

АНАЛИЗ БИЗНЕС ПРОЦЕССОВ НА ОСНОВЕ СТАТИСТИЧЕСКИХ ДАННЫХ

Строителев В.Н.

дтн, профессор, Государственное бюджетное образовательное учреждение высшего профессионального образования Московской области, «Финансово-технологическая академия», г. Королев.

Жидкова Е.А.

старший преподаватель, Государственное бюджетное образовательное учреждение высшего профессионального образования Московской области, «Финансово-технологическая академия», г. Королев.

ANALYSIS OF BUSINESS PROCESSES ON THE BASIS OF STATISTICAL DATA

Stroitelev V.N.,

Zhidkova E.A.

АННОТАЦИЯ

Начальный и основной этап научного анализа бизнес процесса любой направленности включает в себя построение аналитической модели процесса. Для «владельца» процесса подобная модель должна позволять с некоторой степенью доверия проводить оценку и измерение настоящих и будущих (перспективных) характеристик анализируемой продукции, являющейся участником процесса.

SUMMARY

The initial and main stage of the scientific analysis business of process any orientation includes creation analytical model of process. For "owner" process the similar model has to allow to carry out an assessment and measurement of the present and future (perspective) characteristics the analyzed production which is the participant of process with some degree trust.

Ключевые слова: процесс; модель; качество; инновации.

Keywords: process; model; quality; innovations.

Основой для построения аналитической модели процесса служат результаты предыдущего опыта применения аналогичной продукции или результаты планируемых и проведенных экспериментальных испытаний создаваемой продукции. Каким бы сложным ни был процесс, прежде всего, должны быть определены его цели, задачи и конечные результаты. И, в конечном счете, модель должна быть достаточно простой, позволяющей наглядно и с требуемой точностью анализировать процесс, предвидеть его результаты. Следует иметь в виду, что любая модель всегда является лишь приближенной

копией реально происходящих преобразований. При выборе аналитической модели всегда необходимо исходить из возможности выбора простейших вариантов, обеспечивающих требуемую точность аппроксимации анализируемого процесса.

Рассмотрим гипотетический пример.

Предположим, что по результатам проведения и исследования некоторого бизнес процесса получены значения входного (X) и выходного (Y) показателей качества продукции. Соответствующие численные показатели приведены в табл. 1.

Таблица 1

Матрица процесса испытаний продукции

X	Y
1	2
2	5
3	9
4	15
5	9
6	5
7	7

По набору экспериментальных данных желательно построить аналитическую модель процесса, которая бы позволяла с требуемой точностью и достоверностью предсказывать и другие пары значений (x_i , y_i), как внутри анализируемого диапазона значений, так и за его пределами. Имея аналитическую модель, в дальнейшем можно прогнозировать рациональное распределение ресурсов, предвидеть возможные изменения и принимать меры для устранения негативных последствий.

Варианты различного вида аналитических моделей, построенные по результатам экспериментальных исследований, приведены в табл. 1.

Из рисунка видно, что из всех приведенных аппроксимаций статистических данных аналитическими моделями, только одна из них точно проходит через все 7 анализируемых точек. Другие модели в большей или меньшей степени имеют отклонения от наборов экспериментальных точек (x_i , y_i). В общем случае набор из n точек (x_i , y_i) может быть описан полиномом $(n - 1)$ -й степени, кривая которого проходит через все n экспериментальных точек. Наряду с этим, возникает вопрос - стоит ли добиваться такой высокой степени аппроксимации? Во-первых, сами экспериментальные точки, являясь результатами измерений, получены с некоторой погрешностью.

Поэтому любая аналитическая модель не является идеальной. Во-вторых, полиномиальная модель не адаптивна относительно числа точек эксперимента, т.к. любое их изменение приводит к необходимости нового этапа вычислений для определения всех коэффициентов полинома. И наконец, коэффициенты полинома высокой степени должны рассчитываться с предельной точностью.

