

统计实验设计(第2版) 第四章 方差分析试验设计与统计分析

$$1785.0 = 91 \times 19.725 + 0.0023 \times 18.725 + 2$$

第四章 方差分析试验设计与统计分析

第四章 方差分析试验设计与统计分析

第一节 随机单位组设计与 SPSS 应用

在实际中,经常会遇到在多个平均数比较时,供试验用的试验单元不都一样的情况,这就需要进行试验设计,安排试验。如果试验单位有一个方向上的差异,需要进行随机单位组设计。例如,现有 1、2、3 个品种,要比较平均亩产是否一样,试验单元情况如图 4-1 所示。左侧设计是错误的(品种 1 与 3 试验单元不一样),右侧设计是正确的。这就是随机单位组(randomized block design)设计。

地点变化

Ⅲ	3	2	3	3	1	2
Ⅱ	2	3	1	2	3	1
Ⅰ	1	1	2	1	2	3

图 4-1 试验对比图

随机单位组设计也称为随机区组(或窝组)设计。它是根据试验设计的局部控制的原则:组内无差异,组间可有较大差异,将初始条件基本相同的试验单位划归为一组,称为区组,每一区组内的试验单元数等于处理数,再将各处理分配至区组各试验单元上。

一、设计方法 with 要求

(一) 设计方法

将初始条件基本相同的试验单位划归一个区组,共有若干个单位组,每一单位组内的试验单元数等于处理数。

(二) 随机单位组设计要求

同一区组内各个试验单元尽可能一致,不同区组间的试验单元允许存在差异。

二、统计分析方法

(1) 单因素

设有 a 个处理,需要比较它们的平均数一致否,试验单元分为 b 组,所得试验数据见表 4-1。

表 4-1 单因素试验随机区组设计数据

因素(A)	区组(B)				处理合计 $x_{i\cdot}$	处理平均 $\bar{x}_{i\cdot}$
	1	2	...	b		
1	x_{11}	x_{12}	...	x_{1b}	$x_{1\cdot}$	$\bar{x}_{1\cdot}$
2	x_{21}	x_{22}	...	x_{2b}	$x_{2\cdot}$	$\bar{x}_{2\cdot}$
...	...	...	...	...	...	...
a	x_{a1}	x_{a2}	...	x_{ab}	$x_{a\cdot}$	$\bar{x}_{a\cdot}$
合计 $x_{\cdot j}$	$x_{\cdot 1}$	$x_{\cdot 2}$	...	$x_{\cdot b}$	$x_{\cdot\cdot}$	$\bar{x}_{\cdot\cdot}$

平方和与自由度的划分式为

$$SS_T = SS_A + SS_B + SS_e, \quad df_T = df_A + df_B + df_e \tag{4-1}$$

(2) 两因素或多因素

如果是 2 个或 2 个以上个因素时,要比较多个搭配平均数时,每个搭配为一个总体,进行方差分析。

设:A 分为 a 个水平、B 分为 b 个水平,共有 ab 个处理, n 次重复,随机区组设计(n 个区组,区组因素用 C 表示),试验数据见表 4-2。

表 4-2 两因素试验随机区组设计数据

AB 因素	区组(C)			
	1	2	...	n
11	x_{111}	x_{112}	...	x_{11n}
12	x_{121}	x_{122}	...	x_{12n}
...	...	...	...	...
1b	x_{1b1}	x_{1b2}	...	x_{1bn}
⋮	⋮	⋮	⋮	⋮
a1	x_{a11}	x_{a12}	...	x_{a1n}
a2	x_{a21}	x_{a22}	...	x_{a2n}
...	...	...	...	...
ab	x_{ab1}	x_{ab2}	...	x_{abn}

平方和与自由度的划分式为

$$SS_T = SS_{AB} + SS_C + SS_e = SS_A + SS_B + SS_{A \times B} + SS_C + SS_e$$
$$df_T = df_{AB} + df_C + df_e = df_A + df_B + df_{A \times B} + df_C + df_e \tag{4-2}$$

【例 4-1】 有一水稻品种比较试验,供试品种有 A、B、C、D、E、F 共 6 个,要比较它们的平均产量。重复 4 次,随机区组设计,小区计产面积 15m²,其田间排列和产量(kg/15m²)见图 4-2,试比较平均产量的差异。

A 15.3	B 18.0	C 16.6	D 16.4	E 13.7	F 17.0	I
D 17.3	F 17.6	E 13.6	C 17.8	A 14.9	B 17.6	II
C 17.6	A 16.2	F 18.2	B 18.6	D 17.3	E 13.9	III
B 18.3	D 17.8	A 16.2	E 14.0	F 17.5	C 17.8	IV

土壤肥力梯度方向 ↓

图 4-2 水稻品种比较试验的田间排列和产量(kg/15m²)

使用 SPSS 对**【例 4-1】**进行方差分析。

首先运行 SPSS,数据格式、操作方法和统计结果见图 4-3(a)~图 4-3(f)。

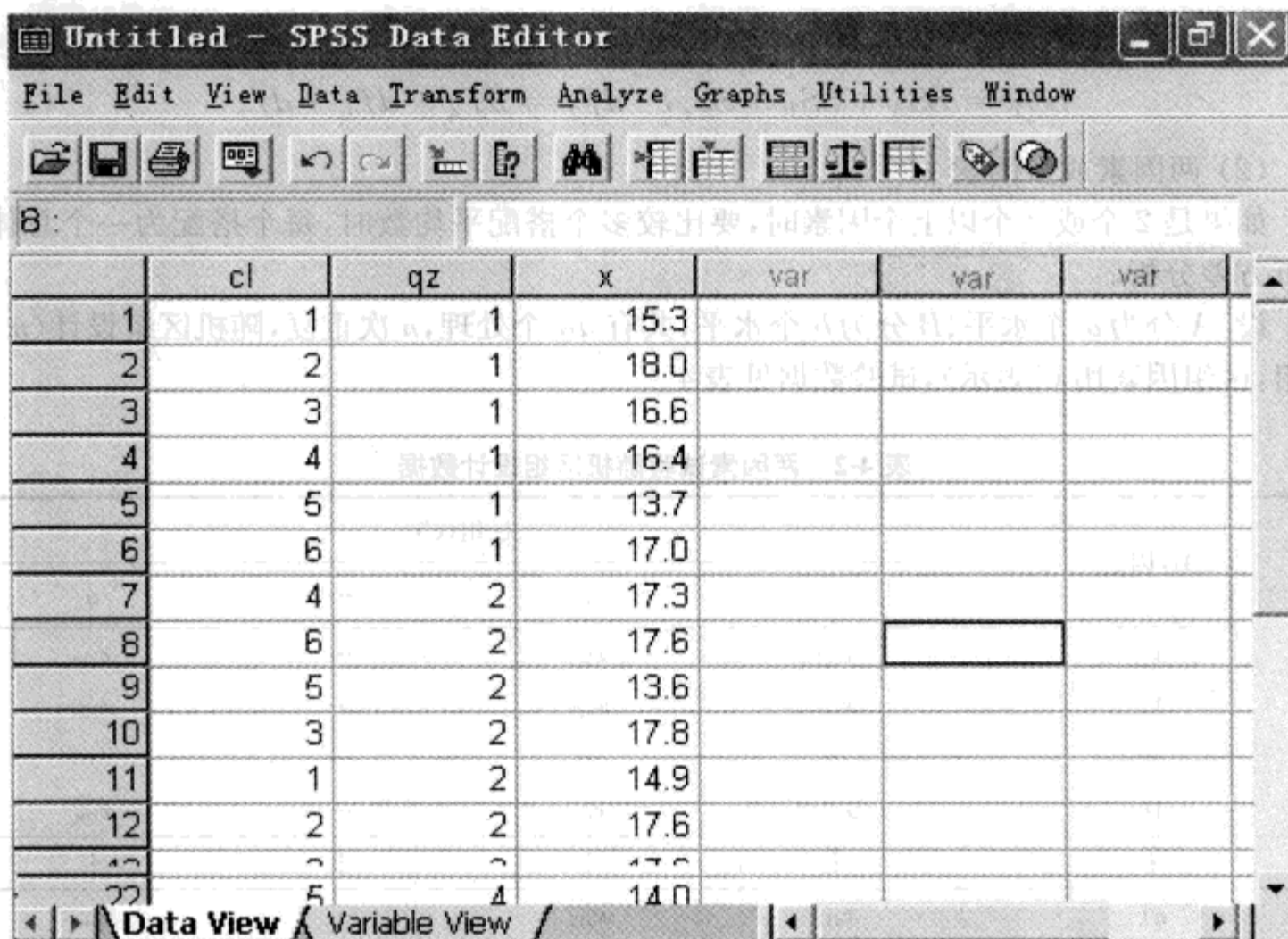


图 4-3(a) 试验数据格式

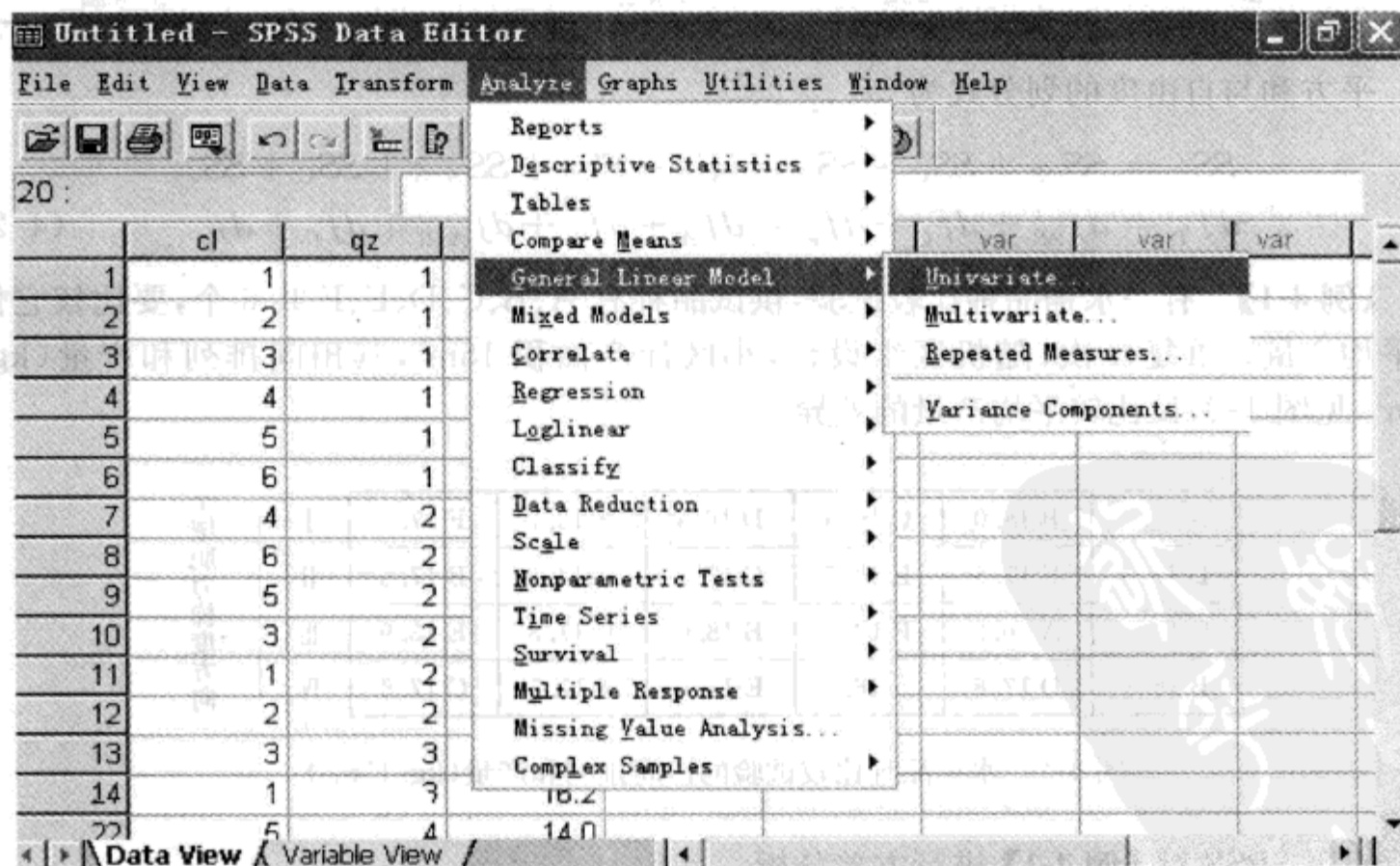


图 4-3(b) 运行方差分析

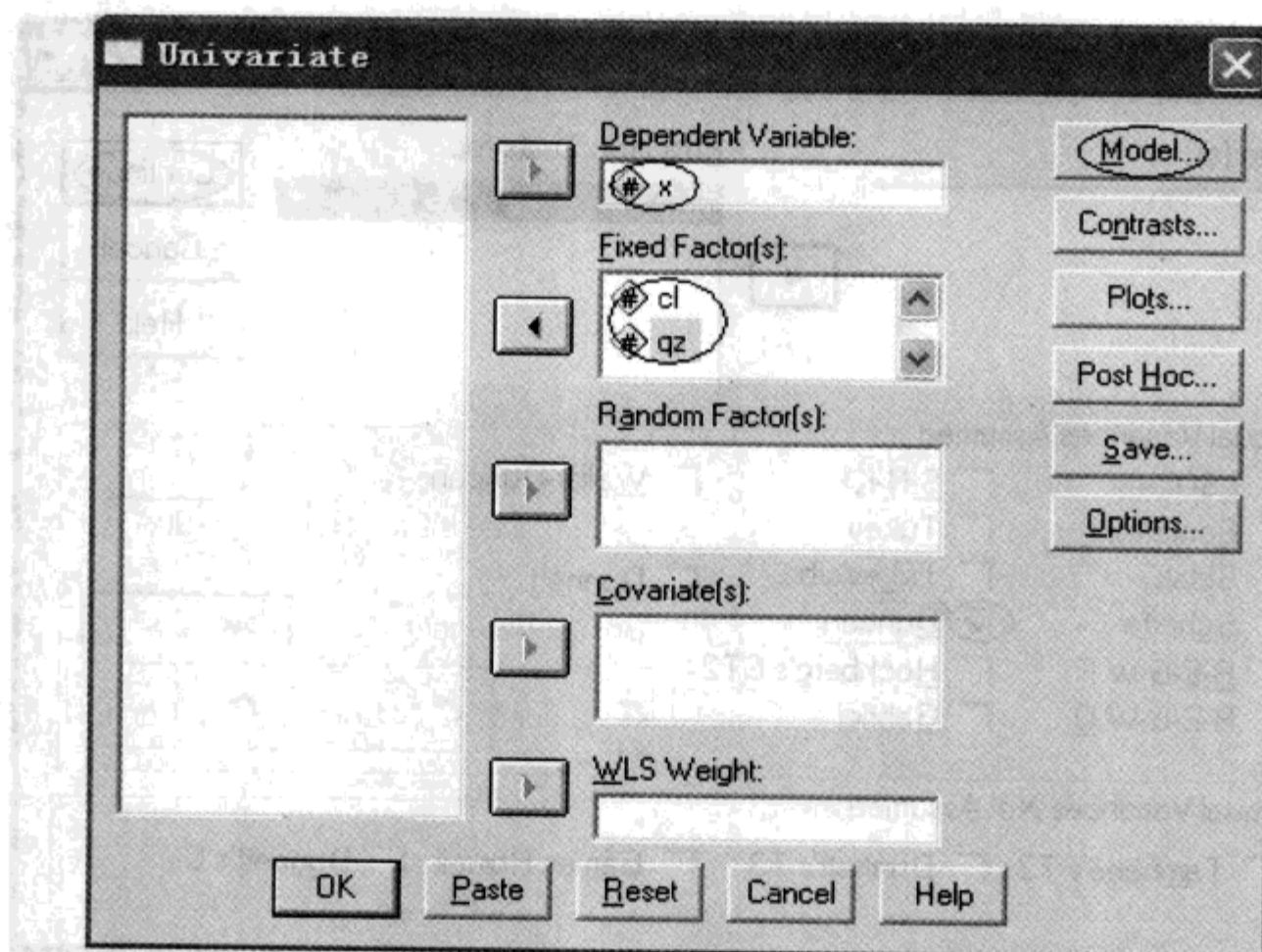


图 4-3(c) 选择依变量、自变量

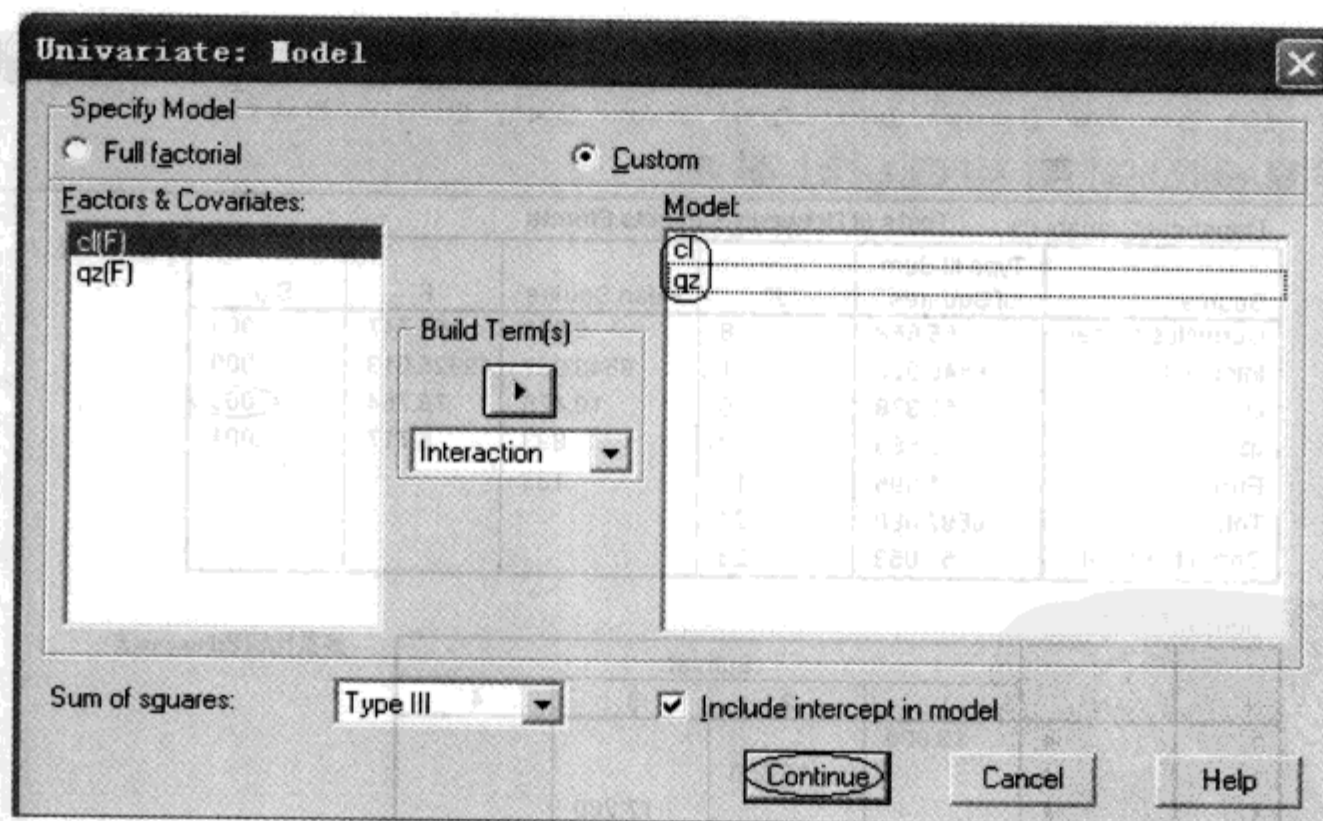


图 4-3(d) 选择没有交互的模型

由图 4-3(f) 方差分析表可知: 6 个品种平均产量极显著不都一样 (cl 项后 Sig. = 0.000, $p < 0.01$), 2 号 (B 品种) 平均产量显著高于其他品种, 品种 4、3、6 平均产量差异不显著, 但高于品种 1、5 平均产量, 品种 1 的平均产量显著高于品种 5 的平均产量。

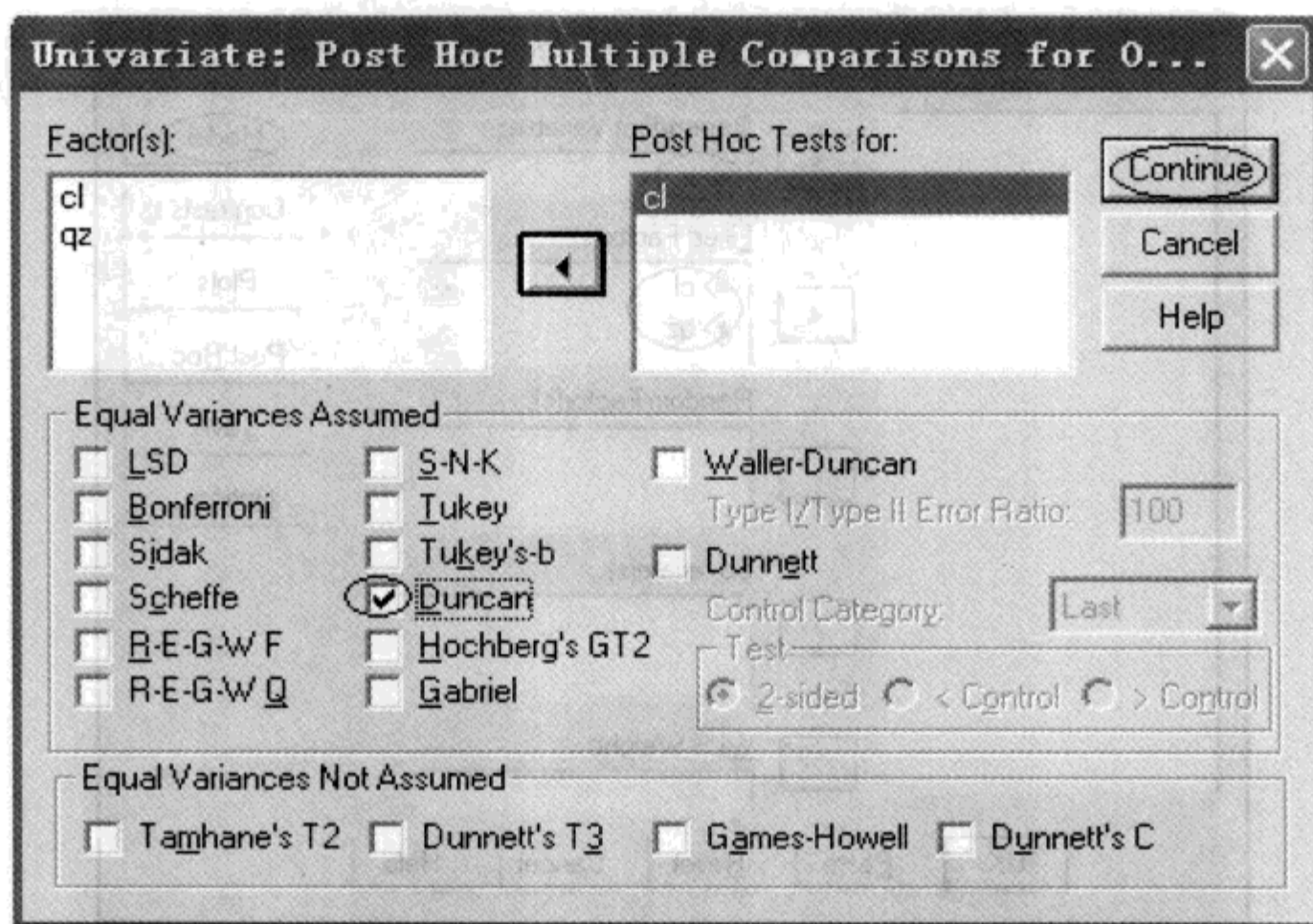


图 4-3(e) 选择比较的因素和比较方法

Output2 - SPSS Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help

Dependent Variable: x Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	55.058	8	6.882	51.747	.000
Intercept	6640.027	1	6640.027	49925.013	.000
cl	52.378	5	10.476	78.764	.000
qz	2.680	3	.893	6.717	.004
Error	1.995	15	.133		
Total	6697.080	24			
Corrected Total	57.053	23			

Duncan^{a,b}

cl	N	Subset			
		1	2	3	4
5	4	13.800			
1	4		15.650		
4	4			17.200	
3	4			17.450	
6	4			17.575	
2	4				18.125
Sig.		1.000	1.000	.187	1.000

多重比较表Duncan法

图 4-3(f) Output 窗口的方差分析结果

第二节 拉丁方设计与 SPSS 应用

在实际中,在多个平均数的比较时会遇到供试验用的试验单位有两个方向上的差异。例如,现有 1、2、3 个品种,要比较平均亩产是否一样,试验单位情况如图 4-4 所示。左侧设计是错误的(品种 3 与 1 试验单位不一样),右侧设计是正确的。这就是拉丁方设计。

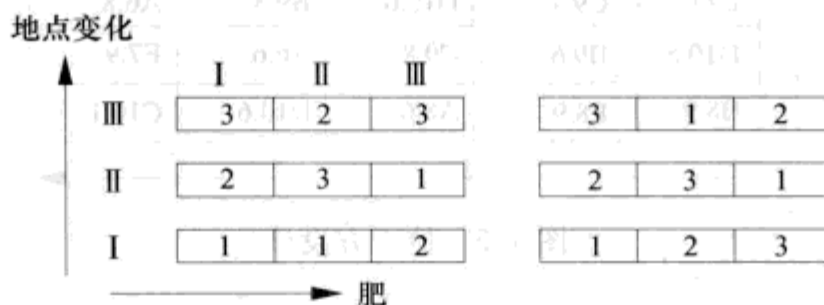


图 4-4 试验设计对比图

一、设计方法

常用拉丁方

常用的有 3×3 、 4×4 、 5×5 、 6×6 阶拉丁方。下面列出部分标准型拉丁方,供进行拉丁方设计时选用。

3×3		4×4			
	(1)	(2)	(3)	(4)	
A B C	A B C D	A B C D	A B C D	A B C D	
B C A	B A D C	B C D A	B D A C	B A D C	
C A B	C D B A	C D A B	C A D B	C D A B	
	D C A B	D A B C	D C B A	D C B A	

