

وزارت جهاد کشاورزی
سازمان تحقیقات و آموزش کشاورزی
مرکز تحقیقات کشاورزی و منابع طبیعی لرستان
(ایستگاه تحقیقات کشاورزی بروجرد)

کاربرد نرم افزار SAS در تجزیه و تحلیل داده‌های آماري

تهیه کننده:

عبدالامير محیسنی

دی ماه ۱۳۸۳

صفحه	موضوع
۱	مقدمه:
۱	آشنایی با نرم افزار SAS:
۱	الف- پنجره های مهم در محیط نرم افزار sas:
۱	ب- عملگرهای محاسباتی و قابلیت های نرم افزار:
۱	برنامه نویسی در SAS
۲	روشهای وارد نمودن داده ها در sas
۲	بخش اول
۲	آمار توصیفی
۲	۱- دستور PROC MEANS
۳	۲- دستور UNIVARIATE:
۳	۳- رسم جدول فراوانی و نمودار فراوانی با استفاده از دستور FREQ و CHART:
۴	۴- رسم جدول فراوانی دو بعدی
۴	۵- رسم نمودار فراوانی بصورت عمودی (VBAR) یا افقی (HBAR):
۴	۶- رسم نمودار نقاط میانی
۴	۷- مشخص نمودن تعداد طبقات در نمودار فراوانی
۴	۸- رسم نمودار فراوانی با دستور PLOT
۵	۹- چنانچه بخواهیم برای هر سطح یک متغیر نمودار فراوانی جداگانه رسم کنیم از متغیر گروه بندی (GROUP) استفاده می کنیم.
۵	۱۰- آزمون t برای مقایسه میانگین دو گروه (دستور TTEST)
۶	۱۱- مقایسه میانگین ها با استفاده از دستور PROC MEANS
۷	بخش دوم
۷	کاربرد SAS در تجزیه و تحلیل طرح های آزمایشات کشاورزی
۷	۱۲- طرح کاملاً تصادفی متعادل
۷	مقایسه میانگین ها با استفاده از نرم افزار SAS
۸	مقایسات متعامد یا ارتوگونال (Contrasts)
۱۰	۱۳- مثال ۱ برای مقایسات متعامد (Orthogonal):
۱۱	۱۴- مثال ۲ برای مقایسات متعامد (Orthogonal):
۱۲	۱۵- طرح کاملاً تصادفی نامتعادل
۱۲	۱۶- طرح بلوک های کامل تصادفی
۱۲	۱۷- طرح بلوک های کامل تصادفی با داده های از دست رفته (طرح بلوک های کامل تصادفی نامتعادل)
۱۳	۱۸- طرح مربع لاتین متعادل

صفحه	موضوع
۱۳	۱۹- طرح مربع لاتین با داده های از دست رفته
۱۴	طرح های مربع لاتین با تکرار مربعات
۱۴	۲۰- طرح مربع لاتین مکرر با مربع های مستقل:
۱۴	۲۱- طرح مربع لاتین مکرر با مربع های جفت شده
۱۴	۲۲- طرح مربع یودن (YUDEN SQUARE DESIGN)
۱۵	طرح های چندمشاهده ای
۱۵	۲۳- طرح بلوک های کامل تصادفی با داده های چند مشاهده ای:
۱۶	۲۴- طرح کاملاً تصادفی با داده های چند مشاهده ای
۱۷	آزمایشات فاکتوریل بدون اختلاط
۱۷	۲۵- آزمایشات فاکتوریل در قالب طرح پایه کاملاً تصادفی
۱۷	۲۶- آزمایشات فاکتوریل در قالب طرح پایه بلوکهای کامل تصادفی
۱۷	۲۷- طرح فاکتوریل نامتعادل
۱۸	تفکیک SS ها
۱۹	۲۸- تفکیک SS به اجزای خطی و درجه دوم در یک آزمایش فاکتوریل
۱۹	۲۹- تفکیک SS برای سطوح فاکتور B در داخل سطوح فاکتور A
۲۰	آزمایشات فاکتوریل با اختلاط
۲۱	۳۰- آزمایشات فاکتوریل با اختلاط کامل
۲۲	۳۱- آزمایشات فاکتوریل با اختلاط ناقص
۲۳	طرح های متداخل (NESTED)
۲۴	۳۲- یک طرح متداخل (NESTED DESIGN) دارای اثرات ثابت و تصادفی
۲۵	۳۳- طرح کرت های خرد شده SPLIT PLOT
۲۶	۳۴- طرح کرت های خرد شده در زمان
۲۷	۳۵- تجزیه جداگانه در یک طرح کرت های خرد شده در زمان برای سطوح فاکتور A
۲۸	۳۶- طرح کرت های دو بار خرد شده (Split-Split Plot)
۲۹	۳۷- طرح کرت های خرد شده فاکتوریل (Split-Plot factorial)
۲۹	۳۸- طرح فاکتوریل کرت های خرد شده (Factorial Split-plot)
۳۱	۳۹- طرح کرت های نواری (Strip-Split-Block)
۳۲	۴۰- طرح کرت های خرد شده نواری (Split-Block)
۳۳	۴۱- طرح کرت های خرد شده نواری با مقایسات متعامد
۳۴	تجزیه مرکب طرح های آزمایشی
۳۴	۴۲- تجزیه مرکب طرح بلوکهای کامل تصادفی در چند مکان

صفحه	موضوع
۳۵	۴۳- طرح بلوکهای کامل تصادفی در چند سال
۳۶	۴۴- طرح بلوکهای کامل تصادفی در چند سال و چند مکان
۳۷	۴۵- تجزیه جداگانه (به تفکیک سال یا مکان) یک طرح بلوکهای کامل تصادفی در چند سال و مکان با داده‌های تجزیه مرکب.
۳۸	۴۶- طرح کرت‌های خرد شده در چند مکان
۳۸	۴۷- طرح کرت‌های خرد شده در چند سال و چند مکان
۳۸	۴۸- طرح مربع لاتین در چند سال و چند مکان
۳۹	برآورد اجزای واریانس
۳۹	۴۹- برآورد اجزای واریانس یک طرح بلوکهای کامل تصادفی
۳۹	۵۰- برآورد اجزای واریانس یک طرح بلوکهای کامل تصادفی در چند سال و چند مکان
۴۰	تجزیه کواریانس در طرحهای آزمایشی
۴۰	۵۱- تجزیه کواریانس در طرح کاملاً تصادفی
۴۲	۵۲- تجزیه کواریانس در یک طرح بلوکهای کامل تصادفی
۴۴	تهیه نقشه‌های تصادفی برای طرحهای آزمایشی
۴۴	۵۳- تهیه نقشه تصادفی برای یک طرح کاملاً تصادفی
۴۴	۵۴- تهیه نقشه تصادفی برای یک طرح کرت‌های خرد شده
۴۵	۵۵- تهیه نقشه تصادفی برای یک طرح بلوکهای کامل تصادفی
۴۵	۵۶- تهیه نقشه تصادفی برای یک طرح مربع لاتین با ۴ ردیف، ۴ ستون و ۴ تیمار
۴۵	۵۷- تهیه نقشه تصادفی برای یک طرح آشیانه‌ای
۴۶	۵۸- تهیه نقشه تصادفی برای یک طرح فاکتوریل یا اسپلیت-اسپلیت پلات
۴۶	۵۹- تهیه نقشه تصادفی برای یک طرح اسپلیت-فاکتوریل
۴۷	
۴۸	بخش سوم
۴۸	کاربرد SAS در تجزیه‌های رگرسیون و همبستگی
۴۸	۶۰- رگرسیون ساده خطی
۴۹	۶۱- رگرسیون ساده خطی با استفاده از دستورات P، CLM و CLI
۴۹	۶۲- رگرسیون ساده خطی با استفاده از دستورات I و XPX
۵۰	۶۳- آزمون فقدان برازش (Lack of fit)
۵۲	۶۴- تعیین و تصحیح عیوب مدل (تجزیه باقیمانده‌ها در تجزیه رگرسیون)
۵۳	۶۵- بررسی نرمال بودن باقیمانده‌ها

صفحه	موضوع
۵۵	۶۶- تعیین و تصحیح عیوب مدل (تعیین نقاط پرت یا OUTLIERS)
۵۷	۶۷- تعیین و تصحیح عیوب مدل (تبدیل متغیرها)
۵۸	۶۸- تعیین و تصحیح عیوب مدل (رگرسیون وزنی)
۵۸	۶۹- رگرسیون چند گانه (Multiple Regression)
۶۰	۷۰- رگرسیون استاندارد شده با تجزیه مسیر
۶۰	۷۱- مدل های چند جمله ای
۶۱	۷۲- مدل های رگرسیون چند جمله ای با استفاده از دستور proc glm
۶۲	۷۳- مدل های رگرسیون با تمرکزدهی (centering)
۶۲	۷۴- مدل رگرسیون برای متغیر نشانگر
۶۲	۷۵- مدل رگرسیون برای متغیر نشانگر با کاربرد گزاره by
۶۴	۷۶- تعیین وجود همخطی در تجزیه رگرسیون (MULTICOLLINEARITY)
۶۵	۷۷- تجزیه رگرسیون با استفاده از دستور ortoreg
۶۶	۷۸- رسم خط رگرسیون
۶۶	۷۹- تجزیه رگرسیون و رسم خط رگرسیون به همراه نمایش پراکنش نقاط
۶۶	گزینش متغیرها و مدل سازی در رگرسیون چند گانه
۶۷	۸۰- گزینش متغیر به روش forward
۶۸	۸۱- گزینش متغیر و مدلسازی به روش backward
۶۹	۸۲- گزینش متغیر و مدلسازی به روش stepwise
۷۱	۸۳- گزینش متغیر و مدلسازی به روش Maxr
۷۳	۸۴- گزینش متغیر به روش Minr
۸۰	۸۵- گزینش متغیر به روش Rsquare
۸۱	۸۶- گزینش متغیر به روش Adjrsq
۸۲	۸۷- گزینش متغیر به روش Cp
۸۳	مدل های رگرسیون غیر خطی
۸۳	۸۸- مدل های رگرسیون غیر خطی لجستیک (مدل های لجستیک) (logistic models)
۸۴	۸۹- مدل های رگرسیون غیر خطی تکه ای (مدل های تکه ای) (Semented models)
۸۶	تجزیه همبستگی
۸۷	۹۰- تجزیه همبستگی به روش پیرسون
۸۷	۹۱- تجزیه همبستگی به روش پیرسون با اضافه نمودن دستورات فرعی NOPROP و NOSIMPLE
۸۸	۹۲- تجزیه همبستگی به روش اسپیرمن
۸۸	بخش چهارم

صفحه	موضوع
۸۸	کاربرد SAS در تجزیه‌های چند متغیره
۸۸	تجزیه به مؤلفه‌های اصلی
۸۹	۹۳- تجزیه به مؤلفه‌های اصلی با استفاده از ماتریس کواریانس
۹۱	۹۴- تجزیه به مؤلفه‌های اصلی با ماتریس ضرایب همبستگی
۹۳	تجزیه به عاملها (FACTOR ANALYSIS)
۹۴	۹۵- تجزیه به عاملها به روش PRINCIPAL COMPONENT
۹۸	۹۶- تجزیه به عاملها به روش ML
۱۰۳	تجزیه کلاستر
۱۰۳	۹۷- تجزیه کلاستر به روش average
۱۰۴	۹۸- تجزیه کلاستر به روش ward
۱۰۴	۹۹- تجزیه همبستگی کانونیک (CANONICAL CORRELATION ANALYSIS)

مقدمه:

SAS مخفف STATISTICAL ANALYSIS SYSTEM می باشد. نسخه های قدیمی تر این نرم افزار مانند نسخه 6.12 با ویندوزهای ME سازگاری نداشتند اما نسخه 8 با همه ویندوزهای رایج سازگار است. نکته مهم و قابل توجه در استفاده از این نرم افزار، تغییر تاریخ رایانه می باشد. به طوریکه برای نسخه 6.12 باید تاریخ رایانه بین 29 اکتبر 1996 لغایت 3 آپریل 1998 و برای نسخه 8 این نرم افزار بین 22 نوامبر 1992 (2 آذر 71) لغایت 31 اکتبر 2002 (10 آبان 81) تنظیم گردد. در غیر این صورت نرم افزار اجرا نخواهد شد.

آشنایی با نرم افزار SAS:

الف- پنجره های مهم در محیط نرم افزار sas:

1- پنجره PG یا PROGRAM EDITOR:

2- پنجره LOG:

3- پنجره OUTPUT:

ب- معرفی عملگرهای محاسباتی و قابلیت های نرم افزار:

در SAS برای جمع از +، برای تفریق از -، برای تقسیم از /، برای ضرب از * و برای توان 2 از ** استفاده می شود. سازگاری این نرم افزار با محیط های EXCEL، MINITAB، SPSS، GS+، WORD، NOTPAD و باعث سادگی کاربرد و گسترش استفاده از آن توسط کاربران شده است.

برنامه نویسی در SAS :

یک برنامه SAS از دو قسمت مهم تشکیل شده است:

الف: DATA STEP: معرفی متغیرها و وارد کردن داده ها در این قسمت از برنامه انجام می گیرد و با کلمه DATA شروع می شود.

```
DATA;
  input block treat$ y1-y2@@;
  cards;
  1 a 30 42.5 2 a 22 34 3 a 31.5 0 4 a 25.5 0
  1 b 0 29.5 2 b 0 23 3 b 0 27 4 b 0 33
  1 c 23 0 2 c 50 0 3 c 54 0 4 c 48 0
  1 d 16.5 0 2 d 35 0 3 d 37 0 4 d 30.5 0
```

ب- PROC STEP: نوشتن دستورالعملهای SAS می باشد که با PROC یعنی مخفف PROCEDURE مشخص می شود که روی داده ها کاری انجام می شود. این قسمت نیز با کلمه PROC شروع می شود.

```
proc anova;
  classes block treat;
  model y1-y2=block treat;
  mean treat/duncan alpha=0.05;
  mean block/duncan alpha=0.05;
run;quit;
```

نکته مهم: * بعد از پایان هر دستور از علامت سمی کولن (;) استفاده می شود*

روشهای وارد نمودن داده ها در sas

الف- روش پشت سر هم:

```
data a;
  input block treat$ y1-y2 @@;
  cards;
    1 a 30 42.5 2 a 22 34 3 a 31.5 0 4 a 25.5 0
    . . . . .
```

ب- روش ستونی:

```
data a;
  INPUT REP TREAT$ Y1;
  CARDS;
  1 A 30
  1 B 9
  1 C 16
  1 D 10
  1 E 30
  1 F 18
  1 G 17
  ;
```

ج- روشهای دیگر (مثلاً داده ها تحت یک نام و بصورت یک فایل ذخیره شده باشند).

بخش اول

آمار توصیفی

. دستور PROC MEANS

```
data a;
  input var$ stage1 stage2 stage3 adult;
  cards;
  A 2.5 3 4 5
  A 1 1.5 3 4.5
  A 2 3 3.6 5
  A 3 3.1 3.5 4
  A 1.5 2 3.8 5.1
  B 3 3.2 4.5 4.9
  B 3.9 4.1 4.4 5.8
  B 3.8 3.9 4.8 6
  B 3.1 3.8 4.1 5.5
  B 3 3.6 4.5 5.8
  ;
  proc means maxdec=2 n mean std sum var T prt stderr MIN MAX CV SKEWNESS KURTOSIS
  RANGE CSS USS;
  var stage1 stage2 stage3 adult;
  class var;
  run;
```

The MEANS Procedure

var	Obs	Variable	N	Mean	Std Dev	Sum	Variance	t Value	Pr > t
A	5	stage1	5	2.00	0.79	10.00	0.63	5.66	0.0048
		stage2	5	2.52	0.73	12.60	0.53	7.76	0.0015
		stage3	5	3.58	0.38	17.90	0.14	21.24	<.0001
		adult	5	4.72	0.47	23.60	0.22	22.66	<.0001
B	5	stage1	5	3.36	0.45	16.80	0.20	16.68	<.0001
		stage2	5	3.72	0.34	18.60	0.12	24.32	<.0001
		stage3	5	4.46	0.25	22.30	0.06	39.73	<.0001
		adult	5	5.60	0.43	28.00	0.18	29.11	<.0001

var	N	Variable	Std Error	Minimum	Maximum	Coeff of Variation	Skewness	Kurtosis
A	5	stage1	0.35	1.00	3.00	39.53	0.00	-1.20
		stage2	0.32	1.50	3.10	28.81	-0.86	-1.81
		stage3	0.17	3.00	4.00	10.53	-0.86	1.09
		adult	0.21	4.00	5.10	9.87	-1.18	0.08
B	5	stage1	0.20	3.00	3.90	13.41	0.60	-3.13
		stage2	0.15	3.20	4.10	9.19	-0.84	0.70
		stage3	0.11	4.10	4.80	5.63	-0.20	1.50
		adult	0.19	4.90	6.00	7.68	-1.38	1.82

var	N	Variable	Range	Corrected SS	USS
A	5	stage1	2.00	2.50	22.50
		stage2	1.60	2.11	33.86
		stage3	1.00	0.57	64.65
		adult	1.10	0.87	112.26

. دستور UNIVARIATE :

```
data a;
input var$ stage1 stage2 stage3 adult;
cards;
;
proc UNIVARIATE;
var stage1 stage2 stage3 adult;
BY var;
Run;
```

۳. رسم جدول فراوانی و نمودار فراوانی با استفاده از دستور FREQ و CHART :

```
DATA A;
INPUT A B;
CARDS;
71 167
69 160
. .
;
PROC FREQ;
TABLE A B;
RUN;
```

The FREQ Procedure

A	Frequency	Percent	Cumulative Frequency	Cumulative Percent
66	1	3.23	1	3.23
67	2	6.45	3	9.68
68	3	9.68	6	19.35
69	6	19.35	12	38.71
70	9	29.03	21	67.74
71	6	19.35	27	87.10
72	3	9.68	30	96.77
73	1	3.23	31	100.00

B	Frequency	Percent	Cumulative Frequency	Cumulative Percent
131	1	3.23	1	3.23
138	1	3.23	2	6.45
140	2	6.45	4	12.90
144	1	3.23	5	16.13
146	1	3.23	6	19.35
151	1	3.23	7	22.58
154	2	6.45	9	29.03
155	1	3.23	10	32.26
157	2	6.45	12	38.71
158	1	3.23	13	41.94
.
.
193	1	3.23	31	100.00

۴. رسم جدول فراوانی دو بعدی

```
DATA A;
INPUT A B;
CARDS;
71      167
.
.
;
PROC FREQ;
TABLE A*B;
RUN;
```

۵. رسم نمودار فراوانی بصورت عمودی (VBAR) یا افقی (HBAR):

```
DATA A;
INPUT A B;
CARDS;
71      167
.
.
;
PROC CHART;
VBAR A;
HBAR B;
RUN;
```

۶. رسم نمودار نقاط میانی

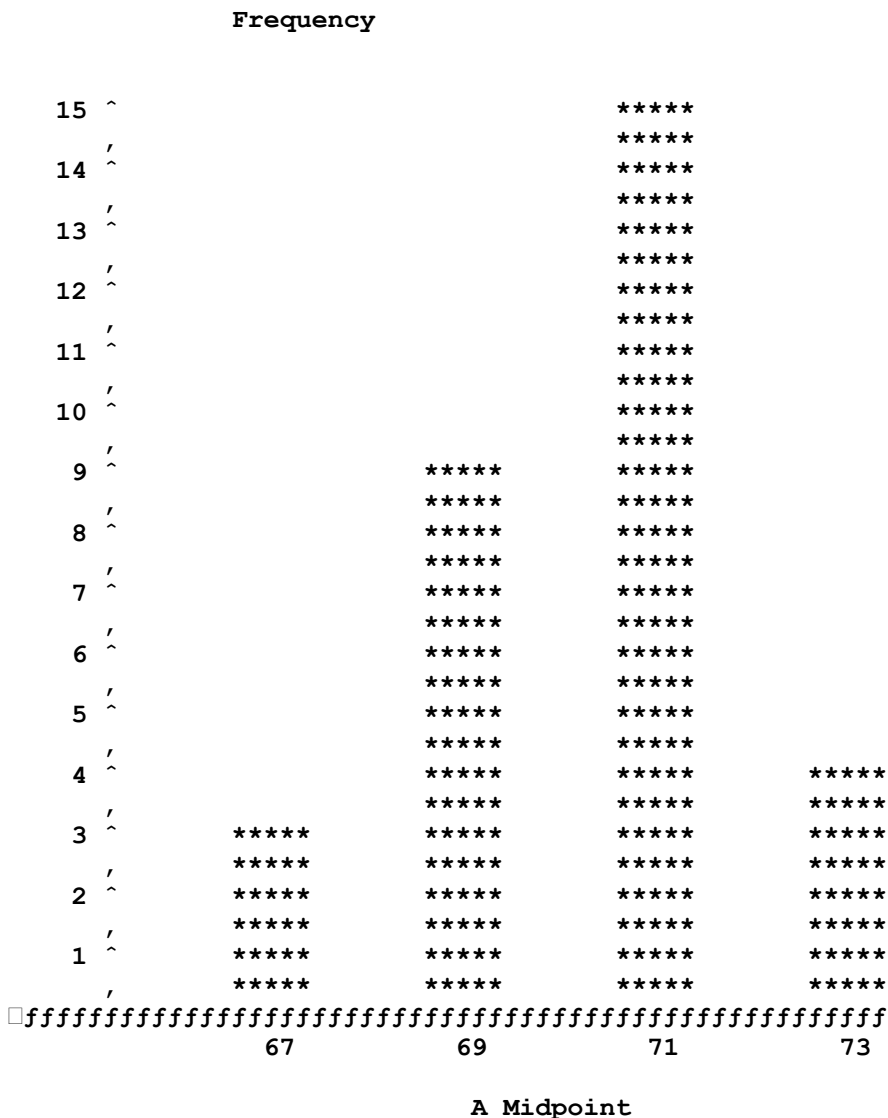
```
DATA A;
INPUT A B;
CARDS;
71      167
.
.
;
PROC CHART;
VBAR A/MIDPOINTS=33 35;
RUN;
```

۷. مشخص نمودن تعداد طبقات در نمودار فراوانی

```
DATA A;
INPUT A B;
CARDS;
71      167
.
.
;
PROC CHART;
VBAR A/LEVELS=4;
RUN;
```

۸. رسم نمودار فراوانی با دستور PLOT

```
DATA A;
INPUT A B;
CARDS;
71      167
69      160
.
.
;
PROC PLOT;
PLOT A*B;
RUN;
```



۹. چنانچه بخواهیم برای هر سطح یک متغیر نمودار فراوانی جداگانه رسم کنیم از متغیر گروه بندی (GROUP) استفاده می کنیم.

```
DATA A;
INPUT SEX$ A B C;
CARDS;
M 71      167
M 69      160
.
;
PROC CHART;
VBAR A/GROUP=SEX;
RUN;
```

۱۰. آزمون t برای مقایسه میانگین دو گروه (دستور TTEST)

```
DATA A;
INPUT SEX$ A B;
CARDS;
.
;
PROC TTEST;
CLASS SEX;
VAR A B;
RUN;
```

Statistics										
Variable	SEX	N	Lower CL		Upper CL		Lower CL		Upper CL	
			Mean	Mean	Mean	Mean	Std Dev	Std Dev	Std Dev	Std Dev
A	F	14	69.045	69.857	70.669	1.0196	1.4064	2.2658	0.3759	
A	M	17	68.783	69.706	70.629	1.3366	1.7946	2.7313	0.4353	
A	Diff (1-2)		-1.053	0.1513	1.3559	1.2998	1.632	2.194	0.589	
B	F	14	158.3	165.57	172.84	9.1246	12.586	20.277	3.3639	
B	M	17	150.47	159.18	167.88	12.607	16.927	25.762	4.1054	
B	Diff (1-2)		-4.777	6.395	17.567	12.054	15.136	20.348	5.4627	

T-Tests					
Variable	Method	Variances	DF	t Value	Pr > t
A	Pooled	Equal	29	0.26	0.7991
A	Satterthwaite	Unequal	28.9	0.26	0.7944
B	Pooled	Equal	29	1.17	0.2513
B	Satterthwaite	Unequal	28.7	1.20	0.2381

Equality of Variances					
Variable	Method	Num DF	Den DF	F Value	Pr > F
A	Folded F	16	13	1.63	0.3804
B	Folded F	16	13	1.81	0.2864

۱۱. مقایسه میانگین ها با استفاده از دستور PROC MEANS

```
DATA A;
INPUT A1 A2 B1 B2;
DIFF1=A1-B1;
DIFF2=A2-B2;
CARDS;
21      7.1      20.5      6.9
.
.
;
PROC MEANS MEAN STDERR T PRT;
VAR DIFF1 DIFF2;
RUN;
```

The MEANS Procedure

Variable	Mean	Std Error	t Value	Pr > t
DIFF1	-1.3125000	0.6473676	-2.03	0.0822
DIFF2	-1.1125000	0.6007250	-1.85	0.1065

بخش دوم

کاربرد SAS در تجزیه و تحلیل طرح های آزمایشات کشاورزی

۱۲. طرح کاملاً تصادفی متعادل

```
DATA;
INPUT REP TREAT$ Y1;
CARDS;
;
PROC ANOVA;
CLASS TREAT;
MODEL Y1=TREAT;
RUN;QUIT;
```

مقایسه میانگین ها با استفاده از نرم افزار SAS

برای مقایسه میانگین تیمارها پس از تعیین MODEL طرح، از دستور MEANS TREAT استفاده می شود. پس از این عبارت علامت / را زده و نوع آزمون را مشخص می نماییم. پس از مشخص نمودن نوع آزمون به دلخواه سطح ALPHA را تعیین می نماییم. لازم به ذکر است که در صورت عدم مشخص نمودن سطح آلفا، برنامه SAS سطح ۵ درصد را منظور خواهد نمود (برای مثال به برنامه شماره ۱۴ توجه نمایید).

```
PROC ANOVA;
CLASS TREAT;
MODEL Y1=TREAT;
MEANS TREAT/LSD ALPHA=0.05;
MEANS TREAT/LSD ALPHA=0.01;
MEANS TREAT/DUNCAN ALPHA=0.05;
MEANS TREAT/DUNCAN ALPHA=0.01;
MEANS TREAT/SNK ALPHA=0.05;
MEANS TREAT/SNK ALPHA=0.01;
MEANS TREAT/TUKEY ALPHA=0.05;
MEANS TREAT/TUKEY ALPHA=0.01;
RUN;QUIT;
```

خروجی SAS در آزمون مقایسه میانگین ها به صورت زیر می باشد:

Alpha	0.05
Error Degrees of Freedom	21
Error Mean Square	۳۵,۶۱۹۰۵
Critical Value of t	2.07961
Least Significant Difference	8.7762
Means with the same letter are not significantly different.	
DUNCAN Grouping	Mean N TREAT
A	26.500 4 A
A	
B	18.250 4 F
B	
B	16.750 4 E
B	
B	15.500 4 C
B	
B	14.250 4 G
B	
B	9.500 4 B
C	
C	5.750 4 D

مقایسات متعامد یا ارتوگونال (Contrasts)

فرض کنید در یک تجزیه واریانس برای اثرات ثابت، فذض یکسان بودن اثر تیمارها رد شد و ما بخواهیم بین تیمارها مقایسه هایی داشته باشیم. این مقایسات فراتر از مقایسات معمول به روش آزمونهای دانکن یا غیره می باشد. در این روش می توان درجات آزادی و مجموع مربعات تیمارها را به اجزای مقایسه های مستقل تفکیک نمود. این اجزا ممکن است مقایسه گروه های تیماری یا رابطه تیمارها با پاسخهای به دست آمده باشند. معمولاً به تعداد درجه آزادی تیمار می توان سؤال مستقل انتخاب و به آنها پاسخ مناسب داد.

فرض کنیم در یک طرح کاملاً تصادفی با ۵ تیمار بخواهیم آزمون زیر را انجام دهیم:

$$H_0 : \mu_1 + \mu_3 = \mu_4 + \mu_5$$

یا اینکه مثلاً فرض زیر را آزمون کنیم:

$$H_0 : \frac{\mu_1 + \mu_3}{2}$$

فرض $H_0 : \mu_1 + \mu_3 = \mu_4 + \mu_5$ ایجاب می کند که:

$$y_{10} + y_{30} - y_{40} - y_{50} = 0$$

در واقع اگر $C = \sum_{i=1}^5 C_i y_{i0}$ باشد، با شرط $\sum_{i=1}^5 C_i = 0$ ، می خواهیم فرض $H_0 : C = 0$ را آزمون کنیم.

به ترکیب خطی $C = \sum_{i=1}^5 C_i y_{i0}$ یک مقایسه **contrast** گفته می شود. این آزمون از طریق به دست آوردن

مجموع مربعات **contrast** یعنی $SSC = \frac{(\sum_{i=1}^a C_i y_{i0})^2}{n \sum_{i=1}^a C_i^2}$ انجام می گردد، که **SSC** دارای یک درجه آزادی

می باشد. چنانچه طرح نامتعادل باشد، رابطه اخیر به صورت $SSC = \frac{(\sum_{i=1}^a C_i y_{i0})^2}{\sum_{i=1}^a n_i C_i^2}$ خواهد شد.

مقایسات متعامد یکی از مهمترین حالت های مقایسه می باشد. اگر در یک طرح متعادل $\sum_{i=1}^a C_i d_i = 0$ یا در یک

طرح نامتعادل $\sum_{i=1}^a n_i C_i d_i = 0$ باشد آنگاه دو مقایسه **contrast** با ضرایب $\{C_i\}$ و $\{d_i\}$ متعامد خواهند بود.

مثلاً در یک طرح کاملاً تصادفی با ۳ تیمار، یک مقایسه متعامد به صورت زیر می باشد:

	d_i	C_i
1	-2	0
2	1	-1
3	1	1

که در $C_i = -2, 1, 1$ تیمار اول را با میانگین تیمار دوم و سوم مقایسه می کند و $d_i = 0, -1, 1$ تیمار دوم و سوم را مقایسه می کند. همانطور که ملاحظه می کنید، مقایسه d_i بیان می کند که آیا چند جمله ای خطی برای تیمارها مناسب است؟ و C_i بیان می کند که آیا چند جمله ای درجه ۲ مناسب است؟ مثلاً برای ۴ تیمار خواهیم داشت:

ضرایب درجه ۳	ضرایب درجه ۲	ضرایب خطی
-1	0	-3
+3	-1	-1
-3	-1	+1
1	+1	3

برای ۵ تیمار این ضرایب به صورت زیر هستند:

ضرایب درجه ۴	ضرایب درجه ۳	ضرایب درجه ۲	ضرایب خطی
+1	-1	2	-2
-4	+2	-1	-1
+6	0	-2	0
-4	-2	-1	1
+1	1	2	2

برای انجام آزمون متعامد از دستور **contrast** به فرم زیر استفاده می شود:

و ضرایب نام اثر "Label(یک اسم)" **Contrast**

مثلاً اگر بخواهیم آزمون $H_0: \mu_3 = \frac{\mu_1 + \mu_2 + \mu_4}{3}$ را انجام دهیم، دستور **Contrast** به صورت زیر خواهد شد:

ضرایب **C**; $-\frac{1}{3} \quad -\frac{1}{3} \quad 1 \quad -\frac{1}{3}$

Contrast " $M_3 = \frac{M_1 + M_2 + M_4}{3}$ " **I** $-\frac{1}{3} \quad -\frac{1}{3} \quad 1 \quad -\frac{1}{3}$;

در مثال فوق برای آزمون چند جمله ای های متعامد به صورت زیر عمل می نمایم:

Contrast " **I Linear** " **I** $-2 \quad -1 \quad 0 \quad 1 \quad 2$; برای چند جمله ای خطی با دستور:

Contrast " **I Quadratic** " **I** $2 \quad -1 \quad -2 \quad -1 \quad 2$; برای چند جمله ای درجه ۲ با دستور:

برای طرح های فاکتوریل نیز می توان از این دستور استفاده کرد. مثلاً طرح تاکتوریل 2^2 را در نظر بگیرید. دو تاکتور **A** و **B** هر کدام در دو سطح (سطوح کمی هستند). در اینجا $a=2$ است و چند جمله ای هایی که می توانیم آزمون کنیم خطی و درجه دو است که ضرایب آن به صورت زیر می باشند:

ضرایب درجه ۲	ضرایب خطی
-2	0
1	-1
1	1

فرض کنیم داده‌ها را نیز وارد کرده باشیم، آنگاه برنامه به صورت زیر نوشته خواهد شد:

```
PROC GLM;
  MODEL Y=A B A*B;
  CONTRAST "A LINEAR" A 0 -1 1;
  CONTRAST "A QUADRATIC" A -2 1 1;
  CONTRAST "B LINEAR" B 0 -1 1;
  CONTRAST "B QUADRATIC" B -2 1 1;
```

برای آزمون خطی بودن اثر A و درجه ۲ بودن اثر B دستور زیر را بکار می‌بریم:

```
CONTRAST "A LINEAR , B QUADRATIC"
  A 0 -1 1 , B -2 1 1;
```

با این دستور SS_{ABL*Q} حساب می‌شود.

```
CONTRAST "A QUADRATIC , B QUADRATIC"
  A 2 1 1 , B -2 1 1;
```

و با دستور فوق SS_{ABQ*Q} را حساب می‌کند.

۱۳. مثال ۱ برای مقایسات متعامد (Orthogonal):

استحکام الیاف مصنوعی به درصد کتان موجود در الیاف بستگی دارد. در یک آزمایش میزان استحکام ۵ سطح درصد کتان (۱۵، ۲۰، ۲۵، ۳۰ و ۳۵ درصد) در ۵ تکرار مورد بررسی قرار گرفت.

```
*\Title "Orthogonal Polynomial;
Data;
Input I$ Y;
CARDS;
15 7
.
35 11
;
PROC GLM;

  CLASS I;
  MODEL Y=I;

  Contrast "I Linear" I -2 -1 0 1 2;
  Contrast "I Quadratic" I 2 -1 -2 -1 2;
  Contrast "I Qubic" I -1 2 0 -2 1;
  Contrast "I Quartic" I 1 -4 6 -4 1;
```

```
RUN;QUIT;
```

```

                                The GLM Procedure
                                Class Level Information
Class      Levels      Values
I          5          15 20 25 30 35
Number of observations      25
```

Dependent Variable: Y

```

                                The GLM Procedure

Source      DF      Sum of Squares      Mean Square      F Value      Pr > F
Model       4      475.7600000      118.9400000      14.76      <.0001
Error      20      161.2000000           8.0600000
Corrected Total      24      636.9600000

R-Square      Coeff Var      Root MSE      Y Mean
0.746923      18.87642      2.839014      15.04000

Source      DF      Type I SS      Mean Square      F Value      Pr > F
I          4      475.7600000      118.9400000      14.76      <.0001
```


Source	DF	Type III SS	Mean Square	F Value	Pr > F
I	4	475.7600000	118.9400000	14.76	<.0001
Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
I Linear	1	33.6200000	33.6200000	4.17	0.0545
I Quadratic	1	343.2142857	343.2142857	42.58	<.0001
I Qubic	1	64.9800000	64.9800000	8.06	0.0101
I Quartic	1	33.9457143	33.9457143	4.21	0.0535

همانگونه که در خروجی آخر ملاحظه می گردد، مقادیر SS ها و همچنین نتیجه آزمون ها مشخص می باشد. در این آزمایش در سطح معنی دار بودن ۵ درصد، خطی بودن و درجه ۴ بودن رد می شود.

۱۴. مثال ۲ برای مقایسات متعامد (Orthogonal):

در یک آزمایش طرح کاملاً تصادفی، اثر مصرف گوگرد در کاهش بیماری گال پوستی سیب زمینی مورد بررسی قرار گرفته است. این آزمایش دارای ۷ تیمار به شرح زیر بوده است:

A: عدم مصرف کود گوگردی، B: به کار گیری ۳۳۶ کیلوگرم گوگرد در پاییز، C: به کار گیری ۶۷۳ کیلوگرم گوگرد در پاییز، D: به کار گیری ۱۳۴۵ کیلوگرم گوگرد در پاییز، E: به کار گیری ۳۳۶ کیلوگرم گوگرد در بهار، F: به کار گیری ۶۷۳ کیلوگرم گوگرد در بهار و G: به کار گیری ۱۳۴۵ کیلوگرم گوگرد در بهار. تعداد تکرار هر تیمار برابر ۴ بوده و صفت مورد اندازه گیری درصد آلودگی در غده های سیب زمینی است. آزمون های زیر را از طریق مقایسات متعامد انجام دهید:

۱. آیا دادن گوگرد در کاهش بیماری تاثیر دارد؟
 ۲. دادن گوگرد در بهار بهتر است یا در پاییز؟
 ۳. آیا مقدار گوگرد کمتر (۳۳۶ کیلوگرم در هکتار) با مقادیر متوسط و زیاد، تاثیر متفاوت داشته است؟
 ۴. آیا به کار گیری گوگرد در مقادیر متوسط و زیاد با هم تفاوت دارند؟
 ۵. آیا بین فصل گوگرددهی و میزان به کار گیری آن اثر متقابل وجود دارد؟
 ۶. آیا بین میزان متوسط مصرف گوگرد و فصل گوگرد دهی اثر متقابل وجود دارد؟
- برنامه مقایسه متعامد برای پاسخ به سؤالات فوق به شرح زیر می باشد:

```
*\Title "Orthogonal Polynomial;
Data;
Input REP TRT$ Y;
CARDS;
;
PROC GLM;

CLASS TRT;
MODEL Y=TRT;
* TRT- ORDER          A B C D E F;
Contrast "QUESTION 1" TRT +6 -1 -1 -1 -1 -1 -1;
Contrast "QUESTION 2" TRT 0 -1 -1 -1 +1 +1 +1;
Contrast "QUESTION 3" TRT 0 -2 +1 +1 -2 +1 +1;
Contrast "QUESTION 4" TRT 0 0 -1 +1 0 -1 +1;
Contrast "QUESTION 5" TRT 0 +2 -1 -1 -2 +1 +1;
Contrast "QUESTION 6" TRT 0 0 +1 -1 0 -1 +1;

RUN;QUIT;
```

The GLM Procedure

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1048.714286	174.785714	4.91	0.0028
Error	21	748.000000	35.619048		
Corrected Total	27	1796.714286			
	R-Square	Coeff Var	Root MSE	Y Mean	
	0.583685	39.22741	5.968170	15.21429	

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TRT	6	1048.714286	174.785714	4.91	0.0028

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRT	6	1048.714286	174.785714	4.91	0.0028

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
QUESTION 1	1	594.3809524	594.3809524	16.69	0.0005
QUESTION 2	1	228.1666667	228.1666667	6.41	0.0194
QUESTION 3	1	0.5208333	0.5208333	0.01	0.9049
QUESTION 4	1	189.0625000	189.0625000	5.31	0.0315
QUESTION 5	1	3.5208333	3.5208333	0.10	0.7563
QUESTION 6	1	33.0625000	33.0625000	0.93	0.3463

همانطور که می بینید پاسخ سؤالات ۱ تا ۶ در خروجی آخر، ستون آخر مشخص می باشد. مثلاً پاسخ سؤال ۱ نشان می دهد که دادن گوگرد در کاهش بیماری تاثیر دارد و یا در خصوص سؤال ۳ نشان می دهد که مقادیر کمتر گوگرد با مقادیر متوسط و زیاد تاثیر متفاوت ندارد. در واقع فقط سؤالات ۱، ۲ و ۴ دارای پاسخ مثبت و معنی دار می باشند.