Иначе ошибки аппроксимации могут превосходить любые разумные пределы.

На рис. 2 приведены полиномиальные модели различной степени – от первой (Y1) до шестой (Y6). Полиномы первой и шестой степени выделены по цвету.

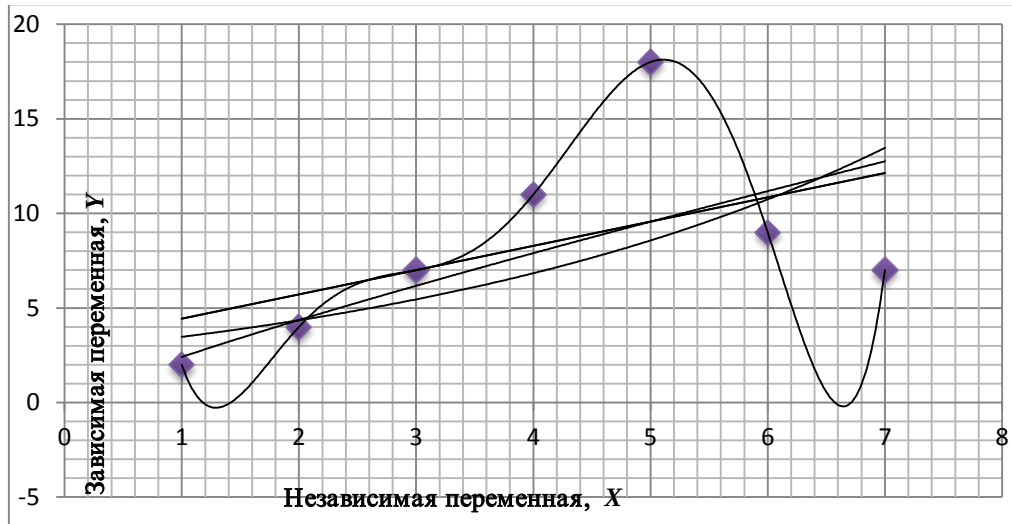


Рис. 1. Варианты аналитических моделей, построенные по результатам эксперимента

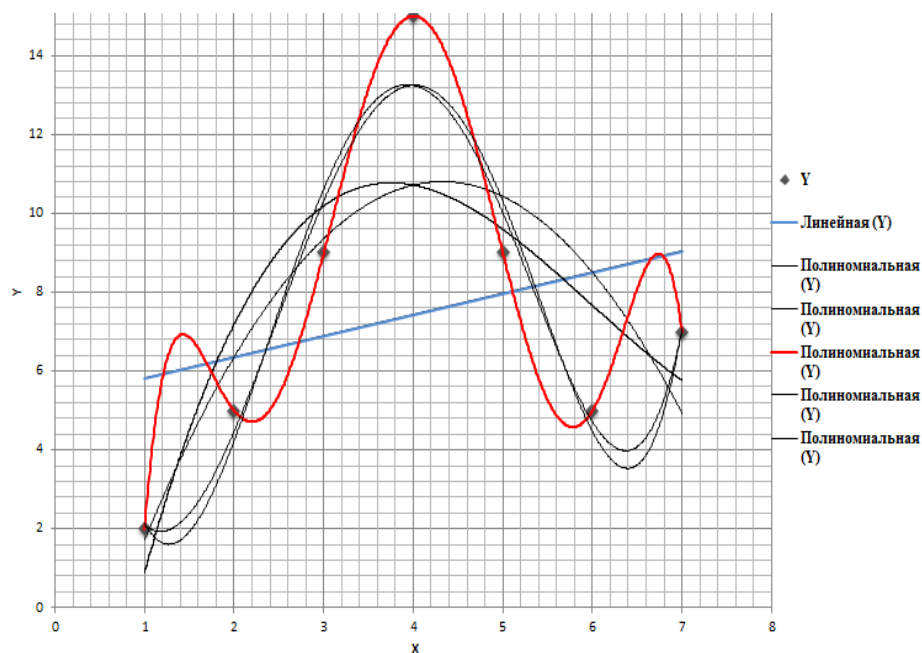


Рис. 2. Полиномиальные модели аппроксимации статистических данных

Ниже приведена система полиномиальных уравнений (1), коэффициенты каждого из которых определены по методу наименьших квадратов (МНК). Это означает, что для каждого полинома i -й степени его коэффициенты гарантируют минимальную сумму квадратов отклонений экспериментальных точек $(x_i, y_i)_{\text{эксп.}}$ от их соответствующих теоретических значений $(x_i, y_i)_{\text{теорет.}}$, получаемых из моделей аппроксимации полиномами первой (Y1),...,шестой (Y6) степени.

$$\begin{aligned} Y_1 &= 5,2857 + 0,5357x \\ Y_2 &= -4,5714 + 7,1071x - 0,8214x^2 \\ Y_3 &= -9,5714 + 12,802x - 2,4881x^2 + 0,1389x^3 \end{aligned} \quad (1)$$

$$Y_4 = 18,286 - 31,376x + 18,792x^2 - 3,8005x^3 + 0,2462x^4$$

$$Y_5 = 10,429 - 15,856x + 8,375x^2 - 0,7102x^3 - 0,1705x^4 + 0,0208x^5$$

$$Y_6 = -140 + 331,22x - 285,7x^2 + 119,56x^3 - 25,687x^4 + 2,7208x^5 - 0,1125x^6$$

Из представленного набора полиномов, попытаемся определить какой из них целесообразно использовать для практического применения. С этой целью, прежде всего, вычислим в заданных точках эксперимента X_i числовые значения зависимой переменной Y_i для каждой из шести моделей аппроксимации. Результаты вычислений приведены в табл. 2.

Таблица 2

