National School of Applied Sciences of Kenitra

ENSA – Kenitra

Statistical Study

Employability of ENSA Kenitra Students

Mini-Project Report

Data Analysis using **R**

Prepared by:

- Wiame EL HABTI
- Assia EL FOUNOUN
- Ranya EL OUARD
- Ihssane EL OTMANI
- Moulaye Abdoule Hady HAIDARA

Supervised by: Pr. Aniss Moumen

Semester : 5

Department: Industrial Engineering

Academic Year: 2025 – 2026

Table of Contents

1 Definition of Data and the Problematic

1.1 Definition and Formulation of the Problematic

1.1.1 Key Terms and Definitions

In a constantly evolving global environment, marked by digital transformation, rapid technological advancements, and shifts in professional roles, employability has become a critical concern for both students and educational institutions. Employability refers to an individual's ability to obtain and maintain employment based on their skills, educational background, experience, and capacity to adapt to the dynamic demands of the labor market.

1.1.2 Central Problematic and Enumeration Questions

Today, employability serves not only as an indicator of the performance of higher education systems but also as a central challenge in preparing young graduates for sustainable professional integration. Among engineering students, employability levels vary considerably: some secure internships or professional opportunities with ease, while others, despite being competent, struggle to convince recruiters. This variability raises questions about the true influence of the different components that shape a student's profile.

1.1.3 Research Question

To gain a deeper understanding of this issue, a targeted statistical analysis is necessary. Such an analysis will identify the key factors influencing employability and model their impact on the employability levels of engineering students.

Research Questions:

1. What are the key factors that influence employability, and to what extent can these factors predict or classify students' employability levels?
2. Which methods do students perceive as effective in enhancing their employability and adapting to the current demands of the labor market?

1.1.4 Documentary Search

A documentary search was conducted using ResearchGate, Google Scholar, SCIRP, and institutional reports from the World Economic Forum.

1.1.5 Literature Review Summary

Concept	Definition	Author	Source / Link
Employability	Employability refers to the capability of an individual to gain initial employment, maintain employment, and obtain new employment when necessary.	Hillage & Pollard	https://www.researchgate.net/publication/225083565
Employability	Employability is defined as a set of achievements, skills, understandings, and personal attributes that increase graduates' chances of employment.	Yorke	https://www.researchgate.net/publication/225083582
Employability	Employability is the ability to demonstrate attributes required for effective performance within organizations.	Harvey	https://www.scirp.org/reference/referencespapers?referenceid=1720258
Technical Skills	Discipline-specific knowledge and the ability to apply technical expertise in professional contexts.	Yorke	https://www.researchgate.net/publication/225083582

Concept	Definition	Author	Source / Link
Soft Skills	Non-technical competencies such as communication, teamwork, adaptability, and leadership.	World Economic Forum	https://www.weforum.org/publications/the-future-of-jobs-report-2020/in-full/
Professional Experience	Practical exposure through internships and projects enabling applied skill development.	Hillage & Pollard	https://www.researchgate.net/publication/225083565
Professional Networking	The development of professional relationships facilitating access to employment opportunities.	Harvey	https://www.scirp.org/reference/referencespapers?referenceid=1720258
Extracurricular Activities	Non-academic activities contributing to complementary skill development.	Knight & Yorke	https://www.scirp.org/reference/referencespapers?referenceid=3190379

1.2 Data Dictionary

Variable Name	Type	Format	Description	Constraints
age	numeric	integer	student age	19-23
gender	text	choice (M/F)	student gender	required (MALE or FEMALE)
field	text	choice (Industrial Eng. / Computer Eng.)	study field	required
acad_lvl	Alphanumeric	choice (CI1 / CI2)	academic level	required
hard_sk	text	Likert scale: Strongly disagree / Disagree / Neutral / Agree / Strongly agree	technical skills	required
soft_sk	text	Likert scale: Strongly disagree / Disagree / Neutral / Agree / Strongly agree	soft skills	required

Variable Name	Type	Format	Description	Constraints
prof_exp	text	Likert scale: Strongly disagree / Disagree / Neutral / Agree / Strongly agree	professional experience	required
stg_rls	text	choice (Yes / No)	internship done	required
prof_net	text	Likert scale: Strongly disagree / Disagree / Neutral / Agree / Strongly agree	professional networking	required
extr_act	text	Likert scale: Strongly disagree / Disagree / Neutral / Agree / Strongly agree	extracurricular activities	required
field_prat	text	Likert scale: Strongly disagree / Disagree / Neutral / Agree / Strongly agree	field practice	required
skill_dev	text	Likert scale: Strongly disagree / Disagree / Neutral / Agree / Strongly agree	skill development	required

1.3 Theoretical Conceptual Model

Based on the literature review conducted in Section 1.1.5 (specifically referencing the interactive employability models), we have designed a theoretical framework to guide our investigation. This model visually represents the hypothesized relationships between various independent determinants and our dependent variable: **Student Employability**.

The model proposes that employability is not a standalone attribute but the result of the interaction between seven key dimensions:

- **Skills Set:** Technical competencies (Hard Skills) and behavioral attributes (Soft Skills).
- **Professional Exposure:** Practical experience (Internships) and Professional Network.
- **Personal Development:** Engagement in Extracurricular activities.
- **Background:** The academic curriculum (Education) and Student demographics.

Each path labeled from **H1 to H7** indicates a **research hypothesis** suggesting a direct influence of these factors on employability. This framework serves as the blueprint for designing our interview guide and survey questionnaire in the subsequent sections.

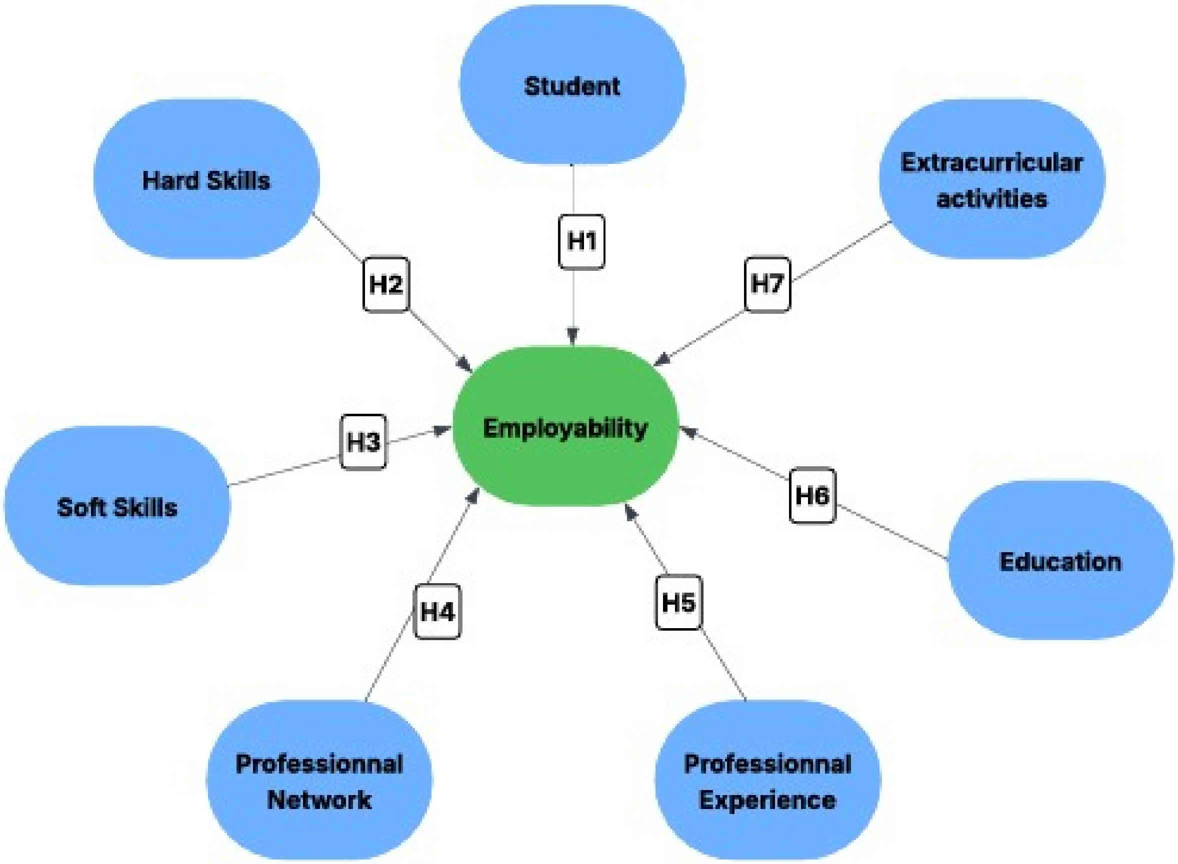


Figure 1.1: Theoretical Conceptual Model of Student Employability

2 Qualitative Study

2.0.1 Interview Guide

Interview Objectives

The purpose of these interviews is to understand the perceptions and experiences of students regarding employability. The interviews aim to:

- Identify the key factors that influence employability from the students' perspective.
- Explore the role of technical skills, soft skills, internships, networking, and extracurricular activities in professional preparation.
- Gather recommendations from students on how the institution can enhance employability support.

Interview Format and Procedure

- Semi-structured interviews conducted individually.
- Duration: 15 to 20 minutes per student.
- Questions are open-ended to allow for detailed responses and follow-up questions.
- Interviews are recorded and transcribed for qualitative analysis while ensuring confidentiality.

Target Participants

The interviews target first and second-year students in Industrial Engineering and Computer Engineering programs. A total of 5 students per program are selected to ensure diversity in perspectives.

Introduction of the Interview

The objective of this interview is to explore students' perceptions and experiences regarding employability at ENSA Kenitra. Hello, my name is [Name], and thank you for agreeing to participate in this interview as part of a mini-project on the employability of students at ENSA Kenitra. This discussion will last approximately 20 minutes. Your participation is not evaluative; the goal is to have an open discussion about your background, experiences, and your perception of employability. All your responses will remain confidential and will be used solely for academic analysis. If anything is unclear, you are free to ask for clarification, expand on your ideas, or skip a question if you prefer.

Identification of the Interviewee

To begin, each participant is asked to briefly introduce themselves by providing the following information:

- Name
- Field of study
- Year of study
- Academic or extracurricular experiences they consider relevant

Interviewee	Full Name	Gender	Major	Study Level
1	BELEMSIGRI WINDYAM JUSTE YANNEL	Male	Computer Engineering	Ci1
2	ABDELJALIL LAABID	Male	Industrial Engineering	Ci1
3	ABIR RATBAOUI	Female	Industrial Engineering	Ci2
4	KARIMA ED-DABARH	Female	Industrial Engineering	Ci1
5	FATIMA ZAHRA ELHALFI	Female	Computer Engineering	Ci1

Table 2.1: Interviewee characteristics

Core of the Interview

The interview focuses on several key areas:

1. **Definition of Employability:**

- How would you define student employability?

- What are the main factors that can improve or limit it?

2. Technical Skills (Hard Skills):

- In your opinion, what role do technical skills play in the employability of future engineers? Please give important examples.
- Which technical skills do you think you have truly mastered thanks to your academic training, and which ones do you consider insufficient or not yet fully developed?

3. Transferable Skills (Soft Skills):

- Which transferable skills (soft skills) do you consider essential?
- Do you think you are developing these skills within your current training program?

4. Influence of Academic and Extracurricular Experiences:

- How do your academic experiences (projects, internships, practical work) and extracurricular activities (associations, clubs, volunteering) influence your preparation for the job market?
- Which experience had the greatest impact on you, and why?

5. Perception of Skills Sought by Employers:

- Are the technical skills and behaviors mentioned so far the most sought after by employers, or are there other skills that are also important?

6. Obstacles to Professional Integration and Practical Experiences:

- What difficulties do you think students in your field face when trying to find a job?
- What obstacles have you personally encountered when seeking internships or practical projects?

7. Role of the Institution and Recommendations:

- Does the training and the institution's initiatives (career workshops, coaching, company forums, partnerships) sufficiently prepare you to face the challenges of the job market?
- What recommendations would you suggest to strengthen the employability of students in your field?

8. Additional Elements:

- Are there any other elements or experiences that you believe could improve students' professional integration?

Conclusion of the Interview

We are reaching the end of this interview. Is there any important point you would like to add regarding your employability or your preparation for the professional world? Acknowledgment: Thank you very much for your participation and the time you have given us.

2.0.2 Relations and Characteristics

Characteristics of the Study Population The characteristics of the study population constitute a fundamental element in analyzing employability among engineering students. This research focuses on students enrolled at the National School of Applied Sciences (ENSA) of Kénitra and explores a set of demographic, academic, and experiential variables that significantly influence employability. By examining these factors, the study aims to better understand the mechanisms through which students develop professional readiness and transition toward the labor market.

The study population is defined by the following characteristics:

- **Age:** Participants are between 19 and 23 years old, representing a crucial period for skill development, career orientation, and professional identity formation.
- **Academic Level:** The research includes students from the first and second engineering cycles (CI1 and CI2), levels that correspond to key stages in the consolidation of academic knowledge and the acquisition of employability-related competencies.
- **Field of Study:** The population consists of students specializing in Industrial Engineering and Computer Science Engineering. These disciplines present distinct professional opportunities and skill requirements, which may influence students' perceptions of employability.
- **Skills Framework:** The analysis addresses both hard skills, encompassing technical and discipline-specific competencies, and soft skills, such as communication, teamwork, leadership, adaptability, and problem-solving, which are increasingly recognized as critical determinants of employability.
- **Professional and Extracurricular Experience:** Employability is examined in relation to students' practical exposure, including internships and extracurricular or co-curricular activities, which contribute to experiential learning, professional behavior, and skill enhancement.
- **Professional Networking:** The study evaluates the role of professional networking, including engagement with industry professionals, participation in career-related events, and involvement in professional platforms, as a strategic factor in facilitating access to employment opportunities.

By integrating these population characteristics, this study situates its findings within the broader context of engineering education and employability, offering valuable insights for higher education institutions seeking to strengthen students' alignment with labor market expectations.

2.0.3 Reformulated Transcription

Question	Student	Answer
Question 1	ABDELJALIL LAABID	So, for me, students' employability is the ability of students to access the job market and also to find a job. In my opinion, the main factors that impact employability are, first of all, soft skills. Why? Because a company cannot teach you behavioral and interpersonal skills, which are soft skills. However, it can teach you practical and technical skills. These technical skills involve repetitive tasks, so the company can train you and you will learn them. Therefore, I believe that balance is essential. I do not neglect academic training; it is also very important. So, the balance between soft skills and academic education greatly helps in finding a job.
	BELEMSIGRI WINDYAM JUSTE YANNEL	I would say that student employability is the process of professional integration of students after their studies. Factors that can influence professional integration include extracurricular activities and technical skills. What can limit it is focusing solely on the theoretical side of learning.
	ABIR RATBAOUI	In my opinion, student employability corresponds to their ability to enter the job market, adapt to it, and advance professionally. It depends on several factors: First, technical skills: mastery of the tools and methods specific to the field of study. Second, transferable skills: everything related to communication, teamwork, and adaptability. Third, practical experience: internships, projects, and group work. And finally, market knowledge: understanding what companies expect and what skills are sought. It can be limited by a lack of concrete experience, a gap between theory and practice, or a lack of career guidance.

Question	Student	Answer
	KARIMA ED-DABARH	To be honest with you, actually, for me, employability means being able to enter the job market, you know, like to adapt to professional life after graduation. It depends a lot on having technical skills, of course, such as the ability to use tools, apply what we learn in class, solve practical problems, but it is also include soft skills, motivation, and the ability to learn. Like when students lack technical skills or practical experience, their employability becomes very more limited.
	FATIMA ZAHRA EL HALFI	Okay, for me, employability is the ability we have to find a job and to adapt to market needs. So, it depends on our technical skills of course, but also on our soft skills, our ability to learn, etc., communication, other factors. So, the factors that improve it, I would say internships, practical projects, languages as well, and especially computer science, staying up to date with new technologies. What limits it, maybe the lack of practical experience, or the gap between what we learn and what companies are really looking for. Honestly, there is a lack in academic training.
Question 2	ABDELJALIL LAABID	So, technical skills play an important role, of course, because they are what we use to solve problems within the factory. They also allow us to optimize work. Therefore, they represent a key area of interest for recruiters.
	BELEMSIGRI WINDYAM JUSTE YANNEL	Yes, I think it plays a very, very important role because it's what employers base their decisions on whether they're going to hire someone or not.
	ABIR RATBAOUI	Technical skills play a crucial role in the employability of future engineers, as they allow them to be immediately operational and to solve concrete problems in a business setting. In Industrial Engineering, this includes, for example, process analysis and optimization, the use of specialized IT tools and software such as advanced Excel or Python, as well as basic knowledge of production, quality, and logistics management. These skills often constitute the primary selection criterion for an employer, as they demonstrate the engineer's ability to contribute quickly and effectively to the company's projects.

Question	**Student**	**Answer**
	KARIMA ED-DABARH	As I said, I think technical skills are very important for future engineers because they allow us to understand and to solve problems. Employers expect engineers to know how to use what they learn in class. For example, basic programming, data analysis, using Excel, I don't know, understanding of your process. These are all important skills.
	FATIMA ZAHRA EL HALFI	Okay, for technical skills, they are the foundation for us as future computer engineers. Without that, we cannot do our job. Companies are looking for people who are operational quickly, who can develop and solve technical problems. For example, a computer engineer cannot be a computer engineer without knowing programming languages, the basics of programming languages, for example.
Question 3	ABDELJALIL LAABID	Thanks to my academic training, I have been able to master several technical skills, such as Python, which is an essential programming language for engineers, especially industrial engineers, as well as the C language. There are also some skills that I still need to strengthen, such as R and Excel.
	BELEMSIGRI WINDYAM JUSTE YANNEL	The skills I believe I've already acquired, as I'm a computer engineer, include website development and some basic cybersecurity knowledge, as well as some basic artificial intelligence. However, regarding the areas where I lack sufficient knowledge, I would say they are the same areas, whether it's cybersecurity or AI, because what I'm learning at school doesn't allow me to work on truly significant projects.
	ABIR RATBAOUI	Thanks to my academic background in Industrial Engineering, I believe I have truly mastered process analysis, the fundamentals of production management, and the use of certain IT and data analysis tools applied to academic projects. However, I feel that some skills remain insufficiently developed, particularly the advanced use of industrial software, the practical application of theoretical concepts in a professional setting, and exposure to real-world business cases, which require more hands-on experience and practical application in the field.

Question	Student	Answer
	KARIMA ED-DABARH	As a first-year student, I truly believe that I just have started to learn basic technical skills, like basic programming concepts and problem-solving methods. However, I feel that many technical skills are still not fully developed, especially advanced tools, you know, like the real industrial applications. I think these skills will improve with more practice and internships in the future.
	FATIMA ZAHRA EL HALFI	Honestly, I think I have well mastered the basics of programming, especially in Python, C, and algorithms, as well as fundamental concepts such as data structures, databases, and problem-solving logic. The training also allowed me to develop an ability to analyze and understand computer systems. On the other hand, what is still missing is deeper knowledge of advanced technologies used in companies, such as certain development frameworks, DevOps tools, cybersecurity, or applied artificial intelligence. I also believe that practice on real projects could be strengthened in order to gain more professional experience. Also, I think that the training remains lacking as long as we are in need of self-training.
Question 4	ABDELJALIL LAABID	For me, all soft skills, or transferable skills, are important, but the ones I consider truly essential are time management, teamwork, and communication. We must master the art of communication—the ability to convey clear and understandable messages
	BELEMSIGRI WINDYAM JUSTE YANNEL	The essential transferable skills for professional integration, I would say communication because in a company, you'll have to communicate with many people, maybe even people you don't like. In quotes, but you'll still have to communicate. The other skill is teamwork because you have to be able to work with other people, even if there are problems. There will inevitably be problems in teamwork more than in individual work. So, that's a very, very important skill, in my opinion. Yes, I would say so because my current training includes some modules on this. For example, some language modules for communication and sometimes practical projects that are given to us for teamwork.

Question	Student	Answer
	ABIR RATBAOUI	The essential transversal skills, in my opinion, are communication, teamwork, time management, adaptability, and initiative-taking. The training at ENSA contributes, on the other hand, to their development through group projects and presentations, but these skills are mainly strengthened through extracurricular activities and personal experiences.
	KARIMA ED-DABARH	Soft skills first are very crucial, which is such a massive word. In my opinion, the importance of skills, I would say, communication first, teamwork, time management, adaptability, problem solving, these all are very important soft skills, because engineers do not work alone, you know, knowing how to communicate and work in a team is very essential. I developed these skills mainly through group projects and presentations and compositions with clubs, but I think there is still room for improvement
	FATIMA ZAHRA EL HALFI	Okay. I think communication is really important, especially for working in a team and explaining technical concepts. Problem-solving as well, adaptability because technologies change very quickly in computer science. Time management to meet deadlines. In our training, well, I think we are in the middle. We have not really reached the level we want in soft skills. We have to look for it ourselves in other environments. We develop them through group projects, presentations that we sometimes do, but it is not formalized enough.
Question 5	ABDELJALIL LAABID	So, academic projects really helped me improve my technical skills. As for clubs and extracurricular projects, they also helped me develop my soft skills, such as teamwork and time management. As I mentioned earlier, both are important for integrating successfully into the professional world.

