

实验概述：

【实验目的及要求】

实验目的：

1. 掌握 SAS 程序设计结构和常规编程技术，并做数值计算应用；
2. 理解 INPUT 语句进阶功能和操作；
3. 掌握 SAS 数组，并且能够结合循环实现数据转置等操作；
4. 掌握通过 SAS 编程读取各种类型外部文件中的数据，然后再对其进行整理与合并的方法。

【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.

一、原始数据文件 sales.data 包含了某季度创下的最高销售记录。具体数据格式如下：

sales.data

		1	1	2	2	3		
1	---	5	---	0	---	5	---	0
101	USA	1-20-2007	3290					
3034	EUR	30JAN2007	1876,30					
101	USA	1-30-2007	2938.00					
128	USA	2-5-2007	2908.74					
1345	EUR	6FEB2007	3145,60					
109	USA	3-17-2007	2789.10					

但不是所有记录都有相同格式，其中美国和欧洲的销售数据的小数点和逗号是相反的，并且销售日期格式也不同，由原始数据创建 SAS 数据集 SalesQ1 并编写程序实现期望输出如下：

Sale ID	Location	<u>Sale</u> Date	Amount
101	USA	17186	3290
3034	EUR	17196	1876.30
101	USA	17196	2938.00
128	USA	17202	2908.74
1345	EUR	17203	3145.60
109	USA	17242	2789.10

二、编程训练：

编制日历表，比如给定 2020 年，在结果查看器输出 12 个月度的日历。

三、（1）已知我国十大城市 2003 年的第三产业增加值数据（变量名为 TIndustry，单位万元）分别存放在 SAS 数据集 RepData1a 和 RepData1b 中，具体如下：

RepData1a:

City	TIndustry
北京	22529400
广州	18788487
杭州	8838612
宁波	6734132
上海	30271100

RepData1b:

City	TIndustry
深圳	11552772
苏州	9442100
天津	11127100
无锡	7499400
重庆	9369000

请通过编写 SAS 程序，将上述两个数据合并成一个名为 RepData1 的 SAS 数据集，数据集 RepData1 中仍只含有 City 和 TIndustry 两个变量。

（2）请按照 RepData1 中的十大城市名，将（1）中的 SAS 数据集 RepData1 和 RepData2 中的数据进行合并，并计算每个城市的第二、第三产业增加值之和，将其存放在变量 STPopulation 中。同时计算每个城市的人均第二第三产业总增加值，将其存放在变量 ASTPopulation 中。再计算这十大城市的第二第三产业总增加值的总量，将其存放在变量 TSTPopulation 中。上述数据均输出到名为 RepData3 的 SAS 数据集中。

四、现有 5 个城市 2019 年末人口总数（万人）和失业率数据存放在名为 data1 的 Excel 文件的 sheet1 和 sheet2 中，其中文件的第一行是变量名：id 代表序号，City 代表城市，Population 代表人口数目，UNERate 代表失业率。请通过编写 SAS 程序，在不改变上述数据的数据结构前提下将这两部分数据读入，根据 id 将这两部分数据合并生成一个新的数据集并打印。

五、创建一个数据集 data1，包含变量 id, name, sex, 再创建一个数据集 data2，包含变量 id, subject, score, 即数据集 data1 中包含的是学生的个人信息，数据集 data2 中包含的是学生的学科得分,利用变量 id 将两个数据集合并起来且按照 id 从小到大的顺序排列，得到新的数据集 data3,打印合并结果，并且分析数据集合并的过程以及合并的过程中需要注意的事项以及可能出现的问题。

data1		
id	name	sex
1005	Tom	M
1003	Lily	F
1001	Alice	F
1002	Jack	M

data2		
id	subject	score
1001	Chinese	95
1003	math	84
1001	English	90
1002	history	92

六、用字符型数组进行表查找

公司安全部门希望所有员工都知道零售店关于顾客服务的新政策和流程。其中每个员工必须参与网络培训，然后参加包括 10 个问题的选择题测试，每个问题都有 5 个选择（A-E），每个测试结果输入到 SAS 数据集 test_answers 中，数据集中每个观测包含一个人的答案。test_answers 列表：

Employee_ ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
121044	A	C	C	B	D	E	D	B	B	A
120145	B	C	C		E	E	D	B	A	A
120761	A	C	C	B	D	D	E	B	B	C
120656	B	C	C	A	D	B	B	C	A	D
121107	A	C	C	B	E	E	D	B	B	A
121038	B	C	C	B	D	D	D	B	B	A
120273	C	C	C	B	E	E	E	B	B	A
120759	A	C	C	B	E	E	D	B	B	A
120798		A	C	B	D	D	D	B	B	A
121030	C	C	C	C	E	E	D	B	B	B
121017	B	B	E	B	E	E	D	B	B	A
121062	A	C	C	B	E	E	D	B	B	A
121119	C	C	C	B	E	E	D	B	B	A
120812	A	C	C	B	E	E	E	B	B	A
120756	A	C	C	B	E	E	D	B	B	A

