В. В. Швырков

Тайна традиционной статистики Запада

THE MYSTERY OF STATISTICAL SCIENCE

G. THROWKOFF PRESS

ТАЙНА ТРАДИЦИОННОЙ СТАТИСТИКИ ЗАПАДА

МОСКВА
"ФИНАНСЫ И СТАТИСТИКА"
1998

УДК 311:33
ББК 65.051
Ш 35

Авторизованный перевод с английского
доктора экономических наук, профессора
В. В. Швыркова

Ответственный редактор
профессор К. Отто

Владислав Васильевич Швырков, доктор экономических наук, профессор, родился в 1931 году в Сибири, учился в Москве. Он работал главным экспертом Госплана СССР, руководил кафедрой экономико-математических методов и статистики во Всесоюзном заочном институте пищевой промышленности в Москве. Профессор Швырков является основателем и руководителем Международного общества статистической науки. В течение 40 лет, занимаясь преподавательской, консультационной и научной деятельностью, он выступал с лекциями в различных университетах мира: Австралии, Болгарии, Канады, Китая, Чехословакии, Финляндии, Германии, Венгрии, Италии, Японии, Панамы, Польши, России, Швеции, Объединенного Королевства и США.

Работы профессора Швыркова посвящены потребительскому спросу, прикладной теории вероятностей, общей теории статистики, прогнозированию, вопросам планирования и управления, анализу временных рядов и проверке однородности статистических данных (15 монографий, 7 учебных пособий и множество журнальных и научных статей).

Биография профессора В.В.Швыркова напечатана в книгах "Кто есть кто на Западе" и "Кто есть кто в Америке". Он был назван Человеком года (1992/93) Международным биографическим центром в Кэмбридже (Англия).

ISBN 0-9420-0466-3(USA)
ISBN 5-279-01946-1 (Россия)

Предисловие

Традиционная статистика Запада отрицает возможность априорного познания причины (генеральной совокупности). Поэтому она вынуждена прибегать к необоснованным предпосылкам (они принимаются на веру без проверок) для анализа следствия (выборочной совокупности). Данная статистическая методология парализовала не только возможность оценки однородности статистической совокупности, но и возможность ее определения. Без решения этой фундаментальной проблемы традиционная статистика оказалась в положении корабля без рулевого. Создавшееся положение в статистике устраивает то большинство, которое использует статистику как служанку для достижения своих корыстных целей.

Критика традиционной статистики Запада в России (проф. П.Маслов), в США (проф. В.Деминг), в Японии (проф. Я.Сакамота), в Индии (проф. Д.Бусу) колышет статистическую общественность, но конструктивного решения фундаментальной проблемы не дает.

Заслуга проф. В.Швыркова заключается в том, что он нашел решение проблемы однородности статистической совокупности, которую традиционная статистика Запада считала неразрешимой в течение более двух веков. Данное решение философско-статистическое. Оно зиждется на четырех аксиомах, т.е. объективных истинах. Статистическая методология, основанная на этих истинах, делает статистику независимой от злоупотреблений и притязаний ее потребителей. Но именно этот факт даст повод в первое время встретить ее в штыки большинством тех, кто любит удить рыбку в мутной воде. Однако истина в конечном счете всегда сильнее лжи, и я полностью разделяю мнение проф. Н.Акаике, основателя японской статистической школы, в том, что научный вклад проф. В.Швыркова "в перспективе приведет к расширению сферы практического применения статистики".

Внедрение в практику новой статистической методологии имеет далеко идущие последствия для всей нашей действительности, которая полна беспринципных компромиссов, двойного стандарта, дуалистических решений и релятивистских объяснений. Универсализм новой статистики, основанной на истинах, будет способствовать общему моральному возрождению, ибо праведная жизнь начинается там, где начинается искание истины.

Проф. В.Швырков посвятил свою статистическую деятельность служению истины, создав новую статистическую методологию, которая знаменует собой новую эпоху статистической мысли для нового XXI века.

Проф. В.Швырков относится к той категории ученых, у которых продукт знания и истина - понятия однозначные. Мы приветствуем публикацию его книги на русском языке и надеемся, что русская статистическая мысль претерпит в будущем радикальные изменения, положив в основу искание истины, и отвергнет необоснованные статистические предпосылки.

*Проф. др. К.Отто**
Берлин, ноябрь 1997 г.

*Проф. К.Отто (выдающийся немецкий статистик) занимал руководящие должности в ООН (Нью-Йорк), в СЭВ (Москва) и в Гумбольдском университете (Берлин). *Ред.*

Введение

"Тайна - это вроде обмана: умные дурачат глупых; но глупые не мирятся с этим и хотят тоже стать умными."

И.А. Ильин

Основная проблема традиционной статистики Запада - количественный анализ законов и закономерностей существующей действительности. Эта действительность представлена в статистике в виде параметров генеральной совокупности. Для оценки этих параметров традиционная статистика пользуется выборочной совокупностью. Однако репрезентативность выборочной совокупности в отношении генеральной совокупности неизвестна. Несмотря на это обстоятельство, традиционная статистика (ТС) "успешно" решает свою основную проблему. Это "успешное решение" представляется для многих пользующихся статистическими методами по меньшей мере таинственным. Данная таинственность строго соблюдается в статистической литературе, особенно в учебной. Причина этому одна: не принято откровенно говорить о своих недостатках, так как это приводит к потере уважения, потере клиентов, а следовательно, и к потере бизнеса.

Однако в чем же тайна традиционной статистики? А тайна ее в том, что ТС двулика. Теория ТС безупречна с математической точки зрения. Практика ТС характеризуется следующими недостатками. Во-первых, она базируется на вере в непогрешимость необоснованных предпосылок, например, относительно таких категорий, как однородность, репрезентативность и случайность. Во-вторых, она отрицает принцип монизма и оперирует принципами дуализма и релятивизма в решении практических задач. В-третьих, ТС является собой яркий пример непоследовательности несогласованности и противоречивости между теорией и практикой.

В настоящее время ТС пребывает в глубоком кризисе. Ее статистические заключения характеризуются неопределенностью, а их научная ценность чрезвычайно мала и условна. Необоснованные предпосылки способствуют быстрому росту неправильных толкований статистических результатов и злоупотреблению статистическими методами. В результате этого снижается профессиональный и этический уровень статистических исследований. Тенденция снижения научного уровня статистических исследований особенно резко стала падать после того, как нестатистики смогли пользоваться "консервированными" программами.

Итак, тайна традиционной статистики сопровождается злоупотреблением и обманом, а всякий обман знаменует собой начало деградации. Для перехода от деградации к прогрессу недостаточно вскрыть свои недостатки и отказаться от статистического обмана. Главное заключается в том, чтобы найти путь к истине. Ибо только знание истины и руководство ее поможет нам освободиться от наших недостатков и создать статистическую науку. Методологические основы этой новой статистики (НС) должны быть, во-первых, построены на истинах, а не на необоснованных предпосылках. Во-вторых, НС должна отвергнуть принципы дуализма и релятивизма, заменив их принципом монизма. В-третьих, НС должна исключить элементы непоследовательности, несогласованности и противоречивости между теорией и практикой. При соблюдении этих трех условий статистика вернет свое первоначальное назначение: "наука, борющаяся за истину" (в переводе с греческого языка).

Итак, статистика как наука, борющаяся за истину, должна быть основана на истинах, т.е. на аксиомах. Формулировка статистических аксиом и разработка критерия репрезентативности однородности (КРО) для проверки необоснованных предпосылок традиционной статистики - главная цель новой статистики.

Критерий репрезентативной однородности не ограничивается анализом одномерных распределений. Он применим и для анализа многомерных распределений, а также для контроля качества продукции; особенно он эффективен при коротких производственных процессах. Преимущество этого критерия в том, что он повышает уровень научных исследований, проводимых с применением статистических методов, например, таких, как регрессионный анализ и корреляционный анализ.

Разработка КРО - результат критического анализа фундаментальных статистических работ Т. Бейеса, В. Лексиса, А. Чупрова, Р. Фишера, В. Шухерта и Г. Тагучи. Критика традиционной статистики такими учеными, как Д. Бусу, В. Деминг, и учениками Х. Акаике (Я. Сакамото, М. Ишигуро, Г. Китагава), а также содержательные философские работы Б. Спинозы воодушевили автора на написание данной работы.

В заключение поясним буквенные и сокращенные обозначения, применяемые в книге:

1. Прописные буквы применяются для обозначения статистических генеральных совокупностей.
2. Прописные буквы (курсив) применяются для обозначения вероятностных генеральных совокупностей.
3. Строчные буквы применяются для обозначения статистических выборок.
4. Строчные буквы (курсив) применяются для обозначения вероятностных выборок.
5. Сокращенные обозначения:

ААККП	- аксиоматический анализ контроля качества продукции
Б	- благоприятные причины
Ис	- интервал сходства
КРО	- критерий репрезентативной однородности
Кс	- коэффициент сходства
Н	- неблагоприятные причины
НВР-ТЗР	- нерепрезентативное вероятностное распределение теоретических значений репрезентативности
НС	- новая статистика
ОНВ	- однородная неделимая выборка
ОНГ	- однородная невидимая генеральная совокупность
ОНП	- однородные невидимые причины
РВР-СС	- репрезентативное вероятностное распределение сложных событий
РВР-ТЗР	- репрезентативное вероятностное распределение теоретических значений репрезентативности
РНСС	- репрезентативный набор сложных событий
РО	- репрезентативная однородность
СС	- сложное событие
ССК	- состояние статистического контроля
ТЗР	- теоретические значения репрезентативности
ТАККП	- традиционный анализ контроля качества продукции
ФНК	- философия нового качества статистических данных
ЭЗР	- эмпирические значения репрезентативности

В.Швырков
Санта Роза, Калифорния
Ноябрь 1997 г.

"Общепринято считать, что предмет познается либо в результате анализа его сущности, либо посредством знания его ближайшей причины."

Бенедикт Спиноза

Часть I
Индуктивный метод мышления

1. Основная проблема статистической науки

Статистика участвует в процессе познания законов и закономерностей нашей действительности. Данный процесс познания осуществляется путем разложения сложного явления на элементарные явления. В статистической терминологии элементарное явление характеризуется одномерной однородной совокупностью, распределение которой одномодальное. Комбинация одномерных однородных совокупностей, т.е. элементарных явлений (причин), образует н-мерную однородную совокупность, т.е. сложное явление (следствие). Это сложное явление является однородным, если объяснимые (известные) элементарные явления оказывают существенное влияние на его формирование, в то время как все необъяснимые (неизвестные) элементарные явления не оказывают существенного влияния. Если это условие не соблюдается, то сложное явление нельзя считать однородным.

С учетом вышеизложенного, наше определение однородной статистической совокупности следующее: статистическая совокупность тогда однородна, когда она представлена элементарным явлением или когда она представлена сложным явлением, изменения которого существенно зависят от объяснимых элементарных явлений.

Традиционная статистика не располагает определением однородности статистической совокупности. Более того, ТС утверждает, что проблема однородности не является статистической проблемой и что "однородность изучаемого статистического материала в принципе не может быть установлена." (Sachs, 1982, стр. 83). Поэтому проблема однородности статистических данных рассматривается ТС как логическая проблема, решение которой зависит от субъективного мнения исследователя.

Значение проблемы однородности трудно переоценить. Статистики различных направлений прекрасно это понимают. Так, например, А. Чупров (1959) еще в 1910 г. считал, что переход от относительных частот к математическим вероятностям возможен только на базе однородных статистических данных. М. Болдрини (Boldrini, 1972, стр. 162), разделяя мнение о важности проблемы однородности в статистике, пишет: "...экспериментальная вероятность - это индукция; она должна вычисляться на основе однородных фактических данных..." Дискутируя о качестве статистической информации, Р. Фишер заметил, что статистик теперь уже не алхимик, ожидающий получить золото из негодного материала, предложенного ему.

Нельзя забывать о том, что потенциальные последствия использования неоднородных статистических данных чрезвычайно серьезны, они полностью аннулируют полученные статистические выводы. В связи с этим, мы полностью согласны со следующим заявлением Е.С. Пирсона и Г.О. Хартли (Pearson, Hartley, 1954, стр. 83): "...функция статистических методов - предупреждать каждого о том, что правильные выводы не могут быть получены на основе неадекватных данных."

В заключение вышеприведенных цитат можно сказать, что наши выводы настолько достоверны, насколько хороши наши данные. К сожалению, статистики не знают, что такое хорошие данные и как проверить однородность статистических данных. Вот почему статистики находятся в положении тех алхимиков, которые, по образному выражению Р. Фишера, надеются извлечь золото из материала неизвестного качества.

Однако положение небезнадежное. Решение проблемы однородности статистических данных следует позаимствовать из теории вероятностей. В теории вероятностей заложены основы статистики. Эти основы зиждутся на принципе однородности, при помощи которого теория вероятностей прослеживает движение следствия от причины. В теории вероятностей принцип однородности известен как принцип равновероятности и, по мнению А. Хинчина (1961, стр. 88), он "...единственный имеющийся у нас метод теоретического предсказания вероятностей событий в отдельных конкретных ситуациях."

Принцип однородности ассоциируется в теории вероятностей также и с понятием однородной генеральной совокупности. Генеральная совокупность считается в теории вероятностей однородной, если соблюдается принцип равновероятности. Следовательно, основная характеристика однородной генеральной совокупности - постоянная вероятность благоприятного события.

Теория вероятностей пользуется дедуктивным методом мышления: следствие познается в результате знания причины, так вероятность благоприятного события есть результат знания однородной генеральной совокупности. Статистика пользуется индуктивным методом мышления: причина познается в результате знания следствия. Для соблюдения согласованности между двумя методами мышления индуктивный метод должен удовлетворять следующим двум условиям: генеральная совокупность - однородная, выборка - репрезентативная.

Индуктивный метод мышления, удовлетворяющий названным выше двум условиям, есть основа философии нового качества (ФНК) статистических данных. ФНК исследует формирование однородных явлений под влиянием воздействия внутренних сил. Это новое направление в исследовании массовых явлений рассматривает статистику как науку и дает ей следующее определение: статистика - универсальная методологическая наука, которая в состоянии количественно измерить причинно-следственные зависимости. Эта количественная оценка только тогда достоверна, когда выборка репрезентативна однородной генеральной совокупности. Согласно такому определению, всякое статистическое исследование должно состоять из двух этапов: на первом этапе проверяется однородная репрезентативность выборки, на втором этапе анализируется статистическая модель и рассчитываются ее характеристики.

Выводы. Основная проблема всех наук - выяснить причину, анализируя следствие, т.е. познать законы и закономерности нашей действительности по данным статистической выборки. Следовательно, качество статистической выборки играет решающую роль в исследованиях всех наук и является основной проблемой статистической науки.

Качество выборки характеризуется ее репрезентативностью в отношении однородной генеральной совокупности. Таким образом, качественная выборка удовлетворяет двум условиям: однородность генеральной совокупности, репрезентативность выборки.

Традиционная статистика утверждает, что проблема однородности не статистическая, а логическая; следовательно она не может быть исследована статистическими методами.

Вероятностное объяснение однородной генеральной совокупности и ее репрезентативного набора выборок есть основа для осмысливания индуктивного метода мышления. Применение этого метода в статистике должно полностью согласовываться с принципами дедуктивного метода мышления, сформулированными теорией вероятностей.

2. Статистические школы и оценка их индуктивного мышления

Философия статистических методов базируется на индуктивном методе мышления. Индуктивный метод мышления следует рассматривать в контексте с дедуктивным методом мышления, формулировка которого дана в теории вероятностей.

В теории вероятностей развитие дедуктивного метода мышления движется от причины к следствию, т.е. от однородной генеральной совокупности к репрезентативному набору выборок. В статистике развитие индуктивного метода мышления происходит от следствия к причине, т.е. от репрезентативной выборки к однородной генеральной совокупности. В этой формулировке индуктивный метод согласуется с дедуктивным методом мышления и характеризуется двумя условиями: генеральная совокупность - однородная, выборка - репрезентативная.

Вопрос о соблюдении вышеназванных двух условий и их статистическая интерпретация по-разному решаются различными статистическими школами. Ниже анализируются решения данного вопроса основными шестью статистическими школами.

2.1. Статистическая школа Бейеса

Английский пресвитерианский священник и математик Томас Бейес (1702-1761) был первым из тех, кто применил индуктивный метод мышления к вероятностным категориям: от условных вероятностей (выборка) к безусловным вероятностям (генеральная совокупность).

Бейес (Bayes, 1764) элегантно вывел формулу функциональной связи между выборкой и генеральной совокупностью. А это значит, что выборка Бейеса всегда репрезентативна генеральной совокупности. Однако однородность генеральной совокупности не оценивалась Бейесом. Следовательно, только одно из двух условий индуктивного метода мышления было соблюдено Бейесом.

2.2. Континентальная статистическая школа

Немецкий статистик Вильям Лексис (1837-1914) был первым из тех, кто попытался оценить однородность статистических данных, изучая стабильность динамических рядов.

Вклад Лексиса в статистику воистину грандиозен. Он выдвинул идею, согласно которой однородная генеральная совокупность известна априорно в виде Бернуллевского вероятностного распределения (Lexis, 1877). Так как Бернуллевское вероятностное распределение однородное, первое условие индуктивного метода мышления (однородность генеральной совокупности) Лексисом было выполнено. Однако второе условие (репрезентативность выборки) не было им выполнено. Лексис не рассчитывал критерий репрезентативности, который бы оценил степень соответствия конфигураций двух распределений, выборки и генеральной совокупности. Более того, Лексис ограничивался точечной оценкой репрезентативности выборки. Поэтому он не мог принять во внимание несущественное влияние необъяснимых причин, образующих допустимые отклонения.

Русский статистик А. Чупров (1874-1926) был активным сторонником теории стабильности Лексиса. Его вклад в развитие этой теории весьма существенный. Он пропагандировал стохастические идеи для решения проблемы однородности. Чупров считал, что для решения этой проблемы фактические данные должны быть проанализированы с позиции случайной переменной с ее вероятностным распределением. К сожалению, Чупров игнорировал вероятность "отдельного случая" (A. Markov, 1911), что является центральным пунктом решения проблемы однородности.

2.3. Английская статистическая школа

Английский ученый Р.А. Фишер (1890-1962) является основателем традиционной математической статистики (Fisher, 1954). Его прекрасно разработанная статистическая теория разрешила проблему интервальной оценки параметров генеральной совокупности, что было слабым звеном в теории стабильности Лексиса. Но, к сожалению, в теории Фишера оба условия индуктивного метода мышления (однородность генеральной совокупности и репрезентативность выборки) не выполняются.

Индуктивный метод Фишера базируется на его теории о выборочных распределениях. Согласно этой теории, оценка параметров генеральной совокупности производится на основе распределения выборочных статистик. Генеральная совокупность по Фишеру не должна быть обязательно однородной для оценки параметров генеральной совокупности. Более того, идея о том, что генеральная совокупность известна априорно, Фишером полностью отвергается. Тем не менее Фишер предполагает, что выборка однородная и что она репрезентативна генеральной совокупности в отношении ее дисперсии. Это предположение является необоснованной предпосылкой. В связи с этим оценка параметров генеральной совокупности по Фишеру не заслуживает доверия.

В заключение следует особо выделить три существенных недостатка теории Фишера о выборочных распределениях. Во-первых, выборка из гипотетической однородной генеральной совокупности Фишера представляет собой смесь совокупностей однородных и разнородных. Во-вторых, так как большинство малых выборок не репрезентативно генеральной совокупности, шансы получения правильных оценок параметров генеральной совокупности на базе малых выборок чрезвычайно малы. В-третьих, тест статистического согласия Фишера неприменим, когда речь идет о выборе одной из двух относительно одинаковых моделей, предназначенных для описания генеральной совокупности.

2 Зак. 28

2.4. Японская статистическая школа Акаике

Японский статистик Х. Акаике (Akaike, 1973) разработал новый конструктивный метод для решения проблемы, которую традиционная статистика была не в состоянии разрешить. Речь идет о выборе одной из двух относительно одинаковых моделей (тест статистического согласия).

Согласно философии Акаике, любая статистическая переменная имеет свой прообраз в виде случайной величины с ее вероятностным распределением. Для определения типа вероятностного распределения случайной величины, который наилучшим образом согласуется со статистической переменной, был разработан критерий под названием "информационный критерий Акаике" (Sakamoto & others, 1986). Этот критерий измеряет "расстояния" между статистической переменной и гипотетическими моделями. Метод "расстояний" не может быть использован для оценки репрезентативности выборки в отношении гипотетических моделей, так как метод "расстояний" не в состоянии уловить различия в конфигурациях сопоставляемых распределений. Более того, генеральная совокупность, т.е. гипотетическая модель, согласно методу Акаике, не обязательно должна быть однородной. В связи с вышеизложенным напрашивается следующий вывод: оба условия индуктивного метода мышления не выполняются.

2.5. Американская статистическая школа контроля качества продукции

Американский статистик В. Шухарт (1891-1967) является основателем статистического контроля качества продукции в промышленности (Shewhart, 1931). Его прогрессивный метод можно охарактеризовать четырьмя пунктами. Первый пункт: Шухарт оценивал стабильность данных, анализируя качество произведенной продукции. Второй пункт: Шухарт строит свои выводы не на одной, а на группе выборок. Третий пункт: первое условие индуктивного метода мышления (однородность генеральной совокупности) выполняется: генеральная совокупность Шухарта представлена нормальным распределением. Однако Шухарт не рассчитывал статистический критерий нормальности по выборочным данным, так как он понимал его ограниченность и условность. Четвертый пункт: Шухарт разработал условия для определения пределов спецификации, которые явились интервальными оценками для установления качества продукции.

1619

Согласно теории Шухарта, производственный процесс стабилен, если индивидуальные значения, средние и стандартные отклонения расположены между пределами спецификации, вычисленными на основе нормального распределения.

Новаторский метод Шухарта был значительно ослаблен необоснованной предпосылкой, согласно которой произведенная продукция считается качественной, если она находится между пределами спецификации. Эта предпосылка ложная, так как не все выборки, расположенные в пределах спецификации, имеют нормальное распределение; другими словами, не все выборки репрезентативны в отношении нормальной генеральной совокупности.

В заключение следует сказать, что второе условие индуктивного метода мышления (репрезентативность выборки) не выполняется Шухартом.

2.6. Японская статистическая школа контроля качества продукции

Японский инженер Г. Тагучи предложил новый прогрессивный метод, который произвел революцию в теории и практике статистического контроля качества продукции (Taguchi, Wu, 1979). Тагучи впервые в статистике попытался выполнить оба условия индуктивного метода мышления (однородность генеральной совокупности и репрезентативность выборки).

Решая практические проблемы каждый день, Тагучи, как инженер, понял, что продукция только тогда качественна, когда она соответствует образцу. Таким образцом Тагучи считал продукцию, которая обладает лучшей сопротивляемостью в отношении влияния внешних факторов. Тагучи заменил пределы спецификации Шухарта этим образцом, так как он понимал, что единообразие продукции гораздо важнее, чем соответствие пределам спецификации.

Метод Тагучи совершенен с теоретической точки зрения, так как он основан на понятии репрезентативности продукции образцу. На статистическом языке это означает, что метод Тагучи удовлетворяет двум условиям индуктивного метода мышления. Его образец есть однородная генеральная совокупность, а произведенная качественная продукция есть репрезентативная выборка.

Но, с практической точки зрения, метод Тагучи обладает двумя недостатками. Первый недостаток: образец (генеральная совокупность) и его допуски спроектированы и оценены на основе выборок, которые подвержены влиянию объяснимых и необъяснимых факторов. Таким образом, образец Тагучи не есть идеальный объект, и в статистической терминологии такой объект может быть неоднородной генеральной совокупностью.

Второй недостаток: репрезентативность произведенной продукции в отношении образца производится методом сокращения отклонений. Недостаток этого метода заключается в его субъективизме при определении размера отклонений. Именно это и служит основанием утверждать, что данный статистический метод в высшей степени сомнителен.

2.7. Выводы

Индуктивный метод мышления в статистике требует выполнения двух условий: однородность генеральной совокупности и репрезентативность выборки. Эти условия были и остаются в центре внимания всех выдающихся представителей традиционной статистики Запада.

1. Т. Бейес установил функциональную связь между выборкой и генеральной совокупностью, т.е. условие репрезентативности выборки было выполнено. Однако первое условие, однородность генеральной совокупности, им не было выполнено.
2. В. Лексис исследовал стабильность временных рядов. Его теория базировалась на *априорном знании генеральной совокупности, представленной* в виде вероятностного распределения Бернулли. Таким *образом, первое условие было выполнено. Второе условие не было* выполнено, так как Лексис не оценивал репрезентативность распределения данных выборки по отношению к вероятностному распределению Бернулли.
3. Р. Фишер разработал методы математической статистики для оценок параметров генеральной совокупности и для проверок гипотез на основе его стройной теории о выборочных распределениях. Однако его индуктивный метод мышления оперирует необоснованными предпосылками, согласно которым выборка однородна и репрезентативна генеральной совокупности в отношении дисперсии. При этом генеральная совокупность Фишера может быть и неоднородной. Все это свидетельствует о том, что оба условия индуктивного мышления Фишера не выполняются.
4. Х. Акаике предложил метод оценки для двух относительно одинаковых моделей. Этот метод позволил решить задачу, которую критерий статистического согласия Фишера был не в состоянии разре-

шить. Акаике, как и Чупров, пропагандирует идею, согласно которой случайная величина с ее *вероятностным распределением* есть прообраз статистической переменной. Однако Акаике не сумел приложить эту идею к *решению* проблемы однородности. Более того, он считает, что выборка и генеральная совокупность не обязательно должны быть *однородными*. Следовательно, метод Акаике явно противоречит первому условию индуктивного метода мышления при желании *соблюдать* второе условие.
5. В. Шухарт, создатель системы контроля качества продукции, предложил проводить статистический анализ не по одной выборке, а по нескольким. Шухарт утверждает, что распределение выборок репрезентативно однородной генеральной совокупности (нормальное распределение), если они находятся между пределами спецификации. Однако пределы спецификации не гарантируют репрезентативность каждой выборки нормальному распределению. Поэтому, при выполнении первого условия, второе условие индуктивного метода мышления не выполняется
6. Г. Тагучи пытается выполнить оба условия индуктивного метода мышления. Его генеральная совокупность - образец, т.е. продукция, обладающая наилучшей сопротивляемостью в отношении влияния внешних факторов. Его метод определения репрезентативности выборки основан на итеративном способе сокращения разностей между выборкой и генеральной совокупностью. Однако образец Тагучи - не идеальная генеральная совокупность, так как он зависит от влияния необъяснимых внешних факторов. Метод Тагучи, применяемый для определения репрезентативности выборки, основан на субъективной оценке разностей между выборкой и генеральной совокупностью. В связи с этим напрашивается вывод: образец не обязательно является однородной генеральной совокупностью, и репрезентативность выборки в отношении этой совокупности не гарантируется. Поэтому оба условия индуктивного метода мышления не выполняются.

"Всякий, кто от истины,
слышит голос мой."

От Иоанна, 18:37

"Знание следствия зависит
от знания причины
и связано с ним."

Бенедикт Спиноза

Часть II
Философия нового качества статистических данных

3. Аксиомы

Каждая истинная наука зиждется на нескольких аксиомах, или общеизвестных истинах, которые доказать невозможно; они воспринимаются как очевидность реального факта.

Традиционная статистика не является истинной наукой, так как она базируется не на аксиомах, а на трех необоснованных предпосылках относительно однородности статистической совокупности, репрезентативности выборки и случайности переменной величины.

Первая необоснованная предпосылка - статистическая совокупность данных, отобранная для анализа, всегда предполагается однородной. Действительно, очень важно иметь дело с качественной информацией, так как наши решения настолько хороши, насколько хороша статистическая информация, на которой они основаны. К сожалению, статистики не знают, что такое хорошая статистическая информация, как проверить ее однородность, т.е. ее пригодность для конкретного анализа.

Вторая необоснованная предпосылка: выборка репрезентативна генеральной совокупности в отношении стандартного отклонения. Это значит, что рассеяние выборки равно рассеянию генеральной совокупности. Эта предпосылка не может быть проверена, так как генеральная совокупность, согласно традиционной статистике, не известна.