Question	Student	Answer
	BELEMSIGRI WINDYAM JUSTE YANNEL	I would say they positively influence my future integration because all these activities allow me to shape my identity. In the end, it allows me to have, I would say, an identity that is attractive to companies. Because you don't just rely on hard skills, you have to start with skills in, how to say, communication, teamwork, and others. The skill that has impressed me the most, I would say, is communication because it's a very important point. It's the skill that some people struggle with in job interviews and other situations. So it's a very important point to consider.
	ABIR RATBAOUI	Academic experiences, such as projects, internships, and practical work, play an important role in preparing me for the job market because they allow me to apply the theoretical knowledge acquired in class and to better understand the realities of the professional world. They also contribute to the development of skills such as problem-solving, teamwork, and meeting deadlines. Furthermore, extracurricular activities, particularly participation in clubs or community projects, strengthen essential transferable skills such as communication, organization, taking responsibility, and initiative, which are highly valued by employers. The experience that has most impacted me is participating in group projects, as it has taught me how to collaborate with people from different backgrounds, manage pressure, and adapt to time constraints. This experience has allowed me to better prepare myself for the demands of working in a company, where cooperation and collective efficiency are essential.

Question	Student	Answer
	KARIMA ED-DABARH	Academic projects help me, as I said before. They help me to understand how theory can be applied in practice, like how to apply what we learned in class in real life, because even small projects help us learn how to organize work and respect deadlines, like the one we're doing right now. So extracurricular activities, especially student clubs, help me also to become more confident, communicate better, and work in a team. So one experience that marked me was that I had joined a debate club, and I had the honor to be a part of it and to participate in a competition that helped me to develop my soft skills. It taught me responsibility.
	FATIMA ZAHRA EL HALFI	So, academic projects allow us to apply what we learn. For example, when we do a database project or a Python project, it gives us a concrete idea of how it works in reality. We learn to manage a project from A to Z, to work in a team, to respect deadlines. For extracurricular activities, in the EIC club for example, I developed my leadership and my time management and a little bit of soft skills. That's it. Before answering this question, I forgot to mention a project that marked me. So, what marked me the most is the Telecom Customer Churn analysis project. It is a Python project for this year, because in this project, I really understood how to exploit real data to make concrete decisions. It allowed me to apply Python for data cleaning, statistical analysis, and visualization, and to understand the importance of indicators such as churn, mean, and median in interpreting results. It also helped me develop an analytical and problem-solving-oriented approach.
Question 6	ABDELJALIL LAABID	Of course, the skills and competencies we have mentioned so far are the most sought after by employers in the job market.
	BELEMSIGRI WINDYAM JUSTE YANNEL	I would say that these are mostly the most important. But recruiters will always have something to say. Because, well, things you wouldn't expect. But the ones we've talked about are quite important. If you can master them, I think a student should be able to find a job or an internship.

Question	**Student**	**Answer**
	ABIR RATBAOUI	In my opinion, the technical and behavioral skills we have mentioned are indeed highly sought after by employers, as they form the basis of an operational engineer's profile. However, other skills are also important, including motivation, the ability to learn quickly, autonomy, and versatility. Employers are also increasingly valuing the ability to adapt to change and thrive in dynamic environments, as these qualities allow engineers to integrate quickly and contribute effectively to the company's growth.
	KARIMA ED-DABARH	In my personal opinion, I do believe that employers look for a balance between technical skills and personal qualities. Technical skills are important, of course, to perform tasks, but employers also value motivation, seriousness, communication skills, the ability to learn quickly, especially for young graduates. Companies do appreciate candidates who show willingness to improve and adapt.
	FATIMA ZAHRA EL HALFI	So, I will answer your question about the skills sought by employers. So, I think that in addition to what we said, employers are looking for autonomous people who can learn quickly and adapt. Mastery of English as well, it has become crucial, especially for reading technical documentation. Curiosity, the desire to learn new technologies by oneself. So, knowing Git or GitHub, versioning tools as well, working in agile mode. These are things we do not really learn at school but that are important in the working world. Honestly, they are very important.
Question 7	ABDELJALIL LAABID	I often hear that there is a limited number of internship opportunities. There are not many internships available, so you have to look for them as early as possible because the places are limited. Therefore, this is the major problem.
	BELEMSIGRI WINDYAM JUSTE YANNEL	The main difficulty students encounter when looking for a job, I would say, is first and foremost the job interview. And also, there's the issue of experience. For a full-time job, you're required to have a certain number of years of experience, which you don't yet have. So that's a very significant obstacle that needs to be taken into account.

Question	Student	Answer
	ABIR RATBAOUI	Students in the Industrial Engineering program may encounter several difficulties in finding employment, including a lack of professional experience upon graduation, strong competition in the job market, and a mismatch between certain skills taught and the actual needs of businesses. Furthermore, insufficient knowledge of the specific career paths available in the field and a lack of professional networks can also complicate job placement, especially for recent graduates.
	KARIMA ED-DABARH	Very interesting question, actually. It's like one of the main difficulties students could face is the lack of professional experience, obviously. They're still a student. They need some expert and some expertise to give. So, many companies ask for experience, even for entry-level positions. There's also strong competition in the job market, as you know, which makes existing jobs more difficult. Sometimes there is also a gap between what we learn at school and what companies expect. That gap is one of the hardest difficulties nowadays.
	FATIMA ZAHRA EL HALFI	So, the main problem is the lack of experience. Companies ask for two to three years of experience even for junior positions, and we, in the first year, have not even done an internship yet. There is also competition, many graduates. And sometimes, there is a gap—we learn the theoretical basics at school, which is important, but the market also looks for practical skills on specific technologies.
Question 8	ABDELJALIL LAABID	I don't have any experience with internships, so I can't speak from personal experience. But what I think, based on my perception, is that internships are very useful.

Question	**Student**	**Answer**
	BELEMSIGRI WINDYAM JUSTE YANNEL	Personally, for internships, I'd say the biggest obstacle was the follow-up. I don't want to submit my application with my CV and everything. They say they'll call you back, but they never do. So that's an important point. Now, for practical work, I'd say it's the lack of knowledge. Because what we're shown at school is very insufficient for doing a real project, I'd say. But well, that's not really a major negative point because that's exactly what engineers are looking for. So if they encounter difficulties, they search for solutions. So that's it.
	ABIR RATBAOUI	The main obstacles I personally encountered when trying to secure internships or practical projects were primarily the difficulty in finding opportunities, especially for first-year engineering students, and the lack of professional experience, which can be a hindrance when faced with company requirements. This is sometimes compounded by a limited professional network and a small number of partner companies, making access to internships and practical projects more competitive.
	KARIMA ED-DABARH	As a CI1 student in industrial engineering and the cycle of industrial engineering, I could say that it is often difficult to find internships because companies usually prefer students in higher years. In addition, students sometimes lack information or guidance. Guidance on how to apply, how to write a CV, or how to prepare for an interview. This can make the process more challenging.
	FATIMA ZAHRA EL HALFI	For internships, it is really difficult, especially in the first year. Companies prefer to take third-year or fourth-year students, those who have more experience and knowledge. For example, this week, we had a company forum. So, we see that companies mostly ask for people who are in the third year of the engineering cycle. Many do not even respond to applications. It is also a question of network—if you do not know anyone, it is complicated to get a foot in the professional environment. In addition, some companies already ask for knowledge of specific technologies that we have not yet learned.

Question	**Student**	**Answer**
Question 9	ABDELJALIL LAABID	So, in my opinion, I believe that the institution offers useful activities such as forums, conferences, and clubs, which help us a lot in developing our skills. However, I also recommend establishing partnerships with companies.
	BELEMSIGRI WINDYAM JUSTE YANNEL	Sufficiently, I'd say, I have some doubts. I'd say they prepare us to about 80 percent. Because we mustn't neglect what the institution does, because it's very important. Through the forums, we get to meet companies. We're already preparing for interviews and other things. So I'd say, it helps, it helps a lot. To strengthen the employability of students in my field, I'd say the curriculum should be better adapted to market demand. Because, for example, we take modules that companies in the IT sector don't require. And in my opinion, it's a bit of a waste of time. Because we could have used that time to learn what companies are looking for. So in my opinion, the curriculum needs to be reviewed.
	ABIR RATBAOUI	In my opinion, the training and initiatives implemented by the institution, such as career workshops, coaching, and company forums, contribute positively to preparing students for the job market, but they remain only partially sufficient to meet the current demands of companies. These initiatives primarily serve to raise students' awareness of professional expectations; however, their impact could be strengthened by a more practical and ongoing approach. To further improve the employability of Industrial Engineering students, I would recommend strengthening partnerships with companies, increasing internship and practical project opportunities, and regularly organizing workshops to prepare students for interviews, CVs, and career guidance. More personalized support and a better connection between academic training and the real needs of the market would allow students to better overcome the challenges of professional integration.

Question	**Student**	**Answer**
	KARIMA ED-DABARH	I think that ENSAK, The National School of Applied Sciences, gives us a good theoretical base. But as I said, theoretical base. So, since we are in the first year, we still need more practical exposure. We need real-life experience. I would recommend more workshops, like company visits, practical projects, like this one that we're working on. This is all a way to shape and grow us as students to better understand the professional world.
	FATIMA ZAHRA EL HALFI	So, the institution makes efforts, there are company forums sometimes, but honestly, I think it is not enough. There should be more partnerships with IT companies to facilitate internships, even for first-year students, even if it is just for observation. And why not projects in collaboration with real companies, or introducing more modern technologies in courses, such as frameworks, web development, Docker, Git, GitHub, etc. Things that we really use in companies and that we really need to learn.
Question 10	ABDELJALIL LAABID	So, in my opinion, I believe that the institution offers useful activities such as forums, conferences, and clubs, which help us a lot in developing our skills. However, I also recommend establishing partnerships with companies.
	BELEMSIGRI WINDYAM JUSTE YANNEL	Yes, there's another element, namely networking. Because you need to know how to, how to put it, make connections. And not just casual connections, but connections that can help you shape your future work environment. For example, you might meet an HR professional today. And in the future, it's that same HR professional who will, I'd say, pull some strings for you so you can get a job easily.

Question	Student	Answer
	ABIR RATBAOUI	Yes, several factors could further improve students' professional integration, including participation in supplementary training and recognized certifications, which strengthen technical skills. Developing a professional network through events, forums, and contact with professionals is also essential. Furthermore, participation in real-world projects in collaboration with companies, as well as involvement in extracurricular activities, can help students gain practical experience and better prepare for the demands of the job market.
	KARIMA ED-DABARH	To sum up, I would say the online courses, participation, certifications. When I mention participation, I mean like as a student, try to participate in activities, extracurricular activities that can strongly improve your employability. Because these experiences help students to develop both at the same time technical and soft skills. And, of course, gain more confidence in themselves.
	FATIMA ZAHRA EL HALFI	So, I think a mentoring system with former students who are already working would be very useful, to guide us and advise us. Encouraging students to do personal projects, to contribute to open source, to create a portfolio on GitHub to show their skills. Developing our professional network from now on via LinkedIn, participating in tech events, etc. And maybe integrating recognized certifications into our curriculum, or at least informing us about which ones to take.
Conclusion	ABDELJALIL LAABID	R.A.S
	BELEMSIGRI WINDYAM JUSTE YANNEL	Yes, the key point is that you really need to develop your skills, technical skills. Because companies, they first look for candidates based on technical skills. They filter candidates based on technical skills first. And then, when they hire you, they conduct the interview to assess your soft skills. So it's an important point. You need to prepare through certifications like Coursera and others, which demonstrate that you have at least a research mindset. You're going beyond what you're taught in school.
	ABIR RATBAOUI	R.A.S

Question	Student	Answer
	KARIMA ED-DABARH	So, to conclude, I believe that employability develops the student. But it should not just develop the student as who he is in the class, but also in the real world. Because we have technical skills, and we have soft skills, and we have also practical experience. All this shapes the student employability. So, as a first year student, I'm still learning. But I think the continuous effort and experience will help me to become more better prepared for the job market.
	FATIMA ZAHRA EL HALFI	So, to conclude, I would just like to add that even if we lack professional experience for now, employability is something that we build progressively. We have to be proactive, learn by ourselves outside of classes, and prepare ourselves from now on for the job market, that's all I think.

2.0.4 Analysis Guide of Ideas in the Temper Student

Saturation Grid - Themes per Student

THEMES / KEY IDEAS	Abdeljalil	Belemsigri	Abir	Karima	Fatima	Saturation
1. DEFINITION & FACTORS						
Access to Job Market	X	X	X	X	X	**5**
Professional Integration	X	X		X		**3**
Importance of Adaptability			X	X	X	**3**
Balance between Tech & Soft Skills	X		X	X		**3**
2. TECHNICAL SKILLS (HARD SKILLS)						
Role: Solving Problems	X		X	X	X	**4**
Role: Selection Filter	X	X	X		X	**4**
Programming Skills (Python, C, Web)	X	X	X	X	X	**5**
Data Analysis Skills (Excel, Data)	X		X	X	X	**4**
Need for Quick Operationality			X		X	**2**
3. SOFT SKILLS						
Communication	X	X	X	X	X	**5**
Teamwork	X	X	X	X		**4**
Conflict Resolution		X				**1**

THEMES / KEY IDEAS	Abdeljalil	Belemsigri	Abir	Karima	Fatima	Saturation
Autonomy		X	X	X	X	**4**
Ability to Learn			X	X	X	**3**
Time Management	X		X	X		**3**
4. EXPERIENCE & LEARNING						
Impact of Clubs	X	X	X	X	X	**5**
Academic Projects	X	X	X	X	X	**5**
Self-learning		X	X	X	X	**4**
Certifications		X	X	X		**3**
5. OBSTACLES & DIFFICULTIES						
Gap between Theory and Practice		X	X	X	X	**4**
Lack of Experience (Junior Paradox)		X	X	X	X	**4**
Scarcity of Internships	X	X	X	X	X	**5**
Lack of Professional Network		X	X		X	**3**
Ghosting / No Response		X			X	**2**
6. SOLUTIONS & RECOMMENDATIONS						
Corporate Partnerships	X		X	X	X	**4**
Company Visits				X	X	**2**
Interview Simulations			X	X		**2**
Modern Tech Updates (Cloud, DevOps)		X			X	**2**
TOTAL THEMES MENTIONED	**14**	**18**	**23**	**22**	**21**	–

2.0.5 Analysis of Consensus Density

Data Integrity Check: We first verified the structural consistency of the grid by calculating the ratio between the sum of saturations and the sum of totals per student: $\frac{98}{98} = \mathbf{100\%}$. This confirms that all items are present and fully accounted for. To evaluate the homogeneity of the responses, we calculated

the **Majority Consensus Rate**, focusing on themes shared by at least three students ($S \geq 3$).

$$\text{Consensus Rate} = \frac{\sum_{S\geq 3} \text{Saturation}}{\text{Total Themes}}$$

Note: The numerator (87) corresponds to the sum of saturation scores for all themes mentioned by a majority of participants (3 or more).

$$\text{Rate} = \frac{98 - 11}{98} = \frac{87}{98} \approx \mathbf{88.8\%}$$

The analysis reveals that **88.8%** of the discourse is focused on themes shared by the majority. This demonstrates a strong homogeneity in the students' perception of employability challenges.

2.0.6 Code Saturation

This script was developed to analyze qualitative interview data and assess whether theoretical saturation was reached in the study on student employability. First, all interview transcripts were collected and organized so that each interview corresponded to one student. The text data were then preprocessed using text mining techniques, including converting text to lowercase, removing numbers, punctuation, stopwords, and extra whitespace, in order to retain only meaningful words.

To evaluate saturation, interviews were added progressively, and the cumulative number of unique words was calculated after each additional interview. This allowed us to observe how the vocabulary evolved as more data were included. A saturation curve was then plotted to visually assess whether the number of new words decreased over time. A flattening curve indicated that new interviews contributed less new information, suggesting that saturation was being approached.

In addition, a saturation percentage was calculated by comparing the number of new words introduced by the last interview to the total number of unique words. This quantitative indicator helped determine whether the data collection was sufficient or if more interviews were needed.

Finally, a keyword frequency analysis was conducted across all interviews to identify the most frequently mentioned terms and highlight the main themes related to employability. The results were exported to a CSV file for further analysis. Overall, this process allowed us to objectively evaluate data saturation and extract key qualitative insights from the interviews

```r
#/////////////// QUALITATIVE STUDY ///////////////////////////////////////////////////////////////////
# ==========================================================================
# SATURATION ANALYSIS FOR QUALITATIVE INTERVIEW DATA
# ==========================================================================
# PURPOSE: This script analyzes interview transcriptions to determine if we've
#          reached "theoretical saturation" - the point where new interviews
#          stop adding new information to our research.
#
# WHAT IS SATURATION?
# - In qualitative research, saturation means we've collected enough data
# - It happens when new interviews repeat the same themes/concepts
# ==========================================================================


# ==========================================================================
# SECTION 1: INSTALL AND LOAD REQUIRED PACKAGES
# ==========================================================================
# We only need to install them ONCE, but load them EVERY TIME we run the script

# The if(!require()) structure means: "If not already installed, then install"
if(!require(tm)) install.packages("tm")                   # tm = Text Mining package
if(!require(SnowballC)) install.packages("SnowballC")     # For word stemming (finding root words)
if(!require(wordcloud)) install.packages("wordcloud")     # For creating word clouds (visual representations)
if(!require(RColorBrewer)) install.packages("RColorBrewer") # For color palettes in visualizations

# Load the packages into memory so we can use their functions
library(tm)             # Main text mining toolkit
library(wordcloud)      # For visual word representations
library(RColorBrewer)   # For choosing nice colors
library(ggplot2)        # For creating professional-looking graphs
```

Figure 2.1: R Script Used for Saturation Analysis

```r
# ==========================================================================
# SECTION 2: INSERT YOUR INTERVIEW DATA
# ==========================================================================

# RULES FOR FORMATTING YOUR DATA:
# 1. Each student's COMPLETE interview goes in quotes "..."
# 2. Separate different students with commas
# 3. Keep all responses from ONE student together in ONE text block

mes_donnees <- c(

  "Okay, for me, employability is the ability we have to find a job and to adapt to market needs. So, it depends on our t
our soft skills, our ability to learn, etc., communication, other factors. So, the factors that improve it, I would say i
languages as well, and especially computer science, staying up to date with new technologies. What limits it, maybe the l
between what we learn and what companies are really looking for. Honestly, there is a lack in academic training.
Okay, for technical skills, they are the foundation for us as future computer engineers. Without that, we cannot do our j
who are operational quickly, who can develop and solve technical problems. For example, a computer engineer cannot be a c
programming languages, the basics of programming languages, for example.
Honestly, I think I have well mastered the basics of programming, especially in Python, C, and algorithms, as well as fur
structures, databases, and problem-solving logic. The training also allowed me to develop an ability to analyze and under
hand, what is still missing is deeper knowledge of advanced technologies used in companies, such as certain development f
or applied artificial intelligence. I also believe that practice on real projects could be strengthened in order to gain
Also, I think that the training remains lacking as long as we are in need of self-training.
Okay. I think communication is really important, especially for working in a team and explaining technical concepts. Prob
because technologies change very quickly in computer science. Time management to meet deadlines. In our training, well, I
We have not really reached the level we want in soft skills. We have to look for it ourselves in other environments. We c
presentations that we sometimes do, but it is not formalized enough.
So, academic projects allow us to apply what we learn. For example, when we do a database project or a Python project, it
works in reality. We learn to manage a project from A to Z, to work in a team, to respect deadlines. For extracurricular
I developed my leadership and my time management and a little bit of soft skills. That's it. Before answering this questi
```

Figure 2.2: R Script Used for Saturation Analysis

```r
# ==========================================================================
# SECTION 3: CALCULATE SATURATION LEVEL
# ==========================================================================
# This is the CORE of the analysis - we calculate how vocabulary evolves
# as we add more interviews, to see if we're reaching saturation
#
# HOW IT WORKS:
# - We start with Student 1's interview and count unique words
# - Then we add Student 2's interview and count unique words again
# - Then Students 1+2+3, then 1+2+3+4, etc.
# - If the number of NEW words keeps decreasing, we're reaching saturation

# STEP 3.1: Prepare an empty storage vector
# This will store the cumulative unique word count after each interview
mots_uniques_cumules <- numeric(length(mes_donnees))


# STEP 3.2: Start the main loop - this processes interviews one by one
# The loop runs from interview 1 to the total number of interviews
for(i in 1:length(mes_donnees)){


  # STEP 3.2a: Combine interviews from Student 1 up to Student i
  # paste() combines text, collapse=" " puts a space between them
  texte_groupe <- paste(mes_donnees[1:i], collapse = " ")
  #"BONJOUR" "BONSOIR" "BONJOUR BONSOIR"
```

