6.1. Первичная обработка информации

Анализ и обобщение данных осуществляется как с помощью описательных, так и с помощью аналитических методов. Среди аналитических методов в маркетинговых исследованиях часто применяются: анализ трендов, методы регрессионно-корреляционного анализа, дискриминантный анализ, кластерный анализ, факторный анализ и другие. Количественная информация измерима и может быть обработана так, как это было показано ранее. При этом следует помнить, что исходная информация может содержать в себе ошибки. По своим свойствам и характеру влияния на результаты наблюдений ошибки подразделяют на *грубые, систематические и случайные*.

Грубые ошибки возникают в случае невнимательности человека, эту информацию записывающего, передающего или получающего. В результате этой невнимательности измеренное число увеличивается на порядок или резко выделяется в общей совокупности данных. В этой связи с такой ошибкой легко бороться – она легко обнаруживается и легко устраняется. Для этого есть несколько возможных процедур:

1. Графический анализ. При этом статистические данные наносят на некоторый график. Если рассматривается некоторая статистическая взаимосвязь, то изучается эта взаимосвязь. Данные, которые выделяются из совокупности наблюдений, отбрасываются. Если изучаются динамические ряды, то строится зависимость изменения показателей во времени. Данные, которые выделяются из общей тенденции, отбрасываются.
2. Табличный анализ. Данные, которые невозможно, или нежелательно изображать графически, располагают в некотором порядке, присущем анализируемому процессу. Те из них, которые выбиваются из этого порядка, отбрасываются.
3. Статистический анализ. Полученные данные проверяются на одну из статистических гипотез относительно характера распределения вероятностей появления анализируемой совокупности данных. В случае выявления некоторого закона распределения вероятностей, данные, противоречащие этому закону, отбрасываются. В самом простом случае, когда измеренные величины колеблются около некоторого значения, рассчитывается средняя арифметическая этого значе-

ния, дисперсия, доверительные интервалы. Данные, выходящие за эти интервалы, отбрасываются.

Источниками систематической ошибки могут являться как инструмент сбора и обработки информации, так и человеческий фактор (желание приукрасить ситуацию или скрыть часть неблагоприятной информации). К сожалению, в маркетинговых исследованиях при работе с вторичными данными (официальной и неофициальной статистикой) очень часто приходится иметь дело именно с такого рода ошибками. Дело в том, что большая часть экономических показателей отражает эффективность деятельности того или иного подразделения, той или иной системы, того или иного региона. Классическим примером ошибки информации такого рода является записываемые год от года в статистические сборники, данные о количественных показателях экономического развития бывшего СССР, которые отражали не столько реальные процессы, сколько желаемые результаты. Известно, что, например, в Узбекской ССР долгие годы шли приписки о сборе невыращенного хлопка, которые попадали в статистические сборники. По отчетам о выполнении плановых заданий по сбору хлопка составлялись планы работы текстильной промышленности, которая из несобранного хлопка не могла, естественно, выпустить несуществующую ткань. В результате этого изменялись нормы расхода хлопка на единицу ткани, нормы электропотребления и т.п. Таким образом, практически все обобщающие данные экономического развития (валовой продукт, национальный доход и т.п.) отдельных регионов и страны в целом оказались засоренными ошибками такого рода.

Сегодня причиной возникновения подобной ошибки может быть, например, желание уменьшить выплаты по платежам в бюджеты и внебюджетные фонды, искажение данных в ходе «информационной войны» с конкурентами и т.п.

Систематические ошибки могут быть также выявлены и исключены, так как имеют примерно одну и ту же величину, один и тот же знак, поэтому исходные данные, содержащие этот тип ошибки, всегда несколько завышены или занижены. Объективным источником этой ошибки служат, в основном, измерительные устройства, приёмы или приборы, вносящие одну и ту же погрешность при измерениях.

Случайные ошибки неизбежны. Причины их появления многообразны. Они вызваны действием множества случайных неконтролируемых факторов и поэтому не поддаются анализу. В результате этого практически любое измерение содержит случайные ошибки, но так как источников возникновения случайных ошибок достаточно много, они, как правило, обладают следующими свойствами[1].

Первое свойство. Для ряда результатов наблюдений с известным параметром распределения абсолютные величины случайных ошибок с заданной вероятностью Р не превосходят определенного предела. Это значит, что влияние случайных ошибок на результат все-таки незначительно.

Второе свойство. Положительные и отрицательные случайные ошибки равновозможны, т.е. они одинаково часто встречаются при наблюдениях. Из этого вытекает и следующее свойство.

Третье свойство. Математическое ожидание случайной ошибки равно нулю.

Четвертое свойство. Малые по абсолютной величине случайные ошибки встречаются при наблюдениях чаще, чем большие.

Следовательно, можно предполагать в большинстве случаев, что случайные ошибки подчиняются закону нормального распределения вероятностей и их математическое ожидание равно нулю. Таким образом маркетинговая информация содержит ошибки наблюдений, но они в общем случае несоизмеримы по сравнению с самими наблюдениями.

С учетом перечисленных свойств, создается ситуация, когда проявляются условия действия центральной предельной теоремы теории вероятностей (закон больших чисел), в соответствии с которой «совокупное действие большого числа случайных факторов приводит при некоторых, достаточно широких условиях к результату, почти не зависящему от случая»[2]. Таким образом, избежать влияния случайных ошибок можно, если увеличить объем выборки.

После того, как будут устранены ошибки информации, перед исследователем возникает проблема ее систематизации и обработки. В достаточно редких случаях необходимая исследователю информация представлена в систематизированном виде и в виде, пригодном для последующего анализа и обработки. Чаще всего информация представлена виде некоторой неупорядоченной совокупности. Для того чтобы ее обработать и сделать соответствующие выводы, она нуждается в упорядочении и систематизации.

Систематизация информации заключается в ее представлении в виде таблиц, графиков, диаграмм и других формах, удобных для исследователя и показывающих некоторые наиболее очевидные закономерности. В большинстве случаев маркетологи предпочитают сведение информации в статистические таблицы – при этом возмо-

---

[1] Большаков В.Д. Теория ошибок наблюдений. - М.: Недра, 1983. - 223 с.

[2] Дунин-Барковский И.В., Смирнов Н.В. Теория вероятностей и математическая статистика в технике. – М.: Гос. изд-во технико-теоретической литературы, 1955. - С.104.

жен их последующий формализованный анализ с помощью математических методов.

Для того чтобы неупорядоченную совокупность данных можно было свести в таблицу, необходимо определить признак упорядочивания данных. Такими признаками могут являться:

- время – период наблюдения или конкретные наблюдения в зафиксированный момент времени,
- номера экспертов, дававших оценку объекту исследования;
- ранги, полученные для свойств товара по шкале отношений или интервалов;
- товарный ряд и т.п.

В первом случае осуществить систематизацию достаточно просто. Информация рассматривается в качестве зависимого от времени фактора. Тогда в первой колонке таблицы указывается время или промежуток времени, а в последующих колонках – количественная информация об объекте исследования в эти моменты времени или промежутки времени. Точно также систематизируются данные по и другим указанным критериям.

Довольно часто на практике встречаются случаи, когда сведенная в таблицы информация оказывается неполной – часть данных отсутствует. Это может возникнуть, например, в случае, когда при опросе один из респондентов ответил не на все вопросы. В подобном случае возникает необходимость восстановления утерянной в процессе сбора и обработки наблюдений информации. Определить неизвестную величину внутри статистического ряда можно с помощью одного из методов интерполяции.

Теория интерполяции является одним из старейших разделов математики и начиналась она работами И.Ньютона, Ж.Лагранжа, Н.Абеля, Ш.Эрмита и др. По определению интерполирование – это способ нахождения какой-либо величины по известным отдельным значениям этой же или других величин, связанных с ней[1].

Теория интерполяции является одним из наиболее разработанных разделов численных методов, и поэтому поставленная задача может быть решена с той или иной степенью точности.

Проще всего воспользоваться методом разностей, хотя это и не самый точный метод экстраполяции. Его суть заключается в следующем.

Первая производная функции, как известно, находится по формуле:

[1] Янович Л.А. Интерполирование // Математическая энциклопедия, т.2. – М.: Советская Энциклопедия, 1979. – С.622.

$$Y_t^{/} = \lim_{x \to 0} \frac{\Delta Y_t}{\Delta X_t}. \quad (6.1.1)$$

Эта производная остаётся постоянной и не равной нулю, если между двумя переменными существует линейная функциональная зависимость. Если между двумя переменными существует линейная регрессионная зависимость, то в каждой конкретной точке наблюдения за переменными в момент t первая производная будет иметь значения, в общем случае отличающиеся от значений первой производной в другие моменты времени. Эти отклонения вызваны действием множества случайных факторов и поэтому значения первой производной в разных точках будут колебаться вокруг своего математического ожидания, лучшей оценкой которой в данном случае является средняя.

Как вычислить первую производную регрессионной зависимости, если не известны коэффициенты линейной функции, которая описывает эту зависимость? При этом у исследователя имеются в распоряжении только эмпирические значения X_t и Y_t. Для решения поставленной задачи первую производную заменяют отношением конечных разностей:

$$Y_t^{/} \approx \frac{\Delta Y_t}{\Delta X_t}. \qquad (6.1.2)$$

Такая замена возможна только в том случае, когда приращения ΔX_t (конечные разности первого порядка) достаточно малы. Обычно с такой ситуацией и приходится иметь дело на практике. Поэтому легко найти первые разности ΔX_t и ΔY_t:

$$\Delta X_t = X_t - X_{t-1}, \ \ \Delta Y_t = Y_t - Y_{t-1}, \qquad (6.1.3)$$

а затем найти их отношение

$$\Delta Y_t / \Delta X_t = (Y_t - Y_{t-1})/(X_t - X_{t-1}). \qquad (6.1.4)$$

Если это отношение при разных значениях t действительно колеблется около некоторого значения, можно найти одну из оценок этого значения, а именно, среднюю арифметическую:

$$\overline{\Delta} = \frac{1}{n-1} \sum \frac{\Delta X_t}{\Delta Y}, \qquad (6.1.5)$$

которая будет в среднем характеризовать первую производную зависимости, а значит, будет являться одной из оценок коэффициента линейной регрессии Y на X. С помощью этого коэффициента можно решать задачи интерполяции подобных линейных зависимостей. Однако точность такой интерполяции не очень высока, к тому же линейные зависимости встречаются крайне редко в маркетинговой практике. Поэтому метод вычисления конечных разностей для целей интерполяции в настоящее время почти не применяется на практике, но именно конечные разности легли в основу интерполяции методом полинома Ньютона.

Интерполяционная формула Ньютона применяется в том случае, когда упорядоченные значения X_t находятся на равном расстоянии друг от друга, то есть, когда $\Delta X_t = X_{t+1} - X_t = h = const$ для всех t. Константа h получила название шага наблюдений (шаг таблицы наблюдений). С учётом этого свойства значения функции двух переменных X_t и Y_t характеризуются только изменением переменной Y_t. Эти изменения можно определить, вычислив значения конечных разностей. Сами разности можно осуществлять с шагом назад, как это было сделано в (6.1.3), а можно делать и с шагом вперёд, как это предусмотрено методом Ньютона. При этом формулы для расчёта первых разностей будет иметь вид:

$$\Delta Y_t = Y_{t+1} - Y_t;$$

вторых разностей:

$$\Delta^2 Y_t = \Delta Y_{t+1} - \Delta Y_t;$$

третьих разностей:

$$\Delta^3 Y_t = \Delta^2 Y_{t+1} - \Delta^2 Y_t$$

и так далее для других конечных разностей. Этот процесс показан в таблице 6.1.

Таблица 6.1. Конечные разности различных порядков

X_t	Y_t	ΔY_t	$\Delta^2 Y_t$	$\Delta^3 Y_t$	$\Delta^4 Y_t$	$\Delta^5 Y_t$
X_1	Y_1	ΔY_1	$\Delta^2 Y_1$	$\Delta^3 Y_1$	$\Delta^4 Y_1$	$\Delta^5 Y_1$
X_2	Y_2	ΔY_2	$\Delta^2 Y_2$	$\Delta^3 Y_2$	$\Delta^4 Y_2$	
X_3	Y_3	ΔY_3	$\Delta^2 Y_3$	$\Delta^3 Y_3$		
X_4	Y_4	ΔY_4	$\Delta^2 Y_4$			
X_5	Y_5	ΔY_5				
X_6	Y_6					

Легко убедиться в том, что если будет не шесть наблюдений, как это показано в табл. 6.1, а, например, восемь, то можно будет вычислить ещё и разности шестого и седьмого порядков; если будет *n* наблюдений, то можно вычислить разности (*n –1)*– го порядка. Разность каждого порядка в определённой степени характеризует производную этой степени, соответствующую данному порядку.