۱۵. طرح کاملاً تصادفی نامتعادل

```
DATA;
INPUT REP TREAT$ Y1;
CARDS;
;
PROC GLM;
CLASS TREAT;
MODEL Y1=TREAT;
MEANS TREAT/DUNCAN ALPHA=0.01;
RUN;QUIT;
```

۱۶. طرح بلوک های کامل تصادفی

```
data a;
input block treat$ y1-y2;
cards;
;
proc anova;
classes block treat;
model y1-y2=block treat;
means treat/duncan;
run;quit;
```

۱۷. طرح بلوک های کامل تصادفی با داده های از دست رفته (طرح بلوک های کامل تصادفی نامتعادل)

```
data a;
input block treat$ y1-y2;
cards;
;
proc anova;
```

```

classes block treat;
  model y1-y2=block treat;
  means treat/duncan;
run;quit;

```

۱۸. طرح مربع لاتین متعادل

```

data;
input row col treat$ y1;
CARDS;
;
PROC ANOVA;
CLASSES ROW COL TREAT;
MODEL Y1=ROW COL TREAT;
MEAN TREAT/DUNCAN ALPHA=0.01;
RUN;QUIT;

```

Dependent Variable: y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	301782.3333	20118.8222	18.78	<.0001
Error	20	21424.8889	1071.2444		
Corrected Total	35	323207.2222			
R-Square		Coeff Var	Root MSE	y1 Mean	
0.933712		7.150597	32.72987	457.7222	

Source	DF	Anova SS	Mean Square	F Value	Pr > F
row	5	60482.2222	12096.4444	11.29	<.0001
col	5	38918.2222	7783.6444	7.27	0.0005
treat	5	202381.8889	40476.3778	37.78	<.0001

Duncan's Multiple Range Test for y1

Alpha	0.01				
Error Degrees of Freedom	20				
Error Mean Square	1071.244				
Number of Means	2	3	4	5	6
Critical Range	53.76	56.08	57.61	58.73	59.58

Means with the same letter are not significantly different.

Duncan Grouping	Mean	N	treat
A	571.67	6	F
A			
B	520.17	6	E
B			
B	477.83	6	C
B			
D	426.50	6	B
D			
D	405.17	6	D
D			
E	345.00	6	A

۱۹. طرح مربع لاتین با داده های از دست رفته

```

data;
input row col treat$ y1;
CARDS;
;
PROC GLM;
CLASSES ROW COL TREAT;
MODEL Y1=ROW COL TREAT;
MEAN TREAT/DUNCAN ALPHA=0.01;
RUN;QUIT;

```

طرح های مربع لاتین با تکرار مربعات

لازم به یادآوری است که ضعف طرح های مربع لاتین با تیمارهای کم، پایین بودن درجه آزادی اشتباه است. مثلاً در یک طرح مربع لاتین با ۳ تیمار درجه آزادی اشتباه برابر $2=3-1$ می شود. لذا با تکرار مربعات این مشکل برطرف می گردد.

۲۰. طرح مربع لاتین مکرر با مربع های مستقل

در اینگونه طرح ها مربعات دارای توزیع تصادفی جداگانه هستند. لذا منابع تغییرات ما علاوه بر سطرها، ستونها و تیمارها شامل مربع ها نیز می باشد.

```
data;
input SQUAR row col treat$ y1;
CARDS;
;
PROC ANOVA;
CLASSES SQUAR ROW COL TREAT;
MODEL Y1=SQUAR ROW(SQUAR) COL(SQUAR) TREAT;
MEAN TREAT/DUNCAN ALPHA=0.01;
RUN;QUIT;
```

۲۱. طرح مربع لاتین مکرر با مربع های جفت شده

در اینگونه طرح ها دو یا چند مربع در کنار یکدیگر قرار گرفته و تشکیل یک مستطیل را می دهند. لذا هر تیمار یک بار در هر ستون و دو بار یا بیشتر در هر ردیف ظاهر می شود. لذا منابع تغییرات در این طرح ها شامل مربع ها نمی شود.

```
data;
input row col treat$ y1;
CARDS;
;
PROC ANOVA;
CLASSES ROW COL TREAT;
MODEL Y1=ROW COL TREAT;
MEANS TREAT/DUNCAN ALPHA=0.01;
RUN;QUIT;
```

۲۲. طرح مربع یودن (YUDEN SQUARE DESIGN)

هنگامیکه در یک طرح مربع لاتین تعداد سطر و ستون و تیمار برابر نباشند، یک طرح مربع یودن بوجود می آید. به عنوان مثال اگر از یک طرح مربع لاتین یک سطر یا یک ستون را حذف کنیم، مربع یودن به دست می آید. در حالت کلی، یک طرح مربع یودن یک طرح بلوک تصادفی ناقص است که هر تیمار فقط یک بار در سطر یا ستون (موقعیت) ظاهر می شود.

```
DATA;
INPUT I$ J$ T$ Y;
CARDS;
1 1 A 8
1 2 B 6
1 3 C 3
1 4 D 5
2 1 B 5
2 2 C 5
2 3 D 4
```

```

2 4 E 12
3 1 C 4
3 2 D 5
3 3 E 10
3 4 A 3
4 1 D 4
4 2 E 11
4 3 A 9
4 4 B 5
5 1 E 10
5 2 A 7
5 3 B 6
5 4 C 4
;
PROC GLM;
  CLASS I J T;
  MODEL Y=I J T;
RUN;QUIT;

```

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	125.0000000	11.3636364	4.73	0.0181
Error	8	19.2000000	2.4000000		
Corrected Total	19	144.2000000			

	R-Square	Coeff Var	Root MSE	Y Mean
	0.866852	24.59037	1.549193	6.300000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
I	4	9.7000000	2.4250000	1.01	0.4564
J	3	2.6000000	0.8666667	0.36	0.7829
T	4	112.7000000	28.1750000	11.74	0.0020

Source	DF	Type III SS	Mean Square	F Value	Pr > F
I	4	5.7000000	1.4250000	0.59	0.6772
J	3	2.6000000	0.8666667	0.36	0.7829
T	4	112.7000000	28.1750000	11.74	0.0020

چون اثر متقابل نداریم، لذا باید برای تجزیه و تحلیل از نتایج مربوط به جدول Type I SS استفاده نماییم. قاعده کلی در مورد اینکه برای تجزیه و تحلیل باید از کدام Type SS استفاده نمود به شرح زیر می باشد:

باید مجموع SS ها در هر تیپ به اضافه SS اشتباه برابر SST باشد. مثلاً در مثال بالا به صورت زیر عمل می نماییم:

$$SSI + SSJ + SST + SSE = SST$$

$$9.7 + 2.6 + 112.7 + 19.2 = 144.2 = SST \quad (\text{TYPE I SS})$$

$$5.7 + 2.6 + 112.7 + 19.2 = 140.2 \neq SST \quad (\text{TYPE III SS})$$

طرح های چندمشاهده ای

هنگامیکه در یک آزمایش صفتی بیش از یک بار اندازه گیری گردد، در آن صورت داده ها را چند مشاهده ای می گویند. مشاهدات ممکن است: الف- در یک زمان ب- در n زمان مختلف انجام شده باشند. در دو برنامه آتی مشاهدات در یک زمان انجام شده اند.

۲۳. طرح بلوک های کامل تصادفی با داده های چند مشاهده ای:

```

data a;
input block treat$ OBS y1;
cards;
;
proc anova;
  classes block treat OBS;

```

```

model y1=block treat OBS BLOCK*TREAT OBS*TREAT;
TEST H=BLOCK TREAT E=BLOCK*TREAT;
means treat/duncan alpha=0.01 E=BLOCK*TREAT;
run;quit;

```

TEST: در تجزیه واریانس از SAS می خواهد که برای آزمون F یعنی بررسی معنی دار بودن واریانس بلوک و تیمار، از منبع تغییر بلوک × تیمار به عنوان مخرج کسر F استفاده نماید، در صورتیکه از این دستور استفاده نشود، در مخرج کسر F برای آزمون این منابع تغییر، از اشتباه باقیمانده (در این برنامه موسوم به اشتباه نمونه برداری) استفاده خواهد شد که نتیجه صحیحی نخواهد داد.

در گزاره MEANS نیز جمله E=BLOCK*TREAT از SAS می خواهد که برای مقایسه میانگین ها، از میانگین مربعات بلوک × تیمار به عنوان میانگین مربعات اشتباه (MSE) استفاده کند (در واقع برای محاسبه Sd و Sx). خروجی SAS برای برنامه ۲۲ بصورت زیر می باشد:

Dependent Variable: y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	19	5287.033333	278.264912	4.50	0.0094
Error	10	618.333333	61.833333		
Corrected Total	29	5905.366667			
	R-Square	Coeff Var	Root MSE	y1 Mean	
	0.895293	15.97173	7.863417	49.23333	
Source	DF	Anova SS	Mean Square	F Value	Pr > F
block	2	206.466667	103.233333	1.67	0.2368
treat	4	4162.866667	1040.716667	16.83	0.0002
OBS	1	32.033333	32.033333	0.52	0.4882
block*treat	8	768.533333	96.066667	1.55	0.2525
treat*OBS	4	117.133333	29.283333	0.47	0.7545

Tests of Hypotheses Using the Anova MS for block*treat as an Error Term					
Source	DF	Anova SS	Mean Square	F Value	Pr > F
block	2	206.466667	103.233333	1.07	0.3860
treat	4	4162.866667	1040.716667	10.83	0.0026

همانگونه که ملاحظه می گردد، در خروجی برنامه، جدول تجزیه واریانس شامل همه مواردی است که در قسمت MODEL خواسته شده است و همانگونه که ملاحظه می گردد شامل بلوک، تیمار، مشاهده، بلوک × تیمار و مشاهده × تیمار می باشد. (model y1=block treat OBS BLOCK*TREAT OBS*TREAT).

نکته قابل توجه اینکه در خروجی SAS به جز برای تیمارها، برای سایر اثرات متقابل بین منابع تغییرات، فقط مقادیر F محاسبه شده است و مقایسه میانگین ها انجام نگرفته است. این موضوع از خصوصیات برنامه SAS می باشد. یعنی این برنامه به طور همزمان قادر به ارائه مقایسات میانگین برای اثرات متقابل تیمارها و... نیست.

۲۴. طرح کاملاً تصادفی با داده های چند مشاهده ای

```

DATA;
INPUT treat$ rep obs Y1;
CARDS;
;
PROC ANOVA;
CLASSES treat rep obs;
MODEL Y1=treat rep(treat);
TEST H=treat E=rep(treat);
MEANS treat/DUNCAN ALPHA=0.01 E=rep(treat);
RUN;QUIT;

```

TEST: در تجزیه واریانس از SAS می خواهد که در مخرج کسر F در آزمون تیمارها از منبع تغییرات تیمار استفاده نماید.

آزمایشات فاکتوریل بدون اختلاط

۲۵. آزمایشات فاکتوریل در قالب طرح پایه کاملاً تصادفی

```
data a;
input REP N$ P$ K$ Y1-Y3;
cards;
1      N1      P1      K1      7.9      93.75      740.625
.
.
4      N4      P2      K2      11.8      100        1180
;
proc anova;
classes N P K;
model y1-y24=N P K N*P N*K P*K N*P*K;
means N P K N*P N*K P*K N*P*K/duncan alpha=.01;
run;quit;
```

۲۶. آزمایشات فاکتوریل در قالب طرح پایه بلوکهای کامل تصادفی

```
***Data: Mohiseni & Dashadi.DERAKHSHAN 8/5/1381;
data a;
input block N$ P$ K$ Y1-Y7;
cards;
1      N0      P0      K1      61      101      47      16      10      235      44
.
.
.
4      N2      P2      K2      110     19      95      4      10      238      248
;
proc anova;
classes block N P K;
model y1-y7=block N P K N*P N*K P*K N*P*K;
means N P K N*P N*K P*K N*P*K/duncan alpha=.01;
run;quit;
```

۲۷. طرح فاکتوریل نامتعادل

```
DATA;
INPUT I$ J$ Y;
CARDS;
1 V 9
1 V -
.
3 C 4
;
PROC GLM;
CLASS I J;
MODEL Y=I J I*J;
LSMEANS I J I*J/STDRR;
RUN;QUIT;
```

گزاره LSMEANS میانگین ها و STDRR انحراف از میانگین های متغیرهای خواسته شده را چاپ می کند.

توجه: در وارد سازی داده های مربوط به یک طرح نامتعادل با کرت های گم شده، به چند طریق می توان عمل نمود:

- ۱- به جای داده گم شده از علامت - قرار دهیم.
- به جای داده گم شده از علامت ممیز (.) استفاده نمود.
- می توان ردیف های مربوط به داده های گم شده را کلاً حذف نمود.

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	47.52549020	9.50509804	13.88	0.0002
Error	11	7.53333333	0.68484848		
Corrected Total	16	55.05882353			
	R-Square	Coeff Var	Root MSE	Y Mean	
	0.863177	12.23343	0.827556	6.764706	

Source	DF	Type I SS	Mean Square	F Value	Pr > F
I	2	45.80168067	22.90084034	33.44	<.0001
J	1	0.92878465	0.92878465	1.36	0.2688
I*J	2	0.79502488	0.39751244	0.58	0.5759
Source	DF	Type III SS	Mean Square	F Value	Pr > F
I	2	35.88756219	17.94378109	26.20	<.0001
J	1	0.82816901	0.82816901	1.21	0.2950
I*J	2	0.79502488	0.39751244	0.58	0.5759

تذکر مهم اینکه برای تجزیه و تحلیل از نتایج موجود در **Type III SS** استفاده می شود. نتایج حاصل از دستور

LSMEANS I J I*J/STDRR به شرح زیر می باشد:

```

NESTED DESIGN
The GLM Procedure
Least Squares Means
Standard
I      Y LSMEAN      Error      Pr > |t|
1      8.58333333      0.37772578      <.0001
2      7.66666667      0.37772578      <.0001
3      5.05000000      0.34619140      <.0001

Standard
J      Y LSMEAN      Error      Pr > |t|
C      6.86666667      0.25680415      <.0001
V      7.33333333      0.33784821      <.0001

Standard
I      J      Y LSMEAN      Error      Pr > |t|
1      C      8.66666667      0.47778952      <.0001
1      V      8.50000000      0.58517027      <.0001
2      C      7.33333333      0.47778952      <.0001
2      V      8.00000000      0.58517027      <.0001
3      C      4.60000000      0.37009417      <.0001
3      V      5.50000000      0.58517027      <.0001

```

تفکیک SS ها

در آزمایشات فاکتوریل و در طرحهای آزمایشی مختلف، گاهاً می خواهیم بدانیم که آیا مقادیر Y وابستگی خطی یا غیر خطی به فاکتور X نشان می دهد. به عبارت دیگر آیا بین مقادیر X و پاسخهای اندازه گیری شده Y رگرسیون معنی داری وجود دارد یا نه؟ تفکیک SS ها این موضوع را نشان می دهد. برای این منظور می توان با کمک جدول ضرایب چند جمله ای ارتوگونال مجموع مربعات فاکتورها را به اجزای خطی و درجه دوم تفکیک نمود.

توجه: فاکتوری که قرار است مجموع مربعات آن با کمک جدول ضرایب چند جمله ای ارتوگونال به اجزای خطی و درجه دوم تفکیک شوند باید فواصل سطوح آن یکسان باشد.

مثال: در یک آزمایش، راندمان علوفه ای در یک آزمایش فاکتوریل 3×3 در قالب طرح پایه بلوکهای کامل تصادفی بررسی شده است. فاکتور A دارای ۳ سطح (سه وارسته) و فاکتور B دارای سه فاصله ردیف (۱۰، ۲۰، ۳۰ سانتیمتر) بوده اند (ولی زاده و مقدم ۱۳۷۳).

۲۸. تفکیک SS به اجزای خطی و درجه دوم در یک آزمایش فاکتوریل

```
DATA;
INPUT A B REP Y1;
CARDS;
;
PROC GLM;
CLASSES REP A B;
MODEL Y1=REP A B A*B;
CONTRAST "LINEAR" B -1 0 +1;
CONTRAST "QUADRATIC" B +1 -2 +1;
RUN;QUIT;
```

در برنامه فوق دستوره های SAS برای تجزیه واریانس و تفکیک SS فاکتور B نشان داده شده است. در این برنامه همچنانکه مشاهده می گردد، از رویه GLM برای تجزیه واریانس و تفکیک SS فاکتور B استفاده شده است. گزاره CONTRAST اول از SAS می خواهد که تفکیک SS فاکتور B به جزء خطی را انجام دهد. اما دستور دوم CONTRAST موجب تفکیک SS فاکتور B به جزء درجه دوم می شود.

Dependent Variable: Y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	2203.527778	200.320707	11.34	<.0001
Error	24	424.111111	17.671296		
Corrected Total	35	2627.638889			
	R-Square	Coeff Var	Root MSE	Y1 Mean	
	0.838596	7.272180	4.203724	57.80556	
Source	DF	Type III SS	Mean Square	F Value	Pr > F
REP	3	255.638889	85.212963	4.82	0.0091
A	2	1027.388889	513.694444	29.07	<.0001
B	2	155.055556	77.527778	4.39	0.0238
A*B	4	765.444444	191.361111	10.83	<.0001
Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
LINEAR	1	155.0416667	155.0416667	8.77	0.0068
QUADRATIC	1	0.0138889	0.0138889	0.00	0.9779

در آزمایش فوق همچنین می توان رگرسیون فاصله ردیف و راندمان علوفه را برای هر یک از وارسته ها نیز محاسبه و آزمون کرد (برنامه ۲۷).

۲۹. تفکیک SS برای سطوح فاکتور B در داخل سطوح فاکتور A

```
DATA;
INPUT A B BLOCK Y1;
CARDS;
;
PROC GLM;
CLASSES BLOCK B;
MODEL Y1=BLOCK B;
CONTRAST "LINEARB" B -1 0 +1;
CONTRAST "QUADRATICB" B +1 -2 +1;
RUN;QUIT;
```

خروجی SAS به صورت زیر خواهد بود:

Dependent Variable: Y1 (برای واریته اول)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	519.5000000	103.9000000	32.53	0.0003
Error	6	19.1666667	3.1944444		
Corrected Total	11	538.6666667			
	R-Square	Coeff Var	Root MSE	Y1 Mean	
	0.964418	3.481755	1.787301	51.33333	

Source	DF	Type III SS	Mean Square	F Value	Pr > F
BLOCK	3	381.3333333	127.1111111	39.79	0.0002
B	2	138.1666667	69.0833333	21.63	0.0018
Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
LINEARB	1	136.1250000	136.1250000	42.61	0.0006
QUADRATICB	1	2.0416667	2.0416667	0.64	0.4545

Dependent Variable: Y1 (برای واریته دوم)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	248.1666667	49.6333333	35.04	0.0002
Error	6	8.5000000	1.4166667		
Corrected Total	11	256.6666667			
	R-Square	Coeff Var	Root MSE	Y1 Mean	
	0.966883	2.063997	1.190238	57.66667	

Source	DF	Type III SS	Mean Square	F Value	Pr > F
BLOCK	3	44.0000000	14.6666667	10.35	0.0087
B	2	204.1666667	102.0833333	72.06	<.0001
Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
LINEARB	1	200.0000000	200.0000000	141.18	<.0001
QUADRATICB	1	4.1666667	4.1666667	2.94	0.1372

Dependent Variable: Y1 (برای واریته سوم)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	711.0833333	142.2166667	9.09	0.0091
Error	6	93.8333333	15.6388889		
Corrected Total	11	804.9166667			
	R-Square	Coeff Var	Root MSE	Y1 Mean	
	0.883425	6.139100	3.954604	64.41667	

Source	DF	Type III SS	Mean Square	F Value	Pr > F
BLOCK	3	132.9166667	44.3055556	2.83	0.1285
B	2	578.1666667	289.0833333	18.48	0.0027
Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
LINEARB	1	578.0000000	578.0000000	36.96	0.0009
QUADRATICB	1	0.1666667	0.1666667	0.01	0.9211

همانگونه که ملاحظه می گردد، هر یک از سطوح A یعنی واریته ها به صورت یک طرح بلوکهای کامل تصادفی تجزیه شده اند و مقایسه سطوح B در داخل هر کدام از واریته ها انجام شده است. گزاره BY از SAS می خواهد که برای هر یک از سطوح A، داده ها را در قالب یک طرح بلوکهای کامل تصادفی تجزیه کند و مقایسه ها (تفکیک SSها) را انجام دهد. نکته اینکه در هنگام استفاده از گزاره BY باید دقت نمود که سطوح فاکتور A نیز باید با عدد نوشته شوند، وگرنه برای SAS قابل شناسایی نیستند.

آزمایشات فاکتوریل با اختلاط

در آزمایشات فاکتوریل، با افزایش تعداد فاکتور و تعداد سطوح آن، تعداد تیمارها به سرعت افزایش می یابد. بنابراین در یک بلوک تعداد واحدهای آزمایشی افزایش یافته و در نتیجه بلوک یکنواختی خود را از دست می دهد و اشتباه آزمایش بزرگ می شود. برای رفع این مشکل می توان از روش اختلاط که توسط فیشر ابداع شده است، استفاده نمود.

در این روش هر بلوک به دو یا چند بلوک کوچکتر تقسیم می شود. به این ترتیب هر بلوک کوچک تعدادی از تیمارها را در خود جای می دهد. اختلاط بر دو نوع است. اختلاط کامل و ناقص. در اختلاط کامل توزیع تیمارها در بین دو بلوک در همه تکرارها یکسان می باشد. به عبارت دیگر اثر یا اثر متقابل که اختلاط می یابد برای همه تکرارها یکسان است. فرض کنید در یک آزمایش فاکتوریل 2^3 سه سطح ABC اختلاط یافته باشند (ولی زاده و مقدم ۱۳۷۳). تجزیه و تحلیل این نوع طرح ها در برنامه شماره ۲۸ آمده است. در اختلاط ناقص در بعضی از تکرارها یک اثر اختلاط یافته و در تکرارهای دیگر، اثرات دیگری اختلاط پیدا می کنند.

۳۰. آزمایشات فاکتوریل با اختلاط کامل

داده های مربوط به ولی زاده و مقدم ۱۳۷۳ که در صفحه ۶۹ کتاب سلطانی ۱۳۷۷ آمده است، به شرح صفحه بعد می باشد:

b	abc	c	a	شاهد	bc	ac	ac	تکرار اول
3	9	6	5	2	2	6	8	
شاهد	ab	ac	bc	a	c	b	abc	تکرار دوم
2	8	5	6	6	7	3	10	
ac	شاهد	bc	ab	a	b	c	abc	تکرار سوم
6	3	4	9	7	5	4	10	
abc	b	a	c	bc	ac	شاهد	ab	تکرار چهارم
11	4	8	5	5	7	2	8	

```
DATA;
INPUT BLOCK A$ B$ C$ Y1;
CARDS;
1 A1 B2 C1 3
1 A2 B2 C2 9
1 A1 B1 C2 6
1 A2 B1 C1 5
2 A1 B1 C1 2
2 A1 B2 C2 2
2 A2 B1 C2 6
2 A2 B2 C1 8
3 A1 B1 C1 2
3 A2 B2 C1 8
3 A2 B1 C2 5
3 A1 B2 C2 6
4 A2 B1 C1 6
4 A1 B1 C2 7
4 A1 B2 C1 3
4 A2 B2 C2 10
5 A2 B1 C2 6
5 A1 B1 C1 3
5 A1 B2 C2 4
5 A2 B2 C1 9
6 A2 B1 C1 7
6 A1 B2 C1 5
6 A1 B1 C2 4
6 A2 B2 C2 10
7 A2 B2 C2 11
```

```

7      A1      B2      C1      4
7      A2      B1      C1      8
7      A1      B1      C2      5
8      A1      B2      C2      5
8      A2      B1      C2      7
8      A1      B1      C1      2
8      A2      B2      C1      8
;
PROC GLM;
CLASSES BLOCK A B C;
MODEL Y1=BLOCK A B C A*B A*C B*C;
RUN;QUIT;

```

همانطور که ملاحظه می گردد، در این برنامه نیز از رویه GLM استفاده شده است. نکته قابل توجه در برنامه فوق این است که در قسمت گزاره MODEL اثرات متقابل A*B*C نوشته نشده است. به عبارت دیگر در اختلاط کامل باید از نوشتن نام اثری که اختلاط یافته است خودداری نمود. خروجی SAS در زیر آمده است:

Dependent Variable: Y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	179.7500000	13.8269231	11.78	<.0001
Error	18	21.1250000	1.1736111		
Corrected Total	31	200.8750000			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
BLOCK	7	18.3750000	2.6250000	2.24	0.0800
A	1	112.5000000	112.5000000	95.86	<.0001
B	1	18.0000000	18.0000000	15.34	0.0010
C	1	12.5000000	12.5000000	10.65	0.0043
A*B	1	15.1250000	15.1250000	12.89	0.0021
A*C	1	3.1250000	3.1250000	2.66	0.1201
B*C	1	0.1250000	0.1250000	0.11	0.7479

۳۱. آزمایشات فاکتوریل با اختلاط ناقص

داده های زیر مربوط به ولی زاده و مقدم ۱۳۷۳ می باشد که در صفحه ۷۰ کتاب سلطانی ۱۳۷۷ به شرح زیر آمده است:

P	PK	NK	N	شامد	K	NP	NPK	تکرار اول
66	69	64	69	67	71	61	71	NP
NP	P	NK	K	NPK	N	PK	شامد	تکرار دوم
72	82	73	81	77	79	86	69	PK
N	PK	NP	K	P	NPK	NK	شامد	تکرار سوم
77	90	71	86	96	95	81	79	NK
N	P	K	NPK	شامد	NP	PK	NK	تکرار چهارم
67	75	64	82	64	62	70	65	NPK

```

DATA;
INPUT BLOCK N$ P$ K$ Y1;
CARDS;
1      N1      P2      K1      66
1      N1      P2      K2      69
1      N2      P1      K2      64
1      N2      P1      K1      69
2      N1      P1      K1      67
2      N1      P1      K2      71
2      N2      P2      K1      61
2      N2      P2      K2      71
3      N2      P2      K1      72
3      N1      P2      K1      82

```

```

3      N2      P1      K2      73
3      N1      P1      K2      81
4      N2      P2      K2      77
4      N2      P1      K1      79
4      N1      P2      K2      86
4      N1      P1      K1      69
5      N2      P1      K1      77
5      N1      P2      K2      90
5      N2      P2      K1      71
5      N1      P1      K2      86
6      N1      P2      K1      96
6      N2      P2      K2      95
6      N2      P1      K2      81
6      N1      P1      K1      79
7      N2      P1      K1      67
7      N1      P2      K1      75
7      N1      P1      K2      64
7      N2      P2      K2      82
8      N1      P1      K1      64
8      N2      P2      K1      62
8      N1      P2      K2      70
8      N2      P1      K2      65

```

```

;
PROC GLM;
CLASSES BLOCK N P K;
MODEL Y1=BLOCK N P K N*P N*K P*K N*P*K;
RUN;QUIT;

```

در این طرح نیز از رویه GLM استفاده شده است. اما تفاوت آن با طرح اختلاط کامل این است که چون در تمامی تکرارها یک اثر خاص اختلاط نیافته است، لذا تمامی اثرات ساده و متقابل قابل برآورد هستند. بنابراین در قسمت دوم برنامه در جلوی گزاره MODEL نام تمامی این اثرات به استثنای باقیمانده (اشتباه) ذکر می گردد.

Dependent Variable: Y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	14	2415.937500	172.566964	9.12	<.0001
Error	17	321.781250	18.928309		
Corrected Total	31	2737.718750			
	R-Square	Coeff Var	Root MSE	Y1 Mean	
	0.882464	5.847180	4.350668	74.40625	

Source	DF	Type I SS	Mean Square	F Value	Pr > F
BLOCK	7	1726.468750	246.638393	13.03	<.0001
N	1	75.031250	75.031250	3.96	0.0628
P	1	148.781250	148.781250	7.86	0.0122
K	1	148.781250	148.781250	7.86	0.0122
N*P	1	63.375000	63.375000	3.35	0.0849
N*K	1	0.666667	0.666667	0.04	0.8534
P*K	1	60.166667	60.166667	3.18	0.0925
N*P*K	1	192.666667	192.666667	10.18	0.0054

طرحهای متداخل (NESTED)

اگر در یک آزمایش فاکتوریل سطوح یک تیمار در داخل سطوح مختلف تیمار دیگر آزمایش نشود، یک طرح متداخل بوجود می آید که لزوماً سطوح عامل B در داخل عامل A ممکن است یکسان نباشد. اگر سطوح عامل B در داخل عامل A یکسان و ثابت باشند، یک طرح متداخل متعادل داریم.

۳۲. یک طرح متداخل (NESTED DESIGN) دارای اثرات ثابت و تصادفی

مثال: قرار است چهار ماشین در یک کارخانه مورد آزمایش قرار گیرند. با هر یک از این ماشین‌ها ۳ کارگر کار می‌کنند و دو قطعه ساخته شده از هر کدام انتخاب می‌شود. (توجه اینکه کارگرانی که روی ماشین‌ها کار می‌کنند متفاوتند).

```
Title "NESTED DESIGN";
DATA;
INPUT A$ B$ Y;
CARDS;
1 1 79
1 1 62
2 1 94
2 1 74
3 1 46
3 1 57
1 2 92
1 2 99
2 2 85
2 2 79
3 2 76
3 2 68
1 4 36
1 4 53
2 4 40
2 4 56
3 4 62
3 4 47
;
PROC ANOVA;
  CLASSES A B;
  MODEL Y=A B(A);
  TEST H=A E=B(A);
RUN;QUIT;
```

طرح فوق یک طرح متداخل دو مرحله‌ای است. همانطور که دیده می‌شود، اثر متداخل B در داخل A را با B(A) نشان می‌دهیم. باید توجه داشت که ماشین‌ها با اثر ثابت و کارگران با اثر تصادفی هستند. لذا برای آزمون معنی دار بودن اثر ماشین‌ها (B)، در کسر F از MSE (۹۴/۰۰۰) و برای آزمون F برای کارگران (A) از $MSB(A) / (78/17)$ استفاده شده است. این مسئله با استفاده از دستور TEST به اجرا در آمده است.

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	5254.444444	656.805556	6.84	0.0046
Error	9	864.500000	96.055556		
Corrected Total	17	6118.944444			

R-Square	Coeff Var	Root MSE	Y Mean
0.858717	14.64019	9.800794	66.94444

Source	DF	Anova SS	Mean Square	F Value	Pr > F
A	2	525.444444	262.722222	2.74	0.1180
B(A)	6	4729.000000	788.166667	8.21	0.0031

Tests of Hypotheses Using the Anova MS for B(A) as an Error Term					
Source	DF	Anova SS	Mean Square	F Value	Pr > F
A	2	525.444444	262.722222	0.33	0.7290

۳۳. طرح کرت های خرد شده SPLIT PLOT

```

data;
input a$ b$ block y1;
cards;
;
proc anova;
classes block a b;
model y1=block a block*a b a*b;
test h=a e=block*a;
means a/duncan alpha=0.01 e=block*a;
means b/duncan alpha=0.01;
means a*b;
run;quit;

```

در این طرح از رویه ANOVA استفاده می شود. در جلوی گزاره CLASSES متغیرهای طبقه بندی داده ها یعنی بلوک، فاکتور A و فاکتور B معرفی می شوند. گزاره MODEL شامل مدل تجزیه واریانس طرح است که در جلوی آن منابع تغییرات طرح به استثنای باقیمانده (اشتباه) ذکر می گردند.

گزاره TEST از SAS می خواهد که برای آزمون F منبع تغییر A، از منبع تغییر BLOCK*A (یعنی اثر متقابل بلوک در فاکتور A) به عنوان مخرج کسر F استفاده نماید (test h=a e=block*a).

در این برنامه سه بار از گزاره MEANS استفاده شده است. گزاره اول MEAN از SAS می خواهد که میانگینهای سطوح مختلف فاکتور A را با استفاده از آزمون دانکن در سطح احتمال ۱ درصد آزمون کند. اما در ادامه همین گزاره نشان داده شده است که برای محاسبه Sd و Sx و مقایسه میانگین، از اثرات متقابل بلوک $A \times B$ به عنوان اشتباه استفاده کند (e=block*a). گزاره دوم MEAN از SAS می خواهد که میانگینهای سطوح B را با استفاده از آزمون دانکن در سطح احتمال ۱ درصد با یکدیگر مقایسه کند. در این گزاره از کلمه اختیاری E=EFFECT استفاده نشده است. بنابراین برای محاسبه Sd و Sx از میانگین مربعات باقیمانده (اشتباه B) استفاده خواهد شد. گزاره MEAN سوم باعث چاپ میانگینهای اثرات متقابل A*B می شود. جدول تجزیه واریانس در خروجی SAS برای برنامه فوق بصورت زیر می باشد:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	313.2216667	15.6610833	35.52	<.0001
Error	27	11.9050000	0.4409259		
Corrected Total	47	325.1266667			

R-Square	Coeff Var	Root MSE	y1 Mean
0.963384	1.801146	0.664023	36.86667

Source	DF	Anova SS	Mean Square	F Value	Pr > F
block	3	2.3883333	0.7961111	1.81	0.1700
a	2	20.6216667	10.3108333	23.38	<.0001
block*a	6	1.5916667	0.2652778	0.60	0.7265
b	3	45.5450000	15.1816667	34.43	<.0001
a*b	6	243.0750000	40.5125000	91.88	<.0001

Tests of Hypotheses Using the Anova MS for block*a as an Error Term

Source	DF	Anova SS	Mean Square	F Value	Pr > F
a	2	20.62166667	10.31083333	38.87	0.0004

در جدول تجزیه واریانس فوق همچنانکه ملاحظه می گردد، میانگین مربعات اشتباه A (0.2652778) کوچک تر از مقدار مربوط به اشتباه B (باقیمانده) (0.4409259) است. این موضوع نشان می دهد که امکان وجود اثر متقابل معنی دار بین فاکتور B و بلوک وجود دارد. بنابراین باید اثر متقابل فاکتور B × بلوک از مجموع مربعات باقیمانده جدا شود. برای این منظور در انتهای گزاره MODEL، باید نام این اثر متقابل (BLOCK*B) را اضافه نموده و مجدداً تجزیه را انجام دهیم.

```
proc anova;
classes block a b;
model y1=block a block*a b a*b block*b;
test h=a e=block*a;
means a/duncan alpha=0.01 e=block*a;
means b/duncan alpha=0.01;
means a*b;
run;quit;
```

Dependent Variable: y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	29	320.6283333	11.0561494	44.24	<.0001
Error	18	4.4983333	0.2499074		
Corrected Total	47	325.1266667			

R-Square	Coeff Var	Root MSE	y1 Mean
0.986164	1.355988	0.499907	36.86667

Source	DF	Anova SS	Mean Square	F Value	Pr > F
block	3	2.3883333	0.7961111	3.19	0.0489
a	2	20.6216667	10.3108333	41.26	<.0001
block*a	6	1.5916667	0.2652778	1.06	0.4206
b	3	45.5450000	15.1816667	60.75	<.0001
a*b	6	243.0750000	40.5125000	162.11	<.0001
block*b	9	7.4066667	0.8229630	3.29	0.0150

Tests of Hypotheses Using the Anova MS for block*a as an Error Term

Source	DF	Anova SS	Mean Square	F Value	Pr > F
a	2	20.62166667	10.31083333	38.87	0.0004

چنانچه ملاحظه می گردد، اثر متقابل بلوک × B وجود دارد و در سطح ۵ درصد معنی دار است.

