

Однофакторный дисперсионный анализ с повторными измерениями в пакете Statistica: 2 способа

на примере данных grergi: <http://forum.disser.ru/index.php?showtopic=3481&st=0>

Нохрин Д.Ю. (nokh для читателей forum.disser.ru)

В интактной группе животных определяется уровень гормонов (время t0), затем на протяжении 10 дней им вводится вещество А, после чего на 11 день повторно измеряется уровень гормонов (время t1). Спустя пять дней отмены – последнее измерение (время t2). Поскольку фактор здесь один – «время» - это однофакторный анализ с повторными измерениями одних и тех же субъектов.

Субъект	t0	t1	t2
1	10,78	4,28	9,15
2	10,73	3,45	5,08
3	9,61	6,33	8,12
4	10,66	4,35	7,79
5	9,63	7,83	7,27
6	8,74	6,43	8,19
7	8,08	5,48	4,46
8	7,91	3,49	4,72

Наиболее важные моменты выделены **жёлтым**.

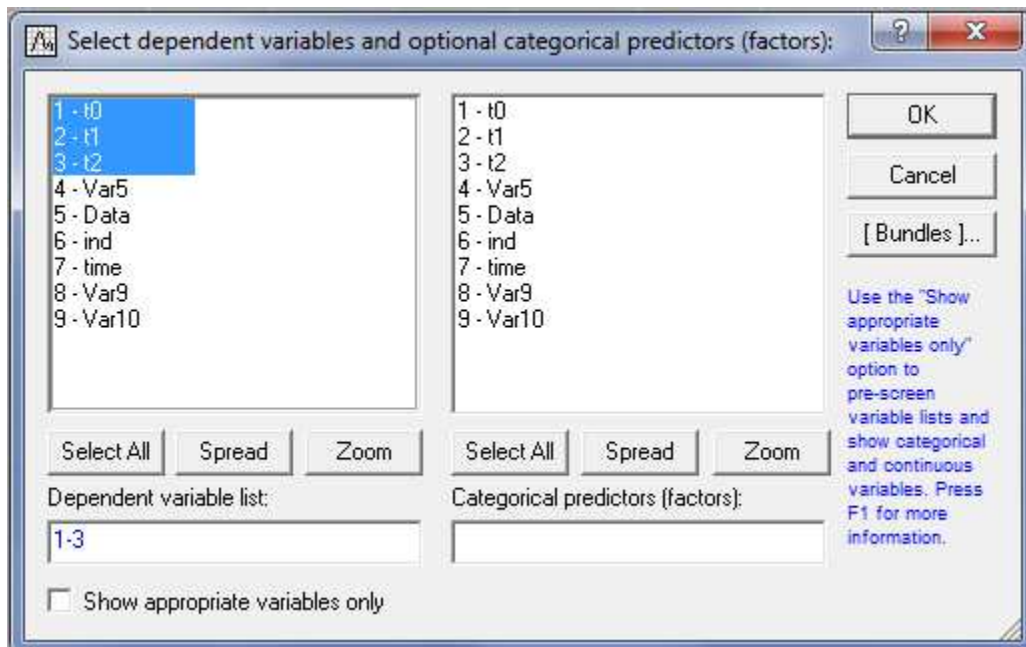
Способ 1. Repeated measures ANOVA

1). Данные заносятся как в выделенной области. В каждой строке – конкретный индивид.

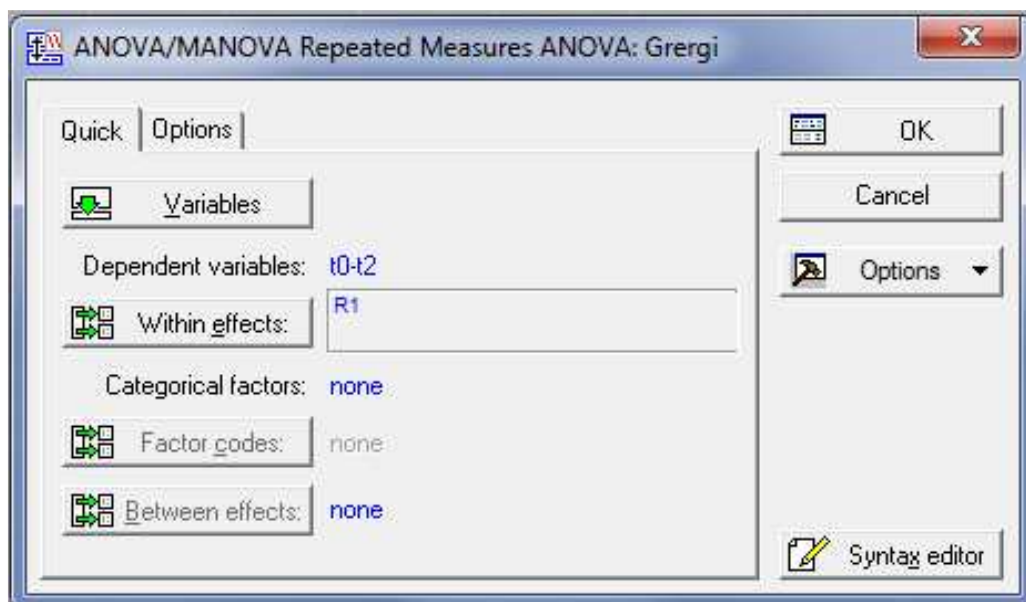
STATISTICA - [Data: Grergi (9v by 24c)]										
File Edit View Insert Format Statistics Data Mining Graphs Tools Data Window Help										
Add to Workbook Add to Report Add to MS Word										
Arial 10 B I U										
	1	2	3	4	5	6	7	8	9	
	t0	t1	t2	Var5	Data	ind	time	Var9	Var10	
1	10.78	4.28	9.15		10.78	1	1			
2	10.73	3.45	5.08		10.73	2	1			
3	9.61	6.33	8.12		9.61	3	1			
4	10.66	4.35	7.79		10.66	4	1			
5	9.63	7.83	7.27		9.63	5	1			
6	8.74	6.43	8.19		8.74	6	1			
7	8.08	5.48	4.46		8.08	7	1			
8	7.91	3.49	4.72		7.91	8	1			
9					4.28	1	2			
10					3.45	2	2			
11					6.33	3	2			
12					4.35	4	2			
13					7.83	5	2			
14					6.43	6	2			
15					5.48	7	2			
16					3.49	8	2			
17					9.15	1	3			
18					5.08	2	3			
19					8.12	3	3			
20					7.79	4	3			
21					7.27	5	3			
22					8.19	6	3			
23					4.46	7	3			
24					4.72	8	3			