Теоретические оценки результатов эксперимента

Числовые значения							
Х исходные	У исходные	У1 теорет.	У2 теорет.	У3 теорет.	У4 теорет.	У5 теорет.	У6 теорет.
1	2	5,821	1,719	0,881	2,148	2,088	2,001
2	5	6,357	6,366	7,191	4,237	4,473	4,994
3	9	6,893	9,371	10,192	10,615	10,305	8,975
4	15	7,428	10,734	10,717	13,249	13,203	14,947
5	9	7,964	10,455	9,599	10,019	10,186	8,913
6	5	8,500	8,534	7,671	4,709	4,163	4,869
7	7	9,036	4,971	5,768	7,017	6,428	6,806

Очевидно, что из всех шести моделей простейшей моделью является линейная аппроксимация результатов эксперимента. Другие модели более сложные, и эта сложность будет проявляться в результате дальнейшего использования и практического применения того или иного аналитического выражения.

Учитывая тот факт, что все полином описывают один и тот же набор статистических данных, применим к набору чисел У, представленных в семи столбцах табл.2, методологию дисперсионного анализа. При этом мы ясно понимаем, что в дальнейшем необходимо обосновать корректность принимаемых решений. Результаты однофакторного дисперсионного анализа приведены в табл. 3.

Таблица 3

Однофакторный дисперсионный анализ

ИТОГИ						
Группы	Счет	Сумма	Среднее	Дисперсия		
Столбец У исх.	7	52	7,428571	17,28571		
Столбец У1теор.	7	51,999	7,428429	1,33964		
Столбец У2теор	7	52,15	7,45	10,80747		
Столбец У3теор	7	52,019	7,431286	11,48407		
Столбец У4теор	7	51,994	7,427714	16,06284		
Столбец У5теор	7	50,846	7,263714	16,32554		
Столбец У6теор	7	51,505	7,357857	17,23523		
Дисперсионный анализ						
Источник вариации	SS	df	MS	F	P-Знач.	F крит.
Между группами	0,183394	6	0,030566	0,00236	0,9999	2,324
Внутри групп	543,2431	42	12,93436			
	.					
Итого	543,4265	48				

