

实验概述：

【实验目的及要求】

实验目的：通过实际案例分析，使学生掌握如何使用 SAS 进行编程，完成方差分析和数据分析管理。

一、为了研究刺激对人血压的影响，研究人员随机选取了 36 位志愿者，在施加刺激之前和施加刺激之后分别测量每个人的血压值。得到数据如下：

施加刺激前	119	123	129	117	139	127
施加刺激后	127	130	130	126	131	124
施加刺激前	139	124	125	124	125	126
施加刺激后	140	136	117	133	129	134
施加刺激前	139	124	125	139	145	136
施加刺激后	141	132	117	129	129	134
施加刺激前	139	134	125	125	125	126
施加刺激后	140	136	117	133	129	134
施加刺激前	139	134	125	129	125	126
施加刺激后	140	136	117	138	119	134
施加刺激前	129	123	119	127	129	127
施加刺激后	137	130	140	136	131	134

在 5% 的显著性水平下，试分析

(1) 施加刺激前后，志愿者的血压有无显著差异？

(2) 假设以 130（含 130）为分界线将人体血压区分为正常和异常，我们想要研究该项刺激对人体血压异常的影响。试运用 McNemar 检验方法进行检验，分析二者之间的关系。

二、有五种不同牌号的铁锈防护剂（简称防锈剂），为了比较其防锈能力，进行了如下试验：

制作一些大小形状和材质都相同的铁件（试验样品），然后把他们依次随机地抽取，在其上涂上某一牌号的防锈剂，5 种牌号的防锈剂分别涂了 4 件。最后，把涂好防锈剂的样品放在同一个广场上让其经受日晒、风吹和雨打。过一段时间后，再观察铁件生锈的情况。通过专家对其防锈能力进行评分：样品上无锈迹的评 5 分，全锈了评 0 分。

五位受聘专家进行独立评分，把一个样品的 5 个评分的平均值作为该样品的防锈能力。数据列于下表 1 中。

表 1： 防锈能力数据

因子 A（防锈剂）		A1	A2	A3	A4	A5
数 据 y_{ij}	1	2.3	2.2	2.2	2.4	2.3
	2	2.1	2.3	2.0	2.7	2.5
	3	2.4	2.4	1.9	2.6	2.3
	4	2.5	2.6	2.1	2.7	2.4

请通过编写 SAS 程序，对上述数据进行分析。

问：在显著性水平 5%的条件下，上述四种不同牌号的防锈剂 A1, A2, A3, A4, A5，其防锈能力是否存在差异？哪个牌号的防锈剂的防锈能力最强？

三、某产品为促进销量新设计了四种广告文案方案，选取了 3 个商店将广告在店内不同货架上摆放。观察一定时期的销量。厂商希望比较四种文案在不同商店的营销能力。

文案方案	商店 B1		商店 B2		商店 B3	
A1	12	14	18	16	17	18
A2	14	15	12	14	13	15
A3	19	20	17	18	21	22
A4	24	23	30	28	29	28