2). Путь: Statistics – ANOVA – Repeated measures ANOVA

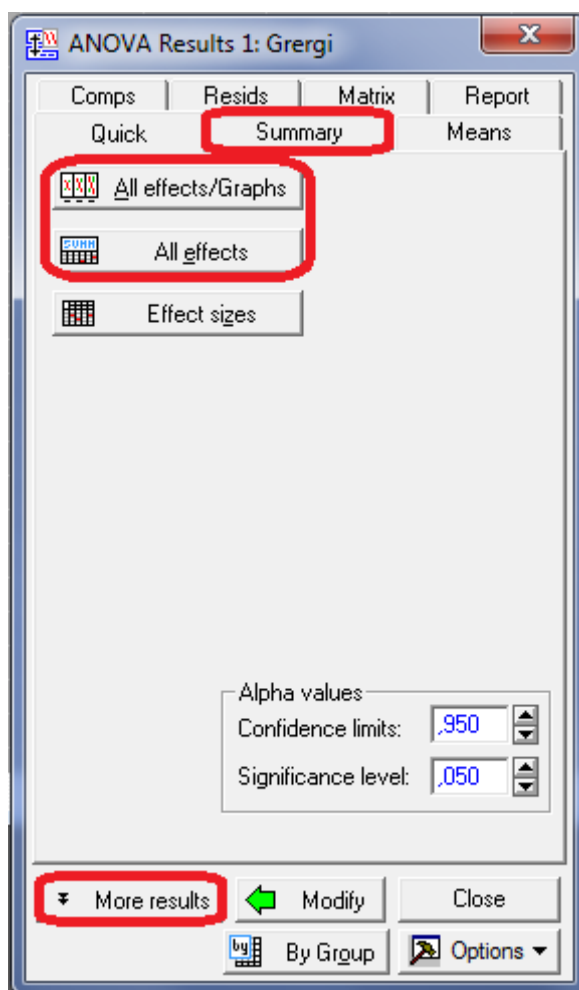
3). В «Variables» задаётся список зависимых переменных. Если бы групп было несколько, например контрольная и экспериментальная, соответствующую метку нужно было бы выбрать в качестве фактора.



4). Нужно просто зайти в «Within effects», чтобы соответствующая модель была задана. Напротив появится «R1». OK



5). В появившейся форме последовательно смотрим выделенные красным объекты.

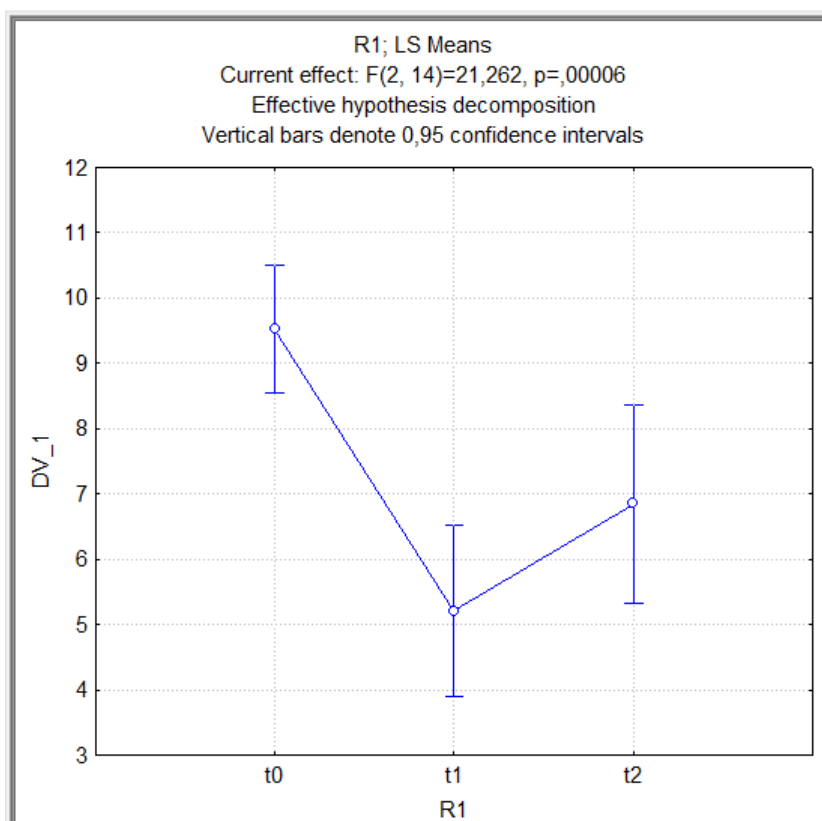


5.1. В «All effects» - таблицу результатов дисперсионного анализа

Repeated Measures Analysis of Variance (Gergi)					
Sigma-restricted parameterization					
Effective hypothesis decomposition					
Effect	SS	Degr. of Freedom	MS	F	p
Intercept	1240,706	1	1240,706	345,0359	0,000000
Error	25,171	7	3,596		
R1	75,798	2	37,899	21,2617	0,000057
Error	24,955	14	1,783		

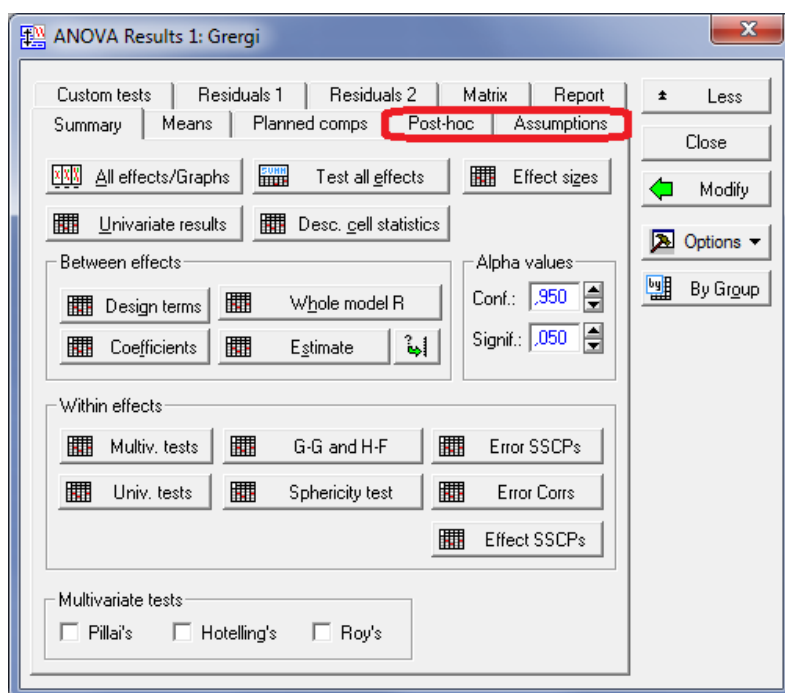
Эффект различий уровня гормонов на разных сроках высоко статистически значим: $F_{(2; 14)}=21,26; p<0,001$.

