

AN EVALUATION OF STUDENT PERFORMANCE AND COMPARISON OF TEACHING METHODS USING A DEPARTMENTAL FINAL EXAM

MICKEY DUNLAP

Department of Mathematics and Statistics, University of Tennessee at Martin, Martin, TN 38238

ABSTRACT—This paper discusses attempts to utilize results from the departmental final exam for an elementary statistics course to understand how well the students are performing in the course as well as comparing 6 instructors. Over the course of three semesters, final exam results from 884 students are used in the analysis. Contingency tables are used to understand the relationship between pairs of variables. Correspondence analysis is used as a descriptive method to graphically understand the relationships among all variables of interest. Multiple tests of independence, as well as tests comparing proportions, are used to further verify the findings from correspondence analysis. As a result, many relationships among instructors, topics, and student classification are found to be highly significant.

There are surely many ways to assess the performance of students as well as professors. In one study by Johnson and Kuennen (2006), much information was collected about the student such as gender, race, and many others. In their paper, instructors were allowed to give their own exams, and grades of students in the course were used to make comparisons. In another study by Gebotys (1990), teaching methods were compared, such as whether or not participants were taught the theory and whether or not participants were taught with examples. While an experimental design would be ideal to control certain factors, they could also restrict some as well. Some instructors may feel more comfortable with their style of teaching but if they are forced to teach differently, the results may favor one instructor over another. There can be differences in learning depending on whether the instructor presents only a definition, an example or a combination of the two (Klausmeier and Feldman, 1975).

The goal of this paper is to assess the performance of students as well as compare the instructors in a department without forcing the instructors to adhere to specific teaching methods. The instructors were not forced to teach in a manner they were not comfortable with. Even though instructors may have varying measures of assessment, such as the number of exams, homework, quizzes, projects, etc., the departmental final exam is an exam that is common to all instructors in the department. The results from the departmental final exam will hopefully allow us to understand which topics the students find most difficult. Furthermore, this will allow us to understand which instructors are most successful in terms of explaining the various topics.

KEY COMPONENTS

The departmental final exam has remained relatively unchanged for a number of semesters. There are 50 multiple choice questions with 4 choices per question. One instructor from the department creates the exam and the remaining

instructors are more than welcome to view the exam and suggest corrections. Topics that are commonly on the final exam are listed below with the average number of questions from each in parentheses:

1. Summarizing and Comparing Data (8)
2. Probability (5)
3. Discrete Probability Distributions (7)
4. Continuous Probability Distributions (7)
5. Confidence Intervals (4)
6. One Sample Hypothesis Testing (8)
7. Inferences for Two Samples (5)
8. Correlation and Regression (2)
9. Multinomial Experiments and Contingency Tables (4)

Data were compiled for three semesters: Fall 2006, Spring 2007, and Fall 2007. The instructors being compared have all taught Elementary Statistics over the course of these semesters. The instructors that teach Elementary Statistics in the department have varying disciplines. Two of the instructors are Statisticians, a few are Mathematicians, and another has a background in industry. Some prefer to give quizzes while others tend to give homework. Some give multiple choice exams, some prefer their students show their work, and others give a mixture of the two. Other differences include the use of PowerPoint notes, the use of a course webpage, and also the use of a calculator, primarily a TI-83. In a given semester, some instructors may have 2 or 3 sections while others may have only 1 section of an Elementary Statistics course.

In this study, there are 884 students. Students come from varying disciplines, such as Biology, Chemistry, Political Science, Business, and many others. In the semesters being considered, 14% of the students were freshmen, 42% were sophomores, 25% were juniors, and 19% were seniors. In this analysis, there is the possibility that a student has taken the final exam in successive semesters. The previous scores of these students were not removed.

TABLE 1. Contingency table for instructor and questions correctly answered.

Instructor	Correct	Incorrect	Total
Instructor 1	1,832	1,568	3,400
Instructor 2	6,353	4,397	10,750
Instructor 3	7,213	4,237	11,450
Instructor 4	1,930	2,420	4,350
Instructor 5	4,736	2,664	7,400
Instructor 6	3,841	3,009	6,850
Total	25,905	18,295	44,200

BASIC ANALYSIS

To begin to understand relationships between the variables, we can create contingency tables. For a particular question that was answered on the final exam, we know who the instructor was, the topic the question came from, the classification of the student, and whether the question was answered correctly. With 50 questions on a final exam and 884 students, this equates to 44,200 observations. For all observations, 25,905 or approximately 58.6% were answered correctly (Table 1).

It would be just as useful, if not more so, to create tables, such as Tables 2–4, containing the proportion of questions answered correctly.

From Table 2, since 58.6% of all questions were correctly answered, we can now understand which instructors have percentages above the overall percentage. Rather than attempt to create a multi-way table for instructor, topic, and whether the student correctly answered the question or not, it will be more useful to create a table with the percentages within the cell for instructor and topic as was done in Table 2. Graphical representations of the above tables can more clearly depict the relationships among the variables. Fig. 1 illustrates that the most difficult topic pertains to Chi-Square tests. This is likely due to the fact that students are asked to determine test statistics, sometimes an involved process. The correlation and regression scores are high because there are usually only two questions on the exam. The material in the middle of the semester is also very difficult, and we should probably make an effort to emphasize this more. Fig. 2 compares the performance of students by their classification. Freshmen and sophomores tend to perform better than juniors and seniors. Perhaps juniors and seniors are already apprehensive about this course and have postponed taking it until near the end of their college career while freshmen and sophomores have likely just completed a Math course in high school or perhaps a college algebra course. Tables 5, 6, and 7 summarize the percentage of questions answered correctly for instructors, topics, and student classifications. Fig. 3 compares the instructors and topics. For a few of the topics, such as 1, 2, 6, and 7, most of the instructors have students that perform about the same. However, there are some obvious differences. Instructor 3 has students that perform better for topics 3, 4, and 5. Instructors 2 and 5 have students that perform better on the correlation and regression questions. Instructor 4 has students that perform very low for topics 5, 6, and 7.