۳۴. طرح کورت های خرد شده در زمان

به آزمایشاتی گفته می شود که بیش از یک مشاهده بر روی یک صفت خاص، در زمانهای مختلف از هر کورت انجام گیرد.

```
data;
input b a$ block y1;
cards;
;
proc anova;
classes block a b;
```



```

model y1=block a block*a b a*b block*b;
test h=a e=block*a;
means a/duncan alpha=0.01 e=block*a;
means b/duncan alpha=0.01;
means a*b;
run;quit;

```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	22	482.5000000	21.9318182	30.96	<.0001
Error	9	6.3750000	0.7083333		
Corrected Total	31	488.8750000			

Source	DF	Anova SS	Mean Square	F Value	Pr > F
block	3	28.3750000	9.4583333	13.35	0.0012
a	3	185.1250000	61.7083333	87.12	<.0001
block*a	9	21.3750000	2.3750000	3.35	0.0430
b	1	15.1250000	15.1250000	21.35	0.0013
a*b	3	205.1250000	68.3750000	96.53	<.0001
block*b	3	27.3750000	9.1250000	12.88	0.0013

Tests of Hypotheses Using the Anova MS for block*a as an Error Term

Source	DF	Anova SS	Mean Square	F Value	Pr > F
a	3	185.1250000	61.7083333	25.98	<.0001

۳۵. تجزیه جداگانه در یک طرح کرفتهای خرد شده در زمان برای سطوح فاکتور A

در صورتیکه بخواهیم برای هر تاریخ برداشت، یا مثلاً برای هر نوبت نمونه گیری از یک آفت یا بیماری، یک تجزیه واریانس جداگانه در قالب طرح بلوکهای کامل تصادفی انجام دهیم، می توان با همان داده های فوق، و با استفاده از گزاره BY این تجزیه و تحلیل ها را انجام داد.

```

data;
input b a$ block y1;
cards;
;
proc sort;by b;
proc anova;
classes block a;
model y1=block a;
means a/duncan alpha=0.01;
by b;
run;quit;

```

----- b=1 -----
Dependent Variable: y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	24.5000000	4.0833333	5.07	0.0154
Error	9	7.2500000	0.8055556		
Corrected Total	15	31.7500000			

Source	DF	Anova SS	Mean Square	F Value	Pr > F
block	3	23.2500000	7.7500000	9.62	0.0036
a	3	1.2500000	0.4166667	0.52	0.6808

Alpha	0.01
Error Degrees of Freedom	9
Error Mean Square	0.805556
Number of Means	2 3 4
Critical Range	2.062 2.148 2.201

Duncan Grouping	Mean	N	a
A	22.7500	4	a2
A			
A	22.5000	4	a1

Class	Levels	Values
block	4	1 2 3 4
a	4	a1 a2 a3 a4

Dependent Variable: y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	421.5000000	70.2500000	30.84	<.0001
Error	9	20.5000000	2.2777778		
Corrected Total	15	442.0000000			

R-Square	Coeff Var	Root MSE	y1 Mean
0.953620	7.186814	1.509231	21.00000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
block	3	32.5000000	10.8333333	4.76	0.0297
a	3	389.0000000	129.6666667	56.93	<.0001

Alpha 0.01
Error Degrees of Freedom 9
Error Mean Square 2.277778

Number of Means	2	3	4
Critical Range	3.468	3.612	3.702

Duncan Grouping	Mean	N	a
A	25.250	4	a4
A	24.250	4	a3
A	21.750	4	a2
B	12.750	4	a1

۳۶. طرح کرت های دو بار خرد شده (Split-Split Plot)

```
data;
input a b c block y1;
cards;
;
proc anova;
classes block a b c;
model y1= block a block*a b a*b block*b(a) c a*c b*c a*b*c;
test h=a e=block*a;
test h=b e=block*b(a);
means a/duncan alpha=0.01 e=block*a;
means b/duncan alpha=0.01 e=block*b(a);
means c/duncan alpha=0.01;
run;quit;
```

در برنامه فوق a، b و c به ترتیب سطوح فاکتورهای a، b و c را نشان می دهند.

Dependent Variable: y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	23	347.3125000	15.1005435	58.77	<.0001
Error	24	6.1666667	0.2569444		
Corrected Total	47	353.4791667			

R-Square	Coeff Var	Root MSE	y1 Mean
0.982554	6.853817	0.506897	7.395833

Source	DF	Anova SS	Mean Square	F Value	Pr > F
block	2	0.6666667	0.3333333	1.30	0.2918
a	1	143.5208333	143.5208333	558.57	<.0001
block*a	2	2.1666667	1.0833333	4.22	0.0270
b	1	38.5208333	38.5208333	149.92	<.0001
a*b	1	6.0208333	6.0208333	23.43	<.0001
block*b(a)	4	0.3333333	0.0833333	0.32	0.8589
c	3	116.2291667	38.7430556	150.78	<.0001
a*c	3	27.8958333	9.2986111	36.19	<.0001
b*c	3	10.2291667	3.4097222	13.27	<.0001
a*b*c	3	1.7291667	0.5763889	2.24	0.1091

Tests of Hypotheses Using the Anova MS for block*a as an Error Term

Source	DF	Anova SS	Mean Square	F Value	Pr > F
a	1	143.5208333	143.5208333	132.48	0.0075

Tests of Hypotheses Using the Anova MS for block*b(a) as an Error Term

Source	DF	Anova SS	Mean Square	F Value	Pr > F
b	1	38.52083333	38.52083333	462.25	<.0001
a*b	1	6.02083333	6.02083333	72.25	0.0011

۳۷. طرح کرت های خرد شده فاکتوریل (Split-Plot factorial)

در این طرح با اینکه سه فاکتور به کار می رود، اما طرح مورد نظر کرت های خرد شده می باشد (نه کرت های دو بار خرد شده). در این طرح یک فاکتور به عنوان فاکتور اصلی در کرت های اصلی پیاده می شود و دو فاکتور دیگر به صورت فاکتوریل در کرت های فرعی پیاده می شوند. فرض کنیم در برنامه شماره ۳۳ (برنامه قبل)، فاکتور A میزان آب (با دو سطح دیم و آبیاری) باشد و دو فاکتور دیگر به صورت فاکتوریل تیمارهای فرعی را تشکیل می دهند (بصیری ۱۳۷۰).

```
data;
input a b c block y1;
cards;
;
proc anova;
classes block a b c;
model y1=block a block*a b c a*b a*c b*c a*b*c;
test h=a e=block*a;
means a/duncan alpha=0.01 e=block*a;
means b/duncan alpha=0.01;
means c/duncan alpha=0.01;
means a*b a*c b*c a*b*c;
run;quit;
```

Dependent Variable: y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	19	346.9791667	18.2620614	78.67	<.0001
Error	28	6.5000000	0.2321429		
Corrected Total	47	353.4791667			

R-Square 0.981611 Coeff Var 6.514642 Root MSE 0.481812 y1 Mean 7.395833

Source	DF	Anova SS	Mean Square	F Value	Pr > F
block	2	0.6666667	0.3333333	1.44	0.2549
a	1	143.5208333	143.5208333	618.24	<.0001
block*a	2	2.1666667	1.0833333	4.67	0.0178
b	1	38.5208333	38.5208333	165.94	<.0001
c	3	116.2291667	38.7430556	166.89	<.0001
a*b	1	6.0208333	6.0208333	25.94	<.0001
a*c	3	27.8958333	9.2986111	40.06	<.0001
b*c	3	10.2291667	3.4097222	14.69	<.0001
a*b*c	3	1.7291667	0.5763889	2.48	0.0814

Tests of Hypotheses Using the Anova MS for block*a as an Error Term

Source	DF	Anova SS	Mean Square	F Value	Pr > F
a	1	143.5208333	143.5208333	132.48	0.0075

۳۸. طرح فاکتوریل کرت های خرد شده (Factorial Split-plot)

در این طرح دو فاکتور به عنوان تیمارهای اصلی در کرت های بزرگ پیدا می شوند و فاکتور سوم به عنوان تیمار فرعی در کرت های کوچک پیاده می شود. البته قالب طرح همان طرح کرت های خرد شده می باشد. فرض کنیم در همان مثال

قبل فاکتورهای A و B به صورت فاکتوریل در کترهای اصلی پیاده شده اند و فاکتور C به عنوان تیمار فرعی قرار گرفته است (بصیری ۱۳۷۰).

```
data;
input a b c block y1;
cards;
;
proc anova;
classes block a b c;
model y1=block a b a*b block(a b) c a*c b*c a*b*c;
test h=a e=block(a b);
means a/duncan alpha=0.01 e=block(a b);
means b/duncan alpha=0.01 e=block(a b);
means c/duncan alpha=0.01;
means a*b a*c b*c a*b*c;
run;quit;
```

یادآوری مجدد: یک مشکل اساسی که در برنامه فوق وجود دارد آن را پیدا کنید؟ block(a b) نشان دهنده اثر بلوک داخل A و B می باشد. همانگونه که ملاحظه می گردد، برای محاسبه F مربوط به A، B و AB باید در کسر از اثر بلوک داخل A و B یعنی block(a b) استفاده نمود. لذا برنامه به صورت زیر تصحیح می گردد.

```
proc anova;
classes block a b c;
model y1=block a b a*b block(a b) c a*c b*c a*b*c;
test h=a e=block(a b);
test h=b e=block(a b);
test h=a*b e=block(a b);
means a/duncan alpha=0.01 e=block(a b);
means b/duncan alpha=0.01 e=block(a b);
means c/duncan alpha=0.01;
means a*b a*c b*c a*b*c;
run;quit;
```

خروجی SAS برای برنامه فوق به صورت زیر است:

Dependent Variable: y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	23	347.3125000	15.1005435	58.77	<.0001
Error	24	6.1666667	0.2569444		
Corrected Total	47	353.4791667			
	R-Square	Coeff Var	Root MSE	y1 Mean	
	0.982554	6.853817	0.506897	7.395833	
Source	DF	Anova SS	Mean Square	F Value	Pr > F
block	2	0.6666667	0.3333333	1.30	0.2918
a	1	143.5208333	143.5208333	558.57	<.0001
b	1	38.5208333	38.5208333	149.92	<.0001
a*b	1	6.0208333	6.0208333	23.43	<.0001
block(a*b)	6	2.5000000	0.4166667	1.62	0.1844
c	3	116.2291667	38.7430556	150.78	<.0001
a*c	3	27.8958333	9.2986111	36.19	<.0001
b*c	3	10.2291667	3.4097222	13.27	<.0001
a*b*c	3	1.7291667	0.5763889	2.24	0.1091
Tests of Hypotheses Using the Anova MS for block(a*b) as an Error Term					
Source	DF	Anova SS	Mean Square	F Value	Pr > F
a	1	143.5208333	143.5208333	344.45	<.0001
b	1	38.5208333	38.5208333	92.45	<.0001
a*b	1	6.0208333	6.0208333	14.45	0.0090

۳۹. طرح کرت های نواری (Strip- Plot)

طرح کرت های نواری برای آن دسته از آزمایش های دو عاملی مناسب است که دقت مورد نظر برای اندازه گیری اثر متقابل بین دو عامل بیشتر از دقت انداز رزه گیری هر یک از اثرات اصلی عاملها باشد. این طرح در سه اندازه کرت مورد استفاده قرار می گیرد:

۱. کرت نواری عمودی (vertical- Strip plot): برای عامل اول- فاکتور عمودی.

۲. کرت نواری افقی (Horizontal- Strip plot): برای عامل دوم- فاکتور افقی.

۳. کرت متقاطع یا مشترک (Intersection plot): برای اثر متقابل بین دو عامل.

کرت های نواری عمودی و افقی همیشه بر هم عمود هستند، اما بر خلاف کرت های اصلی و کرت های فرعی، در طرح کرت های خرد شده، رابطه ای بین اندازه نوارهای عمودی و افقی وجود ندارد. کرت متقاطع کوچکتر از همه است، بنابراین در یک طرح کرت های نواری، دقت اثرات اصلی عوامل کمتر از دقت اثری است که در اثر متقابل وجود دارد.

```
data wheat;
  input cultivar $ row block yield;
cards;
a      1      1      6
a      2      1      8
a      3      1      8
a      1      2      6
a      2      2      7
a      3      2      8
a      1      3      8
a      2      3      8
a      3      3     10
b      1      1     10
b      2      1     12
b      3      1     15
b      1      2     10
b      2      2     11
b      3      2     14
b      1      3     10
b      2      3     13
b      3      3     15
c      1      1     10
c      2      1     15
c      3      1     12
c      1      2     10
c      2      2     14
c      3      2     10
c      1      3     12
c      2      3     16
c      3      3     12
d      1      1      8
d      2      1     12
d      3      1     10
d      1      2      7
d      2      2     10
d      3      2      8
d      1      3      9
d      2      3     12
d      3      3     11
;
```

```
proc glm data=wheat;
```

```

class cultivar row block;
model yield = block cultivar row cultivar*row block*cultivar row*block;
test H=block cultivar E=block*cultivar;
test H=row E=block*row;
means cultivarN /duncan e=block*cultivar;
means row/duncan e=row*block;
means row*cultivar;
run;

```

The GLM Procedure
Class Level Information

Class	Levels	Values
cultivar	4	a b c d
row	3	1 2 3
block	3	1 2 3
Number of observations		36

The GLM Procedure

Dependent Variable: yield

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	23	245.9722222	10.6944444	42.78	<.0001
Error	12	3.0000000	0.2500000		
Corrected Total	35	248.9722222			

R-Square	Coeff Var	Root MSE	yield Mean
0.987950	4.774536	0.500000	10.47222

Source	DF	Type I SS	Mean Square	F Value	Pr > F
block	2	18.3888889	9.1944444	36.78	<.0001
cultivar	3	135.4166667	45.1388889	180.56	<.0001
row	2	49.3888889	24.6944444	98.78	<.0001
cultivar*row	6	38.8333333	6.4722222	25.89	<.0001
cultivar*block	6	2.5000000	0.4166667	1.67	0.2124
row*block	4	1.4444444	0.3611111	1.44	0.2790

Source	DF	Type III SS	Mean Square	F Value	Pr > F
block	2	18.3888889	9.1944444	36.78	<.0001
cultivar	3	135.4166667	45.1388889	180.56	<.0001
row	2	49.3888889	24.6944444	98.78	<.0001
cultivar*row	6	38.8333333	6.4722222	25.89	<.0001
cultivar*block	6	2.5000000	0.4166667	1.67	0.2124
row*block	4	1.4444444	0.3611111	1.44	0.2790

Tests of Hypotheses Using the Type III MS for cultivar*block as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
block	2	18.3888889	9.1944444	22.07	0.0017
cultivar	3	135.4166667	45.1388889	108.33	<.0001

Tests of Hypotheses Using the Type III MS for row*block as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
row	2	49.3888889	24.6944444	68.38	0.0008

۴۰. طرح کرت های خرد شده نواری (Strip-Split-Block)

مواقع استفاده از این طرح به شرح زیر است:

- ۱- یا هر یک از فاکتورهای موجود در آزمایش نیاز به کرت های بزرگ داشته باشند. یا اینکه ۲- اطلاعات مربوط به خود فاکتورها برای محقق کم اهمیت بوده در صورتیکه اثرات متقابل آنها بسیار مهم می باشد. در این طرح نیز یک

فاکتور را به عنوان تیمار اصلی و فاکتور دیگر را تیمار فرعی می نامند. نکته اینکه در این طرح هر یک از فاکتورها و اثر متقابل آنها دارای اشتباه جداگانه هستند.

مثال: فرض کنیم فاکتور A یک کود ازته در ۳ سطح ۵۰، ۱۰۰ و ۱۵۰ کیلوگرم در هکتار و فاکتور B عمق شخم در سطوح ۱۰، ۲۰، ۳۰ و ۴۰ سانتیمتر قبل از کشت بوده و آزمایش در ۴ تکرار یا بلوک در یک مزرعه چغندر قند اجرا شده باشد.

```
data;
input a b block y1;
cards;
;
proc anova;
classes block a b;
model y1=block a block*a b block*b a*b;
test h=a e=block*a;
test h=b e=block*b;
means a/duncan alpha=0.01 e=block*a;
means b/duncan alpha=0.01 e=block*b;
means a*b;
run;quit;
```

قابل توجه است که برای آزمون F فاکتور A باید از اثر بلوک $A \times B$ در مخرج کسر F و برای آزمون F فاکتور B باید از اثر بلوک $B \times A$ در مخرج کسر F استفاده شود.

Dependent Variable: y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	29	8491.750000	292.818966	17.68	<.0001
Error	18	298.166667	16.564815		
Corrected Total	47	8789.916667			

R-Square	Coeff Var	Root MSE	y1 Mean
0.966079	8.229128	4.069990	49.45833

Source	DF	Anova SS	Mean Square	F Value	Pr > F
block	3	68.250000	22.750000	1.37	0.2828
a	2	7191.166667	3595.583333	217.06	<.0001
block*a	6	148.500000	24.750000	1.49	0.2360
b	3	660.916667	220.305556	13.30	<.0001
block*b	9	254.083333	28.231481	1.70	0.1604
a*b	6	168.833333	28.138889	1.70	0.1787

Tests of Hypotheses Using the Anova MS for block*a as an Error Term

Source	DF	Anova SS	Mean Square	F Value	Pr > F
a	2	7191.166667	3595.583333	145.28	<.0001

Tests of Hypotheses Using the Anova MS for block*b as an Error Term

Source	DF	Anova SS	Mean Square	F Value	Pr > F
b	3	660.916667	220.305556	7.80	0.0071

۴۱. طرح کرت های خرد شده نواری (Strip-Split-Block) با مقایسات متعامد

```
data neda;
input plant spray harvest block yield;

cards;
1 1 1 1 25.7
.
3 2 3 4 40.2
;
```

```
proc glm data=neda;
```

```

class plant spray harvest block;
model yield = block plant spray harvest plant*spray plant*harvest spray*harvest
plant*spray*harvest block*plant block*plant*spray;
test H=block plant E=block*plant;
test H=spray spray*plant E=block*plant*spray;

means plant/duncan e=block*plant;
means spray/duncan e=block*plant*spray;
means harvest/duncan;

contrast 'plant1 vs plant2' plant -1 1 0/e=block*plant;
contrast 'plant 2 vs plant3' plant 0 1 -1/e=block*plant;
contrast 'no-spray vs spray' spray 1 -1/e=block*plant*spray;
contrast 'harvest lin' harvest -1 0 1/e;
contrast 'harvest qua' harvest 1 -2 1/e;
run;

```

تجزیه مرکب طرح های آزمایشی

۴۲. تجزیه مرکب طرح بلوک های کامل تصادفی در چند مکان

```

data;
input cultivar place block y1;
cards;
;
proc anova;
classes cultivar place block;
model y1=place block(place) cultivar cultivar *place;
test h=place e=block(place);
means place/duncan alpha=0.01 e=block(place);
means cultivar /duncan alpha=0.01;
means cultivar *place;
run;quit;

```

گزاره **classes** متغیرهای طبقه بندی داده ها را نشان می دهد. که شامل رقم، مکان و بلوک می باشد. گزاره

MODEL منابع تغییرات طرح را بجز باقیمانده را در بر دارد و گزاره **test** از **SAS** می خواهد که در آزمون **F**

برای منبع تغییر مکان از منبع تغییر بلوک داخل مکان $e=block(place)$ به عنوان مخرج کسر **F** استفاده کند.

Dependent Variable: y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	4070527.778	203526.389	13.69	<.0001
Error	24	356786.667	14866.111		
Corrected Total	44	4427314.444			

R-Square	Coeff Var	Root MSE	y1 Mean
0.919412	9.495024	121.9267	1284.111

Source	DF	Anova SS	Mean Square	F Value	Pr > F
place	2	3142807.778	1571403.889	105.70	<.0001
block(place)	6	190896.667	31816.111	2.14	0.0856
a	4	340625.556	85156.389	5.73	0.0022
a*place	8	396197.778	49524.722	3.33	0.0105

Tests of Hypotheses Using the Anova MS for block(place) as an Error Term

Source	DF	Anova SS	Mean Square	F Value	Pr > F
place	2	3142807.778	1571403.889	49.39	0.0002

۴۳. طرح بلوکهای کامل تصادفی در چند سال

طرحهایی که تا به این قسمت مورد بررسی قرار گرفتند عمدتاً دارای فاکتورهای ثابت بودند، اما در اینجا فاکتور سال یک فاکتور تصادفی می باشد. بنابراین نحوه تجزیه آن قدری متفاوت می باشد. در این برنامه ها از رویه GLM به جای ANOVA استفاده می شود.

```
data;
input a year block y1;
cards;
;
proc glm;
classes a year block;
model y1=year block(year) a a*year;
random year block(year) a a*year/test;
means year/duncan alpha=0.01 e=block(year);
means a/duncan alpha=0.01;
means a*year;
run;quit;
```

در برنامه فوق از گزاره RANDOM استفاده شده است. در طرحهایی که دارای اثرات تصادفی باشند، از این گزاره استفاده می شود تا این اثرات تصادفی را به SAS معرفی کند. چنانچه پس از این گزاره از گزاره اختیاری test استفاده شود، برنامه SAS آزمونهای F مناسب برای هر منبع تغییر را مشخص و اجرا می کند.

Class	Levels	Values
a مثلاً رقم	5	1 2 3 4 5
year	3	1 2 3
block	3	1 2 3

Number of observations 45

Dependent Variable: y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	4070527.778	203526.389	13.69	<.0001
Error	24	356786.667	14866.111		
Corrected Total	44	4427314.444			

R-Square	Coeff Var	Root MSE	y1 Mean
0.919412	9.495024	121.9267	1284.111

Source	DF	Type I SS	Mean Square	F Value	Pr > F
year	2	3142807.778	1571403.889	105.70	<.0001
block(year)	6	190896.667	31816.111	2.14	0.0856
a	4	340625.556	85156.389	5.73	0.0022
a*year	8	396197.778	49524.722	3.33	0.0105

Source	DF	Type III SS	Mean Square	F Value	Pr > F
year	2	3142807.778	1571403.889	105.70	<.0001
block(year)	6	190896.667	31816.111	2.14	0.0856
a	4	340625.556	85156.389	5.73	0.0022
a*year	8	396197.778	49524.722	3.33	0.0105

The GLM Procedure

Source	Type III Expected Mean Square
year	Var(Error) + 3 Var(a*year) + 5 Var(block(year)) + 15 Var(year)
block(year)	Var(Error) + 5 Var(block(year))
a	Var(Error) + 3 Var(a*year) + 9 Var(a)
a*year	Var(Error) + 3 Var(a*year)

Tests of Hypotheses for Random Model Analysis of Variance

Dependent Variable: y1

Source	DF	Type III SS	Mean Square	F Value	Pr > F
year	2	3142808	1571404	23.64	0.0002
Error	9.1204	606275	66475		
Error: MS(block(year)) + MS(a*year) - MS(Error)					

Source	DF	Type III SS	Mean Square	F Value	Pr > F
block(year)	6	190897	31816	2.14	0.0856
a*year	8	396198	49525	3.33	0.0105

Error: MS (Error)	24	356787	14866			
Source	DF	Type III SS	Mean Square	F Value	Pr > F	
a	4	340626	85156	1.72	0.2382	
Error: MS (a*year)	8	396198	49525			

۴۴. طرح بلوکهای کامل تصادفی در چند سال و چند مکان

```
data;
input a place year block y1;
cards;
;
proc glm;
classes a place year block;
model y1=place year place*year block(place year) a a*place a*year a*place*year;
random place year place*year block(place year) a a*place a*year a*place*year/test;
run;quit;
```

این برنامه برای طرحهایی کاربرد دارد که مثلاً چند رقم گندم را در ۳ سال و ۳ مکان مورد آزمایش قرار داده و در نهایت بخواهیم بر محصولترین رقم را معرفی نماییم. توجه فرمایید که منابع تغییرات تصادفی در گزاره RANDOM معرفی شده اند. قابل توجه است که: در منابع تغییرات اثرات متقابل سال × رقم، چنانچه یکی از اثرات تصادفی باشند، آن اثر متقابل را نیز می توان تصادفی در نظر گرفت.

Class Level Information		
Class	Levels	Values
a	5	1 2 3 4 5
place	3	1 2 3
year	3	1 2 3
block	3	1 2 3

Number of observations 135

Dependent Variable: y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	62	12842643.70	207139.41	2.77	<.0001
Error	72	5389870.00	74859.31		
Corrected Total	134	18232513.70			

R-Square	Coeff Var	Root MSE	y1 Mean
0.704381	22.81021	273.6043	1199.481

Source	DF	Type I SS	Mean Square	F Value	Pr > F
place	2	524193.704	262096.852	3.50	0.0354
year	2	6769342.593	3384671.296	45.21	<.0001
place*year	4	331820.741	82955.185	1.11	0.3594
block(place*year)	18	2318696.667	128816.481	1.72	0.0551
a	4	398935.926	99733.981	1.33	0.2663
a*place	8	257156.296	32144.537	0.43	0.8997
a*year	8	1409390.741	176173.843	2.35	0.0262
a*place*year	16	833107.037	52069.190	0.70	0.7891

Source	DF	Type III SS	Mean Square	F Value	Pr > F
place	2	524193.704	262096.852	3.50	0.0354
year	2	6769342.593	3384671.296	45.21	<.0001
place*year	4	331820.741	82955.185	1.11	0.3594
block(place*year)	18	2318696.667	128816.481	1.72	0.0551
a	4	398935.926	99733.981	1.33	0.2663
a*place	8	257156.296	32144.537	0.43	0.8997
a*year	8	1409390.741	176173.843	2.35	0.0262
a*place*year	16	833107.037	52069.190	0.70	0.7891

The GLM Procedure

Source	Type III Expected Mean Square
place	Var(Error) + 3 Var(a*place*year) + 9 Var(a*place) + 5 Var(block(place*year)) + 15 Var(place*year) + 45 Var(place)
year	Var(Error) + 3 Var(a*place*year) + 9 Var(a*year) + 5 Var(block(place*year)) + 15 Var(place*year) + 45 Var(year)
place*year	Var(Error) + 3 Var(a*place*year) + 5 Var(block(place*year)) + 15 Var(place*year)
block(place*year)	Var(Error) + 5 Var(block(place*year))

```

a          Var(Error) + 3 Var(a*place*year) + 9 Var(a*year) + 9 Var(a*place) + 27
          Var(a)
a*place   Var(Error) + 3 Var(a*place*year) + 9 Var(a*place)
a*year    Var(Error) + 3 Var(a*place*year) + 9 Var(a*year)
a*place*year Var(Error) + 3 Var(a*place*year)

```

The GLM Procedure

Tests of Hypotheses for Random Model Analysis of Variance

Dependent Variable: y1

Source	DF	Type III SS	Mean Square	F Value	Pr > F
place	2	524194	262097	4.16	0.1965
Error	1.9677	124027	63031		
Error: MS(place*year) + MS(a*place) - MS(a*place*year) + 27E-16*MS(Error)					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
year	2	6769343	3384671	16.35	0.0019
Error	7.4311	1538686	207060		
Error: MS(place*year) + MS(a*year) - MS(a*place*year) + 34E-16*MS(Error)					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
place*year	4	331821	82955	0.78	0.5626
Error	9.6152	1019460	106026		
Error: MS(block(place*year)) + MS(a*place*year) - MS(Error)					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
block(place*year)	18	2318697	128816	1.72	0.0551
a*place*year	16	833107	52069	0.70	0.7891
Error: MS(Error)	72	5389870	74859		
Source	DF	Type III SS	Mean Square	F Value	Pr > F
a	4	398936	99734	0.64	0.6548
Error	5.8431	912972	156249		
Error: MS(a*place) + MS(a*year) - MS(a*place*year) + 22E-17*MS(Error)					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
a*place	8	257156	32145	0.62	0.7514
a*year	8	1409391	176174	3.38	0.0182
Error	16	833107	52069		
Error: MS(a*place*year)					

۴۵. تجزیه جداگانه (به تفکیک سال یا مکان) یک طرح بلوکهای کامل تصادفی در چند سال و مکان با داده های

تجزیه مرکب.

```

data;
input a place year block y1;
cards;
;
proc sort;by place year;
proc anova;
classes a block;
model y1=a block;
by place year;
run;quit;

```

در برنامه فوق رویه SORT داده ها را بر اساس مکان و سال مرتب می کند و سپس با استفاده از گزاره BY تجزیه واریانس برای هر مکان و سال به طور جداگانه انجام می گیرد.

----- place=1 year=1 -----

Dependent Variable: y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	125643.3333	20940.5556	1.70	0.2379
Error	8	98450.0000	12306.2500		
Corrected Total	14	224093.3333			
	R-Square	Coeff Var	Root MSE	y1 Mean	
	0.560674	11.54756	110.9335	960.6667	
Source	DF	Anova SS	Mean Square	F Value	Pr > F
a	4	74560.00000	18640.00000	1.51	0.2856
block	2	51083.33333	25541.66667	2.08	0.1879

به همین ترتیب برای هر نوبت نمونه گیری، به ترتیب تجزیه واریانس جداگانه را انجام می دهد. در اینجا فقط مکان ۱ تاریخ ۱ و آخرین مرحله یعنی مکان ۳ تاریخ ۳ آورده شده است.

```
----- place=3 year=3 -----
```

Dependent Variable: y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1299953.333	216658.889	3.23	0.0646
Error	8	537140.000	67142.500		
Corrected Total	14	1837093.333			

	R-Square	Coeff Var	Root MSE	y1 Mean
	0.707614	19.16086	259.1187	1352.333

Source	DF	Anova SS	Mean Square	F Value	Pr > F
a	4	170910.000	42727.500	0.64	0.6510
block	2	1129043.333	564521.667	8.41	0.0108

۴۶. طرح کرت های خرد شده در چند مکان

```
data;
input place a b block y1;
cards;
;
proc anova;
classes place a b block;
model y1=place block(place) a a*place block*a(place) b a*b b*place a*b*place;
test h=a place e=block*a(place);
run;quit;
```

همانطوریکه ملاحظه می گردد، در برنامه بالا همه اثرات ثابت در نظر گرفته می شوند. لذا از رویه ANOVA استفاده شده است.

۴۷. طرح کرت های خرد شده در چند سال و چند مکان

```
data;
input place year a b block y1;
cards;
;
proc glm;
classes place year a b block;
model y1=place year place*year block(place year) a a*place a*year a*place*year
a*block(place year) b a*b b*year b*place
b*place*year b*a*year b*a*place b*a*place*year;
random year place*year block(place year) a*year a*place a*place*year a*block(place
year) b*year b*place*year b*a*year b*a*place*year;
run;quit;
```

در این برنامه همان طور که مشاهده می گردد، سال تصادفی در نظر گرفته شده است و همانطور که در برنامه ۳۹ توضیح داده شد، در منابع تغییرات اثرات متقابل سال × رقم، چنانچه یکی از اثرات تصادفی باشند، آن اثر متقابل را نیز می توان تصادفی در نظر گرفت.

۴۸. طرح مربع لاتین در چند سال و چند مکان

```
data;
input place year row col treat yi;
cards;
;
proc glm;
```

```

classes place year row col treat;
model y1=place year place*year row(place year) col(place year) treat treat*place
treat*year treat*place*year;
random year place*year row(place year) col(place year) treat*year treat*place*year;
run;quit;

```

برآورد اجزای واریانس

در تحقیقات کشاورزی و دامپروری گاهی نیاز است که واریانس یک یا چند منبع تغییر به اجزای آن تفکیک شود. بهترین مثال در این مورد، برآورد واریانس ژنتیکی در طرح‌های مقایسه ژنوتیپهاست.

۴۹. برآورد اجزای واریانس یک طرح بلوکهای کامل تصادفی

در یک آزمایش میزان علوفه تولیدی ۵ رقم یونجه در قالب طرح بلوکهای کامل تصادفی بررسی شده است ولی زاده و مقدم (۱۳۷۳). چنانچه بخواهیم واریانس ژنتیکی را برای متغیر علوفه تولیدی بدانیم باید از برنامه زیر استفاده نماییم.

```

data;
input block treat$ y1;
cards;
;
proc varcomp method=type1;
classes block treat;
model y1=block treat;
run;

```

برای برآورد اجزای واریانس از رویه VARCOMP استفاده می‌گردد. پس از نوشتن دستور VARCOMP باید نوع روش نیز نوشته شود. مثلاً در این برنامه از روش type1 استفاده شده است (method=type1). روش VARCOMP از SAS می‌خواهد که اجزای واریانس را با روش TYPE1 محاسبه کند. در این گزاره CLASSES همانطور که قبلاً گفته شد متغیرهای طبقه‌بندی، یعنی بلوک و تیمار را معرفی می‌کند.

Variance Components Estimation Procedure

Class Level Information

Class	Levels	Values
block	4	1 2 3 4
treat	5	a b c d e

Number of observations 20

Dependent Variable: y1

Type 1 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square
block	3	8.585655	2.861885	Var(Error) + 5 Var(block)
treat	4	20.273570	5.068392	Var(Error) + 4 Var(treat)
Error	12	5.000670	0.416722	Var(Error)
Corrected Total	19	33.859895	.	.

Type 1 Estimates

Variance Component	Estimate
Var(block)	0.48903
Var(treat)	1.16292
Var(Error)	0.41672

۵۰. برآورد اجزای واریانس یک طرح بلوکهای کامل تصادفی در چند سال و چند مکان

```

data;
input a place year block y1;
cards;
;
proc varcomp method=type1;

```

```

classes a place year block;
model y1=place year place*year block(place year) a a*place a*year
a*place*year/fixed=1;
run;quit;

```

در برنامه فوق فرض بر این است که به استثنای مکان، سایر اثرات تصادفی باشند. در این برنامه که مانند برنامه ۴۴ می‌باشد، رویه VARCOMP، CLASSES و MODEL مانند توضیحات قبل می‌باشد. تنها مورد جدید، استفاده از کلمه اختیاری fixed=1 می‌باشد. این کلمه به SAS اعلام می‌کند که اولین منبع تغییر، یعنی مکان (place)، تصادفی نیست و ثابت می‌باشد و سایر اثرات تصادفی هستند.

Type 1 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square
place	2	524194	262097	Var(Error) + 3 Var(a*place*year) + 9 Var(a*place) + 5 Var(block(place*year)) + 15 Var(place*year) + Q(place)
year	2	6769343	3384671	Var(Error) + 3 Var(a*place*year) + 9 Var(a*year) + 5 Var(block(place*year)) + 15 Var(place*year) + 45 Var(year)
place*year	4	331821	82955	Var(Error) + 3 Var(a*place*year) + 5 Var(block(place*year)) + 15 Var(place*year)
block(place*year)	18	2318697	128816	Var(Error) + 5 Var(block(place*year))
a	4	398936	99734	Var(Error) + 3 Var(a*place*year) + 9 Var(a*year) + 9 Var(a*place) + 27 Var(a)
a*place	8	257156	32145	Var(Error) + 3 Var(a*place*year) + 9 Var(a*place)
a*year	8	1409391	176174	Var(Error) + 3 Var(a*place*year) + 9 Var(a*year)
a*place*year	16	833107	52069	Var(Error) + 3 Var(a*place*year)
Error	72	5389870	74859	Var(Error)
Corrected Total	134	18232514	.	.

Type 1 Estimates

Variance Component Estimate

Var(year)	70613.6
Var(place*year)	-1538.1

Variance Components Estimation Procedure

Type 1 Estimates

Variance Component Estimate

Var(block(place*year))	10791.4
Var(a)	-2093.2
Var(a*place)	-2213.9
Var(a*year)	13789.4
Var(a*place*year)	-7596.7
Var(Error)	74859.3

تجزیه کواریانس در طرح‌های آزمایشی

عمده‌ترین کاربرد تجزیه کواریانس در کشاورزی، کنترل اشتباه آزمایش و تصحیح میانگین تیمارها می‌باشد.

۵۱. تجزیه کواریانس در طرح کاملاً تصادفی

برنامه زیر مربوط به تجزیه کواریانس یک طرح در اثر خسارت ملخ می‌باشد.

```

data;
input treat rep x y1;
cards;
;
proc glm;
class treat;
model y1=treat x;

```

```
manova h=treat/printh printe;
run;
proc glm data=a;
class treat;
model y1=treat x;
estimate 'b coeff.': x1;
lsmeans treat;
run;quit;
```

در برنامه فوق رویه PROC GLM از SAS می‌خواهد که یک تجزیه توام واریانس و کواریانس را بر روی متغیرهای Y و X انجام دهد. دستور MANOVA از رویه GLM می‌خواهد که تجزیه کواریانس را برای تیمار انجام دهد و ماتریسهای SSCP را برای منابع تغییر تیمار (PRINTH) و اشتباه (PRINTE) چاپ کند. روی متغیرهای Y و X انجام دهد.

در قسمت دوم برنامه نیز متغیر X نیز به عنوان یک متغیر مستقل به مدل طرح وارد شده است. گزاره ESTIMATE از SAS می‌خواهد که ضریب رگرسیون اشتباه را محاسبه و تحت برچسب "b coeff." گزارش نماید. گزاره Lsmeans موجب چاپ میانگینهای تصحیح شده تیمار (TREAT) پس از وارد کردن اثر تعداد پنجه اولیه می‌شود.

