# Using Multivariate Rank Sum Tests to Evaluate Effectiveness of Computer Applications in Teaching Business Statistics

by

Yeong-Tzay Su, Professor
Department of Mathematics
Kaohsiung Normal University
Kaohsiung, TAIWAN 80264

Chyi-Lyi (Kathleen) Liang, Assistant Professor
Department of Community Development and Applied Economics
The University of Vermont
Burlington, Vermont

Selected Paper presented in the American Agricultural Economics Association Annual Meeting, Nashville, Tennessee, August 8-11, 1999.

## ABSTRACT

Arguments about using computer facilities to classroom teaching have gained a lot of attention over time. Using the computer facilities will be helpful to demonstrate real world applications, while poor data or inappropriate case studies might reduce the confidence in applying computer programs in classroom teaching. In this article we examine the results of using computer facilities to teach Business statistics to a group of Management students sampled from Krannert School of Management, Purdue University. This study shows that students are attracted to the interactive computer programs designed for the Business Statistics course, and students are more motivated to attend classes when using computer facilities are applied in teaching. Furthermore, computer programs help students to understand confusing topics (such as the Central Limit Theorem), and students feel that teaching them to use computer facilities really improves students' ability to apply similar programs in analyzing real world problems.

## INTRODUCTION

Since 1980 computer facilities have been popularly applied to classroom teaching in order to improve the quality of teaching and learning, especially for those classes involving statistical concepts. At the very beginning of the computer era, many people had doubtes in the effectiveness of applying computer facilities in classroom teaching on campus. Previous study showed that more and more computer facilities and programs had been adapted in classroom teaching, especially for teaching statistics (Evans, 1973). A previous survey concluded that more than 86% of the schools have used computer facilities in teaching statistics in the M.B.A. programs (Rose, Machak and Spivey, 1998). Recently the Department of Education in Taiwan set a side special funding to support the universities to develop programs for computer applications in teaching every course including statistics. Even though computer applications become more and more popular in different schools in different countries, the effectiveness of the computer applications still puzzle most of the teachers. There is limited information on the evaluation of the effectiveness of the computer programs applied in classroom teaching. Teachers usually ask: How well do the students learn from applying computer programs in classroom teaching? Do computer programs really improve the quality of learning and teaching? How significant the impacts are on students who attend the courses? This article presented the results of the case study which examined the effectiveness of applying computer facilities to teach Business Statistics to several groups of undergraduate Management students (including Juniors and Seniors) sampled from Krannert School of Management at Purdue University. The following sections will describe the data of the case study, methodology applied to this study to analyze the data, results from the study, and finally the implications and concluding remarks.

## DISCUSSION OF THE DATA

The data of this study was collected from the teaching evaluation of the same instructor who taught Business Statistics every semester between 1994 and 1996. This instructor exposed three groups of the students to different frequencies of the computer applications: never use computer programs, moderately applied computer programs, or frequently applied computer programs. These three groups of the students were randomly assigned to each section taught by the same instructor when they registered for the Business Statistics course in the beginning of each semester. Then the instructor randomly chose sections and decided the frequency to apply computer programs in teaching each section. The instructor kept the same teaching style, the same homework assignments, the same instruction procedures, and the same examples demonstrated in the class. The only variable in teaching each section was “the frequency of the computer applications in classroom”.

By the end of each semester, students were asked to evaluate the instructor as well as the course. The complete teaching evaluation contained twenty questions, each question had five possible answers: strongly agree, agree, uncertain, disagree, and strongly disagree. Each answer had five different grades: strongly agree (SA = 5), agree (AG = 4), uncertain (UC = 3), disagree (DA = 2), and strongly disagree (SD = 1). We can rank the students’ evaluation towards the instructor or the course by students’ preference, namely SA > AG > UC > DA > SD. There were seven questions in the evaluation form which were directly or indirectly related to the usage of the computer facilities, and they were:

1. This instructor stimulated interest in the course,
2. Explanation of the material was clear and to the point,

3. This instructor used meaningful examples and applications,

4. Overall, this course was very useful to me and my career,

5. My instructor motivated me to do my best work,

6. My instructor explained difficult material clearly,

7. Overall, this course is among the best I have ever taken.

Totally 202 students took the evaluation. Among these 202 students, 36 students were from the class in which computer programs were never applied ($n_1$ = 36, group 1), 60 students were from the classes in which computer programs were moderately applied ($n_2$ = 60, group 2), and 106 students were from the classes in which computer programs were frequently applied ($n_3$ = 106, group 3). Table 1 to Table 3 summarized the number of the students who chose to answer "strongly agree", "agree", "uncertain", "disagree", or "strongly disagree" for each question. For example, 36 students were in group 1 (computer aids were never used), and 5 students answered "strongly agree", 14 students answered "agree", 10 students answered "uncertain", 5 students answered "disagree", and 2 students answered "strongly disagree" for <u>question 1</u>.

Table 1. Computer aids were never used.

| | SA | AG | UC | DA | SD |
|---|---|---|---|---|---|
| 1 | 5 | 14 | 10 | 5 | 2 |
| 2 | 8 | 16 | 5 | 7 | 0 |
| 3 | 10 | 15 | 8 | 2 | 1 |
| 4 | 3 | 12 | 13 | 5 | 3 |
| 5 | 9 | 11 | 8 | 6 | 2 |
| 6 | 6 | 17 | 8 | 3 | 2 |
| 7 | 2 | 9 | 14 | 8 | 3 |

$n_1 = 36$

Table 2. Computer aids were moderately used.

| | SA | AG | UC | DA | SD | |
|---|---|---|---|---|---|---|
| 1 | 16 | 33 | 9 | 2 | 0 | |
| 2 | 19 | 31 | 7 | 3 | 0 | |
| 3 | 23 | 29 | 8 | 0 | 0 | |
| 4 | 14 | 20 | 19 | 6 | 0 | $n_2 = 60$ |
| 5 | 17 | 25 | 11 | 6 | 0 | |
| 6 | 16 | 37 | 6 | 1 | 0 | |
| 7 | 5 | 19 | 29 | 7 | 0 | |

Table 3. Computer aids were frequently used.

| | SA | AG | UC | DA | SD | |
|---|---|---|---|---|---|---|
| 1 | 48 | 50 | 7 | 1 | 0 | |
| 2 | 52 | 45 | 7 | 2 | 0 | |
| 3 | 56 | 45 | 3 | 2 | 0 | |
| 4 | 24 | 42 | 31 | 6 | 3 | $n_3 = 106$ |
| 5 | 46 | 41 | 18 | 1 | 0 | |
| 6 | 54 | 41 | 10 | 1 | 0 | |
| 7 | 29 | 36 | 29 | 8 | 3 | |

## METHODOLOGY - THE MULTIVARIATE RANK SUM TEST

Classic Chi-Square test would not be appropriate for testing the variations of the students' answers in this study, due to different number of observations in each groups. A Multivariate Rank Sum Test had been developed to test the variability of the students' answers between three groups. Assume for each $i = 1,2,3, \{Y_{ij}, j = 1,\ldots, n_i\}$ are identically and independently distributed (i.i.d) random variables with $P\{Y_{ij} = k\} = p_{ik} > 0$, where

$$k = 1,\cdots,5$$

$$\sum_{k=1}^{5} p_{ik} = 1$$

$$i = 1,2,3$$

Consider the testing hypothesis with at least one inequality strict:

$$H_o: \tilde{p}_1 = \tilde{p}_2 = \tilde{p}_3 \quad \text{versus} \quad H_1: \tilde{p}_1 \le \tilde{p}_2 \le \tilde{p}_3$$
$$\tilde{p}_l = (\tilde{p}_{i1}, \tilde{p}_{i2}, \cdots, \tilde{p}_{i5})$$
$$\tilde{p}_l \le \tilde{p}_{l'} \Rightarrow \sum_{k' \ge k} p_{ik'} \le \sum_{k' \ge k} p_{i'k'} \qquad \forall k = 1,2,3,4,5$$

