第 1 章　练习题参考答案

1.1 解：

操作步骤：

Analyze→Descriptive Statistics→Descriptives，将 X 选入 Variable(s)，单击 Options 选中 Mean 和 Variance

输出结果：

表 1-1 Descriptive Statistics

	N	Mean	Variance
x	10	24.7000	8.011
Valid N (listwise)	10		

结果分析：

由表 1-1 可知，样本均值 $\bar{x}$ 为 24.70，样本方差 s^2 为 8.011。

1.2 证明：

设 $X\sim N(0,1)$，$Y\sim\chi^2(n)$，且 X 和 Y 独立，则称随机变量 $T=\dfrac{X}{\sqrt{Y/n}}$ 服从自由度为 n 的 t 分布。记为 $T\sim t(n)$。

设 $U\sim\chi^2(n_1)$，$V\sim\chi^2(n_2)$，且 U 和 V 独立，则称随机变量 $F=\dfrac{U/n_1}{V/n_2}$ 服从自由度为 (n_1,n_2) 的 F 分布。记为 $F\sim F(n_1,n_2)$。

$T^2=\dfrac{X^2/1}{Y/n}$，$X^2\sim\chi^2(1)$，$Y\sim\chi^2(n)$，X 和 Y 独立，因此 X^2 和 Y 独立，则 $T^2\sim F(1,n)$。

第 2 章　练习题参考答案

2.1 **解：**（1）首先将顾客态度分别用代码 1、2、3 表示，然后在数据文件的 Varible View 窗口 Values 栏定义变量值标签：　1 代表“喜欢并愿意购买”；2 代表“不喜欢”，3 代表“喜欢并愿意购买”。操作步骤：

依次点击 File→点击 open→点击 Data→打开数据文件 ex2.1→点击 Analyze→点击 Descriptive Statistics→点击 Frequencies→将“态度”选入 Variable 框→点击 OK。输出结果如表 2.1 所示：

表2.1　30名顾客对某种产品满意程度的频数分布表

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1　喜欢并愿意购买	14	46.7	46.7	46.7
	2　不喜欢	5	16.7	16.7	63.3
	3　喜欢但不愿意购买	11	36.7	36.7	100.0
	Total	30	100.0	100.0	

（2）根据表 2.1 频数分布表资料建立的数据文件为 ex2.1.1。

绘制条形图操作步骤：依次点击 File→点击 open→点击 Data→打开数据文件 ex2.1.1→Graphs→点击 Bar→选中 Simple，选中 Summaries for groups of cases→单击 Define→选中 Other Summary function→将“人数”选入 Variable（纵轴），将“态度分类”选入 Category Axis（横轴）→点击 OK。输出结果如图 2.1 所示：

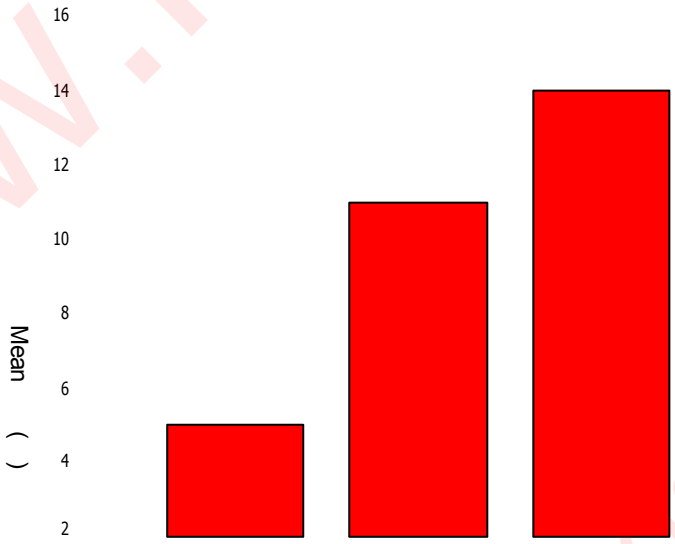


图2.1　30名顾客满意程度分布条形图

绘制饼图操作步骤：依次点击 File→点击 open→点击 Data→打开数据文件 ex2.1.1→点击 Graphs→点击 Pie→选中 Values of individual cases→点击 Define→将“人数”选入 Slices Represent 栏，将“态度分类”选入 Variable 栏→点击 OK。输出结果如图 2.2 所示：

图2.2 30名顾客满意程度分布饼图

2.2 解:首先列计算表如表 2.2 所示：

表 2.2 120 名学生英语成绩的均值、中位数、众数、偏态系数、峰度系数计算表

按成绩分组（分）	组中值（分） x_i	学生数（个） f_i	$x_i f_i$	$(x_i-\bar{x})^2 f_i$	$(x_i-\bar{x})^3 f_i$	$(x_i-\bar{x})^4 f_i$
60 以下	55	19	1045	5932.35	-104824.61	1852250.83
60～70	65	30	1950	1764.87	-13536.53	103825.18
70～80	75	42	3150	228.01	531.27	1237.86
80～90	85	18	1530	2736.52	33741.29	416030.16
90 以上	95	11	1045	5484.92	122478.22	2734938.58
合　计	—	120	8720	16146.67	38389.64	5108282.61

（1）均值 $\bar{x}=\dfrac{\sum_{i=1}^{5} x_i f_i}{\sum_{i=1}^{5} f_i}=\dfrac{8720}{120}=72.67$（分）

表 2.2 中，分布次数最多的组是“40～50”组，这就是众数所在组；$\dfrac{N}{2}$=60，中位数大约在第 60 位，可确定中位数也在“40～50”组。

$$\text{众数 } M_0 = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i = 70 + \frac{42-30}{(42-30)+(42-18)} \times 10 = 73.33(\text{分})$$

$$\text{中位数 } M_e = L + \frac{\frac{N}{2} - S_{m-1}}{f_m} \times i = 70 + \frac{\frac{120}{2} - 49}{42} \times 10 = 72.62(\text{分})$$

（2）首先计算标准差：$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 f_i}{\sum_{i=1}^{k} f_i - 1}} = 11.65(\text{分})$

$$SK = \frac{\sum_{i=1}^{k}(x-\bar{x})^3 f_i / \sum_{i=1}^{k} f_i}{s^3} = \frac{38389.64/120}{11.65^3} = 0.2023$$

由计算结果可看出，偏态系数为正值，但与零的差距不大，说明 120 名大学生英语成绩为轻微右偏分布，成绩较低的同学占有一定的比例，但偏斜程度不大。

$$K = \frac{\sum_{i=1}^{k}(x-\bar{x})^4 f_i / \sum_{i=1}^{k} f_i}{s^4} - 3 = \frac{5108282.61/120}{11.65^4} - 3 = -0.6891$$

由计算结果可看出，峰度系数为负值，说明 120 名大学生英语成绩为平峰分布，成绩较低的同学占一定比例，但低成绩区域的集中程度并不很高。

2.3 解(1)整理的组距数列如表 2.3.1 所示:

表 2.3.1　连续 60 天计算机销售量频数分布表

按销售量分组（台）	天数（天）	比重（%）
10～20	2	3.33
20～30	6	10.00
30～40	23	38.33
40～50	29	48.34
合 计	60	100.00

(2) 下面使用 SPSS16.0 绘制图形：

绘制直方图操作步骤：点击 File→点击 open→点击 Data→读取数据文件 ex2.3→点击 Graphs→点击 Histogram→将“销售量”选入 Variable 栏→点击 OK。若选中 Display normal curve 选项，则在生成的直方图上还显示正态曲线。输出结果如图 2.3.1 所示：

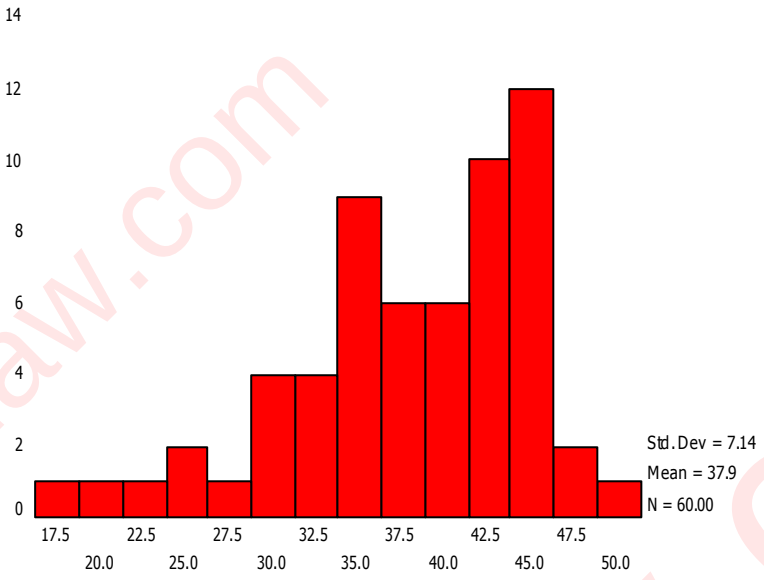


图 2.3.1　连续 60 天计算机销售量直方图

从图 2.3.1 可以看出，连续 60 天中，销售量为 40～50 台的天数较多。

绘制简单箱线图操作步骤：点击 File→点击 open→点击 Data→读取数据文件 ex2.3→点击 Graphs→点击 Boxplot→选中 simple，选中 summaries of separate Variables→点击 Define→将“销售量”选入 Boxes Represent 栏→点击 Ok。输出结果如图 2.3.2 所示：

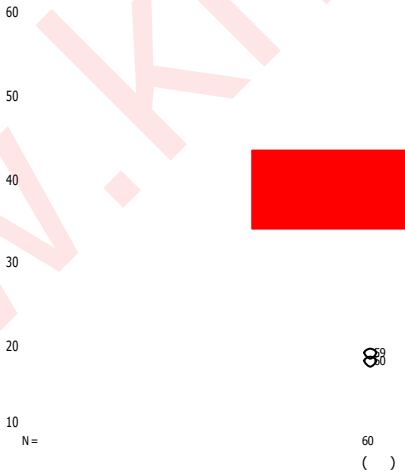


图 2.3.2　连续 60 天计算机销售量箱线图

从图 2.3.2 可看出，下横线之外有两个点，它们分别是第 59 和第 60 个观测值，原始数据中可查到，这两天的销售量分别为 19 台和 18 台，它们是离总体观测数据较远的离群点。

绘制茎叶图操作步骤：点击 File→点击 open→点击 Data→读取数据文件

ex2.3→点击 Analyze→点击 Descriptive Statistics-Explore→将“销售量”选入“Dependent List”栏→点击 Ok。输出结果如图 2.3.3 所示：

销售量（台） Stem-and-Leaf Plot

Frequency	Stem & Leaf
2.00 Extremes	(=<19)
1.00	2 . 2
5.00	2 . 56899
8.00	3 . 00223344
15.00	3 . 555666677788899
19.00	4 . 0001222222333344444
10.00	4 . 5555666789

Stem width: 10

Each leaf: 1 case(s)

图2.3.3 连续60天计算机销售量茎叶图

（3）描述统计分析操作步骤：点击 File→点击 open→点击 Data→读取数据文件 ex2.3→点击 Analyze→点击 Descriptive Statistics→Descriptives→将“销售量”选入 Variable 栏→点击 Options，选中 Mean（均值）、Std.（标准差）、Minimum（最小值）、Maximum（最大值）、Kurtosis（峰态系数）、Skwness（偏态系数），选中 Variable list（变量顺序排列显示输出结果）→点击 Continue→点击 Ok。输出结果如表 2.3.2 所示：

表2.3.2 连续60天计算机销售量的描述统计量

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std.	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std.Error	Statistic	Std.Error
XSHL 销售量（台）	60	18	49	37.88	7.140	-.879	.309	.422	.608
Valid N (listwise)	60								

表 2.3.2 显示，连续 60 天中计算机销售量的最小值为 18 台，最大值为 49 台，平均销售量为 37.88 台，标准差为 7.14 台；偏态系数为-0.879，表明计算机销售量为左偏分布，在连续 60 天中，销售量较高的天数占一定比例，具有一定的偏斜度；峰态系数为 0.608，表明计算机销售量为尖峰分布，销售量较高的天数具有一定的集中度。

2.4 解：总平均价格=采购金额÷采购量

四批西红柿的总平均价格为：

$$\bar{x}_h=\frac{\sum_{i=1}^{k}m_i}{\sum_{i=1}^{k}\frac{m_i}{x_i}}=\frac{100+90+70+40}{\frac{100}{2.5}+\frac{90}{3}+\frac{70}{3.5}+\frac{40}{4}}=\frac{300}{100}=3(\text{元/千克})$$

2.5 解：粮食总平均亩产量=粮食总产量÷播种面积

甲村粮食总平均亩产量为：

$$\bar{x}_{甲}=\frac{\sum_{i=1}^{k}m_i}{\sum_{i=1}^{k}\frac{m_i}{x_i}}=\frac{30000+27000+30000}{\frac{30000}{200}+\frac{27000}{300}+\frac{30000}{500}}=\frac{87000}{300}=290(\text{千克/亩})$$

乙村粮食总平均亩产量为：

$$\bar{x}_{乙}=\frac{\sum_{i=1}^{k}x_if_i}{\sum_{i=1}^{k}f_i}=\frac{200\times120+300\times90+500\times90}{120+90+90}=\frac{96000}{300}=320(\text{千克/亩})$$

计算结果表明：乙村的粮食总平均亩产高。原因是两村不同亩产水平的地块结构不同。亩产水平较高的丘陵在乙村总耕地面积中占 30%（90÷300），而在甲村只占 20%（60÷300）；相反，亩产水平较低的山地在乙村总耕地面积中只占 40%（120÷300），而在甲村则占 50%（150÷300）。高产地块比重大，低产地块比重小，导致乙村的粮食总平均亩产偏高。

