金融数据处理技术

**课程名称：**CUR341 金融数据处理技术 **Financial Data Processing Technology**

**课程性质：**本学院选修,实验课程

**学分课时**：2学分，32课时，其中实验课时15

**主讲教师**：冯建芬副教授

**所属院系**：金融学院金融工程系

电话：64495048,E-mail：jianfen\_feng@UIBE.EDU.CN

**教学对象**：金融学院大学二年级学生

**考核方式**：课下作业（10%），每章一次，提交程序和结果数据集

课堂实验（10%），每章1-2次，共计8次实验，提交程序和结果数据集，

期中考试（小组数据处理报告）（30%）

期末考试：开卷上机笔试（50%）

其中平时成绩包括课下作业和课堂实践，占20%, 期中成绩占30%，期末考试占50%

**学术诚信：**本课程对于学生的学术诚信的要求遵从《对外经济贸易大学学生违纪处分条例》、《对

外经济贸易大学学生学习违纪处分实施细则》、《对外经济贸易大学考场纪律》的规定。

**教学方式：**课堂讲授占比30%，上机实验占70%，教学中强调动手能力，采用实验教学、实践教学和研讨教学相结合的方式。教学全程在金融实验室完成，结合SAS软件、Ocale实验教学平台、金融数据库和PPt进行教学。

**出勤要求：**遵从《对外经济贸易大学本科生课堂学习规范》，要求学生关闭一切电子设备；不能

无故缺席上课；上课专心听讲，积极参与课堂讨论；课后认真复习课堂上讲授内容，独立完成教师布置的任务；并预习新课。学生缺勤不得多于总课时的四分之一。教师可以根据考勤情况决定学生是否可以参加考试、是否扣分。

**一、课程简介：**

本课程通过大量案例，借助数据处理和分析软件SAS，学习处理金融数据的方法和技术,是一门金融和计算机紧密结合的实践性非常强的综合课程。内容涉及：金融数据处理与分析软件（SAS）的基本应用、金融数据的批处理技术、常见统计分析方法的应用及金融数据处理案例分析等。主要教学内容有：数据文件在不同数据库间的格式转换、数据筛选与清洗技术、循环处理技术、同时处理多个文件的数据批处理技术（宏编程技术）等。

**二、教学目标**

本课程的教学目的是弥补学生在数据处理环节的不足，提高其金融数据处理与分析能力，为金融计算、金融时间序列、金融风险定量分析等后续课程学习及金融实践提供技术支持，为提高本科生的科研能力和实践能力排除障碍。

**三、课程学习资料**

**1.教材**

《SAS编程技术教程》（第2版），朱世武， 清华大学出版社，2013。

**2.参考资料**

《SAS编程技术教程》（第1版），朱世武， 清华大学出版社，2007。

《The little SAS book》, Delwiche,Lora D. and Slaughter, Susan J., 2003, NC: SAS Institute Inc.

《BASE SAS软件使用手册》，高惠璇等，编译，中国统计出版社，1997年

《SAS编程与数据挖掘商业案例》，姚志勇，机械工业出版社，2010.

《Using SAS in Financial Research》， Boehmer,Ekkehar,John Paul Broussard and Juha-Pekka Kullunki,2002, NC: SAS Institute Inc.

《金融计量学》，张宗新，中国金融出版社，2012.

《SAS 统计分析与数据挖掘》，谢龙汉，尚涛， 电子工业出版社，2013.

**四、学习效果及达成途径**

 **1.学习效果：**

 **通过本课程的学习，希望达成的学习效果如下：**

1）.掌握SAS的BASE SAS模块的语法，理解SAS处理数据的原理，能够熟练的编制SAS完整的程序。

2).能够使用SAS对金融数据进行各种文件格式间的转换，并通过SAS熟练地访问和生成各种数据库文件、PC文件。

3).能够通过SAS查询金融数据库数据，远程访问金融数据终端。

4).掌握基于SAS软件的金融数据的各种清洗技术，包括研究数据的筛选、数据查询、缺失值处理、日期格式数据处理、多个数据表的纵向合并、横向合并、多表查询、单表查询、数据表记录的添加、删除、更新。

5).掌握SAS帮助系统的查询和学习方法，能够自主利用SAS帮助系统学习其强大的统计分析功能、金融计量功进行金融数据挖掘；

6).掌握基于SAS的海量数据的批处理技术

7）掌握SQL数据库查询语言在SAS中的应用，能够灵活运用SQL语言进行多表数据的查询和链接。

**2.达成学习效果的途径**

 善于动手，课前完成例题程序的上机任务；上课跟着老师思路走，积极参与课堂讨论；充分利用实验教学平台的讨论区和教师指定的答疑时间及时答疑解惑；按时完成课堂实践任务和课下作业，协同完成小组金融实践任务，认真准备期末考试。

