Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Владимирский государственный университет
имени Александра Григорьевича и Николая Григорьевича Столетовых»

Кафедра информационных систем и программной инженерии

МЕТОДОЛОГИЯ НАУЧНЫХ ИССЛЕДОВАНИЙ

Методические указания к лабораторным работам

Составитель
Р. И. МАКАРОВ

Владимир 2013

УДК 001 (076)
ББК 72.4 я 7
М54

Рецензент
Доктор технических наук,
профессор Владимирского государственного университета
имени Александра Григорьевича и Николая Григорьевича Столетовых
В. Н. Ланцов

Печатается по решению редакционно-издательского совета ВлГУ

М54 **Методология** научных исследований : метод. указания к лаб. работам / Владим. гос. ун-т им. А. Г. и Н. Г. Столетовых ; сост. Р. И. Макаров. – Владимир : Изд-во ВлГУ, 2013. – 35 с.

Знакомят с методами организации и проведения научных исследований, а также позволяют выработать навыки подбора, анализа и обработки научной информации по теме исследования; учат формулировать цель и задачи исследования, планировать и проводить эксперимент (физический или вычислительный), обрабатывать результаты экспериментов, оценивать их достоверность; рассматривают методы сопоставления теории (концепции, рабочей гипотезы) и эксперимента, основы формулирования научных выводов; излагают комплекс вопросов, относящихся к составлению отчетов, докладов и статей по результатам научного исследования. Лабораторные работы позволяют освоить методы выполнения научных исследований: получение научных фактов, анализ одномерных и многомерных статистических совокупностей, выявление причинно-следственных связей между тремя и более переменными; построение выводов, структурирование объекта исследования.

Предназначены для подготовки магистров по дисциплине профессионального цикла по направлению 231000 «Программная инженерия» по профилю «Информационные системы и технологии».

Рекомендованы для формирования профессиональных компетенций в соответствии с ФГОС 3-го поколения.

Ил. 3. Табл. 22. Библиогр.: 4 назв.

УДК 001 (076)
ББК 72.4 я 7

Лабораторная работа № 1
МЕТОДЫ ПОЛУЧЕНИЯ НАУЧНЫХ ФАКТОВ

1.1. Цель работы. Освоение методов получения научных фактов наблюдением за функционированием сложной системы и обработкой статистических данных.

1.2. Теоретические сведения. Наблюдение представляет собой элементарный познавательный процесс и определяется как целенаправленное, организованное, систематическое восприятие предметов и явлений внешнего мира [1]. Особым видом наблюдения является измерение. В экономико-управленческих исследованиях наблюдением можно считать работу со статистическими данными, а также организацию всевозможных опросов, получение экспертных оценок.

Эксперимент в отличие от наблюдения характеризуется активным вмешательством исследователя в ход изучаемых процессов и явлений, в положение изучаемых объектов. Эксперименты могут быть весьма дорогостоящими, поэтому нуждаются в тщательном планировании, в том числе с помощью методов математической статистики.

Если изучаемый объект отличается большими масштабами (предприятие, национальная экономика, действующий промышленный объект, общество или его часть и т. п.), то чаще всего приходится прибегать к *моделированию.* В настоящее время в связи с бурным развитием вычислительных систем натурные эксперименты все чаще заменяются компьютерным моделированием, что существенно удешевляет экспериментальные работы и одновременно обеспечивает необходимый уровень надежности выводов.

Пример. Даны наблюдения в виде статистических данных функционирования сложного объекта управления – технологической системы производства закаленного автомобильного стекла [3]. Данные измерений процесса закалки стекла приведены в файле *steclo.exe.* Процесс закалки многостадийный, содержит последовательные стадии: нагрева заготовок стекла, прессования для придания стеклу гнутой формы, закалки путем быстрого охлаждения нагретого стекла струей воздуха и стадии медленного охлаждения стекла до комнатной температуры. Процесс закалки характеризуется 47 входными режимными переменными, в числе их 32 переменные характеризуют температурный режим нагрева заготовок. Качество закаленного стекла характеризуется 8 выходными переменными, среди которых пять переменных $y1 - y5$ характеризуют отклонение гнутых стекол от заданной формы и три переменные $y6 - y8$ механическую прочность.

Необходимо провести многомерный разведочный анализ статистических данных объекта управления. Исследования будем проводить с использованием пакета для прикладного статистического анализа данных *STATISTICA* [2]. В данной лабораторной работе ознакомимся с методами основной статистики и многомерного разведочного анализа.

1) В программе *STATISTICA* создадим новую электронную таблицу и заполним ее данными наблюдений из файла *steclo.exe.*

2) Проведем многомерный разведочный анализ данных. Для уменьшения числа входных переменных по температуре нагрева заготовок (32 переменных) объединим их в группы с использованием кластерного анализа. Иерархическая классификация выдала два класса (рис. 1.1): первый класс объединяет 19 переменных (*Var*22 – *Var*24, *Var*26 – *Var*41), второй-13 (*Var*10 – *Var*21, *Var*25).

3) В каждом классе выбираем информативную переменную по температуре. Для этих целей строим диаграммы размахов переменных для каждого класса. В качестве примера приводим диаграмму размахов для входных переменных класса 2 (рис. 1.2). В качестве информативной переменной выбираем переменную *Var*11, имеющую наибольший размах.

В качестве информативной переменной в классе 1 можно выбрать переменные *Var*27 или *Var*39, имеющие наибольший размах.

4) Для уменьшения размерности объекта управления по выходным переменным проведем многомерный разведочный анализ переменных $y1 - y8$.

