Fergusson College

(Autonomous), Pune

TYBSc Statistics Project

2023

NOMOPHOBIA

A Comparative Approach Through Statistical Learning

" The fear of being without or losing your mobile phone"

“HUMANS ARE THE REPRODUCTIVE ORGANS OF TECHNOLOGY”

-KEVIN KELLY

Deccan Education Society's

Fergusson College (Autonomous), Pune

Department of Statistics

STS3609: Statistics Practical III

CERTIFICATE

This is to certify that Mr./Ms. ____

Roll No.___________, has satisfactorily completed the project work entitled "**NOMOPHOBIA: A Comparative Approach Through Statistical Learning**" towards the partial fulfilment of B.Sc. (Statistics), Semester VI during the academic year 2022-23.

Place: Fergusson College, Pune

Date: / / 2023

Prof. Deepa Kulkarni
(Project Guide)

Prof. Subhash S. Shende
(Head of Department)

Internal Examiner

External Examiner

GROUP MEMBERS

SR.NO	NAMES	ROLL NO.
1.	DNYANESHWAR PAWAR	213115
2.	DHRUV SURYAVANSHI	213137
3.	GAYATRI PATIL	213276
4.	RITESH PATIL	213164
5.	ONKAR BANGAR	213250
6.	SHRAVANI TULASKAR	213139

Project Guide: Prof. Deepa Kulkarni

INDEX

ACKNOWLEDGEMENT

We wholeheartedly thank our project mentor, **Prof. Deepa Kulkarni** for providing valuable guidance and support in all stages of our project work, her meticulous insights helped us to improve the quality of the project we have been making. Our project greatly benefited from her statistical knowledge and experience which were useful at the critical stages of the project. This project could not have been possible without her.

We would like to show our gratitude towards the Head of the Department of Statistics, **Dr Subhash Shende** for providing us with this opportunity to work on this project. We extend our heartfelt gratitude to all the faculty members for their support, suggestions, and encouragement.

Finally, we extend our heartfelt gratitude to all the people who directly or indirectly are involved with our project.

MOTIVATION

In the past couple of decades, the usage of mobile phones has seen a dramatic increase. Although the developments in technology have made our lives and activities easier, it is believed that problematic and excessive use of technology could have some negative effects on people. One of these negative effects is Nomophobia. Nomophobic tendencies can change individuals' daily habits. Negative emotions due to Nomophobic tendencies like fear and anxiety especially in young people are thought to affect their school lives and academic achievements while in other sectors it can affect one's life in various aspects. The purpose of this project is to analyse the prevalence of Nomophobia among individuals and to spread awareness about it.

Nomophobia is very prevalent in young adult men and women, and it is closely associated with insomnia, regardless of mobile phone screen size, suggesting that mobile phone screen size should not be used as a proxy for hazardous usage. The heightened level of nervousness, fear, or anxiety crystallizes with the inability to communicate instantly. The condition has been linked to various mental health problems and poor functioning.

In a technologically advancing world, it has become quite impossible to separate ourselves from the use of gadgets like smartphones. However, increased indulgence in the usage of smartphones without thinking of the effects it has on mental as well as physical health is dangerous. Against the backdrop of this scenario, we aim to study various factors that could contribute to the condition of Nomophobia. Analysis of such factors can help understand this condition in a better way and the findings can be put to use for designing interventions to control or prevent the occurrence of Nomophobia. Thus, through the application of Statistical techniques we intend to study the relationship between factors in the prevalence of Nomophobia.

ABSTRACT

In the era of science and technology as far as communication is concerned it is a very important and trending field of research. Since the period of telegram to Long Term Evolution (LTE) and IP-based smartphone, the pattern of growth is very interesting, and it implicates the amazing transformation in human life. Even though innovation in communication made human life very easy, fast, and productive but on the other hand it's excessive mining can be a reason for deteriorating human health. Although numerous studies have examined factors that influence smartphone addiction, few have analysed the potential protective factors inherent in an individual that may benefit future intervention programs for smartphone addiction. In this project, our concern is the usability of smartphones and their future consequences. In this project, we are going to analyse the presence or absence of Nomophobia in individuals. We did an online questionnaire-based survey, which include sexes, education, and usage of mobile phones. The questionnaire set to calculate the Nomophobic score contains **18** different questions on mobile usage patterns. Based on the responses from questionnaires we assigned each answer a specific point according to Likert Scale (0 = less concerned and 4 =very much concerned). Nomophobia is considered a modern-age disorder introduced to our lives as a by-product of the interaction between people and mobile information and communication technologies, especially smartphones.

Keywords: Nomophobia, Questionnaire, smartphone, Addiction, Phobia, Fear, Anxiety.

INTRODUCTION

Mobile phones are invented and introduced to make human life easier, but if the same mobile phone becomes the reason for the deterioration of human health, then definitely it is not a good signal. In a study by the Boston Medical College on some families during having a meal in restaurants in 2004, it was found that one-third of family members were busy on a mobile phone during having a meal. Accessing information using a mobile phone gives a pleasant feeling as having good food or earning good money, this is found in the research done at Harvard University. Scientists found that if a child uses a mobile phone in excess, then his Insula or insular lobe, a part of the brain can be affected badly. *Various apps make our life easier, but it creates digital pollution if it's used in excess.* We should use technology but not be slaves of technology, so to reduce such pollution we should think of a mobile-free day just like a car-free day. Many people begin and end their days by checking their mobile phones, indicating that they are dependent on them.

Nomophobia (NO-MObile phone-PHOBIA) - The fear/anxiety of being away from a mobile phone is considered a disorder of the contemporary digital and virtual society that refers to discomfort, anxiety, nervousness, or anguish caused by being out of contact with a mobile phone.

Smartphone reliance can be said to fall under the umbrella of technology addiction. Such an addiction is present when there is the compulsive use of the technology leading to preoccupation, tolerance, unsuccessful efforts to control or stop using, withdrawals, loss of control, significant impairment, or neglect in any domain of life, lying to family members about the extent of involvement with the device and using the device as an escape or to relieve low mood. While cell phones offer a technologically advanced method of social interaction, the risk of becoming obsessed can hinder happiness. Most impressively, smartphones introduced the 'World of Apps' where applications can be got for almost anything and everything and thus shift a part of work to e-devices. In this way gradually and stealthily smartphones with their all-solutions-at-palm applications and technologies have crept into the human world. They have made users so helplessly dependent that one may be at a loss if the smartphone vanished. It is a disorder that millions of people suffer from around the globe; Nomophobia. The most affected are from 18-24 years of age. A typical Nomophobic individual can be identified by some characteristics such as never turning off the phone, obsessively checking missed texts and calls, bringing the phone everywhere, using phones at inappropriate times, and missing opportunities for face-to-face interaction while preferring over-the-phone contact. In some severe cases, people may also face physical side effects such as panic attacks, shortness of breath, trembling, sweating, accelerated heart rate, pain in the hand joints, neck and back pain, etc. when their phone dies or is otherwise unusable.

SYMPTOMS OF NOMOPHOBIA:

Nomophobia occurs in situations when an individual experiences anxiety due to the fear of not having access to a mobile phone. The "over-connection syndrome" occurs when mobile phone use reduces the number of face-to-face interactions thereby interfering significantly with an individual’s social and family interactions.

The term "techno-stress" is another way to describe an individual who avoids face-to-face interactions by engaging in isolation including psychological mood disorders such as depression.

Anxiety is provoked by several factors, such as the loss of a mobile phone, loss of reception, and a dead mobile phone battery.

In a few words:

Anxiety, Respiratory Alterations, Getting Angry, Hostile, Agitation, Disorientation, Stress, and Lonely Feeling.

OBJECTIVES

1. To analyse the presence or absence of Nomophobia in an individual.
   i.e., To detect Nomophobia in an individual through her/his responses.

2. To identify the impact of smartphone use on the academic performance of students.

3. To analyse which people are more prone to Nomophobia based on various demographic characteristics such as Area (urban / rural), Gender (male / female), Status (students / job professionals / business professionals / homemakers / senior citizens), Marital status (married / unmarried), Level of education (bachelors/masters / PhD) and the Field of education (science/commerce / arts / engineering/ medical).

4. To spread awareness in society about this harmful phobia.

5. To identify the relationship between Myopia w.r.t. Screen timing and Nomophobic score.

6. Make a detailed report of analysis throughout the whole study.

DATA COLLECTION AND METHODOLOGY

1. The collection of our data was based on a validated questionnaire which was issued by the Internal Journal of Applied Engineering Research ISSN 0973-4562 & also validated by the All India Institute of Medical Science (AIIMS), Delhi.

2. The data are collected through **GOOGLE FORM** and collecting the data through Google form was entirely an online process, we have circulated our questionnaire through various social media platforms and also by personally contacting friends and families.

3. The survey consisted of 18 questions that were designed to detect Nomophobia in an individual and other questions which are designed to take the information of participant background for various aspects such as age, gender, status, the field of study, educational qualification, the area of residence, marital status, screen time, etc.

4. After the process of data collection, the next step was to present the raw data concisely and simply, so that the data analysis would be easy and more efficient. In this step cleaning and sorting out the data takes place, and finally, we got the data of **582** individuals.

5. Now with this cleaned data we assigned a 5-point Likert Scale (as shown below) to calculate the Nomophobic score of the individuals by adding the score of all 18 questions, such that the highest score was 18*4=72 and the lowest score was 18*0=0. This gave us a general psychological idea about Nomophobic and Non-Nomophobic individuals.

6. Based on this score, the data was converted into a dichotomic scale of 0 & 1 (NO & YES) where the response of scores above 18*2.5=45 was taken as a YES (1) & those scores below 45 were taken as NO (0). (Note that the factor of dichotomic conversion i.e., 18*2.5=45 is assigned with the help of Psychiatry experts in the research paper we referred.)