**五、教学进度计划表**

**本课程教学周为16 周，具体安排如下**

| **周次** | **内容提要** | **教学方式** | **参考资料** | **作业与考试** |
| --- | --- | --- | --- | --- |
| 1 | 第一讲：认识SAS | 讲授、实验 | 教材第1章 | 课堂上机 |
| 2 | 第二讲：SAS的数据存储与读取 2.1 SAS对数据文件的管理 2.2建立数据集，读入原始数据  | 讲授、实验 | 教材第2,3章，第4章:4.1-4.4，4.14，第8章，第9章 |  |
| 3 |  学生实践：建立逻辑库，建立数据集，读入数据  | 实验 | 课堂实践测验 |
| 4 |  2.3 输出原始数据  | 讲授、实验 |  |
| 5 | 学生实践：外部文件与SAS的数据传输：导入与导出 | 实验 | 课堂实践测验；课下作业 |
| 6 | 第三讲 查询语言SQL 3.1 SQL过程与SQL语言简介3.2 从单个表中查询和加工数据3.3 从多个表中查询和拼接数据 | 讲授、实验 | 教材13，14章 | 课下作业 |
| 7 | 学生实践：利用SQL语言查询和加工数据 | 实验 |  | 课堂实践测验 |
| 8 | 第四讲：数据集的编辑与修改4.1 改变变量或数据集的属性4.2 变量和观测的选择  | 讲授、实验 | 第4章4.5, 4.6, 4.7, 4.8, 4.11, 4.12，第5章，第6章 |  |
| 9 | 学生实践：对数据的初步处理，变量和观测的选择 | 实验 | 课堂实践测验课下作业 |
| 10 | 4.3变量编辑与观测的修改4.3.1 数据集拼接增加变量和观测4.3.2 利用表达式与函数修改观测 | 讲授、实验 |  |
| 11 | 4.3.3 利用循环语句修改观测和变量4.3.4 利用数组修改观测 | 讲授、实验 | 课下作业 |
| 12 |  学生实践：数据的进一步处理技术，变量与观测的修改技术应用 | 实验 | 课堂实践测验；布置小组金融实践任务 |
| 13 | 第五讲：宏编程技术 5.1 宏变量的定义与引用 5.2 宏的定义与调用 5.3 宏程序语句和宏函数5.4 数据步接口程序 | 讲授、实验 | 教材第16章 | 课下作业 |
| 14 | 学生实践：宏技术的应用—宏与数据批处理技术 | 实验 |  | 课堂实践测验 |
| 15 | 第六讲 过程步的应用 | 讲授、实验 | 教材第11,12章 | 课下作业 |
| 16 | 学生小组演示报告 | 研讨 |  | 小组作业研讨 |
| 17-18 | 期末考试（学校统一安排考试时间及地点） |

**六、教学内容：**

**第一章：SAS入门**

【教学目的和要求】

使学生了解SAS的基本操作界面和操作窗口，指导学生如何使用SAS强大的帮助系统，并通过部分案例使学生了解SAS的数据处理功能。

【主要内容】

1.1 SAS软件介绍

 1.2 SAS工作界面

 1.3 SAS窗口操作

 1.4 SAS帮助文档

 1.5 SAS安装

**教学总时数：**2，其中实验课时1课时

参考资料：《SAS编程技术教程》第二版，第一章

**参考资料：**《SAS编程技术教程》第二版，第一章

**作业与练习（课上实验，1课时）：**

1.关闭日志窗口、输出窗口、编辑器窗口，重新打开上述三个窗口；(使用快捷键)

2.关闭资源管理器，停放资源管理器；

3.关闭结果窗口，停放结果窗口；

4.通过菜单建立逻辑库，删除逻辑库；

5. 打开程序”例1-1class程序”，运行程序，另存程序；

6. 建立新的程序，输入如下代码：

options noxwait;

x md F:\data;

libname data "F:\data";

libname data clear;

7. 逐行执行程序；保存新的程序；

8.清楚日志窗口的内容；

**第二章：SAS的数据存储与读取**

【教学目的和要求】

使学生掌握如何建立SAS的数据集，以各种方式读入数据进行存储，包括直接输入方式、数据集导入方式、外部文件（excel文件、txt文件、access文件，ODBC数据源、SPSS文件等）导入方式，并掌握如何将SAS数据集中的数据以各种方式展现或导出到外部文件。

