EXPERIENCE WITH A LARGE CLASS, GRADE DECISION SUPPORT SYSTEM

Marvin B. Troutt, Southern Illinois University at Carbondale
Eric Poon, Southern Illinois University

1. Introduction

The authors' academic department is charged with delivering the basic business statistics course requirement in the College of Business. This course has enrollments of up to 400 per regular semester. At present the course is delivered by way of two large lecture sections of approximately equal size, along with one hour labs in smaller sections. Thus each lecturer is responsible for assigning approximately 200 grades each semester. Due to the large number of students per lecturer and to tight constraints on graduate assistant help (this available support is needed in test preparation and laboratory leadership), the manual scoring of tests and manipulation of final grade records is all but precluded.

Thus as a first step towards an automated grade reporting system, all tests were designed to have one of up to five multiple choice answers per question, and all tests were scored using an Opscan 1000 optical scanner. A campus office processes the test response forms and creates a card deck, along with raw scores and various other nonquantitative information, select a satisfactory letter grade to report for the student. Generally the author has picked the best letter grade based on raw rather than standardized test scores shown for each student except when one of the scales is experimental in nature. The scales used so far have been as follows. The first letter grade scale was a conservative, contract type scale announced in the syllabus of the course. Specifically in terms of percentages of total points available on all tests, it is as follows:

<table>
<thead>
<tr>
<th>Range</th>
<th>Letter</th>
<th>Standard Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>X + s and over</td>
<td>A</td>
<td>0.16</td>
</tr>
<tr>
<td>X + 1/3s up to X + s</td>
<td>B</td>
<td>0.21</td>
</tr>
<tr>
<td>X - 1/3s up to X + 1/3s</td>
<td>C</td>
<td>0.26</td>
</tr>
<tr>
<td>X - s up to X - 1/3s</td>
<td>D</td>
<td>0.21</td>
</tr>
<tr>
<td>X - s and lower</td>
<td>F</td>
<td>0.16</td>
</tr>
</tbody>
</table>

A desirable additional scale currently being added to the report is a historical type percentile grade based on raw rather than standardized individual test scores. By picking the best letter on either scale, students then feel they get the type of break often asked for when some test was well above average, or their raw scores were above some minimum %, or other innovative arguments teachers are presented with.

2. Some Desirable Features of Outputs

The first requirement of the information output was taken to be a presentation of letter grades (A, B, etc.) for each student on at least three bases of determination. Thus the DM would be able to look at a vector of letters on each student at the end of the course, and combining other nonquantitative information, select a satisfactory letter grade to report for the student. Generally the author has picked the best letter score shown for each student except when one of the scales is experimental in nature. The scales used so far have been as follows.

In the subject matter and tests used it has been found that means are usually nearer the D range on this scale so that it is conservative from the grader's point of view. However, it does provide the student with guaranteed results for given performances. That is, regardless of whatever the rest of the class does, a student who gets 65% of all points is assured a passing grade.

The second scale used is a historical one for the course. The percentage of A's, B's, and other grades given over the past in the course determine percentile cut-off values for the various grades. The PROC RANKS procedure of SAS allows the data set to include percentile score equivalents as a new variable. Each component test used in the course is first standardized and a total is obtained as the basis of the percentile score. In future modifications it would be desirable to output letter scores on both the standardized and nonstandardized course totals. Students are informed that generally the best letter on any scale is the one chosen and this approach seems highly satisfactory to students as judged by grading fairness questions on teacher-course evaluations at the end of the course, as well as a complaint rate of about 2-3 per term.

A third scale has recently been added to the report. It was derived so that students could judge their progress on each test by knowledge of only the mean and standard deviation of the test. Its normal percentile clusters roughly match the percentages of the historical grading scale, except for an overstatement of failing grades. The scale and its normal percentages are as follows:

3. General Approach

For each individual test deck, a separate data set is created adjoining a variable for the score on that test. These are merged by student ID number, then to be the control field. The merger operation retains the student name from the last data set at final exam scores. It was
found that students will not or (cannot easily) conform to coding their name identically on all tests. Coding examples of the following type have been frequent: Bill Jones, Jones William, Jones Bill, etc. (despite requests for consistency). However after the merge only the most recently coded name version is retained: This might conceivably cause problems when a student has not taken the final test. For this reason and for additional control a separate report was made listing, in both name and ID orders, the students not taking the final. Most of these were those who dropped late, or will be receiving incomplete grades. Thus an additional advantage in segregating these is that percentile ranks assigned to others are based on those with complete work.

