

Р.И. МАКАРОВ,
Е.Р. ХОРОШЕВА

**Анализ причинно-следственных
зависимостей
по экспериментальным данным**

УДК 62.50

Владимирский
государственный
университет имени
Александра Григорьевича
и Николая Григорьевича
Столетовых,
г.Владимир

Проведен сравнительный анализ методов оценки причинно-следственных зависимостей по экспериментальным данным. Рассмотрен пример по выявлению зависимости между показателем качества закаленного стекла и режимом закалки с использованием аксиоматического анализа и традиционного метода. Показано, что при неоднородной выборке результаты регрессионного анализа могут приводить к ошибочным выводам.

Контрольные карты Шухарта широко используются в системах менеджмента качества [1]. Их рассматривают как средство оперативного управления по «сигналам разладки». Однако в авторском замысле контрольным картам отводилась роль диагностического инструмента для определения статистической управляемости процесса, т.е. наличия или отсутствия специальных причин варибельности. Смысл интерпретации контрольных карт заключается в поиске источников улучшения системы в целом, либо конкретных процессов [1]. Подозрения на существование специальных причин варибельности контролируемого процесса требуют поиска причин и выработки корректирующих воздействий по их устранению.

Причины могут выявляться на основе анализа статистических данных, например с использованием методов корреляционного, дисперсионного и регрессионного анализов. На практике встречаются случаи, когда анализируемые случайные величины представляют собой не однородные статистические совокупности. Последствия использования методов традиционной статистики неоднород-

ных данных серьезны, они могут полностью аннулировать полученные статистические выводы.

Новая статистика основана на аксиомах и разработанном критерии репрезентативности однородности для проверки предпосылок традиционной статистики. Критерий репрезентативности однородности применяется для анализа как одномерных, так и 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 и корреляционный анализы [2].

Рассмотрим пример использования аксиоматического анализа для выявления причинно-следственной связи между механическими свойствами гнутых закаленных автомобильных стекол и режимом закалки. Технологический процесс производства закаленных стекол состоит из последовательных стадий: нагрева заготовок в туннельной печи, придания гнутой формы путем прессования, закалки стекла путем быстрого охлаждения и медленного охлаждения готовых изделий. 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.

Данные анализируемой выборки представлены в виде таблицы частот (табл.1). Этого распределение построено на предпосылке: максимальное количество осколков (y) при испытаниях зависит существенно только от режима настройки пресса – интервал 2 левый (x_1) и высоты пуансона (x_2) [3]. Влияние неучтенных причин не оказывает существенного влияния на максимальное количество осколков.

Таблица 1

Исходные данные испытаний автомобильных стекол

Переменная	x_1					y	x_2				
	0,35_0,36	0,36_0,37	0,37_0,38	0,38_0,39	0,39_0,4		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

Требуется проверить достоверность этой предпосылки, т.е. проверить однородность выборки. Проверка репрезентативной однородности проводится на основе негрупповых данных, которые вычисляются по групповым данным, формируемым с использованием метода накопленных частот. Результаты расчетов (табл.2-4) представлены данными четырех условных испытаний (табл. 5).

Таблица 2.

Накопленные частоты « x_1 »

x_1	F	F%	$\Sigma F\%$
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	-

$X_{1,1}=0,35$, $X_{1,2}=0,37$, $X_{1,3}=0,39$,
 $X_{1,4}=0,4$.

Таблица 3.

Накопленные частоты « x_2 »

x_2	F	F%	$\Sigma F\%$
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	-

$X_{2,1}=7,6$, $X_{2,2}=12,97$, $X_{2,3}=15,2$,
 $X_{2,4}=18$.

Таблица 4.

Накопленные частоты « y »

Col1	f	f%	$\Sigma f\%$
113_141	14	0,4375	0,4375

Таблица 5.

Результаты условных испытаний

№пп	x_1	x_2	y
1	0,356393	7,6	77

141_169	4	0,125	0,5625	2	0,374262	12,97	153
169_197	4	0,125	0,6875	3	0,392131	15,227	233
197_225	5	0,15625	0,84375	4	0,4	18	253
225_253	5	0,15625	1				
итого:	32	1	-				

$y_1=77, y_2=153, y_3=233, y_4=253,$

$x_{1ch}=0,381, sk(x_1)=-0,32$

$x_{2ch}=0,381, sk(x_2)=-0,45$

Трехмерная однородная невидимая совокупность (ОНГ) состоит из двух двухмерных ОНГ, каждая из которых характеризуется независимой (y) и зависимой переменными x_1 и x_2 . Мы имеем данные трехмерной статистической выборки, в которых влияние неучтенных причин на зависимую переменную не нулевое. Такая выборка репрезентативна трехмерной ОНГ, если влияние неучтенных причин не существенно. Влияние не существенно при выполнении следующих условий:

Каждая независимая переменная должна быть репрезентативной соответствующей переменной ОНГ.

Независимые переменные существенно отличаются друг от друга.

Зависимая переменная существенно зависит только от двух независимых переменных.

Проанализируем выполнение выше названных условий на выборке, состоящей из результатов четырех условных испытаний (табл. 5).

Условие 1: однородность независимой переменной x_1 . Для проверки репрезентативной однородности (РО) x_1 (интервал 2 левый) проанализируем выполнение условий трех принципов: порядка, сходства и соответствия.

Условие принципа порядка выполняется. Вычисленные эмпирические значения репрезентативности (ЭЗР) располагаются в пределах интервалов теоретических значений репрезентативности (ТЗР), которые представлены типом С (табл. 6).

Таблица 6.

