

# **A PROJECT REPORT ON INTERNET ACTIVITY OF STUDENTS**

**SUBJECT- STATISTICS**

**PAPER- STSA PAPER VIII PROJECT**

**REGISTRATION NUMBER- 1081711300218 of 2017**

**ROLL NUMBER- 3201307-24214**

# CONTENTS



Subjects	Page number
1. Acknowledgement	02
2. Summary of the project	02
3. Introduction	03-04
4. Questions Arises in Practical life	04-05
5. Objective of the Study	05
6. Data summary and Source of the Data	05-06
7. Computational Methodology	06
a) Descriptive Statistics	
b) Chi- Square test for independence	
c) ANOVA for Two way Classified Data	
8. Graphical Representation	06-11
a) Personal computer	
b) Age when Start Accessing Internet	
c) How often Student Access Internet	
d) Student's Use of internet	
e) Social Networking Sites	
f) Factors that influence to sign up for SN's	
g) Academics and the Internet	
h) Comparing Internet to the Library Use	
i) Time spent on Internet	
9. Test for independence or measure of association:	11-14
a) Dependency between the age students start to access the internet and their basic institution	
b) Dependency between the gender and the time spent accessing the internet.	
c) Association between the gender and the internet activities.	
d) Dependency between gender and how often student access the internet.	
e) Association between gender and student's reason for using SNs	
10. ANOVA Testing:	14-19
a) ANOVA testing for the age students start to access the internet and their basic institution.	
b) ANOVA testing for gender spent time on the internet.	
c) ANOVA testing for different internet activities.	
d) ANOVA testing for how often Student Access the internet.	
e) ANOVA testing for different reasons behind sign up on the SNs.	
11. Conclusion	19
12. Appendix	20-23

## **1. ACKNOWLEDGEMENT:**

The project has given to me is a golden opportunity for learning and self development through collaborative. I consider myself lucky to have worked under Sir Kiranmoy Chatterjee, Sir Soumyadeep Das & Sir Suryasis Chatterjee , whose encouraging words and monitoring individuals care and motivating approach had widened the horizon of my knowledge and stimulated me to work together with my classmates. I am grateful to them. I also want to give my special thanks to our honourable Head of the Department of Statistics Sir ARUP KUMAR HAIT. Without his guidance and persistent help this dissertation would not have been possible.

At last but not the least, I am grateful to all of my friends, family members, and many well wishers, who helped me in numerous ways in completing this project.

Roll number- 1081711300218 of 2017

Registration number- 3201307-24214

Year- 2020

## **2. SUMMARY OF THE PROJECT:**

To gain information, students should not only rely on textbooks provided and materials prepared by the lectures, but the skills in searching information by themselves from the use of internet should be emphasized. A mastery skill in using computer is the basis towards an effort in mastering communication and information skills. One will be considered as computer literate as they have the basic understanding towards computer functions, can use the computer to ease and complete their daily task either for personal usage or for profession efficiently and impressively and also be able to interact with computer by using applications such as internet. However, there is still the concern about the amount of time student spend on the internet and what really keeps them on the internet.

In light of this, the research study is design to ascertain the amount of time students spend on the internet and the activities students undertake on the internet. So we undergo with the ANOVA technique to conclude which are the factors depends on the internet usage. And we also use the chi- square technique to find out the factors is independent or not.

## **3. INTRODUCTION:**

In 21<sup>st</sup> century, students use the internet for almost everything, from communication to research to entertainment. Being a student who spends a lot of time on the internet, I often thought that the amount of time spent or the activities that students perform on the internet may somehow depend on various factors like student's gender, their pocket money etc.

The internet is a computer based global information system composed of many interconnected computer networks and each network may link tens, hundreds, or even thousands of computers, enabling them to share information and processing power. The Internet has made it possible for people all over the world to communicate with one another effectively and inexpensively. Many individuals use the Internet for communicating through electronic mail (e-mail), retrieving news, researching information, shopping, paying bills, banking, listening to music, watching movies, playing games, and even making telephone calls. Educational institutions use the Internet for research and to deliver online courses and

course material to students. The use of internet in educational environment has enabled easy access to many resources; information sharing has, therefore, significantly increased. Moreover, the prevalence of this sharing has brought additional benefits in that these resources can be used in any location and any time. In regard to academic work, the Internet has introduced considerable change for students it has opened up new academic possibilities for students such as online studies and also students and academicians who do scientific research and prepare projects prefer using the Internet because it is the easiest, fastest, and cheapest ways of accessing necessary information.

Among the variety of online tools now available for communication, are social networking sites (SNSs) such as face book, whatsapp and twitter, which are significant tools for connecting people throughout the world. Despite major productive uses of Internet technology in today's digital world, users prefer to spend much more time on social networking sites (SNSs) like Face book & Whatsapp.

Although technology is a very important, useful and indispensable part of the life, effective and appropriate usage of it still needs to be improved in education, most students sometimes, use the computer and Internet for reasons other than educational purposes. The effect of the Internet in education is, hence, relatively limited.

Internet use, especially in education, has been investigated for some time, and many different studies, exist in literature about that subject, the use of this resources by students in relation to their gender and also activities that students perform on the internet has been a major concern.

University students are heavy users of the Internet compared to the general population. Use of the Internet is a part of college students' daily routine, in part because they have grown up with computers. It is integrated into their daily communication habits and has become a technology as ordinary as the telephone or television.

Internet use is a staple of college students' educational experience. They use the Internet to communicate with professors and classmates, to do research, and to access library materials. For most college students the Internet is a functional tool, one that has greatly changed the way they interact with others and with information as they go about their studies. The college experience is not only about learning in the classroom, it is also about encountering new social situations and gaining new social skills. College students use the Internet nearly as much for social communication as they do for their education. But just as they use the Internet to supplement the formal parts of their education, they go online to enhance their social lives.

#### **4. QUESTIONS ARISES IN PRACTICAL LIFE**

This project attempt to answer the following questions.

- What is the most favourite internet activity among students?
- Are there any differences between male and female student's internet activities?
- Are there any difference in male users' preference and attitudes compared to female user's preference and attitude?
- Does the amount of time student spends on the internet depend on gender?
- Does student use of the internet depend on the institution they had their basic education?

- What are some of the factors that influence students to sign up for SNs?
- What are the effects of SNs on student's education?

## 5. OBJECTIVES OF THE STUDY

The objective of the project includes:

1. To determine students favourite activity when using internet.
2. To determine if there is any relation between gender and the amount of time students spend on the internet.
3. Compare student's internet use to library use.
4. To determine whether the age students start to access the internet depends on whether a student had his basic education in a public or private institution.
5. To determine factors that influence students to sign up for SNs.