【主要内容】

2.1 SAS对数据文件的管理；

2.1.1 SAS逻辑库与SAS数据集功能及其属性

2.1.2 建立逻辑库，读入文件；

2.2 建立数据集，读入原始数据；

2.2.1 建立空数据集

2.2.2 直接对变量赋值建立数据集

2.2.3 直接输入数据建立数据集

2.2.4 从其他数据集导入数据建立数据集

2.2.5 从外部文件导入数据建立数据集

2.3 输出原始数据

2.3.1 输出到Log窗口

2.3.2 输出到输出窗口；

2.3.3 导出的外部文件；

**教学总时数：**8，其中实验课时4课时

参考资料：1.《SAS编程技术教程》第二版，第二章，第三章，第四章:4.1-4.4，4.14，第8章，第9章

2.SAS系统帮助目录：SAS 产品🡪Base SAS🡪SAS语言字典🡪语言元素字典

**作业与练习：**

**课堂实验1（2课时）：**

1.通过菜单操作和libname语句分别建立几个逻辑库，分别指向如下文件夹：



2. .建立各种空数据集

3.完成利用x发布主机操作命令的任务

4.完成注释语句的实验任务

5.通过赋值建立数据集的实验任务

6.完成通过数据行读入数据建立数据集的任务

**课堂实验2（2课时）**

1. 利用import过程导入数据的实验任务。

2. 利用 input-infile语句导入文本数据的任务。

3. 利用put语句输出数据到log窗口

4. 利用print过程将结果输出到输出窗口

5. 利用export过程将数据导出到外部文件

6. 利用file-put语句将数据导出到文本文件

**课下作业：**

1)利用x命令，获取data文件夹中文件名前三个字符是stk的文件，将其存储到文件name.txt;

2)查找帮助，尝试在SAS的命令栏使用下列SAS命令，并说明其作用：keys, dlglib, libname, dir, var,submit, recall.

3)将name.txt中的内容读入到SAS数据集hwork.a, 变量名为name, 长度为9($9.)，要求不能在同一行读入两条记录(@)。（infile-input语句）

4)将hwork.a中的变量name输出到外部文件hongstk.txt，每一个变量记录添加字符串，如：stk000001 变为：%stk(stk000001, result);

5) 登陆锐思数据网站(账号：uibe/ 密码：uibe) ：www.resset.cn参考输出设置部分的“格式说明”文档，分别下载如下格式的数据源，下载条数为500条，且将其转换为SAS数据集。要求设置好环境，并给运行结果正确的程序。

逗号分隔文本（\*.CSV）

空格分隔文本（\*.TXT）

Tab键分隔文本 （\*.TXT）

EXCEL电子表格（\*.XLS）

字符型EXCEL电子表格 （\*.XLS）

6)用PUT和FILE语句写SAS程序，将数据集data.class直接输出为外部TXT文本。

7)在数据集中建立100个变量。

**第三章：查询语言SQL**

【教学目的和要求】

SQL是数据库通用的一种强大的查询语言，用它可以简化挑选和处理数据的过程。通过本章的学习学生可以学习从单个数据文件或多个数据文件中进行合并查询数据的技术，以及在查询中进行简单计算和汇总的方法。本章要求学生掌握SQL过程的简单应用；会使用SQL查询语句处理单个数据集和多个数据集的查询；会使用SQL查询语句进行简单的求和、求均值等汇总运算。

【主要内容】

3.1 SQL过程与SQL语言简介

3.1.1 SQL语言简介

3.1.2 SQL过程简介

3.1.3 SQL中的语句及其作用

3.2 从单个表中查询和加工数据；

3.2.1 利用select 语句处理和加工数据；

3.2.2 创建新表存储查询结果；

3.3 从多个表中查询和拼接数据；

3.3.1 从多个数据表查询数据实例

**教学总时数：**4，其中实验课时2课时

参考资料：1.《SAS编程技术教程》第二版，第十三、十四章

2.SAS帮助目录：SAS 产品🡪Base SAS🡪SAS SQL过程用户指南

SAS产品🡪Base SAS🡪SAS过程🡪过程🡪SQL过程

**作业与练习：**

**课堂实验（2课时）**利用SQL过程完成单个表和多个表的金融数据查询和汇总任务；

1.a)从data.exchbdqttn\_1中选择债券代码、日期、债券名称、收盘全价、收盘净价、净价成交金额、成交笔数，并生成year、 qtr、 month三个变量，存储日期的年、季、月，年限范围为2000年至2010年；

b)将上述结果存储到数据集ex.ex3\_1中；

c)对数据集ex.ex3\_1按照年、季、月排序；

d).删除ex.ex3\_1中成交笔数为0和收盘净价缺失的数据，将剩余结果保存到数据集ex.ex3\_2中

2.将数据集ex.ex3\_2中的数据导出到excel表格ex3\_2.XLS

3. 查询data.exchbdqttn\_1中所有不同的股票代码；

4.利用数据集data.dret计算各股票日平均收益，并按平均收益进行排序

5. 对ex.ex3\_1按季度分组计算可转债总交易量和总成交金额，存储在数据集ex.sum\_qtr中,该数据集应该还有年，季两个时间指标；

6.从data.bdinfo中挑选市场标示(mktflg)同时含有“1”,“3”或者同时含有“2”，“3”的观测，并将筛选出的观测按照债券标识（bdid）升序排列，结果存至ex.bdinfo；