Set N = $n_1 + n_2 + n_3$ and replace each $Y_{ij}$ by $r_{ij}$, its rank in the overall sample, use average ranks for ties. Let $M_{ik}$ = # { $j$: $Y_{ij} = k$}, and

$$m_{\bullet k} = \sum_i m_{ik}$$
$$m_{i\bullet} = \sum_k m_{ik}$$
$$R_i = \sum_{j=1}^{n_i} r_{ij}$$
$$\overline{R}_i = R_i / n_i$$

Then

$$r_{ij} = \sum_{k' < k} m_{k'} + (1 + m_k)/2 \quad \text{if} \quad Y_{ij} = k$$
$$and$$
$$R_i = \sum_{k=1}^{5} \sum_{l<k} m_{\bullet l} + m_{ik}(1 + m_{\bullet k})/2$$

Under the null hypothesis the $Y_{ij}$'s are i.i.d. random variables.

On the other hand, if the alternative hypothesis is true, than $R_3$ tends to have a larger value and $R_1$ tends to have a small value. An easy decision rule for testing hypothesis is set by T = $R_2$ + $R_3$ and reject $H_o$ if and only if T is large. Since the sample sizes $n_1$, $n_2$, and $n_3$ are all large, a large sample approximation is appropriate. We will split the alternative hypothesis into three parts - $\tilde{p}_1 < \tilde{p}_2$ $\quad \tilde{p}_2 < \tilde{p}_3$ $\quad \tilde{p}_1 < \tilde{p}_3$ and compare the corresponding pair of rank averages.

If $H_o$ is true, then it is easy to see [cf. Kruskal 1952] that $ER_i = n_i(N+1)/2$ and

$E\overline{R_i} = (N+1)/2$, where

$$VarR_i = \frac{n_i(N-n_i)(N+1)}{12} - \frac{n_i(N-n_{i)}}{12N(N-1)}\gamma$$

$$and$$

$$Cov(R_i, R_{i'}) = -\frac{n_i n_{i'}(N+1)}{12} + \frac{n_i n_{i'}}{12N(N-1)}\gamma$$

$$with$$

$$\gamma = \sum_{k=1}^{5} m_{\bullet k}(m_{\bullet k}-1)(m_{\bullet k}+1)$$

Moreover, the random variables

$$\frac{R_i - n_i(N+1)/2}{\sqrt{(N^2 - N - \gamma)/12}}$$

are approximately multinormal with zero mean and covariance matrix whose i.j term is

$$\frac{\delta_{ij} n_i}{N} - \frac{n_i n_j}{N^2}$$

Random variables X and Y are multinormal with zero mean, unit variances, and correlation coefficient $\rho$ if and only if they have the joint density

$$\frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} e^{-(x^2 - 2\rho xy + y^2)/2\sigma^2(1-\rho^2)}$$

Let Z = X-Y then the joint density of Z and Y is given by

$$g_{ZY}(z,y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} e^{-((z+y)^2 - 2\rho(z+y)y + y^2)/2\sigma^2(1-\rho^2)}$$

$$= \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} e^{-(y-z/2)^2/\sigma^2(1+\rho^2)+(1-(1-\rho)/2)z^2/2\sigma^2(1-\rho^2)}$$

The p.d.f. of Z is

$$f(z) = \int g_{zy}(z,y)dy$$

$$= \frac{1}{2\pi\sigma\sqrt{1-\rho}} e^{-(2-(1-\rho))z^2/4\sigma^2(1-\rho^2)}$$

$$= \frac{1}{2\pi\sigma\sqrt{1-\rho}} e^{-z^2/4\sigma^2(1-\rho)}$$

Therefore Z is normally distributed with zero mean and variance 2 $\sigma^2$ (1 - $\rho$). For $1 \le i \le j \le 3$, set

$$Z'_{ij} = \frac{\overline{R}_i - (N+1)/2}{\sqrt{n_j(N-n_j)/N}} - \frac{\overline{R}_i - (N+1)/2}{\sqrt{n_i(N-n_i)/N}}$$

$$and$$

$$Z_{ij} = \frac{Z'_{ij}}{\sqrt{2(1+n_i n_j / N^2)}}$$

Then if $H_o$ is true, by the above discussion each $Z_{ij}$ is approximately normally distributed with zero mean and unit variance. So we reject $H_o$ if $Z_{13} > z_\alpha$ for the testing hypothesis (1) with level $\alpha$. However, instead of testing the hypothesis (1), we test the alternative hypothesis

$\tilde{p}_1 < \tilde{p}_2$  $\tilde{p}_2 < \tilde{p}_3$  $\tilde{p}_1 < \tilde{p}_3$ separately. Alternatively we compare the p-value

$p_{ij} = P_o\{Z > Z_{ij}\}$ with α and conclude $\tilde{p}_i < \tilde{p}_j$ if $P_{ij} < \alpha$.

