попытка проанализировать и сопоставить рекламные расходы и изменение объемов продаж.

При рассмотрении данного вопроса определенную пользу приносит такие вспомогательные показатели реальной торговли, как:

- отношение прироста объема продаж товара к сумме затрат на его рекламу;
- отношение прироста прибыли, полученной после рекламной кампании, к сумме рекламных затрат;
- динамика уровня рекламных затрат в общем объеме продаж;
- расходы на рекламу на тысячу потребителей, подвергшихся воздействию всехвидов рекламы;
- расходы на рекламу на тысячу потребителей, подвергшихся воздействиюконкретного вида рекламы;
- количество покупок товара, спровоцированных рекламой;
- прирост объема сбыта за период, прошедший после рекламной кампании (следует всегда иметь в виду, что зачастую прибыль проявляется не сразу после проведения рекламной кампании).

Основным материалом для анализа экономической эффективности результатов рекламных мероприятий фирмы служат статистические и бухгалтерские данные о росте товарооборота. На основе этих данных можно исследовать экономическую эффективность одного рекламного средства, рекламной кампании и всей рекламной деятельности в целом.

Как уже было сказано ранее, измерение экономической эффективности рекламы представляет большие трудности, так как, реклама, как правило, не даёт рекламного эффекта сразу. Кроме того, рост товарооборота нередко вызывается другими (нерекламными) факторами, например, изменением покупательской способности населения из-за роста цен и т.п. Поэтому, для того, чтобы получить результаты, которые будут близки к реальным, следует рассматривать изменения экономических показателей деятельности фирмы под действием лишь рекламной кампании, не проводя в это же времядругих мероприятий по продвижению.

### **Методы оценки экономической эффективности рекламы**

Для расчёта экономической эффективности рекламы могут быть использованы следующие методы и формулы.

#### **1. Метод сравнения товарооборота до и после проведения рекламного мероприятия.**

По этому методу экономическая эффективность рекламы определяется либо путем сопоставления товарооборота за определенный отрезок текущего года, когда товар подвергался воздействию рекламы, с данными за аналогичный период прошлого года, когда товар не рекламировался, либо путем сопоставления ежедневного товарооборота до и после проведения рекламного мероприятия в текущем периоде времени.

Последний способ более приемлем в наших условиях, учитывая постоянный рост цен из-за инфляции, что делает сопоставление данных за большие промежутки времени весьма затруднительным.



Окончательные выводы об экономической эффективности рекламы получаются в результате сравнения дополнительной прибыли, полученной в результате рекламы, с расходами, связанными с ее осуществлением.

*Расчёт дополнительного товарооборота под воздействием рекламы:*

$$T = \frac{T_c * P * D}{100}, \text{ где}$$

д

$T_d$  – дополнительный товарооборот, вызванный рекламными мероприятиями, руб.;

$T_c$  – среднечасовой товарооборот до начала рекламного периода, руб.;

$P$  – относительный прирост среднечасового товарооборота за рекламный период по сравнению с дорекламным, %;

$D$  – количество дней учёта товарооборота в рекламном процессе.

*Расчёт экономического эффекта рекламирования, то есть разницы между прибылью, полученной от дополнительного товарооборота, вызванного рекламными мероприятиями, и расходами на рекламу.*

Для расчёта экономического эффекта можно использовать формулу:

$$\mathcal{E} = \frac{T_d * N_T}{\text{где}}$$

где

$\mathcal{E}$  – экономический эффект рекламирования, руб.;

$T_d$  – дополнительный товарооборот под воздействием рекламы, руб.;

$N_T$  – торговая надбавка на единицу товара, в % к цене реализации;

$Z_p$  – затраты на рекламу, руб.;

$R_d$  – дополнительные расходы по приросту товарооборота, руб.

*Расчёт эффекта от рекламы при проведении рекламной распродажи товара производится по формуле:*

$$\mathcal{E}_p = P_p - Z_p,$$

где

$P_p$  – прирост от реализации товаров за период рекламной кампании, руб.;

Экономический эффект рекламных мероприятий может быть положительным – затраты на рекламу меньше дополнительной прибыли; отрицательным – затраты на рекламу выше дополнительной прибыли; нейтральными – затраты на рекламу равны дополнительной прибыли.

*Оценка рентабельности рекламы:*

$$P = \left( \frac{П}{З} \right) * 100\%, \text{ где } P \text{ – рентабельность рекламирования, \%}$$

$П$  – дополнительная прибыль, полученная от рекламирования товара, руб.;  $З$  – общие затраты, руб. ( $З = Z_p + R_d$ ).

*Метод целевых альтернатив.*

Экономическая эффективность рекламы может определяться методом целевых альтернатив путём сопоставления планируемых и фактических показателей, оцениваемых как результат вложения средств в рекламную кампанию:

$$K = \left( \frac{П_f}{П_o} \right) * 100\%,$$

где

$K$  – уровень достижения планируемого уровня прибыли, %;

$П_f$  – фактический объём прибыли за период действия рекламы, руб.;

$П_o$  – планируемый объём прибыли за период действия рекламы, руб.

### *Оценка эффективности рекламных издержек как формы инвестирования методом ROI.*

Прямой анализ рекламы как формы издержек существует в двух видах. В зависимости от конкретной задачи мы можем оценивать:

1. Удельный вес рекламных расходов в выручке (обороте)
2. Отношение рекламных издержек к чистой прибыли

Суть рекламы как формы инвестирования в самом простейшем случае можно рассматривать как вложение средств в продвижение товара или услуги, с целью получения дохода или иных выгод. С этой точки зрения можно проанализировать эффективность рекламных инвестиций с помощью финансового показателя ROI (Return of Investment) – или возврат инвестиций. Как это сделать?

1. Обозначим начальный этап  $t_0$ , характеризующийся совокупными продажами некоторого товара или услуги  $X_0$ , измеряемых в натуральном выражении. В этой же точке имеем  $p$  – цену единицы этого товара или услуги. Тогда текущий продаваемый объём товара или услуги в стоимостном выражении есть  $Y_0 = p * X_0$ .

2. Через  $q$  обозначим удельную прибыль – сумму, получаемую от продажи одной единицы товара или услуги. Зная  $q$ , можем найти совокупную прибыль  $Z$ , полученную от реализации  $X_0$  товаров или услуг  $Z = q * X_0$ .

3. В момент  $t_0$  запустим рекламную кампанию стоимостью  $I$ . Полагаем, что через определённое время (в момент  $t_1$ ), получим изменившийся объём продаж, измеряемый в натуральном  $X_1$  или стоимостном  $Y_1$  выражении.

Если всё прошло успешно, и полученный  $Z_1 = q * X_1$  оказался больше стартового  $Z_0$ , то имеет место положительная разница эти значений  $\Delta Z$ , которую можно считать главным критерием эффективности рекламы. Чем выше  $\Delta Z$ , тем успешнее прошла рекламная кампания. И чем меньше при этом понесённые издержки  $I$ , тем эффективней это сделали.

Соответственно формула определения эффективности будет выглядеть как отношение прироста прибыли к инвестициям в продвижение, то есть **ROI**. Искомая эффективность, выраженная через коэффициент **ROI - EROI**:

$$EROI = \frac{\Delta Z}{I},$$

где

$$\Delta Z = Z_1 - Z_0.$$

1. **EROI < 0** показывает, что продажи после проведения рекламной кампании упали ниже первоначального уровня. Это говорит о безусловной неэффективности рекламы.

2. **0 < EROI < 1** - прирост продаж никак не оправдал понесённых затрат на рекламу, если она не преследовала каких-либо других целей (например, не экономических, а коммуникативных). Поэтому с финансовой точки зрения кампания считается неэффективной.

3. Получается, что эффективное значение **EROI** лежит в области  $[1; \infty)$ .

Итак, анализ показывает, что эффективное значение **EROI** должно быть в пределах  $[1; \infty)$ . Однако, даже получив число, заведомо большее единицы, требуется определить насколько это много или мало; можно ли как-то повысить эффективность или лучше не пытаться и т.д. Другими словами, необходим некий масштаб, чтобы оценить то, что мы получили. В таких случаях обычно сравнивают отдачу от рекламных инвестиций с ROI по другим направлениям деятельности фирмы: технологическое перевооружение, расширение бизнеса, вложение в более квалифицированный персонал и т.д. Но и такая оценка не даёт исчерпывающего решения, поэтому логичнее сравнивать **EROI** с отдачей от инвестирования аналогичных сумм в наиболее надёжные финансовые инструменты, существующие в настоящее время на рынке (государственные ценные бумаги, депозиты в наиболее надёжных банках и проч.). Это даст адекватную, хотя и довольно условную, оценку того, насколько удачно или неудачно прошла данная рекламная кампания.

Попробуем рассмотреть, как это выглядит на практике. Например, крупный московский автомобильный дилер продает автомобили популярной французской марки, а также недорогую модель 5. Большую часть осени (среднего и довольно ровного сезона для реализации автомобилей) продажи составляли около 20 автомобилей в месяц: колебались от 3-4 до 6-7 в неделю (рис. 1). С 15 ноября по 16 декабря была проведена специальная рекламная акция по стимулированию сбыта  $S$  общей стоимостью 30 тыс. долл. ( $I = 30$ ). В результате продажи выросли и достигли в момент пика 16 автомобилей в неделю. Потом продажи вернулись в прежний коридор (от 3-4 до 6-7 машин в неделю). Всего в момент проведения специальной рекламной кампании была продана 51 машина ( $x_1 = 51$ ). При этом, если проследить динамику осенних продаж, прогноз на период с середины ноября до середины декабря ( $x_0$ ) составлял 17 единиц. Разница ( $\Delta X$ ) составляет 34 сверхплановых автомобиля, которые мы (зная, что в этот период не изменялись какие-либо другие параметры, влияющие на продажи: цены, комплектация и др.) реализовали дополнительно к прогнозному числу благодаря локальной рекламной кампании.

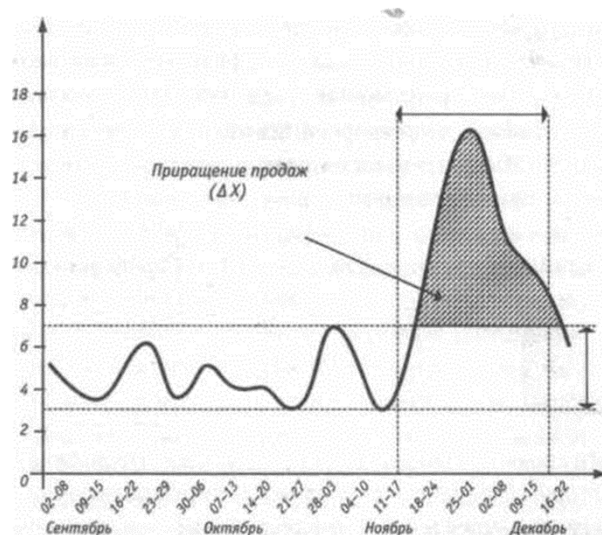


Рис. 1. Продажи и рекламная кампания

Зная удельную прибыль ( $q$ ) от каждого проданного  $S$  - 1,5 тыс. долл., — мы вправе подсчитать совокупную прибыль ( $Z$ ) и подставить эти значения в формулу. В итоге легко рассчитываем  $E_{ROI}$  для конкретного случая.

#### Вопросы для самопроверки:

1. В чем заключаются причины сложностей определения эффективности рекламы?
2. В чем заключается смысл определения экономической эффективности рекламы?
3. В чем заключается смысл определения коммуникативной эффективности рекламы?
4. Коммуникативная эффективность рекламы даёт возможность оценить...
5. Что определяет понятие «эффективность затрат» в рекламном деле?
6. Сбытовая, торговая и коммерческая эффективность — это одно и то же или это разные понятия?
7. Эффективность затрат в рекламном деле обозначает...
8. Можно ли выразить экономическую эффективность рекламы через эффективность затрат на продвижение товара?

## ***Тема 2. Методы и способы оценки эффективности рекламы***

**Методы оценки коммуникативной эффективности рекламы.** Факторы коммуникативной эффективности рекламы могут быть выявлены с помощью специальных опросов и тестирования. *Тестирование* коммуникативной эффективности рекламы обычно проводится на двух этапах: до ее выхода на рынок (предтестирование) и после выхода (посттестирование).

*Предварительное тестирование.* Предтестирование должно обеспечить защиту от ошибок в разработке рекламной коммуникации. Проверка производится по таким параметрам, как:

- форма и содержание сообщения;
- правильность выбора рынка и целевой аудитории;
- оценка средств и каналов передачи рекламных сообщений.

Выбор оптимальных каналов размещения рекламы (рекламных контактов), проводимый с целью достижения максимальной эффективности рекламной кампании происходит в процессе *медиапланирования*.

*Медиаплан* - это план размещения рекламных обращений. В этом документе содержатся ответы на следующие вопросы: где размещать, какую аудиторию охватить, как часто размещать, когда размещать, сколько на это потратить. Медиаплан формируется в результате анализа значительного числа факторов, характеризующих то или иное медиасредство. К медийным каналам распространения рекламы относятся телевидение, пресса, наружные средства, радио, кино.

Определить, какие каналы рекламы следует привлечь, помогает ряд показателей, о которых полезно иметь представление и тем, кто заказывает размещение, и тем, кто его осуществляет. Основные показатели медиапланирования, применяемые для расчета эффективности рекламных кампаний следующие:

1. Рейтинг - это основная характеристика носителя рекламного сообщения, но не самого сообщения. Под рейтингом понимают количество индивидов, составляющих целевую аудиторию данного рекламного сообщения, смотрящих данный носитель в данное время, отнесенное к общей численности людей, имеющих техническую возможность смотреть телевизор, то есть потенциальных телезрителей. Таким образом, рассчитать рейтинг можно по простой формуле

$Rating = \frac{\text{Число телезрит., сост. целевую аудиторию, смотрящих данную передачу}}{\text{Общая численность потенциальных телезрителей}}$

Рейтинг можно считать, как в виде десятичной дроби, так и в процентах:

$Rating = \frac{\text{Число телезрит., сост. целевую аудиторию, смотрящих данную передачу}}{\text{Общая численность потенциальных телезрителей}} \times 100\%$

Рейтинг является характеристикой потенциальной целевой аудитории рекламного сообщения и служит одним из основных показателей при составлении медиаплана и при медиа-анализе результатов рекламных кампаний. Данные о рейтингах отдельных теле- и радиопередач подготавливаются специальными исследовательскими службами на основе регулярных выборочных опросов населения и анализа телепрограмм и публикуются. Существуют различные виды рейтингов, такие как рейтинг временного интервала, рейтинг телепередачи или рейтинг телеканала. Наиболее часто медиапланировщики применяют на практике величину *AQH* - средний рейтинг 1/4 часа.

2. Доля аудитории передачи - это аудитория конкретной рассматриваемой передачи, отнесенная к общей аудитории телезрителей, смотрящих телевизор в данное время, выраженная в процентах. Как следует из определения, *Share* характеризует процент телезрителей, предпочитающих смотреть именно данную передачу в рассматриваемый момент времени. Тем самым она отличается от рейтинга, который характеризует только потенциальную аудиторию носителя рекламы:

$Share = \text{Число телезрит., смотрящих данную передачу} / \text{Общая численность всех телезрителей, смотрящих телевизор в данный момент} \times 100\%$

3. HUT (аббревиатура от англ. *Homes Using Television*) - это доля телезрителей в данный момент времени, т.е. отношение общей численности всех телезрителей, смотрящих телевизор (любую передачу) в данный момент времени, к общей численности потенциальных телезрителей. HUT отличается от рейтинга тем, что здесь в расчет принимаются все телезрители, смотрящие телевизор в данный момент времени. Можно сказать, что это - характеристика популярности данного временного интервала у телезрителей:

$HUT = \text{Общая численность всех зрителей, смотрящих телевизор в данный момент} / \text{Общая численность потенциальных зрителей} \times 100\%$

Как следует из приведенных формул, связь между рассмотренными характеристиками имеет следующий вид:

$$Rating = HUT \times Share$$

4. GRP (от англ. *Gross Rating Point*) - это сумма рейтингов всех размещений рекламы, предусмотренных в медиаплане. Рейтинги могут складываться только при условии, что они определены на одной и той же базе (например, географический регион, демографические данные, уровень доходов и др.). Несмотря на то, что данная величина довольно абстрактная, она имеет важное значение для медиапланирования и при анализе результатов рекламных кампаний. GRP обычно выражается в процентах (хотя сам знак % часто опускается), но при некоторых расчетах берется в виде десятичной дроби. Подчеркнем, что суммирование рейтингов при определении GRP осуществляется независимо от того, что потенциальными телезрителями разных передач могут быть не одни и те же люди. Поэтому значение GRP само по себе не несет информации о частоте восприятия рекламы. Поскольку величина GRP определяется простым суммированием рейтингов, она может превышать 100%. Непосредственно по ее значению оценить суммарную аудиторию рекламной кампании невозможно, так как каждый из составляющих GRP рейтингов дает величину аудитории каждой отдельной трансляции рекламы. GRP - своеобразная характеристика мощности рекламной кампании. Иными словами, чем больше данная величина, тем большая аудитория имеет шансы увидеть/услышать рекламу хотя бы один раз в течение кампании и тем большее число людей из этой аудитории увидит ее более одного раза. GRP - это один из основных показателей при сопоставлении разных вариантов медиапланов между собой.

5. OTS (от англ. *Opportunity To See*) или количество контактов - это количество раз (в абсолютном исчислении), которое данное рекламное сообщение потенциально могли увидеть все потенциальные телезрители, независимо от того, относятся они к целевой аудитории или нет. Данная величина тесно связана с GRP и находится из соотношения:

$$OTS = GRP \times \text{Общая численность потенциальных телезрителей}$$

6. Reach или охват аудитории - это величина, характеризующая аудиторию, которая в ходе рекламной кампании видела/слышала рекламное сообщение. Эта величина может быть представлена в двух типичных формах: как характеристика аудитории, которая видела рекламу определенное количество раз, и как характеристика аудитории, которая видела ее не менее определенного количества раз. Обозначаются эти величины *Reach (n)* и *Reach (n+)* соответственно, где *n* - количество восприятий. Измеряются они в процентах (хотя сам знак % может опускаться). Формулы их расчета таковы:

$Reach (n) = \text{Общая численность зрителей, видевших рекламу n раз} / \text{Общая численность потенциальных зрителей} \times 100\%$

$Reach (n+) = \text{Общая численность зрителей, видевших рекламу не менее n раз} / \text{Общая численность потенциальных зрителей} \times 100\%$

Наибольшее распространение получила величина *Reach (n+)*, показывающая, какой процент потенциальной аудитории по завершении кампании видел рекламное сообщение *n* раз. Особо выделяется величина *Reach (1+)*, представляющая собой процент от общего числа

потенциальных телезрителей, каждый из которых видел/слышал рекламное сообщение в ходе рекламной кампании не менее одного раза.

7. *Frequency* или *Average Opportunity To See* (средняя частота восприятия) - это среднее количество раз, которое произвольно выбранный представитель из числа тех, кого реклама «достигла» хотя бы один раз, оказывается экспонированным ею. Средняя частота восприятия *Frequency* распределена в диапазоне от 1 до  $n$  числа выходов рекламного сообщения в ходе кампании и определяется по формуле:

$$Frequency = GRP/Reach (I+)$$

Из формулы видно, что чем больше средняя частота *Frequency*, тем меньше охват аудитории *Reach (I+)*. Задача оптимального медиапланирования как раз и состоит в том, чтобы, стремясь к наибольшему охвату, обеспечить необходимую среднюю частоту восприятия, ибо, как известно, реклама начинает срабатывать после нескольких (не менее 3-5) воздействий на аудиторию. Кроме всех вышеперечисленных показателей медиапланирования существует еще значительное число авторских методик и методов, имеющих весьма ограниченное применение, основанных на теории математической статистики и в некоторых случаях на теории вероятностей.

Предтестирование способствует разработке обоснованного бюджета рекламной кампании и, наконец, дает прогноз, какова будет ее эффективность.

*Текущее и заключительное тестирование (посттестирование)*. Подобные тесты используются для оценки качества усвоения информации. Обработка информации состоит из анализа реакций потребителя на рекламу: узнаваемость, запоминаемость, убедительность рекламы, влияние на покупательское поведение. Некоторые из перечисленных критериев могут использоваться и для предварительных тестов, а большинство из них применяются на двух этапах тестирования.

1. *Узнаваемость рекламы, оценка узнаваемости*. Узнаваемость рекламы заключается в определении, может ли респондент узнать рекламу, которую он видел некоторое время назад. Часто такие исследования заключаются в проведении оплачиваемого опроса по почте.

Испытание на узнаваемость теле- или радиорекламы может осуществляться следующим образом. Респондентам в течение 10 секунд демонстрируют «выжимку» рекламного сообщения, лишённую идентификаторов компании и марки товара. Респондентов спрашивают, видели (слышали) ли они ранее это рекламное сообщение, какая фирма проводит рекламную кампанию, товар какой марки рекламируется.

Один из самых ранних и все еще популярных тестов на узнаваемость — *тест Старча*, названный так по имени своего создателя Даниеля Старча. Этот тест может проверить только печатные объявления. После просмотра респондентом рекламного сообщения, например, в газете или журнале ему показывают страницу за страницей это издание, спрашивая, видел ли он или читал каждое из этих рекламных объявлений.

Каждому объявлению присвоен номер, и оно разбито на составные части (такие, как иллюстрации, заголовок, логотип или текст), имеющие коды. Если респондент говорит, что помнит или видел определенное объявление в этом издании, ему задают серию вопросов, чтобы точно определить, какую часть объявления респондент видел или читал. Тест Старча дает результат в виде следующих оценок:

- *заметили* (доля респондентов, которые говорят, что они заметили рекламу, когда просматривали журнал раньше);
- *ассоциировали* (доля респондентов, которые говорят, что заметили часть объявления, содержащую название рекламодателя или его логотип);
- прочитали большую часть (доля респондентов, которые говорят, что прочитали половину или больше текста рекламы).

Узнаваемость — необходимое условие для эффективности рекламы. Если реклама не может успешно пройти этот минимальный тест, она, как правило, неэффективна.

2. *Запоминаемость рекламы*. Запоминаемость рекламы позволяет определить, в какой мере

человек способен вспомнить ее содержание. Запоминаемость — один из основных критериев коммуникативной эффективности, по которому можно косвенно судить об экономической эффективности рекламы.

В основе использования этого критерия лежит гипотеза, сформулированная еще классиками рекламного дела Р. Ривсом и А. Политцем и разделяемая многими современными рекламистами: запоминаемость ведет к предпочтению. Эту особенность человеческого поведения А. Политц (*The Politz papers*, 1990) сформулировал в виде принципа «знакомая марка». Он установил, что расширение знаний о марке (или фирме) увеличивает доверие к ней и повышает ее ранг качества в восприятии потребителей. Даже простое знание названия марки (или фирмы) повышает доверие к ней по сравнению с совершенно незнакомой маркой (или фирмой).

Ривс ввел в теорию рекламы два понятия, связанные с эффектом запоминаемости, — *внедрение рекламы* и *вовлечение в потребление*. Внедрение представляет собой отношение числа лиц, запомнивших рекламу, к числу лиц, ее не запомнивших. Показатель вовлечения в потребление рассчитывается как разница между количеством покупателей на каждую сотню лиц, запомнивших рекламу, и числом покупателей на каждую сотню лиц, не знакомых с рекламой.

Таблица 1.

**Пример: эффективная и неэффективная реклама**

*Неэффективная реклама*

| <i>Результат исследований (измерение усвояемости рекламы)</i> | <i>Число респондентов, %</i> |
|---|------------------------------|
| Вспомнили после пересказа содержания                          | 48                           |
| Не вспомнили рекламу  | 41                           |
| Вспомнили по предъявлении марки                               | 7                            |
| Вспомнили спонтанно   | 4                            |

*Эффективная реклама*

| <i>Результат исследований (измерение усвояемости рекламы)</i> | <i>Число респондентов, %</i> |
|---|------------------------------|
| Вспомнили после пересказа содержания                          | 42                           |
| Не вспомнили рекламу  | 29                           |
| Вспомнили по предъявлении марки                               | 19                           |
| Вспомнили спонтанно   | 10                           |

Существует метод измерения запоминаемости рекламы (предложен М. Пикулевой), который базируется на трех составляющих: «спонтанное вспоминание» — реклама вспоминается при упоминании категории товара; «вспоминание при предъявлении товара определенной торговой марки» — реклама вспоминается после произнесения названия конкретного товара или его демонстрации; «вспоминания после пересказа рекламы». Сумма этих трех показателей составляет долю целевой аудитории, которая вспомнила рекламное сообщение. Однако важно учесть соотношение показателей между собой.

Кроме приведенных методик для выявления запоминаемости рекламы используются методика определения читаемости и воспроизведения журнальной рекламы Д. Старча, методика определения воспроизведения и узнаваемости журнальной рекламы Дж. Гэллапа и К. Робинсона, методика «day-after-recall», разработанная для анализа телевизионной рекламы фирмой Burke Marketing Research, Inc., и др.

*3. Убедительность рекламы.* Убедительность оценивает эффективность рекламы по тому, влияет ли она на намерение потребителей купить товар определенной марки. Основная форма теста на убедительность или на изменение отношения такова: потребителей сначала спрашивают, что, вероятно, они купят. Затем им показывают рекламу торговой марки. После этого их снова спрашивают, что они намерены купить. Результаты анализируются, чтобы определить, возросло ли их намерение купить в результате просмотра рекламы.

Побудительность рекламы во многом определяется ее мотивационным воздействием, изучение силы которого представляет чрезвычайно сложную задачу. Исследование в данном случае носит длительный характер в виде неструктурированного интервьюирования. Результаты исследования носят чрезвычайно субъективный характер, в значительной мере зависящий от личности интервьюера, выбранного метода анализа и интерпретации полученных результатов.

*4. Влияние рекламы на покупательское поведение.* В конечном итоге эффективность рекламной кампании определяется совершенными покупками. Известно, что мнение респондента может расходиться с поведением. Поэтому необходимо задавать вопросы о потребительском поведении, например, какую марку категории N вы купили в последний раз? Какую марку предпочитаете? и т.д.

Эффективность влияния рекламы на покупательское поведение можно оценить на основании возврата купонов — специальных талонов, прилагаемых к рекламному объявлению и дающих право на льготную покупку или бесплатное получение товара. Желая получить такой купон, люди следят за рекламными объявлениями и читают их. Когда потребители возвращают эти купоны производителю, выражая желание либо поближе познакомиться с товаром, либо вступить в дискуссию, они тем самым проявляют свой интерес к тому, что рекламируется. Возвращенные купоны, которые свидетельствуют о желании купить товар или получить скидку, — мерило реального покупательского спроса. Но если за возврат купона обещано достаточно соблазнительное вознаграждение, есть опасность, что откликнутся даже те, кто совершенно не заинтересован в товаре. Многим просто хочется хоть что-нибудь получить бесплатно.

Для коммуникативного анализа рекламных сообщений можно использовать различные методы, в том числе:

*Отзыв с помощью* состоит в том, что после просмотра рекламной продукции клиента задаются вопросы о том, повлияла ли реклама на их мнение о фирме и ее продуктах или отношение к фирме было сформировано заранее и осталось неизменным после получения рекламной информации.

*Отзыв без помощи* реализуется в форме анкетного опроса с заранее подготовленными ответами.

*Методы Гэллапа-Робинсона и Старча* позволяют оценить запоминаемость рекламы посредством опроса клиентов через несколько дней после показа рекламы.

*Метод купонов* заключается в том, что в рекламное сообщение включается купон на некоторые льготы, например, скидку. По числу поступивших купонов оценивается сравнительный эффект от размещения обращений в различных СМИ.

Но важно отметить, что имеются еще и другие подходы к оценке эффективности рекламной кампании. Например, такие, которые позволяют оценивать информационные и финансовые показатели.

*Информационная составляющая* включает в себя количество контактов с потенциальными покупателями и эффект от ее воздействия. Она анализируется для оценки влияния рекламного объявления на потребителя, а также для определения того, выделяется ли реклама из общей массы, вызывает ли она желание совершить покупку или задуматься об этом.



Для оценки информационной составляющей используются следующие методы:

- опросы;
- работы с фокус-группами;
- наблюдения со сквозной аналитикой;
- А/В-тесты основной идеи рекламной кампании.

*Финансовая составляющая* выражается величиной прибыли, полученной в результате рекламной кампании. Для оценки ее эффективности используются следующие метрики:

- показатели окупаемости инвестиций ROI и ROMI;
- рентабельность рекламных расходов ROAS;
- жизненный цикл клиента LTV.

Важно отметить, что оценка эффективности рекламы также требует анализа качества самой рекламы, а именно проанализировать ролики, баннеры, креативы и внести необходимые изменения и улучшения.

**Методы оценки рекламы в интернет.** Основное преимущество рекламы в интернете состоит уже в том, что она изначально намного дешевле других видов рекламы. Если для распространения информационной продукции вы используете вместо традиционных – электронные каналы коммуникации, то сразу снижаются затраты на всю сопутствующую деятельность – на телефонные разговоры, пересылку факсов, полиграфию и т.п. А также, за счет предоставления необходимой информации, обеспечивается предпродажная и послепродажная поддержка потребителей, а при создании *виртуального магазина* не нужны все те затраты, что связаны с магазином традиционным (аренда и оборудование торгового помещения и т.п.)

Кроме того, можно утверждать, что, кроме снижения затрат, присутствие фирмы в интернет посредством Web-сервера может принести прибыль за счет:

- повышения имиджа торговой марки компании;
- продвижения товаров фирмы;
- привлечения новых потребителей;
- добавления нового канала распространения продукции, улучшения сервисного обслуживания текущих и потенциальных потребителей.

Разрабатывая рекламную кампанию для начала нужно изучить факторы, влияющие на эффективность рекламы:

- место размещения на странице (реклама может быть не видна пользователю без дополнительной прокрутки экрана в браузере), размер рекламы, формат рекламы - эти параметры влияют на заметность и степень воздействия на пользователя;

- тематика ресурса, демо-портрет аудитории и т.п. влияют на то, насколько мы попадаем в целевую аудиторию, какое отношение у нее изначально будет к рекламе;

- качество работы сервера, от которых зависит качество и скорость загрузки сайта;

- корректная работа службы редиректа на сайт рекламодателя (при работе с баннерными сетями);

- неучтенные посещения, создающие погрешности при оценке (пользователи могут достать страницу и/или баннеры из кэша, и сервер не зафиксирует запрос и показ баннера);

- разрешение экрана пользователя (при более низком, чем в рекламном сообщении – оно может быть не видно).

При выборе места для размещения рекламы компания предоставляет клиенту на рассмотрение набор различных рекламных площадок, в котором рекламодатель имеет доступ к следующей информации:

- посещаемость сайта;
- его индекс цитирования (сколько ссылок на эту страницу ведет с других сайтов);
- стоимость рекламных мест;
- количество партнеров ресурса.

Всё это позволит оценить будущую эффективность рекламного сообщения.

Когда все факторы будут изучены, правильно выбрана аудитория и площадки для размещения, остается вычислить, какую реальную пользу все это приносит.

Чтобы понять насколько эффективно функционирует реклама в интернет, нужно подсчитать соотношение затрат на рекламную кампанию и ее результативности.

#### **Технические данные для анализа:**

*Счётчики* (они размещаются непосредственно на сайте, как правило, в открытом доступе для всех посетителей);

*Cookie-файлы* (файлы, содержащие различную динамическую информацию, сохраняющиеся на компьютере пользователя);

*Log-файлы* (файлы, протоколирующие события разного характера, происходящие на сайте);

Их настройки задаются специальными программами и хранятся на сервере производителя сайта.

Первичными показателями, используемыми при оценке посещаемости рекламируемого сайта, являются ХИТ и ХОСТ.

ХОСТ — это пользователь, просматривающий страницы. С уникальным пользователем ассоциируется уникальный IP-адрес компьютера, с которого выполняется доступ.

ХИТ — просмотр сайта.

Число *хитов* на сайте в единицу времени дает возможность оценки рекламной мощности сайта.

Кроме того, часть данных по рекламной кампании поступает непосредственно от сайтов, размещающих у себя рекламное сообщение. Это:

- график и схема размещения;

- количество показов, кликов (как минимум, с разбивкой по каждому из дней, по каждому из рекламных сообщений, по каждой схеме размещения);

- дополнительные данные. Это информация, полученная о посетителях, собранная в базу данных (к примеру, анкеты, регистрация, база IP-адресов и т.п.).

Внутри компании рекламодатель сам может проводить анализ звонков, поступающих заявок, покупок, контрактов и т.п.

Используя полученные технические данные, о которых говорилось выше, условно можно выделить 3 основных общепринятых обозначения показателей эффективности интернет-рекламы.

1. CTR (синоним — кликабельность, от англ. Click-Through Rate — показатель кликабельности, по-русски может называться откликом) – основной показатель эффективности интернет-рекламы:

$$CTR = (\text{число кликов}) / (\text{число показов}),$$

где клик — одно нажатие на рекламное сообщение, показ — одно предъявление рекламного сообщения посетителю веб-сайта. CTR измеряется в процентах, и является важным показателем эффективности работы рекламного сообщения.

2. СТВ (Click-To-Buy ratio) – показатель эффективности интернет-рекламы, измеряемый как отношение:

$$СТВ = (\text{число покупателей}) / (\text{общее число посетителей})$$

Показатель СТВ отражает конверсию посетителей в покупателей, его иногда называют *коэффициентом конверсии*.

3. СТИ (Click-To-Interest) – показатель эффективности интернет-рекламы, измеряемый как отношение:

$$СТИ = (\text{число заинтересованных посетителей}) / (\text{общее число посетителей})$$

Заинтересованным считается тот посетитель сайта, который пролистал несколько его страниц, либо вернулся сюда снова, либо запомнил адрес сайта и факт его существования.

CTR зависит от вида рекламного сообщения и обстоятельств его показа. СТВ и СТИ зависят от сервера рекламодателя.

Все приведенные выше показатели эффективности исследуют поодиночке, сочетают друг с другом, и в соответствии с результатами анализа воздействуют на ход рекламной кампании интернет-проекта.

Обычно такую работу доверяют специалистам, которые предоставляют по желанию клиента либо отдельные данные, либо уже готовые подсчеты.

Контроль за рекламной кампанией, анализ ее текущих результатов и оценка их эффективности — это неотъемлемая стадия всего процесса планирования рекламных мероприятий. До запуска рекламной кампании фирма помимо плана рекламных мероприятий не располагает ничем, а сам план представляет собой лишь предположение о том, как должна работать реклама. Настоящий специалист захочет проверить это предположение при помощи исследований в течение кампании и после нее. Безусловно, проще всего проверить рекламу, после которой начинается непосредственное выполнение ее целей (быстрый рост сбыта и активности потребителей). Гораздо труднее отслеживать маркетинговые коммуникации, включающие в себя спонсорство, публицити, проведение каких-либо мероприятий и т.д.

Контроль за рекламной кампанией не является обязательной процедурой, и в большинстве случаев компании не контролируют ход самого процесса. Вместо этого менеджеры анализируют показатели сбыта продукции до начала и в течение всей рекламной кампании и затем делают вывод о том, работает реклама или нет.

Крупные фирмы контролируют и оценивают свои рекламные кампании даже тогда, когда все рыночные факторы говорят о том, что реклама (меры стимулирования, комплекс интенсивных маркетинговых коммуникаций) достигает целей по сбыту. Руководитель отдела рекламы и маркетинга, заинтересованный в успехе, всегда должен быть уверен в том, что сбыт продукции обеспечен именно рекламой, а не иными факторами маркетинга, «промахами» со стороны конкурентов или самим рынком (при изменении общеэкономических условий).

Но если оценка эффективности рекламы все-таки нужна, то почему ее не всегда делают даже те компании, для которых это важно? Вряд ли найдется руководитель компании, который, глядя на рекламный бюджет, не задумывался о том, стоит ли тратить столько на рекламу и что в результате получит бизнес.

Отслеживание рекламной кампании требует значительного количества финансовых средств. Для начала требуется провести предварительное, или контрольное, исследование, если оно не проводилось при разработке рекламной стратегии. Затем, уже при проведении кампании, проводятся одна или несколько «волн» исследований, результаты которых сравниваются с «контрольными».

Таким образом, осуществляется слежение за эффективностью рекламы. Стоимость каждого промежуточного исследования зависит от поставленных целей опроса, размера выборки респондентов, объема анкеты или вопросника, вероятности совершения покупки, но в любом случае она остается довольно высокой. Также усложняет оценку многообразие эффектов, целей и задач рекламы, задержка эффекта рекламы и «накапливание» ее воздействия. И, наконец, отсутствие единой (работающей во всех ситуациях) модели воздействия рекламы.

