

УДК 004.67

ДИСПЕРСИОННЫЙ АНАЛИЗ ФАКТОРОВ ПРИ ИМИТАЦИОННОМ МОДЕЛИРОВАНИИ СЛОЖНЫХ СИСТЕМ

Орлова А.А., студент

*Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Программное обеспечение ЭВМ и информационные технологии»*

Научный руководитель: Рудаков И.В., к.т.н, доцент

Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана

irudakov@bmstu.ru

Моделирование сложных систем

Сложная система — система, состоящая из множества взаимодействующих составляющих (подсистем), вследствие чего сложная система приобретает новые свойства, которые отсутствуют на подсистемном уровне и не могут быть сведены к свойствам подсистемного уровня [6].

Моделирование сложных систем на сегодняшний день является актуальной темой. Оно находит применение во многих сферах как научной, так и производственной деятельности.

Моделирование сложных систем в идеальных условиях, т.е. отсутствия внешних факторов, влияющих на процесс работы системы, не представляет особой практической ценности. В связи с чем, при моделировании стараются учесть случайные внешние воздействия, смоделировать их.

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}; \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_p \end{pmatrix}; e = \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix},$$

Где y – результаты наблюдений, β – различные факторы, e – ошибка

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{p1} & x_{p2} & \dots & x_{pn} \end{pmatrix},$$

X – матрица коэффициентов, тогда:

$$y = X'\beta + e. \quad (1)$$

При оценке результатов моделирования встает разумный вопрос, как в таком случае оценить влияние внешних факторов на систему. Для его разрешения используется процедура дисперсионного анализа.

Назначение дисперсионного анализа.

Дисперсионный анализ – анализ изменчивости признака под влиянием каких-либо контролируемых переменных факторов. (В зарубежной литературе именуется ANOVA – «Analysis of Variance»)

Основной целью дисперсионного анализа является исследование значимости различия между средними. Может показаться странным, что процедура сравнения средних называется дисперсионным анализом. В действительности, это связано с тем, что при исследовании статистической значимости различия между средними двух (или нескольких) групп, мы на самом деле сравниваем (т.е. анализируем) выборочные дисперсии.

Обобщенно задача дисперсионного анализа состоит в том, чтобы из общей вариативности признака выделить три частные вариативности:

- вариативность, обусловленную действием каждой из исследуемых независимых переменных;
- вариативность, обусловленную взаимодействием исследуемых независимых переменных;
- вариативность случайную, обусловленную всеми неучтенными обстоятельствами.

Виды дисперсионного анализа.

Дисперсионный анализ схематически можно подразделить на несколько категорий. Это деление осуществляется, в зависимости от того, сколько факторов принимает участие в рассмотрении, сколько переменных подвержены действию факторов и как соотносятся друг с другом выборки значений.

При наличии одного фактора, влияние которого исследуется, дисперсионный анализ называется однофакторным, и распадается на две разновидности:

- Анализ несвязанных (то есть – различных) выборок. Например, одна группа респондентов решает задачу в условиях тишины, вторая – в шумной комнате. (В этом случае, к слову, нулевая гипотеза звучала бы так: «среднее время решения задач такого-то типа будет одинаково в тишине и в шумном помещении», то есть не зависит от фактора шума.)

▪ Анализ связанных выборок. То есть: двух замеров, проведенных на одной и той же группе респондентов в разных условиях. Тот же пример: в первый раз задача решалась в тишине, второй – сходная задача – в условиях шумовых помех. (На практике к подобным опытам следует подходить с осторожностью, поскольку в действие может вступить неучтенный фактор «обучаемость», влияние которого исследователь рискует приписать изменению условий, а именно, - шуму.)

В случае, если исследуется одновременное воздействие двух или более факторов, мы имеем дело с многофакторным дисперсионным анализом, который также можно подразделить по типу выборки.

Многофакторный дисперсионный анализ отличается от однофакторного проблемой взаимодействия факторов, решение которой принципиально не зависит от числа факторов. Анализ с двумя и более факторами требует внимательного отношения и определенных навыков в интерпретации результатов.

Например, интерпретация результатов трехфакторного дисперсионного анализа, требует опыта и умения свободно обращаться со статистическими величинами, четырехфакторного анализа – сложна даже для специалиста высокой квалификации.

Если же воздействию факторов подвержено несколько переменных, - речь идет о многомерном анализе.

Инструменты для моделирования

Существует набор различных сред моделирования, с помощью которых можно проводить исследования сложных систем, в которых используется дисперсионный анализ.

