Study Habits of the student of UIR

Presented by:
Maha EL HANAFI
Younes Diba
Achraf Attary
Imane Rammouch
Aymen Bouhair

Supervised by:
GHOGHO Mounir
SBIHI Nada

Plan

i. Introduction

ii. Data Analysis

iii. Hypothesis Tests

iv. Conclusion

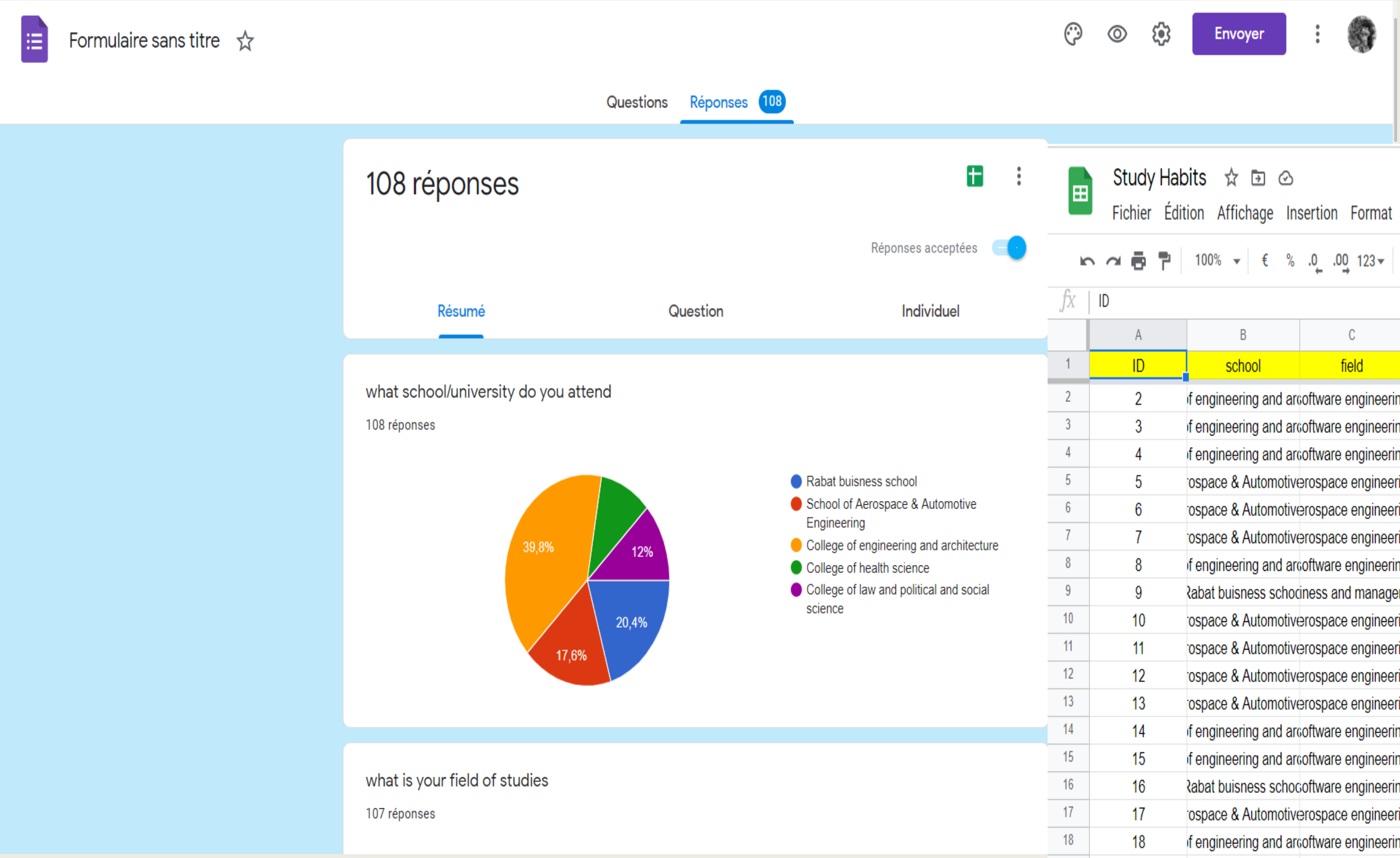
Introduction

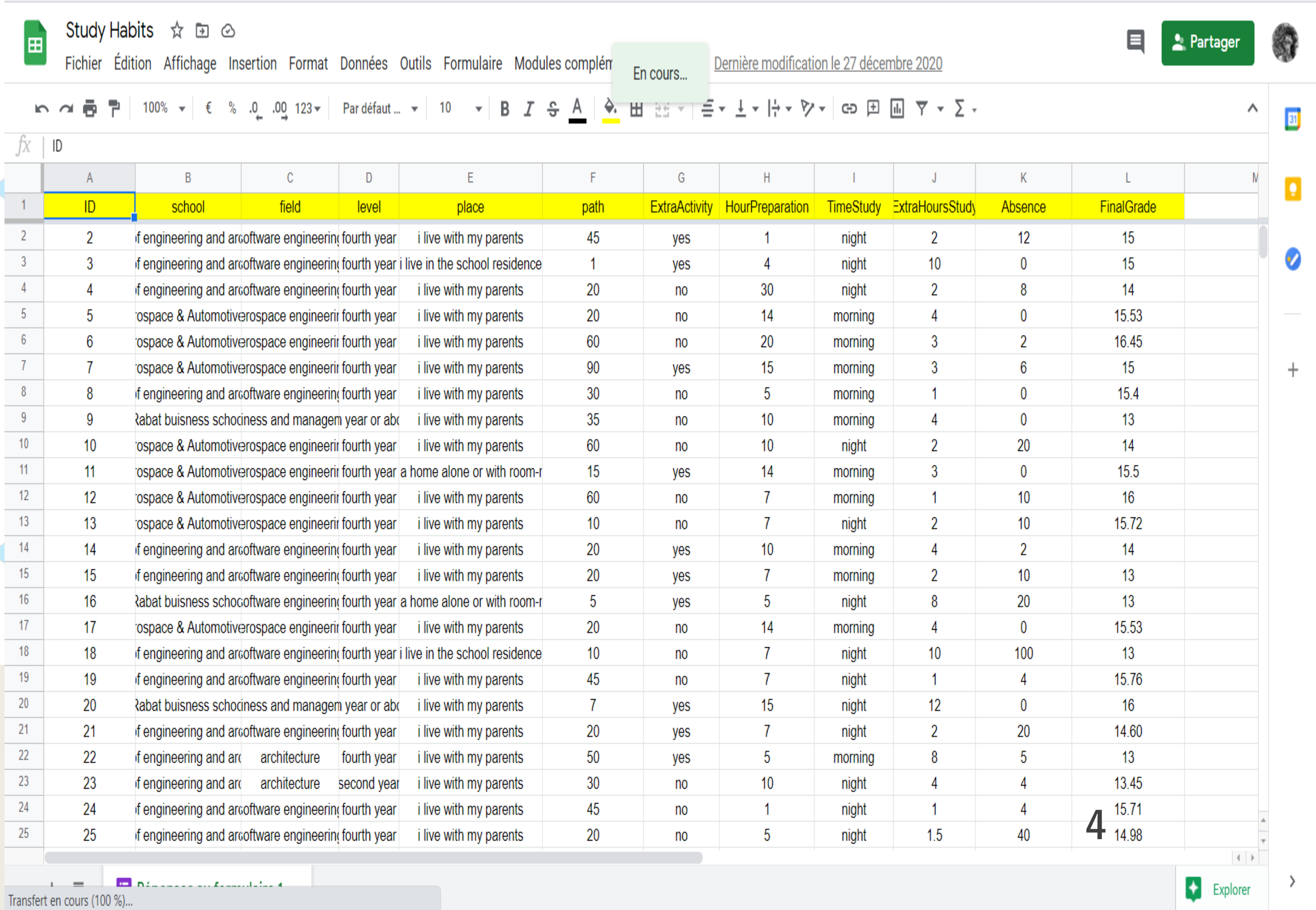
The definition of study habits is the habitual practices one uses to help them study and learn. Good study habits can help students achieve and/or maintain good grades.