Но при этом, когда действительно потребуются оценка эффективности рекламы, учиться этому будет уже поздно. Чтобы она стала полезным инструментом оптимизации рекламных планов и бюджетов, необходимо много вещей, которые не случаются сразу: нужны накопленная компанией статистика, опыт использования полученной информации, «привычка» планировать рекламу, опираясь на цифры и задачи бизнеса, нужны люди, для которых реклама — это не только и не столько творчество. Поэтому хорошее планирование и управление рекламой предполагают оценку её эффективности.

### **Способы оценки эффективности рекламы.**

Соотношение объемов продаж продукта (в денежном выражении) или прибыли до и после проведения рекламной кампании и затраченной на нее суммы.

Изменение процента информированности заданной рекламной аудитории о рекламодателе, его товарном знаке, фирме.

Часто для оценки эффективности рекламной кампании используется способ *опроса по телефону*. При общении клиенту задается вопрос, откуда он получил информацию о фирме и ее продуктах. Затем осуществляется обработка полученных данных и выявляется наиболее эффективный рекламный источник.

Эффективность размещения рекламы в том или ином медиасредстве можно определить с помощью ряда показателей.

**Показатели эффективности размещения рекламы на телевидении:**

*Прогнозы эффективности телевизионных средств:*

- прогноз рейтинга временного интервала телеканала (аудитория канала в определенное время/ аудитория потенциальных телезрителей);
- прогноз рейтинга конкретной телепередачи (аудитория определенной передачи/ аудитория потенциальных потребителей);
- прогноз среднего рейтинга телепередачи (средняя аудитория определенной передачи/аудитория потенциальных телезрителей);
- прогноз GRP (сумма рейтингов всех рекламных сообщений на данном канале).

*Прогнозы эффективности рекламной кампании:*

- оценочная средняя аудитория рекламного ролика (средняя аудитория телепередачи в которой был размещен рекламный ролик);
- оценочная аудитория рекламной кампании (количество индивидов целевой аудитории из потенциальной аудитории телевиденья, которые видели рекламное сообщение хотя бы один раз);
- оценочный рейтинг рекламного ролика (средняя аудитория ролика/целевая аудитория);
- суммарный GRP (сумма рейтингов всех рекламных сообщений за весь период рекламной кампании).

*Прогноз показателей адресности рекламоносителя* – демонстрируют, насколько целевая аудитория рекламоносителя подходит целевой аудитории рекламной кампании:

- Conversion Index или индекс соответствия GRP (вся целевая аудитория/ GRP x 100);
- Affinity Index или индекс соответствия охвата (охват целевой аудитории/охват потенциальной аудитории).

*Прогнозы стоимости показателей рекламной кампании:* прогноз CPT (бюджет рекламной кампании/целевая аудитория в тысячах человек) и прогноз C33 (бюджет рекламной кампании / суммарный рейтинг).

*Показатели эффективности размещения рекламы на радио* практически идентичны телевиденью, однако существует несколько отличных терминов:

- прогноз дневного охвата (индивиды, которые слушали радио хотя бы пять минут в день/ потенциальная аудитория радиостанции);
- прогноз недельного охвата (индивиды, которые слушали радио хотя бы пять минут за неделю/ потенциальная аудитория радиостанции).

**Прогнозы эффективности использования печатных изданий:**

- прогноз CPT (стоимость размещения стандартного модуля/потенциальная аудитория издания);
- AIR – средняя аудитория одного номера (средняя аудитория одного номера/потенциальная аудитория);
- Affinity Index - индекс соответствия (рейтинг рекламоносителя в целевой аудитории/ рейтинг рекламоносителя в потенциальной аудитории);
- ARI коэффициент прочтения одного номера (количество индивидов, прочитавших или пролиставших данное издание, кроме непосредственного покупателя).

**Износ рекламы.** Износ рекламы – это состояние, когда рекламное сообщение перестает быть эффективным и оказывать нужное воздействие на сознание потребителей. С помощью исследований необходимо определить момент износа рекламы и после этого выявить причины возникшего состояния.

*Причины износа рекламы и снижения ее эффективности:*

- устаревание рекламной стратегии;
- потеря актуальности рекламного сообщения;
- износ креативной идеи рекламного сообщения;
- изменение конкурентов;
- изменение потребителей;
- изменение минимальной эффективной частоты;
- уменьшение уровня привлечения внимания;
- уменьшение уровня запоминаемости;
- неприятие рекламы (она начинает вызывать раздражение).

Процесс выявления причин износа рекламы можно разделить на несколько этапов. В первую очередь необходимо выяснить нет ли других причин снижения продаж. Одной из причин может быть устаревший маркетинг. Специалисты выделяют три группы возможных вариантов: изменение имиджевых факторов предприятия и продукта; изменение маркетинга конкурирующих фирм и изменение предпочтений покупателей.

После проверки общего плана маркетинга приступают к анализу рекламной части. Причины неудач могут выявиться на следующих стадиях: контакт аудитории с рекламой, усвоение рекламной информации, стратегия коммуникаций, анализ поведения целевой аудитории. В случае износа рекламы на стадии контакта с аудиторией необходимо повторно оценить эффективность рекламы. Выбранные СМИ могут утратить популярность, рекламное сообщение при той же частоте выхода в свет будет контактировать с меньшей аудиторией. Если график показа рекламы выбран неудачно, необходимо скорректировать его с учетом бюджета.

Часто причиной снижения эффективности является устаревание рекламного сообщения. Это происходит при многократном повторе рекламного объявления.

*Меры борьбы с износом рекламы:*

- при снижении уровня внимания – создание серии рекламных сообщений, содержащих только одну креативную идею;
- при уменьшении уровня запоминания – изменение графика подачи рекламы;
- при неприятии рекламного сообщения – использование укороченных вариантов рекламы, создание новых рекламных сообщений.

Таким образом, процессы оценки эффективности рекламной деятельности зависят от внешних и внутренних факторов, которые выступают в роли переменных величин и подвержены динамике в условиях даже стабильного рынка. Именно динамичность следует рассматривать в качестве характеристики рекламной деятельности, оценка эффективности которой требует налаживания систем статистического учета необходимых показателей. Расчет и анализ показателей рекламной деятельности вполне может быть обоснован использованием только таких показателей, как:

- сводка и группировка;
- расчет относительных величин;
- средние величины и показатели вариации;
- применение выборочного метода;
- расчет показателей динамики;
- корреляционно-регрессионный анализ.

Но для понимания сути происходящих процессов только этих показателей недостаточно. Поэтому в учебнике на основе теории статистики показаны возможности более глубокого количественного анализа эффективности рекламной деятельности.

**Вопросы для самопроверки:**

1. Перечислите факторы коммуникативной эффективности рекламы.
2. Какие Вам известны показатели медиапланирования, применяемые для расчета эффективности рекламных кампаний?
3. Рейтинг, как понятие эффективности рекламы, обозначает...
4. В чем смысл понятия «доля аудитории передачи»?
5. Узнаваемость рекламы – это...
6. Запоминаемость рекламы – это...
7. Что обозначает понятие «внедрение рекламы» в теории рекламы?
8. Что обозначает понятие «вовлечение в потребление» в теории рекламы?

## Раздел 2. Методы обобщения статистической информации

### *Тема 3. Общетеоретическое понимание основных категорий и понятий статистики. Формы представления статистических данных*

В научный обиход 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 велся намного раньше: проводились переписи населения в Древнем Китае, осуществлялось сравнение военного потенциала государств, велся учет имущества граждан в Древнем Риме и пр.

У истоков статистической науки стояли две школы: немецкая описательная и английская школа политических арифметиков. Представители описательной школы (Герман Конринг, Готфрид Ахенваль, Август Людвиг Шленцер) своей задачей считали описание достопримечательностей государства: территории, населения, климата, политического устройства, вероисповедания, торговли и т.п. – без анализа закономерностей и связей между явлениями. Представители школы политических арифметиков (Уильям Петти, Джон Граунт, Эдмунд Галлей) своей главной задачей считали выявление на основе большого числа наблюдений различных закономерностей и взаимосвязей в изучаемых явлениях. Каждая школа развивалась своим путем, используя свои методы в исследованиях, но предмет изучения у них был общий – государство, общество и, в частности, массовые явления и процессы, происходящие в нем. Статистика сформировалась как наука в результате синтеза государствоведения и политической арифметики, причем от последней она взяла больше, поскольку статистика и в настоящее время призвана выявлять прежде всего различного рода закономерности в исследуемых явлениях.

Однако представители этих двух школ не дошли до теоретического обобщения практики учетно - статистических работ, до создания теории статистики. Эта задача была решена позднее, в XIX веке бельгийским ученым Адольфом Кетле, который дал определение предмета статистики, раскрыл суть ее методов. Под влиянием идей Кетле возникло третье направление статистической науки – математико-статистическое, которое получило свое развитие в работах таких ученых как: англичане Фрэнсис Гальтон, Фрэнсис Эджворт, Карл Пирсон, Одни Дж. Юл, Вильям Госсет, Рональд Фишер, Морис Дж. Кендэл, итальянец Коррадо Джини, русские – Пафнутий Львович Чебышёв, Андрей Андреевич Марков, Александр Михайлович Ляпунов, Александр Иванович и Александр Александрович Чупров и пр.

В настоящее время данный термин употребляется в четырех значениях:

1) наука, изучающая количественную сторону массовых явлений и процессов в неразрывной связи с их качественным содержанием – учебный предмет в высших и средних специальных учебных заведениях;

2) совокупность цифровых сведений, характеризующих состояние массовых явлений и процессов общественной жизни; статистические данные, представляемые в отчетности предприятий, организаций, отраслей экономики, а также публикуемых в сборниках, справочниках, периодической печати и в сети Интернет, которые являются результатом статистической работы;

3) отрасль практической деятельности («статистический учет») по сбору, обработке, анализу и публикации массовых цифровых данных о самых различных явлениях и процессах общественной жизни;

4) некий параметр ряда случайных величин, получаемый по определенному алгоритму из результатов наблюдений, например, статистические критерии (критические статистики), применяющиеся при проверке различных гипотез (предположительных утверждений)

относительно природы или значений отдельных показателей исследуемых данных, особенностей их распределения и пр.

Как и любая другая наука, статистика имеет свой предмет и метод исследования. Статистика изучает количественную сторону массовых общественных явлений в неразрывной связи с их качественной стороной или содержанием, а также исследует количественное выражение закономерностей общественного развития в конкретных условиях места и времени. Такое изучение основывается на системе категорий и понятий, отражающих наиболее общие и существенные свойства, признаки, связи и отношения предметов и явлений объективного мира.

Рассмотрим основные понятия, используемые в статистике.

1. Статистическая совокупность – множество социально - экономических объектов или явлений общественной жизни, объединенных качественной основой, но отличающихся друг от друга отдельными признаками, то есть однородных в одном отношении, но разнородных в другом. Таковы, например, совокупность домохозяйств, семей, предприятий, фирм и т.п.

2. Единица совокупности – первичный элемент статистической совокупности, являющийся носителем признаков и основой ведущегося при обследовании счета.

3. Признак единицы совокупности – свойства единицы совокупности, которые различаются способами их измерения и другими особенностями, что дает основание для их классификации.

4. Статистический показатель – понятие, отображающее количественные характеристики (размеры) или соотношения признаков общественных явлений. Статистические показатели можно подразделить на **первичные** (объемные) – характеризуют либо общее число единиц совокупности (объем совокупности), либо сумму значений какого-либо признака (объем признака) и выражаются абсолютными величинами и **вторичные** (расчетные) – задаются на единицу первичного показателя и выражаются относительными и средними величинами. Статистические показатели могут быть плановыми, отчетными и прогностическими.

Таблица 2.

**Основная классификация признаков в статистике**

| Параметр классификации  | Вид признака                | Пример признака                     |
|-------------------------|-----------------------------|-------------------------------------|
| По характеру выражения  | Описательные (атрибутивные) | Цвет волос человека                 |
|                         | Количественные (числовые)   | Рост человека                       |
| По способу измерения    | Первичные (объемные)        | Вес человека                        |
|                         | Вторичные (расчетные)       | Производительность труда            |
| По характеру вариации   | Альтернативные              | Пол человека                        |
|                         | Дискретные                  | Возраст человека                    |
|                         | Интервальные                | Возраст группы людей                |
| По отношению ко времени | Моментные                   | Количество денег в кармане человека |
|                         | Периодные                   | Заработная плата человека за месяц  |

5. Система статистических показателей – совокупность статистических показателей, отражающая взаимосвязи, которые объективно существуют между явлениями. Она охватывает все стороны общественной жизни как на макро-, так и на микроуровне. С изменением условий



жизни общества меняются и системы статистических показателей, совершенствуется методология их расчета.

Совокупность приемов, пользуясь которыми статистика исследует свой предмет, составляет метод статистики. Можно выделить три группы статистических методов (этапов статистического исследования):

- 1) статистическое наблюдение;
- 2) сводка;
- 3) научный анализ исследуемых явлений.

Статистическое изучение тех или иных явлений предполагает, как обязательное условие наличие информации, то есть сведений об этих явлениях. Поэтому первый этап, начало статистического исследования сводится к сбору необходимой информации. Научно организованный сбор сведений, заключающийся в регистрации тех или иных фактов, признаков, относящихся к каждой единице изучаемой совокупности, называется статистическим наблюдением.

В результате статистического наблюдения образуется масса первичной информации (сведений) о каждой единице совокупности. Чтобы получить характеристику всей исследуемой совокупности в целом, первичные данные должны быть подвергнуты обработке, обобщению. Обработка собранных первичных данных, включающая их группировку, обобщение и оформление в таблицах, составляет второй этап статистического исследования, который называется сводкой.

На третьем этапе статистического исследования на основе итоговых данных сводки осуществляется научный анализ исследуемых явлений: рассчитываются различные обобщающие показатели в виде средних и относительных величин, выявляются определенные закономерности в распределениях, динамике показателей и т.п.

Любое законченное статистическое исследование проходит в три этапа, между которыми могут быть перерывы во времени.

#### **Формы представления статистических данных**

Статистические данные должны быть представлены так, чтобы ими можно было пользоваться. Существует три основных формы представления статистических данных:

- 1) текстовая – включение данных в текст;
- 2) табличная – представление данных в таблицах;
- 3) графическая – выражение данных в виде графиков.

Текстовая форма применяется при малом количестве цифровых данных.

Табличная форма применяется чаще всего, так как является более эффективной формой представления статистических данных. В отличие от математических таблиц, которые по начальным условиям позволяют получить тот или иной результат, статистические таблицы рассказывают языком цифр об изучаемых объектах.

*Статистическая таблица* – это система строк и столбцов, в которых в определенной последовательности и связи излагается статистическая информация о социально-экономических явлениях.

Таблица 2.

Пример таблицы: Внешняя торговля РФ за 2000 – 2006 годы, млрд. долл.

| Показатель                     | 2000  | 2001  | 2002  | 2003  | 2004  | 2005  | 2006  |
|--------------------------------|-------|-------|-------|-------|-------|-------|-------|
| Внешнеторговый оборот          | 149,9 | 155,6 | 168,3 | 212   | 280,6 | 368,9 | 468,4 |
| Экспорт                        | 105   | 101,9 | 107,3 | 135,9 | 183,2 | 243,6 | 304,5 |
| Импорт                         | 44,9  | 53,8  | 61    | 76,1  | 97,4  | 125,3 | 163,9 |
| Сальдо торгового баланса       | 60,1  | 48,1  | 46,3  | 59,9  | 85,8  | 118,3 | 140,7 |
| в том числе:                   |       |       |       |       |       |       |       |
| со странами дальнего зарубежья |       |       |       |       |       |       |       |
| экспорт                        | 90,8  | 86,6  | 90,9  | 114,6 | 153   | 210,1 | 261,1 |
| импорт                         | 31,4  | 40,7  | 48,8  | 61    | 77,5  | 103,5 | 138,6 |
| сальдо торгового баланса       | 59,3  | 45,9  | 42,1  | 53,6  | 75,5  | 106,6 | 122,5 |

Например, в табл. 2 представлена информация о внешней торговле России, выразить которую в текстовой форме было бы неэффективным.

Различают *подлежащее* и *сказуемое* статистической таблицы. В подлежащем указывается характеризуемый объект – либо единицы совокупности, либо группы единиц, либо совокупность в целом. В сказуемом дается характеристика подлежащего, обычно в числовой форме. Обязателен *заголовок* таблицы, в котором указывается к какой категории и к какому времени относятся данные таблицы.

По характеру подлежащего статистические таблицы подразделяются на *простые*, *групповые* и *комбинационные*. В подлежащем простой таблицы объект изучения не подразделяется на группы, а дается либо перечень всех единиц совокупности, либо указывается совокупность в целом. В подлежащем групповой таблицы объект изучения подразделяется на группы по одному признаку, а в сказуемом указываются число единиц в группах (абсолютное или в процентах) и сводные показатели по группам. В подлежащем комбинационной таблицы совокупность подразделяется на группы не по одному, а по нескольким признакам.

При построении таблиц необходимо руководствоваться следующими *общими правилами*.

1. Подлежащее таблицы располагается в левой (реже – верхней) части, а сказуемое – в правой (реже – нижней).

2. Заголовки столбцов содержат названия показателей и их единицы измерения.

3. Итоговая строка завершает таблицу и располагается в ее конце, но иногда бывает первой: в этом случае во второй строке делается запись «в том числе», и последующие строки содержат составляющие итоговой строки.

4. Цифровые данные записываются с одной и той же степенью точности в пределах каждого столбца, при этом разряды чисел располагаются под разрядами, а целая часть отделяется от дробной запятой.

5. В таблице не должно быть пустых клеток: если данные равны нулю, то ставится знак «–» (прочерк); если данные не известны, то делается запись «сведений нет» или ставится знак «...» (троеточие). Если значение показателя не равно нулю, но первая значащая цифра появляется после принятой степени точности, то делается запись 0,0 (если была принята степень точности 0,1).

Иногда статистические таблицы дополняются графиками, когда ставится цель подчеркнуть какую-то особенность данных, провести их сравнение. Графическая форма является самой эффективной формой представления данных с точки зрения их восприятия. С помощью графиков достигается наглядность характеристики структуры, динамики, взаимосвязи явлений, их сравнения.

*Статистические графики* – это условные изображения числовых величин и их соотношений посредством линий, геометрических фигур, рисунков или географических карт-схем. Графическая форма облегчает рассмотрение статистических данных, делает их наглядными, выразительными, обозримыми. Однако графики имеют определенные ограничения: прежде всего, график не может включить столько данных, сколько может войти в таблицу; кроме того, на графике показываются всегда округленные данные – не точные, а приблизительные. Таким образом, график используется только для изображения общей ситуации, а не деталей. Последний недостаток – трудоемкость построения графиков. Он может быть преодолен использованием персонального компьютера (например, «Мастером диаграмм» из пакета *Microsoft Office Excel*).

По способу построения графики делятся на *диаграммы, картограммы и картодиаграммы*.

Наиболее распространенным способом графического изображения данных являются диаграммы, которые бывают следующих видов: линейные, радиальные, точечные, плоскостные, объемные, фигурные. Вид диаграмм зависит от вида представляемых данных и задачи построения. В любом случае график обязательно сопровождается заголовком – над или под полем графика. В заголовке указывается, какой показатель изображен, по какой территории и за какое время.

Линейные графики используются для представления количественных переменных: характеристики вариации их значений, динамики, взаимосвязи между переменными. Вариация данных анализируется с помощью *полигона распределения, кумуляты* (кривой «меньше, чем») и *огивы* (кривой «больше, чем»). Для построения кумуляты значения варьирующего признака откладываются по оси абсцисс, а на оси ординат помещаются накопленные итоги частот или частостей (от  $f_1$  до  $\sum f$ ). Для построения огивы на оси ординат помещаются накопленные итоги частот в обратном порядке (от  $\sum f$  до  $f_1$ ). Кумуляту и огиву по данным табл. 4 изобразим на рис. 2.

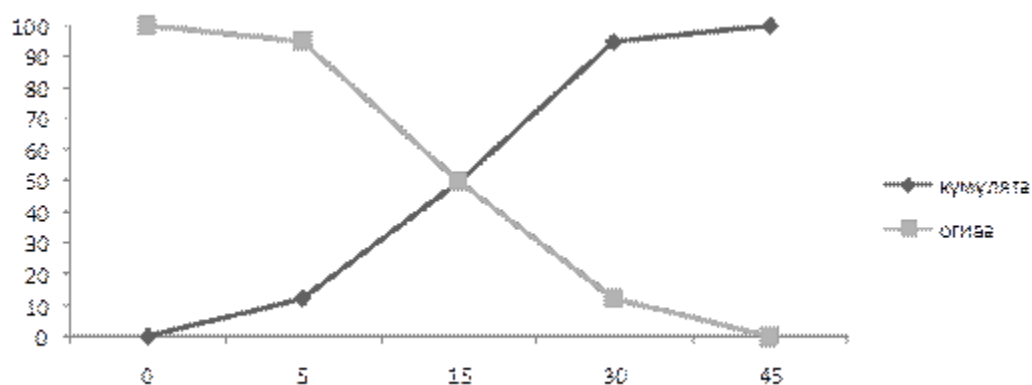


Рис. 2. Кумулята и огива распределения товаров по величине таможенной стоимости

Линейные графики подразделяются на *одномерные*, используемые для представления данных по одной переменной, и *двумерные* – по двум переменным. Примером одномерного линейного графика является полигон распределения, а двумерного – линия регрессии.

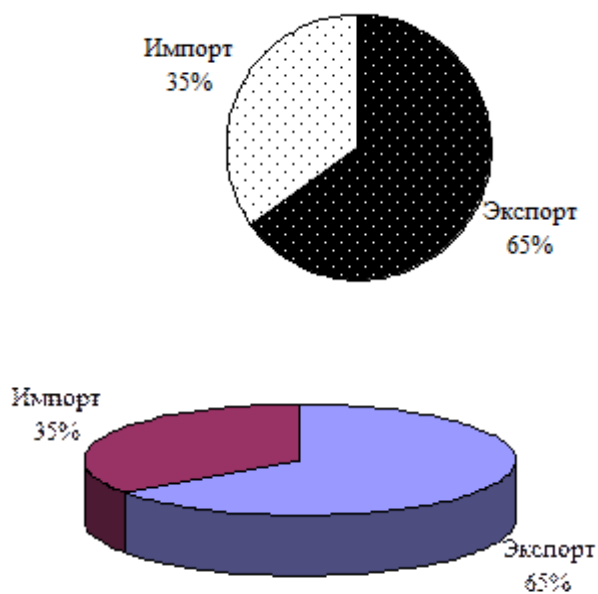
Иногда при больших изменениях показателя прибегают к логарифмической шкале. Например, если значения показателя изменяются от 1 до 1000, то это может вызвать затруднения при построении графика. В таких случаях переходят к логарифмам значений показателя, которые не будут столь сильно различаться:  $lg 1 = 0$ ,  $lg 1000 = 3$ .

Среди *плоскостных* диаграмм по частоте использования выделяются столбиковые диаграммы (гистограммы), на которых показатель представляется в виде столбика, высота которого соответствует значению показателя (рис. 4).

Пропорциональность площади той или иной геометрической фигуры величине показателя лежит в основе других видов плоскостных диаграмм: *треугольных*, *квадратных*, *прямоугольных*. Можно использовать и сравнение площадей круга – в этом случае задается радиус окружности.

*Ленточная диаграмма* представляет показатели в виде горизонтально вытянутых прямоугольников, а в остальном не отличается от столбиковой диаграммы.

Из плоскостных диаграмм часто используется *секторная диаграмма*, которая применяется для иллюстрации структуры изучаемой совокупности. Вся совокупность принимается за 100%, ей соответствует общая площадь круга, площади секторов соответствуют частям совокупности. Построим секторную диаграмму структуры внешней торговли РФ в 2006 году по данным табл. 2 (рис. 3). При использовании компьютерных программ секторные диаграммы строятся в объемном виде, то есть не в двух, а в трех плоскостях (рис. 4).



**Рис. 3. Простая секторная диаграмма**

**Рис. 4. Объемная секторная диаграмма**

Фигурные (картинные) диаграммы усиливают наглядность изображения, так как включают рисунок изображаемого показателя, размер которого соответствует размеру показателя.

При построении графика одинаково важно все – правильный выбор графического изображения, пропорций, соблюдение правил оформления графиков.

Картограммы и картодиаграммы применяются для изображения географической характеристики изучаемых явлений. Они показывают размещение изучаемого явления, его интенсивность на определенной территории – в республике, области, экономическом или административном округе и т.д. Построение картограмм и картодиаграмм рассматривается в специальной литературе.

#### **Вопросы для самопроверки:**

1. Учетно – статистические работы – это...
2. Статистический учет необходим для того, чтобы...
3. Что обозначает статистическая совокупность?
4. Что такое «единица совокупности»?
5. Признак единицы совокупности – это...
6. Статистический показатель обозначает...
7. Система статистических показателей – это..
8. Какие Вам известны основных формы представления статистических данных?
9. Что такое статистическая таблица?
10. Что обозначают подлежащее и сказуемое статистической таблицы? Что это такое?
11. Для чего нужны статистические графики?
12. Что обозначают картограммы и картодиаграммы?

### ***Тема 4. Построение ряда распределения. Расчет структурных характеристик ряда распределения***

Признаки, изучаемые статистикой, варьируются (отличаются друг от друга) у различных единиц совокупности в один и тот же период или момент времени. Например, величина внешнеторгового оборота варьируется по подразделениям ФТС; величина экспорта (импорта) варьируется по направлениям экспорта (по разным странам – партнерам, по внешней торговле), по видам товаров и т.п.

Причиной *вариации* являются разные условия существования разных единиц совокупности. Например, огромное число причин влияет на масштабы внешней торговли различных стран мира.

Для управления и изучения вариации статистикой разработаны специальные методы исследования вариации, система показателей, с помощью которой вариация измеряется, характеризуются ее свойства.

Первым этапом статистического изучения вариации является построение *ряда распределения* (или *вариационного ряда*) – упорядоченного распределения единиц совокупности по возрастающим (чаще) или по убывающим (реже) значениям признака и подсчет числа единиц с тем или иным значением признака.

Существует три *вида* ряда распределения:

1) *ранжированный ряд* – это перечень отдельных единиц совокупности в порядке возрастания изучаемого признака; если численность единиц совокупности достаточно велика ранжированный ряд становится громоздким, и в таких случаях ряд распределения строится с помощью группировки единиц совокупности по значениям изучаемого признака (если признак принимает небольшое число значений, то строится дискретный ряд, а в противном случае – интервальный ряд);

2) *дискретный ряд* – это таблица, состоящая из двух столбцов (строк) – конкретных значений варьирующего признака  $X_i$  и числа единиц совокупности с данным значением признака  $f_i$  – частот; число групп в дискретном ряду определяется числом реально существующих значений варьирующего признака;

3) *интервальный ряд* – это таблица, состоящая из двух столбцов (строк) – интервалов варьирующего признака  $X_i$  и числа единиц совокупности, попадающих в данный интервал (частот), или долей этого числа в общей численности совокупностей (частостей).

Таблица 3.

**Внешнеторговый оборот (ВО) по 35 таможенным постам, млн. у.е.**

| № поста | ВО    | № поста | ВО    | № поста | ВО      |
|---------|-------|---------|-------|---------|---------|
| 1       | 24,16 | 13      | 54,12 | 25      | 65,31   |
| 2       | 27,06 | 14      | 54,91 | 26      | 69,24   |
| 3       | 29,12 | 15      | 55,74 | 27      | 71,39   |
| 4       | 31,17 | 16      | 55,91 | 28      | 77,12   |
| 5       | 37,08 | 17      | 56,07 | 29      | 79,12   |
| 6       | 39,11 | 18      | 56,80 | 30      | 84,34   |
| 7       | 41,58 | 19      | 56,93 | 31      | 86,89   |
| 8       | 44,84 | 20      | 57,07 | 32      | 91,74   |
| 9       | 46,80 | 21      | 58,39 | 33      | 96,01   |
| 10      | 48,37 | 22      | 59,61 | 34      | 106,84  |
| 11      | 51,44 | 23      | 59,95 | 35      | 111,16  |
| 12      | 52,56 | 24      | 62,05 | Итого   | 2100,00 |

Построим ряд распределения *внешнеторгового оборота (ВО)* по таможенным постам России, для чего необходимо провести статистическое наблюдение, то есть собрать первичный статистический материал, который представляет собой величину ВО по таможенным постам.

Результаты наблюдения ВО по 35 таможенным постам региона за отчетный период представим в виде ранжированного по возрастанию величины ВО ряда распределения (табл. 3).

Определим средний размер ВО по формуле (10), приняв за  $X$  величину ВО, а за  $N$  – численность постов:

$$\bar{X} = \frac{\sum X}{N} = 2100/35 = 60 \text{ (млн. долл.)}$$

Дисперсию (о ней будет рассказано чуть позднее – на 4 - м этапе анализа вариации в этой теме) определим по формуле:

$$\sigma^2 = \frac{\sum (X_i - \bar{X})^2}{N} = \frac{(24,16 - 60)^2 + (27,06 - 60)^2 + \dots + (111,16 - 60)^2}{35} = 445,778 \text{ (млн. у.е.)}$$

Построим интервальный ряд распределения ВО по таможенным постам, для чего необходимо выбрать оптимальное число групп (интервалов признака) и установить длину (размах) интервала. Поскольку при анализе ряда распределения сравнивают частоты в разных интервалах, необходимо, чтобы длина интервалов была постоянной. Оптимальное число групп выбирается так, чтобы достаточной мере отразилось разнообразие значений признака в совокупности и в то же время закономерность распределения, его форма не искажалась случайными колебаниями частот. Если групп будет слишком мало, не проявится закономерность вариации; если групп будет чрезмерно много, случайные скачки частот исказят форму распределения.

Чаще всего число групп в ряду распределения определяют по формуле Стерджесса:

$$k = 1 + 3,322 \lg N \quad \text{или} \quad k = 1 + 1,44 \ln N$$

Где  $k$  – число групп (округляемое до ближайшего целого числа);  $N$  – численность совокупности.

Из формулы Стерджесса видно, что число групп – функция объема данных ( $N$ ).

Зная число групп, рассчитывают длину (размах) интервала по формуле:

$$h = \frac{X_{\max} - X_{\min}}{k}$$

Где  $X_{\max}$  и  $X_{\min}$  — максимальное и минимальное значения в совокупности.

В нашем примере про ВО по формуле Стержесса определим число групп:

$$K = 1 + 3,322 \lg 35 = 1 + 3,322 * 1,544 = 6,129 \approx 6.$$

Рассчитаем длину (размах) интервала по формуле:

$$H = (111,16 - 24,16) / 6 = 87 / 6 = 14,5 \text{ (млн.у.е.)}$$

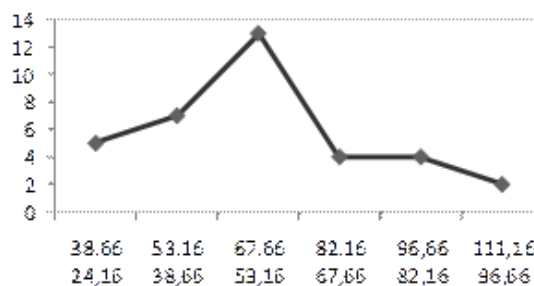
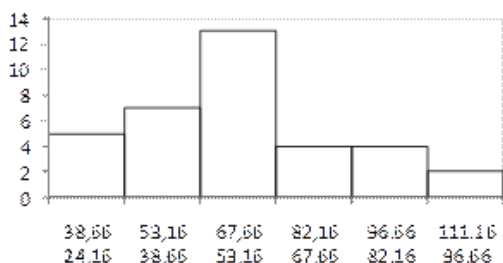
Теперь построим интервальный ряд с 6 группами с интервалом 14,5 млн. у.е. (первые 3 столбца табл. 4).

**Таблица 4.**

**Интервальный ряд распределения ВО по таможенным постам, млн. у.е.**

| $i$ | Группы постов по величине ВО<br>$X_i$ | Число постов<br>$f_i$ | Середина интервала<br>$X_i'$ | $X_i' f_i$ | Накопл. частота<br>$f_i'$ | $ X_i' - \bar{X}  f_i$ | $(X_i' - \bar{X})^2 f_i$ | $(X_i' - \bar{X})^3 f_i$ | $(X_i' - \bar{X})^4 f_i$ |
|-----|---------------------------------------|-----------------------|------------------------------|------------|---------------------------|------------------------|--------------------------|--------------------------|--------------------------|
| 1   | 24,16 – 38,66                         | 5                     | 31,41                        | 157,05     | 5                         | 147,071                | 4326,001                 | -127246,23               | 3742856,97               |
| 2   | 38,66 – 53,16                         | 7                     | 45,91                        | 321,37     | 12                        | 104,400                | 1557,051                 | -23222,31                | 346344,16                |
| 3   | 53,16 – 67,66                         | 13                    | 60,41                        | 785,33     | 25                        | 5,386                  | 2,231                    | -0,92                    | 0,38                     |
| 4   | 67,66 – 82,16                         | 4                     | 74,91                        | 299,64     | 29                        | 56,3                   | 793,629                  | 11178,84                 | 157461,90                |
| 5   | 82,16 – 96,66                         | 4                     | 89,41                        | 357,64     | 33                        | 114,343                | 3268,572                 | 93434,47                 | 2670891,13               |
| 6   | 96,66 – 111,16                        | 2                     | 103,91                       | 207,82     | 35                        | 86,171                 | 3712,758                 | 159966,81                | 6892284,32               |
|     | Итого                                 | 35                    |                              | 2128,85    |                           | 513,714                | 13660,243                | 114110,66                | 13809838,86              |

Существенную помощь в анализе ряда распределения и его свойств оказывает графическое изображение.



**Рис. 5. Гистограмма распределения**  
**Рис. 6. Полигон распределения**

Интервальный ряд изображается столбиковой диаграммой, в которой основания столбиков, расположенные по оси абсцисс, – это интервалы значений варьирующего признака, а высоты столбиков – частоты, соответствующие масштабу по оси ординат. Графическое изображение распределения таможенных постов в выборке по величине ВО приведено на рис. 5. Диаграмма такого типа называется *гистограммой*.

Данные табл. 4 и рис. 5 показывают характерную для многих признаков форму распределения: чаще встречаются значения средних интервалов признака, реже – крайние (малые и большие) значения признака. Форма этого распределения близка к нормальному закону распределения, которое образуется, если на варьирующую переменную влияет большое число факторов, ни один из которых не имеет преобладающего значения.

Если имеется дискретный ряд распределения или используются середины интервалов (как в нашем примере про ВО – в табл. 4 в 4 - м столбце рассчитаны середины интервалов как полусумма значений начала и конца интервала), то графическое изображение такого ряда называется *полигоном* (рис. 6), которое получается соединением прямыми точек с координатами  $X_i$  и  $f_i$ .

При изучении вариации применяются такие характеристики ряда распределения, которые описывают количественно его структуру, строение. Такова, например, *медиана* – величина варьирующего признака, делящая совокупность на две равные части – со значением признака меньше медианы и со значением признака больше медианы. В нашем примере про ВО (табл. 3) медиана – это 18 - й таможенный пост из 35 с величиной ВО 56,8 млн. долл. Из этого примера видно принципиальное различие между медианой и средней величиной: медиана не зависит от значений на краях ранжированного ряда. Даже если бы ВО 35 - го таможенного поста был в 10 раз больше, величина медианы не изменилась бы. Поэтому медиану часто используют как более надежный показатель типичного значения признака, нежели средняя арифметическая, если ряд значений неоднороден, включает резкие отклонения от средней. В интервальном ряду распределения для нахождения медианы применяется формула:

$$Me = X_0 + h \frac{0,5 \sum f - f'_{Me-1}}{f_{Me}}$$

где  $Me$  – медиана;

$X_0$  – нижняя граница интервала, в котором находится медиана;

$h$  – величина (размах) интервала;

$f'_{Me-1}$  – накопленная частота в интервале, предшествующем медианному;

$f_{Me}$  – частота в медианном интервале.

В табл. 4 медианным является среднее из 35 значений, то есть 18 - е от начала значение ВО. Как видно из столбца накопленных частот (6 - й столбец), оно находится в третьем интервале. Тогда по формуле:

$$Me = 53,16 + 14,5 \frac{0,5 * 35 - 12}{13} = 59,30 \quad (\text{млн. у.е.})$$

Аналогично медиане вычисляются значения признака, делящие совокупность на 4 равные по численности части – *квартили*, которые обозначаются заглавной латинской буквой  $Q$  с подписным значком номера квартиля. Ясно, что  $Q_2$  совпадает с  $Me$ . Для первого и третьего квартилей приводим формулы и расчет по данным табл. 4:

$$Q_1 = X_0 + h \frac{0,25 \sum f - f'_{Q_1-1}}{f_{Q_1}} = 38,66 + 14,5 \frac{0,25 * 35 - 5}{7} = 43,43 \quad (\text{млн. у.е.})$$

$$Q_3 = X_0 + h \frac{0,75 \sum f - f'_{Q_3-1}}{f_{Q_3}} = 67,66 + 14,5 \frac{0,75 * 35 - 25}{4} = 72,19 \quad (\text{млн. у.е.})$$

Так как  $Q_2 = Me = 59,30$  млн. у.е., видно, что различие между первым квартилем и медианой (–15,87) больше, чем между медианой и третьим квартилем (12,89). Этот факт свидетельствует о наличии некоторой несимметричности в средней области распределения, что заметно и на рис. 4.



