Late Class Comer in Universiti Teknologi Malaysia

Group 17

Tay Yee Yang, Low Kar Seng,

Muhammad Hakim Bin Md Yussof,

Muhammad Rafiuddin Bin Mar Sidek

***Abstract*— The UTM FKE student come late to the class has been an issue that troubling every lecturer as it might slow down the class progress and influence the image of university. The core reason of student come late to class must be find out in order to take a right action on solving this issue. The data were collected from 47 FKE student through survey form. Chi-square test analysis was conducted to examine the association to find out the core factor of student come late to class. After the analysis, time management skill of the student is proven as the most significance factor for student to come late to class. Actions should be taken to enhance the time management skill of the student such as organizing campaign, enforce punishment, or time management aided technology solution can be created.**

I. INTRODUCTION

Late coming to class is one of the major problems that have plagued many universities. This problem started long ago and keeps spreading. It has contributed immensely in a negative way to the academic achievement of learners and functioning of the universities. The purpose of education in universities is to produce successful person with good personalities. One of the important traits entering into a successful person's character is the habit of punctuality.

Lateness is a function of time and as such time is usually used as the criteria for determining lateness. In simple words, lateness merely means arriving after expected time. The behaviour of coming late to class may become a habit with the individual involved and may have negative consequences not only to the individual itself but also to people surrounding him or her.

There are lots of effects coming late to class for example when students come late to class, it can disrupt the flow of a lecture or discussion, distract other students, impede learning and generally erode class morale. Furthermore, if left unchecked, lateness can become chronic and spread throughout the class. Besides that, lateness also could result into poor academic achievement, reduce the ability to meet the instructional target and damage the universities reputation. Considering which causes are at the root of the problem can help the lecturers or instructors to appropriate responses to this problem and come out with some solutions.

The aim of this study is to investigate different causes of late coming among students in Electrical Engineering Faculty of Universiti Teknologi Malaysia Johor Bahru. Questions were asked on the causes of late-coming including the factors that might contribute to the situation.

II. METHODOLOGY

A survey form was conducted to study the reason of late comer. There 47 survey questionnaires were answer by 47 UTM FKE student. The result of the questionnaire was tabulated and store in an excel.csv file.

Before any analysis procedure started, the data were first undergone transformation in Excel because some of the string attribute variable are hardly analyse statistically. Thus, all the categorical variables were transform into nominal or numeric ordinal variable. The procedure coding of transformation is shown as Figure 2.1 where variable "How did you attend class daily?" is transforming into "Transport_score". The transformation before and after for each variable were also stated in the Table 2.1.

```
=IF(F3="Drive (own Motor/Car/Bicycle)",3,IF(F3="Drive (own Motor/Car/Bicycle/Walking)",3,IF(F3="Car pool with friend",2,IF(F3="Public transport (bus)",1,0))))
```

Figure 2.1: Coding of transformation for one of the variable in Excel.

TABLE 2.1
The transformation of variable

Transformation	
Before	After
How did you attend class daily?	Transport_score
Drive (own Motor/Car/Bicycle/Walking)	3
Car pool with friend	2
Public transport (bus)	1
How frequent you come late to class	Late_score
Late for all class or <80%	1
<60% or <40%	2
<20% or 0% (never late)	3
Distance from room to class?	Distance_score
5-6km	1
4-5km	2
3-4km	3
2-3km	4
1-2km	5
0-1km	6
Are you good in time management?	Time_Management
No	1
Maybe	2
Yes	3

The analysis was conducted to study the association or relationship between the predictor variable and response variable as shown as Table 2.2. The target of the analysis was to verify whether the predictor variable influence the response or not and to find out which variable is the most influential to the student resulting into coming late to the class.

TABLE 2.2
The transformation of variable

Predictor Variable	Response Variable
Gender	Late_score
Study Year	
Distance_score	
Transport_Score	
Time Management	

To perform a formal test of association between two categorical variables, the chi-square test was conducted. There are 2 types of chi-square tests were used in the project which are Pearson chi-square test or simply known as the chi-square test and Mantel-Haenszel chi-square test.