7.从data.exchbdqttn 中挑选出锐思标识（resbdid）在ex.bdinfo中的债券行情数据

课堂实践2. 数据的进一步处理技术，变量与观测的修改技术应用

**课下作业：**

1). 以2008年A股上市公司股票为样本，找出在2008年摘帽的公司 （即公司名称中ST被去掉的）

2). 查找2009年首次被ST的上市公司

3). 查找截止到2009年9月30日，一直没有被特别处理的股票。

4). 计算data.dret中各股票的月累计收益率，增加变量positive,若月累计收益率大于0，值为1，否则值为0.

5). 试由以下的9种德国马克对美元汇率看跌期权和9种英镑对美元汇率的看跌期权产生81种组合。

一年期，DM/USD汇率的9种不同看跌期权的执行价格和成本

执行价格Kdm 成本 Cdm

0.66 0.085855

0.65 0.032191

0.64 0.020795

0.63 0.017001

0.62 0.013711

0.61 0.010851

0.60 0.008388

0.59 0.006291

0.55 0.001401

一年期，BP/USD汇率的9种不同看跌期权的执行价格和成本

执行价格 Kbp 成本 Cbp

1.30 0.137213

1.25 0.082645

1.20 0.045060

1.15 0.028348

1.10 0.016146

1.05 0.007860

1.00 0.003277

0.95 0.001134

0.90 0.000245

6). 很多情况下，读者可能会综合数据库中多张表的内容，来更直观更方便地获得全面信息。比如，需要把同一公司的主要财务比率和财务指标放在一起来研究整个公司。

数据合并举例：

Is\_old数据集：利润及利润分配表\_旧准则。

Frs数据集：代表公司的主要财务比率表

Scf\_old数据集：现金流量表\_旧准则。

需要把这三个表中的以下变量综合起来：

Is\_old表:主营业务利润(Mopeprf),营业利润(Opeprf), 营业外收支净额=营业外收入Noperev -营业外支出Nopeexp，利润总额(Totprf),净利润

FrS表：每股收益摊薄(EPS)，流动比率(Currt)，速动比率(Qckrt)，资产负债率(Dbassrt)，每股净资产(NetassPS)，销售净利率(Netprfrt)，每股经活动产生的现金流量净额(OpeCFPS)，总资产周转率(次)(Totassrat)，净利润（扣除非经常性损益后Netprfcut）

Scf\_old表:经营活动产生的现金流量净额(NCFope)，现金及现金等价物净增加额(NetincrCCE).

**第四章：数据集的编辑与修改**

【教学目的和要求】

学习几乎全部的数据处理技术，包括多个数据表的拼接和匹配合并技术，使用表达式和函数进行数据编辑的技术，使用循环语句通过单个数据生成多条数据的技术，使用数组简化处理多个变量的技术等等，这些技术的综合运用基本能够解决金融数据的初步处理问题。要求学生熟练运用本章的data步程序语句，并灵活运用上述技术解决金融问题。

【主要内容】

4.1 改变变量或数据集的属性

 4.1.1 通过数据集选项改变属性：data语句，set语句

 4.1.2 通过语句改变属性：label语句、rename语句、retain语句;

4.2 变量与观测的选择

 4.2.1 选择变量：keep, drop选项与语句;

 4.2.2 选择观测

 4.3变量编辑与观测的修改

 4.3.1 数据集拼接增加变量和观测

 4.3.2 利用表达式与函数修改观测

 4.3.3 利用循环语句修改观测和变量

 4.3.4 利用数组修改观测

**教学总时数：**10，其中实验课时4课时

参考资料：《SAS编程技术教程》第二版第四章4.5,4.6,4.7,4.8,4.12，第五章，第六章

**作业与练习：**

**课堂实验1（2课时）.** 对数据的初步处理，变量和观测的选择

 1. 建立数据集ex.ex4\_1, 从数据集data.fdbs读入数据：

a)给fdcd加标签为“基金代码“, 给fdnm加标签为“基金名称“，取消date的标签；

b)利用data语句的rename选项，将fdcd更名为code,将fdnm更名为name；

c)为数据集ex.ex4\_1加标签为”基金资产负债表”

2. 建立数据集ex.ex4\_2,从数据集data.class中读入数据，给每个变量加一个中文标签，并删除变量sex；

3. 下段程序的END,LASTOBS和X中，哪个是选项？哪个是自动变量？哪个是普通变量？

data a;

set resdat.a600001 end=lastobs;

if lastobs;

x=lastobs;

put ‘last observation’;

run;