5.2. В «All effects/Graphs» - картинку



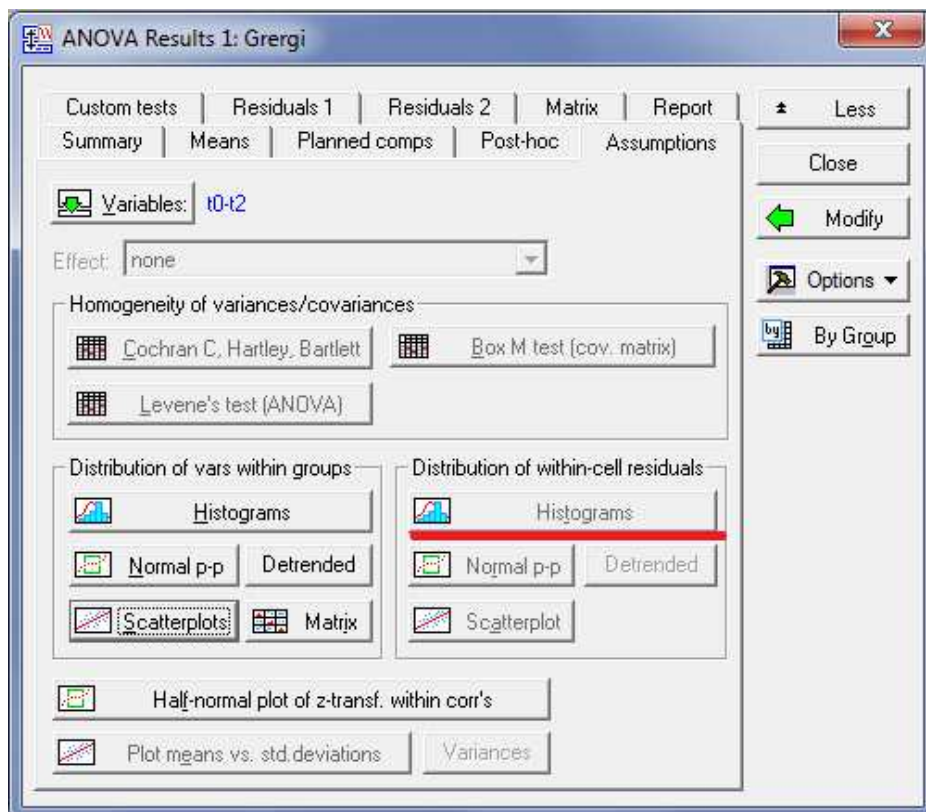
5.3. В «Summary – Cell statistics» - средние и 95%-ные доверительные интервалы (95% ДИ) для оформления таблиц. Для зависимых выборок неплохо было бы рассчитать также среднюю разность с соответствующими ДИ.

5.4. Заходим в «More results». Смотрим выделенные закладки.

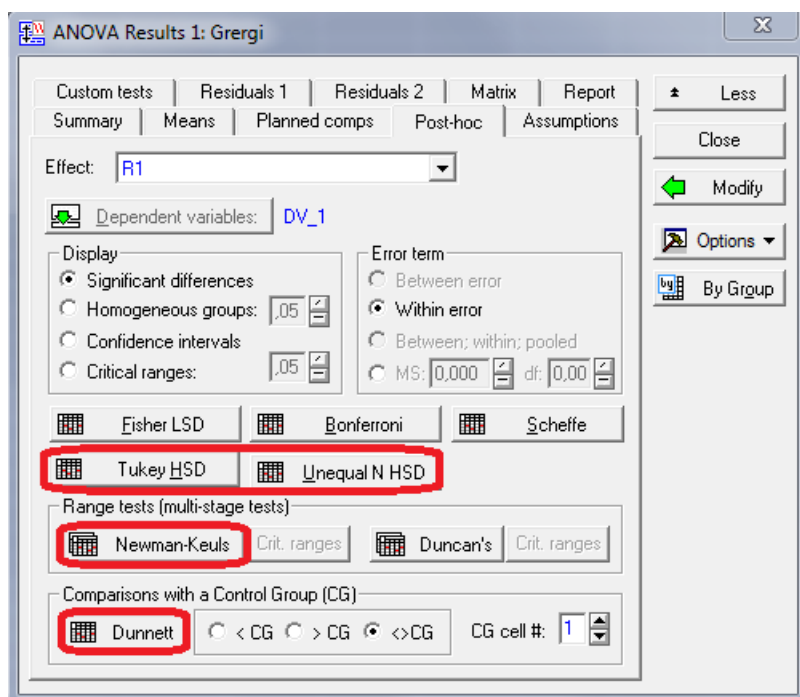


6). Проверяем требования модели ДА – закладка «Assumptions»

! В этом варианте анализа проверять нечего, всё недоступно:



7). Закладка «Post-hoc» - Апостериорные сравнения



Наиболее строгий, классический, но несколько консервативный – метод Тьюки, мощный, но менее консервативный – метод Ньюмена – Кейлса, метод Даннетта – только для сравнения с контролем (например, временем t0).

