A Study of Factors Contributing to the Late Submission of Course Work

Jude .T Lubega [sir02jtl@reading.ac.uk]
Shirley Williams [shirley.williams@reading.ac.uk]
Department of Computer Science, University of Reading, P.O. Box 225,
Whiteknights, Reading, Berkshire, RG6 6AY, United Kingdom
[http://www.rdg.ac.uk]

Abstract

This case study uses log-linear modelling to investigate the interrelationships between factors that may contribute to the late submission of coursework by undergraduate students. A class of 86 computing students are considered. These students were exposed to traditional teaching methods supported by e-learning via a Managed Learning Environment (MLE). The MLE warehouses detailed data about student usage of the various areas of the environment, which can be used to interpret the approach taken to learning. The study investigates the interrelationship between these factors with the information as to whether the student handed in their course work on time or whether they were late. The results from the log-linear modelling technique show that there is an interaction between participating in Discussions within the MLE and the timely submission of course work, indicating that participants are more likely to hand in on time, than those students who do not participate.

Key words

MLE, Log-linear modelling, Data mining, Learning Process.

Topics

Introduction
Learning Context
  The Module
  Characteristics of Learning Programming
  Teaching Methods
Modelling Techniques
Data Collection and Analysis
  Data Collection
  Data Analysis
  The K-Factor Interactions
  Test for Partial and Marginal Associations
  Combinations
Results and Discussion
Conclusions
References

Introduction

During the learning process students commonly need to be assessed. To ensure a timely handing in of work, deadlines are normally set by which the work has to be completed. There are a variety of approaches that are taken to the penalties that are applied to students that are late in handing in work. Within our institution the following rule is applied:

> "For major coursework 10% will be deducted from your mark if your work is submitted up to a week late (that is 10% of the maximum possible mark). A mark of zero will be awarded if your work is submitted more than a week late.

For *minor* coursework a mark of zero will be awarded if your work is submitted late."

Where major and minor items are defined as:

> "A major item of coursework is one that carries more than 10% of the marks available for a module. Anything else is a minor item of coursework."

There are procedures which allow students with extenuating circumstances to be granted appropriate extensions.

A study has been undertaken of students following courses that take blended learning approaches (Lubega and Williams, 2003); this is a continuation of that work looking at the cohort of students who took the course in the academic year 2002-3. In particular the characteristics of students who subsequently hand work in late are investigated.

Learning Context

During the learning process, students use several activities as they try to acquire new skills. Activities that contribute to the learning process include creating learning plan, studying content, participating in discussions, assessment and evaluations. Assessment measures the learning performance of the learners, thus indicating the impact of the activities on the learning process. The interaction of the activities during the learning process is considered important for learning. However, the interactions between the learning activities that can really create a positive impact on learning are unknown.

Using the blended learning method, students are offered several learning activities which they can use. The activities that students get involved in influence their learning habits. When students are set a piece of work, they may use the available activities to facilitate completion. Students are able to participate in activities like collaboration either online/offline where they can communicate with each other using the available means. During the collaboration, cooperation amongst collaborative group members is emphasised because it helps them retain the information longer than the students who learn individually

(Johnson and Johnson 1986). This may help the students in completing their assessment much earlier than they would when working alone. In an online learning environment interaction during the activities is necessary as it is seen to have a significant positive effect on the individual learning over time (Meliha and Denise, 2002). Other research (Nydia, 2002) has shown that discussions during online learning demonstrated a progression in the knowledge to application. The rate at which the students use these learning activities for learning differs depending on factors like computer experience and education qualification (Afzaal *et. al.*, 2002).

Determining what activities are interrelated during the learning process is the subject of this investigation. Several activities which were available to the students have been considered to find out which could have contributed to late handing of a piece of work. The students used the communication including discussions and chartrooms and the instructional content within the managed learning environment during their learning process. They also used other means to learn, such as the library and attending lectures. The tutors played a great role in streamlining the learning process in both online and offline learning environments. Tutors encouraged collaboration amongst the students as they attempted the work. When the work was due, a period of grace was offered for those who could not hand in on time. However the students who failed to hand in on time knew they would be penalised. An investigation was carried out to find out what activities that the students were involved could have had an impact on the hand in of the assessment. Interaction between two activities produces an impact within the learning process. This indicates the interrelationship that co-exists between the learning activities. Therefore the aim of the investigation was to find which activities used by the students had an inter relationship with their submission of the coursework.

The data mining technique that is applied in the investigation is called Log-linear analysis. This technique is further explained in the Modelling Technique section below.

