Снежана Гочева, snow@uni-plovdiv.bg

Учебен пример към темата „Факторен и регресионен анализ” с използване на пакета SPSS

Изследване на ефективността на лазер с пари на меден бромид

Този пример е демонстрационен. Показани са основните стъпки за провеждане на факторен и регресионен анализ за решаване на задачи с помощта на софтуерния пакет SPSS. Препоръчва се студентите да повторят анализите самостоятелно.

Обект на изследване

Използвани са реални данни за 7 типа лазери с пари на меден бромид (CuBr лазери). Устройствата са български патенти, разработени в Института по физика на твърдото тяло към БАН, София.

CuBr лазерите са най-мощните източници във видимата зона. Те излъчват едновременно на две дължини на вълната: 510.6 *nm* и 578.6 *nm*. Имат голямо приложение в медицината, в нанотехнологиите, в развлекателната индустрия, за изследване на атмосферата и океана, и др.

Данни

Изследва се извадка от измерванията от 7 експеримента.

Данните са за 6 основни лазерни величини: D – вътрешен диаметър на лазерната тръба, dr – вътрешен диаметър на пръстените в тръбата, L – разстояние между електродите, Pin – средна входна мощност, P_L – средна входна мощност на единица дължина и EFF - лазерна ефективност (КПД). Измерените стойности са дадени в Табл. 1.

Табл. 1. Експериментални данни за CuBr лазер.

k	Dr, mm	L, cm	Pin, KW	D, mm	PL, KW/cm	EFF
1	4.5	30	1.0	15	1.7	0.7
2	20	50	1.2	40	1.2	1.6
3	30	100	2.5	50	1.2	2.1
4	30	140	2.3	50	0.8	2.0
5	40	50	1.2	40	1.2	2.0
6	40	120	2.7	40	1.1	2.2
7	58	200	3.3	58	0.8	3.0

Приложение на факторен анализ (ФА)

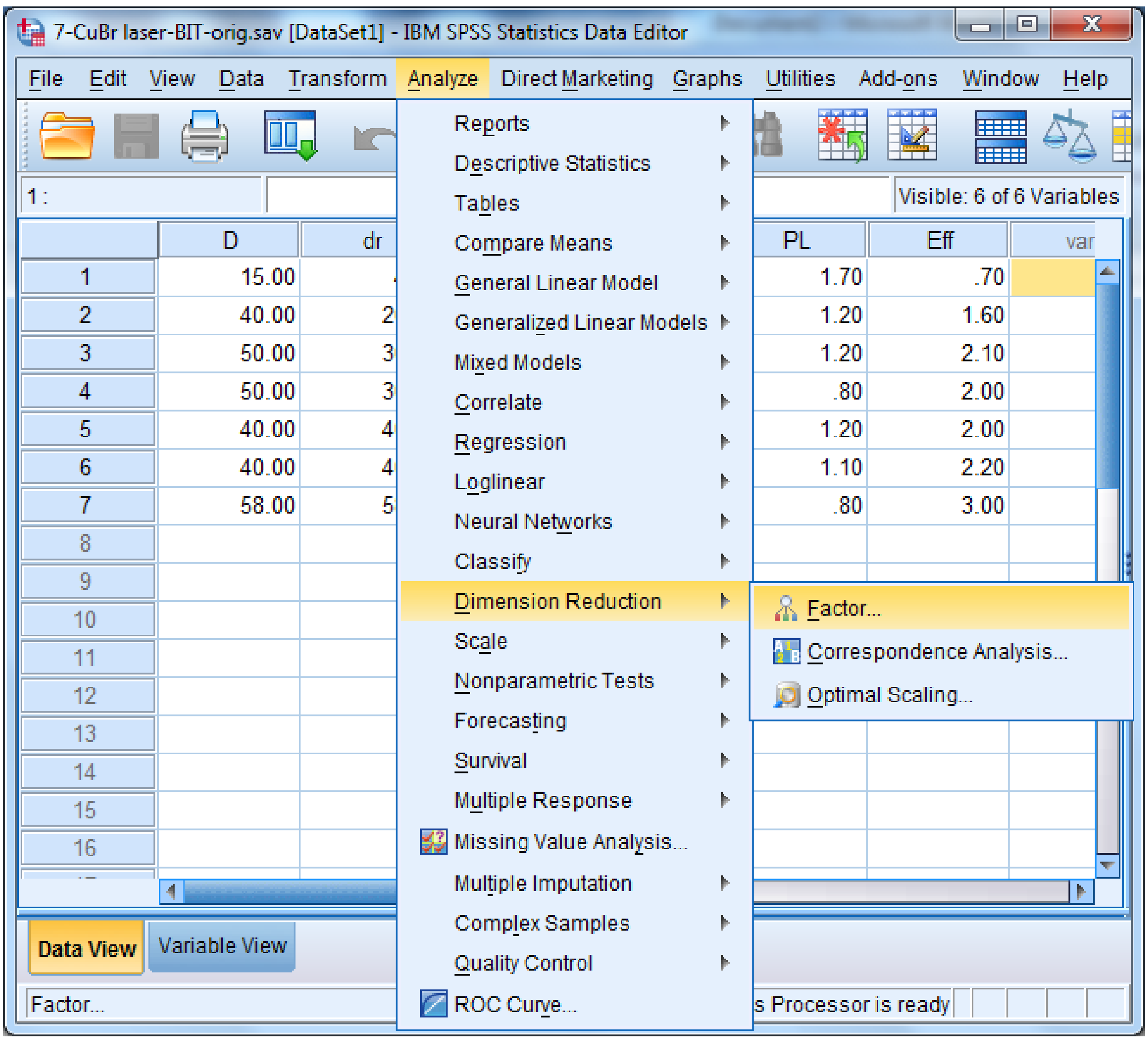
Данните трябва да са от интервален или относителен тип и в общия случай броят наблюдения (експерименти) се препоръчва да е поне 5-10 пъти по-голям от броя на величините.

1) ***<u>Изследване условията за прилагане и адекватност на ФА</u>***

   1.1. Стартираме SPSS и набираме или отваряме файла с данни.

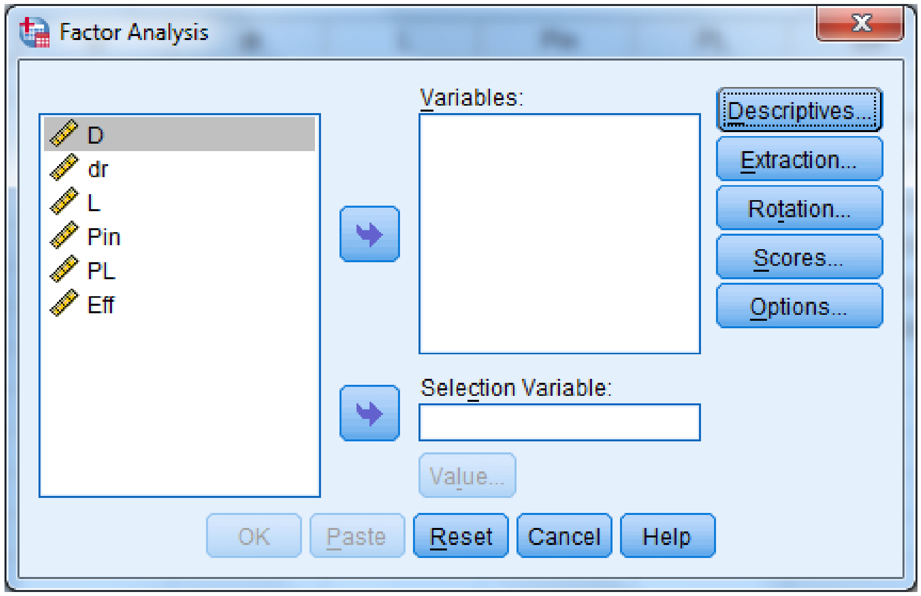
   1.2. От основното меню избираме процедурата за ФА (Фиг. 1):

Analyze/Dimension Reduction/Factor …



Фиг. 1. Стартиране на процедурата за факторен анализ в SPSS.