Значения признака, делящие ряд на 5 равных частей, называются *квинтилями*, на 10 частей – *децилями*, на 100 частей – *перцентилями*. Эти характеристики применяются при необходимости подробного изучения структуры ряда распределения.

Важное значение имеет такая величина признака, которая встречается в изучаемом ряду распределения чаще всего. Такую величину принято называть *модой*. В дискретном ряду мода определяется без вычисления как значение признака с наибольшей частотой. Обычно встречаются ряды с одним модальным значением признака. Если в ряду распределения встречаются два или несколько равных (и даже несколько различных, но больших чем соседние) значений признака, то он считается соответственно бимодальным или мультимодальным. Это свидетельствует о неоднородности совокупности, возможно, представляющей собой агрегат нескольких совокупностей с разными модами. В интервальном ряду распределения интервал с наибольшей частотой является модальным. Внутри этого интервала находят условное значение признака, вблизи которого *плотность распределения* (число единиц совокупности, приходящихся на единицу измерения варьирующего признака) достигает максимума. Это условное значение и считается *точечной модой*. Логично предположить, что такая точечная мода располагается ближе к той из границ интервала, за которой частота в соседнем интервале больше частоты в интервале за другой границей модального интервала. Отсюда получаем обычно применяемую формулу:

$$M_o = X_0 + h \frac{f_{M_o} - f_{M_o-1}}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})},$$

где  $M_o$  – мода;

$X_0$  – нижнее значение модального интервала;

$f_{M_o}$  – частота в модальном интервале;

$f_{M_o-1}$  – частота в предыдущем интервале;

$f_{M_o+1}$  – частота в следующем интервале за модальным;

$h$  – величина интервала.

По данным табл. 4 рассчитаем точечную моду по формуле:

$$M_o = 53,16 + 14,5 \frac{13 - 7}{(13 - 7) + (13 - 4)} = 58,96 \quad (\text{млн. у.е.}).$$

К изучению структуры ряда распределения средняя арифметическая величина также имеет отношение, хотя основное значение этого обобщающего показателя другое. В интервальном ряду распределения ВО по таможенным постам средняя арифметическая рассчитывается как взвешенная по частоте середина интервалов  $X$  (расчет числителя – в 5-м столбце табл. 12) по формуле:

$$\bar{X} = \frac{\sum X \cdot f}{\sum f} = 2128,85/35 = 60,82 \quad (\text{млн. у.е.}).$$

Различие между средней арифметической величиной (60,82), медианой (59,30) и модой (58,96) в нашем примере невелико. Чем ближе распределение по форме к нормальному закону, тем ближе значения медианы, моды и средней величины между собой.

### Вопросы для самопроверки:

1. Вариационный ряд – это...
2. Что обозначает ранжированный ряд?
3. Что обозначает дискретный ряд.
4. Интервальный ряд обозначает...
5. Гистограмма – это...
6. Что обозначает медиана.
7. Что такое квинтиль.

8. Что такое перцентиль.
9. Что обозначает плотность распределения?
10. Точечная мода – это...

## **Тема 5. Методы выявления основной тенденции (тренда) в рядах динамики. Оценка адекватности тренда и прогнозирование**

Одна из основных задач изучения рядов динамики – выявить основную тенденцию (закономерность) в изменении уровней ряда, именуемую *трендом*. Закономерность в изменении уровней ряда в одних случаях проявляется наглядно, в других – может маскироваться колебаниями случайного или неслучайного характера. Поэтому, чтобы сделать правильные выводы о закономерностях развития того или иного показателя, надо суметь отделить тренд от колебаний, вызванных случайными кратковременными причинами. На основании выделенного тренда можно экстраполировать (прогнозировать) развитие явления в будущем. С этой целью (устранить колебания, вызванные случайными причинами) ряды динамики подвергают обработке.

Существует несколько методов обработки рядов динамики, помогающих выявить основную тенденцию изменения уровней ряда, а именно: метод укрупнения интервалов, метод скользящей средней и аналитическое выравнивание. Во всех методах вместо фактических уровней при обработке ряда рассчитываются иные (расчетные) уровни, в которых тем или иным способом взаимопогашается действие случайных факторов и тем самым уменьшается колеблемость уровней. Последние в результате становятся как бы «выравненными», «сглаженными» по отношению к исходным фактическим данным. Такие методы обработки рядов динамики называются *сглаживанием* или *выравниванием* рядов динамики.

Простейший метод сглаживания уровней ряда – *укрупнения интервалов*, для чего определяется итоговое значение или средняя величина исследуемого показателя. Этот метод особенно эффективен, если первоначальные уровни ряда относятся к коротким промежуткам времени. Например, если имеются данные о ежесуточном производстве мороженого на предприятии за месяц, то, естественно, в таком ряду возможны значительные колебания уровней, так как чем меньше период, за который приводятся данные, тем больше влияние случайных факторов. Чтобы устранить это влияние, рекомендуется укрупнить интервалы времени, например, до 5 или 10 дней, и для этих укрупненных интервалов рассчитать общий или среднесуточный объем производства (соответственно по пятидневкам или декадам). В ряду с укрупненными интервалами времени закономерность изменения уровней будет более наглядной. Или, например, имеются ежемесячные данные о производстве мороженого (табл. 5), еще более сильно укрупним интервалы – до трех месяцев (табл. 6).

По своей сути, метод *скользящей средней* похож на метод укрупнения интервалов, но в данном случае фактические уровни заменяются средними уровнями, рассчитанными для последовательно подвижных (скользящих) укрупненных интервалов, охватывающих  $m$  уровней ряда. Например, если принять  $m = 3$ , то сначала рассчитывается средняя величина из первых трех уровней, затем находится средняя величина из 2-го, 3-го и 4-го уровней, потом из 3-го, 4-го и 5-го и т.д., то есть каждый раз в сумме трех уровней появляется новый уровень, а два остаются прежними, что и обуславливает взаимопогашение случайных колебаний в средних уровнях. Рассчитанные из  $m$  членов скользящие средние относятся к середине (центру) каждого рассматриваемого интервала.

Сглаживание методом скользящей средней можно проводить по любому числу членов  $m$ , но удобнее, если  $m$  – нечетное число, так как в этом случае скользящая средняя сразу относится к конкретной временной точке – середине (центру) интервала. Если же  $m$  – четное, то скользящая средняя относится к промежутку между временными точками: например, при

сглаживании по четырем членам ( $m=4$ ) средняя из первых четырех уровней будет находиться между второй и третьей временной точкой, следующая – между третьей и четвертой и т.д. Тогда, чтобы сглаженные уровни относились непосредственно к конкретным временным точкам, из каждой пары смежных промежуточных значений скользящих средних находят среднюю арифметическую, которую относят к временной точке, находящейся между смежными. Такой прием двойного расчета сглаженных уровней называется *центрированием*.

Недостатком метода скользящей средней является то, что сглаженный ряд укорачивается по сравнению с фактическим с двух концов: при нечетном  $m$  на  $(m - 1)/2$ , а при четном  $m$  – на  $m/2$  с каждого конца. Применяя этот метод, надо помнить, что он сглаживает (устраняет) лишь случайные колебания. Если же, например, ряд содержит сезонную волну, она сохранится и после сглаживания методом скользящей средней. Кроме того, этот метод сглаживания, как и метод укрупнения интервалов не позволяет выражать общую тенденцию изменения уровней в виде математической модели.

Наиболее совершенным методом обработки рядов динамики в целях устранения случайных колебаний и выявления тренда является *выравнивание уровней ряда по аналитическим формулам* (или *аналитическое выравнивание*). Суть аналитического выравнивания заключается в замене эмпирических (фактических, исходных) уровней  $y_i$  теоретическими  $\hat{y}_t$ , которые рассчитаны по определенному уравнению, принятому за математическую модель тренда, где теоретические уровни рассматриваются как функция времени:  $\hat{y}_t = f(t)$ .

При этом каждый фактический уровень  $y_i$  рассматривается обычно как сумма двух составляющих:

$$y_i = f(t) + \varepsilon_t,$$

Где  $f(t) = \hat{y}_t$  - систематическая составляющая, отражающая тренд и выраженная определенным уравнением;  $\varepsilon_t$  - случайная величина, вызывающая колебания уровней вокруг тренда.

Задача аналитического выравнивания сводится к следующему:

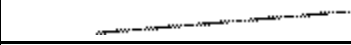

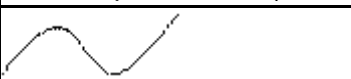

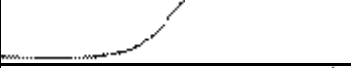
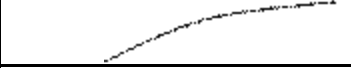

- 1) определение на основе фактических данных формы (вида) гипотетической функции  $\hat{y}_t = f(t)$ , способной наиболее адекватно отразить тенденцию развития исследуемого показателя;
- 2) нахождение по эмпирическим данным параметров указанной функции (уравнения);
- 3) расчет по найденному уравнению теоретических (выравненных) уровней.

В аналитическом выравнивании наиболее часто используются простейшие функции, представленные в табл. 5, где обозначено  $\hat{y}_t$  - теоретические (выравненные) уровни (читается как «игрек, выравненный по  $t$ »);  $t$  - условное обозначение времени (1, 2, 3 ...);  $a_0, a_1, a_2, \dots$  - параметры аналитической функции;  $k$  - число гармоник (при выравнивании по ряду Фурье).

Выбор той или иной функции для выравнивания ряда динамики осуществляется на основании графического изображения эмпирических данных. Если по тем или иным причинам уровни эмпирического ряда трудно описать одной функцией, следует разбить анализируемый период на отдельные части и затем выровнять каждую часть по соответствующей кривой.

Таблица 5.

Виды математических функций, используемые при выравнивании

| Название функции      | Вид функции   | Формула   |
|-----------------------|---|---|
| Прямая линия          |  | $\hat{y}_t = a_0 + a_1 t$   |
| Парабола 2-го порядка |  | $\hat{y}_t = a_0 + a_1 t + a_2 t^2$                                 |
| Парабола 3-го порядка |  | $\hat{y}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$                       |
| Гипербола             |  | $\hat{y}_t = a_0 + \frac{a_1}{t} \quad (94)$                        |
| Показательная         |  | $\hat{y}_t = a_0 a_1^t \quad (95)$                                  |
| Степенная             |  | $\hat{y}_t = a_0 t^{a_1} \quad (96)$                                |
| Ряд Фурье             |  | $\hat{y}_t = a_0 + \sum_{k=1}^{\infty} (a_k \cos kt + b_k \sin kt)$ |

Нередко один и тот же ряд можно выровнять по разным аналитическим функциям и получить довольно близкие результаты. В нашем примере про ВО России можно произвести выравнивание и по прямой линии, и по параболе. Чтобы решить вопрос о том, использование какой кривой дает лучший результат, обычно сопоставляют суммы квадратов отклонений эмпирических уровней от теоретических (*остатки*), рассчитанным по разным функциям, то есть:

$$\sum (\hat{y}_t - y)^2$$

Та функция, при которой эта сумма минимальна, считается наиболее адекватной, приемлемой. Однако сравнивать непосредственно суммы квадратов отклонений можно в том случае, если сравниваемые уравнения имеют одинаковое число параметров. Если же число параметров  $k$  разное, то каждую сумму квадратов делят на разность  $(n - k)$ , выступающую в роли числа степеней свободы, и сравнивают уже квадраты отклонений уровней, рассчитанные на одну степень свободы (то есть, остаточные дисперсии на одну степень свободы).

Параметры искоемых уравнений ( $a_0, a_1, a_2, \dots$ ) при аналитическом выравнивании могут быть определены по-разному, но наиболее распространенным методом является *метод наименьших квадратов* (МНК). При этом методе учитываются все эмпирические уровни и должна обеспечиваться минимальная сумма квадратов отклонений эмпирических значений уровней  $y$  от теоретических уровней  $\hat{y}_t$ :

$$\sum (\hat{y}_t - y)^2 \rightarrow \min$$

В частности, при выравнивании по прямой вида параметры  $a_0$  и  $a_1$  отыскиваются по МНК следующим образом. В этой формуле вместо  $\hat{y}_t$  записываем его конкретное выражение  $a_0 + a_1 t$ . Тогда  $S = \sum (a_0 + a_1 t - y)^2 \rightarrow \min$ . Дальнейшее решение сводится к задаче на экстремум, то есть к определению того, при каком значении  $a_0$  и  $a_1$  функция двух переменных  $S$  может достигнуть минимума. Как известно, для этого надо найти частные производные  $S$  по  $a_0$  и  $a_1$ ,

приравнять их к нулю и после элементарных преобразований решить систему двух уравнений с двумя неизвестными.

В соответствии с вышеизложенным, найдем частные производные:

$$\begin{cases} \frac{\partial S}{\partial a_0} = 2 \sum (a_0 + a_1 t - y) = 0 \\ \frac{\partial S}{\partial a_1} = 2 \sum (a_0 + a_1 t - y)t = 0 \end{cases}$$

Сократив каждое уравнение на 2, раскрыв скобки и перенеся члены с  $y$  в правую сторону, а остальные – оставив в левой, получим систему нормальных уравнений:

$$\begin{cases} na_0 + a_1 \sum t = \sum y \\ a_0 \sum t + a_1 \sum t^2 = \sum yt \end{cases}$$

где  $n$  – количество уровней ряда;  $t$  – порядковый номер в условном обозначении периода или момента времени;  $y$  – уровни эмпирического ряда.

Эта система и, соответственно, расчет параметров  $a_0$  и  $a_1$  упрощаются, если отсчет времени ведется от середины ряда. Например, при *нечетном* числе уровней (как в нашем примере про ВО России – 7 уровней) срединная точка времени (год, месяц) принимается за нуль, тогда предшествующие периоды обозначаются соответственно  $-1, -2, -3$  и т.д., а следующие за средним (центральным) – соответственно  $1, 2, 3$  и т.д. При *четном* числе уровней два срединных момента (периода) времени обозначают  $-1$  и  $+1$ , а все последующие и предыдущие, соответственно, через два интервала:  $\pm 3, \pm 5, \pm 7$  и т.д.

При таком порядке отсчета времени (от середины ряда)  $\sum t = 0$ , поэтому, система нормальных уравнений упрощается до следующих двух уравнений, каждое из которых решается самостоятельно:

$$\begin{cases} na_0 = \sum y \Rightarrow a_0 = \frac{\sum y}{n} \\ a_1 \sum t^2 = \sum yt \Rightarrow a_1 = \frac{\sum yt}{\sum t^2} \end{cases}$$

Как видим, при такой нумерации периодов параметр  $a_0$  представляет собой средний уровень равномерного интервального ряда. Определим по этой формуле параметры уравнения прямой для нашего примера про ВО России, для чего исходные данные и все расчеты необходимых сумм представим в табл. 6.

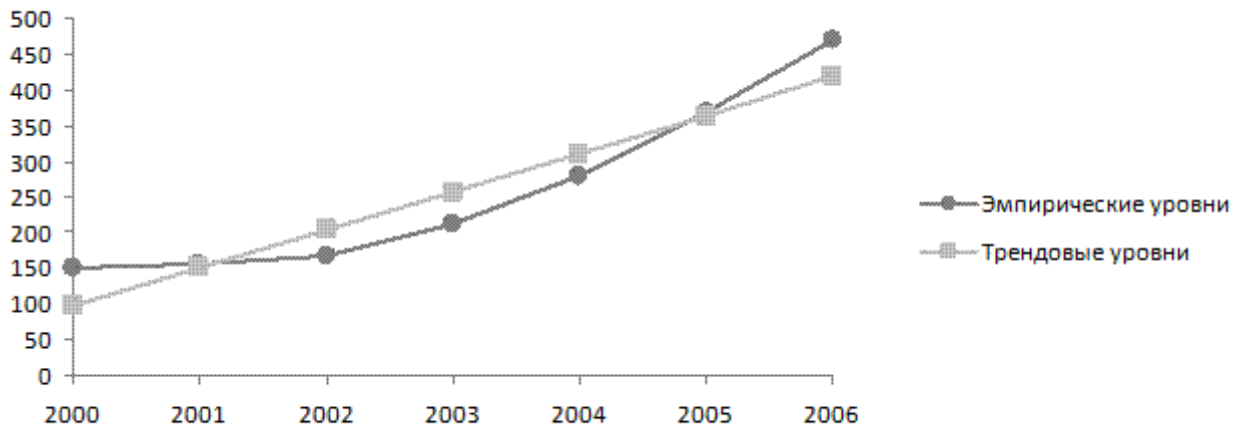
Таблица 6.

**Вспомогательные расчеты для линейного тренда**

| Год   | $y$    | $t$ | $t^2$ | $yt$   | $\hat{y}_t$ | $(\hat{y}_t - y)^2$ | $(\hat{y}_t - \bar{y})^2$ | $(y - \bar{y})^2$ |
|-------|--------|-----|-------|--------|-------------|---------------------|---------------------------|-------------------|
| 2000  | 149,9  | -3  | 9     | -449,7 | 97,557      | 2739,775            | 25636,584                 | 11614,681         |
| 2001  | 155,6  | -2  | 4     | -311,2 | 150,929     | 21,822              | 11394,038                 | 10418,577         |
| 2002  | 168,3  | -1  | 1     | -168,3 | 204,300     | 1296,000            | 2848,509                  | 7987,252          |
| 2003  | 212    | 0   | 0     | 0      | 257,671     | 2085,879            | 0,000                     | 2085,879          |
| 2004  | 280,6  | 1   | 1     | 280,6  | 311,043     | 926,768             | 2848,509                  | 525,719           |
| 2005  | 368,9  | 2   | 4     | 737,8  | 364,414     | 20,122              | 11394,038                 | 12371,795         |
| 2006  | 468,4  | 3   | 9     | 1405,2 | 417,786     | 2561,806            | 25636,584                 | 44406,531         |
| Итого | 1803,7 | 0   | 28    | 1494,4 | 1803,700    | 9652,171            | 79758,263                 | 89410,434         |

Из табл. 6 получаем, что:  $a_0 = 1803,7/7 = 257,671$  и  $a_1 = 1494,4/28 = 53,371$ . Отсюда искомое уравнение тренда:  $\hat{y}_t = 257,671 + 53,371t$ . В 6-м столбце табл. 6 приведены теоретические (трендовые) уровни, рассчитанные по этому уравнению, а в итоге 7 - го столбца

– остатки по формуле. Для иллюстрации построим график эмпирических и трендовых уровней (рис. 7).



**Рис. 7. Эмпирические и трендовые уровни ряда динамики ВО России**

Для найденного уравнения тренда необходимо провести оценку его *надежности (адекватности)*, что осуществляется обычно с помощью критерия Фишера, сравнивая его расчетное значение  $F_p$  с теоретическим (табличным) значением  $FT$ . При этом расчетный критерий Фишера определяется по формуле (102):

$$F_p = \frac{(n-k) \sum (\hat{y}_t - \bar{y})^2}{(k-1) \sum (\hat{y}_t - y)^2}$$

где  $k$  – число параметров (членов) выбранного уравнения тренда.

Для проверки правильности расчета сумм в формуле (102) можно использовать следующее равенство (103):

$$\sum (y - \bar{y})^2 = \sum (\hat{y}_t - y)^2 + \sum (\hat{y}_t - \bar{y})^2$$

В нашем примере про ВО указанное равенство соблюдается (необходимые суммы рассчитаны в трех последних столбцах табл. 6):  $89410,434 = 9652,171 + 79758,263$ .

Сравнение расчетного и теоретического значений критерия Фишера ведется при заданном уровне значимости с учетом степеней свободы:  $\nu_1 = k - 1$  и  $\nu_2 = n - k$ . При условии  $F_p > FT$  считается, что выбранная математическая модель ряда динамики адекватно отражает обнаруженный в нем тренд.

Проверим тренд на адекватность в нашем примере про ВО по формуле:

$$F_p = 79758,263 * 5 / (9652,171 * 1) = 41,32 > FT,$$

значит, модель адекватна и ее можно использовать для прогнозирования ( $FT = 6,61$  находим по Приложению 4 в 1-ом столбце [ $\nu_1 = k - 1 = 2 - 1 = 1$ ] и 5-й строке [ $\nu_2 = n - k = 5$ ]).

Как уже было отмечено ранее, в нашем примере про ВО России можно произвести выравнивание не только по прямой линии, но и по параболе, чего делать не будем, так как уже найденный линейный тренд адекватно описывает тенденцию.

При составлении прогнозов уровней социально-экономических явлений обычно оперируют не точечной, а интервальной оценкой, рассчитывая так называемые *доверительные интервалы прогноза*. Границы интервалов определяются по формуле:

$$\hat{y}_t \pm t_{\alpha} \sigma_y$$

где  $\hat{y}_t$  – точечный прогноз, рассчитанный по модели тренда;  $t_{\alpha}$  – коэффициент доверия по распределению Стьюдента при уровне значимости  $\alpha$  и числе степеней свободы  $\nu = n - 1$ ;  $\sigma_y$  – ошибка аппроксимации, определяемая по формуле:

$$\sigma_y = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - k}}$$

Спрогнозируем ВО России на 2007 и 2008 годы с вероятностью 0,95 (значимостью 0,05), для чего найдем ошибку аппроксимации по формуле (105):  $\sigma_y = \sqrt{9652,171/(7-2)} = 43,937$  и найдем коэффициент доверия по распределению Стьюдента по Приложению 2:  $t_{\alpha} = 2,4469$  при  $\nu = 7 - 1 = 6$ .

Прогноз на 2007 и 2008 годы с вероятностью 0,95 по формуле:

$$Y_{2007} = (257,671 + 53,371 * 4) \pm 2,4469 * 43,937 \text{ или } 363,6 < Y_{2007} < 578,7 \text{ (млрд. у.е.);}$$

$$Y_{2008} = (257,671 + 53,371 * 5) \pm 2,4469 * 43,937 \text{ или } 417,0 < Y_{2008} < 632,0 \text{ (млрд. у.е.).}$$

Как видно из полученных прогнозов, доверительный интервал достаточно широк (из-за достаточно большой величины ошибки аппроксимации). Более точный прогноз можно получить при выравнивании по параболе 2-го порядка.

#### Вопросы для самопроверки:

1. Методы обработки рядов динамики называются...
2. Метод сглаживания уровней ряда обозначает...
3. В чем смысл метода скользящей средней?
4. Выравнивание уровней ряда по аналитическим формулам (или аналитическое выравнивание) необходимо для того, чтобы...
5. Задачи аналитического выравнивания позволяют...
6. Что обозначают доверительные интервалы прогноза?
7. Что обозначает ошибка аппроксимации?
8. Метод наименьших квадратов (МНК) – это...

### Тема 6. Расчет показателей размера и интенсивности вариации. Расчет моментов распределения и показателей его формы

Простейшим показателем является *размах вариации* – абсолютная разность между максимальным и минимальным значениями признака из имеющихся в изучаемой совокупности значений:

$$H = X_{\max} - X_{\min}$$

Поскольку величина размаха характеризует лишь максимальное различие значений признака, она не может измерять закономерную силу его вариации во всей совокупности. Предназначенный для данной цели показатель должен учитывать и обобщать все различия значений признака в совокупности без исключения. Число таких различий равно числу сочетаний по два из всех единиц совокупности (в нашем примере про ВО число сочетаний

$$C_n^k = \frac{n!}{(n-k)!k!} = \frac{35!}{(35-2)!2!} = 595$$

составит ). Однако нет необходимости рассматривать, вычислять и осреднять все отклонения. Проще использовать среднюю из отклонений отдельных значений признака от среднего арифметического значения признака, а таковых в нашем

примере про ВО всего 35. Но среднее отклонение значений признака от средней арифметической величины согласно первому свойству последней равно нулю. Поэтому показателем силы вариации выступает не арифметическая средняя отклонений, а средний модуль отклонений, или *среднее линейное отклонение*:

$$L = \frac{\sum |X - \bar{X}|}{N}$$

В нашем примере про ВО по данным табл. 12 среднее линейное отклонение вычисляется как взвешенное по частоте отклонение по модулю середин интервалов от средней арифметической величины (расчет числителя произведен в 7-м столбце табл. 12), то есть по формуле:

$$L = \frac{\sum |X' - \bar{X}|f}{\sum f} = 513,714 / 35 = 14,678 \quad (\text{млн. у.е.})$$

Это означает, что в среднем величина ВО в изучаемой совокупности таможенных постов отклонялась от средней величины ВО в РФ на 14,678 млн. долл.

Простота расчета и интерпретации составляют положительные стороны показателя  $L$ , однако математические свойства модулей «плохие»: их нельзя поставить в соответствие с каким-либо вероятностным законом, в том числе и с нормальным распределением, параметром которого является не средний модуль отклонений, а *среднее квадратическое отклонение*, обозначаемое малой греческой буквой сигма ( $\sigma$ ) или  $s$  и вычисляемое по формуле – для ранжированного ряда и по формуле – для интервального ряда:

$$\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{N}};$$

$$\sigma = \sqrt{\frac{\sum (X' - \bar{X})^2 f}{\sum f}}$$

В примере про ВО по данным табл. 4 среднее квадратическое отклонение величины ВО по формуле составило:

$$\sigma = \sqrt{\frac{13660,243}{35}} = \sqrt{390,293} = 19,756 \quad (\text{млн. у.е.})$$

Среднее квадратическое отклонение по величине в реальных совокупностях всегда больше среднего модуля отклонений. Разница между ними тем больше, чем больше в изучаемой совокупности резких, выделяющихся отклонений, что служит индикатором «засоренности» совокупности неоднородными с основной массой элементами. Для нормального закона распределения отношение  $\sigma/L \approx 1,25$ . В нашем примере про ВО:  $\sigma/L \approx 19,756/14,678 = 1,35 > 1,25$ , то есть в изучаемой совокупности наблюдаются некоторое число таможенных постов с отличающимися от основной массы величинами ВО.

Квадрат среднего квадратического отклонения представляет собой *дисперсию* отклонений, на использовании которой основаны практически все методы математической статистики (простая дисперсия) и (взвешенная дисперсия):

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

$$\sigma^2 = \frac{\sum (X' - \bar{X})^2 f}{\sum f} = \overline{X^2} - \bar{X}^2$$



Еще одним показателем силы вариации, характеризующим ее не по всей совокупности, а лишь в ее центральной части, служит *среднее квартильное расстояние (отклонение)*, то есть средняя величина разности между квартилями, определяемая по формуле:

$$q = \frac{(Q_3 - Q_2) + (Q_2 - Q_1)}{2} = \frac{Q_3 - Q_1}{2}$$

В нашем примере про ВО по формуле:

$$q = \frac{72,19 - 43,43}{2} = 14,38 \quad (\text{млн. у.е.}).$$

Сила вариации в центральной части совокупности, как правило, меньше, чем в целом по всей совокупности. Соотношение между средним линейным отклонением и средним квартильным расстоянием служит для изучения структуры вариации: большое значение такого соотношения свидетельствует о наличии слабо варьирующего «ядра» и сильно рассеянного вокруг него окружения в изучаемой совокупности. Для нашего примера про ВО соотношение  $L/q = 1,021$ , что говорит о совсем незначительном различии силы вариации в центральной части совокупности и на ее периферии.

Для оценки интенсивности вариации и для сравнения ее в разных совокупностях и тем более для разных признаков необходимы *относительные показатели вариации*, которые вычисляются как отношение абсолютных показателей силы вариации, рассмотренных ранее, к средней арифметической величине признака:

– относительный размах вариации:  $\rho = \frac{H}{\bar{X}}$ ;

– линейный коэффициент вариации:  $\lambda = \frac{L}{\bar{X}}$ ;

– квадратический коэффициент вариации:  $v = \frac{\sigma}{\bar{X}}$ ;

– относительное квартильное расстояние:  $d = \frac{q}{\bar{X}}$ .

В нашем примере про ВО эти показатели составляют:

$$\rho = 87/60,82 = 1,43, \text{ или } 143\%; \quad \lambda = 14,678/60,82 = 0,241, \text{ или } 24,1\%;$$

$$v = 19,756/60,82 = 0,32, \text{ или } 32\%; \quad d = 14,38/60,82 = 0,236, \text{ или } 23,6\%.$$

Оценка степени интенсивности вариации возможна только для каждого отдельного признака и совокупности определенного состава, она состоит в сравнении наблюдаемой вариации с некоторой обычной ее интенсивностью, принимаемой за норматив. Так, для совокупности таможенных постов вариация величины ВО может быть определена как слабая, если  $v < 25\%$ , умеренная при  $25\% < v < 50\%$  и сильная при  $v > 50\%$ .

Различная сила, интенсивность вариации обусловлены объективными причинами, поэтому нельзя говорить о каком-либо универсальном критерии вариации (например, 33%), так как для разных явлений и признаков этот критерий различен.

Для дальнейшего изучения характера вариации используются средние значения разных степеней отклонений отдельных величин признака от его средней арифметической величины. Эти показатели называются *центральные моменты распределения* порядка, соответствующего степени, в которую возводятся отклонения (табл. 7) или просто моментов (нецентральные моменты в таможенной статистике практически не используются).

Таблица 7.

Центральные моменты

| Порядок момента   | Формула  |   |
|-------------------|--|---|
|                   | по несгруппированным данным  | по сгруппированным данным   |
| Первый $\mu_1$    | $\frac{\sum(X - \bar{X})}{N} = 0$                                  | $\frac{\sum(X - \bar{X})f}{\sum f} = 0$                                   |
| Второй $\mu_2$    | $\frac{\sum(X - \bar{X})^2}{N} = \bar{X}^2 - \bar{X}^2 = \sigma^2$ | $\frac{\sum(X - \bar{X})^2 f}{\sum f} = \bar{X}^2 - \bar{X}^2 = \sigma^2$ |
| Третий $\mu_3$    | $\frac{\sum(X - \bar{X})^3}{N}$                                    | $\frac{\sum(X - \bar{X})^3 f}{\sum f}$                                    |
| Четвертый $\mu_4$ | $\frac{\sum(X - \bar{X})^4}{N}$                                    | $\frac{\sum(X - \bar{X})^4 f}{\sum f}$                                    |

Величина третьего момента  $\mu_3$  зависит, как и его знак, от преобладания положительных кубов отклонений над отрицательными кубами либо наоборот. При нормальном и любом другом строго симметричном распределении сумма положительных кубов строго равна сумме отрицательных кубов, поэтому на основе третьего момента строится показатель, характеризующий степень асимметричности распределения – *коэффициент асимметрии*:

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

В нашем примере про ВО показатель асимметрии по формуле составил (расчет числителя произведен в 9 - м столбце табл. 6):

$$As = \frac{114110,66}{19,756^3 * 35} = 0,423 > 0, \text{ то есть асимметрия значительна.}$$

Английский статистик К. Пирсон на основе разности между средней арифметической величиной и модой предложил другой показатель асимметрии:

$$As_{\text{П}} = \frac{\bar{X} - Mo}{\sigma}$$

В нашем примере по данным табл. 12 показатель асимметрии по формуле

$$As = \frac{60,82 - 58,96}{19,756} = 0,09.$$

составил:

Показатель асимметрии Пирсона зависит от степени асимметричности в средней части ряда распределения, а показатель асимметрии – от крайних значений признака. Таким образом, в нашем примере про ВО в средней части распределения наблюдается меньшая асимметрия, чем по краям, что видно и по графику (рис. 8). Распределения с сильной правосторонней и левосторонней асимметрией показаны на рис. 9.

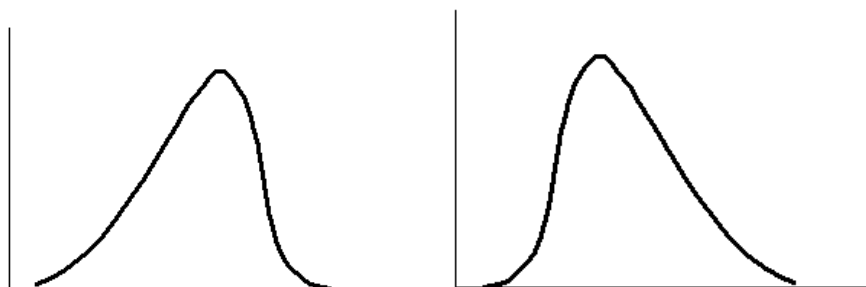


Рис. 8. Асимметрия распределения

С помощью момента четвертого порядка характеризуется еще более сложное свойство рядов распределения – *эксцесс* (от англ. «излишество»). Показатель эксцесса рассчитывается по формуле:

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

Чаще всего эксцесс интерпретируется как «крутизна» распределения, что не совсем верно. График распределения может выглядеть сколь угодно крутым в зависимости от силы вариации признака: чем слабее вариация, тем круче кривая распределения при данном масштабе. Не говоря уже о том, что, изменяя масштабы по осям абсцисс и ординат, любое распределение можно искусственно сделать «крутым» и «пологим». Чтобы показать, в чем состоит эксцесс распределения, и правильно его интерпретировать, нужно сравнить ряды с одинаковой силой вариации (одной и той же величиной  $\sigma$ ) и разными показателями эксцесса. Чтобы не смешать эксцесс с асимметрией, все сравниваемые ряды должны быть симметричными. Такое сравнение изображено на рис. 9.

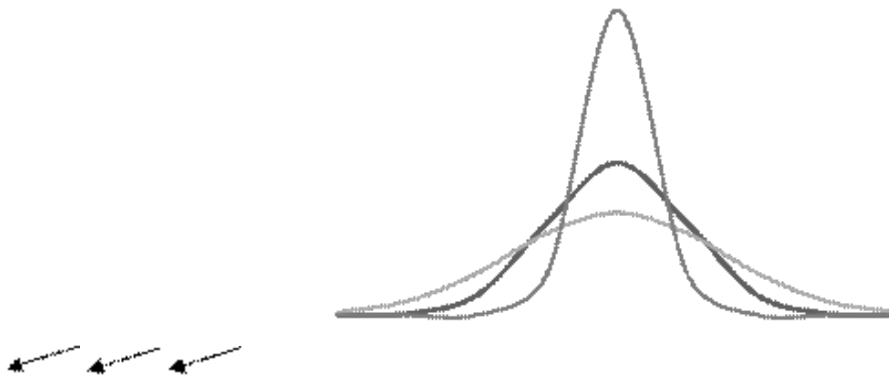


Рис. 9. Эксцесс распределения

Наличие положительного эксцесса означает наличие слабо варьирующего «ядра» и сильно рассеянного вокруг него окружения в изучаемой совокупности. Отрицательный эксцесс означает отсутствие такого «ядра».

В нашем примере по формуле эксцесс составил (расчет числителя произведен в 10-м столбце табл. 7):  $Ex = \frac{13809838,86}{19,756^2 * 35} - 3 = -0,41$ , то есть величина ВО по таможенным постам варьирует сильнее, чем при нормальном распределении.

По значениям показателей асимметрии и эксцесса распределения можно судить о близости распределения к нормальному: показатели асимметрии и эксцесса не должны превышать своих двукратных средних квадратических отклонений, то есть  $|As| < 2\sigma_{As}$  и  $|Ex| < 2\sigma_{Ex}$ . Эти средние квадратические отклонения вычисляются по формулам (39) и (40):

$$\sigma_{As} = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}};$$

$$\sigma_{Ex} = \sqrt{\frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}}.$$

В нашем примере по формулам:

$$\sigma_{As} = \sqrt{\frac{6 * 35 * (35 - 1)}{(35 - 2)(35 + 1)(35 + 3)}} = 0,40 ; \quad \sigma_{Ex} = \sqrt{\frac{24 * 35(35 - 1)^2}{(35 - 3)(35 - 2)(35 + 3)(35 + 5)}} = 0,78$$

Так как показатели асимметрии и эксцесса не превышают своих двукратных средних квадратических отклонений ( $As = |0,423| < 0,4*2$ ;  $Ex = |-0,41| < 0,78*2$ ), можно говорить о сходстве анализируемого распределения с нормальным.

### Вопросы для самопроверки:

1. Что позволяет понять размах вариации?
2. Что обозначает среднее линейное отклонение?
3. Что позволяет понять среднее квадратическое отклонение?
4. Дисперсия отклонений обозначает...
5. Квартильное расстояние (отклонение) – это...
6. Относительные показатели вариации обозначают...
7. Что обозначают центральные моменты распределения?
8. Что отражает в теории статистики эксцесс?

## Тема 7. Проверка соответствия ряда распределения нормальному. Проверка соответствия ряда распределения закону Пуассона

Под теоретической кривой распределения понимается графическое изображение ряда в виде непрерывной линии изменения частот в вариационном ряду, функционально связанного с изменением вариантов, другими словами, теоретическое распределение может быть выражено аналитически – формулой, которая связывает частоты и соответствующие значения признака. Такие алгебраические формулы носят название *законов распределения*. Большое познавательное значение имеет сопоставление фактических кривых распределения с теоретическими.