The Module

The Module studied is a one term (10 week) course called Programming 2. It is a continuation of a first term module Programming 1. These modules are intended for students with a prior experience of computing, for example at a standard equivalent to the English GCE A-Level or BTEC. Topics covered include

- the representation of data, including use of arrays, records, arrays of records, sets, components, basic objects and basic pointers;
- program constructs such as conditionals, loops, functions and procedures and their use;
- design, implementation, testing and debugging of a program to solve a problem;
- use of components from standard, additional and dialogue tabs of Delphi IDE;
- exception handling;
- advanced data structures (static and dynamic);
- object-oriented and modular programming;
- cross-platform development using Kylix and CLX;

The language Object Pascal, in the Delphi Rapid Application Development environment (Williams and Walmsley 1999), is used for practical work.

Characteristics of Learning Programming

Programming is traditionally taught using an approach similar to the 3Ps approach of foreign language learning (Gabrielatos, 1994, Johns et al, 2003). The 3Ps stand for:

- Presentation. Where the teacher gives information. This is often in the form of a lecture.
- Practice. Where the students work with the information they have acquired. This might be by doing simple exercises.
- Production. Where the students consolidate what they have learned, usually doing individual work. For example: producing a program that solves a particular problem.

The learning is supported by student reflection and interaction with the teacher.

Teaching Methods

The module consists of around 30 contact hours over 10 weeks. In a typical term time week there are either two lectures per module, or a lecture and an exercise class, or similar. Each student has a weekly hour-long supervised practical per module, and access to a suitable laboratory at other times. These activities provide the student with the first of the 2Ps (Presentation and Practice).

The on-line Blackboard system (Blackboard, 1994-2004) is used to support student learning and provides the students with a blended learning experience (Lubega and Williams 2003). Material routinely used includes: lecture notes, and assignments; discussion groups; quizzes; general information. This primarily supports the first 2Ps.

In the preceding modules students are given weekly practical exercises to complete, to help consolidate learning (the final P). In this module the students were set a major piece of coursework to be completed by 10am on Monday 10th March 2003.

Modelling Techniques

Data mining (Berry, 1997) is the exploration and analysis of large quantities of data in order to discover meaningful patterns and rules. It is done by the use of a combination of machine learning, statistical analysis, modelling techniques and database technology to find patterns, relationships and future predictions. There are several ways used to extract knowledge within this process and one of them is Log-linear analysis modelling that helps to tell what interrelationships exist among variables within contingency/frequency tables.

Little et al (Little, 2002) applied log-linear analysis in the process of data mining to find the collusion in the US crop insurance program. They were able to find the node-node arrangements and information on node sharing which displayed the potential of this technique to show the interrelationship between available variables under investigation.

Several researchers have used this type of modelling on data from different disciplines that include environment, ecology, agriculture and other social disciplines. Those who have looked at Environment and ecological studies include: Lakhan and Lavalle (2002), Melville and Swain (1997) and Crook (1997)

among the many. Further applicable areas that have seen this technique of log-linear modelling being used include fisheries (Crook, 1999), applied psychology (Viinamakiet al, 1999) and social work (Western, 1996).

Log-linear analysis (McCutcheon and Mills, 1998) is an extension of the chi-square technique. However this analysis technique differs from the chi-square technique because it is used to test for more than two qualitative variables.

Chi-square analysis (Neter *et al.*, 1990) is used to test the hypothesis ("independence" and "homogeneity of proportions"). It is also known as the *goodness-of-fit statistic* or *Pearson's goodness-of-fit statistic*. The test is known as the *chi-square test* or the *goodness-of-fit test*. Simple chi-square can be described by the following:

Let the observed cell counts be denoted by $\{x_{ij}: i=1,...,r; j=1,...,c\}$ and the expected cell counts under a model of independence or homogeneity of proportions be denoted by $\{e_{ij}: i=1,...,r; j=1,...,c\}$. The test statistic is

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} (x_{ij} - e_{ij})^2/e_{ij} = \sum_{allcells} (\text{Observed} - \text{expected})^2/\text{expected}$$

For the chi-square test statistics the p-value is used for accepting or rejecting the hypothesis. When the p-value is less than the level of significance then the chi-square is significant. Hence reject the null hypothesis and accept the alternative hypothesis.

