10. ОБРАБОТКА И АНАЛИЗ РЕЗУЛЬТАТОВ МОДЕЛИРОВАНИЯ

Решения, принимаемые исследователем по результатам имитационного моделировании, могут быть конструктивными только при выполнении двух основных условий:

- полученные результаты обладают требуемой точностью и достоверностью;
- исследователь способен правильно интерпретировать полученные результаты.

Возможность выполнения первого условия закладывается, в основном, еще на этапе разработки модели и частично – на этапе планирования эксперимента. Достоверность результатов моделирования предполагает, что модель, с помощью которой они получены, не только является «правильной», но и отвечает и некоторым дополнительным требованиям, предъявляемым к имитационным моделям. Эти требования и методы оценки соответствия их созданной модели рассматриваются ниже.

Способность исследователя правильно интерпретировать полученные результаты и принимать на их основе важные решения существенно зависит от степени соответствия формы представления результатов целям моделирования.

Если разработчик модели уверен, что полученные результаты будут использоваться в соответствии с одной, четко определенной целью, форма их представления может быть определена заранее. В этом случае преобразование экспериментальных данных к требуемому виду может производиться либо в ходе эксперимента, либо сразу после его завершения.

Если же заранее конкретизировать цель моделирования сложно или целей несколько, данные должны накапливаться в базе данных и затем уже выдаваться в требуемой форме по запросу пользователя. Как правило, по такому принципу строятся системы автоматизации моделирования.

При правильной организации обработки экспериментальных данных могут быть получены дополнительные сведения о моделируемой системе.

10.1 Оценка качества имитационной модели

Оценка качества модели является завершающим этапом ее разработки и преследует две цели:

- проверить соответствие модели ее предназначению (целям исследования);
- оценить достоверность и статистические характеристики результатов, получаемых при проведении модельных экспериментов.

При аналитическом моделировании достоверность результатов определяется двумя основными факторами:

- корректным выбором математического аппарата, используемого для описания исследуемой системы;
- методической ошибкой, присущей данному математическому методу.

При имитационном моделировании на достоверность результатов влияет целый ряд дополнительных факторов, основными из которых являются:

- моделирование случайных факторов, основанное на использовании датчиков СЧ, которые могут вносить «искажения» в поведение модели;
- наличие нестационарного режима работы модели;
- использование нескольких разнотипных математических методов в рамках одной модели;
- зависимость результатов моделирования от плана экспериментов;
- необходимость синхронизации работы отдельных компонентов модели;
- наличие модели рабочей нагрузки, качество которой зависит, в свою очередь, от тех же факторов.

Пригодность имитационной модели для решения задач исследования характеризуется тем, в какой степени она обладает так называемыми целевыми свойствами. Основными из них являются:

- адекватность;
- устойчивость;
- чувствительность.

Ниже рассмотрены некоторые способы проведения оценки модели по каждому из них.

10.1.1. Оценка адекватности модели.

В обще случае под адекватностью понимают степень соответствия модели тому реальному явлению или объекту, для описания которого она строится.

Вместе с тем, создаваемая модель ориентирована, как правило, на исследование определенного подмножества свойств этого объекта. Поэтому можно считать, что адекватность модели определяется степенью ее соответствия не столько реальному объекту, сколько целям исследования. В наибольшей степени это утверждение справедливо относительно моделей проектируемых систем (т.е. в ситуациях, когда реальная система вообще не существует).

Тем не менее, во многих случаях полезно иметь формальное подтверждение (или обоснование) адекватности разработанной модели. Один из наиболее распространенных способов такого обоснования - использование методов математической статистики. Суть этих методов заключается в проверке выдвинутой гипотезы (в данном случае - об адекватности модели) на основе некоторых статистических критериев.

Замечание: при проверке гипотез методами математической статистики необходимо иметь в виду, что статистические критерии не могут доказать ни одной гипотезы: они могут лишь указать на отсутствие опровержения.

Процедура оценки основана на сравнении измерений на реальной системе и результатов экспериментов на модели и может проводиться различными способами. Наиболее распространенные из них:

- по средним значениям откликов модели и системы;
- по дисперсиям отклонений откликов модели от среднего значения откликов системы;

- по максимальному значению относительных отклонений откликов модели от откликов системы.