The purpose of this project is to help us get valuable data about study habits of the students of the international university of Rabat. The data can be used to find patterns in the behavior of the students, in order for us to identify areas of strength and potential changes. All answers are completely anonymous; no names or email are being collected.

Data Analysis

To collect the data, we used Google Forms.





	A	B	C	D	E	F	G	H	I	J	K	L
1	ID	school	field	level	place	path	ExtraActivity	HourPreparation	TimeStudy	ExtraHoursStudy	Absence	FinalGrade
2	2	f engineering and ar	oftware engineerin	fourth year	i live with my parents	45	yes	1	night	2	12	15
3	3	f engineering and ar	oftware engineerin	fourth year	i live in the school residence	1	yes	4	night	10	0	15
4	4	f engineering and ar	oftware engineerin	fourth year	i live with my parents	20	no	30	night	2	8	14
5	5	ospace & Automotive	rospace engineerin	fourth year	i live with my parents	20	no	14	morning	4	0	15.53
6	6	ospace & Automotive	rospace engineerin	fourth year	i live with my parents	60	no	20	morning	3	2	16.45
7	7	ospace & Automotive	rospace engineerin	fourth year	i live with my parents	90	yes	15	morning	3	6	15
8	8	f engineering and ar	oftware engineerin	fourth year	i live with my parents	30	no	5	morning	1	0	15.4
9	9	Rabat buisness scho	iness and managem	year or abo	i live with my parents	35	no	10	morning	4	0	13
10	10	ospace & Automotive	rospace engineerin	fourth year	i live with my parents	60	no	10	night	2	20	14
11	11	ospace & Automotive	rospace engineerin	fourth year	a home alone or with room-r	15	yes	14	morning	3	0	15.5
12	12	ospace & Automotive	rospace engineerin	fourth year	i live with my parents	60	no	7	morning	1	10	16
13	13	ospace & Automotive	rospace engineerin	fourth year	i live with my parents	10	no	7	night	2	10	15.72
14	14	f engineering and ar	oftware engineerin	fourth year	i live with my parents	20	yes	10	morning	4	2	14
15	15	f engineering and ar	oftware engineerin	fourth year	i live with my parents	20	yes	7	morning	2	10	13
16	16	Rabat buisness scho	oftware engineerin	fourth year	a home alone or with room-r	5	yes	5	night	8	20	13
17	17	ospace & Automotive	rospace engineerin	fourth year	i live with my parents	20	no	14	morning	4	0	15.53
18	18	f engineering and ar	oftware engineerin	fourth year	i live in the school residence	10	no	7	night	10	100	13
19	19	f engineering and ar	oftware engineerin	fourth year	i live with my parents	45	no	7	night	1	4	15.76
20	20	Rabat buisness scho	iness and managem	year or abo	i live with my parents	7	yes	15	night	12	0	16
21	21	f engineering and ar	oftware engineerin	fourth year	i live with my parents	20	yes	7	night	2	20	14.60
22	22	f engineering and ar	architecture	fourth year	i live with my parents	50	yes	5	morning	8	5	13
23	23	f engineering and ar	architecture	second year	i live with my parents	30	no	10	night	4	4	13.45
24	24	f engineering and ar	oftware engineerin	fourth year	i live with my parents	45	no	1	night	1	4	15.71
25	25	f engineering and ar	oftware engineerin	fourth year	i live with my parents	20	no	5	night	1.5	40	14.98

