

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Владимирский государственный университет
имени Александра Григорьевича и Николая Григорьевича Столетовых»

В. И. ДЕНИСЕНКО

ЭМПИРИОМЕТРИЯ

Учебное пособие



Владимир 2017

УДК 330.43 (075.8)

ББК 65.в 6.я73

Д33

Рецензенты:

Доктор экономических наук, профессор
зав. кафедрой экономики Владимирского филиала
Российской академии народного хозяйства и государственной службы
при Президенте Российской Федерации
О. Б. Дигилина

Доктор физико-математических наук, профессор
зав. кафедрой физики и прикладной математики
Владимирского государственного университета
имени Александра Григорьевича и Николая Григорьевича Столетовых
С. М. Аракелян

Печатается по решению редакционно-издательского совета ВлГУ

Денисенко, В. И.

Д33 Эмпириометрия : учеб. пособие / В. И. Денисенко ; Владим. гос.
ун-т им. А. Г. и Н. Г. Столетовых. – Владимир : Изд-во ВлГУ, 2017. –
124 с. – ISBN 978-5-9984-0781-9.

Излагаются способы построения математических моделей, основанных на приведении статистических зависимостей к линейному виду и последующем преобразовании к единой форме математического представления, независимо от степени нелинейности и взаимосвязи исходных данных. Разработаны методы построения моделей для процессов с малым числом наблюдений без применения метода наименьших квадратов. Предложены приемы упрощения решений для многомерных задач. Рассмотрены некоторые задачи для детерминантных процессов в экономике, применимые и к другим отраслям науки.

Предназначено для студентов бакалавров, изучающих дисциплины «Финансовая математика» и «Теория принятия управленческих решений», а также магистров направления 38.04.02 «Методы исследования в менеджменте», «Инструменты менеджмента в экономических системах». Может быть использовано аспирантами, инженерными и научными работниками в различных областях прикладных исследований.

Рекомендовано для формирования профессиональных компетенций в соответствии с ФГОС ВО.

Ил. 31. Табл. 68. Библиогр.: 11 назв.

УДК 330.43 (075.8)

ББК 65.в 6.я73

ISBN 978-5-9984-0781-9

© ВлГУ, 2017

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	5
1. ПОДБОР ФОРМУЛ ДЛЯ ОПИСАНИЯ И ПРОГНОЗИРОВАНИЯ СТАТИСТИЧЕСКИХ ПРОЦЕССОВ	8
1.1. Методы подбора аппроксимирующей линии	8
1.2. Необходимое число наблюдений	10
1.3. Допустимое количество факторных признаков.....	11
1.4. Подбор эмпирических формул	14
1.5. Оценка полученных статистических моделей.....	19
2. ПОСТРОЕНИЕ ЛИНЕЙНЫХ УРАВНЕНИЙ ДЛЯ СЛУЧАЕВ С ЧИСЛОМ НАБЛЮДЕНИЙ НЕ БОЛЕЕ ДВУХ	26
3. МЕТОД ЭКСТРЕМАЛЬНЫХ ОПОРНЫХ ТОЧЕК	27
3.1. Прямолинейная зависимость	27
3.2. Параболическая зависимость.....	33
3.3. Угловая зависимость	35
3.4. Гиперболическая зависимость.....	37
3.5. Степенная зависимость	46
3.6. Многофакторная зависимость	49
4. МЕТОД ЦЕНТРА МАСС.....	55
4.1. Прямолинейная зависимость	55
4.2. Параболическая зависимость.....	58
4.3. Гиперболическая зависимость.....	61
4.4. Степенная зависимость	68
4.5. Показательная зависимость	73
4.6. Многофакторная зависимость	77
5. МЕТОДЫ ПОДГОТОВКИ ИСХОДНЫХ ДАННЫХ	79
5.1. Оптимизация положения начала координат при построении гиперболы.....	79

5.2. Определение положения оси симметрии кривой	81
5.3. Выбор показателя степени для нелинейной зависимости	84
6. ПОСТРОЕНИЕ СЛОЖНЫХ ОДНОФАКТОРНЫХ ЗАВИСИМОСТЕЙ	86
6.1. Метод линеаризации сложных однофакторных зависимостей с помощью дробно-линейной функции	87
6.2. Метод построения модели разделением переменных в уравнении на линейную и нелинейную части.....	89
7. ПОСТРОЕНИЕ МНОГОФАКТОРНЫХ МОДЕЛЕЙ.....	92
7.1. Метод суммирования линеаризованных зависимостей	92
7.2. Сведение пространственной задачи к плоской	96
7.3. Построение многофакторных моделей при минимально допустимом количестве опытов.....	97
8. ПОСТРОЕНИЕ МОДЕЛЕЙ ДЛЯ ДЕТЕРМИНАНТНЫХ ПРОЦЕССОВ.....	100
8.1. Решение задачи о многофункциональном наращивании капитала	101
8.2. Оборачиваемость капитала с равномерно нарастающей массой.....	103
8.3. Задача о выравнивании капиталов	105
8.4. Задача об операциях между банком и его филиалами	106
8.5. Задача об управлении портфелями ценных бумаг	108
ЗАКЛЮЧЕНИЕ	111
БИБЛИОГРАФИЧЕСКИЙ СПИСОК	112
ПРИЛОЖЕНИЯ	113

ВВЕДЕНИЕ

Под эмпириометрией будем понимать научную дисциплину о построении математических моделей на основе опытных данных для стохастических и детерминантных процессов в природе и обществе. В настоящем учебном пособии рассматриваются методы построения одно- и многофакторных линейных и нелинейных моделей, выраженных уравнениями. Условно пособие объединило элементы научных дисциплин: эконометрики, планирования экспериментов, теории графов, механики и других. При этом первые две дисциплины в основном связаны со стохастическими процессами. Каждая из указанных научных дисциплин включает методы, применимые для решения широкого круга задач из различных областей науки и техники. Однако, учитывая, что пособие написано для аудитории, связанной с экономикой и управлением, оно относительно специализировано в соответствующей области задач и терминологии.

В существующей научной и учебной литературе по моделированию опытных данных имеется ряд недоговоренностей и нерешенных вопросов, допускающих различные трактовки и мнения при решении практических задач в области экономики, статистики, организации производства и бизнеса. Так, еще на предварительной стадии постановки эксперимента не определены однозначно необходимое число наблюдений, допустимое количество факторных признаков; нет четких указаний на предназначение различных показателей оценки эконометрических моделей. При подборе эмпирических формул не сказано, когда положение аппроксимирующей линии можно сравнительно точно обусловить просто при обеспечении минимума суммы отклонений от нее опытных точек, подсчитанной по отклонениям, взятым в первой степени, а когда эти отклонения следует возводить в квадрат или более высокую степень. В результате, например, метод наименьших квадратов, несмотря на ряд присущих ему недостатков, применяется без до-

статочного обоснования часто только потому, что он наиболее глубоко разработан и освоен, а альтернативные методы используются редко и практически не развиваются.

Наряду со статистическими задачами возможны ситуации, где решения достигаются с помощью детерминантных моделей, построение которых основано на использовании закономерностей и методов физики, механики, различных разделов математики. При этом путь решения может оказываться более простым и точным, требующим небольшого объема исходных данных.

Основное внимание в настоящей работе уделено специальным методам обработки статистических закономерностей без возведения исходных данных в степень. Соответственно использована возможность развития двух методов: экстремальных опорных точек и центра масс. Для них потребовалось усовершенствовать приемы линеаризации исходных данных, снижения неопределенности на стадии постановки задачи, упрощения алгоритма решения.

Решение отмеченных задач носит актуальный практический характер и служит необходимым дополнением к традиционным методам моделирования. Подобного вида работа актуальна также с теоретической точки зрения при обучении новым научным подходам студентов, магистрантов и аспирантов.

Цель учебного пособия – сформировать и представить научную дисциплину «Эмпириометрию», сочетающую разнообразные методы построения математических моделей при ограниченном объеме опытных данных в приложении к решению задач экономики и управления.

Для достижения цели решены задачи по совершенствованию и развитию существующих методов моделирования, позволяющие снизить возможность грубых ошибок при построении и использовании моделей, что составляет научную и практическую новизну и значимость учебного пособия.

Лично автором привнесен научный вклад в математическое моделирование результатов экспериментов для различных, в том числе экономических, процессов:

– разработана методика подбора формул для описания и прогнозирования статистических процессов и обобщены методы оценки статистических моделей;

– определены методы построения уравнений для процессов с малым числом наблюдений;

– для получения линейных и нелинейных зависимостей разработаны методы центра масс и усовершенствованный метод экстремальных опорных точек, исключаящий из расчетов применения метода наименьших квадратов;

– создан ряд частных методик подготовки исходных данных при построении статистических моделей;

– разработаны методы построения сложных одно- и многофакторных статистических зависимостей, а также моделей для частных детерминантных экономических процессов.

Автор стремился к максимальной доступности излагаемых материалов, поэтому практически все решения, за исключением очевидных, подкреплены численными примерами. Соответственно достигнута возможность использования материалов книги студентами вузов и любыми специалистами, интересующимися вопросами анализа и прогнозирования всевозможных статистических процессов.

Все описанные в книге методы опробованы в ходе многолетней преподавательской деятельности автора в высших учебных заведениях, подтвердили свою высокую эффективность. Можно спорить о доказательности тех или иных математических положений и приемов, о физической интерпретации моделей, но практика показала, что методы «работают», а поэтому имеют право на распространение и использование при изучении многообразных явлений научной и общественной жизни.

В целом работа, несмотря на монографический характер ее изложения, может рассматриваться как дополнение к существующей учебной литературе и использоваться в качестве учебного пособия для изучающих методы моделирования. С этой целью приведены тренировочные задания (прил. 1) и контрольный тест (прил. 2).

1. ПОДБОР ФОРМУЛ ДЛЯ ОПИСАНИЯ И ПРОГНОЗИРОВАНИЯ СТАТИСТИЧЕСКИХ ПРОЦЕССОВ

Процессы, например в экономике, подчиняются большому количеству воздействующих факторов. В результате наблюдаемые и измеряемые количественно величины, требующие оценки, имеют статистический разброс, т. е. не укладываются на какую-либо плавную линию. В таких случаях говорят о вероятностном характере таких величин. Будем их называть результирующими. Однако анализ разбросанного множества результирующих признаков в зависимости от одного или даже нескольких одновременно действующих факторных признаков является сложным и тем более сложно давать имеющимся взаимосвязям прогнозную оценку на будущее, прошлое или на промежуточные моменты сочетания факторных признаков. Поэтому результаты наблюдений стремятся описать уравнениями, системами уравнений, неравенствами и т. д. Наилучший вариант с точки зрения простоты расчетов при анализе – подбор уравнений, другими словами, замена множества наблюдений одной линией, простейшая из которых – прямая.

1.1. Методы подбора аппроксимирующей линии

На рис. 1 множество точек, обозначенных крестиками, заменено прямой линией, проведенной в самой гуще точек. При этом можно говорить о том, что наблюдаемый процесс линейный, а аппроксимирующая (приближающая) функция, т. е. прямая, проходит горизонтально. Однако проведение прямой зависит от воли и сознания человека, как говорят, является субъективным: один проведет линию выше, другой ниже, третий не горизонтально. Чтобы найти истину, можно измерить сумму отклонений S точек от линии, затем перемещать линию и в каждом положении вновь определять такую сумму, найдя в конце концов положение, когда сумма будет наименьшей. Это положение линии объективно считается наилучшим, оно обуславливается соотношением $S = |e_1| + |e_2| + \dots + |e_i| + \dots + |e_n| \rightarrow \min$, где n – число точек; i – номер произвольной точки, вертикальными черточками обозначены модули

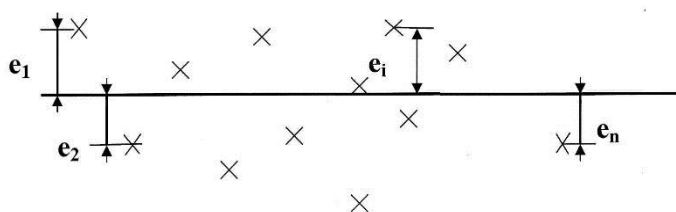


Рис. 1

вертикальными черточками обозначены модули отклонений точек от линии. Однако проведение прямой зависит от воли и сознания человека, как говорят, является субъективным: один проведет линию выше, другой ниже, третий не горизонтально. Чтобы найти истину, можно измерить сумму отклонений S точек от линии, затем перемещать линию и в каждом положении вновь определять такую сумму, найдя в конце концов положение, когда сумма будет наименьшей. Это положение линии объективно считается наилучшим, оно обуславливается соотношением $S = |e_1| + |e_2| + \dots + |e_i| + \dots + |e_n| \rightarrow \min$, где n – число точек; i – номер произвольной точки, вертикальными черточками обозначены модули

вертикальными черточками обозначены модули отклонений точек от линии. Однако проведение прямой зависит от воли и сознания человека, как говорят, является субъективным: один проведет линию выше, другой ниже, третий не горизонтально. Чтобы найти истину, можно измерить сумму отклонений S точек от линии, затем перемещать линию и в каждом положении вновь определять такую сумму, найдя в конце концов положение, когда сумма будет наименьшей. Это положение линии объективно считается наилучшим, оно обуславливается соотношением $S = |e_1| + |e_2| + \dots + |e_i| + \dots + |e_n| \rightarrow \min$, где n – число точек; i – номер произвольной точки, вертикальными черточками обозначены модули

отклонений (означают, что знаки отклонений не учитываются и берутся всегда положительными). Назовем такой метод подбора линии методом наименьших модулей отклонений [1].

Метод удобен тем, что при достаточном количестве наблюдений он малочувствителен к выбросам отдельных точек: эта погрешность теряется в общей сумме равноправных отклонений. Линия всегда будет в пределах границ расположения опытных точек. Но недостатки метода ограничивают его применение. К ним относятся неразработанность вычислительной процедуры; одинаковые суммы отклонений могут получаться при разных углах наклона прямой к осям координат.

Другой подход – определение суммы квадратов тех же отклонений обеспечивает бóльшую чувствительность поиска оптимального (наивыгоднейшего) положения прямой. При этом незначительное смещение линии, возведенное в квадрат, будет более заметным. Соотношение для суммы имеет вид $S = e_1^2 + e_2^2 + \dots + e_i^2 + \dots + e_n^2 \rightarrow \min$. Соответственно метод поиска называется методом наименьших квадратов, теория которого глубоко разработана для подбора и статистической оценки формул практически любого вида [2]. Достаточно громоздкие вычисления по данному методу стандартизованы и имеют программное обеспечение для ЭВМ. Существенные недостатки метода: высокая чувствительность к выбросам точек наблюдений и необходимость проведения вычислений практически без округлений чисел. Все значения должны иметь одинаковую точность, их ошибки следует подчинять нормальному закону распределения. Для обеспечения точности требуются массовые наблюдения. Из-за этих недостатков надежность метода не всегда удовлетворительна и возможны случаи, когда аппроксимирующая линия находится вообще в стороне от множества точек. Соответственно выводы о наблюдаемых процессах будут ошибочными. Например, если из десяти отклонений одно превышает средний уровень на 10 условных единиц, а остальные не превышают одной единицы каждый, то при возведении в квадрат наибольшее отклонение составит $10^2 = 100$ единиц, т. е. возрастет в 10 раз, а прочие – уменьшатся. Так, в случае равенства одного из отклонений 0,1 условных единиц имеем $0,1^2 = 0,01$. Значит, наибольшее отклонение сместит аппроксимирующую линию далеко за пределы расположения опытных точек, в том числе за пределы самого выброса.

Если соблюдать требования метода наименьших квадратов и проводить предварительный отбор исходных данных по равнозначности и другим критериям, то можно получать достоверные результаты. Их еще подвергают дополнительным проверкам в форме корреляционного и регрессионного анализов полученных моделей (уравнений и их систем).

Громоздкость метода наименьших квадратов растет пропорционально числу вводимых в модель факторов. При большом числе факторов в модели система линейных нормальных уравнений для расчета коэффициентов регрессии велика (равна числу факторов плюс единица). На практике может не быть программного обеспечения для решения таких систем, ручной же счет достаточно сложен или практически неосуществим, если факторов десятки.

Чувствительность поиска функции можно поднять еще выше, если искать суммы погрешностей (отклонений), возведенных в более высокую степень: $S = e_1^k + e_2^k + \dots + e_i^k + \dots + e_n^k \rightarrow \min$, но это усложняет расчеты, а недостатки метода наименьших квадратов умножаются многократно.

Таким образом, реальное практическое применение получил для описания экономических процессов метод наименьших квадратов, называемый классическим методом, или подходом.

1.2. Необходимое число наблюдений

Для использования метода наименьших квадратов обязательно определение необходимого числа наблюдений. В литературных источниках по данному вопросу нет четких рекомендаций. Имеются данные о том, что число наблюдений должно быть больше числа оцениваемых параметров, т. е. суммы свободного члена и коэффициентов в регрессионной модели [2]. Отмечается также, что число опытов должно составлять десятки, сотни и значительно превышать количество факторов в модели [6]. Такой приближенный подход неконкретен и требует проработки. Расчет необходимого количества опытов возможен на основе сравнения методов наименьших квадратов и модулей отклонений. Расхождение по ним не должно превышать заданного предела Δ .

Рассмотрим пример по данным оценки результирующего признака у для десяти групп наблюдений. Расчет искомого предела осуществим по формуле

$$\Delta = \frac{\left| \sqrt{\sum_1^n (y - \bar{y})^2} - \sum_1^n |y - \bar{y}| \right|}{\sum_1^n |y - \bar{y}|} 100 \% , \quad (1)$$

где $\bar{y} = \frac{\sum y}{n}$; n – число опытов. Данные наблюдений и промежуточных расчетов сведены в табл. 1.

Расчеты (см. табл. 1) дали результат $\Delta = \frac{|5,1668 - 10,56|}{10,56} 100 \% =$

$= 51,072 \%$. Такая ошибка вряд ли допустима и требует увеличения необходимого числа опытов, чтобы снизить ее до приемлемого уровня, обычно 5 – 10 %. Возможно решение с исключением выброса в наблюдении № 6, если признать его случайным. Тогда для оставшихся девяти опытов имеем

$$\Delta = \frac{|6,2656 - 6,04|}{6,04} 100 \% = 3,735 \% ,$$

т. е. удовлетворительный результат.

Следовательно, необходимое количество опытов обусловлено допустимой погрешностью подбора аппроксимирующей функции и рассеиванием опытных данных, и, конечно, оно должно превышать число оцениваемых параметров из условия возможности решения системы нормальных уравнений для метода наименьших квадратов.

Таблица 1

Номер наблюдения	y	$y - \bar{y}$	$(y - \bar{y})^2$
1	1,5	-0,98	0,9604
2	2	-0,48	0,2304
3	0,8	-1,68	2,8224
4	3	0,52	0,2704
5	2,5	0,02	0,0004
6	7	4,52	20,4304
7	1,3	-1,18	1,3924
8	2,2	-0,28	0,0784
9	1,8	-0,68	0,4624
10	2,7	0,22	0,0484
Сумма	24,8	10,56	26,6960

1.3. Допустимое количество факторных признаков

Для метода наименьших квадратов важно не превысить допустимый предел включаемых в модель факторов, иначе параметры уравнения регрессии оказываются статистически незначимыми, т. е. роль

отдельных факторов во влиянии на результирующий признак становится незаметной, а модель оценивается как неадекватная исследуемым процессам. Данный предел установлен, но без достаточного обоснования и конкретики. Так, по данным [2] число включаемых в модель факторов обычно меньше количества наблюдений в 6 – 7 раз (т. е. менее 7 опытов недопустимо при построении модели), а в работе [7] рекомендуется принимать это количество не более одной трети от числа наблюдений в выборке. Как видим, различие в рекомендациях превышает два раза.

Решение данной задачи о максимально допустимом количестве включаемых в регрессионную модель факторов возможно, исходя из следующих рассуждений.

В простейшем случае аппроксимации с помощью прямой, проводимой по двум точкам опытных данных, очевидна мерность пространства, т. е. количество факторов составит при числе опытов n величину $k = n/2$. Если линию проводить через три точки, что допустимо для монотонной кривой, то имеем соотношение $k = n/3$. Однако известно правило: для построения уравнения монотонной кривой необходимо и достаточно иметь пять точек, откуда $k = n/5$. При нарушении монотонной зависимости вследствие появления перегибов или экстремальных точек правило необходимо соблюдать для каждого участка монотонности. Количество опытов во всех рассмотренных случаях и далее принимается без учета их возможного дублирования с целью повышения статистической достоверности результата. В рассуждениях было принято, что при постановке опытов каждый фактор не зависит от других и принимает вышеупомянутые пять значений при условии неизменности остальных факторов. Аналогично и все остальные факторы изменяются в пределах пяти значений при постоянстве остальной совокупности факторов.

Но возможно планирование экспериментов таким образом, чтобы опыты проводились при всевозможных сочетаниях принятых значений факторов, т. е. когда все факторы изменяются одновременно в установленных пределах. Если факторы задавать только двумя предельными значениями каждого и опыты проводить лишь при всевозможных

сочетаниях этих предельных значений (уровней), то имеем связь $n = 2^k$, где n – количество опытов, k – количество факторов, 2 – число уровней. Тогда для некоторого числа уровней L получаем формулу

$$k = \lg n / \lg L. \quad (2)$$

Например, для двух уровней и числа опытов 10 имеем $k = \lg 10 / \lg 2 = 1/0,3 = 3$ фактора. Соответственно для 100 опытов $k = 30$. Но следует помнить, что в примере речь идет о постановке экспериментов, когда они проводятся при всех сочетаниях каких-то крайних значений факторов и предположении, что искомая зависимость является линейной между этими значениями каждого фактора. Возможно проведение опытов с большим числом уровней, например, при промежуточных средних значениях факторов. Однако данные средние должны участвовать в образовании программы экспериментов наравне с предельными значениями, образуя вместе с ними и между собой все возможные сочетания. В табл. 2, 3 даны примеры программы неповторяющихся опытов соответственно для двух и трех уровней факторов x с обозначениями максимальных, средних и минимальных значений в форме max, ср., min.

Таблица 2

Номер опыта	x_1	x_2	y
1	min	min	
2	max	min	
3	min	max	
4	max	max	

Таблица 3

Номер опыта	x_1	x_2	Y
1	min	min	
2	ср.	min	
3	max	min	
4	min	ср.	
5	ср.	ср.	
6	max	ср.	
7	min	max	
8	ср.	max	
9	max	max	

В обоих случаях число факторов определяется соответственно как $k = \lg 4 / \lg 2 = 0,6/0,3 = 2$; $k = \lg 9 / \lg 3 = 0,95/0,477 = 2$, т. е. программы составлены верно.

Таким образом, в соотношение (2) при определении допустимого числа факторов должны включаться минимально необходимые количества опытов и уровней с обязательным учетом всевозможных сочетаний последних при планировании экспериментов, хотя число промежуточных опытов, в том числе без соблюдения каких-либо правил по их сочетаниям может быть неограниченным.

1.4. Подбор эмпирических формул

В отличие от техники, где формализованные модели строятся, как правило, по данным, полученным на основе экспериментов в достаточном количестве, в экономике преобладает пассивное наблюдение за объективно происходящими трудно воспроизводимыми процессами. При этом фиксируются статистические данные, которые затем используются для построения математических моделей с целью последующего анализа и прогноза процессов. Однако в последнее время все больше в экономику также проникают методы экспериментирования на региональном и отраслевом уровнях. Маркетинговая деятельность предприятий и организаций по существу основана на эксперименте. Конечно, экономический эксперимент обладает определенной степенью риска, но чтобы не потерять в целом и много, приходится ставить локальные опыты в минимально допустимом количестве. Экспериментирование на государственном и даже региональном уровнях обычно приводит к тяжелым, часто необратимым последствиям. Примером тому служат различные реформы в Советском государстве: коллективизация, ликвидация частных хозяйств в городском секторе, посев кукурузы в северных районах, введение «сухого» закона и т. д. Глобальные реформы по России с развалом страны Советов, приватизация, переход к рыночной экономике и прочее привели к катастрофе в целом регионе.

Когда имеется статистический материал, возникает вопрос, какую форму кривой лучше всего для него подобрать. Ответ очевиден: наилучший вариант – это прямая или такая кривая, которая легко может быть преобразована в прямую линию, чтобы все необходимые операции по получению модели выполнялись с прямолинейной функцией.

Причины выбора линейной модели заключаются прежде всего в ее простоте и наглядности, в том числе и в графическом выражении, также в том, что линейная модель подразумевает нормальное распределение случайной величины, обеспечивающее простую и качественную оценку модели и, наконец, линейная модель дает более высокую точность прогноза, особенно в областях, где кривые имеют значительные приращения.

Если точки наблюдений непосредственно не укладываются на прямую линию, а относительно подбора конкретной криволинейной функции нет каких-либо теоретических соображений, то рекомендуется подбирать формулу из наиболее простых и общеизвестных зависимостей, характеристики которых легко определить с помощью линейных преобразований. К таким типам кривых относятся парабола, гипербола, степенная, показательная зависимости, модифицированная экспонента, S-образные кривые.

Процедура получения криволинейной зависимости заключается в преобразовании исходных данных с получением их новых значений, которые графически укладываются на линию, близкую к прямой. То есть речь идет о том, что, например, данные для x и y или только для x (или y) заменяются другими значениями X и Y , получаемыми с помощью специальных функций, обращения, логарифмирования и т. д. Значения X и Y имеют линейную связь, поэтому говорят, что данные для x и y линеаризованы. Далее выводится уравнение прямой $Y = f(X)$, $y = f(X)$ или $Y = f(x)$, по которому рассчитываются значения результирующего признака. Эти значения подвергаются далее обратному преобразованию с получением уже истинных значений, соответствующих нелинейной модели. Например, если для линеаризации пришлось логарифмировать данные для y , получая $Y = \lg y$, то после построения линейной модели рассчитываемые по ней значения Y потенцируются, в результате определяются величины y криволинейной зависимости.

Известны различные методы линеаризации зависимостей между факторным и результирующим признаками. Простейший из них – метод замены переменных. Например, если точки наблюдений укладываются на графике в декартовых простых координатах на кривую, близ-

кую к равноугольной гиперболe, то, учитывая, что такая связь имеет вид $y = f(1/x)$, делается замена $1/x = X$. При этом исходные данные для факторного признака x заменяются на обратные, т. е. на $1/x$, а по ним уже выводится уравнение прямой вида $y = a_0 + a_1X$. После численного определения a_0 и a_1 модель записывается в виде $y = a_0 + a_1 \frac{1}{x}$, которая и используется в дальнейшем для практических целей.

Другой метод – логарифмирование в основном степенных и показательных зависимостей, соответственно имеющих вид $y = ax^b$ и $y = ab^x$. После логарифмирования $\lg y = \lg a + b \lg x$ и $\lg y = \lg a + x \lg b$, а для наглядности и замен получаем линейные уравнения $Y = C + bX$ и $Y = C + xB$. Следовательно, для степенной модели необходимо прологарифмировать данные для x и y , проверить, укладываются ли они реально на прямую линию в простых декартовых координатах, и если да, то найти C и b известным способом (например, методом наименьших квадратов). Затем можно по полученной модели вычислять Y , а по ним значения $y = 10^Y$ или спотенцировать полученную формулу в целом, получив, к примеру, выражение $y = 0,8x^{1,5}$. Для показательной модели логарифмируются только исходные данные y , а величины x используются в неизменном виде. После нахождения C и B полученная модель потенцируется или используется в логарифмической форме с потенцированием только значений Y .

Следующий метод линеаризации степенных моделей – графоаналитический. Суть метода состоит в том, что исходные данные x и y наносят на неравномерные (логарифмические) декартовы координаты, через опытные точки проводят усредненную прямую, измеряют угол ее наклона α к оси x и вычисляют показатель степени $b = \operatorname{tg} \alpha$ (рис. 2). Постоянную модели находят из выражения $a_i = y/b^x$ для нескольких опытов n как среднее арифметическое $a = \sum a_i/n$.

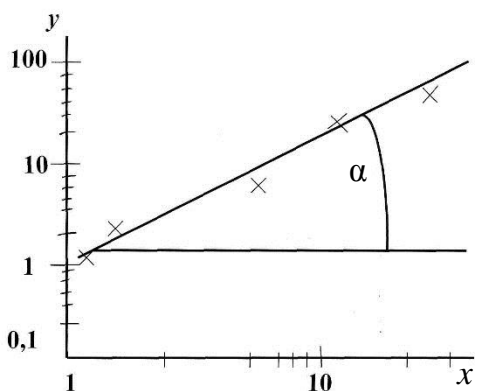


Рис. 2

При моделировании временных рядов используется линеаризация путем определения приростов. Но этот метод пригоден не только для временных рядов, отличающихся постоянным шагом фактора. Если данные любого неравномерного ряда нанести на декартовы координаты, построить точки наблюдения, провести через них усредненную кривую, то, взяв точки x на оси с равномерным расположением, можно из графика определить для них \bar{y} , получить ряд – аналог временного и определить для него средние приросты. Так, первые средние приросты находят из соотношения $\bar{u}_i = (\bar{y}_{i+1} - \bar{y}_{i-1})/2$, где i – порядковый номер точки наблюдения вдоль оси x ($i = 2, 3, \dots, n - 1$). Вторые средние приросты будут $\bar{u}^{(2)}_i = (\bar{u}_{i+1} - \bar{u}_{i-1})/2$. Средние приросты k -го порядка: $\bar{u}^{(k)}_i = (\bar{u}^{(k-1)}_{i+1} - \bar{u}^{(k-1)}_{i-1})/2$. По уровням ряда и средним приростам вычисляют их производные параметры: \bar{u}_i/\bar{y}_i ; $\lg \bar{u}_i$; $\lg(\bar{u}_i/\bar{y}_i)$; $\lg(\bar{u}_i/\bar{y}_i^2)$ и другие, используемые также для выбора вида кривой. Например, если $\bar{u}_i = \text{const}$, то наилучшая кривая – полином первого порядка $\hat{y}_i = a_0 + a_1 i$, т. е. прямая. Если \bar{u}_i растет или убывает линейно, то подходит полином второго порядка $\hat{y}_i = a_0 + a_1 i + a_2 i^2$, выражающий параболу. Линейное изменение второго среднего прироста $\bar{u}^{(2)}_i$ характерно для параболы третьей степени $\hat{y}_i = a_0 + a_1 i + a_2 i^2 + a_3 i^3$. При $\bar{u}_i/\bar{y}_i = \text{const}$ имеем простую экспоненту степенного вида $\hat{y}_i = ab^i$. Для $\lg \bar{u}_i$, изменяющегося линейно, приходим к модифицированной экспоненте $\hat{y}_i = k_0 + ab^i$, где k_0 – асимптота кривой. По условию $\lg(\bar{u}_i/\bar{y}_i)$, изменяющемуся линейно, подбирают функцию $\hat{y}_i = k_0 + a^b i$, имеющую S-образную форму. Таким образом можно обуславливать и более сложные модели. Основные параметры моделей a , b , a_0 , a_j обычно определяют методом наименьших квадратов с помощью специальных систем нормальных уравнений. Вместо графического определения усредненной кривой применяется также сглаживание временных и им аналогичных рядов, например, методами простой скользящей средней, взвешенной скользящей средней, экспоненциального сглаживания. Так, при методе про-

стой скользящей средней после отсеивания аномальных (случайных) уровней ряда статистических данных рассчитывают сглаженные уровни по формулам (для интервала сглаживания, равного трем):

$$\bar{y}_1 = \frac{5y_1 + 2y_2 - y_3}{6}; \bar{y}_i = \frac{y_{i-1} + y_i + y_{i+1}}{3}; \bar{y}_n = \frac{-y_{n-2} + 2y_{n-1} + 5y_n}{6}.$$

Приросты, определяемые как разность соседних уровней ряда, можно заменить производными кривой в различных ее точках, характеризующими скорость изменения результирующего признака во времени (для временных рядов) или на единицу приращения аргумента-фактора. Эту скорость легко определить графическим дифференцированием усредненной кривой [4]. Получив график первой производной, его можно также дифференцировать, выразив вторую производную. Дифференцирование продолжается до тех пор, пока график очередной производной не выразится линией, близкой к прямой. Так завершается линейаризация. Далее по полученной прямой измеряются величины производных для всех уровней ряда – аналоги приростов. Номер последнего графика, не считая исходного, соответствует степени завершающего члена полинома.

Еще один метод линейаризации – с помощью различных функций отображения одних величин в другие. Для этой цели наиболее часто используют дробно-линейные отображения исходных данных, выражающих нелинейные зависимости, в новые значения, укладываемые на прямую. Примером дробно-линейной отображающей функции служит зависимость $Y = (y - y_1)/(x - x_1)$, где x_1 и y_1 – исходные данные какого-либо наблюдения. Если преобразование с помощью данной функции не дало необходимый результат по линейаризации, то новые значения Y можно еще раз преобразовать с помощью такой же или иной функции, и так последовательно по шагам несколько раз, пока не получится удовлетворительная линейная модель. Приведенная функция отображения удобна для линейаризации параболических зависимостей второй степени. Зависимости гиперболического вида (равносторонней гиперболы) лучше преобразовывать с помощью функции $Y = (x - x_1)/(y - y_1)$.

Вид модели многофакторной зависимости заранее установить точно – довольно сложная задача. Но в первом приближении достаточно иногда проверить вид связи результирующего признака с каждым фактором в отдельности и принять в качестве исходной модели сложную зависимость, т. е. имеющую наиболее нелинейный характер.

1.5. Оценка полученных статистических моделей

Комплекс оценочных показателей для статистических моделей достаточно обширен и не ограничивается приведенным в табл. 4 перечнем. В частности, имеются специфические показатели качества систем уравнений, временных и динамических рядов. Да и рассматриваемый перечень может быть дополнен такими показателями, как коэффициент ранговой корреляции для определения тесноты связи между ранжированными по степени выраженности порядковыми переменными, некоторыми статистическими показателями, например, показателями дисперсии и т. д. Каждый показатель имеет свою область применения при анализе качества моделей и физический смысл. Не всегда можно согласиться с дифференциацией имеющихся обобщенных показателей на более мелкие или видоизмененные с помощью простых математических действий (возведение в степень, извлечение корня и т. п.). В таком случае число показателей может быть расширено, например, путем логарифмирования имеющихся показателей, их объединения в интегральные оценки при перемножении и т. д., придавая новым параметрам свой смысл и назначение.

В зависимости от условий и особенностей получения исходных данных возможно отказаться от ряда показателей, используя наиболее важные по мнению исследователя. Предлагая ряд задач по специальным методам эконометрики, мы посчитали нецелесообразным приводить все выполненные по ним проверки и поместили лишь основные, убеждающие в эффективности метода. Желающим дается возможность выполнить полную проверку по данным табл. 4.