Pearson Chi-square test is start with the null hypothesis stated there is no association between the variables and the alternative hypothesis stated there is association between variable[1]. To determine whether there is an association, the Pearson chi-square test measures the difference between the observed cell frequencies and the expected frequencies that there is no association between the variables. In other words, the expected frequencies are the frequencies you expect to get if the null hypothesis is true. The chi-square test calculates the expected frequencies for each cell by multiplying the row total (R) by the column total (C), and then dividing the result by the total sample size (T). If the observed frequencies equal the expected frequencies, there is no association between the variables. If the observed frequencies do not equal the expected frequencies, there is an association between the variables. The chi-square statistic indicates the difference between observed frequencies and expected frequencies. A significant chi-square statistic is robust evidence that an association exists between your variables. The other way to prove the association of variable is examine the p-value[2]. If the p-value is smaller than alpha, α=0.05, the null hypothesis is rejected because the possibility of Type I error is significance low. Thus, we may conclude there is association between variables. The Pearson chi-square statistic and its corresponding p-value indicate only whether an association exists between two categorical variables. The p-value indicates how confident you can be that the null hypothesis of no association is false. However, neither the chi-square statistic nor its p-value concludes the magnitude of an association. Thus, Cramer's V statistic was computed in the project. Cramer's V statistic is derived from the Pearson chi-square statistic and it is one measure of the strength of an association between two categorical variables[3].

Mantel-Haenszel chi-square test is a more powerful test to check the ordinal association when two variables are ordinal variable. When two variables have an ordinal association, an increase in the value of one variable tends to be associated with an increase or decrease in the value of the other variable. Compare to Pearson chi-square test which is only appropriate to check the association between variables, Mantel-Haenszel chi-square can verify the influence of increasing or decreasing of predictor variable to a response variable[4]. For the Mantel-Haenszel chi-square test, the null hypothesis state that there is no ordinal association between the row and column variables, whereas, the alternative hypothesis state that there is an ordinal association between the row and column variables. The Mantel-Haenszel chi-square statistic and its corresponding p-value are similar to the Pearson chi-square statistic and its corresponding p-value in the following ways. They indicate the association exists when p-value is smaller than alpha, α=0.05. To measure the strength of the association between two ordinal variables, Spearman correlation statistic was calculated in the project. Both Cramer's V Statistic for Pearson Chi-square test and Spearman correlation for Mantel-Haenszel chi-square test are rank correlation and has a range between -1 and 1[5]. Values close to 1 indicate that there is a relatively high degree of positive correlation. Values close to -1 indicate that there is a relatively high degree of negative correlation. And values close to zero indicate a weak correlation.

The software was involved in this project is SAS. SAS is an on-cloud data analysis software that provides advance analytical tool. It provides a multifunction analytic algorithm to work with descriptive and inferential statistical test in order to gain valuable insight from the data through SAS technology.

III. Results and Analysis

In this chapter, the analysis was conducted to study the association between each predictor variable to response variable.

A. The association between Gender and Late Score

The analysis was conducted to check whether the gender is the reason contribute to the student late coming late or not. The description of the variable gender and late score is shown as Table 3.1.

TABLE 3.1
The description of variable gender and late score

Variable	Class	Description
Gender	Male	The response is male.
	Female	The response is male.
Late_Score	1	The response is every time or always late to the class.
	2	The response is sometime late coming to class.
	3	The response is least or never late coming to class.

The Chi-square test is run by using SAS data analytical software and the SAS language coding is shown as the Figure 3.1.

```
data Survey_data;
        infile '/folders/myfolders/Engineering Ethic
Assignment/data_survey2.csv' dlm=',' firstobs=2;
        input Gender :$CHAR1. Study_Year
:$CHAR8. Distance_score Transport_Score
Time_management Late_score;
run;
proc freq data=work.survey_data;
                Tables Gender*Late_score/
      chisq expected cellchi2 nocol nopercent
      relrisk;
  title1  'Association between Gender and
Late_Score';
run;
title;
```

Figure 3.1: The SAS coding of Chi-square test for Gender and Late Score.

The variable Gender is nominal. Thus, Pearson Chi-square test was conducted. The analysis result shows the chi square value is 1.1803 and the respective p-value is 0.5542 which is larger than alpha, α=0.05. Thus, we fail to reject null hypothesis and conclude that there is not association between Gender and Late score.

Although there is not significance association between gender and late score but we still can abstract some information from the data. There is another analysis result is shown as Figure 3.2. The crosstabulation table was generated where each cell displays the 4 information which are frequency, expected of frequency, cell chi-square statistic and row percentage. Frequency indicate the number of count of the gender appear at the respective late score. Expected frequency is the expected count under the null hypothesis of no association. Cell chi-square statistic prints each cell's contribution to the total chi-square statistic. The row percent indicates the percent of females or males in each late score category.