Data Analysis : Import Data in Rstudio

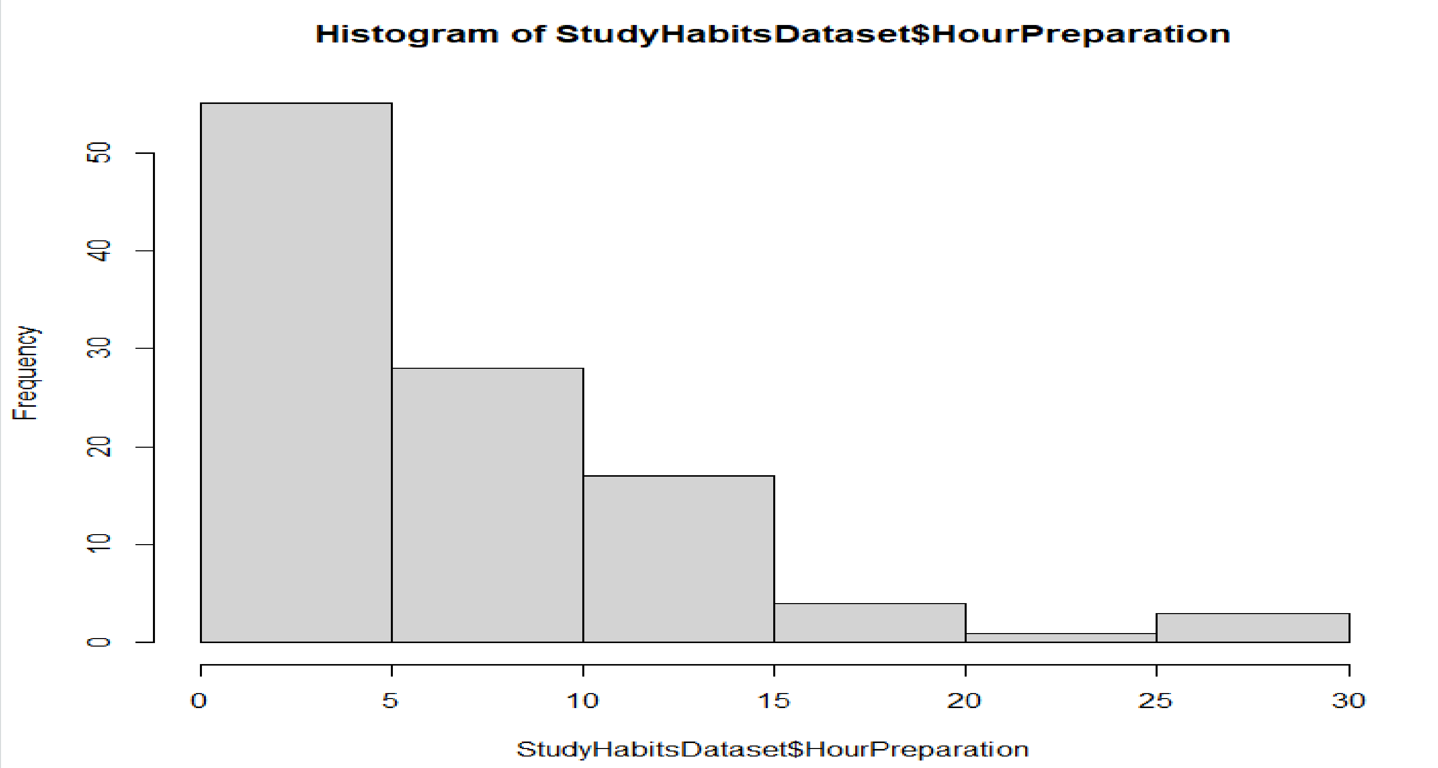
	ID	school	field	level	place	path	ExtraActivity	HourPreparation	TimeStudy	ExtraHoursStudy	Absence	FinalGrade	X	X.1
1	2	College of engineering and architecture	software engineering	fourth year	i live with my parents	45.0	yes	1	night	2.00	12	15.00	NA	
2	3	College of engineering and architecture	software engineering	fourth year	i live in the school residence	1.0	yes	4	night	10.00	0	15.00	NA	
3	4	College of engineering and architecture	software engineering	fourth year	i live with my parents	20.0	no	30	night	2.00	8	14.00	NA	
4	5	School of Aerospace & Automotive Engineering	Aerospace engineering	fourth year	i live with my parents	20.0	no	14	morning	4.00	0	15.53	NA	
5	6	School of Aerospace & Automotive Engineering	Aerospace engineering	fourth year	i live with my parents	60.0	no	20	morning	3.00	2	16.45	NA	
6	7	School of Aerospace & Automotive Engineering	Aerospace engineering	fourth year	i live with my parents	90.0	yes	15	morning	3.00	6	15.00	NA	
7	8	College of engineering and architecture	software engineering	fourth year	i live with my parents	30.0	no	5	morning	1.00	0	15.40	NA	
8	9	Rabat buisness school	Business and management	fifth year or above	i live with my parents	35.0	no	10	morning	4.00	0	13.00	NA	
9	10	School of Aerospace & Automotive Engineering	Aerospace engineering	fourth year	i live with my parents	60.0	no	10	night	2.00	20	14.00	NA	
10	11	School of Aerospace & Automotive Engineering	Aerospace engineering	fourth year	I rent a home alone or with room-mates	15.0	yes	14	morning	3.00	0	15.50	NA	
11	12	School of Aerospace & Automotive Engineering	Aerospace engineering	fourth year	i live with my parents	60.0	no	7	morning	1.00	10	16.00	NA	