Третья необоснованная предпосылка: математическое ожидание отклонений от линейной стохастической регрессионной модели (генеральной совокупности) есть случайная величина или ненаблюдаемые разности. Эта предпосылка случайности априорна, так как нет статистического метода проверки ненаблюдаемых разностей. Анализ наблюдаемых разностей не в состоянии решить проблему проверки предпосылки случайности по двум причинам. Во-первых, распределение наблюдаемых разностей зависит от типа модели. Правильность выбора типа модели не гарантируется. Во-вторых, статистический критерий нормальности, применяемый для анализа распределения наблюдаемых разностей, не есть критерий проверки нормального распределения этих разностей. Роль этого критерия ограничена: он может только принять или отвергнуть гипотезу о том, что наблюдаемые разности есть результат выборки из нормального генерального распределения.

Необоснованные предпосылки - печальная трагедия традиционной статистики. Для избавления ее от этой трагедии необходимо создать статистическую науку, основанную на философии нового качества статистических данных (ФНК). Эта философия должна быть построена на аксиомах, которые формулируются в терминах теории вероятностей.

3. Аксиомы

Каждая истинная наука зиждется на нескольких аксиомах, или общеизвестных истинах, которые доказать невозможно; они воспринимаются как очевидность реального факта.

Традиционная статистика не является истинной наукой, так как она базируется не на аксиомах, а на трех необоснованных предпосылках относительно однородности статистической совокупности, репрезентативности выборки и случайности переменной величины.

Первая необоснованная предпосылка - статистическая совокупность данных, отобранная для анализа, всегда предполагается однородной. Действительно, очень важно иметь дело с качественной информацией, так как наши решения настолько хороши, насколько хороша статистическая информация, на которой они основаны. К сожалению, статистики не знают, что такое хорошая статистическая информация, как проверить ее однородность, т.е. ее пригодность для конкретного анализа.

Вторая необоснованная предпосылка: выборка репрезентативна генеральной совокупности в отношении стандартного отклонения. Это значит, что рассеяние выборки равно рассеянию генеральной совокупности. Эта предпосылка не может быть проверена, так как генеральная совокупность, согласно традиционной статистике, не известна.

Третья необоснованная предпосылка: математическое ожидание отклонений от линейной стохастической регрессионной модели (генеральной совокупности) есть случайная величина или ненаблюдаемые разности. Эта предпосылка случайности априорна, так как нет статистического метода проверки ненаблюдаемых разностей. Анализ наблюдаемых разностей не в состоянии решить проблему проверки предпосылки случайности по двум причинам. Во-первых, распределение наблюдаемых разностей зависит от типа модели. Правильность выбора типа модели не гарантируется. Во-вторых, статистический критерий нормальности, применяемый для анализа распределения наблюдаемых разностей, не есть критерий проверки нормального распределения этих разностей. Роль этого критерия ограничена: он может только принять или отвергнуть гипотезу о том, что наблюдаемые разности есть результат выборки из нормального генерального распределения.

Необоснованные предпосылки - печальная трагедия традиционной статистики. Для избавления ее от этой трагедии необходимо создать статистическую науку, основанную на философии нового качества статистических данных (ФНК). Эта философия должна быть построена на аксиомах, которые формулируются в терминах теории вероятностей.

3.1. Первая аксиома

Знание следствия зависит от знания причины, утверждал Спиноза (Spinoza, 1951, стр. 42). В статистике следствие есть выборка, а причина, в терминах ФНК, есть однородная невидимая генеральная совокупность (ОНГ). Эта генеральная совокупность (причина) должна быть известна априори, если мы намерены определить качество выборки, т.е. ее однородность (следствие). Априорное знание ОНГ - результат логического процесса мышления, в результате которого формулируется определение в виде аксиомы. Так, например, мы предлагаем следующее определение для ОНГ: совокупность двух групп однородных невидимых причин (ОНП), благоприятных ($\mathcal{Б}$) и неблагоприятных ($\mathcal{Н}$), образует ОНГ. Благоприятные причины оказывают положительное влияние на событие. Неблагоприятные причины оказывают отрицательное влияние на событие.

Эти две группы невидимых причин однородные, если соблюдаются три условия. Во-первых, причины должны быть независимыми. Во-вторых, каждая группа причин характеризуется своим постоянным влиянием на следствие. В-третьих, вероятность благоприятного события постоянная.

Таким образом, первая аксиома формулируется в терминах теории вероятностей следующим образом: ОНГ есть вероятностное распределение Бернулли, если однородные невидимые причины (неблагоприятные и благоприятные) обозначаются нулем и единицей (Рис. 3.1.1).

Однородная генеральная совокупность в виде одномерного распределения

характеризуется постоянной вероятностью благоприятного события ($\mathcal{P}$)

↓

Вероятностное распределение Бернулли

$p(x)=\mathcal{P}^x \cdot \mathcal{Q}^{(1-x)}$

Рис. 3.1.1

Данное определение обусловлено принципом монизма - вселенная дискретна и конечна (Ghosh, 1988, стр. 23), а также принципом минимума ($n = 1$).

3.2. Вторая аксиома

Для перехода от невидимых (ненаблюдаемых) причин к статистическим (наблюдаемым) данным необходимо найти связующее звено между ними. Таким звеном служит размер наблюдаемого явления, которое формируется под влиянием невидимых причин (благоприятных и неблагоприятных). Под влиянием неблагоприятных причин размер наблюдаемого явления (индивидуального события) меньше средней, под влиянием благоприятных причин - больше средней, а под влиянием обеих причин - равен средней.

Таким образом сочетания между однородными невидимыми причинами (ОНП) определяют размер статистического индивидуального наблюдения по отношению к средней статистической генеральной совокупности при условии, что она репрезентативна средней ОНГ. Эти сочетания, в соответствии с принципом минимума*, следующие (Рис. 3.2.1):

$$\mathcal{Н} \cap \mathcal{Н} = x_1 < \overline{X}$$
$$\mathcal{Б} \cap \mathcal{Б} = x_2 > \overline{X}$$
$$\mathcal{Н} \cap \mathcal{Б} = x_3 = \overline{X}$$
$$\mathcal{Б} \cap \mathcal{Н} = x_4 = \overline{X}$$

Следовательно, статистическое индивидуальное наблюдение следует рассматривать как сложное событие или как неделимую выборку с соответствующей вероятностью. Вероятности сложных явлений вычисляются при условии, что вероятность благоприятного события ОНГ известна. Это значит, что каждое статистическое индивидуальное наблюдение имеет свою вероятность, или, по терминологии Маркова, “каждый отдельный случай характеризуется своей вероятностью” (Markov, 1911).

В заключение сформулируем вторую аксиому: статистическое индивидуальное наблюдение есть сложное событие или однородная неделимая выборка (ОНВ), формирующаяся под влиянием ОНП.

---

* Принцип минимума соблюдается тогда, когда размер однородной неделимой выборки (ОНВ) равен числу причин ОНГ, т.е. двум.

Рис. 3.2.1

3.3. Третья аксиома

Однородное статистическое индивидуальное наблюдение есть результат влияния сложного события, состоящего из двух ОНП (согласно принципу минимума). Однородная статистическая совокупность (однородная выборка) есть результат влияния нескольких сложных событий, каждое из которых состоит из двух ОНП. Распределение этих сложных событий определяется объяснительным фактором. Этот фактор является вероятностью появления благоприятного события ($\mathcal{P}$) однородной невидимой генеральной совокупности (ОНГ). Данная вероятность определяет размеры средних величин ОНГ и репрезентативной выборки.

В заключение правомерно сделать следующие выводы относительно выборки, репрезентативной ОНГ (Рис. 3.3.1):

1. Средняя величина выборки ($\overline{x}$) равна средней величине ОНГ ($\overline{\mathcal{X}}$)*.
2. Вид статистического распределения (в терминах коэффициента асимметрии, sk) аналогичен виду распределения ОНГ (в терминах вероятности благоприятного события, $\mathcal{P}$).

---

* $\overline{x} = \overline{\mathcal{X}}$, если обе переменные выражены в одинаковых единицах измерения и имеют одинаковые границы колебаний.

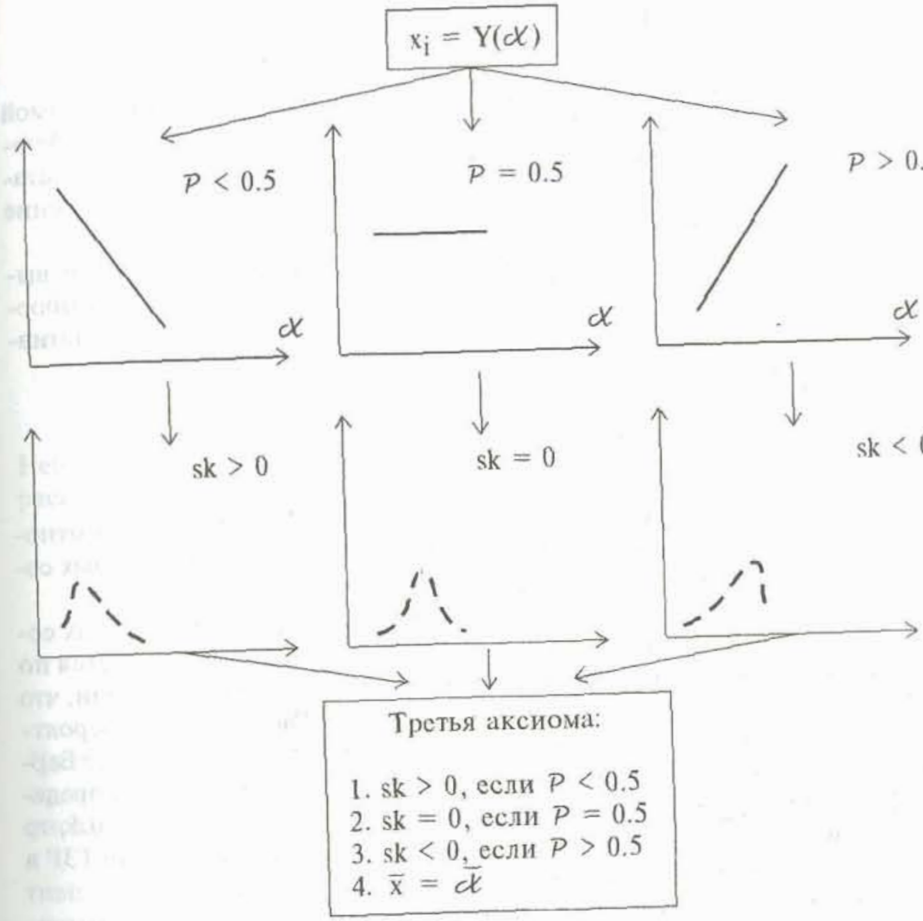


Рис. 3.3.1

Таким образом, третья аксиома однородной статистической совокупности может быть сформулирована следующим образом: однородные невидимые причины определяют размер средней величины и вид распределения репрезентативной статистической выборки. Эта аксиома может быть записана и в статистических терминах:

1. $\overline{x} = \overline{\mathcal{X}}$
2. sk > 0, если $\mathcal{P} < 0.5$
3. sk = 0, если $\mathcal{P} = 0.5$
4. sk < 0, если $\mathcal{P} > 0.5$

3.4. Четвертая аксиома

Первые три аксиомы формулируют определения однородной невидимой генеральной совокупности (ОНГ), статистического индивидуального наблюдения и характеризуют связь между распределением ОНГ и однородным статистическим распределением. Четвертая аксиома формулирует определение однородной статистической совокупности.

Объяснение четвертой аксиомы мы начинаем с анализа ожидаемых выборок ($n = 2$), извлеченных из ОНГ с $\mathcal{P} = 0.5$ (Рис. 3.4.1). Ожидаемые выборки или сложные события выразим в теоретических значениях репрезентативности (ТЗР):

$$\text{Н} \cap \text{Н} = 0$$
$$\text{Б} \cap \text{Б} = 0$$
$$\text{Н} \cap \text{Б} = 1$$
$$\text{Б} \cap \text{Н} = 1$$

Первые два *сложных события не являются репрезентативными* в отношении ОНГ, поэтому они обозначаются нулями. Последние два сложных события репрезентативные, поэтому они обозначаются единицами.

Теперь рассмотрим вопрос о распределении вероятностей сложных событий. *Вероятность появления каждого сложного события вычисляется по* правилу умножения вероятностей для независимых событий при условии, что вероятность благоприятного события ОНГ известна. Распределение вероятностей сложных событий по ТЗР образует вероятностное распределение Бернулли. Вероятность появления благоприятного события (p) этого распределения функционально зависит от $\mathcal{P}$ (Рис. 3.4.1). Так как $p = \mathcal{P} = 0.5$, то репрезентативность распределения вероятностей сложных событий по ТЗР в отношении распределения вероятностей ОНГ равна 100%. Такой же процент репрезентативности имеет место и в отношении двух переменных: переменная ТЗР и переменная ОНГ состоят из одинакового соотношения (50%) благоприятных и неблагоприятных событий. Однако репрезентативная переменная распределения ТЗР состоит только из 50% репрезентативного набора сложных событий (выборок). Следовательно, набор выборок репрезентативен, если он состоит минимум из 50% репрезентативных выборок. Этим определением мы даем количественную характеристику принципу репрезентативности.

В соответствии с вышеизложенным, принцип репрезентативности набора сложных событий сформулируем следующим образом (Рис. 3.4.2). Распределение ТЗР репрезентативно ОНГ, если соблюдаются два условия:

1. Переменная ТЗР состоит из 50% или более репрезентативных сложных событий.

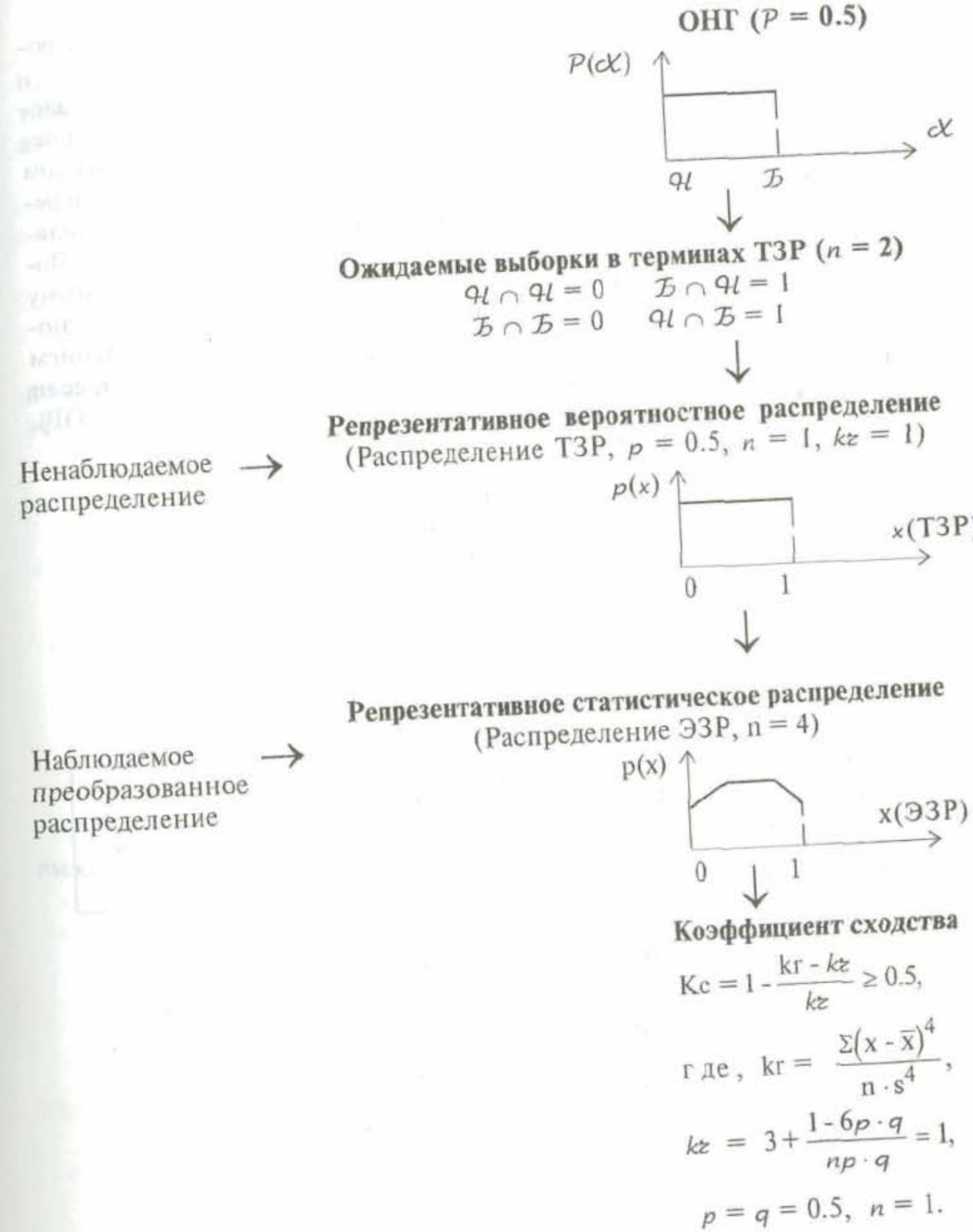


Рис. 3.4.1

2. Расчет вероятностей ТЗР возможен только при условии, что вероятность благоприятного события ОНГ известна.

Эти два условия применимы, когда речь идет о распределении ожидаемых сложных событий (ненаблюдаемом распределении). Данные условия неприменимы, когда речь идет о наблюдаемом распределении (статистическом распределении). Причина этому следующая. Во-первых, наблюдаемая переменная не может быть преобразована в переменную благоприятных и неблагоприятных причин с тем, чтобы быть сопоставимой с переменной ОНГ. Во-вторых, вероятность благоприятного события ОНГ неизвестна, и поэтому невозможно вычислить наблюдаемые частности. Во избежание этих трудностей, наблюдаемое распределение должно сопоставляться с распределением ТЗР. Это распределение сопоставимо с наблюдаемым распределением, если последнее преобразовано в эмпирические значения репрезентативности (ЭЗР).

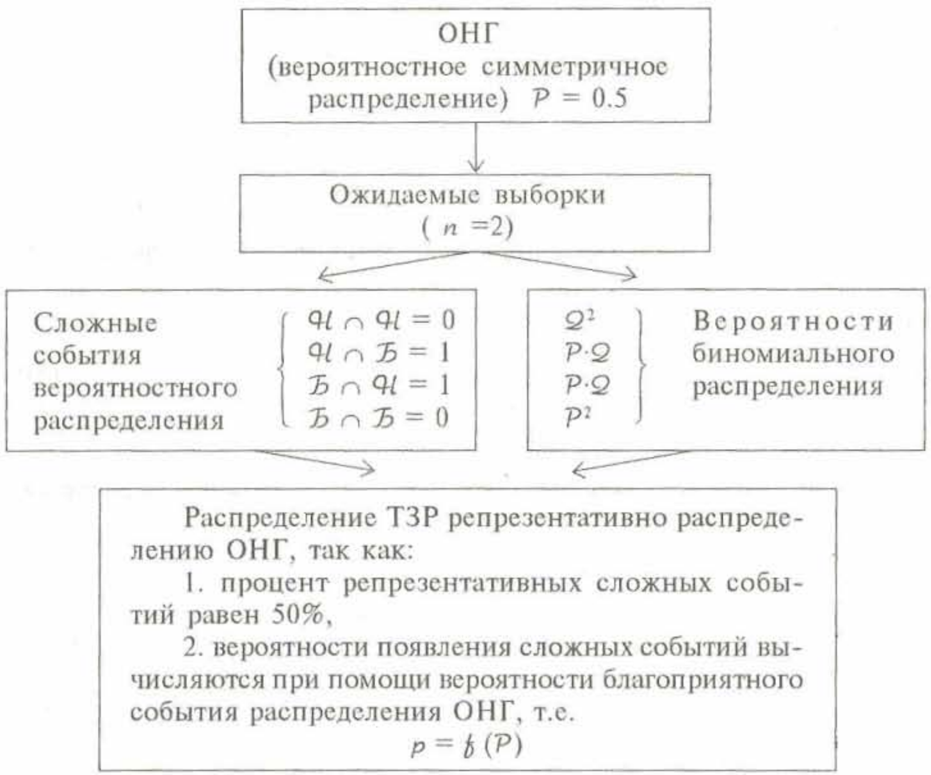


Рис. 3.4.2

Ит

предел ак, четвертая аксиома: статистиче

тивно ение) репрезентативна ОНГ, есл

двум у распределению ТЗР. Данная репре

словиям (Рис. 3.4.1):

1. Переменная ЭЗР и переменная

2 виде репрезентативной совоку

. Сходство между двумя видами

пределения характеризуется к

цесса) должно быть равно 50%

Схо ципом репрезентативности).

помощь дство между вышеназванными

эффицию коэффициента сходства (Кс),

ентов эксцесса, если распределе

$$Kc = 1 - \frac{(k \ldots}{\ldots}$$

где

$$kr = \frac{\Sigma(x - \ldots}{n \cdot s \ldots}$$

$$kz = 3 + \frac{1 - \ldots}{n}$$

$$p = q = 0.5.$$

Если

ем коэфф распределение ЭЗР асимметри

фициентов асимметрии:

$$Kc = 1 - \frac{|s \ldots}{\ldots}$$

где

$$sk = \frac{\Sigma(x - \ldots}{n \cdot \ldots}$$

$$sk = \frac{q - \ldots}{\sqrt{n \cdot \ldots}}$$

$$p \gtrless 0.5, \ldots$$

3 Зак. 3…

3.5. Выводы

1. Однородные невидимые причины (ОНП) образуют однородное статистическое индивидуальное наблюдение, однородную статистическую совокупность (выборку) и однородную статистическую генеральную совокупность (первая аксиома).
2. Статистическое индивидуальное наблюдение, т.е. однородная неделимая выборка, есть категория вероятностная с ее прообразом сложного события. Сложное событие состоит как минимум из двух элементарных событий (причин). Это ограничение необходимо для того, чтобы сохранить ее репрезентативность в отношении ОНГ (вторая аксиома).
3. Однородная статистическая выборка должна состоять как минимум из четырех индивидуальных наблюдений для того, чтобы быть репрезентативной в отношении ОНГ. Эти однородные индивидуальные наблюдения образуются под влиянием ОНП, и вид их статистического распределения тот же, что и вид вероятностного распределения ОНГ; а их средние равны (третья аксиома).
4. Статистическая совокупность данных (выборка) репрезентативна ОНГ, если выполняются следующие условия (четвертая аксиома):

а) Сходство между видом распределения ЭЗР и репрезентативным распределением ТЗР равно 50% и более.

б) Переменные ЭЗР и ТЗР должны иметь один прообраз репрезентативной совокупности сложных событий.

В заключение следует отметить два момента. Во-первых, все четыре аксиомы основаны на принципе монизма: вселенная дискретна и конечна. Во-вторых, главная задача аксиом сводится к проверке однородности статистической совокупности, а следовательно, и к проверке необоснованных предпосылок традиционной статистики.

4. Однородные невидимые выборки

Для проверки репрезентативной однородности (РО) статистических данных, т.е. для расчета критерия РО, нам необходимо разрешить две задачи:

1. Разработка метода для определения функциональной связи между вероятностями двух благоприятных событий ($\mathcal{P}$, p).
2. Группировка наборов невидимых выборок по типам и характеристика репрезентативных типов.

4.1. Функциональная связь между ОНГ и набором ОНВ

Предположим, что мы располагаем двумя репрезентативными наборами ОНВ, извлеченными из ОНГ (Рис. 4.1.1). Анализ этих наборов выполним в два этапа.

Первый этап. Каждый из двух наборов ОНВ состоит из четырех сложных событий (СС). Запишем эти СС в теоретических значениях репрезентативности (ТЗР).

Второй этап. Вычислим вероятности появления каждого СС, применяя $\mathcal{P}$, и построим вероятностные распределения СС и ТЗР. Вероятностное распределение ТЗР есть вероятностное распределение Бернулли, которое характеризуется вероятностью благоприятного события p. Эта вероятность находится в функциональной связи с вероятностью благоприятного события ОНГ ($\mathcal{P}$). Установление функциональной связи между p и $\mathcal{P}$ производится на основе репрезентативного распределения ТЗР следующим образом (Рис. 4.1.1).

Запишем p и q в вероятностных терминах ОНГ:

$$q = (\mathcal{Q}^2 + \mathcal{P}^2)h \quad (1)$$

$$p = 2\mathcal{P}\mathcal{Q}h, \quad (2)$$

где $$h^* = \frac{1}{\Sigma p(\mathcal{Y})}$$

Из равенства (1) определим h:

$$h = \frac{q}{\mathcal{Q}^2 + \mathcal{P}^2} \quad (3)$$

После подстановки (3) в (2) получим:

$$p = \frac{2\mathcal{P}\mathcal{Q}q}{\mathcal{Q}^2 + \mathcal{P}^2} \quad (4)$$

Путем простых преобразований устанавливаем функциональную связь между $\mathcal{P}$ и p:

$$\mathcal{P} = \frac{-2 + \sqrt{4 - 8p}}{-4} \quad (5)$$

* Расчет h производится для соблюдения следующего условия: $q + p = 1$

ОНГ	
$\mathcal{X}$	$\mathcal{P}(\mathcal{X})$
Ч	$0.8=\mathcal{Q}$
Б	$0.2=\mathcal{P}$

Совокупности СС

Ч ∩ Ч = 0
Б ∩ Б = 0
Ч ∩ Б = 1
Б ∩ Ч = 1

Ч ∩ Ч = 0
Ч ∩ Б = 1
Ч ∩ Б = 1
Ч ∩ Б = 1

Вероятностные распределения СС

x	$p(x)$
Ч ∩ Ч	$\mathcal{Q}^2 = 0.64$
Б ∩ Б	$\mathcal{P}^2 = 0.04$
Ч ∩ Б	$\mathcal{Q}\cdot\mathcal{P} = 0.16$
Б ∩ Ч	$\mathcal{P}\cdot\mathcal{Q} = 0.16$
Итого	1.00

x	$p(x)$	$p(x)\cdot h$
Ч ∩ Ч	$\mathcal{Q}^2 = 0.64$	0.571
Ч ∩ Б	$\mathcal{Q}\cdot\mathcal{P} = 0.16$	0.143
Ч ∩ Б	$\mathcal{Q}\cdot\mathcal{P} = 0.16$	0.143
Ч ∩ Б	$\mathcal{Q}\cdot\mathcal{P} = 0.16$	0.143
Итого	1.12	1.00

Вероятностные распределения ТЗР

y	$p(y)$
0	$\mathcal{Q}^2+\mathcal{P}^2 = 0.68 = q$
1	$2\mathcal{P}\cdot\mathcal{Q} = 0.32 = p$
Итого	1.00

y	$p(y)$	$p(y)\cdot h$
0	$\mathcal{Q}^2 = 0.64$	$0.57 = q$
1	$3\mathcal{Q}\cdot\mathcal{P} = 0.48$	$0.43 = p$
Итого	1.12	1.00

$$\mathcal{P} = \frac{-2+\sqrt{4-8p}}{-4} = 0.2, \qquad \mathcal{P} = \frac{p}{3q+p} = 0.2,$$

$$h = \frac{1}{\Sigma p(y)} = 1 \qquad h = \frac{1}{\Sigma p(y)} = 0.893$$

Рис. 4.1.1

Вычислим $\mathcal{P}$, применяя (5) и $p = 0.32$:

$$\mathcal{P} = \frac{-2+\sqrt{4-8\cdot 0.32}}{-4}.$$

Так определяется функциональная связь между $\mathcal{P}$ и p.

4.2. Анализ наборов ОНВ

Мы располагаем 15 наборами ОНВ, извлеченными из ОНГ (Рис. 4.2.1). Первые 9 наборов не являются репрезентативными, следующие 6 наборов - репрезентативные согласно принципу репрезентативности. Не являющиеся репрезентативными наборы делятся на два типа: А и В. Процент репрезентативных сложных событий в этих типах равен 0 и 25. Данные типы наборов образуются под влиянием отрицательного фактора, нарушающего гармонию соотношения между репрезентативными и не являющимися репрезентативными сложными событиями.

Репрезентативные наборы делятся на три типа: С, D, Е. Процент репрезентативных сложных событий в наборе типа С равен 50. Данный тип набора формируется под влиянием случайного фактора, который сохраняет гармонию между репрезентативными и не являющимися репрезентативными СС.