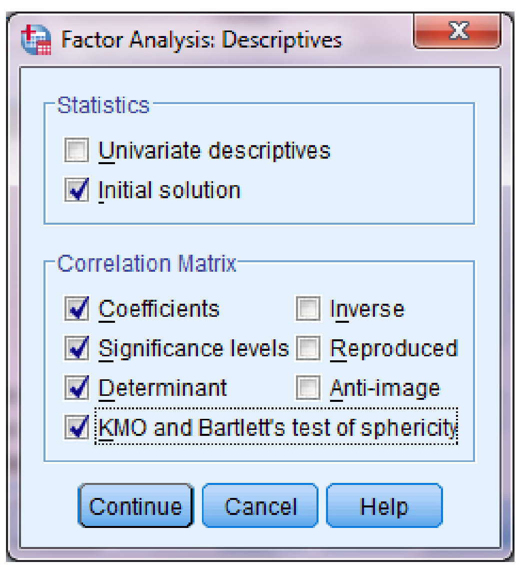
Отваря се прозорецът, показан на Фиг. 2.



Фиг. 2. Изглед на прозореца за ФА.

Най-напред пренасяме всички променливи, вкл. Eff отдясно в панела ***Variables***, като използваме стрелката.

След това отваряме ***Descriptives*** и избираме настройките от Фиг. 3 за да изследваме корелационната матрица.



Фиг. 3. Начални настройки на описателните статистики във ФА.

Потвърждаваме с ***Continue*** и ***OK***. Получаваме резултата в отделен прозорец.

1.3. Изследване на корелационната матрица на всички променливи

В нашия случай корелационната матрица с коефициентите на Пиърсън за всяка двойка величини има вида (Фиг.4):

Correlation Matrix[a]

		D	dr	L	Pin	PL	Eff
Correlation	D	1.000	.806	.800	.763	-.917	.918
	dr	.806	1.000	.785	.757	-.785	.967
	L	.800	.785	1.000	.946	-.847	.855
	Pin	.763	.757	.946	1.000	-.733	.840
	PL	-.917	-.785	-.847	-.733	1.000	-.863
	Eff	.918	.967	.855	.840	-.863	1.000
Sig. (1-tailed)	D		.014	.015	.023	.002	.002
	dr	.014		.018	.025	.018	.000
	L	.015	.018		.001	.008	.007
	Pin	.023	.025	.001		.030	.009
	PL	.002	.018	.008	.030		.006
	Eff	.002	.000	.007	.009	.006	

a. Determinant = 2.10E-006

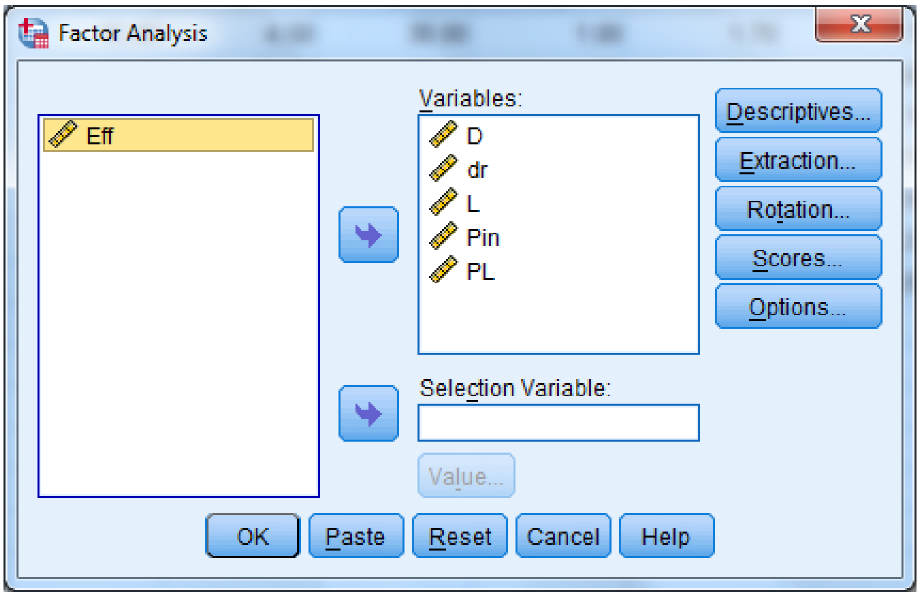
Фиг. 4. Корелационна матрица на всички променливи (в горната половина) и техните статистически значимости (в горната половина на таблицата).

ФА може да се проведе, когато има големи корелационни зависимости (с коефициенти по модул над 0.7). Ако една величина не показва големи корелационни коефициенти с останалите променливи, тя се изключва от ФА. Може да се включи по-нататък в регресионен анализ. А в случая, когато не корелира и със зависимата променлива, такава променлива се отстранява от по-нататъшни анализи и по този начин се намалява размерността на задачата. Остават само променливи, които имат съществено влияние от корелационен тип както върху зависимата променлива, така и с други променливи.

От Фиг. 4 се вижда, че всички независими променливи: D, dr, L, Pin и PL имат големи по абсолютна стойност корелационни коефициенти с ефективността Eff (означени на син фон). Също така те корелират и помежду си. В долната половина на таблицата са дадени и коефициентите на статистическата значимост на тези корелации. Когато един коефициент на значимост (Sig.) е под 0.05 се счита за статистически значим. Това означава, че може да се приеме нулевата хипотеза, че изчислената корелационната зависимост е валидна с вероятност 95%. Когато е над 0.05 (в краен случай над 0.10), Sig. е статистически незначим и в такъв случай не отговаря на изискванията на анализа. Може да се пренебрегне. В нашия случай всички корелационни коефициенти со големи и статистически значими.