**THE TESTING RESULTS AND CONCLUSIONS**

With the significance level α = 0.01, the testing results, p-values, and the decisions are summarized in Table 4 to Table 10.

Table 4. Instructor stimulates interest in the course.

| Computer use | SD | DA | UC | AG | SA |
|---|---|---|---|---|---|
| Never | 2 | 5 | 10 | 14 | 5 |
| Moderate | 0 | 2 | 9 | 33 | 16 |
| Frequent | 0 | 1 | 7 | 50 | 48 |

Testing results and conclusions.

| Testing statistics | Value | p-value | Reject Ho |
|---|---|---|---|
| $Z_{12}$ | 2.210267 | .0136 | No |
| $Z_{23}$ | 3.011938 | .00115 | Yes |
| $Z_{13}$ | 5.311483 | .00000 | Yes |

Table 5. Explanation of the material was clear and to the point

| Computer use | SD | DA | UC | AG | SA |
|---|---|---|---|---|---|
| Never | 0 | 7 | 5 | 16 | 8 |
| Moderate | 0 | 3 | 7 | 31 | 19 |
| Frequent | 0 | 2 | 7 | 45 | 52 |

Testing results and conclusions.

| Testing statistics | Value | p-value | Reject Ho |
|---|---|---|---|
| $Z_{12}$ | 1.286386 | .09965 | No |
| $Z_{23}$ | 2.618263 | .00442 | Yes |
| $Z_{13}$ | 3.986411 | .00005 | Yes |

Table 6. This instructor used meaningful examples and applications.

| Computer use | SD | DA | UC | AG | SA |
|---|---|---|---|---|---|
| Never | 1 | 2 | 8 | 15 | 10 |
| Moderate | 0 | 0 | 8 | 29 | 23 |
| Frequent | 0 | 2 | 3 | 45 | 56 |

Testing results and conclusions.

| Testing statistics | Value | p-value | Reject Ho |
|---|---|---|---|
| $Z_{12}$ | 1.353227 | 0.08743 | No |
| $Z_{23}$ | 2.333013 | 0.0098 | Yes |
| $Z_{13}$ | 3.745691 | 0.0001 | Yes |

Table 7. Overall, this course was very useful to me and my career.

| Computer use | SD | DA | UC | AG | SA |
|---|---|---|---|---|---|
| Never | 3 | 5 | 13 | 12 | 3 |
| Moderate | 0 | 6 | 20 | 20 | 14 |
| Frequent | 3 | 6 | 31 | 42 | 24 |

Testing results and conclusions.

| Testing statistics | Value | p-value | Reject Ho |
|---|---|---|---|
| $Z_{12}$ | 1.906184 | .0283 | No |
| $Z_{23}$ | 0.686453 | .2473 | No |
| $Z_{13}$ | 2.512481 | .0059 | Yes |

Table 8. My instructor motivates me to do my best work.

| Computer use | SD | DA | UC | AG | SA |
|---|---|---|---|---|---|
| Never | 2 | 6 | 8 | 11 | 9 |
| Moderate | 0 | 6 | 12 | 25 | 17 |
| Frequent | 0 | 1 | 18 | 41 | 46 |

Testing results and conclusions.

| Testing statistics | Value | p-value | Reject Ho |
|---|---|---|---|
| $Z_{12}$ | 0.824744 | .2047 | No |
| $Z_{23}$ | 2.563889 | .00517 | Yes |
| $Z_{13}$ | 3.496861 | .0002 | Yes |

Table 9. My instructor explains difficult material clearly.