TABLE 2. Percentage of questions correctly answered for each instructor.

Instructor	1	2	3	4	5	6
Percent Correct	53.88	59.10	63.00	44.37	64.00	56.07

CORRESPONDENCE ANALYSIS

Simple correspondence analysis (CA) is an exploratory technique used for analyzing a two-way contingency table and has been compared to principle components analysis by some and to factor analysis by others. Multiple correspondence analysis (MCA) and joint correspondence analysis (JCA) are extensions of simple correspondence analysis for analyzing multi-way contingency tables. A two-way contingency table for the variables instructor and student classification is given (Table 8). For each question answered, the instructor and classification of the student are recorded. The cell count does not represent the number of questions correctly answered.

After dividing all cells as well as row and column totals by the grand total, we have Table 9.

Following the notation of Khatree and Naik (2000), as well as Greenacre and Nenadić (2007), define matrix P as

$$P = \begin{bmatrix} 0.0181 & 0.0271 & 0.0215 & 0.0102 \\ 0.0373 & 0.1199 & 0.0475 & 0.0385 \\ 0.0204 & 0.0995 & 0.0769 & 0.0622 \\ 0.0068 & 0.0396 & 0.0271 & 0.0249 \\ 0.0192 & 0.0690 & 0.0509 & 0.0283 \\ 0.0271 & 0.0735 & 0.0317 & 0.0226 \end{bmatrix}$$

Define the row profiles as being vector of row sums of P as

$$r = [0.0769 \quad 0.2432 \quad 0.2590 \quad 0.0984 \quad 0.1674 \quad 0.1550]$$

and the vector of column sums as

$$c = [0.1290 \quad 0.4287 \quad 0.2557 \quad 0.1867]$$

Let D_r be a diagonal matrix containing the row sums of P . That is,

$$D_r = \text{diag}(r) = \begin{bmatrix} 0.0769 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.2432 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.2590 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0984 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1674 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.1550 \end{bmatrix}$$

and similarly for D_c . Pearson's chi-square test statistic for testing the independence between instructor and student classification can be expressed as

$$\chi^2 = n \sum_{i=1}^6 \sum_{j=1}^4 \frac{(P_{ij} - r_i c_j)^2}{r_i c_j}$$

where n is the total number of observations 44,200. Denote the matrix S as $S = D_r^{-1/2} (P - r c^T) D_c^{-1/2}$. Then, the ij^{th} element of S is

TABLE 3. Percentage of questions correctly answered for each topic.

Topic	1	2	3	4	5	6	7	8	9
Percent Correct	73.53	55.67	62.33	53.81	59.94	51.61	60.33	73.13	41.20

$$\frac{P_{ij} - r_i c_j}{\sqrt{r_i c_j}}$$

Therefore, $\chi^2 = n \times tr(SS^T)$ where $tr(SS^T)$ refers to the sum of the diagonal elements or trace of the matrix SS^T . The quantity χ^2/n is referred to as the *total inertia*, which is all the information available in the contingency table. Correspondence analysis can be viewed as a technique for decomposing the total inertia. The trace of the matrix SS^T is equal to the sum of its non-zero eigenvalues, and the rank of SS^T is equal to the number of non-zero eigenvalues. It turns out that the matrix SS^T has rank 3, meaning that the total inertia can be represented in 3-dimensional space. The singular value decomposition, a method to break down a matrix into its most important components, is applied to the matrix S , resulting in scores for each category for each dimension. After applying the singular value decomposition to the matrix S , we can express S as $S = UDV^T$ where

$$U = \begin{bmatrix} -0.3450 & 0.7773 & -0.3806 \\ -0.4206 & -0.5061 & 0.0024 \\ 0.5979 & 0.0499 & -0.2344 \\ 0.3873 & -0.1768 & -0.2968 \\ 0.1129 & 0.2930 & 0.8410 \\ -0.4288 & -0.1419 & -0.0696 \end{bmatrix}$$

$$D = \begin{bmatrix} 0.1824 & 0 & 0 \\ 0 & 0.0908 & 0 \\ 0 & 0 & 0.0385 \end{bmatrix}$$

and

$$V = \begin{bmatrix} -0.7074 & 0.4469 & -0.4135 \\ -0.2622 & -0.6151 & 0.3523 \\ 0.3920 & 0.6263 & 0.4455 \\ 0.5265 & -0.1722 & -0.7117 \end{bmatrix}$$

The standard coordinates or dimensions of the rows and columns are found by

TABLE 4. Percentage of questions correctly answered for each student classification.

Student Classification	Student Classification			
	Freshmen	Sophomores	Juniors	Seniors
Percent Correct	59.60	60.11	57.34	56.23

$$X = D_r^{-1/2} U = \begin{bmatrix} -1.2440 & 2.8026 & -1.3721 \\ -0.8529 & -1.0261 & 0.0049 \\ 1.1746 & 0.0981 & -0.4605 \\ 1.2344 & -0.5634 & -0.9460 \\ 0.2758 & 0.7160 & 2.0555 \\ -1.0893 & -0.3604 & -0.1767 \end{bmatrix}$$

and

$$Y = D_c^{-1/2} V = \begin{bmatrix} -1.9698 & 1.2444 & -1.1514 \\ -0.4004 & -0.9395 & 0.5381 \\ 0.7752 & 1.2387 & 0.8810 \\ 1.2188 & -0.3985 & -1.6472 \end{bmatrix}$$

respectively. Each row in the matrix X corresponds to an instructor, and each row in Y corresponds to a student classification. The standard coordinates are scaled in such a way that the weighted average is 0 and the weighted sums of squares is 1. For example, the first standard coordinate of the rows satisfies

$$\sum_{i=1}^6 X_{i1} r_i = 0 \text{ and } \sum_{i=1}^6 X_{i1}^2 r_i = 1.$$

The diagonal elements of the matrix D from the singular value decomposition are referred to as singular values, and the square of these elements are the eigenvalues of SS^T given in the table below are also referred to as principal inertia. The ratio of the principal inertia to the total inertia is a means of understanding the contribution of the respective standard coordinate to the total inertia. Since the first two coordinates

TABLE 5. Percentage of questions correctly answered by topic and instructor.