2.6 解：各车间的平均合格率为：

$$\bar{x}_g=\sqrt[n]{x_1\times x_2\times\cdots\times x_n}=\sqrt[4]{90\%\times95\%\times93\%\times98\%}=93.95\%$$

2.7 解：各工序的平均合格率为：

$$\bar{x}_g=\sqrt[\sum_{i=1}^{n}f_i]{x_1^{f_1}\times x_2^{f_2}\times\cdots\times x_n^{f_n}}=\sqrt[10]{(95\%)^3\times(85\%)^2\times(90\%)^4\times(98\%)^1}=91.21\%$$

2.8 解：首先列出甲车间工人平均奖金及标准差计算表如表 2.8 所示：

表 2.8 甲车间工人平均奖金及标准差计算表

工人按奖金分组（元）	组中值（元） x_i	工人数（人） f_i	x_if_i	$(x_i-x)^2f_i$
60 以下	55	2	110	882
60～70	65	4	260	484
70～80	75	24	1800	24

80～90	85	8	680	648
90 以上	95	2	190	722
合　计	—	40	3040	2760

甲车间工人平均奖金：$\bar{x}_{甲}=\dfrac{\sum_{i=1}^{5}x_i f_i}{\sum_{i=1}^{5}f_i}=\dfrac{3040}{40}=76$（元）

甲车间工人奖金的标准差：$s_{甲}=\sqrt{\dfrac{\sum_{i=1}^{5}(x_i-\bar{x})^2 f_i}{\sum_{i=1}^{5}f_i-1}}=\sqrt{\dfrac{2760}{40-1}}=8.41$(元)

同理可计算得：$\bar{x}_{乙}=80.6$（元），$s_{乙}=9.93$(元)

由于两车间工人平均奖金额不同，因而不能直接用标准差判断奖金额的离散程度，必须进一步计算标准差系数：

$$V_{s甲}=\frac{s_{甲}}{\bar{x}_{甲}}\times 100\%=\frac{8.41}{76}\times 100\%=11.07\%$$

$$V_{s乙}=\frac{s_{乙}}{\bar{x}_{乙}}\times 100\%=\frac{9.93}{80.6}\times 100\%=12.32\%$$

因为$V_{s乙}>V_{s甲}$，所以乙车间工人奖金额的离散程度大，甲车间工人平均奖金额的代表性强。

第 3 章　练习题参考答案

3.1 解：

操作步骤：

Analyze → Descriptive Statistics → Explore，将 x 选入 Dependent List，在 Statistics 中选中 Descriptives

输出结果

表 3-1　Descriptives

			Statistic	Std. Error
x	Mean		180.0500	11.68331
	95% Confidence Interval for Mean	Lower Bound	156.8678	
		Upper Bound	203.2322	
	5% Trimmed Mean		172.3556	
	Median		156.0000	
	Variance		13649.967	
	Std. Deviation		116.83307	
	Minimum		24.00	
	Maximum		510.00	
	Range		486.00	
	Interquartile Range		130.25	
	Skewness		1.112	.241
	Kurtosis		.572	.478

结果分析：

由表 3-1 可知，该超市每位顾客的平均花费金额的点估计 180.0500，可信度为 95%的区间估计为（156.8678，203.2322）（保留两位小数）。

3.2 解：

（1）先估计每位顾客平时的平均花费金额和周末的平均花费金额的置信区间

操作步骤：

Analyze→Descriptive Statistics→Explore，将 x 选入 Dependent List，将顾客总体选入 Factor List，在 Statistics 中选中 Descriptives；

输出结果：

表 3-2-1　Descriptives

	顾客总体		Statistic	Std. Error

x	总体一	Mean	159.8000	17.52340
		95% Confidence Interval for Mean Lower Bound	124.5854	
		Upper Bound	195.0146	
		5% Trimmed Mean	150.3778	
		Median	116.0000	
		Variance	15353.469	
		Std. Deviation	123.90912	
		Minimum	24.00	
		Maximum	480.00	
		Range	456.00	
		Interquartile Range	139.75	
		Skewness	1.264	.337
		Kurtosis	.713	.662
	总体二	Mean	200.3000	15.09183
		95% Confidence Interval for Mean Lower Bound	169.9718	
		Upper Bound	230.6282	
		5% Trimmed Mean	192.8222	
		Median	177.5000	
		Variance	11388.173	
		Std. Deviation	106.71539	
		Minimum	55.00	
		Maximum	510.00	
		Range	455.00	
		Interquartile Range	109.25	
		Skewness	1.237	.337
		Kurtosis	1.059	.662

（2）再估计二者的差的置信区间

操作步骤：

Analyze → Compare means → Independent—Samples T test，将 x 选入 Test Variable(s)，将顾客总体选入 Grouping Variable(s)，在 Define Groups 中分别定义总体一和总体二。

输出结果：

表 3-2-2 Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
x	Equal variances assumed	1.312	.255	-1.751	98	.083	-40.50000	23.12645	-86.39369	5.39369
	Equal variances not assumed			-1.751	95.892	.083	-40.50000	23.12645	-86.40631	5.40631

结果分析：

表 3-2-1 是探索性分析的结果，可知每位顾客平时的平均花费金额的置信区间为（124.5854，195.0146），周末的平均花费金额的置信区间为（169.9718,230.6282）。表 3-2-2 是独立样本 t 检验的结果，Levene's Test for Equality of Variances 为方差检验的结果，F=1.312，其 P 值为 0.255<0.05，拒绝方差相等的原假设，认为两总体的方差不等，因此两总体的差的置信区间为（-86.40631，5.40631）

3.3 解：

操作步骤：

Analyze→Descriptive Statistics→Explore，将 x 选入 Dependent List，在 Statistics 中选中 Descriptives；

输出结果：

表 3-3 Descriptives

			Statistic	Std. Error
x	Mean		68.1875	1.75171
	95% Confidence Interval for Mean	Lower Bound	64.4538	
		Upper Bound	71.9212	

5% Trimmed Mean	68.1528	
Median	67.5000	
Variance	49.096	
Std. Deviation	7.00684	
Minimum	55.00	
Maximum	82.00	
Range	27.00	
Interquartile Range	6.50	
Skewness	.290	.564
Kurtosis	.482	1.091

结果分析：

表 3-3 是探索性分析的结果，由分析结果可知，该市成年男子体重置信度为95%的区间估计为（64.4538，71.9212）。

3.4 解：

由中心极限定理可知：$\frac{n\bar{X}-np}{\sqrt{np(1-p)}}$近似的服从 $N(0,1)$分布，于是有

$$P\left\{-z_{\alpha/2}<\frac{n\bar{X}-np}{\sqrt{np(1-p)}}<z_{\alpha/2}\right\}\approx 1-\alpha \qquad (3.4)$$

该市所有家庭中安装了宽带上网的家庭的百分比为 p 是 $(0,1)$ 分布的参数，$n=200$，$\bar{x}=192/200=0.96$，$1-\alpha=0.95$，$z_{\alpha/2}=1.96$，代入式（3.4）求得 p 的置信水平为 95%的置信区间为（0.92，0.98）

第 4 章　练习题参考答案

4.1 解：

操作步骤：

进行单样本 T 检验，Analyze→Compare means→One—Samples T test，将 x 选入 Test Variable(s)，在 Test Value 中输入 0.618。

输出结果：

表 4-1　One-Sample Test

	Test Value = 0.618					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
x	2.884	15	.011	.072813	.01901	.12662

结果分析：

根据表 4-1 中单样本 T 检验的结果，$P=0.011<0.05$，按显著性水平 $\alpha=0.05$ 无法拒绝 $x=0.618$ 的原假设，认为该厂生产的工艺品框架宽与长的比值符合黄金比率。

4.2 解：

H0：p<50%

H1：p>=50%

p=160/360=44.44%

$$Z=\frac{p-\pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}=\frac{44.44\%-50\%}{\sqrt{\frac{50\%(1-50\%)}{360}}}=-2.11$$

这是一个单侧检验，$z_{\alpha/2}$=1.645

|z|>|$z_{\alpha/2}$|,所以，拒绝原假设，这个专家的论断成立。

4.3 解：

H_0：$\bar{X}\geq X_0=30000$，即该厂家广告可信

H_1：$\bar{X}<X_0=30000$，即该厂家广告不可信

操作步骤：

进行单样本 T 检验，Analyze→Compare means→One—Samples T test，将 x 选入 Test Variable(s)，在 Test Value 中输入 0.618。

输出结果：

表 4-3-1 One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
x	12	30905.8333	1888.37332	545.12642

表 4-3-2 One-Sample Test

	Test Value = 30000					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
x	1.662	11	.125	905.83333	-293.9818	2105.6485

结果分析：

由表 4-3-1 可知，样本均值为 30905.8333，表 4-3-2 是单样本双侧 T 检验的结果，可知平均寿命 95%的置信区间为（-293.9818，2105.6485），根据平均寿命大于 0 以及双侧检验和单侧检验的关系，95%的单侧置信区间应为（0，2105.6485），该置信区间与显著性水平 0.05 的本题的左边检验问题相对应，而 X_0=30000 并不在置信区间（0，2105.6485）内，因此拒绝 H_0，认为该厂家的广告不可信。

4.4 解：

操作步骤：

进行独立样本 T 检验，Analyze→Compare means→Independent—Samples T test，将 x 选入 Test Variable(s)，将顾客总体选入 Grouping Variable(s)，在 Define Groups 中分别定义总体一和总体二。

输出结果：

表 4-4 Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
x Equal variances assumed	1.312	.255	-1.751	98	.083	-40.50000	23.12645	-86.39369	5.39369
Equal variances not assumed			-1.751	95.892	.083	-40.50000	23.12645	-86.40631	5.40631

结果分析：

表 4-4 是单样本 T 检验的结果。首先根据该表方差齐性检验的结果，$F=1.312$，$Sig.=0.255$，即 $P>0.05$，无法拒绝方差齐性的原假设，认为两个总体的方差相等。在方差齐性的假设下，独立样本 t 检验结果为：$t=-1.751$，$Sig.(2-tailed)=0.083$，$P>0.05$，无法拒绝两个总体的平均花费金额相同的原假设，认为该超市每位顾客平时（周一至周五）的平均花费金额与周末（周六和周日）的平均花费金额是相同的$(\alpha=0.05)$。

4.5 解：

操作步骤：

进行配对样本 T 检验，Analyze→Compare means→Paired—Samples T test，将 X1 选入 Test Variable 1，将 X2 选入 Variable 2。

输出结果：

表 4-5-1 Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	促销前销售量	34.9000	10	4.81779	1.52352
	促销后销售量	38.6000	10	4.94862	1.56489

表 4-5-2 Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	促销前销售量 & 促销后销售量	10	.954	.000

表 4-5-3 Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1 促销前销售量 - 促销后销售量	-3.70000	1.49443	.47258	-4.76905	-2.63095	-7.829	9	.000

结果分析：

由表 4-5-1 可知，促销前的平均销售量为 34.9000，促销后的平均销售量为 38.6000。由表 4-5-2 可知促销前后的相关系数为 0.954，P=0.000<0.05，两者相关性显著。根据表 4-5-3 中配对样本 t 检验的结果，得到 $t=-7.829$，$Sig.(2-tailed)=0.000$，$P<0.05$，在 $\alpha=0.05$ 水平下，拒绝原假设，认为促销后销售量有明显的提高。

第 5 章　练习题参考答案

5.1　（数据文件 ex5.1）

（1）、$y_{ij}=\mu+\alpha_i+\varepsilon_{ij}, i=1,2,\cdots,4, j=1,2,\cdots,6$；

（2）、方差分析表如下所示：

表 1

ANOVA

时间

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.540	3	.847	12.700	.000
Within Groups	1.333	20	.067		
Total	3.873	23			

由于其 p-值小于 0.05，因此说明不同的安眠药对安眠时间有显著影响。

（3）、采用了三种方法 LSD、Bonferroni、Turkey 比较前三种与第 4 种药物的比较，结果如下：

表 2

Multiple Comparisons

Dependent Variable:时间

			Mean Difference			95% Confidence Interval	
	(I) 编号	(J) 编号	(I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
Tukey HSD	4	1	-.04000	.15912	.994	-.4874	.4074
		2	-.54000*	.15912	.015	-.9874	-.0926
		3	-.74000*	.15912	.001	-1.1874	-.2926
LSD	4	1	-.04000	.15912	.804	-.3730	.2930
		2	-.54000*	.15912	.003	-.8730	-.2070
		3	-.74000*	.15912	.000	-1.0730	-.4070
Bonferroni	4	1	-.04000	.15912	1.000	-.5084	.4284
		2	-.54000*	.15912	.018	-1.0084	-.0716
		3	-.74000*	.15912	.001	-1.2084	-.2716

前三种与第四种的比较结果：如果显著性水平为 $\alpha=0.05$，在三种方法中，第4种与第1种没有显著差异，与第二种和第三种有显著差异。

5.2 （数据文件ex5.2）

方差分析表如下所示：

表 3

ANOVA

寿命

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	892.111	2	446.056	94.237	.000
Within Groups	71.000	15	4.733		
Total	963.111	17			

由于其 p-值小于 0.05，因此说明三种品牌的电池的平均寿命有显著影响。用LSD法检验方法给出多重比较的结果如下：

表 4

Multiple Comparisons

寿命

LSD

(I) 品牌	(J) 品牌	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	2	-10.000*	1.256	.000	-12.68	-7.32
	3	-17.167*	1.256	.000	-19.84	-14.49
2	1	10.000*	1.256	.000	7.32	12.68
	3	-7.167*	1.256	.000	-9.84	-4.49
3	1	17.167*	1.256	.000	14.49	19.84
	2	7.167*	1.256	.000	4.49	9.84

*. The mean difference is significant at the 0.05 level.

