

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Кузбасский государственный технический университет имени Т. Ф. Горбачева»

Кафедра информационных и автоматизированных
производственных систем

Составители
В. В. Зиновьев
А. Н. Стародубов
П. И. Николаев
И. С. Кузнецов

**АНАЛИЗ ЧУВСТВИТЕЛЬНОСТИ
ПРИ МОДЕЛИРОВАНИИ ПРОЦЕССОВ И СИСТЕМ**

**Методические указания к лабораторной работе
по дисциплине «Моделирование процессов и систем»**

Рекомендовано учебно-методической комиссией направления подготовки
09.03.02 «Информационные системы и технологии» в качестве электронного
издания для использования в учебном процессе

Кемерово 2017

Рецензент

Чичерин И. В., кандидат технических наук, заведующий кафедрой информационных и автоматизированных производственных систем

Анализ чувствительности при моделировании процессов и систем: методические указания к лабораторной работе по дисциплине «**Моделирование процессов и систем**» для студентов направления подготовки 09.03.02 «Информационные системы и технологии» очной формы обучения / сост. В. В. Зиновьев; А. Н. Стародубов; П. И. Николаев; И. С. Кузнецов; КузГТУ. – Кемерово, 2017.

Приведены основные теоретические положения, порядок выполнения лабораторной работы, вопросы для самоконтроля, рекомендуемая литература.

© КузГТУ, 2017

© В. В. Зиновьев; А. Н. Стародубов;
П. И. Николаев; И. С. Кузнецов,
составление, 2017

1. Цель работы

Цель работы – изучить технологию дисперсионного анализа для оценки чувствительности результатов моделирования к изменению входных параметров и изменению вида вероятностных распределений, используемых в модели.

В ходе выполнения лабораторной работы студент должен научиться формировать таблицу дисперсионного анализа при помощи программы табличных вычислений Microsoft Excel и выявлять зависимость результатов имитационных экспериментов от изменения входных параметров модели.

Примечание

Предполагается, что студент, знаком с основами работы с Microsoft Excel 2007 или с более поздними версиями, а также с основами имитационного моделирования систем в среде компьютерного моделирования GPSS World.

2. Теоретические положения

2.1. Задачи дисперсионного анализа при моделировании систем

Дисперсионный анализ (ANOVA – ANalysisOfVAriance) позволяет установить влияние отдельных факторов на изменчивость какого-либо отклика модели, значения которого могут быть получены путем проведения экспериментов с моделью.

Задачи дисперсионного анализа хотя и являются относительно простыми, но, тем не менее, часто используются на практике.

Например, стохастическая имитационная модель не может быть абсолютно точной, и необходимо как-то измерить степень этой неточности. Это измерение можно проделать (частично) путем анализа чувствительности отклика модели к изменению уровней факторов. Также важно определение чувствительности отклика к изменению вида вероятностных распределений, используемых в модели. Это делается для исключения эффектов, которые являются не более чем случайной флуктуацией. Более того, если при несущественных изменениях уровней факторов значение отклика модели меняется очень сильно, то необходимо проводить дополнительные исследования для получения более точных оценок. Если значение отклика при значительных изменениях уровней факторов изменяется не существенно, то дальнейшее экспериментирование нецелесообразно. Также анализ чувствительности посредством дисперсионного анализа может показать, как изменить (заменить) эти факторы, чтобы оптимизировать реальную систему.

Дисперсионный анализ широко используется при валидации модели путем выявления значимых факторов в системе и сравнении их с данными реальной системы или с экспертными оценками специалистов.

Центральная проблема проектирования эксперимента – выбор ограниченного набора комбинаций факторных уровней, которые будут фактически моделироваться при экспериментировании с моделью. Здесь также применяется дисперсионный анализ [1].

В отличие от экспериментов на реальных системах, эксперимент на имитационной модели является контролируемым. Можно варьировать любой фактор и судить о поведении модели по наблюдаемым откликам.

Дисперсионный анализ модели требует выполнения набора прогонов, при этом факторы в течение прогона моделирования изменяются от прогона к прогону. Каждый фактор имеет, по крайней мере, два уровня в эксперименте. Фактор может быть качественным (например, приоритетные правила или закон распределения). Дисперсионный анализ, основанный на статистической модели, заканчивается построением таблицы ANOVA, в которой анализируется влияние факторов А, В, взаимодействие между факторами АВ и случайные помехи наблюдения [2]. С помощью этой таблицы проверяется гипотеза об отсутствии влияния фактора. Если справедлива гипотеза об отсутствии влияния фактора, то считается, что все наблюдения получены из одной генеральной совокупности. Для проверки гипотезы используется F-распределение Фишера. Критерий Фишера определяет отношение двух выборочных дисперсий. Если фактор существенно влияет на отклик, то значения F-распределения принимает большие значения и F-статистика становится значимой. Таким образом, большие значения F приводят к отбрасыванию гипотезы об отсутствии влияния фактора, т.е. фактор является значимым [3].