Log-linear analysis is a technique that focuses on the association of grouped data looking at all levels of possible interaction effects. Therefore it is used to find out what variables are related to each other during interaction. Log-linear analysis may be applied when there are more than two nominal independent variables that are to be tested for inter-relationships. The main goal of this analysis technique is to reduce the number of significant effects it takes to explain the data. Hence the test does not finish until it reaches non significance for the level of effect being tested. Log-linear is a goodness-of-fit test that allows testing all the effects (the main effects, the association effects and the interaction effects) at the same time. Unlike the analysis of the variance (ANOVA) (Berk and Carey, 2000), log-linear analysis is most useful when considering in interactions among variables rather than in the main effects. Because this analysis allows testing for independence and interactions among the variables, it may be used for testing for the independence of learning activities during a learning process. It can also be used to test for the significant learning activity interactions within a learning process.

The models that can be produced from the log-linear analysis can be of two types, direct models where the expected value can be calculated directly from the marginal sums or the indirect model where the expected value must be estimated through interaction. The log-linear model can be represented in the following way using variables A, B and C.

$$\text{Ln}(F_{ijk}) = m + l_i^A + l_j^B + l_k^C + l_{ij}^{AB} + l_{ik}^{AC} + l_{jk}^{BC} + l_{ijk}^{ABC}$$

Where:

$\text{Ln}(F_{ijk})$ = is the log of the expected cell frequency of the cases for cell ijk in the Contingency table.

μ= is the overall mean of the natural log of the expected frequencies,

λ= represent the effects which the variables have on the cell frequencies,

A, B and C are the variables under investigation,

i, j and k refer to the categories within the variables

λ_i^A = the main effect for variable A

λ_j^B = the main effect for variable B

λ_k^C = the main effect for variable C

λ_{ij}^{AB} = the interaction effect for variable A and B

λ_{ik}^{AC} = the interaction effect for variable A and C

λ_{jk}^{BC} = the interaction effect for variable B and C

λ_{ijk}^{ABC} = the interaction effect for variables A B and C

The above illustration of the log-linear model is of the 2x2x2 contingency table. The log-linear models become more complex as the number of variables increase. This makes the model interpretation very difficult.

Data Collection and Analysis

The Blackboard system stores information about all users, their usage and their activities. From a course statistics area instructors can access their warehoused data and generate reports.

Data Collection

At the end of the module all usage data was extracted from the Blackboard system, and combined with the information of which students had handed in their work late. The data was cleansed (Pyle, 1999), this included making data anonymous (including removal of the names of the students). A total of 86 students' anonymous data was set for mining to find out which learning activities inter related.

During the learning process, some students actively participate and some are dormant. The active students participate in carrying out particular learning activities such as: contributing to the discussion while some visit the discussion area to see what was discussed. The dormant students do not participate in learning activities that involve other learners but prefer to learn individually. The data during the cleansing exercise was categorised according to how the students participated in the online learning process.

The data was then summarised and a frequency table created of four variables:

1. Groups. Based on the relative usage of two areas of Blackboard: the Communication and the Contents. Three groups were formed based on the total hits used during learning. Group One (0-400), Group Two (400-800) and Group Three (800 above).
2. Discussion Status. This indicated whether a student participated in the online discussion (Discussed) or did not participate in the discussion (Not Discussed).
3. Hand in status: This represented the status of handing in the course work. These included, Hand In on time and Hand In Late.
4. Rate of Use of the Managed Learning system. This category contained the Low usage (below average hits in that group) and High usage (above average hits in that group) within each group.

From this frequency table the variable (Hand In Status) is regarded as the major response variable of interest. The other three variables Groups, Discussion Status and Rate of Use are regarded as design variables. The frequencies were in a 2x2x2x3 contingency table as in the sample shown below.

Group one (0-400) sample data is shown in Table 1 and is summarised:

- 8 Students who had low hits, handed in on time and discussed
- 14 Students who had low hits, handed in on time and did not discuss
- 0 Students who had low hits, did not hand in on time and discussed
- 7 Students who had low hits, did not hand in on time and did not discuss
- 8 Students who had high hits, handed in on time and discussed
- 6 Students who had high hits, handed in on time and did not discuss
- 2 Students who had high hits, did not hand in on time and discussed
- 7 Students who had high hits, did not hand in on time and did not discussed

Table 1. Frequency Layout

	Discussed	**Not Discussed**
Grp One Low Yes	8	14
Grp One Low No	0	7
Grp One High Yes	8	6
Grp One High No	2	7

The interaction between the variables is seen as the combination of the two or more notations to form a combination, e.g. Discussion Status (2) and Hand In Status (3) making 23, Discussion Status (2) and Rate of Use (4) as 24, Discussion Status (2), Groups (1) and Hand In Status (3) as 213 which meaningfully denotes that the two or more variables are associated between each other.

