**Study the effects of some factors on the better performance of online semester examination**

PROJECT SUBMITTED FOR PARTIAL FULFILLMENT

OF BACHELOR’S DEGREE IN STATISTICS HONOURS



## VIDYASAGAR UNIVERSITY

Registration no.-1160499

Roll-1126116

No-180167

Session-2018-2019

Acknowledgment

The success and final outcome of this project required a lot of guidance and assistance

from many people and I am extremely privileged to have got this all along the completion of my project. All that I have done is only due to such supervision and assistance

and I would not forget to thank them. I respect and thank , for providing me an opportunity to do the project work in department of Statistics, Haldia Government College

and giving us all support and guidance which made me complete the project duly. I am

extremely thankful to him for providing such a nice support and guidance, although

he had busy schedule managing the corporate affairs. I owe my deep gratitude to our

project guides Dr. Shyamsundar Sahoo, Mr. Sibsankar Karan, Mr. Tanmay Kumar

Maity and Mr. Tuhinsubhra Bhattacharya, who took keen interest on our project work

and guided us all along, till the completion of our project work by providing all the

necessary information for developing a good system. I would not forget to remember

my parents, and also my friends for their encouragement and more over for their timely

support and guidance till the completion of our project work. I am thankful to and fortunate enough to get constant encouragement, support and guidance from all Teaching

staffs of Statistics department which helped us in successfully completing our project

work. Also, I would like to extend our sincere esteems to all staff in laboratory for their

timely support.

Contents

* Introduction
* Objective
* Data collection
* Data description
* Methodology
* Results
* Conclusion
* References
* Appendix

INTRODUCTION:

In early 2020, a global pandemic (COVID-19) broke out and severely affected the progress of education in various countries’ universities and institutions, which promoted the progress of online courses at the same time. The aims of this project to conduct a comparative analysis of student surveys between online live teaching and traditional off-line teaching.

As the Covid outbreak continues in India, students are caught in the eye of a storm. Administrative bodies have been dithering over whether or not online examinations should be conducted .The University Grants Commission has reportedly recommended scrapping the online exams for for all semester students, while human resource development minister Ramesh Pokhriyal  tweeted that the “foundation for revisited guidelines shall be health and safety students, teachers and staff”. Earlier in April, the UGC had said that final-year examinations could happen in July.

A New development as a result of the Covid pandemic is online exam. The Covid-19 pandemic has altered everyone’s lives significantly and students certainly weren’t exempt .We were unable to return to University and sit our examinations as normal and so many University resorted to online exams .Several students have perfect scores with some help from Google, their friends .In our current global situation, exams might not seem a priority. But for students, exam stress combined with pandemic stress makes for a pretty hectic time.

\*\*Here the aim of this study is the analysis of online examination system and offline examination system. Also want to know is the online examination system is truly helpful for students or not in future.

**Objective of this study is shown as given below:**

Check the changes of marks of different categories. Try to find the reason behind the changes of result (If difference of marks is too high).Also want to check whether the difference of marks depend on some variable such as online class percentage, communication with teacher ,exam time etc., by some statistical tools .Find which variable does effect the data. Then want to fit a model for difference category by the significant factor.

**Data Collection:**

For this study the marks are collected from student of offline examination. According to marks the students are divided into 4 categories such as bad, medium, good, very good. Then I collect marks of online exam from those students. Compare the result between offline and online .

For this COVID-19 pandemic, the data is collected through online by sharing a google form to students and they circulate it as well.A google form is created along with the questions mentioned in the appendix part. My teacher helped out for creating those questions.

**Description of The Data:**

The data is collected from all the students (Male & Female) who are now pursuing B.sc/B.A Honors from Haldia Government College by sharing google form and call them to fill the form.

The target sample is >70 students and got total 149 observations.

**Methodology:**

This is to be checked whether the difference of marks depend on any factor or not. If so then want to fit a model by significant independent factor to predict the response variable (Difference of marks), so that if anyone give the value of independent factor one can tell whether his/her marks change or not.

**The following tools have been used**

1. To find the dependency between factors:

* **Chi square independence test.**
* **Kendall’s tau b measure of association.**
* **Dummy variable for converting the categorical variable to numeric variable.**

1. To fit the model:

* **Multinomial Logistic Model.**

**Chi-square Test for Independence:**

The Chi-square test of independence checks whether two variables are likely to be related or not. We have counts for two categorical or nominal variables. We also have an idea that the two variables are not related. The test gives us a way to decide if our idea is plausible or not.