从检验统计量p-值可以看出品牌A与品牌B、品牌A与品牌C、品牌B与品牌C之间都有显著差异。

5.3 （数据文件ex5.3）

双因素方差分析的结果如下：

表 5

Tests of Between-Subjects Effects

Dependent Variable:销售数据

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1041.778[a]	4	260.444	16.863	.009
	18860.444	1	18860.444	1221.180	.000
商场	50.889	2	25.444	1.647	.301
促销方式	990.889	2	495.444	32.079	.003
Error	61.778	4	15.444		
Total	19964.000	9			
Corrected Total	1103.556	8			

从表中看到输出结果没有交互作用。从p-值可以看出，在0.05的显著性水平下，促销方式的主效应显著，但商场的主效应不显著。因此说明如果仅考虑可加模型，促销方式对销售量有显著影响，但商场对销售量没有显著影响。

5.4 （数据文件ex5.4）

双因素方差分析的结果如下：

表 6

Tests of Between-Subjects Effects

Dependent Variable:销售量

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3157.333[a]	8	394.667	12.611	.000
	33496.333	1	33496.333	1070.297	.000
广告方案	2277.556	2	1138.778	36.387	.000
广告媒体	643.556	2	321.778	10.282	.001
广告方案 * 广告媒体	236.222	4	59.056	1.887	.157
Error	563.333	18	31.296		
Total	37217.000	27			
Corrected Total	3720.667	26			

a. R Squared = .849 (Adjusted R Squared = .781)

从表中看到此时有交互作用。从p-值可以看出，在0.05的显著性水平下，广告方案和广告媒体均显著。但是交互效应的p-值明显大于0.05，因此交互效应不

显著。从而我们得到结论：广告方案和广告媒体对产品的销售量有显著影响，但不同广告方案在不同媒体上所获得的销售量没有显著差异（即交互效应不显著）。

5.5 （数据文件ex5.5）

先考虑有交互效应的双因素方差分析模型，方差分析表如表7所示，易知因素A与因素B的交互效应不显著。事实上我们也可以做交互作用的图形分析，分析也表明因素A与因素B的交互作用不显著。

表 7

Tests of Between-Subjects Effects

Dependent Variable:生存时间

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.204[a]	11	.200	9.010	.000
	11.030	1	11.030	495.919	.000
因素A	1.033	2	.517	23.222	.000
因素B	.921	3	.307	13.806	.000
因素A * 因素B	.250	6	.042	1.874	.112
Error	.801	36	.022		
Total	14.035	48			
Corrected Total	3.005	47			

a. R Squared = .734 (Adjusted R Squared = .652)

因此只需要考虑因素A与因素B的主效应即可。接着我们对医疗急救措施进行多重比较分析，结果如下：

表 8

Multiple Comparisons

生存时间

LSD

(I) 因素B	(J) 因素B	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	2	-.3625*	.06089	.000	-.4860	-.2390
	3	-.0783	.06089	.206	-.2018	.0451
	4	-.2200*	.06089	.001	-.3435	-.0965
2	1	.3625*	.06089	.000	.2390	.4860
	3	.2842*	.06089	.000	.1607	.4076
	4	.1425*	.06089	.025	.0190	.2660
3	1	.0783	.06089	.206	-.0451	.2018

	2	-.2842*	.06089	.000	-.4076	-.1607
	4	-.1417*	.06089	.026	-.2651	-.0182
4	1	.2200*	.06089	.001	.0965	.3435
	2	-.1425*	.06089	.025	-.2660	-.0190
	3	.1417*	.06089	.026	.0182	.2651

Based on observed means.

The error term is Mean Square(Error) = .022.

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

从表中我们可以看到，若给定显著性水平$\alpha = 0.05$，四种急救措施只有第2种和第4种与其他方法全都有显著差异。另一方面，我们可以运用Option选项中Estimated Maginal Means 来计算边际均值的估计值。结果如下表所示：

表 9

因素B

Dependent Variable:生存时间

			95% Confidence Interval	
因素B	Mean	Std. Error	Lower Bound	Upper Bound
1	.314	.043	.227	.401
2	.677	.043	.589	.764
3	.392	.043	.305	.480
4	.534	.043	.447	.621

从表中我们看到，第2种和第4种急救措施的均值要偏大且第2种措施的均值为0.677大于第4种方法。因此第2种急救措施疗效较好。

第 6 章　练习题参考答案

6.1 解：

（1）对因变量和解释变量进行相关性分析， Analyze→Correlate→Bivariate Correlations，将 y、x1、x2、x3、x4 和 x5 选入 Variables，选中 Person，Two-tailed 和 Flag significant correlations。

表 6-1-1 是相关分析的结果：民航客运量 y 与国民收入 x_1、消费额 x_2、民航航线里程 x_4 和来华旅游入境人数 x_5 相关系数较高，相关性显著，Sig. (1-tailed)=0.000<0.01，而民航客运量 y 与铁路客运量 x_3 相关系数较低，仅为 0.266，Sig. (1-tailed)=0.160>0.01，相关性不显著。

表 6-1-1　Correlations

		民航客运量	国民收入(亿元)	消费额(亿元)	铁路客运量 (万人)	民航航线里程 (万公里)	来华旅游入境人数(万人)
民航客运量	Pearson Correlation	1	.961**	.985**	.266	.987**	.924**
	Sig. (2-tailed)		.000	.000	.320	.000	.000
	N	16	16	16	16	16	16
国民收入 (亿元)	Pearson Correlation	.961**	1	.972**	.231	.960**	.878**
	Sig. (2-tailed)	.000		.000	.389	.000	.000
	N	16	16	16	16	16	16
消费额 (亿元)	Pearson Correlation	.985**	.972**	1	.284	.978**	.942**
	Sig. (2-tailed)	.000	.000		.286	.000	.000
	N	16	16	16	16	16	16
铁路客运量 (万人)	Pearson Correlation	.266	.231	.284	1	.232	.358
	Sig. (2-tailed)	.320	.389	.286		.388	.173
	N	16	16	16	16	16	16
民航航线里程(万公里)	Pearson Correlation	.987**	.960**	.978**	.232	1	.882**
	Sig. (2-tailed)	.000	.000	.000	.388		.000
	N	16	16	16	16	16	16

来华旅游入境人数(万人)	Pearson Correlation	.924**	.878**	.942**	.358	.882**	1
	Sig. (2-tailed)	.000	.000	.000	.173	.000	
	N	16	16	16	16	16	16

**. Correlation is significant at the 0.01 level (2-tailed).

（2）将所有解释变量引入建立多元线性回归模型，进行全变量回归。

Analyze→Regression→Linear Regression，将 y 选入 Dependent，将 x1、x2、x3、x4 和 X5 选入 Independent(s)，Method 选择 Enter。

表 6-1-2 是回归模型统计量：复相关系数 R 为 0.994，解释变量和因变量的相关性很强；可决系数 R^2 为 0.988，用自变量可以解释因变量变异的程度为 98.8%，调整后的可决系数为 0.982，模型整体的拟合效果很好。

表 6-1-3 是回归模型的方差分析表，F 值为 162.787，显著性概率是 0.000，从而拒绝原假设，认为解释变量和因变量之间的线性关系非常显著，可以建立线性模型。

表 6-1-4 是回归模型的回归系数表，可以得出建立的回归模型为：

$$\hat{y} = -401.224 + 0.014x_1 - 0.021x_2 + 5.810 \times 10^{-5} x_3 + 30.440x_4 + 0.200x_5$$

可以发现，仅有民航航线里程数的回归系数显著性检验（t 检验）的 p 值分小于 0.05，认为其显著，其他变量不显著，说明这些变量之间存在共线性。

表 6-1-2 Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.994[a]	.988	.982	129.62078

a. Predictors: (Constant), 来华旅游入境人数 (万人), 铁路客运量 (万人), 国民收入 (亿元), 民航航线里程 (万公里), 消费额 (亿元)

表 6-1-3 ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.368E7	5	2735071.255	162.787	.000[a]
	Residual	168015.477	10	16801.548		
	Total	1.384E7	15			

a. Predictors: (Constant), 来华旅游入境人数 (万人), 铁路客运量 (万人), 国民收入 (亿元), 民航航线里程 (万公里), 消费额 (亿元)

表 6-1-3 ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.368E7	5	2735071.255	162.787	.000[a]
	Residual	168015.477	10	16801.548		
	Total	1.384E7	15			

a. Predictors: (Constant), 来华旅游入境人数 (万人), 铁路客运量 (万人), 国民收入 (亿元), 民航航线里程 (万公里), 消费额 (亿元)

b. Dependent Variable: 民航客运量

表 6-1-4 Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-401.224	143.142		-2.803	.019
	国民收入 (亿元)	.014	.024	.104	.601	.561
	消费额 (亿元)	-.021	.087	-.092	-.241	.814
	铁路客运量 (万人)	5.810E-5	.001	.002	.040	.969
	民航航线里程 (万公里)	30.440	8.298	.748	3.668	.004
	来华旅游入境人数 (万人)	.200	.111	.260	1.798	.102

a. Dependent Variable: 民航客运量

（3）进行逐步回归，Analyze→Regression→Linear Regression，将 y 选入 Dependent，将 x1、x2、x3、x4 和 x5 选入 Independent(s)，Method 选择 Stepwise。

结果见表 6-1-5、表 6-1-6、表 6-1-7 和表 6-1-8。

由结果可知，最终建立两个模型，

模型一：$\hat{y} = -382.508 + 40.147x_4$，$R^2 = 0.974$

模型二：$\hat{y} = -401.074 + 31.471x_4 + 0.187x_5$，$R^2 = 0.987$

两个模型的拟合效果都很好，可决系数 R^2 大于 0.970，且模型中因变量的系数通过了显著性检验。表 6-1-8 的结果显示了排除在模型之外的变量。

表 6-1-5 Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate

1	.987[a]	.974	.973	159.25684
2	.994[b]	.987	.985	115.99148

a. Predictors: (Constant), 民航航线里程 (万公里)

b. Predictors: (Constant), 民航航线里程 (万公里), 来华旅游入境人数 (万人)

表 6-1-6 ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.349E7	1	1.349E7	531.815	.000[a]
	Residual	355078.396	14	25362.743		
	Total	1.384E7	15			
2	Regression	1.367E7	2	6834234.721	507.970	.000[b]
	Residual	174902.308	13	13454.024		
	Total	1.384E7	15			

a. Predictors: (Constant), 民航航线里程 (万公里)

b. Predictors: (Constant), 民航航线里程 (万公里), 来华旅游入境人数 (万人)

c. Dependent Variable: 民航客运量

表 6-1-7 Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-382.508	77.808		-4.916	.000
	民航航线里程 (万公里)	40.147	1.741	.987	23.061	.000
2	(Constant)	-410.074	57.168		-7.173	.000
	民航航线里程 (万公里)	31.471	2.688	.774	11.706	.000
	来华旅游入境人数 (万人)	.187	.051	.242	3.660	.003

a. Dependent Variable: 民航客运量

表 6-1-8 Excluded Variables[c]

Model	Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
					Tolerance

1	国民收入 (亿元)	.167[a]	1.105	.289	.293	.079
	消费额 (亿元)	.463[a]	2.741	.017	.605	.044
	铁路客运量 (万人)	.039[a]	.881	.394	.237	.946
	来华旅游入境人数 (万人)	.242[a]	3.660	.003	.712	.222
2	国民收入 (亿元)	.076[b]	.647	.530	.184	.074
	消费额 (亿元)	.062[b]	.236	.817	.068	.015
	铁路客运量 (万人)	.000[b]	-.010	.992	-.003	.840

a. Predictors in the Model: (Constant), 民航航线里程 (万公里)

b. Predictors in the Model: (Constant), 民航航线里程 (万公里), 来华旅游入境人数 (万人)

c. Dependent Variable: 民航客运量

6.2 解：

(1) 先做散点图

Graphs→Scatter/Dot→Simple Scatterplot，将 y 选入 Y Axis，将 x 选入 X Axis；

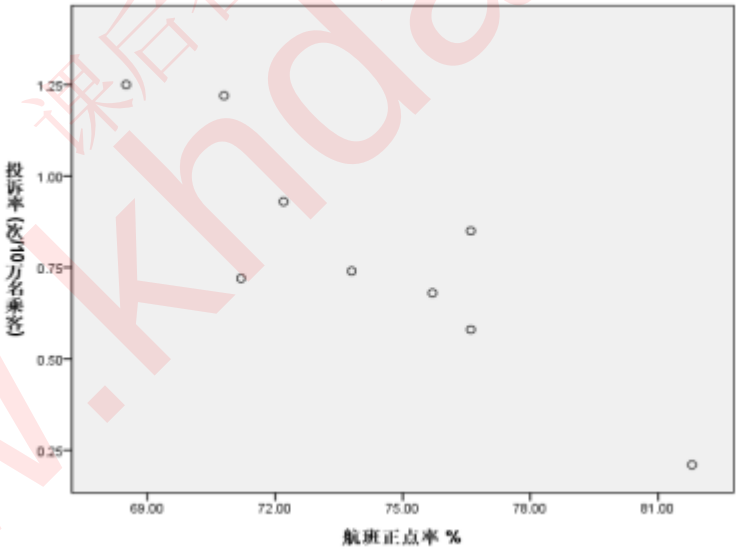


图 6-2-1

图 6-2-1 显示的是航班正点率和投诉率的散点图，由图形可以看出两者大致呈线性关系。因此以航班正点率为自变量，投诉率为因变量建立线性回归模型。

（2）计算相关系数

Analyze→Correlate→Bivariate Correlations，将 y 和 x 选入 Variables，选中 Person，Two-tailed 和 Flag significant correlations。