Assigning a 5-point Likert scale:

Never	0
Rarely	1
Sometimes	2
Often	3
Always	4

SOFTWARES USED

R software

R is a programming language for statistical computing and graphics supported by the R Core Team and the R Foundation for Statistical Computing. Created by statisticians Ross Ihaka and Robert Gentleman, R is used among data miners and statisticians for data analysis and developing statistical software. R-Software was used to perform Exploratory Data Analysis, Regression analysis, and Testing of Hypotheses.

Excel

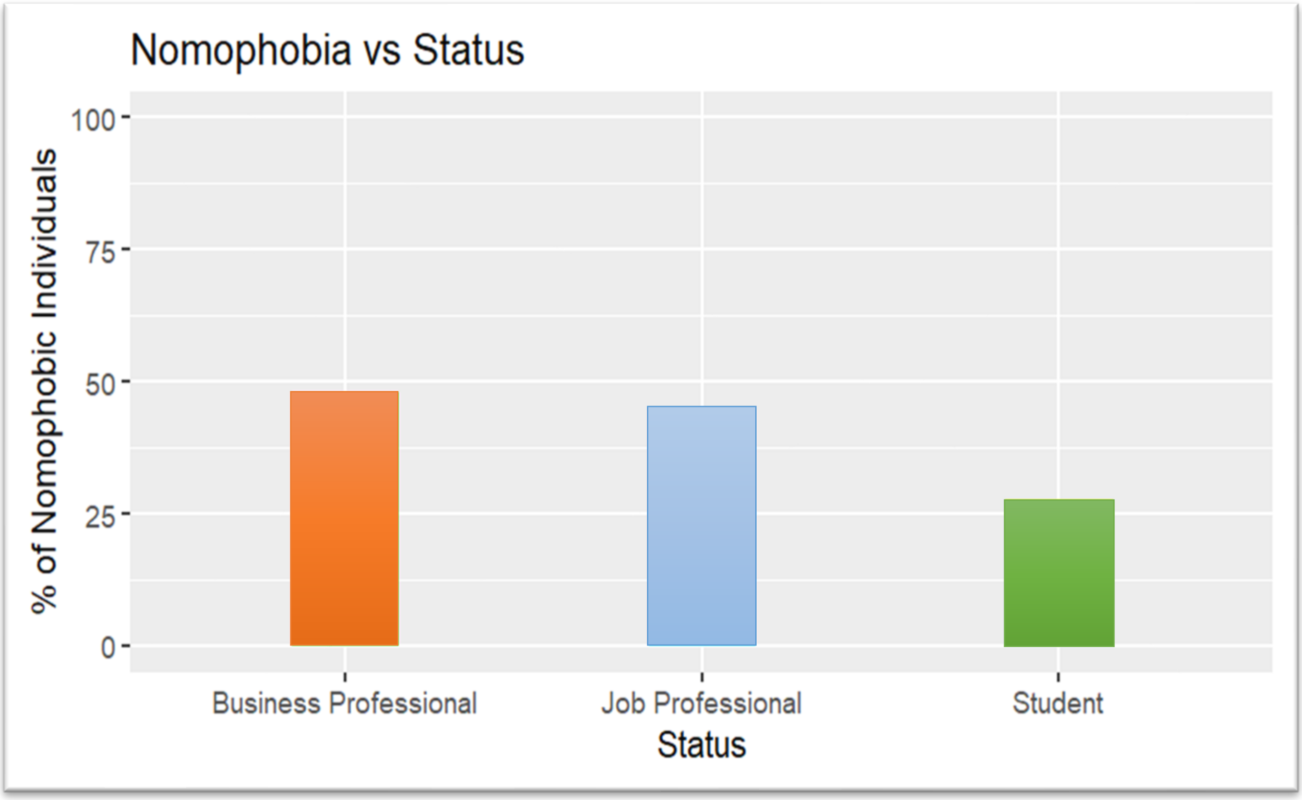
Excel is a spreadsheet program from Microsoft and a component of its Office product group for business applications. Microsoft Excel enables users to format, organize and calculate data in a spreadsheet. By organizing data using software like Excel, data analysts, and other users can make information easier to view as data is added or changed. Excel contains a large number of boxes called cells that are ordered in rows and columns. Data is placed in these cells. Excel was used mainly in sorting and filtering the data, and different functions from Excel such as *sort(), if()*, and *nested if()* were utilised.

R-Markdown

R Markdown is a flexible type of document that allows you to seamlessly combine executable R code, and its output, with text in a single document. These documents can be readily converted to multiple static and dynamic output formats, including PDF (.pdf), Word (.docx), and HTML (.html). The benefit of a well-prepared R Markdown document is full reproducibility. This also means that you can add more data to your analysis, and you will be able to recompile the report without making any changes to the actual document. The *rmarkdown* package comes pre-installed with RStudio, so no action is necessary.

EXPLORATORY DATA ANALYSIS

1) Working Status v/s Nomophobic Score :



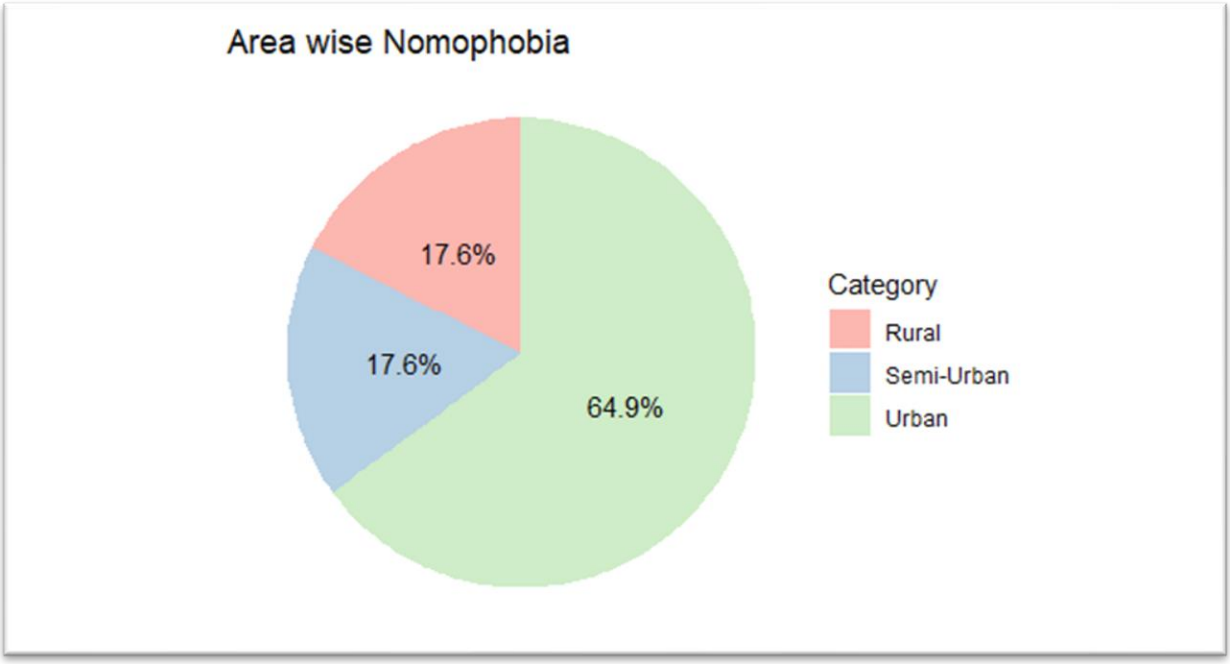
Inferences :

- The proportion of Nomophobic students is significantly lesser than other social statuses affected by Nomophobia.
- Business professionals are seen to have a higher proportion of Nomophobic individuals.

Conclusion :

- It is clear that Nomophobia is most common among students and is also prevalent among working professionals.
- Job and business professionals are more likely to use their mobiles for working purposes while students are mostly using mobiles for entertainment and study purposes.

2) Area of Residence v/s Nomophobia :



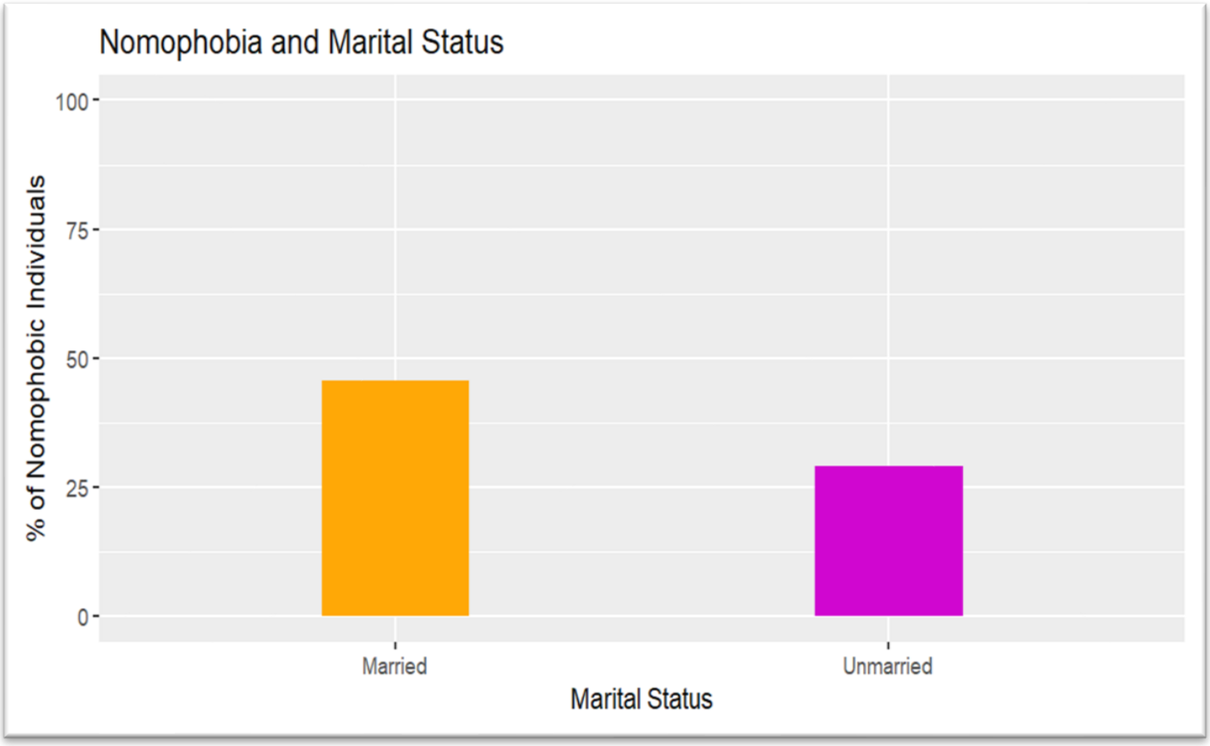
Inference :

- We find that approximately 65% of Nomophobic people live in metropolitan regions, while one-third of Nomophobic people are located in rural and semi-urban areas.

