

用微型计算机统计高考成绩

广西计算中心

符华儿

摘要

1981年5月,我们用一台微型电子计算机(C.M.TECHNOLOGIES, Z-80B档机,仅带一个终端)对南宁市一万一千五百九十一名考生的高考预选成绩进行了统计处理,所得的各项处理结果准确无误。本文总结了用微型计算机处理大量数据的一些经验体会。

一、概况

1981年5月,我们用一台微型电子计算机(C.M.TECHNOLOGIES, Z-80B档机,仅带一个终端)对南宁市高考预选的成绩数据进行了统计处理,取得了较好的效果,对用微型机处理大量数据取得了一些经验。

这次高考预选,南宁市共有考生一万一千五百九十一人参加,其中理工科七千二百七十人,文科三千九百五十七人,外语三百六十四人。所有分数统计工作,共用计算机时100小时,其中输入考生成绩85小时,机器运算15小时(加总分、打印成绩单等10小时,排序〔把考生的成绩按总分从高分至低分排队〕5小时)。

二、计算机完成的工作

计算机在这次高考数据统计中,所做的工作如下:

- 1、计算每个考生的中专总分和大学总分。
- 2、打印每个考生的成绩通知单。
- 3、累计总分各档人数(按招生办要求划为10分一档)。
- 4、按理工、文科和外语三大类统计参加考试总人数。
- 5、统计各单科各分数段人数。
- 6、计算总平均分、各单科平均分。
- 7、将全部考生成绩按高分到低分顺序排队、编号、打印成绩册。
- 8、根据南宁市招生办划定的分数线,统计上线人数、上线考生的总分平均分和各科平均分。

9、统计工作结束后，应南宁招生办要求，增加了南宁市各中学考生成绩的质量分析和统计功能。新增加部分的程序已通过，准备在下一次用计算机统计高考分数时使用。

三、用计算机处理高考分数，我们体会有如下一些优点

1、在计算中专总分和大学总分的过程中，计算机代替了人的脑力劳动，运算速度快，结果完全正确。

2、避免了人工的重复性劳动，提高了工作效率，原来用人工进行统计时，加总分要把所有分数累计一次，发考生成绩单时又把所有分数抄一遍，划分分数段时，则又把已累计出的总分进行粗略的排列。如还做其它的统计工作，则还要有更多次的重复性劳动，改用计算机后，所有数据仅输入计算机一次，上面所列的各项工作，都可以由计算机自动完成。

3、用人工划总分分数段，由于工作量和时间紧迫，只能比较粗糙地划分，一般是10分一档或5分一档。改用计算机统计则是把所有考生的总分，按照从高分到低分的顺序依次排列，并给出了每个人总分成绩的名次。这是过去历年来人工统计从未实现过的。这种排列方式，给划分录取分数线带来了极大的方便和好处。

4、用计算机还做了其它大量的分数统计工作，如算平均分、分数分布情况等，所有这些工作，如果用人工来算，计算量将是非常浩大的。因为统计工作时间上的限制，这些属于考生质量分析方面的工作也从来没有如此细致而及时地开展过。

四、程序粗框

我们把整个分数统计工作用几个程序来完成。它们是：输入数据并检查数据错误程序；计算总分并打印成绩单程序；总分排序程序；考生成绩质量分析程序。分段的结果使程序结构层次清晰，并减小了一旦局部出错互相干扰的可能性。另外再加上调行打印机程序、初始化程序、随机数发生器产生分数程序（用于试验）、重打成绩单程序等，组成了这次分数统计处理的应用软件。后面以理科为例，列出主要程序的框图。

五、一些经验体会

使用微型计算机处理大量的数据，必须重视和解决存贮量和运算速度的问题。我们使用的这种微型计算机，运算速度约每秒数百次，内存是64K，外存使用的是8寸软磁盘，若用双密度，存储量是500K。

1、存储量问题

我们在处理这次高考预选成绩前，先对南宁市体育高考成绩进行了处理。南宁市参加体育高考人数约为五百，其中男生三百多，女生一百多。我们用数组来存储数据。我们所用机器的CBASIC语言规定，数组的最大体积是二千多个元素，每个人所要存储的数据约十项，故一个数组只能存储二百多人的成绩。因此，我们把男生的成绩存入两个数组，女生的成绩存入一个数组。由于统计前不知道确切的考试的人数，因此采取了预留较大体积的动态数组方式。