Таблица 4

Показатель	Физический смысл	Расчетная формула		Назначение показателя
		линейная связь	нелинейная связь	
1. Парный коэффициент корреляции	Теснота связи (отклонений от линии регрессии) признаков x и y	$r = \frac{\overline{yx} - \bar{y}\bar{x}}{\sqrt{x^2 - \bar{x}^2} \sqrt{y^2 - \bar{y}^2}}$	Иногда применяется	Установление функциональной зависимости между одной переменной и условным математическим ожиданием другой
2. Индекс корреляции	Теснота связи рассматриваемых признаков	Иногда применяется	$R = \sqrt{1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}}$	Установление надежности уравнения регрессии. $R = r$ означает правильность выбора типа нелинейного уравнения
3. Частный коэффициент корреляции	Теснота связи результата и одного из совокупности факторов модели при постоянстве остальных факторов	$r_{yx_i(x_1x_2 \dots x_{i-1}x_{i+1} \dots x_p)} = \sqrt{1 - \frac{1 - R_{yx_1x_2 \dots x_i \dots x_p}^2}{1 - R_{yx_1x_2 \dots x_{i-1}x_{i+1} \dots x_p}^2}}$		Ранжирование факторов множественной линейной регрессии и целесообразность включения того или иного фактора в модель
4. Индекс множественной корреляции	Теснота связи набора факторов с результатом	Применяется при стандартизованных коэффициентах регрессии	$R = \sqrt{1 - \frac{\sum (y - \hat{y}_{x_1x_2 \dots x_p})^2}{\sum (y - \bar{y})^2}}$	Практическая значимость уравнения множественной регрессии
5. Линейный (совокупный) коэффициент множественной корреляции	Теснота связи набора факторов с результатом для линейной регрессии	$R_{yx_1x_2 \dots x_p}^2 = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}}; \Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_p} \\ r_{yx_1} & 1 & r_{x_1x_2} & \dots & r_{x_1x_p} \\ r_{yx_2} & r_{x_2x_1} & 1 & \dots & r_{x_2x_p} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_p} & r_{x_px_1} & r_{x_px_2} & \dots & 1 \end{vmatrix};$ $\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & \dots & r_{x_1x_p} \\ r_{x_2x_1} & 1 & \dots & r_{x_2x_p} \\ \dots & \dots & \dots & \dots \\ r_{x_px_1} & r_{x_px_2} & \dots & 1 \end{vmatrix}$		Для расчета не требуется уравнение регрессии, достаточно парных коэффициентов корреляции. Всегда $R_{\text{мн. кор.}} \geq r_{\text{парной кор.}}$

6. Скорректированный индекс множественной корреляции	Содержит поправку на число степеней свободы	$\bar{R} = \sqrt{1 - \frac{\sum (y - \hat{y})^2 / (n - m - 1)}{\sum (y - \bar{y})^2 / (n - 1)}}$		Не допустить преувеличения тесноты связи при числе параметров m (коэффициентов) при x_j , близком к объему наблюдений n
7. Коэффициент детерминации	Доля дисперсии результативного признака, объясняемая регрессией, в его общей дисперсии	r_{xy}^2	–	Оценка адекватности регрессионной модели, мера ее качества и прогностической силы
8. Индекс детерминации	То же	–	R^2	То же. Если $(R^2 - r^2) \leq 0,1$, то оправдана линейная форма связи и не надо усложнять модель нелинейной формой
9. Индекс детерминации (квази- R^2)	»	–	квази $R^2 =$ $= 1 - \frac{\sum (y - \text{анти lg(lg } \hat{y}))^2}{\sum (y - \bar{y})^2}$	Оценка функций, содержащих логарифмические преобразования
10. Стандартная ошибка параметра a	Значимость свободного члена в модели	$m_a = \sqrt{\frac{\sum (y - \hat{y})^2 \sum x^2}{n(n-2) \sum (x - \bar{x})^2}}$	–	Определение целесообразности включения в модель свободного члена
11. t -критерий для параметра a	То же	$t_a = \frac{a}{m_a}$	–	
12. Стандартная ошибка параметра b	Значимость коэффициента регрессии b	$m_b = \sqrt{\frac{\sum (y - \hat{y})^2}{(n-2) \sum (x - \bar{x})^2}}$	–	Определение необходимости учета в модели факторного признака и расчет доверительных интервалов коэффициента b
13. t -критерий для параметра b	То же	$t_b = \frac{b}{m_b}$	–	

Продолжение табл. 4

Показатель	Физический смысл	Расчетная формула		Назначение показателя
		линейная связь	нелинейная связь	
14. Стандартная ошибка коэффициента корреляции	Значимость линейного коэффициента корреляции	$m_r = \sqrt{\frac{1-r^2}{n-2}}$	–	Определение величины и причин ошибки коэффициента корреляции при большом числе наблюдений и если r не близко $k \pm 1$
15. t -критерий для коэффициента корреляции	То же	$t_r = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$	–	
16. Вспомогательный коэффициент корреляции	»	$z = \frac{1}{2} \ln \frac{1+r}{1-r}$	–	Оценка существенности r при его близости $k \pm 1$
17. Стандартная ошибка величины z	»	$m_z = \frac{1}{\sqrt{n-3}}$	–	–
18. t -критерий для z	»	$t_z = \frac{z}{m_z}$	–	–
19. Ошибка разности между индексом и коэффициентом корреляции	Установление различия между индексом и коэффициентом корреляции для нелинейной и линейной связей	$m_{ R-r } = 2\sqrt{\frac{(R^2-r^2) - (R^2-r^2)^2(2-(R^2+r^2))}{n}}$		Установление возможности замены нелинейной формы связи на линейную при $(R^2-r^2) > 0,1$
20. t -критерий для $ R-r $	То же	$t = \frac{R^2-r^2}{m_{ R-r }}$		

21. Средняя относительная ошибка аппроксимации	Приближение теоретических значений, предсказанных моделью, к эмпирическим данным	$\bar{A} = \frac{1}{n} \sum \left \frac{y - \bar{y}}{y} \right 100 \%$		Определение точности модели. Величина $\bar{A} = 5...7 \%$ свидетельствует о хорошем подборе модели к исходным данным
22. F -критерий	Значимость (степень) влияния факторов на резуль- тативный признак	$F = \frac{r^2(n-2)}{1-r^2}$ $F = t^2$	$F = \frac{R^2}{1-R^2} \frac{n-m-1}{m}$	Оценка существенности уравне- ния регрессии в целом
23. Частный F -кри- терий	Значимость влияния на результат отдельных факторов	$F_i = \frac{R_{yx_1x_2...x_p}^2 - R_{yx_1x_2...x_{i-1}...x_p}^2}{1 - R_{yx_1x_2...x_p}^2} (n - m - 1)$		Оценка влияния фактора x_i по- сле включения его в модель на результат
24. Коэффициент эластичности	Процент изменения ре- зультата при изменении фактора на 1 %. Показа- тель степени для сте- пенной модели	$\Theta = f'(x) \frac{x}{y}$		Определение изменения резуль- тирующего признака в процен- тах
25. Частный коэф- фициент эластич- ности	Устанавливает изоли- рованное влияние фак- тора на результат в про- центах	$\Theta_{y_{x_i}} = b_i \frac{x_i}{\hat{y}_{x_i(x_1x_2...x_{i-1}x_{i+1}...x_p)}}$	–	Определение коэффициентов эластичности для частных ли- нейных уравнений регрессии с целью их выделения из совокуп- ности факторов
26. Средний показа- тель эластичности	Средний процент изме- нения результата под влиянием среднего зна- чения фактора	$\bar{\Theta}_{y_{x_i}} = b_i \frac{\bar{x}_i}{y_{x_i}}$	–	Определение средних измене- ний процессов под влиянием силы воздействия факторов в процентах

Показатель	Физический смысл	Расчетная формула		Назначение показателя
		линейная связь	нелинейная связь	
27. Бета-коэффициент	Влияние фактора на результат в долях их среднеквадратического отклонения	$\beta_i = b_i \sqrt{\frac{\sum (x - \bar{x})^2}{\sum (y - \bar{y})^2}}$		Определяется изменение среднего значения результирующего признака при изменении фактора на величину его среднеквадратического отклонения (в долях этой величины)
28. Дельта-коэффициент	Доля влияния фактора в суммарном влиянии всех факторов модели	$\Delta_i = \frac{r\beta_i}{R^2}$		Оценивается степень влияния фактора на результат в сравнении с суммарным влиянием всех факторов модели
29. Коэффициент устойчивости связи	Нарушение связи факторов и результата под влиянием отклонений фактических и теоретических результатов	$k = 1 - \frac{\sum y - \hat{y} }{\sum y}$		Оценивается теснота связи наблюдаемых и расчетных значений результирующего признака

Относительно приведенных в табл. 4 формул коэффициента и индекса корреляции и большинства показателей их производных дадим некоторые практические замечания. В формулах присутствуют прямо или косвенно возводимые в квадрат разности вида $(x - \bar{x})$ и $(y - \bar{y})$, которые затем суммируются. Точки (x, y) , расположенные вблизи среднего значения (\bar{x}, \bar{y}) – при разностях менее единицы измерения, в случае возведения в квадрат дают уменьшение результата по сравнению с величиной разности (например, для разности 0,9 имеем $0,9^2 = 0,81 < 0,9$). Если же разность больше единицы измерения, то результат увеличивается при возведении в квадрат ($2^2 = 4 > 2$). Чем удаленнее точка наблюдения от средней, тем большим будет увеличение. Следовательно, имеем неравнозначность результатов вычислений для разноудаленных точек как по знаку воздействия (уменьшение – увеличение), так и по удельному весу влияния на результат (более удаленные точки имеют бóльший вес). Другой случай: сколько бы ни было опытных точек на линиях $x = \text{const}$ или $y = \text{const}$, они дают нулевые разности с соответствующими координатами средней точки и не принимаются в расчет, а например, единственная точка в опыте с координатой, отличной от средней, будет учтена, что не совсем логично. Поэтому оценку корреляции целесообразнее было бы давать без учета средних величин.

Что могло бы служить в качестве единого коэффициента корреляции для любых моделей? По нашему мнению, отражать корреляцию может площадь, занимаемая точками наблюдения в пределах теоретической (аппроксимирующей) линии модели. Приближенным показателем здесь может быть отношение ширины разброса опытных точек относительно теоретической линии δ к ее длине L . Тогда коэффициент корреляции определится как $r = 1 - \delta/L$. Величина δ представляется как сумма положительных и отрицательных средних отклонений результирующего признака y модели от теоретической линии

$$\delta = \left[\sum_1^{n_1} (y_i - \hat{y}_i)/n_1 + \sum_1^{n_2} (\hat{y}_j - y_j)/n_2 \right] \cos \alpha,$$

где n_1 – число опытов, в которых опытные значения y_i больше им соответствующих теоретических \hat{y}_i ; n_2 – число опытов, в которых опытные значения y_j меньше им соответствующих теоретических \hat{y}_j ; α – угол наклона теоретической линии к оси координат для факторного признака x (см. рис. 2).

Длина теоретической линии зависит от угла ее наклона к оси x :

$$L = (x_{\max} - x_{\min}) / \cos \alpha.$$

Если угол не учитывать, а брать в расчет только проекцию линии на ось x , то с увеличением α отношение величины разброса опытных точек к длине проекции растет безгранично, что является абсурдным.

Можно считать, что корреляционная связь между x и y существует, если опытные точки укладываются в прямоугольную фигуру, вытянутую в направлении линии аппроксимации, т. е. при $L > \delta$. Практически приемлемо $\delta/L = 0,5$.

2. ПОСТРОЕНИЕ ЛИНЕЙНЫХ УРАВНЕНИЙ ДЛЯ СЛУЧАЕВ С ЧИСЛОМ НАБЛЮДЕНИЙ НЕ БОЛЕЕ ДВУХ

Пусть имеются два наблюдения A и B с соответствующими значениями факторных и результирующих признаков (x_1, y_1) и (x_2, y_2) . Тогда уравнение прямой, проходящей через эти точки наблюдений

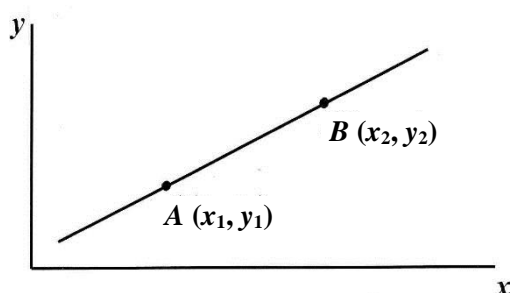


Рис. 3

(рис. 3), определится из известного в аналитической геометрии соотношения

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} \quad [4].$$

После подстановки в данное соотношение известных координат x_1, x_2, y_1, y_2 и арифметических преобразований получаем искомого уравнение $y = f(x)$ в виде $y = a + bx$,

где $a = \frac{x_2 y_1 - x_1 y_2}{x_2 - x_1}$; $b = \frac{y_2 - y_1}{x_2 - x_1}$. Например, для координат $x_1 = 5, y_1 = 10, x_2 = 20, y_2 = 30$ имеем $a = 3,33; b = 1,33$ и уравнение вида $y = 3,33 + 1,33x$.

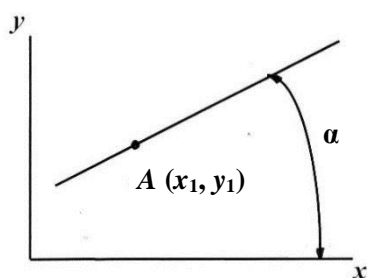


Рис. 4

Возможен случай, когда известны данные лишь для одного наблюдения и дано направление прямой с помощью угла α или углового коэффициента $b = \beta = \operatorname{tg} \alpha$ (рис. 4). Например, бета-коэффициент публикуется для акций, обращающихся на фондовых биржах. При этом в качестве факторного признака выступает доходность рыночного портфеля, а результирующего признака – избыточная доходность акций отдельной компании,

входящих в портфель. Для этого случая известно уравнение $y - y_1 = b(x - x_1)$. Если дано для некоторого периода функционирования биржи $x_1 = 400$, $y_1 = 700$ и $\beta = b = 0,25$, то в соответствии с приведенным уравнением получаем конкретную линейную зависимость $y = 600 + 0,25x$.

Рассмотрим еще один аналогичный вариант построения линейной зависимости, когда даны лишь координата $y_1 = a$ и угловой коэффициент (например, для временного ряда в его начале, где $x_1 = 0$). Тогда, исходя из предыдущей формулы, получаем $y = a + bx$.

Может иметь место частный случай задания координат прямой точками $(x_1 = 0, y_1)$ и $(x_2, y_2 = 0)$, расположенными на осях координат.

Этот вариант дает $y = y_1 - \frac{y_1}{x_2}x$. Данный случай называется уравнением прямой в отрезках.

3. МЕТОД ЭКСТРЕМАЛЬНЫХ ОПОРНЫХ ТОЧЕК

Построенные обычными (классическими) способами уравнения (линейные и нелинейные) не имеют в своем составе ограниченных какими-то пределами переменных, и если такие ограничения необходимы практически, то их обычно обуславливают просто текстом о границах применения формул. Это не совсем удобно с точки зрения пользователей. Поэтому целесообразно было бы включать граничные, т. е. экстремальные, значения непосредственно в вычислительные формулы.

Рассмотрим наиболее распространенные случаи прямолинейной, параболической и гиперболической зависимостей, которые охватывают большинство статистически наблюдаемых закономерностей. Сначала получим зависимости для однофакторных процессов, а затем перейдем к многофакторным.

3.1. Прямолинейная зависимость

Дадим решение в общем виде, выбрав произвольную точку наблюдения в качестве базовой, а затем запишем частные случаи формул, где базовыми являются экстремальные точки. Обозначим базовую

точку буквой М (рис. 5). Уравнение прямой, проходящей через точку М с координатами x_M, y_M [4]: $y - y_M = b_0(x - x_M)$ или $y = y_M + b_0(x - x_M)$, где b_0 – тангенс угла α наклона прямой к оси x .

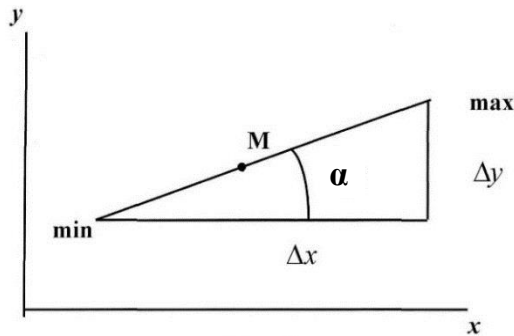


Рис. 5

Между b и $b_0 = \text{tg } \alpha$ очевидна зависимость $b_0 = b y_M/x_M$, которую можно использовать для определения угла α . Для многофакторной линейной зависимости вида $y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n$ достаточно приближенной формулы $\text{tg } \alpha = (a_1 + a_2 + \dots + a_n)/n$.

Очевидно, что $b_0 = \frac{\sum_{i=1}^n |y_i - y_M|}{\sum_{i=1}^n |x_i - x_M|}$, где x_i, y_i – координаты точек наблюдения от первой до n . Преобразуем исходное уравнение $y = y_M \left[1 + \frac{1}{y_M} b_0(x - x_M) \right]$. Обозначим $g_x = (x - x_M)/y_M$ – отношение для прямолинейной функции; $\Delta_{yi} = y_i - y_M, \Delta_{xi} = x_i - x_M$ – разности координат. При этом $b_0 = \frac{\sum_{i=1}^n |\Delta_{yi}|}{\sum_{i=1}^n |\Delta_{xi}|}$. Таким образом, получаем расчетную формулу в общем виде

$$\hat{y} = y_M(1 \pm b_0 g_x), \quad (3)$$

где $g_x = (x - x_M)/y_M$; $b_0 = \frac{\sum |\Delta_y|}{\sum |\Delta_x|}$; $\Delta_y = y - y_M$; $\Delta_x = x - x_M$, знак минус берется для обратно пропорциональной зависимости x и y . В качестве точки М можно принимать любую из наблюдаемых опытных точек (максимальную, минимальную) или какую-либо теоретическую точку (например, среднеарифметическую). Чтобы максимизировать единство обозначений в формуле, избавимся от обозначения Δ с помощью следующего преобразования составляющей $b_0 g_x$:

$$\frac{1}{y_M} \frac{\sum |y_i - y_M|}{\sum |x_i - x_M|} (x_i - x_M) = \frac{\sum |(y_i - y_M)/y_M| \cdot x_i - x_M}{\sum |(x_i - x_M)/x_M| \cdot x_M} = \frac{\sum |g_y|}{\sum |g_x|} g_x = b g_x,$$

где $b = \frac{\sum |g_y|}{\sum |g_x|}$; $g_y = (y_i - y_M)/y_M$; $g_x = (x_i - x_M)/x_M$.

Формула (3) принимает вид

$$\hat{y} = y_M(1 \pm bg_x)^*, \quad (4)$$

где параметр b уже не является тангенсом угла наклона прямой, а есть некоторая условная величина. Теперь из общей формулы можем записать четыре частных ее варианта, где в качестве базовой принимаются максимальная или минимальная по оси y из наблюдаемых точек:

а) прямая линейная зависимость с границей в минимальной точке наблюдений $\hat{y} = y_{\min}(1 + bg_x)$; $g_x = \frac{x_i}{x_{\min}} - 1$; $g_y = \frac{y_i}{y_{\min}} - 1$;

б) прямая линейная зависимость с границей в максимальной точке наблюдения $\hat{y} = y_{\max}(1 - bg_x)$; $g_x = 1 - \frac{x_i}{x_{\max}}$; $g_y = 1 - \frac{y_i}{y_{\max}}$;

в) обратная линейная зависимость с границей в минимальной точке наблюдений $\hat{y} = y_{\min}(1 + bg_x)$; $g_x = 1 - \frac{x_i}{x_{\max}}$; $g_y = \frac{y_i}{y_{\min}} - 1$;

г) обратная линейная зависимость с границей в максимальной точке наблюдений $\hat{y} = y_{\max}(1 - bg_x)$; $g_x = \frac{x_i}{x_{\min}} - 1$; $g_y = 1 - \frac{y_i}{y_{\max}}$.

Во всех случаях $b = \sum g_y / \sum g_x$. Заметим, что эти частные случаи совпадают с формулами, приведенными в работе [3], проверенными в расчетах и давших удовлетворительные результаты.

Приведем сопоставительные расчеты рассмотренным специальным методом и классическим методом наименьших квадратов. Результаты расчетов сведены в табл. 5, 6 для первого частного случая прямой линейной зависимости с границей в минимальной точке наблюдений, т. е. рассматривается гипотеза о линейной зависимости y от x . Экономическое содержание переменных x и y опустим.

* Здесь и далее будем обозначать \hat{y} , y , \bar{y} – теоретические, опытные и средние значения соответственно, $\varepsilon = \hat{y} - y$ – отклонение фактического значения от теоретического.

Таблица 5

Номер опыта	x	y	g_x	g_y	b	bg_x	\hat{y}	ε
1	5,49	0,78	0,209	2,714	14,974	3,133	0,87	0,09
2	4,68	0,38	0,031	0,089		0,461	0,31	-0,07
3	4,67	0,21	0,029	0		0,428	0,30	0,09
4	4,54	0,51	0	1,429		0	0,21	-0,30
5	5,56	0,95	0,225	3,524		3,365	0,92	-0,03
6	6,02	1,05	0,326	4,000		4,882	1,23	0,18
7	5,72	0,83	0,260	2,952		3,892	1,03	0,20
8	5,43	0,98	0,196	3,667		2,935	0,83	-0,15
<i>Итого</i>	-	5,70	1,275	19,095	-	-	5,70	0,00

Таблица 6

Номер опыта	x	y	yx	x^2	y^2	σ_x^2	b	$b\bar{x}$	a	\hat{y}	ε
1	5,49	0,78	4,282	30,140	0,608	0,267	0,487	2,563	-1,851	0,823	0,04
2	4,68	0,38	1,778	21,902	0,144					0,428	0,05
3	4,67	0,21	0,981	21,809	0,044					0,423	0,21
4	4,54	0,51	2,315	20,612	0,260					0,360	-0,15
5	5,56	0,95	5,282	30,914	0,902					0,857	-0,09
6	6,02	1,05	6,321	36,240	1,102					1,080	0,03
7	5,72	0,83	4,748	32,718	0,689					0,935	0,10
8	5,43	0,98	5,321	29,485	0,960					0,793	-0,19
<i>Итого</i>	42,11	5,70	31,028	223,82	4,709	-	-	-	-	5,70	0,00
Среднее значение	5,264	0,712	3,878	27,977	0,589	-	-	-	-	-	-

Используемые соотношения в методе наименьших квадратов:

$$y = a + bx, \text{ где } b = (\overline{yx} - \bar{y} \bar{x}) / \sigma_x^2; \sigma_x^2 = \overline{x^2} - \bar{x}^2; a = \bar{y} - b\bar{x}.$$

Итоговые уравнения:

$$- \text{специальный метод } \hat{y} = 0,21(1 + 14,974g_x); g_x = \frac{x}{4,54} - 1;$$

$$- \text{метод наименьших квадратов } \hat{y} = -1,851 + 0,487x.$$

Во всех случаях получено равенство $y = \hat{y}$. Коэффициент корреляции для заданных условий $r = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\sigma_x \sigma_y} = \frac{0,13}{0,5167 \cdot 0,2720} = 0,925$ указывает на тесную связь x и y .

Индекс корреляции определяется по формуле $R = \sqrt{1 - \frac{\sum \varepsilon^2}{\sum (y - \bar{y})^2}}$:

– специальный метод дает $R = \sqrt{1 - \frac{0,2069}{0,6641}} = 0,83$;

– метод наименьших квадратов – $R = \sqrt{1 - \frac{0,1258}{0,6641}} = 0,90$, что го-

ворит о лучшей аппроксимации классическим методом.

Средняя ошибка аппроксимации выражается формулой $\bar{A} = \frac{1}{n} \sum \left| \frac{\varepsilon}{y} \right| 100 \%$, откуда имеем для специального метода

$\bar{A} = 1,9135/8 \cdot 100 \% = 23,92 \%$, для метода наименьших квадратов

$\bar{A} = 1,6205/8 \cdot 100 \% = 20,26 \%$, что в обоих случаях превышает допустимый предел 10 %.

В то же время видим, что специальный метод требует значительно меньших вычислений, дает достаточно надежный результат, а зависимость получена с ограничением по y_{\min} . Заметим, что при расчете индекса корреляции в качестве базовой точки в формулу включено среднее арифметическое значение \bar{y} . Если для специального метода принять реальный факт использования не средней, а экстремальной точки (в рассмотренном примере минимальной), то можно скорректировать расчет, заменив соответственно величины y на g_y .

Известная формула выглядит так: $R = \sqrt{1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}}$. По аналогии

для специального метода имеем $R = \sqrt{1 - \frac{\sum (g_y - g_{\hat{y}})^2}{\sum (g_y - g_{y_{\min}})^2}}$. Очевидно, что

$g_{y_{\min}} = \frac{y_{\min}}{y_{\min}} - 1 = 0$, тогда получаем окончательно $R = \sqrt{1 - \frac{\sum (g_y - g_{\hat{y}})^2}{\sum g_y^2}}$,

где $g_{\hat{y}} = \frac{\hat{y}_i}{y_{\min}} - 1$. Расчет по этой формуле дает результат

$$R = \sqrt{1 - \frac{4,69}{60,64}} = 0,96, \text{ т. е. выше, чем для классического метода.}$$

Использование полученной формулы индекса корреляции требует лишь определения $g_{\hat{y}}$ и позволяет применять уже полученные при выводе модели параметры g_x и g_y . Так же находится коэффициент корреляции

$$r = \frac{\frac{1}{n} \sum g_x g_y - g_{x_{\min}} g_{y_{\min}}}{\sqrt{\frac{1}{n} \sum g_x^2 - g_{x_{\min}}^2} \sqrt{\frac{1}{n} \sum g_y^2 - g_{y_{\min}}^2}}. \text{ При } g_{x_{\min}} = 0, g_{y_{\min}} = 0 \text{ } r = \frac{\sum g_x g_y}{\sqrt{\sum g_x^2} \sqrt{\sum g_y^2}},$$

откуда для рассматриваемого примера имеем $r = \frac{4,18}{\sqrt{0,31 \cdot 60,64}} = 0,96$.

Равенство коэффициента и индекса корреляции подтверждает правильность выбранной гипотезы о линейном характере связи опытных данных.

Приведем еще один показатель оценки, полученной специальным методом модели – коэффициент устойчивости связи [3] $k = 1 - \frac{\sum |g_y - b g_x|}{\sum g_y}$. Устойчивость означает, что малое изменение фактора x во всем исследованном и предсказываемом моделью диапазоне не может вызвать резких больших изменений решения y .

Таблица 7

k	Устойчивость связи
До 0,5	Очень низкая
0,5 – 0,6	Низкая
0,6 – 0,7	Заметная
0,7 – 0,8	Средняя
0,8 – 0,9	Высокая
0,9 и более	Очень высокая

Качественная оценка устойчивости по величине коэффициента приведена в табл. 7. Для условий нашей задачи вычисления дали результат $k = 1 - 5,34/19,1 = 0,72$, означающий средний уровень устойчивости связи.

Анализ модели можно продолжать и по другим критериям, но и выполненного уже достаточно для ее положительной оценки. В последующих задачах ограничимся в основном оценками коэффициента и индекса корреляции.

3.2. Параболическая зависимость

Для параболы в качестве базовой экстремальной точки следует выбирать ее вершину с координатами x_0 и y_{\max} или x_0 и y_{\min} – для обратной параболы (рис. 6). Факторный признак в параболических зависимостях возводится во вторую степень, поэтому при том же виде основной формулы изменение касается лишь составляющей g_x , т. е. имеем:

$$\text{а) } \hat{y} = y_{\max} (1 - b g_x); \quad g_x = \frac{(x_i - x_0)^2}{x_0^2};$$

$$g_y = 1 - \frac{y_i}{y_{\max}} \text{ – парабола;}$$

$$\text{б) } \hat{y} = y_{\min} (1 + b g_x); \quad g_x = \frac{(x_i - x_0)^2}{x_0^2}; \quad g_y = \frac{y_i}{y_{\min}} - 1 \text{ – обратная парабола.}$$

бола.

Параметр b в обоих случаях $b = \sum g_y / \sum g_x$. Рассмотрим численный пример для специального метода построения математической модели прямой параболы, используя для расчета соотношения "а". Данные вычислений сведены в табл. 8, 9.

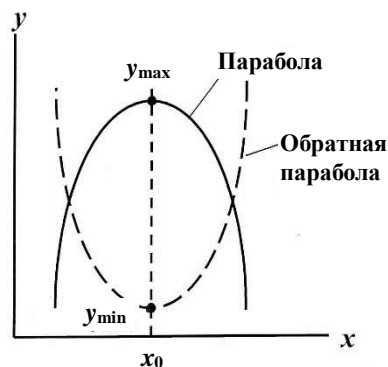


Рис. 6

Таблица 8

Номер опыта	x	y	g_x	g_y	b	$b g_x$	\hat{y}	$g_{\hat{y}}$
1	20	40	0,309	0,714	2,434	0,752	34,72	0,752
2	25	70	0,198	0,500		0,482	72,52	0,482
3	30	100	0,111	0,286		0,270	102,20	0,270
4	35	125	0,049	0,107		0,119	123,34	0,119
5	40	135	0,012	0,036		0,029	135,94	0,033
6	45	140	0	0		0	140	0
7	50	137	0,012	0,021		0,029	135,94	0,033
8	55	120	0,049	0,143		0,119	123,34	0,119
9	60	95	0,111	0,321		0,270	102,20	0,270
10	65	75	0,198	0,464		0,482	72,52	0,482
11	70	40	0,309	0,714		0,752	34,72	0,752
<i>Итого</i>	–	1077	1,358	3,306	–	–	1077,44	–

Таблица 9

Номер опыта	$g_x g_y$	g_x^2	g_y^2	$(g_y - g_{\hat{y}})^2$
1	0,2206	0,0955	0,5098	0,0014
2	0,0990	0,0392	0,2500	0,0003
3	0,0317	0,0123	0,0818	0,0003
4	0,0052	0,0024	0,0114	0,0001
5	0,0004	0,0001	0,0013	0
6	0	0	0	0
7	0,0003	0,0001	0,0004	0,0001
8	0,0070	0,0024	0,0204	0,0006
9	0,0356	0,0123	0,1030	0,0026
10	0,0919	0,0392	0,2153	0,0003
11	0,2206	0,0955	0,5098	0,0014
<i>Итого</i>	0,7123	0,2990	1,7032	0,0074

Итоговое уравнение $\hat{y} = 140(1 - 2,434g_x)$; $g_x = \frac{(x_i - 45)^2}{2025}$, расчетное $g_{\hat{y}} = 1 - y_i/140$. Коэффициент корреляции $r = \frac{0,7123}{\sqrt{0,299 \cdot 1,7032}} = 0,998$ указывает на тесную связь x и y .

$$\text{Индекс корреляции } R = \sqrt{1 - \frac{\sum (g_y - g_{\hat{y}})^2}{\sum g_y^2}} = \sqrt{1 - \frac{0,0074}{1,7032}} = 0,998.$$

Коэффициент и индекс корреляции совпали, что подтверждает параболическую связь x и y .

Рассмотрим аналогичный пример для обратной параболы, используя соотношения ”б”. Основные расчеты даны в табл. 10.

Таблица 10

Номер опыта	x	y	g_x	g_y	b	bg_x	\hat{y}	$g_{\hat{y}}$
1	80	1520	0,1185	0,6000		0,7000	1615	0,7000
2	88	1320	0,0777	0,3895		0,4590	1386	0,4589
3	95	1160	0,0490	0,2211		0,2894	1225	0,2895
4	108	1040	0,0132	0,0947		0,0780	1024	0,0779
5	122	950	0	0	5,9068	0	950	0
6	139	1060	0,0194	0,1158		0,1146	1059	0,1147
7	147	1200	0,0420	0,2632		0,2480	1186	0,2484
8	156	1370	0,0777	0,4421		0,4590	1386	0,4589
9	152	1500	0,0605	0,5789		0,3574	1290	0,3579
<i>Итого</i>	–	11120	0,4580	2,7053	–	–	11121	–

В результате получена расчетная формула $\hat{y} = 950(1 + 5,9068g_x)$; $g_x = (x_i - 122)^2 / 14884$. Расчетное $g_y = \frac{y_i}{950} - 1$. Результаты промежуточных вычислений для проверочных расчетов сведены в табл. 11.

Таблица 11

Номер опыта	$g_x g_y$	g_x^2	g_y^2	$(g_y - g_{\hat{y}})^2$	$ g_y - b g_x $
1	0,0711	0,0140	0,3600	0,0100	0,1000
2	0,0303	0,0060	0,1517	0,0048	0,0695
3	0,0108	0,0024	0,0489	0,0047	0,0683
4	0,0012	0,0002	0,0090	0,0003	0,0167
5	0	0	0	0	0
6	0,0022	0,0004	0,0134	0	0,0012
7	0,0111	0,0018	0,0693	0,0002	0,0152
8	0,0344	0,0060	0,1955	0,0003	0,0169
9	0,0350	0,0037	0,3351	0,0488	0,2215
<i>Итого</i>	0,1961	0,0345	1,1829	0,0691	0,5093

Коэффициент корреляции составляет $r = \frac{0,1961}{\sqrt{0,0345 \cdot 1,1829}} = 0,97$, т. е. указывает на тесную связь x и y . Индекс корреляции $R = \sqrt{1 - \frac{0,0691}{1,1829}} = 0,97$ совпадает с коэффициентом корреляции, что указывает на правильность выбора типа уравнения (параболы). Коэффициент устойчивости связи $k = 1 - \frac{0,5093}{2,7053} = 0,812$, означающий высокую устойчивую связь между x и y . На хороший результат, как и для предыдущих задач, указывает также равенство сумм опытных и теоретических значений результирующего признака.

3.3. Угловая зависимость

Возможен вариант расположения опытных данных, когда точки наблюдений образуют вершину, но парабола для их аппроксимации не подходит, так как обе ветви укладываются не на кривые, а на прямые линии (рис. 7). В таком случае симметричное расположение прямых дает основание совместить в одном уравнении два –

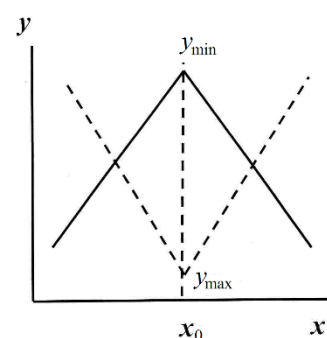


Рис. 7

для прямой и обратной зависимостей, но с базовой точкой в точке пересечения прямых. Исходные и основные расчетные данные приведены в табл. 12, 13.

Таблица 12

Номер опыта	x	y	g_x	g_y	b	bg_x	\hat{y}	$g_{\hat{y}}$
1	20	50	0,500	0,5		0,56	44	0,56
2	25	60	0,375	0,4		0,42	58	0,42
3	30	70	0,250	0,3		0,28	72	0,28
4	35	80	0,125	0,2		0,14	86	0,14
5	40	100	0	0	1,12	0	100	0
6	45	80	0,125	0,2		0,14	86	0,14
7	50	70	0,250	0,3		0,28	72	0,28
8	55	60	0,375	0,4		0,42	58	0,42
9	60	50	0,500	0,5		0,56	44	0,56
<i>Итого</i>	–	620	2,500	2,8	–	–	620	–

Таблица 13

Номер опыта	$g_x g_y$	g_x^2	g_y^2	$(g_y - g_{\hat{y}})^2$	$ g_y - bg_x $
1	0,250	0,2500	0,25	0,0036	0,06
2	0,148	0,1406	0,16	0,0004	0,02
3	0,075	0,0625	0,09	0,0004	0,02
4	0,025	0,0156	0,04	0,0036	0,06
5	0	0	0	0	0
6	0,025	0,0156	0,04	0,0036	0,06
7	0,075	0,0625	0,09	0,0004	0,02
8	0,148	0,1406	0,16	0,0004	0,02
9	0,250	0,2500	0,25	0,0036	0,06
<i>Итого</i>	0,996	0,9374	1,08	0,016	0,32

Результат расчета дал формулу вида ”а”

$$\hat{y} = y_{\max} (1 - bg_x); g_x = \left| 1 - \frac{x_i}{x_0} \right|; g_y = 1 - \frac{y_i}{y_{\max}}.$$

Формула в конкретных значениях для примера

$$\hat{y} = 100(1 - 1,12g_x); g_x = \left| 1 - \frac{x_i}{40} \right|; g_y = 1 - \frac{y_i}{100}.$$

Оценка качества модели показала результаты: $y = \hat{y}$;
 $r = \frac{0,996}{\sqrt{0,9374 \cdot 1,08}} = 0,99$; $R = \sqrt{1 - \frac{0,016}{1,08}} = 0,99$; $k = 1 - \frac{0,32}{2,8} = 0,89$, т. е. закономерность удовлетворительна по всем параметрам.