Процент репрезентативных СС в наборах типа D и Е равен 75 и 100. Эти наборы образуются под влиянием положительного фактора, увеличивающего процент репрезентативных СС. В дальнейшем мы будем анализировать наборы с процентом репрезентативных СС меньше 100, т.е. типы С и D. Тип набора Е анализироваться не будет, так как колеблемость ТЗР здесь отсутствует.

Таким образом, из 14 наборов ОВН только 5 репрезентативные (№10-14). Для этих 5 репрезентативных ОНВ построим вероятностные распределения по двум признакам: СС и ТЗР (Рис. 4.2.2 - 4.2.4). Вероятности этих распределений вычисляются по трем заданным значениям $\mathcal{P}$: $\mathcal{P} < 0.5$, $\mathcal{P} = 0.5$, $\mathcal{P} > 0.5$. В результате этих расчетов вероятности благоприятных событий распределений ТЗР (p) характеризуются четырьмя конкретными значениями (0.00, 0.25, 0.50, 0.75), а также неконкретными, находящимися в интервалах: 0.0-0.5, 0.5-1.0. По найденным значениям p и заданным величинам $\mathcal{P}$ рассчитаны функциональные связи между p и $\mathcal{P}$ (Рис. 4.2.2 - 4.2.4).

Заключительный анализ выполненных выше расчетов представлен в табличной форме, где все 5 наборов репрезентативных вероятностных распределений ТЗР (РВР-ТЗР) разгруппированы по двум признакам: $\mathcal{P}$ и p.

(Табл. 4.2.1). Эта таблица применяется для определения репрезентативности статистической выборки при заданных значениях P и p.

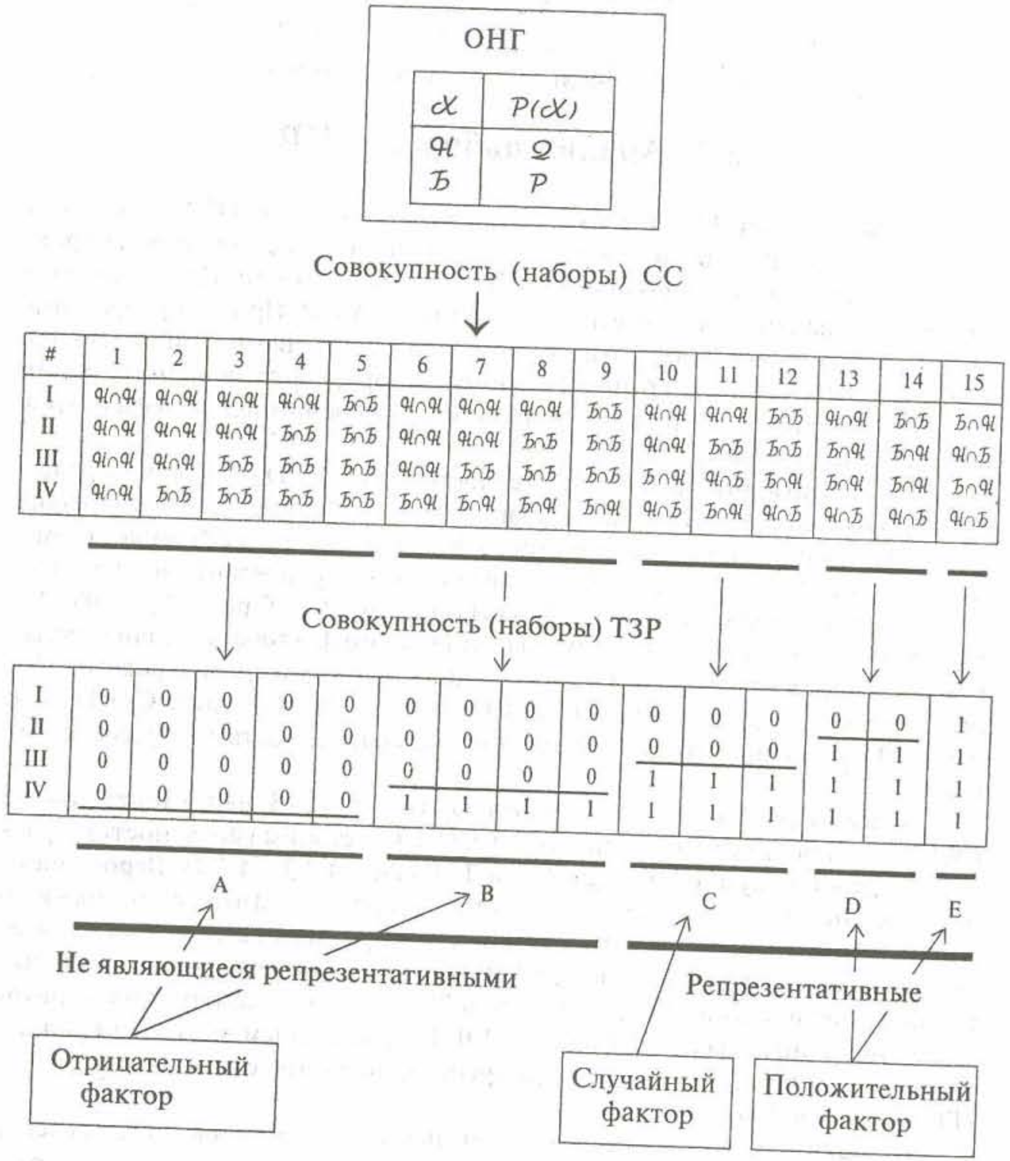


Рис. 4.2.1

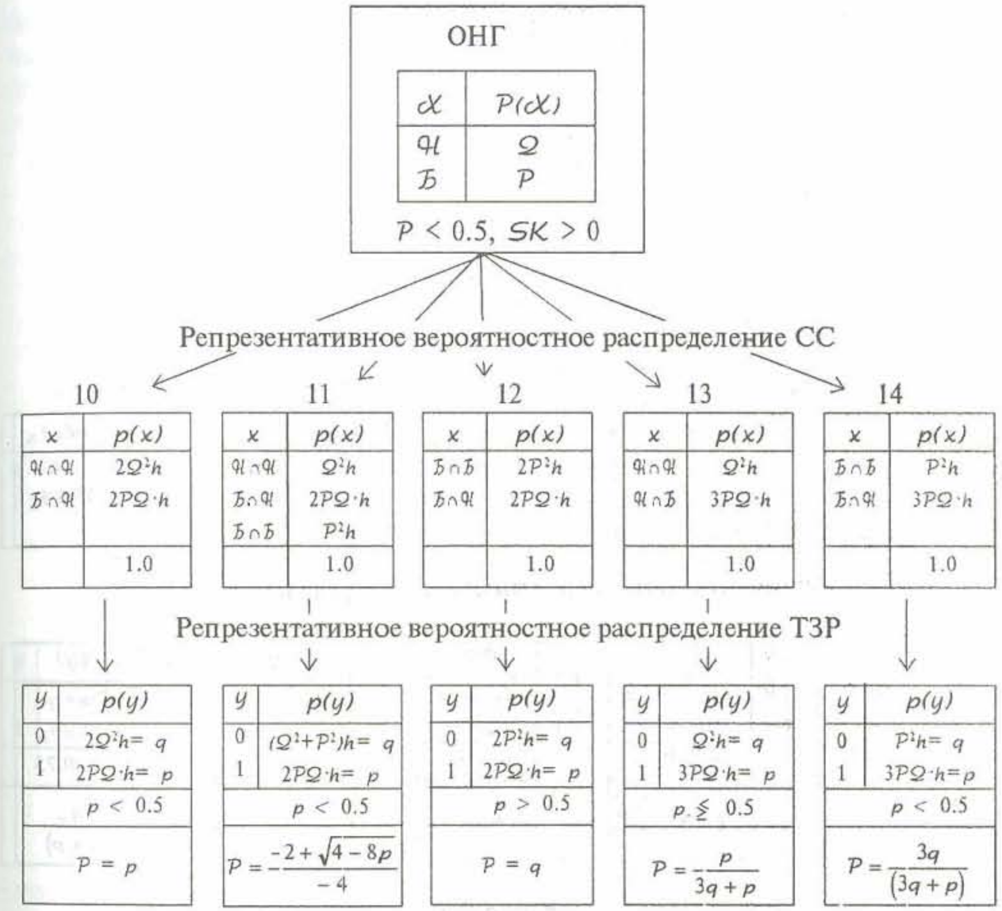


Рис. 4.2.2

4.3. Выводы

Наборы СС, извлеченные из ОНГ, делятся на два типа: репрезентативные и не являющиеся репрезентативными. Репрезентативные наборы СС состоят из 50% и более репрезентативных выборок. Эти наборы есть результат влияния случайного фактора, создающего равное соотношение между репрезентативными и не являющимися репрезентативными выборками; и положительного фактора, увеличивающего долю репрезентативных выборок.

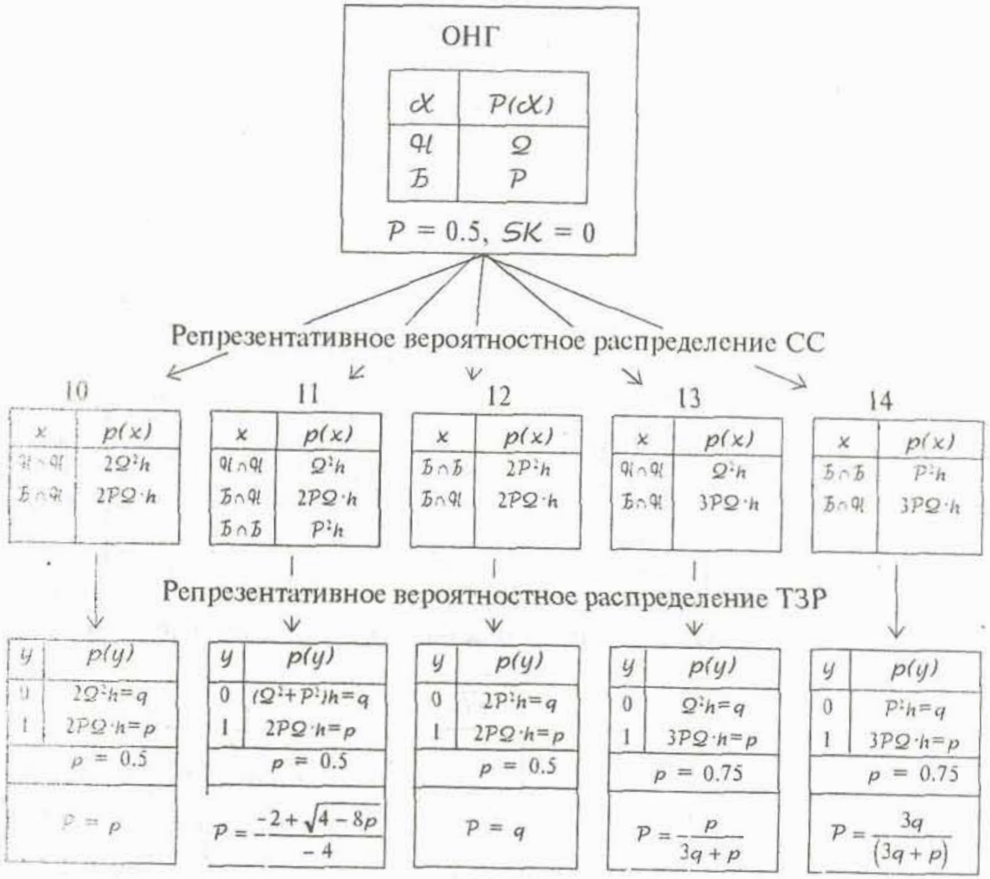


Рис. 4.2.3

Не являющиеся репрезентативными наборы СС состоят из менее чем 50% репрезентативных выборок. Эти наборы есть результат влияния отрицательного фактора.

Репрезентативные вероятностные распределения СС и ТЗР (РВР-СС, РВР-ТЗР) репрезентативны ОНГ, так как они удовлетворяют следующим условиям:

1. Переменная РВР-СС состоит из репрезентативного набора СС.
2. Зависимость p от P функциональная.

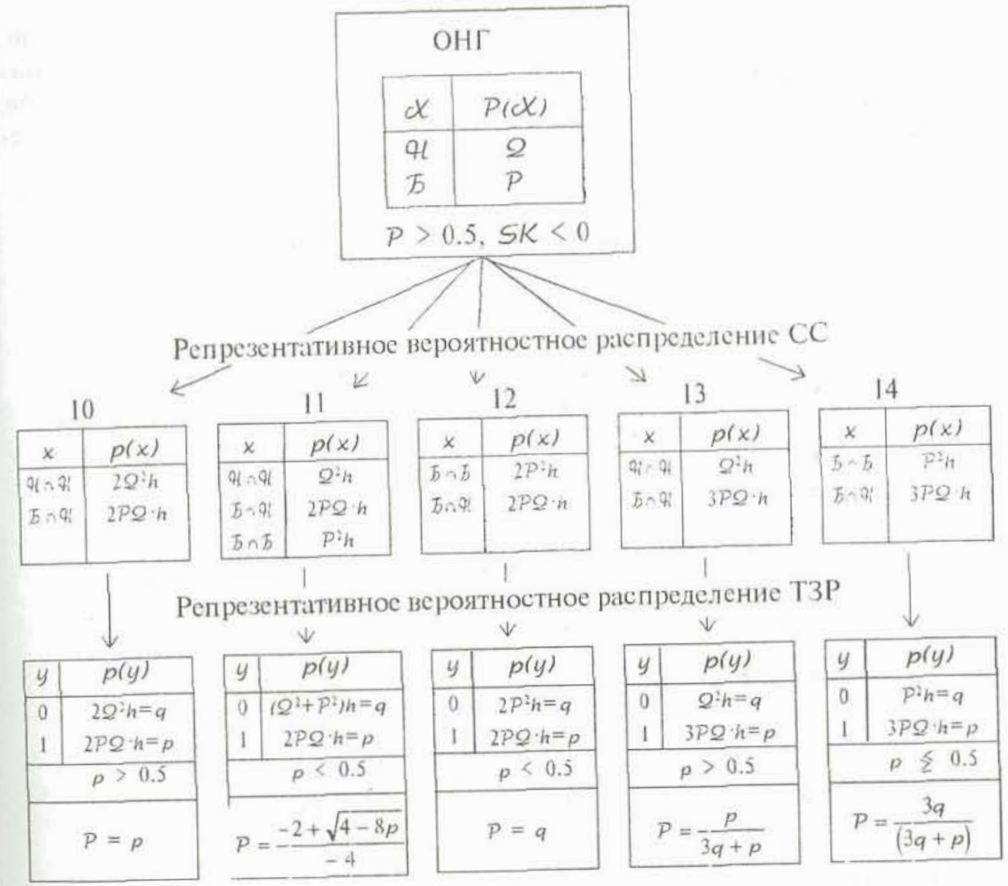


Рис. 4.2.4

5. Однородная статистическая совокупность

Однородная невидимая генеральная совокупность состоит из однородных невидимых причин. Доля благоприятных невидимых причин в общем числе всех причин есть вероятность благоприятного события или объяснительный фактор. Этот фактор образует вероятностное распределение СС, которое возникает в результате извлечения СС из ОНГ. Данное распределение репрезентативно ОНГ, если вероятности СС вычисляются с помощью P и набор СС репрезентативен ОНГ. Правомерно назвать РВР-СС ненаблюдаемой однородной совокупностью данных.

Наблюдаемая однородная совокупность данных есть статистическая переменная, которая образуется в результате существенного влияния объяснительного фактора и несущественного влияния прочих неизвестных факторов. Объяснительный фактор определяет тип однородного наблюдаемого распределения.

Таблица 4.2.1

Репрезентативное распределение ТЗР

(а)

$\mathcal{P}$ \ p	$p=0.00$	$p=0.25$	$p=0.50$	$p=0.75$
$\mathcal{P} < 0.5$	-	-	13	-
$\mathcal{P} = 0.5$	-	-	10, 11, 12	13, 14
$\mathcal{P} > 0.5$	-	-	14	-

(б)

$\mathcal{P}$ \ p	$p < 0.5$ $(sk > 0)$	$p > 0.5$ $(sk < 0)$
$\mathcal{P} < 0.5$ $(\mathcal{SK} > 0)$	10, 11, 13, 14	12, 13
$\mathcal{P} > 0.5$ $(\mathcal{SK} < 0)$	11, 12, 14	10, 13, 14

Тип этого распределения характеризуется коэффициентом асимметрии

$$sk = \frac{\Sigma(x - \bar{x})^3}{n \cdot s^3},$$

если распределение скошенное. Если распределение симметричное, применяется коэффициент эксцесса:

$$kr = \frac{\Sigma(x - \bar{x})^4}{n \cdot s^4}.$$

Для определения однородности статистических данных вычисляется критерий репрезентативной однородности (КРО). Расчет КРО основан на проверяемой предпосылке: наблюдаемые данные репрезентативны ОНГ. Если эта

предпосылка верна, то (согласно третьей аксиоме) однородные данные с положительной асимметрией формируются под влиянием благоприятного события ОНГ, вероятность которого меньше 0.5, т.е. $\mathcal{P} < 0.5$ (Рис. 5.1).

Схема проверки однородной репрезентативности статистических данных

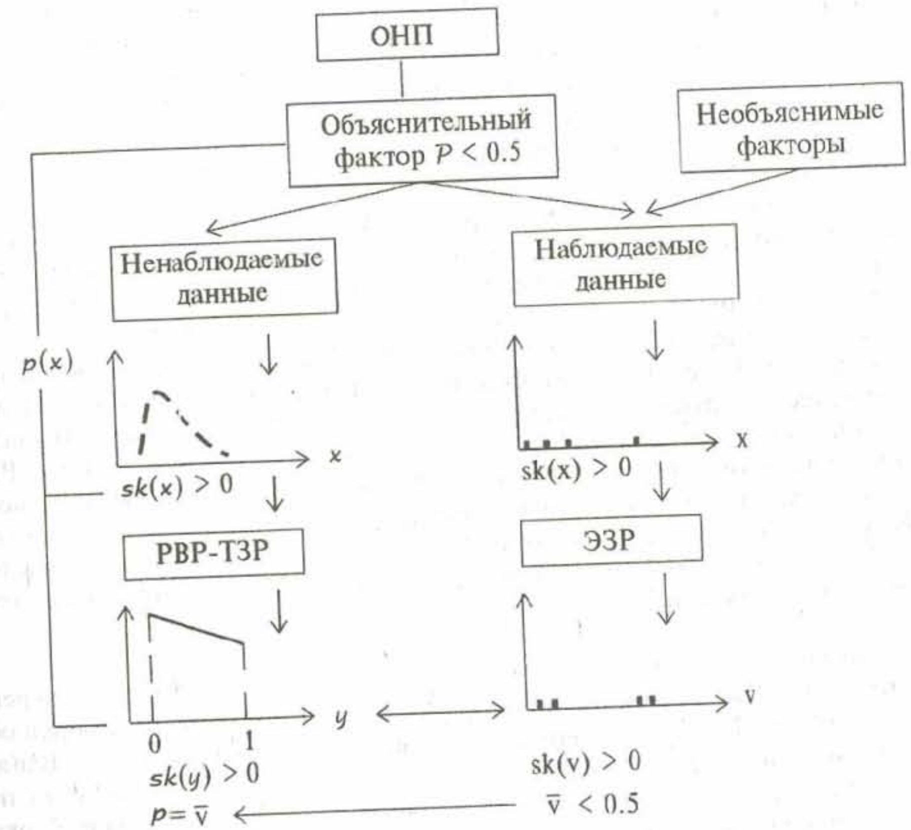


Рис. 5.1

Для проверки нашей предпосылки наблюдаемые данные преобразовываются в эмпирические значения репрезентативности (ЭЗР), которые должны удовлетворять требованиям трех принципов: порядка, сходства и соответствия.

Требование принципа порядка: ЭЗР должны находиться в пределах интервалов ТЗР.

Требование принципа сходства: между распределениями ЭЗР и ТЗР сходство должно быть существенное, т.е. оно должно достигать 50% и более.

Требование принципа соответствия: переменные ЭЗР и ТЗР должны соответствовать одному прообразу в виде репрезентативного набора сложных событий.

Каждый из вышеназванных трех принципов имеет свою специфическую задачу. Так, задача принципа порядка - определить тип репрезентативного набора сложных событий, который соответствует переменной ЭЗР. Задача принципа сходства - рассчитать степень репрезентативности распределения ЭЗР в отношении репрезентативного распределения ТЗР (Рис. 5.1.). Задача принципа соответствия - определить тип репрезентативного набора сложных событий (РНСС), который служит прообразом для переменной распределения ТЗР.

Как уже было сказано выше, расчет КРО выполняется при условии справедливости проверяемой предпосылки. Эта предпосылка предполагает, что распределение ЭЗР, так же как и распределение наблюдаемых данных (непреобразованное), должно быть одномодальным. Эти распределения являются одномодальными, если распределение ЭЗР репрезентативно распределению ТЗР. Для оценки этой репрезентативности коэффициент сходства (Кс) вычисляется с помощью формул 3.4.1 или 3.4.2. Интерпретация этих вычислений зависит от типа РНСС, т.е. от типа С или от типа D. Если РНСС представлен типом С, распределение ЭЗР должно быть репрезентативно распределению ТЗР при условии (согласно принципу последовательности), что Кс равен или больше 0.5. Если РНСС представлен типом D, распределение ЭЗР должно быть репрезентативно распределению ТЗР при условии, что Кс равен или больше 0.75.

Выводы. Статистическая выборка является однородной, если она репрезентативна ОНГ. Распределение статистических данных такой выборки образуется под существенным влиянием объяснительного фактора (P). Влияние необъяснимых (неизвестных) факторов не является существенным. Если же влияние неизвестных факторов существенно, то статистическая выборка не пригодна для анализа, так как она неоднородная. Для оценки пригодности статистической выборки рассчитывается КРО. Этим критерием проверяются требования трех принципов: порядка, сходства и соответствия. Выполнение требований этих трех принципов свидетельствует о том, что данная статистическая выборка репрезентативна ОНГ.

Концепция КРО лаконично может быть сформулирована так: статистическая выборка репрезентативна ОНГ, если распределение ЭЗР репрезентативно распределению ТЗР, которое репрезентативно ОНГ.

"Всякий кто желает пользоваться
научными заключениями,
применяя их к реальным фактам,
не может отвергать их основные принципы
и не допускать определенную степень
реальности их объектов,
не оказавшись в противоречии
даже с самим собой."

А.А. Марков

Часть III
Критерий репрезентативной однородности одномерных статистических совокупностей

6. Негрупповые статистические данные ($P \neq 0.5$)

6.1. Введение

Согласно второй аксиоме, любое статистическое индивидуальное наблюдение есть сложное событие, извлеченное из ОНГ. Совокупность же статистических наблюдений есть совокупность сложных событий. Можно ли считать эту совокупность репрезентативной в отношении ОНГ? Если ответ на данный вопрос утвердительный, то наша совокупность статистических наблюдений однородная.

Предположим, что наша совокупность статистических наблюдений однородная. Для проверки этой предпосылки необходимо:

1. Преобразовать наблюдаемые данные в эмпирические значения репрезентативности (ЭЗР).
2. Проверить сходство между распределениями (наблюдаемыми и ненаблюдаемыми).
3. Проверить соответствие между переменными двух распределений (наблюдаемых и ненаблюдаемых).

Проверка нашей предпосылки производится в результате анализа условий трех принципов: порядка, сходства и соответствия (Рис. 6.1.0).

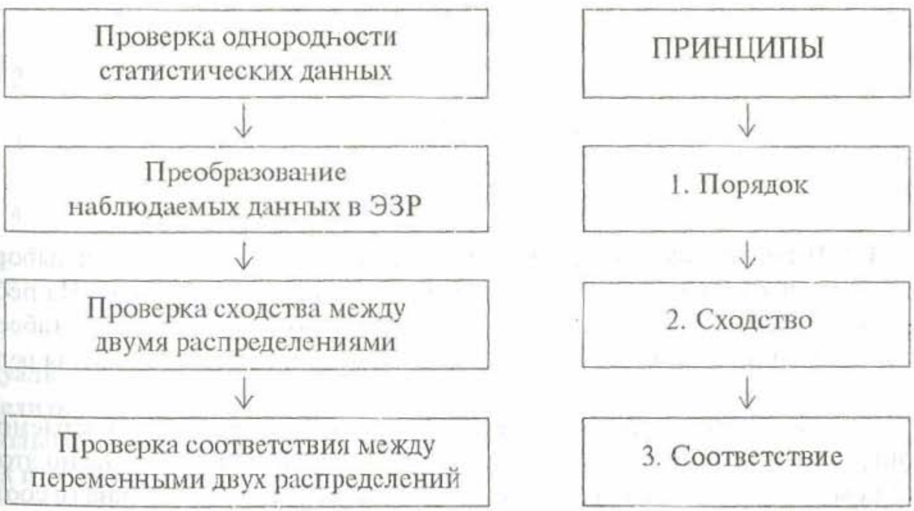


Рис. 6.1.0

Пусть наша выборка состоит из четырех семей, проживающих в квартирах со следующим числом спален (x): 1, 2, 3, 5 (Табл. 6.1.1). Распределение этих семей по числу спален репрезентативно распределению ОНГ, если оно одномодальное и формируется под влиянием одного существенного фактора (дохода или размера семьи), влияние же других факторов не существенно. Проверка репрезентативной однородности (РО) выборки производится путем анализа условий трех принципов: порядка, сходства и соответствия.

Таблица 6.1.1

Выборка

#	x
1	1
2	2
3	3
4	5

$$\bar{x} = \frac{\Sigma x}{n} = 2.75,$$

$$s = \sqrt{\frac{\Sigma\left(x - \bar{x}\right)^2}{n}} = 1.479,$$

$$sk(x) = \frac{\Sigma\left(x - \bar{x}\right)^3}{n \cdot s^3} = 0.43.$$

6.1.1. Принцип порядка. Проверка однородности статистической выборки производится на базе ЭЗР. Расчет ЭЗР выполняется в два этапа. На первом этапе шкала "g" (относительные величины) заменяет шкалу "x" (абсолютные величины). На втором этапе шкала "v" (эмпирические значения репрезентативности, ЭЗР) заменяет шкалу "g" (Рис. 6.1.1).