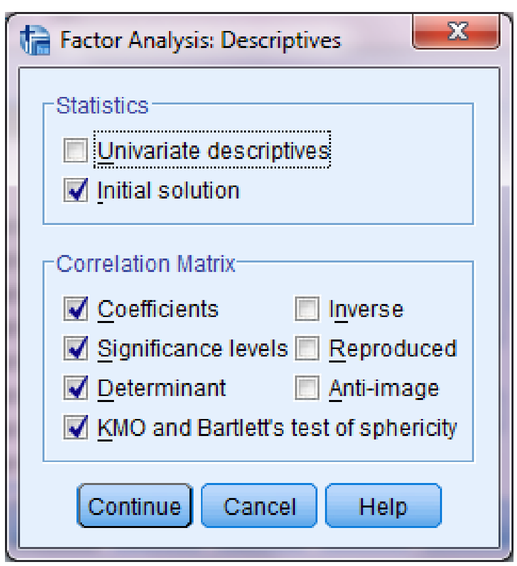
1.4. Изследване на корелационната матрица на независимите променливи, останали след провеждане на 1.3

В прозореца на ФА (Фиг. 2) оставяме в панела на *Variables* променливите без Eff, както е показано на Фиг. 5.



Фиг. 5. Основен прозорец на ФА, с избрани променливи.

Избираме ***Descriptives***, настройките и ***Continue*** от Фиг. 6.



Фиг. 6. Настройки на статистическите индекси за независимите променливи.

Получаваме последователно:

- Корелационната матрица на 5-те променливи (Фиг. 7):

Correlation Matrix[a]

		D	dr	L	Pin	PL
Correlation	D	1.000	.806	.800	.763	-.917
	dr	.806	1.000	.785	.757	-.785
	L	.800	.785	1.000	.946	-.847
	Pin	.763	.757	.946	1.000	-.733
	PL	-.917	-.785	-.847	-.733	1.000
Sig. (1-tailed)	D		.014	.015	.023	.002
	dr	.014		.018	.025	.018
	L	.015	.018		.001	.008
	Pin	.023	.025	.001		.030
	PL	.002	.018	.008	.030	

a. Determinant = .001

Фиг. 7. Корелационна матрица на независимите променливи.

Тази матрица е подматрица на тази от Фиг. 4. Детерминантата й е 0.001, т.е. - достатъчно малка, да означава наличие на мултиколинеарност на променливите, необходимо условие за ФА.

- Тест КМО и Бартлет тест за сферичност – тестове за адекватност на ФА

В нашия случай резултатите от тези тестове са показани на Фиг.8. Първият е обща индикация за многомерно разпределение на данните, близко до нормалното, когато КМО ≥ 0.5. Тестът на Бартлет проверява нулевата хипотеза, че данните не лежат на една права в пространството, т.е. облакът от данни има някаква "сферичност". В нашия случай тестът КМО=0.696>0.5 и тестът на Бартлет има Sig.=0.006<0.05.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.696
Bartlett's Test of Sphericity	Approx. Chi-Square	24.853
	df	10
	Sig.	.006

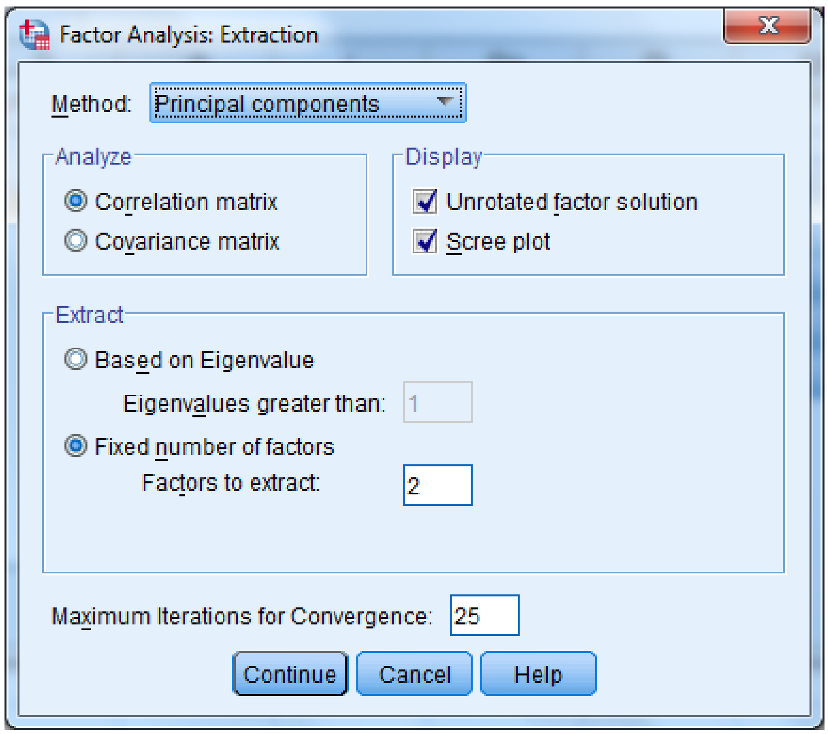
Фиг. 8. Проверка на тестовете за адекватност на ФА.

Заключение. ФА е адекватен и може да се проведе.

2) <u>Провеждане на ФА</u>

1.5. Избор на метод за извличане на факторите и определяне броя на факторите

От основния прозорец на ФА (Фиг. 5) и досега избраните настройки в *Descriptives*, отваряме панела ***Extraction*** (Fig. 9). Избираме посочените полета. Метод може да се избере от най-горното падащо меню. В случая е избран стандартният – Principal Component Analysis, Scree plot и фиксиран брой на факторите (Fixed number of factors) – 2 фактора. След това потвърждаваме с вуртуалния бутон ***Continue***.

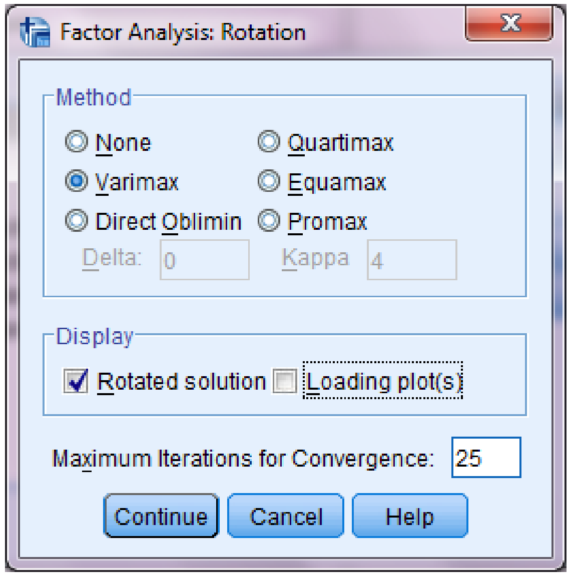


Фиг. 9. Избор на метод и брой фактори във ФА.

1.6. Избор на метод за въртене на факторите

От основния прозорец на ФА (Фиг. 5) отваряме панела ***Rotation*** (Fig. 10). От предлаганите 6 метода можем да изберем желан вариант. Най-често използваният метод е Varimax с нормализация на Кайзер. Този метод преобразува факторите във взаимноперпендикулярни базиси на факторното пространство, в случая – двумерно. От полето *Display* избираме *Rotated solution*.

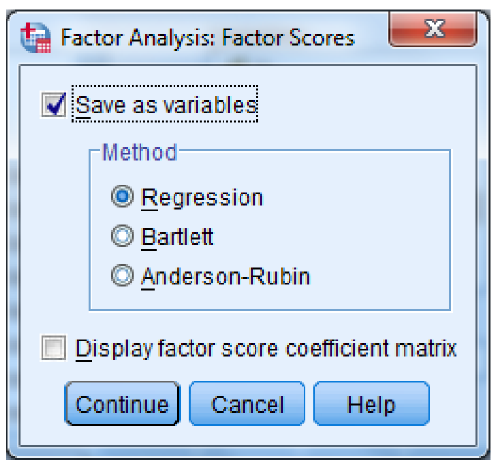
Потвърждаваме избора с вуртуалния бутон ***Continue***.



