

《概率论与数理统计》 SPSS22 的应用

王学民

读者可从

<https://anyshare.sufe.edu.cn/#/link/B9F2F217DF9A179950462AF6B590145F?path=>
下载《概率论与数理统计》配书资料，下载的资料中有一个“《概率论与数理统计》SPSS 数据”文件夹，本文均从该文件夹中打开数据表。启动 SPSS22 软件。

第一部分 第二章

一、例 2.2.4 中的 (1) 和 (2)

在 SPSS22 的空白数据编辑器窗口中，选择编辑→插入变量；编辑→插入个案→转换→计算变量...，出现“计算变量”对话框（见图 1.1）⇒在“目标变量”列表框中输入例 2.2.4 (1)，在“数字表达式”列表框中输入 $\text{PDF.BINOM}(4, 6, 0.3)$ →确定；在“目标变量”列表框中输入例 2.2.4 (2)，在“数字表达式”列表框中输入 $\text{CDF.BINOM}(3, 6, 0.3)$ （见图 1.2）→确定；在“目标变量”列表框中输入例 2.2.4 (3)，在“数字表达式”列表框中输入 $1-\text{PDF.BINOM}(0, 6, 0.3)$ （见图 1.3）→确定。随即在数据编辑器窗口中，出现三个概率计算的结果，如图 1.4 所示。



图 1.1



图 1.2



图 1.3

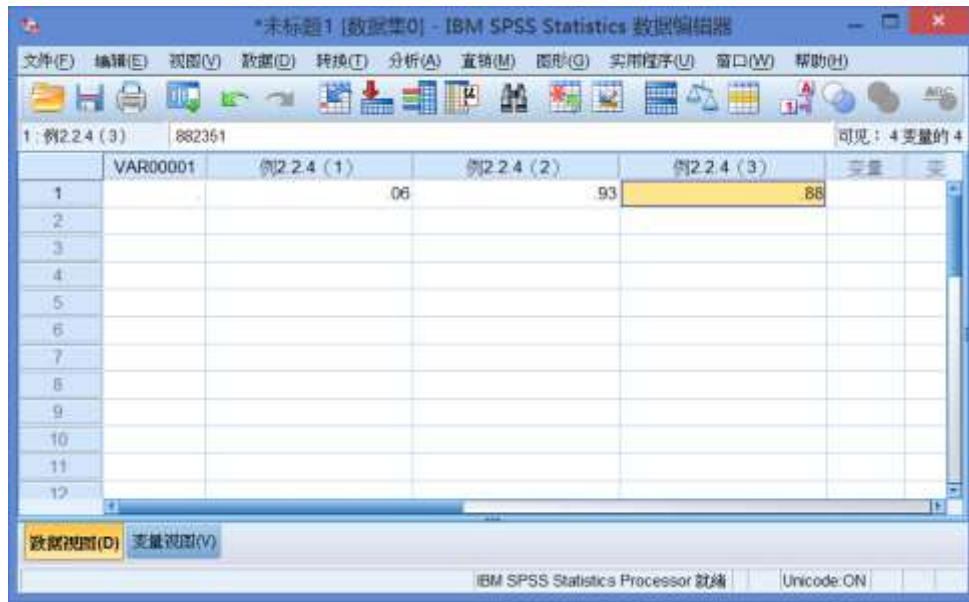


图 1.4

二、例 2.2.6 中的 (1) 和 (3)

对例 2.2.6 中的 (1) 和 (3)，作类似上述操作，并在“变量视图”中将各变量的小数位数由缺省时的“2”改为“4”。在“计算变量”对话框的“数字表达式”列表框中分别输入函数“CDF.BINOM(2, 50, 0.05)”、“CDF.POISSON(2, 2.5)”、“PDF.BINOM(3, 50, 0.05)”和“PDF.POISSON(3, 2.5)”，得到如图 1.5 所示的结果。