Как уже неоднократно отмечалось, часто пользуются типом распределения, которое называется *нормальным*. Формула функции плотности нормального распределения имеет следующий вид:

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-\bar{X})^2}{2\sigma^2}} \quad \text{или} \quad \varphi(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

Где  $X$  – значение изучаемого признака;

$\bar{X}$  – средняя арифметическая ряда;

$\Sigma$  – среднее квадратическое отклонение;

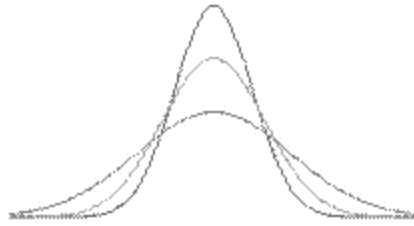
$$t = \frac{X - \bar{X}}{\sigma} \quad \text{– нормированное отклонение;}$$

$\pi = 3,1415$  – постоянное число (отношение длины окружности к ее диаметру);

$e = 2,7182$  – основание натурального логарифма.

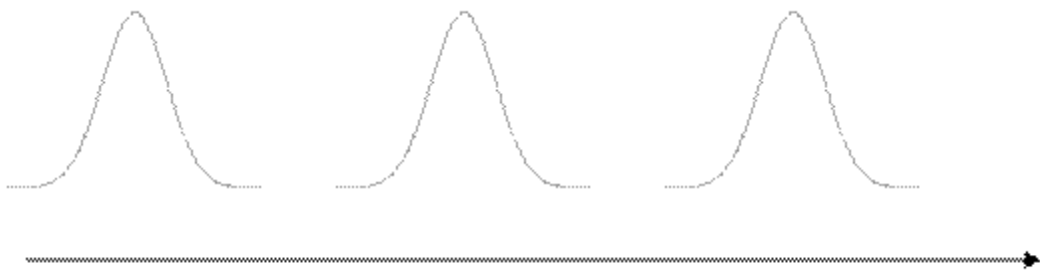
Следовательно, кривая нормального распределения может быть построена по двум параметрам – средней арифметической и среднему квадратическому отклонению. Поэтому важно выяснить, как эти параметры влияют на вид нормальной кривой.

Если  $\bar{X}$  не меняется, а изменяется только  $\sigma$ , то чем меньше  $\sigma$ , тем более вытянута вверх кривая и наоборот, чем больше  $\sigma$ , тем более плоской и растянутой вдоль оси абсцисс становится кривая нормального распределения (рис. 10).



**Рис. 10. Влияние величины  $\sigma$  на кривую нормального распределения**

Если  $\sigma$  остается неизменной, а  $\bar{X}$  изменяется, то кривые нормального распределения имеют одинаковую форму, но отличаются друг от друга положением максимальной ординаты (вершины) (рис. 11).



**Рис. 11. Влияние величины  $\bar{X}$  на кривую нормального распределения**

Итак, выделим *особенности кривой нормального распределения*:

- 1) кривая симметрична и имеет максимум в точке, соответствующей значению  $\bar{X} = Me = Mo$ ;
- 2) кривая асимптотически приближается к оси абсцисс, продолжаясь в обе стороны до бесконечности (чем больше отдельные значения  $X$  отклоняются от  $\bar{X}$ , тем реже они встречаются);
- 3) кривая имеет две точки перегиба на расстоянии  $\pm \sigma$  от  $\bar{X}$ ;
- 4) коэффициенты асимметрии и эксцесса равны нулю.

Гипотезы о распределениях заключаются в том, что выдвигается предположение о том, что распределение в изучаемой совокупности подчиняется какому - то определенному закону. Проверка гипотезы состоит в том, чтобы на основании сравнения фактических (эмпирических) частот с предполагаемыми (теоретическими) частотами сделать вывод о соответствии фактического распределения гипотетическому распределению.

Под гипотетическим распределением необязательно понимается нормальное распределение. Может быть выдвинута гипотеза о логнормальном, биномиальном распределениях, распределении Пуассона и пр. Причина частого обращения к нормальному распределению состоит в том, что, как уже было замечено ранее, в этом типе распределения выражается закономерность, возникающая при взаимодействии множества случайных причин, когда ни одна из не имеет преобладающего влияния.

В нашем примере про ВО близость значений средней арифметической величины (60,82), медианы (59,30) и моды (58,96) указывает на вероятное соответствие изучаемого распределения нормальному закону.

Проверка гипотезы о соответствии теоретическому распределению предполагает расчет теоретических частот этого распределения.

Для нормального распределения порядок расчета этих частот следующий:

1) по эмпирическим данным рассчитывают среднюю арифметическую ряда  $\bar{X}$  и среднее квадратическое отклонение  $\sigma$ ;

2) находят нормированное (выраженное в  $\sigma$ ) отклонение каждого эмпирического значения от средней арифметической:

$$t = \frac{X - \bar{X}}{\sigma};$$

3) по формуле (41) или с помощью таблиц интеграла вероятностей Лапласа находят значение  $\varphi(t)$ ;

4) вычисляют теоретические частоты  $m_i$  по формуле:

$$m_i = Nh_i\varphi(t),$$

Где  $N$  – объем совокупности,  $h_i$  – длина (размах)  $i$ -го интервала.

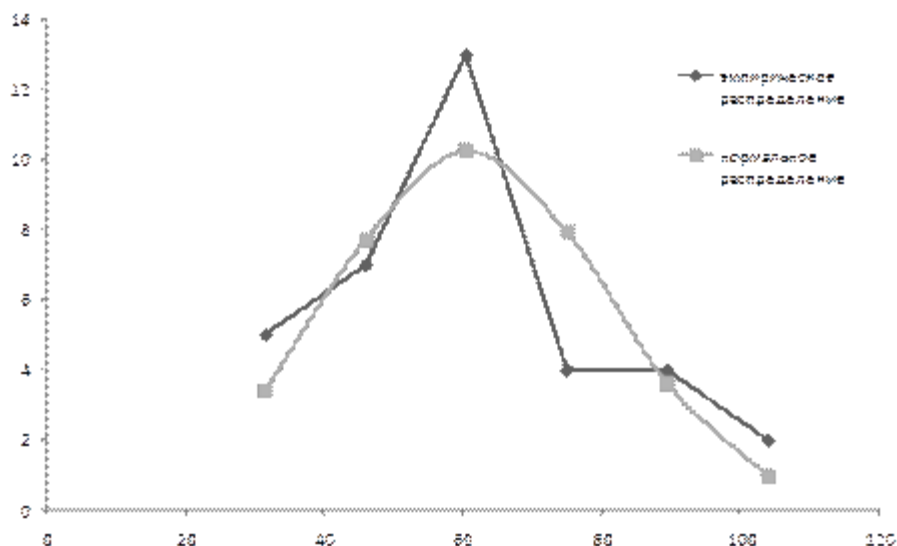
Определим теоретические частоты нормального распределения в нашем примере про ВО по данным табл. 6, для чего построим вспомогательную табл. 8. Средняя арифметическая величина и среднее квадратическое отклонение нами уже найдены ранее ( $\bar{X} = 60,82$ ;  $\sigma = 19,756$ ); значения нормированных отклонений  $t$  рассчитаны в 5-м столбце табл. 8, а значения плотностей  $\varphi(t)$  – в 8-м столбце (в 6-м и 7-м столбцах приведены промежуточные расчеты); в последнем столбце – теоретические частоты нормального распределения.

**Таблица 8.**

**Расчет теоретических частот нормального распределения**

| $i$ | $X_i$          | $f_i$ | $X_i'$ | $t = \frac{X_i - \bar{X}}{\sigma}$ | $-\frac{t^2}{2}$ | $e^{-\frac{t^2}{2}}$ | $\varphi(t)$ | $m_i$  |
|-----|----------------|-------|--------|------------------------------------|------------------|----------------------|--------------|--------|
| 1   | 24,16 – 38,66  | 5     | 31,41  | 1,4889                             | -1,1084          | 0,3301               | 0,0067       | 3,383  |
| 2   | 38,66 – 53,16  | 7     | 45,91  | 0,7549                             | -0,2850          | 0,7520               | 0,0152       | 7,707  |
| 3   | 53,16 – 67,66  | 13    | 60,41  | 0,0210                             | -0,0002          | 0,9998               | 0,0202       | 10,246 |
| 4   | 67,66 – 82,16  | 4     | 74,91  | 0,7130                             | -0,2542          | 0,7756               | 0,0157       | 7,948  |
| 5   | 82,16 – 96,66  | 4     | 89,41  | 1,4470                             | -1,0468          | 0,3510               | 0,0071       | 3,598  |
| 6   | 96,66 – 111,16 | 2     | 103,91 | 2,1809                             | -2,3782          | 0,0927               | 0,0019       | 0,950  |
|     | Итого          | 35    |        |                                    |                  |                      |              | 33,832 |

Сравним на графике эмпирические  $f$  (ВО по таможенным постам) и теоретические  $m$  (нормальное распределение) частоты, полученные на основе данных табл. 8 (рис. 12). Близость этих частот очевидна, но объективная оценка их соответствия может быть получена только с помощью критериев согласия.



**Рис. 12. Распределение ВО по таможенным постам (эмпирическое) и нормальное**

Критерии согласия, опираясь на установленный закон распределения, дают возможность установить, когда расхождения между теоретическими и эмпирическими частотами следует признать несущественными (случайными), а когда – существенными (неслучайными). Таким образом, критерии согласия позволяют отвергнуть или подтвердить правильность выдвинутой гипотезы о характере распределения в эмпирическом ряду и дать ответ, можно ли принять для данного эмпирического распределения модель, выраженную некоторым теоретическим законом распределения.

Существует ряд критериев согласия, но чаще всего применяют критерии Пирсона  $\chi^2$ , Колмогорова и Романовского.

*Критерий согласия Пирсона  $\chi^2$  (хи - квадрат)* – один из основных критериев согласия, рассчитываемый по формуле (44):

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - m_i)^2}{m_i}$$

где  $k$  – число интервалов;

$f_i$  – эмпирическая частота  $i$ -го интервала;

$m_i$  – теоретическая частота.

Для распределения  $\chi^2$  составлены таблицы, где указано критическое значение критерия согласия  $\chi^2$  для выбранного уровня значимости  $\alpha$  и данного числа степеней свободы  $\nu$ .

*Уровень значимости  $\alpha$*  – это вероятность ошибочного отклонения выдвинутой гипотезы, то есть вероятность ( $P$ ) того, что будет отвергнута правильная гипотеза. В статистических исследованиях в зависимости от важности и ответственности решаемых задач пользуются следующими тремя уровнями значимости:

1)  $\alpha = 0,10$ , тогда  $P = 0,90$ ;

2)  $\alpha = 0,05$ , тогда  $P = 0,95$ ;

3)  $\alpha = 0,01$ , тогда  $P = 0,99$ .

*Число степеней свободы  $\nu$*  определяется по формуле:

$$\nu = k - z - 1,$$

где  $k$  – число интервалов;

$z$  – число параметров, задающих теоретический закон распределения.

Для нормального распределения  $z = 2$ , так как нормальное распределение зависит от двух параметров – средней арифметической ( $\bar{X}$ ) и среднего квадратического отклонения ( $\sigma$ ).

Для оценки существенности расхождений расчетное значение  $\chi^2$  сравнивают с табличным  $\chi^2$  табл. Расчетное значения критерия должно быть меньше табличного, то есть  $\chi^2 < \chi^2_{\text{табл}}$ , в противном случае расхождения между теоретическим и эмпирическим распределением не случайны, а теоретическое распределение не может служить моделью для изучаемого эмпирического распределения.

Использование критерия  $\chi^2$  рекомендуется для достаточно больших совокупностей ( $N > 50$ ), при этом частота каждой группы не должна быть менее 5, в противном случае повышается вероятность получения ошибочных выводов.

В нашем примере про ВО для расчета критерия  $\chi^2$  построим вспомогательную табл. 9.

Таблица 9.

**Вспомогательные расчеты критериев согласия**

| $i$ | $X_i$          | $f_i$ | $m_i$  | $\frac{(f_i - m_i)^2}{m_i}$ | $f_i'$ | $m_i'$ | $ f_i' - m_i' $ |
|-----|----------------|-------|--------|-----------------------------|--------|--------|-----------------|
| 1   | 24,16 – 38,66  | 5     | 3,383  | 0,773                       | 5      | 3,383  | 1,617           |
| 2   | 38,66 – 53,16  | 7     | 7,707  | 0,065                       | 12     | 11,090 | 0,910           |
| 3   | 53,16 – 67,66  | 13    | 10,246 | 0,740                       | 25     | 21,336 | 3,664           |
| 4   | 67,66 – 82,16  | 4     | 7,948  | 1,961                       | 29     | 29,284 | 0,284           |
| 5   | 82,16 – 96,66  | 4     | 3,598  | 0,045                       | 33     | 32,882 | 0,118           |
| 6   | 96,66 – 111,16 | 2     | 0,950  | 1,160                       | 35     | 33,832 | 1,168           |
|     | Итого          | 35    | 33,832 | 4,744                       |        |        |                 |

Теперь по формуле:  $\chi^2 = 4,744$ , что меньше табличного (значения  $\chi^2$  табл. = 7,8147 при уровне значимости  $\alpha = 0,05$  и числе степеней свободы  $\nu = 6 - 2 - 1 = 3$ , значит с вероятностью 0,95 можно говорить, что в основе эмпирического распределения величины ВО по таможенным постам лежит закон нормального распределения, то есть выдвинутая гипотеза не отвергается, а расхождения объясняются случайными факторами.

*Критерий Романовского КР* основан на использовании критерия Пирсона  $\chi^2$ , то есть уже найденных значений  $\chi^2$  и числа степеней свободы  $\nu$ , рассчитывается по формуле:

$$K_P = \frac{|\chi^2 - \nu|}{\sqrt{2\nu}}$$

Он используется в том случае, когда отсутствует таблица значений  $\chi^2$ . Если  $K_P < 3$ , то расхождения между теоретическим и эмпирическим распределением случайны, если  $K_P > 3$ , то не случайны, и теоретическое распределение не может служить моделью для изучаемого эмпирического распределения.

$$K_P = \frac{|4,744 - 3|}{\sqrt{2 \cdot 3}} = 0,712 < 3$$

В нашем примере про ВО по формуле:  $K_P = 0,712 < 3$ , что подтверждает несущественность расхождений между эмпирическими и теоретическими частотами.

*Критерий Колмогорова  $\lambda$*  основан на определении максимального расхождения между накопленными частотами эмпирического и теоретического распределений ( $D$ ), рассчитывается по формуле:

$$\lambda = D \cdot \sqrt{N}$$



Рассчитав значение  $\lambda$ , по таблице  $P(\lambda)$  определяют вероятность, с которой можно утверждать, что отклонения эмпирических частот от теоретических случайны. Вероятность  $P(\lambda)$  может изменяться от 0 до 1. При  $P(\lambda) = 1$  (то есть при  $\lambda < 0,3$ ) происходит полное совпадение частот, при  $P(\lambda) = 0$  – полное расхождение.

В нашем примере про ВО в последних трех столбцах табл. 9 приведены расчеты накопленных частот и разностей между ними, откуда видно, что в 3 - ей группе наблюдается максимальное расхождение (разность)  $D = 3,664$ . Тогда по формуле (47):  $\lambda = 3,664 / \sqrt{35} = 0,619$ . По таблице находим значение вероятности при  $\lambda = 0,6$ :  $P = 0,86$  (наиболее близкое значение к 0,619), то есть с вероятностью, близкой к 0,86, можно говорить, что в основе эмпирического распределения величины ВО по таможенным постам лежит закон нормального распределения, а расхождения эмпирического и теоретического распределений носят случайный характер.

Подтвердив правильность выдвинутой гипотезы с помощью известных критериев согласия, можно использовать результаты распределения для практической деятельности. Какое же практическое значение может иметь произведенная проверка гипотезы? *Во-первых*, соответствие нормальному закону позволяет прогнозировать, какое число таможенных постов (или их доля) попадет в тот или иной интервал значений величины ВО. *Во-вторых*, нормальное распределение возникает при действии на вариацию изучаемого показателя множества независимых факторов. Из чего следует, что нельзя существенно снизить вариацию величины ВО, воздействуя только на один - два управляемых фактора, скажем число работников таможенного поста или степень технической оснащенности.

Таможенная инспекция провела проверку после выпуска товаров. В результате получен следующий дискретный ряд распределения числа нарушений, выявленных в каждой проверке (табл. 10).

Таблица 10.

**Ряд распределения числа нарушений, выявленных таможенной инспекцией**

|                 |    |   |   |   |
|-----------------|----|---|---|---|
| Число нарушений | 0  | 1 | 2 | 3 |
| Число проверок  | 24 | 4 | 2 | 1 |

Проведем анализ этого ряда распределения. Сначала рассчитаем среднее число нарушений в выборке, а также его дисперсию, для чего построим вспомогательную табл. 11.

Таблица 11.

**Ряд распределения числа нарушений, выявленных таможенной инспекцией**

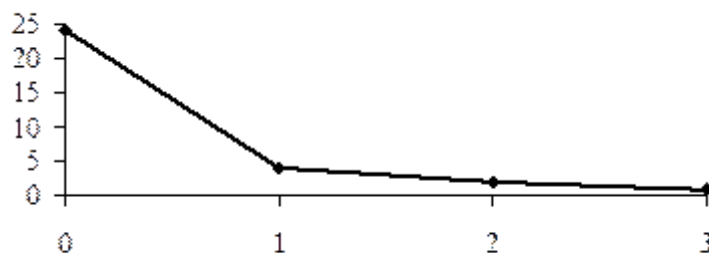
| Число нарушений<br>$X$ | Число проверок<br>$f$ | $Xf$ | $(X - \bar{X})2f$ | $m$  | $\frac{(f - m)^2}{m}$ | $f'$ | $m'$ | $ f' - m' $ |
|------------------------|-----------------------|------|-------------------|------|-----------------------|------|------|-------------|
| 0                      | 24                    | 0    | 3,022             | 21,7 | 0,244                 | 24   | 21,7 | 2,3         |
| 1                      | 4                     | 4    | 1,665             | 7,7  | 1,778                 | 28   | 29,4 | 1,4         |
| 2                      | 2                     | 4    | 5,413             | 1,4  | 0,257                 | 30   | 30,8 | 0,8         |
| 3                      | 1                     | 3    | 6,997             | 0,2  | 3,200                 | 31   | 31   | 0           |
| Итого                  | 31                    | 11   | 17,097            | 31   | 5,479                 |      |      |             |

Среднее число нарушений в выборке по формуле (11):  $\bar{X} = 11/31 = 0,355$  (нарушений).

$$\sigma^2 = \frac{\sum (x_i - \bar{X})^2}{N} = \frac{17,097}{31} = 0,552 \text{ (нарушений } 2).$$

Дисперсию определим по формуле:

Построив график этого распределения (полигон) – рис. 12, видно, что данное распределение не похоже на нормальное.



**Рис. 12. Кривая распределения числа нарушений, выявленных таможенной инспекцией**

Из структурных характеристик ряда распределения можно определить только моду:  $Mo = 0$ , так как по данным табл. 17 такое число нарушений чаще всего встречается ( $f=24$ ).

По формуле определим размах вариации:  $H = 3 - 0 = 3$ , что характеризует вариацию в 3 нарушения.

По формуле найдем среднее линейное отклонение:

$$\bar{L} = \frac{\sum |X - \bar{X}| f}{\sum f} = \frac{|0 - 0,355| 24 + |1 - 0,355| 4 + |2 - 0,355| 2 + |3 - 0,355| 1}{31} = 17,035 : 31 = 0,550$$

Это означает, что в среднем число нарушений отклоняется от среднего их числа на 0,55.

Среднее квадратическое отклонение рассчитаем не по формуле, а как корень из дисперсии, которая уже была рассчитана нами выше:  $\sigma = \sqrt{0,552} = 0,743$ , тогда  $\sigma/\bar{L} \approx 0,743/0,550 = 1,35 > 1,25$ , то есть в изучаемом распределении наблюдается некоторое число выделяющихся нарушений (с большим числом нарушений, выявленных в одной проверке).

Поскольку квартили на предыдущем этапе не определялись, на данном этапе расчет среднего квартильного расстояния пропускаем.

Теперь рассчитаем *относительные показатели вариации*:

– относительный размах вариации по формуле:  $R^0 = 3/0,355 = 8,45$ ;

– линейный коэффициент вариации по формуле:  $\lambda = 0,550/0,355 = 1,55$ ;

– квадратический коэффициент вариации по формуле:  $v = 0,743/0,355 = 2,09$ .

Все расчеты на данном этапе свидетельствуют о значительных размере и интенсивности вариации нарушений, выявленных таможенной инспекцией.

Не имеет практического смысла расчет моментов распределения, так как видно из рис. 12, что в изучаемом распределении симметрия отсутствует вовсе, поэтому и расчет эксцесса также бесполезен.

Выдвинем гипотезу о соответствии изучаемого распределения распределению Пуассона, которое описывается формулой:

$$P(x) = \frac{a^x e^{-a}}{x!},$$

где  $P(X)$  – вероятность того, что признак примет то или иное значение  $X$ ;

$e = 2,7182$  – основание натурального логарифма;

$X!$  – факториал числа  $X$  (то есть произведение всех целых чисел от 1 до  $X$  включительно);

$a = \bar{X}$  – средняя арифметическая ряда распределения.

Из формулы видно, что единственным параметром распределения Пуассона является средняя арифметическая величина. Порядок определения теоретических частот этого распределения следующий:

1) рассчитать среднюю арифметическую ряда, то есть  $= a$ ;

- 2) рассчитать  $e - a$ ;
- 3) для каждого значения  $X$  рассчитать теоретическую частоту по формуле:

$$m_i = N \frac{a^x e^{-a}}{x!} = N * P(X)$$

Поскольку  $a = \bar{X} = 0,355$  найдем значение  $e - 0,355 = 0,7012$ . Затем, подставив в формулу значения  $X$  от 0 до 3, вычислим теоретические частоты:

$$m_0 = 31 \frac{0,355^0 * 0,7012}{0!} = 21,7 \quad (\text{так как } 0! = 1);$$

$$m_1 = 31 \frac{0,355^1 * 0,7012}{1!} = 7,7;$$

$$m_2 = 31 \frac{0,355^2 * 0,7012}{2!} = 1,4;$$

$$m_3 = 31 \frac{0,355^3 * 0,7012}{3!} = 0,2$$

Полученные теоретические частоты занесем в 5-й столбец табл. 17 и построим график эмпирического и теоретического распределений (рис. 13), из которого видна близость эмпирического и теоретического распределений.

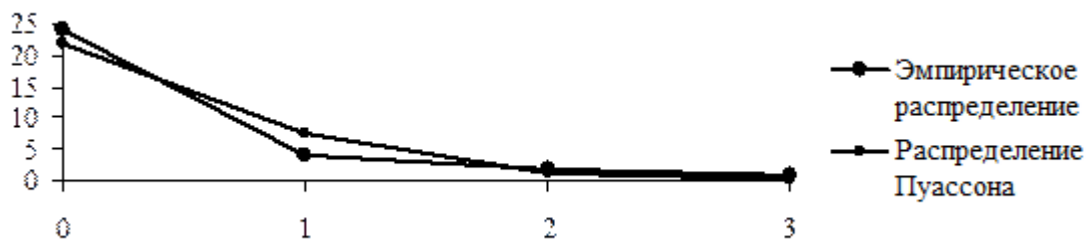


Рис. 13. Эмпирическая и теоретическая (распределение Пуассона) кривые распределения

Проверим выдвинутую гипотезу о соответствии изучаемого распределения закону Пуассона с помощью критериев согласия.

Рассчитаем значение критерия Пирсона  $\chi^2$  по формуле в 6-м столбце табл. 17:  $\chi^2 = 5,479$ , что меньше табличного значения  $\chi^2_{табл} = 5,9915$  при уровне значимости  $\alpha = 0,05$  и числе степеней свободы  $\nu = 4 - 1 - 1 = 2$ , значит с вероятностью 0,95 можно говорить, что в основе эмпирического распределения лежит закон распределения Пуассона, то есть выдвинутая гипотеза не отвергается, а расхождения объясняются случайными факторами.

Определим значение критерия Романовского по формуле:

$$K_p = \frac{|5,479 - 2|}{\sqrt{2 * 2}} = 1,74 < 3, \text{ что подтверждает несущественность расхождений между эмпирическими и теоретическими частотами.}$$

Для расчета критерия Колмогорова в последних трех столбцах таблицы 17 приведены расчеты накопленных частот и разностей между ними, откуда видно, что в 1-ой группе наблюдается максимальное расхождение (разность)  $D = 2,3$ . Тогда по формуле:  $\lambda = 2,3 / \sqrt{31} = 0,413$ . По таблице находим значение вероятности при  $\lambda = 0,4$ :  $P = 0,9972$  (наиболее близкое значение к 0,413), то есть с вероятностью, близкой к единице, можно говорить, что в основе эмпирического распределения величины нарушений, выявленных

таможенной инспекцией, лежит закон распределения Пуассона, а расхождения эмпирического и теоретического распределений носят случайный характер.

### Вопросы для самопроверки:

1. Что позволяют понять законы распределения?
2. В чем заключаются особенности кривой нормального распределения?
3. Критерий Колмогорова  $\lambda$  – это...
4. Что обозначают относительные показатели вариации?

## Тема 8. Абсолютные и относительные показатели изменения структуры. Ранговые показатели изменения структуры

Развитие статистической совокупности проявляется не только в количественном росте или уменьшении элементов системы, но также и в изменении ее структуры. *Структура* – это строение совокупности, состоящее из отдельных элементов и связей между ними. Например, экспорт страны (совокупность) состоит из различных видов товаров (элементов), стоимость которых различается по видам и по странам. Кроме того, происходит постоянное изменение структуры экспорта в динамике. Соответственно возникает задача изучения структуры совокупностей и их динамики, для чего разработаны специальные методы, которые будут рассмотрены далее.

Здесь излагаются показатели, характеризующие изменение структуры в целом. Практическое применение этих показателей рассмотрим на двух примерах, представленных в табл. 12 и 13 (первые 4 столбца, выделенные полужирным шрифтом, – исходные данные, а остальные – вспомогательные расчеты).

Таблица 12.

### Распределение населения России по величине среднедушевых денежных доходов (СДД)

| № группы (j) | СДД, руб./чел. в месяц | Доли населения |               | d1-d0 | $d_0^2$ | $d_1^2$ | (d1-d0)2 | (d1+d0)2 | $\left(\frac{d_1 - d_0}{d_1 + d_0}\right)^2$ |
|--------------|------------------------|----------------|---------------|-------|---------|---------|----------|----------|--|
|              |                        | 2005 год (d0)  | 2006 год (d1) |       |         |         |          |          |  |
| 1            | до 1500                | <b>0,032</b>   | <b>0,018</b>  | 0,014 | 0,0010  | 0,0003  | 0,0002   | 0,0025   | 0,0784                                       |
| 2            | 1500-2500              | <b>0,088</b>   | <b>0,058</b>  | 0,030 | 0,0077  | 0,0034  | 0,0009   | 0,0213   | 0,0422                                       |
| 3            | 2500-3500              | <b>0,113</b>   | <b>0,085</b>  | 0,028 | 0,0128  | 0,0072  | 0,0008   | 0,0392   | 0,0200                                       |
| 4            | 3500-4500              | <b>0,114</b>   | <b>0,094</b>  | 0,020 | 0,0130  | 0,0088  | 0,0004   | 0,0433   | 0,0092                                       |
| 5            | 4500-6000              | <b>0,149</b>   | <b>0,135</b>  | 0,014 | 0,0222  | 0,0182  | 0,0002   | 0,0807   | 0,0024                                       |
| 6            | 6000-8000              | <b>0,149</b>   | <b>0,149</b>  | 0,000 | 0,0222  | 0,0222  | 0,0000   | 0,0888   | 0,0000                                       |
| 7            | 8000-12000             | <b>0,174</b>   | <b>0,197</b>  | 0,023 | 0,0303  | 0,0388  | 0,0005   | 0,1376   | 0,0038                                       |
| 8            | более 12000            | <b>0,181</b>   | <b>0,264</b>  | 0,083 | 0,0328  | 0,0697  | 0,0069   | 0,1980   | 0,0348                                       |
|              | Итого                  | <b>1,000</b>   | <b>1,000</b>  | 0,212 | 0,1420  | 0,1687  | 0,0099   | 0,6114   | 0,1909                                       |

Таблица 13.

Распределение численности безработных России по уровню образования в 2006 г.

| № группы (j) | Имеют образование                     | Мужчины (d0) | Женщины (d1) | $ d1-d0 $ | $d_0^2$ | $d_1^2$ | $(d1-d0)^2$ | $(d1+d0)^2$ | $\left(\frac{d_{1j}-d_{0j}}{d_{1j}+d_{0j}}\right)^2$ |
|--------------|---------------------------------------|--------------|--------------|-----------|---------|---------|-------------|-------------|--|
| 1            | Высшее профессиональное               | 0,087        | 0,130        | 0,043     | 0,0076  | 0,0169  | 0,0018      | 0,0471      | 0,0393   |
| 2            | Неполное высшее профессиональное      | 0,019        | 0,023        | 0,004     | 0,0004  | 0,0005  | 0,0000      | 0,0018      | 0,0091   |
| 3            | Среднее профессиональное              | 0,130        | 0,221        | 0,091     | 0,0169  | 0,0488  | 0,0083      | 0,1232      | 0,0672   |
| 4            | Начальное профессиональное            | 0,200        | 0,149        | 0,051     | 0,0400  | 0,0222  | 0,0026      | 0,1218      | 0,0214   |
| 5            | Среднее (полное) общее                | 0,398        | 0,338        | 0,060     | 0,1584  | 0,1142  | 0,0036      | 0,5417      | 0,0066   |
| 6            | Основное общее                        | 0,148        | 0,121        | 0,027     | 0,0219  | 0,0146  | 0,0007      | 0,0724      | 0,0101   |
| 7            | Начальное общее, не имеют образования | 0,018        | 0,018        | 0,000     | 0,0003  | 0,0003  | 0,0000      | 0,0013      | 0,0000   |
| Итого        |                                       | 1,000        | 1,000        | 0,276     | 0,2455  | 0,2177  | 0,0171      | 0,9092      | 0,1536   |

Обобщающим абсолютным показателем изменения структуры может служить *сумма модулей абсолютных изменений долей*, определяемая по формуле:

$$\Delta d = \sum_{j=1}^k |d_{1j} - d_{0j}|$$

Где  $d_{1j}$  – доля j - ой группы элементов в отчетном периоде;  $d_{0j}$  – доля j - ой группы элементов в базисном периоде.

По данным табл. 12 в 5-м столбце произведен расчет по формуле (50):  $\Delta d = 0,212$ , то есть суммарное изменение долей в распределении россиян по доходам составило 21,2%. Аналогично по той же формуле по данным табл. 20:  $\Delta d = 0,276$ , то есть различие структуры безработных среди женщин и мужчин по уровню образованию составляет 27,6%.

Расчет среднего абсолютного изменения, приходящегося на одну долю (группу, элемент совокупности) не дает никакой дополнительной информации. Зато можно определить, насколько сильно произошедшее изменение структуры в сравнении с предельно возможной величиной суммы модулей, которая равна 2. Для этого используется показатель *степени интенсивности абсолютного сдвига* (или *индекс Лузмора-Хэнби*), который определяется по формуле:

$$K_{\Delta d} = 0,5 \sum_{j=1}^k |d_{1j} - d_{0j}|$$

По данным табл. 12 по формуле:  $K_{\Delta d} = 0,106$ , то есть интенсивность изменения долей в распределении россиян по доходам составила 10,6% от максимально возможного. Аналогично по той же формуле по данным табл. 20:  $K_{\Delta d} = 0,138$ , то есть различие структуры безработных среди женщин и мужчин по уровню образованию составляет 13,8% от максимально возможного.

Обобщенная оценка степени структуризации явления в целом обычно выполняется по формуле уровня концентрации (или *коэффициент Герфиндаля*), который более чувствителен к изменению долей групп с наибольшим удельным весом в итоге, определяемый по формуле:

$$H = \sum_{j=1}^k d_j^2$$

где  $d_i$  – доля  $i$ -го объекта в общем итоге изучаемого показателя;  
 $k$  – количество объектов.

По данным табл. 12 в 6-м и 7-м столбцах произведен расчет коэффициента Герфиндаля по формуле:  $H_{2005}=0,142$  и  $H_{2006}=0,1687$ , то есть уровень концентрации в распределении россиян по доходам увеличился в 2006 году по сравнению с 2005 годом. Аналогично по той же формуле по данным табл. 20:  $H_{\text{муж}} = 0,2455$  и  $H_{\text{жен}} = 0,2177$ , то есть уровень концентрации в распределении безработных по уровню образованию среди мужчин выше, чем среди женщин (влияние уровня образования на статус безработного среди мужчин выше, чем среди женщин).

Обратная индексу Герфиндаля величина – это *эффективное число групп* в структуре, которое показывает количество групп без учета групп, имеющих ничтожно малые доли, определяется по формуле:

$$E = 1/H$$

По данным табл. 12 эффективное число групп по формуле:  $E_{2005} = 1/0,142 = 7,0$  и  $E_{2006} = 5,9$ , то есть эффективное число групп в распределении россиян по доходам уменьшилось с 7 в 2005 году до 6 в 2006 году, что свидетельствует о необходимости пересмотра интервалов распределения россиян по доходам в будущем году. Аналогично по той же формуле по данным табл. 20:  $E_{\text{муж}}=1/0,2455=4,07$  и  $E_{\text{жен}} = 1/0,2177 = 4,59$ , то эффективное число групп в распределении безработных по уровню образованию среди мужчин выше и среди женщин – 4 у мужчин и 5 у женщин.

Еще один вариант оценки степени структуризации явления в целом – *индекс Грофмана*, который представляет собой сумму модулей абсолютных изменений долей, приходящихся на одну эффективную группу:

$$I_{\text{Грофман}} = \frac{\Delta d}{E_0} = H_0 \Delta d$$

По данным табл. 12 в по формуле:  $I_{\text{Грофман}}=0,212*0,142=0,030$ , то есть изменение долей, приходящихся на одну эффективную группу в распределении россиян по доходам незначительно (3,0%). Аналогично по той же формуле по данным табл. 13:  $I_{\text{Грофман}} = 0,2455*0,276 = 0,068$ , то есть различие структуры в расчете на одну эффективную группу среди безработных женщин и мужчин по уровню образованию слабое (6,8%).

Для оценки изменений двух наибольших долей (доминантные доли) применяется *индекс Липхарта*:

$$I_{\text{Липхарт}} = 0,5 \sum_{m=1}^2 |d_{1m} - d_{0m}|$$

где  $d_{1m}$  и  $d_{0m}$  – доля  $m$ -ой группы элементов в отчетном периоде и базисном периодах;  
 $m$  – максимальная доля в совокупности.

По данным табл. 12 по формуле:  $I_{\text{Липхарт}} = 0,5*(0,083+0,023) = 0,053$ , то есть среднее изменение долей в двух доминантных группах распределения россиян по доходам составило 5,3%. Аналогично по той же формуле по данным табл. 20:  $I_{\text{Липхарт}} = 0,5*(0,060+0,051)=0,056$ , то есть различие структуры в двух доминантных группах среди безработных женщин и мужчин по уровню образованию составляет 5,6%.

Рассмотренные показатели основаны на средней арифметической в различных вариантах, и из-за их линейности по отклонениям они одинаково учитывают большие и малые отклонения. *Квадратические индексы* позволяют сравнивать различные структуры, неразличимые с точки зрения суммы изменений.

Квадратический индекс структурных сдвигов *Казинца*:

$$I_{\text{Казинца}} = \sqrt{\frac{\sum (d_{1j} - d_{0j})^2}{k}}$$

По данным табл. 12 по формуле:  $I_{\text{Казинца}} = \sqrt{0,0099/3} = 0,035$ , то есть среднее изменение долей в группе в распределении россиян по доходам составило 3,5% (незначительно). Аналогично по той же формуле по данным табл. 13:  $I_{\text{Казинца}} = \sqrt{0,0171/7} = 0,049$ , то есть различие в группах в структуре безработных среди женщин и мужчин по уровню образованию составляет 4,9% (несущественно).

Аналогичен индексу Казинца *индекс наименьших квадратов* (или *индекс Галлахера*), при расчете которого, в отличие от формулы, малые разности долей слабее влияют на индекс, чем большие, определяется по формуле:

$$I_{\text{LSQ}} = \sqrt{0,5 \sum_{j=1}^k (d_{1j} - d_{0j})^2}$$

По данным табл. 12 по формуле (57):  $I_{\text{LSQ}} = \sqrt{0,5 * 0,0099} = 0,070$ , то есть интенсивность изменения долей в распределении россиян по доходам составила 7,0%. Аналогично по той же формуле по данным табл. 13:  $I_{\text{LSQ}} = \sqrt{0,5 * 0,0171} = 0,092$ , то есть различие структуры безработных среди женщин и мужчин по уровню образованию составляет 9,2%.