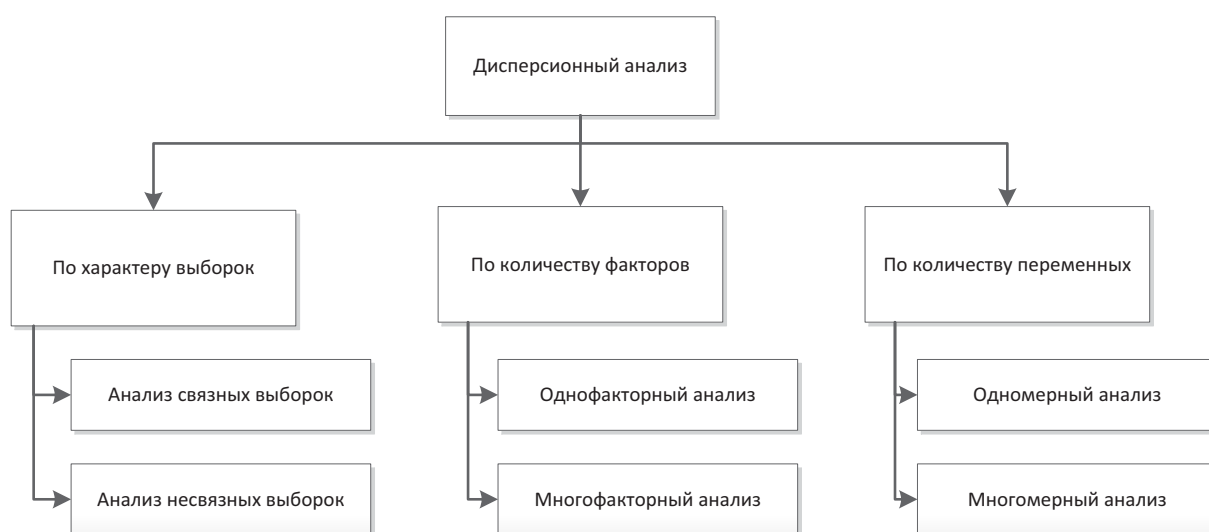


Рис. 1. Классификация дисперсионного анализа

В зависимости от способа использования дисперсионного анализа среды можно разделить на:

- Среда, позволяющие проводить моделирование систем и оценку результатов с помощью процедуры дисперсионного анализа (например, GPSS World);
- Среда, позволяющие проводить моделирование систем и оценку результатов, неявно используя процедуру дисперсионного анализа (например, AnyLogic).

Кроме того, существует обширный ряд программ, выполняющих статистическую обработку информации с использованием дисперсионного анализа. Условно к классификации можно так же добавить:

- Среда статистической обработки информации с явным использованием дисперсионного анализа (SPSS Statistic)

Дисперсионный анализ в GPSS.

GPSS¹ — язык моделирования, используемый для имитационного моделирования различных систем, в основном систем массового обслуживания (существуют бесплатные среды моделирования).

В GPSS существует встроенная процедура – ANOVA, которая выполняет дисперсионный анализ, определяет значимость факторов по критерию Фишера (F-критерию) и выполняет расчет доверительных интервалов исследуемых величин. Если значение F-критерия не превышает критическое значение, то наблюдаемый фактор незначимый, выборочные последовательности принадлежат одной генеральной совокупности.

При выполнении дисперсионного анализа для каждого уровня обработки исследуемого фактора необходимо выполнить несколько прогонов модели. Количество прогонов устанавливается в зависимости от задаваемой точности. Для уменьшения корреляции результатов прогонов рекомендуется, чтобы каждый из прогонов использовал свою последовательность псевдослучайных чисел.

Главная цель процедуры ANOVA состоит в создании стандартной таблицы ANOVA, где указывается окончательная F-статистика и ее критическое значение.

Неявно в ANOVA используется **аддитивная математическая модель**, которая определяет компоненты изменения в наблюдениях. Ее называют статистической моделью. Самая простая статистическая модель:

$$y_{ig} = \mu + e_{ig} \quad (2)$$

¹ англ. *General Purpose Simulation System* — система моделирования общего назначения

т.е. каждое i -е наблюдение представляет собой общее среднее по всем опытам μ , и случайную ошибку e_{ig} . В этой модели общее среднее не изменяется от опыта к опыту, в отличие от ошибки.

Статистическая модель для анализа данных экспериментов с одним фактором **A** имеет следующий вид:

$$y_{ig} = \mu + \alpha_i^A + e_{ig}, \quad (3)$$

где α_i^A – главный эффект фактора **A** на уровне i . Все наблюдения на данном уровне обработки анализируются, используя то же самое значение для α^A . Так как в этом эксперименте имеется только один фактор, число комбинаций обработки определяется числом уровней I этого фактора.

Для двух факторов общая модель факторного плана такова:

$$y_{ig} = \mu + \alpha_i^A + \alpha_j^B + \alpha_{ij}^{AB} + e_{ig}, \quad (4)$$

где α_j^B – главный эффект фактора **B** на уровне j ; α_{ij}^{AB} – взаимодействие фактора **A** на уровне i и фактора **B** на уровне j . Сумма эффектов двух факторов не равна сумме их отдельных эффектов из-за взаимодействия между ними. Главный эффект фактора определяет долю участия фактора в значении функции отклика во время перехода его с нижнего уровня к верхнему.