Фиг. 10. Избор на метод за въртене на факторите във ФА.

1.7. Съхраняване на факторните стойности

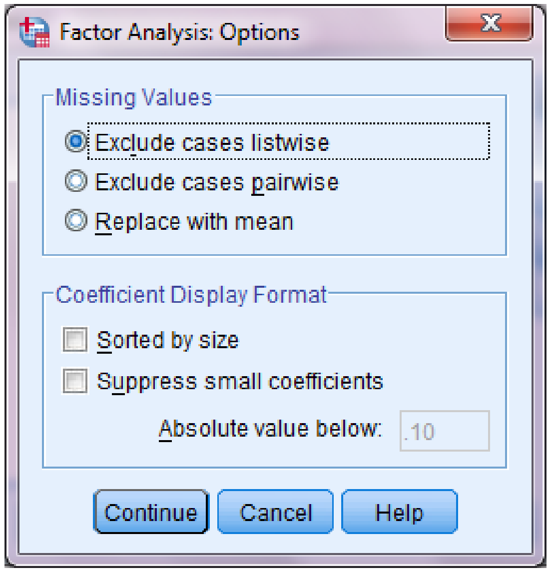
Ако желаем да запомним като променливи получените фактори, от основния прозорец на ФА (Фиг. 5) отваряме панела ***Scores...*** (Fig. 11). Избираме показаните настройки и потвърждаваме с ***Continue***.



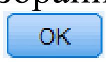
Фиг. 11. Настройки по запомняне на факторните стойности във ФА.

1.8. Избор на допълнителни настройки

За прецизиране на анализа, от основния прозорец на ФА (Фиг. 5) отваряме панела ***Options...*** (Fig. 12). Полето *Missing Values* дава възможност да изберем как да се постъпва при липсващи данни. Долното поле дава възможност за форматиране на факторните тегла. Избираме показаните настройки и потвърждаваме с ***Continue***.



Фиг. 12. Панел за избор на допълнителни опции във ФА.

След като са избрани всички елементи на ФА, от прозореца на ФА (Фиг. 5) избираме 

за да се стартира процедурата на ФА.

3) ***<u>Получени резултати от ФА</u>***

След извличане на 2-та фактора по избрания метод на Главните компоненти (Principal Component Analysis - PCA), ще получим т.н. начално факторно решение в *Component Matrix* с 2 фактора (components). То е показано на Фиг. 13. Нямаме ясно изразени факторни тегла. Началното факторно решение показва 5-те независими променливи, групирани в 2 компонента (фактора). Вижда се, че в него само първият компонент има големи тегла (първата колона), а вторият – доста по-ниски. Засега променливите са концентрирани по първия фактор (направление).

Component Matrix[a]

	Component	
	1	2
D	.929	-.272
dr	.894	-.116
L	.950	.256
Pin	.910	.391
PL	-.929	.262

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Фиг. 13. Началното факторно решение с два фактора.

Съответното разпределение на собствените стойности на корелационната матрица са дадени на Фиг. 14. Общата им сума е винаги равна на 1. Първият фактор обяснява 85.135% от данните, с добавяне на втория (Cumulative %), ще се обяснят 92.626%, ако вземем напр. 3 фактора, ще получим 97.500% и т.н. до 100%. Ние сме фиксирали 2 фактора и ще продължим нататък да опитаме построяването на модел с 2 фактора. Така игнорираме останалата част от 7.374% вариация в данните. Изборът на броя фактори зависи от изследователя.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.257	85.138	85.138	4.257	85.138	85.138	2.508	50.169	50.169
2	.374	7.488	92.626	.374	7.488	92.626	2.123	42.457	92.626
3	.244	4.874	97.500						
4	.105	2.095	99.595						
5	.020	.405	100.000						

Extraction Method: Principal Component Analysis.

Фиг. 14. Разпределение на общата вариация между компонентите.

- Scree графика на компонентите

Използва се, за да се разбере колко на брой фактори най-добре ще опишат изменчивостта на извадката. Много често се избират толкова фактори, колкото е броят на собствените стойности, по-големи от 1 (Принцип на Кайзер). През последните години години обаче, бяха установени доста случаи, за които това правило не е вярно.

- След въртенето на двата фактора с метода varimax и Kaiser normalization получаваме завъртяното решение (Фиг. 15).

Rotated Component Matrix[a]

	Component 1	Component 2
D	.872	.422
dr	.740	.514
L	.532	.827
Pin	.413	.901
PL	-.864	-.430

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 3 iterations.

Фиг. 15. Факторните тегла на завъртяното решение с 2 фактора.

От Фиг. 15 се вижда, че в първия фактор F1 (компонент 1) се групират отчетливо променливите D, dr и PL, с тегла съответно 0.872, 0.740 и (-0.864). Към втория фактор F2 се групират L(0.827) и Pin(0.901). Желателно е ясното разграничаване на променливите така, че да са с високи тегла само към един от факторите. Колкото по-малки са теглата им в другите фактори, толкова по-добър е факторният модел. Счита се стандартно, че прагът на теглата при това разграничаване е 0.5, но последното зависи силно и от размера на извадката. За нашия учебен пример, с извадка с размер n=7, полученото решение е напълно удовлетворително. Ако някоя променлива участва с високи тегла в повече от един фактор се счита, че броят на факторите не е добър и се отхвърля съответният факторен модел. Ако пък има променлива, чието участие е с ниски тегла във всички фактори, тя се счита за слабовлияеща в облака от данни и следва да отпадне от факторния модел. По-късно може да се включи в други статистически анализи, ако преценим, че е важна.

Така за нашите данни за лазера, от извадката получихме следните 2 фактора:

F1= {D, dr, PL}
F2= {L, Pin}

Влиянието на факторите в разпределянето на общата вариация е дадено на Фиг. 14, в най-дясната колона с „% of Variance”. Първият фактор обяснява 50.169% от извадката, а вторият - 42.457%.

- Запомняне на факторите

За да можем да използваме факторите в по-нататъшни изследвания, ги запомняне като променливи – виж настройките от Фиг. 11. Получените факторни стойности са дадени в Табл. 2. (сравни с Табл. 1).

Табл. 2. Факторни стойности за модела с 2 фактора за CuBr лазер.

k	F1	F2
1	-2.05584	-.21121
2	.27949	-1.19499
3	-.06327	.34520
4	.69144	.08725
5	.70904	-1.29897
6	-.33457	.88075
7	.77371	1.39197

Приложение на многомерен линеен регресионен анализ (ЛРА)

Статистическият метод на регресионен анализ е описан в лекциите. Тук ще го приложим в средата на SPSS за намиране на зависимостта на лазерната ефективност (променлива Eff) от факторите F1 и F2.