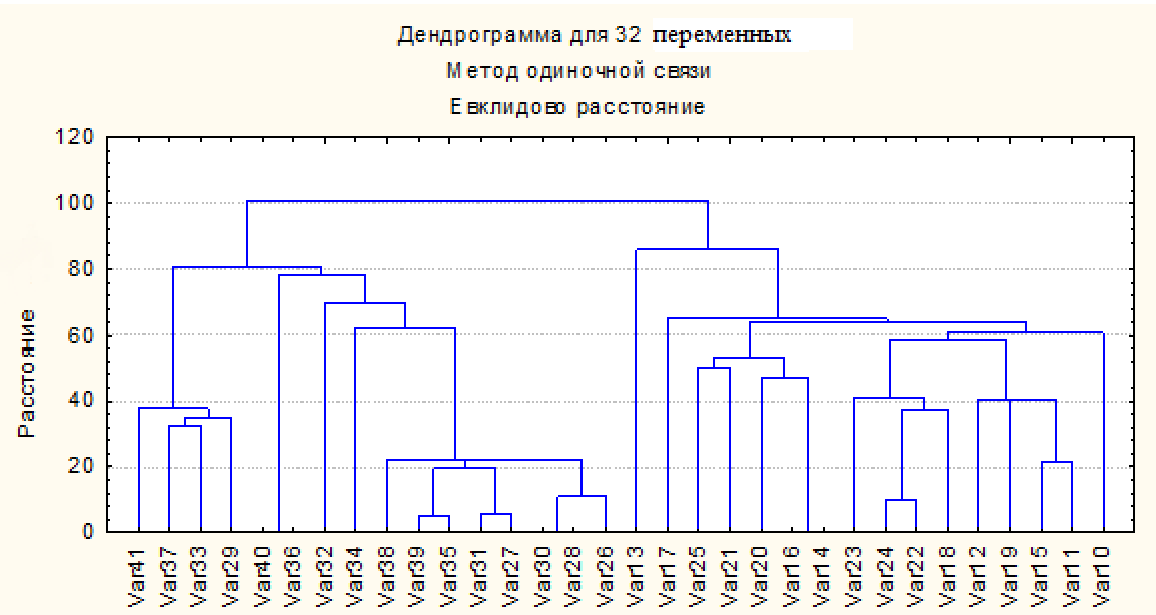

Рис. 1.1. Классификация 32 температур нагрева заготовок в печи

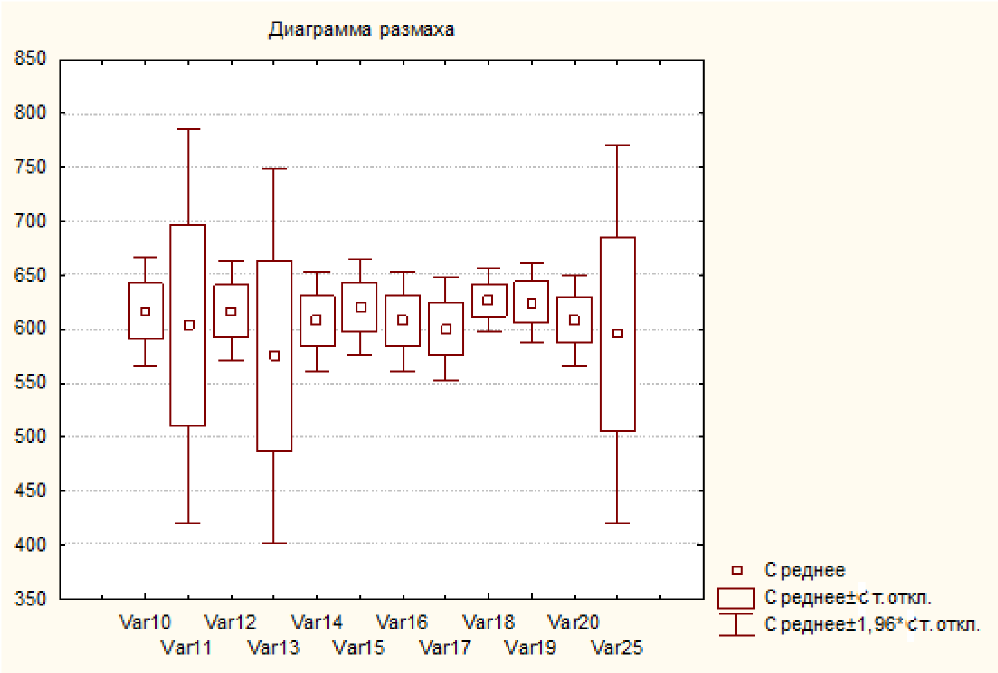

Рис.1.2. Диаграмма размаха входных переменных класса 2

Как следует из дендрограммы (рис. 1.3), можно выделить три класса выходных переменных: класс 1 образуют переменные *Var*2 – *Var*6, характеризующие отклонение гнутых стекол от заданной формы ($y1 - y5$); класс 2 – переменная *Var*7, характеризующая механическую прочность ($y6$ – максимальное количество осколков при испытаниях закаленного стекла); класс 3 – переменные *Var*8, *Var*9, характеризующие также механическую прочность ($y7$, $y8$ – минимальное количество осколков и максимальная длина осколков при испытаниях).

Для выбора информативной переменной в классе 1 проводим анализ размаха переменных $y1 - y\,5$. Наибольший размах имеют переменные *Var*3 и *Var*5, в качестве информативной выбираем переменную *Var*3 ($y2$) – отклонение формы стекла (неприлегание) на стороне *B-C*.

5) Зависимость выходных переменных объекта управления от входных может описываться системой уравнений либо совокупностью уравнений по каждой выходной переменной в отдельности. Для обоснования структуры модели проведем корреляционный анализ тесноты взаимосвязи между выходными переменными. Матрица парных коэффициентов корреляции приведена в таблице.

Матрица парных коэффициентов корреляции

Корреляции (*steclo*)	Отмеченные корреляции значимы на уровне p<0.05		
	Var3	Var7	Var8
Var3	1,00	-0,27	0,43
Var7	-0,27	1,00	0,71
Var8	0,43	0,71	1,00

Рис.1.3. Дендрограмма для восьми переменных

Так как расчетные коэффициенты корреляции меньше 0,75, то зависимость между анализируемыми переменными можно считать умеренной и слабой [2]. Полученные результаты подтверждают возможность построения взаимно независимых моделей объекта управления по каждой выходной переменной.

1.3. Задание к лабораторной работе

1) Ознакомиться с особенностями объекта управления по учебному пособию [2, с. 179 – 227].

2) Провести многомерный разведочный анализ данных для уменьшения числа входных переменных, объединив их в заданное число классов.

3) Выбрать информативные входные переменные по температуре нагрева стекла в каждом классе.

4) Провести многомерный разведочный анализ выходных данных. Для уменьшения числа выходных переменных объединить их в заданное число классов.

5) Выбрать информативные выходные переменные в каждом классе.

6) Обосновать структуру модели объекта управления, описывающей зависимость выходных переменных от входных.

7) Сделать выводы по результатам многомерного разведочного анализа данных объекта управления.

1.4. Порядок выполнения лабораторной работы

1) Получить от преподавателя индивидуальное задание.

2) Ознакомиться с исследуемым объектом управления – технологическим процессом производства закаленного автомобильного стекла по учебному пособию [2, с. 179 – 227].

3) Провести многомерный разведочный анализ данных для уменьшения числа входных переменных, объединив их в заданное число классов.

4) Выбрать информативные входные переменные в каждом классе.

5) Выбрать информативные выходные переменные в каждом классе.

6) Обосновать структуру модели объекта управления, описывающей зависимость выходных переменных от входных.

7) Сделать выводы по результатам многомерного разведочного анализа статистических данных функционирования технологического процесса производства закаленных автомобильных стекол.

8) Составить отчет по лабораторной работе.

1.5. Содержание отчета

1) Индивидуальное задание на лабораторную работу.

2) Диаграмма классификации входных переменных температуры нагрева заготовок на заданное число классов.

3) Диаграммы размаха входных переменных, построенные для каждого класса. Выбор информативных входных переменных.

4) Диаграммы размаха выходных переменных, построенные для каждого класса. Выбор информативных выходных переменных.

5) Матрица взаимных коэффициентов корреляции выходных переменных объекта управления.

1.6. Вопросы для самоконтроля

1) Методы получения научных сведений о сложной системе.

2) С какой целью проводится многомерный разведочный анализ статистических данных функционирования сложных объектов?

3) Методы уменьшения размерности входных и выходных переменных сложных систем.

4) Содержание кластерного анализа. Иерархическая классификация.

5) Обоснование структуры модели сложного объекта со многими выходными переменными.

6) Корреляционный анализ, парная корреляция, корреляционная матрица. Назначение корреляционного анализа.

1.7. Список рекомендуемой литературы

1) Пузыня, К. Ф. Технология научных исследований : учеб. пособие / К. Ф. Пузыня, А. Н. Цветков. – СПб. : СПбГИЭУ, 2002. – 54 с. – ISBN 5-88996-287-6.

2) Халафян, А. А. *STATISTICA* 6. Статистический анализ данных : учебник / А. А. Халафян. – 3-е изд. – М. : Бином-Пресс, 2007. – 512 с. – ISBN 978-5-9518-0215-6.

3) Макаров, Р. И. Информационные технологии в управлении качеством автомобильного стекла / Р. И. Макаров [и др.].– Владимир : Изд-во ВлГУ, 2010. – 275 с. – ISBN 978-5-9984-0038-4.

1.8. Варианты заданий к лабораторной работе № 1

Вариант задания	Число классов входных переменных по температуре нагрева	Число классов входных переменных по режиму прессования	Число классов входных переменных по режиму закалки	Число классов выходных переменных
1	2	1	1	2
2	4	1	1	3
3	2	2	1	3
4	2	1	2	3
5	4	2	1	3
6	4	2	2	3
7	4	1	2	2
8	4	1	2	2
9	2	2	2	2
10	2	2	1	2

Лабораторная работа № 2
АНАЛИЗ ОДНОМЕРНЫХ СТАТИСТИЧЕСКИХ СОВОКУПНОСТЕЙ

2.1. Цель работы. Освоение методов выполнения научного исследования сложных систем на основе обработки статистических данных. Анализ одномерных статистических совокупностей.

2.2. Теоретические сведения. Статистика как наука, борющаяся за истину, должна основываться на истинах, т.е. аксиомах. Формулировка статистических аксиом и разработка критерия репрезентативности однородности (КРО) – главная задача новой статистики [1].

КРО не ограничивается анализом одномерных распределений, он применим для анализа многомерных распределений и контроля качества продукции. Преимущество этого критерия в том, что он повышает уровень научных исследований, проводимых с применением статистических методов, например, таких как регрессионный и корреляционный анализ.

В книге [1] описан новый подход к статистической науке, основанный на философии нового качества статистических данных. Философия построена на аксиомах, которые формулируются в терминах теории вероятностей.