Newman-Keuls test; variable DV_1 (Gergi) Approximate Probabilities for Post Hoc Tests Error: Within MS = 1,7825, df = 14,000				
Cell No.	R1	{1}	{2}	{3}
1	t0	9,5175	5,2050	6,8475
2	t1	0,000209	0,000209	0,027625
3	t2	0,001461	0,027625	

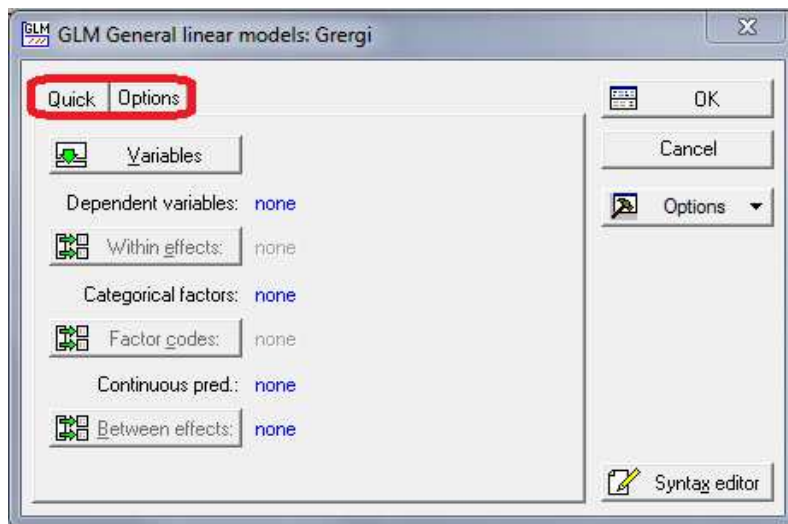
Все сроки высоко статистически значимо отличаются друг от друга.

Способ 2. General Linear Models

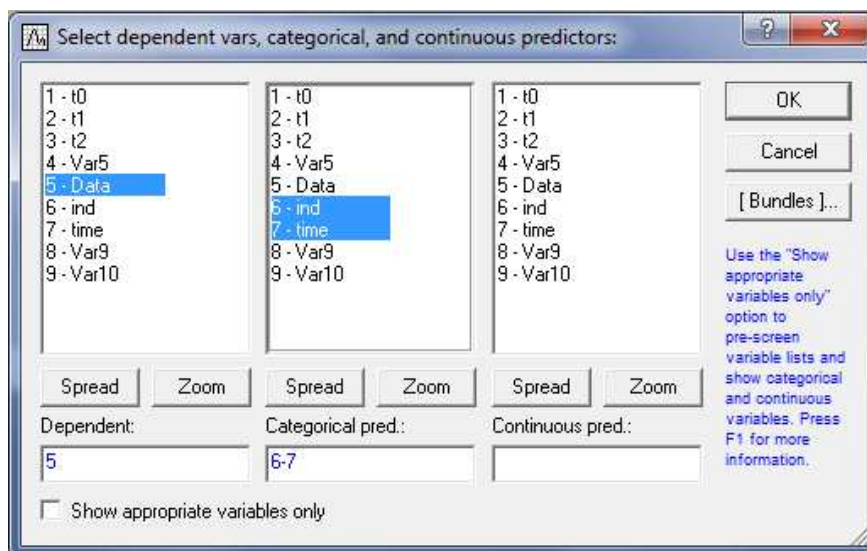
1). Данные заносятся как в выделенной области. Индивидуальные метки животных дублируются в колонке ind.

STATISTICA - [Data: Gergi (9v by 24c)]									
File Edit View Insert Format Statistics Data Mining Graphs Tools Data Window Help									
Add to Workbook Add to Report Add to MS Word									
10 B I U									
	1 t0	2 t1	3 t2	4 Var5	5 Data	6 ind	7 time	8 Var9	9 Var10
1	10,78	4,28	9,15		10,78	1	1		
2	10,73	3,45	5,08		10,73	2	1		
3	9,61	6,33	8,12		9,61	3	1		
4	10,66	4,35	7,79		10,66	4	1		
5	9,63	7,83	7,27		9,63	5	1		
6	8,74	6,43	8,19		8,74	6	1		
7	8,08	5,48	4,46		8,08	7	1		
8	7,91	3,49	4,72		7,91	8	1		
9					4,28	1	2		
10					3,45	2	2		
11					6,33	3	2		
12					4,35	4	2		
13					7,83	5	2		
14					6,43	6	2		
15					5,48	7	2		
16					3,49	8	2		
17					9,15	1	3		
18					5,08	2	3		
19					8,12	3	3		
20					7,79	4	3		
21					7,27	5	3		
22					8,19	6	3		
23					4,46	7	3		
24					4,72	8	3		

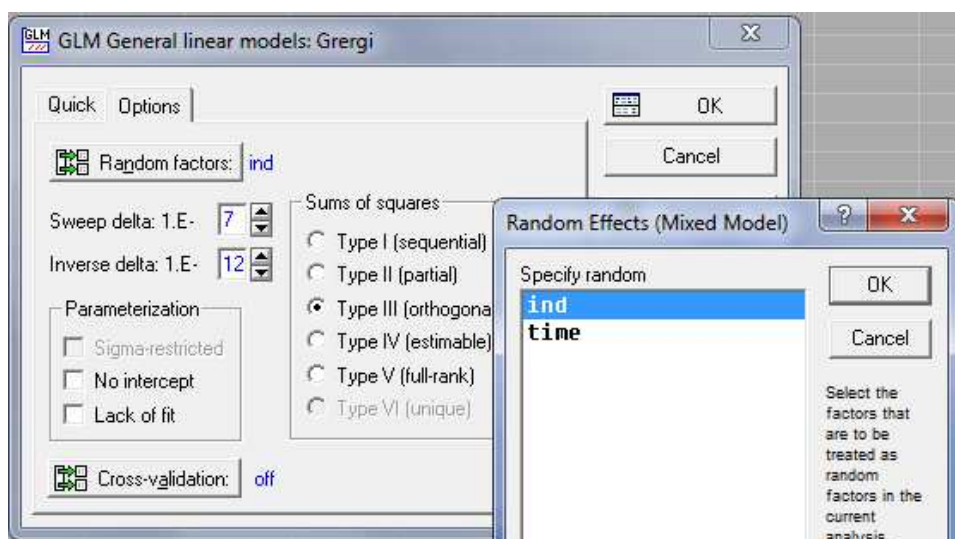
2). Путь: Statistics – Advanced Linear/Nonlinear Models – General Linear Models – General Linear Models.