Многие системы имитационного моделирования содержат встроенные библиотечные процедуры ANOVA. Также стандартные статистические методы обработки данных включены в состав программ табличных вычислений Microsoft Excel, Quatro Pro; в математические пакеты Mathcad, Matlab, Mapl. Еще более мощными возможностями статистической обработки обладают специализированные пакеты STATISTICA, STATGRAPHICS, SPSS.

Для изучения и понимания технологии дисперсионного анализа и, учитывая универсальность и распространенность табличного процессора Microsoft Excel (с встроенным пакетом анализа), используем именно этот программный продукт.

Опустим из рассмотрения теорию дисперсионного анализа, т.к. ей посвящено множество литературы по математической статистике. Рассмотрим практическое приложение этого метода к моделированию систем и технологию дисперсионного анализа в Microsoft Excel.

2.2. Технология дисперсионного анализа в Microsoft Excel

В зависимости от числа оказывающих влияние факторов различают однофакторный и многофакторный (двухфакторный и т.д.) дисперсионный анализ.

Режим работы «Однофакторный дисперсионный анализ» служит для выяснения факта влияния контролируемого входного параметра модели(фактора) A на результат моделирования (отклик) Y на основе выборочных данных результатов имитационных экспериментов. Режимы работы «Двухфакторный дисперсионный анализ без повторений» и «Двухфакторный дисперсионный анализ с повторениями» служат для выяснения на основе выборочных данных результатов экспериментов факта влияния контролируемых факторов A и B на отклик модели Y . При этом в режиме «Двухфакторный дисперсионный анализ без повторений» каждому уровню факторов A и B соответствует только одна выборка результатов экспериментов, а в режиме «Двухфакторный дисперсионный анализ с повторениями» каждому уровню одного из факторов A (или B) соответствует более одной выборки данных. В последнем случае число выборок для каждого уровня должно быть одинаковым.

Так как задачи однофакторного дисперсионного анализа являются самыми простыми в своем классе, а проведение имитационных экспериментов – управляемый процесс, при котором не составляет труда проводить множество прогонов модели (увеличивая тем самым статистическую значимость результатов), рассмотрим двухфакторный дисперсионный анализ с повторениями.

В диалоговом окне (рис. 2.1) задаются следующие параметры:

1. Входной интервал.
2. Число строк для выборки – вводится число выборок, приходящихся на каждый уровень одного из факторов.
3. Альфа – вводится уровень значимости α , равный вероятности возникновения ошибки первого рода (отвержение нулевой гипотезы).
4. Выходной интервал/Новый рабочий лист/Новая рабочая книга.

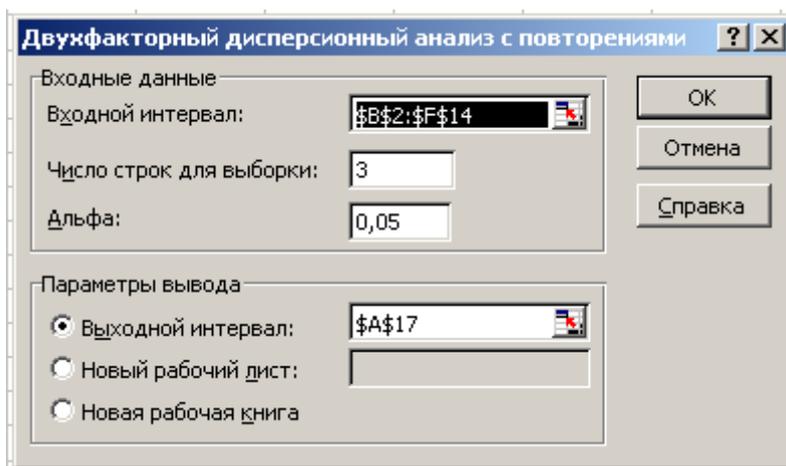


Рис. 2.1

2.3. Пример дисперсионного анализа

Предположим, что на имитационной модели некоторой системы проведен ряд экспериментов. Для анализа чувствительности отклика модели Y к

факторам A и B используем встроенный в Microsoft Excel режим «Двухфакторный дисперсионный анализ с повторениями». Для этого необходимо:

1. По результатам имитационных экспериментов сформировать на рабочем листе Microsoft Excel выборочные данные об отклике модели, при изменении факторов A и B на 4-х уровнях, используя результаты трех прогонов для каждого уровня.

