

统计软件习题

李东风

2014.12.

目录

第一章 SAS 初阶	5
第二章 SAS 语言与数据管理	9
第三章 SAS 功能基础	13
第四章 SAS 的基本统计分析功能	17
第五章 SAS 多元统计分析	23
第六章 S 语言介绍	25

第一章 SAS 初阶

1. (1) 启动 SAS, 认识界面。
 - (2) 下载本习题集配套的数据下载包 data.zip, 取出其中的 SAS 例子数据集压缩包 sampled.zip 并在适当位置释放, 然后用快捷图标“New Library”或运行适当的 libname 语句给下载的数据包中的 sampled 子目录定义 SAS 库名 (libref) 为 SAMP。
 - (3) 查看 SAMP 库中生成的例子数据集的情况。
 - (4) 输入如下例子程序, 在运行记录窗口查看有无错误, 有错时回到程序窗口调回程序修改。

```
title '95 级 1 班学生成绩排名';
data c9501;
  input name $ 1-10 sex $ math chinese;
  avg = math*0.5 + chinese/120*100*0.5;
  cards;
李明      男 92 98
张红艺    女 89 106
王思明    男 86 90
张聪      男 98 109
刘颖      女 80 110
;
run;
proc print;run;
proc sort data=c9501;
  by descending avg;
run;
```

```
proc print;run;
```

(5) 用 INSIGHT 数据窗口输入 C9501 数据集。

2. 调查了 20 个牌子的牛肉热狗的卡路里, 得到如下数据:

186, 181, 176, 149, 184, 190, 158, 139, 175, 148,
152, 111, 141, 153, 190, 157, 131, 149, 135, 132

把数据输入为 SAS 数据集。假设这些数据为某 $N(\mu, \sigma^2)$ 分布的独立样本, 概述其分布, 计算 μ 的 90% 置信区间。

3. (1) 启动 SAS/INSIGHT, 打开 SAMP.GPA 数据集。作各变量的直方图, 查看其分布情况并简答 (变量类型、中心位置、分散程度、分布形状、极端值等)。
 - (2) 重点研究 GPA 分数的分布。说明极端值情况。在纸上画出 GPA 的盒形图, 并说明如何解释。通过直方图、盒形图、各统计量、分布检验结果简述 GPA 分布的特点。
 - (3) 把 GPA 数据集按性别排序, 同性别内按 GPA 分数由高到低排序。
 - (4) 给男女生观测指定不同颜色。画 GPA 对 HSM 的散点图。画各数值型变量的散点图矩阵。画 HSM、HSS、HSE 的三维散点图。简述 GPA 数据集各变量间的直观的相互关系。
4. 例子数据集 SAMP.BUSINESS 数据集保存了若干个大公司的在 1993 年的情况: 公司名称 (COMPANY)、所属国家 (NATION)、所属部门 (INDUSTRY)、雇员人数 (EMPLOYS, 单位: 千人)、销售额 (SALES, 单位: 百万美元)、利润 (PROFIT, 单位: 百万美元)。
- (1) 在 INSIGHT 中打开这个数据集, 把变量 COMPANY 作为 Label 变量。
 - (2) 用雇员人数、销售额、利润作散点图矩阵, 查找其中突出的大公司; 查找其中利润占销售额比例较高和较低的公司; 查找其中人均利润较高的公司。
 - (3) 把 EMPLOYS 和 SALES 做对数变换得到两个新变量, 用这两个新变量画散点图并评价两个变量的关系。

- (4) 计算各公司利润占销售额比例，画并排盒形图比较各国家的利润情况；计算各公司人均创造利润，画并排盒形图比较各国家的劳动生产率。
5. 例子数据集 SAMP.FITNESS 中保存了 31 名男子的一些试验记录，这些人分为三个组 (GROUP)，研究的主要指标是一个利用氧的能力的指标 (OXYGEN)。另外每个人记录了年龄 (AGE)、体重 (WEIGHT)，还有 1.5 英里跑的时间 (RUNTIME)、休息时的心率 (RSTPULSE)、跑步时的心率 (RUNPULSE)、跑步时达到的最快心率 (MAXPULSE)。
- (1) 在 INSIGHT 中打开 SAMP.FITNESS，把 GROUP 的类型指定为名义型。
 - (2) 按组分类然后在每一组内按年龄由高到低排序。
 - (3) 简述各变量的分布情况。
 - (4) 作三个组的年龄的并排的盒形图。
 - (5) 作三个心率的并排盒形图。
 - (6) 作 OXYGEN 对其他变量的散点图。