4.下段程序能否正常运行？等式END=END两边的含义是什么？

data a;

set resdat.a600001 end=end;

if end;

proc print;

title2 'last observation';

run;

5. 建立数据集ex.ex4\_2\_1,读去数据集data.monret1的前100条记录，但从第5条开始读入。

6. 建立数据集app1\_d,获取数据集data.hf000012的第1000条记录。（尝试寻找效率最高的方式）

7. 剔除data.exchbdqttn\_1中成交笔数缺失和为0的记录。

8. 输出数据集data.hf000012的总的记录数到log窗口。

9. 根据data.hf000012中记录的不同月份将相同月份的数据输出到一个数据集。

**课堂实验2（2课时）.**

1. 纵向合并data.stk000001, data.stk000002, data.stk00004, data.stk000005的数据到数据集ex.stock,并将该数据集输出到stock.xls.

2.利用data.stk000001的收盘价clpr计算五日百分比收益率,将其存到数据集ex. r\_day5\_000001;

3. 将data.stk000001和data.stk000002中的数据按照年份的先后顺序进行纵向合并,结果存为ex.ex4\_3\_3。

4.利用input函数将data.lstkinfo中的字符型变量stkcd的值转变为数值型变量code,结果存为ex.ex4\_3\_4。

5.利用字符匹配函数选取data.lstkinfo的股票名称中含有“中”的记录,存储在ex.ex4\_3\_5。

6.求100以内的偶数之积,存储在ex.ex4\_3\_6；

7.创建包含日期变量DATE的SAS数据集，日期值从1900年1月1日到2000年1月1日,存储在ex.ex4\_3\_7。

8.利用锐思数据的首次发行与上市数据，选取2006年在沪市上市的公司，保留股票代码、上市日期、发行价和首日收盘价；从月累积收益率数据中提取这些公司在上市后3年的月度数据，与其首次发行数据进行合并，然后将其月度的收益率与上证指数相应月度收益率合并，计算其月度超额收益率与3年来的月均超额收益率。结果存储在ex.ex4\_3\_8（首次发行与上市数据在iissultst中，月累积收益率数据在moncret中，上证指数月度收益率在idxmonret中）

9.数据集A有一个变量n，5个观测值1,2,3,4,5。数据A1由下面程序2产生，同样有一个变量n，5个观测值1,2,3,4,5。试分析下面两段程序中，PUT语句在Log窗口输出结果的差异，为什么？

程序1：

Data a;

Set a;

Put n=;

Run;

程序2:

data a1;

do n=1 to 5;

output;

end;

put n=;

run;

10. 建立数据集ex.4\_3\_1,其含有100个变量x1-x100，100个观测，分别用DATA步，DATA步数组语句实现。结果存储在ex.ex4\_3\_10

11. 产生100个独立同分布的随机变量e1-e100，每个随机变量服从标准正态分布，对每个随机变量抽取1000个随机数；结果存储在ex.ex4\_3\_11

12.对于上面生成的100个随机变量，产生另外100个随机变量sum1-sum100,对i=1,…,100,其对应的值分别为： Sumi=，结果存储在ex.ex4\_3\_12

**课下作业：**

**第一次作业：**

1. 给下段程序的主要语句加注释。STOP语句能否删除？为什么？

data a;

do obsnum=1 to last by 20;

set ResDat.stk000001 point=obsnum nobs=last;

output;

end;

stop;

run;

2. 现有一个数据流：a 2 b 3 c d 4 6，按下面要求创建SAS数据集。用语句input id $ no; 变量id取值上面数据流里的a,b,c,d，变量no取值2,3,4,6。但是这个数据流存在问题：如有的id没有no，有的no没有id。创建SAS数据集，删除只有id没有no或者只有no没有id的观测，即把上面的c和6去掉，最后得到三个观测，a 2, b 3与d 4。

3. 现有一个数据流：a23 223 bc4 36 3c5 11d 400 620，按下面要求创建SAS数据集。

用语句input id $ no; 变量id取值上面数据流里的a23, bc4, 11d ，变量no取值233,36,400。但是这个数据流存在问题：如有的id没有no，有的no没有id。创建SAS数据集，删除只有id没有no或者只有no没有id的观测，即把上面的3c5和620去掉，最后得到三个观测：

a23 223

bc4 36

11d 400

4.建立数据集block, 在数据集data.hf000012中挑选(成交额-成交量\*成交价），其绝对值小于10000的数据。即删除那些因为数据传输错误导致的成交额与成交量不匹配的数据。

5.将data.monret中的2000年以后的记录按照年份分成不同的数据集，同一年份的放在一个数据集，如2000年的记录放在数据集data\_2000中，2001年的记录放在数据集data\_2001中，依次类推。