Так как значения X_t в рассматриваемом случае представляют собой арифметическую прогрессию, то, введя обозначение:

$$q = \frac{(x - x_1)}{h}, \quad (6.1.6)$$

получим интерполяционную формулу Ньютона на основе вычисленных значений конечных разностей:

$$P_n(x) = y_1 + q\Delta y_1 + \frac{q(q-1)}{2!}\Delta^2 y_1 + ... + \frac{q(q-1)...(q-n+1)}{n!}\Delta^n y_1 . \quad (6.1.7)$$

Подставляя в (6.1.6) известное значение X_k, легко найти интерполируемое значение Y_k.

Формула (6.1.7) называется интерполяционной формулой Ньютона для интерполирования вперёд[1]. Существует также формула Ньютона для интерполирования назад, которая использует разности, вычисленные по принципу (6.1.3).

В случаях, когда необходимо получить более точные результаты интерполяции, рекомендуется использовать и более сложные нелинейные интерполяционные формулы, в первую очередь, интерполяционные формулы Лагранжа и Ньютона. Методика интерполяций этими методами исходит из необходимости построения интерполирующей функции, проходящей через все точки X_t и Y_t. Подобной функцией является многочлен (*n-1*)–ой степени, который, очевидно, пройдёт через все *n* точек:

$$L(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + ... + a_{n-1} x^{n-1} . \quad (6.1.8)$$

Однако достаточно часто построение подобных функций оказывается излишним, поскольку с подобной задачей успешно могут справиться и функции с более низкими степенями. Именно эту задачу решает методы вычисления интерполяционного многочлена (полинома) Лагранжа, который рассчитывается по формуле[2]:

[1] Самарин М.К. Ньютона интерполяционная формула // Математическая энциклопедия, т.3. – М.: «Советская энциклопедия», 1982. - С.1092.
[2] Кудрявцев Л.Д., Самарин М.К. Лагранжа интерполяционная формула // Математическая энциклопедия, т.3. – М.: «Советская энциклопедия», 1982. - С. 170-171.

$$L_n(x)=\sum_{t=1}^{n}\frac{(x-x_1)(x-x_2)...(x-x_{t-1})(x-x_{t+1})...(x-x_n)}{(x_t-x_1)(x_t-x_2)...(x_t-x_{t-1})(x_t-x_{t+1})...(x_t-x_n)}y_t\,. \quad (6.1.9)$$

Этот интерполяционный многочлен вычисляется для имеющихся пар значений, а затем, по известному значению X_k ($k\in\ t$)интерполируют значение Y_k, подставляя это значение X_k в формулу (6.1.9).

Каждая из интерполяционных формул (6.1.7) и (6.1.9) даёт при вычислении ошибки, которые при необходимости можно вычислить и учесть в расчётах[1]. Однако, величина этих ошибок достаточно мала, поэтому их влиянием пренебрегают, тем более, что эмпирические данные маркетинговых исследований загрязнены другими многочисленными ошибками.

Уже в ходе обработки маркетинговой информации могут возникнуть две принципиально различные ошибки:

- связанные с неточностью исходной информации и
- вызванные несовершенством инструментария обработки данных.

Первые могут быть названы ошибками наблюдений, вторые - ошибками инструментария.

В современной экономической и математической науке наибольшее внимание уделяется анализу и устранению ошибок первого рода. При этом считается, что специалист, использующий математическую статистику настолько грамотен, что просто не допустит появления ошибок второго рода. Значительно сложнее дело с ошибками инструментария. И здесь можно выделить два типа ошибок.

Первый тип ошибок связан с ошибками в применении инструментария, разработанного для изучаемого процесса. Такие ошибки вызваны неграмотностью исследователя и могут быть очень просто удалены заменой исследователя или прикреплением к нему специалиста, владеющего методологией данных исследований.

Второй тип ошибок связан с несоответствием инструментария исследований характеру и свойствам исследуемой системы, что вызвано неразработанностью самого инструментария, так как эти проблемы находятся на рубеже «знания-незнания» и являются предметом научного исследования. Более подробно о влиянии ошибок инструментария на результаты обработки маркетинговых данных будет сказано ниже в следующих параграфах.

К числу методов интерполирования относят также и интерполирование методами моделирования с использованием методов ма-

---
[1] Хемминг Р.В. Численные методы для научных работников и инженеров. – М.: Наука, 1968. – С. 110 –112.

тематической статистки – чаще всего с помощью метода наименьших квадратов. Суть этого метода заключается в следующем. По имеющимся статистическим данным двух переменных X_t и Y_t строится регрессионная модель зависимости Y_t от X_t или наоборот. Затем, по известным данным X_k с помощью модели рассчитывается интерполируемая величина Y_k или наоборот. Методы моделирования маркетинговой информации будут рассмотрены ниже.

6.2. Корреляционный анализ

Выявление причинно-следственных связей служит главным объектом маркетинговых исследований, нацеленных на информационное обоснование маркетинговых решений. Так, например, если в ходе маркетинговых исследований выяснилось, что на объём приобретения товара главное влияние оказывает его физический вес, то маркетинговое решение будет нацелено на вопросы управления весом товара, изменением именно этой характеристики для того, чтобы товар получил конкурентные преимущества на рынке. Если эта взаимосвязь оказалась ошибочной и в ходе маркетинговых исследований и обработки данных была допущена ошибка, то решения, основанные на этом выводе, принесут большие убытки для предприятия.

Поэтому в связи с особой важностью результатов этой процедуры для последующего принятия маркетинговых решений, необходимо более подробно остановиться на вопросе выявления степени взаимосвязи между факторами.

В науке часто приходится иметь дело с функциональными зависимостями, когда величина Y однозначно определяется через другую величину X:

$$Y=f(X). \qquad (6.2.1)$$

Неважно, является ли X единственной переменной или вектором переменных, главное то, что величина результирующего фактора Y вполне определена значениями фактора X.

Однако при наблюдениях за объектами в реальной жизни значительно более часто приходится иметь дело с другой формой взаимосвязи - стохастической. Эта зависимость характерна для двух случайных величин, когда одна из них реагирует на изменение другой изменениями своего закона распределения вероятностей.

Задачи, связанные с изучением зависимостей между величинами, отличных от функциональных, достаточно многообразны. Если же предположить, что эти зависимости носят стохастический характер, то тогда можно использовать для исследований основы теории корреляции.

Корреляционный анализ, по определению - это совокупность методов оценки коэффициентов, характеризующих корреляцию между случайными величинами или признаками, и методов проверки гипотез об их значениях.

Корреляция - понятие, характеризующее взаимную зависимость двух случайных величин.

Взаимосвязь между двумя случайными факторами, выраженная в явном виде, может быть названа регрессионной. Регрессионная функция чаще всего упоминается как линия регрессии. Линия регрессии представляет собой математическое ожидание взаимосвязи, а отклонения от нее - случайные величины, как правило, с незначительной дисперсией и нулевым математическим ожиданием.

Тогда случайная величина Y может быть выражена через другую случайную величину X с помощью линии регрессии f(X):

$$Y=f(X)+e_t, \quad (6.2.2)$$

где e_t - случайная величина, характеризующая отклонения от линии регрессии.

В указанных выше определениях специально выделено слово "случайный" - именно для случайных и только для случайных процессов предназначен корреляционный анализ. Однако для того, чтобы определить случайность или не случайность процесса, следует обратиться к точным математическим формулировкам этого понятия, потому что обыденное понимание этого слова совсем не соответствует его математическому смыслу.

Случайный процесс, стохастический процесс, вероятностный процесс, случайная функция времени - процесс (т.е. изменение во времени состояния некоторой системы), течение которого зависит от случая и для которого определена вероятность того или иного его течения[1].

Следующее важное определение: случайная функция - функция произвольного аргумента, такая, что ее значения определяются с помощью некоторого испытания и в зависимости от его исхода могут быть различными, причем для них существует определенное распределение вероятностей.

---

[1] А.М.Яглом. Случайный процесс //Математическая энциклопедия, т.5. - М.: Советская энциклопедия, 1984. – с.890 - 891

Из указанных определений следует, что случайным может быть назван процесс, имеющий вероятностную природу.

Взаимосвязь между двумя случайными величинами можно определить с помощью ряда расчетных коэффициентов и методов, которые и объединены в теории корреляции в единый математический аппарат. В первом параграфе данной главы было показано соответствие того или иного математического аппарата каждой из используемых в практике маркетинговых исследований шкале измерения информации.

При определении наличия взаимосвязи рекомендуется в начале изобразить анализируемые показатели на графике. Для шкалы наименований это сделать не удастся, а остальные шкалы такую возможность исследователю предоставляют. Если множество сгруппированных таким образом пар будет представлено в форме некоторой графической зависимости или некоторого ориентированно не горизонтально и не вертикально облака, то можно делать вывод о наличии зависимости между показателями.

Графический анализ имеет важным преимуществом то, что он позволяет выявить наличие нелинейной взаимосвязи. Недостаток графического анализа заключается в том, что его результаты во многом определяются выбранным масштабом измерения. В любом случае графический анализ даёт исследователю много различной информации, которую трудно получить формальным путём.

Для подтверждения выводов графического анализа используют аппарат корреляционного анализа, который в достаточной степени обширен. Чаще всего в практике маркетинговых исследований используют метрическую шкалу и шкалу интервалов. Для анализа статистических данных, измеренных с помощью этих двух шкал, применяют два расчетных коэффициента - корреляционное отношение и коэффициент парной (или множественной) корреляции.

Свойства корреляционных отношений хорошо изучены - они характеризуют степень приближения зависимости между случайными факторами к функциональной. Одним из свойств корреляционного отношения является равенство его единице при строгой функциональной зависимости. Это свойство можно обратить и утверждать, что, если при расчетах оказалось, что корреляционное отношение $\eta = 1$, то Y есть однозначная функция X.

Корреляционное отношение будет меньше единицы, если существуют внутригрупповые дисперсии - при заданном фиксированном значении X случайная величина Y имеет разные значения, отличающиеся от математического ожидания, вызванные влиянием случайных факторов. В практике маркетинговых исследований случаи, когда используется корреляционное отношение крайне редки. Обычно встречаются зависимости, когда каждому значению одного фактора соответствует только одно значение другого фактора. Поэтому используют другие коэффициенты.

Чаще всего из перечисленных в первом параграфе данной главы коэффициентов рассчитывают коэффициент парной корреляции. Как следует из (6.1.6), коэффициент парной корреляции представляет собой среднее геометрическое коэффициентов регрессии Y на X и X на Y. Эти коэффициенты находят с помощью метода наименьших квадратов (МНК). Непосредственное использование МНК приведёт к необходимости решения системы нормальных уравнений и очень громоздкой формуле для вычисления коэффициента парной корреляции. Поэтому проводят предварительное центрирование исходных данных, то есть от каждого Y_i и X_i вычитают их средние арифметические $\bar{Y}$ и $\bar{X}$. Тогда коэффициенты будут легко вычисляться следующим образом:

$$b_1 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}, \quad (6.2.3)$$

$$a_1 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}. \quad (6.2.4)$$

Подставляя полученные формулы в формулу коэффициента корреляции Пирсона[1] - среднее геометрическое коэффициентов регрессий Y_i на X_i и X_i на Y_i - получим искомую формулу коэффициента парной корреляции:

$$r = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \sum_{i=1}^{n}(X_i - \bar{X})^2}}. \quad (6.2.5)$$

Квадрат коэффициента парной корреляции называют коэффициентом детерминации. Считают, что чем ближе модуль коэффициента парной корреляции к единице, тем сильнее приближается изучаемая взаимосвязь к линейной. Чем ближе модуль коэффициента парной корреляции к нулю, тем слабее линейная взаимосвязь.

[1] См. §2.4

В литературе по маркетингу приходится сталкиваться со следующим совершенно не верным толкованием коэффициента парной корреляции – «коэффициент парной корреляции характеризует степень взаимосвязи между факторами». Основываясь на таком понимании коэффициента, исследователи практически повсеместно заходят в тупик - коэффициенты парной корреляции бывают близки к единице даже в тех случаях, когда взаимосвязи быть не может. Исследованию этой проблемы посвящено много работ, а сама проблема получила название "ложной" корреляции.

Наиболее часто встречается следующее определение ложной корреляции: ложными корреляциями называют такие корреляции между переменными, которые не могут быть объяснены с точки зрения сколько-нибудь "разумной" теории; подобные корреляции следует рассматривать как результат простого совпадения.