Topic	Percentage of Questions Answered Correctly					
	Instructor					
	1	2	3	4	5	6
1	65.69	74.02	76.13	67.47	79.56	69.59
2	54.34	57.78	52.06	53.85	59.10	56.57
3	55.14	59.78	69.82	52.08	65.75	60.36
4	43.49	50.70	66.81	44.17	51.83	43.90
5	53.85	58.54	71.46	30.49	63.06	60.67
6	55.66	55.32	51.43	30.90	55.61	53.24
7	57.47	55.48	69.17	26.42	71.10	64.39
8	66.91	85.81	63.54	63.79	82.09	68.61
9	32.72	45.00	42.47	24.71	59.12	28.47

TABLE 6. Percentage of questions correctly answered by topic and student classification.

Topic	Percentage of Questions Answered Correctly			
	Student Classification			
	Freshmen	Sophomores	Juniors	Seniors
1	73.67	74.05	73.61	72.15
2	56.27	57.12	52.80	55.81
3	62.57	64.90	60.77	58.46
4	51.63	53.26	52.40	53.16
5	61.45	63.39	55.57	57.12
6	55.36	53.16	49.35	48.56
7	61.83	62.41	60.10	54.92
8	75.44	76.65	69.91	67.88
9	42.11	40.90	43.47	38.18

account for a reasonable percentage of the total inertia, 96.6%, it would be feasible to analyze only these rather than all three dimensions (Table 10).

A plot of the first two standard coordinates can be seen (Fig. 4). Notice that the coordinates are negative of those above which has no effect on the interpretation. To understand this graph, consider the following (Table 11), a slight modification of Table 8.

For Instructor 1, $800/3,400 = 23.53\%$ of the students were freshmen. The column average for freshman is found as $5,700/44,200 = 12.71\%$. For each instructor, we are interested in the instructor(s) that correspond to percentages that are above the average. For juniors, Instructors 1, 3, 4 and 5 all have percentages (27.94%, 29.69%, 27.59%, and 30.41%, respectively) which are above the average (25.57%). Notice the relation of juniors to these 3 instructors in the figure. Similarly, Instructors 2 and 6 have above average percentages for sophomores. Likewise, for freshmen, Instructor 4 has the lowest percentage, and we see that these 2 categories are located at opposing sides of the graph. Furthermore, compare the percentages across student classification for Instructors 2 and 6. We see that they have fairly similar percentages (15.35% vs. 17.52%, 49.3% vs. 47.45%, etc.). These categories are also very close in relation to each other. In Fig. 5A, the categories for student classification have been placed at the vertices of a

TABLE 7. Percentage of questions correctly answered by student classification and instructor.

Student Classification	Percentage of Questions Answered Correctly					
	Instructor					
	1	2	3	4	5	6
Freshmen	59.50	55.76	68.56	50.00	68.59	54.25
Sophomores	54.25	61.26	64.20	46.86	64.07	58.25
Juniors	48.84	59.05	61.15	42.17	64.36	53.00
Seniors	53.56	55.65	61.53	41.27	60.08	55.50

Comparison of Topics

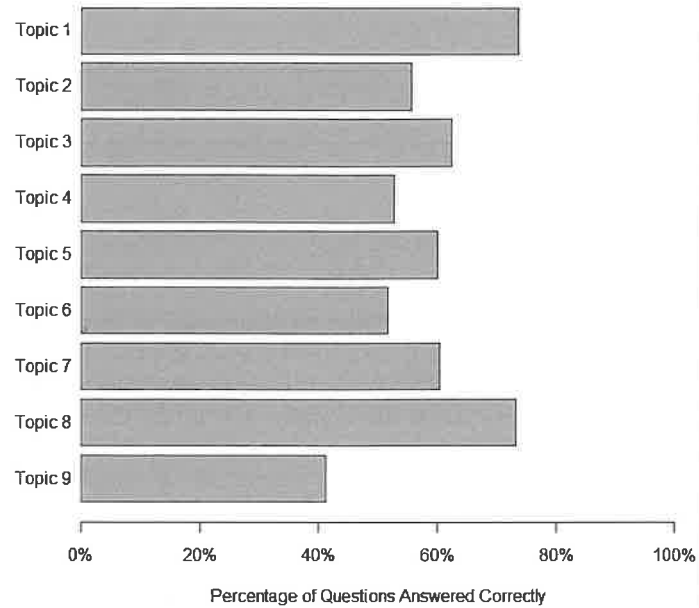


FIG. 1. Bar chart comparing the percentage of correctly answered questions for each topic.

three-dimensional tetrahedron that are one unit apart. These vertices are (0,0,0) for freshman, $(0.5, \sqrt{3}/2, 0)$ for sophomore, $(1,0,0)$ for junior, and $(0.5, \sqrt{3}/6, \sqrt{6}/3)$ for senior. The positions of the points for instructor are weighted averages of the vertices determined by student classification. For example, using the first row of percentages from the previous table, the X coordinate for instructor 1 is