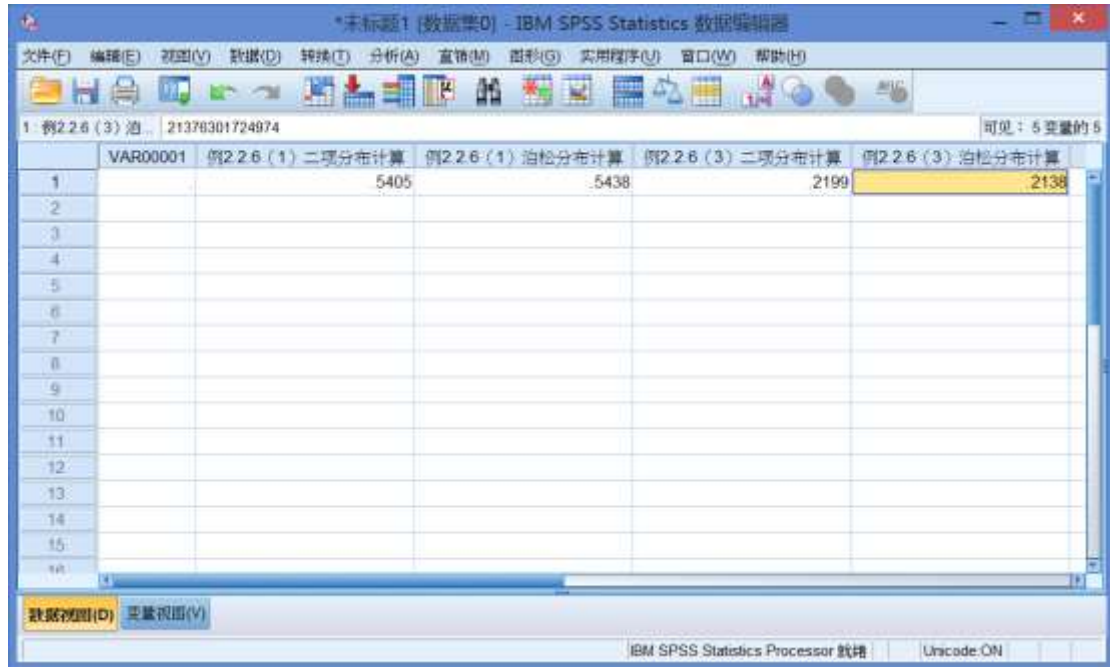


图 1.5

三、例 2.4.7 中的 $P(X < 130)$

对例 2.2.8 中的 (1)，作上述类似操作，在“计算变量”对话框的“数字表达式”列表框中输入函数“CDF.NORMAL(130, 140, 12.2)”，可得 $P(X < 130) = 0.2062$ 。