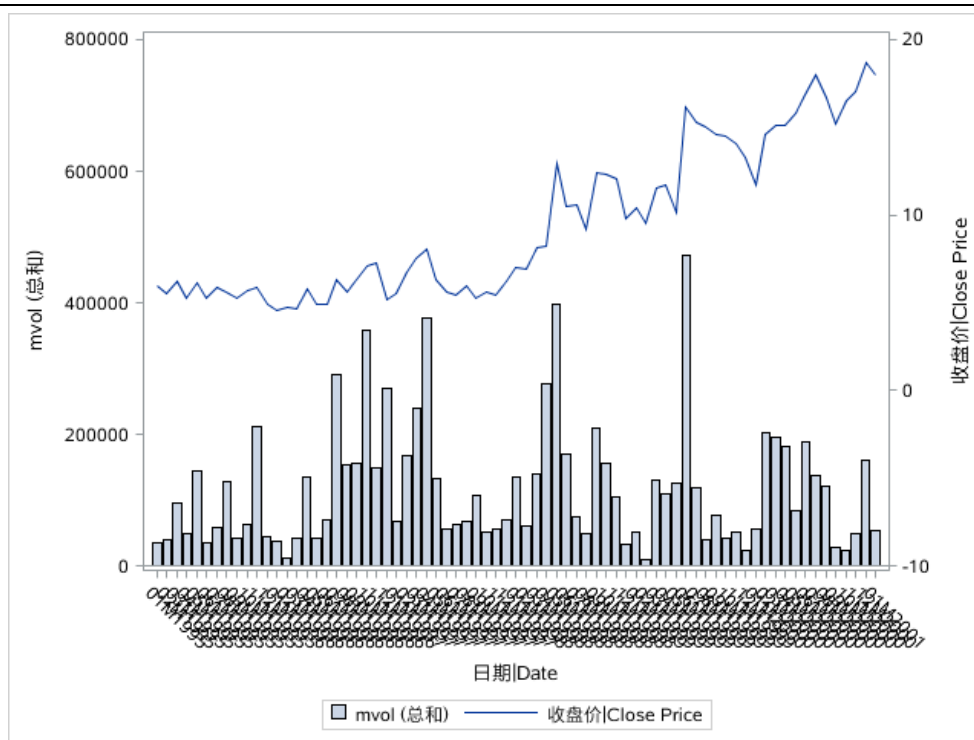
假定满足方差分析的假设条件，在显著性水平 0.05 下，分析不同文案方案和商店下的销量是否有显著差异，以及文案方案与商店的交互作用是否显著，如果交互有差异，分析在商店确定的情况下，文案方案对产出的影响。

四、数据步提升训练

a600605 所给数据中包含上证股票 600605，1995-2001 年的行情信息。其数据信息如下所示。

	开盘价 Open Price	最高价 High Price	最低价 Low Price	收盘价 Close Price	交易量 Volume Traded	交易金额 sums	日期 Date
1	6.48	6.65	6.33	6.51	815.00	531000.00	1995-01-03
2	6.60	6.78	6.50	6.78	1559.00	1035000.00	1995-01-04
3	6.74	6.80	6.58	6.60	1638.00	1089000.00	1995-01-05
4	6.59	6.59	6.40	6.53	2358.00	1526000.00	1995-01-06
5	6.50	6.50	6.30	6.34	2345.00	1491000.00	1995-01-09
6	6.12	6.30	6.10	6.21	3019.00	1865000.00	1995-01-10
7	6.18	6.27	6.13	6.22	2392.00	1484000.00	1995-01-11
8	6.22	6.27	6.08	6.10	1568.00	969000.00	1995-01-12
9	6.05	6.14	6.01	6.05	2611.00	1583000.00	1995-01-13
10	6.01	6.15	5.88	6.10	2800.00	1684000.00	1995-01-16
11	6.18	6.18	6.05	6.07	1891.00	1156000.00	1995-01-17
12	6.10	6.24	6.08	6.22	2158.00	1332000.00	1995-01-18
13	6.20	6.21	5.95	5.96	2065.00	1250000.00	1995-01-19
14	5.86	6.14	5.80	6.09	2725.00	1615000.00	1995-01-20
15	6.00	6.00	5.82	5.88	1361.00	801000.00	1995-01-23
16	5.85	5.85	5.81	5.87	1171.00	687000.00	1995-01-24

- 1) 使用 Data 步来计算 a600605 这支股票在 1995-1998 年的市场收益，即该股票的月收益率（个股月收益率=[（本月收盘价-上月收盘价）/上月收盘价]×100%），过程中不要使用 dif 和 lag 函数。其中 date 的格式设置为‘1995-01’的形式，并删除 1995 年 1 月的观测数据。
- 2) 将数据处理为月度数据，新数据集保留月份（date）、月收盘价（clpr）和月交易量（mvol）三个变量。
- 3) 编写 graph，绘制 a600605 这支股票的收益率曲线，横坐标标签改为“日期”，纵坐标标签改为“月收益率”
- 4) 根据第 2 步骤结果绘制如下月度股线图。



【实验原理】

一、可先使用 SAS 的数据步建立数据集，然后使用 SAS 的 XXX 过程编写程序，通过单因素方差分析获得相应结果。

二、可先使用 SAS 的数据步建立数据集，然后使用 SAS 的 XXX 过程编写程序，通过多因素方差分析获得相应结果。

三、可先使用 SAS 的数据步建立数据集，然后使用 SAS 的 REG 和 GLM 过程建立相应的回归方程。但在使用 REG 和 GLM 建立多项式回归方程时，要注意其实现方法的区别。

【实验环境】（使用的软硬件）

硬件：IBM PC 或其兼容机

软件：Microsoft Windows, Microsoft Word 2003 或更高版本, SAS 8.x 或更高版本.