其中正确答案如下：

问题:	1	2	3	4	5	6	7	8	9	10
答案:	A	C	C	B	E	E	D	B	B	A

（1）阅读 test_answers 并确定哪些员工通过，哪些没有。

提示：计算变量 score，这个变量包含了每个员工正确答案的个数；如果员工得分大于等于 7，那么这个观测写入到数据集 passed 中，否则保存到数据集 failed 中。

（2）打印 passed 数据集来验证 12 个观测。

PROC PRINT 输出：

Passed												
Obs	Employee_ ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Score
1	121044	A	C	C	B	D	E	D	B	B	A	9
2	120145	B	C	C		E	E	D	B	A	A	7
3	121107	A	C	C	B	E	E	D	B	B	A	10
4	121038	B	C	C	B	D	D	D	B	B	A	7
5	120273	C	C	C	B	E	E	E	B	B	A	8
6	120759	A	C	C	B	E	E	D	B	B	A	10
7	121030	C	C	C	C	E	E	D	B	B	B	7
8	121017	B	B	E	B	E	E	D	B	B	A	7
9	121062	A	C	C	B	E	E	D	B	B	A	10
10	121119	C	C	C	B	E	E	D	B	B	A	9
11	120812	A	C	C	B	E	E	E	B	B	A	9
12	120756	A	C	C	B	E	E	D	B	B	A	10

（3）打印 failed 数据集，验证该数据集包含三个观测。

PROC PRINT 输出：

Failed												
Obs	Employee_ ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Score
1	120761	A	C	C	B	D	D	E	B	B	C	6
2	120656	B	C	C	A	D	B	B	C	A	D	2
3	120798		A	C	B	D	D	D	B	B	A	6

实验内容:

【实验方案设计】

【实验过程】(实验步骤、记录、数据、分析)

实验一	
实验代码	<pre>data sales_data; input Sale_ID Location\$ Sale_Date_r\$10-19 Amount_r\$; datalines; 101 USA 1-20-2007 3295 3034 EUR 30JAN2007 1876,30 101 USA 1-30-2007 2938.00 128 USA 2-5-2007 2908.74 1345 EUR 6FEB2007 3145,60 109 USA 3-17-2007 2789.10 ; run; data sales; set sales_data; format Sale_Date 5.0 Amount 7.2; drop Sale_Date_r Amount_r; if Location="USA" and substr(Amount_r,-1,1) ne '.' then do; Sale_Date=input(Sale_Date_r,mmddy10.); Amount=input(Amount_r,7.0); output; end; else if Location="USA" then do; Sale_Date=input(Sale_Date_r,mmddy10.); Amount=input(Amount_r, 7.2); output; end; else if Location="EUR" then do; Sale_Date=input(Sale_Date_r,date10.); Amount=input(Amount_r,commax7.2); output; end; run;</pre>

	<pre>title "Answer to Question 1"; proc print data=sales; run; title;</pre>																																			
实验结果	<div><div>Answer to Question 1</div><table><tr><th>Obs</th><th>Sale_ID</th><th>Location</th><th>Sale_Date</th><th>Amount</th></tr><tr><td>1</td><td>101</td><td>USA</td><td>17186</td><td>3295.00</td></tr><tr><td>2</td><td>3034</td><td>EUR</td><td>17196</td><td>1876.30</td></tr><tr><td>3</td><td>101</td><td>USA</td><td>17196</td><td>2938.00</td></tr><tr><td>4</td><td>128</td><td>USA</td><td>17202</td><td>2908.74</td></tr><tr><td>5</td><td>1345</td><td>EUR</td><td>17203</td><td>3145.60</td></tr><tr><td>6</td><td>109</td><td>USA</td><td>17242</td><td>2789.10</td></tr></table></div>	Obs	Sale_ID	Location	Sale_Date	Amount	1	101	USA	17186	3295.00	2	3034	EUR	17196	1876.30	3	101	USA	17196	2938.00	4	128	USA	17202	2908.74	5	1345	EUR	17203	3145.60	6	109	USA	17242	2789.10
Obs	Sale_ID	Location	Sale_Date	Amount																																
1	101	USA	17186	3295.00																																
2	3034	EUR	17196	1876.30																																
3	101	USA	17196	2938.00																																
4	128	USA	17202	2908.74																																
5	1345	EUR	17203	3145.60																																
6	109	USA	17242	2789.10																																
实验分析	<div><div><div>1. 通过 DATA INPUT … DATALINES 语句在 SAS 程序中流输入相应数据；其中字符型变量后面要以\$标识，例如：Location\$；</div><div>2. 通过 DATA FORMAT 语句对进行处理的语句设置相应的数据格式；</div><div>3. 通过 INPUT(input_data, informat)语句对数据的某列变量的格式进行修改；</div><div>4. 对于不同种类的 numeric 数据，包括日期数据，通过 informat 的参数不同设置可以分别进行识别与处理，例如 DATAw. 为形如 01JAN2020 的数据，且 w. 的长度可以“向下兼容”。</div></div></div>																																			