12	13	School of Aerospace & Automotive Engineering	Aerospace engineering	fourth year	i live with my parents	10.0	no	7	night	2.00	10	15.72	NA	
13	14	College of engineering and architecture	software engineering	fourth year	i live with my parents	20.0	yes	10	morning	4.00	2	14.00	NA	
14	15	College of engineering and architecture	software engineering	fourth year	i live with my parents	20.0	yes	7	morning	2.00	10	13.00	NA	
15	16	Rabat buisness school	software engineering	fourth year	I rent a home alone or with room-mates	5.0	yes	5	night	8.00	20	13.00	NA	
16	17	School of Aerospace & Automotive Engineering	Aerospace engineering	fourth year	i live with my parents	20.0	no	14	morning	4.00	0	15.53	NA	
17	18	College of engineering and architecture	software engineering	fourth year	i live in the school residence	10.0	no							
18	19	College of engineering and architecture	software engineering	fourth year	i live with my parents	45.0	no							
19	20	Rabat buisness school	Business and management	fifth year or above	i live with my parents	7.0	yes							
20	21	College of engineering and architecture	software engineering	fourth year	i live with my parents	20.0	yes							
21	22	College of engineering and architecture	architecture	fourth year	i live with my parents	50.0	yes							
22	23	College of engineering and architecture	architecture	second year	i live with my parents	30.0	no							
23	24	College of engineering and architecture	software engineering	fourth year	i live with my parents	45.0	no							
24	25	College of engineering and architecture	software engineering	fourth year	i live with my parents	20.0	no							
25	26	School of Aerospace & Automotive Engineering	automotive engineering	fourth year	i live with my parents	80.0	yes							
26	27	College of engineering and architecture	software engineering	fourth year	i live with my parents	15.0	no							