实验内容：

【实验方案设计】

第一题 (1)

```
data bp_data;
input type$ bp @@;
datalines;
bf_iter 119 af_iter 127 bf_iter 123 af_iter 130 bf_iter 129
af_iter 130 bf_iter 117 af_iter 126 bf_iter 139 af_iter 131
bf_iter 127 af_iter 124
bf_iter 139 af_iter 140 bf_iter 124 af_iter 136 bf_iter 125
af_iter 117 bf_iter 124 af_iter 133 bf_iter 125 af_iter 129
bf_iter 126 af_iter 134
bf_iter 139 af_iter 141 bf_iter 124 af_iter 132 bf_iter 125
af_iter 117 bf_iter 139 af_iter 129 bf_iter 145 af_iter 129
bf_iter 136 af_iter 134
bf_iter 139 af_iter 140 bf_iter 134 af_iter 136 bf_iter 125
af_iter 117 bf_iter 125 af_iter 133 bf_iter 125 af_iter 129
bf_iter 126 af_iter 134
bf_iter 139 af_iter 140 bf_iter 134 af_iter 136 bf_iter 125
af_iter 117 bf_iter 129 af_iter 138 bf_iter 125 af_iter 119
bf_iter 126 af_iter 134
bf_iter 129 af_iter 137 bf_iter 123 af_iter 130 bf_iter 119
af_iter 140 bf_iter 127 af_iter 136 bf_iter 129 af_iter 131
bf_iter 127 af_iter 134
;
run;
proc print data=bp_data;
run;
/*conduct ttest to determine wheter significant difference
occurs between before_iteration and after_iter.*/
proc sort data=bp_data;
by type;
run;
proc ttest data=bp_data alpha=0.05;
class type;
var bp;
run;
```

SAS 系统

TTEST 过程

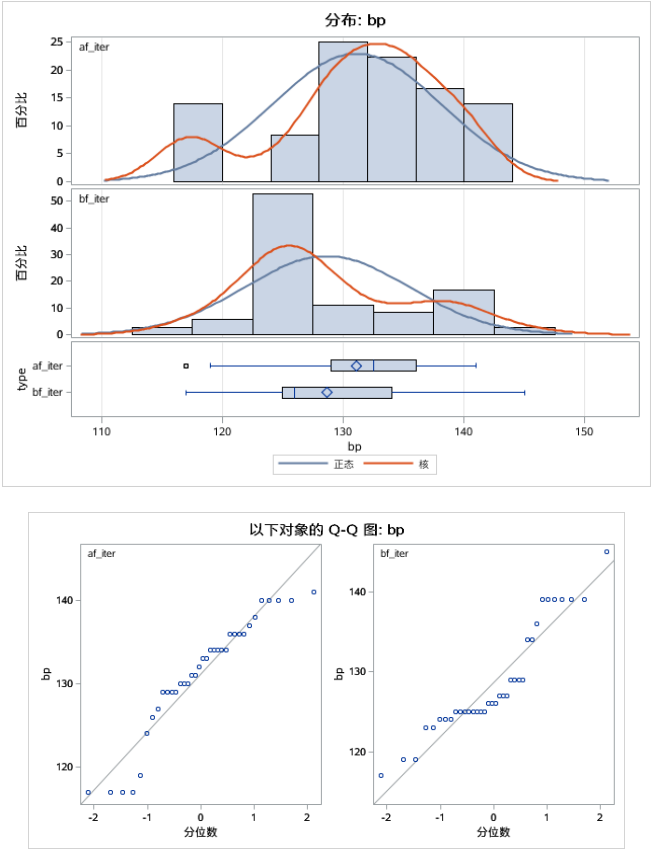
变量: bp

type	N	均值	标准差	标准误差	最小值	最大值
af_iter	36	131.1	6.9684	1.1614	117.0	141.0
bf_iter	36	128.6	6.7788	1.1298	117.0	145.0
差 (1-2)		2.4722	6.8742	1.6203		

type	方法	均值	95% 置信限均值	标准差	95% 置信限标准差
af_iter		131.1	128.8 133.5	6.9684	5.6519 9.0899
bf_iter		128.6	126.3 130.9	6.7788	5.4981 8.8425
差 (1-2)	汇总	2.4722	-0.7593 5.7038	6.8742	5.9001 8.2367
差 (1-2)	Satterthwaite	2.4722	-0.7594 5.7038		