Conclusion :

- In the fast-paced urban environment, we observe more reliance on mobile phones for everyday chores such as ordering groceries and food.
- Even though the share of Nomophobic cases in rural and semi-urban areas is smaller, they are significant. Urbanization in suburban areas could be a major factor in the increasing mobile phone addiction.
- Nomophobic cases in a rural areas could be due to limited access to entertainment, online education/work, etc.

3) Marital Status v/s Nomophobia :



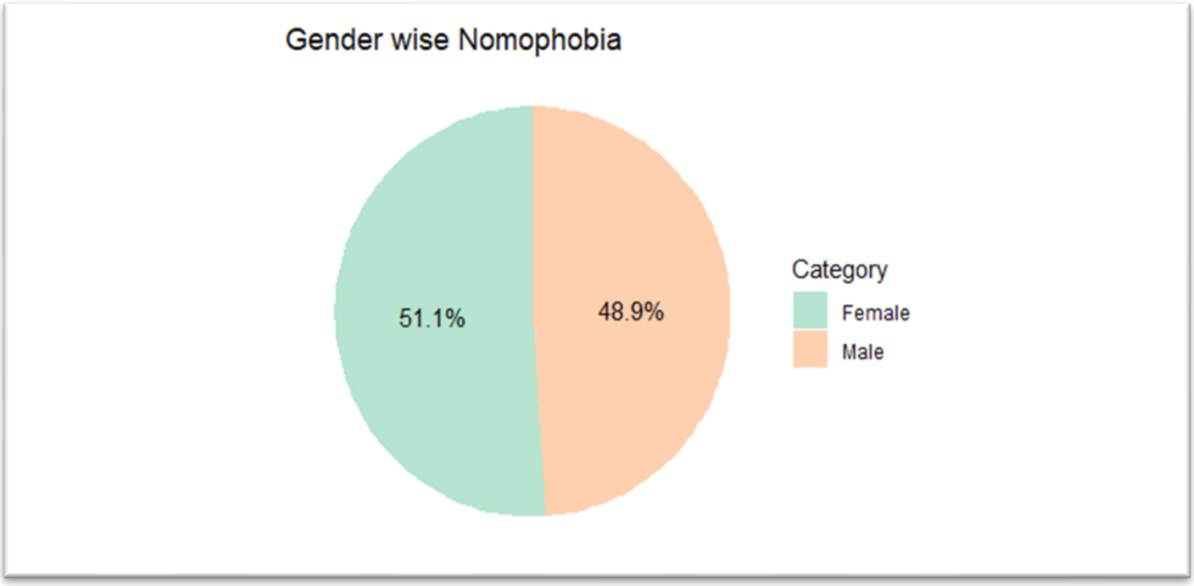
Inferences :

- We observe that almost 50% of married individuals are Nomophobic, while, more than a quarter of unmarried individuals are Nomophobic.

Conclusion :

- This indicates that married people, including business professionals and job professionals, are significantly more vulnerable to Nomophobia than unmarried individuals.
- Higher usage of mobiles by married folks could be attributed to their fascination with new and upcoming technology, work requirements, etc.

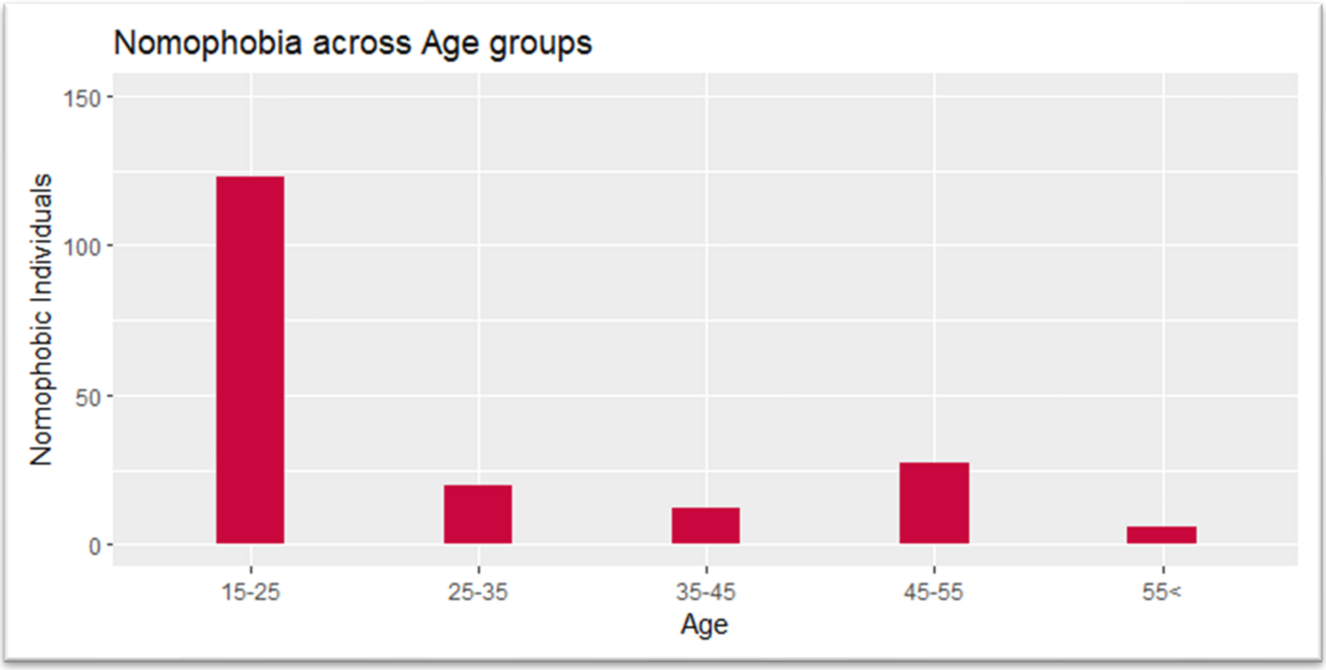
4) Gender–wise distribution of Nomophobic cases :



Conclusion :

- This pie chart demonstrates that Nomophobia affects both men and women almost equally, with slightly more cases in women.

5) Age-wise Segregation of Nomophobic Individuals :



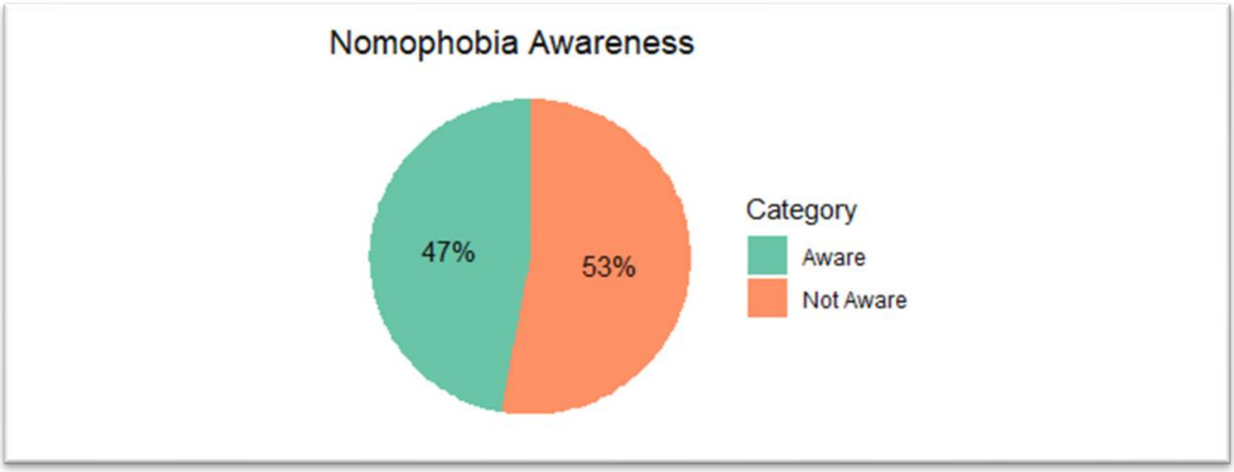
<u>Inferences</u> :

- The largest share of Nomophobic individuals is of the age group 15-25 which consists mainly of students and working professionals.
- The ages of 25-35 and 35-45 have shown merely 10-20 cases.
- The age group 45-55 has a significantly larger number of Nomophobic cases while the senior citizen age group has negligible cases.

<u>Conclusion</u> :

- The prevalence of Nomophobia among the smaller age groups is a topic of concern as this addiction leads to various mental and physical health issues such as anxiety, hypertension, stress, obesity, etc.
- Nomophobia in higher age groups could be associated with increased leisure time, using social media to connect with friends, and/or entertainment purposes.

6) Awareness among respondents :



Inferences :

- The awareness of Nomophobia is less among the respondents.
- Only 47% of people are aware of Nomophobia while 53% aren't aware of this.

Conclusion :

- Need to spread more awareness of Nomophobia in our society to make people cautious about this, because we all know that prevention is better than cure.

TESTING OF HYPOTHESIS

Before performing ANOVA i.e., for CRD on various treatments w.r.t. Nomophobic score and SGPA, we need to check the normality of the data.