Showing 1 to 28 of 108 entries, 14 total columns

```r
#habit=unique(Study.Habits..Dataset$ID)
download.packages("http://r-forge.r-project.org/projects/inference/")


#Peek at our data
#before we begin working on the dataset, lets have a good look at the raw data

View(Study.Habits..Dataset)

View(StudyHabitsDataset)


head(Study.Habits..Dataset, 10) #gives top 10 rows of the data set
tail(Study.Habits..Dataset) #bottom 5 rows of the data set

names(Study.Habits..Dataset)
str(Study.Habits..Dataset) #helps to understand the structure of data

#summary()one of the most important function that help in summarising each attribute
#In case of character variables -> Gives the length and the class.
#In case of a Numerical Variable -> Gives Mean, Median, Mode, Range and Quartiles.
summary(Study.Habits..Dataset)
```

Data Analysis: Now we look at the number of days before students start preparing for an exam

```
> hist(StudyHabitsDataset$HourPreparation)
> 
```



- ✓ The distribution is right skewed. The majority of people study less than 10 days before an exam.

➢ Hypothesis Tests:

1, Hypothesis test and confidence interval on a mean

```
> #Test d'hypothèse et intervalle de confiance sur une moyenne
> head(StudyHabitsDataset)
  ID                                school                 field       level                         place path ExtraActivity HourPreparation
1  2       College of engineering and architecture  software engineering fourth year       i live with my parents   45           yes               1
2  3       College of engineering and architecture  software engineering fourth year i live in the school residence    1           yes               4
3  4       College of engineering and architecture  software engineering fourth year       i live with my parents   20            no              30
4  5 School of Aerospace & Automotive Engineering Aerospace engineering fourth year       i live with my parents   20            no              14
5  6 School of Aerospace & Automotive Engineering Aerospace engineering fourth year       i live with my parents   60            no              20
6  7 School of Aerospace & Automotive Engineering Aerospace engineering fourth year       i live with my parents   90           yes              15
  TimeStudy ExtraHoursStudy Absence FinalGrade  X X.1
1     night               2      12      15.00 NA
2     night              10       0      15.00 NA
3     night               2       8      14.00 NA
4   morning               4       0      15.53 NA
5   morning               3       2      16.45 NA
6   morning               3       6      15.00 NA
>
```

Let us find a 98-confidence interval for the mean Absence of students. For this to be valid, we are assuming that we have a *random sample* from all students and that the average of absence of students is normally distributed.

The R command that finds a confidence interval for the mean in this way is:

```
> t.test(StudyHabitsDataset$Absence, conf.level = .98)

        One Sample t-test

data:  StudyHabitsDataset$Absence
t = 8.2409, df = 107, p-value = 4.642e-13
alternative hypothesis: true mean is not equal to 0
98 percent confidence interval:
  9.77646 17.63095
sample estimates:
mean of x
  13.7037
```

- ✓ We get a lot of information, but we can pull out what we are looking for as the confidence interval [9.78, 17.64]. So, we are 98 confident that the mean Absence of students is greater than 9.78 and less than 17.64.

➢ Hypothesis Tests:

2, Hypothesis test and confidence interval on the difference between two means

	morning	night
Mean_Final_grade	14.44293	14.62701
SD_Finale_grade	1.821781	1.436984
n	41	67

```
> Total_morning
[1] 41 14
> Total_night
[1] 67 14
>
> mean(StudyHabitsDataset$FinalGrade)
[1] 14.55713
> #mean
>
> mean_morning=mean(morning$FinalGrade, na.rm = TRUE)
> mean_night=mean(night$FinalGrade, na.rm = TRUE)
>
> mean_morning
[1] 14.44293
> mean_night
[1] 14.62701
>
> #standard deviation
>
> sd_morning=sd(morning$FinalGrade, na.rm = TRUE)
> sd_night=sd(night$FinalGrade, na.rm = TRUE)
>
> sd_morning
[1] 1.821781
> sd_night
[1] 1.436984
> |
```

Hypothesis Tests:

Hypothesis test and confidence interval on the difference between two means

Parameter of interest: µm-µn

Point estimate: Xm-Xn

Hypotheses:

H0: µm = µn

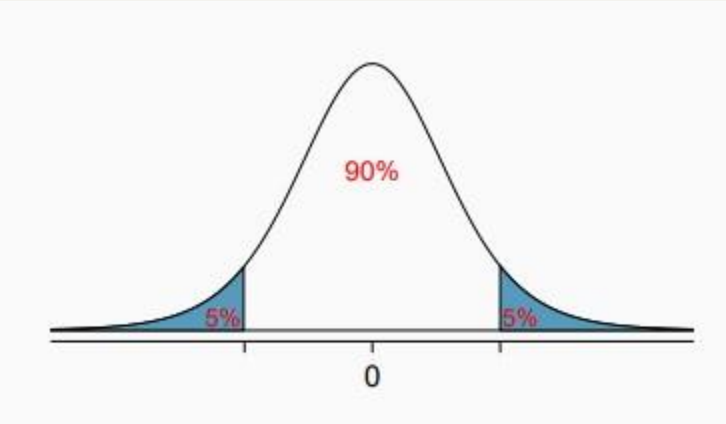
HA: µm < µn

Conditions:

- Independence between those groups
- Both sample sizes should be at least 30

Test statistic :

```
> #we found the p-value is high, so we fail to reject the null hypothesis.
> #variance
> variance<- ((Total_morning-1)*sd_morning**2+(Total_night-1)*sd_night**2)/(Total_morning+Total_night-2)
> variance**0.5
[1] 1.593147 1.640703
> #test
> t<-(mean_morning-mean_night)/((variance*(1/Total_morning+1/Total_night))**0.5)
> t
[1] -0.5827567 -0.2968553
> #degre de liberté
> df<-Total_morning+Total_night-2
> df
[1] 106  26
>
> #Pour calculer notre probabilité, nous avons besoin de la fonction de distribution:
> pt(t,df)
[1] 0.2806474 0.3844678
>
> #On peut regarder la différence entre le test et l'utilisation d'une loi normale :
>
>
> pbis<-2-2*pnorm(abs(t))
> pbis
[1] 0.5600571 0.7665770
>
> #we found the p-value is high, so we fail to reject the null hypothesis.
> |
```



Critical value:

```
>
> #critical value
> qt(p = 0.95, df = 26)
[1] 1.705618
>
```

What is the equivalent confidence level for a one-sided hypothesis test at α = 0.05? **90%**

➢ Hypothesis Tests:

confidence interval on the difference between two means

point estimate ± ME

```
> #Confidence interval
> #SE
> se <- sqrt(sd_morning*sd_morning/Total_morning+sd_night*sd_night/Total_night)
> se
[1] 0.3343175 0.6201273
> #margin error
> error <- qt(0.975,df=pmin(Total_morning,Total_night)-1)*se
> error
[1] 0.6756809 1.3397036
> #we use a 95% confidence interval:
>
> left <- (mean_morning-mean_night)-error
> right <- (mean_morning-mean_night)+error
> left
[1] -0.859769 -1.523792
> right
[1] 0.4915928 1.1556155
>
```

This gives the confidence intervals for each of the three tests. For example, in the first experiment the 95% confidence interval is between -

0.86 and 0.5 assuming that the random variables are normally distributed, and the samples are independent.

➢ Hypothesis tests:

3, Test d’hypothèse et intervalle de confiance sur une proportion

We conduct this test on the proportion of people who take part in extracurricular activities.

H0: p equal to 0.5

HA: p not equal to 0.5

```
> table(StudyHabitsDataset$ExtraActivity)

 no yes
 62  46
```

```
> prop.test(46,46+62,p=0.5)

	1-sample proportions test with continuity correction

data:  46 out of 46 + 62, null probability 0.5
X-squared = 2.0833, df = 1, p-value = 0.1489
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.3324496 0.5247389
sample estimates:
        p
0.4259259
```

The p-value is greater than 0.05, so

we fail to reject the null hypothesis.