This is the motivation behind the hypothesis for the Chi-Square Test of Independence:

* H0: In the population, the two categorical variables are independent.
* Ha: In the population, the two categorical variables are dependent

**Chi-Square Test Statistic**

The Chi-Square test statistic is calculated as follows:

χ2=

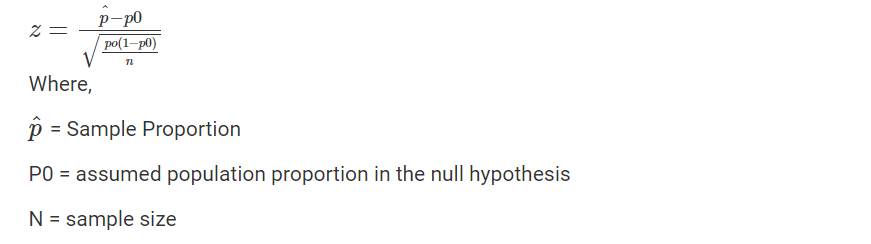
Under the null hypothesis and certain conditions (discussed below), the test statistic follows a Chi-Square distribution with degrees of freedom equal to (r−1)(c−1), where r is the number of rows and c is the number of columns. We leave out the mathematical details to show why this test statistic is used and why it follows a Chi-Square distribution.

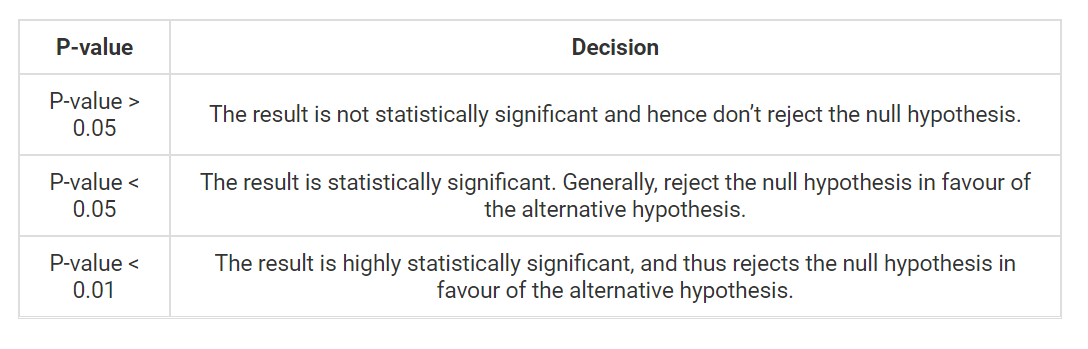
As we have done with other statistical tests, we make our decision by either comparing the value of the test statistic to a critical value (rejection region approach) or by finding the probability of getting this test statistic value or one more extreme (p-value approach).

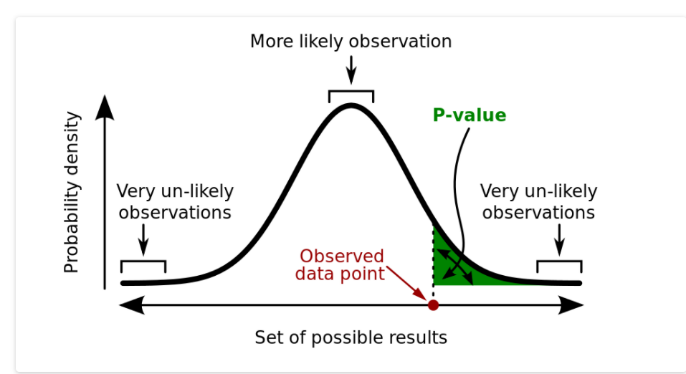
## **What Is P-Value?**

In statistics, the p-value is the probability of obtaining results at least as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is correct. The p-value is used as an alternative to rejection points to provide the smallest level of significance at which the null hypothesis would be rejected. A smaller p-value means that there is stronger evidence in favour of the alternative hypothesis.

We Know that P-value is a statistical measure, that helps to determine whether the hypothesis is correct or not. P-value is a number that lies between 0 and 1. The level of significance(α) is a predefined threshold that should be set by the researcher. It is generally fixed as 0.05. The formula for the calculation for P-value is