**第二次作业：**

1. 为比较不同股票间的相对价格变化，从data.idx.399106中提取data.stk000001, data.stk000002, data.stk00004共有的交易日数据，存储在ex.price中，将data.idx.399106 ，data.stk000001 data.stk000002, data.stk00004的收盘价clpr分别改名为clpr0-clpr3，在ex.price中保存date,clpr0-clpr3。

2.将样本数据文件中文件名为“stk”+股票代码的SAS数据集纵向合并成一个数据集。如stk000001.SAS7bdat, stk000002.SAS7bdat等

3. 将第三章的第6个作业用data步语句实现。

4． 数据集Calendar只有日期变量，test包含股票代码、日期与收盘价三个变量。合并两个数据集，用Calendar中的日期数据替代test数据集中每只股票的日期数据。

5. 设master\_trans是公司所有客户的历史交易总额，为主数据集，day\_trans是公司每天所有客户（包括新客户和老客户）的交易流水表。要求：将每天产生的数据表day\_trans加载到历史表中，汇总每一个客户在截至最新日期时的累积交易总额。

6.建立数据集ex.hm3\_4\_1, 从data.stk000001中读入数据，如果oppr(开盘价）高于clpr（收盘价)color=‘red’,否则color=‘green’，并根据prevclpr（前收盘价)和clpr计算今天的涨跌幅度fluctuate。最终保留变量date, prevclpr,oppr,clpr, color, fluctuate。

7.从data.hf000012中挑选交易量(tvolume）大于10万的记录，存储在ex.blocktrade中，并统计每天大额交易的笔数，如果一天内大额交易笔数超过5笔，删除当天所有的大额交易记录。

8.数据集stk000001日期变量DATE包含有缺失值，创建包含日期变量DATE的数据集B，并填充开始到结束日之间的所有日期值。

9. 将同一个公司的A,B股合并。

在处理公司信息表的时候，通常我们会遇到数据源将同一公司的A，B股分开处理，并写成两条观测的情形。如公司代码为C000002的股票，在这里是以两条记录的形式记录下了该公司发行了A股和B股的情况，在公司信息表中，需要将其整合成一条信息，并分别以A，B股的名称和形式来表示。而且有的公司先发行A股，有的公司却是先发行B股，这也是需要注意的问题。

**原始样式:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **公司代码** | **日期** | **公司中文名称** | **股票代码** | **股票名称** | **股票类型** |
| … | … | … |  |  | … |
| C000001 | 2005-05-18 | 深圳发展银行股份有限公司 | 000001 | 深发展A | A |
| C000002 | 2000-08-15 | 万科企业股份有限公司 | 000002 | 万科A | A |
| C000002 | 2005-06-21 | 万科企业股份有限公司 | 200002 | 万科B | B |
| C000003 | 2000-08-29 | 金田实业(集团)股份有限公司 | 200003 | PT金田B | B |
| C000003 | 2004-04-30 | 金田实业(集团)股份有限公司 | 000003 | PT金田A | A |
| C000004 | 2005-07-09 | 深圳中国农大科技股份有限公司 | 000004 | 国农科技 | A |
| C000005 | 2005-05-27 | 深圳世纪星源股份有限公司 | 000005 | ST星源 | A |
| C000006 | 2005-05-19 | 深圳市振业(集团)股份有限公司 | 000006 | 深振业A | A |
| C000007 | 2005-07-30 | 深圳市赛格达声股份有限公司 | 000007 | 深达声A | A |
| C000008 | 2005-05-21 | 广东宝利来投资股份有限公司 | 000008 | 宝利来 | A |
| … | … | … |  |  | … |

**最终需要形成的样式：**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 公司代码 | 日期 | 公司中文名称 | 股票代码 | A股名称 | A股股票代码 | B股名称 | B股股票代码 |
| … | … | … |  |  |  |  | … |
| C000001 | 2005-05-18 | 深圳发展银行股份有限公司 | 000001 | 深发展A | A |  |  |
| C000002 | 2005-06-21 | 万科企业股份有限公司 | 200002 | 万科A | A | 万科B | B |
| C000003 | 2004-04-30 | 金田实业(集团)股份有限公司 | 000003 | PT金田A | A | PT金田B | B |
| C000004 | 2005-07-09 | 深圳中国农大科技股份有限公司 | 000004 | 国农科技 | A |  |  |
| C000005 | 2005-05-27 | 深圳世纪星源股份有限公司 | 000005 | ST星源 | A |  |  |
| C000006 | 2005-05-19 | 深圳市振业(集团)股份有限公司 | 000006 | 深振业A | A |  |  |
| C000007 | 2005-07-30 | 深圳市赛格达声股份有限公司 | 000007 | 深达声A | A |  |  |
| C000008 | 2005-05-21 | 广东宝利来投资股份有限公司 | 000008 | 宝利来 | A |  |  |
| C000009 | 2005-07-01 | 中国宝安集团股份有限公司 | 000009 | 深宝安A | A |  |  |
| C000010 | 2005-07-22 | 深圳市华新股份有限公司 | 000010 | 深华新 | A |  |  |
| C000011 | 2005-06-29 | 深圳市物业(发展)集团股份有限公司 | 000011 | 深物业A | A | 深物业B | B |
| … | … | … |  |  |  |  | … |