Названные способы оценки достаточно близки по сути, поэтому ограничимся рассмотрением первого из них.

При этом способе проверяется гипотеза о близости среднего значения наблюдаемой переменной У среднему значению отклика реальной системы Y*.

В результате No опытов на реальной системе получают выборку Y*. Выполнив N_M экспериментов на модели, также получают множество значений наблюдаемой переменной Y.

Затем вычисляются оценки математического ожидания и дисперсии откликов модели и системы, после чего выдвигается гипотеза о близости средних значений величин У* и У (в статистическом смысле). Основой для проверки гипотезы является t-статистика (распределение Стьюдента). Ее значение, вычисленное по результатам испытаний, сравнивается с критическим значением $£_{кр}$, взятым из справочной таблицы. Если выполняется неравенство $t < £_{кр}$, то гипотеза принимается.

Необходимо еще раз подчеркнуть, что статистические методы применимы только в том случае, если оценивается адекватность модели существующей системе. На проектируемой системе провести измерения, естественно, не представляется возможным. Единственный способ преодолеть это препятствие заключается в том, чтобы принять в качестве эталонного объекта концептуальную модель проектируемой системы. Тогда оценка адекватности программно реализованной модели заключается в проверке того, насколько корректно она отражает концептуальную модель. Данная проблема сходна с проверкой корректности любой компьютерной программы, и ее можно решать соответствующими методами, например с помощью тестирования.

10.1.2. Оценка устойчивости модели.

При оценке адекватности модели как существующей, так и проектируемой системе реально может быть использовано лишь ограниченное подмножество всех возможных значений входных параметров (рабочей нагрузки и внешней среды). В связи с этим для обоснования достоверности получаемых результатов моделирования большое значение имеет проверка устойчивости модели. В теории моделирования это понятие трактуется следующим образом.

Устойчивость модели — это ее способность сохранять адекватность при исследовании эффективности системы на всем возможном диапазоне рабочей нагрузки, а также при внесении изменений в конфигурацию системы.

Универсальной процедуры проверки устойчивости модели не существует. Разработчик вынужден прибегать к частичным тестам и здравому смыслу. Часто бывает полезна апостериорная проверка. Она состоит в сравнении результатов моделирования и результатов измерений на системе после внесения в нее изменений. Если результаты моделирования приемлемы, уверенность в устойчивости модели возрастает.

В общем случае можно утверждать, что чем ближе структура модели структуре системы и чем выше степень детализации, тем устойчивее модель.

Устойчивость результатов моделирования может быть также оценена методами математической статистики. Здесь уместно вспомнить основную задачу математической статистики. В данном случае устойчивость результатов моделирования можно рассматривать как признак, подлежащий оценке. Для проверки гипотезы об устойчивости результатов может быть использован критерий Уилкоксона.

Критерий Уилкоксона служит для проверки того, относятся ли две выборки к одной и той же генеральной совокупности (т. е. обладают ли они одним и тем же статистическим признаком). Например, в двух партиях некоторой продукции измеряется определенный признак, и требуется проверить гипотезу о том, что этот признак имеет в обеих партиях одинаковое распределение; другими словами, необходимо убедиться, что технологический процесс от партии к партии изменяется несущественно.

При статистической оценке устойчивости модели соответствующая гипотеза может быть сформулирована следующим образом: при изменении входной (рабочей) нагрузки или структуры ИМ закон распределения результатов моделирования остается неизменным.

Проверку указанной гипотезы Н проводят при следующих исходных данных:

· есть две выборки $X = (x_1 \ldots, x_n)$ и $Y = (y_1 \ldots, y_m)$, полученные для различных значений рабочей нагрузки; относительно законов распределения X и Y никаких предположений не делается.

· значения обеих выборок упорядочиваются вместе по возрастанию. Затем анализируется взаимное расположение x_i и y_j В случае $y_j < x_i$ говорят, что пара значений $(x_i\ y_j)$ образует инверсию.