第二章 SAS 语言与数据管理

- (1) 用 SAS 数据步列出 10000 以下的素数, 写出程序。
(2) 把上述问题改写成可以求出 N 以下的素数的宏。
- (1) 写出一个程序, 可以给出 280 的素因子分解。
(2) 把上述问题改写成可以求出任意正整数素因子分解的宏。
- 生成 t 分布的双侧分位数表。水平取 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.10, 0.20, 自由度取 1-100, 分位数精确到小数点后 3 位。表格应为行、列对齐的形式, 并有列标题。写出生成这样的表格并存放到一个文本文件中的 SAS 程序 (用数据步 FILE 和 PUT 语句)。

注: 一个随机变量 X 的双侧 p 分位数是指 x_p 使得 $P(|X| > x_p) = p$ 。

- 写出计算从自己生日到 2000 年初经过的天数的程序。
- 下表为某邮购服务部的部分顾客记录:

姓名	性别	地区	日期	金额
章文	男	华东	1996-3-20	1099
王国铭	男	华东	1996-5-19	39
童子敏	女	华北	1996-1-5	986
刘念新	男	东北	1997-10-1	3581
李思今	女	华北	1997-4-4	659
关昭	女	东北	1996-11-5	358
赵霞	女	东北	1998-9-6	2010

- 用数据步把此表输入到 SAS 数据集;
- 用程序找出男性顾客购买金额超过 1000 的哪些人;

- (3) 把数据拆分为包含姓名、性别、地区的一个数据集和包含姓名、日期、金额的一个数据集;
- (4) 用 MERGE 和 BY 合并上一步拆开的两个数据集。
6. 设有三位老师分别带三个兴趣小组, 人员情况如下:

老师情况		一组情况		二组情况		三组情况	
组号	老师姓名	学号	姓名	学号	姓名	学号	姓名
1	张民	101	刘娜	201	方开祥	301	何明
2	李秀丽	102	张秀敏	202	耿火清	302	孙成秀
3	曾维薇	103	于长江	203	李芳	303	张克成

- (1) 输入这四个表到数据集 teachers, g1, g2, g3 中;
- (2) 用这四个表合并生成包括所有兴趣组成员的数据集, 每个学生为一个观测, 变量包括学生学号 (SID)、学生姓名 (SNAME)、指导老师 (TNAME)。要求所写的程序不能依赖于数据集内容。
7. 设某校的兴趣小组名单格式如下:

组名	指导教师	小组成员
数学	张民	刘娜, 张秀敏
美术	李秀丽	方开祥, 耿火清, 李芳
英语	曾维薇	何明

- (1) 把这些原始数据输入成每组一个观测的数据集 A。
- (2) 把数据集 A 转换成每个组员一个观测的数据集 B, 变量为组名 (GNAME)、指导教师 (TNAME)、组员姓名 (SNAME)。
8. (1) 生成一个数据集 D1, 包含 50 行, 101 个变量, 其中变量 X1—X100 用标准正态分布随机数产生, 变量 R 为行号;
- (2) 用宏按 (1) 的办法生成 20 个数据集 D1—D20;
- (3) 对每个数据集, 计算每个观测的 100 个变量的平均值、标准差、最小值、最大值, 以及 100 个值落入 ± 1.96 之间的比例, 各数据集的结果合并存入一个新的数据集。
9. 在下载数据包的文件 data.zip 中有一个 ex-midsch.zip 数据包, 其中包含了三个中学毕业班某次考试的一些情况:

- student.csv: 学生名单;
- scores.csv: 各科考试成绩表;
- subayw.csv: 语文各小题的题型;
- subsyw.csv: 学生的语文小题成绩, 每个 subayw.csv 中出现的小题对应于这个文件的一列。

输入这四个文件为 SAS 数据集, 用如下两种方法解决:

- (1) PROC IMPORT;
- (2) 数据步, INFILE 和 INPUT 配合读入 CSV 格式数据, 变量名在程序中输入;

10. 设有如下格式的联系方式数据 (数据量很大, 所以不能考虑另行输入的方法), 用数据步把数据输入为 SAS 数据集, 要求每个联系人一个观测。注意数据中每个联系人第一行都是姓名, 但是电话、电子邮件和单位则不是所有人都有, 每个人的信息项数不定。

```
姓名: 张三
电话: 01012345678
电子邮件: zs@abc.com
```

```
姓名: 李四
工作单位: 清华大学
```

```
姓名: 王五
电子邮件: ww@efg.com
```

```
.....
```