10. **.**利用随机数函数RANUNI对某数据集设计返回抽样方案。

11. 股票连续涨停和跌停的收益求和计算。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Stkcd** | **Date** | **Dayret** |  | **Dayret\_update** |
| 600003 | 01-27-00 | -0.98% |  |  |
| 600003 | 01-28-00 | 0.60% |  |  |
| 600003 | 02-14-00 | 10.06% | (1) |  |
| 600003 | 02-15-00 | 10.04% | (2) |  |
| 600003 | 02-16-00 | 9.93% | (3) | =(1)+(2)+(3) |
| 600003 | 02-17-00 | 10.07% | (4) |  |
| 600003 | 02-18-00 | 2.56% | (5) |  |
| 600003 | 10-18-01 | -0.98% |  |  |
| 600003 | 10-19-01 | -1.19% |  |  |
| 600003 | 10-22-01 | -2.40% |  |  |
| 600003 | 10-23-01 | 10.04% | (6) |  |
| 600003 | 10-24-01 | 2.23% | (7) |  |
| 600003 | 10-25-01 | -3.64% |  |  |
| 600003 | 10-26-01 | -0.76% |  |  |
| 600026 | 06-04-02 | 0.70% |  |  |
| 600026 | 06-05-02 | -1.40% |  |  |
| 600026 | 06-06-02 | 3.31% |  |  |
| 600026 | 06-07-02 | 0.92% |  |  |
| 600026 | 06-10-02 | -0.23% |  |  |
| 600026 | 06-11-02 | 0.91% |  |  |
| 600026 | 06-12-02 | -0.23% |  |  |
| 600026 | 06-13-02 | -0.90% |  |  |
| 600026 | 06-14-02 | -1.37% |  |  |
| 600026 | 06-17-02 | 0.23% |  |  |
| 600026 | 06-18-02 | 2.53% |  |  |
| 600026 | 06-19-02 | -0.90% |  |  |
| 600026 | 06-20-02 | 0.91% |  |  |
| 600026 | 06-21-02 | 3.15% |  |  |
| 600026 | 06-24-02 | 10.02% | (8) |  |
| 600026 | 06-25-02 | 10.10% | (9) |  |
| 600026 | 06-26-02 | 10.07% | (10) |  |
| 600026 | 06-27-02 | 9.97% | (11) |  |
| 600026 | 06-28-02 | 0.74% | (12) | =(8)+(9)+(10)+(11)+(12) |
| 600026 | 07-01-02 | 5.16% |  |  |
| 600026 | 07-02-02 | -1.40% |  |  |
| 600026 | 07-03-02 | 2.13% |  |  |
| 600026 | 07-04-02 | -1.67% |  |  |
| 600026 | 07-05-02 | 1.56% |  |  |
| 600026 | 07-08-02 | -0.42% |  |  |
| 600026 | 07-09-02 | 5.32% |  |  |
| 600038 | 12-24-03 | -0.47% |  |  |
| 600038 | 12-25-03 | -1.49% |  |  |
| 600038 | 12-26-03 | 1.03% |  |  |
| 600038 | 12-29-03 | -0.07% |  |  |
| 600038 | 12-30-03 | -0.48% |  |  |
| 600038 | 12-31-03 | 0.14% |  |  |
| 600038 | 01-02-04 | -9.97% | (13) |  |
| 600038 | 01-05-04 | -10.02% | (14) |  |
| 600038 | 01-06-04 | -10.03% | (15) |  |
| 600038 | 01-07-04 | 10.03% | (16) | +(13)+(14)+(15)+(16) |
| 600038 | 01-08-04 | 9.97% | (17) |  |
| 600038 | 01-09-04 | 4.65% | (18) | =(16)+(17)+(18) |
| 600038 | 01-12-04 | -4.74% |  |  |
| 600038 | 01-13-04 | -2.87% |  |  |
| 600038 | 01-14-04 | -4.32% |  |  |
| 600038 | 01-15-04 | -3.68% |  |  |
| 600038 | 01-16-04 | -1.30% |  |  |
| 600038 | 01-29-04 | 10.03% | (19) |  |
| 600038 | 01-30-04 | 9.99% | (20) |  |
| 600038 | 02-02-04 | 0.44% |  | =(19)+(20)+(21) |
| 600038 | 02-03-04 | -1.30% |  |  |
| 600038 | 02-04-04 | 5.94% |  |  |
| 600038 | 02-05-04 | 0.83% |  |  |