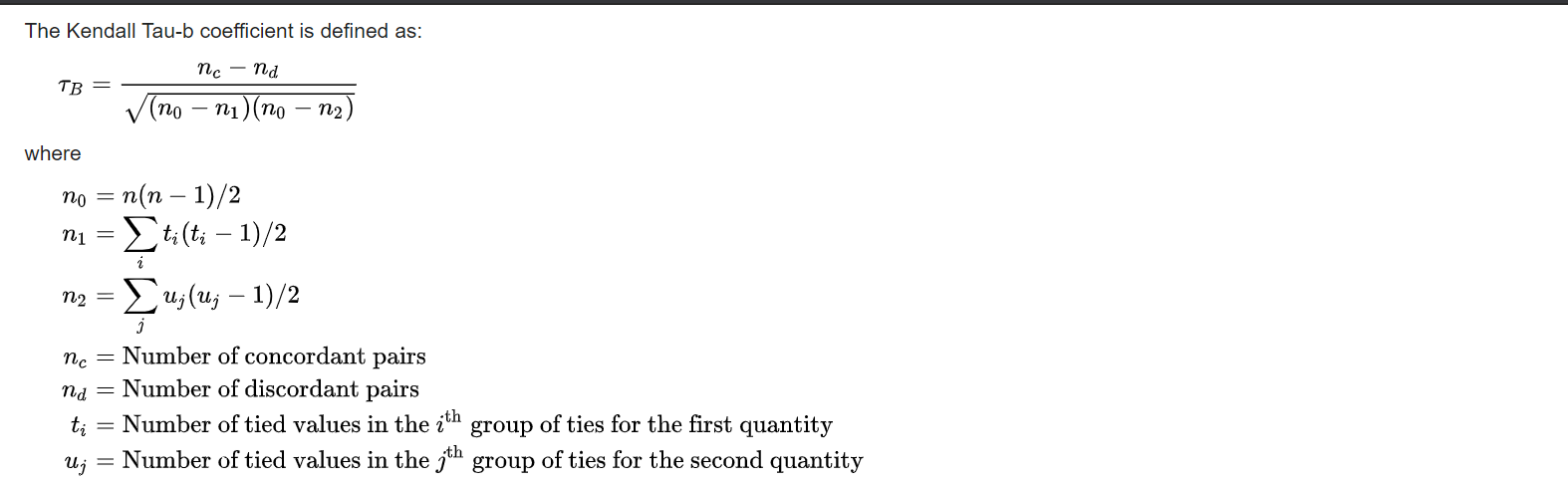


**Kendall's Tau-b**

In statistics the **Kendall’s tau-b** is a statistic used to measure the ordinal association between two measured quantities. A **τ test** is a no parametric hypothesis for statistical dependence based on the τ coefficient.

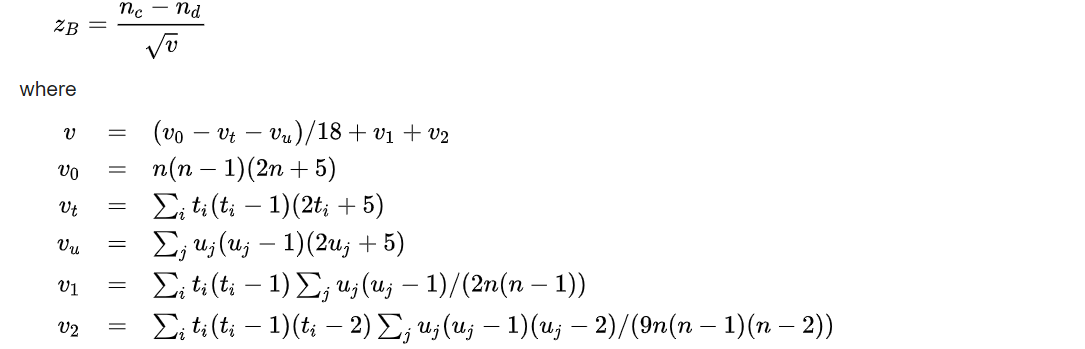
**Concordance and Discordance Pair:**

Let (x1,y1),…..,(xn,yn) be a set of observations of the joint random variables X and Y, such that all the values of (xi) and (yi) are unique (ties are neglected for simplicity).Any pair of observations (xi,yi) and (xj,yj), where i<j, are said to be concordant if the sort order 0f (xi,yj) and (yi,yj) agrees: that is, if either both xi > xj and yi > yj holds or both xi < xj and yi < yj; otherwise they are said to be discordant.

****

**Significance Test:**

When two quantities are statistically independent, the distribution of tau{\displaystyle \tau } is not easily characterizable in terms of known distributions. However, for {\displaystyle \tau \_{A}}TT the following statistic, {\displaystyle z\_{A}}ZB, is approximately distributed as a standard normal when the variables are statistically independent:

****

**What is dummy variable?**

**A dummy variable is a variable that takes values of 0 and 1, where the values indicate the presence or absence of something (e.g., a 0 may indicate a placebo and 1 may indicate a drug). Where a categorical variable has more than two categories, it can be represented by a set of dummy variables, with one variable for each category.**Numeric**variables can also be**dummy coded**to explore**nonlinear effects**.**Dummy variables**are also known as**indicatorvariables**,** designvariables***,***contrasts***,***one-hot coding***,* and**binary basis variables.

**Multinomial Logistic Model:**

Multinomial logistic regression is used when the dependent variable in question is nominal (equivalently *categorical*, meaning that it falls into any one of a set of categories that cannot be ordered in any meaningful way) and for which there are more than two categories.

**Assumption:**

The multinomial logistic model assumes that data are case-specific; that is, each independent variable has a single value for each case. The multinomial logistic model also assumes that the dependent variable cannot be perfectly predicted from the independent variables for any case. As with other types of regression, there is no need for the independent variables to be statistically independent from each other (unlike, for example, in a naive Bayes classifier); however, collinearity is assumed to be relatively low, as it becomes difficult to differentiate between the impact of several variables if this is not the case.

**Connection with baseline category logit model:**

For nominal scale response variables, the standard logits are the baseline category logits. The odder of the response categories is then irrelevant ,and we contrast an arbitrary baseline category against each of the other categories. When category c is the baseline category ,these logits are

C:\Users\Akash\Pictures\Screenshots\Screenshot (44).png

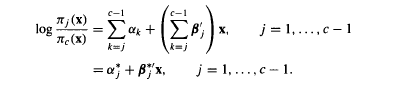
As noted in section ,the adjacent categories logits are a basic set of logits that are equivalent to the base line category logits. For baseline category c

C:\Users\Akash\Pictures\Screenshots\Screenshot (45).png

Models using adjacent categories logits can be expressed as baseline category logit models. For the general adjacent categories logit model,

C:\Users\Akash\Pictures\Screenshots\Screenshot (46).png

The equivalent baseline category logit model is

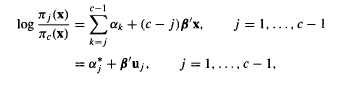


This model has the form of an ordinary baseline category logits model. Because it does not assume a common effect for each j, this model dose not utilize the ordinary of y.