在这次高考预选中，理工科考生有七千多人；采用数组存储考生成绩必然会产生很大困难，而且将要浪费大量存储单元。因此我们改用文件结构来存储考生成绩，解决了上述的问题。为了使存储和修改更方便，对文件我们采用了固定结构而没有用流结构。

由于微型计算机的存储量相对的比较小，而数据量较大，会不会溢出？虽然通过理论上的计算，知道这种状况下还不至于溢出，但剩余存储空间也不多了。因为高考成绩统计，牵涉千家万户，必须做到万无一失，为慎重起见，还应按照实际情况做大规模的试验才行。我们用随机数发生器模拟产生了与实际问题相仿佛的大量数据，按实战要求进行了大规模的试验，未产生溢出。实际运算的情况与实验相符合。

2、排序的计算量问题

存储问题解决之后，我们就转向考虑计算量问题，主要是排序的计算。对理工科考生七千多人，如用气泡漂浮法进行大排序，则计算量惊人，需要比较的次数大约为 $7000 \times 7000 / 2$ ，这对于运算速度较慢的微型计算机来说几乎是不能实现的。为解决这一问题，我们把排序部分计算改为，先把总分由高到低初步分成32个文件，然后在每个文件中用气泡漂浮法进行排序比较。这种方法，比较的次数大约在 $32 \times 200 \times 200 / 2$ 次以上，是大排序计算量的几十分之一。但估计用这种方法对七千人进行排序，计算时间仍要20小时以上，还是太久。当然最好是刚开始时分成1分一个文件，这样就不需要进行任何比较了。但由于我们的磁盘，最多允许有64个文件目录。因此分数段不能划得太细，否则文件目录将溢出。最后我们改为在每个分数段文件中，从高分到低分直接读出，如七千人分成五十八个文件，则比较次数是 7000×10 ，后在做7000人大规模试验时，文件目录溢出。故把原来的五十八个文件合并为二十九个文件，则比较次数是 7000×20 ，计算量是分段气泡漂浮法的五分之一。实际运算时，七千多人的排序，用机时3个半小时。预考统计工作完成后，我们又对排序算法做了进一步的改进，估计七千多人成绩的排序，用2小时左右可完成。

3、其他

一 自动查错

由于输入的数据量大，虽然输入时采取了许多人工校对措施，仍难免会产生错误。我们在程序中另排了数据查错语句，让计算机自动查找数据错误，并指出出错位置。

二 调行打印机

Z-80B档机宽行打印机每打印63行就空3行，如果不调整好，就会在不该空行的地方留出空行。为解决这个问题，我们编了一个小程序来调行打印机，后又把这一功能并入打印成绩单的主程序中，让计算机自动调整行打印机。