(*Например, пусть для* $n = m = 3$. *После упорядочивания получилась такая последовательность значений:* $y_1\ x_1\ y_3,\ x_2,\ y_2,\ x_3$. *Тогда имеем инверсии:* $(x_1, y_1), (x_2, y_1), (x_2, y_3), (x_3, y_1), (x_3, y_2), (x_3, y_3)$).

· подсчитывают полное число инверсий U. Если гипотеза верна, то Uдолжно сильно отклоняться от своего математического ожидания M:

$$M=(п*т)/2$$

От гипотезы отказываются, если $| U - M | > U_{КР}$ ($U_{КР}$ определяют по таблице для заданного уровня значимости).

10.1.3. Оценка чувствительности ИМ.

Очевидно, что устойчивость является положительным свойством модели. Однако если изменение входных воздействий или параметров модели (в некотором заданном диапазоне) не отражается на значениях выходных параметров, то польза от такой модели невелика (ее можно назвать «бесчувственной»). В связи с этим возникает задача оценивания чувствительности модели к изменению параметров рабочей нагрузки и внутренних параметров самой системы.

Такую оценку проводят по каждому параметру X_k в отдельности. Основана она на том, что обычно диапазон возможных изменений параметра известен. Одна из наиболее простых и распространенных процедур оценивания состоит в следующем.

1) вычисляется величина относительного среднего приращения параметра X_k:

$$\Delta X_k = \frac{(X_{k\max} - X_{k\min})}{(X_{k\max} + X_{k\min})} \cdot 100\%$$

2) проводится пара модельных экспериментов при значениях $X_k = X_{kmax}$ и $X_k = X_{kmin}$ и средних фиксированных значениях остальных параметров. Определяются значения отклика модели $Y_1=f(X_{kmax})$ и $Y_2=(X_{kmin})$;

3)вычисляется ее относительное приращение наблюдаемой переменной У:

$$\Delta Y = \frac{|Y_1 - Y_2|}{(Y_1 + Y_2)} \cdot 100\%$$

В результате для k-го параметра модели имеют пару значений $(\Delta X_к, \Delta У_к)$, характеризующую чувствительность модели по этому параметру.

Аналогично формируются пары для остальных параметров модели, которые образуют множество $\{\Delta X_к, \Delta У_к\}$.

Данные, полученные при оценке чувствительности модели, могут быть использованы, в частности, при планировании экспериментов: большее внимание должно уделяться тем параметрам, по которым модель является более чувствительной.

10.1.4. Калибровка модели.

Если в результате проведенной оценки качества модели оказалось, что ее целевые свойства не удовлетворяют разработчика, необходимо выполнить ее калибровку, т. е. коррекцию с целью приведения в соответствие предъявляемым требованиям.

Как правило, процесс калибровки носит итеративный характер и состоит из трех основных этапов:

- глобальные изменения модели (например, введение новых процессов, изменение типов событий и т. д.);
- локальные изменения (в частности, изменение некоторых законов распределения моделируемых случайных величин);
- изменение специальных параметров, называемых калибровочными.

Целесообразно объединить оценку целевых свойств ИМ и ее калибровку в единый процесс. Именно такая стратегия принята в статистическом методе калибровки, описанном ниже.

Процедура калибровки состоит из трех шагов, каждый из которых является итеративным (рис. 2.13).

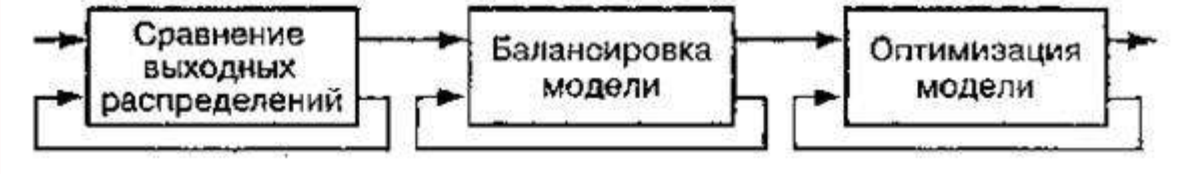

Рис 10.1. Схема процесса калибровки ИМ

Шаг 1. Сравнение выходных распределений.