Of greater interest for ordinal responses is the proportional odds form of the adjacent categories logits model. With common effect for each logit.

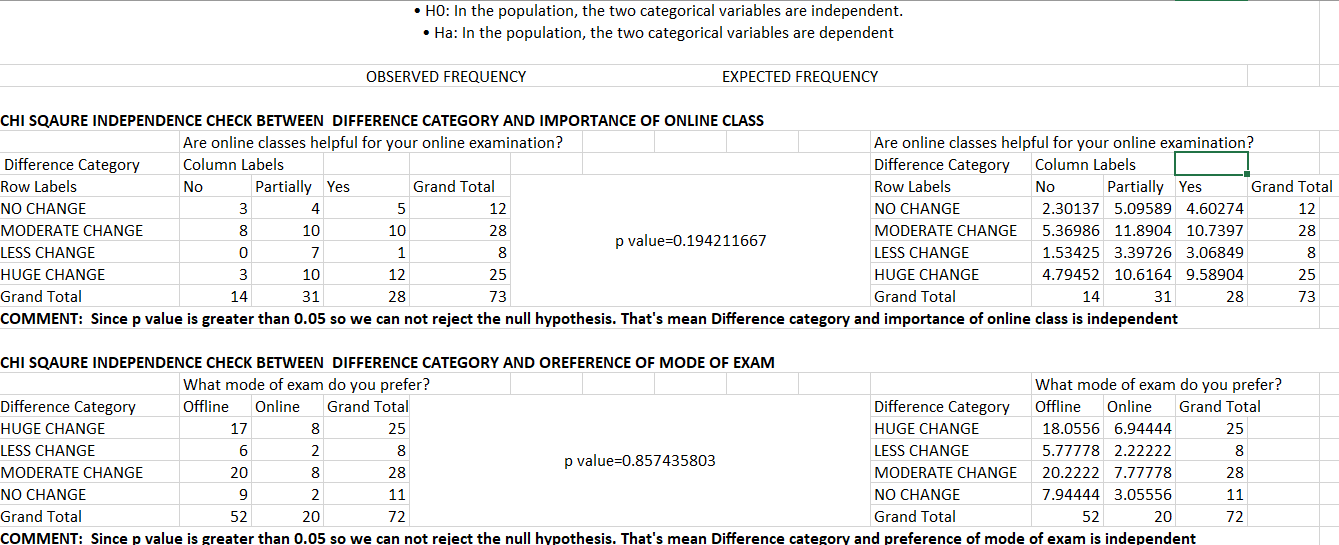
C:\Users\Akash\Pictures\Screenshots\Screenshot (48).png

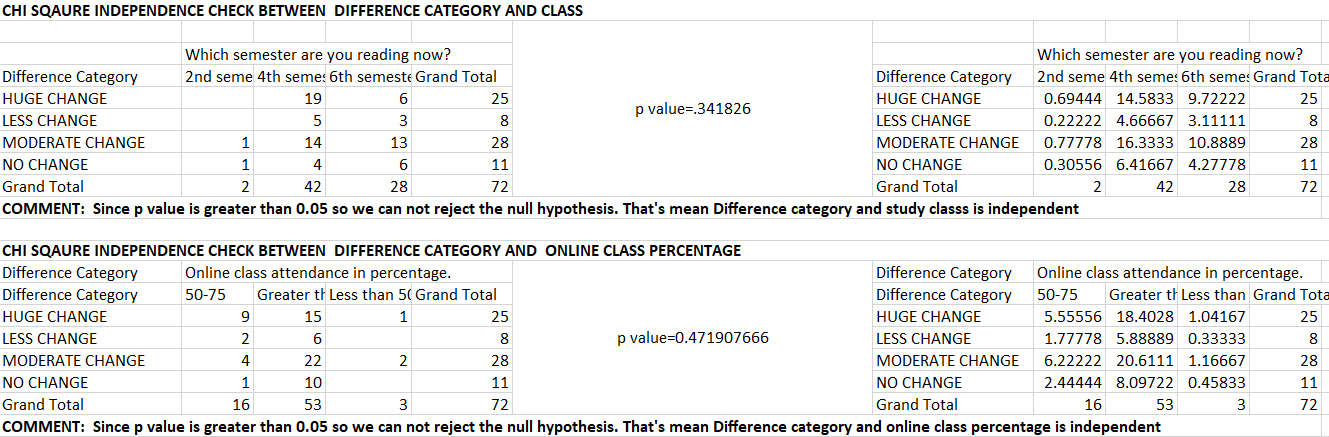
The equivalent baseline category logit model is