Аналогично представлена обратная угловая зависимость "б"

$$\hat{y} = y_{\min} (1 + b g_x); g_x = \left| 1 - \frac{x_i}{x_0} \right|; g_y = \frac{y_i}{y_{\min}} - 1.$$

Заметим, что высокие коэффициенты корреляции, рассчитанные через параметры g_x и g_y , могут вызвать возражения в связи с их несоответствием расчетным значениям, получаемым при проверке решений, полученных на основе метода наименьших квадратов. Но там пороговые значения r и R , которые считаются приемлемыми, установлены достаточно условно. Поэтому и для специального метода возможен условный допустимый уровень r и R , например, не ниже $0,9 \dots 0,92$.

3.4. Гиперболическая зависимость

Особенность уравнения равносторонней гиперболы, как известно из [2], состоит в том, что переменный фактор находится в знаменателе, в результате асимптоты кривых проходят параллельно осям координат. В качестве базовой экстремальной точки для прямой гиперболы берется y_{\min} , а при обратной гиперболе – y_{\max} (рис. 8). Отсюда при том же виде основной формулы и g_y имеем лишь специфическое выражение для параметра g_x . Поэтому можем записать:

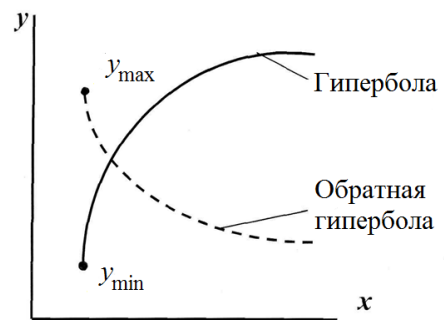


Рис. 8

а) $\hat{y} = y_{\min} (1 + b g_x); g_x = \frac{x_i - x_{\min}}{x_i} = 1 - \frac{x_{\min}}{x_i}; g_y = \frac{y_i}{y_{\min}} - 1$ – гипербола;

б) $\hat{y} = y_{\max} (1 - b g_x); g_x = \frac{x_i - x_{\min}}{x_i} = 1 - \frac{x_{\min}}{x_i}; g_y = 1 - \frac{y_i}{y_{\max}}$ – обратная гипербола.

ная гипербола.

Рассмотрим пример построения гиперболы, введя исходные и основные расчетные данные, приведенные в табл. 14, 15.

Таблица 14

Номер опыта	x	y	g_x	g_y	b	bg_x	\hat{y}	$g_{\hat{y}}$
1	5	42	0	0		0	42,00	0
2	10	170	0,500	3,048		3,721	198,28	4,721
3	15	240	0,667	4,714		4,963	250,45	5,963
4	20	275	0,750	5,548		5,581	276,40	6,581
5	25	280	0,800	5,667	7,441	5,953	292,03	6,953
6	30	310	0,833	6,381		6,198	302,32	7,198
7	35	325	0,857	6,738		6,377	309,83	7,377
8	40	320	0,875	6,619		6,511	315,46	7,511
9	45	335	0,889	6,976		6,615	319,83	7,615
10	50	330	0,900	6,857		6,697	323,27	7,697
<i>Итого</i>	–	2627	7,062	52,548	–	–	2630	–

Таблица 15

Номер опыта	$g_x g_y$	g_x^2	g_y^2	$(g_y - g_{\hat{y}})^2$	$ g_y - bg_x $
1	0	0	0	0	0
2	1,524	0,250	9,290	2,799	0,673
3	3,144	0,445	22,222	1,560	0,249
4	4,161	0,563	30,780	1,067	0,033
5	4,534	0,640	32,115	1,654	0,286
6	5,315	0,694	40,717	0,667	0,183
7	5,774	0,734	45,401	0,408	0,361
8	5,792	0,766	43,811	0,796	0,108
9	6,202	0,790	48,665	0,408	0,361
10	6,171	0,810	47,018	0,706	0,160
<i>Итого</i>	42,617	5,692	320,019	10,065	2,414

По табличным данным можно записать искомую модель

$$\hat{y} = 42(1 + 7,441g_x); \quad g_x = 1 - \frac{5}{x_i}; \quad g_y = \frac{y_i}{42} - 1.$$

Данные оценки качества модели

$$r = \frac{42,617}{\sqrt{5,692 \cdot 320,019}} = 0,999; \quad R = \sqrt{1 - \frac{10,065}{320,019}} = 0,984; \quad k = 1 - \frac{2,414}{52,548} = 0,954$$

свидетельствуют об очень тесной устойчивой связи x и y и правильности выбора типа уравнения – гиперболы. Для сравнения дадим расчет с помощью метода наименьших квадратов (табл. 16).

Таблица 16

Номер опыта	x	y	$1/x$	$(1/x)^2$	$(1/x)y$	\hat{y}
1	5	42	0,2	0,04	8,4	28,35
2	10	170	0,1	0,01	17	194,09
3	15	240	0,0667	0,00445	16,008	249,28
4	20	275	0,05	0,0025	13,75	276,96
5	25	280	0,04	0,0016	11,2	293,53
6	30	310	0,0333	0,00111	10,323	304,64
7	35	325	0,0286	0,00082	9,295	312,43
8	40	320	0,025	0,00063	8	318,39
9	45	335	0,0222	0,00049	7,437	323,04
10	50	330	0,02	0,0004	6,6	326,69
<i>Итого</i>	–	2627	0,5858	0,062	108,013	2627,4

Вид гиперболической модели $\hat{y} = a_0 + a_1 \frac{1}{x}$,

$$\text{где } a_0 = \frac{\sum y \sum (1/x)^2 - \sum (1/xy) \sum (1/x)}{n \sum (1/x)^2 - (\sum 1/x)^2}; \quad a_1 = \frac{n \sum (1/xy) - \sum (1/x) \sum y}{n \sum (1/x)^2 - (\sum 1/x)^2}.$$

Расчеты дали $\hat{y} = 359,83 - 1657,41 \frac{1}{x}$. Вычисления по полученной формуле, как показано в таблице, и без корреляционного анализа подтвердили высокую точность приближения теоретических значений к наблюдаемым. На это указывает оценка средней ошибки аппроксимации $\bar{A} = 0,1 \sum |\varepsilon/y| 100 \% = 0,1 \cdot 0,668 \cdot 100 \% = 6,68 \%$. Специальный метод показал $\bar{A} = 0,1 \sum (\varepsilon/y) 100 \% = 0,1 \cdot 0,391 \cdot 100 \% = 3,91 \%$, т. е. несколько лучший результат.

Рассмотрим пример получения модели для обратной гиперболы. Чтобы несколько усложнить задачу, дадим для обеих наблюдаемых переменных неравномерный ход (в предыдущем примере приращение независимой переменной было постоянным, равным 5 единицам). Увеличим также разброс значений, в том числе введем отдельные «выбросы». Результаты расчетов приводим в табл. 17, 18.

Таблица 17

Номер опыта	x	y	g_x	g_y	b	bg_x	\hat{y}	$g_{\hat{y}}$	ε
1	8	550	0	0		0	550	0	0
2	14	440	0,429	0,200		0,227	425,15	0,227	-14,85
3	21	360	0,619	0,345		0,327	370,15	0,327	10,15
4	36	325	0,778	0,409	0,52864	0,411	323,95	0,411	-1,05
5	53	340	0,849	0,382		0,449	303,05	0,449	-36,95
6	77	270	0,896	0,509		0,474	289,30	0,474	19,3
7	95	260	0,916	0,527		0,484	283,80	0,484	23,8
<i>Итого</i>	–	2545	4,487	2,372	–	–	2545,4	–	0,4

Таблица 18

Номер опыта	$g_x g_y$	g_x^2	g_y^2	$(g_y - g_{\hat{y}})^2$	$ g_y - bg_x $	$ \varepsilon/y $
1	0	0	0	0	0	0
2	0,0858	0,1840	0,0400	0,00073	0,027	0,03370
3	0,2136	0,3832	0,1190	0,00032	0,018	0,02819
4	0,3182	0,6053	0,1673	0,000004	0,002	0,00323
5	0,3243	0,7208	0,1459	0,00449	0,067	0,10868
6	0,4561	0,8028	0,2591	0,00123	0,035	0,07148
7	0,4827	0,8391	0,2777	0,00185	0,043	0,09154
<i>Итого</i>	1,8807	3,5352	1,009	0,008624	0,192	0,33682

По полученным данным запишем искомую модель

$$\hat{y} = 550(1 - 0,52864g_x); \quad g_x = \frac{x_i - 8}{x_i}; \quad g_y = \frac{550 - y_i}{550}.$$

Проверочные результаты модели:

$$r = \frac{1,8807}{\sqrt{3,5352 \cdot 1,009}} = 0,996; \quad R = \sqrt{1 - \frac{0,008624}{1,009}} = 0,996.$$

Эти данные свидетельствуют о тесной связи x и y , а также о правильности выбора типа уравнения (обратной гиперболы) для опытных значений.

$$k = 1 - \frac{0,192}{2,372} = 0,919 \text{ свидетельствует об очень высокой устойчи-}$$

вости связи x и y . $\bar{A} = \frac{0,33682}{7} 100 \% = 4,8 \% < 10 \%$, т. е. средний раз-

брос значений в допустимых пределах. Об этом же свидетельствует малое значение $\varepsilon = 0,4$. Положительный факт – практически равенство сумм опытных и теоретических значений y .

В экономике нередко наблюдаются процессы, когда после быстрого развития возникает процесс стабилизации, а затем через некоторое время наступает равномерно ускоряющийся спад. Такое возможно, например, в закономерностях развития жизненного цикла (производства, товара и пр.), кризисных явлений, спроса и т. д. Для подобных процессов в качестве модели описания парабола не подходит из-за слишком пологой формы участка стабилизации точек наблюдения. Здесь подходящим вариантом аппроксимации могли бы быть две гиперболы для возрастающего и ниспадающего участков, которые принимаются симметрично расположенными (рис. 9).

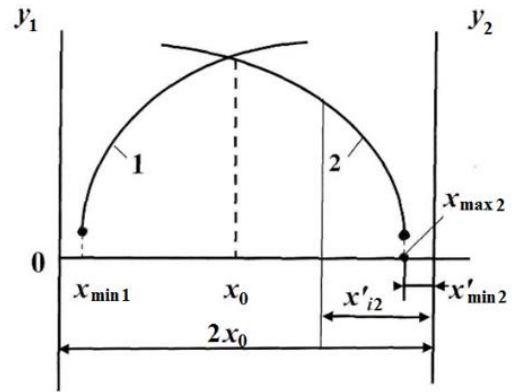


Рис. 9

Обозначим координату оси симметрии x_0 для гипербол 1 и 2. Соответственно в качестве асимптоты для второй гиперболы выступает ось y_2 на расстоянии $2x_0$ от оси y_1 , а отсчет точек для этой кривой будет осуществляться от оси y_2 (точки $x'_{\min 2}$, x'_{i2}). Уравнение второй гиперболы

$$\hat{y}_2 = y_{\min 2} (1 + b_2 g_{x2}), \text{ где } g_{x2} = \frac{x'_{i2} - x'_{\min 2}}{x'_{i2}}. \text{ Очевидно, что } x'_{i2} = 2x_0 - x_{i2},$$

$$x'_{\min 2} = 2x_0 - x_{\max 2}.$$

После подстановки получаем
$$g_{x2} = \frac{2x_0 - x_{i2} - 2x_0 + x_{\max 2}}{2x_0 - x_{i2}} = 1 - \frac{2x_0 - x_{\max 2}}{2x_0 - x_{i2}},$$

т. е. окончательно
$$g_{x2} = 1 - \frac{2x_0 - x_{\max 2}}{2x_0 - x_{i2}}.$$
 Значение для g_y сохраняется как

для прямой гиперболы
$$g_{y2} = \frac{y_i}{y_{\min}} - 1.$$

Таким образом, получили расчетное выражение для встречной гиперболы:

$$\hat{y}_2 = y_{\min 2} (1 + b_2 g_{x2}); \quad g_{x2} = 1 - \frac{2x_0 - x_{\max 2}}{2x_0 - x_{i2}}; \quad g_{y2} = \frac{y_i}{y_{\min}} - 1; \quad b_2 = \frac{\sum g_{y2}}{\sum g_{x2}}.$$

Здесь x_0 принимается ориентировочно, например, на основе графического анализа кривой по установлению точки пересечения гипербол. Остальные значения берутся из опытных данных или получаются расчетом.

Рассмотрим пример построения уравнения встречной гиперболы, симметричной по опытным данным для уже рассмотренной ранее прямой гиперболы (см. табл. 14). Результаты расчетов и исходные симметричные данные при $x_0 = 50$ сводим в табл. 19.

Таблица 19

Номер опыта	x_2	y_2	g_{x2}	g_{y2}	b_2	$b_2 g_{x2}$	\hat{y}_2
10	50	330	0,900	6,857	7,441	6,697	323,27
11	55	335	0,889	6,976		6,615	319,83
12	60	320	0,875	6,619		6,511	315,46
13	65	325	0,857	6,738		6,377	309,83
14	70	310	0,833	6,381		6,198	302,32
15	75	280	0,800	5,667		5,953	292,03
16	80	275	0,750	5,548		5,581	276,40
17	85	240	0,667	4,714		4,963	250,45
18	90	170	0,500	3,048		3,721	198,28
19	95	42	0	0		0	42
<i>Итого</i>	–	2627	7,062	52,548	–	–	2630

Результаты расчетов дали модель гиперболы $\hat{y}_2 = 42(1 + 7,441g_{x2})$;

$g_{x2} = 1 - \frac{5}{100 - x_{i2}}$; $g_{y2} = \frac{y_{i2}}{42} - 1$, по которой определены \hat{y}_2 . Сравнение

табл. 14 и 19 показывает их полную идентичность, начиная с y_2 , но в противоположном порядке, т. е. получили симметричную гиперболу.

Если встречная гипербола имеет несимметричные данные отно-

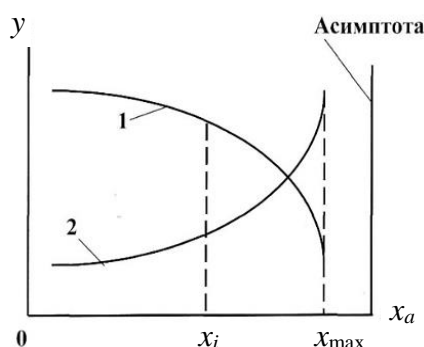


Рис. 10

сительно первой гиперболы, то ее уравнение можно построить независимо от первой. Для этого необходимо, построив гиперболу 1 по опытным данным графически, выбрать ее асимптоту с координатой $x_a > x_{\max}$ приближенно (рис. 10). Тогда выражение для g_x

можно брать в виде $g_x = 1 - \frac{x_a - x_{\max}}{x_a - x_i}$. Точка

пересечения двух гипербол определяется из

равенства $\hat{y}_1 = \hat{y}_2$, откуда выводится формула квадратного многочлена

$(y_{\min 2} + y_{\min 2} b_2 - y_{\min 1} b_1) x_0^2 + (y_{\min 1} x_a - y_{\min 2} x_{\max 2} b_2 + y_{\min 1} b_1 x_a + y_{\min 1} x_{\min 1} b_1) x_0 - y_{\min 1} x_{\min 1} b_1 x_a = 0$, из которой и находится искомая точка x_0 . Для симметричных гипербол достаточно равенства $g_{y1} = g_{y2}$,

позволяющего записать $x_0 = \frac{x_a x_{\min 1}}{x_{\min 1} - x_{\max 2} + x_a}$. Откуда для рассмотренных примеров имеем $x_0 = \frac{100 \cdot 5}{5 - 95 + 100} = 50$, т. е. принятое значение x_0 .

Приводим расчет обратной гиперболы, сведя исходные данные и промежуточные результаты в табл. 20, 21.

Таблица 20

Номер опыта	x	y	g_x	g_y	b	bg_x	\hat{y}	$g_{\hat{y}}$	ε
1	23	82	0	0	0,9886	0	82	0	0
2	27	58	0,148	0,293		0,1463	70	0,1463	12
3	42	44	0,452	0,463		0,4468	45,4	0,4468	1,4
4	49	36	0,531	0,561		0,5249	39	0,5249	3
5	63	33	0,635	0,598		0,6278	30,5	0,6278	-2,5
6	69	28	0,667	0,659		0,6594	28	0,6594	0
7	80	31	0,712	0,622		0,7039	24,3	0,7039	-6,7
8	85	30	0,729	0,634		0,7207	22,9	0,7207	-7,1
<i>Итого</i>	-	342	3,874	3,830	-	-	342,1	-	0,1

Таблица 21

Номер опыта	$g_x g_y$	g_x^2	g_y^2	$(g_y - g_{\hat{y}})^2$	$ g_y - bg_x $	$ \varepsilon/y $
1	0	0	0	0	0	0
2	0,0434	0,0219	0,0858	0,0216	0,1467	0,207
3	0,2093	0,2043	0,2144	0,0003	0,0162	0,032
4	0,2979	0,2820	0,3147	0,0013	0,0361	0,083
5	0,3797	0,4032	0,3576	0,0009	0,0298	0,076
6	0,4395	0,4449	0,4343	0	0,0004	0
7	0,4429	0,5069	0,3869	0,0067	0,0819	0,216
8	0,4622	0,5314	0,4020	0,0075	0,0867	0,237
<i>Итого</i>	2,275	2,395	2,196	0,0383	0,3978	0,851

Таким образом, получили модель $\hat{y} = 82(1 - 0,9886g_x)$; $g_x = 1 - \frac{23}{x_i}$;

$g_y = 1 - \frac{y_i}{82}$; $b = \frac{3,83}{3,874} = 0,9886$. Оценочные показатели модели:

$$r = \frac{2,275}{\sqrt{2,395 \cdot 2,196}} = 0,992; \quad R = \sqrt{1 - \frac{0,0383}{2,196}} = 0,991; \quad k = 1 - \frac{0,3978}{3,83} = 0,896;$$

$$\bar{A} = \frac{0,851}{8} 100 \% = 10,6 > 10 \%$$

Данные означают удачную аппроксимацию опытных результатов и устойчивую связь переменных, хотя средняя ошибка аппроксимации несколько больше допустимой. Встречная обратная гипербола 2, несимметричная относительно первой обратной гиперболы, также при построении требует выбора асимптоты $x_a > x_{\max}$ (см. рис. 10). В результате уравнение встречной обратной гиперболы запишется так:

$$\hat{y}_2 = y_{\max 2}(1 - b_2 g_{x2}); \quad g_{x2} = 1 - \frac{x_a - x_{\max 2}}{x_a - x_{i2}}; \quad g_{y2} = 1 - \frac{y_{i2}}{y_{\max 2}}; \quad b_2 = \frac{\sum g_{y2}}{\sum g_{x2}}.$$

Если такая гипербола проходит симметрично к первой, то по аналогии с рассмотренным случаем для обычной встречной гиперболы получаем формулу $g_{x2} = 1 - \frac{2x_0 - x_{\max 2}}{2x_0 - x_{i2}}$, где x_0 – координата оси симметрии. Точку пересечения несимметричных друг другу обратных прямой и встречной гипербол также можно найти из условия $\hat{y}_1 = \hat{y}_2$, как это было рассмотрено для обычных гипербол.

Рассмотрим пример расчета встречной обратной гиперболы (табл. 22, 23) для принятого значения $x_a = 125$.

Таблица 22

Номер опыта	x_2	y_2	g_{x2}	g_{y2}	b_2	$b_2 g_{x2}$	\hat{y}_2	$g_{\hat{y}_2}$	ε_2
1	56	290	0,8116	0,6282		0,6310	287,82	0,6310	-2,18
2	61	280	0,7969	0,6410		0,6195	269,79	0,6195	16,79
3	68	390	0,7719	0,5000		0,5977	313,79	0,5977	-76,21
4	82	330	0,6977	0,5769	0,7774	0,5424	356,93	0,5424	26,93
5	90	400	0,6286	0,4872		0,4887	398,81	0,4887	-1,19
6	105	530	0,3500	0,3205		0,2721	567,76	0,2721	37,76
7	112	780	0	0		0	780,00	0	0
<i>Итого</i>	-	3000	4,0567	3,1538	-	-	3001,9	-	1,9

Получена модель $\hat{y}_2 = 780(1 - 0,7774 g_{x2}); \quad g_{x2} = 1 - \frac{13}{125 - x_{i2}};$
 $g_{y2} = 1 - \frac{y_{i2}}{780}; \quad b_2 = \frac{3,1538}{4,0567} = 0,7774$, по которой определены теоретические значения \hat{y}_2 .

Таблица 23

Номер опыта	$g_{x2}g_{y2}$	g_{x2}^2	g_{y2}^2	$(g_{y2} - g_{\hat{y}2})^2$	$ g_{y2} - b_2g_{x2} $	$ \varepsilon_2/y_2 $
1	0,5098	0,6587	0,3946	0,000008	0,0028	0,00752
2	0,5108	0,6350	0,4109	0,000462	0,0215	0,05996
3	0,3859	0,5958	0,2500	0,009545	0,0977	0,19541
4	0,4025	0,4868	0,3328	0,001190	0,0345	0,08161
5	0,3063	0,3951	0,2374	0,000002	0,0015	0,00297
6	0,1122	0,1225	0,1027	0,002343	0,0484	0,07125
7	0	0	0	0	0	0
<i>Итого</i>	2,2275	2,8939	1,7284	0,01355	0,2064	0,41872

Показатели качества модели: $r = \frac{2,2275}{\sqrt{2,8939 \cdot 1,7284}} = 0,996$;

$$R = \sqrt{1 - \frac{0,01355}{1,7284}} = 0,996; k = 1 - \frac{0,2064}{3,1538} = 0,935; \bar{A} = \frac{0,41872}{7} 100 \% = 5,98 \% < 10 \%$$

Полученные данные, в том числе практическое равенство $\sum y_2 = 3000$ и $\sum \hat{y}_2 = 3001,9$, свидетельствуют о высокой точности аппроксимации, несмотря на значительный разброс опытных значений от аппроксимирующей линии. В дополнение дадим выражения для расчета прямых и обратных гипербол с противоположными рассмотренному границами наблюдений:

а) гиперболическая зависимость с границей в максимальной точке наблюдений $\hat{y} = y_{\max} (1 - bg_x)$; $g_x = \frac{x_{\max}}{x_i} - 1$; $g_y = 1 - \frac{y_i}{y_{\max}}$;

б) обратная гиперболическая зависимость с границей в минимальной точке наблюдений $\hat{y} = y_{\min} (1 + bg_x)$; $g_x = \frac{x_{\max}}{x_i} - 1$; $g_y = \frac{y_i}{y_{\min}} - 1$;

в) встречная гиперболическая зависимость с границей в максимальной точке наблюдений $\hat{y}_2 = y_{\max 2} (1 - b_2 g_{x2})$; $g_{x2} = \frac{x_a - x_{\min 2}}{x_a - x_{i2}} - 1$;

$$g_{y2} = 1 - \frac{y_{i2}}{y_{\max 2}};$$

г) встречная обратная гиперболическая зависимость с границей в минимальной точке наблюдений $\hat{y}_2 = y_{\min 2}(1 + b_2 g_{x2})$; $g_{x2} = \frac{x_a - x_{\min 2}}{x_a - x_{i2}}$;

$$g_{y2} = \frac{y_{i2}}{y_{\min 2}} - 1.$$

Сходными с гиперболическими зависимостями могут быть и другие типы кривых или их отдельные участки. В частности, возможно использование для аппроксимации различных кривых n -го порядка, в том числе параболического и гиперболического типов. Выбор кривой обуславливается лишь расположением опытных точек на графике. Хотя возможно и экспертное решение на основе предыдущего опыта описания подобных процессов. Рассмотрим два примера такого описания для параболы и гиперболы третьей степени.

3.5. Степенная зависимость

Приведем обработку данных с помощью параболы третьей степени. Описание результатов наблюдений возможно следующими моделями:

а) парабола третьей степени $\hat{y} = y_{\max}(1 - b g_x)$; $g_x = \frac{(x_i - x_{\min})^3}{x_{\min}^3}$;

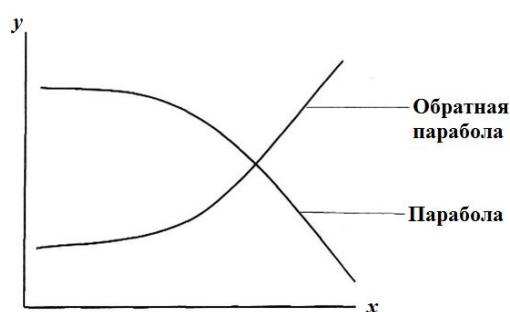


Рис. 11

$$g_y = 1 - \frac{y_i}{y_{\max}}; b = \frac{\sum g_y}{\sum g_x};$$

б) обратная парабола третьей степени $\hat{y} = y_{\min}(1 + b g_x)$; $g_x = \frac{(x_i - x_{\min})^3}{x_{\min}^3}$;

$$g_y = \frac{y_i}{y_{\max}} - 1; b = \frac{\sum g_y}{\sum g_x}.$$

Дадим численное решение для обратной параболы третьей степени (рис. 11), но не в полном ее симметричном виде, а лишь для одной ее правой ветви. Материалы по исходным данным и промежуточным расчетам приведены в табл. 24.

Таблица 24

Номер опыта	x	y	g_x	g_y	b	bg_x	\hat{y}
1	10	140	0	0		0	140
2	20	145	1	0,036		0,01094	141,5
3	25	142	3,4	0,014		0,03718	145,2
4	30	150	8	0,071		0,08752	152,3
5	40	175	27	0,250		0,29538	181,4
6	50	240	64	0,714	0,01094	0,700016	238,0
7	60	350	125	1,500		1,36750	331,4
8	70	470	216	2,357		2,36304	470,8
9	80	660	343	3,714		3,75242	665,3
10	85	790	421,9	4,643		4,61559	786,2
11	90	920	512	5,571		5,60128	924,2
12	100	1250	729	7,929		7,97526	1256,5
<i>Итого</i>	–	5432	2450,3	26,799	–	–	5432,8

Полученные результаты при сравнении значений y и \hat{y} как индивидуально, так и по суммам даже без корреляционного анализа свидетельствуют о высокой точности приближения. Модель при этом имеет вид $\hat{y} = 140(1 + 0,01094g_x)$; $g_x = \frac{(x_i - 10)^3}{1000}$; $g_y = \frac{y_i}{140} - 1$. Результаты расчетов для получения модели в виде обратной гиперболы третьей степени приведены в табл. 25.

Таблица 25

Номер опыта	x	y	g_x	g_y	b	bg_x	\hat{y}
1	22	340	0	0		0	340
2	28	250	0,51494	0,26471		0,32893	228,2
3	36	198	0,77178	0,41765		0,49300	172,4
4	40	185	0,83363	0,45588		0,53251	158,9
5	45	167	0,88315	0,50882		0,56414	148,2
6	53	142	0,92848	0,58235	0,63878	0,59309	138,3
7	62	140	0,95532	0,58824		0,61024	132,5
8	75	121	0,97476	0,64412		0,62266	128,3
9	90	79	0,98539	0,76765		0,62945	126,0
10	120	75	0,99384	0,77941		0,63485	124,2
<i>Итого</i>	–	1697	7,84129	5,00883	–	–	1697

При расчетах предполагалось построение модели обратной гиперболы вида

$$а) \hat{y} = y_{\max}(1 - bg_x); g_x = \frac{x_i^3 - x_{\min}^3}{x_i^3} = 1 - \frac{x_{\min}^3}{x_i^3}; g_y = 1 - \frac{y_i}{y_{\max}}. \text{ Вели-}$$

чина параметра b определялась традиционно $b = \frac{\sum g_y}{\sum g_x}$. В итоге полу-

$$\text{чена модель } \hat{y} = 340(1 - 0,63878g_x); g_x = \frac{x_i^3 - 10648}{x_i^3}; g_y = 1 - \frac{y_i}{340}.$$

Оценочные показатели модели приведены в табл. 26.

Таблица 26

Номер опыта	ε	$ \varepsilon/y $	g_y^2	$g_{\hat{y}}$	$(g_y - g_{\hat{y}})^2$
1	0	0	0	0	0
2	-21,8	0,09	0,070	0,329	0,0041
3	-25,6	0,13	0,174	0,493	0,0057
4	-26,1	0,14	0,211	0,533	0,0059
5	-18,8	0,11	0,259	0,564	0,0030
6	-3,7	0,03	0,339	0,593	0,0001
7	-7,5	0,05	0,346	0,610	0,0005
8	7,3	0,06	0,415	0,623	0,0004
9	27,0	0,34	0,589	0,629	0,0192
10	49,2	0,66	0,607	0,635	0,0208
<i>Итого</i>	-20	1,61	3,01	-	0,0597

$$\text{Индекс корреляции } R = \sqrt{1 - \frac{0,0597}{3,01}} = 0,99. \text{ Относительная ошибка}$$

аппроксимации $\bar{A} = \frac{1,61}{10} 100 \% = 16,1 > 10 \% \text{ слишком завышена. Тео-}$

ретические значения \hat{y} оказались достаточно близкими к опытным величинам y . Отклонения объяснимы тем, что теоретически кубическая обратная гипербола имеет сильно вогнутую форму у вершины, в которую не вписываются исходные данные для x и y (значения \hat{y} занижены в средней части и завышены внизу табличных данных) [4]. Однако суммы значений y и \hat{y} совпадают, что свидетельствует о возможной приемлемости расчетов.

Гипербола третьей степени представляется формулой

$$б) \hat{y} = y_{\min}(1 + bg_x); g_x = \frac{x_{\max}^3 - x_i^3}{x_i^3}; g_y = \frac{y_i}{y_{\min}} - 1.$$

3.6. Многофакторная зависимость

Результирующий фактор y экономической модели является наблюдаемой величиной, не всегда объяснимой полностью ее происхождением. Например, инфляция – вполне конкретная фиксируемая в различные моменты времени t оценка экономического процесса. Если ее рассматривать только во времени, то описание инфляционного процесса представляется однофакторной моделью вида $y = f(t)$. Можно ту же инфляцию изучать в зависимости от совокупности ряда факторов – внешних и внутренних (чрезмерных военных расходов, завышенных объемов инвестиций, необоснованного роста оплаты труда и т. д.), но независимо от числа этих объясняющих переменных, величина инфляции в полном ее временном наборе будет той же. То есть фактически вместо одного временного фактора будут в модель вводиться другие факторы x_j ($j = 1 \dots k$) с получением связи вида $y = f(x_1, x_2, \dots, x_k)$, в том числе в модели может остаться и время, например, в качестве фактора x_1 .

Следовательно, многофакторная модель независимо от вводимого в нее количества объясняющих факторов не меняет существа наблюдаемого результирующего признака. Фактически с введением в модель дополнительных факторов происходит перестройка параметров в правой части уравнения, но их суммарное воздействие на признак y остается одинаковым независимо от того, будет ли введен в модель один фактор или их станет несколько. Каждый фактор x_i во влиянии на y вносит свою долю, однако суммарное количество этих долей – величина постоянная. Отсюда следует весьма важный вывод о том, что существует равенство $g_x = g_{x1} + g_{x2} + \dots + g_{xk}$, которое позволяет дать общую формулу метода экстремальных опорных точек для любого числа факторов

$$\hat{y} = y_{\min(\max)}(1 \pm Bg_x); g_x = g_{x1} + g_{x2} + \dots + g_{xk}; B = \frac{\sum_1^n g_y}{\sum_1^n g_{x1} + \sum_1^n g_{x2} + \dots + \sum_1^n g_{xk}},$$

где параметры g_{xj} и g_y определяются из рассмотренных выше соотношений для различного вида зависимостей. Здесь B называется совокупным параметром многофакторной зависимости [3]. В частном случае при $j = 1$ имеем $B = b = \sum g_y / \sum g_x$, т. е. получаем параметр для однофакторной зависимости.

Многофакторная зависимость – это многомерная модель для многомерной системы координат. Чтобы понять ее суть, рассмотрим графические примеры двухфакторных моделей. Два фактора x_1 и x_2 вместе с результирующим признаком y образуют трехмерное пространство в декартовых координатах x_1x_2y .

На рис. 12, *а* показано образование трехмерной плоской параболы в случае соединения прямой на плоскости x_1Oy и параболы на плоскости x_2Oy . На рис. 12, *б* дан случай получения трехмерной кривой на пересечении поверхностей, сформированных образующими гиперболы на плоскости x_1Oy и параболы на плоскости x_2Oy . Если добавить еще один фактор x_3 , то будем иметь четырехмерное пространство, в котором трехмерные полученные кривые приобретут движение вдоль оси x_3 , например, отражающей время.

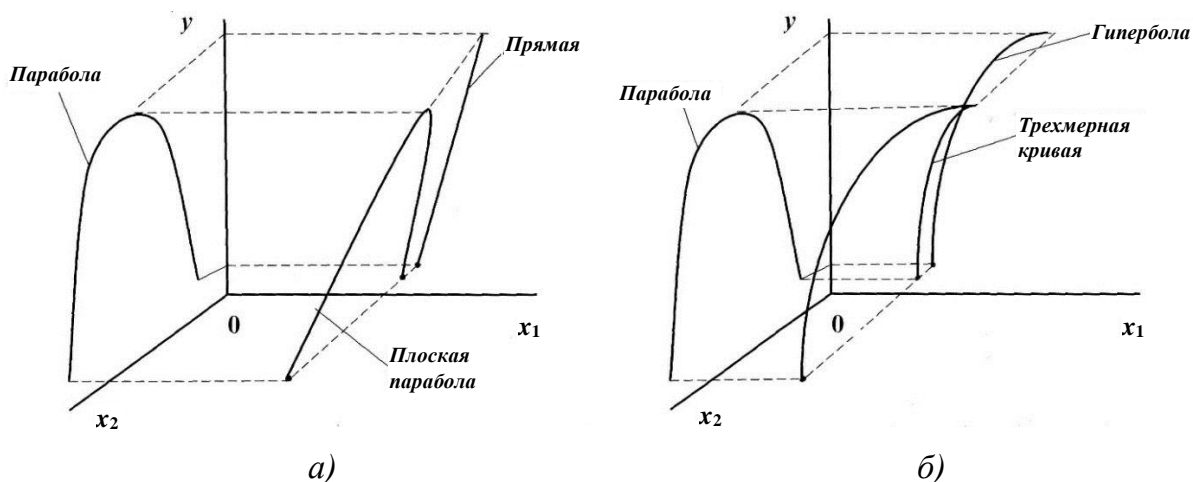


Рис. 12

Однако для описания экономических процессов не требуется такая интерпретация полученных моделей. Рассматривается зависимость y от множества факторов, которые могут быть независимыми друг от друга, и тогда эффект от факторов просто суммируется как в рассмотренных графических примерах и отражен в общей модели для метода экстремальных точек. В то же время графический пример показывает, что объединяться в общей модели могут частные зависимости как линейного, так и нелинейного характера (см. на рис. 12, *б* соединение гиперболы с параболой), т. е. на суммирование параметров g_x не накладывают ограничения виды зависимостей. Впрочем, для метода экстремальных опорных точек одно ограничение все же имеется, а именно:

нельзя в одной модели соединять факторы, ограниченные различными видами экстремумов. Имеется в виду, что факторы должны иметь один вид ограничений – или только максимумы, или только минимумы по признаку y .