Figure 2.3: R Script Used for Saturation Analysis

```r
  # STEP 3.2b: Clean and process the combined text
  # We create a "corpus" - a collection of text documents for analysis
  corpus_temp <- Corpus(VectorSource(texte_groupe))
  corpus_temp <- tm_map(corpus_temp, content_transformer(tolower))
  corpus_temp <- tm_map(corpus_temp, removeNumbers)
  corpus_temp <- tm_map(corpus_temp, removeWords, stopwords("english"))
  corpus_temp <- tm_map(corpus_temp, removePunctuation)
  corpus_temp <- tm_map(corpus_temp, stripWhitespace)



  # STEP 3.2c: Count the unique words
  # Create a Term-Document Matrix - a table showing which words appear
  dtm_temp <- TermDocumentMatrix(corpus_temp)

  # Convert to a regular matrix (table format)
  m_temp <- as.matrix(dtm_temp)

  # Count the number of rows = number of unique words
  # Store this count in our results vector at position i
  mots_uniques_cumules[i] <- nrow(m_temp)
}
```

Figure 2.4: R Script Used for Saturation Analysis

```r
# ==============================================================================
# SECTION 4: CREATE THE SATURATION CURVE VISUALIZATION
# ==============================================================================
# This graph shows how the vocabulary grows as we add more interviews
# If the line starts to flatten out, we've reached saturation

# 3. Affichage du Graphique de Saturation
plot(mots_uniques_cumules,            # The data to plot (our unique word counts)
     type="b",                        # "b" = both lines and points
     pch=19,                          # Point character: 19 = filled circle
     col="blue",                      # Color of the line and points
     lwd=2,                           # Line width: 2 = thicker line
     main="Saturation Curve (Vocabulary Evolution)",  # Graph title
     xlab="Number of Students Interviewed",           # X-axis label
     ylab="Cumulative Unique Words")                  # Y-axis label

# Add a grid to make it easier to read values
grid()

# INTERPRETATION OF THE GRAPH:
# - STEEP upward line = lots of new words being added (NOT saturated)
# - FLATTENING line = fewer new words being added (APPROACHING saturation)
# - HORIZONTAL line = almost no new words (SATURATED)
```

Figure 2.5: R Script Used for Saturation Analysis

```r
# ==========================================================================
# SECTION 5: CALCULATE AND DISPLAY SATURATION PERCENTAGE
# ==========================================================================
# This calculates exactly HOW saturated your data is using percentages

# STEP 5.1: Get the unique word count BEFORE the last student
# length(mots_uniques_cumules)-1 means "second to last position"
# Example: If you have 5 students, this gets the count after Student 4
mots_avant_dernier <- mots_uniques_cumules[length(mots_uniques_cumules)-1]

# STEP 5.2: Get the TOTAL unique word count (after all students)
# This is the final count after including all interviews
mots_total <- mots_uniques_cumules[length(mots_uniques_cumules)]

# STEP 5.3: Calculate how many NEW words the last student added
# Example: If Student 4 had 601 words and Student 5 has 635 words total
#          Then Student 5 added: 635 - 601 = 34 new words
nouveaux_mots <- mots_total - mots_avant_dernier

# STEP 5.4: Calculate what PERCENTAGE of total words are new from last student
# Formula: (new words / total words) × 100
pourcentage_nouveaute <- (nouveaux_mots / mots_total) * 100

# STEP 5.5: Calculate the saturation level
# Saturation = 100% - Percentage of new words
niveau_saturation <- 100 - pourcentage_nouveaute
```

Figure 2.6: R Script Used for Saturation Analysis

```r
# STEP 6.3: Calculate word frequencies
# Create a Term-Document Matrix for all documents
dtm <- TermDocumentMatrix(docs)

# Convert to matrix and calculate total frequency of each word
m <- as.matrix(dtm)

# rowSums() adds up how many times each word appears across ALL interviews
# sort() arranges them from most frequent to least frequent
v <- sort(rowSums(m), decreasing=TRUE)

# Create a data frame (like an Excel table) with words and their frequencies
d <- data.frame(word = names(v), freq=v)

# STEP 6.4: Display the top 20 most common words
cat("TOP 20 MOST FREQUENTLY USED WORDS:\n")
print(head(d, 50))  # head() shows the first 20 rows

# STEP 6.5: Export results to a CSV file
# This creates an Excel-compatible file you can open and analyze further

write.csv(d, "keyword_frequency_results.csv", row.names = FALSE)
cat("\nFile 'keyword_frequency_results.csv' successfully created.\n")
```

Figure 2.7: R Script Used for Saturation Analysis

```
# ================================================================================
# ANALYSIS COMPLETE!
# ================================================================================
cat("\n========== SUMMARY OF WHAT WE DID ==========\n")
cat("1. Loaded", length(mes_donnees), "student interview transcriptions\n")
cat("2. Processed text to extract meaningful words\n")
cat("3. Calculated cumulative unique words after each interview\n")
cat("4. Created a saturation curve visualization\n")
cat("5. Determined saturation level:", round(niveau_saturation, 2), "%\n")
cat("6. Identified most frequent keywords across all interviews\n")
cat("7. Exported detailed results to CSV file\n")
cat("=============================================\n")
```

Figure 2.8: R Script Used for Saturation Analysis

```
+
+   # STEP 3.2c: Count the unique words
+   # Create a Term-Document Matrix - a table showing which words appear
+   dtm_temp <- TermDocumentMatrix(corpus_temp)
+
+   # Convert to a regular matrix (table format)
+   m_temp <- as.matrix(dtm_temp)
+
+   # Count the number of rows = number of unique words
+   # Store this count in our results vector at position i
+   mots_uniques_cumules[i] <- nrow(m_temp)
+ }
Il y a eu 25 avis (utilisez warnings() pour les visionner)
>
```

Figure 2.9: Result of Saturation Analysis

```
>
> # STEP 5.6: Display results in the console (R's output window)
> cat("\n================ SATURATION RESULTS ================\n")

================ SATURATION RESULTS ================
> cat("The last student contributed:", round(pourcentage_nouveaute, 2), "% new content.\n")
The last student contributed: 10.9 % new content.
> cat("SATURATION LEVEL:", round(niveau_saturation, 2), "%\n")
SATURATION LEVEL: 89.1 %
> cat("====================================================\n\n")
====================================================
```

Figure 2.10: Result of Saturation Analysis

```
>
> # STEP 6.2: Clean the entire corpus (same steps as before)
> docs <- tm_map(docs, content_transformer(tolower))        # Lowercase
Message d'avis :
Dans tm_map.SimpleCorpus(docs, content_transformer(tolower)) :
  transformation drops documents
> docs <- tm_map(docs, removeNumbers)                       # Remove numbers
Message d'avis :
Dans tm_map.SimpleCorpus(docs, removeNumbers) :
  transformation drops documents
> docs <- tm_map(docs, removeWords, stopwords("english"))  # Remove common words
Message d'avis :
Dans tm_map.SimpleCorpus(docs, removeWords, stopwords("english")) :
  transformation drops documents
> docs <- tm_map(docs, removePunctuation)                   # Remove punctuation
Message d'avis :
Dans tm_map.SimpleCorpus(docs, removePunctuation) :
  transformation drops documents
> docs <- tm_map(docs, stripWhitespace)                     # Remove extra spaces
Message d'avis :
Dans tm_map.SimpleCorpus(docs, stripWhitespace) :
  transformation drops documents
```

Figure 2.11: Result of Saturation Analysis

```
> # STEP 6.4: Display the top 20 most common words
> cat("TOP 20 MOST FREQUENTLY USED WORDS:\n")
TOP 20 MOST FREQUENTLY USED WORDS:
> print(head(d, 50))  # head() shows the first 20 rows
                       word freq
skills               skills   64
technical         technical   26
also                   also   25
say                     say   24
important         important   23
experience       experience   22
companies         companies   21
projects           projects   21
learn                 learn   20
job                     job   18
think                 think   16
practical         practical   15
professional   professional   15
soft                   soft   15
students           students   14
can                     can   13
employability employability   13
like                   like   13
work                   work   13
ability             ability   12
communication communication   12
market               market   12
well                   well   12
student             student   12
engineers         engineers   11
first                 first   11
internships     internships   11
```

Figure 2.12: Result of Saturation Analysis

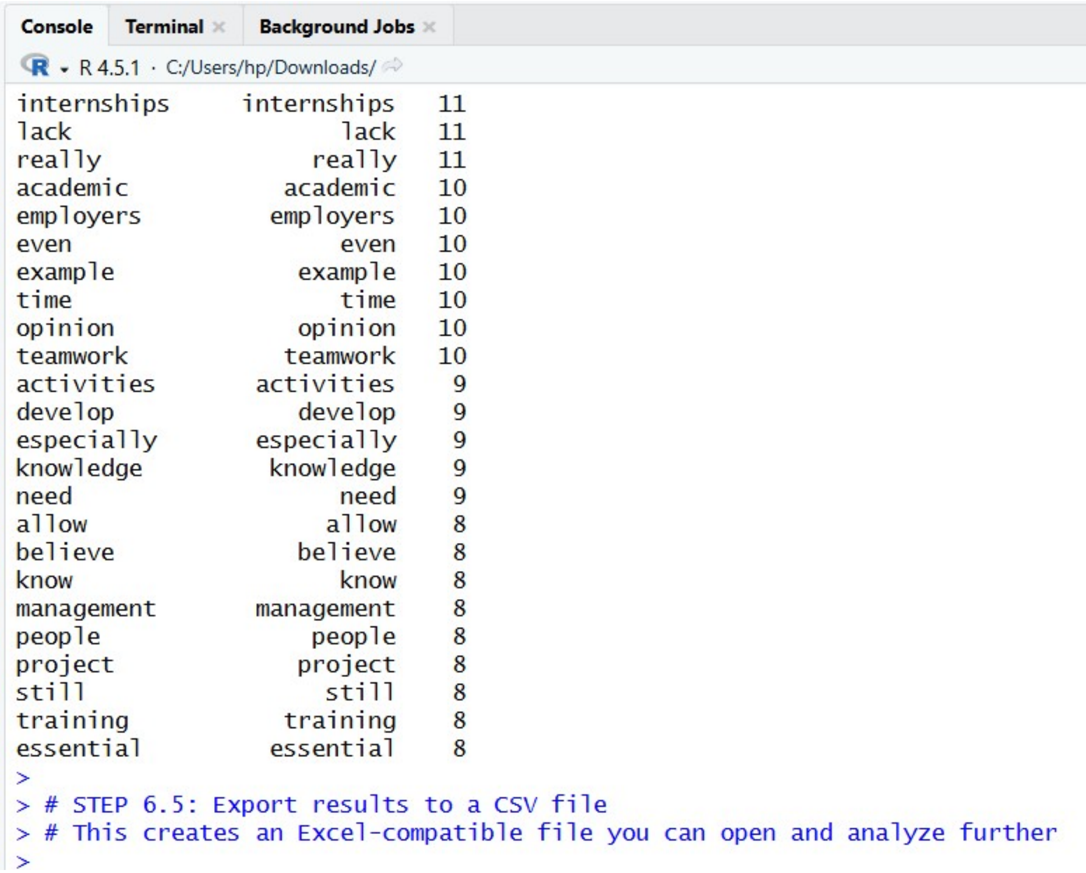
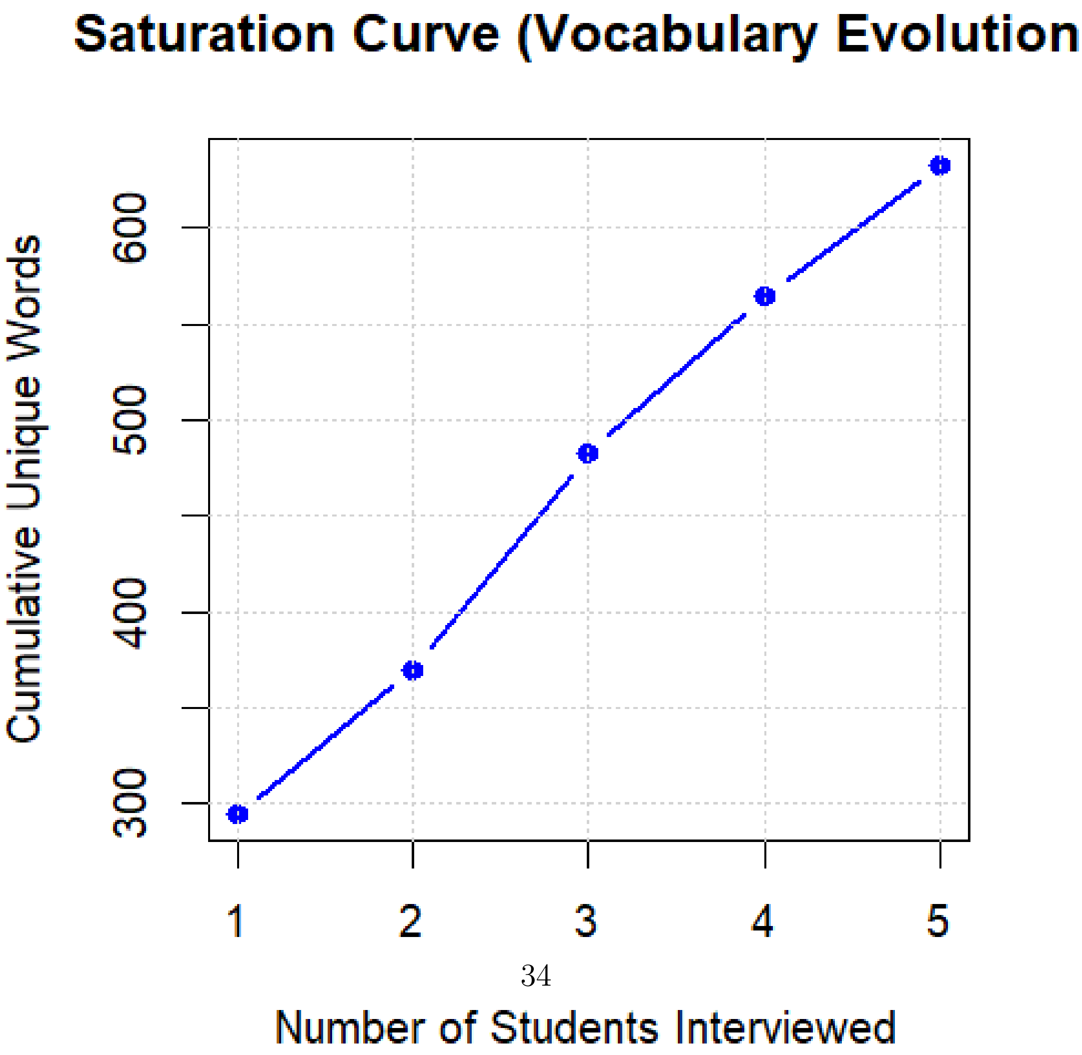


Figure 2.13: Result of Saturation Analysis

```
>
> # STEP 6.5: Export results to a CSV file
> # This creates an Excel-compatible file you can open and analyze further
>
> write.csv(d, "keyword_frequency_results.csv", row.names = FALSE)
> cat("\nFile 'keyword_frequency_results.csv' successfully created.\n")

File 'keyword_frequency_results.csv' successfully created.
>
>
> # ============================================================================
> # ANALYSIS COMPLETE!
> # ============================================================================
> cat("\n========== SUMMARY OF WHAT WE DID ==========\n")

========== SUMMARY OF WHAT WE DID ==========
> cat("1. Loaded", length(mes_donnees), "student interview transcriptions\n")
1. Loaded 5 student interview transcriptions
> cat("2. Processed text to extract meaningful words\n")
2. Processed text to extract meaningful words
> cat("3. Calculated cumulative unique words after each interview\n")
3. Calculated cumulative unique words after each interview
> cat("4. Created a saturation curve visualization\n")
4. Created a saturation curve visualization
> cat("5. Determined saturation level:", round(niveau_saturation, 2), "%\n")
5. Determined saturation level: 89.1 %
> cat("6. Identified most frequent keywords across all interviews\n")
6. Identified most frequent keywords across all interviews
> cat("7. Exported detailed results to CSV file\n")
7. Exported detailed results to CSV file
> cat("============================================\n")
============================================
```

Figure 2.14: Result of Saturation Analysis



2.1 Contextualized Research Model

The **Contextualized Research Model (MCC)** represents the empirical evolution of our initial theoretical framework. By integrating the qualitative findings from the semi-structured interviews, we refined the model to better fit the specific context of engineering students at ENSA Kenitra.

Major adjustments from the theoretical model include:

- **Specification of Variables:** Broad concepts were broken down into measurable components (e.g., distinguishing 'Academic Education' from 'Self-learning').
- **Integration of Barriers:** The qualitative discourse highlighted critical hurdles, leading to the inclusion of the 'Obstacles' dimension (Hypotheses H12, H13, and H14), such as the gap between academic training and professional practice.
- **Moderating Factors:** Student characteristics (Gender, Level, Field) are now explicitly linked to employability (Hypotheses H1 to H4) to measure their specific impact.

This model, presented in Figure 2.16, serves as the foundation for the statistical hypotheses that will be tested in the following quantitative phase.

Figure 2.16: Contextualized Research Model (Derived from Qualitative Results)

3 Quantitative Study

3.1 Introduction

The transition from academic training to the professional world represents a pivotal moment for future engineers. In an increasingly competitive market, the possession of a technical diploma is no longer the sole guarantor of successful professional integration.

The Scope: Defining Employability

This study investigates how students perceive their readiness (Hard & Soft Skills) and how specific factors influence this perception:

- **Demographics:** Age, Gender.
- **Academic Profile:** Specialty, Academic Level.
- **Experience:** Internship history & Familiarity with market trends.

Study Objective

By analyzing student responses through a quantitative approach, we aim to identify **statistical trends** that reveal the current "employability profile" of ENSA Kenitra students and pinpoint gaps between curriculum and recruiter expectations.

3.2 Methodology

The following section provides a visual overview of the questionnaire deployed in this study to evaluate student perceptions of employability at ENSA Kenitra.

These screenshots detail the specific survey items and the Likert scale adopted for data collection, offering insight into how student responses were structured and measured.
Note: It is important to mention that while specific obstacles were discussed during the interviews, they were not explicitly included in this questionnaire.

3.2.1 Questionnaire

Data were collected using a structured questionnaire designed to measure the main factors related to students' employability. The questionnaire covers several dimensions such as technical skills, soft skills, professional experience, networking, and participation in extracurricular activities, which are directly linked to the hypotheses of the study. The survey was administered online through a student forum, allowing for efficient data collection and standardized responses. Most questions are closed-ended in order to facilitate quantitative analysis.

The following figures present screenshots of the questionnaire used in this study.

