
Санкт-Петербургский государственный
политехнический университет

Институт информационных технологий и управления

Кафедра компьютерных систем и программных технологий

Сиднев А.Г.

**Методические указания
к лабораторным работам по теме
«Применение дисперсионного анализа в среде GPSS
World»**

**Санкт-Петербург
2014**

Содержание

1. Задача дисперсионного анализа	3
2. Критерий Фишера	3
3. Однофакторный анализ – One-Way ANOVA	4
4. Многофакторный анализ – Multi-Way ANOVA.....	5
5. Многофакторный анализ (Multi-Way ANOVA) с учетом совместного влияния факторов.....	7
6. Примеры дисперсионного анализа результатов имитационных экспериментов в среде GPSS World	8
6.1. Однофакторный дисперсионный анализ.....	8
6.2. Двухфакторный дисперсионный анализ	13
7. Литература	16

1. Задача дисперсионного анализа

Дисперсионный анализ состоит в выделении и оценке отдельных факторов, вызывающих изменчивость среднего значения наблюдаемой случайной величины [1]. С этой целью производится разложение дисперсии наблюдаемой частичной совокупности на составляющие, порождаемые независимыми факторами. Каждая из этих составляющих получает свою оценку в общей совокупности.

Для проверки значимости факторов оценивают значимость указанных составляющих дисперсии, сравнивая их с так называемой остаточной дисперсией, обусловленной случайным составом частичной совокупности.

Проверка значимости оценок дисперсии производится с помощью критерия Фишера, использующего статистику F (распределение Фишера-Снедекора) [1].

В англо-язычной литературе термину «дисперсионный анализ» соответствует аббревиатура ANOVA – analysis of variation [2]. Дословно «анализ дисперсии».

Если исследуется влияние одного фактора, то речь идет о простой группировке – однофакторном анализе или One-Way ANOVA. При одновременном исследовании влияния нескольких факторов анализ называется многофакторным или Multi-Way ANOVA.

Дисперсионный анализ позволяет принять или опровергнуть гипотезу о значимости влияния исследуемого фактора на изменчивость среднего значения наблюдаемой случайной величины. Более точно речь идет о нулевой гипотезе о том, что все значения в выборке распределены нормально и одинаково, то есть рассматриваемые факторы не оказывают на наблюдаемую величину существенного влияния.

Как правило, после выбора значимых факторов осуществляется построение модели, обычно, линейной регрессионной, связывающей значение наблюдаемой величины со значениями влияющих на нее факторов.

Форма изложения однофакторного и многофакторного анализа заимствована в [2].

2. Критерий Фишера

Критерий Фишера позволяет ответить на вопрос, принадлежат ли две частичных выборки общей нормальной генеральной совокупности с одинаковой дисперсией [1].

Для указанных выборок объемом I_1 и I_2 необходимо получить оценки дисперсии S_1^2 и S_2^2 :

$$S_1^2 = \frac{\sum_{i=1}^{I_1} (x_{1i} - \bar{x}_1)^2}{I_1 - 1}; \quad S_2^2 = \frac{\sum_{i=1}^{I_2} (x_{2i} - \bar{x}_2)^2}{I_2 - 1}.$$

Затем вычисляется статистика Фишера

$$F = \frac{S_1^2}{S_2^2}, \text{ причем } S_1^2 > S_2^2.$$

Пусть $F_{(I_1-1), (I_2-1)} > F_\alpha$, тогда нулевая гипотеза о том, что выборки принадлежат одной генеральной совокупности, отвергается.

3. Однофакторный анализ – One-Way ANOVA

Исследуется влияние фактора A

Обозначения

y_{ij} – отклик, полученный для i -го уровня фактора A в j -м эксперименте (репликации или реплике)

$i = \overline{1, I}$; I – число уровней фактора A

$j = \overline{1, J}$; J – число репликаций для каждого уровня

Общее число опытов, то есть объем выборки IJ

Модель, описывающая влияние фактора A на отклик, следующая:

$$y = \mu + \alpha A + \varepsilon,$$

где

μ – постоянная составляющая,

α – коэффициент регрессии,

ε – случайная составляющая, имеющая смысл ошибки измерения отклика

$y_{i.}$ – среднее значение отклика в i -й группе (для i -го уровня) фактора A

$y_{..}$ – среднее значение отклика в выборке

Введенные обозначения дополнительно поясняются табл. 1.