实验二	
实验代码	<pre> data cal_date; format cdate date9.; cdate='01Jan2020'd; do cdate='01Jan2020'd to '31Dec2020'd; cweekday=weekday(cdate); cmonth=month(cdate); output; end; run; data calendar; set cal_date; array week_day(7) Sunday Monday Tuesday Wednesday Thursday Friday Saturday; retain Sunday Monday Tuesday Wednesday Thursday Friday Saturday; drop cdate cweekday cmonth; format cdate date9. Sunday Monday Tuesday Wednesday Thursday Friday Saturday date5.; /*do cdate = '01Jan2020'd to '31Dec2020'd;*/ week_day(WEEKDAY(cdate))=cdate; </pre>

	<pre>if(WEEKDAY(cdate)=7 DAY(cdate+1)=1) then do; output; call missing (of week_day{*}); end; /*end;*/ run; title "Calendar 2020"; proc print data=calendar noobs; run; title;</pre>																																																																																																																																																																																																																									
实验结果	<table><tr><th>Sunday</th><th>Monday</th><th>Tuesday</th><th>Wednesday</th><th>Thursday</th><th>Friday</th><th>Saturday</th></tr><tr><td>.</td><td>.</td><td>.</td><td>01Jan</td><td>02Jan</td><td>03Jan</td><td>04Jan</td></tr><tr><td>05Jan</td><td>06Jan</td><td>07Jan</td><td>08Jan</td><td>09Jan</td><td>10Jan</td><td>11Jan</td></tr><tr><td>12Jan</td><td>13Jan</td><td>14Jan</td><td>15Jan</td><td>16Jan</td><td>17Jan</td><td>18Jan</td></tr><tr><td>19Jan</td><td>20Jan</td><td>21Jan</td><td>22Jan</td><td>23Jan</td><td>24Jan</td><td>25Jan</td></tr><tr><td>26Jan</td><td>27Jan</td><td>28Jan</td><td>29Jan</td><td>30Jan</td><td>31Jan</td><td>.</td></tr><tr><td>.</td><td>.</td><td>.</td><td>.</td><td>.</td><td>.</td><td>01Feb</td></tr><tr><td>02Feb</td><td>03Feb</td><td>04Feb</td><td>05Feb</td><td>06Feb</td><td>07Feb</td><td>08Feb</td></tr><tr><td>09Feb</td><td>10Feb</td><td>11Feb</td><td>12Feb</td><td>13Feb</td><td>14Feb</td><td>15Feb</td></tr><tr><td>16Feb</td><td>17Feb</td><td>18Feb</td><td>19Feb</td><td>20Feb</td><td>21Feb</td><td>22Feb</td></tr><tr><td>23Feb</td><td>24Feb</td><td>25Feb</td><td>26Feb</td><td>27Feb</td><td>28Feb</td><td>29Feb</td></tr><tr><td>01Mar</td><td>02Mar</td><td>03Mar</td><td>04Mar</td><td>05Mar</td><td>06Mar</td><td>07Mar</td></tr><tr><td>08Mar</td><td>09Mar</td><td>10Mar</td><td>11Mar</td><td>12Mar</td><td>13Mar</td><td>14Mar</td></tr><tr><td>15Mar</td><td>16Mar</td><td>17Mar</td><td>18Mar</td><td>19Mar</td><td>20Mar</td><td>21Mar</td></tr><tr><td>22Mar</td><td>23Mar</td><td>24Mar</td><td>25Mar</td><td>26Mar</td><td>27Mar</td><td>28Mar</td></tr><tr><td>29Mar</td><td>30Mar</td><td>31Mar</td><td>.</td><td>.</td><td>.</td><td>.</td></tr><tr><td>.</td><td>.</td><td>.</td><td>01Apr</td><td>02Apr</td><td>03Apr</td><td>04Apr</td></tr><tr><td>05Apr</td><td>06Apr</td><td>07Apr</td><td>08Apr</td><td>09Apr</td><td>10Apr</td><td>11Apr</td></tr><tr><td>12Apr</td><td>13Apr</td><td>14Apr</td><td>15Apr</td><td>16Apr</td><td>17Apr</td><td>18Apr</td></tr><tr><td>19Apr</td><td>20Apr</td><td>21Apr</td><td>22Apr</td><td>23Apr</td><td>24Apr</td><td>25Apr</td></tr><tr><td>26Apr</td><td>27Apr</td><td>28Apr</td><td>29Apr</td><td>30Apr</td><td>.</td><td>.</td></tr><tr><td>.</td><td>.</td><td>.</td><td>.</td><td>.</td><td>01May</td><td>02May</td></tr><tr><td>03May</td><td>04May</td><td>05May</td><td>06May</td><td>07May</td><td>08May</td><td>09May</td></tr><tr><td>10May</td><td>11May</td><td>12May</td><td>13May</td><td>14May</td><td>15May</td><td>16May</td></tr><tr><td>17May</td><td>18May</td><td>19May</td><td>20May</td><td>21May</td><td>22May</td><td>23May</td></tr><tr><td>24May</td><td>25May</td><td>26May</td><td>27May</td><td>28May</td><td>29May</td><td>30May</td></tr><tr><td>31May</td><td>.</td><td>.</td><td>.</td><td>.</td><td>.</td><td>.</td></tr><tr><td>.</td><td>01Jun</td><td>02Jun</td><td>03Jun</td><td>04Jun</td><td>05Jun</td><td>06Jun</td></tr><tr><td>07Jun</td><td>08Jun</td><td>09Jun</td><td>10Jun</td><td>11Jun</td><td>12Jun</td><td>13Jun</td></tr><tr><td>14Jun</td><td>15Jun</td><td>16Jun</td><td>17Jun</td><td>18Jun</td><td>19Jun</td><td>20Jun</td></tr><tr><td>21Jun</td><td>22Jun</td><td>23Jun</td><td>24Jun</td><td>25Jun</td><td>26Jun</td><td>27Jun</td></tr></table>	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	.	.	.	01Jan	02Jan	03Jan	04Jan	05Jan	06Jan	07Jan	08Jan	09Jan	10Jan	11Jan	12Jan	13Jan	14Jan	15Jan	16Jan	17Jan	18Jan	19Jan	20Jan	21Jan	22Jan	23Jan	24Jan	25Jan	26Jan	27Jan	28Jan	29Jan	30Jan	31Jan	01Feb	02Feb	03Feb	04Feb	05Feb	06Feb	07Feb	08Feb	09Feb	10Feb	11Feb	12Feb	13Feb	14Feb	15Feb	16Feb	17Feb	18Feb	19Feb	20Feb	21Feb	22Feb	23Feb	24Feb	25Feb	26Feb	27Feb	28Feb	29Feb	01Mar	02Mar	03Mar	04Mar	05Mar	06Mar	07Mar	08Mar	09Mar	10Mar	11Mar	12Mar	13Mar	14Mar	15Mar	16Mar	17Mar	18Mar	19Mar	20Mar	21Mar	22Mar	23Mar	24Mar	25Mar	26Mar	27Mar	28Mar	29Mar	30Mar	31Mar	01Apr	02Apr	03Apr	04Apr	05Apr	06Apr	07Apr	08Apr	09Apr	10Apr	11Apr	12Apr	13Apr	14Apr	15Apr	16Apr	17Apr	18Apr	19Apr	20Apr	21Apr	22Apr	23Apr	24Apr	25Apr	26Apr	27Apr	28Apr	29Apr	30Apr	01May	02May	03May	04May	05May	06May	07May	08May	09May	10May	11May	12May	13May	14May	15May	16May	17May	18May	19May	20May	21May	22May	23May	24May	25May	26May	27May	28May	29May	30May	31May	01Jun	02Jun	03Jun	04Jun	05Jun	06Jun	07Jun	08Jun	09Jun	10Jun	11Jun	12Jun	13Jun	14Jun	15Jun	16Jun	17Jun	18Jun	19Jun	20Jun	21Jun	22Jun	23Jun	24Jun	25Jun	26Jun	27Jun
Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday																																																																																																																																																																																																																				
.	.	.	01Jan	02Jan	03Jan	04Jan																																																																																																																																																																																																																				
05Jan	06Jan	07Jan	08Jan	09Jan	10Jan	11Jan																																																																																																																																																																																																																				
12Jan	13Jan	14Jan	15Jan	16Jan	17Jan	18Jan																																																																																																																																																																																																																				
19Jan	20Jan	21Jan	22Jan	23Jan	24Jan	25Jan																																																																																																																																																																																																																				