表 6-2-1 Correlations

		x 航班正点率 %	y 投诉率 (次/10 万名乘客)
x 航班正点率 %	Pearson Correlation	1	-.883**
	Sig. (2-tailed)		.002
	N	9	9
y 投诉率 (次/10 万名乘客)	Pearson Correlation	-.883**	1
	Sig. (2-tailed)	.002	
	N	9	9

**. Correlation is significant at the 0.01 level (2-tailed).

表 6-2-1 是自变量和因变量的相关分析表，两者的 Pearson 相关系数为 -0.883，显著性概率为 0.002<0.01，线性相关性显著。

（3）进行一元线性回归

Analyze → Regression→Linear Regression，将 y 选入 Dependent，将 x 选入 Independent(s)。

表 6-2-2 ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.638	1	.638	24.674	.002[a]
	Residual	.181	7	.026		
	Total	.819	8			

a. Predictors: (Constant), x 航班正点率 %

b. Dependent Variable: y 投诉率 (次/10 万名乘客)

表 6-2-3 Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6.018	1.052		5.719	.001
	x 航班正点率 %	-.070	.014	-.883	-4.967	.002

a. Dependent Variable: y 投诉率 (次/10 万名乘客)

表 6-2-2 是回归模型的方差分析表，F 值为 24.674，显著性概率是 0.002<0.05，从而拒绝原假设，认为解释变量和因变量之间的线性关系非常显著，可以建立线性模型。

表 6-1-4 是回归模型的回归系数表，回归系数的显著性检验统计量 t 统计量

的值为-4.967，对应的显著性水平 Sig.=0.002<0.05，认为方程显著，因此可以得出建立的回归模型为：

$$\hat{y} = 6.018 - 0.070x$$

（4）预测

在 X 列中输入 80，Analyze→Regression→Linear Regression，在 save 选项中 Predicted Values 下选中 Unstandardized，在 Predicted Intervals 同时选中 Mean 和 Individual。数据文件中将输出非标准化的预测值及均值和个体值的预测区间。如果航班正点率为 80%，用回归方程预测的投诉率为 0.38468，均值 95%的预测区间为（0.15071，0.61865），个体值 95%的预测区间为（-0.06180，0.83116），由于投诉率>0，所以个体值 95%预测区间应为（0，0.83116）。因此，如果航班正点率为 80%，每 10 万名乘客投诉的次数为 38468 次，均值 95%的预测区间为（15071， 61865），个体值 95%的预测区间为（0， 83116）。

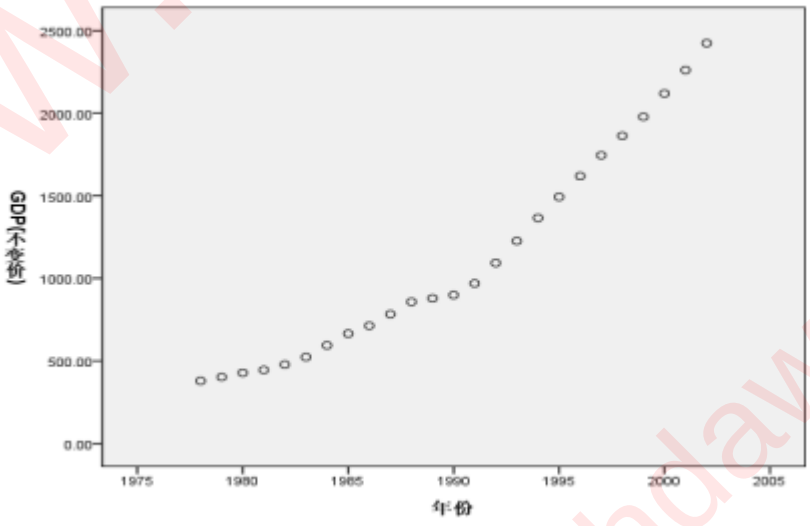
6.3 解：

方法一：曲线拟合

(1) 先做散点图

Graphs→Scatter/Dot→Simple Scatterplot，将 y 选入 Y Axis，将 t 选入 X Axis；

图6-3-1显示了不变价人均GDP（y）与时间（t）的散点图，不变价人均GDP随时间的推移而逐渐增加，当时间到达后期时，不变价人均GDP增长的幅度更加明显，因此用线性回归模型拟合y和t的关系是不恰当的，考虑用曲线模型拟合两者的关系。



（2）进行曲线拟合

Analyze → Regression→Curve Estimation，将 y 选入 Dependent(s)，在 Independent 中选中 Time，在 Models 下选中 Quadratic、Compound、Growth、Cubic 和 Exponential。

根据散点图，选择二次曲线（Quadratic）、混合曲线（Compound）、生长曲线（Growth）、三次曲线（Cubic）和指数曲线表（Exponential）等五种曲线模型进行回归，比较其拟合效果。表6-3-2是五种模型的曲线估计结果，从拟合优度（R Square，即R^2）来看，各个模型的拟合优度均在0.996以上，方差分析的显著性概率P均为0.000，模型的拟合效果均很好。图6-3-2显示的是各种曲线拟合图，五种曲线与原始数据的拟合程度均很好。

因此，五种曲线模型都可以用来建立不变价人均GDP（y）与时间（t）的回归方程：

二次曲线模型（Quadratic），$\hat{y}=387.490+4.571t+3.082t^2$

三次曲线模型（Cubic），$\hat{y}=368.937+12.374t+2.346t^2+0.019t^3$

混合曲线模型（Compound），$\hat{y}=336.814\times(1.084)^t$

生长曲线（Growth），$\hat{y}=e^{(5.820+0.080t)}$

指数曲线表（Exponential），$\hat{y}=336.814\times e^{(0.080t)}$

以上模型时间 t 均从 1 开始

表 6-3-1 Model Summary and Parameter Estimates

Dependent Variable:GDP(不变价)

Equation	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Quadratic	.997	3922.449	2	22	.000	387.490	4.571	3.082	
Cubic	.997	2566.953	3	21	.000	368.937	12.379	2.346	.019
Compound	.996	5959.205	1	23	.000	336.814	1.084		
Growth	.996	5959.205	1	23	.000	5.820	.080		
Exponential	.996	5959.205	1	23	.000	336.814	.080		

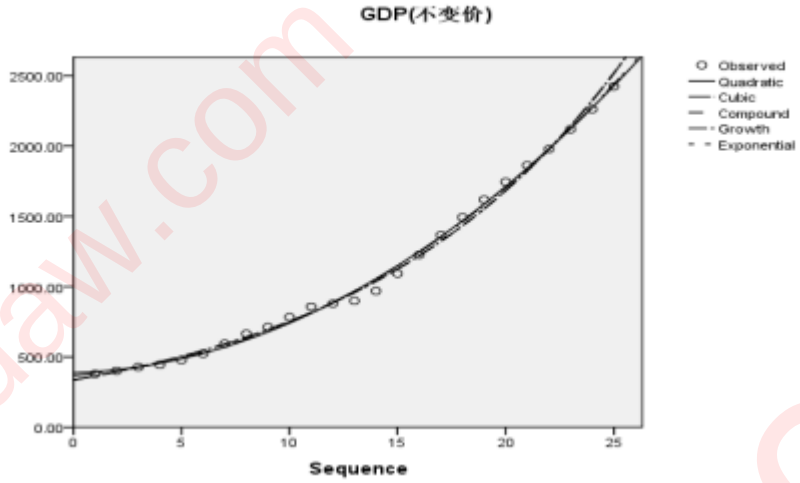


图 6-3-2

方法二：曲线直线化

操作步骤：

(1) 将自变量进行对数转化。

Transform→Compute Variable，在 Target Variable 中输入 lny，在 Numeric Expression 中输入 ln(y)，将不变价人均 GDP（y）对数化得到 lny

(2) 做散点图

Graphs→Scatter/Dot→Simple Scatterplot，将 lny 选入 Y Axis，将 t 选入 X Axis

作出 lny 与时间 t 的散点图，结果见图 6-3-3，发现 lny 与时间 t 有明显的线性相关关系

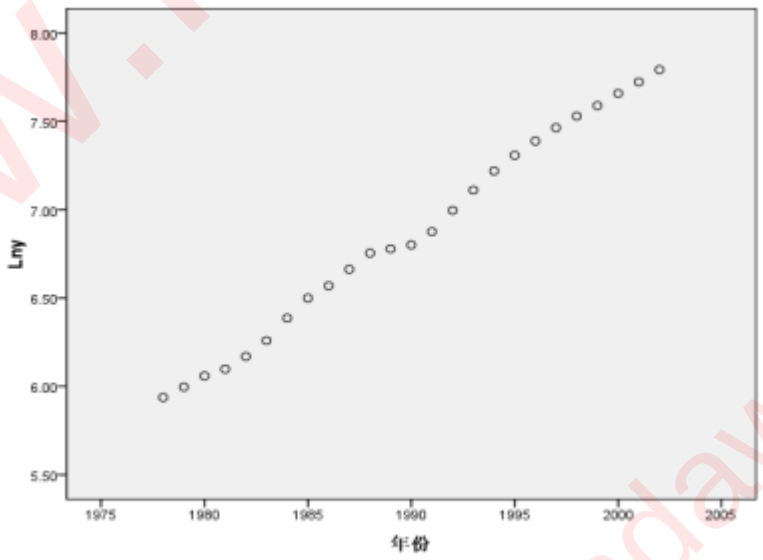


图 6-3-3

(3) 进行线性回归

Analyze→Regression→Curve Estimation，将 y 选入 Dependent(s)，在 Independent 中选中 Time，在 Models 下选中线性模型 Linear，选中 Display ANOVA table。

以 lny 为因变量，时间 t 为解释变量建立线性回归方程，结果见表 6-3-2、表 6-3-3 和表 6-3-4。

表 6-3-2 模型的拟合结果，拟合优度 R^2 为 0.996，拟合效果相当好。表 6-3-3 是回归模型的方差分析表，F 值为 5959.205，显著性概率是 0.000<0.05，认为 lny 与时间 t 的线性关系非常显著，可以建立线性模型。

表 6-3-4 是线性回归模型的回归系数表，回归系数的显著性检验统计量 t 统计量的值为 77.196，对应的显著性水平 Sig.=0.000<0.05，认为方程显著，因此可以得出建立的回归模型为：

$$\hat{y}=5.820+0.080t，\ t从\ 1\ 开始$$

表 6-3-2 Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.998	.996	.996	.038

表 6-3-3 ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	8.415	1	8.415	5959.205	.000
Residual	.032	23	.001		
Total	8.447	24			

表 6-3-4 Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Case Sequence	.080	.001	.998	77.196	.000
(Constant)	5.820	.015		375.605	.000

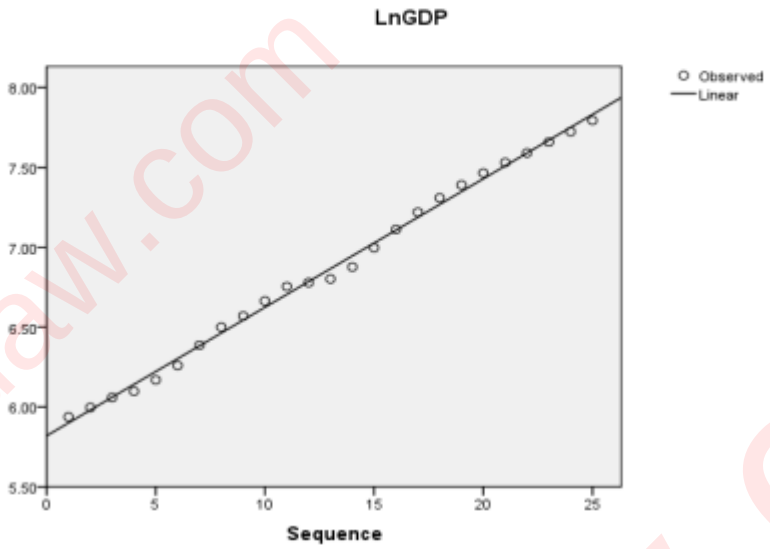


图6-3-4

结果分析：

将不变价人均 GDP（y）对数化得到 lny，作出 lny 与时间 t 的散点图，结果见图 6-3-3，发现 lny 与时间 t 有明显的线性相关关系，因此，

6.4 解：

用SPSS建立Logistic回归模型

操作步骤：

在变量Y列的最后一行输入30，Analyze→Regression→Binary Logistic，将Y选入Dependent，将X选入Covariates，在Save选项中Predicted Values下选中Probabilities。

输出结果：

Block 0: Beginning Block

表 6-4-1 Classification Table[a,b]

Observed			Predicted		
			观点		
			0	1	Percentage Correct
Step 0	观点	0	91	0	100.0
		1	89	0	.0
	Overall Percentage				50.6

a. Constant is included in the model.

b. The cut value is .500

表 6-4-2 Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.022	.149	.022	1	.881	.978

表 6-4-3 Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	X	61.615	1	.000
	Overall Statistics		61.615	1	.000

Block 1: Method = Enter

表 6-4-4 Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	72.088	1	.000
	Block	72.088	1	.000
	Model	72.088	1	.000

表 6-4-5 Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	177.423[a]	.330	.440

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

表 6-4-6 Classification Table[a]

Observed			Predicted		
			观点		
			0	1	Percentage Correct
Step 1	观点	0	65	26	71.4
		1	12	77	86.5
	Overall Percentage				78.9

a. The cut value is .500

表 6-4-7 Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	X	-.123	.018	44.560	1	.000	.884

	Constant	3.634	.542	44.895	1	.000	37.872

a. Variable(s) entered on step 1: X.