Таблица 1

Реплика j	Уровни фактора A						
	1	2	...	i	...	I	
1	y_{11}	y_{12}		y_{1i}		y_{1I}	
2	y_{21}	y_{22}		y_{2i}		y_{2I}	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
J	y_{J1}	y_{J2}		y_{Ji}		y_{JI}	
Среднее	$y_{1.}$	$y_{2.}$		$y_{i.}$		$y_{I.}$	$y_{..}$

Расчет оценок дисперсий производится далее на основании следующих соотношений, позволяющих вычислить составляющие сумм квадратов отклонений.

$y_{ij} - y_{..} = (y_{i.} - y_{..}) + (y_{ij} - y_{i.})$, тогда

$$\underbrace{\sum_{j=1}^J \sum_{i=1}^I (y_{ij} - y_{..})^2}_S = \underbrace{\sum_{j=1}^J \sum_{i=1}^I (y_{i.} - y_{..})^2}_{S_A} + \underbrace{\sum_{j=1}^J \sum_{i=1}^I (y_{ij} - y_{i.})^2}_{S_R}$$

Отметим, что

$$\sum_{j=1}^J \sum_{i=1}^I (y_{i.} - y_{..})(y_{ij} - y_{i.}) = \sum_{i=1}^I (y_{i.} - y_{..}) \sum_{j=1}^J (y_{ij} - y_{i.}) = 0, \text{ а}$$

$$S_A = \sum_{j=1}^J \sum_{i=1}^I (y_{i.} - y_{..})^2 = J \sum_{i=1}^I (y_{i.} - y_{..})^2 \text{ и}$$

Полученные суммы квадратов используются для вычисления оценок дисперсий, которые определяют значение статистики Фишера, табл. 2.

Таблица 2

S – сумма квадратов	Число степеней свободы	Оценка дисперсии	F – статистика Фишера
$S = \sum_{j=1}^J \sum_{i=1}^I (y_{ij} - y_{..})^2$	$IJ - 1$	$MS = \frac{S}{IJ - 1}$	
$S_A = J \sum_{i=1}^I (y_{i.} - y_{..})^2$	$I - 1$	$MS_A = \frac{S_A}{I - 1}$	$\frac{MS_A}{MS_R}$
$S_R = \sum_{j=1}^J \sum_{i=1}^I (y_{ij} - y_{i.})^2$	$IJ - I$	$MS_R = \frac{S_R}{IJ - I}$	

Если $\frac{MS_A}{MS_R} > F_{(I-1),(J-1)}$ для заданного уровня значимости, то нулевая гипотеза о несущественности влияния фактора A отвергается.

4. Многофакторный анализ – Multi-Way ANOVA

Рассмотрим пример двойной группировки данных при проведении двухфакторного эксперимента, устанавливающего степень существенности влияния факторов A и B . Предполагается следующая модель влияния факторов на отклик:

$$y = \mu + \alpha_A A + \alpha_B B + \varepsilon$$

Обозначения

$i = 1, I$; I – число уровней фактора A

$j = \overline{1, J}$; J – число уровней фактора B

В данном случае рассматриваются две нулевые гипотезы, каждая о несущественности влияния соответствующего фактора.

Формула для расчета квадратов отклонений следующая:

$$S = S_A + S_B + S_R$$

$$\underbrace{\sum_{j=1}^J \sum_{i=1}^I (y_{ij} - y_{..})^2}_S = \underbrace{\sum_{j=1}^J \sum_{i=1}^I (y_{i.} - y_{..})^2}_{S_A} + \underbrace{\sum_{j=1}^J \sum_{i=1}^I (y_{.j} - y_{..})^2}_{S_B} +$$

$$\underbrace{\sum_{j=1}^J \sum_{i=1}^I (y_{ij} - y_{.j} - y_{i.} + y_{..})^2}_{S_R}$$

Очевидно, что

$$S_A = J \sum_{i=1}^I (y_{i.} - y_{..})^2, \text{ а}$$

$$S_B = I \sum_{j=1}^J (y_{.j} - y_{..})^2$$

Расчет оценок дисперсий и значений статистики Фишера поясняется содержимым табл. 3.