Preamble

Respondent Identification

Âge : *

Réponse courte

Sexe : *

- ○ H
- ○ F

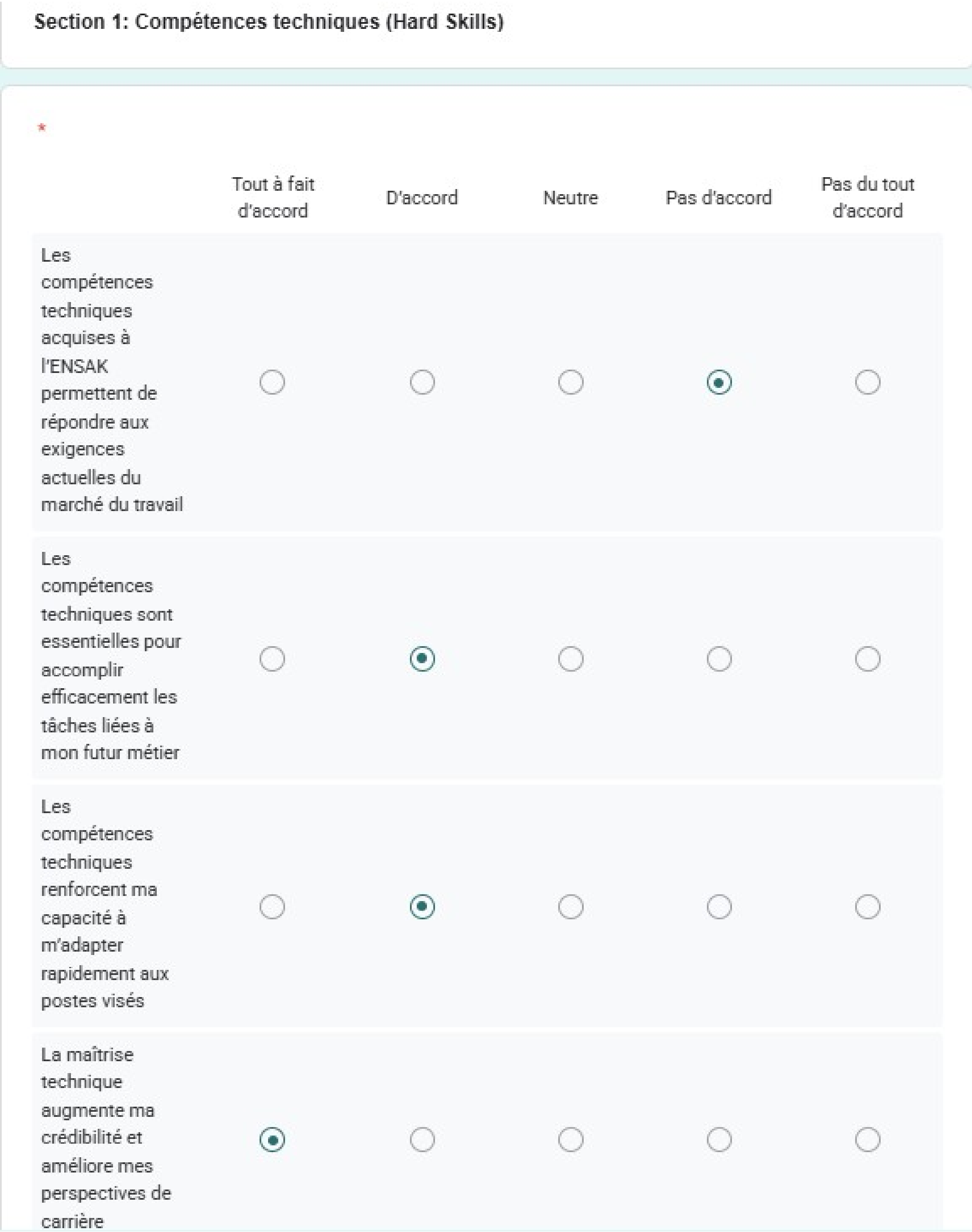
Filière : *

- ○ Génie Industriel
- ○ Génie Informatique

Niveau académique : *

- ○ CI1
- ○ CI2

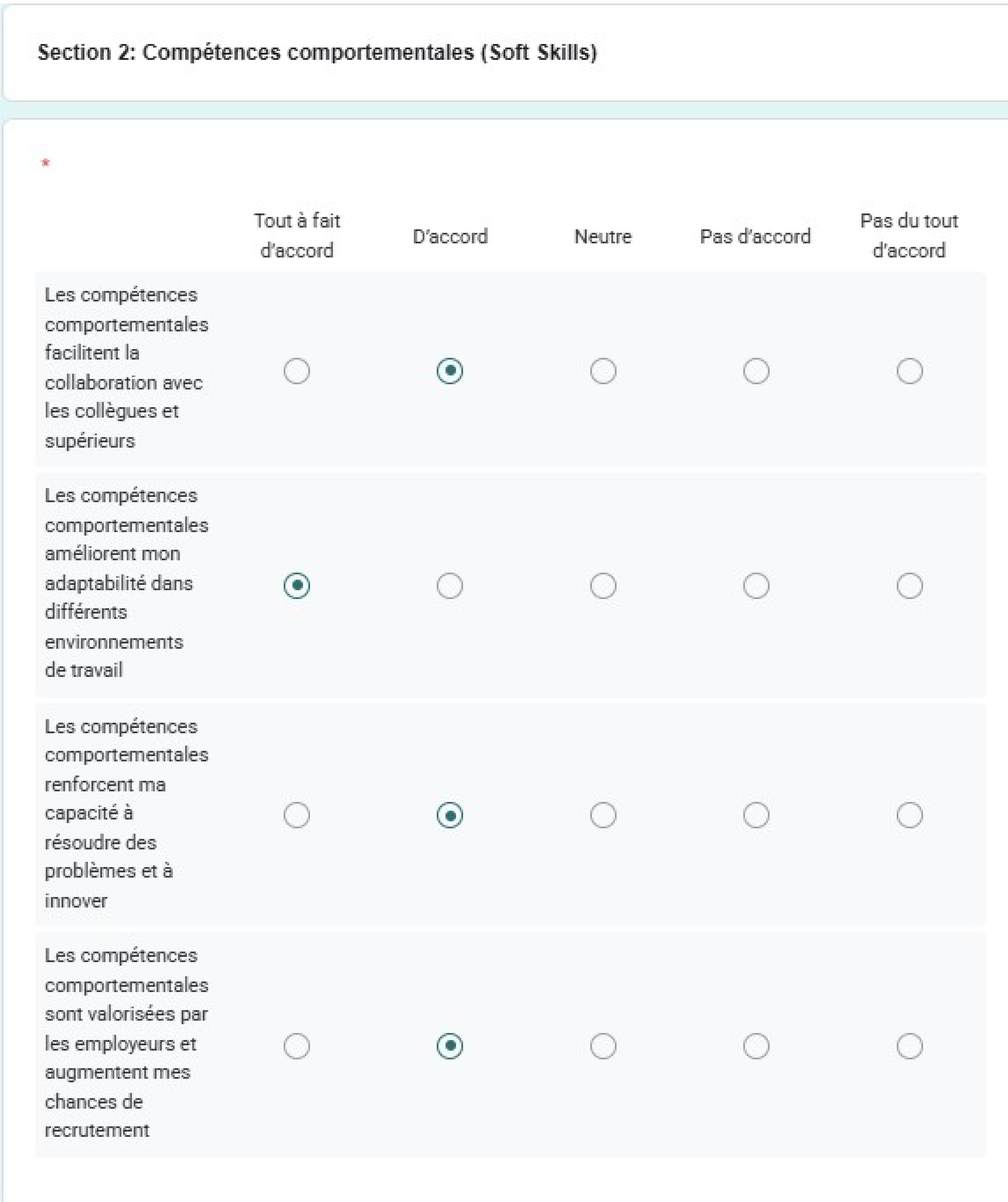
Set of Questions



Section 1: Compétences techniques (Hard Skills)

*

	Tout à fait d'accord	D'accord	Neutre	Pas d'accord	Pas du tout d'accord
Les compétences techniques acquises à l'ENSAK permettent de répondre aux exigences actuelles du marché du travail	○	○	○	◉	○
Les compétences techniques sont essentielles pour accomplir efficacement les tâches liées à mon futur métier	○	◉	○	○	○
Les compétences techniques renforcent ma capacité à m'adapter rapidement aux postes visés	○	◉	○	○	○
La maîtrise technique augmente ma crédibilité et améliore mes perspectives de carrière	◉	○	○	○	○



Section 2: Compétences comportementales (Soft Skills)

*

	Tout à fait d'accord	D'accord	Neutre	Pas d'accord	Pas du tout d'accord
Les compétences comportementales facilitent la collaboration avec les collègues et supérieurs	○	◉	○	○	○
Les compétences comportementales améliorent mon adaptabilité dans différents environnements de travail	◉	○	○	○	○
Les compétences comportementales renforcent ma capacité à résoudre des problèmes et à innover	○	◉	○	○	○
Les compétences comportementales sont valorisées par les employeurs et augmentent mes chances de recrutement	○	◉	○	○	○

Section 3: Expérience professionnelle

*

	Tout à fait d'accord	D'accord	Neutre	Pas d'accord	Pas du tout d'accord
Les stages ou projets pratiques contribuent fortement à ma préparation au marché du travail	○	◉	○	○	○
La pratique sur le terrain me permet d'identifier mes forces et mes axes d'amélioration avant d'entrer sur le marché du travail.	○	◉	○	○	○
L'expérience professionnelle permet de développer mes compétences et mon réseau professionnel	○	◉	○	○	○
Les expériences pratiques me donnent confiance pour postuler et réussir dans mon futur emploi	◉	○	○	○	○

Section 4: Réseautage professionnel

*

	Tout à fait d'accord	D'accord	Neutre	Pas d'accord	Pas du tout d'accord
Le réseautage facilite l'accès à des opportunités professionnelles souvent non publiées	○	●	○	○	○
Les contacts professionnels aident à obtenir des recommandations et des conseils utiles pour ma carrière	○	○	○	●	○
Développer un réseau professionnel est essentiel pour mieux comprendre le marché du travail et ses attentes	○	●	○	○	○
Maintenir des relations professionnelles favorise ma visibilité et ma crédibilité dans le secteur.	○	○	○	●	○

Section 5: Activités parascolaires et développement de compétences

Participez-vous à des activités parascolaires (clubs, associations, projets étudiants, sports,ateliers, etc.)? *

◉ Oui

○ Non

Si oui, quelles activités parascolaires pratiquez-vous ?

*

	Tout à fait d'accord	D'accord	Neutre	Pas d'accord	Pas du tout d'accord
Ces activités facilitent votre préparation à l'insertion professionnelle	○	◉	○	○	○
Ces activités parascolaires contribuent à développer vos compétences professionnelles	○	○	○	◉	○
Participer à des activités parascolaires améliore ma capacité à travailler en équipe et à gérer des projets	◉	○	○	○	○
Les expériences acquises lors des activités parascolaires complètent et enrichissent ma formation académique	○	◉	○	○	○

3.3 Sampling Method and Sample Size

3.3.1 Sampling Method

A. Target Population

The studied population includes :

a. Industrial Engineering: CI1 and CI2

b. Computer Engineering: CI1

Field	Level	Total Population
Industrial Engineering	CI1	67
Industrial Engineering	CI2	70
Computer Engineering	CI1	67
	Total	**204**

Note: CI refers to the Engineering Cycle.

B. Methodology

Chosen method: Stratified sampling

Reasons for the choice:

a. Each subgroup (level and field of study) is proportionally represented.

b. Allows for rigorous comparison of perceptions across levels and fields of study.

c. Reduces the risk of bias related to the overrepresentation of a dominant group.

D. Sample Size

a. Total population (N): 204

b. Confidence interval (CI): 90%

c. Margin of error (ME): 8%

d. Target sample size (n): 71 students

Method : online calculator

C. Proportional Allocation by Stratum

a. Each stratum is a subgroup of the population (here: combination of field of study × level).

b. The proportion of the sample for each stratum is calculated according to the size of the population of that stratum.

This method ensures that the composition of the sample reflects that of the population.

Stratum	**Sample Size**
Computer Engineering (CI1)	24
Industrial Engineering (CI1)	23
Industrial Engineering (CI2)	24
Total	**71**

Table 3.1: Distribution of the sample by stratum

D. Justification of the Plan

- Stratified sampling ensures that all important subgroups (level, field of study) are represented.
- The sample size is sufficient to obtain reliable data with a 90% confidence interval.
- The plan allows for drawing generalizable conclusions about student employability.

3.4 Questionnaire Results

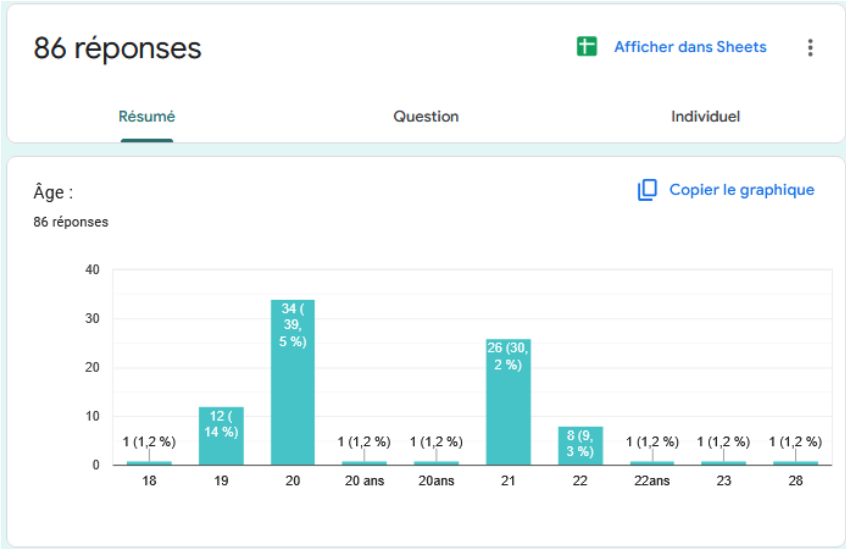


Figure 3.1: Image 1

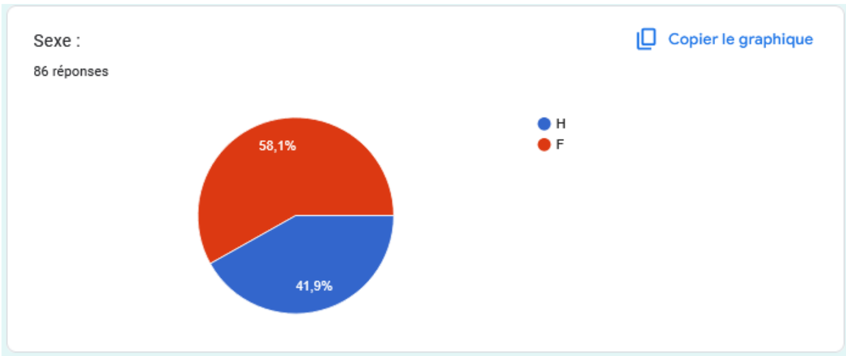


Figure 3.2: Image 2

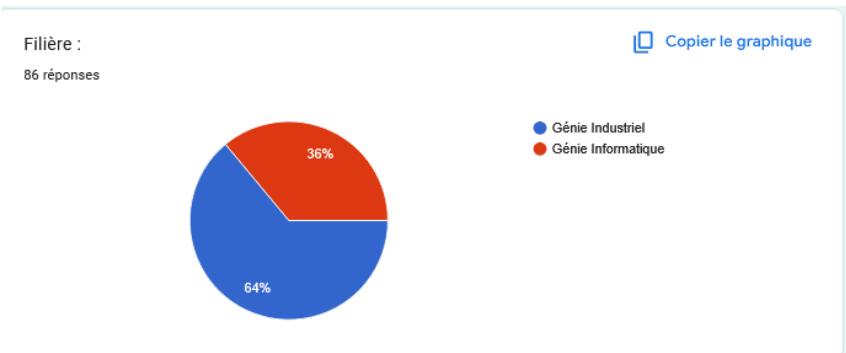


Figure 3.3: Image 3

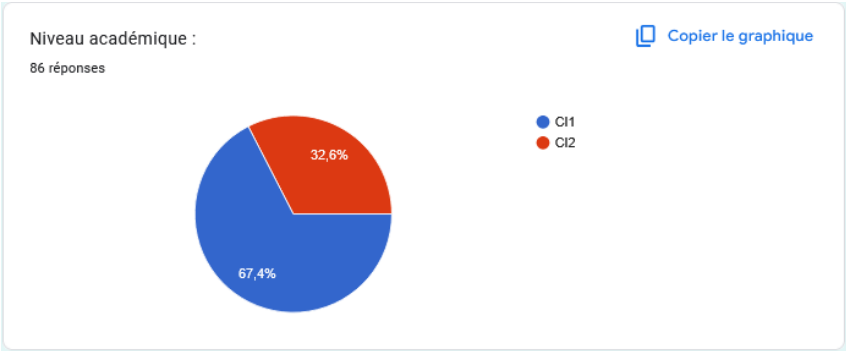


Figure 3.4: Image 4

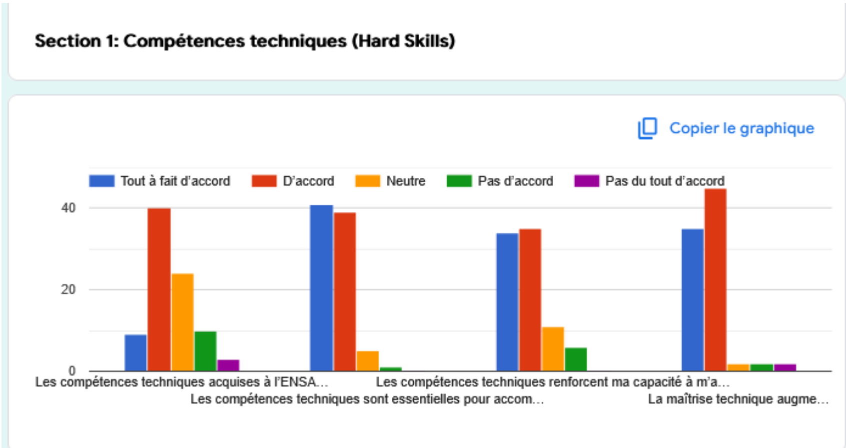


Figure 3.5: Image 5

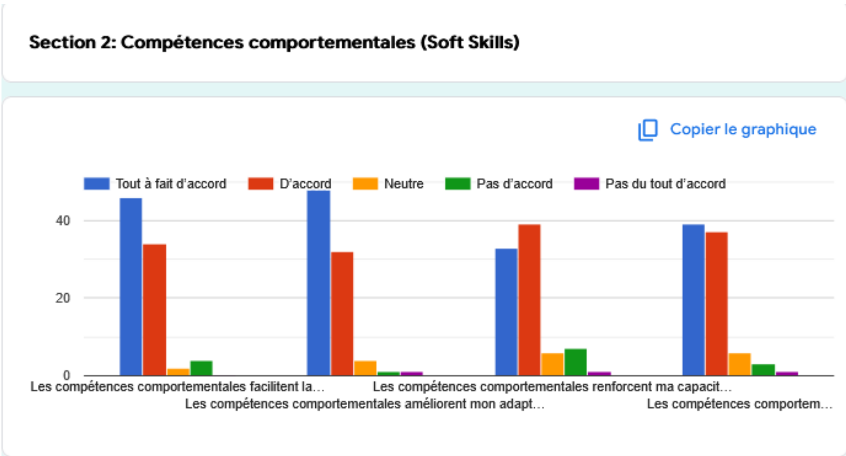


Figure 3.6: Image 6

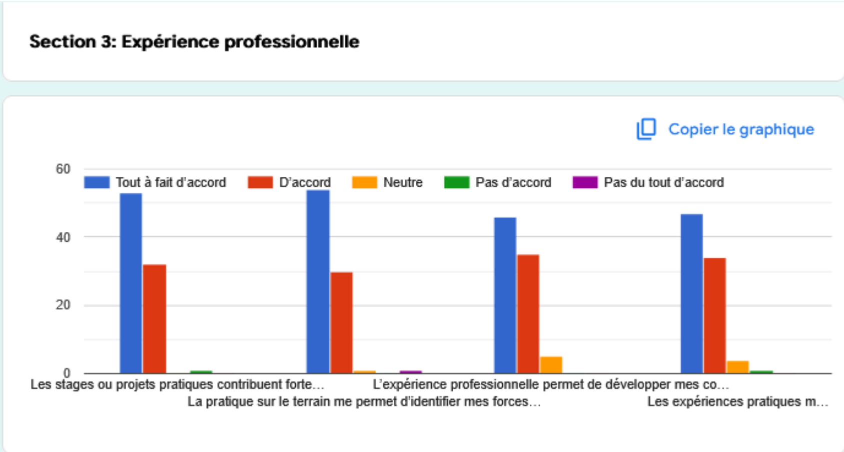


Figure 3.7: Image 7

Figure 3.8: Image 8

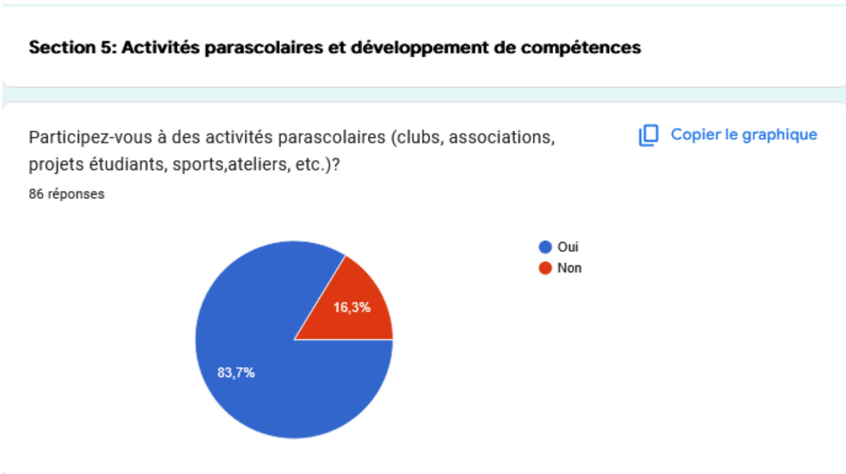


Figure 3.9: Image 9

Figure 3.10: Image 10

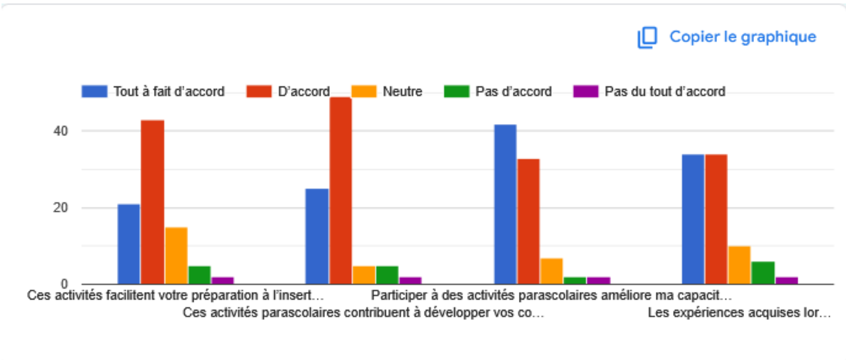


Figure 3.11: Image 11

3.5 Data Collection and Processing

This section describes the main steps followed to prepare and clean the quantitative data before conducting the statistical analysis.