Association between Gender and Late_Score

The FREQ Procedure

Frequency
Expected
Cell Chi-Square
Row Pct

Table of Gender by Late_score

Gender	Late_score 1	2	3	Total
F	2 2.0426 0.0009 12.50	2 3.4043 0.5793 12.50	12 10.553 0.1984 75.00	16
M	4 3.9574 0.0005 12.90	8 6.5957 0.299 25.81	19 20.447 0.1024 61.29	31
Total	6	10	31	47

Figure 3.2: The association between Gender and Late Score.

From the crosstabulation table, odd ratio can be calculated[6]. Odd ratio used to compare the likelihood that a specific event or outcome will occur in one group as compared with another group. The equation odd ratio is shown below and the respective odd ratio for each group were calculated and tabulated in Table 3.2.

$$odd\ ratio = \frac{(possibility\ of\ event\ in\ Group\ A)/(1-possibility\ of\ event\ in\ Group\ A)}{(possibility\ of\ event\ in\ Group\ B)/(1-possibility\ of\ event\ in\ Group\ B)}$$

TABLE 3.2
Shows the odd ratio for each group.

Odd ratio	Late score		
Gender	1	2	3
F	0.9646	0.4106	1.8948
M	1.036699	2.43546	0.52776

We can interpret Table 3.2 and make some conclusions from it. For late score group "1", both male and female are nearly no different where the ratio is approximate equal to 1. In other words, there are equal portion of male and female student every time or always come late to the class. For late score group "2", male student is more likely to be sometime late coming to class where the odd ratio is 2.4355. In other words, every 7 students come late to class frequently, there are 5 must be male. For late score group "3", female student is more likely to be least or never come late to class where the odd ratio is 1.8948. In other words, if there are 14 students always come punctual to the class, 9 of them must be female student. In short, based on the analysis of odd ratio, we can conclude that female student is more likely to be more punctual than male student.

B. The association between Distance_score and Late Score

The analysis was conducted to check whether the year of the student is the reason contribute to the student late coming late or not. The description of the variable study year and late score is shown as Table 3.3.

TABLE 3.3
The description of study year and late score.

Variable	Class	Description
Study year	1st year	The response is 1st year student.
	2nd year	The response is 2nd year student.
	3rd year	The response is 3rd year student.
	4th year	The response is 4th year student.
Late_Score	1	The response is every time or always late to the class.
	2	The response is sometime late coming to class.
	3	The response is least or never late coming to class.

The variable Study year and Late score are ordinal variable. Thus, Mantel-Haenszel Chi-Square test was conducted. The analysis result shows the Mantel-Haenszel Chi-Square value is 0.3988 and the respective p-value is 0.5277 which is larger than alpha, α=0.05. Thus, we fail to reject null hypothesis and conclude that there is not association between Study Year and Late score.

C. The association between Distance score and Late Score

The analysis was conducted to check whether the distance from the room to the class is the reason contribute to the student late coming late or not. The description of the variable distance score and late score is shown as Table 3.4.

TABLE 3.4
The description of distance score and late score

Variable	Class	Description
Distance_score	1	The distance between the response's room to the class is 5-6km
	2	The distance between the response's room to the class is 4-5km
	3	The distance between the response's room to the class is 3-4km
	4	The distance between the response's room to the class is 2-3km
	5	The distance between the response's room to the class is 1-2km
	6	The distance between the response's room to the class is 0-1km
Late_Score	1	The response is every time or always late to the class
	2	The response is sometime late coming to class
	3	The response is least or never late coming to class

The variable Distance_score and Late score are ordinal variable. Thus, Mantel-Haenszel Chi-Square test was conducted. The analysis result shows the Mantel-Haenszel Chi-Square value is 0.0372 and the respective p-value is 0.8470 which is larger than alpha, α=0.05. Thus, we fail to reject null hypothesis and conclude that there is not association between Distance_score and Late score.

D. The association between transport score and Late Score

The analysis was conducted to check whether the type of transport of the student is the reason contribute to the student late coming late or not. The description of the variable transport score and late score is shown as Table 3.5.