Первая аксиома. В статистике следствие есть выборка, а причина в терминах философии нового качества есть однородная невидимая генеральная совокупность (ОНГ). Эта генеральная совокупность (причина) должна быть известна априори, если мы намерены определять качество выборки, т.е. ее однородность (следствие). Предлагается следующее определение для ОНГ: совокупность двух групп однородных невидимых причин (ОНП), благоприятных (Б) и неблагоприятных (Н), образует ОНГ. Благоприятные причины оказывают положительное влияние на событие, неблагоприятные – отрицательное. Эти две группы невидимых причин однородные, если соблюдаются три условия:

- во-первых, причины должны быть независимы;
- 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.

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

$Н \cap Н = x_1 < \bar{X}$

$Б \cap Б = x_2 < \bar{X}$

$Н \cap Б = x_3 = \bar{X}$

$Б \cap Н = x_4 = \bar{X}$

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».

Вторая аксиома формулируется так: статистическое индивидуальное наблюдение есть сложное событие или однородная неделимая выборка, формирующаяся под влиянием ОНП.

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 (P) однородной невидимой генеральной совокупности. Данная вероятность определяет размеры средних величин ОНГ и репрезентативной выборки.

Исходя из этого делаются следующие выводы относительно выборки, репрезентативной ОНГ:

1) Средняя величина выборки ($\bar{x}$) равна средней величине ОНГ (X) ($\bar{x} = \bar{X}$), если обе переменные выражены в одинаковых единицах измерения и имеют одинаковые границы колебаний.

2) Вид статистического распределения (в терминах коэффициента асимметрии sk) аналогичен виду распределения ОНГ (в терминах вероятности благоприятного события P).

$$\bar{x} = \bar{X}$$

sk > 0, если $P < 0{,}5$

sk = 0, если $P = 0{,}5$

sk < 0, если $P > 0{,}5$

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 в статистических терминах.

Четвертая аксиома формулирует определение однородной статистической совокупности. Объяснение четвертой аксиомы начнем с анализа ожидаемых выборок (n = 2), извлеченных из ОНГ с P = 0,5. Ожидаемые выборки или сложные события выражаются в теоретических значениях репрезентативности (ТЗР). Первые два сложных события не являются репрезентативными в отношении ОНГ, поэтому они обозначаются нулями. Последние два сложных события репрезентативны, поэтому они обозначаются единицами.

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

Распределение вероятностей сложных событий по ТЗР образует вероятностное распределение Бернулли. Вероятность появления благоприятного события (p) этого распределения функционально зависит от P. Так как p=P=0,5, то репрезентативность распределения вероятностей сложных событий по ТЗР в отношении распределения вероятностей ОНГ равна 100 %. Такой же процент репрезентативности имеет место 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. Распределение ТЗР репрезентативно ОНГ, если соблюдаются два условия:

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

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).

Четвертая аксиома гласит: статистическая совокупность (одномерное распределение) репрезентативна ОНГ, если ее распределение эмпирического значения репрезентативности (ЭЗР) репрезентативно распределению ТЗР. Данная репрезентативность должна удовлетворять двум условиям:

1) Переменная ЭЗР и переменная ТЗР должны иметь один прообраз в виде репрезентативной совокупности сложных событий.

2) Сходство между двумя видами распределений ЭЗР и ТЗР (вид распределения характеризуется коэффициентом асимметрии и эксцесса) должно быть ровно 50 % или более (в соответствии с принципом репрезентативности).

Сходство между вышеназванными видами распределений измеряется с помощью коэффициента сходства (Кс), рассчитываемого с применением коэффициентов эксцесса, если распределение ЭЗР симметричное:

$$Kc = 1 - (kr - kz)/kz\ ,$$

где
$$kr = \sum(x - x_{\text{ср}})^4/(ns^4), \tag{2.1}$$
$$kz = 3 + (1 - 6p{\cdot}q)/(npq),$$
$$p = q = 0{,}5,\ n = 1.$$

Если распределение ЭЗР асимметричное, Kc вычисляется с применением коэффициентов асимметрии:

$$Kc = 1 - |sk - sk|\ /\ |sk|, \tag{2.2.}$$

где
$$sk = \sum(x - x_{\text{ср}})^3/(\mathrm{ns}^3),$$
$$sk = (q - p)/\sqrt{npq}, \tag{2.3}$$
$$p <> 0{,}5\ ,\ n = 1.$$

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.

Пример. Провести статистический анализ технологического процесса производства закаленного автомобильного стекла [2]. Данные наблюдений приведены в файле *steclo.*

На практике встречаются случаи, когда анализируемые случайные величины представляют собой неоднородные статистические совокупности. Последствия использования методов традиционной статистики неоднородных данных серьезны, они могут полностью аннулировать полученные статистические выводы. Качество статистической выборки характеризуется ее репрезентативностью в отношении однородной генеральной совокупности и должно удовлетворять двум условиям: однородность генеральной совокупности и репрезентативность выборки [1].

Рассмотрим использование аксиоматического анализа для оценки механических свойств гнутых закаленных автомобильных стекол. Готовые стекла подвергаются испытаниям на характер разрушений. Техническими условиями регламентируются количество осколков при разрушении и их размеры. Максимальное количество осколков не должно превышать установленных требований.

Исследования будем проводить на групповых данных испытаний закаленных стекол. Проверка репрезентативности этих данных зависит от достоверности проверяемой предпосылки: индивидуальные статистические данные испытаний равномерно распределены в пределах групп, т.е. в интервалах.

Анализируемая выборка состоит из 32 испытаний, распределенных по пяти группам по количеству осколков в испытаниях (табл. 2.1).

Таблица 2.1. Анализируемая выборка максимального числа осколков

Результаты испытаний, y	Интервал	Количество испытаний, f	Относительное количество испытаний f, д.е.	Сумма $\sum f$, д.е.
113 – 141	28	14	0,4375	0,4375
141 – 169	28	4	0,125	0,5625
169 – 197	28	4	0,125	0,6875
197 – 225	28	5	0,15625	0,84375
225 – 253	28	5	0,15625	1
Итого:		32	1	–

Для проверки репрезентативной однородности выборки групповые данные необходимо представить в виде негрупповых данных. Число негрупповых данных может равняться 4, 8, 16, 20, 24, 28, 32. Для упрощения расчетов размер выборки ограничим результатами четырех испытаний. Вторичная выборка представляется максимальным количеством осколков следующих условных испытаний: 8, 16, 24, 32. Количество осколков этих четырех испытаний $y1$–y 4 вычисляем методом накопленных частот (табл. 2.2).

Предполагаем, что испытания равномерно распределены в группах по максимальному количеству осколков. Предполагается также, что количество осколков отобранных четырех испытаний репрезентативны количеству осколков 32 испытаний. Эта предпосылка непроверяемая.

Таблица 2.2. Расчет количества осколков в четырех испытаниях методом накопленных частот

28 – 0,4375 $y_t \rightarrow (0{,}25 - 0{,}4375)$	$y1 = y_{(i)} + y_{t1}$ = 113-36=	77
28 – 0,125 $y_t \rightarrow (0{,}5 - 0{,}4375)$	$y2 = y_{(i)} + y_{t2}$ = 141+12=	153
28 – 0,15625 $y_t \rightarrow (0{,}75 - 0{,}5625)$	$y3 = y_{(i)} + y_{t3}$ = 197+36=	233
28 – 0,15625 $y_t \rightarrow (1{,}0 - 0{,}84375)$	$y4 = y_{(i)} + y_{t4}$ = 225+28=	253

Вычисления проводим по следующему алгоритму:

- определяем разности между теоретическими значениями накопленных частот и фактическими значениями (сумма $\sum f$) по первым трем группам

0,25 - 0,4375 = -0,1875; 0,5 - 0,4375 = 0,0625; 0,75 - 0,5625 = 0,1875;

- суммируем абсолютные значения разностей частот по первым трем группам

$$0{,}1875 + 0{,}0625 + 0{,}1875 = 0{,}4375;$$

- определяем длину первых трех интервалов 225 – 113 = 112;

- вычисляем ширину первых трех интервалов пропорционально отклонению накопленных частот: $y_{t1} = ((112\text{-}28)/0{,}4375)(-0{,}1875) = -36$, $y_{t2} = 12$, $y_{t3} = 36$. Величина y_{t4} принимается равной величине начального интервала разбиения $y_{t4} = 28$;

- вычисляем количество осколков в четырех испытаниях: $y1 = 113 - -36 = 77$; $y2 = 141+12 = 153$; $y3 = 197+36 = 233$; $y4 = 225+28 = 253$;

- проверяем ошибку выполненных расчетов путем сравнения графиков распределения частот по исходным пяти интервалам и расчетным четырем значениям:

- для исходной таблицы частот

$I_1 = 28(0{,}4375 + 0{,}5625 + 0{,}6875 + 0{,}84375 + 1{,}0) = 98{,}875$;

- для таблицы частот по четырем группам $I_2 = (153\text{-}77)(0{,}25 + 0{,}5)/2+ +(233\text{-}153)(0{,}5\text{-}0{,}75)/2 + (253\text{-}233)(1{,}0\text{-}0{,}75)/2 = 96$;

- погрешность вычисления по методу накопленных частот составляет $\delta = 100(I_1 - I_2)/I_1 = 2{,}9$ %, что приемлемо для дальнейших расчетов.