Из таблицы дисперсионного анализа следует, что все аналитические зависимости примерно с равной степенью успеха характеризуют среднее арифметическое значение набора исходных статистических данных. На уровне значимости $\alpha = 0,05$ оценки средних значений, получаемые различными моделями, статистически не различимы:

$$F = 0,00236 \ll F_{\text{крит.}} = 2,324.$$

Однако при этом выполнении дисперсионного анализа было нарушено одно из необходимых условий его применения, заключающееся в однородности дисперсий. Данные табл. 3 свидетельствуют о том, что дисперсия выборочных значений для У1теор. существенно меньше дисперсий выборок, полученных по другим моделям.

Для проверки однородности дисперсий в группах выборок используется критерий Левенэ или F-критерий. Результаты оценки существенности различий дисперсий для двух выборок (У исх. и У1теор.), выполненные с помощью программы Microsoft Excel, приведены в табл.4.

О чем свидетельствуют результаты табл. 4, и какие выводы нужно сделать при принятии решений о качестве бизнес процесса? Во-первых, линейная модель является простейшей моделью из всех рассматриваемых и ее целесообразно использовать для анализа бизнес процессов. Особенно часто возникает подобная необходимость при выполнении ориентировочных предположений о качестве анализируемой продукции. Во-вторых, дисперсии двух выборок, полученные экспериментально и рассчитанные по линейной модели, существенно различаются:

$F = 12,90322489 \gg F_{кр.} = 4,283865714$.

Поэтому, даже в случаях крайней необходимости при использовании линейной модели для аппроксимации реального бизнес процесса, характеризуемого данными табл. 1, следует делать поправки на диапазон рас-

пределения возможных результатов. Учитывая, что доверительный интервал для точечной оценки прямо пропорционален выборочному стандартному отклонению S , то доверительный интервал для анализируемого процесса (см. табл. 1), определяемый по линейной модели необходимо умножить на коэффициент 3,592.

Таблица 4

Двухвыборочный F-тест для дисперсии

	Y исх.	Y1теор.
Среднее	7,428571429	7,428428571
Дисперсия	17,28571429	1,339642952
Наблюдения	7	7
df	6	6
F	12,90322489	
P(F<=f) одностороннее	0,003331048	
F критическое одностороннее	4,283865714	

Этот коэффициент получен из данных, представленных в табл. 4:

$$\frac{\sqrt{17,28571}}{\sqrt{1,33964}} = 3,592.$$

Метод расчета доверительного интервала для предсказываемого значения зависимой переменной для линейной модели подробно изложен в работе [4], в разделе «Регрессионный анализ».

Еще раз следует подчеркнуть, что для процессов, подобных анализируемому, использование линейной модели весьма проблематично. Приведенный выше материал иллюстрирует только ход и возможные модели анализа, которые необходимо использовать для подтверждения адекватности аналитической модели набору реальных статистических данных. Некорректность непосредственного применения линейной модели лучше всего иллюстрируют данные табл. 5, в которой показано, что корреляция между исходным набором данных и теоретических данных линейной модели крайне низкая - коэффициент корреляции Пирсона $K = 0,278$.

Таблица 5

Парный двухвыборочный t-тест для средних

	Y исх.	Y1
Среднее	7,428571	7,428429
Дисперсия	17,28571	1,339643
Наблюдения	7	7
Корреляция Пирсона	0,278245	
Гипотетическая разность средних	0	
df	6	
t-статистика	9,46E-05	
P(T<=t) одностороннее	0,499964	
t критическое одностороннее	1,94318	
P(T<=t) двухстороннее	0,999928	
t критическое двухстороннее	2,446912	