Цель - оценка адекватности ИМ. Критерии сравнения могут быть различны. В частности, может использоваться величина разности между средними значениями откликов модели и системы.

Устранение различий на этом шаге основано на внесении глобальных изменений.

Шаг 2. Балансировка модели.

Основная задача - оценка устойчивости и чувствительности модели. По его результатам, как правило, производятся локальные изменения (но возможны и глобальные).

Шаг 3. **Оптимизация модели.**

Цель этого этапа — обеспечение требуемой точности результатов. Здесь возможны три основных направления работ:

- дополнительная проверка качества датчиков СЧ;
- снижение влияния переходного режима;
- применение специальных методов понижения дисперсии.

10.2. Подбор параметров распределений

В некоторых случаях имитационная модель сложной системы может быть реализована в виде набора отдельных моделей ее подсистем. При проведении экспериментов с такой моделью в целях сокращения затрат времени бывает необходимо заменять моделирование работы одной из подсистем некоторым числовым параметром (вспомните принцип параметризации), либо случайной величиной, распределенной по заданному закону. Чтобы такая замена была выполнена корректно, исследователь должен располагать описанием зависимости данного числового параметра от времени и других факторов, фигурирующих в модели.

При имитационном моделировании подбор законов распределений выполняется на основе статистических данных, полученных в ходе эксперимента.

В основе процедуры отыскания закона распределения некоторой величины по экспериментальным данным лежит проверка статистических гипотез.

Статистическая гипотеза — это утверждение относительно значений одного или боле параметров распределения некоторой величины или о самой форме распределения.

Обычно выбирают две исходные гипотезы: основную—H_0 и альтернативную ей – H_1

Статистическая проверка гипотезы - это процедура выяснения, следует ли принял основную гипотезу H_0 или отвергнуть ее.

Если в результате проверки гипотеза H_0 ошибочно отвергается, то имеет место ошибка 1-го рода (характеризующаяся более тяжелыми последствиями); если гипотеза H_0 принимается при истинности H_1 — это ошибка второго рода.

Вероятности ошибок I и II рода (α и β) зависят от критерия, на основание которого будет выбираться одна из гипотез. Очевидно, что вероятности этих двух ошибок взаимосвязаны, то есть чем больше значение α, тем меньше β и наоборот.

Обычное решение этой дилеммы состоит в том, что выбирают некоторое фиксированное значение α (как правило 0.05, 0.01, 0.001) и надеются, что β будет также мало. Фиксированное значение а называется **уровнем значимости.**

Для выбранного значения α определяется так называемая критическая область B, удовлетворяющая условию

$$P(Z \in B \mid H_0 верна) \le \alpha$$

где Z— контрольная величина (критерий), представляющая собой некоторую функцию от выборки (результатов эксперимента).

Проверка гипотезы состоит в следующем. Производится выборка (проводится эксперимент), на основании чего вычисляется z — частное значение критерия Z. Если $z \in B$, то от гипотезы H_0 отказываются. Если z не лежит в B, то говорят, что полученные наблюдения **не противоречат принятой гипотезе.**

Для наиболее часто используемых статистических гипотез разработаны критерии, позволяющие проводить их проверку с наибольшей достоверностью. Рассмотрим основные из них.

t-критерий служит для проверки гипотезы о равенстве средних значений двух нормально распределенных СВ (X и Y) в предположении, что дисперсии их равны (хотя и неизвестны). Сравниваемые выборки могут иметь разный объем.

В качестве критерия используют величину

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{(n_1 - 1)D_x + (n_2 - 1)D_y}} \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 - 2)}{n_1 + n_2}}.$$

Величина T подчиняется t-распределению Стьюдента.

Критическое значение для t-критерия определяется по таблице для выбранного значения α и числа степеней свободы $k = n_1+n_2-2$.

Если вычисленное значение T превышает $t_{кр}$, найденное по таблице, то гипотезу H_0 отвергают.

По отношению к предположению о «нормальной распределенности» величин x и y t-критерий не очень чувствителен. Его можно применять, если распределения СВ не имеют нескольких вершин и не слишком асимметричны.