Максимальное количество осколков в четырех испытаниях принимаем равным

$y\ 1 = 77$, $y\ 2 = 153$, $y\ 3 = 233$, $y\ 4 = 253$.

Однородность количества осколков четырех испытаний – вторичная, но уже проверяемая предпосылка. Проверку репрезентативной однородности (РО) выборки проводим путем анализа условий трех принципов: порядка, сходства и соответствия.

Проверку репрезентативной однородности этих четырех испытаний начнем с принципа порядка.

а) *Принцип порядка*. Задача принципа порядка состоит в определении типа репрезентативного набора результатов испытаний. Условие этого принципа выполняется, т.к. вычисленное эмпирическое значение репрезентативности (ЭЗР) расположено в пределах интервалов теоретических значений репрезентативности (ТЗР) (табл. 2.3). В нашем случае ТЗР представлено типом *C* [1]. Расчет ЭЗР выполнялся в два этапа: сначала шкала относительных величин *g* заменяла шкалу абсолютных величин *y*, затем шкала эмпирических значений репрезентативности ν заменяла шкалу *g*.

Таблица 2.3. Эмпирическое значение репрезентативности

y	$d = \lvert y - y_{ср}\rvert$	$g = 1 - d/d_{max}$	$\nu = g/g_{max}$	ТЗР
77	102	0	0	0
153	26	0,74	1	1
233	54	0,47	0,63	1
253	74	0,27	0,37	0

Условие этого принципа выполняется, т.к. вычисленные ЭЗР (см. табл.2.3) расположены в пределах интервалов ТЗР (табл. 2.4). В рассматриваемом примере ТЗР представлены типом *C* [1].

Таблица 2.4. Принцип порядка

ТЗР	ЭЗР
0	0
0-0,5	0,37
0,5-1,0	0,63
1,0	1,0

Коэффициент асимметрии (2.3) исходных данных sk(y) = -0,38, среднее значение эмпирических расчетных данных $\nu_{ср} = 0,5$, среднеквадратичное отклонение $s_\nu = 0,36$, коэффициент асимметрии эмпирических значений sk(ν) = 0, коэффициент эксцесса (2.1) эмпирических значений kr (ν) = 1,76.

б) *Принцип сходства.* Выполнение этого принципа необходимо для установления сходства между распределениями ЭЗР и ТЗР. Репрезентативное распределение ТЗР представляет ожидаемое распределение в виде вероятностного распределения Бернулли. Вид этого распределения обусловливается условием $p = \nu_{ср}$ (табл. 2.5).

Таблица 2.5. Распределение ТЗР

y^*	$p(y^*)$
0	0,5 = q
1	0,5 = p
Итого	1,00

Коэффициент эксцесса ТЗР равен kz (y) = 1,0. Сходство между двумя распределениями, ЭЗР и ТЗР, измеряется коэффициентами эксцесса, т.к. коэффициент асимметрии равен нулю sk(v) = 0. Сходство между распределениями ЭЗР и ТЗР считается существенным, если отклонение коэффициента эксцесса ЭЗР kr(ν) от коэффициента эксцесса ТЗР kz (y^*) равно 50 % или меньше (2.2):

$$Кс = 1- (1,76 – 1,0)/ 1,0 = 0,24.$$

Коэффициент сходства равен 24 %. Следовательно, отклонение распределения ЭЗР от распределения ТЗР существенно, оно равно 76 %. Условие принципа сходства не выполняется. Поэтому отвергается гипотеза о том, что распределение ЭЗР одномодальное и репрезентативно распределению ТЗР.

Анализируемая выборка результатов испытаний неоднородная. Поэтому отвергается предпосылка о репрезентативной однородности выборки.

В качестве учебного примера рассмотрим выполнение принципа соответствия.

в) *Принцип соответствия.* Переменные распределений ЭЗР и ТЗР соответствуют одному типу набора однородных невидимых выборок, т.е. сложному событию (СС). В примере располагаем двумя совокупностями данных: наблюдаемые и ненаблюдаемые. Наблюдаемая совокупность представлена в абсолютных величинах (y) и в относительных величинах (ЭЗР). Ненаблюдаемая совокупность представлена в условных единицах (y^*) и в относительных величинах (ТЗР). Согласно проверяемой предпосылке наблюдаемая совокупность данных репрезентативна ОНГ. ОНГ характеризуется вероятностью благоприятного события (P).

В связи с этим мы можем утверждать (согласно третьей аксиоме), если P < 0,5, то наблюдаемая совокупность имеет положительную асимметрию, если P > 0,5 – отрицательную асимметрию, а при P = 0,5 распределение симметричное (табл. 2.6).

Таблица 2.6. Репрезентативное распределение ТЗР

P	p			
	0,0	0,25	0,5	1,0
P<0,5	-	-	D	-
P=0,5	-	-	C	D
P>0,5	-	-	D	-

или

P	p	
	p<0,5 (sk>0)	p>0,5 (sk<0)
P<0,5 (sk>*0)*	C,D	C,D
P>0,5 (sk<*0)*	C,D	C,D

В анализируемом примере наши данные имеют отрицательную асимметрию sk(y) = -0,38. Если эти данные репрезентативны ОНГ, то $P > 0{,}5$. Зная вероятности благоприятных событий ОНГ и ТЗР ($P > 0{,}5$ и $p = 0{,}5$), определяем тип набора СС, который формирует переменную распределения ТЗР. Этот тип набора соответствует типу *D*. Таким образом, наблюдаемые и ненаблюдаемые данные соответствуют разным типам наборов СС (*C* и *D*). Следовательно, переменные распределений ЭЗР и ТЗР также соответствуют разным типам СС. Из этого следует, что условие принципа соответствия не выполняется. Наша предпосылка о репрезентативности неверна, ее следует отвергнуть.

Можно заявить, что распределение результатов испытаний закаленного стекла по максимальному количеству осколков образуется под влиянием не одного существенного фактора (режима закалки), а под влиянием двух и более существенных факторов.

2.3. Задание к лабораторной работе

1) Ознакомиться с особенностями объекта управления – технологическим процессом производства закаленного автомобильного стекла по учебному пособию [2, с. 179-227].

2) Представить результаты 32 испытаний, содержащиеся в файле *steclo.exe* в виде 4 негрупповых данных.

3) Проверить репрезентативную однородность выборки выходной переменной объекта управления – показателя качества вырабатываемого закаленного стекла.

4) Составить отчет по выполненной лабораторной работе.

2.4. Порядок выполнения лабораторной работы

1) Получить от преподавателя индивидуальное задание.

2) Ознакомиться с исследуемым объектом управления – технологическим процессом производства закаленного автомобильного стекла по учебному пособию [2, с. 179-227].

3) Распределить результаты 32 испытаний по пяти группам.

4) Групповые данные представить в виде 4 негрупповых данных.

5) Проверить репрезентативную однородность 4 негрупповых данных путем проверки выполнения условий трех принципов: порядка, сходства и соответствия.

6) Сделать выводы о репрезентативной однородности данных испытаний.

2.5. Содержание отчета

1) Индивидуальное задание на лабораторную работу.

2) Таблица частот результатов измерений выходной переменной.

3) Расчеты количества осколков вторичной выборки – четырех негрупповых испытаний, выполненные методом накопленных частот.

4) Результаты проверки выполнения трех принципов: порядка, сходства и соответствия.

5) Выводы по результатам исследований.