Теоретически любой набор точек можно аппроксимировать линейной моделью, но адекватность полученной модели статистическому набору данных подлежит дополнительному исследованию. А если уж имеется настоятельная потребность опираться только на линейные модели, то можно попытаться разбить анализируемый интервал на перекрывающиеся участки более или менее близкие к линейным. Например, в табл. 6 представлены результаты аппроксимации статистических данных, приведенных в табл. 1, двумя прямыми, подобранными

по МНК соответственно для четырех первых и четырех последних эксперимента. Уравнения этих прямых приведены в первой строке таблицы, а их теоретические значения при 7-и фиксированных значениях $X_{исх}$ – в третьем и шестом столбцах табл. 6.

Результаты сравнений статистических данных и теоретических значений, получаемых линейной аппроксимаций представленного набора двумя прямыми с различными углами наклона, приведены в табл. 7 и табл. 8. Из рассмотрения таблиц можно сделать вывод о том, что

экспериментальные и теоретические данные статистически не различимы и между ними имеется высокий уровень корреляции.

В заключении попытаемся установить, какая простейшая модель, из приведенных в системе уравнений (1), позволяет адекватно описать характер изменения статистических данных. Анализ начнем с, простейшего из всех оставшихся, уравнения:

$$Y_2 = -4,5714 + 7,1071x - 0,8214x^2 \quad (2)$$

Как и ранее, проведем парные сравнения теоретических значений ($Y_{2\text{теор}}$) с исходными опытными данными ($Y_{\text{исх}}$) при соответствующих значениях независимой переменной ($X_{\text{исх}}$). Исследования по сравнению статистических и теоретических результатов будем проводить, опираясь на опции пакета «Анализ данных» программы Microsoft Excel.

Прежде всего, определим, статистически различимы и нет дисперсии двух пар данных (см. табл. 2). Результаты вычислений с помощью программы Microsoft Excel приведены в табл.9.

Таблица 6

Сплайновая аппроксимация результатов эксперимента

$Y_{11} = -3 + 4,3X$			$Y_{12} = 24,4 - 2,8x$		
X исх.	Y исх.	Y ₁₁ тер.	X исх.	Y исх.	Y ₁₂ тер
1	2	1,3	4	15	13,2
2	5	5,6	5	9	10,4
3	9	9,9	6	5	7,6
4	15	14,2	7	7	4,8

Таблица 7

Парный двухвыборочный t-тест для средних

	Y исх.	Y ₁₁ тер.
Среднее	7,75	7,75
Дисперсия	31,58333	30,81667
Наблюдения	4	4
Корреляция Пирсона	0,987788	
Гипотетическая разность средних	0	
df	3	
t-статистика	3,8E-16	
P(T<=t) одностороннее	0,5	
t критическое одностороннее	2,353363	
P(T<=t) двухстороннее	1	
t критическое двухстороннее	3,182446	

Таблица 8

Парный двухвыборочный t-тест для средних

	Y исх.	Y ₁₂ тер
Среднее	9	9
Дисперсия	18,66667	13,06667
Наблюдения	4	4
Корреляция Пирсона	0,83666	
Гипотетическая разность средних	0	
df	3	
t-статистика	1,88E-16	
P(T<=t) одностороннее	0,5	
t критическое одностороннее	2,353363	
P(T<=t) двухстороннее	1	
t критическое двухстороннее	3,182446	

Таблица 9

Двухвыборочный F-тест для дисперсии

	Y исх	Y ₂ теор
Среднее	7,428571	7,45
Дисперсия	17,28571	10,8074727
Наблюдения	7	7
df	6	6
F	1,599422	
P(F<=f) одностороннее	0,291356	
F критическое одностороннее	4,283866	

Полученные результаты:

$$F = 1,599422 < F_{кр} = 3,238867 -$$

свидетельствуют, что между дисперсиями внутри каждой группы существенной разницы нет. Дополни-

тельные исследования по проверке статистического равенства средних, приведенные в табл. 10, подтверждают: гипотеза H_0 отклонена быть не может, т.к. $Y_{исх}$ и $Y_{2теор}$ статистически не различимы ни по дисперсиям, ни по средним значениям и между существует довольно высокий уровень корреляции.