Вышеназванные этапы замен начинаются с формулировки проверяемой предпосылки: статистическая выборка репрезентативна ОНГ. Согласно этой предпосылке, выборочная средняя ($\bar{x} = 2.75$) также репрезентативна (в соответствии с третьей аксиомой), т.е. она равна средней ОНГ, а в терминах ТЗР она равна единице. Степень репрезентативности определяется и для индиви-

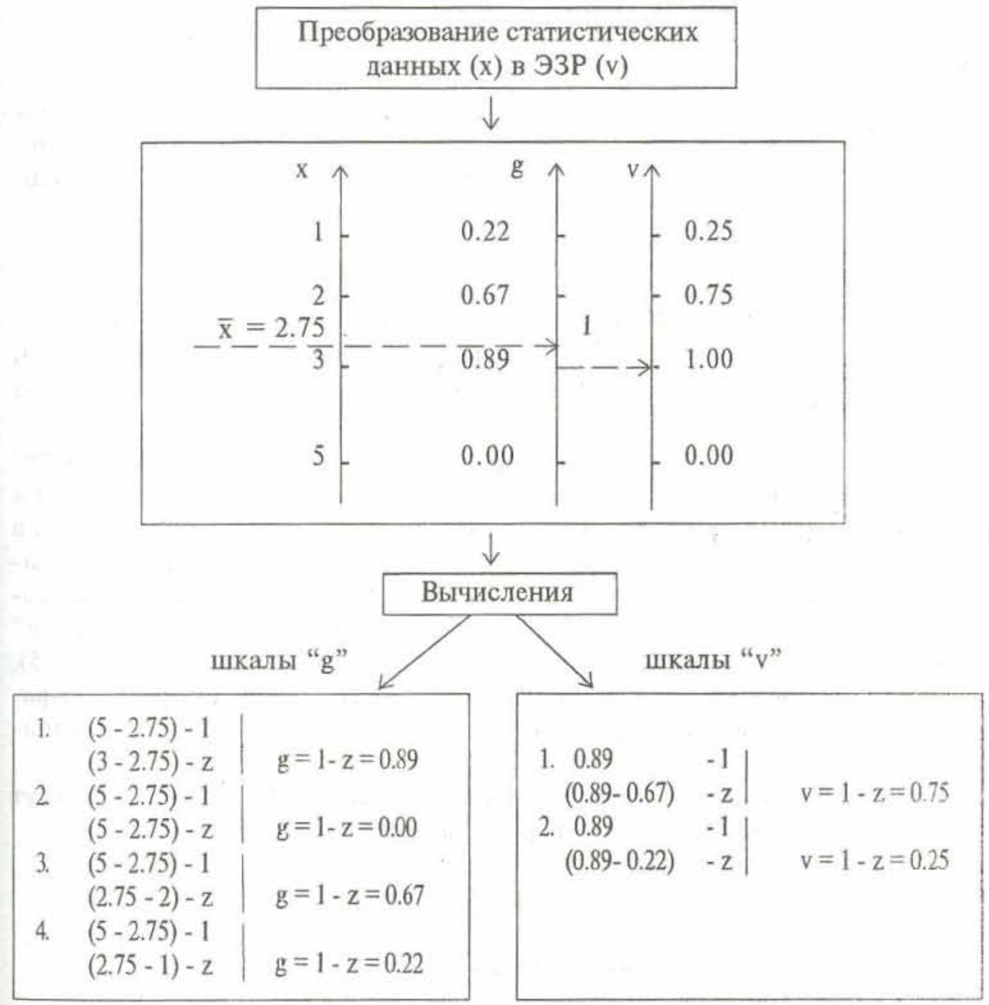


Рис. 6.1.1

дуального статистического наблюдения. В данном случае мы должны определить степень репрезентативности наблюдения (x = 5), находящегося в наибольшей удаленности от средней ($\bar{x} = 2.75$). Расчет Кс для x = 5 производится по следующим формулам:

$$Кс = 1 - \frac{|d_1 - d_2|}{d_2},$$

$$Kc = 1 - \frac{|0.25 - 2.25|}{2.25} = 0.11,$$

где $d_l = |\bar{x} - x_{min}| = |2.75 - 3| = 0.25$ *(данное отклонение считается несущественным, поскольку x = 3 находится в наименьшей удаленности от $\bar{x}$ = 2.75),*

$$d_2 = |\bar{x} - x_{max}| = |2.75 - 5| = 2.25.$$

Так как Kc < 0.5, то индивидуальное наблюдение (x = 5) нерепрезентативно, т.е. в терминах ТЗР оно равно нулю. Расстояние между $\bar{x}$ (2.75) и x_{max}(5) обозначается единицей и применяется для измерения расчета шкалы “g” (Рис. 6.1.1).

Таков первый этап анализа. Второй этап анализа - замена шкалы “g” шкалой “v” для достижения полной сопоставимости со шкалой ТЗР. Условия сопоставимости выполняются, если индивидуальные значения колеблются в пределах 0 и 1 включительно. Для выполнения этого условия g_{max} (0.89) заменяется 1, так как оно образуется под влиянием репрезентативного сложного события $\mathcal{A} \cap \mathcal{B}$. В соответствии с этой заменой остальные значения “g” (0.67 и 0.22) пересчитываются в единицах измерения шкалы “v” (0.75 и 0.25).

Следует отметить, что конфигурация распределения по шкале “v” сохраняется прежней, т.е. асимметрия распределения по шкалам “g” и “v” одинаковая.

Эмпирические значения репрезентативности (ЭЗР) (шкала “v”), могут быть рассчитаны и в табличной форме (Табл. 6.1.2).

Таблица 6.1.2

Вычисления ЭЗР (v)

x	$\lvert x - \bar{x}\rvert = d$ $\bar{x}$ = 2.75	$1 - \frac{d}{d_{max}} = g$ d_{max} = 2.25	$\frac{g}{g_{max}} = v$ g_{max} = 0.89	ТЗР (тип С)
1	1.75	0.22	0.25	0
2	0.75	0.67	0.75	1
3	0.25	0.89	1.00	1
5	2.25	0.00	0.00	0

Вычисленные ЭЗР пригодны для последующего анализа статистической выборки, если они удовлетворяют условию принципа порядка. Это условие удовлетворяется, если ЭЗР распределены в пределах интервалов ТЗР (Табл. 6.1.3, примеры №1 и 2).

Таблица 6.1.3

Пример №1

Интервалы ТЗР (тип С)		ЭЗР			
0		0	0	0	0
0	0.0 - 0.50	0.5	0.2	0.1	0
1	0.51 - 1.00	0.51	0.99	1	0.52
1		1	1	1	1

Пример №2

Интервалы ТЗР (тип D)		ЭЗР		
0		0	0	0
1	0.51 - 1.00	0.51	0.60	0.99
1	0.51 - 1.00	0.99	1	1
1		1	1	1

В первом примере ЭЗР распределены в пределах интервалов типа С. Во втором примере ЭЗР распределены в пределах интервалов типа D.

В нашем примере принцип порядка соблюдается (Табл. 6.1.4), так как вычисленные ЭЗР соответствуют типу С (Рис. 4.2.1).

Таблица 6.1.4

Принцип порядка

ТЗР	ЭЗР
0	0.00
0 [0.0 - 0.5]	0.25
1 [0.5 - 1.0]	0.75
1	1.00

$$\bar{v} = \frac{\Sigma v}{n} = 0.50, \quad s = 0.3953, \quad kr(v) = \frac{\Sigma(v - \bar{v})^4}{4 \cdot s^4} = 1.36$$

6.1.2. Принцип сходства. Выполнение условия этого принципа крайне необходимо для установления сходства между распределениями ЭЗР и ТЗР.

Проверка принципа сходства сопровождается решением трех проблем: проблема сопоставимости между двумя распределениями, проблема выбора статистического показателя для расчета степени сходства между двумя распределениями, проблема размера интервала сходства.

Первая проблема. Репрезентативное распределение ТЗР есть ожидаемое распределение, представленное в виде вероятностного распределения Бернулли. Вид этого распределения обусловлен условием $p = \bar{v}$ (Табл. 6.1.5). Это условие обязательное. Оно гарантирует полную сопоставимость двух совокупностей, так как размах их колебаний также одинаковый: от 0 до 1.

Таблица 6.1.5

Распределение ТЗР

y	$p(y)$
0	$0.50 = q$
1	$0.50 = p$
Итого	1.00

$$p = \bar{v} = 0.5,$$

$$kz(y) = 3 + \frac{1 - 6pq}{pq} = 1.0$$

Вторая проблема. Сходство между двумя распределениями, ЭЗР и ТЗР, измеряется коэффициентами асимметрии или эксцесса. В нашем примере применяется коэффициент эксцесса, так как sk(y) = 0.

Третья проблема. Размер интервала сходства (Ис) устанавливается в соответствии с правилом, сформулированным принципом репрезентативности. С учетом этого правила, сходство между распределениями ЭЗР и ТЗР считается существенным, если отклонение коэффициента эксцесса ЭЗР [kr(v)] от коэффициента эксцесса ТЗР [$kz(y)$] равно 50% или меньше.

Принимая во внимание вышесказанное, расчет коэффициента сходства в нашем примере выполняется с применением следующей формулы:

$$\text{Kc} = 1 - \frac{(\text{kr} - kz)}{kz},$$

$$\text{Kc} = \frac{(1.36 - 1)}{1} = 0.64$$

Коэффициент сходства равен 64%. Следовательно, отклонение распределения ЭЗР от распределения ТЗР не существенно, оно равно 36%.

6.1.3. Принцип соответствия. Анализируя наш пример, следует сделать следующие выводы:

1. Переменная распределения ЭЗР соответствует типу С.
2. Распределения ЭЗР и ТЗР существенно не различаются.

Следующий шаг - проверка условия принципа соответствия. Это условие формулируется так: переменные распределений ЭЗР и ТЗР соответствуют одному типу набора однородных невидимых выборок, т.е. СС.

В нашем примере мы располагаем двумя совокупностями данных: наблюдаемые и ненаблюдаемые. Наблюдаемая совокупность данных представлена в абсолютных величинах (x) и в относительных величинах (ЭЗР). Ненаблюдаемая совокупность данных представлена в условных единицах (y) и в относительных величинах (ТЗР) (см. Рис. 6.1.2).

Принцип соответствия

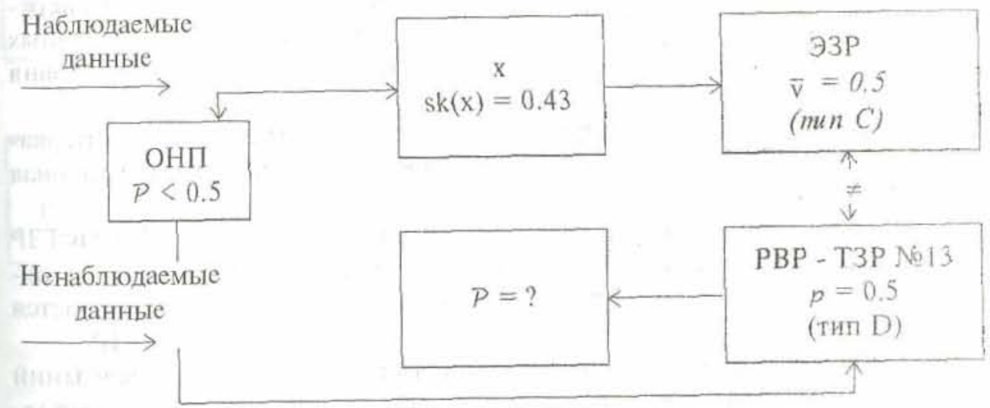


Рис. 6.1.2

Согласно нашей проверяемой предпосылке, наблюдаемая совокупность данных репрезентативна ОНГ. Известно также, что ОНГ характеризуется ве-

роятностью благоприятного события (P). *В связи с этим мы можем утверждать* (принимая во внимание третью аксиому), что если $P < 0.5$, то наблюдаемая совокупность данных имеет положительную асимметрию. Если $P > 0.5$, то наши данные имеют отрицательную асимметрию. Если $P = 0.5$, то они симметричны. В анализируемом примере наши данные имеют положительную асимметрию, sk = 0.43. Если эти данные репрезентативны ОНГ, то $P < 0.5$ (согласно третьей аксиоме).

Зная вероятности благоприятных событий ОНГ и ТЗР ($P < 0.5$, $p = 0.5$), определяем тип набора СС, который формирует переменную распределения ТЗР. *Этот тип набора* соответствует типу D (Табл. 4.2.1). Таким образом, наблюдаемые и ненаблюдаемые данные соответствуют разным типам наборов СС. Следовательно, переменные распределений ЭЗР и ТЗР также соответствуют разным типам наборов СС. Из этого следует, что условие принципа соответствия не выполняется и наша предпосылка о репрезентативности неверна, ее следует отвергнуть. Поэтому правомерно заявить, что распределение семей по числу спален в квартире образуется под влиянием двух или более факторов, т.е. данное распределение не одномодальное, а следовательно, и неоднородное (Рис. 6.1.2).

6.1.4. Выводы. Логика расчета КРО заключается в следующем. Наш *отправной пункт - предпосылка о репрезентативности статистических данных* относительно ОНГ. Данная предпосылка верна, если выполняются условия трех принципов.

1. Условие принципа порядка: ЭЗР расположены в пределах интервалов ТЗР. В нашем примере это условие выполняется, и переменная ЭЗР соответствует переменной типа С.
2. Условие принципа сходства: между распределениями ЭЗР и ТЗР нет существенного различия. В нашем примере это условие выполняется, коэффициент эксцесса распределения ЭЗР не отличается существенно от коэффициента эксцесса распределения ТЗР.
3. Условие принципа соответствия: тип переменных распределений ЭЗР и ТЗР *одинаковый. В нашем примере это условие не выполняется*; переменная распределения ЭЗР соответствует типу С, а переменная распределения ТЗР соответствует типу D. Следовательно, наша предпосылка о репрезентативной однородности выборки не подтверждается.

Известно три типа ТЗР, которые являются эталоном для сравнения с ЭЗР (Рис. 6.1.3):

1. Статистические данные не являются репрезентативными ОНГ, если ЭЗР соответствуют типам А или В.
2. Статистические данные являются репрезентативными ОНГ, если ЭЗР соответствуют типу Е.
3. Статистические данные могут быть или не быть репрезентативными, если ЭЗР соответствуют типам С и D.

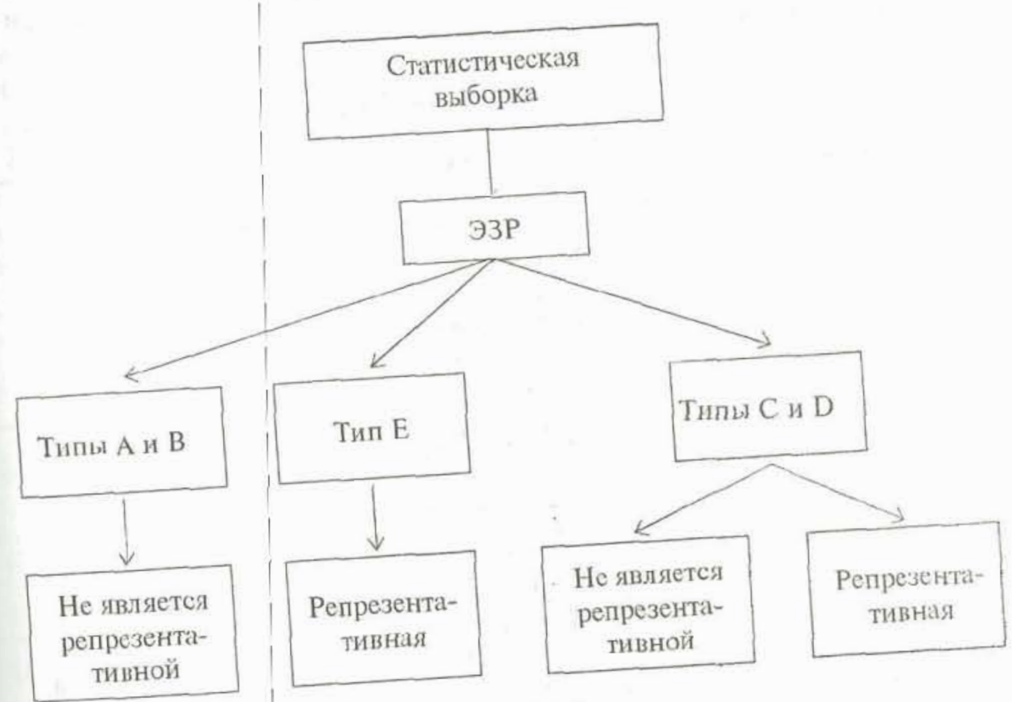


Рис. 6.1.3

В заключение перечислим основные преимущества КРО:

1. КРО применим к большим и малым выборкам, минимальный размер выборки равен четырем.
2. Статистические выводы КРО конкретны в отношении индивидуальной выборки.
3. КРО создает научную основу для правильного применения статистического анализа как-то: регрессионного, корреляционного, табличного.

6.2. Негрупповые статистические данные
$(P = 0.5)$

Пусть наша выборка состоит из четырех служащих разного роста (Табл. 6.2.1). Если распределение служащих по росту симметричное, то можно быть уверенным в том, что данное распределение не формируется под влиянием какого-либо одного или нескольких существенных факторов, например, национальность и пол.

Таблица 6.2.1

Выборка

#	Рост (в дюймах) x
1	66.4
2	67.7
3	69.7
4	71.7

$$\bar{x} = \frac{\Sigma x}{n} = 68.875,$$

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n}} = 2.01,$$

$$sk = \frac{\Sigma(x - \bar{x})^3}{4 \cdot s^3} = 0.19$$

Наша предпосылка - служащие распределены симметрично по росту. Данная предпосылка проверяется путем расчета критерия нормальности. Этим термином мы будем пользоваться условно, так как ФНК имеет дело только с дискретными распределениями. Расчет критерия проходит проверку условий трех принципов: порядка, сходства и соответствия.

6.2.1 Принцип порядка. Условие принципа порядка выполняется, так как вычисление ЭЗР (Табл. 6.2.2) расположены в пределах интервалов ТЗР (Табл. 6.2.3). В данном случае ТЗР представлены типом С.

Таблица 6.2.2

ЭЗР

x	$\lvert x - \bar{x} \rvert = d$ $\bar{x} = 68.875$	$1 - \frac{d}{d_{max}} = g$ $d_{max} = 2.825$	$\frac{g}{g_{max}} = v$ $g_{max} = 0.708$	ТЗР
66.4	2.475	0.124	1.175	0
67.7	1.175	0.584	0.825	1
69.7	0.825	0.708	1.000	1
71.7	2.825	0.000	0.000	0

Таблица 6.2.3

Принцип порядка

ТЗР	ЭЗР
0	0.000
0.0 - 0.5	0.175
0.5 - 1.0	0.825
1	1.000

$$\bar{v} = 0.5,\ sk = 0,\ kr = \Sigma(x - \bar{x})^4 \div 4 \cdot s^4 = 1.16$$

6.2.2. Принцип сходства. Сходство между двумя распределениями, ЭЗР (Табл. 6.2.3) и ТЗР (Табл. 6.2.4), измеряется при помощи коэффициентов эксцесса, так как коэффициент асимметрии равен нулю. Расчет коэффициента сходства выполняется по следующей формуле:

$$Kc = 1 - \frac{\lvert kr - kz \rvert}{kz},$$

$$Kc = 1 - \frac{\lvert 1.16 - 1 \rvert}{1} = 0.84$$

Условие принципа сходства выполнено, Кс равен 0.84. Это означает, что нет существенных различий в конфигурации двух распределений, ЭЗР и ТЗР.

Таблица 6.2.4

Распределение ТЗР

y	$p(y)$
0	0.5
1	$0.5 = p$

$$p = \bar{v}, \quad sk = (q-p) \div \sqrt{q \cdot p} = 0,$$

$$kz(y) = 3 + (1 - 6 \cdot p \cdot q) \div p \cdot q = 1$$

6.2.3. Принцип соответствия. Условие этого принципа выполняется, так как оба вида переменных двух распределений, ЭЗР и ТЗР, соответствуют одному типу С (Рис. 6.2.1). Следовательно, наша предпосылка верна: распределение служащих по росту симметричное.

Принцип соответствия

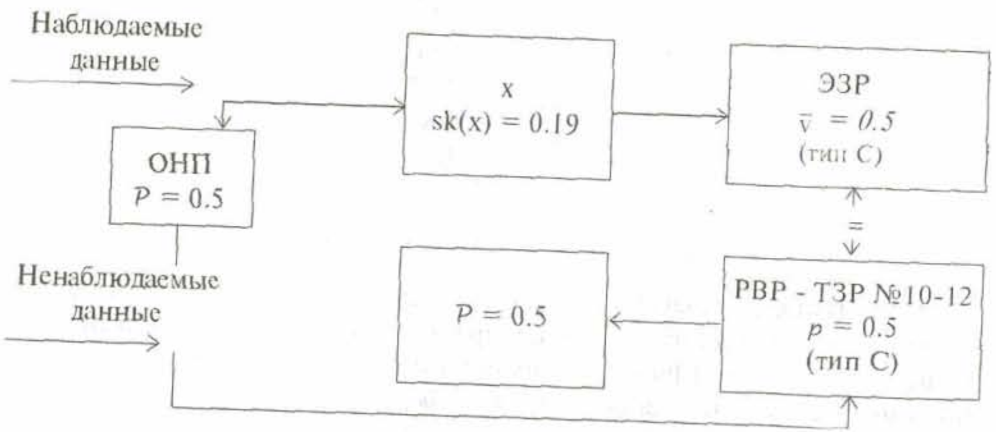


Рис. 6.2.1

6.2.4. Выводы. Проверка нормального распределения статистических данных начинается с формулировки проверяемой гипотезы: статистические данные распределены симметрично. Проверка этой предпосылки производится на основе анализа условий трех принципов: порядка, сходства и соответствия. В нашем примере условия этих принципов выполняются. Следовательно, наша предпосылка верна: распределение служащих по росту репрезентативно ОНГ с $P = 0.5$.

6.3. Негрупповые статистические данные ($n = 4$, $sk \neq 0$, $\bar{v} \neq 0.5$)

Пусть выборка состоит из четырех семей с разным доходом за год в тыс. долларов (Табл. 6.3.1). Данная выборка репрезентативна ОНГ, если распределение семей по доходу одномодальное, т.е. оно формируется под влиянием одного существенного фактора, например, образование. Проверку репрезентативной однородности выборки начинаем с анализа условия принципа порядка.

Таблица 6.3.1

Выборка

#	x
1	13.5
2	15.8
3	19.6
4	31.4

$$\bar{x} = \frac{\Sigma x}{n} = 20.075,$$

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = 6.892,$$

$$sk(x) = \frac{\Sigma(x-\bar{x})^3}{n \cdot s^3} = 0.82$$

6.3.1. Принцип порядка. Условие этого принципа выполняется, так как вычисленные ЭЗР (Табл. 6.3.2.) расположены в пределах интервалов ТЗР (Табл. 6.3.3). В данном примере ТЗР представлены типом С.

6.3.2. Принцип сходства. Сходство между двумя распределениями, ЭЗР (Табл. 6.3.3) и ТЗР (Табл. 6.3.4), измеряется при помощи коэффициентов асимметрии, так как $\bar{v}$ не равно 0.5. Это свидетельствует о том, что распределение ТЗР асимметрично. Коэффициент этого распределения равен -0.08

(Табл. 6.3.4). Сравнивая данный коэффициент с коэффициентом асимметрии распределения ЭЗР, который равен -0.15 (Табл. 6.3.3), мы вычисляем коэффициент сходства:

$$Kc = 1 - \frac{|sk(v) - sk(y)|}{|sk(y)|},$$

$$Kc = 1 - \frac{|-0.15 + 0.08|}{|-0.08|} = 0.125$$

Таблица 6.3.2

ЭЗР

x	$\|x - \bar{x}\| = d$ $\bar{x} = 20.075$	$1 - \frac{d}{d_{max}} = g$ $d_{max} = 11.325$	$\frac{g}{g_{max}} = v$ $g_{max} = 0.96$	ТЗР
13.5	6.575	0.42	0.44	0
15.8	4.275	0.62	0.65	1
19.6	0.475	0.96	1.00	1
31.4	11.325	0.00	0.00	0

Таблица 6.3.3

Принцип порядка

ТЗР	ЭЗР
0.00	0.00
0.00 - 0.50	0.44
0.50 - 1.00	0.65
1.00	1.00

$$\bar{v} = \frac{\Sigma v}{n} = 0.52, \; s = 0.362, \; sk(v) = \frac{\Sigma(v - \bar{v})^3}{4 \cdot s^3} = -0.15$$

Таблица 6.3.4

Распределение ТЗР

y	$p(y)$
0	$0.48 = q$
1	$0.52 = p$
Итого	1.00

$$p = \bar{v} = 0.52,$$

$$sk(y) = \frac{q - p}{\sqrt{q \cdot p}} = -0.08$$

Коэффициент сходства равен 12.5%. Так как Kc < 50%, мы должны отвергнуть гипотезу о том, что распределение ЭЗР одномодальное и репрезентативно распределению ТЗР. Мы также должны отвергнуть и гипотезу о том, что наша выборка однородная.

6.3.3. Выводы. Для проверки репрезентативной однородности (РО) выборки мы должны вначале сформулировать проверяемую предпосылку: выборка однородная. Анализируя сходство между распределениями ЭЗР и ТЗР, применяются коэффициенты асимметрии, так как $\bar{v} \neq 0.5$. В нашем примере вычисленный коэффициент сходства меньше 0.5. Поэтому мы отвергаем предпосылку о РО выборки.

6.4. Негрупповые статистические данные ($n = 8$, $sk \neq 0$, $\bar{v} \neq 0.5$)

Пусть наша выборка состоит из восьми семей с разным доходом за год в тыс. долларов (Табл. 6.4.1). Требуется проверить РО выборки. Решение данной задачи выполняется в результате проверки условий трех принципов: порядка, сходства, соответствия.

6.4.1. Принцип порядка. Условие этого принципа выполняется, так как вычисленные ЭЗР (Табл. 6.4.2.) расположены в пределах интервалов ТЗР (Табл. 6.4.3). В данном примере ТЗР представлены типом С.

Таблица 6.4.1

Выборка

#	x
1	20.1
2	20.7
3	21.5
4	24.6
5	24.7
6	28.8
7	35.9
8	36.1

$\bar{x} = 26.55$
$s = 6.04$
$sk = 0.62$

Таблица 6.4.2

ЭЗР

x	$\lvert x - \bar{x} \rvert = d$ $\bar{x} = 26.55$	$1 - \frac{d}{d_{max}} = g$ $d_{max} = 9.55$	$\frac{g}{g_{max}} = v$ $g_{max} = 0.81$	ТЗР
20.1	6.45	0.32	0.40	0
20.7	5.85	0.39	0.48	0
21.5	5.05	0.47	0.58	1
24.6	1.95	0.80	0.99	1
24.7	1.85	0.81	1.00	1
28.8	2.25	0.76	0.94	1
35.9	9.35	0.02	0.02	0
36.1	9.55	0.00	0.00	0

6.4.2. Принцип сходства. Сходство между двумя распределениями, ЭЗР (Табл. 6.4.3) и ТЗР (Табл. 6.4.4), измеряется при помощи коэффициентов асимметрии, так как $\bar{v} > 0.5$.

Таблица 6.4.3

Принцип порядка

ТЗР	ЭЗР
0	0.00
0 - 0.5	0.02
0 - 0.5	0.40
0 - 0.5	0.48
0.5 - 1.0	0.58
0.5 - 1.0	0.94
0.5 - 1.0	0.99
1.0	1.00

$$\bar{v} = \frac{\Sigma v}{n} = 0.55,$$

$$s = \sqrt{\frac{\Sigma(v - \bar{v})^2}{n}} = 0.381, \quad n = 8,$$

$$sk = \frac{\Sigma(v - \bar{v})^3}{n \cdot s^3} = -0.19, \quad n = 8.$$

Таблица 6.4.4

Распределение ТЗР

y	$p(y)$
0	0.45
1	$0.55 = p = \bar{v}$
Итого	1.00

$$sk = \frac{q - p}{\sqrt{q \cdot p}} = -0.20$$

Вычисленный коэффициент сходства:

$$Kc = 1 - \frac{|sk(v) - sk(y)|}{|sk(y)|},$$

$$Kc = 1 - \frac{|-0.19 + 0.20|}{|-0.20|} = 0.95$$

- больше 0.5. Следовательно, сходство между рядами распределения существенно и условие принципа сходства выполняется.

6.4.3. Принцип соответствия. Условие этого принципа выполняется, так как переменные распределений, ЭЗР и ТЗР, имеют одинаковый тип С (Рис. 6.4.1). Данный вывод основан на следующем анализе.

Принцип соответствия

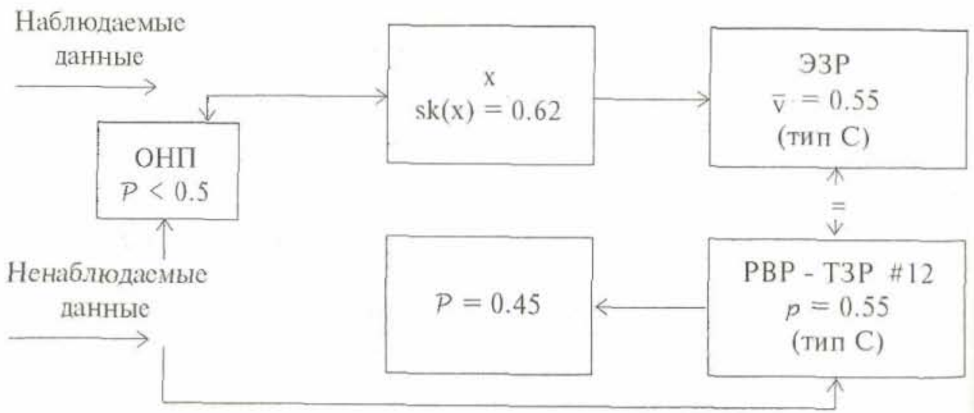


Рис. 6.4.1

В нашем примере выборка характеризуется положительной асимметрией, sk(x) = 0.62. Расчет коэффициента асимметрии основан на проверяемой предпосылке: выборка репрезентативна ОНГ. Из этого следует, что наша выборка образуется под влиянием объяснительного фактора ($\mathcal{P}$) и, согласно третьей аксиоме, этот фактор меньше 0.5. Вероятность благоприятного события распределения ТЗР также известна, она равна 0.55 (Табл. 6.4.4).