二、统计分析方法

拉丁方设计试验结果的统计分析,是将两个单位组因素与试验因素一起,按三因素试验单独观测值的方差分析法进行。将处理因素记为 A,行单位组因素记为 B,列单位组因素记为 C,行单位组数、列单位组数与处理数记为 n 。

平方和与自由度划分式为

$$\begin{aligned} SS_T &= SS_A + SS_B + SS_C + SS_e \\ df_T &= df_A + df_B + df_C + df_e \end{aligned} \quad (4-3)$$

F 检验: $F_A = MS_A / MS_e$, $F_B = MS_B / MS_e$, $F_C = MS_C / MS_e$ 。

现举例说明。

【例 4-2】 有一冬小麦施氮肥时期试验, 5 个处理为: A 不施氮肥; B 播种期(10 月 29 日)施氮; C 越冬期(12 月 13 日)施氮; D 拔节期(3 月 17 日)施氮; E 抽穗期(5 月 1 日)施氮。采用 5×5 拉丁方设计, 小区计产面积 32m^2 , 其田间排列和产量($\text{kg}/32\text{m}^2$)结果见图 4-5, 试做方差分析。

C10.1	A7.9	B9.8	E7.1	D9.6
A7.0	D10.0	E7.0	C9.7	B9.1
E7.6	C9.7	D10.0	B9.3	A6.8
D10.5	B9.6	C9.8	A6.6	E7.9
B8.9	E8.9	A8.6	D10.6	C10.1

图 4-5 拉丁方设计

使用 SPSS 对【例 4-2】进行方差分析。

首先运行 SPSS, SPSS 的数据格式、操作方法和统计结果见图 4-6(a)~图 4-6(f)。

	cl	hqz	lqz	x	var	var	var
1	3	1	1	10.1			
2	1	1	2	7.9			
3	2	1	3	9.8			
4	5	1	4	7.1			
5	4	1	5	9.6			
6	1	2	1	7.0			
7	4	2	2	10.0			
8	5	2	3	7.0			
9	3	2	4	9.7			
10	2	2	5	9.1			
11	5	3	1	7.6			
12	3	3	2	9.7			
13	4	3	3	10.0			
14	2	3	4	9.3			
15	1	3	5	6.8			

图 4-6(a) 试验数据格式

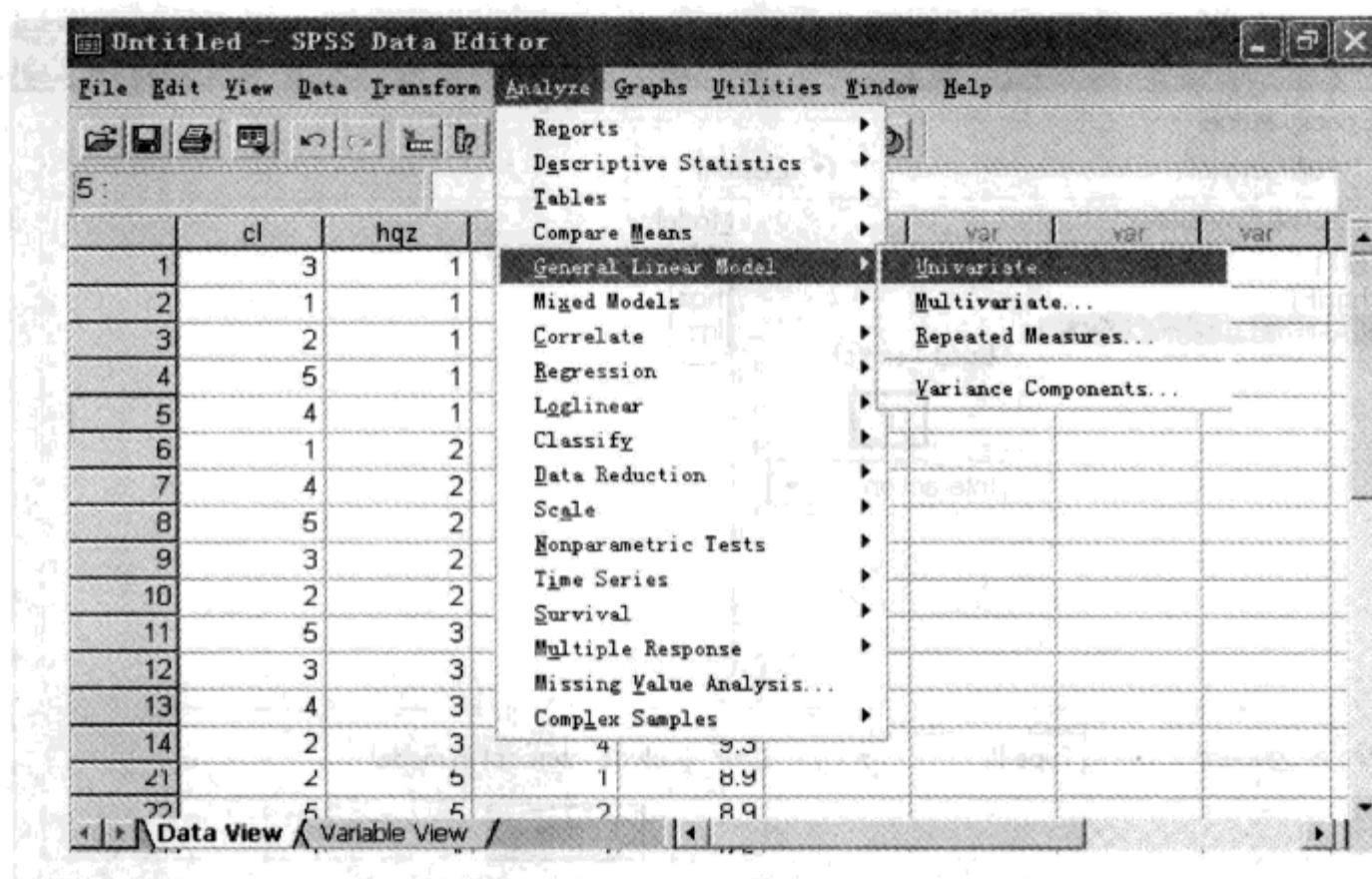


图 4-6(b) 运行方差分析

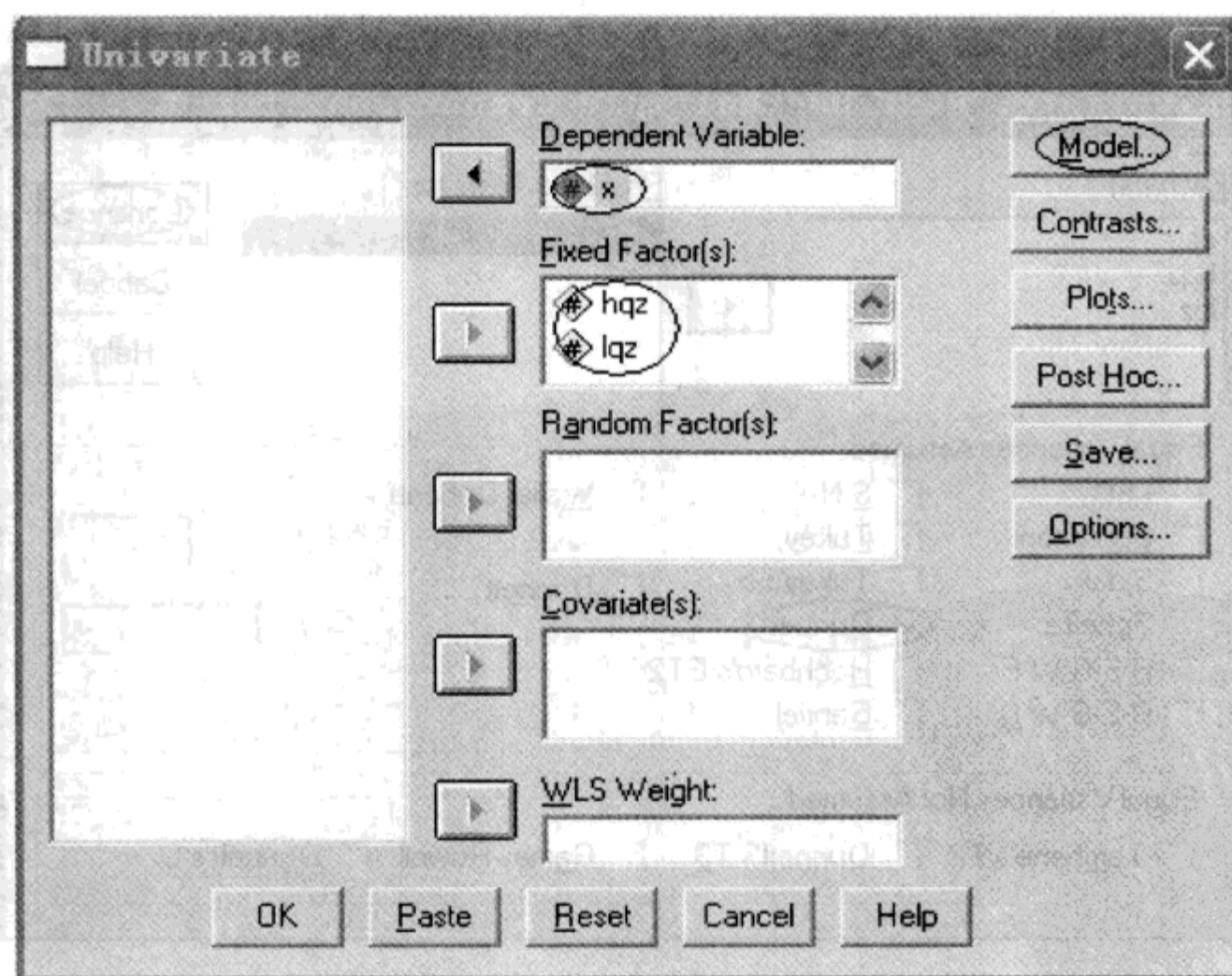


图 4-6(c) 选择依变量、自变量

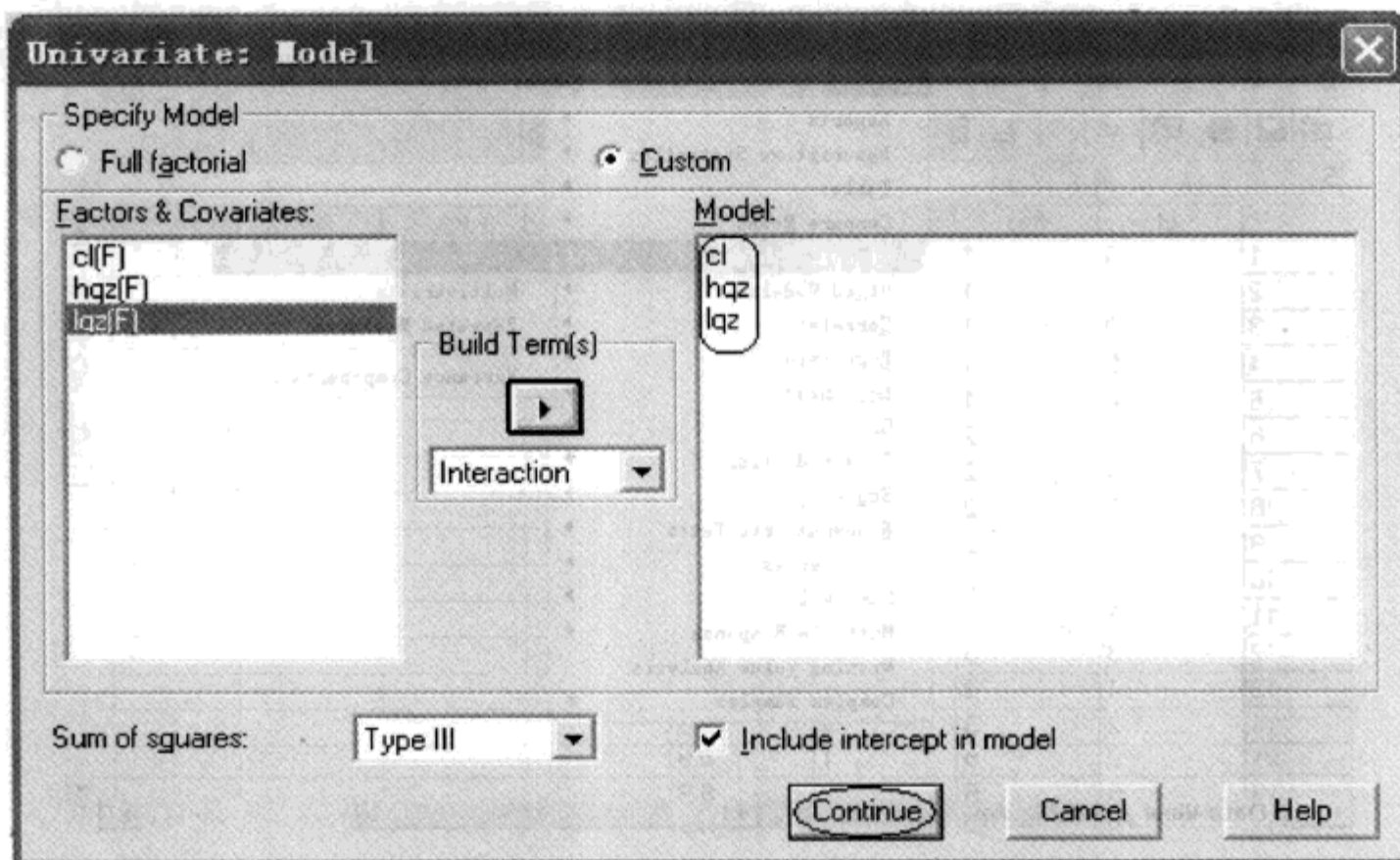


图 4-6(d) 选择没有交互的模型

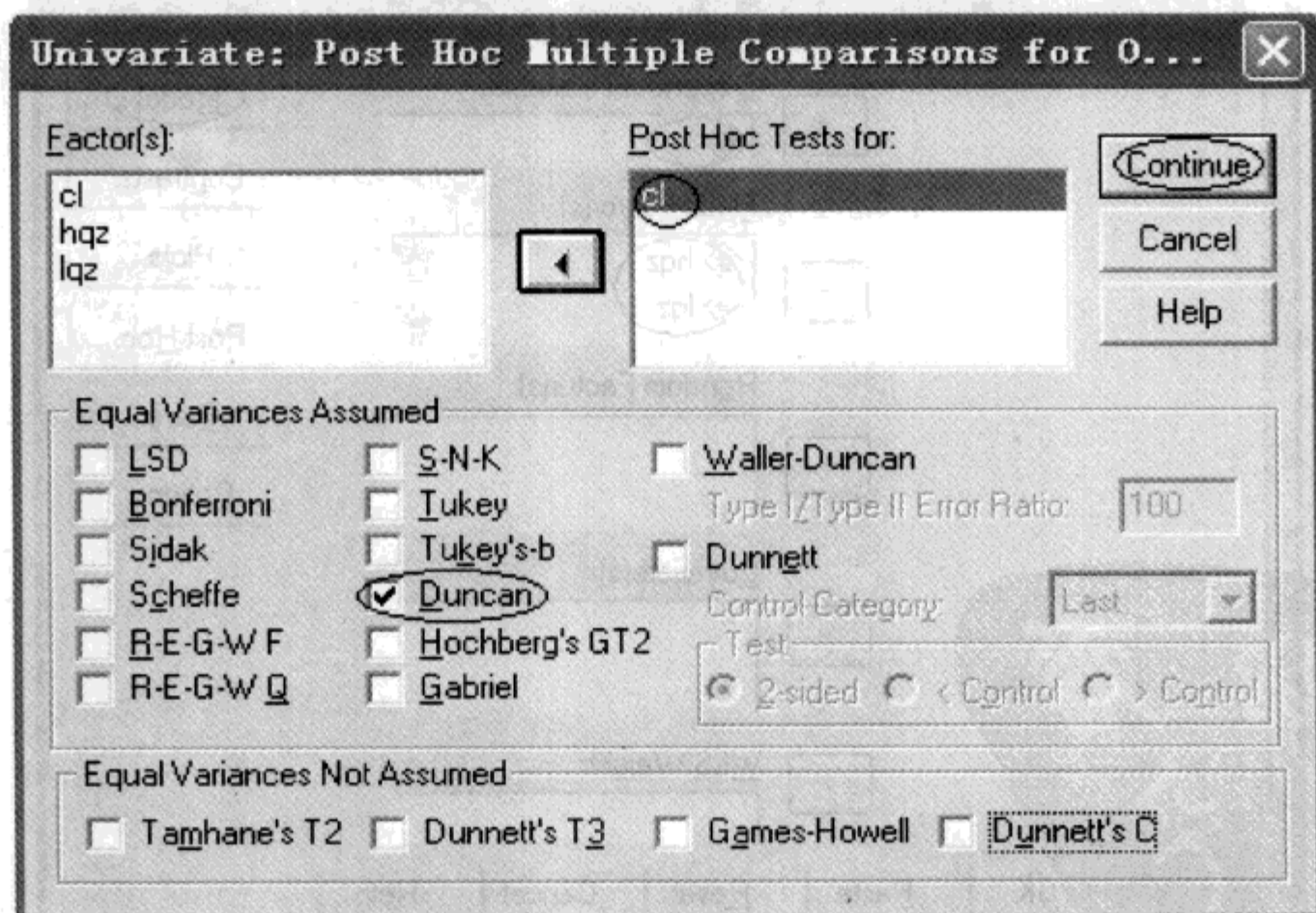


图 4-6(e) 选择比较的因素和平均数的比较方法

Dependent Variable: x Tests of Between-Subjects Effects						方差分析表
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	
Corrected Model	35.503	12	2.959	10.880	.000	
Intercept	1974.914	1	1974.914	7262.492	.000	
cl	32.206	4	8.052	29.609	.000	
hqz	2.170	4	.543	1.995	.159	
lqz	1.126	4	.282	1.036	.429	
Error	3.263	12	.272			
Total	2013.680	25				
Corrected Total	38.766	24				

Duncan ^{a,b}					多重比较表
cl	N	Subset			
		1	2	3	
1	5	7.380			
5	5	7.700			
2	5		9.340		
3	5		9.880	9.880	
4	5			10.140	
Sig.		.351	.128	.446	

图 4-6(f) Output 窗口的方差分析结果

由图 4-5(f) 方差分析表、多重比较表可知: 5 个品种平均产量极显著不都一样 (cl 项后 Sig. = 0.000, $p < 0.01$), 施肥方法 4 平均产量与施肥方法 3 平均产量差异不显著, 但施肥方法 4 平均产量显著高于施肥方法 1、2、5 平均产量。施肥方法 3 平均产量与施肥方法 2 平均产量差异不显著, 但施肥方法 3 和施肥方法 2 平均产量显著高于施肥方法 1、5 平均产量。施肥方法 5 平均产量与施肥方法 1 平均产量差异不显著。

第三节 交叉设计与 SPSS 应用

一、交叉设计的概念

1. 概念

交叉设计 (cross over design) 亦称反转试验设计, 用以比较多个处理的平均数是否一致。将试验分期进行。现仅就两个总体平均数的比较方法进行介绍。

2. 特点

- (1) 特别适用于来源较困难的大动物;
- (2) 节省试验动物;
- (3) 可以较好地消除试验动物个体之间的差异对试验结果的影响。

二、设计方法 (单因素试验)

将条件一致的若干试验动物随机分成头数相等的两个群组, 对于每个群组每头动物使其在不同时期接受不同处理。

1. 2×2 交叉设计 (两个处理两个时期)

2×2 交叉设计见表 4-3(a)。

表 4-3(a) 2×2 交叉设计

群 别	动物编号	时 期		$d_{ij}=x_{ij1}-x_{ij2}$	
		C_1	C_2		
1	11	处理 A_1	x_{111}	x_{112}	d_{11}
	12		x_{121}	x_{122}	d_{12}
	$\vdots$		$\vdots$	$\vdots$	$\vdots$
	1n		x_{1n1}	x_{1n2}	d_{1n}
2	21	对照 A_2	x_{211}	x_{212}	d_{21}
	22		x_{221}	x_{222}	d_{22}
	$\vdots$		$\vdots$	$\vdots$	$\vdots$
	2n		x_{2n1}	x_{2n2}	d_{2n}

2. 2×3 交叉设计(两个处理三个时期)

2×3 交叉设计见表 4-3(b)。

表 4-3(b) 2×3 交叉设计

群 别	时 期			$d_{ij}=C_1-2C_2+C_3$
	C_1	C_2	C_3	
1	处理 A_1	对照 A_2	处理 A_1	d_1
2	对照 A_2	处理 A_1	对照 A_2	d_2

三、统计分析方法

平方和与自由度划分式

$$SS_T = SS_A + SS_e, df_T = df_A + df_e$$

将 $d_{11}, \dots, d_{1n}$ 看成来自一总体的样本, 将 $d_{21}, \dots, d_{2n}$ 看成来自另一总体的样本, 进行总体平均数的比较。

1. 2×2 交叉设计

- (1) 计算各群各动物的差值 $d_{ij} (d_{ij}=C_1-C_2)$;
- (2) 对差值 d_{ij} 按单因素二水平进行方差分析。

2. 2×3 交叉设计

- (1) 计算各群各动物的差值 $d_{ij} (d_{ij}=C_1-2C_2+C_3)$;
- (2) 对差值 d_{ij} 按单因素二水平进行方差分析, 试验数据见表 4-3(c)。

表 4-3(c) 2×3 交叉设计资料统计分析表

群别	动物编号	C_1	C_2	C_3	d_{ij}	合 $d_{i\cdot}$
		A_1	A_2	A_1	$C_1-2C_2+C_3$	
1	B_{11}	x_{111}	x_{112}	x_{113}	d_{11}	$d_{1\cdot}$
	B_{12}	x_{121}	x_{122}	x_{123}	d_{12}	
	...	...	...	...	...	
2	B_{21}	x_{211}	x_{212}	x_{213}	d_{21}	$d_{2\cdot}$
	B_{22}	x_{221}	x_{222}	x_{223}	d_{22}	
	...	...	...	...	...	
总和						$d_{\cdot\cdot}$

设:处理数为 k , 每组动物数为 n :

平方和与自由度计算:

$$C = \frac{d_{..}^2}{kn}$$
$$SS_T = \sum d_{ij}^2 - C, \quad df_T = kn - 1$$
$$SS_A = \frac{1}{n} \sum d_{i.}^2 - C, \quad df_A = k - 1$$
$$SS_e = SS_T - SS_A, \quad df_e = df_T - df_A$$

(4-4)

【例 4-3】 俞渭江等应用蜂王浆滴喂母鸡,提高产蛋量研究。采用交叉设计,两个处理是:处理、对照。从 1979 年 10 月开始隔日口服(8mg/只),第一期 30 天,中间间隔为 10 天。第二期从 11 月 10 日开始,时间 30 天。选择前 3 个月产蛋平均 11 个左右的母鸡 20 只,分为两组,试验数据见表 4-4。

表 4-4 试验数据

组 别	鸡号	时 期		$d=C_1-C_2$
		C_1	C_2	
		处理 1	处理 2	
1	1	13	7	6
	2	19	13	6
	3	18	18	0
	4	21	10	11
	5	13	7	6
	6	13	9	4
	7	18	13	5
	8	12	14	-2
	9	10	10	0
	10	11	8	3
2	鸡号	处理 2	处理 1	
	11	11	6	5
	12	15	16	-1
	13	13	18	-5
	14	17	18	-1
	15	9	12	-3
	16	4	6	-2
	17	9	15	-6
	18	7	15	-8
	19	11	17	-6
	20	8	17	-9

使用 SPSS 对【例 4-3】进行方差分析。
首先运行 SPSS,SPSS 的数据格式、操作方法和统计结果见图 4-7(a)~图 4-7(c)。

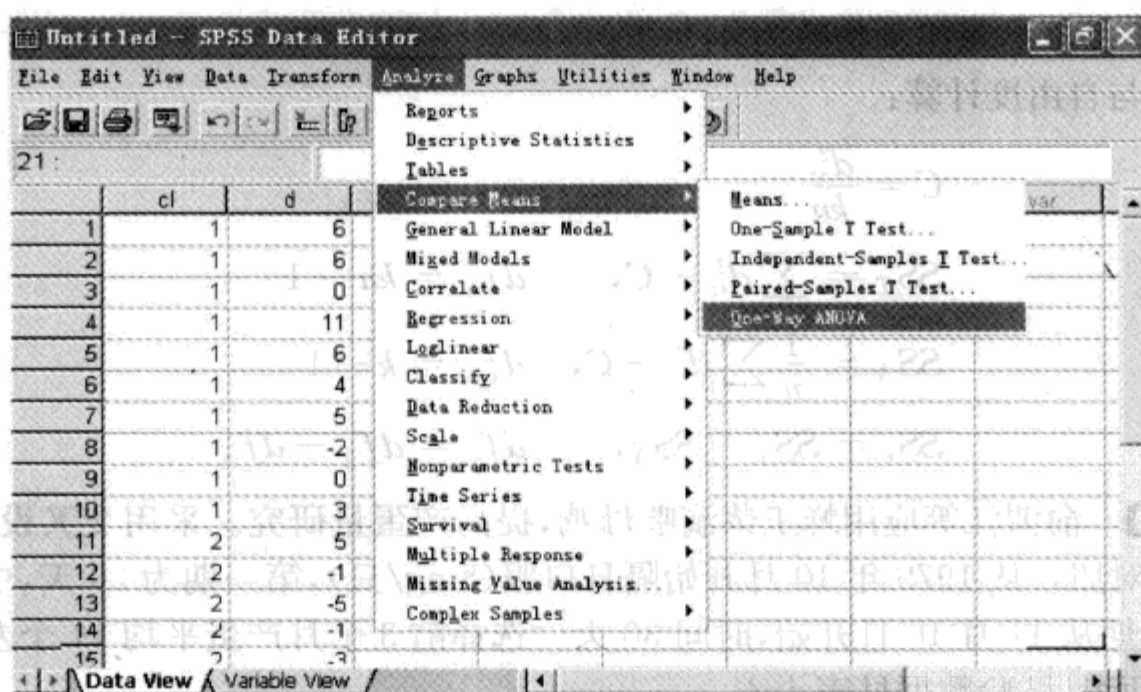


图 4-7(a) 试验数据格式与运行方差分析

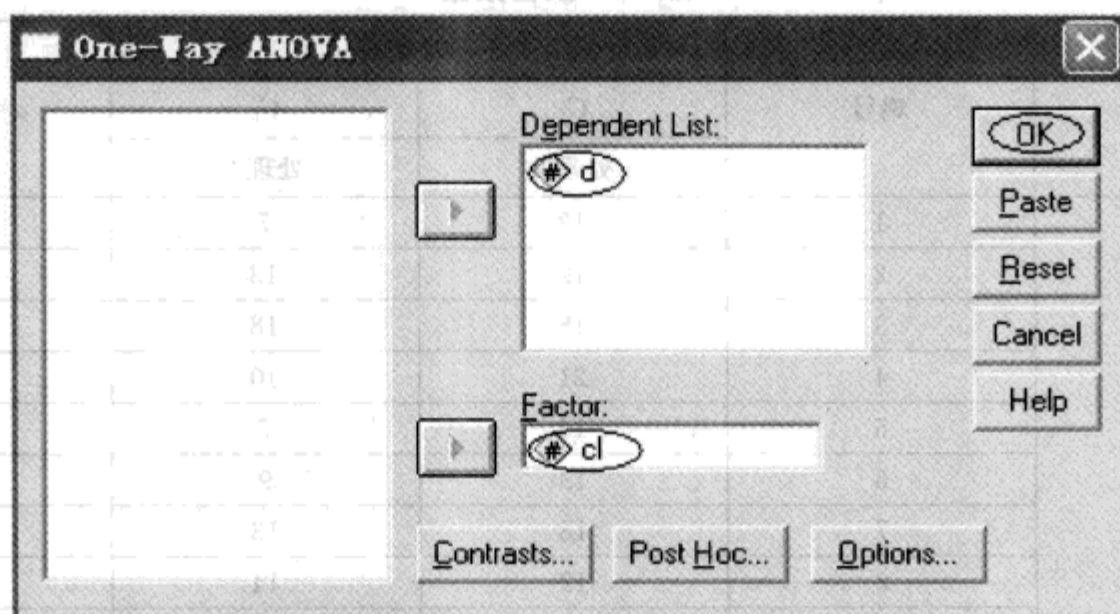


图 4-7(b) 选择因变量、自变量

Output5 - SPSS Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help

→ Oneway

ANOVA

d	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	281.250	1	281.250	17.870	.001
Within Groups	283.300	18	15.739		
Total	564.550	19			