2. Сформировать в Microsoft Excel соответствующую таблицу и при уровне значимости $\alpha=0,05$ выяснить, влияют ли на отклик Y факторы A и B .

Выборочные данные для отклика Y при различных уровнях факторов A и B , приведены на рис. 2.2.

	A	B	C	D	E	F
1	Прогоны	Фактор А		Фактор В		
2	модели		Уровень 1	Уровень 2	Уровень 3	Уровень 4
3	Прогон 1	Уровень 1	21,4	20,9	19,6	17,6
4	Прогон 2		21,2	20,3	18,8	16,6
5	Прогон 3		20,1	19,8	16,4	17,5
6	Прогон 1	Уровень 2	12	13,6	13	13,3
7	Прогон 2		14,2	13,3	13,7	14
8	Прогон 3		12,1	11,6	12	13,9
9	Прогон 1	Уровень 3	13,5	14	12,9	12,4
10	Прогон 2		11,9	15,6	12,9	13,7
11	Прогон 3		13,4	13,8	12,1	13
12	Прогон 1	Уровень 4	12,8	14,1	14,2	12
13	Прогон 2		13,8	13,2	13,6	14,6
14	Прогон 3		13,7	15,3	13,3	14

Рис. 2.2

Для решения задачи используем режим работы «Двухфакторный дисперсионный анализ с повторениями». Значения параметров, установленных в одноименном диалоговом окне, представлены на рис. 2.2, а необходимая часть рассчитанных в данном режиме показателей – на рис. 2.3.

	A	B	C	D	E	F	G	H
50								
51	Дисперсионный анализ							
52	Источник вари	SS	df	MS	F	P-Значение	критическое	
53	Выборка	309,259	3	103,0863	123,6418	1,11E-17	2,901118	
54	Столбцы	9,970625	3	3,323542	3,986257	0,016084	2,901118	
55	Взаимоде	25,67854	9	2,853171	3,422095	0,00473	2,188763	
56	Внутри	26,68	32	0,83375				
57								
58	Итого	371,5881	47					
59								

Рис. 2.3

Как видим, расчетное значение F -критерия фактора $A - F_p^A = 123,64$, а критическая область образуется правосторонним интервалом $(2,90; +\infty)$. Так как F_p^A попадает в критическую область, то гипотезу H_A о равенстве выборочных средних отвергаем, т.е. считаем, что фактор A влияет на отклик Y .

Выборочный коэффициент детерминации для фактора A

$$\tilde{\rho}_A^2 = \frac{\tilde{\sigma}_A^2}{\tilde{\sigma}_Y^2} = \frac{309,26}{371,59} \approx 0,83$$

показывает, что 83% общей выборочной вариации отклика Y связано с влиянием фактора A .

Расчетное значения F -критерия фактора $B - F_p^B = 3,99$, а критическая область образуется правосторонним интервалом $(2,90; +\infty)$. Так как F_p^B попадает в критическую область, то гипотезу H_B отвергаем, т.е. считаем, что фактор B влияет на отклик Y .

Выборочный коэффициент детерминации для фактора B

$$\tilde{\rho}_B^2 = \frac{\tilde{\sigma}_B^2}{\tilde{\sigma}_Y^2} = \frac{9,97}{371,59} \approx 0,03$$

показывает, что только около 3% общей выборочной вариации отклика Y связано с влиянием фактора B .

Значимость фактора взаимодействия F_p^{AB} ($F_p^{AB} = 3,42$ и попадает в критический интервал $(2,19; +\infty)$) указывает на то, что фактор A варьируется при различных уровнях фактора B .

3. Задание к лабораторной работе

Описание системы

Вычислительная система состоит из n каналов и буфера, где могут одновременно находиться не более двух пакетов данных. Каждый пакет данных передаётся по каждому каналу за случайное время, распределенное по экспоненциальному закону в среднем за x миллисекунд. Если в буфере уже находятся два пакета данных, то следующий уходит на резервный канал вычислительной системы. Интервалы поступления пакетов данных распределены экспоненциально с интенсивностью – λ пакетов в секунду.

Значения n , λ и x по вариантам представлены в табл. 3.1.