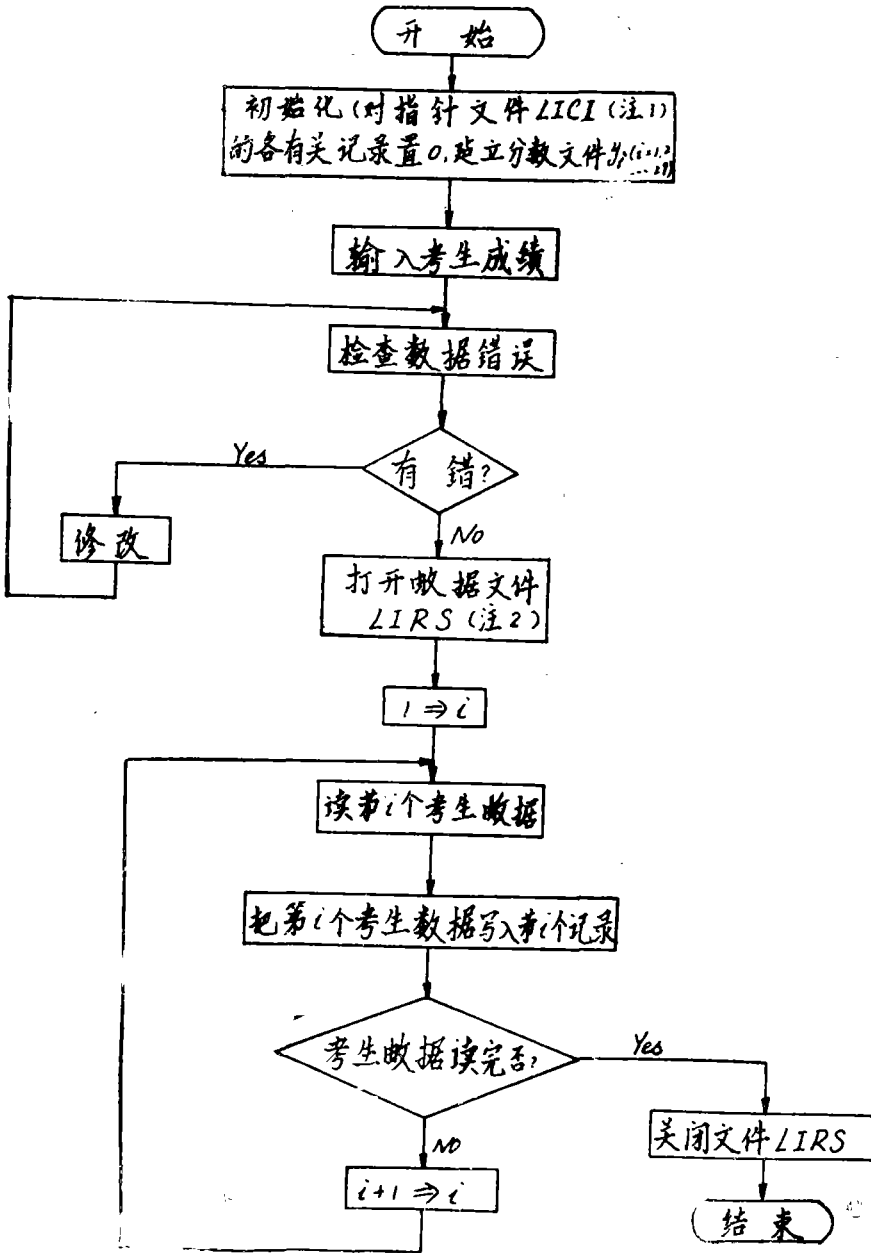
三 注（程序框图中的注）

注1：指针文件LICI存放各个分数文件 y_i ($i=1, 2, \dots, 29$)的最末记录的位置 C_i ；存放第 j 门课 ($j=1, 2, \dots, 7$)的分数累加值 H_j ；存放第 j 门课第 k 档分数（每10分一档，100分自占一档，共11档，即 $k=1, 2, \dots, 11$ ； $j=1, 2, \dots, 7$)的人数 $G_{j, k}$ 。初始化时它们全被置了0。