11. 编写一个宏子程序, 可以自动输入一个 CSV 文件为 SAS 数据集。难点包括: 自动确定变量个数和变量名; 自动确定变量是数值型还是字符型; 确定字符型变量的长度; 自动识别和转换日期、时间。

第三章 SAS 功能基础

1. 对 SAMP.BUSINESS 数据集:
 - (1) 列出数据集的内容, 要求给各列加上合理的中文列标题, 不要观测序号。
 - (2) 把观测按国家、部门、销售额 (由高到低) 排序后按国家、部门列出公司名、销售额、利润, 计算销售额的小计和总计。
 - (3) 用 TABULATE 过程绘制按国家和部门交叉分类的销售额及总计, 使用中文标签。
2. (1) 把第二章习题8中的数据集 D1 的 X1—X100 转置, 保存为 E1;
(2) 把 D1 中每个观测的 X1—X100 合并到一列中, 结果数据集取名为 F1。
3. 对 SASUSER.GPA:
 - (1) 用 UNIVARIATE 过程分析 SATM 分布并简述结果;
 - (2) 叙述性别分布;
 - (3) 计算 HSS、HSM、HSE 的相关系数。
 - (4) 绘制 SATM 的直方图;
 - (5) 画 SATV 对 SATM 的散点图;
4. 对 SAMP.FITNESS:
 - (1) 用 UNIVARIATE 过程分析 OXYGEN 分布并简述结果。
 - (2) 按 GROUP 分组计算各变量的简单统计量。
 - (3) 计算各变量的相关系数。

5. 把第二章习题8中的数据集中的数据集 D1 每一行的 X1-X100 看作一组样本，作直方图。

提示: (1) 用转置和 BY 语句; (2) 用宏、SET 语句选项、转置解决。

6. 绘制 F 分布自由度为 (1,30), (2,30), (3,30), (4,30), (5,30), (10,30) 的密度曲线图，画在同一坐标系中。

提示: F(1,30) 的密度在左端有奇点。

7. 长表和宽表的互相转换时经常遇到的问题。设我们有 X 和 Y 两个变量，10 个人的观测，每个人在 4 个时间点上观测。数据集 LONG 有 $10 \times 4 = 40$ 个观测，两个测量值变量 X 和 Y。我们希望:

(1) 计算后一期减前一期的值，如 X 的第 2 期减去 X 的第 1 期，存为 DX12。

(2) 每个人的所有信息汇集在一个观测内，作成宽表 WIDE，包括 X01, X02, X03, X04 即 X 的四个不同时间点值，Y01, Y02, Y03, Y04, 以及差分值 DX12,DX23,DX34, DY12,DY23,DY34. 用 PROC TRANSPOSE, 横向拆分，横向合并解决，或用数据步和宏解决。

其中生成表 LONG 的代码如下:

```
data long;
  nsubjects=10;
  ntimes=4;
  array vars(2) x y;
  do subject=1 to nsubjects;
    do time=1 to ntimes;
      x = ceil(uniform(111333)*10);
      y = ceil(uniform(111333)*10);
      output;
    end;
  end;
  keep subject time x y;
run;
```

8. 在第二章的习题9中读入了语文的小题题型和学生的小题分。
- (1) 把数据集 subsyw 中 Y1-Y18 合并为一列，另外增加一列变量表示小题序号。设结果数据集为 S1。
 - (2) 把 S1 与 subayw 按小题序号横向合并。
 - (3) 计算每个学生四种不同题型的分数，生成包含学号、T1、T2、T3、T4 变量的结果数据集，其中 T1、T2、T3、T4 是每个学生四种题型的分数。
9. (选做) 编写 SAS 宏程序，对任一数据集，可以分析每个数据集变量，对字符型，如果其不同值的个数不超过 10 个（此界限可更改），显示所有不同值及频数、百分比、缺失个数和比例，否则列出其前 10 个出现最多的值的情况。对数值型，如果其不同值的个数不超过 10 个（此界限可更改），显示所有不同值及频数、百分比、缺失个数和比例，否则显示其均值、中位数、标准差、最小值、最大值、缺失个数和比例。提示: PROC DATASETS 可以获取变量列表以及变量是数值型还是字符型的信息。