На практике очень сложно определить, какая же корреляция является истинной, а какая - ложной. Например, корреляция между производством обуви промышленностью нашей страны и численностью населения Молуккских островов будет явно ложной. Корреляция между производством обуви промышленностью и численностью населения России уже близка к истинной (по крайней мере, с точки зрения "разумной" теории).

Однако на практике исследователь, получивший высокие значения коэффициентов корреляции, спешит делать вывод о высокой степени взаимосвязи между факторами. Таких примеров в экономической практике можно привести достаточно много.

О ложной корреляции заговорили давно, впервые этот термин употребил в начале века известный математик-статистик К.Пирсон. Он показал, что ложная корреляция может возникать в том случае, когда на взаимосвязь двух переменных влияет некоторая третья переменная, называемая в факторном анализе латентной (т.е. скрытой) переменной.

С тех пор изучению явления ложной корреляции придавалось большое значение. Уже в 1969 г. Н.С.Четвериков рассматривал пять возможных случаев ложной корреляции[1]. Совершенно правильно отмечая, что сущность ложной корреляции коренится в первую очередь в логических ошибках, совершаемых в неправильном пользовании методом корреляции, одной из причин ложной корреляции Н.С.Четвериков, в частности, считал и наличие тенденций в исследуемых рядах, а также наличие периодических или сезонных волн.

Развитие идей К.Пирсона привело к появлению нового раздела в математике - факторному анализу. Факторный анализ возник из

[1] Четвериков Н.С. О ложной корреляции //Применение методов корреляции в экономических исследованиях: Ученые записки по статистике. Т. XVI. - М.: Наука, 1969. - с. 203-229.

стремления обнаружить скрытую основу нескольких явлений или свойств, встречающихся одновременно. Он основан на предположении о существовании латентной переменной, корреляция которой с наблюдаемыми переменными и является причиной взаимной корреляции наблюдаемых факторов.

Наблюдаемые в экономической динамике факторы в действительности могут и не быть взаимосвязанными, и если бы не влияние этой латентной переменной, коэффициент парной корреляции между ними был бы очень мал - так рассуждает большинство эконометристов, занятых борьбой с ложной корреляцией. На основе такого вывода, в основном все практические рекомендации по борьбе с ложной корреляцией в классической эконометрии сводятся к выявлению этого третьего "мешающего" фактора и его исключению.

В подавляющем большинстве случаев в экономике таким "мешающим" фактором считают время, влияние которого проявляется в виде некоторой тенденции, которую предлагают выявить, описать математически и исключить из фактических наблюдений. Однако такая процедура глубоко ошибочна – ложная корреляция возникает тогда, когда исследователь по значениям коэффициента парной корреляции пытается выявить взаимозависимость.

Как следует из вывода коэффициента парной корреляции, он характеризует степень приближения зависимости между двумя факторами к линейной, если эта зависимость действительно существует. Если зависимости между факторами нет, то коэффициент парной корреляции показывает лишь на то, насколько график взаимозависимости будет приближаться к линейной форме, если анализируемые два показателя разместить на одном графике. Этот вывод означает, что маркетолог сначала должен высказать и обосновать гипотезу о наличии взаимосвязи между факторами, а потом использовать в качестве дополнительного, но отнюдь не решающего аргумента расчет значений коэффициента корреляции. Любой ряд, имеющий линейную тенденцию роста или убывания, будет коррелировать с любым другим возрастающим или убывающим рядом вне зависимости от того, есть ли между рядами действительная взаимосвязь или её нет[1].

Если же этот коэффициент окажется мал, то это говорит только о том, что зависимость (если ее наличие у исследователя не вызывает сомнений) имеет другой, более сложный нелинейный характер.

Покажем это на простом примере[2]. Пусть ряд {Yt} представляет собой синусоиду, изменяющуюся во времени, а ряд {Xt} представляет собой косинусоиду с той же амплитудой и фазой:

[1] Светуньков С.Г. Эконометрические методы прогнозирования спроса (на примере промышленной энергетики) /Под ред. Г.Л.Багиева. - М.: Изд-во МГУ, 1993. – С. 57.

[2] Светуньков С.Г. Количественные методы прогнозирования эволюционных составляющих экономической динамики. – Ульяновск: Изд-во УлГУ, 1999. - С.71

$$Y_t = sin(t), \ X_t = cosin(t).$$

Для 100 точек, снятых в течение одного периода этих гармонических функций, коэффициент парной корреляции составил величину, равную –0,0013674, что свидетельствует об очевидном отсутствии взаимосвязи между рядами. Тем не менее, между рядом $\{Y_t\}$ и рядом $\{X_t\}$ имеется нелинейная функциональная взаимосвязь, так как

$$sin(t) = cosin\ (\pi/2 - t)$$

Поэтому, если бы коэффициент парной корреляции действительно показывал наличие или отсутствие взаимосвязи, он в данном случае был бы равен единице.

К сожалению, интерпретация коэффициента парной корреляции, как мерила взаимосвязи вообще между случайными факторами встречается и в литературе по маркетинговым исследованиям, например [1]: «абсолютная величина коэффициента корреляции характеризует тесноту связи, а знак указывает на её направление». Ещё раз поэтому следует подчеркнуть, что он характеризует лишь степень приближения зависимости между двумя факторами к линейной, если эта зависимость действительно существует.

В том случае, когда приходится иметь дело со случайными процессами, любые расчётные величины являются выборочными и представляют собой некоторое приближение к истинному значению. Для того, чтобы определить насколько близко это выборочное значение к истинному, используют методику определения доверительных границ (интервалов). Не является исключением из этого правила и коэффициент парной корреляции, который вычисляется по выборочным значениям эмпирических данных, а, следовательно, сам является выборочным значением.

Исследования показали, что выборочные значения коэффициента парной корреляции не подчиняются нормальному закону распределения вероятностей. Поэтому непосредственное определение его доверительных границ с помощью дисперсии и выборочного значения коэффициента невозможно. Математическая статистика рекомендует для этого воспользоваться одним из двух подходов.

Первый подход заключается в вычислении величины

$$t_r = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}, \qquad (6.2.6)$$

[1] Голубков Е.П. Маркетинговые исследования: теория, методология, практика. – М.: Изд-во «Финпресс», 1998. - С.244.

которая имеет *t*–распределение с *n–2* степенями свободы[1]. Проверка нулевой гипотезы на отсутствие линейной корреляции в генеральной совокупности состоит в сравнении значения t_r с табличным значением t при заданной вероятности α и числе степеней свободы *n–2*. Если $t_r>t$, нулевую гипотезу нельзя принимать, а, следовательно, в генеральной совокупности существует линейная корреляция, достоверной характеристикой которой выступает выборочное значение коэффициента парной корреляции *r*. При этом, однако, открытым остаётся вопрос о том, в каких пределах находится истинное значение коэффициента парной корреляции для генеральной совокупности. Для ответа на этот вопрос используют второй подход.

Второй подход связан с Z-преобразованием Фишера, в соответствии с которым преобразование

$$Z=\frac{1}{2}\ln\frac{1+r}{1-r} \qquad (6.2.7)$$

даёт величину, распределение которой приближается к нормальному распределению с дисперсией, зависящей только от числа наблюдений[2]:

$$\sigma_Z{}^2=\frac{1}{n-3}. \qquad (6.2.8)$$

В случае использования данного метода для проверки эмпирического значения коэффициента парной корреляции вычисляется величина (6.2.7), а затем с помощью (6.2.8) и t-статистики Стьюдента при заданной доверительной вероятности определяются доверительные границы:

$$Z-t_\alpha\sigma_Z<M(Z)<Z-t_\alpha\sigma_Z\ , \qquad (6.2.9)$$

после чего, воспользовавшись (6.2.7), осуществляется обратное преобразование из *Z* в коэффициент парной корреляции *r*, для чего, как легко определить из (6.2.7), следует осуществить вычисление:

$$r=\frac{\ell^{2Z}-1}{\ell^{2Z}+1}. \qquad (6.2.10)$$

[1] Вайну Я.Я.-Ф. Корреляция рядов динамики. – М.: Статистика, 1977. – С.93.
[2] Кендалл М., Стьюарт А. Статистические выводы и связи. – М.: Наука, 1973. – С. 390.

При необходимости можно воспользоваться соответствующими таблицами преобразования[1], но в настоящее время такой необходимости не возникает, поскольку вычисление указанной величины не представляет никаких проблем даже с помощью карманных калькуляторов.

Если достоверность полученных результатов подтверждается подобными способами, исследователь убеждается в наличии линейной взаимосвязи между факторами и может приступить к моделированию.

Что делать в том случае, когда маркетолог предполагает наличие между факторами более сложной нелинейной зависимости? Анализ взаимосвязи с помощью коэффициента парной корреляции здесь бессмыслен. Хорошие результаты даёт применение коэффициента согласия в динамике[2], который анализирует соответствие конечных разностей двух исследуемых рядов, но его вычисление требует определённых навыков и математической подготовки, поэтому для широкой маркетинговой практики всё же рекомендуется применять экспертную оценку.

6.3. Моделирование зависимости между показателями

После того, как выявлена взаимосвязь и примерно определен ее характер (линейная, нелинейная, прямая, обратная), возникает необходимость ее описания с помощью математических моделей. Эта необходимость вызвана следующими соображениями:

- математическая модель компактно описывает большие массивы информации и вполне их заменяет,
- модель имеет расчетные коэффициенты, которые, как правило, имеют ясное математическое толкование,
- модель вскрывает присущие процессу закономерности, которые вполне могут быть использованы для целей интерполяции или экстраполяции.

Задачей построения математической модели занимается регрессионный анализ. По определению, регрессионный анализ - это раздел математической статистики, объединяющий практические

---

[1]Например: Длин А.М. Математическая статистика в технике. – М.: Советская наука, 1958. - С.432.

[2] Светуньков С.Г. Количественные методы прогнозирования эволюционных составляющих экономической динамики. – Ульяновск: Изд-во УлГУ, 1999. – С.69

методы исследования регрессионной зависимости между величинами по статистическим данным.

В общем случае регрессионный анализ можно представить в виде следующих взаимосвязанных этапов:
- исследование формы зависимости;
- определение параметров зависимости;
- определение значимости полученных соотношений.

В настоящее время для определения формы зависимости используются в основном два подхода: графический и аналитический.

В маркетинговой практике их используют последовательно: сначала строят графики зависимости с тем, чтобы определить класс возможных взаимосвязей, а затем аналитически определяют различные расчетные коэффициенты, характеризующие каждую из возможных зависимостей.

Наиболее разработанными и широко применяемыми в практике обработки экспериментальных данных являются три группы методов: классические, робастные и непараметрические. В практической эконометрии в основном применяются методы классического регрессионного анализа. При этом рассматривается параметрическая модель, коэффициенты которой определяются на некотором выборочном множестве динамических статистических наблюдений $\{Y_t;X_t\}$. При этом реальный ряд динамических наблюдений $\{Y_t\}$ описывают с помощью линии регрессии $\hat{Y}(X_t)$ следующим образом:

$$Y_t = \hat{Y}(X_t) + \varepsilon_t . \qquad (6.3.1)$$

где ε_t - ошибка аппроксимации, характеризующая совокупное влияние случайных факторов.

В регрессионном анализе предполагается, что случайная составляющая ε_t обладает следующими основными свойствами:

- она имеет нормальное распределение вероятностей,
- математическое ожидание этой случайной составляющей равно нулю,
- ее дисперсия конечна,
- сама случайная величина не подвержена автокорреляции.

Впрочем, современная математическая статистика дает решения и при нарушении некоторых из этих априорных предположений.

Несмотря на то, что методов оценки параметров регрессионных моделей достаточно много, в вычислительной практике наибольшее распространение получил метод наименьших квадратов (МНК). Это объясняется хорошей разработанностью аппарата, удобством в использовании и вполне понятными для неподготовленного исследователя свойствами оценок. Действительно, оценки МНК при свойствах случайной составляющей, перечисленных выше, обладают рядом замечательных особенностей[1]:

- оценки параметров являются несмещенными, т.е. их математическое ожидание равно искомому параметру;
- оценки параметров являются состоятельными, т.е. при увеличении объема выборки N они стремятся к своим истинным значениям;
- оценки параметров являются эффективными, т.е. имеют минимальную дисперсию.

Пусть некоторый исследуемый процесс имеет линейную тенденцию, описываемую в виде зависимости:

$$\hat{Y}(X_t) = a_0 + a_1 X_t. \quad (6.3.2)$$

Эта зависимость описывает реальный ряд наблюдений с некоторой ошибкой аппроксимации ε_t:

$$Y_t = \hat{Y}(X_t) + \varepsilon_t, \quad (6.3.3)$$

относительно которой не будем делать никаких априорных предположений.