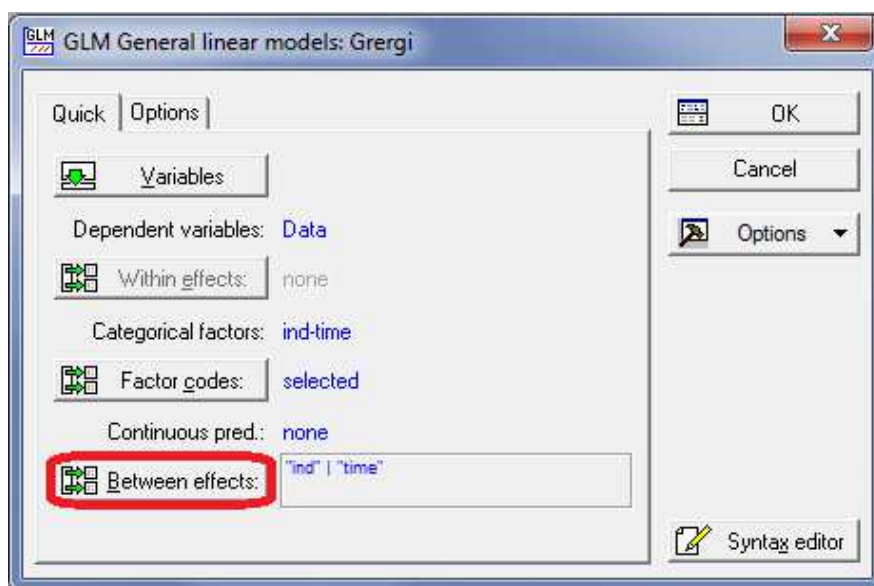
3). В «Variables» задаём зависимую и независимые переменные.



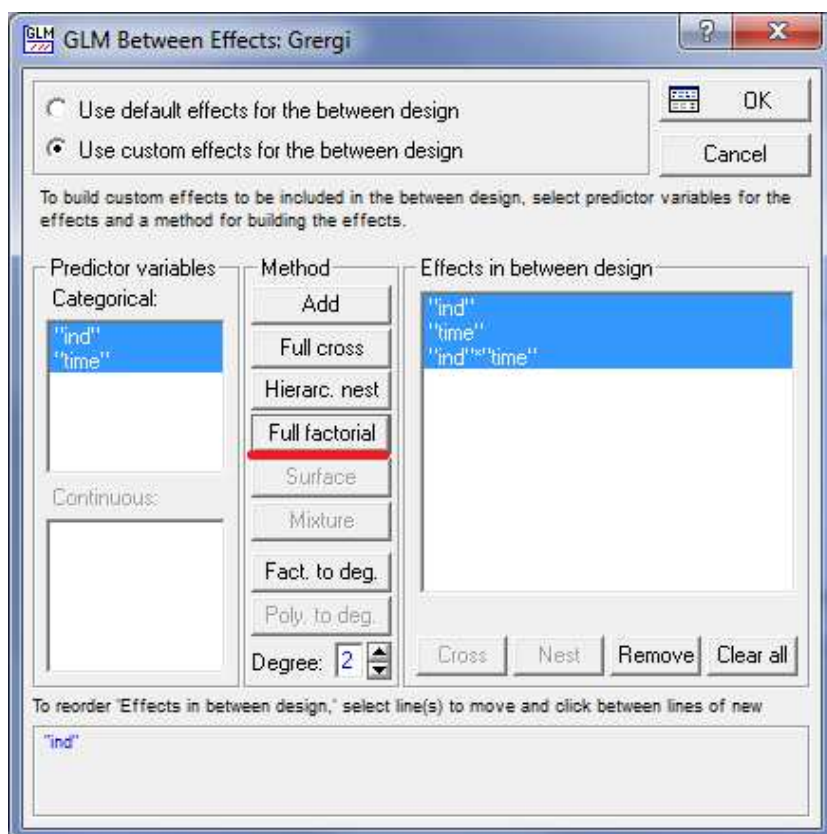
4). На закладке «Options» указываем случайный фактор – ind.



5). В «Factor codes» выбираем везде «All» и переходим в меню «Between effects».

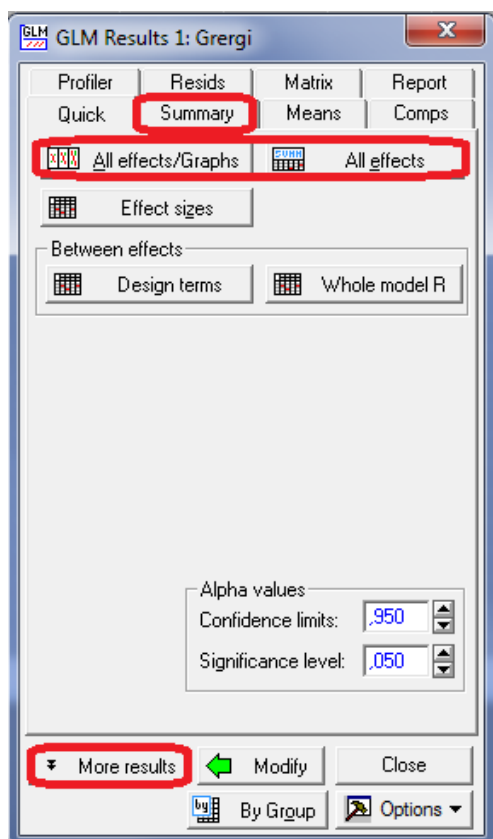


6). В меню «Between effects» задаём модель ДА. **Всегда контролируем: «Use custom...»!** В данном случае всё просто, но в случае сложных дисперсионных комплексов нужно повозиться со вложенными эффектами. Если по сложной модели нужно обчислить сразу несколько показателей – полезно сохранить синтаксис – «Syntax editor»: можно загружать сохранённый и менять в нём названия переменных, пропуская пункты 3-5. ОК.



! Сложные комплексы могут обсчитываться долго, в конце может быть предупреждающее окно, если какие-то эффекты невозможно оценить.

7). В появившейся форме последовательно смотрим выделенные красным объекты.