第四章 SAS 的基本统计分析功能

1. 下面是一组草原隼的鸟巢高度的数据, 试检验其分布是否正态。

15 3.5 3.5 7 1 7 5.75 27 15 8 4.75 7.5
4.25 6.25 5.75 5 8.5 9 6.25 5.5 4 7.5 8.75 6.5
4 5.25 3 12 3.75 4.75 6.25 3.25 2.5

2. 调查了 41 个城市三年平均的二氧化硫含量 (单位: mg/m^3), 试检验:
数据是否服从正态分布; 数据是否服从对数正态分布。

10 13 12 17 56 36 29 14 10 24 110
28 17 8 30 9 47 35 29 14 56 14
11 46 11 23 65 26 69 61 94 10 18
9 10 28 31 26 29 31 16

3. 有若干人参加了一个减肥锻炼, 在一年后测量了他们的身体脂肪含量,
结果如下 (身体脂肪含量的百分数):

男性组: 13.3 19 20 8 18 22 20 31 21 12 16 12 24
女性组: 22 26 16 12 21.7 23.2 21 28 30 23

比较这些人中男性和女性的身体脂肪含量有无显著差异 (检验水平
0.05。)

4. 下表为某基础统计课程两次考试的学生成绩。两次考试考同样的知识。
试比较这两次考试难易程度有无显著差异 (检验水平 0.05)。

学号	1	2	3	4	5	6	7	8	9	10
第一次	93	88	89	88	67	89	83	94	89	55
第二次	98	74	67	92	83	90	74	97	96	81
学号	11	12	13	14	15	16	17	18	19	20
第一次	88	91	85	70	90	90	94	67	87	83
第二次	83	94	89	78	96	93	81	81	93	91

5. 为研究溶菌酶水平在患胃溃疡的病人与正常人之间有无显著差异, 测量了一组病人和一组正常人的溶菌酶水平, 结果见下表。试检验两者的溶菌酶水平有无显著差异 (水平 0.05)。

胃溃疡病人组:	0.2	10.4	0.3	10.9	0.4	11.3	1.1	12.4	2.0	16.2
	2.1	17.6	3.3	18.9	3.8	20.7	4.5	24.0	4.8	25.4
	4.9	40.0	5.0	42.2	5.3	50.0	7.5	60.0	9.8	
对照组:	0.2	5.4	0.3	5.7	0.4	5.8	0.7	7.5	1.2	8.7
	1.5	8.8	1.5	9.1	1.9	10.3	2.0	15.6	2.4	16.1
	2.5	16.5	2.8	16.7	3.6	20.0	4.8	20.7	4.8	33.0

提示: 要考虑分布是否正态。

6. 假设对随机选取的 8 位病人使用 A 药物, 对随机选取的 6 位病人使用 B 药物, 过一段时间后测量每位病人体细胞中的药物含量, A、B 药物的测量数据见下表:

A 药物:	1.23	1.42	1.41	1.62	1.55	1.51	1.60	1.76
B 药物:	1.76	1.41	1.87	1.49	1.67	1.81		

在 0.10 水平下检验 B 药物的含量是否高于 A 药物的含量。

7. 为了考察两种测量萘含量的液体层析方法: 标准方法和高压方法的测量结果有无显著差异, 取了 10 份试样, 每份分为两半, 一半用标准方法测量, 一半用高压方法测量, 每个试样的两个结果如下表, 试检验这两种化验方法有无显著差异 (水平 0.05):

标准:	14.7	14.0	12.9	16.2	10.2	12.4	12.0	14.8	11.8	9.7
高压:	12.1	10.9	13.1	14.5	9.6	11.2	9.8	13.7	12.0	9.1

8. 为了研究药物补钙对高血压是否有疗效, 随机选取了 10 个人服用补钙药物, 11 个人服用安慰剂, 预先记录这些人的血压。12 周后测量每个人的血压并减去原来的血压, 得到如下的血压变化数据:

补钙组: 7 -4 18 17 -3 -5 1 10 11 -2

安慰剂组: -1 12 -1 -3 3 -5 5 2 -11 -1 -3

在 0.10 水平下检验服用补钙药物与服用安慰剂相比是否血压降低更多。

9. 下表列出了若干种鱼产品在 1970 年和 1980 年价格。建立用 1970 年价格预报 1980 年价格的回归公式。

1970	1980	1970	1980
13.1	27.3	26.7	80.1
15.3	42.4	47.5	150.7
25.8	38.7	6.6	20.3
1.8	4.5	94.7	189.7
4.9	23	61.1	131.3
55.4	166.3	135.6	404.2
39.3	109.7	47.6	149