*** قابل توجه است که مجموع مربعات تصحیح شده باید به وسیله دست محاسبه شوند زیرا SAS این مقادیر را مستقیماً چاپ نمی‌کند.

Dependent Variable: y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	219.2427907	21.9242791	5.05	0.0001
Error	39	169.2372093	4.3394156		
Corrected Total	49	388.4800000			

R-Square	Coeff Var	Root MSE	y1 Mean
0.564361	24.44984	2.083126	8.520000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
treat	9	36.8800000	4.0977778	0.94	0.4991
x	1	182.3627907	182.3627907	42.02	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
treat	9	45.9849667	5.1094407	1.18	0.3361
x	1	182.3627907	182.3627907	42.02	<.0001

Multivariate Analysis of Variance

E = Error SSCP Matrix

y1

y1 169.2372093

Multivariate Analysis of Variance

H = Type III SSCP Matrix for treat

y1

y1 45.98496667

Characteristic Roots and Vectors of: E Inverse * H, where

H = Type III SSCP Matrix for treat

E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector	V'EV=1
Root		y1	
0.27171901	100.00	0.07686915	

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall treat Effect
 H = Type III SSCP Matrix for treat
 E = Error SSCP Matrix

S=1 M=3.5 N=18.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.78633723	1.18	9	39	0.3361
Pillai's Trace	0.21366277	1.18	9	39	0.3361
Hotelling-Lawley Trace	0.27171901	1.18	9	39	0.3361
Roy's Greatest Root	0.27171901	1.18	9	39	0.3361

۵۲. تجزیه کواریانس در یک طرح بلوکهای کامل تصادفی

```
data;
input treat block x y1;
cards;
;
proc glm;
class treat block;
model y1 x=treat block;
manova h=treat block/printh printe;
means treat;
run;
proc glm data=a;
class treat block;
model y1=treat block x;
estimate "b coeff." x1;
lsmeans treat;
run;quit;
```

Class	Levels	Class Level Information														
		Values														
treat	15	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
block	3	1	2	3												
		Number of observations										45				

Dependent Variable: y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	69.06666667	4.316666667	8.80	<.0001
Error	28	13.73333333	0.49047619		
Corrected Total	44	82.80000000			

R-Square	Coeff Var	Root MSE	y1 Mean
0.834138	16.41422	0.700340	4.266667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
treat	14	68.13333333	4.866666667	9.92	<.0001
block	2	0.93333333	0.466666667	0.95	0.3983
Source	DF	Type III SS	Mean Square	F Value	Pr > F
treat	14	68.13333333	4.866666667	9.92	<.0001
block	2	0.93333333	0.466666667	0.95	0.3983

The GLM Procedure

Dependent Variable: x

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	12.22222222	0.76388889	1.99	0.0540
Error	28	10.75555556	0.38412698		
Corrected Total	44	22.97777778			

R-Square	Coeff Var	Root MSE	x Mean
0.531915	11.11159	0.619780	5.577778

Source	DF	Type I SS	Mean Square	F Value	Pr > F
treat	14	10.31111111	0.73650794	1.92	0.0694
block	2	1.91111111	0.95555556	2.49	0.1013

Source	DF	Type III SS	Mean Square	F Value	Pr > F
treat	14	10.31111111	0.73650794	1.92	0.0694
block	2	1.91111111	0.95555556	2.49	0.1013

Multivariate Analysis of Variance

E = Error SSCP Matrix

	y1	x
y1	13.73333333	7.4
x	7.4	10.75555556

Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

	y1	x
y1	1.000000	0.608874
x	0.608874	1.000000

The GLM Procedure

Multivariate Analysis of Variance

H = Type III SSCP Matrix for treat

	y1	x
y1	68.13333333	19.4
x	19.4	10.31111111

Characteristic Roots and Vectors of: E Inverse * H, where

H = Type III SSCP Matrix for treat

E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector y1	V'EV=1 x
5.70316002	90.26	0.32848768	-0.14679008
0.61529202	9.74	-0.08837222	0.35525054

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall treat Effect

H = Type III SSCP Matrix for treat

E = Error SSCP Matrix

S=2 M=5.5 N=12.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.09235690	4.42	28	54	<.0001
Pillai's Trace	1.23173353	3.21	28	56	0.0001
Hotelling-Lawley Trace	6.31845205	5.91	28	43.429	<.0001
Roy's Greatest Root	5.70316002	11.41	14	28	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

H = Type III SSCP Matrix for block

	y1	x
y1	0.93333333	1.26666667
x	1.26666667	1.91111111

Multivariate Analysis of Variance

Characteristic Roots and Vectors of: E Inverse * H, where

H = Type III SSCP Matrix for block

E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector y1	V'EV=1 x
0.17783582	94.25	-0.01019064	0.31179283
0.01084463	5.75	0.34001464	-0.22480103

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall block Effect

H = Type III SSCP Matrix for block

E = Error SSCP Matrix

S=2 M=-0.5 N=12.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.83990629	1.23	4	54	0.3088
Pillai's Trace	0.16171352	1.23	4	56	0.3079
Hotelling-Lawley Trace	0.18868045	1.26	4	31.389	0.3066
Roy's Greatest Root	0.17783582	2.49	2	28	0.1011

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

Level of treat	N	The GLM Procedure			
		-----y1-----		-----x-----	
		Mean	Std Dev	Mean	Std Dev
1	3	3.00000000	1.00000000	5.66666667	0.57735027
2	3	3.33333333	0.57735027	5.33333333	0.57735027
3	3	3.66666667	0.57735027	5.00000000	0.00000000
4	3	3.00000000	0.00000000	5.33333333	0.57735027
5	3	6.33333333	0.57735027	6.66666667	0.57735027
6	3	3.33333333	0.57735027	5.33333333	0.57735027
7	3	3.00000000	0.00000000	5.66666667	0.57735027
8	3	6.33333333	0.57735027	6.33333333	0.57735027
9	3	3.66666667	0.57735027	5.66666667	1.15470054
10	3	6.66666667	0.57735027	6.33333333	1.15470054
11	3	4.66666667	0.57735027	5.33333333	0.57735027
12	3	3.66666667	1.15470054	5.33333333	0.57735027
13	3	4.33333333	0.57735027	5.33333333	0.57735027
14	3	4.00000000	1.00000000	5.00000000	0.00000000
15	3	5.00000000	1.00000000	5.33333333	0.57735027

تهیه نقشه های تصادفی برای طرح های آزمایشی

در برنامه SAS می توان با استفاده از رویه PLAN اقدام به تهیه نقشه های تصادفی طرحها نمود.

۵۳. تهیه نقشه تصادفی برای یک طرح کاملاً تصادفی

```
TITLE"COMPLETELY RANDOMIZED DESIGN";
PROC PLAN SEED=123456;
FACTORS UNIT=20;
RUN;
```

در این برنامه گزاره TITLE، تیترا خروجی برنامه را مشخص می کند. PROC PLAN از SAS می خواهد که از رویه PLAN برای تهیه نقشه طرح استفاده کند. کلمه SEED یک عدد تصادفی در اختیار رویه قرار می دهد. گزاره FACTORS، فاکتورهای آزمایشی را مشخص می کند. همچنین عبارت UNIT=20 از SAS می خواهد که ۲۰ شماره تصادفی را بین ۱ تا ۲۰ برای ۲۰ واحد آزمایشی تولید کند. چنانچه فرضاً در این طرح ۵ تیمار به کار رفته باشد، اعداد خروجی SAS را می توان ۴ تا ۴ تا به این تیمارها اختصاص داد. خروجی برنامه به صورت زیر می باشد:

```
COMPLETELY RANDOMIZED DESIGN
The PLAN Procedure
Factor      Select      Levels      Order
-----
UNIT              20              20      Random
-----
15  7  1  9 10  2  6 14 13  5 18 20 11 16 17  3 19 12  8  4
```

یا به صورت ساده تر:

```
proc plan ;
title 'completly randomized design with 6 reps, and 4 treatments';
factors number=24;
run;
```

```
completly randomized design with 6 reps, and 4 treatments
The PLAN Procedure
Factor      Select      Levels      Order
number      24              24      Random
-----
23 16  3  7  1  9 20 21 14  4 13  2 24 22 18 19  5  6 12 10  8 17 15 11
```

۵۴. تهیه نقشه تصادفی برای یک طرح کرت های خرد شده

```
TITLE ,SPLIT PLOT DESIGN,;
PROC PLAN SEED=234567;
```

```
FACTORS BLOCK=3 ORDERED A=5 B=3;
RUN;
```

قابل توجه است که بلوک ها تصادفی نمی شوند و به ترتیب (ORDERED) در خروجی چاپ می شوند.

```
          SPLIT PLOT DESIGN,
          The PLAN Procedure

          Factor      Select      Levels      Order
          BLOCK       3           3           Ordered
          A           5           5           Random
          B           3           3           Random
```

```
          BLOCK      A      --B--
          1          1      3 1 2
                   4      2 3 1
                   3      1 2 3
                   2      1 2 3
                   5      3 1 2
          2          3      3 2 1
                   1      1 2 3
                   4      3 1 2
                   5      3 1 2
                   2      1 2 3
          3          3      2 1 3
                   4      3 1 2
                   1      2 3 1
                   2      1 2 3
                   5      2 1 3
```

۵۵. تهیه نقشه تصادفی برای یک طرح بلوکهای کامل تصادفی

```
proc plan;
title 'randomized complete design with 6 blocks and 4 treatments';
factors blocks=6 ordered treatment=4;
run;
```

latin square design with 4 rows, 4 columns and 4 treatments

```
          The PLAN Procedure

          Factor      Select      Levels      Order
          rows        4           4           Ordered
          treatment   4           4           Random
```

```
          rows      --treatment--
          1          1  4  3  2
          2          3  4  2  1
          3          3  1  2  4
          4          4  1  2  3
```

latin square design with 4 rows, 4 columns and 4 treatments

```
          The PLAN Procedure

          Factor      Select      Levels      Order
          columns     4           4           Ordered
          treatment   4           4           Random
```

```
          columns   --treatment--
          1          1  4  3  2
          2          2  1  3  4
          3          3  4  2  1
          4          4  2  1  3
```

۵۶. تهیه نقشه تصادفی برای یک طرح مربع لاتین با ۴ ردیف، ۴ ستون و ۴ تیمار

```
proc plan;
title 'latin square design with 4 rows, 4 columns and 4 treatments';
```

```
factors rows=4 ordered treatment=4;
factors columns=4 ordered treatment=4;
run;
```

```
latin square design with 4 rows, 4 columns and 4 treatments 1
The PLAN Procedure
```

Factor	Select	Levels	Order
rows	4	4	Random

```
--rows-
4 1 2 3
```

```
latin square design with 4 rows, 4 columns and 4 treatments
The PLAN Procedure
```

Factor	Select	Levels	Order
columns	4	4	Random

```
--columns--
4 2 3 1
```

۵۷. تهیه نقشه تصادفی برای یک طرح آشیانه ای

```
TITLE ,COMPLETELY RANDOMIZED DESIGN,;
PROC PLAN SEED=123456;
FACTORS A=3 B=4 C=3;
RUN;
```

برنامه فوق برای طراحی نقشه یک طرح آشیانه ای با سه فاکتور A، B و C طراحی شده است.

Factor	Select	Levels	Order
A	3	3	Random
B	4	4	Random
C	3	3	Random

```
A B --C--
```

```
3 2 3 2 1
1 1 3 2
3 3 1 2
4 2 1 3
1 2 3 1 2
4 3 2 1
1 1 2 3
3 1 3 2
2 4 2 1 3
3 2 1 3
2 2 3 1
1 1 3 2
```

۵۸. تهیه نقشه تصادفی برای یک طرح فاکتوریل یا اسپلیت-اسپلیت پلات

```
proc plan;
title 'factorial or split split plot in RCD with 4 A, 3 B, 2 C and 4 blocks';
factors blocks=4 ordered A=4 B=3 C=2;
run;
```

```
factorial or split split plot in RCD with 4 A, 3 B, 2 C and 4 blocks 1
06:33 Friday, November 10, 2000
```

```
The PLAN Procedure
```

Factor	Select	Levels	Order
blocks	4	4	Ordered
A	4	4	Random
B	3	3	Random
C	2	2	Random

```
blocks A B -C-
```

```
1 1 1 1 2
3 2 1
2 2 1
3 2 1 2
3 2 1
1 2 1
2 2 2 1
```

			3	2	1
			1	2	1
	4		3	1	2
			2	1	2
			1	2	1
2	4		2	2	1
			1	2	1
			3	1	2
	1		2	2	1
			1	2	1
			3	1	2
	2		1	2	1
			2	1	2
			3	2	1
	3		3	1	2
			1	1	2
			2	2	1
3	2		1	1	2
			3	1	2
			2	1	2
	4		3	1	2
			1	2	1
			2	2	1
	3		1	1	2
			3	1	2
			2	2	1
	1		1	1	2
			3	1	2
			2	2	1
4	3		2	2	1
			1	1	2
			3	2	1
	4		1	2	1
			3	2	1
			2	1	2
	1		3	1	2
			2	2	1
			1	1	2
	2		3	2	1
			2	2	1
			1	2	1

۵۹. تهیه نقشه تصادفی برای یک طرح اسپلیت-فاکتوریل

```
proc plan;
  title 'split-factorial in RCD with 4 A, 3 B, 2 C and 4 blocks, A in main plots
and BC in sub-plots';
  factors blocks=4 ordered A=4 BC=6;*/
run;
split-factorial in RCD with 4 A, 3 B, 2 C and 4 blocks, A in main plotsand BC in sub-plots
```

The PLAN Procedure

Factor	Select	Levels	Order
blocks	4	4	Ordered
A	4	4	Random
BC	6	6	Random

blocks	A	-----BC-----
1	2	1 3 4 2 5 6
	1	5 2 4 1 3 6
	4	2 4 6 3 5 1
	3	1 5 4 2 3 6
2	2	4 2 3 1 5 6
	4	4 6 5 3 2 1
	1	1 2 6 3 4 5
	3	5 4 2 3 6 1
3	3	2 4 1 5 3 6
	2	5 2 3 1 6 4
	4	3 4 2 6 1 5
	1	4 6 2 5 1 3
4	1	2 6 3 1 4 5
	3	6 3 5 1 4 2
	4	5 3 2 4 6 1
	2	6 5 1 4 2 3

بخش سوم

کاربرد SAS در تجزیه های رگرسیون و همبستگی

در SAS از رویه های `GLM`، `REG`، `NLIN` و `ORTHOREG` می توان برای تجزیه رگرسیون استفاده کرد. تجزیه همبستگی با استفاده از رویه `CORR` انجام می شود. البته تمامی رویه های SAS برای تجزیه رگرسیون، می توانند ضرایب همبستگی بین متغیرها را محاسبه و چاپ کنند. گزاره `PROC REG` الزامی است. کلمه های اختیاری موجود در `PROC REG` عبارت اند از:

- ۱- `DATA=SASdataset`: نام مجموعه داده ای را که باید به وسیله رویه `REG` تجزیه شود، معرفی می کند.
 - ۲- `ALL`: از SAS می خواهد که خروجی برنامه را به طور کامل چاپ کند. استفاده از این کلمه اختیاری معادل درخواست تمامی کلمات اختیاری موجود در گزاره `MODEL` می باشد.
 - ۳- `CORR`: ماتریس همبستگی را برای تمامی متغیرهای فهرست شده و در `MODEL` یا `VAR` چاپ می کند.
 - ۴- `SIMPLE`: جمع، میانگین، واریانس، انحراف معیار و مجموع مربعات تصحیح نشده برای تمامی متغیرهای مورد استفاده در رویه `REG` را چاپ می کند.
 - ۵- `USSCP`: ماتریس مجموع مربعات تصحیح نشده و حاصلضربها را برای تمامی متغیرهای مورد استفاده در رویه چاپ می کند.
- کاربرد رویه `BY` مانند اصول گفته شده در صفحات قبل می باشد.
- گزاره `OUTPUT`: یک مجموعه داده خروجی حاوی آماره های محاسبه شده برای هر مشاهده، تولید می کند. در استفاده از این زاره، باید برای هر آماره مورد نظر یک اسم انتخاب شود.
- گزاره `PLOT`: نمودارهای پراکنش را چاپ می کند.

۶۰. رگرسیون ساده خطی

```
DATA;
INPUT Y X;
CARDS;
;
PROC REG;
MODEL Y=X;
PLOT Y*X;
RUN;QUIT;
```

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	90.83551	90.83551	16.23	0.0038
Error	8	44.76849	5.59606		
Corrected Total	9	135.60400			

Root MSE	2.36560	R-Square	0.6699
Dependent Mean	93.56000	Adj R-Sq	0.6286
Coeff Var	2.52843		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	55.26328	9.53489	5.80	0.0004
X	1	7.69010	1.90873	4.03	0.0038

مدل رگرسیونی داده های فوق $Y = 7.6901X + 77.26328$ می باشد.

۶۱. رگرسیون ساده خطی با استفاده از دستورات P، CLI و CLM

```
DATA;
INPUT Y X;
CARDS;
;
PROC REG;
MODEL Y=X/P CLI CLM;
RUN;QUIT;
```

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	90.83551	90.83551	16.23	0.0038
Error	8	44.76849	5.59606		
Corrected Total	9	135.60400			

Root MSE	2.36560	R-Square	0.6699
Dependent Mean	93.56000	Adj R-Sq	0.6286
Coeff Var	2.52843		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	55.26328	9.53489	5.80	0.0004
X	1	7.69010	1.90873	4.03	0.0038

Output Statistics							
Obs	Dep Var	Predicted Value	Std Error Mean	95% CL Mean	95% CL Predict	Residual	
1	92.1000	91.4068	0.9194	89.2867 93.5268	85.5542 97.2593	0.6932	
2	95.5000	94.4828	0.7823	92.6787 96.2869	88.7371 100.2285	1.0172	
3	99.3000	95.2518	0.8579	93.2736 97.2301	89.4491 101.0545	4.0482	
4	93.4000	92.9448	0.7635	91.1842 94.7054	87.2126 98.6770	0.4552	
5	94.4000	94.4828	0.7823	92.6787 96.2869	88.7371 100.2285	-0.0828	
6	87.1000	90.6378	1.0420	88.2350 93.0405	84.6769 96.5986	-3.5378	
7	93.1000	94.4828	0.7823	92.6787 96.2869	88.7371 100.2285	-1.3828	
8	89.8000	92.1758	0.8232	90.2775 94.0741	86.3998 97.9517	-2.3758	
9	91.4000	89.0997	1.3361	86.0187 92.1808	82.8347 95.3648	2.3003	
10	99.5000	100.6349	1.9087	96.2333 105.0364	93.6255 107.6443	-1.1349	
Sum of Residuals				0			
Sum of Squared Residuals				44.76849			
Predicted Residual SS (PRESS)				74.93502			

۶۲. رگرسیون ساده خطی با استفاده از دستورات XPX و I

در این برنامه اختیاری XPX و I بعد از علامت / در گزاره MODEL موجب چاپ ماتریسهای $(XX)^{-1}$, XY , XX را در جریان تجزیه رگرسیون و برازش یک مدل خطی ساده می شوند.

```
DATA;
INPUT Y X;
CARDS;
;
```

```
PROC REG;
MODEL Y=X/XPX I;
RUN;QUIT;
```

Variable	Model	Crossproducts	X'X	X'Y	Y'Y
Intercept					
Intercept	10			49.8	935.6
X	49.8			249.54	4671.1
Y	935.6			4671.1	87670.34

X'X Inverse, Parameter Estimates, and SSE

Variable	Intercept	X	Y
Intercept	16.24609375	-3.2421875	55.26328125
X	-3.2421875	0.6510416667	7.6901041667
Y	55.26328125	7.6901041667	44.768489583

$$X'X = \begin{bmatrix} 10 & 49.8 \\ 49.82 & 49.54 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 935.6 \\ 467.1 \end{bmatrix}$$

$$(X'X)' = \begin{bmatrix} 16.24609375 & -3.2421875 \\ -3.2421875 & 0.651041678 \end{bmatrix}$$

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	90.83551	90.83551	16.23	0.0038
Error	8	44.76849	5.59606		
Corrected Total	9	135.60400			

Root MSE	2.36560	R-Square	0.6699
Dependent Mean	93.56000	Adj R-Sq	0.6286
Coeff Var	2.52843		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	55.26328	9.53489	5.80	0.0004
X	1	7.69010	1.90873	4.03	0.0038

۶۳. آزمون فقدان برازش (Lack of fit)

در اینجا فرض بر این است که نرمال بودن، استقلال و ثبات واریانس همگی صادق هستند و فقط مستقیم الخط یا درجه اول بودن این رابطه مورد تردید می باشد. لازمه انجام این آزمون وجود چند مشاهده از تابع Y در یک سطح X است. توجه داشته باشید که بایستی برای Y تکرار داشته باشیم.

مثال: مشاهدات زیر مربوط به قد و وزن چند نفر می باشد. قد (H) متغیر مستقل و وزن (W) متغیر وابسته می باشد. برای قدهای ۶۰ تا ۶۹ اینچ مشاهدات زیر انجام شده است. می خواهیم رابطه بین $T = \sqrt[3]{W}$ و H را بررسی کنیم.

چنانچه مدل $T = \alpha + \beta H + e$ را فیت کنیم، ضریب تعیین برابر 0.2786 ($R^2 = 0.2786$) می شود. یعنی فقط 0.2786 درصد از تغییرات T توسط H توضیح داده می شود و به نظر می رسد که مدل خوبی نیست. می خواهیم آزمون کنیم که آیا T تابعی دلخواه از H است یا نه.

$$H_1 : Y = \alpha + \beta H + e$$

$$H_1 : Y = f(H) + e$$

نخست آماره F را به صورت زیر تشکیل می دهیم.

$$F = \frac{n-k}{k-p} * \frac{RSS_0 - RSS}{RSS} \quad RSS_0 = \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}X_i)^2 \quad RSS = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$$

با دستور PROC REG مقدار RSS_0 و با دستور proc glm مقدار RSS را از جدول آنالیز واریانس به دست می آوریم. لازم به ذکر است که RSS_0 همان مجموع مربعات اشتباه خالص است که در جدول تجزیه واریانس به با رویه PROC REG آمده است و همچنین مقدار RSS نیز همان مقدار مجموع مربعات اشتباه می باشد که در جدول تجزیه واریانس حاصل از رویه proc glm آمده است ($RSS_0 = 0.82603$ و $RSS = 0.6904888$). ضمناً با دستور proc glm یک مدل خطی بدون عرض از مبدا بر داده ها فیت می کنیم.

در مثال فوق مقادیر تعداد وزنهایی که مورد بررسی قرار گرفته اند ($k = 10$)، تعداد پارامترها برابر ۲ می باشد ($p = 2$) و تعداد ۲۹ نفر مورد بررسی قرار گرفته اند ($n = \sum_{i=1}^{10} n_i = 29$)، در نتیجه مقدار F را محاسبه می کنیم.

$$F = \frac{29-10}{10-2} * \frac{0.82603 - 0.6904888}{0.6904888} = 0.46$$

به جدول توزیع F مراجعه می کنیم و مقدار محاسبه شده را با مقدار جدول $F(8,19)$ مقایسه می کنیم و خواهیم دید که معنی دار نمی باشد.

```
DATA;
INPUT H W @@;
T=W**(1/3);
CARDS;
60 120 60 95 61 132 62 117 62 118 62 118 62 116 62 110 63 125 64 165 64 138 64 105
64 126 64 118 65 140 65 155 65 116 65 130 66 140 66 120 66 138 67 145 67 140 67 125
68 125 68 155 69 135 69 132 69 137
;
PROC REG;
MODEL T=H;
PROC GLM;
CLASS H;
MODEL T=H/NOINT;
RUN;QUIT;
```

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.31904	0.31904	10.43	0.0033
Error	27	0.82603	0.03059		
Corrected Total	28	1.14507			
Root MSE		0.17491	R-Square	0.2786	
Dependent Mean		5.04270	Adj R-Sq	0.2519	
Coeff Var		3.46859			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	2.44736	0.80435	3.04	0.0052
H	1	0.04012	0.01242	3.23	0.0033

The GLM Procedure						
Class Level Information						
Class	Levels	Values				
H	10	60 61 62 63 64 65 66 67 68 69				
Number of observations			29			

Dependent Variable: T

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	737.8917921	73.7891792	2030.44	<.0001
Error	19	0.6904888	0.0363415		
Uncorrected Total	29	738.5822809			

R-Square 0.396989 Coeff Var 3.780402 Root MSE 0.190635 T Mean 5.042705

Source	DF	Type I SS	Mean Square	F Value	Pr > F
H	10	737.8917921	73.7891792	2030.44	<.0001
Source	DF	Type III SS	Mean Square	F Value	Pr > F
H	10	737.8917921	73.7891792	2030.44	<.0001

۶۴. تعیین و تصحیح عیوب مدل (تجزیه باقیمانده ها در تجزیه رگرسیون)

نتایج حاصل از تجزیه رگرسیون وقتی معتبر و دارای مفهوم هستند که فرضهای مربوط به باقیمانده های (RESIDUAL) مدل صادق باشند. در نتیجه بررسی ساختار باقیمانده ها و الگوی داده ها با استفاده از نمودار، اهمیت زیادی دارد. این نکته بسیار حائز اهمیت است که مقدار بالا برای R^2 یا معنی دار بودن آماره t حتماً حاکی از آن نیست که برازش به خوبی انجام شده است. تجزیه باقیمانده ها روشی ساده و مؤثر برای روشن کردن کاستیهای مدل در تجزیه رگرسیون می باشد. مثال زیر تجزیه باقیمانده ها را برای مثال مصرف غذا و وزن بدن مرغ نشان می دهد. در این مجموعه داده متغیر YHAT متغیر وابسته پیش بینی شده به وسیله مدل است و متغیر YRESID شامل باقیمانده های مدل می باشد. با ترکیب (MERGE) دو مجموعه داده A و B مجموعه جدیدی به نام C تشکیل می شود. در پایان رویه PLOT نمودار پراکنش باقیمانده ها را در مقابل YHAT و X رسم می کند.

```
DATA;
INPUT Y X;
CARDS;
;
PROC REG;
MODEL Y=X;
OUTPUT OUT=B P=YHAT R=YRESID;
DATA C;MERGE A B;
PROC PLOT;
PLOT YRESID*YHAT YRESID*X;
RUN;QUIT;
```

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	90.83551	90.83551	16.23	0.0038
Error	8	44.76849	5.59606		
Corrected Total	9	135.60400			

Root MSE 2.36560 R-Square 0.6699
 Dependent Mean 93.56000 Adj R-Sq 0.6286
 Coeff Var 2.52843

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	55.26328	9.53489	5.80	0.0004
X	1	7.69010	1.90873	4.03	0.0038

۶۵. بررسی نرمال بودن باقیمانده ها

یکی از فرضهای مدل رگرسیون خطی همانطور که در بالا عنوان گردید، فرض نرمال بودن مانده ها می باشد. برای بررسی ایم فرض از نمودار نرمال (NORMAL PLOT) استفاده می کنیم. برای رسم نمودار نرمال دو روش شرح داده می شود:

۱- برای تولید امتیازهای نرمال (NORMAL SCORE) از دستور زیر استفاده می کنیم:

```
PROC RANK NORMAL=VW;
```

```
VAR نام متغیرها شامل مانده ها
```

```
RANKS NSCORE;
```

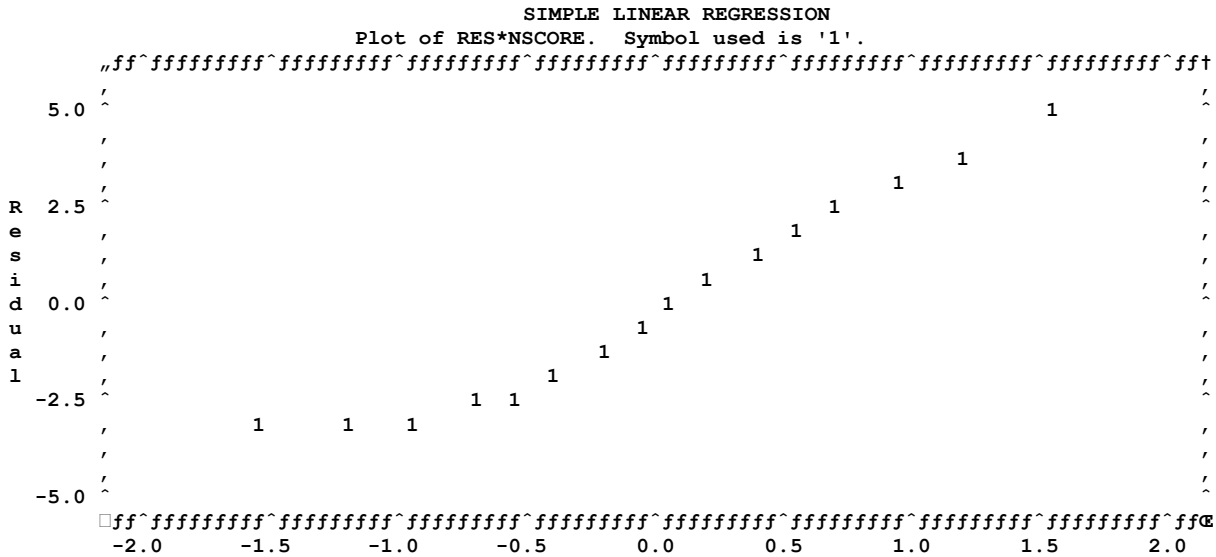
بعد با دستور PLOT مقادیر باقیمانده را در مقابل امتیازهای نرمال رسم می کنیم.

۲- روش دوم استفاده از دستور PROC UNIVARIATE می باشد که قبلاً توضیح داده شده است.

برای مثال فرض کنیم می خواهیم نرمال بودن باقیمانده ها را در مورد مقادیر متغیر X (فشار) و متغیر وابسته Y (حجم) به شرح زیر مورد بررسی قرار دهیم:

```
TITLE "SIMPLE LINEAR REGRESSION";
DATA;
INPUT X Y;
CARDS;
48 29.1
46 30.6
44 31.9
42 33.5
40 35.3
38 37
36 39.3
34 41.6
32 44.2
30 47.1
28 50.3
26 54.3
24 58.8
23 61.3
22 64.1
21 67.1
;
PROC REG;
MODEL Y=X/P NOPRINT;
OUTPUT PREDICTED=PRED RESIDUAL=RES;
PROC RANK NORMAL=VW;
VAR RES;
RANKS NSCORE;
PROC PLOT;
PLOT RES*NSCORE="1"/BOX VPOS=18;
PROC UNIVARIATE NORMAL PLOT;
VAR RES;
RUN;QUIT;
```

در مثال فوق از هر دو روش نمودار نرمال را رسم کرده ایم تا بتوان آنها را با هم مقایسه کرد.



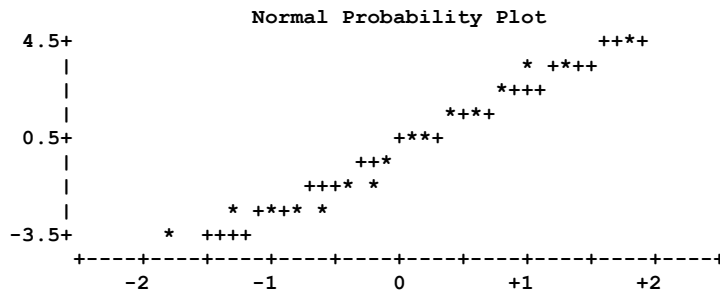
(1)

چنانچه نمودار حاصل از دستور PROC RANK NORMAL=VW (نمودار شماره ۱) یک خط راست را نشان دهد، مانده ها نرمال هستند. در غیر اینصورت از نرمال بودن انحراف وجود دارد که بایستی با انجام تبدیلاتی روی متغیرهای اصلی، آنها را نرمال کرد.

نمودار شماره ۲ که حاصل از دستور PROC UNIVARIATE NORMAL PLOT هستند، را به صورت زیر مورد استفاده قرار می دهیم:

در نمودار شماره ۲ علامتهای *** نشان دهنده باقیمانده ها هستند و علامتهای +++ یک خط معیار را نشان می دهند که با استفاده از میانگین و واریانس باقیمانده ها رسم شده اند. هر چه علامتهای * نسبت به علامتهای + انحراف بیشتری داشته باشند، انحراف از نرمال بودن بیشتر است و برعکس. البته تشخیص و استفاده از این نمودارها بیشتر احتیاج به تجربه دارد که چه وقت قبول کنیم که نرمال هستند یا خیر.

Stem Leaf	#	Boxplot
4 7	1	
3 19	2	
2 6	1	+-----+
1 27	2	
0 05	2	+
-0 9	1	*-----*
-1 2	1	
-2 99440	5	+-----+
-3 0	1	



(2)

۶۶. تعیین و تصحیح عیوب مدل (تعیین نقاط پرت یا OUTLIERS)

یکی از مسائل مهم در رگرسیون، شناسایی نقاط پرت (OUTLIERS) و بررسی تاثیر هر کدام از مشاهدات در مدل فیت شده می باشد. معیارهایی که می توان با استفاده از آنها تاثیر هر کدام از مشاهدات را مشخص کرد و نسبت به پرت بودن مشاهدات تصمیم گیری کرد با دستور فرعی INFLUENCE محاسبه می شوند. در زیر خروجی این برنامه را ملاحظه می نمایید. البته جداول تجزیه رگرسیونی نیز در خروجی برنامه وجود داشت که در اینجا حذف شده اند.

```
DATA;
INPUT Y X;
CARDS;
;
PROC REG;
MODEL Y=X/INFLUENCE;
OUTPUT OUT=B P=YHAT R=YRESID;
PROC PLOT;
PLOT YRESID*(YHAT X);
RUN;QUIT;
```

ستون دوم خروجی فوق باقیمانده ها (Residual) را نشان می دهد. ستون سوم که با **Rstudent** مشخص شده است، مقادیر باقیمانده استیودنت (student residual)، ستون چهارم ماتریس هت (Hat matrix) است و مقادیر **Leverage** یا نفوذ مشاهدات را نشان می دهند. در مثال فوق متغیرهای ۹ و ۱۰ بالاترین **leverage** را دارند. ستون پنجم **covratio**، نشان دهنده تاثیر هر کدام از مشاهدات روی دقت برآوردها می باشد. نقاطی که **Leverage** آنها زیاد باشد، **covratio** آنها نیز بالا خواهد بود. ستون ششم آماره **DFBETAS** می باشد که تغییر در مقدار \hat{Y} را برای **I** امین مشاهده محاسبه می کند. متغیر هفتم آماره **DFBETAS** می باشد که تغییر در هر یک از ضرایب رگرسیون را با حذف **I** امین مشاهده اندازه گیری می کند. اگر n تعداد مشاهدات در داده ها و p تعداد پارامترها در مدل باشند، به طور کلی برای تعیین اینکه آیا یک مشاهده خاص نقطه پرت به حساب می آید یا خیر، می توان از معیارهای زیر استفاده کرد:

$$1- اگر \text{Hat Diag or HAT MATRIX} > \frac{2p}{n} \text{ داده پرت می باشد.}$$

$$2- اگر |RStudent| > 2 \text{ باشد داده پرت است.}$$

$$3- اگر |covratio - 1| \geq \frac{3p}{n} \text{ باشد، داده پرت است.}$$

$$4- اگر مقادیر |DFBETAS| و |DFBETAS| بزرگتر از ۲ باشند.$$

$$5- به طور دقیق تر مقدار |DFBETAS| بزرگتر از $\frac{2\sqrt{p}}{n}$ بوده یا مقدار |DFBETAS| از $2\sqrt{n}$ بیشتر شود.$$

با توجه به موارد عنوان شده، در مثال فوق مشاهدات ۷، ۸، ۹ و ۱۰ را می توان نقاط پرت در نظر گرفت. نتایج تجزیه رگرسیون قبل حذف نقاط پرت که در خروجی **sas** آمده است به صورت زیر می باشد:

Obs	Output Statistics					-----DFBETAS-----	
	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITs	Intercept	X
1	0.9686	0.7066	0.0609	1.1040	0.1800	-0.0733	0.1211
2	-0.7644	-0.5581	0.0688	1.1287	-0.1516	0.0703	-0.1089
3	-0.1975	-0.1441	0.0776	1.1641	-0.0418	0.0214	-0.0316
4	0.0694	0.0509	0.0874	1.1782	0.0158	-0.0087	0.0124
5	-0.4637	-0.3428	0.0982	1.1822	-0.1131	0.0666	-0.0919
6	-1.9467	-1.5048	0.1099	1.0285	-0.5289	0.3281	-0.4415
7	-2.3699	-1.8701	0.1099	0.9470	-0.6573	-0.6344	0.5487
8	-2.3030	-1.7975	0.0982	0.9512	-0.5931	-0.5660	0.4820
9	2.9363	2.3828	0.0982	0.8141	0.7862	-0.4633	0.6389
10	2.3033	1.8110	0.1099	0.9607	0.6365	-0.3948	0.5313
11	1.0993	0.7936	0.0364	1.0658	0.1542	0.0887	-0.0448
12	-1.0337	-0.7444	0.0344	1.0694	-0.1406	-0.0672	0.0252
13	-0.2168	-0.1545	0.0335	1.1107	-0.0288	-0.0106	0.0017
14	-1.1999	-0.8667	0.0335	1.0532	-0.1613	-0.0411	-0.0098
15	0.9671	0.6955	0.0344	1.0749	0.1314	0.0181	0.0235
16	-1.0160	-0.7322	0.0364	1.0730	-0.1423	-0.0034	-0.0413
17	-1.3491	-0.9812	0.0393	1.0437	-0.1986	0.0164	-0.0776
18	-0.7822	-0.5634	0.0433	1.0981	-0.1198	0.0216	-0.0574
19	0.4348	0.3127	0.0482	1.1217	0.0703	-0.0187	0.0390
20	-2.0983	-1.5795	0.0540	0.9527	-0.3776	0.1291	-0.2337
21	0.4301	0.3199	0.1099	1.1992	0.1124	0.1085	-0.0939
22	0.5970	0.4419	0.0982	1.1755	0.1458	0.1392	-0.1185
23	2.1639	1.6663	0.0874	0.9689	0.5156	0.4849	-0.4055
24	1.0308	0.7598	0.0776	1.1176	0.2204	0.2033	-0.1664
25	1.7978	1.3482	0.0688	1.0138	0.3663	0.3296	-0.2630
26	0.3647	0.2640	0.0609	1.1393	0.0672	0.0585	-0.0452
27	0.6316	0.4567	0.0540	1.1195	0.1092	0.0910	-0.0676
28	-0.7014	-0.5060	0.0482	1.1087	-0.1138	-0.0896	0.0632
29	1.0655	0.7714	0.0433	1.0761	0.1640	0.1197	-0.0786
30	-0.4176	-0.2989	0.0393	1.1121	-0.0605	-0.0399	0.0236

Sum of Residuals 0
Sum of Squared Residuals 55.02597
Predicted Residual SS (PRESS) 65.50947

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	36.11570	36.11570	18.38	0.0002
Error	28	55.02597	1.96521		
Corrected Total	29	91.14167			

Root MSE 1.40186 R-Square 0.3963
Dependent Mean 5.03333 Adj R-Sq 0.3747
Coeff Var 27.85154

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	1.70654	0.81715	2.09	0.0460
X	1	0.66536	0.15521	4.29	0.0002

چنانچه نقاط پرت را حذف نموده و سپس اقدام به تجزیه رگرسیون نماییم، نتیجه جدول تجزیه واریانس و ضرایب رگرسیون به صورت زیر خواهد شد:

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	3.94421	3.94421	4.61	0.0421
Error	24	20.53733	0.85572		
Corrected Total	25	24.48154			

Root MSE 0.92505 R-Square 0.1611
Dependent Mean 5.01154 Adj R-Sq 0.1262
Coeff Var 18.45845

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	3.71325	0.63135	5.88	<.0001
X	1	0.25966	0.12095	2.15	0.0421

۶۷. تعیین و تصحیح عیوب مدل (تبدیل متغیرها)

چون تجزیه رگرسیون برای مدل های خطی ساده است، به همین دلیل در مواردی که رابطه بین دو متغیر خطی نیست، سعی می شود آنها را به نحوی تغییر داد که رابطه خطی حاصل شود. در این برنامه نحوه خطی کردن یک مدل غیر خطی را با استفاده از تبدیل متغیرها در برنامه SAS ارائه خواهیم داد. در یک آزمایش تعدادی باکتری زنده پس از قرار گرفتن در معرض ۲۰۰ کیلو ولت اشعه X برای یک تا ۱۵ دوره ۶ دقیقه ای شمارش شده است. ملاحظات تئوریک حاکی از آن است که بین این دو متغیر رابطه $Y = Y_0 e^{BX}$ برقرار است (تعداد باکتری و X زمان) و همچنین Y_0 (تعداد باکتری در شروع آزمایش) و B (شدت از بین رفتن باکتریها در اثر اشعه X) پارامترهای مدل هستند. با گرفتن لگاریتم از مدل زیر می توان آن را به صورت $\ln Y = \ln Y_0 + BX$ خطی کرد.

```
DATA;
INPUT Y X;
y2=log(y);
CARDS;
;
PROC REG;
MODEL Y=X;
OUTPUT P=YHAT R=YRESID;
PROC PLOT;
PLOT YRESID*(YHAT X);
run;
proc reg data=a;
model y2=x;
output p=yhat2 r=yresid2;
proc plot;
plot yresid2*(yhat2 x);
RUN;QUIT;
```

با دقت در برنامه ملاحظه می گردد که بعد از گزاره **input** متغیر **y2** که لگاریتم **y** است تعریف شده است. توجه

اینکه تبدیل ها همه باید بعد از گزاره **input** و قبل از گزاره **CARDS** انجام گیرند.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	106080	106080	60.62	<.0001
Error	13	22749	1749.95201		
Corrected Total	14	128830			
Root MSE		41.83243	R-Square	0.8234	
Dependent Mean		103.86667	Adj R-Sq	0.8098	
Coeff Var		40.27512			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	259.58095	22.72999	11.42	<.0001
X	1	-19.46429	2.49997	-7.79	<.0001

۶۸. تعیین و تصحیح عیوب مدل (رگرسیون وزنی)

در بعضی از مدل های فیت شده که واریانس ثابت نیست، بایستی با وزن دادن واریانس را ثابت کنیم. به عبارت دیگر مدل های رگرسیون خطی با واریانس غیر ثابت را می توان با استفاده از روش کمترین توانهای دوم موزون (WEIGHTED LEAST SQUARES) برازش داد.

```
DATA;
INPUT Y X;
CARDS;
;
PROC REG;
MODEL Y=X;
OUTPUT out=b P=YHAT R=YRESID;
data c;merge a b;
w=1/(yhat**2);
PROC reg;
model y=x;
weight w;
output out=d p=yhatw r=yresidw;
data e;merge c d;
residw=sqrt(w)*yresidw;
wyhat=sqrt(w)*yhatw;
xw=sqrt(w)*x;
proc plot;
PLOT YRESID*(YHAT X);
plot residw*(wyhat xw);
RUN;QUIT;
```

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	49200278140	49200278140	607.53	<.0001
Error	28	2267537268	80983474		
Corrected Total	29	51467815407			
Root MSE		8999.08183	R-Square	0.9559	
Dependent Mean		147092	Adj R-Sq	0.9544	
Coeff Var		6.11798			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	49443	4288.89131	11.53	<.0001
X	1	8.04844	0.32653	24.65	<.0001

مدل رگرسیونی این برنامه به صورت $Y = 49443 + 8.0484X$ می باشد.