4. Input Data

Students code name and ID number on their multiple choice test answer sheets. Processing by the OPSCAN reader creates one card for each student for each test. The card records contain the student identifiers above as well as lecturer, test number, the number of test items, the raw score, and the students response to each item. Although a grading key is submitted with the OPSCAN work and a raw score computed it was found to be all but necessary to be able to modify the raw scores. With multiple choice questions in a large technical subject it has been difficult to modify questions or answer choice during actual examination conditions. Typographical errors are especially frequent due to typists unfamiliarity with the meaning of formulas and technical nuances. Also more often than not the number of instructions. Prior to the use of this system element, it would have been necessary to manually adjust all scores or submit the data again with a new scoring key. However in the latter event it is not possible to give credit for multiple answers directly with the OPSCAN grades in any event. Hence this aspect has greatly contributed to reduction of problems and key work in the grade management of the course.

5. Other Editing and data problems

Make up exam results must be added by way of editing statements. For this and other editing described below it was expedient to create a variable from the students full name (called LNAME). This new variable is an abbreviation of the first 5 letters of the student name card called simply NAME. The following statement included with the data get step for 2 adds a makeup score for a student named ZILLMEN:

IF NAME = 'ZILLM' THEN SCR2 = 20;

Among the most important problems is mis-coding of ID. Typically these errors are due to omission of a digit or column shifting. Despite frequent reminders to code this item carefully, there remains about a 1-3% error rate on this key field. These errors are spotted on a preliminary run of the main report. The name order report is scored for the same name appearing with two distinct ID's after merger. Also on the report of individual test scores the offending miscoded ID will show zeroes on all tests except the one using the miscoded ID. After selection of the correct ID for the student the following type of editing statements are included in the appropriate test data get step:

IF NAME = 'SMITH' THEN ID = 333445555; or
IF ID = '361' THEN ID = 99994361;

In a recent term, a situation of the following type was encountered. An apparent double submission of grading sheets was made by the same student on several of the tests. Apparently the student hoped the program added scores. Perhaps there was time at the end of the test to start copying a second version of the answer sheet. It is not clear whether this was a fraud attempt or motivated for some other reason. However it was not difficult to rectify, no confrontation of the student was made. In the name order report of individual scores the same student would be duplicated with different raw scores on one or more of the tests. An editing statement such as the following threw out the extra identified on the one with the single additional raw score (eg. 5):

IF NAME = 'SMITH' and SCR1 = 5 THEN DELETE;

It is possible that the student misinterpreted the meaning of merger of test records in the description of the grading system in the course syllabus. Currently students are advised that multiple copy submissions will only retain one of the records.

By having the item responses available it is possible to revise each raw score appropriately for situations such as those above. Following are two illustrations of typical modifications used.

SCRI = Raw score on test 1
SCR3 = Raw score on test 3
ITEM14 = Response to item (question) 14
ITEM12 = Similarly for item 14.

Example 1: Item 12 on test 1 was scored as correct for response a (1) but should be scored correct for item d (4). Then the following coding is included in the data set for Test 1:

IF ITEM14 = 1 THEN SCRI = SCRI - 1;
IF ITEM14 = 4 THEN SCRI = SCRI + 1;

Example 2: Item 12 on test 3 was scored as correct for response b (2). Later it is desired to give credit for response c (3) as well. The instruction is:

IF ITEM12 = 3 THEN SCR3 = SCR3 + 1;

Generally any combination of revised keying and raw scores may be obtained with a modest number of instructions.
Occasionally students would take tests with a different lecturer than the correct one. These would show up as having several zero test entries under the correct BY LECTURER listing. The correct lecturer would be determined and an appropriate editing step was inserted. It has been found about equally convenient to manually move the cards to the appropriate data step due to unreliability of coding of lecturer. In future runs of the system it has been decided to operate it separately for each lecturer since the class sizes are each large enough for reliable comparable grading.