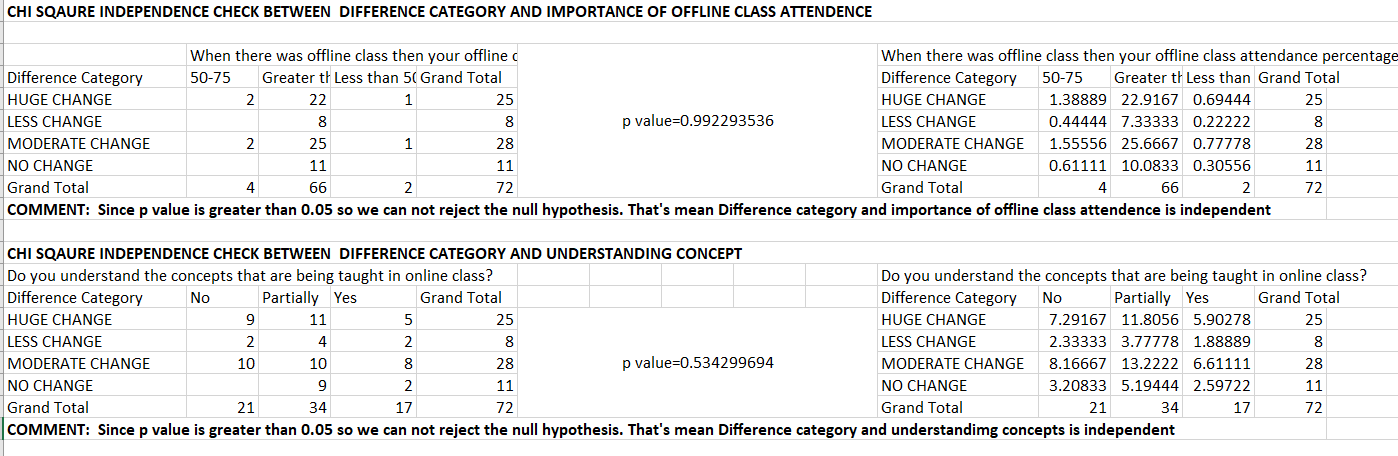


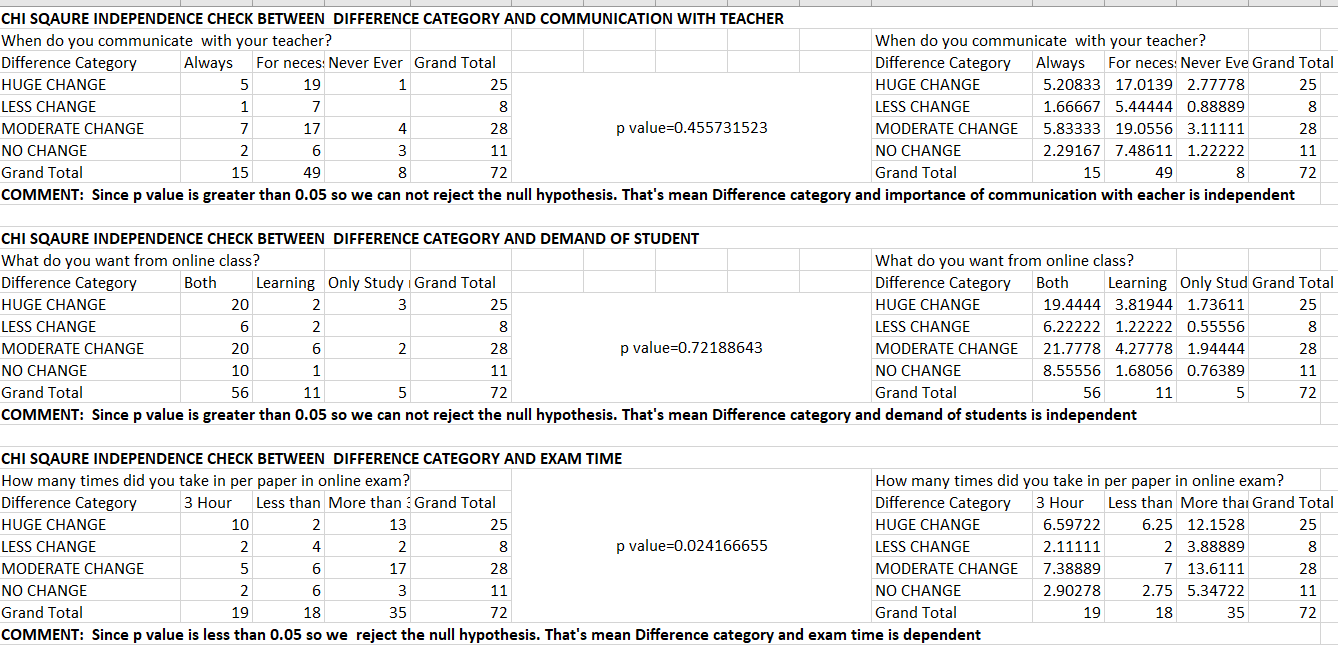
**Result:**

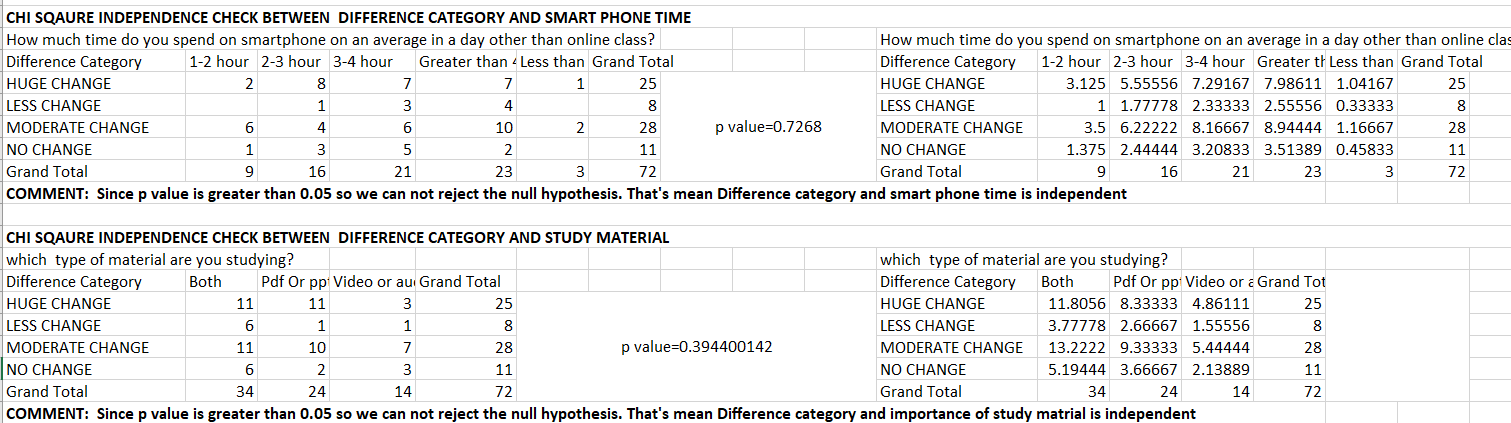
**Output of chi square independence :**











**Output of KENDALL TAU b measure of association:**

**1.Association between Difference of category and online class attendance .**

|  |  |  |  |
| --- | --- | --- | --- |
| Online class attendance in percentage. | | | |
| Difference Category | Less than 50 | 50-75 | Greater than 75 |
| NO CHANGE | 0 | 1 | 10 |
| LESS CHANGE | 0 | 2 | 6 |
| MODERATE CHANGE | 2 | 4 | 22 |
| HUGE CHANGE | 1 | 9 | 15 |

**Concordant pair, discordant pair, total tied pairs:**

|  |  |  |  |
| --- | --- | --- | --- |
| C | 509 | T(x) | 761 |
| D | 158 | T(y) | 1501 |

**Kendall’s tau b :**

|  |  |
| --- | --- |
| Tau b | 0.255064 |

**Significant test:**

**H0:** **The two ordered categorical variables are independent and sigma(C-D) depends only on the sample size and the true marginal proportions.**

**H1:They are dependent.**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | z | 2.355798 |  | > z\_crit | 1.96 | |  |  |  |  |

**H0 is rejected.**

**Comment:**

**Difference of category and online class attendance are positively associated.**

**2.Association between Difference of category and offline class attendance.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Offline class attendance percentage |  |  |
| Difference Category | Less than 50 | 50-75 | Greater than 75 |
| NO CHANGE | 0 | 0 | 11 |
| LESS CHANGE | 0 | 0 | 8 |
| MODERATE CHANGE | 1 | 2 | 25 |
| HUGE CHANGE | 1 | 2 | 22 |

**Concordant pair, discordant pair, total tied pairs:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| C | 191 |  | T(x) | 761 |
| D | 68 |  | T(y) | 2152 |

**Kendall’s tau b :**

|  |  |
| --- | --- |
| Tau b | 0.144438 |

**Significant test:**

**H0:** **The two ordered categorical variables are independent and sigma(C-D) depends only on the sample size and the true marginal proportions.**

**H1:They are dependent.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Significant test:  H0: The two ordered categorical variables are independent and sigma(C-D) depends only on the sample size and the true marginal proportions.  H1:They are dependent.  Z=1.38<z\_crit=1.96   |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | |  |  |  |  |  | | --- | --- | --- | --- | --- | |  |  |  |  |  | |  |  |  |  |   H0 is accepted. |  |  |  |  | |  |  |  |  |

Comment:

Difference of category and offline class attendance are not significant pair.