۶۹. رگرسیون چند گانه (Multiple Regression)

یک مدل رگرسیون با بیش از یک متغیر مستقل را رگرسیون چند گانه می گویند.

```
data;
input x1 x2 y;
cards;
;
proc reg;
model y=x1 x2/p clm cli xpx i;
```



```
output out=b p=yhat r=yresid;
proc plot;
plot yresid*(yhat x1 x2);
run;quit;
```

Model Crossproducts X'X X'Y Y'Y

Variable	Intercept	x1	x2	y
Intercept	25	1312	506	235.6
x1	1312	76097.82	26296.3	11792.722
x2	506	26296.3	10460	4831.86
y	235.6	11792.722	4831.86	2284.1102

X'X Inverse, Parameter Estimates, and SSE

Variable	Intercept	x1	x2	y
Intercept	2.756210284	-0.011015371	-0.105638519	9.0314030737
x1	-0.011015371	0.000144131	0.0001705225	-0.071583584
x2	-0.105638519	0.0001705225	0.0047771491	0.2050051098
y	9.0314030737	-0.071583584	0.2050051098	9.9209535991

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	53.89485	26.94742	59.76	<.0001
Error	22	9.92095	0.45095		
Corrected Total	24	63.81580			

Root MSE	0.67153	R-Square	0.8445
Dependent Mean	9.42400	Adj R-Sq	0.8304
Coeff Var	7.12574		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	9.03140	1.11486	8.10	<.0001
x1	1	-0.07158	0.00806	-8.88	<.0001
x2	1	0.20501	0.04641	4.42	0.0002

Output Statistics

Obs	Dep Var	Predicted Value	Std Error Mean	95% CL Mean	95% CL Predict	Residual
1	10.9800	10.6046	0.1949	10.2004 11.0088	9.1545 12.0547	0.3754
2	11.1300	11.0055	0.2296	10.5292 11.4817	9.5336 12.4773	0.1245
3	12.5100	11.5417	0.2362	11.0519 12.0316	10.0654 13.0181	0.9683
4	8.4000	8.9224	0.1433	8.6253 9.2195	7.4984 10.3464	-0.5224
5	9.2700	8.9413	0.1597	8.6101 9.2724	7.5098 10.3728	0.3287
6	8.7300	8.4376	0.2298	7.9609 8.9143	6.9656 9.9096	0.2924
7	6.3600	5.9606	0.4495	5.0284 6.8928	4.2848 7.6365	0.3994
8	8.5000	8.2561	0.2878	7.6591 8.8530	6.7408 9.7713	0.2439
9	7.8200	8.2756	0.2073	7.8456 8.7055	6.8180 9.7331	-0.4556
10	9.1400	9.0154	0.1401	8.7250 9.3059	7.5928 10.4381	0.1246
11	8.2400	9.8100	0.1442	9.5110 10.1090	8.3856 11.2344	-1.5700
12	12.1900	11.2677	0.2295	10.7918 11.7437	9.7960 12.7395	0.9223
13	11.8800	11.3250	0.2347	10.8384 11.8117	9.8498 12.8003	0.5550
14	9.5700	10.3423	0.2047	9.9178 10.7669	8.8864 11.7983	-0.7723
15	10.9400	10.3964	0.1848	10.0133 10.7796	8.9520 11.8408	0.5436
16	9.5800	9.6597	0.1391	9.3713 9.9481	8.2375 11.0819	-0.0797
17	10.0900	9.2966	0.1720	8.9399 9.6533	7.8590 10.7342	0.7934
18	8.1100	8.5307	0.2223	8.0697 8.9916	7.0637 9.9976	-0.4207
19	6.8300	6.2756	0.4439	5.3551 7.1961	4.6062 7.9450	0.5544
20	8.8800	8.4135	0.2744	7.8444 8.9827	6.9091 9.9180	0.4665
21	7.6800	7.9703	0.2061	7.5430 8.3977	6.5136 9.4271	-0.2903
22	8.4700	9.1775	0.1483	8.8700 9.4851	7.7513 10.6037	-0.7075
23	8.8600	9.9389	0.1500	9.6279 10.2499	8.5119 11.3658	-1.0789
24	10.3600	10.7406	0.2062	10.3129 11.1683	9.2838 12.1975	-0.3806
25	11.0800	11.4942	0.2352	11.0065 11.9820	10.0186 12.9698	-0.4142

Sum of Residuals	0
Sum of Squared Residuals	9.92095
Predicted Residual SS (PRESS)	12.67146

۷۰. رگرسیون استاندارد شده با تجزیه مسیر

مقایسه ضرایب رگرسیون اغلب دشوار می باشد. زیرا بزرگی و کوچکی این ضرایب، بازتابی از واحدهای اندازه گیری متغیر مستقل هستند. در اکثر موارد استفاده از متغیرهای مستقل و وابسته مقیاس بندی شده (SCALED)، که به ضرایب رگرسیون فاقد واحد منجر می شوند، مفید خواهد بود. در SAS امکان به دست آوردن این ضرایب وجود دارد. قابل ذکر است که مقادیر ضرایب استاندارد شده، همان مقادیر ضرایب مسیر (PATH COEFFICIENTS) در تجزیه مسیر (PATH ANALYSIS) هستند.

```
data;
input x1 x2 y;
cards;
;
proc reg;
model y=x1 x2/stb;
run;quit;
```

در این برنامه کلمه STB باعث می شود که علاوه بر ضرایب رگرسیون معمولی، ضرایب رگرسیون استاندارد شده نیز چاپ شوند.

Analysis of Variance						
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	
Model	2	53.89485	26.94742	59.76	<.0001	
Error	22	9.92095	0.45095			
Corrected Total	24	63.81580				
	Root MSE	0.67153	R-Square	0.8445		
	Dependent Mean	9.42400	Adj R-Sq	0.8304		
	Coeff Var	7.12574				

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	9.03140	1.11486	8.10	<.0001	0
x1	1	-0.07158	0.00806	-8.88	<.0001	-0.76268
x2	1	0.20501	0.04641	4.42	0.0002	0.37939

همانطور که مشاهده می گردد، در این مثال ضریب رگرسیون استاندارد شده X_1 برابر -0.76268 و ضریب رگرسیون استاندارد شده X_2 برابر 0.37939 می باشند.

۷۱. مدل های چند جمله ای

در تحقیقات کشاورزی و سایر علوم این مدلها رایج است. برای مثال یک محقق می خواهد ارتباط بین مقاومت کاغذ و درصد سختی چوب (X) در خمیر کاغذ را بیابد. اگر با استفاده از رویه PLOT پراکنش مقادیر Y و X نسبت به هم بررسی شود، ممکن است یک مدل درجه ۲ رابطه بین مقاومت کاغذ و سختی چوب را به خوبی توصیف کند. در برنامه زیر بعد از گزاره INPUT، متغیر X_2 به صورت $X_2 = X^2$ تعریف شده است. در گزاره MODEL نیز مدل رگرسیون مشخص شده است.

```
data;
input x y;
x2=x**2;
cards;
```

```

;
proc reg;
model y=x x2;
output p=yhat r=yresid;
proc plot;
plot yresid*(yhat x);
run;quit;

```

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3104.24697	1552.12349	79.43	<.0001
Error	16	312.63829	19.53989		
Corrected Total	18	3416.88526			

Root MSE	4.42040	R-Square	0.9085
Dependent Mean	34.18421	Adj R-Sq	0.8971
Coeff Var	12.93110		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-6.67419	3.39971	-1.96	0.0673
x	1	11.76401	1.00278	11.73	<.0001
x2	1	-0.63455	0.06179	-10.27	<.0001

مدل رگرسیونی این برنامه به صورت $Y = -6.674 + 11.764X - 0.6345X^2$ می‌باشد.

۷۲. مدل‌های رگرسیون چندجمله‌ای با استفاده از دستور proc glm

در صورت استفاده از این رویه، نیازی به تعریف متغیر X^2 نیست، بلکه در مدل برنامه، جمله درجه ۲ به صورت $X \times X$ به رویه معرفی می‌شود. سایر دستورات مانند برنامه‌های قبل بوده و معادله رگرسیونی نیز مانند برنامه قبل است.

```

data;
input x y;
x2=x**2;
cards;
;
proc glm;
model y=x x*x;
output p=yhat r=yresid;
proc plot;
plot yresid*(yhat x);
run;quit;

```

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3104.246975	1552.123487	79.43	<.0001
Error	16	312.638288	19.539893		
Corrected Total	18	3416.885263			

R-Square	0.908502	Coeff Var	12.93110	Root MSE	4.420395	y Mean	34.18421
----------	----------	-----------	----------	----------	----------	--------	----------

Source	DF	Type I SS	Mean Square	F Value	Pr > F
x	1	1043.427434	1043.427434	53.40	<.0001
x*x	1	2060.819541	2060.819541	105.47	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
x	1	2689.176960	2689.176960	137.62	<.0001
x*x	1	2060.819541	2060.819541	105.47	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-6.67419161	3.39970751	-1.96	0.0673
x	11.76400567	1.00278222	11.73	<.0001
x*x	-0.63454917	0.06178832	-10.27	<.0001

مدل رگرسیونی این برنامه به صورت $Y = -6.674 + 11.764X - 0.6345X^2$ می باشد.

۷۳. مدل های رگرسیون با تمرکزدهی (centering)

در مدل های رگرسیون چندجمله ای، ضعف های مدل را که در اثر انتخاب دلخواه مبدا ایجاد می شوند، می توان به وسیله تمرکزدهی متغیرهای مستقل (یعنی تصحیح X برای میانگین) بپرطرف کرد. برای استفاده از چنین برنامه هایی، ابتدا باید میانگین متغیر مستقل (در اینجا غلظت سختی چوب) را بدانیم. با استفاده از رویه PROC MEANS و کلمه اختیاری MEAN (یادآوری گذشته)، می توان میانگین متغیر مستقل را حساب کرد.

```
data;
input x y;
x1=x-7.263;
x2=(x-7.263)**2;
cards;
;
proc reg;
model y=x1 x2;
output p=yhat r=yresid;
proc plot;
plot yresid*(yhat x);
run;quit;
```

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3104.24697	1552.12349	79.43	<.0001
Error	16	312.63829	19.53989		
Corrected Total	18	3416.88526			
Root MSE		4.42040	R-Square	0.9085	
Dependent Mean		34.18421	Adj R-Sq	0.8971	
Coeff Var		12.93110			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	45.29457	1.48286	30.55	<.0001
x1	1	2.54654	0.25384	10.03	<.0001
x2	1	-0.63455	0.06179	-10.27	<.0001

مدل رگرسیونی این برنامه به صورت $Y = 45.295 + 2.546(X - 7.2632) - 0.635(X - 7.2632)^2$ می باشد.

۷۴. مدل رگرسیون برای متغیر نشانگر

گاهی اوقات نمی توان یک مقیاس پیوسته را برای متغیرهای مورد آزمایش در نظر گرفت و این متغیرها شامل مقادیر ناپیوسته هستند. در این حالت باید به متغیرها اعدادی را اختصاص دهیم تا بتوان اثرات اینگونه متغیرها را بر روی متغیر وابسته مورد بررسی قرار داد. به اینگونه متغیرها، متغیرهای نشانگر می گویند.

```

data;
input x1 y x2 x3;
cards;
;
proc reg;
model y=x1 x2 x3;
output p=yhat r=yresid;
proc plot;
plot yresid*(yhat x1);
run;quit;

```

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3104.24697	1552.12349	79.43	<.0001
Error	16	312.63829	19.53989		
Corrected Total	18	3416.88526			
		Root MSE	4.42040	R-Square	0.9085
		Dependent Mean	34.18421	Adj R-Sq	0.8971
		Coeff Var	12.93110		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	45.29457	1.48286	30.55	<.0001
x1	1	2.54654	0.25384	10.03	<.0001
x2	1	-0.63455	0.06179	-10.27	<.0001

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	38.60575	12.86858	142.78	<.0001
Error	9	0.81118	0.09013		
Corrected Total	12	39.41692			
		Root MSE	0.30022	R-Square	0.9794
		Dependent Mean	12.78462	Adj R-Sq	0.9726
		Coeff Var	2.34827		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	1.43088	0.65744	2.18	0.0575
x1	1	0.48676	0.02574	18.91	<.0001
x2	1	-1.91838	0.20180	-9.51	<.0001
x3	1	-2.19191	0.21143	-10.37	<.0001

مدل رگرسیونی این برنامه به صورت $Y = 1.43088 + 0.48676X_1 - 1.91838X_2 - 2.19191X_3$ می باشد.

۷۵. مدل رگرسیون برای متغیر نشانگر با کاربرد گزاره **by**

```

data;
input x1 y x2 x3;
cards;
;
proc sort;
by x2 x3;
proc reg;
model y=x1;
output p=yhat r=yresid;
by x2 x3;
proc plot;
plot yresid*(yhat x1);
run;quit;

```

```
----- x2=0 x3=0 -----
Analysis of Variance
Source              DF          Sum of          Mean
                   Squares          Square          F Value      Pr > F
Model                1          7.92100          7.92100        62.70      0.0042
Error                3          0.37900          0.12633
Corrected Total      4          8.30000
                   Root MSE          0.35543      R-Square      0.9543
                   Dependent Mean    13.60000      Adj R-Sq      0.9391
                   Coeff Var          2.61349
```

```
Parameter Estimates
Variable            DF      Parameter      Standard
                   Estimate      Error      t Value      Pr > |t|
Intercept           1      2.47500      1.41394      1.75      0.1783
x1                   1      0.44500      0.05620      7.92      0.0042
```

```
----- x2=0 x3=1 -----
Analysis of Variance
Source              DF          Sum of          Mean
                   Squares          Square          F Value      Pr > F
Model                1          1.10450          1.10450        12.07      0.0738
Error                2          0.18300          0.09150
Corrected Total      3          1.28750
                   Root MSE          0.30249      R-Square      0.8579
                   Dependent Mean    12.62500      Adj R-Sq      0.7868
                   Coeff Var          2.39596
```

```
Parameter Estimates
Variable            DF      Parameter      Standard
                   Estimate      Error      t Value      Pr > |t|
Intercept           1     -0.30000      3.72320     -0.08      0.9431
x1                   1      0.47000      0.13528      3.47      0.0738
```

```
----- x2=1 x3=0 -----
Analysis of Variance
Source              DF          Sum of          Mean
                   Squares          Square          F Value      Pr > F
Model                1          23.30332          23.30332       323.26     0.0031
Error                2          0.14418          0.07209
Corrected Total      3          23.44750
                   Root MSE          0.26849      R-Square      0.9939
                   Dependent Mean    11.92500      Adj R-Sq      0.9908
                   Coeff Var          2.25150
```

```
Parameter Estimates
Variable            DF      Parameter      Standard
                   Estimate      Error      t Value      Pr > |t|
Intercept           1     -0.97912      0.73016     -1.34      0.3119
x1                   1      0.50604      0.02815     17.98      0.0031
```

۷۶. تعیین وجود همخطی در تجزیه رگرسیون (MULTICOLLINEARITY)

در تجزیه رگرسیونی، متغیرهای مستقل باید نسبت به هم متعامد (ارتوگونال) باشند. متاسفانه در اکثر موارد متغیرهای مستقل متعامد نیستند. گاهی این امر مشکلی ایجاد نمی کند. اما در برخی موارد همبستگی متغیرهای مستقل بالاست و در چنین مواردی نتیجه گیری از مدل رگرسیون می تواند گمراه کننده یا اشتباه باشد. در این حالت گفته می شود که چند هم خطی یا هم راستایی وجود دارد. در SAS امکان بررسی وجود همخطی در متغیرهای مستقل وجود دارد. به همین منظور پس از معرفی MODEL یک علامت / قرار داده و سپس از گزینه های `collinoint vif` استفاده می شود. این سه گزینه، آماره های مربوط به تعیین وجود چند همخطی را چاپ می کند. دستور

collinooint باعث می شود که ریشه های مشخصه (**CHARACTERISTIC ROOT**) مربوط به هر متغیر مستقل چاپ شوند. ریشه مشخصه کوچک (صفر یا نزدیک به آن) نشان از وجود چند همخطی است. **Vif** موجب محاسبه و چاپ فاکتور افزایش واریانس (**VARIANCE INFLATION FACTOR**) برای هر متغیر موجود در مدل می شود. اگر هر یک از مقادیر **Vif** از ۵ تا ۱۰ بزرگتر باشند، امکان ضعف برآورد ضرایب رگرسیون در اثر چند همخطی وجود دارد. **Tol** آماره ای عکس **vif** است ($TOL = \frac{1}{VIF}$).

```
data;
input yimport xport xinvest xconsum;
cards;
;
proc reg;
model yimport=xport xinvest xconsum/collinooint vif tol;
run;quit;
```

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	204.77614	68.25871	285.61	<.0001
Error	7	1.67295	0.23899		
Corrected Total	10	206.44909			
Root MSE		0.48887	R-Square	0.9919	
Dependent Mean		21.89091	Adj R-Sq	0.9884	
Coeff Var		2.23321			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Tolerance	Variance Inflation
Intercept	1	-10.12799	1.21216	-8.36	<.0001	.	0
xport	1	-0.05140	0.07028	-0.73	0.4883	0.00538	185.99747
xinvest	1	0.58695	0.09462	6.20	0.0004	0.98144	1.01891
xconsum	1	0.28685	0.10221	2.81	0.0263	0.00537	186.11002

Collinearity Diagnostics (intercept adjusted)

Number	Eigenvalue	Condition Index	-----Proportion of Variation-----		
			xport	xinvest	xconsum
1	1.99915	1.00000	0.00134	0.00092898	0.00134
2	0.99815	1.41522	0.00000686	0.98135	0.00000359
3	0.00269	27.25683	0.99865	0.01772	0.99865

۷۷. تجزیه رگرسیون با استفاده از دستور **orthoreg**

زمانیکه داده های تجزیه رگرسیون شرایط مساعدی نداشته باشند، مثلاً بین متغیرهای مستقل چند همخطی وجود داشته باشد، از این رویه استفاده می شود. این رویه در چنین مواردی می تواند ضرایب رگرسیون را با دقت بیشتری برآورد کند.

```
data;
input yimport xport xinvest xconsum;
cards;
;
proc orthoreg;
model yimport=xport xinvest xconsum;
run;quit;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	204.77614145	68.258713816	285.61	<.0001
Error	7	1.6729494613	0.2389927802		
Corrected Total	10	206.44909091			

Root MSE 0.4888688783
R-Square 0.9918965521

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-10.127988155231	1.2121599585	-8.36	<.0001
xport	1	-0.0513961597134	0.0702799871	-0.73	0.4883
xinvest	1	0.5869490426644	0.0946184207	6.20	0.0004
xconsum	1	0.28684867563942	0.1022081143	2.81	0.0263

۷۸. رسم خط رگرسیون

```
data;
input y x;
cards;
;
proc GPLOT;
  PLOT Y*X;
  SYMBOL I=RL;
run;quit;
```

۷۹. تجزیه رگرسیون و رسم خط رگرسیون به همراه نمایش پراکنش نقاط

```
data;
input y x;
cards;
;
PROC REG;
MODEL Y=X/P;
OUTPUT PREDICTED=FIT RESIDUAL=RES;
proc GPLOT;
  PLOT Y*X Y*X="O" Y*FIT="P"/OVERLAY; ;
  SYMBOL I=RL;
run;quit;
```

گزینش متغیرها و مدل سازی در رگرسیون چندگانه

در بسیاری از موارد در رگرسیون چندگانه با تعداد زیادی متغیر مستقل مواجه هستیم و در قدم اول باید زیر مجموعه ای از مناسب ترین متغیرها را انتخاب و وارد مدل رگرسیونی نماییم. برای گزینش متغیرها و تجزیه رگرسیونی در SAS از همان رویه PROC REG استفاده می شود. اما در پس از نوشتن MODEL پس از یک علامت / روش گزینش متغیرها را به صورت **SELECTION=name** می آوریم. به دنبال نوشتن این دستور، از گزاره اختیاری **SLE=p** و **SLS=p** نوشته می شود. **SLE=0.1** که در روشهای FORWARD و STEPWISE به کار می رود، به این معنی است که باید سطح معنی دار بودن متغیرهایی که وارد مدل می شوند، ۱۰ درصد یا کمتر باشد. همچنین **SLS=0.1** که فقط در زمان استفاده از گزاره BACKWARD استفاده می شود، به این معنی است که سطح معنی دار بودن برای حفظ یک متغیر در مدل ۱۰ درصد می باشد. در sas برای گزینش متغیرها گزاره های مختلفی وجود دارد که در ذیل با ذکر مثال ارائه می گردند:

۸۰. گزینش متغیر به روش forward

در این روش، متغیرهایی که سطح معنی دار بودن آنها در مدل کمتر از مقدار اختصاص داده شده به گزاره sle (در اینجا ۱۰ درصد) هستند، یکی یکی وارد مدل می شوند. در پایان نیز پارامترهای مدل ارائه می گردد.

```
data;
input y x1-x6;
cards;
;
proc reg;
model y=x1-x6/selection=forward sle=0.1;
run;quit;
```

متغیرهایی وارد مدل می شوند که سطح معنی دار بودن متغیرهایی که وارد مدل می شوند، ۱۰ درصد می باشد. به نحوه گزینش متغیرها که در خروجی برنامه به شرح زیر آمده است توجه فرمایید:

Forward Selection: Step 1

قدم اول ورود متغیر سوم به مدل

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1763.33890	1763.33890	19.49	0.0001
Error	28	2533.62776	90.48671		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	27.30933	8.63151	905.80266	10.01	0.0037
x3	0.66021	0.14956	1763.33890	19.49	0.0001

Bounds on condition number: 1, 1

Forward Selection: Step 2

Variable x1 Entered: R-Square = 0.5393 and C(p) = 7.2706

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2317.28741	1158.64370	15.80	<.0001
Error	27	1979.67926	73.32145		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	16.00455	8.79121	243.00796	3.31	0.0798
x1	0.36722	0.13360	553.94850	7.56	0.0105
x3	0.43688	0.15725	565.97534	7.72	0.0098

Forward Selection: Step 2

Bounds on condition number: 1.3643, 5.457

مثلاً بر اساس خروجی فوق مدل رگرسیونی به صورت زیر خواهد بود:

$$Y = 0.36722X_1 + 0.43688X_3 + 16.00455$$

خلاصه مراحل گزینش متغیر به صورت زیر می باشد:

No other variable met the 0.1000 significance level for entry into the model.

Summary of Forward Selection							
Step	Variable Entered	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x3	1	0.4104	0.4104	14.0206	19.49	0.0001
2	x1	2	0.1289	0.5393	7.2706	7.56	0.0105

برنامه فوق نخست متغیر سوم و پس از آن متغیر اول را وارد مدل نموده است. زیرا تنها همین دو متغیر هستند که سطح معنی دار بودن آنها در مدل ۱۰ درصد یا کمتر می باشد. بدیهی است که چنانچه سطح ۱۰ درصد را مثلاً به ۲۰ درصد افزایش دهیم، در آن صورت احتمالاً تعداد متغیر بیشتری وارد مدل خواهند شد.

۸۱. گزینش متغیر و مدلسازی به روش backward

در این روش نخست همه متغیرها وارد مدل می شوند. سپس متغیرهایی در مدل حفظ می شوند، که سطح معنی دار بودن آنها با مدل کمتر از ۱۰ درصد باشد. به همین دلیل از گزاره **SLS** به جای **SLE** استفاده شده است.

```
data;
input y x1-x6;
cards;
;
proc reg;
model y=x1-x6/selection=backward sls=0.1;
run;quit;
```

نخست همه متغیرها وارد مدل می شوند.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	2840.88079	473.48013	7.48	0.0002
Error	23	1456.08588	63.30808		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	20.14214	12.46359	165.34231	2.61	0.1197
x1	0.26828	0.16155	174.58215	2.76	0.1104
x2	0.10242	0.16068	25.72088	0.41	0.5302
x3	0.32199	0.18994	181.92831	2.87	0.1035
x4	0.46685	0.21265	305.13235	4.82	0.0385
x5	-0.18692	0.14040	112.21026	1.77	0.1961
x6	-0.30865	0.19845	153.14683	2.42	0.1335

Bounds on condition number: 2.3619, 71.204

Backward Elimination: Step 1

Variable x2 Removed: R-Square = 0.6552 and C(p) = 5.4063

قدم اول: متغیر شماره ۲ از مدل خارج می شود. زیرا همانگونه که ملاحظه می گردد، سطح اختلاف آن با مدل بیش از همه متغیرها می باشد و در کل از ۱۰ درصد بیشتر است.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	2815.15991	563.03198	9.12	<.0001
Error	24	1481.80676	61.74195		
Corrected Total	29	4296.96667			

Backward Elimination: Step 1

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	21.73314	12.05910	200.53735	3.25	0.0841
x1	0.30956	0.14616	276.96140	4.49	0.0447
x3	0.34146	0.18514	210.01982	3.40	0.0775
x4	0.46702	0.21000	305.34924	4.95	0.0358
x5	-0.19726	0.13773	126.64838	2.05	0.1650
x6	-0.28840	0.19345	137.22961	2.22	0.1490

Bounds on condition number: 2.3619, 48.363

Backward Elimination: Step 2
Variable x5 Removed: R-Square = 0.6257 and C(p) = 5.4068
Analysis of Variance

قدم دوم: متغیر شماره ۵ از مدل بیرون می رود.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2688.51153	672.12788	10.45	<.0001
Error	25	1608.45514	64.33821		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	11.04559	9.67009	83.94330	1.30	0.2642
x1	0.22954	0.13787	178.33978	2.77	0.1084
x3	0.41463	0.18165	335.21494	5.21	0.0312
x4	0.46852	0.21437	307.31916	4.78	0.0384
x6	-0.35362	0.19192	218.41352	3.39	0.0773

Bounds on condition number: 2.3619, 31.469

Backward Elimination: Step 3

Backward Elimination: Step 3

Variable x1 Removed: R-Square = 0.5842 and C(p) = 6.2238

قدم سوم: متغیر شماره ۱ از مدل بیرون می رود.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2510.17175	836.72392	12.18	<.0001
Error	26	1786.79492	68.72288		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	15.00526	9.68718	164.88972	2.40	0.1335
x3	0.51333	0.17746	575.01654	8.37	0.0076
x4	0.61186	0.20290	624.94229	9.09	0.0057
x6	-0.44398	0.19026	374.21667	5.45	0.0276

Bounds on condition number: 1.9809, 16.403

و در نهایت متغیرهای ۳، ۴ و ۶ در مدل باقی می ماند.

All variables left in the model are significant at the 0.1000 level.

Summary of Backward Elimination

Step	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x2	5	0.0060	0.6552	5.4063	0.41	0.5302
2	x5	4	0.0295	0.6257	5.4068	2.05	0.1650
3	x1	3	0.0415	0.5842	6.2238	2.77	0.1084

۸۲. گزینش متغیر و مدل سازی به روش **stepwise**

تشابه این روش با روش **forward** بدین ترتیب است که متغیرها را یکی یکی وارد مدل می کند. اما با روش **backward** نیز از این نظر تشابه دارد که ممکن است برخی از متغیرهایی را که وارد مدل شده اند، مجدداً از مدل خارج نماید. زیرا در این روش، برخلاف روش **forward**، وارد شدن متغیرها در مدل شرط لازم و کافی برای باقی ماندن آنها در داخل مدل نیست زیرا متغیرهای وارد شده هر چند دارای سطح معنی دار کمتر از ۱۰ درصد با مدل

بوده اند، اما ممکن است با ورود متغیرهای دیگر، سطح معنی داری آنها با مدل افزایش یافته و از مدل خارج شوند. به همین دلیل برای تعیین سطح معنی دار بودن ورود یا خروج متغیرها در این مدل، از هر دو گزاره **sle** و **sls** استفاده شده است.

```
data;
input y x1-x6;
cards;
;
proc reg;
model y=x1-x6/selection=stepwise sle=0.1 sls=0.1;
run;quit;
```

قدم اول: متغیر ۳ وارد مدل می شود.

Variable x3 Entered: R-Square = 0.4104 and C(p) = 14.0206

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1763.33890	1763.33890	19.49	0.0001
Error	28	2533.62776	90.48671		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	27.30933	8.63151	905.80266	10.01	0.0037
x3	0.66021	0.14956	1763.33890	19.49	0.0001

Bounds on condition number: 1, 1

Stepwise Selection: Step 2

Variable x1 Entered: R-Square = 0.5393 and C(p) = 7.2706

قدم دوم: متغیر ۱ وارد مدل می شود.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2317.28741	1158.64370	15.80	<.0001
Error	27	1979.67926	73.32145		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	16.00455	8.79121	243.00796	3.31	0.0798
x1	0.36722	0.13360	553.94850	7.56	0.0105
x3	0.43688	0.15725	565.97534	7.72	0.0098

Stepwise Selection: Step 2

Bounds on condition number: 1.3643, 5.457

به این ترتیب متغیرهای ۱ و ۳ وارد مدل شده اما و هیچیک از متغیرهای وارد شده از مدل خارج نشدند.

All variables left in the model are significant at the 0.1000 level.
No other variable met the 0.1000 significance level for entry into the model.

Summary of Stepwise Selection

Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x3		1	0.4104	0.4104	14.0206	19.49	0.0001
2	x1		2	0.1289	0.5393	7.2706	7.56	0.0105

۸۳. گزینش متغیر و مدل سازی به روش Maxr

این روش تقریباً شبیه روش FORWARD عمل می کند.

```
data;
input y x1-x6;
cards;
;
proc reg;
model y=x1-x6/selection=maxr;
run;quit;
```

Maximum R-Square Improvement: Step 1
Variable x3 Entered: R-Square = 0.4104 and C(p) = 14.0206

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1763.33890	1763.33890	19.49	0.0001
Error	28	2533.62776	90.48671		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	27.30933	8.63151	905.80266	10.01	0.0037
x3	0.66021	0.14956	1763.33890	19.49	0.0001

Bounds on condition number: 1, 1

The above model is the best 1-variable model found.

Maximum R-Square Improvement: Step 2
Variable x1 Entered: R-Square = 0.5393 and C(p) = 7.2706

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2317.28741	1158.64370	15.80	<.0001
Error	27	1979.67926	73.32145		
Corrected Total	29	4296.96667			

Dependent Variable: y					
Maximum R-Square Improvement: Step 2					
Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	16.00455	8.79121	243.00796	3.31	0.0798
x1	0.36722	0.13360	553.94850	7.56	0.0105
x3	0.43688	0.15725	565.97534	7.72	0.0098

Bounds on condition number: 1.3643, 5.457

The above model is the best 2-variable model found.

Maximum R-Square Improvement: Step 3
Variable x5 Entered: R-Square = 0.5787 and C(p) = 6.5948

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2486.68762	828.89587	11.90	<.0001
Error	26	1810.27905	69.62612		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	29.31976	12.09385	409.22655	5.88	0.0226
x1	0.44925	0.14041	712.76496	10.24	0.0036
x3	0.39329	0.15576	443.90622	6.38	0.0180
x5	-0.22046	0.14134	169.40021	2.43	0.1309

Bounds on condition number: 1.5868, 12.482

Maximum R-Square Improvement: Step 4

Variable x3 Removed: R-Square = 0.5957 and C(p) = 5.4409
Variable x4 Entered

Maximum R-Square Improvement: Step 4

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2559.73481	853.24494	12.77	<.0001
Error	26	1737.23186	66.81661		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	25.96209	12.24456	300.38447	4.50	0.0437
x1	0.44247	0.13596	707.69545	10.59	0.0031
x4	0.47170	0.16958	516.95342	7.74	0.0099
x5	-0.28408	0.13621	290.64911	4.35	0.0470

The above model is the best 3-variable model found.

Maximum R-Square Improvement: Step 5

Variable x3 Entered: R-Square = 0.6232 and C(p) = 5.5739

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2677.93031	669.48258	10.34	<.0001
Error	25	1619.03636	64.76145		
Corrected Total	29	4296.96667			

Maximum R-Square Improvement: Step 5

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	22.38356	12.34237	212.99896	3.29	0.0818
x1	0.38451	0.14056	484.63090	7.48	0.0113
x3	0.23717	0.17556	118.19549	1.83	0.1888
x4	0.33529	0.19511	191.24269	2.95	0.0981
x5	-0.24559	0.13709	207.83229	3.21	0.0853

Maximum R-Square Improvement: Step 6

Variable x5 Removed: R-Square = 0.6257 and C(p) = 5.4068

Variable x6 Entered

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2688.51153	672.12788	10.45	<.0001
Error	25	1608.45514	64.33821		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	11.04559	9.67009	83.94330	1.30	0.2642
x1	0.22954	0.13787	178.33978	2.77	0.1084
x3	0.41463	0.18165	335.21494	5.21	0.0312
x4	0.46852	0.21437	307.31916	4.78	0.0384
x6	-0.35362	0.19192	218.41352	3.39	0.0773

Bounds on condition number: 2.3619, 31.469

The above model is the best 4-variable model found.

Maximum R-Square Improvement: Step 7

Maximum R-Square Improvement: Step 7

Variable x5 Entered: R-Square = 0.6552 and C(p) = 5.4063

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	2815.15991	563.03198	9.12	<.0001
Error	24	1481.80676	61.74195		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	21.73314	12.05910	200.53735	3.25	0.0841
x1	0.30956	0.14616	276.96140	4.49	0.0447
x3	0.34146	0.18514	210.01982	3.40	0.0775
x4	0.46702	0.21000	305.34924	4.95	0.0358
x5	-0.19726	0.13773	126.64838	2.05	0.1650
x6	-0.28840	0.19345	137.22961	2.22	0.1490

 The above model is the best 5-variable model found.
 Maximum R-Square Improvement: Step 8
 Variable x2 Entered: R-Square = 0.6611 and C(p) = 7.0000

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	2840.88079	473.48013	7.48	0.0002
Error	23	1456.08588	63.30808		
Corrected Total	29	4296.96667			

Maximum R-Square Improvement: Step 8					
Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	20.14214	12.46359	165.34231	2.61	0.1197
x1	0.26828	0.16155	174.58215	2.76	0.1104
x2	0.10242	0.16068	25.72088	0.41	0.5302
x3	0.32199	0.18994	181.92831	2.87	0.1035
x4	0.46685	0.21265	305.13235	4.82	0.0385
x5	-0.18692	0.14040	112.21026	1.77	0.1961
x6	-0.30865	0.19845	153.14683	2.42	0.1335

Bounds on condition number: 2.3619, 71.204

 The above model is the best 6-variable model found.
 No further improvement in R-Square is possible.

۸۴ گزینش متغیر به روش **Minr**

این روش تقریباً شبیه روش **BACKWARD** عمل می کند.