6. Report Formats and Usage

Below is the format used for name orders listing:

LECTURER = T

<table>
<thead>
<tr>
<th>NAME</th>
<th>LNAME</th>
<th>IDD</th>
<th>SCR1</th>
<th>SCR2</th>
<th>SCR3</th>
<th>SCR4</th>
<th>SCR5</th>
<th>AVE</th>
<th>LETTER1</th>
</tr>
</thead>
<tbody>
<tr>
<td>JONES</td>
<td>JONESAL</td>
<td>-</td>
<td>14</td>
<td>18</td>
<td>12</td>
<td>20</td>
<td>22</td>
<td>65.75</td>
<td>D</td>
</tr>
<tr>
<td>GONE</td>
<td>GONEDILL</td>
<td>-</td>
<td>16</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td>12.3</td>
<td>F</td>
</tr>
</tbody>
</table>

SUMMARY REPORT = NAME ORDER

<table>
<thead>
<tr>
<th>NAME</th>
<th>LNAME</th>
<th>IDD</th>
<th>AVE</th>
<th>RANK</th>
<th>STD.AVE</th>
<th>STD.RANK</th>
<th>LETTER1</th>
<th>LETTER2</th>
<th>LETTER3</th>
</tr>
</thead>
<tbody>
<tr>
<td>JONES</td>
<td>JONESAL</td>
<td>-</td>
<td>65.75</td>
<td>48</td>
<td>52.92</td>
<td>51</td>
<td>D</td>
<td>C</td>
<td>C</td>
</tr>
</tbody>
</table>

After a preliminary run to clean the data as described above, grades may be transferred to the official university grading forms as a name by name. Generally the best letter shown is the official grade.

As a safeguard against unreasonable grade comparisons it has been found useful to compare lowest C and highest D for instance on the basis of raw test average alone. Some experimentation with clustering techniques have been attempted also. By requesting five clusters on one of more of the variables, the actual grade assignment can be compared with the cluster results. Two problems are found in this general approach however. First the clustering procedure consumes a great amount of CPU time for files this large. Also it is apparently not possible to treat the cluster number as a new variable for listing a comparison. Hence this technique can not be generally recommended from our experience.

A promising technique for future study is that of discriminant analysis. By inputting actual grade assigned to the files it is possible to determine optimal linear discrimination criteria to apply to the hour exam tests, and final exam for instance. This technique might be more valuable if extra criteria are input such as lab scores, or dummy variable type scores for subjective extra credit, now being added to the system.

7. Recent System Augmentations

Besides tests and exams, the students were also graded on their group assignment performances. Each individual's contribution was evaluated by the other team members. Thus, in a group of five, each member would receive four ratings (on a scale of 1 to 9) from their peers.

This requires two decks of input cards. The first deck consists of grades on the team performance, one card per team. On each card, the team number is specified, and the scores for the team performance on the assignments are listed. An average score was calculated for each team to obtain an overall team score on all the assignments.

The second deck consists of peer evaluation scores for each student, one card per student. On each card, the student ID is specified and his ratings by his peers are listed. An average of the peer evaluations were calculated for each member, and for the team as a whole.

The average peer evaluation score for each member was divided by the average peer evaluation for the whole team to obtain the adjusting ratio. In assigning grades of the students, the overall team score was adjusted by the adjusting ratio of each individual. For example, John belongs to team #10, the overall team score was 80, the average peer evaluation for the whole team was 7 and John's peer evaluation score was 8, then the grade assigned to John was = 80 x (8/7) = 91.

The author's mailing address is:

Marvin D. Troutt
Department of Administrative Sciences
211A Rehn Hall
Southern Illinois University
Carbondale, Illinois 62901
USA