При стандартном общепринятом подходе к оцениванию параметров эконометрической модели зависимости показателя Y_t от фактора X_t, саму задачу, представляя графически, изображают на плоскости этих двух факторов. Обосновывается это тем, что изменение расположения модели, при изменении ее параметров, демонстрируются очень наглядно. Сама прямая аппроксимирует эти точки с ошибкой аппроксимации ε_t.

Мерилом «степени точности» аппроксимации фактических данных расчетными могут являться ошибки аппроксимации:

$$\varepsilon_t = Y_t - \hat{Y}(X_t). \quad (6.3.4)$$

Чем меньше эти ошибки, тем точнее аппроксимация исходных данных. Можно предложить целый ряд критериев отбора лучших оценок параметров моделей, основанный на вычислении указанной разности и ее обработке.

Самым простым, оказалось, найти значение параметров модели для критерия:

[1] Маленво Э. Статистические методы эконометрии. Вып. 2. - М.: Статистика, 1976. - 325 с.

$$Q = \sum_{t=1}^{n} \varepsilon_t^2 \to \min. \qquad (6.3.5)$$

Так как при этом минимизируется сумма квадратов отклонений, а не сами отклонения, то влияние знака ошибки аппроксимации на результат оценивания сведено к нулю, поэтому оценки параметров, полученные таким образом, будут вполне приемлемыми, а кроме того, уже из самого критерия следует, что для этих оценок дисперсия (мера колеблемости) будет минимальна.

Функция Q зависит от двух переменных - a_0 и a_1. Известно, что функция достигает своего минимума в точке, в которой первые ее производные по переменным равны нулю:

$$\partial Q/ \partial a_0 = 0; \qquad \partial Q/ \partial a_1 = 0, \qquad (6.3.6)$$

После исчисления частных производных получается система двух уравнений с двумя неизвестными, которая получила название "система нормальных уравнений", имеющая вид:

$$\begin{cases} \Sigma Y_t = a_0 N + a_1 \Sigma X_t \\ \Sigma Y_t X_t = a_0 \Sigma X_t + a_1 \Sigma X_t^2 \end{cases} \qquad (6.3.7)$$

Метод наименьших квадратов позволяет найти оценки параметров такой линейной модели, для которой дисперсия отклонений расчетных значений от фактических ε_t будет минимальна.

Любые другие методы оценивания параметров модели будут иметь другую интерпретацию. Так, например, если в качестве критерия оценивания параметров взять минимизацию квадратов расстояний ε'_t между прямой и фактическими наблюдениями, то эти расстояния от точек до прямой будут характеризоваться отрезками, перпендикулярными к модели, и искомая в соответствии с новым критерием модель, пройдет на графике иначе.

В любом случае, при выборе способа оценивания параметров модели, разные значения оценок параметров модели соответствуют различному положению каждой прямой на плоскости факторов. Иначе говоря, на плоскости можно нанести бесконечное множество прямых, каждая из которых будет тем или иным образом описывать фактические значения и характеризоваться различными значениями дисперсии, суммы квадратов расстояний, средними абсолютными отклонениями и т.п.

Параметр a_0 линейной модели характеризует тот отрезок, который отсекает прямая на оси показателя Y_t, когда фактор X_t принимает значение, равное нулю; параметр a_1 характеризует тангенс угла наклона прямой линии к оси $0\text{-}X_t$.

Однако задачу можно представить по-другому. Действительно, задача построения регрессионной модели сводится по сути к поиску единственной пары оптимальных значений a_0 и a_1 из бесконечного множества значений этих параметров. То есть неизвестными переменными данной задачи являются параметры a_0 и a_1, а не значения Y_t и X_t, которые собраны в ходе маркетинговых исследований. Обозначим критерий оптимизации параметров модели (минимум суммы квадратов отклонений, минимум суммы абсолютных отклонений и т.п.) буквой Q. Очевидно, критерий достигает своего оптимального значения в точке, где первая производная функции этого критерия по переменным равна нулю. Это в общем случае дает основание найти расчетные формулы для определения численных значений коэффициентов модели.

Задачу построения модели возможно рассматривать и на плоскости параметров модели, а не только на плоскости факторов. При этом минимум, например, для оценок критерия МНК на плоскости a_0 - a_1 будет представлять собой точку Q', которая имеет координаты ($a_0';a_1'$). Во всех остальных точках данной плоскости первая производная функции по параметрам не равна нулю и значения критерия Q будут большими, чем в точке ($a_0';a_1'$). Множество точек, в каждой из которых критерий отбора имеет одно и то же значение $Q_i=const$, в большинстве случаев представляет собой замкнутые гладкие выпуклые кривые. По мере приближения значения критерия отбора Q к оптимальной величине, размер этих кривых уменьшается, как уменьшается значение критерия выбора ($Q3>Q2>Q1>Q*$). При достижении оптимума кривая превращается в точку, а значение критерия отбора принимает оптимальное значение $Q*$.

Впрочем, возможны и другие случаи изображения задачи оценивания на плоскости параметров. Для других критериев возможны случаи, когда экстремум функции достигается в двух и более точках.

Пусть, например, необходимо найти параметры линейной однофакторной модели, для которой минимальна следующая сумма:

$$Q' = \Sigma(\varepsilon_t^3 + \varepsilon_t^2 + \varepsilon_t) \rightarrow min. \qquad (6.3.4)$$

Находя частные производные этой функции по переменным, которыми в данном случае вновь являются параметры линейной однофакторной модели a_0 и a_1, получим, что для определения оптимальных данному критерию значений параметров необходимо решить систему двух нелинейных уравнений. В общем случае эта система нелинейных уравнений имеет два корня[1]:

Число критериев оценки параметров модели бесконечно. Оценки МНК являются лишь одной из возможных оценок параметров модели (6.3.1), представляемой в виде точки на плоскости параметров, причём оценка МНК вовсе не является лучшей оценкой модели. Особенно это касается моделирования нестационарных динамических процессов. Представление задачи оценивания параметров эконометрической прогнозной модели на плоскости параметров позволяет решить целый ряд интересных проблем, решение которых в стандартной постановке было бы невозможным. Так, становится вполне очевидным то обстоятельство, что уравнения системы МНК на плоскости параметров будут представлять собой две прямые, точка пересечения которых и является оценкой МНК.

Для нанесения уравнений прямых известной системы МНК на плоскость параметров a_0-a_1, необходимо привести эти уравнения к виду, называемому «уравнениями в отрезках». В рассматриваемом случае для этого левые и правые части уравнений делятся на их левые части, а константы, на которые умножаются неизвестные параметры a_0 и a_1, из числителя переносятся в знаменатель. В итоге подобных преобразований система нормальных уравнений МНК (6.3.7) в отрезках примет следующий вид:

$$1 = \frac{a_0}{\dfrac{\sum_t Y_t}{N}} + \frac{a_1}{\dfrac{\sum_t Y_t}{\sum_t X_t}}, \qquad (6.3.5)$$

$$1 = \frac{a_0}{\dfrac{\sum_t Y_t X_t}{\sum_t X_t}} + \frac{a_1}{\dfrac{\sum_t Y_t X_t}{\sum_t X_t^2}}. \qquad (6.3.6)$$

Полученные уравнения называются «уравнениями в отрезках», так как величины, находящиеся в знаменателях представляют собой ни что иное, как отрезки, отсекаемые прямыми линиями на осях плоскости, то есть на осях параметров. Эти уравнения будут более подробно рассмотрены в следующем параграфе при анализе явления мультиколлинеарности.

[1] Светуньков С.Г. Эконометрические методы прогнозирования спроса (на примере промышленной энергетики) - М.: Изд-во МГУ, 1993. – С. 49.

Нанеся указанные отрезки на оси a_0 и a_1, и соединив эти точки отрезками прямых, получим изображение искомых уравнений системы МНК на плоскости оценок параметров модели. Пересечением этих уравнений является точка Q^* с координатами $(a_0';a_1')$. Очевидно, что эти координаты и есть решение системы МНК, то есть - значения параметров линейной модели, найденных с помощью МНК.

Из элементарной геометрии известно, что через одну точку можно провести бесконечное множество различных прямых. Значит, ту же самую точку Q^* можно найти и разными другими путями, а прямые системы МНК являются лишь частью этого множества различных путей. МНК имеет важные преимущества по сравнению с множеством других методов оценивания параметров моделей:

- метод очень прост в применении,
- в большинстве пакетов прикладных программ статистической обработки данных используется МНК,
- оценки параметров модели, полученные с помощью МНК, имеют ясный статистический смысл.

К числу недостатков метода следует отнести его некоторую громоздкость, а также неустойчивость к ошибкам округления в определенных условиях. Тем не менее, МНК является наиболее часто используемым на практике методом построения регрессионных моделей. В том случае, когда выявленные закономерности носят нелинейный характер, в подавляющем большинстве из них удаётся привести эти нелинейные модели к линейному виду (6.3.2) посредством замены переменных, логарифмирования, экспонирования и т.п. Например, при оценивании параметров нелинейной производственной функции Кобба-Дугласа, представленной в виде модели:

$$Y = \alpha K^{\beta} L^{1-\beta}$$

логарифмируют модель по любому основанию и получают аддитивную форму записи модели:

$$\ln Y = \ln \alpha + \beta \ln K + (1-\beta) \ln L,$$

параметры которой легко найти с помощью МНК.

Аналогично поступают и с нелинейными моделями других форм. Правда, необходимо при этом иметь в виду, что полученные оценки будут смещены, и будут не совсем правильно отражать выявленную закономерность[1]. Тем не менее, это смещение в подавляющем большинстве случаев не оказывают особого влияния на результаты анализа или прогноза, которые выполняются с помощью построенных моделей.

На практике приходится иметь дело с выборочными значениями переменных Y_t и X_t , так как собрать все возможные значения случайных переменных, составляющих генеральную совокупность, невозможно. С учётом того, что маркетолог в этом случае работает не со всей совокупностью данный, а только лишь с некоторой выборкой из неё, то и математические модели, которые он построит, будут, поэтому являться лишь приближением к верному описанию моделируемой зависимости. Поэтому необходимо оценить, насколько полученной модели можно доверять. Для этого в математической статистике используют процедуру построения доверительных интервалов. В общем, виде она имеет следующий вид.

Расчётное значение $\hat{Y}$, является лишь одной из оценок истинного значения Y. Границы, в которых находится это истинное значение, может быть оценено с заданной доверительной вероятностью α по указанному расчётному значению $\hat{Y}$, дисперсии σ^2 и числу наблюдений, на основании которых вычислено $\hat{Y}$. Эти границы определяются интервалом:

$$\hat{Y} - t_\alpha \frac{\sigma}{\sqrt{n}} < Y < \hat{Y} + t_\alpha \frac{\sigma}{\sqrt{n}} . \qquad (6.3.8)$$

Здесь t_α определяется при заданной доверительной вероятности по имеющимся в распоряжении исследователей таблиц для интеграла вероятностей. Если число наблюдений n невелико, не превышает трёх десятков, вместо нормального закона распределения вероятностей, на основании которого и определяются интервалы (6.3.8), используется распределение Стьюдента[2].

Математическая статистика предлагает варианты построения доверительных границ для регрессионных моделей и с помощью процедуры оценивания границ для каждого коэффициента регрессии в отдельности и оценивании после этого совокупного доверительного интервала[3].

Для того чтобы определить, насколько корректно применение аппарата математической статистики в том или ином случае маркетинговых исследований и какова опасность возникновения ошибки

---

[1] Светуньков С.Г. Количественные методы прогнозирования эволюционных составляющих экономической динамики. – Ульяновск: Изд-во УлГУ, 1999. – С. 82 - 84.

[2]Дунин-Барковский И.В., Смирнов Н.В. Теория вероятностей и математическая статистика в технике. – М.: Гос. изд-во технико-теоретической литературы, 1955. - С. 277.

[3] См., например: Андерсон Т. Статистический анализ временных рядов. - М.: Мир, 1976. - 756с.

инструментария, о которой говорилось ранее, разберем более подробно некоторые основы математической статистики.

Статистические данные о развитии любой системы или явления сами по себе не являются достаточным источником научного познания. Необходимо, чтобы эти данные отражали некоторую закономерность развития или давали представление о вероятном направлении этого развития. Выяснением таких свойств наблюдений занимается теория вероятностей и, основанная на ней, математическая статистика.