26Jan	27Jan	28Jan	29Jan	30Jan	31Jan	.																																																																																																																																																																																																																				
.	01Feb																																																																																																																																																																																																																				
02Feb	03Feb	04Feb	05Feb	06Feb	07Feb	08Feb																																																																																																																																																																																																																				
09Feb	10Feb	11Feb	12Feb	13Feb	14Feb	15Feb																																																																																																																																																																																																																				
16Feb	17Feb	18Feb	19Feb	20Feb	21Feb	22Feb																																																																																																																																																																																																																				
23Feb	24Feb	25Feb	26Feb	27Feb	28Feb	29Feb																																																																																																																																																																																																																				
01Mar	02Mar	03Mar	04Mar	05Mar	06Mar	07Mar																																																																																																																																																																																																																				
08Mar	09Mar	10Mar	11Mar	12Mar	13Mar	14Mar																																																																																																																																																																																																																				
15Mar	16Mar	17Mar	18Mar	19Mar	20Mar	21Mar																																																																																																																																																																																																																				
22Mar	23Mar	24Mar	25Mar	26Mar	27Mar	28Mar																																																																																																																																																																																																																				
29Mar	30Mar	31Mar																																																																																																																																																																																																																				
.	.	.	01Apr	02Apr	03Apr	04Apr																																																																																																																																																																																																																				
05Apr	06Apr	07Apr	08Apr	09Apr	10Apr	11Apr																																																																																																																																																																																																																				
12Apr	13Apr	14Apr	15Apr	16Apr	17Apr	18Apr																																																																																																																																																																																																																				
19Apr	20Apr	21Apr	22Apr	23Apr	24Apr	25Apr																																																																																																																																																																																																																				
26Apr	27Apr	28Apr	29Apr	30Apr	.	.																																																																																																																																																																																																																				
.	01May	02May																																																																																																																																																																																																																				
03May	04May	05May	06May	07May	08May	09May																																																																																																																																																																																																																				
10May	11May	12May	13May	14May	15May	16May																																																																																																																																																																																																																				
17May	18May	19May	20May	21May	22May	23May																																																																																																																																																																																																																				
24May	25May	26May	27May	28May	29May	30May																																																																																																																																																																																																																				
31May																																																																																																																																																																																																																				
.	01Jun	02Jun	03Jun	04Jun	05Jun	06Jun																																																																																																																																																																																																																				
07Jun	08Jun	09Jun	10Jun	11Jun	12Jun	13Jun																																																																																																																																																																																																																				
14Jun	15Jun	16Jun	17Jun	18Jun	19Jun	20Jun																																																																																																																																																																																																																				
21Jun	22Jun	23Jun	24Jun	25Jun	26Jun	27Jun																																																																																																																																																																																																																				