Comparison of Student Classification by Topic

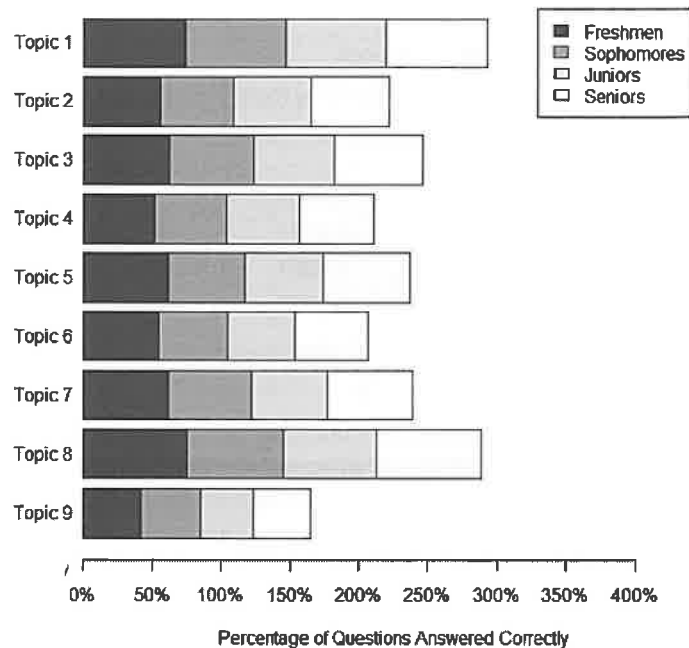


FIG. 2. Bar chart comparing the percentage of correctly answered questions for each student classification by topic.

Comparison of Instructors by Topic

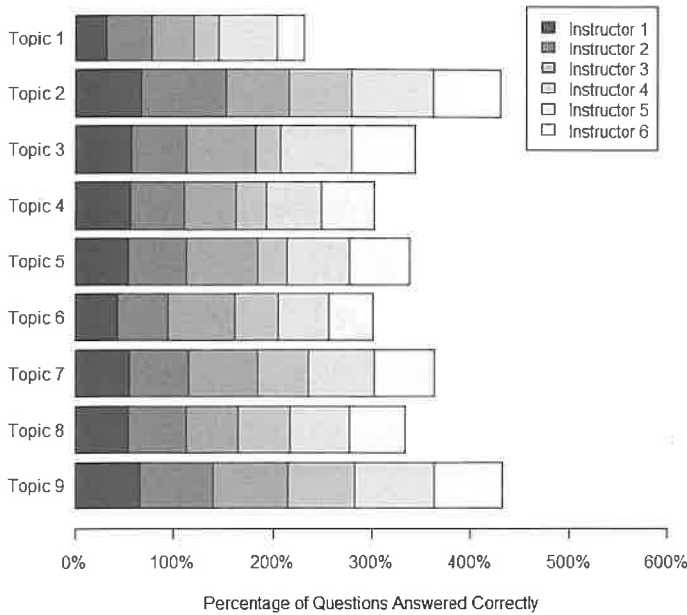


FIG. 3. Bar chart comparing the percentage of correctly answered questions for each student classification by topic for each instructor.

$$\begin{aligned}
 &0.2353 \times \text{freshman} + 0.3529 \times \text{sophomore} \\
 &+ 0.2794 \times \text{junior} + 0.1324 \times \text{senior} \\
 &= 0.2353 \times 0 + 0.3529 \times 0.5 + 0.2794 \times 1 + 0.1324 \times 0.5 \\
 &= 0.52205
 \end{aligned}$$

Continuing in this fashion, the position for instructor 1 is (0.52205, 0.3438, 0.1081). Suppose that instructor 1 only has sophomore students, then the coordinates would be (0.5, $\sqrt{3}/2$, 0), the same as the vertex for sophomore. On the other hand, suppose instructor 1 has an equal percentage of students from each classification. Then the position of instructor 1 would be the average of the vertices of all categories. In other words, the points for instructors are in relation to the categories for student classification. If there was

TABLE 8. Contingency table for instructor and student classification for each question.

Instructor	Student Classification				Total
	Freshmen	Sophomores	Juniors	Seniors	
Instructor 1	800	1,200	950	450	3,400
Instructor 2	1,650	5,300	2,100	1,700	10,750
Instructor 3	900	4,400	3,400	2,750	11,450
Instructor 4	300	1,750	1,200	1,100	4,350
Instructor 5	850	3,050	2,250	1,250	7,400
Instructor 6	1,200	3,250	1,400	1,000	6,850
Total	5,700	18,950	11,300	8,250	44,200

TABLE 9. Percentages for instructor and student classification divided by the grand total.

Instructor	Percentage of Questions Answered Correctly				
	Student Classification				
	Freshmen	Sophomores	Juniors	Seniors	Total
Instructor 1	1.81	2.71	2.15	1.02	7.69
Instructor 2	3.73	11.99	4.75	3.85	24.32
Instructor 3	2.04	9.95	7.69	6.22	25.90
Instructor 4	0.68	3.96	2.71	2.49	9.84
Instructor 5	1.92	6.90	5.09	2.83	16.74
Instructor 6	2.71	7.35	3.17	2.26	15.50
Total	12.90	42.87	25.57	18.67	100

a weak association between instructor and student classification, the positions for instructors would be in the middle of the tetrahedron. Since this is a class for second-year students and due to the fact that the percentages in the previous table are highest for sophomore, it is not surprising that the positions for all instructors are closest to sophomore. Fig. 5B is identical to Fig. 5A with the viewpoint changed. The viewpoint was specifically chosen so that the locations of student classification are similar to those in Fig. 4. Also notice that the locations of instructors are also similar as well. Essentially, Fig. 4 is a projection of Fig. 5B to two dimensions. In much the same way, the standard coordinates for the row and column variables are weighted averages.

When there are more than two categorical variables, we can use either multiple correspondence analysis or joint correspondence analysis. For MCA, there are a few ways to perform the analysis and in particular the Burt matrix is involved. For the variables instructor and student classification, the Burt matrix is given below (Table 12).