7.1. В «All effects» - таблицу результатов дисперсионного анализа

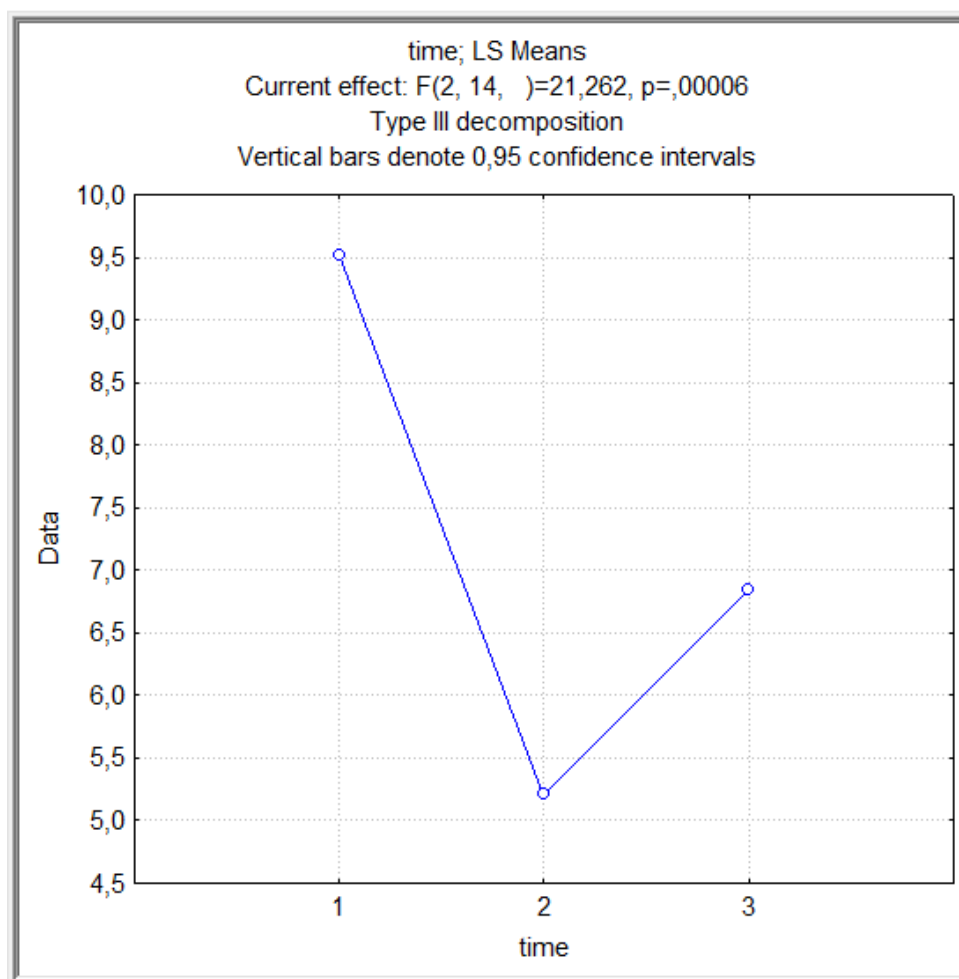
Univariate Tests of Significance for Data (Gergi)								
Over-parameterized model								
Type III decomposition								
Effect	Effect (F/R)	SS	Degr. of Freedom	MS	Den.Syn. Error df	Den.Syn. Error MS	F	p
Intercept	Fixed	1240,706	1	1240,706	7,00000	3,595876	345,0359	0,000000
ind	Random	25,171	7	3,596	14,00000	1,782512	2,0173	0,124947
time	Fixed	75,798	2	37,899	14,00000	1,782512	21,2617	0,000057
ind*time	Random	24,955	14	1,783	0,00000	0,000000		
Error			0					

Эффект различий уровня гормонов на разных сроках высоко статистически значим: $F_{(2; 14)}=21,26; p<0,001$.

! Для интересующего эффекта выписываем вручную средний квадрат и число степеней свободы ошибки: Error MS и Error df. В нашем случае это фактор «time»; выписываем Error MS = 1,782512 и Error df = 14. Видно, что его значимость оценивается относительно взаимодействия факторов «ind x time», характеризующего индивидуальную динамику показателя во времени.

Следует обратить внимание, что данный дисперсионный комплекс имеет нулевую ошибку. Она нулевая – потому что наименьшая ячейка комплекса содержит только одно наблюдение. **Это не является ошибкой**, но не позволяет оценить статистическую значимость взаимодействия «ind x time», что может «не понравится» программе и о чём она сообщит в предупреждающем окне. В данном случае у меня всё прошло гладко, но в более сложных комплексах предупреждающее окно появляется. Чтобы эта ошибка не была нулевой нужно, чтобы в каждой временной точке индивид измерялся дважды, т.е. в нашем примере – чтобы уровень гормонов определялся в двух параллельных пробах. Но нужно это лишь для того, чтобы выразить ошибку и оценить взаимодействие «ind x time», **для ответа на вопрос исследования этого не нужно**, и даже более – параллельные пробы в 2 раза увеличат расходы.

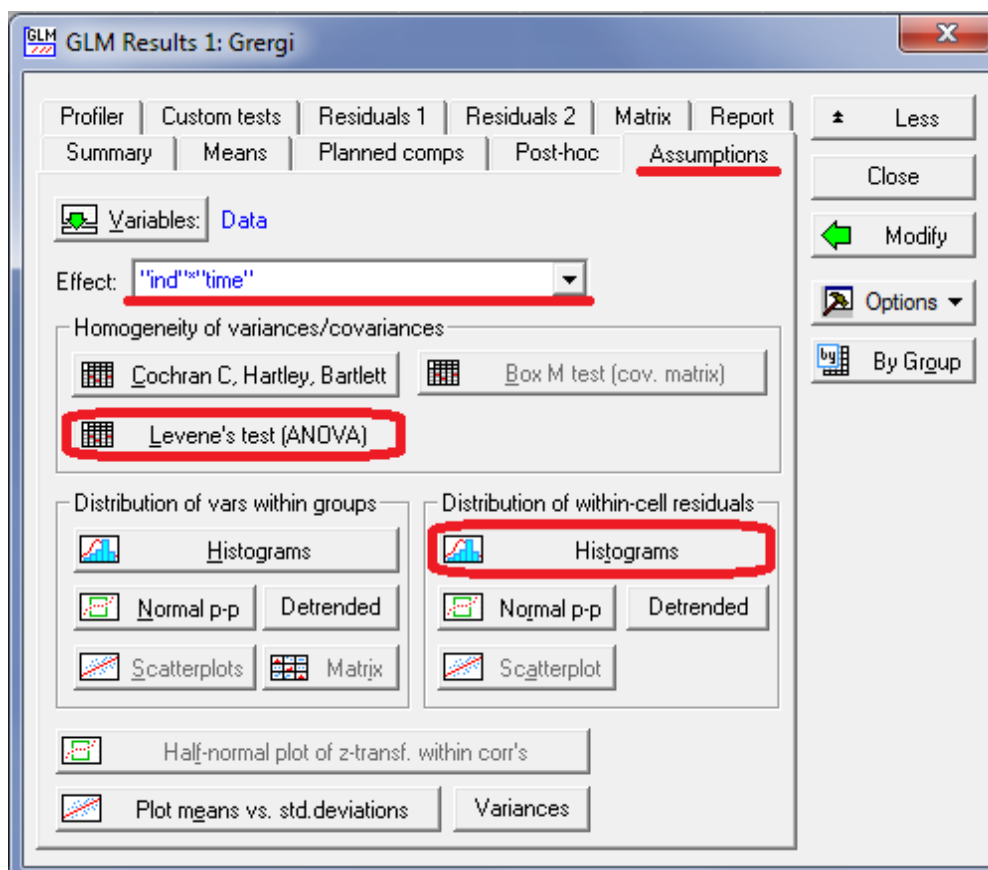
7.2. В «All effects/Graphs» смотрим картинку



Картинка без 95% ДИ нас не устраивает! Построим позже.