图 4-7(c) Output 窗口的方差分析结果

由图 4-7(c)方差分析可知:处理、对照平均产蛋个数差异极显著($\text{Sig.} = 0.001$)。

第四节 正交试验设计与 SPSS 应用

一、正交试验设计的概念与特点

(一) 正交试验设计的概念

利用正交表安排多因素多水平的试验设计,即为正交试验设计(orthogonal experimental design)。

(二) 正交试验设计的特点

用部分水平组合试验代替全部水平组合试验,设计重点是安排试验点,试验目的主要是挑选最优水平组合(最佳配方、最佳工艺条件等)或重要因素。

二、正交表

(一) 正交表结构

如正交表 $L_9(3^4)$ (表 4-5)。

表 4-5 正交表 $L_9(3^4)$

试验号(处理号)	列号(因素)			
	1(A)	2(B)	3(C)	4(D)
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

注: L : 正交表; 9: 要做 9 次试验;

3^4 : 指数 4 表示该表有 4 列, 最多可安排 4 个因素; 底数 3 表示该表适用于每因素 3 水平试验, 每列都由数字(水平号)1、2、3 组成;

$L_9(3^4)$: 用这张正交表安排试验, 最多可安排 4 个因素、每个因素都为 3 水平, 一共做 9 次试验。

(二) 正交表的类别

(1) 相同水平正交表: 如 $L_9(3^4)$ 、 $L_8(2^7)$ 。

(2) 混合水平正交表: 如 $L_{16}(4^3 \times 2^4)$ 。

三、正交设计方法

(一) 根据因素与水平数选用正交表

正交表选用原则：表中每列水平数=该列安排因素的水平数；表的列数≥安排因素的个数。

(二) 作表头设计

把各因素分别放在选用的正交表的表头适当列上。

- (1) 若不考虑因素间的交互作用，则每个因素任意占一列即可；
- (2) 若考虑因素间的交互作用或防止效应的混杂，则应根据正交表的使用表，把因素放在正交表的适当列上。

(三) 列出试验方案

对表头上有因素的各列，将表中的数字(水平号)换成该因素的具体水平，得到试验方案。

四、正交试验资料统计分析方法

(一) 无重复正交试验资料

设：因素 A、B、C 水平数分别为 k_A, k_B, k_C ；

因素 A、B、C 水平重复数分别为 a, b, c ；

试验次数(正交表的行数)为 n 。

1. 平方和与自由度划分式

$$\begin{aligned} SS_T &= SS_A + SS_B + SS_C + SS_e \\ df_T &= df_A + df_B + df_C + df_e \end{aligned} \quad (4-5)$$

[假定 A、B、C 水平数均为 3，选用 $L_9(3^4)$ 正交表。]

2. 无重复正交试验结果数据整理(表 4-6)

表 4-6 $L_9(3^4)$ 无重复试验结果基本计算表

试验号(处理号)	列号(因素)				试验指标 x_i
	1(A)	2(B)	3(C)	4(空)	
1	1	1	1	1	x_1
2	1	2	2	2	x_2
3	1	3	3	3	x_3
4	2	1	2	3	x_4
5	2	2	3	1	x_5
6	2	3	1	2	x_6
7	3	1	3	2	x_7
8	3	2	1	3	x_8
9	3	3	2	1	x_9

续表

试验号(处理号)	列号(因素)				试验指标 x_i
	1(A)	2(B)	3(C)	4(空)	
T_1	T_{A1}	T_{B1}	T_{C1}		T
T_2	T_{A2}	T_{B2}	T_{C2}		
T_3	T_{A3}	T_{B3}	T_{C3}		
$\bar{x}_1$	$\bar{x}_{A1}$	$\bar{x}_{B1}$	$\bar{x}_{C1}$		$\bar{X}$
$\bar{x}_2$	$\bar{x}_{A2}$	$\bar{x}_{B2}$	$\bar{x}_{C2}$		
$\bar{x}_3$	$\bar{x}_{A3}$	$\bar{x}_{B3}$	$\bar{x}_{C3}$		

3. 平方和与自由度的计算

$$C = T^2/n$$
$$SS_T = \sum x^2 - C, \quad df_T = n - 1$$
$$SS_A = \frac{1}{a} \sum T_{Ai}^2 - C, \quad df_A = k_A - 1$$
$$SS_B = \frac{1}{b} \sum T_{Bi}^2 - C, \quad df_B = k_B - 1$$
$$SS_C = \frac{1}{c} \sum T_{Ci}^2 - C, \quad df_C = k_C - 1$$
$$SS_e = SS_T - SS_A - SS_B - SS_C, \quad df_e = df_T - df_A - df_B - df_C$$

4. F 检验

计算 F_A 、 F_B 、 F_C ，均以 MS_e 为分母。

【例 4-4】 水稻栽培试验，选择了 3 个水稻优良品种(A)：二九矮、高二矮、窄叶青，3 种密度(B)：15 万、20 万、25 万苗/666.7m²；3 种施氮量(C)：3、5、8kg/666.7m²，试采用正交设计安排一个试验方案。正交设计见表 4-7。

表 4-7 正交试验方案

试验号	因 素		
	A	B	C
	1	2	3
1	1(二九矮)	1(15)	1(3)
2	1(二九矮)	2(20)	2(5)
3	1(二九矮)	3(25)	3(8)
4	2(高二矮)	1(15)	2(5)
5	2(高二矮)	2(20)	3(8)
6	2(高二矮)	3(25)	1(3)
7	3(窄叶青)	1(15)	3(8)
8	3(窄叶青)	2(20)	1(3)
9	3(窄叶青)	3(25)	2(5)

根据表 4-7，1 号试验处理是 $A_1B_1C_1$ ，即品种二九矮、密度 15 万苗/666.7m²、施氮量 3kg/666.7m²；2 号试验处理是 $A_1B_2C_2$ ，即品种二九矮、密度 20 万苗/666.7m²、施氮量 5kg/666.7m²；…；9 号试验处理为 $A_3B_3C_2$ ，即品种窄叶青、密度 25 万苗/666.7m²、施氮量 5kg/666.7m²。试验结果见表 4-8。

表 4-8 正交试验结果表

试验号	因素			产量/(kg/666.7m ²)
	A	B	C	
	1	2	3	
1	1	1	1	340.0(x_1)
2	1	2	2	422.5(x_2)
3	1	3	3	439.0(x_3)
4	2	1	2	360.0(x_4)
5	2	2	3	492.5(x_5)
6	2	3	1	439.0(x_6)
7	3	1	3	392.0(x_7)
8	3	2	1	363.5(x_8)
9	3	3	2	462.5(x_9)
T_1	1201.5	1092.0	1142.5	3711.0(T)
T_2	1291.5	1278.5	1245.0	
T_3	1218.0	1340.5	1323.5	
$\bar{x}_1$	400.50	364.00	380.83	
$\bar{x}_2$	430.50	426.17	415.00	
$\bar{x}_3$	406.00	446.83	441.17	

由表 4-8 试验数据得出以下两点。

(1) 因素水平的不同造成试验指标的变异 $R=\max(\bar{x}_i-\bar{x}_j)$
因为

$$R_A = 430.5 - 400.5 = 30, \quad R_B = 446.83 - 364.00 = 82.83, \\ R_C = 441.17 - 380.83 = 60.34$$

对试验指标影响大小因素的排序是 B、C、A，因此，考虑次序也是 B、C、A。如果希望试验指标大，优化搭配是 $A_2B_3C_3$ 。

(2) 因素水平的不同，平均数是否一致，进行方差分析，见表 4-9。

表 4-9 方差分析表

变异来源	SS	df	MS	F	$F_{0.05(2,2)}$
品种(A)	1 530.50	2	765.25	<1	19.00
密度(B)	11 153.17	2	5 576.59	3.64	
施氮量(C)	5 492.17	2	2 746.09	1.79	
误差	3 062.16	2	1 531.08		
总变异	21 238.00	8			

使用 SPSS 对【例 4-4】进行方差分析。

首先运行 SPSS，SPSS 的数据格式、操作方法和统计结果见图 4-8(a)~图 4-8(e)。

5 :

	A	B	C	X	var	var var
1	1	1	1	340.0		
2	1	2	2	422.5		
3	1	3	3	439.0		
4	2	1	2	360.0		
5	2	2	3	492.5		
6	2	3	1	439.0		
7	3	1	3	392.0		
8	3	2	1	363.5		
9	3	3	2	462.5		
10						

22

Data View / Variable View

图 4-8(a) 数据格式

20 :

	A	B
1	1	1
2	1	2
3	1	3
4	2	1
5	2	2
6	2	3
7	3	1
8	3	2
9	3	3
10		
11		
12		
13		
14		

22

Data View / Variable View

Reports

- Descriptive Statistics
- Tables
- Compare Means
- General Linear Model
 - Univariate
 - Multivariate...
 - Repeated Measures...
 - Variance Components
- Mixed Models
- Correlate
- Regression
- Loglinear
- Classify
- Data Reduction
- Scale
- Nonparametric Tests
- Time Series
- Survival
- Multiple Response
- Missing Value Analysis...
- Complex Samples

图 4-8(b) 运行方差分析

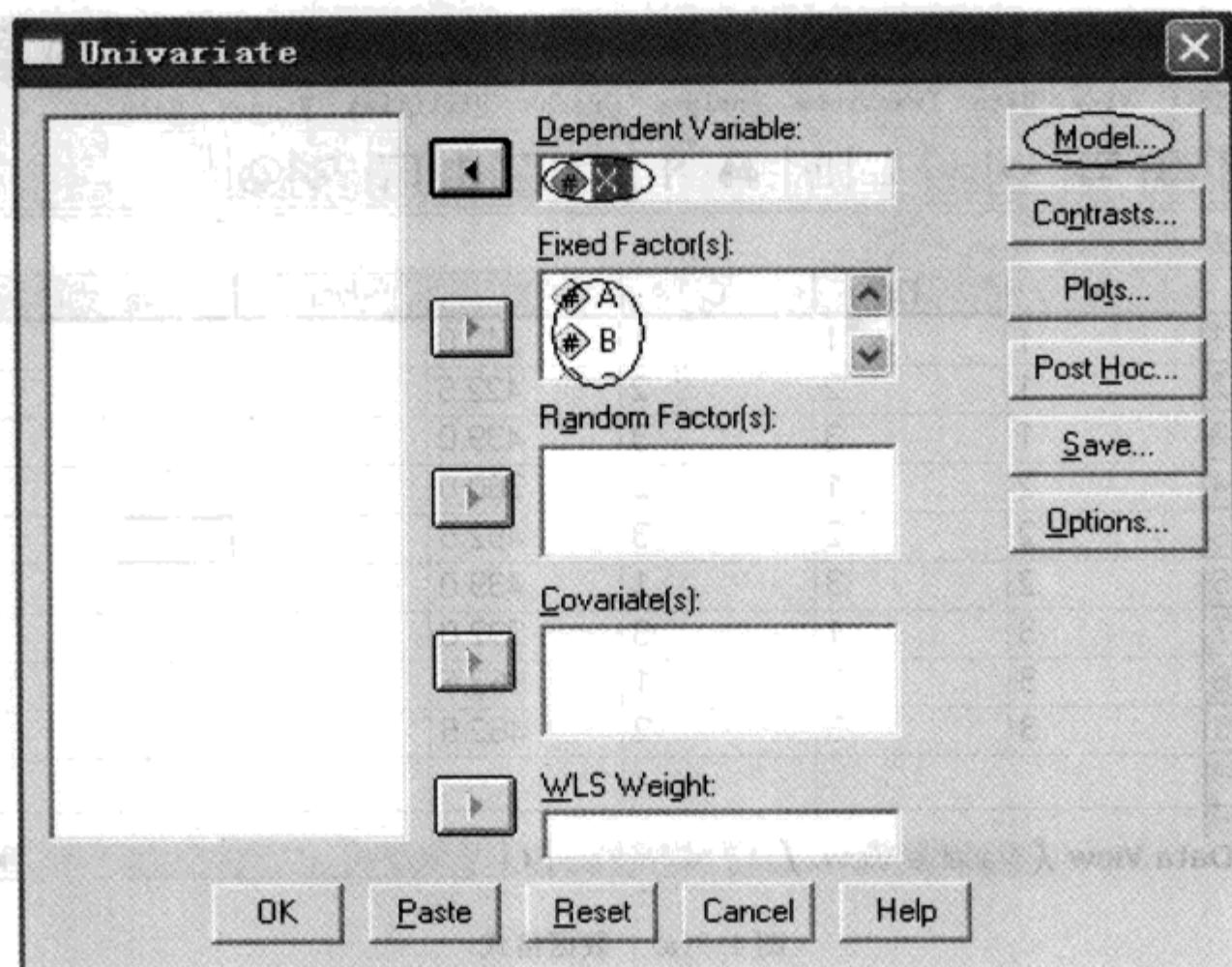


图 4-8(c) 选择处理号和试验数据

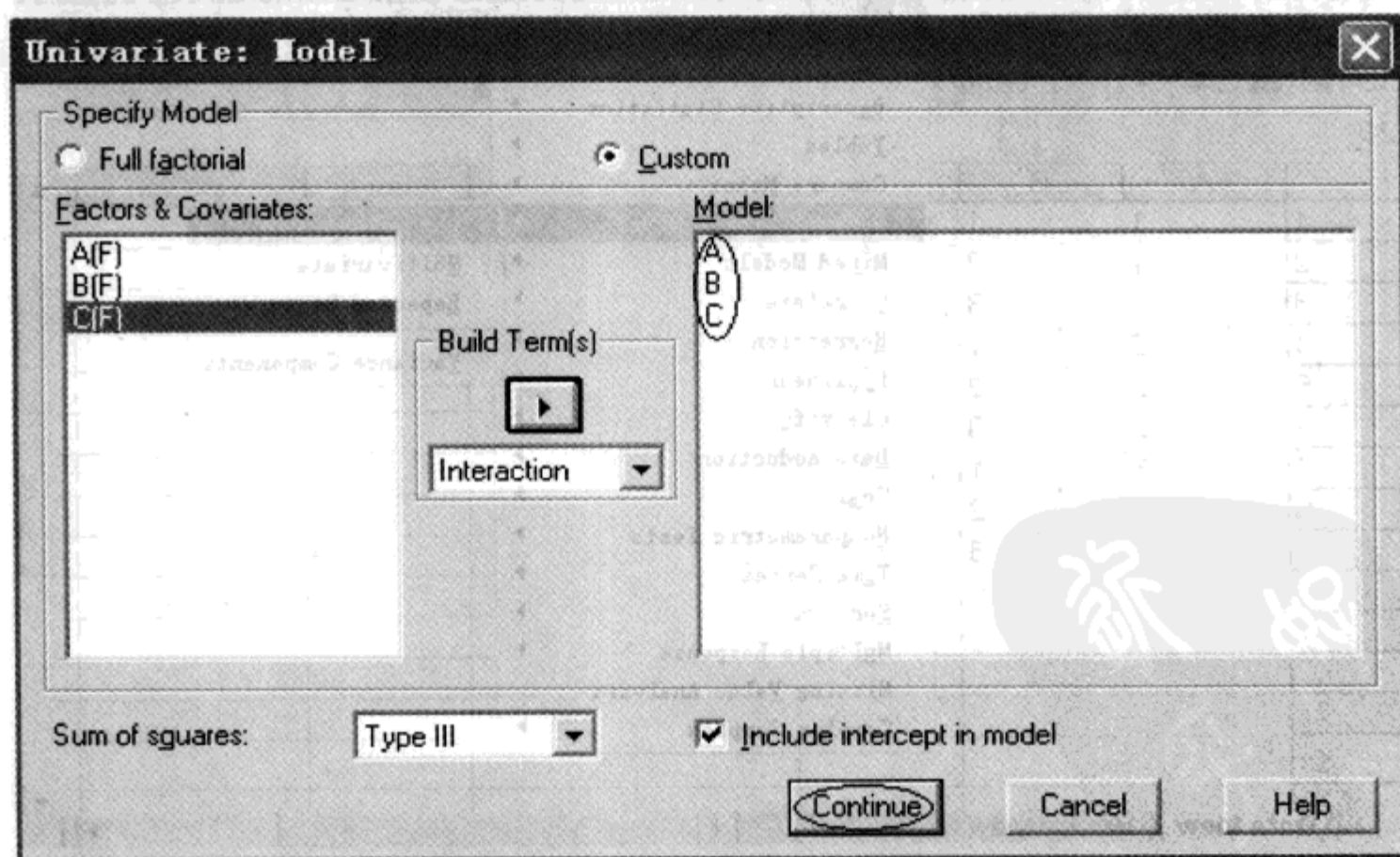


图 4-8(d) 选择模型

Output6 - SPSS Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help

Univariate Analysis of Variance

Tests of Between-Subjects Effects

Dependent Variable: X

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	18175.833 ^a	6	3029.306	1.979	.373
Intercept	1530169.000	1	1530169.000	999.403	.001
A	1530.500	2	765.250	.500	.667
B	11153.167	2	5576.583	3.642	.215
C	5492.167	2	2746.083	1.794	.358
Error	3062.167	2	1531.083		
Total	1551407.000	9			
Corrected Total	21238.000	8			

方差分析表

图 4-8(e) Output 窗口的方差分析结果

由图 4-8(e) 方差分析表明, 每个因素水平的不同, 其平均产量差异不显著。此时, 可从表 4-8 中选择平均数大的水平 A_2 、 B_3 、 C_3 组合成最优水平组合 $A_2B_3C_3$, 所得到的最优水平组合 $A_2B_3C_3$ 在试验方案中不存在, 因此, 应将最优水平组合 $A_2B_3C_3$ 与试验方案中产量最高第 5 号试验处理的水平组合 $A_2B_2C_3$ 再做一次验证性试验。

(二) 有重复正交试验资料(重复采用随机单位组设计)

设因素 A、B、C 水平数分别为 k_A 、 k_B 、 k_C ;

因素 A、B、C(所在列)水平重复数分别为 a 、 b 、 c ;

单位组数 r , 试验次数为 rn 。

假定 A、B、C 水平数均为 3, 选用 $L_9(3^4)$ 正交表。

1. 平方和与自由度划分式

$$SS_T = SS_t + SS_r + SS_e, \quad df_T = df_t + df_r + df_e$$

而

$$SS_t = SS_A + SS_B + SS_C + SS_{A \times B \times C}, \quad df_t = df_A + df_B + df_C + df_{A \times B \times C}$$

$$SS_T = SS_A + SS_B + SS_C + SS_{A \times B \times C} + SS_r + SS_e$$

$$df_T = df_A + df_B + df_C + df_{A \times B \times C} + df_r + df_e \quad (4-6)$$

式中, SS_t 为处理间(水平组合)离差平方和; SS_r 为单位组间离差平方和; $SS_{A \times B \times C}$ 为交互作用平方和; SS_e 为试验误差平方和。

2. 有重复正交试验结果数据整理(重复采用随机单位组设计)(表 4-10)

表 4-10 $L_9(3^4)$ 有重复试验结果基本计算表(重复采用随机单位组设计)

试验号	列号(因素)				试验指标 x_{ij}		处理总和 $T_i \cdot$
	1(A)	2(B)	3(C)	4(空)	单位组 1	单位组 2	
1	1	1	1	1	x_{11}	x_{12}	$T_1 \cdot$
2	1	2	2	2	x_{21}	x_{22}	$T_2 \cdot$
3	1	3	3	3	x_{31}	x_{32}	$T_3 \cdot$
4	2	1	2	3	x_{41}	x_{42}	$T_4 \cdot$
5	2	2	3	1	x_{51}	x_{52}	$T_5 \cdot$
6	2	3	1	2	x_{61}	x_{62}	$T_6 \cdot$
7	3	1	3	2	x_{71}	x_{72}	$T_7 \cdot$
8	3	2	1	3	x_{81}	x_{82}	$T_8 \cdot$
9	3	3	2	1	x_{91}	x_{92}	$T_9 \cdot$
T_1	T_{A1}	T_{B1}	T_{C1}		单位组总和		
T_2	T_{A2}	T_{B2}	T_{C2}		$T \cdot 1$	$T \cdot 2$	T
T_3	T_{A3}	T_{B3}	T_{C3}				
$\bar{x}_1$	$\bar{x}_{A1}$	$\bar{x}_{B1}$	$\bar{x}_{C1}$				
$\bar{x}_2$	$\bar{x}_{A2}$	$\bar{x}_{B2}$	$\bar{x}_{C2}$				
$\bar{x}_3$	$\bar{x}_{A3}$	$\bar{x}_{B3}$	$\bar{x}_{C3}$				

3. 平方和与自由度的计算

$$C = T^2 / (r \cdot n)$$
$$SS_T = \sum x^2 - C, \quad df_T = m - 1$$
$$SS_t = \frac{\sum_{i=1}^n T_i^2}{r} - C, \quad df_t = n - 1$$
$$SS_r = \frac{\sum_{j=1}^r T_{.j}^2}{n} - C, \quad df_r = r - 1$$
$$SS_A = \frac{\sum T_{A.}^2}{ar} - C, \quad df_A = k_A - 1$$
$$SS_B = \frac{\sum T_{B.}^2}{br} - C, \quad df_B = k_B - 1$$
$$SS_C = \frac{\sum T_{C.}^2}{cr} - C, \quad df_C = k_C - 1$$
$$SS_{A \times B \times C} = SS_t - SS_A - SS_B - SS_C, \quad df_{A \times B \times C} = df_t - df_A - df_B - df_C$$
$$(2-1) \quad SS_e = SS_T - SS_t - SS_r, \quad df_e = df_T - df_t - df_r$$

(4-7)

注:对于有重复,但重复采用完全随机化设计的正交试验,在变异剖分中则不含有单位组项(SS_r, df_r)。

4. F 检验

(1) 首先用 $F = MS_{A \times B \times C} / MS_e$ 检验交互作用是否显著,若交互作用不显著,则由因素 F 值的大小,确定因素的重要性,从而估计优化搭配。

(2) 若交互作用显著,则将交互看成因素,依重要程度作出判断。

【例 4-5】 为了探讨花生锈病药剂防治效果的好坏,进行了药剂种类(A)、浓度(B)、剂量(C)3 因素试验,各有 3 个水平,选用正交表 $L_9(3^4)$ 安排试验。试验重复 2 次,随机区组设计。正交试验方案及试验结果(产量 kg/小区,小区面积 133.3m^2)见表 4-11,对试验结果进行方差分析。

表 4-11 防治花生锈病药剂种类、浓度、剂量正交试验的方案及结果计算表

试验号	因素				产量(x)/(kg/小区)		T_i	$\bar{x}_i$
	A (1)	B (2)	C (3)	空 (4)	区组 I	区组 II		
1	1(百菌情)	1(高)	1(80)	1	28.0	28.5	56.5	28.25
2	1(百菌情)	2(中)	2(100)	2	35.0	34.8	69.8	34.90
3	1(百菌情)	3(低)	3(120)	3	32.2	32.5	64.7	32.35
4	2(敌锈灵)	1(高)	2(100)	3	33.0	33.2	66.2	33.10
5	2(敌锈灵)	2(中)	3(120)	1	27.4	27.0	54.4	27.20
6	2(敌锈灵)	3(低)	1(80)	2	31.8	32.0	63.8	31.90
7	3(浓尔多)	1(高)	3(120)	2	34.2	34.5	68.7	34.35
8	3(浓尔多)	2(中)	1(80)	3	22.5	23.0	45.5	22.75
9	3(浓尔多)	3(低)	2(100)	1	29.4	30.0	59.4	29.70
T_1	191.0	191.4	165.8	170.3	237.5	275.5	549.0	
T_2	184.4	169.7	195.4	202.3				
T_3	173.6	187.9	187.8	176.4				
$\bar{x}_1$	31.83	31.90	27.63					
$\bar{x}_2$	30.73	28.28	32.57					
$\bar{x}_3$	28.93	31.32	31.30					

由表 4-11 可得:如果没有交互作用,3 因素的重要程度排序是 $C(R=4.94)$ 、 $B(R=3.62)$ 、 $A(R=2.9)$, $A_1B_1C_2$ 搭配最佳,但 $\bar{x}_{B1}$ 与 $\bar{x}_{B3}$ 的差很小(0.58),所以 $A_1B_1C_2$ 与 $A_1B_3C_2$ 再进行试验比较之,找出最优。