3.5.1 Data Import and Initial Setup

The dataset was imported from an Excel file using the `readxl` package in `R`. Several libraries were loaded to support the analysis, including tools for descriptive statistics, reliability analysis, and distribution assessment. The variables were then renamed to ensure clarity and facilitate manipulation during subsequent analyses.

3.5.2 Data Preprocessing and Sampling Strategy

The age variable was standardized by removing textual suffixes to ensure a numeric format. A filtering process was applied to define the target population: students in Industrial Engineering at levels CI1 and

CI2, and students in Computer Science at level CI1 only. Computer Science students at level CI2 were excluded to maintain consistency with the study design.

To achieve a balanced sample across programs and academic levels, a stratified rebalancing strategy was implemented. A priority-based function was used to remove observations with neutral responses, and when necessary, random removal was applied. The final dataset respected the predefined sampling quotas and was verified using descriptive tables.

3.5.3 Variable Coding and Type Conversion

Categorical variables such as gender, program, academic level, and participation in extracurricular activities were converted into numeric formats suitable for statistical analysis. Ordinal variables measured on a Likert scale were recoded using a predefined hierarchical order and transformed into numeric scores ranging from 1 to 5. This ensured consistency and comparability across all questionnaire dimensions.

3.5.4 Data Cleaning Procedures

Outliers in the age variable were identified using the interquartile range (IQR) method. Detected outliers were replaced with missing values. The proportion of missing values was then evaluated: when the percentage exceeded 5%, mean imputation was applied for the age variable, while observations with missing values in key categorical variables were removed.

3.5.5 Final Dataset and Export

After cleaning and validation, only the variables relevant to the analysis were retained. The final dataset was exported as a CSV file to ensure reproducibility and facilitate further statistical analyses. At this stage, the data were considered clean, balanced, and ready for inferential statistical procedures.

```r
#////////////// QUANTITATIVE STUDY ////////////////////////////////////////
# ===========================================================================
# STEP 1: DEFINITION OF THE PROBLEMATIC & CONTEXT
# ===========================================================================
#
# 1. OBJECT OF THE STUDY
#    Statistical Study Employability of ENSA Kenitra Students
#
# 2. PROBLEMATIC
#    Employability is the ability to obtain, keep, and evolve in a job.
#    Despite high-quality academic training, young graduates often face
#    difficulties accessing the labor market.
#
# 3. RESEARCH QUESTION
#    1. What are the key factors that influence employability, and to what exten
#can thesefactors predict or classify students' employability levels?
#    2. Which methods do students perceive as effective in enhancing their
#employability and adapting to the current demands of the labor market?

# ===========================================================================
# POPULATION AND SAMPLE
# ===========================================================================
# - Target      : ENSA students, 1st and 2nd year engineering cycle.
# - Age         : 19 to 30 years old.
# - Gender      : Male (H), Female (F).
# - Majors      : GIND, GI.
#
# ===========================================================================
# DATA DICTIONARY (STUDY VARIABLES)
# ===========================================================================
```

Figure 3.12: Data Import and Initial Setup

```r
# ===========================================================================
# DATA DICTIONARY (STUDY VARIABLES)
# ===========================================================================
# A. EXPLANATORY VARIABLES (Socio-demographic):
#    - Age, Sex, Major (Filiere), Level
#
# B. VARIABLES TO EXPLAIN (Measured Items):
#    - Hard Skills
#    - Soft Skills
#    - Professional Experience
#    - Professional Networking
#    - Extracurricular Activities
# ===========================================================================
# STEP 2: IMPORT AND INITIAL PREPARATION
# ===========================================================================

# Load necessary libraries
library(readxl)
library(moments) # For skewness and kurtosis
library(psych)   # For Cronbach's alpha

# Import raw data
data <- read_excel("PROJET MOUMEN/data.xlsx")
View(data)
```

Figure 3.13: Data Import and Initial Setup

```r
# --------------------------------------------------------------------------------
# 1. PRELIMINARY TEXT CLEANING & RENAMING
# --------------------------------------------------------------------------------
# Rename variables for easier manipulation in R
colnames(data) <- c(
  "Horodateur",
  "Age",
  "Sexe",
  "Filiere",
  "Niveau",

  # Hard Skills (Technical Competencies)
  "hard_sk_1", "hard_sk_2", "hard_sk_3", "hard_sk_4",

  # Soft Skills (Behavioral Competencies)
  "soft_sk_1", "soft_sk_2", "soft_sk_3", "soft_sk_4",

  # Professional Experience
  "prof_exp_1", "prof_exp_2", "prof_exp_3", "prof_exp_4",

  # Professional Networking
  "prof_net_1", "prof_net_2", "prof_net_3", "prof_net_4",

  # Extracurricular Activities
  "participation_activites",
  "type_activites",
  "extr_act_1", "extr_act_2", "extr_act_3", "extr_act_4"
)
```

Figure 3.14: PRELIMINARY TEXT CLEANING and RENAMING

```r
# ==============================================================================
# STEP 3: DATA PRE-PROCESSING & SAMPLING STRATEGY
# ==============================================================================
# Standardize the 'Age' variable by removing text suffixes (e.g., "ans" or "years")
data$Age <- gsub(" ans", "", data$Age, ignore.case = TRUE)
data$Age <- gsub("ans", "", data$Age, ignore.case = TRUE)


# Check initial response count
nrow(data)

# ------------------------------------------------------------------------------
# 3.1 TARGET POPULATION FILTERING
# ------------------------------------------------------------------------------
# Target Population Definition:
# - Industrial Engineering: CI1 and CI2 (Levels 1 & 2)
# - Computer Science: CI1 Only
# Action: Remove CI2 Computer Science students from the dataset.

data_clean <- subset(
  data,
  !(Filiere == "Génie Informatique" & Niveau == "CI2")
)
```

Figure 3.15: DATA PRE-PROCESSING and SAMPLING STRATEGY and TARGET POPULATION FILTERING

```r
# Verification after filtering
table(data_clean$Filiere, data_clean$Niveau)
nrow(data_clean)
# -----------------------------------------------------------------------------
# 3.2 SPriority function for removing “Neutral” responses
# -----------------------------------------------------------------------------
#First removes all “Neutral” responses
#If insufficient, completes the removal by random selection

remove_neutral <- function(df, n_remove, var){

  # Identify neutral responses
  neutre <- df[df[[var]] == "Neutre", ]

  if(nrow(neutre) >= n_remove){
    # Case in which there are enough neutral responses "Neutre"
    df <- df[!(rownames(df) %in% rownames(neutre)[1:n_remove]), ]
  } else {
    # Case in which neutral responses are insufficient
    reste <- n_remove - nrow(neutre)

    # Removal of all neutral responses
    df <- df[!(rownames(df) %in% rownames(neutre)), ]

    # Random removal of the remaining number
    autres <- df[sample(nrow(df), reste), ]
    df <- df[!(rownames(df) %in% rownames(autres)), ]
  }

  return(df)
```

Figure 3.16: SPriority function for removing “Neutral” responses

```r
# -----------------------------------------------------------------------------
# 3.3 Stratified rebalancing
# -----------------------------------------------------------------------------
# A) CI1 Industrial Engineering: Remove 8 observations
ci1_indus <- subset(
  data_clean,
  Filiere == "Génie Industriel" & Niveau == "CI1"
)

ci1_indus <- remove_neutral(ci1_indus, 8, "hard_sk_1")

# B) CI1 Computer Science: Remove 4 observations
ci1_info <- subset(
  data_clean,
  Filiere == "Génie Informatique" & Niveau == "CI1"
)

ci1_info <- remove_neutral(ci1_info, 3, "hard_sk_1")
# C) CI2 Industrial Engineering: No removal needed (stratum fits quota)
ci2_indus <- subset(
  data_clean,
  Filiere == "Génie Industriel" & Niveau == "CI2"
)
# -----------------------------------------------------------------------------
# 3.4 FINAL MERGE & VERIFICATION
# -----------------------------------------------------------------------------

data_final <- rbind(ci2_indus, ci1_indus, ci1_info)

# -----------------------------------------------------------------------------
```

Figure 3.17: Stratified rebalancing

```r
# ------------------------------------------------------------------------------
# 3.5 FINAL VERIFICATION
# ------------------------------------------------------------------------------

# Final sample size
nrow(data_final)

# Final distribution by program and level
table(data_final$Filiere, data_final$Niveau)

##############################################################
#  DISPLAY AND VERIFICATION OF THE CLEANED DATASET
#  Object: data_final
#  Stage: After cleaning / Before conversion
##############################################################
# 1) Display of the cleaned dataset (full visualization)
View(data_final)

# 2) Dataset dimensions
#    -> Checks that the final sample contains exactly 71 observations
dim(data_final)

# 3) Total number of observations
nrow(data_final)
# 4) Structure of the variables
#    -> Variable types (text, numeric, etc.)
str(data_final)
```

Figure 3.18: FINAL VERIFICATION

```r
# 5) Final distribution by program and level
#    -> Proof of compliance with the sampling plan
table(data_final$Filiere, data_final$Niveau)

# 6) Verification of the variable used for filtering
#    -> Check of remaining categories (Neutral, Agree, etc.)
table(data_final$hard_sk_1)
# 7) Quick preview of the first rows
head(data_final)
##############################################################
# - Balanced distribution by program and level
# - Dataset ready for conversion and analysis
##############################################################

# ------------------------------------------------------------------------------
# 3.1 : TYPE CONVERSION AND CODING
# ------------------------------------------------------------------------------

# A. IDENTIFICATION AND CATEGORICAL VARIABLES
# Ensure that age is numeric for mean calculations
if(!is.numeric(data_final$Age)) {
  data_final$Age <- as.numeric(data_final$Age)
}

# --- GENDER CODING ---
# Conversion into a numeric binary variable (Male=1, Female=2) for analysis
if(!is.character(data_final$Sexe)) { data_final$Sexe <- as.character(data_final$Sexe) }
data_final$Sexe <- as.numeric(factor(data_final$Sexe, levels = c("H", "F"), ordered = TRUE))
```

Figure 3.19: TYPE CONVERSION AND CODING

```r
# --- PROGRAM CODING ---
# Conversion into an ordered factor variable
if(!is.character(data_final$Filiere)) { data_final$Filiere <- as.character(data_final$Filiere) }
data_final$Filiere <- as.numeric(factor(data_final$Filiere,
                                        levels = c("Génie Industriel", "Génie Informatique"),
                                        ordered = TRUE))

# --- LEVEL CODING ---
# Conversion respecting the academic hierarchy (CI1 < CI2)
if(!is.character(data_final$Niveau)) { data_final$Niveau <- as.character(data_final$Niveau) }
data_final$Niveau <- as.numeric(factor(data_final$Niveau,
                                       levels = c("CI1", "CI2"),
                                       ordered = TRUE))

# --- PARTICIPATION CODING ---
# Binary conversion (No=1, Yes=2)
if(!is.character(data_final$participation_activites)) {
  data_final$participation_activites <- as.character(data_final$participation_activites)
}
data_final$participation_activites <- as.numeric(factor(data_final$participation_activites,
                                                        levels = c("Non", "Oui"),
                                                        ordered = TRUE))
```

Figure 3.20: TYPE CONVERSION AND CODING

```r
# B. ORDINAL QUALITATIVE VARIABLES (LIKERT SCALE)
# Creation of a function to automate the conversion of the 20 items
# Objective: Convert text responses ("Agree") into numeric scores (4)

ordre_likert <- c("Pas du tout d'accord", "Pas d'accord", "Neutre", "D'accord", "Tout à fait d'a

convertir_likert <- function(colonne) {
  # 1. Enforce the hierarchical order of responses
  colonne_factor <- factor(colonne, levels = ordre_likert, ordered = TRUE)
  # 2. Convert ranks into numeric values (1 to 5)
  return(as.numeric(colonne_factor))
}

# Apply coding to the 5 questionnaire dimensions:

# 1. Hard Skills
data_final$hard_sk_1 <- convertir_likert(data_final$hard_sk_1)
data_final$hard_sk_2 <- convertir_likert(data_final$hard_sk_2)
data_final$hard_sk_3 <- convertir_likert(data_final$hard_sk_3)
data_final$hard_sk_4 <- convertir_likert(data_final$hard_sk_4)

# 2. Soft Skills
data_final$soft_sk_1 <- convertir_likert(data_final$soft_sk_1)
data_final$soft_sk_2 <- convertir_likert(data_final$soft_sk_2)
data_final$soft_sk_3 <- convertir_likert(data_final$soft_sk_3)
data_final$soft_sk_4 <- convertir_likert(data_final$soft_sk_4)
```

Figure 3.21: TYPE CONVERSION AND CODING

```r
# 3. Professional Experience
data_final$prof_exp_1 <- convertir_likert(data_final$prof_exp_1)
data_final$prof_exp_2 <- convertir_likert(data_final$prof_exp_2)
data_final$prof_exp_3 <- convertir_likert(data_final$prof_exp_3)
data_final$prof_exp_4 <- convertir_likert(data_final$prof_exp_4)

# 4. Networking
data_final$prof_net_1 <- convertir_likert(data_final$prof_net_1)
data_final$prof_net_2 <- convertir_likert(data_final$prof_net_2)
data_final$prof_net_3 <- convertir_likert(data_final$prof_net_3)
data_final$prof_net_4 <- convertir_likert(data_final$prof_net_4)

# 5. Extracurricular Activities
data_final$extr_act_1 <- convertir_likert(data_final$extr_act_1)
data_final$extr_act_2 <- convertir_likert(data_final$extr_act_2)
data_final$extr_act_3 <- convertir_likert(data_final$extr_act_3)
data_final$extr_act_4 <- convertir_likert(data_final$extr_act_4)

str(data_final)
```

Figure 3.22: TYPE CONVERSION AND CODING

```r
# 3.2 : DATA PROCESSING AND CLEANING
# ----------------------------------------------------------------------------
# --- 3.2.1 OUTLIER TREATMENT ---
# Method used: Interquartile Range (IQR).
# Values beyond 1.5 * IQR are considered outliers.

# Visualization (boxplot) before cleaning
boxplot(data_final$Age,
        main = "Age distribution (Before cleaning)",
        ylab = "Age (Years)",
        col = "lightblue",
        border = "darkblue")

# Identification and replacement
aberrantes_age <- boxplot.stats(data_final$Age)$out

cat("Outliers detected for age:", aberrantes_age, "\n")
cat("Number of outliers:", length(aberrantes_age), "\n")

# Replacement
aberrantes_age <- boxplot.stats(data_final$Age)$out

if (length(aberrantes_age) > 0) {
  # Loop through the data to replace outliers with NA
  for (i in 1:length(data_final$Age)) {
    if (!is.na(data_final$Age[i]) && data_final$Age[i] %in% aberrantes_age) {
      data_final$Age[i] <- NA
    }
  }
}
```

Figure 3.23: DATA PROCESSING AND CLEANING

```
# ============================================================================
# STEP 4 : EXPORT OF CLEAN DATA
# ============================================================================

# Selection of relevant columns for final analysis
donnees_finales <- data_final[, c(
  "Age", "Sexe", "Filiere", "Niveau",
  "hard_sk_1", "hard_sk_2", "hard_sk_3", "hard_sk_4",
  "soft_sk_1", "soft_sk_2", "soft_sk_3", "soft_sk_4",
  "prof_exp_1", "prof_exp_2", "prof_exp_3", "prof_exp_4",
  "prof_net_1", "prof_net_2", "prof_net_3", "prof_net_4",
  "participation_activites",
  "extr_act_1", "extr_act_2", "extr_act_3", "extr_act_4"
)]

# Saving
write.csv(donnees_finales, "donnees_nettoyees_finales.csv", row.names = FALSE)
```

Figure 3.24: DATA PROCESSING AND CLEANING

3.5.6 Assessment of Normality Assumptions

Before conducting parametric statistical analyses, the assumption of normality was carefully examined, as it constitutes a fundamental requirement for procedures such as t-tests, ANOVA, Pearson correlation, and linear regression. Given that most variables in this study are measured using Likert scales— which rarely achieve perfect normality in social science research— a pragmatic yet rigorous approach was adopted. Normality was assessed for all continuous and ordinal variables in order to determine the appropriateness of parametric versus non-parametric analytical techniques.

Assessment Framework

The normality assessment protocol consisted of three complementary components:

1. **Shapiro–Wilk Test**: A formal hypothesis test of normality where the null hypothesis assumes that the data follow a normal distribution. A significant result ($p < 0.05$) indicates deviation from normality.
2. **Skewness Analysis**: Measurement of distributional asymmetry. Values close to zero indicate symmetry, while values exceeding ± 1 suggest substantial asymmetry.
3. **Kurtosis Analysis**: Measurement of distributional "tailedness" or peakedness. For a normal distribution, kurtosis approximates 3; values substantially above or below this benchmark indicate distributional abnormalities.

Decision Criteria

Indicator	Threshold	Interpretation
Shapiro–Wilk p-value	$p > 0.05$	Normal distribution
Skewness	$[-1; +1]$	Acceptable asymmetry
Kurtosis	$[-3; +3]$	Acceptable flatness

Table 3.2: Normality assessment decision criteria

The following sections present detailed results for each tested variable, beginning with the demographic variable *Age*, followed by a systematic evaluation of all questionnaire items across the five measurement scales: Hard Skills, Soft Skills, Professional Experience, Professional Networking, and Extracurricular Activities.

```r
#  STATISTICAL ANALYSIS - NORMALITY TESTS
# ============================================================================

# Age test
cat("=== NORMALITY TEST FOR AGE ===\n\n")
# H0: The age variable follows a normal distribution
# H1: The age variable does not follow a normal distribution

# 1. Shapiro-Wilk Test
shapiro_result <- shapiro.test(data_final$Age)
cat("1. Shapiro-Wilk Test:\n")
cat("   W =", shapiro_result$statistic, "\n") # W= 0.8679304 !=1   0 < W < 1 measures how closely your data resembles a normal distribution. If W=1 -> NORMALITY
cat("   p-value =", shapiro_result$p.value, "\n") # p-value = 3.939179e-06
sprintf("p-value = %.12f ",shapiro_result$p.value) # p-value = 0.000003939179 in decimal
if(shapiro_result$p.value < 0.05) {
  cat("    Conclusion: NOT strictly normal (p < 0.05)\n\n")
} else {
  cat("    Conclusion: NORMAL distribution (p >= 0.05)\n\n")
}
# Accept H1 -> There is a significant difference between the distribution and the theoretical normal distribution
# Age variable is not strictly normal -> check for quasi-normality

library(moments)  # For skewness and kurtosis

# 2. Skewness
skew <- skewness(data_final$Age, na.rm = TRUE)
cat("2. Skewness:\n")
cat("   Value =", round(skew, 3), "\n") # Value = 0.21
# 3. Kurtosis
kurt <- kurtosis(data_final$Age, na.rm = TRUE)
cat("3. Kurtosis:\n")
cat("   Value =", round(kurt, 3), "\n") # Value = 2.432
# Skewness = 0.21 is within[-1,1] and Kurtosis = 2.432 is within [-3; +3] -> quasi-normality
```

Figure 3.25: Normality test

```r

# List of all items (NORMALITY TEST LOOP)

variables_to_test <- c(
  "hard_sk_1", "hard_sk_2", "hard_sk_3", "hard_sk_4",
  "soft_sk_1", "soft_sk_2", "soft_sk_3", "soft_sk_4",
  "prof_exp_1", "prof_exp_2", "prof_exp_3", "prof_exp_4",
  "prof_net_1", "prof_net_2", "prof_net_3", "prof_net_4",
  "extr_act_1", "extr_act_2", "extr_act_3", "extr_act_4"
)

cat("=== START OF NORMALITY TESTS ===\n\n")

for(var_name in variables_to_test) {

  # Safety check: Ensure the column exists and is numeric
  if (!is.null(data_final[[var_name]]) && is.numeric(data_final[[var_name]])) {

    cat("--------------------------------------------------\n")
    cat("VARIABLE:", var_name, "\n")

    # 1. Shapiro
    # Note: Shapiro may fail if all values are identical (e.g., everyone answered 5)
    tryCatch({
      shapiro <- shapiro.test(data_final[[var_name]])
      p_val <- shapiro$p.value
      stat_w <- shapiro$statistic

      cat("  > Shapiro: p-value =", sprintf("%.12f", p_val), "\n")

      # Shapiro decision
      status_shapiro <- ifelse(p_val > 0.05, "NORMAL DISTRIBUTION","NON-NORMAL DISTRIBUTION")
      cat("    Statistical Conclusion:", status_shapiro, "\n")

```

Figure 3.26: Normality test

```r
    }, error = function(e) {
      cat("  > Shapiro: Impossible (Zero variance or insufficient data)\n")
      p_val <- 0 # Force 0 for subsequent checks
    })

    # 2. Skewness & Kurtosis (Quasi-normality)
    skew <- skewness(data_final[[var_name]], na.rm = TRUE)
    kurt <- kurtosis(data_final[[var_name]], na.rm = TRUE)

    cat("  > Skewness             : ", round(skew, 3))
    if(abs(skew) < 1) cat(" (Good)\n") else cat(" (High)\n")

    cat("  > Kurtosis             : ", round(kurt, 3))
    if(abs(kurt) < 3) cat(" (Good)\n") else cat(" (High)\n")
    # Note: moments::kurtosis returns normal around 3
    # We check that it's not too extreme
    cat("\n")

    # 3. FINAL VERDICT (Quasi-normality)
    # Normality is accepted if p > 0.05 OR if Skewness and Kurtosis are reasonable
    # (Skewness between -3 and 3 is often accepted in social sciences)

    if (exists("p_val") && p_val > 0.05) {
      cat("  ==> FINAL RESULT: Normal distribution\n")
    } else if (abs(skew) < 1 && abs(kurt) < 3) {
      cat("  ==> FINAL RESULT: Quasi-normal distribution (acceptable for parametric tests)\n")
    } else {
      cat("  ==> FINAL RESULT: Non-normal distribution (use non-parametric tests)\n")
    }

    cat("\n")

  } else {
```

Figure 3.27: Normality test

```r
  } else {
    cat("ERROR: Variable", var_name, "is not numeric or does not exist.\n")
  }
}

cat("=== END OF NORMALITY TESTS ===\n")
```

Figure 3.28: Normality test

Normality Test for the Age Variable

Observed Results

- **Shapiro–Wilk Test**: $W = 0.868$, $p = 0.000003939$
- **Skewness**: 0.21
- **Kurtosis**: 2.432

Interpretation The Shapiro–Wilk test rejects the hypothesis of strict normality ($p < 0.05$), indicating that the age distribution deviates significantly from a theoretical normal distribution. However, the skewness (0.21) and kurtosis (2.432) indices remain within acceptable thresholds, suggesting an approximately symmetric and moderately peaked distribution.