```
data;
input y x1-x6;
cards;
;
proc reg;
model y=x1-x6/selection=minr;
run;quit;
```

Minimum R-Square Improvement: Step 1
 Variable x5 Entered: R-Square = 0.0010 and C(p) = 41.8072

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4.22088	4.22088	0.03	0.8694
Error	28	4292.74579	153.31235		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	67.00288	14.45858	3292.39982	21.48	<.0001
x5	-0.03225	0.19438	4.22088	0.03	0.8694

Minimum R-Square Improvement: Step 2
Variable x5 Removed: R-Square = 0.0427 and C(p) = 38.9751

Variable x6 Entered
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	183.52046	183.52046	1.25	0.2732
Error	28	4113.44620	146.90879		
Corrected Total	29	4296.96667			

Minimum R-Square Improvement: Step 2
Parameter Standard

Variable	Estimate	Error	Type II SS	F Value	Pr > F
Intercept	54.09849	9.68190	4586.65857	31.22	<.0001
x6	0.24330	0.21768	183.52046	1.25	0.2732

Minimum R-Square Improvement: Step 3
Variable x6 Removed: R-Square = 0.2676 and C(p) = 23.7116

Variable x2 Entered
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1149.81955	1149.81955	10.23	0.0034
Error	28	3147.14711	112.39811		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	36.35822	9.04976	1814.21910	16.14	0.0004
x2	0.52556	0.16432	1149.81955	10.23	0.0034

Minimum R-Square Improvement: Step 4
Variable x2 Removed: R-Square = 0.4076 and C(p) = 14.2106

Variable x1 Entered
Minimum R-Square Improvement: Step 4
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1751.31206	1751.31206	19.26	0.0001
Error	28	2545.65460	90.91624		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	28.20396	8.48083	1005.50531	11.06	0.0025
x1	0.55902	0.12737	1751.31206	19.26	0.0001

Minimum R-Square Improvement: Step 5
Variable x1 Removed: R-Square = 0.4085 and C(p) = 14.1499

Variable x4 Entered
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1755.15251	1755.15251	19.33	0.0001
Error	28	2541.81416	90.77908		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	17.20466	10.92577	225.09917	2.48	0.1266
x4	0.72855	0.16569	1755.15251	19.33	0.0001

Minimum R-Square Improvement: Step 6
Variable x4 Removed: R-Square = 0.4104 and C(p) = 14.0206

Variable x3 Entered
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1763.33890	1763.33890	19.49	0.0001

Error 28 2533.62776 90.48671
Corrected Total 29 4296.96667

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	27.30933	8.63151	905.80266	10.01	0.0037
x3	0.66021	0.14956	1763.33890	19.49	0.0001

The above model is the best 1-variable model found.

Minimum R-Square Improvement: Step 7

Variable x5 Entered: R-Square = 0.4128 and C(p) = 15.8534

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1773.92265	886.96133	9.49	0.0008
Error	27	2523.04401	93.44607		
Corrected Total	29	4296.96667			

Minimum R-Square Improvement: Step 7

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	30.98054	13.99776	457.74309	4.90	0.0355
x3	0.66167	0.15205	1769.70177	18.94	0.0002
x5	-0.05109	0.15182	10.58375	0.11	0.7391

Bounds on condition number: 1.0008, 4.0033

Minimum R-Square Improvement: Step 8

Variable x3 Removed: R-Square = 0.4310 and C(p) = 14.6195

Variable x4 Entered

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1852.03936	926.01968	10.23	0.0005
Error	27	2444.92730	90.55286		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	26.68766	14.25214	317.51444	3.51	0.0720
x4	0.76023	0.16829	1847.81848	20.41	0.0001
x5	-0.15715	0.15193	96.88685	1.07	0.3101

Minimum R-Square Improvement: Step 9

Variable x5 Removed: R-Square = 0.4504 and C(p) = 13.3066

Variable x6 Entered

Minimum R-Square Improvement: Step 9

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1935.15520	967.57760	11.06	0.0003
Error	27	2361.81146	87.47450		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	19.13970	10.80956	274.24318	3.14	0.0879
x4	0.89616	0.20027	1751.63474	20.02	0.0001
x6	-0.29669	0.20682	180.00269	2.06	0.1629

Minimum R-Square Improvement: Step 10

Variable x6 Removed: R-Square = 0.4719 and C(p) = 11.8411

Variable x2 Entered

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2027.93713	1013.96857	12.07	0.0002
Error	27	2269.02953	84.03813		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	11.36631	11.00042	89.72155	1.07	0.3106
x2	0.28816	0.15994	272.78462	3.25	0.0828
x4	0.58009	0.17946	878.11758	10.45	0.0032

Minimum R-Square Improvement: Step 10
 Bounds on condition number: 1.2672, 5.0687
 Minimum R-Square Improvement: Step 11
 Variable x2 Removed: R-Square = 0.4971 and C(p) = 10.1348
 Variable x3 Entered

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2135.95507	1067.97754	13.34	<.0001
Error	27	2161.01159	80.03747		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	13.20960	10.42123	128.59822	1.61	0.2158
x3	0.40250	0.18453	380.80256	4.76	0.0380
x4	0.44039	0.20410	372.61617	4.66	0.0400

Minimum R-Square Improvement: Step 12
 Variable x3 Removed: R-Square = 0.5281 and C(p) = 8.0319
 Variable x1 Entered

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2269.08570	1134.54285	15.11	<.0001
Error	27	2027.88096	75.10670		
Corrected Total	29	4296.96667			

Minimum R-Square Improvement: Step 12

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	10.35826	10.27685	76.30133	1.02	0.3224
x1	0.36127	0.13811	513.93319	6.84	0.0144
x4	0.47207	0.17980	517.77364	6.89	0.0141

Minimum R-Square Improvement: Step 13
 Variable x4 Removed: R-Square = 0.5393 and C(p) = 7.2706
 Variable x3 Entered

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2317.28741	1158.64370	15.80	<.0001
Error	27	1979.67926	73.32145		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	16.00455	8.79121	243.00796	3.31	0.0798
x1	0.36722	0.13360	553.94850	7.56	0.0105
x3	0.43688	0.15725	565.97534	7.72	0.0098

The above model is the best 2-variable model found.
 Minimum R-Square Improvement: Step 14
 Variable x2 Entered: R-Square = 0.5453 and C(p) = 8.8628
 Minimum R-Square Improvement: Step 14

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2343.10506	781.03502	10.39	0.0001
Error	26	1953.86160	75.14852		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	14.66335	9.18951	191.33852	2.55	0.1226
x1	0.33228	0.14781	379.78946	5.05	0.0333
x2	0.10074	0.17188	25.81766	0.34	0.5628
x3	0.40501	0.16822	435.57864	5.80	0.0235

Minimum R-Square Improvement: Step 15
Variable x2 Removed: R-Square = 0.5542 and C(p) = 8.2611
Variable x6 Entered

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2381.19237	793.73079	10.77	<.0001
Error	26	1915.77430	73.68363		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	19.34887	9.51646	304.60087	4.13	0.0524
x1	0.35056	0.13512	495.96291	6.73	0.0154
x3	0.52908	0.18614	595.26608	8.08	0.0086
x6	-0.17253	0.18527	63.90496	0.87	0.3603

Minimum R-Square Improvement: Step 15

Minimum R-Square Improvement: Step 16
Variable x6 Removed: R-Square = 0.5748 and C(p) = 6.8568
Variable x4 Entered

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2470.09801	823.36600	11.72	<.0001
Error	26	1826.86865	70.26418		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	8.49061	10.00117	50.64185	0.72	0.4037
x1	0.30137	0.13820	334.14294	4.76	0.0384
x3	0.30254	0.17887	201.01231	2.86	0.1027
x4	0.29800	0.20207	152.81061	2.17	0.1523

Minimum R-Square Improvement: Step 17
Variable x4 Removed: R-Square = 0.5787 and C(p) = 6.5948
Variable x5 Entered

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2486.68762	828.89587	11.90	<.0001
Error	26	1810.27905	69.62612		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	29.31976	12.09385	409.22655	5.88	0.0226
x1	0.44925	0.14041	712.76496	10.24	0.0036
x3	0.39329	0.15576	443.90622	6.38	0.0180
x5	-0.22046	0.14134	169.40021	2.43	0.1309

Minimum R-Square Improvement: Step 18
Variable x3 Removed: R-Square = 0.5957 and C(p) = 5.4409
Variable x4 Entered

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2559.73481	853.24494	12.77	<.0001
Error	26	1737.23186	66.81661		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	25.96209	12.24456	300.38447	4.50	0.0437
x1	0.44247	0.13596	707.69545	10.59	0.0031
x4	0.47170	0.16958	516.95342	7.74	0.0099
x5	-0.28408	0.13621	290.64911	4.35	0.0470

Minimum R-Square Improvement: Step 18
 Bounds on condition number: 1.5504, 12.301

The above model is the best 3-variable model found.

Minimum R-Square Improvement: Step 19
 Variable x2 Entered: R-Square = 0.6029 and C(p) = 6.9504

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2590.78745	647.69686	9.49	<.0001
Error	25	1706.17922	68.24717		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	23.90555	12.74498	240.10659	3.52	0.0724
x1	0.39756	0.15268	462.72380	6.78	0.0153
x2	0.10804	0.16017	31.05264	0.46	0.5062
x4	0.44552	0.17573	438.65696	6.43	0.0179
x5	-0.27218	0.13879	262.47671	3.85	0.0611

Minimum R-Square Improvement: Step 20
 Variable x2 Removed: R-Square = 0.6063 and C(p) = 6.7237

Variable x6 Entered

Minimum R-Square Improvement: Step 20

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2605.14009	651.28502	9.62	<.0001
Error	25	1691.82657	67.67306		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	26.45373	12.33739	311.13075	4.60	0.0419
x1	0.41613	0.14055	593.18206	8.77	0.0066
x4	0.57380	0.21134	498.86269	7.37	0.0118
x5	-0.26736	0.13859	251.84351	3.72	0.0651
x6	-0.15360	0.18751	45.40528	0.67	0.4205

Minimum R-Square Improvement: Step 21

Variable x6 Removed: R-Square = 0.6232 and C(p) = 5.5739
 Variable x3 Entered

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2677.93031	669.48258	10.34	<.0001
Error	25	1619.03636	64.76145		
Corrected Total	29	4296.96667			

The REG Procedure

Model: MODEL1

Dependent Variable: y

Minimum R-Square Improvement: Step 21

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	22.38356	12.34237	212.99896	3.29	0.0818
x1	0.38451	0.14056	484.63090	7.48	0.0113
x3	0.23717	0.17556	118.19549	1.83	0.1888
x4	0.33529	0.19511	191.24269	2.95	0.0981
x5	-0.24559	0.13709	207.83229	3.21	0.0853

Minimum R-Square Improvement: Step 22
 Variable x5 Removed: R-Square = 0.6257 and C(p) = 5.4068

Variable x6 Entered
 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2688.51153	672.12788	10.45	<.0001
Error	25	1608.45514	64.33821		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	11.04559	9.67009	83.94330	1.30	0.2642
x1	0.22954	0.13787	178.33978	2.77	0.1084
x3	0.41463	0.18165	335.21494	5.21	0.0312
x4	0.46852	0.21437	307.31916	4.78	0.0384
x6	-0.35362	0.19192	218.41352	3.39	0.0773

The above model is the best 4-variable model found.

Minimum R-Square Improvement: Step 23
 Minimum R-Square Improvement: Step 23
 Variable x2 Entered: R-Square = 0.6350 and C(p) = 6.7724

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	2728.67053	545.73411	8.35	0.0001
Error	24	1568.29614	65.34567		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	9.76579	9.88129	63.82697	0.98	0.3329
x1	0.18350	0.15084	96.70317	1.48	0.2356
x2	0.12712	0.16215	40.15900	0.61	0.4407
x3	0.38572	0.18675	278.76654	4.27	0.0498
x4	0.46822	0.21604	306.92096	4.70	0.0404
x6	-0.37451	0.19525	240.41741	3.68	0.0671

Minimum R-Square Improvement: Step 24
 Variable x2 Removed: R-Square = 0.6552 and C(p) = 5.4063

Variable x5 Entered
 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	2815.15991	563.03198	9.12	<.0001
Error	24	1481.80676	61.74195		
Corrected Total	29	4296.96667			

The SAS System 12:01 Saturday, November 3, 2001 29
 The REG Procedure
 Model: MODEL1

Dependent Variable: y

Minimum R-Square Improvement: Step 24

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	21.73314	12.05910	200.53735	3.25	0.0841
x1	0.30956	0.14616	276.96140	4.49	0.0447
x3	0.34146	0.18514	210.01982	3.40	0.0775
x4	0.46702	0.21000	305.34924	4.95	0.0358
x5	-0.19726	0.13773	126.64838	2.05	0.1650
x6	-0.28840	0.19345	137.22961	2.22	0.1490

Bounds on condition number: 2.3619, 48.363

The above model is the best 5-variable model found.

Minimum R-Square Improvement: Step 25
 Variable x2 Entered: R-Square = 0.6611 and C(p) = 7.0000

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	2840.88079	473.48013	7.48	0.0002
Error	23	1456.08588	63.30808		
Corrected Total	29	4296.96667			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	20.14214	12.46359	165.34231	2.61	0.1197
x1	0.26828	0.16155	174.58215	2.76	0.1104
x2	0.10242	0.16068	25.72088	0.41	0.5302
x3	0.32199	0.18994	181.92831	2.87	0.1035
x4	0.46685	0.21265	305.13235	4.82	0.0385
x5	-0.18692	0.14040	112.21026	1.77	0.1961
x6	-0.30865	0.19845	153.14683	2.42	0.1335

Bounds on condition number: 2.3619, 71.204

Minimum R-Square Improvement: Step 25
The above model is the best 6-variable model found.
No further improvement in R-Square is possible.

۸۵ گزینش متغیر به روش Rsquare

در این برنامه و دو برنامه بعد، کلمات اختیاری **ADIRSQ**، **CP** و **MSE** در گزاره **MODEL** از **SAS** می خواهند که برای هر مدل گزینش شده، مقادیر R^2 تعدیل شده، آماره مالو و میانگین مربعات اشتباه را چاپ کند.

```
data;
input y x1-x6;
cards;
;
proc reg;
model y=x1-x6/selection=rsqure adjrsq cp mse;
run;quit;
```

Number in Model	R-Square Selection Method				
	R-Square	Adjusted R-Square	C(p)	MSE	Variables in Model
1	0.4104	0.3893	14.0206	90.48671	x3
1	0.4085	0.3873	14.1499	90.77908	x4
1	0.4076	0.3864	14.2106	90.91624	x1
1	0.2676	0.2414	23.7116	112.39811	x2
1	0.0427	0.0085	38.9751	146.90879	x6
1	0.0010	-0.0347	41.8072	153.31235	x5
2	0.5393	0.5052	7.2706	73.32145	x1 x3
2	0.5281	0.4931	8.0319	75.10670	x1 x4
2	0.4971	0.4598	10.1348	80.03747	x3 x4
2	0.4754	0.4365	11.6066	83.48834	x1 x5
2	0.4719	0.4328	11.8411	84.03813	x2 x4
2	0.4569	0.4167	12.8618	86.43152	x2 x3
2	0.4504	0.4096	13.3066	87.47450	x4 x6
2	0.4439	0.4027	13.7431	88.49779	x1 x2
2	0.4387	0.3972	14.0953	89.32360	x3 x6
2	0.4310	0.3889	14.6195	90.55286	x4 x5
2	0.4156	0.3723	15.6638	93.00150	x1 x6
2	0.4128	0.3693	15.8534	93.44607	x3 x5
2	0.2739	0.2201	25.2842	115.55875	x2 x5
2	0.2686	0.2144	25.6456	116.40625	x2 x6
2	0.0473	-0.0232	40.6617	151.61511	x5 x6
3	0.5957	0.5491	5.4409	66.81661	x1 x4 x5
3	0.5842	0.5362	6.2238	68.72288	x3 x4 x6
3	0.5787	0.5301	6.5948	69.62612	x1 x3 x5
3	0.5748	0.5258	6.8568	70.26418	x1 x3 x4
3	0.5542	0.5027	8.2611	73.68363	x1 x3 x6
3	0.5477	0.4955	8.7018	74.75654	x1 x4 x6
3	0.5453	0.4928	8.8628	75.14852	x1 x2 x3
3	0.5418	0.4890	9.0964	75.71754	x1 x2 x4
3	0.5258	0.4711	10.1842	78.36626	x2 x4 x6
3	0.5237	0.4687	10.3279	78.71610	x2 x3 x4
3	0.5104	0.4539	11.2290	80.91028	x3 x4 x5
3	0.5008	0.4433	11.8793	82.49370	x1 x2 x5
3	0.4952	0.4370	12.2595	83.41935	x2 x4 x5
3	0.4920	0.4334	12.4814	83.95966	x2 x3 x6
3	0.4902	0.4314	12.6036	84.25728	x1 x5 x6

3	0.4682	0.4069	14.0935	87.88495	x4 x5 x6
3	0.4617	0.3996	14.5344	88.95856	x2 x3 x5
3	0.4452	0.3812	15.6583	91.69522	x1 x2 x6
3	0.4391	0.3744	16.0720	92.70256	x3 x5 x6
3	0.2758	0.1922	27.1551	119.68903	x2 x5 x6

Number in Model	R-Square Selection Method				
	R-Square	Adjusted R-Square	C (p)	MSE	Variables in Model
4	0.6257	0.5658	5.4068	64.33821	x1 x3 x4 x6
4	0.6232	0.5629	5.5739	64.76145	x1 x3 x4 x5
4	0.6125	0.5505	6.2999	66.59997	x2 x3 x4 x6
4	0.6063	0.5433	6.7237	67.67306	x1 x4 x5 x6
4	0.6029	0.5394	6.9504	68.24717	x1 x2 x4 x5
4	0.5907	0.5252	7.7811	70.35073	x3 x4 x5 x6
4	0.5841	0.5175	8.2295	71.48624	x1 x3 x5 x6
4	0.5827	0.5160	8.3208	71.71753	x1 x2 x3 x5
4	0.5791	0.5117	8.5700	72.34854	x1 x2 x3 x4
4	0.5701	0.5014	9.1758	73.88251	x1 x2 x4 x6
4	0.5636	0.4938	9.6205	75.00868	x1 x2 x3 x6
4	0.5438	0.4708	10.9647	78.41273	x2 x4 x5 x6
4	0.5391	0.4654	11.2806	79.21256	x2 x3 x4 x5
4	0.5065	0.4276	13.4932	84.81572	x1 x2 x5 x6
4	0.4933	0.4122	14.3922	87.09222	x2 x3 x5 x6

5	0.6552	0.5833	5.4063	61.74195	x1 x3 x4 x5 x6
5	0.6350	0.5590	6.7724	65.34567	x1 x2 x3 x4 x6
5	0.6255	0.5475	7.4191	67.05136	x1 x2 x3 x4 x5
5	0.6205	0.5414	7.7577	67.94450	x2 x3 x4 x5 x6
5	0.6188	0.5394	7.8737	68.25059	x1 x2 x4 x5 x6
5	0.5901	0.5047	9.8198	73.38409	x1 x2 x3 x5 x6

6	0.6611	0.5727	7.0000	63.30808	x1 x2 x3 x4 x5 x6

۸۶. گزینش متغیر به روش Adjrsq

```
data;
input y x1-x6;
cards;
;
proc reg;
model y=x1-x6/selection=adjrsq adjrsq cp mse;
run;quit;
```

Number in Model	Adjusted R-Square Selection Method				
	Adjusted R-Square	R-Square	C (p)	MSE	Variables in Model
5	0.5833	0.6552	5.4063	61.74195	x1 x3 x4 x5 x6
6	0.5727	0.6611	7.0000	63.30808	x1 x2 x3 x4 x5 x6
4	0.5658	0.6257	5.4068	64.33821	x1 x3 x4 x6
4	0.5629	0.6232	5.5739	64.76145	x1 x3 x4 x5
5	0.5590	0.6350	6.7724	65.34567	x1 x2 x3 x4 x6
4	0.5505	0.6125	6.2999	66.59997	x2 x3 x4 x6
3	0.5491	0.5957	5.4409	66.81661	x1 x4 x5
5	0.5475	0.6255	7.4191	67.05136	x1 x2 x3 x4 x5
4	0.5433	0.6063	6.7237	67.67306	x1 x4 x5 x6
5	0.5414	0.6205	7.7577	67.94450	x2 x3 x4 x5 x6
4	0.5394	0.6029	6.9504	68.24717	x1 x2 x4 x5
5	0.5394	0.6188	7.8737	68.25059	x1 x2 x4 x5 x6
3	0.5362	0.5842	6.2238	68.72288	x3 x4 x6
3	0.5301	0.5787	6.5948	69.62612	x1 x3 x5
3	0.5258	0.5748	6.8568	70.26418	x1 x3 x4
4	0.5252	0.5907	7.7811	70.35073	x3 x4 x5 x6
4	0.5175	0.5841	8.2295	71.48624	x1 x3 x5 x6
4	0.5160	0.5827	8.3208	71.71753	x1 x2 x3 x5
4	0.5117	0.5791	8.5700	72.34854	x1 x2 x3 x4
2	0.5052	0.5393	7.2706	73.32145	x1 x3
5	0.5047	0.5901	9.8198	73.38409	x1 x2 x3 x5 x6
3	0.5027	0.5542	8.2611	73.68363	x1 x3 x6
4	0.5014	0.5701	9.1758	73.88251	x1 x2 x4 x6
3	0.4955	0.5477	8.7018	74.75654	x1 x4 x6

4	0.4938	0.5636	9.6205	75.00868	x1 x2 x3 x6
2	0.4931	0.5281	8.0319	75.10670	x1 x4
3	0.4928	0.5453	8.8628	75.14852	x1 x2 x3
3	0.4890	0.5418	9.0964	75.71754	x1 x2 x4
3	0.4711	0.5258	10.1842	78.36626	x2 x4 x6
4	0.4708	0.5438	10.9647	78.41273	x2 x4 x5 x6
3	0.4687	0.5237	10.3279	78.71610	x2 x3 x4
4	0.4654	0.5391	11.2806	79.21256	x2 x3 x4 x5
2	0.4598	0.4971	10.1348	80.03747	x3 x4
3	0.4539	0.5104	11.2290	80.91028	x3 x4 x5
3	0.4433	0.5008	11.8793	82.49370	x1 x2 x5
3	0.4370	0.4952	12.2595	83.41935	x2 x4 x5
2	0.4365	0.4754	11.6066	83.48834	x1 x5
3	0.4334	0.4920	12.4814	83.95966	x2 x3 x6
2	0.4328	0.4719	11.8411	84.03813	x2 x4
3	0.4314	0.4902	12.6036	84.25728	x1 x5 x6
4	0.4276	0.5065	13.4932	84.81572	x1 x2 x5 x6
2	0.4167	0.4569	12.8618	86.43152	x2 x3
4	0.4122	0.4933	14.3922	87.09222	x2 x3 x5 x6
2	0.4096	0.4504	13.3066	87.47450	x4 x6
3	0.4069	0.4682	14.0935	87.88495	x4 x5 x6
2	0.4027	0.4439	13.7431	88.49779	x1 x2
3	0.3996	0.4617	14.5344	88.95856	x2 x3 x5
2	0.3972	0.4387	14.0953	89.32360	x3 x6
1	0.3893	0.4104	14.0206	90.48671	x3
2	0.3889	0.4310	14.6195	90.55286	x4 x5
1	0.3873	0.4085	14.1499	90.77908	x4
1	0.3864	0.4076	14.2106	90.91624	x1
3	0.3812	0.4452	15.6583	91.69522	x1 x2 x6
3	0.3744	0.4391	16.0720	92.70256	x3 x5 x6
2	0.3723	0.4156	15.6638	93.00150	x1 x6
2	0.3693	0.4128	15.8534	93.44607	x3 x5
1	0.2414	0.2676	23.7116	112.39811	x2
2	0.2201	0.2739	25.2842	115.55875	x2 x5
2	0.2144	0.2686	25.6456	116.40625	x2 x6
3	0.1922	0.2758	27.1551	119.68903	x2 x5 x6
1	0.0085	0.0427	38.9751	146.90879	x6
2	-.0232	0.0473	40.6617	151.61511	x5 x6
1	-.0347	0.0010	41.8072	153.31235	x5

۸۷. گزینش متغیر به روش Cp

```
data;
input y x1-x6;
cards;
;
proc reg;
model y=x1-x6/selection=cp adjrsd cp mse;
run;quit;
```

Number in Model	C(p) Selection Method			MSE	Variables in Model
	C(p)	R-Square	Adjusted R-Square		
5	5.4063	0.6552	0.5833	61.74195	x1 x3 x4 x5 x6
4	5.4068	0.6257	0.5658	64.33821	x1 x3 x4 x6
3	5.4409	0.5957	0.5491	66.81661	x1 x4 x5
4	5.5739	0.6232	0.5629	64.76145	x1 x3 x4 x5
3	6.2238	0.5842	0.5362	68.72288	x3 x4 x6
4	6.2999	0.6125	0.5505	66.59997	x2 x3 x4 x6
3	6.5948	0.5787	0.5301	69.62612	x1 x3 x5
4	6.7237	0.6063	0.5433	67.67306	x1 x4 x5 x6
5	6.7724	0.6350	0.5590	65.34567	x1 x2 x3 x4 x6
3	6.8568	0.5748	0.5258	70.26418	x1 x3 x4
4	6.9504	0.6029	0.5394	68.24717	x1 x2 x4 x5
6	7.0000	0.6611	0.5727	63.30808	x1 x2 x3 x4 x5 x6
2	7.2706	0.5393	0.5052	73.32145	x1 x3
5	7.4191	0.6255	0.5475	67.05136	x1 x2 x3 x4 x5
5	7.7577	0.6205	0.5414	67.94450	x2 x3 x4 x5 x6
4	7.7811	0.5907	0.5252	70.35073	x3 x4 x5 x6
5	7.8737	0.6188	0.5394	68.25059	x1 x2 x4 x5 x6
2	8.0319	0.5281	0.4931	75.10670	x1 x4
4	8.2295	0.5841	0.5175	71.48624	x1 x3 x5 x6
3	8.2611	0.5542	0.5027	73.68363	x1 x3 x6
4	8.3208	0.5827	0.5160	71.71753	x1 x2 x3 x5

4	8.5700	0.5791	0.5117	72.34854	x1 x2 x3 x4
3	8.7018	0.5477	0.4955	74.75654	x1 x4 x6
3	8.8628	0.5453	0.4928	75.14852	x1 x2 x3
3	9.0964	0.5418	0.4890	75.71754	x1 x2 x4
4	9.1758	0.5701	0.5014	73.88251	x1 x2 x4 x6
4	9.6205	0.5636	0.4938	75.00868	x1 x2 x3 x6
5	9.8198	0.5901	0.5047	73.38409	x1 x2 x3 x5 x6
2	10.1348	0.4971	0.4598	80.03747	x3 x4
3	10.1842	0.5258	0.4711	78.36626	x2 x4 x6
3	10.3279	0.5237	0.4687	78.71610	x2 x3 x4
4	10.9647	0.5438	0.4708	78.41273	x2 x4 x5 x6
3	11.2290	0.5104	0.4539	80.91028	x3 x4 x5
4	11.2806	0.5391	0.4654	79.21256	x2 x3 x4 x5
2	11.6066	0.4754	0.4365	83.48834	x1 x5
2	11.8411	0.4719	0.4328	84.03813	x2 x4
3	11.8793	0.5008	0.4433	82.49370	x1 x2 x5
3	12.2595	0.4952	0.4370	83.41935	x2 x4 x5
3	12.4814	0.4920	0.4334	83.95966	x2 x3 x6
3	12.6036	0.4902	0.4314	84.25728	x1 x5 x6
2	12.8618	0.4569	0.4167	86.43152	x2 x3
2	13.3066	0.4504	0.4096	87.47450	x4 x6
4	13.4932	0.5065	0.4276	84.81572	x1 x2 x5 x6
2	13.7431	0.4439	0.4027	88.49779	x1 x2
1	14.0206	0.4104	0.3893	90.48671	x3
3	14.0935	0.4682	0.4069	87.88495	x4 x5 x6
2	14.0953	0.4387	0.3972	89.32360	x3 x6
1	14.1499	0.4085	0.3873	90.77908	x4
1	14.2106	0.4076	0.3864	90.91624	x1
4	14.3922	0.4933	0.4122	87.09222	x2 x3 x5 x6
3	14.5344	0.4617	0.3996	88.95856	x2 x3 x5
2	14.6195	0.4310	0.3889	90.55286	x4 x5
3	15.6583	0.4452	0.3812	91.69522	x1 x2 x6
2	15.6638	0.4156	0.3723	93.00150	x1 x6
2	15.8534	0.4128	0.3693	93.44607	x3 x5
3	16.0720	0.4391	0.3744	92.70256	x3 x5 x6
1	23.7116	0.2676	0.2414	112.39811	x2
2	25.2842	0.2739	0.2201	115.55875	x2 x5
2	25.6456	0.2686	0.2144	116.40625	x2 x6
3	27.1551	0.2758	0.1922	119.68903	x2 x5 x6
1	38.9751	0.0427	0.0085	146.90879	x6
2	40.6617	0.0473	-0.0232	151.61511	x5 x6
1	41.8072	0.0010	-0.0347	153.31235	x5

مدلهای رگرسیون غیر خطی

مدلهای رگرسیونی که تا به حال مورد بحث قرار گرفته اند همگی از نوع خطی بودند. گاهی بین Y و X یک رابطه غیر خطی وجود دارد. هرچند برخی از این رابطه های غیر خطی را به کمک تبدیلاتی می توان به خمدلهای خطی تبدیل نمود، اما گاهی این رابطه غیر خطی بوده و به سادگی به رابطه خطی تبدیل نمی شود. در چنین مواردی فقط رویه **NLIN** می تواند مقادیر مجهول پارامترهای مدل را برآورد کند. در اینجا نحوه تجزیه و تحلیل یک مدل لجستیک و یک مدل تکه ای را در **SAS** با ذکر مثال مورد بررسی قرار می دهیم.

۸۸. مدل های رگرسیون غیر خطی لجستیک (مدلهای لجستیک) (logistic models)

فرض کنیم یک محقق می خواهد رابطه بین رشد و گسترش سطح برگ یک گیاه و درجه روز رشد را بداند. بدین منظور ۹ اندازه گیری بر روی سطح برگ گیاه (**TPLA** متغیر وابسته) و درجه روز رشد (**TTLF** متغیر مستقل) انجام می دهد. ملاحظات عملی حاکی از آن است که یک مدل لجستیک می تواند این رابطه را به خوبی توجیه کند.

بنابراین محقق می خواهد این مدل را بر داده های خود برازش دهد و پارامترهای آن را برآورد کند. مدل لجستیک به شرح زیر می باشد:

$$TPLA = \frac{TPLAMAX}{1 + EXP(-a_1(TTLF - b_1))}$$

```
data;
input tpla ttlf;
cards;
15      395
.
975     1290
;
proc nlin method=dud;
parms tplamx=1200 a1=0.01 b1=700;
model tpla=tplamx/(1+exp(-a1*(ttlf-b1)));
output p=yhat r=yresid;
proc plot;
plot tpla*ttlf=,a, yhat*ttlf=,p,/overlay;
plot yresid*(yhat ttlf);
run;quit;
```

The NLIN Procedure

NOTE: An intercept was not specified for this model.

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Regression	3	4202247	1400749	607.26	<.0001
Residual	6	13840.1	2306.7		
Uncorrected Total	9	4216087			
Corrected Total	8	1551575			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
tplamx	1000.8	30.3950	926.4	1075.2
a1	0.0147	0.00218	0.00934	0.0200
b1	815.9	11.6323	787.5	844.4

همانطور که در جدول فوق ملاحظه می گردد، متغیرهای لازم برای ارائه مدل لجستیک محاسبه شده است. لذا می توان با جایگذاری آنها در فرمول بالا، مدل لجستیکی که بر این داده ها برازش می شود را به شرح زیر ارائه داد:

$$TPLA = \frac{1000.8}{1 + EXP(-0.0147(TTLF - 815.9))}$$

۸۹. مدل های رگرسیون غیر خطی تکه ای (مدل های تکه ای) (Semented models)

گاهی اوقات با تغییر متغیر مستقل، مقدار متغیر وابسته به طور ناگهانی تغییر می کند. به عبارت دیگر نمودار پراکنش بین Y روی X نشان می دهد که با دو یا چند خط شکسته، می توان رابطه بین Y و X را نشان داد. در SAS رویه PROC NLIN از SAS می خواهد که یک مدل غیر خطی را با روش DUD برازش دهد.

مثال: فرض کنیم کمی کردن وزن دانه یک گیاه بعد از گرده افشانی آم مورد نظر باشد و بدین منظور در روزهای مختلف پس از گرده اتشانی (DAA)، وزن دانه (SW) اندازه گیری شده باشد. اگر نمودار پراکنش SW بر روی DAA با

استفاده از رویه **PLOT** رسم شود، ملاحظه خواهد شد که تا تعداد روز معینی، وزن دانه به طور خطی افزایش و بعد به ناگهان در حد ثابتی باقی می ماند. برای چنین حالاتی می توان یک مدل تکه ای بر داده ها برازش داد. در این مورد مدل به صورت زیر است:

$$\begin{cases} Y = a + bX & X < X_0 \\ Y = a + bX_0 & X \geq X_0 \end{cases}$$

گزاره **PARMS** پارامترهای مدل یعنی **a**، **GFR** و **PM** را همراه با مقادیر اولیه آنها به رویه معرفی می کند. سپس از رویه خواسته می شود که برای مقادیر متغیر **DAA** که کوچکتر از **PM** هستند، مدل $SW = a + GFR \times DAA$ را برازش دهد و برای مقادیر متغیر **DAA** که بزرگتر از **PM** هستند، مدل $SW = a + GFR \times PM$ را امتحان کند.