Таблица 3

S – сумма квадратов	Число степеней свободы	Оценка дисперсии	F – статистика Фишера
$S = \sum_{j=1}^J \sum_{i=1}^I (y_{ij} - y_{..})^2$	$IJ - 1$	$MS = \frac{S}{IJ - 1}$	
$S_A = J \sum_{i=1}^I (y_{i.} - y_{..})^2$	$I - 1$	$MS_A = \frac{S_A}{I - 1}$	$\frac{MS_A}{MS_R}$
$S_B = I \sum_{j=1}^J (y_{.j} - y_{..})^2$	$J - 1$	$MS_B = \frac{S_B}{J - 1}$	$\frac{MS_B}{MS_R}$
$\underbrace{\sum_{j=1}^J \sum_{i=1}^I (y_{ij} - y_{.j} - y_{i.} + y_{..})^2}_{S_R}$	$(I - 1)(J - 1)$	$MS_R = \frac{S_R}{(I - 1)(J - 1)}$	

Отметим, что в данном варианте двухфакторного анализа для каждого сочетания факторов A и B получена всего одна реплика.

Обращаем внимание также на то, что результаты двухфакторного анализа могут быть поданы как результаты однофакторного анализа для отдельного учета влияния факторов A и B . В последнем случае число

уровней фактора B (A) используется как число репликаций при анализе фактора A (B).

Результаты двухфакторного и однофакторных анализов могут оказаться различными, так как при двухфакторном анализе остаточная сумма квадратов S_R получается существенно меньшей, а число степеней свободы уменьшается незначительно: $IJ - I - J + 1$ против $IJ - I$.

5. Многофакторный анализ (Multi-Way ANOVA) с учетом совместного влияния факторов

Предполагается следующая модель влияния факторов на отклик:

$$y = \mu + \alpha_A A + \alpha_B B + \alpha_{AB} AB + \varepsilon$$

Обозначения

$i = \overline{1, I}$; I – число уровней фактора A

$j = \overline{1, J}$; J – число уровней фактора B

$m = \overline{1, M}$; M – число реплик для каждого сочетания ij уровней факторов A и B

Рассматриваются три нулевые гипотезы: две о несущественности влияния фактора A и фактора B и одна о несущественности совместного влияния факторов A и B (фактора AB).

Расчет сумм квадратов отклонений осуществляется по следующим очевидным формулам.

$$S = \sum_{i=1}^I \sum_{j=1}^J \sum_{m=1}^M (y_{ijm} - y_{...})^2$$

$$S_A = \sum_{i=1}^I \sum_{j=1}^J \sum_{m=1}^M (y_{i..} - y_{...})^2 = \sum_{i=1}^I JM (y_{i..} - y_{...})^2$$

$$S_B = \sum_{i=1}^I \sum_{j=1}^J \sum_{m=1}^M (y_{.j.} - y_{...})^2 = \sum_{j=1}^J IM (y_{.j.} - y_{...})^2$$

$$S_{AB} = \sum_{i=1}^I \sum_{j=1}^J \sum_{m=1}^M (y_{ij.} - y_{i..} - y_{.j.} + y_{...})^2 = \sum_{i=1}^I \sum_{j=1}^J M (y_{ij.} - y_{i..} - y_{.j.} + y_{...})^2$$

$$S_R = S - S_A - S_B - S_{AB},$$

где $\left. \begin{matrix} y_{ij} \\ y_{i..} \\ y_{.j.} \\ y_{...} \end{matrix} \right\}$ – означают осреднение по индексам, замененным точками.

Расчет оценок дисперсий и значений статистики Фишера поясняется содержимым табл. 4.

Таблица 4

Сумма квадратов	Число степеней свободы	Оценка дисперсии	F – статистика Фишера
S	$IJM - 1$	$MS = \frac{S}{IJM - 1}$	
S_A	$I - 1$	$MS_A = \frac{S_A}{I - 1}$	$\frac{MS_A}{MS_R}$
S_B	$J - 1$	$MS_B = \frac{S_B}{J - 1}$	$\frac{MS_B}{MS_R}$
S_{AB}	$(I - 1)(J - 1)$	$MS_{AB} = \frac{S_{AB}}{(I - 1)(J - 1)}$	$\frac{MS_{AB}}{MS_R}$
S_R	$IJ(M - 1)$	$MS_R = \frac{S_R}{IJ(M - 1)}$	

Важно учитывать, что совместное влияние факторов имеет смысл рассматривать только при условии $M > I$. В противном случае остаточная сумма квадратов S_R становится равной нулю.