The Burt matrix is essentially a symmetric 2×2 matrix of tables. Each variable is crossed with itself to form the diagonals while the off diagonals are the same as Table 7 above. With four categorical variables, the Burt matrix will be a symmetric 4×4 matrix of tables. Multiple correspondence analysis is the result of performing simple correspondence analysis on the entire Burt matrix. In the Burt matrix above, the inertia data resulting from performing simple correspondence analysis separately on each of the diagonal tables for instructor and student classification are 5 and 3 respectively.

TABLE 10. Breakdown of total inertia for all three standard coordinates.

Standard Coordinate	Principal Inertia	Percentage of Inertia	Cumulative Percentage of Inertia
1	0.033263	77.4	77.4
2	0.008252	19.2	96.6
3	0.001483	3.4	100

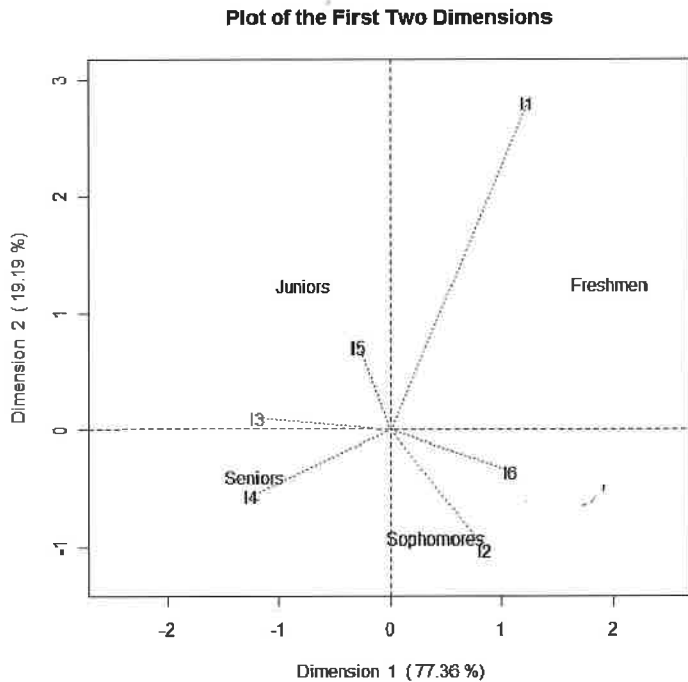


FIG. 4. Plot of the first two dimensions from correspondence analysis for instructor and student classifications.

The inertia for each off-diagonal table is 0.0429. As a result, the total inertia from performing MCA on the entire Burt matrix is the average of the inertias $(5 + 3 + 0.0429 \times 2)/4 = 2.02145$. Furthermore, only 32.01% of the total inertia is explained by the first two dimensions. According to Greenacre (2007), by including the diagonal blocks, the total inertia will be inflated and that joint correspondence analysis improves measures of total inertia. Joint correspondence analysis involves an iterative algorithm in which CA is performed on the Burt matrix just as with MCA except that the diagonal tables are updated at each step. From the discussion of simple correspondence analysis, $S = D_r^{-1/2}(P - rc^T)D_c^{-1/2}$, and from the singular value decomposition of S , $S = UDV^T$. Equating these expressions and solving for P gives

TABLE 11. Percentage of students for each instructor.

Instructor	Percentage of Students			
	Student Classification			
	Freshmen	Sophomores	Juniors	Seniors
Instructor 1	23.53	35.29	27.94	13.24
Instructor 2	15.35	49.30	19.53	15.81
Instructor 3	7.86	38.43	29.69	24.02
Instructor 4	6.90	40.23	27.59	25.29
Instructor 5	11.49	41.22	30.41	16.89
Instructor 6	17.52	47.45	20.44	14.60
Average	12.90	42.87	25.57	18.67

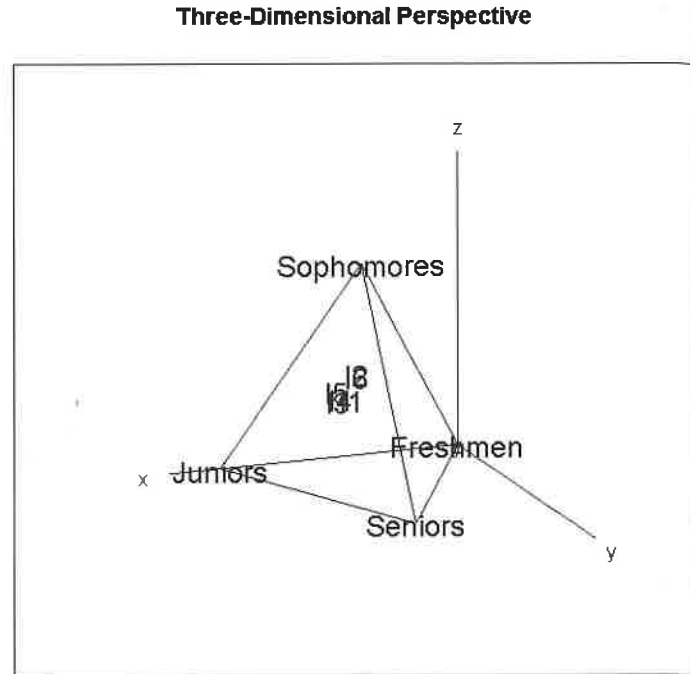


FIG. 5A. Three-dimensional tetrahedron with student classifications placed at the vertices. The positions of instructor are directly related to these vertices.

