

Дисперсионный анализ результатов моделирования.

Однородность статистического материала

Выполнили: студенты гр. ИТм-191

Филиппов С. С.

Чернышов И.В.

Понятие дисперсионного анализа.

Дисперсионный анализ – анализ изменчивости признака под влиянием каких-либо контролируемых переменных факторов.

В основе дисперсионного анализа лежит предположение о том, что одни переменные могут рассматриваться как причины (факторы, независимые переменные), а другие как следствия (зависимые переменные).

Независимые переменные называют иногда регулируемыми факторами именно потому, что в эксперименте исследователь имеет возможность варьировать ими и анализировать получающийся результат.

Цель и задачи дисперсионного анализа

Основной целью дисперсионного анализа является исследование значимости различия между средними с помощью сравнения (анализа) дисперсий.

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.

Гипотезы при дисперсионном анализе

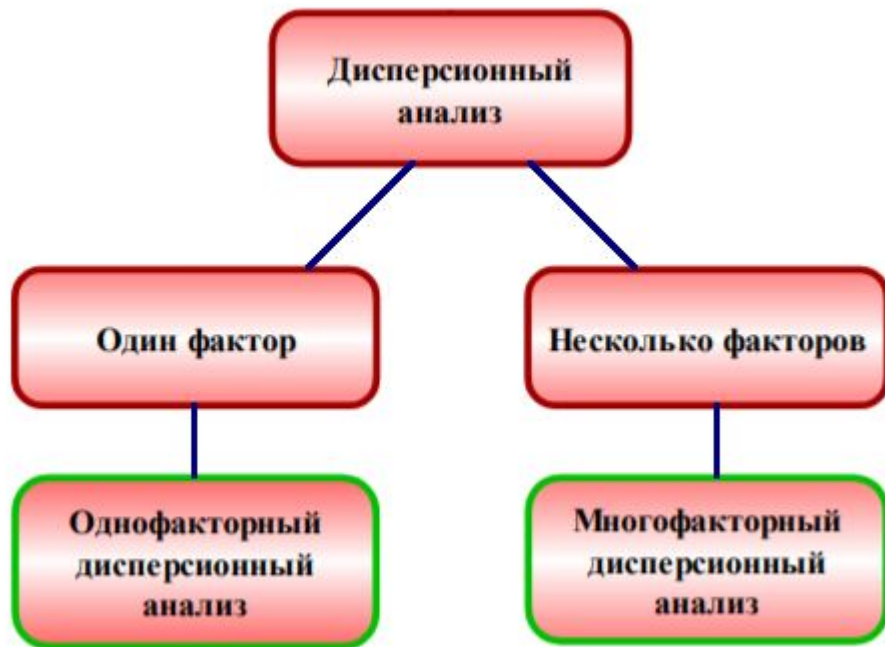
При дисперсионном анализе определяют удельный вес суммарного воздействия одного или нескольких факторов. Существенность влияния фактора определяется путём проверки гипотез:

- $H_0: \mu_1 = \mu_2 = \dots = \mu_a$, где a - число классов градации - все классы градации имеют одно значение.
- H_1 : не все μ_i равны - не все классы градации имеют одно значение средних.

Проверка значимости, разброс данных.

Проверка значимости в дисперсионном анализе основана на сравнении дисперсии, обусловленной межгрупповым разбросом (называемой *средним квадратом эффекта* или $MS_{\text{эффект}}$) и дисперсии, обусловленной внутригрупповым разбросом (называемой *средним квадратом ошибки* или $MS_{\text{ошибка}}$). Если верна нулевая гипотеза (равенство средних в двух популяциях), то можно ожидать сравнительно небольшое различие в выборочных средних из-за случайной изменчивости. Поэтому при нулевой гипотезе внутригрупповая дисперсия будет практически совпадать с общей дисперсией, подсчитанной без учета группой принадлежности. Полученные внутригрупповые дисперсии можно сравнить с помощью F -критерия, проверяющего, действительно ли отношение дисперсий значимо больше 1.

Виды дисперсионного анализа.



Однофакторный дисперсионный анализ

Однофакторный дисперсионный анализ основан на том, что сумму квадратов отклонений статистического комплекса возможно разделить на компоненты:

$SS = SSa + SSe$, где

SS - общая сумма квадратов отклонений,

SSa - объяснённая влиянием фактора a сумма квадратов отклонений,

SSe - необъяснённая сумма квадратов отклонений или сумма квадратов отклонений ошибки.