Незначительную модификацию индекса наименьших квадратов представляет *индекс Монро*:

$$I_{\text{Монро}} = \sqrt{\frac{\sum_{j=1}^k (d_{1j} - d_{0j})^2}{1 + H_0}}$$

По данным табл. 12 по формуле (58):  $I_{\text{Монро}} = \sqrt{0,0099/(1 + 0,142)} = 0,093$ , то есть интенсивность изменения долей в распределении россиян по доходам по формуле Монро составила 9,3%. Аналогично по той же формуле по данным табл. 13:  $I_{\text{Монро}} = \sqrt{0,0171/(1 + 0,2455)} = 0,117$ , то есть различие структуры безработных среди женщин и мужчин по уровню образованию по формуле Монро составляет 11,7%.

*Интегральный коэффициент структурных сдвигов Гатева*, который различает структуры с равными суммами квадратов отклонений (принимает более высокие значения, когда группы имеют примерно одинаковые доли):

$$I_{\text{Гатев}} = \sqrt{\frac{\sum (d_{1j} - d_{0j})^2}{\sum (d_{1j}^2 + d_{0j}^2)}}$$

По данным табл. 12 по формуле:  $I_{\text{Гатев}} = \sqrt{0,0099/(0,142 + 0,1687)} = 0,179$ , то есть интенсивность изменения долей в распределении россиян по доходам по методике Гатева составила 17,9% (незначительно). Аналогично по той же формуле по данным табл. 13:  $I_{\text{Гатев}} = \sqrt{0,0171/(0,2455 + 0,2177)} = 0,192$ , то есть различие структуры безработных среди женщин и мужчин по уровню образованию по методике Гатева составляет 19,2% (незначительно).

Индекс *Рябцева*, отличающийся только знаменателем, принимает обычно более низкие значения, рассчитывается по формуле:

$$I_{\text{Рябцев}} = \sqrt{\frac{\sum (d_{1j} - d_{0j})^2}{\sum (d_{1j} + d_{0j})^2}}$$

По данным табл. 12 по формуле:  $I_{\text{Рябцев}} = \sqrt{0,0099/0,6114} = 0,127$ , то есть интенсивность изменения долей в распределении россиян по доходам по методике Рябцева составила 12,7% (незначительно). Аналогично по той же формуле по данным табл. 13:  $I_{\text{Рябцев}} = \sqrt{0,0171/0,9092} = 0,137$ , то есть различие структуры безработных среди женщин и мужчин по уровню образованию по методике Рябцева составляет 13,7% (достаточно значительно).

Индекс структурных различий *Салаи*, особенностью которого является то, что чем больше доля  $j$ -ой группы, тем большее значение будет принимать  $(d_{1j} + d_{0j})^2$ , что ведет к уменьшению вклада  $j$ -ой группы в общей сумме, тем самым увеличивая значимость изменения долей малых групп:

$$I_{\text{Салаи}} = \sqrt{\frac{\sum (d_{1j} - d_{0j}) / (d_{1j} + d_{0j})^2}{k}}$$

По данным табл. 12 по формуле:  $I_{\text{Салаи}} = \sqrt{0,1909/8} = 0,154$ , то есть средняя интенсивность изменения долей в распределении россиян по доходам по методике Салаи составила 15,4%. Аналогично по той же формуле по данным табл. 13:  $I_{\text{Салаи}} = \sqrt{0,1536/7} = 0,148$ , то есть среднее различие долей в группах безработных среди женщин и мужчин по уровню образованию по методике Салаи составляет 14,8%.

Для оценки структуры распределения доходов применяются специфические индексы: *индекс Джини*, *индекс Аткинсона*, *индекс обобщенной энтропии*, которые будут рассмотрены в курсе социально - экономической статистики в теме «Статистика уровня жизни».

Для измерения различий структуры часто используют менее точные, но более простые по расчету показатели, которые основаны на оценки различий не самих значений долей, а их рангов, то есть порядковых номеров. Для этого чаще всего используются 2 показателя – линейный и квадратический коэффициенты изменения (различия) рангов долей. Эти показатели, как правило, применяются для анализа структуры распределения описательных (атрибутивных) признаков (например, табл. 13), а также для оценки вотумов (голосований).

В 5-м и 6-м столбцах табл. 14 определены ранги по данным табл. 13, а в последующих приведены вспомогательные расчеты, необходимые в дальнейшем.

**Таблица 14.**

**Вспомогательные расчеты для определения ранговых показателей изменения структуры**

| № группы (j) | Имеют образование                     | d0    | d1    | Ранг мужчин R0 | Ранг женщин R1 | R1 - R0 | (R1 - R0) <sup>2</sup> |
|--------------|---------------------------------------|-------|-------|----------------|----------------|---------|------------------------|
| 1            | Высшее профессиональное               | 0,087 | 0,130 | 5              | 4              | 1       | 1                      |
| 2            | Неполное высшее профессиональное      | 0,019 | 0,023 | 6              | 6              | 0       | 0                      |
| 3            | Среднее профессиональное              | 0,130 | 0,221 | 4              | 2              | 2       | 4                      |
| 4            | Начальное профессиональное            | 0,200 | 0,149 | 2              | 3              | 1       | 1                      |
| 5            | Среднее (полное) общее                | 0,398 | 0,338 | 1              | 1              | 0       | 0                      |
| 6            | Основное общее                        | 0,148 | 0,121 | 3              | 5              | 2       | 4                      |
| 7            | Начальное общее, не имеют образования | 0,018 | 0,018 | 7              | 7              | 0       | 0                      |
| Итого        |                                       | 1,000 | 1,000 |                |                | 6       | 10                     |



*Линейный коэффициент различия рангов долей (ЛК<sub>Р</sub>)* – это отношение фактической суммы модулей изменения рангов к предельно возможной сумме модулей при  $k$  элементах структуры. Для четного  $k$  определяется по формуле,

$$ЛК_R = \frac{\sum |R_{1j} - R_{0j}|}{k^2 / 2},$$

а для нечетного  $k$  – по формуле:

$$ЛК_R = \frac{\sum |R_{1j} - R_{0j}|}{(k^2 - 1) / 2},$$

где  $R_{1j}$  и  $R_{0j}$  – ранги доли  $j$ -го элемента структуры (группы) в сравниваемых совокупностях.

Так по данным табл. 14, где в предпоследнем столбце рассчитана сумма модулей различий рангов, по формуле:

$$ЛК_R = \frac{6}{(7^2 - 1) / 2} = 6 / 24 = 0,25,$$

то есть различие структуры безработных среди женщин и мужчин по уровню образованию ощутимо и составляет 25% от максимально возможного.

*Квадратический коэффициент различия рангов долей (КК<sub>Р</sub>)* основан на коэффициенте корреляции рангов Спирмена, особенностью которого является то, что он позволяет определить корреляцию по таким признакам, которые нельзя выразить численно, но можно проранжировать. При полном совпадении рангов долей в базисном и отчетном периодах коэффициент Спирмена равен +1, а при максимальном различии рангов (первый становится последним, порядок рангов «переворачивается») коэффициент Спирмена составит –1, следовательно, максимальное значение изменения коэффициента Спирмена равно 2. Чтобы получить показатель степени (существенности) различия рангов элементов структуры, следует отклонение фактического коэффициента Спирмена от единицы разделить на 2:

$$КК_R = \frac{1 - r_{\text{Спирмена}}}{2} = \frac{1 - \left( 1 - \frac{6 \sum (R_{1j} - R_{0j})^2}{k^3 - k} \right)}{2} = \frac{3 \sum (R_{1j} - R_{0j})^2}{k^3 - k}.$$

Для расчета квадратического коэффициента различия рангов долей необходима сумма квадратов различий рангов, которая рассчитана в последнем столбце табл. 14, тогда по формуле:

$$КК_R = \frac{3 * 10}{7^3 - 7} = 30 / 336 = 0,089,$$

то есть различие структуры безработных среди женщин и мужчин по уровню образованию составляет 8,9% от максимально возможного.

### Вопросы для самопроверки:

1. Что обозначает структура статистической совокупности?
2. Сумма модулей абсолютных изменений долей – это...
3. Степени интенсивности абсолютного сдвига (или индекс Лузмора-Хэнби) обозначает...
4. Что позволяет понять эффективное число групп в структуре?
5. Что обозначают квадратические индексы?
6. Что позволяет понять индекс наименьших квадратов (или индекс Галлахера)?
7. Интегральный коэффициент структурных сдвигов Гатева – это..
8. Поясните смысл индекса обобщенной энтропии в теории статистики.

## **Тема 9. Понятие выборочного наблюдения. Способы формирования выборки. Средняя ошибка выборки. Предельная ошибка выборки. Необходимая численность выборки**

Выборочный метод используется, когда применение сплошного наблюдения физически невозможно из-за огромного массива данных или экономически нецелесообразно. Физическая невозможность имеет место, например, при изучении пассажиропотоков, рыночных цен, семейных бюджетов. Экономическая нецелесообразность имеет место при оценке качества товаров, связанной с их уничтожением. Например, дегустация, испытание кирпичей на прочность и т.п. Выборочное наблюдение используется также для проверки результатов сплошного.

Статистические единицы, отобранные для наблюдения, составляют *выборочную* совокупность или *выборку*, а весь их массив – *генеральную совокупность* (ГС). При этом число единиц в выборке обозначают  $n$ , во всей ГС –  $N$ . Отношение  $n/N$  называется относительный размер или *доля выборки*.

Качество результатов выборочного наблюдения зависит от *репрезентативности* выборки, то есть от того, насколько она представительна в ГС. Для обеспечения репрезентативности выборки необходимо соблюдать принцип случайности отбора единиц, который предполагает, что на включение единицы ГС в выборку не может повлиять какой-либо иной фактор кроме случая.

1. *Собственно случайный* отбор: все единицы ГС нумеруются, а выпавшие в результате жеребьевки номера соответствуют единицам, попавшим в выборку, причем число номеров равно запланированному объему выборки. На практике вместо жеребьевки используют генераторы случайных чисел. Данный способ отбора может быть *повторным* (когда каждая единица, отобранная в выборку, после проведения наблюдения возвращается в ГС и может быть вновь подвергнута обследованию) и *бесповторным* (когда обследованные единицы в ГС не возвращаются и не могут быть обследованы повторно). При повторном отборе вероятность попадания в выборку для каждой единицы ГС остается неизменной, а при бесповторном отборе она меняется (увеличивается), но для оставшихся в ГС после отбора из нее нескольких единиц, вероятность попадания в выборку одинакова.

2. *Механический* отбор: отбираются единицы генеральной совокупности с постоянным шагом  $N/n$ . Так, если она генеральная совокупность содержит 100 тыс. ед., а требуется выбрать 1 тыс. ед., то в выборку попадет каждая сотая единица.

3. *Стратифицированный* (расслоенным) отбор осуществляется из неоднородной генеральной совокупности, когда ее предварительно разбивают на однородные группы, после чего производят отбор единиц из каждой группы в выборочную совокупность случайным или механическим способом пропорционально их численности в генеральной совокупности.

4. *Серийный* (гнездовой) отбор: случайным или механическим способом выбирают не отдельные единицы, а определенные серии (гнезда), внутри которых производится сплошное наблюдение.

После завершения отбора необходимого числа единиц в выборку и регистрации предусмотренных программой наблюдения изучаемых признаков этих единиц, переходят к расчету обобщающих показателей. К ним относят среднюю величину изучаемого признака и долю единиц, обладающих каким-либо значением этого признака. Однако, если ГС произвести несколько выборок, определив при этом их обобщающие характеристики, то можно установить, что их значения будут различными, кроме того, они будут отличаться и от реального их значения в ГС, если такое определить с помощью сплошного наблюдения. Другими словами, обобщающие характеристики, рассчитанные по данным выборки, будут отличаться от их реальных значений в ГС, поэтому введем следующие условные обозначения (табл. 15).

Таблица 15.

Условные обозначения

| Показатель   | Совокупность |                  |
|--|--------------|------------------|
|  | генеральная  | выборочная       |
| Число единиц совокупности                                  | $N$          | $n$              |
| Среднее значение   | $\bar{X}$    | $\bar{x}$        |
| Доля единиц, обладающих каким-либо значением признака      | $d$          | $\bar{d}$        |
| Доля единиц, не обладающих каким - либо значением признака | $1-d$        | $1-\bar{d}$      |
| Дисперсия  | $\sigma^2$   | $\bar{\sigma}^2$ |

Разность между значением обобщающих характеристик выборочной и генеральной совокупностей называется *ошибкой выборки*, которая подразделяется на ошибку *регистрации* и ошибку *репрезентативности*. Первая возникает из-за неправильных или неточных сведений по причинам непонимания существа вопроса, невнимательности регистратора при заполнении анкет, формуляров и т.п. Она достаточно легко обнаруживается и устраняется. Вторая возникает из-за несоблюдения принципа случайности отбора единиц в выборку. Ее сложнее обнаружить и устранить, она гораздо больше первой и потому ее измерение является основной задачей выборочного наблюдения.

Для измерения ошибки выборки определяется ее средняя ошибка по формуле для повторного отбора

$$\mu = \sqrt{\frac{\sigma^2}{n}};$$

и по формуле – для бесповторного:

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

Из формул видно, что средняя ошибка меньше у бесповторной выборки, что и обуславливает ее более широкое применение.

Учитывая, что на основе выборочного обследования нельзя точно оценить обобщающую характеристику ГС, необходимо найти пределы, в которых он находится. В конкретной выборке разность  $|\bar{X}_i - \bar{X}|$  может быть больше, меньше или равна  $\mu$ . Каждое из отклонений  $|\bar{X}_i - \bar{X}|$  от  $\mu$  имеет определенную вероятность. При выборочном обследовании реальное значение  $\bar{X}$  в ГС неизвестно. Зная среднюю ошибку выборки, с определенной вероятностью можно оценить отклонение выборочной средней от генеральной и установить пределы, в которых находится изучаемый параметр (в данном случае среднее значение) в генеральной совокупности. Отклонение выборочной характеристики от генеральной называется *предельной ошибкой выборки*  $\Delta$ . Она определяется в долях средней ошибки с заданной вероятностью, то есть

$$\Delta = t \mu$$

где  $t$  – коэффициент доверия, зависящий от вероятности, с которой определяется предельная ошибка выборки.

Вероятность появления определенной ошибки выборки находят с помощью теорем теории вероятностей. Согласно теореме Чебышёва, при достаточно большом объеме выборки и ограниченной дисперсии генеральной ГС вероятность того, что разность между выборочной средней и генеральной средней будет сколь угодно мала, близка к единице:

$$P(|\bar{X} - \bar{X}| \leq \epsilon) \rightarrow 1 \text{ при } n \rightarrow \infty.$$

А.М. Ляпунов доказал, что независимо от характера распределения генеральной ГС при увеличении объема выборки распределение вероятностей появления того или иного значения выборочной средней приближается к нормальному распределению (центральная предельная теорема). Следовательно, вероятность отклонения выборочной средней от генеральной средней, то есть вероятность появления заданной предельной ошибки, также подчиняется указанному закону и может быть найдена как функция от  $t$  с помощью интеграла вероятностей Лапласа:

$$P(|\bar{X} - \bar{X}| \leq t\mu) = \frac{1}{\sqrt{2\pi}} \int_{-\frac{t}{\mu}}^{\frac{t}{\mu}} e^{-\frac{t^2}{2}} dt,$$

$$t = \frac{\bar{X} - \bar{X}}{\mu}$$

где  $\mu$  – нормированное отклонение выборочной средней от генеральной средней. Значения  $P$  (интеграла Лапласа) для разных  $t$  рассчитаны и имеются в специальной.

Вероятность, которая принимается при расчете выборочной характеристики, называется *доверительной*. Чаще всего принимают вероятность  $P = 0,950$ , которая означает, что только в 5 случаях из 100 ошибка может выйти за установленные границы. Задавшись конкретным уровнем вероятности, выбирают величину нормированного отклонения  $t$  и рассчитывают предельную ошибку выборки по формуле.

После расчета предельной ошибки находят *доверительный интервал* обобщающей характеристики ГС совокупности по формуле – для среднего значения,

$$\bar{X} = \bar{X} \pm \Delta \quad \text{или} \quad (\bar{X} - \Delta) \leq \bar{X} \leq (\bar{X} + \Delta)$$

и по формуле – для доли единиц, обладающих каким - либо значением признака:

$$d = \bar{d} \pm \Delta \quad \text{или} \quad (\bar{d} - \Delta) \leq d \leq (\bar{d} + \Delta)$$

Следовательно, при выборочном наблюдении определяется не одно, точное значение обобщающей характеристики ГС, а лишь ее доверительный интервал с заданным уровнем вероятности. И это серьезный недостаток выборочного метода статистики.

Разрабатывая программу выборочного наблюдения, задаются конкретным значением предельной ошибки и уровнем вероятности. Неизвестной остается минимальная численность выборки, обеспечивающая заданную точность. Ее можно получить из формул средней и предельной ошибок в зависимости от типа выборки.

Вариация ( $\sigma^2$ ) значений признака к началу выборочного наблюдения как правило неизвестна, поэтому ее берут приближенно одним из способов:

- 1) берется из предыдущих выборочных наблюдений;
- 2) по правилу «трех сигм», согласно которому в размахе вариации укладывается примерно

6 стандартных отклонений  $\sigma$  ( $H/\sigma = 6$ , отсюда  $\sigma^2 = H^2/36$ );

- 3) если приблизительно известна средняя величина изучаемого признака, то  $\sigma^2 = \bar{X}^2/9$ ;

- 4) если неизвестна дисперсия доли единиц, обладающих каким - либо значением признака,

то используется ее максимально возможная величин  $\sigma^2 = 0,25$ .

### Вопросы для самопроверки:

1. Что обозначает выборочная совокупность?
2. Что обозначает генеральная совокупность?
3. Доля выборки – это...
4. Репрезентативность выборки обозначает...
5. В чем смысл понятия «собственно случайный отбор.»?
6. Что отражает понятие «механический отбор»?
7. Стратифицированный (расслоенный) отбор – это...
8. Серийный (гнездовой) отбор обозначает...
9. Ошибка выборки – это...
10. Что обозначает ошибка репрезентативности?

### ***Тема 10. Понятие корреляционной зависимости. Методы выявления и оценки корреляционной связи. Коэффициенты корреляции рангов. Особенности коррелирования рядов динамики. Показатели тесноты связи между качественными признаками***

Один из наиболее общих законов объективного мира – закон всеобщей связи и зависимости между явлениями. Естественно, что, исследуя явления в самых различных областях, статистика неизбежно сталкивается с зависимостями как между количественными, так и между качественными показателями, признаками. Ее задача – обнаружить (выявить) такие зависимости и дать им количественную характеристику.

Среди взаимосвязанных признаков (показателей) одни могут рассматриваться как определенные факторы, влияющие на изменение других (*факторные*), а вторые (*результативные*) – как следствие, результат влияния первых.

Существует два вида связи между отдельными признаками: функциональная и стохастическая (статистическая), частным случаем которой является корреляционная.

Связь между двумя переменными  $x$  и  $y$  называется *функциональной*, если определенному значению переменной  $x$  строго соответствует одно или несколько значений другой переменной  $y$ , и с изменением значения  $x$  значение  $y$  меняется строго определенно. Такие связи обычно встречаются в точных науках. Например, известно, что площадь квадрата равна квадрату его стороны ( $S = a^2$ ). Это соотношение характерно для каждого единичного случая (квадрата), это так называемая *жестко детерминированная* связь. Такие связи можно встретить и в области экономических явлений. Например, при простой сдельной оплате труда связь между оплатой труда  $y$  и количеством изготовленных изделий  $x$  при фиксированной расценке за одну деталь, например 5 руб., легко выразить формулой  $y = 5x$ . Для изучения функциональных связей применяется *индексный метод*.

Существуют и иного рода связи, где взаимно действуют многие факторы, комбинация которых приводит к вариации значений результативного признака (показателя) при одинаковом значении факторного признака. Например, при изучении зависимости величины таможенных платежей, поступающих в федеральный бюджет, от количества товаров, перемещаемых через таможенную границу государства, (или от стоимостного товарооборота) последние будут рассматриваться как факторный признак, а величина таможенных платежей – как результативный. Между ними нет жестко детерминированной связи, то есть при одном и том же количестве перемещенных через таможенную границу товаров (или стоимости товарооборота) величина таможенных платежей, перечисленных разными таможами будет различной, так как кроме количества товаров, перемещаемых через таможенную границу государства, (или стоимость товарооборота) на величину таможенных платежей влияет много других факторов (различная номенклатура товаров, для которых применяются различные

таможенные пошлины, сборы и льготы; различные таможенные режимы перемещения товаров через таможенную границу и др.), комбинация которых вызывает вариацию величины таможенных платежей.

Там, где взаимодействует множество факторов, в том числе и случайных, выявить зависимости, рассматривая единичный случай, невозможно. Такие связи можно обнаружить только при массовом наблюдении как статистические закономерности. Выявленная таким образом связь именуется *стохастической*.

Корреляционная связь – понятие более узкое, чем стохастическая связь, это ее частный случай. Именно корреляционные связи являются предметом изучения статистики.

*Корреляционная связь* – это связь, проявляющаяся при большом числе наблюдений в виде определенной зависимости между средним значением результативного признака и признаками - факторами. Другими словами, корреляционную связь условно можно рассматривать как своего рода функциональную связь средней величины одного признака (результативного) со значением другого (или других). При этом, если рассматривается связь средней величины результативного показателя  $y$  с одним признаком-фактором  $x$ , корреляция называется *парной*, а если факторных признаков 2 и более ( $x_1, x_2, \dots, x_m$ ) – *множественной*.

По характеру изменений  $x$  и  $y$  в парной корреляции различают *прямую* и *обратную* связь. При прямой связи значения обоих признаков изменяются в одном направлении, то есть с увеличением (уменьшением) значений  $x$  увеличиваются (уменьшаются) и значения  $y$ . При обратной связи значения факторного и результативного признаков изменяются в разных направлениях.

Изучение корреляционных связей сводится в основном к решению следующих задач:

1) выявление наличия (отсутствия) корреляционной связи между изучаемыми признаками;

2) измерение тесноты связи между двумя (и более) признаками с помощью специальных коэффициентов (эта часть исследования именуется корреляционным анализом);

3) определение уравнения регрессии – математической модели, в которой среднее значение результативного признака  $y$  рассматривается как функция одной или нескольких переменных – факторных признаков (эта часть исследования именуется регрессионным анализом).

Общий термин «*корреляционно - регрессионный анализ*» подразумевает всестороннее исследование корреляционных связей (то есть решение всех трех задач).

Таблица 16.

Величина внешнеторгового оборота и таможенных платежей

| Месяц    | Оборот, млрд.у.е. | Платеж, млрд.руб. |
|----------|-------------------|-------------------|
| Январь   | 27,068            | 172,17            |
| Февраль  | 29,889            | 200,90            |
| Март     | 34,444            | 231,83            |
| Апрель   | 33,158            | 232,10            |
| Май      | 37,755            | 233,40            |
| Июнь     | 37,554            | 236,99            |
| Июль     | 37,299            | 246,53            |
| Август   | 40,370            | 253,62            |
| Сентябрь | 37,909            | 256,43            |
| Октябрь  | 38,348            | 261,89            |
| Ноябрь   | 39,137            | 259,36            |
| Декабрь  | 46,298            | 278,87            |

Корреляционно - регрессионный анализ находит широкое применение в статистике. Рассмотрим его практическое применение на примере данных таможенной статистики внешней торговли России в 2006 году (табл. 16).

В качестве факторного признака  $x$  примем стоимостной внешнеторговый товарооборот в млрд. долл. США, а в качестве результативного признака  $y$  – величину таможенных платежей в федеральный бюджет в млрд. руб.

| $x$    | $y$    |
|--------|--------|
| 27,068 | 172,17 |
| 29,889 | 200,90 |
| 33,158 | 232,10 |
| 34,444 | 231,83 |
| 37,299 | 246,53 |
| 37,554 | 236,99 |
| 37,755 | 233,40 |
| 37,909 | 256,43 |
| 38,348 | 261,89 |
| 39,137 | 259,36 |
| 40,370 | 253,62 |
| 46,298 | 278,87 |

Для выявления наличия и характера корреляционной связи между двумя признаками в статистике используется ряд методов.

**1. Рассмотрение параллельных данных** (значений  $x$  и  $y$  в каждой из  $n$  единиц). Единицы наблюдения необходимо расположить по возрастанию значений факторного признака  $x$  (как в таблице справа) и затем сравнить с ним (визуально) поведение результативного признака  $y$ .

В нашей задаче в 6 случаях по мере увеличения значений  $x$  увеличиваются и значения  $y$ , а в 5 случаях этого не происходит, поэтому затруднительно говорить о прямой связи между  $x$  и  $y$ .

**2. Графический метод** – это графическое изображение корреляционной зависимости. Для этого, имея  $n$  взаимосвязанных пар значений  $x$  и  $y$  и пользуясь прямоугольной системой координат, каждую такую пару изображают в виде точки на плоскости с координатами  $x$  и  $y$ . Совокупность полученных точек представляет собой *корреляционное поле* (рис. 14), а соединяя последовательно нанесенные точки отрезками, получают ломаную линию, именуемую *эмпирической линией регрессии* (рис. 15).

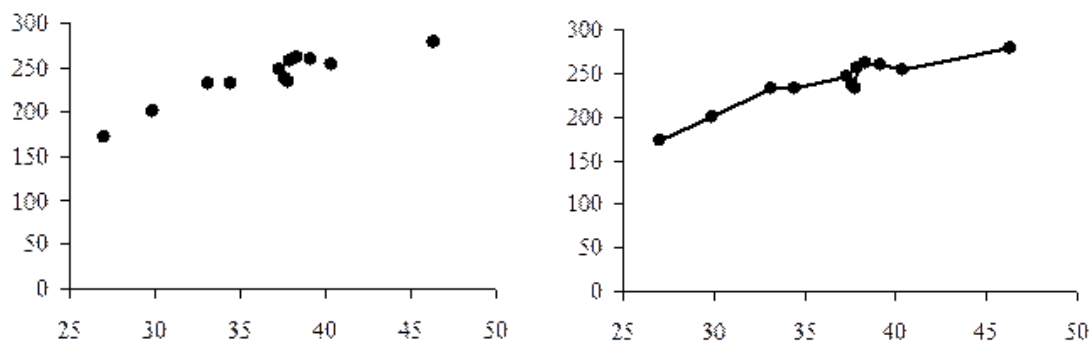


Рис. 14. Корреляционное поле  
Рис. 15. Эмпирическая линия регрессии

Визуально анализируя график, можно предположить характер зависимости между признаками  $x$  и  $y$ . В нашей задаче эмпирическая линия регрессии (рис.15) похожа на восходящую прямую, что позволяет выдвинуть гипотезу о наличии прямой зависимости между величиной стоимостного внешнеторгового товарооборота и величиной таможенных платежей в федеральный бюджет.

**3. Метод аналитических группировок** используется при большом числе наблюдений для выявления корреляционной связи между двумя количественными признаками. Чтобы выявить наличие корреляционной связи между двумя признаками, проводится группировка единиц совокупности по факторному признаку  $x$  и для каждой выделенной группы рассчитывается среднее значение результативного признака  $\bar{y}_j$ . Если результативный признак  $y$  зависит от факторного  $x$ , то в изменении среднего значения  $\bar{y}_j$  будет прослеживаться определенная закономерность. Примером такой группировки могут служить данные об издержках обращения предприятий оптовой торговли с различным товарооборотом (табл. 17).

Таблица 17.

Условные пример аналитической группировки

| Оптовый товарооборот, млн.руб. | Количество предприятий | Издержки обращения, % к оптовому товарообороту |
|--------------------------------|------------------------|--|
| менее 25                       | 9362                   | 46,0   |
| 26-50                          | 3633                   | 26,5   |
| 51-100                         | 3618                   | 24,4   |
| 101-200                        | 3261                   | 23,0   |
| 201-500                        | 3031                   | 17,6   |
| более 501                      | 3100                   | 16,9   |

В последнем столбце табл. 17 приведены средние величины, рассчитанные на основе индивидуальных данных об издержках отдельных предприятий каждой группы. Данные табл. 17 свидетельствуют, что чем крупнее товарооборот, тем меньше издержки обращения. Таким образом, с помощью простой аналитической группировки можно выявить наличие зависимости между рассматриваемыми показателями: объемом товарооборота как показателем размера предприятий и средним уровнем издержек обращения.

**4. Метод корреляционных таблиц** предполагает комбинационное распределение единиц совокупности по двум количественным признакам. Такая таблица строится по типу «шахматной», то есть в подлежащем (строках) таблицы выделяются группы по факторному признаку  $x$ , а в сказуемом (столбцах) – по результативному  $y$  (или наоборот), а в клетках таблицы на пересечении  $x$  и  $y$  показано число случаев совпадения каждого значения  $x$  с соответствующим значением  $y$ . Общий вид такой таблицы показан на условном распределении 40 единиц по признакам  $x$  и  $y$ , где  $x$  – стаж работы,  $y$  – производительность труда (число изделий, вырабатываемых в час одним рабочим) – табл. 18. Среднее значение по группам определяется по средней арифметической взвешенной по серединам группировочных интервалов.

Таблица 18.

Условные корреляционной таблицы

| Значение признака $x_j$ | Значение признака $y_i$ |          |           |            | Итого | Среднее значение по группам $\bar{y}_j$ |
|-------------------------|-------------------------|----------|-----------|------------|-------|---|
|                         | менее 7,5               | 7,5-12,5 | 12,5-17,5 | более 17,5 |       |   |
| менее 2                 | 1                       | 3        | –         | –          | 4     | 8,75                                    |
| 2 – 4                   | 2                       | 3        | 7         | –          | 12    | 12,08                                   |
| 4 – 6                   | –                       | 3        | 9         | 4          | 16    | 15,31                                   |
| 6 – 8                   | –                       | –        | 5         | 3          | 8     | 16,87                                   |
| Итого                   | 3                       | 9        | 21        | 7          | 40    | 14,00                                   |



Как видно из табл. 18, по мере увеличения значений  $x$  итоговые групповые средние  $\bar{y}_j$  тоже увеличиваются от группы к группе, что свидетельствует о том, что между  $x$  и  $y$  существует корреляционная связь. О наличии и направлении связи можно судить и по «внешнему виду» таблицы, то есть по расположению в ней частот: если частоты расположены в клетках таблицы беспорядочно, то это чаще всего свидетельствует об отсутствии связи между группировочными признаками (или о незначительной зависимости); если частоты сконцентрированы ближе к одной из диагоналей и центру таблицы, образуя своего рода эллипс, то это почти всегда свидетельствует о наличии зависимости между  $x$  и  $y$ , близкой к линейной. Расположение по диагонали из верхнего левого угла в нижний правый свидетельствует о прямой линейной связи, а из нижнего левого угла в верхний правый – об обратной.

На основе аналитических группировок и корреляционных таблиц можно не только выявить наличие зависимости между двумя коррелируемыми показателями, но и измерить тесноту этой связи, в частности, с помощью *эмпирического корреляционного отношения*.

$$\eta_{эмп} = \sqrt{\frac{D_{ик}}{D_{общ}}}$$

$$D_{ик} = \frac{\sum_{j=1}^m (y_j - \bar{y})^2 f_j}{\sum f_j}$$

$$D_{общ} = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 f_i}{\sum f_i}$$

где  $m$  – число групп по факторному признаку  $x$ ;  
 $k$  – число групп по результативному признаку  $y$ ;  
 $\bar{y}_j$  – средние значения результативного признака по группам;  
 $\bar{y}$  – общее среднее значение результативного признака;  
 $y_j$  – индивидуальные значения результативного признака;  
 $f_j = f_x$  – частота в  $j$ -й группе  $x$ ;  
 $f_i = f_y$  – частота в  $i$ -й группе  $y$ .

Рассчитаем это отношение для нашего примера

(таблица 41):

$$\bar{y} = \frac{\sum y_i f_i}{\sum f_i}$$

$$= (5*3 + 10*9 + 15*21 + 20*7) / 40 = 14$$

$$D_{ик} = \frac{(8,75 - 14)^2 * 4 + (12,08 - 14)^2 * 12 + (15,31 - 14)^2 * 16 + (16,87 - 14)^2 * 8}{40}$$

$$= 6,19599;$$

$$D_{общ} = \frac{(5 - 14)^2 * 3 + (10 - 14)^2 * 9 + (15 - 14)^2 * 21 + (20 - 14)^2 * 7}{40}$$

$$= 16,5;$$

$$\eta_{эмп} = \sqrt{\frac{6,19599}{16,5}}$$

$$= 0,613.$$

Полученное значение  $\eta = 0,613$  позволяет утверждать, что существует заметная связь между стажем работы и производительностью труда.

**5. Коэффициент корреляции знаков (Фехнера)** – простейший показатель тесноты связи, основанный на сравнении поведения отклонений индивидуальных значений каждого признака ( $x$  и  $y$ ) от своей средней величины. При этом во внимание принимаются не величины отклонений  $(x_i - \bar{x})$  и  $(y_i - \bar{y})$ , а их знаки («+» или «-»). Определив знаки отклонений от средней величины в каждом ряду, рассматривают все пары знаков и подсчитывают число их совпадений ( $C$ ) и несовпадений ( $H$ ). Тогда коэффициент Фехнера рассчитывается как отношение разности чисел пар совпадений и несовпадений знаков к их сумме, то есть к общему числу наблюдаемых единиц:

$$K\Phi = \frac{\sum C - \sum H}{\sum C + \sum H}$$

Очевидно, что если знаки всех отклонений по каждому признаку совпадут, то  $K\Phi=1$ , что характеризует наличие прямой связи. Если все знаки не совпадут, то  $K\Phi= - 1$  (обратная связь). Если же  $\sum C = \sum H$ , то  $K\Phi=0$ . Итак, как и любой показатель тесноты связи, коэффициент Фехнера может принимать значения от 0 до  $\pm 1$ . Однако, если  $K\Phi=1$ , то это ни в коей мере нельзя воспринимать как свидетельство функциональной зависимости между  $x$  и  $y$ .

Средние значения факторного и результирующего признаков определяем по формуле средней арифметической простой:

$$\bar{x} = \frac{\sum x}{n} = \frac{439,229}{12} = 36,602; \quad \bar{y} = \frac{\sum y}{n} = \frac{2864,09}{12} = 238,674$$

В двух последних столбцах табл. 19 приведены знаки отклонений каждого  $x$  и  $y$  от своей средней величины. Число совпадений знаков – 10, а несовпадений – 2, тогда определяем коэффициент корреляции знаков (Фехнера) по формуле:

$$K\Phi = \frac{10 - 2}{10 + 2} = \frac{8}{12} = \frac{2}{3} = 0,667.$$

Таблица 19.

Вспомогательная таблица для расчета коэффициента Фехнера

| № п/п | $x$     | $y$     | $x - \bar{x}$ | $y - \bar{y}$ |
|-------|---------|---------|---------------|---------------|
| 1     | 27,068  | 172,17  | -             | -             |
| 2     | 29,889  | 200,90  | -             | -             |
| 3     | 33,158  | 232,10  | -             | -             |
| 4     | 34,444  | 231,83  | -             | -             |
| 5     | 37,299  | 246,53  | +             | +             |
| 6     | 37,554  | 236,99  | +             | -             |
| 7     | 37,755  | 233,40  | +             | -             |
| 8     | 37,909  | 256,43  | +             | +             |
| 9     | 38,348  | 261,89  | +             | +             |
| 10    | 39,137  | 259,36  | +             | +             |
| 11    | 40,370  | 253,62  | +             | +             |
| 12    | 46,298  | 278,87  | +             | +             |
| Итого | 439,229 | 2864,09 |               |               |

Обычно такое значение показателя тесноты связи характеризует заметную прямую зависимость между  $x$  и  $y$ , однако, следует иметь в виду, что поскольку  $K\Phi$  зависит только от знаков и не учитывает величину самих отклонений  $x$  и  $y$  от их средних величин, то он практически характеризует не столько тесноту связи, сколько ее наличие и направление.

**6. Линейный коэффициент корреляции** – самый популярный измеритель тесноты линейной связи между двумя количественными признаками  $x$  и  $y$ . Он основан на предположении, что при *полной независимости признаков*  $x$  и  $y$  отклонения значений факторного признака от средней ( $x - \bar{x}$ ) носят случайный характер и должны случайно сочетаться с различными отклонениями ( $y - \bar{y}$ ). При наличии значительного перевеса совпадений или несовпадений таких отклонений делается предположение о наличии связи между  $x$  и  $y$ .