**3.Association between Difference of category and understanding concepts.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Do you understand the concepts that are being taught in online class? |  |  |
| Difference Category | No | Partially | Yes |
| NO CHANGE | 0 | 9 | 2 |
| LESS CHANGE | 2 | 4 | 2 |
| MODERATE CHANGE | 10 | 10 | 8 |
| HUGE CHANGE | 9 | 11 | 5 |

**Concordant pair, discordant pair, total tied pairs:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| C | 687 |  | T(x) | 761 |
| D | 465 |  | T(y) | 907 |

**Kendall’s tau b :**

|  |  |
| --- | --- |
| Tau b | 0.129036 |

**Significant test:**

**Z=1.23<z\_crit=1.96**

**H0 is accepted.**

**Comment:**

**Difference of category and understanding the concept are not a significant pair.**

**4.Association between Difference of category and communication with teacher.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | When do you communicate with your teacher? |  |  |
| Difference Category | Always | For necessity | Never Ever |
| HUGE CHANGE | 5 | 19 | 1 |
| MODERATE CHANGE | 7 | 17 | 4 |
| LESS CHANGE | 1 | 7 | 0 |
| NO CHANGE | 2 | 6 | 3 |

**Concordant pair, discordant pair, total tied pairs:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| C | 511 |  | T(x) | 761 |
| D | 359 |  | T(y) | 1309 |

**Kendall’s tau b :**

**Tau b=0.1012**

**Significant test:**

**Z=-3.608<z\_crit=-1.96**

**H0 is rejected**

**Comment:**

**Difference of category and communication with teacher are positively associated.**

**5. Association between Difference of category and online exam time**

|  |  |  |  |
| --- | --- | --- | --- |
|  | How many times did you take in per paper in online exam? |  |  |
| Difference Category | Less than 3 hour | 3 Hour | More than 3 hour |
| NO CHANGE | 6 | 2 | 3 |
| LESS CHANGE | 4 | 2 | 2 |
| MODERATE CHANGE | 6 | 5 | 17 |
| HUGE CHANGE | 2 | 10 | 13 |

**Concordant pair, discordant pair, total tied pairs:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| C | 381 |  | T(x) | 761 |
| D | 775 |  | T(y) | 919 |

**Kendall’s tau b :**

**Tau b=-0.23**

**Significant test:**

**Z=-2.20<z\_crit=-1.96**

**H0 is rejected**

**Comment:**

**Difference of category and online exam time are positively associated.**

**6. Association between Difference of category and spending time on smart phone.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | How much time do you spend on smartphone on an average in a day other than online class? |  |  |  |  |
| Difference Category | Greater than 4 hour | 3-4 hour | 2-3 hour | 1-2 hour | Less than 1 hour |
| HUGE CHANGE | 7 | 7 | 8 | 2 | 1 |
| MODERATE CHANGE | 10 | 6 | 4 | 6 | 2 |
| LESS CHANGE | 4 | 3 | 1 | 0 | 0 |
| NO CHANGE | 2 | 5 | 3 | 1 | 0 |
| Grand Total | 23 | 21 | 16 | 9 | 3 |

**Concordant pair, discordant pair, total tied pairs:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| C | 609 |  | T(x) | 761 |
| D | 675 |  | T(y) | 622 |

**Kendall’s tau b :**

Tau b =-0.035

Significant test:

Z=-0.35>z\_crit=-1.96

H0 is accepted.

Comment:

Difference of category and spending time on smart phone are not a significant pair.

Multinomial logistic model for fitting:

The factor, communication with teacher is not a ordered categorical variable so it should be convert into dummy variable .

The fitted model is given by

log(P())= -0.01623 -2.8203\*x1+2.9878\*x2.A-0.1488\*x2.FN--2.8552\*x2.N+ 2.7716\*x3

log(P())= 0.1381-0.4521\*x1+ 1.0417\*x2.A -0.0049\*x2.FN -0.8986318\* x2.N+0.8666\*x3

log(P())= -0.0999-2.9396\*x1+ 2.7599\*x2.A -0.4712\*x2.FN -2.3887\*x2.N+2.6397\*x3

Where the factors are

X1: student’s Exam time

X2.A: students are always communicate with teacher.

X2.FN: students are communicate with teacher for necessity **.**

X2.N: students are never communicate with teacher.

The p value of each independent factor is given by

**(Intercept) x1 x2.A x2.FN x2.N x3**

**LESS CHANGE 0.87131708 1.344206e-06 7.971811e-05 0.008050549 0.002260521 6.268090e-06**

**MODERATE CHANGE 0.08738511 2.290870e-01 4.114412e-02 0.009891976 0.060753191 1.167484e**

**NO CHANGE 0.36240867 1.149685e-05 3.839121e-04 0.04668241 0.011366716 4.847036e-05**

So here all the p value are less than 0.05 so the factors are significant except x2.FN .

Multinomial model predicts the model correctly on an average 78%.

**Conclusion:**

Difference category depends on how much time students take for exam in each paper, whether they communicate with teacher or not and lastly whether they attend online class regularly or not.

Using the three significant factor as an independent factor the multinomial logistic model predict the model 78% correctly.

**Reference:**

Analysis of ordinal categorical data, Second addition, Alan Agresti, WILEY.

Fundamental of Mathematical Statistics, Goopta kapoor.

**Appendix:**

Questions for data:

1. In which course are you studying?

2.Your subject

3.Give your H.S percentage (10+2)

4.Which method is used by your teachers for online class?

5.Are online classes helpful for your online examination?