Data Analysis

The aim of the log-linear analysis technique is to find the best simple model that relates variables to the variable of interest (Hand In Status). It is not essential that all variables are included in the model, only those that best fit the data should be included. The process is computationally intensive when there are several variables and combinations of variables under investigation. Several models are usually derived and then the one that best fits the data is selected. The statistical package Statistica for Windows by StatSoft Inc (1984-2004) was used in this work to derive the K factors log-linear models from which the best model can be selected.

Log-linear modelling was used within the study to identify what interaction and inter-relationships existed within the variables that are being investigated. The main role of the log-linear analysis technique is to derive a model that contains different combination effects that will describe and fit the data in the frequency table. From the model an inter- relationship amongst the variables under investigation can be reflected basing on the combinations made.

The K-Factor Interactions

The K factor analysis is a method that is used within log-linear analysis to classify variables by reducing them in number through detecting the relationship between them. The K factor analysis is used for reducing a big group of variables by structure detection, first referred to by Thurstone (1931). The interaction that occurs when the variables are seen to have a relationship depends on the number of variables. This is referred to as K-factor interaction and the resulting effect may be termed as k-factor effect. In log-linear analysis it is the interaction that is used within the analysis.

Having used the K factor (number of interactions in the frequency table) for the log-linear analysis, it is possible to derive several log-linear models but not all are essential. The best simplest model that can best describe data is the one that is needed. The interpretation of the tests when a K-factor is either increased or decreased will give a clear judgement on the model improvement. Table 2 describes results from the K-factor interactions for the variables under investigation where Discussion Status (1), Hand In Status (2), Rate of Use (3) and Group (4).

Table 2. Describing the K-Factor Interactions

	Results of Fitting all K-Factor Interactions (Spreadsheet32) These are simultaneous tests that all K-Factor Interactions are simultaneously Zero.				
K-Factor	Degrs.of Freedom	Max.Lik. Chi-squ.	Probab. p	Pearson Chi-squ	Probab. p
1	5	54.55796	0.000000	57.69432	0.000000
2	9	21.96159	0.009002	18.97551	0.025402
3	7	4.62607	0.705488	4.42464	0.729772
4	2	0.66619	0.716703	0.66066	0.718688

From the table 2 it is noted that the figures are related to chi-square, significance and it is these figures that will be considered in order to select what models are of interest. The statistical significance of the goodness-

of-fit of a particular model can also be evaluated via a Chi-square test.

There are two types of Chi-squares that can be computed: the Pearson Chi-square and the maximum likelihood ratio Chi-square statistic. The term likelihood ratio was first introduced by Neyman and Pearson (1931); the term maximum likelihood was first used by Fisher (1922). In practice, the interpretation and magnitude of those two Chi-square statistics are essentially identical. Both tests evaluate whether the expected cell frequencies under the respective model are significantly different from the observed cell frequencies. If so, the respective model for the table is rejected.

Table 2 shows the simultaneous test where all k-factor interactions are equal to Zero. The interpretation for this table is that there is an improvement in fit as the factors are increased during the process of interaction. When the model is highly significant, then it provides a very poor fit. Having K factor =1, p=0.00000 (highly significant hence poor fit), if the factor is increased to K factor = 2, p=0.009002 hence indicating that there was an improvement in the model derivation. Although both still provide a poor fit K factor = 2 provides a better fit than K factor = 1. When factors are increased to 3 way interactions in the same model K factor = 3, it is noted that the significance is lost with p=0.7055. This indicates that the model provides an adequate fit to the data set. Therefore a conclusion is made that, the least complex model that will fit the table need not contain any 3 way associations but instead will contain one or two way associations. So from this it is noted that the models to be used can have effects as pairs or singly for example: Discussion Status, Hand In Status, Rate of Use as stand alones or Discussion Status + Hand In Status as combined learning activities to produce an effect. However a 3 way combination such as: Discussion Status + Rate of Use+ Hand In can not be produced in the model.

Test for Partial (Prt) and Marginal (Mrg) Association

Table 3 describes the association between the four variables under investigation from which several models have been derived to fit the data.