## 6. SOURCE OF DATA & DATA SUMMARY:

**1. Source of data:** The purpose of this research is to determine the amount of time students spend and the activities they undertake on the internet relative to their gender. Secondary data will be used during the research process. This data was gathered from the websites given below. <https://www.statista.com/topics/2157/internet-usage-in-india/> <https://www.academia.edu/7827980>. I also used the journal published by Osam Sarfo Estephan, professor of the University of the Development Studies. Simple random sampling was used to ensure that the sample is not biased. For the purpose of their study University students were identified as the population of interest. The reason for this choice is that previous studies have revealed that University students are heavy users of internet and that they are likely to have computer and internet experience than other groups.

### 2. Data Summary:

**TABLE 01: SOCIO DEMOGRAPHIC CHARACTERISTICS**

As can be seen from the table 57 % represents males and 43 % represent females.

VARIABLE	NUMBER OF RESPONDENTS (%)
<b>Gender:</b>	
Male	57
Female	43
<b>Age range:</b>	
15-20	12
21-25	71
26-30	17
Above 30	---
<b>Internet use:</b>	
Entertainment	25
News	3
Research	49
Communication	23
<b>Department:</b>	
Physics	7
Chemistry	13

Biology	41
Mathematics	34
Earth & Environmental science	5

The full data set is given below in the appendix section.

**7. COMPUTATIONAL METHODOLOGY:** The researcher is interested in answering questions about students internet use in relation to their differences. The following statistical tools will be use for analysis and interpret the data obtain.

**A) EXPLORATORY DATA ANALYSIS:**

This statistical method can be used to summarize data and the most familiar of this method is the finding of the average. Average means the centre of distribution or the most typical case. Measures of average are also called measures of central tendency and include the mean, median, mode and midrange. It helps to gain useful information from raw data by organizing them into frequency distribution and then presenting the data by using various graphs.

**B) CHI-SQUARE TEST FOR INDEPENDENCE:**

The chi-square statistics is used to determine whether a frequency distribution follows a specific pattern and it can also be used to show preference for a specific style. The test of independency of variables is used to determine whether two variables are independent of or related to each other or to test if there is a significant association between two categorical variables when a single sample is selected.

*Assumptions for Chi- Square test:*

1. Large expected frequencies. The chi-square test is based on an approximation that works best when the expected frequencies are fairly large. No expected frequency should be less than 1 and no more than 20% of the expected frequencies should be less than 5.
2. Mutually exclusive row and column variable categories that include all observations. The chi-square test of association cannot be conducted when categories overlap or do not include all of the observations.
3. Independent observations. A critical assumption for chi-square is independence of observations. One person's response should tell us nothing about another person's response. Observations are independent if the sampling of one observation does not affect the choice of the second observation.

**C) ANOVA FOR TWO WAY CLASSIFIED DATA:**

The ANOVA consists in the estimation of the amount of the variation due to each of the independent factors(causes) separately and then comparing these estimates due to assignable factors(causes) with the estimate due to chance factor(causes), the latter being known as the experimental error or simply error.

*Assumptions for ANOVA:*

ANOVA test is based on the test statistics F (or Variance Ratio).For the validity of the F-test in ANOVA, the following assumptions are made.

1. The observations are independent.
2. Parent population from which the observations are taken is normal, and
3. Various treatment and the environmental effects are additive in nature.

**8. GRAPHICAL REPRESENTATION:**

**A) PERSONAL COMPUTER**

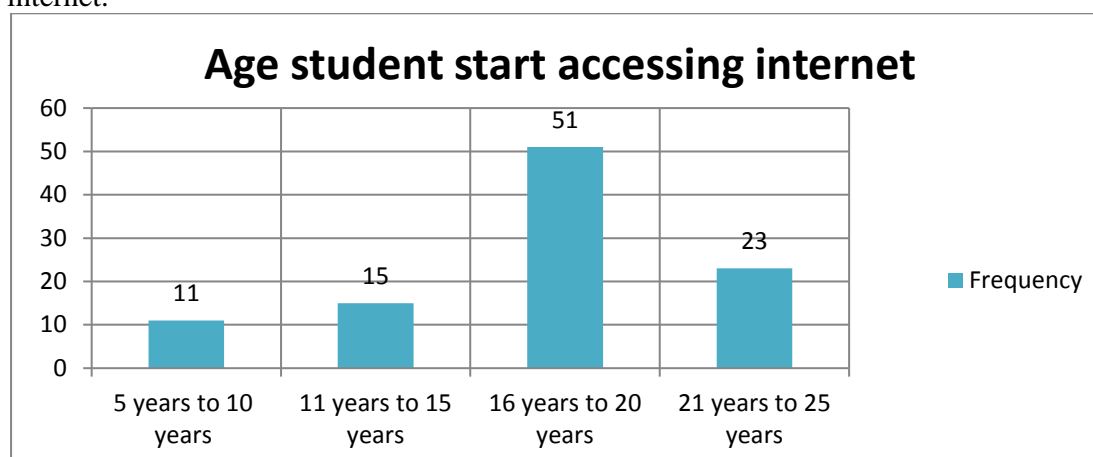
	NUMBER OF RESPONDANTS (%)
YES	53
NO	47

TOTAL 100

Students were asked whether they own a personal computer or not, and out of the 100 students interviewed 53% responded yes while 47% responded no. Most of the students, who responded yes, use laptops more than the desktop computers and some of the reasons being that you can easily take it anywhere especially in the internet coverage areas to access the internet. The table above gives the number of respondents who responded to the question.

#### B) AGE STUDENT START ACCESSING INTERNET

The bar chart below represents the age at which students started using the internet. The figure below illustrates the age at which respondents started using the internet, with the age range 16-20 having the maximum of 51% while the range 5-10 is having the minimum of 11% indicating that students do not get exposed to the internet early, also the range from 21-25 is having 23% indicating that most students also get to the university before they start accessing the internet.



#### C) HOW OFTEN STUDENTS ACCESS THE INTERNET:

Most students access the internet on a daily basis while others like to access the internet during weekends, but some students also access the internet when they have a project or an assignment to work on.