F-критерий служит для проверки гипотезы о равенстве дисперсий D_x и D_y при условии, что x и y распределены нормально.

Гипотезы такого рода имеют большое значение в технике, так как дисперсия есть мера таких характеристик, как погрешности измерительных приборов, точность технологических процессов, точность наведения при стрельбе и так далее.

В качестве контрольной величины используется отношение дисперсий $F = D_x / D_y$ (или D_y / D_x - большая дисперсия должна быть в числителе).

Величина F подчиняется F-распределению (Фишера) с (m_1, m_2) степенями свободы ($m_1=n_1-1$, $m_2=n_2$ - l). Проверка гипотезы состоит в следующем.

Для величины $\gamma = \alpha/2$ и величин m_1 m_2 по таблице F-распределения выбирают значения $F_{\gamma,m1,m2}$. Если статистика F*, вычисленная по выборке, больше этого критического значения, гипотеза должна быть отклонена с вероятностью ошибки α.

Критерии согласия — это критерии, с помощью которых проверяют, удовлетворяет ли рассматриваемая СВ данному закону распределения.

Критерий согласия Пирсона **(формула)** и *Критерий Колмогорова — Смирнова* были подробно рассмотрены в курсе ТВиМС.

10.3. Оценка влияния и взаимосвязи факторов

Как правило, количественная оценка степени влияния того или иного фактора на значения наблюдаемой переменной (показателя эффективности) вызывает значительную сложность, особенно при наличии взаимного влияния факторов. Наиболее простой и доступный способ решения этой проблемы состоит в использовании результатов оценки чувствительности модели.

Однако эти результаты сложно представить в форме аналитической зависимости. Такое представление может оказаться весьма полезным для многих практических задач, связанных как с разработкой моделей (речь опять-таки идет о принципе параметризации), так и непосредственно с принятием решений по экспериментальным данным.

Отыскание аналитических зависимостей, связывающих между собой различные параметры, фигурирующие в модели, может быть основано на совместном использовании группы методов математической статистики: дисперсионного, корреляционного и регрессионного анализа. Подробному и строгому описанию соответствующих процедур посвящено огромное количество книг учебного, научного и справочного характера. Поэтому основная цель изложения последующего материала сводится к тому, чтобы показать роль и место указанных методов при проведении анализа данных, полученных в ходе имитационного эксперимента.

10.3.1. Однофакторный дисперсионный анализ

Его суть сводится к определению влияния на результат моделирования одного выбранного фактора.

Пусть, например, исследователя интересует средняя интенсивность отказов компьютера, и в созданной им модели учтены следующие факторы: интенсивность поступления заданий пользователей, интенсивность обращений в оперативную память, временные характеристики решаемых задач и интенсивность обращений к жесткому диску. Если предварительные данные говорят о том, что основной причиной отказов является ненадежная работа жесткого диска, то в качестве анализируемого фактора целесообразно выбрать интенсивность обращений к нему. Задача факторного анализа в данном случае состоит в том, чтобы оценить влияние указанного фактора на среднее число отказов.

Формально постановка задачи однофакторного дисперсионного анализа состоит в следующем. Пусть интересующий нас фактор x имеет 1 уровней. Для каждого из них получена выборка значений наблюдаемой переменной y: $y_j(1)$, $y_j(2)$, ... $y_j(l)$, j=l,...,n, n — объем выборки (число наблюдений).

Необходимо проверить гипотезу H_0 о равенстве средних значений выборок (т.е. о независимости значений y от значений исследуемого фактора x). Уравнение однофакторного дисперсионного анализа имеет вид:

$$y_{ij} = m+a_i+e_{ij},$$

где y_{ij} - j-е значение y в i-й серии опытов,

m - генеральное среднее случайной величины y,

a_i - неизвестный параметр, отражающий влияние фактора x («эффект» i-го значения фактора x),

e_{ij} - ошибка измерения y.

Для проверки гипотезы H_0 используют F-критерий и переходят от проверки значимости различий средних к проверке значимости различий двух дисперсий:

- генеральной (обусловленной погрешностями измерений) - D_0;
- факторной (обусловленной изменением фактора x) - D_x.