2.6. Вопросы для самоконтроля

1) Содержание нового подхода к статистической науке, основанного на философии нового качества статистических данных.
2) Первая аксиома философии нового качества статистических данных.
3) Вторая аксиома философии нового качества статистических данных.
4) Третья аксиома философии нового качества статистических данных.
5) Четвертая аксиома философии нового качества статистических данных.
6) Алгоритм представления групповых данных в виде негрупповых данных.
7) Алгоритм вычисления эмпирического значения репрезентативности выборки.
8) Алгоритм проверки сходства между двумя распределениями, ЭЗР и ТЗР.
9) Алгоритм проверки соответствия переменных распределений ЭЗР и ТЗР одному типу набора однородных невидимых выборок.

2.7. Список рекомендуемой литературы

1) Швырков, В. В. Тайна традиционной статистики запада / В. В. Швырков. – М. : Финансы и статистика, 1998. – 144 с. – ISBN 5-279-01946-1.

2) Макаров, Р. И. Информационные технологии в управлении качеством автомобильного стекла / Р. И. Макаров [и др.]. – Владимир : Изд-во ВлГУ, 2010. – 275 с. – ISBN 978-5-9984-0038-4.

2.8. Варианты заданий к лабораторной работе № 2

Вариант задания	Проверить репрезентативную однородность выходной переменной
1	*Var*3 (*y*2) – неприлегание стекла на стороне *B-C* к шаблону
2	*Var*8 (*y*7) – минимальное количество осколков стекла при испытаниях
3	*Var*9 (*y*8) – максимальная длина осколков стекла при испытаниях
4	*Var*6 (*y*5) – отклонение образующей цилиндра поверхности стекла
5	*Var*2 (*y*1) – неприлегание стекла на стороне *A-B* к шаблону
6	*Var*4 (*y*3) – неприлегание стекла на стороне *C-D* к шаблону
7	*Var*5 (*y*4) – неприлегание стекла на стороне *D-F* к шаблону
8	*Var*7 (*y*6) – максимальное количество осколков стекла при испытаниях

Лабораторная работа № 3
АНАЛИЗ ДВУМЕРНЫХ СТАТИСТИЧЕСКИХ СОВОКУПНОСТЕЙ

3.1. Цель работы. Освоение методов выполнения научного исследования сложных систем на основе обработки статистических данных. Установление связи между двумя качественными переменными и четырьмя скрещивающимися классификациями, т.е. по статистической таблице размерами 2x2.

3.2. Теоретические сведения. При помощи статистической таблицы размерами 2x2 традиционная статистика устанавливает связь между двумя качественными переменными. Основная задача таблиц сопряженности признаков (таблиц объяснительного фактора ТОФ) заключается в анализе влияния объяснительного фактора на зависимую переменную. Анализ ТОФ основан на априорном знании генеральной совокупности [1]. Генеральная совокупность простых реальных категорий должна быть представлена дискретным распределением Бернулли, которое представляет собой однородную невидимую генеральную совокупность (ОНГ) [1]. Эта совокупность состоит из двух видов однородных невидимых причин: благоприятных (Б) и неблагоприятных (Н). Каждому виду причин присуще свое постоянное влияние на следствие и вероятность благоприятного события.

В наблюдаемых двумерных таблицах частот совместные вероятности неучтенных причин не равны нулю. Эти вероятности меньше, чем совместная вероятность учтенных причин – двумерная таблица частот будет однородной. Если влияние неучтенных причин существенно сильнее влияния учтенных причин, наблюдаемая двумерная таблица частот не будет однородной.

Пример. Рассмотрим использование аксиоматического анализа для выявления причинно-следственных связей количества потоков вырабатываемого стекла ($x13$) с максимальным количеством осколков при испытаниях изделий на характер разрушения (y 6). Представим результаты 32 испытаний в виде таблицы частот (табл. 3.1).

Проведем статистический анализ испытаний закаленных стекол, представленных в виде двумерной таблицы частот (табл. 3.1).

Построение этого распределения основано на следующей гипотезе: количество вырабатываемых потоков стекла ($x13$) – объяснительный фактор, оказывает влияние на максимальное количество осколков стекла при испытаниях (y 6).

Таблица 3.1. Двумерное распределение вырабатываемых стекол по числу потоков ($x13$) и максимальному количеству осколков при испытаниях (y 6)

$x13$	y 6		Итого
	B	$\overline{B}$	
$\overline{A}$	3	15	18
A	14	0	14
Итого:	17	15	32

Пояснения: A – два вырабатываемых потока стекла; $\overline{A}$ – один вырабатываемый поток стекла; B – количество осколков при испытаниях в диапазоне 113 – 150 шт.; $\overline{B}$ - количество осколков при испытаниях в диапазоне 151 – 253 шт.

Влияние неучтенных факторов на y 6 не должно оказывать существенного влияния. Проверка этой гипотезы есть проверка условий двух принципов: порядка и сходства. Проверку условия принципа соответствия не проводим, т.к. это условие всегда выполняется в двумерной таблице частот [1].

а) *Принцип порядка*. Условие этого принципа следующее: сумма совместных вероятностей, образованная под влиянием учтенной причины (вырабатываемые потоки стекла), должна превышать сумму совместных вероятностей, образованную под влиянием неучтенных причин (табл. 3.2).

$$\overline{A}\cap\overline{B}+A\cap B>(A\cap\overline{B}+\overline{A}\cap B). \qquad (3.1)$$

Условие этого принципа соблюдается, т.к. выполняется неравенство (3.1)

$$(0{,}46875 + 0{,}4375) > (0 + 0{,}09375);\quad 0{,}906 > 0{,}09375.$$

Таблица 3.2. Вероятностное распределение вырабатываемых стекол по числу потоков ($x13$) и максимальному количеству осколков при испытаниях (y 6)

$x13$	y 6		Итого
	B	$\bar{B}$	
$\bar{A}$	0,09375	0,46875	0,5625
A	0,4375	0	0,4375
Итого:	0,53125	0,46875	1,00

б) *Принцип сходства*. Задача состоит в определении влияния неучтенных причин на зависимую переменную (y 6). Влияние неучтенных причин считается несущественным, если отклонение наблюдаемого распределения от ненаблюдаемого (измеренное в коэффициентах асимметрии или эксцесса) меньше или равно 50 %. Этот процент установлен в соответствии с принципом репрезентативной однородности. Он равен проценту репрезентативных сложных событий в переменной типа *C*.

В рассматриваемом примере коэффициент асимметрии наблюдаемого распределения, рассчитанный по данным двумерного распределения вырабатываемых стекол, равен sk = 0,062 (табл. 3.3).

Коэффициент асимметрии ненаблюдаемого однородного распределения, рассчитанный по данным наблюдаемых безусловных вероятностей, распределенных по причинному признаку, т.е. числу потоков, равен *sk* (χ) = 0,178 (табл. 3.4). Эти безусловные вероятности использовались в качестве параметров однородной невидимой генеральной совокупности (ОНГ), т.е. *P* и *Q*.

Таблица 3.3. Наблюдаемое распределение

$x13$	p(x13)
$\bar{A}\cap\bar{B}=0$	0,46875
$\left.\begin{matrix} A\cap\bar{B} \\ \bar{A}\cap B \end{matrix}\right\} =1$	0,09375
$A\cap B$	0,4375

$\bar{x}=0{,}968,\ s=0{,}95,\ sk=0{,}062$

Таблица 3.4. Биноминальное распределение (n=2, p=0,4375)

χ	$p(\chi)$
0	0,191
1	0,492
2	0,316

$sk(\chi)$ = (g-p)/√(ngp) = 0,178

На основе этих параметров построено вероятностное биноминальное распределение с n = 2, которое свободно от влияния неучтенных причин и сопоставимо с наблюдаемым распределением.

Зная оба распределения, наблюдаемое и ненаблюдаемое, определяем природу наблюдаемого распределения, т.е. его репрезентативную однородность. Для этого вычисляем коэффициент сходства

$$\text{К}_c = 1 - abs\left(sk(x13) - sk(\chi)\right) / abs\left(sk(\chi)\right). \qquad (3.2)$$

В рассматриваемом примере коэффициент сходства, вычисленный по (3.2), равен К*c* = 0,35 (или 35 %). Так как расчетное значение К*c* меньше 50 %, то можно утверждать, что сходство между распределениями не существенно. Это означает, что влияние неучтенных причин на зависимую переменную *y* 6, количество осколков при испытаниях существенно, поэтому анализируемая выборка неоднородна и нерепрезентативна ОНГ.