对于有重复,且重复采用随机区组设计的正交试验,总变异可以划分为处理间、单位组间和误差变异 3 部分,而处理间变异可进一步划分为 A 因素、B 因素、C 因素与交互 4 部分。此时,总平方和与自由度可分解为

$$SS_T = SS_t + SS_r + SS_e, \quad df_T = df_t + df_r + df_e$$

而

$$SS_t = SS_A + SS_B + SS_C + SS_{A \times B \times C}, \quad df_t = df_A + df_B + df_C + df_{A \times B \times C}$$

于是

$$SS_T = SS_A + SS_B + SS_C + SS_r + SS_{A \times B \times C} + SS_e$$

$$df_T = df_A + df_B + df_C + df_r + df_{A \times B \times C} + df_e$$

式中, SS_r 为区组间离差平方和; $SS_{A \times B \times C}$ 为交互作用平方和; SS_e 为试验误差平方和; SS_t

为处理间平方和; df_r 、 $df_{A \times B \times C}$ 、 df_e 、 df_t 为相应自由度。

注意,对于重复采用完全随机设计的正交试验,在平方和与自由度分解式中无区组间平方和与自由度 SS_r 、 df_r 项。

由表 4-11 可得方差分析表(表 4-12)。

表 4-12 有重复观测值正交试验结果方差分析表

变异来源	SS	df	MS	F	$F_{0.05}$	$F_{0.01}$
A	25.72	2	12.86	214.33**	4.10	7.55
B	45.24	2	22.62	377.00**		
C	78.77	2	39.39	656.50**		
区组	0.22	1	0.22	3.67	4.96	10.01
交互	96.23	2	48.12	802.00**		
试验误差	0.44	8	0.06			
总计	246.62	17				

** 表示因素水平的不同,平均数极显著不都相同;不标记表示因素水平的不同,平均数差异不显著

交互表示的是 A、B、C 的交互作用,交互作用极显著,因此在选择优化搭配上要进一步探讨,因为对试验指标影响最大的是交互作用,其他依次是 C、B、A,要考虑因素的交互作用的重要程度次序,找出最佳,否则找出的搭配不一定是最佳的(对于此问题可根据试验数据求得回归方程,由回归方程求得最大值点,找出最佳搭配,见后响应面分析章节)。

现使用 SPSS 对【例 4-5】进行方差分析。

首先运行 SPSS,SPSS 的数据格式、操作方法和统计结果见图 4-9(a)~图 4-9(e)。

	A	B	C	QZ	x	valvar
1	1	1	1	1	28.0	
2	1	2	2	1	35.0	
3	1	3	3	1	32.2	
4	2	1	2	1	33.0	
5	2	2	3	1	27.4	
6	2	3	1	1	31.8	
7	3	1	3	1	34.2	
8	3	2	1	1	22.5	
9	3	3	2	1	29.4	
10	1	1	1	2	28.5	
11	1	2	2	2	34.8	
12	1	3	3	2	32.5	
13	2	1	2	2	33.2	

图 4-9(a) 数据格式

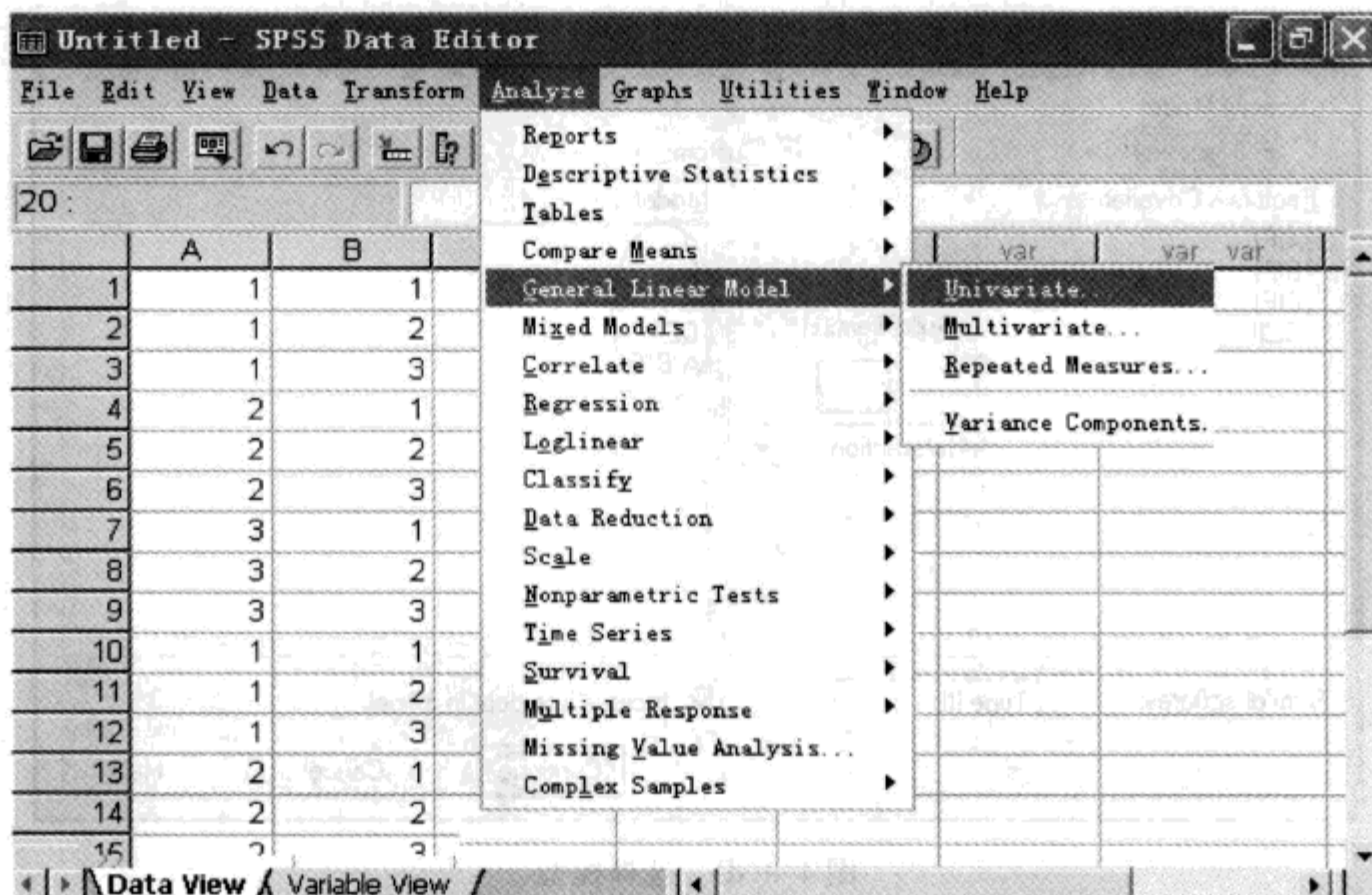


图 4-9(b) 运行方差分析

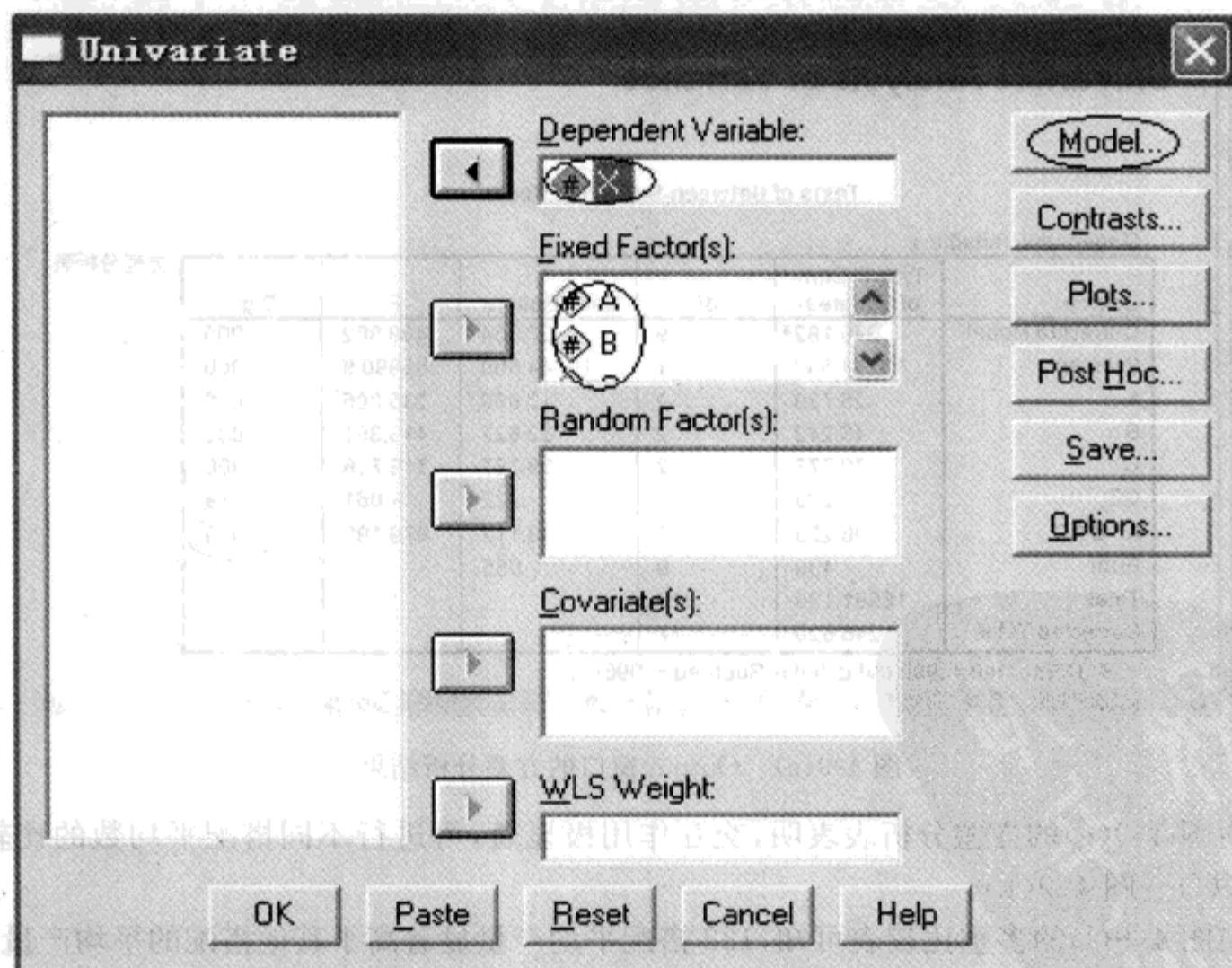


图 4-9(c) 选择处理号和试验数据

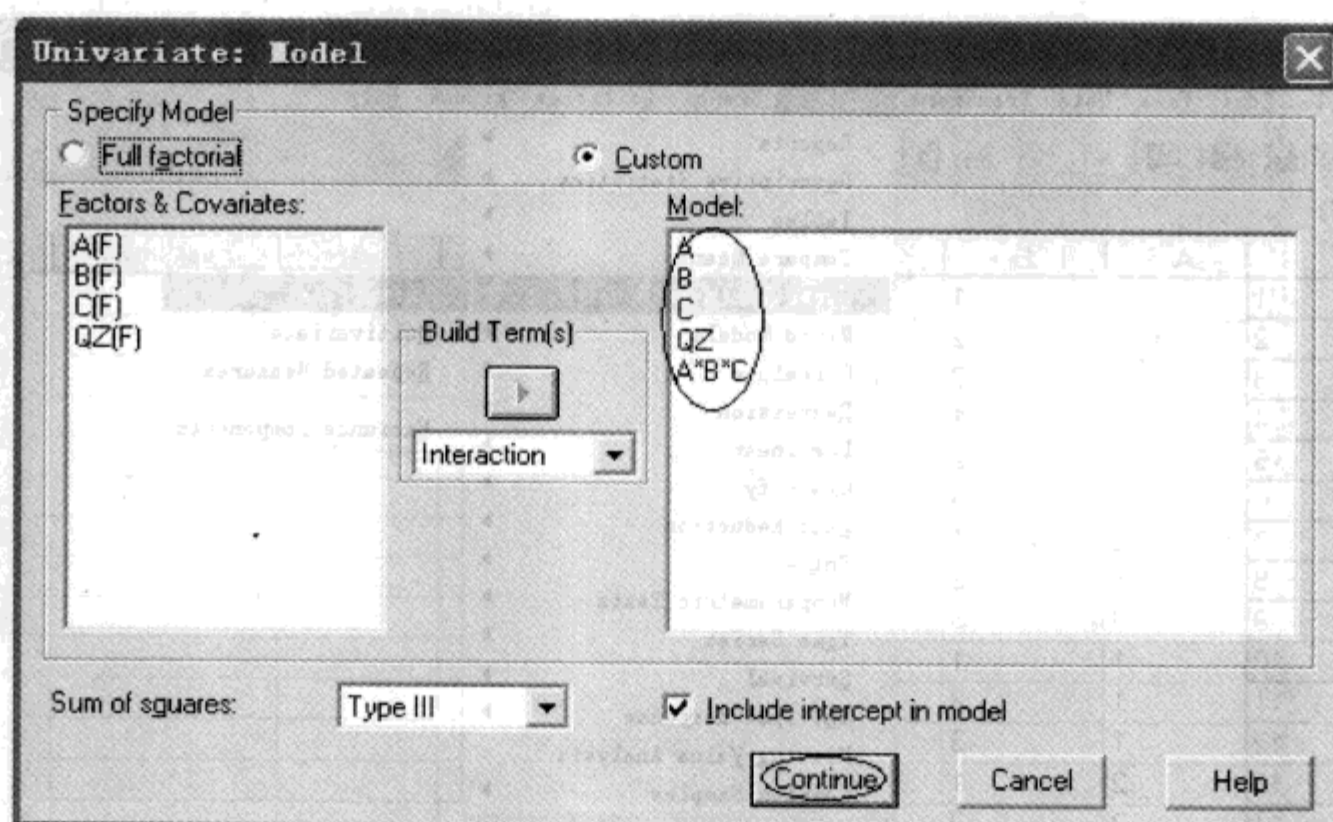


图 4-9(d) 选择模型

Output8 - SPSS Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help

Univariate Analysis of Variance

Tests of Between-Subjects Effects

Dependent Variable: x

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	246.182 ^a	9	27.354	499.862	.000
Intercept	16744.500	1	16744.500	305990.9	.000
A	25.720	2	12.860	235.005	.000
B	45.243	2	22.622	413.391	.000
C	78.773	2	39.387	719.756	.000
QZ	.222	1	.222	4.061	.079
A*B*C	96.223	2	48.112	879.198	.000
Error	.438	8	.055		
Total	16991.120	18			
Corrected Total	246.620	17			

a. R Squared = .998 (Adjusted R Squared = .996)

方差分析表

图 4-9(e) Output 窗口的方差分析结果

由图 4-9(e)的方差分析表表明,交互作用极显著,可进行不同搭配平均数的比较见图 4-9(f)~图 4-9(k)。

由图 4-9(l)的多重比较表可知,122 搭配平均产量显著高于其他搭配的平均产量。

注意,以上只是比较了已做几个搭配试验的平均数。采用回归分析可估计出最佳搭配是 $A_1B_1C_{2.75}$ 。

	ABC	QZ	X	var	var	va	var
1	111	1	28.0				
2	122	1	35.0				
3	133	1	32.2				
4	212	1	33.0				
5	223	1	27.4				
6	231	1	31.8				
7	313	1	34.2				
8	321	1	22.5				
9	332	1	29.4				
10	111	2	28.5				
11	122	2	34.8				
12	133	2	32.5				

图 4-9(f) 数据格式

	ABC	QZ	X	var	var	va	var
1	111	1	28.0				
2	122	1	35.0				
3	133	1	32.2				
4	212	1	33.0				
5	223	1	27.4				
6	231	1	31.8				
7	313	1	34.2				
8	321	1	22.5				
9	332	1	29.4				
10	111	2	28.5				
11	122	2	34.8				
12	133	2	32.5				