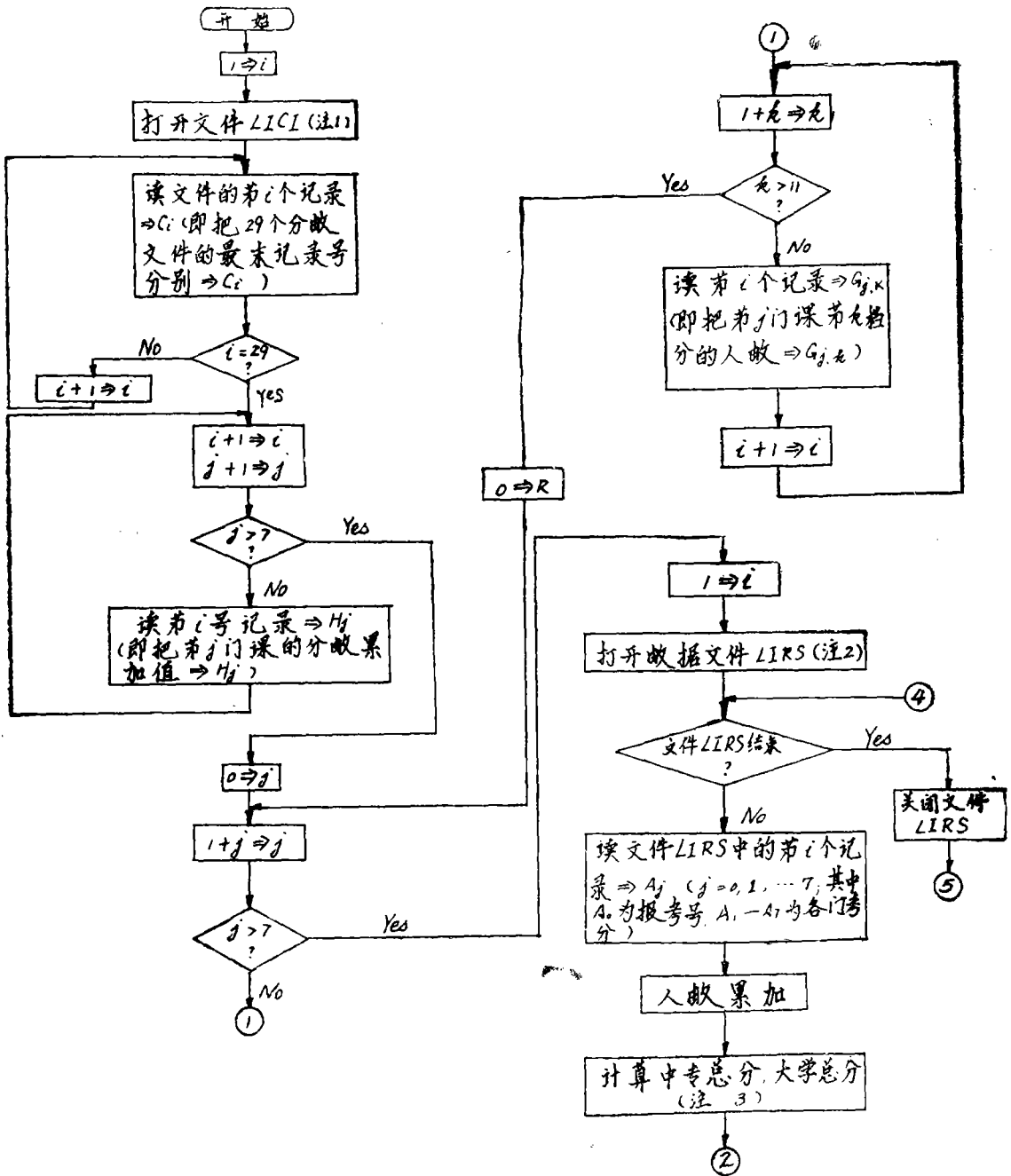
注2：数据文件LIRS存放考生原始数据（报考号、各门课考分），这个文件中的考生数据是经进行检查后一批批输入、生成，一批批处理的。