Основой всех методов математической статистики является выборочный метод. Действительно, задачей математической статистики является разработка методов получения обоснованных выводов об исследуемых процессах и явлениях из данных наблюдений о них. Причем эти выводы относятся вовсе не к данным наблюдениям, а к более общим свойствам, характеризующим сам процесс или явление - утверждения о вероятностях, законах распределения, математических ожиданиях, дисперсиях и т.п. Все эти утверждения можно получить, лишь используя фактические данные наблюдений.

Пусть исследователю известен ряд наблюдений о расходах каких-либо материалов на единицу выпускаемой продукции по данному технологическому процессу. Очевидно, что исследователю известны не все возможные расходы, а только их часть, которая была выявлена в процессе расчета норм материалов. Эта информация, конечно же, представляет интерес при определении количества расхода материального ресурса на данном оборудовании при данном технологическом процессе, но ради этого нет смысла собирать всю статистику, делать расчеты и вообще нести какие-либо расходы. Все эти действия имеют смысл только тогда, когда исследователя интересует расходы материалов не по этому технологическому процессу, а по всем аналогичным процессам с подобной технологией производства. При этом можно говорить об удельном расходе материалов на единицу продукции в среднем для всего процесса.

В этом случае собранный материал рассматривается как некоторая выборка из множества различных возможных вариантов, которые встречаются при наблюдении массового явления или процесса в данной обстановке. Очевидно при этом, что выводы и оценки, основанные на ограниченном материале, могут считаться лишь приближенными оценками реального процесса.

Суть выборочного метода заключается в следующем.

Пусть имеется многочисленная совокупность однородных элементов, каждый из которых может обладать каким-либо признаком. Из этой совокупности выбирается наугад один из элементов, причем при выборе элемента из совокупности принимаются все меры к тому, чтобы вероятность быть выбранным была одинаковой для всех элементов. Группа из N элементов, наблюдаемых при испытаниях, называется в этом случае случайной выборкой, число элементов, отобранных для исследований - объемом выборки, а сам процесс отбора - простым случайным выбором.

Когда маркетолог обрабатывает результаты маркетинговых исследований потребителей, он имеет дело именно с выборочными значениями – из заданного сегмента рынка выбирается несколько потребителей, мнение которых представляет одну из возможных оценок мнения всех потребителей сегмента. Причём чем больше число изученных потребителей, тем точнее оценка мнения всех.

В этом случае выборочный метод методологически совместим с изучаемым объектом. Но если рассматривается изменение, например потребительских предпочтений во времени, то выборочный метод далеко не всегда будет методологически совместим со свойствами изучаемого объекта. Потребительские предпочтения в краткосрочном периоде не претерпевают принципиальных качественных изменений – изменения могут иметь только количественный характер. Но в случае среднесрочного и долгосрочного периодов потребительские предпочтения претерпевают и количественные, и качественные изменения. А последнее обстоятельство делает невозможным применение методологии выборочного метода. Значительная часть динамических процессов, протекающих в социально-экономических системах, имеет в качестве результата динамики качественное изменение и применение в этом случае классических методов математической статистики может привести к возникновению ошибки инструментария.

В общем случае можно выделить два типа динамики социально-экономических систем и процессов: динамика стационарная и нестационарная. Обработку статистических данных стационарных процессов следует осуществлять с помощью методов математической статистики. Для работы с рядами второго типа – нестационарной динамики – математическая статистика неприемлема.

Для обоснования этого вывода необходимо выяснить ряд моментов, которые проще сформулировать в виде вопросов[1].

Вопрос первый. Является ли совокупность статистических наблюдений нестационарной динамики социально-экономических систем однородной?

Вопрос второй. Являются ли все прошлые, настоящие и будущие статистические наблюдения нестационарной динамики социаль-

[1] Светуньков С.Г. Эконометрические методы прогнозирования спроса (на примере промышленной энергетики) /Под ред. Г.Л.Багиева. - М.: Изд-во МГУ, 1993. – С. 37.

но-экономических систем, меняющихся в том числе и качественно, генеральной совокупностью?

Вопрос третий. Является ли выборка статистических наблюдений нестационарной динамики социально-экономических систем случайной выборкой?

Получив ответ на эти вопросы, мы сделаем вывод о том, насколько применим выборочный метод в данном случае. Если все три раза ответ будет утвердительным, то применение методов математической статистики методологически оправдано. Если хотя бы на один из этих вопросов будет получен отрицательный ответ, то выборочный метод оказывается методологически не пригодным для обработки статистических данных такого процесса.

Для ответа на первый вопрос воспользуемся некоторыми размышлениями и выводами, сделанными С.А.Смоляк и Б.П.Титаренко[1]. Отмечая сложность попытки дать однозначное толкование термину «однородная выборка», они вводят «рабочее» определение, удобное для практического применения в их работе - однородной называется такая совокупность, элементы которой формируются под воздействием общих основных причин и условий, а их законы распределения имеют простую структуру. Очевидно, что неоднородной будет являться совокупность наблюдений, элементы которой формируются под влиянием разных причин и условий и законы распределения имеют сложную структуру.

Так, например, если динамика некоторого показателя X описывается моделью постоянного прироста со случайными отклонениями

$$X_t = X_{t-1} + a + e_t, \qquad (6.3.9)$$

где a - постоянный прирост,

e_t - не зависящие друг от друга случайные отклонения, имеющие нормальный закон распределения с нулевым математическим ожиданием и конечной дисперсией, то все элементы совокупности $X_1, X_2, ...X_T$ имеют различные законы распределения, но сама совокупность, в соответствии с «рабочим» определением, будет однородной.

Сложность использования такого «рабочего» определения однородности совокупности заключается в том, что понятия «общие причины и условия» недостаточно конкретны. Например, рассматривая динамику развития промышленности нашей страны на протяжении тридцати лет, можно говорить, что основной причиной формирования этого динамического ряда является спрос на товары про-

[1] С.А.Смоляк, Б.П.Титаренко. Устойчивые методы оценивания (статистическая обработка неоднородных совокупностей). - М.: Статистика, 1980. - 208 с.

мышленности, а условием – наличие трудовых и материальных ресурсов. На основе этого можно утверждать об однородности совокупности. Если же говорить о том, что и причина - спрос на продукцию, и условия – структура и количественные характеристики материальных и трудовых затрат, непрерывно меняются и качественно, и количественно, то необходимо будет признать, что в соответствии с этим определением совокупность будет являться неоднородной.

Воспользуемся другим определением однородности[1]. Там определялась однородность при рассмотрении экономических систем с помощью понятий и подходов кибернетики. В нашем понимании совокупность будет являться однородной только в том случае, когда ее элементы $\{Y_t\}$ формируются под воздействием общих основных неизменных причин и условий так, чтобы при возникновении условий и причин, равных $X_t = X_1$ элемент Y_t будет равен

$$Y_t = Y_1 + e_t, \qquad (6.3.10)$$

где e_t - не зависящие друг от друга случайные отклонения, имеющие нормальный закон распределения с нулевым математическим ожиданием и конечной дисперсией.

Это определение является более приемлемым и более точным. Используя его для случая (6.3.9), можно убедиться в том, что все элементы совокупности $X_1, X_2, \ldots$ однородны. Действительно, при достижении $X_t = X_1$, величина X будет определяться как

$$X_t = X_1 = X_0 + a + e_t,$$

что полностью соответствует определению однородности.

Теперь можно с полным основанием ответить на первый вопрос об однородности совокупности статистических наблюдений нестационарной динамики показателей социально-экономических систем. Условия и причины формирования этих показателей непрерывно меняются во времени, так как меняются и внутренняя структура системы, и ее взаимосвязи с другими системами. В результате этого, например, если предложить промышленности России выпустить продукцию в количестве и качестве, соответствующем 1960 г. $(X_t = X_1)$, то количество потребленного, например, при этом угля Y_t будет значительно отличаться от его потребления, соответствующего 1960 г. (Y_1), причем

$$Y_t - Y_1 = e_t + u_t,$$

[1] Светуньков С.Г. Эконометрические методы прогнозирования спроса (на примере промышленной энергетики) /Под ред. Г.Л.Багиева. - М.: Изд-во МГУ, 1993. – 123 с.

где u_t - составляющая, вызванная результатами нестационарного развития, в том числе непрерывного воздействия научно-технического прогресса.

Таким образом, можно сделать однозначный вывод - совокупность наблюдений нестационарной динамики экономических систем не является однородной.

На второй вопрос также несложно ответить. По определению, генеральная совокупность - это совокупность однородных элементов, характеризующаяся некоторой функцией распределения, математическим ожиданием, дисперсией и т.п. С учетом того, что наблюдения неоднородны, они не относятся к некой генеральной совокупности. Значит, генеральной совокупности для экономической динамики не существует.

Кстати, если предположить обратное, то мы вынуждены будем признать наличие некоторой раз и навсегда заданной величины, называемой математическим ожиданием. Даже если ее представить в виде некоторой очень сложной функции времени, то предположение о наличии некоторой заданной, не меняющей в зависимости от изменяющегося мира своей структуры функции, означает статический, а не динамический подход. А такой подход современной экономикой отвергается.

Ответим теперь на третий вопрос, - каким образом получена выборка? Элементы выборки должны быть получены путем случайного выбора из некоторого множества потенциально возможных значений. Насколько случайны данные, приведенные в любой статистической таблице экономической динамики? Взглянув на цифры таблицы, мы увидим, что все они упорядочены, следуют друг за другом подряд и выбраны отнюдь не случайным образом, а на основе систематизации и агрегировании данных. Таким образом, можно сделать вывод о том, что полученная выборка не является случайной, и, строго говоря, не является выборкой как таковой.

Всё сказанное означает, что свойства статистических данных нестационарной динамики социально-экономических систем совершенно не соответствуют основным требованиям выборочного метода, поэтому можно сделать вывод о том, что методы, основанные на использовании выборочного метода, непригодны для моделирования нестационарных процессов. В большинстве случаев для прикладных целей маркетинговых исследований в этом случае пригодны методы адаптивного моделирования, часть из которых будет рассмотрена в следующей главе.

Всё вышесказанное говорит о том, что маркетолог, собравший необходимую маркетинговую информацию, прежде чем приступить к её математической обработке, должен тщательно выбрать инструмент этой обработки. Только в том случае, когда используемый математический аппарат является методологически совместимым со свойствами обрабатываемой совокупности маркетинговой информации, можно избежать появления ошибки инструментария, вносящей существенные искажения в результаты агрегирования данных.

6.4. Моделирование в условиях мультиколлинеарности

При моделировании социально-экономической динамики объективно приходится иметь дело с многофакторными зависимостями, когда значение показателя или группы показателей определяется поведением нескольких факторов. Любая однофакторная модель в этой ситуации настолько условна, что ее применение может давать лишь очень приближенные ориентиры, а порой просто ошибочные результаты. Однако применение многофакторных моделей вовсе не гарантирует их полную адекватность реальной ситуации - основной проблемой, возникающей при построении многофакторных моделей, является наличие мультиколлинеарности при оценивании параметров модели и как следствие этого - неустойчивость оценок параметров.

Мультиколлинеарность, как следует из самого названия проблемы, возникает тогда, когда факторы модели имеют одинаковые, монотонные относительно друг друга тенденции в динамике. Именно в этом случае и возникает ряд проблем, которые ученые, занимающиеся ею, видят в следующем.

Во-первых, возникают осложнения при вычислениях, так как при построении многофакторных моделей приходится работать с матрицами, а при мультиколлинеарности появляется эффект слабой обусловленности матрицы системы нормальных уравнений - её определитель очень близок к нулю.

Во-вторых, оценки параметров многофакторных моделей будут очень неточными из-за указанной выше причины.

В-третьих, так как оценки параметров оказываются неточными, то интерпретация влияния факторов на прогнозируемый показатель будет совершенно не той, которая есть на самом деле.

В-четвертых, оценки параметров модели оказываются крайне неустойчивыми - малейшее изменение исходных данных или даже ошибки округления, приводят к очень значительным изменениям параметров.

В-пятых, ценность таких моделей крайне низка, так как неустойчивая модель дает очень сильную вариацию расчётных значений.

О наличии мультиколлинеарности судят в основном по матрице парных коэффициентов корреляций между факторами модели. Если коэффициенты парной корреляции этой матрицы превышают величину 0.8, то считается, что модель приходится строить в условиях мультиколлинеарности. Поэтому на первоначальном этапе построения многофакторных моделей проверяется условие наличия мультиколлинеарности - строится корреляционная матрица, и если ее значения оказываются близкими к единице, считается, что последующее построение многофакторных моделей приходится осуществлять в условиях мультиколлинеарности.