[illegible]

	<p>尾作为一行的结束，下一天的日期输出要换行；因此在日期达成上述两种条件之一时要通过 OUTPUT 输出并换行；</p> <p>6. 同时为了保证数据在后面的读取时不被覆盖，需要加上 RETAIN 语句，以保证在对原始数据的每一个 obs 读入后不再抹除该数据；</p> <p>7. 此外由于 DATA 步自身在运行时便包含循环结构，因此在进行输出转化时不需要额外再进行循环操作。</p>
实验三	
(3.1)	
实验代码	<pre> data RepData1a; input City\$ TIndustry; datalines; 北京 22529400 广州 18788487 杭州 8838612 宁波 6734132 上海 30271100 ; run; data RepData1b; input City\$ TIndustry; datalines; 深圳 11552772 苏州 9442100 天津 11127100 无锡 7499400 重庆 9369000 ; run; data RepData1; set RepData1a RepData1b; run; title "RepData1"; proc print data=RepData1; run; title; </pre>
实验结果	

	<div>RepData1</div> <table><tr><th>Obs</th><th>City</th><th>TIndustry</th></tr><tr><td>1</td><td>北京</td><td>22529400</td></tr><tr><td>2</td><td>广州</td><td>18788487</td></tr><tr><td>3</td><td>杭州</td><td>8838612</td></tr><tr><td>4</td><td>宁波</td><td>6734132</td></tr><tr><td>5</td><td>上海</td><td>30271100</td></tr><tr><td>6</td><td>深圳</td><td>11552772</td></tr><tr><td>7</td><td>苏州</td><td>9442100</td></tr><tr><td>8</td><td>天津</td><td>11127100</td></tr><tr><td>9</td><td>无锡</td><td>7499400</td></tr><tr><td>10</td><td>重庆</td><td>9369000</td></tr></table>	Obs	City	TIndustry	1	北京	22529400	2	广州	18788487	3	杭州	8838612	4	宁波	6734132	5	上海	30271100	6	深圳	11552772	7	苏州	9442100	8	天津	11127100	9	无锡	7499400	10	重庆	9369000
Obs	City	TIndustry																																
1	北京	22529400																																
2	广州	18788487																																
3	杭州	8838612																																
4	宁波	6734132																																
5	上海	30271100																																
6	深圳	11552772																																
7	苏州	9442100																																
8	天津	11127100																																
9	无锡	7499400																																
10	重庆	9369000																																
(3.2)																																		
实验代码	<pre>proc import datafile="C:\Users\31949\iCloudDrive\By's Cloud\SUFE-lectures\大三\统计软件\Assignment\hw2\RepData2.xlsx" out=RepData2 dbms=excel replace; run; proc sort data=Repdata1 out=Repdata1_sort; by City; run; proc sort data=Repdata2 out=Repdata2_sort; by City; run; data data_collection; merge Repdata1_sort(IN=InT1) Repdata2_sort(IN=InT2); by City; if (InT1=1 and InT2=1) then do; output; end; run; data RepData3; set data_collection; STPopulation=TIndustry+SIndustry; ASTPopulation=STPopulation/TPopulation; TSTPopulation+STPopulation; output;</pre>																																	