6. Примеры дисперсионного анализа результатов имитационных экспериментов в среде GPSS World

6.1. Однофакторный дисперсионный анализ

Рассматривается вариант использования команды ANOVA на примере программы Anova2.gps, приведенной в папке Sample Models установленной системы моделирования GPSS World. Стандартный путь, ведущий к указанной папке с примерами программ моделирования: c:\Program Files\Minuteman Software\GPSS World Student Version\Sample Models.

Команда ANOVA обеспечивает анализ влияния до шести факторов, включая взаимодействия 2-го и 3-го порядков и оперирует с данными, содержащимися в матрице, которая автоматически формируется в результате выполнения экспериментальных прогонов имитационной модели. План эксперимента и формат матрицы данных задаются файлом с расширением «txt», в общем случае содержащим команды и блоки GPSS.

Программа Anova.gps, в данном случае, подвергнута некоторому упрощению (Anova2.gps), моделирует работу разомкнутой одноканальной системы массового обслуживания с неограниченной очередью.

Исследуется влияние величины среднего времени обслуживания заявки на время её ожидания в очереди.

Дисперсионному анализу подвергаются любые величины, наблюдаемые в ходе исполнения программы моделирования. Выбор задачи моделирования одноканальной СМО обусловлен её простотой.

Текст программы ANOVA2.GPS

```
; GPSS World Sample File - ANOVA2.GPS
*****
*
*           Barber Shop Simulation
*           Time is in minutes
*
*****

GENERATE      5,1.7           ;Create next customer.
QUEUE         Barber         ;Begin queue time.
SEIZE         Barber         ;Own or wait for barber.
DEPART       Barber         ;End queue time.
ADVANCE       Cut_Time       ;Cut takes a few min.
RELEASE       Barber         ;Give up the barber.
```

```
; INCLUDE      "CTLANOVA2.TXT"
TERMINATE
;SHOW ANOVA(Results,2,1)
```

Команда INCLUDE "CTLANOVA2.TXT" подключает упомянутый выше текстовый файл, определяющий параметры эксперимента. Файл CTLANOVA2.TXT должен находиться в той же папке, что и программа Anova2.gps.

Рассмотрим структуру файла "CTLANOVA2.TXT"

файл "CTLANOVA2.TXT"

```
; GPSS World Sample File - CTLANOVA2.GPS

RESULTS MATRIX ,2,3      ; Объявление матрицы (2x3) - 2 уровня фактора по 3
                          ; репликации для каждого

Cut_Time EQU 6.8         ; Задание среднего времени обл.
Treatment EQU 1
RMULT 411
Start 100,NP             ; Отчет не создается
MSAVEVALUE RESULTS,1,1,QT$Barber ; Запись времени ожидания в матрицу

Clear Off                ; Удаление транзактов, обнуление счетчиков
                          ; Сохраняемые величины и матрицы не обнуляются!

RMULT 421
Start 100,NP
MSAVEVALUE RESULTS,1,2,QT$Barber

Clear Off
RMULT 431
Start 100,NP
MSAVEVALUE RESULTS,1,3,QT$Barber

Clear Off
Cut_Time EQU 5
Treatment EQU 2
```

```
RMULT 411
Start 100,NP
MSAVEVALUE RESULTS,2,1,QT$Barber
```

```
Clear Off
RMULT 421
Start 100,NP
MSAVEVALUE RESULTS,2,2,QT$Barber
```

```
Clear Off
RMULT 431
Start 100,NP
MSAVEVALUE RESULTS,2,3,QT$Barber
```

Этот файл содержит совокупность команд и блоков GPSS World.

Для корректного исполнения программы Anova2.gps под управлением файла CTLANOVA2.TXT необходимо придерживаться следующей последовательности действий.

1. Провести трансляцию – команда Create Simulation
2. Выполнить команду INCLUDE CTLANOVA2.TXT" обязательно из командного окна (команда INCLUDE CTLANOVA2.TXT" не располагается в тексте программы Anova2.gps – закомментирована «;»), рис. 1:

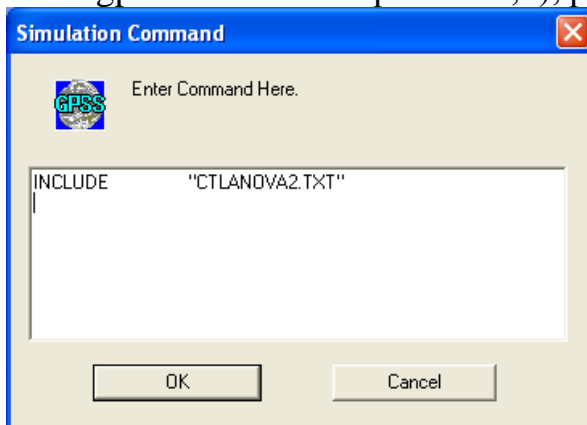


Рис.1

В результате выполнения этой команды в журнале появится следующая запись, рис.2:

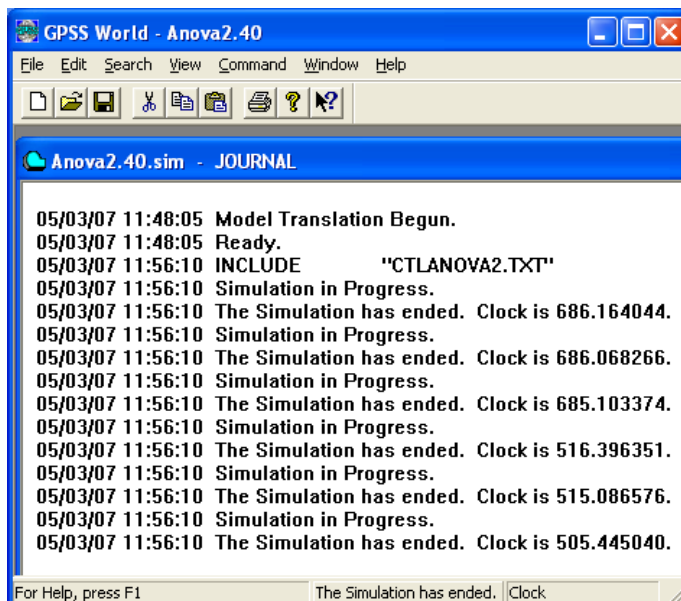


Рис.2

3. Выполняется команда SHOW ANOVA(Results,2,1), рис.3:

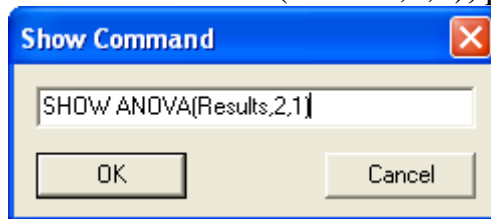


Рис.3

Результат выполнения этой команды появляется в виде записи в журнале, рис.4, а интерпретация полученных данных приводится в табл.4.