Однофакторный дисперсионный анализ

Чтобы провести однофакторный дисперсионный анализ данных статистического комплекса, нужно найти фактическое отношение Фишера - отношение дисперсии, объяснённой влиянием фактора (межгрупповой), и необъяснённой дисперсии (внутригрупповой):

$$F = \frac{MS_a}{MS_e}$$

и сравнить его с критическим значением Фишера

$$F_{\alpha, \nu_a, \nu_e}$$

Дисперсии рассчитываются следующим образом:

$$MS_a = \frac{SS_a}{a - 1} \text{ - } \underline{\text{объяснённая дисперсия}},$$

$$MS_e = \frac{SS_e}{n - a} \text{ - } \underline{\text{необъяснённая дисперсия}},$$

при этом

$\nu_a = a - 1$ - число степеней свободы объяснённой дисперсии,

$\nu_e = n - a$ - число степеней свободы необъяснённой дисперсии,

$\nu = n - 1$ - общее число степеней свободы.

Двухфакторный дисперсионный анализ

Двухфакторный дисперсионный анализ применяется для того, чтобы проверить возможную зависимость результативного признака от двух факторов - A и B . Тогда a - число градаций фактора A и b - число градаций фактора B . В статистическом комплексе сумма квадратов остатков разделяется на три компоненты:

Двухфакторный дисперсионный анализ

$$SS = SSa + SSb + SSe,$$

$SS = \sum_{i=1}^a \sum_{j=1}^b (X_{ij} - \bar{X})^2$ - общая сумма квадратов отклонений,

$SS_a = b \sum_{i=1}^a (\bar{X}_i - \bar{X})^2$ - объяснённая влиянием фактора A сумма квадратов отклонений,

$SS_b = a \sum_{j=1}^b (\bar{X}_j - \bar{X})^2$ - объяснённая влиянием фактора B сумма квадратов отклонений,

$SS_e = \sum_{i=1}^a \sum_{j=1}^b (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X})^2$ - необъяснённая сумма квадратов отклонений или сумма квадратов отклонений ошибки,

$\bar{X} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b X_{ij}$ - общее среднее наблюдений,

$\bar{X}_i = \frac{1}{b} \sum_{j=1}^b X_{ij}$ - среднее наблюдений в каждой градации фактора A,

$\bar{X}_j = \frac{1}{a} \sum_{i=1}^a X_{ij}$ - среднее число наблюдений в каждой градации фактора B.

Двухфакторный дисперсионный анализ

Дисперсии вычисляются следующим образом:

$$MS_A = \frac{SS_A}{a-1} - \text{дисперсия, } \underline{\text{объяснённая}} \text{ влиянием фактора } A,$$

$$MS_B = \frac{SS_B}{b-1} - \text{дисперсия, } \underline{\text{объяснённая}} \text{ влиянием фактора } B,$$

$$MS_e = \frac{SS_e}{(a-1)(b-1)} - \underline{\text{необъяснённая}} \text{ дисперсия или дисперсия ошибки}$$

$\nu_a = a - 1$ - число степеней свободы дисперсии, объяснённой влиянием фактора A ,

$\nu_b = b - 1$ - число степеней свободы дисперсии, объяснённой влиянием фактора B ,

$\nu_e = (a - 1)(b - 1)$ - число степеней свободы необъяснённой дисперсии или дисперсии ошибки,

$\nu = ab - 1$ - общее число степеней свободы.

Двухфакторный дисперсионный анализ

Если факторы не зависят друг от друга, то для определения существенности факторов выдвигаются две нулевые гипотезы и соответствующие альтернативные гипотезы:

для фактора A :

$$H_0: \mu_{1A} = \mu_{2A} = \dots = \mu_{aA},$$

H_1 : не все μ_{iA} равны;

для фактора B :

$$H_0: \mu_{1B} = \mu_{2B} = \dots = \mu_{aB},$$

H_1 : не все μ_{iB} равны.