四、常用分布的分位数

在图 1.6 的对话框中，可根据给定的分布函数值，容易得到许多常用分布的分位数。

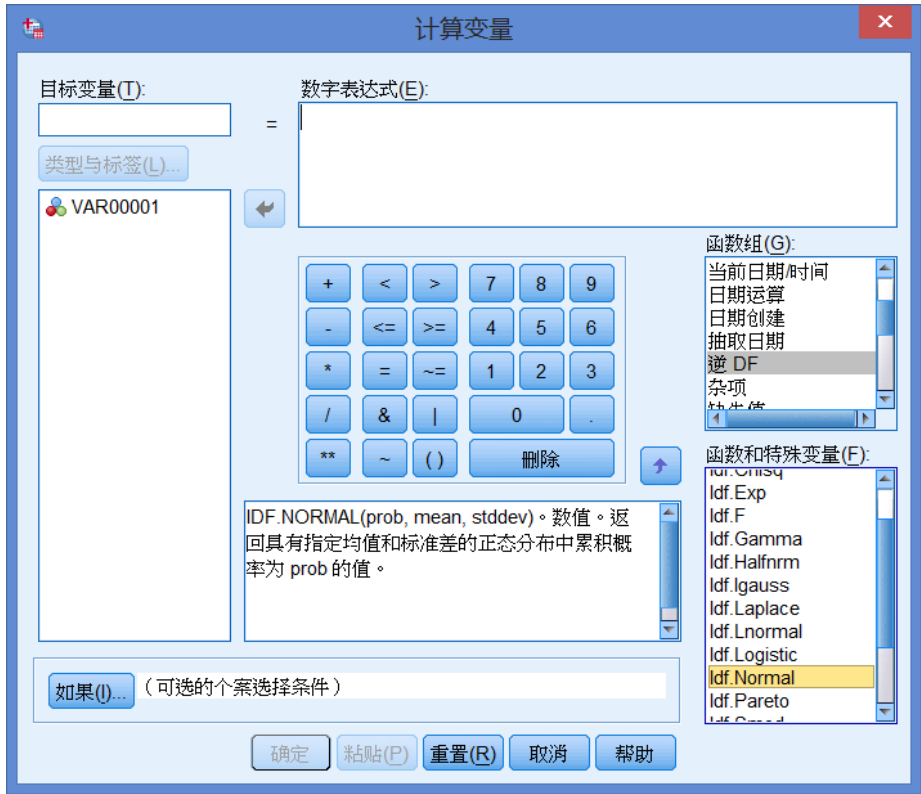


图 1.6

第二部分 第六章和第七章的 § 7.1– § 7.3

一、对例 6.4.2 中的数据进行统计推断

在 SPSS22 的数据编辑器窗口中，选择文件⇒打开>⇒数据...，即出现如图 2.1 所示的“打开数据”窗口⇒在“查找范围”列表框中选择“《概率论与数理统计》SPSS 数据”所在的文件夹。

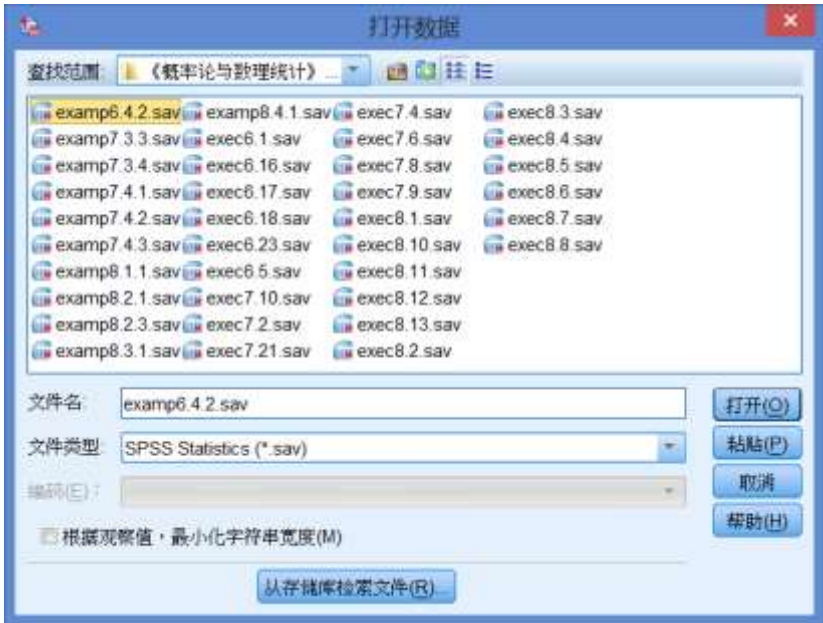


图 2.1

打开 examp6.4.2.sav 数据表（见图 2.2）⇒ 分析 ⇒ 比较平均值 ⇒ 单样本 T 检验...，随即出现“单样本 T 检验”对话框（见图 2.3）⇒ 将压力读数[x]选入“检验变量”列表框中，在“检验值”框中填入 4.2 ⇒ 选项... ⇒ 在出现的图 2.4 的框中填入 95（这是缺省值，已有）⇒ 继续 ⇒ 确定，生成图 2.5。