To test

H_0: Nomophobic score is normally distributed

H_1: Nomophobic score is non-normal

Using Shapiro test in R to check the normality of the Nomophobic score the output is as follows:

```
library(readxl)

Nomo_Project <- read_excel("C:/Users/DELL/Desktop/NOMOPHOBIA PROJECT
/Nomo_Project.xlsx")
d=data.frame(Nomo_Project)
shapiro.test(d[,26])

##
##  Shapiro-Wilk normality test
##
## data:  d[, 26]
## W = 0.97198, p-value = 4.174e-09
```

Conclusion:

As, p-value < 0.05

Hence, we reject H_0 at 5% LOS.

i.e., Nomophobic Score is non-normal.

To test

H_0: SGPA score is normally distributed

H_1: SGPA score is non- normal

Using Shapiro test in R to check the normality of the SGPA score the output is as follows:

```
library(readxl)

Nomo_Project <- read_excel("C:/Users/DELL/Desktop/NOMOPHOBIA PROJECT
/Nomo_Project.xlsx")
d=data.frame(Nomo_Project)
shapiro.test(d[,30])

##
##  Shapiro-Wilk normality test
##
## data:  d[, 30]
## W = 0.76002, p-value = 2.2e-16
```

Conclusion:

As, p-value < 0.05

Hence, we reject H_0 at 5% LOS.

i.e., SGPA Score is non-normal.

Hence, as the data of the Nomophobic score and SGPA score both is non-normal so we can use the Kruskal Wallis test to compare group (treatment) medians.

KRUSKAL WALLIS TEST

This is a non-parametric test that is used when the assumption of normality is not true. This can be used for one-way analysis of variance for non-normal data i.e., CRD. It is based on Chi-Square distribution.

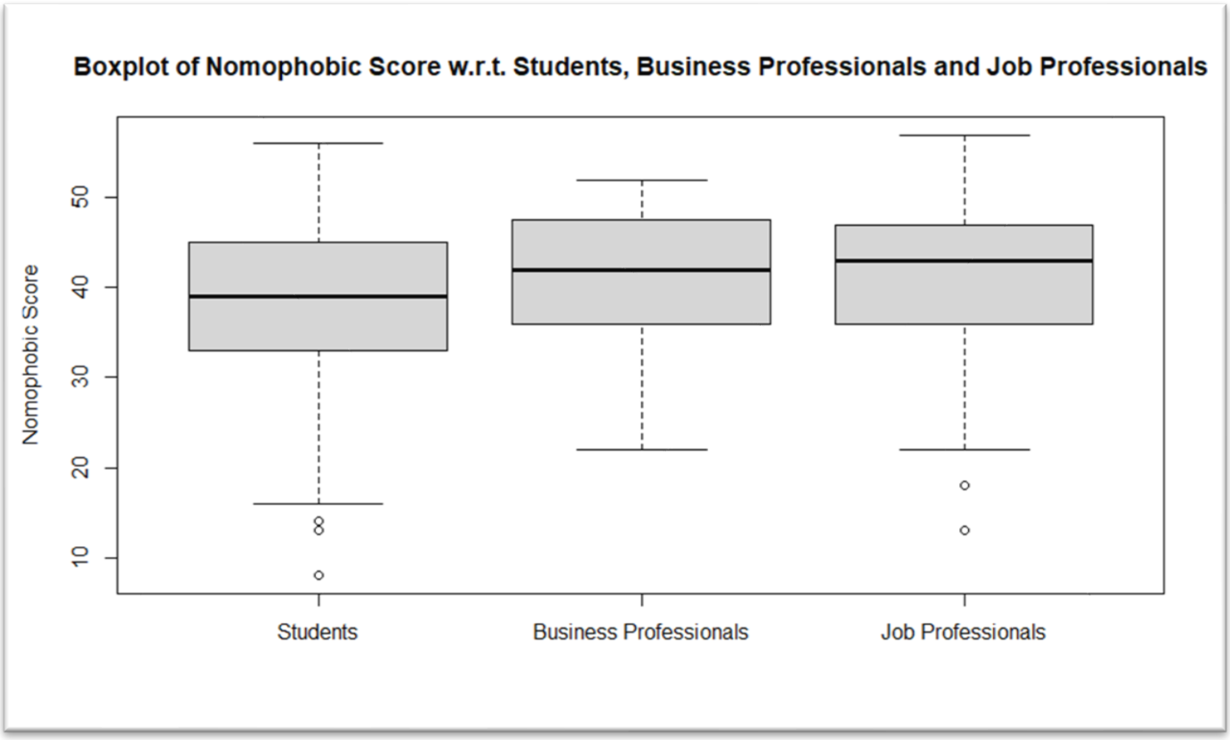
- ✓ Consider the following groups as treatments w.r.t. the Nomophobic score

Group 1: Students

Group 2: Business Professionals

Group 3: Job Professionals

We want to check whether the above groups have the same median Nomophobic score or not.



Conclusion:

The median Nomophobic score is higher in job professionals followed by business professionals and students respectively.

Now to confirm the above conclusion by using the Kruskal Wallis test.

That is to test the following-

H_o: Group medians are the same

H_1: Not all group medians are the same

```
	Kruskal-Wallis rank sum test

data:  list(y11, y22, y33)
Kruskal-Wallis chi-squared = 12.546, df = 2, p-value = 0.001886
```

```
qchisq(0.05,2,lower.tail = F)     #critical value
## [1] 5.991465
```

Conclusion:

We observe that, $\chi^2_{calc} > \chi^2_{critical}$

Hence, we reject H_0 at 5% LOS.

i.e., there is a significant difference in the Nomophobic score of students, business professionals, and job professionals.

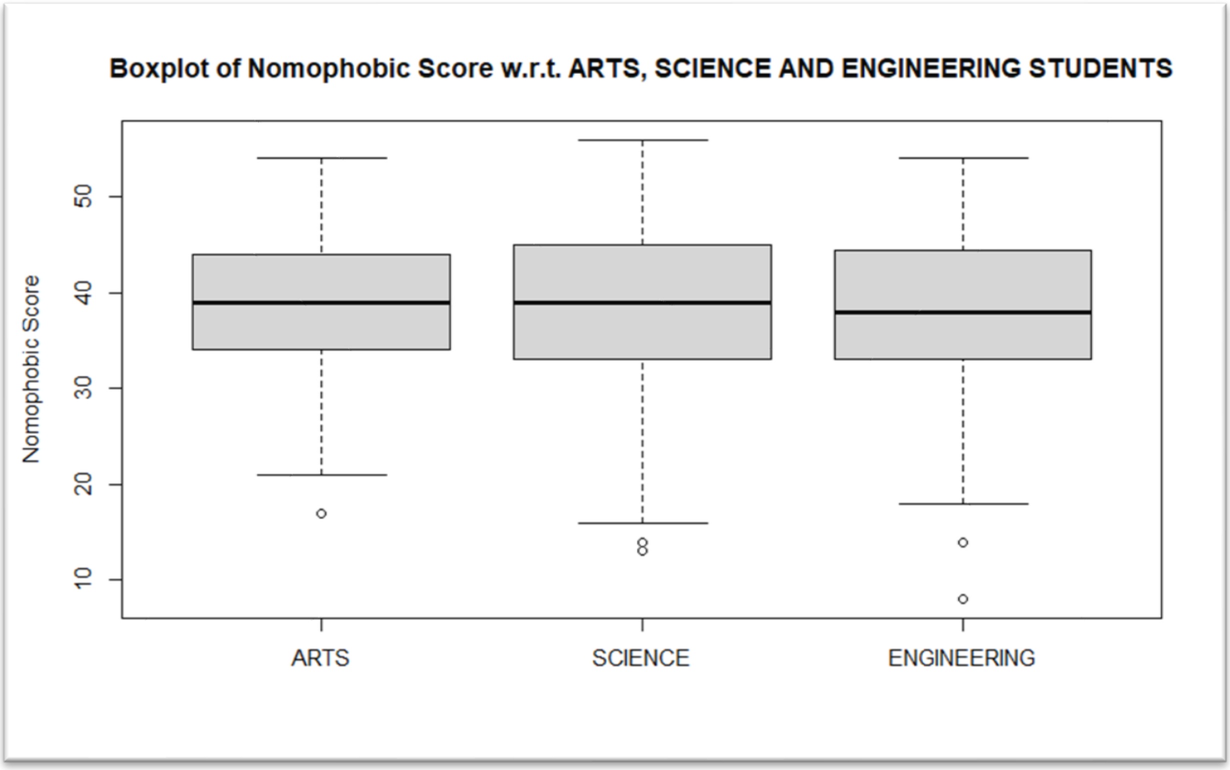
- ✓ Consider the following groups as treatments w.r.t. Nomophobic score

Group 1: Arts students

Group 2: Science students

Group 3: Engineering students

We want to check whether the above groups have the same median Nomophobic score or not.



Conclusion:

We can say that the median Nomophobic score is the same for all three groups which are Arts, Science, and Engineering students.

Now to confirm the above conclusion by using the Kruskal Wallis test.

That is to test the following-

H_o: Group medians are the same

H_1: Not all group medians are the same

```
	Kruskal-Wallis rank sum test

data:  list(z11, z22, z33)
Kruskal-Wallis chi-squared = 0.4782, df = 2, p-value = 0.7873
```

```
qchisq(0.05,2,lower.tail = F)     #critical value
## [1] 5.991465
```

Conclusion:

We observe that, $\chi^2_{calc} < \chi^2_{critical}$

Hence, we accept H_0 at 5% LOS.

i.e., there is no significant difference in the Nomophobic score of Arts, Science, and Engineering students.