ЭЗР x_1 -интервал 2 левый

X	$D= X-X_{CP} $	$G=1-D/D_{МАК}$	$N=G/G_{МАК}$	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
-----	----------	------------	----------	---

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

Условие принципа сходства выполняется, распределение ЭЗР не отличается существенно от распределения ТЗР (табл. 7).

Таблица 7.

Распределение ТЗР

X	P(X)
0	0,5
1	0,5

$$P = N_{cp} \quad KZ(X) = 1$$

Коэффициент сходства (K_c) ЭЗР с ТЗР составляет 54,4%:

$$K_c = 1 - |kr(v) - kz(\chi)| / |kz(\chi)| = 1 - |1,456 - 1| / 1 = 0,544.$$

Условие принципа соответствия выполняются, т.к. типы переменных двух распределений, ЭЗР и ТЗР, одинаковые – тип С.

Проверка РО второй независимой переменной x_2 (высоты пуансона) показала выполнение условий всех трех принципов: порядка, сходства и соответствия.

Условие 2: независимые переменные x_1 и x_2 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 не соблюдение условия 2. Проверим различие между статистическими совокупностями x_1 и x_2 . Для этого проверяемые переменные объединим в одну агрегированную переменную t (табл.8, табл. 9). Объединение проводится с использованием относительных величин t_1 и t_2 .

Таблица 8

Пересчет переменных в относительные величины

п/п	x_1	x_2	$t_1 = x_1 / x_{1cp}$	$t_2 = x_2 / x_{2cp}$
1	0,356393	7,6	0,936159203	0,565087
2	0,374262	12,97	0,983096794	0,964366
3	0,392131	15,227	1,030034384	1,132182
4	0,4	18	1,050704366	1,338365
Среднее	0,380697	13,44925	0,999998687	1

Таблица 9

Агрегированная независимая переменная t

t	$d= t-t_{cp} $	$g=1-d/d_{max}$	$v_i=g/g_{max}$	ТЗР
0,565087	0,434912	0	0,000	0
0,936159	0,06384	0,853212	0,888	1
0,964366	0,035633	0,918068	0,955	1
0,983097	0,016902	0,961136	1,000	1
1,030034	0,030035	0,930939	0,969	1
1,050704	0,050705	0,883412	0,919	1
1,132182	0,132183	0,69607	0,724	1
1,338365	0,338366	0,22199	0,231	0

Далее проверяется РО агрегированной переменной t путем анализа трех принципов: порядка, сходства и соответствия по выше описанной методике. Условие принципа порядка выполняется. Вычисленные ЭЗР располагаются в пределах интервалов ТЗР, который соответствует типу D [2]. Принцип сходства выполняется, т.к. распределения ЭЗР и ТЗР не отличаются существенно, коэффициент сходства равен $K_c=0,82$. Следовательно, переменные x_1 и x_2 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 была отвергнута.

Условие 3: влияние неучтенных причин на зависимую переменную «у» несущественно. Это условие соблюдается, если 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 переменных t_1 и t_2 .

Влияние неучтенных причин на зависимую переменную формирует переменную остаточных величин r (табл. 10).

Таблица 10

Расчет остаточных величин

п/п	t		h	$r = t - h$
1	0,565087	0,750624	0,430168	0,320456
2	0,93616			
3	0,964366	0,973732	0,854749	0,118983
4	0,983098			
5	1,030036	1,081109	1,301676	-0,22057
6	1,132182			

п/п	t	h	$r = t - h$
7	1,050706	1,194536	1,413408
8	1,338365		

При несущественном влиянии неучтенных причин распределение r остаточных величин должно быть симметричным. Проверка РО переменной остаточных величин является проверкой критерия «нормальности». Расчет этого критерия выполняется в результате проверки условий трех принципов: порядка, схождения и соответствия [2].

Принцип порядка. Условие этого принципа выполняется, т.к. вычисленные ЭЗР (табл. 11) располагаются в пределах интервалов ТЗР, представленных типом С.

Условие принципа схождения не выполняется, т.к. коэффициент схождения между распределениями ЭЗР и ТЗР (табл.11) меньше 50%: $K_c = 1 - |kr(v) - kz(\chi)| / kz(\chi) = 0,0003$.

Таблица 11

Расчет ЭЗР и сравнение с ТЗР

r	$d = r - r_{cp} $	$g = 1 - d/d_{max}$	$v = g / g_{max}$	ТЗР
0,320456	0,320456	0	0	0
0,118983	0,118983	0,628707	1	1
-0,22057	0,22057	0,311701	0,495781	0
-0,21887	0,21887	0,317006	0,504219	1

$$sk(r) = 0,27, v_{cp} = 0,5, kr(v) = 1,999, kz(\chi) = 1.$$

Условия принципа соответствия не требуют проверки, т.к. не выполнены условия принципа схождения.

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 переменными « y » и независимыми переменными « x_1 и x_2 ».

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

$$y = 408,88 - 293,26x_1 - 9,2x_2.$$

Уравнение адекватно описывает результаты испытаний. Коэффициент детерминации уравнения регрессии высокий, равен $R^2=87,8\%$. Он свидетельствует о значимой стохастической связи между данными максимального количества осколков при испытаниях и режимов 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.

Литература

1. Ю.П. Адлер, В.Л. Шпер. Контрольные карты Шухарта в действии. ж. Методы менеджмента качества, №2, 2004. – С.34-37.
2. Швырков В.В. Тайна традиционной статистики Запада. – М.: Финансы и статистика, 1998. – 144 с.
3. Управление качеством автомобильного стекла: монография / Р.И. Макаров [и др.]; Владимир: Изд-во Владим. гос. ун-та, 2009.-280 с.