```
data;
input daa sw;
cards;
;
proc nlin method=dud;
parms a=-8 gfr=1 pm=34;
if daa<pm then do;
model sw=a+gfr*daa;
end;
else do;
model sw=a+gfr*pm;
end;
output p=yhat r=yresid;
proc plot;
plot sw*daa=a,yhat*daa=p,/overlay;
plot yresid*(yhat daa);
run;quit;
```

The NLIN Procedure
Dependent Variable sw
DUD Initialization

DUD	a	gfr	pm	Sum of Squares
-4	-8.0000	1.0000	34.0000	270.2
-3	-8.8000	1.0000	34.0000	344.7
-2	-8.0000	1.1000	34.0000	128.0
-1	-8.0000	1.0000	37.4000	215.3

Iterative Phase

Iter	a	gfr	pm	Sum of Squares
0	-8.0000	1.1000	34.0000	128.0
1	3.1402	0.7385	37.0946	12.3268
2	3.1893	0.7359	37.9406	12.0019
3	3.2085	0.7372	37.6747	11.8692
4	2.9483	0.7544	37.2203	11.4918
5	2.6457	0.7669	36.9675	11.4195
6	2.6457	0.7669	36.9749	11.4194
7	2.6457	0.7669	36.9747	11.4194

NOTE: Convergence criterion met.

Estimation Summary

Method	DUD
Iterations	7
Objective	4.496E-9
Objective	11.41943
Observations Read	8
Observations Used	8
Observations Missing	0

NOTE: An intercept was not specified for this model.

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Regression	2	4555.3	2277.7	1196.73	<.0001
Residual	6	11.4194	1.9032		
Uncorrected Total	8	4566.8			
Corrected Total	7	453.5			

The NLIN Procedure

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
a	2.6457	1.5683	-1.1918	6.4832
gfr	0.7669	.	.	.
pm	36.9747	1.6734	32.8800	41.0693

در نتیجه مدل آزمایش تکه ای فوق به صورت زیر می باشد:

$$SW = 2.6457 + 0.7667DAA \quad DAA < 36.9747$$

$$SW = 2.65 + 0.7667PM \quad DAA \geq 36.9747$$

تجزیه همبستگی

در نرم افزار آماری SAS برای تجزیه و تحلیل همبستگی، از دستور PROC CORR استفاده می شود. شکل کلی این دستور به صورت زیر می باشد:

```
PROC CORR options;
VAR variables name;
Run;
```

به جای عبارت options می توان از دستورات زیر استفاده نمود:

Spearman: تجزیه همبستگی را با استفاده از ضرایب اسپیرمن انجام می دهد.

PEARSON: تجزیه همبستگی را با استفاده از ضرایب پیرسون انجام می دهد. توجه اینکه به کارگیری این دستور

اختیاری است زیرا در صورتیکه روش تجزیه و تحلیل همبستگی را مشخص نماییم، sas خود به خود از روش همبستگی پیرسون استفاده خواهد نمود.

NOSIMPLE: اگر بخواهیم خلاصه آماری چاپ نشود، از این دستور استفاده می کنیم.

NOPROP: اگر بخواهیم مقادیر احتمال (P-VALUE) چاپ نشوند، از این دستور استفاده می کنیم.

COV: ماتریس کواریانس را هم چاپ می کند.

۹۰. تجزیه همبستگی به روش پیرسون

```
data;
input yimport xport xinvest xconsum;
cards;
15.9      149.3      4.2      108.1
16.4      161.2      4.1      114.8
19        171.5      3.1      123.2
```

```

19.1    175.5    3.1    126.9
18.8    180.8    1.1    132.1
20.4    190.7    2.2    137.7
22.7    202.1    2.1    146
25.5    212.4    5.6    154.1
28.1    226.1    5      162.3
27.6    231.9    5.1    164.3
26.3    239      0.7    167.6

```

```

;
proc corr PEARSON COV;
var yimport xport xinvest xconsum;
run;quit;

```

```

The CORR Procedure
4 Variables:  yimport  xport  xinvest  xconsum
Covariance Matrix, DF = 10

```

	yimport	xport	xinvest	xconsum
yimport	19.8140000	129.7940000	1.6490000	89.6820000
xport	129.7940000	899.9709091	1.2790000	617.3263636
xinvest	1.6490000	1.2790000	2.7200000	1.2140000
xconsum	89.6820000	617.3263636	1.2140000	425.7785455

```

Simple Statistics

```

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
yimport	11	21.80000	4.45129	239.80000	15.90000	28.10000
xport	11	194.59091	29.99952	2141	149.30000	239.00000
xinvest	11	3.30000	1.64924	36.30000	0.70000	5.60000
xconsum	11	139.73636	20.63440	1537	108.10000	167.60000

```

Pearson Correlation Coefficients, N = 11
Prob > |r| under H0: Rho=0

```

	yimport	xport	xinvest	xconsum
yimport	1.00000	0.97197 <.0001	0.22462 0.5067	0.97640 <.0001
xport	0.97197 <.0001	1.00000	0.02585 0.9399	0.99726 <.0001
xinvest	0.22462 0.5067	0.02585 0.9399	1.00000	0.03567 0.9171
xconsum	0.97640 <.0001	0.99726 <.0001	0.03567 0.9171	1.00000

۹۱. تجزیه همبستگی به روش پیرسون با اضافه نمودن دستورات فرعی **NOSIMPLE** و **NOPROP**

```

data;
input yimport xport xinvest xconsum;
cards;
;
proc corr PEARSON NOSIMPLE NOPROP;
var yimport xport xinvest xconsum;
run;quit;

```

```

The CORR Procedure
4 Variables:  yimport  xport  xinvest  xconsum
Pearson Correlation Coefficients, N = 11

```

	yimport	xport	xinvest	xconsum
yimport	1.00000	0.97197	0.22462	0.97640
xport	0.97197	1.00000	0.02585	0.99726
xinvest	0.22462	0.02585	1.00000	0.03567
xconsum	0.97640	0.99726	0.03567	1.00000

۹۲. تجزیه همبستگی به روش اسپیرمن

```
data;
input yimport xport xinvest xconsum;
cards;
;
proc corr SPEARMAN NOSIMPLE;
var yimport xport xinvest xconsum;
run;quit;
```

The CORR Procedure

4 Variables: yimport xport xinvest xconsum

Spearman Correlation Coefficients, N = 11
Prob > |r| under H0: Rho=0

	yimport	xport	xinvest	xconsum
yimport	1.00000	0.93636 <.0001	0.20501 0.5454	0.93636 <.0001
xport	0.93636 <.0001	1.00000	-0.00456 0.9894	1.00000 <.0001
xinvest	0.20501 0.5454	-0.00456 0.9894	1.00000	-0.00456 0.9894
xconsum	0.93636 <.0001	1.00000 <.0001	-0.00456 0.9894	1.00000

بخش چهارم

کاربرد SAS در تجزیه های چند متغیره

تجزیه به مؤلفه های اصلی

هدف از تجزیه و تحلیل مؤلفه های اصلی تشریح تغییرات مجموعه ای از داده های چندمتغیره در قالب مجموعه ای از متغیرهای غیر همبسته است که ترکیب خطی از متغیرهای اصلی هستند. به این ترکیبات خطی غیر همبسته مؤلفه هم گفته می شود. اولین مؤلفه برای حداکثر تغییرات و به همین ترتیب آخرین مؤلفه برای حداقل تغییرات خواهد بود.

هدف از تجزیه و تحلیل مؤلفه های اصلی این است که چند مؤلفه اصلی را به عنوان ملاک تشخیص تهیه نموده و با توجه به اینکه این مؤلفه ها تقریباً ۱۰۰ درصد تغییرات موجود بین متغیرهای اصلی را می شمارند، بدون اینکه اطلاعات زیادی از دست بدهیم، آنها را به جای متغیرهای اصلی در تجزیه و تحلیل های بعدی بکار می بریم.

در برنامه **sas** تجزیه به مؤلفه های اصلی با استفاده از رویه **PRINCOMP** انجام می گیرد. ساختار کلی این برنامه به صورت زیر است:

```
PROC PRINCOMP option;
VAR variables;
By variables;
```

برخی از کلمات اختیاری که در گزاره **PROC PRINCOMP** مورد استفاده قرار می گیرند به شرح زیر می باشند:

۱- دستور **COV**: از رویه می خواهد که تجزیه به مؤلفه ها را از طریق ماتریس کواریانس انجام دهد. اگر این کلمه نوشته نشود، ماتریس همبستگی مورد تجزیه قرار خواهد گرفت. از این کلمه اختیاری نباید استفاده شود، مگر اینکه متغیرها با یکدیگر قابل مقایسه باشند یا به نوعی استاندارد شده باشند.

۲- دستور **N=n**: تعداد مؤلفه های اصلی را که باید محاسبه شوند، مشخص می کند. در صورت عدم استفاده از این کلمه اختیاری، تعداد مؤلفه ها برابر با تعداد متغیرها خواهد بود.

۳- دستور **NOINT**: از رویه می خواهد که ماتریس کواریانس یا همبستگی برای میانگین تصحیح نشود، یعنی در مدل از عرض از مبدا استفاده نشود.

۴- دستور **std** یا **STANDARD**: از رویه می خواهد که مقادیر مؤلفه های اصلی در مجموعه داده خروجی استاندارد شوند تا واریانس برابر با ۱ پیدا کنند.

۵- دستور **OUT=name** که نتایج تجزیه و تحلیل یا خروجی را که شامل میانگینها، انحراف معیارها، تعداد شواهدات، همبستگیها یا کواریانسها، مقادیر ویژه و بردارهای ویژه را در یک مجموعه داده ها (**SAS dataset**) قرار می دهد.

همچنین با کمک رویه **PROC CORR** همبستگی بین متغیرهای اصلی و مؤلفه های اول و دوم و سوم را چاپ می کند. هدف از این کار این است که بدانیم، هر یک از مؤلفه ها بیشتر به کدام متغیر بستگی دارد.

برای اینکه بدانیم که آیا مؤلفه ها نرمال هستند یا خیر؟ یک روش این است که باید آنها را در مقابل همدیگر رسم کنیم تا ببینیم که آیا درون یک بیضی قرار می گیرند یا خیر؟ چنانچه درون یک بیضی قرار گرفتند، نرمال دو متغیره هستند. روش دیگر استفاده از نمودار نرمال است. برنامه رسم نمودار نرمال به صورت زیر است:

```
PROC RANK NORMAL=VW;
```

```
VAR نام متغیرها (PRIN1-PRIN3) ;
```

```
RANKS NSCORE;
```

۹۳. تجزیه به مؤلفه های اصلی با استفاده از ماتریس کواریانس

```
TITLE "PRINCIPAL COMPONENT";
```

```
data;
```

```
input x1-x6;
```

```
cards;
```

```
1.103 1.052 2.139 2.238 0.873 0.872
```

```
0.842 0.859 1.873 1.741 0.590 0.744
```

```
0.925 0.873 1.887 1.809 0.767 0.713
```

```
0.857 0.744 1.739 1.547 0.706 0.674
```

```
0.795 0.809 1.734 1.715 0.549 0.654
```

```
0.787 0.779 1.509 1.474 0.782 0.571
```

```
0.933 0.880 1.695 1.656 0.737 0.803
```

```
0.799 0.851 1.740 1.777 0.618 0.682
```

```
0.945 0.876 1.811 1.759 0.853 0.777
```

```
0.921 0.906 1.954 2.009 0.823 0.765
```

```
0.792 0.825 1.624 1.657 0.686 0.668
```

```
0.815 0.751 2.204 1.846 0.678 0.546
```

```
0.755 0.724 1.508 1.458 0.662 0.595
```

```
0.880 0.866 1.786 1.811 0.810 0.819
```

```
0.900 0.834 1.902 1.606 0.723 0.677
```

```
0.764 0.757 1.743 1.794 0.586 0.541
```

```
0.733 0.748 1.863 1.869 0.672 0.752
```

```
0.932 0.898 2.028 2.032 0.836 0.805
```

```
0.856 0.786 1.390 1.324 0.578 0.610
```

```
0.890 0.950 2.187 2.087 0.758 0.718
```

```
;
```

```
proc princomp COV out=prin;
var x1-x6;
PROC CORR NOPROB NOSIMPLE;
VAR X1 X2 X3 X4 X5 X6 PRIN1 PRIN2 PRIN3;
proc plot;
plot prin*prin2="1" PRIN1*PRIN3="1" PRIN2*PRIN3="1"/BOX;
run;quit;
```

The PRINCOMP Procedure
Observations 20
Variables 6

Simple Statistics						
	x1	x2	x3	x4	x5	x6
Mean	0.861200000	0.838400000	1.815800000	1.760450000	0.714350000	0.699300000
Std	0.0871318420	0.0806750467	0.220188101	0.223972619	0.0983732286	0.0946150650

خروجی شماره ۱ خلاصه آماری شامل میانگین و انحراف معیار را برای متغیرهای اصلی نشان می دهد.

Covariance Matrix						
	x1	x2	x3	x4	x5	x6
x1	0.0075919579	0.0060146000	0.0091454632	0.0102274842	0.0060854526	0.0059046737
x2	0.0060146000	0.0065084632	0.0097070316	0.0130491263	0.0048818526	0.0058643474
x3	0.0091454632	0.0097070316	0.0484828000	0.0425835684	0.0091050211	0.0082296947
x4	0.0102274842	0.0130491263	0.0425835684	0.0501637342	0.0110816237	0.0124703316
x5	0.0060854526	0.0048818526	0.0091050211	0.0110816237	0.0096772921	0.0058223105
x6	0.0059046737	0.0058643474	0.0082296947	0.0124703316	0.0058223105	0.0089520105

خروجی شماره ۲ ماتریس کواریانس متغیرهای اصلی را نشان می دهد.

Total Variance		Eigenvalues of the Covariance Matrix			
Eigenvalue	Difference	Proportion	Cumulative		
1	0.10308003	0.08736001	0.7846	0.7846	
2	0.01572002	0.00925837	0.1197	0.9043	
3	0.00646165	0.00300749	0.0492	0.9535	
4	0.00345416	0.00128086	0.0263	0.9797	
5	0.00217330	0.00168620	0.0165	0.9963	
6	0.00048710		0.0037	1.0000	

خروجی شماره ۳ مقادیر ویژه ماتریس کواریانس که در واقع واریانس مؤلفه ها می باشند را نشان می دهد. همچنین درصد تغییراتی را که هر کدام از مؤلفه ها توضیح می دهند را نشان می دهد به طوریکه دیده می شود، مؤلفه اول بیش از ۷۸ درصد تغییرات را در بر دارد. مؤلفه های اول و دوم را مؤلفه های اصلی در نظر می گیریم.

Eigenvectors

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6
x1	0.168249	0.431081	0.335519	0.373565	0.409895	-.604751
x2	0.186491	0.343954	-.062154	0.345525	0.403003	0.749169
x3	0.645270	-.475746	0.532670	0.167485	-.176557	0.119716
x4	0.676635	-.001853	-.646174	-.235042	0.167778	-.203044
x5	0.175051	0.471154	0.407669	-.750497	-.022260	0.132045
x6	0.179060	0.497535	-.126854	0.307441	-.780869	-.005263

خروجی شماره ۴ مؤلفه ها را نشان می دهد. مثلاً برای مؤلفه اول به صورت: $Y_1 = 0.168X_1 + 0.186X_2 + \dots + 0.179X_6$ و برای مؤلفه دوم $Y_2 = 0.431X_1 + 0.344X_2 + \dots + 0.497X_6$ می باشد. مؤلفه اول یک میانگین وزنی است و مؤلفه دوم یک مقایسه بین متغیر X_3 و X_5 است.

The CORR Procedure

9 Variables: x1 x2 x3 x4 x5 x6 Prin1 Prin2 Prin3

Pearson Correlation Coefficients, N = 20

	x1	x2	x3	x4	x5	x6	Prin1	Prin2	Prin3
x1	1.00000	0.85564	0.47669	0.52408	0.70997	0.71624	0.61996	0.62031	0.30954
x2	0.85564	1.00000	0.54645	0.72218	0.61513	0.76828	0.74217	0.53455	-0.06193
x3	0.47669	0.54645	1.00000	0.86348	0.42035	0.39503	0.94088	-0.27090	0.19446
x4	0.52408	0.72218	0.86348	1.00000	0.50296	0.58847	0.96994	-0.00104	-0.23191
x5	0.70997	0.61513	0.42035	0.50296	1.00000	0.62554	0.57131	0.60050	0.33312
x6	0.71624	0.76828	0.39503	0.58847	0.62554	1.00000	0.60761	0.65931	-0.10777
Prin1	0.61996	0.74217	0.94088	0.96994	0.57131	0.60761	1.00000	0.00000	0.00000
Prin2	0.62031	0.53455	-0.27090	-0.00104	0.60050	0.65931	0.00000	1.00000	0.00000
Prin3	0.30954	-0.06193	0.19446	-0.23191	0.33312	-0.10777	0.00000	0.00000	1.00000

(5)

خروجی شماره ۵ ضرایب همبستگی بین سه مؤلفه اول و متغیرهای اصلی را نشان می دهد که توسط دستور PROC CORR به دست آمده اند. همانطور که عنوان گردید بر اساس خروجی شماره ۴، مؤلفه اول بیشتر روی متغیرهای X3 و X4 تکیه دارد. اگر به ضریب همبستگی بین مؤلفه اول با این دو متغیر دقت نمایم، خواهیم دید که ضریب همبستگی Y1 و X3 برابر ۰/۹۴ و ضریب همبستگی Y1 و X4 برابر ۰/۹۶ است.

۹۴. تجزیه به مؤلفه های اصلی با ماتریس ضرایب همبستگی

برای مقایسه بهتر این دو روش از همان داده های برنامه قبل استفاده می کنیم.

```
TITLE "PRINCIPAL COMPONENT";
data;
input x1-x6;
cards;
;
proc princomp out=prin;
var x1-x6;
PROC CORR NOPROB NOSIMPLE;
VAR X1 X2 X3 X4 X5 X6 PRIN1 PRIN2 PRIN3;
proc plot;
plot prin*prin2="1" PRIN1*PRIN3="1" PRIN2*PRIN3="1"/BOX;
run;quit;
```

	Observations						20
	Variables						6
	Simple Statistics						
	x1	x2	x3	x4	x5	x6	
Mean	0.8612000000	0.8384000000	1.815800000	1.760450000	0.7143500000	0.6993000000	
Std	0.0871318420	0.0806750467	0.220188101	0.223972619	0.0983732286	0.0946150650	

(1)

خروجی شماره ۱ آماره های میانگین و انحراف از میانگین را ارائه می دهد.

Correlation Matrix

	x1	x2	x3	x4	x5	x6
x1	1.0000	0.8556	0.4767	0.5241	0.7100	0.7162
x2	0.8556	1.0000	0.5465	0.7222	0.6151	0.7683
x3	0.4767	0.5465	1.0000	0.8635	0.4203	0.3950

(2)

x4	0.5241	0.7222	0.8635	1.0000	0.5030	0.5885
x5	0.7100	0.6151	0.4203	0.5030	1.0000	0.6255
x6	0.7162	0.7683	0.3950	0.5885	0.6255	1.0000

در خروجی شماره ۲ ماتریس همبستگی بین متغیرها را نشان می دهد.

Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	4.12908325	3.20219079	0.6882	0.6882
2	0.92689246	0.49537373	0.1545	0.8427
3	0.43151873	0.12095502	0.0719	0.9146
4	0.31056370	0.15459340	0.0518	0.9663
5	0.15597030	0.10999874	0.0260	0.9923
6	0.04597156		0.0077	1.0000

(3)

خروجی شماره ۳ بار عوامل را نشان می دهد. مثلاً مؤلفه های اول و دوم به ترتیب ۶۸/۸۲ و ۱۵/۴۵ درصد تغییرات را در بر دارند. در صورتیکه درصد تغییرات مؤلفه اول و دوم در مثال قبل (به روش COV) به ترتیب ۷۸/۴۶ و ۱۱/۹۷ درصد بود.

	Eigenvectors					
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6
x1	0.427653	-.331033	-.022184	-.603636	0.358597	-.462676
x2	0.450762	-.127997	-.375344	-.306120	-.529596	0.515137
x3	0.360850	0.656720	0.180116	-.129262	0.481509	0.396869
x4	0.413455	0.500792	-.081505	0.226273	-.415666	-.589611
x5	0.382198	-.313867	0.812371	0.215064	-.193668	0.108193
x6	0.408282	-.305692	-.399491	0.654028	0.385745	0.060931

(4)

خروجی شماره ۴ مؤلفه ها را نشان می دهد. مثلاً برای مؤلفه اول $Y_1 = 0.427X_1 + 0.450X_2 + \dots + 0.408X_6$ و برای مؤلفه دوم $Y_2 = -0.331X_1 - 0.127X_2 + \dots - 0.305X_6$ می شود. در اینجا مؤلفه اول باز هم یک میانگین وزنی است (زیرا ضرایب خیلی به هم نزدیک هستند، لذا این مؤلفه را می توان میانگین در نظر گرفت).

The CORR Procedure									
9 Variables:	x1	x2	x3	x4	x5	x6	Prin1	Prin2	
	Prin3	Pearson Correlation Coefficients, N = 20							
	x1	x2	x3	x4	x5	x6	Prin1	Prin2	Prin3
x1	1.00000	0.85564	0.47669	0.52408	0.70997	0.71624	0.86900	-0.31870	-0.01457
x2	0.85564	1.00000	0.54645	0.72218	0.61513	0.76828	0.91595	-0.12323	-0.24656
x3	0.47669	0.54645	1.00000	0.86348	0.42035	0.39503	0.73325	0.63226	0.11832
x4	0.52408	0.72218	0.86348	1.00000	0.50296	0.58847	0.84015	0.48214	-0.05354
x5	0.70997	0.61513	0.42035	0.50296	1.00000	0.62554	0.77663	-0.30218	0.53365
x6	0.71624	0.76828	0.39503	0.58847	0.62554	1.00000	0.82963	-0.29431	-0.26243
Prin1	0.86900	0.91595	0.73325	0.84015	0.77663	0.82963	1.00000	0.00000	0.00000
Prin2	-0.31870	-0.12323	0.63226	0.48214	-0.30218	-0.29431	0.00000	1.00000	0.00000
Prin3	-0.01457	-0.24656	0.11832	-0.05354	0.53365	-0.26243	0.00000	0.00000	1.00000

(5)

در خروجی شماره ۵ ضرایب همبستگی بین مؤلفه ها و متغیرهای اصلی را نشان می دهد، برای اینکه معلوم شود که هر یک از مؤلفه ها بیشتر به کدام متغیر وابسته هستند.

تجزیه به عاملها (FACTOR ANALYSIS)

تجزیه به عاملها در sas با استفاده از رویه FACTOR صورت می گیرد. شکل کلی این دستور بصورت زیر است:

PROC FACTOR options;

VAR نام متغیرها

PRIORS کامیونیتی ها (در صورتیکه آنها را از قبل داشته باشیم)

RUN;QUIT;

به جای **OPTIONS** یک سری دستورات فرعی به کار می رود که بعضی از آنها به شرح زیر هستند:

۱-دستور **METHOD=** که روش تجزیه و تحلیل را مشخص می کند. در حالت عدم استفاده از این دستور، از روش

PRINCIPAL یعنی روش تجزیه به مؤلفه های اصلی استفاده خواهد شد. رو شهای موجود عبارتند از:

الف. روش **ALPHA**: موجب تجزیه به عاملهای آلفا می شود.

ب. روش **HARRIS**: باعث تجزیه مؤلفه های $S^{-1}RS^{-1}$ می شود.

ج. روش **IMAGE**: تجزیه مؤلفه های اصلی را روی نگاره ماتریس همبستگی انجام می دهد.

د. روش **ML** تجزیه مؤلفه های اصلی را به روش حداکثر درستنمایی انجام می دهد.

ه. روش **PRINCIPAL**: تجزیه به مؤلفه های اصلی را انجام می دهد، مشروط به اینکه از کلمه اختیاری **PRIORS**

استفاده نشده باشد یا به عبارت دیگر از **PRIORS=one** استفاده شود.

و. دستور **PRINIT**: تجزیه به عاملهای اصلی تکراری را اجرا می کند.

ز. روش **ULS**: تجزیه به عاملها را به روش کمترین توانهای دوم ناموزون انجام می دهد.

۲- دستور **COV** تجزیه و تحلیل را با ماتریس کواریانس انجام می دهد. در صورتیکه از این دستور استفاده نکنیم برنامه

SAS خود به خود ماتریس ضرایب همبستگی را به کار می برد.

۳-دستور **NFACTOR=** حداکثر تعداد فاکتورها(عاملها) را مشخص می کند. در صورت عدم استفاده از این کلمه

اختیاری، تعداد عاملها برابر تعداد متغیرها خواهد بود.

۴-دستور **HEYWOOD**: وقتی از روش **ML** استفاده می شود. چون این روش بر اساس روشهای عددی و تکرار، مسئله

را حل می کند، ممکن است مقادیر کامیونیتی بیشتر از ۱ دربیاید. بنابراین چنانچه در هنگام استفاده از روش **ML** به

پیغام «کامیونیتی ها بیشتر از ۱ می شوند» مشاهده گردید، از دستور **HEYWOOD** استفاده می شود تا کامیونیتی های بیشتر

از ۱ را به ۱ تبدیل کند و لذا باعث تداوم تکرار برنامه می شود.

۵-دستور **ROTATE=** روش چرخش (دوران) را مشخص می کند. که می تواند یکی از روشهای **VARIMAX**،

PROMAX، **QUARTMAY**، **PROCRUSTER**، **HK**، **EQUIMAX**، **ORTHOMAX** یا معمولاً از روش

استاندارد **VARIMAX** استفاده می شود.

- ۶- دستور **CORR** ماتریس ضرایب همبستگی را چاپ می کند.
- ۷- دستور **NPLOT=n** تعداد فاکتورهای را که می بایست رسم شوند را مشخص می کند.
- ۸- دستور **PLOT** فاکتورها را در برابر یکدیگر رسم می کند.
- ۹- دستور **REORDER** مقادیر **Li j** ها را بر اساس نزولی مرتب می کند.
- ۱۰- دستور **PREPLOT**: اگر از دوران (دستور شماره ۵) استفاده می کنیم، چنانچه بخواهیم نمودار عاملهای قبل از دوران را نیز رسم کند، از این دستور استفاده می کنیم.
- ۱۱- دستور **RES (RESIDUAL)** ماتریس باقیمانده ها و همبستگی جزئی را محاسبه می کند.
- ۱۲- دستور **SIMPLE**: خلاصه آماری شامل میانگین و انحراف معیار متغیرها را محاسبه می کند.
- ۱۳- دستور **NOIT**: مدل بدون عرض از مبدا (**intercept**) را در نظر می گیرد.
- ۱۴- دستور **SCREE**: نمودار مقادیر ویژه را رسم می کند. برای حدس **n** (تعداد فاکتورها) مفید است.
- ۱۵- دستور **SCORE**: امتیاز عاملها را چاپ می کند.
- ۱۶- دستور **PRIORS=**: روشی را برای برآورد کامیونیتی ها مشخص می کند. روشهای موجود عبارتند از:
one: تمامی کامیونیتی ها را برابر با ۱ قرار می دهد.
Smc: برای هر متغیر مقدار کامیونیتی آن را برابر با جذر همبستگی چندگانه آن با تمامی دیگر متغیرها قرار می دهد.
چنانچه کامیونیتی ها را از قبل داشته باشیم، به عنوان مثال کامیونیتی های **X1**، **X2** و **X3** را به ترتیب ۰/۷، ۰/۸ و ۰/۹ داشته باشیم برنامه به صورت زیر خواهد شد:

```
PROC FACTOR;
  VAR    X1 X2 X3;
  PRIORS 0.7 0.8 0.9;
```

مثال: درصد نیروی کارگری در ۹ نوع صنعت مختلف از ۲۶ کشور اروپایی جمع آوری شده است. این ۹ صنعت عبارتند از: کشاورزی (**agr**)، معدن (**min**)، صنایع (**man**)، تامین نیرو (**ps**)، ساختمان (**con**)، صنایع خدماتی (**ser**)، امور مالی (**fin**)، خدمات عمومی و خصوصی (**sps**) و ترابری (**tc**). در این مورد تجزیه به عاملها ممکن است برای تفکیک کشورها به گروههایی که دارای توابع شغلی مشابه هستند، مفید باشد.

۹۵. تجزیه به عاملها به روش **PRINCIPAL COMPONENT**

```
data;
input entry$ agr min man ps con ser fin sps tc;
cards;
;
proc factor COV METHOD=PRIN NFACT=2 SCREE RESIDUAL REORDER PLOT;
var agr min man ps con ser fin sps tc;
run;
```

Initial Factor Method: Principal Components
 Prior Communality Estimates: ONE
 Eigenvalues of the Covariance Matrix: Total = 666.708043 Average = 74.0786715

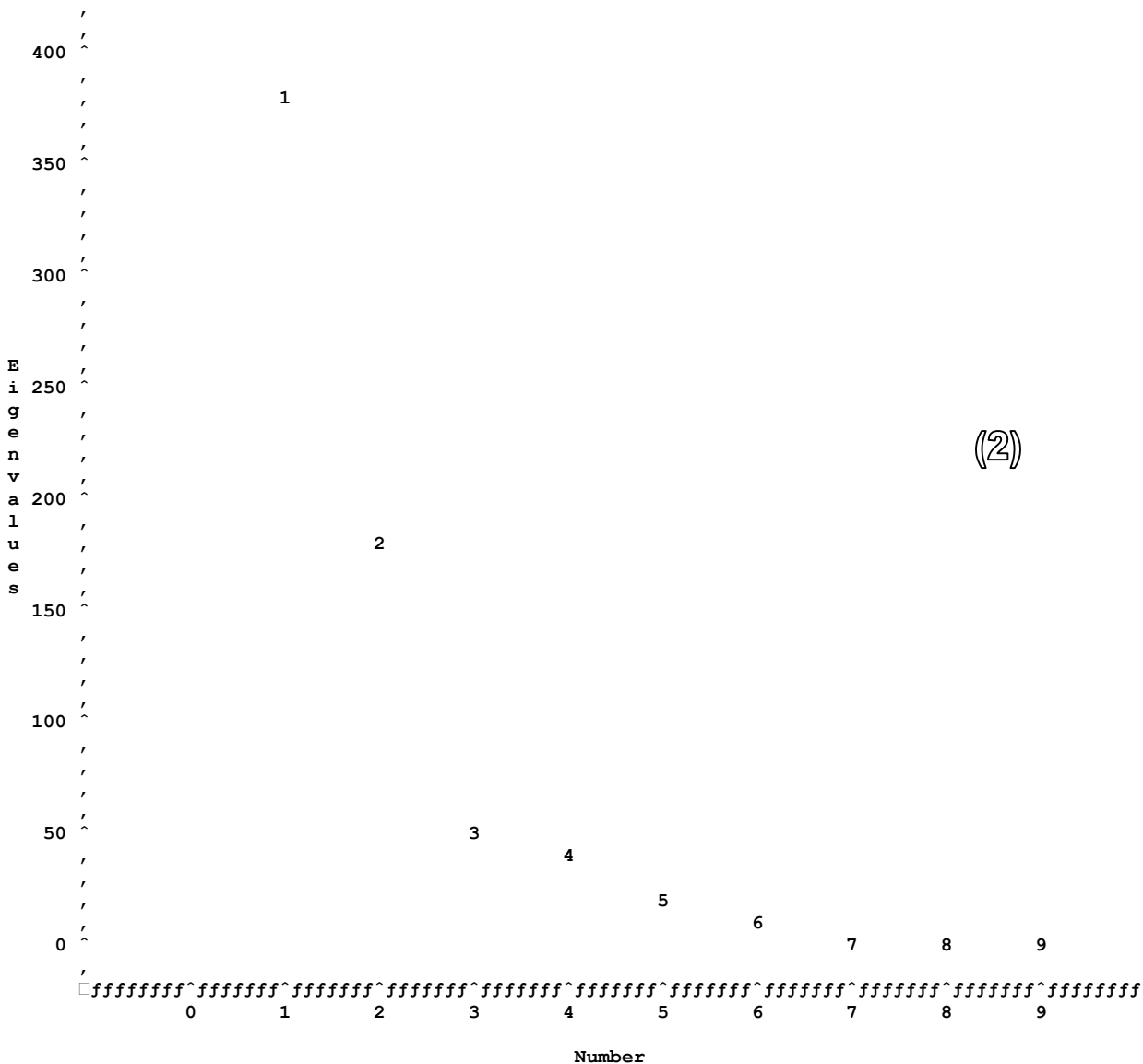
	Eigenvalue	Difference	Proportion	Cumulative
1	378.538196	201.753459	0.5678	0.5678
2	176.784738	127.293192	0.2652	0.8329
3	49.491546	12.971544	0.0742	0.9072
4	36.520001	18.976196	0.0548	0.9619
5	17.543805	10.897692	0.0263	0.9883
6	6.646113	5.898831	0.0100	0.9982
7	0.747282	0.470763	0.0011	0.9993
8	0.276519	0.116675	0.0004	0.9998
9	0.159844		0.0002	1.0000

2 factors will be retained by the NFACTOR criterion.

(1)

در خروجی ۱ کامیونیتی های اولیه ۱ در نظر گرفته شده است. در این قسمت، مقادیر ویژه ماتریس کواریانس را مشاهده می کنید. فاکتور اول بیش از ۵۶ درصد تغییرات را توضیح می دهد. فاکتور اول و دوم روی هم ۸۳/۲۹ درصد از تغییرات را در بر دارند. مقادیر ویژه عوامل اول و دوم تقریباً بزرگ هستند و بقیه تقریباً نزدیک به صفر هستند. (n را ۲ در نظر گرفته ایم).

Initial Factor Method: Principal Components
 Scree Plot of Eigenvalues



(2)

با استفاده از دستور **SCREE** نمودار مقادیر ویژه را در خروجی شماره ۲ خواهیم داشت.

The FACTOR Procedure
Initial Factor Method: Principal Components
Factor Pattern

	Factor1	Factor2
fin	0.98771	0.14803
ps	0.21681	-0.18121
min	-0.30696	0.07931
ser	-0.32321	0.94259
tc	0.02944	0.38302
con	0.11643	0.16408
agr	-0.01182	-0.16928
sps	-0.21238	-0.30741
man	-0.00069	-0.32848

(3)

خروجی شماره ۳ بار عوامل (L_{ij}) را نشان می دهد. در فاکتور اول متغیرهای **fin** و **ps** نسبت به بقیه بار بیشتری دارند. در فاکتور دوم متغیرهای **ser**، **tc** و **con** بار بیشتری دارند. باید دید که چه عاملی متغیرهای **fin** و **ps** را به هم نزدیک کرده و آنها را در یک گروه قرار داده آن عامل همان عامل پنهانی است. همینطور برای فاکتور دوم.

Variance Explained by Each Factor

Factor	Weighted	Unweighted
Factor1	378.538196	1.28094095
Factor2	176.784738	1.35419305

(4)

خروجی ۴ واریانس توضیح داده شده توسط هر عامل را نشان می دهد.

Final Communalities Estimates and Variable Weights
Total Communality: Weighted = 555.32293 Unweighted = 2.635134

Variable	Communality	Weight
agr	0.02879653	9.252632
min	0.10051594	40.871053
man	0.10789806	0.647263
ps	0.07984691	0.381553
con	0.04047783	15.752517
ser	0.99293241	182.989474
fin	0.99748618	361.726316
sps	0.13960564	54.246947
tc	0.14757450	0.840289

(5)

خروجی شماره ۵ برآورد کامیونیتی ها را نشان می دهد.

The FACTOR Procedure
Initial Factor Method: Principal Components
Residual Correlations With Uniqueness on the Diagonal

	agr	min	man	ps	con
agr	0.97120	0.04146	-0.35843	-0.49002	0.36708
min	0.04146	0.89948	0.29937	0.06247	-0.04554
man	-0.35843	0.29937	0.89210	0.50607	-0.44040
ps	-0.49002	0.06247	0.50607	0.92015	-0.37381
con	0.36708	-0.04554	-0.44040	-0.37381	0.95952
ser	0.00890	-0.02986	-0.00698	0.00808	-0.00416
fin	0.00417	0.02472	0.01036	0.00818	-0.00531
sps	0.08391	-0.06645	-0.03223	0.09548	0.05304
tc	-0.09771	0.16371	-0.17868	-0.22638	0.02961
agr	0.00890	0.00417	0.08391	-0.09771	
min	-0.02986	0.02472	-0.06645	0.16371	
man	-0.00698	0.01036	-0.03223	-0.17868	
ps	0.00808	0.00818	0.09548	-0.22638	
con	-0.00416	-0.00531	0.05304	0.02961	
ser	0.00707	0.00251	0.07334	-0.03072	
sps	0.07334	0.03730	0.86039	-0.28000	
tc	-0.03072	-0.00855	-0.28000	0.85243	

(6)

Root Mean Square Off-Diagonal Residuals: Overall = 0.19884350

agr	min	man	ps	con
0.25537766	0.12749271	0.29607175	0.29587995	0.24347546

Root Mean Square Off-Diagonal Residuals: Overall = 0.19884350

ser	fin	sps	tc
0.03047799	0.01696324	0.11704372	0.15804074

خروجی شماره ۶ ماتریس باقیمانده ها را نشان می دهد. عناصر خارج از قطر تقریباً اعداد نزدیک به صفر هستند.

Initial Factor Method: Principal Components
Partial Correlations Controlling Factors

	agr	min	man	ps	con
agr	1.00000	0.04436	-0.38507	-0.51836	0.38025
min	0.04436	1.00000	0.33419	0.06866	-0.04902
man	-0.38507	0.33419	1.00000	0.55856	-0.47600
ps	-0.51836	0.06866	0.55856	1.00000	-0.39783
con	0.38025	-0.04902	-0.47600	-0.39783	1.00000
ser	0.10745	-0.37445	-0.08790	0.10024	-0.05054
fin	0.08436	0.51988	0.21887	0.17013	-0.10805
sps	0.09180	-0.07554	-0.03679	0.10730	0.05838
tc	-0.10739	0.18696	-0.20490	-0.25561	0.03274

Partial Correlations Controlling Factors

	ser	fin	sps	tc
agr	0.10745	0.08436	0.09180	-0.10739
min	-0.37445	0.51988	-0.07554	0.18696
man	-0.08790	0.21887	-0.03679	-0.20490
ps	0.10024	0.17013	0.10730	-0.25561
con	-0.05054	-0.10805	0.05838	0.03274
ser	1.00000	0.59570	0.94052	-0.39584
fin	0.59570	1.00000	0.80209	-0.18481
sps	0.94052	0.80209	1.00000	-0.32695
tc	-0.39584	-0.18481	-0.32695	1.00000

(7)

Root Mean Square Off-Diagonal Partial: Overall = 0.34492256

agr	min	man	ps	con
0.27435334	0.26737680	0.33494455	0.32777909	0.26233433

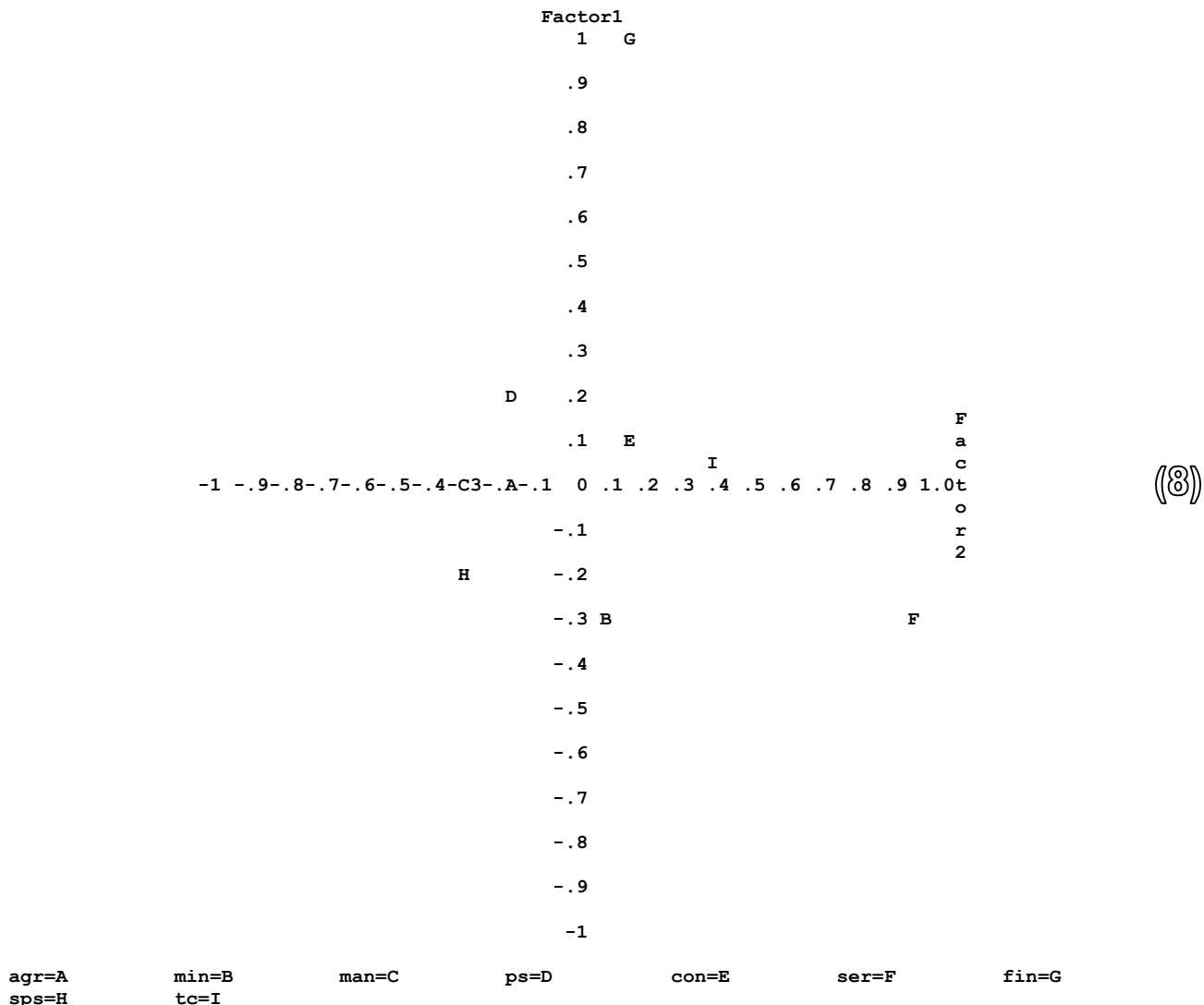
Root Mean Square Off-Diagonal Partial: Overall = 0.34492256

ser	fin	sps	tc
0.44274733	0.41807175	0.45623670	0.23785844

خروجی شماره ۷ همبستگی های جزئی (نسبی) بین متغیرها را نشان می دهد.

The FACTOR Procedure
Initial Factor Method: Principal Components

Plot of Factor Pattern for Factor1 and Factor2



خروجی شماره ۸ نمودار عامل اول را در مقابل عامل دوم رسم می کند.

۹۶. تجزیه به عاملها به روش ML

در این مثال با استفاده از ماتریس ضرایب همبستگی و روش حداکثر درستنمایی (ML) و با **NFACT=2** به همراه دوران به روش واریمکس مسئله تجزیه و تحلیل می شود.