6.What do you want from online class?

7.What mode of exam do you prefer?

8.How many times did you take in per paper in online exam?

9.Which semester are you reading now?

10. SGPA obtained in semester 1 or semester 3 ( 𝐎𝐟𝐟𝐥𝐢𝐧𝐞 result)

11. SGPA obtained in semester 2 or semester 4.(latest 𝗢𝗻𝗹𝗶𝗻𝗲 result)

12. When do you communicate with your teacher?

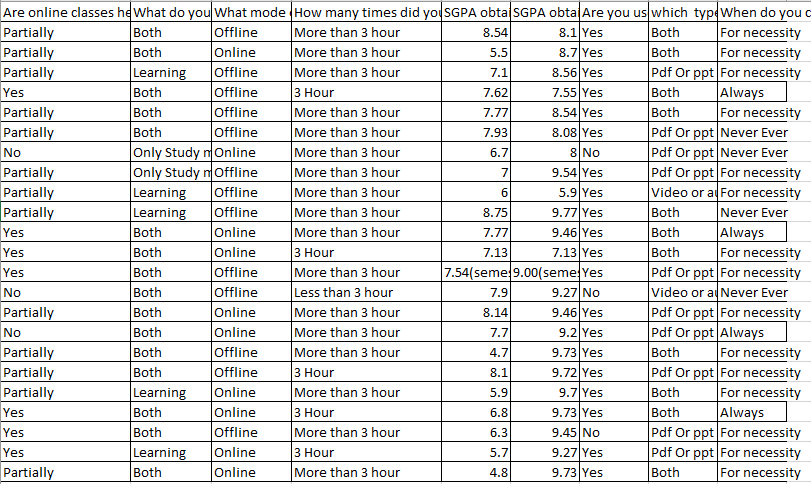
13. Online class attendance in percentage.

14. When there was offline class then your offline class attendance percentage

15. How much time do you spend on smartphone on an average in a day other than online class?

16. Do you understand the concepts that are being taught in online class?

**For modeling a sample transform data is**

****

**After converting into dummy variable**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| y | x1 | x2.A | x2.FN | x2.N | x3 |
| LESS CHANGE | 1 | 1 | 0 | 0 | 2 |
| HUGE CHANGE | 2 | 0 | 1 | 0 | 1 |
| MODERATE CHANGE | 2 | 0 | 1 | 0 | 1 |
| HUGE CHANGE | 2 | 0 | 1 | 0 | 1 |
| LESS CHANGE | 1 | 1 | 0 | 0 | 2 |
| LESS CHANGE | 1 | 1 | 0 | 0 | 2 |
| HUGE CHANGE | 2 | 0 | 1 | 0 | 1 |
| LESS CHANGE | 2 | 1 | 0 | 0 | 1 |
| HUGE CHANGE | 2 | 0 | 1 | 0 | 1 |
| MODERATE CHANGE | 2 | 1 | 0 | 0 | 1 |
| HUGE CHANGE | 2 | 0 | 1 | 0 | 1 |
| NO CHANGE | 1 | 1 | 0 | 0 | 2 |
| MODERATE CHANGE | 2 | 1 | 0 | 0 | 1 |
| HUGE CHANGE | 1 | 0 | 0 | 1 | 2 |
| LESS CHANGE | 1 | 1 | 0 | 0 | 2 |
| LESS CHANGE | 1 | 1 | 0 | 0 | 2 |
| MODERATE CHANGE | 1 | 0 | 0 | 1 | 2 |
| HUGE CHANGE | 1 | 0 | 1 | 0 | 2 |
| HUGE CHANGE | 2 | 0 | 1 | 0 | 1 |
| HUGE CHANGE | 2 | 0 | 1 | 0 | 1 |
| NO CHANGE | 2 | 1 | 0 | 0 | 1 |
| MODERATE CHANGE | 2 | 0 | 1 | 0 | 1 |
| MODERATE CHANGE | 2 | 0 | 1 | 0 | 1 |
| NO CHANGE | 1 | 0 | 1 | 0 | 2 |

Where

X1: student’s Exam time

X2.A: students are always communicate with teacher.

X2.FN: students are communicate with teacher for necessity .

X2.N: students are never communicate with teacher.

Multinomial logit model:

#multinomial logit model

library(nnet)

getwd()

setwd("C:/Users/Akash/Desktop")

df<- read.csv("me.csv",header=TRUE)

y<-as.factor(df$y)

#dividing the data into train data and test data

index<-sample(c(0,1), nrow(df), prob = c(0.8,0.2), replace = T)

train\_data<-df[index == 0,]

dim(train\_data)

test\_data<-df[index == 1,]

dim(test\_data)

#logistic model

model<-multinom(y~x1+x2.A+x2.FN+x2.N+x3,,order=TRUE,data=train\_data)

summary(model)

#p value

z<-summary(model)$coefficients/summary(model)$standard.errors

p<-(1-pnorm(abs(z),0,1))\*2

p

ynew<-predict(model,newdata=test\_data,type='class')

#confusion matrix

t<-table(test\_data$y,ynew)

sum(diag(t))/sum(t)