| Computer use | SD | DA | UC | AG | SA |
|---|---|---|---|---|---|
| Never | 2 | 3 | 8 | 17 | 6 |
| Moderate | 0 | 1 | 6 | 37 | 16 |
| Frequent | 0 | 1 | 10 | 41 | 54 |

Testing results and conclusions.

| Testing statistics | Value | p-value | Reject Ho |
|---|---|---|---|
| $Z_{12}$ | 1.712704 | .0434 | No |
| $Z_{23}$ | 2.892640 | .0019 | Yes |
| $Z_{13}$ | 4.676796 | .0000 | Yes |

Table 10. Overall, this course is among the best I have ever taken.

| Computer use | SD | DA | UC | AG | SA |
|---|---|---|---|---|---|
| Never | 3 | 8 | 14 | 9 | 2 |
| Moderate | 0 | 7 | 29 | 19 | 5 |
| Frequent | 3 | 8 | 30 | 36 | 29 |

Testing results and conclusions.

| Testing statistics | Value | p-value | Reject Ho |
|---|---|---|---|
| $Z_{12}$ | 1.056293 | .1457 | No |
| $Z_{23}$ | 3.021289 | .0013 | Yes |
| $Z_{13}$ | 4.199651 | .0000 | Yes |

After comparing the answers for all seven questions, there was no significant variability between the students in group 1 and students in group 2. However students revealed significant variability between group 2 and group 3, as well as between group 1 and group 3. From the above tables, all p-values of $Z_{12}$ are greater than 0.01. This means that there is no significant improvement in learning or responses for students exposed to moderate usage of the computer programs, comparing to the students who had never been exposed to computer programs.

The results were quite different when comparing students in group 2 (moderately use computer programs ) to students in group 3 (frequently use computer programs), and also when comparing students in group1 (never use computer programs) to students in group 3. To compare the variability in responses between group 2 and group 3, all p-values expect the one for the 4$^{th}$ question (overall this course was very helpful to me and my career) is smaller than 0.01. This means that students who have been frequently exposed to computer programs tend to be more interested in the course contents comparing to students who have been exposed to moderate usage of the computer programs. Students exposed to frequent computer usage also fell that they understand the materials better, and they agree that they have been motivated to do their best work. Generally speaking students exposed to frequent computer usage perform better in the course, comparing to students exposed to moderate computer usage.

Comparing students in the class involving frequent computer usage to students in the class never introducing computer programs, all p-values for $Z_{13}$ are small than 0.01. This provides a very strong support to the conclusion: frequent usage of computer programs really improves the teaching and learning quality in Business Statistics courses.

## SUMMARY AND IMPLICATIONS

Over years teachers have been debating about the effectiveness of the computer applications in classroom teaching. For some courses as Statistics, computer programs could be very helpful to explain difficult contents such as regression analysis, Central Limit Theorem, and probability distributions. Students will be benefit from the demonstration of the computer programs in classroom, and they will be more attracted and more interested in learning Statistics. They will also be used to the applications of the computer programs, so that they will be able to apply similar computer programs in analyzing real world problems. However to apply computer programs frequently in the classroom teaching may or may not help students get good grades. Whether students really learn more or perform better through out the semester, this can not be answered easily from this study. Further research need to be focused on evaluating the relationship between “students’ overall performance” and “the frequency of the computer applications in classroom teaching”.

**REFERENCES**

[1] Evans, D.A. (1973). The influence of computers on the teaching of statistics. *J.R. Statistics Soc. A* **136**(2), 153-190.

[2] Hollander, M. and Wolfe, D.A. (1973). *Nonparametric statistical methods.* John Wily & Sons, New York.

[3} Kruskal, W.H. (1952). A nonparametric test for several sample problems. *Ann. Math. Statistics* **23**, 525-540.

[4] Mead, R. and Stern, R.D. (1973). The use of a computer in the teaching of statistics. *J.R. Statistics Soc. A.* **136**(2), 153-190.

[5] Rose, E.L., Machak, J.A. and Spivey, W.D. (1988). A survey of the teaching of statistics in M.B.A. programs. *Journal of Business and Economics Statistics* **6**(2), 273-282.