В отличие от  $K\Phi$  в линейном коэффициенте корреляции учитываются не только знаки отклонений от средних величин, но и значения самих отклонений, выраженные для сопоставимости в единицах среднего квадратического отклонения  $t$ :

$$t_x = \frac{x - \bar{x}}{\sigma_x} \quad \text{и} \quad t_y = \frac{y - \bar{y}}{\sigma_y}$$

Линейный коэффициент корреляции  $r$  представляет собой среднюю величину из произведений нормированных отклонений для  $x$  и  $y$ :

$$r = \frac{\sum \left( \frac{x - \bar{x}}{\sigma_x} \right) \left( \frac{y - \bar{y}}{\sigma_y} \right)}{n} = \frac{\sum t_x t_y}{n}, \quad \text{или} \quad r = \frac{\sum (x - \bar{x})(y - \bar{y})}{n \sigma_x \sigma_y}$$

Числитель формулы, деленный на  $n$ , представляющий собой среднее произведение отклонений значений двух признаков от их средних значений, называется *коэффициентом ковариации* – это мера совместной вариации факторного  $x$  и результативного  $y$  признаков:

$$\text{cov}(x, y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{n} = \overline{(x - \bar{x})(y - \bar{y})}$$

Недостатком коэффициента ковариации является то, что он не нормирован, в отличие от линейного коэффициента корреляции. Очевидно, что линейный коэффициент корреляции представляет собой частное от деления ковариации между  $x$  и  $y$  на произведение их средних квадратических отклонений:

$$r = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}$$

Путем несложных математических преобразований можно получить и другие модификации формулы линейного коэффициента корреляции, например:

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x \sigma_y},$$

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

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

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

Линейный коэффициент корреляции может принимать значения от  $-1$  до  $+1$ , причем знак определяется в ходе решения. Например, если  $\overline{xy} > \bar{x}\bar{y}$ , то  $r$  по формуле будет положительным, что характеризует прямую зависимость между  $x$  и  $y$ , в противном случае ( $r < 0$ ) – обратную связь. Если  $\overline{xy} = \bar{x}\bar{y}$ , то  $r=0$ , что означает отсутствие линейной зависимости между  $x$  и  $y$ , а при  $r=1$  – функциональная зависимость между  $x$  и  $y$ . Следовательно, всякое промежуточное значение  $r$  от  $0$  до  $1$  характеризует степень приближения корреляционной связи между  $x$  и  $y$  к функциональной. Существует эмпирическое правило (шкала Чэддока) для оценки тесноты связи, представленное в табл. 20.

Таблица 20.

Шкала Чэддока

| <i>  r  </i> | <i>Теснота связи</i>       |
|--------------|----------------------------|
| менее 0,1    | отсутствует линейная связь |
| 0,1 ÷ 0,3    | слабая                     |
| 0,3 ÷ 0,5    | умеренная                  |
| 0,5 ÷ 0,7    | заметная                   |
| более 0,7    | сильная (тесная)           |

Таким образом, коэффициент корреляции при линейной зависимости служит как мерой тесноты связи, так и показателем, характеризующим степень приближения корреляционной зависимости между  $x$  и  $y$  к линейной. Поэтому близость значения  $r$  к 0 в одних случаях может означать отсутствие связи между  $x$  и  $y$ , а в других свидетельствовать о том, что зависимость не линейная.

В этой задаче для расчета  $r$  построим вспомогательную табл. 21.

Таблица 21.

Вспомогательные расчеты линейного коэффициента корреляции

| № п/п | $x$     | $y$     | $(x - \bar{x})^2$ | $(y - \bar{y})^2$ | $tx$  | $ty$   | $tx ty$ | $(x - \bar{x})(y - \bar{y})$ | $xy$       |
|-------|---------|---------|-------------------|-------------------|-------|--------|---------|------------------------------|------------|
| 1     | 27,068  | 172,17  | 90,905            | 4422,804          | 1,993 | -2,408 | 4,799   | 634,078                      | 4660,298   |
| 2     | 29,889  | 200,90  | 45,070            | 1426,888          | 1,403 | -1,368 | 1,919   | 253,594                      | 6004,700   |
| 3     | 33,158  | 232,10  | 11,864            | 43,220            | 0,720 | -0,238 | 0,171   | 22,644                       | 7695,972   |
| 4     | 34,444  | 231,83  | 4,659             | 46,843            | 0,451 | -0,248 | 0,112   | 14,773                       | 7985,153   |
| 5     | 37,299  | 246,53  | 0,485             | 61,714            | 0,146 | 0,284  | 0,041   | 5,472                        | 9195,322   |
| 6     | 37,554  | 236,99  | 0,906             | 2,836             | 0,199 | -0,061 | -0,012  | -1,603                       | 8899,922   |
| 7     | 37,755  | 233,40  | 1,328             | 27,817            | 0,241 | -0,191 | -0,046  | -6,079                       | 8812,017   |
| 8     | 37,909  | 256,43  | 1,707             | 315,270           | 0,273 | 0,643  | 0,176   | 23,199                       | 9721,005   |
| 9     | 38,348  | 261,89  | 3,047             | 538,975           | 0,365 | 0,841  | 0,307   | 40,525                       | 10042,958  |
| 10    | 39,137  | 259,36  | 6,424             | 427,904           | 0,530 | 0,749  | 0,397   | 52,430                       | 10150,572  |
| 11    | 40,37   | 253,62  | 14,195            | 223,378           | 0,788 | 0,541  | 0,426   | 56,310                       | 10238,639  |
| 12    | 46,298  | 278,87  | 94,004            | 1615,705          | 2,027 | 1,455  | 2,950   | 389,722                      | 12911,123  |
| Итого | 439,229 | 2864,09 | 274,594           | 9153,353          |       |        | 11,241  | 1485,066                     | 106317,681 |

В нашей задаче:  $\sigma_x = \sqrt{274,594/12} = 4,784$ ;  $\sigma_y = \sqrt{9153,353/12} = 27,618$ .

Тогда линейный коэффициент корреляции по формуле:

$$r = 11,241/12 = 0,937.$$

Аналогичный результат получаем по формуле:

$$r = 1485,066/(12*4,784*27,618) = 0,937$$

Или по формуле:

$$R = (106317,681/12 - 36,602*238,674) / (4,784*27,618) = 0,937,$$

Найденное значение свидетельствует о том, что связь между величиной стоимостного внешнеторгового товарооборота и величиной таможенных платежей в федеральный бюджет очень близка к функциональной (сильная по шкале Чэддока).

*Проверка коэффициента корреляции на значимость (существенность).* Интерпретируя значение коэффициента корреляции, следует иметь в виду, что он рассчитан для ограниченного числа наблюдений и подвержен случайным колебаниям, как и сами значения  $x$  и  $y$ , на основе которых он рассчитан. Другими словами, как любой выборочный показатель, он содержит случайную ошибку и не всегда однозначно отражает действительно реальную связь между изучаемыми показателями. Для того, чтобы оценить существенность (значимость) самого  $r$  и, соответственно, реальность измеряемой связи между  $x$  и  $y$ , необходимо рассчитать среднюю квадратическую ошибку коэффициента корреляции  $\sigma_r$ . Оценка существенности (значимости)  $r$

основана на сопоставлении значения  $r$  с его средней квадратической ошибкой:  $\frac{|r|}{\sigma_r}$ .

Существуют некоторые особенности расчета  $\sigma_r$  в зависимости от числа наблюдений (объема выборки) –  $n$ .

1. Если число наблюдений достаточно велико ( $n > 30$ ), то  $\sigma_r$  рассчитывается по формуле:

$$\sigma_r = \frac{1-r^2}{\sqrt{n}}$$

Обычно, если  $\frac{|r|}{\sigma_r} > 3$ , то  $r$  считается значимым (существенным), а связь – реальной. Задавшись определенной вероятностью, можно определить *доверительные пределы (границы)*  $r = (r \pm t\sigma_r)$ , где  $t$  – коэффициент доверия, рассчитываемый по интегралу Лапласа.

2. Если число наблюдений небольшое ( $n < 30$ ), то  $\sigma_r$  рассчитывается по формуле:

$$\sigma_r = \frac{\sqrt{1-r^2}}{\sqrt{n-2}}$$

а значимость  $r$  проверяется на основе  $t$ -критерия Стьюдента, для чего определяется расчетное значение критерия по формуле и сопоставляется с  $t$  ТАБЛ.

$$t_{\text{расч}} = \frac{|r|}{\sigma_r} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Табличное значение  $t$  ТАБЛ находится по таблице распределения  $t$ -критерия Стьюдента при уровне значимости  $\alpha = 1 - \beta$  и числе степеней свободы  $\nu = n - 2$ . Если  $t_{\text{расч}} > t_{\text{ТАБЛ}}$ , то  $r$  считается значимым, а связь между  $x$  и  $y$  – реальной. В противном случае ( $t_{\text{расч}} < t_{\text{ТАБЛ}}$ ) считается, что связь между  $x$  и  $y$  отсутствует, и значение  $r$ , отличное от нуля, получено случайно.

В нашей задаче число наблюдений небольшое, значит, оценивать существенность (значимость) линейного коэффициента корреляции будем по формулам:

$$\sigma_r = \frac{\sqrt{1-0,937^2}}{\sqrt{12-2}} = 0,349/3,162 = 0,110;$$

$$t_{\text{расч}} = \frac{|r|}{\sigma_r} = 0,937/0,110 = 8,482.$$

При числе степеней свободы  $\nu = 12 - 2 = 10$  (в 10-й строке) и вероятности  $\beta = 95\%$  (уровень значимости  $\alpha = 1 - \beta = 0,05$ )  $t_{\text{табл}} = 2,2281$ , а при вероятности 99% ( $\alpha = 0,01$ )  $t_{\text{табл}} = 3,169$ , значит,  $t_{\text{расч}} > t_{\text{ТАБЛ}}$ , что дает возможность считать линейный коэффициент корреляции  $r = 0,937$  значимым.

**7. Подбор уравнения регрессии** представляет собой математическое описание изменения взаимно коррелируемых величин по эмпирическим (фактическим) данным. Уравнение регрессии должно определить, каким будет среднее значение результирующего признака  $y$  при том или ином значении факторного признака  $x$ , если остальные факторы, влияющие на  $y$  и не связанные с  $x$ , не учитывать, то есть абстрагироваться от них. Другими словами, уравнение регрессии можно рассматривать как вероятностную гипотетическую функциональную связь величины результирующего признака  $y$  со значениями факторного признака  $x$ .

Уравнение регрессии можно также назвать *теоретической линией регрессии*. Рассчитанные по уравнению регрессии значения результирующего признака называются *теоретическими*. Они обычно обозначаются  $\hat{y}_x$  или  $\hat{y}_x$  (читается: «игрек, выравненный по  $x$ ») и рассматриваются как функция от  $x$ , то есть  $\hat{y}_x = f(x)$ .

Найти в каждом конкретном случае тип функции, с помощью которой можно наиболее адекватно отразить ту или иную зависимость между признаками  $x$  и  $y$ , — одна из основных задач регрессионного анализа. Выбор теоретической линии регрессии часто обусловлен формой эмпирической линии регрессии; теоретическая линия как бы сглаживает изломы эмпирической линии регрессии. Кроме того, необходимо учитывать природу изучаемых показателей и специфику их взаимосвязей.

Для аналитической связи между  $x$  и  $y$  могут использоваться виды уравнений, приведенные в табл. 30 (при условии замены  $t$  на  $x$ ). Обычно зависимость, выражаемую уравнением прямой, называют *линейной* (или *прямолинейной*), а все остальные — *криволинейными зависимостями*.

Выбрав тип функции (табл. 30), по эмпирическим данным определяют параметры уравнения. При этом отыскиваемые параметры должны быть такими, при которых рассчитанные по уравнению теоретические значения результирующего признака  $\hat{y}_x$  были бы максимально близки к эмпирическим данным.

Существует несколько методов нахождения параметров уравнения регрессии. Наиболее часто используется *метод наименьших квадратов* (МНК). Его суть заключается в следующем требовании: искомые теоретические значения результирующего признака  $\hat{y}_x$  должны быть такими, при которых бы обеспечивалась минимальная сумма квадратов их отклонений от эмпирических значений, то есть

$$S = \sum (y - \hat{y}_x)^2 \rightarrow \min$$

Поставив данное условие, легко определить, при каких значениях  $a_0$ ,  $a_1$  и т.д. для каждой аналитической кривой эта сумма квадратов отклонений будет минимальной. Данный метод уже использовался нами в теме 6 «Статистическое изучение динамики ВЭД», поэтому, воспользуемся формулой для нахождения параметров теоретической линии регрессии, заменив параметр  $t$  на  $x$ :

$$\begin{cases} na_0 + a_1 \sum x = \sum y \\ a_0 \sum x + a_1 \sum x^2 = \sum xy \end{cases}$$

Выразив из первого уравнения системы  $a_0$ , получим:

$$a_0 = \frac{\sum y}{n} - a_1 \frac{\sum x}{n} = \bar{y} - a_1 \bar{x}$$

Подставив (134) во второе уравнение системы (133), затем разделив обе его части на  $n$ , получим:

$$(\bar{y} - a_1 \bar{x}) \frac{\sum x}{n} + a_1 \frac{\sum x^2}{n} = \frac{\sum xy}{n}$$

Применяя 3 раза формулу средней арифметической, получим:

$$(\bar{y} - a_1 \bar{x})\bar{x} + a_1 \bar{x}^2 = \overline{xy}$$

Раскрыв скобки и перенеся члены без  $a_1$  в правую часть уравнения, выразим  $a_1$ :

$$a_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\bar{x}^2 - \bar{x}^2} = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x^2}$$

Параметр  $a_1$  в уравнении линейной регрессии называется *коэффициентом регрессии*, который показывает, на сколько изменится значение результативного признака  $y$  при изменении факторного признака  $x$  на единицу.

Исходные данные и расчеты для нашего примера представим в табл. 22.

Таблица 22.

**Вспомогательные расчеты для нахождения уравнения регрессии**

| № п/п | x       | y       | x <sup>2</sup> | xy         | $\hat{y}_x$ | $(y - \hat{y}_x)^2$ | $(\hat{y}_x - \bar{y})^2$ |
|-------|---------|---------|----------------|------------|-------------|---------------------|---------------------------|
| 1     | 27,068  | 172,17  | 732,677        | 4660,298   | 187,124     | 223,612             | 2657,453                  |
| 2     | 29,889  | 200,90  | 893,352        | 6004,700   | 202,377     | 2,181               | 1317,497                  |
| 3     | 33,158  | 232,10  | 1099,453       | 7695,972   | 220,05      | 145,147             | 346,774                   |
| 4     | 34,444  | 231,83  | 1186,389       | 7985,153   | 227,006     | 23,274              | 136,153                   |
| 5     | 37,299  | 246,53  | 1391,215       | 9195,322   | 242,443     | 16,706              | 14,202                    |
| 6     | 37,554  | 236,99  | 1410,303       | 8899,922   | 243,821     | 46,669              | 26,495                    |
| 7     | 37,755  | 233,40  | 1425,440       | 8812,017   | 244,908     | 132,441             | 38,864                    |
| 8     | 37,909  | 256,43  | 1437,092       | 9721,005   | 245,741     | 114,256             | 49,940                    |
| 9     | 38,348  | 261,89  | 1470,569       | 10042,958  | 248,115     | 189,761             | 89,122                    |
| 10    | 39,137  | 259,36  | 1531,705       | 10150,572  | 252,381     | 48,710              | 187,871                   |
| 11    | 40,370  | 253,62  | 1629,737       | 10238,639  | 259,048     | 29,459              | 415,076                   |
| 12    | 46,298  | 278,87  | 2143,505       | 12911,123  | 291,100     | 149,580             | 2748,498                  |
| Итого | 439,229 | 2864,09 | 16351,437      | 106317,681 | 2864,115    | 1121,795            | 8027,945                  |

$$a_1 = \frac{106317,681/12 - 36,602 * 238,674}{4,784^2} = 5,407.$$

По формуле:

$$a_0 = 238,674 - 5,407 * 36,602 = 40,767.$$

Отсюда получаем уравнение регрессии:  $\hat{y}_x = 40,767 + 5,407x$ , подставляя в которое вместо  $x$  эмпирические значения факторного признака (2-й столбец табл. 22), получаем выравненные по прямой линии теоретические значения результативного признака  $\hat{y}_x$  (6-й столбец табл. 22). Для иллюстрации различий между эмпирическими и теоретическими линиями регрессии построим график.

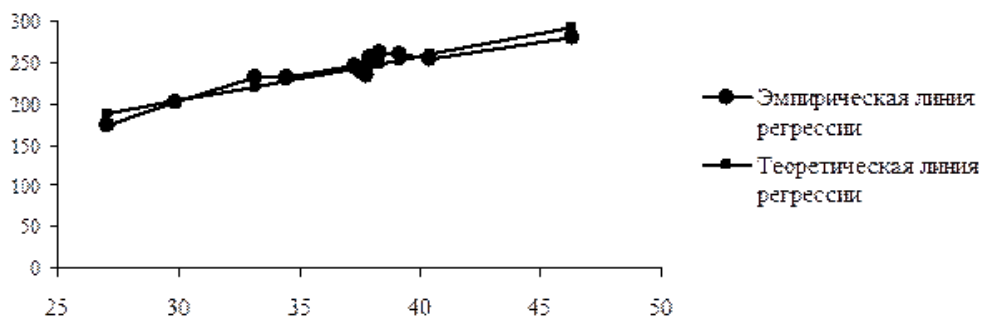


Рис. 16. График эмпирической и теоретической линий регрессии

Из рис. 16 видно, что небольшие различия между эмпирической и теоретической линиями регрессии существуют, поэтому необходимо *оценить существенность* коэффициента регрессии и уравнения связи, для чего определяют среднюю ошибку параметров уравнения регрессии и сравнивают их с этой ошибкой.

Расчет ошибок параметров уравнения регрессии основан на использовании остаточной дисперсии, характеризующей расхождение (отклонение) между эмпирическими и теоретическими значениями резульативного признака. Для линейного уравнения регрессии ( $\hat{y}_x = a_0 + a_1x$ ) средние ошибки параметров  $a_1$  и  $a_2$  определяются по формулам, соответственно:

$$\mu_{a_1} = \frac{\sigma_{ост}}{\sqrt{n-2}},$$

$$\mu_{a_2} = \frac{\sigma_{ост}}{\sigma_x \sqrt{n-2}},$$

$$\sigma_{ост} = \sqrt{\frac{\sum (y - \hat{y}_x)^2}{n}}$$

Значимость параметров проверяется путем сопоставления его значения со средней ошибкой. Обозначим это соотношение как  $t$ :

$$t_{a_i} = \frac{a_i}{\mu_{a_i}},$$

При большом числе наблюдений ( $n > 30$ ) параметр  $a_i$  считается значимым, если  $t_{a_i} > 3$ .

Если выборка малая ( $n < 30$ ), то значимость параметра  $a_i$  проверяется путем сравнения с табличным значения  $t$  - критерия Стьюдента при числе степеней свободы  $\nu = n - 2$  и заданном уровне значимости  $\alpha$ . Если рассчитанное по формуле значение больше табличного, то параметр считается значимым.

В нашем примере по формуле:

$$\sigma_{ост} = \sqrt{\frac{1121,795}{12}} = 9,669.$$

Находим среднюю ошибку параметра  $a_0$  по формуле :

$$\mu_{a_0} = \frac{9,669}{\sqrt{12-2}} = 3,06.$$

Теперь находим среднюю ошибку параметра  $a_1$  по формуле:

$$\mu_{a_1} = \frac{9,669}{4,784\sqrt{12-2}} = 0,639.$$

Теперь по формуле для параметра  $a_0$ :

$$t_{a_0} = \frac{40,767}{3,06} = 13,3.$$

И по той же формуле для параметра  $a_1$ :

$$t_{a_1} = \frac{5,407}{0,639} = 8,46.$$

Так как выборка малая, то, задавшись стандартной значимостью  $\alpha = 0,05$  находим табличное значение  $t_{\alpha} = 2,23$ , которое значительно меньше полученных значений 13,3 и 8,46, что свидетельствует о значимости обоих параметров уравнения регрессии.

Наряду с проверкой значимости отдельных параметров осуществляется *проверка значимости уравнения регрессии* в целом или, что-то - же самое, проверка адекватности модели с помощью критерия Фишера. Данный метод уже использовался нами для проверки



адекватности уравнения тренда в предыдущей теме, поэтому, воспользовавшись формулой

$$F_p = \frac{(12 - 2)8027,945}{(2 - 1)1121,795} = 71,56$$

(102) в нашем примере, получим:

Сравнивая расчетное значение критерия Фишера  $F_p = 71,56$  с табличным  $F_m = 4,96$ , определяемое при числе степеней свободы  $\nu_1 = k - 1 = 2 - 1 = 1$  и  $\nu_2 = n - k = 12 - 2 = 10$  (то есть 1 - й столбец и 10 - я строка) и стандартном уровне значимости  $\alpha = 0,05$ , можно сделать вывод, что уравнение регрессии значимо.

**8. Коэффициент эластичности** показывает, на сколько процентов изменяется в среднем результативный признак  $y$  при изменении факторного признака  $x$  на 1%. Он рассчитывается на основе уравнения регрессии:

$$\varepsilon = \frac{\partial y_x}{\partial x} \frac{x}{y_x},$$

$$\frac{\partial y_x}{\partial x}$$

где  $\frac{\partial y_x}{\partial x}$  – первая производная уравнения регрессии  $y$  по  $x$ .

Коэффициент эластичности – величина переменная, то есть изменяется с изменением значений фактора  $x$ . Так, для линейной зависимости  $y_x = a_0 + a_1 x$ :

$$\varepsilon = a_1 \frac{x}{a_0 + a_1 x}$$

Применительно к рассмотренному уравнению регрессии, выражающему зависимость величины таможенных платежей в федеральный бюджет от величины стоимостного внешнеторгового оборота ( $y_x = 40,767 + 5,407x$ ), коэффициент эластичности по формуле:

$$\varepsilon = \frac{5,407x}{40,767 + 5,407x}$$

Подставляя в данное выражение разные значения  $x$ , получаем и разные значения  $\varepsilon$ . Так,

$$\varepsilon = \frac{5,407 * 40}{40,767 + 5,407 * 40} = 0,84,$$

например, при  $x = 40$  коэффициент эластичности

$$\varepsilon = \frac{5,407 * 50}{40,767 + 5,407 * 50} = 0,87$$

соответственно и т.д. Это значит, что при увеличении внешнеторгового товарооборота  $x$  с 40 до 40,4 млрд.долл. (то есть на 1%), величина таможенных платежей возрастет в среднем на 0,84% прежнего уровня; при увеличении  $x$  с 50 до 50,5 млрд.долл. (то есть на 1%)  $y$  возрастет на 0,87% и т.д.

**9. Теоретическое корреляционное отношение** как универсальный *показатель тесноты связи*. Измерить тесноту связи между коррелируемыми величинами – значит определить, насколько вариация результативного признака обусловлена вариацией факторного (факторных) признака. Ранее были рассмотрены показатели, с помощью которых можно выявить наличие корреляционной связи между двумя признаками  $x$  и  $y$  и измерить тесноту этой связи. Наряду с ними существует универсальный показатель – *корреляционное отношение* (или коэффициент корреляции по Пирсону), применимое ко всем случаям корреляционной зависимости независимо от формы этой связи. Следует различать эмпирическое и теоретическое корреляционное отношение. *Эмпирическое корреляционное отношение* рассчитывается на основе правила сложения дисперсий как корень квадратный из отношения межгрупповой дисперсии к общей дисперсии, то есть

$$\eta_{\text{эмпи}} = \sqrt{\frac{\delta^2}{\sigma^2}}$$

Теоретическое корреляционное отношение  $\eta_{теор}$  определяется на основе выравненных (теоретических) значений результативного признака  $\bar{Y}_x$ , рассчитанных по уравнению регрессии.  $\eta_{теор}$  представляет собой относительную величину, получаемую в результате сравнения среднего квадратического отклонения в ряду теоретических значений результативного признака со средним квадратическим отклонением в ряду эмпирических значений. Если обозначить дисперсию эмпирического ряда игреков через  $\sigma_y^2$ , а теоретического ряда –  $\delta^2$ , то каждая из них выразится формулами

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

Сравнивая вторую дисперсию с первой, получим *теоретический коэффициент детерминации*:

$$\eta_{теор}^2 = \frac{\delta^2}{\sigma_y^2} = \frac{\sum (\bar{y}_x - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

который показывает, какую долю в общей дисперсии результативного признака занимает дисперсия, выражающая влияние вариации фактора  $x$  на вариацию  $y$ . Извлекая корень квадратный из коэффициента детерминации, получаем *теоретическое корреляционное отношение*

$$\eta_{теор} = \sqrt{\frac{\delta^2}{\sigma_y^2}} = \sqrt{\frac{\sum (\bar{y}_x - \bar{y})^2}{\sum (y_i - \bar{y})^2}}$$

Оно может находиться в пределах от 0 до 1, чем ближе его значение к 1, тем теснее связь между вариацией  $y$  и  $x$ . Для оценки тесноты связи обычно применяется шкала Чэддока (табл. 22). Корреляционное отношение применимо как для парной, так и для множественной корреляции независимо от формы связи. В этом смысле его можно назвать универсальным показателем тесноты связи. При линейной зависимости  $\eta_{теор} \equiv r$ .

Покажем расчет  $\eta_{теор}$  на условном примере. Исходные данные и расчет дополнительных показателей приведен в табл. 23.

Таблица 23.

**Исходные данные и вспомогательные расчеты для нахождения теоретического корреляционного отношения**

| Внесено удобрений, ц/га | Урожайность |                                     | $y - \bar{y}$ | $(y - \bar{y})^2$ | $\bar{y}_x - \bar{y}$ | $(\bar{y}_x - \bar{y})^2$ | $y - \bar{y}_x$ | $(y - \bar{y}_x)^2$ |
|-------------------------|-------------|-------------------------------------|---------------|-------------------|-----------------------|---------------------------|-----------------|---------------------|
|                         | фактическая | рассчитанная по уравнению регрессии |               |                   |                       |                           |                 |                     |
| $x$                     | $y$         | $\bar{y}_x$                         |               |                   |                       |                           |                 |                     |
| 1                       | 16          | 16,2                                | -4            | 16                | -3,8                  | 14,44                     | -0,2            | 0,04                |
| 2                       | 19          | 18,5                                | -1            | 1                 | -1,5                  | 2,25                      | 0,5             | 0,25                |
| 3                       | 20          | 20,4                                | 0             | 0                 | 0,4                   | 0,16                      | -0,4            | 0,16                |
| 4                       | 22          | 21,9                                | 2             | 4                 | 1,9                   | 3,61                      | 0,1             | 0,01                |
| 5                       | 23          | 23                                  | 3             | 9                 | 3                     | 9                         | 0               | 0                   |
| $\Sigma$                | 100         | 100                                 |               | 30                |                       | 29,46                     |                 | 0,46                |

$$\bar{y} = \frac{\sum y}{n} = \frac{100}{5} = 20$$

В данном примере общая средняя урожайность: (ц/га).

Общая дисперсия:  $\sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n} = 30/5 = 6,$

факторная дисперсия:  $\delta^2 = \frac{\sum (\bar{y}_x - \bar{y})^2}{n} = 29,46/5 = 5,892.$

$$r_{теор} = \sqrt{\frac{\delta^2}{\sigma_y^2}} = \sqrt{\frac{5,892}{6}} = 0,99.$$

Отсюда теоретическое корреляционное отношение:  $r_{теор} = 0,99$ . Данное значение характеризует очень тесную зависимость изменения урожайности от изменения количества внесенных удобрений. В нашем примере незначительные расхождения ( $30 \neq 29,46 + 0,46$  – это правило сложения дисперсий) объясняются округлением значений параметров уравнения регрессии и самих  $\bar{y}_x$ .

*Коэффициенты корреляции рангов* – это менее точные, но более простые по расчету непараметрические показатели для измерения тесноты связи между двумя коррелируемыми признаками. К ним относятся коэффициенты Спирмэна ( $\rho$ ) и Кендэла ( $\tau$ ), основанные на корреляции не самих значений коррелируемых признаков, а их *рангов* – порядковых номеров, присваиваемых каждому индивидуальному значению  $x$  и  $y$  (отдельно) в ранжированном ряду. Оба признака необходимо ранжировать (нумеровать) в одном и том же порядке: от меньших значений к большим и наоборот. Если встречается несколько значений  $x$  (или  $y$ ), то каждому из них присваивается ранг, равный частному от деления суммы рангов (мест в ряду), приходящихся на эти значения, на число равных значений. Ранги признаков  $x$  и  $y$  обозначают символами  $R_x$  и  $R_y$  (иногда  $N_x$  и  $N_y$ ). Суждение о связи между изменениями значений  $x$  и  $y$  основано на сравнении поведения рангов по двум признакам параллельно. Если у каждой пары  $x$  и  $y$  ранги совпадают, это характеризует максимально тесную связь. Если же наблюдается полная противоположность рангов, то есть в одном ряду ранги возрастают от 1 до  $n$ , а в другом – убывают от  $n$  до 1, это максимально возможная обратная связь. Подходы для оценки тесноты связи у Спирмэна и Кендэла несколько различаются. Для расчета *коэффициента Спирмэна* значения признаков  $x$  и  $y$  нумеруют (отдельно) в порядке возрастания от 1 до  $n$ , то есть им присваивают определенный ранг ( $R_x$  и  $R_y$ ) – порядковый номер в ранжированном ряду. Затем для каждой пары рангов находят их разность (обозначается как  $d = R_x - R_y$ ), и квадраты этой разности суммируют.

$$\rho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \sum d^2}{n^3 - n},$$

где  $d$  – разность рангов  $x$  и  $y$ ;

$n$  – число наблюдаемых пар значений  $x$  и  $y$ .

Коэффициент  $\rho$  может принимать значения от 0 до  $\pm 1$ . Следует иметь в виду, что, поскольку коэффициент Спирмэна учитывает разность только рангов, а не самих значений  $x$  и  $y$ , он менее точен по сравнению с линейным коэффициентом. Поэтому его крайние значения (1 или 0) нельзя безоговорочно расценивать как свидетельство функциональной связи или полного отсутствия зависимости между  $x$  и  $y$ . Во всех других случаях, то есть когда  $\rho$  не принимает крайних значений, он довольно близок к  $r$ .

Эта формула применима строго теоретически только тогда, когда отдельные значения  $x$  (и  $y$ ), следовательно, и их ранги не повторяются. Для случая повторяющихся (связанных) рангов есть другая, более сложная формула, скорректированная на число повторяющихся рангов. Однако опыт показывает, что результаты расчетов по скорректированной формуле для связанных рангов мало отличаются от результатов, полученных по формуле для неповторяющихся рангов. Поэтому на практике эта формула успешно применяется как для неповторяющихся, так и для повторяющихся рангов.

Коэффициент корреляции рангов Кендэла  $\tau$  строится несколько по-другому, хотя его расчет также начинается с ранжирования значений признаков  $x$  и  $y$ . Ранги  $x$  ( $R_x$ ) располагают строго в порядке возрастания параллельно записывают соответствующее каждому  $R_x$  значение  $R_y$ . Поскольку  $R_x$  записаны строго по возрастанию, то ставится задача определить меру соответствия последовательности  $R_y$  «правильному» следованию  $R_x$ . При этом для каждого  $R_y$  последовательно определяют число следующих за ним рангов, превышающих его значение, и число рангов, меньших по значению. Первые («правильное» следование) учитываются как баллы со знаком «+», и их сумма обозначается буквой  $P$ . Вторые («неправильное» следование) учитываются как баллы со знаком «-», и их сумма обозначается буквой  $Q$ . Очевидно, что максимальное значение  $P$  достигается в том случае, если ранги  $y$  ( $R_y$ ) совпадают с рангами  $x$  ( $R_x$ ) и в каждом ряду представляют ряд натуральных чисел от 1 до  $n$ . Тогда после первой пары значений  $R_x = 1$  и  $R_y = 1$  число превышения данных значений рангов составит  $(n - 1)$ , после второй пары, где  $R_x = 2$  и  $R_y = 2$ , соответственно  $(n - 2)$  и т.д. Таким образом, если ранги  $x$  и  $y$  совпадают и число пар рангов равно  $n$ , то

$$P_{\max} = (n-1) + (n-2) + \dots + 3 + 2 + 1 = \frac{n(n-1)}{2}$$

Если же последовательность рангов  $x$  и  $y$  имеет обратную тенденцию по отношению к последовательности рангов  $x$ , то  $Q$  будет такое же максимальное значение по модулю:

$$|Q_{\max}| = \frac{n(n-1)}{2}$$

Если же ранги  $y$  не совпадают с рангами  $x$ , то суммируются все положительные и отрицательные баллы ( $S = P + Q$ ); отношение этой суммы  $S$  к максимальному значению одного из слагаемых и представляет собой коэффициент корреляции рангов Кендэла  $\tau$ , то есть:

$$\tau = \frac{S}{n(n-1)/2} = \frac{2S}{n(n-1)}$$

Формула коэффициента корреляции рангов Кендэла применяется для случаев, когда отдельные значения признака (как  $x$ , так и  $y$ ) не повторяются и, следовательно, их ранги не объединены. Если же встречается несколько одинаковых значений  $x$  (или  $y$ ), то есть ранги повторяются, становятся *связанными*, коэффициент корреляции рангов Кендэла определяется по формуле:

$$\tau = \frac{S}{\sqrt{\left(\frac{n(n-1)}{2} - U_x\right) \left(\frac{n(n-1)}{2} - U_y\right)}}$$

где  $S$  – фактическая общая сумма баллов при оценке +1 каждой пары рангов с одинаковым порядком изменения и -1 каждой пары рангов с обратным порядком изменения;

$U_x = U_y = 0,5 \sum t(t-1)$  – число баллов, корректирующих (уменьшающих) максимальную сумму баллов за счет повторений (объединений)  $t$  рангов в каждом ряду.

Отметим, что случаи следования одинаковых повторяющихся рангов (в любом ряду) оцениваются баллом 0, то есть они не учитываются при расчете ни со знаком «+», ни со знаком «-».

Преимущества ранговых коэффициентов корреляции Спирмэна и Кендэла: они легко вычисляются, с их помощью можно изучать и измерять связь не только между количественными, но и между качественными (описательными) признаками, ранжированными определенным образом. Кроме того, при использовании ранговых коэффициентов корреляции не требуется знать форму связи изучаемых явлений.

Если число ранжируемых признаков (факторов) больше двух, то для измерения тесноты связи между ними можно использовать предложенный М. Кендэлом и Б. Смитом коэффициент конкордации (множественный коэффициент ранговой корреляции):

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

где  $S$  — сумма квадратов отклонений суммы  $m$  рангов от их средней величины;  
 $m$  — число ранжируемых признаков;  
 $n$  — число ранжируемых единиц (число наблюдений).

Формула применяется для случая, когда ранги по каждому признаку не повторяются. Если же есть связанные ранги, то коэффициент конкордации рассчитывается с учетом числа таких повторяющихся (связанных) рангов по каждому фактору:

$$W = \frac{12S}{m^2(n^2 - n) - m \sum_{i=1}^m (t_i^2 - t_i)},$$

где  $t$  — число одинаковых рангов по каждому признаку.

Коэффициент конкордации  $W$  может принимать значения от 0 до 1. Однако, необходимо проверить его на существенность (значимость) с помощью критерия  $\chi^2$  при отсутствии связанных рангов по формуле,

$$\chi^2 = \frac{12S}{mn(n-1)},$$

а при их наличии — по формуле:

$$\chi^2 = 12S \left( \frac{mn(n-1)}{n-1} - \frac{\sum_{i=1}^m (t_i^2 - t_i)}{n-1} \right).$$

Фактическое значение  $\chi^2$  сравнивается с табличным, соответствующим принятому уровню значимости  $\alpha$  (0,05 или 0,01) и числу степеней свободы  $\nu = n - 1$ . Если  $\chi^2_{\text{факт}} > \chi^2_{\text{табл}}$ , то  $W$  — существенен (значим).

Коэффициент конкордации особенно часто используется в экспертных оценках, например, для того, чтобы определить степень согласованности мнений экспертов о важности того или иного оцениваемого показателя или составить рейтинг отдельных единиц по какому-либо признаку. В формулах в этих случаях  $t$  означает число экспертов, а  $n$  — число ранжируемых единиц (или признаков).

Во многих исследованиях приходится изучать динамику нескольких показателей одновременно, то есть рассматривать параллельно несколько рядов динамики. В этом случае возникает необходимость измерить зависимость между ними, вернее, определить, насколько изменения уровней одного ряда зависят от изменения уровней другого ряда. Эта задача решается путем коррелирования рядов динамики.

Однако при этом возникает следующая проблема: если показатели ряда  $x$  и ряда  $y$  рассматривать как функцию времени, то есть  $x = f(t)$  и  $y = f(t)$ , то при однонаправленности их трендов можно получить большое значение коэффициента корреляции между  $x$  и  $y$  даже тогда, когда они независимы, именно в силу однонаправленности их изменения.