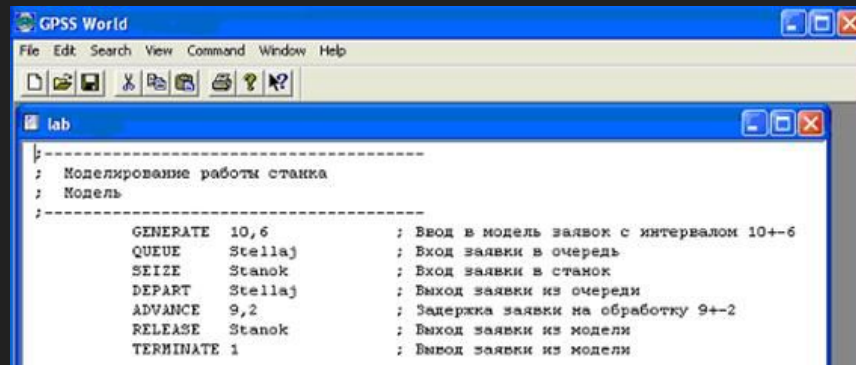
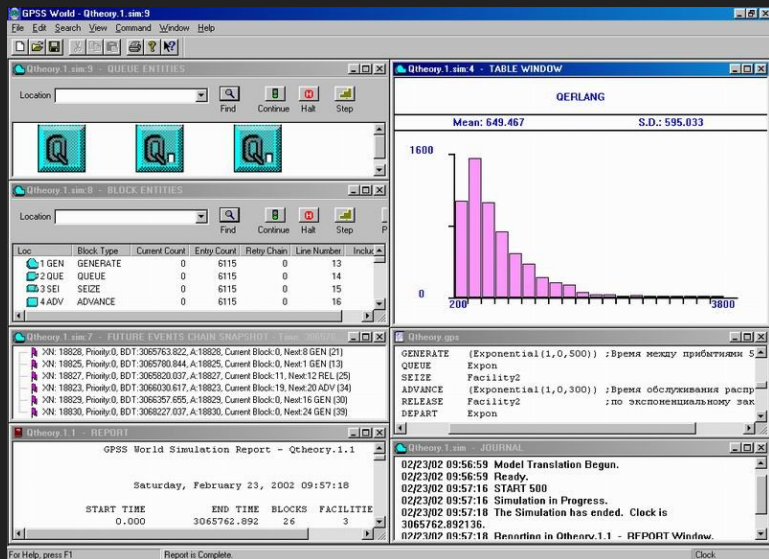
Чтобы определить влияние фактора A , нужно фактическое отношение

Фишера $F_a = \frac{MS_a}{MS_e}$ сравнить с критическим отношением Фишера $F_{\alpha, a, n-k}$.

Чтобы определить влияние фактора B , нужно фактическое отношение

Фишера $F_b = \frac{MS_b}{MS_e}$ сравнить с критическим отношением Фишера $F_{\alpha, b, n-k}$.

Примеры дисперсионного анализа результатов имитационных экспериментов в среде GPSS World



Однофакторный дисперсионный анализ

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.

```
; 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.

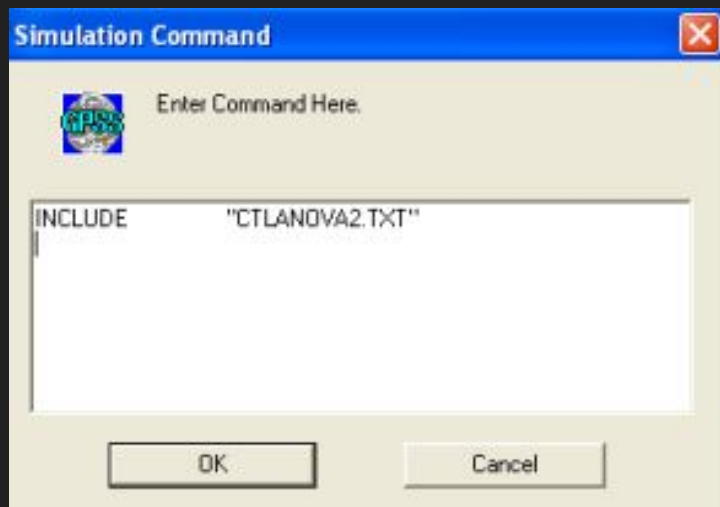
; INCLUDE           "CTLANOVA2.TXT"
TERMINATE
;SHOW ANOVA(Results,2,1)
```