Зная $\mathcal{P}$ и p можно определить вид репрезентативного вероятностного распределения ТЗР (РВР-ТЗР). Для $\mathcal{P} < 0.5$ и $p > 0.5$ возможны следующие два вида РВР-ТЗР: №12 и13 (Табл. 4.2.1, Рис. 6.4.1). Из этих двух видов мы должны отобрать один. Процесс отбора производится в два этапа.

Первый этап: расчет двух коэффициентов асимметрии ОНГ. Первый коэффициент асимметрии вычисляется с применением $\mathcal{P} = 0.45$, расчет которого произведен на основе 12 вида распределения ТЗР, который соответствует типу С (Рис. 4.2.2):

$$\mathcal{SK} = \frac{\mathcal{Q} - \mathcal{P}}{\sqrt{\mathcal{Q} \cdot \mathcal{P}}} = 0.20, \quad \text{где } \mathcal{P} = q = 0.45$$

Второй коэффициент вычисляется с применением $\mathcal{P} = 0.26$, расчет которого произведен на основе 13 вида распределения ТЗР, который соответствует типу D (Рис. 4.2.2):

$$\mathcal{SK} = \frac{\mathcal{Q} - \mathcal{P}}{\sqrt{\mathcal{Q} \cdot \mathcal{P}}} = 1.09, \quad \text{где } \mathcal{P} = \frac{p}{3q + p} = 0.26$$

Второй этап: выбор одного из двух типов ТЗР, С и D. Выбор типа ТЗР осуществляется согласно третьей аксиоме: вид репрезентативного статистического распределения определяется видом распределения ОНГ. Поэтому при выборе типа ТЗР мы руководствуемся наименьшим отклонением между коэффициентом асимметрии статистического ряда распределения (sk = 0.62) и двумя коэффициентами асимметрии ряда распределения ОНГ ($\mathcal{SK}(\#12) = 0.20$, $\mathcal{SK}(\#13) = 1.09$):

$$| sk(x) - \mathcal{SK}(\#13)| > | sk(x) - \mathcal{SK}(\#12)|$$

$$| 0.62 - 1.09| > | 0.62 - 0.20|$$

$$| -0.47| > | 0.42 |$$

Так как отклонение между sk(x) и $\mathcal{SK}(\#12)$ наименьшее, тип С выбран для характеристики переменной ряда распределения ТЗР. Поэтому переменные двух распределений ЭЗР и ТЗР, характеризуются одним и тем же типом С. Следовательно, условие принципа соответствия выполняется, и наша предпосылка (выборка репрезентативна ОНГ) верна. Также верно и то, что выборка одномодальная, т.е. она формируется под влиянием одного существенного фактора.

6.4.4. Выводы. Проверка репрезентативной однородности выборки применима, если размер выборки делится на четыре. В нашем примере размер выборки равен восьми. Данная выборка репрезентативна ОНГ, так как условия трех принципов (порядка, сходства и соответствия) выполняются. Проверка условия принципа соответствия усложнилась в связи с необходимостью анализа двух типов РВР-ТЗР и выбора одного из них.

7. Групповые статистические данные
$(n = 32,\ sk > 0,\ \overline{v} \neq 0.5)$

Большая часть исследований базируется на групповых данных. Проверка РО этих данных зависит от достоверности нашей предпосылки: индивидуальные статистические данные равномерно распределены в пределах групп, т.е. в интервалах.

Наша выборка состоит из 32 семей, распределенных по пяти группам дохода (Табл. 7.1). Для проверки РО выборки групповые данные необходимо

Таблица 7.1

Выборка

x ($1000)	x	f
5.4 - 10.6	8.0	5
10.6 - 15.8	13.2	11
15.8 - 21.0	18.4	11
21.0 - 26.2	23.6	3
26.2 - 31.4	28.8	2
Итого	-	32

$$\overline{x} = \frac{\Sigma xf}{\Sigma f} = 16.125,$$

$$s(x) = \sqrt{\frac{\Sigma(x-\overline{x})^2 f}{f}} = 5.51,$$

$$sk(x) = \frac{\Sigma(x-\overline{x})^3 f}{\Sigma f \cdot [s(x)]^3} = 0.47$$

представить в виде негрупповых данных. В нашем примере число негрупповых данных может быть равно: 4, 8, 12, 16, 20, 24, 28, 32. В целях упрощения расчетной процедуры размер выборки ограничен четырьмя семьями. Вторичная выборка представлена доходами следующих условных семей: 8, 16, 24, 32. Доходы этих четырех семей вычислены методом накопленных частот (Табл. 7.2). Данный метод предполагает, что семьи равномерно распределены в группах по доходу. Предполагается также, что доходы отобранных четырех семей репрезентативны доходам 32 семей. Эта предпосылка непроверяемая. Однородность доходов четырех семей - вторая, но уже проверяемая предпосылка. Проверка РО этих семей начинается с принципа порядка.

Таблица 7.2

Вторичная выборка
(негрупповые данные)

$x_{(i)}$	i	f	f(%)	Σf(%)
5.4 - 10.6	5.2	5	0.1562	0.1562
10.6 - 15.8	5.2	11	0.3438	0.5000
15.8 - 21.0	5.2	11	0.3438	0.8438
21.0 - 26.2	5.2	3	0.0937	0.9375
26.2 - 31.4	5.2	2	0.0625	1.0000
Итого	-	32	1.0000	-

$$x_t \to \left|\begin{matrix} 5.2 - 0.3438 \\ (0.25 - 0.1562) \end{matrix}\right| \quad x_1 = x_{(i)} + x_t = 10.6 + 1.42 = 12.02$$

$$x_t \to \left|\begin{matrix} 5.2 - 0.3438 \\ (0.5 - 0.1562) \end{matrix}\right| \quad x_2 = x_{(i)} + x_t = 10.6 + 5.20 = 15.80$$

$$x_t \to \left|\begin{matrix} 5.2 - 0.3438 \\ (0.75 - 0.50) \end{matrix}\right| \quad x_3 = x_{(i)} + x_t = 15.8 + 3.78 = 19.58$$

$$x_t \to \left|\begin{matrix} 5.2 - 0.0625 \\ 1.0 - 0.9375) \end{matrix}\right| \quad x_4 = x_{(i)} + x_t = 26.2 + 5.20 = 31.40$$

7.1. Принцип порядка. Условие этого принципа выполняется, так как вычисленные ЭЗР (Табл. 7.3) расположены в пределах интервалов ТЗР (Табл. 7.4). В данном примере ТЗР представлены типом С.

Таблица 7.3

ЭЗР

x	$\lvert x-\bar{x}\rvert = d$ $\bar{x} = 19.7$	$1-\frac{d}{d_{max}} = g$ $d_{max} = 11.70$	$\frac{g}{g_{max}} = v$ $g_{max} = 0.98$	ТЗР
12.02	7.68	0.34	0.35	0
15.80	3.90	0.67	0.68	1
19.58	0.12	0.98	1.00	1
31.40	11.70	0.00	0.00	0

$sk(x) = 0.71$

Таблица 7.4

Принцип порядка

ТЗР	ЭЗР
0	0.00
0 - 0.5	0.35
0.5 - 1.0	0.68
1.0	1.00

$$\bar{v} = \frac{\Sigma v}{n} = 0.5075, \quad s = 0.3724,$$

$$sk(v) = \frac{\Sigma(v-v)^3}{n \cdot s^3} = -0.05$$

7.2. Принцип сходства. Сходство между двумя распределениями, ЭЗР (Табл. 7.4) и ТЗР (Табл. 7.5), измеряется при помощи коэффициентов асимметрии, так как $\bar{v} > 0.5$ (Табл. 7.4).

Вычисленный коэффициент сходства:

$$Kc = 1 - \frac{\lvert sk(v) - sk(y)\rvert}{\lvert sk(y)\rvert},$$

$$Kc = 1 - \frac{\lvert -0.05 + 0.03\rvert}{\lvert -0.03\rvert} = 0.33$$

- меньше 0.5. Следовательно, сходство между нашими рядами распределения не существенно, и поэтому условие принципа сходства не выполняется. В связи с этим мы должны отвергнуть предпосылку о репрезентативности выборки в отношении ОНГ.

Таблица 7.5

Распределение ТЗР

y	$p(y)$
0	0.4925
1	0.5075
Итого	1.0000

$$sk(y) = \frac{q-p}{\sqrt{q \cdot p}} = -0.03,$$

$$p = \bar{v} = 0.5075$$

7.3 Выводы. Проверка РО групповых статистических данных производится при условии, что индивидуальные данные распределены равномерно в группах по доходу. При соблюдении данного условия групповые статистические данные трансформируются в индивидуальные условные данные, образующие вторичную выборку. Эта выборка и подвергается проверке в отношении РО. В заключение следует заметить, что трансформация групповых данных основана на непроверяемой предпосылке. Поэтому все выводы, основанные на ней, подлежат сомнению.

*"Априорная истина очевидна
по своей природе и познается
путем анализа концепции."*

К.И. Луис

Часть IV
Критерий репрезентативной однородности н-мерных статистических совокупностей

8. Двумерные статистические таблицы частот

Наш предмет - статистическая таблица сопряженности признаков с двумя качественными переменными и четырьмя скрещивающимися классификациями, т.е. статистическая таблица размером 2x2. При помощи таких таблиц традиционная статистика устанавливает связь между двумя качественными переменными. На первом этапе формулируется нулевая гипотеза: качественные переменные независимые, т.е. связь между ними отсутствует. На втором этапе проверяется гипотеза независимости.

Для проверки гипотезы независимости переменных предлагается много различных методов, например, критерии согласия Пирсона, Фишера, Макнемара, Кохрана, Мантеля и Хаенсзеля, Фримана и Туки, а также статистика правдоподобия (Everitt, 1977). Особое место отводится критерию информации Акаике (Sakamoto and others, 1986), о нем будет идти речь ниже.

Методология всех критериев согласия (за исключением критерия Акаике) основана на философии традиционной статистики. Традиционная статистика сопоставляет наблюдаемые совместные частоты с ожидаемыми совместными частотами для того, чтобы проверить гипотезу независимости между двумя переменными. Проверку гипотезы независимости проиллюстрируем на примере критерия Пирсона:

$$\varkappa^2 = \Sigma \frac{(f - F)^2}{F} \qquad (8.1)$$

Философия традиционной статистики отрицает априорное знание генеральной совокупности. Следовательно, частоты распределения генеральной совокупности (F) должны быть неизвестными. А если это так (а это действительно так), то частоты распределения выборки (f) не могут быть проверены в отношении их согласия с частотами распределения генеральной совокупности. В чем же тогда смысл критерия согласия? В действительности же согласия как такового нет. Есть нагромождение необоснованных гипотез. Содержание этих гипотез проиллюстрируем на следующем примере.

Предположим, что формула (8.1) применяется для расчета статистического критерия нормальности. Для расчета этого критерия формулируется первая необоснованная гипотеза: генеральная совокупность нормальная. Вторая необоснованная гипотеза о выборке: выборка репрезентативна в отношении генеральной совокупности. Третья необоснованная гипотеза об условной генеральной совокупности: параметры этой генеральной совокупности равны аналогичным обобщающим показателям выборки. Четвертая гипотеза проверяемая: выборка произведена из условной генеральной совокупности. Следовательно, проверка производится не в отношении согласия между распределениями данных выборочной и генеральной совокупностей, а для выяснения гипотезы: выборка извлечена из условной нормальной генеральной совокупности. Результат проверки этой гипотезы правдоподобен только при условии, что вышеназванные необоснованные гипотезы верны.

8.1. Ненаблюдаемая двумерная таблица частот

Основная задача таблиц сопряженности признаков (с практической точки зрения) заключается в анализе влияния объяснительного фактора на зависимую переменную. Поэтому мы предлагаем заменить название "таблицы сопряженности признаков" названием "таблицы объяснительного фактора" (ТОФ).

Анализ ТОФ должен быть основан на априорном знании генеральной совокупности. Знание генеральной совокупности должно быть совместимо с пониманием вселенной со статистической точки зрения. В этом вопросе мы согласны с доктором Д. Бусу, который утверждает, что наша познаваемая и зависящая от обстоятельств вселенная характеризуется финитивной совокупностью чисел (Ghosh, 1988, стр. 23). Поэтому генеральная совокупность финитивных реальных категорий должна быть представлена дискретным распределением. Согласно принципу минимума, данное распределение есть вероятностное распределение Бернулли. Это распределение назовем однородной невидимой генеральной совокупностью (ОНГ). Эта совокупность состоит из двух видов однородных невидимых причин: благоприятных ($\mathcal{Б}$) и неблагоприятных ($\mathcal{Н}$). Однородные невидимые причины обладают следующими свойствами: они независимы, каждому типу причин присуще свое постоянное влияние на следствие и вероятность благоприятного события постоянная.

ОНГ может быть представлена в виде одномерного или н-мерного распределения. Одномерное ОНГ состоит из одной переменной. Эта переменная выполняет роль причины или следствия. Двумерное ОНГ состоит из двух переменных. Одна из них причина, а другая - следствие. Связь между этими двумя переменными - функциональная (Рис. 8.1.1). Это значит, что, во-первых, безусловное распределение частот переменной причины равно безусловному распределению частот переменной следствия. Во-вторых, влияние неучтенных причин на следствие полностью отсутствует. В результате этого со-

вместные вероятности, находящиеся под влиянием неучтенных причин, равны нулю (Рис. 8.1.1):

$$\mathcal{Н} \cap \mathcal{Б} + \mathcal{Б} \cap \mathcal{Н} = 0$$

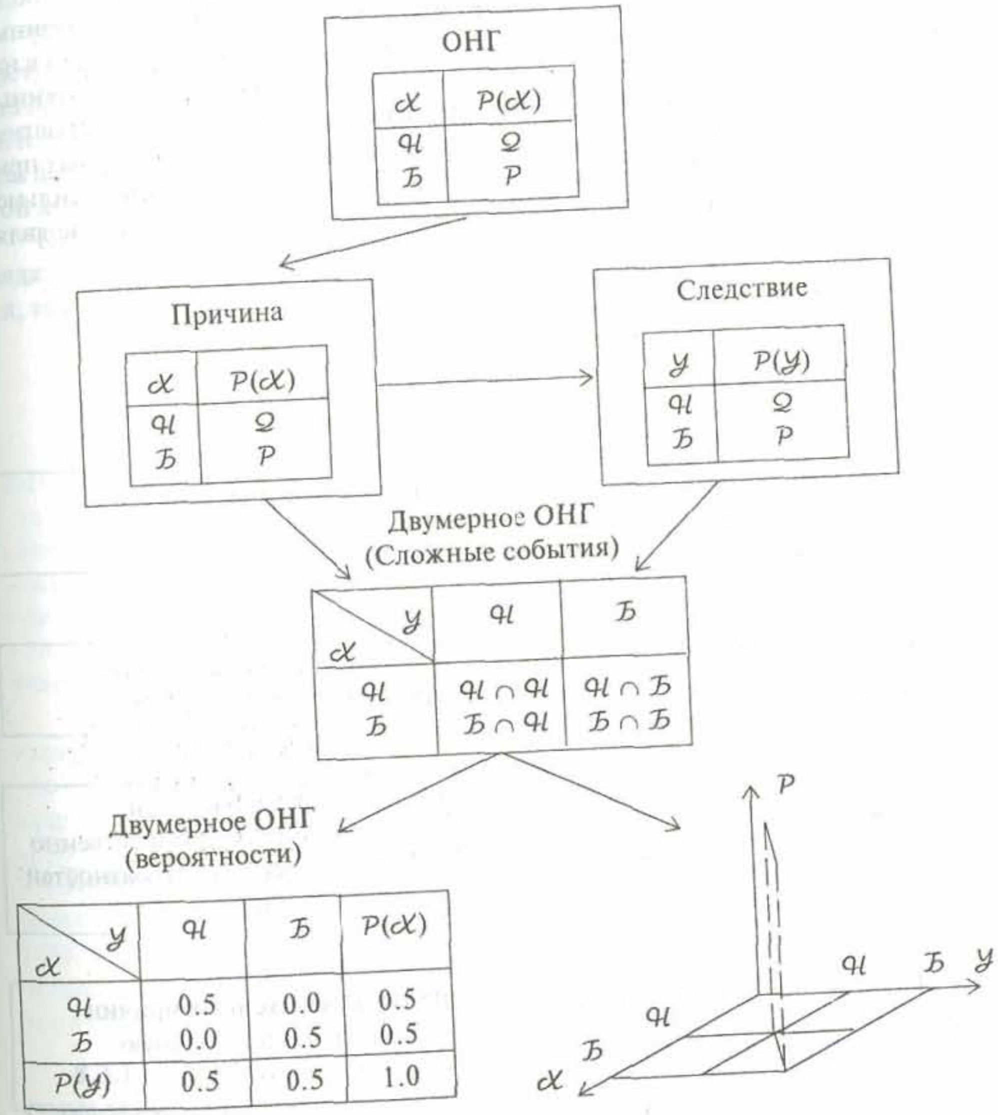


Рис. 8.1.1

8.2. Наблюдаемые двумерные таблицы частот

Известно, что совместные вероятности неучтенных причин в ненаблюдаемой двумерной таблице частот (ОНГ) равны нулю, т.е. влияние неучтенных причин полностью отсутствует (Рис. 8.2.1). Иная ситуация имеет место в наблюдаемых двумерных таблицах частот: совместные вероятности неучтенных причин не равны нулю. Эти вероятности меньше, чем совместные вероятности учтенной причины (объяснительного фактора), если влияние первых причин слабее вторых. Если влияние неучтенных причин существенно сильнее влияния учтенной причины, наблюдаемая двумерная таблица частот не является однородной.

Рис. 8.2.1

8.3. Проверка репрезентативной однородности: образование и штатная должность

Наш пример - двумерная таблица частот (Табл. 8.3.1). Данная таблица построена на основе следующей гипотезы: образование (x) - причина, штатная должность (y) - следствие. Согласно этой гипотезе, образование - объяснительный (учтенный) фактор, оказывающий влияние на получение штатной должности инспектора. Влияние неучтенных факторов на получение штатной должности инспектора не должно оказывать существенного влияния.

Проверка нашей гипотезы есть проверка условий двух принципов: порядка и сходства. Проверка условия принципа соответствия не производится, так как данное условие всегда выполняется в двумерной таблице частот.

Таблица 8.3.1

Двумерное распределение служащих
по образованию (x) и штатной должности (y)

x \ y	$\overline{B}$	B	Итого
$\overline{A}$	0.656	0.156	0.812
A	0.063	0.125	0.188
Итого	0.719	0.281	1.000

Пояснения: A - окончившие колледж

$\overline{A}$ - не окончившие колледж

B - инспекторская должность

$\overline{B}$ - не инспекторская должность

8.3.1. Принцип порядка. Условие этого принципа следующее: общая сумма совместных вероятностей, образованная под влиянием учтенной причины (образование), превышает общую сумму совместных вероятностей, образованную под влиянием неучтенных причин (Табл. 8.3.2):

$$(\overline{A} \cap \overline{B} + A \cap B) > (A \cap \overline{B} + \overline{A} \cap B) \quad (1)^*$$
$$(0.656 + 0.125) > (0.063 + 0.156)$$

Таблица 8.3.2

Принцип порядка

x \ y	$\overline{B}$	B	Итого
$\overline{A}$	0.656	0.156	0.812
A	0.063	0.125	0.188
Итого	0.719	0.281	1.000

$$(\overline{A} \cap \overline{B} + A \cap B) > (A \cap \overline{B} + \overline{A} \cap B)$$
$$0.781 > 0.219$$

8.3.2. Принцип сходства. Наша задача - определить влияние неучтенных причин на зависимую переменную, т.е. узнать, является ли это влияние существенным или нет. Влияние неучтенных причин считается несущественным, если отклонение наблюдаемого распределения от ненаблюдаемого (измеренное в коэффициентах асимметрии или эксцесса) равно 50% или меньше. Этот процент установлен в соответствии с принципом репрезентативности. Он равен проценту репрезентативных сложных событий в переменной типа С.

В нашем примере коэффициент асимметрии наблюдаемого распределения (Табл. 8.3.3) равен 1.16. Коэффициент асимметрии ненаблюдаемого однородного распределения (Табл. 8.3.4) равен 1.13. Первое распределение рассчитано по данным двумерного распределения служащих (Табл. 8.3.1). Второе распределение рассчитано по данным наблюдаемых безусловных вероятностей, распределенных по причинному признаку, т.е. по образованию (Табл. 8.3.1). Эти безусловные вероятности используются в качестве параметров ОНГ, т.е. $\mathcal{P}$ и $\mathcal{Q}$. На основе этих параметров строится вероятностное биномиальное распределение с $n = 2$, которое свободно от влияния неучтенных причин и полностью сопоставимо с наблюдаемым распределением.

* Неравенство (1) справедливо, если между причиной и следствием прямая связь. Если эта связь обратная, то следует поменять местами колонки таблицы с тем, чтобы выполнить условие (1).

Таблица 8.3.3

Наблюдаемое распределение

x	p(x)
$\overline{A} \cap \overline{B} = 0$	0.656
$\left.\begin{matrix} A \cap \overline{B} \\ \overline{A} \cap B \end{matrix}\right\} = 1$	0.219
$A \cap B = 2$	0.125

$\overline{x} = 0.469, \ s = 0.706, \ sk = 1.16$

Таблица 8.3.4

Вероятностное биномиальное распределение
($n = 2$, $p = 0.188$)

x	p(x)
0	0.659
1	0.305
2	0.036

$$sk(x) = \frac{q - p}{\sqrt{n \cdot q \cdot p}} = 1.13$$

Зная оба распределения, наблюдаемое и ненаблюдаемое, мы можем определить природу наблюдаемого распределения, т.е. его репрезентативную однородность. Для этого вычислим коэффициент сходства:

$$Kc = 1 - \frac{|sk(x) - sk(x)|}{|sk(x)|},$$

$$Kc = 1 - \frac{|1.16 - 1.13|}{|1.13|} = 0.97$$

В нашем примере коэффициент сходства равен 97%. Так как коэффициент больше 50%, то можно утверждать, что сходство между распределениями существенно. Это значит, что влияние неучтенных причин на зависимость переменной "y" от переменной "x" не существенно, и поэтому выборка однородна и репрезентативна ОНГ.

8.4. Проверка репрезентативной однородности: аскорбиновая кислота и простуда

Наш пример - двумерная таблица частот (Табл. 8.4.1). Данные для этой таблицы были собраны Т.В. Андерсоном в 1971-1972 гг. (Anderson and others, 1972). Таблица Андерсона основана на гипотезе: потребление аскорбиновой кислоты (причина, x) предотвращает простуду (следствие, y). Проверка этой гипотезы производится путем анализа условий принципа порядка и принципа сходства.

Таблица 8.4.1

Аскорбиновая кислота и простуда

Простуда y / Лечение x	Да $\overline{B}$	Нет B	Итого
Плаcибо $\overline{A}$	0.322	0.180	0.502
Аскорбинка A	0.238	0.260	0.498
Итого	0.560	0.440	1.000

8.4.1. Принцип порядка. Условие этого принципа соблюдается (Табл. 8.4.2), так как выполняется следующее неравенство:

$$(\overline{A} \cap \overline{B} + A \cap B) > (A \cap \overline{B} + \overline{A} \cap B)$$
$$(0.582 > 0.418)$$

Таблица 8.4.2

Принцип порядка

Простуда y / Лечение x	Да $\overline{B}$	Нет B	Итого
$\overline{A}$ A	0.322 0.238	0.180 0.260	0.502 0.498
Итого	0.560	0.440	1.000

$$(\overline{A} \cap \overline{B} + A \cap B) > (A \cap \overline{B} + \overline{A} \cap B)$$
$$0.582 > 0.418$$

8.4.2. Принцип сходства. Для измерения влияния неучтенных факторов строятся два распределения: наблюдаемое (Табл. 8.4.3) и ненаблюдаемое (Табл. 8.4.4) в виде вероятностного биномиального распределения. На основе этих распределений рассчитываются коэффициенты асимметрии и коэффициент сходства:

$$Kc = 1 - \frac{|sk(x) - sk(x)|}{|sk(x)|},$$

$$Kc = 1 - \frac{|0.10 - 0.006|}{0.006} = -14.7$$

Полученный результат свидетельствует о том, что условие принципа сходства не выполняется. Следовательно, выборка неоднородная, и она формируется в результате существенного влияния неучтенных причин. К этим причинам можно отнести, например, состояние здоровья индивидуума и объем потребляемой аскорбиновой кислоты. Относительно второй причины д-р. Л. Полинг писал, что простуду (y) можно лечить аскорбиновой кислотой, потребляя ее в оптимальных размерах, при этом оптимум зависит от состояния здоровья индивидуума (L. Pauling, 1986, страницы 2-8). Следует иметь в виду, что оптимальный размер потребляемой аскорбиновой кислоты (x) находится в обратной связи со здоровьем индивидуума (з). Эта обратная связь элиминирует корреляцию между "y" и "x" в случае, если влияние "з" не исключено.

Таблица 8.4.3

Наблюдаемое распределение

x	p(x)
$\overline{A}\cap\overline{B} = 0$	0.322
$\left.\begin{matrix}A\cap\overline{B}\\ \overline{A}\cap B\end{matrix}\right\} = 1$	0.418
$A\cap B = 2$	0.260

$\overline{x} = 0.938,\ s = 0.760,\ sk = 0.10$

Таблица 8.4.4

Вероятностное биномиальное распределение
($n = 2,\ p = 0.498$)

x	$p(x)$
0 1 2	0.252 0.500 0.248

$$sk(x) = \frac{q-p}{\sqrt{n\cdot q\cdot p}} = \frac{0.502-0.498}{\sqrt{2\cdot 0.502\cdot 0.498}} = 0.006$$

Традиционная статистика не проверяет однородность выборки. Поэтому очень легко сделать ошибку, применяя статистические методы к наблюдаемым данным. Так, например, М. Дайкес и П. Мейер (Dykes and Meier, 1975) исследовали проблему влияния аскорбиновой кислоты на простуду по приведенным выше данным, применяя методы традиционной статистики. В результате своих исследований они пришли к выводу, что лечение простуды аскорбиновой кислотой неэффективно. Данный вывод был бы справедлив, если бы их необоснованная гипотеза была бы верна, т.е. влияние неучтенных причин на простуду не существенно. Однако их гипотеза ложная, так как двумерная статистическая таблица частот не является однородной.

8.5. Проверка репрезентативной однородности: пол и ответы на вопрос - кем ты хочешь быть?

Наш пример - двумерная статистическая таблица частот (Табл. 8.5.1). Эта таблица была составлена и проанализирована японскими статистиками с применением метода Акаике (Sakamoto and others, 1986, стр. 125). Построение таблицы основано на следующей гипотезе: ответы на вопрос (y) зависят от пола (x). Вопрос: если бы вы могли родиться снова, желали бы вы быть мужчиной или женщиной?

Таблица 8.5.1

Пол и ответы на вопрос:
если бы вы могли родиться снова
желали бы вы стать мужчиной или женщиной?

Ответ y / Пол x	М $\overline{B}$	Ж B	Итого
М $(\overline{A})$	0.39	0.04	0.43
Ж A	0.23	0.34	0.57
Итого	0.62	0.38	1.00

Принцип порядка

$$(\overline{A}\cap\overline{B} + A\cap B) > (A\cap\overline{B} + \overline{A}\cap B)$$
$$0.73 > 0.27$$

Связь между вышеназванными двумя переменными "x" и "y" может быть измерена, если статистические данные репрезентативны двумерной ОНГ. Поэтому следует проверить репрезентативность нашей двумерной таблицы частот (Табл. 8.5.1). Проверке подвергнутся условия принципов порядка и соответствия.