图 2.2



图 2.3

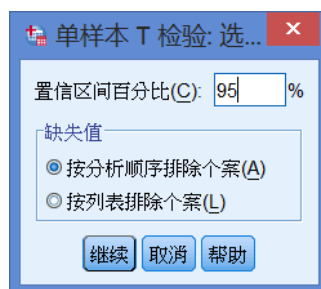


图 2.4

单样本统计				
	数字	平均值(E)	标准偏差	标准误差平均值
压力读数	8	4.0100	.30346	.10729

单样本检验						
	检验值 = 4.2					
	t	自由度	显著性 (双尾)	平均差	差值的 95% 置信区间	
					下限	上限
压力读数	-1.771	7	.120	-.19000	-.4437	.0637

图 2.5

二、对例 7.3.4 中的数据进行比较推断

打开 examp7.3.4sav 数据表 ⇒ 分析 ⇒ 比较平均值 ⇒ 独立样本 T 检验...，出现“独立样本 T 检验”对话框（见图 2.6）⇒ 将袋茶重量[x]选入“检验变量”列表框中；将机器[g]选入“分组变量”列表框中⇒定义组...⇒在“定义组”对话框中（见图 2.7），“组 1”列表框内填入甲，“组 2”列表框内填入乙⇒继续；选择选项...⇒在出现的图 2.8 的框中填入95⇒继续⇒确定，生成图 2.9。



图 2.6

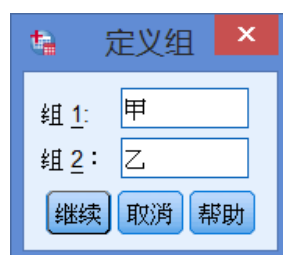


图 2.7

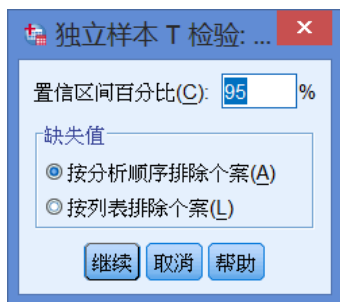


图 2.8

组统计				
机器	数字	平均值(E)	标准偏差	标准误差平均值
袋茶重量 甲	25	3.3124	.13374	.02675
乙	22	3.2782	.07682	.01638

独立样本检验									
		列文方差相等性检验		平均值相等性的 t 检验					
		F	显著性	t	自由度	显著性 (双尾)	平均差	标准误差差值	差值的 95% 置信区间
袋茶重量	已假设方差齐性	6.572	.014	1.056	45	.297	.03422	.03241	-.03106 .09950
	未假设方差齐性			1.091	39.091	.282	.03422	.03136	-.02922 .09765

图 2.9

三、对例 7.3.3 中的成对数据进行比较推断

打开 examp7.3.3sav 数据表 ⇒ 分析 ⇒ 比较平均值 ⇒ 配对样本 T 检验...，即出现“配对样本 T 检验”对话框（见图 2.10）⇒ 将种子 A 种植的谷物产量[x]选入“Variable1”列表框中，将种子 B 种植的谷物产量[y]选入“Variable2”列表框中（⇒ 选项...，在出现的图 2.11 的框中填入 95 ⇒ 继续）⇒ 确定，生成图 2.12。



图 2.10

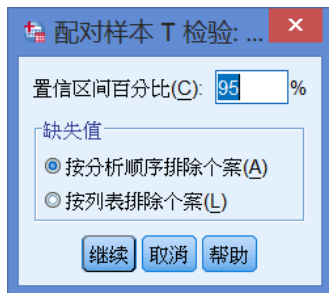


图 2.11

配对样本统计

		平均值(E)	数字	标准偏差	标准误差平均值
配对 1	种子A种植的谷物产量	31.10	10	5.763	1.822
	种子B种植的谷物产量	33.30	10	6.567	2.077