#### NUMBER OF RESPONDENTS (%)

Daily	41
Weekends	34
Other	25

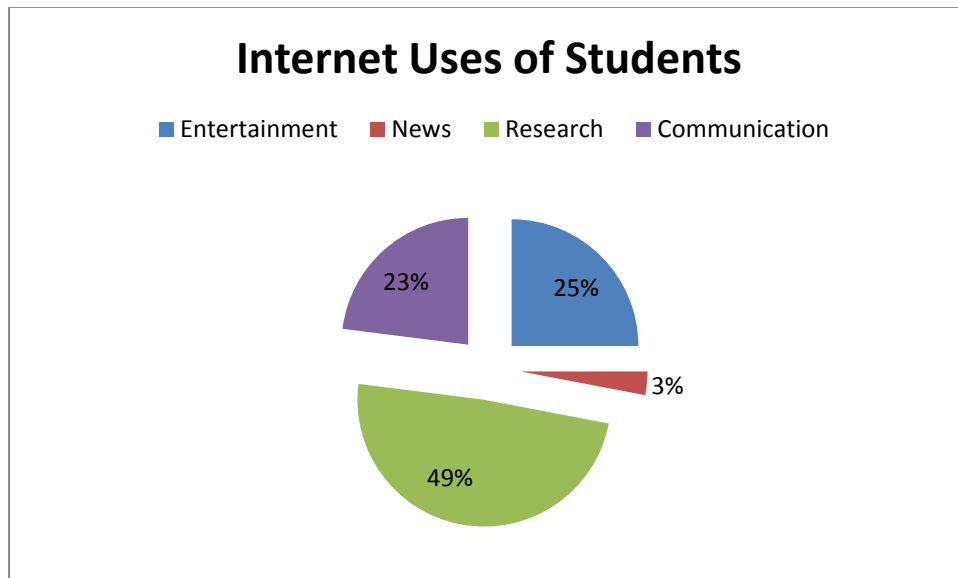
#### D) STUDENT USE OF INTERNET:

University students are heavy users of the internet, most often students use the internet for research work or working on assignment, also activities like downloading games, music and videos as well as chatting and using instant messaging are some of the activities students undertake when they go online.

The table below gives the activities that students undertake when they go online.

Internet use	NUMBER OF RESPONDENTS (%)
Entertainment	25
News	3

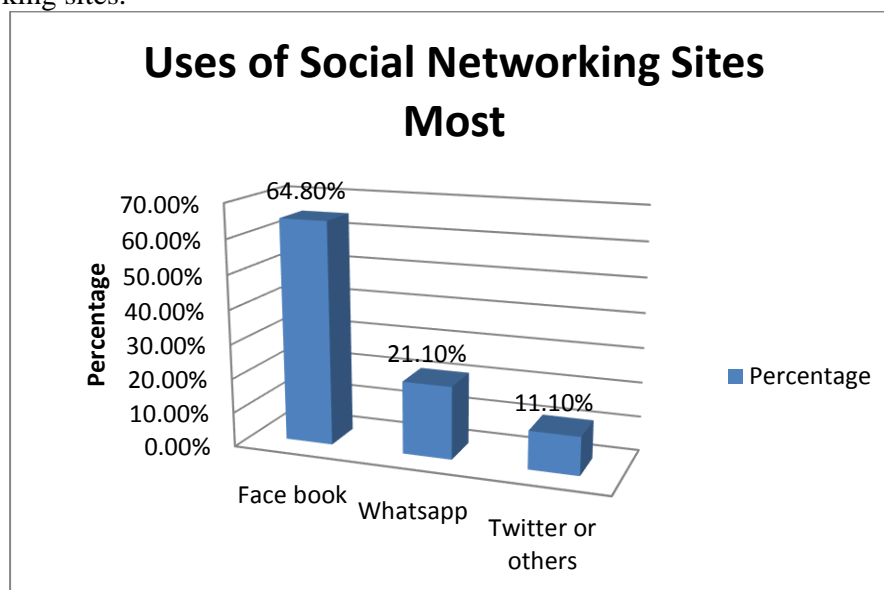
Research	49
Communication	23



#### E) SOCIAL NETWORKING SITES (SNS):

SNSs are virtual spaces that allow individuals to create personal profiles - visible to other users- to establish connections and join an online social network. These websites enable computer-mediated communication (CMC) between people. Social networking has been the tool that brings people together in just a click of a mouse.

Face book a social networking site established in February 2004 has gain much popularity among University students, the survey shows that majority of students signing up for a social networking site will sign up for face book the reason being that they will find most of their friends on that site. The diagram below illustrates the number of students on various social Networking sites.



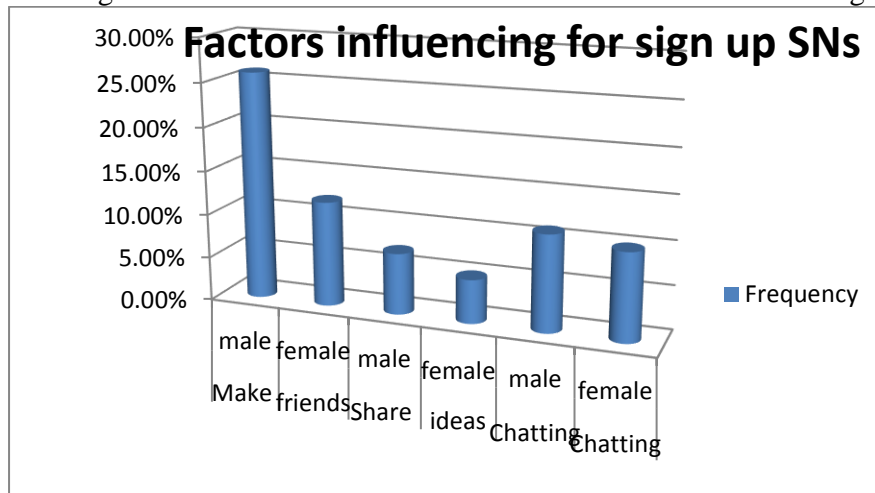
The figure above shows that face book has the maximum number of people i.e. 46 respondents represent 64.8%, whiles Twitter also had 15 respondents representing 21.1%. Almost all users login at least two times a day, in addition friendship was identified as the most favourite motive among all users as well as chatting with friends and also hooking up with old pals.



#### F) FACTORS THAT INFLUENCE STUDENTS TO SIGN UP FOR SNS:

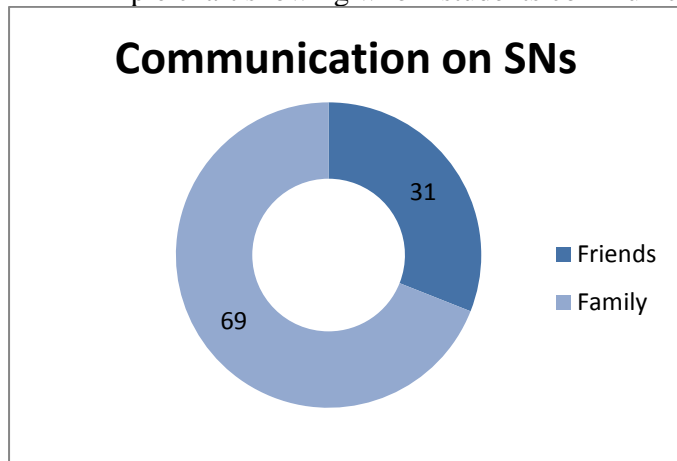
Despite major productive use of the internet most students prefer to spend much time on Social Networking sites (SNS) such as face book. Several factors were identified as the reasons why students sign up for SNS such as chatting, making new friends and sharing ideas.

The diagram below illustrates the factors that influence students to sign up for SNS.