结果分析：

表 6-4-1 输出的是仅含有常数项时计算的预测分类结果，总的准确预测率为 50.6%。

表 6-4-2 和表 6-4-3 输出的是模型的初始迭代结果，模型仅包含常数项，自变量年龄 x 并未包含在模型中。

表 6-4-4、表 6-4-5、表 6-4-6 和表 6-4-7 显示的是引入自变量的模型结果。表 6-4-4 显示 Step、Block、Model 的卡方值相同，均为 72.088，P=0.000，认为模型显著。但由表 6-4-5 显示的 Cox & Snell R Square 和 Nagelkerke R Square 均较小，表示模型的拟合效果不佳，自变量不能很好地解释因变量的变异情况。

表 6-4-6 是引入自变量后重新拟合的回归模型的分类表格，新的预测正确率为 78.9%。表 6-4-7 可知显著性概率为 0.000<0.05，回归模型显著，因此 Binary Logistics 模型为：

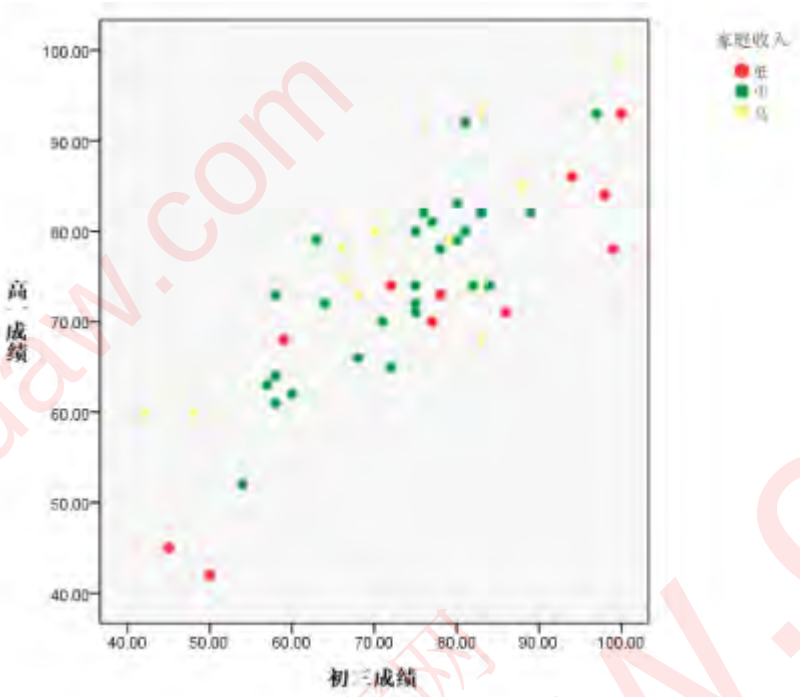
$$Ln(p/1-p) = 3.634 - 0.123X$$

若年龄为 30 岁，运用该模型预测的结果为 0.48681，则年龄为 30 岁的人对该影片持肯定观点的可能性为 48.681%。

6.5 解：

1、观察散点图

Graphs→Scatter/Dot→Simple Scatterplot，将 y 选入 Y Axis，将 x 选入 X Axis，将 d 选入 Set Markers by，调整颜色，结果见下图，发现不同家庭收入中高一成绩和初三成绩呈现大致的线性递增关系，但其截距可能存在不同，因此，考虑设置虚拟变量进行回归。



2、建立回归模型

将家庭收入变量设置为虚拟变量 d1 和 d2，$d_1=0,d_2=0$表示低收入，$d_1=1,d_2=0$表示中等收入，$d_1=0,d_2=1$表示高收入，统计模型设定为

$$y=\beta_0+\beta_1 x+\beta_2 d_1+\beta_3 d_2+\varepsilon$$

（1）设置虚拟变量

Transform→Recode into Different Variables，将家庭收入 d 选入，在 Output Variable 中输入 d1, 在 Old and New Values 中设置当 d=2 时，d1=1, 其他情况时 d1=0, 运行；然后再在 Output Variable 中输入 d2，在 Old and New Values 中设置当 d=3 时，d2=1, 其他情况时 d2=0, 运行

（2）虚拟变量回归

用 SPSS 打开数据文件，选择 Analyze→General Linear Model→Univariate，将因变量高一成绩 y 选入 Dependent Variable 中，把定量自变量初三成绩 x 选入 Covariate 中，把虚拟变量 d_1 和 d_2 选入 Fixed Factor 中，在 Options 中选择 Parameter Estimates，点击 Model，在 Specify Model 中选 Custom，再把定量自变量和虚拟变量选入右边，在 Building Term 中选 Main effect，然后点 Continue 回到主对话框，在 Options 中的 Display 中选择 Parameter estimates，点 Continue→OK 即可得参数估计值表表 6-5-1。

表 6-5-1　Parameter Estimates

Dependent Variable:高一成绩

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	35.095	5.129	6.842	.000	24.770	45.421
[d1=.00]	-6.387	2.265	-2.820	.007	-10.947	-1.828
[d1=1.00]	0[a]	.	.	.	.	.
[d2=.00]	-11.066	2.641	-4.190	.000	-16.382	-5.750
[d2=1.00]	0[a]	.	.	.	.	.
x	.688	.063	10.925	.000	.561	.814

a. This parameter is set to zero because it is redundant.

得到回归方程如下所示：

低收入家庭：　$y=35.095+0.688x-6.387-11.066=17.642+0.688x$

中等收入家庭：$y=35.095+0.688x-11.066=24.029+0.688x$

高收入家庭：$y=35.095+0.688x-6.387=28.708+0.688x$

（3）比较一：直接引入家庭收入变量进行广义线性回归

选择 Analyze→General Linear Model→Univariate，将因变量高一成绩 y 选入 Dependent Variable 中，把定量自变量初三成绩 x 选入 Covariate 中，把家庭收入变量 d 选入 Fixed Factor 中，在 Options 中选择 Parameter Estimates，点击 Model，在 Specify Model 中选 Custom，再把定量和自变量选入右边，在 Building Term 中选 Main effect，然后点 Continue 回到主对话框，在 Options 中的 Display 中选择 Parameter estimates，点 Continue→OK 即可得参数估计值表 6-5-2。

表6-5-2　Parameter Estimates

Dependent Variable:高一成绩

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	28.708	4.904	5.854	.000	18.837	38.579

[d=1.00]	-11.066	2.641	-4.190	.000	-16.382	-5.750
[d=2.00]	-4.679	2.176	-2.150	.037	-9.059	-.299
[d=3.00]	0[a]	.	.	.	.	.
x	.688	.063	10.925	.000	.561	.814

a. This parameter is set to zero because it is redundant.

低收入家庭 $d=1$： $y=28.708-11.066+0.688x=17.642+0.688x$

中等收入家庭 $d=2$： $y=28.708-4.679+0.688x=24.029+0.688x$

高收入家庭 $d=3$： $y=28.708+0.688x$

模型的参数估计结果与引入两个虚拟变量的结果完全一致。

（4）比较二：引入虚拟变量进行线性回归

将虚拟变量 d1 和 d2 作为解释变量引入，与初三成绩共同作为解释变量建立线性回归模型，Analyze→Regression→Linear Regression，将 y 选入 Dependent，将 x、d1 和 d2 选入 Independent(s)，Method 选择 Enter。结果见表 6-3-3。

表 6-5-3 Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.642	5.261		3.354	.002
	初三成绩	.688	.063	.840	10.925	.000
	是否中等收入	6.387	2.265	.273	2.820	.007
	是否高等收入	11.066	2.641	.405	4.190	.000

a. Dependent Variable: 高一成绩

低收入家庭： $y=17.642+0.688x$

中等收入家庭： $y=17.642+0.688x+6.387=24.029+0.688x$

高收入家庭： $y=17.642+0.688x+11.066=28.708+0.688x$

将虚拟变量d1 和d2 作为解释变量进行线性回归的结果和设置虚拟变量的结果完全一样，低收入和中等收入家庭差距为 6.387，中等收入和高收入家庭的差距为 11.066，收入等级越高，差距越大。

第 7 章　练习题参考答案

1、我们首先计算其 ACF 和 PACF，

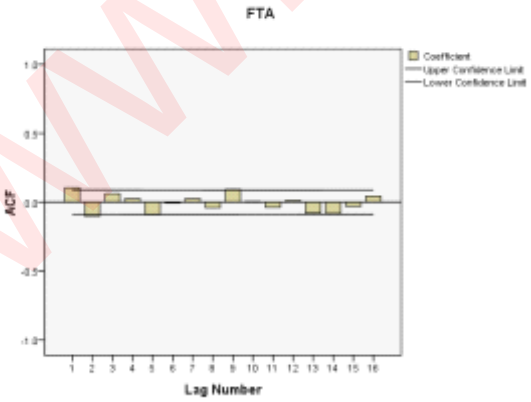
表 1、序列的自相关函数

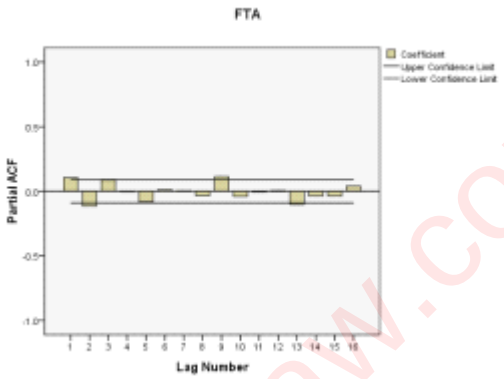
Autocorrelations

Series:FTA

			Box-Ljung Statistic		
Lag	Autocorrelation	Std. Error[a]	Value	df	Sig.[b]
1	.105	.045	5.465	1	.019
2	-.101	.045	10.538	2	.005
3	.061	.045	12.398	3	.006
4	.025	.045	12.718	4	.013
5	-.091	.045	16.851	5	.005
6	-.008	.045	16.880	6	.010
7	.025	.045	17.192	7	.016
8	-.040	.045	18.015	8	.021
9	.093	.045	22.390	9	.008
10	.006	.045	22.406	10	.013
11	-.037	.044	23.087	11	.017
12	.012	.044	23.157	12	.026
13	-.079	.044	26.327	13	.015
14	-.081	.044	29.698	14	.008
15	-.030	.044	30.157	15	.011
16	.045	.044	31.177	16	.013

序列的自相关函数和偏自相关函数图为：





从图中看出；样本序列数据的自相关系数在0附近摆动，且逐渐衰减，所以该时间序列基本是平稳的。从图中可以看出ACF和PACF都有明显的指数衰减特征，因此我们可以认为该数据服从ARMA模型。根据其特征我们考虑了ARMA（1,1）、ARMA（1,2）、ARMA（2,1）、ARMA（2,2）几个模型，并根据BIC准则选择了ARMA（1,1）。

参数估计结果如下：

ARIMA Model Parameters

				Estimate	SE	t	Sig.
FTA-Model_1	FTA	No Transformation	Constant	1.210	.296	4.089	.000
			AR Lag 1	-.464	.206	-2.253	.025
			MA Lag 1	-.604	.185	-3.264	.001

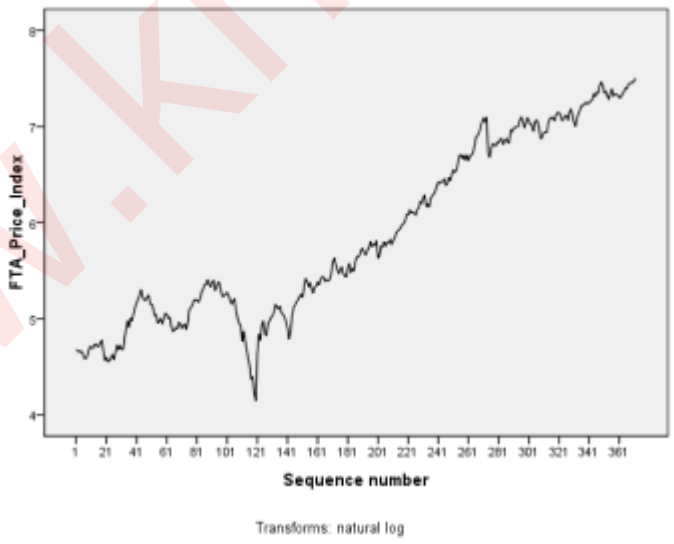
此时 BIC=3.617, $R^2 = 0.025$ 。接着我们考虑存在异常值时的估计方法进行估计，结果如下：

ARIMA Model Parameters

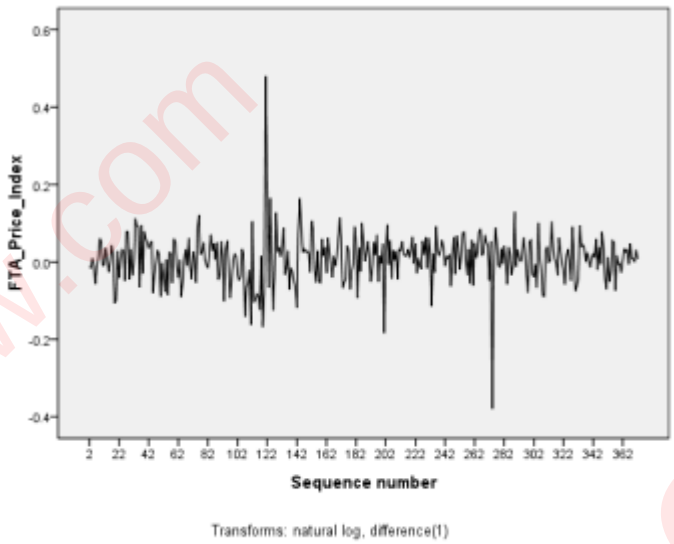
				Estimate	SE	t	Sig.
FTA-Model_1	FTA	No Transformation	Constant	1.102	.229	4.806	.000
			AR Lag 1	-1.000	.012	-83.247	.000
			MA Lag 1	-.999	.021	-46.891	.000