$$D_r^{-1/2}(P - rc^T)D_c^{-1/2} = UDV^T$$

$$D_r^{-1/2}D_r^{-1/2}(P - rc^T)D_c^{-1/2}D_c^{-1/2} = D_r^{-1/2}UDV^TD_c^{-1/2}$$

$$D_r^{-1}(P - rc^T)D_c^{-1} = XDY^T$$

$$P - rc^T = D_rXDY^TD_c$$

$$P = rc^T + D_rXDY^TD_c$$

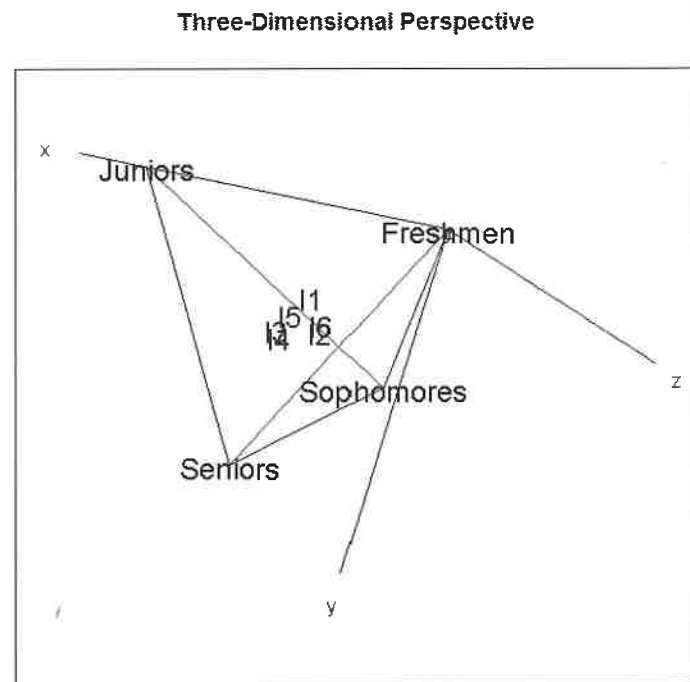


FIG. 5B. A different viewpoint of Fig. 5A. The axes have been rotated to give a different perspective.

TABLE 12. The Burt matrix for instructor and student classification.

	Instructors						Student Classifications			
	I1	I2	I3	I4	I5	I6	Fr.	Soph.	Jun.	Sen.
I1	3,400	0	0	0	0	0	800	1,200	950	450
I2	0	10,750	0	0	0	0	1,650	5,300	2,100	1,700
I3	0	0	11,450	0	0	0	900	4,400	3,400	2,750
I4	0	0	0	4,350	0	0	300	1,750	1,200	1,100
I5	0	0	0	0	7,400	0	850	3,050	2,250	1,250
I6	0	0	0	0	0	6,850	1,200	3,250	1,400	1,000
Fr.	800	1,650	900	300	850	1,200	5,700	0	0	0
Soph.	1,200	5,300	4,400	1,750	3,050	3,250	0	18,950	0	0
Jun.	950	2,100	3,400	1,200	2,250	1,400	0	0	11,300	0
Sen.	450	1,700	2,750	1,100	1,250	1,000	0	0	0	8,250

known as the reconstruction formula. Now, the individual cell counts in the table can be reconstructed by simply multiplying each element in P by the grand total, n . The vectors r and c as well as the matrices D_r and D_c are determined by the diagonal table of interest. Also, $D_r = D_c$ and the diagonal elements of these matrices are those in the corresponding vectors which are also equal. The matrices X , D , and Y are the result of performing simple correspondence analysis on the entire Burt matrix. Once all diagonal tables have been reconstructed, the procedure starts over by performing CA on the updated Burt matrix and reconstructing the cells that correspond to the diagonal tables. The iteration stops when all cells in each of the diagonal tables do not differ by a specified amount from the values in the previous iteration. Ultimately, the tables along the diagonal of the updated Burt matrix will no longer be diagonal, but the row and column totals will be maintained.

The off-diagonal tables remain unchanged throughout the process. Furthermore, performing CA on the newly updated diagonal tables will result in a dimension of one and an inertia that is negligible. Therefore, the total inertia to be explained for the updated Burt matrix is the average of the inertias of the off-diagonal tables.

Taking into account all categorical variables_instructor, student classification, topic, correct/incorrect_with multiple correspondence analysis, only 16.82% of the total inertia is explained by the first two dimensions, and 88% is explained by the first two dimensions when joint correspondence analysis is used. In both instances, the 4×4 Burt matrix described above is used. A plot of the first two dimensions after applying JCA is given (Fig. 6). From this graph, the most obvious characteristic is that correct/incorrect and topic are aligned along one axis while instructor and student classification are

Plot of the First Two Dimensions

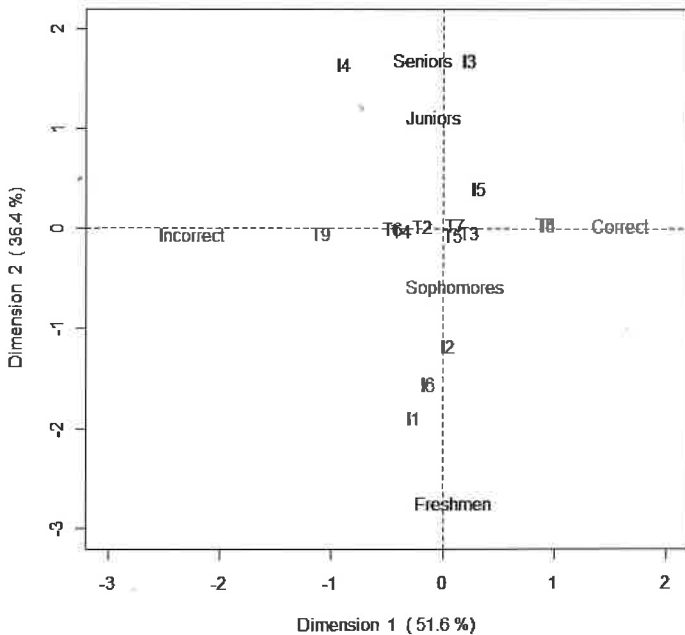


FIG. 6. Plot of first two dimensions from joint correspondence analysis.