From the diagram it is obvious that making new friends having a maximum of 38 respondents representing 53.5% is the major factor that influence student to sign up for SNS, other factors identified include chatting, sharing of ideas and other information and also keeping in touch with friends all the time. Again when students were asked with whom they communicate most on the Internet, (69%) of the students answered “friends,” and much smaller proportions (31%) of students use most for communication with family members.

A pie chart showing whom students communicate with the most.



The diagram above shows that most students will communicate with their friends rather than family members which is no different from their motive for signing up for SNS.

It is clear, however, that the Internet is not the only means of communication for university students, and in some cases it is not even their first choice for it. Most students said they are more likely to use the phone than the Internet to communicate socially, even though most students consider the Internet to be an easy and convenient choice for communicating with friends. The widespread use of cell phones among college students and the telephone's immediacy make it a primary choice for students' social communication. University students' experience with the phone also makes them believe it more useful. They are accustomed to reaching someone right away and to recognizing their voice, elements of interaction the Internet does not readily communicate. Furthermore, since most students use a computer most at home and home computers (as opposed to school computers in computer labs) are not always on or

connected to the Internet, picking up a telephone can be easier and more convenient than dialing up an Internet connection. Email is often considered by these students to be a means of sending short messages or a way to schedule telephone or face-to-face conversations.

#### G) ACADEMICS AND THE INTERNET:

University students seem generally positive about the Internet and its impact on their educational experience. The majority of students (50%) have a positive attitude toward the Internet and its communication tools. They are comfortable with Internet communication, and even report finding enjoyment while using it for academic and personal reasons. Their Internet habits are split between academic and social uses, and they find it functional for both purposes. The table below shows students responses to the impact of internet on education.

RESPONSE	NUMBER OF RESPONDENTS (%)
Agree	50
Neutral	18
Disagree	19
Not sure	13

The table above shows that most student agree with the fact that the internet has had a positive impact on students education.

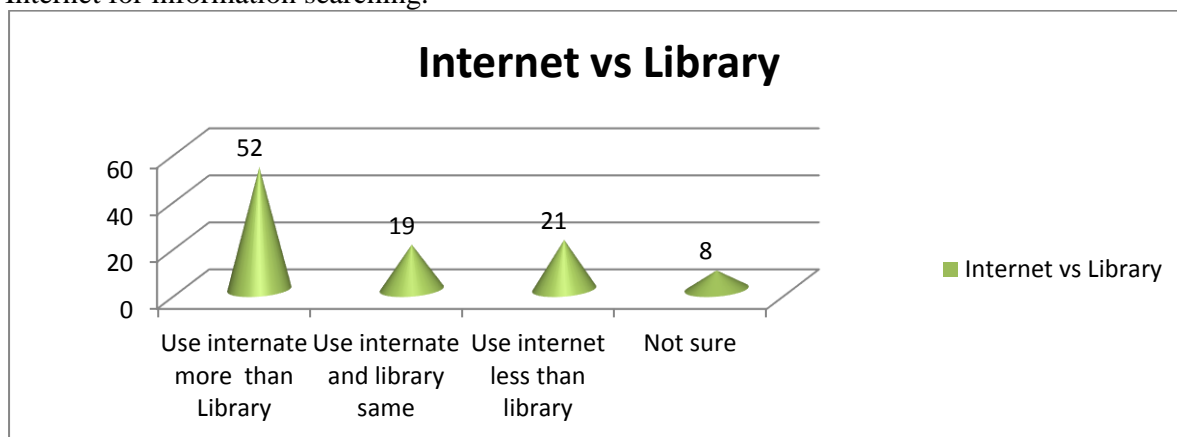
#### H) COMPARING INTERNET TO LIBRARY USE

The convenience of the Internet may be taking some of the trouble out of working on class projects for University students, but some educators and librarians are concerned it may also be creating poor research habits.

Data from the Association of Research Libraries shows that reference queries at university libraries have greatly decreased during and since the late 1990s. The convenience of the Internet is likely tempting students to rely very heavily on it when searching for academic resources. In my own research, an overwhelming number of students reported that the Internet, rather than the library, is the primary site of their information searches. The table below shows the response of students.

RESPONSE	NUMBER OF RESPONDENTS (%)
Use internet more than the library	52
Use internet and library about the same	19
Use internet less than library	21
Not sure	8

The table above shows that out of the 100 students interviewed 52% of the respondents said they use the Internet more than the library, while only 21% said they use the library more than the Internet for information searching.



Traditionally, and ideally, the library has been a place where students go to study and collect materials used for papers, presentations and reports. Of course, people often socialize at the

library, too. Nowadays, the Internet has changed the way students use the library. Students tend to use the Internet prior to going to the library to find information. During direct observations of students' use of the Internet in the library on campus, it was noted that the majority of students' time was not spent using the library resources online. Rather, email use, instant messaging and Web-surfing dominated student's internet activities in the library. Almost every student that was observed checked his or her e-mail while online, but very few were observed surfing university-based or library Web sites. Those students who were using the internet to do academic-related work made use of commercial search engines rather than university and library Web sites. Many students are likely to use information found on search engines and various Websites as research material.

#### I) TIME SPENT ON THE INTERNET:

Students spend a lot of time on the internet nowadays for research work, communication and entertainment. The table below shows the maximum time a student spends on the internet in a day.

Time(minutes)	NUMBER OF RESPONDANTS (%)
10-30	35
40-60	31
Above 60	34

The above table shows that most students spend at least 10-30 minutes on the internet daily while others spend above 60 minutes daily.

### 9. TEST FOR INDEPENDENCE OR MEASURE OF ASSOCIATION:

Here, we want to test  $H_0$ : A and B are independent

$\Rightarrow H_0: p_{ij} = p_{i0} p_{0j}$ , for all (i,j),  $p_{ij}$ =prob of getting a sample from (i,j)th cell.

We consider a sample of size n from the population. Let,  $n_{ij}$  be the frequency of the (i,j)th cell.

$$\chi^2(\text{observed}) = \{n \sum \sum (n_{ij})^2 / n_{i0} n_{0j} - n\} \sim \chi^2 \text{ with df } (K-1)(L-1)$$

Now, we reject  $H_0$ , at level  $\alpha$  level of significance iff-

$$\chi^2(\text{observed}) > \chi^2 \text{ at level } \alpha \text{ with df } (K-1)(L-1)$$

We will use the R programming to find out the value of chi- square.

```
File<-read.table("d:/chi-square.csb",header=TRUE)
```

```
Chi square test<-chisq.test(file)
```

#### TEST 1: Dependency between the age students start to access the internet and their basic institution.

Here we are interested in finding out whether a student had his/her basic education and the age at which they start to access the internet are independent or not. A sample of 100 students is selected and the following data are obtained.