方法	方差	自由度	t 值	Pr > t
汇总	等于	70	1.53	0.1316
Satterthwaite	不等于	69.947	1.53	0.1316

方差等价				
方法	分子自由度	分母自由度	F 值	Pr > F
折叠的 F	35	35	1.06	0.8713



由检验可知， $p\text{-value}>0.05$ ，无法拒绝原假设，且均值的 95%置信区间包含 0，故认为不具有明显差异。

第一题（2）

```
/*group data into normal/abnormal groups*/
data bp_2;
format status $12.;
set bp_data;
if bp <= 130 then status="normal";
```

```

else status="abnormal";
keep type status ;
run;
proc print data=bp_2;
run;
/*generate the dataset*/
proc means data=bp_2;
class type status;
output out=bp_3;
run;
data bp_2_data;
set bp_3;
if _TYPE_ ne 3 then delete;
drop _TYPE_;
run;
proc print data=bp_2_data;
run;
/*bp_2_data is the data_set required for McNemar test*/
/*agree to conduct McNemar test*/
proc freq data=bp_2_data;
    tables type*status / nopercnt norow nocol agree;
    weight _FREQ_;
run;

```

SAS 系统

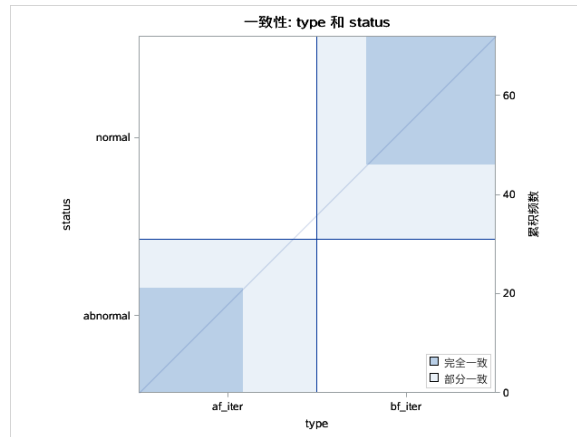
FREQ 过程

频数	表 - type * status			
	type	status		
		abnormal	normal	合计
	af_iter	21	15	36
	bf_iter	10	26	36
	合计	31	41	72

表 "status-type" 的统计量

McNemar 检验	
统计量 (S)	1.0000
自由度	1
Pr > S	0.3173

简单 Kappa 系数	
Kappa	0.3056
ASE	0.1111
95% 置信下限	0.0878
95% 置信上限	0.5234



由 McNemar 检验可知, $p\text{-value} > 0.05$, 故不能拒绝原假设, 认为无显著影响; 而 Kappa 系数认为一致性较差。

第二题

```
data stain;
input chem$ score @@;
datalines;
A1 2.3 A2 2.2 A3 2.2 A4 2.4 A5 2.3
A1 2.1 A2 2.3 A3 2.0 A4 2.7 A5 2.5
A1 2.4 A2 2.4 A3 1.9 A4 2.6 A5 2.3
A1 2.5 A2 2.6 A3 2.1 A4 2.7 A5 2.4
;
run;

proc print data=stain;
run;

proc glm data=stain;
class chem;
model score=chem / alpha=0.05;
run;
```