8.5.1. Принцип порядка. Условие этого принципа соблюдается, так как выполняется следующее неравенство:

$$(\overline{A} \cap \overline{B} + A \cap B) > (A \cap \overline{B} + \overline{A} \cap B)$$
$$0.73 > 0.27$$

8.5.2. Принцип сходства. Условие этого принципа не выполняется (Табл. 8.5.2, 8.5.3), так как коэффициент сходства меньше 50%:

$$Kc = 1 - \frac{\left| sk - |sk| \right|}{|sk|},$$

$$Kc = 1 - \frac{\left| 0.095 - |-0.2| \right|}{|-0.2|} = 0.49$$

Данная формула применима в тех случаях, когда коэффициент асимметрии наблюдаемого распределения положительный, а коэффициент асимметрии ненаблюдаемого распределения отрицательный.

Данный результат свидетельствует о том, что влияние неучтенных причин на зависимую переменную (y) существенное, и, следовательно, двумерная таблица частот не является однородной и по ней нельзя измерять связь между полом и ответом на поставленный выше вопрос.

Таблица 8.5.2

Наблюдаемое распределение

x	p(x)
$\overline{A} \cap \overline{B} = 0$	0.39
$\left.\begin{matrix} A \cap \overline{B} \\ \overline{A} \cap B \end{matrix}\right\} = 1$	0.27
$A \cap B = 2$	0.34

$\overline{x} = 0.95$, $s = 0.85$, $sk = 0.095$

Таблица 8.5.3

Вероятностное биномиальное распределение
($n = 2$, $p = 0.57$)

x	$p(x)$
0	0.185
1	0.490
2	0.325

$$sk(x) = \frac{q - p}{\sqrt{n \cdot q \cdot p}} = -0.2$$

Японские статистики повторили ту же ошибку, которую допустил П. Мейер, делая выводы относительно связи между двумя переменными без предварительного анализа однородности статистической совокупности.

8.6. Выводы

Традиционный статистический анализ двумерной таблицы частот основан на необоснованной предпосылке: влияние неучтенных причин на зависимую переменную не существенно. Наша философия нового качества статистических данных резко отлична от философии традиционной статистики. Она преследует две цели. Во-первых, производится проверка необоснованной предпосылки традиционной статистики. Во-вторых, измеряется влияние неучтенных причин на зависимую переменную. Выполнение этих двух задач возможно только в том случае, если мы располагаем априорным знанием ОНГ.

Проверка репрезентативной однородности двумерной статистической таблицы частот производится в два этапа. На первом этапе проверяется условие принципа порядка. На втором этапе проверяется условие принципа сходства.

Данная проверка основана на сопоставлении двух распределений: наблюдаемого и ненаблюдаемого. Различие между ними объясняется влиянием неучтенных причин. Это влияние измеряется коэффициентом сходства. Если этот коэффициент равен или больше 0.5, влияние неучтенных причин несущественное. В тех случаях, когда влияние неучтенных причин существенно, статистическая выборка не является репрезентативной в отношении ОНГ; она не может быть использована для анализа связи между двумя переменными.

9. Двумерные статистические совокупности

Проверка РО причинно-следственных связей между двумя количественными переменными - тема нашего анализа. Статистическая информация для этой проверки может быть представлена в виде негрупповых и групповых данных.

9.1. Негрупповые статистические данные

Двумерная ОНГ характеризуется двумя идентичными безусловными вероятностными распределениями: распределение причины и распределение следствия (Рис. 9.1.1). Переменные этих двух распределений могут быть выражены в виде СС и ТЗР. Связь между причиной (независимая переменная) и следствием (зависимая переменная) - функциональная. Это свидетельствует о том, что влияние неучтенных причин равно нулю.

В тех случаях, когда мы имеем дело со статистической выборкой, влияние неучтенных причин не равно нулю. Однако это влияние не существенно, если выборка репрезентативна двумерной ОНГ. Такая выборка удовлетворяет двум условиям:

1. Независимая переменная репрезентативна ОНГ.
2. Зависимая переменная существенно зависит только от независимой переменной.

Проверку вышеназванных условий произведем по данным четырех семей (Табл. 9.1.1), проживающих в квартирах разного размера (x) и с разными

Таблица 9.1.1

Количество комнат (x) и расходы на жилье (y)

#	x	y (в $1000)
1	2	2.4
2	3	4.0
3	4	5.3
4	6	8.9

$\bar{x} = 3.75$ $\quad$ $\bar{y} = 5.15$

$s = 1.479$ $\quad$ $s = 2.396$

$sk = 0.43$ $\quad$ $sk = 0.55$

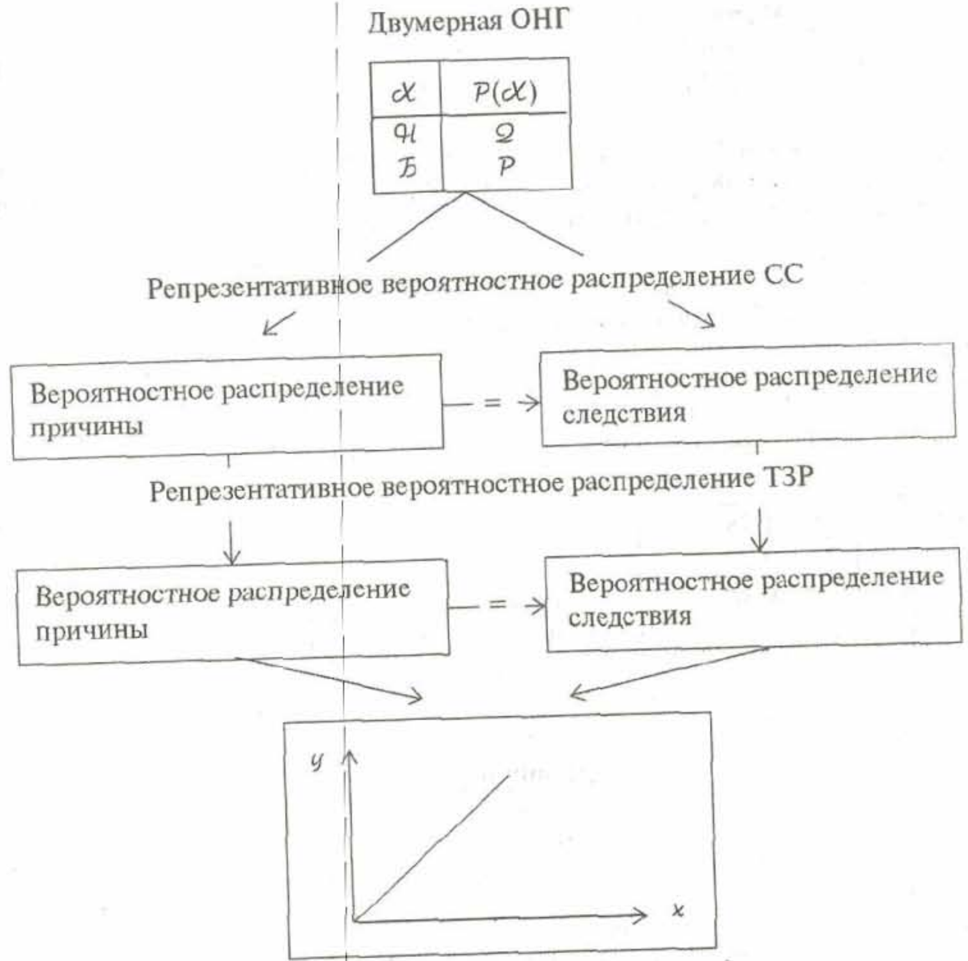


Рис. 9.1.1

расходами на жилье (y). Мы предполагаем, что расходы на жилье существенно зависят от количества комнат в квартире, а неучтенные причины не оказывают существенного влияния на "y". Наша задача - проверить эту гипотезу. Если эта гипотеза верна, то выборка репрезентативна двумерной ОНГ и однородное распределение семей по признаку "x" является единственной существенной причиной, которая оказывает влияние на распределение семей по признаку "y". Проверка нашей гипотезы производится в результате анализа вышеназванных двух условий.

9.1.1. Условие №1: *независимая переменная репрезентативна ОНГ.* Для проверки РО независимой переменной нам необходимо проанализировать выполнение условий трех принципов: порядка, сходства и соответствия.

Принцип порядка. Условие этого принципа выполняется, так как ЭЗР расположены в пределах интервалов ТЗР (Табл. 9.1.2, 9.1.3). В данном примере ТЗР представлены типом С.

Таблица 9.1.2

Эмпирические значения репрезентативности (ЭЗР)

x	$\lvert x-\bar{x}\rvert = d$, $\bar{x} = 3.75$	$1-\dfrac{d}{2.25} = g$	$\dfrac{g}{0.89} = v$	ТЗР
2	1.75	0.22	0.25	0
3	0.75	0.67	0.75	1
4	0.25	0.89	1.00	1
6	2.25	0.00	0.00	0

Таблица 9.1.3

Принцип порядка

ТЗР	v_x
0	0.00
0.0 - 0.5	0.25
0.5 - 1.0	0.75
1	1.00

$$\bar{v}_x = 0.5,\ kr = 1.36$$

Принцип сходства. Условие этого принципа выполняется, так как распределение ЭЗР (Табл. 9.1.3) не отличается существенно от распределения ТЗР (Табл. 9.1.4). Коэффициент сходства между этими распределениями равен 64%:

$$Kc = 1-\left[kr(v)-kz(y)\right] = 1-(1.36-1.00) = 0.64$$

Таблица 9.1.4

Распределение ТЗР

x	$p(x)$
0	$0.5 = q_x$
1	$0.5 = p_x$

$$\bar{v}_x = p_x = 0.5,\ kz = 1.0$$

Вычисленный Кс свидетельствует о том, что влияние неучтенных причин не существенно и равно 36%. Влияние же учтенной причины (учтенного фактора, P) существенно.

Принцип соответствия. Условие этого принципа не выполняется, так как переменные наблюдаемого распределения ЭЗР и ненаблюдаемого распределения ТЗР соответствуют разным типам, типу С и типу D (Рис. 9.1.2). Следовательно, наша выборка не является репрезентативной двумерному ОНГ. Однако мы продолжим наш анализ в целях иллюстрации данного метода.

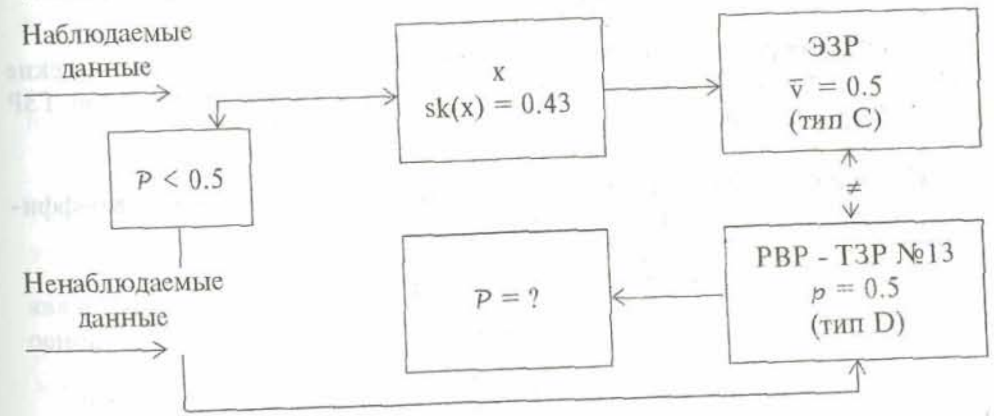


Рис. 9.1.2

9.1.2. Условие №2: зависимая переменная существенно зависит только от независимой переменной. Согласно этому условию, независимая переменная, так же как и зависимая переменная, репрезентативна в отношении одной и той же ОНГ, если выборка однородная. Поэтому и агрегированная переменная, состоящая из зависимой и независимой переменных, также репрезентативна ОНГ.

Объединение двух переменных в одну производится на основе относительных величин. Средние, вычисленные по относительным величинам каждой переменной, должны быть равны для того, чтобы быть сопоставимыми (Табл. 9.1.5). Если вычисленная агрегированная переменная репрезентативна ОНГ, то условие №2 выполняется. Проверка однородности агрегированной переменной производится путем анализа условий трех принципов: порядка, сходства и соответствия.

Таблица 9.1.5

Относительные величины

x	$\frac{x}{\bar{x}} = t_1$, $\bar{x} = 3.75$	y	$\frac{y}{\bar{y}} = t_2$, $\bar{y} = 5.15$
2	0.533	2.4	0.466
3	0.800	4.0	0.777
4	1.067	5.3	1.029
6	1.600	8.9	1.728

Принцип порядка. *Условие этого принципа выполняется. Эмпирические* значения репрезентативности расположены в пределах интервалов ТЗР (Табл. 9.1.6). ТЗР представлены типом С.

Принцип сходства. Условие этого принципа не выполняется. Коэффициент сходства меньше 0.5 (Табл. 9.1.7):

$$Kc = 1 - \frac{|sk(v) - sk(y)|}{|sk(y)|},$$

$$Kc = 1 - \frac{|-0.06 + 0.13|}{-0.13} = 0.46$$

В заключение следует сказать: так как агрегированная переменная не является репрезентативной ОНГ, зависимая переменная "x" находится под существенным влиянием неучтенных причин. В связи с этим второе условие не выполняется, что дает нам полное основание утверждать: наша выборка не

Таблица 9.1.6

Агрегированная переменная ЭЗР

#	t	$\|t - \bar{t}\| = d$, $\bar{t} = 1$	$1 - \frac{d}{0.728} = g$	$\frac{g}{0.96} = v$	ТЗР
1	0.466	0.534	0.266	0.277	0
2	0.533	0.467	0.359	0.374	0
3	0.777	0.223	0.694	0.723	1
4	0.800	0.200	0.725	0.755	1
5	1.029	0.029	0.960	1.000	1
6	1.067	0.067	0.908	0.946	1
7	1.600	0.600	0.176	0.183	0
8	1.728	0.728	0.000	0.000	0

$\bar{v} = 0.532$, $sk(v) = -0.06$

Таблица 9.1.7

Распределение ТЗР

y	$p(y)$
0	$0.468 = q$
1	$0.532 = p$

$\bar{y} = 0.532$, $sk(y) = -0.13$

является репрезентативной двумерной ОНГ и не может служить надежной основой для расчета связи между переменными "y" и "x".

9.1.3. Выводы. Традиционная статистика применяет регрессионный анализ для исследования причинно-следственных связей между переменными без проверки РО выборки. РО выборки характеризуется соблюдением следующих двух условий:

1. Независимая переменная репрезентативна ОНГ.
2. Зависимая переменная существенно зависит только от независимой переменной.

Следует отметить следующий факт: в репрезентативной выборке связь между переменными линейная. Любое отклонение от линейной связи образуется под влиянием неучтенных причин. Поэтому применение нелинейного регрессионного анализа для измерения связи между переменными вызывает сомнение с позиции принципа однородности.

9.2. Групповые статистические данные

Данные нашей выборки представлены в виде двумерного распределения частот (Табл. 9.2.1). Построение этого распределения основано на следующей предпосылке: расходы на жилье (y) зависят существенно только от количества комнат в квартире (x). Требуется проверить достоверность нашей предпосылки, т.е. проверить однородность выборки.

Проверка РО производится на основе негрупповых данных, которые вычисляются по групповым данным. Для определения негрупповых данных применяется метод накопленных частот (см. главу 7). Этот метод применяется к двум безусловным распределениям частот с их переменными: независимой (x) и зависимой (y). Результаты расчетов (Табл. 9.2.2 и 9.2.3) представлены четырьмя условными семьями (вторичная выборка с их расходами на жилье и количеством комнат в квартире (Табл. 9.2.4). Вычисленная выборка (вторичная) репрезентативна двумерной ОНГ, если следующие два условия соблюдаются:

1. Независимая переменная репрезентативна ОНГ.
2. Зависимая переменная существенно зависит только от независимой переменной.

9.2.1. Условие №1: независимая переменная репрезентативна ОНГ. Однородность независимой переменной проверяется путем анализа условий трех принципов: порядка, сходства и соответствия.

Принцип порядка. Условие этого принципа соблюдается (Табл. 9.2.5), *так как ЭЗР расположены в пределах интервалов ТЗР (Табл. 9.2.6). В нашем* примере ТЗР представлены типом С.

Принцип сходства. Условие этого принципа выполняется, так как нет существенного различия между распределениями ЭЗР и ТЗР. Коэффициент сходства между этими распределениями равен 64% (Табл. 9.2.7):

$$Kc = 1 - [kr(v) - kr(y)] = 1 - (1.36 - 1.00) = 0.64$$

Таблица 9.2.1

Двумерное распределение частот

Число комнат x	Расходы на жилье ($1000), y					Итого f(x)
	0.9-2.5	2.5-4.1	4.1-5.7	5.7-7.3	7.3-8.9	
1	2					2
2	1	4	1			6
3	2	6	2			10
4		3	4	2		9
5			1	2	1	4
6					1	1
Итого f(y)	5	13	8	4	2	32

Таблица 9.2.2

Накопленные частоты
(число комнат)

x	Число семей		Σf(%)
	f	f(%)	
1	2	0.0625	0.0625
2	6	0.1875	0.2500
3	10	0.3125	0.5625
4	9	0.2810	0.8440
5	4	0.1250	0.9690
6	1	0.0310	1.0000
Итого	32	1.000	-

$x_1 = 2, \quad x_2 = 3, \quad x_3 = 4, \quad x_4 = 6.$

Принцип соответствия. Условие этого принципа не выполняется, так как типы переменных двух распределений, ЭЗР и ТЗР, разные, а именно - тип С и тип D (Рис. 9.2.1). Этот факт дает нам основание утверждать, что выборка неоднородная. Однако анализ этой выборки будет продолжен в целях иллюстрации данного метода.

Таблица 9.2.3

Накопленные частоты
(расходы на жилье)

y	i	Число семей		Σf(%)
		f	f(%)	
0.9-2.5	1.6	5	0.156	0.156
2.5-4.1	1.6	13	0.406	0.562
4.1-5.7	1.6	8	0.250	0.812
5.7-7.3	1.6	4	0.125	0.937
7.3-8.9	1.6	2	0.063	1.000
Итого	-	32	1.000	-

$y_1 = 2.87, \quad y_2 = 3.86, \quad y_3 = 5.30, \quad y_4 = 8.9$

Таблица 9.2.4

Число комнат (x) и расходы на жилье (y)

#	x	y (в $1000)
1	2	2.87
2	3	3.86
3	4	5.30
4	6	8.90

$\overline{x} = 3.75$ $\quad$ $\overline{y} = 5.23$

$s = 1.479$ $\quad$ $s = 2.287$

$sk = 0.43$ $\quad$ $sk = 0.70$

9.2.2 Условие №2: зависимая переменная существенно зависит только от независимой переменной. Анализ влияния неучтенных причин на зависимую переменную производится на базе агрегированной переменной состоящей из двух переменных, "x" и "y". Эти переменные пересчитаны в относительные величины (Табл. 9.2.8, 9.2.9) в целях их сопоставимости. Однородность агрегированной переменной проверяется путем анализа условий трех принципов: порядка, сходства и соответствия.

Таблица 9.2.5

Расчет ЭЗР (для независимой переменной)

x	$\lvert x - \overline{x}\rvert = d$ $\overline{x} = 3.75$	$1 - \frac{d}{2.25} = g$	$\frac{g}{0.89} = v$	ТЗР
2	1.75	0.22	0.25	0
3	0.75	0.67	0.75	1
4	0.25	0.89	1.00	1
6	2.25	0.00	0.00	0

Таблица 9.2.6

Принцип порядка

ТЗР	v_x
0	0.00
0.0 - 0.5	0.25
0.5 - 1.0	0.75
1	1.00

$\overline{v} = 0.5, \quad kr = 1.36$

Таблица 9.2.7

Распределение ТЗР

x	p(x)
0	$0.5 = q$
1	$0.5 = p_x$

$\overline{v} = p_x = 0.5, \quad kz = 1.0$

Принцип соответствия

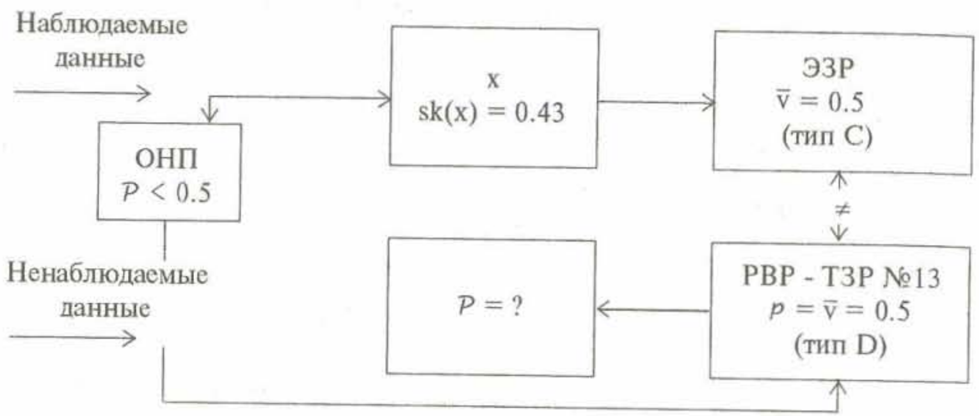


Рис. 9.2.1

Таблица 9.2.8

Расчет относительных величин
для "x" и "y"

x	$\frac{x}{\bar{x}} = t_1$, $\bar{x} = 3.75$	y	$\frac{y}{\bar{y}} = t_2$, $\bar{y} = 5.23$
2	0.533	2.87	0.549
3	0.800	3.86	0.738
4	1.067	5.30	1.013
6	1.600	8.90	1.702

Принцип порядка. Условие этого принципа выполняется, так как ЭЗР расположены в интервалах ТЗР (Табл. 9.2.10). В данном примере ТЗР представлены типом С.

Принцип сходства. Условие этого принципа не выполняется, так как коэффициент сходства меньше 0.5 (Табл. 9.2.11):

$$Kc = 1 - \frac{|sk(v) - sk(y)|}{|sk(y)|},$$

$$Kc = 1 - \frac{|-0.03 + 0.07|}{-0.07} = 0.43$$

Данный результат свидетельствует о том, что агрегированная переменная не является репрезентативной ОНГ, а поэтому проверку условия принципа соответствия производить не следует.

Таблица 9.2.9

Расчет ЭЗР
по агрегированной переменной

#	t	$\lvert t - \bar{t} \rvert = d$, $\bar{t} = 1$	$1 - \frac{d}{0.702} = g$	$\frac{g}{0.981} = v$	ТЗР
1	0.533	0.467	0.335	0.341	0
2	0.549	0.451	0.358	0.365	0
3	0.738	0.627	0.627	0.639	1
4	0.800	0.200	0.715	0.729	1
5	1.013	0.013	0.981	1.000	1
6	1.067	0.067	0.905	0.923	1
7	1.600	0.600	0.145	0.148	0
8	1.702	0.702	0.000	0.000	0

9.2.3. Выводы. Проверка однородности причинной связи между двумя рядами распределения частот производится по вторичной выборке, состоящей из условных несгруппированных данных. Эти данные вычисляются методом накопленных частот на основе первичной выборки, т.е. по рядам распределения частот. Достоверность этого метода проверки однородности статистических данных зависит от соблюдения следующего условия: единицы наблюдения должны быть распределены равномерно в интервалах ряда распределения частот.

7 Зак. 28

Таблица 9.2.10

Принцип порядка

ТЗР	v
0.0	0.000
0 - 0.5	0.148
0 - 0.5	0.341
0 - 0.5	0.365
0.51 - 1.0	0.639
0.51 - 1.0	0.729
0.51 - 1.0	0.923
1.0	1.000

$\bar{v} = 0.518$, $sk(v) = -0.03$

Таблица 9.2.11

Распределение ТЗР

y	$p(y)$
0	$0.482 = q$
1	$0.518 = p$

$\bar{y} = 0.518$, $sk = -0.07$

10. Трехмерные статистические совокупности

Причинно-следственные связи между тремя и более переменными также могут быть подвергнуты проверке репрезентативной однородности. Проверка таких совокупностей будет исследована по негрупповым статистическим данным.

10.1 Трехмерное ОНГ

Трехмерная ОНГ состоит из двух двумерных ОНГ, каждая из которых характеризуется независимой и зависимой переменными (Рис. 10.1). Связь между этими переменными функциональная, т.е. влияние неучтенных причин на зависимую переменную нулевое.

В тех случаях, когда мы имеем дело с трехмерной статистической выборкой, влияние неучтенных причин на зависимую переменную не нулевое. Такая выборка репрезентативна трехмерной ОНГ, если влияние неучтенных причин не существенно. Данное влияние не существенно, если выполняются следующие условия:

1. Каждая независимая переменная должна быть репрезентативной соответствующей переменной ОНГ.
2. Независимые переменные существенно отличаются друг от друга.
3. Зависимая переменная существенно зависит только от двух независимых переменных.

Трехмерные ОНГ

Рис. 10.1

Проанализируем вышеназванные условия на основе выборки, состоящей из четырех семей; каждая семья состоит из мужа, жены и сына разного возраста. Наша выборка базируется на следующей проверяемой гипотезе: потребление хлеба (y) зависит существенно только от дохода, “x”, и возраста сына, “z” (Табл. 10.1). Влияние неучтенных причин не оказывает существенного воздействия на потребление хлеба. Проверка данной гипотезы есть проверка репрезентативной однородности выборки. Эта проверка производится путем анализа вышеназванных трех условий.

Таблица 10.1

Доход (x), возраст сына (z), потребление хлеба (y)

#	x ($1000)	z (годы)	y (фунты)
1	20.1	7.0	309.4
2	24.9	9.2	319.5
3	25.2	11.6	321.6
4	36.1	15.4	336.6

$\bar{x} = 26.757, \quad \bar{z} = 10.8$

$sk = 0.73, \quad sk = 0.32$

10.2. Условие №1: однородность первой независимой переменной.

Для проверки РО первой независимой переменной (доход) проанализируем условия трех принципов: порядка, сходства и соответствия.

Принцип порядка. Условие этого принципа выполняется. Вычисленные ЭЗР расположены в пределах интервалов ТЗР (Табл. 10.2 и 10.3), которые представлены типом С.

Принцип сходства. Условие этого принципа выполняется: распределение ЭЗР не отличается существенно от распределения ТЗР (Табл. 10.4). Коэффициент сходства равен 82%:

$$Kc = 1 - \frac{|sk(v) - sk(y)|}{|sk(y)|} = 1 - \frac{|-0.28 + 0.34|}{|-0.34|} = 0.82$$

Таблица 10.2

ЭЗР (доход)

x	$\lvert x-\bar{x}\rvert = d$, $\bar{x} = 26.575$	$1 - \frac{d}{9.525} = g$	$\frac{g}{0.856} = v_x$	ТЗР
20.1	6.475	0.320	0.374	0
24.9	1.675	0.824	0.963	1
25.2	1.375	0.856	1.000	1
36.1	9.525	0.000	0.000	0

Таблица 10.3

Принцип порядка

ТЗР	ЭЗР
0.0	0.000
0.0 - 0.5	0.374
0.5 - 1.0	0.963
1.0	1.000

$\bar{v} = 0.584, \quad sk = -0.28$

Таблица 10.4

Распределение ТЗР

x	$p(x)$
0	$0.416 = q$
1	$0.584 = p$

$p = \bar{v}, \quad sk(x) = -0.34$

Принцип соответствия. Условие этого принципа выполняется, так как типы переменных двух распределений, ЭЗР и ТЗР, одинаковые - *тип* С (Рис. 10.2).

Принцип соответствия

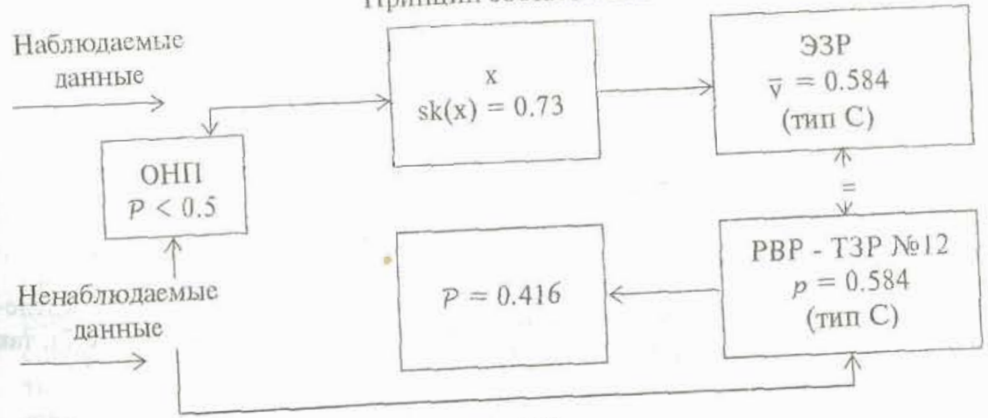


Рис. 10.2

10.3. Условие №1а: однородность второй независимой переменной.