Ще отбележим, че за разглежданите данни пряко намиране на зависимостта между ефективността Eff и независимите променливи D, dr, L, Pin, PL не е статистически възможно, което ще бъде показано по-долу.

1) *Условия за прилагане на РА и валидност на резултатите*

Накратко ще приведем необходимите теоретични постановки за РА.

За прилагането както на едномерен, така и многомерен РА се изисква изпълнението на следните основни условия:

- Изходните данни да имат случаен характер, да бъдат от интервален или относителен тип. Категорийните данни трябва да бъдат предварително трансформирани в двоични.
- Наблюденията да бъдат независими.
- За всяка независима променлива, зависимата променлива трябва да има нормално или близко до нормалното разпределение.
- При линейна регресия се предполага, че общата зависимост има линеен или близък до линеен характер.

Регресионният модел се счита за статистически значим и пригоден за практическо използване и прогнози ако удовлетворява определени статистически оценки. Най-основните от тях при ниво 95% са:

- Нивото на значимост на целия модел, оценяван с F статистика (в таблица ANOVA) трябва да бъде < 0.05
- Коефициентът на детерминация R^2 и абсолютната стойност на коефициента на многомерна корелация R трябва да са близки до 1 и желателно по-големи от 0.5
- Изследването за мултиколинеарност между независимите променливи да е приемливо (за SPSS тестът е VIF < 10)
- Да се направи анализ на остатъците от регресията: разпределението на остатъците да е нормално (напр. с хистограма, P-P, Q-Q тест, тест на Колмогоров-Смирнов, тест на Шапиро-Уилкс), разстоянията на Махалонобис и Кук да са малки, да се изследват отдалечените точки и др.
- Препоръчва се и провеждане на процедура на крос-валидация. Тя се състои в разделяне на дадената извадка на две непресичащи се подмножества, наречени тренировъчно и тестово. С тренировъчното множество се построява и изследва нов модел. С този модел се

предсказват стойностите на зависимата променлива за наблюденията от тестовото множество. Ако резултатите са близки до модела с тренировъчното подмножество и с изходния модел, се приема че изходния модел е валиден.

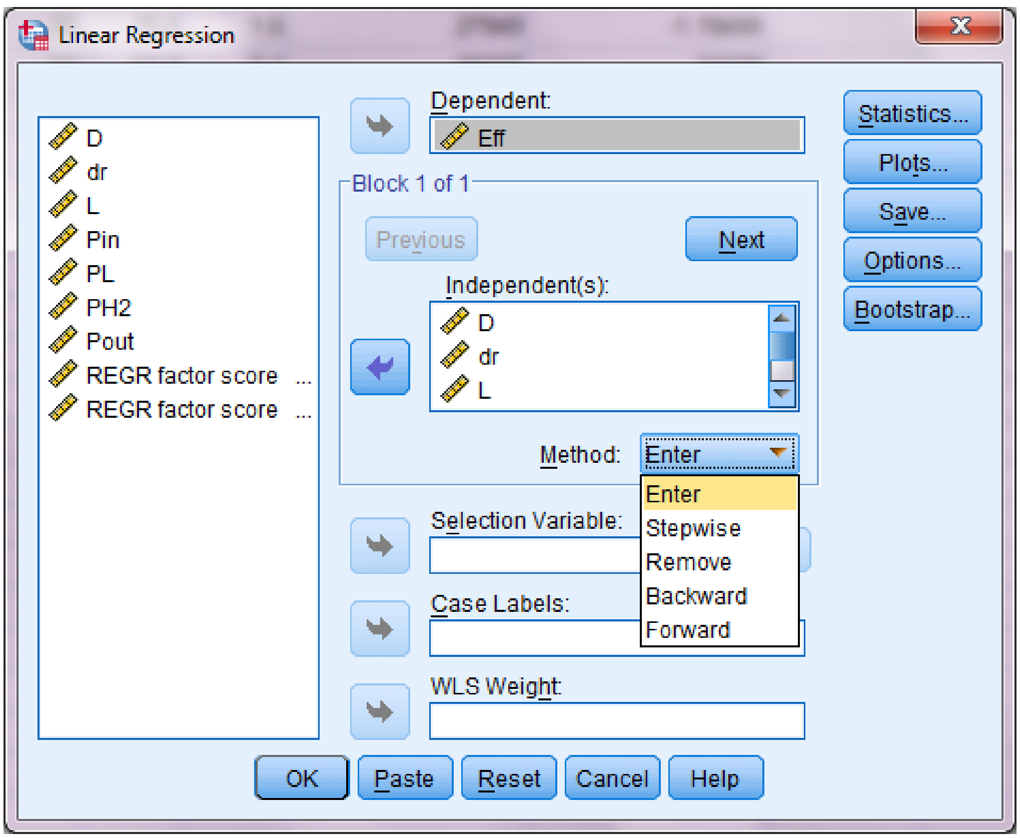
2) ***Прилагане на РА***

Ще приложим многомерна линейна регресия за данните на лазер с меден бромид от Табл. 1.

1.1. Стартиране на процедурата за РА

Analyze/Regression/Linear

Появява се прозорецът, показан на Фиг. 16.

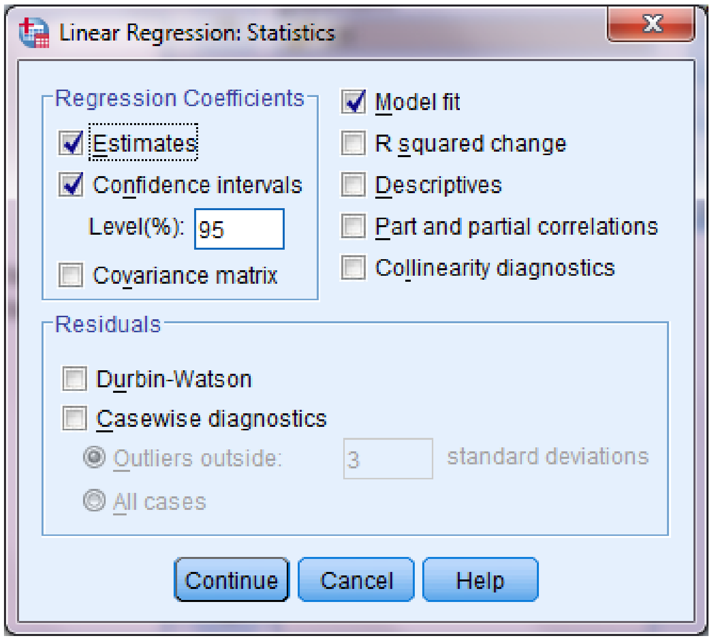


Фиг. 16. Основен прозорец за регресионнен анализ, с падащо меню с възможните методи.

Пренасяме променливата Eff от лявата част на прозореца вдясно в полето *Dependent* (зависима променлива). Отдолу пренасяме желаните независими променливи (предиктори). Тук са показани D, dr, L. Но по-нататък ще дадем конкретните резултати за 2 случая – с петте независими променливи и отделно – с двата фактора F1, F2. Избираме метода *Enter* (означава обикновена линейна регресия).