Table 3. Association between the four variables under investigation (1, 2, 3 and 4)

	Tests of Marginal and Partial Association (Spreadsheet32)				
Effect	Degrs.of Freedom	Prt.Ass. Chi-sqr.	Prt.Ass. p	Mrg.Ass. Chi-sqr.	Mrg.Ass. p
1	1	0.16331	0.686130	0.16331	0.686130
2	1	18.59601	0.000016	18.59601	0.000016
3	1	2.62397	0.105261	2.62397	0.105261
4	2	33.17470	0.000000	33.17470	0.000000
12	1	4.54732	0.032970	4.00310	0.045417
13	1	7.45293	0.006333	4.81966	0.028137
14	2	8.91031	0.011619	8.42509	0.014809
23	1	2.36217	0.124309	1.06570	0.301919
24	2	0.24477	0.884808	1.09691	0.577842
34	2	1.99996	0.367886	0.76208	0.683150
123	1	-0.00376	1.000000	0.33919	0.560296
124	2	1.57586	0.454786	1.90470	0.385834
134	2	1.82506	0.401506	2.64050	0.267069
234	2	0.35800	0.836105	0.29504	0.862847

Based on the method of backward elimination, the partial and marginal associations the best model can be chosen from the 14 models derived.

When table 3 is considered, it contains tests for all marginal (Mrg. Ass.) and partial (Prt. Ass.) associations and will show which two way associations are significant to consider conclusions. The partial and marginal association tests using the concept of backward elimination (Dillon and Goldstein, 1984) ensures the best model selection from model derivations.

In Table 3, the partial association chi-square evaluates the significance of the different effects. This is used on making decisions to either include or exclude an interaction from the model. For example if we consider an interaction between variables 3 and 4 to form 34, there is a two-way interaction that was made. Checking its partial association chi-square, it is 1.9996 with 2 degrees of freedom, and its p = 0.3679<0.05. Therefore excluding this two way interaction from the model would have a detrimental impact on the model fit.

The marginal association chi-square indicates the difference between the models without any two way interactions and the one with the two interactions. Considering models from factors 2 and 4 respectively and the model from the combination of factors 2 and 4, it is noted that the model improves when combined to 0.8848 as compared to the individual factors model of 0.00002 and 0.00000 respectively.

Reviewing the partial and marginal association chi-square, it can be seen that there are other associations that can be included in the model.
These are summarised into 3 columns table 4:

1. Models with both partial and marginal tests significant,
2. Models with both partial and marginal tests not significant,
3. Models with one test significant and the other not significant,

The models of interest are those that have both partial and marginal significant.

Table 4. Summary of the significant tests

Models with both partial and marginal tests significant	Models with both partial and marginal tests not significant	Models with one test significant and the other not significant
Include in analysis	Exclude in analysis	
2, 4, 12, 13, 14	1, 3, 23, 24, 34, 123, 124, 134, 234	None

In order to test for a particular factor, that factor is omitted during the fitting so that main effect can be

realised. If there is no change then that means there was no effect realised by omitting that particular factor. For example if testing the main effect of Hand In Status (2), the log-linear model 134 is fitted to the frequency table and its significance level is noted. The significance of the main effect (2) is then determined by finding the difference between the X^2 values and the degrees of freedom of the model 1234 and 134.

Combinations

Having used the 95% confidence interval and taking our level of significance as 0.05, it is noted from table 4 that only 2, and 4 factors out of the four are significant. However it is found that some other factors when combined, also offer a significant value and these include 12 and 13 and 14. Therefore it is now necessary to derive the best and simplest model that fits the frequency table by considering the significant factors. Table 5 describes the factor effect within the variables. Some factors are tested by eliminating others from the effects and then seeing what changes occur in the values for the significance. Chi-square tests which are not significant are always sufficient to explain the frequencies in the table.

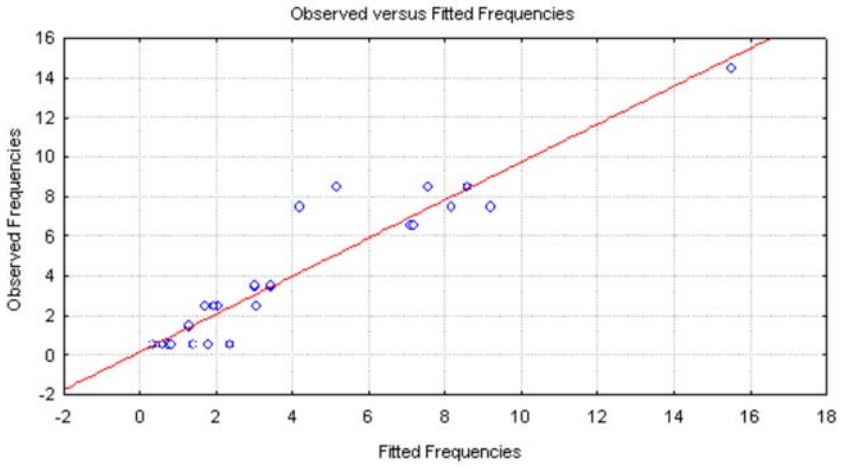
Table 5. Description of the factor effect