10. 测量了在不同温度 x 下制成的某种合金的强度 y , 得到如下数据。拟合线性回归和二次回归并比较。

x : 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0
 y : 40 41 43 42 44 42 43 42

11. 使用放射性金 195 作示踪元素注射到血液中, 下表为注射 x 天后血液内残留的金元素百分比 y , 取了 10 个血样, 对数据分别拟合线性回归、负指数关系 $y = Ae^{-bx}$, 并使用各种非参数曲线拟合方法拟合曲线。比较各结果。

x : 1 1 2 2 2 3 5 6 6 7
 y : 94.5 86.4 71 80.5 81.4 67.4 49.3 46.8 42.3 36.6

12. 对数据集 SASUSER.GPA 中的大学学科平均成绩 GPA 建模, 用高中成绩 HSM、HSS、HSE 作为自变量。简述回归的结果。试改进模型。使用 SAS/INSIGHT 和 REG 过程两种办法。
13. 对 SASUSER.FITNESS, 以 OXYGEN 作为因变量, AGE、WEIGHT、RUNTIME、RSTPULSE、RUNPULSE、MAXPULSE 作为自变量进行线性回归, 研究如何改进。

14. 下面的数据是不同气压（单位：英寸汞柱）下水的沸点（单位：F），建立线性回归方程用沸点估计气压（用气压可以测算海拔高度），并作回归诊断，提出可能的改进意见。

沸点	气压	沸点	气压
194.5	20.79	201.3	24.01
194.3	20.79	203.6	25.14
197.9	22.40	204.6	26.57
198.4	22.67	209.5	28.49
199.4	23.15	208.6	27.76
199.9	23.35	210.7	29.04
200.9	23.89	211.9	29.88
201.1	23.99	212.2	30.06
201.4	24.02		

15. 对 SASUSER.FITNESS，检验 OXYGEN 在不同组 (GROUP) 之间是否有显著差异。
16. 在一个双盲试验 (受试者和操作者都不知道分组情况) 中研究了咖啡因对受试者反映能力的影响。选了 30 个大学生进行按键速度测试，把这 30 人随机分为三组，每组 10 人，分别服用三种不同剂量的咖啡因 (0 mg, 100 mg, 200 mg)。服药后记录每人每分钟按键次数。数据如下：

咖啡因剂量	每分钟按键数										
0 mg	242	245	244	248	247	248	242	244	246	242	
100 mg	248	246	245	247	248	250	247	246	243	244	
200 mg	246	248	250	252	248	250	246	248	245	250	

- (1) 对三个组的数据作并列的盒形图，看各组之间有无显著差异。
- (2) 用方差分析表检验不同剂量的三组的按键次数有无显著差异 (0.10 水平) 并解释结果。
17. 为试制某种化工产品，在三种不同温度、四种不同压力下试验，每一水平组合重复两次，得到产品的收率数据如下 (%)：

温度	压力			
	1	2	3	4
1	52, 57	42, 45	41, 45	48, 45
2	50, 52	47, 45	47, 48	53, 30
3	63, 58	54, 59	57, 60	58, 59

试在 0.05 水平下进行方差分析并简述结果。

18. 为了考察法院判决是否与被告种族有关, 调查了 326 位被告的判决情况:

	黑人	白人
有罪	17	19
无罪	149	141

试在 0.05 水平下检验判决结果与被告种族是否独立。

19. 下表为 100 位被调查者的性别及颜色偏好情况。

性别	颜色偏好		
	红	蓝	绿
男	32	14	4
女	25	17	8

试在 0.05 水平下检验颜色偏好是否与性别有关。

20. 下表为 200 个婴儿的喂养方法 (牛奶、母乳或并用) 及母亲的经济状况的调查情况。试在 0.05 水平下检验喂养方法是否与母亲的经济状况有关。

喂养方法	经济状况			
	贫穷	下	中	上
牛奶	30	15	11	12
母乳	7	18	19	29
并用	5	23	7	19

第五章 SAS 多元统计分析

- 对 SASUSER.GPA 中的变量进行主分量分析并试解释结果。
 - 对 SASUSER.GPA 中的变量进行因子分析并试进行旋转, 对得到的因子进行解释。
- 对 5.1 节例 2 的 CRIME 数据集进行因子分析并试进行旋转, 对得到的因子进行解释。
- 设有三个组, 四个变量, 数据见表 5.1。计算线性判别函数, 简述对训练样本的判别情况。
- 用 SASUSER.BUSINESS 中的雇员数、销售额、利润在 INSIGHT 中作三维散点图;
 - 把这三个变量用主分量方法压缩为两个主分量, 解释这两个主分量的含义, 在 INSIGHT 中作散点图。
 - 用这三个变量作聚类分析, 在 INSIGHT 的散点图中用不同颜色区分得到的类别。
 - 研究这三个变量的分布, 如果某变量严重偏离正态则作适当变换将其分布变为接近正态分布, 然后用变换后的变量作上述分析。
- 对 5.1 中的例子数据集 SOCECON 作聚类分析。