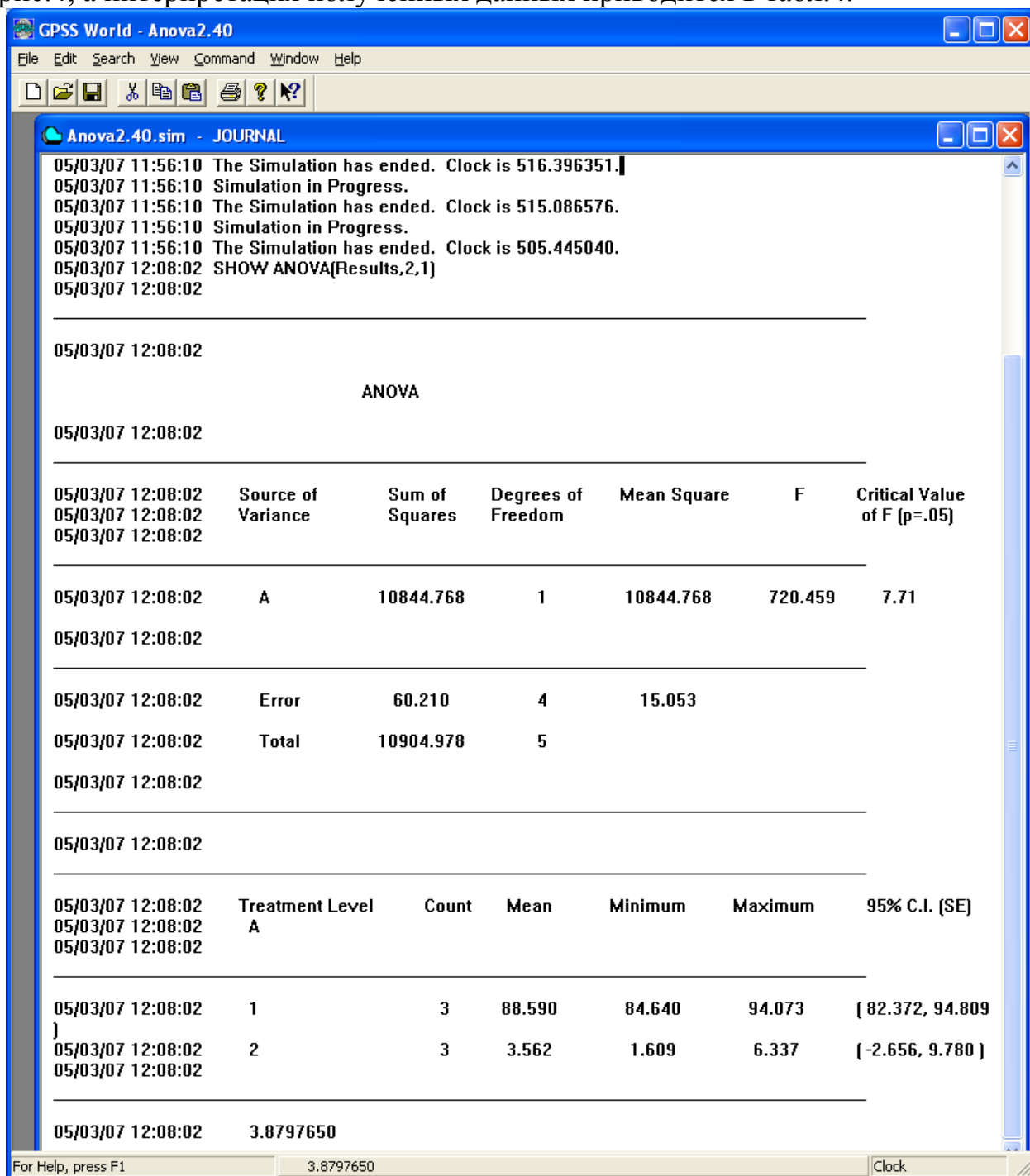


Рис.4

В качестве пояснения синтаксиса команды ANOVA ниже приводится выдержка из Help GPSS World:

ANOVA Syntax

```
StandardError = ANOVA(ResultMatrixName,ReplicateDimension,InteractionLimit)
```

Arguments

ResultMatrixName - The GPSS Matrix containing the results to be analyzed. Required.

May be a GPSS Matrix of up to 6 dimensions of any shape. The argument must be Name.

ReplicateDimension - The dimension of the Result Matrix used for replicates. Use 0 if there are no replicates. Required. Coerced to integer. The argument must be Expression.

InteractionLimit - The Limit of Factor Interactions to be analyzed. This value can be 1, 2, or 3. It is normally used to preserve Degrees of Freedom for the estimate of the Standard Error instead of using them on an interaction presumed to be unimportant. Required. Coerced to integer. The argument must be Expression.

В этом тексте говорится о формате команды ANOVA:

ANOVA(Имя_Матрицы,Номер_размерности_репликации,Порядок взаимодействия_между_факторами)

Номер_размерности_репликации принимает значение 0, если повторение экспериментов отсутствует – каждый набор факторов реализуется в одиночном эксперименте.

Порядок взаимодействия_между_факторами принимает значение от 0 до 3:

1 – взаимодействие не учитывается

2 – учитываются взаимодействия 2-го порядка (например, **AB**, **BC**, **AC**)

3 – учитываются взаимодействия 2-го и 3-го порядков (например, **AB**, **BC**, **AC** и **ABC**)

Взаимодействия более высоких порядков не поддерживаются.

Важно еще раз отметить, что включенные в текстовый файл типа STLANOVA2.TXT блоки GPSS выполняются только в том случае, если команда INCLUDE STLANOVA2.TXT" выполняется из командного окна системы GPSS World. Таким образом, только указанный способ выполнения команды INCLUDE STLANOVA2.TXT" гарантирует её корректное выполнение.

Представленная на рис.4 таблица содержит интерпретацию результатов дисперсионного анализа влияния фактора **A**. Полученное значение статистики

Фишера существенно превышает критическое значение для уровня значимости 5%: $F > (F_{0,05} = 7,71)$.

Таблица 4

S – сумма квадратов	Число степеней свободы	Оценка дисперсии	F – статистика Фишера
$S = 10904,978$	5	не выводится	
$S_A = 10844,768$	1	$MS_A = 10844,768$	$\frac{MS_A}{MS_R} = 720,459$
$S_R = 60,210$	4	$MS_R = 15,053$	$F > (F_{0,05} = 7,71)$

6.2. Двухфакторный дисперсионный анализ

Ниже приводится текст программы Anova_Multi.gps, взаимодействующей с текстовым файлом CTLANOVA_MULTI.TXT.