Test	Chi-sqr	Degrees of freedom	Mrg Ass. (p)
24, 34	25.5582	15	0.04293
34	45.2511	18	0.00038
1, 2, 34	26.4917	16	0.04789
1, 34	45.0878	17	0.00024
31,24,34	20.5752	13	0.08177
1, 2, 42	28.7809	17	0.03660
1,24,34	25.3949	14	0.03086

Comparing Chi-square test of some of the models derived in the table 5, like (34) and (24, 34) it is noted that chi-square difference is 45.2511-25.5582= 19.6929 and the degrees of freedom 18-15=3, the resulting significance being $p<0.05$ which is not sufficient to explain the frequencies in the table. Considering (31, 24, 34) and (24, 34) it is noted that they have a resulting significance $p>0.05$; these combinations are sufficient to explain the frequencies within the table.

Using the Software STATISTICA the best fit was tested for. From the procedure of backward elimination described by Dillon and Goldstein (1984), the software automatically fitted the best model as 21, 31, 41 having Chi-square as 10.0060 with 14 degrees of freedom and a p value of 0.761744>0.05. When a plot of observed versus fitted frequencies (Figure 1) is made to find out if there exists any outliers in that table, it is found out that most of the points are almost plotted on to the straight line. A few points seem to be some distance from the straight line indicating a few outliers within data.

Figure 1. Plot of observed versus fitted frequencies



Results and Discussion

The log-linear analysis results shown here are from an assessment of the factors that influence submission of work by the students. Based on the model (21, 31, 41) that best fits the frequency table, it is clear that the major factor influencing Hand In Status is the Discussion Status. This is due to the fact that in the best model fit (21, 31, 41) it is 21 that contains 2 (Hand In Status) which associates with 1 (Discussion Status). Having 2 as the factor of interest, it is concluded that the factor it associates with (1) causes the effect. The rest of the interactions are not considered because they have no interactions with the factor of interest. Therefore Rate of Use and Groups variables do not have a strong influence during the interaction with the Hand In Status of the course work. The students' Rate of Use that leads to high hits hence a particular group does not imply efficiency so as to complete the course work on time. This is further indicated within the results where students of group one with a general low rate usage and managed to discuss, handed in on time. The total hits leading to the group where students belong, does not necessarily imply that all hits are utilised efficiently for work.

The discussion carried out during the learning process could have contributed to the hand in on time. This can be further indicated by the percentage of those students who discussed and managed to hand in on time as being the highest compared to other student groups. Figure 2 describes the relation ship within

each Group amongst the students who discussed and those who did not discus. It is noted from the graph that majority of the students who discussed managed to hand in on time of the course work. At class level, a Figure 3 is plotted to show the percentage of students who discussed and handed in on time, those students who discussed and handed in late, those students who did not discuss and handed in on time and those students who did not discuss and handed in late.