**注：**

除列出计算式子的行以外，其他各行的收益\_new与原来的收益值相同。

计算的原则是遇到连续的涨停板和跌停板时（即涨停和跌停连续大于等于两天），将所有连续涨停或跌停日的总收益加起来，最后再加上第一个非涨停或跌停日的收益；涨跌停交错时，则最后加上的是涨停或跌停日的收益。详见表格中最后一列的算法。

12. 抽取数据集data.hf000012每5分钟的高频交易记录。（即从9：30开始，每5分钟抽取一次，若在5分钟时点上没有对应记录则以5分钟后离该时点最近的一次记录作为其5分钟交易记录，例如在9:35:00没有记录，但在9:35:01有记录，则抽取9:35:01的记录作为其5分钟交易记录）。

**第五章：宏编程技术**

【教学目的和要求】本章学习程序的简化技术和数据批处理技术，包括宏变量的使用，宏的使用，为程序代码的调用提供简化方式，并通过宏的循环语句简化编程，通过%include语句完成批处理技术。要求学生能够结合本章宏技术与所学所有章节内容，综合运用解决金融数据处理问题。

【主要内容】

 5.1 宏变量的定义与引用

5.1.1 宏变量的定义：%let 语句

5.1.2 宏变量的引用及注意事项：&

5.1.3 从变量取值获得宏变量的值：call symput()

5.2 宏的定义与调用

5.2.1 定义无参数的宏：%macro --%mend

5.2.2 调用无参数的宏；%宏名

5.2.3 定义含参数的宏

5.2.4 调用含参数的宏

5.3 宏程序语句和宏函数

5.3.1 宏程序设计语句及用途

5.3.2 宏函数

5.3.3 综合举例

**教学总时数：**4，其中实验课时2课时

**参考资料：**

1.《SAS编程技术教程》第二版，第16章

2. SAS帮助目录 SAS 产品🡪Base SAS🡪SAS宏参考🡪了解和使用宏（分类解释宏的各项功能）/宏语言字典（所有宏语句、函数的用法说明）

**作业与练习**

**课堂实验（2课时）**

1.书P296-297: 4,9,11,12,18,20;

2. 接例5.22，计算data.idx00001中1999年到2004年间的指数日百分比收益率。将5.22中获得的个股日百分比收益率与指数日百分比收益率按照日期匹配，计算匹配日期的超额百分比收益率

 超额百分收益率=个股百分比收益率－指数的百分比收益率。

 每个个股的收益率存储到一个数据集，保留变量：date, stk\_r, idx\_r,extra\_r四个变量，分别表示：日期，个股日百分收益率，指数日百分比收益率，超额百分比收益率。

3 .将data.monret中的记录按照年份分成不同的数据集，同一年份的放在一个数据集，如2000年的记录放在数据集data\_2000中，2001年的记录放在数据集data\_2001中，依次类推。

4.生成的100个独立的标准正态分布的随机变量的1000个随机数e1-e100，产生另外100个随机变量m1-m100,对i=1,…,100,其对应的值分别为：

 **mi=**

**第六章：过程步的应用**

【教学目标和要求】介绍过程步在统计分析、经济计量、图表制作等方面的应用。重点让学生了解SAS的统计分析功能和图表制作功能，不需要学生掌握每个过程的应用，只需要学生了解如何使用SAS的帮助来了解所需过程的功能以及使用语法，为学生以后使用SAS的这些程序模块提供长远之计。

【主要内容】

6.1 过程步通用语句:proc, var, model, where, class, by, output, quit, label.

 6.2统计量的计算

6.2.1 相关过程：proccorr

6.2.2 频数过程：procfreq

6.2.3 均值过程：proc means

6.2.4 单变量过程：procunivariate

6.2.5 回归分析

6.2.6 回归过程：procreg

6.3 图表制作

6.3.1 作图过程句法；

6.3.2 制作散点图和连线图；

6.3.3 制作柱状图和饼图；

6.3.4 图形存储与输出

**教学总时数：**4，其中实验课时2课时

参考资料：

1.《SAS编程技术教程》第二版，第十一，十二章

2.SAS帮助目录：

1)普通SAS过程：SAS 产品🡪Base SAS🡪SAS过程🡪过程\SAS过程：corr, freq, univeriate

2)统计回归过程：，SAS 产品🡪SAS/STAT🡪SAS/STAT用户指南：GLM\REG\FACTOR等

3)图表过程：SAS 产品🡪SAS/GRAPH🡪SAS/GRAPH参考🡪SAS/GRAPH procedures：GPLOT

**作业与练习：**

**课堂实验（2课时）**

学习利用gplot过程绘制各种图形

  **撰写人：**冯建芬 2016年3月