Conclusion Despite formal statistical rejection of normality, the *Age* variable exhibits sufficient quasi-normality to justify the use of parametric statistical tests in subsequent analyses.

Normality Tests for Questionnaire Items

Methodological Approach For each ordinal variable measured using Likert scales, a triple verification procedure was applied in order to assess distributional normality:

1. **Shapiro–Wilk Test**: A formal statistical test used to evaluate the null hypothesis that the data follow a normal distribution.
2. **Skewness Coefficient**: An indicator of distributional asymmetry, where values close to zero reflect symmetry.
3. **Kurtosis Coefficient**: An indicator of distributional flatness or peakedness relative to the normal distribution.

3.5.7 Methodological Decisions

For each tested variable, methodological decisions regarding the choice of statistical tests were made based on the outcome of the normality assessment, following the criteria defined previously.

Decision Rules

- **Normal or Quasi-normal Distribution**: *"This variable presents a (quasi-)normal distribution, justifying the use of parametric tests such as Pearson correlations, linear regression, and t-tests."*
- **Non-normal Distribution**: *"This variable does not meet normality criteria. Non-parametric tests such as Spearman correlations, Mann–Whitney tests, and Kruskal–Wallis tests are preferred for analyses involving this variable."*

Summary Table of Normality Assessment

Table 3.3: Summary of Normality Assessment for all questionnaire items

Variable	**Shapiro *p*-value**	**Skewness**	**Kurtosis**	**Verdict**	**Recommended Tests**
Age	0.000004	0.21	2.43	Quasi-normal	Parametric
hard_sk_1	0.000018	-0.49	2.69	Quasi-normal	Parametric
hard_sk_2	¡ 0.000001	-0.57	2.40	Quasi-normal	Parametric
hard_sk_3	0.000000032	-0.86	2.84	Quasi-normal	Parametric
hard_sk_4	¡ 0.000001	-1.87	7.67	Non-normal	Non-parametric
soft_sk_1	¡ 0.000001	-1.41	4.93	Non-normal	Non-parametric
soft_sk_2	¡ 0.000001	-1.66	7.07	Non-normal	Non-parametric
soft_sk_3	0.000000008	-1.16	3.82	Non-normal	Non-parametric
soft_sk_4	0.000000002	-1.42	5.23	Non-normal	Non-parametric
prof_exp_1	¡ 0.000001	-0.60	1.36	Quasi-normal	Parametric
prof_exp_2	¡ 0.000001	-2.30	11.90	Non-normal	Non-parametric
prof_exp_3	¡ 0.000001	-0.82	2.64	Quasi-normal	Parametric
prof_exp_4	¡ 0.000001	-1.10	4.42	Non-normal	Non-parametric
prof_net_1	0.000000002	-1.09	4.35	Non-normal	Non-parametric
prof_net_2	0.000000007	-0.64	3.49	Non-normal	Non-parametric
prof_net_3	¡ 0.000001	-1.41	5.43	Non-normal	Non-parametric
prof_net_4	0.000000007	-1.24	4.50	Non-normal	Non-parametric
extr_act_1	0.000000174	-0.92	3.73	Non-normal	Non-parametric
extr_act_2	0.000000001	-1.42	5.39	Non-normal	Non-parametric
extr_act_3	0.000000002	-1.41	5.12	Non-normal	Non-parametric
extr_act_4	0.00000004	-1.13	3.86	Non-normal	Non-parametric

3.5.8 Reliability Analysis: Cronbach's Alpha

Concept

Internal consistency reliability was assessed using Cronbach's alpha coefficient in order to evaluate the degree to which items within each measurement scale consistently reflect the same underlying construct. Following Nunnally's (1978) guidelines, alpha values of 0.70 or higher are considered acceptable, while values above 0.60 are deemed acceptable for exploratory research in the social sciences (Hair et al., 2010). This assessment is crucial to ensure that the multi-item scales used in this study produce reliable and consistent measurements prior to hypothesis testing.

Overall Reliability Assessment

The overall internal consistency of the measurement model was evaluated by computing the mean Cronbach's alpha across all five dimensions. The resulting average alpha coefficient was **0.65**, which meets the recommended threshold for exploratory research in the social sciences.

```r
# ============================================================================
# RELIABILITY TEST - CRONBACH'S ALPHA
# ============================================================================
# install.packages("psych")
# Load the package
library(psych)
# Ensure that the columns exist and are numeric
class(data_final$Age)
class(data_final$Filiere)
class(data_final$Sexe)
class(data_final$`hard_sk_1`) # Hard_sk_1 item chosen arbitrarily

# This code checks the type (class) of certain columns in a dataframe `data`
# to ensure they exist and are numeric.
# This verification is essential to ensure the data is in the correct format before performing analyses.

# Create data frames for the items

# Check column names
cat("Available columns in your dataset:\n")
print(names(data_final))
cat("\n")

# The function names(data) displays the column names of the data frame.
# It also verifies that the columns we want to use (e.g., data$hard_sk_1, data$hard_sk_2, etc.) exist.
# This helps avoid errors and better understand the structure of the data frame before analysis.

# Create a dataframe containing only hard skills items
items1 <- data.frame(
  data_final$hard_sk_1,
  data_final$hard_sk_2,
  data_final$hard_sk_3,
  data_final$hard_sk_4
)

```

Figure 3.29: Reliability test

```
# Create a dataframe containing only soft skills items
items2 <- data.frame(
  data_final$soft_sk_1,
  data_final$soft_sk_2,
  data_final$soft_sk_3,
  data_final$soft_sk_4
)

# Create a dataframe containing only professional experience items
items3 <- data.frame(
  data_final$prof_exp_1,
  data_final$prof_exp_2,
  data_final$prof_exp_3,
  data_final$prof_exp_4
)

# Create a dataframe containing only professional networking items
items4 <- data.frame(
  data_final$prof_net_1,
  data_final$prof_net_2,
  data_final$prof_net_3,
  data_final$prof_net_4
)

# Create a dataframe containing only extracurricular activities items
items5 <- data.frame(
  data_final$extr_act_1,
  data_final$extr_act_2,
  data_final$extr_act_3,
  data_final$extr_act_4
)

# psych::alpha(items, check.keys = TRUE)
# check.keys = TRUE -> automatically corrects reverse-coded items if needed
```

Figure 3.30: Reliability test

```
# na.rm = TRUE -> (optional) ignores missing values
# Compute Cronbach's alpha using psych::alpha to avoid conflicts
alpha_result0 <- psych::alpha(items1, check.keys=TRUE)
alpha_result1 <- psych::alpha(items2, check.keys=TRUE)
alpha_result2 <- psych::alpha(items3, check.keys=TRUE)
alpha_result3 <- psych::alpha(items4, check.keys=TRUE)
alpha_result4 <- psych::alpha(items5, check.keys=TRUE)

# Display results

# These lines display the results of Cronbach's alpha for each group. From these results:
# If alpha ≥ 0.6: The group is reliable, meaning the items are consistent.
# If alpha < 0.6: The group is not reliable.
print(alpha_result0)
# raw_alpha = 0.43 < 0.6 : No significant consistency among Hard Skills items
print(alpha_result1)
# raw_alpha = 0.66 > 0.6 : Acceptable consistency among Soft Skills items
print(alpha_result2)
# raw_alpha = 0.69 > 0.6 : Acceptable consistency among Professional Experience items
print(alpha_result3)
# raw_alpha = 0.63 > 0.6 : Acceptable consistency among Professional Networking items
print(alpha_result4)
# raw_alpha = 0.84 > 0.6 : Good consistency among Extracurricular Activities items
mean(c(0.43,0.66,0.69,0.63,0.84))
# 0.65 > 0.6 : Acceptable overall consistency

cat("\n\n################################################################################\n")
cat("    RELIABILITY SUMMARY AND NEXT STEPS\n")
cat("################################################################################\n\n")

cat("ALPHA SUMMARY:\n")
cat("  Hard Skills      : 0.43 (Insufficient but retained)\n")
cat("  Soft Skills      : 0.66 (Acceptable)\n")
cat("  Professional Exp : 0.69 (Acceptable)\n")
```

Figure 3.31: Reliability test

```
cat("  Professional Exp : 0.69 (Acceptable)\n")
cat("  Professional Net : 0.63 (Acceptable)\n")
cat("  Extracurricular  : 0.84 (Good)\n")
cat("  MEAN             : 0.65 (Acceptable for exploratory research)\n\n")

cat("DECISION: Continue with all dimensions\n")
cat("=== END OF RELIABILITY TESTS ===\n")
```

Figure 3.32: Reliability test

Summary Table of Reliability Results

Dimension	Cronbach's Alpha	Evaluation	Retained?
Hard Skills	0.43	Insufficient but conceptually important	Yes
Soft Skills	0.66	Acceptable	Yes
Professional Experience	0.69	Good	Yes
Professional Networking	0.63	Acceptable	Yes
Extracurricular Activities	0.84	Excellent	Yes
Overall Mean	**0.65**	Acceptable for exploratory research	All retained

Table 3.4: Summary of Cronbach's alpha reliability results

3.5.9 Sample Representativeness Analysis

Introduction

Sample representativeness was assessed to determine whether our study sample accurately reflects the characteristics of the target population ($N = 204$ ENSAK students). We compared sample distributions against known population parameters for three key demographic variables: Major (GI vs. GIND), Gender (Male vs. Female), and Age.

- **Categorical variables** (Major, Gender) were evaluated using *Chi-square goodness-of-fit tests.* - **Continuous variable** (Age) was evaluated using a *one-sample t-test.*

A significance threshold of $\alpha = 0.05$ was applied for all statistical tests.

```r
# ============================================================================
# SAMPLE REPRESENTATIVENESS TEST
# ============================================================================

options(scipen = 999) # Prevents R from displaying numbers in scientific notation.

cat("################################################################################\n")
cat("               SAMPLE REPRESENTATIVENESS TEST\n")
cat("################################################################################\n\n")

cat("OBJECTIVE: Check if the sample accurately represents the total population\n")
cat("METHOD: Compare distributions of Major, Gender, and Age\n\n")


# ============================================================================
# STEP 1: DEFINE POPULATION DATA
# ============================================================================

cat(..., file = "", sep = " ", fill = FALSE, labels = NULL, append = FALSE) ===\n")
cat("STEP 1: POPULATION DATA\n")
cat("========================================\n\n")


# TOTAL POPULATION
pop_total <- 204  # Total number of students at ENSAK

# MAJORS
# If you only have GI and GIND:
pop_gi <- 67      # Number of Computer Engineering students
pop_gind <- 137   # Number of Industrial Engineering students across both levels

# GENDER (TO ADAPT)
pop_men <- 80     # Number of men in the population
pop_women <- 124  # Number of women in the population
```

Figure 3.33: Representativeness test

```r

# AGE (TO ADAPT)
pop_mean_age <- 21  # Average age in the population, e.g., [19,23]

cat("POPULATION DATA:\n")
cat("  Total students :", pop_total, "\n")
cat("  GI             :", pop_gi, "(", round(100*pop_gi/(pop_gi+pop_gind), 1), "%)\n")
cat("  GIND           :", pop_gind, "(", round(100*pop_gind/(pop_gi+pop_gind), 1), "%)\n")
cat("  Men            :", pop_men, "(", round(100*pop_men/pop_total, 1), "%)\n")
cat("  Women          :", pop_women, "(", round(100*pop_women/pop_total, 1), "%)\n")
cat("  Mean age       :", pop_mean_age, "years\n\n")


# ============================================================================
# STEP 2: SAMPLE DATA
# ============================================================================

cat("========================================\n")
cat("STEP 2: SAMPLE DATA\n")
cat("========================================\n\n")
nrow(data_final)
# Sample size
n_sample <- nrow(data_final)

# Major distribution
table_major_sample <- table(data_final$Filiere) # 1 = GIND, 2 = GI
n_gi_sample <- as.numeric(table_major_sample["2"])
n_gind_sample <- as.numeric(table_major_sample["1"])

names(table_major_sample)

# Gender distribution
table_gender_sample <- table(data_final$Sexe)
n_men_sample <- as.numeric(table_gender_sample["1"])
```

Figure 3.34: Representativeness test

```r
n_women_sample <- as.numeric(table_gender_sample["2"])

# Mean age
mean_age_sample <- mean(data_final$Age, na.rm = TRUE)

cat("SAMPLE DATA:\n")
cat("  Size              :", n_sample, "\n")
cat("  Sampling rate     :", round(100*n_sample/pop_total, 1), "%\n")
cat("  GI                :", n_gi_sample, "(", round(100*n_gi_sample/n_sample, 1), "%)\n")
cat("  GIND              :", n_gind_sample, "(", round(100*n_gind_sample/n_sample, 1), "%)\n")
cat("  Men               :", n_men_sample, "(", round(100*n_men_sample/n_sample, 1), "%)\n")
cat("  Women             :", n_women_sample, "(", round(100*n_women_sample/n_sample, 1), "%)\n")
cat("  Mean age          :", round(mean_age_sample, 2), "years\n\n")


# ============================================================================
# STEP 3: REPRESENTATIVENESS TEST - MAJOR (FILIERE)
# ============================================================================

cat("\n")
cat("################################################################################\n")
cat("              TEST 1: MAJOR REPRESENTATIVENESS\n")
cat("################################################################################\n\n")

cat("Question: Does the GI/GIND distribution in the sample\n")
cat("          match that of the population?\n\n")

cat("VISUAL COMPARISON:\n")
cat("  Population   : GI ", round(100*pop_gi/(pop_gi+pop_gind), 1), "% | GIND ",
    round(100*pop_gind/(pop_gi+pop_gind), 1), "%\n")
cat("  Sample       : GI ", round(100*n_gi_sample/n_sample, 1), "% | GIND ",
    round(100*n_gind_sample/n_sample, 1), "%\n\n")

# Difference calculation
```

Figure 3.35: Representativeness test

```r
diff_gi <- abs(100*pop_gi/(pop_gi+pop_gind) - 100*n_gi_sample/n_sample)
cat("  Difference for GI : ±", round(diff_gi, 1), "percentage points\n\n")

# Chi-square test
cat("CHI-SQUARE TEST:\n")
cat("  H0: Sample has the same distribution as the population\n")
cat("  H1: Sample has a different distribution\n\n")

# Expected proportions
prop_gi <- pop_gi / (pop_gi + pop_gind)
prop_gind <- pop_gind / (pop_gi + pop_gind)

test_major <- chisq.test(table_major_sample, p = c(prop_gind, prop_gi))

cat("  χ² statistic =", round(test_major$statistic, 3), "\n")
cat("  p-value      =", sprintf("%.4f", test_major$p.value), "\n\n")

# Decision
if(test_major$p.value >= 0.05) {
  cat("CONCLUSION: Major IS REPRESENTATIVE (p ≥ 0.05)\n")
  cat("→ The GI/GIND distribution of the sample does not differ significantly\n")
  cat("  from the population. Results can be generalized.\n")
  decision_major <- "Representative"
} else {
  cat("CONCLUSION: Major IS NOT REPRESENTATIVE (p < 0.05)\n")
  cat("→ There is over-/under-representation of one major.\n")
  cat("  WARNING: Potential bias in generalizing results.\n")
  decision_major <- "Not representative"
}


# ============================================================================
# STEP 4: REPRESENTATIVENESS TEST - GENDER
# ============================================================================
```

Figure 3.36: Representativeness test

```
cat("\n\n")
cat("################################################################################\n")
cat("              TEST 2: GENDER REPRESENTATIVENESS\n")
cat("################################################################################\n\n")

cat("Question: Does the H/F distribution in the sample\n")
cat("          match that of the population?\n\n")

cat("VISUAL COMPARISON:\n")
cat("  Population  : M ", round(100*pop_men/pop_total, 1), "% | F ",
    round(100*pop_women/pop_total, 1), "%\n")
cat("  Sample      : M ", round(100*n_men_sample/n_sample, 1), "% | F ",
    round(100*n_women_sample/n_sample, 1), "%\n\n")

# Difference calculation
diff_m <- abs(100*pop_men/pop_total - 100*n_men_sample/n_sample)
cat("  Difference for Men : ±", round(diff_m, 1), "percentage points\n\n")

# Chi-square test
cat("CHI-SQUARE TEST:\n")
cat("  H0: Sample has the same distribution as the population\n")
cat("  H1: Sample has a different distribution\n\n")

test_gender <- chisq.test(table_gender_sample,
                          p = c(pop_men/pop_total, pop_women/pop_total))

cat("  χ² statistic =", round(test_gender$statistic, 3), "\n")
cat("  p-value      =", sprintf("%.4f", test_gender$p.value), "\n\n")

# Decision
if(test_gender$p.value >= 0.05) {
  cat("CONCLUSION: Gender IS REPRESENTATIVE (p ≥ 0.05)\n")
  cat("→ The H/F distribution of the sample does not differ significantly\n")
```

Figure 3.37: Representativeness test

```
  cat("  from the population. No gender bias.\n")
  decision_gender <- "Representative"
} else {
  cat("CONCLUSION: Gender IS NOT REPRESENTATIVE (p < 0.05)\n")
  if(n_women_sample/n_sample > pop_women/pop_total) {
    cat("→ OVER-REPRESENTATION of women in the sample.\n")
  } else {
    cat("→ OVER-REPRESENTATION of men in the sample.\n")
  }
  cat("  WARNING: Gender bias should be reported.\n")
  decision_gender <- "Not representative"
}


# ============================================================================
# STEP 5: REPRESENTATIVENESS TEST - AGE
# ============================================================================

cat("\n\n")
cat("################################################################################\n")
cat("              TEST 3: AGE REPRESENTATIVENESS\n")
cat("################################################################################\n\n")

cat("Question: Does the mean age of the sample correspond\n")
cat("          to that of the population?\n\n")

cat("COMPARISON:\n")
cat("  Population mean age :", pop_mean_age, "years\n")
cat("  Sample mean age     :", round(mean_age_sample, 2), "years\n")
cat("  Difference          :", round(abs(mean_age_sample - pop_mean_age), 2), "years\n\n")

# Student t-test
cat("STUDENT T-TEST:\n")
cat("  H0: Sample mean = Population mean\n")
```

Figure 3.38: Representativeness test

```r
cat("  H1: Sample mean ≠ Population mean\n\n")

test_age <- t.test(data_final$Age, mu = pop_mean_age)

cat("  t statistic   =", round(test_age$statistic, 3), "\n")
cat("  p-value       =", sprintf("%.4f", test_age$p.value), "\n")
cat("  95% CI        : [", round(test_age$conf.int[1], 2), " ; ",
    round(test_age$conf.int[2], 2), "]\n\n", sep="")
print(test_age$p.value)
# Decision
if(
cat(..., file = "", sep = " ", fill = FALSE, labels = NULL, append = FALSE)
  c 0.05)\n")
  cat("→ The sample mean age does not differ significantly\n")
  cat("  from the population.\n")
  decision_age <- "Representative"
} else {
  cat("CONCLUSION: Age IS NOT REPRESENTATIVE (p < 0.05)\n")
  if(mean_age_sample > pop_mean_age) {
    cat("→ The sample is significantly OLDER than the population.\n")
  } else {
    cat("→ The sample is significantly YOUNGER than the population.\n")
  }
  cat("  WARNING: Age bias should be mentioned in the report.\n")
  decision_age <- "Not representative"
}


# ============================================================================
# STEP 6: OVERALL SUMMARY
# ============================================================================

# Summary table
summary_table <- data.frame(
  Variable = c("Major", "Gender", "Age"),
```

Figure 3.39: Representativeness test

```r
  Test = c("Chi-square", "Chi-square", "T-test"),
  P_value = round(c(test_major$p.value,
                    test_gender$p.value,
                    test_age$p.value), 4),
  Decision = c(decision_major, decision_gender, decision_age)
)

print(summary_table)

# Number of representative variables
n_rep <- sum(summary_table$Decision == "Representative")

# Overall conclusion
cat("\nSUMMARY: ", n_rep, "/3 representative variables\n")

if(n_rep == 3) {
  cat("→ Sample is overall REPRESENTATIVE\n")
} else if(n_rep == 2) {
  cat("→ Sample is PARTIALLY representative\n")
} else {
  cat("→ Sample is POORLY or NOT representative\n")

}

cat("=== END OF SAMPLE REPRESENTATIVENESS TESTS ===\n")

```

Figure 3.40: Representativeness test

Overall Representativeness Assessment

Variable	Statistical Test	χ^2/t Statistic	p-value	Decision
Major (Filière)	Chi-square	0	0.9980	Representative
Gender	Chi-square	0.389	0.5326	Representative
Age	One-sample t-test	-5.419	0.0000009094249	Not Representative

Table 3.5: Summary of representativeness assessment for key demographic variables

Overall Conclusion

The sample is **partially representative**, with **2** out of 3 variables matching population parameters. While **major and gender** accurately reflect the population, **Age** shows significant deviation from the expected distribution.