	<pre>keep City TIndustry SIndusty TPopulation STPopulation ASTPopulation TSTPopulation; run; title "RepData3"; proc print data=RepData3; run; title;</pre>																																																																													
实验结果	<div>RepData3</div> <table><tr><th>Obs</th><th>City</th><th>TIndustry</th><th>TPopulation</th><th>STPopulation</th><th>ASTPopulation</th><th>TSTPopulation</th></tr><tr><td>1</td><td>北京</td><td>22529400</td><td>1079.22</td><td>35648000</td><td>33031.26</td><td>35648000</td></tr><tr><td>2</td><td>广州</td><td>18788487</td><td>588.26</td><td>33869707</td><td>57576.08</td><td>69517707</td></tr><tr><td>3</td><td>杭州</td><td>8838612</td><td>393.19</td><td>19731854</td><td>50184.02</td><td>89249561</td></tr><tr><td>4</td><td>宁波</td><td>6734132</td><td>206.91</td><td>16746270</td><td>80935.04</td><td>105995831</td></tr><tr><td>5</td><td>上海</td><td>30271100</td><td>1278.23</td><td>61578300</td><td>48174.66</td><td>167574131</td></tr><tr><td>6</td><td>深圳</td><td>11552772</td><td>150.93</td><td>28789326</td><td>190746.21</td><td>196363457</td></tr><tr><td>7</td><td>苏州</td><td>9442100</td><td>216.87</td><td>27160700</td><td>125239.54</td><td>223524157</td></tr><tr><td>8</td><td>天津</td><td>11127100</td><td>758.78</td><td>23580000</td><td>31076.20</td><td>247104157</td></tr><tr><td>9</td><td>无锡</td><td>7499400</td><td>219.60</td><td>18429600</td><td>83923.50</td><td>265533757</td></tr><tr><td>10</td><td>重庆</td><td>9369000</td><td>1010.12</td><td>19142000</td><td>18950.22</td><td>284675757</td></tr></table>	Obs	City	TIndustry	TPopulation	STPopulation	ASTPopulation	TSTPopulation	1	北京	22529400	1079.22	35648000	33031.26	35648000	2	广州	18788487	588.26	33869707	57576.08	69517707	3	杭州	8838612	393.19	19731854	50184.02	89249561	4	宁波	6734132	206.91	16746270	80935.04	105995831	5	上海	30271100	1278.23	61578300	48174.66	167574131	6	深圳	11552772	150.93	28789326	190746.21	196363457	7	苏州	9442100	216.87	27160700	125239.54	223524157	8	天津	11127100	758.78	23580000	31076.20	247104157	9	无锡	7499400	219.60	18429600	83923.50	265533757	10	重庆	9369000	1010.12	19142000	18950.22	284675757
Obs	City	TIndustry	TPopulation	STPopulation	ASTPopulation	TSTPopulation																																																																								
1	北京	22529400	1079.22	35648000	33031.26	35648000																																																																								
2	广州	18788487	588.26	33869707	57576.08	69517707																																																																								
3	杭州	8838612	393.19	19731854	50184.02	89249561																																																																								
4	宁波	6734132	206.91	16746270	80935.04	105995831																																																																								
5	上海	30271100	1278.23	61578300	48174.66	167574131																																																																								
6	深圳	11552772	150.93	28789326	190746.21	196363457																																																																								
7	苏州	9442100	216.87	27160700	125239.54	223524157																																																																								
8	天津	11127100	758.78	23580000	31076.20	247104157																																																																								
9	无锡	7499400	219.60	18429600	83923.50	265533757																																																																								
10	重庆	9369000	1010.12	19142000	18950.22	284675757																																																																								
实验分析	<div><div>1. 通过 PROC IMPORT 语句可以进行 xlsx 数据的读取;</div><div>2. 通过 MERGE data1 data2 ... 语句可以实现几个数据之间的合并查询;</div><div>3. 在使用 MERGE 函数之前注意应先用 PROC SORT ... BY... 函数进行排序，后通过 MERGE ... BY... 进行合并;</div><div>4. 注意到，吉林的数据并不在 10 大城市之中，若只使用 MERGE 函数则将对两组数据进行取并操作，故需要进行剔除；剔除的方法为通过在 MERGE 语句的表格名称后添加（IN=InVar）语句，此时将创建一个新的临时变量 InVar，其可能的取值为 0（表示某变量不在该数据集中）或 1（表示在数据集中）；此时可以通过选择对于所有合并的数据集都有 InVar_i=1 的数据实现取交集的操作;</div><div>5. 通过 sum+inmt 的语句可以实现将每一个 obs 数据的 inmt 数值累加至 sum。</div></div>																																																																													
实验四																																																																														
实验代码	<pre>filename data1 "C:\Users\31949\iCloudDrive\By's Cloud\SUFE-lectures\大三\统计软件\Assignment\hw2\data1.xlsx"; proc import out=data1 datafile=data1 dbms=xlsx replace; sheet='Sheet1'; run;</pre>																																																																													

	<pre>proc import out=data2 datafile=data1 dbms=xlsx replace; sheet='Sheet2' ; run; data data_new; merge data1 data2; by ID; run; title "New Data"; proc print data=data_new noobs; run; title;</pre>																								
实验结果	<div>New Data</div> <table><tr><th>id</th><th>City</th><th>Population</th><th>UNERate</th></tr><tr><td>1</td><td>上海</td><td>2428</td><td>3.6</td></tr><tr><td>2</td><td>江苏</td><td>8070</td><td>3</td></tr><tr><td>3</td><td>浙江</td><td>5850</td><td>2.5</td></tr><tr><td>4</td><td>安徽</td><td>6366</td><td>2.6</td></tr><tr><td>5</td><td>福建</td><td>3973</td><td>3.5</td></tr></table>	id	City	Population	UNERate	1	上海	2428	3.6	2	江苏	8070	3	3	浙江	5850	2.5	4	安徽	6366	2.6	5	福建	3973	3.5
id	City	Population	UNERate																						
1	上海	2428	3.6																						
2	江苏	8070	3																						
3	浙江	5850	2.5																						
4	安徽	6366	2.6																						
5	福建	3973	3.5																						
实验分析																									