Дисперсионный анализ, основанный на статистической модели, заканчивается построением таблицы ANOVA, в которой анализируется влияние факторов **A**, **B**, взаимодействие между факторами **AB** и случайные помехи наблюдения.

На рис. 1 представлена таблица ANOVA, полученная в GPSS World.

В реализации GPSS ANOVA имеет ограничение на количество исследуемых факторов – максимум 6. Данный недостаток может быть критичным в случае необходимости исследования сложной модели с большим количеством внешних воздействий.

Схожая аддитивная математическая модель так же используется в средах AnyLogic и SPSS Statistics, которые будут рассмотрены далее.

ANOVA					
Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F	Critical Value of F (p=.05)
A	28.000	2	14.000	5.600	3.89
B	21.000	3	7.000	2.800	3.49
AB	11.000	6	1.833	0.733	3.00
Error	30.000	12	2.500		
Total	90.000	23			
Treatment Level A B	Count	Mean	Minimum	Maximum	95% C.I. (SE)
1 1	2	3.000	2.000	4.000	{ 0.564, 5.436 }
1 2	2	1.000	1.000	1.000	{ -1.436, 3.436 }

Рис. 2. Результат выполнения процедуры ANOVA в GPSS World.

Дисперсионный анализ в AnyLogic

AnyLogic — программное обеспечение для имитационного моделирования бизнес-процессов (проприетарная лицензия).

Дисперсионный анализ в среде AnyLogic представлен неявно в виде процедуры оптимизации стохастических процессов.

Цель процедуры – нахождение оптимального (минимального или максимального) значения при заданных факторах и разбросу их значений.

Пример задания свойств оптимизационного эксперимента в AnyLogic приведен на рис. 3.