Текст программы ANOVA2.GPS

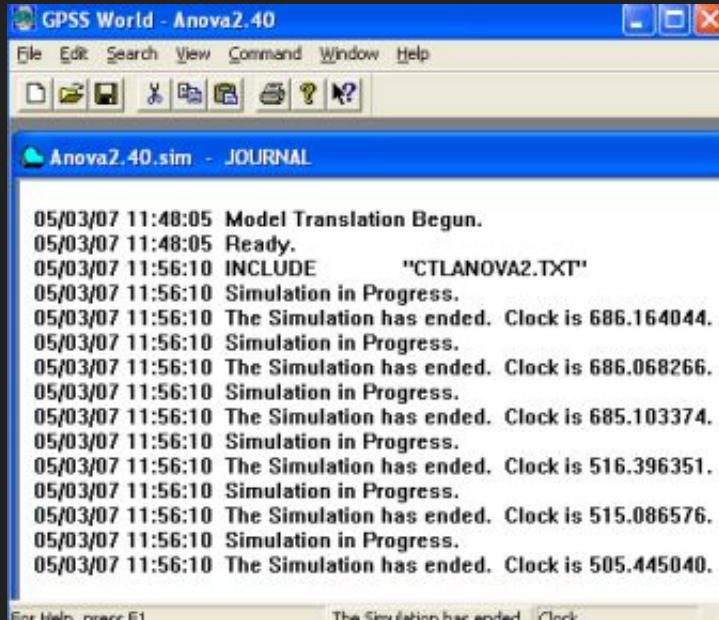
Структура файла "CTLANOVA2.TXT"

1. ; GPSS World Sample File - CTLANOVA2.GPS
2. RESULTS MATRIX ,2,3 ; *Объявление матрицы (2x3)- 2 уровня фактора по 3 ; репликации для каждого*
3. Cut_Time EQU 6.8 ; *Задание среднего времени обл.*
4. Treatment EQU 1
5. RMULT 411
6. Start 100,NP ; *Отчет не создается*
7. MSAVEVALUE RESULTS,1,1,QT\$Barber ; *Запись времени ожидания в матрицу*
8. Clear Off ; *Удаление транзактов, обнуление счетчиков ; Сохраняемые величины и матрицы не обнуляются!*
9. RMULT 421
10. Start 100,NP
11. MSAVEVALUE RESULTS,1,2,QT\$Barber
12. Clear Off
13. RMULT 431
14. Start 100,NP
15. MSAVEVALUE RESULTS,1,3,QT\$Barber
16. Clear Off
17. Cut_Time EQU 5
18. Treatment EQU 2 10
19. RMULT 411
20. Start 100,NP
21. MSAVEVALUE RESULTS,2,1,QT\$Barber
22. Clear Off
23. RMULT 421
24. Start 100,NP
25. MSAVEVALUE RESULTS,2,2,QT\$Barber
26. Clear Off
27. RMULT 431
28. Start 100,NP
29. MSAVEVALUE RESULTS,2,3,QT\$Barber

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 «;»)



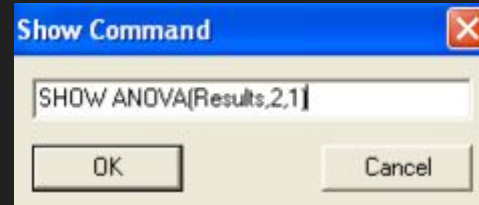
В результате выполнения этой команды в журнале появится следующая запись:



The screenshot shows the 'Anova2.40.sim - JOURNAL' window in GPSS World. The journal contains the following text:

```
05/03/07 11:48:05 Model Translation Begun.
05/03/07 11:48:05 Ready.
05/03/07 11:56:10 INCLUDE      "CTLANOVA2.TXT"
05/03/07 11:56:10 Simulation in Progress.
05/03/07 11:56:10 The Simulation has ended. Clock is 686.164044.
05/03/07 11:56:10 Simulation in Progress.
05/03/07 11:56:10 The Simulation has ended. Clock is 686.068266.
05/03/07 11:56:10 Simulation in Progress.
05/03/07 11:56:10 The Simulation has ended. Clock is 685.103374.
05/03/07 11:56:10 Simulation in Progress.
05/03/07 11:56:10 The Simulation has ended. Clock is 516.396351.
05/03/07 11:56:10 Simulation in Progress.
05/03/07 11:56:10 The Simulation has ended. Clock is 515.086576.
05/03/07 11:56:10 Simulation in Progress.
05/03/07 11:56:10 The Simulation has ended. Clock is 505.445040.
```

Выполняется команда **SHOW ANOVA(Results,2,1)**



Интерпретация полученных данных

Таблица содержит интерпретацию результатов дисперсионного анализа влияния фактора А. Полученное значение статистики Фишера существенно превышает критическое значение для уровня значимости 5%: $F > (F = 7,71)$.