Таблица 3.1. – Значения количества каналов (n), интенсивности поступления сообщений (λ) и времени передачи сообщения x

В а р и а н т														
1			2			3			4			5		
n	λ	x	n	λ	x	n	λ	x	n	λ	x	n	λ	x
1	2	200	2	5	300	1	3	300	2	3	300	1	5	300

При выполнении задания данной лабораторной работы необходимо:

1. Используя среду компьютерного моделирования GPSS World, отобразить работу информационной системы.

2. Определить число пакетов данных, ушедших на резервный канал, при общей передаче 100 пакетов данных.

3. Провести имитационные эксперименты и сформировать на рабочем листе Microsoft Excel выборочные данные о числе пакетов данных, ушедших на резервный канал, при различных значениях интенсивности поступления пакетов (2 и 5 в секунду) и емкости буфера (1 и 2 пакета).

4. Используя режим «Двухфакторный дисперсионный анализ с повторениями» (4 повторения для каждого эксперимента) сформировать в Microsoft Excel соответствующую таблицу и при уровне значимости $\alpha = 0,05$ выяснить, влияют ли на число пакетов данных, ушедших на резервный канал, интенсивность поступления пакетов данных и ёмкость буфера.

5. Изменить имитационную модель так, чтобы пакеты данных передавались по каналу за равномерно распределенное время 300 ± 60 мс, а среднее значение интервалов поступления пакетов данных было в пределах 200 ± 60 мин. (распределение равномерное).

6. Провести имитационные эксперименты и сформировать на рабочем листе Microsoft Excel выборочные данные о числе пакетов, ушедших на резервный канал, при различных видах законов распределения.

7. Используя режим «Двухфакторный дисперсионный анализ с повторениями» (4 повторения для каждого эксперимента), сформировать в Microsoft Excel соответствующую таблицу, и при уровне значимости $\alpha = 0,05$ выяснить, влияет ли на число пакетов данных, ушедших на резервный канал, вид закона распределения.

4. Порядок выполнения работы

1. Ознакомьтесь с методическими указаниями по выполнению данной лабораторной работы.
2. Получите у преподавателя вариант задания.
3. Отобразите работу магистрали передачи данных на языке GPSS World и определите число сообщений, ушедших на резервный канал.
4. Проведите необходимые имитационные эксперименты, по результатам которых сформируйте в Microsoft Excel таблицы дисперсионного анализа и выясните влияние на число пакетов данных, ушедших на резервный канал, ёмкости буфера, интенсивности поступления пакетов данных и вида закона распределения последней.
5. Сделайте выводы.

5. Требования к отчету

Отчет о работе должен содержать:

1. Задание и исходные данные по заданному варианту лабораторной работы.
2. Модель магистрали передачи данных на языке GPSS World (файл.gps).
3. Таблицы дисперсионного анализа, сформированные по результатам имитационных экспериментов, в Microsoft Excel (файлы .xls).
4. Анализ полученных результатов и выводы по работе.

6. Контрольные вопросы

1. В чем суть анализа чувствительности при имитационном моделировании систем?
2. Как дисперсионный анализ помогает в оценке чувствительности при имитационном моделировании систем?
3. Перечислите задачи, возникающие в процессе имитационного моделирования систем, в которых применяется ANOVA.
4. Для чего применяют инструмент «Дисперсионный анализ» встроенный в табличный процессор Microsoft Excel?
5. В чем различие однофакторного и многофакторного дисперсионного анализа?

6. Какие виды двухфакторного дисперсионного анализа можно проводить в Microsoft Excel и в чем их различие?

7. Список рекомендуемой литературы

1. Шеннон Р. Имитационное моделирование систем – искусство и наука. – М.: Мир, 1978.

2. Томашевский В. Имитационное моделирование в среде GPSS. Серия «Факультет» / В. Томашевский, И. Жданова. – М.: Бестселлер, 2003.

4. Зиновьев В. В. Моделирование процессов и систем: учеб. пособие / В. В. Зиновьев, А. Н. Стародубов, П. И. Николаев; КузГТУ. – Кемерово, 2016. – 146 с.

3. Макарова Н. В. Статистика в Excel: учеб. пособие / Н. В. Макарова, В. Я. Трофимец. – М.: Финансы и статистика, 2002.

Оглавление

1. Цель работы.....	0
2. Теоретические положения	2
2.1. Задачи дисперсионного анализа при моделировании систем	2
2.2. Технология дисперсионного анализа в Microsoft Excel.....	3
2.3. Пример дисперсионного анализа	4
3. Задание к лабораторной работе.....	7
4. Порядок выполнения работы.....	8
5. Требования к отчету	8
6. Контрольные вопросы	8
7. Список рекомендуемой литературы	9