TABLE NO: 02  
Student's Age Group & Their Basic institution

	Age Group				
	5-10	11-15	16-20	21-25	Total( $n_{0j}$ )
Public School	5	7	20	11	43
Private School	6	8	31	12	57
Total ( $n_{i0}$ )	11	15	51	23	$n_{00}= 100$

Here we have two way classified data where age (FACTOR A) and Basic institution (FACTOR B) are the two attributes to test whether there exists any dependence between two attributes. To test for dependency, let  $p_{ij}$  be the probability of assuming  $i$ th age group of a  $j$ th basic institution.

**Hypothesis:** Now we test-  $H_0$ : A and B are independent

$$\Rightarrow H_0: p_{ij} = p_{i0} p_{0j} \quad \text{for all } (i,j)$$

$$\text{against } H_1: p_{ij} \neq p_{i0} p_{0j}$$

$j=1$ =public school

$j=2$ =private school

$i=1$ =age group of 5 to 10 years

$i=2$ =age group of 11 to 15 years

$i=3$ =age group of 16 to 20 years

$i=4$ =age group of 21 to 25 years

Now using the above theory we have,

$$\chi^2(\text{observed}) = 0.6258701 \text{ (using the R programming)}$$

**Conclusion:** Now at level  $\alpha=0.05$  level of significance with degrees of freedom  $3*1=3$  the

value of  $\chi^2$  is=7.815. So we accept the null hypothesis that there is no association or dependency i.e. the factor A and the factor B (age of students and their basic institution) are independent to each other.

### TEST 2: Dependency between gender and time spent accessing the internet.

Here we wish to determine whether there is a dependency between the gender of a student and the maximum time spent on the internet. To achieve this sample of 100 students is selected and the following data obtained.

TABLE NO: 03  
Maximum Time spent on the internet (Daily Basis)  
Maximum time spent in minutes

	10min-30min	40min-60min	Above 60 min	Total( $n_{0j}$ )
Male	17	19	21	57
Female	18	12	13	43
Total ( $n_{i0}$ )	35	31	34	$n_{00}= 100$

Here we have two way classified data where Maximum time spent on internet (FACTOR A) and Gender (FACTOR B) are the two attributes to test whether there exists any dependence between two attributes.

To test for dependency, let  $p_{ij}$  be the probability of assuming  $i$ th times pent of  $j$ th gender.

**Hypothesis:** Now we test-  $H_0$ : A and B are independent

$$\Rightarrow H_0: p_{ij} = p_{i0} p_{0j} \quad \text{for all } (i,j)$$

$$\text{against } H_1: p_{ij} \neq p_{i0} p_{0j}$$

$j=1$ =male

$j=2$ =female

$i=1$ =10 to 30 mins

$i=2$ =40 to 60 mins

$i=3$ =above 60 mins

Now using the above theory we have,

$\chi^2(\text{observed}) = 1.5621884$  (using the R programming)

**Conclusion:** Now at level  $\alpha = 0.05$  level of significance with degrees of freedom  $2 \times 1 = 2$  the value of  $\chi^2$  is  $= 5.991$ . So we accept the null hypothesis that there is no association or dependency i.e. the factor A and the factor B (maximum time spent on internet and gender) are independent to each other.

**TEST 3: Association between gender and internet activities.**

Here we are interested to find out whether there is a relation between the gender of a student and what they use the internet for; a random sample of 100 students provides the following data.

TABLE NO: 04

Internet uses of the Students  
Internet Activities

	Entertainment	News	Research	Communication	Total( $n_{0j}$ )
Male	15	2	29	11	57
Female	10	1	20	12	43
Total( $n_{i0}$ )	25	3	49	23	$n_{00} = 100$

Here we have two way classified data where Internet Activities (FACTOR A) and gender (FACTOR B) are the two attributes to test whether there exists any dependence between two attributes.

To test for dependency, let  $p_{ij}$  be the probability of assuming  $i$ th Internet Activity of  $j$ th gender.

**Hypothesis:** Now we test-  $H_0$ : A and B are independent

$$\Rightarrow H_0: p_{ij} = p_{i0} p_{0j} \quad \text{for all } (i, j)$$

against  $H_1: p_{ij} \neq p_{i0} p_{0j}$

$j=1$ =Male  
 $j=2$ =Female  
 $i=1$ =Entertainment  
 $i=2$ =News  
 $i=3$ =Research  
 $i=4$ =Communication

Now using the above theory we have,

$\chi^2(\text{observed}) = 1.091261545$  (using the R programming)

**Conclusion:** Now at level  $\alpha = 0.05$  level of significance with degrees of freedom  $3 \times 1 = 3$  the value of  $\chi^2$  is  $= 7.815$ . So we accept the null hypothesis that there is no association or dependency i.e. the factor A and the factor B (Internet Activities and Gender) are independent to each other.

**TEST 4: Dependency between gender and how often students access the internet.**

Here we are interested in finding out whether there is a Dependency between the gender of a student and how often they access the internet, a random sample of 100 students provides the following data.

TABLE NO: 05

How often Student Access their internet

	Daily	Weekends	Others	Total( $n_{0j}$ )
Male	24	19	14	57
Female	17	15	11	43
Total( $n_{i0}$ )	41	34	25	$n_{00} = 100$

Here we have two way classified data where How often Student Access their internet (FACTOR A) and Gender (FACTOR B) are the two attributes to test whether there exists any dependence between two attributes.

To test for dependency, let  $p_{ij}$  be the probability of assuming  $i$ th times pent of  $j$ th gender.

**Hypothesis:** Now we test-  $H_0$ : A and B are independent

$$\Rightarrow H_0: p_{ij} = p_{i0} p_{0j} \quad \text{for all } (i,j) \quad \begin{array}{l} j=1=\text{male} \\ j=2=\text{female} \end{array}$$

$$\text{against } H_1: p_{ij} \neq p_{i0} p_{0j}$$

$$\begin{array}{l} i=1=\text{Daily} \\ i=2=\text{Weekends} \\ i=3=\text{Others} \end{array}$$

Now using the above theory we have,

$$\chi^2(\text{observed}) = 0.0670239 (\text{using the R programming})$$

**Conclusion:** Now at level  $\alpha=0.05$  level of significance with degrees of freedom  $2*1=2$  the value of  $\chi^2$  is=5.991. So we accept the null hypothesis that there is no association or dependency i.e. the factor A and the factor B (How often Student Access internet and gender) are independent to each other.

#### TEST 5: Association between gender and students reasons for using SNs.

Here we are interested in finding out whether there is a Dependency between the gender of a student and their reason for signing up for social network, a random sample of 71 students provides the following data.

TABLE NO: 06

Reasons Behind sign up for SNs

	Make new Friends	Share ideas/mems	Chatting	Total( $n_{0j}$ )
Male	26	7	11	44
Female	12	5	10	27
Total ( $n_{i0}$ )	38	12	21	$n_{00} = 71$

Here we have two way classified data where Reasons Behind sign up for SNs (FACTOR A) and Gender (FACTOR B) are the two attributes to test whether there exists any dependence between two attributes.

To test for dependency, let  $p_{ij}$  be the probability of assuming  $i$ th times pent of  $j$ th gender.