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	$MS_R = 15,053$	$F > (F_{0,05} = 7,71)$

Двухфакторный дисперсионный анализ

- * 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;

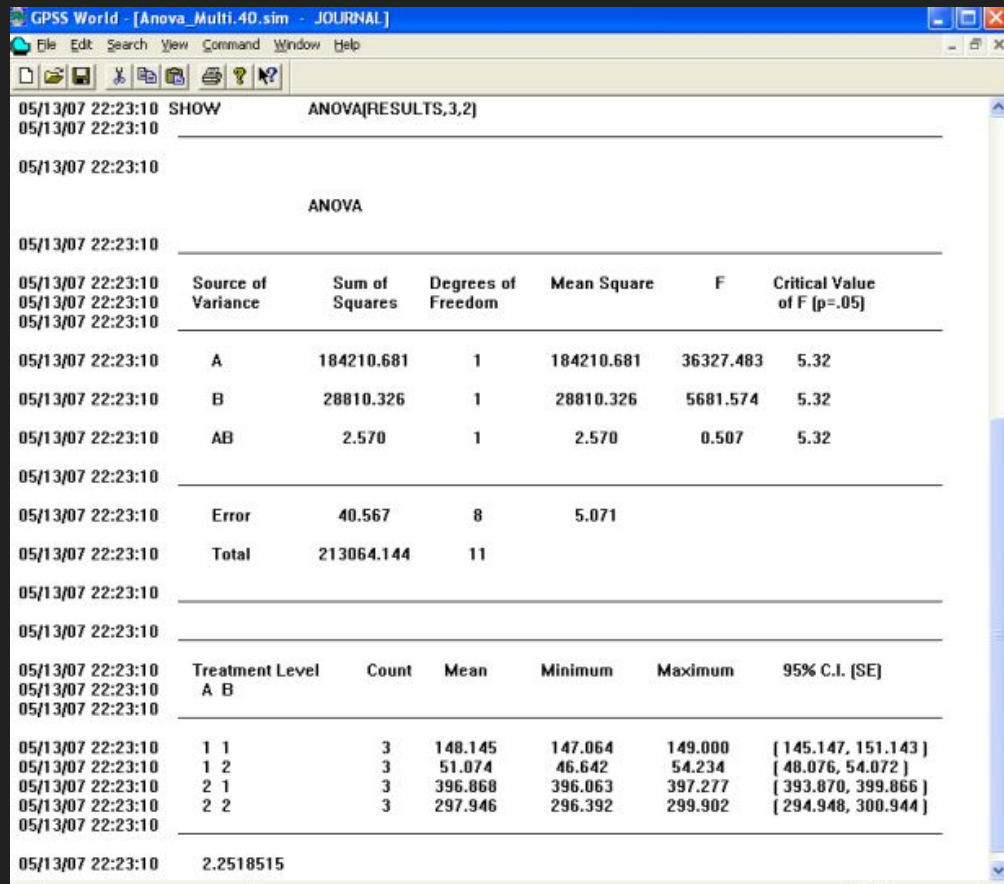
END;
```

Текст программы Anova_Multi.gps, взаимодействующей с текстовым файлом CTLANOVA_MULTI.TXT.

Файл CTLANOVA_MULTI.TXT

- ; GPSS World Sample File - CTLANOVA_MULTI.GPS RESULTS MATRIX
- ,2,2,3 ;
 - Объявление матрицы (2x2x3)- 2 уровня фактора А, 14 ;
 - 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 15
- 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

Результат двухфакторного дисперсионного анализа в GPSS



05/13/07 22:23:10 SHOW ANOVA[RESULTS,3,2]
05/13/07 22:23:10

05/13/07 22:23:10

ANOVA

05/13/07 22:23:10

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F	Critical Value of F (p=.05)
A	184210.681	1	184210.681	36327.483	5.32
B	28810.326	1	28810.326	5681.574	5.32
AB	2.570	1	2.570	0.507	5.32
Error	40.567	8	5.071		
Total	213064.144	11			

05/13/07 22:23:10

05/13/07 22:23:10

Treatment Level	Count	Mean	Minimum	Maximum	95% C.I. (SE)
A B					
1 1	3	148.145	147.064	149.000	[145.147, 151.143]
1 2	3	51.074	46.642	54.234	[48.076, 54.072]
2 1	3	396.868	396.063	397.277	[393.870, 399.866]
2 2	3	297.946	296.392	299.902	[294.948, 300.944]

05/13/07 22:23:10 2.2518515

Спасибо за внимание!