```
data;
input X1-X6;
cards;
1.103 1.052 2.139 2.233 0.873 0.872
0.842 0.859 1.873 1.741 0.590 0.744
0.925 0.873 1.887 1.809 0.767 0.713
0.857 0.744 1.739 1.547 0.706 0.674
0.795 0.809 1.739 1.715 0.549 0.654
0.787 0.779 1.509 1.474 0.782 0.571
0.933 0.880 1.695 1.656 0.737 0.803
0.799 0.851 1.740 1.777 0.618 0.682
0.945 0.876 1.811 1.759 0.853 0.777
0.921 0.906 1.654 2.009 0.823 0.765
0.792 0.825 1.624 1.657 0.686 0.668
0.815 0.751 2.204 1.846 0.678 0.546
0.755 0.724 1.508 1.458 0.662 0.595
```



```
0.880 0.866 1.786 1.811 0.810 0.819
0.900 0.838 1.902 1.606 0.723 0.677
0.764 0.757 1.743 1.794 0.586 0.541
0.733 0.748 1.863 1.869 0.672 0.752
0.932 0.898 2.028 2.032 0.836 0.805
0.856 0.786 1.390 1.324 0.578 0.610
0.890 0.950 2.187 2.087 0.758 0.718
```

```
;
proc factor METHOD=ML NFACT=2 ROTATE=VARIMAX RESIDUAL REORDER HEYWOOD SCORE PREPLOT PLOT;
var X1-X6;
run;
```

دستور **HEYWOOD** به این دلیل به کار رفته است که در هنگام اجرای برنامه بعضی از کامیونیتی ها بیش از ۱ می شدند. لذا این دستور آنها را برابر ۱ قرار می دهد.

The FACTOR Procedure
Initial Factor Method: Maximum Likelihood

Prior Communalities Estimates: SMC

X1	X2	X3	X4	X5	X6
0.84277842	0.87010240	0.68947713	0.82917392	0.57662515	0.64396532

Preliminary Eigenvalues: Total = 22.3037918 Average = 3.71729864

(1)

	Eigenvalue	Difference	Proportion	Cumulative
1	19.4483398	16.3619297	0.8720	0.8720
2	3.0864101	2.6722740	0.1384	1.0104
3	0.4141360	0.1574715	0.0186	1.0289
4	0.2566645	0.5068850	0.0115	1.0404
5	-0.2502205	0.4013177	-0.0112	1.0292
6	-0.6515381		-0.0292	1.0000

2 factors will be retained by the NFACTOR criterion.

Iteration	Criterion	Ridge	Change	Communalities					
1	0.4088298	0.0000	0.1708	0.85929	0.86449	0.52198	1.00000	0.55788	0.64764
2	0.3574314	0.0000	0.0944	0.91988	0.86783	0.61637	1.00000	0.52227	0.64271
3	0.3562141	0.0000	0.0143	0.92080	0.87153	0.61418	1.00000	0.53420	0.62837
4	0.3560868	0.0000	0.0045	0.92531	0.86920	0.61416	1.00000	0.53250	0.62777
5	0.3560710	0.0000	0.0016	0.92563	0.86921	0.61399	1.00000	0.53408	0.62614
6	0.3560689	0.0000	0.0005	0.92617	0.86893	0.61398	1.00000	0.53403	0.62595

Convergence criterion satisfied.

همانطور که در خروجی شماره ۱ ملاحظه می گردد، برای برآورد کامیونیتی ها از روش تکرار استفاده می شود. در این برنامه خود SAS مقدار اولیه کامیونیتی ها را برابر SMC (مربع همبستگی چندگانه هر متغیر با دیگر متغیرها) گرفته است. برای مقادیر ویژه، مقدار اولیه را در نظر گرفته است. روش تکرار در ۶ مرحله به جواب رسیده و کامیونیتی ها برآورد شده اند.

Significance Tests Based on 20 Observations

Test	DF	Chi-Square	Pr > ChiSq
H0: No common factors	15	81.8291	<.0001
HA: At least one common factor			
H0: 2 Factors are sufficient	4	5.2817	0.2596
HA: More factors are needed			

(2)

Chi-Square without Bartlett's Correction	6.7653099
Akaike's Information Criterion	-1.2346901
Schwarz's Bayesian Criterion	-5.2176191
Tucker and Lewis's Reliability Coefficient	0.9280802

خروجی شماره ۲: در روش **ML**، نرم افزار **SAS** برای پی بردن به وجود فاکتور مشترک اقدام به انجام آزمون می کند. اگر این فاکتور وجود داشت. بررسی می کند که آیا یک فاکتور کافیت؟ یا اینکه دو یا چند فاکتور.

خروجی شماره ۲ این آزمون را نشان می دهد. در اینجا فرض H_0 یعنی فاکتور مشترک وجود ندارد، رد می شود. آزمون می شود که آیا دو فاکتور کافیت؟ (فرض H_0) و همانطور که می بینیم در اینجا فرض H_0 رد نمی شود. یعنی دو فاکتور کافیت. پس تجزیه و تحلیل با دو فاکتور انجام می شود.

The FACTOR Procedure
Initial Factor Method: Maximum Likelihood

Squared Canonical Correlations

	Factor1	Factor2
	1.0000000	0.9280208

Eigenvalues of the Weighted Reduced Correlation Matrix: Total = 12.892898 Average = 2.5785796

	Eigenvalue	Difference	Proportion	Cumulative
1	Infty	Infty		
2	12.8928996	12.3568817	1.0000	1.0000
3	0.5360179	0.3193582	0.0416	1.0416
4	0.2166597	0.4740198	0.0168	1.0584
5	-0.2573601	0.2379590	-0.0200	1.0384
6	-0.4953191		-0.0384	1.0000

(3)

خروجی شماره ۳ مقادیر ویژه را نشان می دهد.

Factor Pattern

	Factor1	Factor2
X4	1.00000	0.00000
X3	0.78299	-0.02968
X2	0.71913	0.59312
X6	0.58779	0.52943
X1	0.52212	0.80843
X5	0.50232	0.53092

(4)

Variance Explained by Each Factor

Factor	Weighted	Unweighted
Factor1	10.6914351	3.00065307
Factor2	12.8928996	1.56841134

خروجی شماره ۴ بار عاملها را نشان می دهد.

Final Commnality Estimates and Variable Weights
Total Commnality: Weighted = 23.584335 Unweighted = 4.569064

Variable	Commnality	Weight
X1	0.92617504	13.5445898
X2	0.86893922	7.6297054
X3	0.61395767	2.5905215
X4	1.00000000	Infty
X5	0.53420312	2.1460560
X6	0.62578936	2.6734604

(5)

خروجی شماره ۵ کامیونیتی ها را نشان می دهد.

The FACTOR Procedure
Initial Factor Method: Maximum Likelihood
Residual Correlations With Uniqueness on the Diagonal

	X1	X2	X3	X4	X5	X6
X1	0.07382	0.00190	0.04129	0.00000	0.01848	-0.01866
X2	0.00190	0.13106	-0.05881	0.00000	-0.06072	0.03102
X3	0.04129	-0.05881	0.38604	0.00000	-0.03877	-0.10022
X4	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
X5	0.01848	-0.06072	-0.03877	0.00000	0.46580	0.04920
X6	-0.01866	0.03102	-0.10022	0.00000	0.04920	0.37421

(6)

Root Mean Square Off-Diagonal Residuals: Overall = 0.04039377

X1	X2	X3	X4	X5	X6
0.02190309	0.04027836	0.05781232	0.00000000	0.03988107	0.05248889

خروجی شماره ۶ ماتریس باقیمانده ها را نشان می دهد.

Partial Correlations Controlling Factors

	X1	X2	X3	X4	X5	X6
X1	1.00000	0.01930	0.24461	0.00000	0.09966	-0.11229
X2	0.01930	1.00000	-0.26147	0.00000	-0.24576	0.14006
X3	0.24461	-0.26147	1.00000	0.00000	-0.09142	-0.26368
X4	0.00000	0.00000	0.00000	1.00000	0.00000	0.00000
X5	0.09966	-0.24576	-0.09142	0.00000	1.00000	0.11785
X6	-0.11229	0.14006	-0.26368	0.00000	0.11785	1.00000

(7)

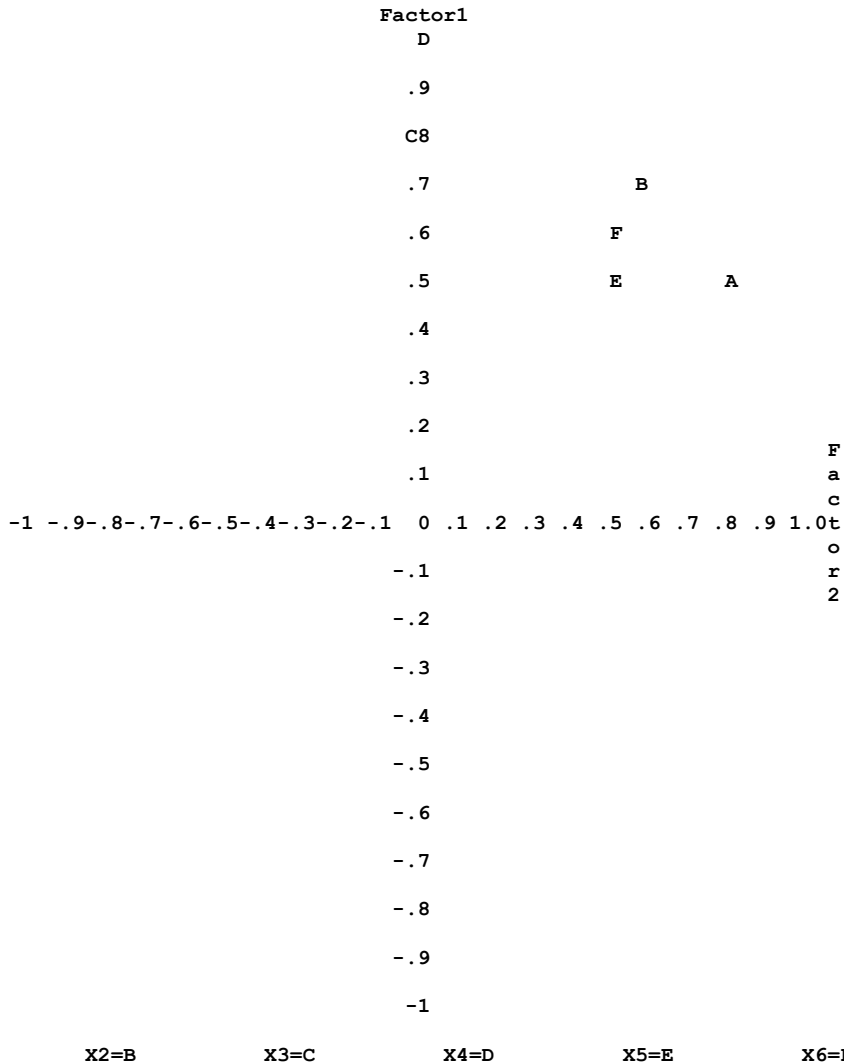
Root Mean Square Off-Diagonal Partials: Overall = 0.14672011

X1	X2	X3	X4	X5	X6
0.12864501	0.17248267	0.20302032	0.00000000	0.13607102	0.15208013

خروجی شماره ۷ همبستگی نسبی (جزئی) بین متغیرها را نشان می دهد.

The FACTOR Procedure
Initial Factor Method: Maximum Likelihood

Plot of Factor Pattern for Factor1 and Factor2



(8)

خروجی شماره ۸ نمودار بار عاملها را قبل از دوران نشان می‌دهد.

```
The FACTOR Procedure
Rotation Method: Varimax
Orthogonal Transformation Matrix
      1          2
1      0.35810      0.93368
2      0.93368     -0.35810
```

```
Rotated Factor Pattern
      Factor1      Factor2
X1      0.94179      0.19800
X2      0.81131      0.45905
X6      0.70480      0.35923
X5      0.67559      0.27888
X4      0.35810      0.93368
X3      0.25267      0.74170
```

(9)

```
Variance Explained by Each Factor
Factor      Weighted      Unweighted
Factor1     19.5085923     2.69043727
Factor2      4.0757424     1.87862715
```

خروجی شماره ۹ ماتریس دوران و بار عاملها را پس از دوران به روش واریمکس نشان می‌دهد. در فاکتور اول بار عاملها برای متغیر $X1$ ، $X2$ ، $X6$ و $X5$ زیاد است. در فاکتور دوم بار عاملها برای متغیرهای $X3$ و $X4$ زیاد است. با مثال قبل که به روش مؤلفه‌های اصلی تجزیه و تحلیل شد مقایسه شود.

```
Final Communality Estimates and Variable Weights
Total Communality: Weighted = 23.584335      Unweighted = 4.569064
```

```
Variable      Communality      Weight
X1      0.92617504     13.5445898
X2      0.86893922     7.6297054
X3      0.61395767     2.5905215
X4      1.00000000           Infty
X5      0.53420312     2.1460560
X6      0.62578936     2.6734604
```

(10)

```
The FACTOR Procedure
Rotation Method: Varimax
Scoring Coefficients Estimated by Regression
Squared Multiple Correlations of the Variables with Each Factor
```

```
Factor1      Factor2
0.93725085     0.99076994
```

خروجی شماره ۱۰ تخمین نهایی کامیونیتی‌ها و همچنین وزن متغیرها را پس از دوران نشان می‌دهد.

```
Standardized Scoring Coefficients
```

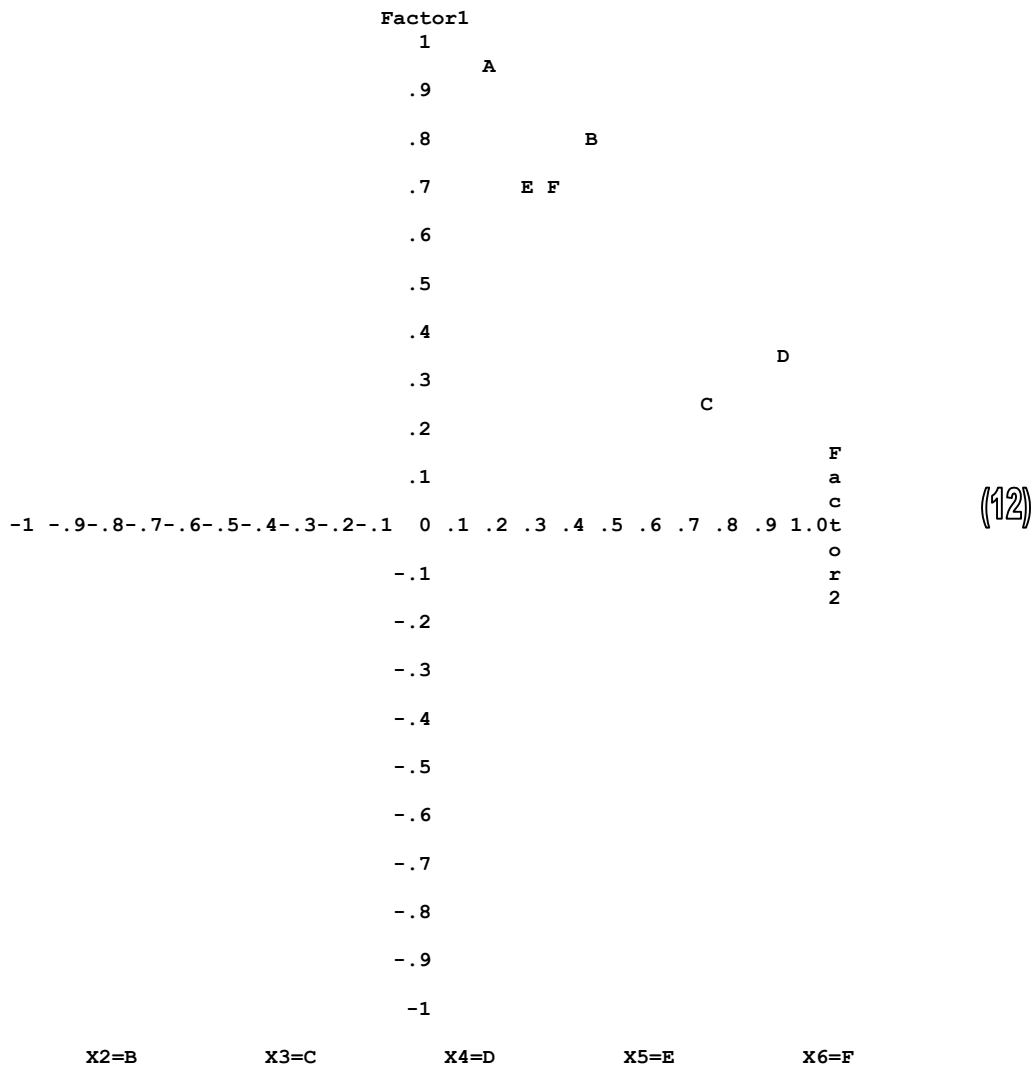
```
Factor1      Factor2
X1      0.73589754     -0.2822382
X2      0.30412995     -0.1166427
X6      0.09512321     -0.0364825
X5      0.07657403     -0.0293684
X4     -0.3351723      1.19957339
X3     -0.0051671      0.00198175
```

(11)

خروجی شماره ۱۱ استاندارد شده عوامل (فاکتورها) را برای متغیرها بعد از گردش نشان می‌دهد.

```
The FACTOR Procedure
Rotation Method: Varimax
```

```
Plot of Factor Pattern for Factor1 and Factor2
```



خروجی شماره ۱۲ نمودار عاملها را پس از دوران نشان می دهد.

تجزیه کلاستر

در تجزیه کلاستر با نمونه‌گیری از n فرد و اندازه‌گیری k متغیر بر روی هر فرد، می‌توان افراد را در کلاس‌ها یا خوشه‌هایی گروه‌بندی کرد که افراد مشابه در داخل یک کلاس قرار گیرند. این روش باید کاملاً عددی بوده و نباید تعداد کلاسهای آن از قبل مشخص باشد. برای تجزیه کلاستر در SAS از رویه **cluster** استفاده می‌شود.

۹۷. تجزیه کلاستر به روش **average**

```
data;
input dog$ x1-x6;
cards;
;
proc cluster method=average;
id dog;
proc tree;
id dog;
run;
```

method= برچسب یا شماره مشاهدات متغیری را که سایر متغیرها به آن مربوط هستند، مشخص می کند. **ID** روش مورد استفاده برای تجزیه کلاستر را مشخص می کند. ۱۱ روش برای تجزیه کلاستر وجود دارد که عبارتند از: **WARD, TWOSTAGE, SINGLE, MEDIAN, MCQUITTY, FLEXIBLE, EML, DENSITY, COMPLETE, CENTROID, AVERAGE** که روش هر یک از آنها برای انجام تجزیه کلاستر متفاوت می باشد. رویه **TREE** از **SAS** می خواهد که دندروگرام حاصل از کلاسترها را رسم کند. برای مثال خویشاوندی بین ۷ نژاد سگ با اندازه گیری ۶ کاراکتر در هر یک از آنها با روشهای **WARD** و **AVERAGE** مورد بررسی قرار گرفته است.

۹۸. تجزیه کلاستر به روش ward

```
data;
input dog$ x1-x6;
cards;
;
proc cluster method=ward;
id dog;
proc tree;
id dog;
run;
```

۹۹. تجزیه همبستگی کانونیک (CANONICAL CORRELATION ANALYSIS)

در رگرسیون چندگانه یک متغیر وابسته (**Y**) با چندین متغیر مستقل (**X**) در ارتباط است اما در همبستگی کانونیک که در واقع حالت عمومیت یافته رگرسیون چندگانه می باشد چندین متغیر **Y** به طور همزمان با چندین متغیر **X** ارتباط داده می شوند. در برخی از مجموعه های مربوط به داده های چند متغیره، متغیرها به طور طبیعی به دو گروه تقسیم می شوند. در چنین مجموعه هایی، همبستگی کانونیک می تواند برای بررسی ارتباط بین این دو گروه مورد استفاده قرار گیرد. در مثال زیر ۱۶ کلنی از پروانه های *Euphydrus editha* در ایالت های کالیفرنیا و اوروگون امریکا مورد بررسی قرار گرفته اند. داده های هر کلنی از لحاظ چهار متغیر محیطی (**Y1-Y4**) و پنج متغیر محیطی ژنتیکی مربوط به فراوانی ژنی (**X1-X5**) مورد بررسی قرار گرفته است. سوال اینجاست که چه رابطه ای بین فراوانیهای ژنی و متغیرهای محیطی وجود دارد؟

```
data;
input eue$ y1-y4 x1-x5;
cards;
ss      500      43      98      17      0      3      22      57      17
.
g1     10500     50      81      12      0      3      1      92      4
;
proc cancorr all;
var x1-x5;
with y1-y4;
run;
```

در SAS رویه **CANCORR** تجزیه همبستگی کانونیک را انجام می دهد. دستور **all** باعث چاپ تمامی خروجیهای ممکن رویه **CANCORR** می شود. دستور **short** باعث می گردد که فقط ضرایب کانونیک و آماره های چند متغیره چاپ شوند. دستور **SIMPLE** میانگینها و انحرافات معیار را چاپ می کند.

```
VAR Variables          5
WITH Variables        4
Observations          16
Means and Standard Deviations
```

Variable	Mean	Standard Deviation
x1	1.750000	4.139243
x2	7.187500	7.494164
x3	18.562500	10.776324
x4	51.187500	20.063960
x5	17.187500	10.777871
y1	2101.875000	2879.359231
y2	28.062500	13.992706
y3	97.250000	6.382267
y4	22.375000	7.135592

Correlations Among the Original Variables

Correlations Among the VAR Variables

	x1	x2	x3	x4	x5
x1	1.0000	0.8548	0.6176	-0.5316	-0.5055
x2	0.8548	1.0000	0.6153	-0.5478	-0.5972
x3	0.6176	0.6153	1.0000	-0.8235	-0.1267
x4	-0.5316	-0.5478	-0.8235	1.0000	-0.2638
x5	-0.5055	-0.5972	-0.1267	-0.2638	1.0000

Correlations Among the WITH Variables

	y1	y2	y3	y4
y1	1.0000	0.5675	-0.8277	-0.7837
y2	0.5675	1.0000	-0.4787	-0.7307
y3	-0.8277	-0.4787	1.0000	0.5336
y4	-0.7837	-0.7307	0.5336	1.0000

Correlations Between the VAR Variables and the WITH Variables

	y1	y2	y3	y4
x1	-0.2025	-0.5303	0.2953	0.2449
x2	-0.1900	-0.4095	0.1732	0.2529
x3	-0.5729	-0.5498	0.5358	0.5459
x4	0.7270	0.6990	-0.7173	-0.7097
x5	-0.4577	-0.1380	0.4383	0.3588

Canonical Correlation Analysis

	Canonical Correlation	Adjusted Canonical Correlation	Approximate Standard Error	Squared Canonical Correlation
1	0.885734	0.827124	0.055635	0.784525
2	0.735391	0.658764	0.118565	0.540801
3	0.231972	.	0.244305	0.053811
4	0.204605	.	0.247390	0.041863

Test of H0: The canonical correlations in the current row and all that follow are zero

Eigenvalues of Inv(E)*H
= CanRsqr/(1-CanRsqr)

	Eigenvalue	Difference	Proportion	Cumulative	Likelihood Ratio	Approximate F Value	Num DF	Den DF	Pr > F
1	3.6409	2.4632	0.7401	0.7401	0.08970230	1.29	20	24	0.2720
2	1.1777	1.1208	0.2394	0.9796	0.41630020	0.70	12	21	0.7330
3	0.0569	0.0132	0.0116	0.9911	0.90657825	0.15	6	18	0.9865
4	0.0437		0.0089	1.0000	0.95813676	0.22	2	10	0.8075

Multivariate Statistics and F Approximations

S=4 M=0 N=2.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.08970230	1.29	20	24.166	0.2720
Pillai's Trace	1.42100000	1.10	20	40	0.3847
Hotelling-Lawley Trace	4.91917578	1.56	20	9.5	0.2444
Roy's Greatest Root	3.64090877	7.28	5	10	0.0041

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

Canonical Correlation Analysis

Raw Canonical Coefficients for the VAR Variables

	V1	V2	V3	V4
x1	-0.141734472	-0.228273375	0.0755890366	-0.39401422
x2	0.1128614375	0.4363151808	-0.296674388	-0.199889756
x3	0.0322875384	0.2270634917	-0.082538242	-0.196407311
x4	0.0715074767	0.1991887857	-0.089935966	-0.220501963
x5	0.0221046005	0.3067807351	-0.094863131	-0.287354329

Raw Canonical Coefficients for the WITH Variables

	W1	W2	W3	W4
y1	-0.000114212	-0.000512714	0.0005366214	0.000520847
y2	0.0328544679	0.0646159111	0.0709787628	0.0383058673
y3	-0.120995419	-0.105652163	0.2595218703	0.0273672276
y4	-0.037713886	0.0344570729	0.1128749815	0.2716649984

Canonical Correlation Analysis

Standardized Canonical Coefficients for the VAR Variables

	V1	V2	V3	V4
x1	-0.5867	-0.9449	0.3129	-1.6309
x2	0.8458	3.2698	-2.2233	-1.4980
x3	0.3479	2.4469	-0.8895	-2.1165
x4	1.4347	3.9965	-1.8045	-4.4241
x5	0.2382	3.3064	-1.0224	-3.0971

Standardized Canonical Coefficients for the WITH Variables

	W1	W2	W3	W4
y1	-0.3289	-1.4763	1.5451	1.4997
y2	0.4597	0.9042	0.9932	0.5360
y3	-0.7722	-0.6743	1.6563	0.1747
y4	-0.2691	0.2459	0.8054	1.9385

Canonical Structure

Correlations Between the VAR Variables and Their Canonical Variables

	V1	V2	V3	V4
x1	-0.5319	-0.4344	-0.6609	-0.3013
x2	-0.3699	-0.1964	-0.9040	0.0789
x3	-0.7056	0.1653	-0.4487	-0.0101
x4	0.9339	-0.1795	0.2493	-0.1767
x5	-0.3928	0.4672	0.7358	0.0570

Correlations Between the WITH Variables and Their Canonical Variables

	W1	W2	W3	W4
y1	0.7821	-0.5978	0.1065	0.1401
y2	0.8394	0.2095	0.4886	-0.1130
y3	-0.8637	0.2461	0.3317	-0.2889
y4	-0.7594	0.3823	-0.2474	0.4647

Correlations Between the VAR Variables and the Canonical Variables of the WITH Variables

	W1	W2	W3	W4
x1	-0.4711	-0.3194	-0.1533	-0.0617
x2	-0.3276	-0.1444	-0.2097	0.0161
x3	-0.6250	0.1215	-0.1041	-0.0021
x4	0.8272	-0.1320	0.0578	-0.0362
x5	-0.3479	0.3436	0.1707	0.0117

Correlations Between the WITH Variables and the Canonical Variables of the VAR Variables

	V1	V2	V3	V4
y1	0.6927	-0.4396	0.0247	0.0287
y2	0.7435	0.1541	0.1133	-0.0231
y3	-0.7650	0.1809	0.0769	-0.0591
y4	-0.6726	0.2812	-0.0574	0.0951

Canonical Redundancy Analysis

Raw Variance of the VAR Variables Explained by
Their Own Canonical Variables The Opposite Canonical Variables

Canonical Variable Number	Proportion	Cumulative Proportion	Canonical R-Square	Proportion	Cumulative Proportion
1	0.6205	0.6205	0.7845	0.4868	0.4868
2	0.0662	0.6867	0.5408	0.0358	0.5226
3	0.2325	0.9192	0.0538	0.0125	0.5351
4	0.0210	0.9402	0.0419	0.0009	0.5360

Raw Variance of the WITH Variables Explained by
Their Own Canonical Variables The Opposite Canonical Variables

Canonical Variable Number	Proportion	Cumulative Proportion	Canonical R-Square	Proportion	Cumulative Proportion
1	0.6117	0.6117	0.7845	0.4799	0.4799
2	0.3573	0.9690	0.5408	0.1932	0.6731
3	0.0114	0.9804	0.0538	0.0006	0.6737
4	0.0196	1.0000	0.0419	0.0008	0.6746

Canonical Redundancy Analysis

Standardized Variance of the VAR Variables Explained by
Their Own Canonical Variables The Opposite Canonical Variables

Canonical Variable Number	Proportion	Cumulative Proportion	Canonical R-Square	Proportion	Cumulative Proportion
1	0.3888	0.3888	0.7845	0.3050	0.3050
2	0.1010	0.4898	0.5408	0.0546	0.3597
3	0.4118	0.9016	0.0538	0.0222	0.3818
4	0.0263	0.9280	0.0419	0.0011	0.3829

Standardized Variance of the WITH Variables Explained by
Their Own Canonical Variables The Opposite Canonical Variables

Canonical Variable Number	Proportion	Cumulative Proportion	Canonical R-Square	Proportion	Cumulative Proportion
1	0.6597	0.6597	0.7845	0.5176	0.5176
2	0.1520	0.8117	0.5408	0.0822	0.5998
3	0.1053	0.9170	0.0538	0.0057	0.6054
4	0.0830	1.0000	0.0419	0.0035	0.6089

The CANCELL Procedure

Canonical Redundancy Analysis

Squared Multiple Correlations Between the VAR Variables and the First M Canonical Variables of the WITH Variables

M	1	2	3	4
x1	0.2220	0.3240	0.3475	0.3513
x2	0.1073	0.1282	0.1722	0.1724
x3	0.3906	0.4054	0.4162	0.4163
x4	0.6842	0.7017	0.7050	0.7063
x5	0.1211	0.2391	0.2682	0.2684

Squared Multiple Correlations Between the WITH Variables and the First M Canonical Variables of the VAR Variables

M	1	2	3	4
y1	0.4799	0.6731	0.6737	0.6746
y2	0.5528	0.5765	0.5894	0.5899
y3	0.5852	0.6180	0.6239	0.6274
y4	0.4524	0.5314	0.5347	0.5438

۱۰۰. آزمون کای اسکور، آزمون دقیق فیشر و محاسبه معیارهای پیوند به روش گودمن-کروسکال در SAS با استفاده از دستورالعمل PROC FREQ جداول توافقی را تجزیه و تحلیل می نمایم. فرم کلی این دستور به صورت زیر است:

```
PROC FREQ;
```

```
TABLES /OPTIONS; فرم جدول
```

```
WEIGHT (متغیری که شامل مقادیر مشاهده شده است) نام متغیر پاسخ
```

فرم جدول را به صورت زیر مشخص می کنیم:

اگر یک جدول دوبعدی بوده، A صفت سطری و B صفت ستونی باشد، جدول به صورت زیر معرفی می شود:

```
TABLES A*B;
```

یا برای یک جدول سه بعدی با سه صفت A، B و C خواهیم داشت:

```
TABLES A*B*C;
```

دستوراتی که به جای OPTIONS به کار می روند به قرار هستند:

۱- دستور EXACT که آزمون دقیق فیشر را برای جداول ۲×۲ چاپ می کند.

۲- دستور CHISQ که آماره هایی را که بر اساس کای اسکور به دست می آیند، محاسبه می کند.

۳- دستور MEASURS معیارهای پیوند (گودمن-کروسکال) و ریسکهای نسبی جداول توافقی را محاسبه

می کند.

```
DATA;
DO I=1 TO 2;
  DO J=1 TO 2;
    INPUT Y @@;
    OUTPUT;
  END;
END;
CARDS;
2 6 18 14
;
PROC FREQ;
WEIGHT Y;
TABLES I*J/EXACT CHISQ;
RUN;
```

مثال: جدول توافقی زیر را در نظر بگیرید:

		J	
		1	2
I	1	2	6
	2	18	14

```
PROC FREQ;
WEIGHT Y;
TABLES I*J/MEASURS;
RUN;
```

```
The FREQ Procedure
Table of I by J
I      J
Frequency,
Percent ,
Row Pct ,
Col Pct ,      1,      2, Total
^^^^^^^^^
1      2      6      8
,      5.00 , 15.00 , 20.00
,      25.00 , 75.00 ,
,      10.00 , 30.00 ,
```

```

                ^^^^^^^^^^
                2 ,      18 ,      14 ,      32
                ,      45.00 ,      35.00 ,      80.00
                ,      56.25 ,      43.75 ,
                ,      90.00 ,      70.00 ,
                ^^^^^^^^^^
Total          20      20      40
                50.00      50.00      100.00
    
```

Statistics for Table of I by J

Statistic	DF	Value	Prob
Chi-Square	1	2.5000	0.1138
Likelihood Ratio Chi-Square	1	2.5943	0.1072
Continuity Adj. Chi-Square	1	1.4063	0.2357
Mantel-Haenszel Chi-Square	1	2.4375	0.1185
Phi Coefficient		-0.2500	
Contingency Coefficient		0.2425	
Cramer's V		-0.2500	

WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Fisher's Exact Test

```

                ^^^^^^^^^^
Cell (1,1) Frequency (F)      2
Left-sided Pr <= F            0.1176
Right-sided Pr >= F           0.9782

Table Probability (P)          0.0958
Two-sided Pr <= P              0.2351

Sample Size = 40
    
```

Statistics for Table of I by J

Statistic	Value	ASE
Gamma	-0.5882	0.2913
Kendall's Tau-b	-0.2500	0.1442
Stuart's Tau-c	-0.2000	0.1225
Somers' D C R	-0.3125	0.1764
Somers' D R C	-0.2000	0.1225
Pearson Correlation	-0.2500	0.1442
Spearman Correlation	-0.2500	0.1442
Lambda Asymmetric C R	0.2000	0.1265
Lambda Asymmetric R C	0.0000	0.0000
Lambda Symmetric	0.1429	0.0872
Uncertainty Coefficient C R	0.0468	0.0559
Uncertainty Coefficient R C	0.0648	0.0760
Uncertainty Coefficient Symmetric	0.0543	0.0643

The FREQ Procedure
 Statistics for Table of I by J
 Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	0.2593	0.0452	1.4861
Cohort (Col1 Risk)	0.4444	0.1288	1.5335
Cohort (Col2 Risk)	1.7143	0.9785	3.0033

Sample Size = 40