图 4-9(g) 运行方差分析

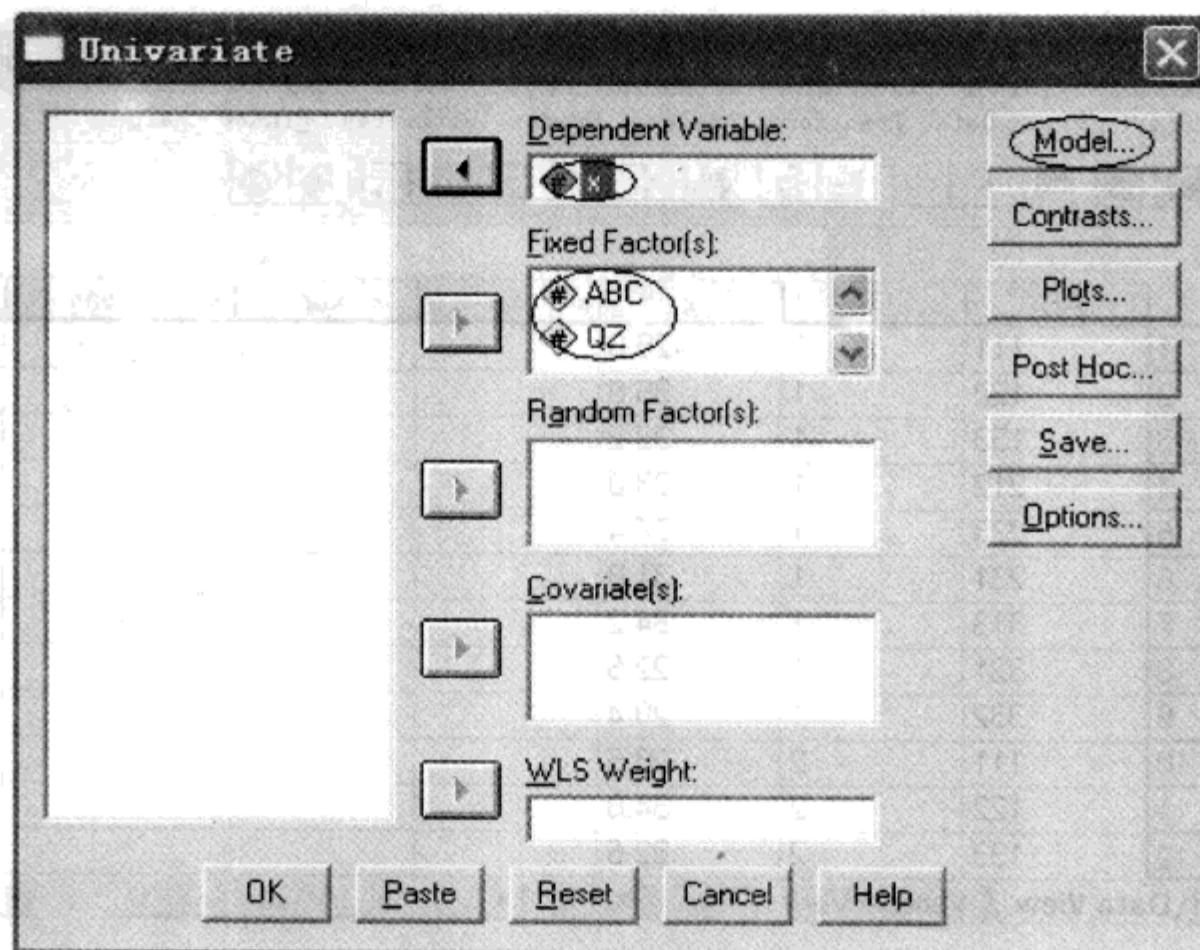


图 4-9(h) 选择因变量和试验数据

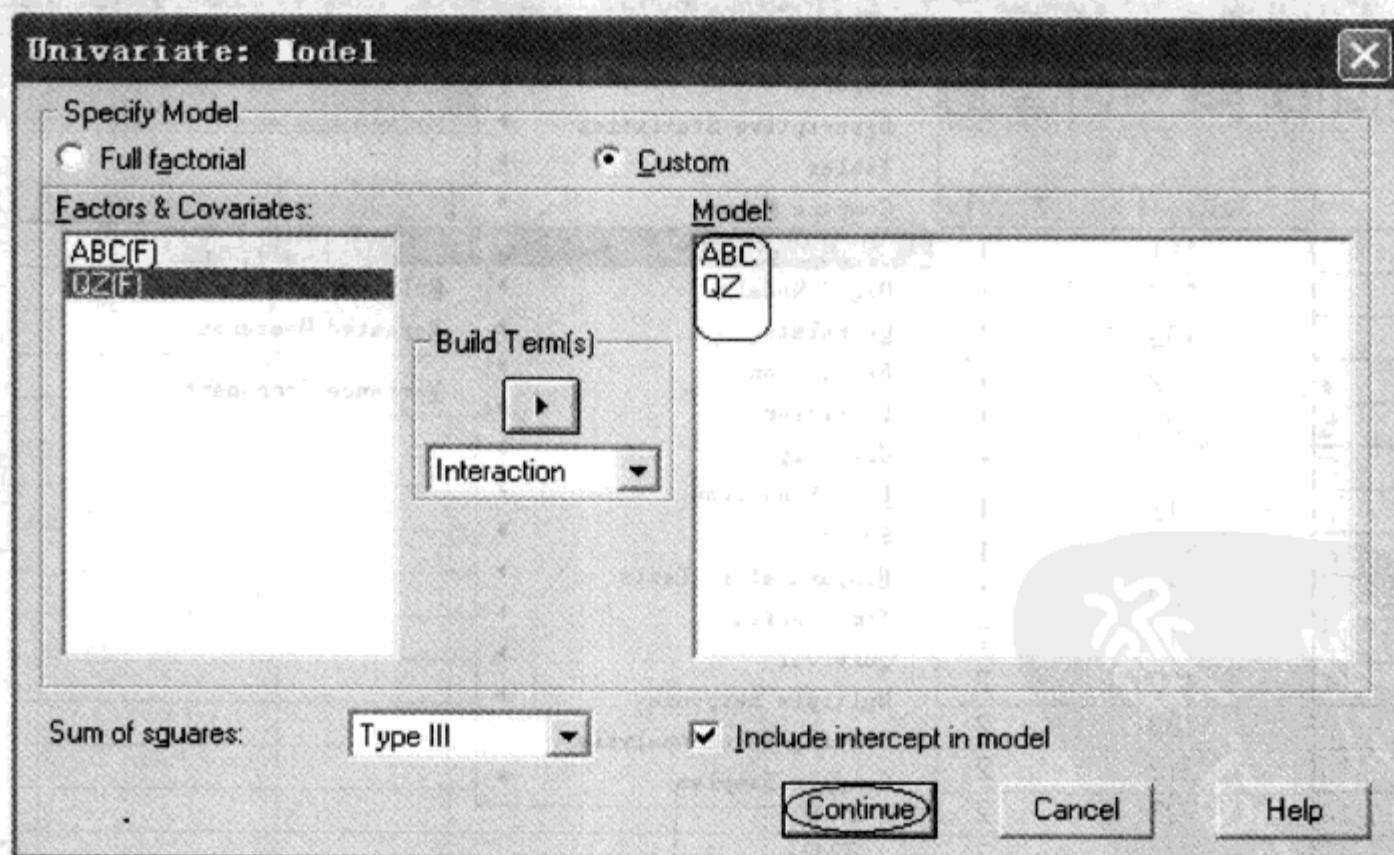


图 4-9(i) 选择模型

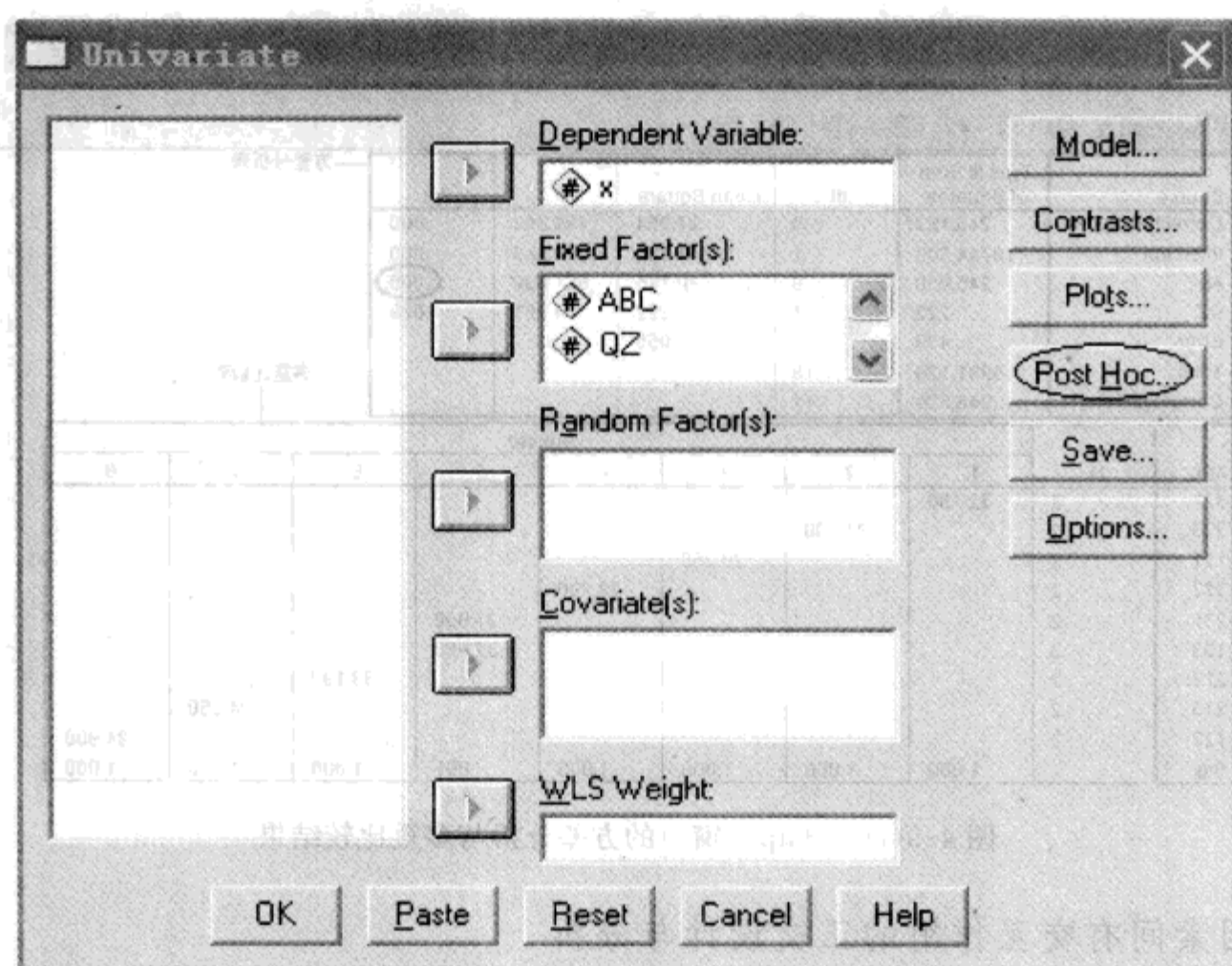


图 4-9(j) 选择多重比较

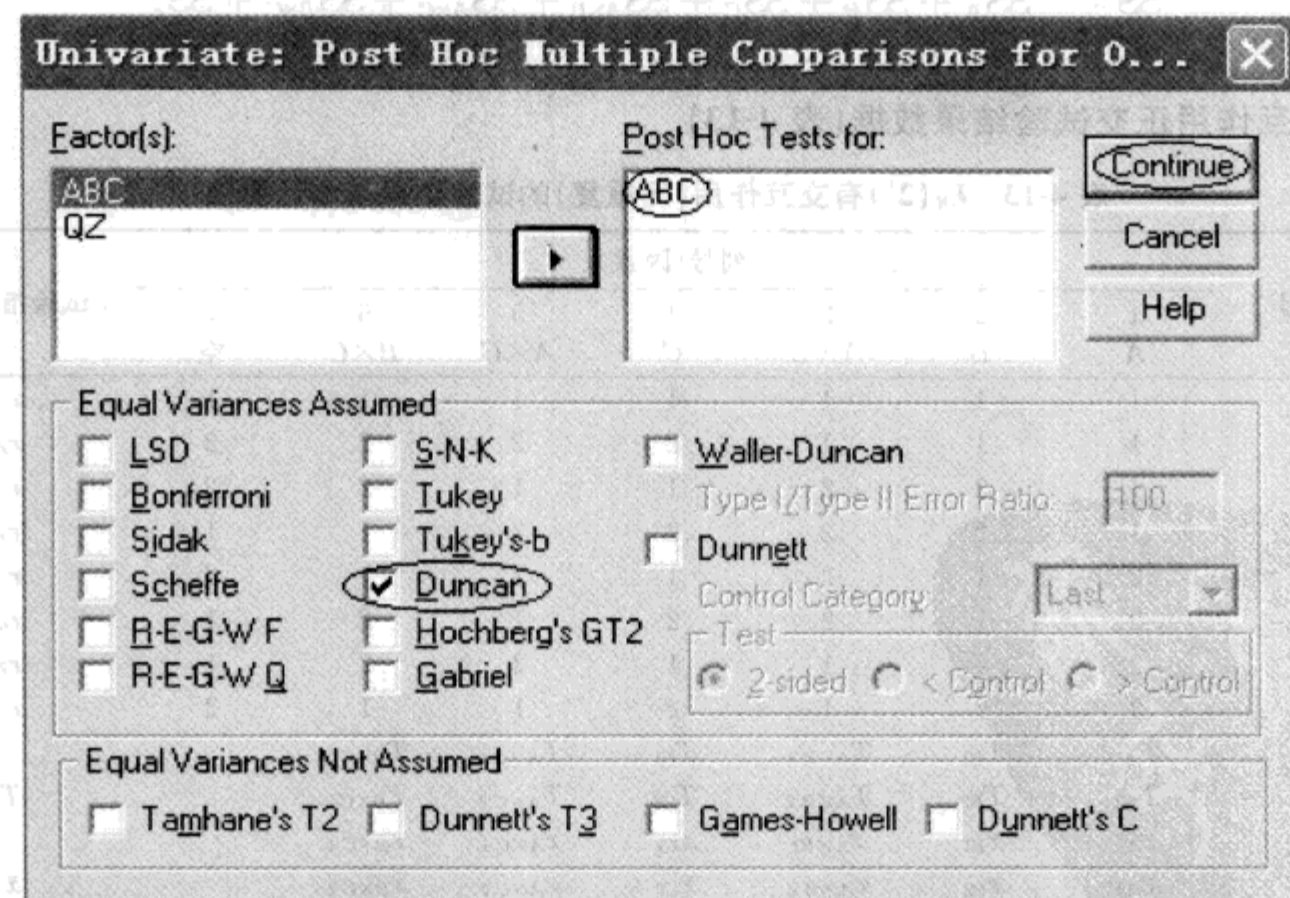


图 4-9(k) 选择多重比较变量及方法

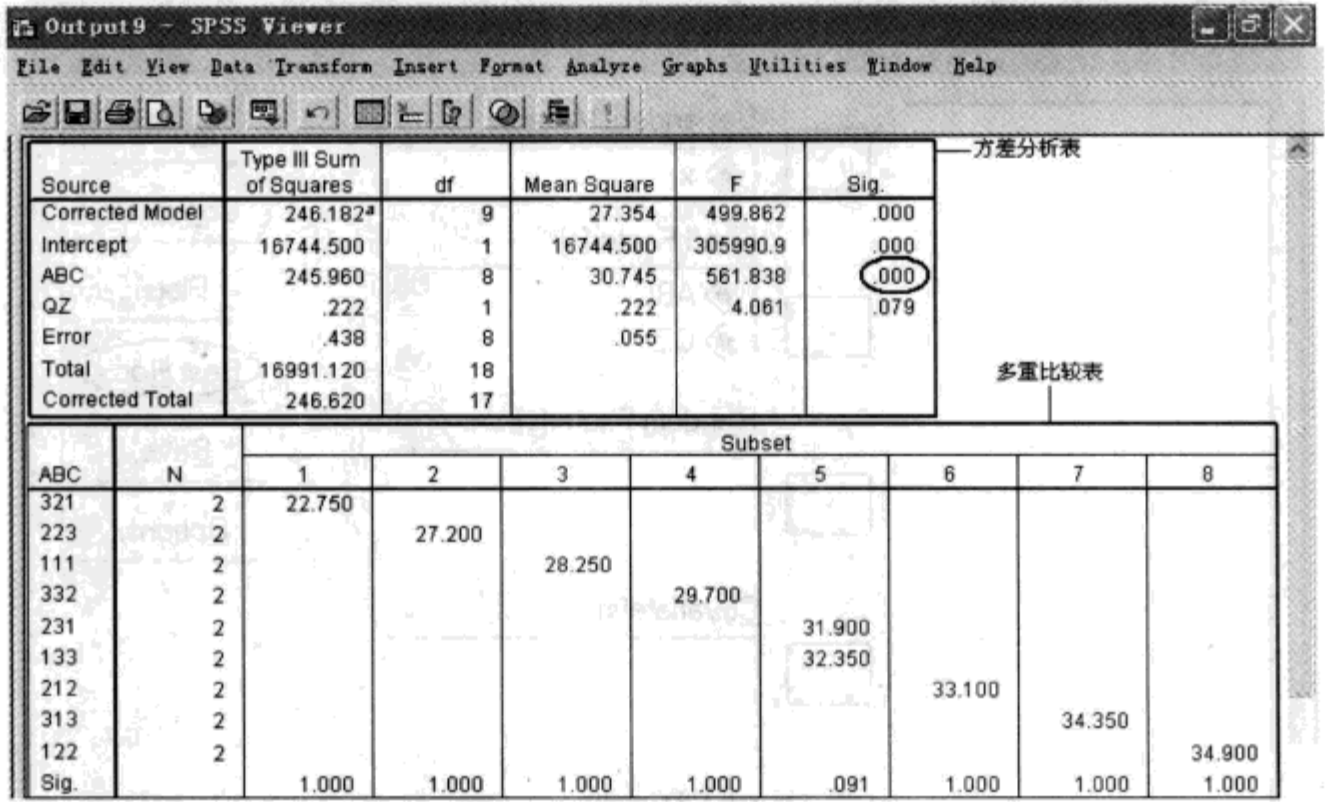


图 4-9(l) Output 窗口的方差分析与多重比较结果

(三) 因素间有交互作用的正交设计与分析

以 3 因素为例说明。

1. 平方和与自由度划分式

$$\begin{aligned} SS_T &= SS_A + SS_B + SS_C + SS_{A \times B} + SS_{A \times C} + SS_{B \times C} + SS_e \\ df_T &= df_A + df_B + df_C + df_{A \times B} + df_{A \times C} + df_{B \times C} + df_e \end{aligned} \quad (4-8)$$

2. 有交互作用正交试验结果数据(表 4-13)

表 4-13 $L_8(2^7)$ 有交互作用(无重复)的试验结果基本计算表

试验号	列号(因素)							试验指标 x_i
	1	2	3	4	5	6	7	
	A	B	A×B	C	A×C	B×C	空	
1	1	1	1	1	1	1	1	x_1
2	1	1	2	2	2	2	2	x_2
3	1	2	2	1	1	2	2	x_3
4	1	2	1	2	2	1	1	x_4
5	2	1	2	1	2	1	2	x_5
6	2	1	1	2	1	2	1	x_6
7	2	2	1	1	2	2	1	x_7
8	2	2	2	2	1	1	2	x_8
T_1	T_{A1}	T_{B1}	$T_{A \times B 1}$	T_{C1}	$T_{A \times C 1}$	$T_{B \times C 1}$		
T_2	T_{A2}	T_{B2}	$T_{A \times B 2}$	T_{C2}	$T_{A \times C 2}$	$T_{B \times C 2}$		T
$\bar{x}_1$	$\bar{x}_{A1}$	$\bar{x}_{B1}$	$\bar{x}_{A \times B 1}$	$\bar{x}_{C1}$	$\bar{x}_{A \times C 1}$	$\bar{x}_{B \times C 1}$		
$\bar{x}_2$	$\bar{x}_{A2}$	$\bar{x}_{B2}$	$\bar{x}_{A \times B 2}$	$\bar{x}_{C2}$	$\bar{x}_{A \times C 2}$	$\bar{x}_{B \times C 2}$		$\bar{x}$

下面通过实例说明。

【例 4-6】 某一抗生素发酵培养基由 A、B、C 3 种成分组成,各有 2 个水平,除考察

A、B、C 3 个因素的主效外,还考察交互作用(A 与 C 交互作用不显著)。试安排一个正交试验方案并对试验结果进行分析。

1. 选用正交表,表头设计

由于本试验有 3 个两水平的因素和 3 个交互作用需要考察,因此可选用正交表 $L_8(2^7)$ 来安排试验方案。

正交表 $L_8(2^7)$ 中有基本列和交互列之分,基本列就是各因素所占的列,交互列为两因素交互作用所占的列。利用 $L_8(2^7)$ 二列间交互作用列表(表 4-14)来安排各因素和交互作用。

表 4-14 $L_8(2^7)$ 二列间交互作用列表

列 号	1	2	3	4	5	6	7
1	(1)	3	2	5	4	7	6
2		(2)	1	6	7	4	5
3			(3)	7	6	5	4
4				(4)	1	2	3
5					(5)	3	2
6						(6)	1

如果将 A 因素放在第 1 列,B 因素放在第 2 列,查表第 1 列(A 因素所在列)与第 2 列(B 因素所在列)的交互作用列是第 3 列,于是将 A 与 B 的交互作用 $A \times B$ 放在第 3 列。这样第 3 列不能再安排其他因素,然后将 C 放在第 4 列。查表第 2 列(B 因素所在列)与第 4 列(C 因素所在列)的交互作用列是第 6 列,于是将 B 与 C 的交互作用 $B \times C$ 放在第 6 列,A 与 C 的交互作用 $A \times C$ 放在第 5 列,余下列为空白,如此可得表头设计,见表 4-15。

表 4-15 表头设计

列号	1	2	3	4	5	6	7
因素	A	B	$A \times B$	C	$A \times C$	$B \times C$	空

2. 列出试验方案

根据表头设计,得出试验方案列于表 4-16。

表 4-16 试验方案表

试验号	因素			试验数据 x
	1(A)	2(B)	3(C)	
1	1(A ₁)	1(B ₁)	1(C ₁)	
2	1(A ₁)	1(B ₁)	2(C ₂)	
3	1(A ₁)	2(B ₂)	1(C ₁)	
4	1(A ₁)	2(B ₂)	2(C ₂)	
5	2(A ₂)	1(B ₁)	1(C ₁)	
6	2(A ₂)	1(B ₁)	2(C ₂)	
7	2(A ₂)	2(B ₂)	1(C ₁)	
8	2(A ₂)	2(B ₂)	2(C ₂)	

3. 结果分析

按表 4-16 所列的试验方案进行试验,其结果见表 4-17。

表 4-17 有交互作用正交试验结果计算表

试验号	因素					试验结果*
	A	B	A×B	C	B×C	
1	1	1	1	1	1	55(x_1)
2	1	1	1	2	2	38(x_2)
3	1	2	2	1	2	97(x_3)
4	1	2	2	2	1	89(x_4)
5	2	1	2	1	1	122(x_5)
6	2	1	2	2	2	124(x_6)
7	2	2	1	1	2	79(x_7)
8	2	2	1	2	1	61(x_8)
T_1	279	339	233	353	327	665(T)
T_2	386	326	432	312	338	
$\bar{x}_1$	69.75	84.75	58.25	88.25	81.75	
$\bar{x}_2$	96.50	81.50	108	78.00	84.50	

* 试验结果以对照为 100 计

由表 4-17 可看出,因素主要程度的排序是: $A \times B(R=49.75)$ 、 $A(R=26.75)$ 、 $C(R=10.25)$ 、 $B(R=3.25)$ 、 $B \times C(R=2.75)$,如果是求最大值, $\bar{x}_{2(A \times B)}$ 大,所以 A 取 2、B 取 1 (21 的平均数大于 12 的平均数);A 取 2;C 取 1;最佳搭配是 $A_2B_1C_1$,最后可将 $A_2B_1C_1$ 、 $A_2B_1C_2$ 各进行一次验证性试验,确定最佳。

下面进行方差分析(表 4-18)。

表 4-18 方差分析表

变异来源	SS	df	MS	F	$F_{0.05}$	$F_{0.01}$
A	1431.1250	1	1431.1250	24.84*	18.51	98.49
B	21.1250	1	21.1250	<1		
C	210.1250	1	210.1250	3.65		
$A \times B$	4950.1250	1	4950.1250	85.90*		
$B \times C$	15.1250	1	12.1250	<1		
误差	115.2500	2	57.6250			
总变异	6742.8750	7				

* 表示因素水平的不同,平均数显著不都相同;不标记表示因素水平的不同,平均数差异不显著

此例为单个观测值正交试验,总变异划分为 A 因素、B 因素、C 因素、 $A \times B$ 、 $B \times C$ 与误差变异 5 部分,平方和与自由度的分解式为

$$\begin{aligned} SS_T &= SS_A + SS_B + SS_C + SS_{A \times B} + SS_{B \times C} + SS_e \\ df_T &= df_A + df_B + df_C + df_{A \times B} + df_{B \times C} + df_e \end{aligned} \quad (4-9)$$

(1) 计算各项平方和与自由度。

矫正数 $C = T^2/n = 665^2/8 = 55\,278.1250$

总平方和 $SS_T = \sum x^2 - C = 55^2 + 38^2 + \cdots + 61^2 - 55\,278.1250 = 6742.8750$

A 因素平方和 $SS_A = \sum T_A^2/a - C = (279^2 + 386^2)/4 - 55\,278.1250 = 1431.1250$

B 因素平方和 $SS_B = \sum T_B^2/b - C = (339^2 + 326^2)/4 - 55\,278.1250 = 21.1250$

C 因素平方和
$$SS_C = \sum T_{C/c}^2 / c - C = (353^2 + 312^2) / 4 - 55\,278.1250 = 210.1250$$

A × B 平方和
$$SS_{A \times B} = \sum T_{A \times B}^2 / 4 - C = (233^2 + 432^2) / 4 - 55\,278.1250 = 4950.1250$$

B × C 平方和
$$SS_{B \times C} = \sum T_{B \times C}^2 / 4 - C = (327^2 + 338^2) / 4 - 55\,278.1250 = 15.1250$$

误差平方和
$$SS_e = SS_T - SS_A - SS_B - SS_{A \times B} - SS_{B \times C} = 6742.8750 - 1431.1250 - 21.1250 - 210.1250 - 4950.1250 - 15.1250 = 115.2500$$

总自由度
$$df_T = n - 1 = 8 - 1 = 7$$

各因素自由度
$$df_A = df_B = df_C = 2 - 1 = 1$$

交互作用自由度
$$df_{A \times B} = df_{B \times C} = (2 - 1) \times (2 - 1) = 1$$

误差自由度
$$df_e = df_T - df_A - df_B - df_C - df_{A \times B} - df_{B \times C} = 7 - 1 - 1 - 1 - 1 - 1 = 2$$

(2) 列出方差分析表(表 4-18),进行 F 检验。

F 检验结果表明:A 因素和交互作用 A×B 显著,B、C 因素及 B×C 交互作用不显著。因交互作用 A×B 显著,对 A 与 B 的水平组合进行多重比较,以选出 A 与 B 的最优水平组合。

(3) A 与 B 各水平组合的多重比较。先计算出 A 与 B 各水平组合的平均数。

A₁B₁ 水平组合的平均数 $\bar{x}_{11} = (55 + 38) / 2 = 46.50$

A₁B₂ 水平组合的平均数 $\bar{x}_{12} = (97 + 89) / 2 = 93.00$

A₂B₁ 水平组合的平均数 $\bar{x}_{21} = (122 + 124) / 2 = 123.00$

A₂B₂ 水平组合的平均数 $\bar{x}_{22} = (79 + 61) / 2 = 70.00$

列出 A、B 因素各水平组合平均数多重比较表,见表 4-19。

表 4-19 A、B 因素各水平组合平均数的多重比较表(SSR 法)

水平组合	平均数 $\bar{x}_{ij}$	$\bar{x}_{ij} - 46.5$	$\bar{x}_{ij} - 70$	$\bar{x}_{ij} - 93$
A ₂ B ₁	123.00	76.5**(**)	53*(**)	30(**)
A ₁ B ₂	93.00	46.5*(**)	23(*)	
A ₂ B ₂	70.00	23.5(*)		
A ₁ B ₁	46.50			

* 表示因素水平的不同,平均数显著不都相同; ** 表示因素水平的不同,平均数极显著不都相同;不标记表示因素水平的不同,平均数差异不显著
括号内为合并误差的多重比较表结果

因为, $S_x = \sqrt{MS_e / 2} = \sqrt{57.625 / 2} = 5.368$,由 $df_e = 2$ 与 $k = 2, 3, 4$ 查临界 SSR 值,并计算出 LSR 值列于表 4-20。

表 4-20 SSR 值与 LSR 值表

df_e	k	$SSR_{0.05}$	$SSR_{0.01}$	$LSR_{0.05}$	$LSR_{0.01}$
2	2	6.09	14.0	32.70	75.18
	3	6.09	14.0	32.70	75.18
	4	6.09	14.0	32.70	75.18

多重比较结果表明,A₂B₁ 极显著优于 A₁B₁,显著优于 A₂B₂;A₁B₂ 显著优于 A₁B₁,

其余差异不显著。最优水平组合为 A_2B_1 。

从以上分析可知, A 因素取 A_2 , B 因素取 B_1 , 若 C 因素取 C_1 , 则本次试验结果的最优水平组合为 $A_2B_1C_1$ 。若 C 因素取 C_2 , 则本次试验结果的最优水平组合为 $A_2B_1C_2$ 。

注意, 此例因 $df_e=2$, F 检验与多重比较的灵敏度低。为了提高检验的灵敏度, 可将 $F<1$ 的 SS_B 、 df_B , $SS_{B \times C}$ 、 $df_{B \times C}$ 合并到 SS_e 、 df_e 中, 得合并的误差方差分析表(表 4-21), 再用合并误差均方进行 F 检验与多重比较(表 4-22)。

表 4-21 合并的误差方差分析表

变异来源	SS	df	MS	F	$F_{0.05(1,4)}$	$F_{0.01(1,4)}$
A	1431.1250	1	1431.1250	37.82**	7.71	21.20
C	210.1250	1	210.1250	5.55		
$A \times B$	4950.1250	1	4950.1250	130.80**		
合并误差	151.3750	4	37.8438			
总变异	6742.8750	7				

**表示因素水平的不同, 平均数极显著不都相同; 不标记表示因素水平的不同, 平均数差异不显著

表 4-22 合并误差的 SSR 值与 LSR 值表

df_e	k	$SSR_{0.05}$	$SSR_{0.01}$	$LSR_{0.05}$	$LSR_{0.01}$
	2	3.93	6.51	17.01	28.32
4	3	4.01	6.80	17.44	29.58
	4	4.02	6.90	17.49	30.01

此时

$$S_x = \sqrt{MS_e/2} = \sqrt{37.8438/2} = 4.3499$$

合并误差的多重比较表结果(列于表 4-19 的括号内)表明, A_2B_1 平均数极显著优于 A_1B_1 、 A_2B_2 、 A_1B_2 平均数, A_2B_1 为最优水平组合。

使用 SPSS 对【例 4-6】进行方差分析。

首先运行 SPSS, SPSS 的数据格式、操作方法和统计结果见图 4-10(a)~图 4-10(e)。

	A	B	AB	C	AC	BC	X	Var
1	1	1	1	1	1	1	55	
2	1	1	1	2	2	2	38	
3	1	2	2	1	1	2	97	
4	1	2	2	2	2	1	89	
5	2	1	2	1	2	1	122	
6	2	1	2	2	1	2	124	
7	2	2	1	1	2	2	79	
8	2	2	1	2	1	1	61	

图 4-10(a) 数据格式

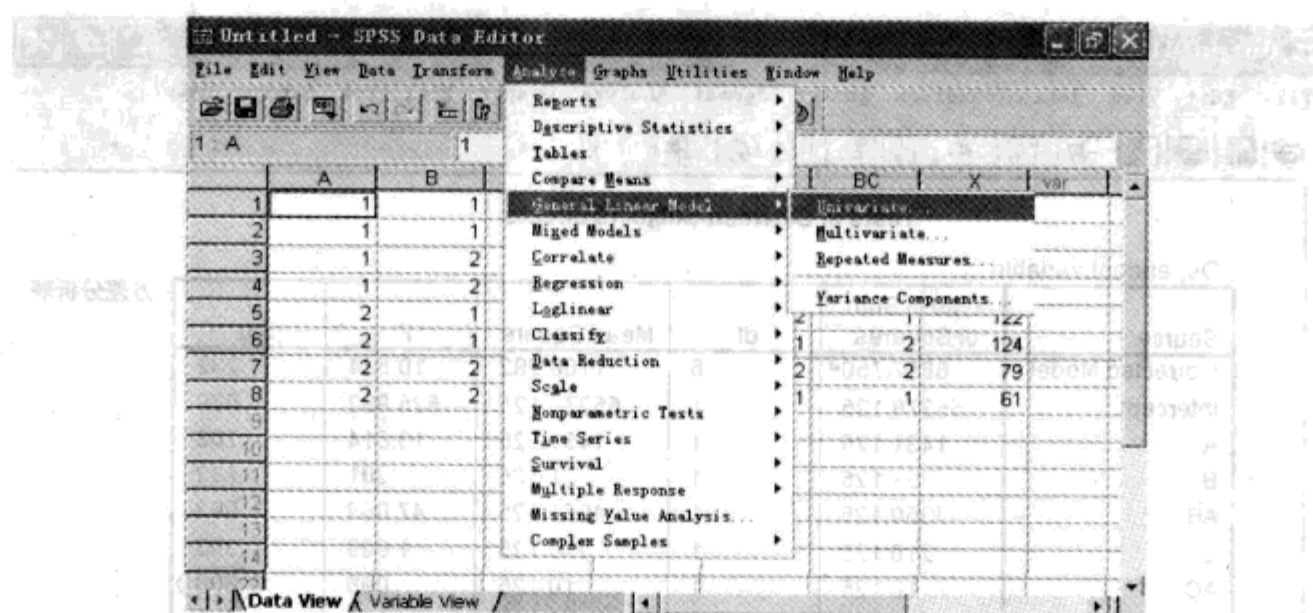


图 4-10(b) 运行方差分析

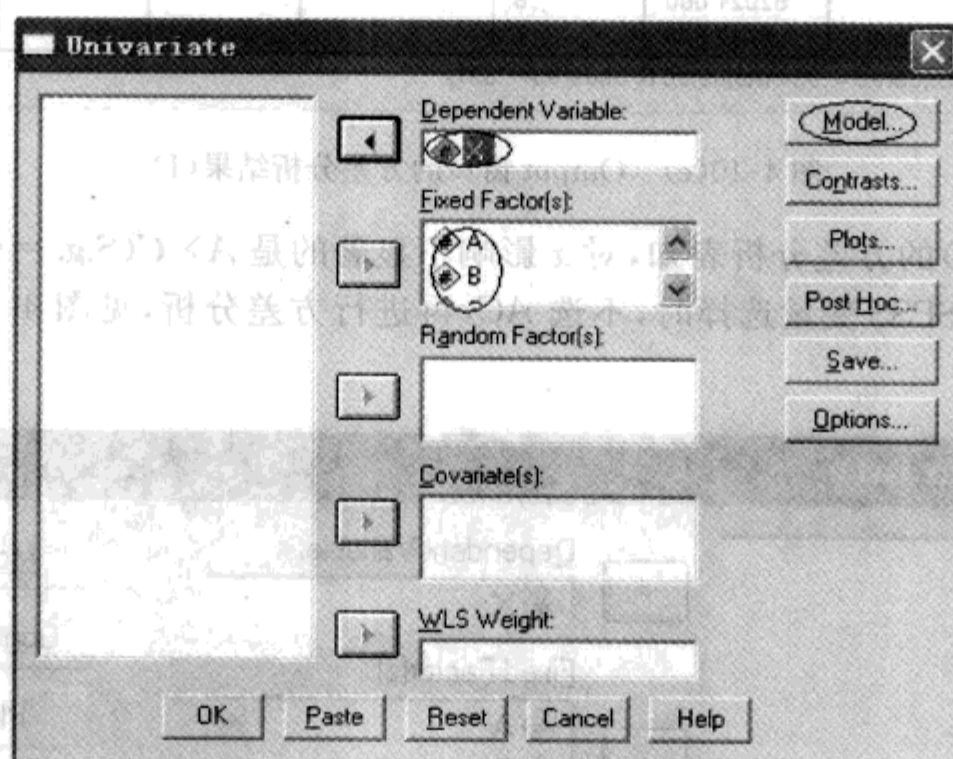


图 4-10(c) 选择处理号和试验数据

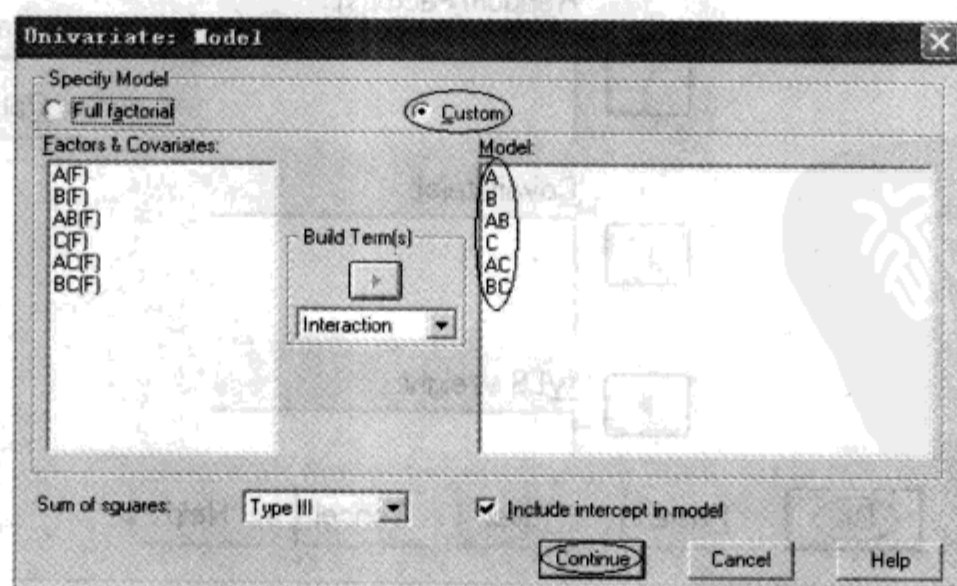


图 4-10(d) 选择模型

Output10 - SPSS Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help

Tests of Between-Subjects Effects

Dependent Variable: X

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6637.750 ^a	6	1106.292	10.524	.232
Intercept	55278.125	1	55278.125	525.832	.028
A	1431.125	1	1431.125	13.614	.168
B	21.125	1	21.125	.201	.732
AB	4950.125	1	4950.125	47.088	.092
C	210.125	1	210.125	1.999	.392
AC	10.125	1	10.125	.096	.808
BC	15.125	1	15.125	.144	.769
Error	105.125	1	105.125		
Total	62021.000	8			
Corrected Total	6742.875	7			

a. R Squared = .984 (Adjusted R Squared = .891)