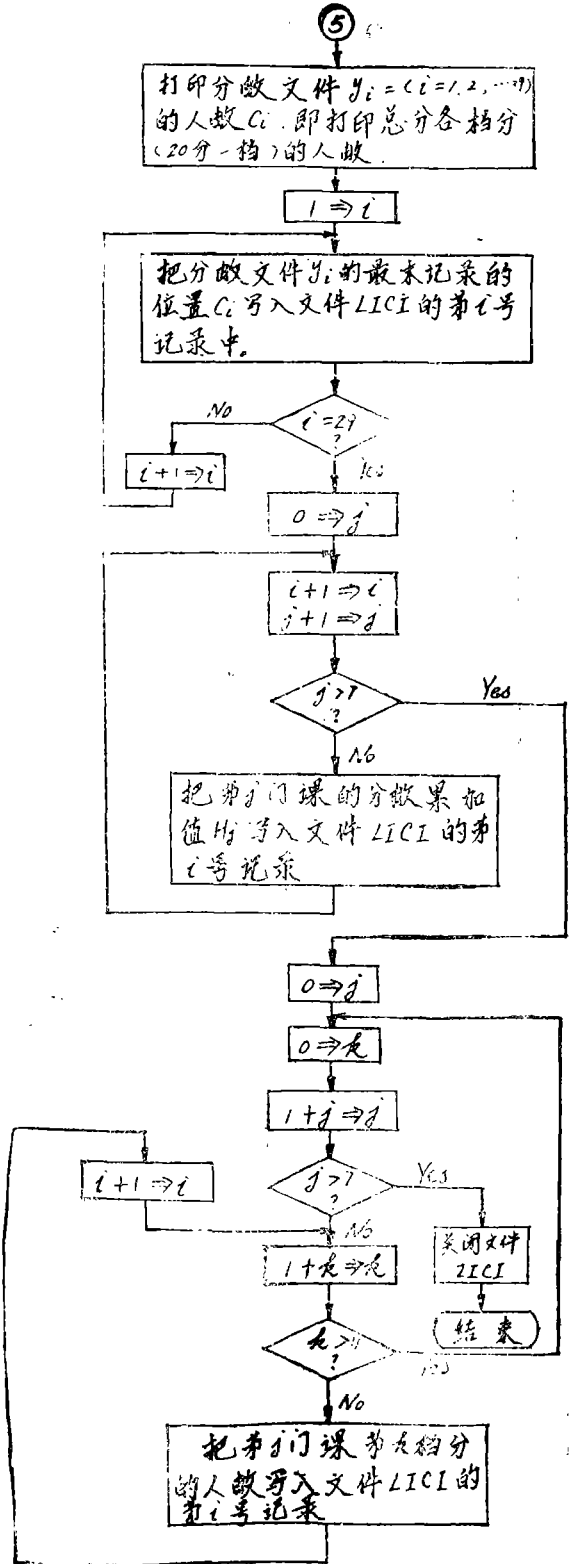
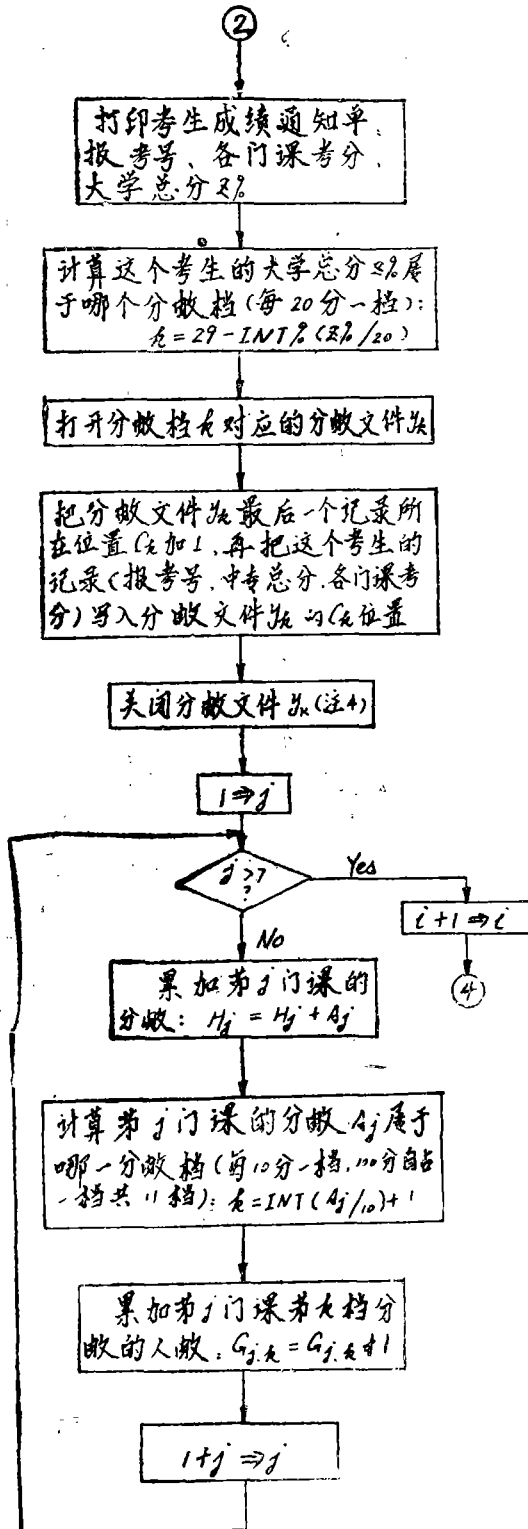
注3：理科的中专总分为各考生政治、语文、数学、物理、化学五门课的成绩累和，大