Для более четкого различия получаемых моделей введем их условную классификацию:

- линейная многофакторная зависимость образована из факторов, имеющих с результирующим признаком линейную связь;
- нелинейная многофакторная зависимость включает факторы, связанные с y по криволинейной закономерности (пример на рис. 12, б);
- комбинационная многофакторная зависимость сочетает факторы, обуславливающие y по линейной и нелинейной формам связи (пример на рис. 12, а).

Рассмотрим пример построения комбинационной двухфакторной модели, в которой сочетаются линейная и гиперболическая зависимости. Исходные данные и результаты расчетов приведены в табл. 27, 28.

Анализ исходных данных показывает, что связь между x_1 и y близка к линейной, а между x_2 и y – к гиперболической, обе зависимости обратные с максимальными опорными точками. Вид комбинационной модели будет $\hat{y} = y_{\max} (1 - Bg_x)$; $g_x = g_{x1} + g_{x2}$; $g_{x1} = \frac{x_i - x_{\min}}{x_{\min}}$ –

для обратной линейной зависимости; $g_{x2} = \frac{x_i - x_{\min}}{x_i}$ – для обратной ги-

перболы; $g_y = \frac{y_{\max} - y_i}{y_{\max}}$; $B = \sum g_y / (\sum g_{x1} + \sum g_{x2})$.

Таблица 27

Номер опыта	y	x_1	x_2	g_{x1}	g_{x2}	g_y	$g_{x1} + g_{x2}$	$B (g_{x1} + g_{x2})$	\hat{y}
1	59	54	270	0	0	0	0	0	59
2	54	78	285	0,44444	0,05556	0,08475	0,50000	0,07935	54,3
3	49	95	375	0,75926	0,38889	0,16949	1,14815	0,18359	48,2
4	41	123	448	1,27778	0,65926	0,30508	1,93704	0,30973	40,7
5	35	128	555	1,37037	1,05556	0,40678	2,42593	0,38791	36,1
6	28	145	645	1,68519	1,38889	0,52542	3,07408	0,49155	30,0
7	22	165	790	2,05556	1,92593	0,62712	3,98149	0,63664	21,4
8	16	180	920	2,33333	2,40741	0,72881	4,74074	0,75804	14,3
<i>Итого</i>	304	968	4288	9,92593	7,88150	2,84745	17,80743	–	304

Таблица 28

Номер опыта	$g_{\hat{y}}$	g_y^2	$(g_y - g_{\hat{y}})^2$	$ g_y - Bg_x $	ε	$\left \frac{\varepsilon}{y} \right $
1	0	0	0	0	0	0
2	0,07966	0,00635	0,00003	0,00480	0,3	0,00556
3	0,18305	0,03351	0,00018	0,01410	-0,8	0,01633
4	0,31017	0,09621	0,00003	0,00465	-0,3	0,00732
5	0,38814	0,15065	0,00035	0,01887	1,1	0,03143
6	0,49153	0,24160	0,00115	0,03387	2	0,07143
7	0,63729	0,40614	0,00010	0,00952	-0,6	0,02727
8	0,75763	0,57400	0,00083	0,02882	-1,7	0,10625
<i>Итого</i>	-	1,50846	0,00267	0,11463	0	0,26559

Параметр $B = \frac{\sum g_y}{\sum g_{x1} + \sum g_{x2}} = \frac{2,84745}{17,80743} = 0,1599$. Показатели

качества модели: $R = \sqrt{1 - \frac{0,00267}{1,50846}} = 0,999$; $k = 1 - \frac{0,11463}{2,84745} = 0,96$;

$\bar{A} = \frac{0,26559}{8} 100 \% = 3,3 \% < 10 \%$. Все показатели очень высокие.

Комбинационная модель в численном выражении:

$$\hat{y} = 59(1 - 0,1599 g_{x1x2}); g_{x1x2} = g_{x1} + g_{x2};$$

$$g_{x1} = \frac{x - 54}{54}; g_{x2} = \frac{x - 270}{270}.$$

Применим для сравнения расчет методом наименьших квадратов. Для двухфакторной модели вида $\hat{y} = a_0 + a_1x_1 + a_2x_2$ система нормальных уравнений

$$\begin{cases} na_0 + (\sum x_1)a_1 + (\sum x_2)a_2 = \sum y, \\ (\sum x_1)a_0 + (\sum x_1^2)a_1 + (\sum x_1x_2)a_2 = \sum yx_1, \\ (\sum x_2)a_0 + (\sum x_1x_2)a_1 + (\sum x_2^2)a_2 = \sum yx_2. \end{cases}$$

Для данной системы требуются дополнительные данные, расчет которых приводим в табл. 29.

Таблица 29

Номер опыта	x_1^2	$x_1 x_2$	yx_1	x_2^2	yx_2	\hat{y}	$(y - \hat{y})^2$	$(y - \bar{y})^2$	$y - \hat{y}$
1	2916	14580	3186	72900	15930	59,4	0,16	441	-0,4
2	6084	22230	4212	81225	15390	53,8	0,04	256	0,2
3	9025	36625	4655	140625	18375	47,8	1,44	121	1,2
4	15129	55104	5043	200704	18368	39,8	1,44	9	1,2
5	16384	71040	4480	308025	19425	36,0	1	9	-1
6	21025	93525	4060	416025	18060	30,0	4	100	-2
7	27225	130350	3630	624100	17380	21,9	0,01	256	0,1
8	32400	165600	2880	846400	14720	15,3	0,49	484	0,7
<i>Итого</i>	130188	589054	32146	2690004	137648	304	0,58	1676	0

Система уравнений в численном выражении будет

$$\begin{cases} 8a_0 + 968a_1 + 4288a_2 = 304, \\ 968a_0 + 130188a_1 + 589054a_2 = 32146, \\ 4288a_0 + 589054a_1 + 2690004a_2 = 137648. \end{cases}$$

Приводим решение этой системы методом определителей

$$D = \begin{vmatrix} 8 & 968 & 4288 \\ 968 & 130188 & 589054 \\ 4288 & 589054 & 2690004 \end{vmatrix} = 8 \cdot 130188 \cdot 2690004 + 968 \cdot 589054 \cdot 4288 +$$

+ 4288 · 968 · 589054 – 4288 · 130188 · 4288 – 8 · 589054 · 589054 –
– 968 · 968 · 2690004 = 0,00148707 · 10¹². Аналогично решаем остальные определители.

$$D_{a_0} = \begin{vmatrix} 304 & 968 & 4288 \\ 32146 & 130188 & 589054 \\ 137648 & 589054 & 2690004 \end{vmatrix} = 0,11605965 \cdot 10^{12}.$$

$$D_{a_1} = \begin{vmatrix} 8 & 304 & 4288 \\ 968 & 32146 & 589054 \\ 4288 & 137648 & 2690004 \end{vmatrix} = -0,0003238143 \cdot 10^{12}.$$

$$D_{a_2} = \begin{vmatrix} 8 & 968 & 304 \\ 968 & 130188 & 32146 \\ 4288 & 589054 & 137648 \end{vmatrix} = -0,0000379976 \cdot 10^{12}.$$

Вычисляем свободный член и коэффициенты регрессионного уравнения по формулам $a_0 = D_{a_0}/D$; $a_1 = D_{a_1}/D$; $a_2 = D_{a_2}/D$.

$$\text{Имеем } a_0 = \frac{0,11605965 \cdot 10^{12}}{0,00148707 \cdot 10^{12}} = 78,046; \quad a_1 = -\frac{0,0003238143 \cdot 10^{12}}{0,00148707 \cdot 10^{12}} = -0,21775; \quad a_2 = -\frac{0,0000379976 \cdot 10^{12}}{0,00148707 \cdot 10^{12}} = -0,02555.$$

Записываем искомое уравнение $\hat{y} = 78,046 - 0,21775x_1 - 0,02555x_2$, с помощью которого рассчитываем теоретические значения \hat{y} , помещенные в табл. 29. Индекс корреляции при $\bar{y} = 38$ $R = \sqrt{1 - \frac{0,58}{1676}} = 0,999$.

Средняя ошибка аппроксимации $\bar{A} = \frac{1}{8} \sum \left| \frac{y - \hat{y}}{y} \right| 100 \% = 2,6 \% < 10 \%$.

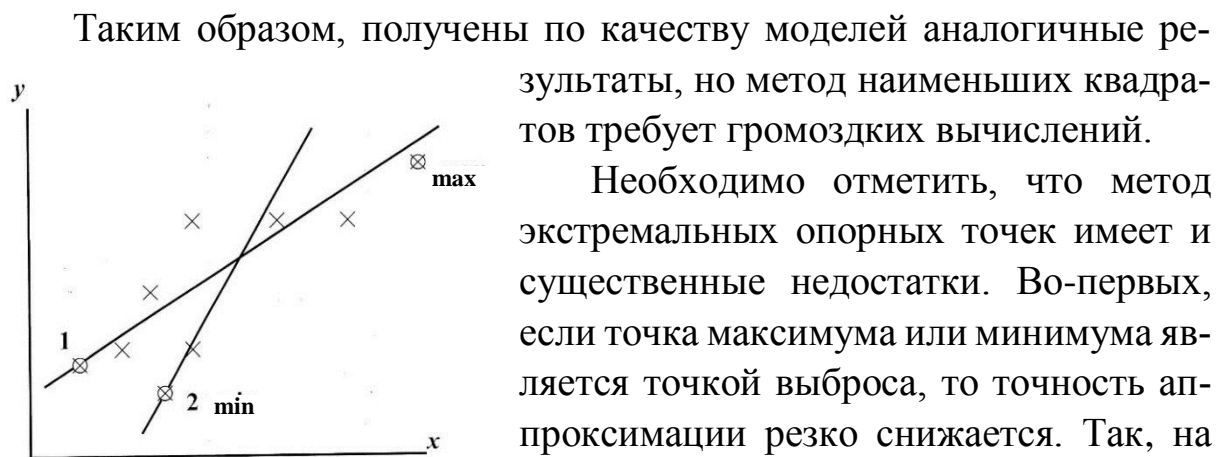


Рис. 13

Необходимо отметить, что метод экстремальных опорных точек имеет и существенные недостатки. Во-первых, если точка максимума или минимума является точкой выброса, то точность аппроксимации резко снижается. Так, на рис. 13 показаны два положения минимальной точки, которые приводят к изменению направления линейной зависимости. Здесь точка 2 является «выбросом». Во-вторых, для параболы обычно трудно установить вершину, так как ее теоретическое положение, как правило, не совпадает с опытной точкой максимума и возможно совпадение нескольких точек по координате $y_{\max(\min)}$. В результате расположение подобранной кривой становится многозначным. Экстремальная точка может быть также точкой «выброса». В-третьих, при расчете параметра b преимущество по «весу» при суммировании разностей координат $x_i, x_{\min(\max)}, y_i, y_{\min(\max)}$ имеют наиболее удаленные точки от исходной опорной точки.

Выход из подобных затруднений, конечно, дает метод наименьших квадратов, но присущие ему недостатки требуют поиска другого приемлемого выхода. Наилучшим вариантом в таком случае будет принятие в

качестве опорной некоторой условной среднеарифметической точки для всей совокупности наблюдений. Тогда вырисовывается еще один метод построения эконометрических моделей, который назовем методом центра масс.

4. МЕТОД ЦЕНТРА МАСС

Введение в качестве опорной точки центра масс всех точек наблюдения вдвое уменьшает погрешность от неравноправности точек при использовании предыдущего метода, так как центр масс находится примерно посередине диапазона координат опытных точек. Координаты этой точки равновесия $x_c = \sum_{i=1}^n x_i/n$, $y_c = \sum_{i=1}^n y_i/n$, где x_i, y_i – координаты i -й точки исходных статистических данных из всей их совокупности от 1 до n .

Надо иметь в виду, что центр масс имеет смысл определять для линеаризованной зависимости координат точек наблюдения, в противном случае он может оказаться далеко в стороне от линии расположения исходных данных, соответственно внося большие погрешности в расчеты при построении модели процесса. Поэтому обязательное условие метода – выполнение операции преобразования исходной зависимости расположения точек наблюдения в линейную форму и проверка качества осуществления линеаризации хотя бы графическим способом. Рассмотрим сначала процесс формирования линейной модели с использованием опорной точки – центра масс.

4.1. Прямолинейная зависимость

Исходя из общей формулы (4), запишем расчетную формулу для построения линейной модели методом центра масс

$$y = y_c (1 \pm b g_x), \quad (5)$$

где $g_x = \frac{x - x_c}{x_c}$; $g_y = \frac{y - y_c}{y_c}$, $b = \frac{\sum |g_y|}{\sum |g_x|}$; знак «минус» берется для обратнo пропорциональной зависимости.

Рассчитав x_c, y_c и b , можно задавать любые значения x , определять g_x и прогнозировать им соответствующие величины y .

Следует обратить внимание на то, что в выражении для параметра b разность $x - x_c$ берется по модулю, а в выражении для $g_x - c$ с учетом

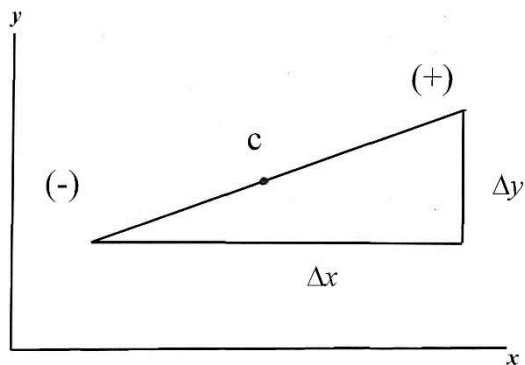


Рис. 14

получаемого знака при вычитании. Действительно, в положительной области линии на графике рис. 14 $x > x_c$, соответственно $\Delta_x = x - x_c > 0$, а в отрицательной области $x < x_c$, т. е. $\Delta_x < 0$. Поэтому при расчете y через g_x следует учитывать знаки Δ_x : прибавлять Δ_y , если Δ_x находится справа от точки c , и вычитать – если слева. При расчете величины b важны абсолютные значения

Δ_y и Δ_x , так как они обуславливают угол наклона прямой к оси x , что требует суммирования разностных отрезков, взятых по модулю.

Для сравнения возьмем пример по выводу прямой линейной зависимости из п. 3.1 и используем его исходные данные. Все расчетные операции отразим в табл. 30.

Таблица 30

Номер опыта	x	y	g_x	g_y	bg_x	\hat{y}	ε	$g_{\hat{y}}$	ε/y
1	5,49	0,78	0,0429	0,0940	0,1720	0,8357	0,0557	0,1720	0,0714
2	4,68	0,38	-0,1109	-0,4670	-0,4447	0,3959	0,0159	-0,4447	0,0419
3	4,67	0,21	-0,1128	-0,7055	-0,4523	0,3905	0,1805	-0,4523	0,8595
4	4,54	0,51	-0,1375	-0,2847	-0,5513	0,3199	-0,1901	-0,5513	0,3727
5	5,56	0,95	0,0562	0,3324	0,2253	0,8737	-0,0764	0,2253	0,0804
6	6,02	1,05	0,1436	0,4727	0,5758	1,1236	0,0735	0,5758	0,0701
7	5,72	0,83	0,0866	0,1641	0,3472	0,9606	0,1306	0,3472	0,1573
8	5,43	0,98	0,0315	0,3745	0,1263	0,8031	-0,1769	0,1263	0,1801
Итого	42,11	5,70	-	-	-	5,70	0,01	-	1,8338

Дополнительные расчетные данные:

$$x_c = \frac{42,11}{8} = 5,264; y_c = \frac{5,70}{8} = 0,713; \bar{y} = 0,7125;$$

$$b = \frac{\sum |g_y|}{\sum |g_x|} = \frac{2,895}{0,722} = 4,0095.$$

Получена модель

$$\hat{y} = 0,713(1 + 4,0095 g_x); g_x = \frac{x - 5,264}{5,264}.$$

Расчетные данные по проверке качества модели приведены в табл. 31.

Таблица 31

Номер опыта	$(g_y - g_{\hat{y}})^2$	g_y^2	$ g_y - bg_x $	ε^2	$(y - \bar{y})^2$
1	0,0061	0,0088	0,078	0,00310	0,00456
2	0,0005	0,2181	0,022	0,00025	0,11056
3	0,0641	0,4977	0,253	0,03258	0,25251
4	0,0711	0,0810	0,267	0,03614	0,04101
5	0,0115	0,1105	0,107	0,00592	0,05641
6	0,0106	0,2234	0,103	0,00541	0,11391
7	0,0335	0,0269	0,183	0,01705	0,01381
8	0,0610	0,1403	0,248	0,03131	0,07156
<i>Итого</i>	0,259	1,3068	1,262	0,13176	0,6643

Индекс корреляции рассчитан двумя методами:

$$R = \sqrt{1 - \frac{\sum \varepsilon^2}{\sum (y - \bar{y})^2}} = \sqrt{1 - \frac{0,13176}{0,6643}} = 0,8953 ;$$

$$R = \sqrt{1 - \frac{\sum (g_y - g_{\hat{y}})^2}{\sum g_y^2}} = \sqrt{1 - \frac{0,259}{1,3068}} = 0,8954 .$$

Получили практически совпадения, так как в обоих случаях сущность оказалась одна – отсчет ведется от среднего значения. Значения R превысили индекс корреляции, определенный для метода экстремальных опорных точек в п. 3.1 и оказались почти равными индексу для метода наименьших квадратов, вычисленному в том же примере.

Средняя ошибка аппроксимации

$$\bar{A} = \frac{1}{n} \sum \left| \frac{\varepsilon}{y} \right| 100 \% = \frac{1,8338}{8} 100 \% = 22,9 \% > 10 \% .$$

Ее величина оказалась ниже, чем при методе экстремальных опорных точек и несколько превышает уровень при методе наименьших квадратов, но в целом она также больше допустимого уровня на 10 %, что связано со значительным разбросом исходных данных. Отсюда и коэффициент устойчивости связи низок

$$k = 1 - \frac{\sum |g_y - bg_x|}{\sum |g_y|} = 1 - \frac{1,2616}{2,8949} = 0,5642 .$$

4.2. Параболическая зависимость

Преобразование параболы в прямую возможно с высокой степенью надежности с помощью специальных функций, в частности так называемых конформных отображений, которые обеспечивают растяжение (сжатие) областей пространства и поворот при сохранении углов между касательными к линиям исходной и отображенной в соответствующих им точках. Применяемый в литературе [3, 4] метод замен переменных может давать положительный результат в отдельных случаях, но он требует графической проверки качества линейризации и часто ее не обеспечивает. Так, замена в параболической зависимости $y^2 = ax^2 + bx + c$ в виде $y^2 = Y$ с получением новой формы записи $Y = ax^2 + bx + c$ не меняет существа исходной формулы.

Для линейных преобразований зависимостей параболического типа наиболее удачной считается дробно-линейная функция вида $Y = \frac{y - y_1}{x - x_1}$, где x_1 и y_1 – координаты выбранной произвольной точки на графике получаемой функции [8]. Рассмотрим пример построения параболы двумя методами по исходным данным табл. 32.

При решении методом центра масс сначала проводим линейризацию связи x и y с помощью дробно-линейной отображающей функции вида $Y = \frac{y - y_1}{x - x_1}$. В качестве произвольной точки (x_1, y_1) примем точку

максимума величины y по данным наблюдений ее координаты $x_1 = 44$, $y_1 = 130,9$. Рассчитываем Y , вводя в отображающую функцию эти координаты и данные опытов x и y . Вид линейной модели $\hat{Y} = Y_c (1 \pm bg_x)$,

где $b = \frac{\sum |\Delta Y|}{\sum |\Delta x|}$; $g_x = \frac{x - x_c}{Y_c}$; $\Delta x = x - x_c$; $\Delta Y = Y - Y_c$; $x_c = \frac{\sum x}{n}$; $Y_c = \frac{\sum y}{n}$.

Результаты расчетов сводим в табл. 32.

Таблица 32

Номер опыта	x	y	Y	$ \Delta x $	$ \Delta Y $	g_x	\hat{Y}	\hat{y}
1	8	34	2,692	30,454	2,042	-46,852	2,769	31,22
2	12	42,5	2,762	26,454	2,112	-40,698	2,491	51,19
3	16	68	2,246	22,454	1,596	-34,545	2,213	68,94
4	20	85	1,912	18,454	1,262	-28,391	1,934	84,48
5	24	110,5	1,020	14,454	0,370	-22,237	1,656	97,78
6	35	119	1,322	3,454	0,672	-5,314	0,890	122,89
7	44	130,9	0	5,546	0,650	8,532	0,264	130,90

Окончание табл. 32

Номер опыта	x	y	Y	$ \Delta X $	$ \Delta Y $	g_x	\hat{Y}	\hat{y}
8	52	127,5	-0,425	13,546	1,075	20,840	-0,293	128,56
9	59	119	-0,793	20,546	1,443	31,609	-0,780	119,20
10	72	93,5	-1,336	33,546	1,986	51,609	-1,685	83,72
11	81	47,6	-2,251	42,546	2,901	65,455	-2,311	45,39
<i>Итого</i>	423	977,5	7,149	231,454	16,109	—	—	964,3

Линейность связи x и Y подтверждается на графике рис. 15. Расчеты дали $x_c = 38,454$; $Y_c = 0,650$; $b = 0,0696$. Таким образом, с учетом обратно-пропорциональной линейной связи x и Y , запишем вид полученной мо-

дели: $\hat{Y} = 0,650(1 - 0,0696g_x)$,

$g_x = \frac{(x - 38,454)}{0,650}$. По отобра-

жающей функции получаем

$\hat{y} = \hat{Y}(x - x_1) + y_1$, откуда име-

ем выражение для обратного

преобразования $\hat{y} = \hat{Y}(x - 44) +$

$+ 130,9$, которое используем

для расчета \hat{y} . Определяем индекс корреляции: $\bar{y} = 88,86$; $\sum(y - \hat{y})^2 =$

$= 362,977$; $\sum(y - \bar{y})^2 = 12877,98$; $S_{\hat{y}x}^2 = 40,331$; $S_y^2 = 1170,73$. Тогда

$r_{\hat{y}} = 0,9826$. При решении методом наименьших квадратов данные рас-

четов сводим в табл. 33.

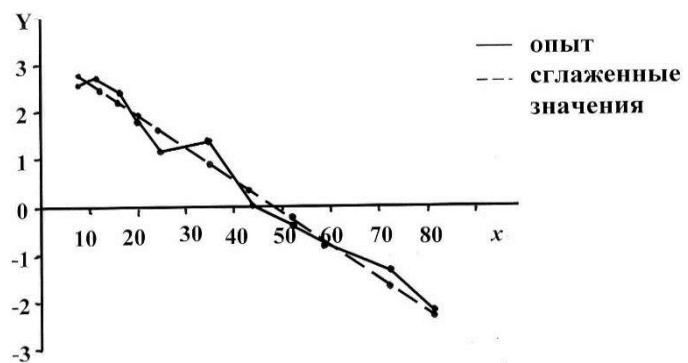


Рис. 15

Таблица 33

Номер опыта	$x = x_1$	y	$x^2 = x_2$	x_2^2	x_1x_2	yx_1	yx_2	\hat{y}
1	8	34	64	4069	512	272	2176	31,15
2	12	42,5	144	20736	1728	510	6120	51,12
3	16	68	256	65536	4096	1088	17408	68,90
4	20	85	400	160000	8000	1700	34000	84,49
5	24	110,5	576	331776	13824	2652	63648	97,88
6	35	119	1225	1500625	42875	4165	145775	123,39
7	44	130,9	1936	3748096	85184	5760	253422	131,39
8	52	127,5	2701	7311616	140608	6630	344760	130,85
9	59	119	3481	12117361	205379	7021	414239	121,41
10	72	93,5	5184	26873856	373248	6732	488470	87,34
11	81	47,6	6561	43046721	531441	3855	312304	50,17
<i>Итого</i>	423	977,5	22531	95180419	1406895	40385	2082322	978,1

Вид модели $y = a_0 + a_1x + a_2x^2$. Система нормальных уравнений для двухфакторной модели:

$$\begin{cases} na_0 + a_1 \sum x_1 + a_2 \sum x_2 = \sum y, \\ a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2 = \sum yx_1, \\ a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2 = \sum yx_2. \end{cases}$$

$$\begin{cases} 11a_0 + 423a_1 + 22531a_2 = 977,5, \\ 423a_0 + 22531a_1 + 1406895a_2 = 40385, \\ 22531a_0 + 140689a_1 + 95180419a_2 = 2082322, \end{cases}$$

откуда $a_0 = -15,39$; $a_1 = 6,366$; $a_2 = -0,0686$. Получили искомую модель $\hat{y} = -15,39 + 6,366x - 0,0686x^2$. Расчеты точности модели дали $\bar{y} = 88,86$; $\sum (y - \hat{y})^2 = 323,84$; $\sum (y - \bar{y})^2 = 11534$; $S_{\hat{y}x}^2 = 35,98$; $S_y^2 = 1049$, соответственно $r_{\hat{y}y} = 0,9827$.

Оба метода дали удовлетворительную точность и практически совпадение индексов корреляции. В качестве базовой точки для параболических зависимостей лучше всего подходит экстремальная точка (min или max в зависимости от типа параболы). Но чтобы не вносить путаницу в знаках для координат точек на смежных участках линий по обе стороны от базовой точки, предлагается выполнять еще одно преобразование вида $Y' \geq Y + |Y_{\max} - Y_{\min}|$.

Рассмотрим пример получения модели параболического вида, представив расчетные данные в табл. 34. Для сопоставления возьмем исходные данные задачи, решенной методом экстремальных опорных точек (см. табл. 8).

Таблица 34

Номер опыта	x	y	$y - y_{\max}$	$x - x_0$	Y	Y'	g_x	$g_{Y'}$	\hat{Y}'	\hat{Y}	\hat{y}
1	20	40	-100	-25	4	9	-0,5556	0,8058	9,243	4,243	33,9
2	25	70	-70	-20	3,5	8,5	-0,4444	0,7055	8,391	3,391	72,2
3	30	100	-40	-15	2,67	7,67	-0,3333	0,5389	7,539	2,539	101,9
4	35	125	-15	-10	1,5	6,5	-0,2222	0,3042	6,687	1,687	123,1
5	40	135	-5	-5	1	6	-0,1111	0,2039	5,836	0,836	135,8
6	45	140	0	0	0	5	0	0,0032	4,984	-0,016	140
7	50	137	-3	5	-0,6	4,4	0,1111	-0,1172	4,132	-0,868	135,7
8	55	120	-20	10	-2	3	0,2222	-0,3981	3,281	-1,719	122,8
9	60	95	-45	15	-3	2	0,3333	-0,5987	2,429	-2,571	101,4
10	65	75	-65	20	-3,25	1,75	0,4444	-0,6489	1,578	-3,422	71,6
11	70	40	-100	25	-4	1	0,5556	-0,7994	0,725	-4,275	33,1
<i>Итого</i>	495	1077	-	-	-0,18	54,82	3,3332	5,1265	-	-0,175	1071,5

Процесс расчета включил следующие соотношения:

$$Y = \frac{y - y_1}{x - x_1}, \text{ где } y_1 = y_{\max} = 140; x_1 = x_0 = 45 \text{ (соответствует } y_{\max});$$

$$g_x = \frac{x - x_c}{x_c}, \text{ где } x_c = \frac{\sum x}{n} = \frac{495}{11} = 45. \quad g_{Y'} = \frac{Y' - Y'_c}{Y'_c}; \quad Y'_c = \frac{\sum Y'}{n} =$$

$$= \frac{54,82}{11} = 4,984; \quad Y' \geq Y + |Y_{\max} - Y_{\min}| = Y + |4 - 0|, \text{ приняли } Y' = Y + 5.$$

$$b = \frac{\sum |g_{Y'}|}{\sum |g_x|} = \frac{5,1265}{3,3332} = 1,538. \text{ Тогда модель прямой линии имеет вид}$$

$$\hat{Y}' = Y'_c(1 - bg_x) = 4,984(1 - 1,538g_x); \quad g_x = \frac{x - 45}{45}.$$

Обратные преобразования: $\hat{Y} = \hat{Y}' - 5$; $\hat{y} = \hat{Y}(x - x_1) + y_1$ или $\hat{y} = \hat{Y}(x - 45) + 140$. Окончательный вид модели параболы: $\hat{y} = [4,984(1 - 1,538g_x) - 5](x - 45) + 140$; $g_x = \frac{x - 45}{45}$. Сопоставление теоретических значений результирующего признака и сумм его опытных и теоретических значений в табл. 8 и 34 показывает их незначительное расхождение. Большинство несовпадений объясняется погрешностями расчета вследствие округлений чисел. Поэтому можно говорить о высокой точности метода центра масс, не уступающего другим методам.

4.3. Гиперболическая зависимость

Гипербола строится на основе обратной связи вида $y = f(1/x)$. Эта функция легко линеаризируется путем замены $X = 1/x$. Применим метод центра масс для рассмотренного примера построения гиперболы методом экстремальных опорных точек (см. табл. 14). Исходные и расчетные данные сводим в табл. 35.

Таблица 35

Номер опыта	x	y	X	g_x	g_y	b	X_c	y_c	\hat{y}
1	5	42	0,2	2,413	-0,840	0,392	0,0586	262,7	14,2
2	10	170	0,1	0,706	-0,353				189,9
3	15	240	0,67	0,143	-0,086				247,9
4	20	275	0,05	-0,147	-0,047				277,8
5	25	280	0,04	-0,317	-0,066				295,4
6	30	310	0,033	-0,437	0,180				307,7
7	35	325	0,29	-0,505	0,237				314,7
8	40	320	0,025	-0,573	0,218				321,7
9	45	335	0,022	-0,625	0,275				327,0
10	50	330	0,02	-0,659	0,256				330,5
<i>Итого</i>	-	2627	0,586	6,525	2,558	-	-	-	2626,8

$$\text{Вид модели } \hat{y} = y_c(1 - bg_x); \quad g_x = \frac{X - X_c}{X_c}; \quad g_y = \frac{y - y_c}{y_c}; \quad b = \frac{\sum |g_y|}{\sum |g_x|}.$$

Получение расчетных данных: $X_c = \frac{0,586}{10} = 0,0586$; $y_c = \frac{2627}{10} = 262,7$;

$b = \frac{2,558}{6,525} = 0,392$. Таким образом, получена модель $\hat{y} = 262,7(1 -$

$- 0,392g_x)$; $g_x = \frac{x - 0,0586}{0,0586}$; где $X = 1/x$. Расчетные данные для оценки

качества модели сведены в табл. 36.

Таблица 36

Номер опыта	ε	$g_x g_y$	g_x^2	g_y^2	$g_{\hat{y}}$	$(g_y - g_{\hat{y}})^2$	bg_x	$\ g_y - bg_x \ $	$ \varepsilon/y $
1	-27,8	-2,027	5,823	0,706	-0,946	0,011	0,946	0,106	0,66
2	19,9	-0,249	0,498	1,125	-0,277	0,001	0,277	0,076	0,12
3	7,9	-0,012	0,020	0,007	-0,056	0,001	0,056	0,030	0,03
4	2,8	-0,007	0,022	0,002	0,057	0	-0,058	0,011	0,01
5	15,4	-0,021	0,100	0,004	0,124	0,003	-0,124	0,058	0,06
6	-2,3	-0,079	0,191	0,032	0,171	0	-0,171	0,009	0,01
7	-10,3	-0,120	0,255	0,056	0,198	0,002	-0,198	0,039	0,03
8	1,7	-0,125	0,328	0,048	0,225	0	-0,225	0,007	0,01
9	-8	-0,172	0,391	0,076	0,245	0,001	-0,245	0,030	0,02
10	0,5	-0,169	0,434	0,066	0,258	0	-0,258	0,002	0
<i>Итого</i>	-0,2	-2,981	8,062	1,122	-	0,019	-	0,368	0,95

Анализ таблиц указывает на неплохие результаты: разность остатков ε незначительна, практически совпадают суммы опытных y и теоретических \hat{y} значений результативного признака. Однако принятый метод линеаризации несовершенен, так как значительно выпадает по всем параметрам точка первого опыта, и при линеаризации точки после опыта № 6 располагаются слишком близко. Такая неравномерность должна отразиться на результатах оценок.

Коэффициент корреляции $r = \frac{-2,981}{\sqrt{8,062 \cdot 1,122}} = -0,991$ указывает на

очень тесную связь x и y . Индекс корреляции $R = \sqrt{1 - \frac{0,019}{1,122}} = 0,983$ бли-

зок к коэффициенту корреляции, что указывает на правильность выбора типа уравнения. Коэффициент устойчивости связи между x и y

$K = 1 - \frac{0,368}{2,558} = 0,856$ подтверждает высокий уровень устойчивости.

Средняя ошибка аппроксимации $\bar{A} = \frac{0,95}{10} 100 \% = 9,5 \% < 10 \% .$

Сравнение полученных результатов с данными для метода экстремальных опорных точек показывает некоторое преимущество последнего, что, по-видимому, связано с отмеченным низким качеством линеаризации в методе центра масс и требует поиска другой отображающей функции. Но в целом метод центра масс дал положительный результат, исключая погрешности «выброса».

Приведем построение обратной гиперболы по исходным данным и с промежуточными результатами расчетов, помещенными в табл. 37, 38. Для сравнения применим классический метод наименьших квадратов (табл. 37) и метод центра масс (табл. 38).

Таблица 37

Номер опыта	x	y	$1/x$	$(1/x)^2$	y/x	\hat{y}
1	5	78,0	0,2000	0,0400	15,6000	83,18
2	10	55,9	0,1000	0,0100	5,5900	47,16
3	15	36,4	0,0667	0,0044	2,4267	35,15
4	20	35,1	0,0500	0,0025	1,7550	29,15
5	25	26,0	0,0400	0,0016	1,0400	25,55
6	30	23,4	0,0333	0,0011	0,7800	23,15
7	35	19,5	0,0286	0,0008	0,5571	21,43
8	40	22,1	0,0250	0,0006	0,5525	20,15
9	45	15,6	0,0222	0,0005	0,3467	19,15
10	50	10,4	0,0200	0,0004	0,2080	18,34
<i>Итого</i>	–	322,4	0,5858	0,0620	28,8560	322,4

Вид модели $\hat{y} = a_0 + a_1 \frac{1}{x}$.