Для проверки однородности второй независимой переменной (возраст сына) проанализируем условия трех принципов: порядка, сходства и соответствия.

Принцип порядка. Условие этого принципа выполняется. Вычисленные ЭЗР расположены в пределах интервалов ТЗР, которые соответствуют типу С (Табл. 10.5 и 10.6).

Таблица 10.5

ЭЗР (возраст сына)

z	$\lvert z-\bar{z}\rvert = d$ $\bar{z} = 10.8$	$1-\frac{d}{4.6} = g$	$\frac{g}{0.826} = v_z$	ТЗР
7.0	3.8	0.174	0.211	0
9.2	1.6	0.652	0.789	1
11.6	0.8	0.826	1.000	1
15.4	4.6	0.000	0.000	0

Таблица 10.6

Принцип порядка

ТЗР	ЭЗР
0.0	0.000
0.0 - 0.5	0.211
0.5 - 1.0	0.789
1.0	1.000

$\bar{v} = 0.5, \quad sk = 0, \quad kr = 1.27$

Принцип сходства. Условие этого принципа выполняется: распределение ЭЗР существенно не отличается от распределения ТЗР (Табл. 10.7), так как коэффициент сходства больше 0.5:

$$Kc = 1 - [kr(v) - kz(z)] = 1 - (1.27 - 1.0) = 0.73$$

Таблица 10.7

Распределение ТЗР

z	$p(z)$
0	$0.5 = q$
1	$0.5 = p$

$p = \bar{v}, \quad sk = 0, \quad kz = 1$

Принцип соответствия. Условие этого принципа не выполняется, так как типы переменных двух распределений, ЭЗР и ТЗР, не одинаковые; они соответствуют типам С и D (Табл. 4.2.1 и Рис. 10.3). Следовательно, вторая независимая переменная не является репрезентативной относительно ОНГ. Однако мы продолжим анализ данной проблемы в целях иллюстрации метода репрезентативной однородности.

Принцип соответствия

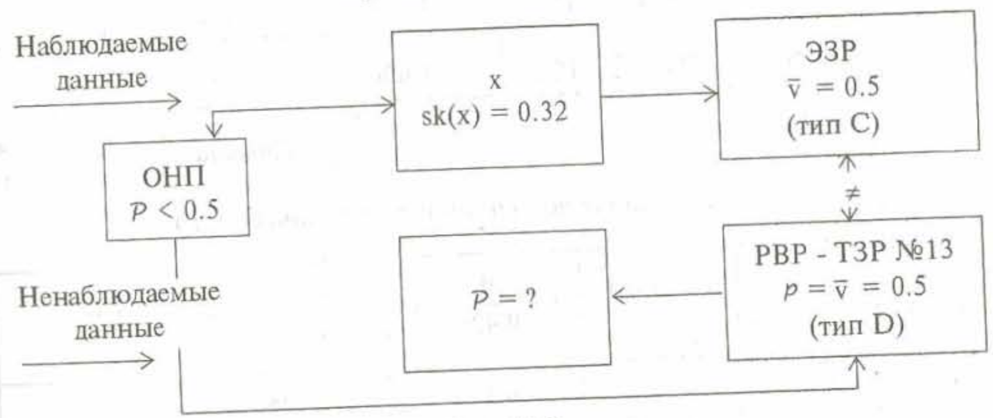


Рис. 10.3

10.4. Условие №2: независимые переменные существенно различаются между собой. Начнем с определения. Если две статистические совокупности репрезентативны в отношении одной и той же ОНГ, то их агрегированная статистическая совокупность также репрезентативна той же ОНГ. Такая агрегированная совокупность свидетельствует о том, что нет существенного различия между двумя статистическими совокупностями, а это значит, что условие №2 не соблюдается.

Метод проверки существенного различия между двумя статистическими совокупностями проиллюстрируем на нашем примере с двумя переменными, "x" и "z". Эти две переменные объединяются в одну агрегированную переменную. Данное объединение производится на основе относительных (t_1 и t_2) с одинаковыми средними в целях выполнения условия сопоставимости (Табл. 10.8 и 10.9). Далее проверяется РО агрегированной переменной, "t", путем анализа условий трех принципов: порядка, сходства и соответствия.

Таблица 10.8

Пересчет переменных в относительные величины

#	x	z	$\frac{x}{\bar{x}} = t_1$	$\frac{z}{\bar{z}} = t_2$
1	20.1	7.0	0.756	0.648
2	24.9	9.2	0.937	0.852
3	25.2	11.6	0.948	1.074
4	36.1	15.4	1.358	1.426
Среднее	26.575	10.8	1.000	1.000

Таблица 10.9

Агрегированная независимая переменная (ЭЗР)

t	$\lvert t-\bar{t} \rvert = d$, $\bar{t}=1$	$1-\frac{d}{0.426} = g$	$\frac{g}{0.878} = v_t$	ТЗР
0.648	0.352	0.174	0.198	0
0.756	0.244	0.427	0.486	0
0.852	0.148	0.653	0.744	1
0.937	0.063	0.852	0.970	1
0.948	0.052	0.878	1.000	1
1.074	0.074	0.826	0.940	1
1.358	0.358	0.160	0.182	0
1.426	0.426	0.000	0.000	0

$sk(t) = 0.46$

Принцип порядка. Условие этого принципа выполняется. Вычисленные ЭЗР расположены в пределах интервалов ТЗР (Табл. 10.10), которые соответствуют типу С.

Таблица 10.10

Принцип порядка

ТЗР	ЭЗР
0.0	0.000
0 - 0.5	0.182
0 - 0.5	0.198
0 - 0.5	0.486
0.5 - 1.0	0.744
0.5 - 1.0	0.940
0.5 - 1.0	0.970
1.0	1.000

$\bar{v}_t = 0.565, \quad sk(v_t) = -0.186$

Принцип сходства. Условие этого принципа выполняется, так как распределения ЭЗР и ТЗР не отличаются существенно между собой. Коэффициент сходства равен 0.71 (Табл. 10.11):

$$Kc = 1 - \frac{\lvert sk(v) - sk(t) \rvert}{\lvert sk(t) \rvert} = 1 - \frac{\lvert -0.186 + 0.262 \rvert}{\lvert -0.262 \rvert} = 0.71$$

Таблица 10.11

Распределение ТЗР

t	$p(t)$
0	$0.435 = q$
1	$0.565 = p$

$p = \bar{v}_t, \quad sk(t) = -0.262$

Следовательно, переменные “x” и “z” несущественно отличаются друг от друга, т.е. условие №2 не выполняется.

Принцип соответствия. Условие этого принципа не проверяется, так как предпосылка о несущественном различии между независимыми переменными была отвергнута.

10.5. Условие № 3: влияние неучтенных причин на зависимую переменную несущественно. Данное условие соблюдается, если переменная остаточных величин (разности между агрегированной независимой переменной и зависимой переменной) репрезентативна ОНГ, вероятность благоприятного события которой равна 0.5.

Для расчета переменной остаточных величин следует в первую очередь преобразовать переменные “x”, “z” и “y” в относительные величины “t_1”, “t_2” и “h” (Табл. 10.12). Эти относительные величины сопоставимы, так как их средние одинаковы и равны единице.

Таблица 10.12

Расчет относительных величин

#	x	z	y	$\frac{x}{\bar{x}} = t_1$	$\frac{z}{\bar{z}} = t_2$	$\frac{y}{\bar{y}} = h$
1	20.1	7.0	309.4	0.756	0.648	0.962
2	24.9	9.2	319.5	0.937	0.852	0.993
3	25.2	11.6	321.6	0.948	1.074	0.999
4	36.1	15.4	336.6	1.358	1.426	1.046
Средн.	26.575	10.8	321.775	1.000	1.000	1.000

Для анализа влияния независимой переменной на зависимую необходимо рассчитать агрегированную независимую переменную. Эта переменная вычисляется как средняя из двух переменных, “t_1” и “t_2” (Табл. 10.13). В нашем примере связь между агрегированной независимой переменной и зависимой переменной не функциональная в результате влияния неучтенных причин. Влияние этих причин формирует переменную остаточных величин, r (Табл. 10.13). Если влияние неучтенных причин в отдельности не существенно (т.е. среди неучтенных причин нет одного или нескольких существенных факторов), то распределение остаточных величин симметричное.

Таблица 10.13

Расчет остаточных величин

#	t		h	t-h=r
1	0.648			-0.260
2	0.756	0.702	0.962	
3	0.852			-0.099
4	0.937	0.894	0.993	
5	0.948			0.012
6	1.074	1.011	0.999	
7	1.358			0.346
8	1.426	1.392	1.046	

Проверка РО переменной остаточных величин есть проверка критерия “нормальности”. Расчет этого критерия выполняется в результате проверки условий трех принципов: порядка, сходства и соответствия.

Принцип порядка. Условие этого принципа выполняется, так как вычисленные ЭЗР (Табл. 10.14) расположены в пределах интервалов ТЗР (Табл. 10.15). В данном примере ТЗР представлены типом С.

Таблица 10.14

Расчет ЭЗР

r	$\lvert r-\bar{r}\rvert = d$ $\bar{r} = 0.00$	$1-\frac{d}{0.346} = g$	$\frac{g}{0.965} = v_r$	ТЗР
-0.260	0.260	0.249	0.258	0
-0.099	0.099	0.714	0.740	1
0.012	0.012	0.965	1.000	1
0.346	0.346	0.000	0.000	0

sk(r) = 0.53

Таблица 10.15

Принцип порядка

ТЗР	ЭЗР
0	0.000
0.0 - 0.5	0.258
0.5 - 1.0	0.740
1	1.000

$\bar{v}_Г = 0.4995, \quad sk = 0.0025$

Принцип сходства. Условие этого принципа выполняется, так как коэффициент сходства между распределениями ЭЗР и ТЗР (Табл. 10.16) равен 75%:

$$Kc = 1 - \frac{|\mathrm{sk} - sk|}{|sk|} = 1 - \frac{|0.0025 + 0.002|}{|0.002|} = 0.75$$

Таблица 10.16

Распределение ТЗР

z	$p(z)$
0	0.5005
1	0.4995

$p = \bar{v}_Г = 0.4995, \quad sk = 0.0020$

Принцип соответствия. Условие этого принципа не выполняется, так как переменная распределения ТЗР не является репрезентативной, поскольку $p \neq 0.5$ (Табл. 4.2.1, Рис. 10.4).

В заключение следует сказать, что наша трехмерная статистическая совокупность не является репрезентативной в отношении ОНГ, так как условия №1, №2 и №3 не выполняются. Перечислим эти условия.

<u>Условие №1:</u> каждая независимая переменная должна быть репрезентативна соответствующей ОНГ. В нашем примере вторая независимая переменная, возраст сына, не является репрезентативной ОНГ.

Принцип соответствия

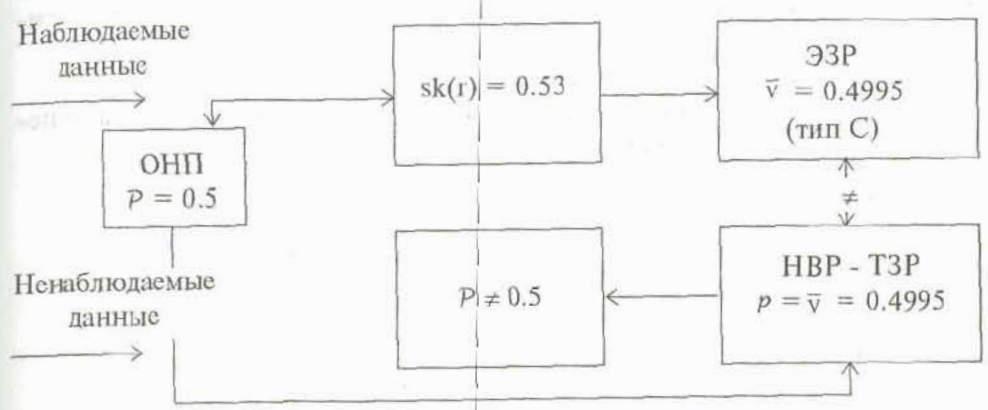


Рис. 10.4

<u>Условие №2:</u> независимые переменные должны быть существенно различны между собой. В нашем примере они существенно зависят от одной или нескольких общих причин. Это обстоятельство элиминирует существенное различие между доходом и возрастом сына.

<u>Условие № 3:</u> зависимая переменная не должна существенно зависеть от неучтенных причин. В нашем примере зависимая переменная существенно зависит от неучтенных причин, в результате этого остаточные величины не распределены симметрично.

10.6. Выводы

Трехмерная ОНГ характеризуется функциональной связью между зависимой переменной и независимыми переменными. Трехмерная статистическая совокупность характеризуется корреляционной связью между зависимой переменной и независимыми переменными в результате влияния неучтенных причин.

Проверка однородности трехмерной статистической совокупности основана на проверяемой предпосылке: статистическая совокупность репрезентативна трехмерной ОНГ. Проверка однородности предполагает выполнение трех условий.

Первое условие: каждая независимая переменная репрезентативна соответствующей ОНГ. Второе условие: независимые переменные существенно отличаются между собой. Третье условие: зависимая переменная существенно зависит только от двух независимых переменных.

Вышеназванные три условия выполняются, если:

1. Связь зависимой переменной от агрегированной независимой переменной линейная.
2. Распределение остаточных величин симметричное.

*"И познаете истину,
и истина сделает
вас свободными."*

От Иоанна 8:32

Часть V
Аксиоматический анализ контроля качества продукции

11. Основные проблемы контроля качества продукции

Производственный процесс считается стабильным, если соблюдаются следующие три предпосылки:

1. Индивидуальные измерения независимы.
2. Индивидуальные измерения характеризуются нормальным распределением. Этот же тип распределения свойственен и индивидуальным измерениям в каждой выборке (подгруппе).
3. Выборочные средние нормально распределены.

Эти предпосылки необоснованные, так как они не проверяются и не могут быть проверены методами традиционной статистики.

Основной статистический метод традиционного анализа контроля качества продукции (ТАККП) основан на анализе контрольной карты, при помощи которой изучается стабильность производственного процесса. Логическая предпосылка ТАККП такова: все производственные изделия считаются пригодными, если индивидуальные измерения, средние величины и средние квадратические отклонения находятся в пределах контрольных линий. Эта предпосылка является необоснованной по двум причинам. Во-первых, отсутствует объективное статистическое определение для такой категории, как качественная, т.е. однородная продукция. Во-вторых, метод расчета контрольных линий основан на предпосылке: индивидуальные измерения распределяются нормально. Однако эта предпосылка не может быть проверена методами традиционной статистики.

Согласно ТАККП, производственный процесс есть результат влияния ряда причин. К этим причинам относятся следующие: отдельный станок, группа станков, сборочный цех, процедура проверки, отдельный рабочий, группа рабочих и т.д. Взаимодействие между этими причинами оказывает существенное влияние на индивидуальные измерения и на обобщающие показатели выборки. Индивидуальные измерения используются для анализа производственного процесса в пределах отдельных выборок. Обобщающие показатели используются для анализа производственного процесса между выборками.

Анализируя производственный процесс между выборками, ТАККП различает следующие его два типа: производственный процесс, находящийся в состоянии статистического контроля (ССК), и производственный процесс, не находящийся в ССК.

Процесс, находящийся в ССК, характеризуется устойчивостью обобщающих показателей выборок и симметричностью их распределения. Процесс, не находящийся в ССК, характеризуется неустойчивостью обобщающих показателей выборок. *Тенденция устойчивости и неустойчивости обобщающих показателей формируется под влиянием невидимых и видимых причин.*

Согласно ТАККП, существуют два типа видимых причин, формирующих неустойчивость обобщающих показателей (Laurenson, 1981). Первый тип *видимых причин оказывает влияние на увеличение размаха колебаний.* Увеличение размаха колебаний вызывается такими причинами, как неустраненные неполадки, небрежная эксплуатация оборудования и ошибки операторов в случае неавтоматизированного производственного процесса. Второй тип *видимых причин отклоняет обобщающие показатели от их постоянного уровня.* Это отклонение (сдвиг) вызывается такими причинами, как плохая наладка оборудования, смена операторов и изменение техники проверки.

Перечисленные выше видимые причины нарушают стабильность производственного процесса. *Для выявления этих причин* ТАККП исследует обобщающие показатели на базе анализа:

1. Изменения колеблемости производственного процесса между подгруппами (выборками).
2. Сдвига центральной тенденции производственного процесса.

Исследование обобщающих показателей методами ТАККП неэффективно, так как игнорируется анализ индивидуальных измерений. Наш метод анализа контроля качества продукции не основан на вышеназванных *необоснованных предпосылках и анализе обобщающих показателей.* Он базируется на аксиомах и анализе индивидуальных измерений. Наши аксиомы - отправная база для проверки необоснованных предпосылок ТАККП.

Аксиоматический анализ контроля качества продукции (ААККП) состоит из двух этапов проверки:

1. Поверка однородности (стабильности) производственного процесса в пределах каждой выборки (в подгруппах).
2. Проверка однородности (стабильности) производственного процесса в пределах нескольких выборок.

12. Анализ стабильности производственного процесса в пределах каждой выборки

Анализ производственного процесса в пределах каждой выборки (в подгруппах) начинается с отбора однородных групп. Процесс отбора групп (выборок) производится на основе логического и статистического анализа. Качество логического анализа зависит от глубины знания технической спецификации производственного процесса. Результаты статистического анализа

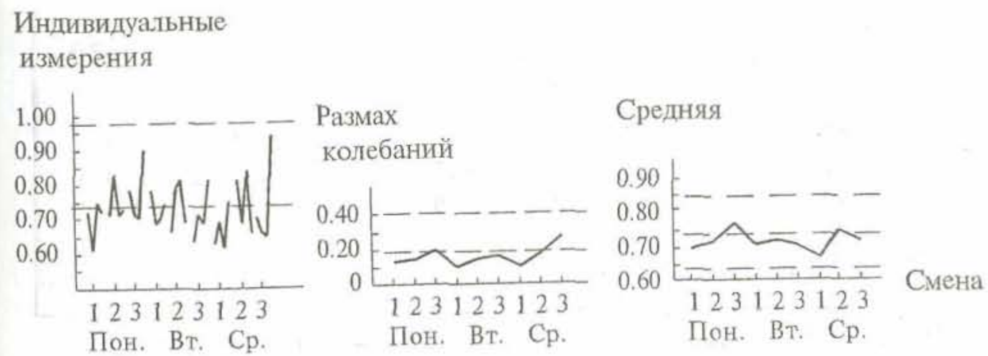
зависят от качества статистических данных, характеризующих стабильность производственного процесса.

Статистическую проверку стабильности производственного процесса проиллюстрируем по данным, опубликованным в "Руководстве по приготовлению статистической информации для анализа контрольной карты" (ASTM, 1976). Мы ограничим статистическую совокупность девятью выборками, размер каждой сокращен до четырех измерений (Табл. 12.1).

Таблица 12.1

Содержание силиция в стали (%)

День недели	Смена	Нагрев 1	Нагрев 2	Нагрев 3	Нагрев 4	Размер выборки n	Размах колебаний r	Средняя $\bar{y}$
Понед.	1	0.72	0.61	0.75	0.73	4	0.14	0.702
	2	0.68	0.83	0.71	0.73	4	0.15	0.738
	3	0.78	0.71	0.70	0.90	4	0.20	0.772
Вторн.	1	0.78	0.68	0.70	0.74	4	0.10	0.725
	2	0.66	0.79	0.81	0.68	4	0.15	0.735
	3	0.64	0.71	0.69	0.81	4	0.17	0.712
Среда	1	0.63	0.69	0.62	0.75	4	0.13	0.672
	2	0.81	0.68	0.84	0.66	4	0.18	0.748
	3	0.70	0.66	0.65	0.93	4	0.28	0.735



Расчленение данных на подгруппы (выборки) - решающий момент в анализе производственного процесса. Это расчленение предполагает, что колеблемость в пределах каждой подгруппы объясняется влиянием ОНП.

Проверка стабильности производственного процесса основана на двух проверяемых предпосылках. Первая предпосылка: измерения в подгруппах распределены симметрично, т.е. они репрезентативны ОНГ, у которой $P = 0.5$. Вторая предпосылка: измерения в подгруппах репрезентативны ОНГ, у которой $P \neq 0.5$. Проверка этих предпосылок производится путем анализа условий трех принципов: порядка, сходства и соответствия.

Таблица 12.2

Понедельник, смена №1

#	x	$\lvert x-\bar{x}\rvert = d$, $\bar{x} = 0.7025$	$1-\dfrac{d}{0.0925} = g$	$\dfrac{g}{0.81} = v$	ТЗР
1	0.61	0.0925	0.00	0.00	0.00
2	0.72	0.0175	0.81	1.00	1.00
3	0.73	0.0275	0.70	0.86	1.00
4	0.75	0.0475	0.49	0.60	0.00
-	$\bar{x} = 0.7025$	-	-	$\bar{v} = 0.615$	-

sk(x) = -1.05 sk(v) = -0.71 $sk(y) = -0.47$

Таблица 12.3

Понедельник, смена №2

#	x	$\lvert x-\bar{x}\rvert = d$, $\bar{x} = 0.7375$	$1-\dfrac{d}{0.0925} = g$	$\dfrac{g}{0.99} = v$	ТЗР
1	0.68	0.0575	0.94	0.95	1.0
2	0.71	0.0275	0.97	0.98	1.0
3	0.73	0.0075	0.99	1.00	1.0
4	0.83	0.0925	0.00	0.00	0.0
-	$\bar{x} = 0.7375$	-	-	$\bar{v} = 0.7325$	-

sk(x) = 0.84 sk(v) = -1.15 $sk(y) = -1.05$

Первая выборка (понедельник, смена №1) не является репрезентативной ОНГ. Коэффициент сходства равен 0.49 (Табл. 12.2).

Вторая выборка (понедельник, смена №2) репрезентативна ОНГ. Вычисленные ЭЗР соответствуют типу D. Этот тип состоит из 75% репрезентативных сложных событий. В соответствии с принципом последовательности, коэффициент сходства тоже должен быть равен или больше 0.75. В нашем примере он равен 0.90. Следовательно, условие принципа сходства выполнено. Условие принципа соответствия также выполнено. В заключение следует сказать, что вторая выборка репрезентативна ОНГ, у которой $P < 0.5$ (Табл. 12.3).

Таблица 12.4

Понедельник, смена №3

#	x	$\lvert x-\bar{x}\rvert = d$, $\bar{x} = 0.7725$	$1-\dfrac{d}{0.1275} = g$	$\dfrac{g}{0.94} = v$	ТЗР
1	0.70	0.0725	0.43	0.46	0.0
2	0.71	0.0625	0.51	0.54	1.0
3	0.78	0.0075	0.94	1.00	1.0
4	0.90	0.1275	0.00	0.00	0.0
-	$\bar{x} = 0.7725$	-	-	$\bar{v} = 0.5$	-

sk(x) = 0.74 kr(v) = 1.97 $kr(y) = 1.0$

Таблица 12.5

Вторник, смена №1

#	x	$\lvert x-\bar{x}\rvert = d$	$1-\dfrac{d}{0.055} = g$	$\dfrac{g}{0.75} = v$	ТЗР
1	0.68	0.045	0.18	0.25	0.00
2	0.70	0.025	0.55	0.75	1.00
3	0.74	0.015	0.73	1.00	1.00
4	0.78	0.055	0.00	0.00	0.00
-	$\bar{x} = 0.725$	-	-	$\bar{v} = 0.5$	-

sk(x) = 0.44 kr(v) = 1.36 $kr(y) = 1.0$

Третья выборка (понедельник, смена №3) не является репрезентативной ОНГ, так как коэффициент сходства равен 0.03 (Табл. 12.4).

Четвертая выборка (вторник, смена №1) репрезентативна ОНГ, у которой $\mathcal{P} = 0.5$. Условия всех трех принципов соблюдаются (Табл. 12.5).

Пятая выборка (вторник, смена №2) репрезентативна ОНГ, у которой $\mathcal{P} = 0.5$. Условия всех трех принципов выполнены (Табл. 12.6).

Шестая выборка (вторник, смена №3) не является репрезентативной ОНГ, так как условие принципа соответствия не выполняется (Табл. 12.7).

Седьмая выборка (среда, смена №1) не является репрезентативной ОНГ, так как коэффициент сходства равен 0.1 (Табл. 12.8).

Таблица 12.6

Вторник, смена №2

#	x	$\lvert x-\bar{x}\rvert = d$	$1-\frac{d}{0.075}=g$	$\frac{g}{0.27}=v$	ТЗР
1	0.66	0.075	0.00	0	0
2	0.68	0.055	0.27	1	1
3	0.79	0.055	0.27	1	1
4	0.81	0.075	0.00	0	0
-	$\bar{x} = 0.735$	-	-	$\bar{v} = 0.5$	-

sk(x) = 0 kr(v) = 1 $kz(y) = 1.0$

Таблица 12.7

Вторник, смена №3

#	x	$\lvert x-\bar{x}\rvert = d$	$1-\frac{d}{0.0975}=g$	$\frac{g}{0.914}=v$	ТЗР
1	0.64	0.0725	0.256	0.263	0.00
2	0.69	0.0225	0.769	0.789	1.00
3	0.71	0.0025	0.974	1.000	1.00
4	0.81	0.0975	0.000	0.000	0.00
-	$\bar{x} = 0.7125$	-	-	$\bar{v} = 0.513$	-

sk(x) = 0.58 sk(v) = -0.055 $sk(y) = -0.052$

Восьмая выборка (среда, смена №2) репрезентативна ОНГ, у которой $\mathcal{P} = 0.5$. Условия всех трех принципов соблюдаются (Табл. 12.9).

Девятая выборка (среда, смена №3) не является репрезентативной ОНГ, так как коэффициент сходства равен 0.22 (Табл. 12.10).

В результате нашего анализа оказалось, что только три выборки репрезентативны ОНГ, у которых $\mathcal{P} = 0.5$ (Табл. 12.11).

Выводы. Проверка стабильности производственного процесса в пределах каждой выборки есть проверка репрезентативности выборки относительно ОНГ. Репрезентативная выборка свидетельствует о том, что производственный процесс стабилен, а произведенная продукция - качественная.

Таблица 12.8

Среда, смена №1

#	x	$\lvert x-\bar{x}\rvert = d$	$1-\frac{d}{0.0775}=g$	$\frac{g}{0.77}=v$	ТЗР
1	0.62	0.0525	0.32	0.42	0.00
2	0.63	0.0425	0.45	0.58	1.00
3	0.69	0.0175	0.77	1.00	1.00
4	0.75	0.0775	0.00	0.00	0.00
-	$\bar{x} = 0.6725$	-	-	$\bar{v} = 0.5$	-

sk(x) = 0.40 kr(v) = 1.90 $kz(y) = 1.0$

Таблица 12.9

Среда, смена №2

#	x	$\lvert x-\bar{x}\rvert = d$	$1-\frac{d}{0.0925}=g$	$\frac{g}{0.324}=v$	ТЗР
1	0.66	0.0875	0.054	0.17	0.00
2	0.68	0.0675	0.270	0.83	1.00
3	0.81	0.0625	0.324	1.00	1.00
4	0.84	0.0925	0.000	0.00	0.00
-	$\bar{x} = 0.7475$	-	-	$\bar{v} = 0.5$	-

sk(x) = 0.03 kr(v) = 1.15 $kz(y) = 1.00$

В нашем примере мы обнаружили четыре репрезентативных выборки, из них три репрезентативны одному типу ОНГ, у которого $P = 0.5$.