MANN- WHITNEY U TEST

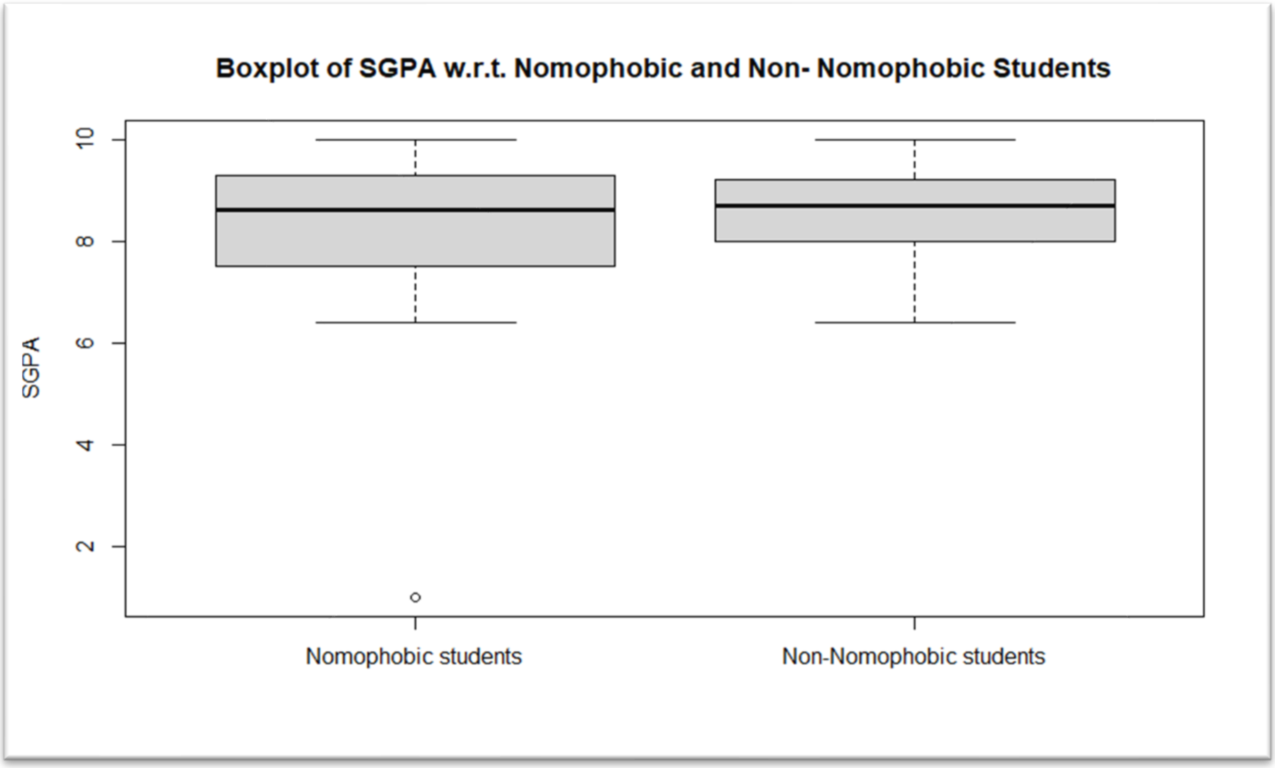
The Mann-Whitney U test, also known as the Wilcoxon rank-sum test, is a non-parametric statistical test used to compare the distributions of two independent samples. It is used to determine if there is a statistically significant difference between the medians of the two samples, without making assumptions about the shape or parameters of the underlying distributions.

- ✓ Consider the following groups w.r.t. the recent SEM SGPA score of students.

Group 1: Nomophobic students

Group 2: Non- Nomophobic students

We want to check whether non-Nomophobic students performed better than Nomophobic students in the recent semester exam or not.



Conclusion:

From the boxplot, we can say that the median SGPA score of both Nomophobic students and non-Nomophobic students is the same.

Now we must verify the above conclusion by confirmatory hypothesis testing (Mann-Whitney U test)

That is to test the following-

H_o: Group medians are the same

H_1: Group medians are not the same

```
> wilcox.test(x11,x22)

        Wilcoxon rank sum test with continuity correction

data:  x11 and x22
W = 5652, p-value = 0.4417
```

Conclusion:

We observe that, $p\text{-value} > 0.05$

Hence, we accept H_0 at 5% LOS.

i.e., there is no significant difference in the performance of Nomophobic and Non-Nomophobic students w.r.t. the recent semester SGPA score.

CHI-SQUARE TEST OF INDEPENDENCE

The Chi-sq. test of independence is a non-parametric test used to check the independence of the two attributes under study. It was first used by Karl Pearson in 1900. The test statistic of this test follows Chi-sq. distribution and hence it is called the Chi-sq. test for checking the independence of attributes.

Association between Area of Residence and Nomophobia

H_0: The attributes are independent

H_1: The attributes are not independent

Area\Phobia	Nomophobic	Non-Nomophobic
Urban	122	212
Rural	33	72
Semi-Urban/Rural	33	110

```
##
## Pearson's Chi-squared test
##
## data:  x1
## X-squared = 8.328, df = 2, p-value = 0.01555
```

```
qchisq(0.05,2,lower.tail = F)     #critical value

## [1] 5.991465
```

Conclusion:

We observe that, $\chi^2_{calc} > \chi^2_{critical}$

Hence, we reject H_0 at 5% LOS

i.e., attributes area of residence and Nomophobia are associated.

Association between Marital Status and Nomophobia

H_0: The attributes are independent

H_1: The attributes are not independent

Marital Status\Phobia	Nomophobic	Non-Nomophobic
Married	53	63
Unmarried	135	331

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  x2
## X-squared = 11.121, df = 1, p-value = 0.0008536
```

```
qchisq(0.05,1,lower.tail = F)     #critical value
```

```
## [1] 3.841459
```

Conclusion:

We observe that, $\chi^2_{calc} > \chi^2_{critical}$

Hence, we reject H_0 at 5% LOS

i.e., there is an association between marital status and Nomophobia.

Association between Gender and Nomophobia

H_0: The attributes are independent

H_1: The attributes are not independent

Gender\Phobia	Nomophobic	Non-Nomophobic
Male	92	192
Female	96	202

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  x3
## X-squared = 2.0514e-30, df = 1, p-value = 1
```

```
qchisq(0.05,1,lower.tail = F)     #critical value
```

```
## [1] 3.841459
```

Conclusion:

We observe that, $\chi^{2}_{calc} < \chi^{2}_{critical}$

Hence, we accept H_0 at 5% LOS

i.e., there is independence between gender and Nomophobia.

Association between Working Status and Nomophobia

H_0: The attributes are independent

H_1: The attributes are not independent

Status\Phobia	Nomophobic	Non-Nomophobic
Student	120	314
Job Professional	46	57
Business Professional	11	12
Homemaker	9	9
Senior Citizen	2	2

```
##
##  Pearson's Chi-squared test
##
## data:  x4
## X-squared = 17.175, df = 4, p-value = 0.001788
```

```
qchisq(0.05,4,lower.tail = F)     #critical value
```

```
## [1] 9.487729
```

Conclusion:

We observe that, $\chi^2_{calc} > \chi^2_{critical}$

Hence, we reject H_0 at 5% LOS

i.e., there is an association between working status and Nomophobia.

Association between Field of Study and Nomophobia

H_0: The attributes are independent

H_1: The attributes are not independent

Study Field\Phobia	Nomophobic	Non-Nomophobic
Arts	34	65
Commerce	19	21
Engineering	23	58
Medical	8	22
Science	97	217
Not Specialised	2	2
Other	5	9

```
##
## Pearson's Chi-squared test
##
## data:  x5
## X-squared = 6.3478, df = 6, p-value = 0.3854
```

```
qchisq(0.05,6,lower.tail = F)      #critical value
```

```
## [1] 12.59159
```

Conclusion:

We observe that, $\chi^2_{calc} < \chi^2_{critical}$

Hence, we accept H_0 at 5% LOS

i.e., there is independence between the field of study and Nomophobia.

Association between Educational Qualification and Nomophobia

H_0: The attributes are independent

H_1: The attributes are not independent

Educational Qualification\Phobia	Nomophobic	Non-Nomophobic
12^{th} and Below	13	34
Bachelors	138	300
Masters	33	55
PhD	4	5

```
##
##  Pearson's Chi-squared test
##
## data:  x6
## X-squared = 2.2839, df = 3, p-value = 0.5156
```

```
qchisq(0.05,3,lower.tail = F)     #critical value
```

```
## [1] 7.814728
```

Conclusion:

We observe that, $\chi^2_{calc} < \chi^2_{critical}$

Hence, we accept H_0 at 5% LOS

i.e., there is independence between educational qualification and Nomophobia.

PROPORTIONALITY TEST

Now as the data of the Nomophobic score is non-normal, hence we cannot apply the Z-test of proportionality. So, we need to use the Chi-Square proportionality test which is non-parametric.

The proportionality test based on Chi-Square Distribution is non-parametric and used to check whether the two populations under consideration differ significantly based on specific characteristics. It is used to test for the difference between two proportions for independent samples.