方差分析表

图 4-10(e) Output 窗口的方差分析结果(1)

由图 4-10(e)的方差分析表知,对 x 影响不显著的是 $A \times C$ (Sig. = 0.808 最大),将 $A \times C$ 去掉,即在 SPSS 变量选择时,不选 AC,再进行方差分析,见图 4-10(f)~图 4-10(g)。

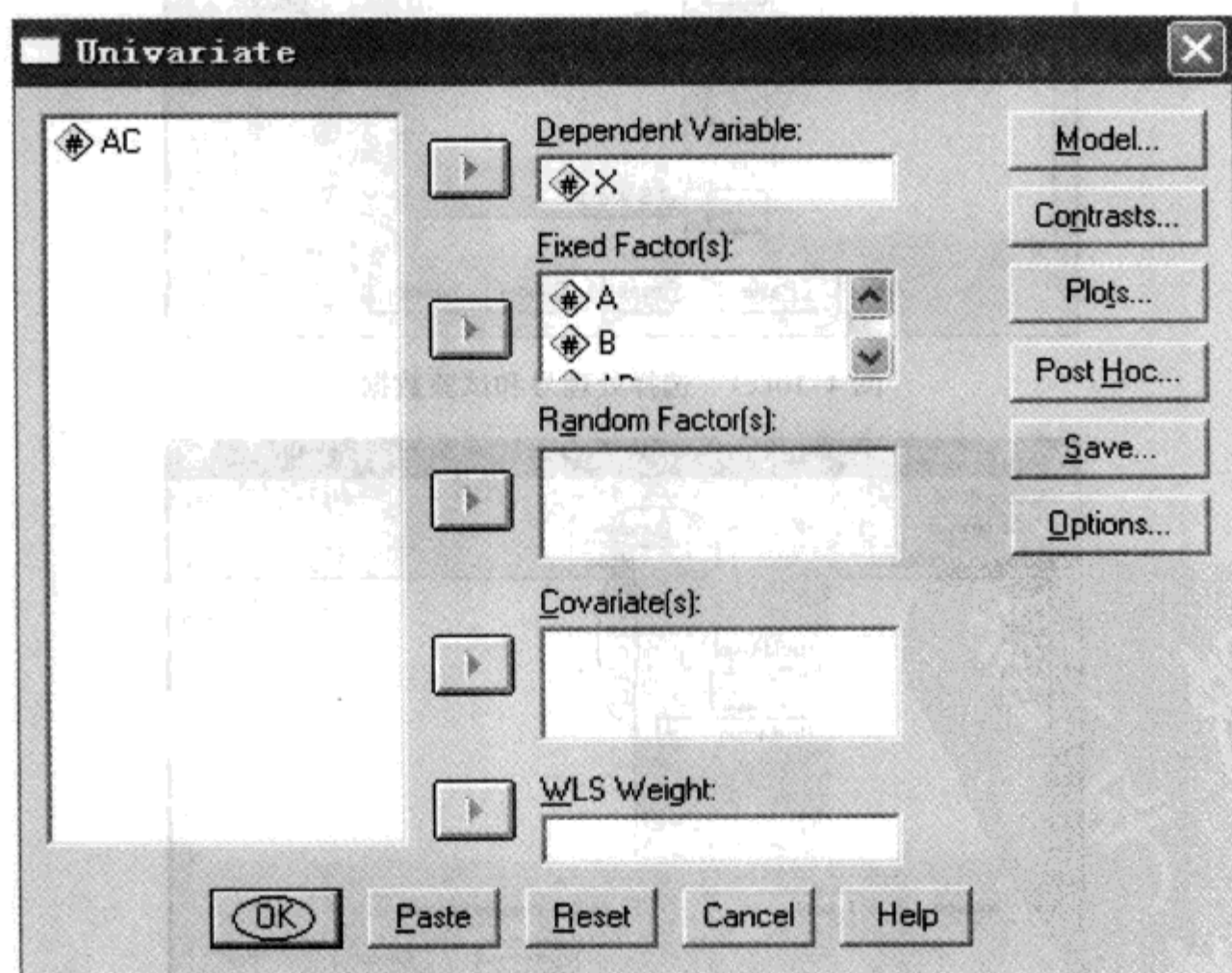


图 4-10(f) 选择变量去掉 AC

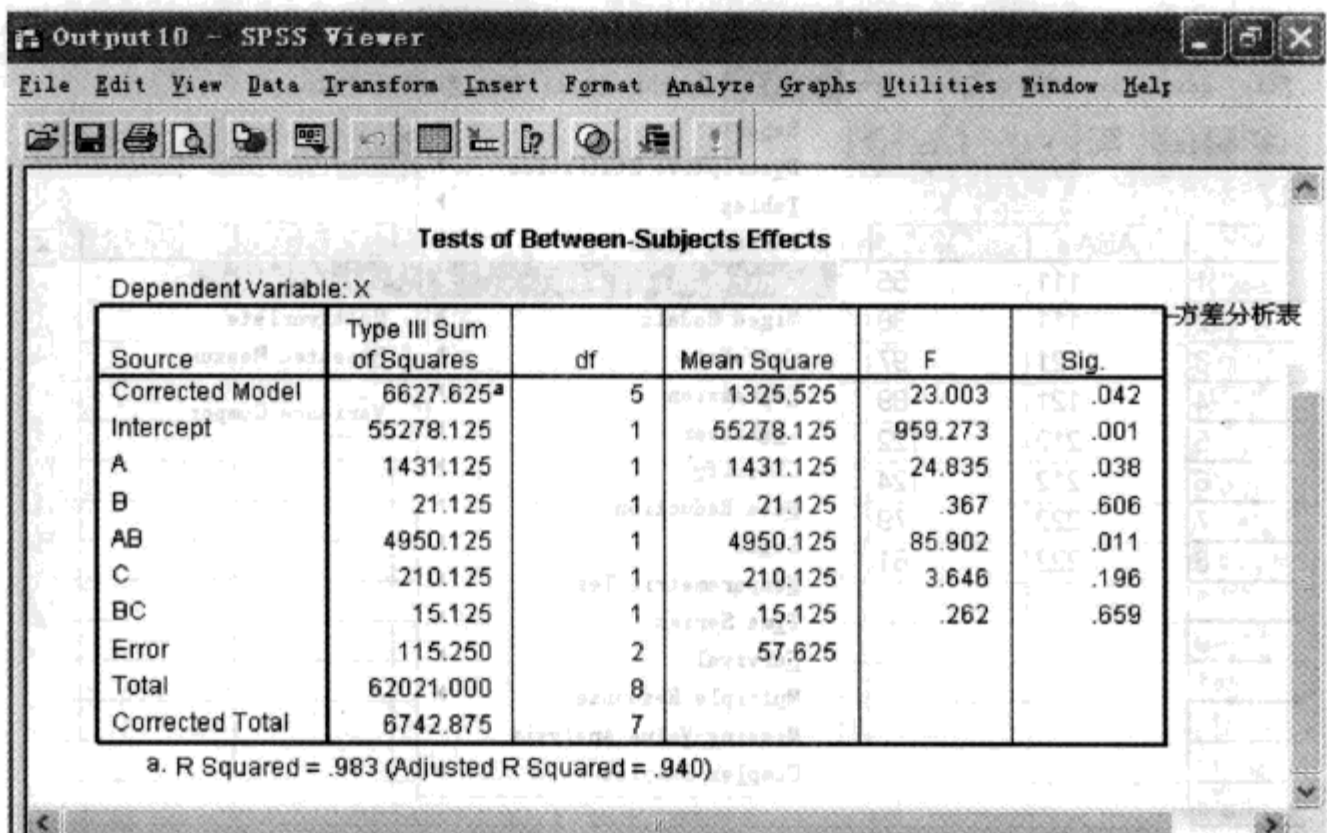


图 4-10(g) Output 窗口的方差分析结果(2)

由图 4-10(g)的方差分析表知,对 x 影响不显著的是 $B \times C$ (Sig. = 0.659 最大),将 $B \times C$ 去掉,直至所选变量对 x 影响显著为止,最后方差分析表见图 4-10(h)。

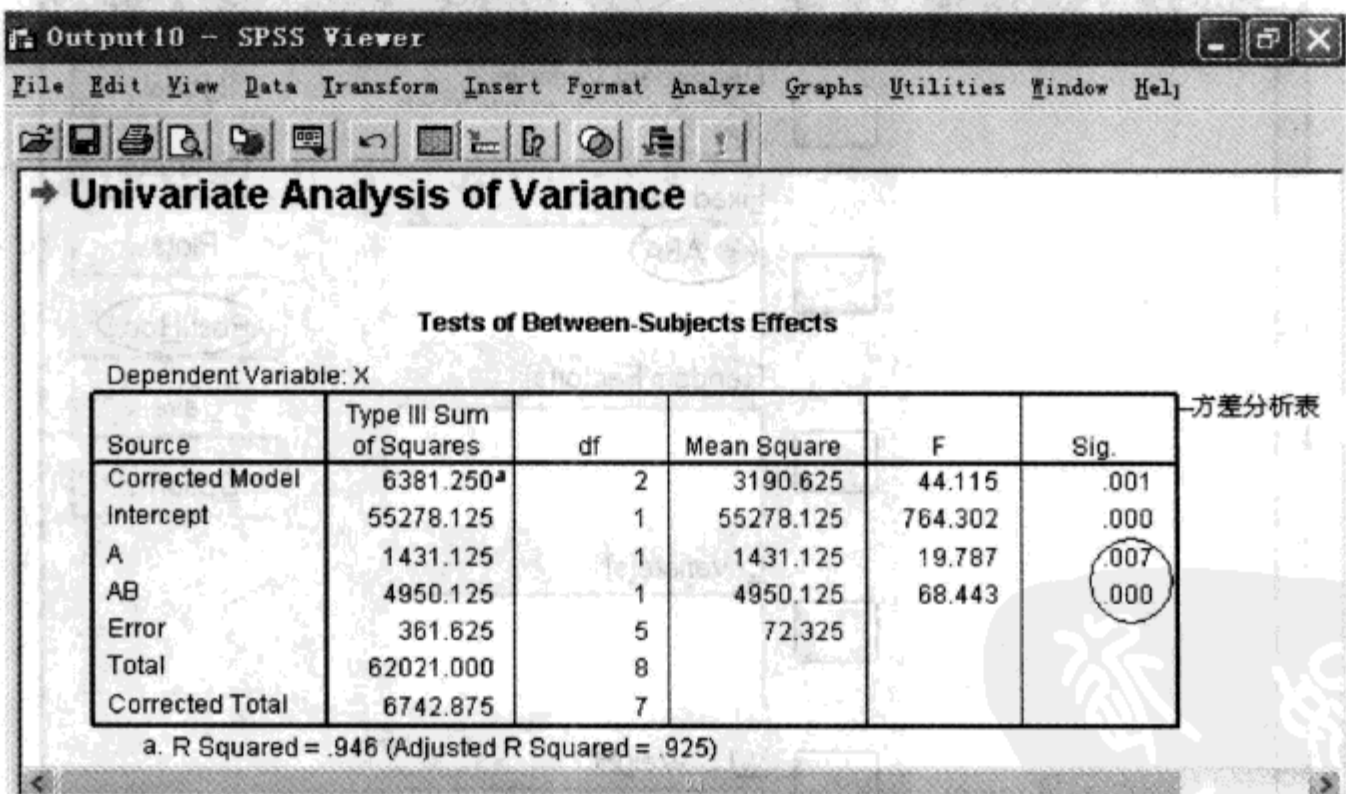


图 4-10(h) Output 窗口的方差分析结果(3)

图 4-10(h)的方差分析表说明 $A \times B$ 与 A 对试验指标有极显著影响,进行不同搭配平均数的比较, $ABA=121$ 表示 AB 搭配各取 1、2, A 取 1,运算方法及结果见图 4-10(i)~图 4-10(l)。

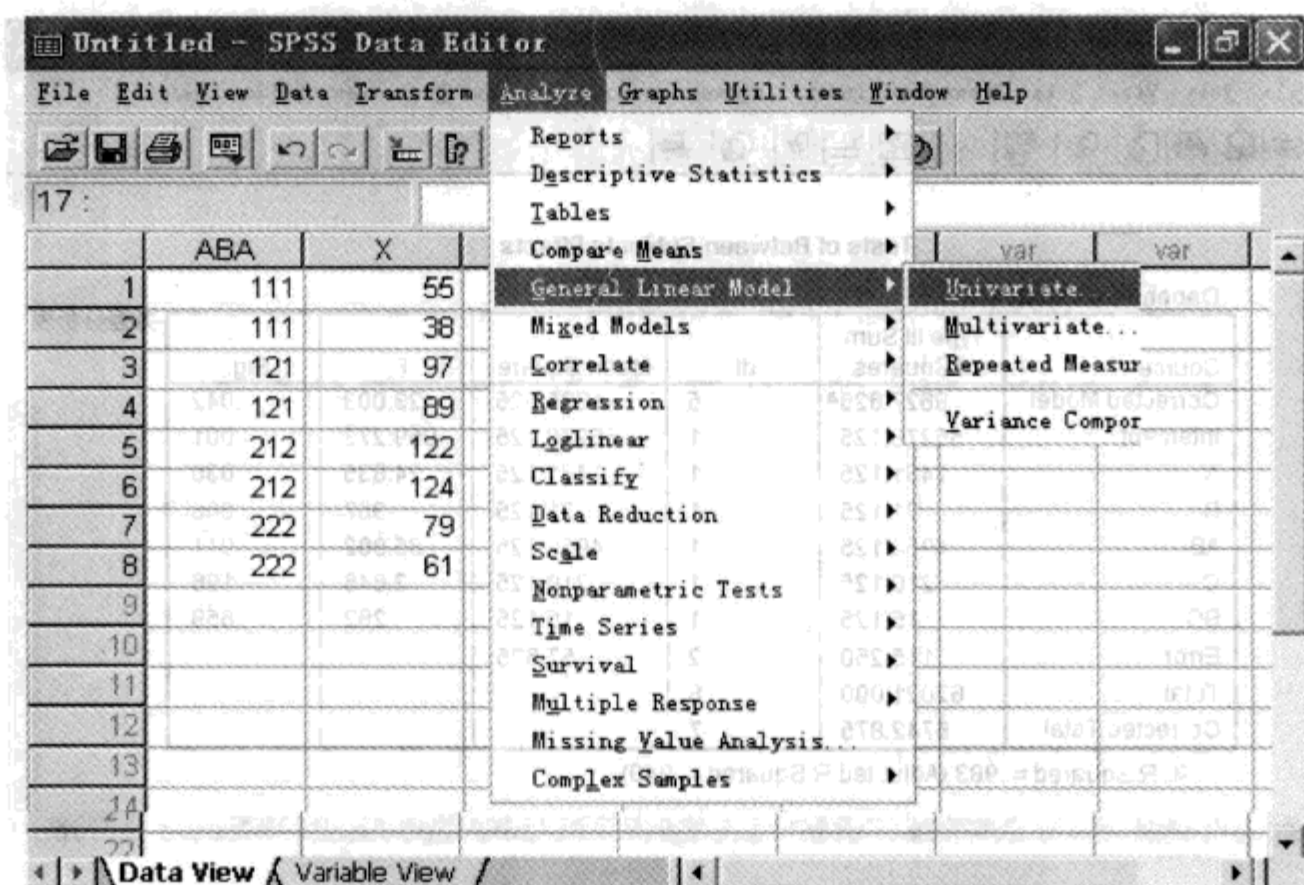


图 4-10(i) 数据格式和运行方差分析

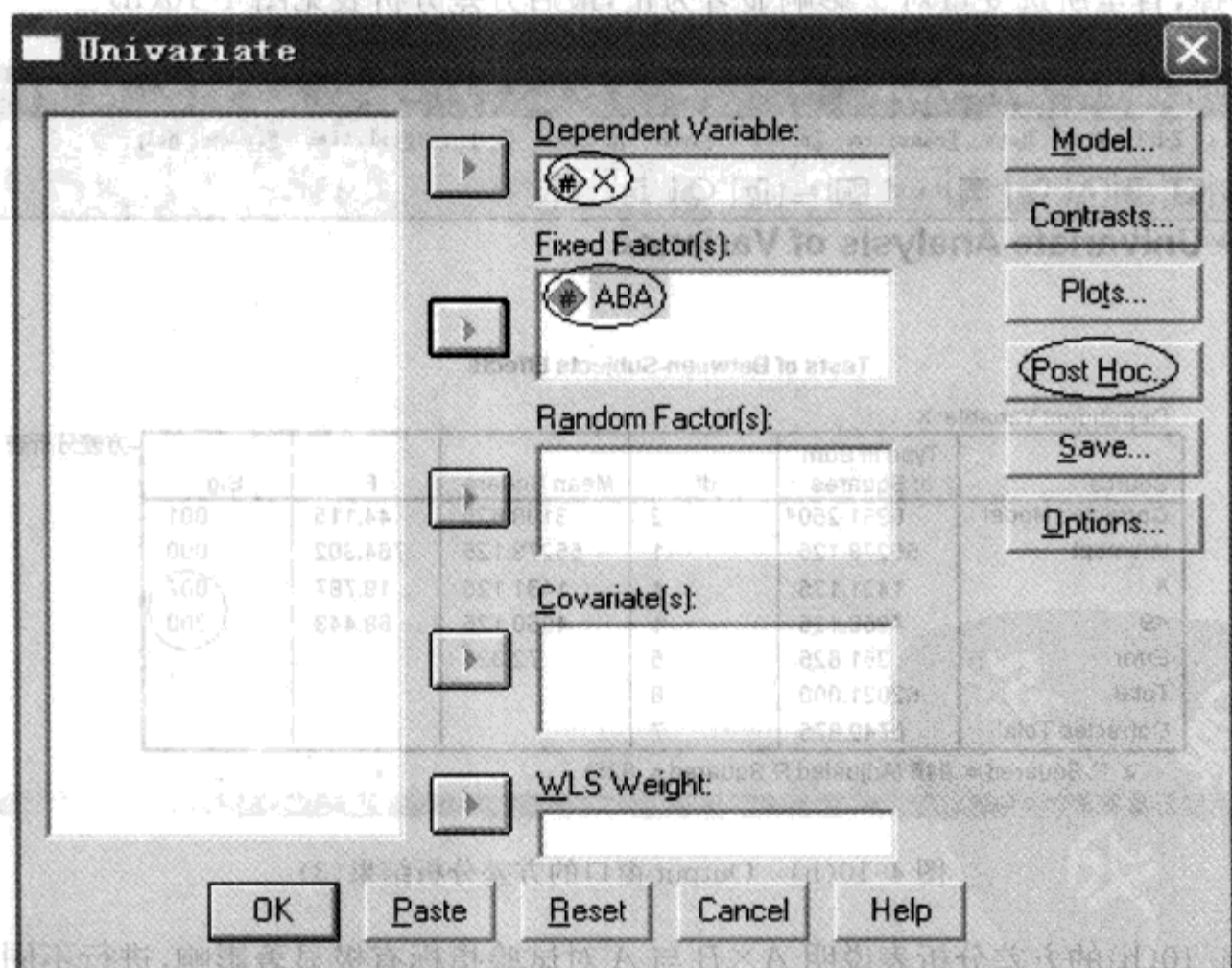


图 4-10(j) 选择处理号和试验数据

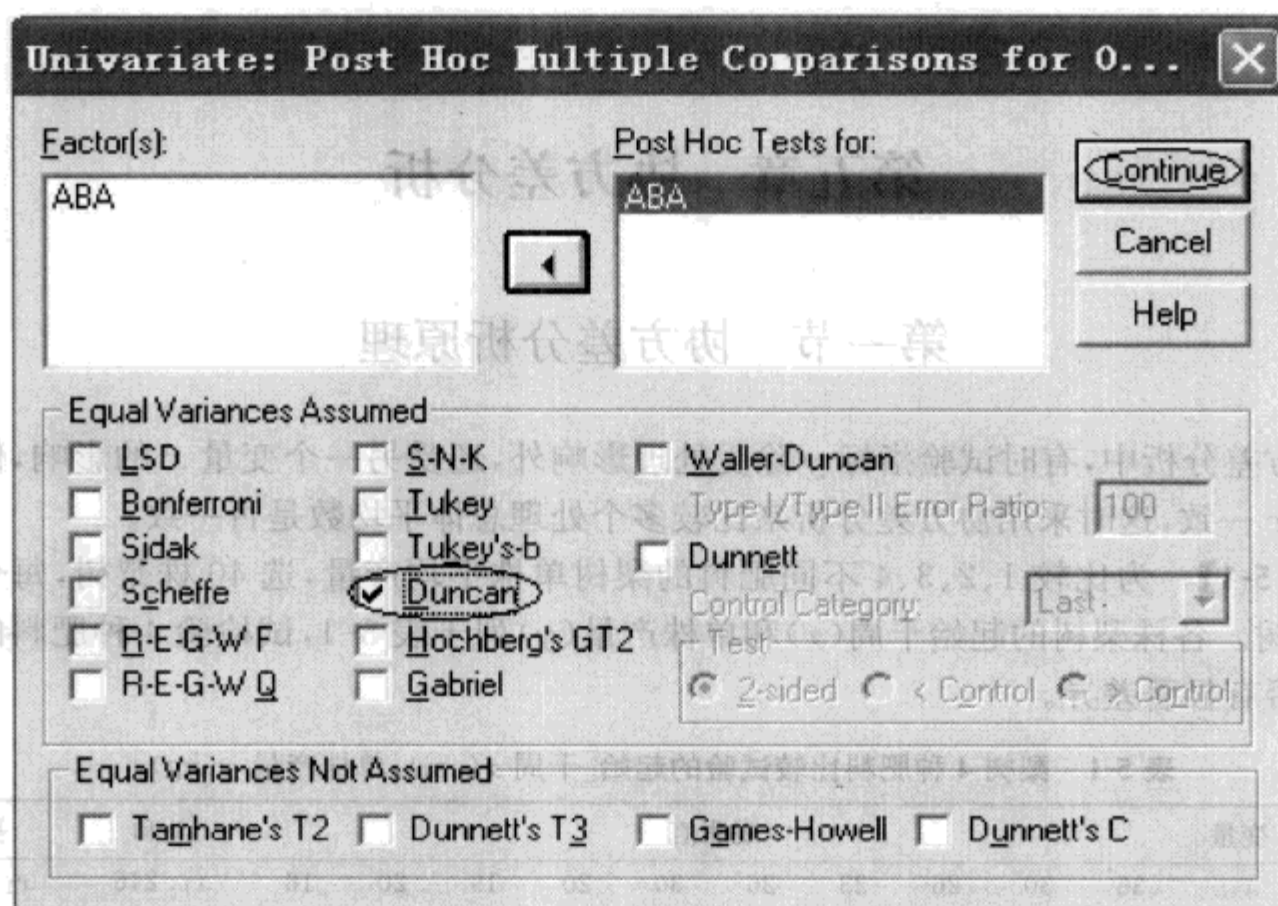


图 4-10(k) 选择多重比较方法

Output 11 - SPSS Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help

Dependent Variable: X Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6402.375	3	2134.125	25.070	.005
Intercept	55278.125	1	55278.125	649.376	.000
ABA	6402.375	3	2134.125	25.070	.005
Error	340.500	4	85.125		
Total	62021.000	8			
Corrected Total	6742.875	7			

Duncan^{a,b}

ABA	N	Subset		
		1	2	3
111	2	46.50		
222	2	70.00	70.00	
121	2		93.00	
212	2			123.00
Sig.		.064	.067	1.000

图 4-10(l) Output 窗口的统计结果

多重比较表表明: ABA=212 的搭配平均数显著高于其他搭配的平均数, 即 A 取 2, B 取 1 时, 培养基抗菌力最强。

第五章 协方差分析

第一节 协方差分析原理

在方差分析中,有时试验指标 y 除受处理影响外,还受另一个变量 x 的影响,但又不能保证 x 一致,这时采用协方差分析来比较多个处理总体平均数是否一致。

【例 5-1】 为比较 1、2、3、4 不同肥料的梨树单株平均产量,选 40 株梨树,每个处理 10 株梨树。各株梨树的起始干周(x)和单株产量(y)列于表 5-1,试检验 4 种肥料的单株产量是否有显著差异。