Таблица 12.10

Среда, смена №3

#	x	$\lvert x - \bar{x} \rvert = d$	$1 - \frac{d}{0.195} = g$	$\frac{g}{0.820} = v$	ТЗР
1	0.65	0.085	0.564	0.69	1.00
2	0.66	0.075	0.615	0.75	1.00
3	0.70	0.035	0.820	1.00	1.00
4	0.93	0.195	0.000	0.00	0.00
-	$\bar{x} = 0.735$	-	-	$\bar{v} = 0.61$	-

$sk(x) = 1.08$ $sk(v) = -0.80$ $sk(g) = -0.45$

Таблица 12.11

Результаты проверки РО отдельных выборок

		sk(x)	ТЗР	$\bar{v}$	Кс	Репрезентативность выборки относительно ОНГ	
						$P = 0.5$	$P < 0.5$
Понед.	#1	-1.05	C	0.62	0.57	нет	нет
	#2	0.84	D	0.73	0.96	нет	да
	#3	0.74	C	0.50	0.51	нет	нет
Вторн.	#1	0.44	C	0.50	0.87	да	нет
	#2	0.00	C	0.50	1.00	да	нет
	#3	0.58	C	0.51	0.94	нет	нет
Среда	#1	0.40	C	0.50	0.55	нет	нет
	#2	0.03	C	0.50	0.92	да	нет
	#3	1.08	D	0.61	0.96	нет	нет

13. Анализ стабильности производственного процесса в пределах нескольких выборок

В качестве примера рассмотрение проблемы стабильности производственного процесса производится по 16 индивидуальным измерениям первых четырех выборок (Табл. 13.1). Из этих четырех выборок первая и третья выборки не являются репрезентативными. Их влияние на стабильность производственного процесса проанализируем путем проверки условий трех принципов: порядка, сходства и соответствия.

Таблица 13.1

Содержание силиция в стали (в %)

#	x	$\lvert x - \bar{x} \rvert = d$	$1 - \frac{d}{0.1656} = g$	$\frac{g}{0.973} = v$
1	0.61	0.1244	0.249	0.256
2	0.68	0.0544	0.671	0.690
3	0.68	0.0544	0.671	0.690
4	0.70	0.0344	0.792	0.814
5	0.70	0.0344	0.792	0.814
6	0.71	0.0244	0.853	0.877
7	0.71	0.0244	0.853	0.877
8	0.72	0.0144	0.913	0.938
9	0.73	0.0044	0.973	1.000
10	0.73	0.0044	0.973	1.000
11	0.74	0.0056	0.966	0.993
12	0.75	0.0156	0.906	0.931
13	0.78	0.0456	0.725	0.745
14	0.78	0.0456	0.725	0.745
15	0.83	0.0956	0.423	0.435
16	0.90	0.1656	0.000	0.000

$\bar{x} = 0.7344$, $s = 0.0644$, $sk = 1.17$

Принцип порядка. Условие этого принципа выполняется, так как ЭЗР расположены в интервалах ТЗР (Табл. 13.2). В нашем примере ТЗР представлены типом D.

Таблица 13.2

Распределение ЭЗР

ТЗР	v						
0	0						
0.51-1.00	0.256,	0.435,	0.690,	0.690,	0.745,	0.745,	0.814
0.51-1.00	0.814,	0.877,	0.877,	0.931,	0.938,	0.993,	1
1	1						

$$\bar{v} = 0.738,\ s = 0.274,\ sk = -1.42$$

Принцип сходства. Условие этого принципа не выполняется, так как коэффициент сходства (в соответствии с принципом последовательности относительно типа D) должен быть равен или больше 0.75. Он равен 0.69 (Табл. 13.3):

$$Kc = 1 - \frac{|-1.42 + 1.08|}{|-1.08|} = 0.69$$

Таблица 13.3

Распределение ТЗР

y	$p(y)$
0	$0.262 = q$
1	$0.738 = p$
Итого	1.000

$$p = \bar{v} = 0.738,\quad sk = -1.08$$

Принцип соответствия. Условие этого принципа также не выполняется: переменные распределений ЭЗР и ТЗР соответствуют разным типам совокупностей сложных событий, а именно типу D и типу C. Определение типа C основано на сопоставлении разностей между коэффициентами асимметрии ОНГ (для типа C: $SK = 1.08$, а для типа D: $SK = 0.06$) и 16 индивидуальных измерений ($sk(x) = 1.17$):

$$1.17 - 1.08 < 1.17 - 0.06$$

Данное неравенство свидетельствует о том, что тип C соответствует переменной распределения ТЗР.

Таким образом, мы пришли к выводу, что изучаемый производственный процесс нестабилен в результате существенного влияния двух не являющихся репрезентативными выборок*.

14. Проверка изменения колеблемости производственного процесса между выборками

Традиционный анализ контроля качества продукции (ТАККП) уделяет особое внимание детальному исследованию колеблемости обобщающих показателей, например, таких, как размах колебаний и средние величины. Эти обобщающие показатели сглаживают действительные колебания индивидуальных измерений и этим искажают результаты анализа стабильности производственного процесса. Ниже мы проиллюстрируем этот процесс искажения, анализируя производственный процесс (с позиции его колеблемости и сдвига центральной тенденции) на основе обобщающих показателей. Эти обобщающие показатели вычисляются по первым четырем выборкам нашего примера, рассмотренного в двух предыдущих главах. Анализ начнем с рассмотрения колеблемости производственного процесса. Данная колеблемость измеряется обобщающим показателем, r, который рассчитывается как разность между минимальным и максимальным индивидуальными измерениями выборки. В нашем анализе эта разность заменена средней величиной, $\bar{r}$, вычисленной по тем же индивидуальным измерениям выборки (Табл. 14.1). Данная замена позволяет нам, с позиции теории ТАККП, утверждать следующее: если $\bar{r}$ распределяются симметрично, то колеблемость производственного процесса находится в пределах нормы, и, следовательно, влияние на него видимых причин (неустраненные неполадки, небрежная эксплуатация оборудования, ошибки операторов) не является существенным.

Для проверки колеблемости производственного процесса вычислим критерий репрезентативной однородности по данным четырех значений $\bar{r}$, а точнее, критерий "нормальности". Расчет этого критерия выполняется в результате проверки условий трех принципов: порядка, сходства и соответствия.

---

* Производственный процесс стабилен (в строгой постановке проблемы), если индивидуальные измерения удовлетворяют требованиям критерия "нормальности" (см. параграф 6.2).

Таблица 14.1

Размах колеблемости, в среднем

День	Смена	r_{min}	r_{max}	$\bar{r}$
Понед.	#1 #2 #3	0.61 0.68 0.70	0.75 0.83 0.90	0.68 0.76 0.80
Вторн.	#1	0.68	0.78	0.73

$$\bar{\bar{r}} = 0.7425,\ s = \sqrt{\frac{\Sigma\left(\bar{r} - \bar{\bar{r}}\right)^2}{n}} = 0.0438,\ sk = \frac{\Sigma\left(\bar{r} - \bar{\bar{r}}\right)^3}{n \cdot s^3} = -0.12$$

Принцип порядка. Условие этого принципа выполняется, так как вычисленные ЭЗР (Табл. 14.2) расположены в пределах интервалов ТЗР (Табл. 14.3). В данном случае ТЗР представлены типом С.

Таблица 14.2

ЭЗР

$\bar{r}$	$\bar{r} - \bar{\bar{r}} = d$ $\bar{\bar{r}} = 0.7425$	$1 - \frac{d}{d_{max}} = g$ $d_{max} = 0.0625$	$\frac{g}{g_{max}} = v$ $g_{max} = 0.80$	ТЗР
0.68 0.73 0.76 0.80	0.0625 0.0125 0.0175 0.0575	0.00 0.80 0.72 0.08	0.00 1.00 0.90 0.10	0 1 1 0

Принцип сходства. Условие этого принципа выполняется: распределение ЭЗР несущественно отличается от распределения ТЗР (Табл. 14.4). Об этом свидетельствует высокий коэффициент сходства, равный 0.95:

$$Kc = 1 - \left[kr(v) - kr(y)\right] = 1 - \left[1.05 - 1.00\right] = 0.95$$

Таблица 14.3

Принцип порядка

ТЗР	v
0.0 0.0 - 0.5 0.5 - 1.0 1.0	0.00 0.10 0.90 1.00

$$\bar{v} = 0.5, \quad sk = 0, \quad kr = \frac{\Sigma\left(v - \bar{v}\right)^4}{n \cdot s^4} = 1.05$$

Таблица 14.4

Распределение ТЗР

y	$p(y)$
0 1	$0.5 = q$ $0.5 = p$

$$p = \bar{v} = 0.5, \quad sk = \frac{q - p}{\sqrt{q \cdot p}} = 0, \quad kr = 3 + \frac{1 - 6p \cdot q}{p \cdot q} = 1$$

Принцип соответствия. Условие этого принципа выполняется, так как тип переменных двух распределений, ЭЗР и ТЗР, одинаковый; это тип С (Рис. 14.1).

Теперь мы можем утверждать, что средние величины, рассчитанные на основе минимальных и максимальных значений размаха колебаний, распределены нормально. Следовательно, влияние видимых причин (неустраненные неполадки, небрежная эксплуатация оборудования, ошибки операторов) не оказывают существенного влияния на производственный процесс. Это влияние образует пятипроцентное отклонение фактических данных от теоретических.

Принцип соответствия

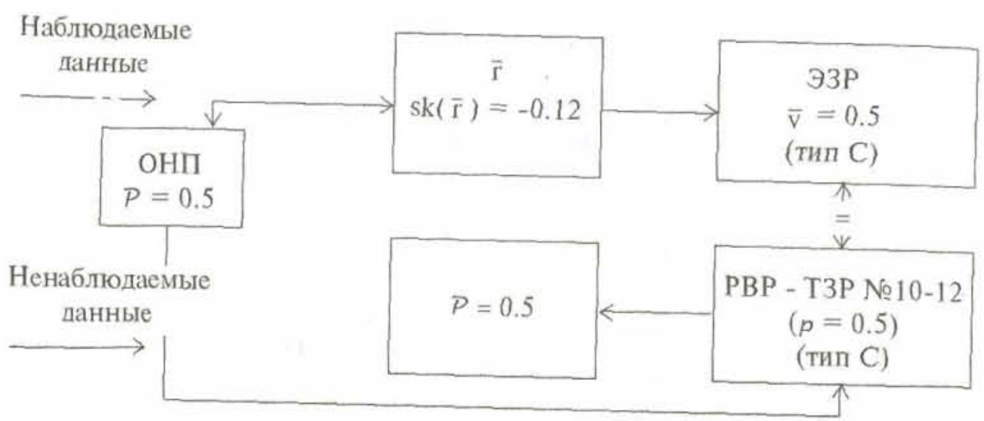


Рис. 14.1

15. Проверка сдвига центральной тенденции производственного процесса

ТАККП анализирует сдвиг центральной тенденции по данным выборочных средних. Цель этого анализа заключается в том, чтобы измерить влияние видимых причин (наладка оборудования, смена операторов, изменение техники проверки) на производственный процесс.

Анализ сдвига центральной тенденции производственного процесса производится на основе этих же четырех выборок (Табл. 15.1). Данный анализ

Таблица 15.1

Выборочные средние

День	Смена	t	$\bar{y}$
Понед.	#1	1	0.702
	#2	2	0.738
	#3	3	0.772
Вторн.	#1	4	0.725

$\bar{\bar{y}} = 0.73425$, $s = 0.02532$, $sk = 0.33$

руководствуется следующей концепцией ТАККП: сдвиг центральной тенденции производственного процесса отсутствует и влияние перечисленных выше причин не является существенным, если распределение выборочных средних нормальное. Наша проверка "нормальности" базируется на анализе условий трех принципов: порядка, сходства и соответствия.

Принцип порядка. Условие этого принципа выполняется, так как вычисленные ЭЗР расположены в пределах интервалов ТЗР (Табл. 15.2 и 15.3). В нашем примере ТЗР представлены типом С.

Таблица 15.2

Расчет ЭЗР

$\bar{y}$	$\left\lvert \bar{y} - \bar{\bar{y}} \right\rvert = d$, $\bar{\bar{y}} = 0.73425$	$1 - \frac{d}{d_{max}} = g$, $d_{max} = 0.03775$	$\frac{g}{g_{max}} = v$, $g_{max} = 0.9007$	ТЗР
0.702	0.03225	0.1457	0.16	0.00
0.725	0.00925	0.7550	0.84	1.00
0.738	0.00375	0.9007	1.00	1.00
0.772	0.03775	0.0000	0.00	0.00

Таблица 15.3

Принцип порядка

ТЗР	v
0.0	0.00
0.0 - 0.5	0.16
0.5 - 1.0	0.84
1.0	1.00

$\bar{v} = 0.5$, $sk = 0$, $kr = 1.13$

Принцип сходства. Условие этого принципа выполняется. Распределение ЭЗР несущественно отличается от распределения ТЗР (Табл. 15.4). Коэффициент сходства равен 0.87:

$$Kc = 1 - [kr(v) - kz(y)] = 1 - (1.13 - 1.00) = 0.87$$

Таблица 15.4

Распределение ТЗР

y	$p(y)$
0	$0.5 = q$
1	$0.5 = p$

$$p = \bar{v} = 0.5, \quad sk = 0, \quad kz = 1$$

Принцип соответствия. Условие этого принципа выполняется, так как тип переменных двух распределений, ЭЗР и ТЗР, одинаковый, тип С (Рис. 15.1).

Так как все условия трех принципов выполняются, распределение выборочных средних нормальное, и влияние вышеназванных видимых причин на производственный процесс не является существенным. Это влияние образует отклонение фактических данных от теоретических в пределах 13%.

Принцип соответствия

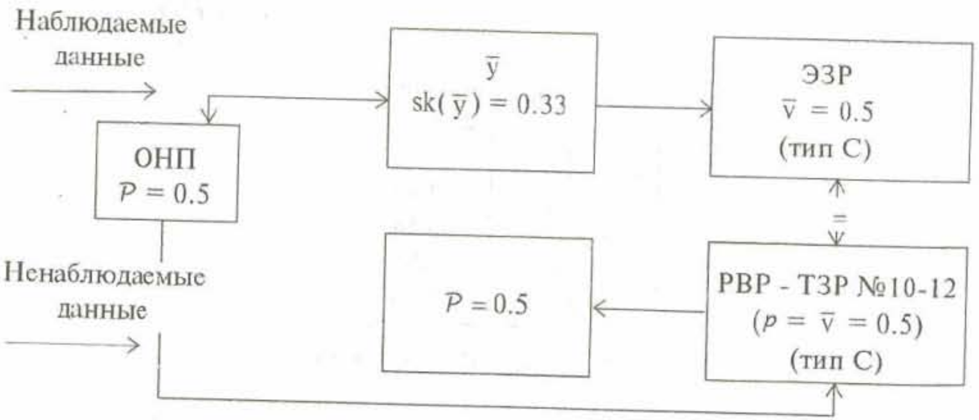


Рис. 15.1

Выводы. Аксиоматический анализ контроля качества продукции (ААККП) существенно отличается от традиционного анализа контроля качества продукции (ТАККП). Перечислим основные отличительные особенности двух методов анализа контроля качества продукции:

1. Принцип построения ТАККП основан на необоснованных предпосылках. Принцип построения ААККП основан на аксиомах. При помощи этих аксиом проверяются необоснованные предпосылки ТАККП.
2. ТАККП анализирует производственный процесс на основе обобщающих показателей. В результате этого сглаживается колеблемость индивидуальных измерений, что приводит к искажению результатов анализа. ААККП анализирует производственный процесс на основе индивидуальных измерений, что является надежной гарантией для результативных выводов.

Мы проанализировали производственный процесс, применяя ААККП и ТАККП. Результаты этих двух анализов оказались прямо противоположными, так как ТАККП базируется на необоснованных предпосылках и обобщающих показателях.

3. Результаты анализа ТАККП субъективны. Они зависят от обоснованности предпосылок и носят на себе отпечаток неопределенности. Результаты анализа ААККП объективны, они основаны на истинах, и их выводы конкретны в каждом случае.
4. Для применения ТАККП необходимо располагать минимум 15 выборками. Для применения ААККП минимальное число выборок равно четырем. В силу этого преимущества существенно сокращаются расходы на проведение анализа стабильности производственного процесса.

Заключение

"Любовь к истине и справедливости вот что делает нас подобными образу Божию."

Демосфен

Статистические методы универсальны: они применяются во всех отраслях знания для количественного измерения причинно-следственных связей. Это направление в статистике способствует познанию нашей действительности, ее законов и закономерностей. Статистическая формулировка законов и закономерностей записывается в виде причинно-следственных моделей. Логика таких моделей была четко сформулирована Б. Спинозой (Spinoza, 1951, стр. 42): "Знание следствия зависит от знания причины и предполагает его". Следовательно, причина должна быть известна. Без этого условия невозможен статистический анализ законов и закономерностей.

Традиционная статистика уделяет большое внимание анализу соотношений между генеральной совокупностью и выборкой. При этом генеральная совокупность представлена причиной, а выборка - следствием. В соответствии с вышесказанным и соблюдая принцип последовательности, следует признать, что генеральная совокупность как причина должна быть известна для того, чтобы познать следствие, т.е. выборку. Однако природа генеральной совокупности такова, что ее можно познать только абстрактно, т.е. априорно в результате анализа ее сущности, на основе абстрактной истины.

Традиционная статистика отрицает возможность априорного познания генеральной совокупности (причины) и, вопреки элементарной логике мышления, претендует на оценку причины (генеральной совокупности) с помощью следствия (выборочной совокупности). Для осуществления этого и была разработана Р. Фишером (Fisher, 1954) теория статистических оценок и выборочных распределений.

Согласно этой теории, возможно оценить некоторые параметры генеральной совокупности (например, средняя, дисперсия) на основе аналогичных обобщающих показателей, вычисленных в среднем по всем ожидаемым выборочным совокупностям, извлеченным из неизвестной генеральной совокупности. Однако данная теория не в состоянии оценить природу генеральной совокупности, т.е. ее однородность. А это корень всех проблем в статистике. Следовательно, традиционной статистике не удалось теоретически обосновать оценку причины на базе следствия.

В практическом плане традиционная статистика, имея дело с причинно-следственными моделями, делает уступку логике мышления, отказываясь от оценки причины (независимой переменной), и оценивает только следствие (зависимую переменную). При этом традиционная статистика признает, что причина должна быть известна и она определяет следствие. Однако эта уступка заячья, так как природа причины определяется не на основе истины, а на базе необоснованных предпосылок. Таким образом, теория и практика традиционной статистики казуистические и "логически несостоятельные" (Ghosh, 1988, стр. 23). Но об этом традиционная статистика скромно умалчивает.

"Наконец, я пришел к печальному заключению, что большинство статистических методов, которые я изучал по трудам таких инициаторов, как Карл Пирсон, Рональд Фишер и Джери Нейман ... логически несостоятельны."

Д. Бусу

Приложение

16. Новый коэффициент асимметрии

Коэффициент асимметрии Пирсона применяется нами для проверки репрезентативной однородности выборки. Однако мы должны иметь в виду, что в некоторых случаях этот коэффициент может ввести нас в заблуждение. Проиллюстрируем это на примере (Рис. 16, Табл. 16). Наши данные, представленные графически, имеют явную отрицательную скошенность. Однако коэффициент Пирсона, вычисленный по этим данным, положительный. Он равен 0.45. Для устранения этого противоречия мы предлагаем рассчитывать коэффициент асимметрии по следующей формуле:

$$sk_n = \frac{\Sigma\left(x - \overline{x}\right)^5 p(x)}{\Sigma p(x) \cdot \left(s_n\right)^5} = -0.13,$$

где

$$s_n = \sqrt[4]{\frac{\Sigma\left(x - \overline{x}\right)^4 p(x)}{\Sigma p(x)}} = 0.65.$$

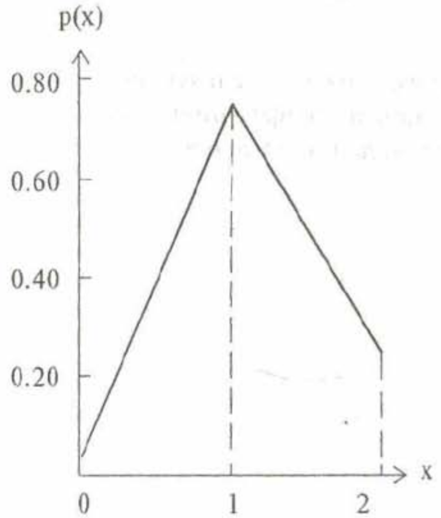


Рис. 16

Таблица 16

Расчет нового коэффициента асимметрии

x	p(x)	xp(x)	$(x-\bar{x})^2p(x)$	$(x-\bar{x})^4p(x)$	$(x-\bar{x})^3p(x)$	$(x-\bar{x})^5p(x)$
0 1 2	0.04 0.73 0.23	0.00 0.73 0.46	0.0566 0.0264 0.1509	0.0802 0.0010 0.0990	-0.0674 -0.0050 0.1223	-0.0955 -0.0002 0.0802
Итого	1.00	1.19	0.2332	0.1802	0.0499	-0.0155

$$\bar{x} = \frac{\Sigma\, xp(x)}{\Sigma\, p(x)} = 1.19$$

$$s = \sqrt{\frac{\Sigma (x-\bar{x})^2 p(x)}{\Sigma\, p(x)}} = 0.48$$

$$sk = \frac{\Sigma (x-\bar{x})^3 p(x)}{s^3} = 0.45$$

$$\bar{x} = \frac{\Sigma\, xp(x)}{\Sigma\, p(x)} = 1.19$$

$$s_n = \sqrt[4]{\frac{\Sigma (x-\bar{x})^4 p(x)}{\Sigma\, p(x)}} = 0.65$$

$$sk_n = \frac{\Sigma (x-\bar{x})^5 p(x)}{(s_n)^5} = -0.13$$

В заключение следует пожелать следующее: применяйте во всех случаях предложенный выше новый коэффициент асимметрии, с тем чтобы избежать искажения реальной действительности.

17. Библиографические ссылки

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," *2nd Inter. Symp. Information Theory* (Petrov, B.N. and Csaki, F. eds). Budapest: Akademiai Kiado.

Anderson, T.W.; Reid, DBW.; Beaton, GH. (1972), "Vitamin C and the Common Cold: A Double-Blind Trial," *Can. Med. Assoc. J.* 107: 503-508.

ASTM, (1976), *Manual on Presentation of Data and Control Chart Analysis*. Philadelphia, Pa: Committee E-11.

Bayes, Th. (1763), "An Essay Towards Solving a Problem in the Doctrine of Chances," *Philosophical Transactions* of the Royal Society of London, 53: 370-418.

Boldrini, M. (1972), *Scientific Truth and Statistical Method*. New York: Hafner Publishing Co.

Chuprov, A.A. (1959), *Essays on the Statistics Theory*. Moscow: Gostatizdat (in Russian).

Dykes, M., Meier, P. (1975), "Ascorbic Acid and the Common Cold: Evaluation of its Efficacy and Toxicity," *J. Amer. Med. Assoc.* 231, 1073-1079.

Everitt, B.S. (1977), *The Analysis of Contingency Tables*, London: Chapman and Hall.

Fienberg, S.E. (1980), *The Analysis of Cross-Classified Categorical Data*, Cambridge, MA: The MIT Press.

Fisher, R.A. (1954), *Statistical Methods for Research Workers*, New York: Hafner Publ. Company Inc.

Ghosh, J.K., editor, (1988), *Statistical Information and Likelihood. A Collection of Critical Essays* by Dr. D. Busu, New York: Springer-Verlag.

Khinchin, A. (1961), "The Frequency Theory of R. von Mises and Modern Ideas of Probability Theory", *Problem of Philosophy*. #1, pp. 91-102; #2, pp. 77-89 (in Russian).

Laurenson, Ch. (1981), *An Introduction to the Understanding, Construction, and Interpretation of Control Charts* (unpublished paper). Santa Rosa, CA: Hewlett-Packard Company.

Lexis, W. (1877), *Zur Theorie der Massenerscheinungen in der Menschlichen Gesellschaft*. Freiburg: Wagner.

Little, R.J.A. (1989), "Testing the Equality of Two Independent Binomial Proportions," *The American Statistician*, 43 283-288.

Markov, A.A. (1911), "On the Basic Principles of the Calculas of Probability and on the Law of Large Numbers," Ondar, K. (Ed.) *The Correspondence Between A.A. Markov and A.A. Chuprov on the Theory of Probability and Mathematical Statistics*, New York: Springer-Verlag, (1981).

Pauling, L. (1986), *How to Live Longer and Feel Better*. New York: Avon Books.

Pearson, E.S., Hartley, H.S. (1954), *Biometrika Tables*, vol. 1, 3rd ed. Cambridge: The University Press.

Sachs, L. (1982), *Applied Statistics*. A Handbook of Tecniques. New York: Springer-Verlag.

Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986), *Akaike Information Criterion Statistics*. Tokyo: KTK Scientific Publishers.

Shewhart, W.A. (1931), *Economic Control of Quality of Manufactured Product*. New York: D. Van Nostrand Co.

Spinoza, B. (1951), *Ethics*. New York: Hafner Publishing Co.

Taguchi, G. and Wu, Y. (1979), *Off-Line Quality Control*. Central Japan Quality Control Association, Nagaya.

Содержание

Швырков В.В.

Ш 35 Тайна традиционной статистики Запада. - М.: Финансы и статистика, 1998. - 144 с., ил.

ISBN 5-279-01946-1

Автор создал новую статистическую методологию на основе четырех аксиом и принципа монизма. Излагаются индуктивный метод мышления, философия нового качества статистических данных, критерий репрезентативной однородности одномерных и многомерных статистических совокупностей. Разработан аксиоматический анализ контроля качества продукции, новый коэффициент асимметрии. В книге подвергаются критике методологические основы традиционной статистики Запада.

Для преподавателей, аспирантов, статистиков.

Ш $\frac{0702000000-}{010(01)-98}$ 103-98 ББК 65.051

Научно-производственное издание

Швырков Владислав Васильевич

ТАЙНА ТРАДИЦИОННОЙ СТАТИСТИКИ ЗАПАДА

Ответственный редактор К.Отто
Редактор Е.В.Стадниченко
Технический редактор О.Р.Князева
Корректор Н.И.Морева

Компьютерный оригинал-макет
в издательстве Н.Ф.Бочкаревой
выполнен О.Р.Князевой
Тел.: (084-2) 57-88-77, 57-08-10

ИБ № 3843

Лицензия ЛР № 010156 от 29.01.97

Подписано в печать 05.02.98. Формат 60х84 1/16. Бумага офсетная №1.
Печать офсетная. Объем 9 усл. печ. л. Тираж 2 000 экз. Заказ № 28.

Издательство "ФИНАНСЫ И СТАТИСТИКА"
101000, г. Москва, ул. Покровка, 7
Тел.: (095) 925-47-08, 925-35-02, факс (095) 925-09-57

Отпечатано с готовых оригинал-макетов в ГУП "Облиздат", 248640, г.Калуга, пл.Старый Торг, 5.

Швырков В.В.

Ш 35 Тайна традиционной статистики Запада. - М.: Финансы и статистика, 1998. - 144 с., ил.

ISBN 5-279-01946-1

Автор создал новую статистическую методологию на основе четырех аксиом и принципа монизма. Излагаются индуктивный метод мышления, философия нового качества статистических данных, критерий репрезентативной однородности одномерных и многомерных статистических совокупностей. Разработан аксиоматический анализ контроля качества продукции, новый коэффициент асимметрии. В книге подвергаются критике методологические основы традиционной статистики Запада.

Для преподавателей, аспирантов, статистиков.

Ш $\frac{0702000000-}{010(01)-98}$ 103-98 ББК 65.051

Научно-производственное издание

Швырков Владислав Васильевич

ТАЙНА ТРАДИЦИОННОЙ СТАТИСТИКИ ЗАПАДА

Ответственный редактор К.Отто
Редактор Е.В.Стадниченко
Технический редактор О.Р.Князева
Корректор Н.И.Морева

Компьютерный оригинал-макет
в издательстве Н.Ф.Бочкаревой
выполнен О.Р.Князевой
Тел.: (084-2) 57-88-77, 57-08-10

ИБ № 3843

Лицензия ЛР № 010156 от 29.01.97

Подписано в печать *05.02.98. Формат 60x84$^1/_{16}$*. Бумага офсетная №1.
Печать офсетная. Объем 9 усл. печ. л. Тираж 2 000 экз. Заказ № 28.

Издательство "ФИНАНСЫ И СТАТИСТИКА"
101000, г. Москва, ул. Покровка, 7
Тел.: (095) 925-47-08, 925-35-02, факс (095) 925-09-57

Отпечатано с готовых оригинал-макетов в ГУП "Облиздат", 248640, г.Калуга, пл.Старый Торг, 5.