配对样本相关性

	数字	相关系数	显著性
配对 1 种子A种植的谷物产量 & 种子B种植的谷物产量	10	.748	.013

配对样本检验

	配对差值					t	自由度	显著性（双尾）
	平均值(E)	标准偏差	标准误差平均值	差值的 95% 置信区间				
				下限	上限			
配对 1 种子A种植的谷物产量 - 种子B种植的谷物产量	-2.200	4.442	1.405	-5.378	.978	-1.566	9	.152

图 2.12

第三部分 第八章的 § 8.1 和 § 8.2

一、例 8.1.2 中的单因素方差分析

打开 examp8.1.1sav 数据表 ⇒ 比较平均值 ⇒ 单因素 ANOVA...，随即出现“单因素方差分析”对话框（见图 3.1）⇒ 将混合原料所需时间[y]选入“因变量列表”框中，将机器[A]选入“因子”列表框中 ⇒ 确定，生成图 3.2。



图 3.1

ANOVA

混合原料所需时间

	平方和	df	均方	F	显著性
组之间	48.111	2	24.056	6.745	.008
组内	53.500	15	3.567		
总计	101.611	17			

图 3.2

二、例 8.2.2 中的两因素方差分析

打开 examp8.2.1.sav 数据表 ⇒ 分析 ⇒ 一般线性模型 ⇒ 单变量...，出现“单变量”对话框（见图 3.3）⇒ 将合成纤维抗断强度[y]选入“因变量”列表框中，将操作员[A]和机器

[B] 选入“固定因子”列表框中⇒确定，生成图 3.4。



图 3.3

主体间因子

	数字
操作员	1 8
	2 8
	3 8
机器	A 6
	B 6
	C 6
	D 6

主体间效应的检验

因变量: 合成纤维抗断强度

源	III 类平方和	自由度	均方	F	显著性
校正的模型	217.458 ^a	11	19.769	5.214	.004
截距	302626.042	1	302626.042	79813.462	.000
A	160.333	2	80.167	21.143	.000
B	12.458	3	4.153	1.095	.389
A * B	44.667	6	7.444	1.963	.151
错误	45.500	12	3.792		
总计	302889.000	24			
校正后的总变异	262.958	23			

a. R 平方 = .827 (调整后的 R 平方 = .668)

图 3.4

三、例 8.2.3 中的两因素方差分析

打开 examp8.2.3sav 数据表⇒分析⇒一般线性模型 >⇒单变量...⇒在“单变量”对话框中，将粘合剂的抗剪强度[y]选入“因变量”列表框中，将压强[A]和温度[B]选入“固定因子”列表框中⇒模型...，即出现“单变量：模型”对话框（见图 3.5）→作图 3.5 中的选择→继续⇒确定，生成图 3.6。



图 3.5

主体间因子

	数字
压强	60
	65
	70
	75
温度	130
	140
	150

主体间效应的检验

因变量: 黏合剂抗剪强度

源	III 类平方和	自由度	均方	F	显著性
校正的模型	5.238 ^a	5	1.048	2.918	.112
截距	1158.957	1	1158.957	3228.468	.000
A	.581	3	.194	.539	.673
B	4.658	2	2.329	6.487	.032
错误	2.154	6	.359		
总计	1166.349	12			
校正后的总变异	7.392	11			

a. R 平方 = .709 (调整后的 R 平方 = .466)

图 3.6

第四部分 第八章的 § 8.3 和 § 8.4

一、对例 8.3.1 中的数据进行一元回归分析

打开 examp8.3.1.sav 数据表 ⇒ 分析 ⇒ 回归 ⇒ 线性... ⇒ 在“线性回归”对话框中（见图 4.1），将每月家庭消费支出[y]选入“因变量”列表框中，将每月家庭收入[x]选入“自变量”列表框中 ⇒ 保存... ⇒ 在弹出的“线性回归：保存”对话框中（见图 4.2），作图中的选择 ⇒ 继续 ⇒ 确定，生成如图 4.3 所示的输出结果和如图 4.4 所示的保存结果。