Plot of the Second and Third Dimensions

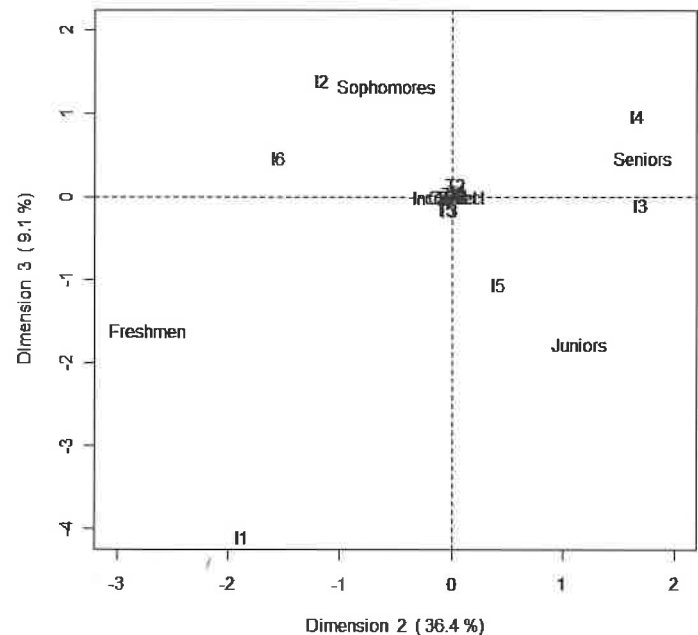


FIG. 7. Plot of second and third dimensions from joint correspondence analysis.

TABLE 13. Results of performing tests of independence.

Test of Independence	Test Statistic	Degrees of Freedom	P-Value
Complete Independence	4977.952	120	< 0.0001
Correct/Incorrect vs. Remaining Variables *	2952.048	215	< 0.0001
Instructor vs. Remaining Variables *	3350.739	355	< 0.0001
Topic vs. Remaining Variables	2295.724	376	< 0.0001
Student Classification vs. Remaining Variables	2196.734	321	< 0.0001
Correct/Incorrect vs. Instructor	593.68	5	< 0.0001
Correct/Incorrect vs. Topic	1509.132	8	< 0.0001
Correct/Incorrect vs. Stud. Class	46.5739	3	< 0.0001
Instructor vs. Topic	8.9138	40	> 0.9999
Instructor vs. Stud. Class	1900.5	15	< 0.0001
Topic vs. Stud. Class	1.6821	24	> 0.9999

* Only one expected cell count is less than 5.

aligned along the other. The topics are arranged from left to right according to the percentage of questions answered correctly. The same is true for instructors and student classifications (Tables 1, 2, and 3). For Fig. 7, a plot of dimension 2 and 3, one characteristic that stands out is that correct/incorrect and topics have all been moved to one location near the origin, leading one to conclude that the variables of interest in this plot are instructor and student classification. As a matter of fact, if we compare this plot to Fig. 4, we see that they are essentially the same, and therefore the conclusion is the same. As mentioned earlier, when the positions of categories for one variable are in the middle of the graph in relation to another variable, a weak association is suggested. Clearly, there is a strong association in Fig. 6 but a weak association in Fig. 7. To determine if there are any significant relationships, we will need to perform statistical inferences.

HYPOTHESIS TESTS

With a multidimensional contingency table, we can test various forms of independence. The Pearson chi-square test was mentioned above in the discussion of correspondence analysis and will continue to be used in these tests. First, some notation needs to be discussed. Let n_{ijkl} be the number of observations in the i -th category for correct/incorrect, the j -th category for instructor, the k -th category for topic, and the l -th category for student classification where $i = 1, 2$; $j = 1, \dots, 6$; $k = 1, \dots, 9$, and $l = 1, \dots, 4$. Indices replaced by ‘.’ mean to sum over that variable. For example, $n_{i...}$ refers to the total number of observations, 44,200. Similarly, $n_{i..}$ refers to the number of observations in the incorrect/correct category. Referring to Table 7, we see that $n_{1..} = 3,400$, $n_{..2} = 18,950$ and $n_{1,2} = 5,300$. Furthermore, the proportion of observations is found by $p_{ijkl} = n_{ijkl}/n_{i...}$.

First, we will be interested in testing complete independence which means that the null hypothesis is $H_0 : p_{ijkl} = p_{i...} \times p_{j..} \times p_{.k.} \times p_{...l}$. If H_0 is true, then we would expect to observe cell frequencies of

$$\frac{n_{ijkl}}{n_{i...}} = \frac{n_{i...}}{n_{i...}} \times \frac{n_{j..}}{n_{j..}} \times \frac{n_{.k.}}{n_{.k.}} \times \frac{n_{...l}}{n_{...l}} \Rightarrow n_{ijkl} = \frac{n_{i...} \times n_{j..} \times n_{.k.} \times n_{...l}}{(n_{i...})^3}$$

Denote these expected counts as \widehat{n}_{ijkl} . The test statistic for testing the claim is

$$\sum_{i,j,k,l} \left(\frac{(n_{ijkl} - \widehat{n}_{ijkl})^2}{\widehat{n}_{ijkl}} \right).$$

When the null is true, the test statistic follows a chi-square distribution with $(I-1)(J-1)(K-1)(L-1) = (2-1)(6-1)(9-1)(4-1) = 120$ degrees of freedom. As a general rule when performing this test, there cannot be too many cells with expected counts less than 5. In this case, there are none. The resulting test statistic is 4,977.952, resulting in a P-value < 0.0001. Clearly, complete independence does not exist among the variables.