Figure 2. A plot of Relation within the different groups in the class

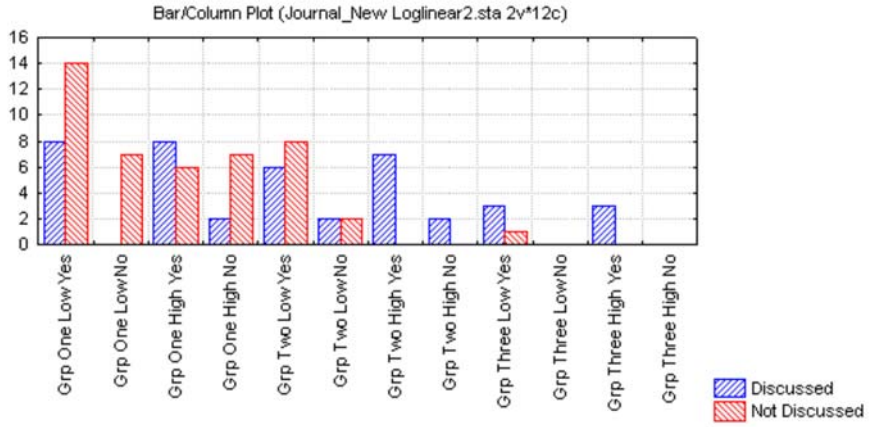
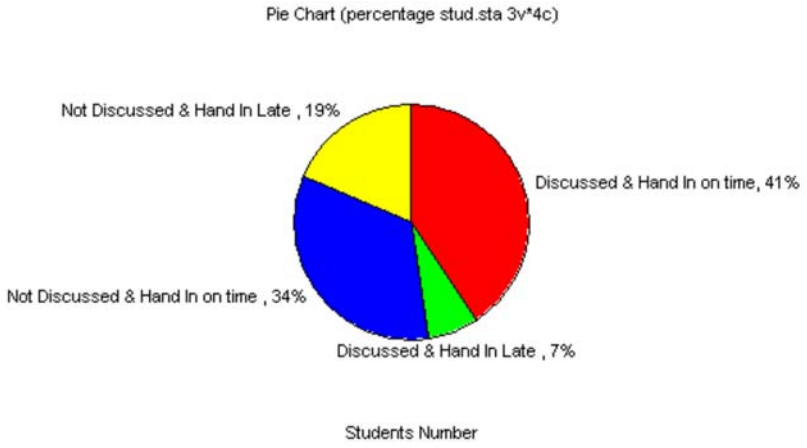


Figure 3. Percentage of the discussed and non discussed students



The two figure (2 and 3) illustrate the distribution of the 86 students and how they related to particular variables used within the study. It is probable that discussing contributed to the hand in of the course as majority of the students who discussed handed in on time. The discussion that is carried out during the learning process helped students to share their understanding with others hence finding solutions to the assessment. From what is discussed, the students were able to formulate solutions to the assessment. Therefore the contents from the discussion helped students in finding solutions to the assessment within the allotted time. The students who did not discuss were more likely to make late submission. Some students who did not discuss but handed in on time are the ones who prefer to learn individually hence finishing on time. The tutors found the discussion board a better place for discussions especially for those students who felt shy in class. The students' contributions to the discussion were related to the assessment topic despite being free to discuss other issues. Staying in the education context during the discussion showed a sense of educational maturity. Something else to note from the results is that Discussion Status was also related to both Group and Rate of Use. This further explains the fact that those students, who had a high usage rate within each group, are ones who were involved in the discussions.

Conclusion

Log-linear modelling technique was used to identify what factor most influences the Hand In Status of a course work by students. This is a very appropriate method for identifying the most significant factors affecting the learning process of students. Therefore this should highlight to the educators that students who participate in discussions are very likely to hand in on time. Those who do not participate in discussions are more likely to be late hence educators should encourage all students to participate in the discussions. Students need to be actively involved in collaboration either by participating in the discussions or viewing what has been discussed. The theory of social constructivism states that knowledge is the result of social interaction and language usage hence is shared rather than an individual experience (Prawatt & Floden, 1994). Therefore when students interact through discussion, they share the knowledge hence coming to an agreed understanding. This may contribute to their learning process because they learn from other people's experiences. However some other social factors could have contributed to the Hand In Status, like sickness, laziness and so on. Discussions among the many learning activities can influence submission of work. However more research of this nature needs to be carried out to further explain this phenomenon of learning activity interactions.

Having carried out a case study that contained only 86 students, larger classes across other disciplines are being considered for future study. This will further highlight to the educators in general what learning activities are more related to each other. Knowing the advantages attributed to particular learning activities, encourages students to use them hence better learning output.

This case study has shown how log-linear analysis is a very good data mining technique for identifying

interacting social factors. This technique can be applied to other scenarios where the aim is to find the relationship that exists between interacting variables under investigation.

References

[1] Afzaal, H. S., Md, Mahbubur. R., & Mohd, N. R. (2002). A study of computer attitudes of non-computing students of technical colleges in Brunei Darussalam. Journal of End User Computing, 14, (2).

[2] Berk, K.N and Carey, P.M. (2000). DATA ANALYSIS with Microsoft EXCEL, Duxbury Press, United Kingdom.

[3] Berry, M.J.A and Linoff, G. (1997). Data Mining Techniques, John Wiley, USA.

[4] Bishop, Y. M., Fienberg, S. E and Holland, P. W. (1975). Discrete Multivariate Analysis: Theory and Practice, MIT Press, Cambridge, Massachusetts.