实验五	
实验代码	<pre>data data1; input id name\$ sex\$; datalines; 1005 Tom M 1003 Lily F 1001 Alice F 1002 Jack M ; run; data data2; input id subject\$ score; datalines; 1001 Chinese 95</pre>

	<pre>1003 math 84 1001 English 90 1002 history 92 ; run; proc sort data=data1 out=data1_sort;; by id; run; proc sort data=data2 out=data2_sort;; by id; run; data data3; merge data1_sort data2_sort; by id; run; title "data3"; proc print data=data3; run; title;</pre>																																				
实验结果	<div>data3</div> <table><tr><th>Obs</th><th>id</th><th>name</th><th>sex</th><th>subject</th><th>score</th></tr><tr><td>1</td><td>1001</td><td>Alice</td><td>F</td><td>Chinese</td><td>95</td></tr><tr><td>2</td><td>1001</td><td>Alice</td><td>F</td><td>English</td><td>90</td></tr><tr><td>3</td><td>1002</td><td>Jack</td><td>M</td><td>history</td><td>92</td></tr><tr><td>4</td><td>1003</td><td>Lily</td><td>F</td><td>math</td><td>84</td></tr><tr><td>5</td><td>1005</td><td>Tom</td><td>M</td><td></td><td>.</td></tr></table>	Obs	id	name	sex	subject	score	1	1001	Alice	F	Chinese	95	2	1001	Alice	F	English	90	3	1002	Jack	M	history	92	4	1003	Lily	F	math	84	5	1005	Tom	M		.
Obs	id	name	sex	subject	score																																
1	1001	Alice	F	Chinese	95																																
2	1001	Alice	F	English	90																																
3	1002	Jack	M	history	92																																
4	1003	Lily	F	math	84																																
5	1005	Tom	M		.																																
实验分析	<div>1. DATA ... INPUT ... DATALINES... 语句可以实现数据的程序内部创建读取；</div> <div>2. 通过 MERGE 语句可以实现不同表格之间的合并；</div> <div>3. 在应用 MERGE 语句之前要先进行 PROC SORT 排序，方可应用 MERGE 语句对排序好的共同元素列用 BY 进行合并查询；</div> <div>4. 注意到 data1 中包含的 id 为 1001、1002、1003、1005；而 data2 中包含的 id 为 1001, 1002, 1003；其中 1001 在 data2 中出现两次而 1005 在 data2 中没有出现，因此合并后的表格中，id=1001 将单独占据两行 obs，id=1005 的 subject 与 score 两列将为空值；</div> <div>5. 在默认情况下，MERGE 语句的合并将对各数据集进行取并集操作，若想要实现取交集，可以在 MERGE var_name 后添加 （IN in_var）语句，该语句将生成临时变量 in var，该变量的可能取值为 0（表示该 obs 不在该数据集中）或 1（表明</div>																																				

	在数据集中), 且该变量并不会在输出中直接显示; 通过逻辑判断语句, 选择同时 在所有数据集中同时有 in_vari=1 的 obs 即可实现数据集合并的取交操作。
--	---

实验六	
(6.1)	
实验 代 码	<pre> libname Lib "C:\Users\31949\iCloudDrive\By's Cloud\SUFE-lectures\ 大三\统计软件\参考\实验报告2\"; data result; set Lib.Test_answers; array key(10) \$1. ('A','C','C','B','E','E','D','B','B','A'); array ans(10) \$1. Q1-Q10; score=0; do num=1 to 10; if key(num)=ans(num) then do; score+1; end; end; output; keep Employee_ID Q1-Q10 score; run; data passed failed; set result; if (score>=7) then do; output passed; end; else do; output failed; end; run; title "Result"; proc print data=result; run; title; </pre>

实验结果

Result												
Obs	Employee_ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	score
1	121044	A	C	C	B	D	E	D	B	B	A	9
2	120145	B	C	C		E	E	D	B	A	A	7
3	120761	A	C	C	B	D	D	E	B	B	C	6
4	120656	B	C	C	A	D	B	B	C	A	D	2
5	121107	A	C	C	B	E	E	D	B	B	A	10
6	121038	B	C	C	B	D	D	D	B	B	A	7
7	120273	C	C	C	B	E	E	E	B	B	A	8
8	120759	A	C	C	B	E	E	D	B	B	A	10
9	120798		A	C	B	D	D	D	B	B	A	6
10	121030	C	C	C	C	E	E	D	B	B	B	7
11	121017	B	B	E	B	E	E	D	B	B	A	7
12	121062	A	C	C	B	E	E	D	B	B	A	10
13	121119	C	C	C	B	E	E	D	B	B	A	9
14	120812	A	C	C	B	E	E	E	B	B	A	9
15	120756	A	C	C	B	E	E	D	B	B	A	10