Поэтому, прежде чем коррелировать ряды динамики, необходимо установить путем логического (качественного) анализа, возможна ли связь между исследуемыми показателями  $x$  и  $y$ . Кроме того, одно из условий корреляции — независимость отдельных значений переменных множества  $x$ , так же как и множества  $y$ . Для рядов динамики, это равнозначно отсутствию *автокорреляции* между уровнями ряда, то есть отсутствию зависимости между последовательными (соседними) уровнями ряда динамики. Другими словами, прежде чем коррелировать ряды динамики, необходимо проверить каждый ряд на автокорреляцию.

Если исходные фактические уровни ряда, относящиеся к определенному моменту (периоду) времени  $t$ , обозначить через  $y_t$ , то сдвинутые на один момент (период) уровни обозначают  $y_{t-1}$ . Тогда, подставив в формулу коэффициента корреляции значения  $y_t$  и  $y_{t-1}$ , получим формулу:

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

а поскольку  $\bar{y}_t \cong \bar{y}_{t-1}$  и  $\sigma_{y_t} \cong \sigma_{y_{t-1}}$ , получим следующие формулы для расчета коэффициента автокорреляции:

$$r_x = \frac{\overline{y_t y_{t-1}} - (\bar{y}_t)^2}{\sigma_{y_t}^2}$$

$$r_x = \frac{\sum y_t y_{t-1} - n(\bar{y}_t)^2}{\sum y_t^2 - n(\bar{y}_t)^2}$$

Или

Сдвинутый (укороченный) ряд условно дополняют, принимая  $y_1 = y_n$  (чтобы сдвинутый ряд не укорачивался и чтобы средний уровень и дисперсия исходного и сдвинутого рядов были одинаковы).

Найденное по формуле (155) или (156) значение коэффициента автокорреляции само по себе еще не говорит о наличии или отсутствии автокорреляции. Его нужно сравнить с критическим.

Существуют специальные таблицы, в которых для разного числа членов ряда  $n$  и разных уровней значимости  $\alpha$  определено критическое значение коэффициента автокорреляции: если найденное по формулам значение окажется меньше критического, то автокорреляция отсутствует.

В нашем примере про внешнеторговый оборот и таможенные платежи проверим оба эти ряда динамики на автокорреляцию с помощью формулы, для чего построим табл. 24.

Таблица 24.

**Вспомогательные расчеты для проверки на автокорреляцию**

| Месяц | $x_t$   | $x_{t-1}$ | $x_t x_{t-1}$ | $x_t^2$   | $y_t$    | $y_{t-1}$ | $y_t y_{t-1}$ | $y_t^2$    |
|-------|---------|-----------|---------------|-----------|----------|-----------|---------------|------------|
| 1     | 27,068  | 46,298    | 1253,194      | 732,677   | 172,170  | 278,870   | 48013,048     | 29642,509  |
| 2     | 29,889  | 27,068    | 809,035       | 893,352   | 200,900  | 172,170   | 34588,953     | 40360,810  |
| 3     | 34,444  | 29,889    | 1029,497      | 1186,389  | 231,830  | 200,900   | 46574,647     | 53745,149  |
| 4     | 33,158  | 34,444    | 1142,094      | 1099,453  | 232,100  | 231,830   | 53807,743     | 53870,410  |
| 5     | 37,755  | 33,158    | 1251,880      | 1425,440  | 233,400  | 232,100   | 54172,140     | 54475,560  |
| 6     | 37,554  | 37,755    | 1417,851      | 1410,303  | 236,990  | 233,400   | 55313,466     | 56164,260  |
| 7     | 37,299  | 37,554    | 1400,727      | 1391,215  | 246,530  | 236,990   | 58425,145     | 60777,041  |
| 8     | 40,370  | 37,299    | 1505,761      | 1629,737  | 253,620  | 246,530   | 62524,939     | 64323,104  |
| 9     | 37,909  | 40,370    | 1530,386      | 1437,092  | 256,430  | 253,620   | 65035,777     | 65756,345  |
| 10    | 38,348  | 37,909    | 1453,734      | 1470,569  | 261,890  | 256,430   | 67156,453     | 68586,372  |
| 11    | 39,137  | 38,348    | 1500,826      | 1531,705  | 259,360  | 261,890   | 67923,790     | 67267,610  |
| 12    | 46,298  | 39,137    | 1811,965      | 2143,505  | 278,870  | 259,360   | 72327,723     | 77768,477  |
| Итого | 439,229 | 439,229   | 16106,951     | 16351,437 | 2864,090 | 2864,090  | 685863,823    | 692737,647 |

Теперь по формуле (155) для ряда  $x$ :  $r_x = \frac{16106,951 - 12 * 36,602^2}{16351,437 - 12 * 36,602^2} = 0,111$ .

$$\frac{685863,823 - 12 * 238,674^2}{692737,647 - 12 * 238,674^2} = 0,249.$$

Аналогично по формуле (155) для ряда  $y$ :  $ra = \frac{685863,823 - 12 * 238,674^2}{692737,647 - 12 * 238,674^2} = 0,249$ .

По таблице определяем критическое (предельное) значение коэффициента корреляции для числа уровней  $n = 12$  и уровне значимости  $\alpha = 0,05$ . Оно равно 0,348. Оба рассчитанных значения оказались меньше критического, значит автокорреляция между уровнями в обоих рядах динамики отсутствует, следовательно, можно коррелировать уровни  $x$  и  $y$ .

*Исключение автокорреляции в рядах динамики.* Если между уровнями ряда (при коррелировании рядов динамики) существует автокорреляция, она должна быть устранена. Есть несколько способов исключения автокорреляции в рядах динамики. Наиболее простой – *коррелирование отклонений от выравненных уровней.* Для этого каждый ряд динамики выравнивают по определенной для него аналитической формуле (то есть находят  $\dot{x}_t$  и  $\dot{y}_t$ ), затем из эмпирических уровней вычитают выравненные (то есть находят остаточные величины, не описываемые уравнением тренда:  $d_x = x - \dot{x}_t$  и  $d_y = y - \dot{y}_t$ ). Так как остаточные величины могут содержать автокорреляцию (например, в случае недостаточно точно подобранного уравнения тренда), необходимо убедиться, что между ними автокорреляция отсутствует. Лишь после этого можно определять тесноту связи между  $dx$  и  $dy$ . Формулу коэффициента корреляции между остаточными величинами можно записать в следующем виде:

$$r = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}}$$

Метод корреляционных таблиц применим не только к количественным, но и к описательным (качественным) признакам, взаимосвязи между которыми часто приходится изучать при проведении различных социологических исследований путем опросов или анкетирования. В этом случае такие таблицы называют *таблицами сопряженности*. Они могут иметь различную размерность. Простейшая размерность – 2 x 2 (таблица «четырёх полей»), когда по альтернативному признаку («да» – «нет», «хорошо» – «плохо» и т.д.) выделяются две группы. В табл. 25 приведены условные данные о распределении 500 опрошенных человек по двум показателям: наличие (отсутствии) у них прививки против гриппа и факт заболевания (незаболевания) гриппом во время его эпидемии.

**Таблица 25.**

**Распределение 500 опрошенных человек**

| Группа лиц            | Число лиц          |                       |       |
|-----------------------|--------------------|-----------------------|-------|
|                       | заболевших гриппом | не заболевших гриппом | Итого |
| Сделавших прививку    | 30 ( <i>a</i> )    | 270 ( <i>b</i> )      | 300   |
| Не сделавших прививку | 120 ( <i>c</i> )   | 80 ( <i>d</i> )       | 200   |
| Итого                 | 150                | 350                   | 500   |

Нетрудно заметить, что среди сделавших прививку подавляющее большинство (270 из 300, или 90%) не заболели гриппом, а среди не сделавших большая часть заболела (120 из 200, или 60%). Таким образом, можно предположить, что прививка положительно влияет на предупреждение заболевания; другими словами, можно предположить, что распределение в таблице (*a*, *b*, *c*, *d*) не случайно и существует стохастическая зависимость между группировочными признаками. Однако выводы о зависимости, сделанные «на глаз», часто могут быть ненадежными (ошибочными), поэтому они должны подкрепляться определенными статистическими критериями, например, *критерием Пирсона  $\chi^2$* . Он позволяет судить о случайности (или неслучайности) распределения в таблицах взаимной сопряженности,

следовательно, и об отсутствии или наличии зависимости между признаками группировки в таблице. Чтобы воспользоваться критерием Пирсона  $\chi^2$ , в таблице взаимной сопряженности наряду с эмпирическими частотами записывают теоретические частоты, рассчитываемые исходя из предположения, что распределение внутри таблицы случайно и, следовательно, зависимость между признаками группировки отсутствует. То есть считается, что распределение частот в каждой строке (столбце) таблицы пропорционально распределению частот в итоговой строке (столбце). Поэтому теоретические частоты по строкам (столбцам) рассчитывают пропорционально распределению единиц в итоговой строке (столбце).

Так, в нашем примере в итоговой строке число заболевших 150 из 500, то есть их доля – 30%, а доля не заболевших – 70%. Следовательно, теоретические частоты в первой строке для заболевших составят 30% от 300, то есть  $0,3 \cdot 300 = 90$ , а для не заболевших –  $0,7 \cdot 300 = 210$ . По второй строке произведем аналогичные расчеты и их результаты занесем в таблицу в скобках.

Таблица 26.

**Эмпирические и теоретические частоты**

| Группа   | I (да)   | II (нет)  |
|----------|----------|-----------|
| I (да)   | 30 (90)  | 270 (210) |
| II (нет) | 120 (60) | 80 (140)  |
| $\Sigma$ | 150      | 350       |

На сопоставлении эмпирических и теоретических частот и основан критерий Пирсона  $\chi^2$ , рассчитываемый по формуле:

$$\chi^2 = \frac{(30-90)^2}{90} + \frac{(270-210)^2}{210} + \frac{(120-60)^2}{60} + \frac{(80-140)^2}{140} = 142,83$$

Рассчитанное (фактическое) значение  $\chi^2$  сопоставляют с табличным (критическим), определяемым по таблице для заданного уровня значимости  $\alpha$  и числа степеней свободы  $\nu = (k_1 - 1)(k_2 - 1)$ , где  $k_1$  и  $k_2$  – число групп по одному и второму признакам группировки (число строк и число столбцов в таблице).

В рассматриваемом примере  $\nu = (2-1)(2-1) = 1$ , а приняв уровень значимости  $\alpha=0,01$ , по таблице Приложения 3 находим  $\chi^2_{\text{табл}} = 6,63$ . Поскольку рассчитанное значение  $\chi^2 > \chi^2_{\text{табл}}$ , значит, существует стохастическая зависимость между рассматриваемыми показателями. При независимости признаков частоты теоретического и эмпирического распределений совпадают, а значит  $\chi^2 = 0$ . Чем больше различия между теоретическими и эмпирическими частотами, тем больше значение  $\chi^2$  и вероятность того, что оно превысит критическое табличное значение, допустимое для случайных расхождений. Аналогично рассчитываются теоретические частоты и  $\chi^2$  в таблицах большей размерности.

В корреляционном анализе недостаточно лишь выявить тем или иным методом наличие связи между исследуемыми показателями. Теснота такой связи может быть различной, поэтому весьма важно ее измерить, то есть определить меру связи в каждом конкретном случае. В статистике для этой цели разработан ряд показателей (коэффициентов), используемых как для количественных, так и для качественных признаков.

Для измерения тесноты связи между группировочными признаками в таблицах взаимной сопряженности могут быть использованы такие показатели, как коэффициент ассоциации и контингенции (для «четырёхклеточных таблиц»), а также коэффициенты взаимной сопряженности Пирсона и Чупрова (для таблиц любой размерности).

Применительно к таблице «четырёх полей», частоты которых можно обозначить через  $a$ ,  $b$ ,  $c$ ,  $d$ , коэффициент ассоциации (Д. Юла) выражается формулой:

$$K_{AC} = \frac{ad - bc}{ad + bc}$$



Его существенный недостаток: если в одной из четырех клеток отсутствует частота (то есть, равна 0), то  $|K_{\chi^2}| = 1$ , и тем самым преувеличена мера действительной связи.

Чтобы этого избежать, предлагается (К. Пирсоном) другой показатель – *коэффициент контингенции*:

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

Теснота связи между двумя и более признаками измеряется с помощью *коэффициентов взаимной сопряженности Пирсона* или *Чупрова*, рассчитываемых на основе показателя  $\chi^2$  :

$$K_{\text{П}} = \sqrt{\frac{\chi^2}{\chi^2 + n}}, \text{ (Пирсона)}$$

$$K_{\text{Ч}} = \sqrt{\frac{\chi^2}{n \sqrt{(k_1 - 1)(k_2 - 1)}}} \text{ (Чупрова)}$$

$$K_{\text{П}} = \sqrt{\frac{142,85}{142,85 + 500}} = 0,47$$

В примере . рассчитывать коэффициент Чупрова для таблицы «четырёх полей» не рекомендуется, так как при числе степеней свободы  $\nu = (2-1)(2-1) = 1$  он будет больше коэффициента Пирсона (в нашем примере  $K_{\text{Ч}}=0,54$ ). Для таблиц же большей размерности всегда  $K_{\text{Ч}} < K_{\text{П}}$ .

#### Вопросы для самопроверки:

1. Функциональная связь в теории статистики обозначает...
2. Что обозначает жестко детерминированная связь?
3. Что такое индексный метод изучения функциональных связей?
4. Стохастическая связь – это..
5. В чем смысл понятия «корреляционная связь»?
6. Что обозначают прямая и обратная корреляционная связь?
7. Корреляционно-регрессионный анализ – это...
8. Графический метод предназначен для...
9. Корреляционное поле – это...
10. Что обозначает эмпирическая линия регрессии?
11. Коэффициент корреляции знаков (Фехнера) – это...
12. Что обозначает понятие «подбор уравнения регрессии»?

## Раздел 3. Статистические инструменты макроэкономического анализа

### *Тема 11. Система национальных счетов (СНС) – инструмент макроэкономического анализа*

Система национальных счетов (СНС) - система взаимосвязанных показателей, применяемая для описания и анализа макроэкономических процессов более чем в 150 странах мира с рыночной экономикой. СНС возникла в наиболее развитых странах в связи с потребностью в информации, необходимой для практического принятия мер по регулированию рыночной экономики, формированию экономической политики. Необходимость в системе макроэкономических показателей была осознана еще английским экономистом Уильямом Петти, который впервые в мире осуществил оценку национального дохода своей страны. Первую макроэкономическую модель национальной экономики создал француз Франсуа Кэне, глава школы физиократов.

Попытки разработать систему макроэкономических показателей, позволяющих определить состояние национальной экономики, начали предприниматься в разных странах еще в годы первой мировой войны для оценки военного и экономического потенциала воюющих держав. Однако особенно сильно потребность в системе макроэкономических показателей проявилась в 20-30-х гг. XX в. в период бурного подъема экономик развитых стран с целью прогнозирования дальнейших тенденций экономического развития. На Западе разработка подобной системы началась после Великой депрессии 1929-1933 гг. группой американских ученых, сотрудников Национального бюро экономических исследований. Ряд важных принципов этой системы был сформулирован еще А. Маршаллом, затем Дж. М. Кейнсом; большой вклад внесли английские экономисты Р. Стоун, К. Кларк, Дж. Хикс и американские экономисты С. Кузнец, М. Джилберт, В. Леонтьев и др.

В СССР была создана система показателей и таблиц, называемая балансом народного хозяйства, которая использовалась уже при составлении первого пятилетнего плана развития народного хозяйства (1928-1932 гг.). В 1953 году ООН был опубликован документ под названием «Система национальных счетов и вспомогательных таблиц», который можно рассматривать как первый международно признанный вариант системы макроэкономических показателей. Эта система пересматривалась, и ныне действует вариант 1993 года. В России с конца 80-х года показатель ВВП стал определяться по методике ООН. Переход к СНС был осуществлен в 1993 году.

Переход на СНС объективно необходим также из-за трансформации команд административной экономики в рыночную, неизбежного формирования единого евроазиатского экономического пространства на рыночных принципах, включения стран в интенсивное международное разделение труда. Все интеграционные мировые процессы, так бурно происходящие в мире подразумевают создание единой системы учета. Именно поэтому изучение данной темы является актуальной на современном этапе.

**Сущность и основные классификации системы национальных счетов.** Система национальных счетов представляет собой совокупность статистических макроэкономических показателей, характеризующих величину совокупного продукта (выпуска) и совокупного дохода, позволяющих оценить состояние национальной экономики. Система национальных счетов - это адекватный рыночной экономике национальный учет, завершаемый на макроуровне системой взаимосвязанных статистических показателей, построенный в виде определенного набора счетов и балансовых таблиц, характеризующих результаты экономической деятельности, структуру экономики и важнейшие взаимосвязи в национальном хозяйстве.

СНС отражает идею общего экономического равновесия в стране, когда наблюдается соответствие (равенство) наличных ресурсов и их использование. Система национальных счетов дает ступенчатую картину экономических процессов в стране, в том числе, информацию по стандартному набору счетов для всех секторов экономики. Сектора экономики (предприятия, финансовые учреждения, органы государственного управления, домашние хозяйства, обслуживающие их частные некоммерческие организации) – это совокупность экономических агентов, занятых одним и тем же видом основной деятельности. Они сводятся в сектора по признаку однородности выполняемых функций или однородности производимой продукции. Понятие «сектор» в этом случае несколько отличается от понятия «отрасль». Фирма или предприятие может относиться к разным отраслям, но согласно статистической группировке входит только в один сектор.

В системе национальных счетов учитываются следующие сферы экономической деятельности:

- отрасли материального производства (в том числе и государственного сектора);
- государственное управление;
- некоммерческие предприятия обслуживания населения и домашние хозяйства.

При этом учетчиками национальных счетов различаются такие понятия как экономическая операция, и экономический агент. Здесь экономический агент - это любой хозяйствующий субъект рыночной экономики, включая домашние хозяйства и некоммерческие организации. Экономическая операция в таком случае есть перемещение денежных средств или иных благ от одного экономического субъекта к другому. При учете, разумеется, используется денежное выражение рыночной стоимости перемещенных благ, для обеспечения базы сравнения.

Сам учет проводится в виде двойной записи по типу принятой в бухгалтерском учете. Это обозначает, что каждая операция имеет плательщика и получателя, и поэтому записана один раз как ресурсы и один раз — как использование. Таким образом, все операции характеризуются в реальном и финансовом аспектах. В общем, упрощенно, можно представить расчет национальных счетов как бухгалтерию гигантского предприятия, которым является экономика страны, а цехами в нем будут тогда являться определенные сектора экономики.

В СНС используются две классификации. Первая классификация, используемая в счетах производства, потребления и капиталообразования, является функциональной, и выглядит следующим образом:

- а) Отрасли производства товаров и услуг;
- б) Органы общего государственного управления;
- в) Частные некоммерческие организации, обслуживающие домашние хозяйства;
- г) Домашние хозяйства.

Вторая классификация — институциональная — используется в счетах расходов и доходов и в счетах финансирования капитальных затрат:

- а) нефинансовые предприятия;
- б) финансовые учреждения;
- в) органы общего государственного управления;
- г) частные некоммерческие предприятия, обслуживающие домашние хозяйства;
- д) домашние хозяйства, включая мелкие предприятия.

Вторая классификация, в отличие от первой, не рассматривает функциональные группы в экономике, а рассматривает предприятия как юридические лица. В ней разделены финансовые и нефинансовые предприятия в графе «отрасли производства товаров и услуг»: производственная деятельность домашних хозяйств и мелких производственных предприятий отражается в графе «домашние хозяйства»; и государственные производственные предприятия в ней относятся к графе «общие государственные органы управления», а не к графе «отрасли производства товаров и услуг».

При расчете показателей СНС накладываются определенные ограничения, связанные с исключением избыточности учета:

а) общий итог ресурсов должен быть равен общему итогу использования как по отдельным операциям, так и по отдельным счетам;

б) счета не должны отражать операций внутри одного и того же экономического агента. Однако на практике такие сделки могут учитываться как фиктивные купли и продажи одного и того же агента. Это необходимо для более точного расчета национального продукта, например, потребление фермером крупной части своей продукции;

в) теоретически счета не должны отражать также операций между экономическими агентами одной группы. На практике такие операции иногда отражаются в ресурсах и использовании данной группы агентов на одну и ту же сумму.

Единицей учета при составлении национальных счетов является экономическая операция. Она регистрируется, как правило, в момент ее оплаты. Поэтому подавляющее большинство операций регистрируется не по фактическим выплатам, а по начисленным суммам. Исключение составляет государственный сектор, в котором учитываются объемы фактических поступивших и выплаченных сумм.

Экономическая операция может быть двусторонней или договорной (выплаты ренты, процента, дивидендов с акций) и односторонней или недоговорной, то есть в ней могут участвовать две стороны или одна. Большая часть экономических операций осуществляется на компенсационной основе, то есть одна сторона получает какое-либо благо в обмен на свои средства; но также встречаются и некомпенсационные сделки — трансферты. Экономические операции также делятся на текущие, идущие на финансовое потребление (прямые налоги, штрафы, сборы, подарки и др.) и капитальные, идущие на финансирование капиталовложений и изменяющие уровень национального продукта (возмещение государством убытков от стихийных бедствий, ассигнования на строительство и др.).

Практически все экономические операции рассчитываются на фактической основе, то есть по рыночным ценам, но в СНС предусмотрена возможность рассчитывать их на расчетной, условной основе. Основные виды счетов группируются по конкретным операциям, осуществляемым участниками хозяйственной деятельности. Каждый счет отражает одну из сторон этой деятельности.

#### **Вопросы для самопроверки:**

1. Понятие о СНС как системе взаимосвязанных показателей, предназначенной для характеристики уровня и динамики экономического развития позволяет...

2. Классификация счетов в СНС и принципы построения счетов необходимы для того, чтобы...

3. Построение счета производства в СНС позволяет...

4. Для чего необходимы показатели и счета образования, распределения и использования доходов в СНС?

5. Валовой национальный доход как важнейший показатель в СНС позволяет представить...

6. Что такое чистый национальный доход?

7. Что такое первичные доходы?

8. В чем статистический смысл понятия «национальный располагаемый доход»?

### ***Тема 12. Методы исчисления ВВП и взаимосвязь основных макроэкономических агрегатов СНС***

**Валовой внутренний продукт** является основным макроэкономическим показателем результатов функционирования экономики в статистике большинства стран, а также международных организаций (ООН, ОЭСР, МВФ, Международный банк реконструкции и развития). Это один из основных показателей системы национальных счетов, который

характеризует конечный результат производственной деятельности экономических единиц - резидентов.

Валовой внутренний продукт представляет собой рыночную стоимость всех товаров и услуг, произведенных резидентами за тот или иной период для конечного использования, то есть он включает стоимости произведенных конечных товаров и услуг и не включает стоимость международных товаров и услуг, использованных в произведенном процессе (сырья, материалов, топлива, услуг транспорта, финансовых услуг и т.п.). В то же время наряду с другими показателями результатов экономической деятельности, рассчитываемыми на валовой основе, ВВП включает потребление основного капитала.

ВВП может быть исчислен на каждой стадии воспроизведенного цикла соответствующим методом:

- производственным (на стадии производства товаров и услуг);
- распределительным (на стадии распределения);
- методом конечного использования (на стадии конечного использования).
- Производственным методом ВВП рассчитывается на основе суммы валовой стоимости отраслей или секторов экономики.

Международный методологический стандарт СНС-93 рекомендует использовать следующую систему цен, различающихся методами отражения налогов и субсидий: товары и услуги, использованные на цели конечного и промежуточного (производственного) потребления, а также накопления оценивают в рыночных ценах конечного потребителя, а товары и услуги, произведённые резидентами, - в ценах производителя или в основных ценах. Цена конечного потребления включает все налоги на продукты и импорт и не включает никаких субсидий на продукты и импорт. Цена производителя в отличие от цен конечного потребителя не включают налога на добавленную стоимость.

Основные цены в отличие от цены производителя, не включают никаких налогов на продукты, но включают субсидии на продукты. Взаимосвязь цены конечного потребителя (ЦКП), цены производителя (ЦП) и основной цены (ОЦ) может быть представлена следующим образом:

$$\text{ЦП} = \text{ЦКП} - \text{НДС} - \text{ТТН}$$

где НДС – налог на добавочную стоимость;

ТТН – торгово-транспортная наценка.

$$\text{ОЦ} = \text{ЦП} - \text{НП} + \text{СП}$$

где НП – налоги на продукты, включенные в цену производителя, то есть, налоги на продукты, за исключением НДС;

СП – субсидии на продукты.

Стоимость конечных товаров и услуг, составляющих валовой внутренний продукт, оценивается в рыночных ценах конечного потребления, выпуск товаров и услуг и соответственно валовая добавленная стоимость отраслей и секторов экономики, на основе которой исчисляются ВВП, могут быть выражены в ценах производителя или в основных ценах.

Следовательно, для определения ВВП в рыночных ценах конечного потребителя к сумме НДС отраслей (или секторов) должна быть прибавлена величина чистых налогов на продукты и импорт (то есть, налогов, не включённых в цены, использованные при оценке выпуска товаров и услуг, за вычетом включенных в эти цены субсидий на продукты и импорт).

В тех случаях, когда оценка выпуска товаров и услуг и соответственно валовой добавленной стоимости получена в основных ценах, взаимосвязь между суммой валовой добавленной стоимости отраслей или секторов экономики и ВВП может быть выражена формулой

$$\text{ВВП} = \sum \text{ВДС} + \text{Н} - \text{С},$$

где  $\sum \text{ВДС}$  – сумма валовой добавленной стоимости в основных ценах всех секторов или отраслей экономики;

Н – сумма всех налогов на продукты и импорт;

$C$  – сумма всех субсидий на продукты и импорт.

В случаях, когда выпуск товаров и услуг и добавленная стоимость выражены на ценах производителя, взаимосвязь суммы валовой добавленной стоимости отраслей и секторов экономики и ВВП имеет вид:

$$\text{ВВП} = \sum \text{ВДС} + \text{НДС} + \text{ЧНИ},$$

где  $\sum \text{ВДС}$  - сумма валовой добавленной стоимости в основных ценах всех секторов или отраслей экономики;

НДС – сумма налога на добавленную стоимость;

ЧНИ – чистые налоги на импорт (сумма налогов на импорт за вычетом субсидий на импорт).

Используемый в расчетах валовой добавленной стоимости показатель промежуточного потребления отраслей и секторов в настоящее время не включает косвенно измеряемых услуг финансового посредничества, не распределяемых между их потребителями и относящихся к промежуточному потреблению условной единицы с нулевым выпуском. Поэтому сумма валовой добавленной стоимости отраслей или секторов экономики должна быть уменьшена на величину стоимости этих услуг, а общая формула расчета ВВП на основе валовой добавленной стоимости может быть представлена следующим образом:

$$\text{ВВП} = \sum \text{ВДС} - \text{КИУФП} + \text{ЧНПИ},$$

где КИУФП – косвенно измеренные услуги финансового посредничества;

ЧНПИ – чистые налоги на продукты и импорт.

ВВП производственным методом может быть также рассчитаны на основе данных о выпуске товаров и услуг и о промежуточном потреблении в целом по экономике. Если выпуск товаров и услуг выражен в основных ценах, то формула исчисления ВВП выглядит так:

$$\text{ВВП} = \text{В} - \text{ПП} + \text{Н} - \text{С} = \text{В} - \text{ПП} + \text{ЧНПИ},$$

где  $\text{В}$  – выпуск товаров и услуг по экономике в целом;

$\text{ПП}$  – промежуточное потребление по экономике в целом, включая косвенно измеряемые услуги финансового посредничества;

$\text{Н}$  – сумма всех налогов на продукты и импорт;

$\text{С}$  – сумма всех субсидий на продукты и импорт.

Если оценка выпуска товаров и услуг по экономике в целом получена в ценах производителя, то ВВП вычисляют по формуле:

$$\text{ВВП} = \text{В} - \text{ПП} + \text{НДС} + \text{ЧНИ},$$

где НДС – сумма налога на добавленную стоимость;

ЧНИ – чистые налоги на импорт.

Распределительным методом ВВП исчисляется как сумма первичных доходов, распределенных производственными единицами - резидентами между непосредственными участниками процесса производства товаров и услуг, то есть как сумма оплаты труда наёмных работников-резидентов и нерезидентов (ОТ), выплаченная производителями - резидентами, чистых налогов на производство (ЧН) и импорт (ЧНИ), а также валовой прибыли и валовых смешанных доходов (ВП), оставшихся у производителя после оплаты труда работников и чистых налогов на производство и импорт.

Формула расчета ВВП распределительным методом имеет вид:

$$\text{ВВП} = \text{ОТ} + \text{ЧН} + \text{ЧНИ} + \text{ВП}.$$

Валовая прибыль и валовые смешанные доходы в СНС определяются балансовым методом как разность между ВВП, рассчитанным производственным доходом, и другими видами первичных доходов (оплата труда – ОТ, чистые налоги на производство – ЧН и чистые налоги на импорт - ЧНИ). Поэтому данный метод не используется для определения объема ВВП и основное его значение состоит в информационном обеспечении анализа процесса распределения валовой добавленной стоимости между ее производителями, состава и структуры доходов, доли различных видов первичных доходов в ВВП.

Методом конечного использования ВВП рассчитывается как сумма конечного потребления товаров и услуг (КП) и валового накопления (ВН), с учётом сальдо экспорта и импорта товаров и услуг (Э-И). Формула расчета ВВП при этом будет выглядеть так:

$$\text{ВВП} = \text{КП} + \text{ВН} + (\text{Э-И})$$

ВВП, рассчитанный на стадии конечного использования, равен сумме расходов на приобретение предназначенных для конечного использования (конечного потребления и накопления) товаров и услуг в ценах потребления и чистого экспорта товаров и услуг. Расходы на конечное потребление включают расходы домашних хозяйств, государственных учреждений (бюджетных учреждений) и некоммерческих организаций, обслуживающих домашние хозяйства. В расходах на конечное потребление выделяют расходы на товары, индивидуальные и коллективные услуги.

Валовое накопление включает валовое накопление основного капитала, изменение запасов материальных оборотных средств и чистое приобретение (то есть, приобретение за вычетом выбытия) ценностей.

Исчисление ВВП методом конечного использования позволяет провести анализ основных пропорций валового внутреннего продукта, определить долю стоимости товаров и услуг, предназначенных для удовлетворения нужд конечных потребителей и для увеличения национального богатства страны.

Расчет ВВП на основе разных методов, как правило, приводит к несовпадению его количественных оценок. В статистических справочниках и сборниках несовпадения между оценками валового внутреннего продукта, полученными различными методами, отражаются в специальной позиции «Статистическое расхождение». Несовпадение оценок объясняется тем, что в расчетах валового внутреннего продукта различными методами используют различные источники информации. Обычно отклонение оценок ВВП не превышают 1-2%.

#### **Вопросы для самопроверки:**

1. Что является центральным показателем СНС?
2. Исчисление ВВП производственным методом, как сумма производственных товаров и услуг, позволяет понять...
3. В чем заключаются особенности определения валового регионального продукта?
4. Взаимосвязь основных показателей и макроэкономических агрегатов в национальных счетах позволяет понять...
5. Поясните вопрос «исчисление ВВП в сопоставимых ценах».
6. Методы международной сопоставимости ВВП заключаются в том, что...

## Заключение

Курс теории статистики является основополагающей дисциплиной, в рамках которой формируется умение конкретного применения статистических методов анализа социально – экономических явлений. Статистическое изучение тех или иных явлений предполагает, как обязательное условие, наличие сведений об этих явлениях, поэтому начало статистического исследования сводится к сбору необходимой информации. Научно организованный сбор сведений, заключающийся в регистрации тех или иных фактов, признаков, относящихся к каждой единице изучаемой совокупности, называется статистическим наблюдением. Студенты бакалавриата и магистратуры по направлению подготовки «Реклама и связи с общественностью» при освоении многих дисциплин сталкиваются с необходимостью понимания сути статистического наблюдения и процессов оценивания эффективности рекламы. Методы и средства аналитики рекламной деятельности позволяют объективно оценивать ее эффективность только при наличии конкретных количественных показателей.

Чаще всего в условиях рекламной деятельности используются различные статистические программы, которые позволяют получить разрозненную количественную информацию. Такую информацию, например, как средняя величина (отдельно) и среднее квадратическое отклонение (отдельно). Но по отдельности оба эти показателя для оценки эффективности рекламы почти ни о чем не говорят: их необходимо рассматривать только вместе. То же самое можно сказать и по расчету критерия достоверности различий или критериев корреляции.

В учебнике показаны методы статистического учета количественных данных, которые почти всегда используются в качестве инструментария оценки рекламы. Акцент обращается на общетеоретическое понимание основных категорий и понятий статистики, формы представления статистических данных. Построение ряда распределения. Расчет структурных характеристик ряда распределения и методы выявления основных тенденции (трендов) в рядах динамики, оценка адекватности тренда и прогнозирования. Эти и другие показатели, которые представлены в учебнике, предназначены для того, чтобы количественная информация была качественно систематизирована для объективного прогнозирования ее динамики. Материал данного учебника обращает внимание на суть статистических закономерностей, без понимания которых объективно оценивать эффективность рекламной деятельности не представляется возможным.



## Глоссарий по теории статистики

**Аналитическая статистика** - процедуры оценки характеристик совокупности по данным выборок.

**Абсолютный прирост** – разностное сопоставление двух уровней ряда динамики; показывает, на сколько единиц увеличился или уменьшился уровень по сравнению с базисным.

**Абсолютный размер 1% прироста** - абсолютный показатель, который показывает, какое содержание имеется в 1% прироста, сколько весом 1%.

**Аналитическое выравнивание** динамического ряда проводится при помощи математической формулы, отражающей общую тенденцию ряда.

**Варианты** - отдельные значения признака, которые он принимает в вариационном ряду.

**Вариация** - колеблемость, многообразие, изменчивость значения признака у отдельных единиц совокупности явлений.

**Величина интервала** - разность между верхней и нижней границами интервала.

**Вес индекса** - величина, служащая для целей соизмерения индексируемых величин.

**Виды выборки** определяют конкретный механизм или процедуру отбора единиц из генеральной совокупности. В практике выборочных обследований наибольшее распространение получили следующие виды выборки: собственно-случайная, механическая, типическая, серийная, комбинированная.

**Вторичная группировка** - операция по образованию новых групп на основе ранее построенной группировки.

**Выборка многофазная** предполагает сохранение одной и той же единицы отбора на всех этапах его проведения. При этом отобранные на каждой стадии единицы подвергаются обследованию. На каждой последующей стадии отбора программа обследования расширяется.

**Выборочная доля** - удельный вес единиц, обладающих данным признаком в выборочной совокупности. Различия между выборочной долей и средним значением признака в выборке (выборочной средней) определяют особенности вычисления необходимого объема, ошибок выборки, доверительных интервалов и др.

**Выборочное наблюдение** - несплошное наблюдение, при котором статистическому обследованию подвергаются единицы изучаемой совокупности, отобранные случайным способом.

**Выборочная совокупность** - совокупность отобранных для обследования единиц.

**Генеральная совокупность** - совокупность единиц, из которых производится отбор.

**Графический образ** - совокупность точек, линий, фигур, с помощью которых изображаются статистические показатели.

**Группировка** - расчленение множества единиц изучаемой совокупности на группы по определенным, существенным для них признакам.

**Группировка аналитическая** - группировка, выявляющая взаимосвязи между изучаемыми явлениями и их признаками.

**Группировка структурная** - разделение однородной совокупности на группы, характеризующие ее структуру по какому-либо варьирующему признаку.

**Группировочный признак** - признак, по которому производится разбиение единиц совокупности на отдельные группы.

**Группировка типологическая** - разделение исследуемой качественно разнородной совокупности на классы, социально-экономические типы, однородные группы единиц в соответствии с правилами научной группировки.

**Децили** - варианты, делящие ранжированный ряд на десять равных частей.

**Дисперсия** - средний квадрат отклонений индивидуальных значений признака от их средней величины.

**Дисперсия общая** измеряет вариацию признака во всей совокупности под влиянием всех факторов, обусловивших эту вариацию.

**Дисперсия межгрупповая** характеризует систематическую вариацию, то есть различия в величине изучаемого признака, возникающие под действием признака-фактора, положенного в основание группировки.

**Дисперсия внутригрупповая** отражает случайную вариацию, то есть часть вариации, происходящую под влиянием неучтенных факторов и не зависящую от признака-фактора.

**Документальный способ наблюдения** - основан на использовании в качестве источника статистической информации различного рода документов, как правило, учетного характера.

**Единица наблюдения** - составной элемент объекта, являющийся носителем признаков, подлежащих регистрации.

**Единица статистической совокупности** - каждый отдельно взятый элемент статистической совокупности, обладающий определенными признаками; это - первичная ячейка, от которой получают необходимые сведения.

**Единовременное обследование** - сведения даются о количественных характеристиках какого-либо явления или процесса в момент его исследования.

**Задача статистического исследования** - получение обобщающих показателей и выявление закономерностей социально-экономических явлений и процессов в конкретных условиях места и времени.