(一) 生成数据文件框图 (理科)

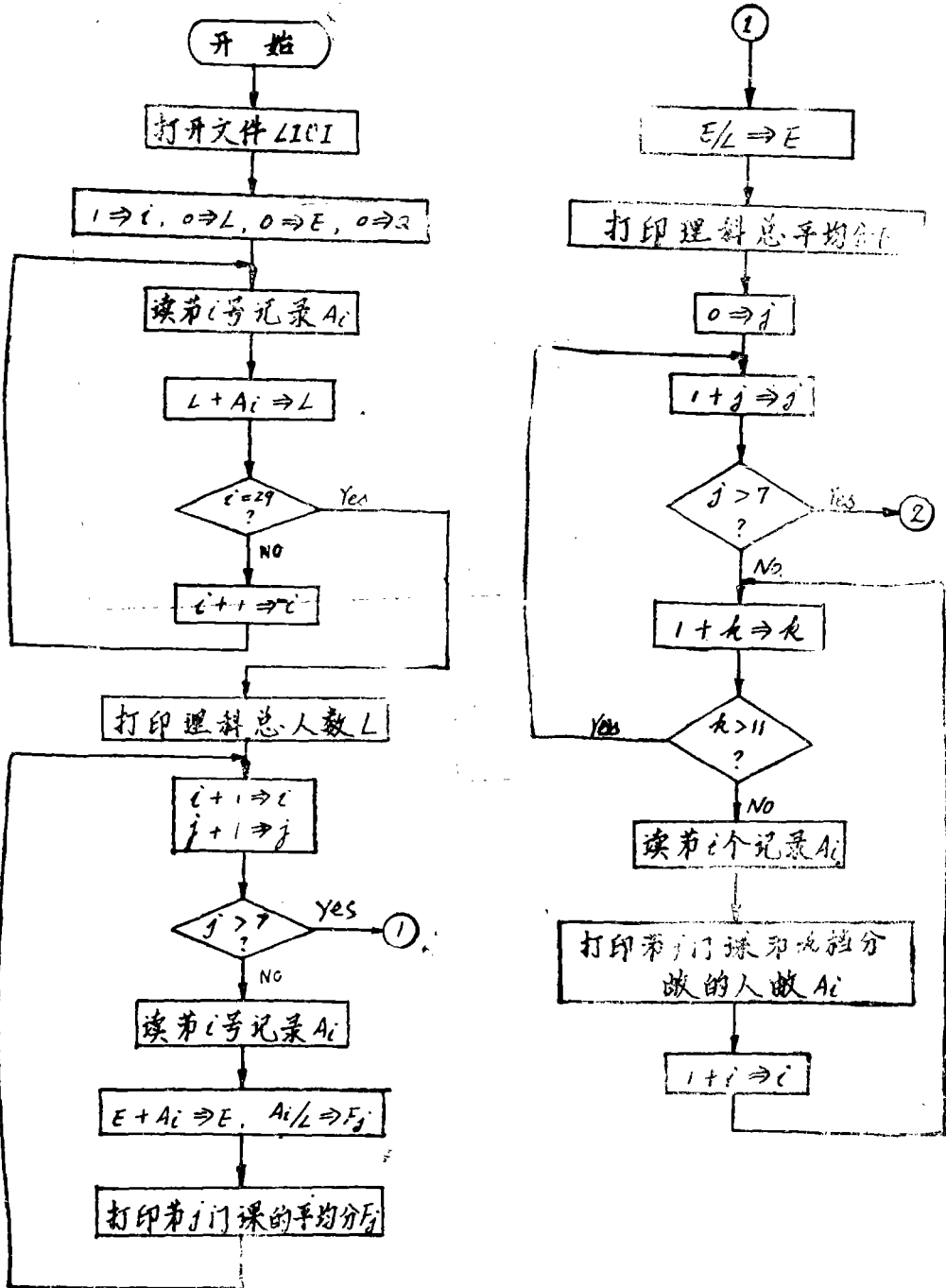


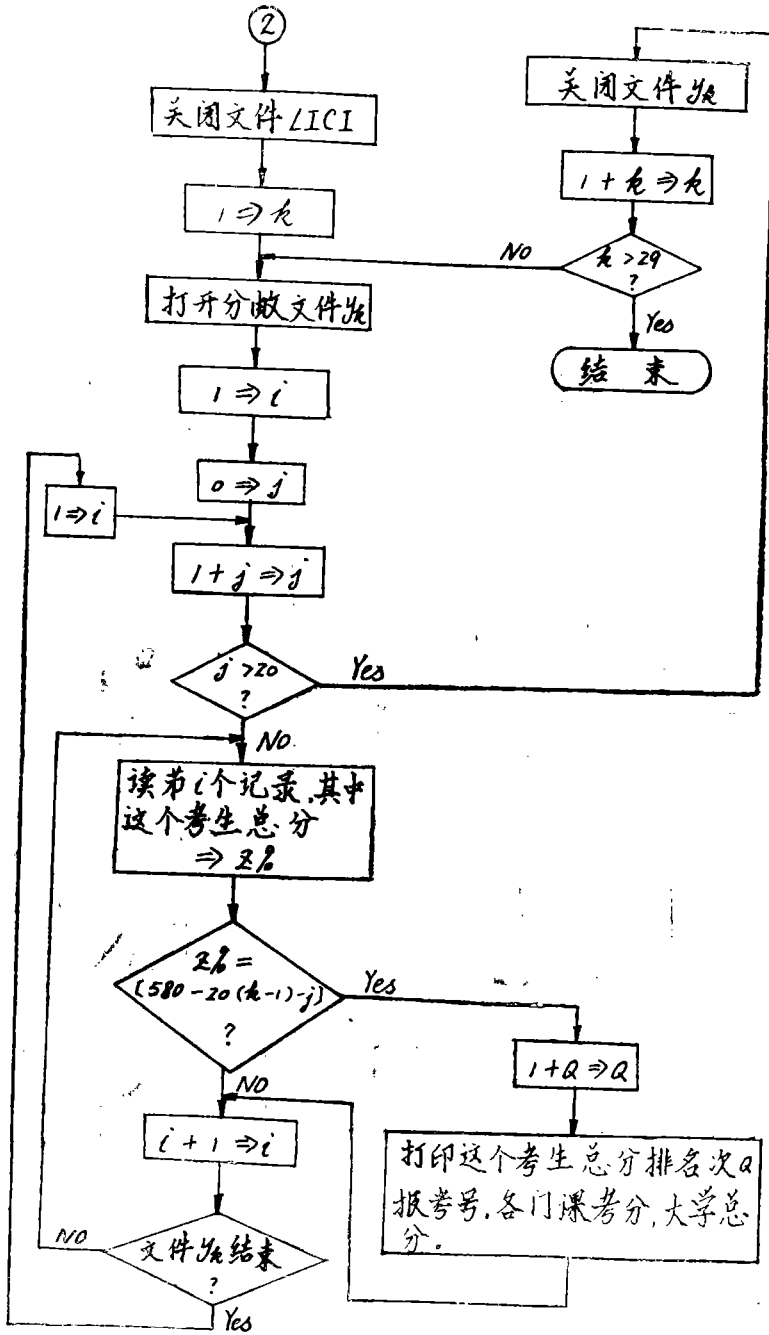
二、对考生分数进行统计,打印考生成绩通知单框图(理科)





(三). 对数据文件. 分数文件进行各种处理. 排序框图(理科):





学总分为如上所述五门课的成绩累和加上30%的生物成绩，再加上50%的外语成绩，满分为580分。

注4：理科考生总分按20分一档分别存入相应的分数文件，共有29个分数文件；文科考生总分按10分一档分别存入相应的分数文件，共有55个分数文件（满分为550分，其中50%的外语分），而Z-80微型机同时只能打开20个文件，为解决此问题，并尽量减少所占内存容量，故使用哪个文件就打开哪个文件，使用完毕便关闭这个文件。

参加本程序编制工作的有符华儿、罗海鹏、曹文全、吴地兴和乔中南。

广西计算中心和南宁市招生办公室部分人员参加了考生数据的输入等工作。