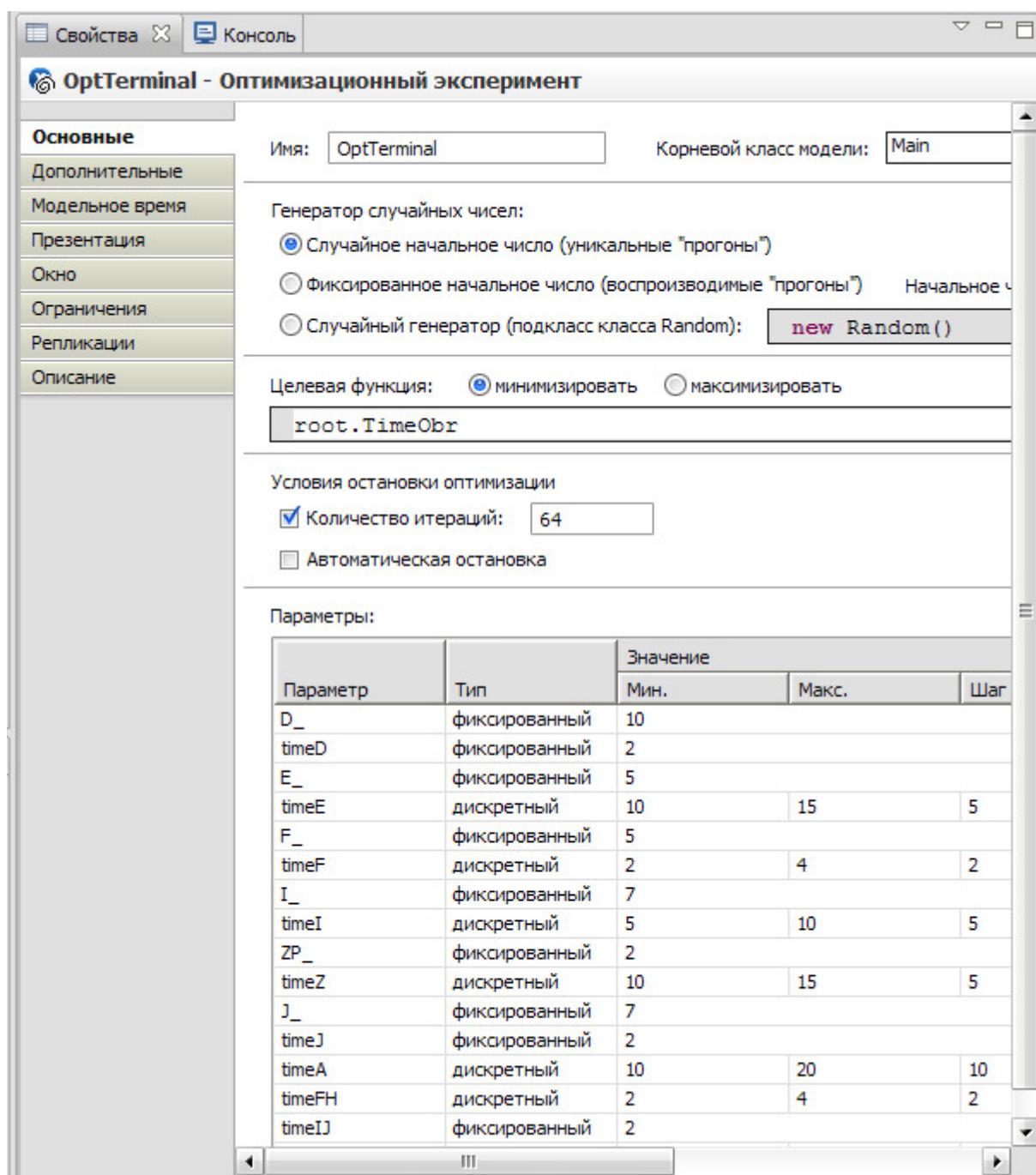


Рис. 3. Пример задания свойств оптимизационного эксперимента в AnyLogic.

Дисперсионный анализ в SPSS

IBM SPSS² Statistics — компьютерная программа для статистической обработки данных, один из лидеров рынка в области коммерческих статистических продуктов, предназначенных для проведения прикладных исследований в социальных науках (проприетарная лицензия).

² англ. «Statistical Package for the Social Sciences» — «статистический пакет для социальных наук»

Основное применение данное ПО нашло в различных маркетинговых и социологических исследованиях, однако SPSS обладает мощнейшими возможностями для статистической обработки информации. Стоит так же отметить, что SPSS не является средством для моделирования в отличие от GPSS и AnyLogic.

Статистический пакет SPSS имеет реализации как для однофакторного дисперсионного анализа (One-Way Anova), так и для многофакторного (Analyze - General Linear Model - Multivariate).

Недостатки дисперсионного анализа

Результаты выполнения дисперсионного анализа в различных средах дают примерно одинаковые результаты, однако все реализации метода обладают схожими недостатками (если не учитывать ограничение GPSS World на количество факторов). На процедуру дисперсионного анализа накладывается несколько ограничений:

1. Размер выборок, используемых в экспериментах, должен быть (примерно) одинаков.
2. Дисперсия во всех выборках должна быть (примерно) одинакова.
3. Результирующая выборка должна подчиняться нормальному закону распределения.

Безусловно, такие ограничения не принимаются во внимание при моделировании, поскольку объемы выборки и методы генерации данных выбираются нами собственноручно, однако при исследовании реальных данных наборы и объемы исходных данных часто различны. Кроме того, при исследовании многофакторных моделей уровень дисперсии может сильно разниться в зависимости от факторов, их взаимодействия и самой модели.

Реализация программного средства моделирования

В рамках студенческой научно-исследовательской работы было спроектировано и разработано программное средство моделирования с анализом выходных данных с помощью модифицированной процедуры дисперсионного анализа, учитывающей вышеприведенные недостатки и позволяющей проводить анализ на более широком массиве данных. На вход программе подается программа на языке GPSS, которая с помощью препроцессора языка преобразуется во внутреннее программное представление для проведения моделирования и анализа полученных данных.

Структура программного комплекса представлена на рис. 4



Рис. 4. Структура программного комплекса

Выводы

Для расширения области применения дисперсионного анализа и снятия ограничений, накладываемых на начальные условия запуска процедуры, разработан программный комплекс, реализующий алгоритм, который обходит эти ограничения и позволяет проводить анализ на более разнообразных выборках данных, например при относительном недостатке данных при определенном уровне фактора и переизбытке при другом.

Список литературы

1. Рудаков И. В., Шляева А. В. Моделирование входных потоков данных для стохастических моделей дискретных систем // Вестник МГТУ им. Н. Э. Баумана. Сер. Приборостроение. - 2008. - № 2. - С. 65-72.
2. В. Дэвид Кельтон, Аверилл М. Лоу. Имитационное моделирование. М.: Питер., 2004, 846с.
3. В. Томашевский, Е. Жданова. Имитационное моделирование в среде GPSS. Серия «Факультет». М.: Бестселлер, 2003.
4. Статсофт. Дисперсионный анализ. URL: <http://www.statsoft.ru/home/textbook/modules/stanman.html#basic> (дата обращения 12.02.2013)
5. Боев.В. Д. Исследование адекватности GPSS World и AnyLogic при моделировании дискретно-событийных процессов: Монография. — СПб.: ВАС, 2011. — 404 с.

6. Лоскутов А. Ю., Михайлов А. С. Основы теории сложных систем. М. — Ижевск: НИЦ «Регулярная и стохастическая динамика», 2007. — 612 с. ISBN 5-93972-558-9.