Система нормальных уравнений:

$$\begin{cases} na_0 + a_1 \sum \frac{1}{x} = \sum y, \\ a_0 \sum \frac{1}{x} + a_1 \sum \left(\frac{1}{x}\right)^2 = \sum y \frac{1}{x}. \end{cases}$$

$$\begin{cases} 10a_0 + 0,5858a_1 = 322,4, \\ 0,5858a_0 + 0,0620a_1 = 28,8560, \text{ откуда } a_0 = 11,141, a_1 = 360,181. \end{cases}$$

Модель имеет вид $\hat{y} = 11,141 + 360,181 \frac{1}{x}$. Для ее оценки расчеты

дали $\bar{y} = 32,24$; $\sum (y - \hat{y})^2 = 223,60$; $\sum (y - \bar{y})^2 = 3815,37$; $S_{yx}^2 = 27,95$; $S_y^2 = 381,537$, соответственно $r_{\hat{y}} = 0,804$.

Таблица 38

Номер опыта	Δ_x	g_x	Δ_y	\hat{y}
1	0,1410	0,0044	45,7600	88,95
2	0,0410	0,0013	23,6600	49,00
3	0,0080	0,0002	4,1600	34,82
4	-0,0090	-0,0003	2,8600	28,37
5	-0,0190	-0,0006	-6,2400	24,51
6	-0,0250	-0,0008	-8,8400	21,93
7	-0,0300	-0,0009	-12,7400	20,64
8	-0,0340	-0,0010	-10,1400	19,35
9	-0,0364	-0,0011	-16,6400	18,06
10	-0,0390	-0,0012	-21,8400	16,77
<i>Итого</i>	0,3824	—	152,88	322,4

Вид модели $y = y_c(1 + bg_x)$, где $b = \frac{\sum |\Delta_y|}{\sum |\Delta_x|}$; $g_x = \frac{X - X_c}{y_c}$;

$\Delta_x = X - X_c$; $\Delta_y = y - y_c$; $X_c = \frac{\sum X}{n}$; $y_c = \frac{\sum y}{n}$; $X = \frac{1}{x}$. Расчеты дали

$X_c = 0,05858$; $Y_c = 32,24$; $b = \frac{152,88}{0,3824} = 399,79$. Тогда модель имеет вид

$\hat{y} = 32,24(1 + 399,79g_x)$; $g_x = \frac{X - 0,05858}{32,24}$. Для нее $\bar{y} = 32,24$;

$\sum (y - \hat{y})^2 = 275,15$; $\sum (y - \hat{y})^2 = 3815,37$; $S_{\hat{y}x}^2 = 34,39$; $S_y^2 = 381,537$ и соответственно парный коэффициент корреляции $r_{x\hat{y}} = 0,954$. Видим, что точность модели, полученной методом центра масс, существенно выше, чем для выведенной методом наименьших квадратов ($0,954 > 0,804$).

Вместе с тем следует заметить, что метод центра масс дает хорошие результаты в тех случаях, когда гипербола по форме близка к прямой линии. Если ее прогиб велик, то центр масс смещается существенно от линии расположения опытных точек, смещая соответственно теоретическую кривую. Чтобы избежать погрешностей при построении модели, превышающих допустимый предел, например, средней ошибки аппроксимации, лучше использовать линейризацию исходных данных. Простейший путь при этом – логарифмирование значений x и y . В табл. 39 даны расчеты средней ошибки аппроксимации для рассмотренных вариантов построения модели.

Таблица 39

Номер опыта	Метод наименьших квадратов		Метод центра масс	
	ε	$ \varepsilon/y $	ε	$ \varepsilon/y $
1	5,18	0,066	10,95	0,140
2	-8,74	0,156	-6,90	0,123
3	-1,25	0,034	-1,58	0,043
4	-5,95	0,170	-6,73	0,192
5	-0,45	0,017	-1,49	0,057
6	-0,25	0,011	-1,47	0,063
7	1,93	0,099	1,14	0,058
8	-1,95	0,088	-2,75	0,124
9	3,55	0,228	2,46	0,158
10	7,94	0,763	6,37	0,612
<i>Итого</i>	0,01	1,632	0	1,570
\bar{A}	–	16,32 %	–	15,70 %

Оба случая дали значение \bar{A} выше допустимого уровня 10 %.

Степенные зависимости гиперболического типа легко линейризуются с помощью логарифмирования всех исходных данных. Рассмотрим пример получения зависимости методом центра масс с использованием такой линейризации для исходных данных x и y с обратной гиперболической связью, приведенных в табл. 25. Эти данные и

результаты преобразований при построении модели сводим в табл. 40. Ниже даны соответствующие зависимости.

Таблица 40

Номер опыта	x	y	$x' = \lg x$	$y' = \lg y$	$g_{x'}$	$g_{y'}$	$bg_{x'}$	\hat{y}'	\hat{y}
1	22	340	1,342	2,531	-0,2117	0,1573	-0,1408	2,495	312,6
2	28	250	1,447	2,398	-0,1500	0,0965	-0,0997	2,405	254,1
3	36	198	1,556	2,297	-0,0860	0,0503	-0,0572	2,312	205,1
4	40	185	1,602	2,267	-0,0590	0,0366	-0,0392	2,273	187,5
5	45	167	1,653	2,223	-0,0290	0,0165	-0,0193	2,229	169,4
6	53	142	1,724	2,152	0,0216	-0,0160	0,0144	2,156	143,2
7	62	140	1,792	2,146	0,0526	-0,0187	0,0350	2,110	128,8
8	75	121	1,875	2,083	0,1014	-0,0476	0,0674	2,040	109,6
9	90	79	1,954	1,898	0,1478	-0,1321	0,0983	1,972	93,8
10	120	75	2,079	1,875	0,2212	-0,1427	0,1471	1,865	73,3
<i>Итого</i>		1697	17,024	21,87	1,0743	0,7143	-	21,86	1677,4

Модель в логарифмической форме $\hat{y}' = y'_c(1 - bg_{x'})$; $g_{x'} = \frac{x' - x'_c}{x'_c}$;
 $g_{y'} = \frac{y' - y'_c}{y'_c}$. Значения ее составляющих $x'_c = \frac{\sum x'}{n} = \frac{17,024}{10} = 1,7024$;
 $y'_c = \frac{\sum y'}{n} = \frac{21,87}{10} = 2,187$; $g_{x'} = \frac{x' - 1,7024}{1,7024}$; $g_{y'} = \frac{y' - 2,187}{2,187}$; $b = \frac{\sum |g_{y'}|}{\sum |g_{x'}|} =$
 $= \frac{0,7143}{1,0743} = 0,6649$. Тогда модель в логарифмической форме в числах бу-
 дет $\hat{y}' = 2,187(1 - 0,6649g_{x'})$. Вычисленные значения \hat{y}' заносят в таб-
 лицу, а затем потенцируют с помощью выражения $\hat{y} = 10^{\hat{y}'}$.

Следует иметь в виду, что линеаризация с помощью логарифми-
 рования экстремальной зависимости дает ломаную линию, которая не
 может быть выражена предложенными методами одной формулой и
 аппроксимируется лишь по отдельным пересекающимся прямым, т. е.
 участкам ломаной линии. Проверочные расчеты для полученной мо-
 дели представлены в табл. 41.

Таблица 41

Номер опыта	ε	ε/y	$g_{y'}$	$g_{y'}^2$	$(g_{y'} - g_{\hat{y}'})^2$
1	-27,4	-0,080	0,141	0,02474	0,00027
2	4,1	0,016	0,100	0,00931	0,00001
3	7,1	0,036	0,057	0,00253	0,00004
4	2,5	0,014	0,039	0,00134	0,00001
5	2,4	0,014	0,019	0,00027	0,00001
6	1,2	0,008	-0,014	0,00026	0
7	-11,2	-0,080	-0,035	0,00035	0,00027
8	-11,4	-0,094	-0,067	0,00227	0,00039
9	14,8	0,187	-0,098	0,01745	0,00114
10	-1,7	-0,023	-0,147	0,02036	0,00002
<i>Итого</i>	-19	0,158	-	0,07888	0,00216

Индекс корреляции $R = \sqrt{1 - \frac{0,00216}{0,078888}} = 0,986$. Средняя ошибка аппроксимации $\bar{A} = \frac{0,158}{10} 100 \% = 1,58 \% < 10 \%$. Эти данные свидетельствуют о высоком качестве подбора модели, не уступающем практически другим методам. Возможна линеаризация исходных данных, укладываемых на кривую, близкую к гиперболе, с помощью логарифмирования только значений результирующего признака (так называемое спрямление кривых в полулогарифмических координатах). Данный прием, в частности, используется при построении показательных моделей [2]. Рассмотрим подобную линеаризацию для условий предыдущей задачи (табл. 42).

Таблица 42

Номер опыта	x	y	$y' = \lg y$	g_x	$g_{y'}$	bg_x	\hat{y}'	\hat{y}
1	22	340	2,531	-0,6147	0,1573	-0,1057	2,418	261,8
2	28	250	2,398	-0,5096	0,0965	-0,0876	2,379	239,3
3	36	198	2,297	-0,3695	0,0503	-0,0635	2,326	211,8
4	40	185	2,267	-0,2995	0,0366	-0,0515	2,300	199,5
5	45	167	2,223	-0,2119	0,0165	-0,0364	2,267	184,9
6	53	142	2,152	-0,0718	-0,0160	-0,0123	2,214	163,7
7	62	140	2,146	0,0858	-0,0187	0,0148	2,155	142,9
8	75	121	2,083	0,3135	-0,0476	0,0539	2,069	117,2
9	90	79	1,898	0,5762	-0,1321	0,0991	1,970	93,3
10	120	75	1,875	1,1016	-0,1427	0,1894	1,773	59,3
<i>Итого</i>	571	1697	21,87	4,1541	0,7143	-	21,87	1673,7

Расчет оценочных показателей представлен в табл. 43.

Таблица 43

Номер опыта	ε	ε/y	$g_{\hat{y}}$	$g_{\hat{y}}^2$	$(g_{\hat{y}} - g_{\hat{y}})^2$
1	-78,2	-0,232	0,106	0,02474	0,00263
2	-10,7	-0,043	0,088	0,00931	0,00007
3	13,8	0,070	0,064	0,00253	0,00019
4	14,5	0,078	0,052	0,00134	0,00024
5	17,9	0,107	0,037	0,00027	0,00042
6	21,7	0,153	0,012	0,00026	0,00078
7	2,9	0	-0,015	0,00035	0,00001
8	-3,8	-0,031	-0,054	0,00227	0,00004
9	14,3	0,181	-0,099	0,01745	0,00109
10	-15,7	-0,209	-0,189	0,02036	0,00214
<i>Итого</i>	-23,3	1,104	-	0,07888	0,00761

В результате имеем $R = \sqrt{1 - \frac{0,00761}{0,07888}} = 0,951$; $\bar{A} = \frac{1,104}{10} 100 \% =$

$= 11,04 \% > 10 \%$. В сравнении с предыдущим расчетом итог получился хуже, но в целом модель приемлема, хотя дает завышенное значение средней ошибки аппроксимации относительно допустимого уровня 10 %.

Модель имеет вид $\hat{y}' = 2,187(1 - 0,17195g_x)$; $g_x = \frac{x - 57,1}{57,1}$.

4.4. Степенная зависимость

В традиционном виде степенная модель представляется зависимостью $\hat{y} = ax^{b_0}$. Формы кривых модели показаны на графиках рис. 16 в простых координатах. В логарифмических координатах – это прямые линии. Поэтому, если прологарифмировать исходную зависимость $\lg y = \lg a + b_0 \lg x$ и сделать замены $\lg y = Y$, $\lg a = C$, $\lg x = X$, то приходим к линейному уравнению $\hat{Y} = C + b_0 X$.

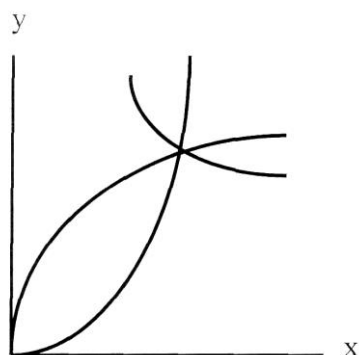


Рис. 16

В методе центра масс линейное уравнение имеет вид (3), где вместо базовой точки М берется точка С центра масс, что в логарифмической форме представляется как

$$Y = Y_c(1 \pm b_0 g_x), \text{ где } g_x = (X - X_c)/Y_c; b_0 = \sum |\Delta_Y| / \sum |\Delta_X|;$$

$$X_c = \sum X/n; Y_c = \sum Y/n; X = \lg x; Y = \lg y; \Delta_X = X - X_c; \Delta_Y = Y - Y_c.$$

После расчетов всех этих составляющих по логарифмам величин исходных данных для x и y и их подстановки в линейное уравнение рассчитываются значения Y , а по ним после взятия антилогарифмов (потенцирования) искомые величины теоретических значений \hat{y} . Для той же цели можно потенцировать последнюю формулу, выполнив предварительно ее преобразование следующим образом:

$$Y = Y_c(1 + b_0 g_x) = Y_c + Y_c b_0 g_x = Y_c + Y_c b_0 (X - X_c)/Y_c = Y_c - b_0 X_c + b_0 X.$$

Сделаем замену $Y_c - b_0 X_c = C$, тогда имеем $\hat{Y} = C + b_0 X$, т. е. получили линейное уравнение в логарифмической форме или без учета замен $\lg y = \lg a + b_0 \lg x$. Рассчитав численно $\lg a = C = \lg y_c - b_0 \lg x_c$, производим потенцирование последней линейной логарифмической функции с получением степенной модели $\hat{y} = ax^{b_0}$, которая используется для прогноза и анализа исследуемого однофакторного процесса.

Рассмотрим пример построения степенной зависимости различными методами со сравнительной оценкой результатов. Данные опытов приняты со значительным разбросом и приближены к гиперболической зависимости: на значительном участке результирующий признак стремится к пределу. Результаты расчетов по методу центра масс сведены в табл. 44.

Таблица 44

Номер опыта	x	y	$X = \lg x$	$Y = \lg y$	g_x	g_y	$\hat{Y} = \lg \hat{y}$	\hat{y}
1	50	252	1,6990	2,4014	-0,2384	-0,1960	2,606	403,6
2	100	693	2,0000	2,8407	-0,1372	-0,0490	2,768	586,1
3	150	840	2,1761	2,9243	-0,0790	-0,021	2,861	726,1
4	200	1155	2,3010	3,0626	-0,0370	0,025	2,928	847,2
5	250	1071	2,3979	3,0298	-0,0043	0,014	2,980	955,0
6	300	1176	2,4771	3,0704	0,0222	0,028	3,022	1052,0
7	350	1155	2,5441	3,0626	0,0446	0,025	3,058	1142,9
8	400	1302	2,6021	3,1146	0,0640	0,043	3,083	1227,4
9	500	1281	2,6990	3,1075	0,0965	0,040	3,141	1383,6
10	600	1323	2,7782	3,1216	0,1230	0,045	3,183	1524,0
11	700	1312,5	2,8451	3,1181	0,1454	0,044	3,219	1622,0
<i>Итого</i>	3600	11560,5	26,5196	32,8536	0,9916	0,530	32,849	11469,9

Расчетные данные: $Y_c = \sum Y/n = 32,8536/11 = 2,9867$; $X_c = \sum X/n = 26,5196/11 = 2,4109$; $b = \sum |g_y| / \sum |g_x| = 0,530/0,9916 = 0,5345$; $g_y = \frac{Y - Y_c}{Y_c}$; $g_x = \frac{X - X_c}{X_c}$. Получена модель $Y = 2,9867(1 + 0,5345g_x)$; $g_x = \frac{X - 2,4109}{2,9867}$. Приводим расчет по тем же исходным данным методом наименьших квадратов (табл. 45).

Таблица 45

Номер опыта	x	y	$\lg x$	$\lg y$	$\lg^2 x$	$\lg y \lg x$	\hat{y}
1	50	252	1,6990	2,4014	2,8866	4,0800	384,1
2	100	693	2,0000	2,8407	4,0000	5,6810	566,2
3	150	840	2,1761	2,9243	4,7354	6,3636	710,5
4	200	1155	2,3010	3,0626	5,2946	7,0470	834,7
5	250	1071	2,3979	3,0298	5,7499	7,2652	945,8
6	300	1176	2,4771	3,0704	6,1360	7,6057	1047,5
7	350	1155	2,5441	3,0626	6,4724	7,7916	1141,9
8	400	1302	2,6021	3,1146	6,7709	8,1046	1230,6
9	500	1281	2,6990	3,1075	7,2846	8,3871	1394,4
10	600	1323	2,7782	3,1216	7,7184	8,6724	1544,3
11	700	1312,5	2,8451	3,1181	8,0946	8,8713	1683,5
<i>Итого</i>	3600	11560,5	26,5196	32,8536	65,1435	79,8694	11483,5

Вид модели $\hat{y} = ax^{b_0}$. В логарифмическом виде $\lg \hat{y} = \lg a + b_0 \lg x$. Система нормальных уравнений для определения $\lg a$ и b_0

$$\begin{cases} n \lg a + b_0 \sum \lg x = \sum \lg y, \\ \lg a \sum \lg x + b_0 \sum \lg^2 x = \sum \lg y \lg x. \end{cases}$$

То же в числах

$$\begin{cases} 11 \lg a + 26,5196 b_0 = 32,8536, \\ 26,5196 \lg a + 65,1435 b_0 = 79,8694. \end{cases}$$

Решение системы: $\lg a = 1,633$; $b_0 = 0,56$. Тогда имеем $\lg \hat{y} = 1,633 + 0,56 \lg x$. После потенцирования с учетом $a = 10^{1,633} = 42,95$ получаем окончательно модель $\hat{y} = 42,95x^{0,56}$. Аналогичную таблицу расчетов и определение расчетных данных приводим для метода

экстремальных опорных точек (табл. 46). В таблицу не включаем исходные данные для x и y , а также их логарифмы. Они имеются в предыдущих таблицах.

Таблица 46

Номер опыта	g_x	g_y	bg_x	\hat{Y}	\hat{y}
1	0,3885	0,2307	0,1229	2,7380	547,0
2	0,2801	0,0900	0,0886	2,8450	699,8
3	0,2167	0,0632	0,0685	2,9078	808,7
4	0,1718	0,0189	0,0543	2,9521	895,6
5	0,1369	0,0294	0,0433	2,9864	969,2
6	0,1084	0,0164	0,0343	3,0145	1034,0
7	0,0843	0,0189	0,0267	3,0383	1092,2
8	0,0634	0,0022	0,0201	3,0589	1145,2
9	0,0285	0,0045	0,0090	3,0935	1240,2
10	0	0	0	3,1216	1323,0
11	-0,0241	0,0011	-0,0076	3,1453	1397,3
<i>Итого</i>	1,5027	0,4753	-	32,9014	11152,2

Вид модели в логарифмической форме:

$$\hat{Y} = Y_{\max} (1 - bg_x); \quad g_x = \left| \frac{X_0 - X}{X_0} \right|; \quad g_y = \frac{Y_{\max} - Y}{Y_{\max}}; \quad b = \frac{\sum |g_y|}{\sum |g_x|}.$$

Модель в числах:

$$\hat{Y} = 3,1216(1 - 0,3163g_x); \quad g_x = \left| \frac{2,7782 - X}{2,7782} \right|; \quad g_y = \frac{3,1216 - Y}{3,1216};$$

$$b = \frac{0,4753}{1,5027} = 0,3163.$$

Значения \hat{y} определены потенцированием $\hat{y} = 10^{\hat{Y}}$. Дадим точностную оценку полученным моделям. Коэффициент корреляции

определим из соотношений [7] $r_{\hat{y}} = \sqrt{1 - \frac{S_{\hat{y}}^2}{S_y^2}}; \quad S_{\hat{y}}^2 = \frac{\sum (y - \hat{y})^2}{n - 2};$

$$S_y^2 = \frac{\sum (y - \bar{y})^2}{n}; \quad \bar{y} = \frac{\sum y}{n} = 1051. \text{ Все результаты сводим в табл. 47.}$$

Таблица 47

Номер опыта	Метод центра масс		Метод наименьших квадратов	Метод экстремальных опорных точек
	$(y - \hat{y})^2$	$(y - \bar{y})^2$	$(y - \hat{y})^2$	$(y - \hat{y})^2$
1	22983	638401	17424	289229
2	11428	128164	16081	20880
3	12973	44521	16765	2153
4	94741	10816	102579	48400
5	13456	400	15670	7744
6	15376	15625	16515	20765
7	146	10816	171	5476
8	5565	63001	5098	29929
9	10527	52900	12860	3025
10	40401	73924	48974	0
11	95790	68382	137641	7482
<i>Итого</i>	323386	1107010	389778	435083
$r_{\hat{y}}$	0,841		0,804	0,779

Расчет средней ошибки аппроксимации дан в табл. 48.

Таблица 48

Номер опыта	Метод центра масс		Метод наименьших квадратов		Метод экстремальных опорных точек	
	ε	ε/y	ε	ε/y	ε	ε/y
1	151,6	0,602	132,1	0,524	295	1,171
2	-106,9	-0,154	-126,8	-0,183	6,8	0,010
3	-113,9	-0,136	-129,5	-0,154	-31,3	-0,037
4	-307,8	-0,266	-320,3	-0,277	-259,4	-0,225
5	-116,0	-0,108	-125,2	-0,117	-101,8	-0,095
6	-124,0	-0,105	-128,5	-0,109	-142	-0,121
7	-12,1	-0,010	-13,1	-0,011	-62,8	-0,054
8	-74,6	-0,057	-71,4	-0,055	-156,8	-0,120
9	102,6	0,080	113,4	0,089	-41	-0,032
10	201,0	0,152	221	0,167	0	0
11	309,5	0,236	370,5	0,282	84,8	0,065
<i>Итого</i>	-90,6	1,906	-77,8	1,968	-408,5	1,930
\bar{A}	17,3 %		17,9 %		17,5 %	

Видим, что по обоим критериям метод центра масс имеет преимущество. Но в целом ни один из методов не дал достаточного приближения аппроксимирующей кривой к опытным данным, так как линеаризация с помощью логарифмирования оказалась неудачной. Более удачным могло бы быть преобразование вида $X = 1/x$ (см. пример в п. 4.3).

4.5. Показательная зависимость

Уравнение вида $\hat{y} = ab_0^x$ представляется кривыми, исходящими из одной точки (рис. 17). После логарифмирования с целью линеаризации получается модель $\lg \hat{y} = \lg a + x \lg b_0$, которая после замен и преобразований приводится к виду $\hat{Y} = Y_c(1 \pm bg_x)$; $g_x = (x - x_c)/Y_c$; $b = \frac{\sum |g_Y|}{\sum |g_x|}$; $x_c = \sum x/n$; $Y_c = \sum Y/n$; $Y = \lg y$; $\Delta_x = x - x_c$; $\Delta_Y = Y - Y_c$. Рассчитанные по исходным данным для x и y значения \hat{Y} потенцируются. Можно использовать и другой вид линейной

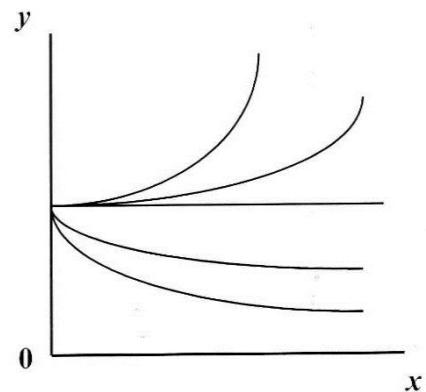


Рис. 17

зависимости: $\hat{Y} = Y_c(1 \pm bg_x)$; $g_x = \frac{x - x_c}{Y_c}$; $g_Y = \frac{Y - Y_c}{Y_c}$; $b = \frac{\sum |g_Y|}{\sum |g_x|}$; $Y = \lg y$; $Y_c = \sum Y/n$.

Рассмотрим вывод показательной модели, представленной в табл. 49.

Таблица 49

Номер опыта	x	y	$Y = \lg y$	g_x	g_Y	bg_x	$\hat{Y} = \lg \hat{y}$	\hat{y}
1	0,2	13,20	1,1206	-2,4081	-0,3040	-0,2596	1,0828	12,10
2	0,5	12,10	1,0828	-2,0591	-0,2600	-0,2220	1,0504	11,23
3	1,0	9,46	0,9759	-1,4774	-0,1350	-0,1593	0,9965	9,92
4	1,5	8,36	0,9222	-0,8958	-0,0730	-0,0966	0,9426	8,76
5	2,0	7,04	0,8476	-0,3141	0,0140	-0,0539	0,8887	7,74
6	2,5	6,60	0,8195	0,2676	0,0466	0,0288	0,8348	6,84
7	3,0	5,94	0,7738	0,8492	0,0998	0,0915	0,7809	6,04
8	3,5	5,06	0,7042	1,4309	0,1808	0,1543	0,7270	5,33
9	4,0	4,84	0,6848	2,0126	0,2034	0,2170	0,6731	4,71
10	4,5	4,62	0,6646	2,5942	0,2268	0,2797	0,6192	4,16
<i>Итого</i>	22,7	77,22	8,596	14,3090	1,5428	—	8,596	76,83

Расчетные соотношения: $x_c = \frac{22,7}{10} = 2,27$; $Y_c = \frac{8,596}{10} = 0,8596$;

$$g_x = \frac{x - x_c}{Y_c} = \frac{x - 2,27}{0,8596}; g_Y = \frac{Y_c - Y}{Y_c} = \frac{0,8596 - Y}{0,8596}; b = \frac{\sum |g_Y|}{\sum |g_x|} = \frac{1,5428}{14,3090} = 0,1078.$$

Получена модель вида $\hat{Y} = 0,8596(1 - 0,1078g_x)$; $g_x = \frac{x - 2,27}{0,8596}$. Рассчитанные значения \hat{Y} потенцируются, и результат \hat{y} заносится в расчетную таблицу. Следует заметить, что потенцировать можно и всю расчетную формулу. Сделаем ее преобразование: $\hat{Y} = 0,8596 - 0,0927g_x$. Потенцирование данного выражения дает $\hat{y} = 10^{0,8596} \cdot 10^{-0,0927} g_x$. После возведения в степени получаем окончательную формулу в показательном виде $\hat{y} = 7,2377 \cdot 0,8078^{g_x}$; $g_x = \frac{x - 2,27}{0,8596}$. Оценим модель (табл. 50).

Таблица 50

Номер опыта	$g_x g_Y$	g_x^2	g_Y^2	$g_{\hat{Y}}$	$(g_Y - g_{\hat{Y}})^2$	$ g_Y - b g_x $	ε	ε/y
1	0,7311	5,7989	0,092	0,2597	0,0019	0,0440	-1,10	-0,083
2	0,5347	4,2399	0,067	0,2220	0,0014	0,0377	-0,87	-0,072
3	0,1999	2,1827	0,018	0,1593	0,0006	0,0240	0,46	0,049
4	0,0652	0,8025	0,005	0,0966	0,0006	0,0238	0,40	0,048
5	0,0044	0,0987	0	0,0339	0,0004	0,0479	0,70	0,099
6	0,0125	0,0716	0,002	0,0289	0,0003	0,0178	0,24	0,036
7	0,0848	0,7211	0,010	0,0916	0,0001	0,0083	0,10	0,017
8	0,2587	2,0475	0,033	0,1543	0,0007	0,0265	0,27	0,053
9	0,4094	4,0506	0,041	0,2170	0,0002	0,0136	-0,13	-0,027
10	0,5884	6,7299	0,051	0,2797	0,0028	0,0529	-0,46	-0,100
<i>Итого</i>	2,8803	26,7434	0,321	1,5430	0,0090	0,2965	-0,39	0,584

Здесь $g_{\hat{Y}} = \frac{Y_c - \hat{Y}}{Y_c}$. Коэффициент корреляции $r = \frac{\sum g_x g_Y}{\sqrt{\sum g_x^2 \sum g_Y^2}} = \frac{2,8803}{\sqrt{26,7434 \cdot 0,321}} = 0,983$ близок к единице, значит между x и Y тесная

связь. Индекс корреляции $R = \sqrt{1 - \frac{\sum (|g_Y| - g_{\hat{Y}})^2}{\sum g_Y^2}} = \sqrt{1 - \frac{0,009}{0,321}} = 0,986$ близок к коэффициенту корреляции, что означает правильность выбора типа уравнения (показательная функция). Коэффициент устойчивости

связи $K = 1 - \frac{\sum |g_Y - bg_x|}{\sum g_Y} = 1 - \frac{0,2965}{1,5430} = 0,81$ свидетельствует о высоком

уровне устойчивости связи x и Y . Средняя ошибка аппроксимации

$$\bar{A} = \frac{\sum |\varepsilon/y|}{n} 100 \% = \frac{0,584}{10} 100 \% = 5,84 \% < 10 \%, \text{ т. е. в пределах допустимого уровня.}$$

Приводим расчет показательной модели методом наименьших квадратов с использованием системы нормальных уравнений

$$\begin{cases} na + \sum xb_0 = \sum Y, \\ a \sum x + b_0 \sum x^2 = \sum Yx, \end{cases}$$

где $Y = \lg y$ (табл. 51).

Таблица 51

Номер опыта	x	y	Y	x^2	Yx	\hat{Y}	\hat{y}
1	0,2	13,20	1,1206	0,04	0,2241	1,0826	12,10
2	0,5	12,10	1,0828	0,25	0,5414	1,0503	11,23
3	1,0	9,46	0,9759	1,00	0,9759	0,9964	9,92
4	1,5	8,36	0,9222	2,25	1,3833	0,9426	8,76
5	2,0	7,04	0,8476	4,00	1,6952	0,8887	7,74
6	2,5	6,60	0,8195	6,25	2,0488	0,8349	6,84
7	3,0	5,94	0,7738	9,00	2,3214	0,7810	6,04
8	3,5	5,06	0,7042	12,25	2,4647	0,7270	5,33
9	4,0	4,84	0,6848	16,00	2,7392	0,6733	4,71
10	4,5	4,62	0,6646	20,25	2,9907	0,6195	4,16
<i>Итого</i>	22,7	77,22	8,596	71,29	17,3847	8,5965	76,83

Видим, что расчетные значения \hat{y} полностью совпали с аналогичными значениями для метода центра масс, поэтому оба метода в данном случае оказались идентичными по качеству модели.

Решение системы нормальных уравнений дало $a = 1,1041$, $b_0 = -0,1077$. Модель в логарифмической форме $\hat{Y} = 1,1041 - 0,1077x$. После потенцирования получаем $\hat{y} = 10^{1,1041} \cdot 10^{-0,1077x}$ или окончательно $\hat{y} = 12,709 \cdot 0,7804^x$. Дадим решение данной задачи методом экстремальных опорных точек (табл. 52).

Таблица 52

Номер опыта	g_x	g_Y	bg_x	\hat{Y}	\hat{y}
1	0	0	0	1,1206	13,20
2	1,5	0,0337	0,0338	1,0828	12,10
3	4,0	0,1291	0,0900	1,0197	10,46
4	6,5	0,1770	0,1463	0,9567	9,05
5	9,0	0,2436	0,2025	0,8937	7,83
6	11,5	0,2687	0,2588	0,8306	6,77
7	14,0	0,3095	0,3150	0,7676	5,86
8	16,5	0,3716	0,3713	0,7046	5,07
9	19,0	0,3889	0,4275	0,6415	4,38
10	21,5	0,4069	0,4838	0,5785	3,79
<i>Итого</i>	103,5	2,3290	–	8,5963	78,51

Вид модели в логарифмической форме:

$$\hat{Y} = Y_{\max} (1 - bg_x); \quad g_x = \frac{x - x_{\min}}{x_{\min}}; \quad g_Y = \frac{Y_{\max} - Y}{Y_{\max}}; \quad b = \frac{\sum g_Y}{\sum g_x}.$$

Модель в численном выражении $\hat{Y} = 1,1206(1 - 0,0225g_x)$. Ее логарифмирование дает $\hat{y} = 10^{1,1206} \cdot 10^{-0,0252g_x}$, что после возведения в степень приводит к окончательной формуле $\hat{y} = 13,2 \cdot 0,9436^{g_x}$; $g_x = \frac{x - 0,2}{0,2}$.

Данные для оценки показательной модели сведены в табл. 53.

Таблица 53

Номер опыта	$g_x g_Y$	g_x^2	g_Y^2	$g_{\hat{y}}$	$(g_Y - g_{\hat{y}})^2$	$ g_Y - bg_x $	ε	ε/y
1	0	0	0	0	0	0	0	0
2	0,0506	2,25	0,0011	0,0337	0	0,0001	0	0
3	0,5164	16,00	0,0167	0,0900	0,0015	0,0391	1,00	0,106
4	1,1505	42,25	0,0430	0,1463	0,0009	0,0307	0,69	0,083
5	2,1924	81,00	0,0593	0,2025	0,0017	0,0411	0,79	0,112
6	3,0901	132,25	0,0722	0,2588	0,0001	0,0090	0,17	0,026
7	4,3330	196,00	0,0958	0,3150	0	0,0055	-0,08	-0,013
8	6,1314	272,25	0,1381	0,3712	0	0,0003	0,01	0,002
9	7,3891	361,00	0,1512	0,4275	0,0015	0,0386	-0,46	-0,095
10	8,7484	462,25	0,1656	0,4838	0,0059	0,0769	0,83	-0,180
<i>Итого</i>	33,6019	1565,25	0,7313	2,3288	0,0116	0,2422	1,29	0,617

Коэффициент корреляции

$$r = \frac{\sum g_x g_y}{\sqrt{\sum g_x^2 \sum g_y^2}} = \frac{33,6019}{\sqrt{1565,25 \cdot 0,7313}} = 0,993.$$

Индекс корреляции

$$R = \sqrt{1 - \frac{\sum (g_y - g_{\hat{y}})^2}{\sum g_y^2}} = \sqrt{1 - \frac{0,0116}{0,7313}} = 0,992.$$

Коэффициент устойчивости связи

$$K = 1 - \frac{\sum |g_y - b g_x|}{\sum g_y} = 1 - \frac{0,2422}{2,3288} = 0,896.$$

Относительная ошибка аппроксимации

$$\bar{A} = \frac{\sum |\varepsilon/y|}{n} 100 \% = \frac{0,617}{10} 100 \% = 6,17 \% < 10 \%.$$

Результаты оценок указывают на преимущество метода экстремальных опорных точек за исключением одного последнего показателя, который оказался здесь несколько хуже, но в допустимых пределах.

4.6. Многофакторная зависимость

Рассмотрим пример построения комбинационной двухфакторной модели, сочетающей линейную и гиперболическую зависимости факторных и результирующего признаков, заданных в табл. 27. Но решение в отличие от уже рассмотренных в п. 3.6 дадим по методу центра масс, сведя результаты вычислений в табл. 54. Линеаризацию выполним путем логарифмирования исходных данных.