7.3. В «Summary – Cell statistics» - средние и 95%-ные доверительные интервалы (95% ДИ) для оформления таблиц. Для зависимых выборок неплохо было бы рассчитать также среднюю разность с соответствующими ДИ.

7.4. Заходим в «More results». Смотрим полезные закладки.



8. Проверяем требования модели ДА – закладка «Assumptions»

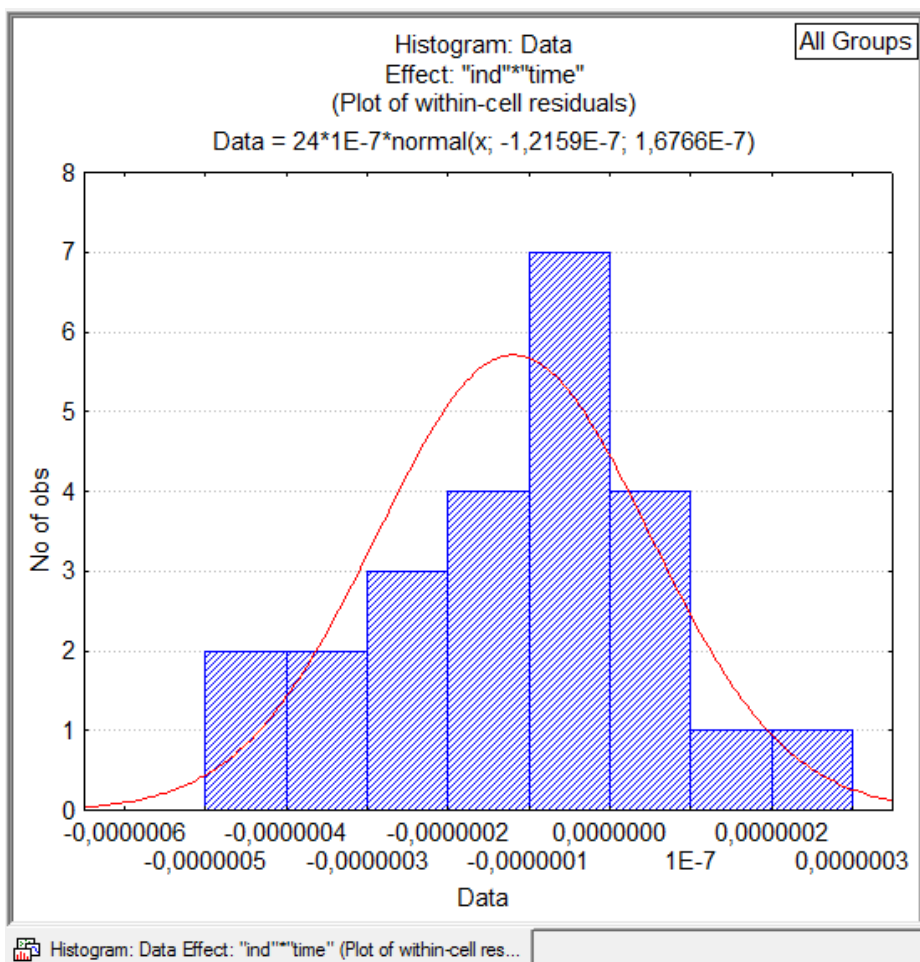
Сначала в окне «Effect» выбираем нужный нам эффект. Нужный эффект это тот, который является дисперсией ошибки комплекса. В нашем случае такой ошибкой является взаимодействие «ind x time»; выше мы выписывали MS и df для этого же эффекта. Выставляем «ind x time».

! В данном способе задания дисперсионного комплекса проверки требований доступны!

Критерий Левне говорит об однородности дисперсий:

Levene's Test for Homogeneity of Variances (Gergi)				
Effect: "time"				
Degrees of freedom for all F's: 2, 21				
	MS Effect	MS Error	F	p
Data	0,762951	0,443066	1,721979	0,203003

Гистограмма показывает унимодальность распределения ошибки. Визуально всё пусть и не идеально, но неплохо:



9). Закладка «Post-hoc» - Апостериорные сравнения

GLM Results 1: Greggi

Profiler | Custom tests | Residuals 1 | Residuals 2 | Matrix | Report

Summary | Means | Planned comps | **Post-hoc** | Assumptions

Effect: "time"

Dependent variables: Data

Display

- ☒ Significant differences
- ☐ Homogeneous groups: .05
- ☐ Confidence intervals
- ☐ Critical ranges: .05

Error term

- ☐ Between error
- ☐ Within error
- ☐ Between: within: pooled
- ☒ MS: 1,7825 df: 14

Fisher LSD | Bonferroni | Scheffe

Tukey HSD | **Unequal N HSD**

Range tests (multi-stage tests)

Newman-Keuls | Crit. ranges | Duncan's | Crit. ranges

Comparisons with a Control Group (CG)

Dunnnett | ☐ < CG | ☐ > CG | ☒ <> CG | CG cell #: 1

9.1. Выставляем нужный эффект для сравнения – временные точки «time»