SAS 系统

GLM 过程

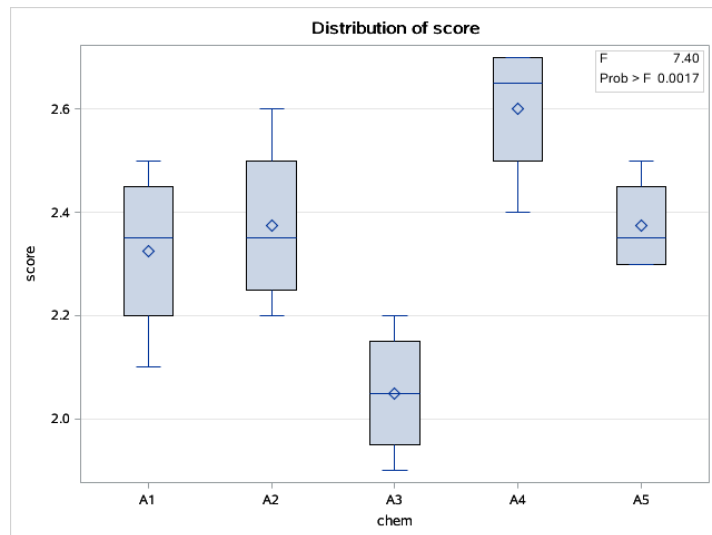
因变量: score

源	自由度	平方和	均方	F 值	Pr > F
模型	4	0.61700000	0.15425000	7.40	0.0017
误差	15	0.31250000	0.02083333		
校正合计	19	0.92950000			

R 方	变异系数	均方根误差	score 均值
0.663798	6.155120	0.144338	2.345000

源	自由度	I 型 平方和	均方	F 值	Pr > F
chem	4	0.61700000	0.15425000	7.40	0.0017

源	自由度	III 型 平方和	均方	F 值	Pr > F
chem	4	0.61700000	0.15425000	7.40	0.0017



由 ANOVA 可知, p-value 较小, 故认为存在显著差异; 又有结果可知 A4 的效果最好。

第三题

```
data ad_data;
input ad$ shop$ sale @@;
datalines;
A1 B1 12 A1 B1 14 A1 B2 18 A1 B2 16 A1 B3 17 A1 B3 17
A2 B1 14 A2 B1 15 A2 B2 12 A2 B2 14 A2 B3 13 A2 B3 15
A3 B1 19 A3 B1 20 A3 B2 17 A3 B2 18 A3 B3 21 A3 B3 22
A4 B1 24 A4 B1 23 A4 B2 30 A4 B2 28 A4 B3 29 A4 B3 28
;
run;
proc print data=ad_data;
run;
proc glm data=ad_data;
class ad shop;
model sale=ad shop ad*shop / alpha=0.05;
run;
```



```
proc anova data=ad_data;
class ad;
model sale=ad;
run;
```

SAS 系统

GLM 过程

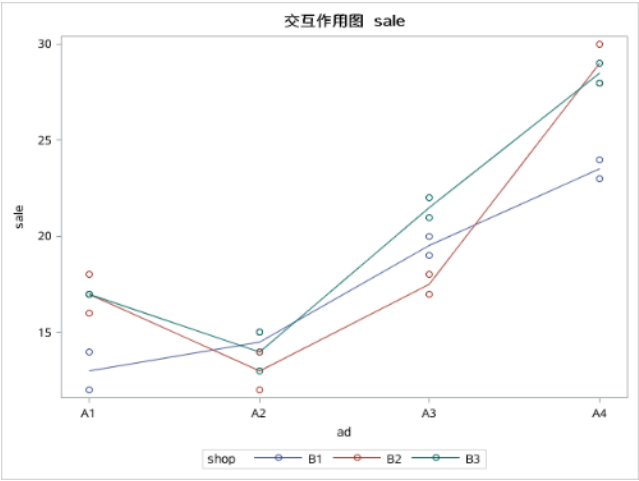
因变量: sale

源	自由度	平方和	均方	F 值	Pr > F
模型	11	689.0000000	62.6363636	57.82	<.0001
误差	12	13.0000000	1.0833333		
校正合计	23	702.0000000			

R 方	变异系数	均方根误差	sale 均值
0.981481	5.478068	1.040833	19.00000

源	自由度	I 型 平方和	均方	F 值	Pr > F
ad	3	612.3333333	204.1111111	188.41	<.0001
shop	2	27.7500000	13.8750000	12.81	0.0011
ad*shop	6	48.9166667	8.1527778	7.53	0.0016

源	自由度	III 型 平方和	均方	F 值	Pr > F
ad	3	612.3333333	204.1111111	188.41	<.0001
shop	2	27.7500000	13.8750000	12.81	0.0011
ad*shop	6	48.9166667	8.1527778	7.53	0.0016