текст программы Anova_Multi.gps

- * Это программа в комплекте с текстовым файлом CTLANOVA_MULTI.TXT,
- * подключаемым командой INCLUDE, обеспечивает проведение 12 серий
- * экспериментов. Для каждой серии СЧА QT\$Barber записывается в соответствующую
- * ячейку матрицы RESULTS. Для этой цели в тексте программы прописана
- * процедура RESULTS(First,Second,Third), позволяющая подобную запись в матрицу,
- * имеющую более 2-х измерений.
- * Обращаем внимание на то, что в представленном файле CTLANOVA_MULTI.TXT,
- * содержатся только одни
- * команды GPSS и ни одного блока.
- * команда SHOW ANOVA(RESULTS,3,2) предполагает анализ влияния факторов А и
- * В, а также совместного влияния этих факторов, то есть фактора АВ.

```

GENERATE INTERVAL,(INTERVAL/4) ;Create next customer.
QUEUE Barber ;Begin queue time.
SEIZE Barber ;Own or wait for barber.
DEPART Barber ;End queue time.
ADVANCE Cut_Time ;Cut takes a few min.
RELEASE Barber ;Give up the barber.
SAVEVALUE 1,(RESULTS(First,Second,Third))
TERMINATE 1

INCLUDE "CTLANOVA_MULTI.TXT" ;Call runtime command file.
; SHOW ANOVA(RESULTS,3,2)

PROCEDURE RESULTS(First,Second,Third) BEGIN
TEMPORARY ANSWER;

RESULTS[First,Second,Third] = QT$Barber;

ANSWER = RESULTS[First,Second,Third];
RETURN ANSWER;

END;
```

Файл CTLANOVA_MULTI.TXT

```

; GPSS World Sample File - CTLANOVA_MULTI.GPS

RESULTS MATRIX ,2,2,3 ; Объявление матрицы (2x2x3)– 2 уровня фактора А,
```

; 2 уровня фактора В
; по 3 репликации для каждого сочетания уровней
; факторов А и В

INITIAL RESULTS, UNSPECIFIED

RMULT 411
Cut_time EQU 5
INTERVAL EQU 2
First EQU 1
Second EQU 1
Third EQU 1 ; (1)1 1 1
Start 100,NP
Clear Off

RMULT 421
Third EQU 2 ; (2)1 1 2
Start 100,NP
Clear Off

RMULT 521
Third EQU 3 ; (3)1 1 3
Start 100,NP
Clear Off

RMULT 621
INTERVAL EQU 4
Second EQU 2
Third EQU 1 ; (4)1 2 1
Start 100,NP
Clear Off

RMULT 531
Third EQU 2 ; (5)1 2 2
Start 100,NP
Clear Off

RMULT 631
Third EQU 3 ; (6)1 2 3
Start 100,NP
Clear Off

RMULT 711
Cut_time EQU 10
INTERVAL EQU 2
First EQU 2
Second EQU 1
Third EQU 1 ; (7)2 1 1
Start 100,NP
Clear Off

RMULT 821
Third EQU 2 ; (8)2 1 2
Start 100,NP
Clear Off

RMULT 121
Third EQU 3 ; (9)2 1 3
Start 100,NP
Clear Off

RMULT 331
INTERVAL EQU 4

```
Second EQU 2
Third EQU 1 ; (10)2 2 1
Start 100,NP
Clear Off
```

```
RMULT 231
Third EQU 2 ; (11)2 2 2
Start 100,NP
Clear Off
```

```
RMULT 931
Third EQU 3 ; (12)2 2 3
Start 100,NP
```

На рис. 5 приводится результат выполнения команды
SHOW ANOVA(RESULTS,3,2).

Дисперсионный анализ свидетельствует о существенном влиянии факторов А и В (среднего времени обслуживания *Cut_time* и среднего интервала поступления *INTERVAL*) ($F = 36327.483 > F_{0,05} = 5.32$ и $F = 5681.574 > F_{0,05} = 5.32$). Совместное влияние указанных факторов незначительно: $F = 0.507 < F_{0,05} = 5.32$ – ожидаемый результат, поскольку совместное одинаковое однонаправленное изменение факторов А и В не приводит к изменению коэффициента загрузки $\rho = \frac{Cut_time}{INTERVAL}$, являющегося определяющим фактором влияния на показатели исследуемой одноканальной СМО.