**Hypothesis:** Now we test-  $H_0$ : A and B are independent

$$\Rightarrow H_0: p_{ij} = p_{i0} p_{0j} \quad \text{for all } (i,j) \quad \begin{array}{l} j=1=\text{male} \\ j=2=\text{female} \end{array}$$

$$\text{against } H_1: p_{ij} \neq p_{i0} p_{0j}$$

$$\begin{array}{l} i=1=\text{Make new Friends} \\ i=2=\text{Share ideas/ memes} \\ i=3=\text{Chatting} \end{array}$$

Now using the above theory we have,

$$\chi^2(\text{observed}) = 1.362072057 (\text{using the R programming})$$

**Conclusion:** Now at level  $\alpha=0.05$  level of significance with degrees of freedom  $2*1=2$  the value of  $\chi^2$  is=5.991. So we accept the null hypothesis that there is no association or dependency i.e. the factor A and the factor B (Reasons Behind sign up for SNs and gender) are independent to each other.

#### 10. ANOVA TESTING:

For the given data we use two-way Analysis of variance model. The underlying model is given by,

$$Y_{ij} = \mu + A_i + B_j + e_{ij}, \text{ where, } e_{ij} \sim N(0, \sigma^2) \text{ for } i = 1, 2, \dots, p \text{ } j = 1, 2, \dots, q.$$

$Y_{ij}$  = the observation corresponding to the  $i$ th level of factor A and the  $j$ th level of factor B.

$\mu$  = the general effect

$A_i$  = the additional effect of  $i$ th level of the factor A

$B_j$  = the additional effect of  $j$ th level of the factor B

$e_{ij}$  = the random error independent of  $\mu, A_i, B_j$ .

Assumptions: Here the Assumptions are-

- 1)  $\sum A_i = 0 \text{ } i=1(1)p$
- 2)  $\sum B_j = 0 \text{ } j=1(1)q$

Hypothesis: Here we want to test the following Hypothesis-

- 1)  $H_{0A}: A_i = 0 \text{ for all } i=1(1)p \text{ against } H_{1A}: A_i \neq 0$
- 2)  $H_{0B}: B_j = 0 \text{ for all } j=1(1)q \text{ against } H_{1B}: B_j \neq 0$

Orthogonal Splitting: The total variation is partitioned as,

$$TSS = SSA + SSB + SSE$$

Where, TSS is the total sum of variation, SSA is the sum of squares due to different levels of factor A, SSB is the total sum of squares due to different levels of factor B and SSE is the sum of squares due to error.

$$TSS = \sum \sum (Y_{ij} - Y_{00})^2 = \sum \sum Y_{ij}^2 - CF$$

$$SSE = \sum \sum (Y_{ij} - Y_{i0} - Y_{0j} + Y_{00})^2$$

$$SSA = q \sum (Y_{i0} - Y_{00})^2 = \{ \sum (T_{i0})^2 / q \} - CF$$

$$SSB = p \sum (Y_{0j} - Y_{00})^2 = \{ \sum (T_{0j})^2 / p \} - CF$$

Where,  $T_{00}$  = Grand total =  $\sum \sum Y_{ij}$

$$CF = (T_{00})^2 / pq$$

$T_{i0}$  = Column means

$T_{0j}$  = Row means

Under  $H_{0A}$ ,  $SSA / \sigma^2 \sim \chi^2$  with df (p-1)  $\Rightarrow MSA = SSA / p - 1$

Under  $H_{0B}$ ,  $SSB / \sigma^2 \sim \chi^2$  with df (q-1)  $\Rightarrow MSB = SSB / q - 1$  &  $MSE = SSE / (p-1)(q-1)$

Statistic: The F statistic is obtained from the ratio  $F_A = MSA / MSE$  &  $F_B = MSB / MSE$

Hence under  $H_{0A}$ :  $F_A = MSA / MSE \sim F_{(p-1), (p-1)(q-1)}$

Hence under  $H_{0B}$ :  $F_B = MSB / MSE \sim F_{(q-1), (p-1)(q-1)}$

Rejection Criterion: And we reject the null hypothesis now at  $\alpha$  level of significances if

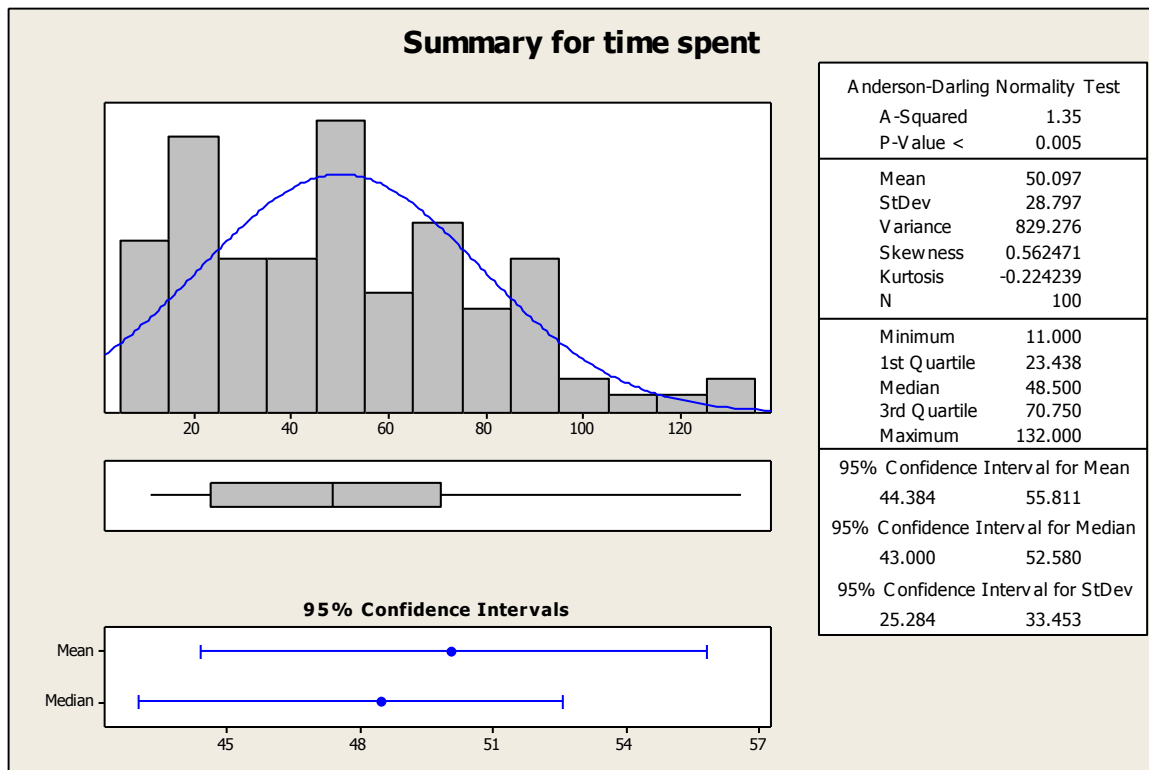
$$F_{obs} > F_{\alpha, (p-1), (p-1)(q-1)}$$

Critical Differences: If we reject our null hypothesis, then to test equality between two levels (suppose  $i$ th level &  $j$ th level), we calculate

$$\text{Critical Difference (CD)} = t_{\alpha/2} (\sqrt{2MSE / q})$$

And compare with  $|T_{i0} - T_{0j}|$ , if it is greater than the difference then there is a significant difference between levels.