Таблица 54

Номер опыта	Y	X ₁	X ₂	g _Y	g _{X1}	g _{X2}	g _{X1X2}	bg _{X1X2}	Ŷ	ŷ
1	1,7709	1,7324	2,4314	0,1477	-0,1567	-0,0964	-0,2531	-0,2023	1,8551	71,6
2	1,7324	1,8921	2,4548	0,1227	-0,0789	-0,0877	-0,1666	-0,1331	1,7484	56,0
3	1,6902	1,9777	2,5740	0,0954	-0,0372	-0,0434	-0,0806	-0,0644	1,6424	43,9
4	1,6128	2,0899	2,6513	0,0452	0,0174	-0,0147	0,0027	0,0022	1,5396	34,6
5	1,5441	2,1072	2,7443	0,0007	0,0258	0,0199	0,0457	0,0365	1,4867	30,7
6	1,4472	2,1614	2,8096	-0,0621	0,0522	0,0442	0,0964	0,0770	1,4242	26,6
7	1,3424	2,2175	2,8976	-0,1300	0,0795	0,0769	0,1564	0,1250	1,3501	22,4
8	1,2041	2,2553	2,9638	-0,2196	0,0979	0,1015	0,1994	0,1594	1,2970	19,8
<i>Итого</i>	12,3441	16,4335	21,5268	0,8234	0,5456	0,4847	—	—	12,3435	305,6

Вид модели в логарифмической форме:

$$\hat{Y} = Y_c(1 - \beta g_{X_1 X_2}); \quad g_{X_1 X_2} = g_{X_1} + g_{X_2}; \quad g_{X_1} = \frac{X_1 - X_{1c}}{X_{1c}}; \quad g_{X_2} = \frac{X_2 - X_{2c}}{X_{2c}};$$

$$g_Y = \frac{Y - Y_c}{Y_c}; \quad X_1 = \lg x_1; \quad X_2 = \lg x_2; \quad Y = \lg y; \quad X_{1c} = \frac{\sum X_1}{n}; \quad X_{2c} = \frac{\sum X_2}{n};$$

$$Y_c = \frac{\sum Y}{n}; \quad B = \frac{\sum |g_Y|}{\sum |g_{X_1}| + \sum |g_{X_2}|}.$$

Комбинационная модель в численном выражении в логарифмической форме:

$$\hat{Y} = 1,543(1 - 0,7992 g_{X_1 X_2}); \quad g_{X_1 X_2} = g_{X_1} + g_{X_2}; \quad g_{X_1} = \frac{X_1 - 2,0542}{2,0542};$$

$$g_{X_2} = \frac{X_2 - 2,6908}{2,6908}.$$

Параметры модели:

$$B = \frac{\sum |g_Y|}{\sum |g_{X_1}| + \sum |g_{X_2}|} = \frac{0,8234}{0,5456 + 0,4847} = 0,7992; \quad Y_c = \frac{\sum Y}{n} = \frac{12,3441}{8} =$$

$$= 1,543; \quad X_{1c} = \frac{\sum X_1}{n} = \frac{16,4335}{8} = 2,0542; \quad X_{2c} = \frac{\sum X_2}{n} = \frac{21,5268}{8} = 2,6908;$$

$$g_Y = \frac{Y - 1,543}{1,543}; \quad g_{X_1} = \frac{X_1 - 2,0542}{2,0542}; \quad g_{X_2} = \frac{X_2 - 2,6908}{2,6908}.$$

Проверочные расчеты внесены в табл. 55.

Таблица 55

Номер опыта	$g_{\hat{Y}}$	g_Y^2	$(g_Y - g_{\hat{Y}})^2$	ε	$ \varepsilon/y $
1	0,2023	0,0218	0,0030	12,6	0,2136
2	0,1331	0,0151	0,0001	2,0	0,0370
3	0,0644	0,0091	0,0010	-5,1	0,1041
4	-0,0022	0,0020	0,0022	-6,4	0,1561
5	-0,0365	0	0,0014	-4,3	0,1229
6	-0,0770	0,0039	0,0002	-1,4	0,0500
7	-0,1250	0,0169	0	0,4	0,0182
8	-0,1594	0,0482	0,0036	3,8	0,2375
<i>Итого</i>	-	0,1170	0,0115	1,6	0,9394

Индекс корреляции $R = \sqrt{1 - \frac{0,0115}{0,1170}} = 0,95$. Относительная ошибка

аппроксимации $\bar{A} = \frac{0,9394}{8} 100 \% = 11,7 \% > 10 \%$. Показанные результаты модели оказались хуже, чем для других методов (см. п. 3.6), что, по-видимому, связано с погрешностями линеаризации гиперболической зависимости путем логарифмирования.

5. МЕТОДЫ ПОДГОТОВКИ ИСХОДНЫХ ДАННЫХ

5.1. Оптимизация положения начала координат при построении гиперболы

Анализ расчетных данных для моделей различных видов показывает, что гиперболическая зависимость дает наиболее неустойчивые результаты вблизи малых значений координаты x , что вполне объяснимо: характерная для гиперболы обратная связь $y = f(1/x)$ приводит при x , стремящемся к нулю, к бесконечно большим значениям y . В результате теоретическое значение \hat{y} сильно отклоняется от фактического y , вызывая рост средней ошибки аппроксимации иногда за предельно допустимый уровень из-за одной-двух точек с малыми значениями x . Например, остаток ε в табл. 36 для опытов № 1 и 2 увеличен, и в результате в общем отношении $\varepsilon/y = 0,95$ он в сумме составляет величину 0,78, т. е. более 82 %, создавая впечатление о неудачном подборе модели. Если заранее все значения x увеличить хотя бы незначительно, то можно существенно улучшить результат. Рассмотрим два примера такого подхода для данных табл. 35.

Увеличим x на 20 единиц и приведем данные расчетов в табл. 56.

Таблица 56

Номер опыта	x	y	$x' = x + 20$	$X = 1/x'$	g_x	g_y	bg_x	\hat{y}	ε	$ \varepsilon/y $
1	5	42	25	0,040	0,7167	-0,840	0,6913	96,7	54,7	1,302
2	10	170	30	0,033	0,4163	-0,353	0,3671	166,3	-3,7	0,022
3	15	240	35	0,029	0,2446	-0,086	0,2157	206,0	-34	0,142
4	20	275	40	0,025	0,0730	0,047	0,0644	245,8	-29,2	0,106
5	25	280	45	0,022	-0,0558	0,066	-0,0492	275,6	-4,4	0,016
6	30	310	50	0,020	-0,1416	0,180	-0,1248	295,5	-14,5	0,047

Окончание табл. 56

Номер опыта	x	y	$x' = x + 20$	$X = 1/x'$	g_x	g_y	bg_x	\hat{y}	ε	$ \varepsilon/y $
7	35	325	55	0,018	-0,2275	0,237	-0,2006	315,2	-9,8	0,030
8	40	320	60	0,017	-0,2704	0,218	-0,2384	325,3	5,3	0,017
9	45	335	65	0,015	-0,3562	0,275	-0,3141	345,2	10,2	0,030
10	50	330	70	0,014	-0,3991	0,256	-0,3519	346,5	16,5	0,050
<i>Итого</i>	-	2627	-	0,233	2,9012	2,558	-	2618,1	-8,9	1,762

Модель гиперболы в общем виде $\hat{y} = y_c(1 - bg_x)$; $g_x = \frac{X - X_c}{X_c}$; $X_c = \frac{\sum X}{n}$; $X = \frac{1}{x'}$; $x' = x + N$, где N – некоторое число (в нашем случае $N = 20$). Имеем $X_c = \frac{0,233}{10} = 0,0233$; $y_c = \frac{2627}{10} = 262,7$; $b = \frac{2,558}{2,9012} = 0,8817$. Полученная модель гиперболы в числах: $\hat{y} = 262,7(1 - 0,8817g_x)$; $g_x = \frac{X - 0,0233}{0,0233}$; $X = \frac{1}{x + 20}$. Средняя ошибка аппроксимации $\bar{A} = \frac{1,762}{10} 100\% = 17,6\% > 10\%$. Видим, что перенос начала координат по оси x на величину $N = 20$ в рассмотренном примере изменил параметры модели гиперболы, ее качество ухудшилось в сравнении с рассчитанной по данным табл. 35. Возьмем другой пример, приняв $N = 5$. Все изменившиеся параметры при этом заносим в табл. 57.

Таблица 57

Номер опыта	$x' = x + 5$	$X = 1/x'$	g_x	bg_x	\hat{y}	ε	$ \varepsilon/y $
1	10	0,100	1,475	0,7862	56,2	14,2	0,340
2	15	0,067	0,658	0,3507	170,6	0,6	0,004
3	20	0,050	0,238	0,1269	229,4	-10,6	0,044
4	25	0,040	-0,001	-0,0005	262,8	-12,2	0,044
5	30	0,033	-0,250	-0,1333	297,7	17,7	0,063
6	35	0,029	-0,282	-0,1503	302,2	-7,8	0,025
7	40	0,025	-0,381	-0,2031	316,1	-8,9	0,027
8	45	0,022	-0,455	-0,2425	326,4	6,4	0,020
9	50	0,020	-0,505	-0,2692	334,4	-0,6	0,002
10	55	0,018	-0,554	-0,2953	340,3	10,3	0,030
<i>Итого</i>	-	0,404	4,799	-	2636,1	9,1	0,599

Модель гиперболы, учитывая $X_c = \frac{0,404}{10} = 0,0404$; $b = \frac{2,558}{4,799} = 0,533$:

$$\hat{y} = 262,7(1 - 0,533g_x); g_x = \frac{X - 0,0404}{0,0404}; X = \frac{1}{x + 5}. \text{ Средняя ошибка ап-}$$

проксимации $\bar{A} = \frac{0,599}{10} 100 \% = 5,99 \% < 10 \%$, ее величина – наименьшая из всех рассмотренных примеров для гиперболы. Все погрешности ε в среднем выровнились.

Опыт показывает, что оптимальное значение N находится в диапазоне, при котором соотношение величин X на начальном участке гиперболы и в области точки ее симметрии составляет $2,5 \dots 3$ (в рассмотренных случаях при $N = 0$ и $N = 20$ эти отношения составили $0,2/0,04 = 5$ и $0,040/0,022 = 1,8$ соответственно, т. е. за пределами указанного рекомендованного диапазона).

5.2. Определение положения оси симметрии кривой

Неполнота информации в связи с недостаточностью статистических наблюдений является причиной неточной аппроксимации опытных данных какой-либо зависимостью. Например, точки на рис. 18 могут принадлежать прямой, различным параболом, синусоиде и т. д. В таких случаях выбор вида кривой обуславливается опытом наблюдения ранее аналогичных процессов и наилучшими оценочными показателями при регрессионном и корреляционном анализе полученных моделей. Если кривая выбрана, например, на основе аналогии с подобными процессами, то ее построение вызывает затруднения в случае отсутствия опорных точек. Для их поиска необходимы специальные приемы. Рассмотрим некоторые из них при построении параболы и гиперболы.

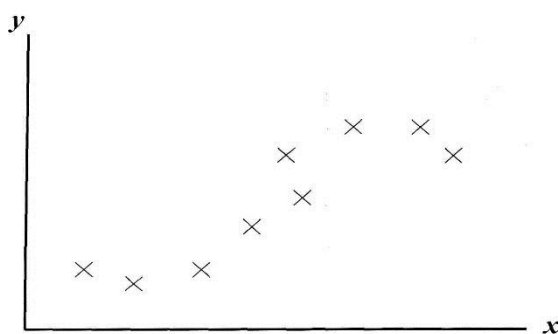


Рис. 18

Для параболы наибольшее затруднение вызывает поиск оси симметрии и экстремальной точки на этой оси. Рассмотрим, как можно решить данную задачу. Пусть имеются точки наблюдения в достаточном

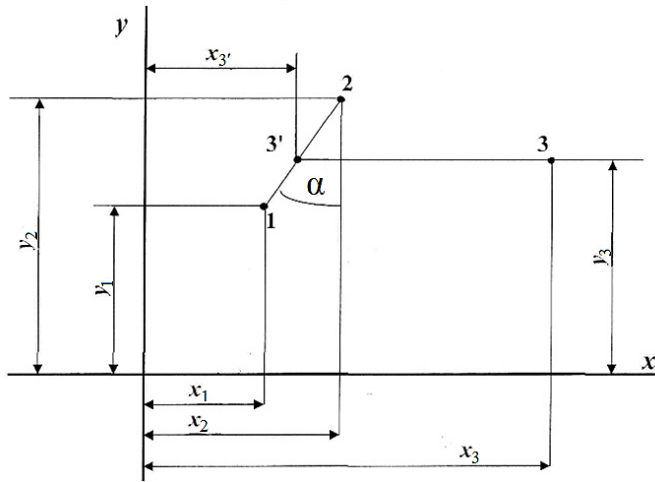


Рис. 19

количестве для левой ветви параболы, а для правой есть лишь одно-два наблюдения (рис. 19). Для построения оси симметрии выберем две точки 1 и 2 на левой ветви и одну – на правой с примерным относительным расположением как показано на рисунке. Соединим точки 1 и 2 прямой и на нее спроектируем точку 3, получив точку 3' на уровне y_3 . По координатам точек находим тангенсы угла α наклона пря-

мой 1 – 2 к оси y $\operatorname{tg}\alpha = \frac{x_2 - x_1}{y_2 - y_1}$; $\operatorname{tg}\alpha = \frac{x_2 - x_{3'}}{y_2 - y_3}$, откуда находим

$x_{3'} = x_2 - \frac{x_2 - x_1}{y_2 - y_1}(y_2 - y_3)$. Координата оси симметрии параболы будет

$x_0 = \frac{x_3 - x_{3'}}{2}$. После подстановки значения $x_{3'}$ имеем окончательное

выражение $x_0 = \frac{1}{2} \left[x_3 - x_2 - \frac{x_2 - x_1}{y_2 - y_1}(y_2 - y_3) \right]$. Экстремальную опорную точку

можно брать на оси x_0 либо как ближайшую к оси наивысшую точку. Если такой точки нет или ближайшая точка ниже соседних точек, то можно сделать графическую экстраполяцию кривой до оси x_0 или же рассчитать y_{\max} по координатам соседних точек (рис. 20). Из очевидных со-

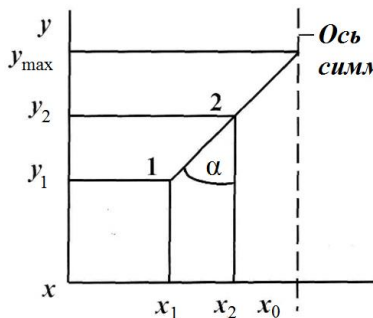


Рис. 20

отношений $\operatorname{tg}\alpha = \frac{x_2 - x_1}{y_2 - y_1}$; $\operatorname{tg}\alpha = \frac{x_0 - x_1}{y_{\max} - y_1}$

имеем $\frac{x_2 - x_1}{y_2 - y_1} = \frac{x_0 - x_1}{y_{\max} - y_1}$, откуда после преобразований находим $y_{\max} =$

$$= \frac{y_2(x_0 - x_1) - y_1(x_0 - x_2)}{x_2 - x_1}.$$

Для симметричной гиперболы вида $y = f(1/x)$ ось симметрии проходит под углом 45° к осям координат. Чтобы найти ее положение, проще всего повернуть оси координат на 45° (рис. 21) и преобразовать гиперболу в параболу. Поворот осей обеспечивают соотношения $x' = x \cos \varphi + y \sin \varphi$; $y' = -x \sin \varphi + y \cos \varphi$ $\cos \varphi = \sin \varphi = 0,7071$. Вид модели параболы

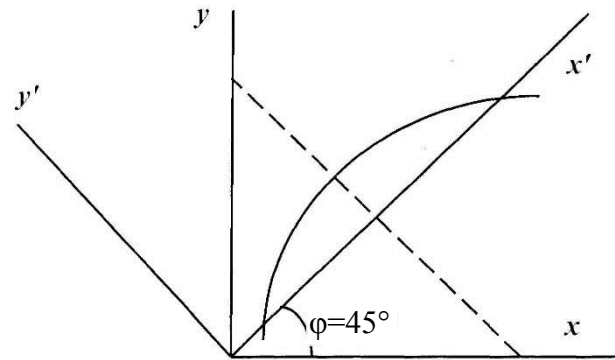


Рис. 21

$$\hat{y}' = y'_{\max} (1 - bg_{x'}); \quad g_{x'} = \frac{(x' - x'_0)^2}{x'^2}, \quad \text{где}$$

x'_0 – координата оси симметрии, которая может определяться по опытными данным в соответствии с y_{\max} либо находиться вышеописанным способом. Рассмотрим пример преобразования по исходным данным табл. 14 и построим модель параболы методом экстремальных опорных точек (табл. 58).

Таблица 58

Номер опыта	x	y	x'	y'	$g_{x'}$	$g_{y'}$	$bg_{x'}$	\hat{y}'	\hat{y}
1	5	42	33,2	26,2	0,76815	0,87226	1,15376	-31,5	1,2
2	10	170	127,3	113,1	0,27693	0,44856	0,41595	119,8	174,7
3	15	240	180,3	159,1	0,10824	0,22428	0,16258	171,8	249
4	20	275	208,6	180,3	0,05003	0,12092	0,07515	189,7	281,6
5	25	280	215,7	180,3	0,03891	0,12092	0,05844	193,1	289,1
6	30	310	240,4	198,1	0,01109	0,03462	0,01666	201,7	312,6
7	35	325	254,6	205,1	0,00275	0	0,00413	204,3	324,5
8	40	320	254,6	198,0	0,00275	0,03462	0,00413	204,3	324,5
9	45	335	268,7	205,1	0	0	0	205,1	335
10	50	330	268,7	198,0	0	0,03462	0	205,1	335
<i>Итого</i>	-	2627	-	1663,3	1,25885	1,8908	-	1663,4	2627,2

Были приняты значения $x'_0 = 268,7$; $y'_{\max} = 205,1$ для опыта № 9.

При этом $g_{y'} = 1 - \frac{y'}{y'_{\max}}$, тогда $b = \frac{1,8908}{1,25885} = 1,502$. Данные для модели

гиперболы получались обратным поворотом осей координат $\hat{y} = x' \sin \varphi + \hat{y}' \cos \varphi = 0,7071(x' + \hat{y}')$. Расчет показателей качества гиперболы сведен в табл. 59.

Таблица 59

Номер опыта	ε	$ \varepsilon/y $	$g_{y'}$	$g_{\hat{y}'}$	$(g_{y'} - g_{\hat{y}'})^2$
1	-40,8	0,97	0,761	1,153	0,079
2	4,7	0,03	0,201	0,416	0,001
3	9,0	0,04	0,050	0,162	0,004
4	6,6	0,02	0,015	0,075	0,002
5	9,1	0,03	0,015	0,059	0,004
6	2,6	0,01	0,001	0,017	0
7	-0,5	0	0	0,004	0
8	4,5	0,01	0,001	0,004	0
9	0	0	0	0	0
10	0	0	0,001	0	0,001
<i>Итого</i>	-4,8	1,11	1,045	-	0,091

Индекс корреляции $R = \sqrt{1 - \frac{0,091}{1,045}} = 0,955$. Относительная ошибка

аппроксимации $\bar{A} = \frac{1,11}{10} 100 \% = 11,1 \% > 10 \%$. Показатели удовлетворительные. В целом модель гиперболы такова:

$$\hat{y} = 0,7071(x' + \hat{y}');$$

$$x' = 0,7071(x + y); \hat{y}' = 205,1(1 - 1,502g_{x'}); g_{x'} = \frac{(x' - 268,7)^2}{72199,7}.$$

Для повышения точности построения модели можно относительно оси симметрии вносить в исходные данные фиктивные точки, расположенные симметрично по отношению к координатам опытных точек. Если координаты опытной точки x_i, y_i , а оси симметрии x_0 , то координаты фиктивной точки будут $x_0 \pm (x_0 \mp x_i); y_i$. Верхние знаки берутся при наличии опытных данных для левой ветви параболы, а нижние – в случае определенности точек для правой ветви. Фиктивные точки могут создаваться и при построении гиперболы с использованием поворота осей координат. Эти точки добавляются к координатам точек получаемой промежуточной параболы (см. предыдущий пример).

5.3. Выбор показателя степени для нелинейной зависимости

Параболы, гиперболы и другие кривые для аппроксимации опытных данных могут быть разных степеней. Возникает вопрос, какую степень лучше принять, какой показатель степени оптимален? В экономической практике обычно используют параболу второй степени, в лучшем

случае вида $y = f(x^2, x)$, в худшем – $y = f(x^2)$, гиперболу представляют в виде обратной функции первой степени вида $y = f(1/x)$. Подобные степенные зависимости были рассмотрены выше и отмечалась возможность построения моделей n -го порядка.

Выбор наилучшего показателя степени возможен с применением графического способа на основе логарифмирования исходных данных. Приведем пример подобного выбора показателя степени для обратной гиперболы. В качестве исходных данных примем пример по табл. 25 и сопоставим результаты при оптимальном показателе степени и для гиперболы третьей степени по п. 3.5.

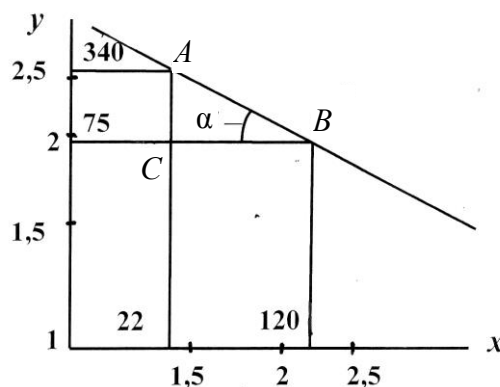


Рис. 22

Нанесем опытные данные для x и y с целью линеаризации зависимости между ними на сетку логарифмических координат, проводим через них прямую и определяем тангенс угла ее наклона α к оси x , который и является искомым показателем степени (рис. 22).

Тангенс угла определен как отношение катетов в прямоугольном треугольнике ABC , измеренных в миллиметрах. Параметры на рисунке и дальнейшие расчеты выполнены по исходным данным для обратной гиперболы, приведенным в табл. 25. Расчеты выполняем по методу экстремальных опорных точек. Так как $\text{tg } \alpha = -0,7$, то модель обратной гиперболы имеет вид $\hat{y} = y_{\max} (1 - bg_x)$; $g_x = \frac{x^{0,7} - x_{\min}^{0,7}}{x^{0,7}}$. Расчетные данные заносим в табл. 60.

Таблица 60

Номер опыта	x	y	g_x	g_y	bg_x	\hat{y}
1	22	340	0	0	0	340
2	28	250	0,15528	0,26471	0,19174	274,8
3	36	198	0,29155	0,41765	0,36000	217,6
4	40	185	0,34189	0,45988	0,42217	196,5
5	45	167	0,39403	0,50882	0,48655	174,6
6	53	142	0,45957	0,58235	0,56748	147,1
7	62	140	0,51580	0,58824	0,63691	123,7
8	75	121	0,57620	0,64412	0,71149	98,1
9	90	79	0,62695	0,76765	0,77416	76,7
10	120	75	0,69501	0,77941	0,85820	48,2
<i>Итого</i>	–	1697	4,05628	5,00883	–	1697,3

Получена модель вида $\hat{y} = 340(1 - 1,2348g_x)$; $g_x = \frac{x^{0,7} - 8,7036}{x^{0,7}}$,

при этом $b = \frac{\sum |g_y|}{\sum |g_x|} = \frac{5,00883}{4,05628} = 1,2348$; $x_{\min}^{0,7} = 22^{0,7} = 8,7036$; $g_y = 1 - \frac{y}{340}$.

Дадим оценку модели, сведя расчетные данные в табл. 61.

Таблица 61

Номер опыта	ε	$ \varepsilon/y $	g_y^2	$g_{\hat{y}}$	$(g_y - g_{\hat{y}})^2$
1	0	0	0	0	0
2	24,8	0,10	0,070	0,194	0,0050
3	19,6	0,10	0,174	0,360	0,0033
4	11,5	0,06	0,211	0,422	0,0014
5	7,6	0,05	0,259	0,486	0,0005
6	5,1	0,04	0,339	0,567	0,0002
7	-16,3	0,12	0,346	0,636	0,0023
8	-22,9	0,19	0,415	0,711	0,0045
9	-2,3	0,03	0,589	0,774	0,0000
10	-26,8	0,36	0,607	0,858	0,0062
<i>Итого</i>	0,8	1,05	3,01	—	0,0234

Индекс корреляции $R = \sqrt{1 - \frac{0,0234}{3,01}} = 0,996$, относительная ошибка

аппроксимации $\bar{A} = \frac{1,05}{10} 100 \% = 10,5 \% > 10 \%$. Все проверочные данные обнаружили для показателя степени 0,7 лучший результат, что позволяет считать его оптимальным.

6. ПОСТРОЕНИЕ СЛОЖНЫХ ОДНОФАКТОРНЫХ ЗАВИСИМОСТЕЙ

Под сложными будем понимать модели, содержащие в одной формуле несколько составляющих из одной и той же переменной, представленных в различной математической форме. Чаще всего сложные модели выражаются квадратным уравнением вида $y = a_0 + a_1x + a_2x^2$ или

приводятся к такому уравнению путем несложных преобразований. Данное уравнение геометрически отображает параболу второй степени – прямую или обратную в зависимости от значений коэффициента a_2 . Эллипс и гипербола могут задаваться уравнением $y^2 = a_0 + a_1x + a_2x^2$. Путем замены $Y = y^2$ приходим к формуле того же вида $Y = a_0 + a_1x + a_2x^2$. Аналогично преобразуются и другие виды сложных кривых [4]:

$y = \frac{1}{a_0 + a_1x + a_2x^2}$ приводится к такому же виду заменой $Y = \frac{1}{y}$; выражение $y = \frac{x}{a_0 + a_1x + a_2x^2}$ – введением переменной $Y = \frac{x}{y}$; зависимость $y = a_0 + a_1/x + a_2/x^2$ – заменой $X = 1/x$ с получением формулы $y = a_0 + a_1X + a_2X^2$; показательная функция $y = ae^{bx+cx^2}$ – путем логарифмирования $\lg y = \lg a + \lg e \cdot bx + \lg e \cdot cx^2$ и замен $Y = \lg y$, $a_0 = \lg a$, $a_1 = \lg e \cdot b$, $a_2 = \lg e \cdot c$ с получением $Y = a_0 + a_1x + a_2x^2$, т. е. того же вида уравнения.

6.1. Метод линеаризации сложных однофакторных зависимостей с помощью дробно-линейной функции

Пусть требуется получить на основе обработки статистических данных наблюдений какого-либо процесса кривую третьего порядка, симметричную относительно вертикальной прямой. Семейство подобных кривых показано на рис. 23. Все кривые могут быть представлены

уравнением вида $\hat{y} = \frac{1}{a_0 + a_1x + a_2x^2}$.

Дадим вывод уравнения методами наименьших квадратов и центра масс.

Рассмотрим сначала метод наименьших квадратов с использованием для вычислений стандартной программы для ЭВМ. Сделаем вначале

замену $Y = 1/y$, получаем $\hat{Y} = a_0 + a_1x + a_2x^2$. Результаты расчетов теоретических значений \hat{Y} , а по ним величин \hat{y} заносим в табл. 62.

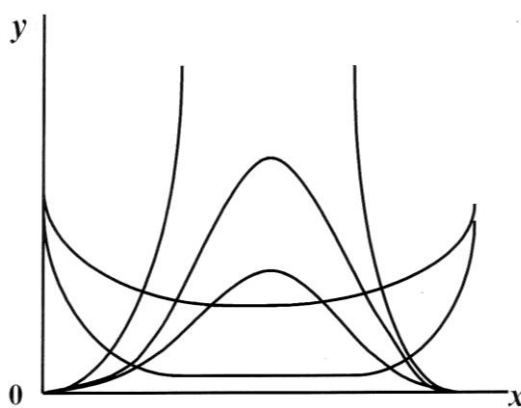


Рис. 23

Таблица 62

Номер опыта	x	y	Y	\hat{Y}	\hat{y}	$(y - \hat{y})^2$	$(y - \bar{y})^2$
1	18	140	0,00714286	0,007848977	127,405	158,63	1190,25
2	20	150	0,00666667	0,005390939	185,496	1259,97	600,25
3	22	260	0,00384615	0,003745092	267,016	49,22	210,25
4	23	300	0,00333333	0,003226740	309,910	98,21	650,25
5	24	390	0,00256410	0,002911436	343,473	2164,76	13340,25
6	25	400	0,00250000	0,002799180	357,247	1827,82	15750,25
7	26	410	0,00243902	0,002889972	346,024	4092,93	18360,25
8	27	360	0,00277778	0,003183812	314,089	2107,82	7310,25
9	28	330	0,00303030	0,003680699	271,687	3400,41	3080,25
10	30	130	0,00769231	0,005283617	189,264	3512,22	20880,25
11	32	150	0,00666667	0,007698726	129,892	404,33	15500,25
<i>Итого</i>	–	3020	–	–	2841,5	19076,29	96872,75

Расчеты на ЭВМ дали значения: $a_0 = 0,0665199$, $a_1 = -0,005086927$, $a_2 = 0,0001015239$. Тогда можно записать $\hat{Y} = 0,0665199 - 0,005086927 x + 0,0001015239 x^2$. После обратного преобразования получили искомую модель

$$\hat{y} = \frac{1}{0,0665199 - 0,005086927 x + 0,0001015239 x^2}.$$

Расчет коэффициента корреляции:

$$\bar{y} = \frac{3020}{11} = 274,5; \quad s_{\hat{y}x}^2 = \frac{\sum (y - \hat{y})}{n - 2} = 2119,59; \quad s_y^2 = \frac{\sum (y - \bar{y})^2}{n} = 8806,61,$$

$$\text{откуда } r_{xy} = \sqrt{1 - \frac{s_{\hat{y}x}^2}{s_y^2}} = \sqrt{0,75932} = 0,8714.$$

Теперь используем метод центра масс. Линеаризацию связи x и y проводим с помощью дробно-линейной функции (см. п. 4.2), приняв в качестве произвольной точки данные для опыта № 7 с максимальным значением $y_1 = 410$ и соответствующим значением $x_1 = 26$. Тогда $Y' = \frac{Y - Y_1}{x - x_1} = \frac{1/y - 1/y_1}{x - x_1}$. Вид модели $\hat{Y}' = Y'_c(1 + b_0 g_x)$, где

$$b_0 = \frac{\sum |\Delta Y'|}{\sum |\Delta x|}; \quad g_x = \frac{x - x_c}{Y_c}; \quad \Delta_x = x - x_c; \quad \Delta Y' = Y' - Y'_c; \quad x_c = \sum x/n;$$

$Y'_c = \sum Y'/n$; $Y = 1/y$. Данные расчетов приведены в табл. 63.

Таблица 63

Номер опыта	Y'	$ \Delta x $	$ \Delta Y' $	g_x	\hat{Y}'	\hat{Y}	\hat{y}
1	-0,00058798	7	0,00064128	-131322,47	-0,00058237	0,0070980	140,9
2	-0,00070461	5	0,00075791	-93801,77	-0,00040075	0,0048947	204,3
3	-0,00035178	3	0,00040508	-56281,06	-0,00021913	0,0033155	301,6
4	-0,00029810	2	0,00035140	-37520,71	-0,00012832	0,0028152	355,2
5	-0,00006254	1	0,00011584	-18760,35	-0,00003751	0,0025140	397,8
6	-0,00006098	0	0,00011428	0	-0,00005330	0,0023857	419,2
7	0	1	0,00005330	18760,35	0,00014412	0,0024390	410
8	0,00033876	2	0,00002855	37520,71	0,00023493	0,0026733	374,1
9	0,00029564	3	0,00002423	56281,06	0,00032574	0,0030905	323,6
10	0,00131332	5	0,00012600	93801,77	0,00050736	0,0044685	223,8
11	0,00070461	7	0,00065130	131322,47	0,00068898	0,0065729	152,1
<i>Итого</i>	0,00058634	36	0,00366920	-	-	-	-

В расчетах использовались значения $x_c = 25$; $Y'_c = 0,000053304$; $b = \sum |\Delta Y'| / \sum |\Delta x| = 0,0000908112$; $\bar{y} = 274,5$. Таким образом, получена модель $\hat{Y}' = 0,000053304(1 + 0,0000908112g_x)$, где $g_x = \frac{x - 25}{0,000053304}$. Из отображающей функции вытекает $\hat{Y} = \hat{Y}'(x - x_1) + Y_1$. Данное выражение использовано для расчетов значений \hat{Y} , занесенных в табл. 63.

При расчете коэффициента корреляции получены данные $s_{\hat{y}x}^2 = 1911,00$; $s_y^2 = 8806,61$. Соответственно вычислено $r_{\hat{y}y} = 0,8849$, что больше коэффициента корреляции, рассчитанного методом наименьших квадратов и означает более высокую точность расчетов.

6.2. Метод построения модели разделением переменных в уравнении на линейную и нелинейную части

Геометрически квадратный трехчлен можно представить схемой, как показано на рис. 24. Кривая 2 пересекает ось y в точке $y_{\min} = a_0$. Касательная 1 к кривой в этой точке наклонена к оси x под углом α , причем $\operatorname{tg} \alpha = a_1$. Для

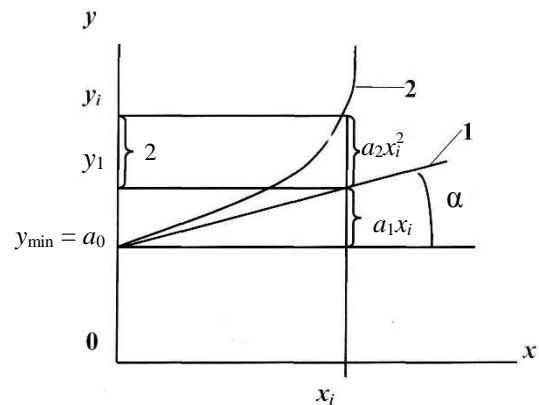


Рис. 24

координаты x_i имеем соотношения $y_i = y_1 + y_2$, где $y_1 = a_0 + a_1x_i$, $y_2 = a_2x_i^2$, т. е. в сумме получаем исходный трехчлен $y_i = a_0 + a_1x_i + a_2x_i^2$. Отсюда следует, что квадратное уравнение сложного вида состоит из уравнений прямой и простой параболы второй степени. Соответственно исходные данные для фактора x возможно разложить на две составляющие, относящиеся к прямой и параболе, т. е. преобразовать однофакторную модель в двухфакторную. При этом независимо строятся две модели: для y_1 и y_2 , которые затем суммируются для определения результирующего признака y . Будем называть y_1 линейной составляющей результирующего признака, а y_2 – нелинейной.

Сделаем важное замечание. Во-первых, уравнение прямой не представляет сложности при выводе. Во-вторых, расстояния между касательной 1 и кривой 2 с ростом x изменяются монотонно, т. е. при получении y_2 нет экстремальной зависимости и преобразования также могут быть простыми (без дробно-линейной функции отображения).

Рассмотрим пример для статистической зависимости y от x , причем зависимость содержит максимум. Подбор теоретической кривой при наличии максимума достаточно трудоемок. Покажем возможность решения более простым изложенным способом. Исходные данные и результаты промежуточных расчетов приводим в табл. 64, 65. Для статистических данных подбираем модель вида $\hat{y} = a_0 + a_1x + a_2x^2$, где $a_0 = y_{\min}$ при $x = 0$, графически находим $a_1 = \operatorname{tg} \alpha = 3,57$.

Таблица 64

Номер опыта	x	y	a_1x	$\hat{y}_1 = y_{\min} + a_1x$	$y_2 = y - y_1$
1	0	188	0	188	0
2	8	215	28,56	216,56	-1,56
3	15	232	53,55	241,55	-9,55
4	21	335	74,97	262,97	-27,97
5	28	233	99,96	287,96	-54,96
6	37	225	132,09	320,09	-95,09
7	44	210	157,08	345,08	-135,08
8	50	190	178,50	366,50	-175,50
9	57	175	203,49	391,49	-216,49
<i>Итого</i>	–	1903	–	–	–

Кривую выравниваем логарифмированием. Так как $x = 0$ не логарифмируется, то далее преобразования проводятся начиная с опыта № 2.