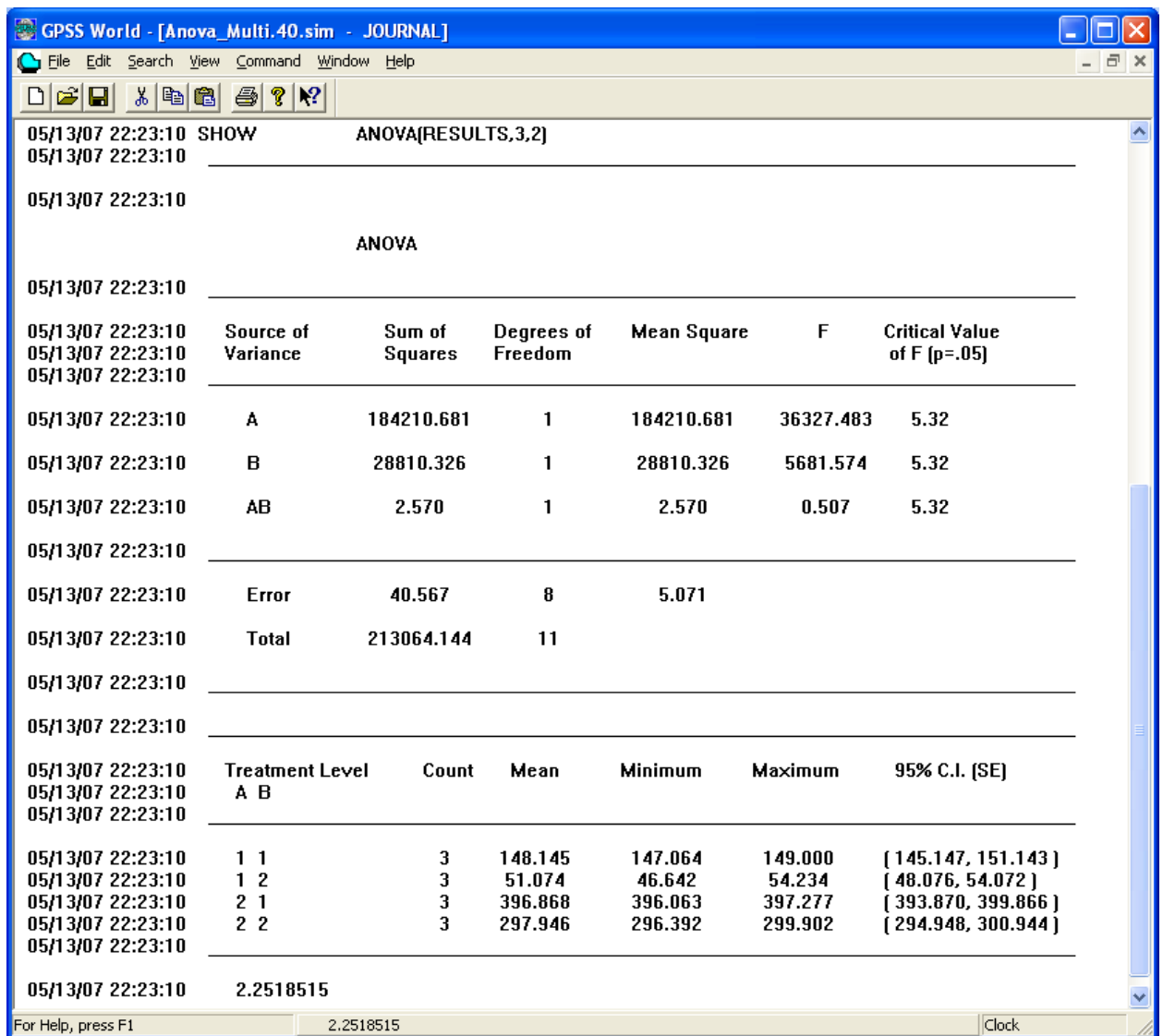


Рис. 5

В нижней части таблицы результатов приводятся данные о средних значениях и доверительных интервалах времени ожидания заявки, полученных в 4-х сериях экспериментов.

7. Литература

1. А.К. Митропольский. Техника статистических вычислений – М.: Наука, 1971. – 576 с.
2. Jerry Banks, John S Carson Barry L. Nelson Discrete-Event System Simulation. 2nd ed. Prentice-Hall, 1996. – 548 p.