**Закономерность** - повторяемость, последовательность и порядок изменений в явлениях.

**Индекс** - относительный показатель, который выражает соотношение величин какого-либо явления во времени, в пространстве или сравнение фактических данных с любым эталоном (план, прогноз, норматив и т.д.).

**Индекс индивидуальный** - относительный показатель, отражающий результат сравнения одотоварных явлений.

**Индекс сводный, или общий** - показатель, измеряющий динамику сложного явления, составные части которого непосредственно несоизмеримы.

**Индекс агрегатный** - сложный относительный показатель, который характеризует среднее изменение социально-экономического явления, состоящего из несоизмеримых элементов.

**Индекс-дефлятор** - отношение фактической стоимости продукции отчетного периода к стоимости объема продукции, структура которого аналогична структуре отчетного года, но определенного в ценах базисного года.

**Индекс переменного состава** - индекс, выражающий отношение средних уровней изучаемого явления, относящихся к разным периодам времени.

**Индекс постоянного (фиксированного) состава** - индекс, исчисленный с весами, зафиксированными на уровне одного какого-либо периода, и показывающий изменение только индексируемой величины.

**Индекс средний** - индекс, вычисленный как средняя величина из индивидуальных индексов.

**Индекс структурных сдвигов** - индекс, характеризующий влияние изменения структуры изучаемого явления на динамику среднего уровня этого явления.

**Индексы динамические** - индексы, которые отражают изменение явления во времени.

**Индексы территориальные** - индексы, которые отражают изменение явления в пространстве.

**Индексируемая величина** - признак, изменения которого изучаются.

**Интервал** - значения варьирующего признака, лежащие в определенных границах.

**Интервалы закрытые** - интервалы, у которых обозначены обе границы.

**Интервалы открытые** - интервалы, у которых указана только одна граница.

**Картограмма** – статистический график, при построении которого на схематическую географическую карту наносится штриховка различной частоты, точки или окраска определенной насыщенности, которая показывает сравнительную интенсивность какого-либо показателя в пределах каждой единицы нанесенного на карту территориального деления.

**Картодиаграмма** представляет собой сочетание диаграмм с географической картой.

**Квартили** - значения признака, делящие ранжированную совокупность на четыре равновеликие части.

**Классификация** - систематическое распределение явлений и объектов на определенные группы, классы, разряды на основании их сходства и различия.

**Корреляция** - статистическая зависимость между случайными величинами, которая не имеет строго функционального характера и при которой изменение одной из случайных величин приводит к изменению математического ожидания другой.

**Критический момент (дата)** - день года, час дня, по состоянию на который должна быть проведена регистрация признаков по каждой единице исследуемой совокупности.

**Корреляционное отношение** показывает связь между двумя признаками.

**Коэффициент вариации** - процентное отношение среднего квадратического отклонения к средней величине признака.

**Коэффициент детерминации** показывает, на сколько процентов вариация результативного признака объясняется вариацией 1-го признака (частный) или всех вошедших в модель факторных признаков (множественный).

**Коэффициент корреляции линейный** определяет тесноту и направление связи между двумя коррелируемыми признаками.

**Коэффициент корреляции множественный** отражает связь между результативным и несколькими факторными признаками.

**Коэффициент корреляции частный** показывает степень тесноты связи между двумя признаками при фиксированном значении остальных факторных признаков.

**Коэффициент осцилляции** - процентное отношение размаха вариации к средней величине признака.

**Коэффициент регрессии** показывает, насколько в среднем изменяется значение результативного признака при изменении факторного на единицу собственного измерения.

**Коэффициент эластичности** показывает, на сколько процентов в среднем изменится значение результативного признака при изменении факторного признака на 1%.

**Линейный коэффициент вариации** - процентное отношение среднего линейного отклонения к средней величине признака.

**Матрица** - прямоугольная таблица числовой информации, состоящая из **Т** строк и **П** столбцов.

**Медиана**- структурная средняя, значение признака, приходящееся на середину ранжированной совокупности.

**Механическое сглаживание** - метод нахождения плавных уровней ряда динамики путем использования скользящих средних. Различают метод невзвешенных и взвешенных скользящих средних.

**Мода** - структурная средняя, значение изучаемого признака, повторяющееся с наибольшей частотой.

**Мультиколлинеарность** - наличие тесной зависимости между факторными признаками.

**Общая теория статистики** - отрасль статистической науки о наиболее общих принципах, правилах и законах цифрового освещения социально-экономических явлений.

**Объект наблюдения** - статистическая совокупность, в которой проистекают исследуемые социально-экономические явления и процессы.

**Описательная (дескриптивная) статистика** - получение статистических показателей, с помощью которых обобщаются характеристики только наблюдаемой совокупности. Задача ее заключается в том, чтобы дать сжатую и концентрированную характеристику изучаемого явления.

**Опрос** - способ наблюдения, при котором наблюдаемые сведения получают со слов респондента.

**Отбор групповой** предполагает формирование выборочной совокупности на основе отбора групп единиц из генеральной совокупности.

**Отбор индивидуальный** предполагает формирование выборочной совокупности на основе отбора отдельных единиц генеральной совокупности.

**Отбор комбинированный** представляет собой сочетание индивидуального и группового отбора

**Отбор многоступенчатый** - отбор, при котором из генеральной совокупности сначала извлекаются укрупненные группы, потом - более мелкие и так до тех пор, пока не будут отобраны те единицы, которые подвергаются обследованию.

**Отчетная единица** — субъект, от которого поступают данные о единице наблюдения.

**Отчетность** - основная форма статистического наблюдения, с помощью которой статистические органы в определенные сроки получают от предприятий, учреждений и организаций необходимые данные в виде установленных в законном порядке отчетных документов, скрепляемых подписями лиц, ответственных за их предоставление и достоверность собираемых сведений.

**Ошибка наблюдения** - расхождение между расчетным и действительным значением изучаемых величин.

**Ошибка выборочного наблюдения** - разность между величиной параметра в генеральной совокупности и его величиной, вычисленной по результатам выборочного наблюдения.

**Перепись** - специально организованное наблюдение, повторяющееся, как правило, через равные промежутки времени, с целью получения данных о численности, составе и состоянии объекта статистического наблюдения по ряду признаков.

**Перцентили** - значения признака, делящие ряд на сто частей.

**Предмет статистики** - количественная сторона качественно определенных массовых социально-экономических явлений и процессов, отображаемая посредством статистических показателей.

**Причинно-следственные отношения** - связь явлений и процессов, когда изменение одного из них - причины - ведет к изменению другого - следствия. Социально-экономические явления - это результат одновременного воздействия большого числа причин.

**Подлежащее статистической таблицы** характеризует объект исследования. В нем дается перечень единиц совокупности либо групп исследуемого объекта по существенным признакам.

**Признак** – характерное свойство изучаемого явления; черта или иная особенность единиц совокупности, которая может быть наблюдаема или измерена.

**Признак атрибутивный (описательный)** – признак, который не поддается прямому количественному (числовому) выражению, может быть выражен только словами.

**Признак количественный** – признак, который поддается непосредственному количественному (числовому) выражению.

**Признак результативный** - признак, изменяющийся под действием факторных признаков.

**Признак факторный** - признак, оказывающий влияние на изменение результативного.

**Программа наблюдения** - перечень признаков (или вопросов), подлежащих регистрации в процессе наблюдения.

**Размах вариации** — разность между наибольшим и наименьшим значениями варьирующего признака.

**Ранг** - порядковый номер значения признака, расположенного в порядке возрастания или убывания величин.

**Ранжирование** - процедура упорядочения объектов изучения, которая выполняется на основе предпочтения значений признака в порядке возрастания или убывания.

**Регрессионный анализ** - аналитическое выражение связи, в котором изменение одной величины - результативного признака - обусловлено влиянием одной или нескольких независимых величин (факторов), а множество всех прочих факторов, также оказывающих влияние на зависимую величину, принимается за постоянные и средние значения.

**Регрессия множественная** - модель связи трех и более признаков.

**Регрессия парная** - аналитическое выражение связи двух признаков.

**Ряд динамики** - ряд числовых значений определенного статистического показателя в последовательные моменты или периоды времени.

**Ряд динамики интервальный** - ряд числовых значений определенного статистического показателя, характеризующего размеры изучаемого явления за определенные промежутки (периоды, интервалы) времени.

**Ряд динамики моментный** - ряд числовых значений определенного статистического показателя, характеризующего размеры изучаемого явления на определенные даты, моменты времени.

**Ряд распределения** - упорядоченное распределение единиц совокупности на группы по определенному варьирующему признаку.

**Ряд распределения атрибутивный** - ряд, построенный по качественному признаку.

**Ряд распределения вариационный** - ряд, построенный по количественному признаку.

**Ряд распределения вариационный дискретный** - распределение единиц совокупности по дискретному признаку.

**Ряд распределения вариационный интервальный** - ряд, который отражает непрерывную вариацию признака.

**Сводка** - комплекс последовательных операций по обобщению конкретных единичных факторов для выявления типичных черт и закономерностей, присущих изучаемому явлению в целом.

**Связь корреляционная** - изменение среднего значения результативного признака, которое обуславливается изменением факторных признаков.

**Связь линейная** - статистическая связь между явлениями, выраженная уравнением прямой линии.

**Связь нелинейная** - статистическая связь между социально-экономическими явлениями, аналитически выраженная уравнением кривой линии (параболы, гиперболы и т.д.).

**Связь обратная** - с увеличением или уменьшением значений факторного признака уменьшается или увеличивается значение результативного.

**Связь прямая** - с увеличением или уменьшением значений факторного признака увеличивается или уменьшается значение результативного.

**Связь стохастическая** - связь, которая проявляется не в каждом отдельном случае, а в общем, среднем или большом числе наблюдений.

**Связь функциональная** — связь, при которой определенному значению факторного признака соответствует одно и только одно значение результативного признака.

**Сезонная компонента** ряда динамики - внутригодовые колебания, имеющие более или менее регулярный характер. Их мерой обычно является индекс сезонности.

**Система индексов** - ряд последовательно построенных индексов.

**Система индексов базисных** - ряд последовательно вычисленных индексов одного и того же явления с постоянной базой сравнения.

**Система индексов с переменными весами** - система сводных индексов одного и того же явления, вычисленных с весами, последовательно меняющимися от одного индекса к другому.

**Система индексов с постоянными весами** - система сводных индексов одного и того же явления, вычисленных с весами, не меняющимися при переходе от одного индекса к другому.

**Система индексов цепных** – ряд индексов одного и того же явления, вычисленных с меняющейся от индекса к индексу базой сравнения.

**Система показателей** - совокупность взаимосвязанных показателей, которые отражают состояние и развитие массовых социально-экономических явлений с разных сторон.

**Сказуемое статистической таблицы** - система показателей, которыми характеризуется объект изучения.

**Смыкание рядов динамики** - один из методов приведения несопоставимых рядов к сопоставимым путем прямого пересчета уровней с помощью специальных коэффициентов или относительных величин.

**Среднее линейное отклонение** - средняя арифметическая из абсолютных значений отклонений вариант признака от их средней.

**Среднее квадратическое отклонение** рассчитывается как корень квадратный из дисперсии.

**Срок (период) наблюдения** - время, в течение которого происходит заполнение статистических формуляров.

**Статистика** - общественная наука, имеющая целью сбор, упорядочивание, анализ и сопоставление данных, относящихся к самым разнообразным массовым явлениям.

**Статистическая закономерность** - форма проявления причинной связи, выражающаяся в последовательности, регулярности, повторяемости событий с достаточно высокой степенью вероятности, если причины, порождающие события, не изменяются или изменяются незначительно.

**Статистическая методология** - система приемов, способов и методов, направленных на изучение количественных закономерностей, проявляющихся в структуре, динамике и взаимосвязи социально-экономических явлений.

**Статистическая совокупность** – множество единиц изучаемого явления, объединённых единой качественной основой. Это множество единиц обладает массовостью, однородностью, определенной целостностью, взаимозависимостью состояний отдельных единиц и наличием вариации.

**Статистическая таблица** - способ рационального изложения и обобщения данных о социально-экономических явлениях при помощи цифр, расположенных в определенном порядке.

**Статистическая таблица простая** - таблица, в подлежащем которой дается простой перечень объектов или территориальных единиц.

**Статистическая таблица групповая** содержит группировку единиц совокупности по одному - количественному или атрибутивному - признаку.

**Статистическая таблица комбинационная** содержит группировку единиц совокупности одновременно по двум и более признакам.

**Статистический показатель** - обобщающая количественная характеристика социально-экономических явлений в конкретных условиях места и времени.

**Статистический показатель абсолютный** - показатель в форме абсолютной величины, отражающий физические свойства, временные или стоимостные характеристики социально-экономических процессов и явлений.

**Статистический показатель относительный** - показатель в форме относительной величины, получаемый как результат деления одного абсолютного показателя на другой и отражающий соотношение между количественными характеристиками изучаемых процессов и явлений,

**Статистический показатель средний** - показатель в форме средней величины, представляющий собой обобщенную количественную характеристику признака в статистической совокупности в конкретных условиях места и времени.

**Статистический формуляр** - документ единого образца, содержащий программу и результаты наблюдения.

**Статистическое наблюдение** - массовое, планомерное, научно организованное наблюдение за явлениями социальной и экономической жизни, которое заключается в регистрации признаков, отобранных у каждой единицы совокупности.

**Статистическое наблюдение несплошное** - обследованию подлежит лишь часть единиц изучаемой совокупности.

**Статистическое наблюдение сплошное** - получение информации о всех единицах исследуемой совокупности.

**Статистическая совокупность** - множество единиц, обладающих массовостью, однородностью, определенной целостностью, взаимозависимостью состояний отдельных единиц и наличием вариации.

**Статистическая карта** - графическое изображение статистических данных на схематической географической карте, характеризующих уровень или степень распространения того или иного явления на определенной территории.

**Статистический график** - чертеж, на котором статистические совокупности, характеризующиеся определенными показателями, описываются с помощью условных геометрических образов или знаков.

**Таблица сопряженности** - таблица, которая содержит сводную числовую характеристику изучаемой совокупности по двум и более атрибутивным признакам или комбинации количественных и атрибутивных признаков.

**Текущее наблюдение** - наблюдение, когда изменения в отношении изучаемых явлений фиксируются по мере их наступления.

**Темп роста** - относительный показатель, показывает, сколько процентов составляет уровень данного периода по сравнению с базисным или предыдущим уровнем.

**Темп прироста** - относительный показатель, характеризующий величину прироста (снижения).

**Точность статистического наблюдения** - степень соответствия величин какого-либо показателя, определяемого по материалам статистического наблюдения, действительной его величине.

**Тренд (основная тенденция)** - достаточно плавное и устойчивое изменение уровня явления во времени, более или менее свободное от случайных колебаний. Основную тенденцию можно представить либо аналитически - в виде уравнения (модели) тренда, либо графически.

**Уровень ряда динамики** — абсолютная (относительная, средняя) величина каждого члена динамического ряда.

**Цель наблюдения** - получение достоверной информации для выявления закономерностей развития явлений и процессов.

**Частоты** - выраженные в долях единицы или в процентах к итогу значения изучаемого признака.

**Экспликация** - словесное описание содержания графика.

**Эмпирическое корреляционное отношение** - корень квадратный из эмпирического коэффициента детерминации.

## Тестовые задания по теории статистики (учебно-ознакомительный тест с ответами)

1. Статистика как наука изучает:
  - а) единичные явления
  - б) массовые явления**
  - в) периодические события
  
2. Термин «статистика» происходит от слова:
  - а) статика
  - б) статный
  - в) статус**
  
3. Статистика зародилась и оформилась как самостоятельная учебная дисциплина:
  - а) до новой эры, в Китае и Древнем Риме
  - б) в 17-18 веках, в Европе**
  - в) в 20 веке, в России
  
4. Статистика изучает явления и процессы посредством изучения:
  - а) определенной информации
  - б) статистических показателей**
  - в) признаков различных явлений
  
5. Статистическая совокупность – это:
  - а) множество изучаемых разнородных объектов
  - б) множество единиц изучаемого явления;**
  - в) группа зафиксированных случайных событий
  
6. Основными задачами статистики на современном этапе являются:
  - а) исследование преобразований экономических и социальных процессов в обществе;
  - б) анализ и прогнозирование тенденций развития экономики;
  - в) регламентация и планирование хозяйственных процессов;
  - а) а, в
  - б) а, б**
  - в) б, в
  
7. Статистический показатель дает оценку свойства изучаемого явления:
  - а) количественную**
  - б) качественную
  - в) количественную и качественную
  
8. Основные стадии экономико-статистического исследования включают:
  - а) сбор первичных данных,
  - б) статистическая сводка и группировка данных,
  - в) контроль и управление объектами статистического изучения,
  - г) анализ статистических данных
  - а) а, б, в
  - б) а, в, г
  - в) а, б, г**
  - г) б, в, г



9. Закон больших чисел утверждает, что:

**а) чем больше единиц охвачено статистическим наблюдением, тем лучше проявляется общая закономерность**

б) чем больше единиц охвачено статистическим наблюдением, тем хуже проявляется общая закономерность

в) чем меньше единиц охвачено статистическим наблюдением, тем лучше проявляется общая закономерность

10. Современная организация статистики включает: а) в России - Росстат РФ и его территориальные органы, б) в СНГ - Статистический комитет СНГ, в) в ООН - Статистическая комиссия и статистическое бюро, г) научные исследования в области теории и методологии статистики

а) а, б, г

**б) а, б, в**

в) а, в, г

11. Статистическое наблюдение – это:

а) научная организация регистрации информации

б) оценка и регистрация признаков изучаемой совокупности

**в) работа по сбору массовых первичных данных**

г) обширная программа статистических исследований.

12. Назовите основные организационные формы статистического наблюдения:

**а) перепись и отчетность**

в) разовое наблюдение

г) опрос

13. Перечень показателей (вопросов) статистического наблюдения, цель, метод, вид, единица наблюдения, объект, период статистического наблюдения излагаются:

а) в инструкции по проведению статистического наблюдения

б) в формуляре статистического наблюдения

**в) в программе статистического наблюдения**

14. Назовите виды статистического наблюдения по степени охвата единиц совокупности:

а) анкета

б) непосредственное

**в) сплошное**

г) текущее

15. Назовите виды статистического наблюдения по времени регистрации:

а) текущее, б) единовременное; в) выборочное; г) периодическое; д) сплошное

а) а, в, д

**б) а, б, г**

в) б, г, д

16. Назовите основные виды ошибок регистрации: а) случайные; б) систематические; в) ошибки репрезентативности; г) расчетные

а) а

б) а, б

**в) а, б, в**

г) а, б, в, г

17. Несплошное статистическое наблюдение имеет виды: а) выборочное; б) монографическое; в) метод основного массива; г) ведомственная отчетность

**а) а, б, в**

б) а, б, г

в) б, в, г

18. Организационный план статистического наблюдения регламентирует: а) время и сроки наблюдения; б) подготовительные мероприятия;

в) прием, сдачу и оформление результатов наблюдения; г) методы обработки данных

а) а, б, г

**б) а, б, в**

19. Является ли статистическим наблюдением наблюдения покупателя за качеством товаров или изменением цен на городских рынках?

а) да

**б) нет**

20. Ошибка репрезентативности относится к:

а) сплошному наблюдению

**б) не сплошному выборочному наблюдению**

21. Статистическая сводка - это:

**а) систематизация и подсчет итогов зарегистрированных фактов и данных;**

б) форма представления и развития изучаемых явлений

в) анализ и прогноз зарегистрированных данных

22. Статистическая группировка - это:

а) объединение данных в группы по времени регистрации

**б) расчленение изучаемой совокупности на группы по существенным признакам**

в) образование групп зарегистрированной информации по мере ее поступления.

23. Статистические группировки могут быть: а) типологическими; б) структурными; в) аналитическими; г) комбинированными

а) а

б) а, б

**в) а, б, в**

г) а, б, в, г

24. Группировочные признаки, которыми одни единицы совокупности обладают, а другие - нет, классифицируются как:

а) факторные

б) атрибутивные

**в) альтернативные**

25. К каким группировочным признакам относятся: образование сотрудников, профессия бухгалтера, семейное положение:

**а) к атрибутивным**

б) к количественным

26. Ряд распределения - это:

**а) упорядоченное расположение единиц изучаемой совокупности по группам**

б) ряд значений показателя, расположенных по каким-то правилам

27. К каким группировочным признакам относятся: сумма издержек обращения, объем продаж, стоимость основных фондов

а) к дискретным

**б) к непрерывным**

28. Какие виды статистических таблиц встречаются:

**а) простые и комбинационные**

б) линейные и нелинейные

29. Статистический показатель - это

а) размер изучаемого явления в натуральных единицах измерения

**б) количественная характеристика свойств в единстве с их качественной определенностью**

в) результат измерения свойств изучаемого объекта

30. Статистические показатели могут характеризовать:

а) объемы изучаемых процессов

б) уровни развития изучаемых явлений

в) соотношение между элементами явлений

**г) а, б, в**

31. По способу выражения абсолютные статистические показатели подразделяются на: а) суммарные; б) индивидуальные; в) относительные; г) средние; д) структурные

а) а, д

б) б, в

в) в, г

**г) а, б**

32. В каких единицах выражаются абсолютные статистические показатели?

а) в коэффициентах

б) в натуральных

**в) в трудовых**

33. В каких единицах будет выражаться относительный показатель, если база сравнения принимается за единицу?

а) в процентах

б) в натуральных

**в) в коэффициентах**

34. Относительные показатели динамики с переменной базой сравнения подразделяются на:

**а) цепные**

б) базисные.

35. Сумма всех удельных весов показателя структуры

**а) строго равна 1**

б) больше или равна 1

в) меньше или равна 1

36. Относительные показатели по своему познавательному значению подразделяются на показатели: а) выполнения и сравнения, б) структуры и динамики, в) интенсивности и координации, г) прогнозирования и экстраполяции

а) а, б, г

б) б, в, г

**в) а, б, в**

37. Статистические показатели по сущности изучаемых явлений могут быть:

а) качественными

б) объёмными

**в) а, б**

38. Статистические показатели в зависимости от характера изучаемых явлений могут быть:

а) интервальными

б) моментными

**в) а, б**

39. Исчисление средних величин - это

**а) способ изучения структуры однородных элементов совокупности**

б) прием обобщения индивидуальных значений показателя

в) метод анализа факторов

40. Требуется вычислить средний стаж деятельности работников фирмы: 6,5,4,6,3,1,4,5,4,5. Какую формулу Вы примените?

**а) средняя арифметическая**

б) средняя арифметическая взвешенная

в) средняя гармоническая

41. Средняя геометрическая - это:

**а) корень из произведения индивидуальных показателей**

б) произведение корней из индивидуальных показателей

42. По какой формуле производится вычисление средней величины в интервальном ряду?

**а) средняя арифметическая взвешенная**

б) средняя гармоническая взвешенная

43. Могут ли взвешенные и невзвешенные средние, рассчитанные по одним и тем же данным, совпадать?

а) да

**б) нет**

44. Как изменяется средняя арифметическая, если все веса уменьшить в А раз?

а) уменьшатся

б) увеличатся

**в) не изменится**

45. Как изменится средняя арифметическая, если все значения определенного признака увеличить на число  $A$ ?
- а) уменьшится
  - б) увеличится**
  - в) не изменится
46. Значения признака, повторяющиеся с наибольшей частотой, называется
- а) модой**
  - б) медианой
47. Средняя хронологическая исчисляется
- а) в моментных рядах динамики с равными интервалами**
  - б) в интервальных рядах динамики с равными интервалами
  - в) в интервальных рядах динамики с неравными интервалами
48. Медиана в ряду распределения с четным числом членов ряда равна
- а) полусумме двух крайних членов
  - б) полусумме двух срединных членов**
49. Что понимается в статистике под термином «вариация показателя»?
- а) изменение величины показателя**
  - б) изменение названия показателя
  - в) изменение размерности показателя
50. Укажите показатели вариации
- а) мода и медиана
  - б) сигма и дисперсия**
  - в) темп роста и прироста
51. Показатель дисперсии - это:
- а) квадрат среднего отклонения
  - б) средний квадрат отклонений**
  - в) отклонение среднего квадрата
52. Коэффициент вариации измеряет колеблемость признака
- а) в относительном выражении**
  - б) в абсолютном выражении
53. Среднеквадратическое отклонение характеризует
- а) взаимосвязь данных
  - б) разброс данных**
  - в) динамику данных
54. Размах вариации исчисляется как
- а) разность между максимальным и минимальным значением показателя**
  - б) разность между первым и последним членом ряда распределения
55. Показатели вариации могут быть
- а) простыми и взвешенными
  - б) абсолютными и относительными**
  - в) а) и б)

56. Закон сложения дисперсий характеризует

**а) разброс сгруппированных данных**

б) разброс неупорядоченных данных

57. Средне квадратическое отклонение исчисляется как

а) корень квадратный из медианы

б) корень квадратный из коэффициента вариации

**в) корень квадратный из дисперсии**

58. Кривая закона распределения характеризует

**а) разброс данных в зависимости от уровня показателя**

б) разброс данных в зависимости от времени

59. Выборочный метод в статистических исследованиях используется для:

**а) экономии времени и снижения затрат на проведение статистического исследования**

б) повышения точности прогноза

в) анализа факторов взаимосвязи.

60. Выборочный метод в торговле используется:

а) при анализе ритмичности оптовых поставок

**б) при прогнозировании товарооборота**

в) при разрушающих методах контроля качества товаров

61. Ошибка репрезентативности обусловлена:

**а) самим методом выборочного исследования**

б) большой погрешностью зарегистрированных данных

62. Коэффициент доверия в выборочном методе может принимать значения:

**а) 1, 2, 3**

б) 4, 5, 6

в) 7, 8, 9

63. Выборка может быть: а) случайная, б) механическая, в) типическая, серийная, д) техническая

**а) а, б, в, г**

б) а, б, в, д

в) б, в, г, д

64. Необходимая численность выборочной совокупности определяется:

**а) колеблемостью признака**

б) условиями формирования выборочной совокупности

65. Выборочная совокупность отличается от генеральной:

а) разными единицами измерения наблюдаемых объектов

**б) разным объемом единиц непосредственного наблюдения**

в) разным числом зарегистрированных наблюдений

66. Средняя ошибка выборки:

**а) прямо пропорциональна рассеяности данных**

б) обратно пропорциональна разбросу варьирующего признака

в) никак не зависит от колеблемости данных

67. Повторный отбор отличается от бесповторного тем, что:

а) отбор повторяется, если в процессе выборки произошел сбой

**б) отобранная однажды единица наблюдения возвращается в генеральную совокупность**

в) повторяется несколько раз расчет средней ошибки выборки

68. Малая выборка - это выборка объемом:

а) 4-5 единиц изучаемой совокупности

б) до 50 единиц изучаемой совокупности

**в) до 30 единиц изучаемой совокупности**

69. Ряд динамики характеризует: а) структуру совокупности по какому-то признаку; б) изменение характеристик совокупности во времени; в) определенное значение признака в совокупности; г) величину показателя на определенную дату или за определенный период

а) а, б

**б) б, г**

в) б, в

70. Ряд динамики может состоять: а) из абсолютных суммарных величин; б) из относительных и средних величин;

а) а

б) б

**в) а, б**

71. Ряд динамики, характеризующий уровень развития социально-экономического явления на определенные даты времени, называется:

а) интервальным;

**б) моментным.**

72. Средний уровень интервального ряда динамики определяется как:

**а) средняя арифметическая;**

б) средняя хронологическая.

5. Средний уровень моментного ряда динамики исчисляется как: а) средняя арифметическая взвешенная при равных интервалах между датами; б) при неравных интервалах между датами как средняя хронологическая, в) при равных интервалах между датами как средняя хронологическая;

а) а

б) б

**в) б, в**

73. Абсолютный прирост исчисляется как: а) отношение уровней ряда; б) разность уровней ряда. Темп роста исчисляется как: в) отношение уровней ряда; г) разность уровней ряда;

а) а, в

**б) б, в**

в) а, г

74. Для выявления основной тенденции развития используется: а) метод укрупнения интервалов; б) метод скользящей средней; в) метод аналитического выравнивания; г) метод наименьших квадратов;

- а) а, г
- б) б, г
- в) а, б, г
- г) **а, б, в**

75. Трендом ряда динамики называется:

- а) **основная тенденция;**
- б) устойчивый темп роста.

76. Прогнозирование в статистике - это:

- а) предсказание предполагаемого события в будущем;
- б) **оценка возможной меры изучаемого явления в будущем.**

77. К наиболее простым методам прогнозирования относят:

- а) индексный метод
- б) метод скользящей средней
- в) **метод на основе среднего абсолютного прироста**

78. Статистический индекс - это:

- а) критерий сравнения относительных величин
- б) сравнительная характеристика двух абсолютных величин
- в) **относительная величина сравнения двух показателей**

79. Индексы позволяют соизмерить социально-экономические явления:

- а) в пространстве
- б) во времени
- в) **в пространстве и во времени.**

80. В индексном методе анализа несуммарность цен на разнородные товары преодолевается:

- а) переходом от абсолютных единиц измерения цен к относительной форме
- б) **переходом к стоимостной форме измерения товарной массы**

81. Можно ли утверждать, что индивидуальные индексы по методологии исчисления адекватны темпам роста:

- а) **можно**
- б) нельзя

82. Сводные индексы позволяют получить обобщающую оценку изменения:

- а) **по товарной группе**
- б) одного товара за несколько периодов

83. Может ли в отдельных случаях средний гармонический индекс рассчитываться по средней гармонической невзвешенной:

- а) может;
- б) **не может.**



84. Индексы переменного состава рассчитываются:

- а) по товарной группе
- б) по одному товару**

85. Может ли индекс переменного состава превышать индекс фиксированного состава:

- а) может**
- б) не может

86. Первая индексная мультипликативная модель товарооборота – это:

- а) произведение индекса цен на индекс физического объема товарооборота**
- б) произведение индекса товарооборота в сопоставимых ценах на индекс средней цены постоянного состава
- в) а, б

87. Вторая факторная индексная мультипликативная модель анализа – это:

- а) произведение индекса постоянного состава на индекс структурных сдвигов**
- б) частное от деления индекса переменного состава на индекс структурных сдвигов
- в) а, б

88. Статистическая связь - это:

- а) когда зависимость между факторным и результирующим показателями неизвестна
- б) когда каждому факторному соответствует свой результирующий показатель**
- в) когда каждому факторному соответствует несколько разных значений результирующего показателя

89. Термин корреляция в статистике понимают, как:

- а) связь, зависимость**
- б) отношение, соотношение
- в) функцию, уравнение

90. По направлению связь классифицируется как:

- а) линейная
- б) прямая**
- в) обратная

91. Анализ взаимосвязи в статистике исследует:

- а) тесноту связи
- б) форму связи
- в) а, б**

92. При каком значении коэффициента корреляции связь можно считать умеренной?

- а)  $r = 0,43$**
- б)  $r = 0,71$

93. Для определения тесноты связи двух альтернативных показателей применяют:

- а) коэффициенты ассоциации и контингенции;**
- б) коэффициент Спирмена.

94. Дайте классификацию связей по аналитическому выражению:

- а) обратная
- б) сильная
- в) прямая
- г) **линейная**

95. Какой коэффициент корреляции характеризует связь между Y и X:

- а) **линейный**
- б) частный
- в) множественный

96. При каком значении линейного коэффициента корреляции связь между Y и X можно признать более существенной:

- а)  $r_{yx} = 0,25$
- б)  $r_{yx} = 0,14$
- в)  **$r_{yx} = - 0,57$**

## Литература

1. Березин И.С. Маркетинговый анализ. Рынок. Фирма. Товар. Продвижение/И.С. Березин. – 3-е изд., перераб. и доп. – М.: Вершина, 2007. – 480 с.: ил., табл. – ISBN 978-5-9626-0335-3.
2. Бернадская Ю.С. Основы рекламы: учебник для студентов вузов/Ю.С. Бернадская [и др.]; под ред. Л.М. Дмитриевой. – М.: ЮНИТИ-ДАНА, 2007. – 351 с. – (Серия «Азбука рекламы»). – ISBN 978-5-238-01252-0.
3. Годин А.М. Статистика: Учебник. – М.: Дашков и К., 2015. – 458 с.
4. Задачник по общей теории статистики/Н.Ю. Агафонова, А.Д. Луныков, А.В. Харламов; Саратов. гос. ун-т им. Н.Г. Чернышевского. - Саратов: Изд-во Саратов. ун-та, 2004. – 69 с. – ISBN 5-292-03190-9.
5. 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.
6. Ефимова М.Р. Практикум по общей теории статистики: учебное пособие для вузов / М.Р. Ефимова и др. – М.: Финансы и статистика, 2007. – 368 с.
7. Котлер Ф. Маркетинг менеджмент/Ф. Котлер, К.Л. Келлер; пер. с англ. С. Жильцов [и др.]. – 12-е изд. – СПб.: Питер, 2010. – 816 с.: ил. – (Серия «Классический зарубежный учебник»). – ISBN 978-5-469-00989-4.
8. Кутлалиев А. Эффективность рекламы/А. Кутлалиев, А. Попов. – М.: Эксмо, 2005. – 416 с. – (Профессиональные издания для бизнеса). – ISBN 5-699-10796-7.
9. Методология экономико-статистического исследования рекламной деятельности: автореф. дис. ... д-ра экон. наук: 08.00.12/А.А. Романов; науч. консультант В. С. Мхитарян; Моск. гос. ун-т экономики, статистики и информатики (МЭСИ). - М., 2007. с.
10. 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.
11. 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.
12. Система национальных счетов: учеб. пособие/В.Н. Салин, С.И. Кудряшова. - М.: Финансы и статистика, 2006. - 269 с.
13. Статистика [Электронный ресурс]: учебник – Эл. изд. - Электрон. текстовые дан. (1 файл pdf: 125 с.). - Чвякин В.А., Козилова Л.В. 2023. – Режим доступа: [http://scipro.ru/conf/statistics\\_23.pdf](http://scipro.ru/conf/statistics_23.pdf). Сист. требования: Adobe Reader; экран 10". ISBN 978-5-907607-26-2
14. 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.

### Нормативно-правовые акты

15. Федеральный закон от 29.11.2007 N 282-ФЗ «Об официальном статистическом учете и системе государственной статистики в Российской Федерации». СПС «Консультант плюс», 2011.
16. Постановление Правительства РФ от 02.10.2006 N 595 (ред. от 21.04.2011) «О федеральной целевой программе «Развитие государственной статистики России в 2007 - 2011 годах» СПС «Консультант плюс», 2011.
17. Постановление Правительства РФ от 02.06.2008 N 420 (ред. от 24.03.2011) «О Федеральной службе государственной статистики» СПС «Консультант плюс», 2011.
18. Постановление Правительства РФ от 26.05.2010 N 367 (ред. от 27.12.2010) «О единой межведомственной информационно-статистической системе» СПС «Консультант плюс», 2011.

19. Распоряжение Правительства РФ от 05.08.2006 N 1086-р (ред. от 22.04.2009) О Концепции федеральной целевой программы «Развитие государственной статистики России в 2007 - 2011 годах» СПС «Консультант плюс», 2011.

20. Постановление Правительства РФ от 18.08.2008 N 620 «Об условиях предоставления в обязательном порядке первичных статистических данных и административных данных субъектам официального статистического учета» СПС «Консультант плюс», 2011.

21. Приказ Росстата от 06.09.2010 N 306 (с изм. от 19.07.2011) «Об утверждении статистического инструментария для организации федерального статистического наблюдения за деятельностью в сфере торговли, услуг, туризма, транспорта и связи, правонарушений» СПС «Консультант плюс», 2011.

22. Приказ Росстата от 20.08.2009 N 179 (ред. от 15.09.2010, с изм. от 19.07.2011) «Об утверждении статистического инструментария для организации федерального статистического наблюдения за внутренней и внешней торговлей» (с изм. и доп., вступающими в силу с отчета за январь - март 2011 года) СПС «Консультант плюс», 2011.

23. Приказ Росстата от 08.07.2011 N 316 «Об утверждении статистического инструментария для организации федерального статистического наблюдения за уровнем жизни населения» СПС «Консультант плюс», 2011.

# ТЕОРИЯ СТАТИСТИКИ И ЭФФЕКТИВНОСТЬ РЕКЛАМНОГО ДЕЛА

В.А. Чвякин

Учебник

**Главный редактор:** Краснова Наталья Александровна – кандидат экономических наук, доцент, руководитель НОО «Профессиональная наука»

**Технический редактор:** Канаева Ю.О.



Fig.3  
12,000 14,000 1,200



ISBN 978-5-907607-56-9



9 785907 607569 >

Усл. печ. Л 6,4.  
Объем издания 7,2 МВ  
Оформление электронного издания:  
НОО Профессиональная наука, mail@scipro.ru  
Дата размещения: 10.01.2024 г.  
URL: [http://scipro.ru/conf/advertising1\\_24.pdf](http://scipro.ru/conf/advertising1_24.pdf)