Традиционная статистика не проверяет однородность выборки. Поэтому легко сделать ошибку, применяя статистические методы к наблюдаемым данным. Поэтому при статистическом анализе небольших выборок для исключения возможных ошибок необходимо проверять выборку на однородность.

3.3. Задание к лабораторной работе

1) Ознакомиться с особенностями объекта управления – технологическим процессом производства закаленного автомобильного стекла по учебному пособию [2, с. 179 – 227].

2) Провести анализ двумерной статистической совокупности с целью установления связи между двумя заданными переменными и четырьмя скрещивающимися классификациями, т.е. по статистической таблице размерами 2x2.

3) Методом регрессионного анализа (парная регрессия) традиционной статистики описать зависимость между результирующей переменной (следствие) и влияющей переменной (причина).

4) Оценить сходимость результатов исследований, полученных с использованием методов аксиоматического подхода и традиционной статистики. Сделать выводы по результатам исследований.

3.4. Порядок выполнения лабораторной работы

1) Получить от преподавателя индивидуальное задание.

2) Ознакомиться с исследуемым объектом управления – технологическим процессом производства закаленного автомобильного стекла по учебному пособию [2, с. 179 – 227].

3) Составить по варианту задания таблицу двумерного распределения режимной переменной (причина) и показателя качества стекла (следствия) в виде таблицы частот размером 2x2.

4) Сформулировать гипотезу к построенному распределению – влияния объяснительного фактора (причины) на следствие.

5) Проверить сформулированную гипотезу, т.е. проверить выполнение условий двух принципов: порядка и сходства.

6) С использованием традиционной статистики (регрессионный анализ) вычислить зависимость следствия от причины.

7) Оценить сходимость результатов исследований, полученных с использованием метода аксиоматического подхода и традиционной статистики. Сделать выводы по результатам исследований.

3.5. Содержание отчета

1) Индивидуальное задание на лабораторную работу.

2) Таблица двумерного распределения режимной переменной (причина) и показателя качества стекла (следствия) в виде таблицы частот размером 2х2.

3) Формулировка гипотезы к построенному распределению – влияния объяснительного фактора (причины) на следствие.

4) Таблица вероятностного распределения вырабатываемых стекол по режимной переменной (причина) и показателю качества стекла (следствие).

5) Результаты проверки выполнения условий двух принципов порядка и сходства.

6) Регрессионное уравнение, описывающее зависимость показателя качества стекла от режимной переменной.

7) Выводы по результатам исследований.

3.6. Вопросы для самоконтроля

1) Содержание нового подхода к анализу двумерных статистических совокупностей.

2) Формулирование гипотезы к двумерному распределению – влияния объяснительного фактора (причины) на следствие.

3) Построение двумерной таблицы частот распределения причина – следствие по данным наблюдений.

4) Построение двумерной таблицы вероятностного распределения причина – следствие по данным наблюдений.

4) Построение двумерной таблицы наблюдаемого распределения причина – следствие по данным таблицы частот.

5) Построение таблицы биноминального распределения по двумерной таблице наблюдаемого распределения.

6) Принцип порядка. Условие этого принципа.

7) Принцип сходства. Условия принципа сходства.

8) В чем ошибка традиционной статистики при анализе двумерных статистических совокупностей?

3.7. Список рекомендуемой литературы

1) Швырков, В. В. Тайна традиционной статистики запада / В. В. Швырков. – М. : Финансы и статистика, 1998. – 144 с. – ISBN 5-279-01946-1.

2) Макаров, Р. И. Информационные технологии в управлении качеством автомобильного стекла / Р. И. Макаров [и др.]. – Владимир : Изд-во ВлГУ, 2010. – 275 с. – ISBN 978-5-9984-0038-4.

3.8. Варианты заданий к лабораторной работе № 3

Вариант задания	Выполнить анализ двумерных статистических совокупностей
1	Количество потоков вырабатываемого стекла $x13$ – максимальная длина осколков стекла при испытаниях y 8
2	Высота пуансона пресса $x19$ – отклонение образующей цилиндра поверхности стекла y 5
3	Высота пуансона пресса $x19$ – неприлегание стекла на стороне *B-C* к шаблону y 2
4	Высота пуансона пресса $x19$ – неприлегание стекла на стороне *A-B* к шаблону y 1
5	Количество потоков вырабатываемого стекла $x13$ – неприлегание стекла на стороне *C-D* к шаблону y 3
6	Высота пуансона пресса $x19$ – максимальное количество осколков стекла при испытаниях y 6
7	Высота пуансона пресса $x19$ – минимальное количество осколков стекла при испытаниях y 7
8	Высота пуансона пресса $x19$ – максимальная длина осколков $y8$

Лабораторная работа № 4
АНАЛИЗ МНОГОМЕРНЫХ СТАТИСТИЧЕСКИХ СОВОКУПНОСТЕЙ

4.1. Цель работы. Освоение методов выполнения научного исследования сложных систем на основе обработки статистических данных. Проверка репрезентативной однородности причинно-следственной связи между тремя и более переменными.

4.2. Теоретические сведения. Причинно-следственные связи между тремя и более переменными могут быть подвергнуты проверке репрезентативной однородности. Трехмерная однородная невидимая генеральная совокупность (ОНГ) характеризуется функциональной связью между зависимой переменной и независимыми переменными. Она состоит из двух двумерных ОНГ, каждая из которых характеризуется независимой и зависимой переменными. В реальных выборках трехмерная статистическая совокупность характеризуется корреляционной связью между зависимой переменной и независимыми в результате влияния неучтенных причин на зависимую переменную.

Проверка однородности трехмерной статистической совокупности основана на проверяемой предпосылке: статистическая совокупность репрезентативна трехмерной ОНГ. Проверка однородности предполагает выполнение трех условий [1]:

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.

Названные три условия выполняются, если:

– связь зависимой переменной от агрегированной независимой переменной линейная;

– распределение остаточных величин симметричное.

Проанализируем вышеназванные условия на примере.

Пример. Рассмотрим пример использования аксиоматического анализа для выявления причинно-следственной связи между механическими свойствами гнутых закаленных автомобильных стекол и режимом закалки. Готовые стекла подвергаются испытаниям на характер разрушений. Техническими условиями регламентируется количество осколков при разрушении и их размеры. Максимальное количество осколков зависит от режима закалки. Мы располагаем результатами ранее проведенных исследований: зависимость максимального количества осколков от режима закалки описывается регрессионным уравнением [2]:

$$y\,6 = 442{,}3 - 26{,}6x13 - 316{,}4x18 - 8{,}0x19,$$

где $x13$ – количество потоков; $x18$ – интервал 2-й левый; $x19$ – высота пуансона.

Проведем проверку однородности трехмерной статистической совокупности: зависимости $y\ 6 = f(x18,\ x19)$. Статистические данные технологического процесса закалки автомобильного стекла приведены в файле *steklo.exe*.

Результаты 32 измерений объединяем в группы и представим в виде двух совмещенных таблиц частот размером 5х5 каждая (табл. 4.1). Наша выборка базируется на следующей проверяемой гипотезе: максимальное количество осколков (y 6) при испытаниях зависит существенно только от режима настройки пресса – интервал 2-й левый ($x18$) и высоты пуансона ($x19$). Влияние неучтенных причин не оказывает существенного влияния на максимальное количество осколков.

Таблица 4.1. Исходные данные испытаний автомобильных стекол

Переменная	$x18$					$y6$	$x19$				
Диапазон	0,3–0,36	0,36–0,37	0,37–0,38	0,38–0,39	0,39–0,4	Диапазон	5–7,6	7,6–10,2	10,2–12,8	12,8–15,4	15,4–18
Количество наблюдений	3	0	0	0	11	113–141	0	0	0	1	13
	3	0	0	0	1	141–169	0	0	0	2	2
	2	0	0	0	2	169–197	0	0	0	4	0
	1	0	0	0	4	197–225	3	0	0	2	0
	3	0	0	0	2	225–253	5	0	0	0	0

Проверку репрезентативной однородности проводим на основе негрупповых данных, которые вычисляются по групповым данным (табл. 4.2 – 4.4), формируемым с использованием метода накопленных частот (см. лаб. работу № 2).