Assumption: Here we assume that all the data of time spent is came from the normal population. From the fitting of the mentioned data in below, the p-value  $< 0.005$ , so, we can conclude that the data is from the normal population.



### TEST: 01: ANOVA Testing for the age students start to access the internet and their basic institution.

Here we are interested in find whether there is any difference between various age groups when they start to access the internet and as well as in different basic institution. Here our factor A is the age students start to access the internet having 4 levels ( $p=4$ ) & factor B is their basic institution having 2 levels ( $q=2$ ).

Here the Basic institution and the age groups are independent to each other, so we are not considering the interaction effect.(from the results of 9.1)

**Hypothesis:** Now we test-

- 1)  $H_{0A}$ : All the age groups when student start accessing internets are same, i.e.  $A_i=0$  for all  $i=1(1)4$  against  $H_{1A}$ :  $A_i \neq 0$
- 2)  $H_{0B}$ : The basic institution student have who accessing internets are same, i.e.  $G=N$  for all against  $H_{1B}$ :  $G \neq N$

G=public school

N=private school

$A_1$ =age group of 5 to 10 years

$A_2$ =age group of 11 to 15 years

$A_3$ =age group of 16 to 20 years

$A_4$ =age group of 21 to 25 year

### General Linear Model: time spent versus SCHOOL, age group

Factor	Type	Levels	Values
SCHOOL	fixed	2	G, N
age group	fixed	4	A1, A2, A3, A4

Analysis of Variance for time spent, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
SCHOOL	1	1786.2	1608.0	1608.0	2.22	0.014
age group	3	11386.9	11386.9	3795.6	5.23	0.002
Error	95	68925.3	68925.3	725.5		



Total 99 82098.4

**Conclusion:** Here the hypothesis under the test  $H_{0A}$ : All the age groups when student start accessing internets are same is accepted. And the hypothesis under  $H_{0B}$ : The basic institution students have who accessing internets are same is accepted, i.e. there is no significant difference for different basic institution in accessing the internet.

### TEST 2: ANOVA testing for gender inequality for use of internet.

Here we are interested in find whether there is any difference between male or female and as well as in different time spent on internet. Here factor A is their gender having 2 levels ( $q=2$ ).

**Hypothesis:** Now we test-

- 1)  $H_{0A}$ : Male and female student have who accessing internets are same, i.e.  $M=F$  against  $H_{1A}$ :  $M \neq F$

M=Male  
F=Female

Analysis of Variance for time spent, gender using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
gender	1	1011.3	1011.3	1011.3	1.22	0.034
Error	98	81087.0	81087.0	827.4		
Total	99	82098.				

**Conclusion:** Here the hypothesis under the test  $H_{0A}$ : Male and female student have who accessing internets are same is also accepted i.e. there is no significant difference between gender when accessing the internet.

### TEST 3: ANOVA testing for different internet activities.

Here we are interested in find whether there is any difference between different internet activities among students. Here our factor A is different internet activities having 4 levels ( $p=4$ ) & factor B is gender which will work like block effects to reduce the SSE for a better testing procedure.

Here the different internet activities and the gender are independent to each other, so we are not considering the interaction effect. (from the results of 9.3)

**Hypothesis:** Now we test-

- 1)  $H_{0A}$ : All different internet activities are same, i.e.  $E=R=N=C$  against  $H_{1A}$ : not  $H_0$

M=Male

F=Female

E=Uses internet mainly for Entertainment

N=Uses internet mainly for News

R= Uses internet mainly for Research

C= Uses internet mainly for Communication

### General Linear Model: time spent versus gender, USES OF INTERNET

Factor	Type	Levels	Values
gender	fixed	2	F, M
USES OF INTERNET	random	4	C, E, N, R

Analysis of Variance for time spent, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
--------	----	--------	--------	--------	---	---

gender	1	1011.3	900.6	900.6	1.33	0.034
USES OF INTERNET	3	16837.0	16837.0	5612.3	8.30	0.251
Error	95	64250.1	64250.1	676.3		
Total	99	82098.4				

*Critical difference:*

We need to calculate Critical Difference (CD) =  $2.365 \times 3.5355 = 8.361537$

Now the mean age when they start using internet is Students Activity on internet

Entertainment	52.5
News	43.5
Research	24.5
Communication	19.5

**Conclusion:** Here the hypothesis under the test  $H_{0A}$ : All different internet activities are is rejected, so there is significant difference between different internet activities among students. For entertainment, research and communication have the most significant differences among students.

#### **TEST 4: ANOVA testing how often students access the internet.**

Here we are interested in find whether there is any difference between how often student accessing internet. Here our factor A is how often student having internet having 3 levels ( $p=3$ ) & factor B is gender which will work like block effects to reduce the SSE for a better testing procedure.

Here different timings of using internet and the gender are independent to each other, so we are not considering the interaction effect. (from the results of 9.4)

**Hypothesis:** Now we test-

- 1)  $H_{0A}$ : All different timings of using internet are same, i.e.  $D=W=O$  against  $H_{1A}$ : not  $H_0$

M=Male

F=Female

D=Daily

W=weekend

O=others

#### **General Linear Model: time spent versus gender, when use internet**

Factor	Type	Levels	Values
gender	fixed	2	F, M
when use internet	fixed	3	D, O, W

Analysis of Variance for time spent, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
gender	1	1011.3	1129.2	1129.2	1.42	0.034
when use internet	2	4901.5	4901.5	2450.8	3.09	0.049
Error	96	76185.5	76185.5	793.6		
Total	99	82098.4				

**Conclusion:** Here the hypothesis under the test  $H_{0A}$ : All different timings of using internet are same is accepted, so there is no significant difference between different timings of using internet.

#### **TEST 5: ANOVA testing for student's reasons for using SNs.**

Here we are interested in find whether there is any difference between Students Reasons for using SNs. Here our factor A is Students Reasons for using SNs having 3 levels ( $p=3$ ) & factor B is gender which will work like block effects to reduce the SSE for a better testing procedure.