TABLE 3.5
The description of transport score and late score

Variable	Class	Description
Transport_score	1	The distance between the response's room to the class is 5-6km
	2	The distance between the response's room to the class is 4-5km
	3	The distance between the response's room to the class is 0-1km
Late_Score	1	The response is every time or always late to the class
	2	The response is sometime late coming to class
	3	The response is least or never late coming to class

The variable Distance_score is nominal variable. Thus, Pearson Chi-square test was conducted. The analysis result shows the chi square value is 1.0955 and the respective p-value is 0.8950 which is larger than alpha, α=0.05. Thus, we fail to reject null hypothesis and conclude that there is not association between Gender and Late score.

E. The association between time management and Late Score

The analysis was conducted to check whether the time management skill of the student is the reason contribute to the student late coming late or not. The description of the variable time management and late score is shown as Table 3.6.

TABLE 3.6
The description of time management and late score.

Variable	Class	Description
Time mangement	1	The response are not good in time management.
	2	The response are partially good in time management.
	3	The response are good in time management.
Late_Score	1	The response is every time or always late to the class
	2	The response is sometime late coming to class
	3	The response is least or never late coming to class

The variable time management and Late score are ordinal variable. Thus, Mantel-Haenszel Chi-Square test was conducted. The analysis result shows the Mantel-Haenszel Chi-Square value is 12.1325 and the respective p-value is 0.0005 which is smaller than alpha, α=0.05. Thus, we have enough evidence to reject null hypothesis as the possibility of Type I error is very low. We can also conclude that there is an ordinal association between time management and Late score.

The Table 3.7 also displays the measures of association including the Spearman Correlation statistic and its 95% confidence limits. The Spearman Correlation statistic indicates the magnitude of an ordinal association. The value is 0.5256 which is far from 0 and it is positive, so there is a positive ordinal association between time management and late coming habits. In other words, we can conclude that as the time management skill better, the student tends to be more punctual to the class. The 95% confidence interval does not contain 0. This also means that the positive ordinal relationship is significant at the 0.05 significance level and we are 95% confident that the

Spearman Correlation significantly different from 0 in other different sample data. The asymptotic standard error(ASE) is used to calculate the confidence intervals for various measures of association. The asymptotic standard error is 0.1034 which indicate there is a small variability of the correlation statistic. In other words, it indicates there is small error we can expect if we use this sample data to estimate the population correlation.

TABLE 3.7
The measures of association.

Statistic	Value	ASE	95% Confidence Limits	
Spearman Correlation	0.5256	0.1034	0.3230	0.7282

IV. Conclusion

The project was conducted to find out the reasons that cause the UTM FKE student to come late to class. We examined the association between each different predictor variable with response variable to find out what is the reason of the UTM FKE student come late to class. In the end of the approach, the most significance association of late comer is the time management skill of the student. We also conclude that the better time management skill of the student, the less frequent to come late to the class for the student.

After using data analysis to verify the core reason of student come late to class is about time management. There are some action and effort can be made to improve this situation by enhancing the time management skill of student. University can organize campaign to teach the techniques of efficient time management. The IOT technology solution or mobile APP also can be created to aid the student to be more punctual to the class. We also can enforce some punishment to cultivate the time management skill of the student.

References

[1] T. M. Franke, T. Ho, and C. A. Christie, "The Chi-Square Test: Often Used and More Often Misinterpreted," *Am. J. Eval.*, vol. 33, no. 3, pp. 448–458, 2012.

[2] T. V. Perneger and C. Combescure, "The distribution of P-values in medical research articles suggested selective reporting associated with statistical significance," *J. Clin. Epidemiol.*, vol. 87, pp. 70–77, 2017.

[3] M. L. Mchugh, "The Chi-square test of independence Lessons in biostatistics," *Biochem. Medica*, vol. 23, no. 2, pp. 143–9, 2013.

[4] M. P. Michaelides, "An Illustration of a Mantel-Haenszel Procedure to Flag," *Pract. Assessment, Res. Eval.*, vol. 13, no. 7, pp. 1–16, 2008.

[5] J. Hauke and T. Kossowski, "Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data," *Quaest. Geogr.*, vol. 30, no. 2, pp. 87–93, 2011.

[6] C. M. Laing and J. A. Rankin, "Odds Ratios and Confidence Intervals," *J. Pediatr. Oncol. Nurs.*, vol. 28, no. 6, pp. 363–367, 2011.