表 5-1 梨树 4 种肥料比较试验的起始[干周 $x(\text{cm})$;单株产量 $y(\text{kg})$]

肥料	变量	观测值										总和	平均
1	x_{1j}	36	30	26	23	26	30	20	19	20	16	$x_1 \cdot 246$	$\bar{x}_1 \cdot 24.6$
	y_{1j}	89	80	74	80	85	68	73	68	80	58	$y_1 \cdot 755$	$\bar{y}_1 \cdot 75.5$
2	x_{2j}	28	27	27	24	25	23	20	18	17	20	$x_2 \cdot 229$	$\bar{x}_2 \cdot 22.9$
	y_{2j}	64	81	73	67	77	67	64	65	59	57	$y_2 \cdot 674$	$\bar{y}_2 \cdot 67.4$
3	x_{3j}	28	33	26	22	23	20	22	23	18	17	$x_3 \cdot 232$	$\bar{x}_3 \cdot 23.2$
	y_{3j}	55	62	58	58	66	55	60	71	55	48	$y_3 \cdot 588$	$\bar{y}_3 \cdot 58.8$
4	x_{4j}	32	23	27	23	27	28	20	24	19	17	$x_4 \cdot 240$	$\bar{x}_4 \cdot 24.0$
	y_{4j}	52	58	64	62	54	54	55	44	51	51	$y_4 \cdot 545$	$\bar{y}_4 \cdot 54.5$
总和												$x_{..} \cdot 947$	$\bar{x}_{..} \cdot 23.675$
												$y_{..} \cdot 2562$	$\bar{y}_{..} \cdot 64.05$

以上实例中 40 株梨树的干周不同。一般干周不同,梨树的产梨量也不同,因此必须考虑梨树的干周对产梨量的影响,这样统计分析的结论才可靠。

在除去处理的效应的情况下,梨树的起始干周 x 和单株产量 y 有关系,设: $y - \bar{y} = b_1(x - \bar{x})$,将每个样本点(x_i, y_i)中的 y_i 进行调整: $y'_i = y_i - b_1(x_i - \bar{x})$,这样的 y'_i 是在 x 一致条件下的依变量,进行多个平均数的比较结论可靠。已经证明在除去处理的效应的情况下, $b_1 = b_e = SP_e / SS_{ex}$,具体计算见表 5-2。

表 5-2 表 5-1 资料的自由度、平方和与乘积和

变异来源	df	SS_x	SS_y	SP
组间	$k-1$	$SS_{tx} = \frac{\sum_{i=1}^k x_i^2}{n} - \frac{x_{..}^2}{kn}$	$SS_{ty} = \frac{\sum_{i=1}^k y_i^2}{n} - \frac{y_{..}^2}{kn}$	$SP_t = \frac{\sum_{i=1}^k x_i y_i}{n} - \frac{x_{..} y_{..}}{kn}$
误差	$kn-k$	$SS_{ex} = SS_{Tx} - SS_{tx}$	$SS_{ey} = SS_{Ty} - SS_{ty}$	$SP_e = SP_T - SP_t$
总变异	$kn-1$	$SS_{Tx} = \sum_{i=1}^k \sum_{j=1}^n x_{ij}^2 - \frac{x_{..}^2}{kn}$	$SS_{Ty} = \sum_{i=1}^k \sum_{j=1}^n y_{ij}^2 - \frac{y_{..}^2}{kn}$	$SP_T = \sum_{i=1}^k \sum_{j=1}^n x_{ij} y_{ij} - \frac{x_{..} y_{..}}{kn}$

有关的乘积和计算如下:

总乘积和

$$SP_T = \sum \sum xy - \frac{x_{..}y_{..}}{nk} = 36 \times 89 + \cdots + 17 \times 51 - \frac{947 \times 2562}{10 \times 4} = 720.650$$

肥料间乘积和

$$SP_t = \frac{\sum x_i y_{i.}}{n} - \frac{x_{..}y_{..}}{nk} = \frac{246 \times 755 + 229 \times 674 + 232 \times 588 + 240 \times 545}{10} - \frac{947 \times 2562}{10 \times 4} = 73.850$$

误差乘积和

$$SP_e = SP_T - SP_t = 720.650 - 73.850 = 646.800$$

结果见表 5-3。

表 5-3 起始干周 x 和单株产量 y 计算表

变异来源	df	x 变量	y 变量	SP
		SS	SS	SP
肥料间	3	17.875	2610.900	73.850
肥料内(误差)	36	878.900	1951.000	646.800
总变异	39	896.775	4561.900	720.650

$$b_e = \frac{SP_e}{SS_{ex}} = \frac{646.800}{878.900} = 0.7359$$

$b_e = 0.7359$ 表示起始干周改变 1cm, 单株产量将平均改变 0.7359kg。

对 b_e 进行显著性检验如下:

$$H_0: \beta_e = 0$$

回归平方和

$$SS_{eR} = \frac{SP_e^2}{SS_{ex}} = \frac{646.8^2}{878.9} = 475.993$$

离回归平方和

$$SS_{er} = SS_{ey} - SS_{eR} = 1951.000 - 475.993 = 1475.007$$

离回归自由度

$$df_{er} = k(n-1) - 1 = 4 \times (10-1) - 1 = 35$$

$$F = \frac{SS_{eR}}{SS_{er}/df_{er}} = \frac{475.993}{1475.000/35} = 11.29 > F_{0.01(1,35)} = 7.42$$

表明 y 与 x 有极显著的直线回归关系。对矫正后的单株产量做方差分析。

1. 求矫正后的单株产量的各项平方和及自由度

利用线性回归关系对单株产量——做矫正, 并由矫正后的单株产量进行方差分析计算量大, 且舍入误差大。统计学已证明, 矫正后的单株产量的总平方和、误差平方和及自由度等于其相应变异项的离回归平方和及自由度, 各项平方和及自由度可直接计算如下。

矫正单株产量的总平方和与自由度, 即总离回归平方和与自由度, 记为 SS'_T, df'_T

$$\begin{aligned} SS'_T &= SS_{Ty} - SS_{Ry} = SS_{Ty} - \frac{SP_{Ty}^2}{SS_{Tx}} \\ &= 4561.900 - \frac{720.650^2}{896.775} = 3982.784 \end{aligned} \tag{5-1}$$

$$\begin{aligned} df'_T &= df_{Ty} - df_{Ry} = 39 - 1 = 38 \\ \text{矫正单株产量的误差项平方和与自由度,即误差离回归平方和与自由度,记为 } SS'_e, df'_e \\ SS'_e &= SS_{er} = 1475.007 \\ df'_e &= df_{er} = 35 \end{aligned} \tag{5-2}$$

$$\begin{aligned} \text{矫正单株产量处理间平方和与自由度,记为 } SS'_i, df'_i \\ SS'_i &= SS'_T - SS'_e = 3982.784 - 1475.007 = 2507.777 \\ df'_i &= df'_T - df'_e = k - 1 = 4 - 1 = 3 \end{aligned} \tag{5-3}$$

2. 对矫正单株产量进行方差分析

对矫正单株产量进行方差分析,见表 5-4。

表 5-4 矫正单株产量的方差分析表

变异来源	df	SS	MS	F 值
肥料间	3	2507.777	835.926	19.835**
肥料内(误差)	35	1475.007	42.143	
总变异	38	3982.784		

** 表示因素水平的不同,平均数极显著不都相同

查 F 值表, $F_{0.01(3,35)} = 4.40$, 由于 $F = 19.835 > F_{0.01(3,35)}$, $p < 0.01$ 表明不同肥料的矫正单株产量间存在极显著的差异,需进一步进行多重比较。

3. 根据线性回归关系计算各肥料的矫正平均单株产量

矫正平均单株产量的计算公式如下:

$$\bar{y}'_i = \bar{y}_i - b_e(\bar{x}_i - \bar{x}_{..}) \tag{5-4}$$

式中, $\bar{y}'_i$ 为第 i 处理矫正单株平均产量; $\bar{y}_i$ 为第 i 处理实际单株平均产量; $\bar{x}_i$ 为第 i 处理实际平均起始干周; $\bar{x}_{..}$ 为全试验 x_{ij} 的平均数; b_e 为误差回归系数。

将各有关数值代入式(5-4)中,即可计算出各肥料的矫正单株平均产量

$$\begin{aligned} \bar{y}'_1 &= \bar{y}_1 - b_e(\bar{x}_1 - \bar{x}_{..}) = 75.5 - 0.7359 \times (24.6 - 23.675) = 74.819 \\ \bar{y}'_2 &= \bar{y}_2 - b_e(\bar{x}_2 - \bar{x}_{..}) = 67.4 - 0.7359 \times (22.9 - 23.675) = 67.970 \\ \bar{y}'_3 &= \bar{y}_3 - b_e(\bar{x}_3 - \bar{x}_{..}) = 58.8 - 0.7359 \times (23.2 - 23.675) = 59.150 \\ \bar{y}'_4 &= \bar{y}_4 - b_e(\bar{x}_4 - \bar{x}_{..}) = 54.5 - 0.7359 \times (24.0 - 23.675) = 54.261 \end{aligned}$$

4. 各肥料矫正单株平均产量间的多重比较

(1) t 检验法。

检验两个处理矫正平均数间的差异显著性,可应用 t 检验法

$$t = \frac{\bar{y}'_i - \bar{y}'_j}{S_{\bar{y}'_i - \bar{y}'_j}}, \quad df = df'_e \tag{5-5}$$

$$S_{\bar{y}'_i - \bar{y}'_j} = \sqrt{MS'_e \left[\frac{2}{n} + \frac{(\bar{x}_i - \bar{x}_j)^2}{SS_e} \right]} \tag{5-6}$$

式中, $\bar{y}'_{i.} - \bar{y}'_{j.}$ 为两个处理矫正平均数间的差数; $S_{\bar{y}'_{i.} - \bar{y}'_{j.}}$ 为两个处理正矫平均数差数标准误; MS'_e 为误差离回归均方; df'_e 为误差离回归自由度; n 为各处理的重复数; $\bar{x}_{i.}$ 、 $\bar{x}_{j.}$ 为处理 i 、 j 的 x 变量的平均数; SS_{ex} 为 x 变量的误差平方和。

例如, 检验 A_1 与 A_2 矫正单株平均产量间的差异显著性: 将有关数值代入式(5-6)得

$$S_{\bar{y}'_{i.} - \bar{y}'_{j.}} = \sqrt{42.143 \times \left[\frac{2}{10} + \frac{(24.6 - 22.9)^2}{878.9} \right]} = 2.927$$

于是

$$t = \frac{\bar{y}'_{i.} - \bar{y}'_{j.}}{S_{\bar{y}'_{i.} - \bar{y}'_{j.}}} = \frac{74.819 - 69.97}{2.927} = 1.657$$

查 t 值表, 当自由度为 35 时, $t_{0.05(35)} = 2.030$, $t < t_{0.05(35)}$, $p > 0.05$ 表明肥料 A_1 与 A_2 矫正单株平均产量间的差异不显著。其余的每两处理矫正平均数间的比较都需另行算出 $S_{\bar{y}'_{i.} - \bar{y}'_{j.}}$, 再进行 t 检验。

(2) LSD 法。

利用 t 检验法进行多重比较, 每一次比较都要算出各自的 $S_{\bar{y}'_{i.} - \bar{y}'_{j.}}$ 。当误差项自由度在 20 以上, x 变量的变异不甚大(即 x 变量各处理平均数间差异不显著), 为简便起见, 可计算一个平均的 $\bar{S}_{\bar{y}'_{i.} - \bar{y}'_{j.}}$, 采用 LSD 法进行多重比较。 $\bar{S}_{\bar{y}'_{i.} - \bar{y}'_{j.}}$ 的计算公式如下:

$$\bar{S}_{\bar{y}'_{i.} - \bar{y}'_{j.}} = \sqrt{\frac{2MS'_e}{n} \left[1 + \frac{SS_{ex}}{SS_{ex}(k-1)} \right]} \quad (5-7)$$

式中, SS_{ex} 为 x 变量的处理间平方和。

然后由误差自由度 df'_e 查临界 t 值 $t_{\alpha(df'_e)}$, 计算出最小显著差数

$$LSD_{\alpha} = t_{\alpha(df'_e)} \bar{S}_{\bar{y}'_{i.} - \bar{y}'_{j.}} \quad (5-8)$$

本例中,

$$\bar{S}_{\bar{y}'_{i.} - \bar{y}'_{j.}} = \sqrt{\frac{2 \times 42.143}{10} \left[1 + \frac{17.875}{878.9 \times (4-1)} \right]} = 2.913$$

由 $df'_e = 35$, 查临界 t 值得 $t_{0.05(35)} = 2.030$, $t_{0.01(35)} = 2.724$

于是

$$LSD_{0.05} = 2.030 \times 2.913 = 5.913$$

$$LSD_{0.01} = 2.724 \times 2.913 = 7.935$$

不同肥料的矫正平均单株产量间的多重比较结果见表 5-5。

表 5-5 不同肥料的矫正平均单株产量间的多重比较表(LSD 法)

肥 料	矫正单株产量平均重 $\bar{y}'_{i.}$	$\bar{y}'_{i.} - 54.261$	$\bar{y}'_{i.} - 59.150$	$\bar{y}'_{i.} - 67.970$
1	74.819	20.558**	15.668**	6.849*
2	67.970	13.709**	8.820**	
3	59.150	4.889		
4	54.261			

* 表示因素水平的不同, 平均数显著不都相同; ** 表示因素水平的不同, 平均数极显著不都相同; 不标记表示因素水平的不同, 平均数差异不显著

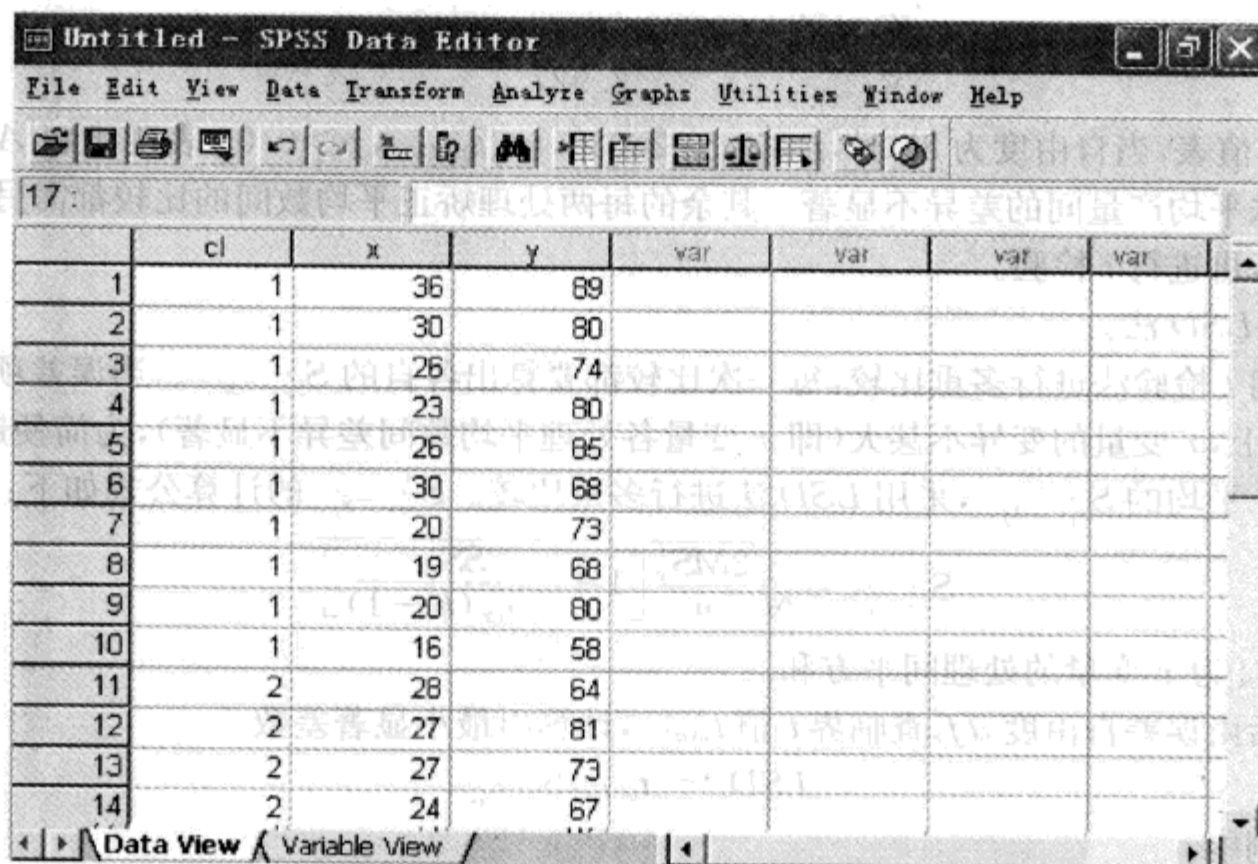
多重比较结果表明: 除肥料 3、4 单株平均产量间的差异不显著外, 其余各种肥料两两单株平均产量间差异显著或极显著, 这里表现为肥料 1 的单株平均产量显著或极显著高

于其余 3 种肥料的单株平均产量;肥料 2 的单株平均产量极显著高于肥料 3、4 的单株平均产量。4 种肥料中以 1 的单株产量最高,2 次之,3、4 的单株产量最低。

第二节 协方差分析与 SPSS 应用

SPSS 上的协方差分析,以【例 5-1】为例。

首先运行 SPSS,SPSS 的数据格式、操作方法和统计结果见图 5-1(a)~图 5-1(g)。



	cl	x	y	var	var	var	var
1	1	36	89				
2	1	30	80				
3	1	26	74				
4	1	23	80				
5	1	26	85				
6	1	30	68				
7	1	20	73				
8	1	19	68				
9	1	20	80				
10	1	16	58				
11	2	28	64				
12	2	27	81				
13	2	27	73				
14	2	24	67				