Значение F-критерия вычисляется как отношение Dx / Do или Do / Dx (в числителе должна стоять большая из дисперсий); затем но по таблице F-распределений находят его критическое значение $f_{кр}$ для заданного уровня значимости и числа степеней свободы.

Если $F > P_{кр}$, то гипотезу H_0 отвергают, т. е. различия являются значимыми (фактор x влияет на значения y).

10.3.2. Многофакторный дисперсионный анализ

Многофакторный дисперсионный анализ (МДА) позволяет оценивать влияние на наблюдаемую переменную уже не одного, а произвольного числа факторов. Точнее, МДА позволяет выбрать ив группы факторов, участвующих в эксперименте, те, которые действительно влияют на его результат.

Методику проведения многофакторного дисперсионного анализа рассмотрим применительно к частичному факторному эксперименту, проводимому в соответствии с латинским планом.

Пусть в эксперименте рассматриваются один первичный фактор и два вторичных, каждый из которых имеет n уровней (т. е. объем испытаний $N = n^2$).

Обозначим через y_{ijk} результат эксперимента при условии, что фактор ***a*** находился на уровне i, фактор ***b*** - на уровне j, фактор ***c*** - на уровне к. Множество значений, которые может принимать упорядоченная тройка (i, j, к), обозначим через L.

В этом случае уравнение дисперсионного анализа выглядит следующим образом:

$$y_{ijk} = m + a_i + b_j + g_k + e_{ijk},$$

где m - генеральное среднее случайной величины y,

a_i, b_j, g_k - неизвестные параметры («эффекты» соответствующих факторов).

Решение задачи дисперсионного анализа заключается в проверке гипотез о независимости результатов измерений от факторов ***a, b, c***:

$$Ha:\ ai=0,\ i=1,\ \dots,\ n;$$

$$Hb:\ bj=0,\ j=1,\ \dots,\ n;$$

$$Hc:\ ck=0,\ k=1,\dots,\ n;$$

Для этого по методу наименьших квадратов (МНК) находят оценки параметров m, a_i, b_j, g_k, минимизируя по указанным переменным (поочередно) функцию

$$SS = \sum_{L} (y_{ijk} - m - a_i - b_j - g_k)^2.$$

Затем по каждому фактору вычисляется f-статистика. Величина F есть мера потерь при принятии гипотезы H_0. Чем больше F, тем хуже модель, отвергающая влияние соответствующего фактора. Таким образом, если вычисленное значение F больше $F_{кр}$, найденного по таблице для некоторого уровня значимости, то гипотеза отвергается.

Необходимо отметить, что дисперсионный анализ может использоваться для оценки влияния факторов, имеющих как количественный характер, так и качественный, поскольку в уравнении дисперсионного анализа фигурируют не сами факторы, а только их «эффекты».

В том случае, если все факторы носят количественный характер, взаимосвязь между ними и наблюдаемой переменной может быть описана с помощью уравнения регрессии.

10.3.3. Корреляционный и регрессионный анализ.

Это два близких метода, которые обычно используются совместно для исследования взаимосвязи между двумя или более непрерывными переменными.

Методы корреляционного анализа позволяют делать статистические выводы о степени зависимости между переменными.

Величина линейной зависимости между двумя переменными измеряется посредством простого коэффициента корреляции, величина зависимости от нескольких - посредством множественного коэффициента корреляции.

В корреляционном анализе используется также понятие частного коэффициента корреляции, который измеряет линейную взаимосвязь между двумя переменными без учета влияния других переменных.

Если корреляционный анализ позволил установить наличие линейной зависимости наблюдаемой переменной от одной или более независимых, то форма зависимости может быть уточнена методами регрессионного анализа.

Для этого строится так называемое уравнение регрессии, которое связывает зависимую переменную с независимыми и содержит неизвестные параметры. Если уравнение линейно относительно параметров (но необязательно линейно относительно независимых переменных), то говорят о линейной регрессии, в противном случае регрессия нелинейна.

Рассмотрим простой корреляционный анализ, т. е. метод определения взаимосвязи между двумя переменными.