Test of proportions between Males and Females:

To test-

H_0: The proportion of Nomophobic individuals is the same in males and females

H_1: The proportion of Nomophobic individuals is not the same in males and females

```
#Between Male and Female
x2=c(92,96)
n2=c(284,298)
prop.test(x2,n2,alternative = "greater")
##  2-sample test for equality of proportions with continuity
correction
##
## data:  x2 out of n2
## X-squared = 3.0594e-30, df = 1, p-value = 0.5
## alternative hypothesis: greater
## 95 percent confidence interval:
##  -0.06378813  1.00000000
## sample estimates:
##    prop 1    prop 2
## 0.3239437 0.3221477
```

```
qchisq(0.05,1,lower.tail = F)     #critical value

## [1] 3.841459
```

Conclusion:

We observe that, $\chi^{2}_{calc} < \chi^{2}_{critical}$

Hence, we accept H_0 at 5% LOS.

i.e., The proportion of Nomophobia is almost the same in males (32.39%) and females (32.21%)

Test of proportions between Married and Unmarried:

To test-

H_0: The proportion of Nomophobic individuals is the same in married and unmarried

H_1: The proportion of Nomophobic individuals is not the same in married and unmarried

```
#Between Married and Unmarried
x3=c(53,135)
n3=c(116,466)
prop.test(x3,n3,alternative = "greater")


##
##  2-sample test for equality of proportions with continuity correc
tion##
## data:  x3 out of n3
## X-squared = 11.121, df = 1, p-value = 0.0004268
## alternative hypothesis: greater
## 95 percent confidence interval:
##  0.07825365 1.00000000
## sample estimates:
##    prop 1    prop 2
## 0.4568966 0.2896996

qchisq(0.05,1,lower.tail = F)     #critical value

## [1] 3.841459
```

Conclusion:

We observe that, $\chi^2_{calc} > \chi^2_{critical}$

Hence, we reject H_0 at 5% LOS.

i.e., the Proportion of Nomophobia is more in married (45.69%) compared to unmarried (28.97%)

Test of proportions between Students and Job Professionals:

To test-

H_0: The proportion of Nomophobic individuals is the same among students and job professionals

H_1: The proportion of Nomophobic individuals is not the same among students and job professionals

```
#Between Students and Job Professionals
x4=c(120,46)
n4=c(434,103)
prop.test(x4,n4,alternative = "less")

##
##  2-sample test for equality of proportions with continuity correc
tion
##
## data:  x4 out of n4
## X-squared = 10.496, df = 1, p-value = 0.9994
## alternative hypothesis: less
## 95 percent confidence interval:
##  -0.2640825  1.0000000
## sample estimates:
##    prop 1    prop 2
## 0.2764977 0.4466019

qchisq(0.05,1,lower.tail = F)      #critical value

## [1] 3.841459
```

Conclusion:

We observe that, $\chi^2_{calc} > \chi^2_{critical}$

Hence, we reject H_0 at 5% LOS.

i.e., the Proportion of Nomophobia is less in students (27.65%) compared to job professionals (44.66%)

Test of proportions between Urban, Rural and Semi-Urban/Rural Areas:

To test-

H_0: The proportion of Nomophobic individuals is the same in Urban, Rural and Semi-Urban/Semi-Rural Areas

H_1: The proportion of Nomophobic individuals is not the same in the areas of residence

```
#Between Urban, Rural and Semi-Urban/Semi-Rural
x1=c(122,33,33)
n1=c(334,105,143)
prop.test(x1,n1)

##
##  3-sample test for equality of proportions without continuity
##  correction
##
## data:  x1 out of n1
## X-squared = 8.328, df = 2, p-value = 0.01555
## alternative hypothesis: two.sided
## sample estimates:
##    prop 1    prop 2    prop 3
## 0.3652695 0.3142857 0.2307692

qchisq(0.05,2,lower.tail = F)     #critical value

## [1] 5.991465
```

Conclusion:

We observe that, $\chi^2_{calc} > \chi^2_{critical}$

Hence, we reject H_0 at 5% LOS.

i.e., the proportion of Nomophobic individuals is more in urban areas (36.53%) followed by rural areas (31.43%) and semi-urban (23.08%) respectively.

Test of proportions between Business Professionals and Job Professionals:

To test-

H_0: The proportion of Nomophobic individuals is the same among business professionals and job professionals

H_1: The proportion of Nomophobic individuals is not the same among business professionals and job professionals

```
#Between Business Professionals and Job Professionals
x5=c(11,46)
n5=c(23,103)
prop.test(x5,n5,alternative = "greater")

##
##  2-sample test for equality of proportions with continuity correc
tion
##
## data:  x5 out of n5
## X-squared = 0.0019474, df = 1, p-value = 0.4824
## alternative hypothesis: greater
## 95 percent confidence interval:
##  -0.1842609  1.0000000
## sample estimates:
##    prop 1    prop 2
## 0.4782609 0.4466019

qchisq(0.05,1,lower.tail = F)       #critical value

## [1] 3.841459
```

Conclusion:

We observe that, $\chi^2_{calc} < \chi^2_{critical}$

Hence, we accept H_0 at 5% LOS.

i.e., Proportion of Nomophobia is almost the same in business professionals (47.83%) and job professionals (44.66%)

Test of proportions between Arts, Science and Engineering Students:

To test-

H_0: The proportion of Nomophobic individuals is the same in Arts, Science and Engineering students

H_1: The proportion of Nomophobic individuals is not the same in Arts, Science and Engineering students

```
#Between Arts, Science and Engineering Students
x6=c(34,97,23)
n6=c(99,314,81)
prop.test(x6,n6,alternative = "greater")

##
##  3-sample test for equality of proportions without continuity
##  correction
##
## data:  x6 out of n6
## X-squared = 0.7667, df = 2, p-value = 0.6816
## alternative hypothesis: two.sided
## sample estimates:
##    prop 1    prop 2    prop 3
## 0.3434343 0.3089172 0.2839506

qchisq(0.05,2,lower.tail = F)

## [1] 5.991465
```

Conclusion:

We observe that, $\chi^2_{calc} < \chi^2_{critical}$

Hence, we accept H_0 at 5% LOS.

i.e., Proportion of Nomophobia is almost the same in Arts students(34.34%), Science students (30.89%) and Engineering students (28.40%)

Test of proportions between Bachelors and Masters students:

To test-

H_0: The proportion of Nomophobic individuals is the same in Bachelor and Masters

H_1: The proportion of Nomophobic individuals is not the same in Bachelor and Masters

```
#Between Bachelors and Masters
x8=c(138,33)
n8=c(438,88)
prop.test(x8,n8,alternative = "greater")

##
##  2-sample test for equality of proportions with continuity correc
tion
##
## data:  x8 out of n8
## X-squared = 0.94198, df = 1, p-value = 0.8341
## alternative hypothesis: greater
## 95 percent confidence interval:
##  -0.1591606  1.0000000
## sample estimates:
##    prop 1    prop 2
## 0.3150685 0.3750000

qchisq(0.05,1,lower.tail = F) #critical value

## [1] 3.841459
```

Conclusion:

We observe that, $\chi^2_{calc} < \chi^2_{critical}$

Hence, we accept H_0 at 5% LOS i.e., the proportion of Nomophobia is the same in Bachelors (31.51%) and Masters (37.50%)

Odds Ratio

Introduction

In statistics, odds are an expression of relative probabilities. Generally coated as odds in favour. The odds in favour of an event are the ratio of the probability that the event will happen to the probability event will not happen. Odds provide a measure of the likelihood of a particular outcome. Odds are calculated as the ratio of the number of events that produce the outcome to the number of events that do not produce the desired outcome. The odds ratio is the ratio of the odds of success between two groups. It is used to find the probability of an outcome of an event when there are two possible outcomes and there is one possible causal effect. The odds ratio is denoted by θ.

Control / Group	Yes	No
Group 1	Π1	1- Π1
Group 2	Π2	1- Π2

Formulae:

$$Odds\ of\ success = \frac{\text{No. of favourable outcome}}{No.of\ non-\ favourable\ outcomes}$$

$$= \frac{\pi}{1-\pi}$$

$$Odds\ ratio(\theta) = \frac{\text{Odds of success of Group 1}}{\text{Odds of success of Group 2}}$$

$$= \frac{\pi 1 * (1-\pi 2)}{\pi 2 * (1-\pi 1)}$$

Interpretation:

Odds ratio(θ)

Case 1: θ=1 => The event is equally likely in both groups.

Case 2: θ>1 => The event is more likely in the group of interest.

Case 3: θ<1 => The event is less likely in the group of interest.

Odds Ratio for Male and Female:

Gender\Phobia	Nomophobic	Non-Nomophobic	Total
Male	92	192	284
Female	96	202	298

Using the package *epitools* in R to calculate the odds ratio, the output is as follows:

```
oddsratio(x1)
```

```
## $data
##          Nomophobic Non Nomophobic Total
## Male             92            192   284
## Female           96            202   298
## Total           188            394   582
##
## $measure
## odds ratio with 95% C.I. estimate     lower    upper
##                     Male 1.000000        NA       NA
##                   Female 1.008293 0.7115593 1.428295
```

Conclusion:

The odds of Nomophobia were 1.0083 [with 95% CI (0.7116,1.4283)] times in males compared to females.
i.e., as the odds ratio is close to 1 which implies that the possibility of Nomophobia is equally likely in males and females.

Odds Ratio for Urban and Rural Areas:

Area\Phobia	Nomophobic	Non-Nomophobic	Total
Urban	122	212	334
Rural	33	72	105

```
oddsratio(x2)
```

```
## $data
##        Nomophobic Non Nomophobic Total
## Urban         122            212   334
## Rural          33             72   105
## Total         155            284   439
##
## $measure
## odds ratio with 95% C.I. estimate     lower    upper
##                     Urban 1.000000        NA       NA
##                     Rural 1.252477 0.7882199 2.021542
```

Conclusion:

The odds of Nomophobia were 1.2525 [with 95% CI (0.7882,2.0215)] times higher in those who are urban compared to rural.

i.e., Nomophobia is more likely in urban areas compared to rural areas.

Odds Ratio for Married and Unmarried:

Marital Status\Phobia	Nomophobic	Non- Nomophobic	Total
Married	53	63	116
Unmarried	135	331	466

```
oddsratio(x3)
```

```
## $data
##              Nomophobic Non Nomophobic Total
## Married              53             63   116
## Unmarried           135            331   466
## Total               188            394   582
##
## $measure
## odds ratio with 95% C.I. estimate    lower    upper
##                  Married 1.000000       NA       NA
##                Unmarried 2.060346 1.354934 3.127639
```

Conclusion:

The odds of Nomophobia were 2.0603 [with 95% CI (1.3149,3.1276)] times higher in those who are married compared to unmarried.

i.e., Nomophobia is more likely in married compared to unmarried.

Odds Ratio for Students and Job Professionals:

Status\Phobia	Nomophobic	Non- Nomophobic	Total
Students	120	314	434
Job Professionals	46	57	103

```
oddsratio(x4)
```

```
## $data
##                       Nomophobic Non Nomophobic Total
## Students                     120            314   434
## Job Professionals             46             57   103
## Total                        166            371   537
##
## $measure
## odds ratio with 95% C.I.  estimate     lower   upper
##        Students          1.0000000        NA      NA
##        Job Professionals 0.4741051 0.3045949 0.73998
```

Conclusion:

The odds of Nomophobia were 0.4741 [with 95% CI (0.3046,0.7399)] times in those who are students compared to job professionals.