(6.2)

实验代码

```
title2 "Passed";  
proc print data=passed;  
run;  
title2;
```

实验结果

Passed												
Obs	Employee_ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	score
1	121044	A	C	C	B	D	E	D	B	B	A	9
2	120145	B	C	C		E	E	D	B	A	A	7
3	121107	A	C	C	B	E	E	D	B	B	A	10
4	121038	B	C	C	B	D	D	D	B	B	A	7
5	120273	C	C	C	B	E	E	E	B	B	A	8
6	120759	A	C	C	B	E	E	D	B	B	A	10
7	121030	C	C	C	C	E	E	D	B	B	B	7
8	121017	B	B	E	B	E	E	D	B	B	A	7
9	121062	A	C	C	B	E	E	D	B	B	A	10
10	121119	C	C	C	B	E	E	D	B	B	A	9
11	120812	A	C	C	B	E	E	E	B	B	A	9
12	120756	A	C	C	B	E	E	D	B	B	A	10

(6.3)

实验代码	<pre>title3 "Failed"; proc print data=failed; run; title3;</pre>																																																				
实验结果	<div><div>Failed</div><table><tr><th>Obs</th><th>Employee_ID</th><th>Q1</th><th>Q2</th><th>Q3</th><th>Q4</th><th>Q5</th><th>Q6</th><th>Q7</th><th>Q8</th><th>Q9</th><th>Q10</th><th>score</th></tr><tr><td>1</td><td>120761</td><td>A</td><td>C</td><td>C</td><td>B</td><td>D</td><td>D</td><td>E</td><td>B</td><td>B</td><td>C</td><td>6</td></tr><tr><td>2</td><td>120656</td><td>B</td><td>C</td><td>C</td><td>A</td><td>D</td><td>B</td><td>B</td><td>C</td><td>A</td><td>D</td><td>2</td></tr><tr><td>3</td><td>120798</td><td></td><td>A</td><td>C</td><td>B</td><td>D</td><td>D</td><td>D</td><td>B</td><td>B</td><td>A</td><td>6</td></tr></table></div>	Obs	Employee_ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	score	1	120761	A	C	C	B	D	D	E	B	B	C	6	2	120656	B	C	C	A	D	B	B	C	A	D	2	3	120798		A	C	B	D	D	D	B	B	A	6
Obs	Employee_ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	score																																									
1	120761	A	C	C	B	D	D	E	B	B	C	6																																									
2	120656	B	C	C	A	D	B	B	C	A	D	2																																									
3	120798		A	C	B	D	D	D	B	B	A	6																																									
实验分析	<div><div><div>1. 对于 SAS 数据的读取，可以通过 LIBNAME 全局语句确定该数据所在的相应物理地址，建立自定义逻辑库，直接通过调用逻辑库读取相应数据；</div><div>2. 对于大量数据的处理，可以通过 ARRAY 建立数组，方便统一的存储与读取；</div><div>3. ARRAY 若读取字符型变量，需要在定义时进行格式声明；</div><div>4. 对于数据表中的连续变量，也可以通过 var1-varn 的缩写进行表示；</div><div>5. 一个 DATA 步骤也可以输出多个数据集，其方法为在声明时首先列出全部可能输出的数据集名称 data_1…data_n，在后续的处理中可以通过逻辑语句等方式进行分支判断，分别通过 OUTPUT data_i 语句将不同的数据输出到不同的数据集中。</div></div></div>																																																				

【小结】

本次的实验主要还是集中在 SAS 的基本语法的应用过程中。在前两道题的编写中,由于仍然对于程序的写法、DATA 步的循环逻辑不太熟悉,因此花费了较长时间。后续随着题目难度的降低与对于程序语言的逐渐熟练,编程效率逐渐提升。

本次实验报告主要涉及到的知识点如下:

- INPUT … DATALINES 语句进行程序内数据读取输入;
- FORMAT 语句对数据集的列格式进行设定;
- IF… THEN DO 的逻辑判断语句;
- INPUT(input, informat)语句对数据集列格式进行修改;
- ARRAY 数组的声明与调用;
- PDV 的循环逻辑;
- RETAIN 函数对数据进行保留;
- CALL MISSING 赋值空值语句的使用;
- OUTPUT 的含义与输出时机的选择;
- PROC PRINT 数据打印;
- MERGE… BY…数据集合并语句的使用;
- (IN=InVar)语句对 MERGE 函数的取交处理;
- PROC IMPORT 对.xlsx 数据的读入与存储;
- Sheet 声明对.xlsx 不同 sheet 的读取;
- PROC SORT… BY 语句的排序功能;
- Sum+inmt 的数据累加;
- INPUT Var\$ 存储字符型变量;
- .sas7bdat 通过自建逻辑库的读取;
- DATA OUTPUT 的分支存储;

20.

希望能够在后续的学习中能够及时对知识进行整理与编程练习,进一步提高 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，做出结论。

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。