Обозначим их x и y. Независимо от способа получения выборки имеются два предварительных шага для определения существования и степени линейной зависимости между x и y. Первый шаг заключается в графическом отображении точек (x_i y_i) на плоскости (x, y) — т. е. в построении диаграммы рассеяния. Анализируя диаграмму рассеяния, можно решить, допустимо ли предположение о линейной зависимости между x и y (см.рис.10.2).

Если r_{xy} не равен нулю, то на втором шаге вычисляется его точное значение.

Чем больше по абсолютному значению r_{xy}, тем сильнее линейная зависимость между переменными. При r =1 имеет место функциональная линейная зависимость между x и y вида y=ax+b, причем если r положительно, то говорят о положительной корреляции, т.е. большие значения одной величины соответствуют большим значениям другой; при r = -1 имеет место отрицательная корреляция; при $0 < r < 1$ вероятна либо линейная корреляция с рассеянием (рис.10.2, в), либо нелинейная корреляция (рис.10.2, г).

При анализе результатов ИМ необходимо иметь в виду, что если даже удалось установить тесную зависимость между двумя переменными, это еще не является прямым доказательством их причинно-следственной связи. Возможно, имеет место стохастическая зависимость, обусловленная, например, коррелированностью последовательностей псевдослучайных чисел, используемых в имитационной модели.

Поэтому результаты корреляционного анализа целесообразно уточнить, проведя регрессионный анализ.

Регрессионный анализ позволяет решать две задачи:

1) устанавливать наличие возможной причинной связи между переменными;

2) предсказывать значения зависимой переменной по значениям независимых переменных. Эта возможность особенно важна в тех случаях, когда прямые измерения зависимой переменной затруднены.

Если предполагается линейная зависимость между x и y, то она может быть описана уравнением вида $y_i=b_0+b_1x+e_i$, которое называется простой линейной регрессией y по оси x.