此时 BIC=3.324， $R^2 = 0.304$ 。显然后者的 R^2 有明显的改善。

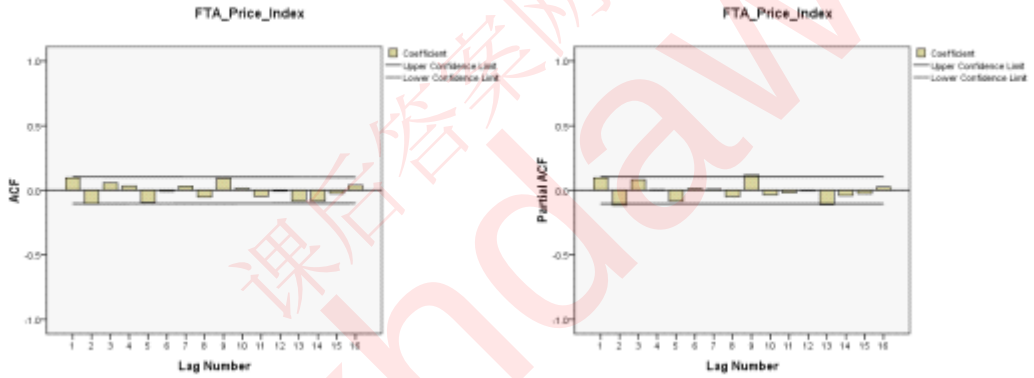
2、我们首先对数据作对数转换，时序图如下



易知所得的序列不平稳，所以我们需要对变换后的数据进行 1 阶差分，差分后的时序图为：



从序列图我们发现该序列有异常点，此时的 ACF 和 PACF 为：



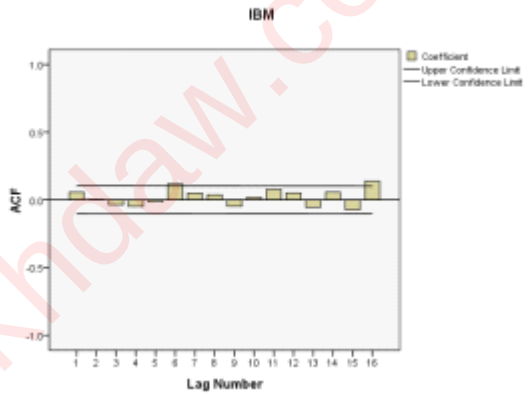
从图中看出；样本序列数据的自相关系数在某一固定水平线附近摆动，且逐渐衰减，所以该时间序列基本是平稳的。从图中可以看出ACF有明显的截尾特征，而PACF具有指数衰减特征，因此我们可以认为该数据服从MA模型。根据其特征我们考虑了ARMA（0,1,1）、ARMA（0,1,2）、ARMA（0,1,3）几个模型，并根据BIC准则选择了ARMA（0,1,2）。参数估计结果如下：

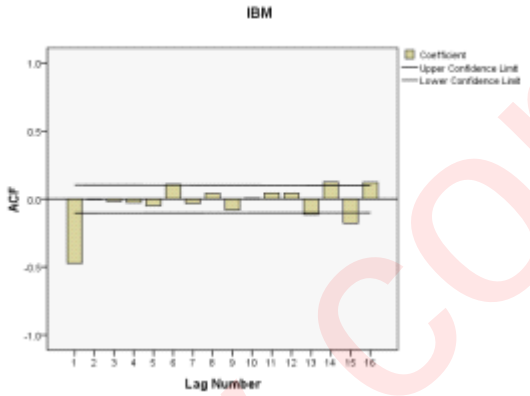
ARIMA Model Parameters

					Estimate	SE	t	Sig.
FTA_Price_Index-Model_1	FTA_Price_Index	No Transformation	Constant		4.612	1.315	3.508	.001
			Difference		1			
			MA	Lag 1	-.315	.053	-5.956	.000
				Lag 2	.252	.053	4.717	.000

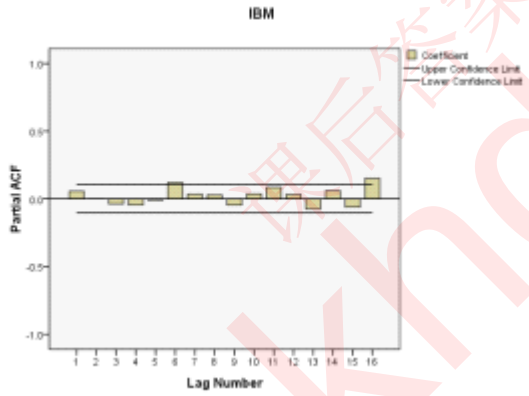
此时 BIC=6.599， $R^2=0.998$。

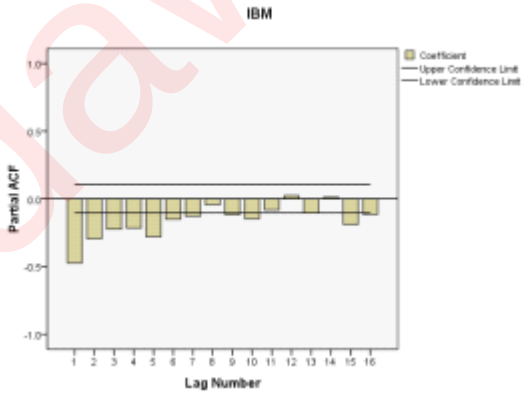
3、从题目中的时序图知该序列不平稳，通过差分平稳性得到了较大的提高。下图给出了 1，2 阶差分后的 ACF 图：





1，2 差分后的 PACF 如图所示：





如果考虑1阶差分，模型拟和的效果并不好，因此我们考虑2阶差分。2阶差分后的ACF有明显的截尾特征，而PACF具有指数衰减特征。因此我们可以认为该数据服从MA模型，经过分析其服从MA模型，经过分析其服从ARIMA(0,2,2)。参数估计结果如下：

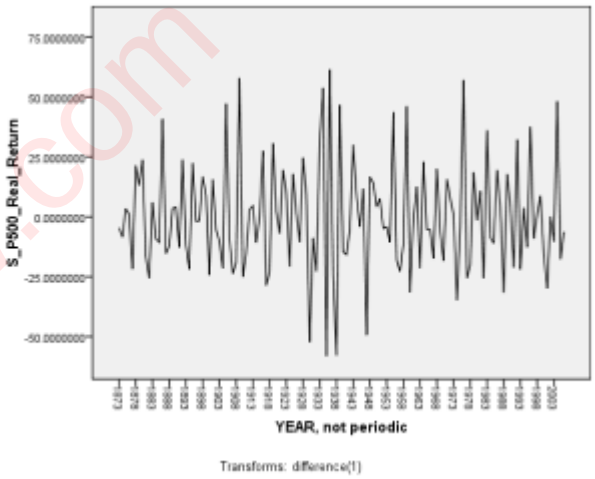
ARIMA Model Parameters

					Estimate	SE	t	Sig.
IBM-Model_1	IBM	No Transformation	Constant		-.006	.005	-1.262	.208
			Difference		2			
			MA	Lag 1	.825	.313	2.633	.009
				Lag 2	.175	.080	2.189	.029

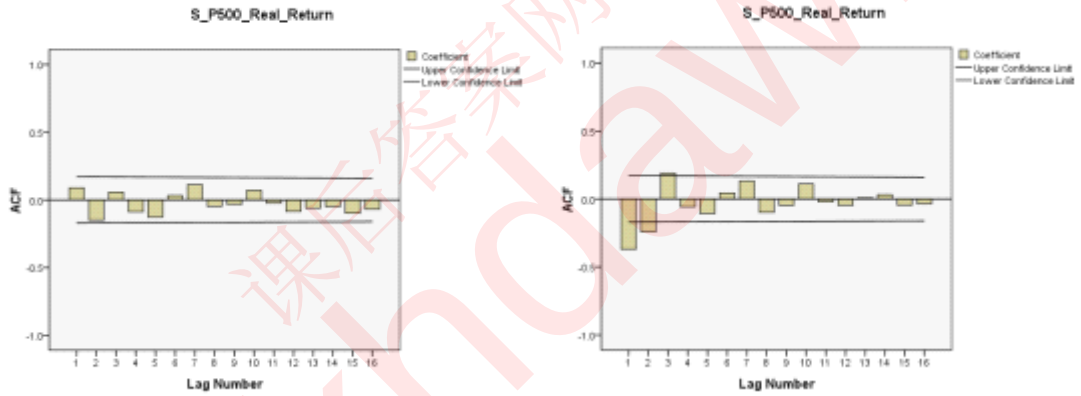
此时 BIC=3.902， $R^2=0.994$。

4、由题目中的序列图知该序列不平稳，通过差分平稳性得到了较大的提高。

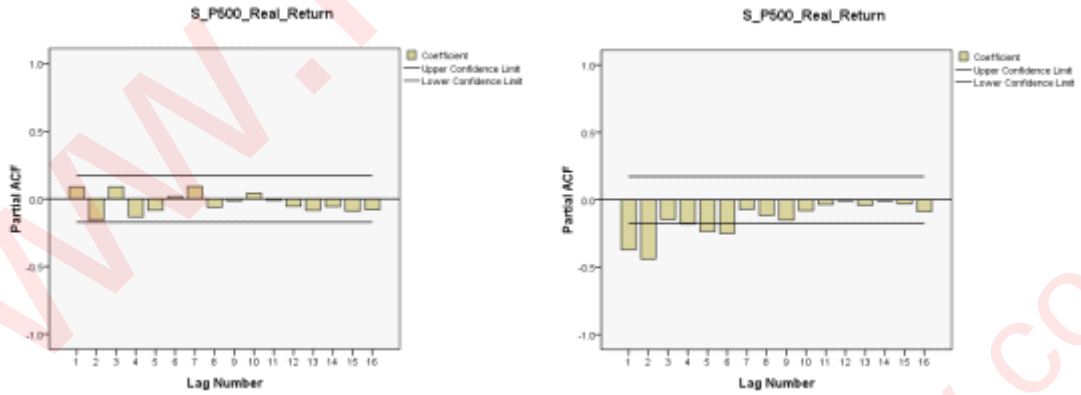
下图给出了 1 阶差分后的时序图：



0，1 差分后的 ACF 如下图所示：



0，1 差分后的 PACF 如下图所示：



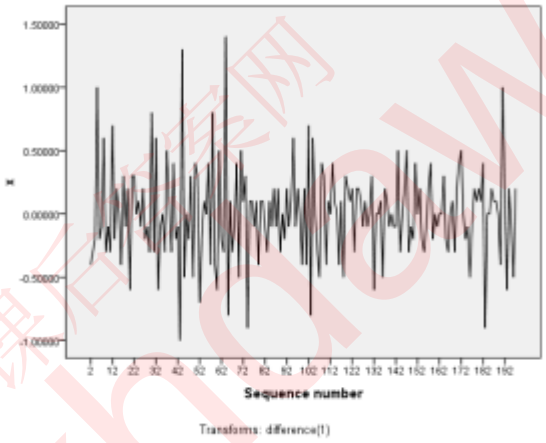
如果考虑0阶差分，模型拟和的效果并不好，因此我们考虑1阶差分。1阶差分后从图中可以看出ACF有明显的指数衰减特征，而PACF具有截尾特征。因此我们可以认为该数据服从AR模型，经过分析其服从ARIMA(2,1,0)。参数估计结果如下：

ARIMA Model Parameters

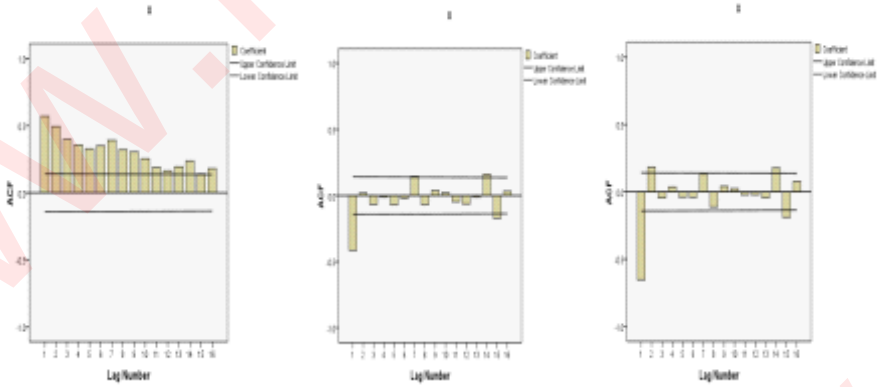
					Estimate	SE	t	Sig.
S_P500_Real_Return-Model_1	S_P500_Real_Return	No Transformation	Constant		.005	.868	.005	.996
			AR	Lag 1	-.530	.079	-6.748	.000
				Lag 2	-.436	.079	-5.540	.000
			Difference		1			