Таблица 4.2. Накопленные частоты $x18$

x 18	F	F, %	$\sum F$, %
0,35–0,36	12	0,375	0,375
0,36–0,37	0	0	0,375
0,37–0,38	0	0	0,375
0,38–0,39	0	0	0,375
0,39–0,4	20	0,625	1
ИТОГО:	32	1	-

Таблица 4.3. Накопленные частоты $x19$

x 19	F	F, %	$\sum F$, %
5–7,6	8	0,25	0,25
7,6–10,2	0	0	0,25
10,2–12,8	0	0	0,25
12,8–15,4	9	0,28125	0,53125
15,4–18	15	0,46875	1
ИТОГО:	32	1	-

Результаты расчетов представим данными четырех условных испытаний (табл. 4.5).

Таблица 4.4. Накопленные частоты $y6$

y 6	F	F, %	$\sum F$, %
113–141	14	0,4375	0,4375
141–169	4	0,125	0,5625
169–197	4	0,125	0,6875
197–225	5	0,15625	0,84375
225–253	5	0,15625	1
ИТОГО:	32	1	-

$y\,6_1 = 77, y\,6_2 = 153, y6_3 = 233, y\,6_4 = 253.$

Таблица 4.5. Результаты условных испытаний

№ п/п	$X18$	$X19$	$Y6$
1	0,356393	7,6	77
2	0,374262	12,97	153
3	0,392131	15,227	233
4	0,4	18	253

$x18_{ср} = 0{,}381$, $sk\,(x18) = -0{,}32$,
$x19_{ср} = 0{,}381$, $sk\,(x19) = -0{,}45$.

Мы получили данные условной трехмерной статистической выборки (см. табл. 4.5), в которой влияние неучтенных причин на зависимую переменную не нулевое. Такая выборка репрезентативна трехмерной ОНГ, если влияние неучтенных причин несущественно. Проверка однородности предполагает выполнение трех условий.

***Условие* 1**: однородность независимой переменной $x18$. Для проверки репрезентативной однородности (РО) $x18$ (интервал 2-й левый) проанализируем выполнение условий трех принципов: порядка, сходства и соответствия.

а) *Условие принципа порядка* выполняется. Вычисленные эмпирические значения репрезентативности (ЭЗР) располагаются в пределах интервалов теоретических значений репрезентативности (ТЗР), которые представлены типом C (табл. 4.6, 4.7).

Таблица 4.6. ЭЗР $x18$ – интервал 2-й левый

$x\,18$	$d=\lvert x18-x18_{CP}\rvert$	$g = 1 - d / d_{MAX}$	$v = g / g_{MIN}$	ТЗР
0,356393	0,024304	0	0	0
0,374262	0,006435	0,73524389	1	1
0,392131	0,011435	0,52951221	0,720186	1
0,4	0,019304	0,20573168	0,279814	0

$v_{CP} = 0{,}5,\ kr\,(v) = 1{,}456$

б) *Условие принципа сходства* выполняется, распределение ЭЗР не отличается существенно от распределения ТЗР (табл. 4.8). Коэффициент сходства ЭЗР с ТЗР равен 54,4 %: $Kc = 1 - \lvert kr(\nu) - kz(\chi)\rvert / \lvert kz(\chi)\rvert = 1 - \lvert 1{,}456 - 1\rvert / 1 = 0{,}544$.

Таблица 4.7. Принцип порядка

ТЗР	ЭЗР
0,0	0,0
0,0 – 0,5	0,279
0,5 – 1,0	0,720
1,0	1,0
$p = v_{CP}$, $kz(x18)=1$	

в) *Условие принципа соответствия* выполняются, т.к. типы переменных двух распределений, ЭЗР и ТЗР, одинаковые – тип C.

***Условие* 1а**: однородность второй независимой переменной $x19$ (высоты пуансона) показала выполнение условий всех трех принципов: порядка, сходства и соответствия. Проверка проводилась аналогично проверке первой независимой переменной.

Таблица 4.8. Распределение ТЗР

χ	$p\ (\chi)$
0	$0{,}5 = q$
1	$0{,}5 = p$

$p = v_{CP}$, $kz\,(\chi)=1$

***Условие* 2**: независимые переменные $x18$ и $x19$ существенно отличаются друг от друга. Если две статистические совокупности репрезентативны в отношении одной и той же ОНГ, то их агрегированная совокупность также репрезентативна той же ОНГ, что свидетельствует об отсутствии существенного различия между этими статистическими совокупностями. Это будет означать несоблюдение условия 2.

Проверим различие между статистическими совокупностями $x18$ и $x19$. Для этого проверяемые переменные объединим в одну агрегированную переменную t (табл. 4.9, 4.10). Объединение проводится с использованием относительных величин $t1$ и $t2$ с одинаковыми средними с целью выполнения условия сопоставимости.

Таблица 4.9. Пересчет переменных в относительные величины

№ п/п	$x18$	$x19$	$t1=x18/x18_{ср}$	$t2=x19/x19_{ср}$
1	0,356	7,6	0,936	0,565
2	0,374	12,97	0,983	0,964
3	0,392	15,227	1,030	1,132182
4	0,4	18	1,051	1,338
Среднее	0,380697	13,44925	1,0	1,0

Таблица 4.10. Агрегированная независимая переменная t

t	$d = \|t-t_{ср}\|$	$g = 1-d/d_{мах}$	$\nu t = g/g_{мах}$	ТЗР
0,565	0,435	0	0,000	0
0,936	0,064	0,853	0,888	1
0,964	0,036	0,918	0,955	1
0,983	0,017	0,961	1,000	1
1,030	0,030	0,931	0,969	1
1,051	0,051	0,883	0,919	1
1,132	0,132	0,696	0,724	1
1,338	0,338	0,222	0,231	0

Далее проверяем репрезентативную однородность агрегированной переменной t путем анализа трех принципов: порядка, сходства и соответствия по вышеописанной методике.

Условие принципа порядка выполняется. Вычисленные ЭЗР располагаются в пределах интервалов ТЗР, который соответствует типу D [1]. Принцип сходства выполняется, т.к. распределения ЭЗР и ТЗР не отличаются существенно, коэффициент сходства $Kc = 0,82$. Следовательно, переменные $x18$ и $x19$ несущественно отличаются друг от друга, т.е. условие 2 не выполняется. Условие принципа соответствия не требует проверки, т.к. предпосылка о несущественном различии между независимыми переменными отвергается.

***Условие* 3:** влияние неучтенных причин на зависимую переменную y 6 несущественно. Это условие соблюдается, если переменная остаточных величин (разность между агрегированной независимой переменной и зависимой переменной) репрезентативна ОНГ, вероятность благоприятного события которой равна 0,5. Независимая агрегированная переменная t вычисляется как средняя арифметическая двух независимых переменных t1 и t2. Зависимую переменную y 6 преобразуем в относительную величину $h = y\ 6/y\ 6_{ср}$.

Влияние неучтенных причин на зависимую переменную формирует переменную остаточных величин r (табл. 4.11).

Таблица 4.11. Расчет остаточных величин

№ п/п	t		$h = y\ 6/y\ 6_{ср}$	$r = t - h$
1	0,565	0,75	0,430	0,320
2	0,936			
3	0,964	0,973	0,855	0,119
4	0,983			
5	1,030	1,081	1,302	-0,22
6	1,132			
7	1,051	1,194	1,413	-0,219
8	1,338			

В нашем примере связь между агрегированной независимой переменной и зависимой переменной нефункциональная в результате влияния неучтенных причин, которое формирует переменную остаточных величин r.

Если влияние неучтенных причин в отдельности несущественно, то распределение r остаточных величин симметричное. Проверка РО переменной остаточных величин является проверкой критерия «нормальности». Критерий рассчитывается в результате проверки условий трех принципов: порядка, сходства и соответствия.

а) *Принцип порядка*. Условие этого принципа выполняется, т.к. вычисленные ЭЗР (табл. 4.12) располагаются в пределах интервалов ТЗР, представленных типом C.