1.2. Избор на необходимите статистики

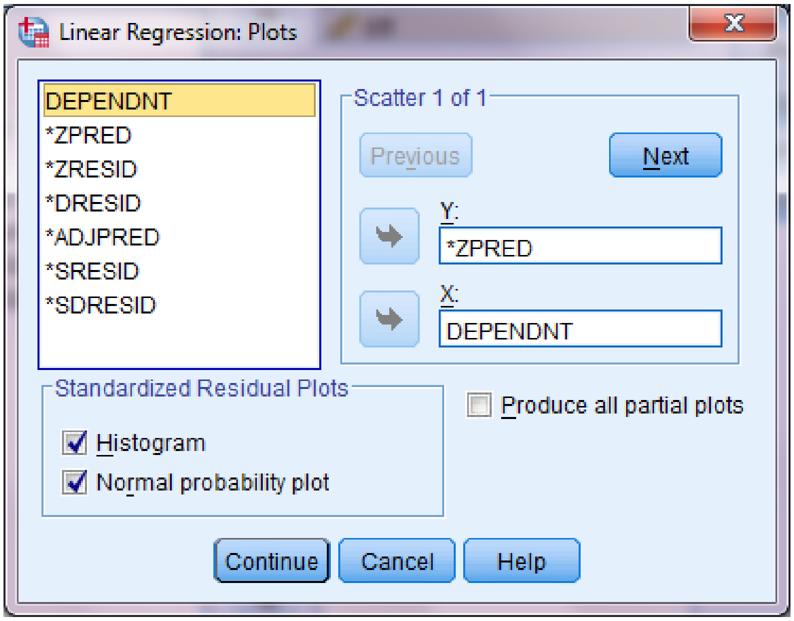
От основния прозорец (Фиг. 16) отваряме ***Statistics*** и правим настройките, показани на Фиг. 17. Ще се визуализират оценките на регресионните коефициенти с доверителните им интервали и коефициентът на детерминация R^2 (Model Fit).



Фиг. 17. Избор на желаните статистики, които ни интересуват.

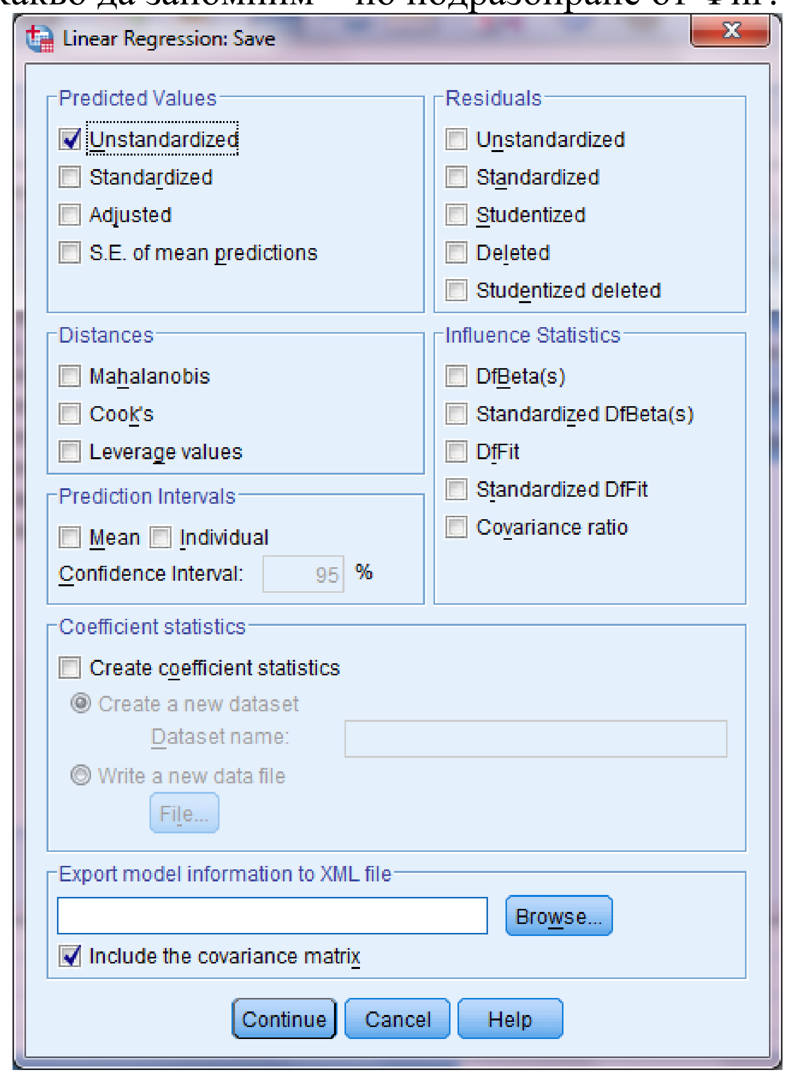
1.3. Показване на графики

От ***Plot(s)*** можем да изберем различни графики на остатъците, стандартизираните остатъци, сравнение на оценката и изходните данни. Тук в лявата колона са съответно: DEPENDNT (зависима променлива, тук *Eff*), ZPRED (стандартизирани предсказани стойности, в нашия случай на $\widehat{Eff}$), ZRESID (стандартизираните остатъци) и др. На Фиг. 18 е избран случаят да се изобрази зависимата променлива по оста X и нейната оценка $\widehat{Eff}$ по оста Y. Също така се на графиката на стандартизираните остатъци ще се добави хистограма и кривата на нормалното разпределение, които се задават подобно с Next.



Фиг. 18. Избор на графики за изследване на остатъците и приближението.

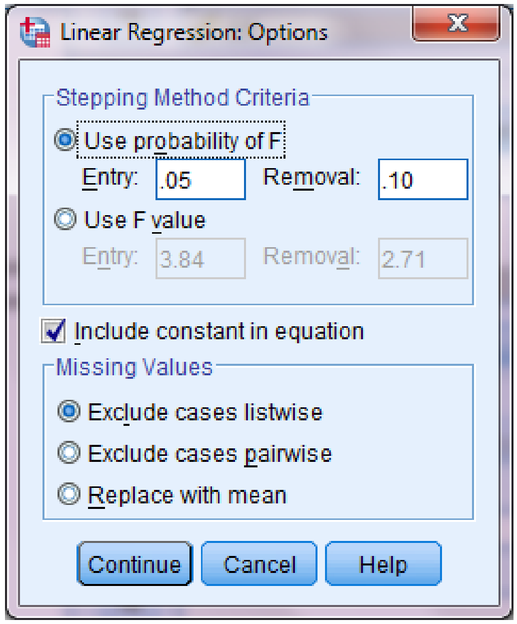
1.4. Какво да запомним – по подразбиране от Фиг. 19.



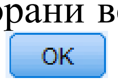
Фиг. 19. Възможности за запомняне на части от анализа.

1.5. Настройки на Опции

Става от ***Options***. Ще оставим настройките по подразбиране, показани на Фиг. 20.



Фиг. 20. Панелът на *Options*.

След като са избрани всички настройки на РА, от основния прозорец (Фиг. 16) избираме 

за да се стартира процедурата на РА.