Таблица 10

Парный двухвыборочный t-тест для средних

	$Y_{исх}$	$Y_{2теор}$
Среднее	7,428571	7,45
Дисперсия	17,28571	10,80747
Наблюдения	7	7
Корреляция Пирсона	0,789909	
Гипотетическая разность средних	0	
df	6	
t-статистика	-0,02224	
$P(T \leq t)$ одностороннее	0,49149	
t критическое одностороннее	1,94318	
$P(T \leq t)$ двухстороннее	0,98298	
t критическое двухстороннее	2,446912	

Уравнение (2) является простейшим из системы (1), предназначенным для аппроксимации исходных данных бизнес процесса.

Список литературы

1. Глоссарий по управлению качеством, стандартизации, метрологии [Текст]: учебное пособие / Под ред. Т.Е. Старцевой, О.А. Горленко. Сост. Н.П. Асташева, Т.Н. Антипова, Е.А. Гришина, А.В. Грекова, А.В. Знайда, М.В. Максимова, И.А. Меркулова, Е.С. Слуднякова, В.Н. Строителей. – Ярославль: Изд-во «Канцлер», 2010. – 184 с.
2. Основы методологии нормирования и оценки показателей качества испытаний технической продукции на соответствие установленным требованиям:

учебное пособие для вузов/ Г.П. Богданов, Е.А. Жидкова, В.Н. Строителей, Л.П. Шумская, А.В. Яшин.-М: РГУИТП, 2012.

3. Управление инновациями и качеством [Текст] / Т.Н. Антипова, Н.П. Асташева, О.А. Горленко, В.Г. Исаев, О.А. Копылов, В.А. Коновалова, Е.А. Жидкова, В.Н. Строителей, А.Г. Суслов: учебное пособие / под ред. Т.Е. Старцевой. – Москва. ФТА, 2013-300с.
4. Управление инновациями на основе априорной и статистической информации: монография / В. Н. Строителей, Е. Ю. Гаврючин, Е. А. Жидкова, Л. П. Шумская, Е. А. Росовская, Е. С. Слуднякова. - М.: РГУИТП, 2011.-260с.

НЕЧЕТКО-МНОЖЕСТВЕННЫЙ ПОДХОД В АНАЛИЗЕ ЭКОЛОГИЧЕСКОГО ВРЕДА ОТ ТЕПЛОЭЛЕКТРОСТАНЦИЙ

Стулова Наталья Вячеславовна

Аспирант, Волгоградский государственный архитектурно-строительный университет, г. Волгоград

FUZZY-MULTIPLE APPROACH IN THE ANALYSIS OF ENVIRONMENTAL HARM FROM THERMAL POWER PLANTS

Stulova Natalya, Postgraduate student, Volgograd state University of architecture and construction, Volgograd

АННОТАЦИЯ

Определено понятие риска загрязнения окружающей среды промышленным предприятием на примере теплоэлектростанций. Представлен порядок формализации исходной качественной информации с помощью аппарата нечетких множеств. Нечеткий логический вывод производится по алгоритму Мамдани.

ABSTRACT

The concept of environmental pollution risk of industrial enterprise on the thermal power plants' example is defined. The procedure of formalization quality information using fuzzy sets is considered. Fuzzy inference is made on the Mamdani algorithm.

Ключевые слова: риск; нечеткий вывод.

Keywords: risk; fuzzy conclusion.

В законе «Об охране окружающей среды» от 10.01.2002 № 7-ФЗ экологический риск понимается как вероятность наступления события, имеющего неблагоприятные последствия для природной среды и вызванного

негативным воздействием хозяйственной и иной деятельности, чрезвычайными ситуациями природного и техногенного характера.