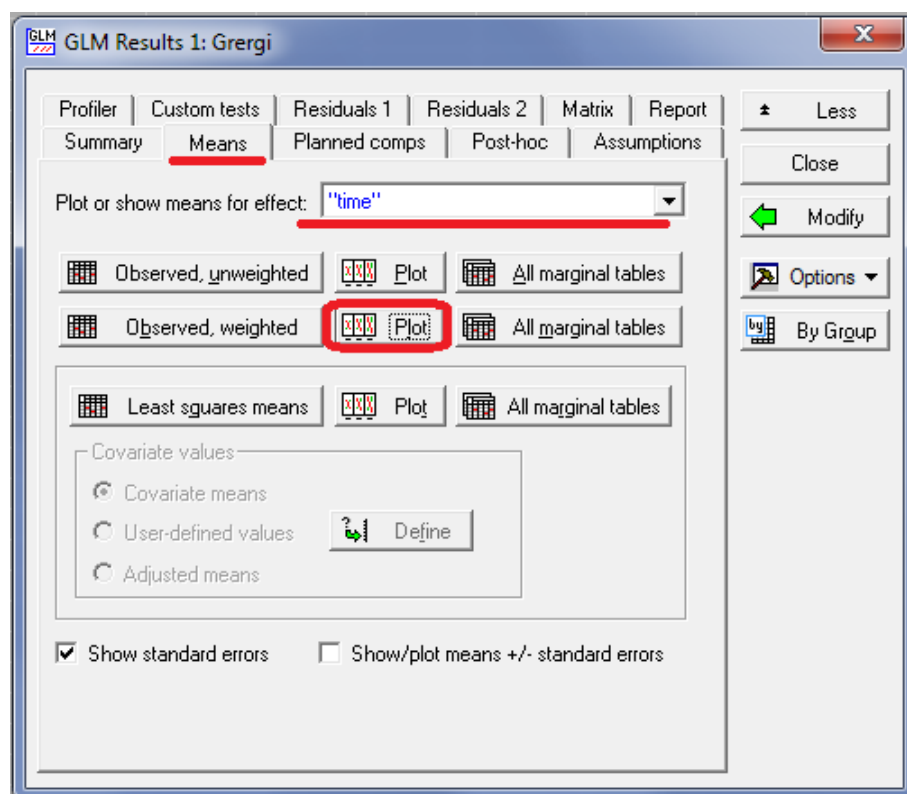
9.2. Активируем радиокнопку в Error term и **вручную** вводим выписанные ранее значения среднего квадрата и числа степеней свободы ошибки эффекта «time»: Error MS = 1,782512 и Error df = 14.

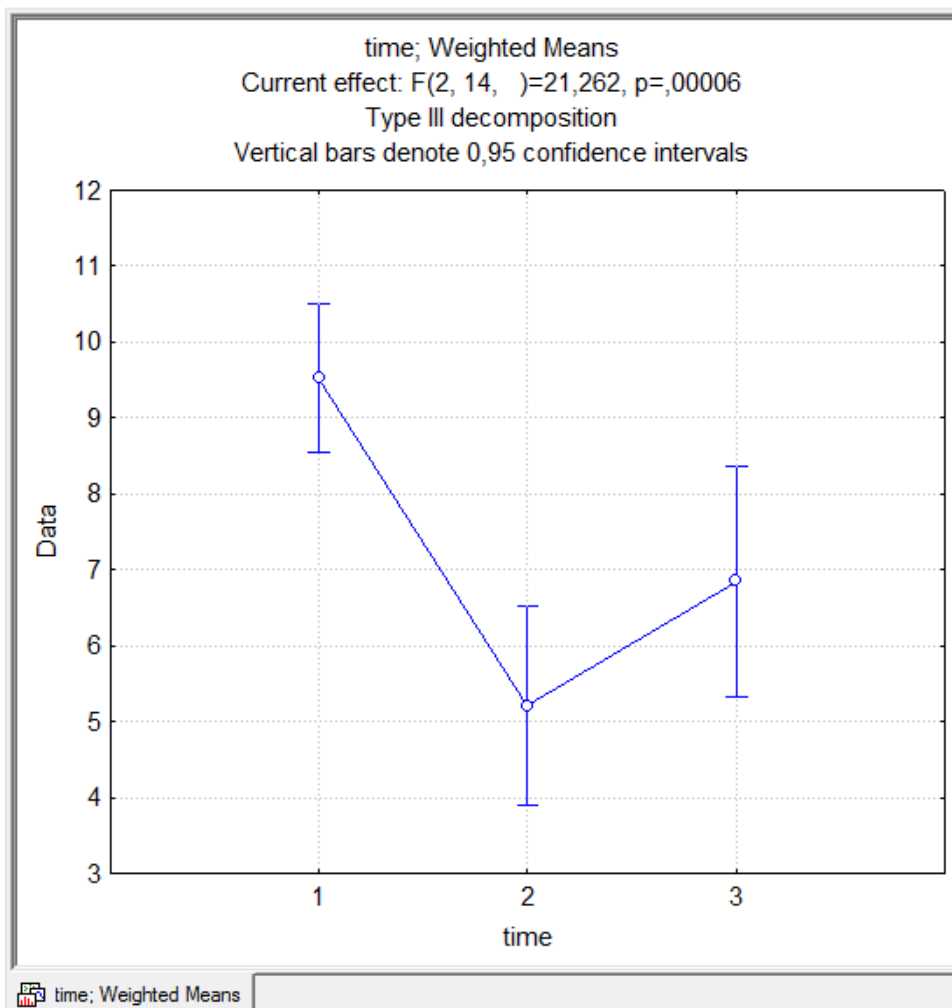
9.3. Проводим интересующие сравнения. Например, методом Ньюмена – Кейлса:

Newman-Keuls test; variable Data (Greggi) Approximate Probabilities for Post Hoc Tests Error: User-defined MS = 1,7825, df = 14,000					
Cell No.	time	{1}	{2}	{3}	
		9,5175	5,2050	6,8475	
1	1		0,000209	0,001461	
2	2	0,000209		0,027625	
3	3	0,001461	0,027625		

Все сроки высоко статистически значимо отличаются друг от друга.

10. Строим график с 95% ДИ:





Выводы:

- 1). Оба способа позволили получить одни и те же:
 - значения F -критерия и p для интересующего эффекта;
 - оценки значимости попарных различий в апостериорных сравнениях;
 - графики средних с 95%-ными ДИ.
- 2). Способ 2 сложнее задаётся, однако позволяет:
 - осмысленно выбирать оцениваемые эффекты на разных стадиях анализа, что способствует более глубокому пониманию принципов дисперсионного анализа;
 - понять, что выступает ошибкой в данном дисперсионном комплексе, а также оценить фактор индивидуальной изменчивости («ind»). В данном случае такой задачи не стояло, но в других может оказаться полезным сравнить изменчивость, привносимую какими-либо воздействиями и индивидуальную изменчивость показателя. Это делается в ходе расчёта и сравнения компонентов дисперсии и для такой задачи фактор «ind», должен быть выражен именно в явном виде.
 - провести проверку выполнения требований дисперсионного анализа.