Для борьбы с этим явлением и его последствиями на практике используют следующие подходы:

- преобразуют факторы в новые переменные, уменьшая тем самым количество переменных (факторный анализ),
- используют методы регуляризации (например, гребневой регрессии, стабилизированных оценок или "сжатых" оценок),
- исключают из совокупности факторов одну или несколько линейно связанных факторных переменных, чтобы вновь полученные элементы корреляционной матрицы были меньше порогового значения 0,8.

Каждый из этих подходов обладает очень серьезными недостатками, которые снижают эффективность их применения.

В первом случае с помощью методов факторного анализа пытаются заменить реальные переменные на те, которые являются скрытой причиной их динамики. Такие переменные неизвестны исследователю и называются латентными. С помощью математических преобразований (наиболее известен метод главных компонент) находят эти новые переменные и строят новую многофакторную модель. При этом, однако, модель теряет какой-либо смысл экономический и становится малопригодной, например, для прогнозирования, так как латентные переменные в данной ситуации не существуют и являются математической абстракцией. Для вычисления каждой из этих новых переменных приходится использовать формулы преобразования с учётом исходных переменных. Всё это вносит дополнительную погрешность в вычисление результирующего признака.

Второй подход представляет собой некоторую попытку «утяжеления» задачи посредством намеренного введения «засоряющей» величины - если какая-либо конструкция неустойчива, то повысить ее устойчивость можно, загрузив конструкцию балластом[1]. В случае многофакторного моделирования построенная таким образом модель, в результате такого утяжеления, будет не самой оптимальной, но вполне устойчивой, что уже следует признать положительным моментом. Этот путь решения проблемы мультиколлинеарности более приемлем, чем первый, однако, он требует от исследователя -маркетолога хорошей математической подготовки, так как аппарат методов регуляризации достаточно сложен. Поэтому на практике предпочитают использовать третий подход.

Сама процедура исключения из совокупности факторов одной или нескольких факторных переменных, весьма сомнительна - так можно последовательно прийти и к простой однофакторной линейной модели. И с экономических позиций эта процедура не может не вызвать недоумения - если специалист отобрал совокупность факторов, объясняющих динамику показателя, то уменьшение их числа только ухудшит свойства модели.

Таким образом, ни один из предлагаемых подходов нельзя признать удовлетворительным и учёные продолжают развивать первые два направления с целью получения более простых и эффективных способов построения многофакторных моделей[2]. Как это не кажется парадоксальным, но для случая нелинейной многофакторной модели эта проблема достаточно просто решается! Если прогнозист использует в качестве прогнозной модели нелинейную многофакторную модель, например, такую:

$$Y'_t = a_0 + a_1 X_{1t} X_{2t} + a_2 X_{1t} X_{2t} X_{3t} + a_3 X^2_{3t},$$

то для нахождения параметров a_0, a_1, a_2 и a_3 осуществляют замену переменных и представляют модель в простой линейной форме:

$$Y'_t = a_0 + a_1 X'_{1t} + a_2 X'_{2t} + a_3 X'_{3t},$$

где $X'_{1t} = X_{1t} X_{2t}$,

$X'_{2t} = X_{1t} X_{2t} X_{3t}$,

$X'_{3t} = X^2_{3t}$.

Для вновь полученной модели и новых переменных строят корреляционную матрицу и если преобразованные новые переменные коррелируют друг с другом и коэффициенты парной корреляции выше по абсолютному значению, чем 0.8, то предлагается для уменьшения взаимной коррелированности переменных центриро-

[1] Дадаян В., Бессараб С. Модель долгосрочного экономического роста // Российский экономический журнал. – 1992, № 4. – С 78.

[2] Седелёв Б.В. Системные свойства объектов и принципы согласования в эконометрии // Известия АН СССР, серия экономическая. - 1991, № 4. – С.15.

вать исходные переменные X_{it}. Действительно, корреляционная матрица новых переменных сильно меняет свои значения в сторону уменьшения коэффициентов парных корреляций. Это доказал еще в 1969 году Г.Ф.Филаретов[1]. Однако, для линейной многофакторной модели центрирование исходных данных (как и любые другие аффинные преобразования) не изменит коэффициентов корреляционной матрицы. Поэтому сложилась парадоксальная ситуация - сложные многофакторные нелинейные модели в условиях мультиколлинеарности можно построить достаточно просто, а элементарные линейные многофакторные модели в этих же условиях построить не удаётся!

Проблему построения моделей в условиях мультиколлинеарности можно рассматривать с двух позиций:

1) параметры многофакторной модели оказываются неустойчивыми из-за того, что факторы мультиколлинеарны,

2) параметры неустойчивы в условиях мультиколлинеарности из-за особенностей применяемого метода их оценивания.

Первая точка зрения превалирует в научной методологии и ее разработка зашла в «тупик». Вторая точка зрения оказывается значительно более продуктивной. Для ее реализации следует воспользоваться представлением уравнений системы МНК как пересечение гиперплоскостей в гиперпространстве коэффициентов модели. Каждое из уравнений системы МНК может быть представлено в этом гиперпространстве как уравнение гиперплоскости. Естественно, графически такую процедуру при числе коэффициентов модели больше трех изобразить невозможно.

Если в однофакторном случае оценки МНК представляют собой точку пересечения на плоскости параметров двух прямых условий МНК, то при числе факторов, равном двум, а числе коэффициентов, равном трем, задачу оценивания следует рассмотреть не на плоскости, а в трехмерном пространстве параметров. Действительно, число неизвестных параметров становится равным трем и их можно изобразить в качестве осей трехмерного пространства. В этом случае условия МНК представляют собой систему из трех уравнений с тремя неизвестными, причем каждое из уравнений представляет собой не что иное, как уравнение плоскости в пространстве. Таким образом, решение системы МНК в данном случае будет представлять собой точку пересечения трех плоскостей в пространстве. Координаты этой точки дают искомые значения коэффициентов мо-

---

[1] Филаретов Г.Ф. К вопросу о построении нелинейной регрессионной модели по данным пассивного эксперимента // Проблемы планирования эксперимента / Под ред. Г.К.Круга. – М.: Наука, 1969.

дели. Уравнения системы МНК описывают плоскости, которые оставляют свои «следы» на каждой из ортогональной плоскостей пространства a_0-a_1, a_0-a_2 и a_1-a_2. Найти графически эти следы совсем не сложно. Для этого необходимо привести систему уравнений к виду в отрезках, как это было сделано в предыдущем параграфе с однофакторной линейной моделью формулами (6.3.5) и (6.3.6), затем найти отрезки на каждой из осей a_0, a_1 и a_2, и соединить их прямыми линиями.

В многофакторном случае система МНК дает гиперплоскости в гиперпространстве оценок параметров многофакторных моделей. Оценки МНК в этом случае будут представлять собой координаты точки в гиперпространстве коэффициентов модели.

Попробуем разобраться в сути эффекта мультиколлинеарности. Для этого воспользуемся статистическими данными таблицы 6.2, в которой приведены некоторые показатели экономической динамики одного из регионов России[1]. Построим многофакторную линейную модель её развития. Она имеет вид:

$$Y_t = -1.8798 + 0.0756X_{1t} - 0.0896X_{2t} + 2.3768X_{3t} + 0.5135X_{4t}.$$

Если попытаться дать экономическое толкование полученным результатам, то мы вынуждены будем утверждать, что увеличение выпуска промышленной продукции было обусловлено положительным влиянием всех факторов за исключением стоимости основных производственных фондов, которая влияла отрицательно на рост производства. Очевидно, что полученный вывод, как и результат не имеют смысла, хотя данная модель хорошо описывает имеющийся ряд наблюдений.

Таблица 6.2.Относительные значения показателей развития организации

Производство продукции, $У_t$	Потребление электроэнергии, X_{1t}	Основные производственные фонды,X_{2t}	Численность занятых, X_{3t}	Фонд оплаты труда, X_{4t}
1.0000	1.0000	1.0000	1.0000	1.0000
1.0410	0.9480	1.0994	1.0145	1.0374
1.0910	0.9800	1.2108	1.0323	1.0878
1.1390	1.1338	1.3346	1.0403	1.1267
1.1902	1.1380	1.4840	1.0565	1.1778
1.2438	1.1335	1.6759	1.0780	1.2276
1.3408	1.4320	1.8039	1.0968	1.3105

[1] Светуньков С.Г., Благов А.А., Инешин К.А., Козлов М.А. Развитие Ульяновской области: проблемы моделирования : Ученые записки экономического факультета // Внутривузовский сборник фМГУ, 1991 г. – С.3.

1.4293	1.5980	1.9623	1.1127	1.3978
1.4936	1.7681	2.0949	1.1109	1.5018
1.5220	2.1061	2.2011	1.1029	1.6133

Если теперь значения показателей табл.6.2 округлить не до четвертого знака после запятой, а до второго и вновь построить многофакторную модель по округленным значениям, она примет вид:

$$Y_t = -1.2152 + 0.3878X_{1t} + 0.1098X_{2t} + 2.2309X_{3t} - 0.4784X_{4t}$$

Если же на основе полученных теперь значений коэффициентов модели попытаться провести анализ, то из него будет следовать, что на рост производства все факторы оказывали положительное влияние за исключением оплаты труда. Она не была связана с конечными результатами труда, и необходимо пересмотреть принципы её исчисления. Выводы, как легко следует из проведённого сравнения двух моделей, оказываются противоречивыми. Таким образом, мы столкнулись с классическим случаем мультиколлинеарности, когда полученные результаты оказались неустойчивыми к незначительным ошибкам округления и привели к бессмыслице. Для подтверждения этого, необходимо построить корреляционную матрицу. Расчет коэффициентов парной корреляции дает следующие значения:

$$r_{12} = 0.9283, \quad r_{13} = 0.8355, \quad r_{14} = 0.9774,$$
$$r_{23} = 0.9736, \quad r_{24} = 0.9811, \quad r_{34} = 0.9194.$$

Из чего следует, что при формальном стандартном подходе многофакторная модель вообще вряд ли может быть успешно построена - все коэффициенты парной корреляции между переменными превышают пороговое значение, равное 0,8 .

Воспользуемся, уравнениями в отрезках для того, чтобы более тщательно изучить причины неустойчивости оценок параметров в условиях мультиколлинеарности. Используя систему МНК для абсолютных значений показателей развития области, приведем эту систему к форме уравнений в отрезках. Такое преобразование системы покажет нам отрезки, которые отсекают гиперплоскости условий многофакторного МНК на осях гиперпространства параметров (табл.6.3).

Таблица 6.3. Отрезки на осях гиперпространства параметров, отсекаемые уравнениями МНК

Оси гиперпространства	a_0	a_1	a_2	a_3	a_4
Первая гиперплоскость	0.2278	0.433	0.853	0.1379	0.301
Вторая гиперплоскость	0.2195	0.449	0.876	0.1339	0.303
Третья гиперплоскость	0.2198	0.445	0.876	0.1343	0.302
Четвертая гиперплоскость	0.2266	0.434	0.856	0.1373	0.302
Пятая гиперплоскость	0.2229	0.441	0.867	0.1356	0.302

Из анализа величин полученных отрезков следует, что отрезки, отсекаемые гиперплоскостями условий МНК на каждой из осей гиперпространства коэффициентов, практически совпадают, т.е. сами гиперплоскости почти параллельны друг другу. Очевидно, поэтому, что решение системы МНК, которое представляет собой точку пересечения этих практически параллельных друг другу гиперплоскостей в гиперпространстве, является чрезвычайно неустойчивым - малейшая ошибка в округлении может привести к тому, что гиперплоскости могут, переместясь незначительно, иметь новую точку их пересечения значительно удаленную от первоначальной. Например, отрезок на оси a_4 первой гиперплоскости величиной 0.301 может стать равным 0.303 (в результате округления, например). При этом первая гиперплоскость практически незаметно изменит свое положение в пространстве, но решение системы МНК, как точка пересечения гиперплоскостей, меняется так, что искажается не только абсолютная величина коэффициентов многофакторной модели, но и сам знак этих коэффициентов, что и было продемонстрировано выше на конкретном примере.

Следовательно, последствия мультиколлинеарности действительно вызваны неприемлемостью существующего алгоритма оценивания параметров многофакторных моделей. Исследования в этом случае должны быть направлены не на борьбу с объективно существующей реальностью - сильной коллинеарностью практически всех показателей и факторов экономической динамики (что подтверждается явлением ложной корреляции), а на улучшение используемого аппарата оценивания таких моделей. Этот вывод имеет место и при построении простых линейных однофакторных моделей, правда неустойчивость полученных оценок, в отличие от многофакторного случая, не особенно заметна - в последнем случае ситуация усугубляется еще и вычислительными сложностями, поэтому и становится такой явной.