Here the different reasons behind sign up for SNS and the gender are independent to each other, so we are not considering the interaction effect. (from the results of 9.5)

**Hypothesis:** Now we test-

- 1)  $H_{0A}$ : All different reasons for sign up SNS are same, i.e.  $C=F=S$  against  $H_{1A}$ : not  $H_0$ .  
 $C$ =communication  
 $F$ =making new Friends  
 $S$ =share new ideas

Factor	Type	Levels	Values
gender	fixed	2	F, M
SNS	random	3	C, F, S

Analysis of Variance for time spent, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
gender	1	1011.3	971.1	971.1	1.16	0.034
SNS	2	398.7	398.7	199.4	0.24	0.021
Error	96	80688.3	80688.3	840.5		
Total	99	82098.4				

**Conclusion:** Here the hypothesis under the test  $H_{0A}$ : All different reasons for sign up SNS are same is accepted, so there is no significant difference between different reasons for sign up SNS.

## 12. CONCLUSIONS:

- The amounts of time spend and the activities students perform on the internet do not depend on gender.
- The age at which students start to access the internet does not depend on the institution they had their basic education.
- The internet has gain much popularity among students nowadays, most students prefer to use the online information searching rather than the library since the internet is found to be fast and easily accessible.
- Although technology is a very important, useful and indispensable part of the life, effective and appropriate usage of it still needs to be improved in education, most students sometimes, use the computer and Internet for reasons other than educational purposes. The effect of the Internet in education is, hence, relatively limited.
- Making new friends, chatting and sharing of ideas are some of the factors that influence students to sign up for SNS, and there is significant difference between different internet activities among students. For entertainment, research and communication have the most significant differences among students.
- SNS has negative impact on students education since students spend most of their time on SNS when they go online.
- To gain information, students should not only rely on textbooks provided and materials prepared by the lectures, but the skills in searching information by themselves from the use of

internet should be emphasized. A mastery skill in using computer is the basis towards an effort in mastering communication and information skills.

### **13.APPENDIX:**

The data set which is used in this project is given below.

Time spent	Gender	SCHOOL	AGE	USES OF INTERNET	Reason behind sign up SNS	POCKET MONEY(monthly)	age group	when use internet
11	M	N	13	R	S	100	A2	D
13	F	G	16	C	S	150	A3	O
27	M	G	5	R	F	100	A1	W
25.5	F	G	18	C	S	300	A3	W
19	F	N	21	C	S	500	A4	O
12.5	M	G	7	R	S	100	A1	D
15	F	N	25	C	F	1500	A4	W
16.5	F	N	24	C	F	1500	A4	W
18.25	M	N	15	R	C	500	A2	D
15.5	F	N	21	C	S	1250	A4	O
23	M	G	20	C	F	1200	A3	W
27.5	F	N	23	C	S	1300	A4	D
28	M	G	16	R	C	500	A3	D
12	M	G	8	R	F	150	A1	D
21	M	N	22	C	C	600	A4	W
22.25	M	G	19	R	F	800	A3	D
25	F	N	12	C	F	200	A2	D
26.25	F	G	18	R	S	1000	A3	O
17	M	N	15	R	F	500	A2	D
16	F	N	25	R	S	1500	A4	D
18	M	G	9	C	F	100	A1	W
13.5	F	G	16	R	F	500	A3	O
14	M	N	13	R	C	150	A2	D
20	F	N	22	R	F	800	A1	D
20	M	G	17	R	C	500	A3	O
29	F	G	19	R	F	850	A3	O
11.5	M	G	20	R	C	1600	A3	D
11.75	F	G	6	R	F	100	A1	D
13.5	M	N	21	C	C	1200	A4	W
27.75	F	G	20	R	F	1000	A3	D
12	M	N	23	R	C	1500	A4	D
30	F	N	10	R	F	100	A1	O
22.25	M	G	16	R	C	800	A3	D
24.75	F	N	7	R	F	100	A1	W
19.75	F	N	25	R	F	750	A4	O
43	M	G	16	R	S	560	A3	D
41.5	F	N	24	R	F	1200	A4	O
45	M	N	11	R	C	450	A2	O
42.75	M	G	19	C	F	600	A3	D
47	M	G	20	R	C	800	A3	W
49.25	F	N	8	R	C	800	A1	W
50	M	G	19	R	F	300	A3	D
52	F	G	23	R	C	1200	A4	W
57	M	G	18	R	F	1500	A3	W
58.5	F	N	5	R	C	1400	A1	O
55	F	G	22	R	S	1100	A4	W
54.25	M	G	18	R	S	400	A3	D

47.25	F	G	25	R	F	300	A4	W
44	M	G	17	C	F	500	A3	W
45.5	M	G	24	R	C	100	A4	D
50	M	G	14	C	F	300	A2	D
49.25	F	G	19	E	F	500	A3	D
51.25	M	N	17	N	C	100	A3	W
52	M	N	7	R	C	1500	A1	D
55	M	G	24	R	F	1500	A4	O
44	F	N	16	E	F	500	A3	D
42.75	M	G	21	E	C	1250	A4	D
43.5	F	N	20	N	F	1200	A3	D
43	M	N	6	R	C	1300	A1	W
46	F	G	23	R	S	500	A4	D
49	M	N	16	E	C	150	A3	D
48	F	G	25	R	F	600	A4	W
58	M	N	11	E	S	800	A2	D
42	F	G	24	R	C	200	A3	W
49.25	M	N	20	E	F	1000	A3	O
50.5	M	N	18	E	F	800	A3	D
65	M	N	18	E	F	500	A3	D
67	F	N	17	R	F	850	A3	O
86	M	N	17	E	S	1600	A3	D
87	F	N	18	C	S	100	A3	W
85	M	N	20	E	C	1200	A3	O
70	F	G	15	C	C	1000	A2	W
77	M	N	18	E	C	1500	A3	W
62	M	N	19	E	F	100	A3	O
75	M	N	17	R	C	800	A3	W
68	M	N	19	R	S	100	A3	O
82	F	N	16	C	C	750	A3	W
83	M	G	13	E	F	1500	A2	W
129	M	N	19	E	F	100	A3	O
132	F	N	19	C	C	500	A3	W
98	M	G	14	E	C	150	A2	W
67	M	N	20	R	F	800	A3	O
72.5	F	N	19	E	C	500	A3	W
65	M	G	12	E	F	850	A2	W
60	F	N	16	E	C	1600	A3	D
71	M	N	16	E	F	100	A3	W
80	F	N	18	E	C	1200	A3	D
87	M	G	11	R	F	1000	A2	O
73	M	N	18	C	F	1500	A3	W
75.5	F	N	20	E	C	100	A3	D
89.5	M	G	24	R	F	800	A4	O
115	M	N	19	C	S	100	A3	W
88	F	G	13	E	S	1200	A2	D

105	M	G	23	C	F	2300	A4	O
92	M	N	20	N	F	1000	A3	O
95	M	G	11	C	F	300	A2	W
93	M	N	17	R	S	400	A3	O
89	F	N	18	E	C	400	A3	D
67.5	F	N	19	E	C	600	A3	D
74.25	F	N	17	E	S	800	A3	D