Таблица 65

Номер опыта	$X = \lg x$	$Y_2 = \lg y_2 $	g_x	g_{Y_2}	\hat{Y}_2	\hat{y}_2	$\hat{y} = \hat{y}_1 + \hat{y}_2$
1	–	–	–	–	–	0	188,0
2	0,90309	0,19312	–0,32874	–0,88162	0,30094	–2,0	214,5
3	1,17609	0,98000	–0,16139	–0,39927	0,97820	–9,5	232,0
4	1,32222	1,44669	–0,07182	–0,11319	1,34069	–21,9	241,1
5	1,44716	1,74005	0,00477	0,06664	1,65064	–44,7	243,3
6	1,56820	1,97813	0,07897	0,21258	1,95093	–89,3	230,8
7	1,64345	2,13059	0,12509	0,30604	2,13758	–137,3	207,8
8	1,69897	2,24674	0,15913	0,37724	2,27534	–188,5	178,0
9	1,75587	2,33544	0,19401	0,43161	2,41649	–260,9	130,6
<i>Итого</i>	11,51505	13,05076	1,12392	2,78819	13,05081	–	1866,1

Расчетные значения: $X_c = \frac{\sum X}{n} = 1,43938$; $Y_{c2} = \frac{\sum Y_2}{n} = 1,63134$;
 $b_0 = \frac{\sum |g_{Y_2}|}{\sum |g_x|} = 2,48077$; $g_x = \frac{X - X_c}{Y_c} = \frac{X - 1,43938}{1,63134}$; $g_{Y_2} = \frac{Y_2 - Y_{c2}}{Y_{c2}} = \frac{Y_2 - 1,63134}{1,63134}$.

Таким образом, получена модель в логарифмической форме

$$\hat{Y}_2 = 1,63134(1 + 2,48077g_x); \quad g_x = \frac{X - 1,43938}{1,63134}.$$

Ее расчетные значения \hat{Y}_2 потенцируются в виде $\hat{y}_2 = 10^{\hat{Y}_2}$, после чего определяются $\hat{y} = \hat{y}_1 + \hat{y}_2$.

Проверка модели по относительной ошибке аппроксимации показала результат $\bar{A} = 0,1|\varepsilon/y| 100\% = 4,3\% < 10\%$, т. е. вполне удовлетворительный, итоговые значения в таблицах также указывают на низкую погрешность вычислений и высокую точность модели.

Описанный путь построения сложных однофакторных зависимостей может применяться для всех видов уравнений, которые сводятся к квадратному трехчлену. Однако метод центра масс применим для любого числа включаемых в модель факторов. Покажем это.

Задача с одним фактором, влияющим на результирующий признак по любой закономерности, является двумерной (плоской) и легко представляется графически. Добавление факторов увеличивает мерность задачи и с увеличением факторов по количеству свыше двух требует нового геометрического представления. Рассмотрим такие случаи и дадим вывод формул.

7. ПОСТРОЕНИЕ МНОГОФАКТОРНЫХ МОДЕЛЕЙ

Многофакторная – это многомерная, т. е. пространственная модель, где несколько факторов имеют свою долю влияния на результирующий признак. Причем влияние может быть раздельным и совместным, когда роль отдельного фактора выделить сложно. Факторы при этом могут усиливать или ослаблять влияние друг друга.

7.1. Метод суммирования линеаризованных зависимостей

Представим, что на результирующий признак y действуют два фактора x_1 и x_2 . Прямая, построенная в пространстве x_1x_2y при конкретных численных значениях x_i и y , имеет центр масс C с координатами x_{1c} , x_{2c} и y_c . Связь между одним фактором x и y выражается формулой (5). Но как должна измениться формула при введении второго фактора – требуется оценить.

По условию неизменности центра масс для совокупности точек наблюдения координата y_c находится на одном уровне для любых значений координат центра масс по осям x_1

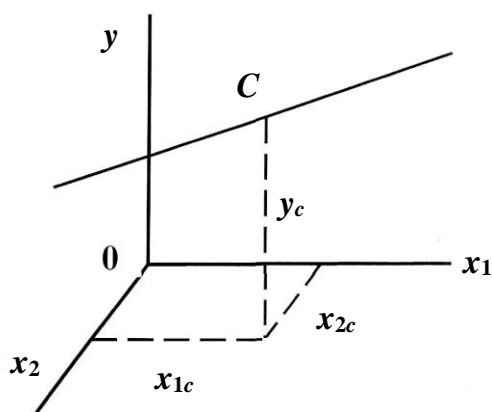


Рис. 25

и x_2 (как показано на рис. 25), т. е. является константой. Значит, влияние дополнительных факторов на положение прямой в пространстве x_1x_2y возможно в пределах, обусловленных только величинами b и g_x . Неизменным также будет положение x_c для одной и той же совокупности точек. Отсюда следует ограничение: прямая вращается вокруг точки C , и если g_x увеличивается, то уменьшается b и наоборот.

Это справедливо для любого числа влияющих факторов, т. е. для любой многомерности пространства. Тогда расчетная формула линейной зависимости имеет вид $y = y_c [1 \pm B_0 (g_{x1} + g_{x2} +$

$$+ \dots + g_{xk} + \dots + g_{xm})], \text{ где } B_0 = \frac{\sum |\Delta_y|}{\sum (|\Delta_{x1}| + |\Delta_{x2}| + \dots + |\Delta_{xk}| + \dots + |\Delta_{xm}|)};$$

$$g_{xk} = \frac{x_{ki} - x_{kc}}{y_c}; \Delta_y = y_i - y_c; \Delta_{xk} = x_{ki} - x_{kc}; x_{kc} = \frac{\sum x_{ki}}{n}; y_c = \frac{\sum y_i}{n}.$$

В преобразованном виде для удобства расчетов

$$y = y_c [1 \pm B(g_{x_1} + g_{x_2} + \dots + g_{x_k} + \dots + g_{x_m})], \quad (6)$$

где $B = \frac{\sum |g_y|}{\sum (|g_{x_1}| + |g_{x_2}| + \dots + |g_{x_k}| + \dots + |g_{x_m}|)}$; $g_{xk} = \frac{x_{ki} - x_{kc}}{x_{kc}}$; $g_y = \frac{y_i - y_c}{y_c}$.

Смысл суммирования g_x арифметически, а не геометрически, как может показаться на первый взгляд, в том, что результирующий фактор u рассматривается под влиянием совокупности m исходных факторов. Фактор u имеет конкретный набор статистически наблюдаемых численных значений, а число влияющих на него m факторов может быть различным, например, в зависимости от интересов исследователя. Некоторые факторы он вообще может проигнорировать. Процентное соотношение влияния на u включаемых в модель m факторов может меняться, но это никаким образом не отражается на уже полученных данных об u . Если, к примеру, x_2 уменьшает величину u , то x_1 должен увеличить u на ту же величину.

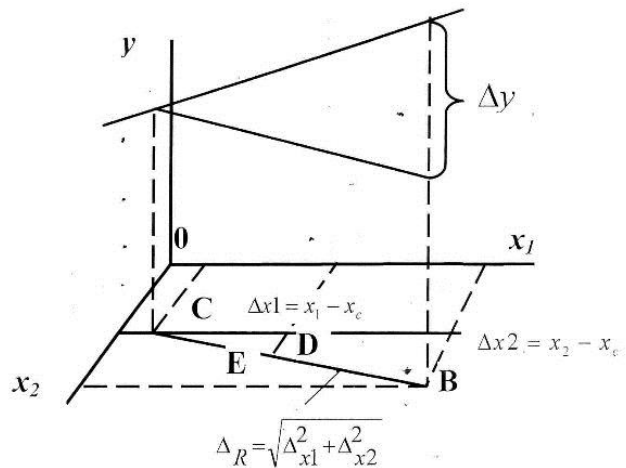


Рис. 26

При арифметическом суммировании имеем для двух факторов

$$y = y_c + \sum |\Delta_y| \frac{(\Delta_{x_1} + \Delta_{x_2})}{\sum (|\Delta_{x_1}| + |\Delta_{x_2}|)}.$$

Для геометрического суммирования формула приобретает вид $y = y_c + \sum |\Delta_y| \frac{\sqrt{\Delta_{x_1}^2 + \Delta_{x_2}^2}}{\sum \sqrt{\Delta_{x_1}^2 + \Delta_{x_2}^2}}$. Наглядно компоненты формул представлены на рис. 26.

Обозначим $\frac{\Delta_{x_1} + \Delta_{x_2}}{\sum (|\Delta_{x_1}| + |\Delta_{x_2}|)} = \frac{\Delta_x}{\sum \Delta_x}$, $\frac{\sqrt{\Delta_{x_1}^2 + \Delta_{x_2}^2}}{\sum \sqrt{\Delta_{x_1}^2 + \Delta_{x_2}^2}} = \frac{\Delta_R}{\sum \Delta_R}$. Фак-

тически надо убедиться в том, что $\Delta_x / \sum \Delta_x = \Delta_R / \sum \Delta_R$. Геометрическое суммирование означает действия с отрезками треугольника

ABC по гипотенузе, а арифметическое – по катетам. Подобие прямоугольных треугольников на плоскости x_1Ox_2 дает равенство отношений $CE/CB = CD/CA = DE/AB$ или $CE/CB = (CD + DE)/(CA + AB)$, т. е. справедливо равенство $\Delta_R/\sum\Delta_R = \Delta_x/\sum\Delta_x$. Левая часть равенства относится к геометрическому суммированию, а правая – к арифметическому.

Получается, что суммирование может быть любым, так как не отражается на отношениях величин, но арифметическое сложение проще, поэтому ему и отдается предпочтение.

Если имеется третий фактор x_3 , то добавляется к оси $x_1 + x_2$, проходящей через линию CB , еще одна ось координат (рис. 27). Отсюда

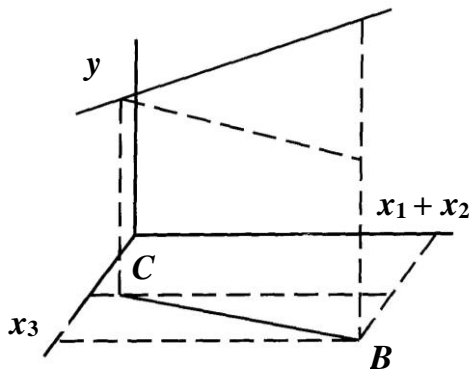


Рис. 27

следует, что в выражение для B_0 войдет $|\Delta_{x3}|$, а в формулу – параметр g_{x3} . Так можно добавлять оси координат до любого числа переменных x . Таким образом, мы убеждаемся в справедливости общей формулы (6).

Например, для двухфакторной модели формула приобретает вид

$$y = y_c [1 \pm B_0 (g_{x1} + g_{x2})], \quad (7)$$

где $B = \frac{\sum |\Delta_y|}{\sum (|\Delta_{x1}| + |\Delta_{x2}|)}$; $g_{x1} = \frac{x_1 - x_{1c}}{y_c}$; $g_{x2} = \frac{x_2 - x_{2c}}{y_c}$; $\Delta_{x1} = x_1 - x_{1c}$;

$$\Delta_{x2} = x_2 - x_{2c}; \Delta_y = y - y_c; x_{ic} = \frac{\sum x_i}{n}; y_c = \frac{\sum y}{n}.$$

То же, но в другом виде

$$y = y_c [1 \pm B (g_{x1} + g_{x2})],$$

где $B = \frac{\sum g_y}{\sum (|g_{x1}| + |g_{x2}|)}$; $g_{x1} = \frac{x_1 - x_{1c}}{x_{1c}}$; $g_{x2} = \frac{x_2 - x_{2c}}{x_{2c}}$; $g_y = \frac{y - y_c}{y_c}$; $x_{1c} = \frac{\sum x_1}{n}$;

$$x_{2c} = \frac{\sum x_2}{n}; y_c = \frac{\sum y}{n}.$$

Рассмотрим пример построения линейной двухфакторной модели. Исходные данные и промежуточные результаты расчетов занесены в табл. 66.

Таблица 66

Номер опыта	x_1	x_2	y	g_{x_1}	g_{x_2}	g_y	\hat{y}	ε	$ \varepsilon/y $
1	108	0,19	36	-0,37371	-0,76914	-0,64	50,1	14,1	0,392
2	123	0,34	67	-0,28672	-0,58688	-0,33	61,9	-5,1	0,076
3	142	0,53	78	-0,17654	-0,35601	-0,22	76,7	-1,3	0,017
4	153	0,64	93	-0,11276	-0,22236	-0,07	85,4	-7,6	0,082
5	176	0,85	106	0,02062	0,03281	0,06	102,3	-3,7	0,035
6	192	1,01	112	0,11340	0,22722	0,12	114,9	2,9	0,026
7	198	1,07	127	0,14820	0,30012	0,27	119,6	-7,4	0,058
8	221	1,30	133	0,28158	0,57959	0,33	137,6	4,6	0,035
9	239	1,48	148	0,38596	0,79830	0,48	151,7	3,7	0,025
<i>Итого</i>	1552	7,41	900	1,89949	3,87243	2,52	900,2	0,2	0,746

Расчет постоянных:

$$x_{1c} = \frac{1552}{9} = 172,4; \quad x_{2c} = \frac{7,41}{9} = 0,823; \quad y_c = \frac{900}{9} = 100;$$

$$B = \frac{2,52}{1,89949 + 3,87243} = 0,4366; \quad g_y = \frac{y - 100}{100}.$$

Получена линейная двухфакторная модель вида $\hat{y} = 100[1 + 0,4366(g_{x_1} + g_{x_2})]$; $g_{x_1} = \frac{x_1 - 172,4}{172,4}$; $g_{x_2} = \frac{x_2 - 0,823}{0,823}$. При значительном разбросе исходных данных относительная ошибка аппроксимации оказалась удовлетворительной $\bar{A} = \frac{0,746}{9} 100 \% = 8,29 < 10 \%$.

Следует заметить, что, если связь между переменными в исходных данных линейна, то линеаризацию переменных необходимо проводить одинаковым способом (или логарифмированием, или с помощью замен, или посредством отображающих функций). Если применять к разным переменным различные способы выравнивания, то для каждой из них будет определен центр масс, отличающийся от реального центра масс системы из-за различия шкал по масштабу и равномерности для каждой из осей координат.

7.2. Сведение пространственной задачи к плоской

Метод центра масс можно модернизировать для двухфакторных процессов, сократив объем расчетов (одного из g_x и комплексной величины B). При этом одному из факторов придают роль массы. Таким фактором может быть составляющая регрессионной модели, имеющая наивысший порядок и требующая значительных вычислений (обычно возведения в степень) в сравнении с другими составляющими, хотя это не обязательно и в принципе возможно брать любую из составляющих или переменных многофакторной модели.

Подобный прием также обеспечивает сокращение модели на один порядок: например, двухфакторная модель сводится к однофакторной или модель второго порядка сводится к регрессии первого порядка, что упрощает использование подобной модели в расчетах. Кроме того, сведение пространственной задачи к плоской позволяет проверить качество решения графически, использовать графики для оценки и прогноза.

Расчетные формулы определения координат центра масс линеаризованной фигуры множества точек наблюдения для модели второго порядка имеют вид $x_c = \frac{\sum x_1 x_2}{\sum x_2}$; $y_c = \frac{\sum y x_2}{\sum x_2}$, где x_2 выполняет функцию массы. В результате двухфакторная модель с факторами x_1 и x_2 сведена к однофакторной с фактором x_1 в виде $y = y_c(1 \pm b g_{x_1})$, где $b = \frac{\sum |g_y|}{\sum |g_x|}$; $g_{x_1} = \frac{x_1 - x_c}{x_c}$; $g_y = \frac{y - y_c}{y_c}$. Знак «минус» берется для обратно пропорциональной зависимости y от x_1 .

Так как центр масс определен в предложенном решении не как среднеарифметическая величина, то он смещен от середины прямой аппроксимации в ту или иную сторону в зависимости от направления приращения составляющей, принятой в качестве массы.

Рассмотрим пример сведения пространственной задачи к плоской для линейной связи различного направления факторов x_1 и x_2 с y по условиям, приведенным в табл. 67. Там же даны результаты промежуточных вычислений. Примем в качестве выполняющего функцию массы фактор x_1 .

Таблица 67

Номер опыта	x_1	x_2	y	x_1x_2	yx_1	g_{x2}	g_y	bg_{x2}	\hat{y}
1	17	45	3080	765	52360	1,3931	0,6026	0,6463	3164,0
2	32	38	2970	1216	95040	1,0209	0,5454	0,4736	2832,1
3	55	33	2838	1815	156090	0,7550	0,4767	0,3502	2594,9
4	74	29	2464	2146	182336	0,5426	0,2821	0,2517	2405,6
5	97	30	2354	2910	228338	0,5957	0,2248	0,2763	2452,9
6	120	25	2024	3000	242880	0,3298	0,0531	0,1530	2215,9
7	125	18	1980	2250	247500	-0,0426	0,0302	-0,0198	1883,8
8	148	14	1925	2072	284900	-0,2553	0,0016	-0,1184	1694,3
9	174	17	1782	2958	310068	-0,0957	-0,0728	-0,0444	1836,5
10	295	15	1650	4425	486750	-0,2021	-0,1415	-0,0938	1741,6
11	320	12	1606	3840	513920	-0,3617	-0,1644	-0,1678	1599,4
<i>Итого</i>	1457	276	24673	27397	2800182	5,5945	2,5952	–	24421

Расчет постоянных: $x_c = \frac{\sum x_1x_2}{\sum x_1} = \frac{27397}{1457} = 18,8037$; $y_c = \frac{\sum yx_1}{\sum x_1} = \frac{2800182}{1457} = 1921,88$; $b = \frac{\sum |g_y|}{\sum |g_{x2}|} = \frac{2,5952}{5,5945} = 0,4639$. Получена модель вида $\hat{y} = y_c(1 + bg_{x2})$; $g_{x2} = \frac{x_2 - x_c}{x_c}$; $g_y = \frac{y - y_c}{y_c}$. Модель в численном выражении $\hat{y} = 1921,88(1 + 0,4639g_{x2})$; $g_{x2} = \frac{x_2 - 18,8}{18,8}$.

Расчетные значения \hat{y} оказались близкими к фактическим и укладываются на прямую в зависимости от x_2 .

7.3. Построение многофакторных моделей при минимально допустимом количестве опытов

Часто решение задач связано с дефицитом ресурсов для построения модели на основе экспериментов: недостаточно число объектов наблюдения, ограниченность во времени проведения опытов, малые возможности по затратам средств и т. д. В таких условиях надо минимально необходимыми возможностями обеспечить получение модели процесса. При этом желательно учесть взаимовлияние факторов, когда каждый из них или некоторые воздействуют на характер зависимостей

остальных факторов и результирующего признака. Прежде чем описывать примеры решения подобных задач методом центра масс, сделаем некоторые важные замечания об особенностях построения многочленов.

Во-первых, свободный член в модели является центром массы для результирующего признака.

Во-вторых, факторные признаки дают приращение к свободному члену как это показано, например, на рис. 24, образуя таким образом в совокупности с ним многочлен.

В-третьих, для простоты решения исходные данные должны быть заранее линеаризованы, обеспечивая получение многочлена первой степени.

В-четвертых, как показывает практика, взаимовлияние факторов легко учитывается путем перемножения взаимодействующих составляющих и введения произведения в состав многочлена [9].

В-пятых, если исходные данные линеаризованы, то нет необходимости учитывать и обрабатывать все их значения, достаточно иметь всего лишь два из них по концам прямой после линеаризации, а отсюда следует, что для построения этой прямой необходимо и достаточно провести всего два опыта с крайними значениями фактора.

В-шестых, опыты для минимизации их общего количества должны проводиться при неповторяющихся сочетаниях задаваемых факторов, а отсюда минимально необходимое общее число опытов составляет $n = 2^k$, где 2 – число задаваемых значений (уровней) фактора, k – количество факторов, учитываемых в модели. Для повышения достоверности результата опыты при тех же сочетаниях исходных данных могут дублироваться многократно.

Дадим описание процесса построения двухфакторной модели с учетом взаимовлияния факторов. Для сокращения расчетов условимся, что связь факторов с результирующим признаком линейная, поэтому факторы заданы на двух уровнях. Порядок опытов определен так, чтобы обеспечить наглядность построения программы их осуществления. Проведение опытов требует больших затрат времени и потери финансовых ресурсов, например, в виде упущенной выгоды от потерь прибыли. Пусть исследуется прибыль y предприятия в зависимости от цены продукции x_1 и числа оборотов капитала x_2 за период. Известно, что рост числа оборотов капитала позволяет в конечном счете снижать цену, а изменение цены, воздействуя на спрос, отражается на скорости

обращения капитала, т. е. между x_1 и x_2 существует взаимовлияние, отражающееся на прибыли y . Минимально необходимое количество опытов составляет для двух факторов, изменяемых на двух уровнях, $n = 2^2 = 4$. Условия и результаты опытов приведены в табл. 68 в условных единицах.

Строится регрессионная линейная модель вида $y = a_0 + a_1x_1 + a_2x_2 + a_{12}x_1x_2$, где a_0 – свободный член; a_1 и a_2 – коэффициенты влияния факторов x_1 и x_2 на результирующий признак y ; a_{12} – коэффициент взаимовлияния факторов.

Таблица 68

Номер опыта	x_1	x_2	y	\hat{y}
1	1	12	785	785,3
2	1,5	12	1305	1305,2
3	1	18,5	1280	1280,7
4	1,5	18,5	1990	1990,7
<i>Итого</i>	–	–	5360	5361,9

Коэффициенты a_1 и a_2 представляют собой тангенсы углов наклона проекций прямой к осям x_1 и x_2 (в пространстве x_1x_2y) на плоскости x_1oy

и x_2oy . Их величина соответственно $a_1 = \frac{\sum(\Delta y_j)/n}{\Delta x_1}$; $a_2 = \frac{\sum(\Delta y_j)/n}{\Delta x_2}$, где

$\Delta y_j = y_{ji} - y_{jc}$; $\Delta x_j = x_{jc} - x_{j\min}$; i – номер опыта; j – номер фактора.

Очевидно, что $a_0 = \sum y/n$, для каждого опыта $\Delta x_{ji} = x_{ji} - x_{jc}$. Тогда общий вид расчетной формулы

$$\hat{y} = \frac{\sum y}{n} + \frac{\sum(\Delta y_j)/n}{x_{1c} - x_{1\min}}(x_{1i} - x_{1c}) + \frac{\sum(\Delta y_j)/n}{x_{2c} - x_{2\min}}(x_{2i} - x_{2c}) + \frac{\sum(\Delta y_j)/n}{(x_{1c} - x_{1\min})(x_{2c} - x_{2\min})}(x_{1i} - x_{1c})(x_{2i} - x_{2c}).$$

В случае прямо пропорциональной зависимости x_j и y их экстремальные значения соответствуют друг другу, а разности с величинами центра масс имеют одинаковый знак. Поэтому для фактора x_1 , имеющего два одинаковых значения $x_{1c} = (1 + 1,5)/2 = 1,25$ между первым и вторым, третьим и четвертым опытами, разности с ним данных первого и третьего опытов будут отрицательны ($1 - 1,25 = -0,25$), а второго и четвертого – положительны ($1,5 - 1,25 = 0,25$). Аналогичные знаки будут иметь и значения y , откуда $\sum(\Delta y_1) = -785 + 1305 - 1280 + 1990 = 1230$. Для фактора x_2 имеется одно значение $x_{2c} = (12 + 12 + 18,5 + 18,5)/4 = 15,25$, соответственно с ним отрицательными будут величины y для первого и второго опытов и положительными – для третьего и четвертого, тогда

$\sum(\Delta y_2) = -785 - 1305 + 1280 + 1990 = 1180$. Следует заметить, что при суммировании Δy_j не включены значения Δy_{jc} , так как они, имея противоположные знаки, взаимно уничтожились.

В последней составляющей расчетной формулы берутся произведения разностей Δx_{ji} с учетом знаков, эти же знаки произведений имеют и величины y , поэтому получаем $\sum(\Delta y_3) = 785 - 1305 - 1280 + 1990 = 190$.

Итак, расчетная формула в численном выражении

$$\hat{y} = \frac{5360}{4} + \frac{1230/4}{1,25-1}(x_1 - 1,25) + \frac{1180/4}{15,25-12}(x_2 - 15,25) + \frac{190/4}{(1,25-1)(15,25-12)}(x_1 - 1,25)(x_2 - 15,25).$$

После преобразований и приведения подобных членов получаем окончательный вид формулы (искомой линейной модели процесса)

$$\hat{y} = -467 + 337,9x_1 + 17,7x_2 + 58,5x_1x_2.$$

Расчеты по этой формуле для всех опытов дали теоретические значения результирующего признака, приведенные в табл. 68, которые практически совпали с опытными величинами. Заметим, что при подобном построении программы опытов и с учетом взаимовлияния факторов ни одна из рассмотренных предыдущих методик расчетов непригодна.

8. ПОСТРОЕНИЕ МОДЕЛЕЙ ДЛЯ ДЕТЕРМИНАНТНЫХ ПРОЦЕССОВ

В экономике многие процессы однозначно обусловлены и не имеют статистической природы. Но следует понимать, что в любой экономической системе практически всегда сочетаются стохастические и детерминантные явления. Например, накопление капитала в процессе его обращения осуществляется на основе рыночных отношений спроса и предложения (преобладают вероятностные процессы) и при инвестировании под определенный стабильный процент с гарантией обеспечения (случайность колебаний капитала, можно сказать, исключена). Поэтому наряду с моделями, формируемыми с помощью

статистических данных, должны развиваться и применяться в экономической практике модели, исключая неопределенности, т. е. детерминантные. Рассмотрим некоторые из примеров построения подобных моделей.

8.1. Решение задачи о многофункциональном наращивании капитала

В массовых обращениях капитала, когда происходит одновременное непрерывное изменение величин добавленного капитала, процентной ставки и числа оборотов капитала, применение только известной формулы сложных процентов недостаточно и необходим переход к многомерной модели. Приведем один из вариантов ее построения в общем виде и некоторые ее частные случаи.

Введем обозначения: k_0 – исходный капитал, k – добавленный капитал, p – процентная ставка, n – число оборотов капитала в единицу времени, t – период времени. Пусть на начальный период t_0 капитал имел размер k_0 и затем наращивался на величину Δk_i за счет вкладов, возврата кредитов и т. п. При этом процентная ставка при инвестировании капитала возрастала на величину Δp_i в период времени t_i . Обращение капитала также ускоряется, например, вследствие совершенствования системы бизнеса и менеджмента на величину Δn_i . Можно считать возможным вариант роста каждого параметра по принципу «капитализации», когда прирост присоединяется к основной массе параметра и дальнейший прирост обуславливается этой суммой.

За период $t_0 \dots t_i$ капитал k_0 возрастет и составит $k_0 + \Delta k_1$ или $k_0(1 + \Delta k_1/k_0)$. Аналогично увеличатся процентная ставка в $(1 + \Delta p_1/p)$ раз и прирост капитала за счет роста числа его оборотов в $(1 + \Delta n_1/n)$ раз. Обозначим $\Delta k/k = \eta$, $\Delta p/p = \vartheta$, $\Delta n/n = \psi$. Тогда величина капитала k составит:

на период $t_0 - k_0$;

$$t_1 - k_0(1 + \eta_1)(1 + \vartheta_1)(1 + \psi_1) = k';$$

$$t_2 - k'(1 + \eta_2)(1 + \vartheta_2)(1 + \psi_2) = k'';$$

$$t_3 - k''(1 + \eta_3)(1 + \vartheta_3)(1 + \psi_3) = k'''; \text{ если вместо } k'' \text{ подставить}$$

предыдущие выражения, то получим

$$k = k_0(1 + \eta_1)(1 + \vartheta_1)(1 + \psi_1)(1 + \eta_2)(1 + \vartheta_2)(1 + \psi_2)(1 + \eta_3)(1 + \vartheta_3)(1 + \psi_3)$$

или, введя знак произведения \prod , имеем $k = k_0 \prod_{t=1}^{t=3} (1 + \eta_t) \prod_{t=1}^{t=3} (1 + \vartheta_t) \prod_{t=1}^{t=3} (1 + \psi_t)$.

В общем случае для периода $t = 1 \dots N$ формула примет вид

$$k = k_0 \prod_{t=1}^N (1 + \eta_t) \prod_{t=1}^N (1 + \vartheta_t) \prod_{t=1}^N (1 + \psi_t).$$

Если отсчет времени начать с t_0 , когда, $\eta, \vartheta, \psi = 0$, то приходим к

выражению $k = k_0 \prod_0^t (1 + \eta_t) \prod_0^t (1 + \vartheta_t) \prod_0^t (1 + \psi_t)$.

Частные случаи для полученной модели:

1. $\eta, \vartheta, \psi = \text{const}$, тогда $k = k_0(1 + \eta)^t (1 + \vartheta)^t (1 + \psi)^t$, т. е. пришли к трехмерной формуле сложных процентов.

2. η, ϑ, ψ изменяются по арифметической прогрессии, тогда $k = k_0(1 + \eta_t)(1 + \vartheta_t)(1 + \psi_t)$ или в выражении через гамма-функцию $k = k_0 \Gamma(2 + \eta_t) \Gamma(2 + \vartheta_t) \Gamma(2 + \psi_t)$.

3. η, ϑ, ψ изменяются непрерывно в функции времени

$$k = k_0 \prod_0^t [1 + f_t(\eta)/k_0] \prod_0^t [1 + f_t(\vartheta)/p] \prod_0^t [1 + f_t(\psi)/n].$$

4. Без учета «капитализации», если известны зависимости изменения параметров во времени, имеем $k = k_0 \int_0^t f_t(\eta) dt \int_0^t f_t(\vartheta) dt \int_0^t f_t(\psi) dt$.

5. Без учета «капитализации», если $\eta, \vartheta, \psi = \text{const}$: $k = k_0(1 + \eta + \vartheta + \psi)t$.

6. Все параметры функционально взаимосвязаны

$$k = k_0 \int_k f(\eta, \vartheta, \psi) dk \quad \text{или} \quad k = k_0 \iiint f(\eta, \vartheta, \psi) d\eta d\vartheta d\psi.$$

7. Сочетание различных вариантов частных случаев, например, по пп. 1, 2, 3

$$k = k_0(1 + \eta)^t (1 + \vartheta_t) \prod_0^t [1 + f_t(\psi)/n].$$

Применяя предложенный расчет, можно дифференцированно подходить к оценке составных частей капитала, не нарушая структуры

формул, а лишь прибавляя к ним соответствующие части, например, разбив k_0 на составляющие k_1, k_2, k_3 и т. д. При этом решение находится с помощью одной формулы за один шаг.

8.2. Оборачиваемость капитала с равномерно нарастающей массой

Для решения задачи применим гипотезу о возможности приложения законов механики к экономическим системам, что позволяет использовать известный совершенный математический аппарат динамических сложных систем. Введем понятия, связанные с движением капитала.

Капитал – это масса, измеряемая в денежных единицах. Подкрепленный хорошей организацией управления, деловыми качествами, перспективной концептуальной идеей, капитал имеет высокий потенциал, а отданный в потребление без перспектив использования в деле – нулевой потенциал. Путь движения капитала – это число его оборотов в процессе воспроизводства. Путь в единицу времени дает скорость оборачиваемости, приращение скорости в единицу времени есть ускорение оборачиваемости. В отличие от механики, где масса при движении с обычными скоростями неизменна, в экономике масса капитала меняется в зависимости от прибыльности бизнеса. Сила капитала обусловлена его движением с ускорением оборачиваемости и может рассматриваться как произведение массы капитала на данное ускорение. При этом масса капитала обладает инерцией – свойством сопротивляться изменению скорости или состоянию покоя.

Пусть первоначальный капитал равен единице и скорость его оборачиваемости в начальный момент равна нулю. Оборот капитала завершается после приращения к нему некоторой массы m_0 . Будем считать, что приращение капитала способствует росту его оборачиваемости за счет организационных мероприятий (роста числа сотрудников фирмы, посредников, совершенствования менеджмента и прочего) и увеличения занимаемой доли рынка. Такое можно допустить до опре-

деленных размеров капитала, пока инерционные процессы уже на макроэкономическом уровне не станут существенными. Иначе зависимость становится обратной: чем больше масса капитала, тем меньше его оборачиваемость.

Определим скорость оборачиваемости капитала v в любой момент времени t в зависимости от числа оборотов капитала и ускорения оборачиваемости. Обозначим через s общее число оборотов капитала в рассматриваемый момент времени, а через m массу капитала в данный момент. Приращение капитала за время dt обозначим dm . В соответствии с понятием силы капитала F и учитывая переменность массы m и скорости оборачиваемости капитала v , приняв равной нулю скорость оборачиваемости присоединяющихся масс dm , имеем

$$\frac{d}{dt}(mv) = F \text{ или } (mv)' = F. \quad (8)$$

Очевидно, что $m = m_0s$, $v = s'$. Если задать стоимостное выражение ρ_0 одного оборота капитала (удельная сила) и ускорение оборачиваемости a , то можно записать $m_0 = \rho_0/a$, тогда $m = (\rho_0/a)s$. Из последнего имеем $F = \rho_0s$. После подстановки в уравнение (8) получим $\frac{d}{dt}\left(\frac{\rho_0}{a}ss'\right) = \rho_0s$. Сократим постоянную ρ_0 , а ускорение перенесем в правую часть. В результате $\frac{d}{dt}(ss') = ss'' + s'^2$. Следовательно,

$$ss'' + s'^2 = as. \quad (9)$$

Для решения данного уравнения введем функцию $u = s'^2$, а s примем в качестве независимой переменной. Несложно получить $s'' = \frac{1}{2} \frac{du}{ds}$.

После подстановок в уравнение (9) приходим к выражению

$$\frac{1}{2}s \frac{du}{ds} + u = as, \text{ т. е. } \frac{du}{ds} + \frac{2}{s}u = 2a. \quad (10)$$

Решение данного уравнения [10] приводит к выражению $u = \frac{2}{3}a \frac{s^3 - 1}{s^2}$, откуда с учетом $u = s'^2$ следует $s' = \sqrt{\frac{2}{3}a \frac{\sqrt{s^3 - 1}}{s}}$. Оконча-

тельный результат $v = \sqrt{\frac{2}{3}a \frac{\sqrt{s^3 - 1}}{s}}$. Если ввести силу воздействия на капитал, то получим $v = \sqrt{\frac{2F}{3m} \frac{\sqrt{s^3 - 1}}{s}}$.

В расчетах можно вводить временной параметр t , используя уравнения равнопеременного движения капитала: $s = s_0 + v_0t + at^2/2$; $v = v_0 + at$, где s_0 и v_0 – начальные значения числа оборотов и скорости оборачиваемости капитала. Из уравнений следует число оборотов капитала в виде $s - s_0 = \frac{v^2 - v_0^2}{2a}$. При неравномерном движении, если известно изменение ускорения на отрезке $s - s_0$ в виде функции $a = f(s)$, среднее ускорение на данном отрезке будет $\bar{a} = \int_{s_0}^s f(s)/(s - s_0)$. Его и следует подставлять в предыдущие формулы.

8.3. Задача о выравнивании капиталов

Пусть имеются две возможности получить инвестиции, каждая из которых реализуется со своей скоростью приращения капитала. Объемы инвестиций различны, их разница составляет Δk в пользу второй возможности. Скорости приращения капиталов v_1 и v_2 , причем $v_2 > v_1$. Определим время T , по истечении которого размеры капиталов сравняются.