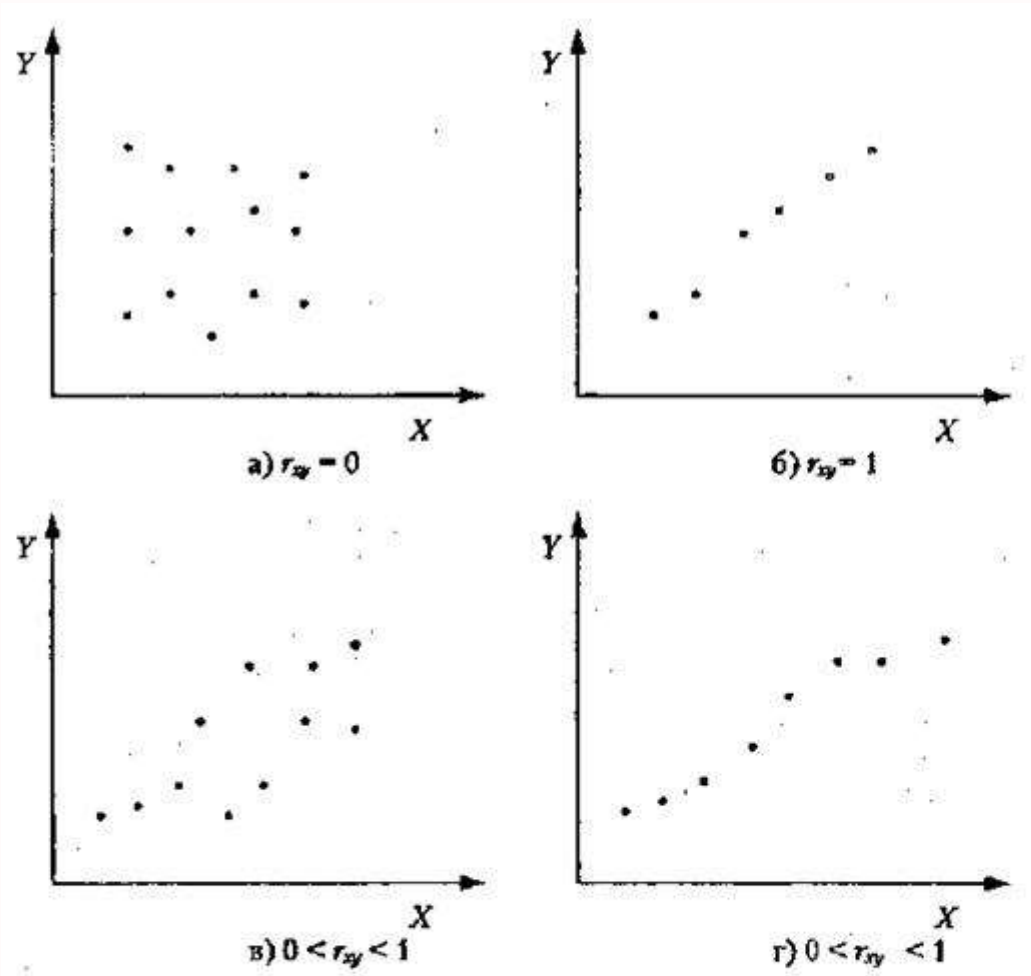

Рис. 10.2. Графическое отображении корреляции переменных

Величины b_0 и b_1 являются неизвестными параметрами, а e_i — случайные ошибки испытаний.

Цель регрессионного анализа — найти наилучшие в статистическом смысле оценки параметров b_0 и b_i (величину b_i обычно называют коэффициентом регрессии).

Разница между наблюдаемым и оцененным значением y при $x = x_i$ называется отклонением (или остатком). Величины отклонений могут быть использованы для проверки адекватности полученной модели. Для этого строится график отклонений d = f(y) или d=f(x) (см. рис.10.2), и по его виду делается предварительное заключение о степени адекватности регрессионной модели.

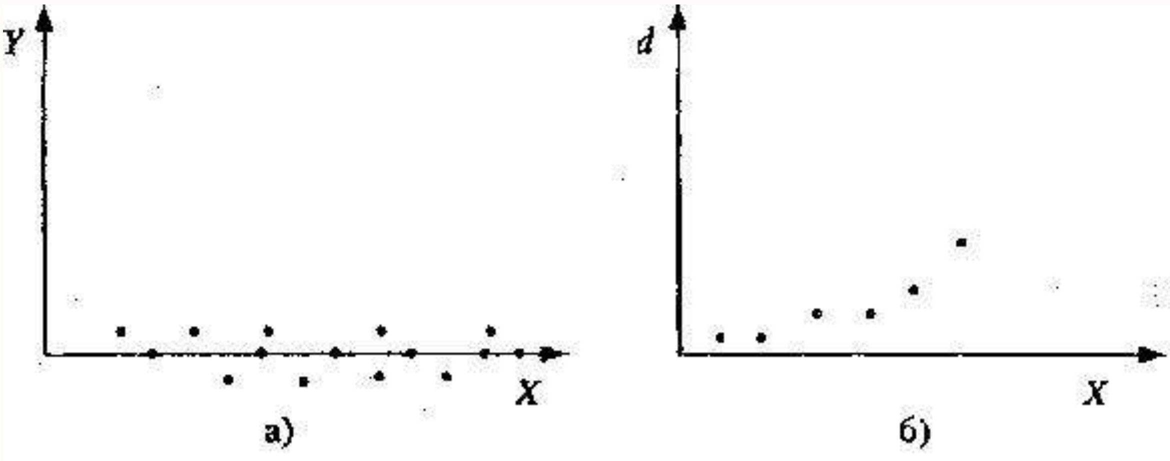

Рис. 10.3 Графическое представление функций отклонений: а – модель адекватна; б – необходимо введение дополнительной независимой переменной.

В случае нескольких независимых переменных имеет место ***множественная линейная регрессия***:

$$y=b_0+b_1x_1+b_2x_2+ \ldots+b_kx_k+e.$$

В этом случае для отыскания оценок b_i также используется метод наименьших квадратов.

В случае нелинейной регрессии основой для построения регрессионной модели опять-таки является МНК. Однако в этом случае для отыскания оценок b строится система нелинейных уравнений (относительно b_i), а для ее решения используются различные итерационные методы.