In other words, the odds of Nomophobia are lowered by about 52.5895% in students compared to job professionals.

LOGISTIC REGRESSION

Introduction :

Logistic regression is used to predict the class (or category) of individuals based on one or multiple predictor variables (x). It is used to model a binary outcome, which is a variable, which can have only two possible values: 0 or 1, yes or no, diseased or non-diseased. Here we use Binary Logistic Regression Model- Used when the response is binary (i.e., it has two possible outcomes).

We are performing Logistic regression on the following response variable and regressor variables given below in the table:

Myopia	Y
Nomophobic Score	X1
Screen time	X2

The model for logistic regression is given as :

$Y = \pi(x) + \varepsilon$

where, $\pi(x) = \frac{e^{\beta 0+\beta 1*X1+\beta 2*X2}}{1+e^{\beta 0+\beta 1*X1+\beta 2*X2}}$ and $\varepsilon \sim B(\pi(x))$

Note that- β0, β1, and β2 are the regression coefficients.

Logistic regression belongs to a family, named Generalized Linear Model (GLM), developed for extending the linear regression model to other situations. Other synonyms are binary logistic regression, binomial logistic regression and logit model. Logistic regression does not return directly to the class of observations. It allows us to estimate the probability (p) of class membership. The probability will range between 0 and 1. You need to decide the threshold probability at which the category flips from one to the other. By default, this is set to p = 0.5, but in reality, it should be settled based on the analysis purpose.

Splitting data into training and testing datasets :

To start with model fitting we need to segregate data into 80% training and 20% test data-set. A logistic regression model is fitted on the training data set, and using it we can proceed to predict the values of the test data set and further calculate the accuracy of our model.

```
#splitting data into training and testing sets
set.seed(123)
sample=sample.split(data$Myopia,SplitRatio = 0.8)
train_data=subset(data,sample=="TRUE")
test_data=subset(data,sample=="FALSE")
```

Model Fitting (on training dataset) :

```
#fitting logistic model
model=glm(Myopia~.,data=train_data,family="binomial")
summary(model)

##
## Call:
## glm(formula = Myopia ~ ., family = "binomial", data = train_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0217  -0.9655  -0.9257   1.3933   1.5378
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -0.83488    0.53514  -1.560    0.119
## Nomophobic.Score  0.00884    0.01176   0.751    0.452
## Screen.Time      -0.01583    0.04304  -0.368    0.713
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 610.61  on 464  degrees of freedom
## Residual deviance: 609.78  on 462  degrees of freedom
## AIC: 615.78
##
## Number of Fisher Scoring iterations: 4
```

By fitting the model, we checked the significance of regressors.

As, (Null deviance – Residual deviance) < $\chi^2_{(2,0.05)}$

Hence, we may conclude that regressors are not significant at a 5% level of significance.

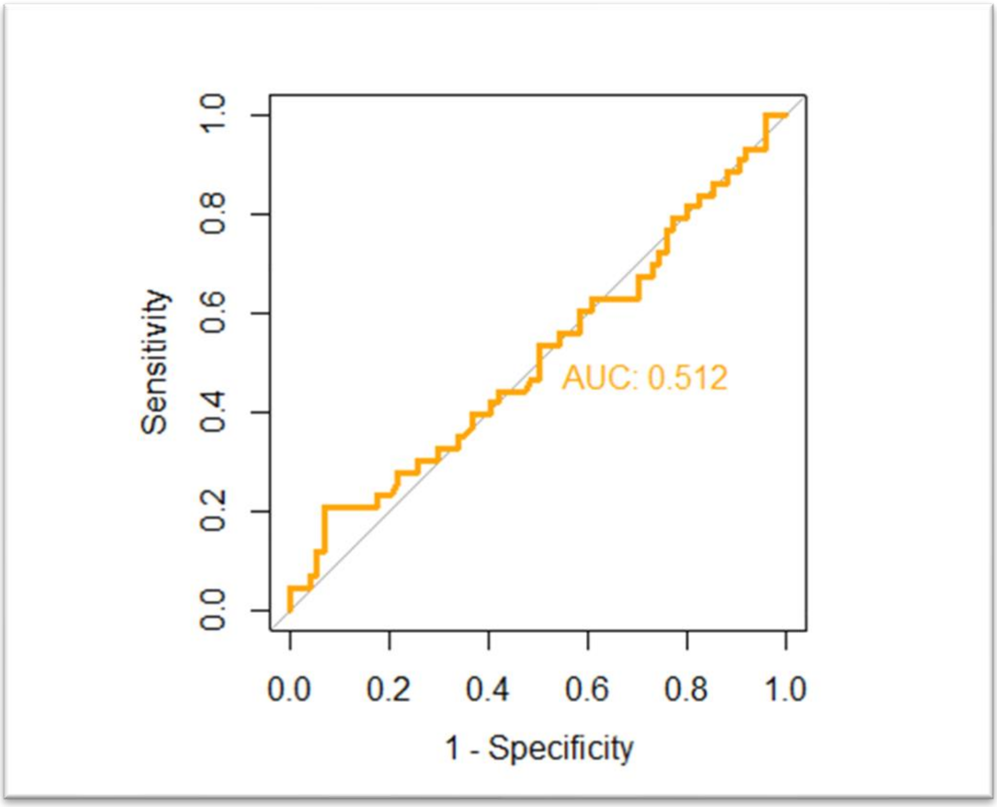
Now we will check the performance of our model by plotting the ROC curve :

ROC Curve: An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters:

- True Positive Rate (Sensitivity)
- False Positive Rate (1- Specificity)

```
#predicted probabilities of test data by using fitted model
predictdata=predict(object=model,newdata = test_data[-1],type="response")
```

```
#plotting ROC curve
par(pty="s")
info_glm=roc(test_data$Myopia,predictdata,plot=TRUE,
             print.auc=TRUE, print.auc.x=0.45,
             legacy.axes=TRUE, col="orange",lwd=3)
```



AUC: **AUC** stands for "Area under the ROC Curve". AUC provides an aggregate measure of performance across all possible classification thresholds.

AUC is an effective way to summarise the overall diagnostic accuracy of the test. It takes a value from 0 to 1, where a value of 0 indicates a perfectly inaccurate test and a value of 1 reflects a perfectly accurate test. A value of 0.5 for AUC indicates that the ROC curve will fall on the diagonal (i.e., 45-degree line) and hence suggest that the diagnostic test has no discriminatory ability.

The AUC value obtained by us is 0.512 which means that our model has a very slight discriminatory ability, which implies that our prediction model is a nearly random classifier which is not better for prediction.

Obtaining threshold probability value for classification:

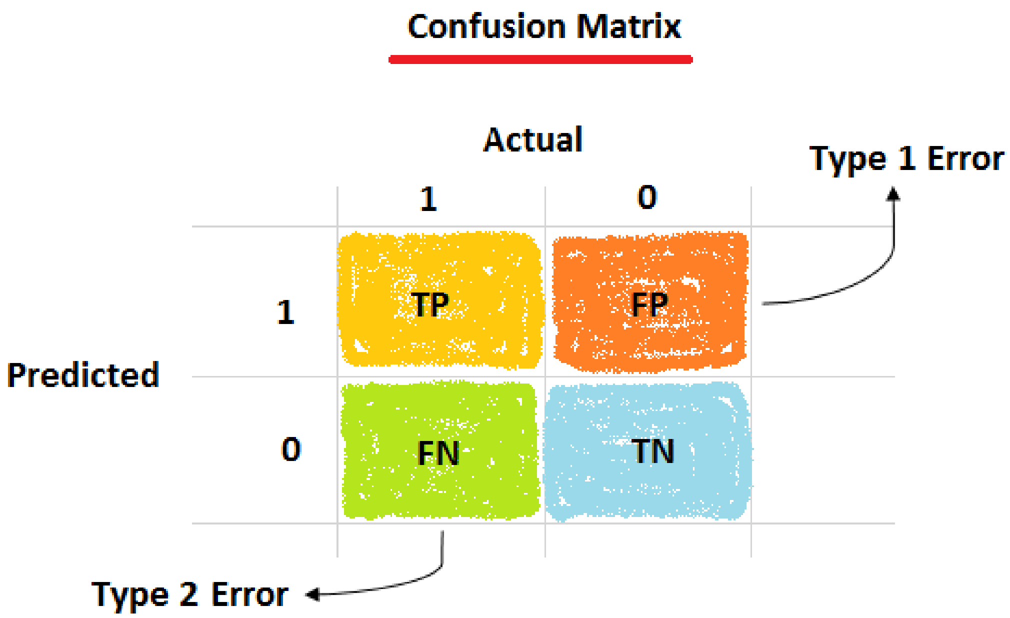
The classification threshold that minimises the distance between the point on the ROC curve i.e., (x=FPR,y=TPR) and the top left corner point i.e., (FPR=0, TPR=1) is the most preferred optimal threshold.

```
#getting TPR, FPR, Thresholds from ROC curve
tpr=info_glm$sensitivities
fpr=1-info_glm$specificities
thresholds=info_glm$thresholds

# Calculate the distance of each point to the top-left corner (Euclidean d
istance)
distances <- sqrt(fpr^2 + (tpr-1)^2)

# Find the index of the point with the shortest distance
optimal_index <- which.min(distances)

# Extract the optimal threshold from the corresponding point on the ROC cu
rve
optimal_threshold <- thresholds[optimal_index]

# Print the optimal threshold
cat("Optimal Threshold:", optimal_threshold, "\n")

## Optimal Threshold: 0.3657959
```

Hence, the optimal threshold for our model as a classifier is 0.3657959

Confusion matrix :

A confusion matrix is a table that is often used to describe the performance of a classification model (or “classifier”) on a set of test data for which the true values are known.