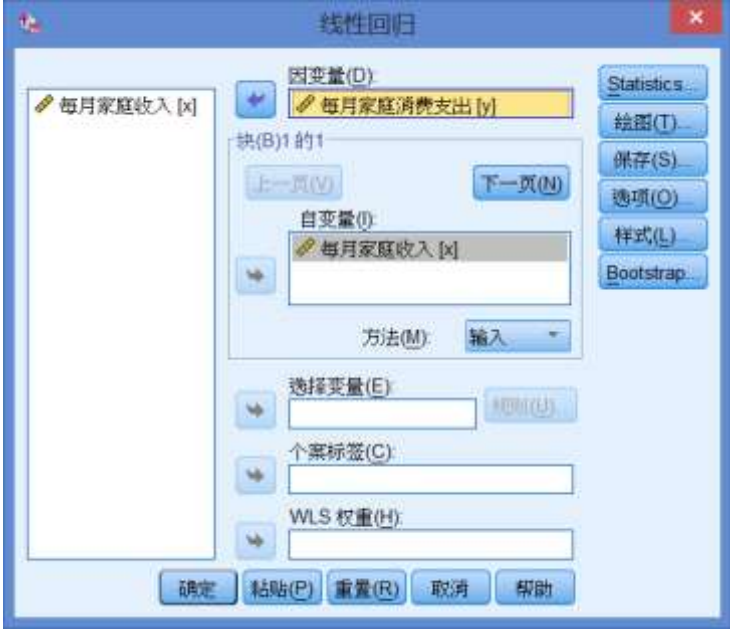


图 4.1



图 4.2

已输入/除去变量^a

模型	已输入变量	已除去变量	方法
1	每月家庭收入 ^b	.	输入

a. 因变量：每月家庭消费支出

b. 已输入所有请求的变量。

模型摘要^b

模型	R	R 平方	调整后的 R 平方	标准估算的误差
1	.938 ^a	.880	.872	368.958

a. 预测变量：（常量），每月家庭收入

b. 因变量：每月家庭消费支出

ANOVA^a

模型		平方和	自由度	均方	F	显著性
1	回归	14010659.11	1	14010659.11	102.921	.000 ^b
	残差	1905820.328	14	136130.023		
	总计	15916479.44	15			

a. 因变量：每月家庭消费支出

b. 预测变量：（常量），每月家庭收入

系数^a

模型		非标准化系数		标准系数	t	显著性
		B	标准错误	贝塔		
1	（常量）	946.184	237.155		3.990	.001
	每月家庭收入	.424	.042	.938	10.145	.000

a. 因变量：每月家庭消费支出

残差统计数据^a

	最小值	最大值(X)	平均值	标准偏差	数字
预测值	1725.76	5160.72	3162.69	966.459	16
标准预测值	-1.487	2.067	.000	1.000	16
预测值的标准误差	92.961	217.477	126.182	34.168	16
调整后的预测值	1751.19	5336.26	3174.53	984.162	16
残差	-493.178	653.540	.000	356.447	16
标准残差	-1.337	1.771	.000	.966	16
学生化残差	-1.386	1.840	-.014	1.024	16
删除的残差	-530.151	705.014	-11.844	402.077	16
学生化删除残差	-1.438	2.036	-.005	1.059	16
马氏距离(D)	.015	4.274	.937	1.112	16
Cook's 距离	.002	.326	.066	.079	16
居中杠杆值	.001	.285	.063	.074	16