SAS 系统

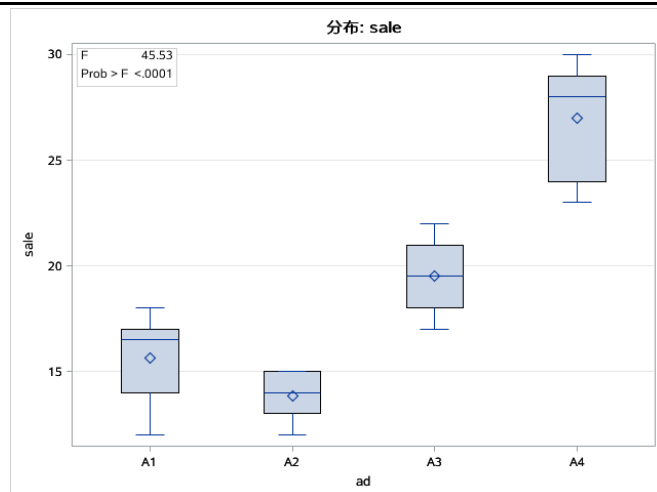
ANOVA 过程

因变量: sale

源	自由度	平方和	均方	F 值	Pr > F
模型	3	612.3333333	204.1111111	45.53	<.0001
误差	20	89.6666667	4.4833333		
校正合计	23	702.0000000			

R 方	变异系数	均方根误差	sale 均值
0.872270	11.14415	2.117388	19.00000

源	自由度	Anova 平方和	均方	F 值	Pr > F
ad	3	612.3333333	204.1111111	45.53	<.0001



SAS 系统

GLM 过程 最小二乘均值

ad*shop 按 shop 划分的效应: sale					
shop	自由度	平方和	均方	F 值	Pr > F
B1	3	138.375000	46.125000	42.58	<.0001
B2	3	284.375000	94.791667	87.50	<.0001
B3	3	238.500000	79.500000	73.38	<.0001

由方差分析表可知, $p\text{-value} < 0.05$, 有理由拒绝原假设, 认为由明显差异, 交互作用显著; 由 ANOVA 及最小二乘过程可知, 文案对销量的影响显著。

第四题 (1)

```
LIBNAME HW "C:\Users\31949\Desktop\sas assignment 4";
/*Problem 1*/
data st1;
set hw.stock;
year=year(Date);
month=month(Date);
day=day(Date);
run;
data st2;
set st1;
by year month;
last_day=last.month;
if last_day ne 1 then delete;
if year<1995 or year>1998 then delete;
run;
data st_income;
set st2;
retain now_close;
last_close=now_close;
now_close=clpr;
income_rate=(now_close-last_close)/last_close;
```

SAS 系统		
Obs	date	income_rate
1	1995-02	-0.07227
2	1995-03	0.13406
3	1995-04	-0.15495
4	1995-05	0.15501
5	1995-06	-0.13421
6	1995-07	0.11531
7	1995-08	-0.04068
8	1995-09	-0.06360
9	1995-10	0.08302
10	1995-11	0.02265
11	1995-12	-0.16525
12	1996-01	-0.06122
13	1996-02	0.02609
14	1996-03	-0.02119
15	1996-04	0.25541
16	1996-05	-0.15517
17	1996-06	0.00612
18	1996-07	0.28803
19	1996-08	-0.11811
20	1996-09	0.12857
21	1996-10	0.12975
22	1996-11	0.01821
23	1996-12	-0.28473
24	1997-01	0.05962
25	1997-02	0.23049
26	1997-03	0.10619
27	1997-04	0.07600
28	1997-05	-0.21933
29	1997-06	-0.11111
30	1997-07	-0.02321
31	1997-08	0.09506
32	1997-09	-0.12187
33	1997-10	0.07414
34	1997-11	-0.03540
35	1997-12	0.15046
36	1998-01	0.11962
37	1998-02	-0.00712
38	1998-03	0.16930
39	1998-04	0.01104
40	1998-05	0.57524
41	1998-06	-0.19337
42	1998-07	0.00955
43	1998-08	-0.12961
44	1998-09	0.34783
45	1998-10	-0.00806
46	1998-11	-0.02033
47	1998-12	-0.18672

```
/*Problem 2*/  
data st3;  
set st1;  
by year month;  
keep date year month vol clpr mvol t_clpr tag;  
if last.month = 1 then do;  
    t_clpr=clpr;  
    tag=1;  
end;  
else t_clpr=0;
```

```
/*Problem 2*/  
  
data st3;  
set st1;  
by year month;  
keep date year month vol clpr mvol t_clpr tag;  
if last.month = 1 then do;  
    t_clpr=clpr;  
    tag=1;  
end;  
else t clpr=0;
```

```

retain vol;
if first.month =1 then do ;
    mvol=0;
    tag=-1;
end;
mvol+vol;
run;
proc print data=st3;
run;
data st_month;
set st3;
if tag ne 1 then delete;
rename t_clpr=clpr;
keep mvol clpr date;
format date yymmdd.;
run;
proc print data=st_month;
run;