Second, we could test whether one variable is independent of the other three. There are 4 tests that can be performed, one of those being $H_0 : p_{ijkl} = p_{i...} \times p_{.jkl}$. Similar to the previous test, the expected cell frequencies would be $n_{ijkl} = \frac{n_{i...} \times n_{.jkl}}{(n_{i...})^2}$. In this case, only one of the 432 expected cell frequencies is less than 5. The test statistic which is found in the same manner is 2,952.048 and the degrees of freedom for the test is $(I-1)(JKL-1) = (2-1)(6 \times 9 \times 4 - 1) = 215$, resulting in a P-value < 0.0001. The variable correct/incorrect is not independent of the other three variables. The usual Pearson chi-square test for two variables will also be of interest. The results of the various tests are summarized (Table 13). Almost all tests are significant at the 0.05 level except when testing for independence for instructor/topic and topic/student classification, but these should not be surprising. If these two tests were significant, it would imply, for example, that some student classifications are tested over a different proportion of topics than others. Now, we know that these are the two weak associations present in Fig. 7. Since the test of independence for correct/incorrect vs. instructor was rejected, it would be interesting to determine which instructors have different proportions.

The results of making multiple comparisons for the proportion of correct answers for instructor, topic, and student classification are given in the following tables. Each cell represents the P-value when comparing the proportion of correct answers for two categories. For example, we see that

TABLE 14a. Multiple comparison results for instructors.

	Instructor 1	Instructor 6	Instructor 2	Instructor 3	Instructor 5
Instructor 4	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Instructor 1		0.5637	< 0.0001	< 0.0001	< 0.0001
Instructor 6			0.0012	< 0.0001	< 0.0001
Instructor 2				< 0.0001	< 0.0001
Instructor 3					> 0.9999

TABLE 14b. Multiple comparison results for topics.

	Topic 6	Topic 4	Topic 2	Topic 5	Topic 7	Topic 3	Topic 8	Topic 1
Topic 9	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Topic 6		>0.9999	0.0048	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Topic 4			0.1101	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Topic 2				0.0053	0.0036	<0.0001	<0.0001	<0.0001
Topic 5					>0.9999	0.8372	<0.0001	<0.0001
Topic 7						>0.9999	<0.0001	<0.0001
Topic 3							<0.0001	<0.0001
Topic 8								>0.9999

the proportion of correct answers is significantly different for instructors 1 and 2 but not for instructors 3 and 5. To make understanding the relationships easier, the categories have been arranged from smallest to largest (Tables 14a, 14b, and 14c). By this arrangement, we see that proportions are significantly different except for instructors 1 and 6, as well as for 3 and 5. For topics, approximately 4 groups can be seen: topic 9; topics 2, 4, and 6 are interconnected; topics 3, 5, and 7; and topics 1 and 8. For student classification, there are two distinct groups juniors/seniors and freshman/sophomore. These results further substantiate the results from correspondence analysis, in particular, the result from Fig. 6. Notice the groupings of topics from the multiple comparisons which also occur in Fig. 6. We see that on the left side of the x-axis, topics 2, 4, and 6 are very close in location and on the right side of the x-axis, topics 3, 5, and 7 are also close. Similar groupings are also seen for student classification. The groupings for instructor are not so obvious but are still there.

TABLE 14c. Multiple comparison results for student classification.

	Juniors	Freshmen	Sophomores
Seniors	0.7599	0.0005	<0.0001
Juniors		0.0304	<0.0001
Freshmen			>0.9999

CONCLUSION

As educators, we should be interested in measuring the performance of our students to find more effective techniques in their understanding of the material. We can change our own technique from semester to semester, or we can compare our own methods with our colleagues' methods. The joint correspondence analysis allows us to understand the relationships between the various variables and how they interact. Tests of independence and multiple comparisons for proportions confirm these relationships.

Obviously, there are disadvantages to using a multiple choice exam to make comparisons. The students may not know how to work the problem or they may know but will make a minor mistake. In both cases, the result is the same in that the question is simply counted as wrong. The advantage of using the departmental final exam is that the difficulty level is entirely the same for all students assuming that the instructor was able to adequately cover the material.

Future considerations could perhaps include a comparison of the students by gender or major course of study. The R programming language (R Development Core Team, 2009) was used extensively in obtaining results.

LITERATURE CITED

GEBOTYS, R. J. 1990. A comparison of teaching methods for probability and statistics. *ASA Proceedings of Statistical Education*: <http://www.amstat.org/publications/jse/v4n3/abbreviations.html>. Accessed 4 December 2009.

- GREENACRE, M. 2007. Correspondence analysis in practice. 2nd ed. Chapman and Hall, Boca Raton, Florida.
- GREENACRE, M., AND O. NENADIC. 2007. Correspondence analysis in R, with two- and three-dimensional graphics: the CA package. *J. of Statistical Software*, 20:1–13, <http://www.jstatsoft.org/v20/i03/>. Accessed 4 December 2009.
- JOHNSON, M., AND E. KUENNEN. 2006. Basic math skills and performance in an introductory statistics course. *J. of Statistics Ed.*, 14, <http://www.amstat.org/publications/jse/v14n2/johnson.html>. Accessed 4 December 2009.
- KHATTREE, R., AND D. N. NAIK. 2000. Multivariate data reduction and discrimination with SAS software. SAS Institute, Inc., Cary, North Carolina.
- KLAUSMEIER, H. J., AND K. V. FELDMAN. 1975. Effects of definition and a varying number of examples and nonexamples on concept attainment. *J. of Ed. Psychol.* 67:174–178.
- R DEVELOPMENT CORE TEAM. 2009. R: A language and environment for statistical computing. Vienna, Austria. <http://www.R-project.org>. Accessed 4 December 2009.