3) <u>Резултати от РА от опита за регресия с петте предиктора D, dr, L, Pin, PL</u>

Най-напред ще направим опит директно да приложим многомерна линейна регресия за данните на лазер с меден бромид от Табл. 1. В случая опитваме да намерим обща зависимост на Eff от независимите променливи D, dr, L, Pin, PL във вида:

$$\widehat{Eff} = b_0 + b_1 D + b_2 dr + b_3 L + b_4 Pin + b_5 PL \ . \qquad (1)$$

Ще покажем че полученият регресионен модел не е валиден.

След изпълнение на процедурата получаваме:

- Описание на използваните променливи и метод на регресия:

Variables Entered/Removed[b]

Mode l	Variables Entered	Variables Removed	Method
1	PL, Pin, dr, D, L	.	Enter

a. All requested variables entered.
b. Dependent Variable: Eff

- Получаваме големи стойности на корелационния коефициент R=0.999 и коефициент на детерминация R2 (квадрат на R), със стойност 0.997.

Model Summary[b]

Mode l	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.999[a]	.997	.985	.0856

a. Predictors: (Constant), PL, Pin, dr, D, L
b. Dependent Variable: Eff

- Таблица ANOVA (ANalysys Of VAriance)

Тук обаче виждаме, че моделът е статистически незначим, тъй като Sig. =0.086>0.05, при ниво на значимост 0.05. Това е така, защото F-статистиката има стойност 78.310>$F_{критично}$. Тук n=7, df=5 степени на свобода. Следователно нулевата хипотеза е приема и моделът не е статистически значим (виж също примера към лекция 8).

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.870	5	.574	78.310	.086[a]
	Residual	.007	1	.007		
	Total	2.877	6			

a. Predictors: (Constant), PL, Pin, dr, D, L
b. Dependent Variable: Eff

- Коефициентите на този статистически незначим модел са:

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	.023	.764		.030	.981
	D	.019	.008	.380	2.433	.248
	dr	.025	.004	.602	6.469	.098
	L	-.001	.003	-.046	-.190	.881
	Pin	.132	.156	.171	.849	.552
	PL	.100	.424	.044	.236	.852

a. Dependent Variable: Eff

Виждаме, че и тук за всички коефициенти имаме Sig.>0.05, т.е. също са статистически незначими.

Заключение. Нямаме право да прилагаме регресионния модел (1).

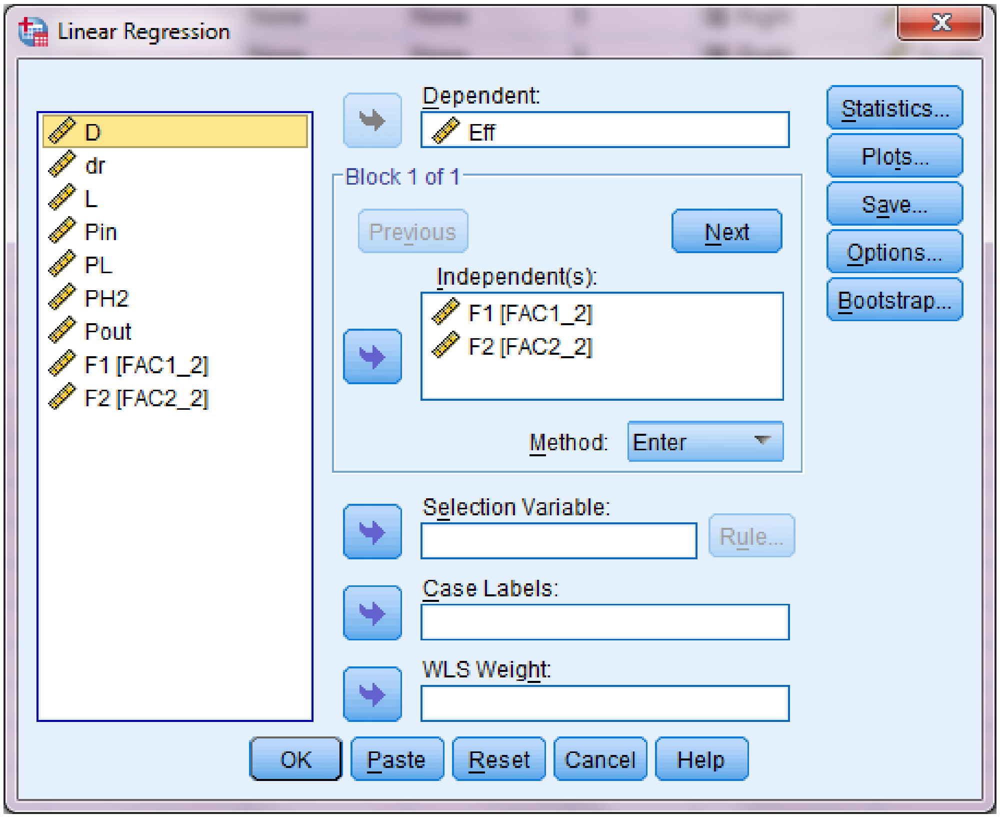
4) Резултати от РА с факторите F1, F2

В случая опитваме да намерим обща зависимост на Eff от независимите променливи F1, F2 във вида:

$$\widehat{Eff} = b_0 + b_1 F1 + b_2 F2 . \qquad (2)$$

Ще покажем че полученият регресионен модел сега е валиден.

От ***Analyse/Regression/Linear ...***
въвеждаме променливата Eff за зависима и F1, F2 за независими променливи (виж Фиг. 21).



Фиг. 21. Избор на нови независими променливи – факторите F1, F2.

След изпълнение на процедурата получаваме последователно:

- Описание на използваните променливи F1, F2 и метод на регресия – Enter (обикновена регресия с МНМК):

Variables Entered/Removed[b]

Mode l	Variables Entered	Variables Removed	Method
1	F2, F1[a]	.	Enter

a. All requested variables entered.
b. Dependent Variable: Eff

- Получаваме големи стойности на корелационния коефициент R=0.968 и коефициент на детерминация R2 (квадрат на R), със стойност 0.937.

Model Summary[b]

Mode l	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.968[a]	.937	.906	.2121

a. Predictors: (Constant), F2, F1
b. Dependent Variable: Eff

- Таблица ANOVA

Тук вече виждаме, че моделът е статистически значим, тъй като Sig. =0.004<0.05, при ниво на значимост 0.05. Тук n=7, df=2 степени на свобода. Следователно нулевата хипотеза се отхвърля и моделът е статистически значим.

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.697	2	1.349	29.969	.004[a]
	Residual	.180	4	.045		
	Total	2.877	6			

a. Predictors: (Constant), F2, F1

b. Dependent Variable: Eff

- Коефициентите на регресионния модел са:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.943	.080		24.232	.000
	F1	.544	.087	.785	6.276	.003
	F2	.393	.087	.567	4.533	.011

a. Dependent Variable: Eff

За всички коефициенти имаме Sig.<0.05, т.е. също са статистически значими.