此时 BIC=6.067， $Adj-R^2=0.304$。

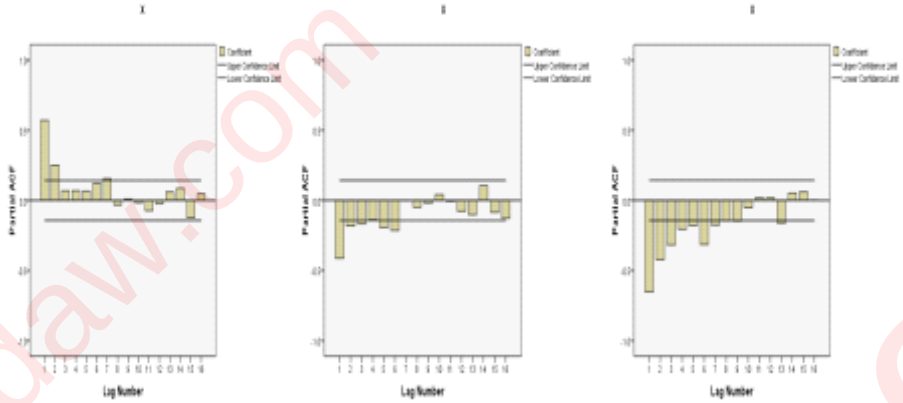
5、由已知的序列图知序列不平稳，通过差分平稳性得到了较大的提高。1 阶差分后的时序图为：



0, 1，2 差分后的 ACF 如下图所示：



0, 1，2 差分后的 PACF 如下图所示：



如果考虑0阶差分，模型拟和的效果并不好，因此我们考虑1,2阶差分。差分后从图中可以看出ACF有截尾特征，而PACF具有指数衰减特征。因此我们可以认为该数据服从MA模型,根据其特征我们考虑了ARMA（0,1,1）、ARMA（0,1,2）、ARMA（0,1,1）,ARMA（0,2,1）、ARMA（0,2,2）、ARMA（0,2,1）几个模型，并根据BIC准则选择了ARMA（0,1,2）。参数估计结果如下：

ARIMA Model Parameters

				Estimate	SE	t	Sig.
x-Model_1 x	No Transformation	Constant		.004	.006	.645	.520
		Difference		1			
		MA	Lag 1	.551	.072	7.663	.000
			Lag 2	.170	.071	2.381	.018

此时 BIC=-2.289， $R^2=0.458$ 。

第 8 章　练习题参考答案

8.1 解：（1） $\mu_{\bar{x}}=\sqrt{\frac{\sigma^2}{n}}=\frac{\sigma}{\sqrt{n}}=\frac{5}{\sqrt{40}}=0.79$

（2） $\Delta_{\bar{x}}=Z_{\alpha/2}\mu_{\bar{x}}=1.96\times0.79=1.55$

8.2（数据文件为 ex8.2）

解：（1） $\bar{x}=\frac{\sum x}{n}=3.32$ （小时）

（2） $s=\sqrt{\frac{\sum\left(x-\bar{x}\right)^2}{n-1}}=1.61$ （小时）

$\mu_{\bar{x}}=\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}=\sqrt{\frac{1.61^2}{36}\left(1-\frac{36}{7500}\right)}=0.27$ （小时）

8.3 解： ∵ $\mu_p=\sqrt{\frac{P(1-P)}{n}}=\sqrt{\frac{23\%\times(1-23\%)}{200}}=2.98\%$

∴ 当置信度为 90%时 $\Delta_p=Z_{\alpha/2}\mu_p=1.645\times2.98\%=4.90\%$

当置信度为 95%时 $\Delta_p=Z_{\alpha/2}\mu_p=1.96\times2.98\%=5.84\%$

8.4 解： $n_{\bar{x}}=\frac{Z_{\alpha/2}^2\sigma^2}{\Delta_{\bar{x}}^2}=\frac{1.96^2\times120^2}{20^2}=138.30\approx139$（人）

8.5 解： ∵ $P_1(1-P_1)=2\%\times98\%=1.96\%$

$P_2(1-P_2)=3\%\times97\%=2.91\%$

$P_3(1-P_3)=4\%\times96\%=3.84\%$

∴ $n_p=\frac{Z_{\alpha/2}^2P_3(1-P_3)}{\Delta_p^2}=\frac{1.96^2\times4\%\times(1-4\%)}{4\%^2}=92.20\approx93$

8.6 解： $\overline{\sigma_i^2}=\frac{\sum\sigma_i^2n_i}{n}=3799.375$（元）

$\bar{x}=\frac{\sum x_in_i}{n}=4275$（元）

重复抽样

$$\mu_{\bar{x}} = \sqrt{\frac{\overline{\sigma_i^2}}{n}} = \sqrt{\frac{3799.375}{400}} = 3.08 \text{（元）}$$

不重复抽样

$$\mu_{\bar{x}} = \sqrt{\frac{\overline{\sigma_i^2}}{n}\left(1-\frac{n}{N}\right)}$$

$$= \sqrt{\frac{3799.375}{400}\left(1-\frac{400}{4000}\right)} = 2.92 \text{（元）}$$

8.7（数据文件为 ex8.7）

解：$R=1440 \quad r=24$

（1）计算样本平均数和样本比例

$$\bar{\bar{x}} = \frac{\sum \bar{x}_i}{r} = 50.11 \text{（千克）}$$

$$p = \frac{\sum p_i}{r} = 97.56\%$$

（2）计算样本平均数群间方差和样本比例群间方差

$$\sigma_{\bar{x}}^2 = \frac{\sum (\bar{x}_i - \bar{\bar{x}})^2}{r} = 0.6491 \text{（千克）}$$

$$\sigma_p^2 = \frac{\sum (p_i - p)^2}{r} = 0.0153\%$$

$$\therefore 49.78\text{（千克）} \le \bar{X} \le 50.44\text{（千克）}$$

$$97.06 \le P \le 98.06\%$$

第 9 章练习题参考答案

9.1 解：（1）三种商品的销售量个体指数分别为：

A：$\frac{8}{10}=80\%$；B：$\frac{7}{7}=100\%$；C、$\frac{20}{15}=133.33\%$

三种商品的销售价格个体指数分别为：

A：$\frac{40}{30}=133.33\%$；B：$\frac{20}{20}=100\%$；C、$\frac{50}{60}=83.33\%$

（2）三种商品的销售量总指数为：

$$L_q=\frac{\sum p_0q_1}{\sum p_0q_0}=\frac{30\times8+20\times7+60\times20}{30\times10+20\times7+60\times15}=\frac{1580}{1340}=117.91\%$$

或 $$P_q=\frac{\sum p_1q_1}{\sum p_1q_0}=\frac{40\times8+20\times7+50\times20}{40\times10+20\times7+50\times15}=\frac{1460}{1290}=113.18\%$$

三种商品的销售价格总指数为：

$$L_p=\frac{\sum p_1q_0}{\sum p_0q_0}=\frac{1290}{1340}=96.27\%$$；或 $$P_p=\frac{\sum p_1q_1}{\sum p_0q_1}=\frac{1460}{1580}=92.41\%$$

比较：由于拉氏指数与帕氏指数的同度量因素固定时期不同，因而对同一指数的计算结果也就不同。本题中，拉氏指数的计算结果大于帕氏指数。

9.2 解：销售额指数 $$V=\frac{\sum p_1q_1}{\sum p_0q_0}=\frac{12\times3+7\times5+4\times6}{10\times2+8\times5+2\times4}=\frac{95}{68}=139.71\%$$

销售额变动绝对额：$\sum p_1q_1-\sum p_0q_0=95-68=27$（万元）

销售量指数 $$L_q=\frac{\sum p_0q_1}{\sum p_0q_0}=\frac{10\times3+8\times5+2\times6}{68}=\frac{82}{68}=120.59\%$$

销售量变动对销售额的影响额：$\sum p_0q_1-\sum p_0q_0=82-68=14$（万元）

销售价格指数 $$P_p=\frac{\sum p_1q_1}{\sum p_0q_1}=\frac{95}{82}=115.85\%$$

销售价格变动对销售额的影响额：$\sum p_1q_1-\sum p_0q_1=95-82=13$（万元）

以上三个指数之间的关系为：

相对数方面：139.71%=120.59%×115.85%

绝对数方面：27(万元)=14(万元)+13(万元)

计算结果表明：报告期与基期相比，由于三种商品的销售量增长 20.59%，使销售额增加了 14 万元；又由于三种商品的销售价格上涨 15.85%，使销售额增

加了 13 万元。两个因素共同作用的结果，使销售额最终增长 39.71%，共增加了 27 万元。

9.3 解：基期加权的算术平均数指数公式编制的产量总指数为：

$$A_q=\frac{\sum\frac{q_1}{q_0}p_0q_0}{\sum p_0q_0}=\frac{1.03\times160+0.95\times100+1.10\times150}{160+100+150}=\frac{424.8}{410}=103.61\%$$

报告期加权的调和平均数指数公式编制的产量总指数为：

$$H_q=\frac{\sum p_1q_1}{\sum\frac{q_0}{q_1}p_1q_1}=\frac{\sum p_1q_1}{\sum\frac{1}{\frac{q_1}{q_0}}p_1q_1}=\frac{200+80+180}{\frac{1}{1.03}\times200+\frac{1}{0.95}\times80+\frac{1}{1.10}\times180}=\frac{460}{442.02}=104.07\%$$

比较：由于 A_q 采用基期总量加权计算总指数，H_q 采用报告期总量加权计算总指数，因而根据同一组资料用两个公式计算的产量总指数存在差异，本题中，$A_q<H_q$。

9.4 解：基期加权的算术平均数指数公式编制的单位成本总指数为：

$$A_p=\frac{\sum\frac{p_1}{p_0}p_0q_0}{\sum p_0q_0}=\frac{1.04\times100+0.94\times120+1.03\times130}{100+120+130}=\frac{350.7}{350}=100.2\%$$

报告期加权的调和平均数指数公式编制的单位成本总指数为：

$$H_p=\frac{\sum p_1q_1}{\sum\frac{1}{\frac{p_1}{p_0}}p_1q_1}=\frac{150+90+130}{\frac{1}{1.04}\times150+\frac{1}{0.94}\times90+\frac{1}{1.03}\times130}=\frac{370}{366.19}=101.04\%$$

比较：比较：由于 A_p 采用基期总量加权计算总指数，H_p 采用报告期总量加权计算总指数，因而根据同一组资料用两个公式计算的单位成本总指数存在差异，本题中，$A_p<H_p$。

9.5 解：（1）三种商品的销售价格总指数为：

$$H_p=\frac{\sum p_1q_1}{\sum\frac{1}{\frac{p_1}{p_0}}p_1q_1}=\frac{500+300+200}{\frac{1}{1.05}\times500+\frac{1}{0.90}\times300+\frac{1}{1.08}\times200}=\frac{1000}{994.71}=100.53\%$$

（2）三种商品的销售额指数为：

$$\frac{\sum p_1q_1}{\sum p_0q_0}=\frac{1000}{800}=125\%$$

则三种商品的销售量总指数为：125%÷100.53%=124.34%

9.6 解：首先列出计算表如下：

某企业工人总平均工资变动分析计算表

工人组别	工人数（人）		平均工资(元)		工资总额（元）		
	基期 f_0	报告期 f_1	基期 x_0	报告期 x_1	基期 x_0f_0	报告期 x_1f_1	假定 x_0f_1
新工人	120	210	1500	2000	180000	420000	315000
老工人	80	90	2300	2600	184000	234000	207000
合计	200	300	—	—	364000	654000	522000

基期总平均工资：

$$\bar{x}_0=\frac{\sum x_0f_0}{\sum f_0}=\frac{364000}{200}=1820(\text{元/人})$$

报告期总平均工资：

$$\bar{x}_1=\frac{\sum x_1f_1}{\sum f_1}=\frac{654000}{300}=2180(\text{元/人})$$

假定总平均工资：

$$\bar{x}_{\text{假定}}=\frac{\sum x_0f_1}{\sum f_1}=\frac{522000}{300}=1740(\text{元/人})$$

(1)总平均工资的可变构成指数：

$$I_{\text{可变}}=\frac{\sum x_1f_1}{\sum f_1}\div\frac{\sum x_0f_0}{\sum f_0}=\frac{\bar{x}_1}{\bar{x}_0}=\frac{2180}{1820}=119.78\%$$

变动差额：$\bar{x}_1-\bar{x}_0=2180-1820=360(\text{元/人})$

(2)总平均工资的结构影响指数：

$$I_{\text{结构}}=\frac{\sum x_0f_1}{\sum f_1}\div\frac{\sum x_0f_0}{\sum f_0}=\frac{\bar{x}_{\text{假定}}}{\bar{x}_0}=\frac{1740}{1820}=95.60\%$$

变动差额：$\bar{x}_{\text{假定}}-\bar{x}_0=1740-1820=-80(\text{元/人})$

总平均工资的固定构成指数：

$$I_{\text{固定}}=\frac{\sum x_1f_1}{\sum f_1}\div\frac{\sum x_0f_1}{\sum f_1}=\frac{\bar{x}_1}{\bar{x}_{\text{假定}}}=\frac{2180}{1740}=125.29\%$$

变动差额：$\bar{x}_1-\bar{x}_{假定}=2180-1740=440$(元/人)

相对数：　　119.78%=95.60%×125.29%

绝对数：　　360（元/人）=－80（元/人）＋440（元/人）

计算结果表明：由于各组平均工资提高，使总平均工资提高了 25.29%，平均每人增加 440 元；又由于各组平均工资水平不同的工人人数结构变动，使总平均工资下降了 4.4%，平均每人减少 80 元。两个因素共同作用的结果，使整个企业总平均工资最终提高了 19.78%，平均每人增加 360 元。

(3)由于总平均工资变动而影响的工资总额为：

$(\bar{x}_1-\bar{x}_0)\times\sum f_1=(2180-1820)\times300=108000$(元)$=10.8$(万元)，表明由于工人总平均工资提高而使工资总额增加了 10.8 万元。