```

SAS 系统

Obs	Clpr	date	mvol
1	5.95	1995-01	36103
2	5.52	1995-02	39198
3	6.26	1995-03	95943
4	5.29	1995-04	49888
5	6.11	1995-05	144985
6	5.29	1995-06	36395
7	5.90	1995-07	59670
8	5.66	1995-08	127337
9	5.30	1995-09	41928
10	5.74	1995-10	63514
11	5.87	1995-11	213047
12	4.90	1995-12	44500
13	4.60	1996-01	37709
14	4.72	1996-02	12933
15	4.62	1996-03	41763
16	5.80	1996-04	134694
17	4.90	1996-05	43091
18	4.93	1996-06	71234
19	6.35	1996-07	291629

20	5.60	1996-08	155051
21	6.32	1996-09	155406
22	7.14	1996-10	357833
23	7.27	1996-11	149754
24	5.20	1996-12	269668
25	5.51	1997-01	69025
26	6.78	1997-02	167234
27	7.50	1997-03	240814
28	8.07	1997-04	376287
29	6.30	1997-05	133065
30	5.60	1997-06	57333
31	5.47	1997-07	63463
32	5.99	1997-08	67755
33	5.26	1997-09	107970
34	5.65	1997-10	51174
35	5.45	1997-11	56593
36	6.27	1997-12	69924
37	7.02	1998-01	135072
38	6.97	1998-02	61101
39	8.15	1998-03	140874
40	8.24	1998-04	276311
41	12.98	1998-05	397593
42	10.47	1998-06	171013
43	10.57	1998-07	74212

45	12.40	1998-09	208931
46	12.30	1998-10	156270
47	12.05	1998-11	104098
48	9.80	1998-12	32698
49	10.40	1999-01	51990
50	9.50	1999-02	10540
51	11.54	1999-03	131536
52	11.69	1999-04	110820
53	10.18	1999-05	125254
54	16.20	1999-06	471971
55	15.28	1999-07	119298
56	14.99	1999-08	39250
57	14.60	1999-09	76293
58	14.51	1999-10	41691
59	14.05	1999-11	50809
60	13.30	1999-12	23215
61	11.76	2000-01	57225
62	14.58	2000-02	203743
63	15.08	2000-03	196016
64	15.12	2000-04	181213
65	15.78	2000-05	84102
66	16.86	2000-06	187732
67	18.00	2000-07	138533
68	16.70	2000-08	120375
69	15.18	2000-09	29032

第四题 (3)

```

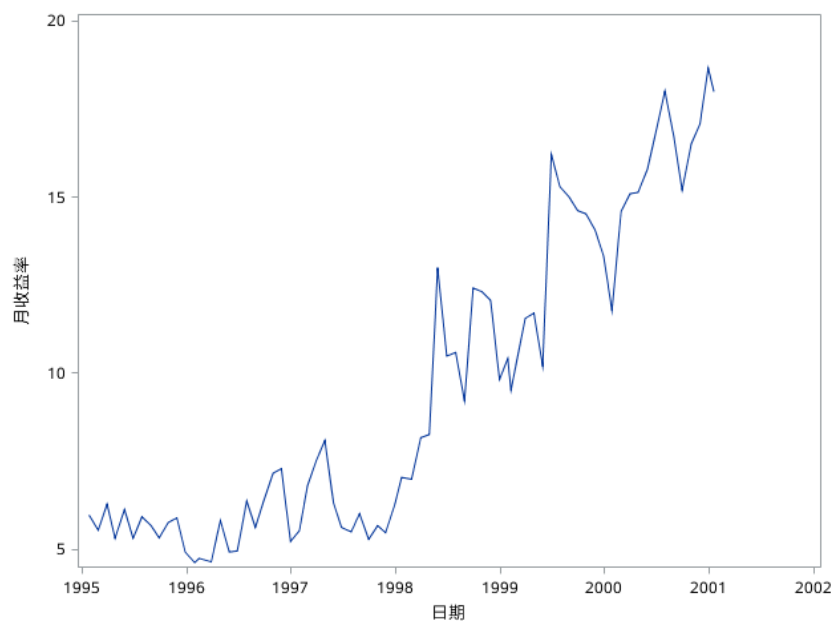
/*Problem 3*/
proc sgplot data=st_month;
series x=date y=clpr;