图 5-1(a) 数据格式

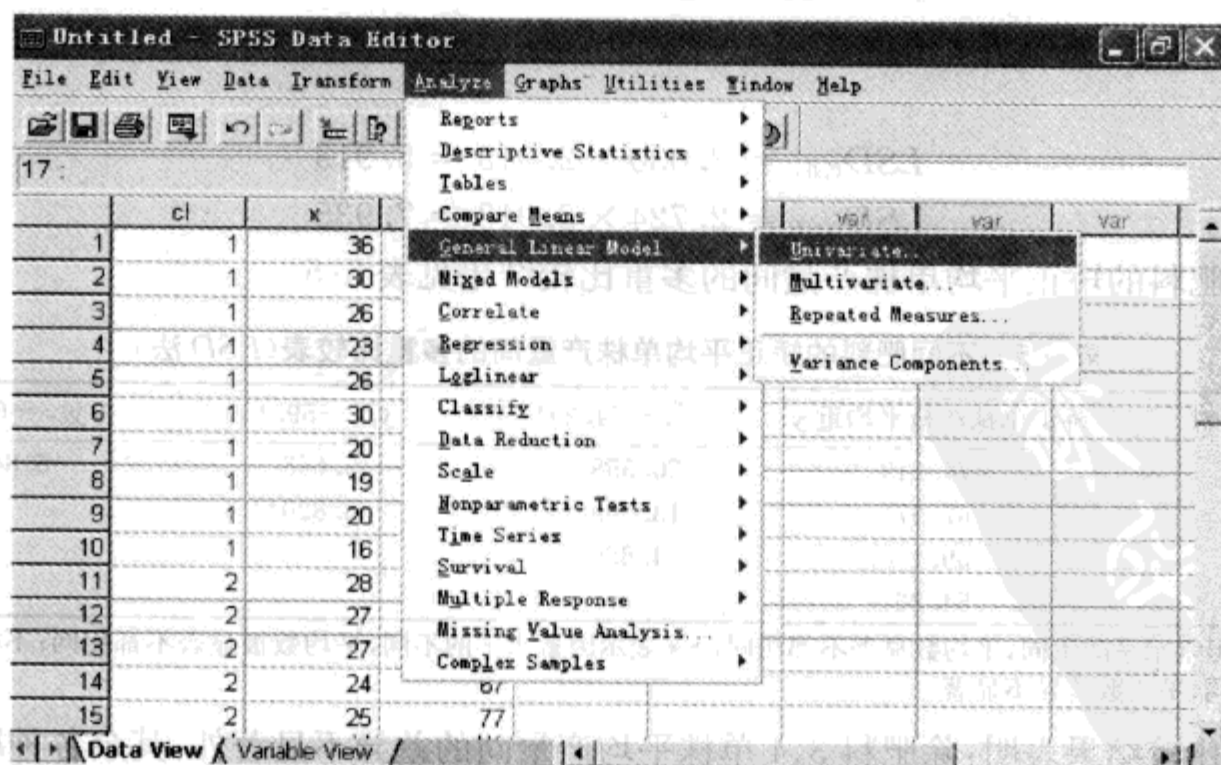


图 5-1(b) 运行方差分析

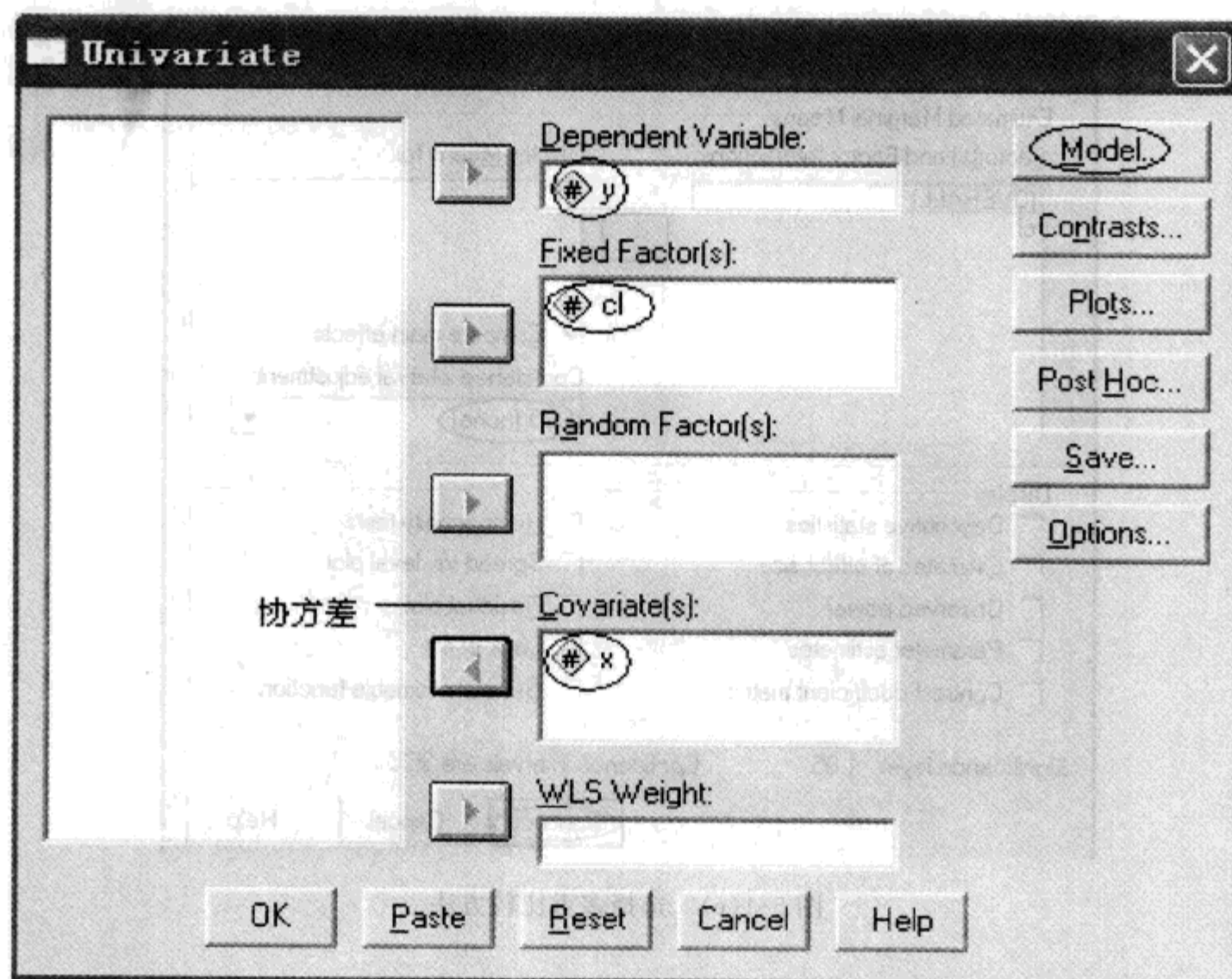


图 5-1(c) 选择变量

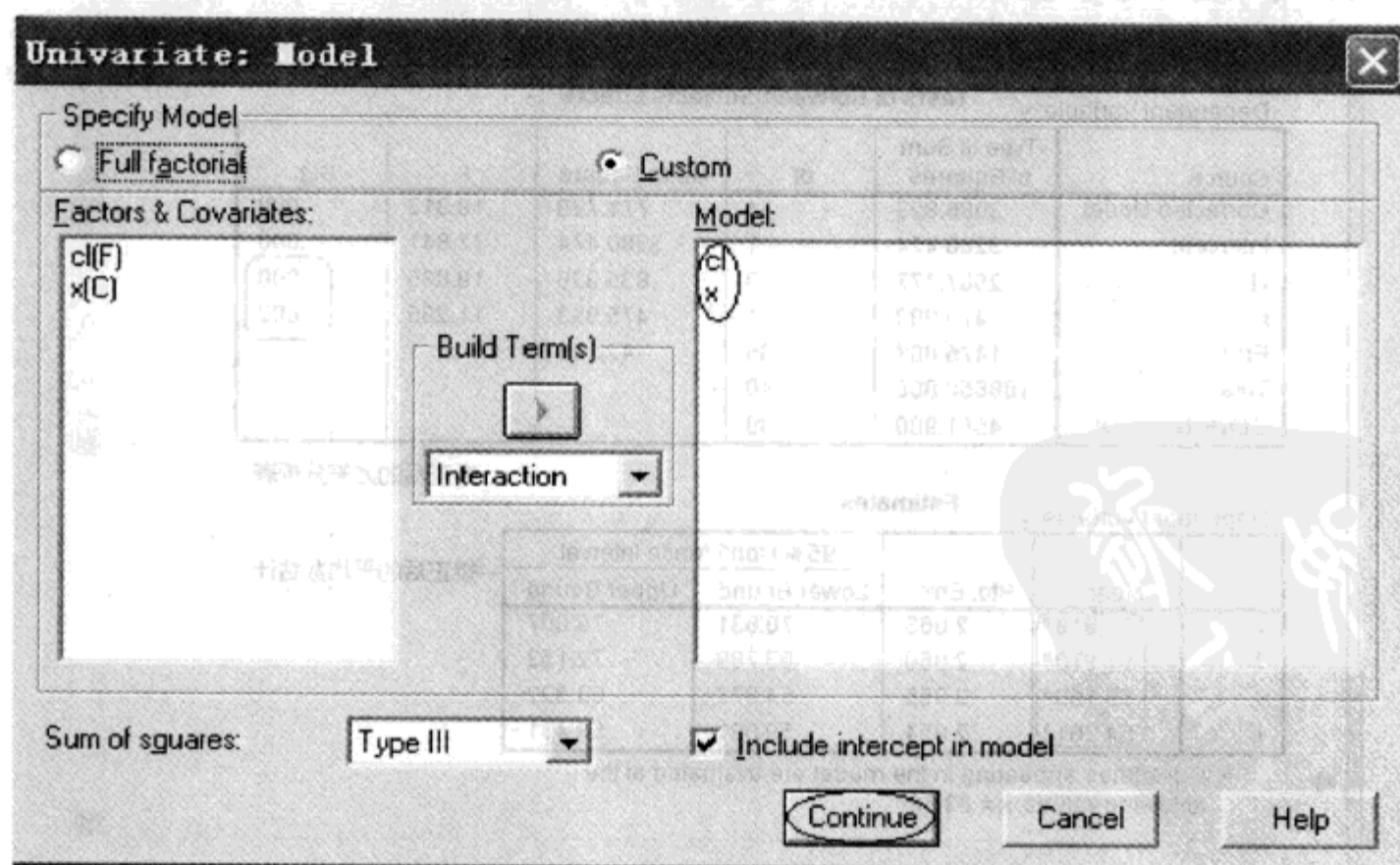


图 5-1(d) 选择模型

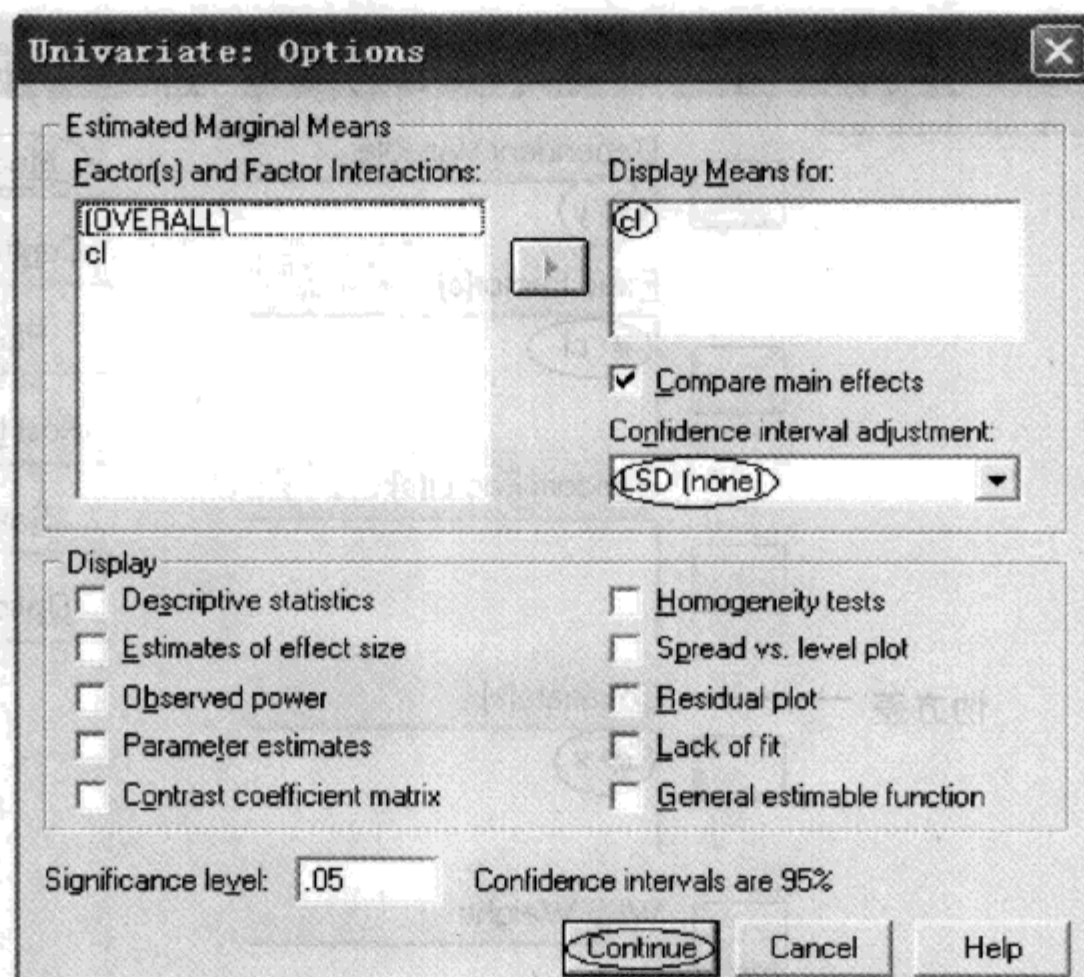


图 5-1(e) 选择多重比较方法

Output 14 - SPSS Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help

Dependent Variable: y

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3086.893	4	771.723	18.312	.000
Intercept	3280.474	1	3280.474	77.841	.000
cl	2507.777	3	835.926	19.835	.000
x	475.993	1	475.993	11.295	.002
Error	1475.007	35	42.143		
Total	168658.000	40			
Corrected Total	4561.900	39			

修正后的方差分析表

Dependent Variable: y

Estimates

cl	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	74.819 ^a	2.063	70.631	79.007
2	67.970 ^a	2.060	63.789	72.152
3	59.150 ^a	2.056	54.977	63.322
4	54.261 ^a	2.054	50.091	58.431

修正后的平均数估计

a. Covariates appearing in the model are evaluated at the following values: x = 23.68.

图 5-1(f) Output 窗口的方差分析结果

Output 14 - SPSS Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help

Dependent Variable: y

Pairwise Comparisons 多重比较表(矫正后)

(I) cl	(J) cl	Mean Difference (I-J)	Std. Error	Sig. ^a	95% Confidence Interval for Difference ^a	
					Lower Bound	Upper Bound
1	2	6.849*	2.927	.025	.907	12.791
	3	15.670*	2.919	.000	9.743	21.596
	4	20.558*	2.906	.000	14.659	26.458
2	1	-6.849*	2.927	.025	-12.791	-.907
	3	-8.821*	2.904	.004	-2.925	-14.716
	4	-13.710*	2.913	.000	-7.795	-19.624
3	1	-15.670*	2.919	.000	-21.596	-9.743
	2	-8.821*	2.904	.004	-14.716	-2.925
	4	-4.889	2.908	.102	-1.016	10.793
4	1	-20.558*	2.906	.000	-26.458	-14.659
	2	-13.710*	2.913	.000	-19.624	-7.795
	3	-4.889	2.908	.102	-10.793	1.016

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

图 5-1(g) Output 窗口的多重比较结果

由图 5-1(f)的方差分析表可知, x 对 y 的影响极显著($\text{Sig.} = 0.002 \leq 0.01$), 处理的不同 y 的平均数极显著的不同($\text{Sig.} = 0.000 \leq 0.01$), 结论是: 肥料的不同, 梨树单株平均产量极显著不都相同。

由图 5-1(g)的多重比较表得出结论: 肥料 1 的梨树单株平均产量显著高于肥料 2 的, 极显著高于肥料 3、4 的(因为肥料 1 与肥料 2 的梨树单株平均产量差异的 $\text{Sig.} = 0.025 > 0.05$, 差异显著, 同理可比较其他的平均数), 肥料 2 的梨树单株平均产量极显著高于肥料 3、4 的, 肥料 3、4 的梨树单株平均产量差异不显著。

将图 5-1(f)、图 5-1(g)用中文表示见表 5-6、表 5-7。

表 5-6 方差分析表(矫正)

变异来源	SS	df	MS	F	Sig.
x	475.993	1	475.993	11.295	0.002
肥料间	2507.777	3	835.926	19.835	0.000
误差	1475.007	35	42.143		
总计	4561.900	39			

表 5-7 多重比较表(矫正)LSD

肥 料	平均产量	$\bar{x}_i - 54.26$	$\bar{x}_i - 59.15$	$\bar{x}_i - 67.97$
1	74.82	20.56**	15.67**	6.85*
2	67.97	13.71**	8.82*	
3	59.15	4.89		
4	54.26			

* 表示平均数显著不同; ** 表示平均数极显著不同; 不标记表示平均数差异不显著

x 项的显著性水平 $\text{Sig.} = 0.002$ 表明 x, y 的线性关系极显著; 肥料间 $\text{Sig.} = 0.000$ 表明肥料的不同梨树的单棵平均产量极显著不都相等。计算结果同前计算一致。多重比较表用三角形法表示见表 5-7。

第三节 随机区组设计协方差分析与 SPSS 应用

【例 5-2】 对 6 个菜豆品种($k=6$)进行维生素 C 含量($y, \text{mg}/100\text{g}$)比较试验, 4 次重复($n=4$), 随机区组设计。根据前人的研究, 菜豆维生素 C 含量不仅与品种有关, 而且与豆荚的成熟度有关。但在试验中又无法使所有小区的豆荚都同时成熟, 所以同时测定了 100g 所采豆荚干物重百分率 x , 作为豆荚成熟度指标。测定结果列于表 5-8, 作协方差分析。

表 5-8 菜豆品种维生素 C 含量 y 与豆荚干物重百分率 x 的测定结果

品 种	区组 I		区组 II		区组 III		区组 IV	
	x_{i1}	y_{i1}	x_{i2}	y_{i2}	x_{i3}	y_{i3}	x_{i4}	y_{i4}
1	34.0	93.0	33.4	94.8	34.7	91.7	38.9	80.8
2	39.6	47.3	39.8	51.5	51.2	33.3	52.0	27.2
3	31.7	81.4	30.1	109.0	33.8	71.6	39.6	57.5
4	37.7	66.9	38.2	74.1	40.3	64.7	39.4	69.3
5	24.9	119.5	24.0	128.5	24.9	125.6	23.5	129.0
6	30.3	106.6	29.1	111.4	31.7	99.0	28.3	126.1

使用 SPSS 进行协方差分析。

首先运行 SPSS, SPSS 的数据格式、操作方法和统计结果见图 5-2(a)~图 5-2(g)。

	cl	qz	x	y	var	var	var
1	1	1	34.0	93.0			
2	2	1	39.6	47.3			
3	3	1	31.7	81.4			
4	4	1	37.7	66.9			
5	5	1	24.9	119.5			
6	6	1	30.3	106.6			
7	1	2	33.4	94.8			
8	2	2	39.8	51.5			
9	3	2	30.1	109.0			
10	4	2	38.2	74.1			
11	5	2	24.0	128.5			
12	6	2	29.1	111.4			
13	1	3	34.7	91.7			
14	2	3	51.2	33.3			

图 5-2(a) 数据格式

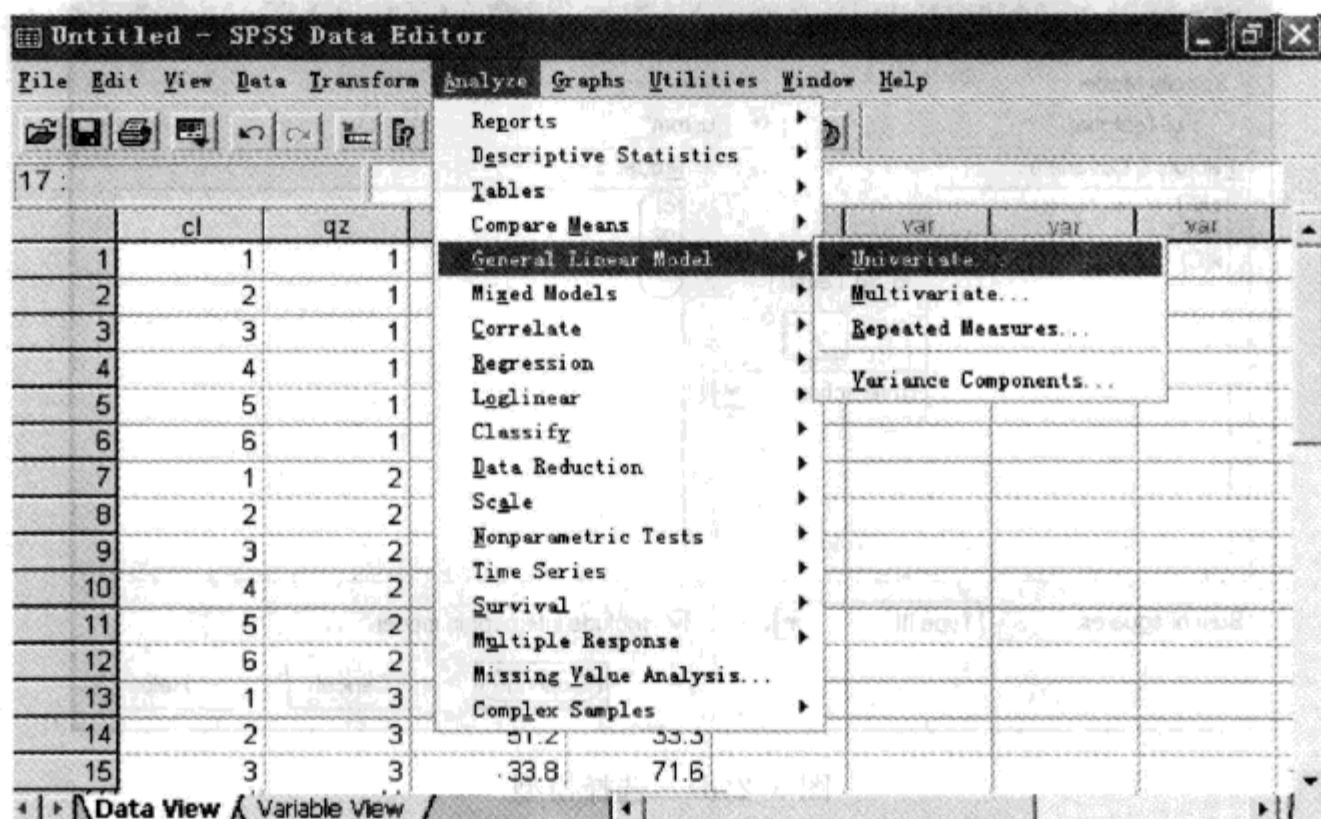


图 5-2(b) 运行方差分析

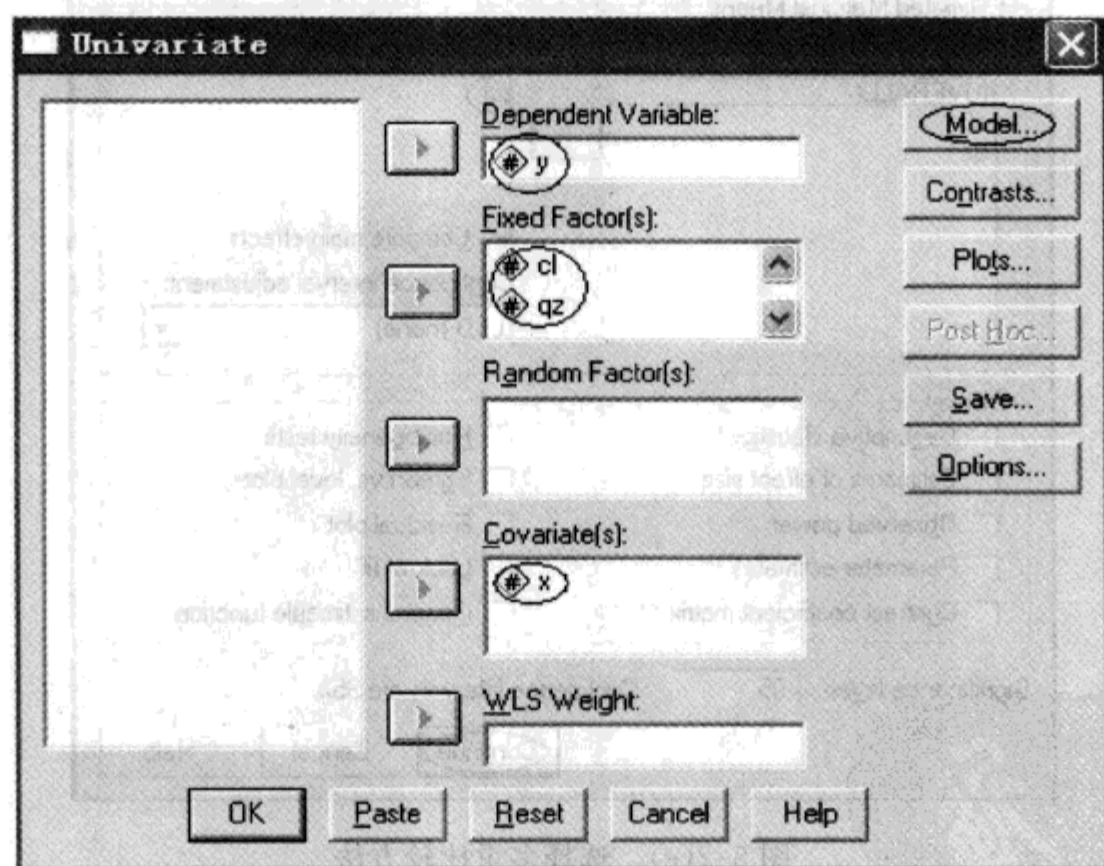


图 5-2(c) 选择变量

由图 5-2(f) 的方差分析表知, x 对 y 的影响极显著 ($\text{Sig.} = 0.001$), 区组的不同对 y 的影响不显著 ($\text{Sig.} = 0.321$), 处理的不同 y 的平均数显著的不同 ($\text{Sig.} = 0.029 < 0.05$)。结论是: 品种的不同, 维生素 C 平均含量显著不都相同。

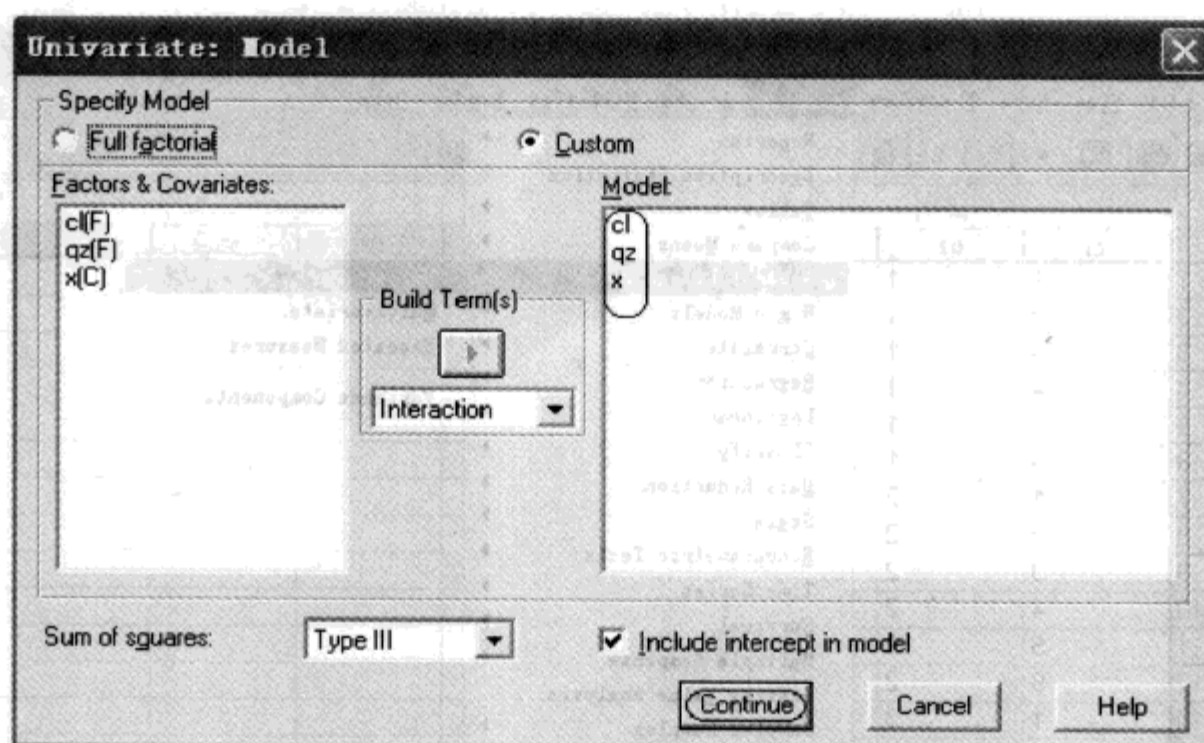


图 5-2(d) 选择模型

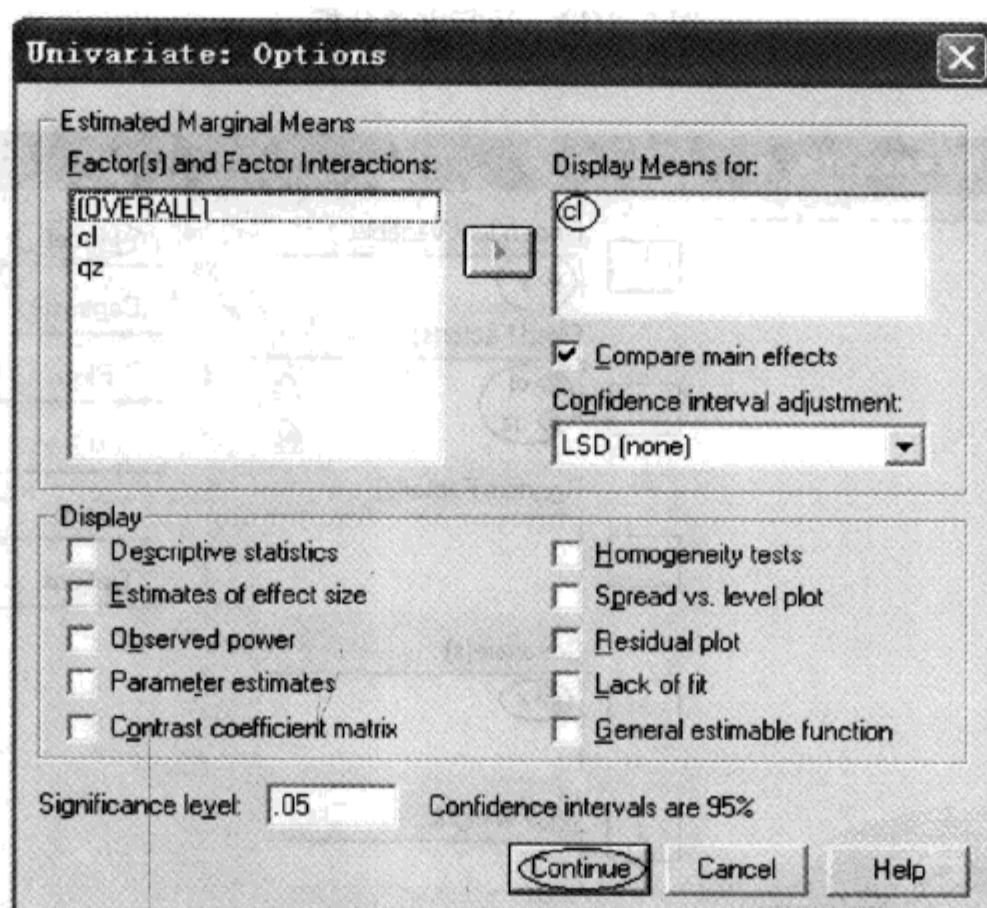


图 5-2(e) 选择多重比较方法

由图 5-2(g)的多重比较表得结论:品种 5 与品种 1、2、4、6 的维生素 C 平均含量差异不显著,但与品种 3 的维生素 C 平均含量差异显著;品种 6 与品种 1 的维生素 C 平均含量差异不显著,但与品种 2、3、4 的维生素 C 平均含量差异显著;品种 1 与品种 4 的维生素 C 平均含量差异不显著,但与品种 2、3 的维生素 C 平均含量差异显著;品种 4 与品种 2、3 的维生素 C 平均含量差异不显著;品种 2、3 的维生素 C 平均含量差异不显著。

Output14 - SPSS Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help

Dependent Variable: y

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	20376.214 ^a	9	2264.024	43.263	.000
cl	916.516	5	183.303	3.503	.029
qz	200.183	3	66.728	1.275	.321
x	960.434	1	960.434	18.353	.001
Error	732.651	14	52.332		
Total	197891.200	24			
Corrected Total	21108.865	23			

修正后的方差分析表

Dependent Variable: y

Estimates

cl	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	91.731 ^a	3.638	83.929	99.533
2	69.224 ^a	7.757	52.586	85.862
3	77.663 ^a	3.654	69.827	85.500
4	80.143 ^a	4.489	70.514	89.772
5	98.163 ^a	7.366	82.365	113.960
6	98.026 ^a	4.684	87.980	108.072

修正后的平均数估计

图 5-2(f) Output 窗口的方差分析结果

Output14 - SPSS Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help

Dependent Variable: y

Pairwise Comparisons

多重比较表(修正后)

(I) cl	(J) cl	Mean Difference (I-J)	Std. Error	Sig. ^a	95% Confidence Interval for Difference ^a	
					Lower Bound	Upper Bound
1	2	22.507*	8.252	.016	4.807	40.207
	3	14.068*	5.194	.017	2.927	25.209
	4	11.588	5.597	.057	-.417	23.594
	5	-6.432	8.511	.462	-24.687	11.823
	6	-6.295	6.121	.321	-19.424	6.834
2	3	-8.439	8.978	.363	-27.696	10.818
	4	-10.919	6.621	.121	-25.119	3.281
	5	-28.939	14.230	.061	-59.459	1.581
	6	-28.802*	11.089	.021	-52.585	-5.019
3	4	-2.480	6.021	.687	-15.393	10.434
	5	-20.500*	7.809	.020	-37.248	-3.752
	6	-20.363*	5.676	.003	-32.537	-8.190
4	5	-18.020	10.418	.106	-40.364	4.324
	6	-17.883*	7.611	.034	-34.207	-1.560
5	6	.137	6.165	.983	-13.085	13.358

图 5-2(g) Output 窗口的多重比较结果