3.6 Data Processing and Statistical Analysis

This section presents the statistical processing and analysis performed on the cleaned dataset. The analysis includes univariate and bivariate descriptive statistics to explore the characteristics of the sample and the relationships between variables.

3.6.1 Univariate Descriptive Statistics

Univariate analysis was conducted to describe each variable individually and to provide an overall understanding of the sample profile.

Quantitative Variables

The quantitative variable *Age* was analyzed using descriptive statistics, including the mean, standard deviation, and summary measures. Graphical representations such as histograms were used to examine the distribution of age, while boxplots allowed the visualization of dispersion and the identification of potential outliers.

Qualitative Variables

Qualitative variables such as gender, academic major, and academic level were analyzed using frequency tables. Their distributions were illustrated using bar charts and pie charts to provide a clear visualization of the composition of the sample.

Ordinal Variables (Likert Scale)

Ordinal variables measured on a Likert scale, such as hard skills and soft skills items, were described using summary statistics. Bar charts were used to show the distribution of responses across different agreement levels, while boxplots were employed to assess variability and central tendency.

3.6.2 Bivariate Descriptive Statistics

Bivariate analysis was carried out to examine relationships between pairs of variables, without making formal statistical inferences at this stage.

Qualitative × Qualitative Analysis

Relationships between qualitative variables, such as academic major and academic level, as well as major and participation in extracurricular activities, were explored using contingency tables. Grouped bar charts were used to visually compare the distributions across categories.

Qualitative × Quantitative Analysis

Comparisons between qualitative and quantitative variables were performed by calculating group means. For example, hard skills scores were compared across academic levels, and soft skills scores were compared across majors. Comparative boxplots were used to illustrate differences between groups.

Quantitative × Quantitative Analysis

The relationship between quantitative variables was examined using scatter plots. In particular, the association between age and hard skills scores was visualized, and a linear trend line was added to highlight the general direction of the relationship.

```
# ================================================================================
# PART 5 - DATA PROCESSING AND ANALYSIS
# ================================================================================
############################################################
# This section includes:
# - Descriptive statistics (univariate and bivariate)
# - Hypothesis testing (inferential statistics)
# - Factor analysis
# - Regression / classification
############################################################


############################################################
# 5.1 UNIVARIATE DESCRIPTIVE STATISTICS
# Objective:
# Describe each variable individually to understand
# general characteristics of the sample.
############################################################
```

```
##############################
# 5.1.1 QUANTITATIVE VARIABLES
##############################

# --- Age ---

# Descriptive statistics
summary(data_final$Age)
mean(data_final$Age, na.rm = TRUE)
sd(data_final$Age, na.rm = TRUE)

# Histogram (distribution)
hist(data_final$Age,
     main = "Age Distribution of Respondents",
     xlab = "Age (years)",
     ylab = "Frequency",
     col = "lavender",
     border = "black")

# Boxplot (visual detection of outliers)
boxplot(data_final$Age,
        main = "Boxplot of Age",
        ylab = "Age",
        col = "palegreen")
```

```r
#############################
# 5.1.2 QUALITATIVE VARIABLES
#############################

# --- Gender ---
table(data_final$Sexe)

# Pie chart
pie(table(data_final$Sexe),
    main = "Gender Distribution of Respondents",
    col = c("lightblue", "lightpink"))



# --- Major ---
table(data_final$Filiere)

barplot(table(data_final$Filiere),
        main = "Distribution of Students by Major",
        ylab = "Count",
        col = "lightblue")

# Pie chart
pie(table(data_final$Filiere),
    main = "Distribution of Students by Major",
    col = c("lightpink","lightgreen"))
```

```r
# --- Level ---
table(data_final$Niveau)

barplot(table(data_final$Niveau),
        main = "Distribution of Students by Level",
        ylab = "Count",
        col = "lightgreen")

# Pie chart
pie(table(data_final$Niveau),
    main = "Distribution of Students by Level",
    col = c("lightskyblue", "khaki"))



############################
# 5.1.3 ORDINAL VARIABLES (LIKERT SCALE)
############################

# Example: Hard Skills 1
summary(data_final$hard_sk_1)

# Bar plot
barplot(table(data_final$hard_sk_1),
        main = "Response Distribution - Hard Skills 1",
        xlab = "Score",
        ylab = "Count",
        col = "lightcoral")
```

```r
# Boxplot (dispersion comparison)
boxplot(data_final$hard_sk_1,
        main = "Boxplot - Hard Skills 1",
        ylab = "Score",
        col = "lightyellow")


# Example: Soft Skills 1
summary(data_final$soft_sk_1)

barplot(table(data_final$soft_sk_1),
        main = "Response Distribution - Soft Skills 1",
        xlab = "Score",
        ylab = "Count",
        col = "lightskyblue")

boxplot(data_final$soft_sk_1,
        main = "Boxplot - Soft Skills 1",
        ylab = "Score",
        col = "lightpink")
```

```r
############################################################
# 5.2 BIVARIATE DESCRIPTIVE STATISTICS
# Objective:
# Study the relationship between two variables without
# making formal statistical inferences yet.
############################################################

############################
# 5.2.1 QUALITATIVE × QUALITATIVE
############################

# Relationship between Major and Level
table_major_level <- table(data_final$Filiere, data_final$Niveau)
table_major_level

# Grouped bar plot
barplot(table_major_level,
        beside = TRUE,
        main = "Major / Level Distribution",
        xlab = "Level",
        ylab = "Count",
        legend = TRUE,
        col = c("lightblue", "lightgreen"))
```

```r
# Relationship between Major and Participation in Activities
table_major_participation <- table(
  data_final$Filiere,
  data_final$participation_activites
)
table_major_participation

barplot(table_major_participation,
        beside = TRUE,
        main = "Major vs Activity Participation",
        xlab = "Participation",
        ylab = "Count",
        legend = TRUE,
        col = c("lightcoral", "lightgray"))
```

```
############################
# 5.2.2 QUALITATIVE × QUANTITATIVE
############################

# Compare Hard Skills score by Level
aggregate(hard_sk_1 ~ Niveau,
          data = data_final,
          mean)

# Comparative boxplot
boxplot(hard_sk_1 ~ Niveau,
        data = data_final,
        main = "Hard Skills by Level",
        xlab = "Level",
        ylab = "Hard Skills Score",
        col = c("lightyellow", "lightblue"))


# Compare Soft Skills score by Major
aggregate(soft_sk_1 ~ Filiere,
          data = data_final,
          mean)

boxplot(soft_sk_1 ~ Filiere,
        data = data_final,
        main = "Soft Skills by Major",
        xlab = "Major",
        ylab = "Soft Skills Score",
        col = c("lightgreen", "lightpink"))
```

```
##############################
# 5.2.3 QUANTITATIVE × QUANTITATIVE
##############################

# Relationship between Age and Hard Skills
plot(data_final$Age, data_final$hard_sk_1,
     main = "Relationship between Age and Hard Skills",
     xlab = "Age",
     ylab = "Hard Skills Score",
     pch = 19)

# Add trend line
abline(lm(hard_sk_1 ~ Age, data = data_final),
       col = "red")
```

3.6.3 Hypothesis Testing

Hypothesis testing was conducted to assess students' perceptions and to examine relationships between variables using inferential statistical methods.

Global Hypothesis Testing for Likert Items

A global hypothesis testing procedure was applied to all Likert-scale items related to hard skills, soft skills, professional experience, networking, and extracurricular activities. For each item, the percentage of neutral responses was first evaluated. When the proportion of neutral responses exceeded 20%, no statistical decision was made. Otherwise, a chi-square goodness-of-fit test was performed under the assumption of a uniform distribution.

Agreement levels (scores 4 and 5) were compared to disagreement levels (scores 1 and 2). Based on the test results and the direction of responses, hypotheses were either accepted, rejected, or considered non-significant.

Comparative Tests on Categorical Variables

Comparative hypothesis tests were carried out between categorical variables. Contingency tables were constructed to analyze relationships such as academic major and hard skills perception. When expected cell counts were sufficient, the chi-square test was used; otherwise, Fisher's exact test was applied. Decisions were made based on a 5% significance level.

Tests Involving Quantitative Variables

Inferential tests involving quantitative variables were also conducted. Analysis of Variance (ANOVA) was used to compare mean hard skills scores across academic levels. In addition, the relationship between age and hard skills was examined using Spearman's rank correlation coefficient to account for the ordinal nature of Likert-scale data.

3.6.4 Factor Analysis

Exploratory factor analysis was performed to identify underlying latent dimensions among the Likert-scale items. Prior to factor extraction, the suitability of the data was assessed using the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's test of sphericity. Eigenvalues were examined using the Kaiser criterion (eigenvalues greater than one) to determine the number of factors to retain.

Principal Component Analysis with varimax rotation was applied to improve interpretability. The resulting factor structure revealed coherent groupings of items corresponding to the main conceptual dimensions of the study, namely hard skills, soft skills, professional experience, professional networking, and extracurricular activities.

3.6.5 Regression Analysis

A multiple linear regression model was developed to examine the determinants of hard skills. The dependent variable was defined as the mean hard skills score, while the independent variables included the mean scores of soft skills, professional experience, professional networking, and extracurricular activities.

The regression results were interpreted using p-values to identify statistically significant predictors, and the coefficient of determination (R^2) was used to evaluate the proportion of variance in hard skills explained by the model. This analysis provided insights into the factors that most strongly influence students' technical skill development.

```
##############################################################################
# 5.3 : HYPOTHESIS TESTING
##############################################################################

cat("========================================\n")
cat("5.3.1: GENERAL HYPOTHESIS TESTING\n")
cat("========================================\n\n")

# =====================
# a) INTRODUCTION
# =====================
cat("Introduction: Verify hypotheses for Likert items and numeric/categorical variables.\n\n")

# =====================
# b) GLOBAL TEST FOR LIKERT ITEMS
# =====================

cat("--------------------------------------------------\n")
cat(" Global test for perception items (Likert)\n")
cat("--------------------------------------------------\n\n")

# List of all Likert items
items_likert <- c(
  "hard_sk_1","hard_sk_2","hard_sk_3","hard_sk_4",
  "soft_sk_1","soft_sk_2","soft_sk_3","soft_sk_4",
  "prof_exp_1","prof_exp_2","prof_exp_3","prof_exp_4",
  "prof_net_1","prof_net_2","prof_net_3","prof_net_4",
  "extr_act_1","extr_act_2","extr_act_3","extr_act_4"
)
```

```
# Function to test each item
tester_item <- function(df, item) {

  responses <- df[[item]]

  # Count categories
  counts <- table(responses)

  # Percentage neutral
  pct_neutral <- ifelse("3" %in% names(counts), counts["3"]/sum(counts)*100, 0)

  # Check if Neutral > 20%
  if(pct_neutral > 20) {
    verdict <- "Neutral > 20%: No decision possible"
  } else {
    # Chi-square test: H0 = uniform distribution
    test <- chisq.test(counts)

    # Check if agreement (4+5) > disagreement (1+2)
    agree <- sum(responses %in% c(4,5))
    disagree <- sum(responses %in% c(1,2))

    if(test$p.value < 0.05 & agree > disagree) {
      verdict <- "Hypothesis accepted (Agreement > Disagreement)"
    } else if(test$p.value < 0.05 & disagree > agree) {
      verdict <- "Hypothesis rejected (Disagreement > Agreement)"
    } else {
      verdict <- "No significant difference"
    }
  }
```

```r
  cat("Item:", item, "\n")
  cat("  % Neutral =", round(pct_neutral,1), "%\n")
  cat("  Verdict:", verdict, "\n\n")
}

# Apply to all items
for(item in items_likert) {
  tester_item(data_final, item)
}

# =====================
# 5.3.2 COMPARATIVE TESTS ON CATEGORICAL VARIABLES
# =====================
cat("--------------------------------------------------\n")
cat(" Comparative tests on categorical variables\n")
cat("--------------------------------------------------\n\n")

# Example: Hard Skills 1 by Major
tab <- table(data_final$Filiere, data_final$hard_sk_1)
cat("Hard Skills 1 by Major:\n")
print(tab)
cat("\n")

# Check for low counts
if(any(tab < 5)) {
  cat("  > Some cells have counts <5, using Fisher's exact test\n")
  test_result <- fisher.test(tab)
} else {
  cat("  > Counts sufficient, using Chi-square test\n")
  test_result <- chisq.test(tab)
```

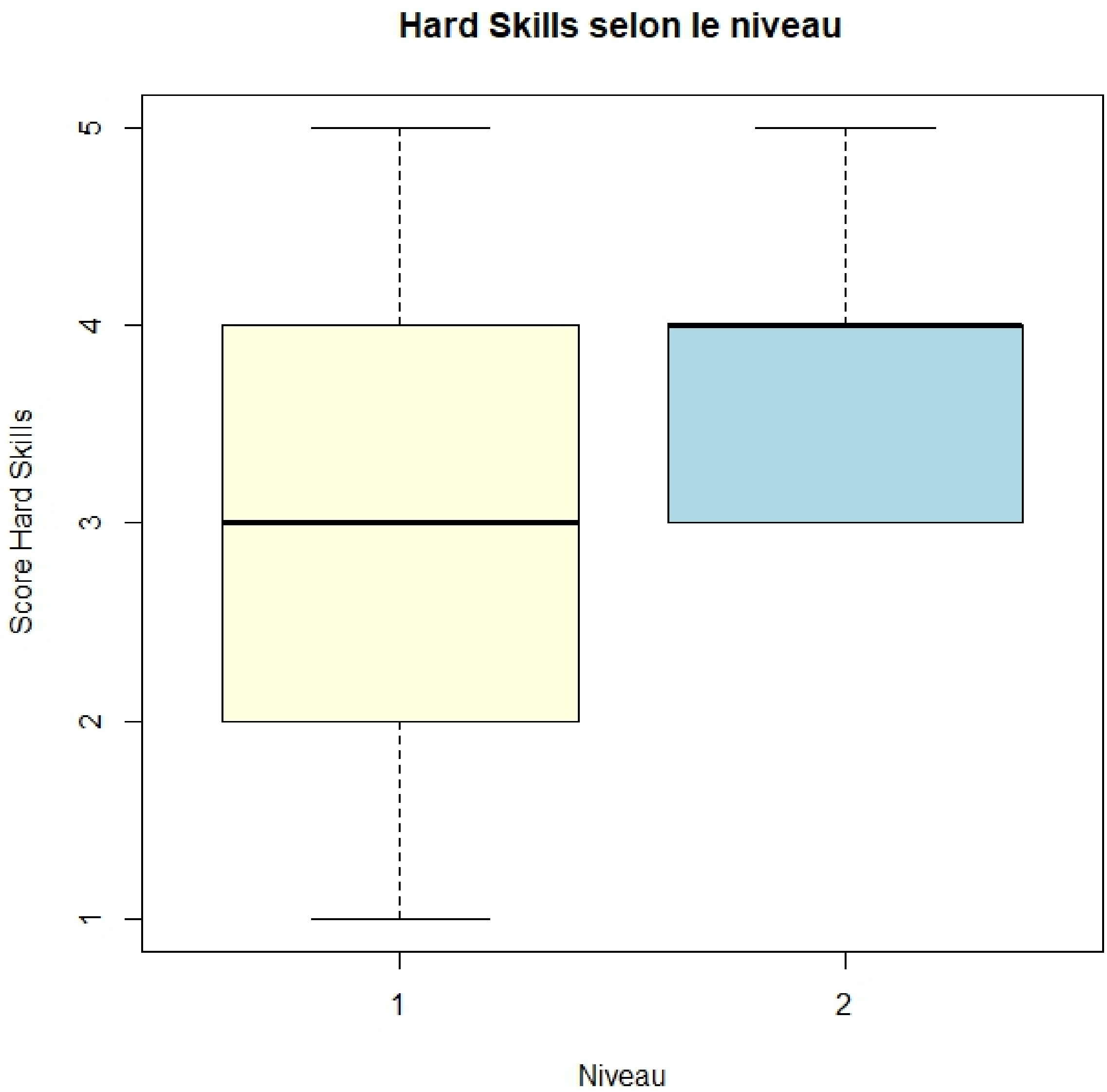
```r
if(any(tab < 5)) {
  cat("  > Some cells have counts <5, using Fisher's exact test\n")
  test_result <- fisher.test(tab)
} else {
  cat("  > Counts sufficient, using Chi-square test\n")
  test_result <- chisq.test(tab)
}

cat("p-value =", sprintf("%.5f", test_result$p.value), "\n")

# Decision
if(test_result$p.value < 0.05) {
  cat("  ==> Hypothesis: REJECTED (significant difference)\n")
} else {
  cat("  ==> Hypothesis: ACCEPTED (no significant difference)\n")
}

# =====================
# 5.3.3 NUMERIC VARIABLE TESTS
# =====================
cat("--------------------------------------------------\n")
cat(" Tests on numeric variables\n")
cat("--------------------------------------------------\n\n")

# Compare mean Hard Skills by Level (ANOVA)
anova_result <- aov(hard_sk_1 ~ Niveau, data = data_final)
cat("ANOVA Hard Skills 1 by Level:\n")
print(summary(anova_result))
cat("\n")
```