The confidence interval is :

(0.3324496 , 0.5247389)

➢ Hypothesis tests:

4, Test d’hypothèse et intervalle de confiance sur deux proportions

We conduct this test to compare two proportions: the proportion of student who got more than 15 on their final grade, who have more than 20 minutes to get to school vs less than 20 minutes to get to school.

```
> Under20Over15 <- subset(StudyHabitsDataset,StudyHabitsDataset$FinalGrade > 15 & Study
HabitsDataset$path <20)
> Under20Under15 <- subset(StudyHabitsDataset,StudyHabitsDataset$FinalGrade < 15 & Stud
yHabitsDataset$path <20)
> Over20Under15 <- subset(StudyHabitsDataset,StudyHabitsDataset$FinalGrade < 15 & Study
HabitsDataset$path >20)
> Over20Over15 <- subset(StudyHabitsDataset,StudyHabitsDataset$FinalGrade > 15 & StudyH
abitsDataset$path >20)
```

Now, we can create a matrix to clearly see the proportions:

```
> myMatrix <- matrix(c(16,21,35,18),ncol=2, nrow=2, byrow=TRUE)
> colnames(myMatrix) <- c("under 20 minutes","over 20 minutes")
> rownames(myMatrix) <- c("more than 15 grade","less than 15 grade")
> myMatrix
                   under 20 minutes over 20 minutes
more than 15 grade               16              21
less than 15 grade               35              18
>
```

We can conduct the hypothesis test, and we assume that success is when a student gets more than 15 on his/her final grade.

Test d’hypothèse et intervalle de confiance sur deux proportions

Hypothesis test:
H0 : p1=p2
HA : p1 != p2

```
> prop.test(x=c(21,16), n=c(21+18,16+35),
+            conf.level=0.95)

        2-sample test for equality of proportions with continuity correction

data:  c(21, 16) out of c(21 + 18, 16 + 35)
X-squared = 3.7289, df = 1, p-value = 0.05348
alternative hypothesis: two.sided
95 percent confidence interval:
 0.0003787567 0.4490933399
sample estimates:
   prop 1    prop 2
0.5384615 0.3137255
```

- ✓ The p-value is fairly small, which means we can reject the null hypothesis.
- ✓ Now we conduct a one sided test to see which proportion is likely to be greater:

```
> prop.test(x=c(21,16), n=c(21+18,16+35),
+           conf.level=0.95,alternative = 'greater')

        2-sample test for equality of proportions with continuity correction

data:  c(21, 16) out of c(21 + 18, 16 + 35)
X-squared = 3.7289, df = 1, p-value = 0.02674
alternative hypothesis: greater
95 percent confidence interval:
 0.03281206 1.00000000
sample estimates:
   prop 1    prop 2
0.5384615 0.3137255
```

We have a p-value of $0.02674 < 0.05$ so we can reject the null hypothesis.

Hypothesis tests:

5, Test Chi-2 :

We want to know if living in the school residence has any effect on one's grades. For this, we use the chi-2 test, to see if the proportion of students that got more than 15 on their final grade changes depending where the student lives.

```
                    over 15   under 15
living with parents 0.4655172 0.5344828
school residence    0.3103448 0.6896552
```

Then we conduct the chi-2 test:

```
> observed <- myData[1,1:2]
> expected <- myData[2,1:2]
> print(chisq.test(x = observed, p = expected))

        Chi-squared test for given probabilities

data:  observed
X-squared = 0.1125, df = 1, p-value = 0.7373
```

The p-value of 0.7373 indicates that there is a 73% chance the the null hypothesis is correct. This fairly high value means that we fail to reject the null hypothesis. In this case, that means that cannot have a high level confidence that living in the school residence has an effect on the final grade.

Conclusion

The initial analysis of our data gave us some pretty encouraging results, like the fact that grades tend to get better with the increase of the time spent studying, or that absences appear to have negative effects on grade once the student gets past a certain threshold. However, once we started conducting hypothesis tests to verify our assumptions about the data, we were pretty disappointed. Almost all tests indicated that study habits don’t have tangible effects on grades, because we got faily high p-values. The only conclusive test was the test on two proportions, which showed us that students who take more time to get to school tend to get better grades, and that was confirmed by the chi-square test.