The following are some of the Terminologies of the Confusion Matrix :

1. True Positives (TP): These are cases in which we predicted yes, and they actually have Myopia.
2. True Negatives (TN): We predicted no, and they don’t actually have Myopia.
3. False Positives (FP): We predicted yes, but they don’t actually have Myopia. (Also known as a “Type I error.”)
4. False Negatives (FN): We predicted no, but they actually have Myopia. (Also known as a “Type II error.”)
5. Accuracy: Overall, how often the classifier is correct? (TP+TN)/(TP+FP+FN+TN)
6. Matthew’s Correlation Coefficient: $\frac{(TP\times TN)-(FP\times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FN)(FP+TN)}}$
   It is the correlation between the actual condition (actual class of dependent variable) and the predicted condition (predicted class of response variable).

```r
#to find confusion matrix and accuracy of our model
binary_predict=(ifelse(predictdata >0.3657959,1,0))
m=as.factor(binary_predict)
n=as.factor(test_data$Myopia)
confusionMatrix(data=m,reference=n)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 37 20
##          1 37 23
##
##                Accuracy : 0.5128
##                  95% CI : (0.4187, 0.6063)
##     No Information Rate : 0.6325
##     P-Value [Acc > NIR] : 0.99691
##
##                   Kappa : 0.0322
##
##  Mcnemar's Test P-Value : 0.03407
##
##             Sensitivity : 0.5000
##             Specificity : 0.5349
##          Pos Pred Value : 0.6491
##          Neg Pred Value : 0.3833
##              Prevalence : 0.6325
##          Detection Rate : 0.3162
##    Detection Prevalence : 0.4872
##       Balanced Accuracy : 0.5174
```

From the confusion matrix, it is clear that the accuracy of the model is 51.28% which is very less.

Further, we obtain the value of Matthew's Correlation Coefficient = 0.033648, which implies that our prediction model is an average random classifier which is not better for classification.

Why regressors are not significant?

Why our model is less appropriate?

Is there any violation of assumptions while model fitting?

Hence, we need to check the assumptions of the model.

Checking Randomness of Residuals

To test -

H_0 : Residuals are Random

H_1 : Residuals are not Random

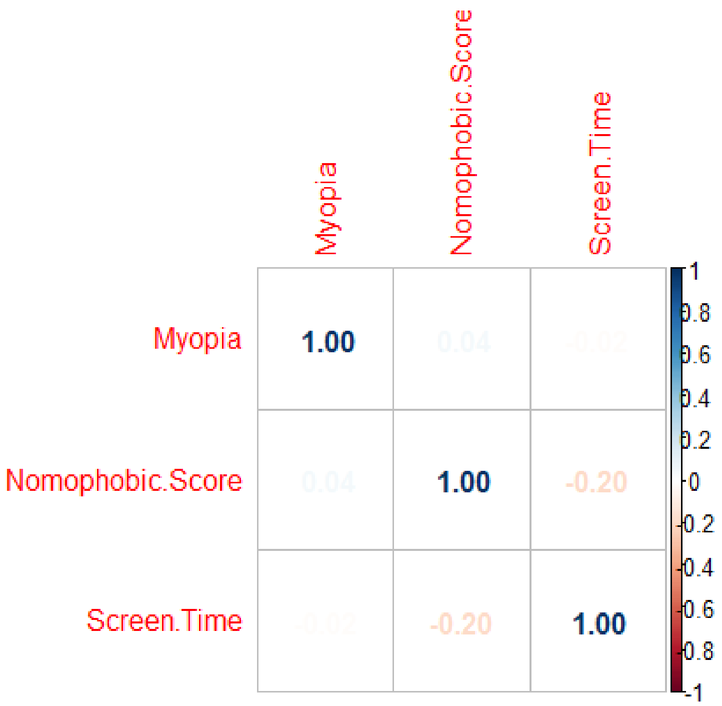
```
#to check residuals are randomly distributed
runs.test(model$residuals)

##
##  Runs Test
##
## data:  model$residuals
## statistic = -1.859, runs = 213, n1 = 232, n2 = 232, n = 464, p-value =
## 0.06303
## alternative hypothesis: nonrandomness
```

As, p-value > 0.05

Hence, the residuals are randomly distributed.

Checking Multicollinearity :

```
#for multicollinearity
corrplot(cor(data),method="number")
```



From the correlation matrix, we can conclude that there is no strong correlation between any pair of variables.

<u>Variance Inflation Factor (VIF) :</u>

```
#vif
vif(model)

## Nomophobic.Score      Screen.Time
##         1.033261         1.033261
```

Hence, as the VIF values are less than 5, thus we can say that there is an absence of multicollinearity in our data for model fitting.

Hence, the assumptions are also satisfied.

<u>Conclusion :</u>

Our binary logistic regression model cannot be used for predicting Myopia in individuals based on the Nomophobic score and Screen timing of individuals. We can use other prediction techniques such as ANN, KNN, Random Forest, SVM, etc., to make our prediction better. (Note that we can predict Hypermetropia rather than Myopia in a better way because hypermetropia is a vision problem in which we cannot see the nearest objects. And which may be caused by more screen timing and more Nomophobic score.)

CONCLUSIONS

1. Nomophobia is most prevalent among students, job professionals and business professionals.
2. Urban people are more prone to Nomophobia than semi-urban people and rural people. (Nomophobia is more likely in urban areas)
3. Married individuals are at more risk of Nomophobia than unmarried ones.
4. There is an equal probability of Nomophobia across gender. i.e., males and females are at an equal risk of Nomophobia.
5. The largest number of Nomophobic individuals are from the age group 15-25 which consists mainly of students and working professionals.
6. Business professionals are at higher risk of Nomophobia followed by job professionals and students.
7. There is no significant difference in the Nomophobic score of Arts, Science and Engineering students.
8. The academic performance of students is not affected by Nomophobia.
9. Area of residence, Marital status and Working status are associated with Nomophobia.
10. The proportion of Nomophobia is almost the same in the following groups- Business professionals and Job professionals, Arts Science and Engineering students, and Bachelor and Masters.
11. Myopia prediction by Nomophobic score and Screen timing based on a binary logistic regression model is not appropriate.

Limitations and Scope

- Due to a time constraint to our project, we had to limit our sample size to 582 which we are using for analysis is relatively small as compared to the population. Perhaps, due to the small amount of data, the results might be skewed.

- Nomophobia, in rare cases, is also associated with some brain tumours like glioma and meningioma which could be studied in detail.

- The results obtained, and the proposals made in the project need to be implemented on the ground to aid and make an important difference in the lives of individuals. For instance, we can assist people in overcoming this phobia with the aid of psychiatric professionals or by taking precautionary measures.

- The younger generation is particularly prone to insomnia, a sleep disorder that makes it difficult to fall or remain asleep and is primarily brought on by late-night mobile phone usage. The analysis of respondents' sleep quality, which could have provided considerably deeper insights, was not a part of the survey.

- A time series analysis could be done to show how the prevalence of Nomophobia among the younger generations has significantly increased over time.

- To acquire the best results, Nomophobia can be thoroughly researched among particular categories such as adolescents, working professionals, homemakers, etc. As we have a very small sample and also children aren't part of the survey, the results may not be appropriate for particular categories, which need to be studied in detail category-wise.

- The association between screen time and numerous issues such as lack of sleep, obesity, hypertension, etc., can be extensively investigated.

REFERENCES

- https://www.ripublication.com/ijaer18/ijaerv13n14_36.pdf
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6510111/#:~:text=The%20signs%20and%20symptoms%20are,very%20judicious%20regarding%20its%20diagnosis.
- http://www.sthda.com/english/wiki/kruskal-wallis-test-in-r
- https://www.rdocumentation.org/packages/questionr/versions/0.7.8/topics/odds.ratio
- https://medium.com/
- https://www.analyticsvidhya.com/
- https://statisticsbyjim.com/
- https://www.iguazio.com/glossary/classification-threshold/

ANNEXTURE: (Source R Codes)

```
#---------------------------------------------------------------------#
                          #Kruskal Wallis Test#
#---------------------------------------------------------------------#
#Between SGPA of Nomophobic and Non-Nomophobic
x1=d[,1]
x11=x1[!is.na(x1)]
x2=d[,2]
x22=x2[!is.na(x2)]
boxplot(x11,x22,names=c("Nomophobic students", "Non-Nomophobic
students"),ylab="SGPA",main="Boxplot of SGPA w.r.t. Nomophobic and Non-
Nomophobic Students")

kruskal.test(list(x111,x22))



#---------------------------------------------------------------------#
             #Chi-Sq test of Independence between attributes#
#---------------------------------------------------------------------#

x1=matrix(c(122,212,33,72,33,110),ncol=2,byrow=T) #b/w area and Nomophobia
chisq.test(x1)

x2=matrix(c(53,63,135,331),ncol=2,byrow=T) #b/w marital status and
Nomophobia
chisq.test(x2)
```

```
#----------------------------------------------------------------------#
                     #Chi-Sq test of Proportionality#
#----------------------------------------------------------------------#
#Between Urban, Rural and Semi-Urban/Rural
x1=c(122,33,33)
n1=c(334,105,143)
prop.test(x1,n1)
qchisq(0.05,2,lower.tail = F)

#Between Male and Female
x2=c(92,96)
n2=c(284,298)
prop.test(x2,n2,alternative = "greater")
qchisq(0.05,1,lower.tail = F)
```

```
#----------------------------------------------------------------------#
                              #Odds Ratios#
#----------------------------------------------------------------------#
library(epitools)

x1 <- matrix(c(92,192,96,202),ncol = 2,byrow=T)
colnames(x1) <- c('Nomophobic','Non-Nomophobic')
rownames(x1) <- c('Male','Female')
oddsratio(x1)

x2 <- matrix(c(122,212,33,72),ncol = 2,byrow=T)
colnames(x2) <- c('Nomophobic','Non-Nomophobic')
rownames(x2) <- c('Urban','Rural')
oddsratio(x2)
```