Для повышения устойчивости оценок параметров многофакторных моделей необходимо «развести» гиперплоскости системы нормальных уравнений МНК и устранить тем самым их практическую параллельность. Это, в свою очередь, означает, что отрезки на осях гиперпространств оценок параметров моделей должны в максимально возможной степени отличаться друг от друга, а не совпадать до такой степени, как это происходит всегда в условиях мультиколлинеарности.

Вычисления показывают, что самый устойчивый вариант оценивания эконометрических моделей в условиях мультиколлинеарности будет тот, когда углы между гиперплоскостями условий МНК в многофакторном случае оценивания многофакторных эконометрических моделей будут равны 90 градусам (гиперплоскости перпендикулярны друг другу). Легко показать, что перпендикулярность гиперплоскостей, определяющих решение задачи оценивания параметров эконометрической модели, означает, что сумма углов пересечения этих гиперплоскостей с любой осью гиперпространства должна быть равна 180 градусам, а произведение тангенсов этих углов - минус единице[1]. Из этого условия следует, что для повышения устойчивости оценок МНК в условиях мультиколлинеарности необходимо осуществить перед построением многофакторной модели элементарное действие – центрирование всех переменных и использовать уже центрированные переменные для оценивания параметров.

Следует напомнить, что операция центрирования представляет собой такое преобразование исходного ряда $\{Y_t\}$, при котором из каждого значения ряда вычитается средняя арифметическая этого ряда:

$$Y_t^{'} = Y_t - \overline{Y}$$

С учетом того, что и для нелинейных многофакторных моделей именно эта процедура является наиболее эффективной, следует отметить, что предварительное центрирование данных при построении многофакторных моделей следует обязательно выполнять. Построенные при этом модели будут точными и устойчивыми.
Продемонстрируем эффективность предлагаемого подхода на данных примера, приведенного ранее. Для этого следует лишь центрировать исходные данные, что и было сделано. Если теперь строить многофакторную линейную модель с помощью нормированных данных Y'_t и X'_{it}, получим следующую многофакторную модель:

$$Y'_t = 0.15633\, X'_{1t} + 0.28614\, X'_{2t} + 1.01790\, X'_{3t} - 0.14803\, X'_{4t}.$$

Если теперь для построения многофакторной модели воспользоваться не исходными данными, а округленными в ней значениями

[1] Светуньков С.Г.. Моделирование в условиях мультиколлинеарности // Промышленная энергетика, 1994, № 6. – С. 28 - 32

до второго знака после запятой (как мы делали ранее без нормирования), получим такую модель:

$$Y'_t = 0.15634\, X'_{1t} + 0.28684\, X'_{2t} + 1.01289\, X'_{3t} - 0.14855\, X'_{4t}.$$

Как видно из сравнения двух последних многофакторных моделей оценки их параметров оказались достаточно устойчивыми в условиях мультиколлинеарности и изменяются лишь в четвертом или пятом знаках после запятых, в то время как до центрирования коэффициенты модели менялись в несколько раз, в том числе и меняя свои знаки.

Таким образом, для построения любых многофакторных моделей в условиях мультиколлинеарности следует осуществлять предварительное центрирование исходных данных, а затем использовать любой из множества методов оценивания параметров эконометрических моделей, в том числе и МНК.

Глоссарий к шестой главе

Агрегирование – объединение, укрупнение показателей по какому-либо признаку. Агрегирование применяется в целях уменьшения объёма информации до приемлемых для использования в вычислениях и анализе размеров. Так, например, информация о всех потребителях одного сегмента преобразуется в модель сегмента, которая характеризуется рядом агрегированных показателей, таких как доход потребителей сегмента, возраст, объёмы приобретения товаров, структура потребления и т.п.

Аддитивность – свойство величин, состоящее в том, что значение величины, соответствующее целому объекту, равно сумме значений величин, соответствующих его частям, каким бы образом ни был разбит объект. Система аддитивна, если она равна сумме тех же характеристик для составляющих систему подсистем и элементов. Свойство аддитивности широко используется в моделировании экономики, поскольку предположение об аддитивности изучаемых процессов существенно облегчает процесс построения моделей. Аддитивные модели, в отличие, например, от моделей в мультипликативной форме, значительно проще построить, а их параметры оказывается легче рассчитать. К сожалению, аддитивные системы и процессы, обладающие этим свойством, встречаются в экономической практике очень редко. Чаще всего экономическая система является результатом очень сложного нелинейного с распределёнными лагами взаимодействия составляющих систему подсистем и элементов. Поэтому аддитивные модели следует рассматривать только в качестве первого приближения к описанию экономической действительности.

Адекватность – соответствие модели моделируемому объекту. В случае неадекватности модели она не рассматривается, поскольку с её помощью невозможно получить необходимые сведения об объекте. В то же время необходимо помнить, что адекватность не может быть абсолютной, поскольку любая модель есть лишь приближённое отображение действительности, основанное на известной степени абстракции. Поэтому степень адекватности может быть различной, и для признания или непризнания адекватности модели используют различные проверочные критерии и оценки (в математической статистике, например, t-статистика Стьюдента). В конечном итоге мерилом адекватности любой модели является соответствие её значений реальной экономической практике.

Апостериори – термин, означающий знание, полученное из опыта. По сути, представляет собой один из элементов индуктивного рассуждения и доказательства: поскольку изучить все возможные эмпирические случаи, как правило, оказывается невозможно, на основе полученных априорных знаний может делаться вывод о свойствах явления в целом. Априорный и апостериорный подходы, также как индукция и дедукция в науке, наибольший эффект приносят при совместном использовании. Модель, построенная по фактографическим данным, имеет апостериорный характер, если при построении модели не высказывалось предварительных предположений.

Априори – термин, означающий знание, полученное до или независимо от опыта. Априорное знание получается путём логических рассуждений, с использованием методов дедукции и индукции. Для получения априорных знаний широко используется аксиоматический подход. Гипотезы, предположения, некоторые посылки любой теории носят априорный характер. В практике маркетинга к априорным допущениям прибегают при невозможности получения апостериорных знаний, например, при использовании в процессе освоения новых рынков метода аналогий. В процессе сегментирования рынка также часто прибегают к априорным положениям, например, при выдвижении гипотезы о признаке сегментирования рынка. В процессе практической реализации сегментирования априорные положения уточняются с помощью апостериорной информации.

Верификация – подтверждение, способы эмпирического обоснования знания, оценка достоверности или точности полученных результатов.

Вероятность – математическая, числовая характеристика степени возможности появления какого-либо события в тех или иных определённых, могущих повторяться неограниченное число раз условиях. Формы проявления вероятности описываются с помощью различных моделей, называемых в теории вероятностей и математической статистики «законами распределения вероятностей». Чаще всего на практике приходится сталкиваться с так называемым «нормальным распределением» вероятностей, когда случайных факторов очень много, а математическое ожидание их суммарного воздействия на событие равно нулю. Иногда приходится встречаться с ситуациями, когда термин «вероятность» используется как синоним понятия «возможность», например: «вероятно, что цена на товар повысится». Такая замена в теории и практике маркетинга недопустима, поскольку вероятность имеет математический и числовой смысл, она может быть измерена как степень возможности появления события.

Выборочный метод – основной метод математической статистики, который заключается в следующем. Пусть имеется многочисленная совокупность однородных элементов, каждый из которых может обладать каким-либо признаком. Из этой совокупности выбирается наугад один из элементов, причем при выборе элемента из совокупности принимаются все меры к тому, чтобы вероятность быть выбранным была одинаковой для всех элементов. Имеющееся множество элементов называют генеральной совокупностью. Группа из N элементов, наблюдаемых при испытаниях, называется в этом случае случайной выборкой, число элементов, отобранных для исследований - объемом выборки, а сам процесс отбора - простым случайным выбором. По выборочным наблюдениям имеется возможность судить о свойствах генеральной совокупности, избежав тем самым необходимости собирать всю информацию об элементах генеральной совокупности, что связано в большинстве случаев с колоссальными затратами. Доказательства, опирающиеся на выборочный метод, относятся к группе индуктивных доказательств. Подавляющее большинство разделов и методов математической статистики опирается на выборочный метод. Маркетологу приходится иметь дело с выборочным методом при анализе сегментов рынка, прогнозировании отдельных показателей, выявлении причинно-следственных связей.

Генеральная совокупность - это полная совокупность всех однородных элементов, характеризующаяся некоторой функцией распределения, математическим ожиданием, дисперсией и т.п. В научной практике собрать информацию обо всех элементах генеральной совокупности не представляется возможным из-за высоких затрат времени и средств на выполнение этой работы. Поэтому о свойствах генеральной совокупности судят по свойствам выборочной совокупности.

Динамический подход – способ изучения объектов и явлений, рассматривающий их в процессе изменения во времени самих объектов, явлений, их элементов и соотношений между ними. С помощью динамического подхода выявляются закономерности в ходе анализа самих изменений. Для динамического подхода основной предпосылкой анализа является признание непрерывности изменений количественных и качественных характеристик процессов, взаимосвязей и показателей во времени. "В тех случаях, когда элементы экономической жизни или их связи подвергаются изменениям, не исчерпывающимся изменением их числа, объема и вообще не сводимых к количественным изменениям, мы говорим о наличии качественных изменений. Сюда относятся, например, изменения в технике производства, в организации хозяйства, в составе и характере общественных потребностей и т.д." (Н.Д.Кондратьев). Качественные изменения, приводящие к необратимости изменений структуры систем во времени, являются основной причиной развития систем. Изучение только количественных изменений без их связи с качественными изменениями характерно для статического подхода.

Доверительная вероятность – показатель математической статистики, означающий вероятность, с которой случайная величина может принимать свои значения в пределах, определённых доверительными границами. Доверительная вероятность определяется как единица минус уровень значимости. В экономике чаще всего используют уровень доверительной вероятности в 0,95 (95%) или 0,97 (97%).

Доказательство – процедура обоснования истинности некоторого утверждения путём приведения тех истинных утверждений, из которых оно логически следует.

Закон – внутренняя существенная и устойчивая связь явлений, обусловливающая их упорядоченное изменение. На основании знания закона возможно достоверное предвидение течение процесса.

Закономерность – совокупность взаимосвязанных по содержанию законов, обеспечивающих устойчивую тенденцию или направленность в изменениях системы.

Идентификация – процесс отождествления некоторой модели объекту. Этот процесс принимает самые различные формы в зависимости от типа используемой модели и характера модели. В криминалистике, например, для идентификации используются отпечатки пальцев, сравнение генетических кодов и т.п. В экономике одним из наиболее часто используемых способов идентификации выступает верификация модели.

Инерционность экономического объекта - свойство объекта, претерпевая количественные изменения под воздействием внешних факторов, незначительно и постепенно изменять при этом свою структуру, направление и степень взаимосвязи между элементами системы. Именно присущая экономическим объектам инерционность позволяет прогнозировать основные характеристики их изменения, осуществлять целеполагания, принимать стратегические решения.

Интерполяция – процесс определения промежуточных данных в неполной статистической базе на основе имеющихся данных.

Мультиколлинеарность – явление множественного роста или убывания факторов, влияющих на некоторый результат. При построении многофакторных эконометрических моделей в условиях мультиколлинеарности возникает ряд проблем. Во-первых, возникают осложнения при вычислениях, так как при построении многофакторных моделей приходится работать с матрицами, а при мультиколлинеарности появляется эффект слабой обусловленности матрицы системы нормальных уравнений - ее определитель очень близок к нулю. Во-вторых, оценки параметров многофакторных моделей будут очень неточными из-за указанной выше причины. В-третьих, так как оценки параметров оказываются неточными, то интерпретация влияния факторов на результируемый показатель будет совершенно не той, которая есть на самом деле. В-четвертых, оценки параметров модели оказываются крайне неустойчивыми - малейшее изменение исходных данных или даже ошибки округления, приводят к очень значительным изменениям параметров. В-пятых, прогнозная ценность таких моделей крайне низка, так как неустойчивая модель дает очень сильную вариацию прогнозных значений. Для устранения этих последствий необходимо осуществить центрирование исходных мультиколлинеарных данных.

Нестационарный процесс – процесс, при котором быстро изменяются условия функционирования экономической системы оказывающие существенное влияние на её характеристики. Подавляющее большинство экономических процессов может быть отнесено к классу нестационарных.

Оптимальность – наилучшее состояние объекта с точки зрения выбранного критерия. Оптимальное состояние в экономике чаще всего достигается для малых систем и объектов. В остальных случаях добиться оптимальных решений трудно, поскольку при осуществлении оптимизации приходится встречаться с ситуациями и факторами, о существовании которых было не известно на момент принятия решения, то есть решение принималось в условиях неопределённости. В подобных ситуациях задача получения оптимального результата заменяется на задачу получения не самого худшего результата.