[5] Blackboard (1994-2004), retrieved October 25, 2004 from http://www.blackboard.com/

[6] Crook, A. C. (1997). Color patterns in a coral reef fish. Is background complexity important? Journal of Experimental Marine Biology and Ecology, 217 (2), 237–252.

[7] Crook, A. C. (1999). Quantitative evidence for assortative schooling in a coral reef fish, Marine Ecology Progress Series, 176, 7–23, retrieved April 13, 2004 from http://www.int-res.com/abstracts/meps/v176 /p17-23.html

[8] Dillon, W. R and Goldstein, M. (1984). Multivariate analysis: Methods and applications, John Wiley, New York.

[9] Fisher, R. A. (1922). On the interpretation of *Chi-square* from contingency tables, and the calculation of p, Journal of the Royal Statistical Society, 85, 87-94.

[10] Gabrielatos, C. (1994). Minding our Ps: A framework for grammar teaching, Current Issues 3, retrieved April 13, 2004 from http://www.gabrielatos.com/MindingOurPs.htm

[11] Johnson, R. T., & Johnson, D. W. (1986). Action research: Cooperative learning in the science classroom. Science and Children, 24, 31-32.

[12] Johns, K., Williams, S., Baker, K and Sun, L. (2003). Uses of a Learning Object, Technology Enhanced Learning, International Conference (T.E.L.'03), Milan, Italy, pp 7.

[13] Lakhan, V. C and Lavalle, P. D. (2002). Use of Log linear Models to Assess Factors Influencing Concern for the Natural Environment, *Environmental Management*, 30 (1), 77-87

[14] Little, B. B., Johnston, W. L., Lovell, A. C., Rejesus, R. M and Steed, S. A. (2002). Collusion in the U. S. Crop Insurance Program: Applied Data Mining, in Proceedings of the Second SIAM International Conference on Data Mining, Arlington, USA, pp 15, retrieved October 24, 2004 from, http://www.siam.org /meetings/sdm02/proceedings/sdm02-34.pdf

[15] Lubega, J. T. and Williams, S. A. (2003). The Effects of a Managed Learning Environment (Blackboard) on the performance of Students, in the 6th International Workshop on Interactive Computer Learning (ICL 2003), Villach, Austria, pp 6.

[16] McCutcheon, A. and Mills, C. (1998). Categorical data analysis: Log-linear and latent class models. In Elinor Scarbrough and Eric Tanenbaum, editors, *Research Strategies in the Social Sciences: A Guide to New Approaches*. Oxford University Press, Oxford.

[17] Meliha, H. and Denise T. (2002). Evaluating an interactive learning environment in management education, *Educational Technology & Society*, 5 (3), 113-122.

[18] Melville, J and Swain, R., (1997). Spatial separation in two sympatric skinks, Niveoscincus microlepidotus and N. metallicus, from Tasmania, Herpetologica, 53, 126-132.

[19] Neter, J., Wasserman, W. and Kutner, M. H. (1990). Applied Linear Statistical Models, 3rd Edition, Irwin, Boston, MA.

[20] Neyman, J., and Pearson, E. S. (1931). On the problem of k samples, Bulletin de l'Academie, Polonaise des Sciences et Lettres, Ser. A, pp 460-481.

[21] Nydia R. H., (2002). Effective Use of a Range of Authentic Assessments in a Web Assisted Pharmacology Course, Educational Technology & Society, 5 (3), 123-137.

[22] Prawatt, R.S., & Floden, R.E. (1994), Philosophical perspective on constructivist views of learning. Educational Psychology, 29 (1), 37-48.

[23] Pyle, D., (1999) Data Preparation For Data Mining, Morgan Kaufmann Publishers, Inc, USA.

[24] StatSoft Inc, Statistica for Windows (1984-2004), retrieved April 12, 2004 from http://www.statsoftinc.com/

[25] Thurstone, L. L., (1931). Multiple factor analyses, Psychological Review, 38, 406-427.

[26] Viinamaki, H., Koivumaa-Honkanen, H., Hintikka, J., Tanskanen, A., and Lehtonen, J. (1998).Association between religious attendance, social support, and depression in psychiatric patients, Journal of psychology and theology, 26 (4), 351-357, retrieved November 10, 2003 from https://wisdom.biola.edu/jpt/issuecont.cfm?volid=29&issueid=2#p158

[27] Western, M. (1996). An overview of topological and multidimensional association log-linear models for contingency table analysis, Australian Journal of Social Research, 2 (1), 85-113

[28] Williams, S. and Walmsley, S. (1999). Discover Delphi, Addison Wesley, UK.