С помощта на коефициентите записваме регресионното уравнение има вида:

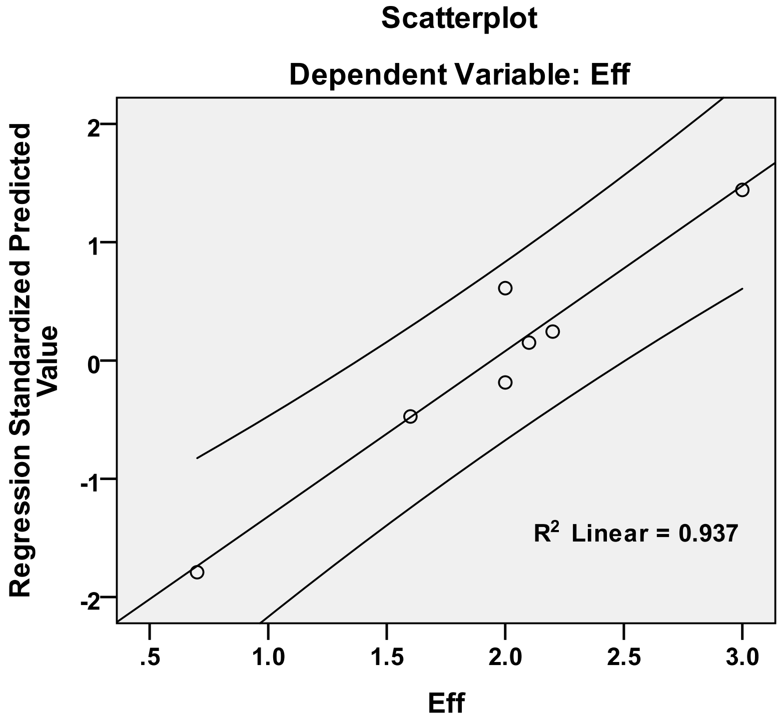
$$\widehat{Eff} = 1.943 + 0.544\,F1 + 0.393\,F2 \qquad (3а)$$

или стандартизирано във вида:

$$\widehat{Eff} = 0.785\,F1 + 0.567\,F2 \qquad (3б)$$

- Графика на приближението.

Тя е показана на Фиг. 22.



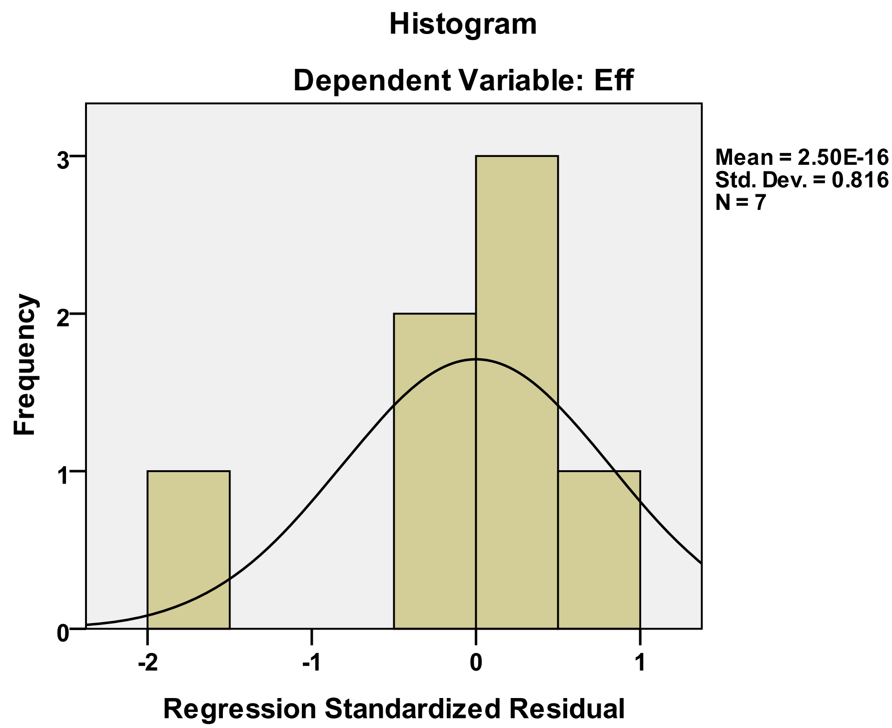
Фиг. 22. Точкова графика на функцията Eff и предсказаните стойности по ординатата с 95% доверителни интервали.

5) Изследване на остатъците

Без да правим пълен статистически анализ даваме тук двете най-основни графики (виж „Условия за прилагане на РА", в началото на регресионнен анализ по-горе).

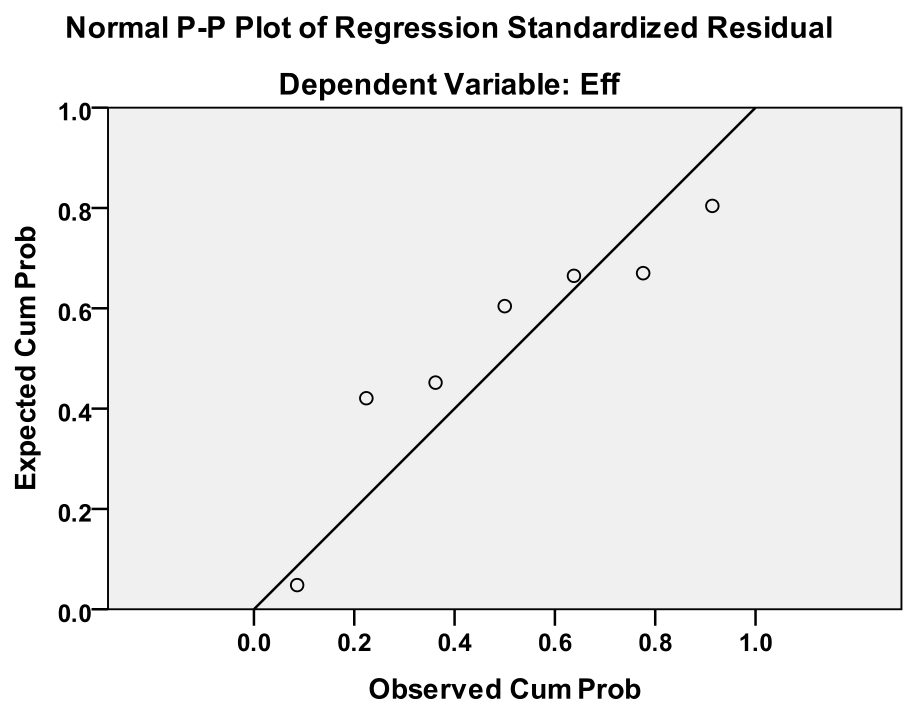
- Хистограма на стандартизираните остатъци

На Фиг. 23 се вижда хистограмата на разпределението на стандартизираните остатъци с добавена нормална крива, както е декларирано в Plot(s) (виж Фиг. 18). Макар малкия брой наблюдения n=7, тя има разпределение,близко до нормалното, като средната е 0 и отклонението е 0.816.



Фиг. 23. Хистограма на остатъците.

- Нормална P-P графика на стандартизираните остатъци (резидиуми) на регресия –Фиг. 24.



Фиг. 24. Нормала на остатъците.

В идеалния случай точките на Фиг. 24 трябва да са възможно по-близо и да съвпадат с 45% наклонената права линия, без „тежки опашки”.

Полученото регресионно уравнение (3) е приемливо, но само условно, тъй като се базира на много малка извадка данни и не удовлетворява всички условия за прилагане на РА.