б) *Условие принципа сходства* не выполняется, т.к. коэффициент сходства между распределениями ЭЗР и ТЗР (см. табл.4.12) меньше 50 %

$$Kc = 1 - |\ kr\ (\nu) - kz\ (\chi)\ |\ /\ kz\ (\chi) = 0{,}0003.$$

Таблица 4.12. Расчет ЭЗР и сравнение с ТЗР

r	$d = \lvert r - r_{ср}\rvert$	$g = 1 - d/d_{мах}$	$\nu = g/\ g_{мin}$	ТЗР
0,320	0,320	0	0	0
0,119	0,119	0,629	1	1
-0,220	0,220	0,312	0,496	0
-0,219	0,219	0,317	0,504	1

$sk(r) = 0{,}27$, $\nu_{ср} = 0{,}5$, $kr(\nu) = 1{,}999$, $kz(\chi)$=1.

в) *Условия принципа соответствия* не требуют проверки, т.к. не выполнены условия принципа сходства.

В заключение следует сказать, что анализируемая трехмерная статистическая совокупность не является репрезентативной в отношении ОНГ, т.к. условия 2 и 3 не выполняются. Рассматриваемая выборка не может служить надежной основой для расчета связи между зависимой переменной y 6 и независимыми переменными x18 и x19.

Традиционная статистика применяет регрессионный анализ для исследования причинно-следственных связей между переменными без проверки РО выборки. Обрабатывая данные испытаний методами регрессионного анализа, получаем уравнение регрессии:

$$y\,6 = 408{,}88 - 293{,}26x18 - 9{,}2x19.$$

Уравнение адекватно описывает результаты испытаний. Коэффициент детерминации уравнения регрессии высокий, R^2 = 87,8 %. Он свидетельствует о значимой стохастической связи между данными максимального количества осколков при испытаниях и режимом прессования. Интерпретация полученной зависимости не согласуется с технологией закалки.

Из рассмотренного примера следует, что перед обработкой экспериментальных данных методами регрессионного анализа всегда необходимо проверять репрезентативную однородность анализируемой выборки для исключения ошибочных выводов, что мы увидели в рассматриваемом примере.

4.3. Задание к лабораторной работе

1) Ознакомиться с особенностями объекта управления – технологическим процессом производства закаленного автомобильного стекла по учебному пособию [2, с. 179 – 227].

2) Проверить репрезентативную однородность причинно-следственной связи между тремя переменными, содержащимися в индивидуальном задании.

3) Методом регрессионного анализа (множественная регрессия) традиционной статистики описать зависимость между результирующей переменной (следствие) и влияющими переменными (причина).

4) Оценить сходимость результатов исследований, полученных с использованием аксиоматического подхода и традиционной статистики. Сделать выводы по результатам исследований.

4.4. Порядок выполнения лабораторной работы

1) Получить от преподавателя индивидуальное задание.

2) Ознакомиться с исследуемым объектом управления – технологическим процессом производства закаленного автомобильного стекла по учебному пособию [2, с. 179 – 227].

3) Составить для исследования таблицу размером 32х3, заполнить ее переменными, указанными в индивидуальном задании, из файла *steclo.exe.*

4) Составить по варианту задания две таблицы частот размером 5х5 с исходными данными. Первая таблица частот содержит распределение первой режимной переменной (причина 1) и показателя качества стекла (следствия). Вторая таблица частот содержит распределение второй режимной переменной (причина 2) и показателя качества стекла (следствия).

5) Составить три таблицы накопленных частот с двумя режимными переменными и показателем качества стекла.

6) Вычислить по групповым данным негрупповые данные для двух режимных переменных и показателя качества стекла с использованием метода накопленных частот (см. лаб. работу № 2). Результаты расчетов представить данными четырех условных испытаний (условная трехмерная статистическая выборка).

7) Проверить однородность распределения двух режимных переменных (проанализировать выполнение условий трех принципов: порядка, сходства и соответствия для каждой режимной переменной).

8) Проверить различие между статистическими совокупностями первой и второй режимных переменных.

9) Проверить влияние неучтенных причин на зависимую переменную – показатель качества вырабатываемого стекла путем проверки репрезентативной однородности переменной остаточных величин (проанализировать выполнение условий трех принципов: порядка, сходства и соответствия для переменной остаточных величин).

10) Обработать данные испытаний методом регрессионного анализа. Получить уравнение регрессии, описывающее зависимость качества стекла от режима закалки.

11) Оценить сходимость результатов исследований, полученных с использованием метода аксиоматического подхода и традиционной статистики. Сделать выводы по результатам исследований.

4.5. Содержание отчета

1) Индивидуальное задание на лабораторную работу.

2) Таблица размером 32х3 с исходными данными для анализа.

3) Таблица трехмерного распределения двух режимных переменных (причин) и показателя качества стекла (следствия) в виде двух таблиц частот размером 5х5.

4) Три таблицы накопленных частот с двумя режимными переменными и показателем качества стекла.

5) Таблица условной трехмерной статистической выборки.

6) Результаты проверки однородности распределения двух режимных переменных (анализ выполнения условий трех принципов: порядка, сходства и соответствия для каждой режимной переменной).

7) Результаты проверки различия между статистическими совокупностями первой и второй режимных переменных.

8) Результаты проверки репрезентативной однородности переменной остаточных величин (анализ выполнения условий трех принципов: порядка, сходства и соответствия для переменной остаточных величин).

9) Уравнение регрессии, описывающее зависимость качества стекла от режима закалки.

10) Выводы по результатам исследований.

4.6. Вопросы для самоконтроля

1) Отличие трехмерной однородной невидимой генеральной совокупности от трехмерной статистической совокупности.

2) На чем основана проверка однородности трехмерной статистической совокупности? Проверяемые предпосылки.

3) Условия выполнения проверяемых предпосылок однородности трехмерной статистической совокупности.

4) Содержание метода проверки однородности трехмерной статистической совокупности.

5) Как определяется выполнение условия принципа порядка?

6) Как определяется выполнение условия принципа сходства?

7) Как определяется выполнение условия принципа соответствия?

8) Как проверяется существенное различие друг от друга двух независимых переменных?

4.7. Список рекомендуемой литературы

1) Швырков, В. В. Тайна традиционной статистики запада / В. В. Швырков. – М. : Финансы и статистика, 1998. – 144 с. – ISBN 5-279-01946-1.

2) Макаров, Р. И. Информационные технологии в управлении качеством автомобильного стекла / Р. И. Макаров [и др.]. – Владимир : Изд-во ВлГУ, 2010. – 275 с. – ISBN 978-5-9984-0038-4.

4.8. Варианты заданий к лабораторной работе № 4

Вариант задания	Выполнить анализ трехмерных статистических совокупностей
1	Количество потоков вырабатываемого стекла $x13$, интервал 2-й левый $x18$ – максимальное количество осколков стекла при испытаниях y 6
2	Количество потоков вырабатываемого стекла $x13$, высота пуансона $x19$ – максимальное количество осколков стекла при испытаниях y 6
3	Высота пуансона пресса $x19$ – неприлегание стекла на стороне *B*-*C* к шаблону y 2
4	Температура свода в камере 2 в зоне 2 x 6, режим прессования интервал 1-й левый $x17$ – неприлегание стекла на стороне *B*-*C* к шаблону y 2
5	Температура свода в камере 2 в зоне 2 x 6, сторона остекления $x22$ – неприлегание стекла на стороне *B*-*C* к шаблону $y2$
6	Режим прессования интервал 1-й левый $x17$, сторона остекления $x22$ – неприлегание стекла на стороне *B*-*C* к шаблону $y2$
7	Температура в камере 4 зона 12 x11, высота пуансона пресса $x19$ – минимальное количество осколков стекла при испытаниях $y7$
8	Температура в камере 4 зона 12 x11, температура в камере 3 зона 11 $x9$ – минимальное количество осколков стекла при испытаниях $y7$
9	Температура в камере 2 зона 2 x6, высота пуансона пресса $x19$ – максимальная длина осколков y 8
10	Температура в камере 2 зона 2 x6, интервал 2-й левый $x18$ – максимальная длина осколков y 8

ОГЛАВЛЕНИЕ

МЕТОДОЛОГИЯ НАУЧНЫХ ИССЛЕДОВАНИЙ

Методические указания к лабораторным работам

Составитель
МАКАРОВ Руслан Ильич

Ответственный за выпуск – зав. кафедрой профессор И. Е. Жигалов

Подписано в печать 22.04.13.
Формат 60х84/16. Усл. печ. л. 2,09. Тираж 75 экз.
Заказ
Издательство
Владимирского государственного университета
имени Александра Григорьевича и Николая Григорьевича Столетовых.
600000, Владимир, ул. Горького, 87.