Величина первого капитала через некоторое время t будет $k_1 = v_1t$, а второго – $k_2 = v_2t + \Delta k$. Выравнивание капиталов произойдет в момент времени T , когда $k_1 = k_2$, т. е. $v_1T = v_2T + \Delta k$. Из этого соотношения получаем $T = \Delta k/(v_1 - v_2)$. Если наращивание капитала начнется неодновременно, а например, для второй возможности произойдет с опозданием по сравнению с первой на промежуток времени τ , то $k_2 = v_2(t - \tau) + \Delta k$. Тогда $v_1T = v_2(T - \tau) + \Delta k$, что дает в итоге $T = (\Delta k - v_2\tau)/(v_1 - v_2)$.

Данная задача близка по смыслу к задаче о точке безубыточности при выравнивании доходов и затрат, но для усложненных исходных условий в соответствии с рассмотренным примером.

8.4. Задача об операциях между банком и его филиалами

Отношения между банками можно представить с помощью графовой модели, состоящей из вершин и ориентированных по направлениям движения активов безмасштабных стрелок (рис. 28).

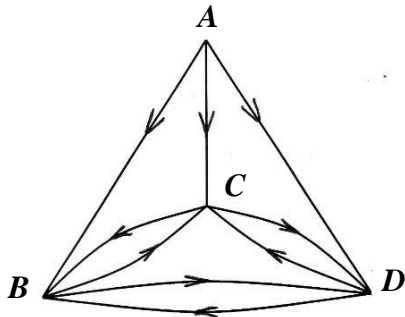


Рис. 28

Головной банк *A* имеет филиалы: инвестиционный банк *B*, валютный банк *C* и сбербанк *D* и осуществляет с ними кредитные операции соответственно под инвестиции, операции с ценными бумагами и депозитные операции. Филиалы между собой также находятся в деловых отношениях. Так, банк *B* инвестирует развитие банков *C* и *D*, но сам пользуется услугами по валютным и депозитным операциям, банк *C* пользуется услугами сбербанка *D*, а тот, в свою очередь, выполняет депозитные валютные операции с помощью банка *C*.

Для исследования системы поставим задачу определить число возможных путей движения активов из головного банка и соответственно схем контроля деятельности филиалов по неповторяющимся вариантам однократного обхода сразу всех филиалов в каждой попытке.

При решении используем задачу из теории графов о прадеревьях [11]. Составим для графа матрицу смежности, учитывая, что движение по стрелке обозначается 1, против стрелки и перемещение внутри вершины графа – 0:

$$M_c = \begin{array}{c|cccc} & A & B & C & D \\ \hline A & 0 & 1 & 1 & 1 \\ B & 0 & 0 & 1 & 1 \\ C & 0 & 1 & 0 & 1 \\ D & 0 & 1 & 1 & 0 \\ \hline \Sigma & 0 & 3 & 3 & 3 \end{array} .$$

Под матрицей даны суммы ее элементов по столбцам, которые заносим в диагональную матрицу

$$M_d = \begin{array}{c|cccc} & 0 & 0 & 0 & 0 \\ \hline 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{array} .$$

Определим разность матриц

$$M_d - M_c = \begin{vmatrix} 0 & -1 & -1 & -1 \\ 0 & 3 & -1 & -1 \\ 0 & -1 & 3 & -1 \\ 0 & -1 & -1 & 3 \end{vmatrix}.$$

Число путей движения активов и схем контроля, т. е. прадеревьев, находится при решении минора элемента B_{11} , выделенного пунктиром:

$$\Delta = 3 \begin{vmatrix} 3 & -1 \\ -1 & 3 \end{vmatrix} + 1 \begin{vmatrix} -1 & -1 \\ -1 & 3 \end{vmatrix} - 1 \begin{vmatrix} -1 & 3 \\ -1 & -1 \end{vmatrix} = 3(9 - 1) + 1(-3 - 1) - 1(1 + 3) = 16.$$

Все варианты обхода показаны на рис. 29.

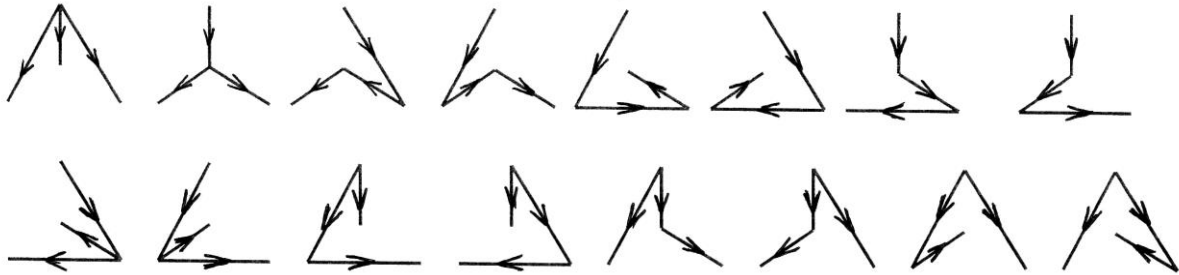


Рис. 29

Если для графа заданы веса, т. е. объемы сделок между банком и филиалами, то можно определить общий объем сделок в банковской системе. Например, укажем веса, означающие объем сделок в условных единицах (рис. 30).

Матрица смежности для графа с учетом направлений стрелок из вершин

$$M_c = \begin{vmatrix} & A & B & C & D \\ A & 0 & 5 & 3 & 6 \\ B & 0 & 0 & 7 & 2 \\ C & 0 & 2 & 0 & 4 \\ D & 0 & 3 & 1 & 0 \\ \hline \Sigma & 0 & 10 & 11 & 12 \end{vmatrix}.$$

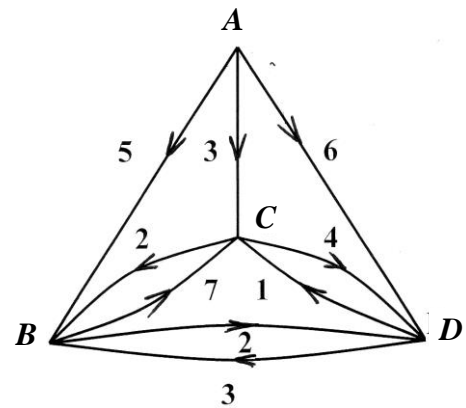


Рис. 30

Диагональная матрица и матрица разности

$$M_{\partial} = \begin{vmatrix} 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 11 & 0 \\ 0 & 0 & 0 & 12 \end{vmatrix}; \quad M_{\partial} - M_c = \begin{vmatrix} 0 & -5 & -3 & -6 \\ 0 & 10 & -7 & -2 \\ 0 & -2 & 11 & -4 \\ 0 & -3 & -1 & 12 \end{vmatrix}.$$

Минор элемента B_{11} , т. е. искомый объем сделок в системе,

$$\Delta = \begin{vmatrix} 10 & -7 & -2 \\ -2 & 11 & -4 \\ -3 & -1 & 12 \end{vmatrix} = 10[11 \cdot 12 - (-4)(-1)] - 7[(-2)(12) - (-4)(-3)] +$$

$$+ (-2)[(-2)(-1) - 11(-3)] = 958.$$

Таким образом, количество возможных путей движения активов из головного банка и схем контроля деятельности филиалов составляет величину, равную 16, а общий объем сделок в системе «головной банк – филиалы» равен 958 условным единицам.

8.5. Задача об управлении портфелями ценных бумаг

Рассмотрим процесс формирования портфелей ценных бумаг для управления ими в крупной холдинговой компании по схеме, показанной на рис. 31. Фирмы a, b, c, d , владеющие ценными бумагами, вкладывают их в банки e, f, g . Те формируют корзины ценных бумаг и отдают их в управление финансово-промышленными группами (ФПГ) i, k . Из ФПГ сформирована холдинг-компания, которая имеет в своем составе чистый холдинг, занимающийся только управлением (покупка – продажа, размещение и т. п.) ценными бумагами. Каждая ФПГ создает свой блок корзин ценных бумаг и передает его в чистый холдинг. Все отношения по передаче ценных бумаг между субъектами холдинг-компания и чистого холдинга predeterminedены весами дуг в условных единицах. Так, в банк e входят x_1 акций фирмы a и x_4 облигаций фирмы b . Из них сформированы y_1 корзин ценных бумаг, направляемых в ФПГ под номером k . В чистый холдинг входят два блока корзин z_1 и z_2 ценных бумаг.

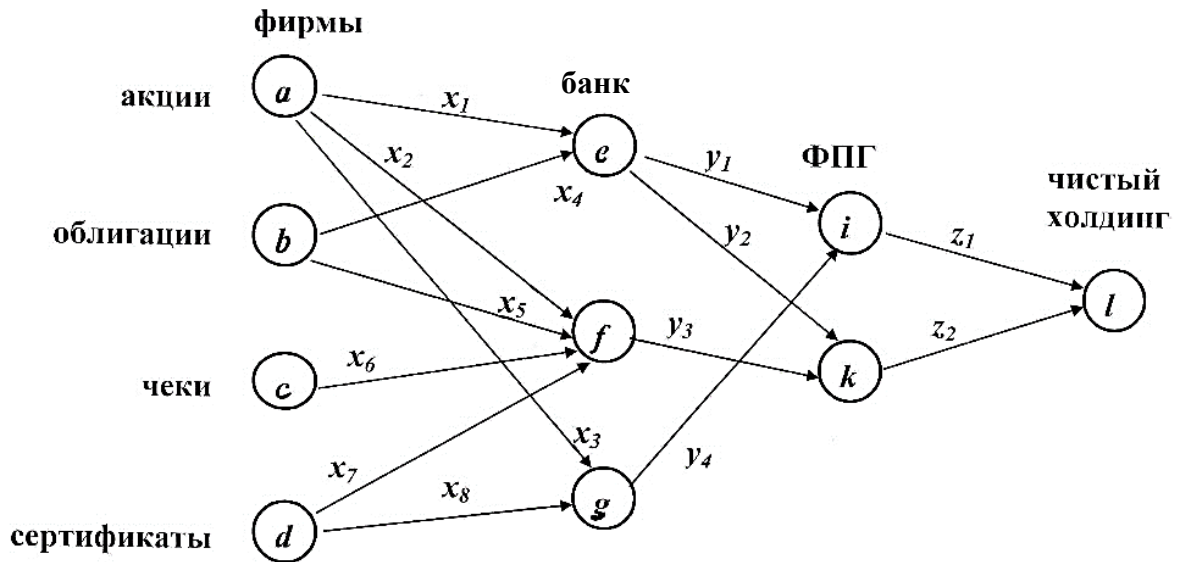


Рис. 31

Требуется определить количество входящих в чистый холдинг ценных бумаг каждого вида.

Решение задачи сводится к построению матрицы смежности и последующему ее умножению самой на себя столько раз, сколько в схеме предусмотрено число шагов на всем маршруте движения ценных бумаг. В нашем случае количество шагов равно трем: от фирм к банкам, от банков к ФПГ и от ФПГ к чистому холдингу.

Матрица смежности имеет вид

$$M_c = \begin{matrix} & \begin{matrix} a & b & c & d & e & f & g & i & k & l \end{matrix} \\ \begin{matrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ i \\ k \\ l \end{matrix} & \begin{vmatrix} 0 & 0 & 0 & 0 & x_1 & x_2 & x_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & x_4 & x_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & x_6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & x_7 & x_8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & y_1 & y_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & y_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & y_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & z_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & z_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix} \end{matrix}$$

Решение получается в виде матрицы $M = M_c^3$, в последнем столбце которой содержится информация о количестве разновидностей ценных бумаг a , b , c и d . По подобной схеме могут решаться задачи о комплектовании наборов товаров для оптовой торговли на различных рынках, о количестве деталей в изделии при узловой сборке, поставках комплектов товаров из разных городов (предприятий), например, при оказании гуманитарной помощи, при подсчете договоров разных видов в деятельности крупных страховых компаний и т. д.

ЗАКЛЮЧЕНИЕ

В учебном пособии обобщены и усовершенствованы методы моделирования стохастических и детерминантных процессов в экономике и управлении. Новые научные результаты работы:

– предложены методы определения необходимого числа наблюдений, допустимого числа факторов в модели и подбора вида зависимости по данным экспериментов; систематизированы оценки статистических моделей;

– разработаны методы моделирования процессов с малым числом наблюдений: экстремальных опорных точек и центра масс, не требующие применения метода наименьших квадратов;

– изложены методы линеаризации сложных однофакторных зависимостей и разработаны методы построения многофакторных зависимостей путем суммирования однофакторных, сокращения мерности пространства и минимизации числа опытов;

– показаны пути решения специальных эконометрических задач с применением аналитических методов и теории графов.

Практические результаты работы:

– определены возможности моделирования экономических процессов при малом объеме опытных данных и сокращенном объеме вычислений, что позволяет снизить потребные временные и человеческие ресурсы;

– простота предложенных методов и приемов позволяет сделать методы моделирования доступными при изучении, а модели более качественными.

Материалы учебного пособия могут найти широкое применение в научно-исследовательских работах студентов, магистрантов, аспирантов и научных работников.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Магнус, Я. Р. Эконометрика. Начальный курс / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий : учебник. – 3-е изд., перераб. и доп. – М. : Дело, 2000. – 400 с. – ISBN 5-7749-0055-X.
2. Эконометрика : учебник / И. И. Елисеева [и др.] ; под ред. И. И. Елисеевой. – М. : Финансы и статистика, 2002. – 344 с. – ISBN 5-279-01955-0.
3. Кулинич, Е. И. Эконометрия / Е. И. Кулинич. – М. : Финансы и статистика, 1999. – 304 с. – ISBN 5-279-02090-7.
4. Выгодский, М. А. Справочник по высшей математике / М. А. Выгодский. – М. : Наука, 1977. – 870 с.
5. Краткий физико-технический справочник. В 2 т. Т. 2 / под ред. К. П. Яковлева. – М. : Изд-во физ.-мат. лит., 1962. – 418 с.
6. Кремер, Н. Ш. Эконометрика : учеб. для вузов / Н. Ш. Кремер, Б. А. Путко ; под ред. проф. Н. Ш. Кремера. – М. : ЮНИТИ-Дана, 2002. – 311 с. – ISBN 5-238-00333-1.
7. Экономико-математические методы и прикладные модели / В. В. Федосеев [и др.] ; под ред. В. В. Федосеева. – М. : ЮНИТИ-Дана, 2002. – 391 с. – ISBN 5-238-00068-5.
8. Лаврик, В. И. Справочник по конформным отображениям / В. И. Лаврик, В. Н. Савенков. – Киев : Наук. думка, 1970. – 252 с.
9. Зажигаев, Л. С. Методы планирования и обработки результатов физического эксперимента / Л. С. Зажигаев, А. А. Кишьян, Ю. И. Романиков. – М. : Атомиздат, 1978. – 232 с.
10. Бать, М. И. Теоретическая механика в примерах и задачах. В 3 т. Т. 1 / М. И. Бать, Г. Ю. Джанелидзе, А. С. Кельзон. – М. : Наука, 1972. – 512 с.
11. Авондо-Бодио, Дж. Применение в экономике теории графов / Дж. Авондо-Бодио. – М. : Прогресс, 1966. – 160 с.

ПРИЛОЖЕНИЯ

Приложение 1

ТРЕНИРОВОЧНЫЕ ЗАДАНИЯ

Задание № 1. Вывести уравнение для описания однофакторной зависимости методом экстремальных опорных точек. Дать графическую и точностную оценку полученной модели. Варианты моделей:

1. Прямая линейная зависимость с границей:
 - а) в минимальной точке наблюдений;
 - б) в максимальной точке наблюдений.
2. Обратная линейная зависимость с границей:
 - а) в минимальной точке наблюдений;
 - б) в максимальной точке наблюдений.
3. Парабола.
4. Обратная парабола.
5. Угловая зависимость.
6. Обратная угловая зависимость.
7. Гипербола с границей:
 - а) в минимальной точке наблюдений;
 - б) в максимальной точке наблюдений.
8. Обратная гипербола с границей:
 - а) в минимальной точке наблюдений;
 - б) в максимальной точке наблюдений.
9. Встречная гипербола с границей:
 - а) в минимальной точке наблюдений;
 - б) в максимальной точке наблюдений.
10. Встречная обратная гипербола с границей:
 - а) в минимальной точке наблюдений;
 - б) в максимальной точке наблюдений.
11. Парабола третьей степени.
12. Обратная парабола третьей степени.
13. Сложная однофакторная зависимость.
14. Обратная сложная однофакторная зависимость.

Применительно к вариантам моделей значения фактора x и результирующего признака y приведены в табл. П1.1. Экономический смысл для x и y опущен, однако преподаватель может его определить самостоятельно, используя рекомендации, приведенные после текста заданий, с учетом закономерностей для реальных процессов.

Однофакторная модель

x	y
Число оборотов капитала	Рентабельность предприятия
Доходы предприятия	Расходы предприятия
Возраст	Производительность труда
Стаж работы	Выработка
Объем выпуска	Издержки
Цена	Объем продаж
Затраты на рекламу	Спрос
Объем налогов	Доход предприятия
Транспортные расходы	Цена товара
Доходы населения	Потребление
Расходы предприятия	Число занятых в производстве
Рост заработной платы	Количество безработных
Объем сделок на бирже	Стоимость акций
Объем капитала фирмы	Страховые резервы
Реальные доходы	Объем сбережений
Размер таможенных пошлин	Рыночная цена
Стоимость совокупных активов предприятия	Прибыль по акциям
Размер заемных средств	Прибыль
Количество работников	Производительность труда
Затраты на инновации	Потери от риска
Стоимость оборудования предприятия	Объем амортизационных отчислений
Объем ипотечных ссуд	Стоимость вводимого жилья
Стоимость активов	Размер дивидендов
Выручка от продаж предприятия	Чистая прибыль предприятия

Задание № 2. Вывести уравнение для описания однофакторной зависимости методом центра масс. Дать графическую и точностную оценку полученной модели. Варианты моделей:

1. Прямая линейная зависимость.
2. Обратная линейная зависимость.
3. Парабола.
4. Обратная парабола.
5. Гипербола.
6. Обратная гипербола.
7. Степенная зависимость.
8. Показательная зависимость.

Значения фактора x и результирующего признака y приведены в табл. П1.2.

Двухфакторная модель

x_1	x_2	y
Ставка заработной платы	Стоимость аренды	Цена продажи
Цена труда	Цена капитала	Производительность труда
Экспорт	Импорт	Объем денежных резервов
Душевой доход в семье	Расходы на детей в семье	Расходы на питание
Размер торговой площади	Товарные запасы	Объем реализации товара

Задание № 3. Построить методом центра масс двухфакторную и трехфакторную линейные модели по данным табл. П1.3 и П1.4 соответственно. Двухфакторную модель построить способом сведения пространственной задачи к плоской. Дать точностную оценку моделей.

Трехфакторная модель

x_1	x_2	x_3	y
Объем промышленного производства	Объем производства в сельском хозяйстве	Объем инвестиций в производство	Валовой внутренний продукт
Расстояние до центра города	Площадь квартиры	Этажность дома	Стоимость жилья
Число работников	Объем минеральных удобрений	Расходы воды на полив	Урожайность
Объем капитала	Оборачиваемость капитала	Численность работников	Чистый доход
Скорость резания на станке	Глубина резания	Подача	Стойкость инструмента

Примечание. Если задания выдаются для большого числа студентов, то разнообразие вариантов данных таблиц П1.1 – П1.3 можно увеличить, например, умножением значений результирующего признака на постоянный для шифра студента коэффициент, выбрав его по табл. П1.5. Шифр студента берется по двум последним цифрам номера зачетной книжки.

Таблица П1.1

Номер варианта модели	1 а		1 б		2 а		2 б		3		4		5		6		7 а		7 б	
	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>
1	0,21	4,9	38	25	0,09	870	24	12,1	0,07	70	40	13,2	360	9,0	17,0	0,158	1,2	15	4,2	0,38
2	0,40	5,2	51	26	0,17	860	30	13,0	0,08	86	60	9,2	365	10,8	23,0	0,130	1,7	24	6,1	0,59
3	0,62	9,0	62	44	0,24	750	32	12,5	0,09	100	105	6,5	470	14,6	24,5	0,095	1,9	31	8,4	0,72
4	0,79	8,2	69	35	0,33	745	37	8,7	0,10	87	165	2,0	490	17,7	24,5	0,157	2,4	37	8,5	0,84
5	0,98	10,0	81	50	0,41	610	40	6,6	0,12	138	200	1,6	510	18,8	27,3	0,066	2,7	39	10,0	0,90
6	1,15	8,4	92	36	0,48	690	48	9,7	0,13	120	235	2,2	600	15,0	31,6	0,045	3,6	45	12,2	0,88
7	1,37	12,5	103	58	0,59	540	53	7,8	0,14	143	275	4,5	680	12,7	31,8	0,068	3,9	46	14,8	0,87
8	1,60	14,1	112	70	0,79	495	56	6,0	0,18	149	330	10,3	790	11,3	36,7	0,112	4,3	45	17,8	0,92
9	1,82	14,7	119	74	0,85	398	59	4,1	0,25	132	340	6,4	800	8,5	36,9	0,138	4,8	50	20,0	0,88
10	2,10	17,6	127	87	0,96	412	64	5,2	0,26	98	370	12,1	–	–	37,4	0,153	5,5	48	–	–
11	2,23	16,8	139	80	1,14	340	76	3,1	–	–	–	–	–	–	–	–	6,4	48	–	–

116

Номер варианта модели	8 а		8 б		9 а		9 б		10 а		10 б		11		12		13		14	
	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>
1	260	202	0,50	7,40	33	0,046	952	18,0	2	0,3	0,35	10,3	60	4	0,4	0,15	14	140	0,14	3,0
2	295	192	0,55	9,03	50	0,047	1000	15,7	7	0,4	0,41	10,4	77	31	0,7	0,40	22	155	0,16	1,9
3	305	169	0,94	6,75	58	0,042	1080	16,6	14	0,4	0,48	10,6	110	47	0,9	0,35	26	195	0,24	1,1
4	322	183	1,06	5,65	69	0,043	1175	16,0	18	0,9	0,56	11,2	140	50	1,2	0,75	29	250	0,37	0,9
5	358	144	1,42	3,80	81	0,042	1206	12,7	25	1,1	0,62	11,9	155	51	1,4	1,45	30	305	0,42	0,7
6	465	132	1,80	2,35	84	0,033	1250	10,9	31	2,1	0,65	12,9	175	50	1,7	2,02	38	355	0,49	0,8
7	557	134	2,47	2,54	95	0,034	1300	10,8	35	2,4	0,70	14,0	195	42	1,7	2,84	42	355	0,64	1,2
8	615	118	2,98	2,00	109	0,024	1300	7,3	37	3,1	0,72	15,1	220	29	1,9	3,50	51	275	0,69	1,5
9	710	124	3,41	2,25	110	0,030	1340	6,6	37	4,0	0,71	16,2	225	14	2,0	4,25	56	155	0,72	1,9
10	796	123	–	–	112	0,020	1350	4,8	38	4,6	0,72	17,2	240	3	2,0	5,40	64	140	0,76	2,4
11	800	118	–	–	124	0,015	1370	3,5	38	5,3	–	–	–	–	–	–	70	135	0,77	3,0

Таблица П1.2

Номер варианта модели	1		2		3		4		5		6		7		8	
	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>
1	0,2	24	0,10	12,2	0,06	69	38	13,5	1,1	0,34	255	13,4	50	120	0,2	6,0
2	0,4	26	0,15	12,9	0,07	81	60	9,4	1,6	0,55	290	9,8	100	330	0,5	5,5
3	0,6	43	0,25	12,6	0,08	95	105	6,6	1,8	0,70	300	6,7	150	400	1,0	4,3
4	0,8	36	0,30	8,5	0,10	88	160	2,0	2,4	0,85	320	5,6	200	550	1,5	3,8
5	1,0	49	0,40	6,5	0,12	135	195	1,5	2,6	0,90	360	3,9	250	510	2,0	3,2
6	1,2	38	0,50	9,8	0,13	122	230	2,1	3,5	0,89	465	2,4	300	560	2,5	3,0
7	1,4	59	0,60	7,9	0,14	145	270	4,5	3,9	0,88	560	2,5	350	550	3,0	2,7
8	1,6	70	0,80	6,1	0,18	147	325	10,2	4,5	0,91	610	1,9	400	620	3,5	2,3
9	1,8	117	0,85	4,1	0,25	130	340	6,5	4,9	0,90	720	2,0	500	610	4,0	2,2
10	2,0	85	0,95	5,1	0,26	96	370	12,5	–	–	–	–	600	630	4,5	2,1

Таблица П1.3

Номер наблюдения	<i>x</i> ₁	<i>x</i> ₂	<i>y</i>
1	17	45	1400
2	32	38	1350
3	55	33	1290
4	74	29	1120
5	97	30	1070
6	120	25	920
7	125	18	900
8	148	14	875
9	174	17	810
10	295	15	750
11	320	12	730

Таблица П1.4

Номер наблюдения	<i>x</i> ₁	<i>x</i> ₂	<i>x</i> ₃	<i>y</i>
1	0,5	12	1150	100
2	1,2	21	1050	240
3	1,4	39	900	500
4	2,5	52	940	470
5	3,1	70	870	900
6	3,6	75	850	930
7	4,0	81	790	1280
8	4,9	94	760	1450
9	5,7	103	700	1600
10	6,2	111	720	1750
11	7,1	127	680	2000

Таблица П1.5

Последние цифры шифра студента	Коэффициент увеличения для <i>y</i>
0	1
1	1,1
2 ... 9	1,2 ... 1,9
10 ... 19	2 ... 2,9
20 ...	3,0 ...

КОНТРОЛЬНЫЙ ТЕСТ

1. Определите значения слов
 1. Метрия 1. Часть сложных слов, имеющих отношение к экономике
 2. Метрика 2. Измерение
 3. Эконо... 3. Часть сложных слов, связанных с измерением
 4. Стихотворный размер
 5. Выписка из книги ЗАГСa
2. Определите то, что относится к содержанию эконометрики
 1. Представление эмпирических закономерностей экономических процессов в математической форме
 2. Анализ экономических процессов
 3. Математическая оценка полученных моделей
 4. Прогнозирование тенденций в экономике
 5. Анализ количественных взаимосвязей между статистическими данными и экономическими показателями
3. Назовите применяемые методы подбора аппроксимирующей линии для множества опытных точек
 1. По минимальной сумме корней квадратных из отклонений по аппроксимирующей линии
 2. По минимальной сумме отклонений в третьей и более степени от аппроксимирующей линии
 3. Графический ручной
 4. По минимальной сумме абсолютных отклонений от аппроксимирующей линии
 5. Наименьших квадратов
4. Чему равно минимально допустимое количество статистических наблюдений в эконометрике? Укажите необходимые ограничения
 1. Превышение суммы свободного члена и числа коэффициентов в регрессионной модели
 2. Допустимая погрешность подбора аппроксимирующей функции
 3. Допустимое рассеивание опытных данных
 4. Значительное превышение количества факторов в модели
 5. Число опытов должно составлять десятки и даже сотни

5. Укажите возможные соотношения для определения числа включаемых в регрессионную модель кривой факторов при заданном количестве опытов n и уровней L .

1. $k = n/(6...7)$ 3. $k = \lg n / \lg L$ 5. $k \leq 0,5n$
2. $k \leq 0,33 n$ 4. $k \leq 0,2 n$

6. Приведите в соответствие методы линеаризации зависимостей между факторным и результирующим признаками и их особенности

- | | |
|--------------------------------------|---|
| 1. Метод замены переменных | 1. Исходные данные для факторного признака преобразуются с помощью зависимости, содержащей этот признак в составляющей регрессионного уравнения |
| 2. Метод логарифмирования | 2. В логарифмических координатах определяют угловой коэффициент модели, затем рассчитывают свободный член |
| 3. Графоаналитический метод | 3. Исходные данные факторного и результирующего признаков предварительно логарифмируются |
| 4. С помощью определения приростов | 4. Исходные данные для нелинейной зависимости отображают в значения, укладываемые на прямую |
| 5. С помощью дробно-линейной функции | 5. Графическим дифференцированием усредненной кривой строят графики производных до тех пор, пока не получат график прямой |

7. Приведите в соответствие показатели корреляции с их назначением

- | | |
|--|--|
| 1. Парный коэффициент корреляции | 1. Не допустить преувеличения тесноты связи при числе параметров при факторах, близком к объему наблюдений |
| 2. Индекс корреляции | 2. Практическая значимость уравнения множественной линейной регрессии |
| 3. Частный коэффициент корреляции | 3. Ранжирование факторов и целесообразность их включения в модель |
| 4. Совокупный коэффициент множественной корреляции | 4. Установление надежности выбора уравнения регрессии |

- | | |
|--|--|
| 5. Скорректированный индекс множественной корреляции | 5. Установление зависимости между одной переменной и математическим ожиданием другой |
|--|--|

8. Приведите в соответствие оценочные показатели с их физическим смыслом

- | | |
|--|---|
| 1. Показатели корреляции | 1. Теснота связи оцениваемых показателей |
| 2. Стандартная ошибка и критерий Стьюдента | 2. Значимость параметра модели или коэффициента корреляции |
| 3. Критерий Фишера | 3. Значимость влияния факторов на результат |
| 4. Коэффициент эластичности | 4. Процентное изменение результата под влиянием фактора |
| 5. Показатели детерминации | 5. Доля дисперсии результата, связанная с регрессионной зависимостью, в его общей дисперсии |

9. Укажите неверное утверждение

Прямую можно построить по исходным данным:

1. По координатам двух точек на плоскости и в пространстве
2. По координатам одной точки и угловому коэффициенту
3. По координате точки на оси координат и углу наклона к одной из осей координат
4. По тангенсу угла наклона к оси абсцисс
5. По двум точкам на осях координат

10. Укажите, каким зависимостям соответствуют выражения для факторных признаков при расчетах по методу экстремальных опорных точек

- | | |
|---|--------------|
| 1. $\frac{x_i}{x_{\min}} - 1$ | 1. Парабола |
| 2. $\frac{(x_i - x_0)^2}{x_0^2}$ | 2. Гипербола |
| 3. $\left 1 - \frac{x_i}{x_0} \right $ | 3. Прямая |

4. $1 - \frac{x_{\min}}{x_i}$ 4. Угловая
5. $1 - \frac{2x_0 - x_{\max}^2}{2x_0 - x_{i2}}$ 5. Встречная гипербола

11. К каким методам относятся виды приведенных формул:

1. $\hat{y} = y_{\min(\max)} (1 \pm b g_x)$ 1. Метод наименьших квадратов
2. $\hat{y} = a_0 + a_1 x_1 + a_2 x_2 + a_{12} x_1 x_2$ 2. Метод построения модели с учетом взаимовлияния факторов
3. $n a_0 + (\sum x_1) a_1 + (\sum x_2) a_2 = \sum y$ 3. Метод экстремальных опорных точек
4. $k = k_0 \prod_0^t (1 + \eta_t) \prod_0^t (1 + \vartheta_t) \prod_0^t (1 + \psi_t)$ 4. Метод центра масс
5. $\hat{y} = y_c (1 \pm b g_x)$ 5. Метод начисления сложных процентов

12. Какие процедуры возможны при подготовке исходных данных в эконометрике? Выберите неправильные ответы

1. Предварительная статистическая оценка модели
2. Определение необходимого числа наблюдений
3. Определение допустимого количества факторов в модели
4. Выбор вида модели
5. Линеаризация модели
6. Оптимизация расположения начала координат
7. Определение положения оси симметрии кривой
8. Потенцирование модели
9. Выбор показателя степени нелинейной зависимости
10. Сведение пространственной задачи к плоской

13. Приведите в соответствие методы построения эконометрических моделей с их преимуществами

- | | |
|-------------------------------|--|
| 1. Метод наименьших квадратов | 1. Малая чувствительность к выбросам экстремальных точек |
| | 2. Высокая чувствительность поиска оптимального положения аппроксимирующей линии |

- | | |
|--------------------------------------|--|
| 2. Метод экстремальных опорных точек | 3. Введение в формулу граничных значений переменных |
| 3. Метод центра масс | 4. Глубоко разработанный математический аппарат и информационная поддержка |
| | 5. Применимость и точность при малом числе наблюдений |
| | 6. Обязательная линеаризация нелинейных связей |
14. Какой из перечисленных недостатков не относится к методу наименьших квадратов?
1. Громоздкость метода растет пропорционально числу вводимых в модель факторов
 2. Высокая чувствительность к «выбросам» точек наблюдения
 3. Значения исходных данных должны иметь одинаковую точность и подчиняться нормальному закону распределения
 4. Требуются массовые эксперименты
 5. Недопустимы сочетания значений исходных данных больше и меньше единицы
 6. Неравнозначность при расчете модели точек по длине аппроксимирующей линии
15. Расположите в нужной последовательности операции для метода центра масс
1. Выбрать вид модели
 2. Определить координаты центра масс
 3. Рассчитать параметры модели и записать модель
 4. При необходимости выполнить линеаризацию модели, проверив графически качество линеаризации
 5. Скорректировать линеаризованные значения исходных данных
 6. Оценить качество модели
 7. Выполнить обратное преобразование модели
16. Составьте алгоритм решения задачи о количестве прадеревьев для ориентированной графовой модели
1. Составить матрицу смежности
 2. Построить диагональную матрицу
 3. Определить суммы элементов матрицы по столбцам
 4. Решить минор элемента B_{11}
 5. Определить разность матриц
 6. Графически представить пути движения по графу

17. Укажите отличительные особенности специальных методов эконометрики в данном учебном пособии

1. Проведение аппроксимирующих линий через выделенные опорные точки
2. Минимальное число статистических наблюдений
3. Использование закономерностей и методов фундаментальных наук
4. Разработка специальных методов оценки эконометрических моделей
5. Дополнение и совершенствование метода наименьших квадратов

18. Выберите метод построения следующих моделей:

- | | |
|------------------------------|--------------------------------------|
| 1. Угловая зависимость | 1. Метод наименьших квадратов |
| 2. Встречная гипербола | 2. Метод экстремальных опорных точек |
| 3. Парабола третьей степени | 3. Метод центра масс |
| 4. Гипербола третьей степени | 4. Метод построения графовых моделей |

19. Укажите, сочетанием каких методов можно построить простую линейную модель, описываемую обычно квадратным трехчленом

1. Методом центра масс
2. Методом экстремальных опорных точек
3. Методом суммирования линеаризованных зависимостей
4. Сведением пространственной задачи к плоской
5. Методом разделения переменных на линейную и нелинейную части

20. Из приведенных параметров постройте формулу парного коэффициента корреляции вида

$$r = \frac{1-2}{\sqrt{3-4\sqrt{5-6}}}$$

Параметры

- | | | | |
|----------------------|---------------------|--------------------|---------------------|
| 1. $\bar{y} \bar{x}$ | 2. \bar{x} | 3. \overline{yx} | 4. $\overline{x^2}$ |
| 5. \bar{y}^2 | 6. $\overline{y^2}$ | 7. \bar{x}^2 | 8. \bar{y} |

Эти параметры следует подставить на месте соответствующих цифр в формуле.

Учебное издание

ДЕНИСЕНКО Владимир Иванович

ЭМПИРИОМЕТРИЯ

Учебное пособие

Редактор А. П. Володина
Технический редактор С. Ш. Абдуллаева
Корректор Е. П. Викулова
Компьютерная верстка Е. А. Герасиной

Подписано в печать 01.08.17.
Формат 60×84/16. Усл. печ. л. 7,21. Тираж 102 экз.
Заказ

Издательство
Владимирского государственного университета
имени Александра Григорьевича и Николая Григорьевича Столетовых.
600000, Владимир, ул. Горького, 87.