a. 因变量：每月家庭消费支出

图 4.3

	x	y	PRE_1	LMCI_1	UMCI_1	LICI_1	UICI_1
1	4530	2556	2886.50541	2658.99836	3074.01246	2048.41480	3684.59602
2	5921	3694	3456.16715	3248.83179	3663.50251	2638.12007	4274.21423
3	3294	1933	2342.55004	2079.48801	2605.61207	1508.63475	3176.46533
4	8150	4565	4401.06724	4072.91697	4729.21752	3544.39013	5257.74436
5	7730	4281	4223.02416	3924.04246	4522.00586	3377.09087	5068.95745
6	2624	2332	2058.52893	1752.53950	2364.51835	1210.09354	2906.96431
7	1839	1630	1725.75792	1363.23338	2088.28246	855.33440	2596.18145
8	5538	3816	3293.80881	3094.04206	3493.57557	2477.64717	4109.97046
9	4952	3244	3045.39632	2846.01427	3244.77837	2229.32675	3861.46389
10	3426	2285	2398.50644	2143.08668	2653.92620	1566.97031	3230.04257
11	9942	4831	5160.71774	4694.27564	5627.15985	4242.14205	6079.29344
12	6134	4200	3546.46043	3332.63541	3760.28546	2726.74450	4366.17637
13	3827	2914	2568.49520	2334.14780	2802.84259	1743.18812	3393.80228
14	7228	3629	4010.22028	3743.30530	4277.13526	3175.08156	4845.35900
15	5980	2988	3481.17797	3272.19937	3690.15657	2662.71286	4299.64307
16	2544	1705	2024.61596	1713.12246	2336.10946	1174.18003	2675.05189
17	4000		2641.83199	2415.41694	2868.24704	1818.74217	3464.92181

图 4.4

二、对例 8.4.1 中的数据进行多元回归分析

打开 examp8.4.1.sav 数据表 ⇒ 分析 ⇒ 回归 ⇒ 线性... ⇒ 在“线性回归”对话框中，将产量[y]选入“因变量”列表框中，将所施肥料[x1]和土质指数[x2]选入“自变量”列表框中 ⇒ 保存... ⇒ 在弹出的“线性回归：保存”对话框（见图 4.2）中，作与图 4.2 中相同的选择 ⇒ 继续；选择绘图... ⇒ 在出现的“线性回归：图”对话框（见图 4.5）中，将“*ZRESID”（标准化残差）选入“Y”列表框，将“*ZPRED”（标准化预测值）选入“X”列表框，如图 4.5 所示 ⇒ 继续 ⇒ 确定，生成图 4.6 和图 4.7。

线性回归: 图

DEPENDENT
*ZPRED
*ZRESID
*DRESID
*ADJPRED
*SRESID
*SDRESID

散点 1 的 1
 上一页(V) 下一页(N)
 Y : *ZRESID
 X : *ZPRED

标准化残差图
☐ 直方图(H)
☒ 正态概率图(R)

☒ 产生所有部分图(P)

继续 取消 帮助

图 4.5

已输入/除去变量^a

模型	已输入变量	已除去变量	方法
1	土质指数, 所施肥料 ^b	.	输入

a. 因变量：产量

b. 已输入所有请求的变量。

模型摘要^b

模型	R	R 平方	调整后的 R 平方	标准估算的误差
1	.990 ^a	.980	.975	1.215

a. 预测变量：（常量），土质指数, 所施肥料

b. 因变量：产量

ANOVA^a

模型		平方和	自由度	均方	F	显著性
1	回归	512.571	2	256.286	173.688	.000 ^b
	残差	10.329	7	1.476		
	总计	522.900	9			

a. 因变量：产量

b. 预测变量：（常量），土质指数, 所施肥料

系数^a

模型		非标准化系数		标准系数	t	显著性
		B	标准错误	贝塔		
1	（常量）	-30.575	5.031		-6.078	.001
	所施肥料	1.041	.191	.482	5.457	.001
	土质指数	.839	.132	.562	6.364	.000