```

```

xaxis label="日期";
yaxis label="月收益率";
run;

```

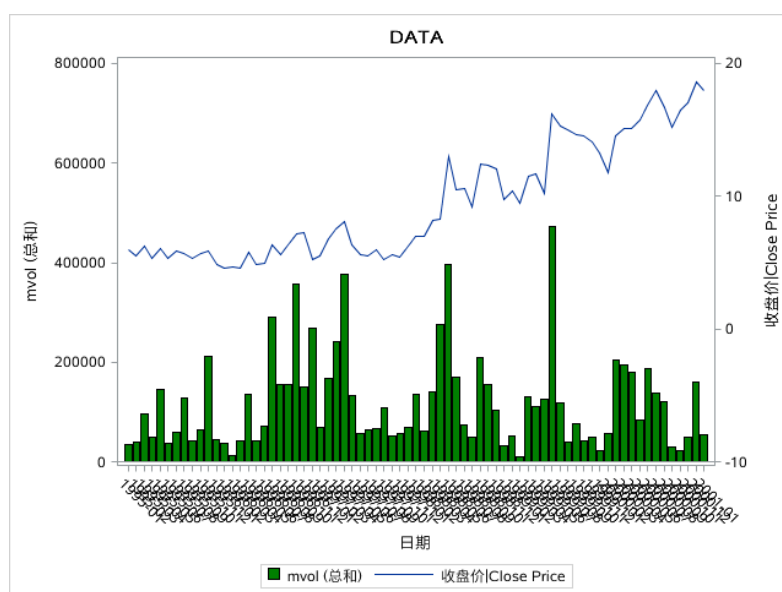


第四题 (4)

```

/*Problem 4*/
proc sgplot data=st_month;
title"DATA";
vbar date / response=mvol fillattrs=(color=green);
vline date / response=clpr y2axis;
xaxis label="日期";
yaxis values=(0,200000,400000,600000,800000);
y2axis values=(-10,0,10,20);
run;

```



【小结】

本次实验主要的实验内容为统计学中的 ANOVA 模型等一系列分析模型。通过本次实验对于模型的各项指标有了更深的了解,对于各个模型的适用情况有了更新的认识,同时也增加了自己的编程熟练程度。希望通过后面的练习继续增强 SAS 的编程能力。

指导教师评语及成绩:

成绩:

指导教师签名:

批阅日期:

附件:

实验报告说明

1. **实验项目名称:** 要用最简练的语言反映实验的内容。
2. **实验类型:** 一般需说明是验证型实验还是设计型实验、综合型实验或其他实验。
3. **实验目的与要求:** 目的要明确,要抓住重点。
4. **实验原理:** 简要说明本实验项目所涉及的理论知识。
5. **实验环境:** 实验用的软硬件环境(配置)。
6. **实验方案设计(思路、步骤和方法等):** 这是实验报告极其重要的内容,概括整个实验过程。

对于**验证型实验**,要写明依据何种原理、何种操作方法进行实验,并写明需要经过哪几个步骤。

对于**设计型和综合型实验**,在上述内容基础上还应该画出流程图、设计思路和设计方法,再配以相应的文字说明。

7. **实验过程(实验中涉及的记录、数据、分析):** 写明具体上述实验方案的具体实施,包括实验过程中的记录、数据和相应的分析。
8. **结论(结果):** 即根据实验过程中所见到的现象和测得的数据,得出结论。
9. **小结:** 对本次实验的心得体会、思考和建议。
10. **指导教师评语及成绩:** 指导教师依据学生的实际报告内容,用简练语言给出本次实验报告的评价和价值。