表 5.1: 练习 3 的数据

组别	X1	X2	X3	X4
1	6	-11.5	19	90
1	-4	-15.0	13	54
1	0	-23.0	5	-35
1	-100	-21.4	7	-15
1	-5	-18.5	15	18
1	10	-18.0	14	50
1	-8	-14.0	16	56
2	90.2	-17.0	17	3
2	0	-14.0	20	35
2	-100	-21.5	15	-40
2	13	-17.2	18	2
3	-11	-18.5	25	-36
3	0.5	-11.5	19	37
3	-10	-19.0	21	-42
3	20	-22.0	8	-20
3	0.6	-13.0	26	21
3	-40	-20.0	22	-50

第六章 S 语言介绍

1. (1) 写出元素为 3, -1.5, 3E-10 的向量。
(2) 写出从 3 开始每次增加 3, 长度为 100 的向量。
(3) 写出 (0, 2) 重复 10 次的向量。
(4) 对向量 x, 写出其元素大于等于 0 小于 1 的条件。
(5) 对向量 x, 写出其元素都等于 0 的条件。
(6) 写出包含 12 个月份名称的向量。
(7) 生成一个包含文件名 tab1.txt 到 tab18.txt 的字符串向量。
(8) 写出包含方程 $z^6 = 1$ 的根的向量, 并写出其辐角的余弦和正弦值。
2. 设 x 为一个长 100 的整数向量。比如, $x < -\text{floor}(100*\text{runif}(100))$ 。
 - (1) 显示 x 第 21 到 30 号元素。
 - (2) 把 x 第 31,35,39 号元素赋值为 0。
 - (3) 显示 x 中除了第 1 号和第 50 号的元素之外的子集。
 - (4) 列出 x 中个位数等于 3 的元素。
 - (5) 列出 x 中个位数等于 3 的元素的下标位置。
 - (6) 给 x 的每一个元素加上名字, 为 x1 到 x100。
 - (7) 求 x 的平均值并求每一个元素减去平均值后的离差, 计算 x 的离差平方和及元素的平方和。
 - (8) 把 x 从大到小排序。计算 x 的 10% 分位数到 90% 分位数之间的距离。

3. 定义一个维数为 (3, 4, 2) 的数组, 其第一层 (第三下标为 1) 取从 1 开始的奇数, 第二层取从 2 开始的偶数。显示每一层的第 2 行元素。把第 (1,1,1),(2,2,2), (2, 2, 1) 号元素赋值为零。把第一层加上 100, 把第二层加上 200。分别计算第一层和第二层的平均值。
4. 对线性模型 $Y = X\beta + \varepsilon$, 写出当 X 满秩时计算 β 的 S 表达式。写出估计 ε 的方差的 S 表达式。
5. 把 SASUSER.GPA 数据中的 SEX, SATM, SATV 分别输入到 S 中。计算不同性别的人数, 并计算每一组的平均 SATM 分。把这些变量组合成一个列表。把 SASUSER.GPA 数据输入为 S 的数据框。
6. 把语句 $x < -\text{floor}(100*\text{runif}(100))$ 所生成的向量保存到一个文本文件中, 数据项用空格和换行分隔。从此文件中读入数据到向量 y 中。
7. 设 x 是一个长度为 n 的向量, 写一段程序, 计算 x 的长度为 s 的滑动和:

$$S_x(t) = \sum_{i=0}^{s-1} x_{t-i}, \quad t = s, s+1, \dots, n$$

8. 写一个 AR(1) 的模拟函数:

$$x_t = a + bx_{t-1} + \varepsilon_t, t = 1, 2, \dots, n, \text{Var}(\varepsilon_t) = \sigma^2$$

函数的参数为 n 、 a 、 b 、 x_0 和 σ , 缺省时 $n=100$, $a=0$, $b=1$, $x_0=0$, $\sigma=1$ 。