其中：

由于各组工人本身平均工资水平提高而影响的工资总额为：

$(\bar{x}_1-\bar{x}_{假定})\times\sum f_1=(2180-1740)\times300=132000$(元)$=13.2$(万元)，表明由于各组工人本身平均工资水平提高而使工资总额增加 13.2 万元；

又由于各组平均工资水平不同的工人人数结构变动而影响的工资总额为：

$(\bar{x}_{假定}-\bar{x}_0)\times\sum f_1=(1740-1820)\times300=-24000$(元)$=-2.4$(万元)；表明由于平均工资水平不同的工人人数结构变动而使工资总额减少 2.4 万元。

以上结果存在如下关系：

$(\bar{x}_1-\bar{x}_o)\times\sum f_1=(\bar{x}_1-\bar{x}_{假定})\times\sum f_1+(\bar{x}_{假定}-\bar{x}_o)\times\sum f_1$

10.8（万元）=13.2（万元）＋（－）2.4（万元）

第 10 章　统计决策练习题参考答案

1、解：

（1）

市场需求	20	30	40	50
概率	0.30	0.40	0.20	0.10
方案一（订购 20 台）	4000	4000	4000	4000
方案二（订购 30 台）	1000	6000	6000	6000
方案三（订购 40 台）	-2000	3000	8000	8000
方案四（订购 50 台）	-5000	0	5000	10000

（2）因为，方案一在各种状态下的最大收益为 4000 元，方案二在各种状态下最大收益为 6000 元，方案三在各种状态下的最大收益为 8000 元，方案四在各种状态下的最大收益为 10000 元，所以，根据最大最大准则，最佳订购计划是 50 台。

因为方案一在各种状态下的最小收益为 4000 元，方案二在各种状态下的最小收益为 1000 万元，方案三在各种状态下的最小收益值为-2000 元，方案四在各种状态下的最小收益值为-5000 元，根据最大最小准则时，最佳订购计划是 20 台。

因为后悔值矩阵表为：

市场需求	20	30	40	50
概率	0.30	0.40	0.20	0.10
方案一（订购 20 台）	0	2000	4000	6000
方案二（订购 30 台）	3000	0	2000	4000
方案三（订购 40 台）	6000	3000	0	2000
方案四（订购 50 台）	9000	6000	3000	0

根据最小最大后悔值准则，最佳订购计划是 30 台。

（3）因为市场需求为 30 台的概率最大，根据最大概率准则时，最佳订购计划是 30 台。

因为

$$E(Q(a_1)) = 0.3\times4000 + 0.4\times4000 + 0.2\times4000 + 0.1\times4000 = 4000$$

$$E(Q(a_2)) = 0.3\times1000 + 0.4\times6000 + 0.2\times6000 + 0.1\times6000 = 4500$$

$$E(Q(a_3)) = 0.3\times(-2000) + 0.4\times3000 + 0.2\times8000 + 0.1\times8000 = 3000$$

$$E(Q(a_4)) = 0.3\times(-5000) + 0.4\times0 + 0.2\times5000 + 0.1\times10000 = 500$$

所以，根据收益期望值准则，最佳订购计划是 30 台。

因为方案四的收益期望值较小，所以可以从备选方案中排除，只需考虑方案一、方案二和方案四。

$$\sigma^2(a_1)=(4000-4000)^2\times0.3+(4000-4000)^2\times0.4+(4000-4000)^2\times0.2+(4000-4000)^2\times0.1=0$$

$$\sigma^2(a_2)=(1000-4500)^2\times0.3+(6000-4500)^2\times0.4+(6000-4500)^2\times0.2+(6000-4500)^2\times0.1=5.25\times10^6$$

$$\sigma^2(a_3)=(-2000-3000)^2\times0.3+(3000-3000)^2\times0.4+(8000-3000)^2\times0.2+(8000-3000)^2\times0.1=1.5\times10^7$$

$$V_1=\frac{\sigma(a_1)}{E(Q(a_1))}=\frac{\sqrt{0}}{4000}=0$$

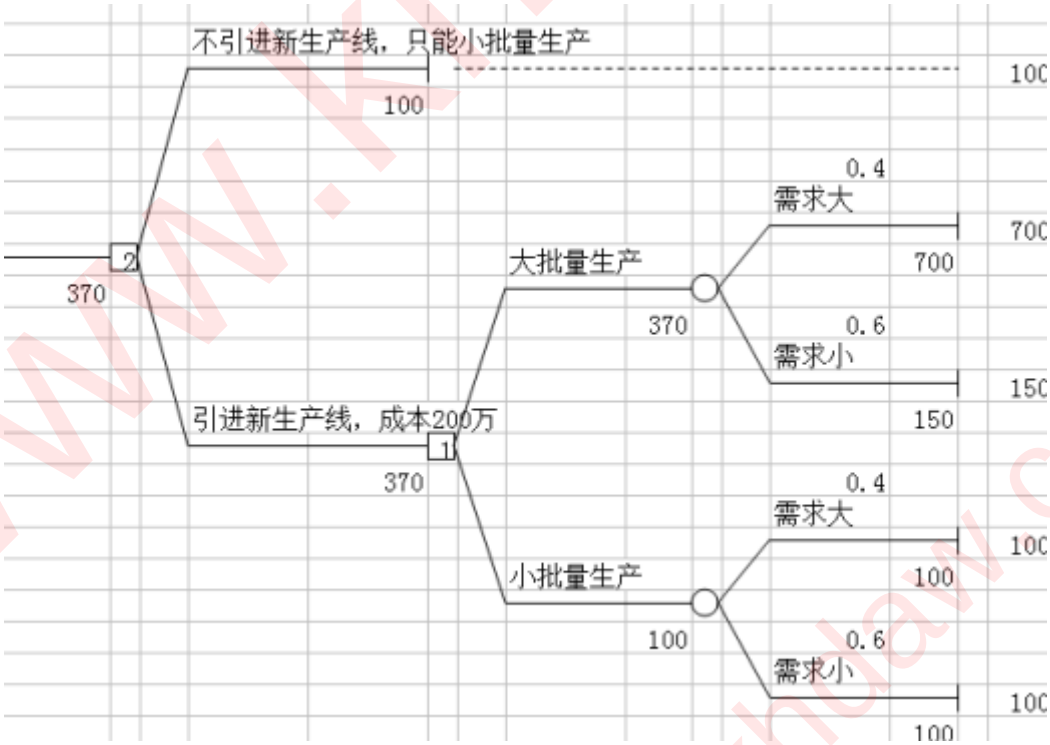
$$V_2=\frac{\sigma(a_2)}{E(Q(a_2))}=\frac{\sqrt{5.25\times10^6}}{4500}=0.5092$$

$$V_3=\frac{\sigma(a_3)}{E(Q(a_3))}=\frac{\sqrt{1.5\times10^7}}{3000}=1.291$$

由于方案一和方案二的期望收益比较接近，根据变异系数准则，最佳订购计划是方案一，即 20 台。

2、解：

（1）依题意，不管市场需求大或小，小批量生产带来的收益均是 100 万，该问题的决策树为：



（2）采用逆推法可知，不管市场的需求大还是小，企业实现大批量的生产获得

的收益值均高于小批量生产的收益值，而要实现大批量生产必须引进新的生产线，除去引进成本 200 万，大批量生产可以获得 170 万的收益。所以，企业应该引进新的生产线，并进行大批量的生产。

3、解：

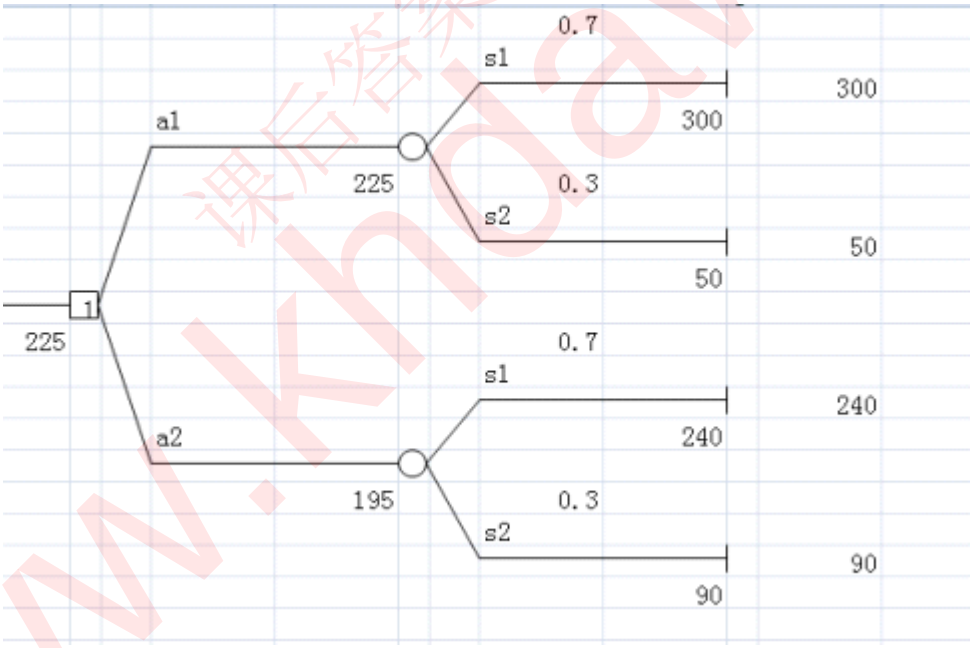
（1）两种销售方案持续 5 年的销售利润如下表

销售状态 概　　率	s_1（销售状况好） 0.7	s_2（销售状况差） 0.3
方案一 a_1（代理销售）	300	50
方案二 a_2（兼营销售）	240	90

方案一的期望销售利润：$300\times0.7+50\times0.3=225$ 万元

方案二的期望销售利润：$240\times0.7+90\times0.3=195$ 万元

(2)



由决策树可知，最佳方案为方案一，即代理销售。

（3）根据销售利润表，如果销售商在销售前已经知道销售状况将如何变化，那么他一定在追求利润最大化原则下选择适合未来实际情况的最佳方案。也就是说，如果他知道销售状况好，那么他就选择方案 a_1，获得利润 300 万元，如果销售状况差，则选择方案 a_2，获得利润 90 万元，从而完全信息条件下的利润期望值为：

$EPPI = 0.7 \times 300 + 0.3 \times 90 = 237$

而由决策树分析可知，在没有完全信息条件下，投资者将选择方案 a_1 作为最佳行动方案，获得平均利润为 225 万元，即 $EMV^* = 225$。

所以，

$EVPI = EPPI - EMV^* = 237 - 225 = 12$

即完全信息的期望价值是 12 万元。

4、解：（1）

$$P(I_1) = \sum_{j=1}^{2} P(s_j)P(I_1|s_j) = P(s_1)P(I_1|s_1) + P(s_2)P(I_1|s_2)$$

$$= 0.8 \times 0.8 + 0.2 \times 0.3 = 0.7$$

$$P(I_2) = \sum_{j=1}^{2} P(s_j)P(I_2|s_j) = P(s_1)P(I_2|s_1) + P(s_2)P(I_2|s_2)$$

$$= 0.8 \times 0.2 + 0.2 \times 0.7 = 0.3$$

（2）

$$P(s_1|I_1) = \frac{P(s_1I_1)}{P(I_1)} = \frac{P(s_1)P(I_1|s_1)}{P(I_1)} = \frac{0.8 \times 0.8}{0.7} = 0.91$$

$$P(s_2|I_1) = \frac{P(s_2I_1)}{P(I_1)} = \frac{P(s_2)P(I_1|s_2)}{P(I_1)} = \frac{0.2 \times 0.3}{0.7} = 0.09$$

$$P(s_1|I_2) = \frac{P(s_1I_2)}{P(I_2)} = \frac{P(s_1)P(I_2|s_1)}{P(I_2)} = \frac{0.8 \times 0.2}{0.3} = 0.53$$

$$P(s_2|I_2) = \frac{P(s_2I_2)}{P(I_2)} = \frac{P(s_2)P(I_2|s_2)}{P(I_2)} = \frac{0.2 \times 0.7}{0.3} = 0.47$$

（3）第一步，进行先验分析。

$E(a_1)) = 0.8 \times 400 + 0.2 \times (-200) = 280$

$E(a_2)) = 0.8 \times 200 + 0.2 \times 70 = 174$

说明在没有完全信息条件下，企业将选择方案 a_1 作为最佳行动方案，获得收益 280 万元，即 $EMV^* = 280$

第二步，进行后验分析。

若信息咨询公司做市场调查，结果是受欢迎，则

$$E(Q(a_1)) = 0.91\times 400 + 0.09\times(-200) = 346$$

$$E(Q(a_2)) = 0.91\times 200 + 0.09\times 70 = 188.3$$

若信息咨询公司做市场调查，结果是不受欢迎，则

$$E(Q(a_1)) = 0.53\times 400 + 0.47\times(-200) = 118$$

$$E(Q(a_2)) = 0.53\times 200 + 0.47\times 70 = 138.9$$

综合来看，后验分析得到的结论为：

若信息咨询公司市场调查的结果是受欢迎，则最优方案是 a_1，预期收益是 346 万元；

若信息咨询公司市场调查的结果是不受欢迎，则最优方案是 a_2，预期收益是 138.9 万元。

因此，由后验分析中数据，得

$$EMV' = 0.7\times 346 + 0.3\times 138.9 = 283.87$$

$$EVSI = EMV' - EMV^* = 283.87 - 280 = 3.87$$

说明企业通过咨询公司市场调查获得附加信息，可以增加投资者 3.87 万元的平均收益，而企业计划用于咨询的费用不超过 2 万元，由此可见，企业应该先请信息咨询公司做市场调查。