Ошибки инструментария – виды погрешностей при обработке информации, которые могут быть двух видов. Первый вид ошибок связан с ошибками в применении инструментария, разработанного для изучаемого процесса. Такие ошибки вызваны неграмотностью исследователя и могут быть очень просто удалены. Второй вид ошибок связан с несоответствием инструментария характеру и свойствам исследуемой системы и обнаруживается в процессе верификации модели.

Ошибки наблюдений – тип ошибок при сборе информации. По своим свойствам и характеру влияния на результаты наблюдений ошибки подразделяют на грубые, систематические и случайные. Грубые ошибки вызваны промахами в наблюдениях из-за невнимательности наблюдателя или неисправности измерительного прибора. В процессе передачи, систематизации и обработки наблюдений такие ошибки сразу же выявляются. Систематические ошибки входят в наблюдения в результате действия некоторого постоянно действующего источника ошибок – неисправности инструмента измерения или преднамеренного искажения информации. Появление случайных ошибок вызвано действием множества случайных причин и факторов.

Ошибки экономической информации – преднамеренное или непреднамеренное искажение информации об экономических процессах, связанные с неточностью исходной информации или же вызванные несовершенством и методологической несовместимостью инструментария моделирования с объектом моделирования. Первые называются ошибками наблюдений, вторые - ошибками инструментария.

Случайная величина – величина, принимающая в зависимости от случая те или иные значения с определёнными вероятностями.

Стационарный процесс – процесс, характеристики которого остаются неизменными во времени. Исследование подобных процессов, если информация о них измеряется в метрической шкале, осуществляется с помощью методов теории вероятностей и математической статистики.

Уровень значимости – один из показателей математической статистики, представляющий собой наибольшую допустимую в исследовании вероятность ошибки. Обычно в экономике принимают уровень значимости, равный 0,03 (3%) или 0,05 (5%). С помощью уровня значимости определяют доверительную вероятность.

Теория хаоса – математическая теория, базирующаяся на теории катастроф и теории ошибок, которая утверждает, что явлениям, кажущимся случайными, присущи особые закономерности развития.

Формализация – представление процесса или объекта посредством символов и формул искусственного формализованного языка, нацеленное на уточнение содержания познания. Изучаемым объектам в процессе формализации определённым образом сопоставляются некоторые материальные конструкции, обладающие относительно устойчивым характером и позволяющие в силу этого выявлять и фиксировать существенные и закономерные стороны рассматриваемых объектов.

Число степеней свободы – понятие математической статистки, используемое для вычисления доверительных границ случайной величины. Представляет собой производную величину от численности выборки. Вычисляется как число единиц в выборке минус число связей, которым они подчинены. Так, при пользовании распределением Стьюдента для средних арифметических в малых выборках объёмом в n единиц число степеней свободы будет равно $n-1$.

РАЗВИТИЕ ТЕМЫ: ИНТЕРПОЛЯЦИОННАЯ ФОРМУЛА НЬЮТОНА

Пусть в распоряжении маркетолога есть данные об объёмах потребления товара в зависимости от уровня дохода потребителя:

Таблица П6.1. Исходные данные для задачи интерполирования

Доход потребителя (в минимальных заработных платах)	Объём приобретения товара
1	3
2	7
3	13
4	21
5	31
6	43
7	57

Перед маркетологом стоит задача узнать объём потребления товара при доходе потребителя, равном *3,1* минимальной заработной платы. По сути, перед ним стоит задача интерполирования значения, поскольку указанный доход находится внутри рассматриваемого интервала значений маркетинговой информации.

Воспользуемся для решения поставленной задачи методом Ньютона. Для этого необходимо составить таблицу разностей:

X_t	Y_t	ΔY_t	$\Delta^2 Y_t$	$\Delta^3 Y_t$	$\Delta^4 Y_t$	$\Delta^5 Y_t$	$\Delta^6 Y_t$
1	3	4	2	0	0	0	0
2	7	6	2	0	0	0	
3	13	8	2	0	0		
4	21	10	2	0			
5	31	12	2				
6	43	14					
7	57						

Так как вторые разности постоянны, а третьи разности равны нулю, в качестве интерполяционного многочлена будет использован полином второй степени (его вторая производная, как известно постоянна, а третья производная равна нулю). Эта информация показывает, что из всех слагаемых многочлена Ньютона для целей интерполяции будут использованы только первые три слагаемых. Действительно, третьи разности и разности более высокого порядка равны нулю, а значит, будут равны нулю и соответствующие слагаемые многочлена, которые содержат в себе эти разности в качестве сомножителя.

Для рассматриваемого случая $x_0 = 1$, $x = 3{,}1$, шаг $h = 1$. Тогда $q = (3{,}1 - 1)/1 = 2{,}1$. Интерполяционный многочлен Ньютона будет равен:

$$P_n(x) = y_1 + q\Delta y_1 + \frac{q(q-1)}{2!}\Delta^2 y_1 + \ldots + \frac{q(q-1)\ldots(q-n+1)}{n!}\Delta^n y_1 = 3 + 2{,}1 * 4 + \frac{2{,}1 * 1{,}1}{2} 2 = 13{,}71$$

Таким образом, маркетолог сделает вывод, что при доходе потребителя, равном *3,1* минимальной заработной платы объёмы приобретения товара будут равны 13,71 единиц.

РАЗВИТИЕ ТЕМЫ: ИНТЕРПОЛЯЦИОННАЯ ФОРМУЛА ЛАГРАНЖА

Применим для рассматриваемого условного случая таблицы П6.1 метод вычисления интерполяционного многочлена (полинома) Лагранжа. Он рассчитывается по формуле:

$$L_n(x) = \sum_{t=1}^{n} \frac{(x - x_1)(x - x_2)...(x - x_{t-1})(x - x_{t+1})...(x - x_n)}{(x_t - x_1)(x_t - x_2)...(x_t - x_{t-1})(x_t - x_{t+1})...(x_t - x_n)} y_t \ ,$$

в которой вычисление конечных разностей не требуется. Подставим исходные данные в приведённую формулу. Получим:

$$L_n(x) = 3\frac{(x-2)(x-3)(x-4)(x-5)(x-6)(x-7)}{(1-2)(1-3)(1-4)(1-5)(1-6)(1-7)} + 7\frac{(x-1)(x-3)(x-4)(x-5)(x-6)(x-7)}{(2-1)(2-3)(2-4)(2-5)(2-6)(2-7)} + ...$$

Подставляя в полученный многочлен значение $x = 3,1$, получим искомое интерполяционное значение объёма приобретения товара, равное *13, 71*. Как видно, результаты интерполяции методом Ньютона и методом Лагранжа совпали.

РАЗВИТИЕ ТЕМЫ: ПРОВЕРКА ВЫБОРОЧНОГО ЗНАЧЕНИЯ КОЭФФИЦИЕНТА ПАРНОЙ КОРРЕЛЯЦИИ

В ходе обработки результатов маркетинговых исследований, состоящих из базы данных из 32 пар наблюдений, был сделан вывод о наличии взаимосвязи между факторами и с помощью корреляционного анализа получено значение коэффициента парной корреляции, равное r = 0,6973. Насколько данное значение коэффициента парной корреляции характеризует генеральную совокупность?

Для ответа на этот вопрос используем два подхода.

Первый подход заключается в вычислении величины (6.2.6):

$$t_r = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}},$$

которая имеет t–распределение с n–2 степенями свободы.

Подставим в данную формулу имеющиеся в нашем распоряжении исходные данные, а именно: $r = 0,6973$ и $n = 32$:

$$t_r = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.6973\sqrt{30}}{\sqrt{1-0,4862}} = 5,32$$

Проверка нулевой гипотезы на отсутствие линейной корреляции в генеральной совокупности состоит в сравнении значения *5,32* с табличным значением t при заданной доверительной вероятности α*=0,95* и числе степеней свободы *30*. Вновь воспользуемся таблицей t-статистики Стьюдента (таблица П4.1). Для удобства приведём её ниже:

Число степеней свободы	Значение t	Число степеней свободы	Значение t
15	2,131	16	2,120
17	2,110	18	2,101
19	2,093	20	2,086
21	2,080	22	2,074
23	2,069	24	2,064
25	2,060	26	2,056
27	2,052	28	2,048
29	2,045	30	2,042
40	2,021	60	2,000
120	1,980	Свыше 120	1,960

Расчётное значение t_r*=5,67* оказывается существенно выше табличного значения t*= 2,040* (при числе степеней свободы n*-2= 30*), а, следовательно, можно сделать вывод о том, что в генеральной совокупности существует линейная корреляция, достоверной характеристикой которой выступает выборочное значение коэффициента парной корреляции r*=0,6973*. Само это значение не близко к единице, следовательно, линейная взаимосвязь не очень тесная.

Теперь для этих же данных используем Z-преобразование Фишера (6.2.7), в соответствии с которым получим:

$$Z = \frac{1}{2}\ln\frac{1+r}{1-r} = \frac{1}{2}\ln\frac{1+0.6973}{1-0.6973} = 0.8620$$

Дисперсия (6.2.8) этой величины будет равна: $\sigma_Z{}^2 = \frac{1}{n-3} = \frac{1}{32-3} = 0.0345$, а среднее квадратичное отклонение - σ_z=0,1857.

Истинное значение величины Z (математическое ожидание) при уровне доверительной вероятности α*=0,95* и числе степеней свободы *30* в будет лежать в пределах:

*0,8620-2,040*0,1857<M(Z)< 0,8620+2,040*0,1857*

или

0,4832<M(Z)< 1,2408

Для того обратного преобразовании из Z в коэффициент парной корреляции r, воспользуемся (6.2.10):

$$r = \frac{\ell^{2Z} - 1}{\ell^{2Z} + 1}.$$

В эту формулу вместо Z подставляется сначала значение нижней границы доверительного интервала $Z= 0,4832$ и получается $r= 0,448$. Затем вместо Z подставляется значение верхней границы доверительного интервала $Z= 1,2408$ и получается $r= 0,845$.

В результате чего получим доверительные интервалы для коэффициента парной корреляции:

$$0,448 < r < 0,845.$$

Из анализа этих границ следует, что изучаемая взаимосвязь может носить и нелинейный характер, поскольку нижняя доверительная граница оказывается менее *0,5*.

РАЗВИТИЕ ТЕМЫ: КОЭФФИЦИЕНТ ПАРНОЙ КОРРЕЛЯЦИИ В ПРИМЕНЕНИИ К ФУНКЦИОНАЛЬНОЙ ЗАВИСИМОСТИ

Коэффициент парной корреляции характеризует только степень приближения зависимости (если она есть) к линейной. В том случае, когда изучаемая зависимость носит нелинейный характер, использование этого коэффициента бессмысленно. К сожалению, в маркетинговой учебной литературе ещё встречаются попытки толкования данного коэффициента как некоторого универсального мерила тесноты связи – он, якобы, показывает наличие или отсутствие связи между случайными величинами. Продемонстрируем ещё раз неправомочность такой интерпретации.

Пусть между двумя изучаемыми факторами Y_t и X_t имеется строгая функциональная зависимость:

$$Y_t = X_t^a$$

Если коэффициент парной корреляции r действительно показывает наличие или отсутствие связи между случайными величинами, то в случае функциональной зависимости он должен принимать значения, равные единице.

Для исследования поведения коэффициента парной корреляции на этой функциональной зависимости значение показателя степени a менялось в пределах от –4 до +4. Для каждого из этих значений вычислялись двадцать пар значений Y_t и X_t и подставлялись в формулу для расчёта коэффициента r. Результаты расчётов приведены ниже в таблице[1].

Показатель степени a	Коэффициент r	Показатель степени a	Коэффициент r

[1] Светуньков С.Г. Количественные методы прогнозирования эволюционных составляющих экономической динамики. – Ульяновск: Изд-во УлГУ, 1999. - С.73

-4,0	-0,122896	-3,0	-0,439260
-2,0	-0,522314	-1,0	-0,707623
-0,7	-0,781567	-0,5	-0,830307
-0,3	-0,875419	-0,2	-0,895841
-0,1	-0,914523	0,1	0,932879
0,2	0,958901	0,3	0,969719
0,5	0,985793	0,7	0,995323
1,0	1,00000	2,0	0,971348
3,0	0,922050	4,0	0,873017

Как видно из данных этой таблицы, только в одном случает коэффициент парной корреляции был равен единице, а именно, когда $a = 1$, то есть между факторами существует линейная функциональная зависимость. Во всех остальных случаях он не равен единице, хотя взаимосвязь между факторами самая тесная из возможных, а именно – функциональная.