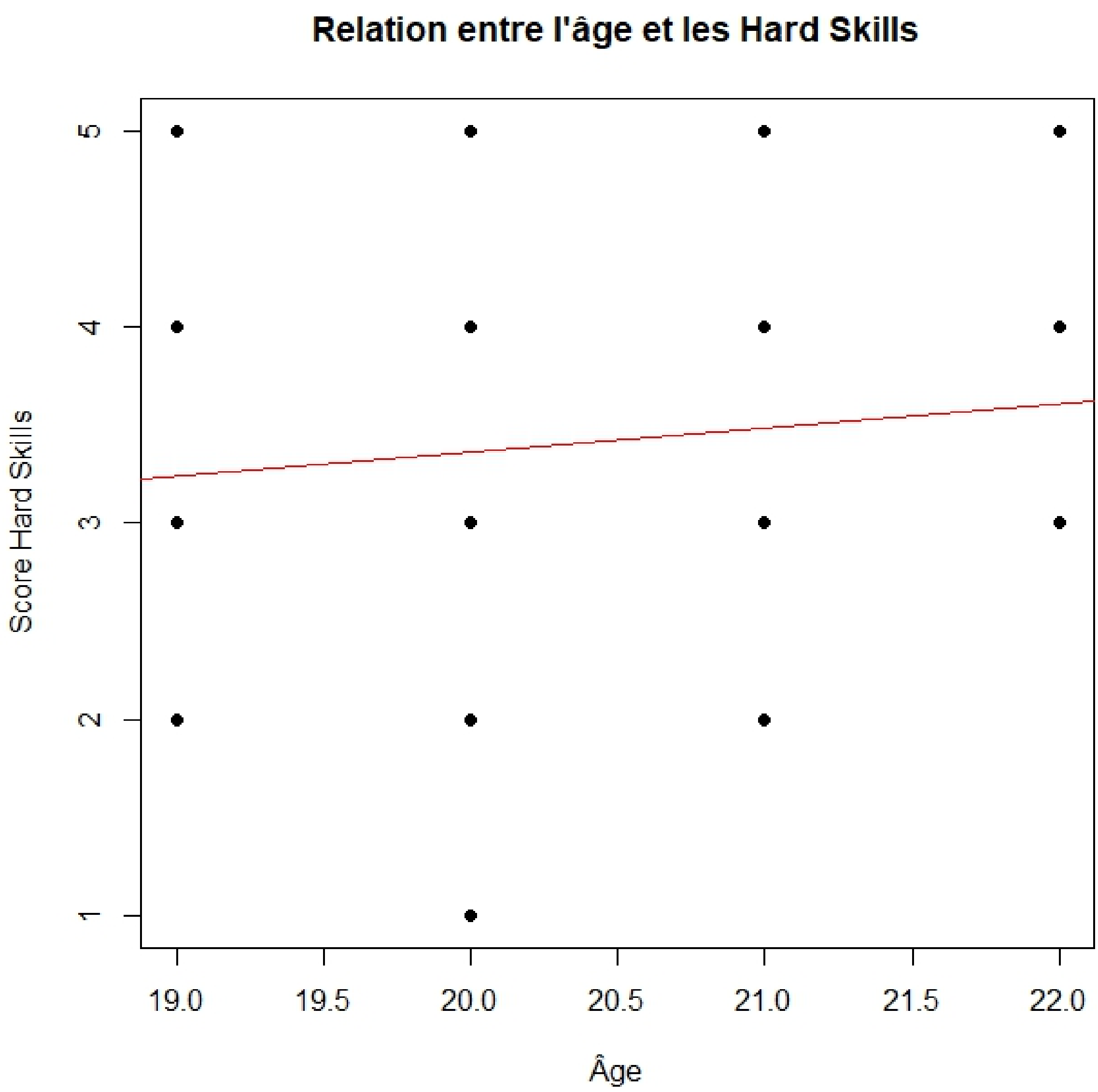
```
Item: extr_act_2
  % Neutral = 5.6 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)

Item: extr_act_3
  % Neutral = 9.9 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)

Item: extr_act_4
  % Neutral = 11.3 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)

> # Example: Hard Skills 1 by Major
> tab <- table(data_final$Filiere, data_final$hard_sk_1)
> print(tab)

     1  2  3  4  5
  1  1  8 12 18  8
  2  2  2  7 12  1
```

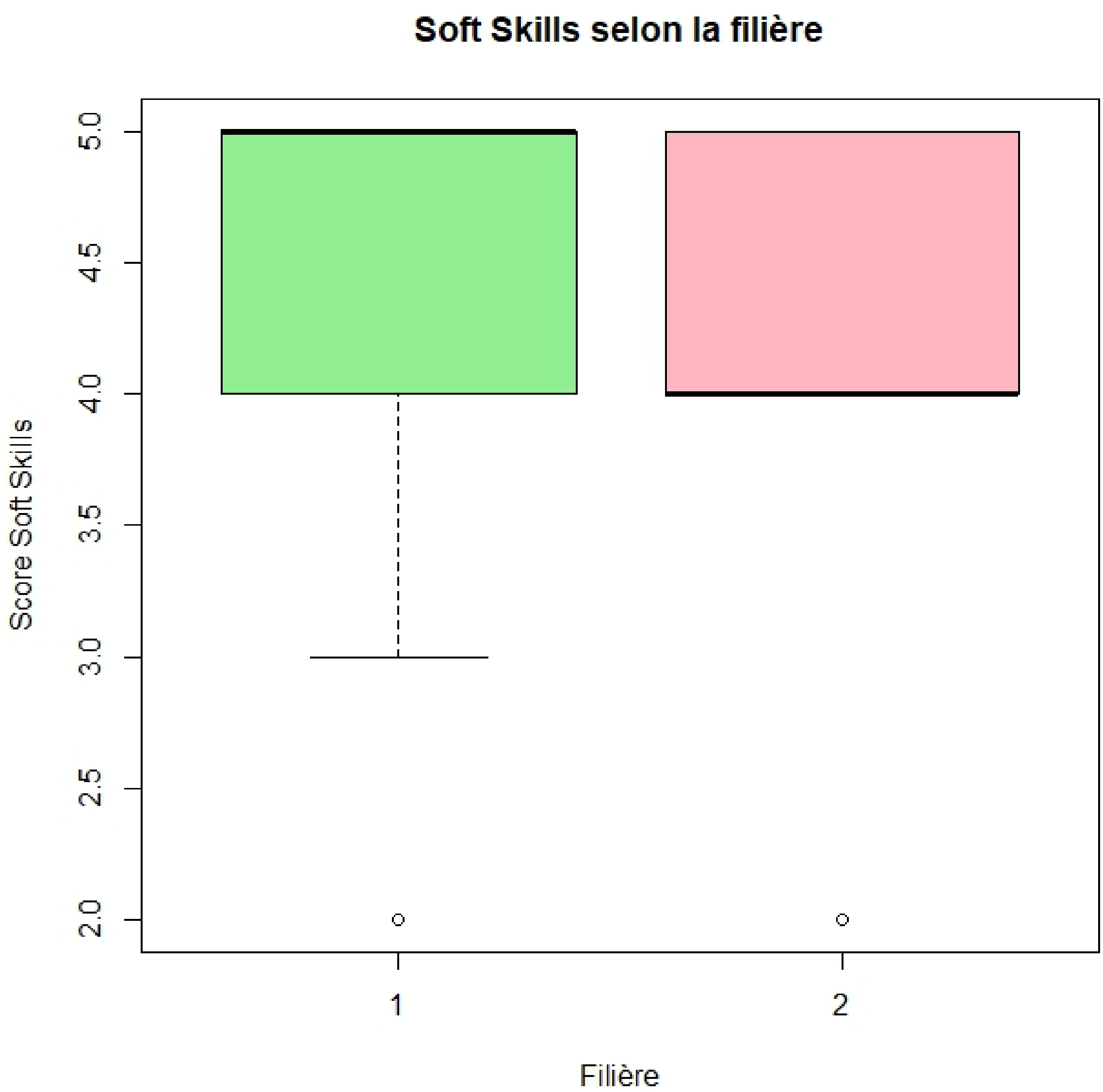
```
Item: prof_exp_1
  % Neutral = 0 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)

Item: prof_exp_2
  % Neutral = 1.4 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)

Item: prof_exp_3
  % Neutral = 7 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)

Item: prof_exp_4
  % Neutral = 4.2 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)

Item: prof_net_1
  % Neutral = 7 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)

Item: prof_net_2
  % Neutral = 8.5 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)

Item: prof_net_3
  % Neutral = 2.8 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)

Item: prof_net_4
  % Neutral = 12.7 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)

Item: extr_act_1
  % Neutral = 14.1 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)
```

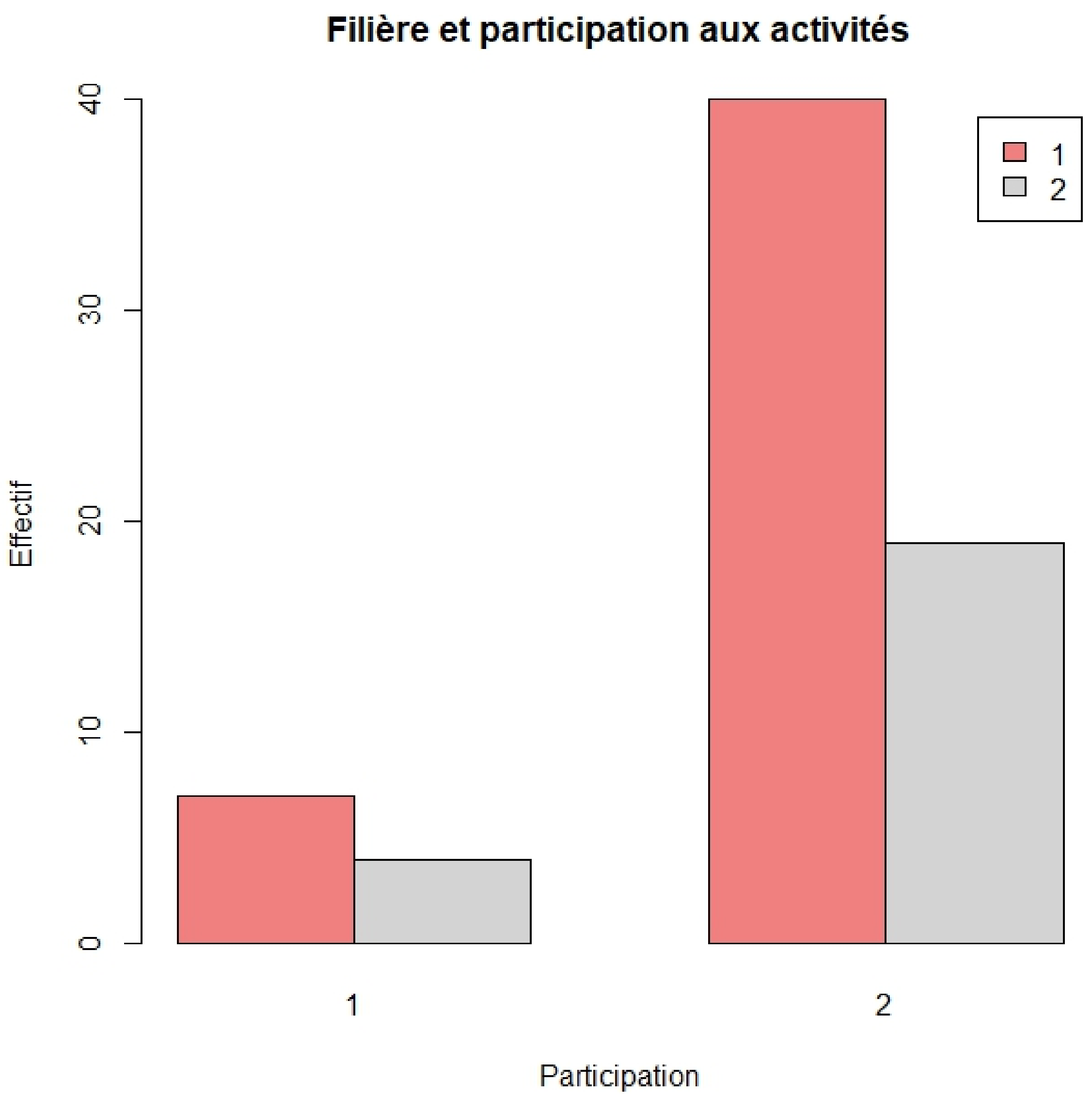
```
> for(item in items_likert) {
+   tester_item(data_final, item)
+ }
Item: hard_sk_1
  % Neutral = 26.8 %
  Verdict: Neutral > 20%: No decision possible

Item: hard_sk_2
  % Neutral = 5.6 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)

Item: hard_sk_3
  % Neutral = 12.7 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)

Item: hard_sk_4
  % Neutral = 1.4 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)

Item: soft_sk_1
  % Neutral = 2.8 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)

Item: soft_sk_2
  % Neutral = 5.6 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)

Item: soft_sk_3
  % Neutral = 5.6 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)

Item: soft_sk_4
  % Neutral = 7 %
  Verdict: Hypothesis accepted (Agreement > Disagreement)
```

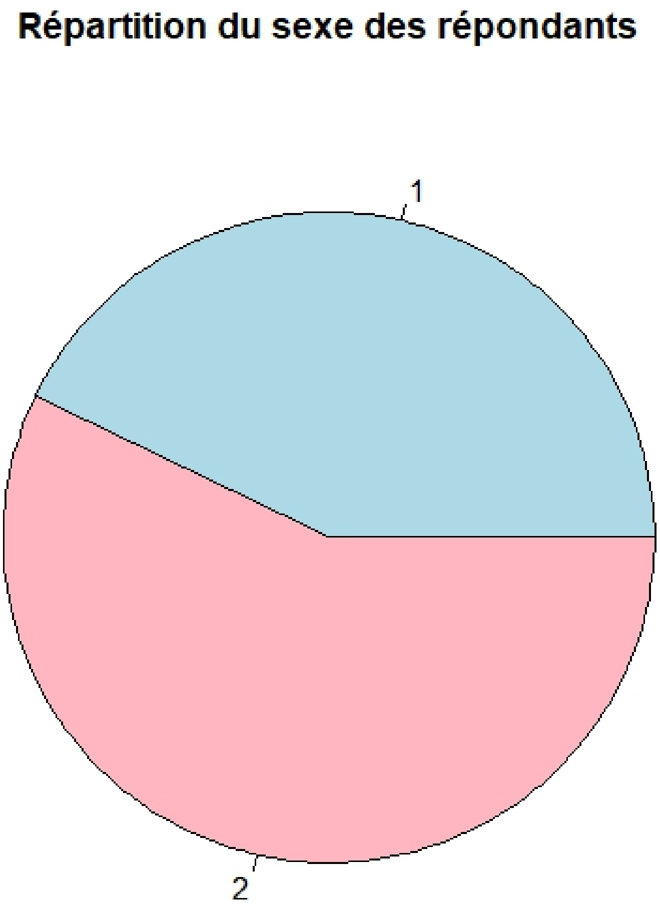
```
> # --- Age ---
>
> # Descriptive statistics
> summary(data_final$Age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  19.00   20.00   20.00   20.41   21.00   22.00       3
> mean(data_final$Age, na.rm = TRUE)
[1] 20.41176
> sd(data_final$Age, na.rm = TRUE)
[1] 0.8679242
```

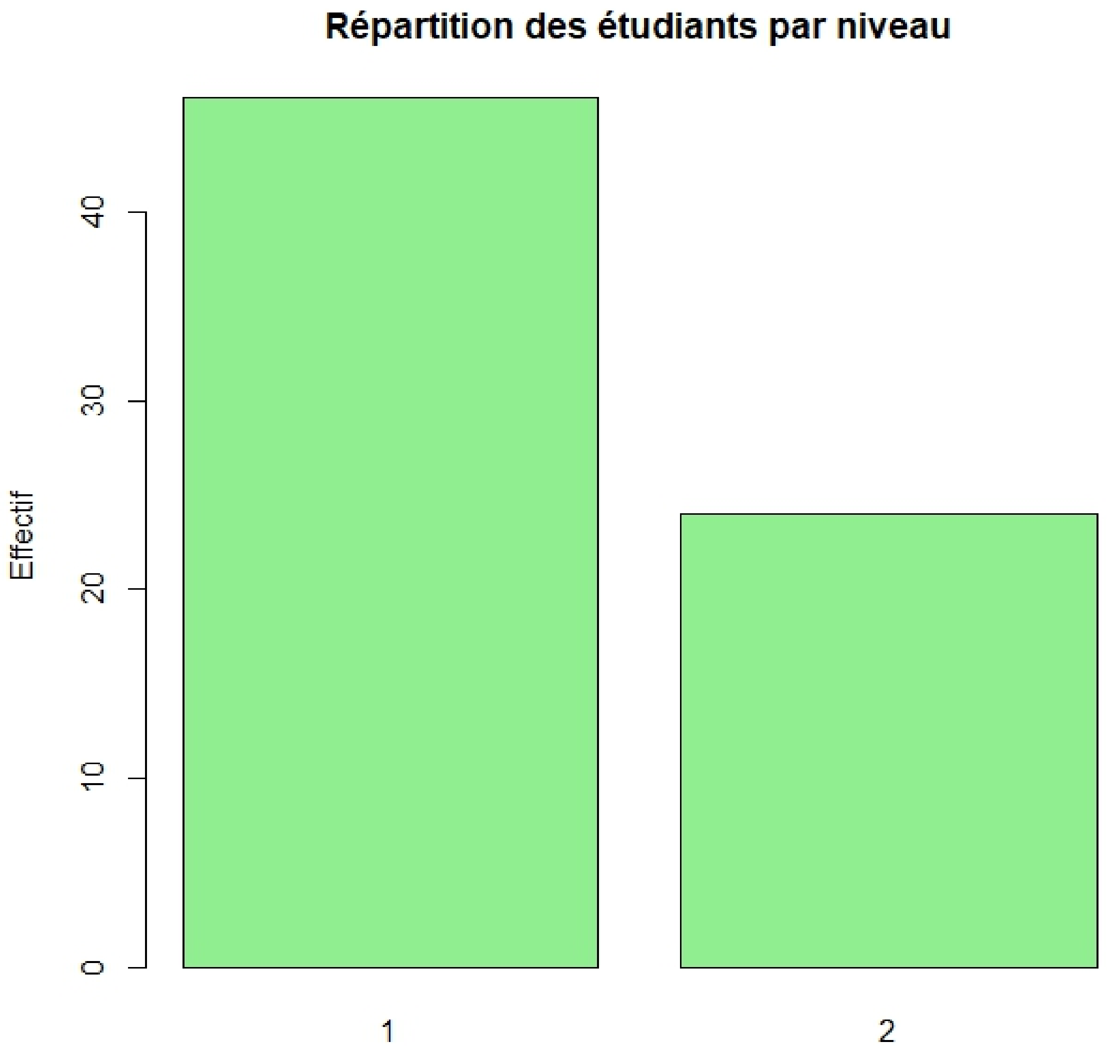

Hard Skills selon le niveau
Score Hard Skills
1
2
3
4
5
1
2
Niveau


Relation entre l'âge et les Hard Skills
Score Hard Skills
1
2
3
4
5
19.0
19.5
20.0
20.5
21.0
21.5
22.0
Âge


Soft Skills selon la filière
Score Soft Skills
2.0
2.5
3.0
3.5
4.0
4.5
5.0
1
2
Filière


Filière et participation aux activités
Effectif
0
10
20
30
40
1
2
Participation
1
2


Répartition du sexe des répondants
1
2



3.7 Discussion

The quantitative analysis conducted in this study provides valuable insights into the factors influencing employability. Using statistical methods implemented in *R*, two main hypotheses were empirically examined and supported.

First, the findings highlight that professional success strongly depends on both technical and transferable skills, as well as the ability to solve problems efficiently. However, the development of these competencies is frequently constrained by several challenges, including insufficient practical experience, rapid technological changes, intense labor market competition, and the financial burden associated with obtaining professional certifications.

Second, the results indicate that employability can be significantly enhanced through targeted strategies such as participation in training programs that prioritize practical internships, the development of transferable skills, and the promotion of an international perspective. These elements play a critical role in facilitating successful integration into the labor market.

Overall, by addressing these challenges and adopting effective employability strategies, engineering students at ENSA Kenitra can better prepare themselves to meet the evolving requirements of today's professional environment.

3.8 General Conclusion

This mixed-method study, combining qualitative and quantitative approaches, provides a comprehensive understanding of employability among engineering students at ENSA Kenitra. The results highlight the intricate relationship between skills, existing challenges, and effective strategies required to succeed in today's dynamic labor market.

The findings confirm that employability is strongly dependent on both technical competencies and transferable skills, including communication, teamwork, and problem-solving abilities. Nevertheless, the acquisition and development of these skills are often limited by several barriers, notably a lack of practical experience, rapid technological evolution, intense labor market competition, and the high financial cost associated with professional certifications.

To overcome these challenges, practical internships emerge as a vital link between academic knowledge and professional requirements. In parallel, the reinforcement of transferable skills enables graduates to perform effectively across diverse sectors. Moreover, fostering an international outlook through global exposure significantly enhances adaptability within an increasingly interconnected professional environment.

Ultimately, higher education institutions play a key role in improving employability by integrating practical training, skills development, and international engagement into their curricula. At the same time, employers can contribute to this ecosystem by providing structured internships and continuous upskilling opportunities, thereby supporting a more effective transition from education to employment.

3.9 General Conclusion

This mixed-method study, combining qualitative and quantitative approaches, provides a comprehensive understanding of employability among engineering students at ENSA

Kenitra. The results highlight the intricate relationship between skills, existing challenges, and effective strategies required to succeed in today's dynamic labor market.

The findings confirm that employability is strongly dependent on both technical competencies and transferable skills, including communication, teamwork, and problem-solving abilities. Nevertheless, the acquisition and development of these skills are often limited by several barriers, notably a lack of practical experience, rapid technological evolution, intense labor market competition, and the high financial cost associated with professional certifications.

To overcome these challenges, practical internships emerge as a vital link between academic knowledge and professional requirements. In parallel, the reinforcement of transferable skills enables graduates to perform effectively across diverse sectors. Moreover, fostering an international outlook through global exposure significantly enhances adaptability within an increasingly interconnected professional environment.

Ultimately, higher education institutions play a key role in improving employability by integrating practical training, skills development, and international engagement into their curricula. At the same time, employers can contribute to this ecosystem by providing structured internships and continuous upskilling opportunities, thereby supporting a more effective transition from education to employment.