a. 因变量：产量

残差统计数据^a

	最小值	最大值(X)	平均值	标准偏差	数字
预测值	50.92	71.39	62.10	7.547	10
标准预测值	-1.482	1.231	.000	1.000	10
预测值的标准误差	.435	.898	.654	.128	10
调整后的预测值	51.41	72.25	62.10	7.535	10
残差	-1.391	2.328	.000	1.071	10
标准残差	-1.145	1.916	.000	.882	10
学生化残差	-1.456	2.353	-.001	1.076	10
删除的残差	-2.249	3.509	-.002	1.599	10
学生化删除残差	-1.614	4.765	.218	1.749	10
马氏距离(D)	.255	4.021	1.800	1.053	10
Cook's 距离	.000	.937	.172	.300	10
居中杠杆值	.028	.447	.200	.117	10

a. 因变量：产量

图 4.6

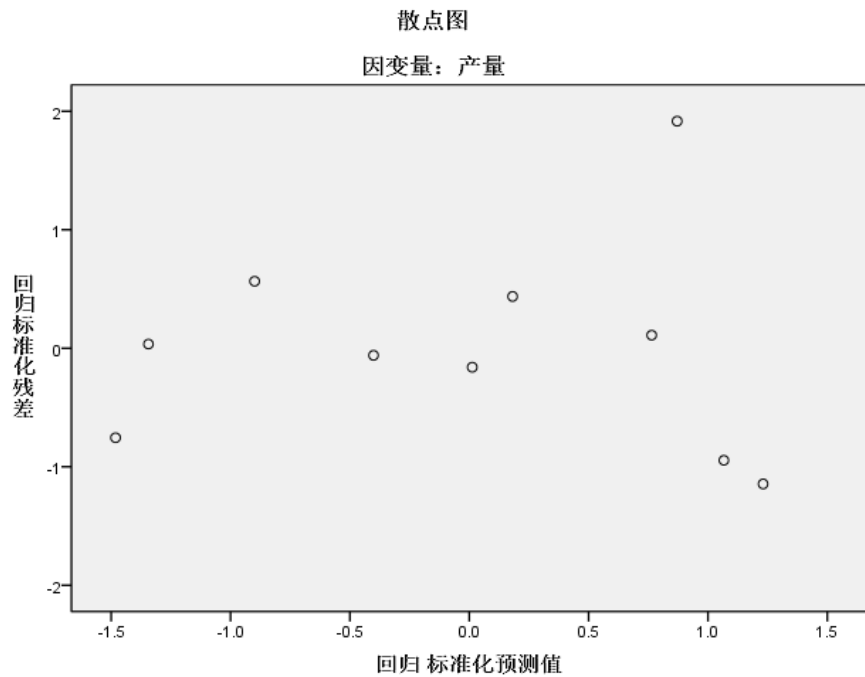


图 4.6 (续)

*examp8.4.1.sav [数据集1] - IBM SPSS Statistics 数据编辑器

文件(F) 编辑(E) 视图(V) 数据(D) 转换(T) 分析(A) 窗口(W) 帮助(H)

1: PRE_1 50.91678224687933 可见: 8 变量的 8

	x1	x2	y	PRE_1	LMCI_1	UMCI_1	LCI_1	UCI_1
1	38	50	50	50.91678	49.22343	52.61013	47.58243	54.25113
2	39	50	52	51.95738	50.27103	53.64373	48.62658	55.28818
3	39	54	56	55.31333	53.93823	56.68842	52.12678	58.49787
4	41	56	59	59.07250	58.04359	60.10141	56.02142	62.12358
5	44	56	62	62.19430	60.84334	63.54526	59.02010	65.36850
6	42	60	64	63.46905	62.08162	64.85647	60.27916	66.65894
7	43	64	68	67.86559	65.74160	69.98958	64.29323	71.43796
8	46	63	69	70.14840	68.76878	71.52803	66.96190	73.33491
9	48	62	70	71.39062	69.61627	73.16497	68.01441	74.76683
10	47	60	71	68.67205	67.00545	70.33864	65.35120	71.99289
11	40	61	-	62.22683	59.85634	64.59732	58.50263	65.95103
12								

数据视图(D) 变量视图(V)

IBM SPSS Statistics Processor 就绪 Unicode: ON

图 4.7